

## IV—15 集落抽出法と考慮したパーソントリップ調査のデータ精度に関する研究

北海道大学工学部 正員 山形 耕一

## 1. はじめに

パーソントリップ調査(以降、PT調査と記す)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基礎調査として定着しつつあり、北海道においても昭和47年度第一次道央都市圏、57年度旭川都市圏、58年度第二次道央都市圏の各調査が実施されている。しかしながら、PT調査によって得られるデータの精度については不明確な部分が残されている。PT調査の標本誤差は、従来、標本として採取されたトリップや人ばらランダムであるという仮定のもとに、標本調査理論にもとづき推定されてきた。しかしながら、同調査の過程では、トリップの標本は、調査対象者の1日の交通行動を観察することにより採取されている。このことは、トリップをトリップメーカーによって集落化して抽出していることとなる。さらに、通常用いられている世帯を抽出単位とする調査では、調査対象者も世帯により集落化して抽出している。集落抽出法により採取された標本では、多くの場合、同一集落に属する要素の標識値が何らかの意味で等質的であるため、推定精度は、要素をランダムに抽出した標本に比べて低下する。そこで、本論文では、集落抽出の影響を考慮に入れ、PT調査の調査形態に即したデータ精度の推定方法を開発することを目的とする。

PT調査における調査項目は、大別すると、トリップを観察単位とし、特定の属性をもつトリップの総数や比率を推定するもの(例えば、発生交通量、分布交通量、交通手段別分担率)と、人を観察単位とし、人のもつ数標識の平均(例えば、生成原単位)や特定の属性をもつ人の総数(例えば、属性別人口)等を推定するものとに分類することができる。トリップ数を観察単位とする推定における集落抽出法の影響については、参考文献1で報告しているので、本論文では、人を観察単位とする推定について論ずる。

## 2. 平均の推定における集落抽出法の影響

PT調査におけるデータの大部分は、家庭訪問調査により収集されるので、以降家庭訪問調査を対象に検討を進める。家庭訪問調査では、通常、住民登録台帳から世帯を系統抽出し、抽出された世帯の全構成員を調査対象者としている。これは、調査対象者の集落抽出である。そこで、集落抽出のデータ精度への影響を、生成原単位を対象に検討する。

いま、母集団 $\pi$ は $N$ 個の集落(世帯)から構成され、各集落は、それぞれ $M_j$ 個の要素(個人)を含んでいるとする。そして、各要素は $x_{jk}$ なる数標識を持っているものとする。このとき平均の推定問題は、

$$\bar{X} = \frac{\sum_{j=1}^N \sum_{k=1}^{M_j} x_{jk}}{\sum_{j=1}^N M_j} \quad (1)$$

を推定することとなる。母集団から、 $n$ 個の集落を抽出し、抽出された集落に属する $\sum_{j=1}^n M_j$ 個の要素から、 $\bar{x}$ を推定する。推定量としては、

$$\bar{x}_d = \frac{\sum_{j=1}^n \sum_{k=1}^{M_j} x_{jk}}{\sum_{j=1}^n M_j} \quad (2)$$

が一般的である。推定量 $\bar{x}_d$ の分散は、各集落に属する要素の数が全て等しい場合、すなわち、 $M_j = \bar{M}$ であれば、

$$V_d = \frac{N-n}{N} \frac{1}{n\bar{M}} S^2 \left\{ \frac{\bar{M}N-1}{(N-1)\bar{M}} + \frac{N(\bar{M}-1)}{N-1} \rho \right\} \quad (3)$$

$$\text{ここに;} \quad \rho = \frac{\sum_{j=1}^N \sum_{k=1}^{M_j} (x_{jk} - \bar{X})(x_{jk'} - \bar{X})}{\bar{M}N(\bar{M}-1)S^2} \quad (4)$$

$$S^2 = \frac{\sum_{j=1}^N \sum_{k=1}^{M_j} (x_{jk} - \bar{X})^2}{(\bar{M}N-1)} \quad (5)$$

と表わされる。一方、標本の要素を、直接、単純ランダム抽出した標本を考える(以降、単純ランダム標本と呼ぶ)。集落抽出法の場合と同じ大きさ、すなわち、 $\bar{M}n$ 個の要素を含む単純ランダム標本から、母平均 $\bar{X}$ を推定

