

従前の横断調査においては、被調査者から報告されない調査事項、チェックによって正当と認められない値は「不詳」として扱われている。もしくは、必当然的にあるべき値が推測されるならば、その値によって回答欄は埋められるであろう。ところが縦断調査においては、これに加えて、途中脱落者によって生じたデータの欠落への対応、スキップされた調査回によって生じたデータの連続性のとぎれへの対応を考慮しなければならない。

欠損値（スキップ回ではすべての項目がこれに該当する。また、回答拒否項目を含んだ未回答項目やチェックによって不詳とされた項目の値も欠損値として扱われる。）の補完のための具体的な方法には、「系列平均」、「周囲値平均」、「周囲値中央値」、「線型補間」、「その点における線型トレンド」等があるが、現時点ではどれも縦断調査では採用されていない。また、どの方法をどのような場合に用いるべきなのかについても検証されていないし、検証の方法も今のところ持っていないから、補完の可能性についても、また、有意性にしても見当識すら持っていない。

しかし、成年者縦断調査において、たとえば結婚意識の情報が欠けている場合、それをあるグループの平均値、あるいは、上述のどれかの方法をもってすれば補完できるのだろうか。このような被調査者の「意識」について欠損が生じた場合は、いわば意識決定の因子に遡って補完するという概念上の操作が必要である場合もあるため困難は一層多いと思われる。

また、例えば配偶関係が欠けているレコ

ードについても、補完できない（補完してはならない）場合が多いのではないだろうか。他の事項からの推定可能性のらち外にあつて、常に直接の質問項目を構成する結婚意識や配偶関係等の事項は、ある調査の分析にとって致命的な項目として、必ず正しい値によって満たされなければならない、という条件は、縦断調査の成立要件になるのではないかと思われる。あるいは、もしもそのような項目について欠損値が生じた場合は、分析に際しては除外されるレコードとして扱われるべきではないのだろうか。また、補完の可能性を最大限認めるとしても、変数が目的変数なのかそうでないのかによって補完はなされたり、なされなかったりするのではないか。

したがって、頻発する、しかも、事項によって出現率に差がある除外レコードへの対応として、分析のための母集団を設定する除外条件は、分析に使用するレコード集団の特性として認めた上で、分析の条件に該当する集団に属するレコード数は、それぞれ異なるものとして扱われなければならないのではないか。

縦断調査では、横断調査のように一義的な「集計に用いたレコード数」はありえず、分析の具体的な目的に即して必要なレコードをそのつど抽出する、しかもどの調査回を選択するかによって同じ目的でも分析対象数が異なること、たとえば同じ調査対象のレコードが、極端な例では、ある特定の条件によって抽出されたグループに含まれた場合は配偶者ありの属性を持っているが、別の抽出軸では配偶関係不詳として除外されることもありうることを認めたほうがよいと思われる。

また、子の成長過程がそのまま履歴データを構成する出生児調査の特質であるが、身長、体重の数値のように年齢による平均値はあるものの、成長の個人差やそもそもの形質差も考慮すれば、少なくともその子の成長記録の過程が明らかになるまでは、その過程の部分に生じた欠損を埋めることは困難であり、それまでの何年間かは不詳項目として扱わざるを得ない。

補完するとは、ある条件によってその値を推測することに他ならないから、補完した値をもって分析するということは、当該レコードについては、言うなれば論点先取の虚偽に陥ることになる場合もあると考えられる。これは、補完が可能と考えられる「収入額」のような純然たる数値情報であって、かつ特定グループの収入額には一定のトレンドがある予測の範囲内において生じる確率が高い項目についても、実は言うのではないか。

これが、補完を施したダミーを置くことよりも、目的別にレコードをスクリーニングする必要があるのではないかと思われる大きな理由である。

しかし、もしも蓋然性すなわち確率によって上述の懸念を否定できる適正な方法が保証されるならば、縦断調査のように、同一者のデータモデルを保持し続けることが究極の調査意義である場合においては、そこに欠損値があるよりは有意な値が代置されるほうがよほど望ましいのであるから、その適否の検証が必要である。

(4) 分析の方法

横断調査においては、様々な条件によっ

てデータを分類すること、すなわち集計することで処理は完了するとみなされる。しかしパネルデータを用いての集計は、おのずと分析にまで踏み込むことになる。なぜなら、変化の諸相においてパネルデータを分類するならば、変化の有無及び変化の前後の状態が判明している限り、かつ、それらの分類が適切な手法を用いて行われるならば、因果の関係性が結論付けられるはずだからである。これがなされないならば、パネルデータを扱う意義は存在しない。

では「適切な方法」とは何なのか。因果関係という場合、原因の原因、さらにその原因は、というふうに因果の連鎖をどこまでたどればよいのか、と極めて単純な疑問を持つことができる。「適切な方法」は、この連鎖に終止符を打ち、これこそが根本の原因と名指すことができるのだろうか。

その原因は、行政にとっては施策が効力を持つという地点であり、心理学者にとっては意識生成の地点であり、というふうに、分析を行う者の立場によって変動するものであろう。現に、たとえば少子化の原因というとき、様々な「原因」が多くの研究者によって述べられている。

例を挙げる。山口一男氏によれば、「出生意向こそが出生行動の最も大きな説明要因である。（「RIETI Discussion Paper Series 04-j-045:少子化の決定要因と対策について：夫の役割、職場の役割、政府の役割、社会の役割」より。）」。また、前国立成育医療センター総長の松尾宣武氏は、「少子化は、子育てに積極的意義を見出せない親世代から構成される社会が辿る当然の帰結であり、育児思想の再構築、家族機能の修復が少子化対策の要である。（「超未熟児、

極小未熟児の出生数が増え続けている。」厚生労働通信No.15。なお引用文は実際の記述を短縮している。）

前者では、夫婦間の役割合意がどのようなものか、職場が家族に優しいか否か、政府の施策が適切かどうかといった点が出生意向の決定要因として挙げられている。これは個人や世帯を対象とする縦断調査が収集できる範囲を超えて情報を収集し、それに影響を受けて行われる個人の意思決定のタイミングを精緻に測定しなければならないことを意味している。たとえば妊娠を知ったことによって女性の出生意向は当然変化を被るであろう。強まることもあれば、弱くなることもあるが、その変化は本人が意識する、しないに関わりなく、彼女の成長過程での様々な環境、彼女の生活環境、特に親子関係に代表される人間関係によって形成されたと見てかまわないであろう。山口氏はそれを「結婚行動や出生行動は、他の社会行動と同様社会的伝播行動で人々の行動は互いに影響しあい、個人の行動モデルでは説明できない」と表現しているが、親子関係の影響は、その強度において一般の社会関係における影響とは比較ができないほど顕著であろうから、両親に関するデータは分析に当たってもっと参照されてしかるべきである。

後者ではさらに困難なことに、「育児思想の崩壊」や、「家族機能の崩壊」を立証しなければなぜ少子化が進むのかが立証できない。事実と意識の間に関係性があることは否定できない。しかし、どのような事実を検知することが、あるいはどのような行動を、どのような言明を捉えることがそれを立証するのか、通例の調査によっては不可

能であると思われる。直接質問するには、多くの人にとってあまりにもこれらの言葉は了解困難である。また、前者に立ち戻れば、調査によって明らかになる、被調査者経由での事実の、その集積の外部にある因子の選択如何によって分析の結果はおそらく大きく変わってくるであろう。このような地点で「適切な方法」は求められている。

