

1. はじめに

パーソントリップ調査は、中心都市人口 20～30 万以上の都市圏における総合的交通計画のためのデータ収集手段として定着しつつある。データの精度は、分析や予測の信頼性、さらには分析・予測プロセスを経て算定される評価指標の信頼性を規定する。それゆえ、パーソントリップ調査によって得られるデータの精度的特性を知ることが、調査自体を設計する見地、データ収集と計画諸作業との精度的適合性を見地、計画案を評価する見地等から極めて重要である。

パーソントリップ調査では、家庭訪問調査、外部コードン調査、営業車調査等により調査対象地域のトリップデータを採取するが、多くの例では、総トリップ数の 95% 以上は家庭訪問調査によって収集されている。それゆえ、パーソントリップ調査のデータ精度は、家庭訪問調査の精度に支配されていると考えよう。家庭訪問調査により収集されたデータは、通常、トリップの標本として扱われ、これを用いて発生集中交通量、分布交通量等、トリップを単位とする統計量の推定がなされる。そして、これらの交通量の推定の精度を考えるにあたっては、上記のトリップの標本では、トリップ自体がランダムに抽出されているものと仮定している。しかしながら、家庭訪問調査は明らかに人を対象とする調査であり、トリップは人によって集落化されて抽出されている。さらに、実際の家庭訪問調査では、調査対象者もまた世帯によって集落化されて抽出されている。すなわち、家庭訪問調査によるトリップの標本は、人および世帯によって二重に集落化されて収集されたトリップデータにより構成されている。

標本調査では、標本の要素が集落化されて抽出されたとき、要素のもつ標識値が集落内において等質的であるならば、この標本を用いての推定の精度は、要素がランダムに抽出されている標本に比べて低下することが知られている。本研究では、トリップの標本作成時にあける人および世帯による集落化の影響を分析し、家庭訪問調査の調査形態に即した調査精度の推定方法について考察することを目的としている。

2. 集落抽出法における標本誤差

母集団の要素を直接ランダムに抽出して標本を作成する抽出法を単純ランダム抽出法という。他方、母集団の要素を何らかの基準により互に重複のない部分集団に分割した後、部分集団をランダムに抽出し、抽出された部分集団に属する全要素をもって標本を構成する方法を集落抽出法 (Cluster Sampling) という。抽出の対象となる部分集団を抽出単位と言う。家庭訪問調査により採取されたトリップの標本を考えると、まず、トリップの母集団をトリップを行った人を基準に部分集団に分割しており、人をランダムに抽出して調査することにより、これらの部分集団をランダムに取り出していると理解することができる。さらに、多くの家庭訪問調査では、世帯をランダムに抽出し、抽出された世帯の全構成員を調査対象者としている。すなわち、家庭訪問調査によるトリップの標本は、人および世帯により二重に集落化されて抽出された標本である。

要素のもつ数標識の値を y_i とし、母集団における y_i の平均 $\bar{y}$ を推定する場合を考える。単純ランダム抽出法による標本では、推定量 $\bar{y} = \sum_{i=1}^n y_i / n$ 、 n : 標本に属する要素数、による推定の分散は、

$$V_r = (S^2/n) \cdot (N-n)/N \quad (1)$$

$$\text{ここに、} S^2: \text{母分散} = \sum_{i=1}^N (y_i - \bar{y})^2 / (N-1)$$

N : 母集団に属する要素の数

で表わされる。集落抽出法の場合には、母集団がそれぞれ M_i 個の要素を含む i 個の集落 ($N = \sum_{i=1}^L M_i$) に分割され、この内 l 個の集落を抽出して推定量 $\bar{y} = \sum_{i=1}^L \sum_{j=1}^{M_i} y_{ij} / \sum_{i=1}^L M_i$ (i : 集落番号, j : 集落内の要素番号)

によって $\bar{Y}$ を推定する。この推定量の分散は、各集落に含まれる要素の数が等しいとき、すなわち、 $M_i = \bar{M} = N/L$ であるときには、

$$V_L = \frac{L-l}{L} \cdot \frac{1}{lM} \cdot S^2 \left\{ \frac{(\bar{M}l-1)}{(L-1)\bar{M}} + \frac{L(\bar{M}-1)}{L-1} \rho \right\} \quad (2)$$

$$\text{ここに: } \rho = \frac{2 \sum_{i=1}^L \sum_{j \neq k} (y_{ij} - \bar{Y})(y_{ik} - \bar{Y})}{L\bar{M}(\bar{M}-1)S^2} \quad (3)$$

