Developing the Curriculum of Undergraduate Courses in Statistics: Indian and American Experiences

C. Radhakrishna Rao
Eberly Professor of Statistics
Pennsylvania State University
University Park, PA 16802, USA
crr1@psu.edu

1. Introduction

Statisticians work at the interface of mathematics, computer science and real world data analysis, and their skills can be applied to a wide range of career fields. There is demand for statisticians in government, business, pharmaceutics, genomics, communications, marketing, banking, finance, internet, courts of law, survey research etc., practically covering all branches of human enterprise. Writing about the importance of statisticians and analyzing the nature of tasks they are called upon to perform, from compiling data to decision making, in various activities of private and public sectors, Marquardt (1987) says: "The bottom line is that the actual content of many jobs throughout society has more to do with statistics than with any other single subject matter field."

Statistics is not a basic subject like mathematics, physics, chemistry or biology. Each of these disciplines has a subject matter of its own and problems of its own which are solved by using the knowledge of the particular discipline. There is nothing like a statistical problem which is solved by using the knowledge of statistics. It is used to solve problems in other disciplines. The subject matter of statistics was motivated by practical problems, and it grew from isolated methods applied to particular problems to the consolidation of different methods under a unified theory based on the concepts of probability.

2. Origin of graduate studies in statistics

When statistics was brought into the academia in the forties of the last century, statistics degrees were widely thought of as *professional degrees* with emphasis on graduate rather than undergraduate study. This was at a time when statistics was acquiring the status of a separate discipline with *the mathematical foundations of theoretical statistics* as developed by R.A. Fisher, J. Neyman and Abraham Wald. This was at a time when computational facilities were limited to the use of a hand driven desk calculator, and exact closed form solutions to problems were sought using simple stochastic models for data. This was also the time when creation of separate departments of statistics were considered as not viable and mathematics departments were deemed as the closest hosts. The creation of the *Institute of Mathematical Statistics* (IMS) with the journal *Annals of Mathematical Statistics* in the US portrayed statistics as a branch of mathematics. Knowledge of advanced mathematics was considered to be a prerequisite for learning theories of estimation, testing of hypothesis and decision making as developed by Fisher, Neyman and Wald. (Fortunately, *Annals of Mathematical Statistics* is renamed as *Annals of Statistics*. When will IMS become IS?). This was also the time when statistical activities were not wide spread and jobs were available to specialists with master's or doctorate degrees in statistics as consultants in agricultural experimental stations or advisors to government departments. Under these circumstances statistics acquired the status of a graduate level course of study.

3. Ubiquity of Statistics

The character of statistics had gradually changed over the last fifty years with the advent of modern computers with enormous computing power, access to unprecedented quality and variety of data due to computerization of science, commerce and government, and the realization that information processing is the key for optimum decision making in any enterprise. The scope of statistics as it is understood, studied and practiced today extends to the whole gamut of natural and social sciences, engineering and technology, management and economic affairs, medicine and law, and arts and literature. In many of these activities, statistics is routinely used requiring a large number of workers for collecting data and summarizing the results for decision making. In some activities such as quality control in industrial production, machine operators are required to know how to use statistical control charts to check on the quality of goods produced. There are, however, situations where experience and knowledge of advanced statistical methodology is required to solve problems. But, once a statistical system is set up in any organization and software is made available to do statistical analysis, we need workers with different levels of knowledge to manage it.

All these new demands on statistics in the information age require rethinking on the education and training of statisticians to work in academic and nonacademic positions.

4. Future education in statistics

4.1 General remarks

A frequently asked question and debated over the last 50 years is: At what stage in a student's educational career should statistics courses be introduced? In my opinion, the study of statistics can begin in the high school, though not as a separate subject like mathematics, physics, chemistry and biology but as an integral part of these subjects. The possibility of introducing statistical concepts in teaching the basic subjects like physics, chemistry and biology in high schools incorporating statistical ideas is indicated in Rao (1969, 1974, 1997a). The study of statistics as a separate discipline could be introduced at the undergraduate level as a compulsory subject to all students. Advanced courses in statistics to turn out practicing statisticians in special areas of statistics could be introduced at the graduate level.

4.2 Statistics in undergraduate courses

In designing undergraduate courses in statistics, we have to keep in view the needs of students who after getting a bachelor's degree want to

1. seek employment in an organization where statistical skills are needed,
2. pursue graduate courses in statistics and prepare themselves for professional careers as statistical consultants, and
3. pursue graduate courses in other subjects where knowledge of statistics is useful.

I shall make a few suggestions based on my experience in designing courses in statistics for the B. Stat (Bachelor in Statistics) degree course at the Indian Statistical Institute (see Rao, 1969), and the experience of the Statistics Department at the Pennsylvania State University in developing courses for B.S. degree with statistics as a major subject.

4.2.1 Compulsory courses for all categories 1, 2, 3 of students

Exploratory data analysis, Concepts of probability, Introduction to estimation and testing of hypothesis, Regression, Sample surveys and Design of experiments as methods of data collection, Database management, Simulation and bootstrap, Computer programming, Use of software, Calculus I (introductory course on Calculus).