すると、分散  $V_r$  は、明らかに

$$V_r = (S^2/\bar{M}n)(\bar{M}n - \bar{M}n) / \bar{M}n \quad (6)$$

である。したがって、 $V_e$  を  $V_r$  で除してやれば、集落抽出により推定の分散がどの位大きくなるか知ることができる。(3)式を(6)式で除すと、 $N$  が十分大きいときには、

$$V_e/V_r \doteq 1 + (\bar{M} - 1)\rho \quad (7)$$

となる。ここで、 $\rho$  は集落間相関係数と呼ばれ、(7)式から明らかなように、 $\rho$  が正のときには集落抽出法の推定の精度は、単純ランダム抽出法に比べて低下するし、 $\rho$  が負の場合には、集落抽出法の方が精度よく推定を行なえることとなる。 $\rho$  の値は、(4)式から分かるように、ある集落  $j$  に属する  $x_{jk}$  の値は、そのほとんどが  $\bar{X}$  より大きく、他の集落  $j'$  に属する  $x_{jk}$  は、ほとんどが  $\bar{X}$  より小さいという如く、 $x_{jk}$  の値が集落内で等質的であるとき正になる。他方、集落内の  $x_{jk}$  の値が  $\bar{X}$  の周りに散在している場合、すなわち、集落内  $x_{jk}$  の値が異質的であるとき、 $\rho$  は負となる。

上述の如く、集落抽出法を用いた場合、推定の精度は、集落内における標識値の等質性・異質性の影響を受ける。この等質性は、世帯内における各構成員の交通行動の関連性から生じていると考えられる。例えば、ある世帯では、各構成員が特定の交通手段をしばしば利用するのに、他の世帯ではほとんど使わないという傾向があれば、その交通手段別生成原単位に係わる集落間相関は正となる。そして、このような場合には、世帯を抽出単位として標本を作成する方法による推定量の精度は、人を直接にランダム抽出する方法による精度よりも低下する。そこで、目的別および交通手段別生成原単位における集落間相関係数を算出することにより、集落抽出を考慮に入れた精度推定を行うと共に、人の単純ランダム標本からの推定に比べ、どの程度の精度低下が生じているかを考察する。また、精度の指標としては、推定量の標準誤差を用いる。

集落間相関係数は、東京都市群第一次パーソントリップ調査(昭和47年実施)の調査結果のうち、5%を再抽出したデータを用い(4)式に基いて計算した。結果を表1に示す。ここで用いている(3)、(4)式は、各集落に属する要素の数は等しいとの仮定のもとに導かれたものである。一方、世帯に属する構成員数は異なるので、上記仮定は成立しない。それゆえ、(3)、(4)式は、本研究の場合、近似式としてしか成立しない。そこで、近似式としての有効性を確かめるために、以下の再抽出実験の方法を採用した。

第一次東京都市群パーソントリップ調査の結果を、ある仮想都市圏における人の母集団と考える。この仮想母集団は、約10万世帯、32万人のデータを含んでいる。この仮想母集団から世帯を抽出単位として、定められた抽出率に従い標本を再抽出する。この再抽出標本を集計し、ゾーン  $l$  に居住する人の生成原単位を推定する。いま、抽出率  $\alpha\%$  の場合、第  $\beta$  番目の標本からの  $l$  ゾーン居住者の生成原単位の値を  $x_{\beta l}^{\alpha}$  と表す。このような、再抽出標本を  $B$  本抽出したとすれば、世帯抽出法による生成原単位  $x_l$  の推定の標準誤差は、

$$SE_{\beta l}^{\alpha} = \sqrt{\frac{\alpha}{B} \sum_{\beta=1}^B (x_{\beta l}^{\alpha} - \mu_l)^2 / B} \quad (8)$$

として推定される。ここで、 $\mu_l$  は仮想母集団の集計値であり、 $x_l$  の真の値である。しかし、 $SE_{\beta l}^{\alpha}$  自体も1つの推定値であり、 $B$  の数が少ない場合には、大きく変動する。本研究では、 $B=5$  である。その不足を補うため、 $\mu_l$  の値がほぼ等しいゾーンをカテゴリーとしてまとめ、カテゴリー毎に  $SE_{\beta l}^{\alpha}$  ( $\beta$  はカテゴリー番号) を算出した。このことにより、 $SE_{\beta l}^{\alpha}$  の推定には、30~60個の  $x_{\beta l}^{\alpha}$  が用いられている。<sup>1)</sup>

一方、個人の単純ランダム標本を用いた場合の推定の標準誤差  $TSE_{\beta l}^{\alpha}$  は、(6)式に  $S^2$  および標本の大きさを代入すれば、理論的に算出できる。 $S^2$  の値は、仮想母集団を対象に(5)式により集計した。

$SE_{\beta l}^{\alpha}$  は、世帯を抽出単位とすることの影響を内在しているので、同じ規模の標本についての  $SE_{\beta l}^{\alpha}$  と  $TSE_{\beta l}^{\alpha}$  の比

$$J_{\beta l}^{\alpha} = SE_{\beta l}^{\alpha} / TSE_{\beta l}^{\alpha} \quad (9)$$

