

IV-23

非集計モデルの推定パラメータの安定性と層別化モデルに関する研究

武蔵工業大学 学生員○久保 透、正員 渡邊 隆、正員 岩崎 征人

1. はじめに

本報文は、非集計モデルのパラメータがサンプル数の大小によってどのように変動するか、またモデル精度にどのような影響を与えるかを検討することを目的として行なった。サンプル数とモデルのパラメータの推定精度については、シュミレーションによる検討¹⁾やモデルの移転可能性検討の一環として検討²⁾がなされている。本報文では、既存のアンケート結果を用いてサンプル数の大小と推定パラメータの安定性の関係、及び層別化したデータを用いた非集計モデルについて検討を加えている。

2. 利用したデータ及びモデルの概要

本報文では、横浜市の新交通システム・金沢シーサイドライン(以下SSLとする)導入予定地域の金沢シーサイドタウン(世帯数約7,500 人口約26,000)で1987年8月に実施したアンケート調査より得られたデータを用いている。この調査では、15歳以上を対象としたもので、世帯、個人の属性、日常利用交通手段の実態、その選択理由、SSLのサービスに対する満足度等について解答してもらっている。調査方法は、アンケート用紙を配布し後日回収した。検討の対象としたのは、このアンケートから得られた641サンプルを用いて作成された5種類の交通機関(SSL、自動車、自二・原付、自転車、徒歩)を対象とした多項選択モデルである。

3. サンプル数とパラメータ推定値の安定性について

ここでは、サンプル数とパラメータ推定値の安定性について検討を行なった。サンプル数の設定は、1987年調査データ(サンプル数641)を100~600まで100刻みとしランダムに選択された10個のデータセットを作成した。このデータセットについて、オリジナルモデルで用いられている9つのパラメータがどのように変化するかを解析した。パラメータの安定性は、①も検定、②10個のデータセットから得られたパラメータの変動係数、によって判定している。まず、各データセットを用いて推定したパラメータの t 値が1.96より小さくなる個数をも一つの目安としている。その結果を表-1に示した。

これによれば「アクセスコスト」、「所要時間」、「トリップ頻度」及び「SSLまでの距離」の4パラメータについては、サンプル数が200以上であれば安定していることが認められた。又サンプル数400以上になると個人あるいは世帯で保有している手段「自二・原付保有」、「自転車保有」も採用できることが判明した。

次に、9つのパラメータの変動係数を求めて検討した結果を表-2に示した。なおここではパラメータの安定性を判定する目安として、既存の研究²⁾で用いられている閾値(変動係数 ≤ 0.2)を用いた。これによれば、各推定パラメータのうち「アクセスコスト」、「所要時間」、「トリップ頻度」及び「SSLまでの距離」は概ねサンプル数が300以上であれば安定したものであることがわかる。

これらも検定及び変動係数についての検討から、客観指標である「アクセスコスト」、「所要時間」、「トリップ頻度」及び「SSLまでの距離」などはサンプル数が比較的少なくても安定したパラメータであることが指摘できそうである。一方、性別や年齢など個人属性に関するパラメータは客観指標に比べると不安定であることが判明した。

表-1 t 値が有意でなかった($t < 1.96$)データセット数

パラメータ \ サンプル数	100	200	300	400	500	600
アクセスコスト	5	0	0	0	0	0
所要時間	0	1	0	0	0	0
運行間隔	7	9	10	10	10	10
トリップ頻度	2	0	0	0	0	0
性別	9	10	10	10	10	10
年齢	10	10	10	9	10	10
自二 原付保有	5	4	2	1	1	0
自転車保有	5	4	2	0	0	0
SSLまでの距離	5	0	0	0	0	0

4. 層別化モデルとパラメータ推定値について

層別化モデルは同じ行動目的や

個人属性を持った人々をまとめて
取り扱うことにより、特定集団の
予測に対して非集計モデルの特徴
をいかしたより高い精度のモデル
が得られることが期待される。

しかしながら、データを層別化
するとサンプル数の減少によって
モデルの信頼度に係わる問題が生
じてくることが考えられる。この
ため本報文ではモデルの作成にあ
たり上述の検討結果を踏まえサン
プル数を200以上のものが得られるような
層別化を行なっている。この結果、層別の
分類では行動目的を「全目的」、「通勤・
通学」、「買い物・その他」とトリップ目
的別に分け、個人属性では「男女別」を考
えることとする。ここで推定されたパラメ
ータのうち「アクセスコスト」はトリップ
目的別、性別間で大きな差異はない。しか
し「トリップ頻度」、「SSLまでの距離」は性別
間で差異が認められた。これは、利用手段
に対する両者の意識の違いに起因するもの
と考えられる。つまり男性は行動目的を優
先させるのに対して、女性は、行動目的と
利便性を同時に考える傾向があるのかもしれ
ない。適中率と層別化に関しては、主た
る予測対象であるSSLではやや精度の向上
が認められるが、自転車に関してはモデルに
よって適中率の著しい低下がみられ、この
ことがモデル全体の適中率の低下に影響を
与えているようである。

5. まとめ

パラメータとサンプル数の関係では、各パラメータの特性による違いもあると思われるが、200から300程度
がモデル作成に必要と思われる。また各パラメータの特性も層別化モデルを作成することによって得られた。
しかし本報文では特定地域の同一データを用いた解析結果であるので同じ地域内の異なるサンプル群や他地
域のデータを用いたモデルによる比較検討も必要であろう。

[参考文献]

- 1) 桐越他「非集計ロジットモデルによる選択比率の推定精度に関する研究」、交通工学, Vol. 18, No. 5
- 2) 森地他「非集計交通選択モデルの地域移転可能性」、土木学会論文集, No. 359/IV-3