

USE OF WEB-BASED PUBLIC DATABASES IN STATISTICS COURSES: EXPERIENCES AND CHALLENGES

Lea Bregar, Irena Ograjenšek, Mojca Bavdaž
University of Ljubljana, Slovenia
lea.bregar@ef.uni-lj.si

During the past decade, national and international organisations have been steadily increasing the number of web-based statistical databases available to general public. While often user-unfriendly (mostly due to poor design and organisation as well as lack of navigation options) in their pioneer years, many of these databases have been gradually transformed into well-managed expansive resources that can greatly enhance the teaching and learning of statistics. A number of illustrative examples are presented in this paper along with discussion of our experiences and identification of future challenges pertaining to their use in statistics courses.

INTRODUCTION

Teaching and learning of statistics can be greatly enhanced by the use of web-based public databases. During the past decade, many of them have been gradually transformed from user-unfriendly compilations of data (characterised by poor design and organisation as well as lack of navigation, search and export options) into well-managed expansive resources, available to general public for free.

Applications of web-based public databases in statistics courses are very heterogeneous and range from in-class use (both to enhance lectures and to facilitate a guided hands-on approach in computer lab seminars) to students' independent exploration (when preparing seminar papers or during a written test).

In this paper, we present and discuss our experiences concerning the above-mentioned applications of web-based public databases in teaching and learning of statistics. Courses on *Economic Statistics* and *Business Statistics*, taught at the *University of Ljubljana's Faculty of Economics* to 2nd year undergraduate students, are used as a framework of discussion.

ECONOMIC AND BUSINESS STATISTICS: BACKGROUND INFORMATION

The course on *Economic Statistics* was introduced into the undergraduate university study programme on *Economics* in the fifties of the previous century. The course on *Business Statistics* became part of the study programme at the *Undergraduate College of Business* about two decades later. Both courses are part of the obligatory curriculum, following the 1st year course on *Introductory Statistics*.

Topics covered in the courses (at a different level of detail and abstraction for university and college students) include statistical units, classifications and registers, selected chapters of demographic statistics, introduction to the theory of index numbers (with emphasis on price, production, wage and productivity indices) as well as introduction to national accounting. One of the principal goals of both courses is to give students enough background knowledge to become informed users of official statistics, capable of carrying out their own research projects based on (or enhanced by) the use of public statistical resources.

During the first few decades of the courses' delivery, students' practical work with statistical data was limited to printed Yugoslav statistical publications (such as *Statistical Yearbook* and selected monthly bulletins). First attempts to use electronic databases in teaching and learning of economic and business statistics started with the application of stand-alone CD-ROMs in the beginning of the nineties. One huge practical benefit of CD-ROM application was of financial and practical nature: copying several hundreds of pages of printed statistical publications became redundant. Computer lab seminar as an ITT-supported form of teaching and learning was introduced at approximately the same time, with the goal of teaching students how to handle electronic data media (Ograjenšek and Bavdaž Kveder, 2003).

The decisive impetus for integration of web-based public databases in both courses was given when we started developing a web-based *Course on European Economic Statistics (CEES)* in 1998 (Bregar *et al.*, 2000; Bregar, 2003). One of the important outcomes of CEES-related

research activities was an inventory of web-based public databases. This inventory has been used in class and as a framework for students' independent data exploration, and continuously updated ever since. A selection of most frequently used links is available on-line [<http://www.ef.unilj.si/enote/sief/ang/povezave.asp>].

CHARACTERISTICS OF WEB-BASED PUBLIC DATABASES

The use of web-based public databases for pedagogic purposes requires an advance examination of their characteristics. The table below summarises key characteristics to be considered.

Table 1: Pedagogically Relevant Characteristics of Web-Based Public Databases

<i>Data Characteristics</i>	<i>Technical Characteristics</i>
– content (general, specific)	– access to data (free, payable)
– geographical coverage (subnational, national, international)	– usability of user interface (transparency of navigation, availability of search options)
– level of data (micro, mezzo, macro)	– data format (html, spreadsheet, pdf, text)
– data series (cross-sectional, longitudinal)	– retrieval features (possibility of variable and other detail selection)
– data accuracy (true values, estimates)	– advanced value-added features (multimedia and interactive applications, applets).
– data frequency and timeliness	
– metadata (availability, clarity and detail)	

The relevant characteristics are categorised into two groups. *Data characteristics* determine which topics may be covered, what types of analyses may be employed, how accurate and up-to-date results may be expected, and what limitations may be encountered from the substantive point of view. *Technical characteristics* determine how easy, fast, attractive and cost-effective will be the inclusion of web-based public databases in pedagogic activities.

