

# Web調査に基づく都市間トリップ発生量の推定

塚井 誠人<sup>1</sup>・片川儀治<sup>2</sup>・北原聡子<sup>2</sup>

<sup>1</sup>正会員 広島大学准教授 大学院工学研究院 (〒739-8527 東広島市鏡山1-4-1)  
E-mail: mtukai@hiroshima-u.ac.jp

<sup>2</sup>学生員 広島大学工学部 大学院工学研究科 (〒739-8527 東広島市鏡山1-4-1)  
E-mail: m126011@hiroshima-u.ac.jp

幹線旅客純流動調査は、実際に都市間トリップを行っている旅客を対象に行われる。しかし純流動調査では、鉄道では十分な回収率が得られない、調査間隔が5年に一度しか確保できない、トリップを行わない潜在的な旅客の特性を把握できない、などの限界を抱えている。本研究では、自宅ベースで都市間トリップ発生量に関する情報を得る方法としてWeb調査に着目し、トリップが発生していない潜在旅客についても一定のサンプルを得る調査法について検討する。具体的には、回顧式の調査によって一定期間内の都市間トリップに関する情報を収集し、これを用いた統計モデルの推計を行う。発生頻度モデルと発生間隔モデルをそれぞれ推定し、両者の精度を比較する。その上で、幹線旅客純流動調査の補完調査としてのWeb調査の可能性を検討する。

**Key Words :** *inter-regional passenger demand, trip generation, propensity score*

## 1. はじめに

全国幹線旅客純流動調査（以下、純流動調査）は、移動中の旅行者に調査票を配布・回収する形式をとっている。しかし従来の調査方法は、1) 鉄道では十分な回収率が得られない、2) 調査間隔が5年に一度しか確保できない、3) トリップを行わない潜在的な旅客の特性を把握できない、という課題を抱えている。<sup>1)</sup>

本研究では、既存の調査手法を補う手法としてWeb調査に着目して、都市間トリップの発生頻度および間隔のモデル化を図る。なおWeb調査では、回答時間帯によるWebを閲覧者層の相違によって、最も重要な調査項目であるトリップ発生間隔（または一定期間内のトリップ発生回数）が異なることが懸念される。そこで回答者の回答時間帯による影響を除いて個人属性がトリップ発生回数に及ぼす純粋な影響を推定し、正しいトリップ発生間隔を推定するための統計手法として、標本の偏りを補正する傾向スコアの適用を検討する。

## 2. 傾向スコア法

自然科学研究は、標本特性（共変量）と条件変数（独立変数）のランダムな割り当てが可能な「実験」と、それが不可能な「調査観察」に大別できる。純流動調査

のようにトリップ中の標本を抽出する場合や、Webによる標本抽出では、標本や選択肢の特性が選択結果に及ぼす影響を因果効果として推定する際に、変数間の交絡が懸念される。そこでRosenbaumらは無作為割り当てが不可能な調査観察において、目的変数への因果効果を効率的に推定する方法として、傾向スコアを提案した。

<sup>2)</sup> $x_i$ を回答者*i*の共変量、 $z_i$ を独立変数とすると、傾向スコア $e_i$ は式(1)で表される。

$$e_i = P(z_i = 1 | x_i) \quad (1)$$

本研究において傾向スコアの目的変数は、回答時間帯とする（回答時間帯が早い場合： $z_i=1$ 、遅い場合： $z_i=0$ ）

（図-1）。したがって、得られる傾向スコアは、回答時間帯が早い群 $z_i=1$ への帰属確率であり、その値はロジスティックモデルによって推定する。その上で、傾向スコアの逆数によってサンプルの重み付けを行うIPW法を用いて、トリップ発生頻度モデルおよびトリップ発生間隔モデルをそれぞれ推定し、傾向スコアの導入効果を検

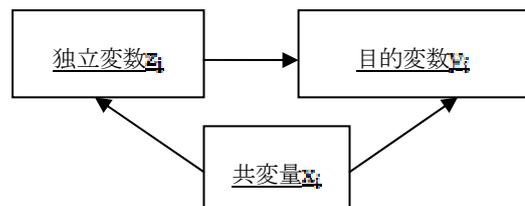


図-1 独立変数・目的変数・共変量の関係

証する。

### 3. 調査概要

本研究では、Web 調査により「旅行に関する調査」を 2012 年 10 月 29 日 19:00～2012 年 10 月 31 日 13:00 にわたって実施した。今回の調査において得られた標本数は 5182 であった。

調査の質問項目を表-1 に示す。調査対象者は 18 歳以上の男女、対象移住地は全国 47 都道府県である。質問項目は、大きく分けて 2 部から構成される。第 1 部は性別や年齢などの個人属性と利用情報端末の種類・IC カード・マイレージ等の各交通機関が行うサービスの利用状況であり、第 3 部は過去 1 年間に行った都市間トリップの回数、そのうち直近の 3 回分の旅行に関する詳細な質問である。

