Tabular data protection

Matthias Templ

2019

Content

Examples

- **Example 1: (Location = Municipality x; interst in soccer x sex) :**

	men	women	toal
soccer fan	12	10	**22**
not a soccer fan	93	85	**178**
total	**105**	**95**	**200**

Examples

- **Example 1: (Location = Municipality x; interst in soccer x sex)** :

	men	women	toal
soccer fan	12	10	**22**
not a soccer fan	93	85	**178**
total	**105**	**95**	**200**

- **Example 1 (cont.): (Location = Municipality x; team preference x sex)** :

	men	women	total
Sk Rapid Wien	12	4	**16**
Sturm Graz	0	3	**3**
SV Ried	0	3	**3**
total	**12**	**10**	**22**

- **Example 2: (Location = Municipality x; education x sex)** :

	men	women	total
primary	49	53	**102**
apprenticeship	34	23	**56**
secondary	22	14	**37**
university	0	5	**5**
total	**105**	**95**	**200**

Examples

- **Example 1: (Location = Municipality x; interst in soccer x sex)** :

	men	women	toal
soccer fan	12	10	**22**
not a soccer fan	93	85	**178**
total	**105**	**95**	**200**

- **Example 1 (cont.): (Location = Municipality x; team preference x sex)** :

	men	women	total
Sk Rapid Wien	12	4	**16**
Sturm Graz	0	3	**3**
SV Ried	0	3	**3**
total	**12**	**10**	**22**

- **Example 2: (Location = Municipality x; education x sex)** :

	men	women	total
primary	49	53	**102**
apprenticeship	34	23	**56**
secondary	22	14	**37**
university	0	5	**5**
total	**105**	**95**	**200**

tables - intro

- tabular data: basis are **Micro data**

tables - intro

- tabular data: basis are **Micro data**
- **magnitude tables** and **frequency tables**

tables - intro

- tabular data: basis are **Micro data**
- **magnitude tables** and **frequency tables**
    - **frequency table:** Counts of categories

tables - intro

- tabular data: basis are **Micro data**
- **magnitude tables** and **frequency tables**
  - **frequency table:** Counts of categories
  - **magnitude table:** sum of all values of a variable.

tables - intro

- tabular data: basis are **Micro data**
- **magnitude tables** and **frequency tables**
  - **frequency table:** Counts of categories
  - **magnitude table:** sum of all values of a variable.
- **Generally...**
  - tabular data have **linear dependencies** between its cells.

tables - intro

- tabular data: basis are **Micro data**
- **magnitude tables** and **frequency tables**
  - **frequency table:** Counts of categories
  - **magnitude table:** sum of all values of a variable.
- **Generally...**
  - tabular data have **linear dependencies** between its cells.
  - tabular data can be **one- or multi-dimensional**, **hierarchical** and/or **linked**.

View on a table

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	h_1	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	h_1	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	y_1	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	h_1	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	y_1	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	h_1	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	y_1	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	h_1	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	y_1	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	4	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	y_1	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

ID	DIM1	DIM2	VALUE
1	I	A	5
2	I	A	7
3	I	A	4
4	I	A	4
5	I	B	13
6	I	B	5
.	.	.	.
.	.	.	.

H	A	B	C	Total
I	4	h_2	h_3	$\mathbf{h_4}$
II	h_5	h_6	h_7	$\mathbf{h_8}$
III	h_9	h_{10}	h_{11}	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	20	y_2	y_3	$\mathbf{y_4}$
II	y_5	y_6	y_7	$\mathbf{y_8}$
III	y_9	y_{10}	y_{11}	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

- proceed ...

H	A	B	C	Total
I	4	6	3	$\mathbf{h_4}$
II	2	5	7	$\mathbf{h_8}$
III	4	5	3	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

H	A	B	C	Total
I	20	50	10	$\mathbf{y_4}$
II	8	19	22	$\mathbf{y_8}$
III	17	32	12	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

From microdata to tables

- proceed ...

H	A	B	C	Total
I	4	6	3	$\mathbf{h_4}$
II	2	5	7	$\mathbf{h_8}$
III	4	5	3	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

H	A	B	C	Total
I	20	50	10	$\mathbf{y_4}$
II	8	19	22	$\mathbf{y_8}$
III	17	32	12	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

- common wording: **marginal totals**

From microdata to tables

- proceed ...

H	A	B	C	Total
I	4	6	3	$\mathbf{h_4}$
II	2	5	7	$\mathbf{h_8}$
III	4	5	3	$\mathbf{h_{12}}$
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

H	A	B	C	Total
I	20	50	10	$\mathbf{y_4}$
II	8	19	22	$\mathbf{y_8}$
III	17	32	12	$\mathbf{y_{12}}$
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

- common wording: **marginal totals**
- im **2-dimensional case**: **row-** and **column sums**.

From microdata to tables

H	A	B	C	Total
I	4	6	3	h_4
II	2	5	7	h_8
III	4	5	3	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

W	A	B	C	Total
I	20	50	10	y_4
II	8	19	22	y_8
III	17	32	12	y_{12}
Total	y_{13}	y_{14}	y_{15}	y_{16}

From microdata to tables

H	A	B	C	Total
I	4	6	3	h_4
II	2	5	7	h_8
III	4	5	3	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

W	A	B	C	Total
I	20	50	10	y_4
II	8	19	22	y_8
III	17	32	12	y_{12}
Total	y_{13}	y_{14}	y_{15}	y_{16}

$$\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12}
\end{aligned}$$

$$\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}
\end{aligned}$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

$$\begin{aligned}\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13}\\ \mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14}\\ \mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12}\end{aligned}$$

$$\begin{aligned}\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80}\\ \mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49}\\ \mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}\end{aligned}$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	$\mathbf{h_{13}}$	$\mathbf{h_{14}}$	$\mathbf{h_{15}}$	$\mathbf{h_{16}}$

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	$\mathbf{y_{13}}$	$\mathbf{y_{14}}$	$\mathbf{y_{15}}$	$\mathbf{y_{16}}$

$$\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12}
\end{aligned}$$

$$\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}
\end{aligned}$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	h_{13}	h_{14}	h_{15}	h_{16}

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	y_{13}	y_{14}	y_{15}	y_{16}

$$\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h_{13}} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h_{14}} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h_{15}} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13}
\end{aligned}$$

$$\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y_{13}} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y_{14}} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y_{15}} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44}
\end{aligned}$$

From microdata to tables

H	**A**	**B**	**C**	**Total**
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	$\mathbf{h_{16}}$

W	**A**	**B**	**C**	**Total**
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	$\mathbf{y_{16}}$

$$
\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h_{13}} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h_{14}} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h_{15}} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y_{13}} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y_{14}} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y_{15}} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44}
\end{aligned}
$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	10	16	13	h_{16}

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	45	101	44	y_{16}

$$\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h_{13}} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h_{14}} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h_{15}} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13}
\end{aligned}$$

$$\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y_{13}} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y_{14}} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y_{15}} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44}
\end{aligned}$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	$\mathbf{h}_{16}$

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	$\mathbf{y}_{16}$

$$
\begin{aligned}
\mathbf{h}_4 &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h}_8 &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h}_{12} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h}_{13} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h}_{14} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h}_{15} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13} \\
\mathbf{h}_{16} &= \mathbf{h}_4(13) + \mathbf{h}_8(14) + \mathbf{h}_{12}(12) = \mathbf{39} \\
\mathbf{h}_{16} &= \mathbf{h}_{13}(10) + \mathbf{h}_{14}(16) + \mathbf{h}_{15}(13) = \mathbf{39}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y}_4 &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y}_8 &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y}_{12} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y}_{13} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y}_{14} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y}_{15} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44} \\
\mathbf{y}_{16} &= \mathbf{y}_4(80) + \mathbf{y}_8(49) + \mathbf{y}_{12}(61) = \mathbf{190} \\
\mathbf{y}_{16} &= \mathbf{y}_{13}(45) + \mathbf{y}_{14}(101) + \mathbf{y}_{15}(44) = \mathbf{190}
\end{aligned}
$$

