

潜在クラス分析を用いた 訪日外国人旅行者の訪問パターン分析

古屋 秀樹¹・劉 瑜娟²

¹正会員 東洋大学教授 国際地域学部 国際観光学科 (〒112-8606 東京都文京区白山5-28-20)

E-mail: furuya@toyo.jp

²学生会員 東洋大学大学院 国際地域学研究科 国際観光学専攻 修士課程 (同上)

E-mail: liuyujuan914@gmail.com

本研究は、観光・レジャーを旅行目的とした訪日外国人旅行者の都道府県への訪問の組み合わせである訪問パターンの特性把握を目的としている。分析では、観光庁が実施した「訪日外国人消費動向調査」データを用いながら、潜在クラスモデルによって類似性から集約される訪問パターン導出を試みた。複数設定したクラス数それぞれの推定結果におけるモデル全体の説明力をAIC値、BIC値によってチェックした結果、24クラスの設定が最も妥当と判断できた。さらに、導出されたクラス別構成比率と主要国籍・地域、旅行形態、旅行時期、訪日回数などの各要因との関連性について一般化 χ^2 乗検定によって明らかにするとともに、訪問地からみた主要マーケットの特徴について明らかにした。

Key Words : *foreign visitors, combination of visiting place, latent class model, generalized chi-squared test*

1. はじめに

2015年の訪日外国人旅行者数は1973.7万人を数え、過去最高を記録した。2020年の東京五輪開催などをひかえ、一層のインバウンド観光促進にむけた様々な取り組みがなされている。外国人旅行者の日本訪問によって日本の認知が一層高まるとともに、訪日外国人旅行者数増加は、外国への情報発信、大きな経済効果発現につながると考えられる。今後は、訪日外国人旅行者数といった量の視点だけではなく、初回訪日者の満足度を高めてリピーターへの変容をはかる質的観点からの取り組みも重要といえる。

そのために、需要側である旅行者に着目して現状の訪日リピーターがどのような観光行動を行い、満足や醸成しているのか、初回訪日者との比較を通じた実態把握が必要と考えられる。一方、着地側からみると、独自のプロモーションを積極的に実施しているものの十分な成果がもたらされていないケースもあり、これは外国人旅行者がどのようなニーズで訪日し、それに合致する立寄り先を選択しているかという行動規範に対応できていないことも一因と考えられる。また、他の地域と連携した魅力的な周遊ルート設定も重要となる。

これらを問題意識として、外国人旅行者の訪問地、訪

日時期、旅行形態（団体旅行、個人旅行の区別）といった観光行動の特徴把握によって、現状のインバウンド観光市場を理解することができ、ひいては効果的なプロモーションの実施、効果的な観光振興策の検討が可能と考えられる。

そこで、本研究は観光・レジャーを旅行目的とした訪日外国人旅行者の訪問パターン把握を目的とする。この訪問パターンとは、日本における訪問地の組み合わせと設定する。この訪問パターンに着目する理由であるが、訪問地を個々に取り上げた場合、旅行者の包括的な訪日意向を把握できないと考えられるためである。すなわち、旅行者は時間・予算制約の中で訪日意向・動機・目的に合致する、より好ましいいくつかの訪問先を決定していることから、訪問地の組み合わせである訪問パターンの抽出によって包括的な訪日意向・動機・目的を類推できるといえる。また、着地側にとっても、他のどの地点と一緒に周遊しているか明らかになることによって、周遊ルートの設定やそれと密接に関連する旅行者セグメントを踏まえたマーケティング戦略立案が容易になると考えられることができる。

次章で、本研究の使用データ、位置づけを説明した後、3章で分析手法、4章で推定結果ならびに妥当性の検証、5章で導出されたクラスと旅行者個人属性等との関連性

や分析結果の活用にもつた考察を示す。

2. 既存研究と本研究の位置づけ

(1) 訪日外国人消費動向調査¹⁾

本研究では、観光庁が実施した「訪日外国人消費動向調査（平成24年、25年）」の個票データを用いて分析を行った。本調査は、11空港の国際線ターミナル搭乗待合ロビーで出国を待つ訪日外国人旅行者に対して、外国語対応のタッチパネル式PCまたは紙調査票を用い、外国語を話せる調査員によって個人属性ならびに訪日旅行に関する訪問地、同行者、旅行支出、土産品の購入実態、旅行情報源、満足度と再訪意向の聞き取りを実施したものであり、サンプル数は57,855人である。

本研究では、その中から、「今回の日本滞在中における訪問地」回答データを利用しているが、宿泊、日帰りの区分はされていないものである。さらに、個人属性として、訪日時期、国籍・地域、性別、年齢、訪日目的、訪日回数、旅行形態（団体、個人旅行）を活用した。

なお、プロモーションへの利用等を考えた場合、訪問者の活動目的・形態が単一となる細かなゾーン設定が望ましいと考えられるものの、本分析ではパラメータ推定環境による制約からゾーニングを都道府県単位とした。そのため、各都道府県の訪問意向・活動形態が複数考えられるため、それらを類推しなければならないことに留意する必要がある。

(2) 既存研究と本論文の位置づけ

国籍等による観光行動の差異を分析した事例として、日本・中国・韓国における年間宿泊観光旅行と年間国外旅行に関する観光行動の実態を分析したもの²⁾や居住地別訪日中国人旅行者の日本国内における観光行動を分析したもの³⁾などがある。近年では、訪日外国人旅行者消費動向調査が実施され、より広範なサンプルの時系列の特性把握が可能となった。この調査では消費に加え、訪問地や満足度なども聞き取られているためインバウンドの経済効果のみならず周遊などの観光行動が詳しく把握できる。さらに、直近では携帯の位置情報データやSNSなどの比較的データ数の多いビッグデータの取得・分析も精力的に行われており⁴⁾、これらデータに対応した知見を蓄積することが望まれている。

一方、観光振興のための組織として近年、DMO (Destination Management Organization)が着目されており、その機能の1つに各種データ等の継続的な収集・分析とそれを参考にした明確なコンセプトに基づいた戦略の策定が示されている⁵⁾。そのために、ヒートマップのようなある1時点・1地点における来訪者数、前後のトリップ目的地

の連関などの実態把握に加えて、同一旅行者がどのような訪問地を訪問しているのか個人単位の訪日旅行に着目し、訪問地点の組合せである訪問パターンの把握を通じた有望なマーケットの同定が必要不可欠といえる。

本論文では訪問地の組み合わせに着目し、それらが異なるとして発現するのは、旅行の動機や志向、ニーズが原因として存在し、それらが複数のクラスに区分できると仮定する。個々の旅行があるクラスに含まれるとすると、旅行訪問地の選択率がクラスによって規定され、選択率の大きさから頻出する訪問先の組み合わせも類推できる。また、クラスへの所属も確率として導出できるものとする。この考えは、以下に示す潜在クラスモデルにもとづくもので、推定の結果、各クラスの構成比率や各訪問先の選択率を明らかにできる。さらに、各クラスと個人属性などの旅行者要因との関連性を明らかにすることによって、より精緻なマーケティング活動に資することができると考えられる。なお、潜在クラスモデルの適用例として、小林他⁶⁾、谷下⁷⁾、古屋・全⁸⁾があるが、これらは被説明変数に対する構造が異なることを明示した手法であるため、今回採用したクラス抽出を主としたものと異なる。

さて、潜在クラスモデルは、旅行をいくつかのクラスに分類する手法の1つといえるが、ここで他の手法との相違点を考える⁹⁾。各旅行の訪問地は、訪問の有無という離散型変数として表現できる。このようなカテゴリ間の関連性を把握する方法の1つとして相関係数がある。しかしながら、2変数間の関連であり、3変数以上の関連性を示すことができない。また、回帰分析、対数線形モデル、正準相関分析では、複数の説明変数の被説明変数への影響を導出できるものの、該当する、もしくは該当しないというカテゴリカルデータの場合は、誤差に正規分布を仮定するため非常に小さい非負の変数を考慮できず、数量化I類についても誤差の設定で同様な問題がある。外的基準がないケースでの適用手法を考えると、因子分析、主成分分析、数量化III類では、多変数からの次元の圧縮を共分散行列のみによって行うため、その組み合わせがどのように発現しているかという構造が明示できない。また、区分分けでしばしば用いられるクラスター分析でも、何故異なるのかという構造を明示できていないとともに、妥当なクラスター数の決定が分析者の主観に依存すること、各サンプルが属するクラスが排他的であること等の問題点・特徴がある。また、特徴的なカテゴリの組み合わせを抽出するマーケットバスケット解析では、特徴的なペアリングを抽出できるものの、生成過程について内生化できていない問題点がある。潜在クラス分析は、クラスの生成過程を明示しているとともに、非負で示される訪問・非訪問の差異を構造を踏まえながら表現することが可能と考えられる。

