RT-MAE-8305

EDUCATING AND TRAINING UNDERGRADUATE APPLIED STATISTICIANS

by

C.A. Peres, P.A. Morettin[1]
and
S.C. Narula[2]

Palavras Chaves: Applied statistics, education,
(key words) training, undergraduate

Classificação AMS: 00A25
(AMS Classification) 01A80

(1) Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil.

(2) School of Business, Virginia Commonwealth University, Richmond, Virginia, USA.

- Setembro de 1983 -

ABSTRACT

The course - Applied Statistics - has been offered at the Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil, since 1978. The course was designed to educate and train statisticians at bachelor's level for jobs in government statistical offices and industry. Such a course is almost mandatory for developing countries and useful for other countries. Our objective, in this paper, is to share our experiences with this course.

1. INTRODUCTION

Most students when they finish their B.S. (or even M.S.) degree, not only in Statistics but other areas also, do not have clear ideas as to what is expected of them in a job? or, what, if anything, can they do for a potential employer? We believe, it will be very useful if universities can offer students some "experience" that will give them a sense of direction and some feeling and clear ideas about answers to questions like: What is expected of them in a job? How will they apply some of the techniques they learn in classrooms? What type of job can they expect on graduation?, etc. This will give students better understanding of reality, more realistic (job) expectations and a little more confidence in themselves.

Our objective, in this paper, is to share our experiences with a course specifically designed to achieve these goals and to offer students more training in writing technical reports and making oral presentations. This course - Applied Statistics - has been offered at the Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brasil, since 1978. In section 2, we give the background which lead to the development of the course. The course description including grading is given in Section 3. We describe the organization and conduct of the course in Section 4. In Section 5, we state our experiences with the course over the last five years and conclude the paper with a few remarks and suggestions in Section 6.

2. BACKGROUND

The following circumstances combined with our desire to provide students with more realistic picture of the job world and training in writing and oral presentation of technical reports lead to the development of the Applied Statistics course.

In Brasil, as in other South American countries, at present, very limited resources are allocated to the graduate education which severely limits opportunities available for graduate education. Further, after graduate education (MS or Ph.D.) employment opportunities are mainly in teaching profession. Thus, most students seek employment after a B.S. degree. For us, it is imperative that we educate and train our undergraduates for jobs in government statistical offices and industry - the major sources of employment. It is clearly in the spirit of the comment by Minton (1980) about the omission of bachelor's degree holder from the Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Industry (1980). He remarked:

> "... the report discusses training for a professional statistician in industry, defining such a person as one trained at the master's level or higher. My limited observation suggests that this is correct for organizations that have begun to use the *professional, consulting* statistician or groups of them. I believe, however, that there are many more industrial organizations that have not, and that these organizations engage in a considerable amount of sta-

tistical activity at the bachelor's degree level, which is carried on by personnel with degrees in everything but statistics ..."

This comment is especially true for South American countries where almost all statistical activity in government statistical offices and industry is carried on by bachelor's degree holders.

Moreover, in Brasil, unlike in the United States, Statistics is a recognized profession just like engineering, medicine, or law. That is, only individuals who have B.S. in Statistics can work professionally. Furthermore, the undergraduate program has to satisfy certain minimum requirements set by Federal regulations. These requirements include a training period in government offices or industry.

Lastly, researchers from other departments, such as Biology, Psychology, Medicine, Agronomy, etc., of the University come to the Department of Statistics for consultation. With time the demand on consulting services increased. This resulted in the creation of Statistical Laboratory in the Department of Statistics. The director of the laboratory, a faculty member, coordinates the in- and out-flow of the projects. The projects are brought to the laboratory where the faculty from the Institute helps researchers (referred to as clients from here on) analyze their data, etc. These projects are generally a part of client's M.S. or Ph.D. research or at times research for possible publication.

3. COURSE DESCRIPTION

The course - Applied Statistics - has been offered as a two course sequence, Applied Statistics I and Applied Statistics II, in the senior year of the undergraduate Statistics program at the University of São Paulo since 1978. Applied Statistics I and Applied Statistics II each carry six credit hours. Applied Statistics I is offered in the Fall semester followed by Applied Statistics II in the Spring semester. For all practical purposes (except for the exception noted in Part III), Applied Statistics II is essentially the same as Applied Statistics I.

Each course is divided into three parts which run concurrently. Part I consists of consulting experience in which a student interacts with a client, analyzes data, submits a written report to the client and makes an oral presentation to the client and class. Part II involves training in some government statistical office or industry. Part III consists of lectures and seminars in statistical methodology used frequently in Parts I and II but not covered in a regular undergraduate course.

