

SPデータのバイアスに関する研究*

ーソウル市を例としてー

The Study about the Bias Problems included in SP Data*

-In the Case of Seoul city-

李 薰基**・原田 昇***・太田勝敏****

By Hunki LEE, Noboru HARATA, Katsutoshi OHTA

1. はじめに

非集計モデルのデータとしては、実際の交通行動を記述したRP(Revealed Preference)データと、人々の選好意識をデータを記述したSP(Stated Preference)データが存在する。SPデータは現存しない交通サービスが分析でき、重共線性の問題を避け、あるファクターを他のファクターと分けて分析することができるという利点があり、最近頻繁に使われるようになった。しかし、選好意識をデータと用いるSPデータには、信頼性の問題が残されており、依存状態(State Dependence)と系列相関(Serial Correlation)によるバイアスの問題が明らかになっている¹⁾。森川らは信頼性を安定性(Stability)と信憑性(Validity)に分け、SPデータで起こりうるバイアスをまとめて整理している²⁾。

SPデータによるバイアスは、異なる実験計画によって生ずるバイアス(安定性)と回答の際に回答者によって生ずるバイアス(信憑性)が存在する。その中で本研究では、①RPデータと比較しながらSPデータの有効性を検討分析し、②SPデータの安定性の中で、属性の数によるパラメータの安定性に関して、③SPデータの信憑性の中で、不完全な代替案の表示によるパラメータの安定性に関して検討する。

2章ではこの論文で取り扱うバイアスの種類について理論的な仮説を立て、各項目について説明を加える。3章では調査の概要を簡略に述べ、得られたデータの特徴を説明する。4章では分析結果を明らかにし、理論的な仮説を検証する。5章では本研究

*キーワード: SPデータ、有効性、バイアス、交通手段選択
**学生員、工修、東京大学工学部都市工学科
(東京都文京区本郷7-3-1、TEL03-3812-2111、FAX03-5800-6958)

***正員、工博、東京大学工学部都市工学科(同上)

****正員、Ph. D、東京大学工学部都市工学科(同上)

の成果をまとめる。

2. 理論的仮説

(1) SPデータの有効性

SPデータは一種の実験計画であってデータの操作性に富んでいる。そこで、現存しない交通サービスの分析ができ、RPデータでは取り扱いにくいデータさえ手軽に取り扱うことができる。本研究ではRPデータでは取り扱いにくい混雑度の分析をSPデータを用いて分析を行い³⁾、RPモデルと比較してSPデータの有効性を明らかにする。SPデータで構築されたSPモデルの比較のために、RPデータでは混雑度の認識値(Perceived Value)を調査している。

(2) 属性の数によるパラメータの安定性

SPデータは異なる属性の数によってバイアスが生ずる可能性がある。理論上、SPデータは実験計画のデータであって、変数間の相関の問題(重共線性の問題)を避けることができる。ということは、異なる属性の数で推定されたSPモデルであってもパラメータの値にはどんな影響も与えないはずである。本研究では属性の数によるバイアスの問題を調べるために次の仮説に従い分析を行う。

基本仮説: すべての変数において推定されたSPモデルにおいて、一つの変数が取り除かれても他の変数にはどんな影響も与えない。

すなわち、総効用

$$U = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (1)$$

から、 x_3 を除けば

$$U' = \beta_1 x_1 + \beta_2 x_2 + \varepsilon' \quad (2)$$

となり、 $\beta_1 = \beta'_1$ 、 $\beta_2 = \beta'_2$ であれば、属性の数によるバイアスは存在しないことになる。

(3) 不完全な代替案の表示によるパラメータの安定性

SPデータは一種の実験データであり、用いられる属性は実験者が任意に決める。そこで、抜け落ちた属性によるバイアスが生ずる可能性が存在する。このバイアスを通常変数省略バイアス (Omitted Variable Bias) と呼ぶ。そこで、不完全な代替案の表示によるパラメータの安定性を調べるために次の仮説に従い分析を行う。

基本仮説：不完全な代替案の表現によるバイアスが全くないと仮定すれば、異なる代替案を含んでいるSP調査票であっても、同じ変数組みに合わせた際、両モデルの変数値が一致しなければならない。