Generally, those institutional data providers (be it national or international organisations) which disseminate macro data, usually grant users free access and publish comprehensive metadata. This was not always the case. In Europe, significant changes occurred when the statistical databases of *Eurostat* were reorganised and opened to public in the autumn of 2004. That same year the *Statistical Office of the Republic of Slovenia (SORS)* considerably improved the on-line availability of its statistical publications and started developing an interactive statistical database *SI-STAT*. Still, the interactivity of *Eurostat* and *SORS* databases is limited by the predefined list of variables (probably due to confidentiality issues). The use of these databases also requires a certain degree of user ability to transform the research problem statement into statistical terms (selection of variables or indicators, observation units, classification levels, time periods, as well as mode of graphical and tabular presentation), which is often too demanding for inexperienced users. However, inclusion of databases in various pedagogic activities represents an opportunity to enhance students' skills through repeated use.

On a more positive note, comprehensive metadata in these databases are organised according to the *Special Data Dissemination Standard (SDDS)*. Also, both *Eurostat* and *SORS* provide users with database alternatives: ready-to-print online publications as well as predefined tables and graphs which are easy to find and explore.

As far as the mezzo and micro data are concerned, the following rule applies: the lower the level of data, the higher the probability of databases' commercial nature. This is true for databases such as *Bloomberg*, *Dialog*, *Lexus-Nexus*, or *Slovenian Business Intelligence GV IN*, offering detailed balance sheet data for individual companies as well as possibilities of direct benchmarking with the largest competitors in a given industry or with the industrial average. Some of these databases are available to our students exclusively on campus.

APPLICATION OF WEB-BASED PUBLIC DATABASES: ILLUSTRATIVE EXAMPLES

It was stated in the introduction that applications of web-based public databases in statistics courses are very heterogeneous and support both teaching and learning of statistics. These can take place in-class or in the process of students' independent exploration.

In our experience, the most important criteria to choose a web-based public database to be used for teaching and learning of statistics are the following: topic relevance; data and metadata availability; free access (from home and/or on campus – depending on the mode of application); database availability (zero down-time occasions); user interface features; as well as what we call the charm factor (interactive features such as the dynamic demographic pyramids provided by the *U.S. Census Bureau* or database of names and surnames by regions provided by the *SORS*).

These criteria vary in their importance for different pedagogic activities. In the framework of the courses on *Economic Statistics* and *Business Statistics* we've been using web-based public databases both in-class and to further students' independent exploration and knowledge creation. More details are given in the following paragraphs.

In-Class Use: Enhancing the Lectures

Lectures provide students with the systematic overview of each course topic (e.g., short-term indicators or labour statistics). Typically, the following elements are included: economic relevance, short methodological description, analysis of international comparability, and selected illustrative practical examples.

In order not to disturb the flow of communication between the teacher and a large group of up to 250 students in a lecture, the use of web-based public databases in such a setting should be limited to short practical illustrations to bring home certain theoretical and/or practical points as well as to add dynamics and attractiveness to the lecture. Therefore, the crucial characteristics of the databases to be used include the contents (topics covered) and the advanced value-added features. Selected illustrative examples are listed below.

Population statistics

Data: Population clock, World & Slovenia.

Goal: To discuss the problem of population growth and ageing.

Source: SORS, Population Clock. [http://www.stat.si/eng/preb_ura.asp], 31.10.2005.

Access: Free; on the SORS homepage, 1 click to methodological explanation, 2 clicks for more details.

Data: Complete life table for the population of Slovenia, 2000-2002.

Goal: To present the indicators of mortality and life expectancy and discuss the values for Slovenia.

Source: SORS, Rapid Reports. [<http://www.stat.si/doc/statinf/2004/si-169.pdf>], 17.6.2004.

