Tabular data protection

Matthias Templ

2019

Content

Examples

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

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

Examples

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

	men	women	total
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; interest in soccer x sex)** :

	men	women	total
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**
- **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	$\mathbf{h}_4$
II	2	5	7	$\mathbf{h}_8$
III	4	5	3	$\mathbf{h}_{12}$
Total	$\mathbf{h}_{13}$	$\mathbf{h}_{14}$	$\mathbf{h}_{15}$	$\mathbf{h}_{16}$

W	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	$\mathbf{y}_{13}$	$\mathbf{y}_{14}$	$\mathbf{y}_{15}$	$\mathbf{y}_{16}$

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

$$\mathbf{y}_4 = y_1(20) + y_2(50) + y_3(10) = \mathbf{80}$$
$$\mathbf{y}_8 = y_5(8) + y_6(19) + y_7(22) = \mathbf{49}$$
$$\mathbf{y}_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}$$

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**	$\mathbf{h_{16}}$

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	$\mathbf{h}_{16}$

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**

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

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

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**

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

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

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: $a = [a_1, \ldots, a_n]$
  - 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

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 $<$ 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 $<$ 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 $<$ 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. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- 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. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- 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. $B_1 + B_2 > \frac{80}{100} \cdot T \Longleftrightarrow 900 > 800$
- 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: $a = [a_1, \ldots, a_n]$
  - 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: $PS = \{i_1, \ldots, i_p\}$

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, e.g., minimizes the number of suppressed cells **and** guarantees protection of the unsafe cells.

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, e.g., minimizes the number of suppressed cells **and** guarantees protection of the unsafe cells.
  - 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, e.g., minimizes the number of suppressed cells **and** guarantees protection of the unsafe cells.
  - 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

- **Example:** Attackers Problem

Cell suppression - attacker problem

- **Example:** Attackers Problem
- **Attacker:** has knowledge on the suppression pattern $SUP = \{1, 3, 5, 7\}$, 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	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

- **Example:** Attackers Problem
- **Attacker:** has knowledge on the suppression pattern $SUP = \{1, 3, 5, 7\}$, 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	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:** $min/max \; y_i \; \forall i \in SUP$ under above given constraints:

Cell suppression - attacker problem

- **Example:** Attackers Problem
- **Attacker:** has knowledge on the suppression pattern $SUP = \{1, 3, 5, 7\}$, 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	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:** $min/max \ y_i \ \forall i \in SUP$ under above given constraints:

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

- **Example:** Attackers Problem
- **Attacker:** has knowledge on the suppression pattern $SUP = \{1, 3, 5, 7\}$, 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	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:** $min/max \; y_i \; \forall i \in SUP$ under above given constraints:

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**

- the estimated primary suppressed cell value y_7 is $[2 : 30]$.

Cell suppression

- **Note:** Cell suppression is nothing else than a publication of intervals.

Cell suppression

- **Note:** Cell suppression is nothing else than a publication of intervals.
- **Protected?** When a cell is protected well enough? Use percentages of the upper and lower estimated cell value given the true value of the cell.

Cell suppression

- **Note:** Cell suppression is nothing else than a publication of intervals.
- **Protected?** When a cell is protected well enough? Use percentages of the upper and lower estimated cell value given the true value of the cell.
- **Example:** the attacker should not be able to estimate the primary suppressed cell value more precisely than $\pm 10\%$.

Cell suppression

- **Note:** Cell suppression is nothing else than a publication of intervals.
- **Protected?** When a cell is protected well enough? Use percentages of the upper and lower estimated cell value given the true value of the cell.
- **Example:** the attacker should not be able to estimate the primary suppressed cell value more precisely than $\pm 10\%$.
- **Information loss:** Cell suppression equals information loss. We want to find a suppression pattern that keeps the information loss as low as possible.