「適切な方法」は、予測を含んでいると思われる。観念的には、あらゆる状況を有効な条件として導き出された結論は、同じ状況と条件の組み合わせが存在するところで導き出された結論と異なっているはずはないからである。したがってシミュレーションが可能になり、条件の強度や方向を変えるもの、すなわち施策の効果測定がその発効前に可能になるであろう。では、そのような時点までやってきてしまうと、調査は不要になってしまうのであろうか。理念的には、その後の調査は実証のために行われるのであろうが、それがそうなるかは、調査をそのような時点まで継続していかなければわからない。

(5) 分析とデータの関係

分析方法は、いうまでもなくその方法が取り扱うデータと相互関係を持っているから、適切な方法は適切なデータを前提としている。

横断調査であれ縦断調査であれ、ある調査については二つの系列のデータがある。上述のとおり、ひとつは被調査者自身の属性として本人または他の被調査者から採集可能なデータであり、もうひとつは彼らを取り巻く環境に関するデータである。

両者に共通した課題は、どこまでのきめ

細かさでデータを収集すべきか、ということである。言い換えればある特定の被説明変数（配偶関係の有無、子どもの数等）の説明変数としてどれほどのものを期待すればよいのかということである。さらにこれを言い換えれば、どこまで仮説を構築するのか、ということである。

では、適切なデータとは、あらかじめ仮説によって項目が具体的に名指しできるようなものであろうか。そうである必然性はないように思える。なぜなら、被調査者のあらゆる属性と彼らを取り巻くあらゆる環境データについて採集することが可能ならば、仮説構築というステップを経ずに、それらあらゆるデータによって相関及び因果の関係性は明らかにされうるはずだからである。

縦断調査が開始されてからこれまでのところ、ふたつのデータ体系のうち、被調査者から採集されたデータに主として重きを置いて分析はなされていると思っている。また、被調査者に影響を与える因子としての環境界を把握することにおいて、特に行動を決定付けるものとしての両親に関する情報や、経済変化のミクロな指標等がまとまったデータとして、欠けている部分がないとはいえない。また、たとえば地域特性を測定するのであれば、得られたデータによって表現される「地域差」によってのみならず、その差を生み出した「地域」がデータ化され、被調査者と同様にモデル化されなければならないであろう。しかし、市区町村レベルで少しずつ違った内容の施策が講じられている場合のようにミクロな分析を実際に行うのであれば、長期の研究に

よって行うしかないと思われる。

これらの、被調査者からみた外部情報は当然被調査者に属するものとして聞き取れる部分もあれば、被調査者から印象として聞き取ったのでは用をなさず、既存の統計から、あるいは新たに調査して統計を作成し、そこから拾いあげなければならない情報もある。いずれにしてもそれらのデータが採集されるタイミングは、ほとんどの事項の場合同一であるべきである。

また、被調査者の属性として、他の属性との関係性において定義される情報もあれば、そうではなくレコード全体にある時間帯、時期においてのみストレスを与え、レコードのある部分を、また部分の組み合わせを変化させるものとして扱われる必要のあるものもある。このようなきめ細かな分析の処理工程はいまだに明確ではなく、したがって方法論として有効な位置づけも与えられておらず、それと相まってのデータの取り扱い方法も、既存の横断調査のそれから未分化であるといわざるを得ない。

(6) コンピューティング

パネルデータの処理に、すなわち毎年蓄積されるデータをクリーニングし、その結果データを使用して履歴データもしくは最新状態を保持した被調査者のデータモデルとその外部情報から成るデータの収蔵体を作成し、それを対象とした分析を行うためにはコンピュータの利用は不可欠である。しかし、処理のアウトプットを一定レベルに保ち、チェックやデータクリーニングの一貫性を保持するためには、コンピュータと共に緻密なマネジメント及びマシナリーが要求される。

マネジメントは、一貫性及び徹底性によって統御されるものであり、マシンリーはそれを実施する組織体である。マネジメントにおいて標準的な手法があるのであれば（例えば ISO9000 規格シリーズのある部分）、その採用によって常に変わらないアウトプットを期待することができる。しかし、そのような文書化は正規化された書法によってなされているわけではないから、(2)で述べたように（ソース）プログラムをそのまま正規文書として扱う方法があれば好都合である。ただし、これは特に縦断調査の統計処理における問題ではなく、より広範なコンピュータ処理における課題である。

また、大量のデータから目的に応じたデータセットが自由自在に構成・抽出できなければ分析において不自由をかこつことになるため、また一貫した処理がなければ均質なアウトプットが期待できないため、コンピューティングにおけるデータセットの構成・抽出等はメニュー化され固定化されていると同時に、そのメニューは更新の自由度を高く持っていなければならない。

自由度の高さに結びつく方法として、操作履歴を常に整理された状態で呼び出せるようにすることがある。膨大な試行錯誤も資産であり、そこから定型へと生成されるものが多いからであり、処理手順の集積を参照体系として方法論を産み出す合理性を持っているからである。しかし、これを合理的に行う方法は VBA のマクロ記録のような簡易なものを除けば、おそらくある程度の資産を投入して新しく開発する必要があると思われる。

3 まとめ

縦断調査は、調査者が被調査者との関係を良好に保って調査の継続を担保すること（しかし、それが被調査者の回答振りにある「傾向」を産んではならない。）、データの蓄積が大量になって調査票の形式でそれを保持していたのでは処理しきれないでのデータ形式を変え、分析に適した形式を新たに作り出すこと（しかし、この処理のために、人的なりソースを無制限に増加させたり、処理時間を延長させたり、予算を自由に使ったりすることが許されるわけではない。）、これらのような新しい課題をもたらした。

新しい課題に対するには、新しい方法論を作り出すしかなく、それは多くの場合、困難と疲れを伴う「走りながら考える」ことになってしまうが、縦断調査は、これまでとはまったく違った新しい調査結果を見せてくれることと期待されており、それが実際にそうなることと、仕事のやりかたが、新しいいくつかの方法論の具体化として調査担当者を支援することとは、まったく同じことなのであるから、なおさら新しい方法は見出される必要があると思われる。

6 縦断調査データ管理システムの検討

金子 隆一
三田 房美
鎌田 健司

1. はじめに

本事業では、縦断調査によって毎年継続して蓄積されて行く統計データに対し、その集計・分析法をも考慮した有効で具体的なデータ管理・分析システムを検討し、開発することを目指している。パネル調査(縦断調査¹)は同一対象(個人)を追跡しながら継続的に調査を実施するものであり、その有効性を十分に引き出すためには横断調査とは異なるデータ管理、および分析が要求される。最も異なる点は、逐次累積される複数の調査回データセットを個人をキーとしながら関係させて集計・分析を行わなくてはならない点であろう。成年者縦断調査においては、男性票・女性票データセットを用いて夫婦単位の集計分析なども行われるため、データ操作はさらに複雑となる。

本稿では、本事業で取り組んでいるパネル調査のデータマネジメントシステムの検討・開発に対して、本年度の課題とされた縦断調査のデータ管理・分析システムとしての基本的課題とその対処法について検討の結果を記述し、またその検討を行うために試作されたエクセル・コード表を中心としたデータ管理・分析システムについて述べた。このシステムは本事業の研究のために新たに作成されたものであるが、実用システムへの発展が構想されているものであり、現行のマネージメントシステムとの連結の仕組みをも含んでいる。

2. データ管理・分析システムの検討

(1) 縦断調査データ管理の課題

縦断調査においては個人として連続するデータを長期に継続して集積することから、データおよびその管理・分析システムには、統一性、汎用性、柔軟性などが要求される。また操作性の観点からは、多数のデータセットを同時に集計分析することから、これらの一括操作、管理が行える必要がある。これらを部分的に備えた既製システムは存在するが、特定の製品に特化したシステムに依存することは汎用性、柔軟性の要請に反することから、ソフトウェアの採用にあたっては、それらの基本的特性や他の部分との親和性についての十分な検討が必要である。すなわち、縦断調査のデータ管理システムにあたっては、具体的なソフトウェアを特定する前に、全体のメタシステム(概念的体系)について十分検討

