

JGSS-2005 を用いた通信機器利用の潜在クラスモデル —統計分析セミナーにおける適用例—

都村 聞人

(大阪商業大学比較地域研究所 JGSS ポスト・ドクトラル研究員)

岩井 紀子

(大阪商業大学総合経営学部)

保田 時男

(大阪商業大学総合経営学部)

穴戸 邦章

(大阪商業大学総合経営学部)

Latent Class Models of the use of communication devices based on JGSS-2005:
Example of Application in the Statistical Analysis Seminar

Mondo TSUMURA Noriko IWAI Tokio YASUDA Kuniaki SHISHIDO
Osaka University of Commerce

This article aims to present examples of application of latent class models using JGSS-2005 data. Latent class models can briefly explain relationship between manifest variables by latent variables. Specifically, this article analyzes how the use of communication devices affects the income level of household. The types of use of communication devices (PC (at home), PC (at work), Fax, Cellular phone) are classified into three categories: "not use", "use all devices" and "use only cellular phone". Such factors of social stratification as work status, educational background and age affect the income level of household. Additionally, the types of use of communication devices also affect household income. An inclusion of latent classes in to the model particularly makes clear the effects of social stratification. Results of this analysis are interesting, but there is still room for improvement of the model.

Key Words: JGSS, Latent Class Analysis, Digital Divide

本稿は、JGSS データを用いた潜在クラスモデルの応用例を示すことを目的としている。潜在クラスモデルは、複数の質的な顕在変数の間の関係を潜在変数によりシンプルに説明できることに特徴がある。本稿では、通信機器の利用傾向と世帯収入レベルの関係を考察した。潜在クラス分析の結果、PC（自宅）、PC（職場）、ファックス、携帯電話といった通信機器の利用傾向は「非利用型」「すべて利用型」「携帯電話のみ利用型」の3類型に分類された。世帯収入レベルに対しては、就労状態、学歴、年齢の社会階層要因が影響を与えていた。さらに、ここで析出した通信機器利用傾向も世帯収入レベルに影響を及ぼしていた。とくに通信機器利用傾向の影響を考慮することにより、学歴、年齢といった社会階層要因が世帯収入レベルに及ぼす影響がより明確になった。分析セミナーの演習事例であるため、モデル改善の余地があるが、興味深い結果が得られたといえよう。

キーワード：JGSS，潜在クラス分析，デジタルデバイド

1. 本稿の目的

JGSS プロジェクトでは、2007 年 9 月 3-4 日に、山口一男教授（シカゴ大学社会学部）を講師として、統計分析のスキルアップを目指す大学院生・研究者を対象とした統計分析セミナーを開催した。セミナーにおいては、(1) 多次元 R C (Row and Column effects) モデルを含む様々な Association Models と Conditional Association Models、(2) ログリニア回帰モデル（多項ロジットモデル、Cumulative Logit (Ordered Logit)モデル、ネステッド・ロジットモデル）(3) 基本的潜在クラスモデル（クラスターモデル、ファクターモデル）(4) 潜在クラス変数を含む回帰モデル、潜在クラス変数が従属変数の多項ロジットモデル、潜在クラス変数を媒介変数にするパス解析モデルの講義が行われた。特に、2 日目の後半では、JGSS プロジェクトが用意したデータと分析枠組を用いて潜在クラス分析の演習を山口教授にお願いした。

本稿は、JGSS データを用いた潜在クラスモデル (latent class analysis) の適用例について、分析の過程をたどり紹介するものである。ただしこの分析演習は、事前に山口教授にご準備いただいたものではなく、いわば初めてデータに触れる段階での試行錯誤のプロセスの一端をお示しいただいたものである。したがって、以下で示す分析課題はあくまでも潜在クラスモデルを容易に説明するためのものであり、分析そのものが完結しているわけではないことをご理解いただきたい。むしろ、読者が再分析できるように、本稿では試行錯誤のプロセスを可能な限り丁寧に記述することにポイントをおいている⁽¹⁾。

2. 分析課題とその背景

近年の急速な情報化社会の進展は、人々の通信機器利用を大きく拡大させた。パソコン、ファックス、携帯電話などは、家庭や仕事の場を問わず広く利用されるようになってきている。またインターネットも急速に普及し、労働・経済・消費・コミュニケーションなどの側面で、人々の社会生活を大きく変化させた。

しかしながら、通信機器の利用が拡大したとはいえ、年齢、性別、学歴、居住地域、所得などによる利用の格差も大きい。パソコンを購入できるゆとりがあるか否かという経済的格差に加え、通信機器を自由に扱えるか否かという情報リテラシーの格差も存在する。そして通信機器の利用の格差は、情報取得における量的・質的差異を生み出し、さらにそのことが経済的格差を拡大するのではないかという懸念もなされている。いわゆるデジタルデバイドの問題である。たとえば、総務省の「通信動向利用調査」の結果を紹介した『情報通信白書（平成 19 年版）』（総務省）によれば、世帯収入が低いほど情報通信機器の保有率が低く、インターネットの利用率も低い傾向にある。

そこで本稿では、通信機器の利用傾向がどのように所得レベルに影響を与えているかという問題をとりあげる。具体的には、図 1 のような分析モデルを設定した。なお、最初にも述べたように、本稿で示す分析モデルは分析セミナーの演習用に考えられたものであり、簡略化された図式となっている。

(1) 分析 1：通信機器利用類型の分析

PC（自宅）、PC（職場）、ファックス、携帯電話という 4 つの通信機器の利用動向を規定する潜在クラスの類型を析出する。

潜在クラス分析の特徴は、複数の顕在変数（応答変数）の間の関係を潜在変数によりシンプルに説明できることにある。顕在変数、潜在変数ともに質的変数である場合に分析が可能な点で、因子分析などとは異なっている⁽²⁾。通信機器利用傾向の類型を探るのに適した分析方法といえよう。

(2) 分析 2：通信機器利用類型を規定する要因の分析

通信機器利用傾向は、社会階層要因によって影響を受けていると考えられる。本稿では、就労状態、学歴、年齢の 3 つの要因が通信機器利用傾向に与える影響を、多項ロジット潜在クラス回帰分析により検討する。仮説としては、就労層ほど、大学卒層ほど、また若年層ほど通信機器利用に積極的と考えられる。

(3) 分析 3：世帯収入レベルを規定する要因の分析

世帯収入レベルは、就労状態、学歴、年齢などの社会階層要因によって規定されていると考えられる。では、通信機器利用傾向の要因を考慮したときに、社会階層要因が世帯収入レベルに与える影響はどのように変化するのだろうか。また、通信機器利用は世帯収入レベルにどのような影響をもたらしているのか。以上の点を二項ロジット回帰分析により検討する。

