A CASE STUDY OF APPLIED STATISTICS EDUCATION IN AN ENVIRONMENTAL SCIENCE DEGREE

Carmen Capilla[1]

[1]*Polytechnic University of Valencia (SPAIN)*
ccapilla@eio.upv.es

Abstract

The objective of this paper is to present a case study of applied statistics education in the undergraduate environmental science degree at the Polytechnic University of Valencia, Spain (PUV). Education in applied statistical methods is of particular value to students who go on to work in careers in environmental management and decision making. In response to a growing need for professionals capable of identifying, analyzing and providing answers to specific issues related to environmental change and environmental problems, the PUV decided to include an applied statistics course in its specialized environmental science program. The course content was designed in accordance with the guidelines set out by the university's official program, while the pedagogical approach and learning activities were developed to reflect recommendations made by mathematicians and statisticians over the past two decades as part of the reform movement in mathematics education. These recommendations include placing a greater emphasis on analyzing and interpreting data, increasing active participation by students, stimulating learning through real-world problem solving, and improving technology and communicating skills regarding data and chance. Student assessment and course evaluation procedures are also an important issue.

Keywords: Cooperative learning, undergraduate students, statistical education, environmental science degree, interdisciplinary.

1 INTRODUCTION

There is a growing need for professionals capable of identifying, analyzing, and providing scientific answers to issues directly related to environmental change and environmental problems. In order to address this need, many universities have increased the number of places offered on their environmental science programs and specialist courses in recent year. The Polytechnic University of Valencia (PUV) launched a Bachelor of Science (BSc) Degree in Environmental Science in 1997. This degree offers a range of specialist subjects taught by faculty staff from 14 academic departments. The minimum time for completion of the degree is two years (four standard semesters) for students who have successfully completed degrees in Agricultural Engineering, Forest Engineering, Mining Engineering, Industrial Engineering, Chemical Engineering, Civil Engineering, Biology, Chemistry, Marine Science, or Geology. Admission to the program is determined on the basis of students' previous marks. The degree, which is run by the School of Civil Engineering, is interdisciplinary in both design and approach.

The main aim of the PUV's Degree in Environmental Science is to prepare students for careers in environmental management and decision making. A scientific approach to environmental issues requires properly applied statistical methodology to ensure well-conducted data collection, analysis and interpretation ([1], [2]). Appropriate training in applied statistical techniques is therefore a valuable asset for professionals that need to deal with environmental problem-solving and management issues. In order to address the growing need for environment professionals with training in statistics, the UPV decided to incorporate a compulsory statistics course into its Environmental Science degree program. The statistics course was tailored designed for the program, and is taught during the first term. This case study will look at the background behind the course, describe its content and structure, and discuss the assessment methods used and the results obtained.

2 BACKGROUND

The application of statistics in areas such as the economy, industry, and social science is well established. However, over the last twenty, the field of environmental statistics, also known as 'environmetrics' [3], has become one of the most rapidly growing areas of statistical applications [4].

This has inspired an increasing number of activities in the field, which has, in turn, given rise to more publications, journals and conference sessions devoted to environmental statistical applications, and improvements in specialized statistics courses. Piegorsch & Edwards [1] discuss the construction of such courses in the light of their experience with the design of a graduate environmental statistics course. In their opinion, it is easier to select material for an undergraduate course. Although there is no consensus on core material, they recommend that an undergraduate course in basic statistics for students whose interests include environmental problem solving should cover such topics as random sampling, basic summary statistics, basic probability and statistical distributions, confidence intervals, significance testing, correlation, and regression. As they indicate: "The challenge in developing an environmental statistics course is to truly engage the practitioner/non-statistics students. The same challenge is present when teaching statistics in other sciences. There the issue has been faced by syllabus designers and this has allowed for consensus to emerge and groups such as Accreditation Board for Engineering and Technology, the American Academy of Environmental Engineers, the Institute of Industrial Engineers, etc., which list a statistics requirement for curriculum certification. A similar effort is only beginning in environmental statistics, and more needs to be done ".

