

大阪商業大学学術情報リポジトリ

JGSS統計セミナー2012ーパネルデータを用いた潜在クラスモデル分析ー

メタデータ	言語: ja 出版者: 日本版総合的社会調査共同研究拠点 大阪商業大学JGSS研究センター 公開日: 2019-07-22 キーワード (Ja): キーワード (En): JGSS, Latent Class, Panel Data 作成者: メールアドレス: 所属:
URL	https://ouc.repo.nii.ac.jp/records/787

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



JGSS 統計セミナー2012

—パネルデータを用いた潜在クラスモデル分析—

武内 智彦

大阪商業大学 JGSS 研究センター

JGSS Statistical Analysis Seminar: Latent Class Model Analysis Using Panel Data

Tomohiko TAKEUCHI

JGSS Research Center

Osaka University of Commerce

JGSS Research Center hosted a statistical analysis seminar on panel data analysis. This paper introduces the analysis of the “latent class model analysis” using panel data. Latent class model relates a set of observed multivariate variables to a set of latent variables. This paper describes mover-stayer model (black-white model), switching regression model, latent trajectory model, and latent growth curve model, and shows applications to the analysis of voting and response stability (mover-stayer model and switching regression model), and the marital satisfaction of wives (latent trajectory model and latent growth curve model).

Key Words: JGSS, Latent Class, Panel Data

JGSS 研究センターでは、「パネルデータ」をテーマに 2012 年度統計分析セミナーを開催した。その中でも、本稿では、「潜在クラスモデル分析」への応用を紹介している。潜在クラスモデルは観測された顕在変数と潜在変数を関連付ける分析である。本稿では、「ムーバー・ステイヤー・モデル（黒白モデル）」「スイッチング回帰モデル」「隣接ロジット潜在軌跡モデル」「隣接ロジット潜在成長曲線モデル」を説明し、その利用例（投票行動と政権支持の安定性に関する分析および妻の夫婦関係満足度の分析）を提示する。

キーワード：JGSS、潜在クラス、パネルデータ

1. はじめに

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 JGSS 研究センター主催の統計分析セミナーが、2012 年 9 月 3 日と 4 日に開催された。セミナーの講師として、2007 年の第 1 回統計分析セミナーから担当していただいているシカゴ大学社会学部の山口一男教授を招聘した。

2012 年の統計セミナーのテーマは「パネルデータ」であった。具体的にはパネル調査データ分析の長所に言及した後に、「固定効果モデル」「ランダム効果モデル」「成長曲線モデル」「ムーバー・ステイヤー・モデル（黑白モデル）」「スイッチング回帰モデル」「隣接ロジット潜在軌跡モデル」「隣接ロジット潜在成長曲線モデル」を取り扱った。なかでも、潜在クラス変数 (latent class variables) を取り入れた分析に力点が置かれた。本稿でもその手法を解説する。なお、セミナーの性格上、その内容を紹介する本稿は Yamaguchi (2008) に多くを負っている。

潜在クラスモデルは、カテゴリカルな観測変数の背後にカテゴリカルな潜在変数があることを仮定して潜在構造を読み解くモデルをいう。例えば、都村・岩井・保田・宍戸 (2008) は JGSS-2005 データを用いて通信機器利用傾向を 3 つの潜在クラス (「非利用型」「全て利用型」「携帯電話のみ利用型」) に分類し、世帯収入との関連を分析した。佐々木 (2010) は JGSS-2009 ライフコース調査を用いて子育ての悩みを 4 つの潜在クラス (「広範型」「不明瞭型」「教育型」「人格形成型」) に分類した。中澤 (2010) は両親と本人の世代間学歴移動のパターンの変化について「社会階層と社会移動に関する全国調査 (SSM 調査)」2005 年版を用いて潜在クラス分析を行い、コーホートによらず高学歴・中学歴・低学歴層の 3 つの潜在クラスが一貫して存在することを示し、両親と本人の学歴関係のパターンを見出した。藤原・伊藤・谷岡 (2012) は地位の非一貫性、格差意識、権威主義的伝統主義に関して潜在クラス分析を適用した分析を行った。例えば格差意識に関しては「日本の国際化と市民の政治参加に関する世論調査」を用いて、潜在クラスとして「積極的不平等・競争志向」「平等・福祉・競争志向」「消極的不平等・競争志向」「平等・福祉・反競争志向」に分類し、その所属に対し人々の社会経済的地位がどのような影響を与えているかを多項ロジット潜在クラス回帰によって分析した。

上記 4 文献は 1 時点データを用いた分析である。潜在クラス分析をパネルデータに適用すると、例えば夫婦関係満足度の潜在クラスごとの時系列推移の違いや、政権支持とその時間的安定性を潜在クラスごとに分析することができる。本稿はこの方法を提示する。

潜在クラスモデル分析のためのソフトウェアとしてはフリーウェアの LEM (Vermunt, 1997) が有名である。都村他 (2008) や三輪 (2009) は LEM を用いた分析の方法を具体的なプログラムと共に記載している。本稿に記載した実証分析への適用例も LEM を用いて推定を行っている。その LEM プログラムは Yamaguchi (2008) に記載されている。

本稿の構成は以下の通りである。第 2 節では 2 値変数が従属変数の場合の応答確率と安定応答確率の同時モデルを、第 3 節ではその実証分析への適用例を紹介する。第 4 節ではカテゴリー変数が従属変数の場合の潜在軌跡モデルと潜在成長曲線モデルを、第 5 節ではその実証分析への適用例を紹介する。最後に全体の内容についてのまとめを述べる。