一方、潜在クラス分析はすべての被験者が同一のクラス別構成割合を有するとする強い仮定であることとともに、モデル推定においてパラメータの事前確率を想定していないことから、データへの過学習（overfitting）の恐れも想定される。

以上から、本論文の特徴として、訪日外国人旅行者の観光行動を約3万人のデータをもとに分析していること、個別地点の立寄りに加えて、立寄り先の組み合わせを考慮していること、そのために適用事例の少ない潜在クラス分析を用いていることがあげられる。

3. 潜在クラスモデル^{10),11),12),13)}

旅行者の訪問地の組み合わせを考えると、訪問パターンを規定するクラスが X 種類存在し、クラス t の構成比率を π_t^X とする。また、ある個人のある旅行 n のゾーン k の立ち寄りの有無を δ_{k1n} , δ_{k2n} （訪問有り： $\delta_{k1n}=1$, $\delta_{k2n}=0$, 訪問無し： $\delta_{k1n}=0$, $\delta_{k2n}=1$ ）で示す。さらに、旅行 n がクラス t に属すると仮定した場合に、ゾーン k の訪問率を $\pi_{k,1,t}^X$, 非訪問率を $\pi_{k,2,t}^X$ とすると、全ゾーン K の訪問の有無の組み合わせを示す同時確率： $P_{n,t}$ は下記のように示すことができる。

$$P_{n,t} = \pi_t^X \cdot \prod_{k=1}^K \pi_{k,1,t}^X \delta_{k,1,n} \cdot \pi_{k,2,t}^X \delta_{k,2,n} \quad (1)$$

ここで、 $0 \leq \pi_t^X$, $\sum_{t=1}^X \pi_t^X = 1$,

$$0 \leq \pi_{k,1,t}^X, \pi_{k,2,t}^X, \pi_{k,1,t}^X + \pi_{k,2,t}^X = 1$$

式(1)で示すように、所属クラス t のもとでは、クラス t の構成比率にクラス t に固有な訪問率・非訪問率を乗じることによって同時確率： $P_{n,t}$ が求められるとする。これより、多数の組み合わせがある訪問パターンを、式(1)で定義した尤度（類似度）に基づき少数パターンに集約するのが潜在クラス分析といえる。なお、式(1)から、訪問率についてゾーン間の交互作用を考慮しないと、クラスの構成比率、訪問率、非訪問率は、訪日外国人旅行者全てで共通とする。

さて、式(1)のパラメータ（ π ）は、式(2)で示される尤度（ L ）の最小化によって推定することができる。

$$L^2 = \sum_{n,t} N \cdot P_{n,t} \cdot \log_e \frac{N \cdot P_{n,t}}{N \cdot \hat{P}_n} \quad (2)$$

ここで、 N ：有効サンプル数

\hat{P}_n ：個人 n の $P_{n,t}$ 最大値

この式(2)は、尤度比 χ^2 乗統計量であり、実績値に推定値の当てはまりが良くなるほど小さくなることが知られている。なお、 $P_{n,t}$ は訪問の実績データならびに推定されるパラメータから導出される確率である。また、 \hat{P}_n は個人 n の $P_{n,t}$ の最大値である。

さて、これらのパラメータの導出方法であるが、構成比率： π_t^X と訪問率パラメータ： $\pi_{k,1,t}^X$, $\pi_{k,2,t}^X$ の2種類があるため、最尤法で推定することが困難な場合がある。そこで、EM アルゴリズム（Expectation-Maximization-step）を用いて推定を行う。まず、 π_t^X に適当な初期値を与え、その条件下で式(2)の最小化をニュートンラフソン法によって行い、訪問率パラメータを導出する。この訪問率パラメータに尤度を乗じることによって π_t^X を更新し、更新前後の差分が所定の値以下の場合に収束されたものと判断するが、これはベイズ更新の考え方に基づくものである。つまり、観測された不完全データをいったん最尤方程式になじみのよい「完全データ」に擬似的に置き換え（E-step）、この「擬似的完全データ」を用いてパラメータの擬似的最尤推定値を求め（M-step）、さらに得られた推定値から擬似的完全データを作り直し、それを用いてまたパラメータ推定値を求めなおす、という手続きの繰り返しといえる。EM アルゴリズムでは、繰り返し最適化を行う必要があるため推定時間が長くなるが、反復ごとに尤度が改善され、一般的条件の下では収束が保障されている。局所最適解が存在する場合には、複数の初期値を用いて推定を行うことで解決できるとされている。

この方法では、複数のクラス数を設定して、各々で推定しながら最適なクラス数を判断することになる。その際は、尤度と自由度（ df ）から導かれる赤池情報量規準（AIC）、ベイズ情報量規準（BIC）の値が最小になるクラス数を最適クラス数と判断する。

$$AIC = L^2 - 2 \times df \quad (3)$$

$$BIC = L^2 - \log_e N \times df \quad (4)$$

この2つの指標は、モデルの複雑さとデータとの適合度とのバランスを取るものと考えられる。なお、各パラメータの t 値も算出されるものの、パラメータ全体を1つの組として χ^2 乗統計量によって p 値を算出して、有意水準を判定することとする。

4. 訪問パターンの推定結果並びに検証

(1) 使用データ

今回の分析では訪日目的が「業務」の場合、訪問地に制約が課されていると考え、訪日目的を業務ならびにト

ランジット、その他を除く「観光・レジャー」に限定したところ、29,067サンプルが抽出できた。

さて、「訪日外国人消費動向調査」では日本滞在中における訪問地の設問があり、最大 10 箇所を回答できる。回答は、地名を記入するようになっているが、全国的な訪問傾向の把握を念頭に、それらを都道府県区分に変換し、さらに各都道府県の訪問回数を考慮しない訪問有無のみにもとづくダミー変数にデータを変換した。

訪問パターンは、訪問の有無に加え、その順序自体を考慮することも考えられるが、本研究では訪問順序は捨象する。これは、テキストマイニングの分野において、単語の順序を無視し、文書を単語の集合として捉える bag-of-words (BOW) の考え方と同一である。

以上の結果、有効サンプル29,067に対して、47都道府県の訪問有無の組み合わせは2,191通り出現し、上位から1位：東京都のみ（14%）、2位：北海道のみ（10%）、3位：大阪府のみ（5%）、4位：沖縄県のみ（5%）、5位：大阪府・京都府（4%）となった。上位10位で累積確率50%を超えるに留まるため、それ以降は多様な訪問パターンが存在するといえる。また、表-1は、立寄り都道府県数分布を示しており、2以上が59%を占める。（都道府県別立寄り率（実績値）は表-3に示す。）

なお、分析対象サンプルの属性別構成比率等は、報告書¹⁾に掲載されているため、分析と関連する部分のみ後段で示すこととする。

(2) 推定結果

上述したデータを用いて、式(2)の最小化を行いながら、パラメータの推定を行った。パラメータ推定は、クラス数 1 から 25 まで各々実施している。表-2 は 15 クラス以上を示しているが、AIC, BIC（情報量基準）はモデルの適合度を示すもので、小さいほうが好ましいと判断できる指標である。なお、AIC はクラス数の多いモデルを、BIC はクラス数の少ないモデルを支持する傾向になると言われている。また、 p 値は、パラメータの組についての尤度比 χ^2 乗検定の有意水準を示しており、0.05 以下の場合に「全クラスそれぞれでゾーン別訪問度数と推定値に差がない」との仮説を棄却できることになる。

表-2 より、BIC で 16 クラス、AIC で 24 クラスが妥当とされるが、より細かな組み合わせを考慮するため、24 クラスが妥当と判断した。なお、分析結果を用いた観光振興策立案までを考えたクラス数の決定手順として、(1)モデル評価指標によるクラス数の決定、(2)各クラスの解釈、(3)実務などへの落とし込み、(4)クラス数決定へのフィードバック、といった手順が想定でき、AIC, BIC 以外の観点からクラス数を決定することも考えられる。本論文では、訪問パターンの抽出に潜在クラスを用いた試みと位置づけできるため、モデルの適合度である

表-1 立寄り都道府県数分布

立寄り都道府県数	サンプル数	構成比率
1	11,800	41%
2	6,857	24%
3	5,186	18%
4	2,929	10%
5	1,385	5%
6	515	2%
7	245	1%
8	100	0%
9	36	0%
10	14	0%
合計	29,067	100%