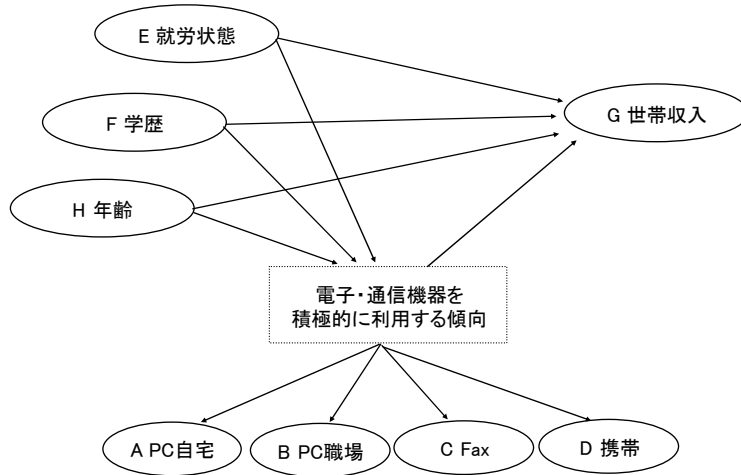


図1 分析モデル

3. 分析に用いるデータ

本稿では、大阪商業大学比較地域研究所が東京大学社会科学研究所と共同で行った JGSS-2005 データを用いて分析を行う。JGSS-2005 の母集団は、全国に居住する 2005 年 9 月 1 日時点で満 20～89 歳の男女である。層化 2 段無作為抽出法により、全国 307 地点の 4,500 人を抽出し、調査対象としている。調査は 2005 年 8 月から 11 月にかけて、全国において実施された。有効回収数は 2,023、回収率は 50.5%となっている。

4. 分析に使用する変数の概要と記述統計

分析には、表 1 の A～H の 8 つの変数を用いた。変数名は、前述の図 1 に対応している。A～D の通信機器の利用は、すべて 2 値変数である。就業状態は、XJOB1WK 変数のうち「仕事をした」「仕事をもっているが、病気・休暇などで先週は仕事を休んだ」を「就労」とし、「仕事をしていない」を「非就労」としている。学歴は、新制中学校、旧制尋常小学校、旧制高等小学校、旧制中学校・高等女学校を「中学卒」、新制高校、旧制実業・商業学校、旧制高校・旧制専門学校・高等師範学校を「高校卒」、新制短大・高専、新制大学、新制大学院、旧制大学・旧制大学院を「大学卒」としている。世帯収入レベルは、「世間一般と比べて、あなたの世帯収入はどれくらいですか」と、主観的評価を 5 点尺度で尋ねている OP5FFINX 変数を用いて、「平均よりかなり少ない」「平均より少ない」「ほぼ平均」を「少ない」とし、「平均より多い」「平均よりかなり多い」を「多い」としている。客観的な収入データを用いないのは、回答拒否や不明による欠損の割合が高いためである。

表 1 分析に使用する変数

変数名	変数ラベル	値の割り当て	JGSS-2005の変数
A	PC自宅で利用	1=非利用 2=利用	DOCOMP
B	PC職場で利用	1=非利用 2=利用	DOCOMPJ
C	ファックス利用	1=非利用 2=利用	DOFAX
D	携帯電話利用	1=非利用 2=利用	DOPHS
E	就業状態	1=非就労 2=就労	XJOB1WK
F	学歴	1=中学卒 2=高校卒 3=大学卒	XXLSTSCH
G	世帯収入レベル	1=少ない 2=多い	OP5FFINX
H	年齢	1=20代 2=30代 3=40代 4=50代 5=60代 6=70代 7=80代	AGEB

表 2 に、各変数の度数分布を示した。分析には、すべての設問に回答のあった 1,987 ケースを用いる。パソコンの利用率は、自宅では 37.0%、職場・学校では 28.9%である。分析の対象となる通信機器においては、携帯電話の利用率がもっとも高く 68.9%となっている。世帯収入のレベルに関しては、「平均より多い」という回答は 11.5%にとどまっている。

表 2 分析に使用する変数の度数分布

	n	%		n	%
A PC(自宅)			F 学歴		
非利用	1251	63.0	中学卒	514	25.9
利用	736	37.0	高校卒	837	42.1
計	1987	100.0	大学卒	636	32.0
B PC(職場・学校)			計	1987	100.0
非利用	1412	71.1	G 世帯収入レベル		
利用	575	28.9	平均+平均より少ない	1759	88.5
計	1987	100.0	平均より多い	228	11.5
C ファックス			計	1987	100.0
非利用	1230	61.9	H 年齢コーホート		
利用	757	38.1	20代	218	11.0
計	1987	100.0	30代	288	14.5
D 携帯電話			40代	301	15.1
非利用	617	31.1	50代	397	20.0
利用	1370	68.9	60代	420	21.1
計	1987	100.0	70代	277	13.9
E 就労状態			80代	86	4.3
非就労	756	38.0	計	1987	100.0
就労	1231	62.0			
計	1987	100.0			

5. 分析の準備

分析セミナーおよび本稿においては、Jeroen K. Vermunt が開発した統計解析ソフト LEM を用いて、潜在クラス分析を行った (Vermunt 1997)。LEM は、下記の Web サイトから、無料でソフトとマニュアルをダウンロードできる。

<http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html>

LEM の利用方法については、太郎丸 (2005) に若干の解説がある。

5.1 データファイルの作成

JGSS は主として統計ソフト SPSS の sav 形式でデータを公開している。LEM で分析を行うためには、変数と変数の間が半角スペースか Tab コードで区切られたテキストファイルでデータファイルを作成しておく必要がある。まず JGSS-2005 データから、表 1 の 8 つの変数を作成し、それ以外の変数を削除する。欠損値を含むケースは削除しておかなければならない。メニューバーの「ファイル」→「名前を付けて保存」と進み、ファイル名を「ABCDEFGH」とし、ファイルの種類から「(*.dat)」を選んで適当な場所に保存する。これで「ABCDEFGH.dat」ファイルが作成される⁽³⁾。本稿では、表 3 の 4 種類のデータファイルを用いた。それぞれのデータファイルは、ファイル名に含まれている変数のみをその順番どおりに含んでいる。LEM ではデータに含まれる一部の変数のみを利用することが難しいので、このようにあらかじめ含まれる変数が異なる複数のファイルを用意する。

表 3 分析に使用したデータファイル

	ファイル名	データに含まれる変数
データ1	ABCDEFGH.dat	全変数
データ2	ABCDEFH.dat	電子・通信機器(ABCD)+就労状態(E)、学歴(F)、年齢(H)
データ3	ABCD.dat	電子・通信機器(ABCD)
データ4	EFGH.dat	就労状態(E)、学歴(F)、世帯収入レベル(G)、年齢(H)

