

A/B テストによる
イールドマネジメントの効果検証方法

松本 涼佑*

Evaluation of Yield Management with Randomized Controlled Trial (A/B Testing)

Ryosuke MATSUMOTO

This study proposes a method to evaluate yield management through sales slot control for discount railway tickets using a randomized controlled trial (A/B test). Implemented on an operating railway line, this method empirically shows how sales slot adjustments influence revenue. The results demonstrate that the appropriate strategy is time-dependent: in certain periods, increasing slots stimulates demand and boosts revenue, while in others, decreasing slots suppresses down-selling for a similar positive effect. As the first published case of A/B testing for yield management in the Japanese railway industry, this paper provides a practical framework for implementation and analysis.

キーワード：イールドマネジメント，発売枠制御，ランダム化比較試験，A/B テスト

1. はじめに

人口減少に伴い利用者，労働者が共に減少傾向にある日本の鉄道では，大規模な設備投資を行わず，また可能な限り少ない業務リソースで運輸収入の確保・増加に資する施策が有効といえる．鉄道や航空等の交通分野では，提供する輸送サービス（商品）を在庫持ち越しすることができない．そのため，価格制御や割引商品（席数限定きっぷ）の発売枠（発売上限数）制御を行うイールドマネジメント（収益向上施策）が有効である．日本の鉄道では，複数の事業者で発売上限数を設けた割引商品が発売されてきたことから，割引商品の発売枠制御によるイールドマネジメントはより伝統的に実施されてきたと考えられる．そこで本稿では，割引商品の発売枠制御によるイールドマネジメントを対象とする．

通常商品と割引商品の2種類の商品があり，このうち割引商品のみが席数限定きっぷ，すなわち発売枠設定がある幹線鉄道を対象に，発売枠制御によるイールドマネジメントを実施することを想定する．ここで，割引商品の発売枠が過剰に小さいと，学生や非正規雇用者などの，相対的に支払意思額が低い旅客層の需要の逸走を招く可能性がある．反対に発売枠が過剰に大きいと，ビジネスマンの出張利用など，相対的に支払意思額の高い旅客層を割引商品に誘導してしまい運輸収入の減少を招く可能性がある．以上のようなメカニズムが働くことが考えられるため，割引商品の発売枠制御によるイールドマネジメントは運輸収入増加の観点から有効であるといえる．しかし適切な発売枠設定や，イールドマネジメントの事後検証に関する研究事例は限定的である．

日本の鉄道業界では発売枠の制御によるイールドマネジメントが伝統的に実施されてきたと考えられるにもかかわらず，学術的には価格制御によるイールドマネジメントの研究が中心^{例えば 1)2)}である．鉄道を対象に発売枠制御によるイールドマネジメントに着目した研究は，筆者の知る限り中川らによる一連の研究^{例えば 3)4)}に限られている．中川らは，運輸収入の最大化を目的に，数理最適化問題を解くことにより，最適な発売枠を算出するアルゴリズムを開発した．しかし中川らは，割引商品が発売枠に達し，その商品が買えなかったときに旅客がとる行動の定式化を，Web アンケートの仮想選択調査に基づいて行っており，

* 情報通信技術研究部 情報解析研究室

実需要データに基づいたものではなかった。また、開発したアルゴリズムが算出した発売枠を実営業路線に適用するケーススタディによって妥当性の検証を行っているが、その対象期間はわずか3日間(三連休)であった。これらのことから、中川らの分析結果の内的妥当性に関しては課題が残っていた。

ここで因果推論分野では、特に高い内的妥当性が得られる手法として、社会実験の一種であるランダム化比較試験(A/Bテスト)がある。例えば文献5では、因果推論手法ごとに得られる結果の内的妥当性の高さを1a, 1b, 2a, 2b, 3, 4の6段階で評価を行っており、最上位の1aには複数のランダム化比較試験(A/Bテスト)の結果を比較したメタアナリシス、次点の1bにはランダム化比較試験(A/Bテスト)と位置付けている。なお、「ランダム化比較試験」と「A/Bテスト」は本質的には同じ手法を指すが、本稿では以降の記述との整合性や可読性を重視し、以下「A/Bテスト」を用いる。

A/Bテストとは、検証したい要素だけが異なる2つのA/Bパターンを用意し、対象(人単位や日単位)をランダムに2つのA/Bグループに振り分け、それぞれのパターンを適用して結果を比較する社会実験の手法である。一方のグループ(比較群)には従来通りの施策を適用し、もう一方のグループ(介入群)には新しい施策を適用し、両グループの目的変数(売上高などのKPI)を比較することで、施策の効果を統計的に評価するという形式が一般的である。対象をランダムに振り分けることで、グループ間の均質性が担保され、結果の差が純粋な施策の効果であると統計的に判断することが可能となる。

しかし、筆者の知る限り、鉄道のイールドマネジメントを対象としたA/Bテストの公表事例はない。その上、A/Bテストの実施・分析事例は、Webサイトを活用したオンライン上でのものを中心に多く報告されている一方で、適切なA/Bグループ振り分けや、ケースバイケースで適切な統計指標を用いないと偽陰性・偽陽性の結果を得てしまうといった落とし穴があることも報告されている^{例えば6)}。

