

# Web 調査を用いたトリップ発生間隔に関する研究

広島大学大学院工学研究科  
 広島大学大学院工学研究院

学生員 ○北原聡子  
 正会員 塚井誠人

## 1. はじめに

全国幹線旅客純流動調査（以下、純流動調査）は、全国一斉（秋季1日）に実際に都市間トリップを行っている人の移動中に調査票を配布・回収する形式をとっているが、特に鉄道では十分な回収率が得られないという課題を抱えている<sup>1)</sup>。

本研究では、全国幹線旅客純流動調査の課題を補足する方法として、Web 調査を利用した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。Web 調査では、回答時間帯により Web を閲覧している回答者層が異なるので、最も重要な調査項目であるトリップ発生回数が異なることが懸念される。回答者の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するための統計手法として、本研究では標本の偏りを補正する傾向スコアを適用する。

## 2. 傾向スコア法

研究デザインは、標本特性（共変量）と実験条件の変数（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。幹線旅客純流動調査をはじめとする多くのトリップ/入込旅客調査は後者に当たり、因果効果を推定する際に説明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研究において、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。

$$e_i = p(z_i = 1 | x_i) \quad (0 \leq e_i \leq 1) \quad (1)$$

本研究では、傾向スコアを推定するときの目的変数は、回答時間帯から作成し（回答時間帯がある時間帯の場合： $z_i=1$ 、それ以外の場合： $z_i=0$ ）（図-1）、傾向スコアは、回答時間帯が特定の時間帯である群  $z_i=1$  への帰属確率である。

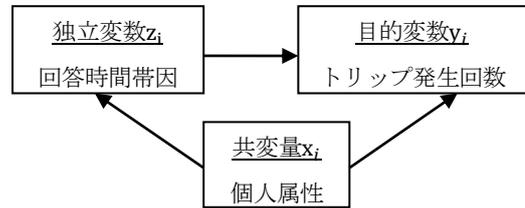


図-1 独立変数・目的変数・共変量の関係

傾向スコア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3部から構成される。第1部は性別や年齢などの個人属性、第2部は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、および第3部は過去1年間に往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. 分析方法

トリップ発生頻度は、ゼロ過剰型の負の二項回帰モデルを用いて推定する。負の二項回帰モデルは、ある一定期間内において稀な事象の生起回数（カウントデータ）を表すモデルとして広く用いられる、ポアソンモデルを拡張したモデルである<sup>3)</sup>。本研究では、調査期間内に都市間トリップの回数を調査する場合に、普段は頻繁に都市間トリップをしているが、たまたま行わなかった者と、普段から都市間トリップをしない人が混在することにより、ポアソン分布で期待される値よりゼロ度数が多くなる可能性がある。このようなカウントデータを扱う際に、ゼロ過剰型のモデルが用いられ、式(2a)、(2b)で表される。

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

$$\Pr(y_i > 0) = (1 - \psi) \frac{\Gamma(y_i + v^{-1})}{y_i! \Gamma(v^{-1})} \left( \frac{v^{-1}}{v^{-1} + \mu_i} \right)^{v^{-1}} \left( \frac{\mu_i}{v^{-1} + \mu_i} \right)^{y_i} \quad (2b)$$

$y_i$  は事象の生起度数を表す確率変数、 $v$  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、目的別に生存時間モデルを用いて推定する。調査データから得られる2種類のトリップ発生間隔は、生存時間解析におけるセンサリング（中途打ち切り）データの扱い方を参照する。本研究では、分布関数として、共変量によって基準分布からトリップ発生間隔が増減すると仮定するワイブル分布の加速故障モデルを適用し、式(3)と表される。

$$S(T_i | X_{ij}) = S_0(T_i \exp(-\sum_{j=1}^J \beta_j X_{ij})) \quad (3)$$

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。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。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、 $P_2 = f(T_i^2)$ 、 $P_1$ 、 $P_3$  はそれぞれ  $F(T_i^1)$ 、 $S(T_i^3)$  で表される。

