

N-STEPS Objectives

Provide regions, states, and tribes with support related to nutrient criteria development

Provide access to expert assistance with issues related to nutrient criteria development and implementation

Improve communication nationwide.

What Is It?

Correlation is a measure of the strength of a relationship between two variables. Correlations do not indicate causality and are not used to make predictions; rather they help identify how strongly and in what direction two variables covary in an environment. In the context of nutrient criteria development, correlation analysis is a powerful tool to explore which variables may be strongly related to nutrient concentrations.

Types:

- Pearson (parametric, assumes linear relationship)
- Spearman (non-parametric, can be non-linear)
- Kendall's Tau (non-parametric, can be non-linear)

Example Question: How strongly is total phosphorus concentration related to the richness of macroinvertebrate taxa?

How is it Applied to Nutrient Criteria Development?

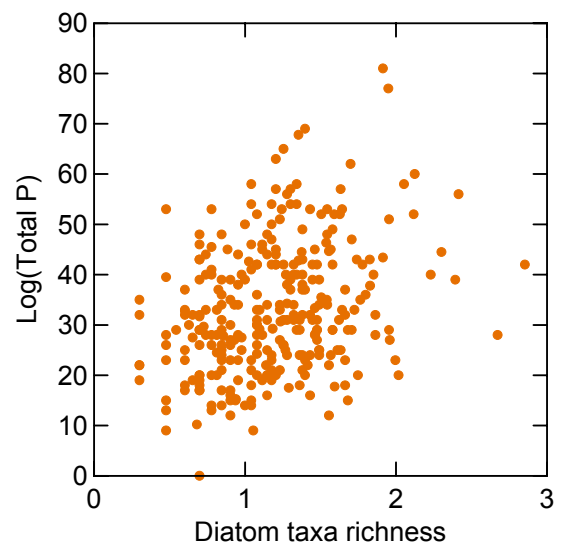
Nutrient criteria development involves three main processes: identifying relationships between biological responses and nutrient stressors, examining these relationships, and establishing nutrient and/or biological thresholds or criteria.

Correlation analysis is a powerful tool to identify the relationships between nutrient variables and biological attributes. The purpose of correlation analysis is to discover the strength of these relationships among a suite of nutrient and biological attributes and to select the most interesting relationships for further analysis. Correlations do not indicate causality and are not used to make predictions; they help identify how strongly and in what direction two variables covary in an environment.

How Does It Work?

Pearson Product Moment - Calculates a correlation coefficient (r) that is the ratio of the covariance of two variables (sums of products of both variables) to their individual variances (square of the sum of each variable). In other words, how much of the change in one variable is associated with changes in the other. Pearson correlation assumes the two variables are approximately normal and related in a linear fashion. Transformations can be used to help meet these assumptions.

Spearman Rank Correlation - Measures the monotonic relationship (one variable simply increasing or decreasing with another) between two variables. It makes no assumptions about the shape of that relationship. Procedure assumes that the two variables are randomly sampled from continuous populations. Values are ranked separately and the ranks among sites are compared using the same correlation coefficient as for the Pearson Correlation.



- Pearson $r = 0.30$, $p < 0.001$, $N=294$
- Spearman $r = 0.29$

Data Requirements

Independently collected numeric data in the form of paired observations are required. These are preferably continuous data, although discrete numeric variables (e.g., taxa richness) can also work. The greater the range of environmental conditions encompassed the better. One way to assure a large range is to use a gradient design and select sites along as large a gradient as possible.

What Should You Look For & Report?

Examine the bivariate plots and the distribution or behavior of the values. This will help choose which method to use. You should see some relationship between the two variables when plotted together if they are correlated. A non-significant correlation does not mean two variables are not related – the relationship may be non-linear or non-monotonic. Similarly, significant relationships may not mean much – often times large sample sizes produce significant correlations. Also, keep an eye on outliers, they can wreak havoc with correlations – especially with small datasets.

Report the correlation coefficient, the number of data pairs or degrees of freedom, and the significance (p-value) or type I error rate (alpha).

Pros

- Effective way to convey simple relationships
- Easy to understand
- Quantitative measure of bivariate association

Cons

- Hard to detect complex relationships (quadratic) without transformations
- Lack of significance does not mean lack of association
- Large sample sizes can lead to significant but small correlations
- Not predictive

Alternatives

- Linear and non-linear regression
- Loess
- Multivariate analyses (for more than one variable at a time)

Citations

EPA Statistical Primer - <http://www.epa.gov/bioindicators/statprimer/index.html>

Ott, R.L. 1993. An introduction to statistical methods and data analysis. 4th edition. Duxbury Press, Belmont, CA.