Furthermore, the reform movement in mathematics education that has taken place over the last decades has influenced statistics education and led to a review of content, pedagogy, and technology at all levels of education [5-7]. Recommended changes in the three areas have gradually been adopted as a means of offering continuing improvement in the teaching of statistics. Above all, experts stress the importance of:

- Highlighting connections between statistics and other sciences
- Understanding and using students' prior conceptions
- Placing emphasis on analyzing and interpreting data
- Encouraging greater active involvement by students
- Using real-world problem scenarios
- Encouraging small-group cooperative learning
- Improving students' technology and communicating skills regarding data and chance.

Both the mathematics and statistics education communities have invested a lot of effort into developing new pedagogical approaches and activities to integrate these recommendations into their courses. A student's previous knowledge is one of the factors that have the greatest influence on learning [8]. All the students that take the environmental science statistics course at the UPV have already studied calculus and algebra during their first-cycle degrees. Their background in statistics, however, ranges from no statistical experience whatsoever to basic skills acquired in an introductory statistics course (descriptive analysis and probability). No assumptions regarding statistical knowledge and experience were therefore made when designing the course for the BSc Degree in Environmental Science. The course was designed to prepare the students "to use statistical thinking and reasoning" [9], and provide them with the necessary skills to interpret and evaluate environmental data analyses. Concepts such as mathematical proof and calculation were given secondary priority. Derivations and hand calculations were replaced by the widespread use of statistical computing methods, and an emphasis was placed on the analysis of real data.

3 THE COURSE SYLLABUS

The official program for the Degree in Environmental Science gives the following keywords for the statistics course: "probability distributions, sampling, hypothesis testing, analysis of variance, regression and correlation". These keywords were an important factor in the design of the course, as were the students' mathematical backgrounds, and the recommendations proposed by [6], [10] and [1]. Table 1 shows the course contents and number of delivery hours.

Table 1. Course contents and number of hours, shown by classroom lectures, in-class activities, and computer laboratory sessions for each module.

Content	**Number of hours**			
	Classroom lectures	In-class activities	Computer room sessions	**TOTAL**
Exploratory data analysis Univariate Bivariate Multivariate	6	1	4	11
Probability distributions Basic probability concepts Discrete distributions Continuous distributions	7	3	4	14
Statistical inference Sampling distributions Confidence Intervals Hypothesis testing	6	3	4	13
Analysis of variance One factor analysis of variance Two-way analysis of variance	6	1	4	11
Regression Simple linear regression Introduction to multivariate Linear regression	5	2	4	11
TOTAL	30	10	20	60

Students attend lectures designed to present basic concepts through the use of illustrative examples, and practical sessions, which consist of in-class activities and computer laboratory classes. Course content covers basic subjects such as univariate descriptive analysis, which looks at different ways of presenting data (e.g. frequency tables, data distributions, graphs, bar charts) Concepts such as population, random samples, data types, and variability are introduced through examples taken from environmental data analyses. Covariance and correlation measures are presented during the analysis of quantitative bivariate variables. Several examples are used to illustrate that correlation and association do not necessarily indicate causation ([8], [9]). Examples are then extended to include multivariate data and introduce the students to the concept of multivariate analysis. Emphasis is placed on the use of graphical data presentation methods to analyze patterns and relationships between variables. Visualization is critical to improving the understanding of environmental data and the results derived from their statistical analysis ([11]). Moreover, as Weldon [12] points out: "While multivariate analysis has formidable mathematical problems associated with it, some very simple graphical strategies are possible to convey in a first course… Use of these plotting methods in an early service course may be one way to alert students to the very common multivariate nature of real-life data".

Statistical distributions are studied in the second module of the course. The introductory lectures focus on helping students to understand the concepts of chance and randomness, which are given greater priority than formal probability [6]. A number of mathematical education researchers have noted that students of all ages have frequent difficulties with these concepts [9]. The course includes a variety of classroom and computer laboratory activities aimed at helping students to overcome misconceptions concerning basic probability. Rather than analyzing the mathematical formalities of probability distributions, students learn about discrete and continuous distributions applied to environmental variables. Particular emphasis is placed on the basic aspects of their application. Part of the work of modeling that is encouraged in the students is to decide when a probability model is more appropriate to the situation than some other mathematical model.

