

【研究ノート】

傾向スコアによるロイヤリティ・プログラムの効果の検証

拜殿 怜奈¹・豊田 秀樹²・池原 一哉³

要約

現在、ロイヤリティ・プログラムの一環としてポイントプログラムがさまざまな企業で実施されているが、その効果の検証方法は確立されていない。本論文では、傾向スコアを用い、ポイントプログラムの顧客ロイヤリティに対する育成効果を推定するための方法を提案する。実際にID-POSデータから得られる共変量を用いて傾向スコアを推定し、星野(2009)の方法により因果効果を推定した結果、値は正となった。したがって、当該企業のポイントプログラムによって顧客ロイヤリティは高められていると考えられ、顧客の消費行動が促進されていることが示唆された。

Key word: 傾向スコア, ポイントプログラム, ロイヤリティ

1. はじめに

現在、日本の小売業やサービス業においては、商品やサービスの多様化に伴い競合する企業が増えている。こうした状況下では、顧客ロイヤリティを高め、消費者を囲い込むことが非常に重要である。顧客の忠誠心(ロイヤリティ)を得ること、すなわち顧客を自社に囲い込んで売上高の増加を達成することを目的とする(海保, 2010) 施策に、ロイヤリティ・プログラムがあり、その1つがポイントプログラムである。ポイントプログラムは、一定期間当該ブランドや企業の商品・サービスを忠実に利用した会員に対してインセンティブを提供するマーケティング施策であり、日本における初めてのポイントプログラムは、1989年に導入されたヨドバシカメラの「ゴールドポイントカード」とされている。現在では、商品を販売する企業にとどまらず、サービスを提供する企業も、そ

の多くがポイントプログラムを導入している。例えば、購入一定金額に対してポイントが付与されたり、顧客の購買行動に応じて変動する評価ステージごとにポイントが付与されたりするなど、その形態はさまざまである。こうしたプログラムには、ポイントを獲得しようとする購買活動が促進される短期的な影響と、ポイント獲得後に購買活動が促進される長期的な影響の2つの効果があるとされている(Taylor & Neslin, 2005)。しかしDowling & Uncles(1997)では、こうしたプログラムでは消費者の行動を根本的に変化させることは難しいという指摘がなされている。また、消費者にとってロイヤリティ・プログラムが当たり前となった現在、「ポイントはもらうものの、ポイントのためだけに商品・サービスや店舗を決めるわけではない」という消費者が増加していることが富田(2013)で示唆されており、適切なプログラムでないものはバラマキとなり、企業にとっては重大な損失となる。したがって、ポイントプログラムなどインセンティブを消費者に与えるプログラムの効果を測定することは非常に重要である。しかし、こうしたプログラムに関して、顧客ロイヤリティへの効果の定まった測定法は確立されていない。

2. 目的

図1は、ある企業における優良顧客(実線)と非優良顧客(点線)の各月の1回の平均購入金額を示したものである。この企業では1年間の顧客の購入金額に応じて評価ステージが変動し、そのステージに応じてポイントが付与されるポイントプログラムを行っている。このとき優良顧客はインセンティブを受けており、非優良顧客はインセンティブを受けていない。このポイントプログラムは2月にステージが一度リセットされ、インセンティブも消失する。

このようなポイントプログラムにおけるロイヤリティへの効果の測定方法として、まず施策を実施する前後で

1 早稲田大学大学院文学研究科

2 早稲田大学文学学術院

3 国際交流基金日本語試験センター

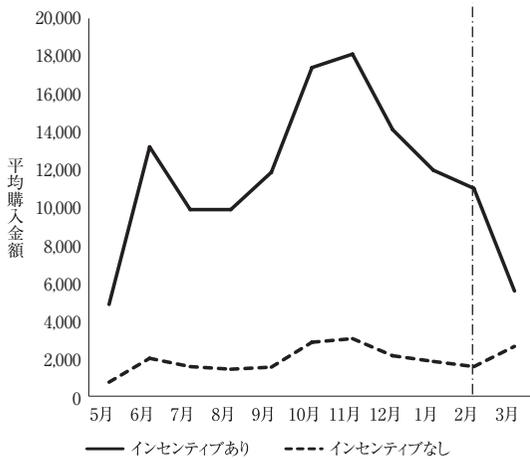


図1 1回の平均購入金額

比較する方法がある。プログラムの実施後と実施前の平均購入額の差を効果と考えるため、算出は容易である。しかし、季節や時期による影響や外部からの影響を取り除くことができないという問題点を抱えている。この問題点を改善する方法として、施策実施の有無で群分けして比較する方法がある。この方法では、同時期のデータを比較するため、季節や時期、外部からの影響は受けない。しかし、全ての時点で群間の差を直接プログラムの効果とすることはできず、さらに群によって特性が異なってしまう場合には、正しく効果を測定することができない。

