

空間データサンプリングのための 交通機関選択モデルの精度検証に関する研究

松原 司¹・塚井 誠人²・津田 敏明³

¹学生会員 広島大学大学院 工学研究科社会基盤環境工学専攻 (〒739-8527広島県東広島市鏡山1-4-1)
E-mail:m110781@hiroshima-u.ac.jp

²正会員 広島大学大学院准教授 工学研究院社会環境空間部門 (〒739-8527広島県東広島市鏡山1-4-1)
E-mail:mtukai@hiroshima-u.ac.jp

³学生会員 広島大学大学院 工学研究科社会基盤環境工学専攻 (〒739-8527広島県東広島市鏡山1-4-1)
E-mail:m121388@hiroshima-u.ac.jp

統計モデルは理想的なデータとして、十分なランダム性と直交性を有するデータを要求する。本研究では、交通機関選択行動データを空間データと考え、交通機関選択モデルの性能をRPデータで検証することを目的とする。広島市PT調査データを用いて、1)入力データの空間類似性をLocal Moran's Iによって指標化し、2)誤差項の定式化が異なる3種類の交通機関選択モデルを推定、3)時間価値(VOT)の算出とcross validation(CV)によるモデル精度の検証を行った。さらに、入力データとモデル精度の関係性を検証した。検証結果より、尤度比、VOT、およびCVで、最も良好なモデルが異なっていた。さらに、空間類似性の強いデータに対して、MNLモデルを用いると、各指標について平均的に安定した結果が得られた。また、バスサービス水準はランダムデータ基準を満たさず、モデル精度を損なっている可能性が明らかとなった。

Key Words : RP data survey, spatial data, mode choice model, model performance

1. はじめに

統計モデルが要求する理想的なデータは、ランダム性と直交性を満たすデータである。しかし、たとえ母集団からの標本抽出がランダムに行われたとしても、実空間の交通行動を調査するRP調査では、交通機関選択モデルなどに用いられる各交通機関の所要時間、費用、移動距離などの説明変数間の直交性を厳密に管理することは不可能であり、これらの要因間に一定の相関が生じる。これは、移動の出発地や目的地などの空間属性が、複数の要因間で交絡することに起因する。この点をふまえれば、RP調査から得られたデータは、非直交性に配慮した空間データとしての取り扱いが必要であると考えられる。

しかし、これまでのところ、RP調査データの空間特性に着目した研究は、筆者の知る限りほとんど蓄積されていない。よって、その空間相関の程度がモデルの再現性を示す的中率に及ぼす影響は全く示されていない。

本研究では、選択肢間の誤差相関を許容する交通機関選択モデルの性能を、交通サービス水準などの空間相関性と非直交性の高い項目を多く含むRPデータの下で検

証する。その際に、RP調査によって得られる交通機関選択行動データの空間類似性を指標化し、誤差相関構造が異なる複数の交通機関選択モデルの精度検証を、メタ分析によって行うことを目的とする。モデルの安定性を計測する指標として、交通時間価値、およびクロスバリデーションによる的中率に着目する。メタ分析では、モデルの的中率を目的変数、入力変数として、モデルの説明変数として用いた各変数のゾーン内分散、および空間類似の指標であるLocal Moran's Iを設定し、入力変数の特性とモデル精度の関係について分析する。

2. 既往研究の整理と本研究の位置づけ

和泉ら¹⁾は、都市圏パーソントリップ調査を基本としたGISを統合的に利用した交通行動分析手法を提案した。具体的には、GISを用いて計算可能な空間情報を交通データと統合し、都市圏の交通現象を時空間的に把握した。この研究では、都市圏活動者の空間分布の時間変化を表現することにより、交通行動の時空間的推移が目的地ごとに可視化した。また、大規模店舗など特定施設立地に

対応した交通目的別のトリップの時間変動情報を用いて、都市圏域の土地利用特性、活動内容、および交通行動の関連性が導出されている。

中井ら²⁾は、交通規制や交通管制の基礎情報となるOD交通量を、入手可能な限られた情報から推定する、空間統計学を援用した新しい手法を提案した。この研究では、未観測地点のデータを近隣の情報を用いて内挿するクリギングを援用して、発生量、および集中量のそれぞれからOD交通量から推計する方法を比較し、それらの手法の有効性を示した。

Miyamotoら³⁾は、目的地選択モデルを構築する場合、明示的に空間データが使用されることを指摘した。その上で、Miyamotoらは、MXLモデルが効用確定項間の空間的自己相関と、ガンベル分布に従う誤差項の空間的自己相関の両者を考慮できることを着目して、最尤法に基づくシミュレーションベースのパラメータ推定法を提案した。さらに、仙台市の目的地選択データを用いて提案手法の有効性を検証したところ、その有用性が明らかとなった。

