

Investigating Factors Affecting House Prices - Chicago as an Example

Chengrui Qian *

Adcote School, Shanghai, Songjiang, 201600, China

* Corresponding Author Email: phaar1121753@mail.wcccd.edu

Abstract. House is a very important necessity in human society and its price is also highly concerned. With the rapid development of the society, the house price is also rising rapidly, how to find some favourable ways to buy a house is very important for people. This paper will firstly analyse the relationship between house price and some factors by grouping them with box-and-line diagrams, then further calculate the correlation coefficients between the factors and house price, and further correlate the factors with house price. After that, this paper will use multiple linear regression to analyse the relationship using house price as y and other eight variables as x. After that, this paper will conclude that Space, Garage, Bathroom, Lot will have a substantial positive impact on the price. The price will have a significant negative effect due to the bedroom. But Room, Condition, Tax will not have any effect on Price. Finally, this paper will summaries the results of the analysis and provide some preferred methods for people to buy houses.

Keywords: House price; box plot; linear regression.

1. Introduction

House is now a necessity of human life and part of our daily consumption. However, with the development of society, the property market is gradually expanding, resulting in the price of housing has continued to rise in the past few years, and gradually difficult for people to afford, which has become a major problem in people's lives. And there are diverse factors affecting the house price, investigating which specific factors affect the house price has become a major goal of people [1].

In order to figure out the factors affecting house prices, many previous scholars have done related investigations. In 1990, the house price was analyzed by Case and Shiller with linear regression analysis, which proved that the house price is affected by the local population as well as the income of the residents, and the larger the values of these factors are, the greater the demand for the house increases, which leads to the greater price of the house price. And the property market is more willing to increase the price of houses to achieve the purpose of increasing its profit. In 1991, the same conclusion was also proved by Poterba, unlike before, he proved that the cost of building houses is also an important influence point, the greater the cost in order to balance the cost of the corresponding house price will be more expensive, so as to profit from it [2]. What's more, Quigley proposed in 1999 that people can learn from the local population, employment and economic prosperity. Quigley suggested in 1999 that one can predict the trend of house price from the local population, employment level and economic prosperity and economic index [3]. The conclusions and opinions of the previous researchers have proved that house prices are affected by some specific factors, and even these data can be used to predict the trend of house prices. And Rosen also proposed in 1974 to investigate the house price factors using hedonic model, which contains the accessibility of transport, and the surrounding natural scenery [4]. HM et al. also suggested that, in the case of Hong Kong, how the external transport environment affects the house price [5]. Poudyal et al. also argued that the price of the residence is greatly influenced by the surrounding environment, such as the proximity of shopping malls and schools, and local transport [6]. And in 2011, Aluko suggested that the location of a home is a determinant of the price of a house [7]. Hao and Ma in 2022 suggested that the price of a home is influenced by the surrounding public service infrastructure, such as healthcare, schools, and other infrastructures [8]. Li et al. suggested that urban water system has a great comfort for human life, so it also has a certain impact on house prices [9], and according to 2022, Emilio et al. suggested that,

in the case of Spain, urban green space also has an inseparable impact on house prices [10]. And all the above researchers have proved that the location's convenience has an impact on house prices. There is a positive relationship between the accessibility of a house and its price. However, these researchers have limitations and cannot analyse them comprehensively. For example, they only investigated the external social conditions that cause differences in house prices and did not consider the impact of the house's own facilities on its price, which may affect the house price to some extent.

Therefore, this paper analyses the house prices in Chicago suburbs by using multiple linear regression to investigate eight factors of the house itself. Bedroom: number of bedrooms, room: number of rooms, space: the house's size. (in square feet), lot: width of a lot, tax: amount of annual tax, bathroom: number of bathrooms, garage: number of garages, condition: condition of house. These factors are taken as variable x and house price is taken as dependent variable y so as to carry out a multivariate linear analysis. The results will first be used to determine whether the data chosen is appropriate and representative. Based on the regression coefficients of each factor and whether the analytical terms are significant, this article will analyse which factors affect the house price and describe the degree of influence, and article will reflect on the limitations of the results and think why the limitations occur, and this article will analyse the data with a high degree of overlap, and in this article will calculate the regression coefficient value and correlation coefficients to illustrate the specificity of the data. At the end of the paper, some insights will be provided on how to buy a house at a discount, and how to combine these influencing factors to purchase a house at a relatively more favourable price.

