

EBPM推進のための自治体税務データ活用プロジェクト

令和6年(2024年)5月

趣旨

デジタル改革関連法の施行を見据え、複数の地方自治体及び東京大学政策評価研究教育センター(CREPE)が連携し、①政策現場におけるEBPM(evidence-based policy making)の推進と、②アカデミアにおける実証研究の発展を目指す。

背景

地方行政のデジタル化・スマート自治体化が主要な政策課題に

国や自治体におけるEBPMの推進

AI(人工知能)等のビッグデータの分析技術の発展

2000年代以降、行政データを用いた実証経済学が発展

目的

① 政策現場におけるEBPMの推進

- ・最新の経済学や機械学習に基づくデータ分析を利用可能にする。
- ・法学者・情報工学者の参画のもと、個人情報保護と両立したデータ利活用を実現する。

② アカデミアにおける実証研究の発展

- ・匿名化されたマイクロデータを基に研究を実施する。

概要

- ・ 参加自治体が、匿名化された個人レベルの税務情報をCREPEに提供
- ・ CREPEが、計量経済学の知見や機械学習を用いて税収予測等のデータ分析を行い、参加自治体にフィードバック
- ・ CREPEが、提供されたデータに基づき学術研究を実施

概要

募集対象自治体

- **全ての市区町村**

- 2024年度は、都道府県の募集は行いません。

参加自治体には、計量経済学の知見を用いた2025年度の個人住民税の税収予測を提供する予定です。性別や年齢等の属性や、所得割・均等割等の区分に応じた記述統計表も提供します。

参加費

- **無料**

応募要件

- **提供いただいたデータに基づく研究結果の公表に同意すること。**

- 学術論文、研究報告資料、ディスカッション・ペーパー、一般向け記事等の研究成果物の公表を想定しています。

なお、自治体の同意がない限り自治体名を明らかにすることはありません。

研究結果の公表についての考え方は、「自治体税務データ活用プロジェクトにおけるデータ取扱規則(第9条)」を参照ください。

- **次ページの必須データを匿名化した上で、CREPEに提供できること。**

- 匿名化については、CREPEが匿名化のツールを配布するため、匿名加工技術やプログラミングの知識は不要です。

ご参加いただいた自治体には、今後、データ分析の勉強会や研修などのご案内も予定しております。

応募要件

必須データ項目

以下を含む、個人単位のデータ(5年分以上)(※1)

1. 年を示す変数
2. 宛名番号(※2)
3. 収入金額 : 営業等、農業、不動産、利子、配当、給与、給与(専従)、雑所得(公的年金)、雑所得(その他)給与、給与(専従)、雑所得(公的年金)以外は、収入金額がなくても所得金額があれば応募可能です。その他、システムで保持していないデータ項目がある場合は、ご相談ください。
4. 控除金額 : 基礎控除、配偶者控除、配偶者特別控除、扶養控除、寡婦控除、ひとり親控除、勤労学生控除、社会保険料控除、生命保険料控除、地震保険料控除、小規模企業共済等掛金控除、雑損控除、寄付金控除、控除合計
5. 総所得金額、合計所得金額、総所得金額等
6. 住民税額 : 均等割、所得割、住民税額合計(利子割、配当割、株式等譲渡所得割を含んだもの)
7. 属性に関する項目 : 生年月、性別、世帯番号、続柄、死亡年月日、転出年月日、住民になった日、住民区分(住登外納税者や死亡者等、1月1日時点の住民以外の者を識別するための変数)、郵便番号、外国人区分、勤務先を示す識別番号(※3)

「1.年を示す変数」「2.宛名番号」「7.属性に関する項目」は、納税の有無にかかわらず当該自治体の全住民を含むものをご提供ください。

その他の変数については、全課税者が含まれていれば全住民を含んでいる必要はありません。

ここに示す以外のデータ項目も、分析において有用である可能性があるため、できる限り幅広くご提供いただけると幸いです。

※1 2023年度の本プロジェクトに簡易な匿名化でご参加いただいた自治体に関しては、本年度は追加1年分のデータ提供で問題ありません。

※2 自治体独自に各個人に一意に割り振った番号のことであり、「個人番号」いわゆる「マイナンバー」とは異なります。世帯番号とともに、ハッシュ化(p. 12Q7、及び参考資料「個人情報保護制度との関係及び具体的匿名化処理の方法について」を参照)をした上で、ご提供いただきます。