Mohammadianら⁴⁾は、MNLモデルの確定項に、距離の減衰関数として他主体の行動結果の観測値を導入することで、選択枝間の空間的自己相関を考慮した選択モデルを開発し、居住地選択に適用したところ、提案モデルは有効と報告している。

以上の研究を概観すると、交通計画において空間情報は、モデルの入力変数の精微化や調査データが得られない地点の空間内挿に用いられている。一方で、多くのデータが正の空間類似性を示す空間データでは、本質的に変数間の多重共線性が避けられないという課題に対して、モデル構築の指針となるような空間サンプリングに関するガイドラインや、入力データの空間特性を踏まえたモデル選択指針に着目した研究は、筆者の知る限りほとんど行われていない。

本研究では、交通機関選択モデルが要求する理想的なデータは、十分なランダム性と直交性を有するデータとの立場から、選択枝間の誤差相関を許容する交通機関選択モデルの性能を、RPデータの下で検証する。

3 使用データの概要

分析に使用するデータは、広島市交通実態調査データと交通機関サービス水準データを統合したものである。H20広島市PT調査は、2010年11月下旬～12月中旬にかけて、郵送配布、郵送回収方式で実施された。調査内容は、性別や年齢、職業などの個人・世帯属性、出発地・到着地や代表交通手段などの移動に関する項目である。本調査では、有効トリップ数として92221トリップが得られ

た。

H20広島市PT調査の住民アンケート調査には、利用した交通手段、ならびに選択可能な交通手段のサービス水準に関する項目が含まれていない。そこで、交通機関サービス水準を算出プログラムを用いて算出し、利用した交通手段と後述する条件で設定した選択可能な交通手段について、トリップデータに付加する。

以下の抽出手順で有効トリップ数(92221)の中から分析対象となるトリップを抽出し、使用データとする。

- ① H20広島市PT調査で設定されたゾーンから、アストラムラインの駅を含むBゾーンから出発するトリップを抽出する。
- ② ①で抽出したトリップの中から、利用された、または選択可能な交通手段が自動車、バス、およびアストラムラインの3種となるトリップを抽出する。この条件を満たすトリップとして、まず、代表交通手段がバイク、タクシー・ハイヤー、自動車(運転)、自動車(同乗)、バス、およびアストラムラインのいずれかのトリップを抽出する。
- ③ 目的地ゾーンは都心部とする。具体的には、アストラムラインの駅である本通、県庁前、城北、および白島を含むゾーン、また、それらに近接するゾーンを含むゾーン内のCゾーンを目的地ゾーンとして設定した。以上の手順により出発地、および目的地ゾーンとしたゾーンを、**図-1**に示す。
- ④ ③で抽出したトリップの中から、自動車、バス、アストラムラインの3つの交通機関のうち、上述した3交通機関が全て選択可能なトリップのみを抽出する。上述の手順によって抽出されたトリップ数は、1215となった。以下、これを使用データとして分析を行う。

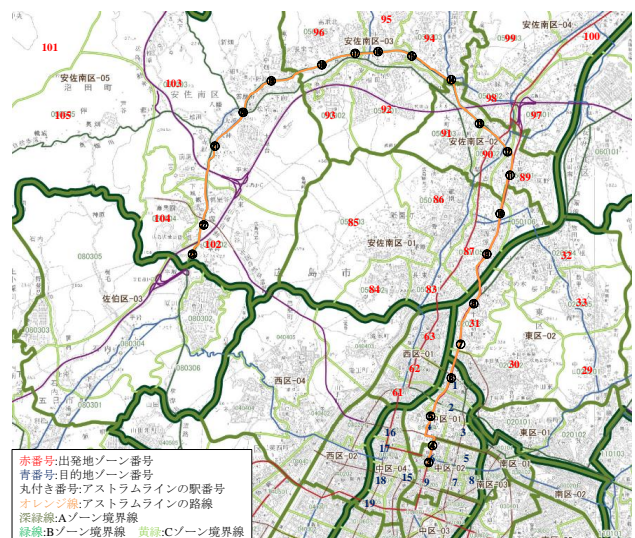


図-1 抽出した出発地ゾーンと目的地ゾーン

4. 空間データの質の評価

本研究では、空間データの質を評価する指標として、Local Moran's I^z ⁹⁾を用いる。

ゾーン z の Local Moran's I は、式(1)より求められる。

$$I_z = \frac{(x_z - \bar{x})}{\sum_{z=1}^N (x_z - \bar{x})^2 / N} \cdot \sum_{j=1}^N w_{zj} (x_j - \bar{x}) \quad (1)$$