In the introductory part of the inference module, the critical role of random sampling in producing predictable sampling distributions is emphasized through various examples and in-class activities. The concept of sampling distribution is key to understanding the basic notions of statistical inference ([8]). Students also study the construction and interpretation of confidence intervals for sample means and proportions, and in the final part of the module, they discuss the applications of one- and two-sample t-tests as useful procedures for examining evidence against null hypotheses.

The course also includes a module on analysis of variance. After a brief discussion of the methodology, students learn about one-way analysis of variance applied to real-life examples. They then discover how to compare 'within variance' and 'between variance' using F-statistics, and finally analyze a case study with two factors that shows the applications of the two-way analysis of variance.

The final module tackles relationship modeling using simple linear regression methods. Students now reanalyze data sets they came across in the first module using the regression approach. They are then briefly introduced to multiple linear regression models and a real-life problem with several regressors is discussed.

They are also taught the different steps involved in building a model: a clear statement of the problem, data collection and initial exploration, model formulation, tentative model fitting and inference, validation of assumptions, and the use of the model to make predictions and decisions regarding the problem. The important role that data plays in the sequential learning process is stressed, with an alternation of induction and deduction steps in the statistical modeling approach. As [13] states: "Discovery of new knowledge requires the use of the scientific paradigm in which the model is continually changing".

This final module is also designed to help students to understand that a model is a useful approximation to reality but that it should never be considered the "final word" ([5]). Students are taught throughout the course that it is very important to bear in mind that some statistical problems can be viewed from different perspectives and that there is not always one right answer ([9]). The course is also designed to encourage students to use a variety of statistical techniques when attempting to resolve problems, or when conducting in-class activities or assignments, an approach which helps them to link the different concepts taught in the different subjects on the course.

4 COURSE ORGANIZATION

Approximately 60 students take the statistics course every year. Theoretical lectures are attended by the entire group while the hands-on computer sessions are taken by smaller groups (four subgroups in total). Classes are taught by a member of the PUV's Department of Applied Statistics, Operational Research, and Quality. Her background and research interests are mainly in statistics applied to engineering and environmental science. Teaching loads vary from three to eleven hours per week. In addition, six regular office hours are kept weekly to attend to students' questions on statistical concepts, in-class activities, or assignments. Students may also ask their instructors questions electronically or post a message in the discussion forum on the course web page. The web page features a collection of students' postings, as well as on-line access to a range of information such as course content, schedules, assessment methods, recommended reading material, and computerized classroom material.

The course consists of 60 hours spread over the 14-week autumn term. Lectures take place on different days of the week and are divided into three types of sessions. The first of these sessions is a weekly two-hour group lecture. In these sessions, the lecturer explains basic concepts related to the subject being dealt with, relates them to the students' existing knowledge, and stimulates the desire to learn new concepts. Each lecture analyzes a real-world problem taken from the field of environmental science, which encourages the students to actively participate in the discussion. In the last part of these sessions, students are arranged in teams of two or three to further analyze the concepts that have been presented. Teams are used to promote cooperative learning in the three types of lecture sessions during the term to help students to realize the benefits of working with others as this will form an important part of their work outside the university.

The second type of session is a one-hour session that is taught over a period of 12 weeks. It involves active participation in group activities and environmental case studies that are introduced by the lecturer using statistical software. The lecture room is equipped with a personal computer connected to the PUV network that has access to the statistical software package used on the course and a video projection system. These are valuable tools since they can be used to display graphs and the results of calculations and thus help students to intuitively understand the theoretical concepts explained.

The third type of session involves ten two-hour sessions spread over the term. The group is divided into four subgroups that work in the computer laboratory at different times. Each subgroup has no more than 15 students. The teams carry out practical problem-solving exercises using statistical software. At the end of each session, they must produce a written report that includes a brief summary of what they have understood about the main concepts involved in the exercise they have done, together with their interpretation and discussion of the analysis.

The teaching material used for the three sessions – lectures, group activities, and computer laboratory tasks – is prepared by the course lecturers and available on the web page, which eliminates the need for extensive copying. Students may further their knowledge of statistical techniques using additional reference material such as [14] and [15], which provide a detailed analysis of statistical applications in environmental systems. Some of the in-class activities, in which students create data themselves or develop an understanding of randomness, have been built around activities proposed by [5] (1996). For instance, at the beginning of the course, students are asked to answer an anonymous questionnaire containing questions about personal aspects such as age, gender, weight, height, their opinion regarding the most serious problems affecting the environment in general, or the area in which they live, etc. The information compiled is then used to create data files, and students are encouraged to discuss whether or not they can be considered a representative sample for the questions posed. They are also asked, for example, to recognize the different types of data involved. This file is used in a number of practical assignments in the computer laboratory.

