Automatically protecting user-defined tables and analytic outputs from the Australian Population Census

Victoria Leaver, Dr James Chipperfield and Melissa Gare

Australian Bureau of Statistics, ABS House, Belconnen, ACT 2614, Australia, victoria.leaver@abs.gov.au, james.chipperfield@abs.gov.au and m.gare@abs.gov.au

Abstract: The Australian Bureau of Statistics (ABS) has developed a method for automatically protecting tables of Census counts. This method has been implemented in a web-based table building system that allows users to define and request their own tables. This method protects against requests for similar tables and repeated requests for identical tables. The approach is currently being extended to key outputs from continuous data items (totals, means, median and quantiles) and weighted estimates from surveys. The ABS is also developing an Analysis Service that will allow a range of user-defined analytical techniques to be applied to Census and other data.

1 Introduction

The Australian Bureau of Statistics (ABS) has been given the authority by legislation to collect statistical information. This legislation requires that any data collected under this authority shall not be released in a manner that is likely to enable the identification of a particular person or organisation.

The ABS has developed a range of policies to assist it in meeting these legislative requirements. For the release of aggregate statistics in tables these policies include specific rules around the suppression of table cells with very small counts and consequential actions to ensure that other data released cannot be used to derive the true value of the suppressed cells. These policies have been developed through consideration of the range of different disclosure risks that may be present when releasing data. A number of these risks are highlighted in Section 2 of this paper.

The ABS conducts the Australian Census of Population and Housing every five years, with the most recent Census held in 2011. The data from this Census allow detailed analysis of small sub-populations and are a valuable source of information that is not available elsewhere. However, it is important that any release of Census data conforms with the legislation and ABS policy. There are a number of different ways that users can access Census data including publication tables, user requested tables (special data service) and through the analysis of confidentialised microdata files referred to as confidentialised unit record files (CURFs). Two CURFs are available – a Basic CURF consisting of a 1% sample of records accessed via CD-ROM and an Expanded CURF based on a 5% sample of records accessible through the ABS online Remote Access Data laboratory (RADL).

Internationally, there are a range of methods used for statistical disclosure control of Census data, including small cell rounding, limits on table sparsity, data swapping and replacing some
values with imputed or synthesised data. Prior to 2006, the ABS used a combination of small cell rounding and limits on table sparsity to protect Census data.

For the 2006 Census of Population and Housing, the ABS developed a perturbation method for automatically protecting tables of Census count data. This method was designed to protect against requests for similar tables, repeated requests for identical tables, and repeated requests for the same table cell within different tables. The method was intended to allow greater access to data for subpopulations, and to enable the development of web-based systems allowing users to define their own tables. The methodology was originally presented in Fraser and Wooton (2005) and is summarised in Section 3. All tabular output from the 2006 Census is protected using the same method, including tables created by ABS staff for publications. For this reason, there are internal systems for applying the method as well as web-based products available to users outside the ABS: CDATA Online and the Census TableBuilder. These products were jointly developed by the ABS and Space-Time Research Pty Ltd.

At the ABS Statistical Disclosure Control research and development is currently focused on expanding the perturbation approach to survey data, to key tabular outputs from continuous data (such as totals, means and medians), and to analytic outputs such as regression parameters. An overview of these current and future directions are provided in Section 4.

2 Overview of potential disclosure risks associated with Census data

A range of different types of disclosure are discussed in the literature. These types of disclosure include:

- **Identity disclosure**: a record in the dataset is associated with the individual unit it describes;
- **Attribute disclosure**: the value of a sensitive variable is disclosed directly;
- **Inferential disclosure**: the value can be inferred with high confidence from the statistical characteristics of the entire database;
- **Predictive disclosure**: A particular type of inferential disclosure where the value of a response variable is disclosed with an unacceptably degree of certainty;
- **Group disclosure**: Membership of a group, for example as defined by cross-classifications, discloses information about an individual; and
- **Key identifiers**: variables which can be used to match to other datasets - for example, age, ethnicity or sex.