2. Methods

2.1. Data Source

The Kaggle website is the source of the data for this literature and was published by Tawfik Elmetwally in 2023. 100 house prices in the suburbs of Chicago were collected.

2.2. Variable Introduction

Table 1. Basic indicators

name	simple size	minimum value	maximum value	average value	standar deviation	meidan
Price	100	34	88	56.61	12.756	55.5
Bedroom	100	1	8	3.22	1.345	3
Garage	100	0	2	0.84	0.813	1
Lot	96	24	50	33	8.58	30
Bathroom	100	1	3	1.48	0.531	1.5
Condition	100	0	1	0.22	0.416	0
Space	94	539	2295	1108.745	454.073	987
Room	100	4	12	6.54	1.684	6
Tax	96	425	2727	910.688	445.819	816

The table 1 above shows the data to be examined in this post, a total of 886 pieces of data, and a preliminary analysis of them.

2.3. Method Introduction

This essay will investigate the relationship between house prices and the factors that may influence them Investigating whether there is an influential relationship between them and to what extent. In the article the house price will be studied as y and the remaining eight possible variables as x.

Firstly, the model equation will first be written using the regression coefficient values calculated as the coefficients of model x and y on the other side of the equation. Next, use the R-squared values

to see if they fit the model. The VIF value or tolerance can also be used to analyze this. The presence of VIF values above 5 and tolerance below 0.2 indicates a covariance problem. Subsequently the significance of X can be analyzed. If the p-value is lower than 0.05 or 0.01, then it means that there is a relationship of effect of X on Y. Next, after stating whether there is an effect or not, this essay will analyze the regression coefficient B value to analyze the extent to which the variable affects the dependent variable. Finally, this essay will summarize the results.

3. Results and Discussion

3.1. Descriptive Analysis

The x-axis in figure 1 represents the number of rooms and the y-axis represents the home price. The graph displays an increase in the number of rooms and house prices, up to \$88,000. When the number of rooms is around five to seven, the price of a room is more even, staying between \$60,000 and \$70. If the number of rooms is higher than eight, the price of the house will have a great increase. In the range of equal to four rooms versus greater than or equal to eight rooms, the price of the home can vary a lot. The price of a home is generally the lowest when the number of rooms is below four. It has an upper limit as large as the 25th percentile of the price of five rooms at \$25,000.

