Practical Stats Newsletter for October 2010

Subscribe and Unsubscribe: http://practicalstats.com/news/
All of our past newsletters:
http://www.practicalstats.com/news/news/bydate.html

In this newsletter:
1. Upcoming Courses
2. Regression doesn't mimic a 1:1 line
3. News from Practical Stats

1. Upcoming Courses
The price increase date for registration for our NADA and Multivariate classes in Atlanta is fast approaching - the end of this month. Don't wait any longer to get the latest on handling data with nondetects, or discerning patterns in multiple chemical and biological variables. If you have benefitted from our newsletters or courses, please tell your contacts about our upcoming courses. It makes a difference. Word of mouth is still a major way that people find out about them. There are still openings in both.

Untangling Multivariate Relationships
Nov 15-16, 2010 $895 registration before Nov 1, 2010
Doubletree Atlanta-Buckhead $995 on or after Nov 1.
Atlanta, GA 30326

Nondetects And Data Analysis
Nov 17-18, 2010 $895 registration before Nov 1, 2010
Doubletree Atlanta-Buckhead $995 on or after Nov 1.
Atlanta, GA 30326

… and if you're in southern CA in early November, sit in on a 1-day introduction to correctly using methods for nondetects (the full story is in our 2-day course, above):

Interpreting Nondetect Data Correctly
Nov 2, 2010 $400 registration
Doubletree Hotel, Ontario, CA
sponsored by the Groundwater Resources Assoc of California

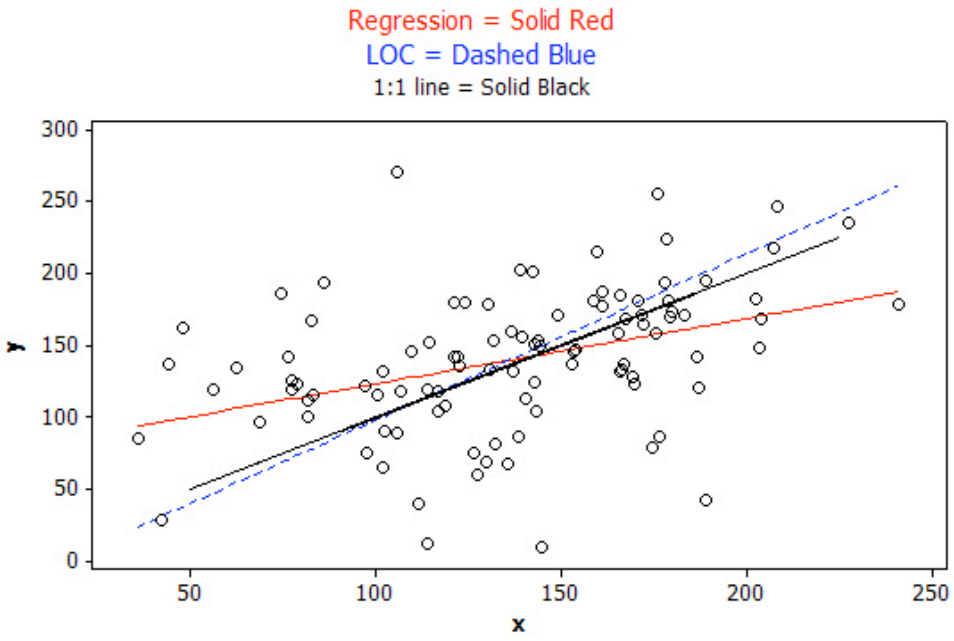
You can always find our complete course listing at
http://www.practicalstats.com/new_classes/classes.html

2. Regression doesn't mimic a 1:1 line
Linear regression is probably the most popular statistical method used by environmental scientists. It fits a line to the relationship between a response variable (y) and one or more explanatory variables (x). It does so by minimizing the sum of squares of residuals

(departures from the fitted line) in the y direction. That last phrase is important - the residuals are in the direction of the y variable, the vertical direction. No errors in the x direction are directly accounted for. One consequence is that a regression line does not necessarily fall close to a 1:1 line in situations where the x and y data estimate the same thing.

We've noticed several reports where two measurements are made on the same object - old and new methods for measuring a chemical, suspended sediment, etc. If the two methods give comparable results, the paired values should fall close to a 1:1 line. The scientist plots the results, computes the regression equation, and almost always finds that the slope is less than 1. The departure is sometimes large enough that they conclude that the two methods therefore produce different data, with the method on the x axis giving larger values than that on the y axis. This is very likely incorrect. The departure below a slope of 1 may well be due to the method chosen to compute the line (regression) rather than to the data itself.

To illustrate, we generated data that are identical, x=y. We added random noise to both the x and y measurements. Then we ran the regression. The graph of the results is below:



The regression line is y = 77.6 + 0.457x. The slope is less than half of the known, original value of 1! This is a consequence of optimizing in only the y direction when solving for a linear fit. It is a property of least-squares regression. The departure from the intrinsic value for the slope is a function of the correlation between the variables, and proportional to the ratio of the variances of each variable. It is part of what regression does, by design. Regression is not designed to provide the 'intrinsic' slope of the relationship between the two variables.

Alternative lines are available for purposes such as recovering the intrinsic slope of the relationship between y and x. One of those is called by several names: the Line of Organic Correlation, or Orthogonal Regression, or the Reduced Major Axis (these and several more names refer to the same line). It is used most commonly in the biological sciences. The LOC slope for the data graphed above is 1.16, much closer to the true 1:1 relationship of the underlying data. Regression is not a good way to try and recover a slope equal to a theoretical or modeled expectation. It does not do this.

The bottom line: know the properties of regression and other methods that you choose to use. For more information on properties of regression and alternative lines, see our November 2005 newsletter or take one of our Applied Environmental Statistics or Time Series and Forecasting courses. Both will be offered in the first half of 2011. Watch our Upcoming Courses page or this newsletter for dates and locations.

3. News from Practical Stats

We teach each of courses directly to agencies and firms, as well as at the open-enrollment locations announced on our website. We would welcome the opportunity to come to your location. Costs per student to hold a course onsite are roughly 60% of the online registration costs, and there's no travel for your employees. If you have 10 or more persons who could use our training, contact us for a quote.

'Til next time,

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