5.2 LEM のインストール

前述の Web サイトからダウンロードした LEM のフォルダは、ディレクトリの途中で半角スペースを含まない場所に保存する。たとえば、C ドライブ直下に「C:\LEMWIN\seminar」と保存する。Windows XP の場合「デスクトップ」は「C:\Documents and Settings\...」にあるため、半角スペースを含んでしまい、正常に動作しないので注意が必要である。

先ほど作成した 4 種類のデータファイルをすべて LEM のフォルダに保存しておく。データファイルをプログラムと同じ場所に保存しないと、分析の際にデータファイルの場所指定が複雑になる。

6. 分析

6.1 分析 1：通信機器利用類型の分析（潜在クラス分析）

まず、通信機器利用傾向をいくつかの類型に分けることを試みる。潜在クラス分析は、潜在クラス変数を含む対数線形モデルである。顕在変数（応答変数）間の関連が、応答パターンの潜在的異質性によって生じると仮定している（局所独立の仮定）⁽⁴⁾。潜在クラスの数を決め、データにもっとも適合するパラメータの値を最尤推定する。推定されたパラメータをもとにクロス表の各セル度数を推定し、適合度検定を行うことでモデルの妥当性が判断できる。なるべく少ない潜在クラスの数で、類型を説明できるモデルを析出することが望ましい。

通信機器利用傾向について、まず潜在クラスを 2 つに設定したモデルを試行した (Model 1)。LEM では、input ウィンドウにプログラム（ここでは Model 1 のプログラム）を書く。「lat」は潜在変数の数を指定する（ここでは 1）。「man」は顕在変数の数（同 4）、「dim」は各変数（潜在変数、顕在変数の順）のカテゴリー数、「lab」は変数名、「mod」はモデル、「rec」はデータのケース数、「dum」はリファレンスカテゴリー、「dat」は使用するデータファイル名を示している。なお、「dum」の「-1」は特定のリファレンスカテゴリーを設けない effect coding であることを示している。

次に、潜在クラスをひとつ増やし 3 つに設定したモデルを試行した (Model 2)。プログラム上は、Model 1 の X の「dim」を「3」に変更しただけである。なお、潜在クラスを 4 つに設定したモデルの場合、自由度がマイナスになる。

```

【Model 1】
lat 1
man 4
dim 2 2 2 2 2
lab X A B C D
mod X
      A|X
      B|X
      C|X
      D|X
rec 1987
dum -1 1 1 1 1
dat ABCD.dat
    
```

```

【Model 2】
lat 1
man 4
dim 3 2 2 2 2
lab X A B C D
mod X
      A|X
      B|X
      C|X
      D|X
rec 1987
dum -1 1 1 1 1
dat ABCD.dat
    
```

各モデルの適合度は表 4 に示す通りである⁽⁵⁾。p 値をみると、潜在クラスを 3 つに設定した場合 (Model 2) に、データに適合的なモデルを推定できることがわかる ($p > .05$)。

潜在クラスの数、つまり通信機器利用類型の数を 3 とした場合の各類型の構成比率と各通信機器利用の条件付確率（すなわち利用率）をみておこう（表 5）⁽⁶⁾。所属率が示すように、1987 ケースのうち 33.4%は類型 1 に、32.5%は類型 2 に、34.1%は類型 3 の通信機器利用パターンをとると推定された。表 5 の利用率は、各類型におけるそれぞれの通信機器の利用率を示している。たとえば、類型 1 の場合、PC（自宅）の利用率が 2.5%、PC（職場）の利用率が 0.1%、ファックスの利用率が 7.2%、携帯

電話の利用率が 21.7%である。

それぞれの類型における通信機器利用傾向は次のようにまとめられる。類型 1 はいずれの通信機器に関しても利用率が低い。したがって、類型 1 は「非利用型」といえる。類型 2 は携帯電話の利用率が非常に高く、PC やファックスの利用率は全体の平均利用率（表 2）を下回るのので、「携帯電話のみ利用型」といえる。類型 3 は、すべての通信機器の利用率が高い。したがって、類型 3 は「すべて利用型」といえる。Model 2 の分析では、全ケースが以上の 3 類型にほぼ 3 分の 1 ずつ割り振られていることになる。

表 4 Model 1~Model 4 の適合度

	潜在変数	潜在クラスの数	自由度 df	尤度比統計量 G ²	p値	BIC
Model 1	1	2	6	17.48	.008	-28.08
Model 2	1	3	1	1.33	.248	- 6.26
Model 3	2	-	1	1.19	.276	- 6.41
Model 4	2	-	2	0.94	.626	-14.25

表 5 Model 2 : 通信機器利用類型・類型への所属率・利用率

	類型1	類型2	類型3
	非利用型	携帯電話のみ利用型	すべて利用型
類型への所属率	.334	.325	.341
PC(自宅)の利用率	.025	.317	.759
PC(職場)の利用率	.001	.197	.660
ファックスの利用率	.072	.378	.686
携帯電話の利用率	.217	.873	.977

Model 1 と Model 2 の場合、1 つの潜在変数のクラス（類型）によって各通信機器の利用が左右されると考えたが（クラスターモデル）、潜在クラス分析では、複数の潜在変数に左右されて各通信機器の利用が決まると考えることもできる（ファクターモデル）。複数の潜在変数を設定する場合、それぞれの潜在変数を 2 値（つまり、それぞれのファクターの有無を表す）と考えることが多い。複数の潜在変数がそれぞれに多くのカテゴリーをもつという設定は複雑すぎてあまり意味がないからである。

Model 3 が潜在変数を 2 つ（それぞれ 2 値）に指定したモデルである。潜在変数の数を示す「lat」を 2 としている。「mod」が示すように、ふたつの潜在変数 X と Y は独立と仮定している。表 4 により、Model 2 と Model 3 を比較すると、適合度の改善はみられない。

```

【Model 3】
lat 2
man 4
dim 2 2 2 2 2 2
lab X Y A B C D
mod XY {X, Y}
      A|XY {A, AX, AY}
      B|XY {B, BX, BY}
      C|XY {C, CX, CY}
      D|XY {D, DX, DY}
rec 1987
dum -1 -1 1 1 1 1
dat ABCD.dat
    
```

```

【Model 4】
lat 2
man 4
dim 2 2 2 2 2 2
lab X Y A B C D
mod XY {X, Y}
      A|XY {A, AX, AY}
      B|XY {B, BX, BY}
      C|XY {C, CX}
      D|XY {D, DX, DY}
rec 1987
dum -1 -1 1 1 1 1
dat ABCD.dat
    
```

表 6 は、Model 3 の各類型の構成比率と各通信機器の利用率を示したものである。各通信機器の利用率に注目すると、潜在変数 X を一定としたとき、潜在変数 Y は C (ファックス) に大きな影響を与えていないことがわかる。そこで、潜在変数 Y と C (ファックス) の関連を仮定しないモデルも検討することにした (Model4)。

