

心理統計法教育におけるExcelとRの活用の有効性

塗師 斌

On the Effect of Applying “Excel” And “R” to the Instruction of Statistical Method of Psychology

Akira Nushi

はじめに

筆者はこれまで約30年間、文系学生（主として心理専攻）を対象とした心理統計やデータ解析に関わる授業を行ってきた。その間の目まぐるしい高度情報化の進展は、これらの授業のありようを、隔世の感を与えるほど大きく変化させてきている。今から約40年前、筆者は大学院生として心理統計データ処理を行うために、東京大学大型計算機センターの受付窓口に、FORTRANで作成したプログラムと、データを穿孔したパンチカードの束（2000枚入りの箱を複数個ということもあった）を置いて、数日後にセンターの大型計算機で計算したアウトプットを受け取っていた。「急行」と「標準」のジョブクラスがあって、「標準」では4～5日を要した。エラーをするとその分伸びてしまう。論文の締切りを間近に控えていると、それこそ神にもすがる思いでカードの束を置いたものであった。それが今ではどうであろうか。同じことが、いやそれよりもはるかに多量の処理が、自分の家庭やオフィス、あるいは研究室のパソコンの統計ソフトの画面をクリックするだけで瞬時の内にできてしまうのである。処理の量や速度も当時とは比較にならない。パンチカードも見かけなくなった。当時1ヶ月以上要したことが、今なら瞬時にあるいは数秒間でできる。しかも自分が勝手に使えるパソコンによってである。筆者の感覚としては、かつてのスーパーコンピュータを机上において独占して使っている感じである。

かつては統計処理を行うためには自分でFORTRANやCOBOL、BASIC等のプログラム言語を用いて統計プログラムを作るか、既成のプログラムを利用するしかなかったが、現在ではExcel等の表計算ソフトやSPSS、SAS等の統計パッケージあるいは近年注目されている統計ソフトR等を利用すれば、パソコン画面のメニューのクリックや簡単なコマンドの入力だけで統計処理を行うことができるようになった。筆者の実感としてその最も大きな転換点は何といても1995年のWindows95の登場である。これによりパソコン画面のメニューをクリックしていくだけで結果が得られるGUI（グラフィカルユーザーインターフェイス）が一般的な処理形態となり、またデータの入出力や変換等のファイル処理が容易にできるようになった。それと同時にWindows95をOSとして作成されたアプリケーションソフトであるExcel等の表計算ソフトやSPSS等の統計パッケージが非常に使いやすくなったのである。

本稿では近年筆者が心理統計やデータ解析の授業で活用してきている表計算ソフトExcelと、フリーソフトでありオープンソースであるという点で今後活用が期待される統計ソフトRの、文系学生を対象とした心理統計法教育における有効性について述べたい。

なお本稿では心理統計法教育において統計ソフトやパソコンを利用することを前提としている。これらを利用しない心理統計法教育も当然考えられる。たとえば約40年前に筆者が大学で受けた教育がそうである。また現代でも数式の板書を中心とした講義形式の授業はかなりあると思うし、理系の統計学の授業では数理統計学のベースで進めるのがむしろ一般的だと思う。しかし筆者は、文系学生を対象とした心理統計法教育においては、以下に述べるような理由で統計ソフトの利用が不可欠であると考えている。

1. 文系学生を対象とした心理統計法教育が困難な理由と解決策

文系学生を対象とした心理統計法教育が如何に困難であるかは、筆者自身が約30年間にわたる授業経験の中でずっと感じてきたことである。同じ立場の他の多くの教員からもそのことを聞かされてきた。最近の学会のシンポジウムでもそのことが取り上げられ論議されている（山田・村井・杉澤ら, 2008；山田・村井・杉澤ら, 2007）。おそらく心理統計法教育の困難さを感じていない教員は一人もいないのではないかと思われる。何故困難であるのかということに関して筆者は以下のような理由が挙げられると思う。

