Disputes in statistical analyses

We sometimes read about disputes over the correctness of some statistical analysis. This note discusses the underlying causes of such disputes, in a non-technical way.

Suppose that we toss two coins, and we ask what the probability is that both coins come up heads. To determine this probability, we will make two reasonable assumptions: (i) the probability that a coin comes up heads is ½ and (ii) one toss has no effect on the other toss. Then the probability of two heads is calculated to be $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

In general, any statistical analysis will consist of two phases. The first phase is to make some assumptions about the process that generates the data (in our example, we made two such assumptions). The second phase is to do some mathematical calculations (in our example, we did a simple multiplication). The calculation phase of a statistical analysis has a special feature: there cannot be opinions or disputes about its validity. For instance, the equation "2 + 2 = 37" is simply incorrect. It is not merely someone's opinion that it is incorrect; rather, the incorrectness is absolute. (There can, though, still be mistakes made in the calculation phase, just as there can be mistakes made in ordinary arithmetic, but that is not the same as having opinions about it.)

The assumption-making phase of a statistical analysis *can* be disputed, unlike the calculation phase. That is, different people might reasonably have different opinions about the validity of the assumptions. A common reason for this is that the process that generates the data is not well understood.

Even when the process that generates the data is well understood, though, the statistical assumptions are often not exactly correct. For instance, in our coin-tossing example, the process by which a coin flips through the air is well understood; yet the assumption that a coin comes up heads with probability ½ is only an approximation. In reality, the two sides of the coin are not the same (which affects air resistance); so the chances that the two sides come up will not be the same. It might really be, for instance, that the probability that the coin comes up heads is 0.500001 and the probability that it comes up tails is 0.499999. Of course, in almost all practical applications, this difference will not matter, and our assumption of a probability of ½ will be fine.

Or so we hope. To properly judge whether or not the assumptions are acceptable, we need knowledge of the relevant statistics *and* of the physics of how the coin moves through the air when it is tossed. More generally, in any statistical analysis, judging the reasonableness of the assumptions usually requires some knowledge of both the relevant statistics and the field of science that studies whatever process is generating the data. This sometimes poses a problem, because there might be few people in the world who have the requisite dual knowledge.

In science, it is fairly common to find researchers criticizing the statistical analyses of other researchers. Despite what was discussed above, however, not all those criticisms are about statistical assumptions. Instead, many of the criticisms are about statistical calculations. Typically, this is because one researcher has made a mistake in a calculation. It is disconcertingly easy to make mistakes in statistical calculations, and all researchers will sometimes make "obvious" mistakes (no opprobrium is due).

Very occasionally, two researchers will still be in dispute about the validity of a statistical calculation after discussing it. This indicates a serious problem: it typically means that one (or both) of the researchers is tendentiously denying due scholarliness.

This work was kindly reviewed by David Banks (editor, *JASA*) and Geert Verbeke (editor, *JRSS*).