

# 神戸大学におけるリモートアクセス型オンサイト利用等 に関する取り組み

勇上和史・中村健太

神戸大学大学院経済学研究科

2018年11月16日

共同研究集会「官民オープンデータ利活用の動向及び人材育成の取組」  
@統計数理研究所



## Contents

- 1 KUMiC の概要
  - 目的と沿革
  - 運営組織
- 2 業務概要
  - サテライト機関としての業務
  - KUMiC 独自の取り組み
- 3 今後の利用促進に向けた課題

# 神戸大学マイクロデータセンター (KUMiC) の概要

## 目的

- 独立行政法人統計センターの西日本で最初のサテライト機関として、公的統計の二次的利用を促進
- 政府統計や自治体統計等の公的統計のマイクロデータ分析拠点の形成

## 沿革

年月	内容
2009年4月	統計法（平成19年法律第53号）施行
2009年11月	国立大学法人神戸大学と独立行政法人統計センターとの間における連携協力に関する基本協定書締結
2010年4月	神戸大学マイクロデータ・アーカイブ (KUMA) 運営開始
2013年10月	兵庫県と神戸大学大学院経済学研究科並びに神戸大学経済経営研究所との間におけるKUMA施設利用に関する基本協定書締結
2016年12月	神戸大学マイクロデータセンター (KUMiC) に発展的に改組

3 / 27

## 運営体制

- 「神戸大学マイクロデータセンターに関する内規」を定め、経済学研究科と経済経営研究所が共同で運営

## KUMiC 運営委員会

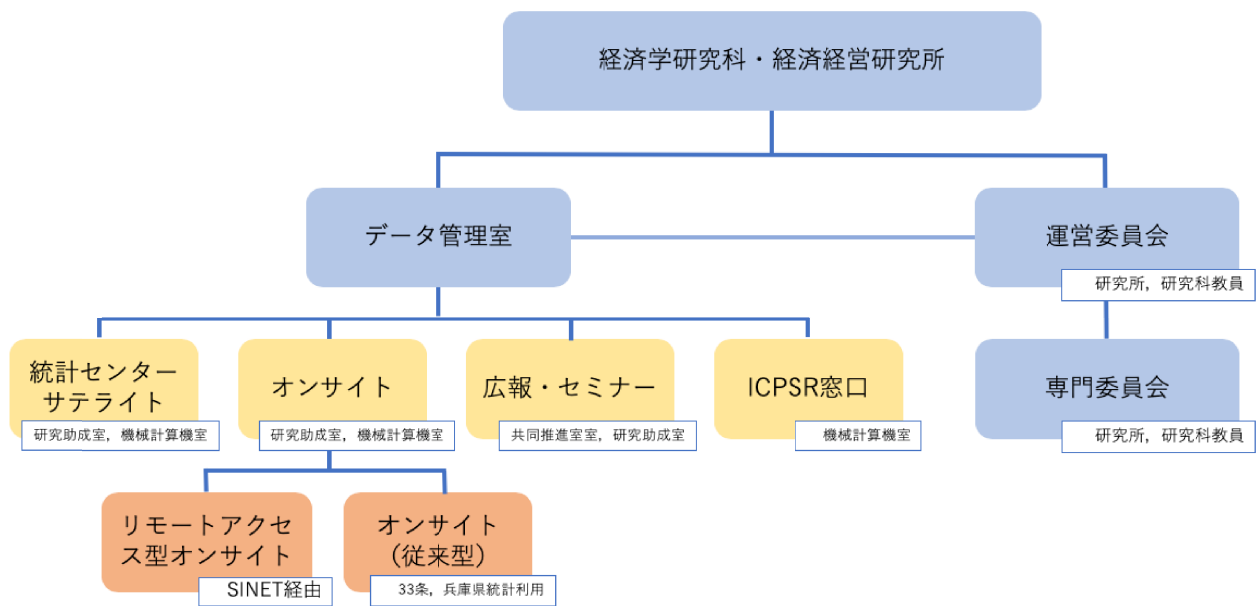
- 両部局長および両部局同数の教員で構成
- 年度計画、予算、連携協定等、運営全般に係る事項の審議