そこで本研究は、A/Bテストによるイールドマネジメントの効果検証方法を提案する。さらに、鉄道事業者協力の下、実営業路線においてA/Bテストを実施し、イールドマネジメントの効果検証を行うことにより、提案方法の有効性を確認するとともに、割引商品の発売枠の増減が運輸収入に与える影響のメカニズムを定量的に評価できることを実証する。

2. A/Bテストの実施・分析方法の提案

本章では、適切なA/Bグループの振り分け単位を示し、その振り分け方法で効果検証を行えることの数学的な証明を与えた後に、A/Bテストによる効果検証の具体的な実施手順を示す。

2.1 A/Bグループの振り分け単位

適切なA/Bグループの振り分け単位を決定するにあたっての要件は、非干渉性の仮定を満たすことであり、これはA/Bグループ間での目的変数が互いに影響しあわずに独立であることが求められる。

本研究で対象とする区間での割引商品の発売枠設定の最小単位は列車単位であり、すなわち「A月B日の列車I号の発売枠は10」といったように枠設定を行うことができる。表1に、列車単位と日単位、それぞれの場合におけるA/Bグループの振り分けをランダムに行ったイメージを示す。ここでは、A/Bパターンの発売枠数をそれぞれ10枠/20枠としている。

表1を基に、列車単位と日単位どちらでのA/Bグループ振り分けが適切かを検討する。まず列車単位でのA/Bグループ振り分けがなされた状況下で、ある旅客が列車3号の割引商品を購入したかったけれども、既に発売枠に達していて購入できなかった状況を想定する。このとき、その旅客がとる行動として前後列車(1号と5号)の割引商品を購入する列車転移が考えられ、需要がA/Bグループ間で干渉してしまっ

表1 列車単位（左）・日単位（右）でのA/Bグループ振り分けイメージ

列車	A/Bグループ (ランダム振り分け)	割引商品の発売枠	日付	A/Bグループ (ランダム振り分け)	割引商品の発売枠
1号	B	20	10月12日	B	全列車一律20
3号	A	10	10月13日	A	全列車一律10
5号	B	20	10月14日	B	全列車一律20
7号	A	10	10月15日	A	全列車一律10
...

いるため、列車単位でのA/Bグループ振り分けは、非干渉性の仮定を満たさない。次に、日単位でのA/Bグループ振り分けがなされた状況下で、10月13日において、上記と同様な旅客が発生したと想定する。このとき、その旅客が一日の中で列車転移行動をとることは考えられるが、（後述する本研究が対象とする平常日において）旅行日自体を前後に変更する行動は少ないと想定されるため、日単位でのA/Bグループ振り分けは非干渉性の仮定を満たすと考えられる。

以上から、イールドマネジメント効果検証のためのA/Bテストの適切なA/Bグループ振り分け単位は日単位であると結論付ける。

2.2 A/Bテストの有効性の数学的証明

ある日にち d において、グループAと比較したグループBの介入効果（Treatment effect） τ_d を式(1)、複数のサンプル日に対する平均介入効果（Average treatment effect） τ^{ATE} を式(2)のとおり定義する。なお、介入効果の用語は、介入グループ（Treatment group）と比較グループ（Control group）への振り分けを前提とした用語であるが、一般的であるため本稿でも用いる。

$$\tau_d \equiv Y_d^B - Y_d^A \tag{1}$$

$$\tau^{ATE} \equiv E[Y_d^B - Y_d^A] \tag{2}$$

Y_d^G は、日にち d 、グループ $G \in \{A, B\}$ における需要や運輸収入等のKPIを表す。ここで、ある日にち d はグループAまたはグループBのどちらかにだけ属するため、 Y_d^A と Y_d^B の両方を実データとして観測することはできない「因果推論の根本問題」がある。したがって特定の日ごとの介入効果 τ_d を計測することは現実社会において不可能である。期待値で表されている τ^{ATE} は本研究で求めるイールドマネジメントによる効果そのものである。しかしランダム化をせずに、分析者がグループAとグループBを自由に振り分けると、必ずしも τ^{ATE} は得られないことを示す。分析者が自由に振り分けた状況下でグループAに属する日にちを $G_d = A$ 、グループBに属する日にちを $G_d = B$ と示す。するとグループAの結果の期待値と、グループBの結果の期待値の差 τ^G は式(3)のとおり表せる。

$$\begin{aligned} \tau^G &\equiv E[Y_d^B | G_d = B] - E[Y_d^A | G_d = A] \\ &= E[Y_d^B | G_d = B] - E[Y_d^A | G_d = A] + E[Y_d^A | G_d = B] - E[Y_d^A | G_d = B] \\ &= E[Y_d^B - Y_d^A | G_d = B] + \mathbf{E[Y_d^A | G_d = B] - E[Y_d^A | G_d = A]} \end{aligned} \tag{3}$$

式(3)の最後の式の第2項と第3項(太字部分)は自己選択バイアス(Self-selection bias)と呼ばれる。自己選択バ

イアスは、グループ B に属する日にちがグループ A に振り分けられた場合の KPI の期待値と、グループ A に属する日にちがグループ A に振り分けられた場合の KPI の期待値の差分を示す。この値は必ずしも 0 になるわけではないため、その意味で τ^G にはバイアスがかかっているといえる。またこの値はデータ観測数が増えてもゼロに収束する訳ではないことに注意が必要である。

