This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Statistical Disclosure Control Methods for Quantitative Tables

Contents

General section

1. Summary

This module is about the protection of quantitative tables. Such tables are typically used to release data on business statistics. There are other forms that are sometimes used (such as microdata, frequency tables), but they are not dealt with here. More in particular we shall focus on a single quantitative table together with its marginals. The general case of linked tables (of which the one with hierarchical tables is a special case) is not treated here. A discussion of this case can be found in the literature. References will be provided.

The main issues with protecting quantitative tables are the identification of the unsafe cells in such tables, and how to protect them. Both issues will be addressed here. The one about actually protecting tables is ultimately rather technical, amounting to the solution of often complicated optimisation problems. How this is done is described in the literature, and references will be provided. We concentrate in this module on two techniques: table restructuring and cell suppression.

2. General description

2.1 Tables of magnitude data

Quantitative tables are tables in which the cell values are composed by summation of a continuous variable over all the contributors to a cell. This is in contrast to frequency tables in which only the *number* of contributors per cell is given. Other rules apply to frequency tables, and other protection methods may be more suitable than those for quantitative tables. In Section 2.2 there is more on frequency tables.

If exactly one or two contributors produce a cell total, it is clear that this cell cannot be published. In the case of a single contributor, individual information is released directly, and in the case of two contributors, one contributor can exactly calculate the other contribution by subtracting his or her own contribution from the cell total.

However, undesirable situations can arise also if there are more than two contributors in a cell. In principle, in the statistical disclosure control of quantitative tables, we must prevent (or at least make it more difficult) that any contribution can be estimated too accurately. This may occur, for example, also in the case that a very large contributor is present in a single cell along with several relatively small contributors. In this case, the second-largest contributor can calculate that the largest contribution does not contribute more than the cell total minus the second-largest contribution to the cell. A relatively good estimation of the contribution of the largest contributor can be obtained as a result, in conflict with the disclosure control rules of any NSI.

The presence of empty cells also requires extra attention. In some cases, an empty cell will be a so-called *structural zero cell*. This means that it is generally known that, logically, it is *impossible* for this cell to have a contribution. Such cells can therefore also not be used in the disclosure control: whatever you do, everyone knows that they must be empty cells.

At the same time, reliable information can sometimes be disclosed using *non-structural zero cells*. If there are contributors in such a cell, there is actually a sort of group disclosure: it is immediately clear that all the contributors to that cell have provided a contribution of zero (assuming that the contributions are non-negative). If there are no contributors in the cell, but it is not logically impossible for a contributor to be in this cell, this in itself also reveals direct information.

2.2 Tables of frequency count data

Frequency tables are tables in which the number of contributors per cell is given. This is in contrast to quantitative tables in which the cell values are created by summation of a continuous variable over all the contributors to a cell. Other rules apply to quantitative tables, and other protection methods may be more suitable than those for frequency tables.

Frequency tables require the protection of recognisable data about statistical units. A violation of statistical confidentiality (a disclosure) may be two-fold: *identity disclosure*, i.e., disclosing the presence of an individual respondent in the table, and *attribute disclosure*, i.e., disclosing additional information about a single respondent. Some statistical laws do not allow identity disclosure on its own, while other statistical laws only care about attribute disclosure (for which identity disclosure is a necessary precondition).

For frequency tables, attribute disclosure can be formulated as follows. The user must first recognise a contributor or group of contributors in the table. This is followed by a statement about these contributor(s) due to the frequency distribution over the cells. The statement that the table makes possible about this group must provide more information about the members of the group than just the group size. In this sense, knowledge that is needed to recognise the members of the group can be considered not to be disclosive information about the members of the group. However, some statistical laws do not allow for disclosing this kind of information nonetheless.

The requirement is satisfied if the table does not provide any information about an individual statistical unit as such. However, the table should not provide information about groups of statistical units that can be identified (*group disclosure*). In particular, that is the case if the table contains variables that could provide harmful or potentially damaging information about these groups, like whether or not an environmental crime has been committed. Such data will be referred to as sensitive data.

2.3 Sensitive cells

The usual approach in SDC for tabular data is to identify the sensitive, or risky cells in a table. These are the ones that need to be protected. Various sensitivity measures are available that can be used for this task. All these measures need to be parameterised.

In Table 1 an overview of some well-known sensitivity rules is given. For a more detailed description, see Hundepool et al. (2012).

The first three sensitivity rules are so-called ‘concentration rules’. For concentration rules it should be borne in mind that in order to apply them, one needs to have information about individual contributions to the various cell values. In particular, one needs to know the *n* largest contributions to each cell.

