

1. はじめに

従来用いられてきた、郵送調査、訪問調査、電話調査などの社会調査手法に代わって、近年では、回収率低下などの問題を改善するためにWeb調査が多く実施される。Web調査では、母集団に合わせて特定の個人属性分布を調整するように標本を抽出する「スクリーニング」が行われるが、スクリーニングによって標本の偏りが改善される程度や、回答結果に与える影響については明らかではない。

また、公正な調査結果を得るために、調査票を設計する際に、質問文の言葉遣い、質問内容、質問形式の違いが回答結果に及ぼす影響に配慮すべきことは、繰り返し指摘されてきた。しかし、従来の研究では、回答者が、日常的生活感覚から判断することが難しい地域公共交通政策について回答する場合に、回答者にどのように説明を行うべきか、研究の蓄積が乏しい。

本研究は、Web調査における標本抽出方法の影響、および異なる質問形式が回答結果に与える影響を把握することを目的とする。なお、標本抽出方法については、スクリーニング標本と非スクリーニング標本を用いて結果を比較する。質問形式による標本個人属性の偏りは傾向スコアを用いた補正を行う。

2. 傾向スコア法

政策への賛否などの主要な設問を、全体の末尾に配した調査票について考えよう。この場合、最終設問に至るまでに、調査票に含まれる予備的な設問が回答者の認識を喚起することによって、最終設問の回答が変化する可能性がある。ただし、政策の理解度の違いによる回答の違いを比較する際、同一の標本では比較できないため、個人属性の異なる標本間の比較をせざるを得ない。本研究では、傾向スコアを用いて、標本の同一性を確保した別の標本間の回答との比較によって把握する。

研究デザインは、原因となる変数（説明変数）の操作が可能な実験と、それが不可能な調査観察に大

別できる。傾向スコアは、後者の場合でも、因果効果を適切に推定できる方法として、Rosenbaum & Rubin¹⁾により提案された。これは、因果効果を推定する際に、標本が異なる共変量条件下で観測されていると、共変量による因果効果と処置割当てによる因果効果が識別できないという問題に対処している。この方法の核となるアイデアは、複数の共変量を一つの変数に集約することができれば、その変数に基づいて標本を層化することで、共変量と割当てが目的変数に及ぼす影響のうち、割当てによる純粋な影響を明らかにできる、というものである。 x_i を回答者 i の共変量、 z_i を割当て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、共変量は性別・職業・家庭環境などの個人属性であり、割当て変数は調査票の種類 ($z_i = 1$: 調査票①, $z_i = 0$: 調査票②③) とする (図-1)。傾向スコアは、 $z_i = 1$ となる確率、すなわちその回答者 i の調査票①への割当てられやすさを共変量で説明した確率である²⁾。各回答者の傾向スコアは、真値を得られないため、ロジットモデルを用いて算出する。

2. 1. マッチング

推定した傾向スコアを元に、処置による因果効果を調べることができるが、この際の主な手法として、マッチングがある。単純に考えると、調査票の違いによる因果効果を調べるには、処置群と非処置群 (対照群) を比較し、その差を計算すればよい。しかし、回収された標本特性に二群間で差があることが原因で、結果に違いが生じる可能性もある。このバイアスを補正するために、二群間で差のある共変量の分布を、可能な限り一致させることが、マッチングの目的である。

本研究では、傾向スコアに基づいて層を作成し、傾向スコアが全く同一の値でなくても、同一層にあ

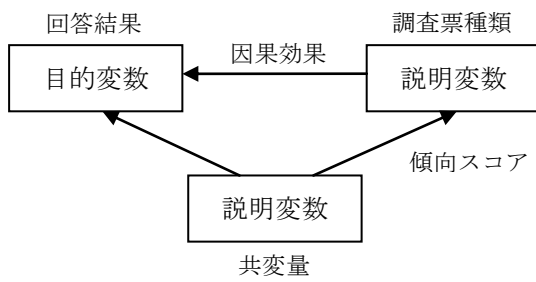


図-1 独立変数・従属変数・共変量の関係

ればペアとした。本研究で用いた Coarsened Exact Matching では、各群で標本の復元抽出が認められおり、必ずしも 1 対 1 のペアではなく、両群からそれぞれ複数回抽出される可能性があるほか、共変量の組合せによっては全く抽出されない（ペアを形成しない）標本が現れる可能性がある。つまり、この手法は標本中で最も類似した処置・非処置の標本をマッチングする方法と比較して、対になるペアの抽出回数そのものを決定していることになり、決定される値によって、処置群と対照群の標本平均を求める際の重みを与える役割を果たしている。