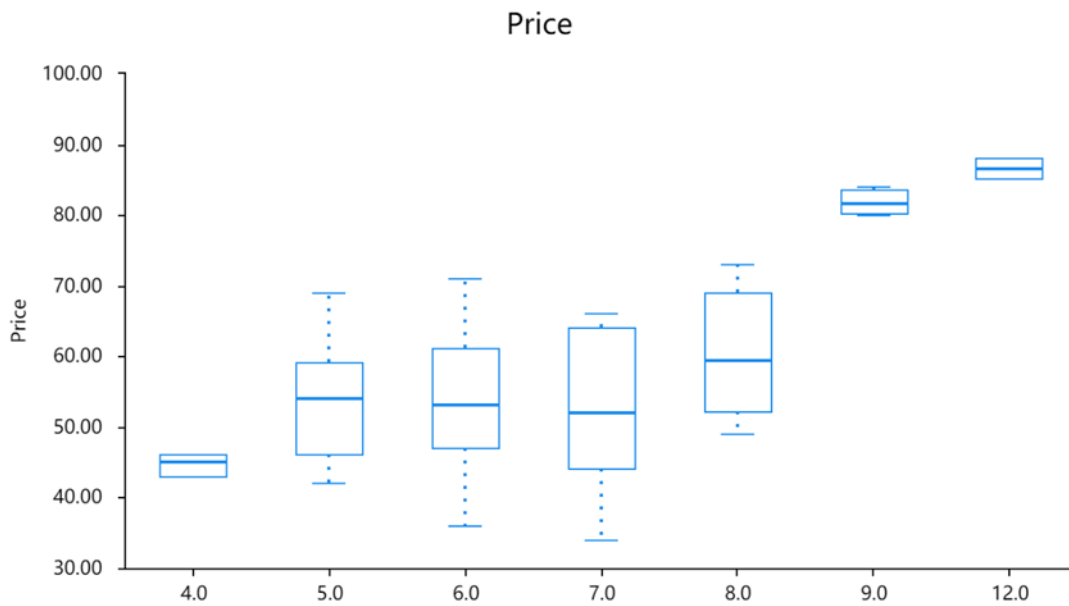


Fig. 1 Price's box plot analysis of number of rooms

In figure 2 the relationship between garage and house price is analysed. In the graph, the x-axis represents the number of garages. In this graph, it can be noticed that the price of a house with a garage is significantly different from the price of a house without a garage. In the case of 0 garage, the minimum house price is \$34,000 which is the lowest price among the four cases. And its upper bound is \$1,000 less than the 75th percentile of home prices with a garage. In the remaining garage scenario, the house with the highest price is a 2-garage house. The price is \$88,000. Based on the above, it is found that the price of a house is linked to the number of garages, i.e., the more the number of garages owned, the higher the house price.

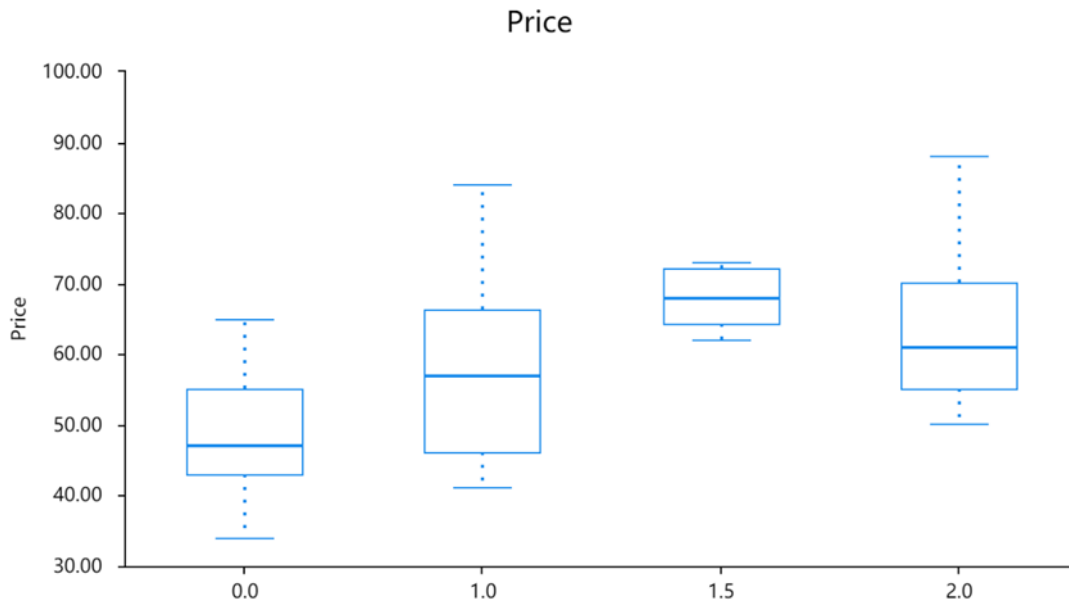


Fig. 2 Price's box plot analysis of number of garages

The correlation between house prices and the number of bathrooms is examined in figure 3. The data shows that prices are relatively flat when the number of bathrooms is between three and five. The median is around \$55,000, while the upper limit is between \$81,000 and \$84,000, with a difference of only \$3,000. For seven to eight bathrooms, there's a slight bump in comparison, with a ceiling of \$88,000, \$4,000 more than the highest price in the previous grouping. At one to two bathrooms, the difference in price is huge, capped at \$70,000, as well as more than \$10,000 different from the previous grouping. But the lowest value of the house price occurs when there are four bathrooms. Based on the above analysis, it is initially found that the house price is related to the number of bathrooms, for instance the more its number the more expensive the house price.

