

# 2020年度の活動紹介（公的ミクロ事業）

社会データ構造化センター / 統計数理研究所  
南 和宏

第5回社会データ構造化シンポジウム  
2021年3月4日

# 公的マイクロデータの利用技術と 研究利用の環境整備

- 公的統計マイクロデータ研究コンソーシアム
  - 公的マイクロデータ分析の普及・啓発
  - 公的マイクロデータの研究利用促進に係る学官産連携の推進
  - オンサイト施設の設置・運用に係る課題の検討
- 公的マイクロデータの利用技術の基盤研究
  - データ統合, リンケージ技術
  - 秘匿処理, 匿名化技術

# マイクロデータ（調査票情報）とは？

総務省など国の行政機関で実施した統計調査の結果は、ホームページ（政府統計の総合窓口 e-Stat）等を通じて広く一般の方にご利用いただいています。

このような調査結果の提供に加え、公益性のある学術研究等にご活用いただくため、**調査対象の秘密の保護を図った上で、世帯単位や事業所単位といった集計する前の個票形式のデータ**を提供しています。

この個票形式のデータを**マイクロデータ（調査票情報）**と言います。

マイクロデータ（調査票情報）を用いることで、研究者の方々は、より自由で多様な分析を行うことが可能となるため、新たな発見につながることを期待されます。



行政機関による集計・公表

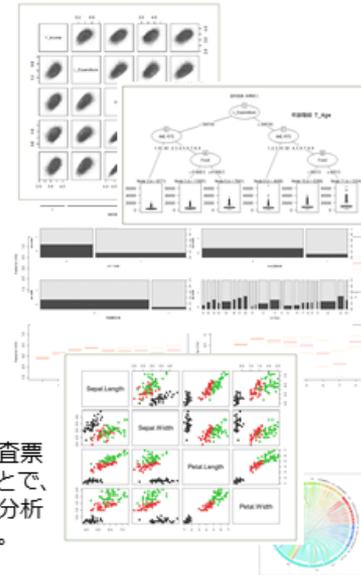
行政機関は、マイクロデータ（調査票情報）を集計して、調査結果を作成しています。調査結果は、「政府統計の総合窓口（e-Stat）」等を通じて公表・提供しています。  
<https://www.e-stat.go.jp/>

## マイクロデータ（調査票情報）のイメージ

	Weight	Y_Income	L_Expnd#Food	Housing	LFW	Furniture	Clothes	Health	Transport	Education
895.2667	3917	201649	47756	16028	9652	6702	8088	726	21546	0
895.2667	6675	166381	34054	7416	26333	17062	6989	7637	20773	0
895.2667	6706	259736	84501	1927	10082	6741	5090	11015	53372	0
895.2667	2790	114511	41684	730	22358	5413	1205	5049	17411	0
895.2667	2577	193505	56981	3779	28747	4812	4243	751	16435	0
895.2667	3452	152109	34924	3418	8131	4164	6970	4247	47698	0
895.2667	3233	136900	49956	203	15429	3659	22843	4365	8684	0
895.2667	9252	192439	68882	2832	23042	2988	5714	2052	37006	0
895.2667	2399	138415	53991	753	13072	5140	1366	5416	11593	0
895.2667	3059	79179	32853	14134	7977	3017	2364	1607	7630	0
895.2667	2324	243835	60528	28118	16392	4652	10759	37060	17666	0
895.2667	4524	241539	104433	5253	40637	7711	13833	3853	43003	0
895.2667	4415	207854	95504	7687	13801	19702	7239	3939	9547	0
895.2667	4162	185110	59798	565	12146	14552	12301	2486	26726	0
895.2667	16647	219935	81572	3704	21164	4944	10282	6573	7103	0
977.1795	6760	176625	43112	5013	11780	1697	3008	4836	12295	848
977.1795	6614	130803	32336	3149	14723	21002	12075	3447	24273	140
977.1795	6849	395294	66755	9524	19962	19777	10598	7398	56370	80
977.1795	6813	284803	86655	13479	15321	22966	1134	3649	88329	551
977.1795	6595	260459	46614	1646	15044	6415	16207	3709	76074	69

