

社会生活基本調査から見た コロナ禍等の影響 ～その1：クラスター非類似度から見た影響～



木村 敦 | Kimura Atsushi

(独)統計センター 理事・CIO

■ NTT にて ICT 関連開発に長年携わり、(株)NTT ファシリティーズ総合研究所 取締役情報技術本部長を経て、2019 年 4 月から現職。1988 年 3 月名古屋大学大学院理学研究科博士課程（前期）修了、修士（理学）、専門統計調査士。

1. はじめに

独立行政法人統計センターでは、公的統計データを統計の学習や学校教育に活用いただけるよう「教育用標準データセット (SSDSE)」^[1]を提供している。公的統計の統計表データはそのままに、学習用に利用しやすく加工した6種類のデータセットをホームページ上で公開している。以前、本誌においては、総務省統計局の家計調査に基づいた SSDSE-C シリーズ（家計消費）を用いた日本の食嗜好の地域性分析事例^[2]をご紹介した。今回は、総務省統計局の社会生活基本調査に基づいた SSDSE-D シリーズ（社会生活）を用いた分析事例をご紹介したい。

2019 年末から世界的に広がったコロナ禍により世の中の生活スタイルは変化を余儀なくされた。緊急事態宣言が発出され、「接触機会」の低減を目的とし「テレワーク」や「時差通勤」が広く行われた。また、旅行や外食なども控えられた。容易に想像できることは、「通勤・通学」の時間が大きく減少し、これらの時間における都道府県間の差が縮小したことである。また、「買い物」や「旅行」、「外食」などの時間も大きく減少したはずであり、これらのばらつ

き縮小も予想される。

SSDSE-D データセットを用いて、コロナ禍前とコロナ禍中の生活スタイルにどのような違いが表れたのかをデータで見てみることにしよう。本稿では、20 種類の行動総平均時間を都道府県毎の 20 次元ベクトルとして、多変量解析を行うことにする。他の地域と比較して「仕事」や「通勤移動」の時間が多いと言われている大都市地域において大きな生活スタイル変化が生じたとの仮説を立てて、データ分析をしてみよう。

なお本稿における見解は筆者個人のものであり、所属する組織を代表するものではない。本稿の誤りは全て筆者の責に帰する。

2. 先行研究について

西郷^[3]では、社会生活基本調査結果から横軸に時間帯を、縦軸に行動者率をとって、行動種別・男女別における時間帯別行動者率の時系列的な変化を観察している。2021 年調査結果を他の年次と比較して、「仕事」の行動者率が低下していることを指摘している。逆に「休養・くつろぎ」については、全ての時間帯において上昇しているが、とりわけ午後7時以降に大きく上

昇しており、「仕事」や「通勤」の時間が少なくなった分が「休養・くつろぎ」に充てられたとみている。「ボランティア活動・社会参加活動」についても低下しており、特に午後の時間帯の低下率が目立つ。「交際・つきあい」も午後2時以降の低下率が大きいことを指摘している。

臼井他 [4] では、内閣府が行った「新型コロナウイルス感染症の影響下における生活意識・行動の変化に関する調査」を用いて、子育て世帯を中心に影響を分析している。その結果、コロナ禍においては、仕事満足度、社会とのつながり、生活の楽しさが大きく落ち込んだことを指摘している。一方で、家庭や生活に由来以上にウェイトを置いたワーク・ライフ・バランスの方向に変わる可能性について言及している。

これらの研究は、コロナ禍が、人々の生活行動に大きな影響を与えていることを明らかにしたものであるが、いずれの研究においても、地域別の影響には着目していない。本稿では、都道府県を単位とした地域別の生活スタイルの影響を明らかにしてみよう。

3. 分析に使用するデータと分析手法について

総務省統計局が実施している社会生活基本調査は、5年に一度行われる基幹統計調査である。直近の調査は2021年調査である。SSDSE-Dシリーズはこの社会生活基本調査に基づいた都道府県別の生活時間データを集めたデータセットであり、このデータセットを用いて分析を行う。

