Package ‘cocor’

September 5, 2013

Type Package

Title Comparing correlations

Author Birk Diedenhofen <mail@birkdiedenhofen.de>

Maintainer Birk Diedenhofen <mail@birkdiedenhofen.de>

Depends methods

Suggests testthat

Enhances rkward

Description This package provides functions to compare two correlations based on either dependent or independent groups.

License GPL (>= 3)

Encoding UTF-8

LazyLoad yes

URL http://r.birkdiedenhofen.de/pckg/cocor/

Version 1.0-0

Date 2013-08-01

Collate '0helper.r' 'as.htest.r' 'cocor.dep.groups.nonoverlap.r' 'cocor.dep.groups.overlap.r' 'cocor.indep.groups.r' 'cocor-package.R' 'cocor.r'

R topics documented:

cocor-package *The cocor Package*

Description

Comparing correlations.

Details

Package:	cocor
Type:	Package
Version:	1.0-0
Date:	2013-08-01
Depends:	methods
Enhances:	rkward
Encoding:	UTF-8
License:	GPL (>= 3)
LazyLoad:	yes
URL:	http://r.birkdiedenhofen.de/pckg/cocor/

This package provides functions to compare two correlations based on either dependent or independent groups.

Author(s)

Birk Diedenhofen `<mail@birkdiedenhofen.de>`

aptitude *Sample dataset: aptitude*

Description

Data of two samples of testees who completed an aptitude test consisting of general knowledge questions, logic tasks, and two measures of intelligence.

Usage

```
data("aptitude")
```

Format

A list that contains two data.frames holding 291 and 334 observations on the following 4 variables.

`knowledge` Score achieved on the general knowledge questions (numeric vector)

`logic` Score achieved on the logic tasks (numeric vector)

`intelligence.a` Intelligence measure A (numeric vector)

`intelligence.b` Intelligence measure B (numeric vector)

Examples

```
data("aptitude")
```

---

as.htest *Convert to a list of class 'htest'*

Description

Convert a cocor result object of the classes 'cocor.indep.groups', 'cocor.dep.groups.overlap', and 'cocor.dep.groups.nonoverlap' to a list of class 'htest'.

Usage

```
as.htest(x)

## S4 method for the classes 'cocor.indep.groups', 'cocor.dep.groups.overlap', and 'cocor.dep.groups.
as.htest(x)
```

Arguments

`x`	A cocor result object of the classes 'cocor.indep.groups', 'cocor.dep.groups.overlap', and 'cocor.dep.groups.nonoverlap'.

Value

Returns a list containing a list of class 'htest' for the result of each test with the following elements:

`data.name`	A character string giving the names of the data.
`estimate`	The two correlations that have been compared and related correlations.
`method`	A character string indicating the test that was performed.
`null.value`	The specified hypothesized value of the difference between the two correlations.
`alternative`	A character string describing the alternative hypothesis.
`parameter`	The degrees of freedom of the distribution of the test statistic.
`statistic`	The value of the test statistic.
`p.value`	The p-value of the test.
`conf.int`	The confidence interval of the difference between the two correlations.

See Also

cocor, cocor.indep.groups, cocor.dep.groups.overlap, cocor.dep.groups.nonoverlap

Examples

```
data("aptitude")

cocor.result <- cocor(~knowledge + intelligence.a | logic + intelligence.a, aptitude$sample1)
as.htest(cocor.result)
```

---

cocor *Compare two correlations based on either dependent or independent groups*

Description

Performs a test of significance for the difference between two correlations based on either dependent or independent groups. Dependent correlations can be either overlapping (they share a variable) or nonoverlapping (they have no variable in common). The function expects raw data input from which the correlations are calculated.

Usage

```
cocor(formula, data, use = "everything",
  alternative = "two.sided", test = "all", alpha = 0.05,
  conf.level = 0.95, null.value = 0,
  return.htest = FALSE)
```

Arguments

`formula` A formula specifying the correlations and their underlying variables (See details).

`data` A list holding two data.frames/matrices for independent groups or a single data.frame/matrix for dependent groups that contain the variables specified in `formula` as columns.

`use` A character string giving a test for computing covariances in the presence of missing values. This must be (an abbreviation of) one of the strings "everything", "all.obs", "complete.obs", "na.or.complete", or "pairwise.complete.obs" (see cor in package stats).

`alternative` A character string specifying whether the alternative hypothesis is two-sided ("`two.sided`"; default) or one-sided ("`greater`" or "`less`", depending on the direction). Optionally, the initial letter of the character strings ("`t`", "`g`", and "`l`)" can be used.

`test` For the tests available, see cocor.indep.groups, cocor.dep.groups.overlap, and cocor.dep.groups.nonoverlap. Use `all` to apply all tests (default).

`alpha` A number defining the alpha level for the hypothesis test. The default value is .05.

`conf.level` A number defining the level of confidence for the confidence interval (if a test is used that calculates confidence intervals). The default value is .95.

`null.value` A number defining the hypothesized difference between the two correlations used for testing the null hypothesis. The default value is 0. If the value is other than 0, only the test `zou2007` that uses a confidence interval is available.

`return.htest` A logical indicating whether the result should be returned as a list containing a list of class 'htest' for each test. The default value is `FALSE`.

Details

Comparison of two correlations based on independent groups The `formula` parameter for the comparison of two correlations based on independent groups can either be `~a + b | a + b`, `~a + b | a + c`, or `~a + b | c + d`. The variables of the first correlation – `a` and `b` before the "`|`" character – must refer to columns in the data.frame/matrix of the first element in the list of the `data` object, whereas the variables of the second correlation – `a`, `b`, `c`, and `d` after the "`|`" character – must refer to columns in the data.frame/matrix of the second element in the list.

