Australian Bureau of Statistics

METHODOLOGICAL NEWS

A QUARTERLY INFORMATION BULLETIN FROM THE METHODOLOGY AND DATA MANAGEMENT DIVISION

June 2011

Quality Management of the Acquisition of Administrative Data

National statistical agencies need to find new ways to reduce respondent burden, maximise available data usage and find efficiencies in the use of taxpayers' money. This has led to the use of data that is collected for purposes other than that of a statistical nature. This type of data is often obtained from records or transactional data from government agencies, businesses or non-profit organisations which use it for administrative purposes. This data is referred to as administrative data.

Administrative data can offer strategic and statistically important advantages when used as part of a statistical process, such as:

- reducing respondent burden which may improve response rates;
- decreasing the costs associated with the collection and processing of statistical data; and
- the ability to impute missing or invalid data based on existing administrative data.

However, there are a number of issues associated with the use of administrative data for statistical purposes. Some of which include:

- limitations associated with the use or supply of administrative data records due to legislative requirements;
- limited ability of a receiving agency to influence the collection of the administrative data (e.g. definitional changes, question wording etc.); and
- incompatibility of technological systems in the sending and receiving of administrative data records.

As part of the Australian Bureau of Statistics commitment to providing leadership in quality management of statistical processes, an information paper, '*Quality Management of Statistical Outputs Produced from Administrative Data (cat. no. 1522.0)*', was released on Thursday, 23rd March 2011. The paper provides information on managing the quality of statistics produced from administrative data. It focuses on some principles and best practices to assist in the management of the acquisition of administrative data. The paper is a continuation of the ABS quality management series which provides guidance on the frameworks and best practice in the use of statistics. Other papers in the ABS quality management series can be used to manage the quality of statistics produced from administrative data, including:

- *ABS Data Quality Framework, May 2009, (cat. no. 1520.0)*;
- *Quality Management of Statistical Risk Using Quality Gates (cat. no. 1540.0)*.

Further information on the quality management strategies discussed in the paper can be obtained from Narrisa Gilbert on (08) 9360 5283 or by email at narrisa.gilbert@abs.gov.au and Andrew Doherty on (03) 9615 7038 or by email at andrew.doherty@abs.gov.au.

National Cancer Institute (NCI) Adjustment Method for Nutrition Data

The ABS has recently commenced collecting detailed food consumption data from the Australian population as part of the new expanded Australian Health Survey (AHS), something not attempted since the National Nutrition Survey (NNS) ran in 1995. Randomly selected participants initially undergo a face to face interview referred to as a "24 hour recall" where they are assisted in comprehensively detailing their consumption of solids and liquids over the previous day.

A key issue is how these 24 hour recalls should be used to draw conclusions regarding the population's *long-term* consumption of foods and nutrients, which is the primary concern of government policy makers in food and nutrition standards.

A general solution, and the one employed by the 1995 NNS, is to collect a second 24 hour recall from participants at a later date so that the component of error due to day to day variation (within-person measurement error) can be modelled across the population and removed to leave the long-term intake distribution as desired.

Methods following this principle are relatively straight-forward for nutrient consumption where data is always non-zero and often well approximated by a normal distribution. However, modelling long-term consumption of certain foods encounters formidable challenges such as high numbers of zero entries (many foods are not consumed on a particular day) and right-skewed consumption patterns (many foods have a smaller number of high-consumers).

Given the lack of developed methods at the time of the 1995 NNS and the small number of second day 24 hour recalls to work with (these were only collected for 10% of the sample), the ABS only published long-term intake distributions for selected nutrients, not foods. Hence quality long-term food intake distributions have long been a significant information gap filled only by food-frequency data with its intractable recall and concept bias issues.

Since this time, methods to overcome such challenges have built on each other to culminate in a comprehensive solution now referred to as the "NCI method" in recognition of the National Cancer Institute in the United States, where the most recent innovation to the method was developed.

Use of the NCI method for the Nutrition component of AHS 2011-12 was proposed by the Household Survey Methodology (HSM) section and justified shedding of the food frequency component of the survey freeing up valuable interview time. HSM has since conducted a trial of the method on the "Kids eat, Kids play" nutrition and physical activity dataset for children collected by CSIRO in 2007 to further understand the method and anticipate issues surrounding its application in practice within the ABS environment in particular.

For further details please contact Ross Watmuff at ross.watmuff@abs.gov.au or on (02) 6252 7084

AWE/Salary Sacrifice Series

The Average Weekly Earnings (AWE) survey will be releasing for the first time in August 2011, coinciding with the May 2011 publication, the cash series spreadsheets, which is earnings that includes salary sacrifice. The PDF publication will continue to publish just the earnings data which by definition excludes salary sacrifice. Salary sacrifice earnings are gross earnings used in Australian Taxation Office (ATO) approved schemes such as novated leases on cars. The advantage to the employee is, for example, not having to pay Goods and Services Tax (GST) on a new car and allows the employee to reduce their taxable income. In the lead up to this, the methodology division was heavily involved in assisting the survey area in quality assuring

the salary sacrifice earnings data. This involved selecting a sub-sample of AWE providers for the survey area to work through potential reporting issues. As the AWE definition of earnings excludes those that were salary sacrificed, the quality assurance programme allowed the survey area to minimise mis-reporting by providers.

