

## STATISTICS FOR THE SOCIAL SCIENCES: A CHALLENGE

Nora Lac Prugent and Liliana Severino

Faculty of Economic Sciences and Statistics, Rosario State University, Argentina

nlacprug@fcecon.unr.edu.ar

*The Creditable Postgraduate Seminar tries to give an answer to the growing demand on training and practice in quantitative analysis for researchers and professionals of the Social Sciences. The pedagogic proposal points the training of theoretical-practice data analysis, linking the acquisition of statistical concepts to the application on reference material that it “makes sense” for the professional field of the Social Sciences. For the practical training the methodological material used is brought by the tutor and the databases are national surveys on conditions of life in urban homes of the Argentina Republic. This integrated approach of social data analysis coming from surveys, tries to enable professionals from this area to use in a critical way official databases that are of public access. In this paper we present some results from our experience of a course for Undergraduates that was given in our University.*

### JUSTIFICATION

The design and development of Creditable Postgraduate Seminar on “*Statistic for the Social Sciences. An integral analysis of social data from surveys*” is interdisciplinary. It combines knowledge in statistics; social sciences and training in statistical software of great use in this area of knowledge; is intended for university graduates trained usage in the thematic of the political, economic sciences, statistics, for accountants and all those that embrace the discipline. Methodological material and databases of the Living Conditions Survey (LCS) of SiEMPRO (Information System Monitoring and Evaluation of Social Programmes) is used for practical training collected in the Republic Argentina.

The different statistical calculations are performed in the Computer Science’s Laboratory available in the Faculty of Economic Sciences and Statistics of the State University of Rosario; workshops are organized with two students per computer. Thus, this Seminar whose first cohort corresponds to the year 2008, integrates:

- the acquisition of basic knowledge in descriptive statistics with an introduction to the inferential statistics and
- the application of that knowledge in the analysis of social survey, through the real working processes training in the statistical programme for social sciences data.

We discussed statistical contents developed in sixteen sessions covering the following concepts: introduction to Statistics, browsing and organization of data, data interpretation, statistical summaries, analysis of categorical data, module tables and relationships between variables.

### OBJECTIVES

The general objective is to show that:

- graduates progress firmly when they are given the opportunity to make and to redo themselves,
- teaching situations based on the reality is a conducive to the troubleshooting development of skills in building and exploration of new knowledge,
- the methodology used holds the student as a protagonist of his learning, larger than the traditional one that puts emphasis on teaching transmitting knowledge.

The specific objectives concern social statistical aspects, familiarization aspects with databases and the use of a computer tool. Some of them are mentioned below.

*In the social statistical aspects:*

- learn the terminology and statistical language for an effective communication with the specialist,
- handle the techniques of the statistics to the correct interpretation of results and fast action on the problem,
- become familiar with the concept of uncertainty, main factor affecting to the process of decision-making,
- develop capacity on the socio-economic analysis situation of the inhabitants of a region or country.

*In the use of databases:*

- handle theoretical methodological documents in the collection of data by surveys: sample, variables dictionary, encoding, interviewer training manual and identify the relevance of the technical documentation to evaluate information sources of data, providing both potential and limitations, for the social knowledge,
- identify the possibilities of comparative analysis between apparently similar data sources; its value to the analysis of trend between moments of measurement and historical series of data,
- promote the use of social official data for the development of social empirical research sources.

*In the field of computer tools:*

- exercise in the management of the program to acquire skills in data analysis,
- perform a learning applied to social databases of public access.

**DICTATION AND EVALUATION MODE**

Seminar dictation is developed in present theoretic and practical classes that combine visualization sessions and analysis of the Survey materials display on the screen, and direct training with practical activities on Computers doing grupal work with two students for computer.

The evaluation is based on semi-present practical which integrate knowledge in the three fields of the course: statistics, use of the software for the analysis of materials and data from sample surveys. The practical consist of the theoretical content applications using the databases provided by the Chair. Wizards have surveys and manuals of the software on Cdrom for training in class and the practical implementation.

Assistance by the Lab availability quota is 30 students, some of them without prior knowledge of Statistics. Required minimum attendance is 75%.

**SAMPLE ACTIVITIES**

Below we present descriptions of two sample activities to provide a better sense of the materials being developed.

The 2001 LCS seeks to become a significant input for the socio-economic analysis in Argentina and a query tool for a variety of researchers, technicians and staff of the social area. The survey was carried out on the basis of a sample of about 26,000 urban housings throughout the country. In a selected housing, households were identified; it follows that housing, households and member's data were sought. Data from six questionnaires resulted from the integration of the proposals and interests of the various social programmes, both of the State Ministry of Social Development as of other dependences of the State. The first two questionnaires search on housing conditions, the physical and social environment, and the demographic make-up of households. The four remaining questionnaires are individual and they belong to the different eateries person's life cycle phases: early childhood, children, youth, adults and older adults. There are three main modules replicated in almost all these questionnaires addressed to persons: Labor, Education and Health. The remaining modules, its specificity, are less extensive and by thematic relevance reasons were not included in all the questionnaires (Objetivos y Metodología, SiEMPRO, 2001).

The LCS collects exploratory questions on common behaviors involving health risk. Accordingly, enquires about smoking to the population's group bigger than 15 years old and on knowledge about AIDS/HIV people of the same age who respond personally the individual questionnaire (autorrespondente people).

Guides that we show below are part of the practical works given to the first cohort of professionals from psychology, education sciences, economics, biochemistry, anthropology and statistics.

