



社会調査のデータ分析について

—カテゴリー的データを中心に—

小野能文

社会調査、とくに大勢の人々を調査対象者とするアンケート調査においては、綿密な調査の計画に基づいた調査票の作成と実施、および調査によって得られたデータの精密な分析が大切であるが、アンケート調査で得られたデータは、年収や満年齢などの連続変量データよりも、性別データや年齢層データ、あるいは5段階評価データなどのカテゴリー的データ (categorical data) が多い。そのためカテゴリー的データの分析法について詳細に考察した。

説明変数がカテゴリー的データの分析を行う数量化理論や要因がすべてカテゴリー的データの分析法である CATDAP (Categorical Data Analysis Program Package) およびクロス集計表におけるカテゴリー的データの分析法である AIC プログラムなどのカテゴリー的データ分析の方法について主に論述した。

CATDAP は数多くの説明変数と目的変数間の関連性について、精密に分析ができる多変量解析法であり、カテゴリー的データが多く含まれる社会調査のデータ分析に有効性が高いことを論じ、実際の調査データに適用して詳細に検討した。

人々の意識や社会現象における要因間の相関関係の分析は大切であるが、因果関係の分析はさらに重要であるので、パス解析や共分散構造分析などの因果関係の分析法について、実際の調査データに適用して考察した。

キーワード：社会調査, カテゴリー的データ (カテゴリカル・データ), CATDAP

1 序

社会調査は、社会生活における人々の意識や生活の実態および社会現象を主に現地調査によって、科学的にデータを収集し、整理し、分析する活動である。

人々の意識や考えを調べるために、調査対象の大勢の人々を個別訪問して、面と向かって質問して調べる個別面接調査や、調査票を配布し、記入のために留め置いた後に回収して回る配票調査、あるいは郵便を利用して調査票を送り、記入後に返送してもらう郵送調査などの調査法が用いられることが多い。

社会現象は、2人以上の人間に関係のある人為的現象や、社会の制度や伝統に関係のある現象であるといえるが、非常に種類が多く、頻繁におこる。たとえば、人々や集団の間の協力や協同、模倣、

競争あるいは指導は社会現象であるが、対立や紛争も社会現象であり、犯罪や非行、いじめ、児童虐待も社会現象である。また、結婚は社会現象であるが、結婚に関する流行やトレンドも社会現象である。

地震は自然現象であって社会現象ではないが、大きな地震の際の人々の被害や人間が作った建物や道路などの被害は社会現象である。また大地震の時にデマが流され、人々がそれに影響されることによって人間に対する被害が出れば、そのような被害は社会現象である。もちろんデマが流されることも社会現象である。

人々の意識や社会現象には様々な要因が複雑に作用しているので、人々の意識や社会現象の解明のために、社会調査を行って様々な要因の各々の効果や様々な要因間の関係を正確かつ精密に分析する必要性が大きい。さらに、因果関係の解明のためには、より厳密な分析が要求され、理論的分析モデルも必要になる。

人々の主観的な意識や考えや好み等を質問調査によって調べると、得られるデータは、カテゴリー的データが非常に多い。カテゴリーは範疇や種類あるいは類別という意味があり、カテゴリー的データには次のようなものがある。

①賛否（賛成、反対）。ある法案や提案に関して、賛成というカテゴリーと反対というカテゴリーのどちらか1つを選んでもらう二者択一式で質問すると、賛成の人が何割、反対の人が何割というような結果が得られる。しかし、二者択一でなく、5段階尺度（非常に賛成、賛成、どちらともいえない、反対、非常に反対）等で質問すると、名義尺度データでなく順序尺度データが得られる。

②くらし方のタイプ（趣味にあったくらし方、金持ちになる、名をあげる、清く正しくくらす、のんきにクヨクヨしないでくらす、社会につくす等）の場合、それぞれのくらし方の人が何%いるか等のデータが得られる。

③階層帰属意識の調査では人々が帰属意識を持つ階層（上の階層、中の上階層、中の中階層、中の下階層、下の階層など）はそれぞれどの程度の割合であるか等のデータが得られる。

④支持政党（自民党、民主党、公明党、社民党、共産党など）の場合、それぞれの政党を支持する人の比率等のデータが得られる。

調査対象者の属性に関して、⑤性別（男性、女性）、⑥年齢層（20歳未満、20代、30代、40代、50代、60歳～64歳、65歳以上）、⑦職業（事務職、販売職、サービス職、管理職、専門職、生産工、農耕作業員など）、⑧学歴（大学院卒・修了、大学卒、短大卒・高専卒、専門学校卒、高校卒、中学卒、旧制小学校卒）、⑨配偶関係（未婚、有配偶、死別・離別）、⑩雇用形態（役員、正規の職員・従業員、パート、アルバイト、労働者派遣事業所の派遣社員、契約社員・嘱託など）のような類別変数データが得られる。

さらに、大勢の人々を対象とする量的調査の自由記述回答においては質的データ（回答者の個人的、主観的な生の意見など）が得られるが、その質的データを何らかの観点や基準に基づいて分類し、カテゴリー分けを行うと、カテゴリー的データや類別変数データが得られる。

