

## 傾向スコアを用いた幹線旅客純流動調査拡大係数設定法の再検討

○ [土] 重白 結一 (広島大学)

[土] 塚井 誠人 (広島大学)

[土] 奥村 誠 (東北大学)

# Reconsideration of the Magnification Factor Setting of Inter-Regional Passenger Travel Survey; by Propensity Score Method

○Yuichi Shigeshiro, Makoto Tsukai, (Hiroshima University)

Makoto Okumura, (Tohoku University)

In the transport demand analysis of inter-regional passenger travel, transfer information over the different modes is acquired from the disaggregated questionnaire survey, and the magnification factor of each sample is calculated by referring to the ratio between net and gross traffic of each link. However, since the conventional procedure does not take count in the socio-demographic characteristics of the sample, the magnified share of them would be biased. Such the bias would cause the unexpected biases on the some estimates of elasticity between modal choice probability and socio-demographic characteristics, or price. Our study discusses the background of the occurrence of such the bias, and proposes a novel magnification factor calculation procedure by using propensity score. The empirical study clarified its applicability.

キーワード：旅客需要，価格弾力性，個人属性構成比

Key Words：passenger travel demand, price elasticity, share of socio-demographic characteristics

## 1. 背景・目的

都市間交通について 1990 年までは、交通機関別に実施された調査データしか存在していなかった。それ以前にも、交通機関別の調査は実施されていたが、異なる調査によって得られたサンプルを整合的に集計する方法の開発が進んでいなかったため、例えば複数の交通機関を乗継いだトリップや複数の交通断面を通過したトリップは、観測断面毎に複数回のトリップと数えられていた。

そこで 1990 年に、全ての交通機関を対象とする全国幹線旅客純流動調査が開始された。この調査では、各交通機関で調査されたデータを（異種）交通機関間乗り継ぎ処理を行い、標本を拡大・集計することによって純流動 OD 表を推定し、旅客の交通目的やトリップの真の発着地を把握できるようになった。ここで、総流動と純流動の違いについて、図 1 に示しておく。

乗り継ぎ補正を含む全国幹線旅客純流動調査の拡大係数



図 1 純流動と総流動のトリップ数

設定手順の概略は、次の通りである。同調査では、効率的に標本を得るために、調査日に都市間トリップを行った人を対象とする choiced-based sampling 方式をとっている。例えば航空便に対しては、調査対象便ごとに標本調査を行う。ここでもし調査対象便に乗り継ぎ旅客が含まれていなければ、その便の総流動旅客数と純流動旅客数は一致し、標本の拡大係数は、調査日の便別輸送旅客数と標本数の逆数として設定すればよい。一方で多くの鉄道利用者のように、異なる運行便間で乗り継ぎを行った旅客が存在すれば、その標本分だけ、断面別の総流動を割り引いて拡大係数を設定される。すなわち、上述した手順によって得られ

る拡大係数は、純流動として正しい値を推計するための値であり、標本の個人属性や交通行動特性を考慮した拡大方法は取られていない。

一方で、交通サービスの質的な改善を行う上では、地域間の旅客需要量を表す集計的な純流動量と非集計的な旅行者の行動特性の関係を明らかにすることが求められる。たとえば、旅客の交通目的、年齢、性別などの個人属性が、交通機関や経路などの選択に及ぼす影響は、交通機関や利用経路選択モデルの推定によって行われる。その際、どの説明変数が交通機関や経路の選択に対して有意となるかは、標本中の個人属性構成比の影響を受けるため、モデルの推定には標本拡大係数が必要となる。しかし先述した理由により、総流動観測データとの整合性を担保するように設定した拡大係数は標本の個人属性に関する情報を用いていないため、そのような拡大係数を用いて旅客行動特性を分析することは、望ましくない可能性がある。

以上より本研究では、非集計的な旅行者の行動特性を明らかにする上で必要な個人属性構成比の偏りの補正に着目して、旅行者の個人属性が示す傾向を統計的に処理する手法である、傾向スコア法の適用可能性を検討する。

## 2. 純流動調査と国勢調査の個人属性比較

本節では、標本調査である純流動データを従来の拡大係数で拡大集計した結果と、全数調査である国勢調査の個人属性構成比を比較する。なお年次は 2005 年、対象は 207 生活圏ゾーン単位、旅行目的は全目的、または業務として、発生トリップについて集計を行う。

