Power to detect a difference in correlations

Applications for twin studies

Carl Higgs[1]

[1]Centre for Epidemiology and Biostatistics
School of Population and Global Health
University of Melbourne

Outline

Pearson correlation coefficient

$$y \sim N(\mu, \sigma)$$

$$x \sim N(\mu, \sigma)$$

The population correlation, an index of linear change in y as x varies is formed of the (assumed) bivariate normal distribution of the respective variables[1]

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

and is estimated by r

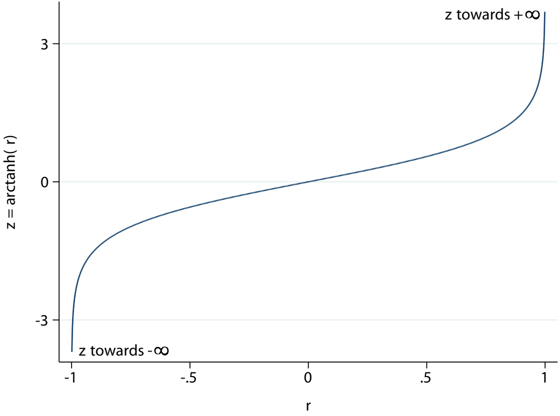
[1] F.N. David. *Tables of the ordinates and probability integral of the distribution of the correlation in small samples*. London: Biometrika, 1938. URL: https://books.google.com.au/books?id=oCa2AAAAIAAJ.

Pearson correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Map r from $(-1, +1)$ to $(-\infty, +\infty)$ as Fisher's z[2]

$$z = \text{arctanh}(r) = \frac{1}{2} \log_e \frac{1+r}{1-r}$$



[2]R. A. Fisher. "Frequency Distribution Of The Values Of The Correlation Coeffients In Samples From An Indefinitely Large Popultion". In: *Biometrika* 10.4 (1915), pp. 507–521. ISSN: 0006-3444. DOI: 10.1093/biomet/10.4.507.

Type 1 and Type 2 error[3]

α

- expected proportion of null hypotheses to be rejected when true
- type 1 error
- the classic '0.05' (5%) but 0.1 or any other number may also be chosen

β

- expected proportion of null hypotheses not rejected when false
- type 2 error
- 0.2 (20%) is a classic choice; gunning for power of $1 - \beta = 80$

[3] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Laurence Erlbaum Associates, 1988. ISBN: 9781134742707.

Hypothesis test for difference in correlations[4]

$$\theta = \operatorname{arctanh}(r_1) - \operatorname{arctanh}(r_2)$$

$$se_\theta = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

$$c_\theta = q / se_q$$

Under the null hypothesis, we evaluate c_θ against the standard normal distribution for the probability of observing an effect size of such mangitude:

$$p(\theta) = \Phi^{-1}_{c_\theta/2}$$

- note: two sided p value; could be onesided

[4] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Laurence Erlbaum Associates, 1988. ISBN: 9781134742707, Christopher L. Aberson. *Applied power analysis for the behavioral sciences*. New York: Routledge, 2010, xiv, 257 p.

Confidence interval for difference in correlations

In the scale of z:

$$\text{CI}_{100(1-\alpha)\%} = \theta \pm c_0 \times se_\theta$$

or back in the scale of r:

$$\text{CI}_{100(1-\alpha)\%} = \tanh\left(\theta \pm c_0 \times se_\theta\right)$$

Power for difference in correlations[5]

$$\theta = \mathrm{arctanh}(r_1) - \mathrm{arctanh}(r_2)$$

$$se_\theta = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

$$c_\theta = q/se_q$$

$$c_0 = \Phi^{-1}_{\alpha/2}$$

$$\beta = \Phi\left(c_0 - c_\theta\right)$$

and

$$\mathrm{power}(\theta) = 1 - \beta$$

[5] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Laurence Erlbaum Associates, 1988. ISBN: 9781134742707, Christopher L. Aberson. *Applied power analysis for the behavioral sciences*. New York: Routledge, 2010, xiv, 257 p.

Power for difference in correlations

$$\text{power} = 1 - \Phi\left(\Phi^{-1}_{\alpha/2} - \frac{\operatorname{arctanh}(r_1) - \operatorname{arctanh}(r_2)}{\sqrt{(n_1 - 3)^{-1} + (n_2 - 3)^{-1}}}\right)$$

Judging magnitude

- r^2 is the proportion of variance of one variable which can be explained by that of the other.
- $r^2 \times 100\%$ of the variance in y is attributable to magnitude of x
- $\tanh(\theta)^2 \times 100\%$ is the magnitude of difference in variance explained by one group compared with the other. Although, by squaring we lose the sign indicating direction of effect; this could be restored.
- Cohen[6] recommends use of r^2 to inform choice of effect size for detection in power calculations (with subject matter knowledge from literature).

[6] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Laurence Erlbaum Associates, 1988. ISBN: 9781134742707.

