

日本学術会議学術フォーラム
「デジタルデータ・社会調査データの公共的な利活用に向けて」

「社会科学における公的データの 現代的意義と可能性 ～農林業センサスの事例から」

2023年9月24日（日）オンライン

村上あかね（桃山学院大学、日本学術会議連携会員）

【謝辞】本研究はJSPS科研費19K02080＜基盤研究（C）「若手農家の参入時における家族の役割の解明—大規模世帯員パネルデータの利用—」（研究代表者・西村教子 公立鳥取環境大学）＞の助成を受けたものです。研究チームメンバー（西村教子先生、仙田徹志先生）、学術会議分科会メンバー、故・佐藤朋彦先生（総務省統計局・東京大学社会科学研究所）に感謝申し上げます。

趣旨

現代において人文・社会科学が扱うデータに大きな変化が起こっています。人文科学ではデジタルヒューマニティーズの興隆、社会科学においてはWeb調査の台頭、公的データの利用可能性の向上、いわゆるビッグデータを用いた計算社会科学の興隆などがあります。

個々の領域で起こっているこれらの変化は公共的な利活用を推進するチャンスでもあります。

本フォーラムでは、様々な分野で新しいタイプのデータを利活用している研究者を演者として招き、相互の交流を通じて公共的な利活用の可能性を探求します。

自己紹介

- 専門 家族社会学、社会階層論、調査方法論
- 問題関心 後期近代の福祉国家再編とハウジングの不平等、家族形成
- 経歴
 - (1) 家計経済研究所
 - 「消費生活に関するパネル調査」の企画・設計
 - 2次分析をするユーザー対応
 - (2) 東京大学社会科学研究所
 - 「働き方とライフスタイルの変化に関する全国調査」の企画・設計
 - SSJデータアーカイブ (SSJDA) 寄託者対応
 - (3) 桃山学院大学
 - 政府統計委員 (内閣府、総務省)
 - 学部生と大学院生にE-statの活用や2次分析を推奨。講演会開催
 - Euro-statの個票データ利用機関認定 (村上 2020)

1.目的

- 公的データの現代的意義と利用可能性について、農林業センサスの2次分析の経験をもとに考察・提言する
- 公的データ = 伝統的な社会調査 + ビッグデータ
 - 伝統的な社会調査
 - 全数調査 / ランダムサンプリング
 - 訪問面接法 / 留置法 / 郵送法
 - 紙と鉛筆 / 紙とコンピューター / インターネット (モニター)
 - ビッグデータ = 「デジタルに記録された人間行動の大規模でマイクロなデータ」 (日本学術会議社会学委員会Web調査の課題に関する検討分科会, 近刊)
 - 2次分析 (または二次利用) = 「既存データを使用して調査研究を行う方法」

1.目的

- 農林業センサスの分析例

- 公立鳥取環境大学と京都大学との共同研究による結果に基づく特別なデータセットを利用
- 計算結果は西村教子・山口幸三・吉田嘉雄・仙田徹志・村上あかね, 2022, 「小規模農家の親族継承の特徴——農業センサスパネルデータによる分析」(2022年度統計関連学会連合大会報告)に基づく

◆本日の問い

- 公的データの利活用にはどのような可能性と意義があるのか？
- 課題は何か？

2. 問題の背景

- 伝統的な社会調査の困難と海外の動向
→ データアーカイブの整備と2次分析の普及

1990年 SORD

1998年 JGSS／NFRJ／SSJDA

2000年頃 家計経済研究所

→ GCOE → 慶應義塾大学パネルデータ設計・解析センター

2007年 統計法改正

2010年 SRDQ／RUDA

2021年 JDCat（人文学・社会科学総合データカタログ）

2次分析の利点と課題

利点

- ①被調査者の負担軽減
- ②データ散逸の防止
- ③研究の発展
 - 新しい問題意識
 - 仮説～分析～解釈に専念
 - 大規模データによる精緻な分析
- ④調査の質の向上

課題

- ①2次分析への理解
 - ②寄託への理解
 - ③アーカイブに限らず、組織の安定的な運営
- 分析結果の再現可能性やそれに関連した懸念は別途要検討

ビッグデータの特徴



サルガニック, 2019『ビット・バイ・ビット』有斐閣
(同社ウェブサイトより)