式(3)の第 1 項 $E[Y_d^B - Y_d^A | G_d = B]$ は、グループ B に属する日にちの平均介入効果 (ATET; Average Treatment Effect on the Treated) であり、この値を τ^{ATE} と示す。 τ^G が τ^{ATE} と等しくなるための条件は、自己選択バイアスが 0、すなわち $E[Y_d^A | G_d = B] = E[Y_d^A | G_d = A]$ となることである。ここでグループ分けがランダムに行われた場合、 G_d の割り付けがランダムであり、 G_d で条件付けることは期待値計算に対して何ら追加的な情報をもたらさないことを意味するため、 G_d により条件付けられた条件付き期待値は、条件付きでない通常の期待値と等しくなる。そのため式(4)が成立し、式(3)と式(4)を組み合わせることで式(5)が成立するため、ランダム化により τ^G と τ^{ATE} は一致する。

$$\begin{aligned} E[Y_d^B | G_d = B] &= E[Y_d^B | G_d = A] = E[Y_d^B] \\ E[Y_d^A | G_d = B] &= E[Y_d^A | G_d = A] = E[Y_d^A] \end{aligned} \tag{4}$$

$$\begin{aligned} \tau^G &\equiv E[Y_d^B | G_d = B] - E[Y_d^A | G_d = A] \\ &= E[Y_d^B - Y_d^A | G_d = B] + E[Y_d^A | G_d = B] - E[Y_d^A | G_d = A] \\ &= E[Y_d^B - Y_d^A | G_d = B] \\ &= E[Y_d^B - Y_d^A] \\ &\equiv \tau^{ATE} \end{aligned} \tag{5}$$

2.3 A/B テストによるイールドマネジメント効果検証の実施手順

図 1 に、1 章での例のように通常商品と割引商品の 2 種類のみの商品があり、2.1 節での例のように発売枠設定の A/B パターンがそれぞれ全列車一律 10 枠/20 枠という想定での、A/B テストによる効果検証のイメージを示す。

図 1 中では、一日平均需要を目的変数として、通常商品では B グループの値から A グループの値を差し引いた値 (以下、「B-A 差分」と記述する) が -45 であることから、A パターンから B パターンに発売枠を拡大することで、45 人がダウンセルしたことを示す。一方、割引商品では B-A 差分の値が 60 で、先のダ

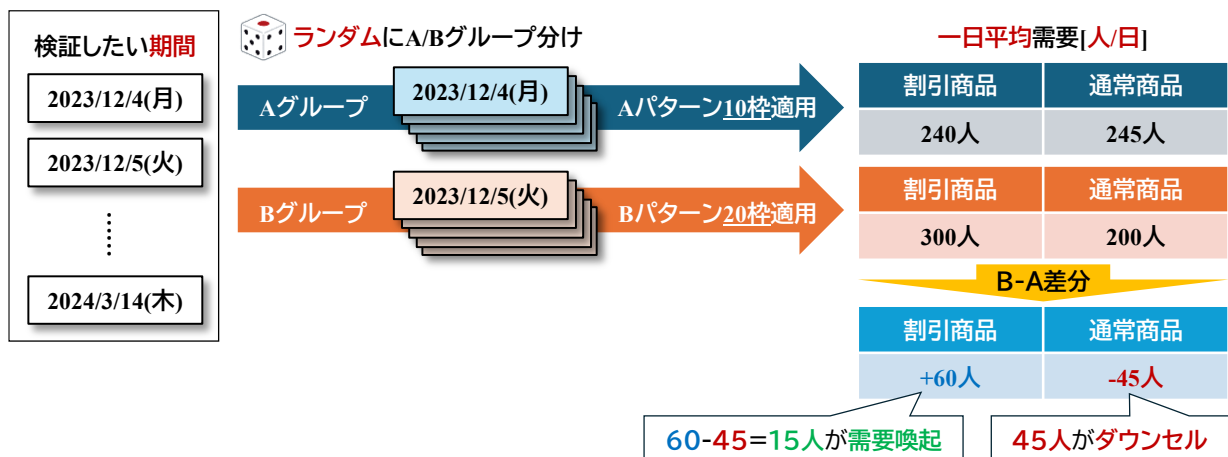


図 1 A/B テストによるイールドマネジメント効果検証のイメージ

ウンセル分 45 を差し引いた値が 15 であることから、発売枠を拡大することで、15 人の需要喚起ができたことを示している。なお、ここでの運輸収入の増減は、45 人のダウンセルによる減収分と、15 人の需要喚起による増収分の大小関係による。

A/B テストによるイールドマネジメントの効果検証の実施手順の概要を示す。まず、実営業路線の特性を踏まえて発売枠の A/B パターンを設定の上、検証期間を日単位で A/B グループ振り分けを行う（第 1 段階）。そして、実営業路線で A/B テストを実施した後に、A/B グループ振り分けが想定どおりランダムに行われたかどうかを確認する（第 2 段階）。ランダム化が無事に行われたことを確認できた後に、図 1 に示したような形で、日単位でのイールドマネジメントの効果検証を行う（第 3 段階）。最後に、時間帯単位でのイールドマネジメントの効果検証を行う（第 4 段階）。各段階の詳細は 3 章で記述する。

