

# 社会調査教育に役立つデータ・アーカイブ

——SRDQ（質問紙法にもとづく社会調査データベース）によるテキスト——

川 端 亮（大阪大学大学院人間科学研究科教授）

## 1

### 研究におけるデータ・アーカイブの普及

データ・アーカイブとは調査データの収集、編集・加工、保管、普及のための専門機関であり、調査データを収集し、それを二次分析が可能な統一的形式へと編集・加工し、公開するものである。この10年間でデータ・アーカイブという言葉もずいぶんと普及してきた印象がある。

データ・アーカイブは研究と教育の2つの面で意義をもっている。研究上の意義としては、データ・アーカイブに保管された過去の調査データを有効に活用できる点があげられる。既存のデータを活用した二次分析が普及すれば、同じような目的の調査を繰り返し実施する必要もなくなり、多額の調査経費が不要になる。また調査対象者は似たような質問の調査に繰り返し回答するといった負担が軽減される。そして計量分析を行う研究者にとっては、従来は無作為抽出による全国データを使うためには、大規模な調査グループに属さなければならなかったが、現在ではデータ・アーカイブを利用すれば、遠方の調査グループに属さなくても、さらには多くの国で実施されるISSP調査やWorld Value Surveyなどの国際比較データであっても分析することが可能となってきた。

しかし、だからといってデータ・アーカイブを利用した二次分析が増加しているかといえ、必ずしもそうではない。たとえば、佐藤ほか編（2000）では、社会学評論に掲載された論文のうち、データの二次分析による論文数が、1998年は25本中1本（4%）、1999年は、23本中3本（13%）であったと紹介されているが、10年を経た2008年は、28本中3本（11%）、2009年は、14本中0本、2010年は、10本中1本（10%）であり（特集論文は除く）、増加傾向にあるとは言え

ない。しかし、一定の利用率は維持されていると言えるだろう。

研究利用を目的としたデータの二次利用はそれなりに進んだと言えるが、もう1つの側面である教育のための利用では、データ・アーカイブの潜在的な能力・機能が、利用者に使いやすいかたちで提供されているとはいいがたい。研究用に優れたデータは、しばしば教育用にも優れたデータである。しかしながら、それらは教育用には使いにくい網羅的な形式で配布される。たとえば数多くの研究業績を生み出してきた、SSM調査データは、1つの調査票で変数が100も200もある。もし授業で使うとしても、1回の授業では少なければ1変数、せいぜい5変数から10変数くらいだろう。大部分は使わない変数を含むデータセットを配布することは、ファカルティ・デベロップメントが流行している大学においては、授業担当者として気が引けるだろう。しかしながら、自分であらかじめ、少数の変数だけを含むデータセットを毎回、毎回作成するのは面倒だ。授業に必要な変数だけを選び、使いやすいかたちのデータセットに編集し、配布するだけでも教員側にはメリットがある。そしてそれと連動する教科書があれば、なおさら楽であることは、言うまでもない。

## 2

### データ・アーカイブの教育での利用

質のよいデータで社会調査法を学習することは、きわめて重要である。多くの学生にとって、さまざまな分野で実施された質のよい調査データを分析しながら社会調査法や計量分析の手法を学習することは、データ・アーカイブが提供するデータによってはじめて可能になる。社会調査法の授業では「有意水準が5%で帰無棄却が棄却」というような推測統計学にもとづいた説明をするが、こ

のような説明は無作為抽出にもとづいた標本調査データを分析したときにはじめて意味をもつ。学生がみずから手で無作為抽出にもとづいた質のよい標本調査を実施することは困難であるから、データ・アーカイブが質のよい調査データを提供することは統計的検定を理解するうえでも意義がある。

