

AS GOOD AS IT GETS: CHALLENGES IN TEACHING APPLIED STATISTICS

Julio M. Singer

Universidade de São Paulo, Brazil

jmsinger@ime.usp.br

The role of mentoring (i.e., direct supervision) in bridging the gap between statistical methodology taught in the academic environment and its actual application to practical problems is considered. The particularities of the Brazilian case are highlighted and the thirty-year experience acquired by the Statistical Laboratory at the University of São Paulo is described. Topics ranging from interaction with clients, problem specification, statistical modelling, report writing to multimedia presentation of results are covered. The main difficulties and positive aspects of the program are identified. Guidelines to improve it are also discussed.

INTRODUCTION

Statistical analysis of practical problems is usually a difficult task. Frequently, investigators who search statistical consulting for their studies do not have an operational definition of their problems at hand and the statistician in charge must employ all efforts to obtain an acceptable formulation of the questions to be answered. This is further aggravated because often the clients have difficulty in understanding fundamental concepts, like those pertaining to the difference between descriptive and inferential Statistics. Additionally, real world problems have their own peculiarities and frequently require the development of new statistical methodology or at least, some novel way of employing standard techniques for their analysis. Finally, reporting the analytical results into terms that are understood by the client may also call for considerable skills. The application of statistical methods in practice generally demands a significant amount of time and may constitute a challenge even to experienced professionals. The question addressed here is how to teach students to overcome such difficulties.

Standard courses in a good Statistics program should provide the methodological background required to analyze a variety of problems. For the reasons outlined above, this, however, may not be sufficient to deal with real world situations and students must be guided through practical cases before they assume steering positions. In many other fields, this task may be completed with the assistance of senior professionals working at the same institution as the recent graduate. In Statistics, this is rather the exception than the rule, especially in developing countries, like Brazil, where the number of experienced professionals in the market is small and where, usually, the graduate under consideration is the only statistician available. To help bypass this problem and bridge the gap between the application of statistical methodology to artificial examples (as is generally the case in the academic environment) and its actual application to practical problems, one must rely on mentoring (i.e., direct supervision).

In the Department of Statistics at the University of São Paulo, Brazil, mentoring has been carried out for over thirty years by means of two one-semester courses in Applied Statistics designed for last year undergraduates. Surprisingly, there are no similar disciplines offered at the graduate level. In each semester, the students are exposed to a concrete practical problem submitted to the Statistical Laboratory. Each project is conducted by one or two students under the direction of a faculty member. Two other faculty members act as general supervisors and reviewers. Additional activities in the courses range from supervised interaction with clients, problem specification, statistical modelling and report writing to multimedia presentation of results.

In this paper, a critical appraisal of such experience is considered. The following sections are devoted to a brief description of the Applied Statistics program, an outline of its positive and negative aspects, and a discussion and the suggestion of possible actions to improve its effectiveness.

THE APPLIED STATISTICS PROGRAM

The program starts with the selection of projects submitted to the Statistical Laboratory. This is accomplished by the two faculty members in charge of the course, before the semester

begins. Selection criteria include scientific interest, amount of available data and a clear need for statistical analysis, besides an acceptable estimate of the time required for completion. Most of the projects involve the analysis of research problems generated by M.Sc. theses or Ph.D. dissertations from investigators (the clients) in the Biological or Medical Sciences. One or two students and a faculty member (the consultant) are assigned to each project. They must provide the complete statistical analysis of the project assigned to them and write a detailed report by the end of the semester; during this period, they must also orally present the initial descriptive analysis as well as the final results. In both occasions, not only the students' colleagues, but also the client, the project consultant and the course supervisor are in the audience along with the client's advisor, in the case of theses and dissertations. Along with the assigned job, students are exposed to regular classes (4 hours/week) on project management, data management, report writing as well as on some specific methodological topics not adequately covered in other courses, but necessary for the analysis of some problems. They are also trained in interviewing new clients who submit their projects for future analysis; here, too, they are required to write a report. The main objective of this activity is to teach them how to identify the problem and to translate the subject matter questions into statistical ones. More recently, they have also been submitted to classes on scientific writing.

The typical projects involve the analysis of data previously collected under experimental conditions designed by the client, with no explicit help from a qualified statistician. Generally, the research questions are related to the comparison of treatments or to the understanding of relations between response and explanatory variables. The most common statistical methods used for the solution of the selected problems include ANOVA and regression analysis, although, longitudinal data, survival or time series analyses are also frequently called for. The following example illustrates the type of problem usually selected for analysis.

