

ウェブスクレイピングによる宿泊料データの取集と分析用データの抽出

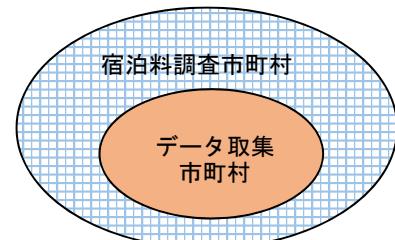
1 データ取集の概要

- (1) データ取集の予約サイトはOTA主要3サイト
- ^{※1}

※1 OTA (Online Travel Agency: オンライン販売に特化した企業)
全体の取扱高の6割弱を占める。

- (2) データ取集市町村は小売物価統計調査（宿泊料調査）における調査市町村99市町村のうち50市町村（図1）

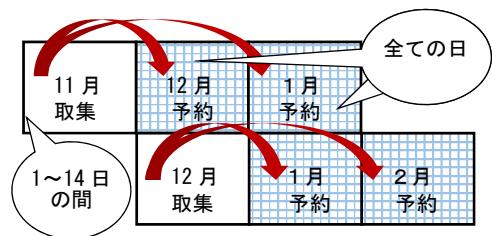
図1 取集市町村



- (3) 取集のタイミングは月の前半（1日～14日）、
-
- 予約対象日は取集月の翌月及び翌々月の全て（図2）

- (4) 宿泊施設ごとに主要な宿泊プラン
- ^{※2}
- （5プラン程度）を取集
-
- ※2 原則、現行の宿泊料調査の調査銘柄に合致するプラン

図2 取集月と予約月



例) 11月は12月と1月の全ての日を予約対象日としてデータを取集

2 分析（試算）用データの抽出

- (1) 宿泊料調査の調査市町村のうち「さいたま市」と「日光市」のデータを抽出

試算には平成29年11月に取集したデータ（宿泊予定日：29年12月及び30年1月）を使用

- (2) 宿泊施設名の標準化（表記揺れの修正）

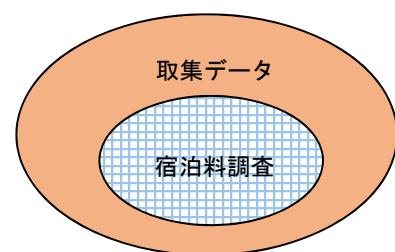
- (3) 宿泊施設の抽出（図3）

現行の宿泊料調査の施設を抽出

- (4) 特異データの削除

「レディースプラン」など特定対象者向けのプラン、「高層階ロイヤルフロア限定」など特異な料金プランを削除

図3 対象宿泊施設



3 分析（試算）結果

2のデータを用いて指数を試算した（別紙参照）。

宿泊料を例にした価格指数の作成案について

1. FEWS(Fixed Effects Window-Splice)指標を参考とした宿泊料指標の算出方法の解説

		○○ホテル				△△ホテル			指標(I _t)	n, m (価格数) 注
		プランA (P ^{dt} ₁)	プランB (P ^{dt} ₂)	プランC (P ^{dt} ₃)	プランD (P ^{dt} ₄)	プランA' (P ^{dt} ₅)	プランB' (P ^{dt} ₆)	プランC' (P ^{dt} ₇)		
		(P ^{dt} ₁)	(P ^{dt} ₂)	(P ^{dt} ₃)	(P ^{dt} ₄)	(P ^{dt} ₅)	(P ^{dt} ₆)	(P ^{dt} ₇)		
2017年 12月1日 (金) (P ¹ _i =P ^{d0} _i)		12,000	—	8,000	23,000	30,000	17,000	9,800	100.000	6
12月2日 (土) (P ² _i)		13,000	—	8,000	25,000	33,500	—	10,000	106.189	5
12月3日 (日) (P ³ _i)		—	16,000	6,000	—	25,000	15,000	8,800	89.478	5
12月4日 (月) (P ⁴ _i)		10,000	15,000	6,000	—	25,000	15,000	8,800	86.856	6
12月5日 (火) (P ⁵ _i)		11,000	15,500	6,000	—	25,000	15,000	8,800	88.730	6
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12月27日 (水) (P ²⁷ _i)		11,000	15,500	6,000	19,000	25,000	17,000	8,800	87.872	7
12月28日 (木) (P ²⁸ _i)		13,000	15,500	8,000	25,000	33,500	18,000	—	104.569	6
12月29日 (金) (P ²⁹ _i)		17,000	—	—	27,000	33,500	19,000	—	115.624	4
12月30日 (土) (P ³⁰ _i)		20,000	—	—	27,000	35,000	20,000	—	123.316	4
12月31日 (日) (P ³¹ _i)		—	—	—	30,000	35,000	20,000	—	117.180	3
プラン別幾何平均価格 (G _i)		11,805	15,347	6,581	24,483	27,681	16,124	9,112		
／(価格数)		26	20	28	13	31	30	27		

