旅行情報サイトから取得されるデータを活用した空き家利活用施設の地域分布・特徴抽出に関する分析

佐々木　碧[1]・森崎　裕磨[2]・藤生　慎[3]

[1]学生会員　金沢大学大学院　自然科学研究科地球社会基盤学専攻（〒920-1192 石川県金沢市角間町）
E-mail: aoto@stu.kanazawa-u.ac.jp

[2]正会員　金沢大学助教　融合研究域融合科学系（〒920-1192 石川県金沢市角間町）
E-mail: morisaki@staff.kanazawa-u.ac.jp (Corresponding Author)

[3]正会員　金沢大学准教授　融合研究域融合科学系（〒920-1192 石川県金沢市角間町）
E-mail: fujiu@se.kanazawa-u.ac.jp

近年わが国では人口減少に伴い，全国のあらゆる地域で空き家が増加している．空き家増加の対策として，国はその利活用を推奨しており課題の解決に取り組んでいる．特に我が国では観光を目的とした空き家利活用の事例が多い．一方で，空き家の増加に伴う利活用事業の活発化により，利活用した観光資源に対する利用者形態の把握が求められている．そこで本研究では，はじめに旅行情報サイトを用いて全国の空き家利活用の観光スポット及び，その他のスポットの利用者データを抽出した．また各利用者が持つ特性を，2 パターンに分けた基礎集計により明らかにした．さらに双方の利用者データを用いた 2 項ロジスティック回帰分析により，利用者が利活用スポットを選択した確率とその要因を明らかにした．

Key Words: *vacant houses, utilization, Travel Information Website , Binomial Logistic Regression Analysis*

1. 本研究の背景と目的

(1) 本研究の社会的な背景

近年，わが国の人口は著しく減少しており，全国の市町村自治体でも全体の約 82.4%で人口の減少が進んでいる [1]．この動きに伴い，各地で空き家の増加が進んでおり地域の重大な問題となっている．現在，全国における空き家の数は約 849 万戸，総住宅数に占める空き家率も約 13.6%と共に過去最高の値となっている [2]．特に，管理の不十分な空き家は，景観の悪化，衛生的な問題，家屋の倒壊，治安の悪化など様々な問題を引き起こし，周辺や地域に悪影響を与えている．

このような状況下で，2015 年 5 月，「空き家等対策の推進に関する特別措置法」が制定され，各自治体が空き家の対策を進める上での骨組みが整備された [3]．また，これに先立ち 2016 年，空き家等対策計画の実施等を支援するために「空き家対策総合支援事業」を開始し，空き家等の除去，利活用，関連事業など総合的な空き家対策に取り組む自治体に対し支援を行う等，空き家の問題解決の推進を図っている．特に利活用には，新築より費用が安いこと，自治体の補助金制度が充実していること等のメリットがあり，近年わが国では空き家の利活用に対する取り組みが積極的に行われている．

利活用の資源となる，古民家や古い住宅には我が国の建築技術や居住習慣が具現化されており，文化的価値が非常に高い．我が国の建築になじみのない都市圏の人々や外国人にとってこの特徴は，新鮮さや懐かしさといった魅力を感じさせている．そのため観光資源としても注目されている．

空き家の利活用の内，観光資源としての事例に関しては，企業や歴史的資源に着目した内閣官房「歴史資源を活用した観光まちづくり事例」[4]により行われている事例が存在するが，取り組み概要や組織形態等が整理されている一方で，利用者から見た事例の実態や一般的な観光地との違い等，ユーザー目線での把握がなされていない．今後，急速な空き家の増加に伴う利活用事業が活発化することが考えられることから，空き家を利活用した観光資源に対する利用者の利用形態を把握しておく必要

があると考えられる．

(2) 本研究の目的

わが国では，空き家の増加に伴う利活用事業が促進されている．一方で，特に集客が見込める観光資源としての空き家利活用事業はその全体把握がなされておらず，ユーザー目線での実態を明らかにする必要がある．本研究では，旅行情報サイトを活用して全国の観光スポットの中から空き家が利活用された観光スポットと空き家が利活用されていないその他の観光スポットの利用者データをそれぞれ抽出し，利用者形態の特徴を明らかにする．さらに前述するデータを用いて，利用者が空き家利活用スポットを選択する確率とその要因を分析により明らかにする．それにより，空き家利活用スポットを選択した利用者の特徴を明らかにすることを目的とする．なお，筆者らは，利活用施設の用途に着目した研究を実施しており，本研究では利用者視点での特徴の把握を目的とする．