2. 2. IPW 推定

マッチングの他に因果効果を推定する方法である IPW 推定は、回答者ごとに傾向スコアによって回答結果に重みづけを行う。y を回答結果、 z_{ij} を回答者 i が調査票 j に回答するダミー変数、 e_{ij} を回答者 i の調査票 j に対する傾向スコアとすると、以下の式で表わされる。

$$E(y_z) = \frac{\sum_{i=1}^N \frac{z_{ij} y_i}{e_{ij}}}{\sum_{i=1}^N \frac{z_{ij}}{e_{ij}}} \quad (2)$$

$$E(y_{1-z}) = \frac{\sum_{i=1}^N \frac{(1 - z_{ij}) y_i}{1 - e_{ij}}}{\sum_{i=1}^N \frac{1 - z_{ij}}{1 - e_{ij}}} \quad (3)$$

調査票①に回答しやすい個人属性をもつ回答者の回答は軽く、反対に、回答しにくい回答者の回答は重く見ることができる。本研究では、賛成・反対で選択する鉄道復旧政策モデルと、ある金額に対して増額・維持・減額のどれに賛成かを選択するバス事業費負担モデルを作成した。前者については二項ロジットモデルを利用し、後者は生存時間モデルを利用

した。さらに、傾向スコアを適切に導入するために、それぞれの対数尤度関数に、式(2)、(3)に基づいて重みづけを行った。

3. 調査概要

表-1 に、「他地域の交通および復興を支援する政策に関する意識調査」の概要を示す。本調査は、2011 年 11 月 17 日（金）～11 月 21 日（月）に Web 調査として実施した。調査対象者は、中国地方在住の 20～69 歳の男女とし、全サンプル数を 900 とした。

本調査では、回答者属性が母集団と異なることによる回答への影響を明らかにするため、半数の 450 サンプルに関して、スクリーニングを行った。以下、本研究においてスクリーニングとは、平成 22 年度国勢調査における中国地方 5 県（鳥取、島根、岡山、広島、山口）の居住人口割合、若年者数（20～39 歳）・中高年者数（40～69 歳）の割合に合う標本が得られるように、回答者を募集することを示す。また、政策への理解度の違いによる回答結果の違いを明らかにするため、異なる質問項目の組合せで構成される調査票を、3 種類作成した。また、政策への理解度の違いによる回答結果の違いを明らかにするため、異なる質問項目の組合せで構成される、3 種類の調査票を使用した。標本数をまとめたものを、表-2 に示す。

設問は大きく分けると 4 部で構成されており、回答者の個人属性、交通不便地地域のバス事業費負担支援に関する質問、東日本大震災の被災地の鉄道復旧支援に関する質問、アンケート評価、の順に問う。これらのうち中心的設問は、B に含まれる山間部バス事業費の増額・維持・減額のうちどれに賛成かを回答する質問と、C に含まれる東日本大震災によって被災した三陸鉄道の復旧政策に賛成か反対かを回答する質問である。調査票②、③に項目 B-1、C-2、C-3 を取り入れることによって、それぞれの政策への理解度を増進させ、政策賛否（B-2、C-4）の回答結果が異なることを確認するためである。なお、C-1 は、被災地に関する訪問経験などの事実を問う設問であって、賛否に影響は無いと考えられるため、調査票間で選択項目としなかった。また、C-3 ではクイズ形式の設問を導入しているが、調査票②と③では、その正答を示すタイミングが異なる。すなわち

前者ではクイズの直後に正答を示す一方で、後者では全質問の最後に正答を示した。これらを比較することによって、クイズの正答に関する知識ではなく、クイズに回答することによって喚起される、被災前の三陸鉄道の運営に関連する状況が政策の賛否に及ぼす影響のみを把握することができると考える。