In tabular outputs, the most common disclosure risks are:

- releasing cells with too few contributors;
- releasing cells that are dominated by a small number of observations (continuous data items);
- releasing rows or columns in a table that are dominated by one cell (predictive disclosure);
- releasing tables that are different by only a small number of units – this may reveal information about a small sub-population or geographical area (differencing attack); and
- repeatedly releasing tables confidentialised using a random process, which could then be averaged to reveal the original values (averaging attack).
The risk of disclosure increases when:

- the underlying data is based on a census of units as opposed to a sample;
- there are other outputs from the microdata, such as publication tables or CURFs;
- there are other auxiliary datasets accessible to users (such as administrative data);
- the underlying data is free from error; and
- the data is very recent.

There are also risks associated with releasing outputs from analysis. Analysis output includes estimated regression coefficients and their associated estimated standard errors, diagnostic plots and test statistics. Releasing these through a remote server without masking the output or imposing restrictions on the models that may be fitted is an unacceptable disclosure risk. There are some well documented data attacks (see for example, Gomatam 2008) that must be protected against.

Attacks can aim to reconstruct the values of one target record or all records in the data set. It is even possible for disclosure to occur for records not on the micro-data. Under a model inferential disclosure arises if the model has highly accurate predictions and the covariates for the target record are known.

Another risk arises from linking data that are collected from two or more agencies. The ABS is establishing itself in Australia to play the role of a data custodian that is trusted to link microdata, collected by two or more agencies, and to release, via a remote server, tabular and model output with an acceptable disclosure risk. There are a number of projects already underway to link data from the 2011 Population Census with other datasets such as past Population Census data and other data including administrative datasets held by other organisations. One disclosure risk unique to this arrangement is that one agency may use the microdata it supplied to the data custodian and the released tabular or model output from the linked dataset, to disclose information about a person or organisation that was collected by the other agency.

3 Census TableBuilder

3.1 Perturbation method

For a Statistical Disclosure Control method to be useful in a web-based product the following design requirements need to be achieved.

1. The method must address the general confidentiality risks associated with tabular output, and specific risks associated with releasing flexible user-defined tables. These specific risks are the differencing risk and the possibility of averaging repeated requests, as described in Section 2.
2. The method must be flexible. That is, users should be able to specify tables using a wide range of data items. The standard geographical structure also has many levels. Ideally, users would be allowed to specify tables at different levels and create their own geographical areas from combinations of standard geographical units.
3. The method must provide tables of high utility. That is, the method should not change the data so much that the tables are significantly less useful for analytical purposes, and ideally the tables should provide a coherent picture of Australia's population.

4. The method must produce tables that are consistent and independent of the requester. That is, the method should produce the same confidentialised output each time the same table is requested, regardless of the user.

5. The method must be quick to execute. For a web-based application the confidentialisation needs to be completed in a reasonable amount of time, so that users are provided with an interactive experience.

The method described in Fraser & Wooton (2005) was developed with these requirements in mind. It attaches an element of a finite group to each unit record. These elements are then combined and mapped to indices that are used to derive the perturbation that will be applied to each cell.

In the ABS's implementation of the method, the elements are permanent numeric values known as “record keys”. When a table is created, the record keys are combined using modular arithmetic to produce a cell-level key. This cell-level key is used to determine the perturbation that is applied to the cell, via a fixed look-up table. Zero cells have no contributing records and do not receive any perturbation. Interior cells and marginal totals are perturbed independently, so that the perturbed tables are not necessarily additive.

This method ensures that each cell receives the same perturbation whenever it appears in any table. Repeated requests for the same table will therefore produce the same results. Also, if the same cell appears in different tables, it will be perturbed in the same way each time, protecting against attempts to use varying results to determine the original value.

The look-up table contains the value of the perturbation that will be applied to the cell, based on the original cell value and the cell-level key. The cell-level key determines the row of the look-up table, and the cell value determines the column of the look-up table. For larger cell values, modular arithmetic is used to determine the column. The distribution of perturbation values within the look-up table determines the range of the perturbation that can be applied and the average magnitude of perturbation that is applied to non-zero cells.