図-1 の本研究の方針に示すように，はじめに，旅行情報サイト「じゃらん net」に掲載されている観光スポットの内，空き家の利活用に関するワードを検索し，ヒットした観光スポットにおける利用者の口コミの投稿に関する特定のデータを抽出する(Step1)．また，上記の観光スポット以外の事例をランダムで抜粋し，同様の利用者の口コミの投稿に関するデータを抽出する(Step2)．3章1節以降で詳しく記述するが，じゃらん net は観光情報サイトとして広く普及しており，特定施設の利用者情報をより具体的に理解することが可能である．その後抽出した利用者データの基礎集計を行い，データの特徴を把握する(Step3)．前述する空き家利活用の観光スポットと空き家利活用以外の観光スポットの各利用者データを組み合わせて，2 項ロジスティック回帰分析を行い，観光スポットにおける利用者データから，利用者が観光時に空き家利活用スポットを選択する確率とその要因を明らかにする．(Step4)．以上の工程により，観光スポットの利用者データから，空き家利活用スポットを選択した利用者の特徴を明らかにする．

2. 既往研究の整理と本研究の位置付け

本研究では旅行情報サイトを用いた空き家利活用の観光スポットにおける利用者に関する基礎集計及び，利活用事例の発生確率の要因分析と予測に関する研究であり，同様に旅行情報サイトや観光地の空き家利活用施設に関する研究はいくつか存在する．本章では本研究の位置付けのために以下の 2 点から既往研究の整理を行う．

ⅰ． 空き家の利活用における全般的な現状分析

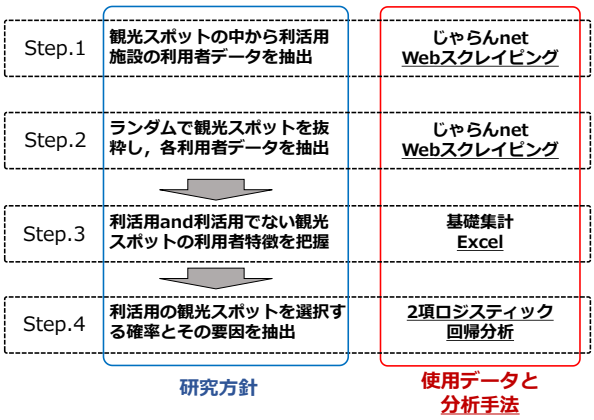


図-1 本研究の分析の流れ

ⅱ． 観光目的の利活用施設に焦点を当てた研究

(1) 既往研究の整理

はじめに，空き家の利活用における全般的な現状分析の研究の整理を行う．金ら [5]は，選定都道府県へのヒアリング調査により空き家利活用事例を収集し，14 事例のマネジメント体制の類型化，可視化を行うとともに，ケーススタディにより事業の実施主体と連携団体との関係性を明らかにしている．中田ら [6]は，関東を中心とした 1 都 9 県の自治体や商工会等から 203 件の古民家の事例を収集し，基礎的な集計により古民家転用の現状と傾向を明らかにすることで，課題と可能性を考察している．

次に，観光を目的とした利活用施設の焦点を当てた研究の整理を行う．溝口ら [7]は，滋賀県近江八幡市の歴史的市街地を対象に，古民家利活用の流れを明らかにし，その促進要因を探っている．対象地の利活用物件の事業者のヒアリング調査を行い，物件成約，改修等の成立要因明らかにするとともに，類型毎の典型例から経緯や成立条件を明らかにしている．

(2) 本研究の位置付け

本研究の位置付けを行うため，前述する既往研究との関係について述べる．まず，空き家の利活用における全般的な現状分析であるが，金らの研究では事業のマネジメント体制や活動内容を明らかにし，適切な方針を考察している．加えて中田らの研究では用途の分類をしているものの，事業者へのヒアリングにより事業の実態を明らかにしている．本研究では，観光情報サイトを用いて，利用者視点からの空き家利活用の特徴に関する集計・分析を行いその特徴を明らかにしており，既往研究と使用データが異なる．また，溝口らの研究に見られるように，観光地や歴史保存地区等に対象地域を絞り空き家利活用の実態把握や将来性を検討している研究は多くみられる．しかし本研究では観光目的の空き家利活用施設に関して，全国に焦点を広げた実態の把握を行っている．筆者らが

