

大規模アンケートデータ×分析手法の紹介

2022/9/7

エム・アール・アイリサーチアソシエイツ株式会社

データサイエンス事業部 池田 陽平

政策立案や商品・サービス開発をする際、受け手となる消費者の実態を把握することなしに企画を進めることはまずありません。実態把握には様々な手法がありますが、大規模な時系列アンケート調査結果は様々な分析に活用できる点で、今なお有効です。

まず、サンプル数が多いことで、詳細な実態が把握できます。例えば、年間で5回以上デイズニードランドを利用するヘビーユーザーの割合や、北海道在住の30台女性といったユーザーの属性・層別の活動実態の把握などが可能になります。

時系列データがあれば、経年変化を捉えられるだけでなく、世の中の変化が時代の潮流によるものか、あるいは世代ごとに異なる価値観によるものかを明らかにすること（コーホート分析）もできます。例えば、若者の車離れが時代の影響か世代の影響かといったこともわかります。さらに、価値観や生活行動に関する多様な調査項目があれば、それぞれの要素間の関係を多角的に分析することによって、例えば、幸福度と食事との関係を、仮説を持たずに明らかにすることも可能になります。

エム・アール・アイリサーチアソシエイツ（株）は、全国30,000人が約2,000問に回答している国内最大規模のアンケートパネル、[生活者市場予測システム（mif）](#)¹を運用しており、2011年の調査開始から今年で12年目となる時系列データを蓄積しています。今回の調査レポートでは「大規模アンケートデータ×分析手法」と題して、生活者市場予測システム（mif）を用いた大規模アンケートデータならではの分析例をシリーズでご紹介します。

第1回は3万人のサンプル数を活かした確率分布分析を通じて、テーマパークの利用者数を分析した結果を紹介します。第2回は時系列データを活かしたコーホート分析、第3回は2,000問の設問を活かした機械学習による分析を予定しています。

「大規模アンケートデータ×分析手法」の調査レポート（予定）

No.	概要	分析手法の観点
第1回	3万人データの確率分布分析	サンプル数を活かした頻度分析
第2回	3万人データのコーホート分析	時系列データを活かした世代効果分析
第3回	3万人データの機械学習分析	2000項目の設問を活かした探索的分析

¹ 生活者市場予測システム（mif）は、サービスの企画・開発を株式会社三菱総合研究所が、販売をエム・アール・アイリサーチアソシエイツ株式会社がそれぞれ行っています。<https://mif.mri.co.jp/>

生活者市場予測システム（mif）の設問項目一覧は以下のリンクから参照ください。

https://dl19w3jlhkm4w.cloudfront.net/2022/2022_basic_questions.pdf

第1回 3万人データの確率分布分析：テーマパークの利用者数予測モデル

ポイント

- ✓ 消費者全体のテーマパーク利用回数の分布は負の二項分布（NBD）で近似できる
- ✓ 利用者の利用回数の分布の形と平均利用回数はテーマパークや時期によらず概ね一定
- ✓ テーマパークの平均利用回数を増やしたければ利用者割合（浸透率）を高めるしかない
- ✓ テーマパークの年間利用者数は利用者割合（浸透率）から簡易予測が可能

キーワード：消費者の購買行動、確率分布、負の二項分布（NBD）、利用者割合（浸透率）、平均利用回数、テーマパークの利用者予測モデル

1 はじめに

第1回ではテーマパークの利用をテーマに、大規模なアンケートデータを用いて消費者の購買行動を確率分布の視点から分析しました。2022年11月には愛知県にジブリパークが開園予定、2025年開催予定の大阪・関西万博跡地利用の一つにエンターテインメントエリアが想定される²等、新たなテーマパーク開発が注目されています。その一方で、足元ではコロナ禍で減少した来園者数の回復や新たな顧客体験価値の創造が既存のテーマパークの課題となっています。

社会調査やマーケティング調査のアンケートでは数百から数千のサンプル数で調査されることが多いですが、出現率が低い人々の行動（例：1年間でディズニーランドを5回以上利用するヘビーユーザーの割合）を詳しく分析するためにはより多くのサンプル数を確保することが有効です。エム・アール・アイリサーチアソシエイツ（株）は、国内最大規模のアンケートパネルである生活者市場予測システム（mif）の全国32,000人（16～69歳）³のデータ（2020年6月及び2021年6月調査）を用いて国内のテーマパークの利用回数を分析しました。このデータからは、例えば「1年間でディズニーランドを5回以上するヘビーユーザー」は全体の1.2%に過ぎないことが分かりますが、そのサンプル数は380人と十分に確保することができます（2020年6月調査）。

本調査レポートでは消費者全体の利用回数の分布はテーマパークや時期によらず負の二項分布（NBD: Negative Binomial Distribution）で近似できることを示し、あわせて、その結果を用いてテーマパークの利用者数を簡易に予測するモデルを紹介します。