表 6 Model 3 : 通信機器利用類型・類型への所属率・利用率

	類型1	類型2	類型3	類型4
	X=1 Y=1	X=1 Y=2	X=2 Y=1	X=2 Y=2
類型への所属率	.195	.396	.135	.274
PC(自宅)の利用率	.857	.465	.107	.017
PC(職場)の利用率	.755	.351	.017	.003
ファックスの利用率	.737	.500	.160	.064
携帯電話の利用率	.980	.938	.486	.224

表 7 は、Model 4 の各類型の構成比率と各通信機器の利用率を示したものである。各通信機器の利用率をみると、類型 1 と類型 2 は「すべて利用型」、類型 3 は「非利用型」、類型 4 は「携帯電話のみ利用型」というパターンと考えられる。Y を一定としたとき、X=1 の場合、X=2 よりも全体に通信機器の利用率が高い。また X を一定としたとき、Y=1 に比して Y=2 の場合、ファックス以外の利用率がやや上昇している。このように X と Y の意味は明確であるが、潜在変数をひとつ追加するほどの意義は見出せない⁽⁷⁾。したがって、より簡潔なモデルで類型の解釈も合理的な Model 2 を採用し、以降の分析を進めることにした。

表 7 Model 4 : 通信機器利用類型・類型への所属率・利用率

	類型1	類型2	類型3	類型4
	すべて利用 (PC自宅中心)型	すべて利用 (PC自宅+職場)型	非利用型	携帯電話 のみ利用型
	X=1 Y=1	X=1 Y=2	X=2 Y=1	X=2 Y=2
類型への所属率	.382	.117	.384	.118
PC(自宅)の利用率	.537	.933	.042	.344
PC(職場)の利用率	.383	.949	.011	.244
ファックスの利用率	.682	.682	.082	.082
携帯電話の利用率	.933	1.000	.256	1.000

6.2 分析 2 : 通信機器利用類型を規定する要因の分析 (多項ロジット潜在クラス回帰分析)

分析 2 においては、分析 1 で析出された通信機器利用類型がどのような社会階層要因によって規定されているかを多項ロジット潜在クラス回帰分析により検討する。具体的には、図 1 で示したように、就労状態、学歴、年齢が通信機器利用傾向 (潜在クラス) に及ぼす影響を考察する。

以下に示す Model 5 は、H 年齢コーホートと潜在変数 X が線形関係にあることを仮定したモデルである。プログラムについて、いくつか説明しておこう。「sta (starting value)」は、パラメータの初期値を指定するコマンドである。これからいくつかの多項ロジット潜在クラス回帰分析を行うが、分析の度に潜在クラスの順番が変化するとわかりにくい。そこで、分析 1 の結果に基づいて、3 つの潜在

クラスを特定の順番にすることができる。分析 2 では、潜在クラス 1=非利用型、潜在クラス 2=すべて利用型、潜在クラス 3=携帯電話のみ利用型と定義したい（表 5 から潜在クラス 2 と潜在クラス 3 の順番を変更した）。指定の方法は以下の通りである。「sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]」は、変数 A（PC（自宅））と潜在変数 X の関係を指定している。[]内の 6 つの値は、左から潜在クラス 1 の際の A=1 の条件付き確率、潜在クラス 1 の際の A=2 の条件付き確率、潜在クラス 2 の際の A=1 の条件付き確率、潜在クラス 2 の際の A=2 の条件付き確率、潜在クラス 3 の際の A=1 の条件付き確率、潜在クラス 3 の際の A=2 の条件付き確率を示している。変数 A（PC（自宅））の各潜在クラスにおける条件付き確率（A=2、すなわち利用率）は、表 5 に示したように、非利用型=0.025、携帯電話のみ利用型=0.317、すべて利用型=0.759 であった。この値に基づき、潜在クラス 1=非利用型、潜在クラス 2=すべて利用型、潜在クラス 3=携帯電話のみ利用型となるように指定する。ここでは、各潜在クラスの A=2 である条件付き確率について、便宜的に 0.025 には 0.1 を、0.759 には 0.8 を、0.317 には 0.3 をわりあてている。「sta」で指定する値は、必ずしも正確である必要はなく近似値で構わない。順番を変更するので、潜在クラス 2 の際の A=2（すべて利用型の PC（自宅）の確率）、すなわち左から 4 つめの値が 0.8 となる。各潜在クラスの A=1 については、1 から A=2 の条件付き確率を引いた値が割り当てられる。したがって、0.9+0.1=1、0.2+0.8=1、0.7+0.3=1 という関係になっている。同様に B と X、C と X、D と X の関係も指定した⁽⁸⁾。

$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7) =$	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><th></th><th>H₁</th><th>H₂</th><th>H₃</th><th>H₄</th><th>H₅</th><th>H₆</th><th>H₇</th></tr> <tr><th>X₁</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><th>X₂</th><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr> <tr><th>X₃</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> </table>		H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇	X ₁	0	0	0	0	0	0	0	X ₂	1	2	3	4	5	6	7	X ₃	0	0	0	0	0	0	0	} desの1行目に あたる
		H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇																										
	X ₁	0	0	0	0	0	0	0																										
	X ₂	1	2	3	4	5	6	7																										
X ₃	0	0	0	0	0	0	0																											
$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7) =$	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><th></th><th>H₁</th><th>H₂</th><th>H₃</th><th>H₄</th><th>H₅</th><th>H₆</th><th>H₇</th></tr> <tr><th>X₁</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><th>X₂</th><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><th>X₃</th><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr> </table>		H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇	X ₁	0	0	0	0	0	0	0	X ₂	0	0	0	0	0	0	0	X ₃	1	2	3	4	5	6	7	} desの2行目に あたる
		H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇																										
	X ₁	0	0	0	0	0	0	0																										
	X ₂	0	0	0	0	0	0	0																										
X ₃	1	2	3	4	5	6	7																											

