

東京大学都市工学教室 正 員 大田 勝 敏

1. 研究目的

ゾーン別発生トリップ数の推定にあたって、従来の地域的(ゾーン)集計値(平均, 合計)を直接に対象とした「ゾーンモデル」と、トリップの行動単位としての世帯を対象とした「世帯モデル」の2方法を理論的および実証的に比較検討すること。

2. モデルの構造とゾーン別トリップ数の推定

ここでいう(非集計)世帯モデルとはトリップ行動に関する意志決定単位としての世帯のトリップ発生量(y_{ij})は、その世帯特性(例えば世帯人員)を独立変数(x_{kij})とする線型モデルで表わされると仮定したものである。一方、ゾーン平均モデルは、ゾーンの世帯当り平均トリップ数($\bar{y}_j$)は、そのゾーンに住む世帯の平均的特性(e.g. 平均世帯人員), またゾーン合計モデルは、ゾーン総発生トリップ数(Y_j)は、これらの世帯の合計的特性(e.g. 居住人口), を独立変数($\bar{x}_{kj}, X_{kj}$)とする線型モデルで表わされると仮定したものである。これらの3モデルが集計レベルは異なるが同一の世帯特性を独立変数として用いるとすると、それぞれ次式のようになる。

$$\text{世帯モデル} \quad y_{ij} = \sum_k a_k x_{kij} + \epsilon_{ij}, \quad x_{0ij} \equiv 1 \text{ (定数項)} \quad (1)$$

$$\text{ゾーン平均モデル} \quad \bar{y}_j = \sum_k b_k \bar{x}_{kj} + \nu_j, \quad \bar{x}_{0j} \equiv 1 \quad (2)$$

$$\text{ゾーン合計モデル} \quad Y_j = \sum_k c_k X_{kj} + \xi_j \quad (3)$$

ただし、小文字は世帯当りの非集計値なり平均値を示し、大文字はゾーン合計値を示す。また「 $-$ 」はゾーン平均値であることを示す。

$i (= 1, 2, \dots, N_j)$, $j (= 1, 2, \dots, J)$, $k (= 0, 1, 2, \dots, P)$ は、それぞれ世帯, ゾーン, 独立変数を表わす。 N_j はゾーン j の総世帯数である。

ϵ_{ij} , ν_j , ξ_j は、かく乱項でこれらの平均値は 0 と仮定する。

a_k , b_k , c_k は、各モデルのパラメータである。

ゾーン別トリップ発生量(Y_j)の推定は、次のようにして行なわれる。

$$\text{世帯モデル} \quad \hat{Y}_j = \left(\sum_{i=1}^{N_j} \hat{y}_{ij} \right) \times N_j / n_j = N_j \cdot \left(\sum_k \hat{a}_k \bar{x}_{kj} \right) \quad (4)$$

$$\text{ゾーン平均モデル} \quad \hat{Y}_j = N_j \times \hat{\bar{y}}_j = N_j \cdot \left(\sum_k \hat{b}_k \bar{x}_{kj} \right) \quad (5)$$

$$\text{ゾーン合計モデル} \quad \hat{Y}_j = \sum_k \hat{c}_k X_{kj} (= N_j \cdot \left(\sum_k \hat{c}_k \bar{x}_{kj} \right)) \quad (6)$$

ただし、「 $\hat{}$ 」は推定値であることを示す。将来の予測では、 $\hat{N}_j, \hat{x}_{kj}$ となる。世帯モデルの場合、ゾーンの標本世帯数を $n_j (\leq N_j)$ としてくる。

3. 理論的考察

ゾーンモデルでは、トリップ行動の主要な意志決定単位である世帯を地域的に集計する過程で各世帯の行動様式の違いがあまりになてしまうという根本的な問題がある。モデルの推定に伴う問題を明らかにするために、世帯モデルをゾーン平均レベルに集計すると次式のようになる。

$$\bar{y}_j = \sum_i y_{ij} / N_j = \sum_k a_k \cdot \left(\sum_i x_{kij} / N_j \right) + \sum_i \epsilon_{ij} / N_j = \sum_k a_k \bar{x}_{kj} + \bar{\epsilon}_j \quad (7)$$

