



Die Informationsmanager

ST35: Statistische Geheimhaltung in Tabellen

Bernhard Meindl, Matthias Templ

- 1 Notwendigkeit statistischer Geheimhaltung
- 2 Methoden zur Geheimhaltung tabellarischer Daten
- 3 Identifizierung schützenswerter Zellen
- 4 Verfahren zum Schutz sensibler Zellen
- 5 Software
- 6 Zusammenfassung

Notwendigkeit statistischer Geheimhaltung

- **unterschiedliche Gründe** statistische Geheimhaltung (auch) in Tabellen zu betreiben

Notwendigkeit statistischer Geheimhaltung

- **unterschiedliche Gründe** statistische Geheimhaltung (auch) in Tabellen zu betreiben
 - gesetzliche Vorgaben
 - Respondentenschutz

Notwendigkeit statistischer Geheimhaltung

- **unterschiedliche Gründe** statistische Geheimhaltung (auch) in Tabellen zu betreiben
 - gesetzliche Vorgaben
 - Respondentenschutz
- **Data privacy** ist ein wichtiges Thema

Notwendigkeit statistischer Geheimhaltung

- **unterschiedliche Gründe** statistische Geheimhaltung (auch) in Tabellen zu betreiben
 - gesetzliche Vorgaben
 - Respondentenschutz
- **Data privacy** ist ein wichtiges Thema
 - wem "gehören" meine Daten
 - kann ich mich sicher sein, dass Information nicht zu meinem Nachteil verwendet wird

Notwendigkeit statistischer Geheimhaltung

- **unterschiedliche Gründe** statistische Geheimhaltung (auch) in Tabellen zu betreiben
 - gesetzliche Vorgaben
 - Respondentenschutz
- **Data privacy** ist ein wichtiges Thema
 - wem "gehören" meine Daten
 - kann ich mich sicher sein, dass Information nicht zu meinem Nachteil verwendet wird
- **Kursziele:**
 - Bewusst machen, wie Angreifer vorgehen
 - welche Möglichkeiten gibt es, (tabellarische) Daten zu schützen

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.
- Beispiel 1: (Ort = Gemeinde x; Fußballinteresse x Geschlecht) :**

	männlich	weiblich	Gesamt
Fußballfan	12	10	22
kein Fußballfan	93	85	178
Gesamt	105	95	200

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.
- Beispiel 1: (Ort = Gemeinde x; Fußballinteresse x Geschlecht) :**

	männlich	weiblich	Gesamt
Fußballfan	12	10	22
kein Fußballfan	93	85	178
Gesamt	105	95	200

- Beispiel 1 (cont.): (Ort = Gemeinde x; Vereinspräferenz x Geschlecht) :**

	männlich	weiblich	Gesamt
Sk Rapid Wien	12	4	16
Sturm Graz	0	3	3
SV Ried	0	3	3
Gesamt	12	10	22

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.
- Beispiel 2: (Ort = Gemeinde x; Bildung x Geschlecht) :**

	männlich	weiblich	Gesamt
keine/Pflichtschule	49	53	102
Lehre	34	23	56
AHS/BHS	22	14	37
Universitäre Bildung	0	5	5
Gesamt	105	95	200

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.
- Beispiel 2: (Ort = Gemeinde x; Bildung x Geschlecht) :**

	männlich	weiblich	Gesamt
keine/Pflichtschule	49	53	102
Lehre	34	23	56
AHS/BHS	22	14	37
Universitäre Bildung	0	5	5
Gesamt	105	95	200

- man unterscheidet **verschiedene Arten von Disclosure**, z.B:

Was versteht man unter "Disclosure"?

- als **Disclosure** versteht man, wenn aus veröffentlichten Daten Information über eine einzelne, spezifische statistische Einheit abgeleitet ("gelernt") werden kann.
- **Beispiel 2: (Ort = Gemeinde x; Bildung x Geschlecht) :**

	männlich	weiblich	Gesamt
keine/Pflichtschule	49	53	102
Lehre	34	23	56
AHS/BHS	22	14	37
Universitäre Bildung	0	5	5
Gesamt	105	95	200

- man unterscheidet **verschiedene Arten von Disclosure**, z.B.:
 - Group disclosure (Beispiel 1)
 - (Negative) attribute disclosure (Beispiel 2)
 - spontane Erkennung (z.B bei Personen mit seltenen Merkmalskombinationen)

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**
- Unterscheidung von **Wertetabellen** und **Häufigkeitstabellen**

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**
- Unterscheidung von **Wertetabellen** und **Häufigkeitstabellen**
 - **Häufigkeitstabelle:** Anzahl der beitragenden Einheiten für jeder Zelle der Tabelle wird ausgewiesen.

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**
- Unterscheidung von **Wertetabellen** und **Häufigkeitstabellen**
 - **Häufigkeitstabelle:** Anzahl der beitragenden Einheiten für jeder Zelle der Tabelle wird ausgewiesen.
 - **Wertetabelle:** Für eine erhobene Variable wird die Summe dieser Variable über alle beitragenden Einheiten in jeder Zelle der Tabelle ausgewiesen.

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**
- Unterscheidung von **Wertetabellen** und **Häufigkeitstabellen**
 - **Häufigkeitstabelle:** Anzahl der beitragenden Einheiten für jeder Zelle der Tabelle wird ausgewiesen.
 - **Wertetabelle:** Für eine erhobene Variable wird die Summe dieser Variable über alle beitragenden Einheiten in jeder Zelle der Tabelle ausgewiesen.
- **Grundsätzlich...**
 - sind Tabellen gekennzeichnet durch **lineare Abhängigkeiten** zwischen Zellen.

Tabellen - Grundsätzliches

- Grundlage für alle Tabellen sind: **Mikrodaten**
- Unterscheidung von **Wertetabellen** und **Häufigkeitstabellen**
 - **Häufigkeitstabelle:** Anzahl der beitragenden Einheiten für jeder Zelle der Tabelle wird ausgewiesen.
 - **Wertetabelle:** Für eine erhobene Variable wird die Summe dieser Variable über alle beitragenden Einheiten in jeder Zelle der Tabelle ausgewiesen.
- **Grundsätzlich...**
 - sind Tabellen gekennzeichnet durch **lineare Abhängigkeiten** zwischen Zellen.
 - können statistische Tabellen **ein- oder mehrdimensional**, **hierarchisch** und/oder **verlinkt** sein.