図 2 Model 5 のデザイン行列の指定

```

【Model 5】
lat 1
man 7
dim 3 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, cov(XH,2)}
    A|X
    B|X
    C|X
    D|X
rec 1987
dum 1 1 1 1 1 2 2 1
des [0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

プログラムの「cov (XH,2)」は、X（潜在変数）と H（年齢）の関係を 2 つのパラメータで設定することを示すコマンドである。各パラメータの内容はそれぞれデザイン行列で表現する。デザイン行列の書き方は多様な方法があるが、本稿では図 2 のように考えている。「des」の 1 行目で潜在クラス 2（すべて利用型）と H の線形関係を、「des」の 2 行目で潜在クラス 3（携帯電話のみ利用型）と H の線形関係を表している。若いコーホートから順に 1 から 7 の値を割り当て、年齢と通信機器利用の

線形関係を表現している。

表 8 の Model 5 の結果⁽⁹⁾が示すように、「すべて利用型」と「非利用型」の対比においては、非就労層、中学卒、年齢がマイナスで有意となっている。つまり、就労層に比べ非就労層では、また高校卒に比べ中学卒では、そして高年齢層であるほど、「非利用型」に比して「すべて利用型」が少ない。他方で、大学卒はプラスで有意となっており、高校卒に比べ「すべて利用型」が多い。就労層、大学卒層、若年層ほど通信機器を積極的に利用している傾向が読み取れる。

「携帯電話のみ利用型」と「非利用型」の対比においては、非就労、中学卒、年齢がマイナスで有意となっている。つまり、就労層に比べ非就労層では、また高校卒に比べ中学卒では、そして高年齢層であるほど、「非利用型」に比して「携帯電話のみ利用型」が少ない。言い方を変えれば、非就労層、中学卒層、高年齢層では通信機器を利用しない傾向がある⁽¹⁰⁾。

次に、年齢コーホートの影響をより詳細に検討するため、年齢変数を 10 歳刻みのコーホートのダミー変数として分析した (Model 6)。プログラム上は、X と H の関係が Model 5 から変化しているだけである。詳しい分析結果は省略するが、必ずしも年齢が高くなるほど通信機器利用が低下するわけではないことが明らかになった。具体的には、20 歳代に比べ 30 歳代で「非利用型」に比べ「すべて利用型」が有意に多く、また「非利用型」に比べ「携帯電話のみ利用型」も有意に多かった。つまり、20 歳代よりも 30 歳代で通信機器利用が活発であることが明らかになった。

表 8 通信機器利用類型の多項ロジットモデル (Model 5)

	すべて利用型 ／非利用型	携帯電話のみ利用型 ／非利用型
変数(参照カテゴリー)	β	β
就労状態(就労)		
非就労	-4.618 ***	-0.588 *
学歴(高校卒)		
中学卒	-2.790 ***	-1.416 ***
大学卒	1.938 ***	0.445
年齢	-1.416 ***	-1.151 ***
G ²	684.21	
自由度	608	
N	1987	

+p<.10, *p<.05, **p<.01, ***p<.001

注:年齢は10歳刻みのコーホートで線形関係を仮定している

```

【Model 6】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, XH}
      A|X
      B|X
      C|X
      D|X
rec 1987
dum 1 1 1 1 1 2 2 1
sta  A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta  B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta  C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta  D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

```

【Model 7】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, XH}
      A|X
      B|X
      C|X
      D|X
rec 1987
dum 1 1 1 1 1 2 2 2
sta  A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta  B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta  C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta  D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

そこで、Model 6 の年齢コーホートのリファレンスカテゴリー (dum) を 30 歳代にして分析を行った (Model 7)。表 9 は、Model 7 の分析結果である。30 歳代に比べ 20 歳代で「すべて利用型」、「携帯電話のみ利用型」が少ないことがわかる。ただし、40 歳代以上では、年齢が高くなるほど通信機器を利用しない傾向にある。また、80 歳代において、「すべて利用型」対「非利用型」に比べ、「携帯電話のみ利用型」対「非利用型」の係数が小さいことから、80 歳代であっても携帯電話はある程度利用されていると考えられる。その他の変数の影響は、Model 5 と同じ傾向である。

Model 7 の分析により、通信機器利用傾向は 30 歳代でもっとも利用が活発になり、以降年齢が高くなるほど利用が少なくなることが明らかになった。そこで、年齢の 2 乗項を変数として追加することにより、モデルの改善を試みた。Model 8 は、「des」が示すように、潜在クラス 2 (すべて利用型) との対比、および潜在クラス 3 (携帯電話のみ利用型) との対比のそれぞれについて、年齢 2 乗項を追加している。

表 9 通信機器利用類型の多項ロジットモデル (Model 7)

変数 (参照カテゴリー)	すべて利用型 / 非利用型	携帯電話のみ利用型 / 非利用型
	β	β
就労状態 (就労)		
非就労	-4.851 ***	-0.732 **
学歴 (高校卒)		
中学卒	-2.897 ***	-1.482 ***
大学卒	1.999 ***	0.477
年齢 (30代)		
20代	-9.079 ***	-8.613 ***
40代	-10.825 ***	-10.706 ***
50代	-12.404 ***	-11.694 ***
60代	-14.050 ***	-12.793 ***
70代	-15.908 ***	-13.743 ***
80代	-442.874 ***	-16.047 ***
G ²	668.00	
自由度	598	
N	1987	

+p<.10, *p<.05, **p<.01, ***p<.001

```

【Model 8】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, cov(XH,4)}
    A|X
    B|X
    C|X
    D|X
rec 1987
dum 1 1 1 1 1 2 2 1
des [0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
     0 0 0 0 0 0 0 1 4 9 16 25 36 49 0 0 0 0 0 0 0
     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 4 9 16 25 36 49]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

表 10 の Model 8 では、「すべて利用型」との対比においてのみ年齢 2 乗項がマイナスで有意となっている。この結果をふまえて、潜在クラス 2 (すべて利用型) との対比にのみ年齢 2 乗項を追加し、潜在クラス 3 (携帯電話のみ利用型) との対比からは年齢 2 乗項を取り除いたのが Model 9 である。

表 10 の「すべて利用型」との対比においては年齢 2 乗項がマイナスで有意、「携帯電話のみ利用型」との対比においては年齢がマイナスで有意という結果になった。つまり、「携帯電話のみ利用型」は年齢とともに一定のペースで減少するが、「すべて利用型」は高齢層で加速度的に減少するということがある。

年齢以外の変数に関しても、いくつかモデルの改善を試行した。まず、Model 10 において、就労と学歴の交互作用を検討した。就労していれば学歴を問わず通信機器の利用に積極的である、あるいは非就労でも大学卒層では通信機器を積極的に利用しているといった仮説が考えられるからである。しかしながら、Model 10 のプログラムを実行した結果、観測度数がゼロのセルが多いため、一部のパラメータが推定できなかった。

また、「PC（職場）」の変数は特別な意味を持っていると考えられる。仕事で PC を使うか否かは、通信機器利用傾向という点で違いが大きいためからである。加えて、PC を職場で利用している層が「携帯電話のみ利用型」になることは考えられない。そこで、B の PC（職場）と潜在クラスの関係性を潜在クラス 2 に限定した Model 11 を検討した⁽¹¹⁾。しかし、その結果は Model 5 とほとんど変わらなかった（詳細な結果は省略）。