2. 2値変数が従属変数の場合の応答確率と安定応答確率の同時モデル

本節ではムーバー・ステイヤー・モデル及びスイッチング回帰モデルについて説明する。

ムーバー・ステイヤー・モデルは、ある回答者の応答が時点を通じて安定的かを検証しつつその応答を分析する。例えばブッシュ政権支持に対して、支持と不支持のみならず、一貫して支持 (安定) をしているのか、それとも支持と不支持を繰り返している (不安定) のかを検証する。問題となるのは、観察者からは t 時点と $t+1$ 時点で支持ないし不支持が一致していても、それが安定的なものなのか、それとも不安定であるけれども偶然 2 時点で一致したのかわからないことである。ムーバー・ステイヤー・モデルはこれを解き明かす。

ムーバー・ステイヤー・モデルは黑白モデルとして定式化される。ここで「黒」は同じ応答を確率 1 で続ける確信的応答者、「白」とは時点 $t+1$ における応答が時点 t の応答とは独立の者である。先述

のように、観察者からは黑白の区別は部分的にしかわからない。Yamaguchi (2008) は黑白モデルを、第一に、説明変数を含む回帰モデルに、第二に、「白」を必ずしも $t+1$ 時点の応答と t 時点の応答が独立でなく、first-order Markov で依存する「灰」として拡張した。つまり「白」を特殊ケースとして含む「黒灰モデル」に拡張した。以下このモデルの説明を行う。

モデルのセットアップは以下のようなになる。

- ・ 応答変数 $\{Y(t)\}$ は各時点で 0 か 1 の 2 値のいずれかを取る。
- ・ X は 0 と 1 の 2 値の潜在クラス変数である。 $X=1$ なら「ステイヤー」であり、確率 1 で同じ Y の値の応答を繰り返す。 $X=0$ なら「ムーバー」であり、 $Y(t)$ の応答は $Y(t-1)$ の値に確率的に依存する。
- ・ 各個人は $\{\pi^{X|Y(1)}, \pi^{Y(t)|Y(t-1)}\}$ という確率の組み合わせで特徴づけられる。
- ・ $\pi^{X|Y(1)}$ は時点 1 での Y の値が与えられているという条件のもとで「ステイヤー」となる条件付き確率である。
- ・ $\pi^{Y(t)|Y(t-1)}$ は $X=0$ (ムーバー) である場合に、 $Y(t-1)$ の値が与えられているという条件のもとで、 $t \geq 2$ で $Y(t) \neq 1$ となる条件付き確率である。

ここで、 $\{\pi^{X|Y(1)}, \pi^{Y(t)|Y(t-1)}\}$ が与えられると、所与の $Y(1)$ に対し $Y(2), Y(3)$ の条件付同時分布 $P_{Y(2)Y(3)|Y(1)}$ は

$$\begin{aligned} P_{11|1} &= \pi_{X|1} + (1 - \pi_{X|1})\pi_{Y(2)|1}\pi_{Y(3)|1} \\ P_{10|1} &= (1 - \pi_{X|1})\pi_{Y(2)|1}(1 - \pi_{Y(3)|1}) \\ P_{01|1} &= (1 - \pi_{X|1})(1 - \pi_{Y(2)|1})\pi_{Y(3)|0} \\ P_{00|1} &= (1 - \pi_{X|1})(1 - \pi_{Y(2)|1})(1 - \pi_{Y(3)|0}) \\ P_{11|0} &= (1 - \pi_{X|0})\pi_{Y(2)|0}\pi_{Y(3)|1} \\ P_{10|0} &= (1 - \pi_{X|0})\pi_{Y(2)|0}(1 - \pi_{Y(3)|1}) \\ P_{01|0} &= (1 - \pi_{X|0})(1 - \pi_{Y(2)|0})\pi_{Y(3)|0} \\ P_{00|0} &= \pi_{X|0} + (1 - \pi_{X|0})(1 - \pi_{Y(2)|0})(1 - \pi_{Y(3)|0}) \end{aligned}$$

となる。例えば一番上の式はステイヤーが安定的に 1 を取り続ける確率と、ムーバーが偶然に 1 を取り続ける確率の和であり、2 番目の式は偶然最初に 1 を取ったムーバーがその後 1 と 0 を取る確率で構成されている。以下 7 番目まではムーバーだけの行動で、一番下の式はステイヤーとムーバーが含まれる。

ここで $Y(t)$ を 0 か 1 を取るダミー変数とする。このとき応答 $\pi^{Y(t)|Y(t-1)}$ に関するロジットモデルおよび潜在クラス $\pi^{X|Y(t-1)}$ に関するロジットモデルのペアは以下のように特定化できる。

$$\log\left[\frac{\pi^{Y(t)|Y(t-1)}}{1 - \pi^{Y(t)|Y(t-1)}}\right] = a_t + bY(t-1) + c'V(t) \quad \text{for } X = 0 \quad (1)$$

$$\log\left[\frac{\pi^{X|Y(1)}}{1 - \pi^{X|Y(1)}}\right] = \alpha + \gamma Y(1) + \beta'U \quad (2)$$

(1) 式は $X=0$ であるムーバーの間では、時点 t における Y の応答確率の対数オッズが、時間ごとの切片 a_t と、 $Y(t-1)$ の影響 b と、説明変数 $V(t)$ の影響の和で表せることを意味している。 $V(t)$ が時間に関して一定であることは排除されない。ここで $bY(t-1)$ が Markov 過程での依存の部分であり、 $b=0$ なら応答が $Y(t-1)$ から独立であるため、「黑白モデル」となる。

(2) 式は応答の安定性に関するもので、潜在クラス $X=1$ のステイヤーである確率は、 Y の初期値の影響 γ と、時間に関して一定の説明変数 U の影響 β の和で表せることを意味している。したがって、ステイヤーには $Y(1)=0$ と $Y(1)=1$ の場合の 2 種類があり、それぞれの割合は U に依存している。

