

多時点パネルのアトリションバイアスに関する基礎的考察：

甲府買物パネルデータを用いて*

An Empirical Analysis of Attritions in Multi-Wave Kofu Shopper Panels*

西井和夫 **, 近藤勝直 ***, 古屋秀樹 ****, 栃本秀典 *****

By Kazuo NISHII **, Katsunao KONDO ***, Hideki FURUYA **** and Hidenori TOCHIGI *****

1. はじめに

交通計画の分野において、パネル分析が最近注目されるようになってきた。その背景には、動的 (dynamic) な特性を表現する交通行動モデルの必要性が認識されてきたという状況がある。従来のクロスセクショナル調査では、断面 (調査時点) ごとに標本 (被験者) が変わるため、「変化の過程」「成長」「習慣」などの動的な影響を十分表現することが困難である。一方、パネル調査は、同一個人に対し複数時点にわたって縦断的 (longitudinal) にデータを収集するものである。パネル分析においては、基本的にはそれぞれの時間断面では従来の静的 (横断的) 分析データとして活用可能であり、かつ個々のサンプルを特定してその経時的变化を縦断的に追跡できる利点をもつ。これにより、人々の活動の変化過程 (連結性) のモニタリングが可能といえ、動的な交通行動特性への理解が深まることが期待できる。

このような特徴を有するパネル分析は、動的な交通行動分析を行おうとするとき、有効な分析ツールとなり得る。しかし、一方で、調査技法としていくつかの課題点があり、調査回数を経るごとにサンプルが減少していくパネルアトリションはその代表的なものといえる。パネル調査では、時点の進行に伴い、転居や回答拒否などにより被験者が次第に脱落 (drop out) してサンプルが減少 (消耗: attrition) してしまうという状況が発生する。その結果、サンプルの属性や行動結果の分布に偏り (bias) が生じ、パネルサンプルが母集団を代表していることを単純に期待することが出来なくなる。これがアトリションの問題であり、パネル

データを用いモデルを推定する場合、この「アトリション (バイアス)」を修正しなければ、母集団とは異なった予測をしてしまう結果となる。筆者らはこれを修正する方法としてWESML法を基礎とした参加確率の逆数で重みづけする方法^{1)~3)}を提案してきた。また、杉恵ら (1993) は参加確率を用いて表現されるバイアス修正項を説明変数の一つとして取り入れたモデル⁴⁾を提案している。また、佐々木ら (1995a, 1995b) は母集団のシェアとサンプルシェアの違いを考慮したモデルの開発⁵⁾や、潜在セグメント⁶⁾を用い個人の嗜好の違いを考慮したモデルの提案を行っている。このようにバイアスを考慮した分析はすでにいくつかの例があるが、パネル調査への参加・不参加を説明する良好なモデルの提案がなされていないのが現状である⁷⁾。

本研究では、現在まで行われてきたアトリションに対する研究事例を踏まえるが、今回は特にパネル参加時点の進行に着目したアトリションモデルの検討結果を紹介するとともに、これを用いた買物場所選択モデルの同定化により、アトリションバイアス修正の有効性を検討する。

2. アトリションの実態

本研究で対象とする甲府パネルデータは、1989年の第1回目の調査 (wave1) に始まり、それ以後毎年秋に実施され、昨年1995年の調査 (wave7) まで計7時点にわたるデータが得られている。この調査は、調査日当日に対象SC来訪者に対して直接配布する来訪調査と、同時に前時点の参加者に対しては郵送により調査票を配布するパネル調査の2種類ある。したがって、各時点ごとに新たな来訪サンプルを加えながら、それぞれのパネルサンプルの時点が進んでいく形式になっている。この時点の進行とともに、この調査特有の問題であるといえるサンプルの消耗が生じ、アトリションを考慮する必要が出てきたのである。

