A Distance Teaching Course in Statistics for the Degree in Economics and Business Administration

Luciano Pieraccini - Rome, Italy

1. Introduction

The project on distance teaching of statistics at university level is part of a wider project for the degree in Economics and Business Administration that is now in progress under the direction of CUD (Distance University Centre - University of Cosenza). The project aims at substituting the first two years' courses for the degree in Economics and Business Administration with distance teaching to cope with the ever-increasing number of students in that field. The project is still in a very early stage as far as the course in statistics is concerned, so that the paper is mainly centred upon the problems of determining its outline, the choice of media to use as teaching support, and the self-contained procedure of learning control. The course in statistics is focussed on so-called "Descriptive Statistics" although its final part introduces the fundamental ideas of statistical inference. A second course in statistics should expand this area of statistical knowledge while deepening the treatment of both univariate and multivariate statistical analysis. The aim of the course is to equip students with the basic methodological tools needed to approach empirical studies in economics and firm management. In this light, sufficient real data analysis is introduced to illustrate statistical procedures. Univariate, bivariate, and multivariate analysis, from a descriptive point of view, are covered by the course, while some optional, more advanced topics are introduced along the course, to open new horizons and give some hints on what else statistics can do. One further main object of the course is to enable students to use standard statistical packages.

Particular attention has been given to the problem of self-evaluation: the student is subject to regular tests whose results are the key for either going back for a new look at earlier concepts or proceeding on to new topics. During the course, students are given examples to explain the meaning of the statistical procedures introduced, while at the end of each unit, tests allow them to evaluate their understanding of the concepts introduced. At the end of each group of units, more general tests, mainly in the form of exercises spanning all the material covered, are proposed, aimed at giving the students a feeling for

how topics developed in the units have a unified meaning and use. The whole procedure is intended to allow students to evaluate their progress themselves and choose whether or not to deepen their study of particular, sometimes optional, topics.

With regard to material and media to be used, preference has been given to written material. Other media are involved in a few videotape introductory lectures, and in computer-assisted teaching, both for lectures and interactive exercise solving.

In this paper an outline of the course is presented in Section 2, while in Section 3 the self-evaluation procedures and intended media are sketched. In Section 4 a few words of conclusion end the communication.

2. Outline of the course

The first problem that a team of full professors of statistics, coming from many Italian universities to work on the project, had to solve was what kind of statistics course to choose. The decision was to concentrate mainly on so-called descriptive statistics, on the grounds that a less formalised introductory course should be of greater help for the student to better understand the kind of empirical problems statistics is called to work on, whereas a second, more formalised course should be devoted to statistical inference. On the other hand, this choice has not been exclusive since some optional, more advanced topics, including some basic ideas on statistical inference in the last part, have been introduced here and there along the course.

The course has been organised in fifteen teaching units grouped into six parts. A brief sketch of topics included in each part is given here.

Part I - Introductory Concepts (units 1 and 2): This part is devoted to the introduction of basic statistical concepts. After outlining the difference between deterministic and statistical analyses of empirical phenomena, concepts of collection of units, observations, and statistical variables are introduced, both with reference to censuses and sample surveys. Some of the more common sampling techniques are briefly sketched here, referring the student to the final part of the course for further developments. In this context, problems linked to the designing of a questionnaire and to its use are briefly illustrated.

The organisation of observations in a database to obtain synthesis of information is considered, with particular regard to the construction and meaning of one-way, two-way, and multi-way tables, stressing the need for self-contained interpretation. Graphical representation of statistical data is introduced as a tool for better understanding the peculiarities of the data, giving a meaningful visual image of the information embodied in them. Several kinds of statistical graphics are introduced: two- and three-dimensional representations, polar diagrams, etc., stressing the importance of the best choice between them and the danger of misleading impressions.

A brief review of the more important national and international statistical data (ISTAT, OCSE, ONU, etc.) ends this part.

Part II - Univariate Statistical Distributions (units 3, 4 and 5): This part is devoted to the study of univariate distributions through several measures characterising their more important features.