また、現実の社会から調査によって得られたデータを分析することは、それ以上に重要である。架空データによる分析結果が示されている統計学のテキストは多い。サンプルデータは、分析の結果が非常にクリアであり、美しい結果が得られるため、分析手法の有効性はわかりやすい。分析手法を手法として学ぶためだけなら、それもよいかも知れない。しかし、データ分析を道具として、社会の様子、実態について考える場合、社会から実際に得られたデータの方がよいのは言うまでもない。2変数の相関係数といっても、それが高い値なのか、低い値なのかは、絶対的な基準、たとえば0.5より大きいとか、0.8より大きいなどによっては、決められない。値の大きさは、実社会の現象を反映して理解することが必要である。若い世代ほど進学率が上昇したことは、おそらくどの学生も知っているだろうが、それを相関係数で表すと、たとえばマイナス0.3であるということや、また、「いい学校に行けば、いい職に就ける」と言われて、大学に進学した人も多いだろうが、学歴と個人所得の相関は、たかだか0.1程度であることなどを知ることが、相関係数の意味を知ることである。サンプルデータの分析では、社会を知るといふ学問的な知的好奇心はあまり刺激されることがないだろう。

分析手法の本を読むだけでは社会に即した計量分析ができるようにはならない。社会の実情を反映するよいデータを使って、自分の手でパソコンを操作し、間違いながらもあれこれ試して、実際にデータ分析をしてみることを繰り返して、計量分析の力がつく。

さらに分析手法を身につけないと計量分析はできないのは確かであるが、計量分析の手法を身につけても、それに見合った専門分野の理論、先行研究を十分に理解していないと、データ分析の結果の正しく、意味のある解釈はできない。これは教える側のものなら誰もがよく知っていることで

はあるが、これまでの計量分析のテキストは、分析手法についての解説だけに限定されることが多かった。もちろんこれだけで1冊のテキストの分量に収まりきらないくらいの分量になるからであるが、計量分析が行われているよい研究事例を取り上げて、その学問的な意義を説明することも、手法を学ぶことと同じくらい計量分析の教育においては、重要なことである。これによって、その分析手法の使い方を研究の理論的な背景も含めてよりよく理解できるであろう。つまり、よい研究で実際に用いられたデータを同じように分析することで、研究分野の中でその手法がどのように使われたときにすばらしい研究が生まれるのかを、よりよく理解できるだろう。

優れた研究を生んだ質の高いデータを実際に分析してみることができると、分析手法の解説だけでなく、その分析手法を用いた論文や本などの研究事例を掲げ、その研究事例のよい点や特徴なども学ぶこと、この2点を兼ね備えた教育を行ううえで一番重要なことは、よい研究事例を生んだデータを誰でもが容易に利用できる環境を整えることである。これを実現できるものこそが、データ・アーカイブである。

「社会調査データベース SRDQ」(Social Research Database on Questionnaires)では、さまざまな社会調査の標本設計や質問文・選択肢を集めて、データベース化している。そして調査データをweb上で直接分析できる機能をもつ。SRDQの際だった特徴は、高価な統計ソフトウェアを各自で準備することなく、インターネット・エクスプローラーなどのブラウザだけで統計分析ができることである。しかもデータをダウンロードすることなく、Webページ上で単純集計だけでなくクロス集計や分散分析、重回帰分析などの多変量解析も行うことができる。また、SRDQは、SPSSをベースに作られているため、SPSSと同じような操作方法でデータ分析が行える。そして利用申し込みやデータ送付の手続きは必要なく、即座に分析を行うことができる。以上の利点をもつため、これから統計分析を学びたい人の教育においては、利用しやすい機能を備えたデータ・アーカイブと言えるだろう。



図1 SIRDQのトップ画面

### 3 ウェブサイトと連動した教科書

本格的な社会調査のデータ・アーカイブは、現在、かなりの数が存在するようになったが、海外の他のデータ・アーカイブと比べても、誰でもがブラウザだけで多様な統計分析が容易にできることがSRDQの特筆すべき特徴である。そのためのデータベース設計上の工夫点は多い。しかし2004年に公開を開始し、すでに7年目に入っているSRDQが、すなわち7年前に企画・設計されたものが、いまだに独創的な仕組みのデータ・アーカイブであるということは、誇るべきことというよりはむしろ、この6年間にデータ・アーカイブの領域で、より使いやすいユーザーフレンドリーなシステムが生み出されなかったという悲しむべきできごとなのかもしれない。

