

# Package ‘r2redux’

May 20, 2022

**Title** R2 statistic

**Version** 1.0.3

**Authors** Sang Hong Lee <hong.lee@unisa.edu.au> and Md Moksedul Momiin

**Maintainer** Moksedul Momiin <cvasu.momin@gmail.com>

**Description** R2 statistic for significance test.

**License** `use\_gpl3\_license()`

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.2

## R topics documented:

cc_trf . . . . .	1
olkin_beta1_2 . . . . .	2
olkin_beta_inf . . . . .	3
r2_diff . . . . .	4
r2_enrich . . . . .	6
r2_enrich_beta . . . . .	7
r2_var . . . . .	9

<b>Index</b>	<b>11</b>
--------------	-----------

---

cc_trf	<i>cc_trf function</i>
--------	------------------------

---

## 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$se1 (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	<i>olkin_beta1_2 function</i>
---------------	-------------------------------

---

**Description**

This function derives Information matrix for  $\beta_1^2$  and  $\beta_2^2$  where  $\beta_1$  and  $\beta_2$  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 $y$ , $x_1$ and $x_2$ , i.e. $\text{omat} = \text{cor}(\text{dat})$ where $\text{dat}$ is N by 3 matrix having variables in the order of $\text{cbind}(y, x_1, x_2)$
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  $\beta_1^2$  and  $\beta_2^2$  where  $\beta_1$  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  $\beta_1^2$ )
0.04146276

output$var2 (variance of  $\beta_2^2$ )
0.1611112

output$var1_2 (variance of difference between  $\beta_1^2$  and  $\beta_2^2$ )
0.03940878
```

---

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

---

## Description

This function derives Information matrix for  $\beta_1$  and  $\beta_2$  where  $\beta_1$  and 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 (see Olkin and Finn 1995).

## Usage

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

## Arguments

omat	3 by 3 matrix having the correlation coefficients between $y$ , $x_1$ and $x_2$ , i.e. $\text{omat}=\text{cor}(\text{dat})$ where $\text{dat}$ is N by 3 matrix having variables in the order of $\text{cbind}(y, x_1, x_2)$
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  $\beta_1^2$  and  $\beta_2^2$  where  $\beta_1$  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  $\beta_1^2$ )
0.2531406

output$var2 (variance of  $\beta_2^2$ )
0.2530269

output$var1_2 (variance of difference between  $\beta_1^2$  and  $\beta_2^2$ )
1.01141
```

---

r2\_diff

*r2\_diff function*


---

## Description

This function estimates  $\text{var}(R^2(y \sim x[,v1]) - R^2(y \sim x[,v2]))$  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.  $v1$  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

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  $R2(y \sim x[,v1])$  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  $R2\_2 = R2(y \sim x[,v1])$ )

```
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$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	<i>r2_enrich</i>
-----------	------------------

---

### 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 <code>cbind(y,x)</code>
nv	sample size
expl	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 (expl))
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

*r2\_enrich\_beta*


---

## Description

This function estimates  $\text{var}(t1/(t1+t2))$  where  $t1 = \text{beta1}^2$  and  $t2 = \text{beta2}^2$ , and  $\text{beta1}$  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).  $y$  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, expl)
```

**Arguments**

dat	N by (M+1) matrix having variables in the order of cbind(y,x)
nv	sample size
expl	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.  
 $\text{var}(t1/(t1+t2))$ , where  $t1 = \text{beta1}^2$  and  $t2 = \text{beta2}^2$ .  
 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

```
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))
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

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

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

output$var_diff (variance of difference, t1/exp - t2/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  $\text{var}(R^2(y \sim x[,v1]))$  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 <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

## 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)
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)
0.01821657

```

# Index

- \*Topic **R2**
  - cc\_trf, 1
  - r2\_diff, 4
  - r2\_enrich, 6
  - r2\_var, 9
- \*Topic **Transformation**
  - cc\_trf, 1
- \*Topic **and**
  - cc\_trf, 1
- \*Topic **a**
  - r2\_enrich\_beta, 7
- \*Topic **beta<sup>2</sup>**
  - r2\_enrich\_beta, 7
- \*Topic **between**
  - cc\_trf, 1
  - r2\_enrich, 6
  - r2\_enrich\_beta, 7
- \*Topic **context**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
- \*Topic **correlation**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
- \*Topic **from**
  - r2\_enrich\_beta, 7
- \*Topic **information**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
  - r2\_diff, 4
  - r2\_var, 9
- \*Topic **in**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
- \*Topic **liability**
  - cc\_trf, 1
- \*Topic **matrix**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
  - r2\_diff, 4
  - r2\_var, 9
- \*Topic **multiple**
  - r2\_enrich\_beta, 7
- \*Topic **observed**
  - cc\_trf, 1
- \*Topic **of**
  - cc\_trf, 1
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
  - r2\_enrich, 6
  - r2\_enrich\_beta, 7
- \*Topic **ratio**
  - r2\_enrich, 6
  - r2\_enrich\_beta, 7
- \*Topic **regression**
  - r2\_enrich\_beta, 7
- \*Topic **scale**
  - cc\_trf, 1
- \*Topic **the**
  - olkin\_beta1\_2, 2
  - olkin\_beta\_inf, 3
- \*Topic **variance**
  - r2\_diff, 4
  - r2\_enrich, 6
  - r2\_enrich\_beta, 7
  - r2\_var, 9
- cc\_trf, 1
- olkin\_beta1\_2, 2
- olkin\_beta\_inf, 3
- r2\_diff, 4
- r2\_enrich, 6
- r2\_enrich\_beta, 7
- r2\_var, 9