

文系学生を対象としたデータ分析支援ツールとしての Visual R Platform の活用

帝京大学経済学部経営学科
横山 暁

1 はじめに

本発表では、2014年6月2日にリリースされた Visual R Platform（以下 VRP）に関して、文系学生を対象とした授業やゼミ等でデータ分析をする際の分析支援ツールとしての利用可能性について論じていく。特に、よく利用される Excel や、VRP の動作のベースとなる R（CUI ベースや R Commander）の利用との比較を行うことで、VRP の利用がどういう場面で効果的であるかを探っていく。

まず 2 節では、現状の授業や設備等の環境面について簡単に紹介する。3 節では実際に R を授業等で利用するにあたり、データのインポートやクロス集計等、幾つかの場面を想定して、メリットやデメリットを考察する。4 節では 3 節で触れた実際の場面に対して VRP を利用することのメリットについて論じ、最後にまとめや今後の発展について論じる。

2 授業と設備

2.1 授業について

大学の文系学部（特に横山の所属する帝京大学経済学部経営学科や、非常勤講師をしている／していた文系学部）における統計学に関する科目やコンピューター操作に関する科目の現状について紹介する。

統計学に関する科目では、統計学の基礎となる記述統計や推測統計に関する授業が設置されていることが多く、また社会調査士資格を取得可能な学部・学科においては「量的データ解析の方法に関する科目」や「質的な解析の方法に関する科目」の授業が設置されている。

コンピューターの操作に関する科目としては、情報リテラシー教育（Word や Excel, PowerPoint 等）の科目に加え、各学部・学科の必要に応じてデータ処理に関する科目等が設置されていることが多いと考えられる。また、いわゆる「ゼミ」においてデータ分析を扱う場合は、適宜分析ソフトウェアを利用することが行われていると考えられる。

帝京大学経済学部経営学科（以下、本学科）においては、

- ・統計に関する科目 「統計学」
- ・データ処理に関する科目 「データ処理演習」（主に Excel を利用）、「データ解析」（SPSS を利用）

が設置されている*1。

データ処理の授業では、Excel を利用することが一般的であると考えられ（本学科では「データ処理演習」が相当する）、基本的な関数やデータの並べ替え、フィルター、ピボットテーブル、分析ツール等を用いたデータ分析の方法を扱い、授業によっては VBA マクロ等も扱うと考えられる。

さらに、その他の分析ソフトウェアを用いた授業（本学科では「データ解析」に相当）では、R や SPSS, SAS を利用し、データの集計や検定、多変量解析を学習することが一般的であると考えられる。

ただし、他に設置される授業の選択肢が多い場合、これらの授業を履修する学生は、さほど多くないという現状も見て取れる*2。

*1 このうち横山が担当しているのは「データ処理演習」である。

*2 「情報リテラシー教育のための科目」は除く。必修になっていることもある。

2.2 設備について

有償の分析用ソフトウェアはライセンスの問題があるため、大学内の（一部の）コンピューター教室での利用に限られることが多く、自由に使うことが出来ないというのが現状であると考えられる。つまり、分析用のソフトウェアを利用する授業で課題を出す場合、自宅学習が困難であるために授業内で終了しう程度の課題しか出すことが出来ないことが想像できる。故に、データ処理に関する科目では、Excel のないしフリーソフトウェアを利用することが無難となる。

ここでフリーソフトウェアとして候補に挙がるのが R であろう。R はオープンソースソフトウェアの統計解析向けのプログラム言語・開発実行環境であり、必要に応じてフリーのパッケージをインストールして用いることも出来る。全てのことがフリー（無料）で出来、Excel では行うことが出来ない高度な分析も可能であるため、授業等では利用しやすい分析用のソフトウェアであり、実際に R を利用した授業も行われ始めている*3。

3 R の授業等での利用について

3.1 インストールについて

R を授業等で利用する場合、大学での利用環境を整える他に、自宅学習のために各学生が所持する PC に R をインストールしてもらう必要があると考えられる。幸いにして、本やインターネット上の情報が充実しているので、R のインストール作業や必要なパッケージのインストールで苦勞するケースは少ないと考えられる。