1. 心理学を学んでいく上で心理統計法の学習の必要性や重要性についての認識が十分でない。
2. 論文やレポートのデータ処理のため心理統計法の学習の必要性は認識しているが、使いやすい統計ソフトを用いてパソコン画面をクリックしていけば簡単に結果を出せるので、心理統計法の理論体系まで理解する必要はないととらえている。
3. 文系学生には数学に対する苦手意識を持つ者が多く、心理統計法を数学と同じとみなして、その苦手意識を心理統計法に対しても適用している。
4. 科目内容が系統的、積み上げ的で、授業内容がステップバイステップとなるため、不注意や理解不足、復習不足、欠席等で途中のステップを踏み外すと、他の多くの文系の科目と違って、その後の授業内容の理解が難しくなる。
5. 文系学生の多くは授業内容に発想の豊かさ（発散性）を望むのに対して、心理統計法の授業内容は解が一つに決まるという意味で収束的（一意的）である。
6. 科目内容が、人間や社会の諸問題を直接的・具体的に扱う他の文系科目と違って、心理統計法の授業内容は非現実的・手段的で抽象度が高いため興味をもちにくい。
7. 科目の授業内容が一方的に与えられるだけで、学習者自身が授業内容を具体的なイメージと関連づけて心理統計学の世界をダイナミックに動かすという学習経験をもてない。

以上の理由はもちろん相互にある程度関連しているわけであるが、この中で理由の1と2については授業者による心理統計法の必要性や重要性を認識させる説得的な指導によるところが大きい。また理由の3については心理統計法が数学とは異なるという認識を持たすとともに、心理統計法に対して学習者が効力感を持てるような指導をすることにより、苦手意識を除去することが必要である。そして理由の4から7までについては、その克服のためにExcelやRの利用が大いに有効であると筆者は考えている。これらのソフトを利用した演習をベースとした授業を行うことにより、ステップをすべての学習者がより登っていきやすいスモールステップにしたり（理由4）、複雑で多量な計算でも驚異的な速度で一瞬のうち解が一意的に決まるすばらしさや見事さを味わわせたり（理由5）、現実ではとても不可能な大規模なシミュレーション実験を行ったり（理由6）、心理統計法の各種の概念や定理等を具体的なイメージと関連づければ条件を変えて動かしてみる（理由7）ことが可能になるのである。こうしたExcelやRの利用の有効性を活かすことができれば、理由の3で述べた苦手意識の解消につながり、さらには理由の1と2で述べた必要性や重要性の認識にも役立つと思われる。

2. 心理統計法教育におけるExcelの活用

Excelは多くのパソコンに標準装備されているので、費用がかからずどこでも利用しやすい。本格的な統計ソフトではないが、GUIによる表計算ソフトとしての操作しやすさや、表やグラフによる結果の視覚的表現、乱数の利用によるシミュレーション実験等は心理統計法教育においてきわめて有効なものである。しかし各種の多変量解析や複雑な分散分析法等を利用したい場合には、市販されているExcel利用の統計ソフトはあるが、Excelの分析ツールでは不可能である。そのような場合、SPSSやSAS等の統計パッケージが一般に用いられるが、筆者の場合にはSPSSを利用してきた。ちなみにExcelとSPSSを連携して利用することは非常に効果的である。たとえばSPSSの出力結果をExcelのシートにコピーして、

その意味をExcelの数式表現で説明するとSPSSの出力結果の理解を非常に深めることができる。

2.1 心理統計法教育においてExcelの利用が有効な理由

文系の学生を対象とした心理統計法教育において、コマンド入力方式の統計ソフトの指導は非常に難しく、極力GUI方式を用いるべきであるというのが筆者の考えである。その観点では少なくとも心理統計法教育の入門期においてはRではなく、利用しやすさ、操作しやすさ、表やグラフによる視覚的表現のしやすさという点で、Excelを利用すべきであるというのが筆者の立場である。なおExcelの関数の問題点の指摘もいくつかあるが、それらは実際のデータにはまず見られない極端なケースにのみ指摘されることである。箱ひげ図や幹葉図等一部描けない図表はあるにしても、記述統計学の各種の図表の作成、変数の代表値・散布度等の基礎統計量の算出、各個人のz得点・偏差値等の計算、変数間の散布図の作成や相関係数の計算、正規分布表のグラフや表の作成、乱数を利用した母集団からの標本抽出、標本平均等の標本統計量の標本分布等の教授・学習にExcelは極めて有効である。