※データは擬似データです

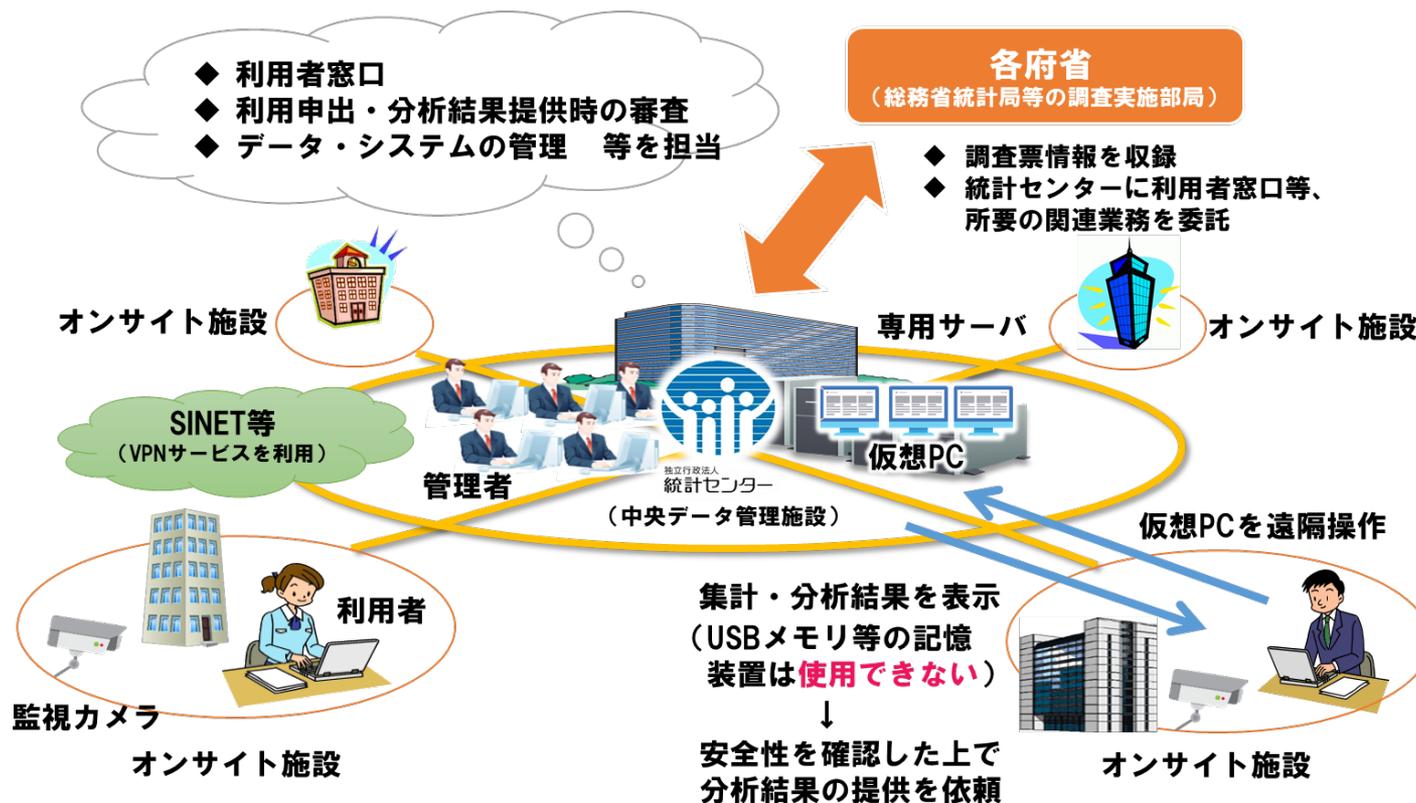
マイクロデータ（調査票情報）を用いることで、より自由で多様な分析が可能になります。