## KUMiC 専門委員会

- 両部局教員、客員教授（総務省統計局）、ネットワーク担当および研究助成担当で構成
- 匿名データの提供、セミナーの開催等、日常的な運営業務に係る事項を審議

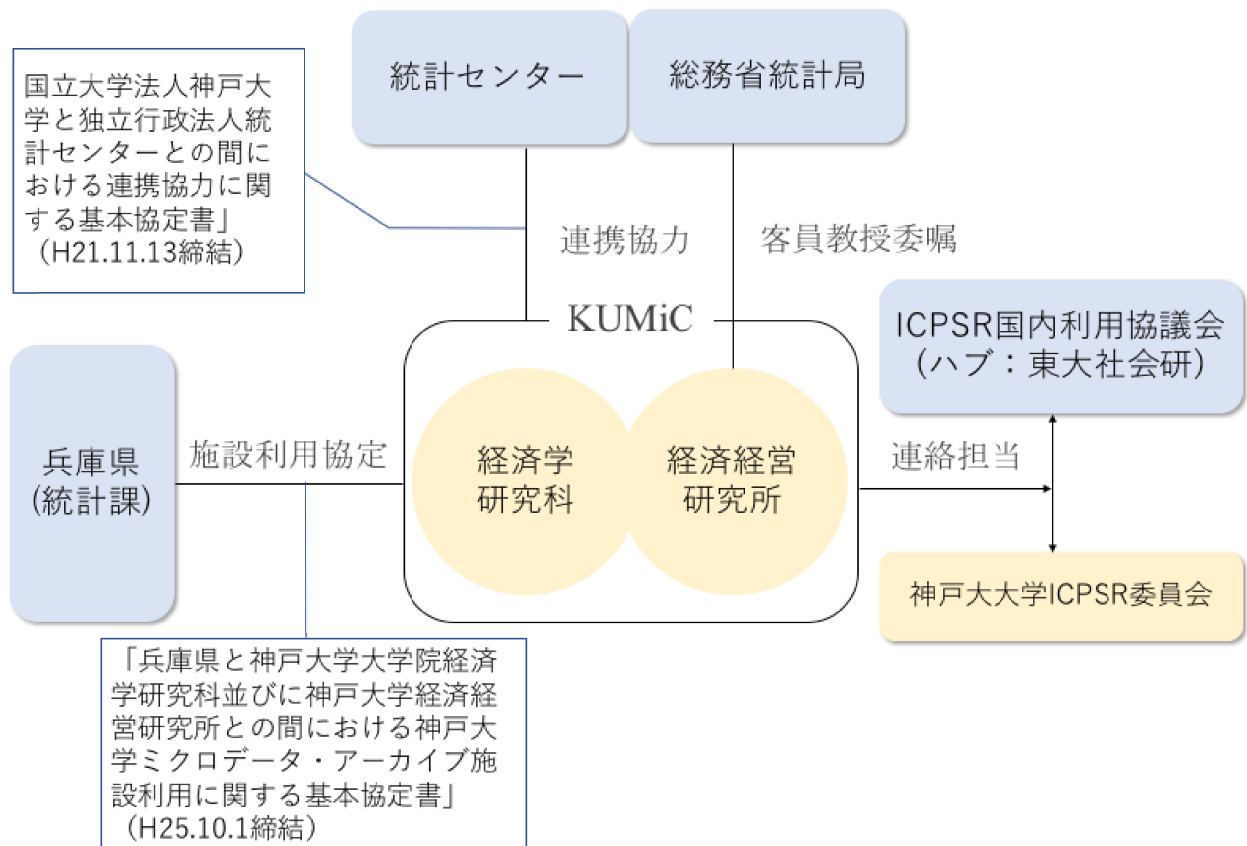
4 / 27

## 運営組織と業務



5 / 27

## 外部との連携協力関係



6 / 27

## 業務概要

### サテライト機関としての業務

- 総務省統計調査の「匿名データ」の提供
- 統計センターと連携した二次的利用の啓発・普及活動

### KUMiC 独自の取り組み

- 匿名データの研究利用の促進
  - 匿名データの加工・集計プログラムの作成
  - ミクロデータ分析に関する講義・演習
  - 匿名データの利用助成
- 公的統計のミクロデータ利用環境の整備
  - データ管理室における（独自の）オンサイト利用
  - (リモートアクセス型) オンサイト利用の運用
  - 兵庫県統計調査の利用に関する県との連携
  - ICPSR 国内利用協議会の神戸大学における連絡担当

7 / 27

## サテライト機関としての業務

### サテライト機関とは

- 法人の要件：国立大学，私立大学，独立行政法人，公共法人，公益法人
- 連携協力事項：統計センター統計データアーカイブの運営／公的統計の二次的利用に関する研究・開発・普及・啓発／人材交流等
- サテライト機関としての事務：匿名データの提供／オンサイト利用環境の提供，のいずれかまたは両方
- 施設等の要件：管理者・担当者の設置，セキュリティを施した施設・PC の設置等
- ※詳細は統計センター HP 「学術研究機関等との連携」を参照

### 現在のサテライト機関

- ①一橋大，②神戸大，③法政大，④情報・システム研究機構，⑤滋賀大。

8 / 27

# 匿名データの提供

## 匿名データ

- 個人や法人等が識別できないよう加工されたマイクロデータ
- 国勢調査, 全国消費実態調査, 社会生活基本調査, 就業構造基本調査, 住宅・土地統計調査, 労働力調査

## 運営体制

- 施設管理者：運営 2 部局の部局長の持ち回り
- 連絡担当者：経済学研究科総務係長ならびに KUMiC 担当教員
- 情報セキュリティ担当者：運営 2 部局の情報セキュリティ担当教員
- 複製担当者：経済学研究科研究助成室・データ管理室担当
  - 統計センターの研修を受講
  - データ提供に係る受付事務, データの複製と提供事務を実施

