

税務データ分析報告書*

東京大学政策評価研究教育センター (CREPE)

2024年5月15日

目次

1	はじめに	2
2	本報告書を適切にご活用いただくために	2
3	税収予測について	4
3.1	はじめに	4
3.2	予測の考え方	5
3.3	データ	7
3.4	予測値の各要素と誤差率について	8
3.5	予測モデルにおける解釈の限界	9
4	記述統計	9
4.1	基本統計	11
4.2	基本統計 (税収)	14
4.3	市町村民税分析 (性別・年齢階級別)	16
4.4	市町村民税分析 (所得階層別)	18
4.5	転入・転出の分析	21
4.6	所得項目の分析 (区分別)	25
4.7	所得分布・格差動向に関する分析	26

* 報告書の公開範囲や活用方法について制限は設けておりません。ただし、報告書の一部を抜粋もしくは引用する場合は、出典として「東京大学政策評価研究教育センター 税務データ分析報告書 (2023年度)」を明記してください。

4.8	男性の給与収入と既婚率の関係	30
4.9	出産前後の女性の給与収入・就業率の変化	31
4.10	高齢者の就業率	32
5	補遺	33
5.1	変数の作成過程について	33
5.2	税収予測のモデルについて	34

要旨

- 報告書の目的
 - 本報告書は、予算編成・折衝における参考資料を提供するとともに予算編成のみならず各政策領域の企画・立案に活用いただくことを目的としています。予算編成・折衝における参考資料として2024年度の市町村民税の賦課総額の予測を行い、企画・立案に活用いただくために足元の動向を把握することのできる記述統計を作成しました。

- 税収予測
 - 2023年度以前のデータを予測アルゴリズムに学習させることで2024年度の市町村民税の賦課総額の予測額を算出しました。
 - 予測アルゴリズムにおいては、マクロ経済動向、年齢階級や性別などの個人の属性、過年度の賦課実績などの個人の税申告に関わる特徴を踏まえて、各個人の税収額を予測しました。加えて人口動態も考慮して各個人の予測額を合算することでCREPE市の市町村民税の賦課総額を計算しました。
 - 予測結果によると、2024年度におけるCREPE市の市町村民税の賦課総額の予測値は**21.07億円**です。
 - 2023年度以前のデータでは予測誤差率の平均値は1.7%です。しかし、本報告書が行なった税収予測は、あくまでも過去のデータの特徴を将来に当てはめた結果に過ぎず、特に感染症の蔓延や自然災害、金融危機など、予測不可能な事態が発生した場合は予測精度が著しく低下する可能性がある点については留意いただく必要があります。

- 記述統計
 - CREPE市の人口構成（性・年齢階層）や税収、所得階層別割合などをまとめています。
 - CREPE市の所得階層別の人口割合をみると、所得なしの割合が過去5年で43%から39%へ横ばいで推移しています。
 - CREPE市の転入（出生を含む）者数は過去5年で1346人から1258人へ横ばいで推移している一方で、転入者の賦課総額はやや増加しています。
 - CREPE市の転出（死亡を含む）者数は過去5年で2196人から1969人へやや減少している一方で、転出者の賦課総額は横ばいで推移しています。

1 はじめに

この度は東京大学政策評価研究教育センターの「EBPM 推進のための自治体税務データ活用プロジェクト」にご参加いただきありがとうございました。ご提供いただいたデータを用いて税収予測や足元の経済状況に関する情報をまとめましたので、貴自治体における政策の企画立案の参考資料としていただければ幸いです。

エビデンスに基づく政策決定 (EBPM) の重要性が指摘されるようになって久しいですが、本プロジェクトはそのための基礎的な資料の提供を目指すものです。今回ご提供いただいた税務データは、本報告書作成のほか、本プロジェクトに協力する研究者チームによるより詳細かつ学術的に厳密な政策評価研究にも活用させていただく予定です。本プロジェクトは、EBPM の観点から、自治体の企画立案する政策の質を高めるための知見を提供することを最終的な目標としております。

次節で詳述いたしますように、エビデンスと呼ばれうる統計資料は大きく二つに分けることができます。本報告書ではそのうち、貴自治体などの政策当局が足元の経済状態を的確に把握するための統計資料の提供を目的としております。今回ご提供いただいたデータがこれまでの統計資料よりも高い解像度で貴自治体の経済状況を把握していただくための足掛かりとなれば幸甚です。

本報告書では、特定の政策を実施したことが、どのような効果を上げたかを検証する、いわゆるプログラム評価までは行っておりません。あくまでもその手前の現状把握のための基礎資料の提供を目指しております。もっとも、今回ご提供いただいた税務データは、本プロジェクトに協力する研究者チームによるプログラム評価を行うための基礎資料としても活用させていただき、研究成果は随時[ホームページ](#)等で公表・発信していく予定です。なお、本プロジェクトは科学研究費補助金(課題番号 22H05007 および 22H05011)の補助を受けています。

東京大学もまだ手探りでプロジェクトを進めている状態で、様々な点で行き届かない部分があったかと思えます。プロジェクトの趣旨にご理解をいただき、ご協力いただきましたこと、あらためて感謝申し上げます。来年度以降もプロジェクトへの参加を積極的にご検討いただけますと大変幸いです。

2 本報告書を適切にご活用いただくために

エビデンスに基づく政策決定 (EBPM) におけるエビデンスとは科学的な根拠のことですが、「科学的な根拠」という言葉は大きく二つの意味で使われます。

一つは貴自治体などの政策当局が現状を的確に把握するための客観的資料という意味です。すでに様々な統計資料や業務資料を利用して貴自治体の足元の経済状況の把握につとめられていることと拝察します。今回ご提供いただいた匿名化された税務情報は、納税者の個人情報や秘匿しつつも個人レベルの収入などを正確に記録したデータであり、適切な分析を施すことによって、平均的な収入がどのように変化しているのみならず、収入の格差がどのように変化しているかを把握することができます。また、仮に平均収入が変化しないときでも、すべての住民の収入が変化していない場合もあるでしょうし、収入が上がる人がいる一方で収入が下がる人がいるという激しい変化が起こっているかもしれません。前者の場合と後者の場合では、現状認識が大きく異なりますので、求められる政策的な対応が異なるかもしれません。

もう一つは、特定の政策を実施したことが、どのような効果を上げたかを、統計データに基づいて検証したものです。このように政策効果を分析することをプログラム評価と呼びます。EBPMに関する学術研究の文脈では科学的根拠といったときにはこのプログラム評価の結果を指すことが多いのですが、本報告書では必ずしも特定の政策を評価しているというわけではありません。

本報告書はあくまでも、エビデンスの第一の意味である「現状を的確に把握するための客観的資料」を提供することを目指しております。税収予測に加えて、世帯や個人の収入に関するいくつかの記述的分析を掲載しておりますので、適切な政策立案を行うための基礎情報としてご活用いただければ幸甚です。

この際にご注意いただきたいのは、本報告書に掲載した記述的分析が示すのはあくまでも相関関係にすぎず、因果関係を検証したものではない点です。たとえば、新しく子供が生まれた世帯の平均年収が、世帯主が同年代の他の世帯の平均よりも高かったとします。これだけをもって、「収入が上がると子供が増える」と結論づけるのは早計です。たとえば、新しく子供が生まれた世帯は、世帯主が同年代の他の世帯よりも、これまで子供がいなかった割合が高く、そのため妻の労働時間が長く収入が高かっただけの可能性もあります。また、実際のところ出生率に影響するのは各世帯の収入ではなく、保育園の空き状況など子育てに適した環境が整っているか否かであり、環境の良い地域は地価も高く高収入世帯が集まるため結果的に世帯収入が出生率に影響しているように見える、といった可能性もあります。

こうした、様々な可能性を排除していくためには、エビデンスの第二の意味である厳密なプログラム評価^{*1}が必要になります。しかし厳密なプログラム評価のためには数多くの前提条件や追加情報が必要となり、短期間に行うことは困難です。一方、相関関係を示した記述的分析や、観測できる要因のみに基づく税収予測であっても、政策立案をするうえでの前提条件の把握には十分に有用であると認識しております。このような点に留意しつつご活用いただけますようお願い申し上げます。

本報告書の範囲からは外れますが、今回ご提供いただいたデータは、他の情報を補完することによって、特定の政策のプログラム評価を行うための基礎資料として活用できる可能性があります。例えば、ある地域で保育定員を拡大した結果として、子供のいる女性の就業率が高まるかどうかを知りたいとしましょう。女性の就業については今回ご提供いただいたデータの中で収入がある人かどうかで判断することができます。ここに、地域の保育所の定員の増減や子供の人口、当該世帯の保育所申し込みの有無や入所の可否といった情報を接続することができれば、より統計的に厳密な方法で、保育所の利用可能性が女性の就業率に与える影響が検証することができます。

^{*1} 特定の政策を実施したことが、どのような効果を上げたかを、統計データに基づいて検証することをプログラム評価といいます。一般的に、ある政策の対象者に特定の傾向がみられること（相関関係）は、かならずしもその政策の効果（因果関係）を意味しません。しかし、様々な統計的手法を駆使して、一定の仮定の下で、観察された相関関係を因果関係として解釈しうるかを判別し、効果の大きさを計測できることがあります。これがプログラム評価です。

3 税収予測について

3.1 はじめに

2024 年度における CREPE 市の市町村民税の賦課総額の予測値は 21.07 億円です*2。個人住民税のうち道府県民税額分を除いた市町村民税額分を予測対象にしています。2023 年度以前のデータで本モデルの予測結果を確認すると、予測誤差率の平均値は 1.7% です。

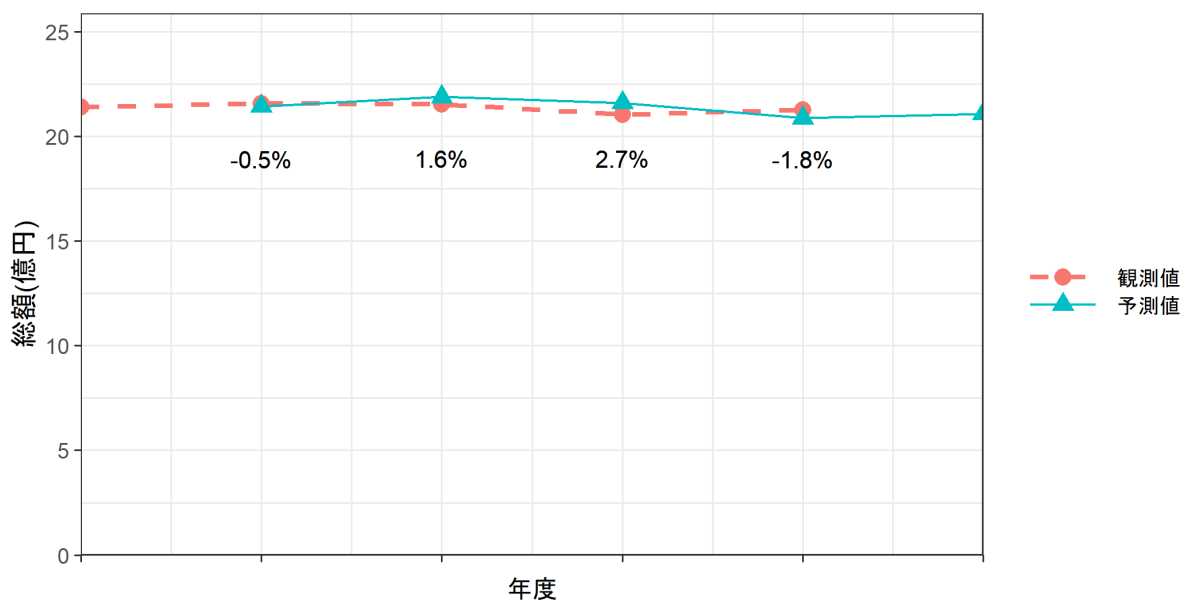


図1 税収予測値の推移

図1は CREPE 市の市町村民税の賦課総額の予測結果を時系列で示しています。横軸は賦課年度を表しており、縦軸のうち赤線は過去の税務データから算出した観測値を、青線は当方が算出した予測値を表しています*3。グラフ内の数値は各年度の予測誤差率です。