「黑白」ないし「黒灰」から「灰白」にモデルを改変すると、2 値のカテゴリー変数を従属変数とするスイッチング回帰モデルとなる。なお、2 値のカテゴリー従属変数のスイッチング回帰モデルは一方の潜在クラスで Y に影響し、他方では影響しない効果的な操作変数がないと安定的なパラメータ推定ができないことが知られている。ここでは $Y(t-1)$ が操作変数となっている。潜在変数が 3 つ以上に増えるとさらに追加的な操作変数が必要となる。

このとき、モデルは以下の2式ようになる。

$$\log\left[\frac{\pi^{Y(t)|Y(t-1)}}{1-\pi^{Y(t)|Y(t-1)}}\right] = a_t + bX + cXY(t-1) + d'V(t) \quad (3)$$

$$\log\left[\frac{\pi^{X|Y(1)}}{1-\pi^{X|Y(1)}}\right] = \alpha + \gamma Y(1) + \beta'U \quad (4)$$

(3)式は $X=1$ である場合、時点 t における Y の応答確率の対数オッズが、時間 t ごとの切片 $a_t + b$ と、 $Y(t-1)$ の影響 c と、説明変数 $V(t)$ の影響の和で表せることを意味する。 $X=0$ のときは、時間 t ごとの切片 a_t と、説明変数 $V(t)$ の影響の和で表せ、 $Y(t-1)$ の値に依存しないことを意味する。

(4) 式は潜在クラス $X=1$ のステイヤーである確率は、 Y の初期値の影響 γ と、時間に関して一定の説明変数 U の影響 β との和で表せることを意味している。

また、ムーバー・ステイヤー・モデル、スイッチング回帰モデルともに、応答の安定性について $Y=1$ の場合の応答の安定性への説明変数 U の影響と $Y=0$ の場合の応答の安定性への U の影響が異なることが考えられる。これは以下のように (2) 式、(4) 式に $Y(1)$ と U の交互作用効果を加えて検定すればよい。

$$\log\left[\frac{\pi^{X|Y(1)}}{1-\pi^{X|Y(1)}}\right] = \alpha + \gamma Y(1) + \beta'U + (\delta U^*)Y(1) \quad (5)$$

(5) 式において U^* は U の一部である。

3. ムーバー・ステイヤー・モデルを用いたブッシュ政権支持率の分析

以下、第2節で提示されたモデルの分析例を示す (Yamaguchi, 2008)。American National Election Studies (ANES) の1990-92 データを用いてブッシュ政権支持とその安定性について分析する。標本は1990年に20-69歳の男女である(839人)。この間のブッシュ政権の支持率は65.7%から79.5%、41.8%と変化した。従属変数は1991年と1992年にブッシュ政権を支持したかどうか(先の $Y(2)$ と $Y(3)$ にあたる)であり、説明変数は1990年の政権支持に対する回答 ($Y(1)$ にあたる)である。それ以外に説明変数として年齢層(5カテゴリー)、人種(2カテゴリー)、教育水準(4カテゴリー)、宗教(3カテゴリー)、1990年における支持政党(3カテゴリー)を用いている。表1に説明変数の分布を示す。

表1 記述統計量

	割合 (%)		割合 (%)		割合 (%)
年齢層		教育		支持政党	
20-29	22.5	12年未満	12.7	共和党	26.7
30-39	27.8	12年未満	37.1	民主党	38.6
40-49	21.1	13-15年	24.3	その他、なし	34.7
50-59	12.7	16年以上	25.9		
60-69	16.0	信仰			
人種		プロテスタント	44.2		
白人	90.1	カソリック	26.0		
黒人	9.9	その他	15.4		
		なし	14.4		
				n	839

表2は各モデルの適合度に関する情報を示したものである。表3にムーバー・ランダム・モデル(上記(3)式と(4)式にあたる)とムーバー・ステイヤー・モデル(上記(1)式と(2)式にあたる)によるパラメーター推定結果を示す。なお、係数推定値を解釈するにあたって、ムーバー・ランダム・モデルにおいては $Y(t)$ が1と-1にコーディングされている点を考慮する必要がある。

表3の結果から以下のことが読み取れる。黒人は白人に比べてブッシュ政権の支持率は低い、しかし応答の安定性は非黒人と差がない。高齢者(60-69歳)は若者に比べて支持について差はない、また、30-39歳も支持に差はない。高齢者の応答の安定性は若者より高い。大学卒(16年以上)は高卒(12年)に比べてブッシュ政権支持率が低く、応答の安定性が高い。プロテスタントと比べて無宗教の者はブッシュ政権支持率が低く、応答の安定性が高い。民主党支持者は共和党支持者よりブッシュ政権支持率は低い。潜在変数 X と前年の応答の交差項は有意であり、潜在クラスによって前年の応答

表2 モデル診断

	パラメーター数	2*LL	p値
Models without heterogeneity in stability			
A. Mover-random model	18	-10,946.50	
B. Mover-stayer models			
B1. Mover-stayer	18	-10,950.50	
B2. Stayer-random	16	-10,959.30	
Models without heterogeneity in stability (main effects only)			
A. Mover-random model	32	-10,904.70	
B. Mover-stayer models			
B1. Mover-stayer	31	-10,913.10	
B2. Stayer-random	30	-10,916.10	
Models that add the interaction effect of party preference and Y(1) on Logit(π^X)			
A. Mover-random model	34	-10,890.30	
B. Mover-stayer models			
B1. Mover-stayer	33	-10,900.60	
B2. Stayer-random	32	-10,901.30	
Model 2-A vs. 1-A	14		41.80 <0.001
Model 2-B1 vs. 1-B1	14		37.40 <0.001
Model 2-B2 vs. 1-B2	14		43.20 <0.001
Model 2-B1 vs. 2-B2	1		3.00 >0.05
Model 3-A vs. 2-A	2		14.40 <0.001
Model 3-B1 vs. 2-B1	2		12.50 <0.01
Model 3-B2 vs. 2-B2	2		14.80 <0.001
Model 3-B1 vs. 3-B2	1		0.70 >0.5