と表わされる。 ρ は集落内相関係数と呼ばれる。集落内相関係数は、(3)式から明らかなように、それぞれの集落内で要素の標識値 y_{ij} が母平均 $\bar{Y}$ よりも大きい傾向にあると小さい傾向にあるとき、言い換えば、標識値が各集落内で等質的であるときには正の値となる。逆に、各集落内で y_{ij} の値が $\bar{Y}$ の周りに分散しており、集落内で標識値が異質的であるとき負の値をとる。

単純ランダム抽出法と集落抽出法における推定量の分散を比較すると、両者の標本規模が等しい、すなわち $n = \bar{M}L$ のときには、母集団に含まれる集落数 L が十分大きければ、

$$V_L/V_r = 1 + (\bar{M}-1)\rho \quad (4)$$

の関係が成立する。すなわち、 ρ が正のときには、集落抽出法は単純ランダム抽出法に比べて、推定量の分散が大きくなり、推定の精度が低下する。

3. 解析の課題

前節に述べた如く、家庭訪問調査によって収集されたトリップの標本は、明らかに、集落抽出法にもとづく標本である。一方、同一個人の行ったトリップ属性の個人内における等質性を考えると、例えば、利用交通機関では、自家用車を使用する人は自家用車利用に特化しがちである。このため、個人内のトリップでは、交通機関属性を自家用車とするトリップが多くなり、等質的となる。このような属性としては、トリップの発地属性、業務等の目的属性、利用交通機関属性等が挙げられよう。一方、世帯内においてトリップ属性の等質性が想定されるものとしては、発地属性、利用交通機関属性等が考えられる。それゆえ、これらの集落内の相関性のため、家庭訪問調査にもとづくトリップの推定量の精度は、通常用いられているトリップの単純ランダム標本を仮定した場合の算定精度に比して低下している恐れがある。従って、以下の事項が検討の課題となる。

(a) 家庭訪問調査によって収集されたトリップ標本にもとづく推定量の標準誤差は、トリップの単純ランダム抽出を仮定した理論的な推定量の標準誤差(後述(6)式)に合致するのか。もし合致しないならば、それは集落抽出の影響であり、今後集落抽出の影響を考慮した精度推定法を用いる必要である。(4)式によれば、両抽出法による推定量の標準誤差の比は $\bar{M}$ および ρ のみに規定されるので、この比の値 ρ を求めておけば、集落抽出法による推定量の標準誤差は、単純ランダム抽出法による標準誤差を用いて

$$\sqrt{V_L} = \rho \sqrt{V_r}$$

によって推定し得ることとなる。

(b) 個々のトリップを単純ランダム抽出することは不可能である。したがって、トリップを人によりクラスター化して抽出することはなめられ得ない。人を抽出する方法には、個人を単純ランダム抽出する個人抽出法と現在の調査で採用されているような世帯抽出法とがある。これら2つの抽出法の精度・費用特性から見た得失を比較し、世帯抽出法の妥当性を検討することが必要である。

4. 再抽出標本を用いた集落抽出の影響の分析

上記の検討を行うために、実際のパーソントリップ調査データを用いた標本の再抽出実験を行なった。すなわち、調査結果として得られているトリップの標本を、ある仮想都市圏におけるトリップの母集団とみなし、これから実際の家庭訪問調査をシミュレートした形で標本の再抽出を行う。このような再抽出標本を集計するならば、その集計値は母数すなわち、原データにおける集計値の周辺にある誤差をもって分布する。したがって、この

ような再抽出標本を複数作成すれば、これらの誤差の散ばりのデータが得られ、設定した抽出方法における推定の標準誤差を実験的に算出することができる。一方、トリップの単純ランダム抽出を仮定した場合には、推定の標準誤差は、原データにおける総トリップ数、推定の対象となる母数および標本の規模から理論的に算出できる。したがって、これらの標準誤差を比較することにより、単純ランダム抽出法と集落抽出法の推定精度と比較することができる。

