Practical Stats Newsletter for March 2014

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

In this newsletter:
1. Upcoming Training
2. The Top 12 Tips from our Applied Environmental Statistics course
3. Webinars on demand, and a partner in training

1. Upcoming Training

In-person courses (see our Training page to register):
Applied Environmental Statistics
"Statistics, Down to Earth"
May 5-9, 2014 $1395 through April 13, $1495 after.
Indianapolis, IN
Applied Environmental Statistics covers statistical methods for analysis of air, water, soils, and bio data. It includes how to build good regression models, a myriad of hypothesis tests including the newer permutation tests, and trend analysis. It enables you to make sense of your data. A full course outline is on our website.

Webinars (see our Training page to register):
Statistics for Managers
March 20, 2014   11am-noon Mountain, 1-2 pm Eastern.   Free.
Why would my employees need to know more than that one-semester course they took 5-10 years ago in college? Think about what's changed over the last 10 years – tablets, ease of access to wi-fi, Facebook and Twitter…… Statistics has changed a lot as well. We'll hit the high points of
   • flexible tests with few requirements for validity,
   • free comprehensive statistics software,
   • new and better methods of finding the best regression line,
   • why "has there been a change in concentration?" has very little to do with a mean,
   • and much more.
Only 15 spots left as of March 5th.

To register and for more information on all of our courses and webinars, see our Training page at http://practicalstats.com/training/

2. The Top 12 Tips from our Applied Environmental Statistics course

Over the last several months we've placed the Top Twelve Tips from our Applied Environmental Statistics course on our blog site,
http://practicalstats.com/news/blog.html

If you haven't seen them yet, here's a list, with the last one spelled out in detail. There's more info on the blog, and even more in our 4 ½ day AES course coming up in May in Indianapolis.

Top Twelve Tip #1:   Plot the Data First!
Top Twelve Tip #2:   Treat Outliers Like Children: correct them when needed, but never throw them out.
Top Twelve Tip #3:   Objectives drive methods -- Totals or typical? Mass or Frequency?
Top Twelve Tip #4:   Decision Time -- Reject 'no signal' when its believability (p-value) is small
Top Twelve Tip #5:   Transforming data requires transforming solutions
Top Twelve Tip #6:   I Walk the Line -- Correlation on and off the straight and narrow
Top Twelve Tip #7:   Maximize the signal to noise ratio
Top Twelve Tip #8:   Meet the demands of regression: LNC.
Top Twelve Tip #9:   "All models are wrong; some models are useful" (G. E. P. Box), so choose the least hopelessly wrong model
Top Twelve Tip #10: Know your target for Trend Analysis.
Top Twelve Tip #11: Nondetects: never substitute. Mine the information in the proportions.
Top Twelve Tip #12: "An approximate answer to the right question is worth a great deal more than a precise answer to the wrong question." -- attributed to John Tukey.

TTT#12
Specific capacity, a standardized measure of yields of water from wells, was measured in hundreds of wells across the Appalachian region of the US in a USGS report of the late 1980s.   Is there a difference in specific capacity among the four major rock types? The figure below connects each group's mean with a straight line.
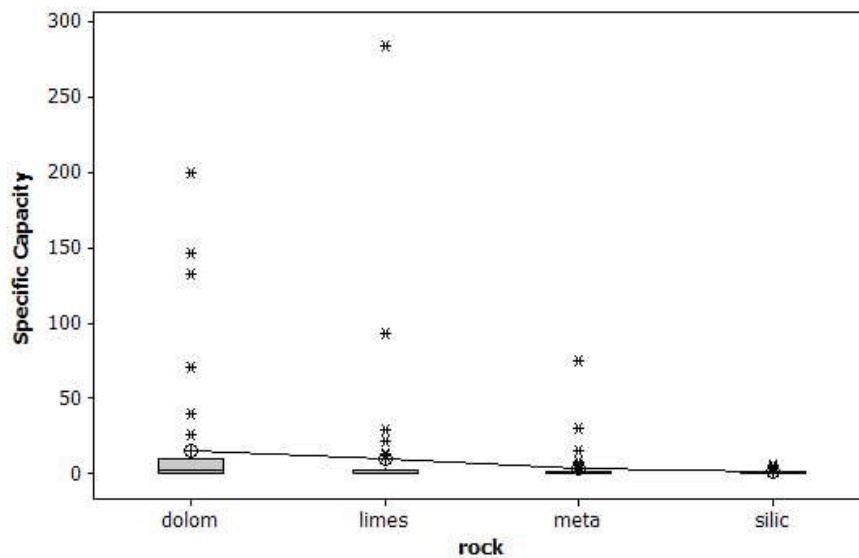
The wrong question: assuming each group's data follows a normal distribution, do the mean concentrations differ among the groups? This can be answered with Analysis of Variance, and the p-value that is produced every time ANOVA is run equals 0.06.

