Practical Stats Newsletter for February 2006

In this newsletter:
1. Upcoming Courses
2. The Quantile Test - Testing percentages
3. For More Information....

Almost didn't get this one out on time! Its been busy around here.

1. Upcoming Courses
Both of our current courses will be taught in the next few months.

Applied Environmental Statistics March 27-31 Portland, OR
Our week-long survey of "how to make sense of your data", will be taught in Portland OR at the Mark Spencer Hotel on March 27-31. Registration is $1495, which covers all course materials, digital textbook, and data. Links for registration can be found at http://www.practicalstats.com/Pages/aes.html

Nondetects And Data Analysis May 4-5 San Jose, CA
The 2-day course on handling nondetect data using modern methods of survival analysis will be taught May 4-5 at the Hotel de Anza in San Jose, CA. This is the Thursday and Friday just prior to the National Water Quality Monitoring Conference, also in San Jose (see www.nwqmc.org for more detail). A great way to save some travel costs and come to both the course and conference in one trip! Registration for the course is found at: http://www.practicalstats.com/Pages/lto.html

Course content for each course is listed on the Practical Stats web site. For all other information, email us at ask[atsign]practicalstats.com.

2. The Quantile Test - Testing percentages

A quantile is a high-tech name for a percentile. Where quantiles range from 0 to 1 (with the median = the 0.5 quantile), percentiles range from 0 to 100 percent (with the median = 50th percentile).

The quantile test determines whether a specified quantile q is similar or different from a specified value. In environmental applications, the test is most commonly used to determine whether a quantile exceeds a legal standard. For example, Section 303(d) of the Clean Water Act declares that stream waters be listed as 'impaired' if more than 10% of samples exceed a standard. This is equivalent to stating that the 90th percentile (0.9 quantile) shall not exceed the standard. Usually determined based only on the observed frequency of exceedances, more than 10% of (especially a small number of) samples could exceed the standard just by chance, when the true percentage in the stream is less than 10%. The stream would be falsely listed as impaired. The quantile test can instead be used to determine whether significantly more than 10% of samples exceed the standard.

The null hypothesis is that the quantile equals or is less than the standard. The one-sided alternative is then that it exceeds the standard. Computation is simple: compare each observation to the standard. Score a + if it exceeds the standard, and a - if the observation is less than the standard. For the null hypothesis to be true, 10% or fewer +s would be expected. Compare the proportion of observed +s to a table of the binomial distribution, and if sufficiently more than the expected 10% are +s, reject the null hypothesis.

Example
In samples from a stream receiving treated wastewaters, do more than 10% of fecal coli concentrations exceed 400?

Null Hypothesis:		the 90th percentile is 400 ug/L or less
	OR	10% or fewer samples are above 400.
Alternative Hyp:		the 90th percentile is greater than 400
	OR	more than 10% of counts are above 400.

Minitab computes the quantile test with its Stat > Basic Statistics > 1-Proportion command. The user must enter the column number for the data and the (1-q)*100 percent of data expected to exceed the standard. Other software packages have similar implementations.

In our course Applied Environmental Statistics, we include a macro with our "Comparing Data to Standards" lecture that performs this test and gives output in a more user-friendly form:

```
---------------------------------------------------
Is the probability of exceeding 400 greater than 10 percent ????

Test and CI for One Proportion
Test of p = 0.1 vs p > 0.1
                                                         Exact
Sample      X      N  Sample p  95.0% Lower Bound  P-Value
1           7     24  0.291667           0.145686    0.007

There were 7 out of 24 observations greater than 400.
To be at least 95% sure, you would have to observe
more than 5 observations greater than 400.
---------------------------------------------------
```

As $p < 0.05$, reject the null hypothesis. The 90th percentile for these data is found to be greater than 400. A 95% lower bound on the percent of observations that are 400 or above is 0.145 (14.5%). More than 10% exceed the standard.

This test is more stringent than simply observing whether greater than 10% of observations exceed a standard. Twelve percent of observations, for example, could exceed a standard in a specific set of collected data even when the true percent of stream concentrations that exceed the standard is less than 10%. The test requires sufficient evidence to demonstrate exceedance beyond what might occur by chance alone.

3. For More Information....

For more information on the mechanics of the quantile test, sometimes called the binomial test, see Practical Nonparametric Statistics (good title!) by W.J. Conover, Wiley, 1999 (third edition).

For more information on using hypothesis tests with legal standards, see Smith, Ye, Hughes and Shabman (2001), Statistical assessment of violations of water quality standards under Section 303(d) of the Clean Water Act. Environmental Science and Technology 35, pages 606-612. They show that using the simple observed percent of exceedances without a test (they call this the "raw score") results in false declarations of impairment -- Type I errrors. They show that the Type I errors are much lower using the quantile (binomial) test, though there are tradeoffs, as always. See their article for much more detail.

If you'd benefit from more personal tutoring on this issue, take our Applied Environmental Statistics course at the end of March, or at another time. We'd love to see you there.

'Til next time,

Practical Stats
http://www.practicalstats.com

-- Make sense of your data

.