

# スマートカードデータを用いた 公共交通の施策実施評価に関する一考察

宮崎 耕輔<sup>1</sup>・柳原 奨<sup>2</sup>・桑野 将司<sup>3</sup>・伊藤 昌毅<sup>4</sup>・谷本 圭志<sup>5</sup>

<sup>1</sup>正会員 香川高等専門学校 准教授 建設環境工学科 (〒761-8058 香川県高松市勅使町 355)  
E-mail: miyazaki@t.kagawa-nct.ac.jp

<sup>2</sup>学生会員 香川高等専門学校 創造工学専攻 (〒761-8058 香川県高松市勅使町 355)

<sup>3</sup>正会員 鳥取大学大学院 准教授 (〒680-8552 鳥取県鳥取市湖山町南 4 丁目 101)

<sup>4</sup>正会員 東京大学生産技術研究所 助教 (〒153-8505 東京都目黒区駒場 4-6-1)

<sup>5</sup>正会員 鳥取大学大学院 教授 (〒680-8552 鳥取県鳥取市湖山町南 4 丁目 101)

本研究は、スマートカードの利用履歴データを用いて、統計的因果推論を援用することによって、公共交通の施策実施評価を行った。具体的には、香川県高松市における高松琴平電気鉄道の鉄道とバスとの乗継ぎ割引拡大の施策実施を対象として、乗継ぎ割引の拡大が幹線系公共交通手段である鉄道の利用回数の増加に寄与したかという観点に着目して分析を行った。

その結果、利用券種別に効果は異なるものの乗継ぎ割引の拡大によって、鉄道の利用回数が増加したという結果が得られた。なお、本研究は統計的因果推論の手法を援用した一試算に過ぎないため、異なる手法を用いた検討結果を比較するなどしつつ更なる検証を行うことが研究課題として残った。

**Key Words:** public transport, smartcard data, transfer, multimodal public transport trips

## 1. はじめに

地方都市においては、利便性の高い自家用車の普及や人口減少などによって、地域公共交通システムが衰退している地域が少なくない。しかしながら、高齢社会の進展については、地方都市の方が深刻な社会問題となっており、地域公共交通を活用した高齢者のモビリティの確保などが課題となっている。そのため、地方都市における地域公共交通システムの確保はもちろん、どのように維持していくかについては、非常に重要な社会問題となっている。

さて、本研究でとりあげる地方都市である香川県高松市においては、地域公共交通である鉄道やバスに対して、利便性向上のための施策を数多く実施しており、最近では利用者数の減少に歯止めがかかっていると言われている。本稿では、2014年3月1日に実施した「乗継ぎ割引の拡大」に焦点をあてる。この「乗継ぎ割引の拡大」とは、高松琴平電気鉄道の鉄道とバスとの乗継ぎにおいて、従来の20円の乗継ぎ割引を100円割引に拡大した施策である。この施策の効果検証に着目し、統計的因果推論の手法を援用することによって、実証分析するものである。本研究では、鉄道とバスの「乗継ぎ割引の拡大」によっ

て、幹線系公共交通手段である鉄道の利用に焦点をあてる。

## 2. 本研究で用いる統計的因果推論

### (1) 統計的因果推論の概要

自然科学の検証対象とされる研究仮説は、原因Aが結果Bに影響するという因果効果として表現される場合がある。研究仮説は原因Aを設定した「処置群」とそれを設定しない「対照群」の標本を収集する実験を行い、原因Aによる結果Bへの影響を観察または統計的に検定することによって検証される。ただし、一連の手順において、検証の対象としない特性Nは、両群で同一でなくてはならない。しかしながら、土木計画学における研究対象となる観察研究については、Nを完全に統制することができない。このようなデータは調査観察データと呼ばれ、統計アプローチによる相関研究の中でも、特別な注意が必要なデータといえる。

一般に観察研究によって、独立変数の従属変数(結果変数)に対する影響を調べる際には、従属変数に影響を与える共変量(剰余変数、または交絡変数、交絡要因と

もいう)の分布が独立変数の値によって異なる(=交絡する)可能性がある<sup>1)</sup>といわれている。共変量の影響を除去する方法としては、様々な手法が挙げられるが、Rosenbaum & Rubin<sup>2)</sup>が提案した傾向スコア (Propensity score) を利用した共変量調整法が応用研究に利用されてきている<sup>1)</sup>。このモデルは、観察データから因果関係を推定するための手法として、観察データに隠れた影響を与える共変量を求め、この共変量によって及ぼされる影響を調整した因果効果を推定することができるものである。