注) 下記例の場合、nは12月1日の価格数(6価格)、mは12月1日以降の価格数(例:12月2日は5価格、12月3日は5価格)

$$\text{固定効果の推計} \quad P^t = G_1 D^t_1 + G_2 D^t_2 + G_3 D^t_3 + G_4 D^t_4 + G_5 D^t_5 + G_6 D^t_6 + G_7 D^t_7 + \varepsilon^t \quad (\text{A})$$

(G_i: 推定固定効果、D_i^t: プランの有無)

① 同一品質の平均的な価格(推定固定効果)の作成

$$G_i = \left(\prod_{k=1}^l P_i^{dk} \right)^{1/l} \quad (\text{B})$$

※ G_i は、基準期間における i プランの幾何平均価格であり、i プランにおける平均的な品質とみなせる
G_i は、基準時と比較時に存在するプランの差異を調整する役割を果たす

例: ○○ホテル プランAの平均的な価格 (26価格) $= (12,000 \times 13,000 \times \dots \times 17,000 \times 20,000)^{\frac{1}{26}} = 11,805$

② 基準時の相対的な日次指標の作成

$$I_{d0}^{(N)} = \left(\prod_{i=1}^n p_i^{d0} \right)^{1/n} \bigg/ \left(\prod_{i=1}^n G_i \right)^{1/n} \quad (\text{C})$$

※ 基準(t=0)時点において価格が存在するn個のプランの各幾何平均価格を、同一プランのG_iで除することで、相対的な基準時の価格指標 I_{d0(N)} を作成

例: 基準時の相対的な指標 (12月1日) = $\frac{(12,000 \times 8,000 \times 23,000 \times 30,000 \times 17,000 \times 9,800)^{\frac{1}{6}}}{(11,805 \times 6,581 \times 24,483 \times 27,681 \times 16,124 \times 9,112)^{\frac{1}{6}}} = 1.061$

③ 比較時の相対的な日次指標の作成

$$I_{dt}^{(M)} = \left(\prod_{j=1}^m p_j^{dt} \right)^{1/m} \bigg/ \left(\prod_{j=1}^m G_j \right)^{1/m} \quad (\text{D})$$