Access: Free; 3 clicks from the SORS homepage; pdf format.

Data: Access to citizen's personal data in population register.

Goal: To illustrate the data collected in population register (e.g., PIN, tax number, name, date and place of birth, residence, citizenship, civil status, right to vote, forthcoming elections).

Source: Ministry of Internal Affairs. Central Register of Population. [<http://vpogled-crp.gov.si/demo/en/>], 4.4.2003.

Access: Demo due to confidentiality reasons; 2 clicks from the ministry homepage.

Data: First and family names of citizens.

Goal: To illustrate the attractiveness of population register data using elementary statistical indicators: frequencies and ranks of names by birth period and by region.

Source: SORS, First Names and Family Names. [<http://www.stat.si/eng/imena.asp>], 30.9.2004.

Access: Free; 1 click from the SORS homepage.

Inflation

Data: Nominal values in US dollars.

Goal: To illustrate the effects of inflation on nominal values in time.

Source: U.S. Department of Labor, Bureau of Labor Statistics. [<http://www.bls.gov/>], 31.10.2005.

Access: Free, one click from the Bureau homepage.

Data: Inflation rate for any chosen time period between January 1st, 1980 and present.

Goal: To illustrate the effects of hyperinflation.

Source: SORS, The Inflation Rate for A Period. [http://www.stat.si/eng/indikatorji_preracun_inflacija.asp], 31.10.2005.

Access: Free, two clicks from the SORS homepage.

Classifications

Data: Classification registry.

Goal: To compare national and international classifications of economic activities and products, their structures and relationships.

Source:

– SORS, Klasje. [<http://www.stat.si/klasje/klasje.asp>], 31.10.2005.

– Eurostat, RAMON – Eurostat’s Classification Server: Classifications. [http://europa.eu.int/comm/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM], 31.10.2005.

– UN Statistics Division, UN Classifications Registry. [<http://unstats.un.org/unsd/cr/registry/default.asp?Lg=1>], 31.10.2005.

Access: Free; 1 click from the SORS homepage, 2 clicks from the Eurostat homepage, 1 click from the UN Statistics Division homepage.

In-Class Use: Facilitating a Guided Hands-On Approach in Computer Lab Seminars

Four academic hours during the semester are reserved for the computer lab seminar with the goals of providing students with an overview of the most useful web-based public databases; showing students how to use these databases (guiding them in their first attempts at exploration and helping them out when they continue working on their own); and providing the platform for discussions about international comparability of the most important socio-economic indicators. In this case, the data and technical characteristics of the database seem to be equally important.

Following are selected illustrative examples.

Population statistics

Data: Demographic data for countries at different developmental levels, 1950-2050.

Goal: To compare demographic indicators in time and across countries and discuss the consequences of different demographic situations.

Source: U.S. Census Bureau, International Data Base: Online Access.

[<http://www.census.gov/ipc/www/idbprint.html>], 26.4.2005.

Access: Free; 4 clicks from Census homepage.

Price indices

Data: Latest indicators of inflation rate.

Goal: To enhance understanding of price indices, to locate and interpret inflation rates in various forms (monthly, yearly).

Source: SORS, Inflation in October. [<http://www.stat.si/>], 31.10.2005.

Access: Free; on SORS homepage, 1 click to more detail (time series also available in a pop-up window).

Data: Metadata on price indices for consumer prices and producer prices.

Goal: Compare metodological data on the compilation and dissemination of price indices across countries.

Source: IMF, Dissemination Standards Bulletin Board: Special Data Dissemination Standard Site.

[<http://dsbb.imf.org/Applications/web/sddshome/>], 31.10.2005.

Access: Free; 4 clicks from IMF homepage.

Students' Independent Exploration: Preparation of Seminar Papers

Seminar papers in our courses are designed according to principles of constructivism and active learning (Garfield, 1995). Working on the seminar papers, students experience all stages of the empirical research process: the initial definition of the research problem; search for, and selection of, appropriate secondary data; selection and application of relevant statistical tools; and, finally, interpretation and communication of results in both written and oral form. In this setting, quality aspects of data used in the analysis are given special emphasis.

