The Statistical Science: Local and Global Directions

ERNIEL B. BARRIOS, ZITA V. ALBACEA, AND JACQUELINE M. GUARTE

STATISTICS EDUCATION IN THE PHILIPPINES

In 1952, the Philippine Statistical Association pursued the idea of putting up an international training center in statistics in Manila. Soon after, in 1953, a bilateral agreement between the United Nations and the Government of the Philippines was signed to create the Statistical Training Centre. The Centre, as a training institution, had the initial mandate of training professional statisticians for the country and for the region in support of reconstruction programs after World War II. After a series of short courses to capacitate the statistical workers in the Philippines and in the region, the Master of Arts in Statistics program was offered in 1954. Ten years later, the Statistical Center was absorbed into the University of the Philippines, initially to continue offering the Master of Arts in Statistics program. The program went through a series of revisions until the current offering of two master level programs, the Master of Statistics and the Master of Science in Statistics. In 1967, the undergraduate program was offered initially as Bachelor of Statistics, then later up to the present, as Bachelor of Science in Statistics. The Doctor of Philosophy in Statistics program was introduced in 1978, but the first graduate (Dr. Sonia P. Formacion) completed only in 1983. The PhD program was enriched by the visit of well-known statisticians such as C.R. Rao, G. Cox, L. Kish, K.C.S. Pillai, L. Chen, J. Rustagi, and R. Mariano.

Since 1953, other UP constituent universities started offering degree programs in statistics at the undergraduate and graduate levels. In other universities, some programs were developed initially to support the research requirements of other disciplines, but later on, as a separate discipline offering formal training in statistics. In all these programs, mathematical statistics is emphasized while some key areas in survey sampling, experimental designs, time series analysis, statistical modeling, and computing are also pursued as strengths in some universities. At present, some of the more active schools offering degree programs in statistics (BS Statistics) are UP Diliman, UP Los Baños, UP Visayas, Visayas State University, Mindanao State University(MSU)-Iligan Institute of Technology, Eastern Visayas State University, and De La Salle University. There are schools which offer BS Applied Statistics or Major / Minor in Statistics, including Polytechnic University of the Philippines, Rizal Technological University, Benguet State University, and Samar State University. At the graduate level, MS

Statistics, Master of Statistics, and Master of Applied Statistics are offered in the following schools: UP Diliman, UP Los Baños, MSU-Iligan Institute of Technology, Polytechnic University of the Philippines, De La Salle University, and Benguet State University. However, PhD Statistics is offered only at UP Diliman and UP Los Baños.

Statistics education is also accessible to practitioners in government and private institutions—either to acquire new skills or for skills updating—through the Statistical Research and Training Center (SRTC), Philippine Statistical Association (PSA), UP School of Statistics (UPSS), and other training institutions. SRTC caters to government workers, PSA provides annual training to its institutional members, and UPSS offers short courses on basic statistics and on recent advances in the discipline.

CURRICULUM

Statistical science is defined as the science and art of uncertainty. Statistics as a discipline roots in a hybrid of methodologies from mathematics, the life sciences, and the social sciences. As such, the philosophical basis of statistics incorporates various principles from the different disciplines. The discipline evolves quickly, and the theory and methods are documented not only in statistics literature but also in mathematics, engineering, computing, biology, economics, and the social sciences. Regardless of the origin of the different BS Statistics programs in the country, the curriculum for the discipline should necessarily be dynamic since there has been dramatic development in the discipline these past few years.

Today, statistics is offered at all levels of higher education: undergraduate, master, and doctoral levels. The BS program usually aims to introduce students to statistical methods along with its theory and applications. At the master's level, students are trained to innovate towards novel use of statistical methods as well as contribute to the development of new methods. At the PhD level, students are trained to be able to contribute new knowledge in the statistical science.

To keep pace with rapid developments in the field, the current curricula of various BS Statistics programs are updated regularly. A typical curriculum should be revised every five years and should be updated as the need arises (in-between the five-year periods).

