

Improving the Use of Statistics in Marine Ecology

Guangli Yu*

Department of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China

Abstract

In marine ecology, linear regression is a statistical method that is commonly employed, either to describe straightforward correlations or as a part of more intricate models. This method's seeming simplicity frequently hides its much more complicated foundations, on which its validity and eventual ecological interpretations completely depend. With an emphasis on proper model specification, the various concepts of linearity, the problems with data transformation, the assumptions that must be respected, and regression model validation, we present a non-technical review of the fundamentals of linear regression and its application in marine ecology. R² and p-values alone do not provide enough data to draw valid conclusions.

Keywords: Marine ecology • Statistics • Cellular ultrastructure

Introduction

Given the limitations of the human brain, no researcher can fully understand all fields of study pertaining to marine ecology. Hence, there are three main groups of statisticians who conduct research that is more or less directly related to marine ecology: Quantitative marine ecology specialists, i.e., those whose primary level of expertise is in sampling and experimental design, data treatment, and interpretation; these workers typically concentrate on the upper levels of ecological organisation, such as populations and communities; Biological marine ecology specialists who work on related biological questions and have advanced familiarity with statistical procedures; their main focus is on the interaction of individual organisms with their environment. Those who operate at the sub-individual level, such as in physiology, chemical ecology, or cellular ultrastructure, and who perceive statistics as a kind of recipe book to follow in order to be considered seriously by journal reviewers. Several eminent "giants" in the field of quantitative marine ecology fall under the first group of scientists; they have emphasised the value of designing studies with statistical robustness as their top priority. For lack of a better name, we will refer to the last two kinds of researchers as "functional" ecologists in this review. These are those that study how humans interact with the environment [1,2].

Literature Review

Even with a sound methodology, the planning stage of the study and the analysis and interpretation of the collected data both pose serious risks to the accuracy of the findings. Biological literature has extensively covered each of these topics, although statistical application has a history of being misapplied and misinterpreted. There are much more fundamental issues that frequently arise in the use of statistics, whether we are aware of them or not, aside from the mundane arguments over which test is best suited for which experimental design. In fact, numerous significant studies and entire books have been written about the historical and contemporary misunderstanding, abuse, and

misinterpretation of statistics in almost every field of study. Many marine ecologists might also be surprised to learn that the social sciences, particularly psychology, and the medical sciences, accepted. Statistical data analysis was first used in the field of ecology in the middle of the 1950s as opposed to the middle of the 1960s, and during the past 20 years, these fields have led the way in new advancements in statistical practise. Terrestrial ecologists have recently made significant strides towards better statistical methods as a result of these developments but these considerations don't seem to be well-represented in the work of all but the most established quantitative ecologists [3].

In marine ecology, there are three possible goals for linear regression analysis

- Defining the type of connection between the two variables. If all we want to do is say, "This is the equation that seems to characterise the relationship," then we don't need to worry about many assumptions or preconditions. In marine ecology, however, where we typically want to be able to predict the value of the dependent variable for a given value of the independent variable (for example, what sardine or tuna weight corresponds to what sardine or tuna length-values that are much quicker and easier to measure shipboard?), this is not a very useful tool.
- Prediction of dependent variables that fall within the scope of actual dependent variables. Now, all we want to do is forecast any y-value that falls between the maximum and minimum observed x-values, for example, what weight for any length that is between the maximum and minimum weight values' x-coordinates. This is a much more worthwhile goal, but it comes at the cost of more tight assumptions.
- Prediction of dependent variables outside the scope of observed dependent variables. Here, we make a brave endeavour to venture beyond the highest and lowest y-values that have been observed. The use of this modelling extension ranges from enzyme kinetics to climate change. Typically, it is an effort to foresee a future y-value, which is what humans do. Because marine ecologists have a general propensity to believe that the abstract, ideal mathematical world can be used to directly simulate the much messier real world, there is a great deal of very genuine misunderstanding and misuse of linear regression. In the actual world, a variety of uncontrolled variables, in addition to the independent variable, might affect the dependent variable, such as individual variances in physiology, the amount of time samples are handled, or even variations in atmospheric pressure. As a result, we are aware that other factors may affect the dependent variable, but we are unable to define or quantify these factors or their influence. Moreover, these variables may have a multiplicative or additive effect on the dependent variables (i.e., add their unknowable positive or negative values to the linear equation). Statisticians have