図 2 に生活圏別に集計した純流動調査の業務トリップ人口と国勢調査の居住人口の散布図を、図 3 に純流動調査の業務トリップ人口と国勢調査の就業人口の散布図を、それぞれ示す。またトリップ人口と居住人口は、概ね比例関係にあるが、人口当たりのトリップ発生頻度（トリップ生成原単位）のばらつきが大きいことがわかる。

図 4 に生活圏別に集計した国勢調査の生活圏別男女別構成比を、図 5、および図 6 に、業務純流動の生活圏別男女別構成比を示す。図 4 より、国勢調査の男性人口構成比は全生活圏ともほぼ 50% 程度であり、生活圏間のばらつきも数ポイントであるのに対して、図 5、図 6 に示す純流動調査では、ほとんどの生活圏で男性が 50% 以上であり、生活圏間のばらつきも大きい。

生活圏間のばらつきは、幹線旅客純流動調査が特定日のみの標本調査のため、必然的に発生する。業務純流動を行う男性が女性よりも多いことは、男性の就業率から予想される傾向である。さらに業務目的の発生トリップ原単位の一部は、明らかに生活圏別の産業立地特性や生活圏のアクセシビリティの影響を受ける。

しかしこれらの他に、純流動調査において、個人属性構成比に大きなばらつきが生じる理由として、拡大係数の設定方法自体に課題がある可能性がある。次節では、調査方法の特性を踏まえた、拡大係数の設定法について論じる。

