

繁忙期と閑散期に対応したホテルの最適価格予測モデル提案

Proposal Forecasting Method for an Optimal Rates of Hotels corresponding to Peak and Off-peak Seasons

2041013 岩崎 巧充

Takumi IWAZAKI

指導教員 秋葉 知昭

This study proposes an optimal rates forecasting model for hotels. I collect data on hotels in the 23 wards of Tokyo and create a price prediction model by mixing multiple regression analysis and quantification theory I. In this study, I also conducted a stratified analysis by dividing the 23 wards of Tokyo into four areas to infer that the average rates of a hotel differs from place to place. The model was evaluated using MAD and RMSE, which are calculated from the error between the forecasted and actual hotel rates. The results showed that the model focused on the western area of Tokyo's 23 wards had the lowest error in forecasting.

1. 緒言

2020年に世界中に蔓延した新型コロナウイルス感染症(COVID-19)は,世界中に大影響を与え,日本の宿泊業界にも影響が出ていた.しかしながら,2022年10月に行われた入国規制や外出自粛の規制緩和の結果,2023年頃にはコロナ前と変わらない量の外国人が日本に来日するようになった[1].こういった背景から,コロナ過を経てたくさんの人々が観光やビジネスなどの目的でホテル需要が高まるのではないかと推測される.

日本の延べ宿泊数は観光地や都心部などに多いが,最も多い場所は東京である[2].その背景を元に,本研究ではホテル価格の予測モデルを作成する.先行研究である北川の研究[3]では考慮していなかったシーズンごとのホテル需要にも対応できるように,本研究は12月や1月などのイベントが多い時期のホテルデータも用いながら,より実用性の高いモデル作成を目指す.

2. データと分析・評価方法

2.1 データの対象と収集

本研究では東京23区内のホテルを対象とし,2023年11月1日から2023年3月31日までのデータを取得する.収集するプラン内容としては,一人使用の一泊二日,食事なしとし,そのホテルの中で一番安いプラン内容を対象とした.データは主にじゃらんnetからPythonのスクレイピングによって収集した.ホテルデータの取得日は全て2023年10月25日となっている.

2.2 分析方法

本研究では,StatWorksを用いて重回帰分析と

数量化理論I類の混合で分析を行う.説明変数の選択は,逐次変数選択機能を利用し,自由度2重調整寄与率が一番高くなるように選択する.今回は東京23区内でもエリアごとにホテルの平均価格が違っていると予測し,四つのエリアに分けて層別して分析を行った.そのため,今回は層別をしないで23区内を全て分析した全体モデル一つと,層別後の分析モデル四つを分析する.

3. 分析結果

表3は東京23区内のホテル全体の分析結果と,層別によってエリアごとの分析結果の寄与率と残差標準偏差である.

表3 分析結果

	寄与率	残差標準偏差
全体	0.568	8030
都心部	0.543	10116
副都心部	0.635	6663
東エリア	0.537	5291
西エリア	0.794	4522

表3より,西エリアの寄与率が,0.794と一番高い値となった.残差標準偏差も4522と西エリアが一番低い値となっており,西エリアのモデルが一番精度の高いモデルであるといえる.次に良いモデルといえるのは副都心部のモデルであった.副都心部は寄与率が0.635と二番目に高く,残差標準偏差も6663と,全体の中で三番目に低い値となった.他の三つのモデルは,西エリアや副都心部のモデルと比較すると,寄与率が低いという結果となった.

4. モデルの評価

本研究では各モデルのホテル価格の予測値を算

出し、実際のホテル価格との誤差からモデル評価を行う。予測に使うホテルデータは、今回取得した 11 月 1 日から 3 月 31 日までのホテルデータからランダムに抽出する。この予測に使うデータは、分析をする前に抽出したため、分析には使用していないデータである。全体のモデルでは 200 件分、各モデルでは 50 件分のホテルデータを抽出し、予測値を算出する。今回は 11 月から 3 月まで十分に予測が出来るように、全体では月ごとに 40 件ずつ、エリアごとでは月ごとに 10 件ずつの予測値を算出した。