In case of magnitude tables, often a combination of a concentration rule and a threshold rule is used to determine the sensitive cells. However, the concentration rules imply a certain threshold by definition.

Table 1. Various sensitivity rules

Sensitivity rule	**Type of table**	**Cell is unsafe if**
(n,k) rule / dominance rule	Magnitude	The n largest contributions to that cell make up for more than k% of the cell total.
(p,q) rule / ambiguity rule / prior posterior rule	Magnitude	Some contributor to that cell is able to derive an estimate of some other contributor to the same cell within p% of the true value, a-priori knowing all the other contributions within q% of their true values.
p % rule	Magnitude	Some contributor to that cell is able to derive an estimate of some other contributor to the same cell within p% of its true value.
Threshold rule	Frequency and Magnitude	The number of contributors is less than a prespecified threshold.

From a methodological point of view, the p% rule is preferred. Moreover, note that a concentration rule implies a certain threshold rule. E.g., under the p% rule a cell will always be unsafe when there are less than 3 contributors to that cell.

2.4 Table protection measures

To protect tabular data several methods are being employed in practice. In Table 2 some of the more important techniques for protecting tables (both magnitude and frequency tables) have been assembled. For detailed descriptions of these methods, we refer to Hundepool et al. (2012).

Table 2. Various SDC methods for tabular data

SDC Method	**Type of table**	**Type of method**	**Short description**
Barnardisation	Frequency	Perturbative	Randomly add/subtract 1 from some cell values.
Table redesign / table restructuring	Magnitude or frequency	Nonperturbative	Collapsing rows and/or columns.
Cell suppression	Magnitude or frequency	Nonperturbative	Completely suppress the value of some cells (put a “cross’).
Rounding • Controlled • Conventional / deterministic • Random	Magnitude or frequency	Perturbative	Round each cell value to a prespecified rounding base.
Controlled Tabular Adjustment (CTA)	Magnitude	Perturbative	Selectively adjust cell values: unsafe cells are replaced by either of their closest safe values. Other cell values are adjusted to restore additivity.
Perturbation / adding noise	Magnitude	Perturbative	Add random noise to cell values.

As "Type of method" a two-fold classification is used: Perturbative or Nonperturbative. Whenever a method is of type Perturbative, this means that certain cell values will be replaced by adjusted cell values, i.e., they will be *perturbed*.

The most commonly used methods are table redesign and cell suppression. CTA is a promising recent technique, but is not used that often in Europe yet.

2.4.1 Table restructuring

In general, cells with a limited number of contributors or a cell with one or two large contributors are the obvious candidates to be characterised as risky. All risky cells must be protected. Before performing suppression on a large scale, restructuring the table can also be considered. By combining rows and/or columns, cells are pooled and the content per cell is increased. The result of this is that fewer cells are identified as risky by a sensitivity rule (such as the *p* % rule).

This method will generally lead to fewer risky cells in the table. Combining an unsafe cell with one or more safe cells may result in a cell that is safe.

There are no methodological conditions for using this method. However, externally imposed obligations sometimes specify what level of detail a table must have when published. This may be a Eurostat obligation. Or an NSI may have a publication policy requesting a certain level of detail for a table when published. So although restructuring could be applied successfully, publication policy might prevent this.

Furthermore, an assessment must be made between the information loss resulting from the larger number of suppressed cells that are needed to protect the table, and the information loss resulting from combining columns/rows, for which fewer crosses are needed.

The software package τ-ARGUS has provisions for recoding rows and/or columns in tables. Two situations are distinguished:

- In the case of a hierarchical spanning variable, the recoding implies that certain splits are omitted at the lowest level.
- In the case of an unstructured spanning variable, users are free to combine the columns or rows of a table as they choose.

Example. Figure 1 presents a fictitious table of turnover according to Region (hierarchical) and SizeClass. The crosses in Figure 1 are cells that are unsafe (or risky) according to some sensitivity rule. Figure 2 and Figure 3 provide two restructuring possibilities for this table.

