Practical Stats Newsletter for February 2018

Subscribe and unsubscribe: http://practicalstats.com/news
Archive of past newsletters http://practicalstats.com/news/archive.html

In this newsletter:
A. Upcoming Courses
B. The Importance of Control Sites
C. A Good Source of Interesting Graphics

A. Upcoming Courses

Our self-paced Applied Environmental Statistics course is available in two parts on our online training site:
http://practicalstats.teachable.com/
The two courses separately are each $650 USD for a 1-year access for one person. Or get both courses together (equivalent to the week-long course) in a bundle for $1200 USD.

B. The Importance of Control Sites

Setting up a sampling design for a study to determine the effect of some change in human action on water quality immediately presents an important question – how can human effects be separated from those resulting from non-human causes such as wet vs dry years? Sampling over time, from before until after the human action has changed, leaves open the possibility that any observed change in water quality might have occurred even if the change in human activity did not occur. How can the one possible cause be separated from other possible causes? The classic answer from experimental design is to sample control sites, sites where the human activity did not change, during the entire study period. Patterns in water quality at the control site are compared to the study site, to see if the study site's human action produces a noticeably different pattern.

As an example, Figure 1 shows a downtrend in concentration over a 6-year period, at the middle of which a change in farming practice occurred upstream (or the human change could have been an improvement in waste treatment, or removal of a dam, or a variety of other possible activities). It is tempting to attribute this change to the activity being investigated. However, a maxim in hydrology is that the wettest (or driest) 3-year period on record is almost guaranteed to occur during the early, or the late, half of the study you've just diligently instrumented and monitored (the hydrologic equivalent of Murphy's Law). Is the decrease in concentration due simply to an increased streamflow during the second half of the study?

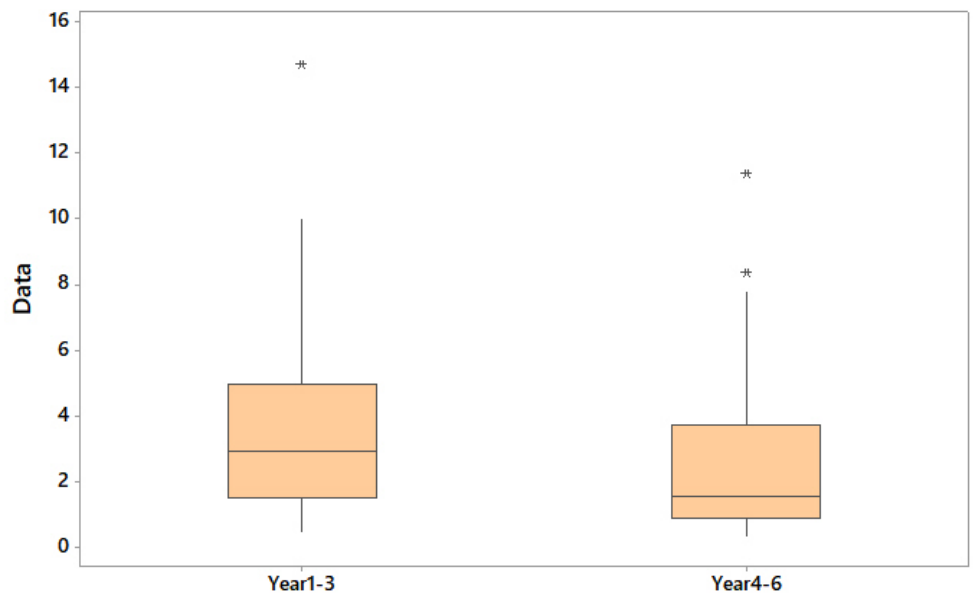

Figure 1. Concentrations at the Study site before vs after the change in human activity

Figure 2 shows the same data along with concentrations at a control site where no change in activity occurred during the time period. It could be a site where farming activity was similar to the early years at the Study site, and remained that way during the final 3 years. Or, the Control site could be an upstream site out of the effects of the dam that was removed during the study period.

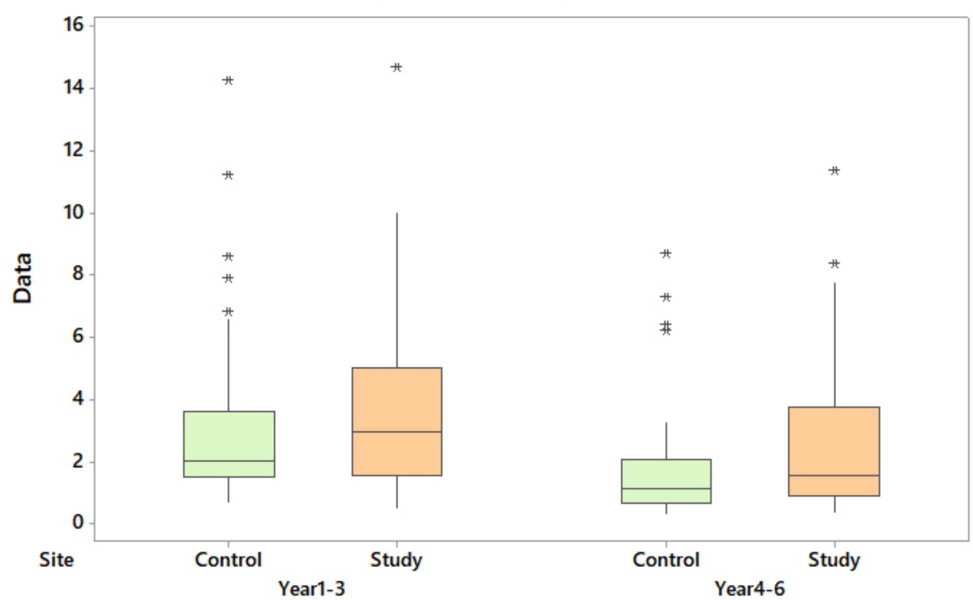

Figure 2. Concentrations at both the Study and Control sites before vs after the change in human activity

