

過去の乗客数データを利用した日本の駅の乗降客数予測法の開発

TRAN Thanh Hai

正会員 戸田建設株式会社 (〒104-0032 東京都中央区八丁堀二丁目)

E-mail: tran.thanh.hai@toda.co.jp

人口が急速に減少しているため、日本の列車利用需要も減少している可能性がある。しかし、都市部の通勤時間帯には、無人駅が急速に増加しているにもかかわらず、電車の混雑は依然としてよく見られる。日本全体で、列車の需要は地域によって異なることが興味深い。需要が高い地域に鉄道の維持や開発の建設活動を集中することがより効果的である可能性がある。したがって、将来的に利用者数が増加する駅がどこに位置しているかを調査することが必要である。

本研究では、過去の利用者データを利用して数値実験を行い、将来的な利用者の傾向を予測する。利用者の推移を理解することで、本研究の結果がその推移に影響を与える要因についての将来的な研究の基礎となることを期待する。

Key Words: train network, Japan depopulation, train infrastructure, population growth, population distribution, prefectures, nationwide, immigration, urban city, countryside

1. はじめに

日本において、電車は通学や通勤などでよく利用される交通手段である。しかし、全国の駅において、維持・展開に関する必要性があるのか否かについては疑問が残る。少子化に伴い、人口が減少していく中で、全国の駅の利用者数が減少することや、展開要求がなくなることが懸念されている。また、都市部においては通勤通学時間帯に混雑する駅が多い一方、地方の鉄道には無人駅が多いという問題も指摘されている。このような背景から、都市部だけで線路に関する建設活動を行えばよいという意見があるが、これが正しいかどうかについては疑問が残る。なぜなら、人口だけではなく、駅の分散、天候、観光客など、他の要素も駅の建設要求に影響を与える可能性があるためである。以上を考慮した上で、本論文では、過去の利用者数データを用いて、未来の利用者数の推移を予測し、その要素について検討することを目的とする。

(1) 論文の構成

本研究では、駅の乗降客数を対象として、数値実験を行う。本論文は、以下の4つの章から構成されている。

1章：はじめにでは、本研究の目的、論文の構成、研究方法、使用データ、使用ソフトについて紹介する。

2章：駅の客数の推移の予測では、使用データの特徴、数値実験、そして実験結果の判断を行う。

3章：結果の議論と影響要素の基礎判断では、実験結果から得た認識と影響を与える要素に関する基礎的な議論を行う。

4章：結論と今後の課題では、本研究のまとめと今後の課題に取り組むための残された課題について述べる。

(2) 研究方法

研究方法は、過去のデータを基に回帰分析の数値実験を行った。回帰分析は、複数の変数の関係を表すために最も適した線を作成し、数式化する分析手法である。作成された線を回帰線と呼び、回帰線から回帰式も算出される。この回帰線と回帰式を用いることで、データの傾向やパターンを予測することができる。

回帰線の作図と回帰式の算出は、エクセルでデータを入力した後、近似曲線の追加機能を使用して行う。散布図を作成した後、表示された各点を選定し、適切な表現タイプを選んで回帰線を描画した。ダイアログボックスには回帰式も表示させる機能がある。

数値分析においては、正確性を判断するために決定係数が使用される。決定係数とは、データに対する推定された回帰式の当てはまりの良さを表す係数であり、一般に R^2 で表されます。0から1までの値をとり、1に近い

ほど回帰式が実際のデータに当てはまっていることを示す。

決定係数は次のような公式によって求められる

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

y_i : 実測値 $_i$

\hat{y}_i : 予測値 $_i$

\bar{y} : 実測値の平均値

得られた回帰線、その傾向がデータを正確に予測できるかどうか、あるいはデータとの当てはまりがどの程度あるかは、決定係数を用いて判断される。

<決定係数 (R^2 値) の判断目安>

0.6以下 : モデルとして意味なしでない

(全然予測できてない)

0.8以上 : すごい良いモデルができた

0.9以上 : 「過学習」してないか?