表-1 調査の質問項目

個人属性	性別、年齢、居住地、職業、年収、 世帯構成 携帯電話、ICカード、 航空会社のマイレージサービス、 鉄道会社の新幹線予約サービスの利用状況
通勤・通学 以外の 旅行・出張 について	過去1年に行った都市間トリップの回数、 過去3回の旅行時期、日程、目的、 出発地と目的地、利用交通機関、 交通機関の割引サービス、費用、 費用の支払い元、旅行先への再訪問頻度

### 4. 分析手法

トリップ発生頻度は、一定期間内の稀事象の発生確率を表すポアソンモデルを拡張したゼロ過剰型の負の二項回帰モデルを用いて推定する。在宅ベースで調査する Web 調査によって都市間トリップの回数のような稀少な発生事象を調査すると、頻繁に都市間トリップを行う者と普段はあまり都市間トリップをしない人が標本中に混在する。その際に、トリップ発生回数が 0 となる者が多く含まれるため、ゼロ度数の標本がポアソン分布の期待値と合致せず、モデル適合度が低下する可能性がある。ゼロ過剰型の負の二項回帰モデルは、事象の生起回数（カウントデータ）を表すモデルとしてポアソンモデルと並んで広く用いられており<sup>3)</sup>、式(2)、(3)で表される。

$$P(y_i = 0) = \psi + (1 - \psi) \left( \frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \quad (2)$$

$$P(y_i > 0) = (1 - \psi) \frac{\Gamma(y_i + \nu^{-1})}{y_i! \Gamma(\nu^{-1})} \left( \frac{\nu^{-1}}{\nu^{-1} + \mu_i} \right)^{\nu^{-1}} \left( \frac{\mu_i}{\nu^{-1} + \mu_i} \right)^{y_i} \quad (3)$$

$y_i$  は事象の生起度数を表す確率変数、 $\nu$  は正のパラメータ、 $\mu_i$  は期待値であり、標本の個人属性に依存する。

トリップ発生間隔は、トリップ目的別に生存時間モデルを推定する。なおトリップ発生間隔は、生存時間解析におけるセンシングを参照する。すなわち、生存関数  $S(T_i)$  の累積分布関数を  $F(T_i)$ 、確率密度関数を  $f(T_i)$  とすると、トリップ発生間隔が  $T_i^1$  以下の確率を  $P_1$ 、 $T_i^2$  の確率を  $P_2$ 、および  $T_i^3$  以上の確率を  $P_3$  として、 $P_2 = f(T_i^2)$ 、 $P_1 = F(T_i^1)$ 、 $P_3 = S(T_i^3)$  として、尤度関数を定義する。

本研究では、分布関数として、共変量によって基準分布としてワイブル分布を仮定し、トリップ発生間隔が共変量によって変化する、加速故障モデル：式(4)を用いてモデル推計する。

$$S(T_i | X_{ij}) = S_0 \left( T_i \exp \left( - \sum_j \beta_j X_{ij} \right) \right) \quad (4)$$

ここで、 $X_{ij}$  は回答者  $i$  の属性  $j$  の共変量ベクトル、 $S_0(T_i)$  は基準の生存関数、 $\beta_j$  は未知パラメータベクトルである。

### 5. 基礎集計

調査開始から終了までの累積観測数の推移を図 2 に、時間帯別の観測数を図 3 に、それぞれ示す。両図から明らかかなように、調査 2 日目の 11:00 以降の標本数が急激に増加しているが、これは調査会社が調査開始後の回答数が伸びないことを懸念して、10:00 に回答を呼びかける再配信を行っているためである。