3. 実営業路線での A/B テストの実施と分析

本章では、A/B テスト実施の前提条件と要領を記述した後に、2.3 節で示した A/B テストによるイールドマネジメント効果検証の実施手順（第 1 段階から第 4 段階）に則り、実営業路線での A/B テストの分析結果を示す。1 章で既述のとおり、この分析を通じて、提案方法の有効性を確認するとともに、割引商品の発売枠の増減が運輸収入に与える影響のメカニズムを定量的に評価できることを実証する。

3.1 A/B テスト実施の前提条件と要領

地方自治体法に基づく大都市の中心駅（X 駅）と中核市の中心駅（Y 駅）を、2 時間程度で直通する特急列車の X 駅と Y 駅の双方向 OD を、A/B テストの実施・検証対象とした。本特急列車は、年に一度のダイヤ改正時に定められたダイヤを基に運行するため、定期列車の運行本数や発着時刻、列車ごとの車両編成数や指定席・自由席車両の編成数、車両定員数等は、遅延や車両故障等の異常が発生しない限りは毎日同じである（一部列車で平休日等の編成差はあるが、本分析は曜日区分ごとに行うため分析単位内では同一となる）。以上から、一日の中では列車単位で多少の条件は異なるが、日単位での条件は同一であるため、非干渉性の仮定を満たす。なお本 OD ペアを移動する主な手段としては、特急列車の他に高速バスや自家用車が想定される。また本稿では、X 駅から Y 駅への流動を下り、Y 駅から X 駅の流動を上りと表現する。

本 OD ペアの旅行商品は、窓口や券売機で購入する①定価商品と、インターネットで購入する商品がある。さらにインターネットで購入する商品には、②通常商品と③割引商品がある。発売枠が設けられているのは割引商品のみであり、A/B テストでは割引商品の発売枠を制御する。割引商品の価格（本稿の価格は全て運賃と特急料金の合計額を指す）は通常商品の約 2 割引であるのに対し、定価商品の価格（指定席基準）は通常商品の約 2 倍である。そのため定価商品のみインターネット購入ではない上に、価格も乖離して高い。以上から、割引商品の発売枠を制御する本 A/B テストにより変動する需要は、通常商品と割引商品のみと想定し、定価商品の需要には着目しない。また割引商品は指定席のみであるが、通常商品には指定席と自由席がある。ただし、通常商品の指定席と自由席の価格は同一であるため、基本的に旅客は指定席を利用し、自由席は A/B テストの影響を大きくは受けないことが予想される。

A/B テストの対象期間は、2023 年 12 月 1 日（金）から 2024 年 6 月 30 日（日）である。期間中の 2024 年 3 月にダイヤ改正があったが、定期列車の運行本数は変わらず、発着時刻も概ね同等であった。

A/B テストの実施要領は後述するが、発売枠は A パターンでは相対的に小さく、B パターンでは相対的に大きく設定し、以下、A パターンを基準に B パターンで発売枠を拡大したと捉えたときに、想定される結果を記述する。まず割引商品の発売枠が拡大されたことによって、指定席の通常商品の需要の一定数は、

割引商品にダウンセルすると想定される。一方、自由席の通常商品の需要は、指定席の通常商品と価格が変わらないにもかかわらず自由席を利用している層であるため、割引商品の発売枠が拡大されても特段変わらないか、もしくは一部の少ない割合が割引商品にダウンセルすると考えられる。また割引商品の発売枠が拡大されたことによって、割引商品の需要が喚起されていれば、全体の需要は増加し、特段喚起されていなければ、全体の需要は変動しないと想定される。ここで運輸収入がどのように増減するかは、ダウンセルによる減収分と割引商品の需要喚起による増収分のバランスによる。したがって、結果的にBパターンの運輸収入が大きければ、割引商品の発売枠の拡大によって需要の喚起が起きたと解釈でき、Aパターンの運輸収入が大きければ、割引商品の発売枠の縮小によって過度な安売りを防いだと解釈することができる。以上を踏まえて、表2に、A/Bテストの結果で想定される正負の符号を示す。

表2 A/Bテストで想定される結果の正負

割引商品の発売枠		需要の差分 (B-A)				運輸収入の差分 (B-A)
A	B	割引商品	指定席の通常商品	自由席の通常商品	全体	全体
小	大	+	-	0 or -	+ or 0	+ or 0 or -

A/Bテスト実施前まで鉄道事業者は、需要の特性に応じて、各日にちを平日（月曜から木曜の非休日）、休日前（主に金曜）、土曜、日曜、中型連休（主に三連休）、最繁忙期（上記期間中では2023年12月29日～2024年1月4日の年末年始と2024年4月27日～5月6日のGW）に区分し、発売枠の設定数を決定してきた。本研究でもこの曜日区分に併せてA/Bテストの検証を行う。ただし中型連休や最繁忙期は予め大きな需要が見込まれるため対象とはせず、A/Bテストの検証対象は平日、休日前、土曜、日曜の4曜日区分とした。またX駅周辺に大規模なイベントが開催できる施設があり、土曜・日曜を中心に人気アーティストによるコンサートなどの大規模なイベントが催されることがある。そのような日程はA/Bグループ分けに関係なく、発売枠をA設定とする運用を行っているため、A/Bテストの検証対象から除外している。表3に、枠設定数と検証日数をA/Bパターン別に示す。