From microdata to tables

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

$$
\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h_{13}} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h_{14}} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h_{15}} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13} \\
\mathbf{h_{16}} &= \mathbf{h_4}(13) + \mathbf{h_8}(14) + \mathbf{h_{12}}(12) = \mathbf{39} \\
\mathbf{h_{16}} &= \mathbf{h_{13}}(10) + \mathbf{h_{14}}(16) + \mathbf{h_{15}}(13) = \mathbf{39}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y_{13}} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y_{14}} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y_{15}} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44} \\
\mathbf{y_{16}} &= \mathbf{y_4}(80) + \mathbf{y_8}(49) + \mathbf{y_{12}}(61) = \mathbf{190} \\
\mathbf{y_{16}} &= \mathbf{y_{13}}(45) + \mathbf{y_{14}}(101) + \mathbf{y_{15}}(44) = \mathbf{190}
\end{aligned}
$$

From microdata to tables

H	**A**	**B**	**C**	**Total**
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

W	**A**	**B**	**C**	**Total**
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

$$\begin{aligned}
\mathbf{h_4} &= h_1(4) + h_2(6) + h_3(3) = \mathbf{13} \\
\mathbf{h_8} &= h_5(2) + h_6(5) + h_7(7) = \mathbf{14} \\
\mathbf{h_{12}} &= h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12} \\
\mathbf{h_{13}} &= h_1(4) + h_5(2) + h_9(4) = \mathbf{10} \\
\mathbf{h_{14}} &= h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16} \\
\mathbf{h_{15}} &= h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13} \\
\mathbf{h_{16}} &= \mathbf{h_4}(13) + \mathbf{h_8}(14) + \mathbf{h_{12}}(12) = \mathbf{39} \\
\mathbf{h_{16}} &= \mathbf{h_{13}}(10) + \mathbf{h_{14}}(16) + \mathbf{h_{15}}(13) = \mathbf{39}
\end{aligned}$$

$$\begin{aligned}
\mathbf{y_4} &= y_1(20) + y_2(50) + y_3(10) = \mathbf{80} \\
\mathbf{y_8} &= y_5(8) + y_6(19) + y_7(22) = \mathbf{49} \\
\mathbf{y_{12}} &= y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61} \\
\mathbf{y_{13}} &= y_1(20) + y_5(8) + y_9(17) = \mathbf{45} \\
\mathbf{y_{14}} &= y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101} \\
\mathbf{y_{15}} &= y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44} \\
\mathbf{y_{16}} &= \mathbf{y_4}(80) + \mathbf{y_8}(49) + \mathbf{y_{12}}(61) = \mathbf{190} \\
\mathbf{y_{16}} &= \mathbf{y_{13}}(45) + \mathbf{y_{14}}(101) + \mathbf{y_{15}}(44) = \mathbf{190}
\end{aligned}$$

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear constraints of the form: $Ma = b$

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \dots, a_n]$
  - linear constraints of the form: $Ma = b$
- **Remarks:**
  - M is a matrix with $M_{ij} \in \{-1, 0, 1\}$ and b is a vector containing 0

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear constraints of the form: $Ma = b$
- **Remarks:**
  - M is a matrix with $M_{ij} \in \{-1, 0, 1\}$ and b is a vector containing 0

$$M = \begin{pmatrix} 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear constraints of the form: $Ma = b$
- **Remarks:**
  - M is a matrix with $M_{ij} \in \{-1, 0, 1\}$ and b is a vector containing 0

$$M = \begin{pmatrix}
1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{pmatrix}$$

  - Each row of $M\ a = b$ referes to a constraint of a row or column sum.

Formalization

- **Gerneralisation:** A (mulit-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear constraints of the form: $Ma = b$
- **Remarks:**
  - M is a matrix with $M_{ij} \in \{-1, 0, 1\}$ and b is a vector containing 0

$$M = \begin{pmatrix} 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}$$

  - Each row of $M\ a = b$ referes to a constraint of a row or column sum.
  - the cells of a table are determined by its (column) index: $i = 1, \ldots, n$

Primary suppressions

- Some rules for determining if a cell is "unsafe":

Primary suppressions

- Some rules for determining if a cell is "unsafe":
  - **Frequency rule:**
    Count of observations contributing to a cell. Unsafe if the count $<$ a threshold k (mostly k is 3 or 4)

Primary suppressions

- Some rules for determining if a cell is "unsafe":
  - **Frequency rule:**
    Count of observations contributing to a cell. Unsafe if the count $<$ a threshold k (mostly k is 3 or 4)
  - **(n,k)-dominance rule:**
    A cell must be protected if the total of n largest of contributers to a cell is larger than $k\%$ of the total cell value.

Primary suppressions

- Some rules for determining if a cell is "unsafe":
  - **Frequency rule:**
    Count of observations contributing to a cell. Unsafe if the count $<$ a threshold k (mostly k is 3 or 4)
  - **(n,k)-dominance rule:**
    A cell must be protected if the total of n largest of contributers to a cell is larger than k% of the total cell value.
  - **p-% rule:**
    total minus the sum of the two largest contributers is smaller than p% of the largest contributor. (the largest contributor is again dominant)

The later two rules are similar (but not the same). We will not go into details here.

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \Longleftrightarrow 900 > 800$

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- which p equals the (2,80)-dominance rule?

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \Longleftrightarrow 900 > 800$
- which p equals the (2,80)-dominance rule?
  $\longrightarrow p = 100 \cdot \frac{100-80}{80} = 25$

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \Longleftrightarrow 900 > 800$
- which p equals the (2,80)-dominance rule?
  $\longrightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- let p=25. Is the cell unsafe when we apply the 25%-rule?

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- which p equals the (2,80)-dominance rule?
  $\longrightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- let p=25. Is the cell unsafe when we apply the 25%-rule?
  $\longrightarrow$ Yes. $T - B_1 - B_2 < \frac{25}{100} \cdot B_1 \iff 100 < 125$

Identification of unsafe cells (examples)

- **Example:** The total T of one cell in a table is 1000.
  The value of the larges contributor is $B_1 = 500$.
  The value of the second largest contributor is $B_2 = 400$.
- Is the cell safe for the (2,80)-dominance rule?
  $\longrightarrow$ Yes. $B_1 + B_2 > \frac{80}{100} \cdot T \Longleftrightarrow 900 > 800$
- which p equals the (2,80)-dominance rule?
  $\longrightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- let p=25. Is the cell unsafe when we apply the 25%-rule?
  $\longrightarrow$ Yes. $T - B_1 - B_2 < \frac{25}{100} \cdot B_1 \Longleftrightarrow 100 < 125$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear contraints of the form: $My = b$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear contraints of the form: $My = b$
  - upper and lower bounds for each cell value expressing the knowledge of an intruder: $lb_i \leq a_i \leq ub_i$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear contraints of the form: $My = b$
  - upper and lower bounds for each cell value expressing the knowledge of an intruder: $lb_i \leq a_i \leq ub_i$
  - A cell in a table is determined by its index: $i = 1, \ldots, n$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \dots, a_n]$
  - linear contraints of the form: $My = b$
  - upper and lower bounds for each cell value expressing the knowledge of an intruder: $lb_i \leq a_i \leq ub_i$
  - A cell in a table is determined by its index: $i = 1, \dots, n$
- **Additionally:**
  - given p primary suppressions: $PS = \{i_1, \dots, i_p\}$

Protection of unsafe cells

- **Repitition:** A (multi-dimensional, hierarchical) table is given by:
  - a data vector: $a = [a_1, \ldots, a_n]$
  - linear contraints of the form: $My = b$
  - upper and lower bounds for each cell value expressing the knowledge of an intruder: $lb_i \leq a_i \leq ub_i$
  - A cell in a table is determined by its index: $i = 1, \ldots, n$
- **Additionally:**
  - given p primary suppressions: $PS = \{i_1, \ldots, i_p\}$
- **Question:** How to protect primary suppressed cells?

