Data-Based Comparative Analysis of Housing Price per Square Meter Across Regions in South Korea - Including Comparisons with International Metropolitan Areas

Hyunsung KIM, Seohyeon PARK, Yuhwan KIM

Korea Institute of Energy Technology (KENTECH)
21 KENTECH-gil, Naju-si, Jeonnam, 58330, Republic of Korea
kimharry0610@kentech.ac.kr
parksh5295@kentech.ac.kr
kyh0711@kentech.ac.kr

1 Introduction

In recent years, South Korea has been experiencing severe overconcentration in the Seoul metropolitan area, leading to a rapid surge in real estate prices and a widening gap in property values between regions [1]. Property prices have persistently risen as population and infrastructure continue to concentrate in the metropolitan area, while non-metropolitan areas have not shown similar trends. This intensification of regional economic imbalance poses a serious national issue. Such disparities are likely to exacerbate socio-economic conflicts over time and could eventually lead to the disappearance of certain regions and hinder balanced national development. Therefore, this project aims to objectively analyze the current real estate market based on quantitative data.

The primary objective of this project is to analyze the distribution of housing prices per square meter across different regions within South Korea using administrative regional units, and to quantitatively evaluate regional disparities using statistical distributional metrics. Through this, the project seeks to objectively identify and compare the characteristics and degrees of regional price disparities within South Korea. Our primary research questions are whether there are significant structural differences in real estate prices in metropolitan areas compared to non-metropolitan areas within Korea, and whether there are structural differences in the distribution of housing prices per square meter between Korea and other countries. Here, "structural differences" refer to distributional features such as mean, variance, skewness, and kurtosis, which reflect the statistical characteristics of regional price gaps.

For the analysis of South Korean real estate prices by region, this project will utilize the Housing Official Appraised Price Data provided by the Ministry of Land, Infrastructure, and Transport (MOLIT). This dataset covers nationwide information on both multi-family and single-family housing, including key variables such as exclusive floor area, appraised price, and administrative location. The data to be analyzed will be based on the most recent available year, and South Korea will be divided into the Seoul Metropolitan Area (Seoul, Gyeonggi Province, and Incheon) and non-metropolitan areas for regional comparison. Comparable datasets from other countries will also be collected, and necessary adjustments will be made using exchange rates or purchasing power parity (PPP) to ensure cross-country comparability.

The dependent variable in this project is the housing price per square meter, while the independent variable is the regional classification. In South Korea, regions will be analyzed at the administrative level (e.g., by province, metropolitan city, or district), allowing for more granular comparison. The project will apply statistical measures such as mean, variance, skewness, and kurtosis to examine the distributional characteristics of housing prices across administrative regions. Through this approach, the project aims to evaluate how the phenomenon of metropolitan concentration in South Korea compares to that of other countries and to provide foundational data and direction for future policies regarding balanced national development.

2 Related Works

Recent studies on house price comparison using official valuation systems have applied diverse methodologies to analyze spatial and temporal patterns. In South Korea, cyclical variations in Seoul's housing market are examined using Granger causality tests and vector autoregression models, revealing short-term price impacts on transaction volumes [2]. This aligns with Tehran's housing market analysis [3], which identified spatial diversity in price-transaction ratios through generalized linear mixed models, emphasizing localized factors like employment rates and housing stock attributes. Both studies underscore the limitations of macroeconomic policies in addressing regional imbalance.

The impact of housing policies has been critically evaluated in comparative frameworks. [4] demonstrated that publicly funded rental housing in Korea, particularly purchased units, stabilizes prices more effectively than private-sector initiatives, highlighting the need for institutional support in supply strategies. This contrasts with other existing mortgage market analyses, which found that low-interest policy loans in Korea reduce prepayment risks in non-metropolitan areas, reflecting regional behavioral differences in housing investment [5].

Collectively, the literature emphasizes two critical insights:

1. Regional price disparities require specified spatial analysis rather than uniform policies.
2. Public-sector interventions show greater price stabilization potential than market-driven approaches.

Based on the insights of related works, we make a hypothesis that regional disparities in South Korea's official housing prices per unit area are significantly greater than those of international metropolitan markets, primarily due to the unique spatial concentration of economic, infrastructural, and policy factors in the Seoul metropolitan area. Throughout this research, we will apply various statistical methods on reference housing price data provided by MOLIT.

3 Methodology

3.1 Dataset Preprocessing

For this study, we utilized the **Housing Official Appraised Price Data** provided by the **Ministry of Land, Infrastructure, and Transport (MOLIT)**. The raw dataset, provided in CSV format, contains nationwide records of residential properties, including attributes such as administrative location (시도, 시군구), exclusive floor area (전용면적), and official appraised price (공시가격). Our primary variable of interest—the unit housing price per square meter (공시가격_㎡)—was derived during preprocessing.

