

東京 23 区内の優良宿泊プラン選択方法の提案

Proposal for the Selecting Method of an Excellent Accommodation Plan of Hotels within the 23 Wards of Tokyo

1641040 北川 汰一

Taichi KITAGAWA

指導教員 秋葉 知昭

In this study, I collected accommodation data of evaluation and facilities information of hotels from the accommodation portal site. So, I proposed formulas which find out optimal price of accommodation plan of hotels within the 23 wards of Tokyo by using multivariate analysis. From the results of evaluation, I showed proposed formulas accuracy about 90%.

1. 緒言

近年はインターネット上にある旅行情報サイトから事前に宿泊施設の情報を得ることができる。調査によると、旅行情報サイトから予約する人が全体の 29.8%と一番多い[1]。そこで本研究では、宿泊料金と宿泊施設の設備や客室の質（部屋の広さ、ベッドの大きさなど）が釣り合っているかを判別できるように、多変量解析で最適な宿泊料金を調べる。そして最適価格と実際の料金を比較し、優良宿泊プランを選択する方法を提案する。

2. 多変量解析

多変量解析とは複数の変数に関するデータをもとに、変数間の相互関係を分析する統計的手法の総称である[2]。本研究では予測を行ったので、予測に使う手法をいくつか紹介する。

2.1 重回帰分析

重回帰分析とは、予測のための手法であり、重回帰式とは複数の説明変数を使用する回帰式である。目的変数に対して、各説明変数がどの程度影響しているのかが分かる。重回帰分析を使用する際は、目的変数と説明変数のどちらもが定量データの時に使う。

2.2 数量化理論 I 類

数量化理論 I 類も重回帰分析と同様に複数の説明変数から一つの目的変数を予測する。数量化理論 I 類の場合は目的変数が定量データ、説明変数が質的データの際に使う。

2.3 重回帰分析と数量化理論 I 類の混合

重回帰分析と数量化理論 I 類の混合は目的変数が定量データ、説明変数が定量データと質的データのときに使う。

3. 宿泊プランのデータ分析

3.1 データの収集

本研究では宿泊予約サイト「じゃらん net」からデータを収集し、掲載されている各宿泊施設の標準プラン（スタンダードプラン）をデータ分析の対象とした。その他の条件としては、東京 23 区内、1泊2日、一人使用をデータ分析の条件とした。集めたデータとその使用用途を下記に記す。

目的変数

- ・各宿泊プランの宿泊料金

説明変数

- ・各プランの最安値
- ・ロコミの数
- ・ベッドの数
- ・施設の評価（総合、朝食、部屋、接客、風呂、清潔感）
- ・ベッドの幅
- ・部屋の広さ
- ・最寄り駅からの時間
- ・ホテルの価格帯
- ・通信環境
- ・食事の有無
- ・曜日

3.2 データ分析

表 1 宿泊プランの分析結果

説明変数名	分散比	P値 (上側)	偏回帰係数
定数項	134.5340	0.000	6050.796
ホテルの価格	5943.1538	0.000	
一万円台			0.000
一万円未満			-4628.305
二万円台			8578.726
三万円台			17258.804
四万円台			25763.776
五万円台			36193.087
六万円台			49682.379
七万円台			52174.547
ベッドの数	435.0719	0.000	2360.127
最安値	296.2429	0.000	0.160
ベッドの幅	199.9519	0.000	33.223
曜日	50.0496	0.000	
月			0.000
火			63.220
水			345.471
木			602.827
金			449.283
土			1844.283
日			122.974
総合	26.1153	0.000	-818.458
風呂	26.5833	0.000	671.615

データ分析の結果を表 1 に示す。「ホテルの価格帯」の影響度が極めて高く、それに続き「ベッドの数」や「最安値」の説明変数の影響度が高い。総合評価の偏回帰係数がマイナスになったので、総合

評価が高い施設ほど、宿泊料金が安くなる傾向になることが分かった。

4. 予測モデルの作成・評価

4.1 予測モデルの作成

表1を参考にモデル式を作成する。

$B(x_i)$: 宿泊プランの最適価格

x_i : 評価するプラン

$\vec{s}(x_i)$: 評価するプラン x_i において、 $s_0(x_i)$: 定数項

$s_1(x_i)$: ベッドの数 $s_2(x_i)$: 最安値 $s_3(x_i)$: ベッドの幅

$s_4(x_i)$: 総合評価 $s_5(x_i)$: 風呂の評価

\vec{a} : 偏回帰係数ベクトル

\vec{b} : 偏回帰係数ベクトル, ここで b_1 : 1万円未満 b_2 :

