# Housing Prices in Fujian Province: A Linear Regression Analysis of Fuzhou

Hanyu Zhuang *

School of No.5 High School, Quanzhou, Fujian, China

* Corresponding Author Email: hanyuzoe_0330@outlook.com

**Abstract.** This study used multiple linear regression to investigate the key points affecting housing prices in Fuzhou, Fujian Province. According to data from 20 residential communities, 5 variables are examined: building area, building age, distance to the nearest metro station, distance to the central business district (CBD), and proximity to key schools. The results show that metro distance has the strongest effect: one kilometer closer to a metro station increases price by 4,856 RMB/m². School proximity adds 8,923 RMB/m², while each additional year of building age reduces price by 413 RMB/m². The model explains 71.2% of price variation ($R^2$ = 0.712), and all variables are statistically significant. These findings support land value capture for metro funding and education equity policies to reduce housing pressure. Limitations include a small sample size (20 communities) and cross-sectional data. Future research should expand the number of the samples and include more detailed values especially variables.

**Keywords:** House price prediction; linear regression; ordinary least squares

## 1. Introduction

There has been a strong fluctuation in housing market of Fujian Province, a coastal economics city. During these decades, the real estate, sales areas and sales values all decreased. From January to August 2025, real estate investment dropped by 30.8%, sales area dropped by 15.9% and sales value decreased by 11.7% from Fujian Provincial Bureau of statistics [1].

Housing is an important part of asset for Chinese urban households. The reason for researching will give some factors as follows. For instance, economics significance, according to the work carried out by Zhang et al. [2], the price of housing almost accounts for around 80% of the share of household's wealth. Furthermore, real estate and related industries contribute around 30% of national GDP. Finally, the housing prices of Fujian is 10,474RMB/m²[3]. Therefore, understanding the factors of impacting housing prices is essential for people of all roles, like policymakers, investors, and families.

Previous studies have examined housing price determinants in Fuzhou. Huang et al. [4] analyzed the impact of Fuzhou Metro Line 1 using a hedonic price model and found that distance to metro stations significantly affects housing prices. Lin et al. investigated 3,386 communities in central Fuzhou and placed primary school quality as the most important factor, followed by building age [5]. However, research has largely focused on specific infrastructure projects or complex specific models, interpretable skeletons using publicly available data.

The main contribution from this study has twofold. Firstly, this study provides many datasets and available methodology to solve the questions, making the research available for other researchers. Secondly, the paper's findings offer some limitations and the direction of future development, like connection between different factors and housing prices, providing policymakers or even other roles of the whole society.

The reminder of this paper is organized as follows. Section 2 demonstrates the dataset and methodology. Section 3 presents experiential results and real findings and key findings. Section 4 discusses the limitations and conclusion.

## 2. Method

### 2.1. Dataset Preparation

The dataset was converted from publicly available sources, like housing prices, building area, building age, and school area information, were obtained from the China Housing Price Platform (creprice.cn), Fang.com, and Anjuke. This guidance provides useful data for residential communities across Fujian Province. Distance measurements to metro stations and the central business district (CBD) were calculated using the Baidu Maps application programming interface (API). Marketing trend data was collected from the China Real Estate Information System (CREIS) monthly report. Official statistics, including actual estate investment and sales figures, selected from the Fujian Provincial Bureau of Statistics [1].

The final dataset contains 20 data items; each represented one residential community. For each community, 6 variables were recorded. The variable is housing price per square meter (RMB/m²), which is a continuous digital variable, making this a regression task. Independent variables include building area (square meters), building age (years), distance to the nearest subway station (km), distance to the central business district (CBD), and a variable indicating whether the community is within the enrollment range of key schools (1 means yes, 0 means no).

Before using the data, some steps were done. First, if a community did not have complete data, it was removed from the list. Second, if a community had a price that was too low or too high (below 5,000 or above 80,000 RMB/m²), it was also removed because these prices are not typical for normal housing. All these steps were done using Excel. After then, the data was ready for the regression analysis.

### 2.2. OLS and Specification Model

Ordinary Least Squares (OLS) is a basic way to find the relationship between one thing (like house price) and several other things (like size, age, location), minimizing the sum of Squared Residuals: OLS provides the Best Linear Unbiased Estimators [6-8].

$$\text{Price} = \alpha + \beta_1(\text{area}) + \beta_2(\text{age}) + \beta_3(\text{Metro-Distance}) + \beta_4(\text{CBD-Distance}) + \beta_5(\text{school}) + \varepsilon(\text{others})$$

Here, $\alpha$ is the starting price when all other factors are zero, just like intercept. The $\beta_1$ to $\beta_5$ tell how much the price changes when the study changes one factor while keeping everything else the same, just like factors.

The main reason to use OLS is that it is simple and easy to understand, which is the advantages [9-11]. These results are easy to explain. However, OLS has some drawbacks. The data might not truly for the real situation, the OLS just like a model which affected by the figures research imported, it assumes the relationship between the variables is a straight line.

In this study, OLS was run using Excel. The model looks at how price changes with area, age, metro distance, CBD distance, and school. The study expected area and school to increase price, and age, metro distance, and CBD distance to decrease price. The study checked how good the model is by looking at $R^2$ (how much of the price change is explained by the model) and p-values (whether the results are meaningful). Table 1 summarizes the variable.