- ビッグデータ = 研究目的のために作られたデータではない
- 「見出されたデータ」
レディーメイド ⇔ カスタムメイド
- 「農林業センサスは、（中略）我が国農林業の生産構造及び就業構造等の実態や農山村地域の現状を把握することにより、農林業に関する諸統計調査に必要な基礎資料を整備するとともに、（中略）各国農林業との比較において我が国農林業の実態を明らかにすることを目的とする」（農林水産省）

ビッグデータの特徴（続き）

ビッグデータの特徴	該当	備考
①「巨大であること」	○	
②「常時オンであること」	—	
③「非反応的であること」	—	
④「不完全性」	○	
⑤「アクセス不能性」	○	一定の制限あり。利益相反の可能性は低い
⑥「非代表性」	△	「外形基準」 標本調査の場合の回収率
⑦～⑨「ドリフト」「汚染」「アルゴリズムによる交絡」	△	データクリーニング
⑩センシティブ	△	
その他 「概念の妥当性」	△	レディーメード

3. 分析事例（2005～2015年）

- 農林業センサス
 - ┌ 農家経営体調査
 - └ 農山村地域調査（農業集落用・市区町村用）

- 農家経営体調査

回答数（全国47都道府県）

2005年 2,085,086

2010年 1,726,751

2015年 1,404,488

分析対象など

- 分析対象

- 家族経営の販売農家「経営耕地面積が30 a 以上又は調査期日前 1 年間に
おける農産物販売金額が50万円以上の農家」
- 2005年の「新規参入農家」≠「新設農家」
- 除く 北海道・福島・沖縄。除く 複数経営。除く「復活」ケース。

- 継続パターンの類型化

- ベースとなるケース数

経営者50歳以上 26,573

経営者50歳未満 5,332

合計 31,905

符号表 (一部)

baa9515符号表_農林業センサ共通調査票(農家)

最終更新日:2018年6月8日

政府統計コード	00500209	実施時期	1995年、2000年、2005年、2010年、2015年	作成日	2018年6月8日	最終更新日	2019年3月25日
統計調査名	農林業センサ	集計区分		ファイル名			
調査票名	共通調査票(農家)	(備考・補足事項)		(照会先等)			
コード体系							
レコード長							

