WP. 35
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Geneva, Switzerland, 9-11 November 2005)

Topic (v): Confidentiality aspects of tabular data, frequency tables, etc.

A PROPOSED METHOD FOR CONFIDENTIALISING TABULAR OUTPUT TO PROTECT AGAINST DIFFERENCING

Supporting Paper

Submitted by the Australian Bureau of Statistics [1]

[1] Prepared by Bruce Fraser (bruce.fraser@abs.gov.au) and Janice Wooten (Janice.wooton@abs.gov.au), Data Access and Confidentiality Methodology Unit.

A proposed method for confidentialising tabular output to protect against differencing

Bruce Fraser* and Janice Wooton*

* Data Access and Confidentiality Methodology Unit, Australian Bureau of Statistics, Locked Bag 10, Belconnen ACT 2616 Australia, bruce.fraser@abs.gov.au, janice.wooton@abs.gov.au

Abstract: The differencing problem puts increased demands on a system of tabular confidentiality. Methods currently in use at the Australian Bureau of Statistics and many other national statistical offices only target small cell values for treatment, and allow large cells values to be released without any perturbation to protect confidentiality. Such methods are vulnerable to differencing attacks which can derive unprotected small cell values as the difference of two unprotected large cell values. This paper proposes a cell perturbation method for confidentialising Australian population Census tables to protect against differencing attacks and any other attempts at identification. The method is a two stage process. At the first stage a perturbation is added to all cells of all tables, including the independent perturbation of table marginals. The perturbation is set to zero for a pre-determined set of key output (e.g. age by sex population counts). This perturbation process produces a non-additive protected table. At the second stage additivity is restored. Record keys are assigned to the microdata and are used to produce consistent perturbations at the first stage of the process, although consistency is lost when additivity is restored.

1 Introduction

The Australian *Census and Statistics Act, 1905* provides the authority for the Australian Bureau of Statistics (ABS) to collect statistical information, and requires that statistical output shall not be published or disseminated in a manner that is likely to enable the identification of a particular person or organisation. This requirement means that the ABS must take care with any statistical information that relates to very small subpopulations or subsamples.

Output from Australia's 5-yearly Population Census is extensively used for studying small subpopulations in Australia. As a complete enumeration of people in Australia on census night, it is one of the few statistical datasets in Australia that can be used to compile meaningful statistics for small subpopulations. It is therefore important that rigorous procedures and techniques are in place to ensure that Population Census output is released in a manner that is not likely to enable identification of any individual, household or family.

The current technique used to guard against identification or disclosure of confidential information in census tables of counts is random rounding to base 3 for cells with small values. However with the 2006 Population Census this method alone is no longer adequate to protect census tables against disclosure. This is due to a number of factors, in particular the introduction of a new small area geographical unit - the mesh block, and a proposed web-based table-builder product which would allow users to produce tailored tables according to their own specifications. The small cell perturbations produced by the random rounding base 3 method are not effective as the sole protection for ad hoc output where a user can specify tailored tables using fine level geography or fine disaggregations of other variables. A further problem is that it is becoming increasingly difficult to keep track of all the Population Census tables released across different mechanisms, and so difficult to protect against differencing problems by tracking the release of output.

An example of a differencing problem is where a user specifies a table for a user-defined geography, compiled from a number of small area building blocks. For example, the Statistical Local Area (SLA) "Remainder of ACT" had a population of approximately 430 at the time of the 2001 Population Census. The SLA consists of 7 Collection Districts (CDs) with populations of approximately 210, 115, 65, 25, 10 and two with populations of less than 5 each. If a user can specify tables for a tailored geography made up from CD building blocks, then they can specify a table for the full SLA, as well as a table for the amalgam of the six CDs with the greatest populations. Differencing the two tables provides information for a single CD with a population of less than 5 persons. The confidentiality protections applied to the SLA table, and the 6 CDs table, must therefore be sufficient to ensure that no information is disclosed through differencing the two tables. For further details regarding the definition of CD's and SLA's see ABS Cat. no. 1216.0 *Australian Standard Geographical Classification (ASGC)- Electronic publication* , July 2004.

The differencing problem is not limited to geography. User-specified tables could be requested for the whole population of "Remainder of ACT", and for the population aged less than 70 years, or for the population born in English-speaking countries, or for the population who only speak English at home. Differencing would again produce results in respect of small subpopulations.

To solve the differencing problem, we propose an alternative to the current census tabular confidentialisation methodology. The new methodology will be discussed in the following sections and is a cell perturbation technique. This technique has advantages in that it provides protection against disclosure through differencing and disclosure through single contributor cells. Like the current method, the new method will ensure small cells are perturbed to protect against disclosure through single contributor cells. But unlike the existing method a small amount of perturbation is introduced to all cells instead of just small cells. This ensures that when two large cells are differenced to produce a small difference enough perturbation has been introduced so that users cannot have much confidence in the accuracy of the differenced value.

The method has been developed in the context of Population Census tables, and has only been designed to protect tables of non-negative integer counts.

2 Deriving a New Cell Perturbation Method

By perturbing all cells instead of just small cells, we can protect census tables against disclosure through differencing.

Denote the i^{th} cell count of a multi-way table as n_i. For each non-zero n_i an independent perturbation d_i is generated from an integer-value distribution that satisfies the following criteria:

(a) mean of zero;

(b) $d_i \geq -n_i$;

(c) fixed variance V for all i and all n_i; and

(d) $|d_i| \leq c$ for some small positive integer c.

d_i is added to n_i to give n_i^*, the cell value for the protected table.

