ON TEACHING STATISTICS AND ECONOMETRICS USING MODERN ICT METHODOLOGIES

MIŞCOI, Gheorghe
Hab., Ph.D. Prof., Member of the Academy of Science of Moldova

POPESCU, Ion
Spiru Haret University, Bucharest, Romania

ALBEANU, Grigore
Spiru Haret University, Bucharest, Romania
g.albeanu.mi@spiruharet.ro

GHICA, Manuela
Spiru Haret University, Bucharest, Romania
m.ghica.mi@spiruharet.ro

Abstract

This paper presents an approach on teaching statistics and econometry using modern ICT methodologies. Details on time series teaching and methods to prepare a statistics course are given.

Keywords: *ICT, E-learning, time series, statistics, applets, simulation*
ACM classification: K.3.1

1. Introduction

In this paper we investigate different levels of e-content structuring both in statistics and in time series modelling, analysis and forecasting. In the second section we cover the e-content structuring for the effective learning of methods, algorithms and tools for statistics, while in the third section we provide details on time series modelling, analysis and forecasting. Not only introductory methods are considered, but also advanced topics including those based on soft computing field. The concluding remarks will close the subject presentation

2. ICT and statistics

It is generally accepted that Information and Communication Technologies (ICT) can be used to add value in teaching. The E-courses have to be more than an electronic presentation of the disciplines. In order to produce good E-courses, which increase the quality of teaching and learning in classrooms, or at distance, the following ideas can be considered:

1. The E-course will be an environment providing opportunities to make conjectures test and modify them, according to the **learning from feedback principle**.

2. The environment will support the making and justifying the generalizations by successive increasing the size of the samples (observing patterns).

3. The changing of representations will be supported for a better understanding of the connections between representations (seeing connections).

4. Generation and editing of images describing the geometrical structure of data, or of different entities during data analysis is required also (working with dynamic images).

5. The environment will permit data interpretation and analysis in order to work with real data (exploring data).

6. Inserting formulas or code to be interpreted by environment will be supported (teaching the environment).

The main chapters of a course in statistics will address (or other references in the field):

- The basics of statistics (descriptive statistics, inferential statistics, sampling, variables, percentiles, measurement, distributions, and linear transformations)
- Graphing (charting) distributions (histograms, frequency polygons, box plots, bar charts, line graph, etc.)
- Distribution summarization (measuring central tendency, mean and median, variance, additional measures, effect of transformations, etc.)
- Bivariate data description (Pearson correlation, properties and computing aspects)
- Basic of probability theory (conditional probabilities, uniform distribution, binomial distribution, the Bayes' Theorem etc.)
- Normal distribution (standard normal, approximations, properties)
- Sampling distributions (of the mean, difference between means, Pearson's correlation, proportion)
- Estimation (degrees of freedom, characteristics of estimators, confidence intervals)
- Logic of hypothesis testing (significance testing, type I and type II errors, one- and two-tailed tests, interpreting results)
- Testing means
- The power of a test
- Prediction (at least linear regression)
- ANOVA (one- and multi-factor)
- Chi square (distribution, one-way tables, contingency tables)

The environment will include case studies and different calculators (or searching through tables).

Moreover, the style and presentation will be considered to permit navigation (meaningful buttons, access to the Table of contents, etc.), understanding (minimum text but maximum information), and progress in learning (considering pedagogic factors, formative evaluation etc.), contiguity (between text and simulations), using meaningful data for demos, self-monitoring (personal reflection and self-assessment).

3. ICT and econometrics

The content of a course in econometric is very rich. In the beginning of the project we will cover time series aspects. Both univariate and multivariate time series will be considered. An interesting feature of discrete univariate time series can be described using autocovariance and autocorrelation functions. The main difference between deterministic data and random time series shows the persistence of the autocorrelation functions along the time displacements in the deterministic case.