² 出典）大阪市「夢洲まちづくり基本方針」（2019年12月18日）

<https://www.city.osaka.lg.jp/osakatokei/page/0000473459.html>（2022年7月19日閲覧）

³ 20～69歳の30,000人（ベーシック調査）に加えて、2020年6月から調査を開始したティーンズ調査（16～19歳のインターネット利用者（有効サンプル約2,000人、ベーシック調査と同様に人口構成比で割付）とあわせた全国の16～69歳男女32,000人

2 分析結果

2.1 テーマパークの利用回数の分析

(1) テーマパークの利用者割合（浸透率）

まず、過去1年間に1回以上利用した人の割合を利用者割合（浸透率）とします。テーマパーク毎の利用者割合（浸透率）を図1に示します（2020年6月調査）。東京ディズニーランドが14.1%と最も多く、次いで東京ディズニーシー（11.8%）、ユニバーサル・スタジオ・ジャパン（7.8%）となっています。利用者割合（浸透率）は、当然ながらテーマパークによって異なり、また時期によっても変化します。本調査レポートの分析は原則として2020年6月調査を元にしてはいますが、2021年6月調査ではコロナ禍の影響でどのテーマパークも利用者割合（浸透率）は大きく減っています。

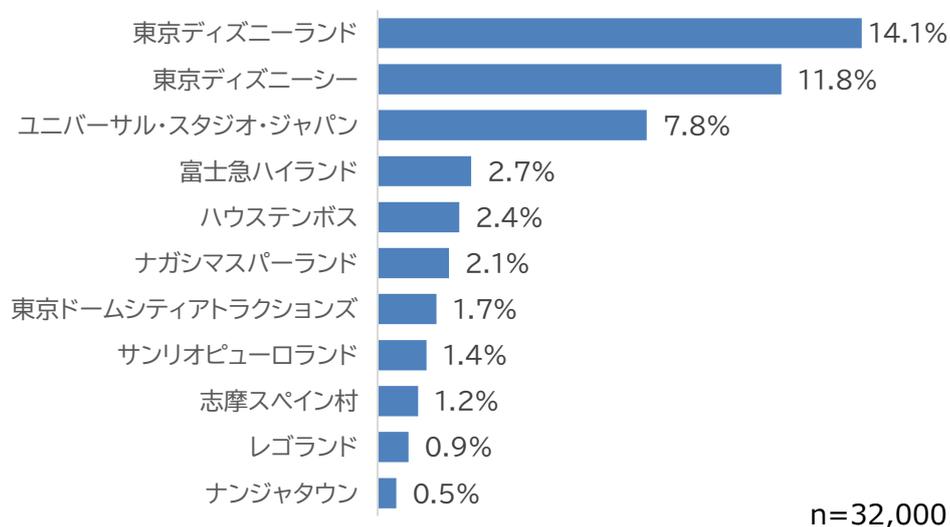


図1 過去1年間のテーマパークの利用者割合（浸透率）（2020年6月調査）

(2) テーマパークの利用回数の分布

消費者全体のテーマパーク利用回数の分布は負の二項分布（NBD）で近似できる

次にテーマパークの利用回数の分布をみてみます。例えば東京ディズニーランドでは非利用者（年0回）が85.9%、利用者が14.1%でそのうち年1回の利用者が9.1%、年2回が2.5%、年3回1.0%、年4回が0.3%、年5回以上が1.2%となっています（図2）。

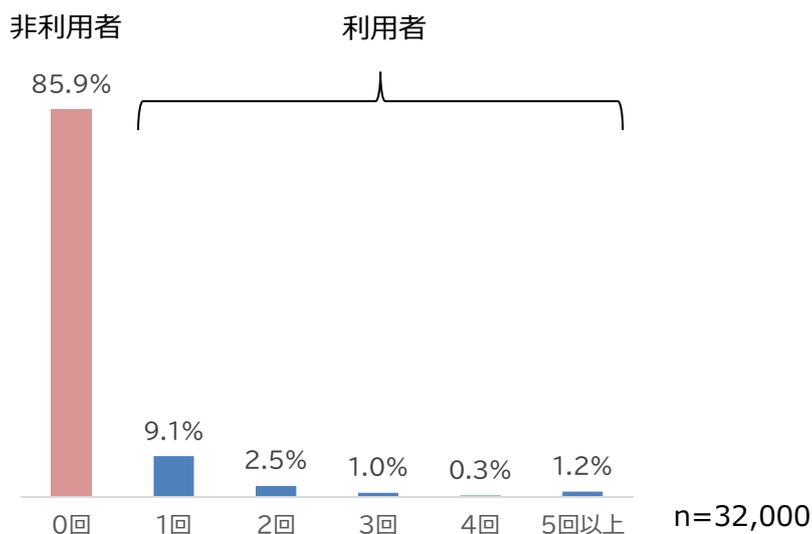


図2 過去1年間の東京ディズニーランドの利用回数の分布 (2020年6月調査)