この結果を対応するゾーン平均モデル(②)と比較すると次の関係式を得る。

$$b_k = a_k, k=0,1,2,\dots,p : \bar{y}_j = \bar{e}_j = \sum_{k=1}^p e_{kj}/N_j \quad (8)$$

同様にして、世帯モデルとゾーン平均モデルをゾーン合計レベルに集計すると次の関係式を得る。

$$c_k = a_k, c_k = b_k, k=0,1,2,\dots,p : \xi_j = \sum_{k=1}^p e_{kj}, \xi_j = N_j \bar{y}_j \quad (9)$$

従って、この3モデルのパラメータには、地域的集計による偏り(aggregation bias)はなく、同一であることがわかる。差は、パラメータ推定に通常の最小二乗法(OLS法)を用いる時のよく乱項の仮定にある。即ち、OLS法ではよく乱項の平均は0、分散は一定(homoscedasticity)、かつ相互に無相関と仮定しているが、この内、平均値が0と無相関性については、この3モデルで符合性があることは、②、④より容易に示される。分散については、トリップ行動単位の世界レベルで一定とするのが妥当であると考えられ、この場合、対応するゾーンモデルでは分散が不均一となり、OLS法によりパラメータを推定すると、偏りはないが推定誤差が大きくなるはずである。

さらに、地域的集計に伴い、いわゆる ecological correlation により変数間の相関係数が行動単位のものとは異なってくることは、ゾーンのとり方によっても変化すること、そして、用いられる集計データに標本誤差が入り、パラメータ推定値の精度が下ることがモデル推定上の主な問題となる。

ゾーン別トリップ数の推定精度は、基本的に同一モデル構造であるため理論的には同じであるが、ゾーンモデルでは、パラメータ推定値の誤差が大きいため精度が劣るはずである。尚、④、⑤、⑥式で明らかにするように、独立変数の推定値の精度は同一である。

4. ケース・スタディ

広島のパソントリップデータ(抽出率5%、拡大前)を用いて、上記の理論的考察・結果の検討を行なった。(下表) ゾーン合計モデルは定数項を含み②式と異なるが、 $b_k = a_k$ が大規模なデータ信頼限界の計算より証明され、またゾーンモデルのパラメータ推定値の精度の劣ることも、大規模なデータ信頼限界の計算より確認された。また、サンプル時のゾーン別トリップ数の推定では、世帯モデルは、標本世帯をさらに10%の抽出率でサブサンプリングしたものを対象としたのにも関わらず、精度の真偽ゾーンモデルの結果と比べ、殆ど遜色のないことがわかった。

5. 結 論

ここで検討した世帯モデルは、その理論的妥当性、目的(ゾーン別トリップ発生量の推定)、推定に伴うコスト(データ、推定手法)の面で、ゾーンモデルよりも一般に勝れていることが以上の理論的および実証的分析により示された。

表 集計レベルの異なるトリップ発生モデルの分析 — 広島・年世帯・全日トリップ—

モデル名	データ数	推定値 平均値	モデルの構造	原 R ²	%SE	ゾーン合計 R ²	%SE
世帯モデル (10%)	1108	8.962	$\bar{y} = 1.655 + 1.929 X_1 + 1.015 X_2$ (9.38) (19.2) (9.92)	.284	52.8%	.981	10.9%
ゾーン平均モデル	102	8.850	$\bar{y} = 2.000 + 1.614 \bar{X}_1 + 2.735 \bar{X}_2$ (1.91) (5.60) (8.06)	.303	10.8	.985	9.9
ゾーン合計モデル	102	975.55	$Y = 7.588 + 2.002 X_0 + 1.372 X_1 + 4.576 X_2$ (1.74) (8.85) (5.32)	—	—	.986	9.8

注 1. 独立変数は $k=1$ 世帯人数, $k=2$ 乗用車数, X_0 はゾーン総世帯数である。

2. () 内は t 値を示す。

3. ゾーンモデルでは 11368 世帯を 102 ゾーンに集計したものをデータとしている。

4. 5% および 1% をサブサンプリングした世帯モデルでは、%SE は 10.7% および 16.1% であった。