Let $k \geq 1$, and r_k be the k^{th} order correlation coefficient between observations separated by k time units. The array of autocorrelation coefficients r_k, plotted with k as abscissa and r_k as ordinate, form the correlogram that provides useful information to identify the type of time series according to the following rules:

Rule 1: If a time series is completely random, then for large N, $r_k = 0$ for every k.

Rule 2: The stationary time series have only short-term correlations.

Rule 3: If successive values of a time series tend to alternate, the correlogram would also tend to alternate.

Rule 4: For stationary time series, when the time series has a trend, the values of rk would not decrease to zero, except for very large values of k. This is not the case for non-stationary time series.

Rule 5: If a time series is characterized by seasonal fluctuations, then the correlogram would also shows oscillations along the same period.

To measure the similarity between two time series the cross-correlation (i.e. the linear correlation coefficient between two time series) can be used, also based on the size of time lag.

Any e-learning tool dedicated to time series modelling and analysis should consider a servlet/CGI module for pre-processing that means a component to transform the data to be suitable as input for some algorithm. As an example, sometimes is necessary to remove the trend (by subtracting the fitted time series model from the original time series) and smoothing the time series in order to stabilize the variance. For trend removing, also, can be used first-differencing between consecutive values, filtering and trend identification. When stationary time series, for trend studying we can use regression models, moving average models, autoregressive models (AR), autoregressive moving average models (ARMA) and state-space models. In general, is necessary to select an appropriate model based on some criteria. To study adaptive AR and ARMA models the following algorithms can be used: the Recursive Least-Squares algorithm (RLS) and the square-root RLS, the Extended Least-Square algorithm (ELS), the Recursive Maximum Likelihood (RML) algorithm, stochastic gradient algorithms based on Gauss-Newton method, Lattice type algorithms based on projection operators etc. When studying time series in the frequency domain, spectral analysis is necessary to decompose a stationary time series into a sum of components from an adequate function space. Let us mention the very used Fourier transform (including the Fast Fourier Transform) and the spectrum estimation by parametric (Yule-Walker autoregressive method, Burg method), nonparametric (Periodogram and the Welch's method), and subspace methods (based on eigen-processing of the correlation matrix). Filtering is based on low, high or band pass filters. The extension of the cross-correlation method to the frequency domain is called cross-spectrum and it is useful for study the relationship between any two time-series. An applet for computing the cross-spectrum is, therefore, required.

4. **Final remarks and conclusions**

Exercises, case studies, simulations, graphical presentations and animation will be used to create an interactive behaviour in order to improve the capacity of understanding the concepts, methods, algorithms and the relevant fields of applications. Offline created files will be uploaded and processed according to the appropriate procedure. Access to statistical databases will be granted for different analysis, in special those mentioned above.

References

1. Albeanu, G., *On Some Methods for Non-stationary Time Series Analysis: a Java Based Software*, *Proceedings of ICCCC 2006*, Baile Felix – Oradea, Romania, pp. 62-67, 2006.
2. Burtschy, B., http://www.infres.enst.fr/~lobstein/MIR/ZZstatsrech.html, *Projet - Analyse de Données et Statistiques.*
3. Green, L. B., McDaniel S. N., Rowel, G. H., *Online Resources for Non-Statisticians Teaching Statistics*, Journal of Online Learning and Teaching, 1, 2, http://jolt.merlot.org/documents/Vol1_No2_rowell.pdf, 2005.
4. Greene H. W., *Econometric Analysis*, Pearson Education (5th ed), 2003.
5. Hamilton, J., *Time Series Analysis*, Princeton University Press, 1994.
6. Sanchez J, Christou N., Dinov ID., *Guidelines for Teaching and Learning Statistics using the Statistics Online Computational Resources: Taking Computer-Based Learning to the Next Step*, http://wiki.stat.ucla.edu/socr (available, 14 March 2008).
7. Shao J., *Mathematical Statistics*, Springer Verlag, 1999.
8. West W.J., Java Applets for Teaching and Research, http://www.stat.sc.edu/~west/javahtml/