※3 自治体独自で管理されている番号で構いません。

個人情報保護・匿名化

参加自治体には、匿名化したデータをCREPEに提供いただきます。本プロジェクトでは、自治体からの多様なニーズに応えるため、個人情報保護及び匿名化の専門家の監修の下、以下の匿名化手法をご用意しています(※1)。

匿名化手法

0. 氏名、個人番号(いわゆるマイナンバー)を事前に削除
1. 宛名番号、世帯番号、勤務先を示す識別番号のハッシュ化
2. 宛名番号をキーにして、複数年のデータ結合
3. 生年月日の月単位への丸め

匿名化方法について確認いただいた後、CREPEから参加自治体に、R(※2)による匿名化プログラムを配布します。マニュアルに従って、自治体側のPCにて匿名加工を実行いただきます。

CREPEでは、個人識別行為の禁止等を定めた「自治体税務データ活用プロジェクトにおけるデータ取扱規則」、
「自治体税務データ活用プロジェクトにおける安全管理措置等措置等に関する規程」を定めており(※3)、個人情報の管理に万全を尽くしています。

※1 詳細は「(参考資料)個人情報保護制度との関係及び具体的匿名化処理の方法について」を参照ください。最終的に、どこまでの匿名化が必要十分かについては、各自治体において判断いただくこととなりますが、その際、これらの資料を参考にしてください。

※2 学術的な統計分析に広く用いられている無料のプログラミング言語です。

※3 https://web.iss.u-tokyo.ac.jp/jichitai_data/data.html

想定スケジュール

| | | |
|--------------|-------|--|
| 2024年5月 | 東京大学 | プロジェクトへの参加を希望する自治体を募集します。 |
| 6月 | 東京大学 | 協力自治体を選定します。 |
| 7月～8月 | 協力自治体 | 保有する税務データのデータ項目・定義を東京大学にお伝えいただきます。この時点で自治体側にデータを抽出(※1)していただき、「必須データ項目」に該当する変数名等を確認していただきます。また、要配慮個人情報(※2)が含まれている場合は、変数名と定義を確認していただきます。 |
| | 東京大学 | 協力自治体の提供データの概要(データの形式・分布)を確認します。 |
| 9月～11月 | 東京大学 | 匿名化ツールを作成し、協力自治体に送付します。 |
| | 協力自治体 | マニュアルに沿って税務データを匿名加工し、東京大学に送付します。 |
| 11月～ 2025年1月 | 東京大学 | 自治体からの提供データを分析し、詳細な分析結果を返します。 |

※1 匿名化を予定しているデータの概要を教えてください。この段階でデータの抽出をお願いします。データ抽出をシステムベンダに依頼する場合、想定よりも時間を要する場合がございます。お早めにご準備をいただけますと幸いです。

※2 要配慮個人情報とは、個人情報保護法2条3項に規定する「要配慮個人情報」と個人情報保護法60条5項に規定する「条例要配慮個人情報」のことを指します。これらの変数につきましては、研究において必要があるときのみ提供をお願いします。

上記のスケジュールは現段階の想定であり、進捗に応じて、変更する場合があります。

また、2023年度に簡易な匿名化でご参加いただいた自治体の場合は、年内に分析報告書を送付できる可能性がございます。

研究体制

総括班

川口 大司
北尾 早霧
近藤 絢子
古川 知志雄
正木 祐輔

所得リスク・格差班(税務データを活用した所得リスクと所得格差の分析)

- 北尾 早霧 政策研究大学院大学 教授
東京大学政策評価研究教育センター 招聘研究員
- 鈴木 通雄 東北大学大学院経済学研究科 准教授
東京大学政策評価研究教育センター 招聘研究員
- 山田 知明 明治大学商学部 教授
東京大学政策評価研究教育センター 招聘研究員

雇用・社会保障班(セーフティーネットと雇用・家庭)

- 近藤 絢子 東京大学社会科学研究所 教授
- 深井 太洋 学習院大学 経済学部 准教授
東京大学政策評価研究教育センター 招聘研究員

税理論・実験班(最適税制理論モデルと徴税率フィールド実験)