すなわち、

第1属性のSPモデル

$$U = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (3)$$

第2属性のSPモデル

$$U' = \beta_1 x_1 + \beta_2 x_2 + \varepsilon' \quad (4)$$

式(3)から x_3 を除けば、

$$U' = \beta_1 x_1 + \beta_2 x_2 + \varepsilon' \quad (5)$$

となり、 $\beta_1 = \beta_1$ 、 $\beta_2 = \beta_2$ であれば、省略変数によるバイアスは存在しないことになる。

3. 調査の概要

本研究で用いたデータはソウルでSP調査を実施し、得たものである。韓国のソウルは深刻な交通問題に伴い、5号線から12号線までの地下鉄が計画され、建設中である。SP調査は、新しく地下鉄が提供される2つの地域と、すでに地下鉄が運行されている1つの地域で同時に行い、3つの地域から都心に通勤している人を都心側の企業で調査したものである。

なお、2章で述べたバイアスの問題を調べるため、属性の種類が異なるようにX調査票とY調査票に分け、2つのタイプの調査票を被験者に半分ずつ割り当てた。表1にX調査票とY調査票に盛り込まれた属性の種類を示す。Y調査票には混雑度の属性が含まれており、混雑度による回答値の変化が分析できる。

3つの地域から選られた調査票ごとの有効サンプル数を表2に示す。

表1 調査票の分類

要因	A 調査票			B 調査票		
	バス	地下鉄	自動車	バス	地下鉄	自動車
アクセス時間	○	○	—	○	○	—
乗車時間	○	○	○	○	○	○
コスト	○	○	○	○	○	○
混雑度	—	—	—	○	○	—

表2 SP調査票ごとの有効サンプル数

	A地域	B地域	C地域
X調査票	495	386	637
Y調査票	509	456	516

4. 分析結果

(1) SPデータの有効性

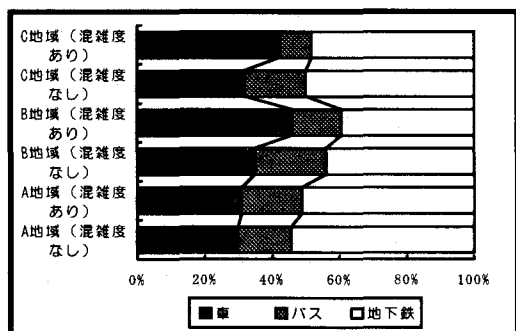
本研究では、RPデータでは取りにくい混雑度をSPデータを用いて分析を行っており、それによってSPデータの有効性が分析できる。まず、X調査票とY調査票のSPデータを用いて、第1交通手段選択率を比較する。車内の混雑度が組み込まれることにより、公共交通手段の選択率は減少し、車の選択率が増加している。ということは混雑度に対して回答者は反応しており、混雑度を意識の中で交通手段の要因として考慮していることがわかる(図1)。

次にSPデータに基づいて構築された多項ロジットモデルの混雑度のパラメータの値が有効であるかどうかを調べるために混雑度を含めたRPモデルと比較する。通常RPデータでは混雑度の値が取りにくいのが、ここでは混雑度に対する一人一人の認識値 (Perceived Value)を取り、それを変数値に用いたものである。表3にSPモデルとRPモデルを比較したモデルを示す。RPモデルは混雑度とコストの相関によりコストの変数が意味のないものになっており、重要な変数でありながらもパラメータの重みが推定できない。それに対してSPモデルでは混雑度のパラメータの符号が正しく、すべての変数のパラメータがよく推定されており、他の変数のパラメータとの相関が存在しないことがわかる。従って、SPデータを用いて推定されたSPモデルは有効であると考えられる。

表3 SPモデルとRPモデルの比較

	RPモデル	SPモデル
乗車時間 (分)	-0.0529 (1.85)	-0.0533 (5.46)
コスト (ウォン)	-0.0004 (0.35)	-0.0015 (6.04)
徒歩アクセス 時間 (分)	-0.0549 (1.31)	-0.0692 (3.14)
バスアクセス 時間 (分)	-0.0473 (0.92)	
混雑度 (%)	-0.0190 (6.12)	-0.0120 (6.06)
車ダミー(RP)	-4.5760 (4.36)	
バスダミー(RP)	-1.3570 (2.49)	
車ダミー(SP)		-0.8026 (1.80)
バスダミー(SP)		-1.5430 (6.64)
車慣性項		1.3170 (4.34)
バス慣性項		0.3679 (0.84)
尤度比	0.4314	0.2606
的中率	76.22	58.72
サンプル数	185	516