Location measures such as median, quartiles, arithmetic mean, etc. are introduced stressing the major unifying approaches to location measures. Particular emphasis is given to the arithmetic mean whose properties are studied in detail, while quadratic,

geometric, and harmonic means are briefly introduced, mainly to show their use in specific contexts. The importance of median and quartiles in discussing outliers is stressed.

Measures characterising the variability of univariate distributions are then introduced, both as measures of dispersion about location parameters, and as measures of differences between observations. In this light, absolute mean deviation and quadratic mean deviation with respect to the principal location parameters are introduced, giving particular emphasis to the standard deviation and variance. As far as measures of differences between observations are concerned, mainly the mean absolute difference is introduced. Finally, measures of asymmetry of distributions are considered with particular regard to Pearson's β.

In the third unit of this part (unit 5), measures of concentration are introduced in the context of income distribution. The concentration ratio and δ index are studied in detail. Finally, the problem of fitting analytic curves to empirical distributions is considered, starting from the choice of the curve, the method to use, and ending up with measures of goodness of fit. In particular, normal and lognormal models are considered.

Part III - Index Numbers and Time Series (units 6 and 7): This part is devoted mainly to two particular topics, both related to univariate analysis, but with an implicit temporal ordering.

The first unit (unit 6) is mainly devoted to definition and properties of index numbers; Laspeyres and Paasche formulae are introduced showing their properties and limits. The more important index numbers of economic series produced by the Italian Central Statistical Institute (ISTAT), such as indices of prices and quantities of goods, cost of living index, industrial production index, etc., are presented.

The second unit (unit 7) is devoted to classical time series analysis starting from moving averages and with particular emphasis on exponential smoothing. The most important methods of classical decomposition of time series into trend, cyclic, and seasonal components are introduced, while a brief optional part on Box and Jenkins methods for the analysis of time series ends the unit.

Part IV - Bivariate Statistical Distributions (units 8, 9 and 10): The study is focussed on the concept of dependence between variables, its consequences, and ways of measuring it.

After introducing marginal and conditional distributions, together with their corresponding moments, the concept of dependence between variables is introduced, focussing on different measures of dependence. The introduction of the contingency table leads to the definition of "Chi square" and the derivation of its more important properties. Then Pearson's correlation ratio is defined as a further measure of dependence. Last but not least, the correlation coefficient is introduced as a measure of linear dependence.

The problem of linear and non-linear regression and least squares fitting is studied, stressing the measurement of goodness of fit. In the context of linear regression the squared correlation coefficient is shown to be the right measure, while for non-linear regression the determination index is introduced. As a further step along this line the analysis of residuals from regression is introduced, mainly through their graphical representation.

Part V - Multivariate Statistical Distributions (units 11 and 12): After a brief introduction to multivariate distributions, the multiple regression problem is studied, making use of matrix algebra. In this context, squared partial correlation coefficients are

introduced to measure the net linear effect of independent variables on the dependent one.

Measures of goodness of fit and criteria for the choice of the model using transformations of variables are introduced, and analyses of residuals are shown. A few hints on regression diagnostics are given at this stage.

The next unit (unit 12) is an optional one and is intended to give some idea of the most important techniques of multivariate analysis, mainly through examples, to show their potential in empirical studies. Cluster analysis and multidimensional scaling are briefly presented as methods of classification, while principal components and correspondence analysis are introduced as ways of analysing the structure of variance and covariances of a set of variables defined on a number of observations. Finally, canonical correlation and multivariate regression are introduced.

Part VI - Probabilistic Models and Statistical Inference (units 13, 14 and 15): This part is largely optional. In the first unit, which is mainly compulsory (unit 13), the need for a step beyond descriptive statistics - namely statistical inference - is stressed. At this point basic concepts of probability are introduced in an axiomatic approach, both from classical and subjective points of view, stressing De Finetti's contributions. Conditional probabilities are then considered and Bayes' theorem is presented, giving particular emphasis to its importance for statistical inference and decision theory. The definition of random variables is then introduced, together with their moments, while the last section of the unit, dealing with types of random variables, is optional.