¹ 調査形態については、本報告書「3 パネル調査における統計分析モデル 2(1)調査法と分析デザイン」を参照。

しておく必要が有るであろう。本章では、以下そうしたシステムの基礎的な課題と要請について、とりわけ縦断調査特有の課題を中心に検討を行う。

実地調査から得られた情報・データを管理するためには、いくつかの種別の情報を維持管理する必要がある。(1) 調査票上に記入された回答をコード化したデータ本体、(2) データの物理的配置を示したデータレイアウト表、(3) コードの意味を対照するためのコード表、(4) データ仕様・クリーニングルール表（および対処表）、(5) 自由回答等リスト、などである。横断調査においては、これらは個票を縦糸、変数を横糸とする2次元のキーにより関係させることになる。しかし、縦断調査においてはさらに時系列が加わり、3次元の関係を扱わなくてはならない。また、成年者縦断調査においては、1回の調査においても、男性票・女性票データセットを用いて夫婦単位の集計分析も行われるため、データセットの関係はさらに複雑なものとなる²。これらひとつひとつに、上記の複数の種別のデータ情報が付随するから、もしそれらを別々の文書、あるいはファイルに散在した形で保管・維持するとすれば、毎回の集計分析に当たって必要なデータ情報を多数のソースから収集する作業が必要となる。それは調査の回数が増すにしたがってさらに煩雑なものとなるだろう³。これに対処するためには、高機能な文書管理システムを用いることが考えられるが、文書管理自体が複雑化して担当者に操作法習得の努力を要求するようなことは望ましくない。その際に最も重要なことは、情報を一元管理することであり、そのことで文書管理システム自体も簡単なものにすることができる。本稿では次節においてコード表を中心としてデータに関する情報を一元管理するシステムを提案する。

次に、データファイルの形式について、やはり長期の継承を行うものなので、汎用性の高いテキスト形式を選択し、これをマスターデータとし、必要に応じて他の形式（統計ソフトウェアなど）を作業ファイル（一時ファイル）として作成して活用すればよいであろう。ここでも同じデータの重複した維持管理は望ましくなく、マスターデータを元としたデータの一元管理が必要である。その際、マスターデータの単位としては、1調査票ごとに1データセットとすることが、自然であり、管理も行い易いものと思われる。集計分析においても、できるだけマスターデータを直接参照する方法が望ましい。その際には再現性の観点から、データの参照から報告結果に至る間に用いられたプログラム等を結果とともに保存する必要がある。また、複雑な集計分析の場合には、対象データセットとして作業用データセットを経る方が、効率的な場合がある。この際にも作業用データセット作成過程のプログラム等を集計分析過程とともに保存する必要がある、さらに作業用データセット自体も参考として保存しておくことが望ましい。ただし、作業用データセットはマスターデータとは明確に区別して保存維持されるべきものであり、管理システム上、マスターデータとの混在があってはならない。

² 実際には、成年者縦断調査では、コーホート追跡の対象となる男性・女性の（コーホート対象外の）配偶者も調査対象としており、これらに対するデータセットが加わるため、データの関係はより複雑である。

³ 縦断調査に特有の問題として、標本の脱落と補充があり（ただし、現在補充は行われていない）、分析におけるデータの関係方法においても、回を重ねるに従いこれらに対する特別の工夫する必要がある。

データセットの作成過程において、調査票に記入された回答を機械入力し、データファイルを作成した後、集計分析用のデータを作成するためにデータ・クリーニングが行われる。データ・クリーニングとは、たとえば設問に用意した選択肢以外の自由回答や、不規則な回答（例として単一選択の設問に対する複数回答）、数値回答の異常値、設問の間での矛盾した回答など、調査の回答をコーディングする際にしばしば遭遇する変則の情報に対して再分類や一定のコーディングを行い、分析可能なデータとすることである。縦断調査のような同一変数を複数回の調査で計測する場合には、比較可能性の確保のため統一的なルールが必要となる。本縦断調査では長期にわたる継続が予定されるため、こうしたルールの詳細な検討と、明確な文書化が必要である。ただし、それらは各設問に依存する課題であるので、本稿ではこれ以上扱わない。

縦断調査では、調査票において設問の最後に自由記述欄（調査に対する感想など）を設けている。自由記述は、対象者に関する重要な追加的情報源であるとともに、とくに縦断調査では対象者と実施者の間の連絡手段としての機能を有するだろう。これらは調査実施上の新たなアイデアやヒントをもたらすことも期待される。しかしながら、機械的なコード化が困難な自由記述は、数が多い場合にはそこから有用な情報を抽出して活用することは難しい。人手によって情報の整理を行うことは、膨大な労力を要するだけでなく、抽出する情報の客観性にも問題が生ずる。その場合、自由記述の情報源としての活用には、処理の省力化・利便性確保とともに、結果の客観性確保のためにもテキストマイニングツールの利用が有効であろう。テキストマイニングツールとは、多量の文書情報の中からキーワードの抽出、類義語の検索、複数の類義語による概念の特定・学習などをもとに、それらのつながり方などによって文章を分類・体系化する機能を持ったソフトウェアである。テキストマイニングの技術は今後も発展して行くものと見られるので、ソフトウェアの選定には、発展性、柔軟性に関する考慮が必要であろう。

本事業におけるデータ管理の課題検討には、横断調査とも共通する調査一般のデータ管理上の課題を含むが、本稿では扱わないこととし、以下、縦断調査に必要なデータセットならびにデータ管理情報の関係の問題を中心に検討を進める。

(2) コード表を用いたデータ情報の一元管理

データ管理の中心となるのは、個票データ本体とコード表である。コード表は原理的には統計処理に必要なデータに関する情報をすべて含めることが可能である。また、できるだけそのようにしてデータ情報の一元管理を行うべきである。データ・クリーニングに用いたルール等も、可能な限りコード表の当該変数の欄から参照出来ることが望ましい。というのは、データは長期に継承され分析に付されるが、新たな調査結果のデータ・クリーニング、集計分析いずれの段階においても、変数の同一性、比較可能性を確保するなどの理由から、過去の調査データ情報について詳細な参照が必要となることが多い。しかし、長期にわたる縦断調査では、必ずしも過去の調査に関して知識のあるユーザがそうした作

業にあたるには限らず、様々な文書、あるいはファイルに散在した情報を参照することは、負担増や誤りの原因となり、そうでなくとも情報の所在の特定などの付加的作業自体が、円滑な作業の流れを妨げる可能性が高い。したがって、情報を一元化し、一括管理することの重要性は高く、調査の回数が増すに当たってそれは必須となると思われる。その際に、データ分析に中心的役割を果たすコード表を中心にデータ情報を集約することが最も有効であると考えられる。本稿では、以下コード表を中心としたデータ管理システムについて検討して行くことにする。