本研究で使用するデータは、鉄道事業者が保持する発券実績データである。表4に、曜日区分・方向・商品別の記述統計を示す。表4から、最も多く日数を確保できた平日106日間と比較すると、土曜・日曜では大規模なイベントが催されやすい影響もあり、それぞれ18日間・19日間と少なくなりましたが、先行研究の3日間（三連休）と比較すると多く確保できていることが分かる。

表3 A/Bテストの条件と日数

曜日区分	一列車あたり発売枠数		検証日数	
	A	B	A	B
平日	10 ^{*1}	20	55	51
休日前	10 ^{*1}	20	15	16
土曜	20 ^{*2}	30	10	8
日曜	20 ^{*2}	30	11	8

※1: 一部臨時列車のみ15

※2: 一部臨時列車のみ30

表4 記述統計

曜日区分	方向	需要[人/日]								対象日数	除外日数 [*]
		割引商品		指定席の通常商品		自由席の通常商品		全体			
		平均値	SD	平均値	SD	平均値	SD	平均値	SD		
平日	下り	289.4	72.4	219.9	68.7	61.1	20.1	570.4	116.9	106	1
	上り	281.1	59.8	203.9	56.5	64.3	18.6	549.3	96.8		
休日前	下り	349.7	98.8	258.7	75.2	96.2	30.9	704.6	111.8	31	1
	上り	380.9	95.2	306.2	77.0	113.1	27.3	800.2	94.4		
土曜	下り	516.3	55.2	251.1	58.7	76.6	18.6	844.1	96.1	18	7
	上り	510.7	47.4	259.9	60.0	128.7	41.2	899.3	103.5		
日曜	下り	525.3	72.2	304.4	70.5	124.6	41.8	954.3	102.3	19	7
	上り	556.4	61.9	244.4	49.1	99.5	27.2	900.3	97.2		

※: 大規模イベントによる除外日数を示す。その他中型連休・最繁忙期であることによる除外日数は22

3.2 ランダム化の確認

A/Bグループ振り分けがランダムに行われたかの確認（第2段階）が必要な理由は、例えば大都市にあるX駅では分析者にとって把握が困難なイベント等による特需が発生する場合に、その特異日がA/Bグ

ループいずれかに偏ると、発売枠制御によるイールドマネジメントの効果を抽出できなくなってしまうためである。そこで本研究では、適切にランダム化が行われたことを確認するために、A/B テスト実施期間中において、X 駅-Y 駅間の割引商品の発売枠制御の影響を受けない共変量の平均値が、A/B グループ間で統計的有意差が確認されないことを示す。

重要な共変量として、所要時間や運行本数等の運行体系が挙げられるが、運行体系が毎日同じであることは3.1節で既述したとおりである。その他の共変量の例として、暦配列、気温や天候等の気象条件、沿線地域でのイベント状況などが挙げられるが、膨大かつ多岐に渡るため、それら全てを網羅的に検証することは現実的には困難である。加えて、イベント状況に関しては、一企業の社員旅行などの個別の情報を完全に把握することは不可能といえる。これらの限界を補う共変量として、割引商品の発売枠制御を行っていない同一線区内の OD 需要がある。なぜなら、観測可能なものに加えて観測不可能なものを含めた要因による偏りが、X 駅-Y 駅間の需要に影響を与えた場合、同じ偏りが他の OD 需要にも表れると考えられるためである。

以上を踏まえ、本研究で適切にランダム化が行われたかを確認する共変量は、X 駅-Y 駅と同一路線内で、割引商品が存在しない OD の内、需要の大きい上位3 OD ペア（合計6 OD）における需要とした。それらを対象に、「グループ B の平均値からグループ A の平均値を差し引いた値（以下、B-A 差分）が0である」という帰無仮説の下、二標本 t 検定（両側検定）を行う。表5にその結果を示す。表5から、自由席の通常商品の土曜・日曜の一部 OD で、有意水準 10%の下で有意差が見られるが、それ以外の全ての商品・OD では、有意水準 10%の下でも有意差は確認されない。なお有意水準 5%の下では、全ての商品・OD で有意差が確認されない。したがって、A/B グループの振り分けがランダムに行われたかを確認する意味で、比較的良好な結果が得られたといえる。