Comparison of two overlapping correlations based on dependent groups The `formula` parameterfor correlations based on dependent groups with overlapping variables must follow the pattern `~a + b | a + c`. The variables of the two correlation – `a`, `b`, and `c` – must refer to columns in the data.frame/matrix of the `data` object.

Comparison of two nonoverlapping correlations based on dependent groups The `formula` for correlations based on dependent groups with nonoverlapping variables must have the form `~a + b | c + d`. The variables of the two correlation – `a`, `b`, `c`, and `d` – must refer to columns in the data.frame/matrix of the `data` object.

Value

Returns an object of the class 'cocor.indep.groups', 'cocor.dep.groups.overlap', or 'cocor.dep.groups.nonoverlap' depending on the invoked comparison function.

See Also

cocor.indep.groups, cocor.dep.groups.overlap, cocor.dep.groups.nonoverlap, as.htest

Examples

```
data("aptitude")

# Compare two correlations based on two independet groups
cocor(~logic + intelligence.a | logic + intelligence.a, aptitude)

# Compare two correlations based on two depenendent groups
# The correlations are overlapping
cocor(~knowledge + intelligence.a | logic + intelligence.a, aptitude$sample1)
cocor(~knowledge + intelligence.a | logic + intelligence.a, aptitude$sample2)
# The correlations are nonoverlapping
cocor(~logic + intelligence.b | knowledge + intelligence.a, aptitude$sample1)
cocor(~logic + intelligence.b | knowledge + intelligence.a, aptitude$sample2)
```

```
# Return result as a list of class 'htest'
cocor(~knowledge + intelligence.b | logic + intelligence.a, aptitude$sample1, return.htest=TRUE)
```

---

cocor.dep.groups.nonoverlap

Compare two nonoverlapping correlations based on dependent groups

---

Description

Performs a test of significance for the difference between two correlations based on dependent groups (e.g., the same group). The two correlations are nonoverlapping, i.e., they have no variable in common. The function tests whether the correlations between j and k (r.jk) and between h and m (r.hm) differ in magnitude. Because the significance depends on the pairwise intercorrelations between all of the variables involved (j, k, h, and m), these intercorrelations have to be provided as additional parameters. The function expects correlation coefficients as input.

Usage

```
cocor.dep.groups.nonoverlap(r.jk, r.hm, r.jh, r.jm, r.kh,
  r.km, n, alternative = "two.sided", test = "all",
  alpha = 0.05, conf.level = 0.95, null.value = 0,
  data.name = NULL, var.labels = NULL,
  return.htest = FALSE)
```

Arguments

r.jk	A number specifying the correlation between j and k (this correlation is used for comparison)
r.hm	A number specifying the correlation between h and m (this correlation is used for comparison)
r.jh	A number specifying the correlation between j and h
r.jm	A number specifying the correlation between j and m
r.kh	A number specifying the correlation between k and h
r.km	A number specifying the correlation between k and m
n	An integer defining the size of the group
alternative	A character string specifying whether the alternative hypothesis is two-sided ("two.sided"; default) or one-sided ("greater" or "less", depending on the direction). Optionally, the initial letter of the character strings ("t", "g", and "l)" can be used.
test	A vector of character strings specifying the tests to be used (pearson1898, dunn1969, steiger1980, raghunathan1996, silver2004, or zou2007). Use all to apply all tests (default). For further information see the tests section below.

`alpha`	A number defining the alpha level for the hypothesis test. The default value is .05.
`conf.level`	A number defining the level of confidence for the confidence interval (if test `zou2007` is used). The default value is .95.
`null.value`	A number defining the hypothesized difference between the two correlations used for testing the null hypothesis. The default value is 0. If the value is other than 0, only the test `zou2007` that uses a confidence interval is available.
`data.name`	A character string giving the name(s) of the data. If `data.name` is `NULL`, the data names of `r.jk`, `r.hm`, `r.jh`, `r.jm`, `r.kh`, and `r.km` are used.
`var.labels`	A vector of 4 character strings specifying the labels for j, k, h, and m (in this order).
`return.htest`	A logical indicating whether the result should be returned as a list containing a list of class 'htest' for each test. The default value is `FALSE`.

Value

Returns an object of the class 'cocor.dep.groups.nonoverlap' with the following slots holding the input parameters described above:

`r.jk`	Input parameter
`r.hm`	Input parameter
`r.jh`	Input parameter
`r.jm`	Input parameter
`r.kh`	Input parameter
`r.km`	Input parameter
`n`	Input parameter
`alternative`	Input parameter
`alpha`	Input parameter
`conf.level`	Input parameter
`null.value`	Input parameter
`data.name`	Input parameter
`var.labels`	Input parameter

For each test a slot of the same name exists with a list containing the following elements:

`statistic`	The value of the test statistic (unless test `zou2007` is used).
`distribution`	The distribution of the test statistic (unless test `zou2007` is used).
`p.value`	The p-value of the test (unless test `zou2007` is used).
`conf.int`	The confidence interval of the difference between the two correlations (if test `zou2007` is used).

