通勤交通手段の予測における集計モデルと非集計モデルの予測精度の比較

広島大学　正員　杉恵頼寧
広島大学　学生員　塚本冬文

1 はじめに

交通量の需要推計手法は4段階推定法を中心に近年急速な進歩をとげ、1972年にはアメリカのWilliamsburgにおいて、これまでの発展に対する総括が行われた。その会議の結果、従来の4段階推定法に対するいくつかの問題点が指摘され、今後の研究としては交通政策に*sensitive*な非集計交通行動モデルの研究の重要性が確認された。[1] それ以降、MITを中心に非集計モデルの研究が盛んに行われ、交通政策の評価モデルとして広く用いられるようになってきている。

非集計モデルと従来の4段階推定法を中心とした集計モデルの大きな違いは、分析の単位を前者は個人とすることに対して、後者はそれをゾーンレベルに集計したものを用いることにある。しかし、政策評価としては最終的に予測値の集計化が必要になるので、その予測プロセスは図-1に示したように、モデルの作成の前に集計するのか、あるいはモデル作成後に集計するかの違いになり、予測プロセスそのものは基本的にあまり差がなくなる。

しかし、この集計化の問題は図-1のどちらの流れを取ろうとも大きな問題を含んでいる。モデル作成の前にデータを集計する流れでは、各ゾーンに属する人は似たような交通行動を取るという前提に立ってモデルの作成を行うが、実際には各ゾーンには所得や乗用車の有無により異なる交通行動をとる人々が含まれているので、ゾーン平均値の関係から作成されるゾーン集計モデルでは、交通行動を正しく把握できない問題を有する。例えば、世帯所得が高いほどトリップを多くするという交通行動がある場合、ゾーン内の世帯の所得分布によっては、ゾーン平均値でみた場合には、図-2で示すような逆の相関関係を示すことがありうる。[2]

一方図-1の下方の流れでは、予測値をどのように集計すればよいかという集計問題が生じる。将来の個人個人の社会経済指標や交通サービス特性が判明していれば、それぞれの選択確率を求め、それを加算すればよいが、実際それが不可能のため、各種の集計方法が現在研究されている。最も簡単な方法は各変数の平均値を用いて集計レベルの選択確率を求めるものであるが、モデルが非線形の場合、予測値は一般に偏りが生じる。[3] これは集計バイアス（*Aggregation Biase*）と呼ばれており、このバイアスを少なくするために、いろいろな手法が提案さ

図-2 ゾーン集計による生態学的相関の例

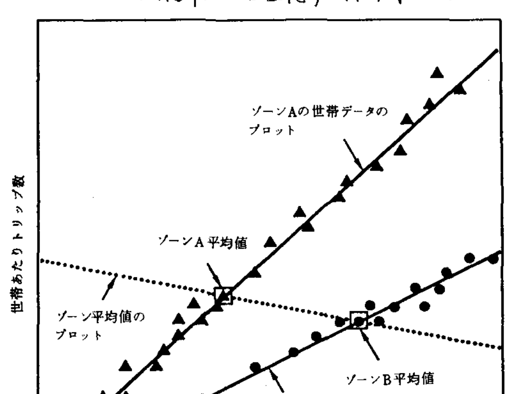
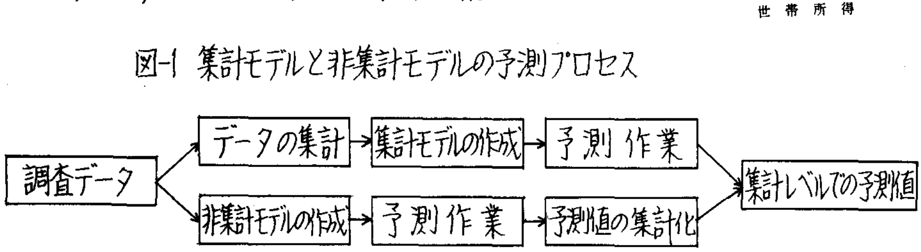