Part I: This part offers students consulting experience. Each student is given a project. A student meets with a client to learn the background of the project and data. Based on the information gathered so far, he/she meets with the client and faculty coordinator (in charge of the project) to determine the best way to analyze the data and present results. After the data have been carefully and thoroughly analyzed, he/she gives

a written report to the client and an oral presentation to the class. At this presentation, the client and the faculty coordinator are also present. When a project is finished, the student is assigned another project. A student may finish as few as four projects and as many as six projects in the two semesters.

Part II: As per the Federal regulations, each student has to take practical training for half a day each day for two semesters in a government statistical office or industry. It is a paid position where a student works on statistical problems.

Part III: This part consists of lectures and seminars in statistical methodology used frequently in Parts I and II but not covered in a regular undergraduate course. Different set of topics are covered in the Fall and the Spring semesters. The major emphasis is on the techniques rather than theory.

A student is evaluated on each of the three parts of the course. The final grade of a student is based on his/her performance in Parts I, II and III.

4. CONDUCT OF THE COURSE

The course, Applied Statistics, is the responsability of one faculty member. Since, Applied Statistics I and II are divided into three parts, for the sake of clear exposition, we will describe the organization and actual conduct of each part separately.

Part I: This part of the course takes place in the Statistical Laboratory. The faculty-in-charge of the course sorts

through the available projects for their appropriateness, usefulness and feasibility and selects the ones that are suitable for student participation. It may be mentioned at this time that each client when he brings a project to the Statistical Laboratory has to fill out a form which includes among other things a complete description of the project.

The class is divided into small groups ranging from two to five students per group. Each group is assigned a project and a faculty coordinator (often different from the faculty-in-charge of the course) within the first two weeks of the semester. During this time, each client presents his project to the class and answers any questions from students and faculty coordinators. Each group meets with its client to learn more about the background of the project and the data. Then the group meets with its faculty coordinator and the client to select an appropriate methodology for data analysis. After the analysis is completed, the group writes a report (under close supervision of the faculty coordinator) which is given to the client. A copy of the original data and a copy of the report (as an Applied Statistics Report) are kept on file in the Statistical Laboratory. The report does not have a fixed format in order to allow students to he innovative and creative. However, the reports usually contain a description of the project, the statistical analysis, a summary of conclusions and a reference list. Each group also gives an oral presentation to the class at which the client and the faculty coordinator are present.

Although most projects, at present, involve data analysis only, a few involve design and data collection.

As soon as one project is completed, the group is assigned another project preferably in a differente field and a different faculty coordinator in an effort to give students breadth of experience. During the two semesters, a group may complete as few as four projects and as many as six projects. Faculty coordinator assigns a grade for each project.

A list of projects completed in 1982 is given in Appendix A.

Part II: Most students start their training period (called Estágio) in the Fall semester of their senior year. However, a student may start it in his/her junior year.

A student may find an acceptable training position himself/herself or through a faculty member. However, an agency - the Centro de Integração Empresa Escola, CIEE - helps students find suitable positions. The list of students who have not been able to secure positions themselves is given to CIEE who then contacts students directly with possible positions.

For two semesters (minimum), a student works in a government statistical office or industry for a half day per day. He/she is paid twice the minimum wages during this period and works on statistical problems generally under the supervision of a qualified statistician or if one is not available, a faculty advisor.

For this part of the course, the performance of a student is evaluated by his/her supervision who comments about the statistical competence and personality of the student. This forms a part of the evaluation for the course.

Part III: This part is devoted to lectures and seminar covering those statistical methodologies often used in Parts

I and II but not developed in other courses in the program. (A list of Statistics courses taken by a student is given in Appendix B.) The emphasis is mainly on techniques and their applications. Students are also asked to read relevant papers in journals like the American Statistician, the Journal of Quality Technology, etc and give seminars.

Different sets of topics are covered in the Fall and the Spring semesters. Topics in multivariate analysis and as detection of outliers, discriminant analysis, principal component analysis, factor analysis, cluster analysis and canonical correlation analysis are covered in the Fall semester. Topics in analysis of categorical data such as chi-square tests for multinomial distributions, likelihood ratio tests for multinomial distributions and analysis of categorical data by linear models and its extensions are covered in the Spring semester. This part of the course is conducted like a regular course and students are assigned a grade.

The final grade for the course is an average of the grades assigned in Parts I, II and III.