一般的に、消費者全体の商品やサービスの購入・利用頻度は、負の二項分布 (NBD) に従う場合があることが知られていますが、図3は、図2の利用回数の分布のうち、年1回以上の利用回数分布を抜き出し、あわせて平均利用回数に基づく負の二項分布の予測値を追加したものです。これを見ると消費者全体の東京ディズニーランドの利用回数の分布は負の二項分布と似ています。負の二項分布は消費者個人の利用回数が一定期間内にある確率で起こると仮定していますから、このことは東京ディズニーランドの利用が消費者個人の確率的な行動の結果であることを示唆しています。

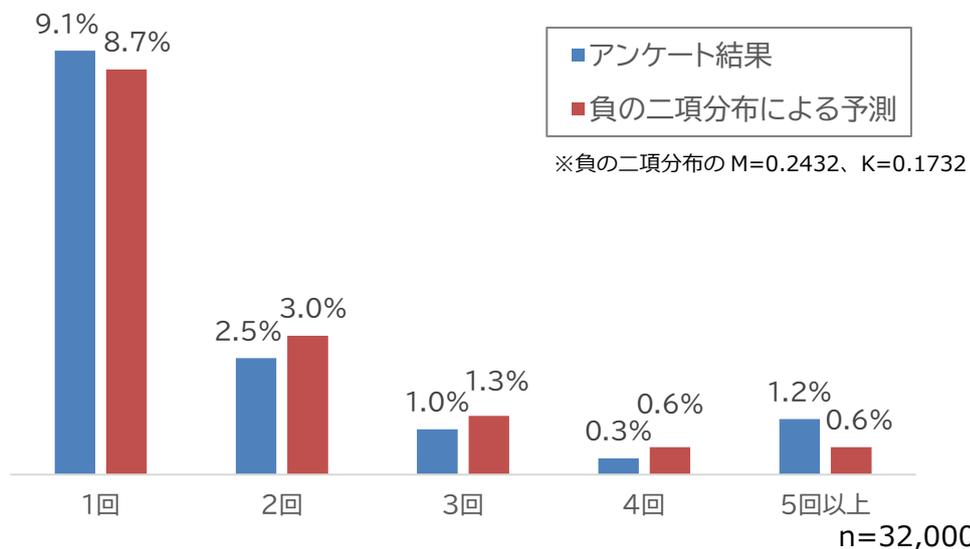


図3 過去1年間の東京ディズニーランドの利用1回以上の分布 (2020年6月調査)

負の二項分布 (NBD : Negative Binomial Distribution)

消費者個人の商品やサービスの購入回数がポアソン分布していてその長期的平均購入回数がガンマ分布しているとき、ある期間における消費者全体の購入回数は負の二項分布に従います。具体的には一定期間の消費者全体の購入回数を r 、一定期間に r 回購入される確率を P_r とすると、 P_r は以下のように負の二項分布で表されます。

$$P_r = \frac{\left(1 + \frac{M}{K}\right)^{-K} \cdot \Gamma(K + r)}{\Gamma(r + 1) \cdot \Gamma(K)} \cdot \left(\frac{M}{M + K}\right)^r$$

ここで M は一定期間の平均購入回数、 K は確率分布の形状を決める変数です。 M は全ての消費者が購入した述べ回数を消費者の数で割ったものです。 K は M と 0 回購入される確率 (P_0) から求まります。

出所：森岡毅、今西聖貴「確率思考の戦略論 USJ でも実証された数学マーケティングの力」P245～258、2016年、株式会社 KADOKAWA より作成

同様に、ユニバーサル・スタジオ・ジャパンと富士急ハイランド、ハウステンボスの年1回以上の利用回数分布と負の二項分布による予測値をそれぞれ図4、図5、図6に示します。いずれも利用回数の分布は負の二項分布に近いことが分かります。この傾向は今回調査した全てのテーマパークに当てはまり、どのテーマパークも利用回数の分布が負の二項分布 (NBD) で近似できることを示しています。