*Keywords: 交通行動分析、パネルデータ

** 正会員 工博 山梨大学工学部土木環境工学科
(甲府市武田 4-3-11 TEL&FAX 0552-20-8533)

*** 正会員 工博・商博 流通科学大学情報学部
(神戸市西区学園西町 3-1 TEL078-794-3517 FAX078-794-3054)

**** 正会員 工修 山梨大学工学部土木環境工学科
***** 学生員 山梨大学工学部土木環境工学専攻

今回考慮するデータは、調査への参加回数及び時点の進行に着目したものであり、今まで行ってきた1時点のアトリションから、多時点を考慮したものになっている。具体的な分析対象としてのデータセットには、

- (1) 来訪→1時点後 (t時点→t+1時点)
- (2) 1時点後→2時点後 (t+1時点→t+2時点)
- (3) 2時点後→3時点後 (t+2時点→t+3時点)

を用いる。これを表-1の甲府買物パネル調査データセットに示す。表中の斜線部は新規調査参加者 (refreshments) であり、時点の経過とともに右方向に進んでいく。下段括弧内はt-1時点 (前時点) のサンプル数に対するt時点での調査への不参加 (drop out) の割合である。なお本研究では、パネルデータとしての最初の母集団である来訪時のサンプルをプールしたものをt時点、以後時点が進むごとにt+1時点、t+2時点、t+3時点としている。筆者らがこれまで行ってきたアトリションに関する研究は調査の参加時点に的を絞ったものであった。そこで今回は、調査参加時点は同質なものとし、経過する時点の進行に着目したアトリション分析を試みる。

3. アトリションモデル

まず、アトリションに関する分析として、パネル調査への参加・不参加行動に関するロジットモデルモデルを構築する。これにより、アトリションの規定要因構造の記述を行う。参加と不参加を外的基準とし、説明変数は表-2に示した計13項目とした。

また具体的な対象サンプルであるが、t時点での次の時点への参加・不参加 (モデル1、wave4,5,6の来訪3時点のサンプルのプーリング)、t+1時点での参加・不参加 (モデル2)、t+2時点での参加・不参加 (モデル3) についてそれぞれロジットモデルパラメータ推計を行った。その結果を表-3に示す。

まず、全体的な適合度を示す尤度比 (ρ^2 値) は3モデルとも小さく、特に来訪時点 (モデル1) で低い値となった。これは、t時点での調査は、パネル調査としては第1回目となり、この被験者の中には、調査日にたまたまSCに来訪し、さほどSCに対して関心・なじみがなく、翌年のパネル調査時に、SCを対象とした本調査にも興味をもたない者が多く存在する可能性が高い。そのため、このような被験者の影響により、結

表-1 甲府買物パネル：アトリションの割合およびデータセット

	wave1	wave2	wave3	wave4	wave5	wave6	wave7
wave1	653 (-)	221 (66.2%)	153 (30.8%)	117 (23.5%)	86 (26.5%)	71 (17.4%)	52 (26.8%)
wave2	323 (-)	150 (53.6%)	101 (32.7%)	67 (33.7%)	53 (20.9%)	39 (26.4%)	
wave3		357 (-)	159 (55.5%)	104 (34.6%)	67 (35.6%)	55 (17.9%)	
wave4			364 (-)	185 (49.2%)	128 (30.8%)	96 (25.0%)	
wave5				371 (-)	146 (60.6%)	101 (30.8%)	
wave6					357 (-)	198 (44.5%)	
wave7						432 (-)	

上段：サンプル数
下段 (括弧内)：t-1時点に対するt時点のアトリション割合

斜線部：refreshments, 網目部：panel sample

分析対象データセット
 t→t+1 (モ*ル1)
 t+1→t+2 (モ*ル2)
 t+2→t+3 (モ*ル3)

表-2 アトリションモデルの説明変数

説明変数	カテゴリ	説明変数	カテゴリ
1. 性別	P 男性=0 女性=1	7. 活動目的	P 単目的=0, 複合目的=1
2. 年齢	P 30歳未満=1, 30歳以上40歳未満=2, 40歳以上50歳未満=3, 50歳以上=4	8. 同伴者	P 同伴者あり=1, なし=0
3. 子供の有無	P 18歳未満人数	9. 滞在時間	P 30分以下=1, 31~60分=2, 61~90分=3, 91~120分=4, 121分以上=5
4. 自家用車保有	P 保有台数	10. 購買品目	P 食料品のみ=0, その他=1
5. 居住年数	P 5年未満=1, 6年~10年=2, 11年~15年=3, 16年~20年=4, 21年以上=5	11. 交通利便性 (SC評価-最寄店評価)	P 評価値
6. 利用交通手段	P 自家用車=1, その他=0	12. 活動利便性 (SC評価-最寄店評価)	P 評価値
		13. 固有ダミー	P 1

