COEFFICIENTS OF ALIENATION AND DETERMINATION

*Hervé Abdi**

The coefficient of correlation evaluates the similarity of two sets of scores obtained from two variables (or one variable measured twice) within a single sample. It indicates the amount of information common to the two variables. Because this coefficient takes values between −1 and +1, it is difficult to interpret or describe its quantitative meaning easily. For example, it is tempting to interpret a positive correlation as a proportion because it is a decimal number between 0 and 1 and looks like a proportion, but that would not be correct. We can square a correlation, however, and that does create a proportion. A squared correlation is called the *coefficient of determination* and gives the proportion of common variance between the two variables. Subtracting the coefficient of determination from 1 gives the proportion of variance *not* shared between two variables. This quantity is called the *coefficient of alienation.*

As an example, imagine a study interested in the conditions that were related to the spread of the coronavirus (COVID-19). In 2020, Denes K. A. Rosario and colleagues conducted a study that looked at the correlation between various weather factors such as solar radiation, temperature, and wind speed measured daily and the number of people infected with COVID-19 daily in six of the cities in the state of Rio de Janeiro in Brazil. The correlation found between wind speed and the infection rate was −.44. Theory suggests that wind helps to dilute droplets that spread viruses, so the negative direction of the correlation makes sense. A correlation of this size is often interpreted as moderate to strong. But what words can one use to describe the size of relationship in a more direct way? Squaring the correlation gives us a coefficient of determination of about .19, so we can say that the two variables (wind speed and number of infections) share 19% of their variance. Other ways of describing this relationship include that there is 19% overlap or the variables share 19% of their information. The coefficient of alienation for this relationship is 1 − .19 or .81. There was still 81% of the daily number of COVID-19 infections that could be explained by other variables besides wind speed.

The significance of the underlying coefficient of correlation used to compute coefficients of determination or alienation can be tested with an *F* or a *t* test. Note that the coefficient of correlation always overestimates the size of the correlation in the population represented by

---

* **In Bruce Frey (Ed.), *The SAGE Encyclopedia of Research Design.* Thousand Oaks, CA: Sage. 2022. pp. 217–218.**

Address Correspondence to Hervé Abdi,
Program in Cognition and Neurosciences, MS: Gr.4.1,
The University of Texas at Dallas,
Richardson, TX 75080—3021 USA
E-mail: herve@utdallas.edu   https://personal.utdallas.edu/~herve.

the sample used and needs to be "corrected" in order to provide a better estimation. The corrected value is called *shrunken* or *adjusted.*

The fact that two variables share variance does not mean that one variable causes the other one. *Correlation is not causation.* Use of the term *determination* can be misleading. In our example of a moderate coefficient of determination (.19) between the number of COVID-19 cases in Rio de Janeiro and wind speed, it does not make theoretical sense that the number of cases could cause wind speed, so we would not even consider that possibility. It is more reasonable to assume that wind speed affects the physical spread of the virus, but we still cannot assume a cause-and-effect relationship there. It is possible that some other factor is related to the two variables (e.g., perhaps high wind speeds keep people safely inside, or hot weather affects both wind speed and the survival of the virus in the air).

There are limitations to the somewhat common use of coefficient of determination as an effect size and some have argued that it is better to use the correlation coefficient. Effect sizes are simple quantitative indices used to describe the strength of relationships among variables. Objections to the use of coefficient of determination are sometimes based on a rejection of variance itself as the best way to assess the amount of variability in scores. These critics would prefer that the mean of the absolute deviations be used (instead of squaring them to remove negative values as we do for variance). If this is used as the measure of variability, then the correlation coefficient represents the proportion of variance accounted for pretty well.

Hervé Abdi.

See also Correlation Coefficient; R; R^2

FURTHER READINGS

Abdi, H., Edelman, B., Valentin, D., & Dowling, W. J. (2009). *Experimental design and analysis for psychology*. Oxford, UK: Oxford University Press.

Barrett, J. P. (1974). The coefficient of determination—Some limitations. *The American Statistician, 28(1),* 19–20. doi: 10.1080/00031305.1974.10479056.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Edwards, A. L. (1985). *An introduction to linear regression and correlation*. New York: Freeman.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78(3)*, 691–692. doi:10.1093/biomet/78.3.691.