表-2 立寄り都道府県数分布

	p 値	AIC	BIC
15クラス	1.00	-36,980	-271,627
16クラス	1.00	-38,203	-272,453
17クラス	1.00	-38,317	-272,168
18クラス	1.00	-38,733	-272,188
19クラス	1.00	-38,922	-271,979
20クラス	1.00	-38,883	-271,543
21クラス	1.00	-39,159	-271,421
22クラス	1.00	-39,540	-271,405
23クラス	1.00	-39,510	-270,978
24クラス	1.00	-39,740	-270,811
25クラス	1.00	-39,690	-270,363

(1)によってクラス数を決定することとし、(3)等の考慮は今後の課題とする。

そこで、表-3 に示す 24 クラスのパラメータを検討する。表-3 の 2 行目に示した「構成比率」は、構成比率： π_t^x に相当する。24 クラスそれぞれの割合を示し、その合計は 100(%)となる。次に、3 行目以下に示した数値が、訪問率(%)： $\pi_{k,1,t}^x$ である。赤が濃いほど、訪問率が高いことを示しているが、例えば、クラス 1（以下、C1）は「北海道：100(%)」となっており、C1 に含まれるサンプルの 100%が北海道内を 1 箇所以上訪問していると判断する。各クラスで都道府県ごとの高い訪問率（1 地方単独訪問型は 25%以上、複数地方訪問型は 30%以上と設定）をもとに、C1「北海道のみ」から C24「四国・広島」まで名称を付した。

表-3 から、C1「北海道のみ」、C2「東京のみ」、C6「大阪・京都」、以上の 3 クラスは 10%以上の割合を占めており、訪日における主要訪問パターンと考えられる。さらに、C1 から C12 までは訪問率 33%以上が「1 地方のみ」（構成比率：78%）に限定されるのに対して、C13 から C24 までは複数地方にまたがる訪問パターン（構成比率：22%）に属する。複数地方周遊型パターンでは東京、京阪への訪問が基調になっている（C15-C20）以

表-3 潜在クラスモデル結果(クラス別構成比率・ゾーン訪問率)

クラス	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	訪問率
	北海道のみ	南関東	東京のみ	東京・神奈川・長野	愛知のみ	大阪・京都	京阪奈・兵庫	福岡のみ	福岡・熊本・大分	福岡・長崎・熊本・大分	九州全域(佐賀を除く)	沖縄のみ	北海道・東京・長野・京阪	北東北・東京	南関東・山梨・愛知・京阪	東京・京阪1	東京・京阪2	南関東・石川・岐阜・京阪・広島	東京・京阪・広島・福岡	東京・京阪・北部九州	昇龍道	東京・昇龍道・大阪	京阪・岡山・広島	広島・四国	
構成比率	11	7	21	4	1	11	7	4	4	4	1	5	0	1	6	5	4	1	1	0	1	1	1	0	
北海道	100	3	2	0	1	1	0	0	0	0	0	0	48	14	1	13	1	2	6	0	0	0	0	0	13
青森県	0	0	0	0	0	0	0	0	0	0	0	0	28	27	0	0	0	0	0	0	0	0	0	0	0
岩手県	0	0	0	1	0	0	0	0	0	0	0	0	12	26	0	0	0	1	0	0	0	0	0	0	0
宮城県	0	0	0	0	0	0	0	0	0	0	0	0	21	56	0	1	0	1	2	0	0	0	0	0	1
秋田県	0	0	0	0	0	0	0	0	0	0	0	0	21	20	0	0	0	0	0	0	0	0	0	0	0
山形県	0	0	0	1	0	0	0	0	0	0	0	0	7	17	0	0	0	0	0	0	0	0	0	0	0
福島県	0	0	0	0	0	0	0	0	0	0	0	0	10	8	0	0	0	0	0	0	0	0	0	0	0
茨城県	0	0	0	3	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	0	2	0
栃木県	0	1	2	11	1	0	0	0	0	0	0	0	15	8	0	1	8	12	10	0	0	0	0	0	2
群馬県	0	1	0	7	0	0	0	0	0	0	0	0	4	0	0	0	1	1	0	0	1	3	0	2	1
埼玉県	0	1	1	7	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	1	0	0	1
千葉県	2	100	0	11	0	0	0	0	0	0	0	0	13	3	45	2	10	9	9	1	0	7	4	0	11
東京都	8	85	100	66	0	0	3	3	0	0	7	3	74	34	91	97	96	98	94	36	1	30	11	7	48
神奈川県	0	29	17	29	1	0	0	0	0	0	0	0	7	0	41	17	26	35	21	3	0	0	0	2	12
新潟県	0	0	1	5	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	4	0	0	0	0	1
富山県	0	0	0	2	0	0	0	0	0	0	0	0	7	1	0	0	1	0	0	0	75	40	0	0	2
石川県	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	4	45	7	1	69	40	1	0	2
福井県	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	3	2	1	0	0
山梨県	0	21	8	13	0	0	0	0	0	0	0	0	7	0	49	12	10	18	14	1	0	9	0	3	8
長野県	0	3	4	27	0	0	0	0	0	0	1	0	40	1	0	1	5	26	13	0	71	45	1	0	4
岐阜県	0	0	0	2	21	0	1	0	0	0	0	0	23	0	1	4	3	76	16	0	89	58	3	0	3
静岡県	0	3	2	8	2	0	0	0	0	0	0	0	4	2	10	6	4	3	3	7	0	2	0	5	2
愛知県	0	1	0	3	100	2	2	0	0	0	1	0	11	1	46	19	8	13	6	4	95	46	6	0	7
三重県	0	0	0	1	10	1	0	0	0	0	0	0	4	0	0	1	1	3	3	0	0	8	3	0	1
滋賀県	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	1	1	0	4	0	0	0
京都府	0	1	5	0	13	42	91	0	0	0	0	0	37	0	60	40	97	93	86	58	0	22	33	9	24
大阪府	1	0	0	1	7	89	94	2	2	0	3	1	30	1	95	66	50	40	42	79	1	36	48	18	29
兵庫県	0	0	0	0	0	14	50	0	0	0	0	0	7	0	5	4	8	8	8	27	0	1	21	27	6
奈良県	0	0	0	0	2	4	44	0	0	0	0	0	12	0	5	0	27	10	15	11	0	2	4	2	5
和歌山県	0	0	0	0	1	6	6	0	0	0	0	0	7	0	0	1	6	16	7	1	0	0	7	3	2
鳥取県	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	16	0	0
島根県	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	1	0	0	0	13	0	0
岡山県	0	0	0	0	0	0	1	0	0	0	2	0	9	0	0	0	0	6	12	0	0	0	35	29	1
広島県	0	0	0	0	0	1	0	1	0	0	3	0	18	0	0	4	27	55	71	5	0	0	50	44	3
山口県	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	2	0	0	0	6	4	0
徳島県	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	1	0	0	0	0	51	0
香川県	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	3	7	1	0	0	7	71	0
愛媛県	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	7	0	0	0	8	53	0
高知県	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	3	0	0	0	0	43	0
福岡県	0	1	0	0	1	0	0	100	88	95	89	1	8	0	0	2	0	1	38	94	0	1	22	15	13
佐賀県	0	0	0	0	0	0	0	3	1	9	7	0	0	0	0	0	0	0	1	0	0	0	2	0	1
長崎県	0	0	0	0	0	0	0	0	0	81	36	0	6	0	0	0	0	0	26	13	0	0	2	0	4
熊本県	0	0	0	0	0	0	0	0	65	59	70	0	7	0	0	0	0	0	19	46	0	0	0	0	6
大分県	0	0	0	0	0	0	5	88	55	23	0	0	0	0	0	0	0	0	19	49	0	0	2	0	7
宮崎県	0	0	0	0	0	0	0	1	0	1	42	0	0	0	0	0	0	0	2	0	0	0	0	0	0
鹿児島県	0	0	0	0	0	0	0	1	1	0	91	0	11	0	0	0	0	0	16	7	0	0	1	4	1
沖縄県	0	0	0	0	0	0	0	0	0	0	1	100	1	0	0	1	3	1	6	0	0	0	0	2	5
1地方単独訪問型：78.1													複数地方訪問型：21.9												

表4 潜在クラスモデル結果(各都道府県における訪問者のクラス構成比率)