- 古川 知志雄 横浜国立大学大学院国際社会科学研究院・経済学部 准教授
東京大学政策評価研究教育センター 招聘研究員
- 別所 俊一郎 早稲田大学政治経済学術院 教授

学術利用基盤整備班(個人情報保護、匿名化を含めた行政記録情報の学術利用基盤整備)

- 川口 大司 東京大学大学院経済学研究科 教授
東京大学公共政策大学院 院長・教授
東京大学政策評価研究教育センター 前センター長
- 佐藤 一郎 国立情報学研究所 教授
- 穴戸 常寿 東京大学大学院法学政治学研究科 教授
- 正木 祐輔 東京大学公共政策大学院 特任准教授
独立行政法人経済産業研究所 コンサルティングフェロー
(総務省から神戸市に派遣中)

※ 2024年5月時点

※ 「○」は班長

※ 教育や福祉データ等、税務データに紐づけが可能なデータの提供がある場合、他の研究者も参画予定

個人情報保護等に関する法的検討体制

- 本プロジェクトでは、匿名化を行っているものの、個人単位の自治体データを扱うため、個人情報保護等に万全を期す必要があります。
- 学術利用基盤整備班では、法学者を中心とする法学班を設け、個人情報保護法や地方税法をはじめとする法的検討を専門的に行っています。

学術利用基盤整備班(個人情報保護、匿名化を含めた行政記録情報の学術利用基盤整備)

統括補佐

○ 川口 大司 佐藤 一郎 穴戸 常寿 正木 祐輔

細田 幸恵

法学班

| | |
|---------|------------------------|
| 小川 亮 | 國學院大學法学部 専任講師 |
| 海道 俊明 | 関西大学大学院法務研究科 准教授 |
| 神山 弘行 | 東京大学大学院法学政治学研究科 教授 |
| 亀川 達哉 | 東京大学大学院法学政治学研究科 博士後期課程 |
| 北山 昇 | 弁護士(森・濱田松本法律事務所) |
| ○ 穴戸 常寿 | 東京大学大学院法学政治学研究科 教授 |
| 巽 智彦 | 東京大学大学院法学政治学研究科 准教授 |
| 平田 彩子 | 東京大学大学院法学政治学研究科 准教授 |
| 藤谷 武史 | 東京大学社会科学研究所 教授 |
| 船渡 康平 | 信州大学経法学部 准教授 |
| 正木 祐輔 | 東京大学公共政策大学院 特任准教授 |
| 山羽 祥貴 | 東京都立大学法学部 准教授 |

※ 2024年5月時点

※ 「○」は班長

EBPMを主とした自治体への付加価値提供のための体制

- これまで本プロジェクトの実施により、多くの分析結果を自治体にフィードバックし、一定の成果を生み出しつつあります。一方、政策現場におけるEBPMを推進するためには、自治体にとってより付加価値の高い成果を創出する必要があります。
- 学術利用基盤整備班では、EBPM企画班を設け、EBPMを主とした自治体への付加価値提供を目的に、様々な活動を行っています。
- 具体的には、自治体ヒアリング等を通じたデータ分析ニーズ調査、データ分析研修など人的支援、様々なデータ活用による発展的分析といった取組を検討しています。

学術利用基盤整備班(個人情報保護、匿名化を含めた行政記録情報の学術利用基盤整備)