ところが、このような社会調査によって得られるカテゴリー的データや類別変数データは、年収（税込み年収額など）や預貯金額のような連続変数の量的データと違って、分析するために連続変数の量的データと違った分析法が必要になる。

2 カテゴリー的データの分析

量的データである年収（税込み年収額など）、預貯金額、利用回数、教育年数、満年齢などの場合、調査によって得られたデータの平均値、中央値、標準偏差といった統計数値が正確に計算でき、他の連続変量の量的データとの関連を相関係数などの統計学的分析をすることができる。さらに、3個以上の変数間の関係を同時に統計学的に分析する多変量解析法で分析することができる。量的データに関しては重回帰分析や主成分分析などの多変量解析法がよく用いられている。

しかしカテゴリー的データは、各カテゴリーに1や2といった数字を割り当てても、くらし方のタイプのような名義尺度データの場合、数字の差は意味を持たない。そのため、カテゴリー的データが含まれる調査データを精密に分析するために工夫された多変量解析法がいくつも開発されて来ている。

多変量解析法は、分析において基準変数や目的変数があるかないか、換言すれば外的基準があるかないか、で分類することができるが、外的基準がある場合に適用できる多変量解析法として、林知己夫が開発・発展させた数量化理論がよく知られ、広く使われている。その中の数量化Ⅱ類は、説明変数も目的変数も共にカテゴリー的データや質的変数の場合に適した多変量解析法である⁽¹⁾。

数量化Ⅱ類は、複数の群に区分される外的基準との間の相関比（ η^2 ）を最大にするように全説明変数（要因アイテム）の全カテゴリーの数量を計算し、さらに各説明変数（要因アイテム）ごとのレンジ（最大カテゴリー数量－最小カテゴリー数量）を計算し、レンジの大きな説明変数ほど、外的基準に対する関連性が大きいとみなす。また、数量化した外的基準と数量化した各説明変数（要因アイテム）との間の偏相関係数を計算し、偏相関係数の値が高いほど外的基準との関連性が大きいとみなす。

相関比の値はマイナス1からプラス1の間の値であるので、相関比の値が0に近すぎると、解が群の判別に関して適切ではないことになる。その場合には、説明変数を追加したり、群の再区分などを行って、分析をやり直す必要がある。

このような数量化Ⅱ類は、カテゴリー的データで得られた生活満足度や支持する政党などに対してどのような要因が影響しているかといったことの解明に有効な分析法である。

目的変数（外的基準）が数量変数で、説明変数がカテゴリー変数の場合は数量化Ⅰ類が分析に適している。

カテゴリー変数を取り扱い、目的変数がない場合、換言すれば外的基準がない場合には、数量化Ⅲ類が適している。

数量化Ⅲ類は個体の様々なカテゴリー（変数）への反応に基づいて、個体とカテゴリーの両方を数量化し、座標平面で、類似したカテゴリーが近くに、似ていないカテゴリーが遠く離れて位置するように示す。

3 CATDAP

説明変数も目的変数も共にカテゴリー的データを精密に分析するために工夫された多変量解析法として、CATDAP（キャットダップ）という分析法がある。

CATDAP (Categorical Data Analysis Program Package) は、カテゴリー的データである目的変数に対する、説明変数の関連性の度合をAICすなわち赤池情報量規準という統計量を用いて解析する多変量解析法である。この方法は坂元慶行らが開発した非常に有効な多変量解析のコンピュータプログラムである。FORTRAN プログラムがCATDAP -01とCATDAP -02であり (2)、BASIC プログラムがCATDAP Jである (3)。

実行プログラムがCATDAPXP.EXEである。

AICの値は、目的変数に対する説明変数の情報の量を表わし、AICの値が小さい程、情報量が大きいと見なされ、説明変数の有効度が大きいと考えられる。また、AICのマイナスの値は、説明変数と目的変数の間の関係が独立ではないということを示している。

AICの値は、次のような公式で計算される (4)。

$$AIC = (-2) \text{ 最大対数尤度} + 2 \text{ (自由なパラメータ数)}$$

AICは、パラメータ数の異なるモデルが競合している時、そのモデルを用いて予測をする場合に期待される誤差が最小になるようなモデルを選ぶことをめざした統計量である (5)。

CATDAPは、調査で得られたデータをコンピュータのディスクファイルから入力して目的変数と全説明変数との間のあらゆるクロス表を作って、AICの値を計算し、AICの値の小さい順に、すなわち有効度の大きい順に説明変数をリストしたり、説明変数の最適組み合わせを検出する分析法である。

この方法では説明変数の数の制限はないので、調査した全変数を分析にかけることが出来て、目的変数に対して関連性のある変数とそうでない変数を選び分けることが出来る。説明変数の中に連続変数 (たとえば収入額や年齢など) が含まれていても、CATDAP -02 では、AICの値が最小となるように自動的にカテゴリー区分をする。