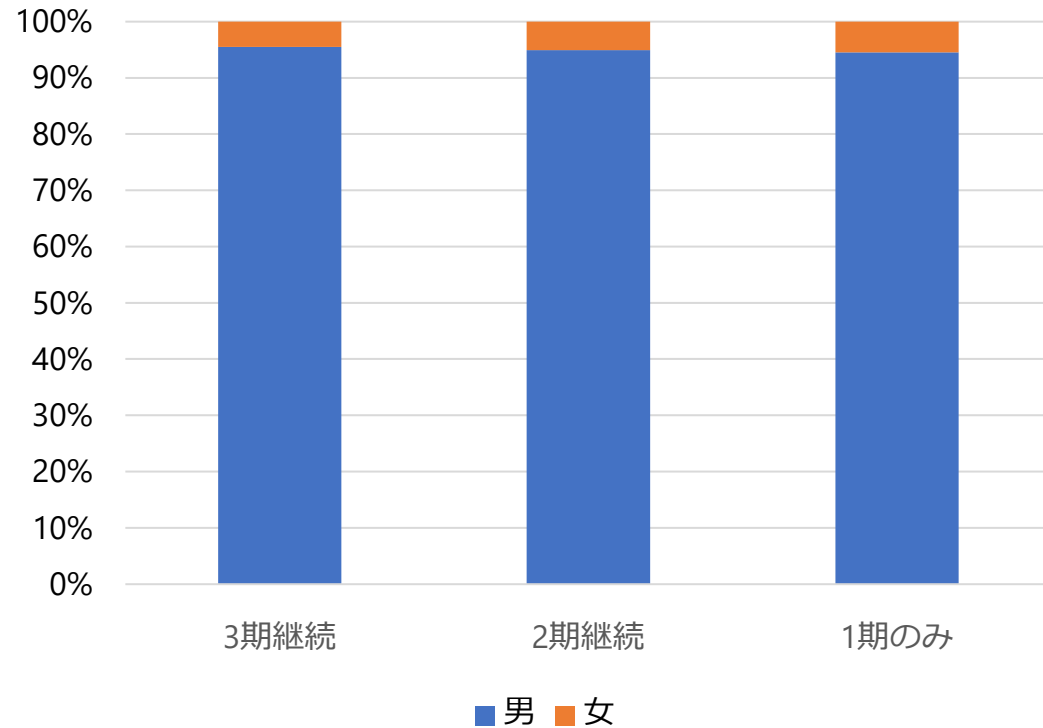
行番号	共通変数名	共通項目名	変数属性	共通符合	共通符合内容	2015変数名	2010変数名	2005変数名	2000変数名	1995変数名	レコードの有無	備考
1	baa9515_0001	新道府県	1			bae15_001	bap10_001	bap05_001	baa00_001	baa95_001		
2	baa9515_0002	支庁	1			bae15_002	bap10_002	bap05_002	baa00_002	baa95_002		
3	baa9515_0003	郡	1			bae15_003	bap10_003	bap05_003	baa00_003	baa95_003		
4	baa9515_0004	市区町村	1			bae15_004	bap10_004	bap05_004	baa00_004	baa95_004		
5	baa9515_0005	旧市区町村	1			bae15_005	bap10_005	bap05_005	baa00_005	baa95_005		
6	baa9515_0006	農業集落	1			bae15_006	bap10_006	bap05_006	baa00_006	baa95_006		
7	baa9515_0007	集計単位地域	26						baa00_007	baa95_007		
8	baa9515_0008	調査区	1			bae15_007	bap10_007	bap05_007	baa00_008	baa95_008		
9	baa9515_0009	経営体番号	5			bae15_008		bap05_008	baa00_009	baa95_010		
10	baa9515_0010	経営番号001	28				bap10_008					
11	baa9515_0011	経営番号002	28				bap10_009					
12	baa9515_0012	区分(調査種類)	30						baa00_010			
13	baa9515_0013	変換前集計単位地域	30						baa00_011			
14	baa9515_0014	変換前調査区	30						baa00_012			
15	baa9515_0015	全国農業地域コード	16					bap05_009	baa00_013	baa95_007		
16	baa9515_0016	地域類型別	30	1	都市的地域				baa00_015			
17				2	平地農業地域							
18				3	中間農業地域							
19				4	山間農業地域							
20	baa9515_0017	地域類型_新市_1次分類	24					bap05_010	baa00_016			
21	baa9515_0018	地域類型_新市_2次分類	24					bap05_011	baa00_017			
22	baa9515_0019	地域類型_旧市_1次分類	24					bap05_012	baa00_018			
23	baa9515_0020	地域類型_旧市_2次分類	24					bap05_013	baa00_019			
24	baa9515_0021	調査客体分類コード	26	1	専従農家				baa00_014	baa95_011	1	
25				2	自給的農家							
26												
27	baa9515_0022	調査客体の状況	27	1	調査区内にいない	bae15_01						
28				2	調査不能							
29				3	調査区内にいる							
30	baa9515_0023	家族による経営	7	1	家族による経営である	bae15_02	bap10_0101	bap05_0101				
31				2	家族による経営でない							
32	baa9515_0024	農業生産継続経営	27	1	相続の農業活動に従事している	bae15_03						
33				2	相続の農業活動に従事していない							
34	baa9515_0025	世帯中複数経営	27	1	該当する	bae15_04						
35				0	該当しない							
36	baa9515_0026	経営耕地面積30a以上	27	1	該当する	bae15_05						
37				0	該当しない							
38	baa9515_0027	経営規模	27	1	該当する	bae15_06						
39				0	該当しない							
40	baa9515_0028	農業サービス	27	1	該当する	bae15_07						
41				0	該当しない							

本資料の著作権は、京都大学農林水産統計デジタルアーカイブ講座に帰属しますので、無断での利用・転載・転用・複製は御遠慮願います。
 (C) 2014 Statistical Digital Archive of Agriculture, Forestry and Fisheries, Kyoto University. All Rights Reserved