仮想の母集団としては、才1次東京都群パーソントリップ調査によって採取されたトリップデータを用いた。このデータは、東京都、神奈川県、埼玉県、千葉県、東京都の大部分を対象地域とし、対象人口1,900万人の約2%と世帯を抽出単位として抽出調査したものであり、約32万人80万トリップを含んでいる。再抽出標本は、母集団データと記録した磁気テープから乱数発生プログラムを用いて抽出した。再抽出標本は、(i) 抽出単位として、個人を用いたものと世帯を用いたもの、(ii) 抽出率が、2%、5%、10%、20%の4種、計8ケース作成し、各ケース毎に5本の標本テープを作成した。これらの再抽出標本は、互に標本の独立性が保たれるように作成されている。

パーソントリップ調査を用いた分析・予測プロセスを考慮し、以下のゾーン単位の集計量について、両抽出法の精度比較を行った。(a) 目的別発生交通量・集中交通量、(b) 交通機関別発生交通量・集中交通量、(c) 目的別交通機関別発生交通量・集中交通量。ここに、ゾーンの規模はほぼ市区町村に対応する。

再抽出標本を上記集計項目について集計した結果を kx_{ij}^l (i: 個人抽出、世帯抽出の別、j: 抽出率の別、k: ケース別再抽出標本番号、l: ゾーン番号) で表す。 kx_{ij}^l は、母集団データの当該項目の集計量 μ_l の推定量であり、 μ_l の周辺に確率分布する。したがって、ij ケースにおける μ_l の推定の標準誤差 SD_{ij}^l は、

$$SD_{ij}^l = \sqrt{\frac{1}{n_l} \sum_{k=1}^{n_l} (kx_{ij}^l - \mu_l)^2} \quad (5)$$

として求められる。しかしながら、 SD_{ij}^l もまた確率変数であり、標本数 m が小さいときには変動が甚しい。そこで、母数 μ_l の大きさが標準誤差の大きさに影響することから、ゾーンと母数 μ_l の大きさに従ってカテゴリー分けし、カテゴリーに属する全ゾーンを用いて、カテゴリー別の標準誤差 SD_{ij}^l (k: カテゴリー番号) を算出した。各 SD_{ij}^l の算出には、集計項目によって異なるが、30~60個の集計値が用いられている。

トリップの単純ランダム抽出を仮定した場合の μ_l の推定の標準誤差 TSD_j^l は、(1) 式と総数推定の分散に変形し、

$$TSD_j^l = \sqrt{(\mu_l^2(N - \mu_l)/n_l) \cdot (N - n_l)/(N - 1)} \quad (6)$$

として理論的に求めることができる。上記のゾーンのカテゴリサイズに対応して、 μ_l の代りに、同一カテゴリーに属するゾーンの母数 μ_l の平均 $\bar{\mu}_l$ を用いて TSD_j^l の算出を行った。

5. 結果の概要

前節の過程を経て算出された再抽出標本からの標準誤差 SD と単純ランダム標本における理論的な標準誤差、 TSD と比較し、家庭訪問調査によるトリップの集落抽出の影響を考察する。家庭訪問調査によるトリップの標本に、集落抽出の影響がはく。単純ランダム標本と見做してよいならば、両者の標準誤差は確率的な範囲で一致し、比 $\rho = SD/TSD$ は、1 周辺に確率分布する。この比の値は、個人抽出と世帯抽出の別に、抽出率のケース数とカテゴリー数の積の数だけ得られる。そこで、(i) 各カテゴリー、各抽出率における標準誤差比 ρ の値の平均値 (ii) ρ の平均値の1からの乖離の統計的な意味付けを行った。(5) 式において、 kx_{ij}^l が分散 TSD_{ij}^l をもって μ_l の周辺に確率分布するならば、 $\sum_{k=1}^{n_l} (kx_{ij}^l - \mu_l)^2 / (TSD_{ij}^l)^2 = n_l (SD_{ij}^l)^2 / (TSD_{ij}^l)^2$ は自由度 m の χ^2 分布となる。従って、 $(SD_{ij}^l)^2 / (TSD_{ij}^l)^2$ は、有意水準 α とすると、作業上の近似の範囲内で、

$$\text{区間} [f/\chi^2(f, 1-\frac{\alpha}{2}), f/\chi^2(f, \frac{\alpha}{2})] \quad (7)$$

ここに、 f : 各カテゴリーにおける集計値の数

に確率1- α で分布する。そして、 ρ の平均値の偶然変動の範囲は、(7) 式において、自由度 f と平均値算出に用いた、抽出率およびカテゴリーについて加え合せて場合の区間となる。