With the exception of zero cells, every cell in each table has a chance of being perturbed. The maximum amount of perturbation that any cell can receive is fixed, which means that larger cell values will receive a proportionally smaller amount of perturbation.

Since all cells have a chance of being perturbed, the difference between any two tables will also be perturbed. This prevents small cells values being rederived by differencing. If small cells can be calculated by differencing, these cells will receive, on average, a proportionally large amount of perturbation. The method does not guarantee that the difference will be different from the true value, but the variance around the true value gives sufficient uncertainty to protect individuals from being identified.
The method does not depend on the user nor the data items or geographical areas that users choose for their tables. While the method can be applied to different levels of data, it does not guarantee exact consistency across these levels. For example, the count of dwellings and the count of persons in the same geographic area may be perturbed in different directions.

3.2 Specification of perturbation parameters

There are a number of parameters that need to be specified in implementing the method. In particular:

1. the distribution used to generate the record keys;
2. the best way to combine the record keys to create the cell-level keys; and
3. the distribution of the perturbation values held in the look-up table;

The choices for some of these parameters have a significant impact on the disclosure risk and utility loss. In particular, the distribution of perturbation values in the look-up table determine the amount of protection applied to the tables, and thus the amount of utility loss. The properties of these distributions were chosen to satisfy the requirements from the ABS's legislation and policy, while minimising the utility loss. As described in Fraser and Wooton (2005), in the ABS implementation, the keys and the look-up table were designed to produce integer-value perturbations $e_i^*$ whose distribution satisfies the following criteria:

a) $E_i(e_i^*) = 0$

b) $Var_i(e_i^*) = \sigma^2$

c) $Cov_i(e_i^*, e_j^*) = 0 \text{ if } i \neq j$

d) Whenever the same set of records contribute to a cell count, the value for $e_i$ will always be the same

e) $e_i^*$ is an integer

where $Var(e_\cdot)$ and $E_i(e_\cdot)$ are the variance and expectation with respect to the perturbation distribution of $e_i^*$. The perturbation look-up table is constructed by choosing a distribution of perturbation values which minimises disclosure risk subject to information loss constraints. More specifically, entropy (a measure of uncertainty) is maximised subject to bias and variance constraints. Marley and Leaver (2011) demonstrate the flexibility of the method through demonstrating the risk and utility loss for three alternative perturbation distributions:

1. A maximum perturbation of +/- 1 to the unconfidentialised cell counts;
2. A maximum perturbation of +/- 5 to the unconfidentialised cell counts and no confidentialised cell values between 1 and 4 inclusive; and
3. A maximum perturbation of +/- 20, with the constraint that no cell values are unchanged, and that all confidentialised cell values are multiples of 5.
Analyses outlined in the paper using various measures of disclosure risk and utility loss demonstrated that perturbation distribution 1 provides very little utility loss, but a large amount of identification risk. Conversely, perturbation distribution 3 provides a significant amount of utility loss with little identification risk.

An option that was considered for Census TableBuilder was to assign the record keys in a way that preserves the value of certain estimates exactly. This option has not been implemented in Census TableBuilder, but could be explored in the future.

3.3 Restoring additivity to tables

One side-effect of the perturbation method is that following perturbation tables are not additive. That is, the sum of the interior cells may not add to the marginal cells of the table as a result of each cell being perturbed independently. To overcome this the ABS developed a second routine, referred to as the Additivity Module, that restores additivity to the table so that the tables would make more sense to users and have greater utility.

The Additivity Module does not share all the properties of the perturbation method. For example, if the same cell appears in two different tables, the cell will receive the same perturbation, however it may be altered in different ways by the additivity module. This is not considered to be a significant confidentiality risk.

The additional changes introduced by the additivity module are generally small in comparison to those applied by the perturbation method. It is usually only on very sparse tables that the additivity module has a significant impact.

It may be possible for a user to undo the changes applied by the additivity module, for example by averaging cell values over a large number of table requests. However, the user will only be able to recover the underlying perturbed table, not the original data. Also, zero cells may be altered by the additivity module for a very small number of unusual tables. This is not necessarily ideal, especially if there are cells that logically should not contain any contributors. However, it was necessary to allow the module to alter some zero cells, particularly in sparse tables, to meet the other constraints.