In Figure 2 the variable SizeClass is recoded such that the categories 2 to 6 are combined into the category MediumSmall, and that the categories 7, 8 and 9 are combined into the category Large. Note that, in this way, all the risky cells are combined to create safe cells. In Figure 3 the recoding of the variable Region is such that the smallest detail level has been removed. This restructuring does not resolve all the problems: the risky cells at region level (for North and East) are still present in the table. This is not necessarily a problem. If the protector is satisfied with the structure of this table, he may decide to eliminate the remaining sensitive cells by, e.g., cell suppression. ■

	tot	2	4	5	6	7	8	9	99
tot	16,847,646.84	20.00	25.00	2,711,808.00	2,320,534.00	2,505,042.58	2,799,074.26	6,510,758.00	385.00
- .North	4,373,664.00	×	×	719,049.00	659,680.00	688,962.00	756,529.00	1,549,049.00	385.00
.. 1	1,986,129.00	×	×	398,062.00	348,039.00	354,711.00	418,778.00	466,529.00	-
.. 2	1,809,246.00	0.00	-	223,990.00	221,332.00	241,913.00	258,233.00	863,393.00	385.00
.. 3	578,289.00	-	-	96,997.00	90,309.00	92,338.00	79,518.00	219,127.00	-
- .East	3,703,896.00	15.00	×	642,238.00	515,003.00	534,147.00	620,392.00	1,392,096.00	-
.. 4	124,336.00	×	-	36,311.00	32,132.00	25,770.00	18,150.00	×	-
.. 5	526,279.00	-	-	93,589.00	94,957.00	110,930.00	81,799.00	145,004.00	-
.. 6	2,234,995.00	×	×	345,803.00	251,358.00	251,188.00	303,377.00	1,083,254.00	-
.. 7	818,286.00	-	-	166,535.00	136,556.00	146,259.00	217,066.00	151,870.00	-
- .West	4,576,115.84	-	-	648,972.00	543,570.00	663,896.58	775,132.26	1,944,545.00	-
.. 8	485,326.00	-	-	63,767.00	75,442.00	87,305.00	59,953.00	198,859.00	-
.. 9	3,664,559.84	-	-	537,911.00	430,851.00	515,019.58	643,762.26	1,537,016.00	-
..10	426,230.00	-	-	47,294.00	37,277.00	61,572.00	71,417.00	208,670.00	-
- .South	4,193,971.00	-	15.00	701,549.00	602,281.00	618,037.00	647,021.00	1,625,068.00	-
..11	2,752,743.00	-	15.00	488,613.00	392,395.00	363,490.00	402,925.00	1,105,305.00	-
..12	1,441,228.00	-	-	212,936.00	209,886.00	254,547.00	244,096.00	519,763.00	-
.99	-	-	-	-	-	-	-	-	-

Figure 1. Quantitative table for turnover according to region and size class

	tot	Large	SmallMedium	99
tot	16,847,646.84	11,814,874.84	5,032,387.00	385.00
- .North	4,373,664.00	2,994,540.00	1,378,739.00	385.00
.. 1	1,986,129.00	1,240,018.00	746,111.00	-
.. 2	1,809,246.00	1,363,539.00	445,322.00	385.00
.. 3	578,289.00	390,983.00	187,306.00	-
- .East	3,703,896.00	2,546,635.00	1,157,261.00	-
.. 4	124,336.00	55,888.00	68,448.00	-
.. 5	526,279.00	337,733.00	188,546.00	-
.. 6	2,234,995.00	1,637,819.00	597,176.00	-
.. 7	818,286.00	515,195.00	303,091.00	-
- .West	4,576,115.84	3,383,573.84	1,192,542.00	-
.. 8	485,326.00	346,117.00	139,209.00	-
.. 9	3,664,559.84	2,695,797.84	968,762.00	-
..10	426,230.00	341,659.00	84,571.00	-
- .South	4,193,971.00	2,890,126.00	1,303,845.00	-
..11	2,752,743.00	1,871,720.00	881,023.00	-
..12	1,441,228.00	1,018,406.00	422,822.00	-
.99	-	-	-	-

Figure 2. Recoding of SizeClass (all risky cells have disappeared)

	tot	2	4	5	6	7	8	9	99
tot	16,847,646.84	20.00	25.00	2,711,808.00	2,320,534.00	2,505,042.58	2,799,074.26	6,510,758.00	385.00
.North	4,373,664.00	×	×	719,049.00	659,680.00	688,962.00	756,529.00	1,549,049.00	385.00
.East	3,703,896.00	15.00	×	642,238.00	515,003.00	534,147.00	620,392.00	1,392,096.00	-
.West	4,576,115.84	-	-	648,972.00	543,570.00	663,896.58	775,132.26	1,944,545.00	-
.South	4,193,971.00	-	15.00	701,549.00	602,281.00	618,037.00	647,021.00	1,625,068.00	-
.99	-	-	-	-	-	-	-	-	-

Figure 3. Recoding of Region (not all risky cells have disappeared)

2.4.2 Cell suppression

2.4.2.1 Short description

A frequently used method to protect risky cells is to suppress (not publish) certain cells. The cell value is then simply replaced by a certain symbol, e.g., a cross (×).

