

調査データ分析の教育における リモート集計システムの活用



保田 時男 | Yasuda Tokio

大阪商業大学総合経営学部講師

■1975年生まれ。97年大阪大学人間科学部卒。99年大阪大学大学院人間科学研究科博士前期課程修了。2003年同博士後期課程満期退学。修士（人間科学）。04年4月より現職。大阪商業大学JGSS研究センター運営委員。専門は計量社会学、家族社会学。社会調査の方法の改善、欠損データの分析、中期親子関係の分析などに関心を持つ。

1. 現実の調査データを教育に利用する

本稿では、リモート集計システムを調査データ分析の教育に活用することについて考察する。無料でどこでも調査データの分析ができるリモート集計システムは、教育への活用が大いに期待されるが、運用上の課題も少なくない。その利点と問題点について、筆者の体験を基に、順に解説する。

東京大学社会科学研究所 SSJデータアーカイブにおけるリモート集計システムでは、いくつかの調査データをオンラインで分析することができるが、その1つにJGSS（Japanese General Social Surveys）がある。JGSSは大阪商業大学JGSS研究センター（旧比較地域研究所JGSS部門）が、東京大学社会科学研究所を主な共同研究機関として実施している全国規模の総合的社会調査である。幅広い関心をカバーする継続調査の個票データを速やかに公開し共同利用することで、社会科学の全般的

な発展を企図している。詳細はJGSSホームページ（<http://jgss.daishodai.ac.jp>）に記されているので参照されたい（図1）。

JGSSプロジェクトは、公開データを教育に活用することをその目的の1つとしており、当初から大学の授業等でJGSSデータを利用することを推奨してきた。現実の調査データを利用した教育は、実際的な社会の分析能力を育てるために非常に有効と考えられるからである。毎年、20～30件程度の授業でJGSSデータが利用されてきたが、データの教育利用をさらに推し進めるために、2003年頃からJGSSデータを利用した調査データ分析のテキストを作成する企画が練られ始めた。有名なボンシュテッド&ノーキのテキスト『社会統計学』で米国GSSのデータが有効に活用されていることを手本とした発想である。筆者は、ちょうどその頃、JGSSプロジェクトの一翼を担うようになり、岩井紀子教授（JGSS研究セ

図1 JGSSホームページ



ンター・センター長) と共にテキストの企画・執筆に参加する機会を得た。

しかし、このテキストの企画において当初から筆者を悩ませたことがあった。それは、学習者にJGSSデータをどこまでどのようにして取り扱わせるか、ということである。せわかつの公開データであるから、やはり学習者が実際にデータに触って分析を行えるという利点を活かしたいが、そのためにSPSS等の統計分析ソフトの利用を前提にすることには、一抹のためらいがあった。なぜならば、その分析ソフトが導入されていない教育機関ではテキストが有効に活用できなくなる上に、ソフトの操作方法の解説に紙面が割かれ、現実の調査データを用いた学習という力点がぼやけてしまうことが懸念されたからである。そ

のようなこともあり、当初、テキスト作成の企画は遅々として進まなかった。

2. リモート集計システムとの出会い

そのような折、共同研究機関の社会科学研究所がリモート集計システムの導入を予定していることを知らされた。それは、まさに渡りに船のシステムであり、懸案が一気に解決できる感触を得たことを、今もはっきりと記憶している。オンラインで誰もが無料で分析できるツールがあれば、特定の統計分析ソフトの導入に悩まされずに済む。また、データの利用申請をする前に簡単な分析でデータ内容を

確認することを想定していたため、操作も簡単に初学者にうってつけのシステムであると感じられた。

リモート集計システムをテキストに利用したい意向を社会科学研究所の担当者に伝えたところ、そのような利用は想定していなかったが、可能なはずである、との回答を得た。そこで、早速、リモート集計システムの利用を前提にして企画を練り直し、予定より大幅に遅れていたテキストの執筆が急ピッチで進められることになった。