表-1 調査概要

調査名	他地域の交通および復興を支援する政策に関する意識調査
調査期間	2011年11月17日（金）～11月21日（月）
調査会社	株式会社クロス・マーケティング
調査型式	Web調査
調査対象者	中国地方在住の20～69歳の男女
居住地・年齢構成割合	中国地方における居住都道府県、若年者・中高年割合（H22国勢調査参考）
全サンプル数	900

表-2 標本数

	スクリーニング無			スクリーニング有		
組合せ	①	②	③	①	②	③
サンプル数	150	150	150	150	150	150

表-3 設問構成

項目番号		質問内容
A	A-1	個人属性、世帯属性 利用交通
	B-1	交通権に対する認知 移動制約者に関する質問
B	B-2	バス事業費の賛否（CVM型式）
	C-1	東北地方居住・訪問経験 慈善活動に関する質問
C	C-2	被災家屋の再建地に関する質問
	C-3	三陸鉄道に関するクイズ
	C-4	三陸鉄道復旧費用の賛否（金額提示）
	D-1	アンケートの評価

4. 分析結果

クロス集計からは、居住地と年代に基づきスクリーニングを行ったことで、両者について分布の差を確認できた（図-2、図-3）。国勢調査による割合と比較して、年代については20代と30代が多く、居住地については鳥取県と島根県の在住者が多く回収された。これは、スクリーニング標本として若年層が集まりやすかったためと考えられる。また、過疎地域を多く保有する鳥取県と島根県在住者が、調査題目に反応しやすかったと考えられる。また、調査票間での回答の違いが生じるか確かめるために、調査票と政策賛成率のクロス集計を行ったところ、調査票間でのわずかな差が見られた。

傾向スコア推定については、有意となる共変量が少なかったものの、モデルの適合度は中程度であっ

た。なお、調査票に組み入れた共変量項目は33あり、それらすべてを用いて傾向スコアを推定すると適合度が悪くなるため、二段階で共変量の選別を行った。まず、その共変量単独でモデル推定した場合にパラメータが10%以下の有意水準を示したものを採用したうえで、t検定を用いて、調査票間において共変量の平均値に差があると判定されたものを、追加で採用した（表-4）。

この傾向スコアを用いてマッチング・IPW推定を行った。なお、表-5に、調査票①～③それぞれの政策への賛成率の差を示す。例えば調査票①は、①と②③の賛成率の差を計算した値である。また、百分率の差であるため、単位にポイント（pt.）を用いている。バス事業費負担政策賛否について有意な調査票間の個人属性の偏りの影響は現れなかった。

さらに、非スクリーニング標本に対して傾向スコアで重みをつけたことで、年齢が回答に及ぼす影響が見られなくなった。一方、スクリーニング標本では、傾向スコアの導入の有無によらず、年齢パラメータは有意であった。（表-6）