Aufbau einer Tabelle

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

ID	DIM1	DIM2	WERT
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	h_4
II	h_5	h_6	h_7	h_8
III	h_9	h_{10}	h_{11}	h_{12}
Total	h_{13}	h_{14}	h_{15}	h_{16}

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

Von Mikrodaten zur Tabelle

- **Wiederholung** dieser Prozedur für alle Ausprägungskombinationen der Dimensionsvariablen.

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}

H	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}

Von Mikrodaten zur Tabelle

- **Wiederholung** dieser Prozedur für alle Ausprägungskombinationen der Dimensionsvariablen.

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}

H	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}

- **Randsummen** entstehen durch Auflösen der linearen Abhängigkeiten in der Tabelle.

Von Mikrodaten zur Tabelle

- **Wiederholung** dieser Prozedur für alle Ausprägungskombinationen der Dimensionsvariablen.

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}

H	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}

- **Randsummen** entstehen durch Auflösen der linearen Abhängigkeiten in der Tabelle.
- im **2-dimensionalen Fall** handelt es sich dabei um **Zeilen-** bzw. **Spaltensummen**.

Von Mikrodaten zur Tabelle

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}

Von Mikrodaten zur Tabelle

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}

$$h_4 = h_1(4) + h_2(6) + h_3(3) = \mathbf{13}$$

$$h_8 = h_5(2) + h_6(5) + h_7(7) = \mathbf{14}$$

$$h_{12} = h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12}$$

$$y_4 = y_1(20) + y_2(50) + y_3(10) = \mathbf{80}$$

$$y_8 = y_5(8) + y_6(19) + y_7(22) = \mathbf{49}$$

$$y_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}$$

Von Mikrodaten zur Tabelle

H	A	B	C	Total
I	4	6	3	13
II	2	5	7	14
III	4	5	3	12
Total	h₁₃	h₁₄	h₁₅	h₁₆

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

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	y₁₃	y₁₄	y₁₅	y₁₆

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

Von Mikrodaten zur Tabelle

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}

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

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}

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

Von Mikrodaten zur Tabelle

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}

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

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}

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

Von Mikrodaten zur 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	h₁₆

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

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	45	101	44	y₁₆

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

Von Mikrodaten zur 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	h₁₆

W	A	B	C	Total
I	20	50	10	80
II	8	19	22	49
III	17	32	12	61
Total	45	101	44	y₁₆

$$h_4 = h_1(4) + h_2(6) + h_3(3) = \mathbf{13}$$

$$h_8 = h_5(2) + h_6(5) + h_7(7) = \mathbf{14}$$

$$h_{12} = h_9(4) + h_{10}(5) + h_{11}(3) = \mathbf{12}$$

$$h_{13} = h_1(4) + h_5(2) + h_9(4) = \mathbf{10}$$

$$h_{14} = h_2(6) + h_6(5) + h_{10}(5) = \mathbf{16}$$

$$h_{15} = h_3(3) + h_7(7) + h_{11}(3) = \mathbf{13}$$

$$y_4 = y_1(20) + y_2(50) + y_3(10) = \mathbf{80}$$

$$y_8 = y_5(8) + y_6(19) + y_7(22) = \mathbf{49}$$

$$y_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = \mathbf{61}$$

$$y_{13} = y_1(20) + y_5(8) + y_9(17) = \mathbf{45}$$

$$y_{14} = y_2(50) + y_6(19) + y_{32}(5) = \mathbf{101}$$

$$y_{15} = y_3(10) + y_7(22) + y_{12}(3) = \mathbf{44}$$

Von Mikrodaten zur 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	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}