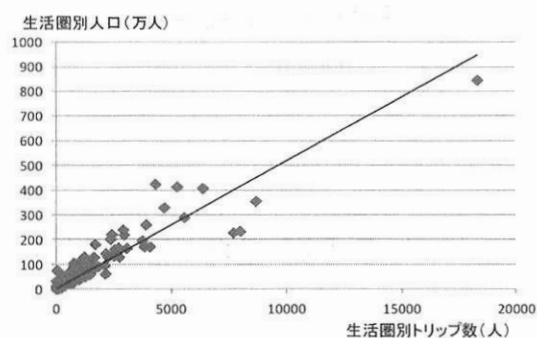


図 2 国勢データと純流動データの比較（人口）

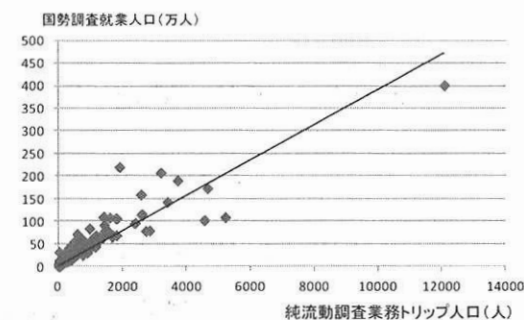


図 3 国勢データと純流動データの比較（就業人口）

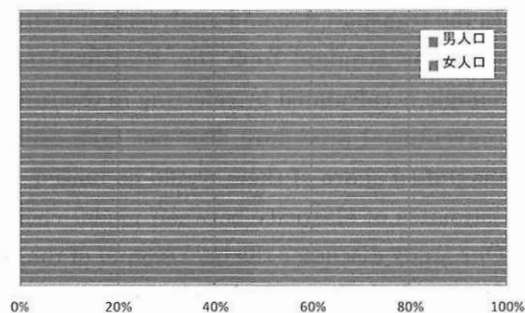


図 4 生活圏別男女比（国勢調査）

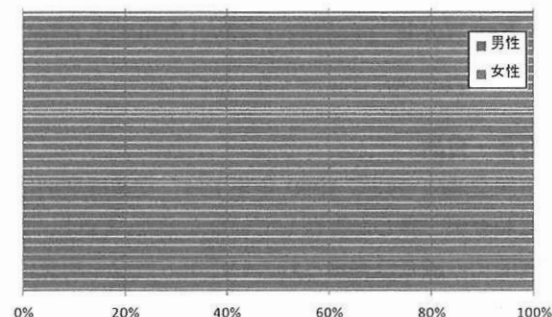


図 5 航空利用生活圏別男女比（純流動調査）

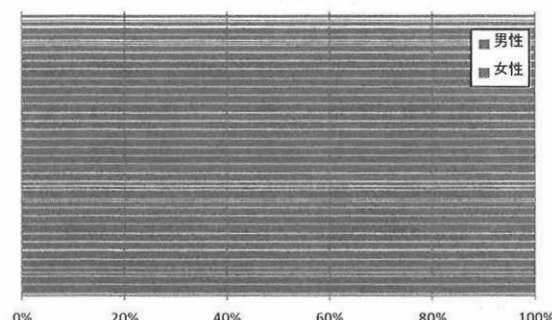


図 6 鉄道利用生活圏別男女比（純流動調査）

### 3. 全国幹線旅客純流動調査の課題

全国幹線旅客純流動調査の個票利用に関する課題は大別して2種類あり、それらはトリップを行っている個人に対して理想的なランダムサンプリングが実施できないことに起因して起こる。以下、調査対象者の設定を「抽出」、調査対象者が調査票に回答して標本となることを、「標本化」と呼ぶ。まず **Choiced based sampling** では、「トリップを行った人」しか抽出されないため、「ゼロトリップの人」に関する情報は、標本に含まれない。よって得られた標本を用いて、「ゼロトリップの人」に対する推測を行うことは、統計的には「外挿」となり、予測結果の安定性は低い。たとえば、両セグメントに属する標本の個人属性構成比が異なる可能性は十分にあるが、その構成比の違いを無視して LOS 変化が「ゼロトリップの人」に新たにトリップを起こさせる誘発効果を推測することは不適切と考えられる。このような問題は、たとえば各旅客が利用した券種情報を個票ベースで調査し、その結果と集計的な需要の関係を分析する分布交通モデルの推定を行って、誘発交通を明らかにする場合などで、生じる。

「標本化」の問題は、交通需要調査を含む社会調査に共通する課題であり、調査対象者自身が回答を行うか否かを判断することに起因する。たとえば、「高頻度でトリップを行う人」と「低頻度でトリップを行う人」の間で調査票回収率が異なっていれば、両者を含む母集団の標本の代表性は低くなる。また、標本抽出率が極端に低い地域が含まれていれば、地域間で標本の個人属性構成比が異なる可能性がある。いずれの場合も、従来法の拡大係数を使って交通手段や経路分担の影響要因を分析すると、価格弾力性の推定バイアスが大きくなる可能性がある。

以上のうち、「抽出」の問題は、**Choiced based sampling** によって得られる標本に加えて、居住地ベースの調査を実施して「ゼロトリップの人」の情報を収集すること無しでは、解決できない。一方、「標本化」の問題は、拡大係数の算出方法を改善することによって、解消することができる。

### 4. 傾向スコアを用いた拡大係数の設定

総流動情報などによって生活圏別にコントロールトータルが得られている場合は、生活圏別の標本回収率の逆数によって拡大係数が設定される。ただしこの方法を適用するためには、標本の個人属性構成比が母集団を代表していないわけではない。一方、土屋<sup>2)</sup>は、ランダムサンプリングを行うことが不可能で標本の母集団代表性が担保されない場合に、属性別のコントロールトータル情報を組み合わせたキャリブレーションによって拡大係数を算出する方法を提案した。以下に示す傾向スコア法は、適切な個人属性の観測確率に基づいて、拡大係数を求める手法である。

傾向スコアは、以下の手順で算出する<sup>2)</sup>。  $z_i$  を目的変数への影響が強く予想される個人属性の個票  $i$  の状態を表す指標（以下割り当て変数）、 $x_{i1}, \dots, x_{iK}$  を、それ以外の

$K$  種類の個人属性とすると、個票  $i$  が割り当て  $z_i$  を有する確率  $\theta_i$  は、式 (1) によって得られる。

$$\theta_i = P[z_i | x_{i1}, \dots, x_{iK}] \quad (1)$$

ここで確率  $\theta_i$  は傾向スコアと呼ばれる。本研究では、その算出を、式(2)に示すロジットモデルによって行う。

$$V_i = \sum_k \beta_k x_{ik} + Con \quad (2)$$

$V_i$  : 個票  $i$  が割り当て  $z_i$  に割り当たる確率効用の確定項、 $\beta_k$  : 個人属性  $k$  のパラメータ、 $x_{ik}$  : 個票  $i$  の個人属性  $k$  の説明変数、 $Con$  : 定数項