4.2.2 Elective courses for category 1 of students. (Some suggested topics are listed below. Other topics may be added depending on the demand for statistical personnel with some knowledge in special areas.)

Economics, Business administration, Accounting and finance, Statistical quality control, Financial management, Actuarial statistics and risk theory, etc. Only basic concepts in these subjects need be taught.

4.2.3 Compulsory additional courses for category 2 of students

Linear algebra, Calculus II (advanced methods), Probability, Generalized linear models, ANOVA, Advanced topics in estimation and testing of hypothesis, Applied Multivariate analysis, Time series, Bayesian statistics.

4.2.4 Elective courses for categories 2 and 3 of students

Econometrics, Biostatistics, Time series, Bayesian statistics, Multivariate analysis, Demography, Applications in Psychology.

4.2.5 Elective courses in fields other than statistics for category 3 of students

The students can choose some field in which they want to pursue in their graduate program.

4.2.6 Lab Course

In addition to the above courses, provision should be made for a Lab Course, as in physics, chemistry and biology, where students get actual training in collecting and analyzing real data using computer software, and writing reports. In India, both at the undergraduate and graduate levels, a three-hour period per week is devoted to the Lab Course where under the supervision of an experienced teacher, students do the actual computations on live data given to them, and write reports on the conclusions that can be drawn from data.

4.2.7 Apprenticeship: Attempts should be made to provide opportunities for students to work in organizations involved in applications of statistics during the vacation period.

The compulsory courses for all categories of students are meant to provide the basic knowledge and skills needed in statistical practice for collecting data, doing exploratory analysis for understanding data and applying appropriate statistical methods using available software and interpreting results.

The courses on sample surveys and design of experiments, two important methodologies in statistics for collection of data from which valid inferences can be drawn, offer an excellent scope for introducing statistical concepts of a random sample and the principle of randomization, and estimation and tests of significance. The best way of teaching these subjects is to let the students conduct a survey and design some experiments. The validity of the method of inference from survey data can be demonstrated by drawing samples from a known population.

There are several text books on probability, statistical inference and statistical methods written with an applied flavor, which can be used in teaching undergraduate courses in statistics. The book (1997b) by the author will be useful for a general reading and also the paper by Rao (1996).

Considering the importance of and the need for developing a good undergraduate program in statistics, the American Statistical Association created an *Undergraduate Statistics Education Initiative* (USEI) to consider what might be done to foster the growth of undergraduate education in statistical sciences. The USEI held a Workshop on April 28-29, 2000 to discuss guidelines for undergraduate statistics. Reference may be made to papers by Moore, Bryce et al and Ritter et al summarizing the deliberations of the Workshop, which appeared in *The American Statistician*, 55, 1-18 (2001). Further details on the curriculum guidelines for Bachelor of Arts degree in statistical science discussed at the Workshop can be found in a paper by Tarpey et al (2001).

I believe all efforts should be made to introduce statistics as a separate discipline at all levels of education from high school to graduate programs to meet the demand for an efficient work force in the new information age.

REFERENCES

Marquardt, D.W. (1987). The importance of statisticians, J. Am. Statist. Assoc. 82, 1-7.

Rao, C.R. (1969). A multidisciplinary approach for teaching probability and statistics, Sankhyâ B, 30, 321-340.

Rao, C.R. (1974). Teaching of statistics at the secondary level – an interdisciplinary approach, In Statistics at the School Level, Almquist and Wiksel Int., 121-140.

Rao, C.R. (1996). Uncertainty, statistics and creation of new knowledge, Chance 9, 5-11.

Rao, C.R. (1997a). A cross disciplinary approach to teaching of statistics, Proc. 51st Session of the Int. Stat. Institute, Istanbul.

Rao, C.R. (1997b). Statistics and Truth: Putting Chance to Work, World Scientific, Singapore.

Tarpey, T., Acuna, C., Cobb, G., and De Veaux, R. (2000). Curriculum guidelines for Bachelor of Arts degree in statistical science, J. Statist. Education.

RESUME

L'étendue de statistique comme il est compris, étudié et exercé aujourd'hui s'étend à la gamme entière des sciences naturelles et sociales, l'ingénierie et la technologie, la direction et les affaires économiques, le médecine et la loi, et les arts et la littérature. Dans beaucoup de ces activités, tandis qu'il y a le besoin pour les consultants de niveau supérieur avec la connaissance avancée de statistique, la demande pour les ouvriers milieu habiles dans les techniques de réception de données et analyse logiciel d'ordinateur seront beaucoup plus grands. Dans ce contexte, il y a un besoin pour un bon programme de non diplômé, avec l'accent sur statistiques appliquées ou sur l'expérience dans la vraie analyse de données. Quelques suggestions pour développer un tel programme sont présentées basé sur l'expérience de l'auteur et d'autres dans l'Inde et dans les Etas Unis.