ところで、2つの項目や変数の間の独立性の検定を行う χ^2 検定の場合、説明変数のカテゴリーの数が一定でないので、そのままの形では、どの説明変数が目的変数に対する有効度が大きいかがわかりにくい。また、データ件数が多くなると、 χ^2 検定では、たいがいの変数が独立でなくなりやすくなる。

χ^2 検定では、クロス表 (分割表) での各セルの期待度数が最低でも1以上いるという制限があるが、CATDAPはそのような制限がない。また、同じくカテゴリー的データや名義尺度データの解析に適した数量化Ⅱ類では、該当するサンプルが少なすぎるカテゴリーがあると、分析結果の信頼性が弱くなるという弱点があるが、CATDAPにはそのような弱点がない。さらにCATDAPでは、カテゴリーのリコード (recode) 機能があり、該当サンプル数の少ないカテゴリーを除外したりカテゴリーを統合することが出来る。たとえば、該当サンプルが少なすぎるカテゴリーを「その他」というカテゴリーに含めてしまうことが簡単に出来る。

さらに、相互の関連性が小さくない説明変数が含まれていると、分析結果の正確性が弱くなるという欠点が数量化Ⅱ類にあるが、CATDAPにはそのような欠点がない。

また、2つの説明変数のカテゴリーの間に交互作用があると、分析結果の正確性が弱くなり、非

常に強い交互作用があると計算出来なくなるという欠点が数量化Ⅱ類にあるが、CATDAPにはそのような欠点がない。

CATDAPは以上のような特徴や長所を備えているが、現代医療と医師に対する人々の意識調査の調査データに関してCATDAPを適用した⁽⁶⁾。

調査対象は、地方都市圏の1つとしての広島市東部の郊外地域である安芸区内の有権者である。この地域の選択理由は、鉄道交通は便利であるが道路整備が不十分であり、人口密度が高くなく、高層化も進んでおらず、農地が残り緑が豊かで伝統的意識が残存と思われる地域において、医療や医師に対する信頼がどの程度のものであるかを検証するためである。

調査法は配票調査法（留め置き調査法）である。

サンプリングは、選挙人名簿から無作為抽出法によって約6万人弱の有権者の中から調査対象者を抽出した。安芸区内から、瀬野地区、中野地区および畑賀地区を選び、全有権者の中から系統抽出法によって500人を抽出した。

調査期間は2004年8月1日から9月30日までである。

有効回収数は257であり、有効回収率は51.4%である。

主に利用する医療機関の医師に対する評価はどのような要因と大きく関連しているか、ということとを、CATDAPを用いて分析したところ、表1のような結果が得られた。最も大きく関連する要因は、診療における医師の説明の評価であった。AICの値が-59.57と最もマイナスの値が大きく、最も関連性が大きい。2番目は大きな病気やけがの有無であり、3番目はかかりつけの医師の有無、4番目は診てもらっている理由である。（CATDAPによる分析では欠損値を除いて242のデータで分析した）

表1 主に利用する医療機関の医師の評価の説明変数順位

- CATDAPによる -

順位	説明変数	カテゴリー数	A I C	自由度
1	診療における説明	5	-59.57	12
2	大きな病気やけがの有無	5	-14.39	12
3	かかりつけの医師の有無	3	-13.93	6
4	診てもらっている理由	9	-8.67	24
5	年齢層	6	1.13	15
6	性別	2	4.51	3
7	就労形態	5	11.92	12
8	医師の職業威信の評価	4	12.14	9
9	診療における安心	7	18.34	18
10	病院のイメージ	8	19.05	21
11	職業	12	36.50	33

主に利用する医療機関の医師の評価と診療における説明の評価との関連をクロス表で詳細にみると表2で示されるように、医師の診療における説明が「十分である」と思っている人（全体の48.6%）は、その80.3%が医師を「良い」と評価し、医師に対する評価が高い。説明が「非常に十分である」と思っている人（全体の8.4%）は、医師を「良い」と評価している人が90.4%もいる。

これは、人々の医療についての知識が増し、患者の権利意識の増大を背景に、「インフォームドコンセント（十分な説明と同意）」が浸透し、病状や病名などについて十分な説明を求める意識が強くなってきていることが影響しているものと考えられる。説明が十分な医師は、患者に対して、誠実で親切であり、患者を見下したりしていない医師であるので、信頼され評価が高くなるのではないかと、と思われる。

表2 医師の診療における説明に対する評価と主に利用する医療機関の医師の評価

医師の診療における説明に対する評価	主に利用する医療機関の医師の評価				合計
	非常に良い	良い	どちらともいえない	良くない	
非常に十分である	15 71.4%	4 19.0%	1 4.8%	1 4.8%	21 100.0%
十分である	13 10.7%	85 69.7%	24 19.7%	0 0.0%	122 100.0%
どちらともいえない	2 2.5%	37 46.8%	40 50.6%	0 0.0%	79 100.0%
少し不十分である	0 0.0%	9 33.3%	18 66.7%	0 0.0%	27 100.0%
非常に不十分である	0 0.0%	0 0.0%	2 100.0%	0 0.0%	2 100.0%
合計	30 12.0%	135 53.8%	85 33.9%	1 0.4%	251 100.0%