ここで、傾向スコアの推定値  $\hat{e}_i$  を、推定された確率効用の確定値  $\hat{V}_i$  より、式 (3) で算出する。

$$\hat{e}_i = \frac{\exp(\hat{V}_i)}{1 + \exp(\hat{V}_i)} \quad (3)$$

式(3)によって得られる傾向スコアは、観測された標本  $i$  が、特定の個人属性を有する確率（割り当て確率）を、その他の個人属性によって説明した値である。一般に、目的変数に影響する複数の個人属性は、相互に独立ではない。観測された標本  $i$  の個人属性組のうち、目的変数への影響度が強い属性の因果関係は正しく推定される一方で、目的変数への影響度が低い属性の因果関係を正しく推定することは難しい。このような場合に目的変数への影響度を正しく推定するための方法として、IPW (Inverse Probability Weighting) と呼ばれる、傾向スコアに基づく重み付け法が利用できる。なお、IPW が有効に働くためには、目的変数  $y$  と割り当て変数  $z$ 、およびその他の個人属性  $x$  の間に、式 (4) に示す「強く無視できる割り当て条件」が成立していなくてはならない。

$$P[y, z | x] = P[y | x] P[z | x] \quad (4)$$

このとき、IPW 推法による個票の拡大係数  $w_i$  は、式 (5) で表される。

$$w_i = \frac{N_m \cdot \frac{z_i}{\hat{e}_i}}{\sum_{i=1}^{N_k} \frac{z_i}{\hat{e}_i}} \cdot \omega_i, \text{ or } w_i = \frac{\bar{N}_f \cdot \frac{1-z_i}{1-\hat{e}_i}}{\sum_{i=1}^{\bar{N}_k} \frac{1-z_i}{1-\hat{e}_i}} \cdot \omega_i \quad (5)$$

ここで、 $N_m, \bar{N}_f$  は、それぞれ生活圏別の  $z_i = 1$ 、または  $z_i = 0$  となる標本数である。また  $\omega_i$  は、生活圏別の標本抽出確率の逆数である。幹線旅客純流動調査の生活圏別流動量は、総流動データに基づいて算出されているので、その値の信頼性は高いと考えられ、生活圏別の標本抽出確率の信頼性は高いと考えられる。式(5)から明らかなように、結局 IPW による重み付け法は、従来の抽出率の逆数を、傾向スコアによって補正した値となっている。

さらに、式(4)に示す「強く無視できる割り当て条件」が

成立するには、割り当て変数 $z$ として、利用可能な属性情報の中で、目的変数への影響が最も強いと考えられる変数を用いるべきことがわかる。すなわち、 $z$ として適切な個人属性を設定するには、実証的検討が必要である。

### 5. 傾向スコアモデルの推定結果

本研究では、割り当て変数 $z$ として性別を仮定する。なお交通機関別に性別構成比が異なることが予想されたため、傾向スコアを与えるモデルは、交通機関別に推定することとした。なおモデルを全 OD について計算するとサンプル数が膨大となるため、目的地が東京の場合について計算を行った。

表 1、表 2 に示すモデルを比較すると、代表交通機関が航空の方が、鉄道よりもモデルの適合度が高い。有意となった変数は、航空については年収と職業、および地域ダミー（近畿圏）であった。鉄道については、年齢と年収は有意となったが、それ以外の変数は有意とはならなかった。さらに図 7、8 に傾向スコアによって重み付けした標本による生活圏別男女構成比を示す。