図1において5月から2月までの優良顧客と非優良顧客の平均購入金額の差は、プログラムにより高められたロイヤリティだけではなく、インセンティブによる影響を含んでいる。したがって、インセンティブを与えている時点での差をプログラムによって高められたロイヤリティの効果とすることはできない。インセンティブによる影響を取り除くために、ステージのリセットに伴いインセンティブが消失した3月時点における優良顧客と非優良顧客の平均購入金額の差を比較することを考えると、その差は2,936円である。しかしこの差は依然として、プログラムによって高められたロイヤリティの効果を純粹に示しているわけではない。優良顧客群と非優良顧客群はランダムに割り当てられておらず、プログラム実施前から持つロイヤリティに違いがあるためである。もともとロイヤリティが高いからインセンティブを得られたのであり、その差はインセンティブがロイヤリティを育成した効果だけで説明されたものではないからである。

	優良顧客 ($z = 1$)	非優良顧客 ($z = 0$)
y_1	(1) 優良顧客でインセンティブを受けたときの結果 $E(y_1 z=1)$	(2) 非優良顧客でインセンティブを受けたときの結果 $E(y_1 z=0)$
y_0	(3) 優良顧客でインセンティブを受けなかったときの結果 $E(y_0 z=1)$	(4) 非優良顧客でインセンティブを受けなかったときの結果 $E(y_0 z=0)$

図2 データの欠測と割り当て

したがって、プログラムの効果を測定するためには優良顧客群と非優良顧客群のもともと持つロイヤリティを調整し、揃える必要がある。本論文では群間の差を調整するために傾向スコアを用い、ポイントプログラムの顧客ロイヤリティに対する育成効果を推定するための方法を提案する。

3. 方法

3.1 因果効果

当該ブランドの優良顧客であるか否かは、無作為に割り当てられるのではなく、さまざまな変数によって影響を受けているため、単純に優良顧客と非優良顧客を比較することではその差を測ることができない。こうした割り当ての問題は欠測データの問題として捉えることができる (Rubin, 1987)。

ここで、ポイントプログラムで優良顧客としてのインセンティブを受けた時の従属変数の値を y_1 、インセンティブを受けなかった時の従属変数の値を y_0 とし、顧客が優良顧客である場合には $z = 1$ 、非優良顧客である場合には $z = 0$ とする割り当て変数 z を考える。また、顧客を $i (i = 1, \dots, N)$ とする。さらに、優良顧客(処置群, $z = 1$)に関して実際には y_1 のみが観測されるが、本来は y_0 も存在しており、 y_0 が欠測したデータであるとする。同様に非優良顧客(対照群, $z = 0$)に関して、 y_1 が欠測しているとする (図2)。

ここで、ルービンの因果効果 (Rubin, 1974) を

$$E(y_1 - y_0) = E(y_1) - E(y_0)$$

によって定義する。優良顧客と非優良顧客の周辺期待値 $E(y_1)$ 、 $E(y_0)$ は、全顧客の期待値によって推定可能だが、欠測値を含むため実際に算出することはできない。また、星野(2009)では、因果効果は処置群と対照群がどのような母集団から抽出されたものであるかに依存する量であることが示されている。例えば、処置群と対照群の構成比率が1対10ならば、因果効果はほとんど対照群を抽出した元の母集団での因果効果とほぼ同値であるとされている。本研究で検討する優良顧客群と非優良顧客群は、一般に1対10よりもその比は大きくなるため、因果効果は非優良顧客がインセンティブを受けた場合の期待値と非優良顧客がインセンティブを受けなかった場合の期待値の差

$$E(y_1|z=0) - E(y_0|z=0)$$

となる。したがって、本研究においては、 $E(y_1|z=0)$ 、つまり、欠測している図2の(2)を算出することでポイントプログラムの効果を推定することが可能となる。

3.2 傾向スコアによる因果効果の推定

傾向スコアは、Rosenbaum & Rubin(1983)によって提案された無作為割り当てを伴わない観察研究において因果効果を推定する方法であり、 i 人目の顧客の傾向スコアは、 i 人目の顧客の共変量の値を x_i 、割り当て変数の値を z_i とすると、1群に割り当てられる確率 e_i

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

で定義される。傾向スコアの推定には、一般にロジスティック回帰やプロビット回帰が使用される。本研究ではロジスティック回帰を用いる。このとき、傾向スコアは割り当てを2つのカテゴリカルな基準変数としたロジスティック回帰分析における共変量による条件付き予測確率である。

