A FIRST COURSE IN STATISTICS FOR COMPUTER ENGINEERING UNDERGRADUATE STUDENTS

Carmen Capilla[1]

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

Abstract

The Polytechnic University of Valencia (PUV) in Spain, has a special School of Engineering in Computer Science. The school offers a Bachelor's Degree in Computer Engineering. A first service course in statistics is a compulsory subject for students in the first semester of the degree. This paper looks at the background behind the course, describes its contents and structure, and presents the assessment methods and results. Statistical education for computer engineers tries to give them a solid foundation in statistics. An emphasis is placed on mastering a wide use of statistical methods in order to allow the students to apply these techniques in many fields of computer science. About 400 students enroll each year in the subject. Basic knowledge is important, but not more important than creating in the student a positive attitude towards statistical methods.

Keywords: Computer engineering studies, statistical education, innovation, integrated higher education area in Europe.

1 INTRODUCTION

The PUV introduced in the year 2010 an undergraduate degree in Computer Engineering. This degree was designed taking into account the recommendations of the Bologna Process for tuning European Higher Education structures. A transversal approach was 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 ([1]). The process has entailed a complete restructuring of the existing Computer Engineering program, as well as the teaching methodology. The resulting present program is a four-year degree which is run by the School of Engineering in Computer Science.

A first service course in statistics is a compulsory subject for students in the second semester of the new degree. It's a single-semester introductory course in statistics. The application of statistical principles and methods is necessary for effective practice in resolving the different problems that arise in the many branches of Computer Science and Engineering activity ([2], [3]). In recent years the Accreditation Board of Engineering and Technology (ABET) has recognized the correct use of appropriate statistical methods, as an important requirement for undergraduate engineering education ([4], [5]).

Statistical education for computer engineers tries to give them a solid foundation in statistics. An emphasis is placed on mastering a wide use of statistical methods in order to allow the students to apply these techniques in many fields of computer science. The challenge in teaching statistics in any science is to truly engage the practitioner/non-statistics student. This paper describes the methodology used in the statistics course of the computer engineering degree of the PUV, the assessment methods and the results obtained.

2 METHODOLOGY

New engineering criteria have been designed to measure the effectiveness of engineering education programs, by focusing on an assessment and evaluation process that assures the achievement of a set of educational objectives and outcomes. The old system of counting course credits was modified and replaced by an outcomes-based process. Outcomes-based assessment focuses on what students have learned, or what they actually can do, at the time of graduation. Professional skills which have to be achieved include communication, teamwork, understanding ethics, professionalism within a global and societal context, lifelong learning, and knowledge of contemporary issues.

At the time of graduation, computer engineering students should have the ability to apply the method they have learned, among which statistical tools are included. During the last decades, statistics education has been involved in a reform process, which has led to a review of content, pedagogy, and technology at all levels of education [6-8]. Recommended changes in these three areas have gradually been adopted as a means of offering continuing improvement in the teaching of statistics. Statistics education experts recommend:

- 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, and
- improving students' technology and communicating skills regarding data and chance.

These recommendations contribute to the achievement of professional skills. One factor that has a great influence on student's learning is his previous knowledge of the subject [9]. All the students that take the statistics course of the Computer Engineering degree, have calculus and algebra courses during the first semester of this program at the university. Their background in statistics, however, ranges from no statistical experience whatsoever to basic skills acquired in a secondary level introductory statistics course (descriptive statistics and probability). No assumptions regarding statistical knowledge and experience were therefore made when designing the statistics course for the Computer Engineering degree. The course was designed to prepare students "to use statistical thinking and reasoning" [10], and provide them with the necessary skills to interpret and evaluate data analysis. Mathematical proofs and calculations are given secondary priority. Derivations and hand calculations are replaced by the widespread use of statistical computing methods, and an emphasis is placed on the analysis of real data.

