

# Package ‘r2redux’

June 16, 2022

**Title** R2 Statistic

**Version** 1.0.4

**Description** R2 statistic for significance test. Variance and covariance of R2 values used to assess the 95% CI and p-value of the R2 difference.

**License** GPL (>=3)

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.2

**NeedsCompilation** no

## R topics documented:

cc_trf . . . . .	1
olkin12_1 . . . . .	2
olkin12_13 . . . . .	3
olkin12_3 . . . . .	3
olkin12_34 . . . . .	4
olkin1_2 . . . . .	4
olkin_beta1_2 . . . . .	5
olkin_beta_inf . . . . .	6
r2_diff . . . . .	7
r2_enrich_beta . . . . .	9
r2_var . . . . .	10

<b>Index</b>	<b>13</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

**Value**

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

**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
```

---

olkin12\_1

*olkin12\_1 function*


---

**Description**

olkin12\_1 function

**Usage**

```
olkin12_1(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

**Value**

This function will be used as source code

---

olkin12_13	<i>olkin12_13 function</i>
------------	----------------------------

---

**Description**

olkin12\_13 function

**Usage**

```
olkin12_13(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

**Value**

This function will be used as source code

---

olkin12_3	<i>olkin12_3 function</i>
-----------	---------------------------

---

**Description**

olkin12\_3 function

**Usage**

```
olkin12_3(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

**Value**

This function will be used as source code

---

olkin12_34	<i>olkin12_34 function</i>
------------	----------------------------

---

### Description

olkin12\_34 function

### Usage

```
olkin12_34(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

### Value

This function will be used as source code

---

olkin1_2	<i>olkin1_2 function</i>
----------	--------------------------

---

### Description

olkin1\_2 function

### Usage

```
olkin1_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

### Value

This function will be used as source code

---

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

## Value

This function will give information (variance-covariance) matrix of  $\beta_1^2$  and  $\beta_2^2$ . Output from the command is the lists of outcomes.

## References

1. Olkin, I. and J.D. Finn, Correlations redux. Psychological Bulletin, 1995. 118(1): p. 155.
2. Momin, M.M., Lee, S. Wray, N. and S. Lee, S.H. Significance tests for  $R^2$  of out-of-sample prediction using polygenic scores. bioRxiv.

## Examples

To get information (variance-covariance) matrix of  $\beta_1^2$  and  $\beta_2^2$  where  $\beta_1$  and  $\beta_2$  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 beta2^2)
0.1611112
```

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

---

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

## Value

This function will generate information (variance-covariance) matrix of beta1 and beta2. Output from the command is the lists of outcomes.

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

<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

### Value

This function will estimate significant difference between two PGS (either dependent or independent and joint or single). Output from the command is the lists of outcomes.

### Examples

```

To get the test statistics for the difference between  $R^2(y \sim x[,v1])$  and
 $R^2(y \sim x[,v2])$ . (here we define  $R2\_1 = R^2(y \sim x[,v1])$  and  $R2\_2 = R^2(y \sim x[,v2])$ )

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$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  $R2(y \sim x[,v1] + x[,v2])$  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$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	<i>r2_enrich_beta</i>
----------------	-----------------------

---

## Description

This function estimates  $\text{var}((t1/exp) - (t2/(1-exp)))$ , where  $t1 = \beta_1^2$  and  $t2 = \beta_2^2$ , and  $\beta_1$  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 (see Olkin and Finn 1995).  $y$  is  $N$  by  $1$  matrix having the dependent variable, and  $x_1$  is  $N$  by  $1$  matrix having the  $i$ th explanatory variables.  $x_2$  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

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

## Value

This function will test the ratio which is significantly different from the expectation. Output from the command is the lists of outcomes.

## 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.  
 $\text{var}[(t1/\text{exp}) - (t2/(1-\text{exp}))]$ , 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.  $y = x1\hat{\alpha} + \hat{\alpha}\text{beta1} + x2\hat{\alpha} + \hat{\alpha}\text{beta2} + e$ , 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$beta1_sq (t1)
0.011118301
```

```
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
```

## 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

<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> , <code>v1=c(1,2)</code> or possibly with more values
<code>nv</code>	sample size

## Value

This function will test the null hypothesis for  $R^2$ . Output from the command is the lists of outcomes.

## 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$rsq (R2)
0.03836254

output$var (variance of R2)
0.0001437583

output$r2_based_p (P-value under the null hypothesis, i.e.  $R^2=0$ )
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  $R^2(y \sim x[,v1]+x[,v2]+x[,v3])$ 

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$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

- \*Topic **R2**
  - cc\_trf, 1
  - r2\_diff, 7
  - r2\_var, 10
- \*Topic **Transformation**
  - cc\_trf, 1
- \*Topic **and**
  - cc\_trf, 1
- \*Topic **a**
  - r2\_enrich\_beta, 9
- \*Topic **beta^2**
  - r2\_enrich\_beta, 9
- \*Topic **between**
  - cc\_trf, 1
  - r2\_enrich\_beta, 9
- \*Topic **context**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
- \*Topic **correlation**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
- \*Topic **from**
  - r2\_enrich\_beta, 9
- \*Topic **information**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
  - r2\_diff, 7
  - r2\_var, 10
- \*Topic **in**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
- \*Topic **liability**
  - cc\_trf, 1
- \*Topic **matrix**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
  - r2\_diff, 7
  - r2\_var, 10
- \*Topic **multiple**
  - r2\_enrich\_beta, 9
- \*Topic **observed**
  - cc\_trf, 1
- \*Topic **of**
  - cc\_trf, 1
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
  - r2\_enrich\_beta, 9
- \*Topic **ratio**
  - r2\_enrich\_beta, 9
- \*Topic **regression**
  - r2\_enrich\_beta, 9
- \*Topic **scale**
  - cc\_trf, 1
- \*Topic **source**
  - olkin12\_1, 2
  - olkin12\_13, 3
  - olkin12\_3, 3
  - olkin12\_34, 4
  - olkin1\_2, 4
- \*Topic **the**
  - olkin\_beta1\_2, 5
  - olkin\_beta\_inf, 6
- \*Topic **variance**
  - r2\_diff, 7
  - r2\_enrich\_beta, 9
  - r2\_var, 10
- cc\_trf, 1
- olkin12\_1, 2
- olkin12\_13, 3
- olkin12\_3, 3
- olkin12\_34, 4
- olkin1\_2, 4
- olkin\_beta1\_2, 5
- olkin\_beta\_inf, 6
- r2\_diff, 7
- r2\_enrich\_beta, 9
- r2\_var, 10