ISSN 2959-6157

Research on the Influencing Factors of Housing Satisfaction

Maimu Yang

Faculty of Science, Jiangsu Ocean University, Lianyungang, 222005, China
Corresponding author: rwfl123@tzc.edu.cn

Abstract:

This article analyzes the impact of housing satisfaction from multiple perspectives. And there are no missing values in the data. Factor analysis is used to reduce the dimensionality of variables, integrating multiple factors into five factors for easy analysis. The meanings of the factors are clear, namely: living conditions, family situation, regional economy, experience situation, and social employment quality. The factor is processed using binomial logistic regression, and the prediction effect is relatively satisfactory. Analysis of the parameters shows that the better the current living conditions, the higher the regional economy, the higher the quality of social employment, and the higher the probability of housing satisfaction. By comparing the full variable binomial logistic regression, it was found that the older the model parameters, the better their age and employment status, the larger their per capita living area, and the lower their education level. Unmarried individuals are more likely to be satisfied with their houses, which is consistent with basic knowledge.

Keywords: Housing satisfaction; influencing factors; binomial logistic regression.

1. Introduction

With the acceleration of urbanization and the improvement of people's living standards, housing issues have gradually become the focus of social attention. Housing, as the infrastructure of people's lives, its quality, environment, supporting facilities, and other factors directly affect people's quality of life and happiness. Conducting in-depth analysis on the testing and influencing factors of housing satisfaction not only helps to understand the actual needs and expectations of residents for housing but also provides a decision-making basis for the government, developers, and relevant departments, promoting the healthy development of the housing market. Therefore, based on the theories of public goods, hierarchy of needs, and fairness and efficiency, identify the influencing factors of household satisfaction in affordable housing and the impact paths on household satisfaction [1].

In the current social context, people's satisfaction with housing presents diverse characteristics. With the development of the economy and the increase in residents' income, people's requirements for the quality, comfort, safety, and other aspects of housing are becoming increasingly high. In particular, livability factors significantly affect the housing satisfaction level of young people. The better the livability of a city, the higher the housing satisfaction [2]. Meanwhile, based on the 2000 population census data of six major cities in China, it was analyzed that the age, industry, occupation, education level, household registration nature, registration status, migration, and total household population have a significant impact on housing conditions [3]. The nature of household registration, migration time, and distance are also important factors affecting the housing situation of urban families in China [4]. The housing situation of families also has a significant impact on the happiness of married groups [5].

Based on the above background, this study aims to comprehensively and systematically understand the housing satisfaction status of residents through scientific testing methods, and deeply explore the influencing factors behind it. By analyzing the data of urban residents in China's Fifth National Population Census (from now on referred to as the Fifth National Population Census), it is possible to understand the situation of different occupational classes in terms of housing property rights, housing area, and housing quality [6]. By collecting and analyzing a large amount of housing satisfaction data, combined with relevant theories and models, this study will reveal the connotation and composition of housing satisfaction, providing strong theoretical support and practical guidance for improving residents' housing satisfaction and living environment.

This study aims to reveal the connotation and composition of housing satisfaction through in-depth investigation and analysis, explore the main factors affecting housing satisfaction, and provide strong theoretical support and practi-

cal guidance for improving residents' housing satisfaction and living environment [7]. Starting from the essential requirements of inclusive urban development, propose path suggestions to improve the satisfaction of Chengdu public rental housing residents. At the same time, the author also aims to understand some issues through this study, such as what choices and constraints mobile populations face in urban housing decision-making and how the housing model of mobile population compares to that of local residents [8].

This article mainly focuses on testing housing satisfaction and the analysis of its influencing factors. A satisfaction seat influencing factor index system was constructed by analyzing relevant literature, and a study was conducted on the influencing factors of residential full competition seats among residents in the community [9]. An ordered multiple logistic regression model was used to determine the key factors affecting the housing satisfaction of new citizens and to understand the gap between their housing expectations and their living conditions [10]. By using factor analysis, the collected housing satisfaction data is preprocessed to test whether variables are suitable for factor analysis, and key factors are extracted for naming and interpretation in order to reveal potential factors that affect housing satisfaction. Using a binomial logistic regression model, the model was constructed and tested under different input strategies to explore the significant impact of each factor on housing satisfaction. Through these studies, this article aims to provide theoretical support and practical guidance for improving resident housing satisfaction.