Several authors have remarked on the importance of having students solve statistical problems involving real data as this helps them to understand the value of using statistical methods in decision-making processes ([5], [6], [9]). Accordingly, the PUV's environmental science statistics course uses environmental case studies to develop certain concepts. The course lecturers have created a collection of environmental data sets – which are updated annually – for use in classroom and laboratory activities. For example, students apply descriptive analysis methods to atmospheric pollution data gathered by the automatic network managed by Valencia City Council's Environmental Service, or use graphical methods and sample statistics to analyze temperature and rain records provided by the National Institute of Meteorology. These activities facilitate the establishment of connections with other compulsory subjects on the syllabus such as Atmospheric Pollution, Meteorology, and Climatology. They are carried out under the guidance of the lecturer, who adopts the role of `facilitator' ([6]), and serve to motivate statistical concepts, activities and problems.

The connections with other degree subjects are established in accordance with the transversal approach needed to express educational objectives in terms of 'competences' and 'learning outcomes'. The identification of these objectives is part of the process required to create an integrated higher education area in Europe [16]. The Spanish educational system has been reformed in line with the recommendations of the Bologna Process for tuning European Higher Education structures.

The use of suitable statistical software in the teaching of statistics is extremely important. Moore [6] notes "A student who understands conceptually what must be done to solve a problem is much more likely to be able to implement that understanding when software is at hand". Statistical software reduces and sometimes even eliminates the time devoted to boring, irrelevant, and often practically impossible arithmetical calculations done by hand. It helps students to focus on questions relevant to the exposition of a problem and on the interpretation and critical analysis of the results. The statistical software of choice for teaching undergraduate courses is Statgraphics *Plus* ® and SPSS. The software is licensed for on-campus use through the PUV's local network. It is used in the laboratory classrooms and the computer rooms where students work on problems and projects outside of class time. The software has the features that [17] considers to be adequate for supporting learning and doing statistics in introductory courses. Statgraphics® are SPSS are easy-to-use and learn menu-driven software with high-quality graphics; they implement the most common statistical methods taught during the course, and also incorporate simulation tools. For example, in a statistical inference laboratory class, students might generate random data from different probability distributions using the software, save them in the program's data editor, and group them according to different sample sizes. The results can then be used to visualize how changing data distribution and sample size affect mean distribution. The students are introduced to the software in the first computer session before being asked to apply univariate descriptive techniques to environmental data analysis.

5 ASSESSMENT METHODS AND IMPLEMENTATION RESULTS

The educational literature recommends the use of a variety of performance assessment methods ([10]). In addition to exams, alternative assessment methods help to monitor the learning/teaching process during the course. This provides students and teachers with feedback on problematic concepts and applications, and helps to identify changes required to improve instruction. [8] note that "the identification of errors and difficulties which students display is needed in order to organize statistical training programs and to prepare didactical situations which allow the students to overcome their cognitive obstacles".

Written results from in-class activities, laboratory tasks, and exams are some of the proposed alternative methods which can be useful for assessing student and course performance. Following these recommendations, the environmental science statistics course assesses performance on the basis of marks obtained in exams, class activities and laboratory tasks. The students' written answers to in-class activities provide a weekly evaluation of what they have learned regarding the main concepts studied in the lectures. The ten laboratory task reports account for 30% of the final mark. The practical classroom and computer laboratory sessions also provide an ongoing evaluation of the students' performance in groups. The lecturers walk around the room, observing the groups and listening to them as they discuss results. This is an informal assessment of understanding and helps to detect problems such as negative group dynamics or unequal participation. The confidence intervals of the mean marks (over 10) obtained during the last academic year in the ten laboratory classes, is shown in Fig.1. The last computer laboratory classes obtained the lowest results given the difficulty of its contents: students have to apply more advance methods, such as descriptive multivariate tools for data analysis.