$$h_4 = h_1(4) + h_2(6) + h_3(3) = 13$$

$$h_8 = h_5(2) + h_6(5) + h_7(7) = 14$$

$$h_{12} = h_9(4) + h_{10}(5) + h_{11}(3) = 12$$

$$h_{13} = h_1(4) + h_5(2) + h_9(4) = 10$$

$$h_{14} = h_2(6) + h_6(5) + h_{10}(5) = 16$$

$$h_{15} = h_3(3) + h_7(7) + h_{11}(3) = 13$$

$$h_{16} = h_4(13) + h_8(14) + h_{12}(12) = 39$$

$$h_{16} = h_{13}(10) + h_{14}(16) + h_{15}(13) = 39$$

$$y_4 = y_1(20) + y_2(50) + y_3(10) = 80$$

$$y_8 = y_5(8) + y_6(19) + y_7(22) = 49$$

$$y_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = 61$$

$$y_{13} = y_1(20) + y_5(8) + y_9(17) = 45$$

$$y_{14} = y_2(50) + y_6(19) + y_{32}(5) = 101$$

$$y_{15} = y_3(10) + y_7(22) + y_{12}(3) = 44$$

$$y_{16} = y_4(80) + y_8(49) + y_{12}(61) = 190$$

$$y_{16} = y_{13}(45) + y_{14}(101) + y_{15}(44) = 190$$

Von Mikrodaten zur 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

$$h_4 = h_1(4) + h_2(6) + h_3(3) = 13$$

$$h_8 = h_5(2) + h_6(5) + h_7(7) = 14$$

$$h_{12} = h_9(4) + h_{10}(5) + h_{11}(3) = 12$$

$$h_{13} = h_1(4) + h_5(2) + h_9(4) = 10$$

$$h_{14} = h_2(6) + h_6(5) + h_{10}(5) = 16$$

$$h_{15} = h_3(3) + h_7(7) + h_{11}(3) = 13$$

$$h_{16} = h_4(13) + h_8(14) + h_{12}(12) = 39$$

$$h_{16} = h_{13}(10) + h_{14}(16) + h_{15}(13) = 39$$

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

$$y_4 = y_1(20) + y_2(50) + y_3(10) = 80$$

$$y_8 = y_5(8) + y_6(19) + y_7(22) = 49$$

$$y_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = 61$$

$$y_{13} = y_1(20) + y_5(8) + y_9(17) = 45$$

$$y_{14} = y_2(50) + y_6(19) + y_{32}(5) = 101$$

$$y_{15} = y_3(10) + y_7(22) + y_{12}(3) = 44$$

$$y_{16} = y_4(80) + y_8(49) + y_{12}(61) = 190$$

$$y_{16} = y_{13}(45) + y_{14}(101) + y_{15}(44) = 190$$

Von Mikrodaten zur 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

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

$$h_4 = h_1(4) + h_2(6) + h_3(3) = 13$$

$$h_8 = h_5(2) + h_6(5) + h_7(7) = 14$$

$$h_{12} = h_9(4) + h_{10}(5) + h_{11}(3) = 12$$

$$h_{13} = h_1(4) + h_5(2) + h_9(4) = 10$$

$$h_{14} = h_2(6) + h_6(5) + h_{10}(5) = 16$$

$$h_{15} = h_3(3) + h_7(7) + h_{11}(3) = 13$$

$$h_{16} = h_4(13) + h_8(14) + h_{12}(12) = 39$$

$$h_{16} = h_{13}(10) + h_{14}(16) + h_{15}(13) = 39$$

$$y_4 = y_1(20) + y_2(50) + y_3(10) = 80$$

$$y_8 = y_5(8) + y_6(19) + y_7(22) = 49$$

$$y_{12} = y_9(17) + y_{10}(32) + y_{11}(12) = 61$$

$$y_{13} = y_1(20) + y_5(8) + y_9(17) = 45$$

$$y_{14} = y_2(50) + y_6(19) + y_{32}(5) = 101$$

$$y_{15} = y_3(10) + y_7(22) + y_{12}(3) = 44$$

$$y_{16} = y_4(80) + y_8(49) + y_{12}(61) = 190$$

$$y_{16} = y_{13}(45) + y_{14}(101) + y_{15}(44) = 190$$

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $Ma = b$

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $Ma = b$
- **Bemerkungen:**
 - M ist eine Matrix mit $M_{ij} \in \{-1, 0, 1\}$ und b ist ein Vektor mit lauter 0ern

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $Ma = b$
- **Bemerkungen:**
 - M ist eine Matrix mit $M_{ij} \in \{-1, 0, 1\}$ und b ist ein Vektor mit lauter 0ern

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

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $Ma = b$
- **Bemerkungen:**
 - M ist eine Matrix mit $M_{ij} \in \{-1, 0, 1\}$ und b ist ein Vektor mit lauter 0ern

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

- Jede Zeile des Gleichungssystems $Ma = b$ entspricht hier der Einschränkung einer Zeilen-/ oder Spaltensumme.

Formalisierung

- **Verallgemeinerung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $Ma = b$
- **Bemerkungen:**
 - M ist eine Matrix mit $M_{ij} \in \{-1, 0, 1\}$ und b ist ein Vektor mit lauter 0ern

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

- Jede Zeile des Gleichungssystems $Ma = b$ entspricht hier der Einschränkung einer Zeilen-/ oder Spaltensumme.
- die Zellen einer Tabelle sind festgelegt durch ihren (Spalten)Index:
 $i = 1, \dots, n$

- Zur Beurteilung, ob eine Tabellenzelle als "unsicher" (und daher schützenswert) gelten soll, kann eine der folgenden Regeln herangezogen werden:

- Zur Beurteilung, ob eine Tabellenzelle als "unsicher" (und daher schützenswert) gelten soll, kann eine der folgenden Regeln herangezogen werden:
 - **Fallzahlregel:**
Die Anzahl der zu einer Zelle beitragenden Einheiten ist $<$ einem festgesetzten Wert (oftmals 3 oder 4)

- Zur Beurteilung, ob eine Tabellenzelle als "unsicher" (und daher schützenswert) gelten soll, kann eine der folgenden Regeln herangezogen werden:
 - **Fallzahlregel:**
Die Anzahl der zu einer Zelle beitragenden Einheiten ist $<$ einem festgesetzten Wert (oftmals 3 oder 4)
 - **(n,k)-Dominanzregel:**
Eine Zelle muss geschützt werden, wenn der Gesamtwert der n größten Beitragenden $k\%$ des gesamten Zellwertes überschreitet.

- Zur Beurteilung, ob eine Tabellenzelle als "unsicher" (und daher schützenswert) gelten soll, kann eine der folgenden Regeln herangezogen werden:
 - **Fallzahlregel:**
Die Anzahl der zu einer Zelle beitragenden Einheiten ist $<$ einem festgesetzten Wert (oftmals 3 oder 4)
 - **(n,k)-Dominanzregel:**
Eine Zelle muss geschützt werden, wenn der Gesamtwert der n größten Beitragenden $k\%$ des gesamten Zellwertes überschreitet.
 - **p-% Regel:**
Der Totalwert minus der 2 größten Beitragenden ist geringer als $p\%$ des größten Beitrages.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p -%-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p -%-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.
 - Nach der p – %-Regel wird das "nicht genug" als Rate ($p\%$) am wahren Wert der größten beitragenden Einheit gemessen.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p -%-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.
 - Nach der p – %-Regel wird das "nicht genug" als Rate ($p\%$) am wahren Wert der größten beitragenden Einheit gemessen.
 - nach der (n,k) -Dominanzregel wird das "nicht genug" als Rate $(100 - k)\%$ am Zellwert gemessen.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p -%-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.
 - Nach der p – %-Regel wird das "nicht genug" als Rate ($p\%$) am wahren Wert der größten beitragenden Einheit gemessen.
 - nach der (n,k) -Dominanzregel wird das "nicht genug" als Rate $(100 - k)\%$ am Zellwert gemessen.
 - **Ausserdem gilt:**
 - jeder Zellwert, der nach der $(2,k)$ -Regel als "sicher" gilt, ist auch "sicher" nach der p %-Regel.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p -%-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.
 - Nach der p – %-Regel wird das "nicht genug" als Rate ($p\%$) am wahren Wert der größten beitragenden Einheit gemessen.
 - nach der (n,k) -Dominanzregel wird das "nicht genug" als Rate $(100 - k)\%$ am Zellwert gemessen.
 - **Ausserdem gilt:**
 - jeder Zellwert, der nach der $(2,k)$ -Regel als "sicher" gilt, ist auch "sicher" nach der $p\%$ -Regel.
 - nicht jeder Zellwert, der nach der $p\%$ -Regel als "sicher" gilt ist auch "sicher" nach der $(2,k)$ -Dominanzregel.