In a quantitative table when the marginals are also provided, however, it is often not sufficient to suppress only the risky cells (i.e., only use so-called primary suppressions). If a suppressed cell is the only suppressed cell in a row, the suppressed value can, after all, simply be calculated by subtracting the other cell values in that row from the corresponding marginal.

To sufficiently protect risky cells, it is therefore also necessary to suppress other cells which, in themselves, are safe. This is called *secondary suppression*. It is not easy to perform this in such a way such that the risky cells are protected sufficiently, while also ensuring that not too much information is removed from the table. Furthermore, account must also be taken of the fact that structural zero cells cannot be used as secondary suppressions: everyone knows that, by definition, these cells are empty.

To prevent a situation where suppressed, risky cells can be (re)calculated exactly, secondary suppressions are therefore necessary. However, also a "too accurate" estimation for a suppressed cell is not desirable. Indeed, what is the difference between the following statements: "This suppressed cell actually has a value of 10000" and "This suppressed cell actually has a value of between 9998 and 10002". Given a suppression pattern, it is always[1] possible to calculate an interval in which a suppressed cell must lie. The method of "Cell Suppression" must then also produce a suppression pattern, for which the intervals that can be calculated are sufficiently large. The size of these intervals is determined by the rule that is used to determine the risky cells.

Fischetti and Salazar (2000) have developed a method to solve the above problem in an optimal manner. Their method is, in theory, applicable to arbitrary, additive tables with non-negative contributions. In practice, however, their solution involves too much computing time if the tables become too large, either in size or complexity. This is why a number of sub-optimal methods have been developed to find suitable suppression patterns for larger and/or more complex tables.

For example, the "modular approach" (also known as HiTaS) splits a hierarchical table into a large number of non-hierarchical sub-tables and applies the optimal method to each individual sub-table. By correctly combining the results, a sub-optimal solution can be obtained for the entire table, with a significantly shorter computing time.

The "hypercube approach" can also protect large tables by protecting the sub-tables in a certain iterative way. The protection of each sub-table also takes place sub-optimally. Consequently, the approach is relatively fast, but, in general, more cells are suppressed than strictly necessary to obtain a protected table.

2.4.2.2 Applicability

This method can be used to adequately protect quantitative tables with cells that do not satisfy the requirements of the NSI's statistical disclosure control policy. In particular, if the table cannot be restructured further or at all, the cell suppression method can be used effectively.

The contributions to the table to be protected must not be negative[2] and the table must be additive. If no marginals are provided, secondary cell suppression is not needed. When marginals are provided, secondary cell suppression is usually needed to properly protect the sensitive cells.

In the modular approach, the table may be at most three-dimensional. Each dimension may be hierarchical. The limit on the dimensionality of the table is due to the fact that for higher dimensional tables, the calculation time would grow exponentially and effectively become too large.

[1] In case the table is composed of non-negative contributions and the marginals are provided as well.

[2] The requirement of non-negativity can be relaxed to the requirement that the values should be uniformly bounded from below. However, this requires an adaptation of the concentration rules. See, e.g., Hundepool et al. (2012).

Linked tables can be protected by copying the suppressions from one table to the other, and then protecting the tables. This should then possibly be performed in an iterative manner. The current version of τ-ARGUS is able to solve certain classes of linked tables problems automatically.

In the hypercube approach as implemented in τ-ARGUS, the table may be at most seven-dimensional. The table may be hierarchical in every dimension. Linked tables are also *possible.*

In theory, neither the modular approach nor the hypercube approach are limited in dimensionality of the tables. It is purely for performance issues, that the dimensionality is limited in the way these approaches are implemented in τ-ARGUS.

Moreover, it should be mentioned that for both approaches, from a performance perspective, the recommendation is to avoid using long, unstructured (non-hierarchical) code lists.

2.4.2.3 Detailed description

To apply statistical disclosure control techniques to tabular data, specialised software is available. In Europe, the most commonly used "generally available" software is τ-ARGUS. For that reason, the following paragraphs are dedicated to explaining methods as implemented in τ-ARGUS.

Other software packages that are available are: sdcTable (R package, no user interface available) and G-Confid (see, e.g., Statistics Canada, 2011). For a general discussion of different software tools, see Giessing (2013).

The software package τ-ARGUS has a provision to apply cell suppression to quantitative tables. If the original microdata is used as input, τ-ARGUS will determine the risky cells with the associated safety intervals.

After this, τ-ARGUS will have to determine a suppression pattern that guarantees the necessary safety intervals. There are various options for this. We will discuss the two approaches that are the most interesting for Statistics Netherlands.