データのレイアウト														
縦 (行の数) 項目情報 (2) + 男女の別 (3) × 全国・都道府県 (48) = 146 行														
横 (列の数) 男女の別情報 (1) + 地域情報 (2) + データ項目 (121) = 124 列														
項目情報	項目コード	SSDSE-D-2023	データの年次情報 (対象年次2021年)											
	項目名	男女の別	地域コード	Prefecture	MA00	MB00	MB01	MB011					MH51	MH52
男女の別 (総数・男・女)	全国	0. 総数	R00000	全国	112462	39.6	14.1	12.8					8:20	18:34
	47都道府県	0. 総数	R01000	北海道	4614	35.0	11.1	9.9					8:19	18:07
		0. 総数	R02000	青森県	1090	25.4	7.4	6.6					8:09	18:01
		0. 総数	R46000	鹿児島県	1377	38.2	10.9	10.1					8:11	17:56
		0. 総数	R47000	沖縄県	1263	36.6	14.4	13.3					8:21	18:01
	男	1. 男	R00000	全国	54829	39.8	14.3	13.5					8:02	19:08
		1. 男	R01000	北海道	2179	34.6	11.1	10.4					8:00	18:39
		1. 男	R02000	青森県	514	25.9	8.5	8.0					7:56	18:31
		1. 男	R46000	鹿児島県	651	34.4	9.9	9.7					7:52	18:28
		1. 男	R47000	沖縄県	621	34.8	12.5	11.6					7:56	18:25
	女	2. 女	R00000	全国	57633	39.5	13.9	12.2					8:44	17:47
		2. 女	R01000	北海道	2434	35.4	11.1	9.5					8:44	17:23
		2. 女	R02000	青森県	579	25.0	6.4	5.4					8:26	17:23
		2. 女	R46000	鹿児島県	725	36.0	11.7	10.6					8:34	17:16
		2. 女	R47000	沖縄県	643	38.4	16.3	14.9					8:53	17:31
※ 前回の版 (SSDSE-D-2021) に比べ2項目増加しています。														

図1 SSDSE-D-2023のデータレイアウト

表1 行動種別20種類

生活時間編 1日に行動した平均時間	20の行動の種類別の平均時間 (総平均時間)	分	10歳以上	過全体
	MG01 睡眠			
	MG02 身の回りの用事			
	MG03 食事			
	MG04 通勤・通学			
	MG05 仕事			
	MG06 学業			
	MG07 家事			
	MG08 介護・看護			
	MG09 育児			
	MG10 買い物			
	MG11 移動 (通勤・通学を除く)			
	MG12 テレビ・ラジオ・新聞・雑誌			
	MG13 休養・くつろぎ			
	MG14 学習・自己啓発・訓練 (学業以外)			
	MG15 趣味・娯楽			
	MG16 スポーツ			
	MG17 ボランティア活動・社会参加活動			
	MG18 交際・付き合い			
	MG19 受診・療養			
	MG20 その他			

コロナ禍における「令和3年社会生活基本調査 (2021年調査)」に基づいたSSDSE-D-2023のデータレイアウトを図1に示す。

コロナ禍前の「平成28年社会生活基本調査 (2016年調査)」に基づくSSDSE-D-2021も同様に分析する。SSDSE-Dの年表示は調査年次ではなくSSDSE公表年次を示しているので注意されたい。

今回の分析では、SSDSE-Dの中の「調査票Aに基づく結果」・「生活時間に関する結果 (地域)」を用いる。47都道府県別の10歳以上の男女区分 (男女総数、男、女) 毎の1日の行動 (20種類) の種類別総平均時間である (表1)。

この20種類の数値データを全て用いて、都道府県を最小単位とした階層的クラスター分析を実施する。分析前処理として生活区分毎の時間の大小を揃えるために生活区分で20区分毎に標準化（平均0、標準偏差1）した後に、階層型クラスター分析（ward法）を用いて分析を行った。RスクリプトをエストレーラWeb^{注1}に掲載しておく。統計センターのホームページ[1]からダウンロードしたSSDSE-D-2023.csvファイルを読み込み、分析前処理を実施した後にクラスター分析を行いデンドログラム作成するま

