

さくっと

テーマ

Rによる 実証分析 (入門編)



「さくっと」とは？

興味のある分野について、さくさくと勉強が進むように作成された調べ方ガイド(パスファインダー)です。みなさんの学習支援を行う図書館学生サポーターが作成しました。ぜひ学習の際に参考にしてください。

図書館学生サポーター 山本

1. はじめに

統計分析ソフトRとは:統計分析ソフトには様々な種類のものがあり、それぞれ利点と欠点がありますが、今回は統計分析ソフトRを紹介します。Rには以下の3つの特徴があります。

- フリーソフトであるため、PCとネット環境さえあれば、誰でもいつでも利用が可能
- 研究者やRユーザーが開発したプログラムを自由に公開ができるため最新の統計分析手法が使用可能
- Rのコマンド解説やユーザー同士の情報交換の場があり使用上のサポートや情報交換が充実している

大きくまとめた以上の利点から多くの研究機関や教育現場で使用が広がっています。今回はこのRのソフトをダウンロードし、ソフトに触れるまでの一連の流れを紹介いたします。

- テーマに関するキーワード
統計分析, 計量経済学, 実証分析

2. Rのダウンロード

PCから「<http://www.r-project.org>」にアクセスし、ページ上部の「download R」をクリックします。すると、ファイルをインストールするサーバーの選択欄に移動するので、最寄りの統計数理研究所(Institute of Statistical Mathematics)を選択します。サーバーを選択し終わったら、PCのOSにあったファイルをダウンロードします。後は、言語選択、インストール先、コンポーネントの選択に移ります。コンポーネントは使用しているPCのOSに合わせてください。OSが32bit版であれば64-bit Fileのチェックは外してください。オプション選択を終えるとダウンロード完了です。

3. 基本を理解しよう

基本はどのプログラムも同様にコンソール画面に数字やコマンドを入力し実行するという流れになっています。ただしR特有の所作として、数式や値を変数に代入する記号として「<-」を利用します。例えば、「1+1」を変数「A」に格納する場合「A <- 1+1」となることだけは注意してください。加えて一点、単純な四則計算や初歩的な回帰分析であればRに初めから実装されているコマンドで事足りますが、最新の統計分析手法を活用したい場合は、一度、配布される「パッケージ」からダウンロードする必要があります。方法は、メニューバーの「パッケージ」から「パッケージのインストール...」を選択し、最寄りのサーバーを選択して必要なパッケージをダウンロードします。そして、同じくメニューバーの「パッケージ」から「パッケージの読み込み...」を選択し、パッケージを読み込みます。あと一つほどコマンドで読み込み方法がありますが、今回紹介したやり方が一番簡単なため、今回はこの方法を記載しておきました。またOSがWindowsかMac OSかで表示方法が異なる部分もあるので注意してください。その他詳細は書籍で調べてください。

4. 理解を深めよう

ダウンロードが完了したら、ソフトを動作させましょう。今回は、統計の基礎知識を実践していきます。まずは、R用の作業フォルダを用意します。PC上に新たなフォルダを作成してください。続いてRを起動して、図1のとおりメニューバーの「ディレクトリの変更...」から作成した作業フォルダを選択すれば完了です。この作業を使用しなければ、Rが調べたいデータにたどり着かず解析できないということになりますので注意してください。