まず、そうしたコード表の形態の要件を考えよう。まず、コード表を記述すべき媒体（ソフトウェア）について考えると、それは汎用的な媒体に記述されなくてはならないだろう。長年にわたる継続的な実施の間に、コンピュータのハード、ソフト技術ともに著しい変化を経験することが予想される。その際に、特定のソフトウェアに依存することは当然ながら望ましくない。できるだけ汎用的で長期に継承されるソフトウェアの媒体を用いる必要がある。最も汎用性が高い情報の保管形式は、テキストファイル形式であろう。現在のソフトウェアのほとんどはテキストファイル形式をデータ交換の暗黙の統一形式としており、その多くはこの形式の直接の入出力に対応しているため、情報の相互変換が将来にわたって保証されている。したがって、情報保管形式としては、テキストファイル形式を採用することが、汎用性の点で最も優れていると言える。実際、調査データの本体はテキストファイル形式によって保管されている（その際、各種ソフトウェアからはできるだけ中間形式を介せず、テキストファイルを直接参照することが望ましい）。しかしながら、コード表に関しては情報間の関係の仕組みを考える上で、テキストファイル形式では不十分である。なぜならば、コード表上の大量の情報を検索したり、一括操作したりする際、そのすべてを手作業で行うことは考えられないが、テキストファイル形式に対しては、表示、編集、その他の操作に対して、支援ツールが多数必要となるだろう。それらのツールはエディタを始め既製のものが多数存在するものの、それらの組み合わせでシステムを構築した場合、個別のツールの維持管理作業が発生し、システム全体の操作性の継承は難しい。また、調査に特化したツールについては新たに作成する必要があるが、アドホックなツールの作成・利用はやはり継承性に問題がある。また、アンケート調査のデータ・情報文書の体系的管理に特化したソフトウェアもすでに存在するが、それは上述の通り汎用性・継承性の要件に反する。

したがって、テキストベースと、既存システムの間位置するような汎用性とある程度の操作性を兼ね備えたソフトウェアの媒体を用いることが考えられる。現在において、そうした要件を満たすのは、Microsoft Excel や、Microsoft Access といった国際的にも標準と見なされるソフトウェアであろう。これらのデータ形式はすでに複数の OS に対応し、また、他の多くのソフトウェアがそれらのファイル形式に対応しているため、長期にわたって汎用性が保たれることが期待される。

コード表の形態に関するその他の主な要件を挙げてみよう。まず、調査データの管理シ

システムにおいて、コード表はマン・マシンインターフェースとしての役割が大きいので、人間が見やすく、また入力や編集などの操作が行いやすいものであることが重要である。また、情報量が多くなることから、検索やフィルター、ソートなど基本的情報操作機能を持つ必要がある。さらに、データ操作や分析の際に、多くの定型的な作業が必要なことから、それを自動化し、あるいは支援するツールが必要であり、コード表の形態はそうしたツールを開発する機構を含むか、少なくともツールとの親和性が高くなければならない。さらに多くのソフトウェアとの連携が必要なことから、外部との情報交換が円滑にできるものでなくてはならない。

以上の点をすべて備えたコード表形態としては、やはり上記 Excel や Access が挙げられるだろう。これらは、情報の操作に対し基本的なデータベース機能とツールを提供している。また、操作の自動化などを行える汎用的なプログラミング環境（VBA）を擁することは、ソフトウェア間の連携のためのツール開発のために有効である。以上から、本事業におけるコード表を中心としたシステム開発においては、Excel ならびに Access ファイルを媒体として検討の対象とすることとした。ただし、本年度の事業において、コード表を中心とした試験的システムの開発を行ったが、その媒体としては表の操作性に優れた Excel ファイルを用いることとした。試験的システムの詳細は次章において説明する。

以上、データ情報の一元管理におけるコード表の役割とその形態の要件について考察した。その結果、最終的には媒体とするソフトウェアの選択を行ったが、長期的な視野に立てば、特定にソフトウェアに依存しないメタシステムについての検討をさらに深めておくことが必要であると考えられる。

3. 縦断調査におけるコード表を中心としたデータ管理・分析システムの開発

(1) エクセル・コード表を中心としたデータ管理・分析システムの概要

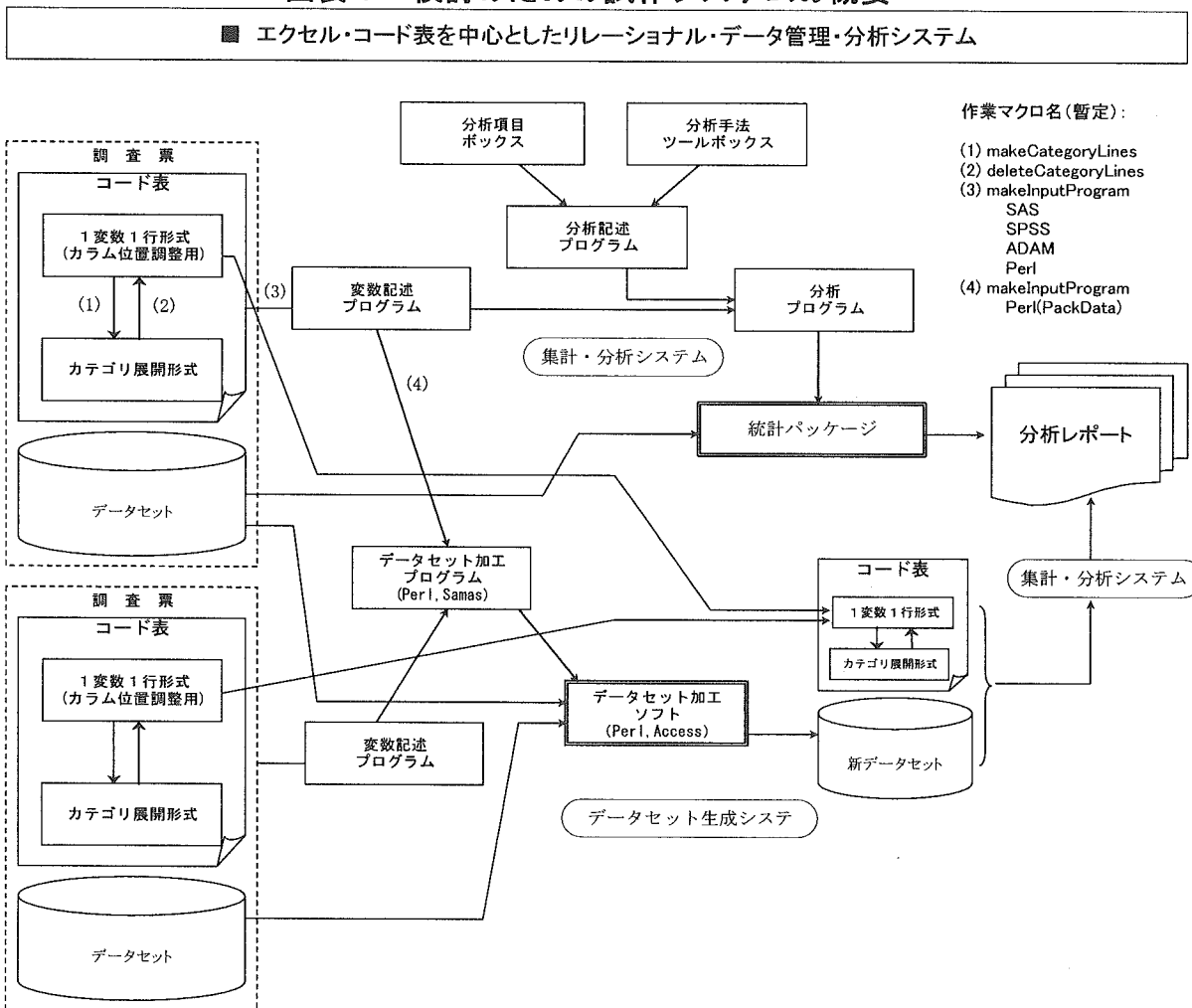
縦断調査におけるデータ管理システムの検討を行うため、エクセル・コード表を中心としたデータ管理・分析システムを試作した。このシステムは、検討されているメタシステムに具体的な形を与えることで、実際的な課題を検討することが主な目的であり、必ずしも完結したシステムではない。したがって、現段階ではそのまま実用の管理システムとして提案するものではない。しかし、今後機能の拡張などにより実用システムに対する開発のベースとしても資することを期待するものである。したがって、その基本的な仕様等は変更される可能性がある。