Table 1 gives the contents of the course syllabus and the number of hours assigned to each part in the lectures and in the computer laboratory classes. The order of contents in the table corresponds to the same that is applied in the sequence of lectures during the semester. The syllabus includes (i) measures of centre, variability, descriptive charts; (ii) probability laws; (iii) binomial and Poisson distributions; (iv) uniform, exponential and normal distributions; (v) sampling distributions and confidence intervals for means and differences, *p*-values; (vi) one way analysis of variance, post hoc tests, and two way analysis of variance, and (v) simple linear regression.

Students attend lectures that present the basic statistical concepts through the use of illustrative examples, and practical sessions which consist of in-class activities and computer laboratory classes. Course content covers subjects such as univariate descriptive analysis, which looks at different ways of presenting data (e.g. frequency tables, bar and pie charts, histograms, box-whisker plots). Concepts such as population, random samples, data types and variability are introduced through examples taken from real data analysis.

Probability and statistical distributions are studied in the second module of the course. The lectures of this part focus on helping students to understand the concepts of chance and randomness, which are given greater priority than formal probability ([7]). A number of statistical education researchers that students of all ages have frequent difficulties with these concepts ([10]). 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 distribution applied to real data. 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.

The third module of the course focuses on basic inference. 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 ([9]). Students also study the construction and interpretation of confidence intervals for means and standard deviations. In the final part of the module, we discuss the applications of one- and two-sample t-tests as useful procedures for examining evidence against null hypotheses.

Table 1- Course contents and number of hours of classroom lectures and computer laboratory sessions for each module.

Content	**Number of hours**	
	Classroom lectures	Computer room sessions
Introduction	1	1:30
Univariate descriptive statistics	3:30	1:30
Basic probability concepts	6	
Discrete distributions	3	1:30
Continuous distributions	6	3
Sampling distributions	3	
Basic inference in one population	3	1:30
Comparison of two populations	3	1:30
Analysis of Variance	6	1:30
Bivariate descriptive statistics	1:30	0:45
Simple linear regression	6	0:45
TOTAL	42	19

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. Before studying the simple regression model, covariance and correlation measures are presented during the descriptive analysis of quantitative bivariate variables. Several examples are used to illustrate that correlation and association do not necessarily indicate causation.

European Credit Transfer and Accumulation System (ECTS) is a standard for comparing the study attainment and performance of students of higher education across the European Union and other collaborating European countries. For successfully completed studies, ECTS credits are awarded. One academic year corresponds to 60 ECTS-credits that are equivalent to 1500–1800 hours of study in all countries respective of standard or qualification type and is used to facilitate transfer and progression throughout the Union. In Spain this system has been recently adopted, and the new degree in Computer Engineering has been designed according to it. The number of hours per ECTS credit point in Spain is 25. The statistics course has been assigned 6 ECTS credits. Table 2 gives the number of hours per each module that students would employ with classroom and private work.

Table 2- Number of hours of student’s classroom and private work for each module.

Content	**Number of hours**	
	Classroom work	Private work
Introduction	1,5	3
Descriptive statistics	10	12
Basic probability concepts	6	8
Probability distributions	21,5	31
Inference (Basic inference, ANOVA,regression)	34,5	48
TOTAL	73,5	102

The classroom work includes the theory lectures, the seminars, the in-class activities, the computer laboratory classes, and the assessment activities (e.g. exams). The difference with the total number of hours given in Table 1, is due to the fact that in Table 2 seminars and assessment activities are taken into account. The private work refers to the task that students perform on their own or working in groups outside of the class.

Approximately 400 students take the statistics course every year. They are divided into 10 groups. In one groups classes are taught in English, in two other ones in valenciano, the local dialect of Valencia region. Theoretical lectures are attended by each entire group, while computer sessions are taken by smaller subgroups. Classes are taught by members of the Polytechnic University´s Department of Applied Statistics and Operations Research and Quality. Teachers´ backgrounds and research interests are mainly in statistics applied to engineering and quality control.