2. Methods

2.1 Data Source

The prediction data used in this paper are derived from individual studies. The raw data were saved in the CSV format. This dataset provides key insights into the correlation between various factors and the degree of satisfaction. The total sample size was 2993, with numerical types and classification types.

2.2 Variable Selection

According to the table, there are a total of 2993 data points, including multiple factors. As shown in Table 1.

Table 1. Describe statistics

Variable Name	Symbolic	Type	Meaning
Area	x_1	Categorical	Five regions a-e
gender	x_2	Categorical	1man/2woman
age	x_3	Numeric	20-65
Educational level	x_4	Categorical	Four educational qualifications from junior high school to graduate school
Employment status	x_5	Categorical	Six employment situations
marriage	x_6	Categorical	1yes/2no
population	x_7	Numeric	1-8
Household registration	x_8	Categorical	1this locality/2Field
Household income	x_9	Numeric	1200-250000
Current living area	x_{10}	Numeric	6-300
Per capita area	x_{11}	Numeric	2.4-115
Property rights	x_{12}	Categorical	Eight types of unit types
Housing satisfaction	y	Categorical	1satisfied/2Dissatisfied

According to the observation data, age, family income, and current living area are quantitative data, while district, gender, marriage, education level, household registration status, and permanent resident population are classified data. This paper calculates the average of quantitative data based on housing satisfaction, and calculates the proportion of housing satisfaction for each specific category. The mean values obtained by dividing the quantitative data into housing satisfaction levels differ significantly. There is not much difference in the proportion of housing satis-

faction based on education level, household registration status, gender, and marital status among different categories.

To verify the above idea, the author conducted a variance test on quantitative data and a chi-square test on categorical data. It was found that at a significance level of 0.05, except for education level, household registration status, and gender, the probability of marriage was less than 0.05. Therefore, all except for them reject the original hypothesis.

2.3 Bartlett's Test

In order to reduce the number of variables, the author used factor analysis to complete the "dimension reduction". The factor analysis requires a certain linear relationship between the original variables, so the author made the reflection image matrix and used Bartlett's test. The results are shown in the table, and the reflection image matrix shows that all variables have a certain correlation. According to Bartlett's test, the p-value is much less than 0.05, so the null hypothesis is rejected at the significance level of 0.05. Therefore, there is a significant difference between the image matrix and the identity matrix, and factor analysis can be conducted (Table 2).

Table 2. Bartlett's Test

Approximate chi square	5237.758
Degree of freedom	28
Statistical significance	0.000

2.4 Data Preprocessing

The author denotes the area, age, employment status, permanent resident population, family income, current living area, per capita area, and housing property rights, respectively, and the factor of birth is recorded as many of the dependent variables are classified variables. Hence, the author uses the main components of classification, and the results are shown in the following table 3.

Table 3. Percentage of variance

1		2	3	4	5	Average value
Area	0.084	0.007	0.391	0.307	0.234	0.205
Age	0.024	0.238	0.148	0.476	0.236	0.225
Employment	0.261	0.043	0.289	0.055	0.197	0.169
Population	0.028	0.796	0.050	0.071	0.006	0.190
Income	0.388	0.112	0.212	0.094	0.108	0.183
Living area	0.754	0.142	0.150	0.104	0.096	0.249
Per capita	0.762	0.454	0.243	0.147	0.154	0.352

It can be seen from the table that these five factors explain 85.116% of the total data. Basically, the information of the original variables is used. The author can also know one coefficient of each factor about each variable, so the author can write the factor score function:

$$u_1 = 0.084x_1 + 0.024x_2 + 0.261x_3 + 0.028x_4 + 0.388x_5 + 0.754x_6 + 0.762x_7 + 0.239x_8 \quad (1)$$

$$u_2 = 0.007x_1 + 0.238x_2 + 0.040x_3 + 0.796x_4 + 0.112x_5 + 0.142x_6 + 0.454x_7 + 0.241x_8 \quad (2)$$

