

Package ‘r2redux’

March 23, 2022

Title R2 statistic

Version 1.0.3

Authors Sang Hong Lee (hong.lee@unisa.edu.au) and Md Moksedul Momin (momin@cvasu.ac.bd)

Description R2 statistic for significance test.

License `use_gpl3_license()`

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2.9000

R topics documented:

olkin_beta1_2	1
olkin_beta_inf	2
r2_diff	3
r2_enrich	4
r2_enrich_beta	6
r2_var	7
Index	10

olkin_beta1_2	<i>olkin_beta1_2 function</i>
---------------	-------------------------------

Description

This function derives Information matrix for β_1^2 and β_2^2 where β_1 and β_2 are regression coefficients from a multiple regression model, i.e. $y = x_1 \cdot \beta_1 + x_2 \cdot \beta_2 + e$, where y , x_1 and x_2 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. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	sample size

References

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

Examples

```
olkin_beta1_2 (omat,nv) (see example file)
dat=read.table("test_ukbb_thresholds_scaled") (see example files)
omat=cor(dat)[1:3,1:3]
omat
1.000 0.195 0.197
0.195 1.000 0.998
0.197 0.998 1.000
```

olkin_beta_inf	<i>olkin_beta_inf function</i>
----------------	--------------------------------

Description

This function derives Information matrix for beta1 and beta2 where beta1 and 2 are regression coefficients from a multiple regression model, i.e. $y = x1 \cdot \text{beta1} + x2 \cdot \text{beta2} + e$, 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. $\text{omat} = \text{cor}(\text{dat})$ where dat is N by 3 matrix having variables in the order of cbind(y,x1,x2)
nv	sample size

References

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

Examples

```
olkin_beta_inf (omat,nv) (see example file)
dat=read.table("test_ukbb_thresholds_scaled") (see example files)
omat=cor(dat)[1:3,1:3]
omat
1.000 0.195 0.197
0.195 1.000 0.998
0.197 0.998 1.000
```

r2_diff

*r2_diff function***Description**

This function estimates $\text{var}(R2(y \sim x[,v1]) - R2(y \sim x[,v2]))$ 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 i th 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

<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

Examples

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

```
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$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 $R2(y \sim x[,v1] + x[,v2])$ and $R2(y \sim x[,v2])$ (here $R2_1 = R2(y \sim x[,v1] + x[,v2])$ and $R2_2 = R2(y \sim x[,v2])$)

```
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$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
```

```
r2_enrich
```

Description

This function estimates $\text{var}(t1/(t1+t2))$ where $t1 = R2(y \sim x[,v1] + x[,v2]) - R2(y \sim x[,v1])$ and $t2 = R2(y \sim x[,v1] + x[,v2]) - R2(y \sim x[,v2])$ 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 i th column in the x matrix ($v1$ or $v2$ should be a single interger between $1 - M$, see Arguments below)

Usage

```
r2_enrich(dat, v1, v2, nv, expl)
```

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. # SNPs for the genomic region of interest/total # 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

Examples

```

To get test statistics for the ratio, i.e.  $t1/(t1+t2)$ .
t1 = R2(y~x[,v1]+x[,v2]) - R2(y~x[,v1]) and
t2 = R2(y~x[,v1]+x[,v2]) - R2(y~x[,v2])
(here we define R2_1=R2(y~x[,v1]), R2_2=R2(y~x[,v2]), R2_12=R2(y~x[,v1]+x[,v2]))

dat=read.table("test_ukbb_enrichment_choles") (see example file)
nv=length(dat$V1)
v1=c(1)
v2=c(2)
expected_ratio=0.04 (# SNPs for the regulatory/total # SNPs)
output=r2_enrich(dat,v1,v2,nv,expected_ratio)
output

r2redux output

output$var1 (variance of R2_1)
8.758455e-05

output$var2 (variance of R2_2)
7.36385e-05

output$var12 (variance of R2_12)
0.000102236

output$var_diff1_2 (var of difference of R2(y~x[,v1]) - R2(y~x[,v2]))
6.074567e-05

output$var_diff12_1 (var of difference of R2(y~x[,v1]+x[,v2]) - R2(y~x[,v1]))
1.184853e-05

output$var_diff12_2 (var of difference of R2(y~x[,v1]+x[,v2]) - R2(y~x[,v2]))
2.650564e-05

output$mean_diff12_1 (difference of R2(y~x[,v1]+x[,v2]) - R2(y~x[,v1]))
0.003048595

output$mean_diff12_2 (difference of R2(y~x[,v1]+x[,v2]) - R2(y~x[,v2]))
0.006845484