表5 ランダム化の確認の結果

曜日 区分	OD	方向	需要[人/日]					
			指定席の通常商品			自由席の通常商品		
			A	B	B-A差分	A	B	B-A差分
平日	X駅-I駅	下り	115.9	110.0	-5.9	37.8	38.4	0.6
		上り	100.3	96.5	-3.8	42.2	44.0	1.8
	X駅-J駅	下り	70.3	69.9	-0.5	15.2	16.0	0.8
		上り	63.0	60.8	-2.3	17.3	18.2	0.8
	X駅-K駅	下り	64.0	66.4	2.4	18.6	17.5	-1.1
		上り	59.3	58.2	-1.1	18.2	18.6	0.5
休日前	X駅-I駅	下り	136.1	150.8	14.7	48.9	51.1	2.1
		上り	119.5	125.4	5.9	59.6	65.3	5.7
	X駅-J駅	下り	84.3	84.1	-0.3	18.4	21.7	3.3
		上り	82.0	77.3	-4.8	25.3	29.1	3.9
	X駅-K駅	下り	83.1	83.1	-0.1	22.6	21.8	-0.8
		上り	71.4	70.7	-0.7	26.4	26.2	-0.2
土曜	X駅-I駅	下り	200.7	195.9	-4.8	65.5	57.8	-7.8
		上り	149.4	145.0	-4.4	84.3	78.0	-6.3
	X駅-J駅	下り	90.3	79.0	-11.3	18.3	15.9	-2.4
		上り	78.4	77.8	-0.7	36.3	28.1	-8.2 *
	X駅-K駅	下り	92.5	93.9	1.4	19.2	17.3	-2.0
		上り	81.4	81.1	-0.3	33.6	27.0	-6.6
日曜	X駅-I駅	下り	169.3	153.4	-15.9	52.8	42.9	-9.9 *
		上り	139.5	131.0	-8.5	78.2	64.3	-13.9 *
	X駅-J駅	下り	78.5	76.0	-2.5	14.5	14.8	0.3
		上り	64.7	66.1	1.4	25.9	20.1	-5.8 *
	X駅-K駅	下り	79.6	76.9	-2.8	16.4	16.6	0.3
		上り	73.4	77.1	3.8	24.2	21.6	-2.6

それぞれ有意水準 ***: 1%、**: 5%、*: 10%の下で有意であることを示す

3.3 A/B テスト：日単位

表6に、X 駅-Y 駅間の日単位の A/B テストの検証結果（第3段階）を示す。なお表2に示したとおり、運輸収入の B-A 差分は正と負の両方が想定される一方で、需要は区分によって正か負がいずれかに偏ると想定されるため、表6の二標本 t 検定は、運輸収入に対しては両側検定、需要は片側検定による結果を示す。

曜日区分によらない、全体的な傾向の解釈を行う。割引商品の B-A 差分では、ほぼ全ての区分で、有意水準 1%の下で有意に正の値で推定されており、指定席の通常商品の A/B グループの差分では、全ての区分で、有意水準 1%の下で有意に負の値で推定されている。この結果から、仮説どおり、発売枠を拡大する

ことによって、指定席の通常商品から割引商品へのダウンセルが起こったことが、統計的に有意に確認された。自由席の通常商品の B-A 差分では、一部の例外を除いては、有意水準 10% の下であっても有意差は確認されないことから、割引商品の発売枠の制御は特段自由席の需要には影響を与えない仮説と整合的な結果が得られた。

表 6 A/B テストの結果

曜日区分	方向	需要[人/日]										運輸収入[円/日]	
		割引商品			指定席の通常商品			自由席の通常商品			全体	全体	
		A	B	B-A差分	A	B	B-A差分	A	B	B-A差分	B-A差分	B-A差分	
平日	下り	249.6	332.3	82.7 ***	250.6	186.8	-63.9 ***	62.3	59.8	-2.5	16.3	+	
	上り	243.7	321.5	77.8 ***	229.7	176.1	-53.6 ***	66.7	61.7	-4.9 *	19.3	+	
休日前	下り	256.7	437.0	180.3 ***	310.9	209.8	-101.1 ***	90.4	101.6	11.2	90.4 **	+	
	上り	287.7	468.3	180.6 ***	362.2	253.8	-108.5 ***	107.9	118.0	10.1	82.3 ***	+	
土曜	下り	495.9	541.9	46.0 **	288.2	204.8	-83.5 ***	81.9	70.0	-11.9	-49.4	-	
	上り	488.2	538.9	50.7 ***	302.5	206.6	-95.9 ***	140.4	114.1	-26.3	-71.5	- *	
日曜	下り	478.1	590.3	112.2 ***	347.7	244.8	-103.0 ***	127.5	120.6	-6.8	2.4	-	
	上り	522.0	603.6	81.6 ***	270.5	208.5	-62.0 ***	97.1	102.9	5.8	25.4	+	

それぞれ有意水準 ***: 1%、 **: 5%、 *: 10%の下で有意であることを示す。データ秘匿性の観点から運輸収入は値の正負のみを示す

曜日区分ごとの解釈を行う。まず平日および日曜では、全体需要の B-A 差分は、いずれにおいても正の値が得られており、この点だけに着目すると、割引商品の発売枠を拡大することによって、需要が喚起されたといえる。しかし、いずれにおいても有意水準 10% の下でも有意差は確認されなかった。次に休日前では、全体需要の B-A 差分が、有意水準 1% と 5% で有意差が正の値で観測されている。これは、割引商品の発売枠を拡大することによって、需要が喚起されたといえる。一方で運輸収入の B-A 差分は正の値で観測されつつも、有意水準 10% の下でも有意差が確認されなかった。最後に土曜では、全体需要と運輸収入の B-A 差分が、ともに負の値で観測された。割引商品の発売枠を拡大することによって、全体需要が減少する関係は想定されない。したがって、土曜では暦配列やイベント等の発売枠以外の要因によって、全体需要の B-A 差分が負の値という非現実的な結果となってしまったと考えられる。表 4 に記載したとおり、休日前・土曜・日曜は、平日と比較して需要が大きいにもかかわらず、検証日数が少ないため、平日と比較すると結果が不安定となることが懸念される。また需要には季節変動や月次変動があり、さらに同じ月の中でも月初と月末で傾向が違うことが想定される。しかし平日以外の曜日区分は、一ヶ月ごとにそれぞれ 4 日間程度しかないため、偶然需要が大きい日がグループ A に集中すれば、上記結果のような事象は発生し得る。