<https://www.e-stat.go.jp/microdata/micro>

# オンサイト利用全体の概要

イメージ

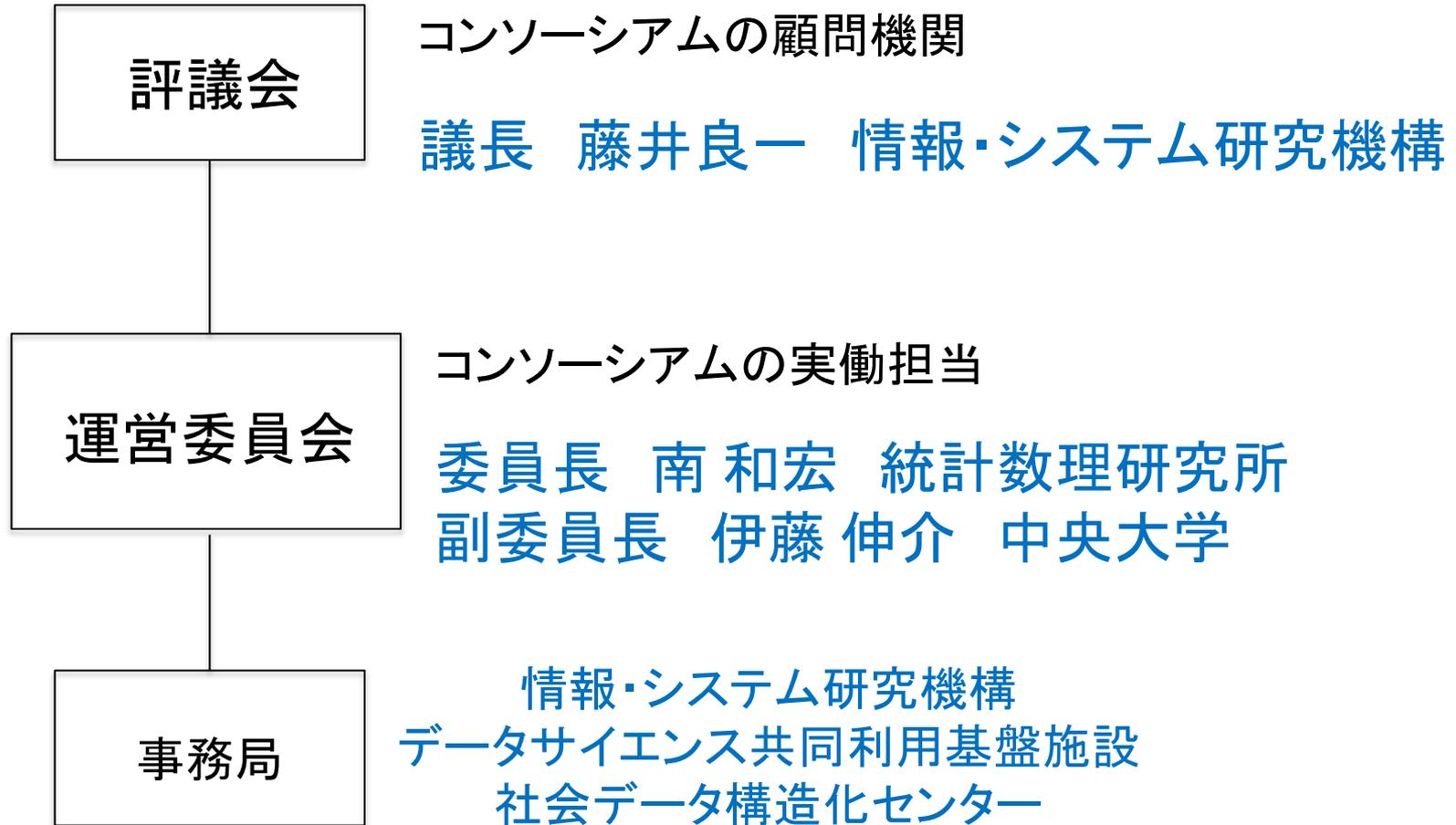


# 公的統計マイクロデータ 研究コンソーシアムの活動

# 公的統計マイクロデータ研究コンソーシアム

- 公的統計マイクロデータの研究利用（二次的利用）を促進するための環境の整備するために2017年に設立
  - 利用手続きの煩雑さ、利用機会の少なさにより、少数の研究者に利用が限定
- 公的統計マイクロデータの研究利用促進に係る**学官産連携**の推進
- 公的統計マイクロデータ分析の普及・啓発
- **オンサイト利用**によるマイクロデータの研究利用の推進