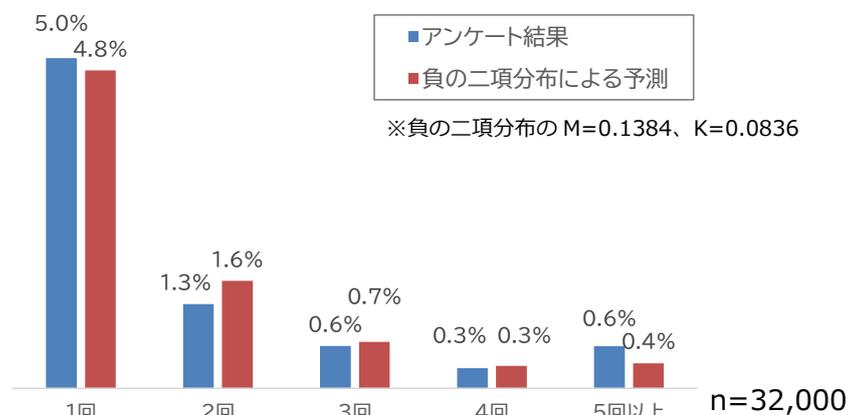


図4 過去1年間のユニバーサル・スタジオ・ジャパンの利用1回以上の分布 (2020年6月調査)

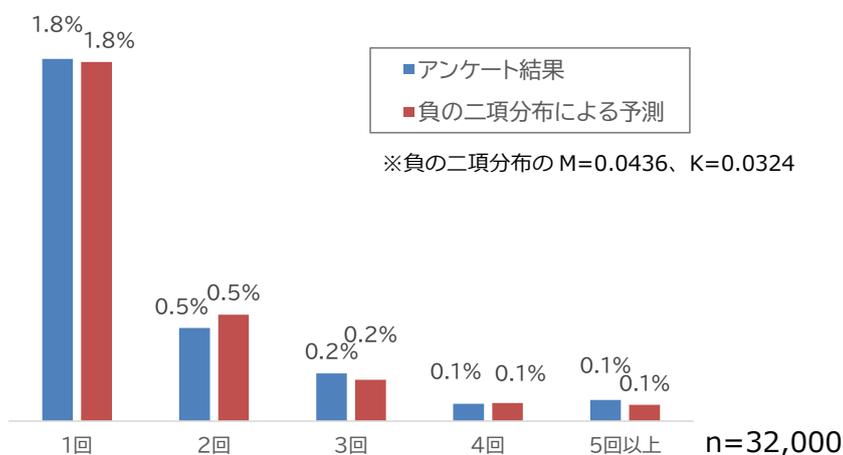


図5 過去1年間の富士急ハイランドの利用1回以上の分布 (2020年6月調査)

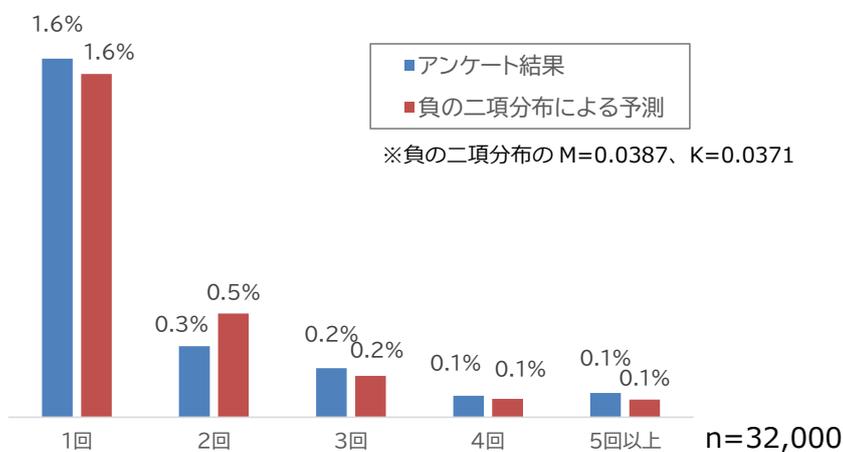


図6 過去1年間のハウステンボスの利用1回以上の分布 (2020年6月調査)

(3) テーマパークの平均利用回数

**利用者の利用回数の分布の形と平均利用回数はテーマパークや時期によらず概ね一定
テーマパークの平均利用回数を増やしたければ利用者割合（浸透率）を高めるしかない**

図3から図6を見ると、年1回以上の利用回数の割合はそれぞれのテーマパークによって異なりますが、驚くべきことにその分布の形はよく似ています。すなわち、どのテーマパークも利用者に限ると、そのうち平均して年1回の利用者が65%程度、年2回の利用者が16%程度、年3回の利用者が9%程度を占めています。利用者の平均利用回数⁴は、利用者割合（浸透率）に関わらずどのテーマパークも概ね1.6~1.8回となっています（表1）。洗剤やシャンプー等の生活必需品のブランドでは浸透率が上がるほど購入者の購買頻度も上がるという法則（ダブルジョパティの法則）⁵が知られていますが、テーマパークには当てはまらないようです。