The curriculum is a balance of various skills (statistics, mathematics, and computing) needed by a functioning statistician in the industry, government, or in the preparation of more advanced work in basic research. The programs include a thorough study of the different schools of thought, e.g., classical parametric, nonparametric, Bayesian, among others. It also includes tools that will allow collection and analysis of data through sample surveys, experimentation, and mining from huge databases. To complement the different sub-areas of statistical science, a good set of mathematics courses—up to calculus and linear algebra—are included along with a working knowledge of programming and statistical software, exposure to other possible areas of applications, and development of communication skills to help them interpret information and share this with stakeholders in other disciplines.

In most programs, students are trained to be capable of conducting independent research. Similarly, a practicum program is also included to ensure that students are exposed to statistical consulting, i.e., the practice of statistics.

The graduate curricula are also updated regularly to incorporate recent developments in the philosophy of the discipline, e.g., the emergence of computational statistics. With the availability of huge databases as a consequence of large storage capabilities among different sectors, there is now a sudden change of perspective. From small, homogeneous data sets, statisticians now mine for valuable information from huge, heterogeneous databases. As this information usually does not conform to traditional assumptions of the probabilistic structure of statistics, characterization is done through intensive computations and simulations, to discover empirical structures rather than impose

analytical models. Optimality in analytical approaches (or in asymptotic methods) has now been replaced by robustness. These and other areas of concentration (social statistics, computational statistics, data mining and business intelligence, risk management, mathematical statistics, and industrial statistics) are included in graduate programs of the discipline. The core graduate programs in mathematical statistics, multivariate analysis, linear models, survey sampling, Bayesian methods, and nonparametric methods are also updated with the new developments in the field.

Outside the courses in the statistical sciences, students are also exposed to different application areas like microarray (genomics), econometrics, risk management, clinical trials, biostatistics, genetics, plant and animal breeding, rural sociology, environmental science, actuarial science, among others.

The statistics curriculum in coming years is expected to continue providing solutions, methods, and theories that use computing technology. Also, other subspecific topics in biostatistics like high dimensional data will continue to trigger new theoretical developments in the discipline.

Statistical thinking has to be developed among young statisticians through curricular provisions. Exposure to "real" data in learning different statistical techniques (via statistical consulting practice) is a way to develop one's statistical thinking. Students should also be trained in the management and analysis of large databases as availability of such is becoming more of a norm than as a "privilege" access in the past.

FACULTY

With the increasing number of new sub-specializations, it is important that a statistics unit should have a reasonable number of faculty members who can keep track of the developments in different areas. It is also a reality that statisticians are being offered more attractive compensation packages in various industries, specifically, in the financial sector. Thus, while the faculty should be the backbone of any statistics program, it is also the most vulnerable element in successfully implementing a statistics program at the higher education level.

To offer a program in statistics at the master's or PhD levels, there should be sufficient number of faculty members with PhD in statistics or in related fields. While those with BS (Statistics) are allowed to teach at the undergraduate level, they are required to continue working for their MS (and subsequently, PhD) degree.

The UP Diliman School of Statistics and UP Los Baños Institute of Statistics have the "largest" faculty complement in the country. The School of Statistics has a total of 28 faculty members and some five to eight lecturers. Eight of the faculty members have PhDs while fifteen have MS degrees; the rest are working on their MS degrees. At the Institute of Statistics, of its 24 faculty members, three have PhDs, six have MS degrees, and the rest are working on their MS. In some "smaller" units like Visayas State University, there are six statistics faculty, three of whom have PhDs. These are similarly observed in other schools like MSU-Iligan Institute of Technology, UP in the Visayas, and De La Salle University.

While the current profile of the faculty is already almost ideal in some schools, this will change soon as they get attracted to more lucrative jobs outside the university. To fill up the vacuum left by these faculty, it is important to continue training new faculty to replace them. Also, the faculty should pursue more advanced studies not only in statistics, but in allied disciplines as well, so that statisticians can contribute significantly in solving certain real-life problems more holistically.