クラス	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
	北海道のみ	南関東	東京のみ	東京・神奈川・長野	愛知のみ	大阪・京都	京阪奈・兵庫	福岡のみ	福岡・熊本・大分	福岡・長崎・熊本・大分	九州全域(佐賀を除く)	沖縄のみ	北海道・東京・長野・京阪	北東北・東京	南関東・山梨・愛知・京阪	東京・京阪1	東京・京阪2	南関東・石川・岐阜・京阪・広島	東京・京阪・広島・福岡	東京・京阪・北部九州	昇龍道	東京・昇龍道・大阪	京阪・岡山・広島	広島・四国
構成比率	11	7	21	4	1	11	7	4	4	4	1	5	0	1	6	5	4	1	1	0	1	1	1	0
北海道	85	1	3	0	0	1	0	0	0	0	0	0	1	1	0	6	0	0	0	0	0	0	0	0
青森県	3	2	6	0	0	0	0	0	0	0	0	0	25	62	1	1	0	0	0	0	0	0	1	0
岩手県	0	1	0	7	0	0	0	0	0	0	0	0	13	76	0	0	1	1	0	0	0	0	0	0
宮城県	0	1	4	0	0	0	0	0	0	0	0	0	10	70	0	9	2	1	3	0	0	0	0	0
秋田県	3	1	4	0	0	1	0	0	0	0	0	0	25	62	0	3	0	0	0	0	0	0	0	0
山形県	0	3	0	11	0	0	0	0	0	0	0	0	11	71	0	1	2	1	0	0	0	0	0	0
福島県	0	1	18	10	0	0	0	0	0	0	0	0	23	46	0	3	0	0	0	0	0	0	0	0
茨城県	2	10	12	48	0	6	0	0	0	0	0	0	2	4	0	4	8	0	3	0	0	0	0	1
栃木県	0	4	23	25	0	0	0	0	0	0	0	0	4	5	0	4	21	6	6	0	0	0	0	0
群馬県	1	7	19	49	0	2	0	0	0	0	0	0	3	0	0	4	7	1	0	0	1	5	0	1
埼玉県	0	7	44	38	0	2	0	0	0	0	0	0	0	2	1	0	3	1	1	0	0	1	0	0
千葉県	2	64	1	4	0	0	0	0	0	0	0	0	0	0	23	1	4	1	1	0	0	1	0	0
東京都	2	13	44	5	0	0	0	0	0	0	0	0	1	1	11	11	8	2	2	0	0	1	0	0
神奈川県	0	18	31	9	0	0	0	0	0	0	0	0	0	0	20	8	9	3	2	0	0	0	0	0
新潟県	0	1	32	39	0	2	0	0	0	0	0	0	14	2	3	0	3	0	0	3	0	1	0	0
富山県	0	0	1	5	0	0	0	0	0	0	0	0	2	1	0	1	2	0	0	0	64	26	0	0
石川県	0	0	2	0	0	0	1	0	0	0	0	0	3	0	0	0	7	18	3	0	45	19	0	0
福井県	0	7	0	0	0	22	0	0	0	0	0	0	8	0	4	6	0	2	0	2	32	12	3	0
山梨県	0	19	21	6	0	0	0	0	0	0	0	0	0	0	35	8	5	2	2	0	0	1	0	0
長野県	0	5	19	23	0	1	0	0	0	0	0	0	4	0	1	2	5	5	3	0	22	11	0	0
岐阜県	0	0	0	2	6	0	2	0	0	0	0	0	3	0	1	7	4	19	4	0	35	17	1	0
静岡県	0	10	20	13	1	0	0	0	0	0	0	0	1	1	27	14	7	1	1	1	0	1	0	0
愛知県	0	1	0	2	13	2	2	0	0	0	0	0	1	0	35	14	4	1	1	0	17	6	1	0
三重県	0	0	0	6	17	20	4	1	1	0	0	0	3	0	3	9	9	5	5	0	1	14	3	0
滋賀県	0	0	3	0	0	15	53	0	0	0	0	0	0	0	4	3	9	0	2	1	0	10	0	0
京都府	0	0	5	0	1	19	26	0	0	0	0	0	1	0	14	9	17	3	3	1	0	1	1	0
大阪府	1	0	0	0	0	32	22	0	0	0	0	0	0	0	19	12	7	1	1	1	0	1	1	0
兵庫県	0	0	0	0	0	24	54	0	0	0	0	0	0	0	5	3	5	1	1	2	0	0	2	1
奈良県	0	0	0	0	0	9	57	0	0	0	0	0	1	0	5	0	21	2	3	1	0	0	1	0
和歌山県	0	0	0	0	0	40	24	0	0	0	0	0	2	0	1	2	14	8	4	0	0	0	3	0
鳥取県	0	1	1	0	0	7	6	0	0	0	0	0	3	0	3	4	4	4	2	0	0	0	66	0
島根県	0	2	4	0	0	0	0	0	0	0	0	0	5	0	0	0	8	12	6	0	0	0	62	0
岡山県	0	0	2	0	0	0	10	0	0	0	2	0	5	0	0	1	1	8	19	0	0	0	42	10
広島県	0	0	0	0	0	2	1	1	0	0	1	0	2	0	0	6	36	15	21	1	0	0	11	3
山口県	0	2	0	0	0	1	0	23	12	25	3	0	0	0	0	1	0	1	9	0	0	0	19	4
徳島県	0	0	0	0	0	19	0	0	0	0	0	0	9	4	2	0	0	0	4	0	0	0	0	62
香川県	0	0	1	1	0	2	1	0	0	0	0	0	1	0	0	3	11	7	17	1	0	0	14	41
愛媛県	0	0	1	0	0	1	0	0	0	0	0	0	8	0	0	3	0	1	25	0	0	0	20	42
高知県	0	0	0	0	0	9	0	0	0	0	0	0	6	0	0	0	0	0	18	0	0	0	0	67
福岡県	0	0	0	0	0	0	0	28	30	29	4	0	0	0	0	1	0	0	3	3	0	0	1	0
佐賀県	0	0	0	2	0	1	0	19	6	59	7	0	0	0	0	0	0	0	2	0	0	0	3	0
長崎県	0	0	1	0	0	0	0	0	0	85	5	0	1	0	0	0	0	0	6	1	0	0	0	0
熊本県	0	0	0	0	0	0	0	0	48	39	6	0	0	0	0	0	0	0	3	3	0	0	0	0
大分県	0	0	0	0	0	0	0	3	57	32	2	0	0	0	0	0	0	0	3	3	0	0	0	0
宮崎県	0	0	0	4	0	0	1	8	3	7	68	0	0	0	0	3	0	0	5	0	0	1	0	0
鹿児島県	0	0	0	2	0	0	0	5	6	0	56	0	5	1	0	2	0	0	17	3	0	0	1	1
沖縄県	0	0	0	0	0	0	0	0	0	0	0	93	0	0	0	1	2	0	1	0	0	0	0	0

外に、C13のような北海道から京阪神まで広範な訪問となっている「北海道・東京・長野・京阪」パターンやC14「北東北・東京」パターンで示される東日本を中心としたものがみられる。さらに、C21、C22は中部地方の愛知県・岐阜県・富山県・石川県を南北に縦断する「昇龍道」ルートが抽出されたり、大阪から西側地域を中心とした周遊であるC20「東京・京阪・北部九州」や京阪神に中国地方が加わったC23「京阪・岡山・広島」、C24のような四国が集塊として抽出された「広島・四国」などがみられる。

また、推定結果と比較して、都道府県別訪問率（実訪問者数/全サンプル数）の実績値(%)を最右欄に示している。訪問率実績値と値の大きいパラメータが類似して出現しており、多くの訪問者を数える訪問地がどのようなクラスに含まれるのか判断することができる。

これらの結果の妥当性検証であるが、(1) p 値=1.00より「クラス別都道府県訪問数（実績値）と推定値に差がない」との帰無仮説を採択できること、(2) クラス別構成比率の実績値と推定値との相関係数が0.992であることがあげられる。さらに、(3) 表-3を用いた推定値から、クラス別構成比率に訪問率を乗じると全サンプルに占める訪問比率を算出できる。そして判別された所属クラスの訪問比率実績値との比較から誤差を算出することができる（式(5)）。この値の標準偏差を算出したところ0.0026となり、大きな誤差を含んでいないと考えられる。

$$\text{訪問比率誤差の標準偏差} = \sqrt{\frac{\sum(\text{実績値}-\text{推定値})^2}{\text{セルの数}}} \quad (5)$$

以上より、クラス数は24クラスが妥当と判断でき、さらに推定値誤差の標準偏差から推定されたパラメータは適当と判断した。

さて、表-3は列方向に着目するもので、需要側・旅行者の訪問特性を示したものとイえる。それに対して、供給側・着地サイドへの知見について考える。例えば、「北海道を訪れる主要クラスはどれか」ということを明らかにできれば、マーケティング対象やプロモーション対策が明確にできると考えられる。式(1)より、構成比率： π_t^x に訪問率を乗じて全クラスにおいて正規化を行うと、各都道府県における各クラスの占める割合が導出できる。表-4はその割合を示したものであり、行方向の合計は100(%)となる。なお、2行目は、推定された構成比率： π_t^x を参考として示している。例えば、北海道では、C1「北海道のみ」訪問パターンが85%を占めており、次にC16「東京・京阪1」訪問パターン(6%)が主要なマーケットと分かる。これより、北海道サイドから考えると、主要なマーケットはC1であるが、それにつづ

