The Comfort Premium: Factors Shaping House Prices

Using R to Explore the Impact of Heating Quality and Air Conditioning on Sale Prices

Introduction

Understanding the drivers of house prices is essential for real estate professionals, policymakers, and homebuyers. Features like **heating quality** and **central air conditioning** play an increasingly significant role as buyers prioritize comfort and energy efficiency. Previous research **(Hahn et al., 2018; Cağlayan & Arikan, 2011; Anselin & Lozano-Gracia, 2008)** has shown that **energy-efficient systems** and **internal features** significantly influence property values, emphasizing the relevance of these aspects.

This study investigates: **How do heating quality and central air conditioning affect house prices, controlling for overall quality, living area, and basement area?**
The goal is to develop an interpretive model that quantifies the influence of these features on house sale prices. By focusing on **interpretability**, **the findings will offer actionable insights for real estate stakeholders**, enabling them to better understand the value added by heating quality and central air conditioning in property pricing.

Methodology

Using **R**, we developed a **Multiple Linear Regression** model to analyze predictors such as **Heating Quality, Air Conditioning, Overall Quality, Basement Area, and Living Area**. The process included:

- **Testing Assumptions**: Checked linearity (scatterplots), multicollinearity **(VIF)**, and residual normality/independence.
- **Evaluating Predictors**: Used **ANOVA, t-tests, and adjusted R** to assess predictor significance and model fit.
- **Addressing Violations**: Applied **Box-Cox transformations** for normality/variance issues and reviewed influential points flagged by **Cook's Distance** and leverage.

A simplified version of this process is shown in **Figure 1** below.

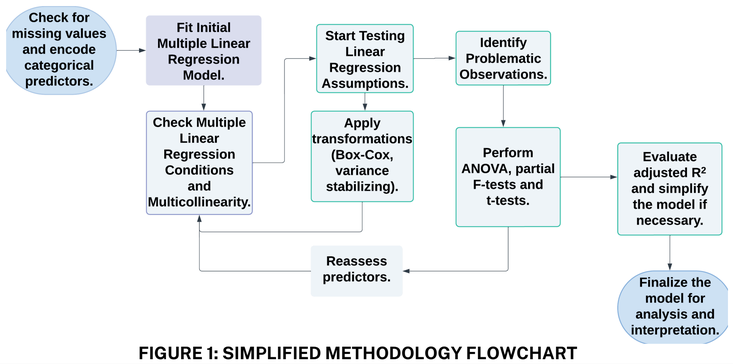

FIGURE 1: SIMPLIFIED METHODOLOGY FLOWCHART

Data Collection

The data for this study was sourced from the **Ames Housing Dataset** (De Cock, 2011), a publicly available dataset hosted on **Kaggle** (Prevek, 2024). It contains detailed property records of 2,930 house sales in Ames, Iowa (2006–2010), collected by the city's Assessor's Office.

Why This Data?

- **Reliable Source**: professional property assessments, ensuring accuracy and consistency in measurements.
- **Population Fit**: Focuses on housing features relevant to real estate markets, allowing generalization to similar populations.
- **Comprehensive Variables**: Includes meaningful predictors like heating quality and central air conditioning, suitable for addressing our research question.

Analysis

Our initial model, fitted without transformations, violated key linear regression assumptions such as normality and linearity, as shown in **Figure 2** on the right. To address this, we applied a **power transformation** to the response variable, SalePrice, based on a **Box-Cox analysis**.

The Box-Cox method suggested a **λ** value of 0.141, leading to a simpler transformation of SalePrice to $\textbf{SalePrice}^{1/4}$. Figure 3 on the right shows the improved diagnostic plots for the transformed model, confirming **normality, linearity, and homoscedasticity**.

Key results are summarized in Table 1 below:

- **Poor heating quality reduces $\text{SalePrice}^{1/4}$ by 1.465 units compared to excellent quality**, holding other variables constant.
- **Central air conditioning increases $\text{SalePrice}^{1/4}$ by 0.865 units** compared to houses without it.

Coefficient	Variable name	Model Coefficient Estimate
β_{13}	Poor Heating Quality	-1.465
β_{14}	Central Air Conditioning	0.8647

TABLE 1: MODEL COEFFICIENTS ESTIMATES

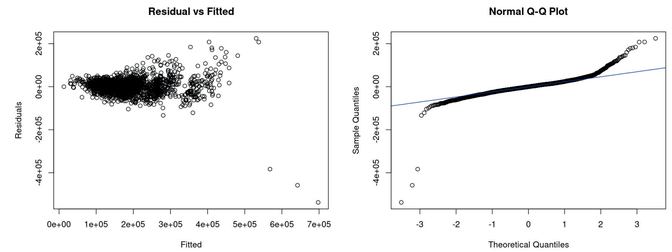

Figure 2: Pre-Transformed Model selected Residual Plots

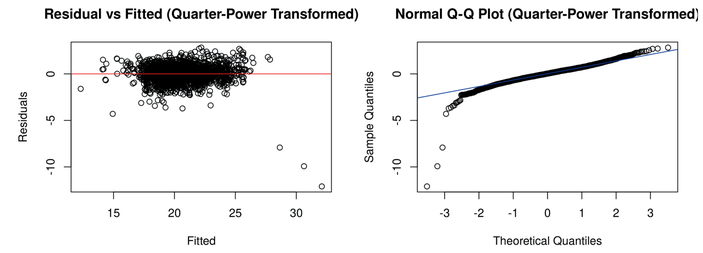

Figure 3: Transformed Model selected Residual Plots

ANOVA results indicate the overall model is significant, while **individual t-tests show that predictors are statistically significant**. Additionally, the **adjusted R Square value of 0.8263** demonstrates a strong balance between model fit and simplicity, supporting the validity of the transformed model. These results are in **Table 2** below:

Diagnostic Type	Value
ANOVA test	2.2e-16
Adjusted R Square	0.8263
t-test for Poor Heating Quality	0.008805
t-test for Air Conditioning	2e-16

Table 2: Selected Model Diagnostics

THESE RESULTS CONFIRM THAT EXCELLENT HEATING QUALITY AND CENTRAL AIR CONDITIONING POSITIVELY INFLUENCE HOUSE PRICES.

Limitations

Our study has limitations. The Ames Dataset, though robust, **is limited to Ames, Iowa**, reducing application to broader markets. **Minor violations of regression assumptions** persist despite refinements. While transformations were applied, residual patterns suggest some non-linear relationships remain. Future studies could explore advanced models like **machine learning** to better capture complex interactions and improve accuracy.

BIBLIOGRAPHY

1. Anselin, L., & Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics, 34(1), 5–34. https://doi.org/10.1007/ s00181-007-0152-3
2. Çağlayan, E., & Arıkan, E. (2011). Determinants of house prices in Istanbul: A quantile regression approach. Quality & Quantity, 45(2), 305–317. https://doi.org/10.1007/ s11135-009-9296-x
3. De Cock, D. (2011). Ames, Iowa: Alternative to the Boston housing data as an end-of-semester regression project. Journal of Statistics Education, 19(3). https://jse.amstat.org/v19n3/decock.pdf
4. Hahn, J., Hirsch, J., & Bienert, S. (2018). Does "clean" pay off? Housing markets and their perception of heating technology. Property Management, 36(5), 575–596. https://doi.org/ 10.1108/PM-08-2017-0051
5. Prevek. (2024). Ames Housing Dataset. Retrieved from https://www.kaggle.com/datasets/ prevek18/ames-housing-dataset