Practical Stats Newsletter for July 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. A Better Flowchart for Hypothesis Testing
3. News from Practical Stats

1. Upcoming Courses

Applied Environmental Statistics
Sept. 13-17, 2010 — $1395 registration before August 30, 2010
Hampton Inn Tropicana — $1495 on or after August 30th.
4975 S. Dean Martin Dr.
Las Vegas, Nevada 89118-1656

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

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

2. A Better Flowchart for Hypothesis Testing

Perhaps you've seen flowcharts to decide what type of hypothesis test to employ, and it goes something like this:

Test for normal distribution. If normal, use parametric test (t-test, etc.)
If non-normal, use nonparametric test.

This type of guidance is way out of date. If you are still following a rule like this one, take a newer statistics course such as our Applied Environmental Statistics offering this coming September. There are several reasons why this type of guideline is flat out wrong.

First, the two classes of tests determine differences in different parameters. Parametric procedures are based on means, and test for differences between means. Thus the familiar t-test and Analysis of Variance procedures determine if group means are the same or different. Nonparametric tests are based on ranks, or equivalently, percentiles. Tests such as the Mann-Whitney and Kruskal-Wallis procedures look for differences in the set of percentiles, or the cdfs, of grouped data. They are frequently discussed as being tests for group medians, and they are if the major difference in groups is an overall shift upward from one group to the next. But they can be used to discern differences occurring at the upper ends of the distributions, say where the bulk of both group's data are nondetects but the upper ends above the reporting limit are generally higher in one group than the other. The two classes of tests determine different differences. When a mean is appropriate, such as when interest is in total amounts, mass, volume, etc., then a test between group means is warranted. When a 'typical value' is of interest, such as the question of whether one group gives generally higher values than the other, a test using percentiles is warranted. The choice of which of these two classes of tests to use is better based upon your objective, rather than on the observed distribution of data.

Second, data collected in the field will only be approximately normally distributed, if that. Nonparametric tests work equally as well as (have the same power as) parametric tests on data that are approximately normally distributed. So there is little need to prefer a parametric test over a nonparametric test if data are approximately normal. The 4 percent power advantage of parametric tests, often cited in basic statistics classes, is only true for data <u>exactly</u> following a normal distribution. While this may occur in simulations, it does not occur in the real world. Transforming non-normal data by logarithms or other power transformations can sometimes produce approximate normality, and so equal power for the two types of tests. And after transformation, parametric tests have determined differences in the mean logarithm, or mean square root, or whatever, rather than the mean of the original data. The result is a test on something close to the median rather than the mean. Why not instead just test for differences in medians with a nonparametric test?

Third, the statistics discipline is moving away from tests requiring distributional assumptions towards tests that do not require them. Parametric tests were developed in the early 1900s as a way of providing p-values, the statement of how likely it is for a true no-difference in means to produce data similar to your observed data. Assuming the data follow a normal distribution, the resulting test statistics follow a known distribution such as the t or F distribution, and an accurate p-value can then be computed. If the data do not follow the assumed normal distribution, the t or F distributions do not accurately describe the distribution of resulting test statistics, and the resulting p-values can be wrong. In the 1940s and 50s, p-values for nonparametric tests were developed in a different way, by enumerating all the possible values of test statistics for given sample sizes. The probability of getting the observed data when there is truly no difference (the p-value) could therefore be accurately stated without assumptions about the shape of the original data. These nonparametric tests were based on the relative positions (ranks, percentiles) of the data and so do not explicitly test for differences in means.

In the 1990s, computing power was sufficient to compute all possible values of test statistics, even those developed when testing differences in means. When the number of all possible statistic values is still too large for all to be computed in a short enough time, a random assortment of thousands of test statistics is computed and used as the basis for p-values. In neither case do these 'permutation tests' require data to follow any distribution. Nonparametric tests always computed p-values in this way, but now tests between group means can also, and so produce accurate p-values even when data are not normally distributed. This frees the scientist to choose a test based on their objectives, rather than on data characteristics. A better, modern flowchart of the decision of which test to use is:

Test for mean (total) difference? If yes -- use permutation test
If no, and instead test for typical difference -- use nonparametric or permutation test

When the early 1900s form of statistical procedures are no longer used, the observed distribution of data has nothing to do with the appropriate choice of test.

3. News from Practical Stats

a) Our journal article “Summing Nondetects”, with application to computing toxic equivalent concentrations for PCB congeners, is finally available in the current issue of Integrated Environmental Assessment and Management.
http://dx.doi.org/10.1002/ieam.31
We will send a pdf copy to you upon request.

b) A new version of the Kaplan-Meier worksheet for Excel is now available on our website: http://www.practicalstats.com/nada/nada/downloads.html

The new version 1.4 allows nondetects higher than the highest detection limit to be ignored, which they should be, bringing the Excel computation in line with how Minitab, R, and other true statistics packages compute the estimates.

'Til next time,

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