# 組織体制



# 評議会 評議委員名簿

- **議長** 藤井 良一 情報・システム研究機構 機構長
- 川崎 茂 日本大学経済学部 特任教授
- 北村 行伸 立正大学経済学部 教授
- 玄田 有史 東京大学社会経済研究所 教授
- 佐和 隆光 公益財団法人国際高等研究所 副所長
- 椿 広計 統計数理研究所 所長
- 岡部 寿男 京都大学学術情報メディアセンター 教授
- 松林 洋一 神戸大学経済経営研究所 研究科長
- 杉山 学 群馬大学社会情報学部 教授
- 南 和宏 統計数理研究所 教授
- 山下 智志 統計数理研究所 副所長/教授
- 渡邊 聡 広島大学 高等教育研究センター 教授

# 運営委員会 委員一覧

- 委員長 南 和宏 統計数理研究所 教授
- 副委員長 伊藤 伸介 中央大学経済学部 教授
- 山下 智志 統計数理研究所 副所長/教授
- 岡本 基 情報・システム研究機構戦略企画本部 主任URA/  
特任准教授
- 稲垣 好展 総務省政策統括官(統計基準担当)室 参事官
- 佐藤 紀明 総務省 統計局 統計調査部 調査企画課 課長
- 高部 勲 総務省 統計局 統計データ利活用センター セン  
ター長
- 三神 均 (独)統計センター 情報技術センター 統計情報提供  
課 課長