筆者もなかなか教育用にデータ・アーカイブを利用する試みを実現できなかった1人であるが、幸いなことに筆者は、代表者として、2007年度に文部科学省の大学院教育改革支援プログラム(大学院GP)「人間科学データによる包括的専門教育」に採択された。そこでは、人間科学研究科という、心理学、教育学、社会学などの異なる研



図2 サポートページのトップ画面

究分野から構成される研究科において、研究分野の学問的背景を踏まえたうえで、架空のデータではなく、研究に使われた質の高いデータを用いた実習によって、分析能力を高めることがめざされた。データ分析技法の習得のみにとどまらず、データ分析の結果を、社会における人間の実際の行動の理解に結びつけるという解釈面の力をのばすことが主要な教育目的に据えられたのであった。そして2009年度にはこのプログラムに沿った取り組みの成果として、計量分析の教科書が執筆され、『データアーカイブSRDQで学ぶ社会調査の計量分析』という教科書が出版された(川端編, 2010)。このテキストは、web上の専用のサポートサイトと連動している。図1のSRDQのトップページ(<http://srdq.hus.osaka-u.ac.jp/>)の左下に「Support Page」というバナーがあるので、そこをクリックすると、図2のサポートページ(<http://srdq.hus.osaka-u.ac.jp/book/>)が開く。ここから各章をクリックするとその章で紹介された分析を行うことができる。

このテキストにおいては、一部の章を除いて、個々の分析の「技法の説明」や、「分析シミュレーション」に先立って、「研究事例の紹介」が置かれている。

第1章は、第2章以下のさまざまな統計分析の手法を学ぶために用いるSRDQの使い方を解説

したページであり、研究事例は紹介されていないが、第2章では、クロス集計表の使い方、とくに3重クロス表によるエラボレーションを、性別役割分業意識を扱った研究を例に取り上げて解説している。第3章は、階層論において基本的な社会移動の移動表の諸指標の求め方と、近年よく使われるようになったログリニアについての解説である。第4章は多元配置の分散分析についてである。多元配置の分散分析の特徴である、2つの要因が重なり合って現れる交互作用効果が明瞭に出ている高齢者の家事分担の研究例を取り上げ、交互作用効果についての説明を行っている。第5章は重回帰分析、第6章はパス解析、第7章はロジスティック回帰分析であり、これらは一連のものである。第5章では、所得を決める要因を探索する研究例を取り上げ、第6章では、「中流意識」を取り上げたパス解析を扱い、第7章では、転職における学歴の影響の研究例をあげて2項ロジスティック回帰分析が解説される。第8章から10章までは、因子分析についてであり、第8章では、well-being（生活満足度、幸福感）について因子分析を行い、その結果を一元配置の分散分析に応用する。第9章では因子分析と主成分分析の違いを説明し、主成分分析を用いて階層意識の新たな次元を抽出した研究例を用いる。第10章では検証的な因子分析について紹介する。この章は、近年使われるようになった共分散構造分析の入門となっており、SRDQ上だけでは実際に分析を行うことができず、Amosというソフトウェアが必要である。第11章はクラスター分析である。地位の非一貫性の研究例にもとづいてクラスター分析の使い方を学び、クロス集計を使ってクラスターの特徴を確かめるという構成になっている。

このように、主としてSSM調査の1985年、95年のデータを用いた「研究事例の紹介」をし、そこで用いられている分析「技法の説明」を行い、研究事例と同じ分析を「分析シミュレーション」する。