通常の分析では、優良顧客と非優良顧客の周辺期待値 $E(y_1)$ と $E(y_0)$ の差を、傾向スコアを用いて推定するが、本研究では先に述べたとおり、 $E(y_1|z=0)$ のみを推定すればよい。IPW推定量(Rubin, 1985)ではなく、星野(2009,p.177)の $E(y_1|z=0)$ の推定量の式

$$\hat{E}(y_1|z=0) = \frac{\sum_{i=1}^N \frac{z_i(1-e_i)}{e_i} y_{1i}}{\sum_{i=1}^N \frac{z_i(1-e_i)}{e_i}}$$

を用いる。

4. 分析

4.1 データ

本研究では、経営科学系研究部会連合協議会主催の平成26年度データ解析コンペティションで提供されたデータを利用する。これは、株式会社良品計画が展開する無印良品の日本国内店舗およびオンラインショップにおける2013年5月から2014年6月までの顧客ID付き購買履歴データである。本データには、商品の購買年月日、時刻、個数、価格、購買店舗、購買した顧客のデモグラフィックス(年代、性別)が同時に記録されていた。また、商品に関しては、商品名のみではなく、部門名やデパートメント名、ライン名(商品の大分類)等も同時に記録されていた。顧客データには、株式会社良品計画が展開するアプリ「MUJI passport」の入会日、MUJIカードの入会日についても記録されていた。

株式会社良品計画では、3月から2月までの1年間の購入金額に応じて変動する評価ステージごとにポイントを付与しており、1年間で20万円以上購入している顧客をダイヤモンドステージ会員(最優良顧客)として位置づけ、1,000ポイントを付与している。なお、会員のステージは毎年2月にすべてリセットされる。

本研究では、1回の平均購入金額によって顧客ロイヤリティを測ることとした。1度の購買活動で家具などの高額商品を購入したり、雑貨等の商品をまとめて購入したりする顧客は無印良品に対するロイヤリティが高いと考えられるためである。今回はデータ利用の制約上、平均購入金額を顧客ロイヤリティの指標として用いたが、その他にも、購買リピート率や顧客満足度、他者への推奨意向などが考えられるだろう。図1の折れ線グラフは、2013年5月から2014年3月までのダイヤモンド会員と一般会員の1回の平均購入金額を示した図である。2月に会員のステージがリセットされるため、3月の平均購入金額を従属変数とし、5月から2月までのデータは共変量の算出に使用した。

分析の対象とした期間は、2013年5月から2014年3月までの11ヵ月であり、分析の対象とした顧客は、2014年2月までに20万円以上購入しているダイヤモンド会員

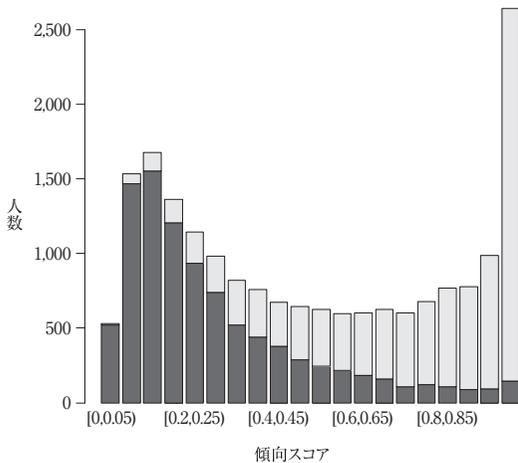


図3 傾向スコアの推定値

9,509名と2014年2月までに5回以上購入している一般会員から無作為抽出した9,509名の計19,018名であった。本研究では共変量として購入間隔を扱うため、購入した回数を5回以上の会員に絞って分析を行った。これは分析上の制約であり、4回以下の会員を考慮できていない。また、そもそも会員でなければ購入データを得ることはできないため、一般的かつ大多数を占める一般顧客を考慮できていないという点には注意が必要である。

4.2 傾向スコアの推定

傾向スコアを算出するために、性別(不明=0, 男性=1, 女性=2)、年代(不明=0, 10代以下=1, 20代=2, 30代=3, 40代=4, 50代=5, 60代=6, 70代以上=7)、MUJIカードの有無(無し=0, 有り=1)、無印良品週間⁴期間中の購入回数、使用店舗数、総購入デパートメント数、実店舗使用率、最大購入間隔、キャンセル回数、平均購入回数、最大購入回数の11項目の共変量を用いた。星野・前田(2006)や星野(2007)では、共変量を選択するための4つの基準が提案されている。本研究では、変数の数や質に制限があることから、2つ目の基準に重点を置き、従属変数と関連があると思われる変数をなるべく多く投入する(星野, 2009)ことで傾向スコアを算出する。なお、傾向スコアの推定にはロジスティック回帰分析を用いた。