2.4.2.4 Modular approach

Generally, the modular approach can be described as follows:

1. Split the hierarchical table into all logical non-hierarchical sub-tables.
2. Group the sub-tables in classes in such a way that all tables in a single class can be protected independently of each other. For a suitable classification, see De Wolf (2002).
3. Protect all tables in class K.
4. If no secondary suppressions are placed in the marginals of the sub-tables of class K, continue with class $K + 1$, including any secondary suppressions in the inside of a table as primary suppressions for class $K + 1$.
5. If secondary suppressions do have to be placed in a marginal of at least one sub-table, go back to class $K - 1$, including only the secondary suppressions in the marginals as primary suppressions.
6. Repeat steps 4 and/or 5 until all sub-tables have been protected at the lowest (most detailed) hierarchical level.

All non-hierarchical sub-tables will be protected using the mixed integer approach from Fischetti and Salazar (2000). In this approach, the required safety intervals are guaranteed, while a certain cost function is minimised. This cost function can be selected in different ways, as a result of which various forms of information loss can be minimised. This minimisation takes place *locally*, so that the ultimate solution for the entire (hierarchical) table does not necessarily also have to be optimal.

Note that in this way, the required safety intervals are guaranteed when using the subset of table relations that define the sub-table. In certain specific situations it might be possible that a required safety interval is not attained when using the complete set of table relations that defines the hierarchical table.

In selecting the cost function in τ-ARGUS, several options can be selected, including:

- A variable from the dataset (such as the quantitative value on which tabulation takes place);
- A constant (so that the number of suppressions is minimised);
- The number of contributors per cell (so that the total number of suppressed contributions is minimised).

In the disclosure control of a sub-table, also the so-called singletons problem must be taken into account: cells with only one contribution. If such cells are in a suppression pattern, the contributors involved can reverse part or all of the suppression pattern. After all, they know what their own contribution is and can therefore fill in that suppressed value, as a result of which it may also be possible to calculate other suppressed cells. In the current implementation of the mixed integer approach in τ-ARGUS, it is not possible to keep each conceivable combination of a singleton with another suppressed cell under control while searching for a suppression pattern. However, it is possible to take account of the combinations within a single row, column or layer[3] in the table. The combinations which must be taken into account consist of exactly two risky cells in a single row, column or layer, of which at least one cell is a singleton. By requiring a small safety interval for the combination of these two cells, it will be made sure that even with knowledge of one of these cells, it is not possible to exactly disclose the other risky cell.

In a similar way, it is ensured that, within a single row, column or layer, all the suppressed cells together contain more than the minimum required number of contributors for a safe cell.

For a detailed description and an elaborated example of the modular approach, see De Wolf (2002). For a detailed description of the adjustments to be able to deal with linked tables, see De Wolf and Giessing (2009).

2.4.2.5 Hypercube approach

In this approach too, a hierarchical table is split into non-hierarchical sub-tables. The non-hierarchical sub tables are then protected in a certain order, where the sub-tables at the highest level are dealt with first.

[3] A row consists of the cells with coordinates (r, k, l) where k and l are fixed. A column consists of the cells with the coordinates (r, k, l) where r and l are fixed. A layer consists of the cells with coordinates (r, k, l) where r and k are fixed.

For each sub-table, all possible hyper cubes are constructed for each risky cell in which that risky cell is one of the corner points. For each hypercube, the interval is calculated around the risky cell if all other corner points of the hypercube are also suppressed. If that interval is large enough (depending on the protection rule used), the associated hypercube is designated as “feasible”. The information loss is then calculated for each feasible hypercube. Finally, the feasible hypercube with the smallest information loss is selected to protect the risky cell concerned.

No linear programming problem needs to be solved in order to calculate the safety intervals resulting from a hypercube. This significantly accelerates the procedure. The hypercube approach is therefore, in general, faster than the modular approach, for which a mixed integer programming problem needs to be solved.

After all sub-tables are protected in this way, the entire procedure is repeated. Secondary suppressed cells from a certain sub-table that also occur in other sub-tables are considered as sensitive cells in those other sub-tables, and dealt with as such. This process is repeated until no more changes take place.

Note that the use of hype cubes to protect risky cells is a sufficient but not necessary condition for a safe suppression pattern. In other words, in some cases, the combination of the different hyper cubes will not lead to an optimal suppression pattern, but it will always produce a safe suppression pattern. Consequently, this approach tends to suppress more cells than necessary for a safe suppression pattern.