での全処理を行うことができる。

4. クラスター分析結果とその特徴について

男女での生活スタイルが異なることが想定されることから、男女別のデータを用いてクラスター分析を行ってみよう。

まず、男性データについて見てみることにする。図2、図3がSSDSE-D-2023の分析結果である^{注1}。さらに図4、図5がSSDSE-D-2021を用いて同様に分析した結果である。以後、日本

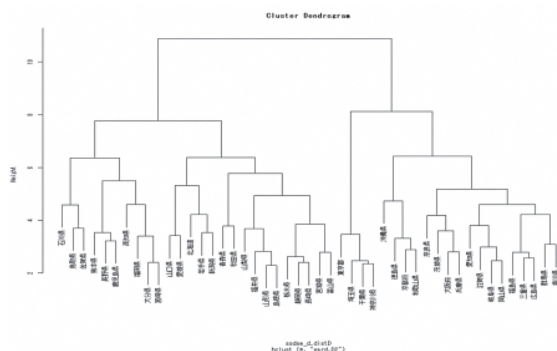


図2 2021年調査（男）による分析結果

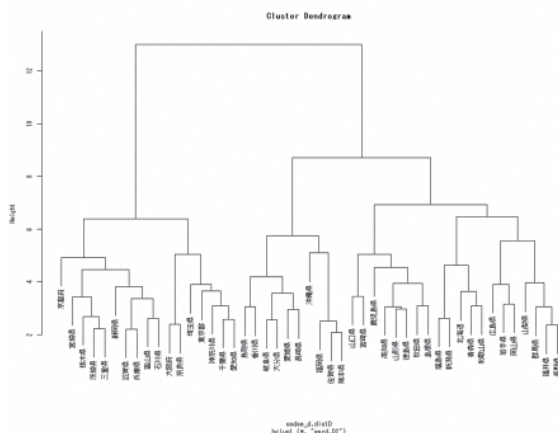


図4 2016年調査（男）による分析結果



図3 2021年調査（男）による区分地図



図5 2016年調査（男）による区分地図



区分地図は全てクラスター数2で表示している。

図3と図5を比較すると、いずれも東京都や大阪府を含むクラスターとそれ以外に分かれることが分かる。本稿では便宜上、クラスター数2の場合における東京を含むクラスターを「東京クラスター」、それ以外の都道府県で構成されるクラスターを「その他クラスター」と呼ぶことにする。

コロナ禍前の2016年では、宮城、栃木、静岡、富山、石川が「東京クラスター」に含まれている。一方、コロナ禍の2021年調査結果分析にお

いては、前述の各県に代わって、福島、群馬、岐阜、和歌山、岡山、広島、香川、徳島が「東京クラスター」に含まれた。2021年の結果では、「東京クラスター」に含まれる都道府県の数が若干増加し、西側の都道府県が少し増えているように見える。

さらに、2021年の分析結果（図2）と2016年の分析結果（図4）のデンドログラムを比べてみよう。「東京クラスター」と「その他クラスター」の間の非類似度（デンドログラムの縦軸）が2016年の分析結果では約13であり、