Protection of unsafe cells

- **Example:**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

Protection of unsafe cells

- **Example:**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	22	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- Let cell II/C $(PS = \{7\})$ be unsafe and to be protected
- Different possibilites to protect this cell, e.g.:
  - **cell suppression**
  - **rounding**
  - **reporting upper and lower bounds**

Cell suppression

- **Example:**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- Most popular method
- **However:** Because of the linear dependencies in tables, it is not enough to protect the unsafe cells only.

Cell suppression

- **Example:**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- Most popular method
- **However:** Because of the linear dependencies in tables, it is not enough to protect the unsafe cells only.
- **E.g.:** $44 - 10 - 12 = 22$ and $49 - 8 - 19 = 22$.
  no protection for the primary suppressed cell.

Cell suppression

- **Example:**

W	A	B	C	Total
I	20	50	10	**80**
II	8	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- Most popular method
- **However:** Because of the linear dependencies in tables, it is not enough to protect the unsafe cells only.
- **E.g.:** $44 - 10 - 12 = 22$ and $49 - 8 - 19 = 22$.
  no protection for the primary suppressed cell.
- $\longrightarrow$ **secondary cell suppression**: suppressing additional cells.

Cell suppression

- **Example:** suppression pattern

Cell suppression

- **Example:** suppression pattern

W	A	B	C	Total
I	20	50	10	**80**
II	S	19	NA	**49**
III	S	32	S	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression

- **Example:** suppression pattern

W	A	B	C	Total
I	20	50	10	**80**
II	S	19	NA	**49**
III	S	32	S	**61**
Total	**45**	**101**	**44**	**190**

W	A	B	C	Total
I	S	50	S	**80**
II	S	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression

- **Example:** suppression pattern

W	A	B	C	Total
I	20	50	10	**80**
II	S	19	NA	**49**
III	S	32	S	**61**
Total	**45**	**101**	**44**	**190**

W	A	B	C	Total
I	S	50	S	**80**
II	S	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- When does a suppression pattern support enough protection to unsafe cells?

Cell suppression

- **Example:** suppression pattern

W	A	B	C	Total
I	20	50	10	**80**
II	S	19	NA	**49**
III	S	32	S	**61**
Total	**45**	**101**	**44**	**190**

W	A	B	C	Total
I	S	50	S	**80**
II	S	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- When does a suppression pattern support enough protection to unsafe cells?
- Is there a optimal suppression pattern?

Cell suppression

- **Example:** suppression pattern

W	A	B	C	Total
I	20	50	10	**80**
II	S	19	NA	**49**
III	S	32	S	**61**
Total	**45**	**101**	**44**	**190**

W	A	B	C	Total
I	S	50	S	**80**
II	S	19	NA	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- When does a suppression pattern support enough protection to unsafe cells?
- Is there a optimal suppression pattern?
- **Generally:** problem is NP-hard for hierarchical, multi-dimensional and linked tables

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution
- **bad news:** The optimal method is much too slow in practice.

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution
- **bad news:** The optimal method is much too slow in practice.
- Methods to find the optimal suppression pattern are based on **linear optimization**

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution
- **bad news:** The optimal method is much too slow in practice.
- Methods to find the optimal suppression pattern are based on **linear optimization**
  - Aim: find a suppression pattern that minimizes the number of suppressed cells and guarantees that the unsafe cells are protected.

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution
- **bad news:** The optimal method is much too slow in practice.
- Methods to find the optimal suppression pattern are based on **linear optimization**
  - Aim: find a suppression pattern that minimizes the number of suppressed cells and guarantees that the unsafe cells are protected.
  - Protection: A primary unsafe cell is protected if the suppressed cell value cannot be estimated well enough, i.e. the attacker can only estimate an upper and lower bound of the cell value (attacker problem). This interval must be large enough.

Cell suppression

- **good news:** there exist algorithms for obtaining the optimal solution
- **bad news:** The optimal method is much too slow in practice.
- Methods to find the optimal suppression pattern are based on **linear optimization**
  - Aim: find a suppression pattern that minimizes the number of suppressed cells and guarantees that the unsafe cells are protected.
  - Protection: A primary unsafe cell is protected if the suppressed cell value cannot be estimated well enough, i.e. the attacker can only estimate an upper and lower bound of the cell value (attacker problem). This interval must be large enough.

Cell suppression - Attacker problem

- **Example:** Attacker problem

Cell suppression - Attacker problem

- **Example:** Attacker problem
- **Attacker:** knows the suppression pattern $SUP = \{5, 7, 9, 11\}$, the protected table, $My = b; lb_i \leq y_i \leq ub_i \; \forall i \in SUP; y_i = a_i \; \forall i \notin SUP$

W	A	B	C	Total
I	20	50	10	**80**
II	y_5	19	y_7	**49**
III	y_9	32	y_{11}	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression - Attacker problem

- **Example:** Attacker problem
- **Attacker:** knows the suppression pattern $SUP = \{5, 7, 9, 11\}$, the protected table, $My = b; lb_i \leq y_i \leq ub_i\ \forall i \in SUP; y_i = a_i\ \forall i \notin SUP$

W	A	B	C	Total
I	20	50	10	**80**
II	y_5	19	y_7	**49**
III	y_9	32	y_{11}	**61**
Total	**45**	**101**	**44**	**190**

- **LP-problem:** $min/max\ y_i\ \forall i \in SUP$ unter contraints:

Cell suppression - Attacker problem

- **Example:** Attacker problem
- **Attacker:** knows the suppression pattern $SUP = \{5, 7, 9, 11\}$, the protected table, $My = b; lb_i \leq y_i \leq ub_i \ \forall i \in SUP; y_i = a_i \ \forall i \notin SUP$

W	A	B	C	Total
I	20	50	10	**80**
II	y_5	19	y_7	**49**
III	y_9	32	y_{11}	**61**
Total	**45**	**101**	**44**	**190**

- **LP-problem:** $min/max \ y_i \ \forall i \in SUP$ unter contraints:

W	A	B	C	Total
I	20	50	10	**80**
II	[0:25]	19	[5:30]	**49**
III	[0:25]	32	[4:29]	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression - Attacker problem

- **Example:** Attacker problem
- **Attacker:** knows the suppression pattern $SUP = \{5, 7, 9, 11\}$, the protected table, $My = b; lb_i \leq y_i \leq ub_i \; \forall i \in SUP; y_i = a_i \; \forall i \notin SUP$

W	A	B	C	Total
I	20	50	10	**80**
II	y_5	19	y_7	**49**
III	y_9	32	y_{11}	**61**
Total	**45**	**101**	**44**	**190**

- **LP-problem:** $min/max \; y_i \; \forall i \in SUP$ unter contraints:

W	A	B	C	Total
I	20	50	10	**80**
II	[0:25]	19	[5:30]	**49**
III	[0:25]	32	[4:29]	**61**
Total	**45**	**101**	**44**	**190**

- the primary suppressed value y_7 is estimated by $[5 : 30]$.

