Explaining Environmental Data (Newsletter)*

March 3, 2014

Most people are familiar with statistical hypothesis tests such as the t-test and ANOVA to analyze whether two or more samples (from a parametric distribution) came from the same population. The nonparametric equivalents (Wilcoxon and Kruskal-Wallis tests) are less familiar but equally robust. What is not always clear is that these models are applied to one or more response variables; e.g., chemical concentrations that result from natural or anthropogenic causes. They do not answer the question of why these values were observed.

Regulators, stake holders, and environmental NGOs question the potential for adverse project effects on natural ecosystems, particularly surface waters. These concerns are expressed at all stages of a project's life cycle. One aspect of effectively addressing these concerns requires including explanatory and response variables in the statistical model to estimate how much response variability is explained by each explanatory variable. Which statistical model to apply depends on the specifics of the concern and the nature of the data (e.g., chemical or biological).

When the question is of the type, "What affect does <activity> have on surface water quality?" start with redundancy analysis, nonmetric multidimensional scaling, or correspondence analysis to reduce the number of meaningful explanatory variables and map their relationships to response variables.

When the question is of the type, "What affect does <activity> have on fish or wildlife?" start with Poisson regression for count data and logistic regression for presence-absence data. When the biotic data do not show a clear linear relationship with potential explanatory variables, mixed effects modeling is a suitable alternative to linear regression. In addition to these frequentist (hypothesis testing) statistical models, there are other statistical models based on maximum likelihood estimation and Bayesian paradigms. MLE and Bayesian approaches require fewer assumptions of the data and produce equally robust results. These models can be extremely useful when the organisms of concern are listed under the Endangered Species Act (or are being considered for listing) and technically sound, legally defensible statistical models are necessary to inform regulatory decisions.

Two other regulatory issues addressed by environmental data analyses are statutory water quality goals and estimating potential impacts of a project in an environmental impact assessment.

Many states have a water quality goal stated as "no degradation" or equivalent term. While laudable in concept, it is difficult (or impossible) to comply with such a standard in rural areas in which many natural resource projects are located because there are no baseline data for comparison. If there has been no regulated activity in the drainage basin there has been no reason to collect water and analyze concentrations of constituents. This means there are no values against which current values can be compared. For water quality, and the broader situation of forecasting effects of an approved project on defined components of the natural environment, it is necessary to quantify variability and separate the inherent natural fluctuations from those attributable to anthropogenic activities using appropriate statistical models. This approach is quantitative and defensible as having taken a hard look at the project and the natural environment in which it is located.

—

All newsletters, white papers, and other technical resources can be freely downloaded from `http://www.appl-ecosys.com/publications/`.