図-1 集計モデルと非集計モデルの予測プロセス



れている。

そこで、本研究はこれらの手法を用いて集計バイアスがどの程度になるかを広島都市圏のデータを用いて検討する。さらに図-1の２つの流れによる最終的な予測誤差を比較し、非集計モデルを実際の問題に適用した場合の有効性を従来の集計モデルと対比させて検討する。非集計モデルとしては乗用車と公共輸送機関（バスと鉄道）の２項選択型のロジットモデルを用いる。集計モデルとしては同じく２項選択型の集計ロジットモデルを用いる。

2 非集計モデルのキャリブレーション

本研究で用いるデータは、公共旅客輸送機関整備基本計画策定を目的として昭和54年９月広島都市圏で調査されたものである。この調査で通勤目的の有効回収票は4,657票であるが、データチェックののち、本研究で用いるのが可能なのは2,395票で、そのうち乗用車通勤は68.1％、公共輸送機関通勤は31.9％である。モデルは次のような２項選択型のロジットモデルを用いる。

$$P_{mi}=\frac{1}{1+\exp\{-a_0-\sum_j a_j(X_{mij}-X_{cij})\}} \qquad (1)$$

ただし、P_{mi}：個人iが公共輸送機関を選択する確率、X_{mij}, X_{cij}：公共輸送機関、乗用車を利用する時の交通機関の選択要因jに対する個人iの値、a_0, a_j：パラメータ（$a_j<0$）

交通機関の選択要因としては、同じデータを用いたこれまでの研究から次のような要因を採用した。[4)]

1) 乗用車の有無（有＝１、無＝０）
2) 通勤先（都心部＝１、それ以外＝０）
3) 総所要時間（自宅から通勤先までの所要時間）
4) 乗車時間（交通機関に乗車している時間）
5) アクセス時間（公共輸送機関の場合は自宅から最寄り駅あるいはバス停までの所要時間、乗用車の場合は零）
6) コスト（公共輸送機関の場合は片道の普通運賃とし、乗用車の場合は次式で求めた。）

$$Y_c = kD + P_r/2 \qquad (2)$$

ただし、Y_c：乗用車のコスト（円）、k：単位走行距離当りのコスト（26円／km）、D：走行距離（ゾーン間の最短距離）（km）、P_r：個人負担の駐車料金（円）

自動車が利用できないためにタクシーを利用する人は、タクシー料金を乗用車のコストとした。

7) エグレス時間（公共輸送機関の場合は降車駅あるいはバス停から勤務先までの所要時間、乗用車の場合は駐車場から勤務先までの徒歩時間とした。）
8) 公共輸送機関の乗換回数
9) 公共輸送機関の運行間隔

これらの変数をいろいろ組み合わせて、最尤法で(1)式のパラメータ、そのt値、的中率、ρ^2値、χ^2値を求めたのが表-1である。

モデル1DIは総所要時間と乗車時間の間には0.847の高い相関があるので、後者と他の物理的指標のみを用いたものである。各変数のパラメータの符号は全てマイナスで論理的となっている。t値は10％の危険率で全て有意である。これらの変数の中でコストのt値が最も高く、交通機関の選択における主要な要因となっている。

モデル2DIはモデル1DIに自動車の有無を加えたものである。この変数のt値は17.2と最も高く、交通機関を選択する上で最も重要な変数になっている。これによってモデルの的中率はモデル1DIよりも５％余り高くなっている。

モデル3DIはモデル2DIにさらに通勤先を加えたものである。この変数のt値も高く、的中率が若干高くなっている。しかし、乗換回数のt値が10％の危険率で有意とならない問題が生じている。

モデル4DIはモデル3DIの乗車時間、アクセス、エグレス時間のかわりに総所要時間を変数として用いたものである。乗換回数のt値はやはり低く、的中率はモデル3DIに比べて若干低下している。