図表 1 にシステムの概略を示す。図では個票データの情報が集計・分析プログラムおよび実行ソフトウェア等を経て結果報告に至るフローが → によって示されている。その際に各種のツールが介在する。このシステムはユーザがそれら作業のほとんどをコード表から操作、実行することを目指している。言い換えれば、コード表をデータ操作、管理、集計分析のフロントエンドとして位置づけていることが特徴である。

コード表は調査票ごとに、1表が作成される（図では左端の波線内）。ただし、表は使用目的によって二つの表示形式を持つ。一つは「一変数一行形式」であり、他方は「カテゴリ展開形式」である。これらは表示形式が異なるだけであり、作業の都度ツールによって相互変換が可能である（図では(1)(2)の →）。それぞれの詳細については、次節で説明する。

集計分析にあたっては、コード表からはまず集計分析ソフトウェアに対するプログラムの変数記述がツールによって自動生成される。集計分析ソフトウェアは、目的によって複数存在することになるが、基本的に既存のものを想定しており、ここではデータ操作ソフトとして JPERL、集計ソフトとして ADAM、集計・統計分析ソフトとして SAS および SPSS を取り上げている。これらはツールの作成追加によりその他のソフトウェアに対応することが可能である。ユーザはこの変数記述プログラムをベースに目的の集計分析プログラムを完成させ、各ソフトウェアで実行して結果を得ることになる。

図表 1 検討のための試作システムの概要



縦断調査の集計分析においてとくに必要となるのは、複数調査回のデータを同時に用いることであるが、本システムではこれをそれぞれのコード表の統合によって行うこととしている。すなわち、それぞれのコード表から必要な変数の記述のみを取り出して合わせることにより、仮想的な統合データのコード表ができたことになる。このコード表と各個票データセットから、用意したツールによって実際に統合データを作成することも可能である。しかし、一時的な集計分析などのためには、集計分析プログラムを生成の際にオリジナルの各個票データセットから情報を得るように指示することができる。本システムでは、このような仕組みにより、調査票単位に作成されたデータセットを元にして、リレーショナルな集計分析を可能とすることを目指している。

なお、ツール類は基本的に Excel VBA (付属のマクロ言語)によって作成されている。これにより、Excel ワークシート上のコード表を簡便に操作することができる。現段階ではツール類は一つの Excel ブック (ファイル) にまとめてあり (暫定ファイル名「コード表変換.xls」本稿では、ツールブックと呼ぶ)、これらを使う場合には、目的のコード表を表示し、用意されたツールボタン、メニューなどを選択することによって、作業が行われる。Excel 外部のソフトウェアによる作業についても、それらを VBA から制御することによってコード表上から行うことが可能であり、今後この仕組みによる作業範囲の拡張 (たとえば統計分析の適用など) を予定している。

(2) コード表形式とその変換

上述のとおり、本システムのコード表には2つの表示形式を用意している。一つは「一変数一行形式」であり、他方は「カテゴリー展開形式」である。これらは使用目的により便宜的に表示形式を変えているだけであり、必要に応じてツールによる相互変換が可能である。図表2に、これらの例を示している。

「一変数一行形式」は、ワークシート上の1行を一つの変数に割り当て、当該変数に関する情報をその行の各セルに格納する形式である。コンパクトな表示形式であるとともに、全変数 (全行) に対する一括処理が行いやすいという特徴を持つ。一方、「カテゴリー展開形式」は、各変数の各カテゴリーの記述 (コードと内容の対照) を1行ずつに展開した形式であり、カテゴリーの構造が一見してわかりやすく、またカテゴリー記述を要するプログラムの生成を行いやすい (ツールを作成しやすい)。なお、コード表の詳しい仕様は、章末を参照されたい。

図表2 コード表のサンプル

1変数1行形式

第1回成年者縦断調査(女性票)

変数名	項目名	位置	幅	コード範囲	変数値	備考
AF001	調査票番号	1	1	1	女性票	
AF002	KEY番号(世帯情報) 地区番号	2	5	01002-47090	地区番号	
AF003	KEY番号(世帯情報) 単位区番号	7	2	01-14	単位区番号	
AF004	KEY番号(世帯情報) 世帯番号	9	2	01-30	世帯番号	
AF005	KEY番号(世帯情報) 該当者番号	11	1	1-5	該当者番号	
AF006	KEY番号(世帯情報) 配偶者番号	12	1	1-6	1-6)配偶者なし、配偶者の有無不詳	非該当=配偶者なし、配偶者の有無不詳
AF007	出生年月 年	13	2	42-57	42-57)昭和##年 \$)不詳	
AF008	出生年月 月	15	2	01-12	01-12)##月 \$)不詳	
AF009	問1 最終学歴	17	1	1-9	1)中学 2)高校 3)専門学校 4)短大・高専 5)大学 6)大学院 7)その他 \$)不詳 \$)不詳	
AF010	問1 卒業・在学の別	18	1	1-3	1)卒業 2)在学中(休学等を含む) 9)不詳 \$)不詳	
AF011	問2(複数回答) 1通院した	19	1	1,0	1)該当ありの該当なし	
AF012	問2(複数回答) 2入院した	20	1	1,0	1)該当ありの該当なし	
AF013	問2(複数回答) 31,2のことはない	21	1	1,0	1)該当ありの該当なし	
AF014	問2 2入院 入院年月 年	22	2	13,14	13-14)平成##年,99)不詳,\$)不詳、入院していない、入院の有無不詳	
AF015	問2 2入院 入院年月 月	24	2	01-12	01-12)##月,99)不詳,\$)不詳、入院していない、入院の有無不詳	
AF016	問2 2入院 退院年月 年	26	2	13,14	13 14)平成##年,99)不詳,\$)不詳、入院していない、入院の有無不詳	
AF017	問2 2入院 退院年月 月	28	2	01-12	01-12)##月,99)不詳,\$)不詳、入院していない、入院の有無不詳	
AF018	問3就業状況 仕事の有無	30	1	1-5	1)仕事あり 2)仕事ありだが休業中 3)家事に従事しているが通院している \$)不詳 \$)不詳	
AF019	問3就業状況 補1 複数の仕事	31	1	1,2	1)複数の仕事あり 2)複数の仕事なし \$)不詳 \$)仕事についていない \$)仕事の有無不詳	
AF020	問3就業状況 補2 就業形態	32	2	01-09	01)会社などの役員・自営業主 02)自家営業の手伝い 03)自営での賃仕事(内職) 04)正規の職員・従業員 05)アルバイト 06)パート 07)労働者派遣事業所の派遣社員 08)契約社員・嘱託 09)その他 99)不詳 \$)仕事についていない、仕事の有無不詳	

Click for Demo ↓

カテゴリ展開形式コード表作成 Demo

1変数1行形式コード表作成 Demo

複数読み込みプログラム生成 Demo (カテゴリ展開形式でのみ有効)

ボタンがあるのは、サンプルシートのみ。一般のシートでは、メニューから各マクロを指定して実行する。

*配偶者生年はどうするか？

(1) makeCategoryLines

(2) deleteCategoryLines

カテゴリ展開形式

第1回成年者縦断調査(女性票)