**Table 1.** Three Scheme comparing.

| Variable | Description | Unit | Expected sign | Source |
|---|---|---|---|---|
| Price | Housing price | RMB/m² | - | CREIS |
| Area | Building floor area | m² | + | Lin et al. [2] |
| Age | Years since construction | years | - | Lin et al. [2] |
| Metro | Distance to metro station | km | - | Huang et al. [1] |
| CBD | Distance to city center | km | - | Huang et al. [1] |
| School | Key school catchment | binary | + | Lin et al. [2] |

## 3. Results and Discussion

The dataset contains 20 residential communities in Fuzhou. Table 2 shows the basic statistics for each variable.

**Table 2.** Descriptive Statistics

| Variable | N | Mean | Std Dev | Min | Max |
|---|---|---|---|---|---|
| Price | 20 | 28247 | 11234 | 14013 | 52000 |
| Area | 20 | 119.7 | 30.2 | 80 | 200 |
| Age | 20 | 11.2 | 6.8 | 2 | 25 |
| Metro | 20 | 1.38 | 0.89 | 0.05 | 3.00 |
| CBD | 20 | 5.23 | 3.15 | 1.2 | 12.0 |
| School | 20 | 0.15 | 0.37 | 0 | 1 |

The study has several conclusions. In terms of Metro Distance, the coefficient for metro distance is -4,856, meaning each kilometer further from a metro station reduces the house price by about 4,856 RMB per square meter.

Living more closer to a metro station, price would be much more. Because people tend to be close to public transportation. As for the key School, the coefficient for key school is +8,923. This means that if a community is in a key school catchment area, the price is about 8,923 RMB per square meter higher. This indicates that parents are willing to invest significantly in sending their children to good schools. This finding aligns with previous studies. Additionally, the coefficient of the building Age is -413. This means that every year the building gets older, the price drops by 413 RMB per square meter. This is about 1.5% of the average price each year. Older buildings cost less because they have more wear and tear. While the coefficient for area is +156. This means that for every additional square meter of space, the price increases by 156 RMB per square meter. Bigger houses cost more, which is what this paper expects. In terms of CBD Distance, its coefficient f is -2,135. This means that every kilometer further from the city center, the price drops by 2,135 RMB per square meter. People want to live close to downtown, where there are more jobs and activities.

There has been some comparison with previous studies. The results from this study can match with earlier research. Huang et al. [4] found that metro distance has a strong negative effect on prices. The result of the study is similar after adjusting for different measurement scales. Lin et al. [5] found that school quality was the most important factor in their large study of 3,386 communities. In the study, school also had a large positive effect, though the metro was slightly more important in this smaller sample. The difference might be because the sample research is smaller or because the study looked at different areas of the city. Still, both studies agree that schools, metro, and building age have a significant effect on house prices.

However, the study still has some obvious limitations that should be noticed. First, the study had a small sample size. Only 20 communities were selected as samples. The number was not large enough; the sample may not be fully persuasive. Hence, it is difficult to generalize the results in all similar situations. Second, the study only used cross-sectional data. All the data was collected at one specific time. Because of this, the study can only show the current situation at that moment, but it cannot predict how things will change over a long period of time or find any development trends. Third, the linearity assumption might be too simple for the study. The study assumed that the relationship between variables was linear, but the relationship between things is often more complicated. Finally, there may be ignored variables in the study. The study did not take every possible factor into account. This missing part affects the accuracy of the results and makes the study less reliable.

## 4. Conclusion

This study used multiple linear regression to analyze housing prices in Fuzhou, Fujian. The results show that metro distance has the strongest effect: one kilometer closer to a metro station increases

price by 4,856 RMB/m². School close to 8,923 RMB/m², while each additional year of building age reduces price by 413 RMB/m². The model explains 71.2% of price variation ($R^2 = 0.712$), and all variables are statistically significant. These findings support land value capture for metro funding and education equity policies to reduce housing pressure. The study has limitations, including a small sample size (20 communities) and cross-sectional data. Future research should expand the number of samples, use detailed present data, and include more detailed variables such as the quality of the wall in the house and floor level.

## References


[1] Fujian Provincial Bureau of Statistics. Fujian real estate market report, January–August 2025. Fuzhou: Fujian Provincial Bureau of Statistics; 2025.

[2] Zhang P, Zhuge AD, Tian Y. Wealth inequality in urban China: Sources and mechanisms. Fin Trade Econ. 2025;(7).

[3] Xu XY, Li CH, Qu N, Wang JW, Wu SD. MGWR analysis of second-hand housing prices in Fuqing City. J China West Norm Univ Nat Sci Ed. 2023;44(1):29.

[4] Huang CC, Wang XW, Li LN. Urban rail transit and housing prices: A case study of Fuzhou Metro Line 1. Geogr Res. 2021;40(10):2808–2822.

[5] Lin Z, Wang WL, Gong J, Lin T. Multi-scale effects of housing price determinants in Fuzhou central area. Trop Geogr. 2023;43(8):1536–1546.

[6] Craven BD, Islam SM. Ordinary least-squares regression. The SAGE dictionary of quantitative management research. 2011; 1:224-8.

[7] Dempster AP, Schatzoff M, Wermuth N. A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association. 1977 Mar 1;72(357):77-91.

[8] Phillips PC, Park JY. Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. Journal of the American Statistical Association. 1988 Mar 1;83(401):111-5.

[9] Pohlmann JT, Leitner DW. A comparison of ordinary least squares and logistic regression (1). The Ohio Journal of Science. 2003 Dec 1;103(5):118-26.

[10] Zdaniuk B. Ordinary least-squares (OLS) model. InEncyclopedia of quality of life and well-being research 2024 Feb 11 (pp. 4867-4869). Cham: Springer International Publishing.

[11] Rzhetsky A, Nei M. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. Journal of molecular evolution. 1992 Oct;35(4):367-75.