Proceedings of Business and Economic Studies, 2024, Volume 7, Issue 6
http://ojs.bbwpublisher.com/index.php/PBES
Online ISSN: 2209-265X
Print ISSN: 2209-2641

Analysis and Evaluation of Housing Price Factors Using Mathematical Modeling

Xing Lyu*

Ningbo Yinzhou Senior High School, Ningbo 315206, Zhejiang Province, China

***Corresponding author:** Xing Lyu, 3472324729@qq.com

Copyright: © 2024 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In recent years, the real estate industry has achieved significant progress, driving the development of related sectors and playing a crucial role in economic growth. However, rapid real estate market expansion has led to challenges, particularly concerning housing prices, which have drawn widespread societal attention. This article explores the theories of housing prices, analyzes factors influencing them, and conducts an empirical investigation of the impact of representative factors on ordinary residential prices. Using regression analysis and the entropy weight method, a mathematical model was developed to examine how various factors affect housing prices.

Keywords: Mathematical modeling; Regression analysis; Housing price; Formation factors; Multiple linear regression; H ypothesis testing; Multiple decision coefficients

Online publication: December 23, 2024

1. Empirical study

1.1. Data source

Housing prices have consistently been a topic of interest, with prospective buyers closely monitoring trends to mitigate potential property value depreciation. Traditional research methods focus on analyzing longitudinal housing price data in specific regions over time to predict future trends and estimate future prices. This study utilized data published by the Singapore government, which included variables such as house location, age, quality, structure, and interior. The primary aim was to analyze the key factors influencing housing prices using multiple linear regression [1-3].

Regression analysis was employed to describe the linear dependency relationship between a dependent variable (housing price) and independent variables (age, area, and quality of the house). This method helps establish a regression model to predict trends in housing prices. Calculations, such as the mean and standard deviation, were conducted using Excel, alongside correlation assessments and scatter plots to visualize the relationships between variables [4-6].

1.2. Symbols and their meaning

(1) α: Housing price

(2) ε: Age of the house

(3) v: Area of the house

(4) μ: Quality of the house

1.3. Data analysis

1.3.1. Correlation analysis of housing price with age, area, and quality

The dataset analyzed includes information on age, area, quality, and price of houses, with a sample shown in **Table 1**:

Table 1. Dataset with information on age, area, quality, and price of houses

Age	Area	Quality	Price
10	1,470	6	177,000
72	1,176	6	114,500
47	816	5	110,000
25	1,842	9	385,000
52	1,360	5	130,000
44	1,425	6	180,500
10	1,739	8	172,500
2	1,720	7	196,500
0	2,945	10	438,780
36	780	5	124,900
2	1,158	6	158,000
87	1,111	5	101,000
1	1,370	8	202,500
89	1,710	7	140,000
...	...	...	...

Basic calculations yielded a mean housing price of $237,500, with an average estimation error of 286,448.96. These values provided initial insights into average housing prices and their fluctuations.

1.3.2. Correlation between housing price and age of the house

A scatter plot and trend line analysis revealed a negative correlation between housing prices and age (**Figure 1**). The regression equation derived was:

The coefficient of determination ($R^2 = 0.3367$) suggests a moderate linear relationship.

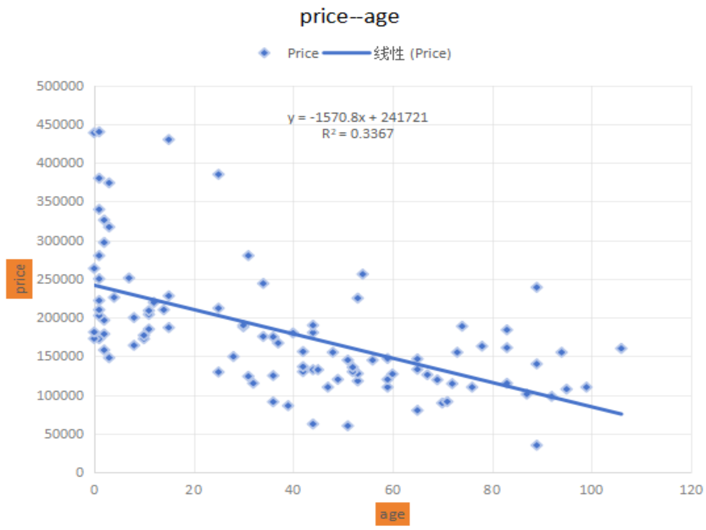

Figure 1. Regression analysis diagram of housing price and age of the house

1.3.3. Correlation between housing price and area of the house

Regression analysis of the housing area demonstrated a positive correlation with housing prices (**Figure 2**). The equation obtained was:

The coefficient of determination ($R^2 = 0.5779$) indicates a relatively strong linear relationship.

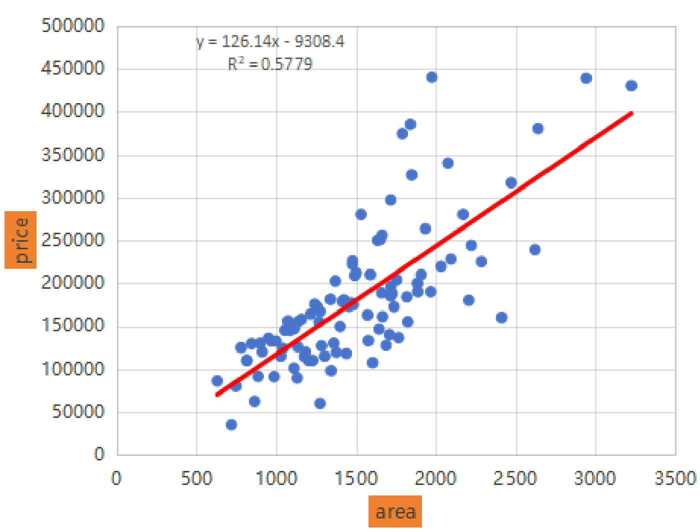

Figure 2. Regression analysis diagram of housing price and area of the house

1.3.4. Correlation between housing price and quality of the house

Analysis of house quality revealed a strong positive correlation with housing prices (**Figure 3**). The regression equation derived was:

The coefficient of determination ($R^2 = 0.6148$) suggests a significant linear relationship.

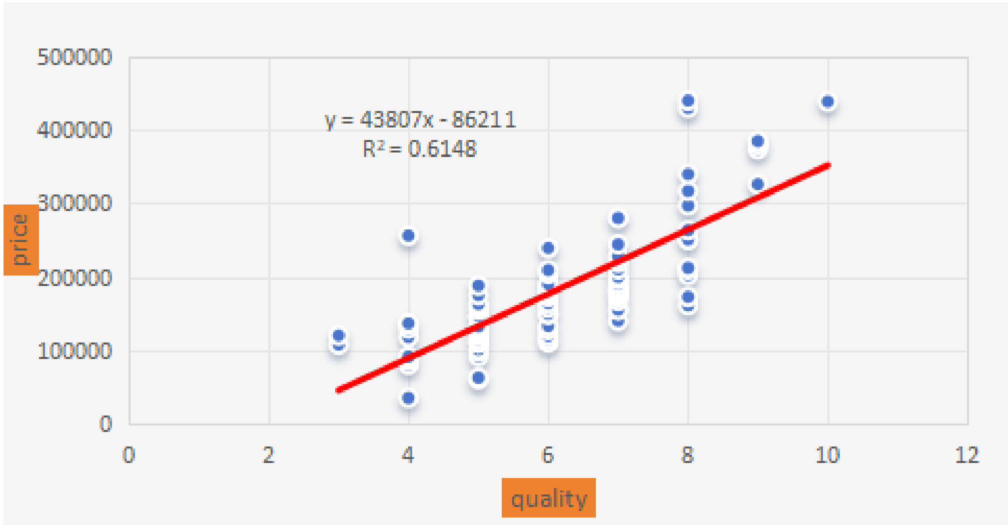

Figure 3. Regression analysis diagram of housing price and quality of the house

1.4. Summary

In summary, the analysis of different variables reveals correlations between the age of the house, area of the house, quality of the house, and housing price. The regression statistics are shown in **Table 2**.