Further work was carried out by the methodology division in assessing the suitability of the current imputation methods used for the earnings and employment data items, in imputing salary sacrifice data items given their tendency to be more volatile. These imputation methods involve using the previously reported value with a growth factor applied where the historical data is available. Where the historical data is not available a ratio factor based on benchmark employment is applied to the unit requiring imputation. The analysis found that the existing imputation methods were suitable and have now been implemented.

For further details please contact Martin Caruso on (08) 9360 5302 or martin.caruso@abs.gov.au

Producing State Level Estimates of Hours Worked

The ABS publishes estimates of total hours worked by employed Australians, which complement the employment rate as a measure of economic activity and employment in Australia. They are also used as a labour input for the productivity estimates in the Australian System of National Accounts. Recently, there has been increased interest in using the these data to measure the impact of, and the recovery from, the global financial crisis and its effect on the Australian economy. In meeting user demands for finer level estimates, the ABS has improved methods of accounting for holiday effects in these time series and has published estimates at the state and territory level. These flow estimates of state level hours worked have been first published in the January 2011 Labour Force publication (cat no. 6202.0).

Accounting for holiday effects is involves a number of stages. In the context of actual hours worked data, holiday correction is a process of removing the effects of regular non-random events and is performed prior to adjusting the data for monthly seasonality. A holiday corrected series is one without known holiday impacts that is intended to reflect the hours worked in a working week unaffected by holidays. The overall process involves taking holiday corrected stock series of hours worked (derived from the ABS Monthly Population Survey) to produce monthly flow estimates of working weeks unaffected by holidays by using a linear interpolation methodology. Holiday effects are then reintroduced into these estimates to give monthly flow estimates of actual hours worked.

The effect of state specific holidays are better estimated using state level series, and such estimates were introduced earlier this year. Likewise, the effects of one-off events can be seen more clearly in these series when the impact of the event is localised geographically. For example, the estimates of hours worked for Queensland last January were noticeably affected by the wide-spread flooding occurring at that time.

The methodology for estimating and accounting for moving holidays in these series was presented at the Australian Statistical Conference in December 2010. For more information, please contact Philip Crouch at philip.crouch@abs.gov.au.

Time Series Analysis Training

"How do I interpret a time series?" "Should I be looking at the seasonally adjusted or trend movements?" "How do I quality assure the seasonal adjustment of my time series?"

If you are asking (or are being asked) questions like these, you may be looking for some training in time series analysis. The Methodology and Data Management Division (MDMD) considers the statistical training of ABS staff to be a key priority, enabling ABS staff to make informed decisions about their collections, and pass on this knowledge to data users.

The Time Series Analysis (TSA) section in the Analytical Services Branch (ASB) provides both formal and informal time series analysis training within the ABS, as well as responding to technical queries from external data users. Our training program includes our suite of formal training courses, which progressively provide greater technical detail about the time series analysis process.

Time Series Made Simple (TSMS)

This half-day course provides participants with a basic understanding of time series analysis, including an understanding of the components of a time series, problems and issues which may arise during seasonal adjustment, and interpretation of time series estimates. This course is ideal for new starters in areas that produce seasonally adjusted time series, or for anyone interested in increasing their understanding of time series.

Understanding Time Series (UTS)

This two day course covers the same material as TSMS in more technical detail, as well as discussing seasonal adjustment 'ABS style', how to resolve issues with time series data, and using the SEASABS (SEASonal analysis, ABS standards) software for data investigations. This course is primarily intended for staff in areas that produce seasonally adjusted time series, but will also be of value to staff who regularly liaise with external data users, or anyone who wants to expand their data analysis capabilities.

Producing Time Series (PTS)

This two day course provides participants with an understanding of periodic time series processing within the ABS, the responsibilities of the subject matter areas and TSA in producing and quality assuring time series estimates, and how to undertake quality assurance of time series output. It includes many exercises using the SEASABS Suite for data analysis. This course is intended for staff involved in the production of ABS time series estimates, and who have attended UTS.

The courses described above have been designed for an ABS audience, but have been presented to staff from external agencies on occasion.

If you are interested in attending time series analysis training, or want more information about our courses or to discuss the time series training needs of your area, please contact Tessa Nunan on (02) 6252 6382 or theresa.nunan@abs.gov.au

How to Contact Us and Subscriber Emailing List

The Methodological Newsletter features articles and developments in relation to methodology work done within the ABS Methodology and Data Management Division. By its nature, the work of the Division brings it into contact with virtually every other area of the ABS. Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

If you would like to be placed on our electronic mailing list, please contact:

Valentin Valdez
Methodology & Data Management Division
Australian Bureau of Statistics
Locked Bag No. 10
BELCONNEN ACT 2617
Tel: (02) 6252 7037
Email: methodology@abs.gov.au