モデル5DIは集計レベルの交通機関別分担の予測によく使われている自動車の有無と総所要時間を変数としたも

のである（たとえば分担率曲線がある）。変数を少なくしたことにより、的中率、ρ^2値が低下している。

モデル6Dも集計レベルの交通機関別分担の予測によく用いられている総所要時間とコストを変数としたものである。モデル5Dよりもさらに的中率、ρ^2値が低くなっているが、t値はモデル5Dと同様いずれの変数も高くなっている。

3 集計モデルのキャリブレーション

前節の非集計ロジットモデルと比較するために、表-1のモデルと同じ変数を用いて集計ロジットモデルのキャリブレーションを行う。(1)式の非集計ロジットモデルと違ってP_{mi}が確率的に与えられるので、(1)式を次のように変換すると重回帰分析によってパラメータa_0、a_iを決定することが可能である。

$$\ln\left(\frac{1}{P_{mn}}-1\right)=-b_0+\sum_j b_j(X_{cnj}-X_{mnj}) \qquad (3)$$

ただし、P_{mn}：ゾーンペアnの公共輸送機関の分担率、X_{cnj}、X_{mnj}：ゾーンペアnの乗用車、公共輸送機関の説明変数jの値、b_0、b_j：パラメータ（$b_j<0$）

(3)式の説明変数は広島都市圏を39ゾーン（昭和53年のパーソントリップ調査のゾーン区分）に分割した時のゾーンレベルの平均値を用いることにする。

その時のデータは前節で用いたのと同じものを用いたが、これをゾーンペアごとに集計すると各ゾーンペアのデータ数は非常に少なくなってしまう問題が生じた。そこで、(3)式の重回帰分析では調査データが10トリップ以上あるゾーンペアのみを回帰分析の対象とした。その結果、全部で1,521のゾーンペアのうち76ゾーンペアのみしか有効とならなかった。このゾーンペアの総トリップ数は1,585で、非集計モデルで用いたデータの66.2%しか

表-1 非集計型ロジットモデルのパラメータ（t値）と適合度

変数＼モデル		1D	2D	3D	4D	5D	6D
1	定数項	−0.101	1.095	0.763	0.726	1.001	−0.313
2	自動車の有無	—	−1.436 (17.2)	−1.415 (16.9)	−1.401 (16.7)	−1.342 (17.9)	—
3	通勤先	—	—	0.603 (10.7)	0.607 (10.8)	—	—
4	総所要時間差(分)	—	—	—	−0.012 (8.4)	−0.014 (10.8)	−0.013 (10.0)
5	乗車時間差(分)	−0.011 (8.3)	−0.012 (8.4)	−0.012 (8.0)	—	—	—
6	アクセス時間差(分)	−0.017 (3.6)	−0.014 (2.7)	−0.014 (2.6)	—	—	—
7	エグレス時間差(分)	−0.017 (4.3)	−0.024 (5.5)	−0.019 (4.3)	—	—	—
8	コスト差(1000円)	−1.091 (12.0)	−1.238 (11.1)	−1.004 (8.9)	−1.027 (9.1)	—	−1.065 (12.2)
9	乗換回数差(回)	−0.115 (2.5)	−0.093 (1.9)	−0.055 (1.1)*	−0.005 (0.1)*	—	—
10	運行間隔差(回)	−0.008 (6.1)	−0.008 (5.3)	−0.006 (3.7)	−0.005 (3.7)	—	—
的中率(%)		75.7	80.8	81.3	80.5	77.0	74.6
ρ^2値		0.125	0.255	0.297	0.293	0.187	0.112
χ^2値		383	772	889	885	563	340

（注1）
（ ）内はt値で符号はパラメータの符号と同じ。
（注2）
*はt検定で有意とならないもの。ただし、危険率は10%。
（注3）
データ数は2,395で公共輸送機関利用は764、乗用車は1,631。