表1では CREPE 市における税収予測結果の概要を示しています。観測値は過去の税務データから算出した市町村民税の賦課総額の実績値です。予測値は当該年度以外の賦課額の実績値を基に当方の予測モデルによって算出した市町村民税の賦課総額の予測値を示しています。NA はデータが欠損していることを示しています。2024 年度における実績値は税務データが存在しない予測対象で

*2 賦課額というのは、個人が納付すべき税額として行政機関が課税する金額のことを示しています。当方にデータ提供いただいた後に修正申告があった場合などは、データ内での賦課額と実際の課税金額との間に差が生じている可能性があります。また、賦課額と収納額は異なるため、実際の市町村民税歳入額とは異なります。

*3 ここでの予測結果・予測誤差率という言葉は、今回頂いたデータのサンプル内での精度や誤差であり、2024 年度の予測の正確性を保証するものではありません。

表1 税収予測結果の概要

年度	観測値 (億円)	予測値 (億円)	予測誤差率	有効求人倍率
2019	21.387	NA	NA	1.571
2020	21.545	21.430	-0.5%	1.762
2021	21.531	21.878	+1.6%	1.783
2022	21.024	21.593	+2.7%	1.308
2023	21.251	20.860	-1.8%	1.133
2024	NA	21.072	NA	NA

* 予測誤差率 (平均):1.7%

あることから、そして2019年度における予測値は、ある年度のデータから次年度の予測を行う性質から、欠損値になっています。予測誤差率は観測値と予測値の誤差を示しています。有効求人倍率はそれぞれの公的統計データベースから取得したCREPE県単位でのマクロ変数であり、景気変動を捉える目的で導入しています*4。

3.2 予測の考え方

本報告書における税収予測は、各個人からの税収額の増減や転入（出生も含む）・転出（死亡も含む）のパターンを踏まえて、2023年以前のデータを用いて構築した予測モデルによるものです。CREPE市の $t+1$ 年度の税収予測には、 t 年までのデータを用いて、個人の転出・転入確率を考慮に入れた予測モデルを考えます。ここで、 $t+1$ 年度とは予測対象となる年度を指し、 t 年度とは予測対象年度を基準とした時の前年度を指します。例えば2024年度の税収予測を行う場合は、 $t+1$ 年度は2024年度を指し、 t 年度は2023年度のことを指します。

CREPE市の税収予測には3つの要素があります。一つめは、 t 年初時点でCREPE市に居住する各個人の $t+1$ 年度の賦課額です。 t 年度までのデータを用いて、平均的に個々の市町村民税の賦課総額がどのように変化するかを推定します。次に、以上のモデルから得られたパラメータの推定値を用いて、それぞれの個人の $t+1$ 年度の賦課額を予測します。

二つめの要素は、それぞれの個人が $t+1$ 年も引き続きCREPE市に居住し続ける確率です。この確率の推定は、 t 年度までのデータを用いて、過去の転出・死亡の確率を推定することから始まります。推定モデルから得られたパラメータを用いて、 $t+1$ 年までに転出・死亡する確率の予測値を算出します。その確率の予測値を p とすると、居住し続ける確率の予測値は $1-p$ になります。

最後に、CREPE市への転入者の賦課額も考える必要があります。この三つめの要素に関しても、 t 年度までのデータを用いて、どのような属性の個人がCREPE市に転入（出生）するかを予測します。そして、入ってくる個人が性別・年齢階級別に平均的にどの程度の市町村民税を賦課されるかを推定します。性別・年齢階級別の過去の平均値を用いて、 $t+1$ 年から新たに住民となる転入者の賦課総額を予測します。

この三つの要素を合わせることで、CREPE市の市町村民税の賦課総額の予測値を算出します。数式で記述すると、

*4

予測賦課総額 = t 年より継続して居住する全個人の $(t+1)$ 年度の予測賦課総額 + 全ての転入者の予測賦課総額
(1)

t 年より継続して居住する全個人の $(t+1)$ 年度予測賦課総額 = $\sum_{i=1}^N [i \text{の居住継続確率} \times i \text{の居住継続を仮定した予測賦課額}]$
(2)

となります。

3.3 データ

本節では、本報告書で使用した変数について簡単にご説明します。税収予測ならびに記述統計で使用した変数は以下の通りです*5。以下の項目に限らず、収入項目や控除項目、その他住民区分などについても幅広く活用させていただく予定です。

■ 税務データ

- 賦課年度
- 宛名番号（ハッシュ化済み）
- 属性に関する項目*6
 - 性別
 - 生年月
- 市町村民税*7
 - 市町村民税 合計額
 - 市町村民税 減免後均等割金額
 - 市町村民税 減免後所得割金額
- 控除前の所得金額*8
 - 営業等所得
 - 農業所得
 - 不動産所得
 - 利子所得
 - 配当所得
 - 給与所得
 - 雑所得（年金）
 - 雑所得（その他）
- 所得合計
 - 総所得金額

*5 なお、分析の目的に合わせて変数を加工した後、分析を行っています。変数の作成過程については補遺をご参照ください。また、各変数につきましては当方が提供した匿名化ツールを用いて匿名化処理を行っています。処理内容は自治体によって異なりますが、簡易な匿名化を選択された自治体は①氏名・個人番号・住所等の事前削除、①宛名番号・世帯番号のハッシュ化、②生年月日の丸め込みを、高度な匿名化を選択された自治体は加えて③特異な世帯の世帯番号の秘匿、④所得・賦課金額等のトップコーディング、⑤3-匿名性、⑥ランダムサンプリングを、中間的な匿名化を選択された自治体は①から②に加えて、③から⑥の各種匿名化処理を行っています。詳しくは公募時に Web ページで公開しております「(参考資料 2) 匿名化についての考え方 (公表版) .pdf」をご参照ください。

*6 匿名化の段階で「続柄」と「世帯番号」を使用しています。

*7 自治体によっては減免額が別途存在する場合がございます。

*8 自治体によっては存在しない所得項目もございます。

また、税収予測のモデルでは税務データに加えてマクロ経済動向を速報的に捉えることのできる公的統計として職業安定業務統計も使用しました。ただし、新型コロナウイルス感染症の蔓延等で社会経済状況が大きく変化し、予測に役立つような一貫性を持たなくなった可能性を踏まえて、マクロ変数を説明変数に加えたモデルと加えないモデルの両方を推計し、精度の高い方を採用しています。^{*9}。

- 有効求職者数（季節調整値）
- 有効求人倍率（季節調整値）

3.4 予測値の各要素と誤差率について

本節では、図1ならびに表1で示した税収予測結果を、各要素に分けて詳述します。3.2. 予測の考え方で説明した通り、本報告書においては前年度居住者の賦課額・転出確率・転入者の賦課額という3つの要素をモデル化しているため、いずれの変動が2024年度の市町村民税の賦課総額の予測値に影響を与えているのか、および各要素の誤差率について報告します。

表2は税収予測結果の内訳を示しており、表内の数値は**全て予測値**です。まず、CREPE市の年度ごとの賦課総額の予測値を2列目に記載しており、この数値は表1「税収予測結果の概要」における予測値の列に対応します。この賦課総額の予測値のうち居住継続者の賦課総額（予測値）と転入者の賦課総額（予測値）の内訳を3列目・4列目に示しています。それぞれの賦課総額（予測値）を算出する際に使用した転入者の総数（予測値）と転出者の総数（予測値）ならびに総人口（予測値）を5・6・7列目に記しています^{*10}。

表2 【予測値】賦課総額&居住継続・転入・転出

年度	賦課総額（億円）	居住継続者の賦課総額（億円）	転入者の賦課総額（億円）	転入者の総数	転出者の総数	総人口
2019	NA	NA	NA	NA	NA	NA
2020	21.430	20.908	0.522	1303	2099	54288
2021	21.878	21.388	0.490	1242	2050	53526
2022	21.593	21.075	0.518	1279	2052	52731
2023	20.860	20.320	0.540	1316	2048	51986
2024	21.072	20.553	0.518	1289	2021	51275

この予測結果については表3の予測誤差率が示す通り、上振れ・下振れする可能性がございます。転入者の総数ならびに転出者の総数、転入者の賦課総額、居住継続者の賦課総額それぞれの誤差率を合算して得られる賦課総額の予測誤差率の平均値は1.7%となっています。

^{*9} ■ 厚生労働省「一般職業紹介状況（職業安定業務統計）」〔macrodata〕

^{*10} 表中のNAはデータが欠損していることを示しており、予測アルゴリズムの適用対象外の年度においてNAとなっています。また、 t 年度における総人口に $t+1$ 年度の転入者の総数を加え、 $t+1$ 年度の転出者の総数を差し引いた数値が $t+1$ 年度における総人口に合致していませんが、年度ごとにそれぞれ独立に予測値を推定しているためです。

表3 【予測誤差率】賦課総額&居住継続・転入・転出

年度	賦課総額（億円）	居住継続者の賦課総額（億円）	転入者の賦課総額（億円）	転入者の総数	転出者の総数	総人口
2019	NA	NA	NA	NA	NA	NA
2020	-0.5%	-0.6%	+1.9%	-3.2%	+0.1%	-0.1%
2021	+1.6%	+1.9%	-7.7%	-5.7%	-4.5%	+0.0%
2022	+2.7%	+2.6%	+6.0%	+2.1%	+0.6%	+0.0%
2023	-1.8%	-1.9%	+1.0%	+4.6%	+4.0%	-0.0%
2024	NA	NA	NA	NA	NA	NA

3.5 予測モデルにおける解釈の限界

本報告書で使用している予測モデルは、マクロ経済動向やCREPE市の政策が住民の行動に与える影響、ならびに個人の嗜好をモデル化したものではありません。そのため、なぜその予測値なのかについて経済学的知見に基づく解釈を加えることはできません。解釈が可能であるのは、前節で説明したように、各予測項目の予測が合計値に与えた影響を特定することまでであり、各予測項目の変動要因を特定することやその要因の影響の大きさを推定することはできません。例えば、居住継続者の賦課金額の予測値が増加した場合、それはマクロ経済動向が改善し給与所得の平均値が上昇することで生じているのかもしれませんが、また、転入者の賦課金額の予測値が増加する場合、移住誘致政策が影響しているのかもしれませんが、ある大企業が新しく拠点を設けたため転入者人口が増加しているのかもしれませんが、これらの因果関係を特定するためには、経済学的知見に基づくモデルで推定することや実験的環境を作り出すことが求められます。そのため、本プロジェクトにおいては、税収予測額に影響を及ぼす因果関係やその大きさの特定を行うことが難しい点についてご理解いただけますようお願い申し上げます。

また、本報告書が行なった税収予測は、あくまでも過去のデータの特徴を将来に当てはめた結果に過ぎません。特に感染症の蔓延や自然災害、金融危機など、予測不可能な事態が発生した場合は予測精度が著しく低下する可能性がある点についても合わせてご留意いただきますようお願いいたします。

4 記述統計

前節では、税収予測の限界について述べました。しかしながら、今回頂いたデータは税収予測のみならず、賦課額の足元での動きを把握するのに役立つと考えております。報告書での分析を予算編成や各政策領域における企画・立案に活用いただくため、税務データを人口・所得階層などの観点から分析した記述統計を本章では提供しております。記述統計での分析は、税収予測結果との因果関係を説明できるわけではございませんが、税収に影響を与え得る要因について足元の動向を把握するためには十分に有用であると認識しております。