とはいえ、できたばかりのリモート集計システムで実際に授業を行ったことはなく、ゲラを作成しては授業で試し、修正する、という手探りの状態で執筆は進められた。今から

思えば、リモート集計システムがまったくうまく機能しなかった場合にはどうするつもりであったのだろうか、やや恐ろしくもあるが、幸いなことに、さまざまな問題を抱えながらも（第5節で触れる）、無事にテキストを完成させることができた。

3. 『調査データ分析の基礎』

こうしてできあがったのが、『調査データ分析の基礎—JGSSデータとオンライン集計の活用—』（岩井紀子・保田時男著、有斐閣、2007年）である。このテキストでは、以下のように基礎的な記述統計から多変量解析の入り口までを解説している。

〈Part I〉

- 1章 調査データとは何か／2章 社会調査の手順の概要／3章 多様な分析の方向性／4章 公開データの利用

〈Part II〉

- 5章 度数分布表／6章 グラフ／7章 代表値とばらつき／8章 複数回答の扱い方／9章 2変数のクロス集計表／10章 関連性を表す統計量／11章 3変数のクロス集計表

〈Part III〉

- 12章 統計的推定の考え方／13章 統計的検定の考え方／14章 独立性の検定と比率の差の検定／15章 平均値に関する推定と検定

〈Part IV〉

- 16章 分散分析／17章 回帰分析の基礎／18章 回帰分析の応用

と一般線型モデル／19章 ロジスティック回帰分析

これらのほぼすべての解説項目にわたって、リモート集計システムによるJGSS-2000の分析例を示している（システムがカバーしていないPart IVを除く）。リモート集計による分析例では、必ず操作手順と実際の出力画面を示し（図2）、学習者が操作面で戸惑うことのないように配慮している。また、リモート集計システムで解答することを前提にした章末問題を付けている。

幸いなことに、リモート集計システムを活用して現実の調査データで効果的に教育を行うという狙いは、おおむね成功したように思える。『調査データ分析の基礎』は、2008年度では、少なくとも11校の授業で教科書指定を受けており、一定の需要に答えている。また、やや意外であったこととして、大学院生（特に、中途からデータ分析の勉強をしなくてはならなくなった大学院生）からの評判がよい。調査研究に関わる環境に恵まれなかつ

図2 リモート集計システムによる例題

例題 人々の月給には、どのくらい格差があるのか（ばらついているのか）。

JGSS-2000で月給のばらつきを算出してみる。ばらつきも、やはり「1. 度数分布と関連統計量の集計」で算出できる。「問01_24s4 月給」を集計対象の変数として指定し、「最小値」「最大値」「4分位」「標準偏差」を集計すると、図7-5のような結果が表示される。

統計量		
実数: 問01_24s4 月給		
度数	有効	955
	欠損値	1938
標準偏差		179807.366
最小値		10000
最大値		2000000
パーセンタイル	25	195000.00
	50	280000.00
	75	400000.00

標準偏差は約179,807円なので、平均値(304,809円)の±179,807円の間、すなわち125,002円～484,616円の間多くの回答(約3分の2)が集まっていることがわかる。分散は表示されていないが、標準偏差の2乗なので、およそ

た大学院生が、独学でデータ分析の知識と技術を学び直すには、自由に現実のデータに触れることができるリモート集計システムおよびそれを基に組み立てられたテキストは、便利な導入ツールとなるようである。昨年の夏に社会調査士資格認定機構が開催した「専門社会調査士取得希望者のための講習会」において、筆者らは『調査データ分析の基礎』をテキストとした5日間の講習を行ったが、好評であったため、今年も開催されることになった。

4. リモート集計システムによる授業の利点

リモート集計システムを授業に活用することには、明らかにいくつかの一般的な利点がある。筆者の体験から感じられた主な利点をいくつか記しておこう。

(1) 分析ソフトが不要

リモート集計システムを利用した授業のもっともわかりやすい利点は、自前で統計分析ソフトを用意する必要がないということであろう。大学の設備環境によっては、この利点は甚大である。一般に統計分析ソフトは決して安価ではないので、受講学生の人数分だけソフトが確保されていないことも多い。リモート集計システムであれば、インターネットに接続されているパソコンさえあれば分析実習が可能になる。