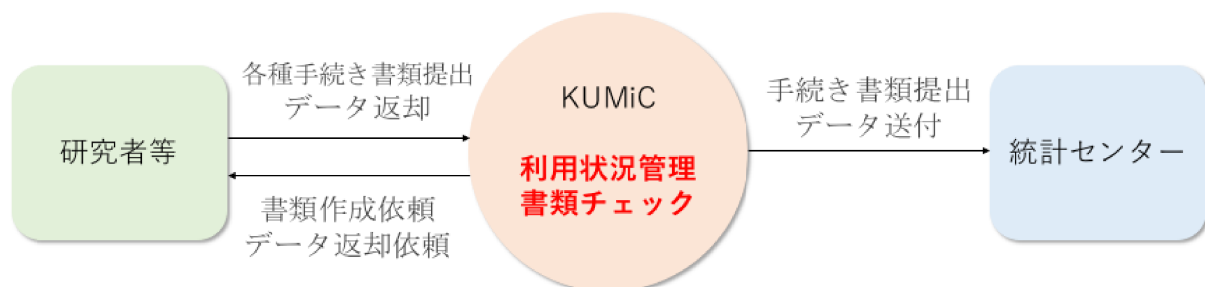
9 / 27

## 匿名データの提供に係る事務の概要

### 利用相談～提供まで（新規申出） ※ 1 件あたり 1～2 か月

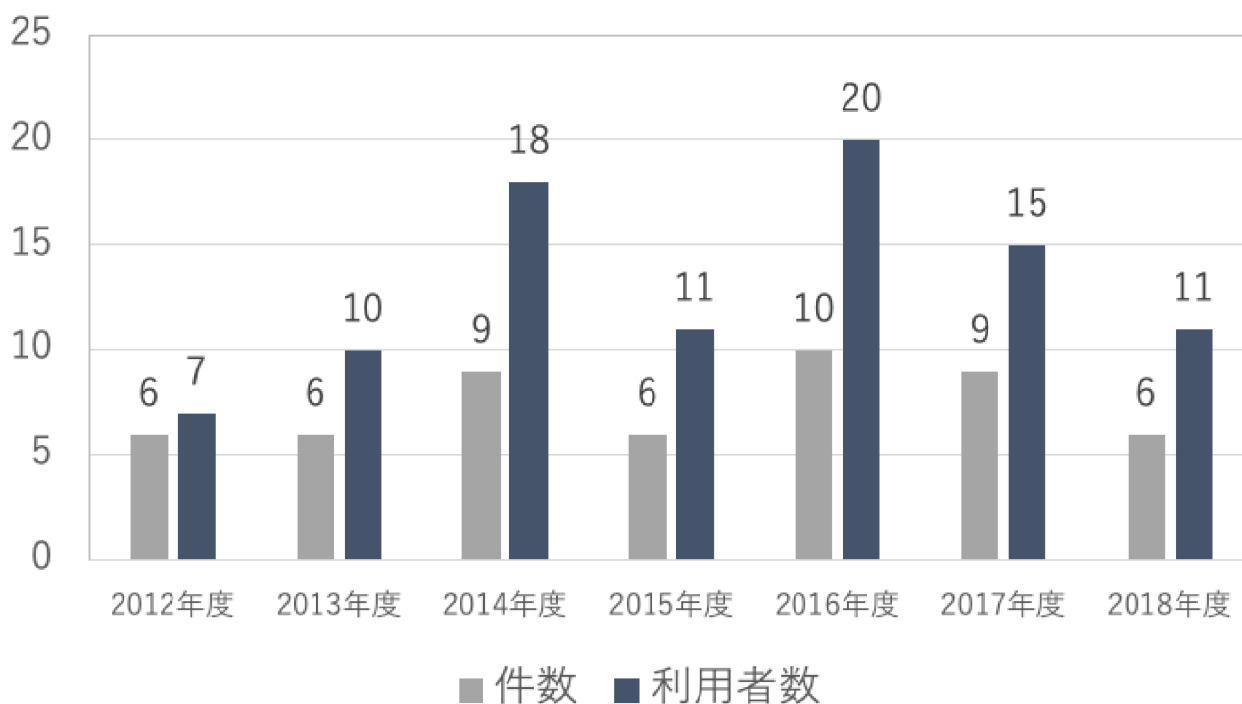


### データ提供後：内容変更, 利用延長・終了, 管理状況報告



10 / 27

## 匿名データの提供実績（新規提供分のみ）



注: 2018年度は10月末時点の実績

11 / 27

## 二次的利用の啓発・普及活動

### 統計センター・統計局と連携した学内イベント

年度	内容
2010	匿名データ利用説明会＋セミナー
2011	連続セミナー（介護関連統計・研究報告）
2012	セミナー
2013	セミナー＋匿名データ利用説明会
2014	連続セミナー（匿名データ＋その他のマイクロデータ）
2015	セミナー＋マイクロデータ利用促進ワークショップ
2016	セミナー＋マイクロデータ利用促進ワークショップ
2017	連続セミナー（匿名データ）＋セミナー＋マイクロデータ利用促進ワークショップ
2018	セミナー＋マイクロデータ利用促進ワークショップ（2019.1）

### 学外におけるイベントの共催

- 「二次利用共同研究集会」（於：統計数理研究所）
- 「匿名データ利用促進ワークショップ」（於：一橋大学）

12 / 27

# 匿名データの研究利用の促進

## 匿名データの加工・集計プログラムの作成

- 目的：KUMiC を通じた匿名データ利用者のデータ整理の簡便化を図る
- 内容：変数のラベル付け，文字列データの変換，ハンドリングしやすいデータへの分割・合併，公表値と整合的な分析用データの構築
- 形態：計量分析ソフト (Stata) のプログラム (do ファイル)
- 対象統計
  - 全国消費実態調査・匿名データ（平成元年～16年，DP 公開済）
  - 就業構造基本調査（平成4年～14年，検証済み提供可能）
  - 社会生活基本調査（調査票 A:平成3年～18年；調査票 B:平成13年～平成18年，検証済み提供可能）