3.4 Additivity Module

The Additivity Module uses iterative methods to restore additivity, under various constraints. These constraints specify that:

1. the resulting table will contain non-negative integer values
2. very small non-zero cell values will not appear in the additive table
3. the perturbed grand total of the table will be preserved
4. the changes to cell values will be minimised.

Denote the internal and marginal cells of a table by \( t = 1, 2, ..., C, C+1, ..., T \), where \( t = 1, 2, ..., C \) denotes the internal cells of the table. Denote the \( t \) th cell count by \( n_t \). Instead of releasing \( n_t \),
Census TableBuilder releases \( n_i^* = n_i + e_i^* + a_i^* \), which is obtained in two steps. The first step involves calculating the preliminary counts \( m_i = n_i + e_i^* \), where \( e_i^* \) has properties a) to e) as listed above. The table’s preliminary counts are not consistent: sums of preliminary counts for internal cells are not guaranteed to equal corresponding preliminary marginal counts. The second step involves finding the value for \( a_i^* \) so that the table with counts \( n_i^* \) is consistent and \( |a_i^*| \leq L_a \) for all \( l = 1, \ldots, L \). This means no preliminary count, for a marginal or internal cell, is changed by more than \( L_a \).

3.5 Web-based systems for generating user-requested tables from the Population Census

The perturbation method and the Additivity Module have been incorporated into the SuperSTAR suite of products developed by Space-Time Research Pty Ltd. Internally, ABS publication areas use SuperCROSS to produce publication estimates, while the SuperWEB product underpins the web-based systems for users outside the ABS: CDATA Online and the Census TableBuilder.

CDATA Online is available free of charge to all users with access to the Internet and one of a range of supported web browsers. CDATA Online contains a large collection of underlying topic-based data cubes, from which users can create customised tables, maps and graphs. While the data cubes are used to define the tables that users can request, the data cubes are not used directly in the confidentialisation process. That is, the tables are not confidentialised by aggregating confidentialised counts in the data cubes. Instead, each table created using CDATA Online is confidentialised using the standard perturbation method from the underlying microdata. Once users have requested tables, maps or graphs, these outputs can then be exported or downloaded. CDATA Online can be accessed from www.abs.gov.au/CDataOnline. This webpage also contains links to the user manual.

Census TableBuilder is similar to CDATA Online, but allows more flexibility. Instead of containing underlying topic-based data cubes, Census TableBuilder allows users to create tables directly from the full range of Census data items. This means that users can include most combinations of Census data items in their tables. As with CDATA Online, users can create maps and graphs as well as tables. Users of Census TableBuilder need to register before they are allowed to use the product. There is more information about Census TableBuilder, and a user manual, at www.abs.gov.au/TableBuilder.
4 Future research directions

Current research and development at the ABS is focused on developing a new Remote Execution Environment for microdata (REEM). REEM has two components – a table builder for survey data, Survey TableBuilder, and an Analysis Service.

4.1 Survey TableBuilder

The Census TableBuilder method has been extended for a new system called Survey Table Builder (STB). In the first stage of STB, weights have been incorporated into the method so that users can request tables of weighted counts from sampled survey data.

Relative Standard Errors (RSEs) are also supplied with the weighted count estimates. Stage 2 of STB extends the method further, to confidentialise estimated totals and means calculated from continuous variables. The confidentialisation is applied by perturbing the values of the continuous variables for the $n$ largest contributors in each cell. Confidentialised quantile boundaries and estimates will also be available in Stage 2 of STB.

Ensuring the counts are consistent after perturbation follows an analogous approach to Census Table Builder. Possible future extensions for Survey TableBuilder include methods for time series data, linked and longitudinal data, administrative data and business data.