Other than teaching, faculty members are expected to engage in statistical consultancy. This is important not only in augmenting the financial resources of faculty members who might otherwise leave for other industries, but also as a stimulus in the conduct of research that will have instant applications outside the academe. Furthermore, statistical consultancy is proven to have enriched teaching of statistics in the classrooms.

It is important for school administration to continue searching for creative strategies of providing/generating incentives for the faculty, including, but not limited to, allowing the faculty to do statistical consulting (limited practice of profession). Statistical consulting will continue to provide motivation for various research problems that the faculty will be working on in coming years.

There should be increasing investments on faculty development not only to abate the threat of a mass exodus of faculty members, but also to have a faculty with enough training on certain emerging subspecializations.

Universities should continue to provide the right environment so that faculty members can do research and extension activities. These activities do not only support instruction but also provide them with appropriate exposure to be developed in the statistics profession.

The faculty should also be encouraged and supported to participate in international conferences to establish linkages with their peers and be updated on new developments in the field.

RESEARCH

The current thrust in graduate education is the production of new knowledge through researches that are subsequently published. The contribution of Filipino statisticians is not only contained in the local journal, *The Philippine Statistician*, but is now becoming prominent in international journals as well.

Research in statistics is straightforward to implement. A high-speed computer, an efficient algorithm, some data sets, and access to the literature would usually suffice to produce a high quality statistical paper.

A statistical paper can either include theoretical contributions or information useful to other disciplines. This is a documentation of new developments or applications of statistical science. Young researchers are sent to international conferences not only to present/disseminate results of their work, but also to expand their exposure to the work of other statisticians. The habit of presenting research results in various venues (seminars, local conferences, international conferences) enriches the research environment.

The cash incentives for publications in ISI-indexed journals in the UP System prove to have stimulated research productivity, with more papers now published in ISI-indexed journals and other media. Other schools like the Visayas State University, De La Salle University, Mindanao State University-Iligan Institute of Technology, among others, are also implementing similar incentive schemes. Study groups/research laboratories are also providing stimulus for an environment conducive for research.

The contributions to the theory of statistics are focused on the areas of computational statistics, nonparametric approaches, spatial-temporal model, analysis of high dimensional data, to name a few. There are also substantial contributions in application areas like risk management, consumer behavior modeling, predictive modeling, data mining and business intelligence, econometrics, and biostatistics.

There is a necessary complementation between research and statistical practice. Problems encountered in statistical practice provide the stimulus needed for research to become relevant. Researches that are motivated by problems in the practice of statistics easily find their way into real-life applications in a short period.

There is a growing pattern of infusing learning from statistical practice into research. The results are innovative ideas, integrated/hybrid of different methods, that are often more "optimal." The resulting papers would then appear not only in statistics journal, but also in journals of allied areas like computing, economics/econometrics, biostatistics, among others.

Academic institutions are seen to further strengthen the capacity in computational statistics and other areas, as well as more regular meetings (and funding) for the different study groups. This should also be accompanied by upgrading to high-speed computing laboratories. There will be greater visibility of academic researchers in the industry (locally and in the region) forging further

collaboration between statisticians and other experts.

The complementation between research and statistical practice will continue to prevail where problems encountered in the practice of statistics will continue to provide the stimulus for research.

RESEARCH OUTSIDE THE UNIVERSITY

SRTC is mandated to conduct research on topics within the research agenda of the Philippine Statistical System (PSS). The research agenda of the PSS is identified in the Philippine Statistical Development Plan developed in the medium term (about six years) by an inter-agency committee coordinated by the National Statistical Coordination Board (NSCB). The themes of these researches usually focus on the enhancement of official statistics with beneficiary agencies like NSCB, National Statistics Office (NSO), Bureau of Agricultural Statistics (BAS), Bureau of Labor and Employment Statistics (BLES), among others. In addition, other agencies like the Department of Economic Statistics of the Bangko Sentral ng Pilipinas, NSO, NSCB, BAS, and BLES, also conduct in-house researches or commission such studies that will resolve some internal problems/issues related to the generation and utilization of official statistics.