これと関連することとして、特定の分析ソフトに依存しないで済むことも1つの利点である。これは特にテキストを利用する場合の問題であるが、テキストで用いられている分

析ソフトと大学で導入されている分析ソフトが異なるために、そのテキストが利用できないといった事態に陥らずに済む。

(2) 分析の理解に集中

一方で、より本質的な教育上の利点として一番にあげられるのは、分析ソフトの操作方法を習得することに授業時間を占領されずに済むことであろう。リモート集計システムでできることは限られているが、その分、習得が必要な操作は少ない。操作方法の説明に多くの時間を費やすことなく、データ分析の考え方そのものの理解に学習を集中できることは、大きな利点である。

データ分析の考え方に学習を集中したい場合、実習を行わなければよいという意見もあるかもしれないが、コンピュータに個票データを集計させるという具体的なイメージがない学生にとって、机上で学習を進めることはやはり難しい。一定の体験によって実感を持つことは大切であろう。自分の操作で多様な集計ができるという体験に素朴な感動を示し、統計的なデータ分析に興味を持つ学生も多い。

また、結局、本格的な分析を行うためには専門的分析ソフトを用いなければならないのであるから、初めからソフトの操作方法を学習しておいた方がよい、という意見もあるであろう。これには一理あるが、同じ目的を複数のルート（リモート集計システムと通常の分析ソフト）で果たすことを経験することには、本格的な分析ソフトに早急に慣れることよりも大きな意義があるように思える。終始1つの分析ソフトに依存することは、しばしば分析の目的・理論を、特定のソフトの操

作・出力と誤って結びつけることになる恐れがあるからである。リモート集計システムを用いて分析のイメージを理解した後に、本格的な分析ソフトで、より発展した分析を習得することは、理想的に思える。

(3) 自学が容易

リモート集計システムによる授業のもう1つの大きな利点は、自学が容易になることである。通常分析ソフトを用いている場合、自宅で分析ソフトを用いた予習・復習をすることはほぼ不可能である。しかし、リモート集計システムであれば、インターネットに繋がるパソコンさえあれば、どこでも自学が可能である。

自学が容易になるということは、データ分析の宿題やレポート課題を出しやすくなるということである。特に、宿題にリモート集計システムを活用することは、非常に大きな教育効果が得られる、という感触を得ている。

データ分析の実習では、どうしてもコンピュータ・リテラシーの格差によって学生の進行スピードに差が出やすい。そのため、単にコンピュータに慣れていないという理由で授業についていけない学生が出たり、逆に慣れた学生が退屈してしまうことが多い。宿題を出せることは、その調整をかなり容易にしてくれる。

同様の利点は、Rなどのフリーソフトを活用することでも得られるかもしれない。フリーソフトであれば、自宅のパソコンでも利用可能だからである。しかし、コンピュータ・リテラシーが高くない学生が、直接の指導なしに本格的な統計分析ソフトで自学を行うこ

とは、やはり難しいであろう。自学が可能であることによって、コンピュータ・リテラシーが高い学生がますます高度な技術を身につけ、格差が開いてしまう可能性もある。適度に簡単なことしかできないリモート集計システムは、コンピュータ・リテラシーの高くない学生を取り残さないためにも有効と考えられる。

(4) 卒業研究等への発展の可能性

自学が容易になるということと関連しているが、リモート集計は卒業研究等に発展する可能性を開いてくれる。これは、現実の調査データを用いた学習一般に言えることであるが、学生がそのデータへの関心を深めれば、現実に意味のある分析を自由に行うことができる。場合によっては、それが本格的な卒業研究等に繋がることもあるであろうし、むしろ、そのような方向に積極的に指導することも可能であろう。

リモート集計では、そのような現実データの利用一般に見られる利点に加えて、分析が自宅でも容易にできるという利点があるので、データへの関心を深める機会がより多い。リモート集計システムを用いたデータ分析の授業を修了した後に、分析のアイデアを思いついた場合や他の授業で学んだことと関連づけて新しい分析に関心をもった場合に、自学でそれを発展させることが期待できる。