「過学習」とは公式が「過去データに対してはきちんと予測できるが、知らないデータに対しては全然当たらない」という実用性を全く伴っていない状態のことである。本研究では、 $R^2 = 0.8$ を目安として選び、数値実験の結果を判断することとした。決定係数を求めるには、エクセルに記載されている機能の中にも求めることができる。

(3) 使用データ、使用ソフト

本研究では、乗降客数データについては国道数値ダウンロードサービスサイト (<https://nlftp.mlit.go.jp>) から提供を受け、参考になる人口データについては政府統計の総合窓口 (<https://www.e-stat.go.jp>) から提供を受けた。

数値実験およびグラフの描画には、エクセルが使用され、事前処理や繰り返しの計算は FORTRAN プログラミング言語で実行された。また、地図上でデータを表示するために QGIS が利用された。

2. 駅の客数の推移の予測

駅の客数は色々な要素により影響を与える。例えば：周辺に施設の発展があり、スポットの展開などのことで一般的な人や観光客を引く要素になり可能性がある。周辺の要素から影響があるのを否定できないが本研究の方法としてそのような要素を考慮せずに、過去のデータを基づいて未来の駅の客数の推移予測を行う。

具体的には、本研究では全国の駅の乗降客数の推移を予測し、県の境界線で数値結果を分割し、境界線内の結果を分析することを目的としている。境界線を用いるこ

とで、県ごとに異なる結果を比較することが可能となり、より詳細な分析が可能になる。

(1) 利用データ

国土交通省が提供してくれた 2011 年から 2020 年までの 10 年間の駅の乗降客数データを利用できる。データ数が多いほど、予想数値実験の精度が向上すると考えられるが、2020 年は新型コロナウイルス感染拡大の影響により、100% の駅で客数が減少した。この減少は、リモートワークの普及や政府による緊急事態宣言など、人為的な要因によるものであり、自然現象ではないため、例外的な影響と言える。そのため、本研究では 2019 年までのデータを用いて、数値実験を行うことにする。

(2) 数値実験

2011 年から 2019 年までのデータが十分に提供されている「お台場舞浜公園」の乗降客数の推移を例として乗降客数の推移の予想を次に説明する。

下記の三つのグラフの図 (図-1, 図-2, 図-3) は東京都の「お台場海浜公園」駅の 2011 年から 2019 年までの一日の乗降客数を示している。図の上にデータの近似曲線の数式と決定係数が書いてある。下記の三つのグラフの図から明らかに同じデータで数式の次数が大きくなれば決定係数も大きくなり、数式と実際のデータの当てはまりがよくなるとわかる。節 1.2 のように過去データを用い、未来のわからない傾向を予測するために決定数 $R^2 = 0.8$ を目安として適当な数式を選定される。「お台場海浜公園」乗降客数データの場合は次数 1 の回帰式で近似曲線を表現したとき $R^2 = 0.53$ の決定係数が得られた。決定係数の大きさは上記の目安に比べると小さく、図-1 から明らかなように近似曲線と乗降客数推移はピッタリではないとわかる。