Cell suppression - attacker problem

- **Beispiel:** Attackers Problem

Cell suppression - attacker problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{1, 3, 5, 7\}$, die gesicherte Tabelle, $My = b; lb_i \leq y_i \leq ub_i \; \forall i \in SUP; y_i = a_i \; \forall i \notin SUP$

W	A	B	C	Total
I	y_1	50	y_3	**80**
II	y_5	19	y_7	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression - attacker problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{1, 3, 5, 7\}$, die gesicherte Tabelle, $My = b; lb_i \leq y_i \leq ub_i \; \forall i \in SUP; y_i = a_i \; \forall i \notin SUP$

W	A	B	C	Total
I	y_1	50	y_3	**80**
II	y_5	19	y_7	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- **Lp-Problem:** $min/max \; y_i \; \forall i \in SUP$ unter obigen NB:

Cell suppression - attacker problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{1, 3, 5, 7\}$, die gesicherte Tabelle, $My = b; lb_i \leq y_i \leq ub_i \; \forall i \in SUP; y_i = a_i \; \forall i \notin SUP$

W	A	B	C	Total
I	y_1	50	y_3	**80**
II	y_5	19	y_7	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- **Lp-Problem:** $min/max \; y_i \; \forall i \in SUP$ unter obigen NB:

W	A	B	C	Total
I	[0:28]	50	[2:30]	**80**
II	[0:28]	19	[2:30]	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

Cell suppression - attacker problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{1, 3, 5, 7\}$, die gesicherte Tabelle, $My = b; lb_i \leq y_i \leq ub_i\ \forall i \in SUP; y_i = a_i\ \forall i \notin SUP$

W	A	B	C	Total
I	y_1	50	y_3	**80**
II	y_5	19	y_7	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- **Lp-Problem:** $min/max\ y_i\ \forall i \in SUP$ unter obigen NB:

W	A	B	C	Total
I	[0:28]	50	[2:30]	**80**
II	[0:28]	19	[2:30]	**49**
III	17	32	12	**61**
Total	**45**	**101**	**44**	**190**

- der prim?r gesperrte Wert y_7 ist im Intervall $[2 : 30]$ gesch?tzt.

Zellunterdr?ckung

- **Bemerkung:** Zellunterdr?ckung ist in Wahrheit eine Form von Intervallpublikation.

Zellunterdr?ckung

- **Bemerkung:** Zellunterdr?ckung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu kl?ren, ab wann eine Zelle ausreichend gesch?tzt ist. Im Normalfall werden f?r ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.

Zellunterdr?ckung

- **Bemerkung:** Zellunterdr?ckung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu kl?ren, ab wann eine Zelle ausreichend gesch?tzt ist. Im Normalfall werden f?r ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen prim?r unsicheren Zellwert auf $\pm 10\%$ genau berechnen k?nnen.

Zellunterdr?ckung

- **Bemerkung:** Zellunterdr?ckung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu kl?ren, ab wann eine Zelle ausreichend gesch?tzt ist. Im Normalfall werden f?r ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen prim?r unsicheren Zellwert auf $\pm 10\%$ genau berechnen k?nnen.
- **Informationsverlust:** Zellunterdr?ckung bewirkt Informationsverlust, der durch ein optimales Unterdr?ckungsschema mittels einer Verlustfunktion minimiert werden soll.

Zellunterdr?ckung

- **Bemerkung:** Zellunterdr?ckung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu kl?ren, ab wann eine Zelle ausreichend gesch?tzt ist. Im Normalfall werden f?r ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen prim?r unsicheren Zellwert auf $\pm 10\%$ genau berechnen k?nnen.
- **Informationsverlust:** Zellunterdr?ckung bewirkt Informationsverlust, der durch ein optimales Unterdr?ckungsschema mittels einer Verlustfunktion minimiert werden soll.
- Wir zeigen das **mathematische Modell** f?r optimale Zellunterdr?ckung:

Zellunterdr?ckung - Modellannahmen

- Wir nehmen an, dass ein Angreifer f?r jeden Zellwert a_i eine untere und obere Grenze (lb_i bzw. ub_i) kennt, mit:

$$lb_i \leq a_i \leq ub_i \ \forall i = 1, \ldots, n$$

Zellunterdr?ckung - Modellannahmen

- Wir nehmen an, dass ein Angreifer f?r jeden Zellwert a_i eine untere und obere Grenze (lb_i bzw. ub_i) kennt, mit:

$$lb_i \leq a_i \leq ub_i \ \forall i = 1, \ldots, n$$

- Man definiert relative ?ussere Grenzen f?r jede Zelle:

$$\begin{aligned} LB_i &:= a_i - lb_i \geq 0 \ \forall i = 1, \ldots, n \\ UB_i &:= ub_i - a_i \geq 0 \ \forall i = 1, \ldots, n \end{aligned}$$

Zellunterdr?ckung - Modellannahmen

- Wir nehmen an, dass ein Angreifer f?r jeden Zellwert a_i eine untere und obere Grenze (lb_i bzw. ub_i) kennt, mit:

$$lb_i \leq a_i \leq ub_i \ \forall i = 1, \ldots, n$$

- Man definiert relative ?ussere Grenzen f?r jede Zelle:

$$\begin{aligned} LB_i &:= a_i - lb_i \geq 0 \ \forall i = 1, \ldots, n \\ UB_i &:= ub_i - a_i \geq 0 \ \forall i = 1, \ldots, n \end{aligned}$$

- F?r alle sensible Zellen werden untere (LPL_i) und obere (UPL_i) Protection Levels definiert, sodass f?r die vom Angreifer berechneten Intervalle gilt:

$$min(y_i) \leq a_i - LPL_i \ \forall i \in PS$$
$$max(y_i) \geq a_i + UPL_i \ \forall i \in PS$$

Zellunterdr?ckung - Modellannahmen

- Wir nehmen an, dass ein Angreifer f?r jeden Zellwert a_i eine untere und obere Grenze (lb_i bzw. ub_i) kennt, mit:

$$lb_i \leq a_i \leq ub_i \; \forall i = 1, \ldots, n$$

- Man definiert relative ?ussere Grenzen f?r jede Zelle:

$$\begin{aligned} LB_i &:= a_i - lb_i \geq 0 \; \forall i = 1, \ldots, n \\ UB_i &:= ub_i - a_i \geq 0 \; \forall i = 1, \ldots, n \end{aligned}$$

- F?r alle sensible Zellen werden untere (LPL_i) und obere (UPL_i) Protection Levels definiert, sodass f?r die vom Angreifer berechneten Intervalle gilt:

$$min(y_i) \leq a_i - LPL_i \; \forall i \in PS$$
$$max(y_i) \geq a_i + UPL_i \; \forall i \in PS$$

- Wir f?hren bin?re Variablen $x_i, \; i = 1, \ldots, n$ ein, f?r die gelten soll:

$$x_i = 0 \; \forall i \notin SUP$$
$$x_i = 1 \; \forall i \in SUP$$

Zellunterdr?ckung - Modellannahmen (2)

- Wir definieren f?r jede Zelle a_i ein Gewicht w_i, das in die Zielfunktion einflie?t, z.B:

$$\begin{aligned} w_i &= a_i \\ w_i &= 1 \\ w_i &= log(1 + a_i) \end{aligned}$$

Zellunterdr?ckung - Modellannahmen (2)

- Wir definieren f?r jede Zelle a_i ein Gewicht w_i, das in die Zielfunktion einflie?t, z.B:

$$\begin{aligned} w_i &= a_i \\ w_i &= 1 \\ w_i &= log(1 + a_i) \end{aligned}$$

- Die Zielfunktion des Optimierungsproblems ist gegeben als:

$$min \sum_{i=1}^{n} w_i \cdot x_i$$

Zellunterdr?ckung - Modellannahmen (2)