傾向スコアは、標本の特性に関する無作為割当てが不可能な相関研究において、結果変数に対する因果効果を推定する方法である。因果効果とは、条件統制ができない調査観察実験のいずれにおいても、ある処置を受けた処置群とその処置を受けなかった対照群との結果の差、すなわち処置効果である。処置群と対照群の結果をそれぞれ、 $y_1$ ,  $y_0$  とすると、推定される平均処置効果  $T$  は、母集団における期待値を用いて、式(1)のように表される。

$$T = E(y_1) - E(y_0) \quad (1)$$

調査観察実験では、処置群と対照群において、統制できない回答者属性等の分布は、特に小標本では同一とみなせないため、式(1)を用いて  $T$  を求めることは難しい。

そこで、バランシングスコアを用いた比較法が提案されている。バランシングスコア  $b(x)$  は、共変量ベクトル  $x$  の条件付き分布が処置群と対照群に対して等しくなるような関数である。すなわち、関数  $b(x)$  は、処置の有無を表す指標 (割当て変数)  $z$  と共変量ベクトル  $x$  の直交性条件として、式(2)のように表される。

$$x \perp z(x) \quad (2)$$

傾向スコアは、観測された共変量  $x$  の分布が、処置群に割当てられる傾向を示す条件付き確率であり、最も粗いバランシングスコアと呼ばれる。傾向スコア  $e(x)$  は、式(3)のように表せる。

$$e(x) = p(z = 1|x) \quad (3)$$

さて、統計的因果推論を用いる最大の目的は、処置の有無による効果の差を定量的に表現することである。これを因果効果あるいは、平均処置効果 (Average Treatment Effect; ATE) と呼んでいる<sup>3)</sup>。

傾向スコアを用いた解析手法は、潜在的な交絡要因となる様々な共変量をロジスティック回帰分析やプロビット回帰分析などで推計した予測確率を傾向スコアという一つの合成変数に一次元化し、その傾向スコアを基準として、交絡因子の影響を除去するものである。すべて二段階推定法であり、以下の2つのステップを踏む必要がある<sup>1)</sup>。

#### 1) 傾向スコアの推定

割付変数を共変量によって説明するモデルを設定し、そのモデルの母数の推定を行う。母数の推定値を用いて、各被験者に対して「条件1に割付けられる予測確率」を計算し、これを傾向スコアの推定値とする。

#### 2) 推定された傾向スコアを用いた調整

上記で推定された傾向スコアを用いて、具体的な調整を行う。

傾向スコアを利用した交絡因子の影響を除去する方法としては、傾向スコアマッチング (propensity score matching), 層別解析 (stratified analysis), 傾向スコアを回帰分析に共変量として用いる方法, 逆確率重み付け法 (inverse Probability Weighting: IPW), 二重にロバストな推定法 (Doubly Robust: DR) などがある<sup>4,5)</sup>。

IPW はそれぞれの標本に、処置群の場合は傾向スコアの逆数、対照群の場合は1から傾向スコアを引いた値の逆数によって、重み付けをして、ATEを補正したものである。一方のDRはそれぞれの標本に対して共変量による回帰モデルを構築することにより、目的変数の予測値を求め、これと目的変数とによって、傾向スコアの逆数によって、加重平均させてATEを補正したものである。IPWの定義式は式(4a)ならびに式(4b)に、ATEの定義式は式(5a)ならびに式(5b)に表される。

$$\hat{E}(y_1) = \sum_{i=1}^N \frac{z_i y_i}{e_i} / \sum_{i=1}^N \frac{z_i}{e_i} \quad (4a)$$

$$\hat{E}(y_0) = \sum_{i=1}^N \frac{(1-z_i) y_i}{1-e_i} / \sum_{i=1}^N \frac{1-z_i}{1-e_i} \quad (4b)$$

$$\hat{E}^{DR}(y_1)$$

$$= \frac{1}{N} \sum_{i=1}^N \left( y_{i1} + \frac{z_i - e(x_i, \hat{\alpha})}{e(x_i, \hat{\alpha})} \right) \left( y_{i1} - g(x_i, \hat{\beta}_1) \right) \quad (5a)$$

$$\hat{E}^{DR}(y_0)$$

$$= \frac{1}{N} \sum_{i=1}^N \left( \frac{(1-z_i) y_{i0}}{1-e(x_i, \hat{\alpha})} + \left( 1 - \frac{1-z_i}{1-e(x_i, \hat{\alpha})} \right) g(x_i, \hat{\beta}_0) \right) \quad (5b)$$