- *The study was designed to evaluate the effects of torrefaction time (0 min = raw beans, 10 min and 20 min), torrefaction temperature (140°C, 160°C and 180°C), type of solvent (water, alcohol and ethyl ether) and solvent concentration (0.47 ppm, 0.93 ppm, 1.86 ppm, 0.24ppm + 25ppm BHT, 0.48ppm + 50ppm BHT, 0.98ppm + 100ppm BHT) on the oxidation capacity of coffee beans.*
- *The response variable was the percent oxidation inhibition (%OI), computed as follows:*
$$\%OI = 100 \times \left(1 - \frac{\text{Variation of A in the test sample between 0 and 120 min}}{\text{Variation of A in the control sample between 0 and 120 min}} \right)$$
where A is the absorbancy.
- *A batch of coffee beans (Arabic variety) was divided into 7 portions and each portion was (randomly?) submitted to one of the following 7 treatments: Torrefaction at 140°C for 10min, at 160°C for 10min, at 180°C for 10min, torrefaction at 140°C for 20min, at 160°C for 20min, at 180°C for 20min or raw beans.*
- *Each of the 7 portions was subdivided into 3 sub-portions.*
- *Each sub-portion was submitted to one of the solvents to obtain an extract.*
- *From each extract, 7 aliquots were obtained.*
- *Each aliquot was diluted to one of the 6 solvent concentrations plus a control (no dilution).*
- *From each aliquot the %OI was obtained in triplicate.*
- *The main questions asked by the client were whether the sample size was sufficient to obtain "reliable" analysis and how should the data be analyzed.*

THE EXPERIENCE SO FAR

There is no doubt about the pre/post intervention improvement in the ability of students to deal with real world problems. As a by-product, the experience also provides a rich source of problems for statistical methodological research. The question is whether the observed gains are sufficient and what could be done to improve the program if the answer is negative. The program's principal achievements and major caveats are discussed next.

It seems clear that at the end of the two semester experience, students begin to distinguish between academic and practical problems and to identify the limitations of the techniques they have learned so far. It is also evident that they become less afraid of exposing themselves to public evaluation. On the other hand, one feels that in many cases, there is no real interest in getting involved to the extent needed to conduct a satisfactory statistical analysis (one must keep in mind that the analysis will be used in theses, dissertations and research papers). Furthermore, one notices almost no improvement in their ability in specifying a statistical problem and even less in their capacity to write an understandable report, i.e., one that not only translates the findings in terms that the investigator may use to answer his/hers questions, but also allows the analysis to be reproduced. I have had the opportunity to read some of the reports in order to provide follow-up to clients after a couple of years and to my surprise, I was not able to understand what had been done. It must be mentioned that reports are reviewed (and in most cases, almost rewritten) by the consultant (when there is interest) and by the supervisor (when there is interest). In some cases, up to ten versions of reports must be reviewed before they are considered acceptable. A few examples of part of reports (my translation) presented by the students follow. The original texts are in the Appendix.

- *These results suggested a model that quantifies the increase in the average score of patients without hepatomegaly when the patients have hepatomegaly.*
- *Analysis of variance is undoubtedly concerned with the variability that the data may present according to the various sources of their classification that may be categorical.*
- *The classification, hypertension state and obesity state are called factors and the classes of each factor, in case of hypertension 2 and obesity 3, are called levels.*
- *Suggestions of the CEA (Centre for Applied Statistics):*
 - *To find out methods in the literature to study the procedure of how to eat beef. For example: If an individual masticates a bone after eating beef, the possibility of having rheumatoid arthritis is larger than a person that does not do this.*
- *We may conclude from the descriptive analysis, that the values of (pulmonary) complacence on the average seem smaller in the individuals that received the anesthetic, the values related to resistance and to alveolar ventilation on the average seem larger in individuals that received the anesthetic. For the study in which different air flows were injected in individuals that received physiologic serum, the values of complacence on the average seem to decrease when a larger flow is injected.*

This is as good as it gets. Unfortunately, the negative points of the program outweigh the positive ones and the main reasons for this are:

- i) clients have a great difficulty in transmitting their problems in an organized way and tend to mingle subject matter and statistical modelling questions;
- ii) consultants and supervisors do not get sufficiently involved to completely understand the problem at hand and guide the students through the appropriate statistical analysis;
- iii) the methodological courses taught at the Department are too technical and provide little motivation for applications;
- iv) students have little interest and dedicate little time to their projects;
- v) clients, students and even consultants and supervisors not seldom have very poor writing abilities including logical concatenation of ideas.

DISCUSSION AND PERSPECTIVES

The presence of the advisor of the graduate student requiring the service is mandatory in all meetings and presentations. Nevertheless, in many cases, this does not happen, generally with the excuse that the advisor is "too busy." The project is cancelled when this occurs at its initial stages, but this may not be done when the project has already taken off. Although the difficulties outlined in item i) are not exclusive of the client, they are certainly worsened when he/she has to handle the problem alone. About 6 two-hour meetings were necessary to obtain a reasonable description of the study presented above. To a certain extent, this has consequences on the loss of

interest by everyone else involved in the project. To overcome this barrier, changes in the emphasis of service courses and more generally, in the dissemination of basic statistical concepts are needed. To emphasize the misunderstanding of fundamental statistical concepts by non-statisticians in general, I reproduce the comments made by a referee on a manuscript submitted to a well known international medical journal. The problem under investigation involved the comparison of an active drug with placebo with respect to the distribution of a response variable.