表 10 通信機器利用類型の多項ロジットモデル（Model 8～9）

変数(参照カテゴリー)	Model 8		Model 9	
	すべて利用型 ／非利用型	携帯電話のみ利用型 ／非利用型	すべて利用型 ／非利用型	携帯電話のみ利用型 ／非利用型
	β	β	β	β
就労状態(就労)				
非就労	-4.847 ***	-0.748 **	-4.854 ***	-0.753 **
学歴(高校卒)				
中学卒	-2.943 ***	-1.483 ***	-2.954 ***	-1.490 ***
大学卒	2.051 ***	0.518	2.044 ***	0.507
年齢				
年齢	-0.059	-0.859	-0.307	-1.095 ***
年齢2乗	-0.199 *	-0.024	-0.173 **	-----
G ²	672.63		672.58	
自由度	606		607	
N	1987		1987	

+p<.10, *p<.05, **p<.01, ***p<.001

注:年齢は10歳刻みのコーホートで線形関係を仮定している

年齢2乗は10歳刻みのコーホートの2乗項で線形関係を仮定している

```

【Model 9】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, cov(XH,3)}
  A|X
  B|X
  C|X
  D|X
rec 1987
dum 1 1 1 1 1 2 2 1
des [0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
      0 0 0 0 0 0 0 1 4 9 16 25 36 49 0 0 0 0 0 0 0
      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

```

【Model 10】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, cov(XH,2), XEF}
    A|X
    B|X
    C|X
    D|X
rec 1987
dum 1 1 1 1 1 2 2 1
des [0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

```

【Model 11】
lat 1
man 7
dim 3 2 2 2 2 2 3 7
lab X A B C D E F H
mod X|EFH {X, XE, XF, cov(XH,2)}
    A|X
    B|X {B, cov(XB,1)}
    C|X
    D|X
rec 1987
dum 1 1 1 1 1 2 2 1
des [0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
    0 0 0 1 0 0]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFH.dat
    
```

表 11 は、Model 5～Model 11 の適合度検定の結果を示したものである。表 11 が示すように、有意水準を 5% とすると、適合的なモデルはひとつもない。また G^2 と χ^2 の値が大きく乖離している。これは、年齢コーホートなどを細かく分けすぎたために、観測度数ゼロのセルが多くなったためと考えられる。したがって、適合度検定の結果からモデルを選択することはできない。そこで、以下の分析はもっともシンプルな Model 5 にしたがって進めることにした。

表 11 Model 5～Model 11 の適合度

	自由度 df	尤度比統計量 G^2	p値	Pearsonの χ^2 値
Model 5	608	684.21	.017	976.44
Model 6	598	668.00	.025	924.98
Model 7	598	668.00	.025	924.97
Model 8	606	672.63	.031	928.77
Model 9	607	672.38	.033	928.77
Model 10	604	668.40	.035	923.88
Model 11	609	684.32	.018	993.53

5.3 分析 3：世帯収入レベルを規定する要因の分析（二項ロジット回帰分析）

分析 3 においては、世帯収入レベルを規定する要因を二項ロジット回帰分析により考察する。まず、社会階層要因が世帯収入レベルをどのように規定しているかを検討した (Model 12)。世帯収入レベルは表 1 で示したように「1 平均以下」「2 平均より多い」の 2 値変数である。また、年齢と世帯収入レベルに線形の関係性を仮定している。分析に用いる変数が E、F、G、H の 4 つであるので、データファイルは EFGH.dat となる。

表 12 は、二項ロジット回帰分析の結果を示したものである。Model 12 をみると、就労状態は非就労がマイナスで有意となっている。就労に比べ非就労では、世帯収入レベルが低い傾向にある。学歴は、中学卒がマイナスで有意、大学卒がプラスで有意となっている。高校卒に比べ中学卒では世帯収入レベルが低く、高校卒に比べ大学卒では世帯収入レベルが高い。年齢はプラスで有意であり、年齢が高いほど世帯収入レベルが高くなっていることを示している。以上のように、社会階層要因は、世帯収入レベルに影響を与えていることが明らかである。

では、社会階層要因に加えて、通信機器利用傾向が世帯収入レベルにどのような影響を与えているかを検討しよう。Model 13 は、Model 12 に通信機器利用傾向の潜在変数を追加したものである。年齢と潜在クラスおよび年齢と世帯収入レベルに線形の関係性を仮定している。

```

【Model 12】
man 4
dim 2 3 2 7
lab E F G H
mod G|EFH {G, GE, GF, cov(GH,1)}
rec 1987
dum 2 2 1 1
des [ 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
dat EFGH.dat
    
```

表 12 世帯収入レベルの二項ロジット回帰分析

	Model 12	Model 13	Model 14	Model 15
変数(参照カテゴリー)	β	β	β	β
就労状態(就労)				
非就労	-0.353 *	0.380	0.369	0.429 +
学歴(高校卒)				
中学卒	-0.902 ***	-0.172	0.478	-0.130
大学卒	0.883 ***	0.664 ***	1.259 ***	0.660 ***
年齢	0.110 *	0.249 ***	0.255 ***	0.948 +
潜在変数(非利用型)				
すべて利用型		2.634 ***	3.004 *	6.668 *
携帯電話のみ利用型		1.724 ***	2.284	6.684 *
潜在変数*学歴				
(すべて利用型*高校卒)			-	
すべて利用型*中学卒			-0.695	
すべて利用型*大学卒			-0.471	
(携帯電話のみ利用型*高校卒)			-	
携帯電話のみ利用型*中学卒			-0.992	
携帯電話のみ利用型*大学卒			-0.718	
潜在変数*年齢				
すべて利用型*年齢				-0.581
携帯電話のみ利用型*年齢				-0.854 +
G ²	57.93	930.87	926.23	923.92
自由度	37	1273	1269	1271
N	1987	1987	1987	1987

+p<.10, *p<.05, **p<.01, ***p<.001

注:年齢は10歳刻みのコーホートで線形関係を仮定している

表 12 が示すように、通信機器利用傾向の潜在変数は、「すべて利用型」「携帯電話のみ利用型」ともにプラスで有意となっている。つまり、「非利用型」に比べ、通信機器を利用する層では、世帯収入レベルが高い。とくに、「すべて利用型」の係数が大きく、通信機器を全般的に利用する層では、世帯収入が平均より多くなっていることがわかる。社会階層要因の影響の変化をみると、Model 12 ではマイナスで有意であった就労状態は、有意ではなくなる。学歴に関しては、中学卒は有意でなくなり、大学卒はプラスで有意であるが、Model 12 に比べその効果を弱めている (Model 12 : 0.883、Model 13 : 0.664)。このことは通信機器利用傾向の要因により、就労状態による世帯収入レベルの格差および学歴による世帯収入レベルの格差が一部説明されていることを示している。つまり、就労層と高学歴層で世帯収入レベルが高いことは、それらの層が通信機器を積極的に利用していることにより一部説明される。