Tests

In the following, r_{jk} and r_{hm} are the two correlations that are being compared; Z_{jk} and Z_{hm} are their Z transformed equivalents. r_{jh}, r_{kh}, r_{jm}, and r_{km} are the related correlations that are also required. n specifies the size of the group the two correlations are based on. Some tests make use of Fisher's r-to-Z transformation (1921, p. 26):

$$Z = \frac{1}{2}(ln(1+r) - ln(1-r)).$$

pearson1898: *Pearson and Filon's (1898) z*

This test was proposed by Pearson and Filon (1898, p. 262, formula xl). The formula for the test statistic z is computed as

$$z = \frac{\sqrt{n}(r_{jk} - r_{hm})}{\sqrt{(1-r_{jk}^2)^2 + (1-r_{hm}^2)^2 - k}}$$

(Raghunathan, Rosenthal, and Rubin, 1996, p. 179, formula 1), where

$$k = (r_{jh} - r_{jk}r_{kh})(r_{km} - r_{kh}r_{hm}) + (r_{jm} - r_{jh}r_{hm})(r_{kh} - r_{jk}r_{jh})$$
$$+(r_{jh} - r_{jm}r_{hm})(r_{km} - r_{jk}r_{jm}) + (r_{jm} - r_{jk}r_{km})(r_{kh} - r_{km}r_{hm})$$

(Raghunathan et al. (1996, p. 179, formula 2). The two formulae can also be found in Steiger (1980, p. 245, formula 2 and p. 246, formula 5).

dunn1969: *Dunn and Clark's (1969) z*

The test statistic z of this test is calculated as

$$z = \frac{(Z_{jk} - Z_{hm})\sqrt{n-3}}{\sqrt{2-2c}}$$

(Dunn and Clark, 1969, p. 370, formula 15), where

$$c = \Big(\frac{1}{2}r_{jk}r_{hm}(r_{jh}^2 + r_{jm}^2 + r_{kh}^2 + r_{km}^2) + r_{jk}r_{hm} + r_{jm}r_{kh}$$
$$-(r_{jk}r_{jh}r_{jm} + r_{jk}r_{kh}r_{km} + r_{jh}r_{kh}r_{hm} + r_{jm}r_{km}r_{hm})\Big)$$
$$\Big/\Big((1-r_{jk}^2)(1-r_{hm}^2)\Big)$$

(Dunn and Clark, 1969, p. 368, formula 9).

steiger1980: *Steiger's (1980) modification of Dunn and Clark's (1969) z using average correlations*

This test was proposed by Steiger (1980) and is a modification of Dunn and Clark's (1969) z. Instead of r_{jk} and r_{hm} the mean of the two is being used. The test statistic z is given by

$$z = \frac{(Z_{jk} - Z_{hm})\sqrt{n-3}}{\sqrt{2-2c}}$$

(Steiger, 1980, p. 247, formula 15), where

$$\bar{r} = \frac{r_{jk} + r_{hm}}{2}$$

(Steiger, 1980, p. 247) and

$$c = \frac{\frac{1}{2}\bar{r}^2(r_{jh}^2 + r_{jm}^2 + r_{kh}^2 + r_{km}^2) + \bar{r}^2 + r_{jm}r_{kh} - (\bar{r}r_{jh}r_{jm} + \bar{r}r_{kh}r_{km} + r_{jh}r_{kh}\bar{r} + r_{jm}r_{km}\bar{r})}{(1-\bar{r}^2)^2}$$

(Steiger, 1980, p. 247, formula 11; in the original article, there are brackets missing around the divisor).

raghunathan1996: *Raghunathan, Rosenthal, and Rubin's (1996) modification of Pearson and Filon's (1898) z*

This test of Raghunathan, Rosenthal, and Rubin (1996) is based on Pearson and Filon's (1898) z. Unlike Pearson and Filon (1898), Raghunathan et al. (1996) use Z transformed correlation coefficients. The test statistic z is computed as

$$z = \sqrt{\frac{n-3}{2}} \frac{Z_{jk} - Z_{hm}}{\sqrt{1 - \frac{k}{2(1-r_{jk}^2)(1-r_{hm}^2)}}}$$

(Raghunathan et al., 1996, p. 179, formula 3), where

$$k = (r_{jh} - r_{jk}r_{kh})(r_{km} - r_{kh}r_{hm}) + (r_{jm} - r_{jh}r_{hm})(r_{kh} - r_{jk}r_{jh})$$

$$+(r_{jh} - r_{jm}r_{hm})(r_{km} - r_{jk}r_{jm}) + (r_{jm} - r_{jk}r_{km})(r_{kh} - r_{km}r_{hm})$$

(Raghunathan et al., 1996, p. 179, formula 2).

silver2004: *Silver, Hittner, and May's (2004) modification of Dunn and Clark's (1969) z using a backtransformed average Fisher's (1921) Z procedure*

The approach to backtransform averaged Fisher's (1921) Zs was first proposed in Silver and Dunlap (1987) and was applied to the comparison of nonoverlapping correlations by Silver et al. (2004). The test is based on Steiger's (1980) approach. The formula of the test statistic z is given by

$$z = \frac{(Z_{jk} - Z_{hm})\sqrt{n-3}}{\sqrt{2-2c}}$$

(Silver et al., 2004, p. 55, formula 5), where

$$c = \frac{\frac{1}{2}\bar{r}_z^2(r_{jh}^2 + r_{jm}^2 + r_{kh}^2 + r_{km}^2) + \bar{r}_z^2 + r_{jm}r_{kh} - (\bar{r}_z r_{jh}r_{jm} + \bar{r}_z r_{kh}r_{km} + r_{jh}r_{kh}\bar{r}_z + r_{jm}r_{km}\bar{r}_z)}{(1-\bar{r}_z^2)^2}$$