使用できなくなってしまった。

前節で用いた変数のうち、通勤先は0、1のダミー変数としたが、自動車の有無についてはゾーンペア間の自動車の保有率とした。

重回帰分析の結果を示すと表-2のようになる。各モデルごとに表-1のモデルと同じ変数を用いており、非集計モデルと集計モデルが対応して比較できるようになっている。全般的にt値が低くなっており、全ての変数が10%の危険率で有意となるのは説明変数を2個としたモデル5A、6Aだけである。これは個人ごとのデータをゾーンレベルに集計することによって各変数が平均化してしまい、ゾーンペア間であまりバラツキがなくなり、交通機関の分担率の変動を十分説明しきれなくなっているものと考えられる。通勤先に関しては、ダミー変数として用いており、ゾーンレベルでの集計による影響が出ないために表-1のモデルと同様t値が他の変数に比べて相対的に高くなっているものと思われる。

また、パラメータの符号はモデル3A、4Aで乗換回数が論理的に逆になる問題が生じている。これは図-2で説明した生態学的相関によるものと思われる。重相関係数は最も変数の多いモデル3Aで0.787とかなり高くなっているが、変数の少ないモデル5A、6Aは各変数のt値が有意にもかかわらずかなり悪くなっている。

表-1の非集計モデルのパラメータと表-2の集計モデルのパラメータを比較すると、当然ながら両者の間には大きな差が見られるが、説明変数の少ない5Dと5Aおよび6Dと6Aのパラメータの間にはあまり大きな差がないことが特徴的である。

4 非集計モデルと集計モデルの推計精度の比較

図-1の最終的なステップ、すなわち集計レベルで非集計モデルと集計モデルを用いた場合のそれぞれの推計誤

表-2 集計型ロジットモデルのパラメータ(t値)と適合度

変数 ＼ モデル	1A	2A	3A	4A	5A	6A
1 定数項	-0.723	1.019	0.309	0.307	1.168	-0.423
2 車保有率(%)	—	-1.401 (3.2)	-0.927 (2.4)	-0.926 (2.5)	-1.552 (3.4)	—
3 通勤先	—	—	0.507 (4.6)	0.502 (4.9)	—	—
4 総所要時間差(分)	—	—	—	-0.010 (2.0)	-0.014 (2.6)	-0.012 (2.3)
5 乗車時間差(分)	-0.009 (1.7)	-0.012 (2.0)	-0.012 (2.2)	—	—	—
6 アクセス時間差(分)	-0.033 (1.5)*	-0.021 (1.0)*	-0.013 (0.7)*	—	—	—
7 エグレス時間差(分)	-0.023 (1.0)*	-0.038 (1.7)	-0.007 (0.4)*	—	—	—
8 コスト差(1000円)	-1.918 (5.0)	-1.741 (4.8)	-0.462 (1.1)*	-0.511 (1.3)*	—	-1.694 (4.6)
9 乗換回数差(回)	-0.020 (0.1)*	-0.045 (0.3)*	0.159** (1.1)*	0.160** (1.2)*	—	—
10 運行間隔差(回)	-0.009 (3.2)	-0.006 (2.0)	-0.004 (1.2)*	-0.003 (1.1)*	—	—
変数の数	6	7	8	6	2	2
重相関係数	0.644	0.705	0.787	0.783	0.457	0.543
F 値	7.8	9.2	13.0	17.4	9.3	14.6

(注1) ( )内はt値で符号はパラメータの符号と同じ。
(注2) *はt検定で有意とならないもの。ただし、危険率は10%。
(注3) **はパラメータの符号が論理的に逆なもの。
(注4) データ数は76。

差を比較してみる。

非集計モデルの集計法は次の4つの方法を用いた。[5)]