年齢は Model 13 においても、有意であり、年齢が高いほど世帯収入レベルが高い。ここで注目すべきことは、通信機器利用傾向の潜在変数を追加したことにより、年齢効果が増大していることである (Model 12 : 0.110、Model 13 : 0.249)。これは以下のように考えることができる。世帯収入レベルが

```

【Model 13】
lat 1
man 8
dim 3 2 2 2 2 2 3 2 7
lab X A B C D E F G H
mod X|EFH {X, XE, XF, cov(XH,2)}
    A|X
    B|X
    C|X
    D|X
    G|XEFH {G, GE, GF, cov(GH,1), GX}
rec 1987
dum 1 1 1 1 1 2 2 1 1
des [0 0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
    0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFGH.dat
    
```

```

【Model 14】
lat 1
man 8
dim 3 2 2 2 2 2 3 2 7
lab X A B C D E F G H
mod X|EFH {X, XE, XF, cov(XH,2)}
    A|X
    B|X
    C|X
    D|X
    G|XEFH {G, GE, GF, cov(GH,1), GX, GXF}
rec 1987
dum 1 1 1 1 1 2 2 1 1
des [0 0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
    0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFGH.dat
    
```

高い層は年齢が高いほど多いが、年齢が高い場合には通信機器は利用しない傾向が強い。通信機器を利用しない層では世帯収入レベルが低いいため、Model 12 では年齢効果が過小評価されてしまっていたのである。しかし、Model 13 で通信機器利用傾向をコントロールした結果、この過小評価が改善され、年齢効果が増大したと考えられる。つまり、通信機器利用傾向の潜在変数を加えたことにより、社会階層要因が世帯収入レベルに及ぼす影響がよりクリアになったといえよう。

潜在変数の追加がもたらした変化についてさらに2つのモデルを試行した。第1は通信機器利用傾向と学歴の関係について、より詳細に考察した Model 14 である。Model 14 は、Model 13 に潜在変数（通信機器利用傾向）と学歴の交互作用項を追加したものである。表 12 が示すように、潜在変数と学歴の交互作用項は有意ではない。係数の符号もすべてマイナスであり、解釈が難しい。大学卒と年齢がそれぞれ効果を増大させているが、交互作用項を加えたことによりモデルが改善されたとはいえないであろう。

第2は、通信機器利用傾向と年齢の関係について考察した Model 15 である。Model 15 は、Model 13 に潜在変数（通信機器利用傾向）と年齢の交互作用項を追加している。「des」については、「mod」の「cov (XH,2)」「cov (GH,1)」「cov (XGH,2)」が示すように、最初の2行がXとHの関係、次の1行がGとHの関係、最後の2行がXとGとHの関係を表している。デザイン行列については、図3を参照されたい。表12により結果をみると、「携帯電話のみ利用型*年齢」がマイナスで有意であり、「携帯電話のみ利用型」では年齢が上がるほど世帯収入レベルが低くなっていることを示している。また、就労形態が有意となっている。しかしながら、全体の解釈は困難であり、交互作用項を加えた意義はあまり見いだせない。

```

【Model 15】
lat 1
man 8
dim 3 2 2 2 2 2 3 2 7
lab X A B C D E F G H
mod X|EFH {X, XE, XF, cov(XH,2)}
    A|X
    B|X
    C|X
    D|X
    G|XEFH {G, GE, GF, cov(GH,1), GX, cov(XGH,2)}
rec 1987
dum 1 1 1 1 1 2 2 1 1
des [0 0 0 0 0 0 0 0 1 2 3 4 5 6 7 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
    0 0 0 0 0 0 0 0 1 2 3 4 5 6 7
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 2 3
    4 5 6 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    0 0 0 0 0 0 0 0 0 0 1 2 3 4 5 6 7]
sta A|X [0.9 0.1 0.2 0.8 0.7 0.3]
sta B|X [0.9 0.1 0.3 0.7 0.8 0.2]
sta C|X [0.9 0.1 0.3 0.7 0.6 0.4]
sta D|X [0.8 0.2 0.1 0.9 0.1 0.9]
dat ABCDEFGH.dat
    
```

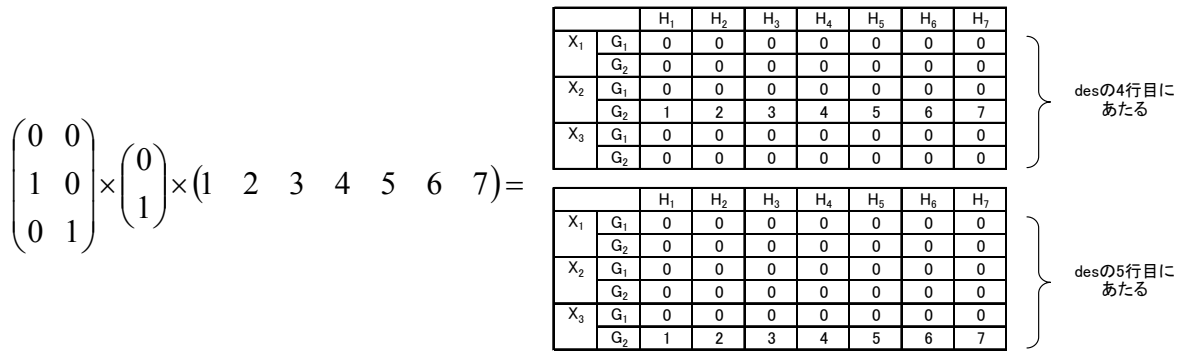


図3 Model 15 のデザイン行列の指定

表 13 に、Model13～15 の適合度検定の結果を提示した。尤度比統計量 G² の p 値が 1.000 であり適合しているものの、G² と χ² の値が乖離している。細かいカテゴリー分類を行っているにもかかわらずデータ数が少ないため、どのモデルも信頼性は乏しい。より適切にモデルの適合度を確かめるためには、変数のカテゴリーを少なくする、度数の少ないセルに小さな値を与える等の技術的な工夫が必要である。ただ、解釈には慎重を要するが、潜在変数を追加した Model 13 の結果は比較的興味深いものといえるであろう。

表 13 Model 13～Model 15 の適合度

	自由度 df	尤度比統計量 G ²	p値	Pearsonの χ ² 値
Model 13	1273	930.87	1.000	1378.54
Model 14	1269	926.23	1.000	1375.82
Model 15	1271	923.92	1.000	1364.26

6. まとめと課題

本稿では、通信機器利用傾向と世帯収入レベルの関連について、潜在クラスモデルにより分析を行った。分析セミナーの演習事例であるため、改善の余地があるが、分析の結果興味深い結果が得られた。

PC（自宅）、PC（職場）、ファックス、携帯電話といった通信機器の利用傾向は、潜在クラス分析の結果、「非利用型」「すべて利用型」「携帯電話のみ利用型」の3類型に分けられることが明らかになった。世帯収入レベルに対しては、就労状態、学歴、年齢の社会階層要因が影響を与えている。さらに、そうした要因に加えて、ここで析出した通信機器利用傾向も世帯収入レベルに影響を及ぼしている。とくに通信機器利用傾向の影響を考慮することにより、学歴、年齢といった社会階層要因が世帯収入レベルに及ぼす影響がより明確になった。

