

❁ 医療職のための統計シリーズ

医療職のための学び直し—研究デザインから論文報告までの生物統計学の道標— 第17回 統計解析ソフトRによる図表の作成と統計解析

カワハラ タクヤ サカマキ ケンタロウ
川原 拓也*1 坂巻 顕太郎*2

I はじめに

過去の連載で、図表によるデータの要約（連載第5・6回）や、検定（連載第8回）、回帰モデル（連載第11～13回）について解説がなされてきた。CSV形式やTSV形式などで入力されたデータを用いてこれらの分析を実行するために、コンピュータソフトウェアを利用することが必須である。代表的なソフトウェアとして、Microsoft Excel（Microsoft Corp.）やNumbers（Apple Inc.）などの表計算ソフトや、JMP（SAS Institute Inc.）、SAS（SAS Institute Inc.）、R（R Core Team）、SPSS（IBM Corp.）、Stata（Stata Corp.）などの統計解析ソフトが挙げられる。今回は、図表の作成と統計解析の実行の点で非常に優れたソフトウェアである、Rの使い方について紹介する。

（1）Rを選択する理由

Excelのような表計算ソフトは、データの単純な要約や可視化（グラフ作成）には便利であるが、統計解析には特化しておらず、研究目的を達成するための統計解析が実行できないことが多い。また、グラフ作成においても柔軟性が低いという問題もある。そのため、論文投稿や研究発表を目指すには統計解析ソフトを用いることが望ましい。統計解析ソフトの選択には、適用したい統計解析方法が実行可能であるか、コスト、使いやすさなどが考慮される¹⁾。数ある統計解析ソフトのなかでRを推奨するのは、次のような理由である²⁾。

- ・ Rは無料で、主要なOS（Windows, Mac, Linux）で利用できる。
- ・ Rの分析結果は、プログラム（とデータ）を

共有すれば、誰でも再現できる。

- ・ データ加工（読み込みやハンドリングなど）、可視化、統計モデリング、機械学習などに対する膨大なパッケージ（作業に必要な関数などをまとめたもの）を有する。
- ・ 統計学や機械学習の研究者は、Rプログラムを公表論文の付録やGitHubなどに公開することが多く、最新方法でも実行可能である。

Rは最低限のプログラムを組むことが必須であり、Excel, JMPやSPSSなどのGUI操作（マウスのみでの操作）によるグラフ作成や統計解析に慣れた研究者にとって、Rを使うことに障壁を感じることもあるかもしれない。しかし、基本的な操作を覚えるだけでも実行できることは無限大である。また、論文投稿・研究発表にそのまま使える表（集団の要約など）を作成できたり（2節）、美しいグラフを作成できたり（3節）、Excelでは不可能なことがRだけで実行できる。いくつかの汎用される関数（パッケージ）に関しては、R Commander³⁾を立ち上げれば、マウス操作でも統計解析が実行できるようになる（4節）。これらが実行できるだけでも、Rが研究の助けになるソフトウェアであることがわかる。

（2）Rの基礎

ここでは、Rの初学者に向けて、RStudio（RStudio, PBC）というソフトウェアを用いた統計解析の実行方法を解説する。2節以降では、本節のプログラムを実行することで付随してダウンロードされるテストデータを用いて説明する。プログラムを転記すれば各自で実行可能であるため、手を動かしながらの学習を勧める。なお、初学者が読み進めやすいように、正確な表現でない場合がある（例えば、データフレー

* 1 東京大学医学部附属病院臨床研究推進センター助教

* 2 横浜市立大学大学院データサイエンス研究科特任准教授