



STATISTICS NEW ZEALAND'S DATA LABORATORY

Sandra McDonald

Project Manager, Data Laboratory

Abstract

Statistical microdata is an essential input to social science and economic research. Demand for microdata has been growing with increased emphasis on policy integration and targeting. As well, technological and methodological advances mean that researchers are increasingly able to undertake microdata analysis. With the introduction of the Data Laboratory, Statistics New Zealand's microdata is more accessible for significant research projects. This presentation will look at eligibility for using the Data Laboratory, the process that researchers will need to follow and some of the conditions that are imposed.

Keywords: *microdata, data laboratory*

Statistics New Zealand's Data Laboratory has been operating for about 18 months. This presentation outlines how the Data Laboratory operates and provides some examples of how researchers have used it in the past.

Traditional methods of accessing statistical data

- reports, information releases, maps
- standard tables
- customised tables
- INFOS, Supermap - access to common databases

All these are using aggregated data and in some cases they are not suitable for the purpose.

Background

- Statistics New Zealand wants more use made of existing datasets collected at considerable expense
- increasing demand for more complex analysis to support policy development and evaluation
- increased computing power and better desktop software allows researchers to use large and complex datasets

Access to microdata

- is access to unidentified data at an individual level, rather than aggregated data

Advantages of microdata

- flexibility to recode, derive variables and subset data
- identify outliers for analysis and remove
- order/sort data for derivation of new variables, graphing
- initial analysis on small sample to establish relationships, then tables on whole sample
- multivariate analysis

Access is enabled by the Statistics Act

- access is not automatic, researchers need to make an application
- must be for *bona fide* research or statistical purposes
- names and addresses are deleted
- researchers are required to make a statutory declaration of secrecy
- tight security to protect information
- published results can not divulge more than the Statistician could publish

Data Laboratory service

- liaison with researcher
- facilitates assessment of proposals
- organises datasets, in accordance with confidentiality requirements
- provides secure facilities for access to data, including records derived from matched datasets
- set up by supplementary funding on the basis that costs are recovered from users

Some examples of previous Data Laboratory projects are included:

Health Services Research Centre

- 1991 census data, updated with 1996 data
- developed an index of deprivation
- proportions of 10 variables within meshblocks weighted to form an index for each area
- index used by wide range of organisations for:
 - health needs assessment
 - asthma research
 - maternity services
 - resource allocation

Department of Labour

- 1981, 1986 and 1996 census data
- assessed immigrants' labour market experiences
- used multivariate techniques to compare the labour market performance of immigrants with native New Zealanders (STATA)

Massey University

- income data from several Household Economic Surveys
- analysed household income distribution over period of economic reforms
- prepared various disaggregated Gini co-efficients
- published as a research paper
- SNZ to include results in a report on Incomes

Wellington School of Medicine

- 1991 census data and NZHIS mortality data
- association of socio-economic factors and mortality
- SNZ linked mortality and census data
- using SAS to assess the relationship of socio-economic factors (income, education, occupational class, housing tenure, car access) with mortality
- if successful planning to extend to other years

Actions taken to protect confidentiality

- deletion of identifying data items
- top coding
- grouping
- removal of rare data records
- random perturbation

What to do if interested

- contact Data Laboratory manager and develop proposal
- approval - Government Statistician gets advice on:
 - validity of the proposed use of the data
 - eligibility of researchers
 - any analytical or confidentiality issues
- contractual arrangement entered into setting out:
 - who has access
 - what outputs are to be produced
 - security obligations
 - price

Author

Sandra McDonald is Project Manager, Data Laboratory,
Statistics New Zealand,
PO Box 2922, Wellington,
E-mail: sandra_mcdonald@stats.govt.nz

USER'S COMMENTARY ON SNZ DATA LAB

Panel Discussion

Three users of the SNZ Data Lab made a number of points based on their own individual experience. **Paul Callister**, currently undertaking doctoral research at Victoria University of Wellington, undertook research on changing work and family patterns using 1986, 1991 and 1996 census data.

Paul argued for as rapid access to the lab as possible after data were collected by SNZ. This speed was particularly important in the case of the quinquennial census of population and dwellings. Once access was gained then there was considerable value to the researcher in interacting with the staff in the lab itself as well as with other specialists within Statistics New Zealand.

Paul was concerned that the high cost of access to the lab might not be recognised by funding agencies. He also made the point that researchers need to have a very good prior understanding of the databases to be used.

Sylvia Dixon, a senior researcher with the Department of Labour used the lab for research projects on the changing distribution of earnings, labour force participation patterns and low wage employment. The access arrangement which the Department of Labour and Statistics came to enabled Sylvia to use the data in the Department of Labour's own offices. This access arrangement allowed two projects to be addressed from the same data set, namely the examination of earnings distribution and factors associated with labour supply. Like Paul, Sylvia pointed to the real benefits in having access to specialist programming skills of SNZ during the data analysis phase as well as being able to call on their resident experts who were experts on the data themselves.

Rainer Winkelmann, a university based researcher, in joint work with **Liliana Winkelmann**, drew selected data from the 1981, 1986 and 1991 censuses in order to study the labour market outcomes of Maori and non-Maori. In a more recent project, they evaluated the relative labour market performance of New Zealand's immigrants, using the 1996 census as an additional observation point.

Rainer found that micro level census data were easy to use, and that there was excellent support from staff. However, he pointed out that low-cost access to micro-data, available in most OECD countries, is essential for applied economic research and policy evaluation. While the data lab was a step in the right direction, further developments, such as the distribution of anonymized sub-samples of census data via CD-Rom, were necessary in order to stimulate more applied research on New Zealand issues and reap the full benefits of these data sources for New Zealand society.

The following points were made in general discussion:

1. The somewhat rigid time frame and prescribed nature of what has to be done did not sit easily with the essentially iterative nature of research itself and the wish to following interesting leads which may not have been in the original proposal. The present constraint to tightly specify what was required from the outset tended to favour deductive over inductive research: the former was oriented to the testing of specific hypotheses whereas the latter required greater opportunity to explore and follow leads. There was a general wish that contractual arrangements to use the lab also allow a revisit at a later date once feedback had been received about the initial research.
2. There is quite a time taken for researchers to negotiate with SNZ on access to the lab although it was recognised that both parties were well aware of the other's requirements.
3. The complexities of the data sets and the technical issues associated with retrieval meant that there was a steep learning curve facing researchers with commensurate claims on their time. This was compounded if researchers lacked expertise in programming. There is therefore a certain risk which researchers take in using the lab which they should be aware of.
4. The ability of researchers to bring in their own software was appreciated and this helped cut down programming time.
5. It was felt that SNZ should go further than at present and encourage the appointment of internal researchers so that for example a university staff member could work for a certain period within SNZ thus gaining access to the lab on a more informal basis but SNZ could still have all the major confidentiality safeguards in place.
6. SNZ could go further in freeing up researcher's access to unit record data and reducing the costs of that access. It was pointed out that in the USA anonymized copies of unit record data collected by the Bureau of the Census are available to researchers on an unrestricted basis and at minimal cost.

Roberta Hill kindly chaired the sessions and took notes. Each of the panelists above had the opportunity of amending the text written from these notes.