⁴ ここでは平均利用回数M（非利用者を含む）と利用者の平均利用回数（非利用者を含まない）を区別しています

⁵ 出典）パイロン・シャープ、前平謙二 訳「ブランディングの科学」P44、2018年、朝日新聞出版

例えば東京ディズニーランドは利用者割合（浸透率）が最も高いですが、年に複数回利用する利用者の割合が特に多いわけではありません。このことは、平均利用回数を増やしたければ利用者割合（浸透率）を高めるしかないとも言い換えられます。逆に、利用者割合（浸透率）を上げずに年に複数回利用する利用者を増やすことは難しいこととなります。

表 1 テーマパークの利用者割合（浸透率）と平均利用回数（2020年6月調査）

n=32,000	利用者割合 (浸透率)	平均利用回数(M)	利用者の 平均利用回数
東京ディズニーランド	14.1%	0.24	1.73
東京ディズニーシー	11.8%	0.20	1.71
ユニバーサル・スタジオ・ジャパン	7.8%	0.14	1.77
富士急ハイランド	2.7%	0.04	1.60
ハウステンボス	2.4%	0.04	1.63
ナガシマスパーランド	2.1%	0.03	1.62
東京ドームシティアトラクションズ	1.7%	0.03	1.78
サンリオピューロランド	1.4%	0.02	1.72
志摩スペイン村	1.2%	0.02	1.49
レゴランド	0.9%	0.02	1.71
ナンジャタウン	0.5%	0.01	1.83

利用者の平均利用回数の差に注目すると、最も多いのはナンジャタウン（1.83回）で、最も少ないのは志摩スペイン村（1.49回）です。この違いには各テーマパークの立地地域の人口が関係していると考えられます。居住地域別の利用者割合（浸透率）を見ると、基本的に各テーマパークが立地する都道府県の居住者が最も高く、遠い居住者ほど低くなる傾向（距離抵抗）があります。図7と図8は大阪府に立地するユニバーサル・スタジオ・ジャパンの地域別の利用者割合（浸透率）を示したもので、大阪府在住者が17.7%と最も高く、大阪府からの距離が離れるほど低くなる傾向が見て取れます。

各テーマパークの立地都道府県人口と利用者の平均利用回数との間には相関があるため（図9）、人口が多い地域に立地していることが利用者のリピート率上昇に寄与していると考えられますが、それでもその差は利用者割合（浸透率）の差に比べれば大きくありません。

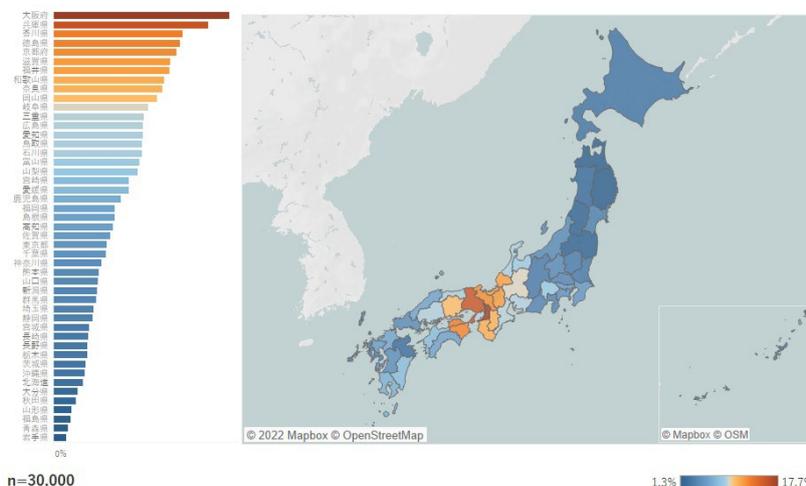


図7 都道府県別のユニバーサル・スタジオ・ジャパンの利用者割合（浸透率）（2020年6月調査）

以下の図 10 は表 1 に示した利用者割合（浸透率）と平均利用回数（M）の関係を示したものです。両者はほぼ線形に相関しており、テーマパークの利用者割合（浸透率）が増加すれば（利用者の平均利用回数がほぼ同じであるため）それに比例して平均利用回数（M）が増加することが分かります。