# 公的統計マイクロデータ 研究コンソーシアム

評議会

議長  
情報システム・研究機構  
藤井 良一

運営委員会

委員長  
統計数理研究所  
南 和宏

副委員長  
中央大学  
伊藤伸介

事務局

プラットフォーム  
分科会

データ構造化  
分科会

利用普及促進  
分科会

研究代表者：椿 広計 統計数理研究所 名誉教授  
研究分担者：山下 智志、南 和宏、岡本 基

統計数理研究所

一橋大学

青山学院大学

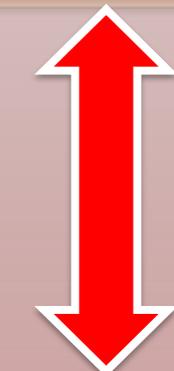
慶應義塾大学

科学研究費補助金 基盤研究 (A)  
「政府統計マイクロデータの構造化と  
研究利用プラットフォームの形成」

情報システム・  
研究機構  
データサイエンス  
共同利用基盤施設

国際マイクロ統計  
データベース

山下 智志、岡本 基、  
馬場 康維 名誉教授



(公財)  
統計情報  
研究開発  
センター

総務省  
政策  
統括官室

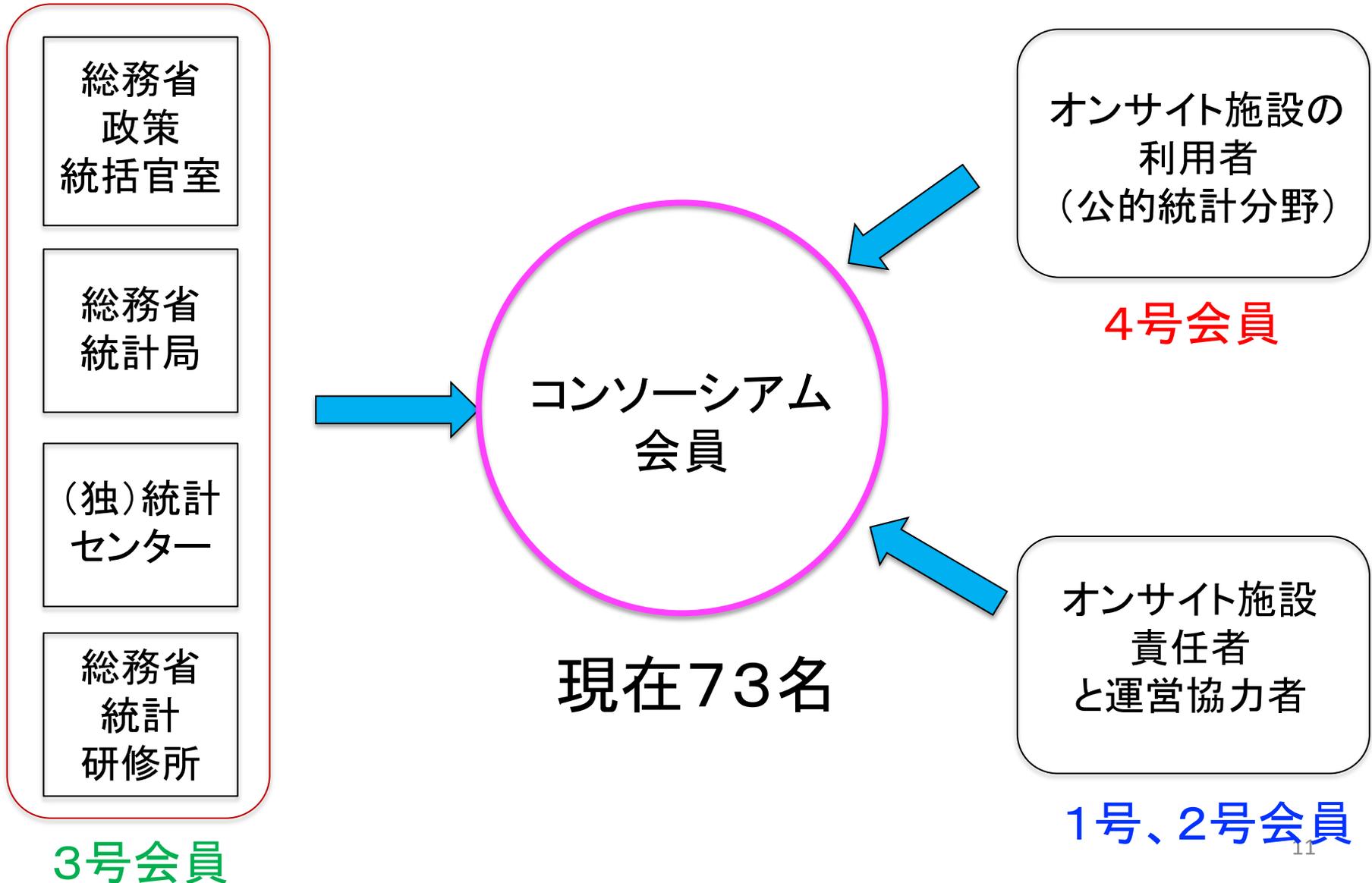
総務省  
統計局

(独) 統計  
センター

総務省  
統計研究  
研修所



# コンソーシアム会員



# 今年度の主な活動

- **コンソーシアム・シンポジウム開催（2020年11月）**
  - Zoomウェビナーによるオンライン開催
  - 参加者：111名
- **Webサイトのリニューアル**
  - URL: [jmodc.org](http://jmodc.org)
- **会員向けのアンケートの実施**
  - Googleフォーム回答によるオンラインアンケート
  - 実施対象：コンソーシアム会員（73名）
  - 回答数：24名
  - 質問内容：公的マイクロデータの利用状況，オンサイト施設の利用意向・経験
- **日本統計学会 第16回統計活動賞受賞**
  - 受賞記念講演を統計関連学会連合大会で実施