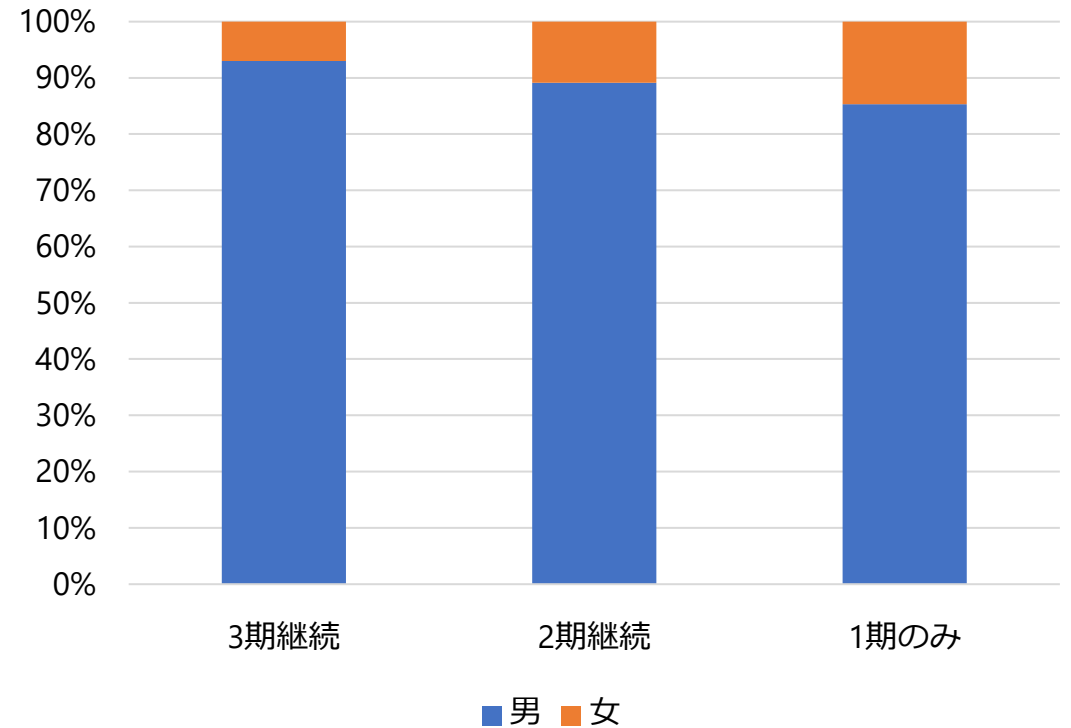
1/150

経営者年齢層×継続パターン：経営者性別 (2005年)