***Address for Correspondence:** Guangli Yu, Department of Medicine and Pharmacy, Ocean University of China, Qingdao 266003, China; E-mail: glyu123@ouc.edu.cn

Copyright: © 2023 Yu G. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 02 January, 2023, Manuscript No. jbes-23-90509; **Editor Assigned:** 04 January, 2023, PreQC No. P-90509; **Reviewed:** 16 January, 2023, QC No. Q-90509; **Revised:** 23 January, 2023, Manuscript No. R-90509; **Published:** 30 January, 2023, DOI: 10.37421/2332-2543.2023.11.462

combined all contributions from unobserved variables into a single phrase and given it the terrible moniker "error term." As much or possibly most of this variance may not be attributable to any true "mistake," but rather to random events, the differences between individual organisms, and tradition, this word is primarily employed for convenience. It is a prime example of incorrect nomenclature in the sciences.

Discussion

There are potential problems to be avoided at each of the three steps in the approach, regardless of how quickly and simply a software programme would do a linear regression. The regression model must be correctly specified as the first and most important stage. The best estimation of the regression parameters is the second step, and model validation is the third (e.g. test whether the regression parameters are statistically different from zero and verify the goodness of fit). Since they relate to the structure of the model, the independent variables it contains, the distribution of residuals, and the first and second steps, the assumptions can be tested both before and after the regression model is built. ε — the error term — this factor is seldom shown in the equations supplied in most marine ecology publications, as it cannot be quantified (and its influence and the consequences of its characteristics are often unknown or unforeseen), but its qualities may enormously alter the equation. It stands for all of the variation in Y that is not accounted for by variance in X. This is how the other real-world influences on the dependent variable are mathematically quantified. This fluctuation will have an impact on each y-value, thus we may define an "error" as the difference between each sample's unique y value and the population's unidentified real value. Because the impacts of might be extremely modest to very large [4].

It's crucial to focus on the size and fluctuation of while performing regression analysis. The only feasible tactic in this situation is to reduce as much as possible through controls and repetition because its scale is by definition uncertain. Similar to how it is typically impossible to pinpoint the precise variation of (also known as the error distribution) in a population, we can test hypotheses about it in our samples, and this will allow us to determine whether or not linear regression is a meaningful way of connecting the dependent and independent variables. We accomplish this by comparing the variances between each sample's unique y-values and the corresponding y-values predicted by the regression model [5,6].

Conclusion

The term "residuals" is frequently used in the same way that the phrase "sample standard deviation" is used to refer to the population standard deviation. Residuals are estimates of the genuine population error. The ideal distribution for residuals is random; in other words, they shouldn't exhibit bias because bias renders statistical methods useless. An essential but frequently overlooked component of linear regression is the examination of residuals. As we present the proper linear regression approach, we will talk about residuals analysis in much more detail.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Armitage, Derek. "Socio-institutional dynamics and the political ecology of mangrove forest conservation in Central Sulawesi, Indonesia." *Global Envir Change* 12 (2002):203-217.
2. Aswani, Shankar and Kenneth Ruddle. "Design of realistic hybrid marine resource management programs in Oceania 1." *Pacific Sci* 67 (2013): 461-476.
3. Aswani, Shankar. "Insights from experimental economics on local cooperation in a small-scale fishery management system." *Global Environ Change* 23 (2013): 1402-1409.
4. Aswani, Shankar, Amy Diedrich and Kitty Currier. "Planning for the future: Mapping anticipated environmental and social impacts in a nascent tourism destination." *Soc Natural Res* 28 (2015): 703-719.
5. Aswani, Shankar and Simon Albert. "One size does not fit all: critical insights for effective community-based resource management in Melanesia." *Marine Policy* 81 (2017): 381-391.
6. Gause, Georgii Frantsevich. "Experimental studies on the struggle for existence: I. Mixed population of two species of yeast." *J Exp Biology* 9 (1932): 389-402.

How to cite this article: Yu, Guangli. "Improving the Use of Statistics in Marine Ecology." *J Biodivers Endanger Species* 11 (2023): 462.