「第9章 主成分分析」の章を例として紹介しよう。第9章は、今田高俊が1995年のSSMデータを使って分析した「社会階層の新次元——ポスト物質社会における地位変数」という1998年に社会学評論に発表された論文を研究事例として取り上げている（今田，1998）。

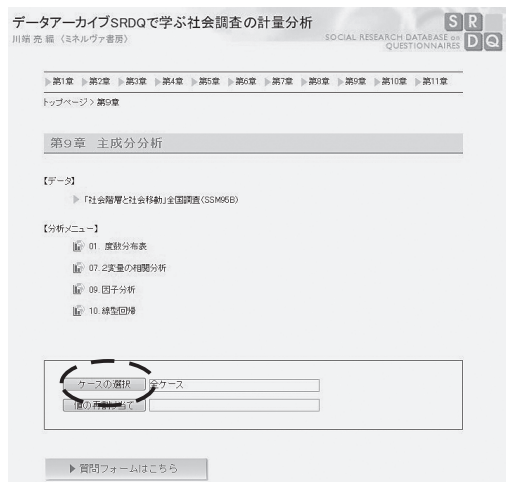


図3 第9章のページ

「研究事例の紹介」として、強調されていることは、2つの因子を抽出する意義である。階層研究では、職業・所得・学歴などから構成される社会的地位を中心に研究されてきた。今田論文では、それとは異なる地位の新しい次元を計量分析によって、示そうと試みる。生活様式に関わる7つの変数から、達成的地位指向と関係的地位指向が主成分分析によって抽出される。地位とは異なる新しい次元、すなわち2番目の因子（関係的地位指向の主成分）を作り出すこと自体は、極端に言えば、さほど難しいことではない。しかし、2番目の因子として、意味ある有効な因子を作り出すことは、じつは非常に難しいのである。今田論文では、関係的地位指向の因子は、「地位競争不安」「現状維持」「ポスト物質指向」の3つの意識との関連が達成的地位指向の因子と異なることを分析によって示している。これが、因子を2つ抽出する意義であり、それをなぞることによって、2つの因子を抽出する価値があることを理解させる仕組みになっている。

今田論文では、男性、女性、男女合わせた全体の3つに分けて、それぞれデータ分析をしているが、テキストでは、女性のデータのみを扱う。図3の第9章のページから「ケースの選択」をクリックする。

すると、図4の画面となり、変数名と値を指定すると女性だけが選択されるようになる。左側のボックスの中に変数名が現れ、そこから変数名を指定するところはSPSSとまったく同じだが、変



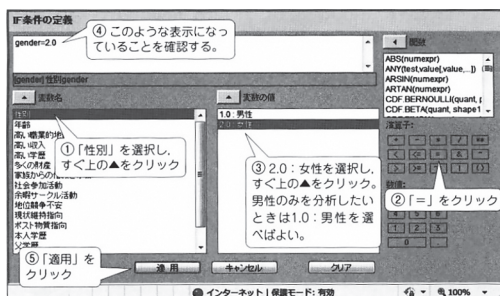


図4 ケースの選択画面

数名を指定すると、その変数の値の一覧が真ん中のボックスに現れる点は、SRDQの便利なところである。とくに計量分析を学び始める人にとって、このような親切さは欠かせないだろう。

主成分分析の分析手順を設定する画面は、次の図5である。テキストでは、操作を吹き出しの①～⑥のように、順に従ってクリックしていけば主成分分析の結果が得られる。

そして最後に「課題」を掲げている（一部課題のない章もある）。「課題」は、本文で説明した「分析シミュレーション」で用いられた変数の一部を入れ替えるなどしたものである。「分析シミュレーション」と同じような分析であり、やらなくてもできると思われるかもしれないが、分析技法を身につけるためには繰り返し練習することが必要であり、「課題」に取り組むことは、計量分析の力をつけるうえでは、欠かせない。また、テキストのサポートページからSRDQ本体に戻れば、「分析シミュレーション」で用いた同じ調査のデータがすべて利用可能なので、それらを用い