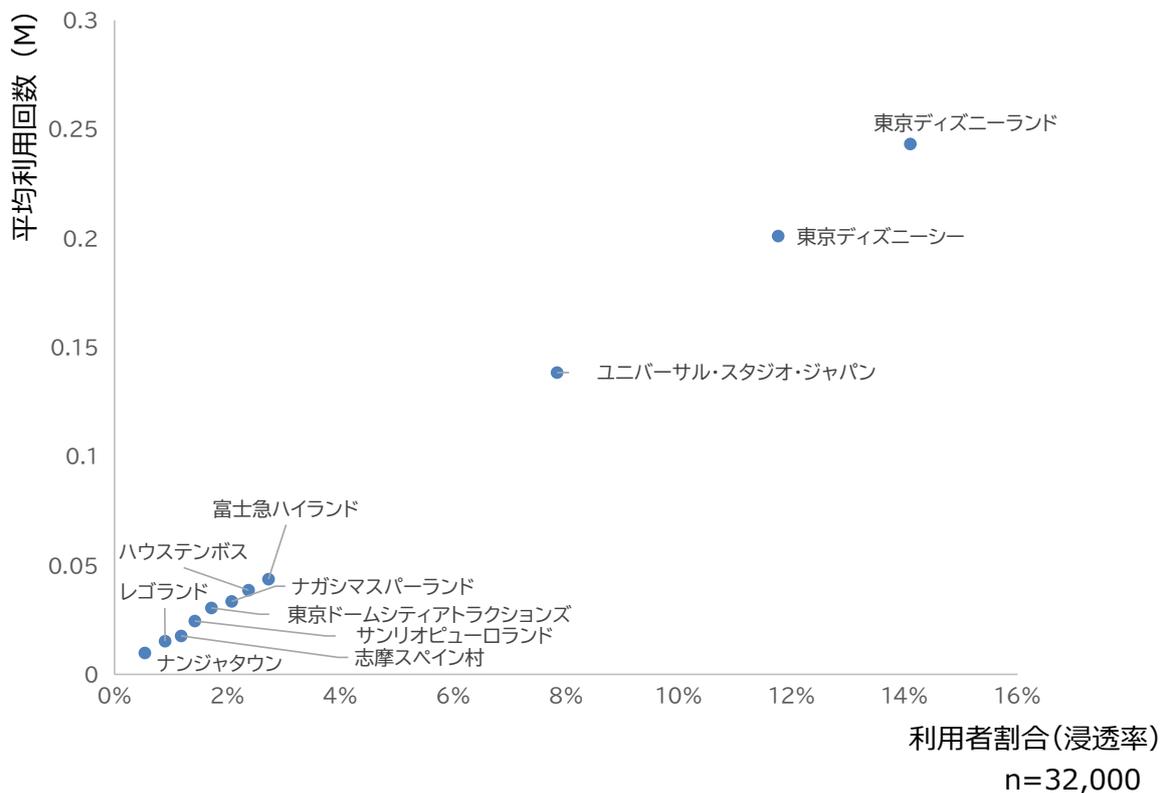


図 10 テーマパークの利用者割合（浸透率）と平均利用回数（M）の関係
（2020 年 6 月調査）

また、上記の傾向は時期を変えても変わりません。前述の通り 2021 年 6 月調査ではコロナ禍の影響でどのテーマパークも利用者割合（浸透率）は大きく減少しましたが、利用者割合（浸透率）と平均利用回数（M）は 2020 年 6 月調査のデータとほぼ同じ直線上で相関します（図 11）。

この結果は、時期によって利用者割合（浸透率）が変動しても、利用者の平均利用回数は大きく変わらないことを意味します。利用者の平均利用回数は、テーマパークや時期によらず一定になっているのです。

