秘密分散・秘密計算技術を用いた公的統計ミクロデータ分析のセキュリティ強化に関する研究

一橋大学経済研究所　　NTTセキュアプラットフォーム研究所

研究概要

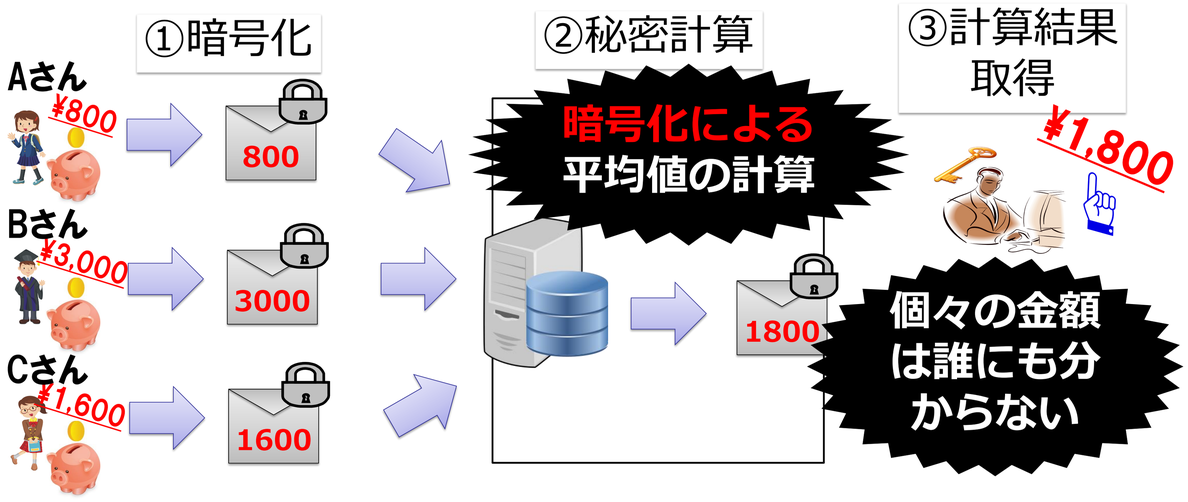
公的統計のミクロデータの二次利用として，場所を問わず即時に分析できる『オフサイト利用』が期待されている．しかしオフサイト利用の実現には，① 分析に用いるミクロデータのセキュリティ，及び ② 分析結果の開示制御，の対策が不可欠である．本研究では前記 ① の対策に絞り，データを暗号化（秘密分散）したまま分析が可能な『秘密計算』システムを一橋大学内に構築し，擬似データを用いた回帰分析のための実験及びその評価を行った．その結果，回帰分析において分析結果の精度が高く，処理速度が一定基準以上であることを実証した．