ここで、 x_z はゾーン z の属性値、 \bar{x} は属性値の平均、 N はゾーン数、 x_j はゾーン j の属性値、 w_{zj} はゾーン zj 間の空間重み付け行列の要素である。本研究では、ゾーン間距離の逆数を行ごとに標準化した値を各要素 w_{zj} とする空間重み付け行列 W を用いる。

Local Moran's I は、Moran's I のゾーン z に関する指標である。Local Moran's I の値が正の場合は、対象ゾーンの属性値が近隣ゾーンと近い値を持つこと、すなわち、その属性値に関して空間類似性があることを示す。逆に、Local Moran's I の値が負の場合は、その属性値に関して空間類似性がないことを示す。

個人属性と交通機関サービス水準に関して Local Moran's I を算出した。その中から、年齢、性別、バスの所要時間、およびバスの費用に関して算出した Local Moran's I の分布を図-2, 3, 4, および5に示す。まず、年齢に関して図-2より、3分の2以上のゾーンが正の値を取っており、多くのゾーンでは空間類似性がある。特に、ゾーン101, 104, および105の Local Moran's I の値が正で大きく、これらのゾーンの年齢の値が近隣ゾーンの値と非常に類似していることがわかる。一方、性別に関して図-3より、Local Moran's I の値が負のゾーンが半数以上を示しており、性別に関しては空間類似性があまりみられないことがわかる。しかしながら、年齢と同様に、ゾーン101, 104, および105では非常に高い値となっており、一部に空間類似性がみられるゾーンも存在する。

次に、図-4と5を着目すると、所要時間に関してはゾーン32, 89, および104以外、料金に関してはゾーン89以外の全てのゾーンで Local Moran's I の値が正となり、空間類似性のあるゾーンが多く見られた。また、空間類似性の高いゾーン数を他の交通機関と比較したところ、バスのサービス水準は他の交通機関よりも空間類似性が高いことが明らかとなった。

5. 交通機関選択モデルの安定性の検証

(1) 使用する交通機関選択モデル

a) 多項ロジット (Multinomial Logit) モデル
多項ロジットモデル (以下、MNLモデル) は選択確

率に積分形が残らず、計算が容易な特徴を持つ。MNLモデルでは、各選択肢 i の誤差項に、個人 n の間についても独立で同一なガンベル分布を仮定する。MNLモデルでは、個人 n が選択肢 i を選択する確率 $P_n(i)$ は、式(2)のように表現される。

$$P_n(i) = \frac{\exp(\mu V_{n,i})}{\sum_{k=1}^K \exp(\mu V_{n,k})}, \quad i=1, \dots, K \quad (2)$$

ここで、 $V_{n,i}$ は効用の確定項、 K は選択肢数、および μ は分布のばらつきを表すスケールパラメータであり、1に固定されることが多い。

b) 多項プロビット (Multinomial Probit) モデル

多項プロビットモデル (以下、MNPモデル) では、各選択肢 i の誤差項 $\varepsilon_{n,i}$ が正規分布に従うと仮定する。MNPモデルは、最も誤差相関構造に関する制約が緩く、選択肢間の類似性を明示的に導入できるモデルであり、その一般形は式(3)で定義される。

$$P_n(i) = \int_{\varepsilon_1=-\infty}^{\varepsilon_1+V_i-V_1} \dots \int_{\varepsilon_i=-\infty}^{\infty} \dots \int_{\varepsilon_K=-\infty}^{\varepsilon_i+V_i-V_K} \Phi(\varepsilon) d\varepsilon_K \dots d\varepsilon_1 \quad (3)$$

ここで、 $\Phi(\varepsilon)$ は誤差項 $\varepsilon_{n,i}$ の密度関数であり、式(4)で表される。さらに、誤差項の分散共分散行列 Σ は、式(5)を満たす。

$$\Phi(\varepsilon) = (2\pi)^{-\frac{K}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \varepsilon \Sigma^{-1} \varepsilon^T\right] \quad (4)$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1K} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \dots & \sigma_K^2 \end{pmatrix} \quad (5)$$

c) ミックスドロジット (Mixed Logit) モデル

ミックスドロジットモデル (以下、MXLモデル) は、MNLモデルとMNPモデルの融合系モデルである。MXLモデルは、効用の誤差項 $\varepsilon_{n,i}$ を2つの確率変数に分解し、式(6)のように定義される。

$$U_{n,i} = V_{n,i} + (\eta_{n,i} + \xi_{n,i}), \quad i=1, \dots, K \quad (6)$$

ここで、 $\eta_{n,i}$ を共分散行列 Ω を持つ多変量正規分布に従う誤差項、 $\xi_{n,i}$ をガンベル分布に従う誤差項とする。 $\eta_{n,i}$ の値が与えられたときの条件付き選択確率は、式(7)のように表現される。

$$P_n(i|\eta_n) = \frac{\exp(V_{n,i} + \eta_{n,i})}{\sum_{k=1}^K \exp(V_{n,k} + \eta_{n,k})} \quad (7)$$