日単位の分析では、時間帯によって異なる需要の傾向を平均化してしまうため、例えば朝はダウンセルが多い一方で正午前後は需要喚起が上回るなどのような傾向があったときに、本来は発生していた効果を捉えられない可能性がある。そこで次節では、より詳細な時間帯単位での分析を行うことで、一日の中での運輸収入の増減を検証する。

3.4 A/B テスト：時間帯単位

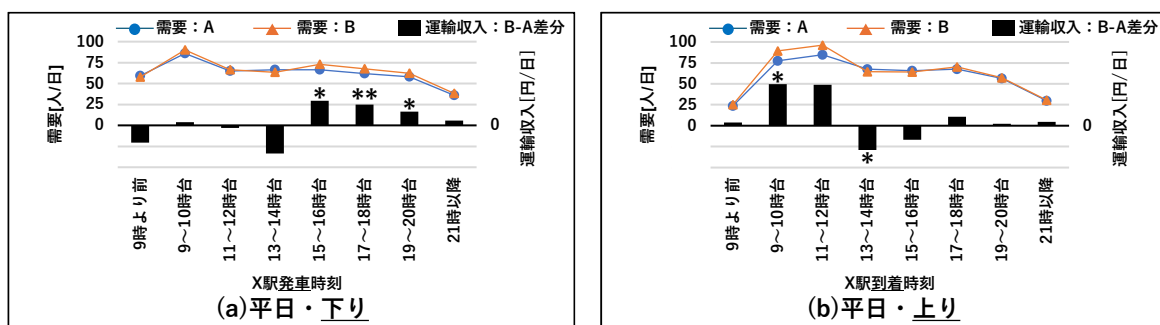
検証日数を最も多く確保でき、日単位の結果も安定している平日を対象とし、3.3 節よりも詳細な時間帯単位での A/B テストの検証を行う（第 4 段階）。ただし、このように A/B グループの振り分け単位（日にち）と分析単位（列車）が異なるときの有意差検定は、二標本 t 検定ではなく、クラスター頑健標準誤差に基づく必要があることが知られる^{（例えば 6）7）}。本研究では式(6)に示す回帰式のパラメータ推定に基づいて、クラスター頑健標準誤差に基づく有意差検定の結果を示す。

$$y_{i,d} = \beta_1 + \sum_{t=2}^T \beta_t S_t(i) + \tau_1 B(d) + \sum_{t=2}^T \tau_t S_t(i) \cdot B(d) + \varepsilon_{i,d} \quad (6)$$

ここで添え字 i は列車, d は日にち, t は区分数 T の時間帯の id (通し番号) を示す. $y_{i,d}$ は日にち d の列車 i の KPI, $B(d)$ は日にち d がグループ B であるときに 1 を取るダミー変数, $S_t(i)$ は列車 i が時間帯 t に属するときに 1 をとるダミー変数, $\varepsilon_{i,d}$ は誤差項, β と τ はパラメータである. 以上から時間帯 t における A/B グループの KPI に統計的有意差があるかの検定は, 「 τ_t が 0 に等しい」を帰無仮説とした有意差検定である.

図 2 に, 時間帯を {9 時より前, 9 時台~10 時台, ..., 19 時台~20 時台, 21 時以降} の 8 区分とした際の, A/B グループにおける需要の値, および運輸収入の B-A 差分の値と統計的有意差検定 (両側検定) の結果を示す. 平日・下りでは, 図 2(a) から 15 時台~16 時台, 17 時台~18 時台, 19 時台~20 時台の連続した 3 区分の時間帯で, 運輸収入の B-A 差分が, 有意水準 5%~10% で統計的有意に正の値が得られていることより, B パターンの枠設定が運輸収入増加の観点で推奨されることが分かる. また同時間帯の需要は, いずれも B パターンが A パターンよりも大きいことが分かる. したがって, この時間帯における運輸収入増加は, B パターンの枠設定とすることにより, ダウンセルによる旅客単価の低下を上回る規模の需要喚起が生じていることに起因するといえる. 平日・上りでは, 図 2(b) から 9 時台~10 時台で運輸収入の B-A 差分が有意水準 10% で統計的有意に正の値が得られた一方, 13 時台~14 時台で運輸収入の B-A 差分が有意水準 10% で統計的有意に負の値が得られていることより, それぞれ B パターン, A パターンの枠設定が推奨されることが分かる. またそれら時間帯の需要は, 9 時台~10 時台では B パターンの方が大きい一方で, 13 時台~14 時台では A/B パターンが概ね同等であることが分かる. したがって, 9 時台~10 時台では, 平日・下りと同様に, 発売枠を拡大することにより, 旅客単価の低下を上回る規模の需要喚起が見られた一方で, 13 時台~14 時台では発売枠を縮小することにより, ダウンセルを抑制できたことが分かる.