具体的には特に以下に挙げる点が有効である。

1. 度数分布の作成を関数frequencyを用いて行うことにより、作成した後でもデータ区間のデータを変えればそれに対応して度数分布は変わる。たとえば、乱数を用いて標本平均の標本分布の性質を調べる実験において、frequencyのこの機能は非常に有効である。後述するF9キーを利用することにより、一連の実験の手順の結果の度数分布をワンタッチで反復して行うことができる。なお「分析ツール」の「ヒストグラム」は度数分布を作成するツールであるが、この機能はない。
2. 記述統計学で最もつまづきやすい箇所は、「標準偏差」の概念の理解であるが、Excelでは同じシートにばらつきの非常に大きいものからまったくばらつきのないものまで、複数のデータ例を任意に作成して、それぞれの標準偏差を求めて比較することができるので直感的な理解がしやすい。
3. オートフィル機能により、ある変数や個人に行った処理を他のすべての変数や個人に瞬時的に繰り返し行うことができる。また乱数を用いた実験を行う場合に、オートフィル機能により任意の数の乱数を瞬時的に発生させることができる。
4. 一様乱数や正規乱数等の乱数を発生させる関数ができるので、これを利用して推測統計学の基本ともいえる母集団と標本との関係、標本平均等の標本統計量の標本分布の性質等をシミュレーションによって調べ明らかにすることができる。
5. F9キーの使用により、たとえば1000個の乱数を発生させてその度数分布を求めヒストグラムを描くといった一連の手順を一つの単位として瞬時的に繰り返し行うことができる。

以上、心理統計法教育の入門期におけるExcel利用の有効性についていろいろ述べてきたが、それらに共通して言えることは、Excelの利用により、壮大な規模の複雑な統計計算を、人力ではとても考えられない超高速で瞬時的に実行できるというダイナミックな感覚を味わうことができるということである。それと同時に心理統計学の学習を受け身ではなく、学習者自らが主体的に心理統計学の世界をダイナミックに動かすという学習経験を持つことができる。これは前節の心理統計法教育が困難な理由の7の解決策でもある。もともと統計学はstatisticsという英語表現が表しているように、状態を記述するという静的な意味合いが強い。しかしExcelを使うと心理統計学の世界をダイナミック（動的）に動かすことができる。それが入門期の学習者の心理統計法への学習意欲や興味を喚起すると思われるのである。

2.2 心理統計法教育におけるExcelの利用例

2.2.1 スロットマシン方式

塗師（2002）は推測統計学の基本をなす最も重要な考え方ともいえる大数の法則と中心極限定理の教授法として、Excelを活用した「スロットマシン方式」という教授法を提案している。この方法は母集団からの無作為標本抽出を、スロットマシンでハンドルを引き3つの窓のボタンを順次押していく操作と見

立てるのである。但し窓の数は一般に3つよりもはるかに多いN個である。このN個が標本の大きさに相当する。たとえば内閣支持率調査でいえば、新聞やテレビ等では通常2千人あるいは千人規模（ここでは一般にN人として話を進める）の無作為標本を対象とした調査が行われるが、スロットマシン方式ではこれをハンドルを引いてN個の窓のボタンを押すものとみなす。それぞれの窓に出現するのは1（支持の場合）か0（非支持の場合）であり、N個の窓の裏側にある仕組み（仕掛け）は同一で（同一母集団）、1の出る確率がわれわれの知りたい未知の値であるp、0の出る確率が1-pのベルヌイ分布となっている。イメージとしては図1のように1と0がp対1-pの割合で画かれている円板の側面が各窓の裏側にあつて、N個のボタンを一つ一つ押すことにより1か0のいずれかが各窓に出現する装置を考えればよい。内閣支持率pの推定値・としてN人の中で内閣を支持した人の比率が用いられるが、これはN個の窓の中で1が出た窓の比率に相当する。各窓は1か0しかとらないわけであるから、・はN個の窓の値の総和を求めて窓の総数Nで割った値つまり標本平均でもある。具体的なイメージとしては図2のように、

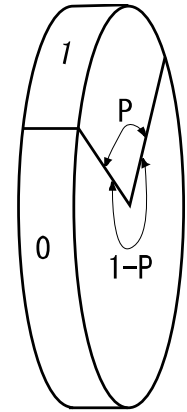


図1 各窓の裏側にあるベルヌイ分布

N個の各窓の裏側にある母集団分布は1の生起確率がp、0の出る確率が1-pの同一のベルヌイ分布で、それぞれの確率変数を X_1, X_2, \dots, X_N で表すと、図2の下に書いてあるような標本平均・を求める公式で表現されることになる（塗師、2002より引用）。ここで各確率変数は1か0しかとらないので、標本平均・は標本比率・と同じである。