くC6は東京や京阪の訪問の一環として北海道を立寄っており、それらの地域との周遊性の確保、プロモーション戦略が必要不可欠との考察が可能といえる。表-4全体から、訪問率の実績値が高いところは、1地方単独訪問型と複数地方訪問型訪問パターンのいずれも一定以上の割合を有しており、その地域における多様な観光テーマの設定による訴求力の高さを示していると考えられる。

5. 訪問パターンと個人属性との関連

(1) 訪問パターンの主要国籍・地域別構成比率

旅行者の国籍・地域によって観光行動が異なると考えられる。その場合、国籍・地域ごとに潜在クラス分析を行い、より細かな訪問パターンも抽出できるが、本論文では全サンプルから訪問パターンを抽出することによって、訪日外国人の全体的な傾向把握や国籍・地域別の違いを明らかにすることを意図した。あわせて、層別化によるサンプルの減少によって訪問パターンとして抽出できなくならないように、全サンプルを用いて潜在クラスの抽出を行った。

それをうけて本章では、国籍・地域をはじめ、旅行者の個人属性や旅行特性と訪問パターンとの関連性について検討を進める。

表-5は、サンプル数の上位14国籍・地域を取り上げ、訪問パターン構成比率を示したものである。その導出方法であるが、個々のサンプルの実際の訪問データをもとに、式(1)によって24クラスの中から最も所属確率の高いクラスをそのサンプルの所属クラスとして導いた。

表-5より、欧米・豪州とアジア各国・地域とで異なることがわかる。前者では、C3「東京のみ」とC17「東京・京阪2」の割合が比較的高い。広島を含む、いわゆるゴールデンルートへの訪問が多いのに対して、後者に該当するシンガポール・マレーシアではC1「北海道」が高くなり、自国では体験できない冬・雪が旅行者への訴求になっていると考えられる。また、韓国では、その近接性から九州関連の訪問パターン、中国では典型的なゴールデンルートであるC15が多いなど、差異が確認できる。

さらに、同じ国籍・地域でも訪日回数によって訪問パターンの変化が考えられるため、上位3ヶ国ならびに米国・カナダについて集計を行った。

図-1より、台湾では訪日回数が増加するに従い、C1「北海道」やC2「南関東」、C7「京阪奈・兵庫」が減少するのに対して、C3「東京のみ」やC6「大阪・京都」、C14「北東北・東京」が増加している。東日本大震災などの影響が考えられるものの、訪日回数によって訪問地性向が異なると考えられるため、それらを考慮し

表-5 国籍・地域別クラス構成比率 (%)

		台湾	韓国	中国	香港	タイ	シンガポール	オーストラリア	マレーシア	米国	加	英	仏	露	独	全体
C1	北海道のみ	15	10	10	13	10	24	15	22	2	4	4	1	6	2	11
C2	南関東	13	3	5	8	10	10	4	12	8	9	5	2	10	3	7
C3	東京のみ	22	16	11	30	31	32	32	15	45	40	47	45	47	35	24
C4	東京・神奈川・長野	2	1	1	2	2	2	4	1	3	2	3	1	5	3	2
C5	愛知のみ	1	2	0	1	1	1	0	0	1	0	1	0	0	2	1
C6	大阪・京都	13	19	7	14	8	5	3	15	7	4	4	5	1	6	12
C7	京阪奈・兵庫	9	6	4	5	5	3	1	4	2	2	1	3	1	2	6
C8	福岡のみ	1	13	1	2	1	1	0	0	1	0	1	1	0	1	4
C9	福岡・熊本・大分	2	18	0	3	1	0	0	0	1	1	0	0	0	0	5
C10	福岡・長崎・熊本・大分	5	5	1	2	2	2	0	1	1	0	0	0	0	0	3
C11	九州全域(佐賀を除く)	1	0	0	1	0	1	0	0	0	0	0	1	1	0	0
C12	沖縄のみ	6	3	6	13	0	1	0	1	2	2	0	0	2	0	5
C13	北海道・東京・長野・京阪	0	0	0	0	0	0	2	0	1	1	1	1	0	2	0
C14	北東北・東京	2	0	0	1	2	1	1	0	1	1	1	1	1	0	1
C15	南関東・山梨・愛知・京阪	0	0	39	1	5	5	2	14	2	2	2	2	2	2	6
C16	東京・京阪1	1	1	12	2	12	5	11	8	7	9	8	7	9	7	5
C17	東京・京阪2	0	0	2	0	4	3	12	2	11	14	14	18	12	19	3
C18	南関東・石川・岐阜・京阪・広島	0	0	0	0	2	0	6	0	3	4	4	5	0	4	1
C19	東京・京阪・広島・福岡	0	0	0	0	0	1	3	1	2	4	3	7	1	7	1
C20	東京・京阪・北部九州	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0
C21	昇龍道	4	0	0	1	0	1	0	0	0	0	0	0	0	0	1
C22	東京・昇龍道・大阪	1	0	0	1	1	2	1	0	0	0	1	0	1	1	1
C23	京阪・岡山・広島	1	0	0	0	1	1	1	0	1	1	1	1	0	2	1
C24	広島・四国	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	合計	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	サンプル数	8,040	6,598	3,560	3,204	1,094	720	936	416	1,491	591	559	505	340	286	29,067

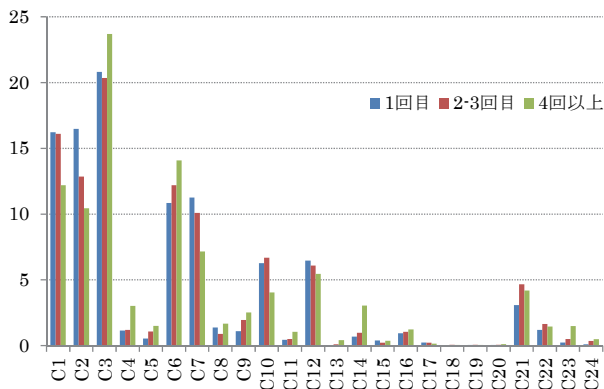


図-1 訪日回数別クラス別構成比率 (縦軸%, 台湾)

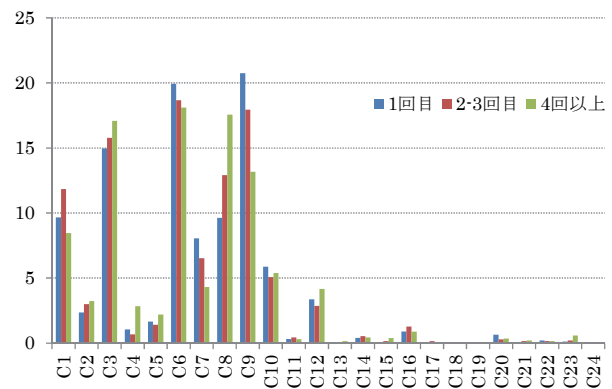


図-2 訪日回数別クラス別構成比率 (縦軸%, 韓国)

たプロモーション、ツアー造成等の必要性が考えられる。この訪日回数が増加するに従い大都市を中心とする訪問パターンに移行する傾向は、韓国（図-2. C3, C8）、中国（図-3. C3, C6）でもみられる。一方、米国・カナダでは、初回訪日においてC3が卓越しているため、徐々に他のパターンが増加する傾向がわかる。

次に、滞在日数や団体ツアーの占める割合、入国時期（四半期別）とクラスとのクロス集計を行った（表-6）。

なお、各項目ごとに数値が大きい場合は赤で、小さい場合は緑で示している。サンプル数が少なくなるため、ここでは国籍・地域を考慮しないが、まず滞在日数をみると C8, C9, C12 で少なく、訪問地点の増加や移動距離が長くなる訪問パターンで日数が増加していることがわかる。また、団体ツアー割合(%)では C21, C15, C22, C1 などが高くなっている。徐々に旅行の FIT (Foreign Independent Tour) 化が進むと言われていることから、こ