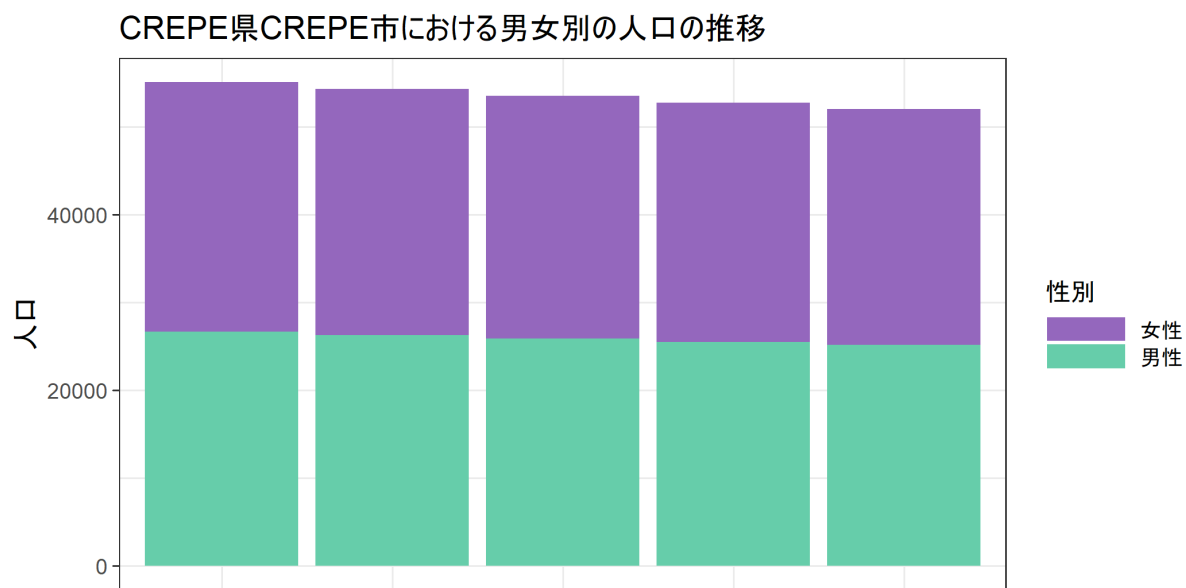
グループ別での賦課額の動きを観察することで、税収の足元での動きを把握できます。例えば、市町村民税の賦課総額に影響する要素として、総人口・年齢階級・所得階層・所得区分などが考えられます。これらの要素が税収に与える影響を計量的に分析することは、正確性を担保できないため本報告書では差し控えますが、これらの要因の基本統計や時系列変化を図表で確認することで、転

入（出生も含む）・転出（死亡も含む）の確率推定による税収予測だけでは捉えられない税収動向の変化を捉えることができます。

なお、重ねての説明となり恐縮ですが、要因の時系列変化は必ずしも市町村民税の賦課総額の予測額の増減に影響を与える「因果関係」を示しているわけではない点についてご留意いただきたく存じます。

4.1 基本統計

4.1.1 人口の推移

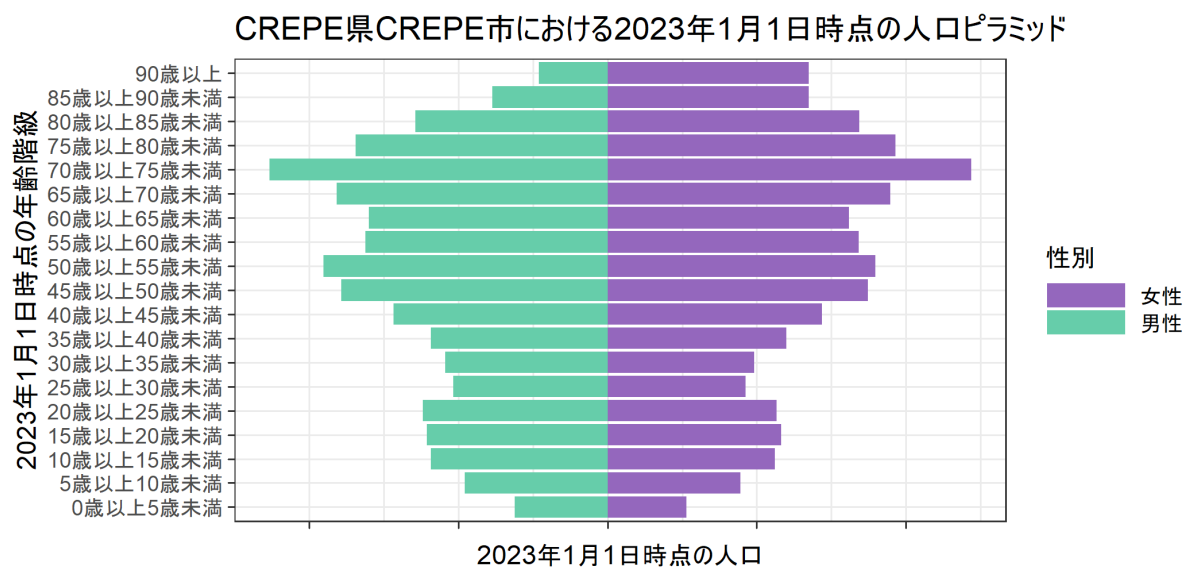


備考：ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。
1月1日時点に合わせた暦年で表しています。

図2 人口の推移

性別	2019年	2020年	2021年	2022年	2023年
女性	28,461	28,095	27,643	27,208	26,867
男性	26,623	26,239	25,861	25,510	25,140

4.1.2 人口ピラミッド（男女・年齢階級別）



備考：男性の人口がマイナスの数値を取るのはプログラミング言語の仕様であり、その絶対値が人口を表します。
 ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。
 1月1日時点に合わせた暦年で表しています。

図3 人口ピラミッド（男女・年齢階級別）

年齢階級	女性	男性
0 歳以上 5 歳未満	530	623
5 歳以上 10 歳未満	892	956
10 歳以上 15 歳未満	1,121	1,182
15 歳以上 20 歳未満	1,162	1,211
20 歳以上 25 歳未満	1,134	1,236
25 歳以上 30 歳未満	925	1,035
30 歳以上 35 歳未満	982	1,086
35 歳以上 40 歳未満	1,200	1,183
40 歳以上 45 歳未満	1,435	1,432
45 歳以上 50 歳未満	1,745	1,785
50 歳以上 55 歳未満	1,793	1,903
55 歳以上 60 歳未満	1,684	1,623
60 歳以上 65 歳未満	1,619	1,600
65 歳以上 70 歳未満	1,895	1,814
70 歳以上 75 歳未満	2,436	2,263
75 歳以上 80 歳未満	1,929	1,688
80 歳以上 85 歳未満	1,688	1,288
85 歳以上 90 歳未満	1,350	771
90 歳以上	1,347	461

4.2 基本統計（税金）

4.2.1 税金項目に関する基本統計

表 4 分析に用いた税金データの基本統計

変数名	年度の数	観察数	平均	最小値	第 1 四分位点	中央値	第 3 四分位点	最大値
市町村民税賦課総額	5	267,647	39,880	0	0	0	51,300	16,211,500
市町村民税均等割額	5	267,647	1,483	0	0	0	3,500	3,500
市町村民税所得割額	5	267,647	33,656	0	0	0	36,700	16,208,000
総所得金額	5	267,647	1,306,807	-6,741,091	0	470,049	1,953,600	503,506,129

* トップコーディングを選択された場合は、最大値は匿名化前のデータとは異なります。

* 平均値は小数第一位を四捨五入し、整数で表しています。

4.2.2 割別（均等割・所得割）の市町村民税の賦課総額の推移

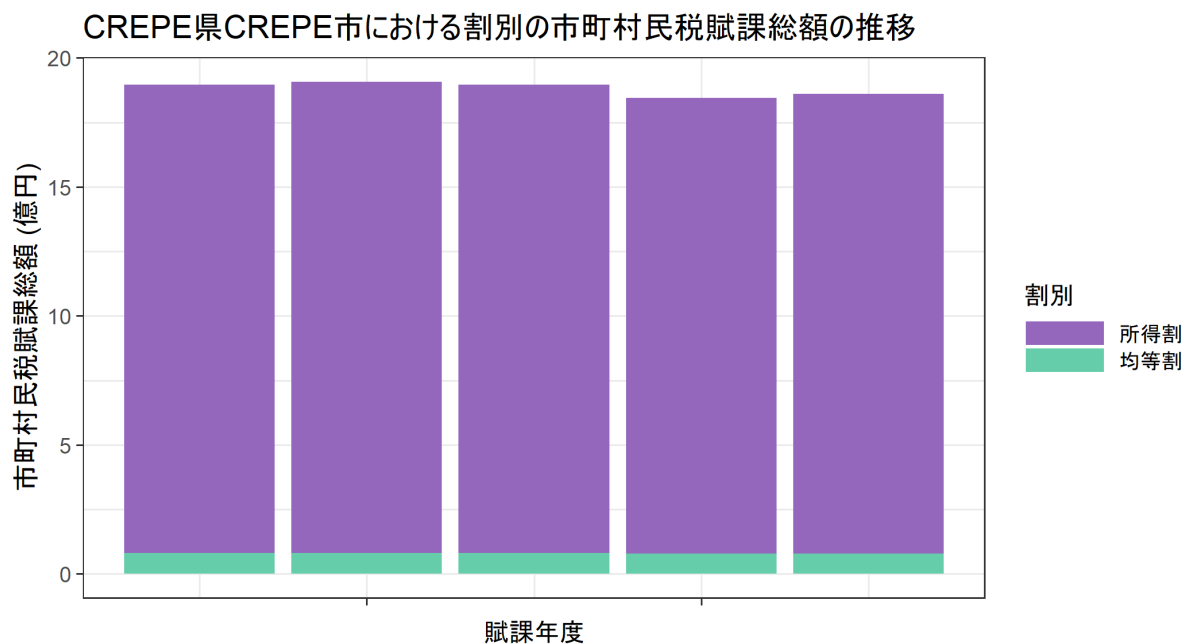


図4 割別（均等割・所得割）の市町村民税の賦課総額の推移

割別	2019年	2020年	2021年	2022年	2023年
均等割	80,465,000	80,236,000	79,826,100	78,853,700	77,521,500
所得割	1,815,004,800	1,828,092,400	1,816,563,700	1,765,698,400	1,782,628,500

4.3 市町村民税分析（性別・年齢階級別）

4.3.1 性別・年齢階級別の市町村民税均等割・所得割額

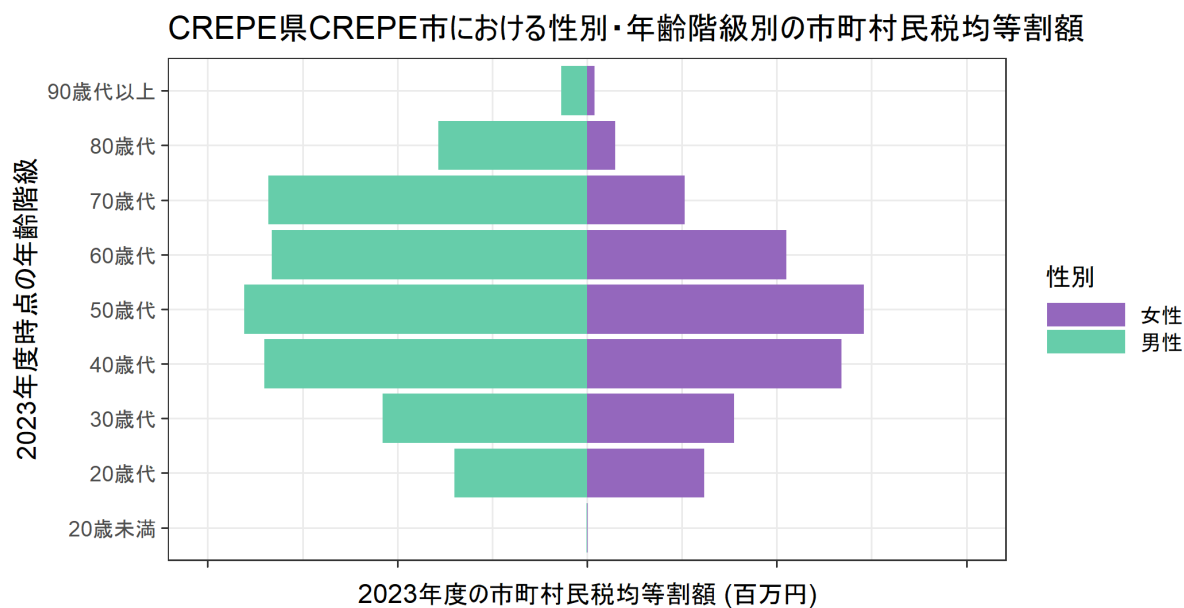


図5 性別・年齢階級別の住民税均等割額

年齢階級	男性	女性
20歳未満	17,500	10,500
20歳代	3,503,500	3,090,500
30歳代	5,390,000	3,874,500
40歳代	8,515,500	6,709,500
50歳代	9,044,000	7,297,500
60歳代	8,305,500	5,257,000
70歳代	8,407,000	2,565,500
80歳代	3,927,000	742,000
90歳代以上	675,500	189,000