を作れば、 $J_{\beta l}^{\alpha}$  は集落抽出に起因する精度の低下を表わしている。 $J_{\beta l}^{\alpha}$  は、抽出率のケース数とカテゴリー数の積の数だけ算出される。これらの  $J_{\beta l}^{\alpha}$  は、期待値は共通であるが、分散は  $\alpha$  および  $\beta$  の値により異なる。そこで、

$SE_{\alpha}$  推定に伴う信頼区間中の逆数を重みとする加重平均 $\bar{P}$ をもって、当該項目の標準誤差比とした。全目的および通勤の生成原単位につき $\bar{P}$ の値を求めた結果を表2に示す。

表1および2の結果をみると、まず、集落間相関係数を算出する方法によった場合の標準誤差比と再抽出実験による標準誤差比が極めてよく一致していることが指摘される。それゆえ、集落間相関係数の算出する方法は、十分な近似性を持っていると判断される。集落抽出の影響をみると、通勤、通学、業務、買物の生成原単位では、標準誤差比はほぼ1であり、単純ランダム抽出による推定に比べて精度の低下は生じていない。これらの目的のトリップは、ある限られた属性の人、例えば、業務は就業者のみが行うが、このような属性をもつ人は世帯構成員の一部であることから、集落間相関係数は0または負となっているものと考えられる。他方、全目的、帰宅、社交目的および交通手段別生成原単位では、標準誤差比は約1.1であり、集落抽出のため、約10%推定の精度が低下している。これらの目的のトリップでは、世帯による行動の活発さの違いの影響を受けていると考えられる。また、交通手段別生成原単位については、世帯構成員が住居を同じくするため、各交通手段へのアクセス条件が共通であること等により、世帯内の交通手段選択に相関性があるためと考えられる。

集落抽出法を考慮した推定の標準誤差は、表1や後述表3に示される標準誤差比 $P$ を、同じ規模の単純ランダム標本からの推定の標準誤差 $TSE$ に乗ずることによって推定し得る。このことは、もし $SE_{\alpha}$ の真の値が $P \cdot TSE_{\alpha}$ であるならば $(SE_{\alpha} / P \cdot TSE_{\alpha})^2$ が自由度 $B^*$ の $\chi^2$ 分布をすることを用いて検定できる。全目的生成原単位の場合、 $\alpha$ 、 $\epsilon$ を要にする20個の $SE_{\alpha}$ の全てが、この検定で受容となった。

### 3. 総数の推定における集落抽出の影響

総数の推定の場合には、第2節の議論において、 $x_{jk}$ を $j$ 要素が着目する属性をもつとき1、然らざるとき0なる値を取る変数として定義すれば、そのまま成り立つ。前節と区別をするため、このような確率変数を $y_{jk}$ とする。このとき、総数の推定は $y_{jk}$ が1である要素を数え上げること、すなわち、 $Y = \sum_j \sum_k y_{jk}$ の推定である。また、 $y_{jk}$ の平均、 $\bar{Y} = Y / \sum_{j=1}^N M_j$ は母比率 $P$ となる。

総数の推定において、集落間相関係数は、前節と同様に $M_j = \bar{M}$ を仮定すると、(4)式より、

$$\rho = \frac{\sum_{j=1}^N \sum_{k \neq l}^M (y_{jk} - P)(y_{jl} - P)}{\bar{M}N(\bar{M} - 1)S^2} \quad (10)$$

となる。ここで、第 $j$ 集落で $y_{jk}$ が1である要素の数を $g_j = \sum_{k=1}^M y_{jk}$ とすると、(10)式の分子は

$$\begin{aligned} \sum_{j=1}^N \sum_{k \neq l}^M \{y_{jk}y_{jl} - (y_{jk} + y_{jl})P + P^2\} &= \sum_{j=1}^N \{g_j(g_j - 1) - 2g_j(\bar{M} - 1)P + M(\bar{M} - 1)P^2\} \\ &= \sum_{j=1}^N (g_j - \bar{M}P)^2 - \bar{M}NP(1 - P) \end{aligned} \quad (11)$$

と表わされる。(11)式の第2項は、 $(\bar{M}N - 1)S^2$ となり、集落の特性とは無関係であるため、 $P$ の大きさは、第1項に支配される。 $\bar{M}P$ は各集落に含まれる $y_{jk}$ が1の要素の期待値であるため、(11)式は、 $y_j$ の期待値周りの分散が大きいき、言い換えれば、集落が $y_j$ が大きいグループと小さいグループに分かれているような場合に大きくなる。(11)式を(3)式に代入し、

$$S_{\alpha}^2 = \sum_{j=1}^N (g_j - \bar{M}P)^2 / (N - 1), \quad S^2 = \bar{M}NP(1 - P) / (\bar{M}N - 1) \quad (12)$$