の影響が異なることがわかる。ここではステイヤーは前年の応答に依存している。なお、安定性については、交差項も含めて解釈すると、共和党のブッシュ政権支持者と民主党のブッシュ政権非支持者は応答の安定性が高く、共和党のブッシュ政権非支持者と民主党のブッシュ政権支持者は応答の安定性が低い。これらはほぼ両モデルで共通した傾向である。

4. カテゴリー変数が従属変数の場合の潜在軌跡モデルと潜在成長曲線モデル

本節では隣接ロジット潜在軌跡モデルと隣接ロジット潜在成長曲線モデルを取り扱う。特徴は潜在変数が量的な側面に与える影響をみることにある。例えば、隣接ロジット潜在成長曲線モデルを用いれば夫婦関係満足度を「高レベル軌跡クラス」「中レベル軌跡クラス」「低レベル軌跡クラス」に分類し、各クラスのダミー変数とある説明変数と時間の交差項を作成することによって、顕在変数の時間的変化の切片と傾きを推定することができる。これにより様々な変化のパターンを発見することができる（ある潜在クラスにおいて最初は夫婦関係満足度が低いが時間の経過に伴い上昇していくなど）。両モデルにおける大きな違いは、隣接ロジット潜在軌跡モデルが共変量の効果を潜在クラスの構成の違いを特徴づけるために導入するのに対し、隣接ロジット潜在成長曲線モデルでは共変量や潜在変数の切片あるいは傾きの違いを特徴づけるために導入されることである（よって当初の状況とその後の時間変化をとらえることができる）。

まずは隣接ロジット潜在軌跡モデルから説明する。これは従属変数が順序の付いたカテゴリー変数 $Y(t)$ の場合の潜在軌跡モデルであり、潜在クラス変数が従属変数の回帰分析の特殊ケースである。従属変数は潜在クラス変数の一般の回帰モデルと以下の3点で異なる。第一に、観察される従属変数 $Y(t)$, for $t=1, \dots, T$ は潜在クラス変数の指標となり、各潜在クラスは $Y(t)$ の確率分布の異なる軌跡を代表する。第二に、潜在クラス変数は時間で一定と仮定されているので、潜在クラスの構成割合に影響を与えると仮定される説明変数は時間に依存しない変数に限られる。時間で変化する変数は直接 $Y(t)$ に影響すると仮定する。第三に、潜在クラス変数 X の $Y(t)$ への影響について、2つの X の値について、各時点で一方が他方より Y のレベルが高いか低いかのいずれかが成り立つという条件を満たすため、 X の Y への影響は等比オッズ（隣接対数オッズに対する線形の影響）の条件を満たすと仮定する。

表3 各モデルの推定結果

	スイッチング回帰モデル (Y coded 1 vs. -1)		ムーバー・ステイヤー・モデル (Y coded 1 vs. 0)	
	応答 (支持)	安定性	応答 (支持)	安定性
年齢層 (基準: 20-29歳)				
30-39	0.035	1.018	0.067	0.896 *
40-49	0.018	0.594	0.005	0.592
50-59	0.010	0.493	0.201	0.428
60-69	-0.166	1.421 ***	-0.362	1.259 **
人種 (基準: 非黒人)				
黒人	-0.645 ***	0.671	-1.280 ***	0.278
教育 (基準: 12年)				
12年以下	-0.176	-0.130	-0.177	-0.331
13-15年	-0.190	0.487	-0.375	0.406
16年以上	-0.569 ***	1.300 ***	-0.781 ***	0.736 *
宗教 (基準: プロテスタント)				
カソリック	-0.160	0.681	-0.213	0.412
その他	-0.124	-0.128	-0.196	-0.229
無宗教	-0.570 ***	0.864 *	-1.324 ***	0.719 *
政党支持 (基準: 共和党)				
民主党	-0.608 **	-0.140	-1.163 **	1.396 *
その他、あるいは支持なし	-1.094 ***	0.908	-1.667 ***	1.051
前年の応答				
Y(t-1)			0.408	
潜在変数				
X	-0.546			
潜在変数と前年度の反応の交差項				
X*Y(t-1)	2.198 ***			
初年度の反応				
Y(1)		0.584		1.328
Y(1)と政党支持の交差項 (基準: 共和党)				
民主党		-1.306 *		-2.965 ***
その他、あるいは支持なし		0.264		-0.968
Y切片				
1991	2.129 ***		3.827 ***	
1992	0.412		0.332	
X切片				
		-2.276 ***		-2.775 ***
対数尤度	-5,445.14		-5,450.29	
パラメーター数	34		33	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