Primärsperungen (Fortsetzung)

- **Zusammenhang** zwischen (n,k) -Dominanzregel und p %-Regel:
 - Nach beiden Dominanzregeln müssen Zellwerte geschützt werden, wenn ein oberer Schätzwert für den Wert des größten Beitragenden konstruiert werden kann, der den wahren Wert **nicht genug** überschätzt.
 - Nach der p – %-Regel wird das "nicht genug" als Rate ($p\%$) am wahren Wert der größten beitragenden Einheit gemessen.
 - nach der (n,k) -Dominanzregel wird das "nicht genug" als Rate $(100 - k)\%$ am Zellwert gemessen.
 - **Ausserdem gilt:**
 - jeder Zellwert, der nach der $(2,k)$ -Regel als "sicher" gilt, ist auch "sicher" nach der $p\%$ -Regel.
 - nicht jeder Zellwert, der nach der $p\%$ -Regel als "sicher" gilt ist auch "sicher" nach der $(2,k)$ -Dominanzregel.
 - Es gilt für den Zusammenhang zwischen $p\%$ -Regel und $(2,k)$ -Regel:
$$p = 100 \cdot \frac{100 - k}{k}$$

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
→ Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
 \rightarrow Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- welchem p entspricht die (2,80)-Dominanzregel?

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
 \rightarrow Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- welchem p entspricht die (2,80)-Dominanzregel?
 $\rightarrow p = 100 \cdot \frac{100-80}{80} = 25$

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
 \rightarrow Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- welchem p entspricht die (2,80)-Dominanzregel?
 $\rightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- sei $p=25$. Ist die Zelle bei Anwendung der 25%-Regel zu schützen?

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
 \rightarrow Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- welchem p entspricht die (2,80)-Dominanzregel?
 $\rightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- sei $p=25$. Ist die Zelle bei Anwendung der 25%-Regel zu schützen?
 \rightarrow Ja. $T - B_1 - B_2 < \frac{25}{100} \cdot B_1 \iff 100 < 125$

Identifizierung sensibler Zellen (Beispiele)

- **Beispiel:** Der Gesamtwert T einer Tabellenzelle sei 1000.
Der Wert des größten Einzelbeitrages sei $B_1 = 500$.
Der zweitgrößte Einzelbeitrag sei $B_2 = 400$.
- Ist die Zelle nach Anwendung der (2,80)-Dominanzregel geheimzuhalten?
 \rightarrow Ja. $B_1 + B_2 > \frac{80}{100} \cdot T \iff 900 > 800$
- welchem p entspricht die (2,80)-Dominanzregel?
 $\rightarrow p = 100 \cdot \frac{100-80}{80} = 25$
- sei $p=25$. Ist die Zelle bei Anwendung der 25%-Regel zu schützen?
 \rightarrow Ja. $T - B_1 - B_2 < \frac{25}{100} \cdot B_1 \iff 100 < 125$

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $My = b$

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $My = b$
 - obere und untere Grenzen für jeden Tabellenwert, die einem Angreifer bekannt sind (zb. nicht-Negativität): $lb_i \leq a_i \leq ub_i$

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $My = b$
 - obere und untere Grenzen für jeden Tabellenwert, die einem Angreifer bekannt sind (zb. nicht-Negativität): $lb_i \leq a_i \leq ub_i$
 - Die Zelle einer Tabelle ist festgelegt durch ihren Index: $i = 1, \dots, n$

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $My = b$
 - obere und untere Grenzen für jeden Tabellenwert, die einem Angreifer bekannt sind (zb. nicht-Negativität): $lb_i \leq a_i \leq ub_i$
 - Die Zelle einer Tabelle ist festgelegt durch ihren Index: $i = 1, \dots, n$
- **Ausserdem:**
 - gegeben sind p Primärsperkungen: $PS = \{i_1, \dots, i_p\}$

Schützen sensibler Zellen

- **Wiederholung:** Eine (mehrdimensionale, hierarchische) Tabelle ist gegeben durch:
 - einen Datenvektor: $a = [a_1, \dots, a_n]$
 - lineare Einschränkungen/Constraints der Form: $My = b$
 - obere und untere Grenzen für jeden Tabellenwert, die einem Angreifer bekannt sind (zb. nicht-Negativität): $lb_i \leq a_i \leq ub_i$
 - Die Zelle einer Tabelle ist festgelegt durch ihren Index: $i = 1, \dots, n$
- **Ausserdem:**
 - gegeben sind p Primärsperkungen: $PS = \{i_1, \dots, i_p\}$
- **Frage:** Wie schützt man primär gesperrte Zellen?

Schützen sensibler Zellen

- **Beispiel:**

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

Schützen sensibler Zellen

- **Beispiel:**

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

- Sei Zelle II/C ($PS = \{7\}$) sensibel und muss geschützt werden.
- Verschiedene Möglichkeiten um den Wert zu schützen. z.B:
 - **Zellsperrung/-unterdrückung**
 - **Runden**
 - **Zellanpassung**

Zellunterdrückung

- **Beispiel:**

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

- Zellunterdrückung ist eine sehr häufig verwendete Methode.
- **Aber:** Wegen linearer Zusammenhänge ist es nicht ausreichend, nur geheimzuhaltende Zellen alleine zu unterdrücken (Primärspernung).

Zellunterdrückung

- **Beispiel:**

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

- Zellunterdrückung ist eine sehr häufig verwendete Methode.
- **Aber:** Wegen linearer Zusammenhänge ist es nicht ausreichend, nur geheimzuhaltende Zellen alleine zu unterdrücken (Primärspernung).
- **Man sieht:** $44 - 10 - 12 = 22$ bzw. $49 - 8 - 19 = 22$.
kein ausreichender Schutz für die primär gesperrte Zelle.

Zellunterdrückung

- **Beispiel:**

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

- Zellunterdrückung ist eine sehr häufig verwendete Methode.
- **Aber:** Wegen linearer Zusammenhänge ist es nicht ausreichend, nur geheimzuhaltende Zellen alleine zu unterdrücken (Primärspernung).
- **Man sieht:** $44 - 10 - 12 = 22$ bzw. $49 - 8 - 19 = 22$.
kein ausreichender Schutz für die primär gesperrte Zelle.
- → **sekundäre Unterdrückung:** zusätzliche Zellen müssen zum Schutz der primär gesperrten Zelle unterdrückt werden.

