

アプリを通じた段階的ロイヤルティシステムの効果検証

1. はじめに

本稿の目的は、スターバックスをはじめ多くの企業で見られるアプリを通じたポイント制度¹が消費者の購買意欲を促進し、企業の売り上げ増加に貢献するという因果効果が存在するかを検討することである。本稿ではまず、仮説とその学術的・実務的意義を明らかにし、次に仮説の根拠となる因果メカニズムを提示する。その後、適切な方法論について検討し、最後に調査の実現可能性と今後の展望について論じる。

2. 仮説と期待される学術的・実務的意義

前述のように、本稿のリサーチクエスション（以下、RQ）は「アプリ内のポイント制度は消費活動の増加に貢献するか」である。ここでいうポイント制度とは、商品を購入するたびに一定のポイントが付与され、後にそのポイントを用いて無料特典が得られるような仕組みを想定している。従来、ロイヤルティプログラムの導入が売上や顧客維持に与える影響については多くの分析がなされてきたが、アプリ特有のユーザーインターフェース（UI）や報酬の有効期限といった要素を踏まえたうえで、それぞれの構成要素が消費行動に与える効果を個別に測定した研究は限られている。したがって、本研究がポイント制度そのものの因果効果を示すことができれば、今後、報酬の水準や有効期限といったより細かい設計要素に関する因果分析の基盤となり得る。これは、企業にとってロイヤルティプログラムの設計を科学的根拠に基づいて最適化するための一助となる点で、実務上の意義も大きい。以上のことから、本稿の分析は、マーケティング実務に資すると同時に、アプリ型ロイヤルティプログラムに関する学術的知見の深化にも貢献することが期待される。

¹ Starbucks Coffee Japan - スターバックス コーヒー ジャパン

https://www.starbucks.co.jp/?srsltid=AfmBOopEdS5KbR-XQ7tDTKySXbdI92qrFFs8ANNSz9VbtZAzm_r8UJm6

3. 変数の選択と因果メカニズム

次に、RQ を具体的な変数に落とし込み、その因果関係に關与する他の変数を導入したうえで、因果メカニズムを考察する。まず、ポイント制度の利用については、ポイント機能を使用したか否かを表す二値変数（使用=1／不使用=0）とする。一方、消費行動は一定期間内の支出総額により測定される。したがって、RQ は以下の図 1 のように、二つの変数間の関係として簡潔に表すことができる。



図 1：処置変数と結果変数の因果関係

図 1 に示したように、ポイント制度の利用は、消費者の支出に因果的な影響を与えると考えられる。その背後にあるメカニズムとしては、将来的に無料特典が得られるという期待が消費者にとって心理的な利益となり、現在の支出に対する心理的な抑制が緩むこと、また、ポイント獲得自体が快感や達成感として報酬化されることによって、消費が促進されるといった要因が挙げられる²。

本研究の RQ を検証するうえで、最も基本的なアプローチは、ポイント機能の利用者と非利用者との間で、平均支出額を単純に比較する方法である。しかし、この単純比較には欠落変数バイアスが生じるリスクが高い。特に以下の 3 つの要因が交絡因子として重要であると考えられる。第一に、ブランド愛着度 (X_1) は、処置変数 (D) と結果変数 (Y) の共通の原因となりうる。ブランドロイヤルティの高い消費者は、そのブランドが提供するアプリを使用する可能性が高く、さらに支出額も大きくなる傾向があると予想される。第二に、所得 (X_2) もまた典型的な交絡因子である。高所得層は金銭的余裕があるため支出額が大きくなりやすい一方で、ポイント制度などのインセンティブ設計にはあまり関心を示さない可能性がある。これにより、ポイント機能の利用と支出額の間、所得を介した見かけ上の相関が生じる。第三に、年齢 (X_3) は、アプリの利用可能性や消費傾向に直接的な影響を与えるだけでなく、所得やブランド愛着度といった他の交絡因子にも影響を