○ 川口 大司 佐藤 一郎 穴戸 常寿 正木 祐輔

統括補佐

細田 幸恵

EBPM企画班

福田 隆巳 東京大学大学院経済学研究科 特任研究員

細田 幸恵 東京大学政策評価研究教育センター招聘研究員

正木 祐輔 東京大学公共政策大学院 特任准教授

○ 由本 聖 東京大学公共政策大学院 特任准教授

※ 2024年5月時点

※ 「○」は班長

応募方法

応募自治体から直接、「応募様式」を電子メールで提出してください。

提出先: jichitai_data[at]e.u-tokyo.ac.jp

* “[at]”の部分を“@”に変えて送信してください。

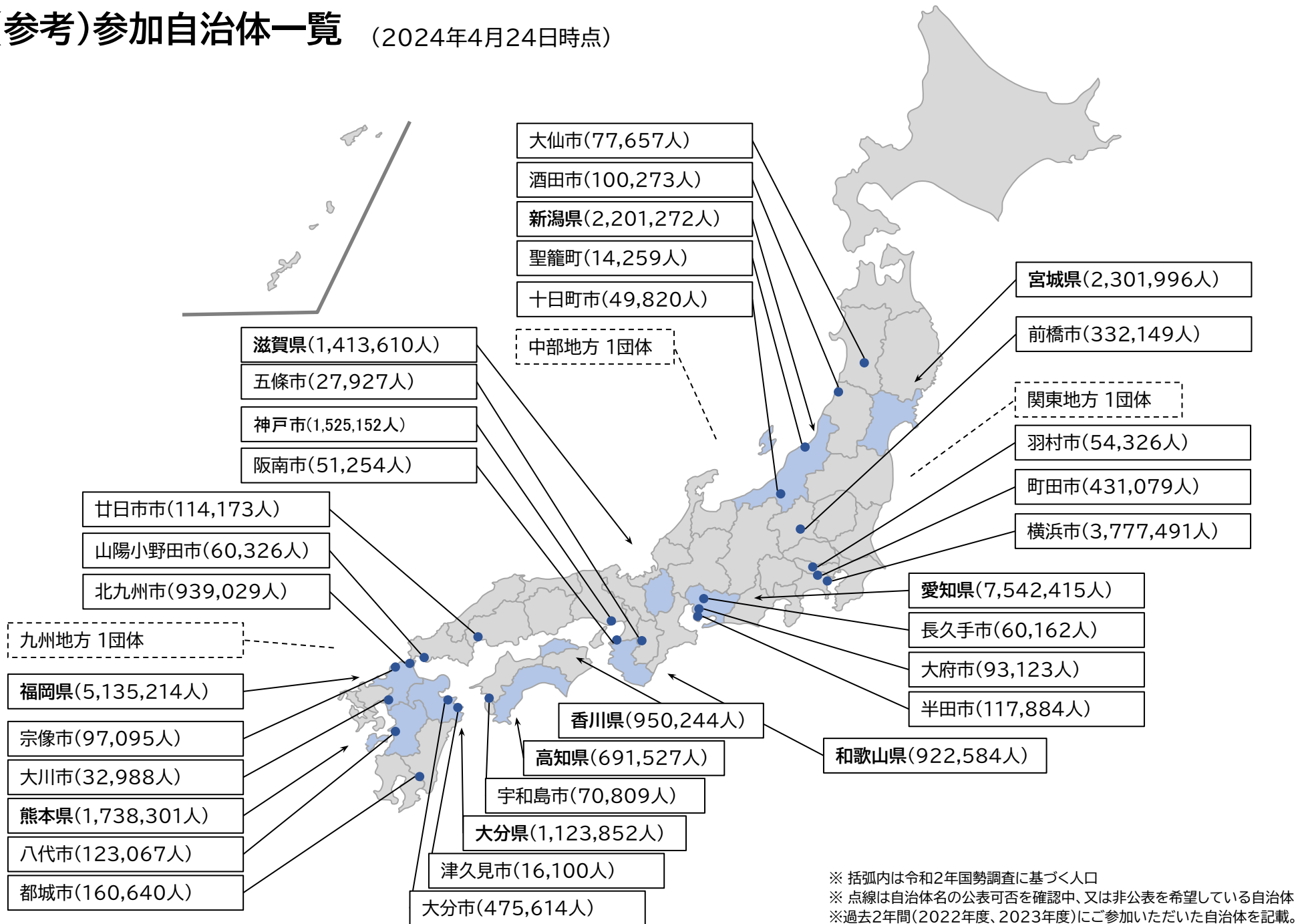
提出期限: 2024年6月12日(水)

※ 応募の期限に間に合わない場合は、期限前にご相談くださいますようお願いいたします。

※ 応募に関するご質問やご不明点、疑問点等がある場合は、遠慮なくお問い合わせください。

※ 応募前に、HPにて公開しております「個人情報保護制度との関係及び具体的匿名化処理の方法について」「自治体税務データ活用プロジェクトにおけるデータ取扱規則」、「自治体税務データ活用プロジェクトにおける安全管理措置等措置等に関する規程」をご一読ください。