Spearman correlation coefficient

The Spearman correlation coefficient is calculated using the rank ordered variables for each group; subsequent steps are as per the Pearons correlation coefficient.

Permutation test[7]

- take order statistic (ranked, no ties) representation of combined group correlations r_1 and r_2
- break ties using a Bernoulli trial
- two vectors:
  - v is vector of order statistics: $v = \{v_1, v_2, \ldots, v_N\}$
  - v is corresponding ordered vector of group membership: $g = \{g_1, g_2, \ldots, g_N\}$

[7] Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap*. Monographs on statistics and applied probability. New York: Chapman Hall, 1993, xvi, 436 p. ISBN: 0412042312.

Permutation test[8]

- Permutation lemma: *"Under $H_0 : \rho_1 = \rho_2$, the vector g has probability $1/\binom{N}{n}$ of equaling any one of its possible values"*
- *so, assuming H_0 of no difference, all permuations of z_1 and z_2 are equally likely*
- *combine $n_1 + n_2$ observations from two groups together*
  - *reduces the two sample situation to a single distribution assumed true under H_0.*
  - *if no difference, should be no discernible pattern of this difference in distributions when re-sampled a sufficiently large number of times*

[8] Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap.* Monographs on statistics and applied probability. New York: Chapman Hall, 1993, xvi, 436 p. ISBN: 0412042312.

Permutation test[9]

- without replacement, take sample of size n_1 to represent first group,
- remaining sample of size n_2 represents second group
- take difference in means
- repeat a large number of times
- Evaluate: *does the original difference lie outside the middle $100 \times (1-\alpha)\%$ of the re-sampled distribution? If yes, reject H_0.*
- *Permutation α is probability that the permutation replication $\hat{\theta} \geq \hat{\theta}$ the sample difference, and is evaluated as the proportion of occurances relative to total number of possible permutations*
- *often approximated using Monte Carlo methods*

[9] Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap.* Monographs on statistics and applied probability. New York: Chapman Hall, 1993, xvi, 436 p. ISBN: 0412042312.

Approach taken by R package cocor

cocor[10] is a recent implementation of a flexible calculator for inferences on differences in r

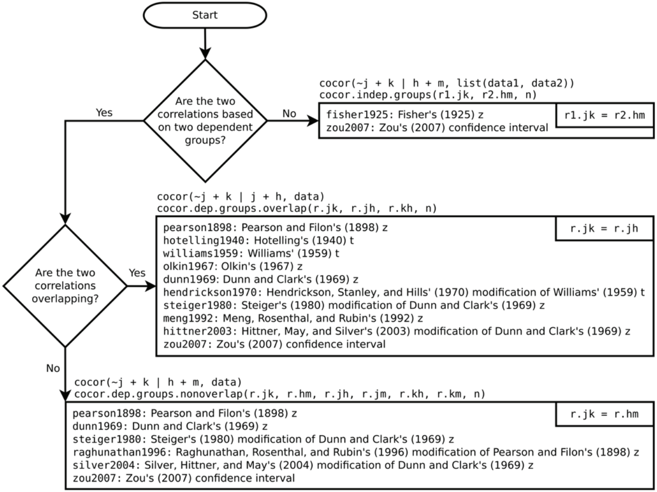


Fig 1. A flowchart of how to use the four main functions of cocor, displaying all available tests. For each case, an example of the formula passed as an argument to the cocor() function and the required correlation coefficients for the functions cocor.indep.groups(), cocor.dep.groups.overlap(), and cocor.dep.groups.nonoverlap() are given. The test label before the colon may be passed as a function argument to calculate specific tests only.

doi:10.1371/journal.pone.0121945.g001

Bibliography I

F.N. David. *Tables of the ordinates and probability integral of the distribution of the correlation in small samples*. London: Biometrika, 1938. URL: https://books.google.com.au/books?id=oCa2AAAAIAAJ.

R. A. Fisher. "Frequency Distribution Of The Values Of The Correlation Coeffients In Samples From An Indefinitely Large Popultion". In: *Biometrika* 10.4 (1915), pp. 507–521. ISSN: 0006-3444. DOI: 10.1093/biomet/10.4.507. URL: http://dx.doi.org/10.1093/biomet/10.4.507.

J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. New York: Laurence Erlbaum Associates, 1988. ISBN: 9781134742707.

Christopher L. Aberson. *Applied power analysis for the behavioral sciences*. New York: Routledge, 2010, xiv, 257 p.

Bibliography II

Bradley Efron and Robert Tibshirani. *An introduction to the bootstrap*. Monographs on statistics and applied probability. New York: Chapman Hall, 1993, xvi, 436 p. ISBN: 0412042312.

B. Diedenhofen and J. Musch. "cocor: a comprehensive solution for the statistical comparison of correlations". In: *PLoS One* 10.3 (2015), e0121945. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0121945.