ただし X の Y への影響の時間的変化は非線形を仮定する。

このとき X を潜在クラス変数、 D_m^X を、もし $X=1$ なら値 1 を取るダミー変数、さらに

$$\pi_{tj}^{Y_t} \equiv \text{Prob}(Y(t) = j)$$

$$\pi_m^X \equiv \text{Prob}(X = m)$$

として、モデルは順序ロジットモデルと多項ロジットモデルを用いて

$$\log(\pi_{tj}^{Y_t} / \pi_{t(j+1)}^{Y_t}) = a_{tj} + \sum_{m=2}^M b_{tm} D_m^X + c'V(t) \quad (6)$$

$$\log(\pi_m^X / \pi_1^X) = \alpha_m + \beta'_m U \quad (7)$$

と書ける。(6) 式において、パラメーター a_{tj} は $Y(t)$ の各時点の周辺分布を固定する。また、 $Y(t)$ の隣接カテゴリー j と $j+1$ の対数オッズに X が与える影響 b_{tm} は j に依存しないと仮定を置いている。ただし b_{tm} は時間 t に対し自由に変化するので、 X の Y への影響の時間的変化は非線形であると仮定されている。この仮定より、潜在クラス L_1 が潜在クラス L_2 に比べて Y のレベルが高いか、同じか、低いかは、 $b_{tL_1} - b_{tL_2}$ の値が正か、ゼロか、負かによって判断する。

続いて隣接ロジット潜在クラス成長曲線モデルの説明に移る。これは Vermunt and van Dijk (2001) にある潜在成長曲線モデルの変形である。ただし以下で説明するモデルは、第一にパラメーター推定値の安定性を得るために制約のないしきい値の非線形主効果を仮定していること、第二に潜在クラス間で共変量の「切片と傾きの成長」の影響が異なりうるスイッチング回帰モデルに拡張している点で Vermunt and van Dijk (2001) と異なる。

基本モデルは以下のように記述できる。

$$\log(\pi_{tj}^{Y_t}/\pi_{t(j+1)}^{Y_t}) = a_{tj} + \sum_{m=2}^M b_{tm} D_m^X + c'U + d'V(t) \quad (8)$$

$$\log(\pi_m^X/\pi_1^X) = \alpha_m \quad (9)$$

しかし(8)では軌跡の形状が共変量に依存しない。そこで時間 t と U の交差項を導入すると

$$\log(\pi_{tj}^{Y_t}/\pi_{t(j+1)}^{Y_t}) = a_{tj} + \sum_{m=2}^M b_{tm} D_m^X + c'U + d'V(t) + te'U \quad (8a)$$

となる。(8a)を成長への共変量の影響が非線形になるように拡張すると

$$\log(\pi_{tj}^{Y_t}/\pi_{t(j+1)}^{Y_t}) = a_{tj} + \sum_{m=2}^M b_{tm} D_m^X + c'U + d'V(t) + te'U + \sum_{m=2}^M f'_m U D_m^X \quad (8b)$$

となる。 D_m^X の影響は $b_{tm} + f'_m U$ となり、切片の軌跡は共変量 U によって変化するだけでなく、この変化が潜在クラスによっても異なる。このモデルは各潜在クラスが共変量の効果に対して異なる回帰係数を持ちうるという点でスイッチング回帰モデルである。さらに共変量の成長曲線の傾きに対する影響も潜在クラスごとに変わるように拡張すると

$$\log(\pi_{tj}^{Y_t}/\pi_{t(j+1)}^{Y_t}) = a_{tj} + \sum_{m=2}^M b_{tm} D_m^X + c'U + d'V(t) + te'U + \sum_{m=2}^M f'_m U D_m^X + \sum_{m=2}^M tg' U D_m^X \quad (8c)$$

となる。 D_m^X の影響は $b_{tm} + (f'_m + tg')U$ となり、切片と傾き双方が共変量と共に変わるだけでなく、この変化が潜在クラス間で異なる。これは共変量 U からみても、その効果は $c + e't + \sum_m (f'_m + tg') D_m^X$ となり、時間 t と共に成長するだけでなく、その成長は潜在クラスによっても異なる。

なお、パネルデータ分析の問題点として欠損値の発生がある。欠損値の発生は MCAR(missing completely at random)、MAR(missing at random)、NIR(nonignorable response)の3タイプに分けられる。NIRはMNAR(missing not at random)とも言い換えられる。MCARは欠損値の発生確率がモデルの全ての変数の値と無関係である状況であり、MARは欠損値の値とは無関係であるが他の変数の値と関係することは許されている。一方でNIRは欠損値の値自体が欠損の発生に関係する状況である。

例えば4時点で観察している場合を考えよう。観測値 $Y(t)$ についていくつかのサンプルにおいて $Y(4)$ が欠損であったとする。このときデータに含まれるのは $Y(1)Y(2)Y(3)Y(4)$ が揃っている場合と $Y(1)Y(2)Y(3)$ に留まる場合である。 $R=1$ or 2 を欠損発生パターンインディケーターとする。MCARは R が全ての $Y(t)$ から独立である場合であり、MARは R が $Y(1), Y(2), Y(3)$ に依存するが $Y(4)$ には依存しない場合である。NIRは R が $Y(4)$ に依存する、あるいは潜在クラス X に依存する場合である。MCARやMARの違いは周辺分布以外のパラメーター推定値に影響しないので、潜在軌跡クラスや潜在成長曲線クラスの確定に影響を与えない。しかしNIRの場合は影響がある。

NIRについては3本目の推定式を立てることで対処する。ここで $R=1$ なら $Y(4^*)=Y(4)$ 、 $R=2$ なら $Y(4^*)=L$ とする。(6)式の $Y(4)$ に代わり $Y(4^*)$ を用いる。欠損の発生が潜在変数の軌跡に依存するNIRを仮定したモデルは

$$\log(\pi^R/(1-\pi^R)) = d + \sum_{m=2}^M e_m D_m^X$$

ここで $\pi^R = \text{Prob}(R=1)$ 、と特定化される。なお、MARと特定化する場合は $\text{logit}(\pi^R)$ が観測された $Y(t)$ に依存するモデルを組む。