- Wir definieren f?r jede Zelle a_i ein Gewicht w_i, das in die Zielfunktion einflie?t, z.B:

$$\begin{aligned} w_i &= a_i \\ w_i &= 1 \\ w_i &= log(1 + a_i) \end{aligned}$$

- Die Zielfunktion des Optimierungsproblems ist gegeben als:

$$min \sum_{i=1}^{n} w_i \cdot x_i$$

- unter folgenden Nebenbedingungen:

$$\begin{array}{cc} Mf = b & Mg = b \\ f_i \geq a_i - LB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i \geq a_i - LB_i \cdot x_i \forall i = 1, \ldots, n \\ f_i \leq a_i + UB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i \leq a_i + UB_i \cdot x_i \forall i = 1, \ldots, n \\ f_i \leq a_i - LPL_i \ \forall i \in PS & g_i \geq a_i + UPL_i \ \forall i \in PS \end{array}$$

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

$$\begin{aligned} Mf &= b & Mg &= b & (1)\\ f_i &\geq a_i - LB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\geq a_i - LB_i \cdot x_i \forall i = 1, \ldots, n & (2)\\ f_i &\leq a_i + UB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\leq a_i + UB_i \cdot x_i \forall i = 1, \ldots, n & (3)\\ f_i &\leq a_i - LPL_i \ \forall i \in PS & g_i &\geq a_i + UPL_i \ \forall i \in PS & (4) \end{aligned}$$

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

$$
\begin{align}
Mf &= b & Mg &= b \tag{1}\\
f_i &\geq a_i - LB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\geq a_i - LB_i \cdot x_i \forall i = 1, \ldots, n \tag{2}\\
f_i &\leq a_i + UB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\leq a_i + UB_i \cdot x_i \forall i = 1, \ldots, n \tag{3}\\
f_i &\leq a_i - LPL_i \ \forall i \in PS & g_i &\geq a_i + UPL_i \ \forall i \in PS \tag{4}
\end{align}
$$

- Es werden zwei m?gliche Tabellen $f = (f_1, \ldots, f_n)$ sowie $g = (g_1, \ldots, g_n)$ gesucht.

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

$$
\begin{aligned}
Mf &= b & Mg &= b & (1)\\
f_i &\geq a_i - LB_i \cdot x_i \; \forall i = 1, \dots, n & g_i &\geq a_i - LB_i \cdot x_i \forall i = 1, \dots, n & (2)\\
f_i &\leq a_i + UB_i \cdot x_i \; \forall i = 1, \dots, n & g_i &\leq a_i + UB_i \cdot x_i \forall i = 1, \dots, n & (3)\\
f_i &\leq a_i - LPL_i \; \forall i \in PS & g_i &\geq a_i + UPL_i \; \forall i \in PS & (4)
\end{aligned}
$$

- Es werden zwei m?gliche Tabellen $f = (f_1, \dots, f_n)$ sowie $g = (g_1, \dots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass f?r f und g alle linearen Abh?ngigkeiten erf?llt sind und dass gilt:

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

$$\begin{aligned} Mf &= b & Mg &= b && (1)\\ f_i &\geq a_i - LB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\geq a_i - LB_i \cdot x_i \forall i = 1, \ldots, n && (2)\\ f_i &\leq a_i + UB_i \cdot x_i \ \forall i = 1, \ldots, n & g_i &\leq a_i + UB_i \cdot x_i \forall i = 1, \ldots, n && (3)\\ f_i &\leq a_i - LPL_i \ \forall i \in PS & g_i &\geq a_i + UPL_i \ \forall i \in PS && (4) \end{aligned}$$

- Es werden zwei m?gliche Tabellen $f = (f_1, \ldots, f_n)$ sowie $g = (g_1, \ldots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass f?r f und g alle linearen Abh?ngigkeiten erf?llt sind und dass gilt:

$$f_i = g_i = a_i \ \forall i \notin SUPP \quad (5)$$

$$lb_i \leq f_i, g_i \leq ub_i \ \forall i \in SUPP \quad (6)$$

Zellunterdr?ckung - Das Modell

- Optimiere: $min \sum_{i=1}^{n} w_i \cdot x_i$ unter

$$
\begin{aligned}
Mf &= b & Mg &= b & (1)\\
f_i &\geq a_i - LB_i \cdot x_i \; \forall i = 1, \ldots, n & g_i &\geq a_i - LB_i \cdot x_i \forall i = 1, \ldots, n & (2)\\
f_i &\leq a_i + UB_i \cdot x_i \; \forall i = 1, \ldots, n & g_i &\leq a_i + UB_i \cdot x_i \forall i = 1, \ldots, n & (3)\\
f_i &\leq a_i - LPL_i \; \forall i \in PS & g_i &\geq a_i + UPL_i \; \forall i \in PS & (4)
\end{aligned}
$$

- Es werden zwei m?gliche Tabellen $f = (f_1, \ldots, f_n)$ sowie $g = (g_1, \ldots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass f?r f und g alle linearen Abh?ngigkeiten erf?llt sind und dass gilt:

$$
\begin{aligned}
f_i = g_i &= a_i \; \forall i \notin SUPP & (5)\\
lb_i \leq f_i, g_i &\leq ub_i \; \forall i \in SUPP & (6)
\end{aligned}
$$

- die Nebenbedingungen (4) erzwingen die Einhaltung der Protection Levels f?r alle prim?r-gesperrten Zellen.

Zellunterdr?ckung - Bemerkungen zum Modell

- Das Modell liefert ein **optimales Sperrmuster** in Bezug auf die Zielfunktion.
- Aber: In der **Praxis** (in dieser Form) **nicht einsetzbar**, da die Anzahl der Hilfsvariablen (f_i, g_i, x_i) und die Anzahl der Nebenbedingungen sehr schnell, sehr gro? wird.

Zellunterdr?ckung - Bemerkungen zum Modell

- Das Modell liefert ein **optimales Sperrmuster** in Bezug auf die Zielfunktion.
- Aber: In der **Praxis** (in dieser Form) **nicht einsetzbar**, da die Anzahl der Hilfsvariablen (f_i, g_i, x_i) und die Anzahl der Nebenbedingungen sehr schnell, sehr gro? wird.
- Es ist m?glich, die Anzahl der f?r das Modell notwendigen Variablen zu verringern, indem man das **Dualit?tsprinzip** jedes lineares Problems ausn?tzt.

Zellunterdr?ckung - Bemerkungen zum Modell

- Das Modell liefert ein **optimales Sperrmuster** in Bezug auf die Zielfunktion.
- Aber: In der **Praxis** (in dieser Form) **nicht einsetzbar**, da die Anzahl der Hilfsvariablen (f_i, g_i, x_i) und die Anzahl der Nebenbedingungen sehr schnell, sehr gro? wird.
- Es ist m?glich, die Anzahl der f?r das Modell notwendigen Variablen zu verringern, indem man das **Dualit?tsprinzip** jedes lineares Problems ausn?tzt.
- Das Modell wird nur mehr mittels der bin?ren Variablen x_i formuliert.

Zellunterdr?ckung - Bemerkungen zum Modell

- Das Modell liefert ein **optimales Sperrmuster** in Bezug auf die Zielfunktion.
- Aber: In der **Praxis** (in dieser Form) **nicht einsetzbar**, da die Anzahl der Hilfsvariablen (f_i, g_i, x_i) und die Anzahl der Nebenbedingungen sehr schnell, sehr gro? wird.
- Es ist m?glich, die Anzahl der f?r das Modell notwendigen Variablen zu verringern, indem man das **Dualit?tsprinzip** jedes lineares Problems ausn?tzt.
- Das Modell wird nur mehr mittels der bin?ren Variablen x_i formuliert.
- Es werden **schrittweise zus?tzliche Nebenbedingungen** ins Modell aufgenommen, die aber nur mehr von x_i abh?ngen.

