

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* Shopping Place Choice Analysis Considering Representativity of Panel Samples

佐々木邦明**, 杉山幸司***, 森川高行****

By Kuniaki SASAKI, Kohji SUGIYAMA and Takayuki MORIKAWA

1. はじめに

個人の行動履歴を表すパネルデータの非集計行動モデルへの適用に関する研究が近年数多くなされ、パネルデータを用いることの利点と共に欠点も明らかになってきた。その中でパネルデータを用いることの問題点としてあげられているのがアトリションバイアスである。一般にパネルサンプルは転居、死亡などの物理的要因や調査拒否などによって調査回数が増すにつれて減少していく。このとき、パネル調査より脱落したサンプルが分析対象となる事象に対し何らかの特徴を持っている場合、パネルサンプルは母集団を正確に表さなくなる。このようなデータを用いてモデルを推定した場合、そのパラメータは不偏性を失ってしまう。通常このバイアスはアトリションバイアスと呼ばれ、パネル分析の創成期よりこの問題は指摘されてきており、アトリションバイアスを考慮した分析手法が開発されてきている¹⁾。しかしパネル分析の歴史自体がまだ浅く、決め手となる手法が確立されていないのが現状である。そこで本研究は、休日における買物目的地選択行動を分析対象とし、どのような属性の人がどのような確率でパネル調査から脱落し、それによってどのようなバイアスが発生するのかを分析し、パネルサンプルの母集団代表性を考慮したパネル分析手法の開発を目的とする。

2. パネルアトリションと母集団代表性

非集計行動モデルの推定に用いられる代表的なサンプリング手法としては、ランダムサンプリング、層別抽出法、選択肢別抽出法があり、それぞれの抽出方法の特性に応じたモデルの推定方法やパラメータの修正方法が提案されている。それらのほとんどは母集団のシェアとサンプル中のシェアの違いに着目したものであり、母集団のシェアが不明であるときには、サンプルを無作為抽出とみなしてモデルを推定する場合が多い。しかし、一般に交通機関選択などの場合は母集団中の交通機関別のシェアなどが容易に判明するために、先に述べた母集団中のシェアを考慮して推定値の修正が行えるが、母集団シェアが不明な適用対象も多い。

ここでパネルサンプルの磨耗について再び考えてみる。サンプルの脱落が属性や選択に対してランダムな場合には、無作為抽出標本によって推定されたモデルはアトリションの影響を受けず、最尤推定法によって推定されたパラメータは漸近的有効性と一致性を保つために問題は生じない。しかし、これまでのパネルサンプルの磨耗についての研究によると、モビリティの高い人が脱落し易いなどある特性を持ったサンプルが脱落してゆくという報告がなされている²⁾。つまりパネルサンプルのアトリションの影響とはある特徴を持ったグループのサンプル中でのシェアが母集団でのシェアに一致しなくなるということであり、無作為抽出がある種の層別抽出標本に変化してゆく現象であると考えることができる。パネルアトリションによりサンプルが減少してゆくことは、そのままモデルを推定する場合にはパネルデータの欠点としてあげられるが、その減少してゆく状況を分析することによって逆に各属性の母集団中でのシェアを推定することが可能になると考えられる。

* Key Words : 交通行動分析, 地区交通計画, 調査論

**正会員 工修 名古屋大学助手 工学部土木工学科
(〒464-01 名古屋市千種区不老町)

Tel. 052-789-3565 Fax. 052-789-3738)

***正会員 工修 建設技術研究所 東京支社道路本部・
道路計画課

(〒103 東京都中央区日本橋本町4-9-11)

Tel. 03-3668-0451 Fax. 03-5695-1883)

****正会員 Ph.D. 名古屋大学助教授 工学部土木工学科
(〒464-01 名古屋市千種区不老町)

Tel. 052-789-3564 Fax. 052-789-3738)

3. パネルアトリションによるサンプルの変化

山梨大学の西井研究室が中心となって行っている甲府パネル調査³⁾によって得られたデータを対象にパネルサンプルの構成の変化を分析する。本調査は郊外立地型の大型ショッピングコンプレックス（以下SC）への来訪者に対して配布され、郵送によって回収されているもので、1989年に第1回の調査が始まり、本年度で七回目の調査が行われている。本研究ではその中のWAVE5までのデータを用いて分析を行った。