### (2) 本研究における統計的因果推論の適用方法

本研究における統計的因果推論の適用に際しては、傾向スコアを2項ロジットモデルによって推計し、推計された傾向スコアを回帰分析の共変量として用いる方法により行った。そして、処置の有無による効果の差については、IPWによる方法を用いた。

### 3. 分析に用いたデータの概要

本研究で利用した分析対象データは、高松琴平電気鉄道において導入されているスマートカードの利用履歴データである。本研究では「乗継ぎ割引拡大」施策の効果

表-1 検討対象データの概要

	利用回数 [回]	利用者数 [人]	一人あたり利用回数 [回/人/14ヶ月]
鉄道	8,515,524	93,109	91.5
バス	2,004,903	55,872	35.9
計	10,520,427	107,412	97.9

※2013年3月1日～2013年9月30日, 2014年3月1日～2014年9月30日

表-2 交通手段を組合せた利用別にみた検討対象データの概要

	利用回数 [回]	利用者数 [人]	一人あたり利用回数 [回/人/14ヶ月]	備考
鉄道→鉄道	3,674,582	77,586	47.4	
鉄道→バス	196,804	24,249	8.1	
バス→鉄道	197,210	25,044	7.9	
バス→バス	656,408	35,661	18.4	
none→鉄道	4,643,732	90,610	51.2	注1)
none→バス	1,151,691	49,489	23.3	注1)
計	10,520,427	—	—	

注1)当該利用日初回の利用

分析を行うため、検討対象期間を施策実施前後の1年間と考えた。しかし、「乗継ぎ割引の拡大」施策実施の概ね半年後の2014年10月1日より実施した高松市在住の高齢者を対象とした料金割引施策による影響が大きいことから、検討対象期間を施策実施前後の3月～9月の7ヶ月間とした。具体的には、2013年3月1日～2013年9月30日を施策実施前（以下、「事前」と記す）、2014年3月1日～2014年9月30日を施策実施後（以下、「事後」と記す）とした。

なお、本研究においては、以下の要件を満たすデータを有効データとした。

- ・鉄道、バスともに、乗車ならびに降車の利用履歴が記録されたデータ
- ・鉄道、バスともに、乗車ならびに降車のデータを1セットとして、利用回数を1回と定義した。

以上より、検討対象とした利用履歴データの概要は表-1のとおりである。鉄道の利用履歴データが8,515,524件、バスの利用履歴データが2,004,903件で、合計10,520,427件となった。

表-3 事前事後別にみた利用状況の概要

利用状況		総利用回数 [回]		乗換利用回数 [回]		乗換なし利用回数 [回]		利用者数 [人]
事前	事後	事前	事後	事前	事後	事前	事後	
利用なし	乗換なし	0	977,719	0	0	0	977,719	19,935
利用なし	乗換あり	0	385,907	0	54,305	0	331,602	5,312
乗換なし	乗換なし	2,566,374	2,256,804	0	0	2,566,374	2,256,804	38,206
乗換なし	乗換あり	371,472	451,390	0	27,065	371,472	424,325	6,134
乗換なし	利用なし	622,733	0	0	0	622,733	0	18,517
乗換あり	乗換なし	427,162	267,701	23,064	0	404,098	267,701	6,147
乗換あり	乗換あり	1,026,858	972,752	130,293	136,594	896,565	836,158	9,700
乗換あり	利用なし	193,555	0	22,692	0	170,863	0	3,461
計		5,208,154	5,312,273	176,049	217,964	5,032,105	5,094,309	107,412

※事前:2013年3月1日～2013年9月30日

※事後:2014年3月1日～2014年9月30日

また、本研究の主題である鉄道とバスを組合せた利用については、「乗継ぎ割引」が適用される乗継ぎの条件に準じて、以下のように設定した。

- ・同一日の利用に限定
- ・鉄道→バスの順番もしくは、バス→鉄道の順番の利用を対象。

以上より、本研究において対象とした鉄道とバスの利用の組合せ別による利用履歴データの概要は表-2のように整理できる。本研究において着目する鉄道とバスを組合せた利用となる「鉄道→バス」「バス→バス」は、あわせて、394,014件となった。

以降においては、事前、事後の各期間において、1度でも鉄道とバスを組合せた利用があれば、「乗換あり」と表現した。また、逆に1度も鉄道とバスを組合せた利用がなければ、「乗換なし」と表現した。さらに、事前、事後の各期間において、1度も鉄道あるいはバスの利用がなければ、「利用なし」として分類した。このように分類し、事前事後別に利用状況を整理したものが表-3となる。表-3によれば、「乗換あり（「鉄道→バス」「バス→鉄道」を合わせた利用）」の利用回数は、事前で176,049件、事後で217,964件となっており、全体では41,915件増加した。なお、この増加分については、乗継ぎ割引の拡大によるものか、それとも単に交通行動が変わって、事後から利用が増えた人であるかについてはわからない。