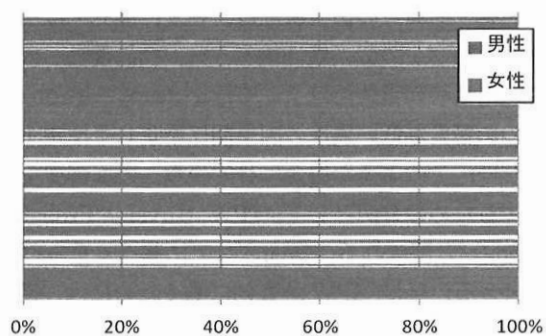


図 7 業務目的航空の傾向スコアによる重み付け

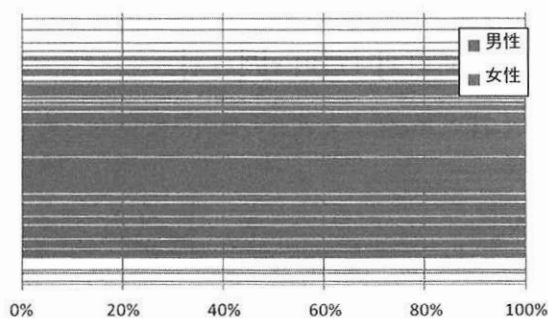


図 8 業務目的鉄道の傾向スコアによる重み付け

### 6. 結論

本稿では、傾向スコア法による標本拡大係数の設定手順についてまとめると共に、交通機関別に傾向スコアモデルを推定した。

今後は、推定した傾向スコアモデルを用いて、拡大係数を設定した場合の、個人属性構成比への影響を明らかにする必要がある。さらに、交通機関選択モデルを推定することによって、価格弾力性の推定値がどの程度変化するかを、明らかにする必要がある。

表 1 代表交通機関航空の傾向スコアモデル

| 説明変数         | 推定値       | t-値    |
|--------------|-----------|--------|
| 年齢           |           |        |
| 60代          | 0.199     | 1.396  |
| 年収           |           |        |
| 300～500万円未満  | -0.884    | ***    |
| 500～700万円未満  | 0.152     | 1.621  |
| 700～1000万円未満 | 0.635     | ***    |
| 職業           |           |        |
| 役職・管理職       | 2.231     | ***    |
| 会社員          | 1.158     | ***    |
| 公務員          | 1.555     | ***    |
| 農林・漁業        | 1.390     | *      |
| 商工・自営業       | 0.376     | *      |
| 居住地          |           |        |
| 北海道          | -0.133    | -1.215 |
| 東北圏          | -0.010    | -0.080 |
| 首都圏          | 0.147     | 1.652  |
| 北陸圏          | -0.133    | -0.771 |
| 中部圏          | 0.711     | 0.941  |
| 近畿圏          | 0.485     | ***    |
| 中国圏          | 0.114     | 0.898  |
| 四国圏          | -0.039    | -0.274 |
| 定数項          | 0.455     | ***    |
| サンプル数        | 9834      |        |
| 初期対数尤度       | -6816.410 |        |
| 最終対数尤度       | -3357.955 |        |
| 自由度調整済み尤度比   | 0.505     |        |

\*, \*\*, \*\*\*: 有意水準 10%, 5%, 1%

表 2 代表交通機関鉄道の傾向スコアモデル

| 説明変数         | 推定値       | t-値    |
|--------------|-----------|--------|
| 年齢           |           |        |
| 60代以上        | 0.364     | *      |
| 年収           |           |        |
| 300～500万円未満  | -0.757    | ***    |
| 500～700万円未満  | 0.236     | *      |
| 700～1000万円未満 | 1.076     | ***    |
| 職業           |           |        |
| 役職・管理職       | 10.787    | 0.054  |
| 会社員          | 0.078     | 0.099  |
| 公務員          | 8.267     | 0.059  |
| 居住地          |           |        |
| 北海道          | 9.380     | 0.114  |
| 首都圏          | 0.071     | 0.414  |
| 北陸圏          | -0.036    | -0.114 |
| 中部圏          | -0.181    | -0.962 |
| 近畿圏          | 0.012     | 0.064  |
| 中国圏          | 0.054     | 0.241  |
| 四国圏          | 0.267     | 0.460  |
| 定数項          | 1.586     | ***    |
| サンプル数        | 4332      |        |
| 初期対数尤度       | -3002.714 |        |
| 最終対数尤度       | -1748.212 |        |
| 自由度調整済み尤度比   | 0.413     |        |

\*, \*\*, \*\*\*: 有意水準 5%, 1%

### 参 考 文 献

- 1) 土屋隆裕:「調査への指向性」変数を用いた調査不能バイアス補正の試み—「日本人の国民性調査」データへの適用—, 日本統計学会誌, Vol.36, No.1, pp.1-23, 2006
- 2) Iannaccione, V.G.: Sequential weight adjustment for location and cooperation propensity for the 1995 national survey of family growth, Journal of Official Statistics, Vol.19, pp.31-43, 2003