Power to detect a difference in correlations

Applications for twin studies

C. Higgs[1]

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

Supervision meeting

Table of Contents

Context: Twin studies

Key terms

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

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

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.

Key terms

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

The distribution of r for any n and ρ^2

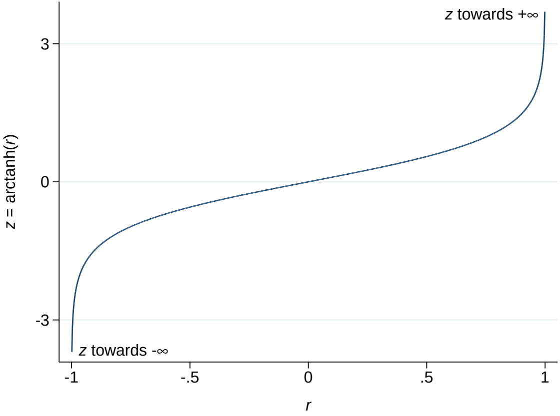
$$p(r|n,\rho) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{(\pi(n-3)!)}(1-r^2)^{\frac{n-4}{2}}\frac{d^{n-2}}{d(r\rho)^(n-2)}\left(\frac{\arccos(-\rho r)}{\sqrt{1-\rho^2 r^2}}\right)$$

which is not so tracactable for statistical inference as,

[2] 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.

Map r from $(-1, +1)$ to $(-\infty, +\infty)$ as Fisher's z[3]

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



[3] 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[4]

$$\alpha$$

- 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: rate of failing to reject the null hypothesis when false (Type 2 error)

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

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

Hypothesis test for difference in correlations[5]

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

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

$$c_\theta = q/se_q$$

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

- note: two sided p value; fill in rest here...

[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. URL: https://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=646521%20Connect%20to%20ebook%20(University%20of%20Melbourne%20only).

Confidence interval for difference in correlations

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

Power for difference in correlations[6]

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

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

$$c_\theta = q/se_q$$

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

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

and

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

[6] 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. URL: https://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=646521%20Connect%20to%20ebook%20(University%20of%20Melbourne%20only).

Power for difference in correlations

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

Bibliography

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. URL: https://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=646521%20Connect%20to%