(Silver et al., 2004, p. 56),

$$\bar{r}_z = \frac{exp(2\bar{Z}-1)}{exp(2\bar{Z}+1)}$$

(Silver and Dunlap, 1987, p. 146, formula 4), and

$$\bar{Z} = \frac{Z_{jk} + Z_{hm}}{2}$$

(Silver et al., 2004, p. 55).

zou2007: *Zou's (2007) confidence interval*

This test calculates the confidence interval of the difference between the two correlations r_{jk} and r_{hm}. If the confidence interval includes zero, the null hypothesis that the two correlations are equal must be retained. If the confidence interval does not include zero, the null hypothesis has to be rejected. A lower and upper bound for the interval (L and U, respectively) is given by

$$L = r_{jk} - r_{hm} - \sqrt{(r_{jk} - l_1)^2 + (u_2 - r_{hm})^2 - 2c(r_{jk} - l_1)(u_2 - r_{hm})}$$

and

$$U = r_{jk} - r_{hm} - \sqrt{(u_1 - r_{jk})^2 + (r_{hm} - l_2)^2 - 2c(u_1 - r_{jk})(r_{hm} - l_2)}$$

(Zou, 2007, pp. 409-410), where

$$l = \frac{exp(2l') - 1}{exp(2l') + 1},$$

$$u = \frac{exp(2u') - 1}{exp(2u') + 1}$$

(Zou, 2007, p. 406),

$$c = \Big(\frac{1}{2} r_{jk} r_{hm} (r_{jh}^2 + r_{jm}^2 + r_{kh}^2 + r_{km}^2) + r_{jk} r_{hm} + r_{jm} r_{kh}$$
$$-(r_{jk} r_{jh} r_{jm} + r_{jk} r_{kh} r_{km} + r_{jh} r_{kh} r_{hm} + r_{jm} r_{km} r_{hm})\Big)$$
$$\Big/ \Big((1 - r_{jk}^2)(1 - r_{hm}^2)\Big)$$

(Zou, 2007, p. 409), and

$$l', u' = Z \pm z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n-3}}$$

(Zou, 2007, p. 406). α denotes the desired alpha level of the confidence interval.

References

Dunn, O. J., & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, *64*, 366-377. doi:10.2307/2283746

Pearson, K., & Filon, L. N. G. (1898). Mathematical contributions to theory of evolution: IV. On the probable errors of frequency constants and on the influence of random selection and correlation. *Philosophical Transactions of the Royal Society of London, Series A*, *191*, 229-311. doi:10.1098/rsta.1898.0007

Raghunathan, T. E., Rosenthal, R., & Rubin, D. B., (1996). Comparing correlated but nonoverlapping correlations. *Psychological Methods*, *1*, 178-183. doi:10.1037//1082-989X.1.2.178

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's Z transformation be used? *Journal of Applied Psychology*, *72*, 146-148. doi:10.1037//0021-9010.72.1.146

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing dependent correlations with nonoverlapping variables: A Monte Carlo simulation. *Journal of Experimental Education*, *73*, 53-69. doi:10.3200/JEXE.71.1.53-70

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251. doi:10.1037//0033-2909.87.2.245

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413. doi:10.1037/1082-989X.12.4.399

See Also

cocor, cocor.indep.groups, cocor.dep.groups.overlap, as.htest

Examples

```
# Compare the difference between the correlations (age, intelligence) and
# body mass (index, shoe size) measured in the same group (all values are fictional):
r.jk <- .2  # Correlation (age, intelligence)
r.hm <- .7  # Correlation (body mass index, shoe size)
r.jh <- .4  # Correlation (age, body mass index)
r.jm <- .5  # Correlation (age, shoe size)
r.kh <- .1  # Correlation (intelligence, body mass index)
r.km <- .3  # Correlation (intelligence, shoe size)
n <- 232  # Size of the group

cocor.dep.groups.nonoverlap(r.jk, r.hm, r.jh, r.jm, r.kh, r.km, n, var.labels=c("age", "intelligence", "body mas
```

---

cocor.dep.groups.overlap

Compare two overlapping correlations based on dependent groups

---

Description

Performs a test of significance for the difference between two correlations based on dependent groups (e.g., the same group). The two correlations are overlapping, i.e., they have one variable in common. The comparison is made between `r.jk` and `r.jh`. The function tests whether the correlations between `j` and `k` (`r.jk`) and between `j` and `h` (`r.jh`) differ in magnitude. Because the significance depends on the intercorrelation between `k` and `h` (r.kh), this intercorrelation has to be provided as an additional parameter. The function expects correlation coefficients as input.

Usage

```
cocor.dep.groups.overlap(r.jk, r.jh, r.kh, n,
  alternative = "two.sided", test = "all", alpha = 0.05,
  conf.level = 0.95, null.value = 0, data.name = NULL,
  var.labels = NULL, return.htest = FALSE)
```

Arguments

`r.jk`	A number specifying the correlation between j and k (this correlation is used for comparison)
`r.jh`	A number specifying the correlation between j and h (this correlation is used for comparison)
`r.kh`	A number specifying the correlation between k and h
`n`	An integer defining the size of the group

alternative — A character string specifying whether the alternative hypothesis is two-sided ("two.sided"; default) or one-sided ("greater" or "less", depending on the direction). Optionally, the initial letter of the character strings ("t", "g", and "l)" can be used.

test — A vector of character strings specifying the tests to be used (pearson1898, hotelling1940, hendrickson1970, williams1959, olkin1967, dunn1969, steiger1980, meng1992, hittner2003, or zou2007). Use all to apply all tests (default). For further information see the tests section below.

alpha — A number defining the alpha level for the hypothesis test. The default value is .05.

conf.level — A number defining the level of confidence for the confidence interval (if test meng1992 or zou2007 is used). The default value is .95.

null.value — A number defining the hypothesized difference between the two correlations used for testing the null hypothesis. The default value is 0. If the value is other than 0, only the test zou2007 that uses a confidence interval is available.

data.name — A character string giving the name(s) of the data. If data.name is NULL, the data names of r.jk, r.jh, and r.kh are used.

var.labels — A vector of 3 character strings specifying the labels for j, k, and h (in this order).

return.htest — A logical indicating whether the result should be returned as a list containing a list of class 'htest' for each test. The default value is FALSE.

