Methodological News

A Quarterly Information Bulletin by ABS Methodology Division

DECEMBER 2018 EDITION

Contents

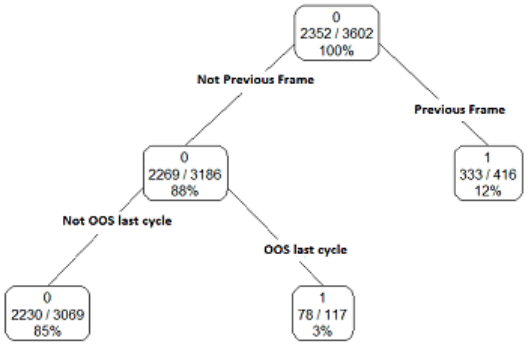
Determining Out of Scope Responses for the Research and Experimental Development Business Survey

The ABS produces statistics about business expenditure on research and development (R&D) via the Research and Experimental Development Business Survey (RDBS). RDBS is a sample survey that is conducted biennially, with approximately 4,000 businesses sampled.

Businesses are grouped into strata according to Industry division and then size. For the purposes of the RDBS, the size of a business is determined by its expected R&D expenditure, and this may come from several sources, for example, previous cycle's frame or from auxiliary data sources. Units that report zero R&D expenditure are deemed as Out of Scope (OOS) and are subsequently cancelled. Reducing the number of OOS units contacted will result in reduced provider burden as well as realised benefits for the ABS through improved sample design.

A project was undertaken to develop a method to scope out likely OOS responders from the frame whilst keeping undercoverage low. Data from 2015/16 was used to devise the methods whilst 2013/14 data was used for testing. Data from 2015/16 was seen as the most up-to-date hence why we used these data to form the methods. A variety of different methods were investigated as part of this project. A decision was made to pursue the use of decision tree models given the relative ease of implementation.

Based on model covariates, the decision tree algorithm attempts to split the data into groups that are more or less likely to report OOS. This can be used to create a set of scoping out rules for the frame. The model covariates included size, final expenditure stream (which source was used as the expected R&D expenditure on the frame), ANZSIC division, and OOSLastCycle (taking a value of 1 if unit reported OOS last cycle, and 0 otherwise). Below is a visual representation of the decision tree:



We found that units that sourced their expected expenditure from the previous frame were more likely to report OOS than those that didn't use that particular expenditure source. Furthermore, units that didn't use the previous frame as their expenditure source were further split with those that reported zero last cycle being more likely to report OOS.

The effectiveness of the scoping rules to reduce the number of OOS units was examined using both the 2015/16 and 2013/14 data. There were two sets of scoping rules tested, one that used the full

decision tree and another that only used the first/main branch.

Main points to note:

1) The use of the full tree resulted in relatively more in-scope units removed from the frame as compared to the main branch.

2) The estimated undercoverage from the main branch was rather small and this approximately doubled when the full tree was used.

For the upcoming cycle of RDBS, we have partially applied the rules from the main branch by scoping out units that have expected R&D expenditure sourced from the 2011/12 frame. The effectiveness of the scoping exercise will be evaluated after the cycle. Future work is also planned to assess scoping out units that reported 0 in the previous cycle.

References

For more information:
Tom Davidson
Tom.Davidson @abs.gov.au

Multi-stage Conditional Selection for Household Surveys

The ABS has recently moved to the use of an address register for selecting samples for household surveys. This has required some changes to the way household survey samples are selected and managed.

Previously, broad geographic areas (First Stage Units or FSUs) were selected every five years, using Karmel or Keyfitz selection to avoid overlap with broad regions from previous selections. Overlap and rotation were managed within these areas by a fairly manual process of ordering Base Frame Units or BFUs (small geographic areas) and clusters of dwellings, and field staff counting through them. With the address register we will be selecting dwellings directly, although to facilitate the hiring of local interviewers and minimise interviewer travel we will still be doing this within the framework of BFUs and FSUs. This means that the manual approach for overlap and rotation control is no longer viable.

Conditional selection is a new method for controlling overlap and rotation for sample surveys that allows the usage of individual selection units to be tracked over time. It avoids selecting units that are undesirable (perhaps because they have been selected before) by assigning them a low 'conditional' probability of selection. This is done in such a way as to guarantee the sample is selected without bias and maintaining the desired probability of selection overall. Conditional selection is a very powerful and flexible method for overlap and rotation control. In particular, it imposes no limits on survey

stratum boundaries or on how probabilities of selection are calculated. However, in its basic form conditional selection only applies to a single stage of selection. The ABS has been working on extending conditional selection to apply to multi-stage selection processes, and in particular the multi-stage approach used in ABS household surveys.

In order to extend conditional selection to two or more stages it is necessary to reflect the conditional selection probabilities of earlier stages of selection in the second or subsequent stages. This is relatively straightforward. Logistically, it is generally preferable for sample enumeration to continue sampling in an area for as long as there are dwellings available to select. As such, it is also necessary to introduce usage measures. The usage measures count sample usage within areas and prompt conditional selection to make an area undesirable for further selection once it no longer has enough sample.

However, a difficulty arises in ensuring that areas get a fair chance of selection in later samples. Because we stay in an area until it is fully used, there is a possibility that the sample will get 'trapped' in larger BFUs and never move on to smaller BFUs, effectively altering their probability of selection. To avoid this we introduce a new generalisation of conditional selection called Multiple Outcome Conditional Selection. This approach allows the sample the possibility of moving out of a BFU and returning later, thus allowing all BFUs a fair chance of selection at any point.

Another issue with multi-stage selections taken over a long period of time is that the geographical areas involved change as the population grows and evolves. We have developed approaches for splitting and/or combining selection histories and recalculating usage measures when geographic boundaries are redrawn.

The first ABS surveys to have their samples selected using conditional selection were the Labour Force Survey, the Survey of Disability and Carers and the National Aboriginal and Torres Strait Islander Health Survey. These began going into the field in July 2018.

For more information:

Claire Clarke
Claire.Clarke@abs.gov.au

How to contact us and Email Subscriber List

Methodological News features articles and developments in relation to methodology work done within the ABS Methodology 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 added to or removed from our electronic mailing list, please contact:

Nick Husek
Methodology Division
Australian Bureau of Statistics
Locked Bag No. 10
BELCONNEN ACT 2617

Email: methodology@abs.gov.au

The ABS Privacy Policy outlines how the ABS will handle any personal information that you provide to us.