² 第 79 回 ロイヤルティ・プログラムの効果について：現金値引きよりもサービスポイント

の方が魅力的なことがあるのだろうか？

<https://www.ogis-ri.co.jp/column/kr/285.html>

与える可能性が高い。たとえば、若年層はアプリに慣れており利用率が高い一方で、可処分所得やブランドへの忠誠心においては世代間で差があることが想定される。したがって、これらの交絡因子を統制しないまま平均の単純比較を行うと、ポイント制度の真の因果効果を歪めてしまう可能性がある。これらの交絡因子も含め、DAG を描画すると次の図 2 のようになる。

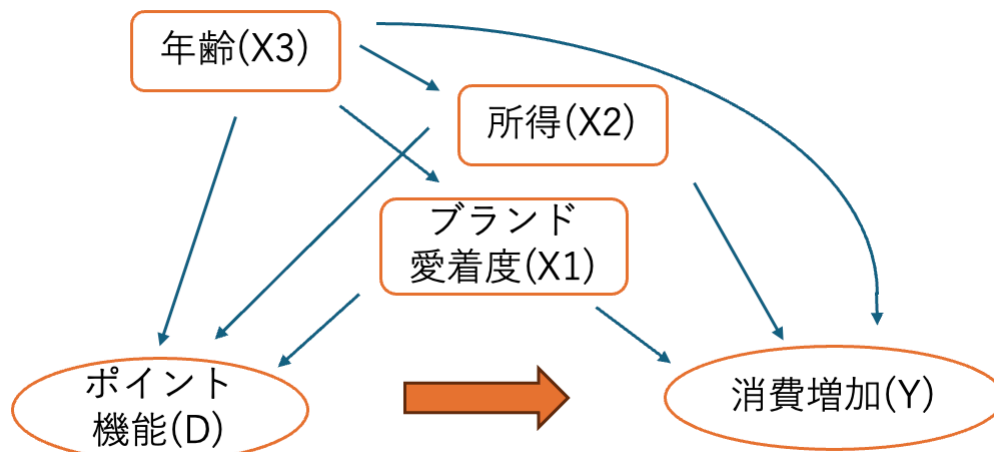


図 2：ブランド愛着度、所得、年齢を含む変数同士の関係

以上を踏まえると、本稿における平均処置効果（ATE）を重回帰分析によって推定する際には、ブランド愛着度（ X_1 ）、所得（ X_2 ）、および年齢（ X_3 ）を交絡因子として統制する必要がある。これらのうち、処置変数（ D ）、結果変数（ Y ）、所得（ X_2 ）、年齢（ X_3 ）はいずれも客観的な数値指標として比較的容易に測定可能である。一方で、ブランド愛着度（ X_1 ）の測定には注意が必要である。たとえば、一定期間内の支出額や購買頻度などの行動指標をもってブランド愛着度を代理しようとする、それらは結果変数（ Y ）と重複し、交絡因子として独立にコントロールすることが困難となる。実際、支出傾向や購買頻度は愛着の表れであると同時に、分析の目的であるアウトカムそのものであるため、両者の明確な切り分けが求められる。したがって、ブランド愛着度を適切に交絡因子として扱うためには、消費行動とは独立した心理的指標として測定する必要がある。そのためには、アンケート調査等を用いて、ブランドに対する親近感やロイヤルティに関する主観的評価を別途収集することが望ましい。また、因果推論の内的妥当性を高めるためには、その他の外的要因への配慮も不可欠である。特に、アプリ内の一時的なキャンペーンによりポイント還元率が通常より高く設定されている場合、消費の増加は制度自体の効果ではなく、キャンペーンによる誘因の結果である可能性がある。このような混合効果を排除するため、データ分析に際してはキャンペーンの実施有無や時期を正確に把握し、分析

対象から除外する、または統制変数としてモデルに組み込むといった対応が求められる。以上の検討を踏まえ、次節では本研究における具体的な調査手法および分析方法について論じることとする。

4. データ取得・方法論の検討

本研究において因果関係を明らかにするためには、理想的にはランダム化比較試験を通じて、アプリを使用する処置群と使用しない非処置群に無作為に割り当てることが考えられる。これにより交絡因子の影響を排除した厳密な因果推論が可能となる。しかし、現実的にはいくつかの制約が存在する。第一に、アプリの使用という介入の性質上、被験者をブラインド化することが困難であり、ホーソン効果などの観察バイアスが生じるリスクがある。第二に、消費行動を一定期間観察する必要があるため、実験室実験による検証は不適切である。さらに、フィールド実験としてランダムにアプリの使用可否を割り当てる場合、一部のユーザーのアクセスを制限することになり、倫理的・実務的な問題が発生する可能性がある。