50歳未満

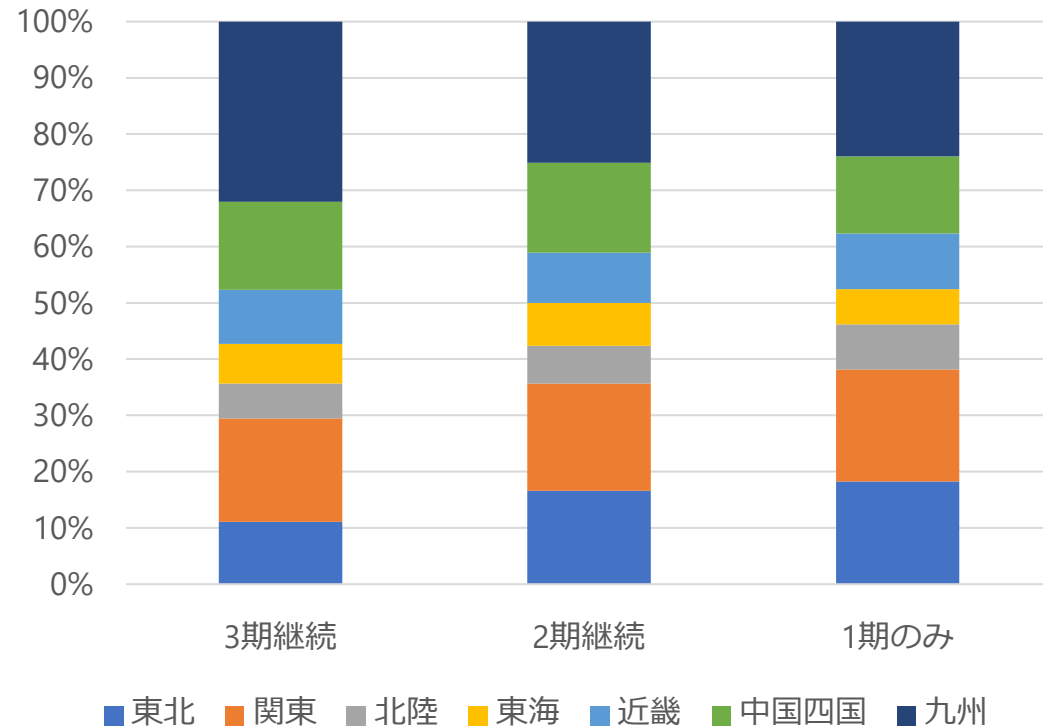


50歳以上

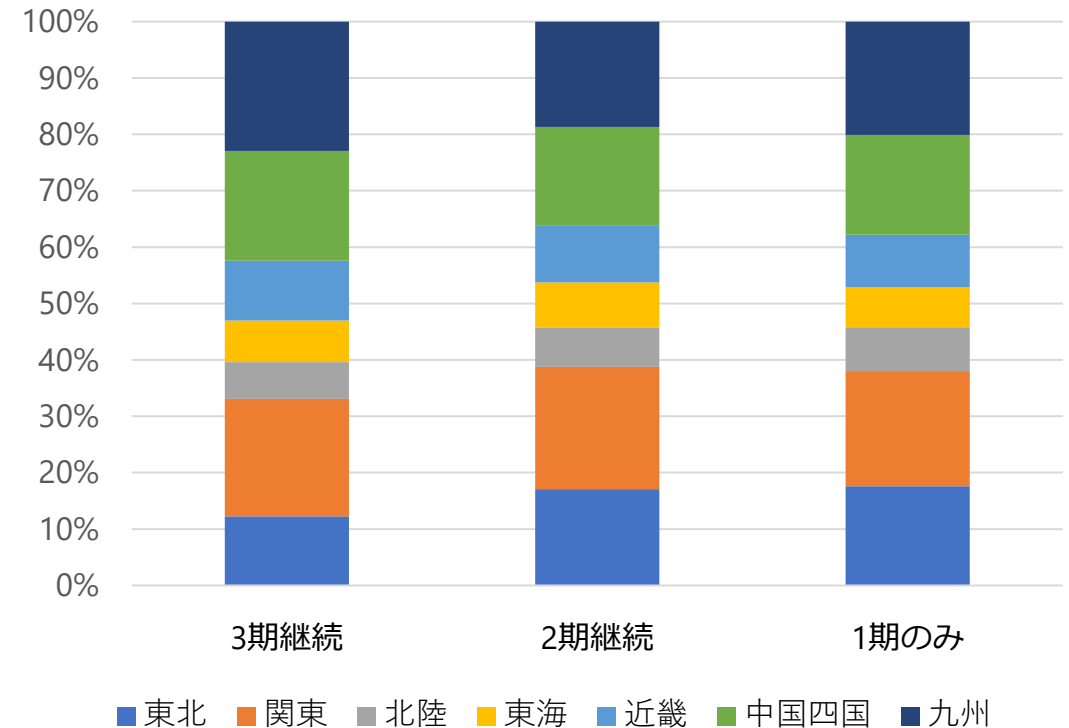


経営者年齢層×継続パターン：地域 (2005年)