確認を行った限り，上記のような研究は存在しない．

3. 利用者データの抽出と基礎集計

(1) 旅行情報サイトを用いた利活者データの抽出

本研究では全国の空き家を利活用した観光スポット及び，それ以外の空き家を利活用していない観光スポットを旅行情報サイト「じゃらん net」[8]を用いて収集する．**図-2** に示すように「じゃらん net」は，国内の各旅行サイトの年間利用者数が最も多く，旅行や観光の情報源として広く活用されている [9]．当サイトを用いた研究には，大川ら [10]の観光地のイメージを口コミからテキストマイニングを用いて言語分析により明らかにした研究等があり，利用者の視点から観光施設の実態を明らかにする際の分析のデータとして広く活用されている．本研究においても，利用者から見た観光目的の上述した 2 種類のスポットの特徴を把握するため，該当する施設を抽出した．

まず空き家を利活用した観光スポットを抽出するために，当サイトに掲載されている約 10.8 万件の観光スポット [8]の内，「空き家」，「利活用」，「再利用」，「古民家」，「リノベーション」の 5 つのワードを検索，Web ページの HTML から情報を抽出するためのライブラリである Beautiful Soup を用いて Web スクレイピングを行った．抽出したスポットの内，本研究の目的と沿わない施設を除いた 292 件の事例を本分析のデータで用いる．ここで「利活用施設」とは，「一度使われていた，又は空き家として放置されていた家並びに施設が，以前とは異なる用途として利用されている施設」と定義する．また，抽出したスポットを訪れた口コミ投稿者（以下，利用者）のデータを抜粋した．**表-1** に，利用者のデータの内訳を示す．データの内訳として，利用者の 10 歳毎の「年代（年齢）」に加え，「観光スポットでの滞在時間」，「観光スポット訪問時の混雑具合」，「観光スポットに対する満足度評価」といったデータが得られる．

続いて，空き家を利活用していない観光スポットを抽出する．じゃらん net に掲載される約 10.8 万件の観光スポットから，ランダムで 300 件の観光スポットを抽出する．また前述した方法と同様に，各スポットの各利用者データを抜粋する．**図-3** に空き家を利活用した観光スポットと利活用していない観光スポットの利用者データのデータ数の割合を表記する．

(2) 利用者データの基礎集計

本節では「じゃらん net」から得られた各観光スポットの利用者データに着目し，空き家を利活用した観光スポットとその他の利活用していない観光スポットの利用者の特徴を，基礎集計によりそれぞれ把握する．

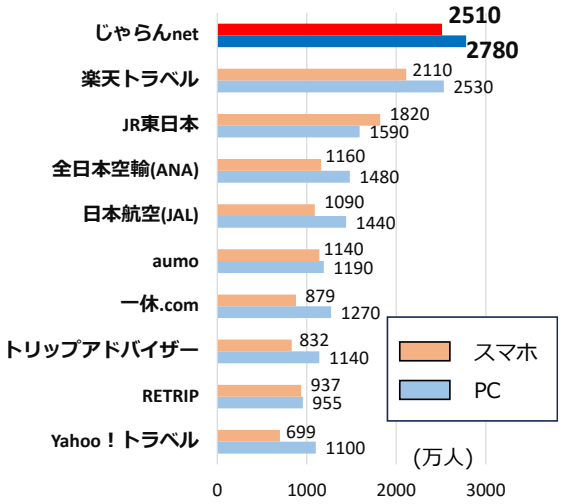


図-2 各旅行サイトの年間利用者数（2021 年）

表-1 「じゃらん net」から得たデータの概要

項目	分類数	データ例
口コミ	（文章データ）	-
年代（年齢）	（量的データ）	30代
観光スポット訪問時の混雑具合	5分類	1：空いていた 2：やや空いていた 3：普通 4：やや混んでいた 5：非常に混んでいた
観光スポットでの滞在時間	4分類	1：１時間未満 2：１～２時間 3：２～３時間 4：３時間以上
観光スポットに対する満足度評価	5分類	1：不満 2：やや不満 3：普通 4：満足 5：大変満足

