

全国 12 地域別賃貸物件価格予測モデルの提案

Proposal for the Leasehold Property Price Prediction Model in the Whole Country 12 Areas

1741079 田村 明洋

Akihiro TAMURA

指導教員 秋葉 知昭

In this study, I collected data of the leasehold property information from SUUMO and the data each municipality. So, I suggest model types discovering the most suitable for the leasehold property price using a multivariate analysis in 12 areas. The predictive precision of the average of 12 areas exceeded 80%.

1. 緒言

近年はインターネットが普及し、多くの人々がインターネット上にある物件情報サイトを利用し、賃貸物件の検索や内見の予約を行っている。だが実際に賃貸契約で後悔したという結果のアンケートがある[1]。そこで本研究では、物件情報サイトである。全国の賃貸物件情報のデータを対象にした。また人口が密集している地域の家賃は高い傾向にあるため、物件の家賃に人口密度や転入が大きく影響していると考えたため市町村の人口や人口密度のデータを加えて重回帰分析と数量化理論 I 類の混合を用いて分析を行う。その結果から 12 地域別の賃貸価格の予測モデルを作成し、実際の賃貸価格とのデータを比較して評価を行う。

2. 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 類の混合とは予測のための手法であり、目的変数に対してどの程度影響しているのかが分かる。重回帰分析も数量化理論 I 類も複数の説明変数から 1 つの目的変数を予測するという点では同じだが、説明変数の種類によって異なる。重回帰分析と数量化理論 I 類の混合では目的変数が定量データ、説明変数が定量データと定性データの時に利用する。

3. 賃貸物件情報のデータ分析

3.1 データの収集

本研究では「SUUMO」に掲載されている賃貸物件情報の家賃や築年数などのデータを分析の対象とした。また、物件の家賃に人口密度や転入(他の市区町村から移り住む人数)が大きく影響していると考えた。そのため各市町村から人口、市の面積、人口密度、転入、転出のデータを分析の対象に加えた。

3.2 スクレイピングを用いたデータの収集[2]

今回のデータ分析に使うデータ数が膨大なため効率的にデータ収集を行うためにスクレイピングを行った。集めた情報とその使用用途を下記に記す。

目的変数

- ・賃貸物件の家賃

説明変数

- ・築年数 ・床面積 ・構造(階建) ・階数 ・管理費 ・敷金 ・礼金 ・人口 ・面積 ・人口密度 ・転入 ・転出 ・間取り(1LDK, ワンルームなど)

4. データ分析の結果

表 1 都心・副都心の分析結果の一部

実数選択	確率	残差の分布	残差の連関	予測		
目的変数名	重相関係数	寄与率R ²	R ²	R ²		
家賃	0.950	0.903	0.903	0.903		
	残差自由度	残差標準偏差				
	30268	2.185				
vNo	説明変数名	分数比	P値(上側)	偏回帰係数	標準偏回帰	トレランス
0	空数項	3167.3485	0.000	5.003		
4	築年数	9292.0617	0.000	-0.089	-0.173	0.790
5	構造	2.8901	0.085 *			
6	階数	2326.6313	0.000	0.137	0.086	0.802
8	管理費	460.5429	0.000	0.000	0.040	0.745
9	敷金	19866.0759	0.000	0.141	0.330	0.461
10	礼金	710.9222	0.000	0.040	0.051	0.701
18	CIT:間取り	100.2266	0.000			
	1DK			0.000		
	1K			-0.385		
	1LDK			0.696		
	1LK			-1.242		
	1SDK			-0.070		
	1SK			-0.078		
	1SLDK			1.454		
	2DK			-0.310		
	2K			-0.083		
	2LDK			2.001		

表 1 はデータ分析の結果の一部の表である。敷金の影響度が大きく、それに続いて築年数と床面積の影響度が高い。築年数度の偏回帰係数がマイナスになった。残差標準偏差が 2.185 なので、実際の価格と予測した価格の平均的な誤差が 2.185 万円となる。

全ての地域で共通して影響度が大きいのは床面積、築年数だった。管理費、市の面積の説明変数は影響度が小さい傾向にあった。

5. 予測モデルの作成・評価

5.1 予測モデルの作成

表 1 都心・副都心の分析結果を参考に最適価格を求める式を作る。

$B(x_i)$:賃貸物件の最適価格

x_i :評価する物件

$s(x_i)$:評価する物件の x_i の定量データ

$$\mathbf{s}(x_i) = (s_0(x_i), s_1(x_i), s_2(x_i), s_3(x_i), s_4(x_i), s_5(x_i), s_6(x_i), s_7(x_i), s_8(x_i))$$