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

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

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

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

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

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

- Was ist ein ausreichendes Sperrmuster?

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

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

- Was ist ein ausreichendes Sperrmuster?
- Gibt es optimale Sperrmuster und wenn ja, was charakterisiert ein optimales Muster?

Zellunterdrückung

- **Beispiel:** alternative Sperrmuster

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

- Was ist ein ausreichendes Sperrmuster?
- Gibt es optimale Sperrmuster und wenn ja, was charakterisiert ein optimales Muster?
- **Grundsätzlich:** Problem der sekundären Unterdrückung wird NP-hard bei hierarchischen, multidimensionalen bzw. verlinkten Tabellen

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.
- **schlechte Nachricht:** Der optimale Algorithmus ist in der Praxis kaum (nicht) verwendbar.

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.
- **schlechte Nachricht:** Der optimale Algorithmus ist in der Praxis kaum (nicht) verwendbar.
- Optimale Algorithmen basieren auf **Linearer Optimierung**

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.
- **schlechte Nachricht:** Der optimale Algorithmus ist in der Praxis kaum (nicht) verwendbar.
- Optimale Algorithmen basieren auf **Linearer Optimierung**
 - Ziel ist es, ein Sperrmuster zu finden, das entweder die Anzahl der zusätzlich zu unterdrückenden Zellen oder die Summe der zusätzlich zu unterdrückenden Tabellenzellen minimiert und gleichzeitig ausreichenden Schutz gewährleistet.

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.
- **schlechte Nachricht:** Der optimale Algorithmus ist in der Praxis kaum (nicht) verwendbar.
- Optimale Algorithmen basieren auf **Linearer Optimierung**
 - Ziel ist es, ein Sperrmuster zu finden, das entweder die Anzahl der zusätzlich zu unterdrückenden Zellen oder die Summe der zusätzlich zu unterdrückenden Tabellenzellen minimiert und gleichzeitig ausreichenden Schutz gewährleistet.
 - ausreichender Schutz: Eine primär zu sichernde Zelle gilt als ausreichend geschützt, wenn der unterdrückte Wert nicht innerhalb eines zu definierenden Intervalls einzuschränken ist.

Zellunterdrückung

- **gute Nachricht:** es existieren optimale Algorithmen für das sekundäre Zellunterdrückungsproblem.
- **schlechte Nachricht:** Der optimale Algorithmus ist in der Praxis kaum (nicht) verwendbar.
- Optimale Algorithmen basieren auf **Linearer Optimierung**
 - Ziel ist es, ein Sperrmuster zu finden, das entweder die Anzahl der zusätzlich zu unterdrückenden Zellen oder die Summe der zusätzlich zu unterdrückenden Tabellenzellen minimiert und gleichzeitig ausreichenden Schutz gewährleistet.
 - ausreichender Schutz: Eine primär zu sichernde Zelle gilt als ausreichend geschützt, wenn der unterdrückte Wert nicht innerhalb eines zu definierenden Intervalls einzuschränken ist.
- **Attackers Problem:** Für ein gegebenes Sperrmuster *SUP* berechnet sich der Angreifer systematisch untere und obere (mögliche) Grenzen für die gesperrten Zellen.

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{5, 7, 9, 11\}$, 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	20	50	10	80
II	y_5	19	y_7	49
III	y_9	32	y_{11}	61
Total	45	101	44	190

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{5, 7, 9, 11\}$, 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	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 obigen NB:

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{5, 7, 9, 11\}$, 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	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 obigen NB:

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

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem
- **Angreifer:** kennt das Sperrmuster $SUP = \{5, 7, 9, 11\}$, 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	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 obigen NB:

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

- der primär gesperrte Wert y_7 ist im Intervall $[5 : 30]$ geschützt.

Zellunterdrückung - Attackers Problem

- **Beispiel:** Attackers Problem

Zellunterdrückung - Attackers 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

Zellunterdrückung - Attackers 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:

Zellunterdrückung - Attackers 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

Zellunterdrückung - Attackers 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 primär gesperrte Wert y_7 ist im Intervall $[2 : 30]$ geschützt.

Zellunterdrückung

- **Bemerkung:** Zellunterdrückung ist in Wahrheit eine Form von Intervallpublikation.

Zellunterdrückung

- **Bemerkung:** Zellunterdrückung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu klären, ab wann eine Zelle ausreichend geschützt ist. Im Normalfall werden für ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.

Zellunterdrückung

- **Bemerkung:** Zellunterdrückung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu klären, ab wann eine Zelle ausreichend geschützt ist. Im Normalfall werden für ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen primär unsicheren Zellwert auf $\pm 10\%$ genau berechnen können.

Zellunterdrückung

- **Bemerkung:** Zellunterdrückung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu klären, ab wann eine Zelle ausreichend geschützt ist. Im Normalfall werden für ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen primär unsicheren Zellwert auf $\pm 10\%$ genau berechnen können.
- **Informationsverlust:** Zellunterdrückung bewirkt Informationsverlust, der durch ein optimales Unterdrückungsschema mittels einer Verlustfunktion minimiert werden soll.

Zellunterdrückung

- **Bemerkung:** Zellunterdrückung ist in Wahrheit eine Form von Intervallpublikation.
- **Ausreichender Schutz:** Es bleibt die Frage zu klären, ab wann eine Zelle ausreichend geschützt ist. Im Normalfall werden für ausreichende obere und untere Grenzen Prozentwerte des Original-Zellwerts verwendet.
- **Beispiel:** der Angreifer darf keinen primär unsicheren Zellwert auf $\pm 10\%$ genau berechnen können.
- **Informationsverlust:** Zellunterdrückung bewirkt Informationsverlust, der durch ein optimales Unterdrückungsschema mittels einer Verlustfunktion minimiert werden soll.
- Wir zeigen das **mathematische Modell** für optimale Zellunterdrückung:

Zellunterdrückung - Modellannahmen

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

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

Zellunterdrückung - Modellannahmen

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

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

- Man definiert relative äussere Grenzen für jede Zelle:

$$LB_i := a_i - lb_i \geq 0 \quad \forall i = 1, \dots, n$$

$$UB_i := ub_i - a_i \geq 0 \quad \forall i = 1, \dots, n$$

Zellunterdrückung - Modellannahmen

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

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

- Man definiert relative äussere Grenzen für jede Zelle:

$$LB_i := a_i - lb_i \geq 0 \quad \forall i = 1, \dots, n$$

$$UB_i := ub_i - a_i \geq 0 \quad \forall i = 1, \dots, n$$

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

$$\min(y_i) \leq a_i - LPL_i \quad \forall i \in PS$$

$$\max(y_i) \geq a_i + UPL_i \quad \forall i \in PS$$

Zellunterdrückung - Modellannahmen

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

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

- Man definiert relative äussere Grenzen für jede Zelle:

$$LB_i := a_i - lb_i \geq 0 \quad \forall i = 1, \dots, n$$

$$UB_i := ub_i - a_i \geq 0 \quad \forall i = 1, \dots, n$$

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

$$\min(y_i) \leq a_i - LPL_i \quad \forall i \in PS$$

$$\max(y_i) \geq a_i + UPL_i \quad \forall i \in PS$$

- Wir führen binäre Variablen x_i , $i = 1, \dots, n$ ein, für die gelten soll:

$$x_i = 0 \quad \forall i \notin SUP$$

$$x_i = 1 \quad \forall i \in SUP$$

Zellunterdrückung - Modellannahmen (2)

- Wir definieren für jede Zelle a_i ein Gewicht w_i , das in die Zielfunktion einfließt, z.B:

$$w_i = a_i$$

$$w_i = 1$$

$$w_i = \log(1 + a_i)$$

Zellunterdrückung - Modellannahmen (2)

- Wir definieren für jede Zelle a_i ein Gewicht w_i , das in die Zielfunktion einfließt, z.B:

$$w_i = a_i$$

$$w_i = 1$$

$$w_i = \log(1 + a_i)$$

- Die Zielfunktion des Optimierungsproblems ist gegeben als:

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

Zellunterdrückung - Modellannahmen (2)

- Wir definieren für jede Zelle a_i ein Gewicht w_i , das in die Zielfunktion einfließt, z.B:

$$w_i = a_i$$

$$w_i = 1$$

$$w_i = \log(1 + a_i)$$

- Die Zielfunktion des Optimierungsproblems ist gegeben als:

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

- unter folgenden Nebenbedingungen:

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

Zellunterdrückung - Das Modell

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

Zellunterdrückung - Das Modell

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

$$Mf = b \qquad Mg = b \qquad (1)$$

$$f_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (2)$$

$$f_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (3)$$

$$f_i \leq a_i - LPL_i \quad \forall i \in PS \qquad g_i \geq a_i + UPL_i \quad \forall i \in PS \qquad (4)$$

Zellunterdrückung - Das Modell

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

$$Mf = b \qquad Mg = b \qquad (1)$$

$$f_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (2)$$

$$f_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (3)$$

$$f_i \leq a_i - LPL_i \quad \forall i \in PS \qquad g_i \geq a_i + UPL_i \quad \forall i \in PS \qquad (4)$$

- Es werden zwei mögliche Tabellen $f = (f_1, \dots, f_n)$ sowie $g = (g_1, \dots, g_n)$ gesucht.

Zellunterdrückung - Das Modell

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

$$Mf = b \qquad Mg = b \qquad (1)$$

$$f_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (2)$$

$$f_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (3)$$

$$f_i \leq a_i - LPL_i \quad \forall i \in PS \qquad g_i \geq a_i + UPL_i \quad \forall i \in PS \qquad (4)$$

- Es werden zwei mögliche Tabellen $f = (f_1, \dots, f_n)$ sowie $g = (g_1, \dots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass für f und g alle linearen Abhängigkeiten erfüllt sind und dass gilt:

Zellunterdrückung - Das Modell

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

$$Mf = b \qquad Mg = b \qquad (1)$$

$$f_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (2)$$

$$f_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (3)$$

$$f_i \leq a_i - LPL_i \quad \forall i \in PS \qquad g_i \geq a_i + UPL_i \quad \forall i \in PS \qquad (4)$$

- Es werden zwei mögliche Tabellen $f = (f_1, \dots, f_n)$ sowie $g = (g_1, \dots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass für f und g alle linearen Abhängigkeiten erfüllt sind und dass gilt:

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

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

Zellunterdrückung - Das Modell

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

$$Mf = b \qquad Mg = b \qquad (1)$$

$$f_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \geq a_i - LB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (2)$$

$$f_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad g_i \leq a_i + UB_i \cdot x_i \quad \forall i = 1, \dots, n \qquad (3)$$

$$f_i \leq a_i - LPL_i \quad \forall i \in PS \qquad g_i \geq a_i + UPL_i \quad \forall i \in PS \qquad (4)$$

- Es werden zwei mögliche Tabellen $f = (f_1, \dots, f_n)$ sowie $g = (g_1, \dots, g_n)$ gesucht.
- Die NB (1, 2, 3) stellen sicher, dass für f und g alle linearen Abhängigkeiten erfüllt sind und dass gilt:

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

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

- die Nebenbedingungen (4) erzwingen die Einhaltung der Protection Levels für alle primär-gesperren Zellen.

Zellunterdrückung - 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 groß wird.

Zellunterdrückung - 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 groß wird.
- Es ist möglich, die Anzahl der für das Modell notwendigen Variablen zu verringern, indem man das **Dualitätsprinzip** jedes lineares Problems ausnützt.

Zellunterdrückung - 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 groß wird.
- Es ist möglich, die Anzahl der für das Modell notwendigen Variablen zu verringern, indem man das **Dualitätsprinzip** jedes lineares Problems ausnützt.
- Das Modell wird nur mehr mittels der binären Variablen x_i formuliert.

Zellunterdrückung - 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 groß wird.
- Es ist möglich, die Anzahl der für das Modell notwendigen Variablen zu verringern, indem man das **Dualitätsprinzip** jedes lineares Problems ausnützt.
- Das Modell wird nur mehr mittels der binären Variablen x_i formuliert.
- Es werden **schrittweise zusätzliche Nebenbedingungen** ins Modell aufgenommen, die aber nur mehr von x_i abhängen.