In industry, research is very active as well, but often, these are supervised by researchers from the academe who are engaged in the practice of the profession outside the university. This is similarly practiced among government agencies. This collaboration fertilizes the kind of research conducted in the academe, as research topics are normally prompted by some real-life problems.

LINKAGES AND EXTENSION

Linkages and extension should be pursued within industry, government, and other academic institutions. Universities conduct a strategic, participatory planning workshop in the formulation of extension programs/agenda. Harmonization of the institutional agenda with the national and regional research and development priorities is also done.

Different schools are currently collaborating with some key industry players in the country (banks and financial institutions, consumer research, manufacturing, product design, etc.) and with universities in the Asia-Pacific region, Europe, and USA. Likewise, collaborative extension work with government agencies is being conducted to improve official statistics. Training, as well as sharing of expertise with the agencies mandated to collect data for the country, can also increase statistical literacy.

There is an increasing collaboration with other universities outside the country and with practitioners in different industries. Collaboration with government agencies will also be beneficial to both parties concerned. Those in the academe get access to nationwide surveys and learn from the people in the "field" in terms of actual data collection. For its counterpart, those in the "field" learn new developments from people in the academe who usually collaborate with them on these activities. Hence, both parties benefit and are able to attain a common goal.

There are emerging statistical methods that require new/additional training among faculty. Initiatives are needed towards collaborative work with other academic institutions offering degree programs in statistics in the region to further strengthen and promote statistics education.

EMPLOYABILITY OF GRADUATES

While the demand for professionally trained statisticians has grown rapdily, there has only been a gradual increase in the number of statistics graduates. Hence, the number of graduates in various statistics programs in the country is not enough to meet the increasing demand.

The UP Diliman School of Statistics produces the majority of graduates at the undergraduate and graduate levels. Table 1 shows that in academic year 2009-2010, 94 graduated at the BS level, 15 at the master's level, and one at the PhD level. The gradually declining enrolment at the BS level is

related to the declining enrolment in the entire UP System, especially in UP Diliman. Table 2, on the other hand, presents the enrolees and graduates of the UP Los Baños Institute of Statistics.

As testament to high demand, graduates of the different programs are usually hired even before graduation. The companies usually request for the list of graduates months before graduation. Statistics graduates are hired in various industries like those generating data, accumulating large amount of data, or using these data in making business decisions. Specifically, graduates are absorbed in government (official statistics generation, planning, forecasting), financial sector, pharmaceuticals, market research, manufacturing (quality control), and recently, in global support service centers (globally processed services, e.g., programming, modeling, interactive decisions, pattern recognition, and general information mining activities from huge databases).

As the Philippines is now rated the number one country in outsourced business process provision, more global companies will continue to relocate their global support services here. As such, more programmers, modelers, data managers, data miners, etc., with training in statistics will be needed. Aside from global support services, more companies will recognize the role of statisticians in their business operations. There is also a growing demand for statisticians in government and the academe.

Graduation rates for the undergraduate and graduate levels will increase to cope with the continuously increasing demand for professionally-trained statisticians. Priority should be given to increase the number of professionally-trained statisticians that are hired in the academe to induce the "domino" effect, i.e., more teachers of statistics will mean more youth will be trained to handle statistical jobs.

While banks, financial institutions, and knowledge process outsourcing firms will likely absorb the most number of statistics graduates, data mining and artificial intelligence will have an increasing demand for statisticians as well. More statisticians will also get involved in information and communication technologies where their primary role will be on pattern recognition and computing. More government agencies will need statisticians as the UN Statistics Department is promoting more evidence-based decision-making in government.