5. 夫婦関係満足度の潜在クラス分析

本節では、第4節で提示されたモデルの分析例を示す。Yamaguchi (2008) は夫婦関係満足度について潜在クラスへの分類や隣接ロジット潜在軌跡モデル及び隣接ロジット潜在成長曲線モデルの推定を行った。利用したデータは公益財団法人家計経済研究所の「消費生活に関するパネル調査(1993-2001)」

である。1993年に24-34歳のサンプルと1997年に追加された24-27歳のサンプルのうち、少なくとも3時点連続で夫婦関係満足度に回答している者を対象に分析している。この変数は2年ごとに調査している。注意として、この調査は女性のみが対象である。

まず、欠損値のMCAR, MAR, NIRモデルをテストする。MARモデルは3分類され、4時点{1,2,3,4}のYの観察値A,B,C,Dについて、MAR1はBCDの出現確率は欠損値Aの出現確率に依存しないが、ABCDとABCの相対出現確率はAの値に依存する。MAR2は、ABCの出現確率はDの値に依存せず、BCDの出現確率はAの値に依存しないが、BとCの値に依存する。MAR3は、ABCの出現確率は欠損値Dの値に依存しないが、ABCDとBCDの相対出現確率はDの値に依存する。MCARはRの出現確率はA,B,C,Dの値に依存せず、NIRはABCの出現確率が欠損値Dの値に依存する、あるいはBCDの出現確率が欠損値Aの値に依存するモデルである。これらのテストもLEMで可能である。テストの結果は表4にまとめてある。単純なモデルから複雑なモデルへとテストを進め、BIC(Bayesian Information Criteria)および対数尤度比検定より潜在クラス数3のMCARモデルが最も良いと判断できる。以下このモデルをもとに記述する。表5に後の分析で利用する変数(夫への経済力信頼度、夫への心の支え信頼度)の記述統計量を示す。表6に潜在クラスと結婚経過期間ごとの夫婦関係満足度の分布を示す。

表4 モデル診断

	L-squared	df	BIC
Two Latent Classes			
MCAR	431.92	851	-4,632.17
MAR-1	430.46	847	-4,600.86
MAR-2	420.96	835	-4,539.07
MAR-3	428.47	847	-4,602.04
NIR	429.17	849	-4,613.73
Three Latent Classes			
MCAR	349.77	846	-4,675.61
MAR-1	348.32	842	-4,653.30
MAR-2	338.83	830	-4,591.51
MAR-3	346.32	842	-4,655.30
NIR	347.50	842	-4,654.12
Four Latent Classes			
MCAR	343.41	841	-4,652.26
MAR-1	342.83	837	-4,629.10
MAR-2	332.44	825	-4,568.20
MAR-3	339.78	837	-4,632.14
NIR	332.85	835	-4,627.20

表5 記述統計量

	夫への信頼度(%)	
	経済力信頼度	「心の支え」信頼度
非常に信頼できる	29.2	55.5
ほどほどに信頼できる	31.0	23.9
普通	34.5	16.6
あまり信頼できない	2.9	2.9
全く信頼できない	1.8	1.1
	妻の就業状態(%)	
有業	42.7	
無業(専業主婦)	57.3	
n	380	

次に隣接ロジット潜在軌跡モデル（上述の（6）式及び（7）式）の推定結果を表7に示す。モデル1とモデル2の違いは交互作用の有無である。どちらのモデルにおいても潜在クラスが高くなるほど夫婦満足度も高まることになる。また、交互作用の係数推定値から、妻が就業している場合は夫への経済力信頼度が高まると「中レベル」ではなく「高レベル」潜在軌跡クラスに属する確率が高まり（1.209）、「低レベル」に属する可能性には影響しない（0.099=1.209-1.110）ことがわかる。妻が専業主婦である場合は夫への経済力信頼度は「中レベル」ではなく「低レベル」潜在軌跡クラスに属する確率（-1.091=-0.394-0.697）を有意に下げ、「中レベル」ではなく「高レベル」潜在軌跡クラスに属する確率には影響しない（-0.394）ことがわかる。この結果をもとに「高レベル」潜在クラスと「低レベル」潜在クラスに属する確率を示したものが表8である。

さらに、同じデータを用いて隣接ロジット潜在軌跡曲線モデル（上述の(8c)および(9)式）を推定し、結果を表9に示す。この分析においても潜在クラス数が3のモデルが採択され、「高レベル」「中レベル」「低レベル」に分類できる。表のモデル1に注目すると、共変量の効果から、専業主婦は当初は夫婦関係満足度が低いけれども、その上昇ぶりは就業している妻よりも高いことがわかる。「経済力信頼度」「心の支え」については、切片は正で傾きが負であることから、結婚後時間が経過するにつれて弱まっていくことがわかる。モデル2を見ると、潜在クラスにより「心の支え」の効果が違うことがわかり、潜在クラス1（高レベルクラス）において他のクラスより大きいことがわかる。

図1はしきい値の非線形主効果を固定し、潜在クラス2に属し「心の支え」信頼度が平均的である回答者を基準点として、夫婦関係満足度が1つ上のカテゴリーに上昇する対数オッズの時間変化を共変量の状態及び潜在クラスごとに描いたものである。なお、「心の支え」信頼度に関して「やや低い」「低い」を選んだ回答者は非常に少ないため（表5）、図からは除いている。この図から、潜在クラス1の回答者は夫婦関係満足度が他のクラスより高いだけでなく、「心の支え」信頼度の差が他のクラスより永続的に影響する

