Practical Stats Newsletter for September 2009

Subscribe and Unsubscribe: http://practicalstats.com/news
All of our past newsletters: http://practicalstats.com/news

In this newsletter:

1. Upcoming Courses

We've just completed a grueling schedule of summer/September courses, both open enrollment classes advertised on our website, and classes taught directly to agencies or organizations. I'm squeezing this newsletter into September, and will catch up by releasing an October newsletter as well. October's will address statistics using Excel. It will also provide details on the first of the 2010 class offerings.

We will be offering our new Time Series and Forecasting class in Golden, Colorado the week of Feb. 15, either Tues-Wed or Thurs-Fri. Another 2-day class, probably our Nondetects And Data Analysis class, will be offered on the other two days. Stay tuned.

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

2. Spearman's rho vs. Kendall's tau nonparametric correlation coefficients

It was brought to our attention at a recent Applied Environmental Statistics class that some persons believe there are advantages to using Spearman's rho over Kendall's tau as a nonparametric measure of correlation. In short, we know of none. Spearman's rho is one example of a procedure described as "perform a parametric test on the ranks of data". These type of procedures were discussed by Conover and Iman in 1981 (Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. The American Statistician, 35, 124-129). These types of tests sometimes result in p-values that are too low -- for the null hypothesis they result in alpha values that are inaccurately lower than 5 percent. Many consider these types of procedures (rho) to be approximations to the true nonparametric equivalent (Kendall's tau). In addition, an entire class of trend tests (Mann-Kendall test, Seasonal Kendall test, Regional Kendall test, partial Mann-Kendall test) have been built using Kendall's tau.

Gottfried Noether, a famous statistician specializing in nonparametric methods, had these comments in the early 1980s to make about the two correlation methods. He prefers Kendall's tau because it is directly interpretable in terms of the data, because tau can be directly related to the odds of increasing versus decreasing values. There is no such 'physical meaning' to Spearman's rho. You can find his full discussion at
http://www.rsscse-edu.org.uk/tsj/bts/noether/text.html

Why Kendall Tau?
G. E. NOETHER
[edited here for brevity]

In the 1980 issue of TEACHING STATISTICS, D. Griffiths supports the use of the Spearman rank correlation coefficient on the grounds that "it is the one which is commonly used." The claim may well be true. But it is a poor excuse for ignoring practical and pedagogical advantages of the Kendall coefficient.

All too often statistics users compute a quantity like a correlation coefficient without asking what the quantity means. In the Griffiths discussion, what operational interpretation can we associate with the value r for the Spearman correlation coefficient between the density of public houses and the density of places of worship? We are given a formal explanation for the rather awkward-looking algebraic structure of the Spearman coefficient, but we are not told anything about how to interpret the result of our computations.

The facts are that it is no easy matter to assign an operational interpretation to the Spearman coefficient. The Kendall coefficient, on the other hand, has an intuitively simple interpretation. What is more, its algebraic structure is much simpler than that of the Spearman coefficient. It can even be computed from the actual observations without first converting them to ranks.

A correlation coefficient is intended to measure "strength of relationship". But different correlation coefficients measure strength of relationship in different ways. A product moment coefficient, a Spearman coefficient, and a Kendall coefficient, all equal to 1/3 mean three rather different things. Only the Kendall coefficient has a simple interpretation.

On practical and pedagogical grounds, the Kendall coefficient has substantial advantages over the Spearman coefficient. This writer is not aware of any theoretical reasons for preferring the Spearman coefficient. Quite the contrary, Kendall S has much greater universality, so much so that a good deal of what is called non-parametric statistics can be built around S. A recent book by C. Leach does exactly that.

3. New Time Series class for frequently-collected data
See last month?s newsletter for a description of our new Time Series and Forecasting course. More and more places are using 5- or 15-minute data recorded electronically, or remotely-sensed or other data collected almost continually. These types of data usually violate the assumption of independence built into all common statistical procedures, including regression and hypothesis tests. This class instructs students how to correctly analyze these types of data.

'Til next time,

Practical Stats
-- Make sense of your data