Practical Stats Newsletter for September 2013

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. Review of Lower-Cost Statistics Software
3. Our 2014 Schedule

1. Upcoming Training
Fall Webinars (see our Webinars page for registration):
R: Free Software for Environmental Statistics
September 25th (Wed) $50
R is open-source, free software for statistical analysis. It is sometimes seen as too
difficult for occasional users such as environmental scientists. This webinar introduces
scientists to R software with pull-down menu systems that make it no harder to use than
any other statistics software. After taking this webinar, scientists should be able to
download, install, and begin to use R software with ease. This could save your
organization a lot in software costs -- a webinar for both managers and scientists.

Permutation tests: Never worry about a normal distribution again!
October 7th (Mon) $250
Permutation tests are increasingly used to provide p-values for testing means without
assuming a normal distribution. Find out how they work, why they are such an advance
over parametric methods like t-tests and analysis of variance, and which software
performs them.

In-person courses (see our Training page for registration):
Applied Environmental Statistics
"Statistics, Down to Earth"
Nov. 18-22, 2013 $1395
Homewood Suites, Littleton, CO 80127
Applied Environmental Statistics covers all of the statistical methods required for routine
analysis of 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.

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

2. Review of Lower-Cost Statistics Software
In fall 2008 we reviewed nine lower cost stat software packages. This month we add two
more packages and update that review, five years later. Some are programs linked in

various ways to Microsoft Excel; others are standalone packages. The cost of commercial stat software continues to be high, typically $1500 or more for scientists without a corporate discount. Yet environmental scientists have sophisticated needs. The shadow of R, the free and very comprehensive statistics software used in many of our training courses, hovers over the field. However, many scientists decide they do not have time to invest in learning R, and look for something simpler. How many necessary procedures, such as regression diagnostics for building a good multiple regression equation, can be found in lower-cost software?

The 11 programs reviewed this month range in price from free to $695.

PAST 2.17c	folk.uio.no/ohammer/past/	free
Fast Statistics 2.0	www.fatesoft.com/excel	$50
Statisti-XL 1.8	www.statistiXL.com	$75
Statistician 2.0	http://www.statisticianaddin.com	$95
WinStat	www.winstat.com	$99
StatPlus 2009	http://www.analystsoft.com	$160
Winks 7	www.texasoft.com	$229
Analyze-It 3.1.5	www.analyze-it.com	$249
NCSS 9	http://www.ncss.com	$329
Stat Tools 6	http://www.palisade.com	$595
xlStat Pro	www.xlstat.com	$695

We tested each package running Windows 7 with Excel 2007, if needed. Each package performs several basic statistical procedures with a menu-driven system. All estimate percentiles, means and other summary statistics. All perform t-tests (paired and 2-sample), the Mann-Whitney, Kruskal-Wallis, and signed-rank tests. Note that in order to get the Mann-Whitney test in Fast Statistics you must perform the Kruskal-Wallis test procedure on two groups of data. All perform ANOVA and estimate regression slopes and intercepts. All compute Pearson's r correlation coefficient and draw scatterplots. All compute contingency table (chi-square) tests on a table of counts.

From there the feature sets of the packages diverge. Fast Statistics does not perform many functions necessary for analysis of environmental data. It cannot plot boxplots by groups on one plot, one of the most helpful procedures for comparing among groups of data. It cannot compute Kendall's tau correlation coefficient, the basis of several tests for trend. It cannot test for differences in variance (lack of precision) by groups. It cannot perform multiple comparison tests as a follow up to ANOVA or Kruskal-Wallis. It has no regression diagnostics such as Mallow's Cp, adjusted r-square or VIFs. It cannot plot partial plots. It does however compute regression residuals, which can be copied and pasted back into the Excel worksheet and then plotted to produce residuals plots (albeit with a great deal of work). However, it only includes the crude chi-square test for normality, so judging the distributional assumption of regression residuals, or of any original set of data, is not really possible. In sum, its feature set is inadequate for even basic analysis of environmental data.

StatistiXL and WinStat add a few more features for a small increase in price. Both perform Tukey's multiple comparisons and Spearmen's rho correlation. For some reason both perform the multivariate methods of discriminant function and factor analysis. StatistiXL displays residual versus predicted plots for regression, and is only one of two packages to perform partial plots. However, StatistiXL and WinStat only perform the KS test for normality, not a powerful test for continuous data. In short, these two packages come up short for scientific applications.

