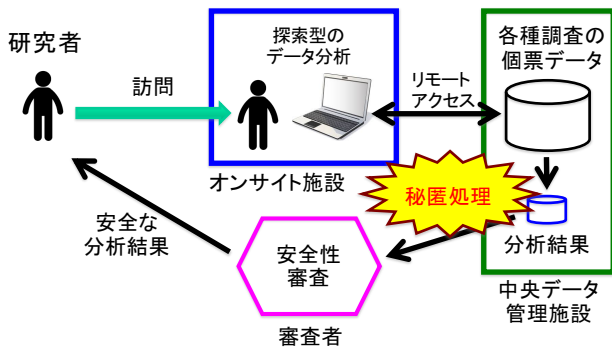


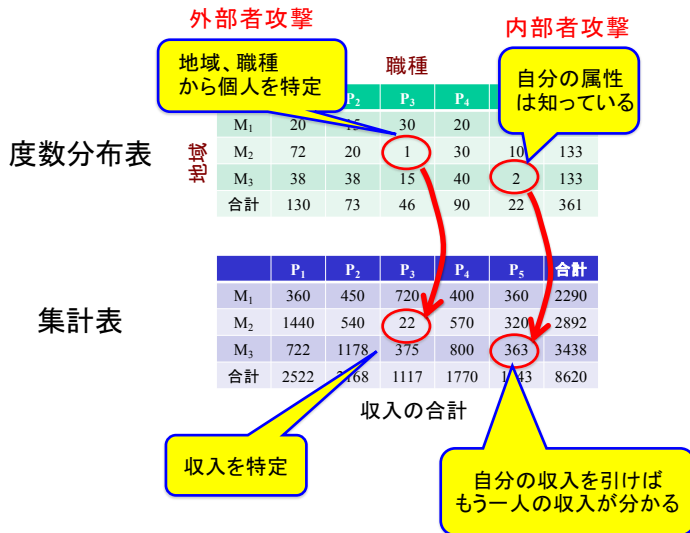
# 公的調査票情報を用いた表データ秘匿処理の実証的評価

南 和宏 モデリング研究系 准教授

## オンサイト利用の安全性審査



## 表データからの情報漏えい



## 表セルの1次秘匿

- 度数分布表**
- 最小度数ルール
- 集計表**
- 占有性ルール
  - (n, k)-ルール
  - p%ルール

## 表セルの2次秘匿

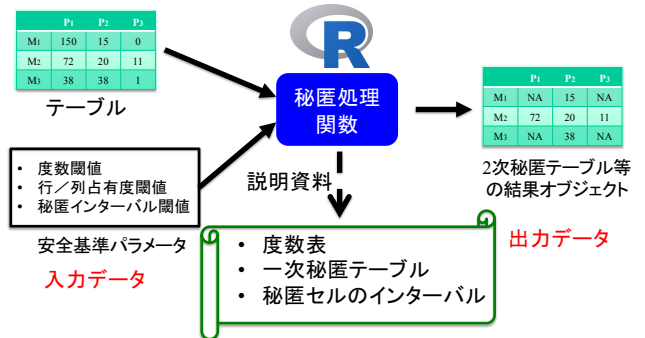
- 秘匿セル変数の可能範囲(秘匿インターバル)の幅wの長さがしきい値t(度数分布表では10)以上であること

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	合計
M <sub>1</sub>	x <sub>11</sub>	24	x <sub>13</sub>	72
M <sub>2</sub>	x <sub>21</sub>	38	x <sub>23</sub>	116
M <sub>3</sub>	40	39	42	121
合計	98	101	110	309

1. 最小値問題  
 $d_{23} = \min x_{23}$   
 拘束条件:  $x_{11} + x_{13} = 72 - 24$   
 $x_{21} + x_{23} = 116 - 38$   
 $x_{11} + x_{21} = 98 - 40$   
 $x_{13} + x_{23} = 110 - 42$   
 $(x_{11}, x_{13}, x_{21}, x_{23}) \geq 0$
2. 最大値問題  
 $d_{23} = \max x_{23}$   
 拘束条件:  $x_{11} + x_{13} = 72 - 24$   
 $x_{21} + x_{23} = 116 - 38$   
 $x_{11} + x_{21} = 98 - 40$   
 $x_{13} + x_{23} = 110 - 42$   
 $(x_{11}, x_{13}, x_{21}, x_{23}) \geq 0$

秘匿インターバル  $w = \max x_{23} - \min x_{23} = 68 - 20 = 48 > 10$

## R言語による秘匿処理ツール SDCLink の開発



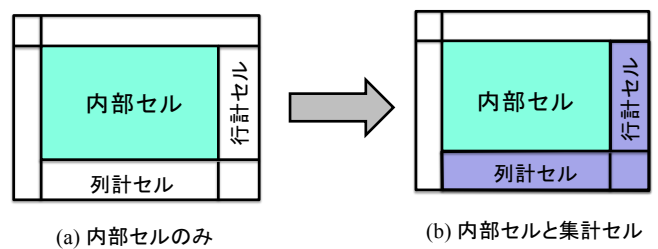
## 全国消費実態調査データによる検討課題

- 占有性(集中度)ルールの審査は必要か?
  - 欧米各国の統計局は企業系データに審査を限定
- 情報損失(秘匿セル)はどの程度生じるか?

## 秘匿セルの消費項目別内訳

消費項目	秘匿セル数
有価証券売却	26
財産購入	25
他の非消費支出	13
土地家屋借入金	12
他の借入金	12
分割払一括払購入借入金	11
保険金	10
住居	10
有価証券購入	10
実支出以外の支払のその他	10

## 秘匿対象セルの拡張



## 秘匿セル数による2つの手法の比較