13 / 27

## 例 1：完全な変数レイアウト表の作成

社会生活基本調査：A 票生活行動編，A 票生活時間編，B 票生活時間編

行番号	項目名	階層	位置	項目番号	バイト数	繰返し	配置	型	小数点	種別	対象	符号	符号内
629	その他の家族	5	409	132	4		3	1				0000～9999	その他の家族と行
630		5	413		1			2					
631	学校・職場・その他の人	5	414	133	4		3	1				0000～9999	学校・職場・その動した時間
632		5	418		1			2					
633	不詳	5	419	134	4		3	1				0000～9999	一緒にいた人が不問
634		5	423		1			2					
635	行動の種類 (EU区分・A区分)	1				36							
636	行動符号	2	22180	5574	3								

「繰返し」があるため、元のレイアウト表では途中の項目番号 (変数番号) がすぐにわからない

8

14 / 27

# 例 1：完全な変数レイアウト表の作成

すべての変数を記述したレイアウト表を作成し、フィルタ機能で変数番号の検索が容易に

1	項目A	項目B	項目C	項目D	項目E	項目番号
1	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	一人で (睡眠等を除	127	
2	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	父	128	
3	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	母	129	
4	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	子	130	
5	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	配偶者	131	
6	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	その他の家族	132	
7	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	学校・職場・その他の	133	
8	行動時間 (主行動+同時行動)	000:総数 (100,200,300,400,500,600)	一緒にいた人	不詳	134	
136	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	行動符号		135	
137	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	行動の有無		136	
138	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	インターネットの利用の		137	
139	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	場所	自宅	138	
140	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	場所	学校・職場	139	
141	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	場所	移動中	140	
142	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	場所	その他	141	
143	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	場所	不詳	142	
144	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	一緒にいた人	一人で	143	
145	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	一緒にいた人	一人で (睡眠等を除	144	
146	行動 主行動+同時行動 (該当するしない)	100:有償労働 (110,120,131,140の合	一緒にいた人	父	145	

# 例 2：時間帯のデータフレームを wide から long に

社会生活基本調査：A 票生活時間編, B 票生活時間編

0時～0時15分 15分～

個体番号	v4	v5777	v5778	v5779	v5780	v5781	v5782	v5783	v5784	v5785	v5786	v5787	v5788	v5789	v5790	v5791	v5792	v5793
1	1	11	1	20	17	12	12	12	14	18	17	14	21A	15	21A	2	13	17
2	2	16	1	18	16	19	12	12	16	16	12	16	17	15	10	2	18	17
3	3	16	1	13	19	17	11	16	19	16	18	19	18	12	15	2	10	11
4	4	15	1	12	17	13	18	13	14	10	17	14	16	14	14	2	12	15
5	5	13	1	18	16	18	11	13	10	16	15	15	20	14	12	2	17	11
6	6	13	1	14	17	13	11	19	13	18	12	19	19	10	12	2	13	11
7	7	11	1	18	20	13	19	17	17	17	16	11	14	12	12	2	19	14
8	8	19	1	13	14	16	19	13	10	15	20	14	12	12	13	2	16	18
9	9	12	1	14	14	18	16	10										
10	10	17	1	14	18	13	13	10										
11	11	14	1	19	13	20	12	16										
12	12	15	1	16	19	19	13	12										
13	13	15	1	12	16	11	19	17										
14	14	12	1	14	14	14	19	11										