#### 4. 本研究における統計的な因果推論の適用結果

##### (1) モデルの構築

本研究の目的は、「乗継ぎ割引の拡大」による施策実施の効果検証である。そのため、「事前において乗換利用をしていない人」を処置群として設定した。一方、対照群は事前に乗換利用をしていた人とした。これにより、

表-4 モデル適用に際してのダミー変数の設定

Case 設定の趣旨 (仮説)	目的関数	処置群	対照群
「鉄道とバスを組合せた利用の促進」によって、鉄道の利用を誘発させた	事後の鉄道の利用回数	事前に鉄道を 56 回(週 1 回程度)以上利用していた人で、「鉄道とバスを組合せた利用」をしていなかった人	事前に「鉄道とバスを組合せた利用」をしていた人

表-5 傾向スコアモデル推定結果

	SF大人	SF学生	SF高齢者	定期大人通勤	SF障害者
(Intercept)	1.544209 ***	1.354414 ***	1.08492 ***	4.01289 ***	1.215455 ***
rate	-30.111 ***	-25.57199 ***	-16.0809 ***	-10.5351 ***	-12.01104 ***
bus_total	0.032983 ***	0.031991 ***	0.01354 ***		0.016726 ***
サンプル数	3,441	749	1,959	8,387	835
初期尤度	-2314.55	-510.22	-1358.42	-1061.14	-575.64
最終尤度	-1518.61	-369.1017	-974.72	-845.87	-414.4396
尤度比	0.34	0.28	0.28	0.20	0.28
AIC	3,043.23	744.20	1,955.45	1695.74	834.88

※ 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

事前に乗換利用をしていなかった人が「乗継ぎ割引の拡大」施策によって、事前に乗換利用をしていた人と比較して、鉄道利用が増加したか否かという観点で評価をするものである。

なお、表-3 より、事前に乗換利用をしていなかった人は 62,857 人であった。同様に事前に乗換利用をしていた人は 19,308 人であった。なお、事前の利用回数でみると、乗換利用については、176,049 回、乗換なしの利用回数は 5,032,105 回となっており、ほとんどが乗換なしの利用である。そこで、今回の検討においては、処置群には鉄道を定期的に利用をしていた人を対象とし、年数回といったほとんど利用していない人は処置群から除外することとした。具体的には、週 1 回程度以上の利用のある人を対象とした。

さて、傾向スコアモデルを構築するに際しては、処置群あるいは対照群の割付は従属変数の測定より時間的に先行しているため、従属変数の値によって割付が決まるということはない。ことから、事前の期間、すなわち 2013 年 3 月 1 日から 2013 年 9 月 30 日において測定されたデータを用いてモデルの構築を行った。具体的には、以下の説明変数を導入することによって傾向スコアモデルを構築することとした。なお、目的変数は処置群ダミー（処置群ならば 1、対照群ならば 0 を示すダミー変数）とした。

rate : 事前の鉄道全体の利用回数に対する鉄道とバスを組合せた利用回数の割合

bus\_total : 事前のバスの利用回数

モデル構築に際しては、利用券種別に効果の違いなどを分析することができるように、利用券種ごとにモデル

の構築を行った。具体的には、乗継ぎ割引が適用される「SF 大人」「SF 学生」「SF 高齢者」「定期大人通勤」である。そして、2014 年 10 月 1 日から運用を開始した高齢者割引のためのカードである「SF 障害者」の計 6 種類のモデルを構築した。モデルの推計に際しては、説明変数間の多重共線性が生じていないことを確認した上で行った。推計の結果を表-5 に示す。すべての傾向スコアモデルにおいて、尤度比が 0.2~0.3 程度を示しており、まずまずの適合度が得られた。また、定期大人通勤についてのみ、モデルの適合度を勘案して、説明変数を rate のみとした。

さて、個別のパラメータについては、すべてのモデルにおいて rate の符号は負、bus\_total の符号は正となり、すべて 0.1% 以下の確率で統計的に有意となった。