まずはじめに、アトリションによって各サンプル中の属性の割合が実際に変化しているかを確認した。その中の一例としてパネル滞留回数別の性別シェアを図-1に示す。

図から分かるようにパネルへの残留という事象は性別に対してランダムに生起するものではない。よってパネルアトリションを考慮せずにモデルを推定した場合、パラメータがバイアスを持つと考えられる。

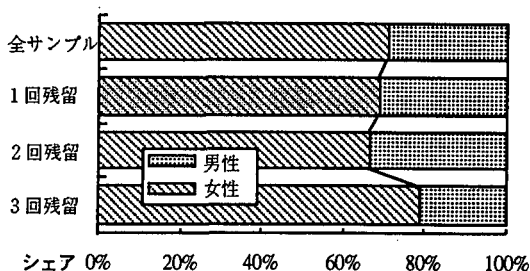


図-1 パネル滞留回数別の性別のシェア

4. アトリションによる影響を考慮した選択モデル

(1) モデルのフレームワーク

3. で用いた甲府パネル調査には以下のようなデータが含まれている。

- 1) 最近1～2カ月の休日に最もよく利用する買物場所
- 2) 回答者の属性や世帯の属性など
- 3) 自宅から買物場所への距離や満足度など

これらのデータを用いてパネル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し、パネルサンプルの母集団代表性について考察を行う。

以下にパネルサンプルの母集団代表性を考慮したモデルの定式化を行う。 t 時点において個人 n がパネル調査に滞留するか否かは既存の研究よりライフサイクル

ステージなどの個人や家庭の属性によって影響を受けていることが報告されている⁴⁾。そこで t 時点で個人 n がパネルに参加する意向を示す潜在変数 A_n^t を $t-1$ 時点の個人属性によって表現し、各 t 時点での個人 n の買物目的地選択行動を t 時点の個人属性と目的地の属性によって定式化する。ただし、これまでの研究によると、同一個人から時系列的に得たデータには系列相関と状態依存性が存在することが指摘されている⁵⁾。本研究では森川ら⁶⁾によって提案された系列相関項として同一個人に対し時点間で共通の誤差項を導入するモデルを採用して系列相関の影響を取り除く。ただし、本モデルではパネル調査残留モデルと買物目的地選択モデルの2つがあり、同一時点でのパネル残留意向を表す変数と買物目的地に対する効用との間に存在する系列相関と、多時点間の買物目的地に対する効用の間に存在する系列相関の2つがあると考えられ、それぞれ独立と仮定し以下のような式になる。

$$A_n^t = \Gamma' X_n^{t-1} + \lambda_n + \gamma$$

$$U_{ni}^t = B' Z_n^t + \lambda_n + \theta_{ni} + \varepsilon_{ni}$$

ただし、

X_n^t : 個人 n の t 時点での個人属性

Γ : 未知パラメータベクトル

γ : 誤差項

U_{ni}^t : 選択肢 i に対する個人 n の t 時点での効用

Z_n^t : 時点 t での選択肢 i と個人 n の属性

B : 未知パラメータベクトル

ε_{ni} : 個人 n の選択肢 i の効用に対する誤差項

λ_n : パネル残留変数と買物目的地の効用の間に存在する系列相関

θ_{ni} : 各時点間の買物目的地の効用に存在する系列相関

ここでそれぞれの誤差項にIID.ガンベル分布を仮定し、ランダム効用理論を適用することによって、2つの系列相関項が既知のもとでの個人 n のパネル残留確率は以下のようにになる。

$$P_{ni}(\text{stay}|\lambda_n) = \frac{[\exp(\Gamma' X_n^{t-1} + \lambda_n)]^{\delta_{ni}}}{\exp(\Gamma' X_n^{t-1} + \lambda_n) + 1} \quad (6)$$

ただし、 δ_{ni} はパネルへの残留を示すダミー変数であ

る。また、個人 n が t 時点で買物目的地 i を選ぶ確率は以下のように表される。

$$P_{ni}(i|\lambda, \theta) = \frac{\exp(\mathbf{B}'\mathbf{Z}'_{ni} + \lambda_n + \theta_n)}{\sum_{k \in C_n} \exp(\mathbf{B}'\mathbf{Z}'_{nk} + \lambda_n + \theta_n)} \quad (5)$$