Table 2. Regression statistics, regression coefficients, and standard errors

Multiple R	R Square	Adjusted R Square	Standard Error	Observed Value
0.88234	0.778523	0.771602	39,162.99	100
	Regression Coefficients		Standard Errors	
Intercept	-26,684.7		24,219.19	
Age	-700.566		159.4284	
Area	79.45805		10.3143	
Quality	18,833.83		4,062.478	

The regression equation describing the relationship between housing price (α), age of the house (ε), area of the house (ν), and quality of the house (μ) is as follows:

This equation quantitatively illustrates how the age, area, and quality of a house influence its price.

2.1. Estimating price using each model

The above four models were utilized to estimate housing prices. Each model was applied as an Excel formula to compute the “estimated” price for the provided data points. **Table 3** illustrates the observed prices and the corresponding estimations for each model.

Table 3. Observed prices and the corresponding estimations for each model

Age	Area	Quality	Price	Model A	Model B	Model C	Model D
10	1,470	6	177,000	226,011	175,912	176,631	196,116
72	1,176	6	114,500	128,609	138,868	176,631	129,320
47	816	5	110,000	167,884	93,508	132,824	99,396
52	1,360	5	130,000	160,029	162,052	132,824	139,118
25	1,842	9	385,000	202,446	222,784	308,052	271,667

Model formulations: Model A, $\alpha = 1{,}570.8\epsilon + 241{,}721$; Model B, $\alpha = 126{,}14\nu - 9{,}308.4$; Model C, $\alpha = 43{,}807\mu - 86{,}211$; Model D, $\alpha = -26{,}684.7 - 700.566\epsilon + 79.45805\nu + 18{,}833.83\mu$.

2.2. Analysis of estimation errors

The estimation errors for each model were calculated as the difference between the observed prices and the estimated prices. **Table 4** summarizes these errors and their corresponding margins.

Table 4. Estimation errors for each model

Price	Error A	Error B	Error C	Error D
177,000	-49,011	1,088	369	-19,116
114,500	-14,109	-24,368	-62,131	-14,820
110,000	-57,884	16,492	-22,824	10,604
130,000	-30,029	-32,052	-2,824	-9,118
385,000	182,554	162,216	76,948	113,333
Margin	40,089	33,590	20,308	23,345

Model C demonstrated the lowest margin of error, suggesting that it may provide the most accurate estimation of housing prices among the four models.

2.3. Result analysis

This study applies stepwise multiple linear regression analysis to evaluate data such as permanent population, area, and housing quality using the established linear model. The analysis demonstrates valid reference values for estimating future housing prices. Among the factors considered, housing quality exhibits a significantly closer linear relationship with housing price compared to other variables.

Certain limitations are inherent in this study. The multiple linear regression model developed for factors influencing housing prices is based on a specific region and timeframe, making it susceptible to regional and temporal particularities. To predict future housing price trends with greater accuracy, it is strongly recommended to consider more comprehensive and effective factors, as well as to integrate the inference model derived in this study. This approach is expected to support government adjustments to the real estate market and provide valuable guidance for purchasers in making informed decisions [7,8].

The mathematical modeling methodology used in this study abstracts real-life problems into mathematical frameworks, translating them into mathematical language and constructing models to address these issues.

The mathematical modeling process includes identifying problems from practical scenarios, proposing and analyzing problems, establishing models, determining parameters, calculating solutions, verifying results, refining models, and ultimately resolving real-world issues.

Data analysis forms an integral part of this process. It involves collecting data from research subjects, organizing and analyzing it using mathematical methods, and drawing inferences to develop insights about the subject of study. The key stages include data collection, organization, information extraction, model construction, inference, and conclusion formation [9-12].

Regression analysis, as applied in this study, is a critical statistical model and a valuable tool for data analysis. By employing Excel for regression analysis, the data processing workflow is streamlined, enabling enhanced problem-solving skills and fostering scientific literacy. The demonstration using Excel highlights the transformation of non-linearity into linearity, illustrating the practicality of this technique [13-18].

