Acting locally: A perspective for statistics training in South Asia

Anil P Gore

CYTEL Statistical Software and Services Pvt. Ltd

goreanil@gmail.com

Summary: Statistics is turning out to be a very widely used science and technology in the twenty first century. Hence a discussion about capacity building in this field is very relevant. The problems in this area differ from one geographical area to another. This note seeks to highlight the problems in South Asia in general and India in particular. This region has a reasonable infrastructure for statistics education. The limitation is lack of interdisciplinary interaction. It can be traced to an insular teaching community. Mentoring statistics teachers to enter applied statistics in general and consulting in particular is proposed as the way out of the present difficulties.

Introduction: According to an oft quoted statement attributed to Hal Varian, Chief Economist in Google, 'With data in huge supply and statisticians in short supply, being a statistician has to be 'the really sexy job for the 2010s'.   This of course refers to the first world (Western Europe, North America, Japan etc.). Here opportunities for statisticians are ample and expanding. This is because for various reasons, availability of data is enormous and the only constraint on its use is digesting and interpreting it using statistical tools. Also, the infrastructure for training manpower is well developed. So the near future prediction is of popularity of the subject among students.

In the 'wild world' of Africa, there is a growing need for statistics capacity building. My own experience is limited to work on the MIKE TAG (Monitoring Illegal Killing of Elephants- Technical Advisory Group) membership for a few years. It turned out that data of some complexity do get collected. But its analysis remains primitive; very little beyond simple tabulation and descriptive statistics. This is due to lack of capacity at various levels from field level to national level. Even confidence intervals are regarded as confusing, let alone more sophisticated indices like PIKE (probability that a killing is illegal). Hence there is a major need to impart training and increase statistics capacity.

The situation in South Asia is somewhere in between. My own experience is mainly from India. I am going make the assumption that conditions in Bangladesh, Pakistan, Nepal, Bhutan and Sri Lanka are similar. I have a feeling that the situation in Indonesia, Thailand, Myanmar, Cambodia, Laos, Vietnam, Philippines etc may also be not too different.

Some generalities about India: There is a considerable and rising demand for statisticians. A large number of fresh graduates are available for employment. However, there is a shortage of experienced workers. Situation with creative workers is worse. There is a draught of such personnel. Do statisticians cater to the urban needs or do they serve rural areas also? Generally, the employment is in the urban area only. But the use of statistics may be related to urban or rural ventures. So, statisticians are needed in the IT

sector, especially in the analytics field. But they are needed in agriculture related organizations also.

Subject of statistics has a history of over 80 years in India. Professor P C Mahalanobis founded the Indian Statistical Institute in 1933 and also started the journal Sankhya in the same year. Many renowned scientists from India made signal contributions to the theory and practice of statistics in these years; P V Sukhatme, R C Bose, S N Roy, C R Rao, V P Godambe, to name a few. Indian Agricultural Statistics Research Institute is yet another major training and innovation center developed during the second half of the twentieth century. Many statisticians trained in this institute went on to work in various third world countries on problems of agricultural development. National Sample Survey Organization (NSSO) and Central Statistical Organization (CSO) are two large departments of the central government dedicated to gathering and digesting data on the national economy.

At present, there is a large educational infrastructure available for training undergraduate and graduate students in statistics. As a ball park/ order of magnitude estimate, there may be about a thousand colleges offering B.Sc. in statistics with an output of say 10 per year. In a similar vein, there may be about 100 centers offering M.Sc. in statistics with about 15 students passing out each year. The output of Ph.D. degree holders however may be rather small.

Content and Quality: So, if the output of statistics degree holders is reasonably large, does it fulfill the societal needs or are there any issues regarding content and quality of the statistics training available in India? If we talk of quality of education in general, clearly there is a group of institutions such as Indian Institutes of Technology (IIT), Indian Institutes of Management (IIM). Indian Statistical Institute is such a place in the field of statistics. Alumni of such centers excel in their respective fields. The ambience in these institutes is suitable for innovation and frontline research. But these are all extreme cases. What about quality in a typical educational facility? Many studies are available regarding education in India. Here is one study from Yale University: http://yaleglobal.yale.edu/content/india-graduates-millions-too-few-are-fit-hire

Here are some of the comments in the report: So few of the high school and college graduates can communicate effectively in English. Schools are hampered by overbearing bureaucracy and a focus on rote learning rather than critical thinking and comprehension. Curriculum in most places is outdated and disconnected from the real world.75% of technical graduates and more than 85% of general graduates are unemployable. Average graduate's "ability to comprehend and converse is very low. Clearly, there is much room for improvement. Now let us consider issues specifically related to statistics.