We have selected these guides to assist appreciation of the value of the computational tool in speed of response, in a process of construction of the very difficult knowledge given the interdisciplinary proposal. The methodological outline is based on the active participation of the graduates using the discussion resource.

*Sample Activity 1: Data interpretation: descriptive statistics*

Reference population: 15 year-old population and over. Risk factors. Smoking. Total of the country. Variables to consider: smokes or smoked ever, daily quantity of cigarettes smoked. Indicate population, who currently smokes, stopped smoking and that never smoke, quintiles of income *per capita* household and age in the smoking.

- For numeric data used display: the value of the arithmetic mean, median, the third percentile, the first quartile, the third quartile, the range, the sixth percentile, the greatest absolute frequency and the value of the cumulative relative frequency to 80%.
- What are the five summarize measures necessary to build a Boxplot? The considered distributions are symmetrical or asymmetric? If so, indicate how. Do you agree with the graphics that the software builds by defect or could they be improved? If so, how?

*Sample Activity 2: Analysis of categorical data: custom tables*

- First part: build the following table; analyze the coefficients of variation corresponding to the country's total population. Present totals for rows and columns. Remember to generate the syntax of all procedures because it will be used in the second part of this activity.

*Title of table:* Occupational category and coverage of social security based on weekly hours worked of 15 to 64 years old. Total of the country. Absolute frequency.

*Variable in rows:* occupational category: employee or worker of a private company, employee or worker of a public institution, employee of the domestic service, owner or employer, own-account worker without wage, NS/NR. The first three modalities disaggregated for social coverage: registered, not registered, NS/NR.

*Variable in columns:* Worked hours: up to 34 hours, 35 to 45 hours, 46 and more, NS/NR.

- Second part: repeat the table previous to the "pampeana" region. Analyze the coefficients of variation for the region, which suggests? Consider, for example, the modifications that seem pertinent to achieve better estimates.

In order to show the graduate's work we present an example of syntax.

```
CTABLES
/VLABELS VARIABLES=hstrab regnoreg Catocu DISPLAY=DEFAULT
/TABLE Catocu [C][COUNT F40.0] BY hstrab [C] + regnoreg [C]
/SLABELS VISIBLE=NO
/CATEGORIES VARIABLES=hstrab regnoreg Catocu ORDER=A KEY=VALUE
EMPTY=INCLUDE TOTAL=YES POSITION=AFTER
/TITLES
TITLE= 'Cuadro N° 3: Ocupados de 15 a 64 años por categoría ocupacional y cobertura de'+
' la seguridad social según horas semanales trabajadas. ' 'Total país. En'+
' absolutos.'
CAPTION= '* Se estima que para valores inferiores a 50.759 el coeficiente de variación es
superior al 17%.' ' Fuente: Encuesta Condiciones de'+
' Vida (ECV- SIEMPRO, 2001) '.
```

## CONCLUSION

Activities that have been proposed involve topics of the Social Sciences, ensuring that the use of the appropriate software facilitates the work of the graduate based learning construction in their active participation. The computer not only enables the reduction of time in routine calculations and the facilitation of students involved in the real working processes, but also graduates came to develop syntax strategies and to transfer them to their specific problems.

The selection of didactic strategies involving the use of the computer focused on the preparation guides motivators; work in the design of structures for reports of Laboratory of Computer Science and in the preparation of assessment tests of the knowledge of the graduate.

Developed workshop classes detected special skills and/or difficulties for the management of the software, the interpretation of the computational statistical language and to unforeseen responses of the computer. The learning student's attitude made it possible to assess: the capacity of induction of new knowledge of the error, the criteria for the analysis of solutions to problems in real situations, the interpretation and use of notation and the statistical language.

It was possible to appreciate the wealth of the interdisciplinary work in processes of construction of the knowledge, impossible of being carried out without the integration proposed in the Seminar.

Other elements to consider in the first release of the given Seminar can be synthesized in some reached achievements: the students' better disposition for the statistical work, great interest in the proposed problem, faster understanding of some abstract concepts. Due to the interest of the assistants and to a great number of requests for the next year, we will work with the second cohort. We are also planning a new seminar embracing topics of the statistical inference allowing updating and improvement of the future participants.

There is a clear need for professionals of the Social Sciences to master some quantitative tools to collaborate in decision-making and implementation of policy; and there is a growing demand of these specialists for training on statistical knowledge for their investigations or studies.

## REFERENCES

- Agresti, A., & Finlay, B. (1997). *Statistical Methods for the Social Sciences*. New Jersey: Prentice Hall.
- Cleveland, W. (1993). *Visualizing Data*. New Jersey: Hobart Press
- Levin, R., & Rubin, D. (2004). *Estadística para Administración y Economía*. (7<sup>a</sup> Edition). México, Pearson Education.
- Manual SPSS, [Online, a seleccionar].
- Rowntree, D. (1984). *Introducción a la estadística : Un enfoque no matemático*. Ed. Norma.
- SIEMPRO (2000). Encuesta de Desarrollo Social. Condiciones de Vida y acceso a Programas y Servicios Sociales 1997 [CD ROM] Buenos Aires: SiEMPRO Base Usuarios, 2000.
- SIEMPRO (2004). Objetivos y metodología ECV 2001. [On line] Buenos Aires: SIEMPRO, 2004 [citado 10 marzo de 2004]. Online: <http://www.siempro.gov.ar>.
- Webster, A.. (2000). *Estadística aplicada a los negocios y la economía*. (3<sup>a</sup> Edition).Bogotá, McGraw-Hill Interamericana S.A.
- Zeisel, H. (1962). *Dígalos con números*. México: Fondo de Cultura Económica.