ここで評価する物件 x_i において、

$s(x_i) = s_0(x_i)$:定数項 $s_1(x_i)$:築年数 $s_2(x_i)$:階数
 $s_3(x_i)$:管理費 $s_4(x_i)$:敷金 $s_5(x_i)$:礼金 $s_6(x_i)$:床面積
 $s_7(x_i)$:人口密度 $s_8(x_i)$:転入

\mathbf{a} :偏回帰係数ベクトル

$$\mathbf{a} = (a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)$$

\mathbf{b} :偏回帰係数ベクトル

$\mathbf{b} = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}, b_{15}, b_{16}, b_{17}, b_{18}, b_{19}, b_{20}, b_{21}, b_{22}, b_{23}, b_{24})$
ここで、 b_1 :1DK b_2 :1K b_3 :1LDK b_4 :1LK b_5 :1SDK
 b_6 :1SK b_7 :1SLDK b_8 :2DK b_9 :2K b_{10} :2LDK
 b_{11} :2SDK b_{12} :2SK b_{13} :2SLDK b_{14} :3DK b_{15} :3K
 b_{16} :3LDK b_{17} :3SDK b_{18} :3SLDK b_{19} :4D
 b_{20} :4K b_{21} :4LDK b_{22} :4SLDK b_{23} :5LDK b_{24} :ワンルームと定義する。

$\mathbf{y}(x_i)$:サンプル x_i における間取り

$$\mathbf{y}(x_i) = (y_1(x_i), y_2(x_i), \dots, y_{23}(x_i), y_{24}(x_i))$$

$y_j(x_i) \in \{0,1\}, j = 1, \dots, 24$

ここで $y_j(x_i)$ はサンプル x_i の示す間取りが1,それ以外は0とする2値変数である。

$A(x_i)$:サンプル x_i における定量データの部分の予測値の合計

$\mathbf{A}^T \mathbf{s}(x_i)$ を $\mathbf{s}(x_i)$ の転置ベクトルとすると $A(x_i)$ は次式(1)で求められる。

$$A(x_i) = \mathbf{a}^T \mathbf{s}(x_i) \quad (1)$$

$C(x_i)$:サンプル x_i における曜日による予測値

$\mathbf{C}^T \mathbf{y}(x_i)$ を $\mathbf{y}(x_i)$ の転置ベクトルとすると $C(x_i)$ は次式(2)で求められる。

$$C(x_i) = \mathbf{b}^T \mathbf{y}(x_i) \quad (2)$$

以上により,サンプル x_i の販売数は次式(3)で求められる。

$$B(x_i) = A(x_i) + C(x_i) \quad (3)$$

5.2 モデル式の評価

間取り毎にMAD値を求めて予測式の精度確認を行う。まずはMAD値を求めるのに必要な相対誤差の求め方を次の式(4)に記す。

$$\text{相対誤差} = \frac{\text{予測値} - \text{実測値}}{\text{実測値}} \quad (4)$$

求めた相対誤差の絶対値の平均がMAD値をなる。表2は12地域別にMAD値を求めた結果である。

表2 12地域別のMAD値比較

	MAD値
都心副都心	0.134
東北・北海道	0.163
北関東・北陸・甲信越	0.146
神奈川	0.154
東京都	0.127
西東京とその他	0.165
埼玉・千葉	0.152
東海	0.158
京都・兵庫	0.17
大阪	0.174
中国・四国	0.198
九州・沖縄	0.178
平均	0.16

最も予測精度が高かったのは都心・副都心の地域でMAD値が0.134となった。これは13.4%の誤差という結果になる。反対に最も予測精度が低かったのは中国・四国の地域であった。予測精度が12地域の平均の実際の価格と予測した価格の誤差は16%という結果となった。

6. 結言

80%以上の精度で予測できているので、賃貸物件の選択の際に参考にするには、十分な精度だと考えられる。しかしサンプル数が少ない間取りでは誤差が大きくなった。サンプル数が少ない他、既存の説明変数では説明しきれなかった要因があると考えられる。

今後の課題は新たに説明変数に加えることで、回帰式の予測精度を上げることができると考えられる。また本研究では12地域別に分析を行ったが間取りを単身向けの物件や4人家族向けの物件などに層別して分析を行うことで定量データの説明変数のみで予測でき、モデル式が使いやすく実用的なものになると考えられる。

文献

[1] 不動産総研：賃貸契約での後悔に関するアンケート

<https://wakearipro.com/rental-contract-regret/#toc4>(2022/1/20 時点)

[2] TAT [コード解説]pythonでSUUMOの賃貸物件情報をスクレイピングする

<https://myfrankblog.com/scraping-suumo-website/#Python%EF%BC%99>(2022/1/20 時点)