Analysis of Boston Housing Data and its Ethical Implications

Aviral Bhardwaj

January 22, 2024

Discussion

> I think we should be suspicious when we find ourselves attracted to data—very, very thin and weak data—that seem to justify beliefs that have held great currency in lots of societies throughout history, in a way that is conducive to the oppression of large segments of the population.
>
> Amia Srinivasan (Cowen 2021)

The following quote by Amia Srinivasan is very relevant to the analysis of the `Boston Housing Data` as collected by Harrison and Rubinfeld in 1979 (Harrison and Rubinfeld 1978). The data is thin and weak because it only contains 506 observations, and it is used to justify the oppression of large segments of the population.

The dataset, designed by Harrison and Rubenfield to study the relationship between housing prices, has inherent assumptions that make it susceptible to being considered thin and weak data. Notably, during the data collection process, a critical assumption was made, assuming that racial self-segregation had a positive impact on house prices. This assumption encoded systemic racism as a factor influencing housing prices, suggesting a correlation between people of color residing in certain areas and lower housing income, thereby implying reduced affordability for them to purchase homes.

This racial bias within the dataset perpetuates the notion that individuals from certain racial backgrounds are associated with lower socioeconomic status, contributing to the reinforcement of stereotypes "justify beliefs that have held great currency in lots of societies" (Cowen 2021). Specifically, it implies that people of color are more likely to be classified as lower-class citizens with limited housing income, potentially leading to discriminatory practices in real estate.

Moreover, it's important to acknowledge that the data was collected in 1978, a period marked by high inflation and interest rates. This temporal context renders the dataset less representative of the contemporary housing market. Economic conditions, societal attitudes, and

housing dynamics have likely evolved over time, making it necessary to interpret the findings with caution when applying them to present-day scenarios.

Additionally, the dataset reflects biases related to racial perceptions, indicating that people of color were considered undesirable neighbors by white individuals. This aspect of the data not only contributes to the perpetuation of discriminatory beliefs but also provides a distorted justification for the oppression of large segments of the population based on race.

In summary, the Boston Housing Dataset, designed to explore property price determinants, lacks representation of the contemporary housing market due to its inherent biases and the influence of historical assumptions. Notably, racial biases embedded in the dataset, assuming positive impacts of racial self-segregation on house prices, underscore the risk of perpetuating discriminatory beliefs. Moreover, the dataset's 1978 context, marked by high inflation and interest rates, limits its applicability to today's housing dynamics.

Additionally, the data should be utilized to comprehend the historical context of the housing market and to identify historical biases, rather than being employed for making predictions about the contemporary housing market.

References

(R Core Team 2023)

Cowen, Tyler. 2021. "Episode 132: Amia Srinivasan on Utopian Feminism." *Conversations with Tyler*, September. https://conversationswithtyler.com/episodes/amia-srinivasan/.

Harrison, David, and Daniel L Rubinfeld. 1978. "Hedonic Prices and the Demand for Clean Air." *Journal of Environmental Economics and Management* 5 (1): 81–102.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.