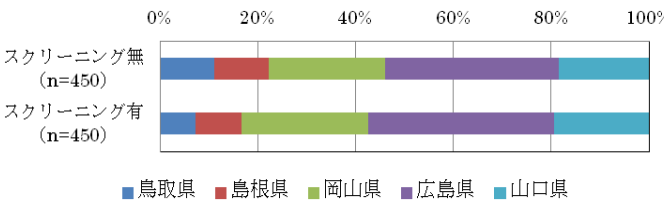


図-2 スクリーニングと居住地の関係

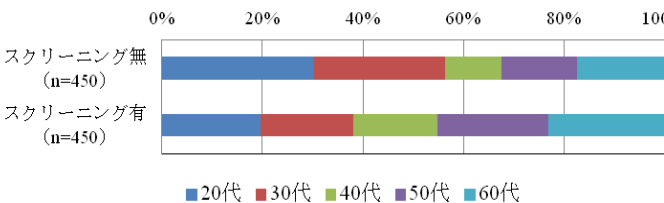


図-3 スクリーニングと年代の関係

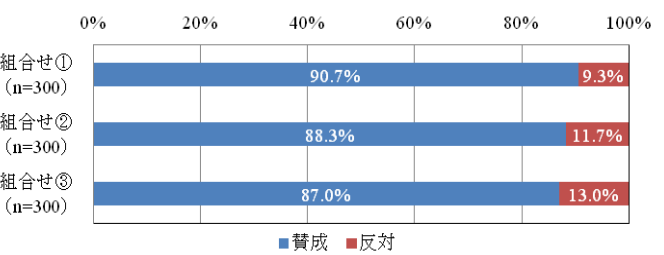


図-4 調査票と賛成率の関係

表-4 10%以下の有意水準を示した共変量項目

傾向スコアの分類	傾向スコア推定 による選択項目	二群間における差 による選択項目
組合せ①	フェリー利用 アストラムライン利用 家族による送迎利用	タクシー利用 徒歩利用 家族形態 世帯内高齢者
組合せ②	家族による送迎利用	家族形態
組合せ③	—	鉄道利用 自動車保有 徒歩利用
スクリーニング	鳥取県居住 世帯内高齢者 居住年数 外出困難 電動カート利用 バイク利用	—

表-6 (a) バス事業費負担モデル推定結果
(非スクリーニング標本)

共変量	スクリーニング無	
	傾向スコア無	傾向スコア有
調査票①	0.041	0.030
調査票②	0.034	-0.003
年齢	0.004 **	0.002
広島県在住	0.073 +	0.054 +
収入 (500~1000万円)	-0.051	-0.058 +
送迎利用	-0.085	-0.123 **
尺度パラメータ	1.206 ***	1.325 ***
形状パラメータ	2.955 ***	3.201 ***
サンプル数	450	450
自由度調整済み尤度比	0.020	0.010

表-6 (b) バス事業費負担モデル推定結果
(スクリーニング標本)

共変量	スクリーニング有	
	傾向スコア無	傾向スコア有
調査票①	-0.024	-0.020
調査票②	0.029	0.033
年齢	0.005 ***	0.005 ***
山口県在住	-0.119 **	-0.113 **
収入 (500万円未満)	0.019	0.018
単身世帯	-0.109 +	-0.109 *
バス利用	0.072	0.064
送迎利用	0.082	0.088
尺度パラメータ	1.134 ***	1.148 ***
形状パラメータ	2.915 ***	2.979 ***
サンプル数	450	450
自由度調整済み尤度比	0.026	0.011

表-5 バス事業費負担政策賛成率の調査票間の差

種類	スクリーニング無			スクリーニング有		
	傾向スコア 導入前	傾向スコア導入後 マッチング	IPW推定	傾向スコア 導入前	傾向スコア導入後 マッチング	IPW推定
調査票① (①-②③)	2.7pt.	2.6pt.	2.6pt.	-3.3pt.	-5.5pt.	3.3pt.
調査票② (②-①③)	-0.3pt.	-0.3pt.	-0.2pt.	0.7pt.	0.2pt.	0.4pt.
調査票③ (③-①②)	-2.3pt.	-2.3pt.	-2.2pt.	2.7pt.	2.8pt.	-3.0pt.

5. 結論

本調査票では、傾向スコアによる補正によらず、調査票間の回答の違いは生じなかった。調査票が回答結果に影響を及ぼす可能性については、今後も検証を重ねる必要がある。また、回答に差は現れなかったが、政策に対する理解度に差が生じているかは不明である。理解度を確認する質問項目を組み込み、質問項目の違いが理解度に与える影響、さらに回答に与える影響を二段階で推定する方法について検討する必要がある。

傾向スコア算出に関しては、共変量選択の明確な規則は提案されていないため、現状では、様々な調査データによる多くの分析を重ねる必要がある。今後はその基準について研究を行う必要がある。

傾向スコアの効果は、調査データを直接モデル化するとき生じる、割当てを介した間接効果の除去にあり、必ずしも調査データの母集団構成比率からの逸脱が補正できる訳ではない。今後は、傾向スコアによる統計的因果の抽出補正と母集団構成比の補正を同時に考慮することにより、より適切な（おそらく時間安定性や空間移転性の高い）モデル構築につながる可能性がある。また、傾向スコアの導入とスクリーニングを組み合わせることにより、信頼性の高い結果が得られる可能性がある。

6. 参考文献

- 1) Rosenbaum, P.R. & Rubin, D.B., The central role of the propensity score in observational studies for casual effects. 70, 41-55, 1983
- 2) 星野 崇宏：調査観察データの統計科学—因果推論・選択バイアス・データ融合、岩波書店、2009