# シンポジウムの開催概要

## <午前部 10:00-11:30> オンサイト利用に関するチュートリアル

司会：岡本 基（情報・システム研究機構）

### 開会挨拶

南 和宏（統計数理研究所）

「統計データ利活用センターにおけるオンサイト利用推進の取組」

高部 勲（総務省統計局 統計データ利活用センター）

「調査票情報のオンサイト利用要件解説と利用事例紹介」

阿部 穂日（独立行政法人統計センター 統計情報提供課）

「公的マイクロデータ、因果推論とEBPM：オンサイト施設で国勢調査を利用した経験談」

張 俊超（統計数理研究所）

## <午後部 13:30-16:55> 公的統計マイクロデータ研究コンソーシアムシンポジウム

総合司会：南 和宏（統計数理研究所）

### 開催挨拶

藤井 良一（情報・システム研究機構）

## <第1セッション「統計行政とEBPM」>

司会：南 和宏（統計数理研究所）

「統計行政の最近の動向」

岩佐 哲也（総務省大臣官房）

「統計委員会の公的統計への取り組みと課題」

北村 行伸（立正大学）

「労働政策のあり方と統計整備—EBPMの役割をまじえて」

神林 龍（一橋大学）

「公的統計マイクロデータやアンケートなどを活用したEBPMの取組」

田邑 聡子、吉田 敦史（長崎県統計課）

## <第2セッション「マイクロデータのデータ結合と利用技術」>

司会：白川 清美（一橋大学）

「大規模低質データと小規模高質データのデータ結合による分析」

山下 智志（統計数理研究所）

「モデルに基づくアプローチによる小領域ごとの推定—公的マイクロデータ活用を目指して」

廣瀬 雅代（九州大学）

「マイクロデータの利活用と国際連携」

白川 清美（一橋大学）

## <参加者数と内訳>

登録区分	人数
大学・研究所	63名
官公庁	31名
民間企業	11名
学生	2名
その他	4名
合計	111名

# コンソーシアム・ウェブサイト (<http://jmodc.org>)

公的統計マイクロデータ  
研究コンソーシアム

トップページ  
コンソーシアム概要  
オンサイトネットワーク  
活動予定・報告  
お問い合わせ

お問い合わせ

お問い合わせはメールにて承ります。下記メールアドレス宛に以下の内容をご記載の上、お送りください。

公的統計マイクロデータの  
利活用推進に向けて

「公的統計マイクロデータ研究コンソーシアム」は、我が国における公的統計マイクロデータの研究利用（二次利用）を促進するために、学官産の関係機関が一体となり、取り組むことを目的とし設立するものです。

個人会員募集中

重要なお知らせ 2021.12.23  
[会員規則](#)を改正しました。

# アンケート集計結果概要

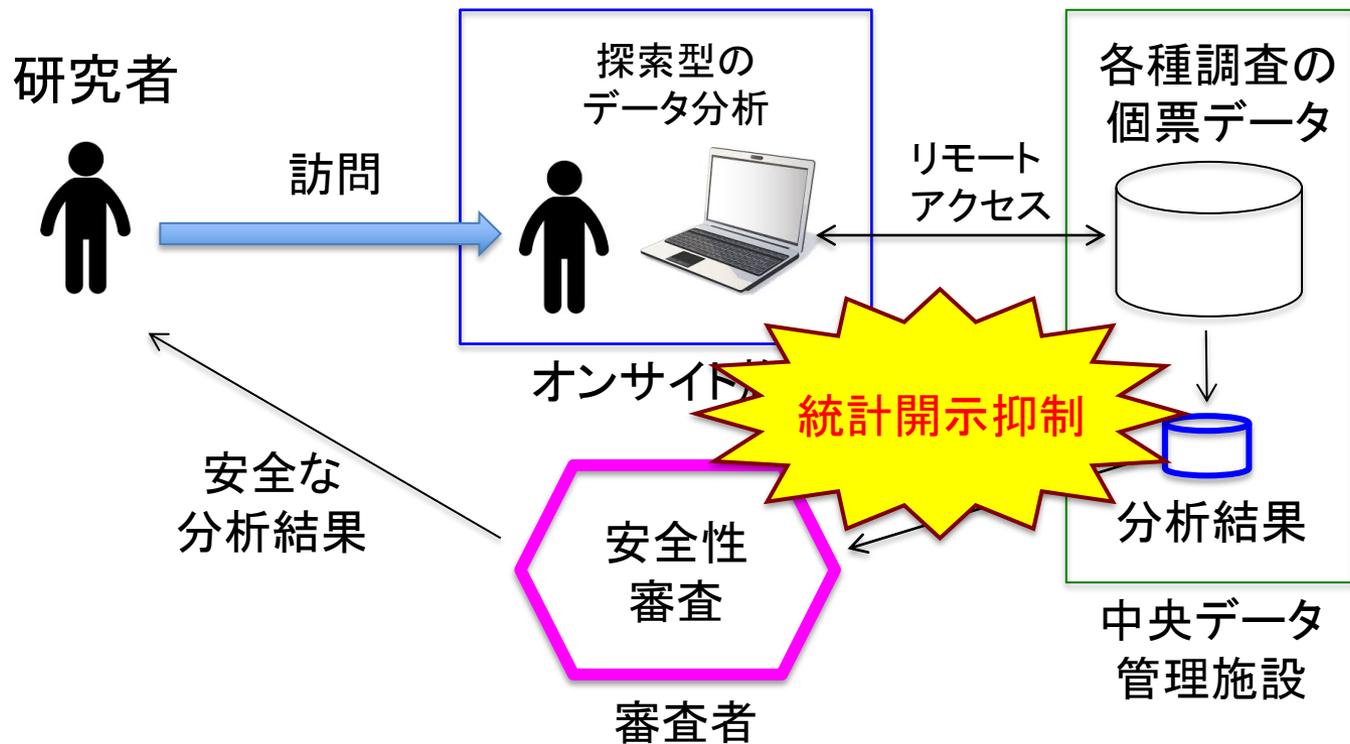
- 利用者が使う分析は、記述統計、グラフ可視化、回帰分析、多変量解析等の基本的手法が多い。
- 利用数が多い調査票情報は、国勢調査、全国消費実態調査、社会生活基本調査。ただし、匿名データでは、社会生活基本調査が一番利用数が多い。
- 分析ソフトは、Excel, RStudioに加え、Stata, SPSSもよく使われている。
- 約9割の回答者がオンサイト施設の利用意向あり。
- オンサイト施設の設置希望は、南関東に続き、北九州が多い。
- オンサイトの利用開始までの所要時間が3ヶ月以上が8件中4件あり、また申請手続きが書類の作成、やり取り等が大変であったという意見が半数程度あった。
- 現状、コンソーシアムに対しては、マイクロデータの具体的な利用方法、メリット、利用体験等に関する情報提供が期待されている。