5. リモート集計システムを授業で利用する上での問題点とその改善

リモート集計システムは、当初、授業での利用を想定せずに構築された。そのため、当

然ではあるが、システムを実際に授業で利用してみると、さまざまな問題が持ち上がった。新しい問題が発覚するごとに、その改善をお願いしたため、担当者には多くの手間を取っていただいたが、おかげで現在のシステム（およびそのJGSSデータ）は、授業での運用が可能なものになったように思える。やや細々としたことであるが、同様のシステムを構築する際に参考になる可能性があるので、授業で利用する上で問題になった点とその改善のいくつかをここに記しておく。

(1) 変数の検索

筆者がリモート集計システムを利用してみですぐに感じた問題は、「集計する変数をリストの中から探し出すことが容易でない」ということであった。リモート集計システムでは、各変数の内容を簡潔に表す「変数ラベル」（例：住居形態、住宅床面積、同居家族人数、……）のリストの中から集計したい変数を選択する。しかし、JGSS-2000のデータには500以上の変数が含まれているので、例えばこの中から「住宅床面積」を目で探し出すことは容易でない。

通常の統計分析ソフトでは、変数ラベルとは別に、形式的な「変数名」（例：SZDWEL）を設定しており、変数名をアルファベット順に表示する機能や変数名を直接入力して分析対象を指定する機能があるので、変数の検索に悩まされることはない。また、変数は基本的に調査票の質問順に並んでいるので、調査票の構造を熟知していれば、変数ラベルだけでも目的とする変数の位置の目星は付けられる。しかし、リモート集計システムを用いた

授業では、学習者が調査票の構造を熟知していることは期待できない。

この問題に対応するために、変数ラベルの先頭に調査票の質問番号を付けることにした（例：問16 住居形態、問17 住宅床面積、問18_1 同居家族人数、……）。これならば、調査票の質問番号と照らし合わせて、容易に目的の変数を探し出すことができる。通常分析ソフトを利用している際にはまったく問題にならない意外な盲点であった。

(2) データの加工ができない

次に、リモート集計システムではデータの加工がまったくできないことが問題になった。この点は、当初、データの加工を必要としない範囲で授業を組み立てれば問題ないと考えていたが、実際にはそれほど単純ではなく、以下のような点で問題が浮上した。

第1に、データの加工ができないということは、欠損値の指定が自由にできないということである。JGSSでは、通常「非該当」のコードのみを欠損値に指定し、「無回答」のコードは欠損値に指定していない。そのため、そのまま平均値等の統計量を算出すると、無回答のコード（9、99、999等）を計算に含めたおかしな結果を出力してしまう。同様に、「答えたくない」「わからない」等の回答コードも欠損値に指定していない。やむを得ないので、収入や年齢などの量的変数については、無回答も欠損値に含めたデータに差し替えた。

データの差し替えによって、適切な統計量が算出されるようにはなったが、「無回答」「答えたくない」「わからない」等が分析対象

から除外されることは、分析目的によっては望ましくない。データの加工を認めないとしても、分析に欠損値を含めるか含めないかを選択できるオプションがあれば、ありがたかった。

第2に、単純なリコーディングすらできないことは、しばしばテキストの記述をわかりにくくすることに繋がった。例えば、何らかの意識設問で、満足(=1)から不満(=5)のコードを、満足(=5)から不満(=1)と逆転したい場面は多い。値が高いほど肯定的である方が、直感的にわかりやすいからである。しかし、データの加工ができない以上、「○○に対する満足度(得点が低いほど満足)」といった記述で説明せざるを得ない。同様に、「非常に満足」と「ある程度満足」の2つをまとめて「満足」とリコーディングすることすらできないので、リモート集計の出力を基にして改めて2つのカテゴリーを合わせる計算をする必要があった。

想定されるリコーディングをすでに行ったデータをリモート集計システムで用いるという方策も考えられるが、それでは調査データの分析におけるリコーディング作業の重要性が隠されてしまう。『調査データ分析の基礎』で用いるJGSS-2000のデータに関しては、極めて必要性が高い「年齢10歳刻み」の変数だけを加える対処をすることにした。