#### 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。これらより、各時間帯によって回答者数が異なっていることがわかる。また、調査会社が調査開始後に回答数が伸

びないことを懸念して、調査2日目の10:00に回答を呼びかける再配信を行なっているため、11:00以降の標本数が急激に増加している。

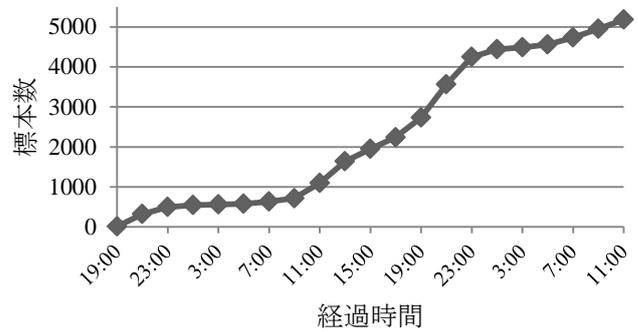


図-2 時間帯別累積観測数

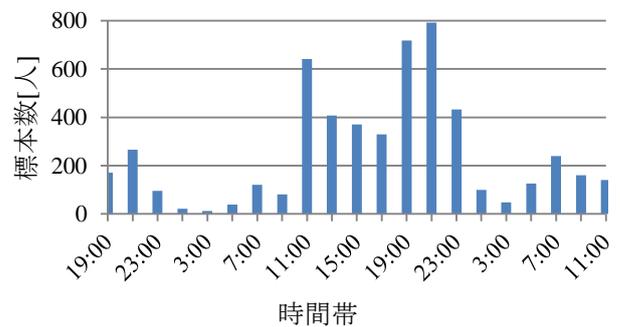


図-3 時間帯別標本数

#### 5. 分析結果

##### 1) トリップ発生頻度モデル

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。有意になったパラメータに着目すると、役員または公務員であり、年収が高く、世帯構成が回答者と配偶者の場合に、トリップ発生頻度が高くなる傾向がみられる。なお、男性および60歳以上のパラメータが有位となったことを除いて、IPW 推定量の導入前後で推定値の値はほぼ変わらなかった。

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		傾向スコアあり	
	推定値	t値	推定値	t値
$\psi$	0.06	1.88	0.07	2.00 *
$\nu$	1.62	12.55 ***	1.62	12.57 ***
男性	-0.11	-1.89	-0.13	-2.08 *
年齢(30-39歳)ダミー	-0.10	-1.58	-0.10	-1.62
年齢(40-49歳)ダミー	-0.14	-2.37 *	-0.13	-2.19 *
年齢(50-59歳)ダミー	-0.05	-0.76	-0.07	-1.01
年齢(60歳以上)ダミー	0.14	1.89	0.15	2.00 **
役員・管理職	0.35	3.02 ***	0.38	3.20 ***
会社員	0.15	1.21	0.14	1.14
公務員	0.32	4.08 ***	0.32	4.15 ***
農林・漁業	0.33	1.02	0.31	0.95
製造業	-0.36	-2.31 *	-0.36	-2.30 *
その他の業種	-0.11	-1.16	-0.11	-1.15
学生(4年制大学・大学院)	0.26	2.02 *	0.26	2.01 *
無職	-0.19	-2.32 **	-0.20	-2.40 **
300~700万円	0.40	6.50 ***	0.40	6.39 ***
700~1500万円	0.97	10.54 ***	0.96	10.50 ***
1500万円以上	1.11	5.29 ***	1.10	5.15 ***
わからない・答えたくない	-0.04	-0.51	-0.05	-0.59
ICカード利用ダミー(有=1,無し=0)	-0.11	-1.57	-0.11	-1.48
マイレージ利用ダミー(有=1,無し=0)	0.28	3.49 ***	0.27	3.34 ***
あなたと配偶者のみ	0.06	0.90	0.07	0.97
あなたと親	-0.12	-1.29	-0.12	-1.25
その他	0.09	0.91	0.09	0.89
定数項	0.67	6.51 ***	0.68	6.57 ***
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

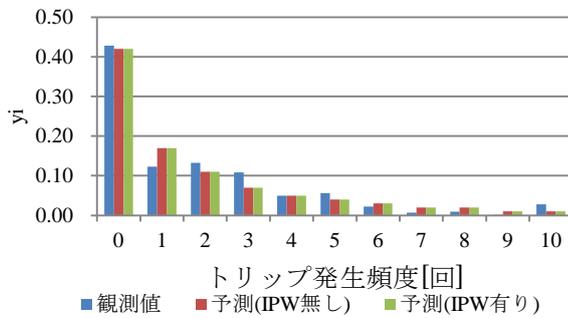


図-3 トリップ発生頻度の分布

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, 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それぞれ示す. 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未導入モデルの分布とほぼ同じ値をとっており, 同図による識別は困難である.