Zellunterdr?ckung - Bemerkungen zum Modell

- Das Modell liefert ein **optimales Sperrmuster** in Bezug auf die Zielfunktion.
- Aber: In der **Praxis** (in dieser Form) **nicht einsetzbar**, da die Anzahl der Hilfsvariablen (f_i, g_i, x_i) und die Anzahl der Nebenbedingungen sehr schnell, sehr gro? wird.
- Es ist m?glich, die Anzahl der f?r das Modell notwendigen Variablen zu verringern, indem man das **Dualit?tsprinzip** jedes lineares Problems ausn?tzt.
- Das Modell wird nur mehr mittels der bin?ren Variablen x_i formuliert.
- Es werden **schrittweise zus?tzliche Nebenbedingungen** ins Modell aufgenommen, die aber nur mehr von x_i abh?ngen.
- Es ergibt sich ein **iterativer Algorithmus**, in dem zwar mehr lineare Probleme gel?st werden m?ssen, die aber weniger komplex und umfangreich sind.

Zellunterdr?ckung in hierarchischen Tabellen

- Gegeben sei folgende Tabelle:

Zellunterdr?ckung in hierarchischen Tabellen

- Gegeben sei folgende Tabelle:

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	8	19	22	**49**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	9	28	5	**42**
56.12	4	7	6	**17**
56.13	27	15	9	**51**
56.1	40	50	20	**110**
56.2	2	20	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

Zellunterdr?ckung in hierarchischen Tabellen

- Sch?tzenswerte Zellen seien identifiziert und gel?scht:

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	8	19	NA	**49**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	9	28	5	**42**
56.12	NA	NA	6	**NA**
56.13	27	15	9	**51**
56.1	40	NA	20	**110**
56.2	NA	20	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

Zellunterdr?ckung in hierarchischen Tabellen

- Aufgabe: Auffinden eines g?ltigen Sperrmusters gegen exakte R?ckrechenbarkeit mit minimaler Anzahl von Sekund?rsperrungen:

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	8	19	NA	**49**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	9	28	5	**42**
56.12	NA	NA	6	**NA**
56.13	27	15	9	**51**
56.1	40	NA	20	**110**
56.2	NA	20	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

Zellunterdr?ckung in hierarchischen Tabellen

- Aufgabe: Auffinden eines g?ltigen Sperrmusters gegen exakte R?ckrechenbarkeit mit minimaler Anzahl von Sekund?rsperrungen:

	R1	R2	R3	Total
55.1	20	50	S	**S**
55.2	8	19	NA	**S**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	S	S	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**S**
56.2	NA	S	18	**S**
56.3	S	S	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

Zellunterdr?ckung in hierarchischen Tabellen

- Aufgabe: Auffinden eines g?ltigen Sperrmusters gegen exakte R?ckrechenbarkeit mit minimaler Anzahl von Sekund?rsperrungen:

	R1	R2	R3	Total
55.1	20	50	S	**S**
55.2	8	19	NA	**S**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	S	S	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**S**
56.2	NA	S	18	**S**
56.3	S	S	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

- Wir mussten insgesamt 13 zus?tzliche Zellen sperren.

Zellunterdr?ckung in hierarchischen Tabellen

- Aufgabe: Auffinden eines g?ltigen Sperrmusters gegen exakte R?ckrechenbarkeit mit minimaler Anzahl von Sekund?rsperrungen:

	R1	R2	R3	Total
55.1	20	50	S	**S**
55.2	8	19	NA	**S**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	S	S	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**S**
56.2	NA	S	18	**S**
56.3	S	S	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

- Wir mussten insgesamt 13 zus?tzliche Zellen sperren.
- Der Informationsverlust der Sekund?rsperrungen betr?gt 485.

Zellunterdr?ckung in hierarchischen Tabellen

- Aufgabe: Auffinden eines g?ltigen Sperrmusters gegen exakte R?ckrechenbarkeit mit minimaler Anzahl von Sekund?rsperrungen:

	R1	R2	R3	Total
55.1	20	50	S	**S**
55.2	8	19	NA	**S**
55.3	17	32	12	**61**
55	**45**	**101**	**44**	**190**
56.11	S	S	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**S**
56.2	NA	S	18	**S**
56.3	S	S	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

- Wir mussten insgesamt 13 zus?tzliche Zellen sperren.
- Der Informationsverlust der Sekund?rsperrungen betr?gt 485.
- Gibt es bessere/alternative Sperrmuster?

Zellunterdr?ckung in hierarchischen Tabellen

- Lösung: es gibt bessere Sperrmuster, z.B das optimale

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	S	19	NA	**49**
55.3	S	32	S	**61**
55	**45**	**101**	**44**	**190**
56.11	S	28	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**110**
56.2	NA	S	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

Zellunterdr?ckung in hierarchischen Tabellen

- Lösung: es gibt bessere Sperrmuster, z.B das optimale

	R1	R2	R3	Total
55.1	20	50	10	80
55.2	S	19	NA	49
55.3	S	32	S	61
55	45	101	44	190
56.11	S	28	5	S
56.12	NA	NA	6	S
56.13	27	15	9	51
56.1	S	NA	20	110
56.2	NA	S	18	40
56.3	20	30	25	75
56	62	100	53	225
Total	107	201	97	415

- Wir mussten insgesamt 7 zus?tzliche Zellen sperren.

Zellunterdr?ckung in hierarchischen Tabellen

- Lösung: es gibt bessere Sperrmuster, z.B das optimale

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	S	19	NA	**49**
55.3	S	32	S	**61**
55	**45**	**101**	**44**	**190**
56.11	S	28	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**110**
56.2	NA	S	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

- Wir mussten insgesamt 7 zus?tzliche Zellen sperren.
- Der Informationsverlust der Sekund?rsperrungen betr?gt 148.

Zellunterdr?ckung in hierarchischen Tabellen

- Lösung: es gibt bessere Sperrmuster, z.B das optimale

	R1	R2	R3	Total
55.1	20	50	10	**80**
55.2	S	19	NA	**49**
55.3	S	32	S	**61**
55	**45**	**101**	**44**	**190**
56.11	S	28	5	**S**
56.12	NA	NA	6	**S**
56.13	27	15	9	**51**
56.1	S	NA	20	**110**
56.2	NA	S	18	**40**
56.3	20	30	25	**75**
56	**62**	**100**	**53**	**225**
Total	**107**	**201**	**97**	**415**

- Wir mussten insgesamt 7 zus?tzliche Zellen sperren.
- Der Informationsverlust der Sekund?rsperrungen betr?gt 148.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekund?r) gesperrt, muss in allen Tabellen die R?ckrechenbarkeit ?berpr?ft werden.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekund?r) gesperrt, muss in allen Tabellen die R?ckrechenbarkeit ?berpr?ft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und f?hrt oft zu langen Rechenzeiten.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekund?r) gesperrt, muss in allen Tabellen die R?ckrechenbarkeit ?berpr?ft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und f?hrt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale L?sungen liefern.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekund?r) gesperrt, muss in allen Tabellen die R?ckrechenbarkeit ?berpr?ft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und f?hrt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale L?sungen liefern.
    - HITAS: Umwandlung von hierarchisch gegliederten Tabellen in einfache, 2-dimensionale Tabellen. Zellsperrung in 2-dimensionalen Tabellen nach einer gewissen Reihenfolge.

Zellunterdr?ckung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abh?ngigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekund?r) gesperrt, muss in allen Tabellen die R?ckrechenbarkeit ?berpr?ft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und f?hrt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale L?sungen liefern.
  - HITAS: Umwandlung von hierarchisch gegliederten Tabellen in einfache, 2-dimensionale Tabellen. Zellsperrung in 2-dimensionalen Tabellen nach einer gewissen Reihenfolge.
  - Quaderverfahren: Algorithmus zum Aufsuchen "geometrischer" Sperrmuster

Runden

- **Runden** ist eine Alternative zur Zellunterdr?ckung.