5. OUR EXPERIENCES

This course has been a very positive and satisfying experience for everyone - students, faculty and clients involved. That this is so, is evident as a student moves from the first to the last project by the changes in the type of questions asked, manner of interaction with a client, solution procedures selected, the quality of written and oral presentations. There is visible growth in students maturity and self confidence.

More often than not, student receives a job offer from the place where he/she completed his/her training period.

The reports from the clients have been very supportive and encouraging. Clients have frequently sent letter of appreciation and gifts to students who worked on their projects.

The Applied Statistics Reports provide excellent real life examples and case studies for other courses and in other disciplines.

For Part I of the course, we experimented with groups of size two to five. In our experience, we found that we needed too many projects for groups of size two, whereas in groups of size five some students could get away without doing their share of the work. We have found a group of size three to be an optimum compromise. Our class size varies from 20 to 24 students.

We would like each group to do more projects in different areas to give them a breadth of experience, however, at times it is not possible. We do not feel it is a big drawback as they all attend all the presentations. Further, at present, students have opportunity to work mainly with projects developed by researchers at the University. They rarely have a chance to work with problems in other areas. To overcome this situation, we encourage students to bring to class discussion the projects they encounter in their training program.

Finally, the set up of the course, imitially, required an excessive amount of time and effort. However, now it does not require any more time or effort than an equivalent regular course. It does require cooperation of a number of faculty members to act as faculty-coordinators for projects.

6. CONCLUSIONS AND SUGGESTIONS

We have had a very positive and satisfying experience with this course. We have observed students learn to have confidence in applying theoretical concepts to real problems, develop skills in writing technical reports, give oral presentations and communicate with researchers from other areas. It gives them a realistic idea of what is expected of them on a job and how well they can perform it. In brief, it gives them a little more self confidence.

Such a course is very important in the education and training of undergraduate statisticians who seek employment upon their graduation. For developing countries, it is almost mandatory since most of the statistical work in government statistical offices and industry is performed by bachelor's level statisticians. In light of the comment by Minton (1980), other countries may also find it useful.

If you decide to try such a course, it is worthwhile to remember: (i) the initial set up time and effort may seem excessive, (ii) make sure you have enough manageable projects, and (iii) enough faculty-coordinators are available and willing to work with the students. The success of such aventure depends upon ones committment.

REFERENCES

ASA COMMITTEE ON TRAINING OF STATISTICIANS FOR INDUSTRY (1980), "Preparing Statisticians for Careers in Industry", The American Statistician, 34, 65-76.

MINTON, P.D. (1980), "Comment", The American Statistician, 34, 77.

APPENDIX A

Research projects completed by undergraduate students in the Applied Statistics course in 1982.

AREA \ PURPOSE	Research Leading to Doctor Degree	Master Degree	Publication	Total
Bus.Adm.	-	1	-	1
Biosciences	-	4	1	5
Biomed.Sc.	1	2	4	7
Communications	1	-	-	1
Education	1	2	-	3
Phys.Educ.	-	1	-	1
Nursing	-	1	-	1
Pharmac.Sc.	-	1	-	1
Geophysics	1	-	-	1
Nuclear Eng.	-	1	-	1
Medicine	2	1	3	6
Veterinary	1	-	1	2
Dentistry	-	-	3	3
Engineering	-	2	1	3
Psychology	2	1	-	3
Chemistry	-	-	1	1
Total	9	17	14	40

APPENDIX B

List of Statistics courses taken by a student in the B.S. program at the University of São Paulo, Brazil.

Probability (2 semesters)

Statistical Inference (2 semesters)

Exploratory Data Analysis

Introduction to Stochastic Processes

Sampling Techniques

Nonparametric Statistics

Regression Analysis

Design of Experiments (2 semesters)

Multivariate Analysis

Time Series Analysis

Applied Stochastic Processes

Sequential Analysis

Quality Control.

RELATÓRIO TÉCNICO

DO

DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS

7901 - BORGES, W. de S. On the limiting distributios of the failure time of composite material. São Paulo, IME-USP,1979,22p.

7902 - GALVES, A.; LEITE, J.G.; ROUSSIGNOL,M. The invariance principle for the one-dimensional symmetric simple exclusion process. São Paulo, IME-USP, 1979. 9p.

8001 - MENTZ, R.P. et al. Exploratory fitting of autoregressive and moving average models to well-behaved time series data. São Paulo, IME-USP, 1980. 16p.