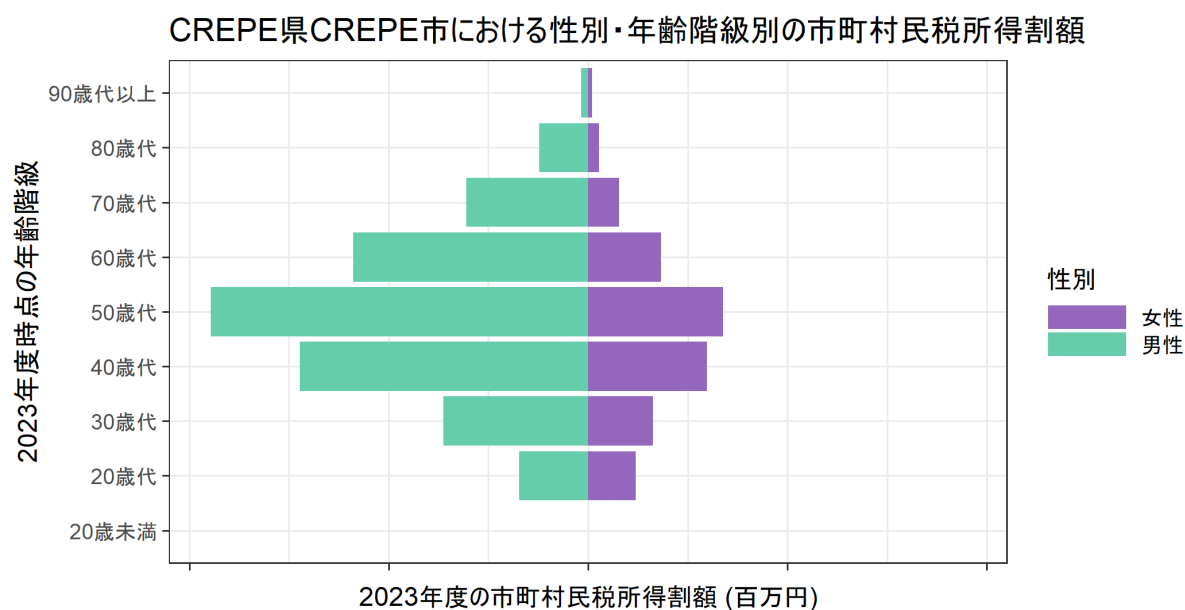
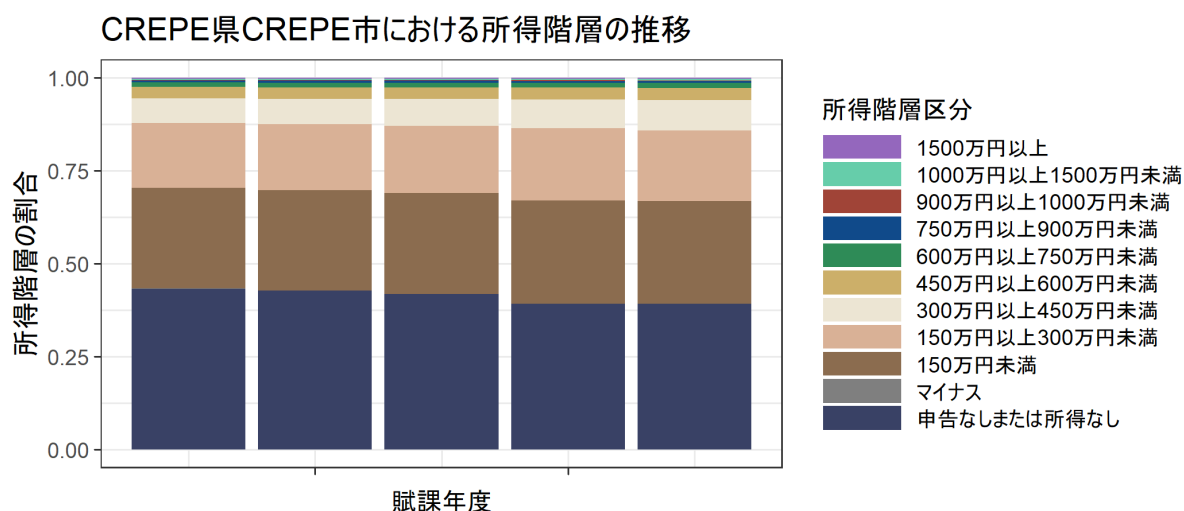


図6 性別・年齢階級別の市町村民税所得割額

年齢階級	男性	女性
20歳未満	261,300	143,800
20歳代	69,059,300	47,839,700
30歳代	144,944,500	65,238,400
40歳代	289,506,600	119,191,100
50歳代	378,486,700	135,334,300
60歳代	235,806,500	73,107,900
70歳代	122,192,300	30,845,300
80歳代	48,922,200	10,714,600
90歳代以上	7,156,300	3,877,700

4.4 市町村民税分析（所得階層別）

4.4.1 所得階層別の人口割合



備考：前年の所得額が賦課対象になっているため、賦課対象となる前年の所得を表しています。

図7 所得階層別の人口の割合

所得階層区分	2019年	2020年	2021年	2022年	2023年
1500万円以上	128	141	147	160	164
1000万円以上 1500万円未満	164	162	157	151	174
900万円以上 1000万円未満	75	85	95	103	98
750万円以上 900万円未満	312	290	279	249	292
600万円以上 750万円未満	667	714	672	701	651
450万円以上 600万円未満	1,654	1,655	1,643	1,672	1,727
300万円以上 450万円未満	3,724	3,724	3,911	4,101	4,227
150万円以上 300万円未満	9,555	9,631	9,648	10,229	9,926
150万円未満	14,937	14,681	14,551	14,671	14,335
マイナス	27	33	31	30	36
申告なしまたは所得なし	23,841	23,218	22,370	20,651	20,377

4.4.2 所得階層別の市町村民税均等割額

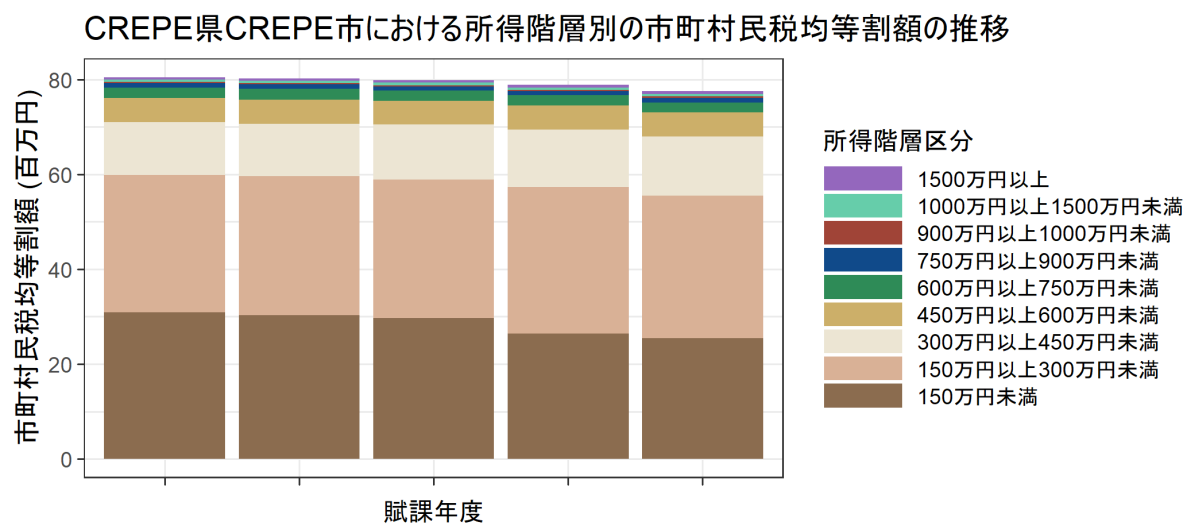


図 8 所得階層別の市町村民税均等割額の推移

所得階層区分	2019年	2020年	2021年	2022年	2023年
1500万円以上	437,500	469,000	490,000	514,500	539,000
1000万円以上 1500万円未満	528,500	535,500	521,500	504,000	574,000
900万円以上 1000万円未満	252,000	283,500	308,000	336,000	301,000
750万円以上 900万円未満	987,000	920,500	885,500	770,000	917,000
600万円以上 750万円未満	2,128,000	2,254,000	2,114,000	2,219,000	2,082,500
450万円以上 600万円未満	5,096,000	5,113,500	5,015,500	5,050,500	5,173,000
300万円以上 450万円未満	11,165,000	11,084,500	11,557,000	12,096,000	12,407,500
150万円以上 300万円未満	28,920,500	29,330,000	29,281,000	30,954,000	30,128,000
150万円未満	30,950,500	30,245,500	29,653,600	26,409,700	25,396,000
マイナス	0	0	0	0	0
申告なしまたは所得なし	0	0	0	0	3,500