2番目に医師に対する評価と関連性の大きな要因の「大きな病気やけがの有無」に関しては、「最近患って治療中」の人が最も評価が高く、100%の人が「良い」と評価している。

他方、全然大きな病気やけがをしていない人は「良い」と評価している人が約61%であった。

3番目に関連性の大きな要因の「かかりつけの医師の有無」については、かかりつけの医師のある人は、「以前あったが今はいない」人や「かかりつけの医師のいない」人と比べて、医師の評価が高い。

4番目に関連性の大きな要因の「診てもらっている理由」に関しては、かかりつけの医師がいない人については9番目のカテゴリとして「不明」を設定した。「人柄が信頼できるから」診てもらっている人は約96%が「良い」と評価し、「説明がわかりやすいから」診てもらっている人は100%の人が「良い」と評価している。

年齢層は5番目であるが、A I Cの値がプラスであるので、関連性があるとはいえない。

以上のように、CATDAPは明確に目的変数に対する説明変数の規定要因の解析をすることができるので、CATDAPによる分析に適した調査事例は比較的多く、CATDAPはもっと活用

されるべき多変量解析法であると思われる。

4 A I Cプログラム

CATDAPは、調査や実験によって得られたデータをコンピュータに入力すると、自動的にクロス表を作り、A I Cの値を算出するが、筆者が1990年に開発したA I Cプログラムは、クロス表でしかデータが得られていない場合に、クロス表のデータを直接入力してA I Cの値を計算するパソコン用のプログラムである⁽⁷⁾。

A I C X P . E X Eは、Windows上で作動する実行プログラムであり、フリーソフトである。

A I Cプログラムは、2次元クロス表だけでなく、3次元クロス表（3重クロス集計表）および4次元クロス表（4重クロス集計表）にも適用できて、操作が非常に簡単な技法である。

表3は、意識調査におけるある質問に対する賛否と性別との間のクロス表であり、表4は賛否と年齢との間のクロス表である。

表3 賛否と性別

賛否 性別	賛成	反対	計
男性	132 66.3%	67 33.7%	199 100.0%
女性	93 46.3%	108 53.7%	201 100.0%
計	225 56.2%	175 43.8%	400 100.0%

表4 賛否と年齢

賛否 年齢	賛成	反対	計
35歳未満	131 66.2%	67 108.0%	198 100.0%
35歳以上	94 225.0%	108 53.7%	202 100.0%
35歳以上	225 56.2%	175 43.8%	400 100.0%

性別で見ると、男子では賛成が3分の2でかなり多いが、女子は反対の方が多い。この表3のデータにA I Cプログラムを適用すると、 $A I C = -14.480$ であり、性別と賛否とが関連性が大きいことが明らかである。

つぎに、年齢別に賛否を見ると、表4のように、35歳未満では賛成が約3分の2でかなり多いが、35歳以上では反対の方が多い。この表4のデータにA I Cプログラムを適用すると $A I C = -13.768$ であり、年齢は性別に劣らず賛否と関連性が大きいことが明らかである。

この2つのクロス表の結果から、男子の35歳未満は、相当賛成が多いと結論してしまいがちだが、正確を期するため性別・年齢別に賛否を見た3重クロス表を作成して分析してみると、表5のように予想に反して、男子の35歳未満は、反対の人が58%もいて、賛成の人よりもかなり多い。さらに、女子の35歳未満は90%以上が賛成であるが、女子の35歳以上は逆に96%以上が反対である。そこで、この表5にA I Cプログラムを適用してみると、 $A I C = -251.919$ と桁違いにA I Cの値が低く、性別と年齢が複雑に関連しあって賛否に非常に大きく関わっていることが明確となる。

このように、3重クロス集計をし、A I Cプログラムによって、調査データの精密な分析を行うことによって、真実が明らかとなった。（なお、表5は、A I Cプログラムの適用に関して、行が4つ、列が2つである。）

表5 性別・年齢別の賛否

		賛 否		計
		賛成	反対	
性別	年齢			
	男性	35歳未満	42 42.0%	58 58.0%
35歳以上		90 90.9%	9 9.1%	99 100.0%
女性	35歳未満	89 90.8%	9 9.2%	98 100.0%
	35歳以上	4 3.9%	99 96.1%	103 100.0%
計		225 56.2%	175 43.8%	400 100.0%

この400サンプルのデータをCATDAP-01プログラムで分析すると、賛否に対する有効性の大きな説明変数のリストでは、第1位が性別×年齢(AIC = -251.92)、第2位が性別(AIC = -14.48)、そして第3位が年齢(AIC = -13.77)となる。

社会調査で得られることの多いカテゴリー的データの多変量解析法として、CATDAPとAICプログラムは、解析力の優れた実用性の大きな多変量解析法であるといえる。まとめてみると以下のような特長を持っている。

①CATDAPは、カテゴリー的データである目的変数に対する、カテゴリー的データの説明変数の有効性の度合をAICすなわち情報量規準という統計量を用いて解析することができる。また、目的変数に対する有効性に関して、2つ以上の説明変数の最適組み合わせを分析することができる。