Value

Returns an object of the class 'cocor.dep.groups.overlap' with the following slots holding the input parameters described above:

r.jk — Input parameter

r.jh — Input parameter

r.kh — Input parameter

n — Input parameter

alternative — Input parameter

alpha — Input parameter

conf.level — Input parameter

null.value — Input parameter

data.name — Input parameter

var.labels — Input parameter

For each test a slot of the same name exists with a list containing the following elements:

statistic — The value of the test statistic (unless test zou2007 is used).

distribution — The distribution of the test statistic (unless test zou2007 is used).

df — The degrees of freedom of the distribution of the test statistic (if test hotelling1940, hendrickson1970, or williams1959 is used).

p.value — The p-value of the test (unless test zou2007 is used).

conf.int — The confidence interval of the difference between the two correlations (if test meng1992 or zou2007 is used).

Tests

In the following, r_{jk} and r_{jh} are the two correlations that are being compared; Z_{jk} and Z_{jh} are their Z transformed equivalents. r_{kh} is the related correlation that is additionally required. n specifies the size of the group the two correlations are based on. Some tests make use of Fisher's r-to-Z transformation (1921, p. 26):

$$Z = \frac{1}{2}(ln(1+r) - ln(1-r)).$$

pearson1898: *Pearson and Filon's (1898) z*

This test was proposed by Pearson and Filon (1898, p. 259, formula xxxvii). The test statistic z is computed as

$$z = \frac{\sqrt{n}(r_{jk} - r_{jh})}{\sqrt{(1-r_{jk}^2)^2 + (1-r_{jh}^2)^2 - 2k}}$$

(Steiger, 1980, p. 246, formula 4), where

$$k = r_{kh}(1 - r_{jk}^2 - r_{jh}^2) - \frac{1}{2}(r_{jk}r_{jh})(1 - r_{jk}^2 - r_{jh}^2 - r_{kh}^2)$$

(Steiger, 1980, p. 245 formula 3).

hotelling1940: *Hotelling's (1940) t*

The test statistic t is given by

$$t = \frac{(r_{jk} - r_{jh})\sqrt{(n-3)(1+r_{kh})}}{\sqrt{2|R|}}$$

(Hotelling, 1940, p. 278, formula 7) with $df = n - 3$, where

$$|R| = 1 + 2r_{jk}r_{jh}r_{kh} - r_{jk}^2 - r_{jh}^2 - r_{kh}^2$$

(Hotelling, 1940, p. 278). The test statistic is also given in Steiger (1980, p. 246), Glass and Stanley (1984, p. 311, formula 15.7), and Hittner, May, and Silver (2003, p. 152).

williams1959: *Williams' (1959) t*

This test is a modification of Hotelling's (1940) t and was suggested by Williams (1959). Two mathematically different formulae for Williams' t can be found in the literature (Hittner et al., 2003, p. 152). This is the version that Hittner et al. (2003, p. 152) labeled as "standard Williams' t":

$$t = (r_{jk} - r_{jh})\sqrt{\frac{(n-1)(1+r_{kh})}{2(\frac{n-1}{n-3})|R| + \bar{r}^2(1-r_{kh})^3}}$$

with $df = n - 3$, where

$$\bar{r} = \frac{r_{jk} + r_{jh}}{2}$$

and

$$|R| = 1 + 2r_{jk}r_{jh}r_{kh} - r_{jk}^2 - r_{jh}^2 - r_{kh}^2.$$

An alternative formula for Williams' t—termed as "Williams' modified t per Hendrickson, Stanley, and Hills (1970)" by Hittner et al. (2003, p. 152)—is implemented in this function as `hendrickson1970` (see below). The test statistic of `williams1959` is also given in Steiger (1980, p. 246, formula 7) and Neill and Dunn (1975, p. 533).

Results of `williams1959` are in accordance with the results of the software DEPCORR by Hittner and May (1998) and DEPCOR by Silver, Hittner, and May (2006). However, we found several typographical errors in formulae that also claim to compute Williams' t. For example, the formula reported by Boyer, Palachek, and Schucany (1983, p. 76) contains an error because the term $(1 - r_{rk})$ is not being cubed. There are also typographical errors in the formula described by Hittner et al. (2003, p. 152). For example, $r_{jk} - r_{jh}$ is divided instead of being multiplied by the square root term, and in the denominator of the fraction in the square root term, there are additional parentheses so that the whole denominator is multiplied by 2. These same errors can also be found in Wilcox and Tian (2008, p. 107, formula 1).