変数名	項目名	位置	幅	コード範囲	変数値	備考
AF001	調査票番号	1	1	1	女性票	
AF002	KEY番号(世帯情報) 地区番号	2	5	01002-47090	地区番号	
AF003	KEY番号(世帯情報) 単位区番号	7	2	01-14	単位区番号	
AF004	KEY番号(世帯情報) 世帯番号	9	2	01-30	世帯番号	
AF005	KEY番号(世帯情報) 該当者番号	11	1	1-5	該当者番号	
AF006	KEY番号(世帯情報) 配偶者番号	12	1	1-6	1-6)配偶者なし、配偶者の有無不詳	非該当=配偶者なし、配偶者の有無不詳
				1-6 \$		
AF007	出生年月 年	13	2	42-57	42-57)昭和##年 \$)不詳	配偶者なし、配偶者の有無不詳
AF008	出生年月 月	15	2	01-12	01-12)##月 \$)不詳	
				\$		
AF009	問1 最終学歴	17	1	1-9	1)中学 2)高校 3)専門学校 4)短大・高専 5)大学 6)大学院 7)その他 \$)不詳 \$)不詳	
				1	中学	
				2	高校	
				3	専門学校	
				4	短大・高専	
				5	大学	
				6	大学院	
				7	その他	
				9	不詳	
				\$	不詳	
AF010	問1 卒業・在学の別	18	1	1-3	1)卒業 2)在学中(休学等を含む) 9)不詳 \$)不詳	
				1	卒業	
				2	在学中(休学等を含む)	
				9	不詳	
				\$	不詳	
AF011	問2(複数回答) 1通院した	19	1	1,0	1)該当ありの該当なし	
				1	該当あり	
				0	該当なし	
AF012	問2(複数回答) 2入院した	20	1	1,0	1)該当ありの該当なし	
				1	該当あり	
				0	該当なし	
AF013	問2(複数回答) 31,2のことはない	21	1	1,0	1)該当ありの該当なし	
				1	該当あり	
				0	該当なし	
AF014	問2 2入院 入院年月 年	22	2	13,14	13-14)平成##年,99)不詳,\$)不詳、入院していない、入院の有無不詳	
				99	不詳	
				\$	不詳、入院していない、入院の有無不詳	

Click for Demo ↓

カテゴリ展開形式コード表作成 Demo

1変数1行形式コード表作成 Demo

複数読み込みプログラム生成 Demo (カテゴリ展開形式でのみ有効)

すでにのべたように、両表示形式はツールによって相互変換が可能である。コード表の構築段階においては、手入力、修正、全変数記述に対する一括操作、カラム位置・幅等の数式による調整などの作業の観点から、一変数一行形式が優れていると思われる。一旦構築が終わり、集計分析段階になった場合には、カテゴリー展開形式が見やすさ、ツールの応用のしやすさなどの点で優れている。したがって、一旦コード表が確定して後は、集計分析のユーザはカテゴリー展開形式のみを用いることが望ましい。というのは、両形式を同時に持つと情報の重複となり、片方のみへの修正などが行われた場合、情報の齟齬が生ずるので、マスターを決め、他の形式は常にこれを変換して用いることが必要となるからである。しかし、少なくともデータクリーニングなどの確定作業の間は、その両方の段階を行き来する必要があるため、その都度変換によって情報を最新に保つ必要があるだろう。

形式相互の変換には VBA によるツール (1) `makeCategoryLines` (一変数一行形式 → カテゴリー展開形式)、(2) `deleteCategoryLines` (カテゴリー展開形式 → 一変数一行形式) が用意されている。これらは目的のコード表を表示し、ツールボタンをクリックすることで変換された新たなコード表が生成追加される仕組みとなっている (情報を一元化するため、変換後元のコード表は自動的に削除されることが望ましいが、現段階では保守のため自動削除は行わない)。

なお、これらとは別に、従来統計情報部において横断調査等に用いられてきたコード表 (近年はエクセル形式である) を本システムで用いるコード表の形式に変換するツールを用意しており、従来 (型) のコード表の継承ないし、本システムの従来調査への応用が可能となっている。

(2) 変数記述プログラムの生成

ツールによって、コード表 (カテゴリー展開形式) から、集計分析などに用いる各種のプログラムを自動生成することができる。以下、集計分析ソフトウェア用の変数記述、ならびに頻度分布表、ヒストグラムなどを作成するプログラムを例に説明する。

まず、プログラム自動生成に必須な情報 (データの所在、目的のソフトウェアなど)、およびオプション (プログラムの保存先など) をツールブック内の「ツール設定」シート上に指定する (図表 3)。

次にコード表 (カテゴリー展開形式) を表示して、プログラム生成のツールボタンをクリックすると、VBA マクロ (3) `makeInputProgram` が実行され、コード表の内容が目的のソフトウェアのプログラムに変換される。

本試験システムでは、SAS (頻度分布 / ヒストグラム)、SPSS (頻度分布 / ヒストグラム)、ADAM (クロス集計用変数記述)、JPerl (データ操作変数記述) の各プログラムを自動生成することができる。図表 4 には、この様子が示されている (コード表 → プログラム (複数ソフト) → 頻度分布表 (SAS))。

図表3 ツールブック内「ツール設定」シート

■ 統計ソフト・データ読み込みプログラムの自動生成マクロ

① 下記、設定項目を設定（必須は(1)～(3)）。

② コード表シート（AF_女性票など）のボタン「プログラム生成 makeInputProgram」をクリックすると実行される。

→ プログラム・シートが追加され、ファイル保存の指定画面となる。
（保存を[キャンセル]すると、シート追加のみで終了）

Note :

- ・すでにプログラムシートが存在する場合は、ユーザに確認の上で更新される。
- ・“AF_女性票”のみ、集計サンプル・プログラム付加（保存したファイルはそのまま実行可）。

設定項目	設定値
(1) プログラム保存先フォルダ (default)	C:\¥Work
(2) データフォルダ	F:\¥dataent¥10
(3) パッケージ	SAS
(4) 作成者	Lexis
(5) タイトル行長さ (>=80)	100
(6) インデント	8
(7) 文字サイズ	10
(8) フォント	MS ゴシック

図表4 コード表（カテゴリ展開形式）からのプログラム自動生成

変数名	変数名	変数値	変数値	変数値	変数値	変数値	変数値
AF001	調査員番号	1	1	1	女性票		
AF002	KEY番号(世帯情報) 地区番号	2	5	01002-47090	地区番号		
AF003	KEY番号(世帯情報) 単位区番号	7	2	01-14	単位区番号		
AF004	KEY番号(世帯情報) 世帯番号	9	2	01-30	世帯番号		
AF005	KEY番号(世帯情報) 該当者番号	11	1	1-5	該当者番号		
AF006	KEY番号(世帯情報) 配偶者番号	12	1	1-6	1-6)M,\$)配偶者なし、配偶者の有無不詳	非該当=配偶者なし、配偶者の有無不詳	
				1-6	配偶者なし、配偶者の有無不詳		
AF007	出生年月 年	13	2	42-57	42-57)昭和##年,\$)不詳		
				\$	不詳		
AF008	出生年月 月	15	2	01-12	01-12)##月,\$)不詳		
				\$	不詳		
AF009	問1 最終学歴	17	1	1-9	1)中学,2)高校,3)専門学校,4)短大・高専,5)大学,6)大学院,7)その他,\$)不詳		
				1	中学		
				2	高校		
				3	専門学校		
				4	短大・高専		
				5	大学		
				6	大学院		
				7	その他		
				9	不詳		
				\$	不詳		
AF010	問1 卒業・在学の別	18	1	1-3	1)卒業,2)在学中(休学等を含む),9)不詳,\$)不詳		
				1	卒業		
				2	在学中(休学等を含む)		
				9	不詳		
				\$	不詳		
AF011	問2(複数回答) 1通院した	19	1	1,0	1)該当ありの該当なし		
				1	該当あり		
				0	該当なし		
AF012	問2(複数回答) 2入院した	20	1	1,0	1)該当ありの該当なし		
				1	該当あり		
				0	該当なし		
AF013	問2(複数回答) 31.2のことばない	21	1	1,0	1)該当ありの該当なし		
				1	該当あり		
				0	該当なし		
AF014	問2 2入院 入院年月 年	22	2	13,14	13-14)平成##年,99)不詳,\$)不詳、入院していない、入院の有無不詳		
				99	不詳		
				\$	不詳、入院していない、入院の有無不詳		