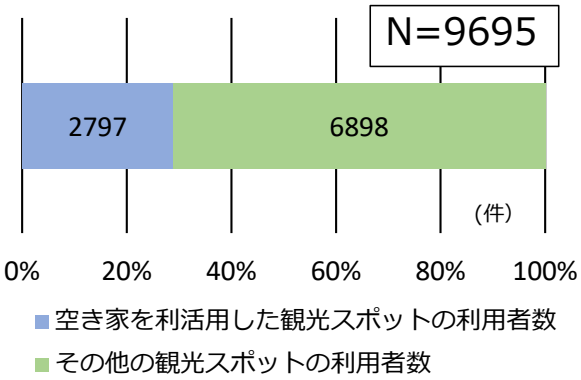


図-3 収集した観光スポットの利用者数の内訳

図-4 に空き家を利活用した観光スポットと利活用していない観光スポットにおける，それぞれの利用者の年代の内訳の集計結果を示す．利活用した観光スポットでは，30 代，40 代，50 代の利用者の割合が多く占めている一方で，利活用していない観光スポットでは，20 代，30

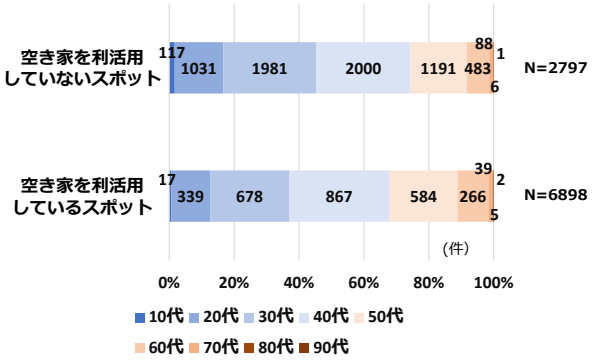


図-4　各観光スポット毎の利用者年代（年齢）の内訳

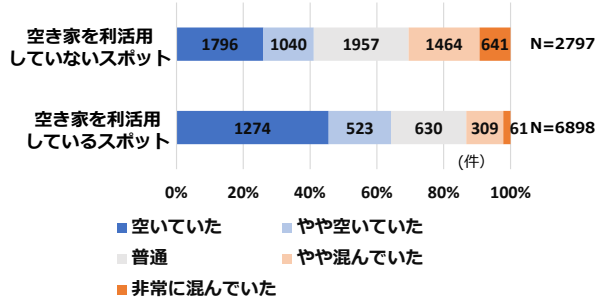


図-5　各観光スポット毎の利用者訪問時の混雑具合の内訳

代，40 代の利用者の割合が多く占めている．全体的に，利活用している観光スポットの利用者の方が，年齢層が高めであることが分かる．

図-5 には，空き家を利活用した観光スポットと利活用していない観光スポットにおける，それぞれの利用者の訪問時の混雑具合の内訳の集計結果を示す．利活用した観光スポットでは，「空いていた」や「やや空いていた」と回答する利用者が全体の 6 割以上を占める一方で，利活用していない観光スポットでは，混雑さを回答する利用者の割合も一定数存在している．利活用した観光スポットの利用者の方が，スポットにおける混雑具合が軽いと回答する割合が多いことが分かる．

図-6 に空き家を利活用した観光スポットと利活用していない観光スポットにおける，それぞれの利用者の滞在時間の内訳の集計結果を示す．双方に大きな相違は見られないが，利活用した観光スポットの利用者の方が，利活用していない観光スポットの利用者よりも，比較的長時間，スポットに滞在していることが分かる．

図-7 には，空き家を利活用した観光スポットと利活用していない観光スポットにおける，それぞれの利用者の満足度評価の内訳の集計結果を示す．利活用した観光スポットの利用者の過半数が「大変満足」と回答している一方で，利活用していない観光スポットの満足さを回答する利用者は多いものの，前者と比較すると割合は少ない．全体的に利活用した観光スポットの利用者の方が，