olkin1967: *Olkin's (1967) z*

In the original article by Olkin (1967, p. 112) and in Hendrickson, Stanley, and Hills (1970, p. 190, formula 2), the reported formula contains a typographical error. Hendrickson and Collins (1970, p. 639) provide a corrected version. In the revised version, however, n in the enumerator is decreased by 1. This function implements the corrected formula without the decrement. The formula implemented in this function is used by Glass and Stanley (1970, p. 313, formula 14.19), Hittner et al. (2003, p. 152), and May and Hittner (1997a, p. 259; 1997b, p. 480):

$$z = \frac{(r_{jk} - r_{jh})\sqrt{n}}{\sqrt{(1 - r_{jk}^2)^2 + (1 - r_{jh}^2)^2 - 2r_{kh}^3 - (2r_{kh} - r_{jk}r_{jh})(1 - r_{kh}^2 - r_{jk}^2 - r_{jh}^2)}}.$$

dunn1969: *Dunn and Clark's (1969) z*

The test statistic z of this test is calculated as

$$z = \frac{(Z_{jk} - Z_{jh})\sqrt{n-3}}{\sqrt{2 - 2c}}$$

(Dunn and Clark, 1969, p. 370, formula 15), where

$$c = \frac{r_{kh}(1 - r_{jk}^2 - r_{jh}^2) - \frac{1}{2}r_{jk}r_{jh}(1 - r_{jk}^2 - r_{jh}^2 - r_{kh}^2)}{(1 - r_{jk}^2)(1 - r_{jh}^2)}$$

(Dunn and Clark, 1969, p. 368, formula 8).

hendrickson1970: *Hendrickson, Stanley, and Hills' (1970) modification of Williams' (1959) t*

This test is a modification of Hotelling's (1940) t and was suggested by Williams (1959). Two mathematically different formulae of Williams' (1959) t can be found in the literature. `hendrickson1970` is the version that Hittner et al. (2003, p. 152) labeled as "Williams' modified t per Hendrickson, Stanley, and Hills (1970)". An alternative formula termed as "standard Williams' t" by Hittner et al. (2003, p. 152) is implemented as `williams1959` (see above). The `hendrickson1970` formula can be found in Hendrickson, Stanley, and Hills (1970, p. 193), May and Hittner (1997a, p. 259; 1997b, p. 480), and Hittner et al. (2003, p. 152):

$$t = \frac{(r_{jk} - r_{jh})\sqrt{(n-3)(1 + r_{kh})}}{\sqrt{2|R| + \frac{(r_{jk} - r_{jh})^2(1 - r_{kh})^3}{4(n-1)}}}$$

with $df = n - 3$. A slightly changed version of this formula was provided by Dunn and Clark (1971, p. 905, formula 1.2), but seems to be erroneous, due to an error in the denominator.

steiger1980: *Steiger's (1980) modification of Dunn and Clark's (1969) z using average correlations*

This test was proposed by Steiger (1980) and is a modification of Dunn and Clark's (1969) z. Instead of r_{jk} and r_{jh}, the mean of the two is used. The test statistic z is defined as

$$z = \frac{(Z_{jk} - Z_{jh})\sqrt{n-3}}{\sqrt{2-2c}}$$

(Steiger 1980, p. 247, formula 14), where

$$\bar{r} = \frac{r_{jk} + r_{jh}}{2}$$

(Steiger, 1980, p. 247) and

$$c = \frac{r_{kh}(1 - 2\bar{r}^2) - \frac{1}{2}\bar{r}^2(1 - 2\bar{r}^2 - r_{kh}^2)}{(1 - \bar{r}^2)^2}$$

(Steiger ,1980, p. 247, formula 10; in the original article, there are brackets missing around the divisor).

meng1992: *Meng, Rosenthal, and Rubin's (1992) z*

This test is based on the test statistic z,

$$z = (Z_{jk} - Z_{jh})\sqrt{\frac{n-3}{2(1-r_{kh})h}},$$

(Meng et al., 1992, p. 173, formula 1), where

$$h = \frac{1 - f\overline{r^2}}{1 - \overline{r^2}}$$

(Meng et al., 1992, p. 173, formula 2),

$$f = \frac{1 - r_{kh}}{2(1 - \overline{r^2})}$$

(f must be ≤ 1; Meng et al., 1992, p. 173, formula 3), and

$$\overline{r^2} = \frac{r_{jk}^2 + r_{jh}^2}{2}$$

(Meng et al., 1992, p. 173). This test also constructs a confidence interval of the difference between the two correlation coefficients r_{jk} and r_{jh}:

$$L, U = Z_{jk} - Z_{jk} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{2(1-r_{kh})h}{n-3}}$$

(Meng et al., 1992, p. 173, formula 4). α denotes the desired alpha level of the confidence interval. If the confidence interval includes zero, the null hypothesis that the two correlations are equal must be retained. If zero is outside the confidence interval, the null hypothesis can be rejected.

hittner2003: *Hittner, May, and Silver's (2003) modification of Dunn and Clark's (1969) z using a backtransformed average Fisher's (1921) Z procedure*

The approach to backtransform averaged Fisher's (1921) Zs was first proposed by Silver and Dunlap (1987) and was applied to the comparison of overlapping correlations by Hittner et al. (2003). The test is based on Steiger's (1980) approach. The test statistic z is calculated as

$$z = \frac{(Z_{jk} - Z_{jh})\sqrt{n-3}}{\sqrt{2-2c}}$$

(Hittner et al., 2003, p. 153), where

$$c = \frac{r_{kh}(1-2\bar{r}_z^2) - \frac{1}{2}\bar{r}_z^2(1-2\bar{r}_z^2 - r_{kh}^2)}{(1-\bar{r}_z^2)^2}$$