50歳未満

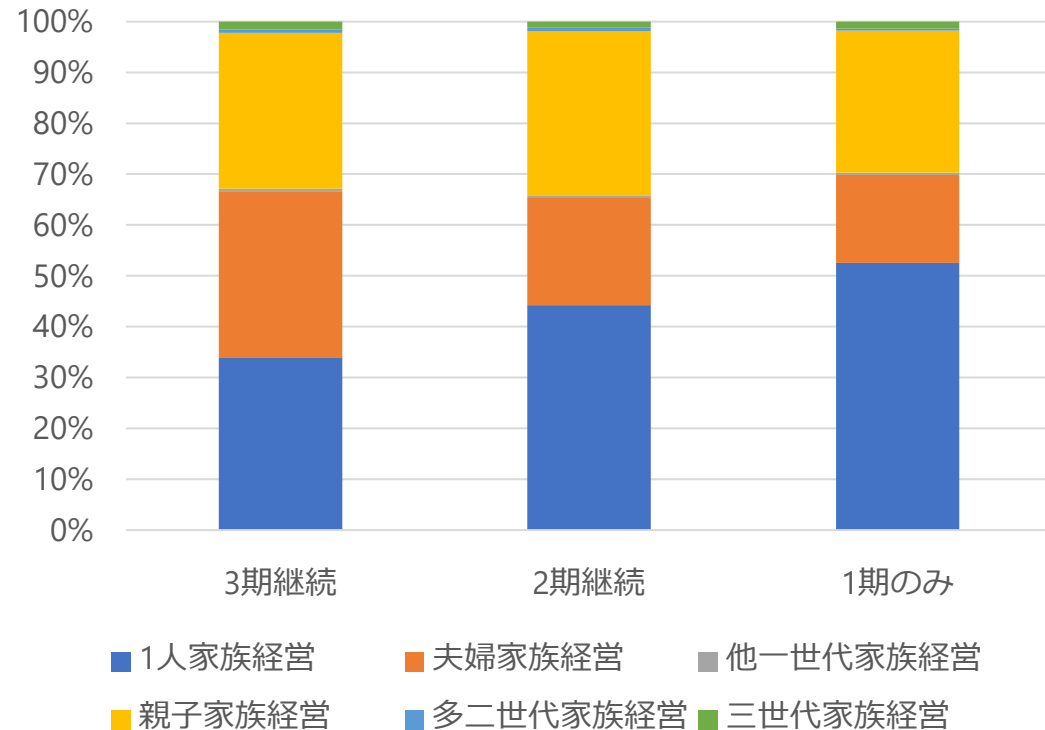


50歳以上

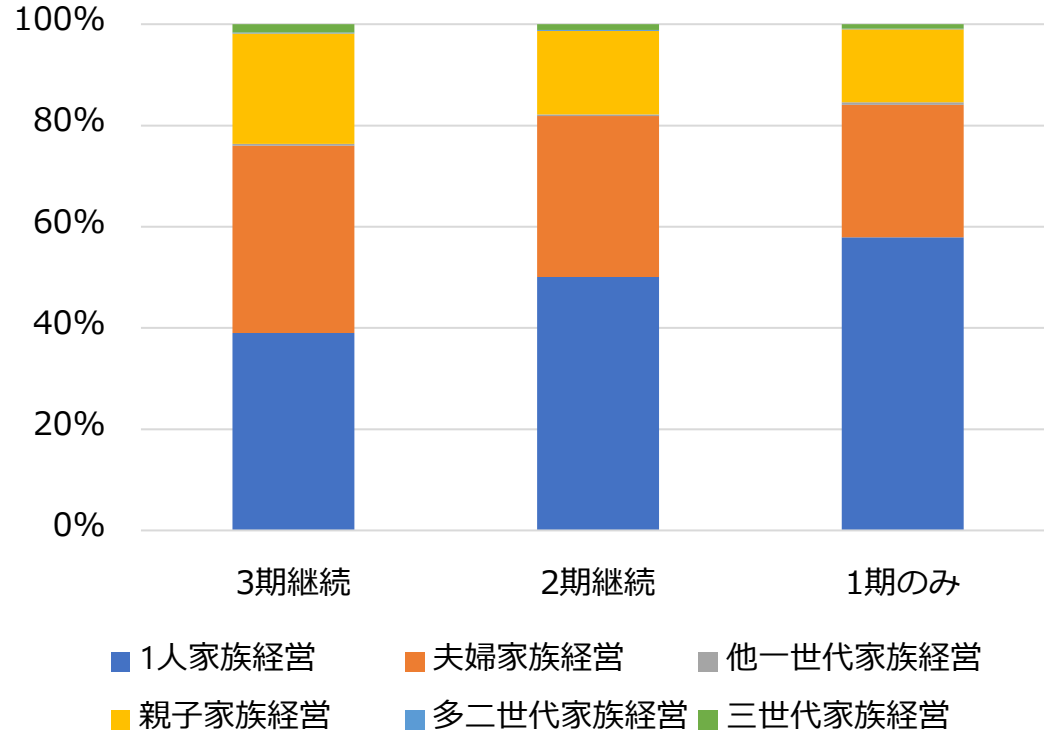


経営者年齢層×継続パターン：家族構造 (2005年)