Teaching loads vary from three to six hours per week and per group. 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 section which is on-line accessible through the platform PoliformaT. This platform has been adopted by the PUV as part of the Sakai project in cooperation with some American universities, and offers resources for teaching the different courses. The teaching material used for the class sessions – lectures, group activities, and computer laboratory tasks – is prepared by the course lecturers and available on PoliformaT, which eliminates the need for extensive copying.

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 [6]. 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 country in general, 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.

Basic knowledge is important, but not more important than creating in the student a positive attitude towards statistical methods. We must convince students of the great value of these methods as tools for data analysis and decision-making in real problems that will arise in their future professional work. The only way to succeed in this is through the formulation and solution of real, or at least realistic, problems of direct interest to students. This must be done using the scientific method and sharing the teacher's experience in real projects. To encourage the students' active participation, we have reduced the time spent in lecture classes and increased individual and team work and discussion. The role of the teacher has changed from that of "source of information" to "facilitator of learning" [7].

All class hours are divided into two types of sessions. The first session lasts three hours every week, and consists of explanation by the teacher and, individual study by the students and discussion of actual problems introduced by the teacher using the Statgraphics statistical package. Some in class-activities are included to work with the methods taught. The second session lasts one hour and thirty minutes, and takes place in a computer room. There are nine laboratory classes during the semester. In these laboratory classes the students work in teams of two or three during one hour and thirty minutes. They carry out practical tasks using Statgraphics to solve actual problems. This software is licensed for campus use through the university´s local network. It is used in the laboratory classrooms and in the computer rooms where students work outside of class time. The software has the features to be adequate for supporting learning and doing statistics in introductory courses [11]. Statgraphics® is an easy-to-use and learn menu-driven software with high-quality graphics; it implements the most common statistical methods taught during the course, and also incorporates simulation tools.

3 ASSESSMENT ACTIVITIES AND RESULTS

The education research literature recommends the use of a variety of performance assessment methods [12]. Assessment of the teaching/learning process is crucial to identify the errors and difficulties which students display, and is needed to organize training programs and to prepare didactical situations to overcome students' cognitive obstacles [9]. 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. On the other hand, the ECTS credits system implies a greater emphasis on continuous assessment activities, giving much less importance to the traditional exams.

Table 3 shows the assessment activities used in the statistics course of PUV degree in Computer Engineering. Students' performance is assessed on the basis of marks obtained in in-class activities, student's behaviour during the classes, computer laboratory tasks, and exams.

Table 3- Assessment activities weight in the final mark.

Activity	Weight in the final mark
Observation	10%
Mid-term exam	27,5%
Final-term exam	27,5%
Computer laboratory evaluations	35%

Observation refers to results in in-class activities and student's behaviour; it accounts for 10% of the final mark. In-class activities provide a continuous assessment during the semester. They consist in exercises that students have to do individually, using the course materials, in the classroom. There are four assessments of this type. Observation of students behaviour during the classes (e.g. if they participate or not, if they answer question done by the teacher to the whole group), is taking into account to rate students' performance.

There is a two-hour midterm exam with questions on descriptive analysis and probability distributions. This exam is for students that follow the course with a learning contract. This mid-term exam may account for 27,5% of the final mark. At the end of the term, there is another three-hour exam that covers the inference module 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-term exam accounts for 27,5% of the final mark. In the time period between semesters, there is an exam to recuperate any of the two other exams that students may have failed.

The computer laboratory sessions provide an ongoing assessment of what they have learned regarding the main concepts studied in the lectures. The ten laboratory tasks that teams answer at the end of these sessions account for 35% of the final mark. The computer laboratory sessions also provide an ongoing evaluation of the students' performance in groups. An on-line evaluation must be completed by each team at the end of the laboratory session, using the PoliformaT platform. The evaluations consist in multiple choice tests. During the computer laboratory classes, 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.