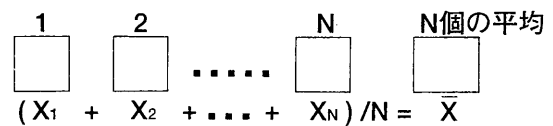


図2 スロットマシン方式と平均の対応関係

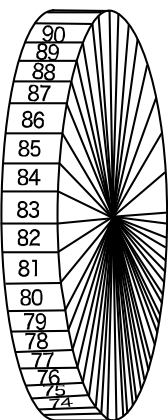


図3 各窓の裏側にある学力テストの分布

なお上記の内閣支持率調査は離散型変数の例であるが、塗師（2002）でも取り上げているように、たとえば学力テストの得点といった連続型変数の場合でも離散型変数と同様に考えることが可能である。その場合には、その学力テストの母集団分布の確率に対応した出やすさあるいは出にくさをもった各得点が円板の側面に書かれた装置が各窓の裏側にあると想定すればよい。簡単のために整数値のみをとる100点満点の学力テストの場合、図3のような装置（塗師、2002より引用）を想定すればよい。

以上のようなスロットマシン方式を用いる利点は、後述するExcel上でのシミュレーション実験により、以下のような推測統計学の基本的な考え方を直感的にわかりやすく示すことができるということである。

1. 図2の式を記述統計レベルの単にN個のデータの和ではなく、N個の確率変数の和をNで除したものであるという推測統計レベルの捉え方ができる。
2. 標本平均（あるいは標本比率）がN個の確率変数の値によって変動することから標本平均（あるいは標本比率）の標本分布という概念が理解しやすくなる。
3. 標本比率・が我々の知りたい値である母集団比率pの推定量として用いることが適切であることを示すことができる。
4. 標本平均（あるいは標本比率）の標本分布のばらつきはNが大きくなるほど小さくなるという大数の法則を示すことができる。
5. 標本平均（あるいは標本比率）の標本分布の形状は、母集団分布の分布がどのようなものであっても正規分布に近づいていくという中心極限定理を示すことができる。

2.2.2 スロットマシン方式へのExcelの活用

ここでは未知の値である内閣支持率 p が母集団では0.4(40%)であると想定して、 $N=25$ の無作為標本の標本比率・の標本分布の平均と標準偏差、および $N=100$ の無作為標本の標本比率・の標本分布の平均と標準偏差を求め、その間の比較も行う。理論的には標本の大きさの大きい $N=100$ の方が母集団比率に近い値が得られ、標本分布の標準偏差は $N=25$ の場合の2分の1に近くなることが期待される。繰り返し回数は多ければ多いほど一般に母集団比率 p に近い値が得られるが、ここでは300回とする。

手順1 Excelのシートの1行目に $N=25$ の場合、変数名X1からX25を、A列(セルA1)からY列(セルY1)までに入力する。Z列(セルZ1)には標本比率を表す変数名としてPHATを入力する。スロットマシン方式との対比で言えば、2行目以下の各行は、図2のように、A列からY列までは母集団の確率変数の実現値が出現する窓、Z列はその左側の25個の実現値の標本比率が計算されて出現する窓と考えればよい。

手順2 シートの2行目のA列(セルA2)に=IF(RAND()<=0.4,1,0)と入力し、この式をオートフィルによってセルB2からY2までコピーする。ここでRAND()は0以上1未満の一樣乱数を発生させる関数で、その値が0.4以下の場合には1、それ以外の場合には0とすることにより、母集団の内閣支持率が0.4であることを表現している。またRAND()は他のセルにコピーした場合、そこで新たに別の独立な一樣乱数を発生させるので、多くの乱数を使用した実験を行ったり、繰り返し実験を行うのに極めて便利である。

手順3 シートの2行目のZ列(セルZ2)に=AVERAGE(A2:Y2)と入力してその左側の25個のセルの平均を求める。この値は標本比率であると同時に母集団比率 p の推定値である。

