Practical Stats Newsletter for September 2015

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

In this newsletter:
1. 2016 Training
2. Tests for Difference in Variance
3. Transitions

1. 2016 Training

In-person courses:
These are likely our final open-enrollment offerings for these two courses. We will be glad to come to your site and teach them, just email us for possibilities.

Permutation Tests
January 11-12, 2016 $995 through Dec. 18, $1095 after.
Golden, Colorado
Permutation test procedures replace parametric tests like t-tests and ANOVA. Learn about these new, important methods for environmental statistics.
http://practicalstats.com/training/

Untangling Multivariate Relationships
January 13-14, 2016 $995 through Dec. 18, $1095 after.
Golden, Colorado
Untangle information in the pattern of chemicals and community structures. Multivariate methods for ecology, hydrology, geology, and other 'ologies.
http://practicalstats.com/training/

Or register for both courses. $1790 through Dec 18, 2015. $1990 after.

Webinars:
Nondetects And Data Analysis
Statistical methods for data with nondetects. Compute means, UCL95s, hypothesis tests and regression and trend equations, all without substituting 1/2DL for nondetects.
Spring 2016. A series of 4 webinars.

2. Tests for Difference in Variance
Not all of the relevant differences between groups are in their location (mean or median). Differences in variance or precision are of great interest as well. People test for difference in variance because the t-test and ANOVA require each group to have the same variance. If not, these tests have lower power and may miss differences between group means that are present. A second reason is that there may be specific interest in the precision of groups – does a new laboratory method have better precision (less variability) than a previous method? For either reason, what test best determines difference in variability between groups of data?

The classic method was Bartlett's test -- also called "The F-test", though of course there are many types of F-tests. Guidance documents today have moved away from this test to Levene's test, a newer parametric test that is far less dependent on the assumption of a normal distribution. Bartlett's test has the disturbing characteristic of rejecting the null hypothesis of equal variance whenever the data are somewhat non-normal, regardless of whether the groups differ in variance. Levene's test is far less sensitive to a normality assumption, and behaves like other parametric tests in that its p-value will increase when data are non-normal, rather than the decrease in p-value of Bartlett's test.

In 1981, Conover and others evaluated a number of tests for heteroscedasticity (unequal variance). Levene's test performed better than any other parametric test. Nonparametric tests also performed well, including the Fligner-Killeen test and the closely-related Squared-ranks test. All three tests in essence determine whether the distance of observations from the median differs between groups. Levene's test is found in most commercial software. Levene's and the Fligner-Killeen test are found in the car or stats packages of R. To observe their performance, we've applied them to two data sets. The first is constructed from a single lognormal distribution, randomly allocating observations to four different groups (see figure 1). Therefore the underlying variance of all four groups is known to be the same.

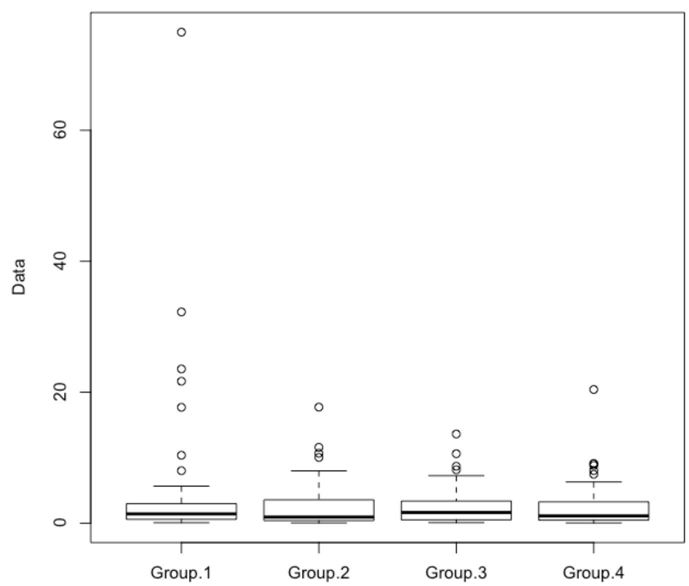

Figure 1. Data generated from the same lognormal distribution

```
> bartlett.test(Data,Group)

      Bartlett test of homogeneity of variances
data:  Data and Group
Bartlett's K-squared = 143.16, df = 3, p-value < 2.2e-16

> leveneTest(Data,Group)
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   3  2.0092  0.114
      195

> fligner.test(Data,Group)

      Fligner-Killeen test of homogeneity of variances
data:  Data and Group
Fligner-Killeen:med chi-squared = 3.9141, df = 3, p-value = 0.2709
```