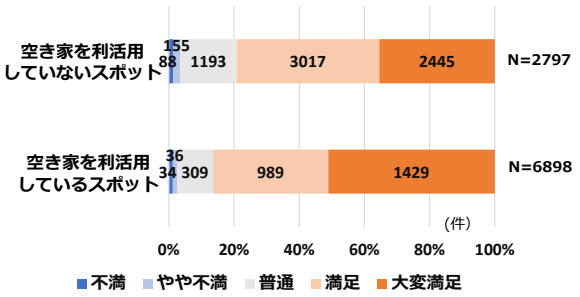


図-6　各観光スポット毎の利用者の滞在時間の内訳

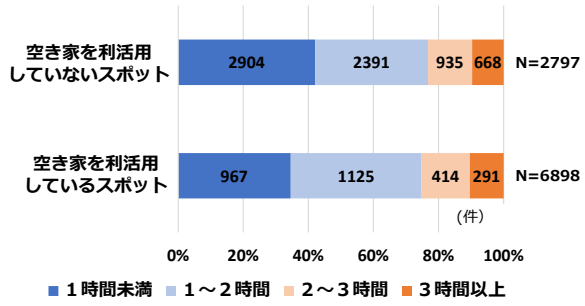


図-7　各観光スポット毎の利用者の満足度評価の内訳

スポットに満足さを感じている割合が高いことが分かる．

4.　利用者データを用いた利活用スポットの選択の是非に関する分析

前章(2)において，空き家利活用している観光スポットとその他の利活用していない観光スポットの利用者データに関する基礎的な集計を行うことで，双方の利用者形態の特徴を把握した．本章では，以上のデータを用いた 2 項ロジスティック回帰分析により，観光時に利用者が空き家利活用スポットを選択した確率とその特徴を明らかにする．

(1)　2 項ロジスティック回帰分析の概要と使用する変数

本分析においては 3 章で分類した利活用している観光スポットの利用者データ（2,797 件）とその他の利活用していない観光スポットの利用者データ（6,898 件）を用いて，2 項ロジスティック回帰分析を実施する．

2 項ロジスティック回帰分析は，2 値の質的変数を目的関数とし，その発生確率，説明変数との影響力を表すことができる多変量解手法[11]である．説明変数の発生確率pは式(1)によって表すことができる．なお，β_1，β_2，β_kは回帰係数，x_1，x_2，x_kは説明変数である．

$$p = \frac{1}{1 + \exp\left(-\{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k\}\right)} \quad (1)$$

本分析において 2 項ロジスティック回帰分析に投入する変数を**表-2** に示す．目的変数は，利活用スポットの利用是非のカテゴリーデータ（利活用スポットを利用した・他スポットを利用した）とした．説明変数は，ロコミ回答者の個人属性である，「年代（年齢）」に加え，「観光スポットでの滞在時間」，「観光スポット訪問時の混雑具合」，「観光スポットに対する満足度評価」の合計 4 変数とした．これらを目的変数，説明変数とし，2 項ロジスティック回帰分析を実施する．

また，2 項ロジスティック回帰分析に使用する説明変数の各相関関係を**表-3** に示す．「観光スポットでの滞在時間」と「観光スポットに対する満足度評価」の変数間で最も高い相関が得られ，相関係数は 0.222 であった．また，「観光スポット訪問時の混雑具合」と「観光スポットでの滞在時間」の間にも相関係数 0.216 の値が得られたが，いずれも弱相関であることから，説明変数に採用した．

(2) 観光スポットにおける利用者の空き家利活用スポットの選択に関する分析結果

統計モデルを作成，統計的な検定を実行し，データを分析するのに用いられる Statsmodels という Python ライブラリを用いて，2 項ロジスティック回帰分析を実施する．まず，変数を増減法により選択する．変数選択過程における各ステップの精度，及び変数選択の結果を**表-4** に示す．本分析においては，モデル精度は AIC（Akaike Information Criterion：赤池情報量基準）[12]を用いて評価する．増減法による変数選択の結果，「観光スポット訪問時の混雑具合」，「観光スポットに対する満足度評価」，「観光スポットでの滞在時間」，「年代（年齢）」の 4 変数が投入されたステップ 4 のモデルで AIC が 10815.93 となり，最も低い値になった．したがって，本研究において観光スポット利用者における空き家利活用スポットの選択モデルは，上記 4 変数のすべてが採用される結果となった．

また，**表-5** に本分析における 2 項ロジスティック回帰分析によって得られた各説明変数および定数項の偏回帰係数，オッズ比，p 値を示す．尤度比検定の結果，p 値が 0.01 未満を示したため本分析で作成したモデルは有意であると言える．また，算出された偏回帰係数から，本分析で得られた利用者の空き家利活用スポットの選択モデルは以下の式(2)になった．なお，x_1 は「観光スポット訪問時の混雑具合」，x_2 は「観光スポットに対する満足度評価」，x_3 は「観光スポットでの滞在時間」，x_4 は「年代（年齢）」の説明変数である．