This approach also takes account of the so-called singletons. A cell with only one contributor would indeed allow all suppressed corner points of a hypercube to be calculated. Therefore the extra requirement in the case of singletons is that this type of cell must be a corner point of at least two different hypercubes.

As said, the hypercube method for hierarchical tables also splits a hierarchical table into non-hierarchical sub-tables. Therefore, the protection that is provided is of a similar level as with the modular approach. I.e., the required safety intervals are guaranteed when using the subset of table relations that defines the sub-table. In certain specific situations it might be possible that a required safety interval is not attained when using the complete set of table relations that defines the hierarchical table.

2.4.2.6 Example

Using τ-ARGUS, it is easy to apply cell suppression to a quantitative table. Both the modular approach and the hypercube approach are implemented in τ-ARGUS. It is also possible to select multiple information loss measures for the cost function that must be minimised. See Section 4 for more information on τ-ARGUS.

Figure 4 shows an example of a table with some sensitive cells suppressed.

It is clear that this is not sufficient: both the cell (East, 4) and the cell (4, 9) can be directly calculated: (East, 4) = 3 703 896 – 15 – 642 238 – 515 003 – 534 147 – 620 392 – 1 392 096 = 5 and (4, 9) = 1 392 096 – 145 004 – 1 083 254 – 151 870 = 11 968.

	Total	2	4	5	6	7	8	9	99
Total	16,847,646.84	20.00	25.00	2,711,808.00	2,320,534.00	2,505,042.58	2,799,074.26	6,510,758.00	385.00
- .Nr	4,373,664.00	X	X	719,049.00	659,680.00	688,962.00	756,529.00	1,549,049.00	385.00
.. 1	1,986,129.00	X	X	398,062.00	348,039.00	354,711.00	418,778.00	466,529.00	-
.. 2	1,809,246.00	0.00	-	223,990.00	221,332.00	241,913.00	258,233.00	863,393.00	385.00
.. 3	578,289.00	-	-	96,997.00	90,309.00	92,338.00	79,518.00	219,127.00	-
- .Os	3,703,896.00	15.00	X	642,238.00	515,003.00	534,147.00	620,392.00	1,392,096.00	-
.. 4	124,336.00	X	-	36,311.00	32,132.00	25,770.00	18,150.00	X	-
.. 5	526,279.00	-	-	93,589.00	94,957.00	110,930.00	81,799.00	145,004.00	-
.. 6	2,234,995.00	X	X	345,803.00	251,358.00	251,188.00	303,377.00	1,083,254.00	-
.. 7	818,286.00	-	-	166,535.00	136,556.00	146,259.00	217,066.00	151,870.00	-
- .Ws	4,576,115.84	-	-	648,972.00	543,570.00	663,896.58	775,132.26	1,944,545.00	-
.. 8	485,326.00	-	-	63,767.00	75,442.00	87,305.00	59,953.00	198,859.00	-
.. 9	3,664,559.84	-	-	537,911.00	430,851.00	515,019.58	643,762.26	1,537,016.00	-
..10	426,230.00	-	-	47,294.00	37,277.00	61,572.00	71,417.00	208,670.00	-
- .Zd	4,193,971.00	-	15.00	701,549.00	602,281.00	618,037.00	647,021.00	1,625,068.00	-
..11	2,752,743.00	-	15.00	488,613.00	392,395.00	363,490.00	402,925.00	1,105,305.00	-
..12	1,441,228.00	-	-	212,936.00	209,886.00	254,547.00	244,096.00	519,763.00	-
.99	-	-	-	-	-	-	-	-	-

Figure 4. Quantitative table for turnover according to region and size class

Figure 5 shows the suppression pattern that was determined with τ-ARGUS using the hypercube approach. Figure 6 shows the same based on the modular approach. Of course, in a publication, it should be impossible to make a distinction between primary and secondary suppressions.

