

Using Tau-Argus and sdcTable to Conduct Secondary Cell Suppression for Linked Tables

2011 Joint Statistical Meeting, Miami Beach, FL,
August 3, 2011 (Session 455)

Amang Sukasih, Donsig Jang, and David Edson

MATHEMATICA
Policy Research, Inc.

Overview

- **Primary and secondary cell suppression**
- **Protection for linked tables**
- **Software: τ -Argus and sdcTable**
- **Comparison**

Primary and Secondary Cell Suppression

■ **Primary cell:**

- **A table cell that could reveal individual respondent information**
- **Determined based on**
 - Dominance or (n,k) rule,
 - P% rule, or
 - Frequency or threshold rule

■ **Secondary cell:**

- **Additional cell(s) suppressed to make sure the primary suppressions cannot be derived by subtraction from published marginal totals**

Primary and Secondary Cell Suppression

- Example of threshold rule:

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

Primary and Secondary Cell Suppression

- Example of threshold rule: suppress < 5

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

 Primary suppression

Primary and Secondary Cell Suppression

- Example of threshold rule: suppress < 5

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5		9	6	5	11
012	5		8			
013		6	9		6	7
020	11	15	26	16	10	26
021		7	10			7
022			7	7		11
023		5	9	5		8
Total	24	28	52	26	22	48

 Primary suppression

Primary and Secondary Cell Suppression

- Example of threshold rule: suppress < 5

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5		9	6	5	11
012	5		8			
013		6	9		6	7
020	11	15	26	16	10	26
021		7	10			7
022			7	7		11
023		5	9	5		8
Total	24	28	52	26	22	48



Primary suppression



Secondary suppression

Primary and Secondary Cell Suppression

- **Secondary cell suppression methods:**
 - Hypercube
 - HiTaS/Modular
 - Optimal
 - Network flow
 - Rounding

Linked Tables

- Examples of link tables

TABLE 2

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			
013			9		6	
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

Linked Tables

■ Examples of link tables

TABLE 1			
VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

TABLE 2						
VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			
013			9		6	
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

VAR1 and VAR2 are common variables

TABLE 1 provides marginal totals of VAR3

Linked Tables

■ Examples of link tables

TABLE 1			
VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

TABLE 2						
VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			4
013			9		6	
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

VAR1 and VAR2 are common variables

TABLE 1 provides marginal totals of VAR3

Linked Tables

■ Examples of link tables

TABLE 1			
VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

TABLE 2						
VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			4
013			9		6	7
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

VAR1 and VAR2 are common variables

TABLE 1 provides marginal totals of VAR3

Linked Tables

■ Examples of link tables

TABLE 1			
VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

TABLE 2						
VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			4
013			9	1	6	7
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

VAR1 and VAR2 are common variables

TABLE 1 provides marginal totals of VAR3

Linked Tables

■ Examples of link tables

TABLE 1			
VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

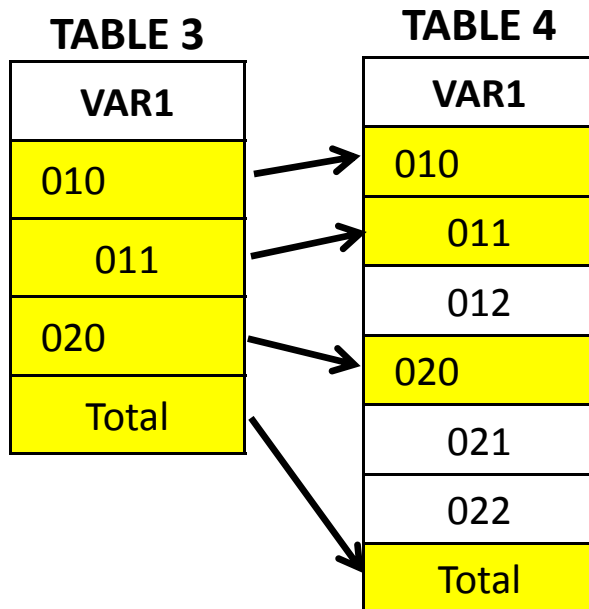
TABLE 2						
VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011			9			11
012			8			4
013	3		9	1	6	7
020	11	15	26	16	10	26
021			10			7
022			7			11
023			9			8
Total	24	28	52	26	22	48