Cell suppression

- **Note:** Cell suppression is nothing else than a publication of intervals.
- **Protected?** When a cell is protected well enough? Use percentages of the upper and lower estimated cell value given the true value of the cell.
- **Example:** the attacker should not be able to estimate the primary suppressed cell value more precisely than $\pm 10\%$.
- **Information loss:** Cell suppression equals information loss. We want to find a suppression pattern that keeps the information loss as low as possible.
- We will show the **mathematical modell** for optimal cell suppression

Cell suppression - model assumptions

- Assumption: the intruder knows a lower and upper bound for each cell value a_i:

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

Cell suppression - model assumptions

- Assumption: the intruder knows a lower and upper bound for each cell value a_i:

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

- Relative outer bounds for each cell:

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

Cell suppression - model assumptions

- Assumption: the intruder knows a lower and upper bound for each cell value a_i:

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

- Relative outer bounds for each cell:

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

- For all sensible cells, define lower (LPL_i) and upper (UPL_i) protection levels, so that for the attackers intervals the following holds:

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

Cell suppression - model assumptions

- Assumption: the intruder knows a lower and upper bound for each cell value a_i:

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

- Relative outer bounds for each cell:

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

- For all sensible cells, define lower (LPL_i) and upper (UPL_i) protection levels, so that for the attackers intervals the following holds:

$$\begin{aligned} min(y_i) \leq a_i - LPL_i \ \forall i \in PS \\ max(y_i) \geq a_i + UPL_i \ \forall i \in PS \end{aligned}$$

- Lets define a binary variable $x_i, \ i = 1, \ldots, n$:

$$\begin{aligned} x_i = 0 \ \forall i \notin SUP \\ x_i = 1 \ \forall i \in SUP \end{aligned}$$

Cell suppression - model assumptions (2)

- For each cell a_i we define a weight w_i, that is included in the objective function to be optimized:

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

Cell suppression - model assumptions (2)

- For each cell a_i we define a weight w_i, that is included in the objective function to be optimized:

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

- The objective function to be optimized is given as:

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

Cell suppression - model assumptions (2)

- For each cell a_i we define a weight w_i, that is included in the objective function to be optimized:

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

- The objective function to be optimized is given as:

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

- under these constraints:

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

Cell suppression - the model

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

Cell suppression - the model

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

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

Cell suppression - the model

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

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

- We search for two possible tables $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$.

Cell suppression - the model

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

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

- We search for two possible tables $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$.
- The constraints (1, 2, 3) define that for f and g all linear dependencies holds and that:

Cell suppression - the model

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

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

- We search for two possible tables $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$.
- The constraints (1, 2, 3) define that for f and g all linear dependencies holds and that:

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

Cell suppression - the model

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

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

- We search for two possible tables $f = (f_1, \ldots, f_n)$ and $g = (g_1, \ldots, g_n)$.
- The constraints (1, 2, 3) define that for f and g all linear dependencies holds and that:

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

- The constraints (4) ensure the protection levels for all primary suppressed cells.

Cell suppression - remarks on the model

- The model result in a **optimal suppression pattern** related to the objective function.
- But: in **practise it never works**, because the amount of utiltiy variables (f_i, g_i, x_i) and the amount of contraints increases fastly.

Cell suppression - remarks on the model

- The model result in a **optimal suppression pattern** related to the objective function.
- But: in **practise it never works**, because the amount of utiltiy variables (f_i, g_i, x_i) and the amount of contraints increases fastly.
- Another formulation of the model allows to reduce the necessary variables using the duality principle.

Cell suppression - remarks on the model

- The model result in a **optimal suppression pattern** related to the objective function.
- But: in **practise it never works**, because the amount of utiltiy variables (f_i, g_i, x_i) and the amount of contraints increases fastly.
- Another formulation of the model allows to reduce the necessary variables using the duality principle.
- We will not go further into details, otherwise we need a lecture on linear mixed integer programming.