式(7)に現れるパラメータ $\eta_{n,i}$ の期待値を取ると、MXLモデルの式(8)のような選択確率の一般形となる。

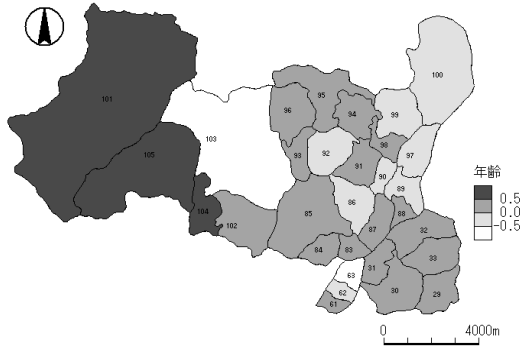


図-2 年齢に関するLocal Moran's Iの分布

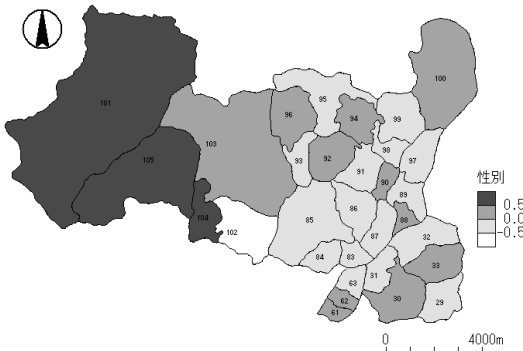


図-3 性別に関する Local Moran's Iの分布

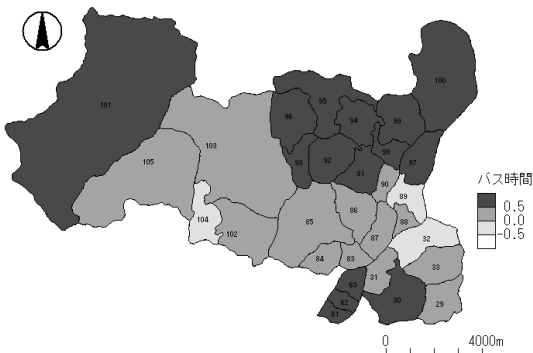


図-4 バスの所要時間に関する Local Moran's Iの分布

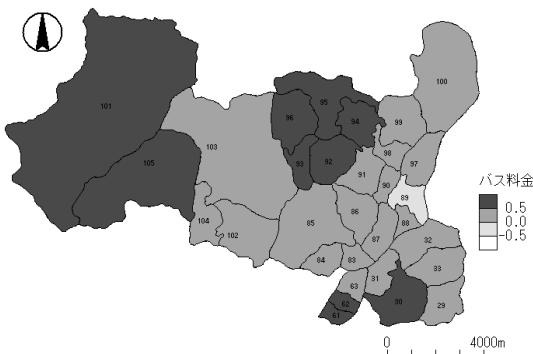


図-5 バスの費用に関する Local Moran's Iの分布

$$\begin{aligned}
 P_n(i) &= \int P_n(i|\eta_n) f(\eta_n|\Omega) d\eta_n \\
 &= \int \frac{\exp(V_{n,i} + \eta_{n,i})}{\sum_{k=1}^K \exp(V_{n,k} + \eta_{n,k})} f(\eta_n|\Omega) d\eta_n
 \end{aligned} \tag{8}$$

ここで、 $f(\eta_n|\Omega)$ は共分散行列 Ω を持つ多変量正規分布の確率密度関数である。

(2) 交通機関選択モデルの推定結果

(1)で示した3種類の交通機関選択モデルの推定結果をまとめて表-1に示す。モデルの当てはまりを表す自由度調整済み尤度比に着目すると、MXLモデルが最も高く、以下にMNLモデル、MNPモデルの順となった。各モデルの再現性は、モデルの当てはまりと一致している。

MNLモデルは、自動車とバスの所要時間、全ての交通機関の費用、自動車免許の有無、年齢に関するパラメータが負で有意となり、所要時間と費用が低下すると、各交通機関の選択確率が高くなり、また、年齢が低い人ほど公共交通機関であるアストラムラインの選択確率が高くなること表示。自動車免許の有無に関するパラメータは負で有意な値を示し、符号条件を満たさない予想に反した結果となった。

MNPモデルの当てはまりは、3モデルの中で最も低いものの、自動車とバスの所要時間、バスとアストラムラインの費用、および年齢は、それぞれ負で有意となり、符号条件を満たす。また、選択肢間の相関を表す σ_{12} 、 σ_{23} 、および σ_{31} は、公共交通機関であるバスとアストラムラインの相関が最も高くなり、妥当な結果である。