Conclusion: there is not sufficient evidence to say that the group means differ. However, we know that at least three of these groups do not follow a normal distribution, so there is likely a loss of power (p-values too high). But you get the same (incorrect) p-value each time the test is run.

The right question: do the mean concentrations differ? (see tip #5 on whether the mean is the best measure of center, but we will use it here). To answer this regardless of the shapes of data within groups, run a permutation test. Permutation tests scramble the group assignments for data, representing situations equally likely when the null hypothesis of no group difference is true. Between 1000 and 10,000 scrambles are usually made; differences in means and the resulting F statistics are calculated. Below is a plot of 999 F statistics as a histogram representing the null hypothesis. The one result from our data is the dashed line. The proportion of scrambled results that equal or exceed the dashed line (proportional area of the bars at or above the line) is the p-value for the permutation test, here at 0.039. The means are declared different, as the p-value is below 0.05. No assumption of data shape was required. In several trials we have obtained F-values at or above 2.512 in 3.6% to 4.0% of the cases. We've sacrificed exactness in order to obtain correctness. The means do differ, and the permutation test, unlike ANOVA, is able to see that.

F VALUES FROM THE PERMUTATIONS
(Observed F-value in red)
p-value = 0.039
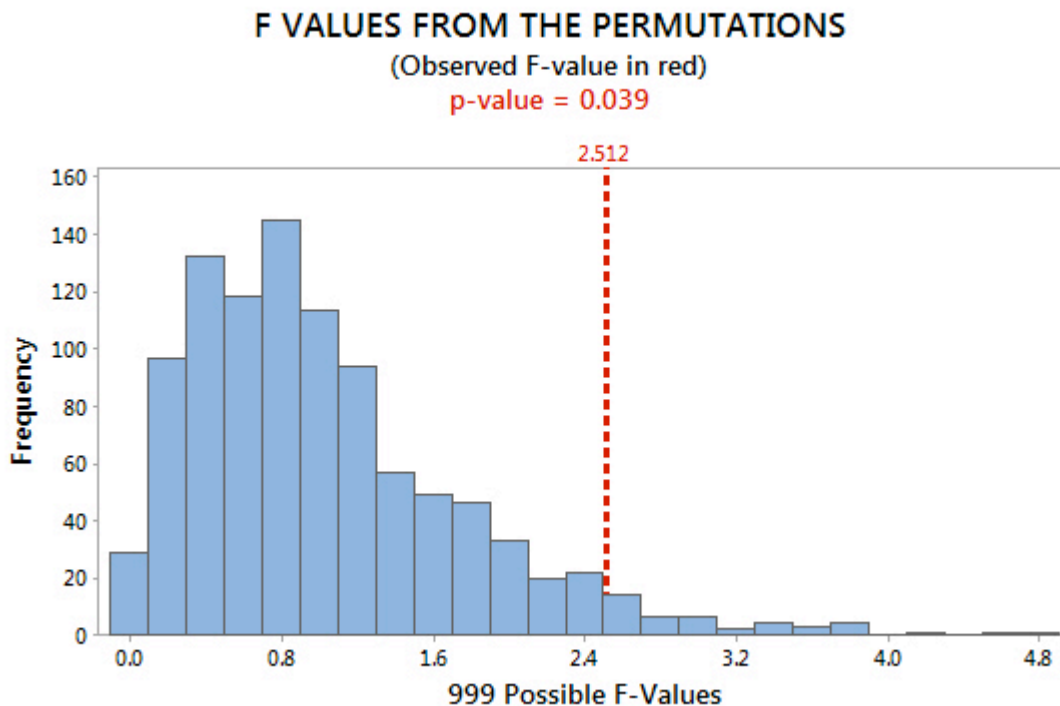
For more on permutation tests, see our January 2008 newsletter, or head to Indy in May.

3. Webinars on demand, and a partner in training

Starting in May we will offer our four-part "Nondetects And Data Analysis" webinar series on demand. You will be able to sign up and view the videos any time of the day, at

your convenience. This will also let those outside the North American time zones view the material and interact with the instructor by email. More webinars will be coming in the same format. Registration will be through our Training page within the month.

GeoEnviroLogic trains environmental professionals in risk assessment and contaminated site investigation/remediation. We will be announcing their webinars and in-person courses (usually held in Canada) in upcoming newsletters. Go to http://geoenvirologic.ca/   and you'll see that there are webinars on the basics of Contaminated Site Investigations coming up in April. In May, a course will be held on Vancouver Island – so who cares what the topic is? Check out their site.

'Til next time,

Practical Stats
-- Make sense of your data