(参考)参加自治体一覧 (2024年4月24日時点)



※ 括弧内は令和2年国勢調査に基づく人口
 ※ 点線は自治体名の公表可否を確認中、又は非公表を希望している自治体
 ※ 過去2年間(2022年度、2023年度)にご参加いただいた自治体を記載。

Q&A (1)

全般

- Q1 匿名化に関する知識は必要ですか。
- A1 いいえ、必要ありません。個人情報保護及び匿名化の専門家の監修の下、デジタル社会形成整備法施行後の匿名加工情報制度において求められる水準の匿名加工の案を、CREPEから考え方とともにご提示します。
- Q2 匿名化等のため、プログラミングのスキルは必要ですか。
- A2 いいえ、必要ありません。匿名加工の案について確認いただいた後、CREPEから匿名化ツールを配布します。自治体様には、CREPEが配布するマニュアルに従って匿名化ツールを実行いただきます。
- Q3 データ分析のスキルは必要ですか。
- A3 いいえ、必要ありません。ご提供いただいた匿名化データはCREPEにて分析し、分析結果をお返しします。
- Q4 税務データと世帯データ、普通徴収のデータと特別徴収のデータなど、自治体側でデータ間の紐付けを行うことが必要ですか。
- A4 いいえ、必要ありません。紐付けのキーとなるデータ項目(宛名番号など)が含まれていれば、CREPEにおいて紐付けをします。

自治体側の負担

- Q5 参加にあたって、費用負担(CREPEへの委託料など)は必要ですか。
- A5 いいえ、必要ありません。ただし、システムからデータを抽出する際にシステムベンダに依頼しなければならない場合は費用が発生する可能性がありますので、事前にご確認をお願いいたします。また、データを郵送していただく場合は、郵送費や電磁的記録媒体等を準備していただく必要がございます。
- Q6 自治体側にどれくらいの作業負担が発生しますか。
- A6 2023年度に参加いただいた自治体の作業時間の中央値は、データの概要報告作業は3時間、匿名化作業は4時間でしたが、それぞれ30時間以上を要した自治体もあり、自治体によって大きな差がありました。2024年度はツールの改善や体制強化を行い、自治体様の作業負担軽減に取り組みます。

Q&A (2)

匿名化

Q7 データの紐付け等のために、宛名番号・世帯番号を提供する必要がありますか。

A7 自治体様にて設定いただくパスワードに基づいて、ハッシュ化された宛名番号・世帯番号を提供いたします。

ハッシュ化とは、元のデータから一定の計算手順に従ってハッシュ値と呼ばれる規則性のない値を求め、その値によって元のデータを置き換えることにより、データを不可逆的に別の形に変える方法をいいます。

例としては、以下のようなイメージです。

<宛名番号> → <ハッシュ化した宛名番号>
12345678 aks;ldfjpawnefdoiewadlksfdajf
23456789 k3298refcsna3489elsjodslerire
34567890 kjlz-9erj8ufdjiawe8re;ofjwloe

ここで「aks;ldfjpawnefdoiewadlksfdajf」から、元の「12345678」を復元することは不可能です。そのため、当方が宛名番号を知りえないまま、年度をまたぐ個人の紐付けが可能となります。

ハッシュ化は、CREPEが配布する匿名化ツールに組み込まれているため、自治体様はパスワードを設定するだけでハッシュ化を行うことができます。

Q8 Rのバージョンについて、マニュアル記載のものと異なるバージョンでも問題なく動作しますか。

A8 マニュアル記載のもの(最新版)と同じバージョンをご使用いただきたく存じます。

Q9 貸出PCで匿名化作業を行う場合、自治体負担の費用は発生しますか。また貸出PCはスタンドアロンでの運用が可能ですか。それとも自治体側のネットワークへの接続が必要でしょうか。

A9 貸出PCについて、費用負担はありません。また、貸出PCはインターネットに接続せずに匿名化作業ができるよう、当センターで用意しております。

Q10 匿名化データやツールなど、容量の大きなファイルの送受信はどのような方法で行いますか。

A10 オンライン上のデータ送信サービス、CD-Rの郵送など、各自治体様にとってご都合のよろしい方法での送受信を予定しております。

Q&A (3)