②CATDAPは、説明変数の数の制限がないので、調査した全変数を分析にかけることが出来て、目的変数に対して関連性のある変数とそうでない変数を選び分けることが出来る。

③CATDAPでは、該当するサンプルが少なすぎるカテゴリーがあっても、分析結果の信頼性が弱くなるということはない。

④CATDAPにはカテゴリーのリコード(recode)機能があり、該当サンプル数の少ないカテゴリーを除いたり2つ以上のカテゴリーを統合することが出来る。

⑤相互の関連性が小さくない説明変数が含まれている場合に、分析結果の正確性が弱くなるという弱点がCATDAPにはない。

⑥2つの説明変数のカテゴリーの間に交互作用があると、数量化Ⅱ類では分析結果の正確性が弱くなり、非常に強い交互作用があると計算出来なくなるという欠点があるが、CATDAPにはそのような欠点がない。

⑦AICプログラムは、Raw Data が得られなくてクロス表でしかデータが得られていない場合に、クロス表のデータを直接入力して情報量規準AICの値を計算することができる。

⑧AICプログラムは、2次元クロス表だけでなく、3次元クロス表(3重クロス集計表)および4次元クロス表(4重クロス集計表)にも適用できる、操作が非常に簡単な技法である。

以上のような特長を備えたCATDAPとAICプログラムは、人々の意識の分析や社会現象の

解明にもっと活用されるべき多変量解析法であると思われる。

5 因果関係の分析法

要因間の関連性の分析について述べてきたが、要因間の因果関係の分析および社会現象の因果関係の分析は社会調査のデータの分析において非常に大切である。因果関係の厳密な分析は、社会現象の研究において条件の統制が出来にくく、実験が困難であり、客観的な比較ができにくいいため、非常に難しいといえる。しかし、いくつかの分析法が開発されて来ている。

(1) パス解析

量的変数を用いて、変数間の相関係数を計算し、相関行列と仮説に基づいてパスダイヤグラムという因果関係の経路(Path)を図示する因果関連図を作成し、標準偏回帰係数を計算して、因果効果を解析するパス解析という方法がある⁽⁸⁾。

パス解析は、社会学などの社会科学と生物学や医学などの自然科学の分野で広く用いられて来た分析法であり、量的変数のデータが得られ、目的変数がはっきりしている場合には有効な方法であるといえる。

そこで、コンビニエンスストアの利用状況に関する意識調査(量的な質問調査)で得られた変数の相関行列データを用いてパス解析をおこなった。

目的変数はコンビニの週利用回数である。週利用回数は、0回～1.0回、1.1回～2.0回、2.1回～3.0回、3.1回～5.0回、5.1回～10.0回、10.1回以上の6つに区分した。

週利用回数に関連する変数として4つの変数を選んだ。すなわち、年齢(20歳未満、20歳以上24歳以下、25歳以上29歳以下、30歳代、40歳代、50歳代、60歳以上の7区分)、近所要因(住居・学校・職場近くに関する合成変数で、徒歩5分以内、徒歩5分～10分、自転車の距離、自動車やバイクの距離、近所に無しの5区分)、同居人数(1人～7人以上の7区分)、自動車の運転(ほぼ毎日運転する、週に3、4日、週に1、2日、月に2、3日、免許を持っているがほとんど乗らない、運転免許を持っていないの6区分)である。

分析ソフトは、筆者が1986年に作成したPATH.BASプログラムを改良したPATHX.EXEである。

このコンビニ利用調査の概要は以下のようなものである⁽⁹⁾。

① 調査対象と内訳

- 学生(広島国際学院大学・H大学・M大学・S大学)
 - 社会人(広島高等技術専門学校・広島国際学院大学科目等履修生とその知人)
- 高等技術専門学校は、キャリアアップや再就職のための職業訓練校である。

② 調査期間

2005年11月～12月

③ 調査方法

集合調査法と託送調査法および留め置き調査法を併用した。

④ 調査員

⑤ 回収率

有効回収率は85.0%である。

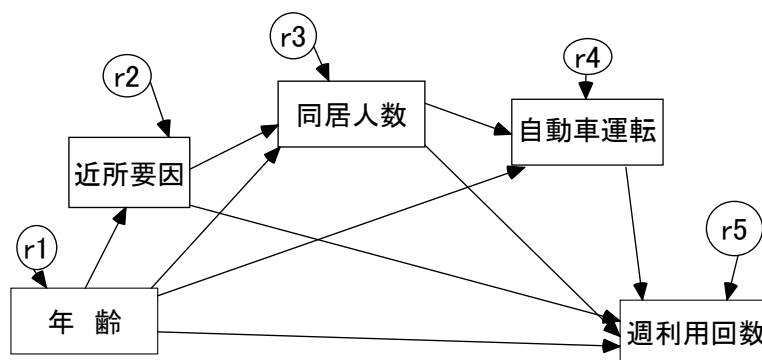
コンビニの週利用回数に関する相関行列データは表6で示される。

表6 週利用回数に関する要因間の相関行列

	年 齢	近所要因	同居人数	自動車運転	週利用回数
年 齢	1	0.038	0.039	0.283	0.263
近所要因	0.038	1	0.105	0.030	0.189
同居人数	0.039	0.080	1	0.088	0.108
自動車運転	0.283	0.032	0.088	1	0.108
週利用回数	0.263	0.189	0.108	0.108	1