1万円台 b_3 : 2万円台 b_4 : 3万円台 b_5 : 4万円台

b_6 : 5万円台 b_7 : 6万円台 b_8 : 7万円台, と定義する。

$\vec{y}(x_i)$: 宿泊プラン x_i における価格帯の選択

$\vec{y}(x_i) = (y_1(x_i), y_2(x_i), y_3(x_i), y_4(x_i), y_5(x_i), y_6(x_i), y_7(x_i), y_8(x_i))$ $y_j(x_i) \in \{0, 1\}$

ここで $y_j(x_i)$ は宿泊プラン x_i の示すホテルの価格帯が1, それ以外は0の2値変数。

\vec{c} : 偏回帰係数ベクトル, ここで c_1 : 月曜日 c_2 : 火曜日 c_3 : 水曜日 c_4 : 木曜日 c_5 : 金曜日 c_6 : 土曜日 c_7 : 日曜日と定義する。

$\vec{z}(x_i)$: 宿泊プラン x_i における曜日の選択

$\vec{z}(x_i) = (z_1(x_i), z_2(x_i), z_3(x_i), z_4(x_i), z_5(x_i), z_6(x_i), z_7(x_i))$ $z_j(x_i) \in \{0, 1\}$

ここで z_j は宿泊プラン x_i の示す曜日が1, それ以外は0の2値変数である。

$A(x_i)$: 宿泊プラン x_i における定量データの予測値の合計 ここで ${}^t\vec{s}(x_i)$ を $\vec{s}(x_i)$ の転置ベクトルとすると $A(x_i)$ は次式(1)で求められる。

$$A(x_i) = \vec{a} {}^t\vec{s}(x_i) \quad (1)$$

$C(x_i)$: 宿泊プラン x_i における価格帯による予測値 ここで ${}^t\vec{y}(x_i)$ を $\vec{y}(x_i)$ の転置ベクトルとすると $C(x_i)$ は次式(2)で求められる。

$$C(x_i) = \vec{b} {}^t\vec{y}(x_i) \quad (2)$$

$D(x_i)$: 宿泊プラン x_i における曜日による予測値 ここで ${}^t\vec{z}(x_i)$ を $\vec{z}(x_i)$ の転置ベクトルとすると $D(x_i)$ は次式(3)で求まる。

$$D(x_i) = \vec{c} {}^t\vec{z}(x_i) \quad (3)$$

以上より宿泊プラン x_i の最適価格は次式(4)で得る。

$$B(x_i) = A(x_i) + C(x_i) + D(x_i) \quad (4)$$

4.2 モデル式の評価

今回は価格帯ごとにMAD値を求めて予測式の精度確認を行う。まずはMAD値を求める際に必要な相対誤差の求め方を次の式(5)に記す。

$$\text{相対誤差} = \frac{\text{予測値} - \text{実測値}}{\text{実測値}} \quad (5)$$

求めた相対誤差の絶対値の平均がMAD値となる。表2は価格帯ごとにMAD値を求めた結果である。

表2 MAD値

	10月と11月	12月	サンプル数
1万円未満	0.123	0.145	1193
1万円台	0.134	0.180	2493
2万円台	0.095	0.154	646
3万円台	0.069	0.093	201
4万円台	0.063	0.086	81
5万円台	0.052	0.070	41
6万円台	0.037	0.037	58
7万円台	0.017	0.031	12
平均	0.073	0.099	

表2の結果から、実際の料金と予測した料金の平均的な誤差は9.9%となった。誤差18%の1万円台の宿泊プランの予測精度が一番低くなった。3万円台から7万円台に関しては、90%以上の精度で予測ができた。

5. 結 言

本研究では、宿泊施設の設備や客室の質(部屋の広さやベッドの幅)などを用いて、宿泊プランの最適価格を導びく式を作成した。説明変数の「ホテルの価格帯」の影響度が非常に大きく、予測精度が90%を超えた。作成した予測モデルはホテル選びの際に参考にするには十分な予測精度であったが、1万円台などの大きな誤差が生じた価格帯もあるので、既存の説明変数では説明しきれなかった要素があったと考えられる。

文 献

- [1] ネットリサーチタイムスドライブ: 国内ホテルに関するアンケート, <http://www.dims.ne.jp/timelyresearch/2007/071121/> (2020/1/5時点)
- [2] インテージ: マーケティング用語集, <https://www.intage.co.jp/glossary/056/> (2020/01/05時点)