Cell suppression in hierachical tables

- Given the table:

Cell suppression in hierachical tables

- Given the table:

	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**

Cell suppression in hierachical tables

- Cells that needs protection are primary suppressed:

	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**

Cell suppression in hierachical tables

- Task: find a secondary suppression pattern so that primary cells cannot be estimated well enough and with a minimal Amount of secondary suppressions:

	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**

Cell suppression in hierachical tables

- Task: find a secondary suppression pattern so that primary cells cannot be estimated well enough and with a minimal Amount of secondary suppressions:

	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**

Cell suppression in hierachical tables

- Task: find a secondary suppression pattern so that primary cells cannot be estimated well enough and with a minimal Amount of secondary suppressions:

	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**

- We suppressed 13 cells in addition to the primary suppressed ones.

Cell suppression in hierachical tables

- Task: find a secondary suppression pattern so that primary cells cannot be estimated well enough and with a minimal Amount of secondary suppressions:

	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

- We suppressed 13 cells in addition to the primary suppressed ones.
- The information loss from the seconardy suppressions is 485.

Cell suppression in hierachical tables

- Task: find a secondary suppression pattern so that primary cells cannot be estimated well enough and with a minimal Amount of secondary suppressions:

	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**

- We suppressed 13 cells in addition to the primary suppressed ones.
- The information loss from the seconardy suppressions is 485.
- Is there a better suppression pattern?

Cell suppression in hierachical tables

- Solution: the optimal suppression pattern

	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**

Cell suppression in hierachical tables

- Solution: the optimal suppression pattern

	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**

- We suppressed 7 cells in addition to the primary suppressed ones.

Cell suppression in hierachical tables

- Solution: the optimal suppression pattern

	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**

- We suppressed 7 cells in addition to the primary suppressed ones.
- The information loss from the seconardy suppressions is 148.

Cell suppression in hierachical tables

- Solution: the optimal suppression pattern

	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**

- We suppressed 7 cells in addition to the primary suppressed ones.
- The information loss from the seconardy suppressions is 148.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.
- **linked tables:** If certain cells can be found different tables. If such a cell is suppression by a secondary suppression, it must be checked if the cell cannot be estimated precisly in other tables.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.
- **linked tables:** If certain cells can be found different tables. If such a cell is suppression by a secondary suppression, it must be checked if the cell cannot be estimated precisly in other tables.
- **Computational complexity:** the optimization problem is very hard to solve.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.
- **linked tables:** If certain cells can be found different tables. If such a cell is suppression by a secondary suppression, it must be checked if the cell cannot be estimated precisly in other tables.
- **Computational complexity:** the optimization problem is very hard to solve.
- **Heuristics** are necessary to find an almost optimal suppression pattern.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.
- **linked tables:** If certain cells can be found different tables. If such a cell is suppression by a secondary suppression, it must be checked if the cell cannot be estimated precisly in other tables.
- **Computational complexity:** the optimization problem is very hard to solve.
- **Heuristics** are necessary to find an almost optimal suppression pattern.
  - HITAS: Transformation of hierarchical tables to 2-dimensional tables and suppression of them given a certain schedule.

Cell suppression - challenges

- **Hierarchical tables:** Variables (z.B NACE, NUTS,...) are usually hierarchical, which makes it difficult to model the linear dependencies ($My = b$) in an automatized manner.
- **linked tables:** If certain cells can be found different tables. If such a cell is suppression by a secondary suppression, it must be checked if the cell cannot be estimated precisly in other tables.
- **Computational complexity:** the optimization problem is very hard to solve.
- **Heuristics** are necessary to find an almost optimal suppression pattern.
  - HITAS: Transformation of hierarchical tables to 2-dimensional tables and suppression of them given a certain schedule.

Rounding

- **Rounding** as an alternative to cell suppression.

Rounding

- **Rounding** as an alternative to cell suppression.
- **Variants for rounding:**

Rounding

- **Rounding** as an alternative to cell suppression.
- **Variants for rounding:**
    - rounding as usual
    - random rounding
    - controlled rounding

Rounding

- **Rounding** as an alternative to cell suppression.
- **Variants for rounding:**
  - rounding as usual
  - random rounding
  - controlled rounding
- All have in common a chosen **rounding basis** (often 3 or 5).