データの加工を認めることは、もしかすると学習者を混乱させることになるかもしれない。また、筆者には詳しいことはわからないが、システムの管理を複雑にするのかもしれない。しかしながら、一般に、社会調査のデータを分析するには、データの加工が必要不

可欠なので、ぜひ一定のデータ加工の機能は搭載してほしい。

(3) ID認証の問題

本質的な問題ではないものの、リモート集計システムの授業活用においてももっとも困った事態に直面したのは、実はID認証に関する問題であった。リモート集計システムを利用するためには、IDとパスワードを取得して認証を受ける必要がある。ID認証自体は、さまざまなインターネットサービスで必要なことであり、学生もその仕組みに慣れているのであるが、当初のシステムではその具体的な手続きにいくつもの問題があった。

利用者のメールアドレスを登録すると、そのアドレスにパスワードがメールで届けられるが、ここでまず問題が起こった。リモート集計システムでは、そのメールアドレスを大学等の機関のアドレス(…….ac.jp等)に限っている。筆者が担当した授業では、ふだん携帯電話のメールしか利用しない学生が大半であったため、大学から割り振られているアドレスを知らなかったり、パスワードを忘れていたりしていることが続出し、その解決に大変な時間を費やさざるを得なかった。

さらに悪いことに、利用者のメールアドレスそのものがユーザーIDとなる仕組みのため、学生は初めのID取得だけでなく、システムを利用するたびに、登録したメールアドレス(=ふだんは使わないメールアドレス)を入力しなければならなかった。大学から割り振られるアドレスは、学籍番号などから構成された無意味な長い文字列であったため、単純に入力ミス等も続出し、授業でシステムを

利用するごとにその解決にあたらなければならなかった。しかも、当初のシステムは現在よりもID認証が不安定で、正しいIDとパスワードを入力してもエラーが出るのがしばしばあったため、問題の解決により手間がかかった。

この事態にはほとんど困り果てたため、担当者に事情を説明し、改善をお願いした。その結果、メールアドレスとは別のIDを設定できるようになり、さらに、1つのアドレスに複数のIDのパスワードを届けてもらえるように改善していただいた。つまり、教員のメールアドレスに各学生の初期パスワードを届けもらうこともできるようになり、学生のメールアドレスに依存する必要がまったくなくなった。この改善は、授業の運用上は本当に助かったので、あえて強調しておきたい。

(4) その他の問題

その他、遠隔操作である以上当然のことではあるものの、授業での運用上は注意すべき点がある。まず、システムにトラブルが起こった場合、利用者側では即時的な対処のしようがないことには、やはり気を付けなければならない。コンピュータにトラブルは付き物であるが、目の前のコンピュータであれば再起動させたり、隣の席に移動したりすれば解決するであろうことでも、遠隔操作ではどうしようもない。

また、授業での利用の場合、どうしてもアクセスが集中しがちになる。設備が十分であれば、多くのアクセスに耐えられるのである

うが、現状ではアクセス過多とならないように注意が必要である。上の2点から考えると、リモート集計システムの活用は、宿題を中心とした利用を想定することがもっとも安全であろう。

もう1つ注意すべきことは、WWWブラウザを用いているので、分析の履歴が残っていないということである。通常の統計分析ソフトであれば、少なくとも分析結果の出力は累積的に残っていくので必要な出力をまとめて保存できるが、リモート集計システムでは、必要な出力はそのつど保存しなければならない。また、どのような分析命令を出したのかも後からではわからない。

6. リモート集計システムへの期待

やや苦言を呈するような形になったが、基本的に、リモート集計システムが非常に有効な教育ツールであることは間違いない。同じような機能をより効果的に果たすツールがあるとすれば、機能を絞った簡便な統計分析ソフトがフリーソフトとして出回ることであるが、開発者側の動機が考えられないので、あまり期待はできないであろう。また、同時に、現実の個票データを誰もが直接利用できることが必要条件になるが、これも現状では障害が多い。

やはり、リモート集計システムを洗練し、磨きをかけることが、現状で最善の方向性であると考えられる。細かな点で改善を施せば、より効果的な活用が可能となるので、ぜひ更なる展開がなされることを期待したい。