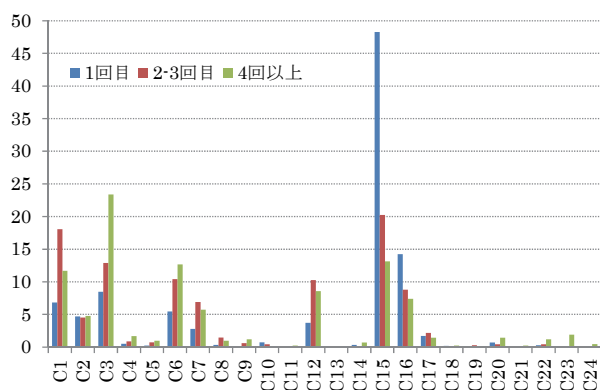


図-3 訪日回数別クラス別構成比率（縦軸:%，中国）

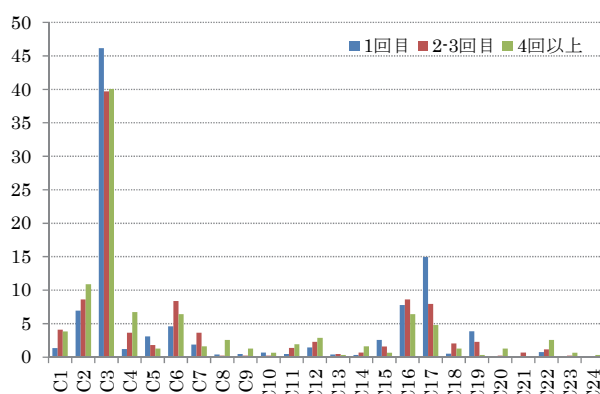


図-4 訪日回数別クラス別構成比率（縦軸:%，米国・カナダ）

これらの訪問パターンが多い地域は、個人旅行に適した環境整備が今後、重要と考えられる。特に、不特定多数を対象としたツアーに満足できず自分で旅程を作成したいというニーズがFIT化の背後にあるとも考えられるため、情報提供や商品の販売などの実務面で考慮が必要といえる。また、各訪問パターンにおける四半期別構成比率をみると、特定の四半期で集中するパターン（C24、C21）や変動の小さいパターン（主に大都市を訪問するパターン）など傾向が異なる。実態の把握に加え、地域資源の活用、プロモーション等についての参考情報としてこれらは位置づけできる。

最後に、サンプル数が1000を超える国・地域の訪日全体に対する満足度（平均値、7段階評価、1：非常に不満～4：どちらともいえない～7：非常に満足）をクラス別に算出した。比較的、時間・費用をかけて訪日する米国・カナダ、英・仏・独の旅行者ではそのコストに見合うようにするためか、満足度が高くなる傾向であるのに対して、隣国である韓国の満足度が低い。また、同一国籍・地域の中でも評価に高低があり、その原因をより細かく分析することが課題として考えられる。

(2) 一般化 χ^2 検定による独立性の検定¹⁴⁾

これまでのクロス集計はおおまかな傾向把握にとどま

表-6 クラス別平均滞在日数(日)・団体ツアー割合・四半期構成比率・満足度

クラス			平均滞在 日数	団体 ツアー 割合	四半期別構成比率				満足度（平均値）								英・仏・独
					Q1	Q2	Q3	Q4	全体	台湾	韓国	中国	香港	米・加	タイ		
C1	北海道のみ		5.3	64	33	18	29	19	6.0	6.0	5.5	6.1	6.1	6.7	6.5	6.8	
C2	南関東		5.2	44	23	26	31	20	6.4	6.3	6.1	6.3	6.3	6.8	6.5	6.7	
C3	東京のみ		6.2	13	28	25	23	23	6.5	6.4	6.1	6.4	6.4	6.8	6.6	6.8	
C4	東京・神奈川・長野		9.3	25	35	19	18	27	6.5	6.4	6.3	6.4	6.6	6.9	6.6	6.7	
C5	愛知のみ		6.0	23	24	21	28	28	6.2	6.2	5.9	5.8	6.3	6.9	6.8	6.8	
C6	大阪・京都		5.1	21	22	25	28	25	6.3	6.4	6.0	6.4	6.3	6.8	6.7	6.7	
C7	京阪奈・兵庫		5.5	42	20	25	27	29	6.3	6.3	6.2	6.4	6.2	6.6	6.6	6.7	
C8	福岡のみ		2.9	18	24	21	28	28	5.9	6.3	5.8	6.4	6.4	6.7	6.4	6.4	
C9	福岡・熊本・大分		2.9	56	28	25	18	29	5.9	6.3	5.8	6.1	6.3	6.4	6.1	7.0	
C10	福岡・長崎・熊本・大分		4.1	63	24	28	26	22	6.1	6.2	5.9	6.2	6.3	6.7	6.3	7.0	
C11	九州全域(佐賀を除く)		6.6	63	17	24	33	26	6.2	6.1	6.0	7.0	6.3	6.5	-	6.0	
C12	沖縄のみ		3.8	50	23	28	30	19	6.0	6.0	5.6	6.0	6.0	6.5	-	6.0	
C13	北海道・東京・長野・京阪		21.0	16	21	28	19	33	6.8	6.4	7.0	-	-	6.9	6.8	7.0	
C14	北東北・東京		7.5	50	26	17	21	35	6.3	6.2	6.3	5.6	6.1	6.7	6.7	6.8	
C15	南関東・山梨・愛知・京阪		6.2	84	18	31	39	13	6.3	6.4	6.5	6.2	6.0	6.5	6.7	6.7	
C16	東京・京阪1		8.8	34	24	23	29	23	6.5	6.4	6.1	6.3	6.4	6.8	6.6	6.9	
C17	東京・京阪2		12.6	18	12	30	25	32	6.8	6.8	7.0	6.4	6.8	6.9	6.9	6.8	
C18	南関東・石川・岐阜・京阪・広島		14.8	29	6	28	23	43	6.8	6.8	-	6.5	6.0	7.0	6.5	6.8	
C19	東京・京阪・広島・福岡		19.7	15	17	29	23	31	6.8	5.6	6.0	7.0	7.0	6.8	7.0	6.8	
C20	東京・京阪・北部九州		11.4	45	21	21	34	23	6.4	6.7	6.0	6.4	6.5	6.6	6.7	6.7	
C21	昇龍道		4.5	89	12	45	16	27	6.3	6.2	6.2	7.0	6.3	-	-	-	
C22	東京・昇龍道・大阪		6.1	67	15	35	22	28	6.4	6.3	6.0	6.4	6.3	6.8	6.3	6.8	
C23	京阪・岡山・広島		10.1	20	25	23	22	30	6.3	6.3	5.4	6.2	6.4	6.6	6.8	6.6	
C24	広島・四国		10.7	39	10	10	21	59	6.4	6.1	-	6.5	6.1	7.0	6.5	6.7	
サンプル全体			6.0	38	25	25	25	25	6.3	6.2	5.9	6.2	6.2	6.8	6.6	6.8	

るものである。そこで、国籍・地域ごとに旅行時期（四半期区分）、訪日回数（初回、2-3回、4回以上）、訪日経験（初回訪日、2回以上）、旅行形態（団体・個人）、旅行者の年齢階層（29歳以下、30歳代、40歳代、50歳代、60歳以上）といった要因と24訪問パターンとの関連性チェックのため、独立性の検定（ χ^2 検定）を行った。さて、比較的多くのサンプルを用いた場合、通常の独立性の検定では僅かに差異があるだけで統計的に有意という結果が導かれる。使用したデータは、訪日外国人旅行者からのサンプリングであるとともに、導出した訪問パターン自体も推定時の初期値設定で異なる結果になる恐れがあるため、それぞれに誤差があることを考慮することが好ましいと考えた。

その対処方法として、推定値の信頼区間算出があるが、独立性検定のような母数推定にはなじまない。また、有意水準をより低く設定することも考えられるが、本論文では医療統計の分野から提案されている一般化 χ^2 独立性検定 (the generalized χ^2 independence test) を用いるものとする。

まず、一般化 χ^2 独立性検定の考え方を概説すると、通常、Pearsonの独立性検定はPearsonの χ^2 統計量を「逸脱量」として用いるが、本検定では、 χ^2 統計量を標本数 n で除した値を用いる（式(6)）。

$$\frac{\chi^2}{n} = \frac{1}{n} \sum_i \frac{(F_i - E_i)^2}{E_i} = \sum_i \frac{\left(\frac{F_i}{n} - P_i\right)^2}{P_i} \quad (6)$$