Click for Demo ↓

カテゴリ展開形式コード表作成 Demo

1変数1行形式コード表作成 Demo

複数読み込みプログラム生成 Demo
 (カテゴリ展開形式でのみ有効)

(3) makeInputProgram

自動生成された SAS プログラム (一部)

```

/*****
*** SAS :
*** 第1回成年者縦断調査(女性票)
*** [コード表ツール1.xls(コード表例 AF_女性票)]
*** <Lexis> 2004/09/27
*****/
filename INF 'D:\Project\RWPanel\Data\sample\adult\data1' LRECL=300;
option device=WIN linesize=256 pagesize=9999 nocenter;

data AF01; infile INF; input
AF001 1 /*| 調査票番号
AF002 2 - 6 /*| KEY番号(世帯情報) 地区番号
AF003 7 - 8 /*| KEY番号(世帯情報) 単位区番号
AF004 9 - 10 /*| KEY番号(世帯情報) 世帯番号
AF005 11 /*| KEY番号(世帯情報) 該当者番号
AF006 12 /*| KEY番号(世帯情報) 配偶者番号
AF007 13 - 14 /*| 出生年月 年
AF008 15 - 16 /*| 出生年月 月
AF009 17 /*| 問1 最終学歴
. . . . .
AF171 238 /*| メールアドレスの記入有無
AF172 239 /*| 除外
;

label
AF001 = '調査票番号'
AF002 = 'KEY番号(世帯情報) 地区番号'
AF003 = 'KEY番号(世帯情報) 単位区番号'
AF004 = 'KEY番号(世帯情報) 世帯番号'
AF005 = 'KEY番号(世帯情報) 該当者番号'
AF006 = 'KEY番号(世帯情報) 配偶者番号'
AF007 = '出生年月 年'
AF008 = '出生年月 月'
AF009 = '問1 最終学歴'
. . . . .
AF171 = 'メールアドレスの記入有無'
AF172 = '除外'
;

run;
proc format;
value AF001f;
value AF002f;
value AF003f;
value AF004f;
value AF005f;
value AF006f
1-6 = '配偶者なし、配偶者の有無不詳';
value AF007f
= '不詳';
value AF008f
= '不詳';
value AF009f
1 = '1 中学'
2 = '2 高校'
3 = '3 専門学校'
4 = '4 短大・高専'
5 = '5 大学'
6 = '6 大学院'
7 = '7 その他'
9 = '9 不詳';
. = '不詳';
. . . . .
value AF171f
1 = '1 メールアドレスへの記入あり'
= 'メールアドレスへの記入なし';
value AF172f
1 = '1 質問項目未記入による集計対象外'
2 = '2 年齢はずれ又は生年月未記入で配偶者がいないため集計対象外'
0 = '0 ペア間の矛盾等による集計対象外'
= '集計対象';
run;

/****[集計]*****/
ODS HTML file='listing.xls';
/*-----*/
data SEL_DATA;
set AF01;
/*-----<<対象標本の限定>>-----*/
if AF034 = 1 then DELETE;

/*| 数値変数の作成 |*/
if AF069 = 99 then cex = .; else cex = AF069;
if cex >= 5 then cex = 5;
run;
/*-----*/
proc tabulate noseps format=22.17;
class
AF009
AF018
AF069
/missing mlf;
format
AF009 af009f.
AF018 af018f.
;
var
cex
;
table (all=総数 AF018='')*(all=総数 AF009='),
(all=総数 *n= AF069= *n= cex= *mean= 平均値) /rts=40;
footnote 'これはHTMLファイルです。Excelファイルにするには「名前を付けて保存」でファイルの
種類を「Excelブック(*.xls)」で保存します。';
run;
/*-----*/
ODS HTML close;

```


自動生成された SPSS プログラム (一部)

```

/*****
/**** SPSS :
/**** 第1回成年者縦断調査 (女性票)
/**** [コード表ツール1.xls (コード表例 AF_女性票)]
/**** <Lexis> 2004/09/27
/*****

DATA LIST FILE='D:\ProjectR\Panel\Data\sample\adult\data1' /
AF001 1 /* 調査票番号
AF002 2 - 6 /* KEY番号 (世帯情報) 地区番号
AF003 7 - 8 /* KEY番号 (世帯情報) 単位区番号
AF004 9 - 10 /* KEY番号 (世帯情報) 世帯番号
AF005 11 /* KEY番号 (世帯情報) 該当者番号
AF006 12 /* KEY番号 (世帯情報) 配偶者番号
AF007 13 - 14 /* 出生年月 年
AF008 15 - 16 /* 出生年月 月
AF009 17 /* 問1 最終学歴

. . . . .

AF171 238 /* メールアドレスの記入有無
AF172 239 /* 除外

CACHE.
EXECUTE.
VARIABLE LABELS
AF001 '調査票番号'
AF002 'KEY番号 (世帯情報) 地区番号'
AF003 'KEY番号 (世帯情報) 単位区番号'
AF004 'KEY番号 (世帯情報) 世帯番号'
AF005 'KEY番号 (世帯情報) 該当者番号'
AF006 'KEY番号 (世帯情報) 配偶者番号'
AF007 '出生年月 年'
AF008 '出生年月 月'
AF009 '問1 最終学歴'

. . . . .

AF171 'メールアドレスの記入有無'
AF172 '除外'

EXECUTE.

VALUE LABELS AF001
VALUE LABELS AF002
VALUE LABELS AF003
VALUE LABELS AF004
VALUE LABELS AF005
VALUE LABELS AF006
1-6 ''
VALUE LABELS AF007
VALUE LABELS AF008
VALUE LABELS AF009
1 '中学'
2 '高校'
3 '専門学校'
4 '短大・高専'
5 '大学'
6 '大学院'
7 'その他'
9 '不詳'

. . . . .

VALUE LABELS AF171
1 'メールアドレスへの記入あり'

VALUE LABELS AF172
1 '質問項目未記入による集計対象外'
2 '年齢はずれ又は生年月未記入で配偶者がいないため集計対象外'
0 'ペア間の矛盾等による集計対象外'

EXECUTE.

/**** [集計] *****/
CROSSTABS
/TABLES=AF009 BY AF069 BY AF018
/FORMAT=AVALUE TABLES
/STATISTIC=CHISQ CORR
/CELLS= COUNT ROW COLUMN TOTAL

