

Rによる因果推論入門: 信頼性改善に向けた quarto, matching, multiple-testing の活用

日時: 2022年9月9日(金) 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

社会・生物医療科学において広く用いられている、統計的推論について、Rでの実装方法とともに紹介する入門コースです。特にコントロール変数法を重点的に解説し、より分析の”信頼性”を担保しやすい実装方法も紹介します。

データ分析へ社会的・学術的関心が高まる中で、その信頼性について厳しい目が向けられています。信頼性向上には、分析のパイプライン(データ収集、識別・推定戦略、事後解釈の是非、コミュニケーション)全体の包括的改善が必須であり、地道な取り組みが必要です(Christensen and Miguel 2018; Christensen, Freese, and Miguel 2019)。個々の研究者においては、“自衛”のためにも、しっかりとしたコンセプトの理解・適用、分析の再現可能性の担保を通じた分析過程の透明性確保が求められています。

本コースでは、(1)因果効果の識別、(2)Matching法を利用したSemiparametricな推定、(3)多重検定の補正、(4)recipesパッケージを利用した包括的なデータ整備、(5)Quartoを利用した分析コード・結果・記述の一体化(Dynamic documentation)を紹介します。これらの手法は多くの受講者の研究に応用可能であり、すぐに活用できる実用的なテクニックです。

■次のような方におすすめです

- ・統計的因果推論に関心のある方
- ・分析の透明性・信頼性を確保するためのRstudioの機能を使いこなしたい方

■注意事項

- ・SSJ データアーカイブのデータを利用するため、大学または公的研究機関所属の研究者・学生(学部生も可)、SSJ データアーカイブへデータを寄託されている民間研究機関の方のみを対象とします。
- ・R・R studio・必要パッケージのインストールを事前に済ませてください。登録・インストール方法を紹介した動画を事前に配布します。

■本コースの日程

「Dynamic Documentation」、「条件付き独立に基づく因果効果の識別」、「回帰分析のみを用いた推定の有効性と限界」、「マッチング法を用いたセミパラメトリック法」、「多重検定問題」、「recipes パッケージを利用したデータ整備」

* 進度によって、内容が若干変わることがあります。