※ t 時点において価格が存在するm個のプランの各幾何平均価格を、同一プランのG_jで除することで、相対的な比較時の価格指標 I_{dt(M)} を作成

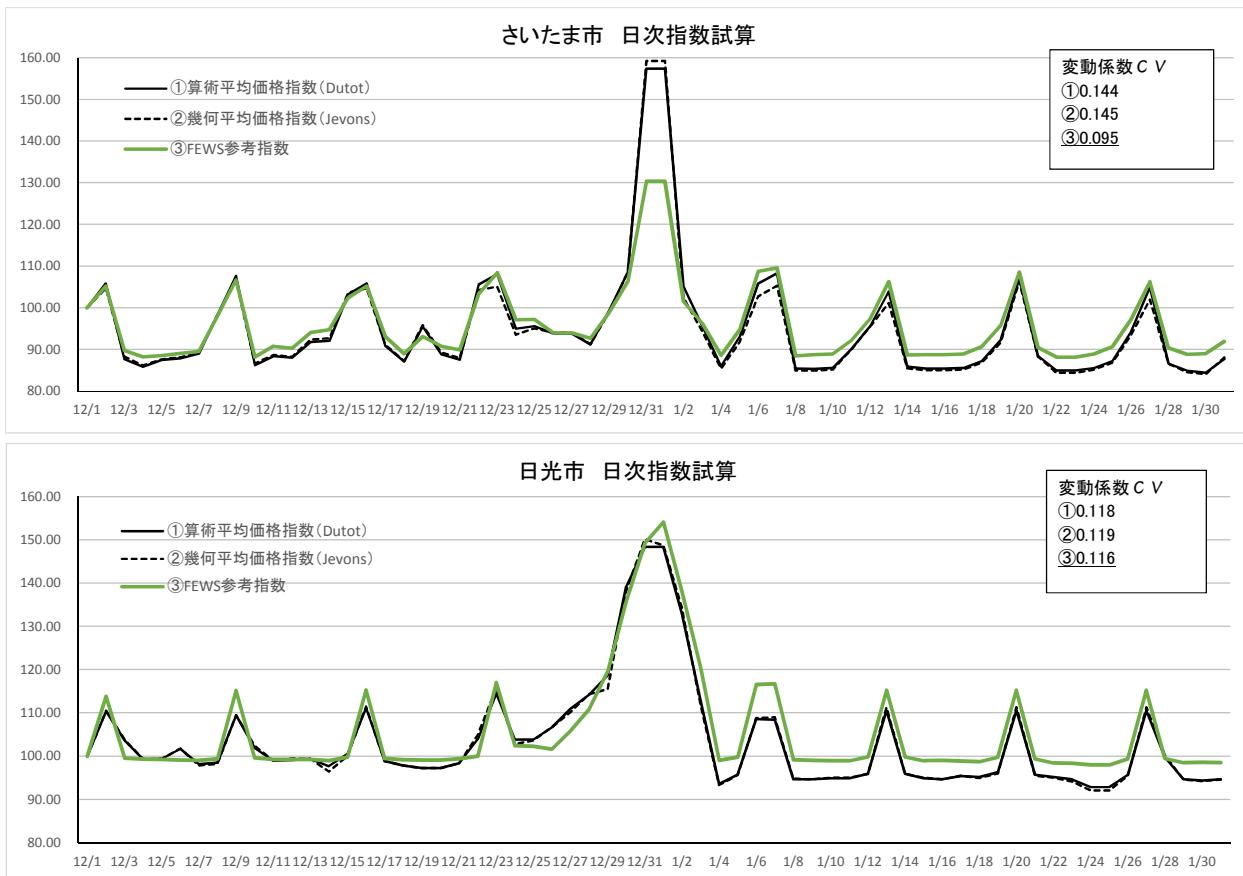
例: 比較時の相対的な指標 (12月2日) = $\frac{(13,000 \times 8,000 \times 25,000 \times 33,500 \times 10,000)^{\frac{1}{5}}}{(11,805 \times 6,581 \times 24,483 \times 27,681 \times 9,112)^{\frac{1}{5}}} = 1.127$

④ 日次指標の作成

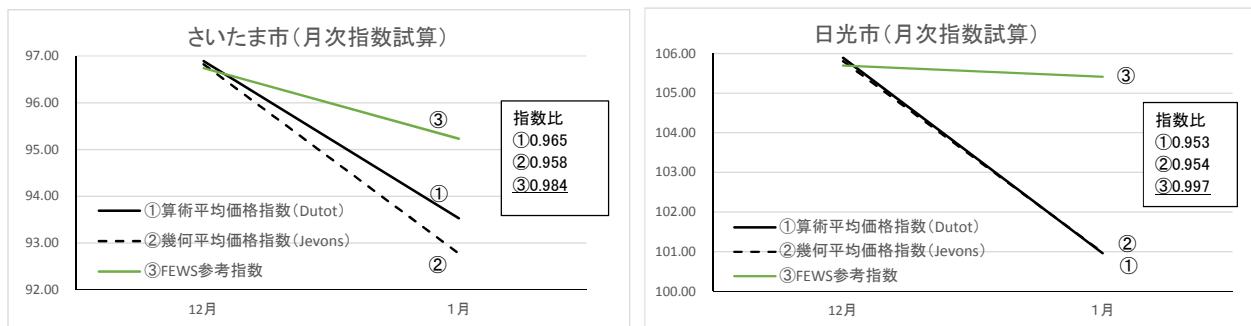
$$I_{dt} = \frac{I_{dt}^{(M)}}{I_{d0}^{(N)}} \times 100 \quad (\text{E})$$