8002 - MORETTIN, P.A. Walsh spectral analysis. São Paulo, IME-USP , 1980. 27p.

8003 - RODRIGUES, J. Robust estimation and finite population. São Paulo, IME-USP, 1980. 13p.

8004 - BORGES, W. de S. & RODRIGUES, F.W. On the axiomatic theory of multistate coherent structures. São Paulo, IME-USP, 1980, 10p.

8005 - MORETTIN, P.A. A central limit theorem for stationary processes. São Paulo, IME-USP, 1980. 5p.

8101 - DANTAS, C.A.B. & COLUCCI, E. A Simulation program for emergency services-II, São Paulo, IME-USP, 1981, 14p.

8102 - ANDJEL, E.D. Invariant measures for the zero range process. São Paulo, IME-USP, 1981, 55p.

8103 - ANDJEL, E.D. The asymmetric simple exclusion process on Z^d. São Paulo, IME-USP, 1981, 13p.

8104 - MORETTIN, P.A. & TOLOI, C.M.C., Accuracy of forecasting with special reference to the Box-Jenkins and Bayesian Methodologies. São Paulo, IME-USP, 1981, 41p.

8105 - PINO, F.A. & MORETTIN, P.A., Intervention analysis applied to Brazilian coffee and milk times series. São Paulo IME-USP, 1981, 36p.

8106 - BORGES, W.S. & RODRIGUES, J., Testing for new better than used in expectation. São Paulo, IME-USP, 1981, 7p.

8107 - FAHMY, S.; PEREIRA, C.A.B.; PROSCHAN, F., The influence of the sample on the posterior distribution. São Paulo, IME-USP, 1981, 17p.

8108 - PERES, C.A., Asymptotic efficiency of the likelihood ratio conditional test for multinomial distributions. São Paulo IME-USP, 1981, 29p.

8109 - PERES, C.A., Testing the effect of blocking in a randomized complete block design (RCBD). São Paulo, IME-USP, 1981, 14p.

8110 - BASU, D. & PEREIRA, C.A.B., On the Bayesian analysis of categorical data: the problem of nonresponse. São Paulo, IME-USP, 1981, 13p.

8201 - BASU, D. & Pereira, C.A.B., Conditional independence in statistics. São Paulo, IME-USP, 1982, 37p.

8202 - BASU, D. & PEREIRA, C.A.B., A note on Blackwell sufficiency and a Skibinsky characterization of distributions. São Paulo, IME-USP, 1982, 12p.

8203 - PERES, C.A., On the interpretation of the parameters of the quadratic model for cell survival ofter irradiation. São Paulo, IME-USP, 1982., 22p.

8204 - GALVES, A., et al. Rescaling the stirring process. São Paulo IME-USP, 1982, 23p.

8205 - RODRIGUES, J., On the asymptotic theory for the fixed size confidence ellipsoids. São Paulo, IME-USP, 1982, 14p.

8206 - PEREIRA, C.A.B. & RODRIGUES, J., Robust linear prediction in finite populations. São Paulo, IME-USP, 1982, 14p.

8207 - MORETTIN, P.A., Walsh-Fourier transforms. São Paulo, IME-USP 1982, 15p.

8208 - PERES, C.A. & MORETTIN, P.A., Building bridges between the academic and real worlds - some observations from South America. São Paulo, IME-USP, 1982, 16p.

8209 - PEREIRA, C.A.B. & ROGATKO, A., The Hardy-Weinberg equilibrium under a Bayesian perspective. São Paulo, IME-USP, 1982, 16p.

8210 - MORETTIN, P.A., The Levinson algorithm and its applications in time series analysis. São Paulo, IME-USP, 1982, 16p.

8211 - RODRIGUES, J., A Note on Maximized and Conditional Likelihood Functions. São Paulo, IME-USP, 1982, 9p.

8301 - PEREIRA, C.A.B., Stopping rules and conditional inference in 2x2 contingence tables. São Paulo, IME-USP, 1983, 7p.

8302 - BOLFARINE, H., PEREIRA, C.A.B. & RODRIGUES, J., Robust Linear Prediction in Finite Populations: A Bayesian Perspective. São Paulo, IME-USP, 1983, 21p.

8303 - MORETTIN, P.A., et al., Rainfall at Fortaleza, Ceará, Brazil Revisited. São Paulo, IME-USP, 1983, 33p.

8304 - MORETTIN, P.A. & TOLOI, C.M.C., Evaluation of Forecasting Procedures: A Case Study. São Paulo, IME-USP, 1983, 30p.