```

自動生成された ADAM プログラム (一部)

```

*****
*** ADAM:
*** 第1回成年者縦断調査(女性票)
*** [コード表ツール1.xls(コード表例 AF_女性票)]
*** <Lexis>
*** 2004/09/27
*****
;:AR
//FDC
FILE NFS11C.T.300.D:\ProjectR\Panel\data\sample\adult\data\
REG MAIN.300

-----
ITEM AF001 , N1/1 ; 調査票番号
LABEL 総数[
-----
ITEM AF002 , N5/2 ; KEY番号(世帯情報) 地区番号
LABEL 総数[
-----
ITEM AF003 , N2/7 ; KEY番号(世帯情報) 単位区番号
LABEL 総数[
-----
ITEM AF004 , N2/9 ; KEY番号(世帯情報) 世帯番号
LABEL 総数[
-----
ITEM AF005 , N1/11 ; KEY番号(世帯情報) 該当者番号
LABEL 総数[
-----
ITEM AF006 , N1/12 ; KEY番号(世帯情報) 配偶者番号
LABEL 総数[
CODE "配偶者なし、配偶者の有無不詳"
CODE "1-0, 配偶者なし、配偶者の有無不詳"=other
-----
ITEM AF007 , N2/13 ; 出生年月 年
LABEL 総数[
CODE "不詳"
CODE "不詳"="不詳"=other
-----
ITEM AF008 , N2/15 ; 出生年月 月
LABEL 総数[
CODE "不詳"
CODE "不詳"="不詳"=other
-----
ITEM AF009 , N1/17 ; 問1 最終学歴
LABEL 総数[
"中学",
"高校",
"専門学校",
"短大、高専",
"大学",
"大学院",
"その他",
"不詳",
"不詳"]
CODE "中学"=1, 2, 3, 4, 5, 6, 7, 9, "不詳"=other
-----
. . . . .

-----
ITEM AF171 , N1/238 ; メールアドレスの記入有無
LABEL 総数[
"メールアドレスへの記入あり",
"メールアドレスへの記入なし"]
CODE "メールアドレスへの記入あり"=1, "メールアドレスへの記入なし"=other
-----
ITEM AF172 , N1/239 ; 除外
LABEL 総数[
"質問項目未記入による集計対象外",
"年齢はずれ又は生年月未記入で配偶者がいないため集計対象外",
"ペア間の矛盾等による集計対象外",
"集計対象"]
CODE "質問項目未記入による集計対象外"=1, 2, 0, "集計対象"=other

END

;===[ 集計 ]=====
//CTC
IF AF034=1
TABLE T1.1. AF069. AF018 * AF009
ENDIF
-----
//TCC
trace on
SHEET 132,132
-----
FORM T1.C16.0
TITLE T1."問1 最終学歴・問9 希望子ども数・問3 就業状況 仕事の有無、加減"
FOOT T1."注：サンプル集計"
LINE T1.2
LBLOPT T1.C.M/4//"/
SPAN T1.1_20/3_22_23_128

COPY TR1=T1
FORM TR1.C8.2
RATE TR1.C.100
TITLE TR1."問1 最終学歴・問9 希望子ども数・問3 就業状況 仕事の有無、加減(割合)"
FOOT TR1."注：サンプル集計"
LINE TR1.2
LBLOPT TR1.C.M/4//"/
SPAN TR1.1_20/3_22_23_128

-----
PRINT T1
PRINT TR1
to123 T1,TR1

```

自動生成された JPerl プログラム (一部)

```

=====
#== P e r l : ==
#== 第 1 回成年者縦断調査 (女性票) ==
#== [ コード表ツール 1.xls (コード表例 AF_女性票) ] ==
#== <Lexis> == 2004/09/27 ==
=====
$|=1;
open(FIN, "<D:¥¥ProjectR¥¥Panel¥¥Data¥¥sample¥¥adult¥¥data1"); # 入力データファイル名
open(FOUT, ">data1.rst"); # 出力データファイル名

while(<FIN>){
    $AF001 = substr($_, 1, 1); # 調査票番号
    $AF002 = substr($_, 2, 5); # KEY番号 (世帯情報) 地区番号
    $AF003 = substr($_, 7, 2); # KEY番号 (世帯情報) 単位区番号
    $AF004 = substr($_, 9, 2); # KEY番号 (世帯情報) 世帯番号
    $AF005 = substr($_, 11, 1); # KEY番号 (世帯情報) 該当者番号
    $AF006 = substr($_, 12, 1); # KEY番号 (世帯情報) 配偶者番号
    $AF007 = substr($_, 13, 2); # 出生年月 年
    $AF008 = substr($_, 15, 2); # 出生年月 月
    $AF009 = substr($_, 17, 1); # 問 1 最終学歴

    . . . . .

    $AF171 = substr($_, 238, 1); # メールアドレスの記入有無
    $AF172 = substr($_, 239, 1); # 除外
}

    substr($_, 6, 4) = " "; # 単位区番号、標本番号をマスク
    print $_, "\n";
}
close(FIN);
close(FOUT);

```

自動生成された SAS プログラムによる頻度分布表（一部）

コード分布表：BK02

FREQ プロシジャ

調査年				
B001	度数	パーセント	累積 度数	累積 パーセント
	2	47015	100	47015
				100

調査番号				
B002	度数	パーセント	累積 度数	累積 パーセント
	17	47015	100	47015
				100

認識番号				
B003	度数	パーセント	累積 度数	累積 パーセント
100001 100001-727034		47015	100	47015
				100

OCR入力番号（入力機器）				
B004	度数	パーセント	累積 度数	累積 パーセント
space	23592	50.18	23592	50.18
1 1	11893	25.3	35485	75.48
2 2	11527	24.52	47012	99.99
9 9	3	0.01	47015	100

OCR入力番号（種別）				
B005	度数	パーセント	累積 度数	累積 パーセント
	1	47015	100	47015
				100

OCR入力番号（処理月）				
B006	度数	パーセント	累積 度数	累積 パーセント
01 01-03月	23423	49.82	23423	49.82
07 07-09月	23592	50.18	47015	100

OCR入力番号（処理県）				
B007	度数	パーセント	累積 度数	累積 パーセント
1 北海道	1839	3.91	1839	3.91
2 青森	561	1.19	2400	5.1
3 岩手	512	1.09	2912	6.19
4 宮城	846	1.8	3758	7.99
5 秋田	432	0.92	4190	8.91
6 山形	473	1.01	4663	9.92
7 福島	818	1.74	5481	11.66
8 茨城	1140	2.42	6621	14.08
9 栃木	744	1.58	7365	15.67
10 群馬	772	1.64	8137	17.31
11 埼玉	2534	5.39	10671	22.7
12 千葉	2049	4.36	12720	27.06
13 東京	3780	8.04	16500	35.1
14 神奈川	3217	6.84	19717	41.94
15 新潟	941	2	20658	43.94
16 富山	439	0.93	21097	44.87
17 石川	503	1.07	21600	45.94
18 福井	326	0.69	21926	46.64
19 山梨	347	0.74	22273	47.37
20 長野	950	2.02	23223	49.39
21 岐阜	860	1.83	24083	51.22
22 静岡	1392	2.96	25475	54.18
23 愛知	2956	6.29	28431	60.47
24 三重	681	1.45	29112	61.92
25 滋賀	557	1.18	29669	63.11
26 京都	961	2.04	30630	65.15
27 大阪	3377	7.18	34007	72.33
28 兵庫	2066	4.39	36073	76.73
29 奈良	532	1.13	36605	77.86
30 和歌山	372	0.79	36977	78.65
31 鳥取	248	0.53	37225	79.18
32 島根	264	0.56	37489	79.74
33 岡山	758	1.61	38247	81.35
34 広島	1087	2.31	39334	83.66
35 山口	547	1.16	39881	84.83
36 徳島	261	0.56	40142	85.38
37 香川	361	0.77	40503	86.15
38 愛媛	572	1.22	41075	87.37
39 高知	279	0.59	41354	87.96
40 福岡	1887	4.01	43241	91.97
41 佐賀	353	0.75	43594	92.72
42 長崎	573	1.22	44167	93.94
43 熊本	714	1.52	44881	95.46
44 大分	476	1.01	45357	96.47
45 宮崎	434	0.92	45791	97.4
46 鹿児島	663	1.41	46454	98.81
47 沖縄	561	1.19	47015	100