完全逐次モデルによるパスダイヤグラムは 図1で示される。

図1 週利用回数のパスダイヤグラム



パス解析の計算結果は表7と表8で示される。

表7 パス係数

	1 年 齢	2 近所要因	3 同居人数	4 自動車運転	5 週利用回数
1 年 齢		0.0380	0.0351	0.2796	0.2460
2 近所要因			0.1037	0.0114	0.1707
3 同居人数				0.0759	0.0782
4 自動車運転					0.0264

表8 残差係数

1 年 齢	2 近所要因	3 同居人数	4 自動車運転	5 週利用回数
	0.9993	0.9939	0.9560	0.9443

パス解析の結果、年齢と週利用回数間の因果効果の総効果は、0.263003 である。直接効果が、0.246 で、間接効果の合計が 0.017003 である。

年齢とコンビニの利用回数は関連性があり、年齢が若い人の方が週利用回数が多い。

近所要因（コンビニが住居・学校・職場近くにある）と週利用回数間の因果効果の総効果は、0.204435 である。直接効果が 0.188868 であり、間接効果の合計が 0.015567 である。

コンビニが住居・学校・職場に近いことと利用回数は関連性があり、コンビニが近くにある人ほど、コンビニの週利用回数が多い。

(2) 共分散構造分析

パス解析をより一般化した分析方法が共分散構造分析という因果関係の分析方法である。

共分散構造分析は、調査で測定した観測変数（observed variables）だけでなく、調査で直接測定していない潜在変数（latent variables）や構成概念（因子）を用いて、因果関係の分析を行う統計的分析法である⁽¹⁰⁾。

共分散構造分析では、要因間における原因と結果に関して、分析モデルを設定し、因果関係のパス（経路）を示したパス図を描き、パス係数や標準化推定値（最尤推定値）等を計算して因果関係や要因間の関連性を精密に分析する。

共分散構造分析は、観測変数間の共分散を用いるので、連続変数データだけでなく各質問項目に関する 5 段階評価のデータ（順序尺度データ）から得られた相関行列データ等を用いることができる。そこで、先述した「現代医療に対する意識調査」における、人々の健康に関する意識と行動についての調査データに対して共分散構造分析を適用した⁽¹¹⁾。

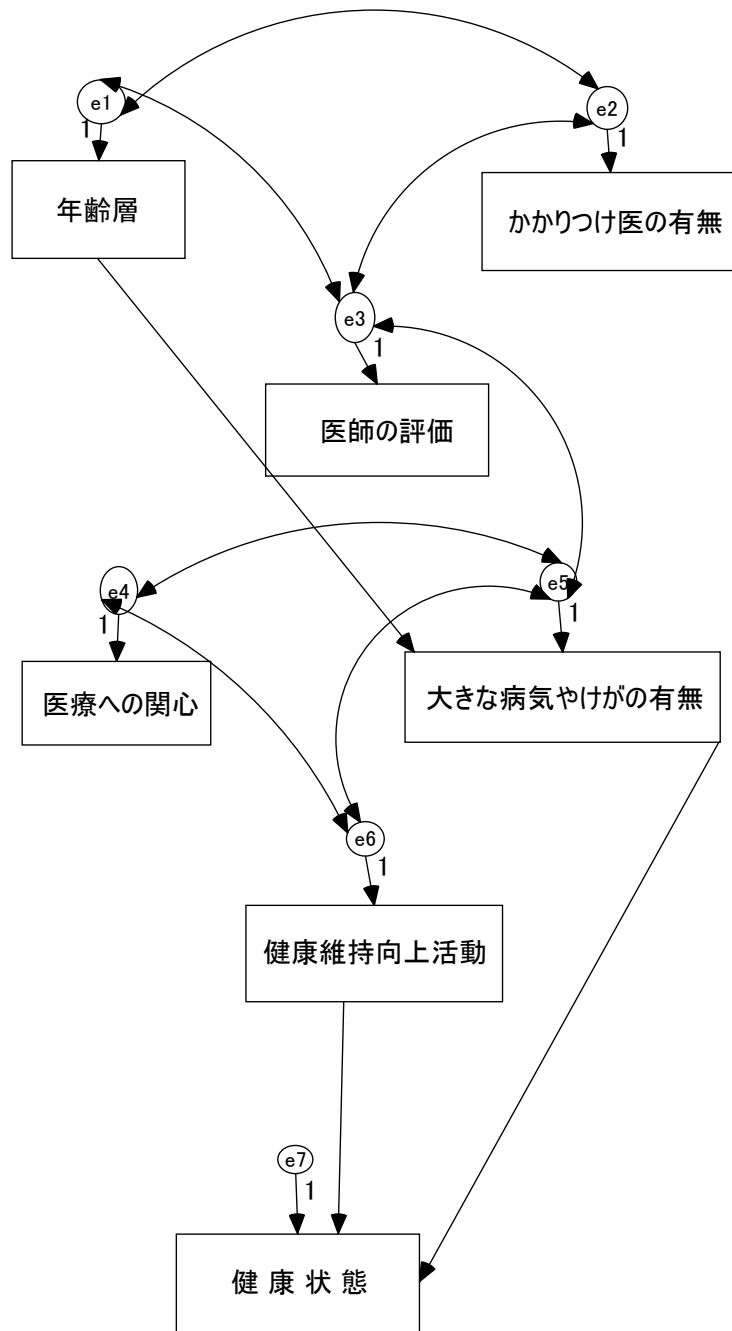
健康状態を結果変数とし、観測変数は年齢層、かかりつけ医の有無、医師の評価、医療への関心、大きな病気やけがの有無、健康維持向上活動である。