Furthermore, an exponential function model, with its larger R^2 value, demonstrates a superior fit compared to a quadratic function model. Additional exploration using Excel allows for the incorporation of higher-order functions, such as cubic and logarithmic functions, thereby broadening perspectives and showcasing the versatility of mathematical approaches. This reinforces the notion that mathematics originates from life, transcends it, and finds application in real-world scenarios [19].

Ultimately, there is no single "best" mathematical model; rather, continuous improvement and refinement of models ensure their ability to serve human needs effectively.

Disclosure statement

The author declares no conflict of interest.

References

[1] Zhu S, Qian Y, Guo H, 2022, Statistical Inference Method of Factors Affecting House Prices. Statistics and Management, 37(5): 58–63.

[2] Li Z, 2018, Research on the Comprehensive Measurement Model of Housing Price Data Based on the Prediction Range Theory, dissertation, Wuhan University of Science and Technology.

[3] Li G, Guo H, 2020, Research on Nanjing Housing Price Forecast Based on GM (1,1) Model. Journal of Jianghan University (Natural Science Edition), 48(2): 4.

[4] Xie S, 2022, Research on the Teaching Strategies of Mathematical Modeling Literacy Training for Secondary Vocational Students. Star of the Century (High School Edition), 2022(10): 196–198.

[5] He J, Zhang C, Jiang R, 2019, Teaching Principles and Strategies of Penetrating Mathematics Modeling in High School Mathematics Classroom. Reading and Writing, 16(25): 178.

[6] Zhou H, 2023, Exploration on the Cultivation Strategy of Core Literacy Ability in High School Mathematics Teaching. Electronic Magazine of the New Education Era (Teacher Edition), 2023(3): 40–42.

[7] Liu S, 2023, Using the Zero-Point Problem of the Derivative Study Function. High School of Mathematics, Physics and Chemistry, 2023(15): 49–50.

[8] Li H, 2015, The Design Framework of "Learning-Oriented" – Takes the Micro-Lesson Design of "Triangular Function" in High School Mathematics as an Example. Middle School Mathematics Teaching Reference, 2015(Z3): 163–167.

[9] Ao L, 2021, Analysis of Mathematics Derivatives, Problem Solving and Teaching Countermeasures, dissertation, East China Normal University.

[10] Zhong G, 2020, A to Z for the Derivative Hidden Zero Problem. Mathematics for Middle school Students, 2020(23): 6–8.

[11] Gu X, Liang R, 2012, Research on Sales Price Prediction of Commercial Housing in Zhengzhou Based on ARIMA Model. Financial Theory and Practice, 2012(1): 51–54.

[12] Luo B, Hong Z, Wang J, 2020, Application of Multiple Linear Regression Statistical Models in House Price Prediction. Computer Age, 2020(6): 51–54.

[13] Zhao J, Ma Z, Tang H, 2020, Comparison of Carbon Emission Predictions Between BP Neural Networks and Multiple Linear Regression Models. Technology and Industry, 20(11): 172–176.

[14] Zhu Z, Jin Z, 2010, An Empirical Study of Decision-Making, Decision-Making Team and Procedural Rationality in the Strategic Decision-Making Process. Nankai Management Review, 13(1): 4–14 + 49.

[15] Li A, 2018, Application of Mathematical Modeling in Data Mining. Information System Engineering, 2018(1): 124–125.

[16] Zhou Y, 2019, Two Conclusions of “Linear Regression Analysis” in High School Mathematics from Different Perspectives. Journal of Chongqing Normal University (Natural Science Edition), 36(4): 131–136.

[17] Zhang H, 2017, The Cultivation of Mathematical Core Literacy – Thinking from the Dimension of Independent Inquiry. Curriculum Education Research, 2017(20): 144.

[18] Yao J, 2022, Research on Cultivating Students’ Mathematical Modeling Ideas Under the Perspective of Core Literacy. Primary School, 2022(4): 78–80.

[19] Liu H, 2021, Teaching Research on the Cultivation of Mathematical Modeling and Data Analysis Literacy Among High School Students – Taking the Teaching of Regression Analysis as an Example. Middle School Teaching Reference, 2021(32): 7–8 + 28.

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.