	Total	2	4	5	6	7	8	9	99
Total	16,847,646.84	20.00	25.00	2,711,808.00	2,320,534.00	2,505,042.58	2,799,074.26	6,510,758.00	385.00
- .Nr	4,373,664.00	X	X	719,049.00	X	688,962.00	756,529.00	1,549,049.00	385.00
.. 1	1,986,129.00	X	X	398,062.00	X	354,711.00	418,778.00	466,529.00	-
.. 2	1,809,246.00	X	-	223,990.00	X	241,913.00	258,233.00	863,393.00	385.00
.. 3	578,289.00	-	-	96,997.00	90,309.00	92,338.00	79,518.00	219,127.00	-
- .Os	3,703,896.00	X	X	642,238.00	X	534,147.00	620,392.00	1,392,096.00	-
.. 4	124,336.00	X	-	36,311.00	X	25,770.00	X	X	-
.. 5	526,279.00	-	-	93,589.00	94,957.00	110,930.00	X	X	-
.. 6	2,234,995.00	X	X	345,803.00	X	251,188.00	303,377.00	1,083,254.00	-
.. 7	818,286.00	-	-	166,535.00	136,556.00	146,259.00	217,066.00	151,870.00	-
- .Ws	4,576,115.84	-	-	648,972.00	543,570.00	663,896.58	775,132.26	1,944,545.00	-
.. 8	485,326.00	-	-	63,767.00	75,442.00	87,305.00	59,953.00	198,859.00	-
.. 9	3,664,559.84	-	-	537,911.00	430,851.00	515,019.58	643,762.26	1,537,016.00	-
..10	426,230.00	-	-	47,294.00	37,277.00	61,572.00	71,417.00	208,670.00	-
- .Zd	4,193,971.00	-	X	701,549.00	X	618,037.00	647,021.00	1,625,068.00	-
..11	2,752,743.00	-	X	488,613.00	X	363,490.00	402,925.00	1,105,305.00	-
..12	1,441,228.00	-	-	212,936.00	209,886.00	254,547.00	244,096.00	519,763.00	-
.99	-	-	-	-	-	-	-	-	-

Figure 5. Suppression pattern for the table from Figure 4, using the hypercube approach

	Total	2	4	5	6	7	8	9	99
Total	16,847,646.84	20.00	25.00	2,711,808.00	2,320,534.00	2,505,042.58	2,799,074.26	6,510,758.00	385.00
- .Nr	4,373,664.00	X	X	719,049.00	X	688,962.00	756,529.00	1,549,049.00	385.00
.. 1	1,986,129.00	X	X	398,062.00	X	354,711.00	418,778.00	466,529.00	-
.. 2	1,809,246.00	0.00	-	223,990.00	221,332.00	241,913.00	258,233.00	863,393.00	385.00
.. 3	578,289.00	-	-	96,997.00	90,309.00	92,338.00	79,518.00	219,127.00	-
- .Os	3,703,896.00	X	X	642,238.00	515,003.00	534,147.00	620,392.00	1,392,096.00	-
.. 4	124,336.00	X	-	36,311.00	32,132.00	X	X	X	-
.. 5	526,279.00	-	-	93,589.00	94,957.00	110,930.00	X	X	-
.. 6	2,234,995.00	X	X	345,803.00	251,358.00	X	303,377.00	1,083,254.00	-
.. 7	818,286.00	-	-	166,535.00	136,556.00	146,259.00	217,066.00	151,870.00	-
- .Ws	4,576,115.84	-	-	648,972.00	543,570.00	663,896.58	775,132.26	1,944,545.00	-
.. 8	485,326.00	-	-	63,767.00	75,442.00	87,305.00	59,953.00	198,859.00	-
.. 9	3,664,559.84	-	-	537,911.00	430,851.00	515,019.58	643,762.26	1,537,016.00	-
..10	426,230.00	-	-	47,294.00	37,277.00	61,572.00	71,417.00	208,670.00	-
- .Zd	4,193,971.00	-	X	701,549.00	X	618,037.00	647,021.00	1,625,068.00	-
..11	2,752,743.00	-	X	488,613.00	X	363,490.00	402,925.00	1,105,305.00	-
..12	1,441,228.00	-	-	212,936.00	209,886.00	254,547.00	244,096.00	519,763.00	-
.99	-	-	-	-	-	-	-	-	-

Figure 6. Suppression pattern for the table from Figure 4, using the modular approach

For a more detailed description of the hypercube approach, see Hundepool et al. (2011, Section 2.8). References to the original literature on this method can also be found there.

2.4.3 Waivers

Sometimes, the need to maintain the confidentiality of the contribution of a particular respondent may result in disastrous results. E.g., just to protect that single respondent, it may be that many additional (secondary) suppressions are needed. In those cases, if the local law permits, it may be good practice to ask the respondent in question for a so-called *waiver*. That is, permission is asked to publish a table cell that contains the contribution of that respondent, even though it may not pass the primary confidentiality rule. According to the Dutch Statistical Law, waivers are permissible in economic surveys, provided a formal agreement of the respondent is present.

When waivers are used, the sensitivity rule that is used to identify the risky cells needs to be adjusted. This follows from the fact that some but not all respondents to a particular cell may have given a waiver. To adjust sensitivity rules in the presence of waivers, see Hundepool et al. (2012, Chapter 4).

