

= Practical Stats

Statistics, down to earth

To: Attendees of the Jan. 26 webinar "Seven Perilous Errors in Environmental Statistics"
From: Dennis Helsel, Practical Stats

Hello everyone! I'm glad you were able to attend last week's webinar. You've sent in some interesting questions, to add to those asked in the chat window during the webinar itself. I've answered all of them below, with two exceptions. The exceptions were outside the scope of the webinar content, asking for statistical consulting. I do provide consulting for firms and government agencies and would be glad to set something up – just contact me at dhelsel@practicalstats.com for rates if you are interested.

1. Questions related to permutation tests.

Did you show the Welch's (non-equal variance) t-test result? Or equal variance t-test?

A. I used Welch's t-test. This was the example where the t-test produced a p-value of 0.14 while the 2-group permutation test p-value was 0.0018. Welch's test includes a correction for any unequal variances in the two groups. That version of the t-test is the best form of the t-test available – if the variances are equal, the correction goes away so there is no downside to using this form of the test, even when variances are equal. It is the form set as the default in commercial statistics software (Minitab, SAS, etc. and I'm including R in this group even though it is free). Don't use the uncorrected version.

You pointed out that a confidence interval may provide better information than a p-value. Can you clarify your recommendation to use a permutation test over a t-test in many circumstances, when a permutation test typically will only provide you a p-value? Permutation based confidence intervals can be difficult to get.

A. Bootstrap confidence intervals are resampling methods, as are permutation tests. Bootstrap intervals do not require a normal distribution, unlike t-intervals. Bootstrap intervals are computed in many software packages, including ProUCL. Whether the output from a permutation test includes an interval or not, bootstrap intervals are widely available, so I don't agree that they are 'difficult to get'.

Can permutation tests be run with covariates?

A. Yes, using the coin package in R and probably other R packages as well.

What R packages perform Permutation Tests?

A. As some answered in the chat window, there are several. The perm package performs tests on means of groups, as shown in the webinar. The coin package performs tests that include blocking and covariates. The boot package computes bootstrap intervals. The asbio package computes the permutation equivalent of a 2-way ANOVA. I cover all of these in our permutation test course.

Can you recommend any website/book/review paper which provides examples of permutation tests for people like me who are still in the "1950s statistics"?

A. There are many. For an applied book, Ken Aho's "Foundational and Applied Statistics for Biologists Using R" is a good one. Later this year you'll be able to get the new edition of "Statistical Methods in Water Resources", which will be authored by Helsel, Hirsch, Archfield, Ryberg and Gilroy, to be published as a free pdf by the US Geological Survey. It will have many examples of permutation tests, and the reasons for using them (my example in the webinar is one of them). Sign up for the free newsletter on PracticalStats.com if you'd like to get an announcement when the book is released.

Much of what you described today targeted environmental samples (water, soil, air). How much of it is applicable to "biological" samples?

A. Essentially all of it. See Ken Aho's book, referenced above, for example. The perm package of R was produced by scientists in the National Cancer Institute. Permutation methods are widely used in genomics. Those are certainly 'biological' fields of study.

I was asked if you have a reference for using $N=70$ as the criterion for a "large" sample (slide #19).

A. The ProUCL users guides states that it is 100. I think that's a bit too large. Seventy is my rule of thumb based on 37 years of environmental data analysis, and on simulations. I can point you to papers that describe that the number of observations required for the Central Limit Theorem (the point at which datasets become 'large enough' to use parametric tests on non-normal data) is a function of skewness, increasing as skewness increases. Most environmental data is 'heavily skewed'. The urban legend of 30 observations has actually nothing to do with the Central Limit Theorem, but has been passed around by environmental scientists for decades. It is way too few for the skewness of data we collect.

2. Questions related to policy and software

I notice that you're using R. One question I have (especially for someone new to the field and with a Stats degree in the last five years) is how do you counter the "Because I said so" requirement by some regulators to use a less flexible software tool, for example ProUCL?

A. As you'll see in the next question, regulators also want better statistical methods from applicants. From your (and a number of others) point of view, applicants want the freedom to perform better statistics from regulators. I don't have a magic solution other than for people to talk to and listen to one another. That's a difficult thing for human beings to do.

I managed to get a number of my colleagues to attend, but my problem is getting the consultants who send us their statistical analyses in environmental reports to do the same. Many times, consulting firms will complain that it is too difficult for them or that their software cannot do what we, the regulators, request of them. ProUCL has been very helpful for us in getting better estimations of UCLs because it is easy to use and doesn't cost the consultants anything. Can you recommend software capable of performing AIC, PRESS, C_p , and permutation test analyses that would be appropriate (cheap and easy to use, similar to ProUCL) for regulators to recommend to consultants?