背景

公的統計ミクロデータにおける利便性向上の推進：『オフサイト利用』の実現がキーポイント

- ◆課題：
  分析に用いるミクロデータのセキュリティ
- ◆解決策：
  データの暗号化（秘密分散）による『秘密計算』技術を適用し，実装および実験評価

秘密計算のイメージ



研究成果

誰にもミクロデータを見せずに分析を実現

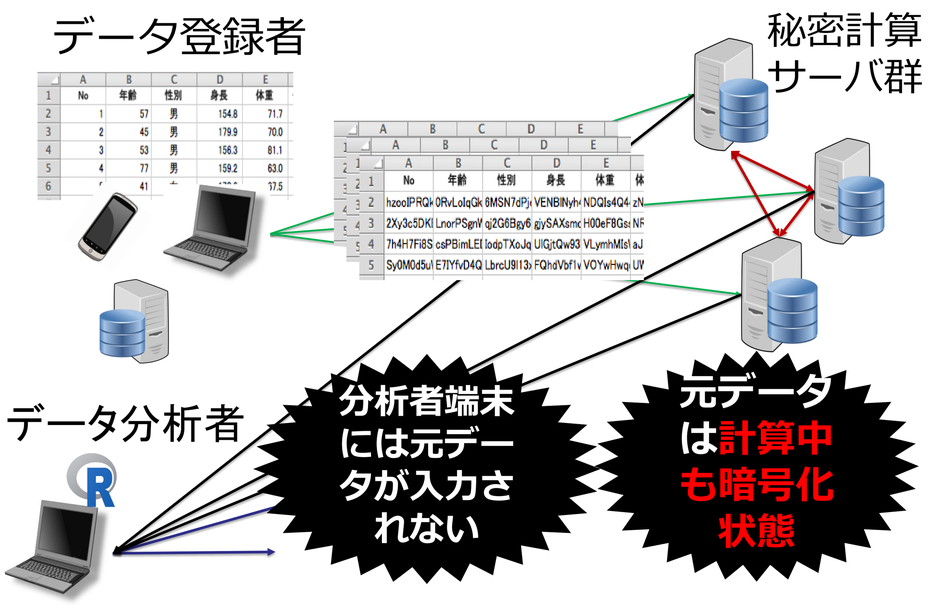
- ◆回帰モデルを秘密計算で高速に処理する手法の開発
- ◆秘密分散ベースの秘密計算を呼び出し実行できるシステムを用いて回帰モデルを実装
- ◆擬似ミクロデータ※を用いた秘密計算を実行し，実用的な処理時間および分析精度を実証

※平成16年全国消費実態調査

Rから実行した秘密計算のイメージ

```
R Console
>Sec.ChowTest(27,rec1,rec2,attrNo1=3,attrNo=4)

        Chow test

data:  就業人員1人 vs 就業人員2人
F = 22.0588, df1 = 2, df2 = 27368, p-value = 2.677e-10
alternative hypothesis: 就業人員1人 != 就業人員2人
95 percent confidence interval:
  0.05129339 2.99606

        Chow test

data:  就業人員1人 vs 就業人員3人
F = 6.7285, df1 = 2, df2 = 16859, p-value = 0.0012
alternative hypothesis: 就業人員1人 != 就業人員3人
95 percent confidence interval:
  0.05129345 2.996265

        Chow test

data:  就業人員2人 vs 就業人員3人
F = 0.3532, df1 = 2, df2 = 16405, p-value = 0.7025
alternative hypothesis: 就業人員2人 != 就業人員3人
95 percent confidence interval:
  0.05129345 6.885869
```

秘密分散ベースの秘密計算システム



実験環境

システム構成	
マシン	秘密計算サーバ 3台, 管理サーバ 1台, クライアント端末 1台
通信環境	1Gbps LAN
秘密計算サーバ	
OS	CentOS 6.4 (仮想マシン)
CPU	Intel Xseon E3-1220 v5 (3.00 GHz, 4コア)
メモリ	6 GB
クライアント端末	
OS	Windows 7 (64 bit)
CPU	Intel Core i7-6600U (2.60 GHz, 2コア)
メモリ	16 GB

実験結果

実支出額に対する収入総額の線形回帰

有業人員	世帯数	パラメータ		残差平方和	実行時間	
		収入総額	切片		暗号化なし	暗号化あり
全体	32,027	0.852	0.981	2313.3	23.3ms	1.02s
1人	13,913	0.851	1.012	992.9	11.1ms	1.07s
2人	13,459	0.864	0.831	947.0	9.5ms	1.05s
3人	2,950	0.854	0.953	224.9	2.0ms	0.46s

関連技術：秘密分散・秘密計算

- ◆秘密分散：
  元の情報からランダムな複数の断片データを生成
  1つの断片では元の情報を復元できない暗号技術
- ◆秘密計算：
  誰にもデータを見せずに計算ができる暗号応用技術
  - 秘密分散や準同型暗号に基づく方式などがある
  - 秘密計算の中では，現状，秘密分散に基づく方式が最も処理時間が短い

今後の予定

- ◆短期目標：
  集計表の開示制御ができるソフトウェア “τ-ARGUS” との連携システムを実装
- ◆中長期目標：
  - 開示制御機能を秘密計算で実現
  - ３台のサーバの管理者・設置拠点を分離することでセキュリティを向上

上記を具備したシステムによる実証実験