データ内容・形式

Q11 現在、システムベンダごとにデータの構造や形式が異なっていますが、提供するデータの構造や形式に指定はありますか。

A11 ありません。データ構造・データ形式の差異については、CREPE側で対応いたします。
具体的には、プロジェクトへの参加決定後、データ構造やデータ形式の詳細をお伺いし、それに基づいて匿名化ツールを作成し、配布いたします。

Q12 何年分のデータが必要ですか。

A12 少なくとも5年分のデータをご提供いただきます。なお、機械学習を用いた予測を行うため、可能な限り多くの年度分のデータをご提供いただけますと予測精度の向上が期待できます。

Q13 ベンダの変更に伴い、いくつかの変更(①各種マスタやデータベースの構成、データ項目が変更 ②新システム移行後は、データ提供時点のデータベース(現行システムのデータベース)からの抽出が制約される等)が想定されますが、プロジェクトへの応募は問題なく行えますか。

A13 変更後も応募要件を満たす場合は応募可能です。

Q14 収入と所得の変数が両方なくても参加可能でしょうか。

A14 給与、給与(専従)、雑所得(公的年金)以外については、所得金額があれば、収入金額がなくても応募可能です。
またシステムの管理上、給与、給与(専従)が分割されていない場合は、給与のみで問題ありません。

Q15 自治体の全個人のデータベースがなくても参加できますか。

A15 大変申し訳ありませんが、自治体の全個人のデータベースをお持ちでない場合は、正確な分析ができないためご参加いただけません。

Q16 全てのデータベースに性別や生年月日、続柄がない場合でも参加可能でしょうか。

A16 全てのデータベースに年度と宛名番号があれば突合可能ですので、性別や生年月日が全てのデータベースに含まれている必要はありません。しかしながら、自治体の全住民が含まれるデータベースには性別、生年月日、続柄が含まれている必要がございます。

Q&A (4)

PCスペック等

- Q17 匿名化を行うPCのスペックはどの程度必要ですか。
- A17 自治体様の人口規模によって変動がありますが、昨年度ご参加いただいた人口10万程度の自治体様で、
プロセッサ: Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz 1.80 GHz、メモリ: 16.0 GB
のPCをお使いになった場合はデータ処理に大きな問題はなかったようです。
応募様式にご記入いただいたスペックを参考に、メモリ16GBの小型PC(モニター・キーボード・マウスは自治体様にご用意いただきます。)を当センターの判断で貸し出す場合がございます。
- Q18 匿名化を行うPCはインターネットに接続できる必要がありますか。
- A18 インターネットに接続できる必要はありません。
- Q19 匿名化データの提出方法はどのようになりますか。
- A19 ファイル転送システム、記録媒体の郵送など、自治体様のご都合のよろしい方法で結構です。送付方法の詳細については改めてご説明いたします。

- Q20 PCに外部のソフトウェアをインストールすることができません。匿名化にRを用いることは必須ですか。
- A20 匿名化等の作業に要する時間を削減するため、必ずRで匿名化作業をしていただきます。自治体様のPCでRをご使用いただけない場合は、貸出PCをご活用いただくことも可能ですので、ご相談ください。

契約・協定

- Q21 契約や協定は必要ですか。
- A21 必ずしも必要ないと考えていますが、必要があると考える自治体様は、応募様式でご回答ください。

分析

Q22 最大何年間の推計が可能ですか。

A22 原則として1年間の推計が最大となります。これ以上の推計は予測精度の確保が難しいためです。

Q23 今年度以降予定されている税制改正について、改正項目ごとに影響額を分析することは可能でしょうか。

A23 現状の税収予測モデルでは、税制改正の影響を正確に推定することはできません。

Q24 前年度の分析結果やサンプル、提供データを活用した研究は公開されていますか。

A24 前年度の分析結果やサンプルはCREPEのHP上で公表しております。また、提供データを活用した研究は完了次第、CREPEのHPで公表いたします。

Q25 税収予測の精度はどれくらいですか。

A25 提供いただけるデータの数(人口)、年数等によりますが、2023年度はほとんどの自治体で平均誤差率は1.5-3%でした(個人住民税)。2024年度は、更なる精度向上を目指しています。