Statistician and PAST are two programs not reviewed in our 2008 survey. Statistician includes VIFs and diagnostics such as Mallow's Cp. It includes Kendall's tau correlation and Tukey's multiple comparison test. It even includes Levene's test for differing variances. Its major weakness is in graphics – it is adding numerical capabilities to your Excel software, but doesn't add graphics capabilities, and Excel is quite weak in scientific graphics. If you had another way to construct good graphs, Statistician could serve your needs for basic scientific statistics. In contrast, the free PAST software is a well-rounded scientific package producing both graphs and numerical tests. Its feature set is in the top two of the bunch. Built originally for analysis of paleontologic data (hence the name), its coverage of univariate and multivariate methods is outstanding for low-cost software, lacking only some capabilities helpful in multiple regression. A version 3 is in beta; download the most recent complete version (2.17) until the new one is ready for prime time.

Software in the $150-$300 range (Winks, Stat Plus and Analyze-It) adds better normality tests (Shapiro-Wilk or Anderson-Darling). All but Winks compute Kendall's tau; only Analyze-It includes residuals plots for multiple regression. Winks computes Dunn's nonparametric multiple comparison test when the Kruskal-Wallis test finds differences in group medians. We found questionable computation accuracy in parts of both Winks and Stats Plus. In general these three packages add a little more but still fall short of a complete set of features for environmental statistics.

Of the three most expensive packages reviewed here, NCSS, Stat Tools and xlSTAT Pro, the least expensive (NCSS) stands out. NCSS provides the most extensive feature set of all of the packages reviewed. It performs all of the features we were looking for in an environmental statistics package, except for Lowess smoothing and Kendall's tau correlation with the associated Theil-Sen line. Even there, NCSS performs an alternative robust line method, and so provides the functionality of Theil-Sen. It includes modern regression diagnostic methods, the newer tests for normality, and both parametric and nonparametric multiple comparisons. For an individual scientist without a corporate license for one of the 'major' statistics packages costing far more, NCSS would provide a fairly complete suite of methods for much lower cost.

The feature set in Stat Tools is more consistent with software in the $0-$200 range, such as Winks and StatPlus. Its website states that Stat Tools uses only Excel functions, and therein lies the problem. Excel doesn't compute nonparametric methods. It provides no regression diagnostics. And so on. The breadth of functions in Excel, and therefore in Stat Tools, is quite inadequate for scientific statistics. Indeed, the free PAST software is

much more complete. XLStat Pro, the most expensive of the 11 packages, is more complete in function coverage than all but PAST and NCSS, both of which are quite less expensive. XLStat's primary deficiency is in lacking partial regression plots, along with the Sen slope and prediction and tolerance intervals.

A few functions are missing in most of these packages. Prediction and tolerance intervals give assessments for not-yet-collected data. Bootstrapping and permutation tests are replacing older parametric methods. Power / sample size analysis help the scientist to plan future data collection. The Sen slope, the trend slope for the Mann-Kendall test for trend, is a common alternative to regression in environmental studies. Partial plots are incredibly important in building multiple regression models. The Shapiro-Wilk and Anderson-Darling tests for normality are standard statistical practice -- the KS and chi-square tests have largely been out of use since the 1950s. All of these methods can be done using R, and most with commercial software costing much more. To fully teach our Applied Environmental Statistics course, we require them.

Our newsletter archive at
http://www.practicalstats.com/news/bydate.html
now includes for download an Excel spreadsheet (that's fitting!) comparing the feature set of all 11 packages in detail. A summary evaluation is below, based on their ability to perform the procedures we find most necessary for the analysis of environmental data. The maximum rating is 6*s. One could look in your price range and find a package that had at least one star per $100. In our judgment, to use one of these software packages as your only statistics package for environmental applications, you would need a package rated at either 5 or 6 stars. You can get this, as you see below, over a wide price range.

Package	Price	Rating
PAST	free	● ● ● ● ● ●
NCSS	$329	● ● ● ● ● ●
XLStat Pro	$695	● ● ● ● ●
Analyze It	$249	● ● ●
StatistiXL	$75	● ●
Statistician	$99	● ●
Stat Plus	$160	● ●
Winks	$229	● ●
Stat Tools	$595	● ●
Fast Stats	$50	●
WinStat	$99	●

3. Our 2014 Schedule
We're pretty booked up for the rest of this year, but are putting together our course and webinar schedule for 2014. We'll hold both the Nondetects And Data Analysis webinars and Applied Environmental Statistics webinars series. We'll offer at least one of each of our in-person courses. We'll scatter some low-cost and hopefully fun webinars

throughout the year on special topics, similar to the R for Environmental Stats introductory webinar later this month. We also hold courses at your location upon arrangement, so "don't be a stranger", call or email us.

'Til next time,

Practical Stats
-- Make sense of your data