

The principles of uncertainty

Statistics and real life? It's messy, writes **Andrew Gelman** on Ian Stewart's analysis.

Uncertainty “isn't always bad”, begins *Do Dice Play God?*, the latest book from celebrated mathematics writer Ian Stewart. It ends: “The future is uncertain, but the science of uncertainty is the science of the future.” In between, Stewart discusses topics from mathematics to meteorology, in which accepting uncertainty is necessary to understand how the world works. He touches on probability theory and chaos (the subject of his 1989 *Does God Play Dice?*). And he probes the connection between quantum entanglement and communication, with interesting excursions into the history of mathematics, gambling and science.

“My favourite aspect of the book is the connections it makes in a sweeping voyage from familiar (to me) paradoxes, through modelling in human affairs, up to modern ideas in coding and much more. We get a sense of the different “ages of uncertainty”, as Stewart puts it.

But not all the examples work so well. The book's main weakness, from my perspective, is its assumption that mathematical models apply directly to real life, without recognition of how messy real data are. That is something I'm particularly aware of, because it is the business of my field — applied statistics.

For example, after a discussion of uncertainty, surveys and random sampling, Stewart writes, “Exit polls, where people are asked who they voted for soon after they cast their vote, are often very accurate, giving the correct result long before the official vote count reveals it.” This is incorrect. Raw exit polls are not directly useful. Before they are shared with the public, the data need to be adjusted for non-response, to match voter demographics and election outcomes. The raw results are never even reported. The true value of the exit poll is not that it can provide an accurate early vote tally, but that it gives a sense of who voted for which parties once the election is over.

It is also disappointing to see Stewart trotting out familiar misconceptions of hypothesis testing, the statistical theory underlying

the familiar $P < 0.05$ (in which P signifies probability) so often used in this and other journals to indicate that a certain empirical result has a statistical seal of approval.

Here's how Stewart puts it in the context of an otherwise characteristically clearly described example of counts of births of boys and girls: “The upshot here is that $p = 0.05$, so there's only a 5% probability that such extreme values arise by chance”; thus, “we're 95% confident that the null hypothesis is wrong, and we accept the alternative hypothesis”. (In general, the null hypothesis is a comparison point in a statistical analysis. Here, it is the supposition that births of boys and girls occur with equal probabilities; in fact, the birth of a boy is slightly more likely.)



Do Dice Play God?: The Mathematics of Uncertainty
IAN STEWART
Profile (2019)

Stewart makes the common mathematical error of transposing the probabilities. He interprets 0.05 as the probability that the hypothesis is true; it is actually a statement about how probable it would be to see the results or something more extreme if the null hypothesis were true (it isn't, in this case).

Later, he erroneously states that a confidence interval indicates “the level of confidence in the results”; in fact, it is a statistical procedure for expressing uncertainty, or a range of values consistent with the data.

Stewart does, however, discuss a mistake all too common among researchers and students: using the statistical rejection of a straw-man null hypothesis to validate a scientific claim about the real world. In simple cases, this might not be an issue. In rejecting the model that births of boys and of girls are equally likely, we at the same time learn the general fact of likelier boy births. But this kind of learning-by-rejection can fail in more complicated settings. A null hypothesis is extremely specific, and the alternative



Drug testing is reliant on statistical models.

KRISZTIAN BOCSI/BLOOMBERG VIA

includes not just one correct answer, but all other possibilities.

In a medical experiment, the null hypothesis might be that a new drug has no effect. But the hypothesis will come packaged in a statistical model that assumes that there is zero systematic error. This is not necessarily true: errors can arise even in a randomized, blinded study, for example if some participants work out which treatment group they have been assigned to. This can lead to rejection of the null hypothesis even when the new drug has no effect — as can other complexities, such as unmodelled measurement error.

To say that $P=0.05$ should lead to acceptance of the alternative hypothesis is tempting — a few million scientists do it every year. But it is wrong, and has led to replication crises in many areas of the social, behavioural and biological sciences.

Statistics — to paraphrase Homer Simpson's thoughts on alcohol — are the cause of, and solution to, all of science's problems. Many difficulties have been associated with the misuse of statistics to make inappropriately strong claims from noisy data, but I don't think that the solution is to abandon formal statistics. Variation and uncertainty are inherent in modern science. Rather, we need to go deeper in our statistical modelling. For example, in polling, we accept that we cannot get clean randomized or representative sampling, so we gather the data necessary to adjust our sample to match the population.

As I recall the baseball analyst Bill James writing somewhere, the alternative to good statistics is not no statistics: it's bad statistics. We must design our surveys, our clinical trials and our meteorological studies with an eye to eliminating potential biases, and we must adjust the resulting data to make up the biases that remain. If we do not, people can take the numbers that are available and draw all sorts of misleading conclusions. One thing I like about Stewart's book is that he faces some of these challenges directly.

In a sense, the answer to Stewart's question, "Do dice play god?", is yes. Probability is an unreasonably effective mathematical model for uncertainty in so many areas of life. I believe that a key future development in the science of uncertainty will be tools to ensure that the adjustments we need to make to data are more transparent and easily understood. And we will develop this understanding, in part, through mathematical and historical examples of the sort discussed in this stimulating book. ■

Andrew Gelman is a professor of statistics and political science at Columbia University in New York City, and has written books on applied regression analysis, Bayesian statistics, statistics education and American politics.
e-mail: gelman@stat.columbia.edu