例: 日次指標 (12月2日) = $\frac{1.127}{1.061} \times 100 = 106.189$

2. FEWS指標を参考とした宿泊料指数の試算



注)上図において、Gは12月1日～1月30日までの期間としている。



注)月次指標は、日次指標の算術平均により算出している。

【補足 上図FEWS参考指標との比較に用いている算式】

① 算術平均価格指標(Dutot平均値補完)

以下の式は、欠価格に算術平均価格を補完して計算した結果と同値

$$I_{dt} = \frac{\sum_{j=1}^{n'} P_j^{dt}}{n'} \Big/ \frac{\sum_{i=1}^n P_i^{d0}}{n} \quad (F)$$

例示)

	1日	2日
A	15,200	欠価格
B	14,500	13,400
C	13,200	欠価格
D	欠価格	9,800
価格数	3	2

	1日	2日
A	15,200	11,600
B	14,500	13,400
C	13,200	11,600
D	14,300	9,800

$$= \text{価格数} \quad \boxed{4} \quad \boxed{4}$$

算術平均価格

14,300	11,600
--------	--------

算術平均価格

14,300	11,600
--------	--------

価格比

0.811

価格比

0.811

② 幾何平均価格指標(Jevons平均値補完)

以下の式は、欠価格に幾何平均価格を補完して計算した結果と同値

$$I_{dt} = \left(\prod_{j=1}^{n'} p_j^{dt} \right)^{1/n'} \Big/ \left(\prod_{i=1}^n p_i^{d0} \right)^{1/n} \quad (G)$$

例示)

	1日	2日
A	15,200	欠価格
B	14,500	13,400
C	13,200	欠価格
D	欠価格	9,800
価格数	3	2

	1日	2日
A	15,200	11,459
B	14,500	13,400
C	13,200	11,459
D	14,276	9,800
価格数	3	2

幾何平均価格

14,276	11,459
--------	--------

幾何平均価格

14,276	11,459
--------	--------

価格比

0.803

価格比

0.803

3. ネット通販価格に係る今後の予定

消費者物価指数におけるインターネット通信販売価格の把握については、「統計改革の基本方針」等において、同価格の更なる捕捉及び2020年基準改定における採用の可否を検討し、2018年度までに結論を得ることとされており、来年度の結論に向け、引き続き検討・研究を進める。

(ア) サービスに係る品目

宿泊料、航空運賃の旅行サービス品目に関しては、F E W S 参考指数により継続的に指数を作成するとともに、以下の内容等について検討・研究を進める。

- ① 同質サービスの料金取集のための宿泊施設・プランの選定
- ② 適切な取集タイミング・頻度の見極め
(満室・満席による料金取集不可の回避、料金設定実態を考慮した取集頻度の検討等)
- ③ 下位類から上位類への指数作成方法 (日次⇒月次⇒市町村⇒全国)

(イ) 財に係る品目

以下の①～⑥の家電6品目に関しては、大手のネット専業販売会社及び家電量販店(店舗価格、ネット通販価格)の実売価格情報(製品の平均価格、販売数量、機能等の情報)やウェブスクレイピングによって取集した日次価格情報を用いて、指数作成に係る検討・研究を進める。

- ① テレビ
- ② ビデオカメラ
- ③ ビデオレコーダー
- ④ プリンタ
- ⑤ 携帯オーディオプレイヤー
- ⑥ 電子辞書

なお、その他の家電品目については、インターネット通信販売価格を含む実売価格情報が市場に販売・提供されていないため、ウェブスクレイピングによって取集した日次価格を用いた指数作成方法について検討・研究を進める。

