Statistics as unbearable longing

Eric Blair

26 August 2008

[Introductory logistics: This is part two; see the beginning in entry #238. Also, I just finished writing a textbook on statistical computing—about three hours ago as I write this, and I'm half relieved about getting my time back, half anxious about how long it will take before I get my first complaint letter about how the betas on page 317 should be in boldface, and half amazed that I managed to write something like that. I suppose this is a bit of catharsis after writing hundreds of pages of math and data technique.]

Statistics—by which I mean all of mathematical inquiry aimed at explaining the real world, and sometimes even plain measurement—has fundamental failings for its intended purpose of allowing us humans to better understand the world.

Picking up from last time, statistics can never prove. The real world is uncertain and messy, but mathematics is pure and certain and unwavering. Mix the two together, and what do you get? An uncertain mess.

Our language is inclined toward our desire to accept things as true. Statistical language makes at least a half-assed effort—maybe even a three-quarter–assed effort—to retain skepticism at all times. A good and pedantic hypothesis test comes up with two outcomes: *reject* or *fail to reject*. This seems appropriately skeptical, and it means that if somebody is snooping around in the data beforehand, and inappropriately failed to reject, then it's no big deal: we just learned nothing from that experiment (in formal stats language, the test had insufficient power).

But this fine point breaks down at every opportunity, because we long for certainty, and statistics just won't give it to us.

The first breakdown [and non-math–geeks are welcome to skip to the next paragraph] is that the system is asymmetric regarding what should be symmetric hypotheses. Given two variables, we typically wind up with H_0: the variables are equal, and H_1: the variables are not equal. The above reject/fail-to-reject language typically refers to H_0. Failing to reject H_0 is appropriately indefinite, but rejecting H_0 is definite, not-squishy language stating that we know the variables are not equal, because we reject the claim of equality. In an ideal world, perhaps we'd say that the test *fails to reject* and *fails to accept*. Then when we fail to reject one hypothesis, we're failing to accept the alternate, which has the same level of confidence on both sides.

The second problem is that even that little bit of legalese, *fail to reject*, is hard to keep in place for long—it turns into *accept* even in many stats textbooks, especially the ones with a 'tude that tries to make statistics fun. And don't expect the phrase to ever appear in the newspaper: my brief search of the NYT turns up one op-ed making exactly the point I'm making here, one correct use of the phrase, and assorted cruft. The longing for certainty is just too strong to let weak language stand.

But there are benefits to accepting the weakness of statistics. If we bear in mind that statistics can not prove, then my lament last time about how all our published positive results are doomed to be too confident is not so bad. An article with a solid result from a statistical test should simply slightly raise our confidence in whatever they found. If the research was especially carefully conducted, then it will raise our confidence a lot. Perhaps another article will come by next year that bolsters our belief or cuts it down a bit.

So after incredibly tedious and careful mathematical contortions, the best result we can get is that the human reader believes the claim a little bit more.

Some people are disappointed by the inability of mathematics to touch the core of what we as humans want, and just reject the entire project. Forget all those studies: they either tell us what we already know or are a pile of sophistry that will be contradicted next week. That's extreme. Our measurements are never perfect, but we make them. We're surrounded by black boxes that we'll never be able to crack open, and situations where we know any measurement will be imprecise. Despite knowing that we'll never be able to fully and truly understand anything, we still try.

Correlation is not causation, but neither is anything else But to really make our model good, we need to tell a story, almost invariably of the form A *causes* B. Unfortunately, *statistics has no concept of causality*.

This is one of those philosophy of science things that you could expound on forever, though I won't go into it too deeply here. But the concept of causation happens only inside the human brain. It's not something we can measure, perhaps with a causality ruler (or a more portable causality tape), and then write down that A causes B with 3.2 causal units, but C causes B with 8.714 causal units. There are intuitive ways to measure a causal claim, like saying that if A always comes before B, then A causes B; in direct correspondence, there are easy ways to break such a simple measure, like how Christmas card sales cause Christmas.

But people like stories. As kids, we're taught how the world works via causal stories, that were not just a list of incidents but were a chain of events. Because granny was ill, Ms Hood took her basket of food and went walking over the river and through the woods; because the wolf was evil, he conspired to eat Ms Hood; because Ms Hood was virtuous, she was saved. A story where a bunch of unconnected, seemingly random things happen is just not satisfying, and correlation without causation is dissatisfying in exactly the same way.

You could take the basic intuition about how causality works and build machinery to draw causal flowcharts, which give a wealth of means to reject the flowchart; look up structural equation modeling or read Perl [2000]. But apply the above rule that statistics can never prove a model of the world: statistics can never prove a causal model of the world—and this case is only worse because we're not even entirely certain about how to measure or even identify causality. As with any model, stats can bolster or cut down our confidence in the causal claim, but that's where it ends.

Of course, people fake it all the time. You will rarely if ever find a newspaper article declaring a correlation without strongly implying (if not directly stating) that the statistical model showed a causal link. Get your favorite researcher drunk and he or she will stop talking about correlations and start talking about causation, even though everybody in the room knows that it's just a mathematical mirage.

There's so much that we want to understand about our world and those around us that we'll never come close to. We're just guessing at reality based on our sadly limited information, and nothing makes that more evident and visceral than statistics.

References

Judea Perl. *Causality*. Cambridge University Press, March 2000.