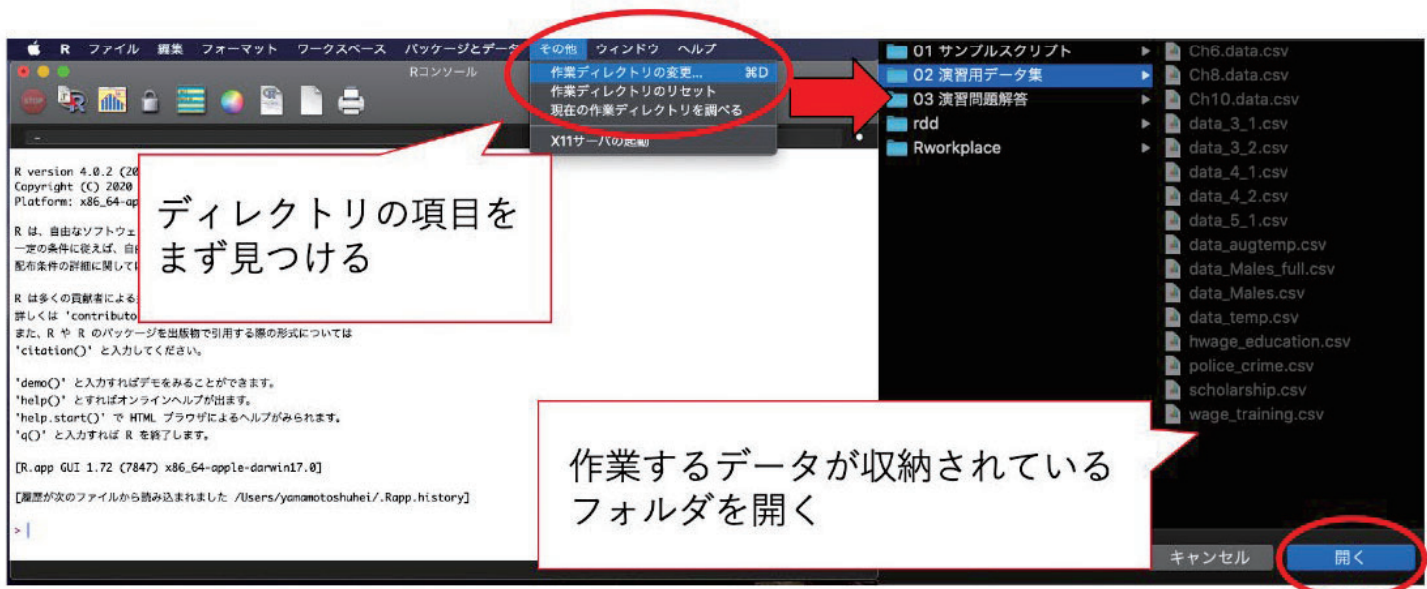


図1 R上での作業ディレクトリー変更の方法

5. Rを用いた実証分析

前章まででRを使用するまでの準備ができたかと思います。この章では実際のRを使った際の結果の見方について述べていきます。詳細は「さくっと:計量経済学」にも記載されていますので用語等を確認されたい方はそちらで勉強して頂けたら幸いです。使用するデータはRに元から組み込まれている「モルモットの歯の成長がビタミンCの摂取方法と量によってどのように異なるか」を調べる目的で作成された実験データです。Rで推定した結果が図2になります。

```
Call:
lm(formula = len ~ dose, subset = (supp == "VC"))

Residuals:
    Min       1Q   Median       3Q      Max
-8.2264 -2.6029  0.0814  2.2288  7.4893

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.295      1.427   2.309  0.0285 *
dose          11.716      1.079  10.860 1.51e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.685 on 28 degrees of freedom
Multiple R-squared:  0.8082, Adjusted R-squared:  0.8013
F-statistic: 117.9 on 1 and 28 DF, p-value: 1.509e-11
```

図2. アスコルビン酸を直接モルモットに与えた場合の歯の成長度合いを回帰分析したもの

推定結果は赤く囲んだ部分に示しています。「Estimate」が推定結果を表しています。「Std. Error」が標準誤差を示し、「t value」がt値、「Pr(>|t|)」はp値を表しています。このp値が小さいほど推定結果に有意性が持てると示されています。また視覚的にも「*」の存在からも有意性の確認ができます。「*」の数が多いほど有意性が高いという認識で良いです。つまり、doseの列では、Pr(>|t|)の値が小さいかつ「*」の数も3つほどあるのでアスコルビン酸を用いたモルモットの歯の成長要因を示す推定結果は有意であることがこの結果より明らかになるということが読み取れます。ただし、上記のような単純な解析ならまだしも、複数の要因が絡む事象では、解析結果では有意性があつたとしても、現実的ではない答えも生じてきます。その点に細心の注意を払うことは意識してください。

6. 論文作成のために

レポートの書き方・学びかた

- 星野匡朗・田中久稔著, Rによる実証分析—回帰分析から因果分析へ—, オーム社(2016)
- 山本勲著, 実証分析のための計量経済学, 中央経済社(2015)

上記の書籍を一読すると論文の書き方、論文作成のために必要な知識を習得できるかと思えます。また「さくっと:計量経済学」にも簡単ではありますが、論文を書く上での手順も紹介しております。合わせてご確認ください。