ここで、 n ：サンプル数、

F_i ：実績値（ゾーン別クラス別度数）

E_i ：モデルから導かれる期待値

P_i ：モデルから導かれる確率

母集団における真の逸脱量を d と表すとする、通常の独立性検定(A)、ならびに一般化 χ^2 独立性検定(B)では、下記のように設定される。

(A) 帰無仮説： $d = 0$ ，対立仮説： $d > 0$

(B) 帰無仮説： $d \leq d_0$ ($d_0 > 0$)，対立仮説： $d > d_0$

一般化 χ^2 独立性検定では、Pearsonの χ^2 統計量を非心 χ^2 分布 (non-central χ^2 -distribution) に照らし合わせることで検定を行うことができる。通常の χ^2 分布 (central χ^2 -distribution) では平均値がその分布の自由度 ν と一致するのに対して、非心 χ^2 分布は中心が偏っている χ^2 分布であり、その分布の非心度 (non-centrality)： δ を加えた $\nu + \delta$ が平均値となる。サンプル数 n が十分に大きいとき、非心度 δ と逸脱量 d との間には $d = \delta/n$ の関係が成り立つ。従って、母集団の真の逸脱量 d が d_0 以下

表-7 24クラスと各要因との独立性の検定結果

	旅行時期	訪日回数	旅行形態	年齢階層	訪日経験
台湾	×	×	○	×	×
韓国	×	×	○	×	○
中国	○	○	○	×	○
香港	×	×	○	×	×
米国	×	×	○	×	×
タイ	○	×	○	×	×

凡例 ○:5%有意水準で仮説を棄却，×:採択

表-8 瀬戸内地域への訪問主要クラスの構成比率

クラス	C6	C7	C8	C9	C10	C17	C18	C19	C23	C24	その他
サンプル全体	11%	7%	4%	4%	4%	4%	8%	9%	7%	2%	0%
兵庫県	24%	54%	0%	0%	0%	5%	1%	1%	2%	1%	0%
岡山県	0%	10%	0%	0%	0%	1%	8%	19%	42%	10%	0%
広島県	2%	1%	1%	0%	0%	36%	15%	21%	11%	3%	0%
山口県	1%	0%	23%	12%	25%	0%	1%	9%	19%	4%	0%
徳島県	19%	0%	0%	0%	0%	0%	0%	4%	0%	62%	0%
香川県	2%	1%	0%	0%	0%	11%	7%	17%	14%	41%	0%
愛媛県	1%	0%	0%	0%	0%	0%	1%	25%	20%	42%	0%

という帰無仮説が正しいとするならば、その母集団から選ばれた n 人のサンプルが示すPearsonの χ^2 統計量は、非心度 $\delta = n \cdot d_0$ の非心 χ^2 分布に近似するとされている。非心 χ^2 分布への近似はサンプル数 n がおおよそ100以上あれば成り立つといわれている。

次に、一般化 χ^2 独立性検定の具体的な手続きについて概説する。まず、どの程度の逸脱量ならば意味のない逸脱とみなすのか、 d_0 の値を定める必要がある。表-2を用いた推定値と、所属クラスを判別された各サンプルの実績値から、推定値の誤差の標準偏差を算出したところ0.0026であったため、本分析では式(6)における $(F_i/n - P_i) = 0.0026$ と設定した。また、分母： P_i は表-3で示された構成比率推定値を用いた。このような考えに基づき、 d_0 を設定した。これらの設定により導かれる χ^2 統計量ならびに非心度 δ ($\delta = n \cdot d_0$) から、非心 χ^2 分布に照らし合わせて、確率 p を算出し、5%有意水準によって検定を行った。

表-7より、旅行時期では中国とタイ、訪日回数では中国、訪日経験の有無では韓国、中国、さらに旅行形態（団体・個人）では6カ国・地域全てで独立性が棄却され、両者に関連性がある可能性が高いと推察できる。

(3) 分析結果の活用

以上の結果から、インバウンド振興に対する考察を行う。ここでは、ハード的整備ではなく、地域の連携のあり方といったソフト的対応に焦点を当てることとするが、その背景としては複数の主体が連携して観光振興に取り組むDMO (Destination Management/Marketing Organization) が着目されていることがある。すなわち、個別の地域、個別主体よりも、互酬性のもとで複数主体によって広域の

連携をはかり、組織として収益を内生化しながら過少となる観光投資を適正水準に高め、それと相まって魅力的な地域づくり、来訪者の満足度向上に取り組む中で、どのようにゾーニングしながら、観光マーケットを捉まえるのかが重要との問題意識に基づく^{15,16)}。

さて、ここでは国土交通大臣に認定された広域周遊ルートの中から瀬戸内・海の道を取り上げ、中国四国7県（兵庫県、岡山県、広島県、山口県、徳島県、香川県、愛媛県）の観光振興方策を検討した。一般的に、地域振興を考える場合、旅行者に対する訴求を地域資源のみで検討する事例がある。しかしながら、旅行者自身は時間・予算制約の中で訪日意向・動機・目的に合致する、より好ましいいくつかの訪問先を決定していると考えられるため、訪日時の訪問先を包括的に把握し、それに対応した周遊ルートや訴求テーマの設定が望ましいといえる。

図-5～7 は、瀬戸内地方での訪問率が比較的高いクラスを示したものである。特に C24 は四国や広島への訪問率が比較的高いため、7 県を包括的にマネジメントする DMO にとって地域的な広がりで整合性があり、有望なマーケットと位置づけられる。しかしながら、C24 は全体の 0.21%を占めるに留まるため、必ずしも大きな効果は考えづらい。さて、表-8 は、表-4 より当該地域部分を抜き出したものであるが、先ほどの C24 は徳島県の訪問者の 62%を占めていること、岡山県、香川県、愛媛県では C19、C23 も一定規模の構成比率を占めることがわかるが訪日全体の割合はいずれも 1%と低いことがわかる。

そこで、サンプル全体の構成比率がある程度みられる C6、C7、C17、C18 は中国四国 7 県におけるクラス別構成比率も一定数存在することから、これらの訪問パターンに着目し、共通の観光テーマ設定、周遊コースの造成などをはかり、7 県全体への周遊を誘発する取り組みが考えられる。一方、C8、C10 は山口県への訪問者においてある程度の構成比率を占めることがわかる。しかしながら、表-3 で確認すると、山口県以外他 6 県への立寄りにつながっておらず、むしろ九州地方への訪問率が高いことから、中国四国としてのシナジーが現状では発揮されていないといえる。そのため、九州地方の観光テーマとの連携を図るなどして、九州地方への訪問者に対して周遊を誘発する仕掛けを考える必要があるといえる。なお、これら訪問パターンに占める国籍では米国、オーストラリア、英国などが高いことから、それら訪日外国人をターゲットとすることが考えられる。

このように、訪問パターンの分析は単一地域への訪問の有無だけでは把握できない地域間の連関、訪問傾向を把握することができ、観光振興のためのゾーニングの検討、テーマ設定に資することができよう。この観点から、