健康状態は、1.非常に良い、2.良い、3.どちらともいえない、4.悪い、という 4 つのカテゴリーであり、健康維持向上活動は、1.本格的な運動、2.ジョギングなどの軽い運動、3.規則正しい生活をする、4.補助健康薬品などを利用、5.栄養に気を配る、6.その他、7.何もしていない、という 7 つの健康維持向上活動に関する複数回答結果から合成した変数である。新しいカテゴリーは、1.非常によくしている、2.よくしている、3.ふつう程度している、4.あまりしていない、5.全然していない の 5 つのカテゴリーである。

健康観という潜在変数を設定して、因果関係分析モデルを検討したが、変数間の関係の強さを表す標準化推定値（最尤推定値）やモデルの適合度などの点で適切な分析結果が出ないので、潜在変数を設定しないモデルを採用した。

健康状態の因果関係分析モデルは以下のようなものである。

図2 健康状態の因果関係分析モデル



モデルの適合度は AGFI（修正適合度指標）が0.896であるので、まずまずの適合度であるといえる。

分析の結果、「健康状態」に最も大きな影響を及ぼす要因は「大きな病気やけがの有無」であり、「大きな病気やけがの有無」と「健康状態」との間の標準化推定値（最尤推定値）が0.307である。これまで大きな病気やけがをしていない人は健康状態が良い人が多い（約71%）けれども、「少し前」や「最近」大きな病気やけがをした人は、健康状態が良い人が少ない（約33%）。

「健康維持向上活動」と「健康状態」との間の標準化推定値は0.208であり、健康維持向上活動をよくしている人（運動をし、規則正しい生活をし、栄養に気を配っている人）は、健康状態が良い。

年齢層と「大きな病気やけがの有無」との間の標準化推定値は0.304であり、関連性が大きく、年齢層が高い人ほど、「大きな病気やけが」をした人の割合が高い。しかし、若い年齢層の人と比べて有病率が高い60歳以上の人々の場合、「大きな病気やけが」をしていない人は「健康維持向上活動」をよくしている人が多く、その結果「健康状態」が良い人が多かった。

共分散構造分析は、このように要因間の因果関係と相関関係を同時に精密に分析することができる。

6 結び

社会調査は、調査票を用いる調査であれば、仮説の構成などの調査の計画と調査票の作成というものが非常に重要であるが、調査で得られたデータの精密な分析も非常に大切であり、要因間の関連性の分析や因果関係の分析が大切である。

量的な社会調査によって得られたデータは、連続変数データとカテゴリー的データ（カテゴリカル・データ）が混ざっているが、カテゴリー的データが多いので、カテゴリー的データの分析法について中心に論じた。

カテゴリー的データの分析法として開発されたCATDAP（Categorical Data Analysis Program Package）やAICプログラムについて詳しく論じた。

CATDAPは、共にカテゴリー的データである目的変数に対する説明変数の関連性の度合をAICすなわち赤池情報量規準という統計量を用いて解析する多変量解析法である。この方法を用いて分析すると、調査で用いたカテゴリー的変数間の関連性が明確に判明する。調査データに名義尺度データが多く含まれている場合に有効性が大きいといえる。連続変数データを一定基準を用いて区分するとカテゴリー的データに変換できるので、CATDAPを用いることができやすい。

AICプログラムは要因間のクロス集計表を用いて、AICの値を計算し、要因間の関連性を判定するので、実数が表示されたクロス表や各セルの%と全体の合計が表示されているクロス表で用いることができる。そのため、汎用性がある。

因果関係の分析法に関しては、パス解析と共分散構造分析について論じた。共分散構造分析は、社会調査によって得られた量的データにおける因果関係の精密な分析ができる有効性の大きな優れた分析法である。

調査データのタイプや性質に合った分析法で調査データを分析することや分析目的に合致した精度の高い分析法で分析することが精密な分析結果を得るために必要であり、社会調査にとって非常に大切であると考えられる。