4.4.3 所得階層別の市町村民税所得割額

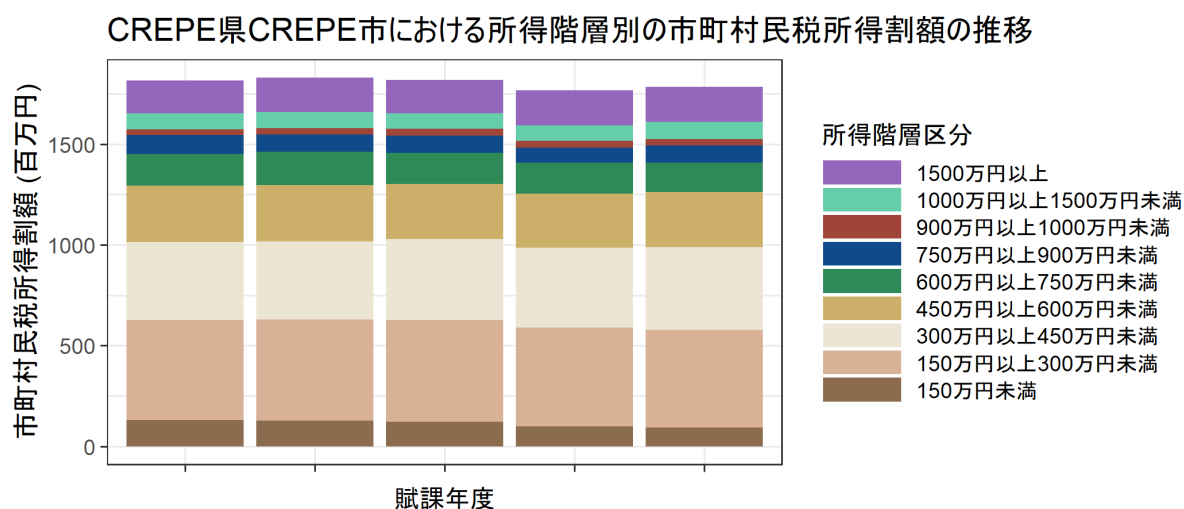
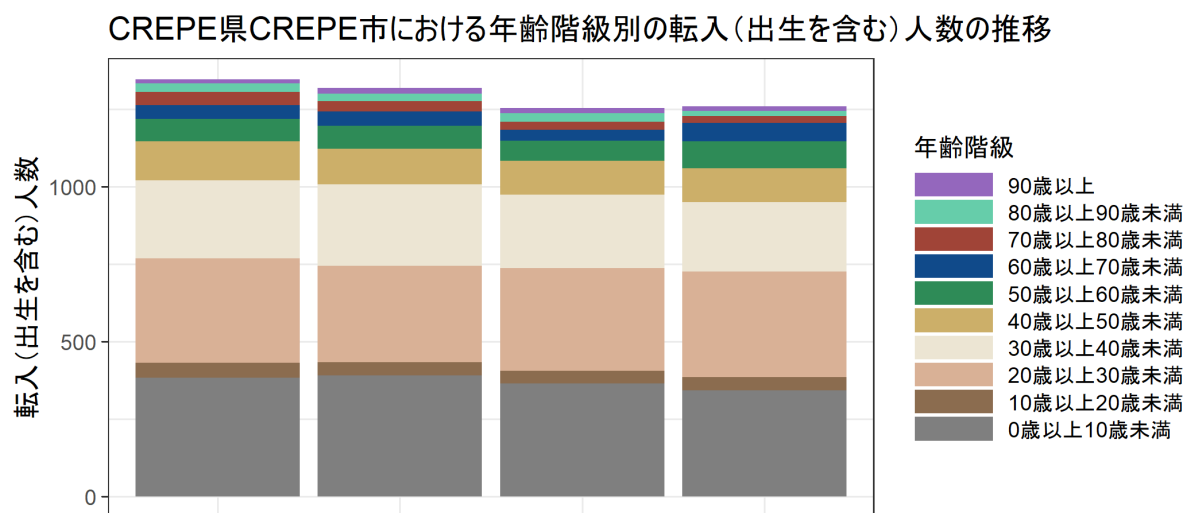


図9 所得階層別の市町村民税所得割額の推移

所得階層区分	2019年	2020年	2021年	2022年	2023年
1500万円以上	162,032,600	170,375,300	164,102,800	174,258,700	172,250,900
1000万円以上 1500万円未満	79,104,400	78,198,100	77,896,200	74,494,500	85,035,000
900万円以上 1000万円未満	29,277,900	32,965,800	34,493,600	36,891,200	32,549,900
750万円以上 900万円未満	94,024,000	86,157,100	83,995,000	71,480,400	84,144,300
600万円以上 750万円未満	157,249,800	165,444,700	154,572,000	156,603,300	146,763,800
450万円以上 600万円未満	279,188,000	279,428,800	273,884,300	266,266,800	272,788,500
300万円以上 450万円未満	387,715,200	385,854,100	400,899,600	395,942,300	412,146,500
150万円以上 300万円未満	496,674,200	502,891,100	503,903,300	491,379,700	482,621,900
150万円未満	129,738,700	126,777,400	122,816,900	98,381,500	94,327,700
マイナス	0	0	0	0	0
申告なしまたは所得なし	0	0	0	0	0

4.5 転入・転出の分析

4.5.1 転入（出生を含む）の年齢階級別の人数

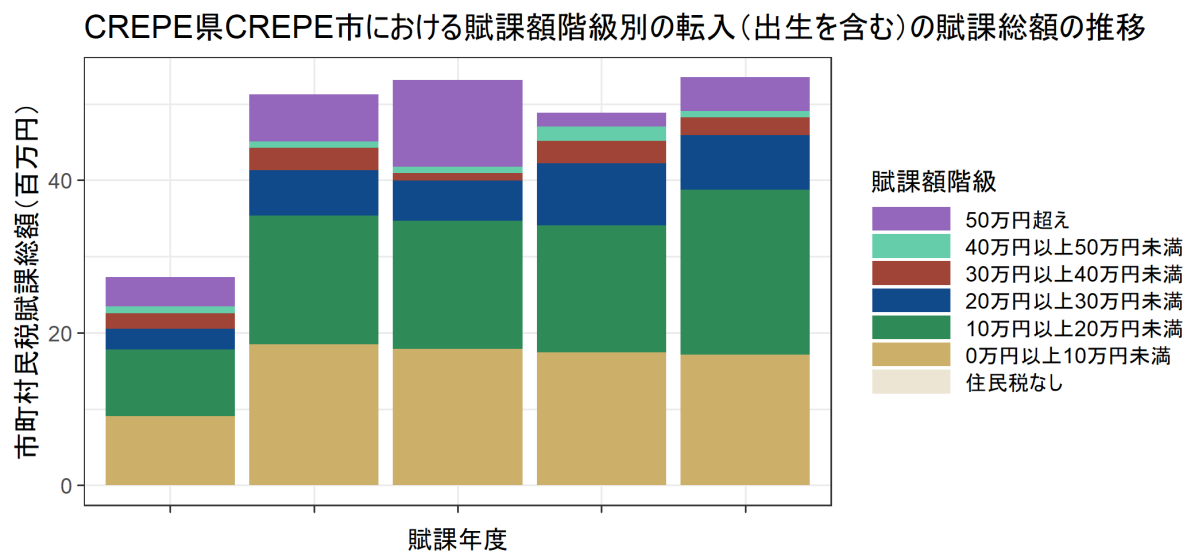


備考：1月1日時点に合わせた暦年で表しています。
ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。

図 10 転入（出生を含む）の年齢階級別の人数

年齢階級	2019年	2020年	2021年	2022年
90歳以上	14	18	17	15
80歳以上90歳未満	28	24	27	16
70歳以上80歳未満	41	34	26	22
60歳以上70歳未満	45	45	35	60
50歳以上60歳未満	72	75	66	86
40歳以上50歳未満	126	115	108	110
30歳以上40歳未満	251	262	238	223
20歳以上30歳未満	337	310	331	340
10歳以上20歳未満	49	44	41	44
0歳以上10歳未満	383	390	364	342

4.5.2 転入（出生を含む）の賦課額階級別の賦課総額



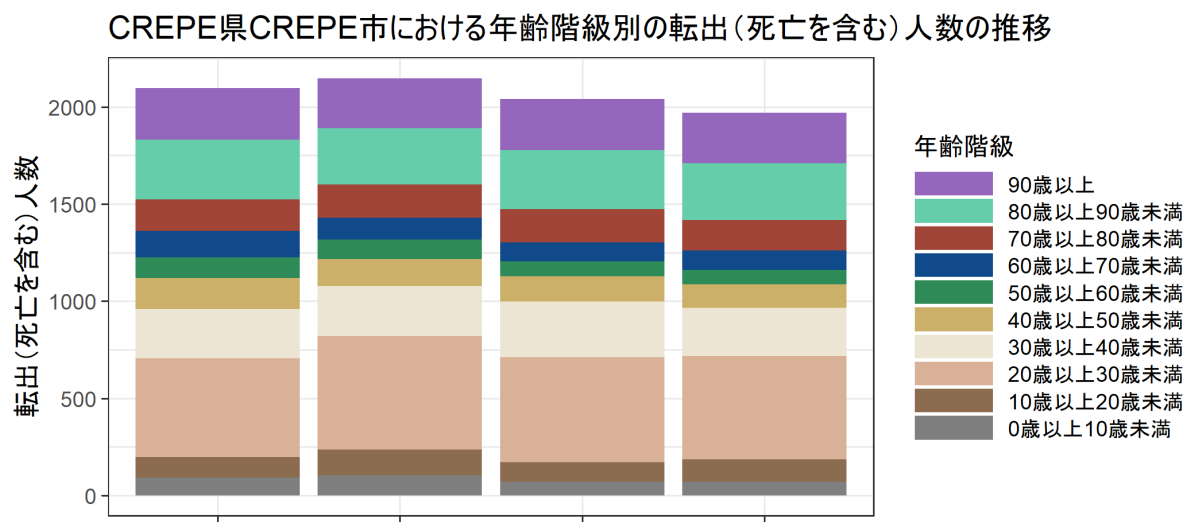
備考：賦課年度の前年の1月1日から12月31日までに自治体に転入した人の賦課税額を表しています。
1月1日時点に合わせた暦年で表しています。
ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。

図 11 転入（出生を含む）の賦課額階級の賦課総額

賦課額階級	2019年	2020年	2021年	2022年	2023年
住民税なし	0	0	0	0	0
0万円以上10万円未満	9,090,000	18,480,000	17,900,000	17,450,000	17,110,000
10万円以上20万円未満	8,700,000	16,910,000	16,740,000	16,600,000	21,610,000
20万円以上30万円未満	2,720,000	5,920,000	5,310,000	8,170,000	7,200,000
30万円以上40万円未満	2,010,000	2,950,000	1,010,000	2,960,000	2,290,000
40万円以上50万円未満	940,000	830,000	810,000	1,860,000	840,000
50万円超え	3,850,000	6,150,000	11,340,000	1,790,000	4,420,000

4.5.3 転出（死亡を含む）の年齢階級別の人数

翌年度にデータが存在しない場合を「転出（死亡を含む）」と定義しているため、例えば2021年の棒グラフに含まれる転出者（死亡を含む）は、翌2022年に転出（死亡を含む）する人です。



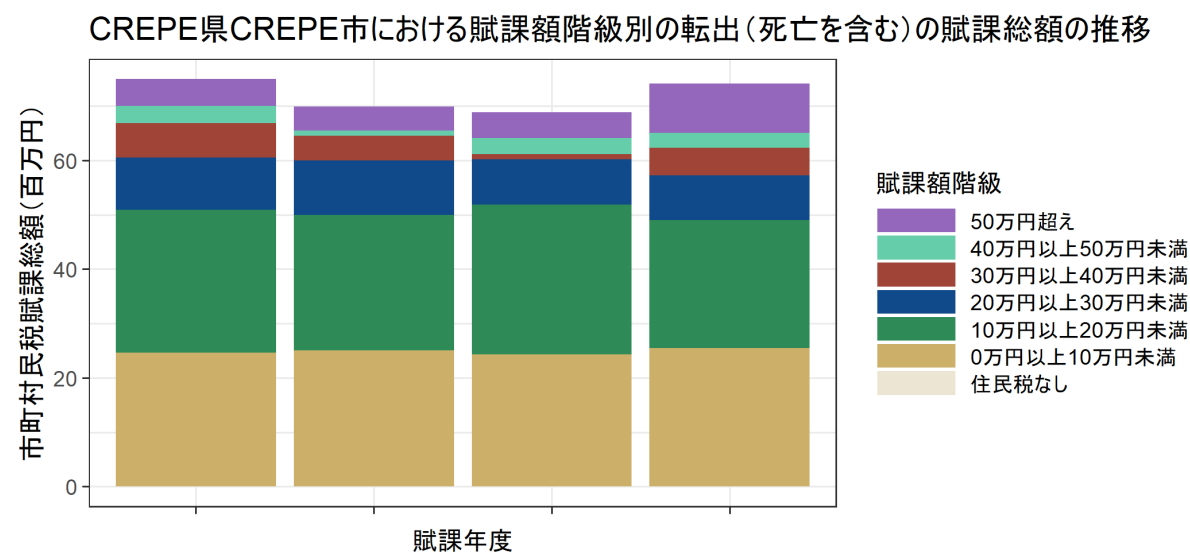
備考：1月1日時点に合わせた暦年で表しています。
ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。