図1 混雑度による交通手段選択率



(2) 属性の数によるパラメータの安定性

属性の数によるパラメータの安定性を調べるために、Y調査票のすべての変数を用いて推定したSPモデルと、Y調査票から混雑度を削除して構築したSPモデルを比較分析した。その結果を表4に示し、混雑度の変数以外の変数を比較してt検定を行った結果を表5に示す。

実験計画により設定された乗車時間、コスト、アクセス時間、混雑度の4つの変数は互いに独立であ

り、どんな影響も与えないはずである。実際に、すべての変数を用いて推定したモデルと、混雑度を削除して推定したモデルの、乗車時間、コスト、アクセス時間のパラメータの値にはどんな有意な差も見られなかった。従って、2章の(2)の基本仮説は採択され、モデルに導入する属性の数を変えても、パラメータの値は変化せず、SPモデルのパラメータの値は安定していると考えられる。

表4 混雑度を組み込んだモデルにおいて混雑度の変数を除いた時の比較

変数	A地域	混雑度 削除	B地域	混雑度 削除	C地域	混雑度 削除
乗車 時間	-0.0253 (2.77)	-0.0258 (2.93)	-0.0367 (3.90)	-0.0376 (4.07)	-0.0504 (5.27)	-0.0501 (5.33)
コスト	-0.0016 (5.92)	-0.0015 (5.88)	-0.0015 (6.07)	-0.0014 (5.89)	-0.0014 (5.77)	-0.0013 (5.71)
アクセス 時間	-0.0676 (3.21)	-0.0672 (3.36)	-0.0368 (1.64)	-0.0395 (1.80)	-0.0643 (2.94)	-0.0518 (2.46)
混雑度	-0.0157 (8.17)		-0.0112 (5.68)		-0.0119 (6.12)	
車 ダミー	-1.958 (4.35)	0.3781 (1.18)	-0.0706 (0.16)	1.4940 (4.41)	-0.7024 (1.60)	1.1000 (3.46)
バス ダミー	-1.374 (6.55)	-1.182 (6.14)	-0.9397 (4.47)	-0.8628 (4.26)	-1.4890 (6.75)	-1.3480 (6.44)
最終対数 尤度	-450.7	-488.1	-413.1	-430.2	-429.6	-449.8
χ^2 値	74.8*	-	34.2*	-	40.4*	-
尤度比	0.1940	0.1272	0.1755	0.1412	0.2421	0.2065
的中率	58.74	53.83	56.58	54.61	57.75	55.81
サンプル	509	509	456	456	516	516

注 *: 1%有意

() 内t値

表5 t検定結果

変数	A地域	B地域	C地域
乗車時間	0.04	0.07	0.02
コスト	0.19	0.24	0.16
アクセス時間	0.01	0.08	0.41
車ダミー	4.23*	2.80*	3.32*
バスダミー	0.67	0.26	0.46

*5%有意差で棄却

(3) 不完全な代替案の表示による安定性

異なる代替案を表示した二つのモデルに対して、混雑度を除いて同じ変数組みを持っているモデルを構築し、t検定により両モデルの変数値を比較分析する。混雑度を除いて同じ変数組みで作成したモデルを表6に、変数値のt検定を行った結果を表7に

示す。

分析結果により、コスト、ダミー変数には有意な差が見られなかった。しかし、乗車時間、アクセス時間の変数には有意な差が見られた。同じ地域から選ばれたSPデータであるにもかかわらず、変数間に有意な差が見られることは、変数の省略によって生じたバイアスと思われる。すなわち、車内で感じる混雑度の抵抗を乗車時間に入れて一緒に考えたり、アクセス時間に入れて一緒に考えたりすることによるバイアスと考えられる。従って、3つの地域において混雑なしのモデルの乗車時間とアクセス時間のパラメータの値が、混雑ありのモデルの乗車時間とアクセス時間のパラメータの値より大きくなっている。これにより、変数が省略された時に生ずるバイアスが明らかに示された。そこで、2章(3)の基本仮説は棄却され、不完全な代替案の表示によるバイアスが存在することが明らかになった。