手順4 シートの2行目のA列からZ列までの範囲をオートフィルによって3行目から301行目までの299行にわたってコピーする。この手順はスロットマシン方式との対比でいえば、ハンドルを引き25個の窓のボタンを順次押すことを300回繰り返すことに相当する。これにより、セルZ2からセルZ301までに300個の標本比率が得られることになる。なお我々の得る調査結果は、この300回中のどれか1回で、得られた標本比率が推定しようとしている母集団比率に近いかどうかは「神のみぞ知る」である。

手順5 シートの適当なセルに、手順4で求めた300個の標本比率の平均を求め、母集団比率である0.4にどの程度近いかを調べる。また300個の標本比率の標準偏差が理論値である $\text{SQRT}(P(1-P)/N)$ とどの程度近いかを調べ、さらに300個の標本比率の分布のグラフを関数frequencyを用いて描き、それが正規分布に近似していることを確認する(中心極限定理)。

表1 N=25とN=100の標本分布の平均と標準偏差

	N=25の場合	N=100の場合
平均	0.397	0.398
標準偏差	0.094	0.046

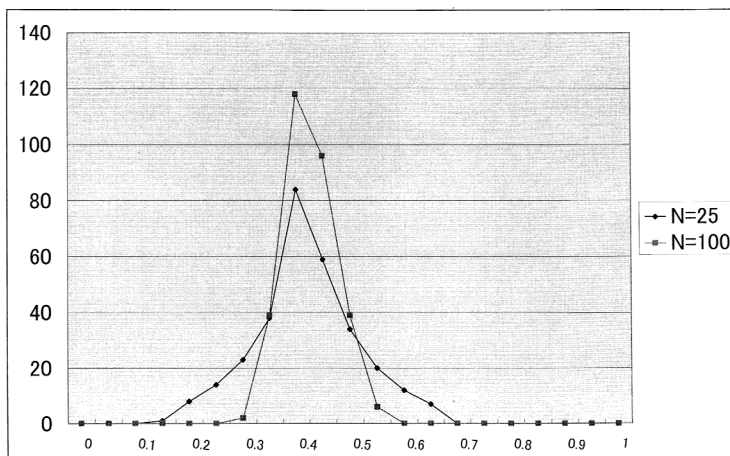


図4 N=25とN=100の標本比率の標本分布

手順6 別のシートで、今度は $N=100$ として手順1から5までの手順を繰り返す。

手順7 $N=25$ の場合と $N=100$ の場合の標本比率の分布を平均、標準偏差、分布形に関して比較する。理論的に $N=100$

の場合の標準偏差は $N=25$ の場合の半分になるはずであるから(大数の法則)、そのような傾向が見られるか確認する。実験結果の一例を図4と表1に示す。

手順8 F9キーを押すことにより、2から7までの手順を繰り返す。このようにF9キーを1回押すだけで一連の複雑な処理の結果を瞬時に行ってくれるところがExcelのすごいところであり、学習者に、心理統計学の世界を自ら動かしているというダイナミックな感覚をもたせてくれる。

3. 心理統計法教育におけるRの活用

ここ2～3年急速に普及してきている本格的な統計ソフトであるRはフリーソフトであることから、パソコンさえあれば誰でも利用可能である。これまであるいは現在、心理統計法教育では統計ソフトとして統計パッケージであるSPSSあるいはSASが一般的に用いられてきているが、それらは操作しやすくかつ高性能・高性能である反面、個人として使用する場合一定の費用がかかるのに対して、Rはフリーソフトであるため無料である。またRはオープンソースプログラムであるためプログラムが公開されており、世界中のユーザーの誰もが開発・拡張に携わることができる極めて発展性が高いソフトである。さらにRはプラットフォームを選ばず、Windows9x/NT/2000/XP/Vista, Unix, MacOS/Os X等で動作させることができるという点や、グラフィックス機能が強力であり、統計計算ソフトウェアとしても非常に優れている等の長所がある。このように書いてくるとRはいいところばかりのようであるが、いくつか難点がある。その一つはGUIではなく、CUI（キャラクターユーザーインターフェイス）すなわちコマンド入力により操作しなければならないという点である。R Commander（Rコマンダー）というマウスでメニュー選択することによりRを操作することのできるGUIのパッケージも開発されているが（舟尾、2007；荒木、2007）、機能や使いやすさという点で未だ十分とはいえない。またRはインタープリターであることもあって、出力の見やすさや出力結果の他のソフトへのコピーペーストといった点において難点がある。しかしRはオープンソースプログラムであることから世界の衆知を集めて今後ますます使いやすくなっていくことが期待される。