Runden

- **Runden** ist eine Alternative zur Zellunterdr?ckung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:

Runden

- **Runden** ist eine Alternative zur Zellunterdr?ckung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
  - normales Runden:
  - zuf?lliges Runden
  - kontrolliertes Runden

Runden

- **Runden** ist eine Alternative zur Zellunterdr?ckung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
  - normales Runden:
  - zuf?lliges Runden
  - kontrolliertes Runden
- Bei allen Varianten muss eine **Rundungsbasis** (oft 3 oder 5) gew?hlt werden.

Runden

- **Runden** ist eine Alternative zur Zellunterdr?ckung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
  - normales Runden:
  - zuf?lliges Runden
  - kontrolliertes Runden
- Bei allen Varianten muss eine **Rundungsbasis** (oft 3 oder 5) gew?hlt werden.
- **normales Runden** (Runden des Zellwertes zum n?chsten Vielfachen der Basis) bringt etwas Schutz, aber nicht genug
  $\rightarrow$ wir vernachl?ssigen diese Variante.

Zuf?lliges Runden

- **Idee:** jeder Zellwert wird in einer zuf?lligen Art und Weise - unabh?ngig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.

Zuf?lliges Runden

- **Idee:** jeder Zellwert wird in einer zuf?lligen Art und Weise - unabh?ngig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht ver?ndert.

Zuf?lliges Runden

- **Idee:** jeder Zellwert wird in einer zuf?lligen Art und Weise - unabh?ngig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht ver?ndert.
- **Randsummen** werden ?blicherweise getrennt von den inneren Tabellenzellen behandelt.

Zuf?lliges Runden

- **Idee:** jeder Zellwert wird in einer zuf?lligen Art und Weise - unabh?ngig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht ver?ndert.
- **Randsummen** werden ?blicherweise getrennt von den inneren Tabellenzellen behandelt.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind m?glich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.

Zuf?lliges Runden

- **Idee:** jeder Zellwert wird in einer zuf?lligen Art und Weise - unabh?ngig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht ver?ndert.
- **Randsummen** werden ?blicherweise getrennt von den inneren Tabellenzellen behandelt.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind m?glich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.
- **Nachteil:** Tabellen sind m?glicherweise (wahrscheinlich) nicht mehr additiv.

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	10	16	13	39

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	10	16	13	39

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

H	A	B	C	Total
I	1	0	0	1
II	1	2	1	2
III	1	2	0	0
Total	1	1	1	0

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

H	A	B	C	Total
I	1	0	0	**1**
II	1	2	1	**2**
III	1	2	0	**0**
Total	**1**	**1**	**1**	**0**

- **Gewichtungsschema:** Wir w?hlen folgendes Gewichtungsschema:

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

H	A	B	C	Total
I	1	0	0	**1**
II	1	2	1	**2**
III	1	2	0	**0**
Total	**1**	**1**	**1**	**0**

- **Gewichtungsschema:** Wir w?hlen folgendes Gewichtungsschema:
  - Divisionsrest = 0: Tabellenwert wird nicht ver?ndert.

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

H	A	B	C	Total
I	1	0	0	**1**
II	1	2	1	**2**
III	1	2	0	**0**
Total	**1**	**1**	**1**	**0**

- **Gewichtungsschema:** Wir w?hlen folgendes Gewichtungsschema:
  - Divisionsrest $= 0$: Tabellenwert wird nicht ver?ndert.
  - Divisionsrest $= 1$: mit Wahrscheinlichkeit $\frac{1}{3}$ wird aufgerundet, mit Wahrscheinlichkeit $\frac{2}{3}$ abgerundet.

Zuf?lliges Runden - Beispiel

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- **Basis:** Wir w?hlen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

H	A	B	C	Total
I	1	0	0	**1**
II	1	2	1	**2**
III	1	2	0	**0**
Total	**1**	**1**	**1**	**0**

- **Gewichtungsschema:** Wir w?hlen folgendes Gewichtungsschema:
  - Divisionsrest $= 0$: Tabellenwert wird nicht ver?ndert.
  - Divisionsrest $= 1$: mit Wahrscheinlichkeit $\frac{1}{3}$ wird aufgerundet, mit Wahrscheinlichkeit $\frac{2}{3}$ abgerundet.
  - Divisionsrest $= 2$: mit Wahrscheinlichkeit $\frac{2}{3}$ wird aufgerundet, mit Wahrscheinlichkeit $\frac{1}{3}$ abgerundet.

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	**15**
II	3	3	6	**12**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	**15**
II	3	3	6	**12**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

- Additivit?t in Spalten 1 und 3 stimmt nicht mehr.

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	**15**
II	3	3	6	**12**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

- Additivit?t in Spalten 1 und 3 stimmt nicht mehr.
- andere M?glichkeit:

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	15
II	3	3	6	12
III	3	6	3	12
Total	9	15	15	39

- Additivit?t in Spalten 1 und 3 stimmt nicht mehr.
- andere M?glichkeit:

H	A	B	C	Total
I	3	6	3	15
II	0	6	6	15
III	3	3	3	12
Total	12	15	15	39

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	15
II	3	3	6	12
III	3	6	3	12
Total	9	15	15	39

- Additivit?t in Spalten 1 und 3 stimmt nicht mehr.
- andere M?glichkeit:

H	A	B	C	Total
I	3	6	3	15
II	0	6	6	15
III	3	3	3	12
Total	12	15	15	39

- z.B: Additivit?t in Spalte 1,3 und 4 sowie Zeile 1-4 stimmt nicht mehr.

Zuf?lliges Runden - Beispiel

- Es ergibt sich z.B folgende Tabelle:

H	A	B	C	Total
I	6	6	3	**15**
II	3	3	6	**12**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

- Additivit?t in Spalten 1 und 3 stimmt nicht mehr.
- andere M?glichkeit:

H	A	B	C	Total
I	3	6	3	**15**
II	0	6	6	**15**
III	3	3	3	**12**
Total	**12**	**15**	**15**	**39**

- z.B: Additivit?t in Spalte 1,3 und 4 sowie Zeile 1-4 stimmt nicht mehr.
- **Achtung:** Problem wenn die gleiche Zelle in verlinkten Tabellen unterschiedlich gerundet wird.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivit?t der Tabelle gewahrt bleibt.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivit?t der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grunds?tzlich) nicht ver?ndert.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivit?t der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grunds?tzlich) nicht ver?ndert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind m?glich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivit?t der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grunds?tzlich) nicht ver?ndert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind m?glich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.
- **Vorteil:** Tabellen sind additiv.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivit?t der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grunds?tzlich) nicht ver?ndert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind m?glich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.
- **Vorteil:** Tabellen sind additiv.
- **Nachteil:** bei kontrolliertem Runden handelt es sich um ein (komplexes) lineares Problem das m?glicherweise unl?sbar ist.

Kontrolliertes Runden - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

Kontrolliertes Runden - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- Tabelle **nach kontrolliertem Runden:**

H	A	B	C	Total
I	3	6	3	**12**
II	3	3	9	**15**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

Kontrolliertes Runden - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	4	6	3	**13**
II	2	5	7	**14**
III	4	5	3	**12**
Total	**10**	**16**	**13**	**39**

- Tabelle **nach kontrolliertem Runden:**

H	A	B	C	Total
I	3	6	3	**12**
II	3	3	9	**15**
III	3	6	3	**12**
Total	**9**	**15**	**15**	**39**

- Alle Randsummen stimmen, die Tabelle ist additiv.

Zellanpassung - CTA

- **Idee:**
  - 1) jeder prim?r gesperrte Zellwert wird durch einen "sicheren" Wert am oberen oder unteren Rand des vorgegebenen Sicherheitsintervals ersetzt.
  - 2) Alle anderen Zellen werden so adjustiert, dass wiederum eine additive Tabelle entsteht.