Bartlett's test rejects the null hypothesis of equal variance, though the data were generated to have identical variance. It is confused because the data do not follow a normal distribution. Levene's and the Fligner-Killeen test correctly do not reject the null hypothesis of equal variance. Because most environmental data are skewed, Bartlett's test is essentially useless in our discipline. It almost always rejects the null hypothesis, whether or not that is correct.

The second example is data on specific capacity of wells in four rock types, found in Helsel and Hirsch (2002) and taken from an original report by Knopman (1990).

The variances of specific capacity in the four groups certainly appear to differ:

```
   dolomite     limestone   metamorphic   siliclastic
1546.745658   1759.180921    129.995499      1.839462
```

with the three test results:

```
> bartlett.test(spcap ~ rock, data=specapic)
 Bartlett test of homogeneity of variances
data:  spcap by rock
Bartlett's K-squared = 327.0692, df = 3, p-value < 2.2e-16

> levene.test(specapic$spcap, specapic$rock)
Levene's Test for Homogeneity of Variance
       Df  Fvalue  Pr(>F)
group   3   2.447 0.06506 .
      196

> fligner.test(spcap,rock)
 Fligner-Killeen test of homogeneity of variances
data:  spcap and rock
Fligner-Killeen:med chi-squared = 39.458, df = 3, p-value = 1.388e-08
```

Bartlett's test declares there is a difference in the variances, but this is uninformative as the data are quite skewed. It could simply be rejecting that the data follow a normal distribution. Levene's test does not find a difference between the four group variances, though close (0.065). Is this also because the data are strongly non-normal, and the test has lost power to discern differences? The Fligner-Killeen test, the only test of the three which doesn't depend on the shape of the data distribution, strongly rejects heteroscedasticity (p = 0.00000001). That indicates that non-normality did adversely affect Levene's test results.

In short, stop using Barlett's test if you haven't already done so. The Fligner-Killeen test works well in a variety of situations. Levene's test is a parametric test that loses power with non-normal data, but somewhat less than a t-test or ANOVA would. It is a good, if not perfect, test. Become familiar with the latter two tests.

References:
Conover, W.J., M.E. Johnson and M.M. Johnson, 1981. A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics 23*, 351-361.

Helsel, D. R. and R. M. Hirsch, 2002, Statistical Methods in Water Resources, *USGS Techniques of Water Resources Investigations, Book 4, Chapter A3*, 510 p. http://water.usgs.gov/pubs/twri/twri4a3/

Knopman, D. S., 1990, Factors Relating to the Water-Yielding Potential of Rocks in the Piedmont and Valley and Ridge Provinces in Pennsylvania. *USGS Water Resources Investigation Report 90-4174,* 52 p. http://pubs.usgs.gov/wri/1990/4174/report.pdf

3. Transitions
As we mentioned last month, our final open-enrollment Permutation Test and Untangling Multivariate Relationships classes will be held this January in Golden CO. The registration website has a good price for sleeping rooms at a very nice hotel only a 45 minute drive from ski and snowshoeing areas. We are fading out our open-enrollment classes that you sign up for on our site, in favor of webinars, and classes taught directly to agencies and companies that can fill them. Our first 2016 webinar series on handling nondetect data will begin in the spring. And we will eagerly come and teach any of our courses directly at your site. We've already scheduled several in-house courses such as our Applied Environmental Statistics course to the Park Service this October in Fairbanks, Alaska. Invite us over to teach to a group at your site. Or listen to our webinars.

'Til next time,

Practical Stats
-- Make sense of your data