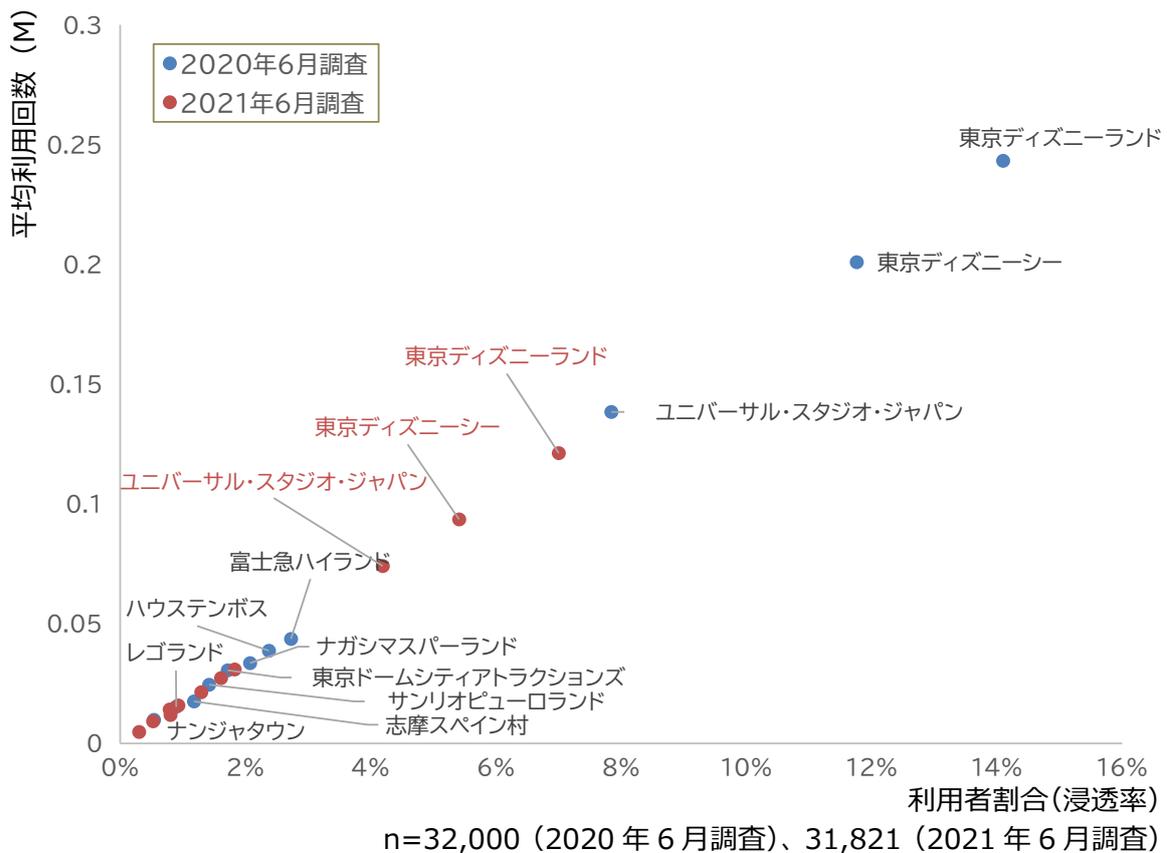


図11 2時点のテーマパークの利用者割合（浸透率）と平均利用回数（M）の関係

※図10に2021年6月調査を重ねたもの、2021年6月調査は上位3テーマパークのみラベル表記

2.2 テーマパークの利用者数予測モデル

テーマパークの年間利用者数は利用者割合（浸透率）から簡易予測が可能

テーマパークのような特定地域のサービス需要は、より詳細には地域別の平均利用回数と地域間距離を用いたモデル（一般には重力モデル）で予測しますが、全体（全国）の平均利用回数が分かっているれば、国内の利用者数は以下の式で計算できます。

国内の年間利用者数

$$= \text{人口} \times \text{利用者割合 (浸透率)} \times \text{利用者の平均利用回数}$$

$$\text{※利用者割合 (浸透率)} \times \text{利用者の平均利用回数} = \text{平均利用回数 (M)}$$

テーマパークについては2.1で示したように、利用者の平均利用回数はテーマパークや時期によらず概ね一定のため定数と仮定でき、上式は利用者割合（浸透率）さえ分かれば計算が可能である、すなわちテーマパークの年間利用者数の予測が可能であるといえます。

上式と 2.1 の結果及び国内 16～69 歳の人口⁸を用いると、テーマパークの利用者数予測モデルは以下のような単純な回帰式になります（図 12）。なお、本調査レポートではサンプル数を性・年代別・地域別インターネット利用人口の構成比に合わせているため、結果は近似的に全国の 16～69 歳男女の予測になっています。