Table 1 Enrolment and Graduation Data of the UP Diliman School of Statistics, 2005-2010

Academic Year	Average Enrolment per semester				Number of Graduates			
	BS	MOS	MS	PhD	BS	MOS	MS	PhD
2005-2006	403	90	70	20	84	3	7	0
2006-2007	424	94	71	21	85	7	2	0
2007-2008	379	82	75	20	83	8	7	1
2008-2009	368	67	64	22	80	7	10	0
2009-2010	360	68	52	22	94	3	12	1

Table 2 Enrolment and Graduation Data of the UP Los Baños Institute of Statistics, 2005-2010

Academic Year	Average Enrolment per Semester			Number of Graduates		
	BS	MS	PhD	BS	MS	PhD
2005-2006	204	17	1	21	0	0
2006-2007	205	17	3	31	2	1
2007-2008	209	16	4	43	4	0
2008-2009	209	11	6	40	3	0
2009-2010	204	19	6	34	1	1

There is also a need to exert substantial efforts in the recruitment of students into the program. The improving strength of the faculty in statistics should be matched with increased student enrolment.

DIRECTION IN THE DISCIPLINE

The direction of statistics as a discipline is spelled out in Tsay(2000), Hirsch (2008), Kettenring (2008), Efron (2000), and LeSage et. al. (2009).

Statistical science will confront the following scenarios:

- *High dimensional data* – more variables are measured than the available observation units. This usually occurs in microarray from genomics, where there are more genetic markers that need to be measured than the number of carriers.
- *Data from unknown selection or from nonprobability samples* – this is resulting from accumulation of databases, for instance, from customer usage or engagement history. These databases are mined for valuable insights and patterns, hence, the necessity of statistical methods that will address this scenario.
- *Nonparametric methods* – as lesser information on the data collection / accumulation process becomes available, the goal is now for robustness rather than statistical optimality, hence, more nonparametric methods will be introduced.
- *Computing intensive* – in lieu of classical assumptions, the data are examined for its distributional properties and methods become iterative, thus, computing intensive methods will dominate statistical science. Statistical theory is developed through repetitive computations.
- *Increase in more complicated dependence structure of observations like those in space-time interactions* – instead of the separate treatment of space and time that both exhibit dependencies, they are treated simultaneously.

During the 58th World Congress of the International Statistical Institute in August 2011 in Dublin, Ireland, Professors Cox, Stigler, Huber, and Rubin reflected on the history and future directions of statistical science. The last 100 years have seen massive developments in statistical methods and theory. These have been driven by the challenges of ever-widening fields of application and by the rapid evolution in techniques of measurement, data capture and analysis, and of the balance between these. The current challenges are raised by having very large amounts of data ranging from particle and astrophysics to genetics to many aspects of public affairs which certainly call for new statistical methods. The question is do these dramatically change the basic principles of statistical thinking? Yes, they do. The advent of computing and its impact on ever larger data sets have remarkable implications on the way we look at statistical theory.

What is the implication on statistics education? What can be transmitted by classroom teaching is limited and insights from statistical consulting should be infused into teaching to enrich appreciation of the statistical science. Thus, the necessary marriage of statistical theory and practice is inevitable.

NOTE

This paper was based on the technical report prepared for CHED by the authors as members of the CHED Technical Committee for Statistics.

REFERENCES

Goodman, A., Kamath, C., Kumas V. (2008). Data analysis in the 21st Century. *Statistical Analysis and Data Mining, 1*, 1-3.

Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association*, *95*(452), 1293-1296.

Hirsh, H. (2008). Data mining: Current status and future opportunities. *Statistical Analysis and Data Mining, 1*(1) , 104-107.

Kettenring, J.R. (2008). A perspective on cluster analysis. *Statistical Analysis and Data Mining, 1*(1), 52-53.

LeSage, J. Banerjee, S., Fischer, M., & Congdon, P. (2009). Spatial statistics: Methods, models and computation (Editorial). *Computational Statistics and Data Analysis*, *53*, 2781-2785.

Tsay, R. (2000). Time series forecasting: Brief history and future research. *Journal of the American Statistical Association*, *95*(450), 638-643.