Students are provided with short instructions about the objectives, structure and main building blocks of the seminar paper. They are also provided with the list of recommended web-based public databases (the already mentioned inventory at <http://www.ef.uni-lj.si/enote/sief/ang/povezave.asp>) and given some practical advice how to approach the secondary data search and evaluation (in addition to the information received in the framework of the computer lab seminar). In this case, the availability of metadata is essential while the data format is important for comfortable data management. Selected sample research topics are listed below.

Sample research topics

- Topics included in the curriculum: comparative analysis of labour market development during the last decade (for Slovenia and another country or for two Slovenian regions); analysis of current economic trends in a selected country on the basis of short-term statistics; comparative analysis of population trends for Slovenia and another country.
- Topics not included in the curriculum yet closely connected to the course (analysis of poverty; structural characteristics of international trade, trends in tourism).
- New or generally interesting topics (time-use surveys; structural indicators of Eurostat; sustainable development indicators; indicators of information society; environmental indicators).
- Topics dealing with evaluation of web-based public databases of selected institutional data providers such as WTO, UNESCO, World Economic Forum, IMF, etc.
- Topics of students' choice.

Students' Independent Exploration: Written Test

Written test based on the exploration of web-based public databases has been introduced in the academic year 2001/2002 for the DE students of *Business Statistics* who traditionally work in small groups (30-50 students) which are relatively easy to manage.

Open test questions are structured in such a way that students need to identify the relevant data source and find (or calculate) the proper indicator(s) and provide its (their) correct interpretation. The database availability is essential due to time constraints. Selected sample questions are listed below.

Sample questions on population statistics

- What is the life expectancy at birth for men and women in Japan in 2004? Why the difference between the genders?
- What is the infant mortality rate for Slovenia and China in 2004? Provide the formula, interpret both indicators and explain the difference in their values.
- For a given country, find the data on population size for the last two censuses. Using the growth rate method estimate the population size for 31.12.2010.

Sample questions on price indices (for Slovenia)

- What was the change of producer prices in 2004?
- What inflation rate have we accumulated this year?
- What was the level of retail prices in 1990 compared to 2000?
- What are the methodological elements relevant for the compilation of the CPI in Slovenia?

Sample questions on unemployment

- Compare two unemployment rates published by the SORS and explain the differences in the underlying concepts.
- What was the registered unemployment rate in Slovenia in the last decade?
- Compare the unemployment rates in the new EU member states.

CONCLUSIONS

Use of the web-based public databases has been one of the cornerstones of the undergraduate courses on *Economic Statistics* and *Business Statistics* at the *University of Ljubljana's Faculty of Economics* for over a decade. Rapid technological advances (especially with regard to navigation, search and extraction options) and the improved accessibility enabled full integration of these databases into the teaching process; a development reported by statistical educators from all around the world.

Given the more than satisfactory level of technological development, the true challenge for statistical educators and students (future decision-makers) now lies elsewhere: in the development of their capacity to evaluate and exploit *the contents* hidden in the accumulated wealth of statistical data. Both simple spreadsheets and advanced statistical software packages make it very easy for users to calculate statistical parameters for a given dataset without understanding the substance of data. This is a problem that should be systematically tackled by joint efforts of academia and official statistics.

REFERENCES

- Bregar, L. (2003). Teaching economic statistics in internet era. In *Statistics Education and the Internet, Proceedings of the IASE Satellite Conference on Statistics Education*, Berlin, August 11-12, CD-ROM. Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education.
- Bregar, L., Ograjenšek, I. and Bavdaž, M. (2000). Teaching economic statistics in a digital environment. In A. Ferligoj and A. Mrvar (Eds.), *New Approaches in Applied Statistics* (pp. 237-249). Ljubljana, Slovenia: Faculty of Social Sciences.
- Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63, 1, 25-34.
- Ograjenšek, I. and Bavdaž Kveder, M. (2003). Student acceptance of itt-supported teaching and internal course administration: Case of business statistics. In *Statistics Education and the Internet, Proceedings of the IASE Satellite Conference on Statistics Education*, Berlin, August 11-12, CD-ROM. Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education.
- SORS, *Statistical Yearbook: Prices*. (2004). http://www.stat.si/letopis/2004/15_04/15-01-04.xls.