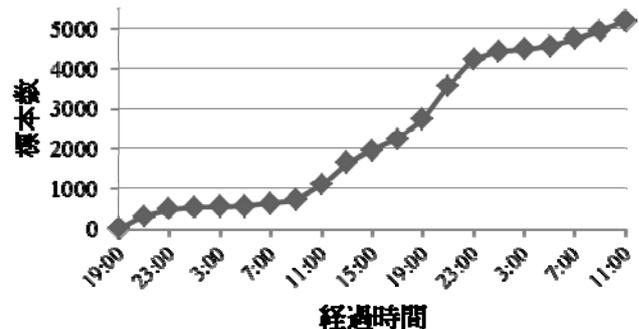


図 2 時間帯別累積観測数

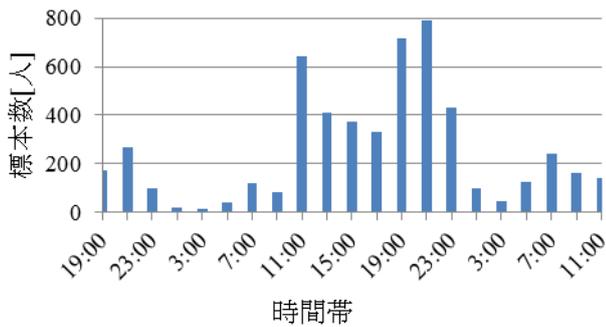


図-3 時間帯別標本数

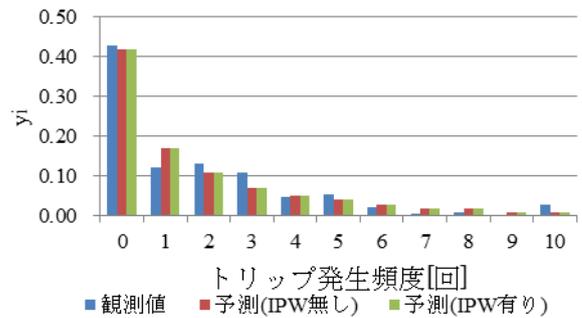


図-3 トリップ発生頻度の分布

表-2 トリップ発生頻度モデル推定結果

説明変数	傾向スコアなし		傾向スコア有り	
	推定値		推定値	
$\psi$	0.06	.	0.07	*
$v$	1.62	***	1.62	***
男性	-0.11	.	-0.13	*
年齢(30-39歳)	-0.10	.	-0.10	.
年齢(40-49歳)	-0.14	*	-0.13	*
年齢(50-59歳)	-0.05	.	-0.07	.
年齢(60歳以上)	0.14	.	0.15	*
役員・管理職	0.35	**	0.38	**
会社員	0.15	.	0.14	.
公務員	0.32	***	0.32	***
農林・漁業	0.33	.	0.31	.
製造業	-0.36	*	-0.36	*
その他の業種	-0.11	.	-0.11	.
学生(大学・大学院)	0.26	*	0.26	*
無職	-0.19	*	-0.20	*
300~700万円	0.40	***	0.40	***
700~1500万円	0.97	***	0.96	***
1500万円以上	1.11	***	1.10	***
年収無回答	-0.04	.	-0.05	.
世帯主ではない	-0.11	.	-0.11	.
回答者と配偶者のみ	0.28	***	0.27	***
回答者と配偶者と子供	0.06	.	0.07	.
あなたと親	-0.12	.	-0.12	.
その他	0.09	.	0.09	.
定数項	0.67	***	0.68	***
サンプル数	5182		5182	
最終対数尤度	-10181.48		-10186.24	
AIC	20410.96		20420.49	

\*\*\*0.1%有意, \*\*1%有意, \*5%有意, .10%有意

## 5. 都市間トリップ発生量の推定

### 1) トリップ発生頻度モデル

回答時間帯に関する傾向スコアモデルを推定したところ、尤度比は0.21程度となり、比較的当てはまりの良いモデルが得られた。なおモデルの推定結果は、紙面の都合上省略する。表-2にトリップ発生頻度モデルの推定結

果を示し、傾向スコアなしのモデルと、IPW推定量により傾向スコアを導入したモデルを比較する。推定結果において有意になったパラメータより、トリップ発生頻度が高くなる旅行者は、役員または公務員、高年収、および世帯構成が回答者と配偶者の回答者であった。なお、男性および60歳以上のパラメータが有意となったことを除いて、IPWを導入してパラメータ推定を行っても、推定値はほぼ変わらなかった。

図-2に観測されたデータから算出したトリップ発生頻度の期待値とゼロ過剰型の負の二項回帰モデルにより推定したトリップ発生頻度の期待値、さらにIPW推定量を導入した負の二項回帰モデルにより推定したトリップ発生頻度の期待値を示す。IPW推定量を導入するとパラメータの精度が少し向上したが、トリップ頻度の発生確率はIPW未導入モデルとほぼ同じ分布となり、明確な導入効果は認められなかった。

### 2) トリップ発生間隔モデル

トリップ発生間隔モデルのパラメータは、正であれば間隔が長くなり、負であればそれが短くなることを表す。

表-3に業務目的のトリップ発生間隔モデルの推定結果を示す。有意となったパラメータは男性のダミー変数であったが、IPW推定量によって重み付けしたモデルにおいて、パラメータの精度は向上しなかった。表-4に観光目的のトリップ発生間隔モデルの推定結果を示す。年齢60歳以上、マイレージ利用、新幹線予約サービス利用のダミー変数が有意になった。また、IPW推定後は新幹線予約サービス利用ダミー変数の精度が向上した。表-5に帰省私用目的のトリップ発生間隔モデルの推定結果を示す。年齢50代、年収不明のパラメータが有意になった。また、IPW推定量を導入することによってモデルの当てはまりを表すAICが改善された。

