

Technical

Long-Form/Newsletter



Correlational Analysis Part 2: Two Tails and Noise Reduction

In an earlier post, we looked at using correlational analysis and Pearson's Coefficient as statistical tools for predicting outcomes. As part of your performance management process, correlational analysis can help you predict the best likelihood for success based on past activities and even identify opportunities for applying resources in more efficient ways.

As a quick review, correlational analysis is essentially an equation based on two statistical variables, identified as P and R. R (or Pearson's coefficient) is the measure of how two elements are related and P is the measure of the data's significance. In other words, what is the strength of the relationship between the effort and the outcome?

Not all correlational analysis is created equally, though. A basic analysis will provide insights, but more advanced analysis techniques like "two-tailed" analysis and noise reduction will show the impact of your efforts even more clearly .

Advanced Pearson's Coefficient: A Tale of Two Tails

When we talk about advanced Pearson's coefficient calculations, it's helpful to start by thinking about a bell curve.

The narrow edges of the curve is where we find our analytical insights because that's where people performing significantly above or below average lie. In workforce development, for example, when people who have received specific interview training lie further from the center point on a chart of successful hires, it can be inferred that that class has a direct impact on hiring.

However, analysts have 2 options when determining how far from that center point you want to look for significance (and in which direction). The closer to the center point that you define as

having statistical impact, the more likely it is that you will show a correlation. However, that doesn't necessarily mean that the correlation is correct. In a "one-tailed" analysis, the correlation is drawn from only one side of the curve:

This means that your analysis will only ever find positive correlations, and it will find them more frequently. This is useful when you are 100% certain that your efforts will either have a positive outcome or none at all, but the in world of social services that is rarely the case.

Fishing for Significance: Imagine that you are a "fisherman" who's out to catch important data points (instead of fish). This method of analysis is a bit like doing your fishing with a spear at a specific spot on the river – you'll catch all the data points (big or small) that go through that sandbar, but who knows what you're missing in the other areas?

The better method of performing correlational analysis for workforce development and other social services organizations is to use "two tails." In this method, significance is sought on both ends of the curve. The lowered area of significance means that you are less likely to get a false correlation and it means that both positive and negative correlations will be found.

Fishing for Significance, Part 2: Two-tailed analysis is like fishing with a net rather than a spear. You'll end up with all of the big data points, no matter where they're swimming on the river – even if it means you end up missing some of the smaller (less significant) ones.

Social Solutions' ETO software follows the two-tailed approach. The math involved is obviously more complicated because it is drawing from more data points to find the significant correlations. But the reliability of and additional insight offered by the results is invaluable.

Muting the "Noise" of Correlational Analysis

There is one other issue related to correlational analysis that makes it challenging to use effectively, and that's the issue of "noise." An extremely large data set might easily involve hundreds of data points – maybe more. Finding the meaningful, and actionable, correlations in that universe of points (especially as they start to clump together) is extremely difficult and time-intensive.

The right performance management software can compensate for a huge volume of data for you, though. Through a technique called discretization or "data binning," the scatter plot is divided into bins of equal width. Representative values (like the average) from each bin are used to show the correlations. This method filters out the less-meaningful correlations without overlooking any available data.

The above screenshot from ETO uses sample information (not actual client data) to show how data binning can be used to cut through the clutter and find areas for improvement in a simulated workforce development environment. The chart represents an accumulation of hundreds of data points that have been reduced to just a few-dozen meaningful elements. This simplification means that decisions can be made quickly about what steps should be taken next. For this example, that means finding out which referrals are more (or less, in this case) helpful for program participants.

Discovering what's working (and what's not) is a key element of your performance management process. The best way to do that is with correlational analysis, especially two tailed analysis that includes noise reduction to make the results easier to see.

Social Solutions provides ETO performance management software systems that streamline WIA reporting and simplify correlational analysis, allowing social services organizations to draw data points from the system and simply plug them in to find meaningful correlations.