主は集計項目における $\bar{P}$ の平均値および $\bar{P}$ の平均値が上述の区間内にあるか否かの判定結果を表1に示している。 $\bar{P}$ の検定結果が棄却であることは、 $\bar{P}$ の1からの乖離は偶然的是なものではなく、トリップの集落抽出の影響が生じていると見すべきことを意味する。

表1では、多くの集計項目において、 $\bar{P}$ は1よりも大きく、家庭訪問調査によって採取されたトリップの標本は、トリップのランダム標本とは見做せないこと、さらには、個人や世帯内におけるトリップ属性の等質性のため

推定の精度が低下していることが分る。世帯抽出法において推定精度の低下の大きい項目は、全目的発生集中量、業務目的発生集中量、全目的交通機関別発生集中量、業務目的交通機関別発生集中量等である。個人抽出法において推定精度の低下の大きい項目もほぼ同様であり、とくに、業務関連の集計項目、鉄道以外の交通機関関連の集計項目で甚しい。一方、通勤関係トリップの如く、個人内の相関が負とみられ、人による集落抽出法によってかえって精度よく推定し得る項目、業務関連トリップの如く、世帯内では属性が異質的であり、世帯抽出の方が個人抽出よりも推定精度がよくなる項目もある。これらの結果は、人の交通行動特性の既知の知識によく一致しており、本研究での解析方法の妥当性と裏付けている。

上述の結果は、家庭訪問調査による交通量推定の精度と、従来の如く、トリップの単純ランダム抽出を仮定した理論的推定式(6)式で行うことには難があることを示している。そして、個人や世帯でのトリップの集落化の影響と考慮し、(家庭訪問調査データからの推定の標準誤差) = $\bar{P} \times$ [単純ランダム標本からの推定の標準誤差] として、精度も表示すべきものである。 $\bar{P}$ の値としては、表1の値と使い得る。抽出率やゾーンカテゴリー別の個々の $\bar{P}$ の値は平均値 $\bar{P}$ の周辺によく集中しており、代表値としての信頼性を有している。

世帯抽出法と個人抽出法の優劣と比較するため、両者の $\bar{P}$ の平均値の比ととり、表1の右欄に示している。表1の値は、世帯抽出による推定量の標準誤差は、個人抽出のそれに比べて増加し勝ちであることを示している。すなわち、同じ大きさの標本では、世帯抽出法は個人抽出法に較べて、最大で15%程度推定精度が低下することと意味する。一方、標本の採取に要する費用と考えると、現在我国で一般に用いられている留置型家庭訪問調査では、一世帯の調査と1人の個人の調査にほぼ同程度の費用と要する。したがって、同じ費用の下では、世帯抽出法は平均世帯構成員数分、すなわち、約3倍の規模の標本と得られる。標本規模の推定の標準誤差への影響は

表1 再抽出標本からの推定量の標準誤差と単純ランダム抽出標本からの推定量の標準誤差との乖離の検定(個人抽出および世帯抽出)

集計項目			個人抽出		世帯抽出		世帯抽出による標準誤差の増加 世帯/個人
			$\bar{P}$ の平均値	平均値の χ^2 検定	$\bar{P}$ の平均値	平均値の χ^2 検定	
目的別発生集中量	全目的	発生集中	1.31	R	1.52	R	1.16
		集中	1.31	R	1.48	R	1.13
	通勤	発生集中	.87	R	.91	R	1.05
		集中	1.02	A	1.00	A	.98
	業務	発生集中	1.50	R	1.48	R	.99
		集中	1.50	R	1.51	R	1.01
	全目的発生集中量	発生集中	.96	R	1.08	R	1.13
		集中	1.00	A	1.12	R	1.11
		バス	1.22	R	1.30	R	1.06
		集中	1.21	R	1.23	R	1.06
		自動車	1.57	R	1.78	R	1.13
		集中	1.56	R	1.77	R	1.14
業務目的発生集中量	鉄道	発生集中	1.42	R	1.64	R	1.15
		集中	1.43	R	1.64	R	1.15
	バス	発生集中	.91	R	.94	A	1.03
		集中	1.02	A	.98	A	.96
	自動車	発生集中	.96	R	1.02	A	1.06
		集中	.98	A	1.06	R	1.07
	徒歩	発生集中	.97	A	1.06	R	1.09
		集中	1.01	A	1.07	R	1.06
	業務関連	発生集中	.98	A	1.08	R	1.10
		集中	1.00	A	1.10	R	1.09
業務目的発生集中量	鉄道	発生集中	1.12	R	1.11	R	.98
		集中	1.08	R	1.01	A	.94
	自動車	発生集中	1.46	R	1.45	R	.99
		集中	1.38	R	1.46	R	1.06
	徒歩	発生集中	1.37	R	1.22	R	.89
		集中	1.32	R	1.25	R	.95

