The new way to report correlations

By Julián Monge-Nájera, julianmonge@gmail.com

ABSTRACT: Scientific communication is a crucial aspect of research that allows scientists to share their findings with the broader scientific community and the public. However, many scientific papers are difficult to understand because of the way they report statistical correlations. As a rule, the way you learned to do it, is not the best. Here I present a better, and valid, way to report correlations, with biological examples. Basically: state what they mean, rather than the raw result; and delete all obvious statements.

KEYWORDS: statistics, results, tests, causation, communication.

FIGURE 1. Writing the meaning of correlations, instead of the raw results, will significantly improve scientific communication. Additionally, do not report non-significant correlations and remember, correlation does not imply causation. Image: Bing Image Creator-J. Monge-Nájera, and Researchgate.net.

Statistical correlation tests are a crucial tool to identify relationships between variables in biological phenomena. The concept of correlation dates back to the early 19th century, when Sir Francis Galton, a British scientist, first introduced the idea of correlation coefficient. However, it was not until the 20th century that statistical correlation tests became widely used in research. One of the pioneers in this field was Ronald Fisher, who developed several statistical methods, including analysis of variance (ANOVA) and regression analysis.

Correlation tests can be broadly classified into two categories: positive and negative correlations. Positive correlations indicate that, as the value of one variable increases, the value of the other variable also tends to increase. In negative correlations, one goes down as the other goes up. Statistically non-significant correlations are normally the result of chance and should not be reported as correlations at all.

Famous examples of correlations are those between smoking and lung cancer; between exercise and cognitive function in older adults, and between number of nesting storks and of babies being born; in other words, correlation does not always mean causation: while it is true that smoking causes lung cancer; storks do not really bring babies [1]; interpret correlations carefully and consider other variables that may be influencing the relationship. In any case, many scientific papers are filled with long, complex reports of statistical correlations that can be challenging for readers to follow [2]. Here, I explain how to improve scientific communication by writing shorter descriptions of any statistical correlations.

The following imaginary example is written in the bad style that is common in scientific articles, and sadly, even in textbooks, leading to hundreds of thousands of bad reports when students learn the wrong way from their teachers:

The study identified a high statistically significant positive correlation between the diameter of the canopy of the R. mangle and A. germinans species and the pH values at the monitoring sites ($r=0.85$, $p=0.03$ and $r =0.7$, $p=0.02$, respectively). Additionally, a statistically significant negative correlation was found between the height of the trees and the phosphorus content ($r =-0.83$, $p=0.04$). Furthermore, a significant positive correlation was identified between the tissue moisture in A. germinans and the clay content in the soil ($r = 0.6$, $p = 0.02$). The analysis also revealed a significant negative correlation between the temperature of the water and the survival rate of the fish larvae in the area ($r =-0.65$, $p < 0.05$). Finally, there was a significant positive correlation between the amount of rainfall and the diversity of plant species in the mangrove forest ($r = 0.55$, $p < 0.05$).

Compare with this improved version:

"In sites with higher pH, R.mangle and A. germinans trees had wider canopies ($r = 0.85$, $p=0.03$ and $r =0.7$, $p =0.02$, respectively). Sites with more phosphorus had shorter trees ($r =-0.83$, $p=0.04$), and, in soils with more clay, A. germinans tissue moisture was higher ($r = 0.6$, $p = 0.02$). As water temperature increased, less fish larvae survived ($r =-0.65$, $p<0.05$), while plant diversity was highest in mangroves with more rainfall ($r=0.55$, $p<0.05$)."

The improved version, nearly 50 % shorter, presents the same information but is far easier to understand because it eliminates redundant statements (the use of "significant differences" or "statistically significant correlation" is redundant, this is implied in the probability value) and because it presents the meaning of the correlation (e.g. "sites with more phosphorus had shorter trees") instead of just presenting the raw result in a lengthy way ("there is an inverse significant correlation between phosphorous concentration at the sites and tree height").

Briefly: Writing the meaning of correlations, instead of the raw results, will significantly improve scientific communication. This will make your findings easier to understand, use and cite. Share this information with your students and colleagues to promote better scientific communication practices. And remember, statistically non-significant correlations should not be reported as correlations at all.

References

[1] Matthews, R. (2000). Storks deliver babies (p= 0.008). Teaching Statistics, 22(2), 36-38.
[2] Psychology Writing Center, University of Washington (2012). Reporting Results of Common Statistical Tests in APA Format (https://psych.uw.edu/storage/writing_center/stats.pdf).

Julián Monge-Nájera is a Costa Rican scientist whose work has been featured by The New York Times, National Geographic, the BBC; Wired, IFLoveScience, The Independent and The Reader's Digest. Panelist of the Apocalypse Clock, curator in Encyclopedia of Life and member of the Red List of Threatened Species team at IUCN (Switzerland). He has published in prestigious outlets like Nature, The Linnean Society and the Royal Society.