MXLモデルは、自動車とバスの所要時間、バスとアストラムラインの費用、年齢に関するそれぞれのパラメータが負で有意となり、予想に沿った結果となった。自由に利用できる自動車の有無に関するパラメータは、正で有意となっている。このパラメータは、MNLモデルと同じく、自動車の効用関数に設定しており、正で有意ということは、妥当な結果と言える。

(3) 交通時間価値の算出

一般に、交通プロジェクトのよって生じる効果のうち、交通時間短縮は特に大きな割合を示すことが多い。そこで、プロジェクトの経済効果分析にあたって、妥当な交通時間価値（以下、VOT: Value of Time）を求めることが重要な問題となる。こうした背景を受けて、近年、国別のVOTの比較が行われるようになり、日本でも加藤らが研究を行っている。なお、式(9)で算出されるVOTは、本節では表-1に示した推定結果から、パラメータの安定性の指標としてVOTを算出する。

表-1 MNL, MNP, およびMXLモデルの推定結果

| 説明変数 | MNL | MNP | MXL |
|---------------------------------|------------|------------|----------|
| | 推定値 | 推定値 | 推定値 |
| 自動車所要時間 | -0.15 ** | -0.13 ** | -0.48 * |
| バス所要時間 | -0.08 ** | -0.06 ** | -0.23 * |
| アストラムライン所要時間 | -0.10 | -0.06 | -0.28 |
| 自動車費用 | -1.E-03 ** | -3.E-04 | -2.E-03 |
| バス費用 | -3.E-03 ** | -1.E-03 ** | -0.01 * |
| アストラムライン費用 | -0.01 ** | -5.E-03 ** | -0.02 * |
| 自動車免許の有無(1) ^{a)} | -0.50 * | -9.E-04 | -1.56 |
| 自由に利用できる自動車の有無(1) ^{a)} | 1.51 ** | 0.01 | 7.44 * |
| 性別(3) ^{a)} | -0.05 | -1.E-03 | 0.12 |
| 年齢(3) ^{a)} | -0.02 ** | -0.01 ** | -0.05 * |
| 定数項(1) ^{a)} | -1.90 ** | -0.01 | -6.71 * |
| 定数項(2) ^{a)} | -0.37 | 0.01 | -0.40 |
| η1 | — | — | -7.95 |
| η2 | — | — | 1.57 |
| η3 | — | — | 3.12 |
| σ12 | — | 0.02 | — |
| σ23 | — | 0.51 | — |
| σ31 | — | -9.E-04 | — |
| 初期対数尤度 | -1334.81 | -1334.81 | -1334.81 |
| 最終対数尤度 | -1086.73 | -1141.38 | -1071.54 |
| サンプル数 | 1215 | 1215 | 1215 |
| 自由度調整済み尤度比 | 0.18 | 0.14 | 0.19 |
| 再現性(%) | 57.6 | 52.3 | 58.2 |

a): 説明変数が入力された選択肢 1:自動車, 2:バス, 3:アストラムライン
 **):1%有意 *)5%有意 +):10%有意

表-2 各交通機関の交通時間価値

| 交通機関\モデル | MNLモデル | MNPモデル | MXLモデル |
|----------|--------|--------|--------|
| 自動車 | 152.6 | 420.4 | 217.5 |
| バス | 25.4 | 39.7 | 35.6 |
| アストラムライン | 13.3 | 13.2 | 14.2 |

$$\text{交通時間価値} = \frac{\beta_{time}}{\beta_{cost}} \quad [\text{円/分}] \quad (9)$$

ここで、 β_{time} は所要時間のパラメータ、 β_{cost} は費用のパラメータである。得られたVOTをまとめて表-2に示す。表-2より、どのモデルにおいても自動車のVOTが非常に高い結果となっている。また、モデル間でその値を比較すると、MNPモデルが最も高い値を示した。バスのVOTに着目すると、MNLモデルが最も小さく、以下、MXLモデル、MNPモデルの順となった。アストラムラインのVOTは、MXLモデルから算出された値が他のモデルよりもやや大きい、自動車やバスと比較してVOTのばらつきは小さかった。

(3) クロスバリデーション (CV) による再現性の検証

全データをあるブロックのデータとあるブロックを除いたデータに分割し、前者を検証用、後者を推定用として、推定用データで検証用データの予測を行い、その再現性を検証する。除くブロックは、3章で設定した出発地ゾーンで、アストラムラインの駅を含むゾーンを抽出し、それらのゾーンをPT調査のゾーン区分であるBゾーンごとに統合したブロックである。各ブロックを除いたデータの自動車、バス、およびアストラムラインの分担率が極端に偏らないようにブロックを選定した結果、