- *Use of SE is not appropriate. This is a measure of variation in means from different studies. Standard deviation is a measure of deviation in one data set. This was a single data set.*

The topics covered in the principal service course taught by the Department of Statistics at the University of São Paulo consist of an overview of standard Probability results and Descriptive Statistics with the major efforts concentrated on interval estimation and tests of hypotheses. This material is covered in 16 weekly four-hour classes. Students finish the course with no idea about the difference between standard errors and standard deviations, about what it means to say that a variable follows a normal distribution and even worse, about the difference between the statistical description of a dataset and the statistical inference that may be carried out there from.

A shift towards teaching data management, variable coding, study design and basic statistical concepts (emphasizing the difference between sample and population and on the role of Statistics in bridging the two) in service courses would certainly improve the ability of potential clients to pose the subject matter questions in a way that facilitates statistical analysis. The ideas underlying a t-test (what should be taught and required on exams), for example, are far more important and interesting than its actual computation (what is really taught and required on exams).

The questions described in items ii) and iii) essentially derive from the enormous amount of time demanded to prepare good methodological courses and even more to supervise a program like the one described here. It is worthwhile to mention that although supervisors do get credits for the Applied Statistics courses, consultants do not, for the projects they direct. These problems stem from the same cause, namely the practically null incentive assigned to teaching activities in our academic environment. Promotions are exclusively based on research activities, i.e., publications (books not included) on methodological journals only and no rewards are given for excellence in teaching. Faculty positions are filled on a who-has-more-publications basis rather than by looking at specific areas in which the Department lacks experienced professionals. This may also explain the absence of advisors in some phases of the project as discussed above, since they might be limited by similar restrictions. I am obviously not placing research on a secondary plan; in fact, research activities are almost mandatory to generate good courses. A policy of hiring faculty with specific methodological and applied interests coupled with incentives for good teaching performance and publications in multidisciplinary journals along with promotions based on a balance between the two activities would certainly help in enhancing both student methodological background and faculty interest in the Applied Statistics program.

Last year undergraduate students provide cheap labour force in a market where there are few statisticians available and many of them start working part-time, mainly in private institutions. Thus, they do not have enough time to devote to their projects and given the other reasons listed previously, end up losing interest in them. In theory, this would be resolved or at least diminished if they could bring the problems they are faced with in their jobs for analysis by the Statistical Laboratory. Unfortunately, this does not happen for various reasons. Firstly because they lack the ability to appropriately identify the problems and there is usually no senior statistician to help them with such a task; secondly because the companies they work for generally do not want to pay for the services they are supposedly hired to do and finally, because of confidentiality reasons. The development of joint research projects with private companies or public institutions would certainly provide funding to hire the students and thus increase their interest in the projects. This type of activity is still not considered in our environment mainly because of bureaucratic reasons.

Poor writing skills are certainly becoming a worldwide problem. Scientific writing, in particular, requires both concision and precision and such characteristics are neglected by many

faculty members in their courses, where exams cover mainly very technical material. This reflects directly in the Applied Statistics program, where technical details play a secondary although not less important role. Good writing skills demands long training periods and the inclusion of a few lectures on scientific writing does not help much. In fact, faculty members should attend these lectures periodically and should consider asking for written reports on their courses. This could not only help students but also improve the quality of their own work and even facilitate publications.

APPENDIX

- *Esses resultados sugeriram um modelo que quantifica o aumento do escore médio dos pacientes sem hepatomegalia quando os pacientes possuem hepatomegalia.*
- *A análise de variância é sem dúvida concernente à variabilidade que os dados podem apresentar de acordo com as várias fontes de classificação dos mesmos que podem ser categorizados.*
- *A classificação, estado de hipertensão e de obesidade são chamados fatores e as classes em cada fator, no caso hipertensão 2 e obesidade 3, são chamados de níveis.*
- *Sugestões do CEA*
 - *Verificar métodos na literatura para estudar o procedimento de como comer carne. Por exemplo: Se um indivíduo mastiga um osso depois de comer a carne, a possibilidade de se ter artrite reumatóide é maior do que uma pessoa que não faz isto.*
- *Podemos concluir pela análise descritiva, que os valores de complacência em média parecem ser menores nos indivíduos que receberam anestésico; os valores relacionados à resistência e à ventilação alveolar em média parecem maiores nos indivíduos que receberam anestésico. Para o estudo em que foram injetados diferentes fluxos de ar nos indivíduos que receberam soro fisiológico, os valores de complacência em média parecem diminuir quando é injetado um fluxo maior.*