ことがわかる。これは信頼度の異なる回答者の間で、初期状態における夫婦関係満足度の差がクラス1は他の2クラスより大きいからである。

表6 潜在クラス分類

	1st-2nd Year	3rd-4th Year	5th-6th Year	7th-8th Year
Latent class 1: Size=19.1%				
非常に満足している	0.704	0.741	0.913	0.620
まあまあ満足している	0.280	0.255	0.086	0.372
ふつう	0.015	0.004	0.001	0.018
あまり満足していない	0.001	0.000	0.000	0.000
まったく満足していない	0.000	0.000	0.000	0.000
Latent class 2: Size=57.0%				
非常に満足している	0.351	0.167	0.108	0.063
まあまあ満足している	0.527	0.707	0.584	0.515
ふつう	0.110	0.123	0.278	0.331
あまり満足していない	0.011	0.003	0.029	0.082
まったく満足していない	0.001	0.000	0.001	0.009
Latent class 3: Size=23.9%				
非常に満足している	0.072	0.001	0.005	0.012
まあまあ満足している	0.415	0.102	0.156	0.252
ふつう	0.332	0.518	0.471	0.409
あまり満足していない	0.127	0.313	0.315	0.254
まったく満足していない	0.054	0.066	0.053	0.073

表7 隣接ロジット潜在軌跡モデルによる夫婦関係満足度の推定結果

	モデル1		モデル2	
潜在クラス構成への効果				
	L1 vs. L2	L3 vs. L2	L1 vs. L2	L3 vs. L2
(1) 就業状態 (専業主婦 対 有業の妻)	-0.236	-0.256	4.479 *	2.233
(2) 夫の経済力信頼度	0.593 *	-0.778 **	1.209 **	-0.394
(3) 交互作用効果: (1) × (2)			-1.110 *	-0.697
(4) 夫の「心の支え」信頼度	1.989 *	-1.451 ***	2.135 *	-1.462 ***
The Effects of L1 versus L2				
第1期	1.792 **		1.832 **	
第2期	2.929 ***		2.823 ***	
第3期	3.021 ***		3.104 ***	
第4期	2.367 ***		2.489 ***	
The Effects of L3 versus L2				
第1期	-1.810 ***		-1.841 ***	
第2期	-3.777 ***		-3.701 ***	
第3期	-1.513 ***		-1.526 ***	
第4期	-0.801 ***		-0.804 ***	
対数尤度	-2,472.36		-2,468.85	
パラメータ数	34		36	
L1: 夫婦関係満足度の「高レベル軌跡」潜在クラス L2: 夫婦関係満足度の「中レベル軌跡」潜在クラス L3: 夫婦関係満足度の「低レベル軌跡」潜在クラス				
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$				

表8 各潜在クラスに属する割合

	経済力信頼度						
	専業主婦			有業の妻			
	非常に	ほどほどに	普通以下	非常に	ほどほどに	普通以下	
「高レベル」潜在クラスの割合							
「心の支え」信頼度	非常に	0.32	0.29	0.25	0.57	0.28	0.10
	ほどほどに	0.05	0.04	0.03	0.12	0.03	0.01
	普通以下	0.01	0.00	0.00	0.01	0.00	0.00
「低レベル」潜在クラスの割合							
「心の支え」信頼度	非常に	0.01	0.04	0.11	0.03	0.06	0.11
	ほどほどに	0.07	0.19	0.41	0.19	0.29	0.38
	普通以下	0.26	0.51	0.76	0.54	0.64	0.73

実証分析に関して注意点を記す。第一に、表6と表7のモデル1と2について、パラメータ数が同じであり、対数尤度から表7の潜在成長曲線モデルのほうが説明力は高く、モデルとして採択したくなるかもしれない。しかしながら、モデル選択の判断は分析の目的に照らして行うべきである。第二に、分類されたクラスがどのような特徴を持つかは推定された構成割合から読み取るものであり、潜在クラス1に必ず指標が高い集団が分類されるわけではない。その分類はパラメータ推定でどのような初期値が与えられるかにも依存する。

6. まとめ

本稿では、2012年のJGSS統計セミナーで取り扱った内容のうち、パネルデータを用いた潜在変数モデルについてモデルの紹介と分析例の提示を行った。ムーバー・ステイヤー・モデル(黒白モデル)とその発展モデル(黒灰モデル、白灰モデル)の紹介においてはAmerican National Election Studiesのデータを用いてブッシュ政権支持とその安定性を、潜在軌跡モデルと潜在成長曲線モデルの紹介においては家計経済研究所の「消費生活に関するパネルデータ」を用いて夫婦関係満足度の潜在クラス別

の変遷を分析した。いずれにおいても、潜在モデル分析が母集団の中での個体差を許す（選挙の例においては政権支持が安定的かどうか、夫婦関係満足度の例においては満足度の軌跡が高・中・低のどのレベルにあるか）フレームワークであり、興味深い結果を提示する分析手法であることが示された。本稿を契機に潜在変数モデルに関心を持っていただければ幸いである。

表9 潜在成長曲線モデルによる夫婦関係満足度の推定結果

	モデル1	モデル2
妻の就業状態（専業主婦 対 有業の妻）		
切片	-0.251	-0.106
傾き	0.220 *	0.175 †
夫への経済力信頼度		
切片	0.590 ***	0.538 ***
傾き	-0.118 *	-0.010 †
夫への「心の支え」信頼度		
切片	1.491 ***	1.026 ***
傾き	-0.331 ***	-0.297 ***
交互作用効果；「心の支え信頼度」×「潜在クラス」× t		
L1 versus L2		1.544 ***
L3 versus L2		-0.081
潜在クラスの主効果		
The effects of L1 versus L2		
第1期	0.484	-1.049 *
第2期	1.977 ***	-0.771 †
第3期	2.990 ***	-0.293
第4期	1.973 ***	0.389
The effects of L3 versus L2		
第1期	-1.326 **	-0.775
第2期	-2.746 ***	-2.389 ***
第3期	-2.363 ***	-2.136 ***
第4期	-1.448 ***	-1.186 *
潜在クラスの構成割合		
L1	0.261	0.264
L2	0.643	0.611
L3	0.106	0.125
対数尤度	-2,444.71	-2,440.93
パラメータ数	34	36