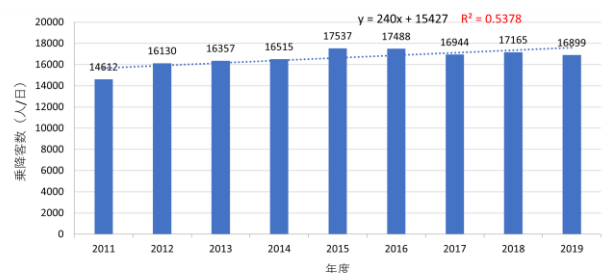


図-1 1 次数の回帰式で「お台場海浜公園」駅の乗降客数推移の表現

次、次数 2 の回帰式で近似曲線を表現すれば $R^2 = 0.89$ の決定係数が得られ、近似曲線の形も客数推移を表現できるようになる (図-2)。

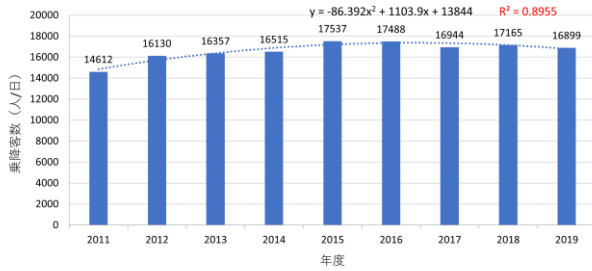


図-2 2 次数の回帰式で「お台場海浜公園」駅の乗降客数推移の表現

$$y(x) = -86.392x^2 + 1103.9x + 13844 \quad (2)$$

回帰式 (式(2)) の微分によって、ある関数の傾きの変化を解析することができる。特に、1 回微分の数式を 0 と等しい方程式 (式(3)) を解けば、その関数の極値や増減の変化点を求めることができる (表-1)。この解を基にし、ある時点におけるグラフが増加しているか減少しているかを判断することができる。

$$y^1(x) = -172.784x + 1103.9 = 0 \quad (3)$$

$$\rightarrow x = 6.389$$

表-1 関数の極値と増減の変化点を表現表。

	2011	2012	...	2016	2019
x	1			6.389	$+\infty$
$y^1(x)$		+		0	-
$y(x)$	↗			↔	↘

上記の分析結果 (表-1) から、「お台場海浜公園」駅の乗客数は 2011 年から着実に増加しており、2016 年頃にピークを迎えたことが明らかとなった。その後は減少傾向を示しており、2019 年時点でも減少傾向が見られることが分かった。この傾向が今後も続くことが予想されるため、同駅の利用者数が減少する傾向にあることが結論される。

次数 3 の回帰式で近似曲線を表現すれば決定係数は $R^2 = 0.90$ になる (図-3)。回帰式の次数を上げることで、決定係数が大きくなり実際のデータとの当てはまりがよくなることが示された。ただし、決定係数が 1 に近くなりすぎると「過学習」が発生し、未来の傾向を正確に予測することができなくなることが前章で議論された。本研究では、「過学習」を避けつつも傾向判断を正確に行うため、回帰式の次数を 1 から上げ、回帰式の決定係

数が $R^2 \geq 0.8$ になり、その回帰式を選択することとした。「お台場海浜公園」の場合、次数 2 の回帰式が選択され、傾向判断に用いられることとなった。