2006年B票生活時間編の場合  
元のデータフレームでは、行動時間に関する変数は1248個

個人を1レコードに収めるために、変数をすべて横に広げており取り扱いが困難

※表示されているデータは、ランダムに発生させたダミー・データです



## 例 2：時間帯のデータフレームを wide から long に

3	4	5	0
項目C	項目D	項目E	項目番号
午前0時～午前0時15分	時間区分		5778
午前0時～午前0時15分	インターネットの利用		5779
午前0時～午前0時15分	場所		5780
午前0時～午前0時15分	一緒にいた人	一人で（一緒にいた	5781
午前0時～午前0時15分	一緒にいた人	父（一緒にいた人）	5782
午前0時～午前0時15分	一緒にいた人	母（一緒にいた人）	5783
午前0時～午前0時15分	一緒にいた人	子（一緒にいた人）	5784
午前0時～午前0時15分	一緒にいた人	配偶者（一緒にいた	5785
午前0時～午前0時15分	一緒にいた人	その他の家族（一緒に	5786
午前0時～午前0時15分	一緒にいた人	学校・職場・その他の	5787
午前0時～午前0時15分	主行動・行動符号		5788
午前0時～午前0時15分	同時行動の有無		5789
午前0時～午前0時15分	同時行動・行動符号		5790
午前0時15分～午前0時30分	時間区分		5791
午前0時15分～午前0時30分	インターネットの利用		5792
午前0時15分～午前0時30分	場所		5793
午前0時15分～午前0時30分	一緒にいた人	一人で（一緒にいた	5794
午前0時15分～午前0時30分	一緒にいた人	父（一緒にいた人）	5795
午前0時15分～午前0時30分	一緒にいた人	母（一緒にいた人）	5796
午前0時15分～午前0時30分	一緒にいた人	子（一緒にいた人）	5797
午前0時15分～午前0時30分	一緒にいた人	配偶者（一緒にいた	5798
午前0時15分～午前0時30分	一緒にいた人	その他の家族（一緒に	5799
午前0時15分～午前0時30分	一緒にいた人	学校・職場・その他の	5800
午前0時15分～午前0時30分	主行動・行動符号		5801
午前0時15分～午前0時30分	同時行動の有無		5802
午前0時15分～午前0時30分	同時行動・行動符号		5803
午前0時30分～午前0時45分	時間区分		5804

変数が多くレコード数が少ないwide型

⇒変数が少なくレコード数が多いlong型に変換

17 / 27

## 例 3：調査年をまたいだ定義の統一

### 就業構造基本調査の家族類型

#### 1992年 v10

1	夫婦のみの世帯
2	夫婦と子供からなる世帯
3	夫婦と親からなる世帯
4	夫婦と子供と親からなる世帯
5	その他の世帯+単身世帯

#### 1997年 v10

1	夫婦のみの世帯
2	夫婦と両親からなる世帯
3	夫婦と片親からなる世帯
4	夫婦と子供からなる世帯
5	夫婦、子供と両親からなる世帯
6	夫婦、子供と片親からなる世帯
7	上記以外

#### 2002年 v10

01	夫婦のみの世帯
02	夫婦と両親からなる世帯
03	夫婦とひとり親から成る世帯
04	夫婦と子供から成る世帯
05	夫婦、子供と両親から成る世帯
06	夫婦、子供とひとり親から成る世帯
07	ひとり親と子供からなる世帯
08	兄弟姉妹からなる世帯
09	単身世帯
10	母子世帯
11	父子世帯
12	上記以外

同じ変数だが年次により定義が異なる

⇒変数の値の階級が大きな年次のものに合わせる

18 / 27

## マイクロデータ分析に関する講義・演習

- 目的：経済学研究科院生ならびに経済学部生を対象として、**公的統計等のマイクロデータ利用を直接・間接的に支援する**
- 内容
  - 公的統計のメタデータと匿名データの解説
  - ミクロデータ分析モデルの講義
  - リサーチ・プロポーザルの指導
  - R や Stata を用いたデータの加工・推計の実践的指導
- 実績
  - 2015 年度 (D3 名, M3 名, 学部 1 名)
  - 2016 年度 (D1 名, M3-5 名, 学部 1-4 名)
  - 2017 年度 (大学院 58 名\*, 学部 126 名\*) \*のべ人数
  - 2018 年度 (大学院 55 名, 学部 40 名\*) \*のべ人数