The issues are partly cultural and partly structural. Typical teaching style emphasizes theory; theorems and proofs. On the other hand there is lack of independent/innovative work. There is hardly any work on real life problems. Of course some aspects of the situation are beyond the control of teachers. For example students prefer to study engineering or business management or computer science rather than statistics. Traditionally science education is split into two streams, mathematical science and

biological sciences. Students in one stream cannot take up subjects in the other stream. So, statistics (mathematical sciences) students cannot learn biology, the field with very high potential for use of statistics. Examination system in colleges and universities is rigid and discussion on innovation in syllabus etc can get bogged down with the question- how will we examine students on this? In general in a typical educational institute, there is hardly any interdisciplinary interaction. So statistics students do not learn what the needs of biologists are.

A key lacuna: Perhaps the key problem with content and quality is that there are few teachers with experience of practice of statistics. This leads to many distortions in the system. In particular, there is lack of appreciation about which result/method is more important and which is not. Is robustness more important or second order efficiency? Is an optimum property crucial or is model dependence a serious weakness? These judgments are not easy if one is not practicing the art/science of statistics. People seriously need to learn John Tukey's dictum that an approximate answer to a relevant problem is better than an accurate answer to an irrelevant problem.

Rectification: The only way this can be corrected is through active interaction of teachers with scientists, doctors, agriculturists, bankers, traders etc. I like to use the phrase 'FFFF transforms' to describe this. There is no mistake and no mystery. I believe that 'Following the Footsteps of our Founding Fathers' transforms us as statisticians. Who are they? In India the names would be P C Mahalanobis and P V Sukhatme. For statistics as a discipline, there can be many candidates. My favorite names are Karl Pearson and R A Fisher. They all have one thing in common; a passion to use statistical tools to solve scientific and social problems. Take home lesson from their professional lives is that statistics comes alive when in practice. So how do we go about persuading the teachers to imbibe this value? There is no point in suggesting some revolutionary changes in the system. We need a 'jugaad'; in other words, inexpensive innovation.

Jugaad: Let us recall that in our assessment the weak link is teachers. They lack contacts with industry. They also lack confidence. So we need to arrange mentors. A three way dovetailing with a company, a mentor and a teacher as the triad may do the trick. The company proposes problem of interest. Mentor proposes analysis plan/action plan. Teacher and students do the actual analysis. Finally Mentor helps in interpretation and report writing.

Ideally, professionals in the industry would make the best mentors. But they may not be able to devote enough time. That is where friends and well wishers all over can play a useful role. Since we are discussing the possible role that International Statistical Institute can play, let us visualize an action sequence with ISI as the initiator. So, as a first step ISI approaches the government/industry interested in promoting statistics education and use. Then a specific task identified and a task force is formed. So there are three parties. Industry representative, mentor in North America or any other part and an Indian college with statistics teaching program.

Clearly in the process visualized, mentor contributes expertise and credibility, Industry gives problem, data and some resources while teachers provide manpower. If the task is completed well, Industry gets problem solved, teachers get training and experience and mentors get gratitude.

Of course, we have to be clear about one thing. Each society must solve its own problems. Others can help. But the main thrust must come from within the society. However, ISI does want to get actively involved in such capacity building activity. They already have a mentoring program for students (Ph D) in developing countries. My contention is that a similar program is needed for teachers; not about how to teach but about how to consult. ISI can be particularly useful to bring various parties to the table.

Conclusion: The analysis presented suggests that for India and South Asia, improving training in applied statistics is essential. Difficulty is lack of application experience in teachers. Hence a special kind of mentoring can be helpful. Lastly, the emphasis is on teacher training rather than student training. This is because while one student trained equals one qualified worker, one teacher trained equals a stream of qualified workers .