最後に、本稿で行った分析に関し若干の課題をあげておこう。第1に、潜在クラスモデルを行ううえで、分析方法のさらなる工夫が必要である。すでに指摘しているように、本稿で行った分析には適合度の点でやや問題が残るものが多い。繰り返しになるが、変数のカテゴリーの構成変更やゼロセルの処理に関して、工夫が必要である。

第2に、通信機器利用傾向と世帯収入レベルに影響を及ぼす他の要因についてより考慮すべきである。たとえば、本稿では性別の影響は検討されていない。通信機器利用、就業形態に関し性差が大きいはずである。この点に関連し、就業・非就業の区別についても慎重であるべきであろう。同じ通信機器利用傾向と世帯収入レベルの関連といっても、有職の男性回答者と無職の女性回答者では大きく意味が異なる。また「PC（職場）」も厳密には就労者のみを対象とする質問項目と考えられる。さらにこれはより本質的な問題であるが、通信機器利用傾向そのものがホワイトカラーであるか否かといった職種の違いを示している可能性がある。

第3に、モデルそのものに関しても、一考の余地がある。本稿では、図1のように、社会階層要因と通信機器利用傾向が世帯収入レベルに影響を与えていると考えている。しかし、世帯収入が通信機器利用を規定している側面も否定できない。

豊富な変数を有する JGSS データと、モデルを自在にデザインできる LEM の特徴を生かし、本稿の潜在クラスモデルを発展させることが期待される。

[Acknowledgement]

日本版 General Social Surveys (JGSS) は、大阪商業大学比較地域研究所が、文部科学省から学術フロンティア推進拠点としての指定を受けて (1999-2008 年度)、東京大学社会科学研究所と共同で実施している研究プロジェクトである (研究代表: 谷岡一郎・仁田道夫、代表幹事: 岩井紀子、副代表幹事: 保田時男)。東京大学社会科学研究所附属日本社会研究情報センターSSJ データアーカイブがデータの作成と配布を行っている。

[注]

- (1) JGSS では、今後も分析セミナーを開催する計画をたてている。分析セミナーや講習会に関しては JGSS の Web ページで告知を行う (<http://jgss.daishodai.ac.jp/>)。
- (2) 潜在クラス分析については、Goodman (1974a[1978],1974b[1978])、Clogg and Goodman (1984)、Clogg (1995)、Agresti (2002)、Hagenaars et al (2002)、Vermunt and Magidson (2005)、松田 (1988) が詳しい。また、潜在クラス分析の適用例としては、山口 (1998)、Yamaguchi (2000) を参照されたい。
- (3) ファイルの拡張子が表示されない場合は、マイコンピュータを開き、メニューバーの「ツール」→「フォルダオプション」→「表示」と進み、「詳細設定」の下の方にある「登録されている拡張子は表示しない」のチェックを外す。
- (4) 分析セミナーにおける山口先生の配布資料を参照した。
- (5) 本稿では、プログラムと分析結果を提示する。ただし、LEM ではパラメータの初期値が乱数であるため、同じプログラムを実行しても必ずしも同一の分析結果になるわけではないことに留意されたい。
- (6) 構成比率と条件付き確率は、Output の「LATENT CLASS OUTPUT」に表示される。
- (7) ファクター・モデルでは通常2つの潜在ファクター (X と Y) の独立性を仮定しないモデルから検討をするが、ここでは X と Y が独立であることを仮定したモデル (Model 3、4) のみ検討した。本稿で扱っているデータで model 3 から独立性の仮定を取り除いた場合 (mod XY {XY}とした場合)、自由度が0になるためである。潜在ファクターY と C (ファックス) の関連を仮定しないモデル (Model 4) では独立性の仮定を取り除いた分析が可能であるが、df=1、G²=0.48、p 値=.491、BIC=-7.12 であり、有意なモデルの改善は見られない。
- (8) なお、Model 2 と Model 5 の「LATENT CLASS OUTPUT」を比較するとわかるように、モデルが変わると構成比率と条件付き確率も変化する。
- (9) Output の「LOG-LINEAR PARAMETERS」の結果をまとめたものである。
- (10) 「LATENT CLASS OUTPUT」には、E、F、H の変数の各カテゴリーの潜在クラス所属率も推定されている。それを参照することにより、各クラスの特徴を把握できる。
- (11) Model 11 は、B | X のデザインを指定しているため、フリーなパラメータではなくなっている。そのため、「sta」で B | X の初期値は指定しない。

[参考文献]

- Agresti, Alan, 2002, *Categorical data analysis* [2nd ed.], Wiley.
- Clogg, Clifford C. and Leo A. Goodman, 1984, "Latent Structure Analysis of a Set of Multidimensional Contingency Tables," *Journal of the American Statistical Association*, 79 : 762-771.
- Clogg, Clifford C, 1995, "Latent Class Models," in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, ed. by G. Arminger et al., Perseus Publishing, 311-359.

- Goodman, Leo A, 1974a[1978], “The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I:A Modified Latent-Structure Approach,” in *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent Structure Analysis*, Abt Books, 281-361. (Reprinted from *American Journal of Sociology*, 7)
- Goodman, Leo A, 1974b[1978], “Exploratory Latent-Structure Analysis Using Both Identifiable and Unidentifiable Models.” in *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent Structure Analysis*, Abt Books, 403-419. (Reprinted from *Biometrika*, 61)
- Hagenaars, Jacques A. and Allan L. McCutcheon ed., 2002, *Applied latent class analysis*, Cambridge University Press.
- 松田紀之, 1988, 『質的情報の多変量解析』朝倉書店 (絶版であるが、次の Web ページで公開されている <http://www.sci.kagoshima-u.ac.jp/~ebsa/matsuda01/index.html>)
- 太郎丸博, 2005, 『人文・社会科学のためのカテゴリーカル・データ解析入門』ナカニシヤ出版.
- Vermunt, Jeroen K. 1997, *LEM: A General Program for the Analysis of Categorical Data*, Department of Methodology and Statistics, Tilburg University.
- Vermunt, Jeroen K. and Jay Magidson, 2005, “Factor Analysis with Categorical Indicators: A Comparison Between Traditional and Latent Class Approaches” in *New developments in categorical data analysis for the social and behavioral sciences*, ed. by L. Andries van der Ark et al., L. Erlbaum Associates, 41-62.
- 山口一男, 1998, 「女性における性別役割意識と社会階層, 職歴, ライフスタイルとの関連」『1995 年 SSM 調査シリーズ 3 社会移動とキャリア分析』123-155.
- Yamaguchi, Kazuo, 2000, “Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender-Role Attitudes among Japanese Women, ” *American Journal of Sociology* 105 : 1702-1740.