P：参加期への固有変数

表-3 アトリションモデルのパラメータ推計結果

	モデル1	モデル2	モデル3
説明変数	t→t+1	t+1→t+2	t+2→t+3
1. 性別	0.167 (0.84)	0.504 (1.35)	0.458 (1.06)
2. 年齢	0.399 (4.13)	0.427 (2.23)	0.592 (2.59)
3. 子供の有無	-0.010 (-0.13)	0.187 (1.11)	0.276 (1.34)
4. 自家用車保有	0.024 (0.27)	0.048 (0.26)	0.325 (1.44)
5. 居住年数	0.011 (0.19)	0.187 (1.54)	0.010 (0.07)
6. 利用交通手段	-0.195 (-0.74)	0.042 (0.10)	0.581 (1.16)
7. 活動目的	0.094 (0.56)	-0.216 (-0.63)	0.196 (0.49)
8. 同伴者	0.097 (0.46)	-0.329 (-0.87)	-0.204 (-0.44)
9. 滞在時間	-0.121 (-1.51)	0.191 (1.07)	-0.493 (-2.77)
10. 購買品目	0.146 (0.74)	-0.899 (-2.33)	0.462 (1.13)
11. 交通利便性	0.048 (1.31)	-0.030 (-0.47)	-0.077 (-0.89)
12. 活動利便性	0.002 (0.05)	-0.132 (-2.03)	0.027 (0.30)
13. 固有ダミー	-0.765 (-1.81)	-0.701 (-0.87)	-1.456 (-1.50)
L(0)	-450.79	-144.72	-111.48
L(θ)	-435.30	-130.57	-100.88
ρ^2 値	0.033	0.092	0.088
的中率	59.6%	66.1%	75.4%
有効サンプル数	653	227	187

() 内の値はt値

果としてはモデルの適合度が低くなると考えられる。そして、時点が進めば、このような被験者は減少し、次第に適合度も高くなる傾向にある。(ただし、モデル3はモデル2より良くない。)

また、説明変数に着目してみると、3モデルのすべてにおいて「年齢」が有意なt値を示している。これは、アトリションを記述していく上で被験者の年齢が重要な要因となっていることがわかる。これは、参加・不参加の年齢構成を見ると明らかである。また、時点が進んだモデル2では「購買品目」と「活動利便性」が、モデル3では「滞在時間」といった変数も有意となり、パネル調査からの離脱の要因が年齢といった個人属性を表すものと同時に買物活動や個人のSCへの評価も要因となることがわかる。

このように、時点の進行により参加・不参加行動に影響を与える要因とその構造に違いがあることが示された。また、全体的な結果については、適合度の悪さに問題はあるものの、ある程度はアトリションの規定要因を表現できたと考えられる。

4. アトリションを考慮した買物場所選択モデル

次に、アトリションの影響を考慮した買物場所選択モデルを構築する。このモデルは、外的基準がSCと非SC(具体的には、調査において普段良く利用する買物場所がどこであるのかを聞いており、それが対象SC