これらの制約を踏まえると、RCT を実施する場合の現実的な代替手段としては、シナリオベースのサーベイ実験が挙げられる。この方法では、参加者に仮想的な購買状況やアプリ使用条件を提示し、行動意図や選好を測定することが可能である。実施のハードルは低い一方で、観測されるのはあくまで意図や自己申告ベースのデータであり、実際の行動との乖離が生じるおそれがあるという限界もある。

一方、観察研究に基づくアプローチは、実務的観点から現実性が高い。特に、企業に蓄積された POS（販売時点情報管理）データを活用することで、アプリの使用状況と消費行動を直接観察することが可能である。ただし、これらのデータを分析に用いるには個人情報保護の観点から制約があるほか、支払い手段によっては個人との紐づけが困難であり、結果としてサンプルが特定の決済手段（e.g. アプリ決済、クレジットカード）に偏る可能性もある。そのため、POS データを用いる場合であっても、補完的にアンケート調査を実施し、消費者の属性や心理的要因（e.g. ブランド愛着度）などを収集することが望ましい。以上の議論より、本研究では観察研究に基づき、実際の購買行動データとアンケートによる補足情報を統合した分析手法を採用することが、実現可能性と測定妥当性の両面から最も現実的かつ説得力のあるアプローチであると考えられる。よって、アンケート調査から処置変数、結果変数、交絡因子を含む共変量の値が得られることを前提とし、以下により具体的な分析手法の選定を行う。

まず最も基本的な手法として、重回帰分析が挙げられる。これは、処置変数（アプリの使用）と結果変数（消費額）の関係を交絡因子（ブランド愛着度、所得、年齢など）で統制しながら推定するものである。交絡因子がすべて観測可能であり、線形性の仮定が成立する場合にはシンプルかつ有効なアプローチとなる。ただし、観測されていない交絡因子が存在する場合や、処置群と非処置群で属性が大きく異なる場合にはバイアスのリスクが残る。

次に、操作変数法（IV 法）は、未観測の交絡因子が存在すると想定される場合に有効な方法である。しかし、本研究では「アプリを使うか否か」という処置に対して適切な操作変数を現実的に設定することは極めて困難である。例えば、アプリの利用可能性や地域別の店舗特性といった外的要因が考えられるが、それらが消費行動と無関係であるという強い独立性の仮定を満たす保証がない。ゆえに、操作変数法は理論的には有効だが、本研究では採用が難しい。

差分の差分法（DID）は、前後の消費変化に注目する方法であり、時間的な処置前後の比較が可能な場合に適している。しかし、アプリ導入のタイミングが一律でないことや、導入前後で外的条件（キャンペーンなど）が変化する可能性を統制することが難しいことから、適用には慎重を要する。

最後に、傾向スコアマッチング（PSM）は、観察データにおいて処置群と非処置群の属性のバランスを調整することで、処置の効果を推定する手法である。本研究においては、アプリの使用確率をロジスティック回帰などで推定し、そのスコアに基づいて類似の非処置群をマッチングさせることで、処置群との比較が可能となる。ブランド愛着度や所得、年齢など、処置決定に影響する共変量が事前に十分に観測できる前提があれば、PSM は重回帰よりも処置・非処置群の比較可能性を高める手段として有効である。

以上を踏まえると、今回の研究においては、観察可能な交絡因子が比較的限定的であること、サンプル数が仮に十分に確保されると想定できること、非線形性への柔軟な対応が可能であることなどの点から、傾向スコアマッチングを主たる手法とし、必要に応じて重回帰分析による頑健性の確認を行うというアプローチが、方法論的にも実現可能性の面でも最適と考えられる。

5. 調査・分析実施に向けた流れ

実際に研究を進める際の流れを分かりやすくするため、実際に集まったアンケート調査から実際に集まったデータのイメージを元に分析の流れを示す。

ID	アプリ利用 (D)	支出額 (Y)	年齢 (X ₁)	所得 (X ₂)	ブランド愛着度 (X ₃)
1	1	3500	22	300	5.0
2	0	2400	21	320	3.5
3	1	4100	28	400	4.8
4	0	2000	30	260	2.0