3.2 R Commander の利用について

R を実際に授業で利用するにあたり、最も学生が躓く点としては R が CUI 環境である点と考えられる。Excel やその他の分析用のソフトウェアは GUI 環境で操作するものであり、データを手入力する場合を除けば、データの読み込みから集計・分析、グラフ作成までほぼ全ての作業をマウス操作で行うことが出来る。しかし、R の場合は CUI 環境であるためコマンドの入力が必要である。この場合、コマンド入力を極力減らすために、R を GUI 環境で利用可能とする「R Commander」というパッケージを利用する方法が考えられる。R Commander の画面では、マウス操作で各種集計・分析が可能であり、出力結果も R Commander の画面上に表示されるため、基本的にこの画面内で処理を完結させることが出来る。

3.3 実際の利用について

本節では、授業等でのデータの集計・分析の具体的な場面における R の利用を考える。

3.3.1 データの入力・読み込み

データを手入力する場合、Excel であれば直接ワークシートに入力することが出来るが、R の場合は多少困難が伴う。R Commander であれば、「データ」→「新しいデータセット」を行い、出てきた「Data Editor」画面で手入力することは可能ではある。

CSV ファイル等のテキスト系のファイルから読み込む場合、`read.table` や `read.csv` コマンドを使うことになる。R Commander では「データ」→「データのインポート」→「テキストファイルまたはクリップボード、URL から」を選択することでファイルから読み込むことが出来る。なお、SPSS 等の分析用のソフトウェア

*3 本学では少なくとも文学部心理学科において R を利用した授業が行われている。

のファイルも同様に読み込むことが出来る*4.

3.3.2 各種集計

データの各変数や変数間において、基本統計量の計算やクロス集計等を行うことがある。Excel の場合は、各種関数や分析ツール、ピボットテーブルを使うことになる。R の場合も各種関数のコマンドを入力することで実現可能であるが、文系学生が対象であることを考えるとコマンドの入力は現実的ではない。R Commander の場合は、「統計量」から各種集計等を行うことが出来、非常に簡単である。

しかし R Commander を利用する場合、集計結果を書き出す際に困難を伴う。集計結果をグラフ等にまとめる場合、Excel を利用するケースが多いと考えられ、R で集計した結果を直接もしくは間接的に Excel に読み込む（コピーアンドペーストする）が必要になる。R Commander の画面では、各種集計を行った結果は「出力」部分が表示される。この結果は「ファイル」→「出力を保存する」等を行うことで結果を書き出し、テキストエディタ等で開いた上で対象範囲をコピーし Excel に貼り付けることになる。しかし、貼り付けたい部分は表形式である場合でも、Excel に貼り付ける際にはそのまま貼り付けることが出来ず、「テキスト ファイル ウィザードを使用」する必要がある、手間がかかってしまうという点がデメリットであると考えられる。

3.3.3 各種分析

各種分析についても集計と同様である。R Commander を利用することで、回帰分析や主成分分析、因子分析、クラスター分析等、基本的な分析を簡単に行うことが出来る。しかし結果を出力し、Excel 等で利用する際に手間がかかってしまう点がデメリットである。

4 VRP の利用について

3.3.2 節や 3.3.3 節で言及したデータ集計・分析の各場面のそれぞれにおいて、VRP の利用のメリットについて考察する。

4.1 データの読み込み

VRP では R のデータ形式である.RData を基本データ型としており、保存してある.RData ファイルを VRP のプロジェクトウィンドウ上へドラッグ&ドロップするだけでデータ（正確には保存してあるオブジェクト）を読み込むことが出来る。また、CSV や Excel ファイルを VRP に読み込む際にも、プロジェクトウィンドウ上へドラッグ&ドロップするだけで良い。CSV ファイルの場合は、メニューが表示され、「カンマ区切り表形式データ」か「.RData に変換」かを選択することが出来る。