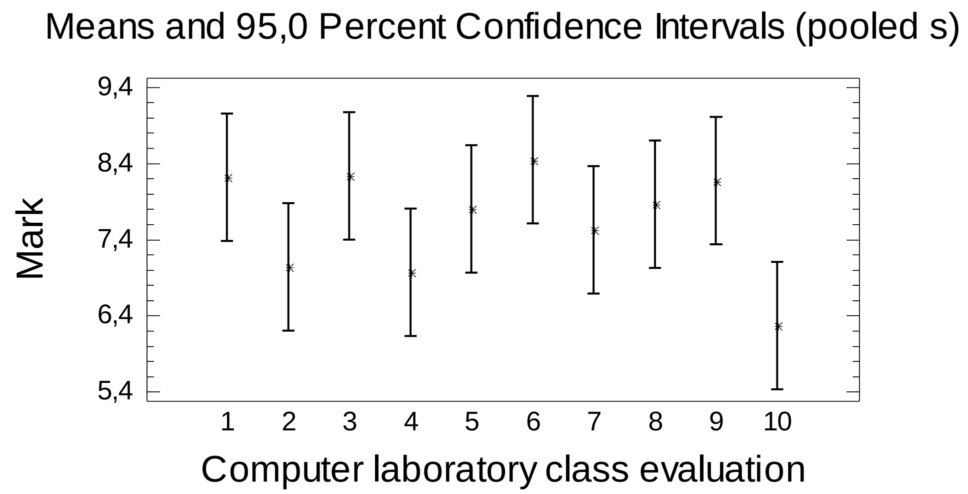

Fig. 1- Confidence intervals of computer laboratory evaluations marks.

At the end of the semester, students take a three-hour exam that covers the entire content of the course. In both exams, emphasis is placed on statistical analysis and interpretation skills. Some of the questions require students to interpret output from the statistical software used in the laboratory classes. The final exam accounts for 70% of the final mark. The pass rate is over 60% and two fifths of those who pass do so with high marks. Fig. 2 shows the evolution of the pass rate during the years in which the course has been taught.

Students are highly motivated and play an active part in classes. Before taking the statistics course, they have already completed three years of higher-level education and are more mature than students starting out at university. In addition, there are other courses on the degree program that involve active learning, so they readily accept the teaching methodology employed. About 40% of the students have worked in engineering positions outside the university, and a number have experience with environmental issues which require the application of statistical methods. These students contribute greatly to the course; they are able to provide real examples of environmental problems encountered during their work, and they also tend to participate more in the discussions. Students who have completed an introductory statistics course (descriptive analysis and probability) in their first-cycle studies acquire the knowledge somewhat faster and obtain better marks on average than those with no experience in statistics. However the progress of the latter is not problematic as the course moves through the syllabus and keep up.

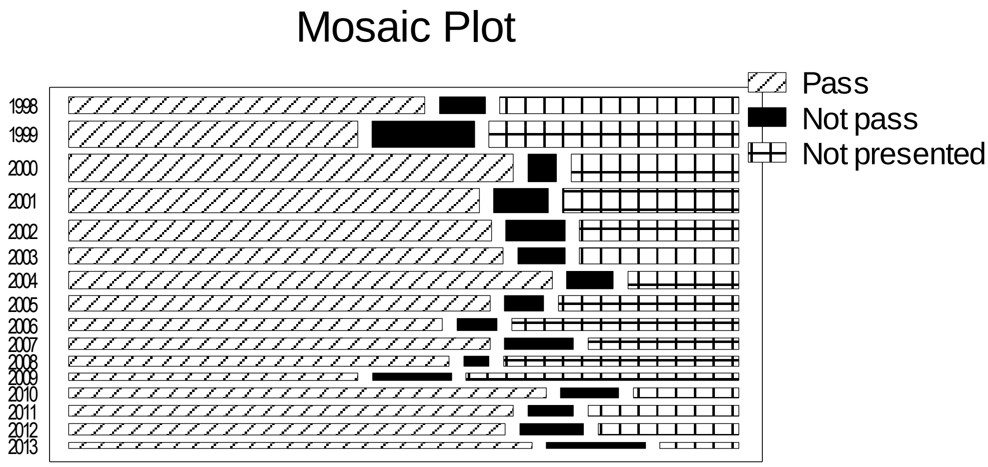

Fig.2- Evolution of pass, not pass and not presented rates since 1998.