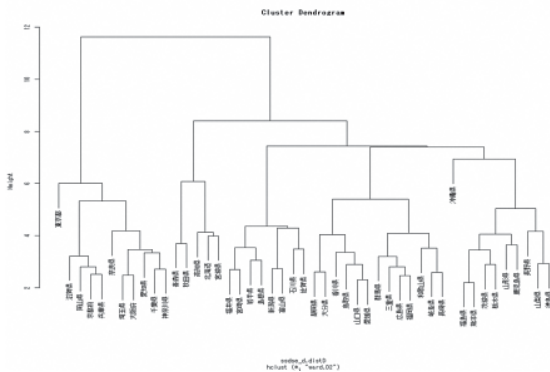


図6 2021年調査（女）による分析結果

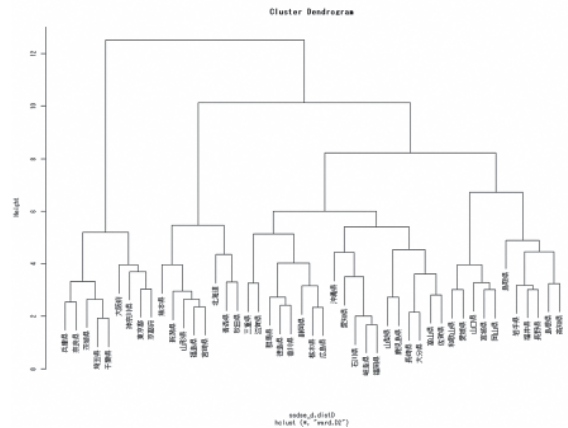


図8 2016年調査（女）による分析結果



図7 2021年調査（女）による区分地図



図9 2016年調査（女）による区分地図

2021 年の分析結果の約 11 よりも大きい。コロナ禍という状況下では、どの都道府県においても同様な行動制約が課されたため、平常時と比較してどの都道府県でも似かよった生活スタイルになったと解釈できるだろう。なお、このデンドログラムは R 言語の hclust 関数に method パラメータ “ward.D2” として作成している。このためデンドログラムの縦軸の数値は、全て「Ward が定義した非類似度」の平方根であることに注意されたい [5, 6]。

次に、女性データについて見てみることにする。図 6、図 7 が 2021 年調査の分析結果、図 8、

図 9 は 2016 年調査の分析結果である。図 6 と図 8 を比較してみると、「東京クラスター」と「その他クラスター」の間の非類似度についての傾向は男性の分析結果と同様に 2021 年調査結果分析の非類似度が小さくなっている。