ここで，オッズ比を見ると，「観光スポットに対する満足度評価」が 1.390 であったため，観光スポットにおける利用者の空き家利活用スポットの選択確率に最も影響を及ぼす変数であると言える．オッズ比が正の値であることから，満足度評価が高いほど空き家利活用スポットを選択した確率も高くなることを示している．また，「観光スポットでの滞在時間」のオッズ比は 1.228 であったため，滞在時間が長いほど空き家利活用スポットを選択した確率も高くなることを示している．さらに，「年代（年齢）」のオッズ比は 1.015 であり，わずかではあるが，年代が高いほど空き家利活用スポットを選択した確率が高くなることを示している．これら 3 つの説明変数は，オッズ比の 95%信頼区間の下限値が 1 以上であるため，有意な変数であると言える．一方，「観光スポット訪問時の混雑具合」のオッズ比は 0.629 で，1 未満であった．これは，説明変数の設定より，混雑具合が小さいほど，空き家利活用スポットを選択した確率が高くなることを示している．この変数もまた，オッズ比の

表-2 本分析で使用する変数一覧

	項目	データタイプ	データ例
目的変数	利活用スポット利用の是非	質的データ	1：利活用している 0：利活用していない
説明変数	年代（年齢）	量的データ	30（代）（10~90代まで）
	観光スポット訪問時の混雑具合	質的データ	1：空いていた 2：やや空いていた 3：普通 4：やや混んでいた 5：非常に混んでいた
	観光スポットでの滞在時間	質的データ	1：1時間未満 2：1～2時間 3：2～3時間 4：3時間以上
	観光スポットに対する満足度評価	質的データ	1：不満 2：やや不満 3：普通 4：満足 5：大変満足

表-3 説明変数間の相関関係

	年代（年齢）	観光スポット訪問時の混雑具合	観光スポットでの滞在時間	観光スポットに対する満足度評価
年代（年齢）	1.000	-0.069	-0.063	-0.066
観光スポット訪問時の混雑具合	-0.069	1.000	0.216	-0.012
観光スポットでの滞在時間	-0.063	0.216	1.000	0.222
観光スポットに対する満足度評価	-0.066	-0.012	0.222	1.000

表-4 変数選択結果

ステップ	AIC	変数選択の結果 投入した変数	変数選択の結果 除去した変数
ステップ0	-	観光スポット訪問時の混雑具合	なし
ステップ1	11099.62	観光スポットに対する満足度評価	なし
ステップ2	10937.22	観光スポットでの滞在時間	なし
ステップ3	10878.17	年代（年齢）	なし
ステップ4	10815.93	なし	なし

表-5 観光スポット利用者における空き家利活用スポットの選択に影響する要因

		偏回帰係数	オッズ比	オッズ比の95%信頼区間		P値	判定
				下限値	上限値		
説明変数	観光スポット訪問時の混雑具合	-0.463	0.629	0.605	0.655	0.000	＊＊
	観光スポットに対する満足度評価	0.330	1.390	1.310	1.475	0.000	＊＊
	観光スポットでの滞在時間	0.206	1.228	1.168	1.291	0.000	＊＊
	年代（年齢）	0.015	1.015	1.011	1.019	0.000	＊＊
定数項		-2.166	0.115	0.084	0.157	0.000	＊＊
回帰式の有意性（尤度比検定による）				＊＊			

0.01未満：＊＊

$$p = \frac{1}{1 + \exp\left(-\{-2.166 - 0.463x_1 + 0.330x_2 + 0.206x_3 + 0.015x_4\}\right)} \quad (2)$$

95%信頼区間の上限値が 1 以下であるため，有意な変数であると解釈できる．

5. まとめと今後の課題

本研究では，近年増加する空き家対策の一環としての利活用事業に焦点を当て，観光を目的とした事例の利用者データから，基礎的特徴と「利活用スポットを選択した」確率の要因を明らかにした．全国的な観光を目的としたスポットのデータの収集を行うために，旅行情報サイトを用いて空き家の利活用スポットと，利活用していないスポットの利用者データを Web スクレイピングにより抽出した．続いて，それらのデータを前述する 2 パターンに分け，基礎集計により利用者形態を明らかにした．また，双方の利用者データを用いた 2 項ロジスティック回帰分析により，観光時に空き家利活用スポットを選択した利用者の特徴を明らかにした．