# 今後の活動予定

- 会員向けの情報発信の強化
  - ニュースレターの発行
- 擬似データを用いたマイクロデータ分析のワークショップ開催
- 会員の要望(欲しいデータ等)を官側に伝える機会の提供
  - アンケート等による意見の集約
  - シンポジウム等による意見交換会の開催
- 技術的課題を検討するタスクフォースの設置
  - EBPM – どのような情報、分析結果がエビデンスとなりうるか？また対象領域は？
  - データリンケージ(構造化) – パネルデータ、行政記録情報との結合の実現性
  - 匿名化技術等

# 公的マイクロデータの 分析結果の秘匿処理

# オンサイト利用におけるプライバシー保護と安全性審査



# 統計表の例

度数表  
(職業・地域が  
一致する  
レコード数)

地域

		職業					
		P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	合計
地域	M <sub>1</sub>	20	15	30	20	10	95
	M <sub>2</sub>	72	20	1	30	10	133
	M <sub>3</sub>	38	38	15	40	2	133
	合計	130	73	46	90	22	361

数量表  
(収入の総和)

地域

		職業					
		P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	合計
地域	M <sub>1</sub>	360	450	720	400	360	2290
	M <sub>2</sub>	1440	540	22	570	320	2892
	M <sub>3</sub>	722	1178	375	800	363	3438
	合計	2522	2168	1117	1770	1043	8620

# 外部者による攻撃

度数表  
(職業・地域が  
一致する  
レコード数)

職業

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	合計
地域 M <sub>1</sub>	20	15	30	20	10	95
地域 M <sub>2</sub>	72	20	1	30	10	133
地域 M <sub>3</sub>	38	38	15	40	2	133
合計	130	73	46	90	22	361

数量表  
(収入の総和)

職業

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	合計
地域 M <sub>1</sub>	360	450	720	400	360	2290
地域 M <sub>2</sub>	1440	540	22	570	320	2892
地域 M <sub>3</sub>	722	1178	375	800	363	3438
合計	2522	2168	1117	1770	1043	8620

# 機密セルの秘匿には 2次秘匿処理が不可欠

集計表

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	合計
M <sub>1</sub>	20	24	28	72
M <sub>2</sub>	38	38	40	116
M <sub>3</sub>	40	39	42	121
合計	98	101	110	309

1次  
秘匿

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	合計
M <sub>1</sub>	20	24	28	72
M <sub>2</sub>	38	38	NA	116
M <sub>3</sub>	40	39	42	121
合計	98	101	110	309

2次  
秘匿

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	合計
M <sub>1</sub>	NA	24	NA	72
M <sub>2</sub>	NA	38	NA	116
M <sub>3</sub>	40	39	42	121
合計	98	101	110	309

占有性ルール  
を侵害

# 安全性の検証には、秘匿セル変数の可能区間(秘匿インターバル)の確認が必要

- 1次秘匿したセルの値の可能範囲がしきい値 $t$ 以上であることが必要
- 変数間の依存関係のチェックだけでは不十分

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

## 1. 最小値問題

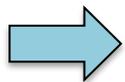
$$a_{23} = \min x_{23}$$

拘束条件:  $x_{11} + x_{13} = 72 - 24 = 48$   
 $x_{21} + x_{23} = 116 - 38 = 78$   
 $x_{11} + x_{21} = 98 - 40 = 58$       and  
 $x_{13} + x_{23} = 110 - 42 = 68$   
 $(x_{11}, x_{13}, x_{21}, x_{23}) \geq 0$