5ブロックを設定することとなった。aブロックはゾーン30, 31, bブロックはゾーン87, cブロックはゾーン89~91, dブロックはゾーン94~96, およびeブロックはゾーン102, 103をそれぞれ統合した。以下の分析では、これらの5ブロック単位ごとにCVを行う。

再現性の指標として、的中率を用いる。予測結果 $\hat{y}_{n,i}$ の右側に示す括弧内において、右側は推定用のデータを表し、-a, -b, -c, -d, および-eと記し、一方の左側は予測するブロックを表すので、a~eの除くブロック名が入る。たとえば、aブロックを除くデータで推定し、aブロックの再現性を検証する場合、ゾーンzが出発地である個人 n_z が選択肢iを選択した実際の選択結果 $y_{n_z,i}$ と、予測結果 $\hat{y}_{n_z,i}(a|-a)$ を用いて、ゾーンzにおける的中率 $HitR_z$ は、式(10)と(11)で算出される。

$$S_{n,i} = \begin{cases} 1; & y_{n_z,i} = \hat{y}_{n_z,i}(a|-a) \text{ のとき} \\ 0; & y_{n_z,i} \neq \hat{y}_{n_z,i}(a|-a) \text{ のとき} \end{cases} \quad (10)$$

$$HitR_z = \frac{1}{N_z} \sum_{n_z=1}^{N_z} \sum_{i=1}^3 S_{n_z,i}, \quad z=1, \dots, 30 \quad (11)$$

各交通機関選択モデルのCV結果を、表-3~5に示す。

表-3より、MNLモデルでは、a, d, およびeブロックの再現性が全データよりも高くなった。一方で、bブロックでは43.5%, cブロックでは42.6%となり、全データの再現性よりも、大幅に低くなった。表-4に着目すると、MNPモデルでは、a, c, およびdブロックで全データの再現性の値を上回り、逆に、b, cブロックでは下回った。表-5より、全データの再現性よりも高い再現性となった。ブロックは、a, dの2ブロックであり、他の3ブロックの再現性は全データよりも低くなった。モデル間で比較すると、5ブロックのCVのうち、最も再現性の高いブロックが多かったモデルはMNLモデルとなり、次にMNPモデル、MXLモデルであった。以上より、CV結果の良好な順を、尤度比の順序と比較すると、両者は異なる結果となった。また、表-3~5より、どのモデルを用いても、bブロックの再現性は全データよりも低下した。逆に、a, およびdブロックの再現性は全データの再現性よりも上昇し、その中でもdブロックの再現性は全データよりも大きく上昇した。さらに、a~eブロックの再現性の加重平均を計算した結果、MNLモデル、とMXLモデルでは全体よりも再現性は低くなり、逆にMNPモデルでは全体よりも高くなった。再現性が大きく低下するブロックが生じた理由は、bブロックでは、3つの交通機関サービス水準と交通機関選択の傾向が他のゾーンとは異なる傾向を示し、これらのゾーンを除いたデータで推定を行うと、パラメータが安定しなかったためと考えられる。

表-3 MNLモデルによるCVの結果

| 予測ブロック | トリップ数 | $y_{n,i} = \hat{y}_{n,i}$ となる トリップ数 | 再現性(%) | 全データとの 比較(%) |
|--------|-------|--|--------|-----------------|
| 全データ | 1215 | 636 | 57.6 | 0 |
| a | 197 | 118 | 59.9 | +2.3 |
| b | 85 | 37 | 43.5 | -14.1 |
| c | 122 | 52 | 42.6 | -15.0 |
| d | 123 | 81 | 65.9 | +8.3 |
| e | 61 | 36 | 59.0 | +1.4 |
| 加重平均 | | | 55.1 | -2.5 |

表-4 MNPモデルによるCVの結果

| 予測ブロック | トリップ数 | $y_{n,i} = \hat{y}_{n,i}$ となる トリップ数 | 再現性(%) | 全データとの 比較(%) |
|--------|-------|--|--------|-----------------|
| 全データ | 1215 | 636 | 52.3 | 0 |
| a | 197 | 113 | 57.4 | +5.1 |
| b | 85 | 38 | 44.7 | -7.6 |
| c | 122 | 73 | 59.8 | +7.5 |
| d | 123 | 80 | 65.0 | +12.7 |
| e | 61 | 29 | 47.5 | -4.8 |
| 加重平均 | | | 56.6 | +4.3 |