The second unit (unit 14), which is the only compulsory one, deals in its first section with basic ideas of samples and sampling distributions of estimators of population parameters. In sections 2 and 3, both optional, properties of point estimators are considered, together with methods for getting them, such as least squares and maximum likelihood. Finally, interval estimation is presented.

The third unit (unit 15) begins with a first compulsory section devoted to the general principles of testing statistical hypotheses, definition of first and second type errors, and procedures of testing.

Section 2 introduces the principles of hypothesis testing in a more formal decision theory approach, while in sections 2 and 3, likelihood ratio tests are introduced formally with derivation of tests for the difference between two means and analysis of variance. The Chi square tests for independence and goodness of fit are also presented.

3. Self-evaluation procedure and media to be used

While a final written and oral examination has to be passed by the student, a procedure has been introduced to allow students to evaluate by themselves their understanding and ability to solve exercises. To this end, three kinds of tests are planned.

The first is a preliminary one to be done at the beginning of each part of the course to establish whether the student already has the kind of knowledge needed to understand the concepts introduced. Prerequisites can be specific, with regard to what has already been introduced in previous parts of the course, or of a general kind, in the sense that knowledge of particular topics of other disciplines (mainly mathematics) may be needed.

The second kind of test is planned for the end of each unit, to control how much has been understood of the topics presented, both from theoretical and practical points of view. In each unit, examples help the student to understand better the meaning of

statistical procedures while, at its end, exercises of the same type are given for the student to solve.

The third kind of test is a more general one that is planned at the end of each part, with the purpose of summarising what has been introduced in the part itself and those which precede it. The attempt is to give the student a feeling of how topics developed so far can have a unified meaning and use.

The results of the tests are the key for either going back to have a new look at concepts introduced formerly, or to proceed to new topics. For concepts introduced previously, reference is given to the unit in which they were presented and to those related to it; for example, from other disciplines, reference is made to the part of the specific course (for instance mathematics) in which that topic is presented.

If the test results are particularly good or bad, advice is given to the students to go into particular optional topics or to skip them. In this way it has been possible to design several paths of learning according to the ability shown by the student: from a minimal one, below which the examination cannot be passed, to an optimal one, in which new horizons are opened and some hints are given on what else statistics can do.

As mentioned before with respect to media, written material has been given greatest importance in supporting distance teaching. At the present stage of the project it is almost complete. Great attention has been given to all those devices that can be used to make learning easy.

This does not mean of course that other media will not be used. In point of fact, an introductory videotape lecture for each one of the six parts of the course is planned, to give a general presentation and to illustrate the kinds of problems which will be dealt with. The use of graphic representation and animated cartoons will be largely adopted.

Great importance will be given to the use of computers both for assisted learning and interactive exercise solving. Computer programs for several hours of assisted learning (at least two hours for each part) are planned to run in parallel with written lectures. The main purpose is to stress general concepts and to show how to proceed in empirical studies with the statistical tools introduced. On the other hand, interactive exercise solving should be, for the student, a kind of guide on how to approach a problem and how to solve it, what kind of choices to make, and what kind of tools to use.

In this context it has to be pointed out that one of the main objects of the course is to enable students to use one of the standard statistical packages such as SAS, SPSS, etc. without difficulty. Students will be introduced to their use through exercise solving, at first without even knowing they are using a statistical package. Afterwards, it will be made explicit, together with all the explanations needed to learn how to use commands for input, data processing, and output. Particular attention will be given to correct interpretation of output in the light of theoretical knowledge already acquired.

4. Conclusions

The early stage of the project as already pointed out does not yet allow any sensible evaluation of its effectiveness to be made. The only thing that can be said on the work done so far is that the project seems to have started in the right way and to be proceeding on the right track. Its performance in practice can only be judged after its implementation, but we hope that it will perform well.