19 / 27

## 公的統計の匿名データの利用助成

- 目的：経済学研究科院生ならびに経済学部生の**公的統計の匿名データを利用した実証研究を支援する**
- 内容
  - 匿名データの利用に係る手数料を（一部）補助
  - 匿名データ：総務省の 6 調査, ならびに厚生労働省「国民生活基礎調査」
  - 経済学研究科へのご寄附（「高原正之奨学基金」ならびに「小林義雄奨学寄附金」）によるもの
- 実績
  - 2017 年度：1 名（博士後期課程, 社会生活基本調査）
  - 2018 年度：1 名（学部生, 社会生活基本調査）

20 / 27

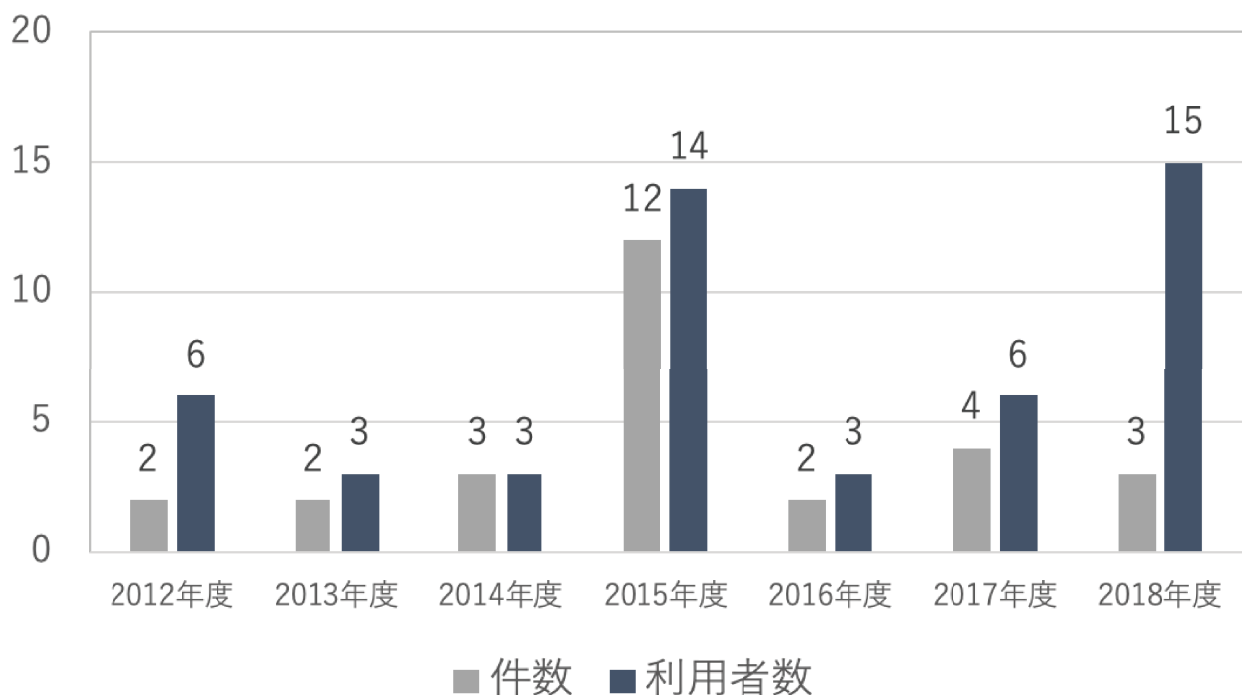
# 公的統計のマイクロデータ利用環境の整備

## データ管理室における（独自の）オンサイト利用

- 目的：統計法第 33 条による公的統計の調査票情報の利用等，**機密性の高いマイクロデータの利用環境を提供**
- 対象：学内教員，大学院生（学外者の兵庫県統計調査の二次利用も可）
- 内容
  - データ管理室内クライアント PC にてマイクロデータを分析
  - 「データ管理室管理要領」，「同利用要領」により運用
  - 利用時間：月～金（10 時～17 時），土（事前申出により対応）
- 利用環境
  - セキュリティ環境：カードキーによる入退室管理，監視カメラ
  - データ・サーバー：施錠可能なブース内に設置
  - クライアント PC（6 台）：外部ネットワークから遮断
  - ソフトウェア：Stata, Eviews, MS-Office