以上の分析から, 日にち単位では, 全体需要, 運輸収入ともに統計的有意差が見られなかった平日でも, 時間帯単位ではクラスター頑健標準誤差に基づく統計的有意差が見られることが分かった. これは一日一律の枠設定ではなく, 時間帯別や列車別に発売枠の設定数を変える施策が有効であることを示唆する結果である. 具体的には, 発売枠の拡大がダウンセルによる減収を上回る需要喚起をもたらす運輸収入が増加する時間帯がある一方, 発売枠の縮小がダウンセルを抑制し運輸収入が増加する時間帯があるなど, 異なるメカニズムが働いていることを示した. 以上から, 提案方法に則った実営業路線での A/B テスト実施を通じて, 割引商品の発売枠の増減が運輸収入に与える影響のメカニズムを定量的に確認できることを実証した.



それぞれ有意水準 ***: 1%, **: 5%, *: 10%の下で有意であることを示す. データ秘匿性の観点から運輸収入の縦軸の目盛値は0のみを示す

図 2 時間帯別の A/B テスト結果

4. おわりに

本研究では、割引商品の発売枠制御に着目し、A/B テストによるイールドマネジメントの効果検証方法を提案した。さらに、提案方法に則った実営業路線での A/B テストの実施と効果検証を通じて、提案方法の有効性を確認するとともに、割引商品の発売枠の増減が運輸収入に与える影響のメカニズムを定量的に評価できることを実証した。得られた結果の特徴として、時間帯によって適切な発売枠の設定戦略が異なることを実証した点にある。具体的には、発売枠の拡大が旅客単価の低下を上回る需要喚起を通じて増収をもたらす時間帯と、逆に発売枠の縮小がダウンセルの抑制を通じて増収をもたらす時間帯が存在することを示した。この結果は、鉄道事業者が時間帯別のイールドマネジメント戦略を策定する上での重要な意思決定支援情報となり得る。

今後の課題として、平日と比較して十分な検証日数の確保が困難な土曜・日曜で、より頑健な結果が得られる補正方法を開発することが挙げられる。なお本研究の限界として、A/B テストにより高い内的妥当性を確保できている反面、実験を行っていない期間や場所に結果を援用できるかの外的妥当性の検証ができない点が挙げられる。この点は、A/B テストの実施を継続し、また他の線区にも拡張、分析を更新していく必要がある。

以上のような課題を残しつつも、交通経済学のイールドマネジメントにおいて未解明であった割引商品の発売枠の制御が需要と運輸収入に与えるインパクトを実証した点で、本研究には学術的な寄与がある。また本研究は日本の鉄道に A/B テストを実施した初の公表事例であり、適切な A/B テストを行うための A/B グループの振り分け方法や、A/B グループの振り分け単位（日単位）と分析単位（列車）が異なるときの分析方法を示した点で実務的な寄与もある。本稿が、今後の鉄道等の交通分野での A/B テストの実施マニュアルとしての役割を担うことを期待する。

本稿は、『交通学研究』で発表した内容⁸⁾を基に、実務的含意を詳説したものである。

謝 辞

A/B テストの実施と本稿の完成は、鉄道事業者の多大なるご協力があったて実現した。厚く御礼申し上げる。

文 献

- 1) McGill, J. I., and Van Ryzin, G. J., "Revenue management: Research overview and prospects", *Transportation science*, Vol. 33, No. 2, pp. 233-256, 1999.
- 2) 佐藤公俊, 沢木勝茂: レベニューマネジメント: 収益管理の基礎からダイナミックプライシングまで, 共立出版, 2020
- 3) NAKAGAWA, S., SUZUKI, T., MATSUMOTO, R., FUKASAWA, N., and OZAKI, N., "Yield Management Method Focused on Discount Ticket Sales Quota", *Quarterly Report of RTRI*, Vol. 61, No. 4, pp. 261-266, 2020.
- 4) 中川伸吾, 鈴木崇正, 松本涼佑, 深澤紀子, 尾崎尚也: 割引商品の発売上限数に着目したイールドマネジメント手法, 鉄道総研報告, Vol. 34, No. 2, pp. 11-16, 2020
- 5) 内閣府 政策統括官 経済社会システム担当: 海外における E B P M の事例とエビデンス集作成について, 2020
- 6) Kohavi, Ron, Diane Tang, and Ya Xu. "Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing", Cambridge University Press, 2020. (大杉直也訳: A/B テスト実践ガイド: 真のデータドリブンへ至る信用できる実験とは, ドワンゴ, 2021)
- 7) 川口康平, 澤田真行: 因果推論の計量経済学, 日本評論社, 2024
- 8) 松本涼佑: 鉄道の割引商品の発売枠制御によるイールドマネジメントの効果検証: A/B テストに基づいて, *交通学研究*, No. 6, pp. 33-40, 2025