基本的にドラッグ&ドロップでデータの読み込みが出来るという点において、VRP を利用するメリットは大きいと考えられる。

4.2 各種集計や分析

VRP で各種集計や分析を行う場合、VRP の画面の左側にある「Object Browser」に存在しない集計や分析を行うことは困難である。しかし、VRP は開発が継続されており、ユーザーの声によって利用可能になる分析や集計は増えていくと考えられる。

Object Browser に存在しない集計や分析でも、Object Browser に存在する処理を組み合わせたり、「R スクリプト」を用いることで実行可能とすることが出来る。例えばクロス集計は VRP のバージョン 1.1 では実

*4 Excel ファイルの読み込みに関しては、read.xlsx 等のコマンドがあるが、一旦 CSV ファイルにしてから読み込んだほうが無難である。

装されていないが、「データハンドリング」と「マトリクス化」を組み合わせることで実行可能である。R スクリプトは、R のコマンドを実行するための処理であり、Object Browser にドラッグ&ドロップすることで現在実行しているプロジェクト登録することが出来る。授業等での利用の場面を考えると、授業で扱う集計や分析を R スクリプトとして教員が準備し、Object Browser に登録したプロジェクトを保存して学生に配布するということが考えられる。これにより、学生はコマンドを入力することなく分析を行うことが可能になる。

4.3 結果の出力

各集計や分析の結果は、アイコンを右クリックして「データビュー」を選ぶと、データビューアーで結果を見ることが出来る。このデータビューで表示された表は、値を選択してコピーをすることが可能であり、結果をコピーし、Excel 等に直接貼り付けるということを非常に簡単に行うことが出来る。3.3.2 節や 3.3.3 節で述べたように、R Commander の利用では結果を Excel 等に貼り付けるためには手間がかかってしまうことを考えると、VRP の利用は有効であるといえる。

4.4 R ソースのエクスポート

VRP には「R ソースのエクスポート」という機能がある。これは VRP のプロジェクトフローを R のスクリプトファイルとして出力することが出来る機能であり、R がインストールされてる環境であれば出力したプロジェクトを単独で実行することが出来る。

エクスポートの手順は、エクスポートしたい結果を含むアイコンを右クリックして、「R ソースエクスポート」を選択するだけである。授業等で VRP を利用する場合、VRP 環境で実行した結果を授業終了時にエクスポートし、学生は R がインストールされた PC でエクスポートしたスクリプトを読み込んで実行することで、VRP がインストールされていない環境でも授業内容を再現することが出来、自宅学習が可能になる*⁵というメリットがある。

ただし、Excel ファイルを読み込む処理がフローに入っている場合、R ソースエクスポートが出来ない。これは Excel ファイルを読み込む処理が VAP の処理であり、VAP の処理が含まれていると R ソースエクスポートが出来ないためである。R ソースエクスポートを行うことが前提で Excel ファイルを読み込む場合は、一旦 CSV ファイルにしておき、R スクリプトで `read.table` (もしくは `read.csv`) コマンドを利用すると良いようである。

5 まとめ

本稿では、文系学生を対象とした授業等でのデータ分析の際の分析支援ツールとしての VRP の利用可能性について、実際に想定されるいくつかのデータ分析の場面について R および R Commander の利用と比較をしつつ VRP の利用のメリットについて論じてきた。R や R Commander だけの利用では困難もしくは手間がかかる処理でも VRP を利用することで簡単に行うことが出来、また同時に VRP がインストールされていない環境でもある程度の実習が可能であるという点で VRP は授業支援ツールとして非常に有効であると考えられる。

また、4.2 節でものべたように、VRP は開発が継続されており、今後、新たに利用可能となる分析が増えていくと考えられる。これに伴い、授業での利用可能性も増していくと考えられ、今後の発展が楽しみである。

発表では、本稿で述べたいくつかの場面について、実際に VRP を操作しながら授業等での活用について説明していきたい。

*⁵ R や R Commander を利用してさらに集計や分析を進めることが出来る。ただし進めた分析を再び VRP で利用するには工夫が必要である。