Zellanpassung - CTA

- **Idee:**
  - 1) jeder prim?r gesperrte Zellwert wird durch einen "sicheren" Wert am oberen oder unteren Rand des vorgegebenen Sicherheitsintervals ersetzt.
  - 2) Alle anderen Zellen werden so adjustiert, dass wiederum eine additive Tabelle entsteht.
- **Vorteil:** es entstehen im Gegensatz zu Zellunterdr?ckung keine l?ckenhaften Tabellen, ausserdem sind die Anpassungen der nicht prim?r sensiblen Zellen meist gering.

Zellanpassung - CTA

- **Idee:**
  - 1) jeder prim?r gesperrte Zellwert wird durch einen "sicheren" Wert am oberen oder unteren Rand des vorgegebenen Sicherheitsintervals ersetzt.
  - 2) Alle anderen Zellen werden so adjustiert, dass wiederum eine additive Tabelle entsteht.
- **Vorteil:** es entstehen im Gegensatz zu Zellunterdr?ckung keine l?ckenhaften Tabellen, ausserdem sind die Anpassungen der nicht prim?r sensiblen Zellen meist gering.
- **weiterer Vorteil:** optimale Algorithmen existieren.

Zellanpassung - CTA

- **Idee:**
  - 1) jeder prim?r gesperrte Zellwert wird durch einen "sicheren" Wert am oberen oder unteren Rand des vorgegebenen Sicherheitsintervals ersetzt.
  - 2) Alle anderen Zellen werden so adjustiert, dass wiederum eine additive Tabelle entsteht.
- **Vorteil:** es entstehen im Gegensatz zu Zellunterdr?ckung keine l?ckenhaften Tabellen, ausserdem sind die Anpassungen der nicht prim?r sensiblen Zellen meist gering.
- **weiterer Vorteil:** optimale Algorithmen existieren.
- **Nachteile:** optimale Algorithmen nur brauchbar f?r sehr kleine Tabellen, Heuristiken existieren, garantieren aber keine L?sung.

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Fixieren der Werte** f?r die sensitiven Zellen

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Fixieren der Werte** f?r die sensitiven Zellen

H	A	B	C	Total
I		0*		
II				
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*		
II				
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	
II				
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II				
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71			
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51		
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*		
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total				

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total	**146**			

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total	**146**	**80***		

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total	**146**	**80***	**151**	

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- **Anpassung der** nicht sensitiven Zellen

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total	**146**	**80***	**151**	**377***

Zellanpassung (CTA) - Beispiel

- **urspr?ngliche Tabelle:**

H	A	B	C	Total
I	74	17 [0:37]	85	**176**
II	71	51	30	**152**
III	1[0,21]	9[0,29]	36	**46**
Total	**146**	**77**	**151**	**374**

- Tabelle **nach Zellanpassung:**

H	A	B	C	Total
I	75*	0*	85	**160***
II	71	51	30	**152**
III	0*	29*	36	**65***
Total	**146**	**80***	**151**	**377***

- **Implementation** ist wiederum basierend auf linearer Optimierung (komplexes Formelwerk).

Zellanpassung (ABS)

- die vom **ABS** (Australian Bureau of Statistics) entwickelte Methode ist ein Spezialfall der **Tabellenverschmutzung**

Zellanpassung (ABS)

- die vom **ABS** (Australian Bureau of Statistics) entwickelte Methode ist ein Spezialfall der **Tabellenverschmutzung**
- **Idee:** Konsistente, aber zuf?llige Verschmutzung von Tabellenzellen basierend auf
  - Record-Keys
  - Cell-Keys
  - LookUp-Tabelle

Zellanpassung (ABS)

- die vom **ABS** (Australian Bureau of Statistics) entwickelte Methode ist ein Spezialfall der **Tabellenverschmutzung**
- **Idee:** Konsistente, aber zuf?llige Verschmutzung von Tabellenzellen basierend auf
  - Record-Keys
  - Cell-Keys
  - LookUp-Tabelle
- **Vorteil:** Konsistenz

Zellanpassung (ABS)

- die vom **ABS** (Australian Bureau of Statistics) entwickelte Methode ist ein Spezialfall der **Tabellenverschmutzung**
- **Idee:** Konsistente, aber zuf?llige Verschmutzung von Tabellenzellen basierend auf
  - Record-Keys
  - Cell-Keys
  - LookUp-Tabelle
- **Vorteil:** Konsistenz
- **Nachteil:** fehlende Tabellenadditivit?t

Zellanpassung (ABS)

- die vom **ABS** (Australian Bureau of Statistics) entwickelte Methode ist ein Spezialfall der **Tabellenverschmutzung**
- **Idee:** Konsistente, aber zuf?llige Verschmutzung von Tabellenzellen basierend auf
  - Record-Keys
  - Cell-Keys
  - LookUp-Tabelle
- **Vorteil:** Konsistenz
- **Nachteil:** fehlende Tabellenadditivit?t

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche
- `sdcTable`:
  - freie Implementierung von Algorithmen (vor allem) der Sekundärsperrung in `R`

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche
- `sdcTable`:
  - freie Implementierung von Algorithmen (vor allem) der Sekundärsperrung in `R`
  - Code ist frei verfügbar, modifizierbar,...

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche
- `sdcTable`:
  - freie Implementierung von Algorithmen (vor allem) der Sekundärsperrung in `R`
  - Code ist frei verfügbar, modifizierbar,...
  - kleinerer Funktionsumfang ohne graphischer Oberfläche

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche
- `sdcTable`:
  - freie Implementierung von Algorithmen (vor allem) der Sekundärsperrung in `R`
  - Code ist frei verfügbar, modifizierbar,...
  - kleinerer Funktionsumfang ohne graphischer Oberfläche
  - gut adaptierbar, flexibel anpassbar an aktuelle Problemstellungen

Software zur Geheimhaltung von Tabellen

Software zur Geheimhaltung tabellarischer Daten:

- τ-`Argus`:
  - τ-Argus ist das Produkt eines Europäischen Research-Projektes im Bereich Statistischer Geheimhaltung
  - Code wurde erst nach $\approx$ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
  - großer Funktionsumfang mit graphischer Oberfläche
- `sdcTable`:
  - freie Implementierung von Algorithmen (vor allem) der Sekundärsperrung in `R`
  - Code ist frei verfügbar, modifizierbar,...
  - kleinerer Funktionsumfang ohne graphischer Oberfläche
  - gut adaptierbar, flexibel anpassbar an aktuelle Problemstellungen
  - wird von der Methodik entwickelt und gewartet.

Weitere Informationen

?ber statistische Geheimhaltung finden Sie etwa bei:

- Allgemeines ?ber SDC: `http://neon.vb.cbs.nl/casc`
- Handbuch: `http://neon.vb.cbs.nl/casc/SDC_Handbook.pdf`
- R-Paket zur Geheimhaltung (hierarchischer) Tabellen: `http://cran.r-project.org/web/packages/sdcTable`
- oder nat?rlich auch unter der Klappe 7988 (fast immer)

Conclusio

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.

Conclusio

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.
- Varianten f?r Geheimhaltung von Tabellen sind:

Conclusio

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.
- Varianten f?r Geheimhaltung von Tabellen sind:
  - Zellunterdr?ckung
  - Runden
  - Zellanpassung

Conclusio

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.
- Varianten f?r Geheimhaltung von Tabellen sind:
  - Zellunterdr?ckung
  - Runden
  - Zellanpassung
- alle vorgestellten Methoden besitzen sowohl f?r den Anwender als auch f?r den Datenproduzenten Vor- und Nachteile.

Conclusio

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.
- Varianten f?r Geheimhaltung von Tabellen sind:
  - Zellunterdr?ckung
  - Runden
  - Zellanpassung
- alle vorgestellten Methoden besitzen sowohl f?r den Anwender als auch f?r den Datenproduzenten Vor- und Nachteile.
- noch keine (brauchbaren) und flexiblen L?sungen f?r komplexe, verlinkte Tabellen vorhanden.