でない人は非SCとなる)の2項ロジットモデルである。そして、モデル構築の過程において、先ほどのアトリションモデルの推計結果を用いて、サンプルに重みをつけていく重みづけ法を用いる。この方法は以下の式に示すように、買物場所選択モデルの尤度関数の定式化において、パネル調査への参加確率の小さい被験者はどうエイトが大きく推計されるというものである。これにより、アトリションの影響を考慮することができる。

$$W_n = 1/P_n$$

W_n : 個人nのもつウエイト

$$P_n: \text{次調査への参加確率} = \frac{1}{1 + \exp(\theta \bar{Z}_n)}$$

θ : アトリションモデルのパラメータ

Z_n : アトリションモデルの説明変数

$$L^* = \prod_{n=1}^N \left(\frac{1}{1 + \exp(\beta X_n)} \right)^{W_n \delta_{in}} \left(\frac{\exp(\beta X_n)}{1 + \exp(\beta X_n)} \right)^{W_n \delta_{2n}}$$

$$= \prod_{n=1}^N \left(\frac{\exp(\beta X_n)}{1 + \exp(\beta X_n)} \right)^{W_n (1 - \delta_{in})}$$

β : 買物場所選択モデルのパラメータ

X_n : 買物場所選択モデルの説明変数

データセットは、アトリションモデルに対応して、t時点での買物場所選択(モデル4)、t+1時点(モデル5)、t+2時点(モデル6)とした。これより、買

表-4 買物場所選択モデルのパラメータ推計結果

()内の値はt値

説明変数	モデル4(t→t+1)		モデル5(t+1→t+2)		モデル6(t+2→t+3)	
	Non-Weighted	Weighted	Non-Weighted	Weighted	Non-Weighted	Weighted
1. 18歳未満人数	0.011 (0.11)	-0.014 (-0.22)	-0.041 (-0.23)	-0.009 (-0.09)	0.262 (1.24)	0.421 (3.99)
2. 自家用車保有	0.227 (2.20)	0.294 (4.33)	-0.367 (-1.90)	-0.353 (-3.44)	0.250 (1.18)	0.299 (3.07)
3. 活動目的	0.252 (1.31)	0.249 (1.96)	1.085 (3.22)	1.072 (5.81)	1.283 (3.54)	1.666 (9.10)
4. 同伴者	-0.043 (-0.48)	-0.031 (-0.50)	0.209 (1.18)	0.324 (3.35)	-0.113 (-0.57)	-0.220 (-2.19)
5. 購買品目	-0.121 (-0.54)	-0.038 (-0.26)	0.589 (1.60)	0.622 (3.36)	1.436 (3.66)	1.359 (6.93)
6. 自宅～買物場所距離	-0.331 (-8.69)	-0.344 (-13.35)	-0.216 (-3.35)	-0.179 (-4.85)	-0.170 (-2.44)	-0.229 (-6.05)
7. 交通利便性	0.200 (4.12)	0.172 (5.39)	0.346 (4.54)	0.394 (8.26)	0.259 (2.60)	0.501 (8.30)
8. 活動利便性	0.028 (0.60)	0.011 (0.36)	0.024 (0.36)	0.098 (2.50)	0.062 (0.69)	0.124 (2.85)
9. 固有ダミー	-0.660 (-2.27)	-0.543 (-2.87)	0.289 (0.60)	0.610 (2.54)	2.004 (3.61)	2.185 (7.93)
L(0)	-401.86	-401.86	-152.97	-152.97	-129.29	-129.29
L(θ)	-333.47	-334.08	-120.03	-121.52	-95.22	-99.00
ρ ² 値	0.167	0.166	0.206	0.198	0.251	0.223
的中率	72.2%	72.6%	75.4%	73.7%	72.7%	74.9%
有効サンプル数	594	594	224	224	187	187

物場所選択の経年的変化も表現できる。表-4にそのパラメータ推計結果を示す。

まず、重みづけしないケース間の比較をすると、パネルの時点が進むにつれて、全体の適合度 (ρ^2 値) が高くなっていることがわかる。今、ケース間ではパネルの進行とともに多時点残留者の割合が増加する傾向にあることから、この結果は、これらの対象サンプルの買物場所選択行動についてここで取上げている説明変数群がよく説明できていることを示唆している。具体的には、 $t \rightarrow t+1$ 時点で有意な説明要因としては、距離、交通利便性、固有ダミー、車保有であったが、 $t+1 \rightarrow t+2$ 時点では、固有ダミーの代わりに活動目的となり、さらに $t+2 \rightarrow t+3$ 時点では、購買品目、活動目的、交通利便性、距離、固有ダミーが規定力の大きな要因となっている。