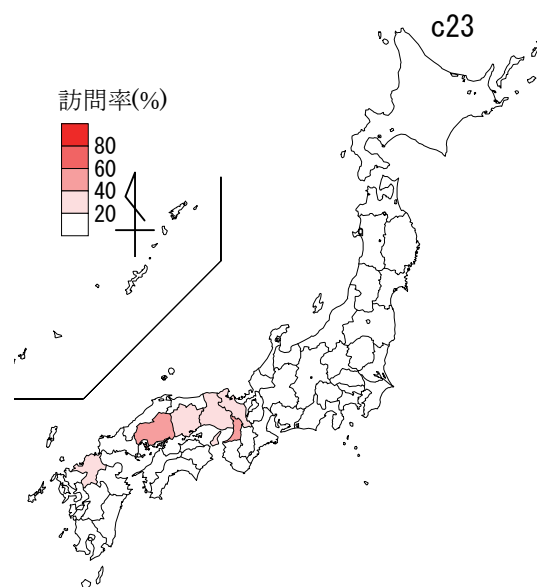


図-5 C23 の都道府県別訪問率

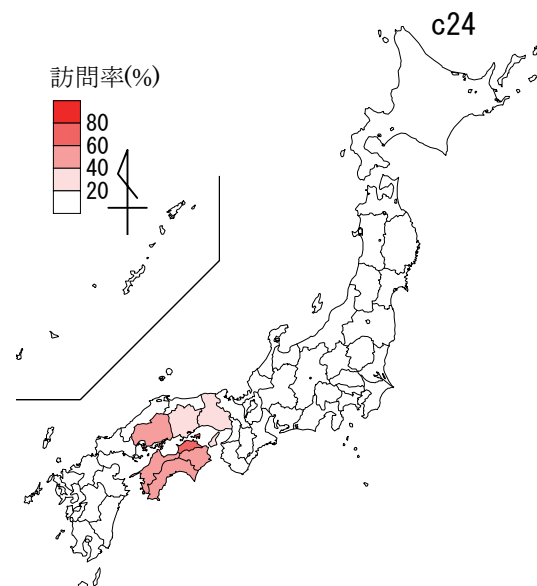


図-6 C24 の都道府県別訪問率

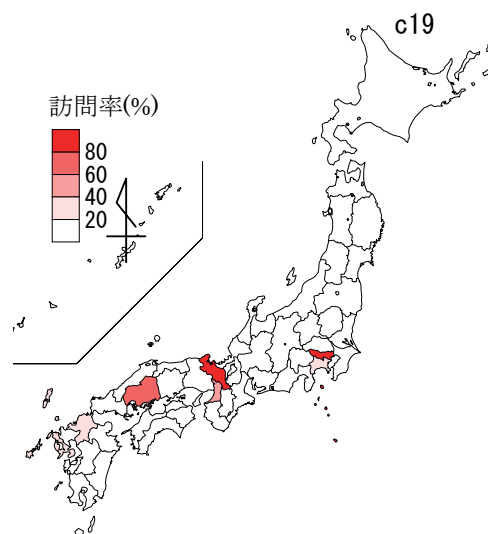


図-7 C19 の都道府県別訪問率

中国・四国地域を限定的に考えるのではなく、京阪との連関が強さ、ゴールデンルートとの結び付きの強さを抽出でき、それらの地域と関連付けたプロモーションが必要不可欠であることが考察できた。また、中国・四国のすべてを訪問することに加え、そのいずれかに訪問したその次の機会に中四国全域に誘うような観光戦略も考えられる。これらから、訪日外国人旅行者の訪問パターンを地域にとらわれず包括的に行う意義を見出すことができる。施設整備のみならず地域連携のあり方から地域への誘客、地域振興に資する情報提供の一例と位置づけできる。

6. まとめ

本研究は、観光・レジャー目的の訪日外国人旅行者の訪問パターンの特性把握を目的として、「訪日外国人消費動向調査（平成 24, 25 年）」データを用いて分析を行った。数多く存在する訪問パターンの把握・集約のために潜在クラスモデルを用いて分析を行ったところ、24 クラスの設定が最も妥当と判断できた。分析結果より、1 地方のみの訪問が 78%を占めるのに対して、複数地方にまたがる周遊型パターンは 22%であること、サンプル数の多い国籍・地域に着目した分析では、台湾、韓国、中国でクラス別構成比率が異なること、特に中国で訪日回数別に構成比率が大きく異なることを確認できた。また、一般化 χ^2 検定から訪問パターンへの影響要因として、旅行形態、旅行時期、訪日回数、訪日経験が該当することを明らかにした。

類型化が困難であった数多くの組み合わせを適当に区分できたことにより、各訪問パターンと国籍・地域、訪日経験などの来訪者属性との関連性把握が容易になり、来訪者へのプロモーション活動への示唆、ゾーンから見たターゲット設定に対して有益な情報になると考えられる。そして、本分析はビッグデータを活用した事例として位置づけできるとともに、継続的なデータ収集により時系列変化の把握、新たな訪問パターンの抽出などの可能性も考えられる。

今後の課題として、まず分析手法の改善が考えられる。本論文では、全サンプルを分析した結果を用いながら、国籍・地域別の違いを明らかにした。国籍・地域と訪問パターンとに関連があるとすると、国籍・地域を予めクラス導出に用いる改良も考えられ、それが可能である共変量を用いた潜在クラスモデルの適用が考えられる。また、すべての被験者が同一のクラス別構成割合を有するとする強い仮定であることとともに、モデル推定においてパラメータの事前確率を想定していないことから、データによる過学習（overfitting）の恐れがあるなどの問題

点への対処、クラス間の類似性を考慮したクラス導出など分析ツールとしての改良が考えられる。

次に、サンプル数とゾーン区分、クラス数の設定については、継続的な検討が必要といえる。研究の目的からゾーン区分についてはある程度設定できるが、より多くのクラス数を設定すると実務面における対策立案をはじめとする検討が困難になる一方、少ないクラス数であると旅行行動の分解能が粗くなるトレードオフが存在する。そのため、数多くの分析事例の比較を通じて、これら 3 者の関連性を明確にする必要がある。

そして、最後の課題としては分析結果を活用したより効果的なプロモーションの実施、効果的な観光振興策への落とし込みが考えられる。

謝辞：本研究の分析にあたり、データを提供いただきました観光庁ならびに有益なコメントを頂きました西井和夫教授（流通科学大学）、岡本直久教授（筑波大学）をはじめとする関係諸氏に深謝の意を表します。

参考文献

- 1) 観光庁 HP, 訪日外国人消費動向調査, 2012, 2013, <http://www.mlit.go.jp/kankoch/siryoutoukei/syouthityousa.html> (2016.2.6 閲覧)
- 2) 古屋秀樹, 西井和夫, 野瀬元子, 呉戈, 金賢: 日本・中国・韓国における観光行動の実態分析, 第 37 回土木計画学研究・講演集, CD-ROM, 2008.
- 3) 日比野直彦, 森地茂, 島田貴子: 居住地域別訪日中国人旅行者の日本国内における観光行動—インバウンド戦略検討のため基礎的分析—, 交通学研究, Vol. 54, pp. 55-64, 2011.
- 4) 観光庁 HP, 観光ビッグデータを活用した観光振興/GPS を利用した観光行動の調査分析, 2015, <http://www.mlit.go.jp/kankoch/shisaku/kankochi/gps.html> (2016.2.6 閲覧)
- 5) 観光庁, 日本版 DMO とは, 2015, http://www.mlit.go.jp/kankoch/page04_000048.html (2016.4.23 閲覧)
- 6) 小林克行, 西村雅史, 岩倉成志: 観光需要におけるトライアル層とリピート層の嗜好性に関する考察, JRail 2005, pp. 407-410, 2006.
- 7) 谷下雅義: 世帯の多様性を考慮した人口密度が自動車走行量に及ぼす影響分析—オーナーインタビュー調査を用いた潜在クラス回帰分析—, 第 42 回土木計画学研究発表会・講演集, No. 176, CD-ROM, 2011.
- 8) 古屋秀樹, 全相鎮: 潜在クラスモデルを用いた宿泊観光旅行回数の基礎的分析, 第 50 回土木計画学研究発表会講演集, CD-ROM, 2014.
- 9) 古屋秀樹: 旅行要因の関連性を考慮した宿泊観光旅行のクラス構築に関する基礎的分析, 都市計画論文集, Vol. 50, No. 3, pp. 337-344, 2015.
- 10) 三輪哲: 潜在クラスモデル入門, 理論と方法, Vol. 24, No. 2, pp. 345-356, 2010.
- 11) 里村卓也: マーケティング・モデル(第 6 章 セグメンテーションと潜在クラスモデル), 共立出版, 2015.
- 12) 劉瑜娟, 古屋秀樹: 潜在クラス分析を用いた訪日外

- 客の訪問パターンに関する基礎的分析，第 52 回土木計画学研究発表会講演集（277），No. 52, 2015.
- 13) 古屋秀樹，劉瑜娟：訪日外客の 47 都道府県の訪問パターン分析，日本観光研究学会第 30 回全国大会研究発表論文集，CD-ROM, 2015.
- 14) 保田時男：大規模サンプルに対する一般化 χ^2 適合度検定－JGSS データへの適用例－，JGSS 研究論文集，No. 3, pp. 175-186, 2004.
- 15) 観光庁，日本版 DMO，http://www.mlit.go.jp/kankocho/page04_000053.html（2016.2.6 閲覧）
- 16) 瀬戸内ブランド推進連合事務局 HP，瀬戸内ブランドプロジェクト，<http://www.setouchiweb.jp/>（2016.2.6 閲覧）
- (2016. 2. 26 受付)

ANALYSIS OF COMBINATION OF VISITING PLACES BY FOREIGN VISITORS IN JAPAN BY LATENT CLASS MODEL

Hideki FURUYA and Yujuan LIU

The purpose of this paper is to identify the combination of visiting places by foreign visitors in Japan. Zoning is set as prefectural level. The number of combination of visiting prefecture is 2191. The Latent Class Model is useful tool to summarize the tour courses from the view point of their similarities. Actual 29,067 trips data collected by Japan Tourism Agency in 2012 and 2013 was used for estimation of the model. It is suitable to assume that there are twenty four classes using foreign tourist visiting prefecture patterns by AIC, BIC and p-Value. These patterns are divided into two groups. One is the group which includes only one region of visiting prefecture. These patterns occupied 78.1%. The other is the visiting places spread into some regions. Typical round trip called Golden Route are included in this pattern and the occupied ratio is 21.9%. It is clear that there are some differences of component ratio in visitors' nations and region, the frequencies of visit Japan. Finally generalized chi-squared test is applied for the independent test between some factors and the visiting places patterns.