ただし、 C_n は個人 n の選択肢集合

また、これらより、個人 n のパネルへの残留と買物目的地選択の同時確率は以下の式ようになる。

$$P_n = \int \int \cdots \int \int (Q_n^1)^{d_1} \cdot (C^2)^{d_1} \cdots (Q_n^{t-1})^{d_{t-1}} \cdot (C^t)^{d_{t-1}} \cdot f(\lambda) d\lambda d\theta \quad (7)$$

ただし

$$Q_n^t : P_{ni}(\text{stay}|\lambda_n)$$

$$C_n^t : P_{ni}(i|\lambda_n, \theta_n)$$

d_t : 1, t 時点でサンプルが存在 0, その他

$f(\lambda)$: λ に関する同時確率密度関数

(7)式を尤度関数として最尤推定法により未知パラメータを推定する。ここで、各属性ベクトルごとの磨耗確率が各WAVEで一定だと仮定すると、以下のような関係が成立する。

$$N_{ij} = H_j \times Q_j \quad (8)$$

ただし

N_{ij} : t 時点での j グループの残留人数

H_j : ランダムサンプリングされた場合のグループ j のサンプル数

(8)式より各 H_j をもとめ、各グループのウェイトを次の式で計算する。

$$w_{ij} = \left(H_j / \left(\sum_{j=1}^J H_j \right) \right) / \left(N_{ij} / \left(\sum_{j=1}^J N_{ij} \right) \right) \quad (9)$$

これを用いてWESML推定量と同様な重み付きロジットモデルを推定することが可能になる。

$$L^* = \sum_{j=1}^J \sum_{n=1}^{N_j} \sum_{i \in C_n} \delta_{i,n} w_{ij} \ln P(i|\mathbf{X}_n) \quad (10)$$

(2) 推定結果とその考察

甲府パネルデータを用いて(1)で定式化したモデルの有効性を検討する。買物目的地選択モデルはSC、甲府市中心部の商店街、最寄りのスーパーの3項選択で、その説明変数は同じデータを用いたこれまでの研

究7)と同様のものを用いる。磨耗モデルは3、でサンプル滞留回数に応じたサンプル中の属性シェアの変動に特徴が見られるものをいずれもダミー変数として説明変数に用いたモデルの推定結果を表一に示す。

表一 系列相関を考慮したサンプル滞留と買物目的地選択モデルの推定結果

説明変数	推定値	t 値
買物目的地選択	SC定数項3	2.68
	中心街定数項3	-0.763
	SC定数項4	2.47
	中心街定数項4	-1.34
	SC定数項5	2.41
	中心街定数項5	-1.58
	距離	-0.0201
	交通利便性満足度	0.191
パネル滞留	買物利便性満足度	0.212
	買物金額ダミー	0.293
	自動車複数保有ダミー	0.0359
	居住年数ダミー	-0.408
	性別ダミー	-0.242
その他の統計量	年齢ダミー	-0.215
	サンプル数1133	$\rho^2: 0.330$

ただし

買物金額ダミー： 買物金額7500円以上 1
買物金額7500円未満 0
自動車複数保有ダミー： 自動車保有台数2台以上 1
自動車保有台数1台以下 0
居住年数ダミー： 居住年数10年以上 1
居住年数10年未満 0
年齢ダミー： 40歳以上 1, 40歳未満 0

この推定結果の特徴は以下のようにまとめられる。

- 1) 目的地選択モデルのパラメータ推定値は合理的な符号を持ち、t値も良好な値を示している。
- 2) SCの定数項が正の符号を、中心街の定数項が負の符号を持っている。つまり、甲府市中心街に対する潜在的評価は、総合的には最寄りスーパーより低いと考えるサンプルが多いことになる。
- 3) 距離に関するパラメータ推定値がそれほど大きくなく有意性も他の変数に比べて低い、これは甲府などの地方中核都市では自動車の利用を前提に都市の構造が決定されており、自由度の高い買物目的地選択に対して距離がそれほど影響していないことを示していると考えられる。
- 4) パネル滞留モデルのパラメータは、t値の高いも