図12 転出（死亡を含む）の年齢階級別の人数

年齢階級	2019年	2020年	2021年	2022年
90歳以上	267	259	263	260
80歳以上90歳未満	306	287	303	293
70歳以上80歳未満	161	171	170	156
60歳以上70歳未満	137	112	99	100
50歳以上60歳未満	105	101	76	74
40歳以上50歳未満	161	140	130	121
30歳以上40歳未満	253	256	287	247
20歳以上30歳未満	507	584	540	532
10歳以上20歳未満	107	132	101	115
0歳以上10歳未満	92	105	70	71

4.5.4 転出（死亡を含む）の賦課額階級別の賦課総額

翌年度にデータが存在しない場合を「転出（死亡を含む）」と定義しているため、例えば2021年の棒グラフに含まれる転出者（死亡を含む）は、翌2022年に転出（死亡を含む）する人です。



備考：賦課年度の1月1日から12月31日までに自治体に転出した人の賦課税額を表しています。1月1日時点に合わせた暦年で表しています。ご提供いただいたデータから作成しているため、実際の状況とは異なる可能性があります。

図13 転出（死亡を含む）の賦課額階級別の賦課総額

賦課額階級	2019年	2020年	2021年	2022年
住民税なし	0	0	0	0
0万円以上10万円未満	24,630,000	25,050,000	24,340,000	25,430,000
10万円以上20万円未満	26,270,000	24,950,000	27,490,000	23,520,000
20万円以上30万円未満	9,560,000	9,940,000	8,350,000	8,230,000
30万円以上40万円未満	6,350,000	4,540,000	960,000	5,140,000
40万円以上50万円未満	3,160,000	960,000	2,990,000	2,700,000
50万円超え	4,940,000	4,460,000	4,660,000	9,050,000

4.6 所得項目の分析（区分別）

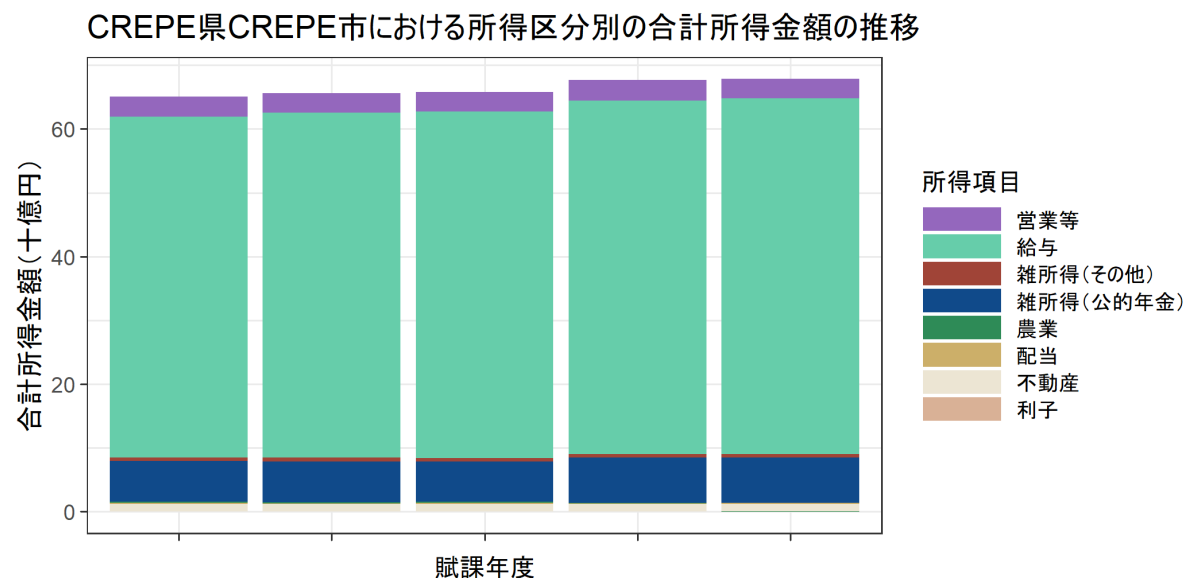


図 14 合計所得金額に占める所得区分別割合

産業区分	2019年	2020年	2021年	2022年	2023年
営業等	3,139,718,244	3,077,711,036	3,076,776,072	3,226,766,374	3,114,526,849
農業	179,995,165	140,455,711	146,584,393	73,825,018	-55,205,075
不動産	1,184,523,327	1,172,713,301	1,217,238,265	1,155,592,350	1,176,885,604
利子	357,726	625,644	560,787	507,812	434,318
配当	143,159,280	124,686,771	144,791,857	116,433,585	180,178,460
給与	53,504,281,975	54,113,668,769	54,385,639,599	55,503,965,710	55,808,275,370
雑所得（公的年金）	6,390,151,735	6,421,062,482	6,379,652,031	7,157,064,994	7,155,849,503
雑所得（その他）	545,979,980	573,984,493	491,226,274	479,592,498	464,318,077

4.7 所得分布・格差動向に関する分析

4.7.1 性別・年齢ごとの就業率

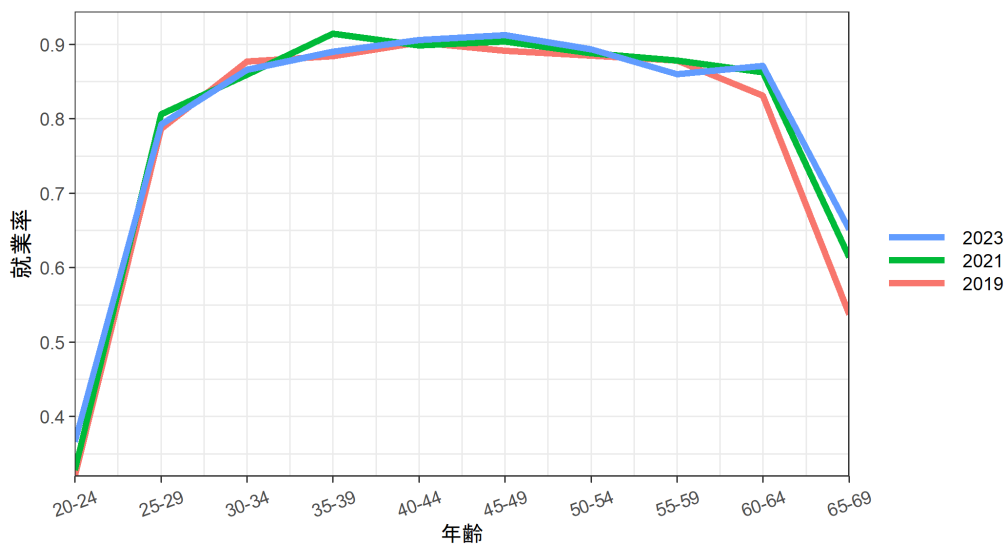


図 15 就業率 (男性)

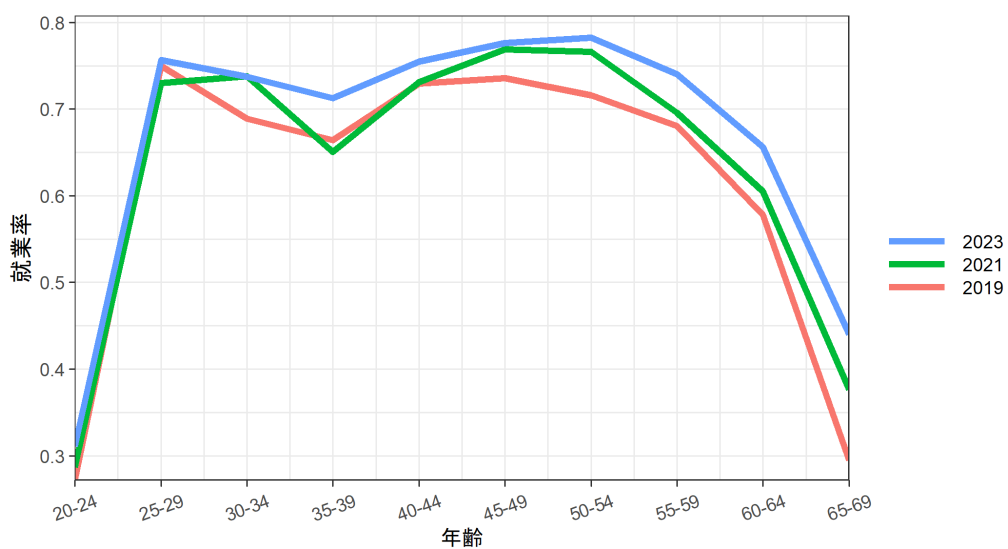


図 16 就業率 (女性)

4.7.2 性別・年齢ごとの平均所得

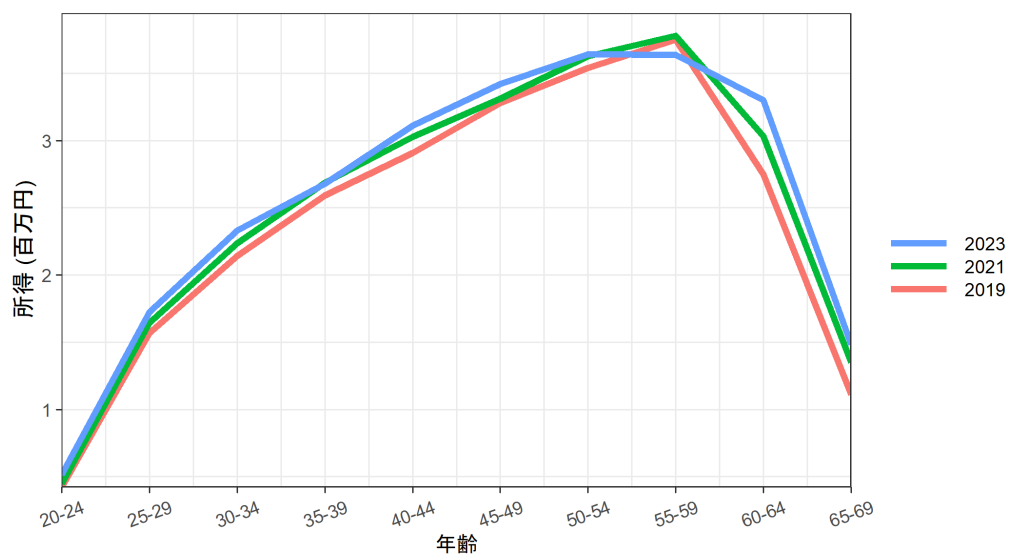


図 17 平均所得 (男性)

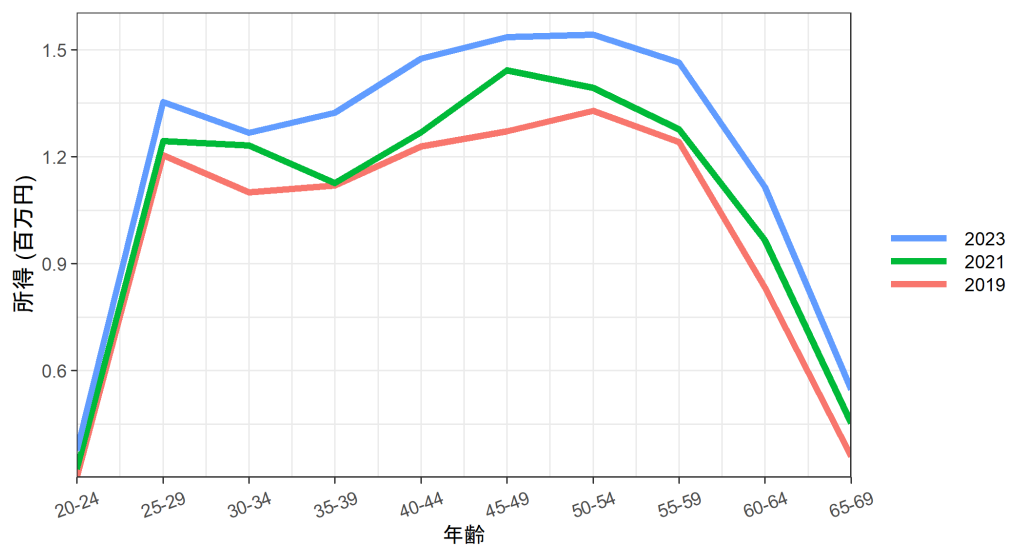


図 18 平均所得 (女性)