4.2 Analysis Service

4.2.1 A model for a remote server

A simple model for a remote analysis server is:

A. An analyst submits a query, via the internet, to the agency’s analysis server

B. The analysis server processes the analyst’s query on the sensitive micro-data. The statistical output (e.g. regression coefficients) from the query is modified for the purpose of protecting confidentiality. Some output may be restricted on the basis that it could allow an analyst to reconstruct the attributes of an arbitrary record.

C. The analysis server sends the modified output, via the internet, to the analyst.

Some advantages of a remote analysis server are:

- although the statistical output is modified, it is based on the real microdata. This means complex relationships in the microdata are essentially retained.
- the degree to which a particular output is modified can depend upon the output itself. For example, estimates at a broad level may require proportionally less modification than estimates at a fine, or small area, level. Since an analyst is restricted from viewing the attributes of any record, less modification is needed than would otherwise be the case.
- the impact of the modifications on the output can be broadly indicated to the analyst. If the impact is large the analyst may decide to ignore the results altogether.
- once the server is set up, it can process multiple analyses in real time.
• all submitted programs can be logged and audited. If an audit concludes an attempt at disclosure was made, the agency can revoke the analyst's access to the server and take legal action.

There are some disadvantages of a remote analysis server:

• Some statistical outputs may be aggregated (e.g. record-level residual plots may be replaced with box plots) or perturbed (e.g. regression coefficients), and others may be restricted altogether.
• The analyst may be restricted to use only analysis techniques supported by the server.

Analysis through a remote server may take longer than if the microdata were available on the analyst’s personal computer.

4.2.2 An example of defending against attacks on an analysis server: regression coefficients

Ways of protecting the underlying values on the microdata are well documented (e.g. Gomatam et. al 2008). These include imposing restrictions on models that may be fitted (e.g. restrictions on creation of new variables, sub-setting records before model fitting and transformations). This may also include aggregating or smoothing graphical output which could otherwise reveal data values (e.g. residual plots or QQ plots). Some level of “tracking” or “monitoring” of queries may also be useful to identify attacking analysts. Finally, perturbation of outputs is also a useful protective measure. Below, we summarise the ABS approach to perturbing regression coefficients. Further details of the approach are described in Chipperfield and Yu (2011).

Consider fitting a regression model with parameter $\beta$ using an unbiased estimating function (see Chambers and Skinner, 2003)

$$H(\beta) = \sum_{i=1}^{n} G_j(\beta) \left( y_j - f_j(\beta) \right),$$

where $f_j(\beta) = E\left(y_j | x_j\right)$ and $G_j(\beta)$ is a vector of order $K$ with $k$ th element $G_k(\hat{\beta})$ which is a function of $\beta$ and $x_j$ but not of $y_j$. The solution to $H(\beta) = 0$ gives the standard estimate, $\hat{\beta}$, of the regression coefficients. Instead of solving $H(\beta) = 0$ and releasing $\hat{\beta}$, an alternative is to solve

$$H(\beta) = E^*$$

and releases the resulting estimator $\hat{\beta}^*$, where $E^* = (E^*_1, E^*_2, \ldots, E^*_K)'$ are perturbations introduced for the purpose of SDC, $E^*_k = u^*_k e_k$, $u^*_k$ is the uniform distribution on the range (-1,1), and $e_k = \max_j \left\{ G_j(\hat{\beta}) \left( y_j - f_j(\hat{\beta}) \right) \right\}$ is the maximum influence a record may have on the $k$ th estimating equation. The distribution of the perturbations, $E^*_k$, are independent and if the same
model is fitted the same value of $E^*$ is used—this stops an analyst estimating $\hat{\beta}$ by fitting the same model a number of times and averaging over the regression parameters obtained from solving (1).

This perturbation makes it much more difficult for an attacker to solve for the underlying data values on the microdata.

5. **Summary**

The Web-based user applications developed by the ABS for the 2006 Census of Population and Housing have been a great success, and will again form a key component of the dissemination strategy for results from the 2011 Census.

The underlying Census TableBuilder perturbation methodology is flexible. It can be adapted to provide different levels of perturbation and utility loss through alteration of the underlying perturbation look-up table. This enables the method to be tailored to the confidentiality requirements of different organisations interested in utilising the method.

**References**