Students react well to the educational approach. The activities and laboratory classes are updated annually on the basis of the previous year's experience and students' recommendations for improvements. During the semester, lecturers use brief anonymous surveys (consisting of two or three questions) to ask students questions such as "How did you like this activity?" or "What did you find most difficult to understand?" This allows the school to monitor lectures and activities and detect problems well in advance. At the end of the semester, students complete a 45-question survey designed by the Institute of Educational Sciences at the UPV. The Institute analyzes the survey and sends its findings to the lecturers every year. The questions cover aspects such as class atmosphere, teaching materials, assessment methods, and lecturer performance. The environmental science statistics course scores well in these opinion surveys. Students consider that the course arouses interest and that the concepts and methods learned will be very useful for their future careers. Lecturers also receive informal comments from students about how much they have enjoyed the course or how much they have learned from the activities. These results and comments serve to increase the lecturers' own levels of motivation and satisfaction.

Feedback from former students is another good way of assessing the course. These students frequently comment that they use statistical methods to solve problems in other subjects or in their own work outside the university. This proves that they have incorporated statistical concepts and techniques into their knowledge and use them in their problem-solving and decision-making processes.

REFERENCES

[1] Piegorsch, W.W. & Edwards, D. (2002). What Shall We Teach in Environmental Statistics?. Environmental and Ecological Statistics, 9 (2), pp. 125-150.

[2] Barnett, V. (2004). Environmental Statistics, Chichester, UK: Wiley.

[3] Hunter, J.S. (1994). Environmetrics: an Emerging Science. In G.P. Patil & C.R. Rao (Eds.), Handbook of Statistics XII: Environmental Statistics (pp 1-7). Amsterdam: North Holland.

[4] Guttorp, P. (2003). Environmental Statistics- a Personal View. International Statistical Review, 71(2), pp. 169-179.

[5] Scheaffer, R.L., Gnanadesikan, M., Watkins, A. & Witmer, J. (1996). Activity-Based Statistics. New York: Springer-Verlag.

[6] Moore, D.S. (1997). New Pedagogy and New Content: The case of Statistics. International Statistical Review, 65(2), pp. 123-137.

[7] McGinnis, J.R., Wanatabe, T. & McDuffie, A.M. (2005). University Mathematics and Science Faculty Modelling their Understanding of Reform Based Instruction in a Teacher Preparation Program: Voices of Faculty and Teacher Candidates. International Journal of Science and Mathematics Education, 3 (3), pp. 407-428.

[8] Batanero C., Godino, J.D., Vallecillos, A., Green, D.R. & Holmes, P. (1994). Randomness, its Meanings and Educational Implication. International Journal of Mathematical Education in Science and Technology, 25(4), pp. 527-547.

[9] Garfield, J. (1995). How Students learn Statistics. International Statistical Review, 63(1), pp. 25-34.

[10] Gal, I. & Garfield, J. (Eds.) (1997). The Assessment Challenge in Statistics Education. Amsterdam: IOS press.

[11] Patterson, N.D. & Norwood, K.S. (2004). A Case Study of Teacher Beliefs on Students' Beliefs about Multiple Representations. International Journal of Science and Mathematics Education, 2(1), pp. 5-23.

[12] Weldon, K.L. (2002). Advanced Topics for a First Service Course in Statistics. In B. Phillips (Ed.), Proceedings of the Sixth International Conference on Teaching Statistics, CDROM, Cape Town, South Africa. International Statistical Institute.

[13] Box, G.E.P. (1999). Statistics as a Catalyst to Learning by Scientific Method. Part II- A discussion. Journal of Quality Technology, 31(1), pp. 16-29

[14] Gilbert, R.O. (1987). Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold.

[15] Berthouex, P.M. & Brown, L.C. (2002). Statistics for Environmental Engineers, Second Edition. Boca Raton, FL: CRC Press.

[16] Berlin Summit on Higher Education (2003). http://www.bologna-berlin2003.de.

[17] Biehler, R. (1997). Software for Learning and for Doing Statistics. International Statistical Review, 65(2), pp. 167-189.