Housing in Ames*

Yanzun Jiang, Siyuan Lu, Yi Tang

November 24, 2024

Table of contents

*Code and data supporting this proposal is available at: https://github.com/Stary54264/Housing-in-Ames

1 Introduction

The research question we aim to answer is: what are the factors that influence house prices in Ames from 2006 to 2010? We would set up a linear regression model to answer this with house prices as the outcome (response). Factors that might affect the outcome (predictors) include area, quality, year of construction, facilities, etc.

By setting up the model, we can identify the factors that influence house prices, since linear regression allows us to quantify the relationship between predictors and responses, making it easier to interpret the impact of each factor alone on house prices. Our primary goal is understanding the factors that influence historical house prices, so our focus would be on description rather than prediction.

We found several peer-reviewed articles that focus on similar problems with this paper. "Influencing Factors Analysis of House Prices Based on Multiple Linear Regression" concludes that housing prices are negatively correlated with completion costs, land acquisition prices, residents' disposable income, and population density (Wang (2013)). This article provides some characteristics other than what we use that can also influence house price in national scope.

In "Dynamic Relationships Between Commodity Prices and Local Housing Market", the researchers examines the significant nonlinear relationship between agricultural commodity prices and the housing prices (Liang, Fan, and Hu (2021)). Another research, "Non-Linear Relationships Between House Size and Price", clarifies the non-linear relationship between size and price (Feng et al. (2021)). These two researches explain the non-linear relationship, between house price and other factors, providing more insights into ways that factors might affect house prices.

2 Method

3 Results

4 Conclusion and Limitations

A Appendix

A.1 Detailed Process of Checking Assumptions

A.2 Ethics Discussion

Our data is collected from Ames City Assessor's Office (De Cock (2011)), then we cleaned the data to only keep some necessary key factors that is highly relavant to house prices. Raw and processed versions of the data from De Cock is published on Journal of Statistics Education in 2011. The cleaned data we are using includes some detailed information about housing characteristics, but does not contain personal identifiers.

The Ames housing dataset has been used widely, especially in the context of academic projects and machine learning competitions. It is often considered a modern alternative to the Boston Housing dataset. The dataset is well-vetted and trusted by the data science community for its comprehensiveness and relevance.

The use of automated selection tools in academic research brings both opportunities and ethical challenges. Automated tools can significantly speed up research processes, but their use must be transparent, including acknowledging the specific tools and algorithms employed, as well as their limitations.

Ensuring fairness and avoiding skewed results require careful selection of training data and ongoing monitoring to detect and mitigate biases presented in AI training data. While automated tools can enhance productivity, they should not replace human judgment. By addressing these ethical considerations, we can leverage the benefits of automated selection tools while maintaining the integrity and fairness during our research practices.

A.3 Contributions

Group contribution is available at https://github.com/Stary54264/Housing-in-Ames/graphs/contributors. Below is a more specific version of group contribution.

- Yanzun Jiang: Organized discussions and meetings; assigned tasks to group members; set up Github workspace for collaborating; refined the model by model selection tools; checked multicollinearity of the models; made the reference list; revised group member's work; combined group member's work together.
- Siyuan Lu: Set research question; searched and read peer-reviewed articles; introduced the project; checked data ethics.
- Yi Tang: Built linear regression model; checked conditions for performing linear regression; checked extra conditions for performing multiple linear regression; showed the results of the linear regression model.

References

De Cock, Dean. 2011. "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project." *Journal of Statistics Education* 19 (3).

Feng, S.-T., C.-W. Peng, C.-H. Yang, and P.-W. Chen. 2021. "NON-LINEAR RELATIONSHIPS BETWEEN HOUSE SIZE AND PRICE." *International Journal of Strategic Property Management* 25 (3): 240–53. https://doi.org/10.3846/ijspm.2021.14607.

Liang, J., Q. Fan, and Y. Hu. 2021. "Dynamic Relationships Between Commodity Prices and Local Housing Market: Evidence for Linear and Nonlinear Causality." *Applied Economics* 53 (15): 1743–55. https://doi.org/10.1080/00036846.2020.1845295.

Wang, Cheng. 2013. "Influencing Factors Analysis of House Prices Based on Multiple Linear Regression." *Applied Mechanics and Materials* 415: 722–25.