今後の課題として，本研究では旅行情報サイト「じゃらん net」を用いて，観光を目的とした空き家利活用施設を抽出したが，観光という分野の中の事例のみに焦点を当てた実態把握となった．今後はじゃらん net のグルメスポットに関する空き家利活用施設や，アンケート調査を用いた更なる空き家利活用事例に関しても集計を検討する必要がある．加えて，空き家問題の解決策の観点から，観光における空き家の利活用事例の需要や維持に関して着目し，各事例で発生した課題や継続性に関しても具体的なデータを取得し，現状の把握を行う必要がある．さらに本研究では利用者の形態に焦点をあてて事例の実態把握を行っている．本研究では深く行わなかった，事業の地理的特性や近隣住民の課題把握に関して，解決に向けたアプローチを行っていく．

参考文献

1) 総務省：令和 2 年国勢調査　人口速報集計　結果の要約，https://www.stat.go.jp/data/kokusei/2020/kekka/pdf/summary.pdf，2023 年 10 月 8 日閲覧
2) 国土交通省：平成 30 年住宅・住宅土地統計調査の集計結果, https://www.mlit.go.jp/common/001314574.pdf，2023 年 10 月 8 日閲覧
3) 国土交通省：空き家対策の現状について，https://www.mlit.go.jp/jutakukentiku/house/content/001429587.pdf，2023 年 10 月 8 日閲覧
4) 内閣官房「歴史資源を活用した観光まちづくり事例」：https://www.cas.go.jp/jp/seisaku/kominkasupport/file/jirei_210426.pdf，2023 年 10 月 8 日閲覧
5) 金大一，姫野由香，牛苗，太田裕喜：地方都市における空き家利活用のマネジメント体制に関する研究，都市住宅学，Vol.103，No.AUTUMN，p.169-p.176，2018
6) 溝口徳昭，谷川陸，山口敬太，川崎雅史：近江八幡旧市街地における古民家利活用のプロセスと成立要件に関する研究，日本都市計画学会関西支部研究発表会講演概要集，Vol.19，p.41-p.44，2021
7) 竹鼻紫，大村謙二郎，有田智一，藤井さやか：伝建地区とその周辺における空き家実態とその利活用可能性に関する研究，日本都市計画学会都市計画論文集，Vol.45，No.3，p.25-p.30，2010
8) じゃらん net：https://www.jalan.net/，2023 年 10 月 8 日閲覧
9) PRTIMES:「調査リリース」2021 年観光関連サイト閲覧者数ランキング，https://prtimes.jp/main/html/rd/p/000000137.000007396.html，2023 年 10 月 8 日閲覧
10) 大川慎：テキストマイニングを利用した観光地イメージの言語間比較に関する試み，日本情報経営学会誌，Vol.39，No.2，p.85-p.94，2019
11) 内田治：SPSS によるロジスティック回帰分析 第 2 版，オーム社，2016
12) 赤池弘次，甘利俊一，北川源四郎，樺島祥介，下平英寿：赤池情報量規定 AIC―モデリング・予測・知

識発見一，共立出版，2007

(Received ???)
(Accepted ???)

ANALYSIS OF REGIONAL DISTRIBUTION AND FEATURE EXTRACTION OF FACILITIES UTILIZING VACANT HOUSES USING DATA OBTAINED FROM TRAVEL INFORMATION WEBSITES

Aoto SASAKI, Yuma MORISAKI, Makoto FUJIU

In recent years, the number of vacant houses has been increasing in all regions of Japan as a result of population decline. As a countermeasure to the increase in vacant houses, the government has recommended the utilization of vacant houses and is working to solve the problem. In Japan in particular, there are many cases of utilization of vacant houses for the purpose of tourism. On the other hand, the increase in the number of vacant houses has led to the activation of utilization projects, and there is a need to understand the types of users of the utilized tourism resources. In this study, we first extracted user data for tourist spots utilizing vacant houses and other spots in Japan using a travel information website. The characteristics of each user were clarified through basic tabulation by dividing the data into two patterns. Furthermore, a binomial logistic regression analysis using both types of user data was conducted to clarify the probability that a user selected a spot for utilization and its factors.