図 3：アンケート結果のイメージ

① 傾向スコアの推定

まず、処置変数 (D) を従属変数として、交絡因子 (X₁, X₂, X₃) を説明変数とするロジスティック回帰モデルを推定する。

$$\text{logit}(P(D=1)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

このモデルによって、各個体がアプリを利用する確率（傾向スコア）が算出される。

② マッチング

推定された傾向スコアに基づいて、アプリ利用者（処置群）と非利用者（対照群）をマッチングする。例えば、最近傍法³を用いて、各処置群に対して傾向スコアが最も近い対照群を 1 対 1 でマッチさせる。

③ 処置効果の推定

マッチされたペア間の支出額の差 (Y の差) を計算し、それらの平均を取ることで、アプリ利用の平均処置効果 (ATE) を推定する。本稿の RQ は、推定された ATE が統計的に有意に正の値を示した場合に支持される。具体的には、信頼区間に 0 を含まない正の値が観測されれば、ポイント制度の利用が支出額を有意に増加させると解釈できる。

³ 最近傍法と k 近傍法: Python での実装や違いについて

<https://pydocument.hatenablog.com/entry/2023/10/14/004008#最近傍法 Nearest-Neighbor>

④ 頑健性の確認

得られた推定値の信頼性を担保するため、以下の追加分析が考えられる。

- バランスチェック：マッチ後の共変量（年齢、所得、ブランド愛着度）において、処置群と対照群に有意差がないか確認。
- 感度分析：マッチング手法（1 対 1、半径マッチング、カリパー調整など）⁴を変えても処置効果が一貫しているか確認。
- 重回帰分析による頑健性チェック：マッチング後サンプルで改めて重回帰分析を行い、交絡因子の影響を統制した上で効果の一貫性を確認する。

6. 本稿のリサーチデザインの実現可能性と今後の展望

以上の議論を踏まえ、本稿では、アンケート調査とその集計結果に基づく傾向スコアマッチングの手法を用いることで、年齢・所得・ブランド愛着度といった潜在的な交絡因子の影響を統制しつつ、アプリによるポイント制度が消費行動に与える因果効果を推定できる可能性を示した。

序論でも述べたように、アプリを通じたポイント付与システムには、ユーザーインターフェースの設計、ポイントの有効期限、通知頻度など、企業が調整可能な多くの要素が含まれている。したがって、本稿で扱ったリサーチクエスションの妥当性が確認できれば、実務的にはこうした構成要素ごとの効果測定といった、より精緻なマーケティング戦略の設計につながると期待される。

しかし、アンケート調査に依存する分析には限界もある。例えば、回答の信頼性、回答者のサンプルバイアス、あるいは本稿で想定していない新たな交絡因子の存在によって、推定結果に歪みが生じるリスクは否定できない。今後の研究では、こうした不確実性に対する感度分析や、複数のデータソースを組み合わせた検証が求められるだろう。

⁴ 医学統計セミナー第 4 回-傾向スコア分析-

<https://waidai-csc.jp/updata/2019/05/7e3a9ea92b282ede82550e81a73c6b54.pdf>

7. 参考文献

- Starbucks Coffee Japan - スターバックス コーヒー ジャパン

https://www.starbucks.co.jp/?srsltid=AfmBOOpEdS5KbR-XQ7tDTKySXbdI92qrFfS8ANNSz9VbtZAzm_r8UJm6

- 第 79 回 ロイヤルティ・プログラムの効果について：現金値引きよりもサービスポイントの方が魅力的なことがあるのだろうか？

<https://www.ogis-ri.co.jp/column/kr/285.html>

- 最近傍法と k 近傍法: Python での実装や違いについて

<https://pydocument.hatenablog.com/entry/2023/10/14/004008#最近傍法 Nearest-Neighbor>

- ¹医学統計セミナー第 4 回-傾向スコア分析-

<https://waidai-csc.jp/updata/2019/05/7e3a9ea92b282ede82550e81a73c6b54.pdf>