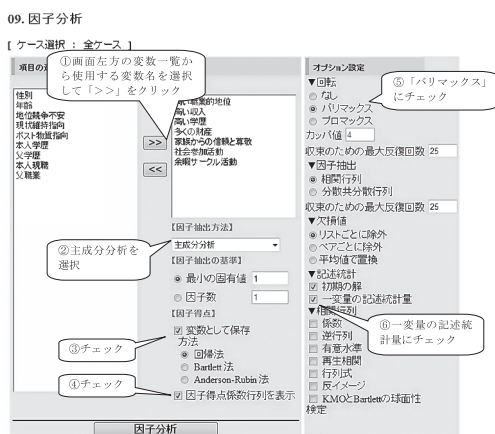


図5 主成分分析の画面

て、さらに自分で課題をたて、練習することも可能である。

## 4 自宅学習の重要性

1コマ90分の授業15回で、計量分析を教えることは難しい。計量分析は、分析を繰り返して練習してはじめて身につくが、授業時間だけでは不十分であることは、教える側の立場に立ったものならば、誰もが思うことだろう。大学設置基準第21条に「一単位の授業科目を四十五時間の学修を必要とする内容をもつて構成することを標準とし」と書かれている。本来、2単位科目は、大学等の授業時間内だけではなく、授業準備や復習に60時間を割かなければならないというのが、建前である。単位の実質化を強調する文部科学省の意向や中央教育審議会の答申はさておき、計量分析を実質的にできる能力を身につけるためには、授業時間以外に、それと同じ程度の自習が必要であり、そのための学習環境を整備することは不可欠だろう。計量分析は、比較的高価な統計ソフトウェアが必要であり、学生に統計ソフトウェアを買わせることは、難しい。大学外で統計分析の練習を可能にする手段の1つは、本誌5号の104-108頁で紹介されたフリーのRを使うことである。しかし、Rは初心者には敷居が高いであろう。大学は、学生がいつでも自習に利用できる統計ソフトがインストールされたパソコンを備えられればよいが、すべての大学でそれが可能ではないだろうし、学生にとってみれば、自宅で深夜でも復習できるに越したことはないだろう。SRDQとサポートページは、パソコンさえあればどこからでも利用可能であり、自宅で課題に取り組んで、復習や自習に活用することができる。大学教育を実質化するうえで、e-Learningの教材開発に力を入れている大学も多いように見受けられるが、同じように社会調査においては、本当の計量分析の力を身につけさせるためには、データ・アーカイブをより活用した教材やテキストを開発していくことが必要であろう。

### 注

・1 SRDQは、2003年、直井優大阪大学教授（当時）により作成が始められ、樋口耕一立命館大学准

教授のアイデアと尽力によって、調査の基本情報の提示、質問項目の検索、web上で調査データが直接分析できるシステムとして公開された。本稿では触れなかったが、SRDQでは、収録されている144件の調査の質問文が、全文テキスト化されており、検索機能を用いることで、検索語を含む質問文を探すことができる。また調査データの分析を行う際にも、その調査に含まれている質問文の一覧を別の画面で表示したり、同じ質問文が他の調査にも含まれていないかどうか確認しながら、分析するというかたちで、検索機能を活用できる。

・2 SRDQの詳細は、松川ほか（2005）参照。

#### 文献

- 今田高俊，1998，「社会階層の新次元——ポスト物質社会における地位変数」『社会学評論』48(4): 419-37。
- 川端亮編，2010，『データアーカイブSRDQで学ぶ社会調査の計量分析』ミネルヴァ書房。
- 松川太一・樋口耕一・川端亮，2005，「社会調査データ・アーカイブSRDQの公開」『大阪大学大学院人間科学研究科紀要』31: 61-84。
- 佐藤博樹・石田浩・池田謙一編，2000，『社会調査の公開データ——2次分析への招待』東京大学出版会。