心理統計法教育においてRを活用していく際には、以上述べてきたような長所短所を踏まえて対処していく必要がある。しかしやはり最大の問題はRが基本的にはCUIであるという点である。将来研究者やSE等の情報産業を目指している学生を除けば、GUIの使いやすい統計ソフトがある中で文系の学生にコマンド入力をベースとした授業を行うのは大変であるというのが実感である。RコマンダーのようなGUIのパッケージを充実させ使いやすくしていくことが重要であると考えられる。

現時点でRによる心理統計法の本として渡辺(2005)、山田・杉澤・村井(2008)の2冊が挙げられる。前者は分散分析の各種の方法、多変量解析の各種の方法について詳細にわかりやすく書かれた本であり、後者は心理統計法の全般にわたり入門者が読んでいけるようにも配慮された懇切丁寧な本である。このような良書が今後数多く出版されることによって、心理統計法教育におけるRの活用も次第に多くなっていくのではないかと思われる。

Rを用いると、非常にコンパクトに統計プログラムを表現することが可能である。たとえば、前章でExcelにより説明したN=25とN=100の標本比率の標本分布を求め、その平均や標準偏差を比較する手順は、以下のように短いプログラムで表現可能である。（なお+は入力しない）

```
>標本比率25人<-numeric(length=300)
>標本比率100人<-numeric(length=300)
>for(i in 1:300) {
+ 標本<-runif(n=25)
+ 標本01<-ifelse(標本<=0.4, 1, 0)
+ 標本比率25人[i]<-mean(標本01)
+ 標本<-runif(n=100)
+ 標本01<-ifelse(標本<=0.4, 1, 0)
+ 標本比率100人[i]<-mean(標本01)
+ }
>mean(標本比率25人)
```

```
>sd (標本比率25人)
>mean(標本比率100人)
>sd (標本比率100人)
```

以上のプログラムの実行結果の一例として、「標本比率25人」の平均と標準偏差はそれぞれ0.398と0.972、「標本比率100人」の平均と標準偏差はそれぞれ0.401と0.497となった。この結果から標本の大きさが大きいN=100の場合の「標本比率100人」の方が、N=25の「標本比率25人」に比べ、より母集団比率0.4に近い値が得られ、標準偏差も理論値どおり、N=25の場合の約2分の1に近くなっていることがわかる。

4. まとめ

以上、心理統計法教育におけるExcelとRの活用の有効性について具体例を交えて述べてきたが、実際に文系の学生を対象とした授業場面でどのように用いていくかということになると、筆者としてはやはり入門期から本稿で述べた推測統計の基本的考え方である大数の法則や中心極限定理あたりまではExcelを利用したほうがよいと思う。ExcelはGUIで使いやすいのみならず、スロットマシン方式で示したように、心理統計法の基本的な考え方を直感的に理解しやすいように説明できるからである。しかもExcelでは学習者が自ら心理統計の世界をダイナミックに動かすことができる。このようにしてある程度心理統計の基本的な考え方が理解でき学習意欲を喚起できた段階で、Rを扱ったほうがよいと考える。

文 献

- 荒木孝治（編著） 2007 RとR Commanderではじめる多変量解析 日科技連
舟尾暢男 2007 R Commanderハンドブック 九天社
塗師 斌 2002 統計教育におけるExcelの活用の可能性—スロットマシン方式による大数の法則と中心極限定理の教授法— 横浜国立大学教育人間科学部教育実践研究指導センター紀要第18号
渡辺利夫 2005 フレッシュマンから大学院生までのデータ解析・R言語 ナカニシヤ出版
山田剛史、村井潤一郎、杉澤武俊ほか 2007 文系学生に対する心理統計教育の実践 日本教育心理学会第49回総会発表論文集 S26
山田剛史、村井潤一郎、杉澤武俊ほか 2008 文系学生に対する心理統計教育の実践2 日本教育心理学会第50回総会発表論文集 S98
山田剛史、杉澤武俊、村井潤一郎 2008 Rによるやさしい統計学 オーム社