4.7.3 性別・年齢ごとの所得パーセンタイル (2023 年)

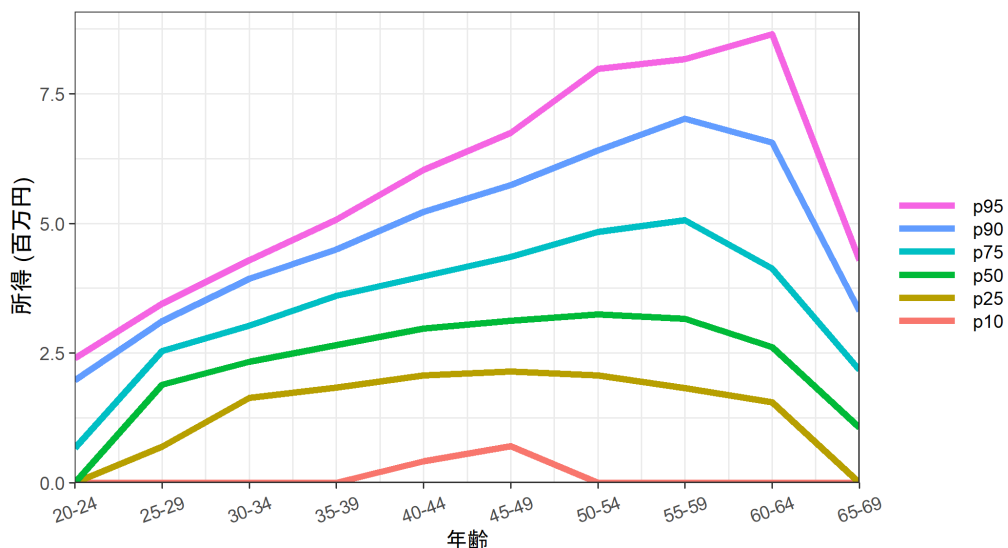


図 19 所得パーセンタイル (男性)

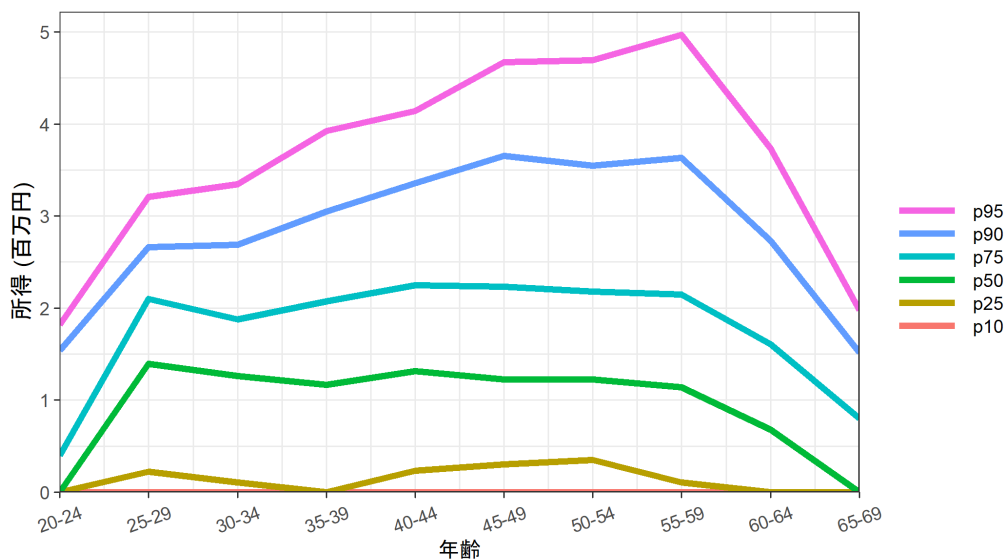


図 20 所得パーセンタイル (女性)

4.7.4 性別・年齢ごとの所得ジニ係数

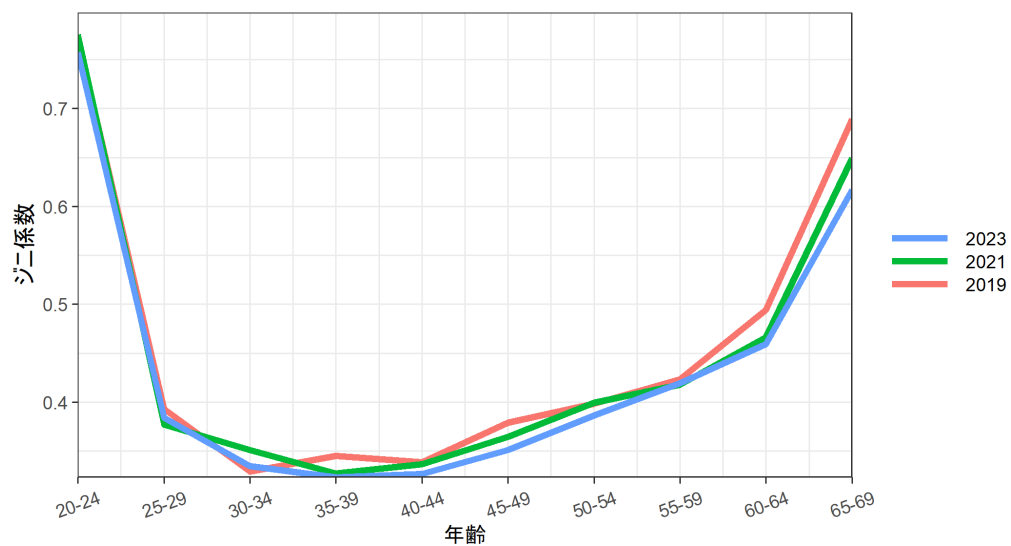


図 21 所得ジニ係数 (男性)

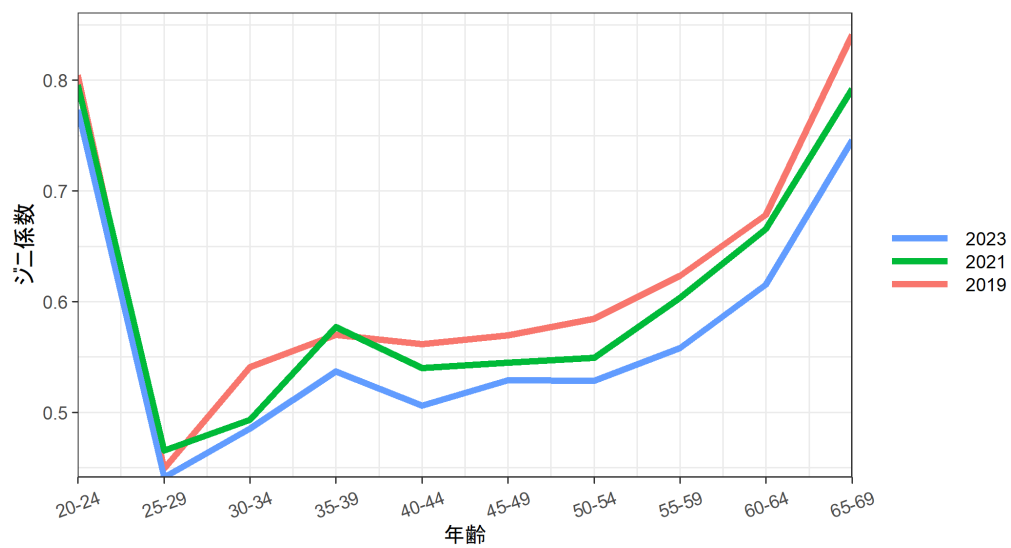


図 22 所得ジニ係数 (女性)

4.8 男性の給与収入と既婚率の関係

4.8.1 全年齢における男性の給与収入と既婚率

** 実際の報告書ではグラフが挿入されます。 **

4.8.2 男性の年齢別・給与収入と既婚率

** 実際の報告書ではグラフが挿入されます。 **

4.9 出産前後の女性の給与収入・就業率の変化

4.9.1 出産前後の女性の給与収入の変化

** 実際の報告書ではグラフが挿入されます。 **

4.9.2 出産前後の女性の就業率の変化

** 実際の報告書ではグラフが挿入されます。 **

4.10 高齢者の就業率

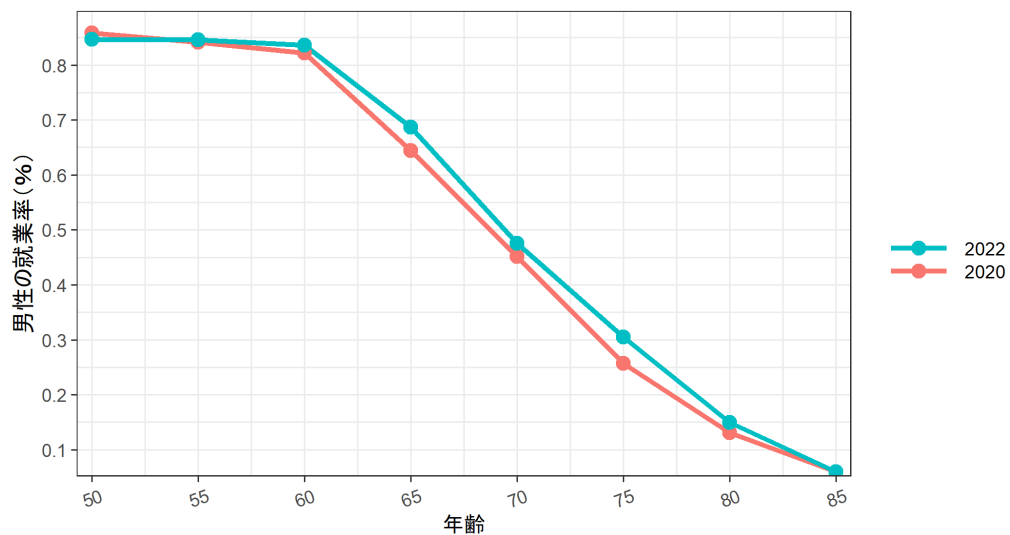


図 23 高齢者の就業率 (男性)

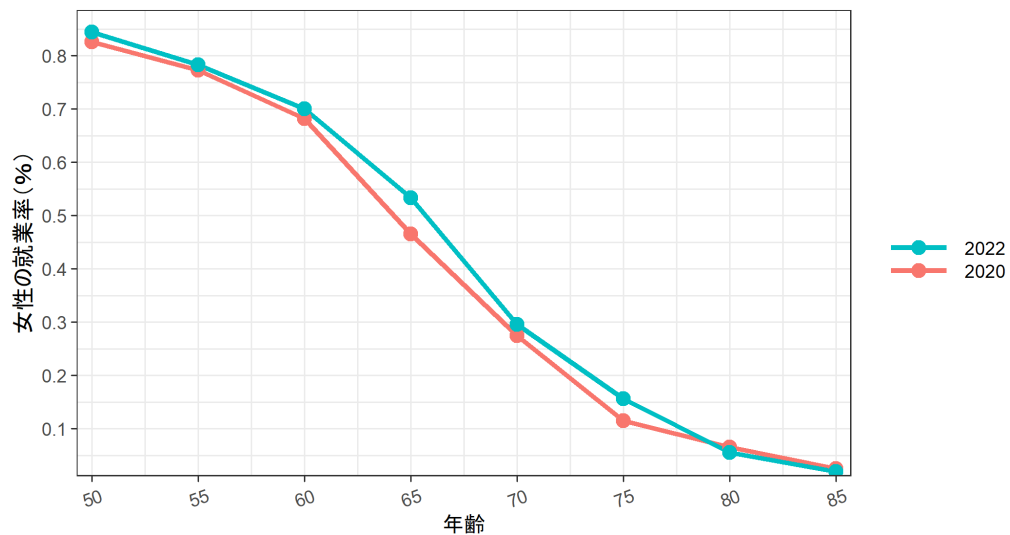


図 24 高齢者の就業率 (女性)