とすると、集落抽出法による総数推定の分散は、

表1 生成原単位推計における  
世帯抽出の影響

| 項目       | $\rho$ | $SE_{\alpha}/SE_{\alpha}$ |
|----------|--------|---------------------------|
| 目的別原単位   |        |                           |
| 全目的      | .1376  | 1.11                      |
| 通勤       | .0278  | 1.02                      |
| 通学       | -.0231 | .98                       |
| 業務       | .0208  | 1.02                      |
| 買物       | -.0358 | .97                       |
| 社交・娯楽    | .0775  | 1.07                      |
| 帰宅       | .1290  | 1.11                      |
| 交通手段別原単位 |        |                           |
| 鉄道       | .1097  | 1.09                      |
| バス       | .0672  | 1.06                      |
| 自家用車     | .0708  | 1.06                      |
| 二輪車      | .1141  | 1.00                      |
| 徒歩       | .0742  | 1.06                      |

表2 再抽出実験による世帯抽出法  
と個人抽出法との標準誤差比

| 項目      | $SE_{\alpha}/SE_{\alpha}$ |
|---------|---------------------------|
| 生成原単位   |                           |
| 全目的     | 1.11                      |
| 通勤      | .99                       |
| 職業別人口   |                           |
| 管理職     | 1.01                      |
| 専門職・事務職 | 1.11                      |
| 販売業     | 1.27                      |
| サービス業   | 1.18                      |
| 生産工程    | 1.12                      |
| 主婦      | 0.77                      |

$$V_{cl} = (N-n)/N \cdot (1/\bar{M}n) (S_{cl}^2/\bar{M}) \quad (13)$$

と表わされる。一方、標本の規模が  $\bar{M}n$  のときの単純ランダム標本からの推定の分散は、

$$V_r = (N-n)/N \cdot (1/\bar{M}n) S^2 \quad (14)$$

であるから、2つの抽出法における標準誤差比は、次式で表わされる。

$$SE_{cl}/SE_r = V_{cl}/V_r = (1/\bar{M}) (S_{cl}^2/S^2) \quad (15)$$

(15)式を用いた場合の職業別および産業別人口推定における標準誤差比を表3に示す。また、再抽出実験にもとづく、検証のための職業別人口推定における標準誤差比を表2に示す。まず、表3と表2の標準誤差比は極めてよく一致し、(15)式による標準誤差比の推定が、集落に属する要素数が等しいという仮定にも係わらず、十分信頼のおけるものであることが分かった。集

表3 属性別人口推定における世帯抽出の影響

落抽出法による推定精度の低下は、生成原単位の場合に比べて大きくなっている。特に、販売業、農林漁業従事者、学生・生徒などで甚しい。販売業や農林漁業では、世帯全員で事業を営んでいる場合が多く、世帯内での職業、産業構成は均質的である。また、学生・生徒に関しては、子供のいる世帯とない世帯を二分できることが影響していると考えられる。他方、主婦に関しては、ほとんどの世帯では主婦が1人いるため、世帯内における標識値が均質的になっているためと考えられる。このように、表3の結果は、世帯に関する既存の知識と符合している。

#### 4. 結論

本研究では、パーソントリップ調査における調査項目のうち、人を観察単位とする項目について、世帯を抽出単位とする集落抽出のデータ精度への影響を考察した。主な成果は

以下のようにまとめられる。第一には、生成原単位や属性別人口推定のような人を観察単位とする推定において、世帯抽出法によって得られた標本からの推定量の標準誤差は、人をランダムに抽出して作成した標本からの標準誤差よりも大きい。すなわち、パーソントリップ調査によって得られた標本を単純ランダム標本と仮定して精度を考えることには問題がある。第二には、集落抽出の影響を検討するために、集落間相関係数を算出する方法が簡便かつ信頼のおける方法であることが分かった。第三には、集落抽出を考慮に入れたデータの精度は、上記の方法により算出された集落間相関係数を(7)式に代入して標準誤差比  $\rho$  を求め、これを単純ランダム標本における標準誤差に乗じることによって推定し得ることが分かった。集落間相関係数は、世帯内の人の行動や属性の相関によって生じるため、地域や時代によって異なる可能性があるため、今後、集落間係数の地域的、時間的安定性について検討していきたい。

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次第である。

参考文献 1) 山形耕一、パーソントリップ調査の調査精度に対する抽出単位の影響について、第5回土木計画学研究発表会講演集、2) 中山伊知郎編、現代統計学大辞典、東洋経済新報社、3) 宮沢光一、近代統計概論、培風館、4) 東京都市群パーソントリップ調査報告書