表-3 トリップ発生間隔推定結果（業務）

説明変数	傾向スコアあり	傾向スコア無し
	推定値	推定値
男性	-0.66 **	-0.63 **
役員・管理職	0.00	0.00
会社員	0.01	0.00
公務員	-0.14	-0.16
農林・漁業	0.22	0.16
製造業	0.26	0.35
その他業種	-0.07	-0.07
学生（大学・大学院）	0.47	0.56
年収300～700万	0.09	0.09
年収700～1500万	-0.02	-0.02
年収1500万以上	0.04	0.06
年収不明など	0.23	0.28
ICカード利用有無	-0.03	-0.03
航空MC利用有無	-0.20	-0.20
新幹線予約利用有無	-0.11	-0.14
$\lambda$	8.70 ***	8.56 ***
$\gamma$	0.68 ***	0.68 ***
サンプル数	1250	1250
最終対数尤度	-940.79	-943.11
AIC	1915.58	1920.22

\*\* : 5%有意, \*\*\* : 1%有意

表-4 トリップ発生間隔推定結果（観光）

説明変数	傾向スコアあり	傾向スコア無し
	推定値	推定値
男性	-0.06	-0.05
30～39歳	0.02	0.01
40～49歳	-0.08	-0.11
50～59歳	-0.03	-0.03
60歳以上	-0.26 ***	-0.29 ***
年収300～700万	-0.06	-0.06
年収700～1500万	0.03	-0.01
年収1500万以上	-0.36	-0.39
年収不明など	0.01	-0.01
ICカード利用有無	-0.04	-0.06
航空MC利用有無	-0.26 ***	-0.25 ***
新幹線予約利用有無	0.20 *	0.21 **
$\lambda$	3.80 ***	3.90 ***
$\gamma$	1.03 ***	1.03 ***
サンプル数	3823	3823
最終対数尤度	-2941.36	-3026.62
AIC	5910.73	6081.23

\* : 10%有意, \*\* : 5%有意, \*\*\* : 1%有意

## 6. おわりに

本研究では、Web 調査を利用した都市間トリップ発生頻度と間隔のモデルを推定し、回答時間帯の影響を除去する傾向スコアの現況再現性の検証を行った。

基礎集計からは、回答時間帯によって得られる標本数、個人属性の構成比、および一人あたりの平均トリップ回数が異なることが確認された。しかし、回答時間帯の違いが個人属性構成比やトリップ発生回数に明確に影響す

表-5 トリップ発生間隔推定結果（帰省・私用）

説明変数	傾向スコアあり	傾向スコア無し
	推定値	推定値
男性	-0.01	-0.02
30～39歳	-0.3	-0.28
40～49歳	-0.1	-0.11
50～59歳	-0.35 *	-0.36 *
60歳以上	-0.17	-0.18
年収300～700万	-0.06	-0.08
年収700～1500万	-0.08	-0.09
年収不明など	-0.39	-0.41
ICカード利用有無	-0.1	-0.08
航空MC利用有無	-0.02	-0.03
新幹線予約利用有無	-0.18	-0.21
$\lambda$	5.78 ***	5.78 ***
$\gamma$	1.35 ***	1.37 ***
サンプル数	748	748
最終対数尤度	-491.08	-488.28
AIC	1008.16	1002.55

\* : 10%有意, \*\*\* : 1%有意

るという結果はみられなかった。

傾向スコアモデルの適合度は比較的高い結果が得られたが、発生頻度モデルに IPW 推定量を導入しても十分な補正効果を確認できなかった。なおモデルの再現性に関しては、観測値が 0 付近の期待値を再現することができたが、1 以上の観測値に現れる期待値のピークは再現できなかった。これは、トリップ数が 10 回以上の標本がもたらす大きな分散に影響されたためと考えられる。

トリップ発生間隔モデルに関しては、帰省・私用目的においては、IPW 推定量の導入によるパラメータ有意水準の向上が確認された。

本研究で実施した Web 調査では、回収された標本の男女年齢別構成比が、国勢調査の構成比と合致しなかった。今後は国勢調査の構成比に割りつけるように有効標本数を推定する方法か、または母集団の男女構成に合わせて逆重みをかけることによって、母集団代表性を確保できると考えられる。また、これにより傾向スコアの割り当てによる影響がより強く現れる可能性がある。

## 参考文献

- 1) 奥村誠：全国幹線旅客純流動調査の意義と課題，運輸と経済，Vol72，No.6，pp.21-30，2012
- 2) Rosenbaum, P.R. & Rubin, D.B. : The central role of the propensity score in observational studies for casual effects, 70, 41-55, 1983
- 3) 爲季和樹，堤盛人：固有ベクトル空間フィルタリングを用いたゼロ過剰重力モデル，土木計画学研究・講演集，Vol.45，2012