Working Paper No. 25
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 ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (i): New theories and emerging methods

THE DETERMINATION OF INTERVALS OF SUPPRESSED CELLS IN AN *n*-DIMENSIONAL TABLE

Contributed Paper

Submitted by the Technical University Ilmenau (Germany)[1]

[1] Prepared by Karl Luhn (karl.luhn@tu-ilmenau.de).

The Determination of Intervals of Suppressed Cells in an *n*-dimensional Table

We consider statistical tables with sensitive cells. These cells are suppressed, either primary or secondary. The difference between them is well known but insignificant for the attacker problem. It is only import to the attacker to know which cells are suppressed cells. The solution of the attacker problem will be found by the means of linear programming. The suppressed cells are not included in the published table. In the model of linear programming one equation, dependent on these variables, belongs to each total or subtotal of the table. Hence we have objective functions twice as variables (primary and secondary suppressed cells). This demonstrates the possibility - but the high efforts too – to solve the attacker problem by linear programming.

Let us consider an example as a 4-dimensional table:

		$l = 1$			$l = 2$			$\mathbf{a}_k^T$		
		$j = 1$	$j = 2$	$a_{ik,l=1}$	$j = 1$	$j = 2$	$a_{ik,l=2}$	$a_{i,j=1,k}$	$a_{i,j=2,k}$	a_{ik}
$k = 1$	$i = 1$		32	**P**		16		40	48	88
	$i = 2$			44	8				46	**P**
	$i = 3$								57	
	$b_{j,k=1,l}$	45	76	121	59	75	134	104	151	255
$k = 2$	$i = 1$		39			17		50	56	106
	$i = 2$			48	75				64	
	$i = 3$								91	
	$b_{j,k=2,l}$	56	110	166	135	101	236	191	211	402
$\mathbf{b}_l$	$b_{i=1,jl}$	20	71	91	70	33	103	90	104	194
	$b_{i=2,jl}$	48	44	92	83	66	149	131	110	241
	$b_{i=3,jl}$	33	71	104	41	77	118	74	148	222
	b_{jl}	101	186	287	194	176	370	295	362	657

“P” means a primary suppressed cell, each other empty cell is a secondary suppressed one. There are 2 primary and 34 secondary suppressed cells, thus 72 objective functions.

The first reflection to make the problem easier was to solve it with parameters, belonging to the **primary** suppressed cells. But it is impossible to define these primary suppressed cells. Therefore we solve the problem in two steps: At first we find a solution of the system of linear equations.

We will show this with a small example:

i \ j	1	2	3	4	Total
1	100	80	120	100	400
2	50	20	170	60	300
3	150	100	110	40	400
Total	300	200	400	200	1100

After suppression (P – primary, S – secondary):

i j	1	2	3	4	Total
1	100	80	120	100	400
2	50	**P**	170	60	**S**
3	150	**S**	110	40	**P**
Total	300	200	400	200	1100

Hence it follows the system of linear equations with the cells x_{ij} and the totals $x_{\cdot j}$ resp. $x_{i\cdot}$:

$$-x_{22} + x_{2\cdot} = 280$$
$$-x_{32} + x_{3\cdot} = 300$$
$$x_{22} + x_{32} = 120$$
$$x_{2\cdot} + x_{3\cdot} = 700$$

The solution of 8 linear programming problems is as follows:

	min	max
x_{22}	0	120
x_{32}	0	120
$x_{2.}$	280	400
$x_{3.}$	300	420

The second step is to solve the remaining system by the means of linear programming. We receive intervals of the remaining variables and determine with them the intervals of all variables (primary an secondary suppressed cells) of the original system.

If we are solving the linear equation system at first we get:

$$x_{2\cdot} = 400 - x_{32}$$
$$x_{3\cdot} = 300 + x_{32}$$
$$x_{22} = 120 - x_{32}$$

It is to interpret as a task of linear programming with the objective functions $x_{32} \to \min$ and $x_{32} \to \max$ under the restrictions $x_{32} \le 400; x_{32} \ge -300; x_{32} \le 120$. The solution is $\min x_{32} = 0$ and $\max x_{32} = 120$. The other intervals we receive by inserting in the above-mentioned equations.

This solution delivers a system dependent on not more variables than primary suppressed cells exist.

In the first example we can determine by this means the equations:

		$l = 1$			$l = 2$			$\mathbf{a}_k^{\mathrm{T}}$		
		$j = 1$	$j = 2$	$a_{ik,l=1}$	$j = 1$	$j = 2$	$a_{ik,l=2}$	$a_{i,j=1,k}$	$a_{i,j=2,k}$	a_{ik}
$k = 1$	$i = 1$	X1111	32	X1.11	X1112	16	X1.12	40	48	88
	$i = 2$	X2111	X2211	44	8	X2212	X2.12	X211.	46	X2.1.
	$i = 3$	X3111	X3211	X3.11	X3112	X3212	X3.12	X311.	57	X3.1.
	$b_{j,k=1,l}$	45	76	121	59	75	134	104	151	255
$k = 2$	$i = 1$	X1121	39	X1.21	X1122	17	X1.22	50	56	106
	$i = 2$	X2121	X2221	48	75	X2222	X2.22	X212.	64	X2.2.
	$i = 3$	X3121	X3221	X3.21	X3122	X3222	X3.22	X312.	91	X3.2.
	$b_{j,k=2,l}$	56	110	166	135	101	236	191	211	402
$\mathbf{b}_l$	$b_{i=1,jl}$	20	71	91	70	33	103	90	104	194
	$b_{i=2,jl}$	48	44	92	83	66	149	131	110	241
	$b_{i=3,jl}$	33	71	104	41	77	118	74	148	222
	b_{jl}	101	186	287	194	176	370	295	362	657

The system consists of 32 horizontal and 34 vertical equations and 36 variables.