5 補遺

5.1 変数の作成過程について

本節では、税収予測ならびに記述統計に用いた変数の作成方法について記述します。3.3 データに記述した通り本報告書ではご提供いただいた税務データの中でも、賦課年度や個人属性に関する変数のみならず、市町村民税や控除前の所得金額、所得合計などの変数を分析に使用しました。ただし、税収予測や記述統計などの目的に基づいて分析を行う際には、変数を加工して新たに変数を作成する必要があります。具体的には、本報告書では個人の税収を予測する際に、個人の属性（例：年齢や性別）や賦課状況（例：過去年度の賦課金額）を利用しておりますが、分析に適したデータにするためにデータクリーニングを行う必要があります。メールにて変数の定義や対応関係についてご質問差し上げたのはこの工程を正確に実行するためです。本節では特に、ご提供いただいた税務データの状態からどのように目的に合わせて変数を作成したのかに焦点を当て解説し、税収予測や記述統計の分析結果が何を意味しているのかをより正確に把握していただくことを目指しました。今後変数に関してお問合せをさせていただく際に、その意図を正しくご理解いただくための補助資料として活用いただけましたら幸いです。

全ての処理に共通して使用するのは賦課年度と宛名番号です。宛名番号は匿名化処理の段階でハッシュ化されていますが、税務データ上の同一個人を識別するために用います。

税収予測には①個人の年齢区分、②性別、③ある年にある個人が転入（出生を含む）したか・転出（死亡を含む）したかのダミー変数、④賦課金額が未申告であるか・0であるかのダミー変数を、ご提供いただいたデータから使用しています。ここでダミー変数とは、0か1かのどちらかを取る変数のことを指します。

①個人の年齢区分は、ある個人がある年度において何歳代に属するのを示すカテゴリー変数です*¹¹。個人の生年月と賦課年度を用いて算出した年齢を元に、個人を5歳ごとの年齢区分ごとのカテゴリーに割り当てます。年少者ならびに高齢者は所得が限りなく0に近いため、税収予測の精度を向上させる目的で、15歳以下の個人は15歳～20歳のカテゴリーに、90歳以上の個人は90歳以上のカテゴリーに割り振っています。

②性別は欠損値や男女以外のカテゴリーが存在する場合、その他のカテゴリーに置き換えています。③転入（出生を含む）のダミー変数は、ある宛名番号が前年度に税務データに存在するのかを、転出（死亡を含む）のダミー変数は次年度に税務データに存在するのかを判定し、作成しています。なお、住基データと税務データを突合してデータセットを作成している関係上、住基登録外であるが税務データに宛名番号が存在する個人については住民として扱っています。ここでの転入・転出は住基データまたは課税データにレコードが新規登録されることとレコードが削除されることを指します。また、出生・死亡フラグ等の変数を用いて識別しているわけではないため、転入と出生、転出と死亡は切り分けられておりません。

最後に④賦課金額が未申告であるか・0であるかのダミー変数は、ある宛名番号の市町村民税について、データの中身が未申告を表すNAまたは0である場合に、1を取るかどうかを判定し、変数を作成しています。

*¹¹ 高度な匿名化水準を選択された場合や中間的な匿名化水準のうち3匿名性を選択された場合は、四半期単位に丸めた生年月月を利用しています。

次に、記述統計には⑤各種収入・所得金額⑥所得合計⑦市町村民税均等割・所得割額を使用しています。全ての変数について欠損値を0に置き換える処理を行った上で、⑤各種収入・所得金額については給与所得（一般）が給与所得（専従）の金額を足し合わせた変数であるか、足し合わせる前の変数であるかを自治体の税務データの実態に合わせて条件分岐の処理をかけることで、給与収入・所得の合計金額を算出しています。⑦市町村民税均等割・所得割額については、減免額を差し引いた数値でない場合は各割別で賦課額から差し引いています。減免額が割別で存在しない場合については、市町村民税所得割額から先に差し引き、残りを市町村民税均等割額から差し引いています。

5.2 税収予測のモデルについて

税収予測の章における予測方法の説明では予測モデルの内容の説明である回帰式やパラメータの推定方法を割愛いたしました。本節では、税収予測の予測モデルの詳細を説明します。

5.2.1 継続居住者のデータを用いた次年度の賦課額予測

まず、ある年に居住している個人の次年度の賦課額を予測します。

$$y_{i,t+1} = \alpha + \beta y_{i,t} + X_{i,t} \gamma + M_{m,t} \tau + u_{i,t+1} \quad (3)$$

$y_{i,t+1}$ は、個人 i の $t+1$ 年の賦課額の観測値です。同様に、 $y_{i,t}$ は個人 i の t 年の賦課額の観測値です。 $X_{i,t}$ は、 t 年における個人 i の特徴で賦課額 $y_{i,t+1}$ に影響を与えうる変数を選択しています。CREPE 市では、課税対象外ダミー変数と市町村民税賦課額 0 円ダミー変数、性別、また年齢区分の交差項を $X_{i,t}$ として用います。 $M_{m,t}$ は CREPE 市が属する都道府県のマクロ変数（有効求人倍率など）、 $u_{i,t+1}$ は誤差です。 α 、 β 、 γ 、 τ はこの回帰式のパラメータで、真の値は観察できませんが最小二乗法により推定値を得ることができます。パラメータの推定値をそれぞれ $\hat{\alpha}$ 、 $\hat{\beta}$ 、 $\hat{\gamma}$ 、 $\hat{\tau}$ とします。 $\hat{\tau}$ については、マクロ変数を入れないモデル、すなわち $\tau = 0$ のケースと比較してより予測精度の高いものを採用するアルゴリズムとなっています*12。

このモデルの推定には、 $t+1$ 年と t 年の 2 年分のデータが必要となります。よって、この予測モデルに用いるサンプルは、 $t+1$ 年だけでなく、 t 年にも税収を観察できる個人に限定します。

式 (3) で推定した $\hat{\alpha}$ 、 $\hat{\beta}$ 、 $\hat{\gamma}$ 、 $\hat{\tau}$ を用いて、以下の式 (4) で、個人 i の次年度の賦課額 $\hat{y}_{i,t+1}$ を予測します。

$$\hat{y}_{i,t+1} = \hat{\alpha} + \hat{\beta} y_{i,t} + X_{i,t} \hat{\gamma} + M_{m,t} \hat{\tau} \quad (4)$$

式 (4) で得られる $\hat{y}_{i,t+1}$ は、居住を継続した場合の納税額の予測です。当然、個人が転出又は死亡する可能性があるため、考慮する必要があります。また、転出や死亡の可能性は個人の属性によって異なります。例えば、税収の多い人が転出する可能性が高いかもしれません。よって、個人が転出または死亡する確率を求める必要があります。

*12

5.2.2 居住継続確率の計算

転出・死亡の確率を計算するために次の式を考えます。

$$\mathbb{P}(out_{i,t+1}) = \frac{e^{\alpha + X_{i,t}\beta}}{1 + e^{\alpha + X_{i,t}\beta}} \quad (5)$$

$out_{i,t+1}$ は $t+1$ 年に個人 i が CREPE 市から転出・死亡したかどうかのダミー変数です。転出か死亡したら、 $out_{i,t+1}$ が 1 をとり、それ以外の場合は 0 をとります。式 (3) での左辺が連続値である場合とは異なり $out_{i,t+1}$ の観察値は 0 か 1 しかとりません。よって、ここでは二項ロジスティック回帰を用います。 $\mathbb{P}(out_{i,t+1})$ は $out_{i,t+1} = 1$ になる確率です。 e はネイピア数を表しています。 $X_{i,t}$ は、性別と年齢区分の交差項や課税対象外ダミー変数と市町村民税賦課額 0 円ダミー変数です。この式のパラメータ α 、 β は最尤法により推定されます。予測対象以外のデータから推定したパラメータ ($\hat{\alpha}$ 、 $\hat{\beta}$) を利用して、以下で各個人 i の翌年度の転出確率・死亡を計算します。

$$\mathbb{P}(out_{i,t+1}) = \frac{e^{\hat{\alpha} + X_{i,t}\hat{\beta}}}{1 + e^{\hat{\alpha} + X_{i,t}\hat{\beta}}} \quad (6)$$

$t+1$ 年における個人の転出・死亡の確率 $\mathbb{P}(out_{i,t+1})$ を計算することができます。よって、居住し続ける確率は $1 - \mathbb{P}(out_{i,t+1})$ となります。

5.2.3 継続居住者の税収総額の予測

以上では、居住継続すると仮定した場合の予測賦課額と居住継続確率を別々に計算しました。最後に以下の式で個人の賦課額の期待値を足し合わせて、賦課総額の予測を行います。

$$\hat{stay}_{t+1} = \sum_{i=1}^{N_t} \{ \hat{y}_{i,t+1} \times [1 - \mathbb{P}(out_{i,t+1})] \} \quad (7)$$

ここでの、 $y_{i,t+1} \times [1 - \mathbb{P}(out_{i,t+1})]$ は各個人 i の $t+1$ 年度の賦課金額の期待値を表します。 N_t は t 年度に居住する個人の数を表します。全個人の期待値を合計することで、 t 年度に居住している個人が $t+1$ 年度に賦課総額が予測できました。

5.2.4 転入者の税収予測

ここまで、居住継続者の税収予測を行いました。CREPE 市を去る個人がいれば、CREPE 市に来る個人も存在します。ここでは、CREPE 市の全転入者の賦課総額の予測方法を説明します。

転入者の税収予測では、二段階の予測を行い、それらを組み合わせることで転入する住民の予測賦課総額が計算できます。一つ目に、性別年齢階級別の各グループ (j) の $t+1$ 年度における転入人数 (\hat{N}_{t+1}^j) を予測します。次に、各グループ (j) 内の平均的な個人の $t+1$ 年度の納税額 (\hat{y}_{t+1}) を予測します。

まず、性別年齢階級別の転入人数 (\hat{N}_{t+1}^j) を予測します。CREPE 市では、それぞれのグループで、次の回帰式のパラメータを最小二乗法で推定します。

$$N_{t+1}^j = \alpha^j + \beta^j N_t^j + u_t^j \quad (8)$$

N_t^j は t 年におけるグループ j の CREPE 市への転入者数です。 u_t^j は誤差項です。つまり、過去の転入の時系列的な傾向を 1 次の自己回帰モデルとして推定します。この推定式から得られた $\hat{\alpha}^j$ や $\hat{\beta}^j$ とを用いて、2023 年の転入者数を \hat{N}_{t+1}^j に代入すると、以下の式のように 2024 年に転入するグループ j の人数を予測することができます。

$$\hat{N}_{t+1}^j = \hat{\alpha}^j + \hat{\beta}^j N_t^j \quad (9)$$

次に、性別年齢階級グループごとに税収を予測します。ここでは、各グループの過去の賦課額の平均値を予測値として使用します。つまり予測値の \hat{y}_{t+1} は $\mathbb{E}[y_t|j, t]$ とみなすこととなります。最後に各グループの合計を表します。

$$\widehat{MoveIn}_{t+1} = \sum_{j \in J} (\mathbb{E}[y_t|j, t] \times \hat{N}_{t+1}^j) \quad (10)$$

5.2.5 税収額の合計

ここまで、居住継続者・転入者の税収総額の予測を別々に行いましたが、最後に合計します。この式 (10) は式 (2) に対応しています。

$$\hat{Tax}_{t+1} = \widehat{MoveIn}_{t+1} + \hat{Stay}_{t+1} \quad (11)$$

このようにして $t+1$ 年度の予測賦課総額が算出されます。