のが少なく、推定値もさほど大きなものはないが、推定値の符号は3. のシェアの変化から予想された通りである。

このパネル滞留モデルより式(9)を用いてサンプルの母集団シェアを求め(10)式を用いて未知パラメータを求めた結果と、比較のため同じスペシフィケーションで重み付けしないモデルの推定結果を表-2に示した。二つの推定結果を比較して、最も特徴的な点は、適合度を示す尤度比が重みを付けたモデルの方が格段に向上している点である。これは脱落傾向の強いサンプルが高いウェイトを持つため、見かけ上サンプル数が増加したように計算され、初期尤度が低下することによるものと考えられる。また2つのスケールパラメータについてみると、いずれも絶対値が1に近づき、各WAVEの誤差分散の比が小さくなった。一方、定数項の減少傾向に変化はないが、それぞれの推定値の絶対値が小さくなり減少する幅が逆に大きくなった。また、他のパラメータは一様にその絶対値が小さくなった。これは、真の誤差の分散が増加したことを示していると考えられる。

5 まとめ

本研究はパネルアトリションを利用して個人属性の母集団での分布を推定し、母集団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を行った。しかし、事例研究の推定結果にそれほど大きな変化はみられず、最終尤度もほぼ同じであった。これは、本研究で提案した方法が、磨耗確率を時点に関わらず一定と仮定しているため、第一回目の調査から第二回目の調査に移るときに観測される、サンプルの大幅な減少をうまく説明できず、その結果サンプルの磨耗の特質をうまく表現できていないことによるものと考えられる。また本研究の課題としては、先に述べたように第一回目から第二回目の調査でみられる大きな減少をうまく記述できるモデルの開発や、WESML推定量に代わる母集団代表性を考慮した推定方法の開発、および本研究で提案した手法では、選択モデルを2回推定することになるが、最初に推定されたモデルをどう評価するかなどがあげられる。

表-2 サンプルの母集団シェアで重み付けした買物目的地選択モデルの推定結果

説明変数	重み付けモデル	重みなしモデル
	推定値 (t 値)	推定値 (t 値)
SC定数項3	0.372 (2.8)	0.336 (2.4)
中心街定数項3	-0.651 (-3.2)	-0.778 (-3.7)
SC定数項4	0.175 (1.7)	0.139 (1.3)
中心街定数項4	-0.894 (-2.8)	-0.981 (-3.0)
SC定数項5	-0.0841 (-0.4)	-0.130 (1.3)
中心街定数項5	-1.22 (-3.1)	-1.30 (-3.5)
距離	-0.0135 (-2.1)	-0.0137 (-2.5)
交通利便性満足度	0.143 (4.3)	0.154 (5.4)
買物利便性満足度	0.113 (3.2)	0.158 (4.0)
スケールパラメータ34	1.46 (3.3)	1.31(3.9)
スケールパラメータ45	0.879 (4.4)	0.917 (4.8)
諸統計量	サンプル数575 $\rho^2: 0.547$	サンプル数575 $\rho^2: 0.547$

参考文献

- 1) 例えば、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，pp311-318，1991。
- 2) J.M. Golob and L. J. M. Schreurs：The Design and Policy Applications of a Panel for Studying Changes in Mobility Over Time, Behavioural Research for Transportation Policy, pp.81-95, 1986。
- 3) 西井和夫・岩本哲也：ショッピングコンプレックス来訪者の買物行動特性の基礎的分析，土木計画学研究・講演集，No.13，pp975-982，1990。
- 4) 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・講演集，No.17，pp.39-42，1995。
- 5) Daganzo, C.F and Y. Shefi: Multinomial Probit with Time-Series Data: Unifying State Dependence and Serial Correlation Models, Environment and Planning A, Vol.14, pp.1377-1388, 1982。
- 6) 森川高行・山田菊子：系列相関を持つRPデータとSPデータを同時に用いた離散型選択モデルの推定法，土木学会論文集，No.476/IV-21，pp.11-18，1993。
- 7) 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・講演集，No.17，pp.43-46，1995。