つぎに、傾向スコアモデルにより推計された傾向スコアを説明変数とする重回帰モデルの構築を行った。

モデル構築に際して、本研究の目的である鉄道とバスの乗継ぎ割引の拡大施策の実施によって、幹線系公共交通手段である鉄道の利用回数が増加したかを検証するため、目的変数は事後の鉄道利用回数とした。説明変数には、処置群ダミーである iv と傾向スコアの ps を設定した。

iv : 処置群ダミー。処置群ならば 1、対照群ならば 0 を示すダミー変数（傾向スコアモデルの目的変数）

ps : 傾向スコアモデルで推計された傾向スコア（構築したモデルによって異なる）

このようにして推計した重回帰モデルの推計結果ならびに IPW 推計量を求めた（表-7）。

表-6 処置群, 対照群別にみた検討対象サンプル数

	SF大人	SF学生	SF高齢者	定期大人通勤	SF障害者
対照群	1,372	315	992	232	378
処置群	2,070	435	968	8,156	458
計	3,442	750	1,960	8,388	836

表-7 重回帰モデルによるパラメータ推定結果

	SF大人	SF学生	SF高齢者	定期大人通勤	SF障害者
(Intercept)	2.52E+01 ***	2.92E+01 ***	2.77E+01 ***	2.86E+01	3.23E+01 ***
iv	6.95E+01 ***	2.28E+01 ***	7.15E+01 ***	1.41E+02 ***	1.01E+02 ***
傾向スコア	4.26E-01	-3.69E+00	-1.22E+00	6.06E+00	1.70E+00
サンプル数	3,441	749	1,959	8,387	835
決定係数	0.1864	0.04296	0.2965	0.03313	0.2443
調整済み決定係数	0.186	0.04039	0.2958	0.0329	0.2425
IPW_e1	111.4526	123.1802	111.9157	184.1874	88.85418
IPW_e0	19.16428	25.42992	25.35828	34.55536	37.57739
IPW推定量	92.28829	97.75031	86.55737	149.632	51.27679

※ 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

その結果, すべてのモデルにおいて, 決定係数が極めて低く, モデルの推計精度はあまりよくない. しかしながら, 処置群ダミーはすべてのモデルにおいて正の符号を示し, かつ 0.1%以下の確率で有意となった. 以上より, 今回の乗継ぎ割引の拡大によって, 鉄道利用の回数が向上したといえる. また, その回数は, 利用券種によって異なり, IPW 推計量の試算によると, 定期大人通勤で約 150回, SF 学生で約 98回, SF 大人で約 92回と推計された.

## 5. まとめ

本研究では, スマートカードの利用履歴データを用いて, 統計的因果推論の手法を援用することによって, 政策の効果検証を行うことを目的とした. 具体的には, 鉄道とバスの乗継ぎ割引の拡大施策の実施によって, 幹線系公共交通手段である鉄道の利用回数が増加したかという観点に着目して分析を行った.

その結果, 今回検討対象とした利用券種においては, すべての利用券種で鉄道の利用回数が統計的に有意に増加したということがわかった. またその定量的な効果は利用券種によって異なっていることが明らかとなった.

本研究では, 統計的因果推論の考え方を援用して施策の実施効果の検証の一例を示したに過ぎず, 本研究で示した以外にも様々な考え方がある. これらの考え方を援用しつつ, またケース設定の工夫などにより, より詳細な効果の把握をすることが可能であると考えられる. これらの複数のケース結果を比較しつつ, 乗継ぎ割引の拡大施策の効果がどのように波及していったかなど, 詳細

な知見を得ることなどが課題として残った.

**謝辞:** 本研究は高松琴平電気鉄道株式会社の多大なるご協力をいただいた. ここに感謝の意を表する. 本研究はJSPS 科研費 JP16K06548 の助成を受けたものである.

## 参考文献

- 1) 星野崇宏, 岡田謙介: 傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について, 保健医療科学 55(3), pp.230-243, 2006.
- 2) Paul R. Rosenbaum and Donald B. Rubin: The central role of the propensity score in observational studies for causal effects, *Biometrika*, Vol.70, No.1, pp. 41-55, 1983.
- 3) 岩波データサイエンス刊行委員会 編: 岩波データサイエンス Vol. 3, 岩波書店, 2016年.
- 4) 山下和哉, 塚井誠人, 桑野将司, 増田裕元: 過疎地域整備に関する認知喚起型調査手法の提案, 土木学会論文集 D3 (土木計画学), Vol.68, No.5 (土木計画学研究・論文集第 29 卷), pp.I\_207-I\_221, 2012.
- 5) 大林 準: ロジスティック回帰分析と傾向スコア(propensity score) 解析, 天理医学紀要 19(2), pp.71-79, 2016.