SCIENTIFIC AMERICAN™

Permanent Address: **http://www.scientificamerican.com/article.cfm?id=the-mind-reading-salmon**

The Mind-Reading Salmon: The True Meaning of Statistical Significance

By Charles Seife | Friday, August 12, 2011 | 16 comments

See Inside

If you want to convince the world that a fish can sense your emotions, only one statistical measure will suffice: the p-value.

The p-value is an all-purpose measure that scientists often use to determine whether or not an experimental result is "statistically significant." Unfortunately, sometimes the test does not work as advertised, and researchers imbue an observation with great significance when in fact it might be a worthless fluke.

Say you've performed a scientific experiment testing a new heart attack drug against a placebo. At the end of the trial, you compare the two groups. Lo and behold, the patients who took the drug had fewer heart attacks than those who took the placebo. Success! The drug works!

Well, maybe not. There is a 50 percent chance that even if the drug is completely ineffective, patients taking it will do better than those taking the placebo. (After all, one group has to do better than the other; it's a toss-up whether the drug group or placebo group will come up on top.)

The p-value puts a number on the effects of randomness. It is the probability of seeing a positive experimental outcome even if your hypothesis is wrong. A long-standing convention in many scientific fields is that any result with a p-value below 0.05 is deemed statistically significant. An arbitrary convention, it is often the wrong one. When you make a comparison of an ineffective drug to a placebo, you will typically get a statistically significant result one time out of 20. And if you make 20 such comparisons in a scientific paper, on average, you will get one significant result with a p-value less than 0.05—even when the drug does not work.

Many scientific papers make 20 or 40 or even hundreds of comparisons. In such cases, researchers who do not adjust the standard p-value threshold of 0.05 are virtually guaranteed to find statistical significance in results that are meaningless statistical flukes. A study that ran in the February issue of the American Journal of Clinical Nutrition tested dozens of compounds and concluded that those found in blueberries lower the risk of high blood pressure, with a p-value of 0.03. But the researchers looked at so many compounds and made so many comparisons (more than 50), that it was almost a sure thing that some of the p-values in the paper would be less than 0.05 just by chance.

The same applies to a well-publicized study that a team of neuroscientists once conducted on a salmon. When they presented the fish with pictures of people expressing emotions, regions of the salmon's brain lit up. The result was statistically significant with a p-value of less than 0.001; however, as the researchers argued, there are so many possible patterns that a statistically significant result was virtually guaranteed, so the result was totally worthless. p-value notwithstanding, there was no way that the fish could have reacted to human emotions. The salmon in the fMRI happened to be dead.