50歳未満

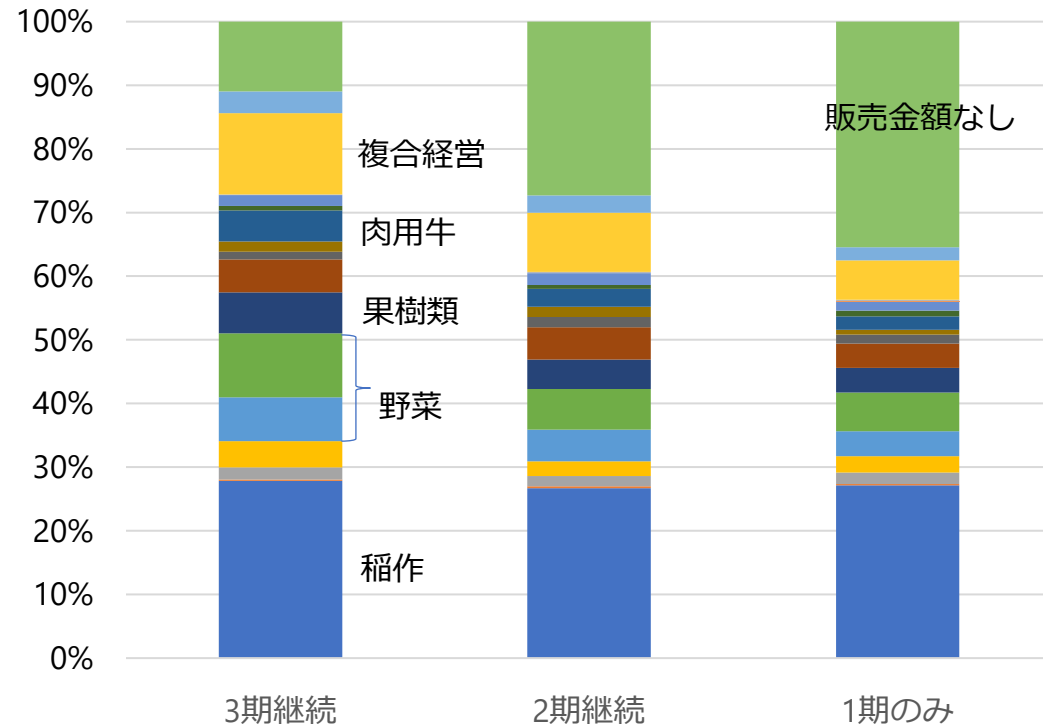


50歳以上

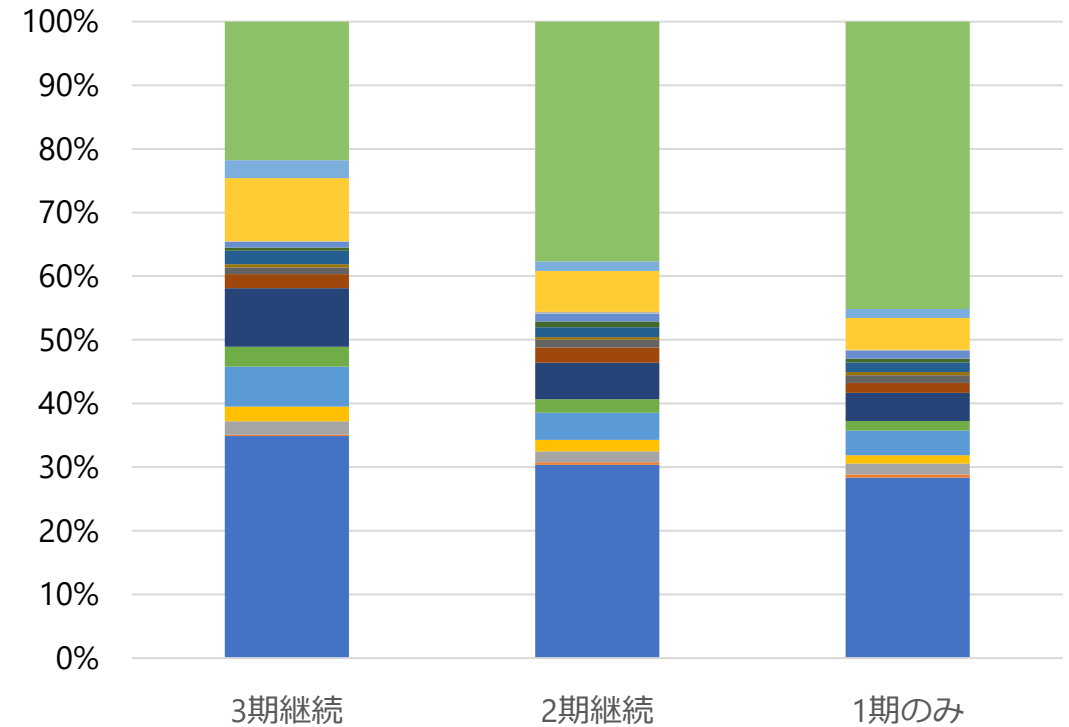


経営者年齢層×継続パターン：組織体と作物（2005年）

50歳未満



50歳以上



分析事例からみえた意義と課題

意義

- 大規模なサンプルサイズ
- 「少数」事例（⇔ウェブ調査・インタビュー）
- 秘匿性の問題
- 利用方法やオンサイトの増加

課題

- レディーメイドによる制約
- 社会学的関心とのずれ
 - 参入や退出の状況
 - 他の職業との比較
 - 半農+Xと外形基準
- 調査方法や項目の変更
- 実質的理解と「慣れ」が必要
 - 「手引き」+利用経験者+公的統計ミクロデータ研究コンソーシアムセミナー

4. 結論

◆本日の問い 公的データの利活用にはどのような可能性と意義があるのか？課題は何か？

①可能性と意義

- 現代社会における農林業の重要性
- 研究者（含むプロジェクト）や民間企業では収集が難しい対象・規模・内容のデータを長期間繰り返し収集
- 民間企業のデータに比べるとアクセス可能性は高く、利益相反や秘匿性が問題となる可能性は低い

②課題

- レディーメード→提案+複数の研究方法の組み合わせ