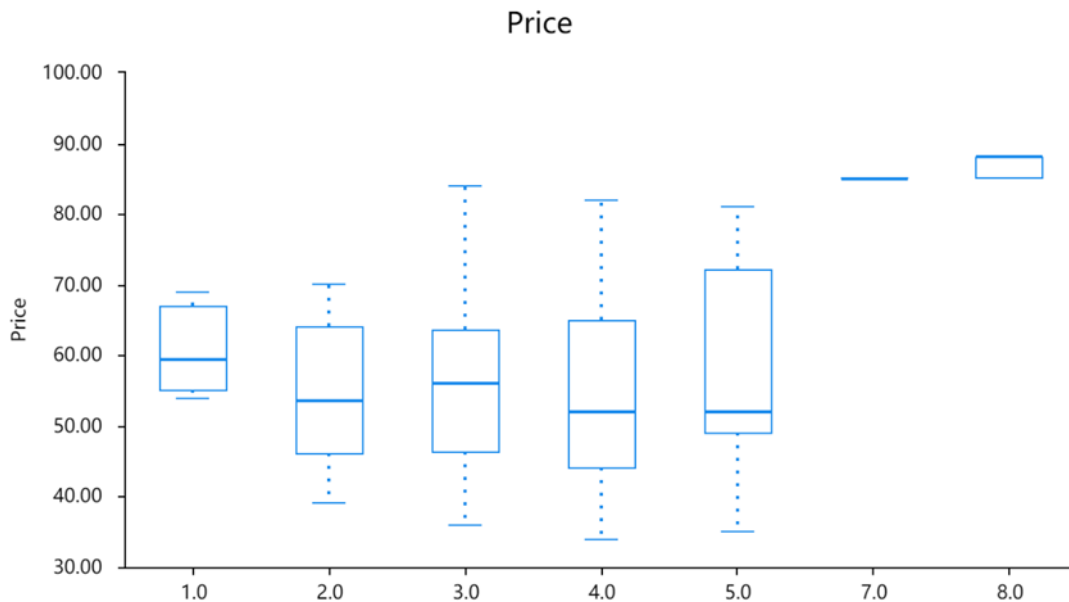


Fig. 3 Price's box plot analysis of number of bathrooms

3.2. Correlation Coefficients

From table 2, this paper uses correlation analysis to investigate how the eight variables are related to house prices. From the figure, it can be seen that the correlation coefficient between Price and Bedroom is 0.321, which is greater than 0, this indicates that there is a significant positive correlation

between price and bedroom. Space also has a positive correlation with price at the same time. The value between them is 0.722 which is also greater than 0. The correlation coefficient value between Price and Garage is also greater than 0 which is 0.553. Price and Garage have a significant positive relationship, as indicated by this. The correlation coefficient values for Price and Room are 0.574 and show a significance of 0.01, Price and room are directly proportional. However, the figure of Price and Condition is 0.094 and value of p is 0.351>0.05, so there is no relationship of influence between Price and Condition. The value of p between Price and Bathroom, lot, and tax are 0.544, 0.458, and 0.499 which are all greater than 0. 0.499, all of which are larger than 0, indicating the relationship is proportional between Price and all three of these variables.

Table 2. Pearson correlation coefficient

	Bedroom	Space	Garage	Room	Condition	Bathroom	Lot	Tax
Price	0.321**	0.722**	0.553**	0.574**	0.094	0.544**	0.458**	0.499**
* p<0.05 ** p<0.01								

3.3. Regression Model Results

In table 3, from the above table, it can be seen that the eight influencing factors are set as Price is taken as the dependent variable for the linear regression analysis. From the above table, it can be seen that the model equation is:

$$Price = 22.378 - 3.218 * Bedroom + 0.012 * Space + 4.331 * Garage + 1.186 * Room + 0.786 * Condition + 6.144 * Bathroom + 0.275 * Lot + 0.003 * Tax \tag{1}$$

Table 3. Linear regression analysis results

	Nonnormalized coefficient		standardization coefficient	t	p	collinearity diagnostics	
	B	S.E.	Beta			VIF	tolerability
constant	22.378	5.172	-	4.327	0.000**	-	-
Bedroom	-3.218	1.206	-0.341	-2.669	0.009**	4.406	0.227
Space	0.012	0.005	0.411	2.438	0.017*	7.66	0.131
Garage	4.331	1.177	0.269	3.679	0.000**	1.435	0.697
Room	1.186	1.488	0.154	0.797	0.428	10.086	0.099
Condition	0.786	2.234	0.025	0.352	0.726	1.31	0.763
Bathroom	6.144	2.587	0.255	2.375	0.020*	3.113	0.321
Lot	0.275	0.114	0.18	2.416	0.018*	1.499	0.667
Tax	0.003	0.004	0.088	0.737	0.463	3.825	0.261
R ²	0.714						
adjusted R ²	0.684						
F	F (8,77)=24.035, p=0.000						
D-W value	1.948						
dependent variable: Price							
* p<0.05 ** p<0.01							