表6 混雑度の変数を省略した時の比較

変数	A地域 (混雑 なし)	A地域 (混雑 あり)	B地域 (混雑 なし)	B地域 (混雑 あり)	C地域 (混雑 なし)	C地域 (混雑 あり)
乗車 時間	-0.05872 (6.02)	-0.0258 (2.93)	-0.0595 (5.67)	-0.0376 (4.07)	-0.0585 (5.67)	-0.0501 (5.33)
コスト	-0.00155 (5.67)	-0.0015 (5.88)	-0.0015 (5.18)	-0.0014 (5.89)	-0.0016 (5.18)	-0.0013 (5.71)
アクセス 時間	-0.09931 (4.57)	-0.0672 (3.36)	-0.1067 (4.37)	-0.0395 (1.80)	-0.1103 (4.37)	-0.0518 (2.46)
車 ダミー	0.3811 (1.19)	0.3781 (1.18)	0.7321 (2.07)	1.4940 (4.41)	0.5343 (2.07)	1.1000 (3.46)
バス ダミー	-1.092 (5.61)	-1.182 (6.14)	-0.5133 (2.43)	-0.8628 (4.26)	-0.8692 (2.43)	-1.3480 (6.44)
尤度比 的中率	0.1913 59.39	0.1272 53.83	0.1242 53.11	0.1412 54.61	0.1642 53.73	0.2065 55.81
サンプル	495	509	386	456	637	516

表7 t検定結果

変数	A地域	B地域	C地域
乗車時間	2.51*	1.57	0.67
コスト	0.20	0.17	1.02
アクセス時間	1.09	2.05*	2.08*
車ダミー	0.01	1.55	1.32
バスダミー	0.33	1.19	1.81

*5%有意差で棄却

5. まとめと今後の課題

本研究を通して選ばれた結果を以下にまとめる。

①RPデータでは取りにくい混雑度の変数をSPデータを用いて分析することができ、混雑度の認識値を用いたRPモデルと比較分析した結果、SPデータの有効性が確認できた。

②SPデータにおいて、モデルに導入する属性の数を変えてもパラメータの値は変化しないという仮説の下で、分析を行った結果、パラメータ間には有意な差はなく、属性の数を変えてもパラメータは安定していることが示された。

③抜け落ちた変数による省略バイアスを調べた結果、混雑度を盛り込んでいない(X調査票)SPモデルと、混雑度を盛り込んだ(Y調査票)SPモデルとのパラメータの値には有意な差が見られた。すなわち、変数が省略されることによる、不完全な代替案の表示によるバイアスが存在することが確認できた。

以上のことにより、SPデータは、RPデータで取り扱いにくい変数を分析し、重要な変数の重みを推定するに適していると考えられる。しかし、SPデータにはバイアスが含まれており、なるべくそのバイアスを取り除くための工夫が必要と考えられる。

本研究では変数の省略によるバイアスが明らかに現れており、SP調査票の作成の際に重要な変数が抜け落ちないような工夫が必要であり、今後の研究課題でもある。

参考文献

1. 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,pp11-18,1993
2. 森川高行：ステイティド・プリファレンス・データの交通需要予測モデルの適用に関する整備と展望、土木学会論文集、No.413/IV12,pp9-18,1990
3. 李 薫基、原田 昇、太田勝敏：SP調査において混雑度が交通手段選択に及ぼす影響、49回年次学術講演会概要集第4部、pp888-889、平成6年
4. Moshe Ben-Akiva and S.R.Lerman:Discrete Choice Analysis:Theory and Application to Travel Demand,The MIT Press,1985
5. Bradley,M.:Realism and Adaptation in Designing Hypothetical Travel Choice Concepts,Journal of Transport Economics and Policy,Vol.22,No.1,pp.121-137,1988
6. Bradely,M.,Daly A.:Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data,Transportation,Vol.21,pp.167-184,1994