

Lies, damn lies and analytics?

Over 200 years ago Benjamin Disraeli¹ said: ‘There are three kinds of lies: lies, damned lies and statistics’. This aphorism is still immensely popular. Every time someone wants to ridicule statistics and statistical thinking, he quotes Disraeli. Then, with more relish, he tells you that other one linking statistics to a bikini.

Most laypersons think of statistics as something that starts with ‘mean–mode–median’ and then wanders off to ‘variance or standard deviation’ with probability coming in somewhere. Very few realize, or accept, that statistics can be a powerful analytical tool. Since it is hard to change perceptions, statistical analysis must be sold with a new name: ‘analytics’. This new name works better; the top management ‘buys’ it more easily. And, one day, when they stop buying ‘business analytics’, it will be re-packaged as ‘business intelligence’.

The major reason why analytics has risen appreciably in public esteem is because computers and computer software can now deliver their power much more effectively. A market researcher, worried that his customer is no longer excited by percentages and pie charts, is now asking for key driver analysis; which really is regression in the ‘analytics’ garb, and which can now be accomplished in a split second. The financial analyst, who does not want to look beyond his Excel spreadsheet, is asking how he can make more intelligent profit predictions. He has seven deterministic variables (such as sales during last three months, or the Sensex value on 1 January 2007) that he can manipulate well, but also three stochastic variables (e.g. market volatility) that he just cannot get a handle on. So why not simulate these variables, especially now that a very powerful pseudo-random number generator like the Mersenne twister is available?

Or look at all those telecom operators who have mushroomed in the last five years. During the heady, early days it was all about selling an exciting dream to the common man. It was about telling him how he can carry an affordable mobile telephone in his pocket with no ugly wires sticking out. But, with growing competition, the battle now is to achieve

the highest ARPU (average revenue per unit). So operators now worry about ‘churn’ (what fraction of your connections will you lose to your rival), revenue ‘leakages’ (when someone avails of a service without paying for it) and customer ‘segmentation’ (is your user a chic socialite, a rich man’s spoiled daughter who calls up every hour from Greece, or a taxi driver who only uses his handset to make ‘missed calls?’)... all situations crying out for ‘analytics’.

This clamour for analytics can only grow as computer networks proliferate, and more and more data becomes digital. We are currently living through times when data capture and transmission is getting continually cheaper and more efficient. Soon we will face a situation where there is a data overload, but no one knows what to do with all these data. The retailers have their SCM (supply chain management) data, the ‘pharma’ companies have their clinical trials data, the customs and excise boards have their data, the financial institutions have their investor data and airline operators have their data about flights and passenger load factors! Only statistical analysis, or analytics, appears to be capable of spotting a method in this madness.

Indeed analytics can adapt itself admirably to every offering on the plate; for example, cricket that is very much the current flavour. The cricket fan is a rather curious specimen: his surreptitious eye is glancing at the score, asking rates, or D/L par scores, but his public pretence is to say that he only cares for the artistry or lazy elegance of a batsman; not cold and unfriendly numbers².

One example of cricket analytics that we are attempting right now involves the ‘pressure index (PI)’; to quantify the ball-by-ball probability of a team chasing a target in an ODI game. If the odds are even at the start of a chase, the PI is exactly 100. If a batsman hits four boundaries in the first over, the PI might drop to 93 or 94. If he gets out in the following over, the PI would rise to 102 or 105. We find that the PI gives a realistic picture of the ‘pressure’ felt by the chasing team.

Our formulation essentially tracks the ratio of ‘runs needed to win’ and ‘runs

that the team can potentially score’ on a ball-by-ball basis. The potential score is derived from a resource table not dissimilar to the D/L or Jayadevan tables. The ratio is wrapped in a symmetric function, that also ensures that the pressure approaches zero if a win is imminent, and 200 if defeat is imminent. We also add a tweaking parameter to make the PI sufficiently perky and responsive.

The reaction to the PI has been mixed so far; and of course many have called it lies and damn lies. We were also asked: ‘how can one number hope to capture the variability and romance of ODI match situations?’.

It is indeed ridiculous to make any such claim; no single number is usually bestowed with such magical properties. And, yet, it is amazing how often we are guilty of making such single numbers hallowed and sacrosanct: the impact factor of publications is one number that enjoys an undeserved halo; the ARPU in the telecom sector is perhaps another, and for many years CSIR’s key parameter to measure the performance of its R&D establishments was simply its external cash flow (ECF).

1. There’s some confusion about who really said this; it has been also attributed to Mark Twain and a French politician named Labouchère.
2. I am reminded of the time when the Indian leg spinner Narendra Hirwani was playing an ODI match in Australia in 1991–92. His batting ability was pathetic, and he could bat only at No. 11. As Hirwani came out to bat, his fellow Australian radio commentator asked Harsha Bhogle: ‘How bad is Hirwani as a batsman?’. Harsha chose to go analytical. ‘Well, Tim, if you were to make a playing eleven of only No. 11 batsmen, then Hirwani would bat No. 11 in this team!’.

Srinivas Bhogle, Cranes Software International Ltd, 4th Floor, Shankaranarayana Building, 25, M.G. Road, Bangalore 560 001, India
e-mail: srinivas.bhogle@cranesoftware.com