評価は MAD と RMSE の二つの値で行う。表 4.1 は各モデルの MAD を比較した表である。

表 4.1 MAD の比較

MAD	全期間	11月	12月	1月	2月	3月
全体モデル	0.314	0.317	0.327	0.307	0.315	0.305
都心部	0.287	0.207	0.327	0.277	0.453	0.171
副都心部	0.291	0.222	0.308	0.272	0.434	0.221
東エリア	0.306	0.421	0.240	0.177	0.240	0.452
西エリア	0.238	0.242	0.259	0.165	0.226	0.300

表 4.1 は各モデルの MAD を月ごとに算出したものである。また全期間とは、月ごとに算出した全ての MAD を平均した値である。MAD は値が小さいほど誤差が少ないことを表している。表 4.1 の全期間から、西エリアのモデルが一番誤差を少なく予測できたといえる。また、他の層別後のエリアも、全体のモデルよりも比較的誤差が少なく予測できていた。これは、層別後のモデルの方がホテルデータに対して専門性があるため、誤差が少なかったと推測される。

表 4.2 RMSE

RMSE	全期間	11月	12月	1月	2月	3月
全体モデル	7148	6339	10535	4924	6186	6463
都心部	7245	7072	5966	4107	11412	5453
副都心部	6341	9911	4656	6395	5393	3340
東エリア	9801	19251	2655	2277	8911	4250
西エリア	4684	6143	2385	4019	5137	4880

表 4.2 は各モデルの RMSE をまとめた表である。RMSE は、予測値と実測値との誤差がどのくらいになるかを表している。表 4.2 の全期間より、西エリアの誤差が一番少ないことがわかる。また、東エリアの 11 月は全体的に見てもとても高い RMSE になっている。原因としては、RMSE の特徴として外れ値の影響を大きく受けてしまうという欠点がある。その結果、東エリアの 11 月の誤差が大きくなってしまったのではないかと考えられる。

5. 考察

MAD と RMSE の値から、今回は西エリアのモ

デルが一番誤差が少なく予測できていることが分かった。また層別をした結果、寄与率にそこまで左右されることなくある程度予測が出来ており、全体モデルよりも誤差が少なく予測で来たのではないかと考えられる。その理由としては、層別後のモデルの方が全体モデルよりもホテルデータが絞られているため、各エリアのホテルに対して専門性を手に入れたと推測される。

また本研究は月ごとの予測も行ったが、どのモデルでも比較的 1 月の予測が一番誤差が少なく予測できていた。これは年末年始などの繁忙期であるシーズンのホテル価格予測が上手くいった結果と考察できる。

6. 結言

本研究では、11 月から 3 月末までのホテルデータから、重回帰分析と数量化理論 I 類の混合を用いてホテル価格の予測を行った。また、場所ごとにホテルの平均価格に違いがあると推測し、エリアごとの層別分析を行った。分析結果としては、特に西エリアに高い寄与率が見られた。

また各モデルの MAD と RMSE を用いて評価を行った。評価結果としても、西エリアの予測誤差が一番少ない結果となった。分析では今回西エリア以外のモデルで高い寄与率を出すことが出来なかった。

今後の展望として、より詳細なホテルの場所を示すデータを説明変数に加える事や、ホテルの予約日と宿泊日の差から生まれる価格変動を説明変数で示す事によって、さらに誤差の少ない価格予測が実現できると考えられる。また本研究は 11 月から 3 月までのデータ期間で分析を行ったが、5 月のゴールデンウィークや 10 月のハロウィンなどの他の繁忙期のホテル価格予測を実現するためには、一年分のホテルデータの取得が必要だと考えられる。

文 献

- [1] JTB 総合研究所：月別訪日外国人の推移内訳
<https://www.tourism.jp/tourism-database/stats/inbound/>
(2024/1/24 時点)
- [2] 国土交通省：延べ宿泊数内訳
<https://www.mlit.go.jp/kankochou/siryou/toukei/shukuhakutoukei.html> (2024/1/24 時点)
- [3] 北川汰一：東京 23 区内の優良宿泊プラン選択方法の提案、千葉工業大学社会システム科学部経営情報科学科 2019 年度卒業論文