(Hittner et al., 2003, p. 153),

$$\bar{r}_z = \frac{exp(2\bar{Z}-1)}{exp(2\bar{Z}+1)}$$

(Silver and Dunlap, 1987, p. 146, formula 4), and

$$\bar{Z} = \frac{Z_{jk} + Z_{jh}}{2}$$

(Silver and Dunlap, 1987, p. 146).

zou2007: *Zou's (2007) confidence interval*

This test calculates the confidence interval of the difference between the two correlation coefficients r_{jk} and r_{jh}. If the confidence interval includes zero, the null hypothesis that the two correlations are equal must be retained. If the confidence interval does not include zero, the null hypothesis has to be rejected. A lower and upper bound for the interval (L and U, respectively) is given by

$$L = r_{jk} - r_{jh} - \sqrt{(r_{jk} - l_1)^2 + (u_2 - r_{jh})^2 - 2c(r_{jk} - l_1)(u_2 - r_{jh})}$$

and

$$U = r_{jk} - r_{jh} - \sqrt{(u_1 - r_{jk})^2 + (r_{jh} - l_2)^2 - 2c(u_1 - r_{jk})(r_{jh} - l_2)}$$

(Zou, 2007, p. 409), where

$$l = \frac{exp(2l') - 1}{exp(2l') + 1},$$

$$u = \frac{exp(2u') - 1}{exp(2u') + 1}$$

(Zou, 2007, p. 406),

$$c = \frac{(r_{kh} - \frac{1}{2}r_{jk}r_{jh})(1 - r_{jk}^2 - r_{jh}^2 - r_{kh}^2) + r_{kh}^3}{(1-r_{jk}^2)(1-r_{jh}^2)}$$

(Zou, 2007, p. 409), and

$$l', u' = Z \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n-3}}$$

(Zou, 2007, p. 406). α denotes the desired alpha level of the confidence interval.

References

Boyer, I. E., Palachek, A. D., & Schucany. W. R. (1983). An empirical study of related correlation coefficients. *Journal of Educational Statistics*, *8*, 75-86. doi:10.2307/1164871

Dunn, O. J. & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, *64*, 366-377. doi:10.2307/2283746

Dunn, O. J. & Clark, V. A. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association*, *66*, 904-908. doi:10.2307/2284252

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 1-32.

Glass, G. V., & Stanley, J. C. (1970). *Statistical methods in eduction and psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Glass, G. V., & Stanley, J. C. (1984). *Statistical methods in eduction and psychology (2nd ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Hendrickson, G. F., Stanley J. C., & Hills, J. R. (1970). Olkin's new formula for significance of r13 vs. r23 compared with Hotelling's method. *American Educational Research Journal*, *7*, 189-195. doi:10.2307/1162159

Hendrickson, G. F., & Collins, J. R. (1970). Note correcting the results in "Olkin's new formula for the significance of r13 vs. r23 compared with Hotelling's method". *American Educational Research Journal*, *7*, 639-641. doi:10.2307/1161847

Hittner, J. B., & May, K. (1998). DEPCORR: A SAS program for comparing dependent correlations. Applied Psychological Measurement, 22, 93-94. doi:10.1177/01466216980221010

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, *130*, 149-168. doi:10.1080/00221300309601282

Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Mathematical Statistics*, *11*, 271-283. doi:10.1214/aoms/1177731867

May, K., & Hittner, J. B., (1997a) - A note on statistics for comparing dependent correlations. *Psychological Reports*, *80*, 475-480. doi:10.2466/pr0.1997.80.2.475

May, K., & Hittner, J. B. (1997b). Tests for comparing dependent correlations revisited: A Monte Carlo study. *The Journal of Experimental Education*, *65*, 257-269. doi:10.1080/00220973.1997.9943458

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172-175. doi:10.1037//0033-2909.111.1.172

Neill, J. J., & Dunn, O. J. (1975). Equality of dependent correlation coefficients. *Biometrics*, *31*, 531-543. doi:10.2307/2529435

Olkin, I. (1967). Correlations revisited. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis* (pp. 102-128). Chicago, IL: Rand McNally.

Pearson, K., & Filon, L. N. G. (1898). Mathematical contributions to theory of evolution: IV. On the probable errors of frequency constants and on the influence of random selection and correlation. *Philosophical Transactions of the Royal Society of London, Series A*, *191*, 229-311. doi:10.1098/rsta.1898.0007

Silver, N. C , & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's Z transformation be used? *Journal of Applied Psychology*, *72*, 146-148. doi:10.1037//0021-9010.72.1.146

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing dependent correlations with nonoverlapping variables: A Monte Carlo simulation. *Journal of Experimental Education*, *73*, 53-69. doi:10.3200/JEXE.71.1.53-70

Silver, N. C., Hittner, J. B., & May, K. (2006). A FORTRAN 77 program for comparing dependent correlations. *Applied Psychological Measurement*, *30*, 152-153. doi:10.1177/0146621605277132

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251. doi:10.1037//0033-2909.87.2.245

Wilcox, R. R., & Tian, T. (2008). Comparing dependent correlations. *The Journal of General Psychology*, *135*, 105-112. doi:10.3200/GENP.135.1.105-112

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B*, *21*, 396-399. Retrieved from http://www.jstor.org/stable/2983809

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413. doi:10.1037/1082-989X.12.4.399

See Also

cocor, cocor.indep.groups, cocor.dep.groups.nonoverlap, as.htest

Examples

```
# Compare the difference between the correlations (age, intelligence) and
# (age, shoe size) measured in the same group (all values are fictional):
r.jk <- .2  # Correlation (age, intelligence)
r.jh <- .5  # Correlation (age, shoe size)
r.kh <- .1  # Correlation (intelligence, shoe size)
n <- 315  # Size of the group

cocor.dep.groups.overlap(r.jk, r.jh, r.kh, n, var.labels=c("age", "intelligence", "shoe size"))
```

---

cocor.indep.groups *Compare two correlations based on independent groups*

Description

Performs a test of significance for the difference between two correlation coefficients based on independent groups. The function expects correlation coefficients as input.