Zellunterdrückung - 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 groß wird.
- Es ist möglich, die Anzahl der für das Modell notwendigen Variablen zu verringern, indem man das **Dualitätsprinzip** jedes lineares Problems ausnützt.
- Das Modell wird nur mehr mittels der binären Variablen x_i formuliert.
- Es werden **schrittweise zusätzliche Nebenbedingungen** ins Modell aufgenommen, die aber nur mehr von x_i abhängen.
- Es ergibt sich ein **iterativer Algorithmus**, in dem zwar mehr lineare Probleme gelöst werden müssen, die aber weniger komplex und umfangreich sind.

Zellunterdrückung in hierarchischen Tabellen

- Gegeben sei folgende Tabelle:

Zellunterdrückung 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

Zellunterdrückung in hierarchischen Tabellen

- Schützenswerte Zellen seien identifiziert und gelöscht:

	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

Zellunterdrückung in hierarchischen Tabellen

- Aufgabe: Auffinden eines gültigen Sperrmusters gegen exakte Rückrechenbarkeit mit minimaler Anzahl von Sekundärspernungen:

	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

Zellunterdrückung in hierarchischen Tabellen

- Aufgabe: Auffinden eines gültigen Sperrmusters gegen exakte Rückrechenbarkeit mit minimaler Anzahl von Sekundärsperungen:

	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

Zellunterdrückung in hierarchischen Tabellen

- Aufgabe: Auffinden eines gültigen Sperrmusters gegen exakte Rückrechenbarkeit mit minimaler Anzahl von Sekundärsperren:

	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 zusätzliche Zellen sperren.

Zellunterdrückung in hierarchischen Tabellen

- Aufgabe: Auffinden eines gültigen Sperrmusters gegen exakte Rückrechenbarkeit mit minimaler Anzahl von Sekundärsperren:

	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 zusätzliche Zellen sperren.
- Der Informationsverlust der Sekundärsperren beträgt 485.

Zellunterdrückung in hierarchischen Tabellen

- Aufgabe: Auffinden eines gültigen Sperrmusters gegen exakte Rückrechenbarkeit mit minimaler Anzahl von Sekundärsperungen:

	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 zusätzliche Zellen sperren.
- Der Informationsverlust der Sekundärsperungen beträgt 485.
- Gibt es bessere/alternative Sperrmuster?

Zellunterdrückung 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

Zellunterdrückung 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 zusätzliche Zellen sperren.

Zellunterdrückung 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 zusätzliche Zellen sperren.
- Der Informationsverlust der Sekundärsperren beträgt 148.

Zellunterdrückung 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 zusätzliche Zellen sperren.
- Der Informationsverlust der Sekundärsperren beträgt 148.

Zellunterdrückung - Herausforderungen

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

Zellunterdrückung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abhängigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekundär) gesperrt, muss in allen Tabellen die Rückrechenbarkeit überprüft werden.

Zellunterdrückung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abhängigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekundär) gesperrt, muss in allen Tabellen die Rückrechenbarkeit überprüft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und führt oft zu langen Rechenzeiten.

Zellunterdrückung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abhängigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekundär) gesperrt, muss in allen Tabellen die Rückrechenbarkeit überprüft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und führt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale Lösungen liefern.

Zellunterdrückung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abhängigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekundär) gesperrt, muss in allen Tabellen die Rückrechenbarkeit überprüft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und führt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale Lösungen liefern.
 - HITAS: Umwandlung von hierarchisch gegliederten Tabellen in einfache, 2-dimensionale Tabellen. Zellspernung in 2-dimensionalen Tabellen nach einer gewissen Reihenfolge.

Zellunterdrückung - Herausforderungen

- **Hierarchische Tabellen:** Dimensionsvariablen (z.B NACE, NUTS,...) sind hierarchisch gegliedert. Es ist komplex, die linearen Abhängigkeiten ($My = b$) (automatisch) zu modellieren.
- **verlinkte Tabellen:** Man spricht von verlinkten Tabellen, wenn einzelne Zellen in unterschiedlichen Tabellen auftreten. Wird eine solche Zelle (sekundär) gesperrt, muss in allen Tabellen die Rückrechenbarkeit überprüft werden.
- **Rechenzeit:** das Optimierungsproblem ist hochgradig komplex und führt oft zu langen Rechenzeiten.
- **Heuristiken:** es ist notwendig, Heuristiken zu entwickeln/verwenden, die quasi/optimale Lösungen liefern.
 - HITAS: Umwandlung von hierarchisch gegliederten Tabellen in einfache, 2-dimensionale Tabellen. Zellspernung in 2-dimensionalen Tabellen nach einer gewissen Reihenfolge.
 - Quaderverfahren: Algorithmus zum Aufsuchen "geometrischer" Sperrmuster

Runden

- **Runden** ist eine Alternative zur Zellunterdrückung.

Runden

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

Runden

- **Runden** ist eine Alternative zur Zellunterdrückung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
 - normales Runden:
 - zufälliges Runden
 - kontrolliertes Runden

Runden

- **Runden** ist eine Alternative zur Zellunterdrückung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
 - normales Runden:
 - zufälliges Runden
 - kontrolliertes Runden
- Bei allen Varianten muss eine **Rundungsbasis** (oft 3 oder 5) gewählt werden.

Runden

- **Runden** ist eine Alternative zur Zellunterdrückung.
- **Varianten:** Es gibt verschiedene Varianten zum Runden von Tabellen:
 - normales Runden:
 - zufälliges Runden
 - kontrolliertes Runden
- Bei allen Varianten muss eine **Rundungsbasis** (oft 3 oder 5) gewählt werden.
- **normales Runden** (Runden des Zellwertes zum nächsten Vielfachen der Basis) bringt etwas Schutz, aber nicht genug
→ wir vernachlässigen diese Variante.

Zufälliges Runden

- **Idee:** jeder Zellwert wird in einer zufälligen Art und Weise - unabhängig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.

Zufälliges Runden

- **Idee:** jeder Zellwert wird in einer zufälligen Art und Weise - unabhängig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht verändert.

Zufälliges Runden

- **Idee:** jeder Zellwert wird in einer zufälligen Art und Weise - unabhängig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht verändert.
- **Randsummen** werden üblicherweise getrennt von den inneren Tabellenzellen behandelt.

