

Data in a post-truth age

Trust in official statistics is vital for democracy — the new policy must avoid centralisation



SONALDE DESAI

David Spiegelhalter, president of Royal Statistical Society in the U.K., gave a most unusual presidential address in 2017. Instead of talking about esoteric statistical techniques, he talked about declining trust in numbers in a post-truth society bombarded by fake news and alternative facts. He recommended to the statistical community that the best way of inspiring trust was to be trustworthy by demonstrating competence, reliability and honesty.

India has been fortunate in inheriting a statistical system from stalwarts like P.C. Mahalanobis and C.R. Rao that has historically demonstrated all three. However, with the growing demand for statistics and increasingly challenging data collection environment, the move by Ministry of Statistics and Programme Implementation (MOSPI) towards developing a National Policy on Official Statistics is most welcome.

There is much to like in this policy. It notes increasing data needs, lays down the groundwork for ethical data collection, highlights the importance of data quality and addresses the need for documentation and durable data storage. However, it also remains rooted within the confines of governmental administrative structures and does not directly address the criteria identified by Mr. Spiegelhalter. In the Indian context, each of these presents great challenge.

Competence

Sample surveys, the bedrock of Indian statistical systems, must make explicit choices about who to ask various questions as well as

what to ask and how to ask. In a statistical system developed by renowned statisticians and econometricians, it is not surprising that much attention has been directed towards identifying the universe of respondents and sample selection. However, this is only a small part of the challenge. Given the increasing need for statistics in diverse areas, it is important that scholars from many different disciplines be involved.

The National Sample Survey (NSS) collects data on occupations and industries of workers. In 2009, it suddenly switched from older codes designed in 1968 to new series of codes developed in 2004. This change makes it difficult to differentiate between farmers and farm managers and shopkeepers and sales managers via occupational codes alone. This leaves out such a large portion of the Indian workforce that it is mind-boggling. Why? We decided to adopt international standards developed for industrial societies where self-employed farmers and shopkeepers have been swallowed up by large corporations. I suspect that if a sociologist interested in occupations was involved in overseeing this change, it might not have passed the scrutiny.

Reliability

How surveys are designed and questions are developed has evolved into a science that transcends the skill set usually employed by our statistical systems. The Reserve Bank of India has adopted an inflation-targeting approach that relies on data on inflation expectations of individuals. In a country where ASER (Annual Status of Education Report) surveys repeatedly document extremely low mathematical skills, how reliable are the data when individuals are asked to compare their expectations of inflation rates over the coming year with that in the future? We have little understanding of reliability and

validity of these data, and yet they form the bedrock of our policy. Experiments designed by cognitive anthropologists, educational assessment experts and survey design specialists are needed to arrive at the correct questions. And even then, we will need some way of estimating uncertainty surrounding these results.

Honesty

The draft policy as well as many other reports have paid great attention to the fact that data collection is increasingly being done by contractual employees and for-profit organisations. Supervising them and ensuring their honesty remains challenging. While improved technology for monitoring fieldwork such as random segment audio recording of interviews and real-time checks for detecting frauds and errors may help increase honesty, there is no substitute for empathy and experience. Whenever I talk about interviewer errors and fraud, I recall doing a health-related interview in a mosquito-infested locality. I was bravely suffering through mosquito bites until my respondent told me her husband was recovering from malaria and I simply wanted to flee her home. We expect interviewers to work under challenging circumstances and often send them out to collect data with little training and support. A nimble survey management structure that understands the difficulties of on-the-ground data collectors and responds appropriately to find

ways of ensuring quality and honesty must form the cornerstone of good data collection.

The draft policy on official statistics engages with these challenges only tangentially. Instead, it chooses to follow the report of the C. Rangarajan-led National Statistical Commission (NSC) submitted in 2001 and focusses largely on coordination within different ministries at the Centre and between State governments and the Centre. A tendency to centralise authority and decision-making within well-defined structures such as the NSC forms the core of the policy statement. It also recommends that a registered society under the oversight of MOSPI be set up with ₹2,000 crore endowment that will be tasked with all government data collection and statistical analyses.

Instead of creating a statistical data ecosystem that harnesses the energy of diverse institutions and disciplines in which innovative thinking on data collection and analysis could be undertaken, this tendency towards centralisation may well isolate official statistical systems. This is quite a departure from India's illustrious history. Mahalanobis was instrumental in setting up both the Indian Statistical Institute (ISI) and what was to become the National Sample Survey Organisation. Most of the early innovations implemented in the NSS emerged from work by academics at the ISI. However, as former member of the NSS Governing Council, T.J. Rao, notes, the collaboration between academics and the NSS has weakened substantially in recent years. The proposed move would lead to even further alienation of official statistical systems from the academic and research infrastructure of the nation.

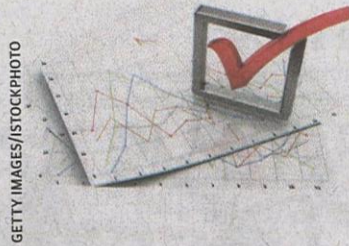
Harness diverse energies

If we are to revitalise India's statistical infrastructure, it is vitally important to harness diverse ener-

gies from academic and research institutions such as the ISI, the Indian Agricultural Statistics Research Institute, National Council of Applied Economic Research, the Tata Institute of Social Sciences, the International Institute for Population Sciences, the Delhi School of Economics, the Madras Institute of Development Studies and the National Institute of Rural Development and Panchayati Raj. Smaller, technology-savvy private sector organisations may also make important contributions in technology-driven data collection. Around the world, in diverse countries such as China, South Africa, Brazil, the U.K. and the U.S., statistical ecosystems consist of universities, research institutions and government agencies working synergistically. The proposed policy on official statistics is timely and thoughtful but it is also isolationist. Creative thinking about building synergies with diverse communities such as academic and research institutions would strengthen it and reduce the burden on the NSC, leaving it free to devote greater attention to developing quality control parameters and to play an oversight and coordination role.

The phrase 'figures don't lie, but liars figure' seems to sum up the motif of a post-statistics society. A report in *The Guardian* in 2017 noted declining trust in official statistics around the world and argued that it damages democracy by jeopardising public knowledge and public argument. The draft National Policy on Official Statistics offers a great start for fostering trust in statistics but enhancing its inclusiveness will go a long way towards encouraging competence, reliability and honesty in public statistics.

Sonalde Desai is Professor of Sociology at the University of Maryland and Senior Fellow and Centre Director, NCAER-National Data Innovation Centre. The views expressed are personal



GETTY IMAGES/ISTOCKPHOTO