L1：夫婦関係満足度の「高レベル軌跡」潜在クラス

L2：夫婦関係満足度の「中レベル軌跡」潜在クラス

L3：夫婦関係満足度の「低レベル軌跡」潜在クラス

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$

[Acknowledgement]

日本版 General Social Surveys (JGSS) は、大阪商業大学 JGSS 研究センター(文部科学大臣認定日本版総合的社会調査共同研究拠点)が、東京大学社会科学研究所の協力を受けて実施している研究プロジェクトである。

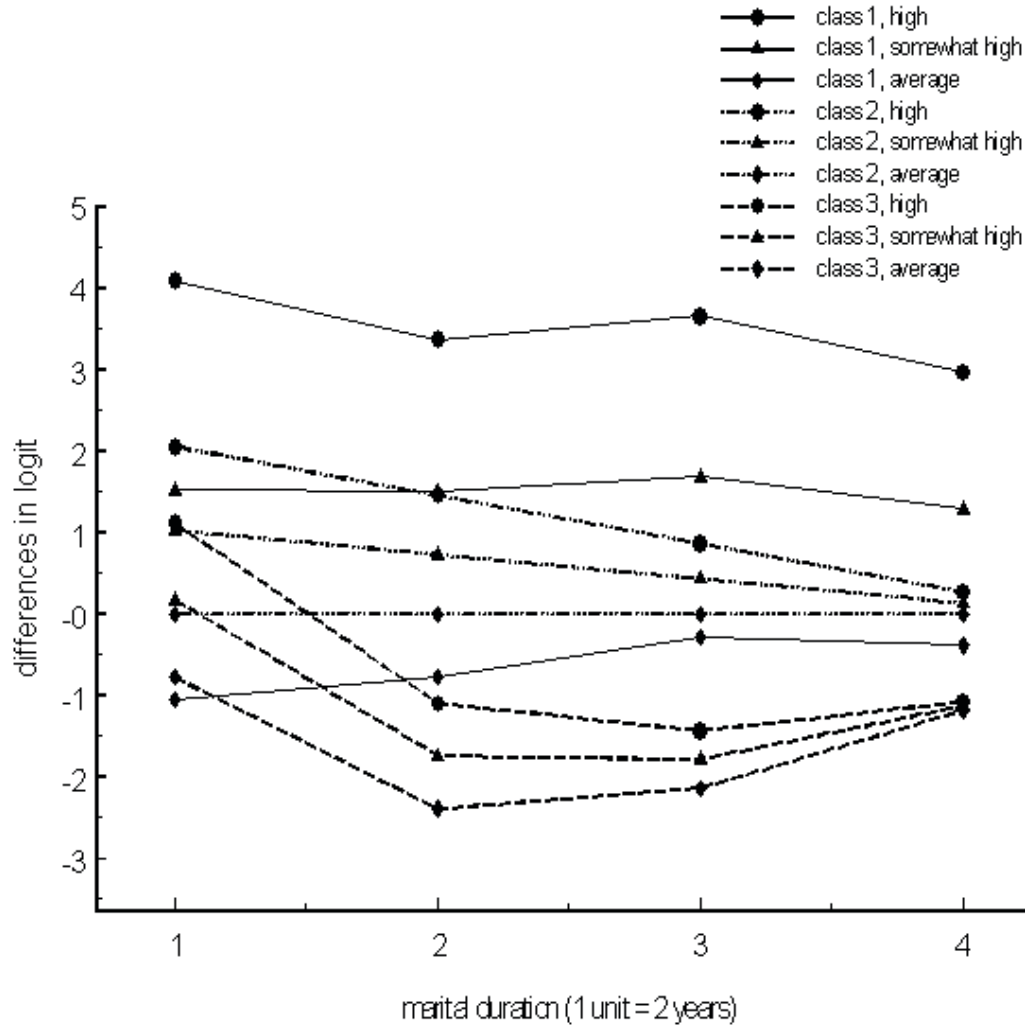


図1 潜在クラスごとの夫婦関係満足度の対数オッズの推移

[参考文献]

藤原翔・伊藤理史・谷岡謙, 2012, 「潜在クラス分析を用いた計量社会的アプローチ：地位の非一貫性、格差意識、権威主義的伝統主義を例に」『年報人間科学』 33:43-68.

三輪哲, 2009, 「潜在クラスモデル入門」『理論と方法』 24:345-356.

中澤渉, 2010, 「学歴の世代間移動の潜在構造分析」『社会学評論』 61:112-129.

佐々木尚之, 2010, 「子育ての悩みの類型：JGSS-2009 ライフコース調査による人間発達学的検証」『日本版総合的社会調査共同拠点研究論文集』 10:261-272.

都村聞人・岩井紀子・保田時男・宍戸邦章, 2008, 「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」『日本版総合的社会調査共同拠点研究論文集』7:233-249.

Vermunt, Jeroen K., 1997, *LEM: A General Program for the Analysis of Categorical Data*, Department of Methodology and Statistics, Tilburg University.

Vermunt, Jeroen K., & van Dijk, Liesbet A., 2001, "A Non-Parametric Random Coefficient Approach: The Latent-Class Regression Model," *Multilevel Modeling Newsletter*, 13:6-13.

Yamaguchi, Kazuo, 2008, "Four Useful Finite Mixture Models for Regression Analysis of Panel Data with a Categorical Dependent Variable," *Sociological Methodology*, 38:283-328.