2) トリップ発生間隔モデル

本モデルの, パラメータは正であればトリップ発生間隔が長くなり, 負であればトリップ発生間隔が短くなること表す. 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. なお, 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-3 トリップ発生間隔推定結果 (業務)

説明変数	傾向スコアなし		傾向スコアあり	
	推定値	t値	推定値	t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	
その他の業種	-0.07		-0.07	
学生(4年制大学・大学院)	0.47		0.56	
300~700万円	0.09		0.09	
700~1500万円	-0.02		-0.02	
1500万円以上	0.04		0.06	
わからない・答えたくない	0.23		0.28	
ICカード利用ダミー(有=1,無し=0)	-0.03		-0.03	
マイレージ利用ダミー(有=1,無し=0)	-0.20		-0.20	
新幹線予約サービス利用ダミー(有=1,無し=0)	-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	

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

表-4 トリップ発生間隔推定結果 (観光)

説明変数	傾向スコアなし		傾向スコアあり	
	推定値	t値	推定値	t値
男性	-0.06		-0.05	
年齢(30-39歳)ダミー	0.02		0.01	
年齢(40-49歳)ダミー	-0.08		-0.11	
年齢(50-59歳)ダミー	-0.03		-0.03	
年齢(60歳以上)ダミー	-0.26	***	-0.29	***
300~700万円	-0.06		-0.06	
700~1500万円	0.03		-0.01	
1500万円以上	-0.36		-0.39	
わからない・答えたくない	0.01		-0.01	
ICカード利用ダミー(有=1,無し=0)	-0.04		-0.06	
マイレージ利用ダミー(有=1,無し=0)	-0.26	***	-0.25	***
新幹線予約サービス利用ダミー(有=1,無し=0)	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	

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

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 量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あてはまりを表す AIC が改善された.

表-4 トリップ発生間隔推定結果（帰省私用）

説明変数	傾向スコアなし	傾向スコアあり
	推定値	推定値
男性	-0.01	-0.02
年齢(30-39歳)ダミー	-0.30	-0.28
年齢(40-49歳)ダミー	-0.10	-0.11
年齢(50-59歳)ダミー	-0.35 *	-0.36 *
年齢(60歳以上)ダミー	-0.17	-0.18
300～700万円	-0.06	-0.08
700～1500万円	-0.08	-0.09
わからない・答えたくない	-0.39 *	-0.41 *
ICカード利用ダミー(有=1,無し=0)	-0.10	-0.08
マイレージ利用ダミー(有=1,無し=0)	-0.02	-0.03
新幹線予約サービス利用ダミー(有=1,無し=0)	-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

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

#### 4. おわりに

##### 1) 研究成果

本研究では、更新期間が長く、トリップ発生についての分析が困難な全国幹線旅客純流動調査の課題への対応のため、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。以下に分析結果についてまとめる。

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が、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るという結果はみられなかった。これはアンケート受付期間中に行われた、回答を促すメールの再配信の影響が強くあらわれているためと考えられる。

傾向スコア推定については、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。これは、集計レベルで確認されたように、回答時間帯別の群のトリップ発生回数に、個人レベルでも顕著な差異がなかったためと考えられる。

トリップ発生頻度モデルに関しては、ゼロ過剰型の負の二項回帰モデルを用いたため、観測値が 0 付近の旅行者について、トリップ発生頻度の期待値を再現することができた。しかし、観測値が 1 以上の位置に現れる期待値のピークを再現することはできなかった。これは、負の二項分布の性質として、複数の峰を持つ分布の推定が難しいことや、10 回以上のトリップを行っている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、生存時間解析におけるセンサリングデータの取り扱い方を参照することで、目的別のトリップ発生間隔を推定した。その結果、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。

##### 2) 今後の課題

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。これは、都市間トリップについて調査するという、本調査内容に対して一部の回答者の興味が集中したためと考えられる。この問題に関しては、得られた標本のうち、予定より多く集まった層の標本を、最も回収率が低い層の標本数に合わせて国勢調査の構成比に割りつけて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り加重標本とすることによって、母集団代表性を確保できると推測される。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。

トリップ発生に関しては、一定期間のトリップ発生頻度とトリップ発生間隔に着目してモデル化を行ったが、頻度と間隔が逆数の関係にあることを利用した、モデル推計精度の検証は行わなかった。今後は目的別のトリップ発生間隔からトリップ発生頻度を推定することにより、信頼性の高いモデルを推定できる可能性がある。

ゼロ過剰型の負の二項分布を適用したトリップ発生頻度の分析について、観測値の分布が 0 付近と 2 付近の 2 箇所の極大値から構成されていることを考慮すると、混合分布型による尤度関数の特定化を検討する必要がある。さらに本分析は年間のトリップ発生データを用いているが、より詳細な目的別データを用いた分析が必要である。

##### 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