The preprocessing steps were as follows:

1. **Variable Selection:** We retained only the relevant columns: 시도, 시군구, 전용면적, and 공시가격.
2. **Missing Value Removal:** Rows containing null values in any of the selected columns were removed to ensure data quality.
3. **Outlier Removal for Official Price (공시가격):** We applied the Interquartile Range (IQR) method to exclude extreme values. Observations outside the range
$$[Q1 - 1.5 \times \text{IQR},\ Q3 + 1.5 \times \text{IQR}]$$
were discarded, where $Q1$ and $Q3$ denote the first and third quartiles, respectively. This helped to reduce the impact of luxury housing on subsequent statistical analysis. The resulting distribution is shown in Figure 1.
4. **Unit Price Calculation:** We computed the unit price per square meter as:
$$\text{공시가격\_㎡} = \frac{\text{공시가격}}{\text{전용면적}}$$
5. **Outlier Removal for Unit Price (공시가격_㎡):** A second IQR-based filtering step was applied to the newly computed 공시가격_㎡ variable. This was essential to account for extreme values that could still remain even after total price filtering. The resulting distribution after this step is shown in Figure 2.

Since the outlier calculated for Official Price and Unit Price differ, we saved the preprocessing output in separated csv files.

The results of the two preprocessing steps are visualized in Figures 1 and 2. As shown, the distributions of both the total official price and the unit price per square meter exhibit strong right-skewness due to high-priced properties concentrated in specific metropolitan regions. This observation motivates the use of logarithmic transformation and distributional diagnostics, which will be discussed in Section 3.2.

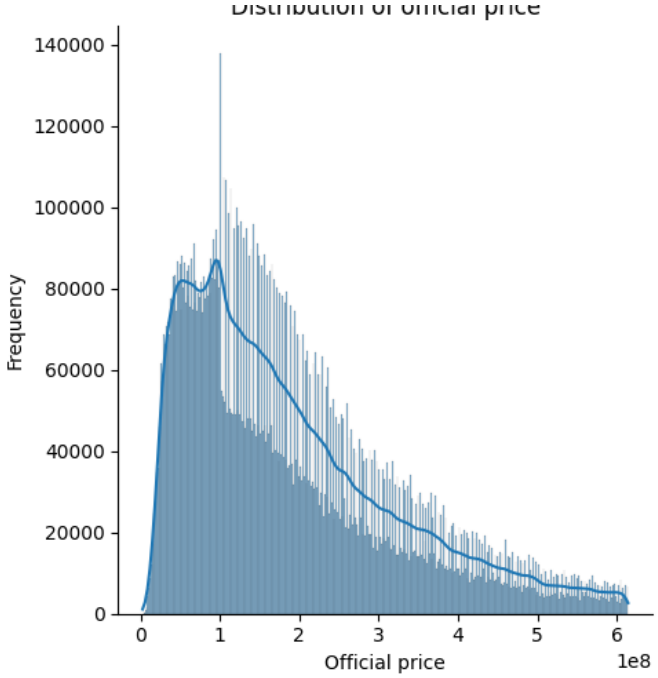

Figure 1: Distribution of official price (공시가격) after outlier removal

3.2 Application of Statistics [6]

3.2.1 Distributional Diagnostics: Skewness and Kurtosis

To assess the statistical characteristics of housing price distributions between the Seoul Metropolitan Area and other non-metropolitan regions, we compute:

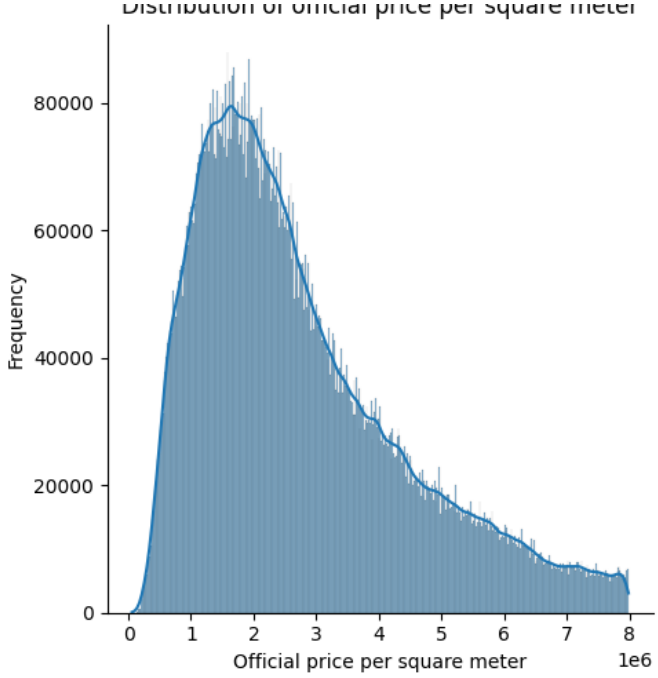

Figure 2: Distribution of official price per square meter (공시가격_㎡) after outlier removal