Fig.1 gives the scatterplot matrix of the marks in these four activities, of students of one of the groups, during the academic year 2012-2103. Table 4 contains the significative linear correlations between the marks.

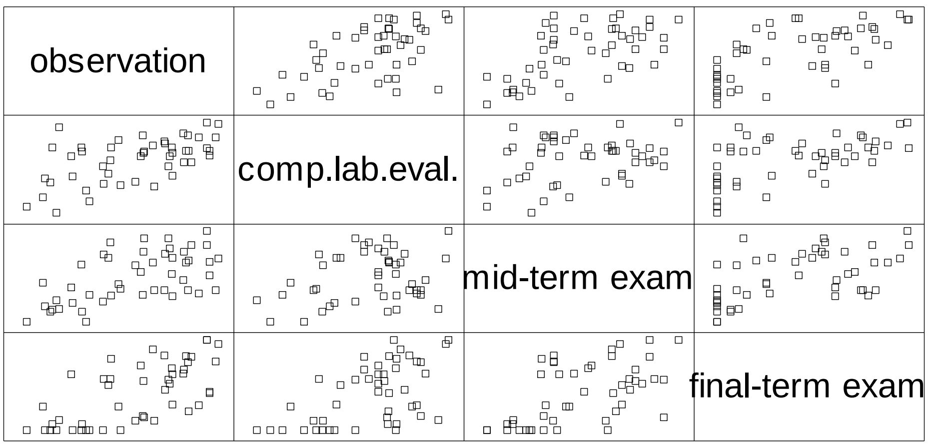

Figure 1- Scatterplot matrix of the marks obtained in the statistics course (year 2012-2103)

Table 4- Correlations between the marks of the assessment activities in the year 2012-2013.

	Observation	Comp.lab.eval.	Mid-term exam	Final-term exam
Observation	1			
Comp.lab.eval.	0,57	1		
Mid-term exam	0,59	NS	1	
Final-term exam	0,67	0,56	0,57	1

The correlation between the computer laboratory evaluations and the mid-term exam marks is not significative. The other ones indicate an intermediate relationship between the results. There were eight on-line evaluations during the semester in the year 2012-2013. Fig.2 plots the confidence intervals of the mean marks obtained (over 10) in those evaluations, in the same group of Table 4 and Fig. 1.

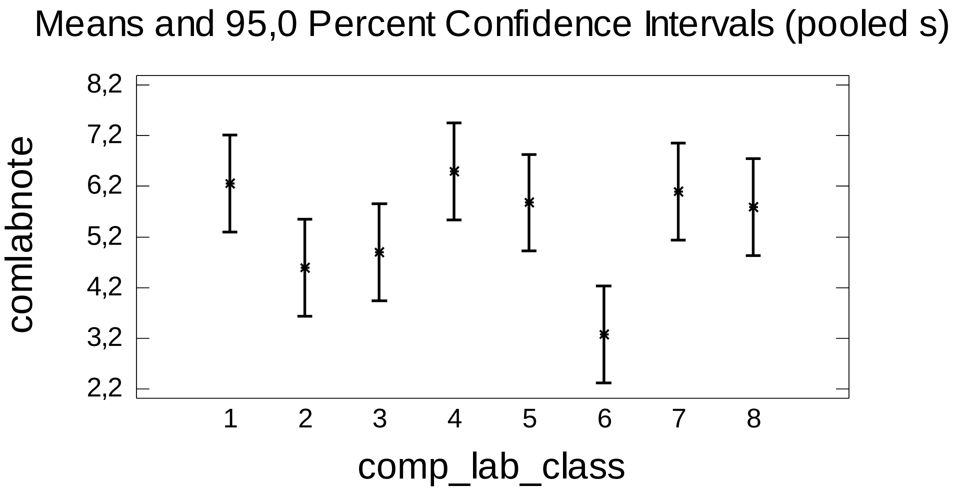

Figure 2- Confidence intervals for the mean marks of the computer class evaluations.