表-5 MXLモデルによるCVの結果

| 予測ブロック | トリップ数 | $y_{n,i} = \hat{y}_{n,i}$ となる トリップ数 | 再現性(%) | 全データとの 比較(%) |
|--------|-------|--|--------|-----------------|
| 全データ | 1215 | 707 | 58.2 | 0 |
| a | 197 | 118 | 59.9 | +1.7 |
| b | 85 | 37 | 43.5 | -14.7 |
| c | 122 | 50 | 41.0 | -17.2 |
| d | 123 | 80 | 65.0 | +6.8 |
| e | 61 | 25 | 41.0 | -17.2 |
| 加重平均 | | | 52.7 | -5.5 |

6. 空間データ特性とモデル精度の関係

(1) 交通機関選択モデル的中率に関するメタ分析

モデル的中率に対して、入力変数のゾーン内分散、および空間類似性の指標であるLocal Moran's Iがどの程度寄与するかに関してメタ分析を行う。分析には、以下の式(12)のロジスティック回帰モデルを考える。

$$\ln \frac{HitR_z}{1-HitR_z} = \beta_0 + \sum_{l=1}^L \beta_l I_{z,l} + \sum_{m=1}^M \beta_m x_{z,m} + \beta_p \delta_p + \beta_x \delta_x + \varepsilon_z \quad (12)$$

ここで、 $HitR_z$ はゾーン z における的中率、 $I_{z,l}$ はゾーン z における属性 l のLocal Moran's Iの値、および $x_{z,m}$ はゾーン z における属性 m のゾーン内分散である。また、 δ_p はMNPモデルを用いて推定したときに1、他のモデルで推定したときに0となるダミー変数、 δ_x はMXLモデルを用いて推定したときに1、他のモデルで推定したときに0となるダミー変数である。 β_0 、 β_l 、 β_m 、 β_p 、および β_x はパラメータを表す。パラメータは、式(12)の残差二乗和 $\sum \varepsilon_z^2$ の最小化によって推定する。 $I_{z,l}$ と $x_{z,m}$ となるゾーン属性として、年齢、性別、自動車免許保有の有無、および自由に利用できる自動車保有の有無の個人属性のゾーン平均値と、自動車、バス、お

よびアストラムラインの各交通機関サービス属性の全てを用いて、分析を行う。

(2) メタ分析の結果

式(12)のロジスティック回帰モデルを用いてパラメータ推定を行った結果を表-6に示す。表-6に示す決定係数は0.63であり、モデルの当てはまりは中程度であった。以下に有意となったパラメータに関して、考察を述べる。なお、サンプル数は30ゾーン×3のため、90である。

自動車の所要時間は、Local Moran's Iのパラメータが負、ゾーン内分散は有意でなかった。道路ネットワークは密度が高く、実際の所要時間に近い値が得られているため、期待された符号が得られたと考えられる。バスの所要時間は、Local Moran's Iのパラメータは正、一方でゾーン内分散のパラメータは負で、それぞれ有意となり、ともに基準に反する結果になった。本分析の入力データとして用いた公共交通サービス水準データはランダム性が満たされていないため、ランダムデータ仮説に反する結果が得られた可能性がある。一方で、バスの費用は、ゾーン内分散のパラメータが正で有意となった。バスの費用は、所要時間と異なり、待ち時間要因による精度低下がないため、基準を満たす結果が得られたと考える。アストラムラインの所要時間は、Local Moran's Iのパラメータが正となり、基準に反する結果が得られた。アストラムラインの運行頻度は、バスよりも1日のうちの変動が低いいため、以上の結果はアストラムラインの駅までのアクセス圏の設定に関する誤差が影響している可能性がある。自動車とアストラムラインの費用に関しては、いずれのパラメータも有意にならなかった。よって、これらのデータについて、明示的にランダム性基準を満たすとは言えないものの、比較的ランダム性が確保されたデータであったと考えられる。

年齢に関して、Local Moran's Iとゾーン内分散は、それぞれのパラメータが正で有意な値を示し、Local Moran's Iのパラメータはランダムデータ基準に反する結果となった。この結果から、年齢に関しては、交通機関選択行動が異質であって、説明変数として年齢を設定するよりも、年齢を基準とするデータセグメントによってモデル構築を行った方が、再現性の高いモデルが得られる可能性がある。