VAR1 and VAR2 are common variables

TABLE 1 provides marginal totals of VAR3

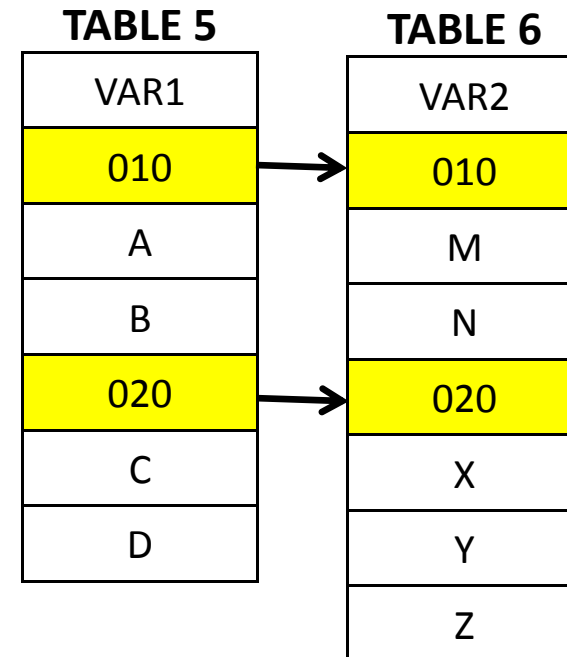
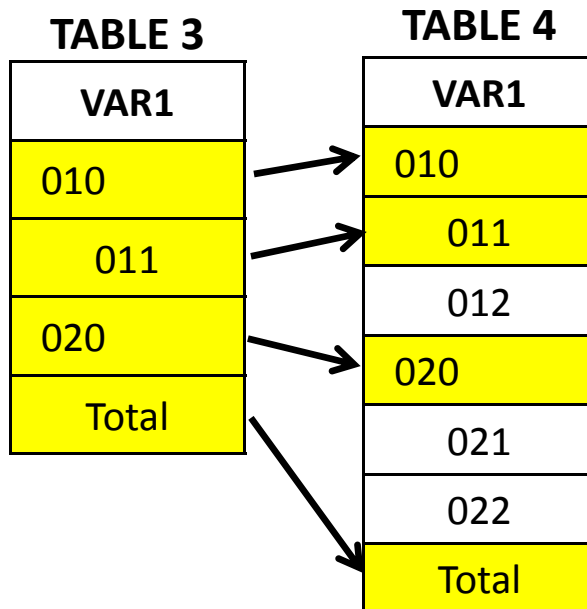
Linked Tables

- Examples of link tables (continued)



Linked Tables

- Examples of link tables (continued)



Software for Secondary Suppression

■ **τ-Argus by Statistics Netherland**

- Freeware, available for download from <http://neon.vb.cbs.nl/casc/tau.htm>
- Windows platform only
- Menu driven user interface
- Microdata or tabular data as input
- Performs tabulation when the input data is a microdata
- Capability to recode categories on the fly, put cost to each cell, and keep track suppression history
- Available safety rule for primary suppression:
 - Dominance or (n,k) rule
 - P% rule
 - Frequency or threshold rule

Software for Secondary Suppression

■ **τ -Argus (continued)**

- **Available methods for secondary suppression:**
 - Hypercube, HiTaS/Modular, Optimal, Network, Rounding
 - HiTas and Optimal require external commercial optimization solver.
- **Disclosure protection for linked tables (version 3.5.0, build 4, May 9, 2011)**
- **Complete manual (116 pages), but PDF is not searchable**
- **Proprietary software; no help, error messages are not intuitive, support only from author/maintainer (Anco Hundepool)**

Software for Secondary Suppression

- **sdcTable by Bernhard Meindl (Statistik Austria)**
 - Freeware, *R* package (<http://www.r-project.org/>)
 - Require *R* programming knowledge
 - Can access the codes (open source software)
 - Available for Windows, Linux, Mac
 - Available safety rule for primary suppression:
 - Dominance or (n,k) rule
 - P% rule
 - Frequency or threshold rule
 - Available methods for secondary suppression:
 - Hypercube, HiTaS

Software for Secondary Suppression

- **sdcTable (continued)**
 - Disclosure protection for linked tables (version 0.6.4, April 4, 2011)
 - Limited manual; however, can see the source codes
 - Extra support from *R* community (*R* forum)

Software for Secondary Suppression

■ Ichim and Franconi (2009):

