

Practical Stats Newsletter for May 2013

Subscribe and Unsubscribe: <http://practicalstats.com/news/>

Download any of our past newsletters:

<http://www.practicalstats.com/news/bydate.html>

In this newsletter:

1. Upcoming Events
2. Which of Many Variables Are Most Important?
3. Discount on Course Registration!

## 1. Upcoming Events

### **In-person courses**

Untangling Multivariate Relationships

August 13-14, 2013

Allmendinger Center, Washington State Univ.

Puyallup, WA 98371

Time Series Methods for Frequently Collected Data

August 15-16, 2013

Allmendinger Center, Washington State Univ.

Puyallup, WA 98371

Applied Environmental Statistics

*“Statistics, Down to Earth”*

Nov. 18-22, 2013

Homewood Suites

Littleton, CO 80127

### **Webinars:**

We're finishing up two series of webinars, *Applied Environmental Statistics* and *Nondetects And Data Analysis*. You may find the remaining individual webinars of interest. And if you missed these, this fall we will host a few individual webinars on hot topics (if you have any suggestions, send them along!). These will be announced in July's newsletter, or see the web page for more information.

To register and for more information on all of our courses and webinars, see our [Training](http://www.practicalstats.com/training/) page at <http://www.practicalstats.com/training/>

## 2. Which of Many Variables Are Most Important?

Our Untangling Multivariate Relationships course highlights several methods for determining which of a string of variables you've measured provides the greatest ability to discriminate between groups, or to explain the community structure of organisms. The

idea is the same as with tools such as AIC that choose the several best subsets of explanatory variables in multiple regression. The difference is that the Y variable is now also multivariate – counts of 35 different species at each site, for example.

Figure 1 shows an NMDS (see our April 2007 newsletter) of streams across the middle US with atrazine concentrations either above 1 ug/L (shown in green) or below 1 ug/L (blue). Each point has a number of measured attributes associated with it, such as the flow percentile of runoff and the number of days since a crop was planted (a surrogate for days since atrazine was last applied).

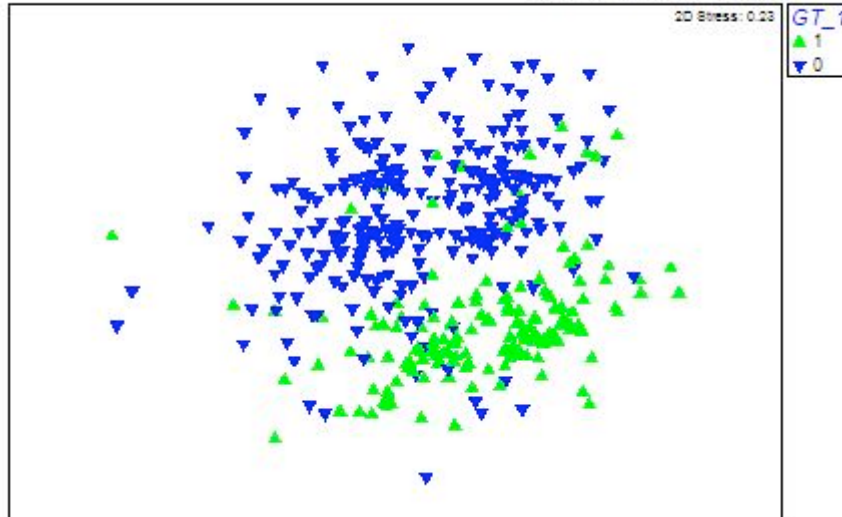


Figure 1. NMDS of atrazine categories

A Mantel-type test (see our August 2011 newsletter) found that the pattern of environmental attributes differed significantly between the two groups. Which of the attributes best distinguished the high and low atrazine sites? The Simper routine in Primer provides the answer:

	Group 1	Group 0				
Variable	Av.Value	Av.Value	Av.Sq.Dist	Sq.Dist/SD	Contrib%	Cum.%
DYPLANT	-0.841	0.496	2.85	1.09	24.63	24.63
FPCTL	0.709	-0.418	2.66	1.03	23.05	47.68
PRECIP	0.255	-0.151	2.05	0.74	17.71	65.39
SOILGP	0.147	-8.67E-2	2.04	0.71	17.66	83.05
CORN%	0.13	-7.7E-2	1.96	0.74	16.95	100.00

The Contrib% column shows that DYPLANT, the number of days since a crop was planted, explained the greatest amount of the difference between the two groups, while flow percentile (FPCTL) was second. The other listed attributes were quite a bit lower.

Another Mantel-type test run on a different data set found that the community structure of 140 nematodes at 19 sites was related to a suite of physical characteristics of those sites. So for this relationship, which of the 6 characteristics were most important? How much information is still present if the best 4 characteristics are used instead of 6, and how much information remains using only these 4? These questions are answered with the BEST procedure.

Variables

- 1 Med Part Diam
- 2 Dep Water Tab
- 3 Dep H2S layer
- 4 Shore height
- 5 %Organics
- 6 Interstitial Salinity

Global Test

Sample statistic (Rho): 0.812  
Significance level of sample statistic: 1%  
Number of permutations: 99 (Random sample)  
# of permuted statistics >= Rho: 0

Best results

No. Vars	Corr.	Selections
3	0.812	1,3,6
4	0.803	1,3,5,6
5	0.800	1-3,5,6
5	0.800	1,3-6
4	0.792	1,3,4,6
4	0.790	1-3,6
5	0.776	1-4,6
4	0.775	1,3-5
2	0.774	3,6
4	0.773	1-3,5

The characteristics that best explained the pattern of nematode abundance are #1,3 and 6: median particle diameter, depth to the sulfide (H2S) layer, and interstitial salinity. The multivariate correlation of the pattern of just these 3 characteristics and the pattern of nematode abundance is 0.812. Using only these three, rather than the entire six, explains most of why some nematodes occur at some sites, and others at other sites.

Powerful methods such as these allow the scientist to make sense out of a confusing pattern of multiple biological and chemical attributes. That is the goal of our Untangling Multivariate Relationships course, on deck for this August in Washington State.

### 3. Discount on Course Registration!

We publish a newsletter of tips on environmental statistics -- this pdf is the May newsletter, for example. Subscribing saves us a lot of time NOT answering email questions about "when is the next course/webinar being held?" So to say thank you in a Practical way, when you subscribe to our newsletter at <http://practicalstats.com/news/> you'll receive a discount code for 10% off registration costs to either our Untangling Multivariate Relationships or Time Series Methods for Frequently-Collected Data course this coming August.

This offer is also available to friends and co-workers of yours who sign up. Point them to our Newsletter page to sign up, and the discount is theirs also!

'Til next time,

Practical Stats (Dennis Helsel)  
-- Make sense of your data