3. Design issues

The issue here is to make the necessary preparations for protecting tables to be issued by an NSI. In other to facilitate the production of safe (enough) tables relatively quickly, it is mandatory that standardised procedures are available for the staff responsible to protecting tables. These persons will be typically scattered over an NSI, working in different departments. In this section we discuss what elements are important for such rules. However, we will not go into this matter exhaustively nor discuss the choice of the various parameters. This is impossible and depends on local circumstances in a country, and the statistical laws and practices that have to be taken into account.

3.1 Sensitivity rules

The criteria to test the safety of tables typically operate at the cell level. So they can be used to test which cells are considered safe and which not. These criteria have parameters that have to be specified by the NSI responsible for the disclosure control of its tables. They should be specified along with the criteria, and should be part of the disclosure control policy of the NSI. The specification, apart from the choice of the kind of sensitivity measure, is a choice for the parameters to use

3.2 Choice of table protection methods

To protect the sensitive cells in tabular data, the NSI has to specify what SDC methods will be used to protect the tables they want to release. There may be a choice of techniques available, but which one(s) are to be applied in a particular case depends also on the user demand.

3.3 Longitudinal aspects

Special attention needs to be paid to longitudinal data or panel data, in which the same entities (say businesses) yield data at several points in time. It is then not sufficient to protect the data at each point in time as if they are cross-sectional data.

4. Available software tools

τ-ARGUS is a package intended to protect tabular data by various techniques, such as table redesign, various versions of cell suppression, rounding and controlled tabular adjustment. For more information see Hundepool et al. (2011). This package requires a commercial LP-solver (either Xpress or Cplex) for certain techniques (like cell suppression and rounding). The τ-ARGUS package itself, however, is free of charge. See also http://neon.vb.cbs.nl/casc/index.htm. Currently, non-commercial Open Source LP Solvers are investigated to be included in future versions of τ-ARGUS. In τ-ARGUS one can apply cell-suppression to unstructured tables and hierarchical tables, to single tables and sets of lined tables.

There are other packages for the protection of tabular data, such as sdcTable (R package, no user interface available) and G-Confid (see, e.g., Statistics Canada, 2011). For a general discussion of different software tools, see Giessing (2013).

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

De Wolf, P. P. (2002), HiTaS: a heuristic approach to cell suppression in hierarchical tables. *Proceedings of the AMRADS meeting in Luxembourg 2002*.

De Wolf, P. P. and Giessing, S. (2009), How to make the τ-ARGUS modular approach to deal with linked tables. *Data & Knowledge Engineering* **68**, 1160–1174.

Fischetti, M. and Salazar Gonzales, J. J. (2000), Models and Algorithms for Optimizing Cell Suppression in Tabular Data with Linear Constraints. *Journal of the American Statistical Association* **95**, 916–928.

Giessing, S. (2013), Software tools for assessing disclosure risk and producing lower risk tabular data. Data Without Boundaries Deliverable 11.1 – Part B, February 2013. (http://www.dwbproject.org/export/sites/default/about/public_deliveraples/dwb_d11-1b_software-tools-disclosure-risk-assessment.pdf)

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K., and De Wolf, P. P. (2012), *Statistical disclosure control*. Wiley-Blackwell.

Hundepool, A., Van de Wetering, A., Ramaswamy, R., De Wolf, P. P., Giessing, S., Fischetti, M., Salazar, J. J., Castro, J., and Lowthian, P. (2011), *τ-ARGUS user manual 3.5*. Statistics Netherlands, Voorburg.

Statistics Canada (2011), *G-Confid User Manual*. Internal report.

Interconnections with other modules

8. Related themes described in other modules

1. Statistical Disclosure Control – Main Module

9. Methods explicitly referred to in this module

1.

10. Mathematical techniques explicitly referred to in this module

1. Linear programming
2. Mixed integer programming

11. GSBPM phases explicitly referred to in this module

1. 6.4 Apply disclosure control

12. Tools explicitly referred to in this module

1. τ-ARGUS
2. sdcTable
3. G-Confid

13. Process steps explicitly referred to in this module

1. Statistical disclosure control

Administrative section

14. Module code

Statistical Disclosure Control-T-Methods for Quantitative Tables

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	04-09-2013	first version	Leon Willenborg, Peter-Paul de Wolf	Statistics Netherlands (CBS)
0.2	24-01-2014	revised version after review	Leon Willenborg, Peter-Paul de Wolf	Statistics Netherlands (CBS)
0.3	04-02-2014	minor revision after review by EB	Peter-Paul de Wolf	Statistics Netherlands (CBS)
0.4	05-02-2014	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	26-3-2014 13:30