一方、重みづけケースの結果に触れよう。重みづけしないケースはパネル残留者の買物場所選択行動特性をいずれも表わしているのに対して、重みづけケースは、アトリションバイアスを除去した母集団に対応する買物場所選択行動を表現するものと考えている。

重みづけ3ケースの推定結果をみると、全体の適合度 (ρ^2 値) は、重みづけしないケースのそれぞれとほとんど差異はないが、 $t+1 \rightarrow t+2$ 時点と $t+2 \rightarrow t+3$ 時点の2ケースで各パラメータの t 値が大きく推定されていることがわかる。これは、前節の定式化における尤度関数からわかるように、重みをつけた場合に t 値が大きくなることが一般的であるといえる。

ここで重要なのは、重みづけの有無で推計される買物場所選択モデルのモデル構造が有意に異なるかどうかである。まず $t \rightarrow t+1$ 時点では、重みづけの有無による主たる説明要因の各パラメータ値に違いはなく、したがって両者の意志決定構造の差異は認められない。これに対して、 $t+1 \rightarrow t+2$ 時点では、有意なパラメータ値を示す変数のうちで、活動利便性や固有ダミーの係数値が大きくなっており、また $t+2 \rightarrow t+3$ 時点でも交通利便性、活動利便性、18歳未満人数といった変数の推計パラメータ値が大きくなっている。これらより、 $t \rightarrow t+1$ 時点は、アトリションモデルの同定化が必ずしも十分できていないことも影響していると考えられるが、アトリションバイアスを修正した前後での差異は小さいこと、一方パネル時点が進むにつれてアトリションバイアスを考慮した修正法の適用によって修正前に比較して、個々の説明要因に対してより敏感な、

そして SC 評価や買物活動といった要因の規定力が無視できない選択構造の推計結果を得ることができたといえる。

5. おわりに

本研究では、多時点パネルにおけるアトリションバイアスの修正について、パネル参加時点の進行に着目したアトリションモデルを提案した。これにより、アトリションを規定する要因が時点の進行とともに変化することが示された。また、重みづけによる買物場所選択モデルにおいては、個々の説明要因に対してより敏感な推計結果を得ることができた。しかし、アトリションモデルに関しては、未だに参加・不参加行動を十分に記述するものが開発されているとはいいがたい。

今後の課題としては、このパネルアトリションによる影響をより明確に表現できる洗練されたモデルの開発が望まれる。アトリションを規定する潜在要因間の関係を取り込んだ因果構造モデルの開発などの別途の分析ツールの適用もその一つであると考えられ、今後の研究課題としたい。

<参考文献>

- 1) Kazuo NISHII, Katsunao KONDO: Panel Analysis of Shopping Destination Choice Behavior in Japan, Paper presented at The First US Conference on Panels in Transportation, Arrowhead, CA., 1992.
- 2) 西井和夫, 古屋秀樹, 鈴木隆: 甲府買物パネル調査データにおけるアトリション分析, 日本行動計量学会第22回大会発表論文抄録集, pp64-67, 1994.
- 3) 西井和夫, 近藤勝直, 古屋秀樹, 鈴木隆: パネルアトリションを考慮した買物場所選択モデル, 土木計画学研究・論文集, No.12, pp389-396, 1995.
- 4) 杉恵頼寧, 藤原章正, 山根啓典: 選好意識パネルデータに潜在する消耗バイアスの修正, 土木計画学研究・論文集, No.11, pp311-318, 1993.
- 5) 佐々木邦明, 杉山幸司, 森川高行: パネルサンプルの母集団代表性を考慮した買物目的地選択モデル, 土木計画学研究・講演集, No.18 (2), pp493-496, 1995a.
- 6) 佐々木邦明, 森川高行, 杉本直: 潜在セグメントを考慮した動的な休日買物目的地選択分析, 土木計画学研究・論文集, No.12, pp397-404, 1995b.
- 7) 平成5・6年度科研総合研究 (A) (飯田恭敬): 交通計画におけるパネル調査データの適用性に関する研究, 1995.