output$ratio (ratio = t1/(t1+t2))
0.6918768

output$ratio_var (variance of ratio, var(t1/(t1+t2)))
0.1324076

```

```

output$enrich_p (p-value for testing the ratio significantly different
from the expectation (exp1))
0.07321821

output$upper_ratio (upper limit of 95% CI for the ratio)
1.405079

output$lower_ratio (lower limit of 95% CI for the ratio)
-0.02132515

```

r2_enrich_beta	<i>r2_enrich_beta</i>
----------------	-----------------------

Description

This function estimates $\text{var}(t1/(t1+t2))$ where $t1 = \text{beta1}^2$ and $t2 = \text{beta2}^2$, and beta1 and beta2 are regression coefficients from a multiple regression model, i.e. $y = x1 \cdot \text{beta1} + x2 \cdot \text{beta2} + e$, 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 i th explanatory variables. $x2$ is N by 1 matrix having the j th explanatory variables. $v1$ and $v2$ indicates the i th and j th 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

<code>dat</code>	N by $(M+1)$ matrix having variables in the order of <code>cbind(y,x)</code>
<code>nv</code>	sample size
<code>exp1</code>	The expectation of the ratio (e.g. ratio of # SNPs in genomic partitioning)
<code>v1/v2</code>	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/(t1+t2)), where  $t1 = \text{beta1}^2$  and  $t2 = \text{beta2}^2$ .
 $\text{beta1}$  and  $\text{beta2}$  are regression coefficients from a multiple regression model,
i.e.  $y = x1 \cdot \text{beta1} + x2 \cdot \text{beta2} + e$ , 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$var1 (variance of t1)
7.072931e-05

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

output$var1_2 (difference between t1 and t2)
0.000162113

output$beta1_sq (t1)
0.01118301

output$beta2_sq (t2)
0.004980285

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

output$ratio (ratio = t1/(t1+t2_2))
0.6918768

output$ratio_var (variance of ratio, var(t1/(t1+t2)))
0.1324076

output$enrich_p (p-value for testing the ratio significantly different from
the expectation (exp1))
0.07321821

output$upper_ratio (upper limit of 95% CI for the ratio)
1.405079

output$lower_ratio (lower limit of 95% CI for the ratio)
-0.02132515

```

r2_var

r2_var function

Description

This function estimates $\text{var}(R^2(y \sim x[,v1]))$ where R^2 is the R squared value of the model, where R^2 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 i th 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 $R^2(y \sim x[,v1])$

```
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$var (variance of R2)
0.0001437583
```

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

```
output$mean_r2 (R2)
0.03836254
```

```
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 $R^2(y \sim x[,v1]+x[,v2]+x[,v3])$

```
dat=read.table("test_ukbb_thresholds_scaled") (see example file)
nv=length(dat$V1)
v1=c(1,2,3) ##1=R2(y~x[,v1]+x[,v2]+x[,v3])
outout=r2_var(dat,v1,nv)
output
```

r2redux output

```
output$var (variance of R2)
0.0001499374
```

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

```
output$mean_r2 (R2)
0.03917668
```

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

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


r2_var

9

0.01821657

Index

- *Topic **R2**
 - r2_diff, 3
 - r2_enrich, 4
 - r2_var, 7
- *Topic **a**
 - r2_enrich_beta, 6
- *Topic **beta^2**
 - r2_enrich_beta, 6
- *Topic **between**
 - r2_enrich, 4
 - r2_enrich_beta, 6
- *Topic **context**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
- *Topic **correlation**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
- *Topic **from**
 - r2_enrich_beta, 6
- *Topic **information**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
 - r2_diff, 3
 - r2_var, 7
- *Topic **in**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
- *Topic **matrix**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
 - r2_diff, 3
 - r2_var, 7
- *Topic **multiple**
 - r2_enrich_beta, 6
- *Topic **of**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
 - r2_enrich, 4
 - r2_enrich_beta, 6
- *Topic **ratio**
 - r2_enrich, 4
 - r2_enrich_beta, 6
- *Topic **regression**
 - r2_enrich_beta, 6
- *Topic **the**
 - olkin_beta1_2, 1
 - olkin_beta_inf, 2
- *Topic **variance**
 - r2_diff, 3
 - r2_enrich, 4
 - r2_enrich_beta, 6
 - r2_var, 7
- olkin_beta1_2, 1
- olkin_beta_inf, 2
- r2_diff, 3
- r2_enrich, 4
- r2_enrich_beta, 6
- r2_var, 7