【注】

- (1) 林知己夫監修、駒沢勉、1982、『数量化理論とデータ処理』朝倉書店。
柳井晴夫・高木廣文、1986、『多変量解析ハンドブック』現代数学社。
- (2) 坂元慶行、1985、『カテゴリーカルデータのモデル分析』共立出版。
- (3) 坂元慶行、1988、「最適なクロス表の選択法」村上征勝・田村義保編、『パソコンによるデータ解析』朝倉書店、155-166ページ。
- (4) 坂元慶行、1988、「最適なクロス表の選択法」村上征勝・田村義保編、『パソコンによるデータ解析』朝倉書店、155-166ページ。
- (5) 小野能文、1995、「CATDAP」井上文夫・井上和子・小野能文・西垣悦代『よりよい社会調査をめざして』創元社、193-203ページ。
- (6) 中野秀一郎・小野能文、2005、「現代医療に対する「信頼」についての実証的研究」、『現代社会学』第6号、広島国際学院大学、27-46ページ。
- (7) 小野能文、1995、「CATDAP」井上文夫・井上和子・小野能文・西垣悦代『よりよい社会調査をめざして』創元社、193-203ページ。
- (8) 小野能文、2003、「パス解析」、大西正和編著小野能文他著『パソコンによるデータ分析』建帛社、99-104ページ。
- (9) 小野能文、2006、『コンビニエンスストアの利用に関する人々の意識と行動』広島国際学院大学現代社会学部。
- (10) 山本嘉一郎・小野寺孝義編、2002、『Amosによる共分散構造分析と解析事例 [第2版]』ナカニシヤ出版。
- (11) 小野能文、2005、『現代医療に対する人々の意識 現代医療に対する意識調査報告書』広島国際学院大学現代社会学部。

【参考文献】

- Johnson,R.A. & Wichern,D.W.,1992, *APPLIED MULTIVARIATE STATISTICAL ANALYSIS, 3rd.Ed., Prentice-Hall.*
- Alan Bryman & Duncan Craner,1994, *Quantitative data analysis for social scientist, Revised edition, Routledge.*
- 数理社会学会監修 / 与謝野有紀・栗田宣義・高田洋・安田雪編、2006、『社会の見方、測り方 計量社会学への招待』勁草書房。
- 太郎丸 博、2005、『人文・社会科学のためのカテゴリカル・データ解析入門』ナカニシヤ出版。
- ハンス・ザイゼル著 / 佐藤郁哉訳、2005、『数字で語る』新曜社。
- 大西正和編著小野能文他著、2003、『パソコンによるデータ分析』建帛社。
- 柳井晴夫・高木廣文、1986、『多変量解析ハンドブック』現代数学社。
- 林知己夫監修、駒沢勉、1982、『数量化理論とデータ処理』朝倉書店。
- 坂元慶行、1985、『カテゴリカルデータのモデル分析』共立出版。
- 坂元慶行、1988、「最適なクロス表の選択法」村上征勝・田村義保編、『パソコンによるデータ解析』朝倉書店、155-166ページ。
- 坂元慶行・石黒真木夫・北川源四郎、1983、『情報量統計学』共立出版。
- Katura,K. & Sakamoto,Y., 1980, "*CATDAP, A Categorical Data Analysis Program Package Computer Science Monographs No. 14, The Institute of Statistical Mathematics, pp. 1-73.*
- 坂元慶行、1981、「カテゴリカルデータにおける変数選択 — プログラム CATDAP を中心に一」『統計数理研究所彙報』第28巻第1号、135-155ページ。
- 小野能文、1995、「CATDAP」井上文夫・井上和子・小野能文・西垣悦代『よりよい社会調査をめざして』創元社、193-203ページ。
- 山本嘉一郎・小野寺孝義編、2002、『Amosによる共分散構造分析と解析事例 [第2版]』ナカニシヤ出版。
- 豊田秀樹編著、2007、『共分散構造分析 [Amos編]』東京書籍。
- 中野秀一郎・小野能文、2005、「現代医療に対する「信頼」についての実証的研究」、『現代社会学』第6号、広島国際学院大学、27-46ページ。
- 小野能文、2005、『現代医療に対する人々の意識 現代医療に対する意識調査報告書』広島国際学院大学現代社会学部。
- 小野能文、2006、『コンビニエンスストアの利用に関する人々の意識と行動』広島国際学院大学現代社会学部。

On the Data Analysis of Social Research : Mainly on Categorical Data

Yoshibumi ONO

I thought that making the questionnaire based on precise plan of social research and doing it and the precise analysis of data provided by questionnaire research are important, but there is much categorical data such as age group or evaluation data than the continuous variate data such as yearly income. Therefore I considered the analysis of categorical data in detail.

I considered a method of the analysis of the quantification theory that explanation variables are categorical data and CATDAP (Categorical Data Analysis Program Package) that all the factors are categorical data and the AIC program that are the analysis of categorical data in the cross table mainly.

CATDAP is the multivariate analysis method that can analyze in detail the relevance between the much explanation variables and purpose variables. I precisely discussed on the effectiveness being high to the data analysis of the social research that there are many categorical data.

The analysis of the correlation between factors on the consciousness of people or social phenomenon is important, but I considered the analyzing method of causation such as path analysis or the covariance structure analysis and I applied it to real research data because I think it was more important.

Keywords : social research , categorical data, CATDAP