テーマパークの年間利用者数予測モデル（万人）

$$= \text{人口（7,781）} \times \text{利用者割合（浸透率）} \times \text{利用者の平均利用回数（1.72）}$$

※2020年6月調査（図 10）の単回帰分析で作成、予測対象は 16～69 歳のみ

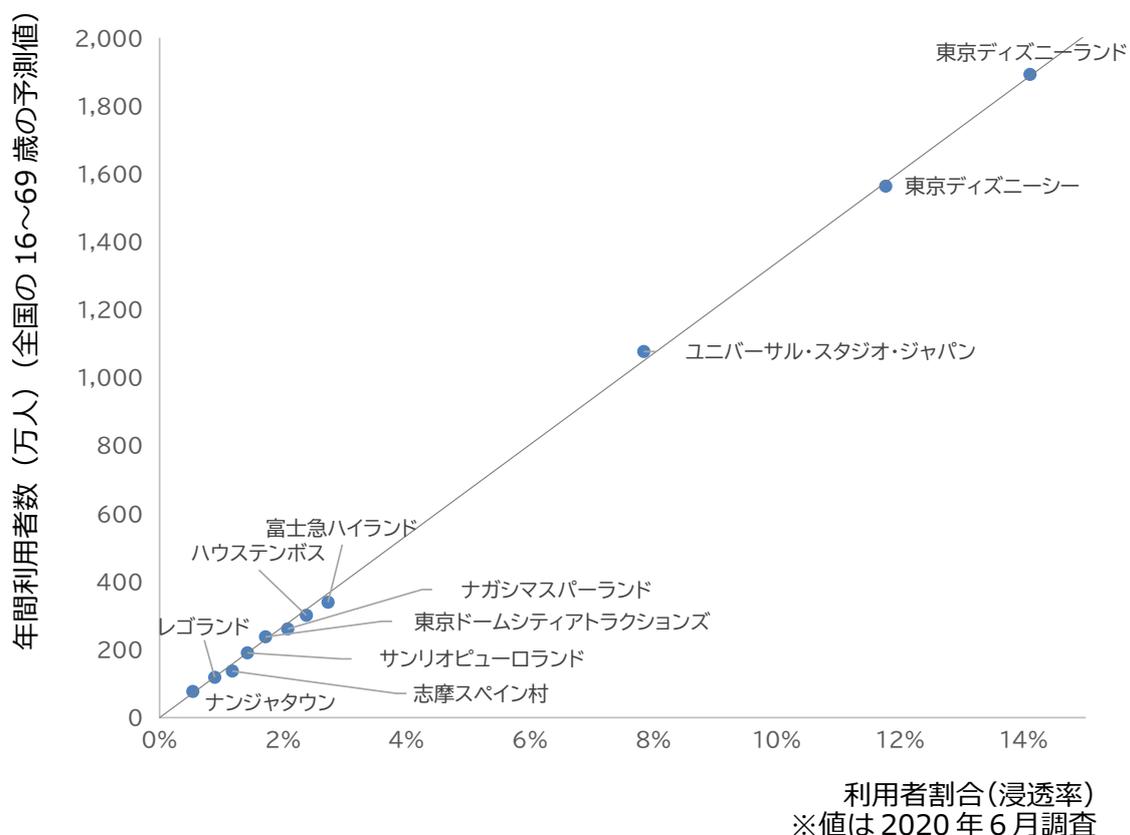


図 12 テーマパークの利用者割合（浸透率）と年間利用者数（予測値）の関係

なお、上記の予測モデルの結果は当然ながら実際の年間利用者数に対して誤差があります。例えばこのモデルでは、東京ディズニーランドの年間利用者は約 1,887 万人、東京ディズニーシーの年間利用者は約 1,574 万人と予測されますが（いずれも 2020 年 6 月調査の利用者割合（浸透率）による）、両テーマパークを合計した 2019 年度の入園者数の実績

⁸ 出所) 令和 2 年国勢調査 人口等基本集計 第 2 - 1 表

<https://www.stat.go.jp/data/kokusei/2020/kekka.html> より作成 (2022 年 7 月 19 日閲覧)

は2,900万人⁹となっており、時期の違いを踏まえても予測値はやや過大と考えられます。アンケートデータは様々な原因でバイアスが生じているので予測精度を高めるには実績値を用いた利用者の平均利用回数の補正を行う必要があります。なお、今回は本調査レポートの調査対象外である15歳以下と70歳以上は予測に含んでおらず、訪日外国人の利用者数も含んでいない点にも注意してください。

3 おわりに

本調査レポートでは、全国32,000人(16~69歳)のアンケート分析により、消費者全体の利用回数の分布はテーマパークや時期によらず負の二項分布(NBD)で近似できることを示しました。また、利用者の平均利用回数はテーマパークや時期によらず概ね一定であることが分かりました。これは、平均利用回数を増やしたければ利用者割合(浸透率)を高めるしかないということの意味しており、マーケティング戦略において重要な事実と言えるでしょう。最後に、それらの結果を用いてテーマパークの利用者数を利用者割合(浸透率)から簡易に予測するモデルを作成しました。

なお、同じアンケートパネルの他の調査項目を分析すると、負の二項分布(NBD)による利用回数の近似や利用者の平均利用回数が一定になる傾向はテーマパークだけでなく商業施設や文化施設でも成り立つ場合があることが分かっています。このため、利用者割合(浸透率)から利用者数を予測するモデルは、テーマパークや各種施設、イベント等の新規開発時の需要予測にも活用可能と考えられます。こうした施設、イベント等の新規開発に際して、個々人の平均利用回数の予測は難しいですが、利用意向はアンケート等により比較的容易に把握可能です。利用意向から利用者割合(浸透率)を推計すれば、あらかじめ類似の既存施設、イベント等の利用状況から作成したモデルを適用することで利用者数を簡易的に推計することができます。

本調査レポートでは[生活者市場予測システム\(mif\)](#)のサンプル数を活かし、テーマパークの利用回数を詳細に分析しました。次回以降は定点調査としてのデータを活かした時系列分析や、2000問の設問項目を活かしたライフスタイル分析を行い、紹介します。

本調査結果及び「生活者市場予測システム(mif)」に関する問い合わせ先

エム・アール・アイ リサーチアソシエイツ株式会社

データサイエンス事業部 担当：池田

TEL 03-6858-3526 / メール mif@mri.co.jp

URL: <https://mif.mri.co.jp/>

〒100-0014 東京都千代田区永田町二丁目10番3号 東急キャピトルタワー6階

⁹ 出典) 株式会社オリエンタルランド「FACT BOOK 2022」<http://www.olc.co.jp/ja/ir/library/factbook.html>
(2022年6月29日閲覧)