Criterion (a) ensures that the perturbations do not add a bias to the table, criterion (b) ensures that no negative numbers are created as a result of perturbation, criterion (c) ensures that any cell derived by differencing two perturbed cells has a fixed variance, and also that relatively more noise is added to the smallest cells (smallest n_i), and criterion (d) is applied to ensure that no perturbation is ever greater than c in magnitude.

Note that no perturbation is added to zero cells ($n_i = 0$) in order to maintain any structural zeroes. The method we propose also independently perturbs every non-zero cell in a table, in particular this means table margins are perturbed independently from interior cells and so table additivity is lost. An alternative approach would be to perturb interior cells only, and then generate marginals by adding interior cells. However this approach would result in a final perturbation of marginal cells with a high level of variance, and indeed a variance that increases as the number of interior cells increases. In order to prevent large perturbations on the margins we instead perturb them independently of interior cells to create a non-additive table, then restore table additivity as a separate step.

3 Improving consistency by assigning random numbers or record keys to microdata records

A technique for improving consistency between tables is to assign permanent random numbers to each record on the microdata file, and to use these permanent random numbers to generate the random perturbations. In the following we discuss two possible ways of doing this.

The first way is to assign each record a key in the form of a 32-bit binary number (the keys are assigned randomly to each observation on the census microdata file). This record key can be used as a seed for a pseudo-random number generating function, which in turn could be used for generating d_i at record level. Record keys can also be combined across records to guarantee consistent results are applied to aggregates of records. This can be done using the XOR (exclusive or) function. The XOR function will return another 32-bit binary number, and will always return the same result from the input, regardless of the order in which the individual keys are XORed together. This means any aggregate of n records will correspond to a unique 32-bit aggregate key, obtained by XORing the keys of the individual records. The key of the aggregate can be used to seed a pseudo-random number, which in turn can be used to determine the aggregate's perturbed value. This gives the property that whenever the same set of units are together in an interior cell, the perturbed cell value will always be the same.

The second method is to assign an independent random discrete uniform number to each unit in the microdata file. The interval for the discrete uniform random variables will be from 0, 1, 2, ...m-1, where m is a sufficiently large integer value. Let y_i denote the discrete uniform random variable assigned to the j^{th} unit on the microdata file. It can be proved that the sum mod m of any combination of the y_i values is also a discrete uniform random variable on the interval 0, 1, 2, ..., m-1. Whenever a set of units are present together in an internal cell, we can combine their discrete uniform random numbers using the mod function described above and use the result to decide on the perturbation to be applied to the cell. This guarantees that whenever the same units contribute to an interior cell, the perturbation will always be the same.

4 Allowing for zero perturbation counts

The assignment of record keys or random numbers to microdata records discussed in section 3 can be modified slightly to allow for a set of predefined counts for which no perturbation error will be introduced. For example, age by sex population counts at a particular level of geography can be defined as zero perturbation counts. (In this example the level of geography chosen must be broad enough to ensure that it is not vulnerable to differencing).

This can be done by assigning record keys or random numbers independently to all but one of the records contributing to each zero perturbation count. The final record in each group is assigned the particular record key or random number that is required to ensure that the aggregation of the record keys or random numbers of all records in the group (through mod m addition, XOR, or some other method of aggregation), gives a final result of zero. The perturbation process is then constrained to ensure that a result of zero will always produce a perturbation of zero.

5 Restoring additivity to perturbed tables

There are a number of ways in which additivity might be restored to the non-additive table that results from the independent perturbation of interior and marginal cells. A key criterion for an additivity algorithm is that it can process large tables quickly. This is necessary in order to be able to offer a responsive web-based table-builder service, whereby a user can specify a tailored table, and have it constructed, confidentialised and delivered to the user via the web interface in a short period of time. A secondary criterion is to only make small adjustments to the table marginals, with larger adjustments in the interior cells, if necessary.

Work is currently in progress to determine a suitable algorithm for restoring additivity to perturbed tables.

6 Balancing information loss and disclosure risk

The Australian Bureau of Statistics is required by law not to release information in a manner that is likely to enable the identification of any respondent. Furthermore, it is essential to retaining the trust of providers that they are confident their personal information will be safeguarded. A tabular confidentiality system therefore must reduce the disclosure risk to a low level. But this should be done in a way that minimises the information loss in the data, and preserves as far as possible the analytical value and integrity of the results.

The method we have outlined has flexibility, primarily in the distribution chosen for the cell perturbation, and to a lesser extent in the algorithm chosen to restore additivity to a table. These characteristics can be varied to change both the amount of protection provided (reduction in disclosure risk) and the amount of damage done to the output (amount and characteristics of information loss). Work is still in progress assessing disclosure risk and information loss resulting from different choices of perturbation distributions. While the intention is to measure information loss and disclosure risk using a number of metrics, there are two key measures being used in the evaluation.

A measure of disclosure risk is the probability that an observed count of 1 corresponds to a true count of 1. It is a simple matter to specify a perturbation distribution that will ensure any cell value of 1 in a table will be perturbed to a value other than 1, however counts of 1 can also be observed through the differencing of two table. Therefore one measure that is used in the evaluation is the probability that an observed difference of 1 equates to a true difference of 1.

The primary measure of information loss is provided by a comparison of the results of a chi-squared test of association on each of the protected and unprotected tables.

References

Australian Bureau of Statistics, (2004). *Australian Standard Geographic Classification (ASGC): Electronic Publication* . Cat. no. 1216.0.