$$u_3 = 0.391x_1 + 0.148x_2 + 0.289x_3 + 0.050x_4 + 0.212x_5 + 0.150x_6 + 0.243x_7 + 0.070x_8 \quad (3)$$

$$u_4 = 0.307x_1 + 0.476x_2 + 0.055x_3 + 0.071x_4 + 0.094x_5 + 0.104x_6 + 0.147x_7 + 0.031x_8 \quad (4)$$

$$u_5 = 0.234x_1 + 0.236x_2 + 0.197x_3 + 0.006x_4 + 0.108x_5 + 0.096x_6 + 0.154x_7 + 0.146x_8 \quad (5)$$

Through analysis, factor one mainly contains the current living area and per capita area. The author explains factor one as living conditions, factor two as permanent population, factor two as family situation, factor three as district and employment status, family income, factor three as regional economy, factor four as experience, and factor five as district and age, The proportion of variance in employment status is relatively high, so it is explained as the quality of social employment. The variance of the area, age, permanent population, and per capita area are almost all represented by the total proportion of variance at the end. The minimum is property ownership, but it is also very close to 70%.

2.5 Method Introduction

Binomial logistic regression was performed as an input strategy when housing satisfaction was used as the dependent variable and the independent variable. The author

took 0.5 as the cut-off value. If the probability predicted value was greater than 0.5, the author considered the predicted value of the explanatory variable to be 2, which is not satisfied with the housing. If the probability predicted value is less than 0.5, the author denoted the predicted value of 1 as satisfactory.

3. Results and Discussion

3.1 Preliminary Work

As Table 4 shows, at the 0.05 significance level, the probability P value of each step is less than 0.05, so the linear relationship of the explanatory variables in this model and Logit P is significant, and the model is reasonable.

Table 4. Omnibus test for model coefficients

Chi-square			Degree of freedom	Significance
Step 1	Step	406.574	1	0.000
	Piece	406.574	1	0.000
	Model	406.574	1	0.000
Step 2	Step	120.178	1	0.000
	Piece	526.752	2	0.000
	Model	526.752	2	0.000
Step 3	Step	29.920	1	0.000
	Piece	556.672	3	0.000
	Model	556.672	3	0.000

From table 5, it can be seen that the explanatory variables were screened three times, and the final model contained these three variables. Under the significance level of 0.05, the P-values corresponding to the Wald test of their regression coefficients were all lower than the significance level. Therefore, the null hypothesis was rejected.

Table 5. Variables in the equation

B		Wald	Degree of freedom	Significance	Exp(B)
Step1 u1	-0.832	342.23	1	0.000	0.435
Constant	0.581	199.00	1	0.000	1.788
Step2 u1	-0.870	353.07	1	0.000	0.419
u3	-0.457	114.10	1	0.000	0.633
Constant	0.611	207.77	1	0.000	1.843
Step3 u1	-0.884	356.58	1	0.000	0.413
u3	-0.466	115.97	1	0.000	0.628
u5	0.229	29.324	1	0.000	1.257
Constant	0.621	211.04	1	0.000	1.862

The final Logit regression equation is

$$LogitP = 0.621 - 0.884 - 0.466 + 0.229u_1u_3u_5 .$$

From table 5, it can be seen that the explanatory variables were screened three times, and the final model contained these three variables. Under the significance level of 0.05, the P-values corresponding to the Wald test of their regression coefficients were all lower than the significance level. Therefore, the null hypothesis that their linear relationship with Logit P was significant should be retained in the equation.

3.2 Hosmer-Lemeshaw Test

The model was subjected to the Hosmer Lemeshow test, and the results are shown in table 5. The observed value of the Hosmer Lemeshow statistic in the final model is 18.159, with a probability P-value of 0.020. At a significance level of 0.01, the null hypothesis cannot be rejected; that is, it is considered that there is no significant difference between the distribution of actual class values of the

explained variable and the distribution of predicted class values (table 6).