R:棄却 A:受容

(6)式にも明らかた如く、 $1/\sqrt{n}$ に比例する。すなわち、3倍の標本規模があるならば、標準誤差は、約 $1/1.73$ となり、このことの影響は個人と世帯によって集落化することによる精度低下よりも大きい。それゆえ、推定量の精度面から考えると、世帯抽出法を用いた家庭訪問調査は妥当であると言える。

6. 生成原単位における世帯抽出の影響

生成原単位推定における世帯内での個人の交通行動特性の等質性の影響を見るため、(3)式に従い集落内相関係数を算出した。すなわち、オセ帯のオセ構成員の1日のトリップ数を y_{ij} として ρ を算出している。目的別および交通機関別生成原単位における集落内相関係数を表2に示す。ここで、世帯内における構成員数は異なるため、集落内で要素数が等しいとの仮定が成り立たない。したがって、表2の標準誤差比は(4)式と近似式として用いて算出している。すなわち、

$$\rho' = \sqrt{V_a/V_r} \approx \sqrt{1/(n-1)}\rho \quad (8)$$

である。生成原単位集計においても、世帯抽出の影響は現われ、大よそ10%の標準誤差の増加が生じている。一方、4.5節で述べた再抽出標本による方法では、全目的および勤務先への生成原単位における標準誤差比は、それぞれ、1.11および0.99である。これは、両方法における算出プロセスを考えると十分は一致性をもっていると考えられる。表2の結果は、納得し得る傾向を示しているし、また、表1の結果とも符合した傾向を示しており、本研究で用いた2種の標準誤差比算出法は十分妥当なものとして理解してよいであろう。

7. おわりに

本研究では、パーソントリップ調査のデータ精度に及ぼす集落抽出の影響を考察し、以下の結論を得た。

- (1) パーソントリップ調査における推定量の標準誤差は、従来、の如く、トリップがランダムに抽出されていると仮定して標準誤差を算定するのでは不十分であり、人や世帯を抽出単位とすることによるトリップの集落抽出の影響を考慮に入れることが必要である。
 - (2) 人や世帯を抽出単位とするトリップ標本からの推定の標準誤差は、トリップの単純ランダム標本からの推定の標準誤差に、集計項目によって定まる定数 ρ を乗じて算定することができる。これらの定数の値を表1の如く求めた。
 - (3) 抽出単位としては、個人抽出と世帯抽出が考えられる。この両者と比較すると、同じ規模の標本では、多くの集計項目において個人抽出法の方が10%程度推定精度が高い。他方、標本の採取に要する費用を考慮すると、同じ費用のもとでは、世帯抽出法の方が高い推定精度をもつ標本を供すると考えられる。
- 人や世帯内におけるトリップ属性の集落内相関は、人あるいは世帯の交通行動様式の影響を受けると考えられる。それゆえ、時間的あるいは地域的な意味での集落内相関の安定性について、今後検討する必要がある。

最後に、本論文作成にあたって御助言頂いた北海道大学五十嵐日出夫教授に謝意を表する。

主な参考文献： 有藤金一郎、浅井晃；標本調査の設計、培風館、宮沢光一；近代統計概論、培風館、中山伊知郎編；現代統計学大辞典、東洋経済新報社、東京都市群パーソントリップ調査報告書 他

表2 生成原単位における世帯内相関係数の計算値

項 目	ρ の計算値	分散比 V_a/V_r	標準誤差比
全目的トリップ(数/人)	.1376	1.241	1.11
目的別トリップ(数/人)			
勤務先へ	.0278	1.048	1.02
通学先へ	— .0231	.960	.98
帰宅	.1290	1.226	1.11
業務先へ	.0208	1.036	1.02
買 物	— .0358	.937	0.97
社交・その他	.0775	1.136	1.07
交通手段別 トリップ(数/人)			
鉄 道	.1097	1.192	1.09
バ ス	.0672	1.118	1.06
自家用車	.0708	1.124	1.06
二輪車	.1141	1.200	1.10
徒 歩	.0742	1.130	1.06