MNPモデルダミーのパラメータは負で有意な値を示した。つまり、MNPモデルの再現性は平均的に低い傾向がある。

表-6 的中率に関するメタ分析結果

| 説明変数 (Local Moran's I) | 推定値 | t値 |
|-------------------------------|---------|----------|
| 自動車の所要時間 | -1.63 | -2.66 ** |
| バスの所要時間 | 1.68 | 3.77 ** |
| アストラムラインの所要時間 | 0.72 | 2.90 ** |
| 自動車の費用 | 0.28 | 0.97 |
| バスの費用 | -0.61 | -1.04 |
| アストラムラインの費用 | -0.24 | -0.77 |
| 自動車免許の有無 | 0.51 | 1.54 |
| 自由に利用できる自動車の有無 | -1.13 | -3.11 ** |
| 性別 | -0.06 | -0.13 |
| 年齢 | 1.45 | 4.24 ** |
| 説明変数 (ゾーン内分散) | | |
| 自動車の所要時間 | 0.01 | 0.19 |
| バスの所要時間 | -0.01 | -3.12 ** |
| アストラムラインの所要時間 | 0.06 | 1.63 |
| 自動車の費用 | -2.E-06 | -0.46 |
| バスの費用 | 1.E-04 | 3.16 ** |
| アストラムラインの費用 | -2.E-04 | -1.24 |
| 自動車免許の有無 | 3.86 | 1.05 |
| 自由に利用できる自動車の有無 | -4.90 | -0.94 |
| 性別 | 10.95 | 1.03 |
| 年齢 | 3.E-03 | 2.04 * |
| 説明変数 (その他) | | |
| MNPモデルダミー (MNPモデル: 1, その他: 0) | -0.28 | -2.40 * |
| MXLモデルダミー (MXLモデル: 1, その他: 0) | 0.04 | 0.33 |
| 定数項 | -3.06 | -1.36 |
| 決定係数 | 0.63 | |

**):1%有意 *)5%有意 +):10%有意

7. 結論

本研究では、交通機関選択行動データの入力データとして用いる交通機関サービス水準をはじめとする諸変数を空間データと考えると、その空間類似性を Local Moran's I 指標によって表し、RP 調査から推定した交通機関選択モデルのゾーンの的中率に着目したモデルの安定性に関する検証を行った。以下に得られた知見をまとめる。

RP データの空間特性を Local Moran's I によって指標化したところ、多くの属性で入力データの空間類似性が存在することが明らかとなった。また、個人属性と交通機関サービス水準の空間類似性を比較すると、後者の空間類似性が高いことが明らかとなった。

検証の結果、尤度比、VOT、および CV で、最も良好な性質を示すモデルが異なることが明らかとなった。さらに、空間類似性の強い交通機関サービス水準データを

用いた場合、MNL モデルを用いることによって、各指標について平均的に安定した結果が得られることが明らかとなった。

メタ分析により、自動車サービス水準はランダムデータ基準を満たす一方で、バスサービス水準はランダムデータ基準を満たさず、モデル精度を損なっている可能性が明らかとなった。これらの結果から、空間データサンプリングによる重み調整の必要性が示唆される。

今後の課題として、都心部から郊外へのトリップといった双方のトリップを含むデータセットによる検討を行う必要がある。さらに、本研究では対象としなかった Nested Logit、Cross Nested Logit モデルなど、選択肢間の誤差相関を一部の類似性の高い選択肢のみに認めるモデルや、地理的加重回帰 (GWR) モデルなどの空間特性を考慮した選択モデルを含めた検証が必要である。以下の成果を踏まえて、モデルの空間移転性 (安定性) を考慮したサンプル重み決定方法の開発を進める必要がある。

参考文献

- 1) 和泉範之, 奥嶋政嗣, 秋山孝正: 空間情報を利用した交通行動の時間的推移の表現方法, 土木計画学研究・論文集, Vol.22, No.3, 2005.
- 2) 中井周作, 菊池輝, 北村隆一: 空間統計学を援用した OD 交通量推計手法の開発, 土木学会第 60 回年次学術講演会, pp.57-58, 2005.
- 3) Kazuaki Miyamoto, Varameth Vichiensan, Naoki Shimomura, and Antonio Páez: Discrete choice model with structuralized spatial effects for location analysis, Transportation Research Record, Vol.1898, pp.183-190, 2004.
- 4) Mohammadian, A. K., Haider, M. and Kanaroglou, P.S: Incorporating spatial dependencies in random parameter discrete choice models, presented at the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2005.
- 5) 丹後俊郎, 横山徹彌, 高橋邦彦: 空間疫学への招待, 朝倉書店, 2007.
- 6) 古谷知之: R による空間データの統計解析, 朝倉書店, 2011.

(2013.5.7 受付)

Comparison in Model Performance of Transportation Mode Choice Model for Spatial Data Sampling

Tsukasa MATSUBARA, Makoto TSUKAI and Toshiaki TSUDA

RP survey data obtained in PT Data with statistical modeling in transportation behaviors are expected to give rich implications about the transportation policies. RP survey data should be considered in its spatial characteristics of the survey field, while conventional studies in transportation behavior modeling or PT survey ignored them.

The purpose of this study is to test the accuracy of mode choice models focusing on spatial data characteristics.

As a result, likelihood ratio, value of time, and cross validation indicated that MNL model shows the best performance over MNP and MXL. Meta regression analysis showed that level of service in road satisfied with expected randomness, while that in bus did not satisfied with the expected randomness.