The mean mark of evaluation 6 was lower. The contents of that evaluation were related to inference methods to compare two normal populations. The identification of the difficulties that students' had with the application of these methods, will be useful to redesign the teaching methodology on the subject.

This teaching methodology has increased class attendance to 90%. Academic results have also improved, in spite of the difficulty and complexity of the evaluations questions. More than 90% of the students attend the exams, and the pass rate is around 70%. Figure 3 represents the final results statistics of the three academic years in which the new degree has been taught.

The most positive result of the experience, however, is the student attitude towards the subject. At the end of the semester, students complete a survey designed by the Institute of Educational Sciences at the Polytechnic University of Valencia. 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 statistics course of the Computer Engineering degree 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.

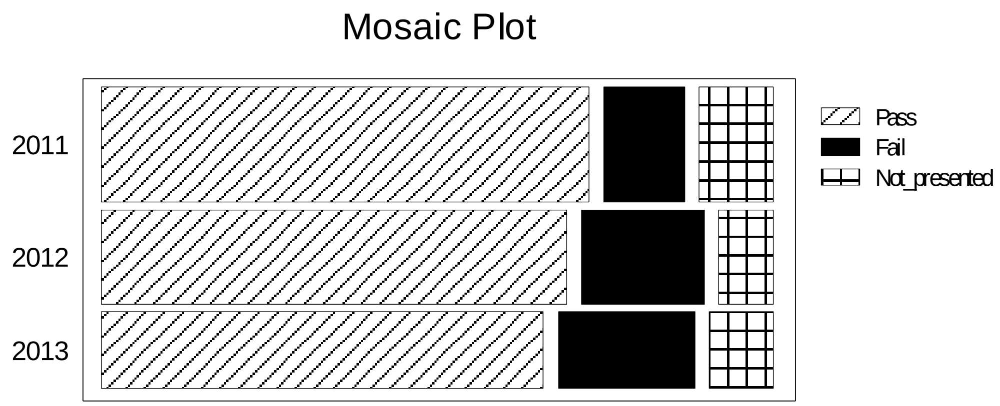

Figure 3- Statistics of the final academic results of the course.

4 CONCLUDING REMARKS

This paper has presented the statistics course that is currently taught in the Computer Engineering degree at the PUV. The course follows the guideline established in the process to create an integrated higher education area in Europe. More emphasis is placed on the students work outside the classroom, and in continuous assessment methods. 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 application of statistical principles and methods is necessary for effective practice in resolving the different problems that arise in the many branches of Computer Science activity. Statistical education for computer engineers tries to give them a solid foundation in statistics. An emphasis is placed on mastering a wide use of statistical methods in order to allow the students to apply these techniques in many fields of computer science.

Students that follow the course are highly motivated and play an active part in classes. Their marks and pass rate are higher than those of past system students. Students react well to this educational approach. In addition, the other courses on the degree program that involve active learning and the use of continuous assessment, so they readily accept the teaching methodology employed. The activities and laboratory classes are updated annually on the basis of the previous year's experience and students' recommendations for improvements. The pass rate is around 70%. The most positive result of the experience, however, is the student positive attitude towards the subject. At the end of the year the students complete a survey on every subject. Statistics obtains very good results in these opinion surveys and is considered one of the most useful subjects because of the interest it arouses and its usefulness in the computer engineers' future careers.

REFERENCES

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

[2] Jain, R. (1991). The Art of Computer Systems Performance Analysis, New York, John Wiley & Sons.

[3] Ryan, P.T. (2007). Modern Engineering Statistics, Hoboken, NJ, Wiley.

[4] ABET (2012). About ABET. Available on-line http://www.abet.org/about-abet/.

[5] Al-yahya, S.A. & Abdel-halim, M.A. (2013). A Successful Experience of ABET Accreditation of an Electric Engineering Program. IEEE Transactions on Education, 56(2), pp.165-173.

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

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

[8] 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.

[9] 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.

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

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

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