Table 6. Hosmer Lemeshow test

Step	Chi-square	Degree of freedom	Significance
1	8.446	8	0.391
2	24.158	8	0.002
3	18.159	8	0.020

3.3 Model Prediction

From table 7, the accuracy of the first model is 68.0%, with high accuracy for unsatisfactory predictions but low accuracy for satisfactory predictions. The second model was 71.0%, which compared with model 1, about 0.6%, and satisfactory prediction accuracy. The overall accuracy of the third model was 71.3%.

Table 7. Classification Table

Satisfied			Dissatisfied	Correct percentage
Step1	Satisfied	476	652	42.2
	Dissatisfied	306	1559	83.6
Overall percentage				68.0
Step2	Satisfied	556	572	49.3
	Dissatisfied	295	1570	84.2
Overall percentage				71.0
Step3	Satisfied	558	570	49.5
	Dissatisfied	288	1577	84.6
Overall percentage				71.3

According to the Logistic regression equation: Logit P=0.621-0.884-0.466+0.229 $u_1u_3u_5$. The author knows that housing satisfaction is mainly related to growth positively and negatively; that is, the better the current living conditions, the higher the regional economy, the higher the quality of social employment, and the higher the probability of housing satisfaction.

4. Conclusion

Education level, hukou status, gender, and marriage had no significant effect on housing satisfaction. The better the current living conditions, the higher the regional economy, the higher the quality of social employment, and the higher the probability of housing satisfaction. The overall prediction accuracy rate of binary logistic regression after excluding insignificant variables is about 1% higher than that after factor analysis, and the gap is not large, indicating that the information lost in factor analysis is not much, and the factor analysis is relatively successful. Although it improved by 1.6% over the model after factor analysis, the satisfactory prediction accuracy decreased by about 2.6%. And their unsatisfactory predictions are very high, but their satisfactory predictions are very low.

Moreover, the model has five variables, compared with only three variables after the factor analysis. Therefore, the factor analysis is forward: the LR binomial logistic regression model is better. The overall percentage of binary logistic regression without excluding variables is lower than the overall percentage when removing variables, and the number of variables is also one more. Therefore, it can also indicate that excluding non-significant variables has little impact on the correct rate of the final binary logistic regression model. The older the age, the better the working conditions; the larger the per capita living area, the lower the education level, unmarried, outside the city; the more high-end the housing type, the higher the possibility of satisfaction with the house, which is consistent with the basic common sense.

References

[1]Li Wenjing. Research on the influencing factors of household satisfaction with affordable housing based on SEM. Nanjing Forestry University, 2022.

[2]Xian Xizhu. Research on the factors influencing housing satisfaction among young people in small and medium-sized cities is based on empirical evidence from SH city. Guangzhou University, 2023.
[3]Yi Chengdong. Factors influencing the housing situation of urban families: an empirical study based on Logit model and data from six cities' "Five Popularizations". Journal of Nanjing Population Management Cadre College, 2007, 2: 24-27.
[4]Yi Chengdong. Household Registration, Migration, and Housing Status of Urban Families in China: An Empirical Study Based on 2000 Population Census Data. Journal of Nanjing Population Management Cadre College, 2006, 4: 20-22+44.
[5]Wei Wanqing. Resource Advantage or Security Link - Housing Status, Property Ownership, and Happiness. Sociological Research, 2023, 38(3): 159-179+229.
[6]Bian Yanjie, Liu Yongli. Social Stratification, Housing Property Rights, and Housing Quality: An Analysis of China's "Five Popular Education" Data. Sociological Research, 2005, 3: 82-98+243.
[7]Wang Lu. Research on the influencing factors of household satisfaction with public rental housing in Chengdu. Southwest Jiaotong University, 2021.
[8]Wu Weiping, Wang Hansheng. Analysis of the Housing Status of Floating Population in Beijing and Shanghai as Residences in Metropolitan Areas. Sociological Research, 2002, 3: 92-110.
[9]Hao Qingting, Wang Haiyan, Hao Jiangjiang. A Study on the Factors Influencing Housing Satisfaction of Community Residents: A Case Study of Yijing Community in Dazhou, Sichuan Province. Jiangxi Building Materials, 2022, 7: 322-325+328.
[10]Li Jiafei. Analysis of Factors Influencing Satisfaction with Affordable Housing for New Citizens in Taiyuan City. Shanxi University of Finance and Economics, 2023.