The R-squared value is 0.714, which indicates that these variables explain 71.4% of the variation in Price. The model's F-test revealed that (F=24.035, p=0<0.05), which implies that at least one of them will have an impact on Price. In addition, this paper tested the model for multicollinearity and The VIF value in the model was found to exceed 10, which suggests that there is a problem with covariance. The regression coefficient value of Bedroom is -3.218 (t=-2.669, p=0.009<0.01), which means that Bedroom has a significant negative impact on Price. The regression coefficient value of Space is 0.012 (F=0.012), which means that at least one of these eight factors, thus the price will be negatively affected. The value of Space is 0.012 (t=2.438, p=0.017<0.05) which means that there is a significant positive impact on the price. The value of regression coefficient for Garage is 4.331 (t=3.679, p=0.000<0.01) which means that Garage will have a significant positive effect on Price.

The value of Room is 1.186 ($t=0.797$, $p=0.428>0.05$), it implies that the relationship between room and price is not influenced. The value of coefficient for Condition is 0.786 ($t=0.352$, $p=0.726>0.05$), implying that the condition has no impact. The value of regression coefficient of Bathroom is 6.144 ($t=2.375$, $p=0.020<0.05$), implying that Bathroom will have a significant positive impact on Price. the value of regression coefficient of Lot is 0.275 ($t=2.416$, $p=0.018<0.05$), assuming that Lot will have a substantial positive impact on Price. The value of regression coefficient of Tax is 0.003 ($t=0.737$, $p=0.463>0.05$), which means that Tax does not affect Price. The analysis can be summarized as follows, Space, Garage, Bathroom, Lot will have a significant positive impact on Price. Bedroom has a significant negative impact on Price. But Room, Condition, Tax does not have any effect on Price.

4. Conclusion

This article concludes that Space, Garage, Bathroom, Lot will have a major impact on the price. But the price of suburban Chicago will be significantly affected by Bedroom. However, Room, Condition, and Tax do not have any effect on Price. However, there are some limitations in this paper, such as only choosing the price of Chicago suburbs, but not combined with the price of urban areas, if also combined with the price of urban areas, the conclusion will be more convincing. Finally, to sum up this article, if people need to buy a house but have insufficient money, they can choose a house with a smaller space and lot, then a smaller amount of garage and bathroom, and finally a house with more bedrooms. This combination will buy a house more favourably.

References

- [1] Case K E, Shiller R J. Is there a bubble in the housing market?. *Brookings papers on economic activity*, 2003, (2): 299-342.
- [2] Poterba J M, et al. House price dynamics: the role of tax policy and demography. *Brookings Papers on Economic Activity*, 1991, (2): 143-203.
- [3] Quigley J M. Economic fundamentals in local housing markets: evidence from US metropolitan regions. *Journal of Regional Science*, 2006, 46(3): 425-453.
- [4] Rosen S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Polit. Econ.*, 1974, 82: 34-55.
- [5] So H M, et al. Estimating the influence of transport on house prices: evidence from Hong Kong. *Journal of Property Valuation and Investment*, 1997, 15: 40-47.
- [6] Poudyal N C, et al. A hedonic analysis of the demand for and benefits of urban recreation parks. *Land Use Policy*, 2009, 26(4): 975-978.
- [7] Aluko O. The effects of location and neighborhood attributes on housing values in metropolitan Lagos. *Ethiopian Journal of Environmental Studies and Management*, 2011, 4(2): 69-82.
- [8] Zhao Lili, Jiao Jiwen. Grey correlation analysis of factors affecting housing prices. *Statistics and Decision*, 2007, (23): 2.
- [9] Luo Yubo. Analysis of Factors Influencing House Prices: Quantile Regression Method. *Statistics and Decision Making*, 2011, (6): 2.
- [10] Fan Yunqi, Wang Yiming. Research on Regional Differences and Temporal Changes of Factors Influencing Housing Prices in China. *Journal of Guizhou University of Finance and Economics*, 2014.