1) 総あたり法

$$P_{mn}=\sum_{i=1}^{T_n} P_{mi}/T_n \qquad (4)$$

ただし、P_{mn}：ゾーンペアnでの交通機関mの分担率、P_{mi}：(1)式によって得られる個人iの交通機関mに対する選択確率、T_n：ゾーンペアnに属する全ての人

表-3 非集計ロジットモデルの集計レベルでの推計誤差（%RMS誤差）

(%)

集計法	交通機関	モデル 1D	モデル 2D	モデル 3D	モデル 4D	モデル 5D	モデル 6D
総あたり法	公共輸送機関	43.6	40.4	32.3	32.4	49.2	46.3
	乗用車	20.3	18.7	15.0	15.1	22.8	21.5
平均値法	公共輸送機関	43.7	—	—	—	—	46.5
	乗用車	20.3	—	—	—	—	21.6
モーメント法	公共輸送機関	43.0	—	—	—	—	45.6
	乗用車	19.9	—	—	—	—	21.1
分類法	公共輸送機関	—	41.0	33.1	33.3	49.2	—
	乗用車	—	19.0	15.3	15.5	22.8	—

2) 平均値法

$$P_{mn}=P_{mi}(\bar{X}_n) \qquad (5)$$

ただし、$\bar{X}_n$：ゾーンペアnの説明変数の平均値

表4 集計ロジットモデルの推計誤差（%RMS誤差）

(%)

交通機関	モデル 1A	モデル 2A	モデル 3A	モデル 4A	モデル 5A	モデル 6A
公共交通機関	38.4	36.3	31.5	31.6	47.9	43.5
乗用車	18.9	17.9	15.5	15.6	23.6	21.4

3) モーメント法

$$P_{mn}=P_{mi}(\bar{X}_n)+Var\{G(X_n)\}\cdot P_{mi}(\bar{X}_n)\cdot\{P_{mi}(\bar{X}_n)-1\}\cdot\{P_{mi}(\bar{X}_n)-\tfrac{1}{2}\} \qquad (6)$$

ただし、$Var\{G(X_n)\}$：$G(X)=a_0+\sum_j a_j(X_{mij}-X_{cij})$の変動

4) 分類法 $$P_{mn}=\sum_{s=1}^{S} P_{mi}(\bar{X}_{ns})\cdot T_{ns}/T_n \qquad (7)$$

ただし、X_{ns}：ゾーンペアnのサブグループsの説明変数Xの平均値、T_{ns}：ゾーンペアnのサブグループsに属する個人の総数、S：サブグループの数

総あたり法は全ての個人の確率をゾーンペアレベルに集計したもので、集計方法の中では最も精度の高いものと考えられている。平均値法はゾーンペアレベルの平均値を非集計モデルに代入して求めるので、集計バイアスが生じる。モーメント法は非集計モデルをテーラ展開により線型近似するので、平均値法より集計バイアスが小さくなる。分類法は母集団をいくつかの同質のグループに分類し、各グループの選択確率を変数の平均値より推定する。全体の分担率はグループの選択確率の重みつき平均値より求める。これによって集計バイヤスをできるだけ小さくしようとする方法である。

これらの4手法を用いて、前節の回帰分析で用いた76ゾーンペアでの各交通機関の分担率を推計し、次式で76ゾーンペア全体のパーセントRMS誤差を計算すると表-3のようになる。

$$\text{パーセントRMS誤差}=\sqrt{\frac{\sum_n (P_{mn}-\hat{P}_{mn})^2}{N}}\Big/\bar{P}_m \times 100\ \% \qquad (8)$$

ただし、P_{mn}、$\hat{P}_{mn}$：ゾーンペアnの交通機関mの分担率とその推計値、N：ゾーンペアの数、$\bar{P}_m$：交通機関mの分担率の平均値