一方、図 7 と図 9 の日本区分地図を比較してみると、女性の場合は男性（図 3、図 5）ほど変化が大きくないことが分かる。

この非類似度についての傾向がコロナ禍独自のものなのかどうかを、さらに他の調査年次データも用いて確認してみることにしよう。

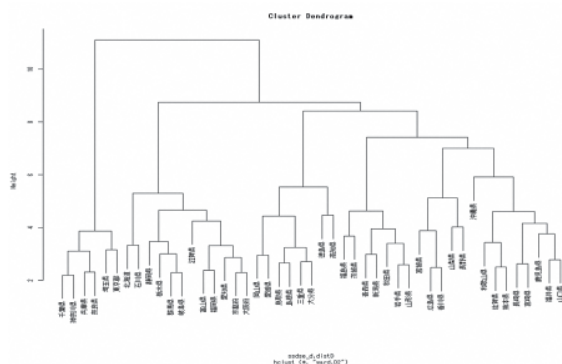


図 10 2011 年調査（男）による分析結果

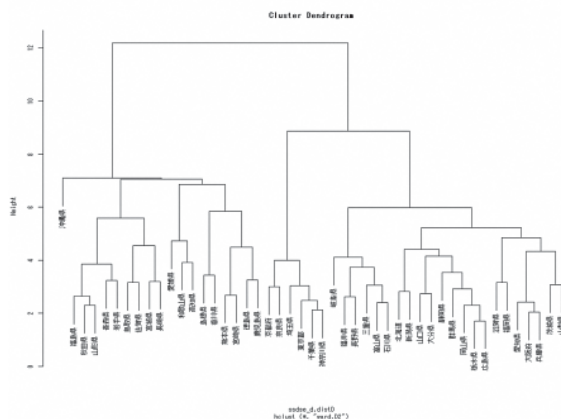


図 12 2001 年調査（男）による分析結果

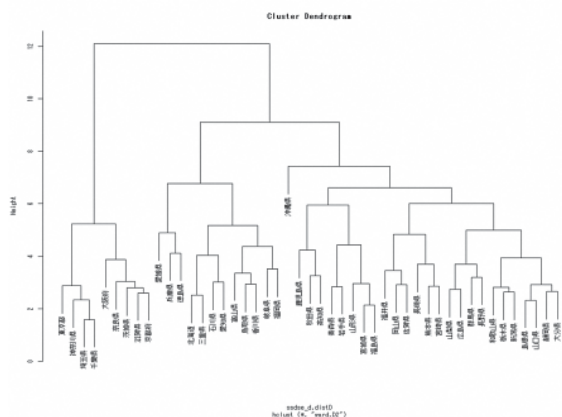


図 11 2006 年調査（男）による分析結果

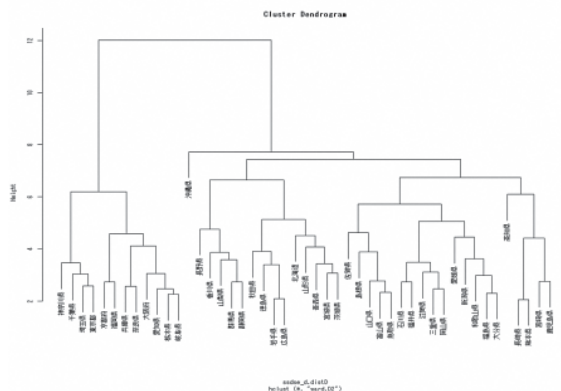


図 13 1996 年調査（男）による分析結果

5. 調査年次をさらに広げて分析してみる

「政府統計の総合窓口（e-Stat）」から同じ条件のデータとして2011年、2006年、2001年、1996年のものが利用できる。

まず、男性データから見てみよう（図10～図13）。「東京クラスター」と「その他クラスター」間の非類似度は、2011年を除けば全て約12程度であり、2016年の結果とほぼ同じである。2021年のものが約11と小さいことが分かる。コロナ禍においては都道府県間の非類似度が小さくなるという仮説は、分析結果と矛盾しない。2011年については非類似度が約11程度であり、

2021年の分析結果と同程度となっている。つまり、コロナ禍以外でも非類似度が小さい時期がある。2011年においては別の要因が影響している可能性がある。これについては別途考察する。

次に、女性データについて見てみると（図14～図17）、全ての非類似度は12以上である。コロナ禍の2021年の分析結果が他の年次と比較して非類似度が小さいことは間違いないようだ。

しかし、2001年と1996年においては非類似度が約14程度あり、他の年次の分析結果と比べて大きな値となっている。この点については、コロナ禍の影響とはまた別の影響がありそうである。