小売物価統計調査宿泊料調査の現状

参考
2-1

1. 宿泊料調査の方法

(1) 調査市町村数及び調査施設数

全国 99 調査市町村、320 施設 (下表「調査市町村及び調査宿泊施設数一覧」参照)

(2) 調査日

毎月 5 日を含む週の金曜日 (休日の前日の場合は翌週の月曜日) 及び土曜日

(3) 調査方法

以下の 4 種類の価格を都道府県職員が確認

	客室タイプ	調査価格 (税・サービス料込み)		施設数
		泊・食事の回数	平日・休前日	
①	和室	1 泊 2 食付	平日	201
			休前日	
②	洋室	1 泊朝食付	平日	119
			休前日	

2. 調査市町村及び調査施設の選定方法

(1) 調査市町村

観光庁等の調査結果に基づき統計局において選定

(2) 調査施設

以下の 2 つの基準を満たす施設を都道府県において選定

① 宿泊者数が多いこと

② 主に観光旅行等を目的とした一般の旅行者が宿泊する施設であること

※上記基準により難い場合は、施設の規模 (部屋数、敷地面積など) を参照すること

表 調査市町村及び調査宿泊施設数一覧

(1) 和室 (1 泊 2 食付)

平成 30 年 1 月現在

市町村 符号	調査 市町村	宿泊 施設数									
01100	札幌市	2	07202	会津若松市	4	20214	茅野市	4	29201	奈良市	3
01202	函館市	2	09206	日光市	7	20561	山ノ内町	4	30401	白浜町	3
01203	小樽市	2	09407	那須町	4	21203	高山市	7	30421	那智勝浦町	3
01206	釧路市	2	10208	渡川市	3	21220	下呂市	3	31202	米子市	3
01230	登別市	2	10426	草津町	4	22130	浜松市	3	32201	松江市	2
01400	俱知安町	2	14382	箱根町	4	22205	熱海市	3	34213	廿日市市	2
01457	上川町	2	14383	真鶴町	2	22208	伊東市	3	35204	萩市	3
01545	斜里町	2	14384	湯河原町	5	22219	下田市	3	38201	松山市	4
01584	洞爺湖町	2	15224	佐渡市	3	22304	南伊豆町	3	41209	嬉野市	3
01631	音更町	2	15461	湯沢町	4	22301	東伊豆町	2	42213	雲仙市	3
02206	十和田市	2	17202	七尾市	3	22302	河津町	2	43214	阿蘇市	3
03201	盛岡市	3	17206	加賀市	3	22222	伊豆市	2	44202	別府市	4
03301	零石町	3	18208	あわら市	2	24211	鳥羽市	4	44213	由布市	3
03205	花巻市	3	18210	坂井市	3	24215	志摩市	3	46210	指宿市	2
04100	仙台市	4	19211	笛吹市	3	28100	神戸市	4	46218	霧島市	2
04401	松島町	2	19430	富士河口湖町	4	28205	洲本市	3	計	66	201
06201	山形市	4	20206	諏訪市	3	28209	豊岡市	3			

(2) 洋室 (1 泊朝食付)

市町村 符号	調査 市町村	宿泊 施設数									
01100	札幌市	4	13100	東京都区部	4	33100	岡山市	3	43100	熊本市	4
01229	富良野市	2	14100	横浜市	4	33202	倉敷市	3	45201	宮崎市	3
04100	仙台市	4	16201	富山市	3	34100	広島市	4	46201	鹿児島市	4
05201	秋田市	3	17201	金沢市	4	36201	徳島市	3	47201	那覇市	4
08201	水戸市	3	23100	名古屋市	4	37201	高松市	3	47207	石垣市	3
11100	さいたま市	3	25201	大津市	3	39201	高知市	4	47209	名護市	2
12100	千葉市	3	26100	京都市	4	40130	福岡市	4	47214	宮古島市	3
12211	成田市	2	27100	大阪市	4	42201	長崎市	4	47311	恩納村	2
12227	浦安市	3	28100	神戸市	4	42202	佐世保市	3	47324	読谷村	2
									計	36	119