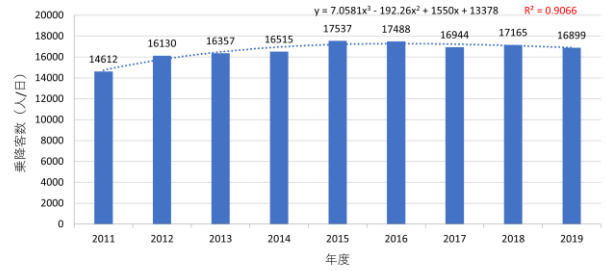


図-3 3 次数の回帰式で「お台場海浜公園」駅の乗降客数推移の表現

上記の例と同じように、本研究では全国の駅の乗降客数の予測を行う。全国を対象として、以下のステップで数値実験を行い、各駅の乗降客数の推移を予測する。

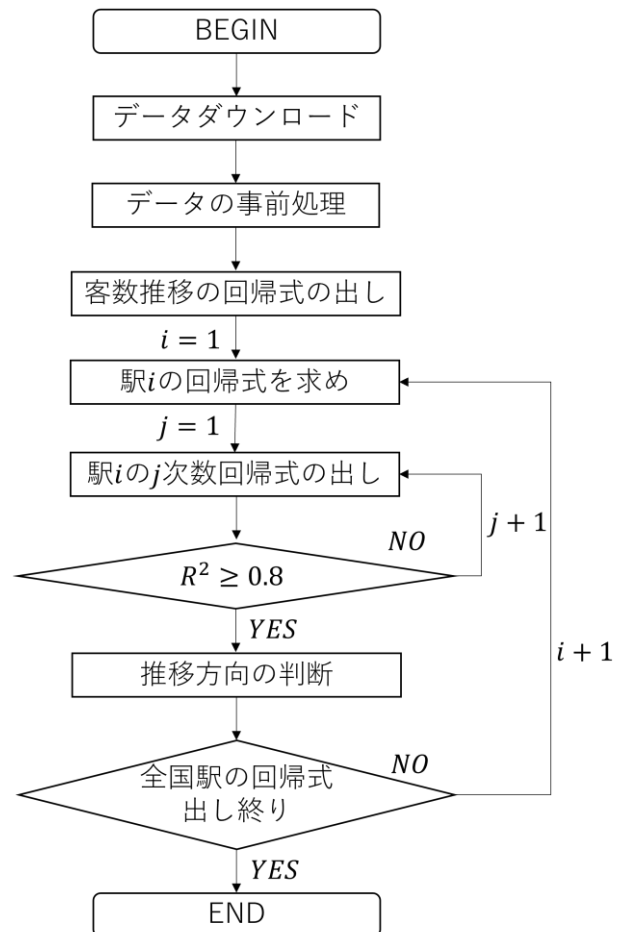


図-4 全国駅の規模で推移予測のプロセス

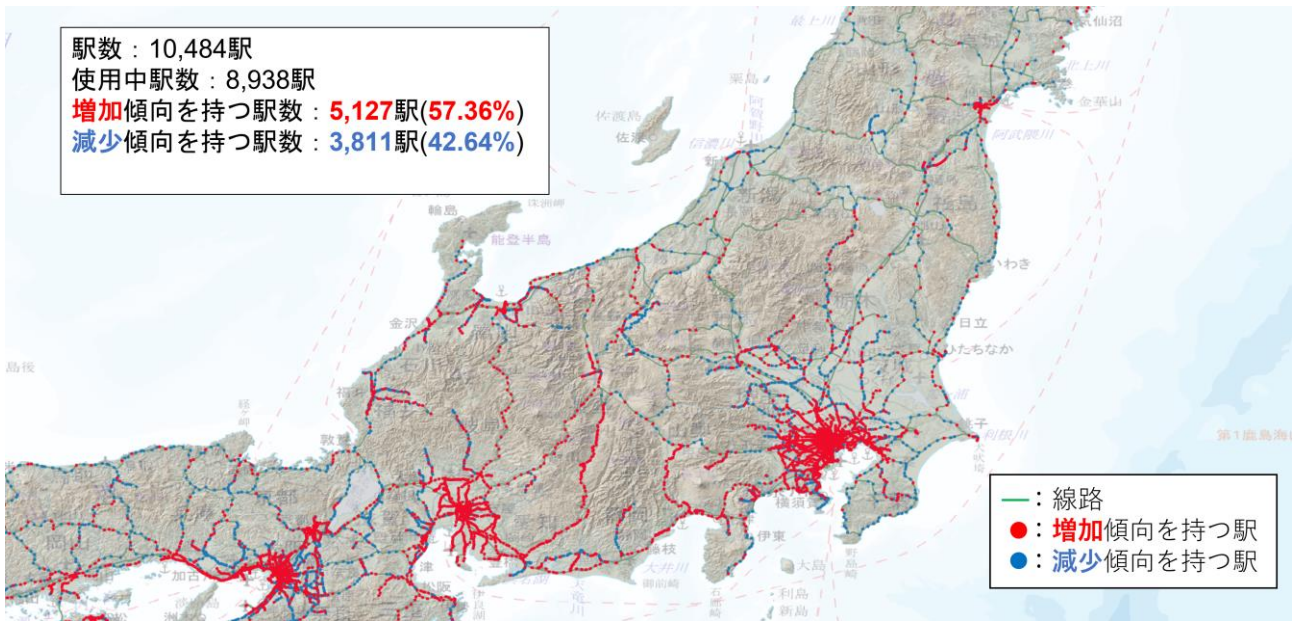


図-5 予測の結果

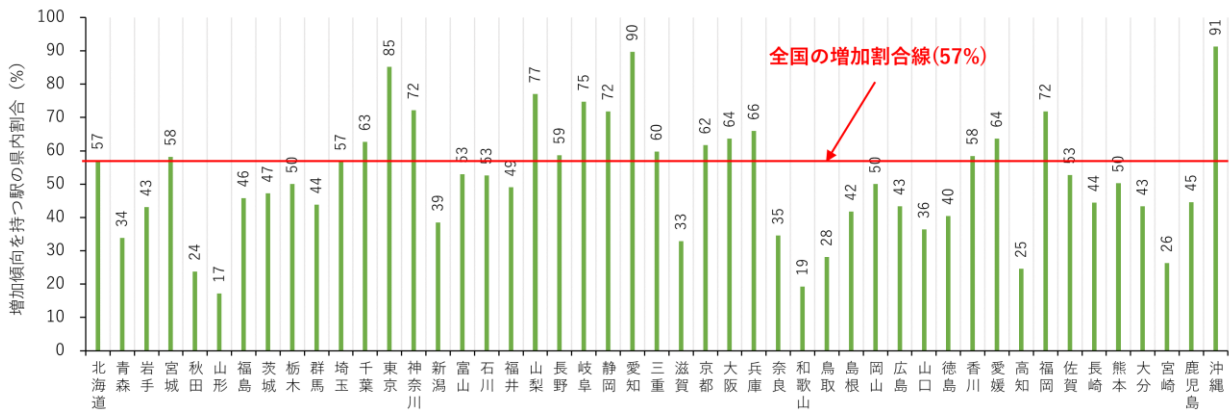


図-6 県内の増加傾向を持つ駅の割合

「事前処理」では、現在使用されていない駅のデータを削除した。事前処理を行う前には 10,484 駅のデータがダウンロードされたが、事前処理後は現在まだ使用中の駅数が 8,938 駅に減少し、前と比較して 14.75%減少した。全国使用中の 8,938 駅のデータを使用して、各駅の客数推移の回帰式が求められ、客数推移を予測する。一方、「お台場海浜公園」の例とは異なり、2019年度以降に上昇と下降を繰り返すケースもあるが、本研究では 2019年度直後の傾向だけを考慮し、推移方向を決定する。

予測結果を地図上で表現することにより、全国の駅の客数推移を示すことができる(図-5)。図-5は、線路を緑の線で表し、増加傾向を持つ駅を赤い点、減少傾向を持つ駅を青い点で表したマップの結果を示している。また、県の単位で結果を分け、統計を行えば結果は図-6に表される。