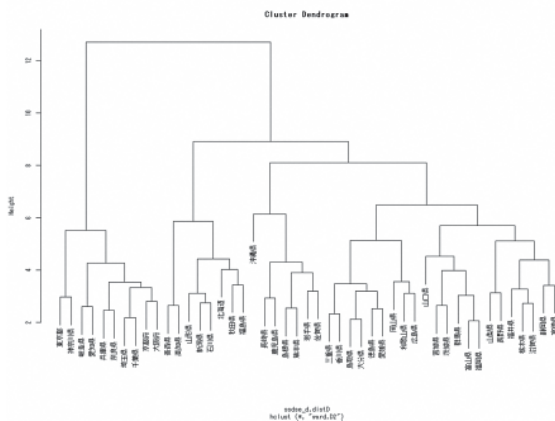


図14 2011年調査（女）による分析結果

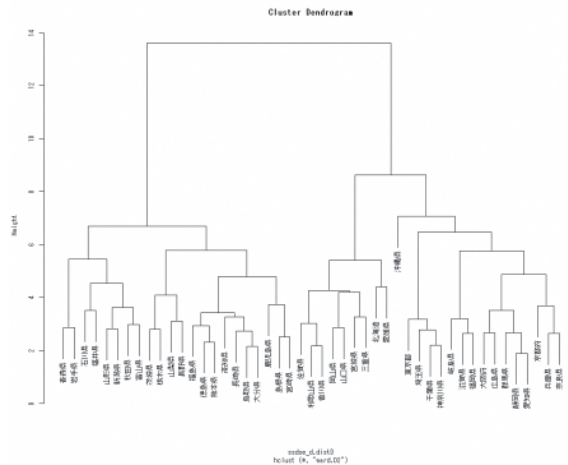


図16 2001年調査（女）による分析結果

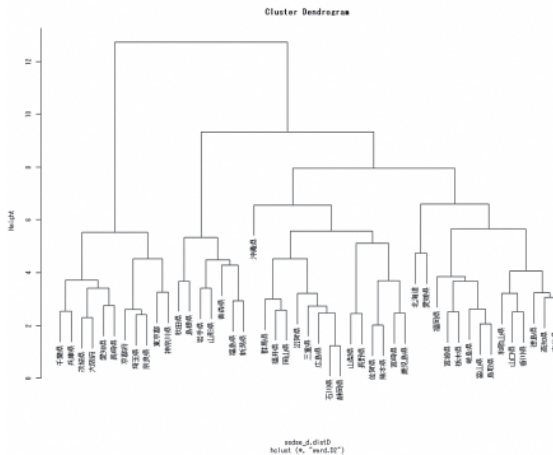


図15 2006年調査（女）による分析結果

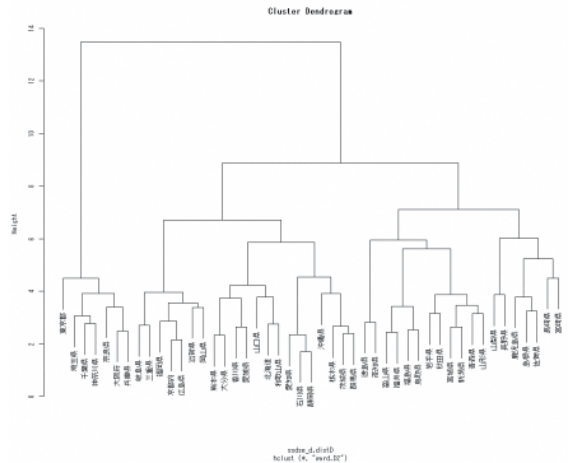


図17 1996年調査（女）による分析結果

6. 2011 年における非類似度縮小 (男性) について

男性データの分析結果では、2011 年についてもコロナ禍の 2021 年の分析結果と同様に非類似度が小さくなっている。2011 年 3 月には東日本大震災が発生している。内閣府の平成 23 年度年次経済報告書によれば、「東日本大震災の経済的影響が、電力供給の制約やサプライチェーンの寸断によって、被災地以外にも広く及んでいる」と記載されている。企業側の影響としても計画停電や電力使用制限など、生活スタイルにも大きな影響を与えた。コロナ禍とは異なり、男性のデータだけに非類似度の縮小が表れたのは、主に企業活動の全国的な変化に起因する生活スタイルへの影響が大きかったことを示しているのかもしれない。

7. おわりに

社会生活基本調査のクラスター分析から、日本全国を 2 つのクラスターに分けた場合の結果について今回は非類似度を中心に特徴を見てみた。

コロナ禍の影響が「東京クラスター」と「その他クラスター」の非類似度の小ささとして表れていることが分かった。コロナ禍による各種の行動制限によって、全国における生活スタイルの都道府県別の違いが近づいたということはある程度理解しやすい結果であろう。

男性データにおいてのみ特徴的だったのは、東日本大震災の直後の 2011 年調査においても 2021 年と同様に生活スタイルの違いが少なくなっていることである。女性データのみの特徴としては、2001 年と 1996 年の調査データの分析結果で非類似度が大きくなっていることである。次稿では、クラスターサイズの観点から分析してみたい。

*参考文献

- [1] (財)統計センター「SSDSE(教育用標準データセット)」
<https://www.nstac.go.jp/SSDSE/index.html>
- [2] 木村敦・高部勲 (2021) 「家計消費データから見る日本の食料嗜好地域性 ～人文社会系知見との連携も見据えて～」『ESTRELA』No.324 (2021 年 3 月号), pp.36-42, 統計情報研究開発センター.
- [3] 西郷浩 (2023) 『公的統計による統計入門』日本統計協会.
- [4] 白井恵美子・佐藤蘭香・松下美帆 (2022) 「新型コロナウイルス感染症の影響下におけるワーク・ライフ・バランス」『経済研究』Vol. 73, No. 4, pp. 358-391, 一橋経済研究所.
- [5] 木村敦 (2022) 「R の hclust 関数による正しい ward 法分析の方法 ～その 1: ward 法の特徴と検証実験～」『ESTRELA』No.337 (2022 年 4 月号), pp.22-27, 統計情報研究開発センター.
- [6] Joe H. Ward, Jr.(1963) "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, Vol. 58, No. 301(Mar., 1963), pp.236-244.