4 無印良品メンバーであれば10%割引で購入できるキャンペーン

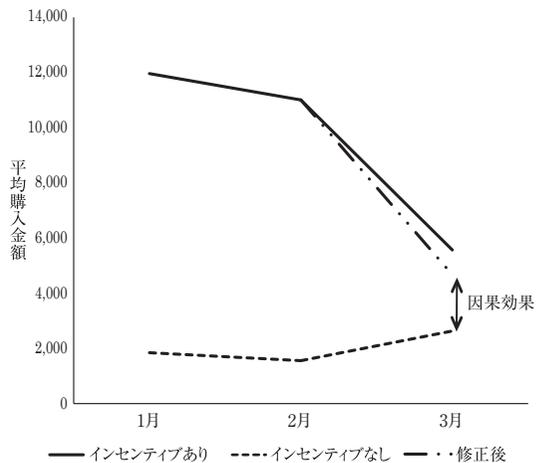


図4 ポイントプログラムと効果検証結果

5. 結果

はじめに、推定された傾向スコアを図3に示す。黒が一般会員を、灰色がダイヤモンド会員を表している。図3より傾向スコアの値は、一般会員の方が小さく、ダイヤモンド会員の方が大きい傾向にあることが分かる。ロジスティック回帰分析における正判別率は81.39%であり、概ね良好な推定であった。なお、傾向スコアが0.05未満の場合には、因果効果を算出する際に逆数をとったときの影響力が大きくなりすぎるため、以降の分析では0.05に固定して推定に利用した。

続いて、平均購入金額の期待値を算出した結果、ダイヤモンド会員がインセンティブを受けたときの平均購入金額が5,568.149円(図2における(1))、一般会員がインセンティブを受けたときの平均購入金額が4,666.461円(図2における(2))、一般会員がインセンティブを受けなかったときの平均購入金額が2,631.806円(図2における(4))となった。また、図4には2月から3月への変化に関して、推定された傾向スコアを用いて修正された平均購入金額(図2における(2))を二点鎖線で示した。

以上の数値より、因果効果は2,034.655円(=4,666.461 - 2,631.806)と推定された。一方、インセンティブを受けたダイヤモンド会員(修正前の平均購入金額:5,568.149円)とインセンティブを受けなかった一般会員(2,631.806円)の差は、2,936.343円である。修正によって約900円の差が見られるが、依然因果効果は正の値であり、1回の購入における平均金額は、インセンティブを受けた一般会員の方が受けなかった一般会員よりも

約1.8倍(≒4,666,461÷2,631,806)高いことが示された。したがって、無印良品のポイントプログラムによって、顧客ロイヤリティは高められており、顧客の消費行動は促進されていると言える。

6. まとめ

本研究では、傾向スコアを用いてロイヤリティ・プログラムの顧客ロイヤリティに対する効果の推定方法を検討した。分析に使用したデータは、一般的なID-POSデータから容易に加工が可能な変数であるため、共変量や従属変数を各企業のプログラムに応じて変更することで、本研究における分析はさまざまなプログラムの効果測定に使用することが可能である。

引用文献

Dowling, G. R. & Uncles, M.(1997) Do customer loyalty programs really work?, *Sloan Management Review*, 38(4), 71-82.

星野崇宏・前田忠彦(2006) 傾向スコアを用いた補正法の有意抽出による標本調査への応用と共変量の選択法の提案. *統計数理*, 54, 1, 191-206.

星野崇宏(2007) インターネット調査に対する共変量調整法のマーケティングリサーチへの適用と調整効果の再現性の検討. *行動計量学*, 34, 1, 33-48.

星野崇宏(2009) 「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

海保英孝(2010) ポイント・プログラムをめぐる経営の諸問題について, *成城・経済研究*, 187, 119-148.

Taylor, G & Neslin, S. A.(2005) The current and future sales impact of a retail frequency reward program, *Journal of Retailing*, 81(4), 293-305.

富田勝己(2013) 効果測定に基づくポイントプログラムの有効活用, *知的資産創造*, 2013年11月号, 12-21.

Rosenbaum, P. R. & Rubin, D. B.(1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55.

Rubin D. B.(1974) Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 688-701.

Rubin D. B.(1985) The use of propensity scores in applied bayesian inference. In J.M. Bernardo, M. H. De Groot, D.V. Lindley, & A.F.M. Smith(eds.), "Bayesian Statistics," 2 463-472. North-Holland: Elsevier Science Publisher B.V.

Rubin D. B.(1987) "Multiple Imputation for Nonresponse in Surveys", New York, NY : Wiley. (2016年6月27日採択)