(3) 実験結果の判断

図-6は、県内の増加傾向を持つ駅の割合を示している。全国の駅数の 57%にあたる 5,127 駅が増加傾向を持ち、43%にあたる 3,811 駅が減少傾向を持っている。図-6から明らかなように、沖縄県が最も高い増加傾向を持つ割合であることが分かる(91%)。次いで、愛知県(90%)、東京都(85%)、山梨県(77%)が続く。一方、最も低い増加割合を示す県は山形県であり、その割合は 17%である。和歌山県(19%)、秋田県(24%)、高知県(25%)が続く。全国の増加割合(57%)を上回る割合を示す県は 19 県あり、逆に下回る割合を示す県は 28 県ある。

3. 結果の議論と影響要素の基礎判断

(1) 結果の議論

これらの結果から、県内の鉄道需要は増加傾向にあると言える。特に沖縄県、愛知県、東京都、山梨県は高い増加傾向を示しており、今後の交通政策の方針決定において重要な考慮要因になると思う。一方で、増加割合が低い県も存在し、鉄道需要の伸びに関しては地域差があることも示唆される。これらの県では、より線路に関する展開が必要がないと言える。また、鉄道使用の需要の伸びを促進するために、鉄道運行におけるサービス向上や周辺施設の整備など、より具体的な対策が必要であることが考えられる。

(2) 影響要素の基礎判断

人口に関する要素：徒歩で一番近い駅までかかる時間は、アパートを探すときによく考慮される要素の一つである。駅の周辺に住んでいる人口は、直接的に駅の客数に関係があるため、都市部では客数増加駅の割合が高いことが図5と図6から分かる。