Rounding

- **Rounding** as an alternative to cell suppression.
- **Variants for rounding:**
  - rounding as usual
  - random rounding
  - controlled rounding
- All have in common a chosen **rounding basis** (often 3 or 5).
- **rounding as usual** (rounding to the next multiple of the basis) is not the best approach
  $\rightarrow$ we skip to apply this approach.

Random rounding

- **Idea:** a cell value is round to a multiple of the basis, but ceiling or floor is decided randomly.

Random rounding

- **Idea:** a cell value is round to a multiple of the basis, but ceiling or floor is decided randomly.
- **Disadvantage:** hierarchical tables are no longer be additiv.

Random rounding - example

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

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its basis:

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its 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**

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its 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**

- **Weighting scheme:**

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its 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**

- **Weighting scheme:**
  - rest of division = 0: cell value stays untouched.

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its 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**

- **Weighting scheme:**
  - rest of division $= 0$: cell value stays untouched.
  - rest of division $= 1$: with probability $\frac{1}{3}$ we apply ceiling, with prob. $\frac{2}{3}$ floor.

Random rounding - example

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:** Lets choose 3 and calculate the krest of the division through its 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**

- **Weighting scheme:**
  - rest of division $= 0$: cell value stays untouched.
  - rest of division $= 1$: with probability $\frac{1}{3}$ we apply ceiling, with prob. $\frac{2}{3}$ floor.
  - rest of division $= 2$: with probability $\frac{2}{3}$ we apply ceiling, with prob. $\frac{1}{3}$ floor.

Random rounding - example

- One possible solution:

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**

Random rounding - example

- One possible solution:

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**

- problem with additivity in colum 1 and 3.

Random rounding - example

- One possible solution:

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**

- problem with additivity in colum 1 and 3.
- another solution:

Random rounding - example

- One possible solution:

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**

- problem with additivity in colum 1 and 3.
- another solution:

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**

Random rounding - example

- One possible solution:

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**

- problem with additivity in colum 1 and 3.
- another solution:

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**

- additivity in colum 1,3 and 4 and rows 1-4 stimmt is violated.

Random rounding - example

- One possible solution:

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**

- problem with additivity in colum 1 and 3.
- another solution:

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**

- additivity in colum 1,3 and 4 and rows 1-4 stimmt is violated.
- **Attention:** this causes problems when the same cell is rounded different in linked tables.

Controlled rounding

- **Idea:** each cell value is rounded on a specific manner, so that additivity of tables are not violated.

Controlled rounding

- **Idea:** each cell value is rounded on a specific manner, so that additivity of tables are not violated.
- **Advantage:** tables stays (almost) additive.

Controlled rounding

- **Idea:** each cell value is rounded on a specific manner, so that additivity of tables are not violated.
- **Advantage:** tables stays (almost) additive.
- **Disadvantage:** complex problem which is often practically unsolvable.

Controlled rounding - example

- **original table:**

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

Controlled rounding - example

- **original table:**

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 **after controlled rounding:**

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**

Controlled rounding - example

- **original table:**

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 **after controlled rounding:**

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**

- All marginal totals are valid, the table is additive.

Controlled tabular adjustment - CTA

- **Idea:**
  - 1) each primary suppressed cell is replaced by an (large enough) interval.
  - 2) All cells are adjusted in a way that the tables stays additive.

Controlled tabular adjustment - CTA

- **Idea:**
  - 1) each primary suppressed cell is replaced by an (large enough) interval.
  - 2) All cells are adjusted in a way that the tables stays additive.
- **Advantage:** no suppressions! Adustments for non-primary protected cells are often minor.