- データアーカイブに比べアクセスの問題は残るが、世界でも同様

3つの提言

①研究

- 適切な分析手法を選択する
- データ収集と利活用に関する倫理的問題を意識する
- 理論に基づき意味のある解釈をする
- 人間や社会の理解にどうつなげるか考える

3つの提言（続き）

②教育

- 『白書』やE-statの紹介
 - 大学に限らず卒業後も利活用できると説明
- いくつかの工夫
 - 日常的な概念と統計用語のずれを説明
(例：少子化 = 合計特殊出生率)
 - CSVファイルの使い方
 - 総務省ウェブサイト 教材や指導手引きあり
- 民主主義社会で必要な知識・教養



3つの提言（続き）

③政府統計に期待すること

- オンライン化（迅速な公開、諸コスト削減）
- 社会の短期かつ長期変化を追えるような調査設計
- 国際性（海外への提供・国際比較）（村上 2020）
- さまざまなデータの提供（含む行政登録データ）

公的データの公共的な利活用が可能にする未来

- 人口減少、農業に限らず担い手減少、格差の拡大などさまざまな社会的課題
→エビデンス（大規模データの分析）に基づく対策が必要
- 多くの人に参加できる民主主義社会の実現のために公的統計の意義を考え、公共財である公的統計を支えましょう

ご清聴
ありがとうございました