Usage

```
cocor.indep.groups(r1.jk, r2.hm, n1, n2,
  alternative = "two.sided", test = "all", alpha = 0.05,
  conf.level = 0.95, null.value = 0, data.name = NULL,
  var.labels = NULL, return.htest = FALSE)
```

Arguments

`r1.jk`	A number specifying the correlation between j and k measured in group 1
`r2.hm`	A number specifying the correlation between h and m measured in group 2
`n1`	An integer defining the size of group 1
`n2`	An integer defining the size of group 2
`alternative`	A character string specifying whether the alternative hypothesis is two-sided (`"two.sided"`; default) or one-sided (`"greater"` or `"less"`, depending on the direction). Optionally, the initial letter of the character strings (`"t"`, `"g"`, and `"l"`)" can be used.
`test`	A vector of character strings specifying the tests to be used (`fisher1925` or `zou2007`). Use `all` to apply all tests (default). For further information see the tests section below.
`alpha`	A number defining the alpha level for the hypothesis test. The default value is .05.
`conf.level`	A number defining the level of confidence for the confidence interval (if test `zou2007` is used). The default value is .95.
`null.value`	A number defining the hypothesized difference between the two correlations used for testing the null hypothesis. The default value is 0. If the value is other than 0, only the test `zou2007` that uses a confidence interval is available.
`data.name`	A character string giving the name(s) of the data. If `data.name` is `NULL`, the data names of `r1.jk` and `r2.hm` are used.
`var.labels`	A vector of 4 character strings specifying the labels for j, k, h, and m (in this order).
`return.htest`	A logical indicating whether the result should be returned as a list containing a list of class 'htest' for each test. The default value is `FALSE`.

Value

Returns an object of the class 'cocor.indep.groups' with the following slots holding the input parameters described above:

`r1.jk`	Input parameter
`r2.hm`	Input parameter
`n1`	Input parameter
`n2`	Input parameter
`alternative`	Input parameter
`alpha`	Input parameter
`conf.level`	Input parameter
`null.value`	Input parameter
`data.name`	Input parameter
`var.labels`	Input parameter

For each test a slot of the same name exists with a list containing the following elements:

statistic	The value of the test statistic (if test `fisher1925` is used).
distribution	The distribution of the test statistic (if test `fisher1925` is used).
p.value	The p-value of the test (if test `fisher1925` is used).
conf.int	The confidence interval of the difference between the two correlations (if test `zou2007` is used).

Tests

The tests make use of Fisher's r-to-Z transformation (1921, p. 26):

$$Z = \frac{1}{2}(ln(1+r) - ln(1-r)).$$

fisher1925: *Fisher's (1925) z*

This significance test was first described in Fisher (1925, pp. 161-168) and its test statistic z is calculated as

$$z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}.$$

Z_1 and Z_2 are the two Z transformed correlations that are being compared. n_1 and n_2 specify the size of the two groups the correlations are based on. The equation is also given for example in Peters and van Voorhis (1940, p. 188) and Cohen, Cohen, West, and Aiken (2003, p. 49, formula 2.8.11).

zou2007: *Zou's (2007) confidence interval*

This test calculates the confidence interval of the difference between the two correlation coefficients r_1 and r_2. If the confidence interval includes zero, the null hypothesis that the two correlations are equal must be retained. If the confidence interval does not include zero, the null hypothesis has to be rejected. A lower and upper bound for the interval (L and U, respectively) is given by

$$L = r_1 - r_2 - \sqrt{(r_1 - l_1)^2 + (u_2 - r_2)^2}$$

and

$$U = r_1 - r_2 - \sqrt{(u_1 - r_1)^2 + (r_2 - l_2)^2}$$

(Zou, 2007, p. 409). A lower and upper bound for the confidence interval of r_1 (l_1 and u_1) and r_2 (l_2 and u_2) are calculated as

$$l = \frac{exp(2l') - 1}{exp(2l') + 1},$$

$$u = \frac{exp(2u') - 1}{exp(2u') + 1}$$

(Zou, 2007, p. 406), where

$$l', u' = Z \pm z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n-3}}$$

(Zou, 2007, p. 406). α denotes the desired alpha level of the confidence interval, whereas n specifies the size of the group the correlation is based on.

References

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3rd ed.)*. Mahwah, NJ: Erlbaum.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 1-32.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd. Retrieved from http://psychclassics.yorku.ca/

Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases*. New York: McGraw-Hill.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413. doi:10.1037/1082-989X.12.4.399

See Also

cocor, cocor.dep.groups.overlap, cocor.dep.groups.nonoverlap, as.htest

Examples

```
# Compare the difference between two correlations based
# on two independent groups:
r1.jk <- .7  # Correlation between age and intelligence measured in group 1
n1 <- 305    # Size of group 1

r2.hm <- .6  # Correlation between age and intelligence measured in group 2
n2 <- 210    # Size of group 2

cocor.indep.groups(r1.jk, r2.hm, n1, n2, var.labels=c("age", "intelligence", "age", "intelligence"))
```

Index