Controlled tabular adjustment - CTA

- **Idea:**
  - 1) each primary suppressed cell is replaced by an (large enough) interval.
  - 2) All cells are adjusted in a way that the tables stays additive.
- **Advantage:** no suppressions! Adustments for non-primary protected cells are often minor.
- **Additional advantage:** optimal algorithms exists.

Controlled tabular adjustment - CTA

- **Idea:**
  - 1) each primary suppressed cell is replaced by an (large enough) interval.
  - 2) All cells are adjusted in a way that the tables stays additive.
- **Advantage:** no suppressions! Adustments for non-primary protected cells are often minor.
- **Additional advantage:** optimal algorithms exists.
- **Disadvantage:** optimal algorithms are only feasable in computational time for small tables. Again we need non-optimal heuristics which do not guarantee a solution of the problem.

CTA - Example

- **original table:**

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**

CTA - Example

- **original table:**

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**

- **fix the values** for the sensitive cells

CTA - Example

- **original table:**

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**

- **fix the values** for the sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of the non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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

CTA - example

- **original table:**

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**

- **Adjustment** of non-sensitive cells

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***

CTA - example

- **original table:**

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**

- Table **after CTA:**

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** is based on linear optimization (complex formulas).

Cell Adjustment by ABS

- method from **ABS** (Australian Bureau of Statistics) is a special case of **table perturbation**

Cell Adjustment by ABS

- method from **ABS** (Australian Bureau of Statistics) is a special case of **table perturbation**
- **Idea:** konsistent and random perturbation of cells in a table based on
  - Record-keys
  - Cell-keys
  - LookUp-tables

Cell Adjustment by ABS

- method from **ABS** (Australian Bureau of Statistics) is a special case of **table perturbation**
- **Idea:** konsistent and random perturbation of cells in a table based on
  - Record-keys
  - Cell-keys
  - LookUp-tables
- **Advantage:** konsistent tables results

Cell Adjustment by ABS

- method from **ABS** (Australian Bureau of Statistics) is a special case of **table perturbation**
- **Idea:** konsistent and random perturbation of cells in a table based on
  - Record-keys
  - Cell-keys
  - LookUp-tables
- **Advantage:** konsistent tables results
- **Disadvantage:** tables are non-additive anymore

Cell Adjustment by ABS

- method from **ABS** (Australian Bureau of Statistics) is a special case of **table perturbation**
- **Idea:** konsistent and random perturbation of cells in a table based on
  - Record-keys
  - Cell-keys
  - LookUp-tables
- **Advantage:** konsistent tables results
- **Disadvantage:** tables are non-additive anymore

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code
  - Well established software, but mostly it works only for Windows 32-bit version and it includes a lot of workarounds (cause the orginal code nobody touched so far)

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code
  - Well established software, but mostly it works only for Windows 32-bit version and it includes a lot of workarounds (cause the orginal code nobody touched so far)
- `sdcTable`:
  - open-source in `R`

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code
  - Well established software, but mostly it works only for Windows 32-bit version and it includes a lot of workarounds (cause the orginal code nobody touched so far)
- `sdcTable`:
  - open-source in `R`
  - easySdcTable package as front-end (we will hear a presentation, I do not know if its implemented well)

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code
  - Well established software, but mostly it works only for Windows 32-bit version and it includes a lot of workarounds (cause the orginal code nobody touched so far)
- `sdcTable`:
  - open-source in `R`
  - easySdcTable package as front-end (we will hear a presentation, I do not know if its implemented well)
  - adaptable, flexible and cool, but still in heavy development

Software for tabular protection

- τ-`Argus`:
  - τ-Argus resulted from research funds of the EU
  - Code ist partly opened after $\approx$ 15 years of the development of the code
  - Well established software, but mostly it works only for Windows 32-bit version and it includes a lot of workarounds (cause the orginal code nobody touched so far)
- `sdcTable`:
  - open-source in `R`
  - easySdcTable package as front-end (we will hear a presentation, I do not know if its implemented well)
  - adaptable, flexible and cool, but still in heavy development

Minimal example

For cell suppression:

- Problem must be specified in a structured manner (the most difficult task) (**makeProblem()**)
- Primary suppression according to a primary suppression rule (**primarySuppression()**)
- Apply one of the tabular protection rules (**protectTable()**)

Minimal example

```
> library(sdcTable)
> data("microData1", package="sdcTable")
> # having a look at the data structure
> str(microData1)
```

```
'data.frame':        100 obs. of  3 variables:
 $ region: chr  "C" "C" "A" "A" ...
 $ gender: chr  "male" "male" "male" "male" ...
 $ val   : num  9 11 10 11 18 10 6 9 5 12 ...
```

→ two spanning variables ('region' and 'gender'), one numeric ('val'). Specify hierarchical structure of 'region', levels 'A' to 'D' sum up to a total

```
> dim.region <- data.frame(
+  levels=c('@','@@','@@','@@','@@'),
+  codes=c('Total', 'A','B','C','D'),
+  stringsAsFactors=FALSE)
```

Minimal example

Specify structure of hierarchical variable 'gender'

```
> dim.gender <- hier_create(root = "Total",
+                 nodes = c("male", "female"))
> hier_display(dim.gender) # see result in R
```

```
Total
```

Minimal example

create a named list with each element being a data-frame containing information on one dimensional variable

```
> dimList <- list(region = dim.region, gender = dim.gender)
> numVarInd <- 3
```

Minimal example

In this example, no variables holding counts, numeric values, weights or sampling

```
> p1 <- makeProblem(
+   data = microData1,
+   dimList = dimList,
+   numVarInd = "val" # third variable in `data`
+ )
> print(class(p1))
```

```
[1] "sdcProblem"
attr(,"package")
[1] "sdcTable"
```

Minimal example

```
> p1 <- primarySuppression(
+   object = p1,
+   type = "freq",
+   maxN = 2
+ )
```

Minimal example

Problem is set up

```
> df1 <- sdcProb2df(p1, addDups = TRUE,
+   addNumVars = TRUE, dimCodes = "original")
> print(df1)
    strID freq sdcStatus  val region gender
 1:  0000  100         s 1284  Total  Total
 2:  0001   45         s  482  Total female
 3:  0002   55         s  802  Total   male
 4:  0100   20         s  198      A  Total
 5:  0101    2         u   20      A female
 6:  0102   18         s  178      A   male
 7:  0200   33         s  344      B  Total
 8:  0201   19         s  204      B female
 9:  0202   14         s  140      B   male
10:  0300   22         s  224      C  Total
11:  0301   10         s  106      C female
```

Minimal example

We now can apply an algorithms (several can be chosen) to receive protected tables

```
> protectedData <- protectTable(p1,
+                     method='HITAS')
```

Minimal example

```
> summary(protectedData)
######################################################
### Summary of the result object of class 'safeObj' ###
######################################################
--> The input data have been protected using algorithm HITAS.
--> The algorithm ran for 1 second.
--> To protect 1 primary sensitive cells, 3 cells need to be a
--> A total of 11 cells may be published.

###################################
### Structure of protected Data ###
###################################
Classes 'data.table' and 'data.frame':        15 obs. of  5 va
 $ region   : chr  "Total" "Total" "Total" "A" ...
 $ gender   : chr  "Total" "female" "male" "Total" ...
 $ Freq     : num  100 45 55 20 2 18 33 19 14 22 ...
```

Conclusio

- Anonymization of tabular data is very **complex**.

Conclusio

- Anonymization of tabular data is very **complex**.
- Methods:

Conclusio

- Anonymization of tabular data is very **complex**.
- Methods:
  - cell suppression
  - rounding
  - CTA
  - ABS

Conclusio

- Anonymization of tabular data is very **complex**.
- Methods:
  - cell suppression
  - rounding
  - CTA
  - ABS
- all methods have its advantages and disadvantages

Guide for sdcTable: https://cran.r-project.org/web/packages/sdcTable/vignettes/sdcTable.html