Seeing a similar pattern at the two sites (as in Figure 2) would indicate that the decrease in concentration is likely caused by something other than the change being investigated, or at least that any effect due to the change was small compared to that from other causes such as a changing flow regime. Testing the significance of this difference can be done by looking for an 'interaction' between site and time in a regression or analysis of variance model. For these data (in log units – these data were non-normal), the interaction term (Years*Site) was not significant, indicating the decreasing concentration pattern at the control and study sites were not significantly different – no measurable effect of the human change.

Analysis of Variance for lnData

Source	DF	SS	MS	F	P
Years	1	8.8179	8.8179	12.14	0.001
Site	1	1.9079	1.9079	2.63	0.107
Years*Site	1	0.5422	0.5422	0.75	0.389
Error	140	101.6742	0.7262		

Figure 3 presents a different story. Concentrations at the Control site stay similar during the entire study period (actually increase a slight bit), while the Study site concentrations drop as before. This difference in the patterns at the two sites indicates a probable effect due to the change in activity. The interaction test statistic and p-value for the accompanying ANOVA (in log units – these data were non-normal) is significant, indicating a different pattern occurs at the two sites. The drop at the Study site was not evident at the Control site.

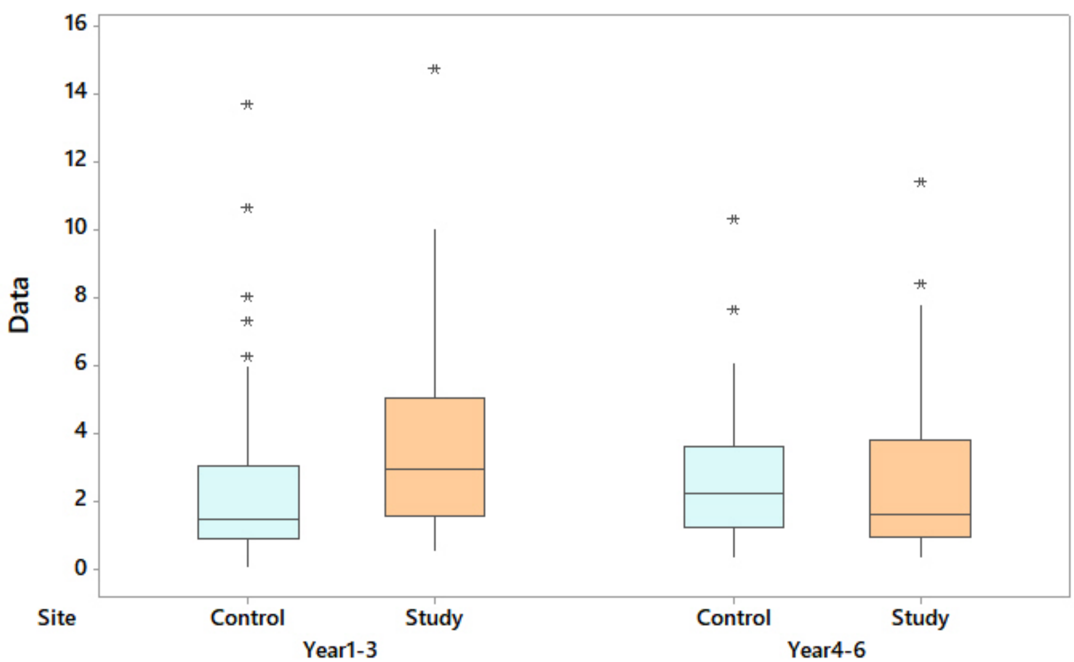

Figure 3. Concentrations at the Study site decreases after the change in human activity, while those at the Control site do not.

Analysis of Variance for lnData

Source	DF	SS	MS	F	P
Years	1	0.000	0.000	0.00	0.991
Site	1	2.727	2.727	2.66	0.105
Years*Site	1	4.937	4.937	4.81	0.030
Error	140	143.643	1.026		
Total	143	151.308			

Without an adequate control site, it is easy to mistakenly attribute observed changes to a cause that isn't in fact the driver of the observed pattern. Costs of sampling a control site throughout the entire extent of the study should always be included when investigating 'the effects of X on water quality' measured over time. Conditions other than the one being studied will be changing over that time. Deal with it, with control sites.

C. A Good Source of Interesting Graphics

If you haven't seen the website fivethirtyeight.com, it provides articles accompanied by often excellent and innovative graphics for data analysis. Use of color is excellent. Data techniques often clearly present interactions of many more than two variables on a graphic. Even their tables look interesting! A recent, though not the most complex of examples is:
https://fivethirtyeight.com/features/we-used-broadband-data-we-shouldnt-have-heres-what-went-wrong/

And if you like baseball you'll like this article, my personal fave read in all of 2017:
https://fivethirtyeight.com/features/goose-egg-new-save-stat-relief-pitchers/
The author creates a new stat to measure something intrinsic that has previously been missed. Perhaps you can do the same for the environmental studies you're an expert on? Don't just repeat what others do!

Take a look at their graphs on a regular basis and you're bound to learn ways you can improve your data presentation skills.

'Til next time,

Practical Stats
-- Make sense of your data