The first step enables a reduction on 2 independent variables (non-basic variables) and 34 inequalities (basic-variables, slack) to determine the 2 intervals by linear programming:

	b_i	x_{1111}	x_{2211}
$x_{1.11}$	32	-1	0
$x_{1.12}$	56	1	0
$x_{3.12}$	24	-1	-1
x_{2121}	4	0	-1
$x_{2.12}$	54	0	1
$x_{2.1.}$	98	0	1
$x_{211.}$	52	0	1
x_{2212}	46	0	1
x_{3111}	1	1	-1
x_{3112}	11	-1	0
x_{1121}	20	1	0
$x_{3.11}$	45	1	0
x_{3211}	44	0	1
$x_{1.22}$	47	-1	0
$x_{212.}$	79	0	-1
$x_{2.2.}$	69	0	-1
x_{3121}	32	-1	1
$x_{3.2.}$	153	0	1
x_{3221}	27	0	-1
$x_{3.21}$	59	-1	0
x_{1112}	40	1	0
x_{2111}	44	0	1
$x_{311.}$	12	0	-1
$x_{3.1.}$	69	0	-1
$x_{1.21}$	59	1	0
x_{3212}	13	0	-1
x_{3122}	30	1	0
x_{1122}	30	-1	0
x_{3222}	64	0	1
x_{2222}	20	0	-1
$x_{3.22}$	94	1	1
$x_{312.}$	62	0	1
x_{2221}	44	0	1
$x_{2.22}$	21	0	-1
x_{1111}	0	-1	0
x_{2211}	0	0	-1

	b_i	x_{3111}	x_{2211}
$x_{1.11}$	33	1	-1
$x_{1.12}$	55	-1	1
$x_{3.12}$	25	1	-2
x_{2121}	4	0	-1
$x_{2.12}$	54	0	1
$x_{2.1.}$	98	0	1
$x_{211.}$	52	0	1
x_{2212}	46	0	1
x_{1111}	1	1	-1
x_{3112}	12	1	-1
x_{1121}	19	-1	1
$x_{3.11}$	44	-1	1
x_{3211}	44	0	1
$x_{1.22}$	48	1	-1
$x_{212.}$	79	0	-1
$x_{2.2.}$	69	0	-1
x_{3121}	33	1	0
$x_{3.2.}$	153	0	1
x_{3221}	27	0	-1
$x_{3.21}$	60	1	-1
x_{1112}	39	-1	1
x_{2111}	44	0	1
$x_{311.}$	12	0	-1
$x_{3.1.}$	69	0	-1
$x_{1.21}$	58	-1	1
x_{3212}	13	0	-1
x_{3122}	29	-1	1
x_{1122}	31	1	-1
x_{3222}	64	0	1
x_{2222}	20	0	-1
$x_{3.22}$	93	-1	2
$x_{312.}$	62	0	1
x_{2221}	44	0	1
$x_{2.22}$	21	0	-1
x_{1111}	1	1	-1
x_{2211}	0	0	-1

	b_i	x_{3111}	x_{1121}
$x_{1.11}$	52	0	1
$x_{1.12}$	36	0	-1
$x_{3.12}$	63	-1	2
x_{2121}	23	-1	1
$x_{2.12}$	35	1	-1
$x_{2.1.}$	79	1	-1
$x_{211.}$	33	1	-1
x_{2212}	27	1	-1
x_{1111}	20	0	1
x_{3112}	31	0	1
x_{2211}	19	-1	1
$x_{3.11}$	25	0	-1
x_{3211}	25	1	-1
$x_{1.22}$	67	0	1
$x_{212.}$	98	-1	1
$x_{2.2.}$	88	-1	1
x_{3121}	33	1	0
$x_{3.2.}$	134	1	-1
x_{3221}	46	-1	1
$x_{3.21}$	79	0	1
x_{1112}	20	0	-1
x_{2111}	25	1	-1
$x_{311.}$	31	-1	1
$x_{3.1.}$	88	-1	1
$x_{1.21}$	39	0	-1
x_{3212}	32	-1	1
x_{3122}	10	0	-1
x_{1122}	50	0	1
x_{3222}	45	1	-1
x_{2222}	39	-1	1
$x_{3.22}$	55	1	-2
$x_{312.}$	43	1	-1
x_{2221}	25	1	-1
$x_{2.22}$	40	-1	1
x_{1111}	20	0	1
x_{2211}	19	-1	1

For instance we receive the interval $0 \le x_{1111} \le 20$ from the minimum and the maximum solution. In similar way we must solve the problems of the variable x_{2211}, and we get the other 34 intervals by inserting of these solutions into the first table.

We assume that the number of independent variables in the first solution is corresponding with the number of primary suppressed cells.