乗用車と公共輸送機関の分担率の比率は約2：1であり、パーセントRMS誤差としてはその逆に1：2となっている。総あたり法では、的中率および$\bar{\rho}^2$値が一番高かったモデル3Dの誤差が一番小さくなっている。しかし、非集計レベルではモデル6Dの方がモデル5Dよりも適合度は悪かったが、集計レベルではモデル6Dの方が精度が良くなっている。同様に、モデル2Dは非集計レベルの的中率がモデル3Dと4Dに比べてほとんど変らないのに、集計レベルの誤差は大きくなっている。これから、的中率の良い非集計モデルが集計レベルでは必ずしも誤差が少ないとは言えないことがわかる。

平均値法はゾーンの平均値をモデルに代入するため、モデル2D、3D、4D、5Dのようにダミー変数を含んだモデルには適用できず、モデル1Dと6Dのみに適用した。初期の予想に反して、総あたり法とほとんど同じ誤差となり、集計バイアスの影響はほとんど生じなかったことがわかる。

モデル1Dと6Dに対してはモーメント法を適用したが、平均値法に比べて誤差はほとんど改善されなかった。これは当然ながら集計バイアスの影響がほとんどないためである。

分類法はダミー変数を有するモデル2D、3D、4D、5Dのみに適用した。モデル2Dと5Dは車保有、非保有による2グループ、モデル3Dと4Dはそれに勤務先を加えた4グループ分類となる。精度としては、総あたり法に比べてほとんど誤差に差がなく、やはり集計バイアスの影響はあまり大きく生じていないことがわかる。

集計ロジットモデルについても同様に、分析に用いた76ゾーンペアを対象にパーセントRMS誤差を計算すると表-4のようになる。表-3の総あたり法に比べて集計ロジットモデルの方が少し誤差は小さくなっているが、全体的に相対するモデル間ではあまり大きな差は見られない。このことから、図-1のどちらの流れに沿っても最終的な段階では、精度に関してあまり大差のないことがわかる。

5 まとめ

本研究で得られた成果をまとめると次のようになる。

1) 従来の集計モデルによる予測と非集計モデルによる予測の違いは、どの段階で集計化を行うかであるが、最終的なゾーンペアの集計レベルでは推計誤差にあまり大きな差はなかった。
2) 非集計モデルを集計モデルとして用いた場合、ゾーンペアレベルでの集計バイアスの影響は非常に小さかった。
3) 非集計モデルは従来長期予測には不向きとされてきたが、表-1のモデル5D、6Dのような変数を用いれば、説明変数の値が長期的に予測しやすいので、長期予測も可能である。その時の推計誤差は集計モデルと比べてあまり大きな差はない。
4) 集計モデルで説明変数を多くすると、有意とならない変数が増えるのに対して非集計モデルは多くの変数を導入することが可能で、人の交通行動を分析するのに有効である。その結果として、非集計モデルの方が交通政策の評価モデルとしての適用範囲が広いと言える。

以上の結果より、非集計交通行動モデルは短期予測のみならず長期予測も可能であり、しかも政策評価モデルとして汎用性が広く、集計バイアスも実用的な面においてはさほど問題のないことがわかった。また、このモデルは非集計レベルの調査データがそのままモデルのキャリブレーションに使用できるため、集計モデルに比べて調査データが少なくてもよいという利点も有しており、今後各方面での活用が望まれる。

参考文献

1) Highway Research Board : Urban Travel Demand Forecasting, Special Report 143, National Academy of Sciences, 1973.
2) Spear, B.D. : Application of New Travel Demand Forecasting Techniques to Transportation Planning, U.S.DOT, pp12-13, 1977.
3) 運輸省大臣官房政策計画官：大都市圏における交通機関選択に関する調査、PP194~202、昭和53年3月.
4) 杉恵頼寧：非集計型ロジットモデルによる短期交通政策の評価、第15回日本都市計画学会学術研究発表会論文集、PP367~372、1980.
5) 太田勝敏：非集計行動モデルの交通計画への適用に関する研究(II)、東京大学工学部都市工学科、PP11~14、1981.