21 / 27

## オンサイト利用の実績（新規利用分のみ）



注：2018 年度は 10 月末時点の実績

22 / 27

## 公的統計の調査票情報の（リモートアクセス型）オンサイト利用

- オンサイト施設の認証
  - オンサイト施設運用要綱の締結
  - オンサイト施設のセキュリティに係る認証
- オンサイト利用の試行運用（平成 29 年 2 月～12 月）
  - 統計法 33 条に基づく調査票情報のオンサイト利用
  - 利用データ：就業構造基本調査（平成 24 年）
  - 内容：東日本大震災の影響による避難・移動と、調査時点における就業の関係の検証
  - 利用ソフトウェア：R（R Studio）
  - **いくつかの課題のフィードバック**
    - 中央データ管理施設（統計センター）の仮想 PC の処理能力
    - 統計ソフトの利用環境整備の必要性，等
- オンサイト利用の（本格）運用（平成 30 年 1 月～）
  - KUMiC 関連の研究プロジェクトにて，チームでのデータ利用を検討中

23 / 27

## 兵庫県統計調査の利用に関する県との連携

「兵庫県と神戸大学大学院経済学研究科並びに神戸大学経済経営研究所との間における神戸大学マイクロデータセンター施設利用に関する基本協定書」（平成 25 年 10 月 1 日締結）

### 連携内容

- 調査票情報の二次的利用に関する KUMiC の活用
- 調査票情報の二次的利用に関する研究・開発
- その他本協定の目的を達成するために必要な事項

⇒ **兵庫県独自の統計調査について，県統計調査条例に基づき利用が承認された場合は，KUMiC データ管理室にてオンサイトが利用**

24 / 27

## 今後の利用促進に向けた課題

### 匿名データの利用活用に向けて

- 調査の種類が抱負、より「リアルな」データ
- 匿名化処理により一部の情報に制限があるが、特に学部上回生～修士程度にとっては、本格的な実証研究への架け橋になりうる

⇒ **高等教育利用がしやすい条件（料金等）の整備**

### リモートアクセス型オンサイト利用の普及に向けて

- 中央データ施設
  - 仮想 PC の処理能力の向上
  - ソフトウェアの利用環境の整備
- オンサイト施設：ヒト・モノ・カネ
  - 利用時間の課題、特に夜間や土日
  - 複数キャンパスでの施設設置・運用

25 / 27

## 統計調査への理解を深めるためのデータ利用

### 匿名データの利用要件

- 統計の作成または統計的研究のみに利用されること
- 学術研究または高等教育目的
- **現状、統計調査やデータ自体の分析・研究のための貸出は受けられない**
  - KUMiC の匿名データの加工・集計プログラムの作成過程では、ダミー・データを作成して一部利用

### 利用目的や公益性認定の緩和に期待

- 「試験研究の例外」的な発想の導入
- リモート型オンサイトの利用規程についても同様

26 / 27

神戸大学大学院経済学研究科研究助成室  
神戸大学マイクロデータセンター担当

- 住所：〒 657-8501 兵庫県神戸市灘区六甲台町 2-1
- 直通電話：078-803-6854
- E-mail：satellite@econ.kobe-u.ac.jp
- URL： <http://www.econ.kobe-u.ac.jp/kumic/>
- 開設期間：土，日，祝日，年末年始の期間及びホームページ等で事前に周知した休業日を除く平日  
※ただし，申出の受付期間は 4 月 1 日～翌年 2 月末日
- 利用時間：10:00～17:00 (12:00～13:00 を除く)