

Python で学ぶデータによる社会分析”超”入門

日時:2023年3月8日(水)10:30~17:00

場所:オンライン開催

料金:一般 5,000 円、学生 2,500 円

講師:川田恵介(東京大学)

定員:35名

■本コースの内容

本コースでは、これまで量的・数理的実証研究に触れたことがない学生・研究者を念頭にサーベイデータの代表的分析方法を紹介します。このため数式などは用いず、直感的なコンセプトの紹介に特化します。

具体的には研究関心である社会集団(母集団)の特徴を、特定の理論モデルではなく、データの収集方法に依拠し推論する手法を紹介します(OLS, 漸近正規性に基づく信頼区間の近似、など)。データの収集方法は、多くの学術・公的データにおいて明示されており、社会科学における実証分析と相性が良いことが改めて強調されています。

また統計・計量理論・因果推論の発展や機械学習などとの融合が進み、より発展的な分岐を行う上での重要な基礎ともなっています。

さらに実際のデータに対して、Python を用いて分析を実装する方法も解説します。

具体的には、pandas や statmodels、seaborn などの有力パッケージを用いた実習を行います。Python は、R と並び、現代のデータ分析で最も幅広く用いられているオープンソース・無料言語です。有料ソフトである STATA や Eviews, SPSS, SAS などと比べて、(1) ユーザーコミュニティが広い、(2) 最新の手法が迅速に実装されやすい、(3) 幅広い学生・研究者・実務家と協働が容易、といった利点があります。このため教育機関からの需要も拡大しており、教員としてのキャリアを形成する上で重要となっています。また R と比べて、深層学習などについては明確な比較優位があります。すでに R に習熟されている方にとっても、分析の幅を広げる魅力的な”第2の分析言語”の有力候補です。

量的分析・Python とともに、学び始めるハードルが高い、と過剰に考えられがちです。しかしながらしっかりと内容を絞って順序立てて習得すれば、そのハードルは他の手法や言語と比べて決して難しいものではありません。幅広い学生・研究者の参加をお待ちしております!!!

■次のような方におすすめです

- 初めて定量的実証研究を学ぶ方
- キャリアアップのために Python を学びたい方

■注意事項

- Python のクラウド環境である Google colab を用いた、実習を行います。このため事前に Google ID の作成、Colab の動作確認を行なってください。動作確認用の動画を事前に公開いたします。

■本コースの日程

「社会の特徴を推論するとは?」、「データの特徴を記述・可視化」、「母集団の推定」、「推計誤差の評価」、「多重検定問題」、「pandas によるデータ整備」

* 進度によって、内容が若干変わることがあります。