## 2. 最大値問題

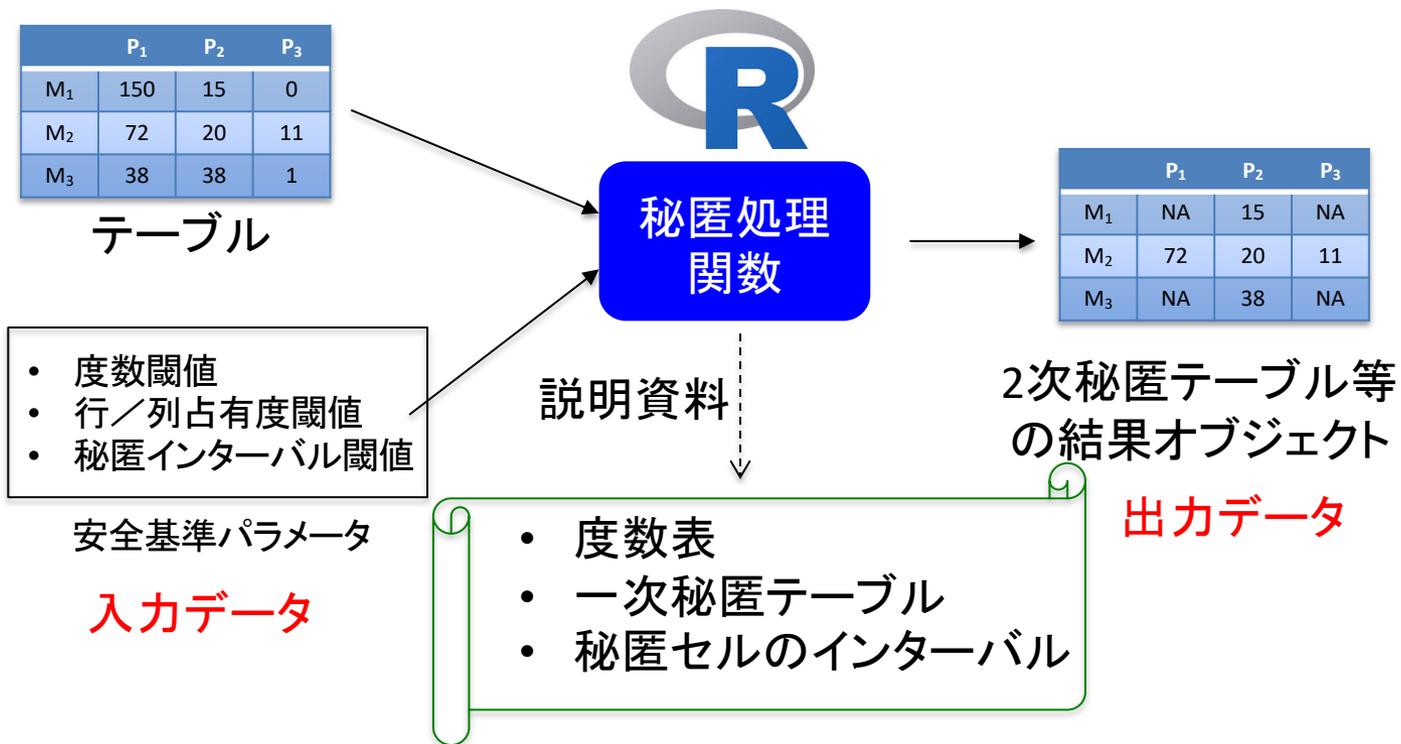
$$a_{23} = \max x_{23}$$

拘束条件  $x_{11} + x_{13} = 72 - 24 = 48$   
 $x_{21} + x_{23} = 116 - 38 = 78$   
 $x_{11} + x_{21} = 98 - 40 = 58$   
 $x_{13} + x_{23} = 110 - 42 = 68$   
 $(x_{11}, x_{13}, x_{21}, x_{23}) \geq 0$



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

# Rによる説明資料も作成する 秘匿処理ツールの開発



- 今後はオンサイト施設への展開を目指す
- 一部機能は、「簡易集計システム」で既に提供