人口と駅の分布に関する要素：北海道は日本の都会の一つであるが、北海道の人口は全国において最も急速に減少している。しかしながら、北海道においては、人口が急速に減少していても増加傾向を持つ駅の割合が他地域と比べて高くなっており（57%）、人口と駅の分布の要素は増加の割合にも影響を与える可能性があると考えられる。一方、滋賀県の人口は増減せず、急速ではないが、徐々に増加している。しかし、滋賀県においては、増加傾向を持つ駅の割合が多くなると、割合が低いことが分かる。（図6）

気候条件に関する要素：図5と図6から分かるように、日本海側においては増加傾向が非常に低い地域が複数存在している。特殊な気候条件により、建設条件が困難であることが、乗降客数の推移に悪影響を与える可能性がある。

観光客に関する要素：一方、沖縄県は、日本全国において人口が最も少ない地域であるが、最高の増加割合（91%）を示している。これは、沖縄県における人口の急速な増加や、観光時期において、ローカル住民だけでなく、観光客も駅を利用するためである。

4. 結論と今後の課題

(1) 結論

本研究では、県を単位とし、鉄道建設要求のある地域を特定するために、県内において乗降客数が増加傾向にある駅の割合を分析した。2011年から2019年までの乗降客数のデータを使用し、過去のデータから未来の傾向を予測するための適切な回帰式を求めた。都市部では、乗降客数が増加傾向にある駅の割合が高いことが明らかになった。一方、乗降客数が増加傾向にない地域では、様々な要素が影響していることが議論された。

(2) 今後の課題

本研究では、過去のデータを使用して回帰式を求め、駅の乗降客数の増加傾向を予測した。回帰式と実際の値の当てはまりを確保するため、適当な決定係数を用いたが、未知のデータに対しては予測誤差が発生する可能性があることを認識している。そのため、より信頼性の高い予測を得るために、3.2節で示した複数の要素を考慮に入れてから回帰式の結果を調整し、比較する必要があると考える。

人口の推移は、乗降客数に直接的な影響を与える要素の一つであると考えられる。この理由から、今後の研究の課題として、各県の人口推移を始めとする要素を検討し、本研究の結果と比較して、より正確な予測を得ることが必要であると考えています。

謝辞：本研究における分析のデータは国土交通省国土数値情報ダウンロードサイトと総務省統計局ホームページより提供いただいた。ここに記して感謝の意を表す。

参考文献

- 1) 浅見 均：地域鉄道の駅勢圏人口分析等に関する事例研究，土木計画学研究発表会・講演集，Vol.61，2020。
- 2) 浅見 均：地域鉄道の駅勢圏人口分析時系列分析と将来予測，土木計画学研究発表会・講演集，Vol.62，2020。
- 3) 上野 雄貴：将来人口分布と乗降客数減少に対応した東京都市圏における鉄道駅周辺の土地利用，土木計画学研究発表会・講演集，Vol.63，2021。
- 4) 清田 成毅，伊藤 雅：鉄道の中心駅とLRT路線が都市中心部の人口増減に及ぼす影響に関する研究，土木計画学研究発表会・講演集，Vol.64，2021。
- 5) 国土交通省：国土数値情報(駅別乗降客数 2.7 版)，<https://nlftp.mlit.go.jp/ksj/other/agreement.html>，最終閲覧 2022/12/26。
- 6) 政府統計の総合窓口：国政調査(2020年，2015年，2010年，2005年，2000年)<https://www.e-stat.go.jp/gis/statmap-search?page=1&type=1&toukeiCode=00200521>，最終閲覧 2023/01/14。

PREDICTING THE FUTURE TREND OF TRAIN PASSENGERS BY DOING STATISTIC NUMBERIAL EXPERIMENT WITH PAST DATA

TRAN Thanh Hai

In this research, we do statistic numberial experiments basing on the data of passenger number in past years to predict the trend of passenger number in the future (increase or decrease) in one particular station in Japan. Same experiments are done to every station (still in use) in Japan. Finally, we count the number of station having increasing trend in one particular prefecture, calculate the proportion to the number of station in that prefecture. This result is tend to be the base for futher researches that evaluate the prefecture having the need of construction in maintaining or scale developing in train system.