Zufälliges Runden

- **Idee:** jeder Zellwert wird in einer zufälligen Art und Weise - unabhängig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht verändert.
- **Randsummen** werden üblicherweise getrennt von den inneren Tabellenzellen behandelt.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind möglich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.

Zufälliges Runden

- **Idee:** jeder Zellwert wird in einer zufälligen Art und Weise - unabhängig von allen anderen Zellen - zu einem **Vielfachen der Basis** auf- oder abgerundet.
- **Vielfache** der Basis werden nicht verändert.
- **Randsummen** werden üblicherweise getrennt von den inneren Tabellenzellen behandelt.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind möglich, jedoch soll keine Tendenz zum Auf- oder Abrunden durch das Gewichtungsschema implizit gegeben werden.
- **Nachteil:** Tabellen sind möglicherweise (wahrscheinlich) nicht mehr additiv.

Zufälliges 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

Zufälliges 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 wählen 3 und berechnen die Reste der Division der Zellwerte durch die Basis:

Zufälliges 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 wählen 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

Zufälliges 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 wählen 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 wählen folgendes Gewichtungsschema:

Zufälliges 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 wählen 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 wählen folgendes Gewichtungsschema:
 - Divisionsrest = 0: Tabellenwert wird nicht verändert.

Zufälliges 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 wählen 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 wählen folgendes Gewichtungsschema:
 - Divisionsrest = 0: Tabellenwert wird nicht verändert.
 - Divisionsrest = 1: mit Wahrscheinlichkeit $\frac{1}{3}$ wird aufgerundet, mit Wahrscheinlichkeit $\frac{2}{3}$ abgerundet.

Zufälliges 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 wählen 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 wählen folgendes Gewichtungsschema:
 - Divisionsrest = 0: Tabellenwert wird nicht verändert.
 - 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.

Zufälliges 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

Zufälliges 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

- Additivität in Spalten 1 und 3 stimmt nicht mehr.

Zufälliges 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

- Additivität in Spalten 1 und 3 stimmt nicht mehr.
- andere Möglichkeit:

Zufälliges 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

- Additivität in Spalten 1 und 3 stimmt nicht mehr.
- andere Möglichkeit:

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

Zufälliges 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

- Additivität in Spalten 1 und 3 stimmt nicht mehr.
- andere Möglichkeit:

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: Additivität in Spalte 1,3 und 4 sowie Zeile 1-4 stimmt nicht mehr.

Zufälliges 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

- Additivität in Spalten 1 und 3 stimmt nicht mehr.
- andere Möglichkeit:

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: Additivität 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 Additivität 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 Additivität der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grundsätzlich) nicht verändert.

Kontrolliertes Runden

- **Idee:** jeder Zellwert wird in einer Art und Weise zu einem Vielfachen der Basis auf- oder abgerundet, sodass die Additivität der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grundsätzlich) nicht verändert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind möglich, 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 Additivität der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grundsätzlich) nicht verändert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind möglich, 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 Additivität der Tabelle gewahrt bleibt.
- **Vielfache** der Basis werden (grundsätzlich) nicht verändert.
- **Wichtig:** unterschiedliche Gewichtungsschemata sind möglich, 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 möglicherweise unlösbar ist.

Kontrolliertes Runden - Beispiel

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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 primär 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 primär 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 Zellunterdrückung keine lückenhaften Tabellen, ausserdem sind die Anpassungen der nicht primär sensiblen Zellen meist gering.

Zellanpassung - CTA

- **Idee:**

- 1) jeder primär 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 Zellunterdrückung keine lückenhaften Tabellen, ausserdem sind die Anpassungen der nicht primär sensiblen Zellen meist gering.

- **weiterer Vorteil:** optimale Algorithmen existieren.

Zellanpassung - CTA

- **Idee:**

- 1) jeder primär 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 Zellunterdrückung keine lückenhaften Tabellen, ausserdem sind die Anpassungen der nicht primär sensiblen Zellen meist gering.
- **weiterer Vorteil:** optimale Algorithmen existieren.
- **Nachteile:** optimale Algorithmen nur brauchbar für sehr kleine Tabellen, Heuristiken existieren, garantieren aber keine Lösung.

Zellanpassung (CTA) - Beispiel

- ursprüngliche 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

- ursprüngliche 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 für die sensitiven Zellen

Zellanpassung (CTA) - Beispiel

- ursprüngliche 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 für die sensitiven Zellen

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

Zellanpassung (CTA) - Beispiel

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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

- ursprüngliche 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 zufällige 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 zufällige 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 zufällige Verschmutzung von Tabellenzellen basierend auf
 - Record-Keys
 - Cell-Keys
 - LookUp-Tabelle
- **Vorteil:** Konsistenz
- **Nachteil:** fehlende Tabellenadditivität

Zellanpassung (ABS)

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

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 ≈ 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 ≈ 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 ≈ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
 - großer Funktionsumfang mit graphischer Oberfläche
- sdcTable:
 - freie Implementierung von Algorithmen (vor allem) der Sekundärspernung 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 ≈ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
 - großer Funktionsumfang mit graphischer Oberfläche
- sdcTable:
 - freie Implementierung von Algorithmen (vor allem) der Sekundäersperrung 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 ≈ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
 - großer Funktionsumfang mit graphischer Oberfläche
- sdcTable:
 - freie Implementierung von Algorithmen (vor allem) der Sekundäersperrung 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 ≈ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
 - großer Funktionsumfang mit graphischer Oberfläche
- sdcTable:
 - freie Implementierung von Algorithmen (vor allem) der Sekundäersperrung 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 ≈ 15 Jahren Entwicklung in 2015 (teilweise) veröffentlicht
- großer Funktionsumfang mit graphischer Oberfläche

- sdcTable:

- freie Implementierung von Algorithmen (vor allem) der Sekundäersperrung 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.

über statistische Geheimhaltung finden Sie etwa bei:

- Allgemeines über 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 natürlich auch unter der Klappe 7988 (fast immer)

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

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

- Geheimhaltung von tabellarischen Daten ist ein sehr **komplexer Prozess**.
- Varianten für Geheimhaltung von Tabellen sind:
 - Zellunterdrückung
 - Runden
 - Zellanpassung

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

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