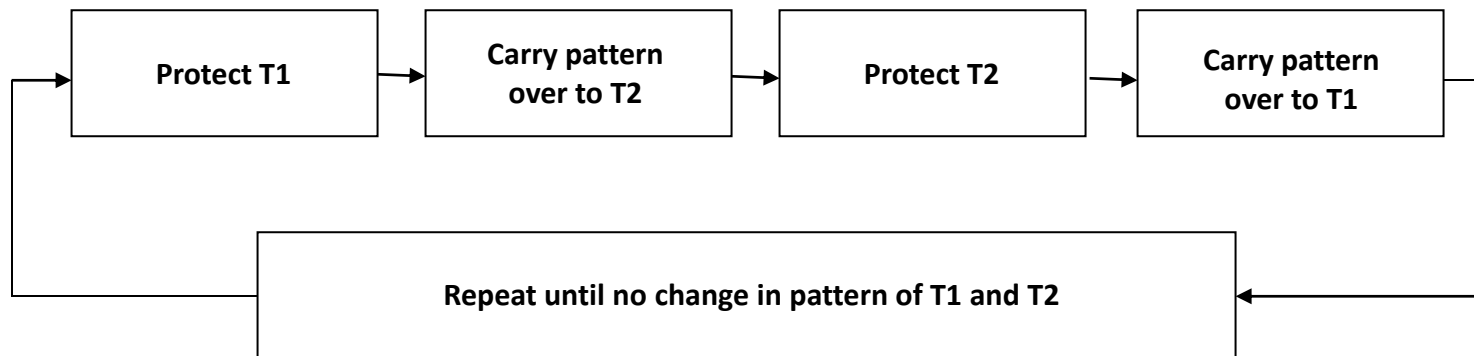
Feature	Tau-Argus	sdcTable
Possibility to check/control/modify/adapt	NO	YES
Coordinated development	YES	NO
Predictable results	YES	NO
Development agenda	YES	NO
SDC people involved	YES	NO
Documentation	YES	YES/NO
Help	NO	YES/NO
Modular architecture	YES	YES
Extensibility	YES	YES
Platform dependent	YES	NO
Unique maintainer	YES	YES
Personalisation	NO	YES
User-friendly	YES/NO	NO
Programming skills required	NO	YES
Free	YES/NO	YES
Open source	NO	YES
Designed for official statistics	YES	NO/YES
Mirror sites	NO	YES
Consortium	YES	NO
Test reports	YES	NO

Linked Table Suppression

- **Suppression methods for linked tables (de Wolf and Giessing 2009):**
 - **Complete modular approach:**
 - Modular approach to protect the complete cover table
 - **Adapted modular approach:**
 - Modular approach on the cover table, but only consider those subtables that are also subtables of at least one of the specified tables $T_1, \dots, T_N$, and disregard the other subtables; i.e., any simple subtable that is not a subtable of any tables in the set $\{T_1, \dots, T_N\}$ is skipped

Linked Table Suppression

- **Suppression methods for linked tables (continued):**
 - **Linked subtables modular approach:**
 - more complex approach deals with sets of linked, simple subtables at a time
 - **Traditional approach:**
 - Iterative backtracking procedure



Linked Table Suppression

- **Suppression methods for linked tables (continued):**

de Wolf and Giessing (2009) advocate the Adapted Modular approach and the Linked Subtables Modular approach. However, the **Adapted Modular approach is preferable for practical reasons:**

- less complex algorithm
- good result
- less computation

Linked Table Suppression

- **Methods directly available in software**
 - **Tau-Argus**
 - Adapted modular approach (optimization using HiTaS)
 - Linked table suppression using Hypercube is also available
 - **sdCTable**
 - Iterative backtracking procedure
 - Stop criterion: all common cells have the same suppression status
 - Optimization using Hypercube or HiTaS

Comparison of TauArgus and sdcTable

- Example of a pair of link tables
- Implement threshold rule: suppress < 5

TABLE 1

VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

TABLE 2

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

Comparison of TauArgus and sdcTable

- **Tau-Argus linked tables suppression**
 - Based on adapted modular

TABLE 1

VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

 Primary suppression
 Secondary suppression

TABLE 2

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

Comparison of TauArgus and sdcTable

■ Individual vs. linked table suppression

Individual suppressed (Table 1 is not given)

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subt	01	02	Subt
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

Linked table suppressed

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subt	01	02	Subt
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

 Primary suppression

 Secondary suppression

Comparison of TauArgus and sdcTable

- **sdcTable linked tables suppression**
 - Based on iterative modular

TABLE 1

VAR1	VAR2		
	01	02	Total
010	23	25	48
011	11	9	20
012	8	4	12
013	4	12	16
020	27	25	52
021	7	10	17
022	11	7	18
023	9	8	17
Total	50	50	100

 Primary suppression
 Secondary suppression

TABLE 2

VAR1	VAR3					
	01			02		
	VAR2			VAR2		
	01	02	Subtot	01	02	Subtot
010	13	13	26	10	12	22
011	5	4	9	6	5	11
012	5	3	8	3	1	4
013	3	6	9	1	6	7
020	11	15	26	16	10	26
021	3	7	10	4	3	7
022	4	3	7	7	4	11
023	4	5	9	5	3	8
Total	24	28	52	26	22	48

Summary

- **Software packages for table protection that implement optimization technique are publicly available**
- **Options for linked tables continue to be developed**
- **In sdcTable common cells can come from different variables; user specifies the common cells**
- **In Tau-Argus common cells have to be subset of categories in a common variable**

Summary

- **Current release of sdcTable only deals with two linked tables. Tau-Argus can handle more than two linked tables as long as the cover table does not have more than 4 dimensions.**
- **Current release of sdcTable may not be able to give results due to non-converged iteration (bug?)**
- **protectLinkedTables in the sdcTable package can easily lead to over protection**
- **Different results are due to implementation of different methods. Protection for linked tables is limited by the availability of software.**

For more information

- **Contact:**

Amang Sukasih

asukasih@mathematica-mpr.com