Skewness Measures the asymmetry of a distribution. A positively skewed distribution (In general, Skewness ¿ 0) indicates a long right tail, often caused by extremely high-priced properties in metropolitan clusters (e.g., Gangnam). Skewness helps evaluate whether the mean can serve as a reliable measure of central tendency, and whether parametric tests such as the t-test are appropriate without transformation.

$$\text{Skewness} = \frac{1}{n}\sum_{i=1}^{n}(\frac{X_i - \bar{X}}{s})^3$$

Kurtosis Measures the peakedness and tail heaviness of the distribution. High kurtosis (In general, Kurtosis ¿ 3) suggests the presence of extreme values and a higher likelihood of outliers. This is especially relevant in housing data where a small number of luxury properties can significantly distort distributional shape and variance.

$$\text{Kurtosis} = \frac{1}{n}\sum_{i=1}^{n}(\frac{X_i - \bar{X}}{s})^4$$

X_i is the unit housing price or unit housing price per square meter, $\bar{X}$ is the sample mean, and s is the sample standard deviation.

If we look at the overall distribution, we see that there is a very thin and long outiler, and if we remove the outiler, we see that the overall quantity of houses itself is skewed to the left. Skewness and kurtosis thus serve as diagnostic tools to determine whether data transformations (e.g., log transformation) are necessary, and whether parametric tests are valid or should be replaced by non-parametric alternatives. These diagnostic statistics are therefore not just descriptive, but instrumental in guiding preprocessing decisions and selecting appropriate statistical tests for hypothesis evaluation in the next stages. Additionally, by comparing these values across Seoul and non-metropolitan regions, we can quantify structural inequality and detect whether the housing market in metropolitan areas is not only more expensive but also more volatile and unequal in its distribution.

3.2.2 Logarithmic Transformation and Normality Testing

Apply a logarithmic transformation to normalize data that is heavily skewed to one side, has large units, and to stabilize the variance:

$$X_i^{\log} = \log(X_i + 1)$$

This transformation allows you to define values close to zero while compressing extreme values. After the transformation, re-evaluate the skewness and kurtosis and apply the Shapiro-Wilk test, Kolmogorov-Smirnov (K-S) test to ensure that the transformed variables satisfy the normality assumptions required for subsequent parametric tests.

3.2.3 Hypothesis Testing: Regional Comparison

After transforming the data appropriately, we perform two statistical hypothesis tests to evaluate home price disparities by region:

Two-sample t-test Used to evaluate whether the mean housing price per square meter is significantly higher in the Seoul Metropolitan Area than in non-metropolitan regions.

- Null hypothesis $H_0 : \mu_{\text{Seoul}} = \mu_{\text{Local}}$
- - Alternative hypothesis $H_1 : \mu_{\text{Seoul}} > \mu_{\text{Local}}$

This one-tailed test provides statistical evidence for urban concentration of housing costs if $p <$ Significance level.

Kolmogorov-Smirnov (K-S) test Used to compare the entire distribution of housing prices across the two regions. This non-parametric test evaluates differences in shape, variance, and tail behavior.

- Null hypothesis H_0 : The two samples come from the same distribution.
- Rejection implies structural disparity in housing price distribution beyond average differences.

References

[1] "Resident population by city, county, and district," tech. rep., Ministry of the Interior and Safety, 2025.

[2] h. b. m. Jun and j. h. Koo, "A study on the causality of housing price and trading by cyclical variation: Focused

on housing market of seoul," *Korea Real Estate Society*, vol. 71, p. 211–226, Mar 2024.

[3] M. Roohani Qadikolaei, Y. Hatami, S. Nikmard Namin, and A. Soltani, "Spatial heterogeneity in housing price-transaction ratios: A historical analysis of tehran," *International Journal of Housing Markets and Analysis*, Nov 2024.

[4] H. J. Kwon and M. H. Choi, "Comparative analysis of the effects of rental housing supply provider and method on the housing market price stabilization," *Korea Real Estate Institute*, vol. 34, p. 7–22, Mar 2024.

[5] D. Lim, C. Lee, and S. Kim, "Analysis of housing mortgage prepayment in korea: Focus on shared equity/appreciation mortgage and didimdol loans," *Korean Association for Housing Policy Studies*, vol. 32, p. 5–28, Nov 2024.

[6] D. C. Montgomery and G. C. Runger, *Applied Statistics and probability for Engineers*. Wiley, 2018.