A. You missed one item in your definition of appropriate – of sufficient quality. Quality assurance programs have emphasized that data collection and lab analysis must meet specified objectives. It is no less important for data interpretation procedures. The argument for 'cheap and easy to use' (meaning it can be done by someone who knows nothing about statistics) isn't convincing. EPA certifies labs to insure quality of chemical analyses. It would be cheaper and easier to hire people who don't know anything about chemistry and provide them with 'recipes' to do chemical analyses. But it wouldn't produce data of sufficient quality. Do you really want just cheap data analysis? The cost of statistical software is nothing compared to the cost of misunderstanding the results of a study spending many thousands of dollars for sampling, lab analysis, and salaries of the project personnel and reviewers and regulators. But to finally answer what you asked – PAST is free, easy to use and performs bootstrapping, permutation tests, and many other modern methods. R is free, does everything, and can be made easy to use with one day of good training.

Are permutation tests acceptable to state regulatory agencies and EPA?

A. Obviously, they should be. I'd love to see a poll to answer that question. Someone let me know the results, if that is done.

The Unified Guidance was written after permutation tests were available, but were not included. Was there a political decision to suppress their inclusion?

A. The Unified Guidance is the best government statistics guidance manual that I have seen. Yes, one weakness is the lack of permutation tests. I sincerely doubt there was any intent to suppress anything. Since these tests have more power to detect differences such as contamination than do traditional parametric tests, regulatory agencies would benefit from employing them because better protection would result. The environmental community simply hears about things 20+ years after other scientific disciplines do. It is a structural problem that I've tried to address through Practical Stats by providing better training for environmental scientists.

Does Minitab do Permutation tests for testing means of two groups?

A. Not that I'm aware of. Your best answer for "does XXX package do this?" is to call their Help Desk or Sales Department.

I regularly use software (MAROS) that lets you choose the substitution value for non-detects. Is substitution appropriate for a Mann Kendall trend analysis?

A. Not really. Choosing the substitution value is like choosing poison versus a firing squad. The end result is the same no matter what is chosen. An analogue to Mann-Kendall that handles censored data (nondetects) at multiple detection limits is the Akritas-Theil-Sen (ATS) test. See my 2012 book "Statistics for Censored Environmental Data using Minitab and R" for more detail on ATS. And on what could be done with the Mann-Kendall test with nondetects.

When I use the cenboxplot command in NADA for R on a dataset with groups with multiple reporting limits, what is the method used in estimating the quartile boxes? Is it ROS? How about the Minitab censored boxplots?

A. Yes, ROS is used for drawing boxplots both in NADA for R and in the Minitab NADA macros available on our website. There is a 'user manual' on our downloads page for NADA for R, and a read.me file is included with the Minitab macros on our website that describes all methods used.

3. Questions about training.

Will you be hosting a non-detects training course anytime soon? Thanks for the webinar!

A. There will be an online course later this year. It will be announced in our newsletter.

Which permutation tests do you recommend?

A. That would take some time to answer. It is the topic of our Permutation Test online class, which will occur this April.

Would you consider teaching a class on using "R" for environmental statistics?

A. I have been teaching one for many years -- "Applied Environmental Statistics", which doubles as an introduction to R. After teaching it in person since 1990 (though using R for only the last 10 years), I am moving our classes online, or will come to your site to teach them. I also teach a one-day R class prior to any of our two-day classes (Nondetects, Time Series, Multivariate) on request.

Is there a place I can listen online to a recording of your presentation today on Common Errors in Stats?

A. Yes, soon, at no cost. Check our website in about a week, or subscribe to our newsletter, to get the URL.

4. The definition of Mallow's Cp.

Many people asked or commented about the sign within the definition of Mallow's Cp on slide 34! It wasn't visible on the webinar slides -- not to me either, and that came as a bit of a shock. I stated at that time that there should be a plus sign between the two parts of the formula, when in truth it should be a minus sign. I apologize! Chalk that up to bad memory and an unplanned event in a live webinar ☺. Yes, the minus sign is correct, as the pdf of the webinar slides shows.

Again, thanks for attending the Practical Stats webinar "Seven Perilous Errors in Environmental Statistics". Based on your questions, I hope that it starts a dialogue between regulators and the regulated over ways to achieve high quality data analysis.

Dennis Helsel
PracticalStats.com