

IMPACTO SOCIOECONÔMICO E PROXIMIDADE AO RIO NOS VALORES IMOBILIÁRIOS: APLICAÇÃO DE MODELO NÃO LINEAR

Yasmin Rissato Pichinini (PIBIC/CNPq/FA/UEM), Vanderley Janeiro (Orientador, e-mail: vjaneiro@uem.br)

Universidade Estadual de Maringá, Centro de Ciências Exatas, Maringá, PR.

Área e subárea do conhecimento: Probabilidade e Estatística/ Probabilidade e Estatística Aplicadas.

Palavras-chave: Decaimento Exponencial; Modelagem; Análise Socioeconômica.

RESUMO

Este estudo aplica um modelo de decaimento exponencial para analisar com o status socioeconômico e a proximidade ao Rio Charles afetam o valor das propriedades residenciais na Área Estatística Metropolitana de Boston, com base em dados do Censo dos EUA de 1970. O modelo não linear revela que a proximidade ao rio aumenta o valor mediano das casas, enquanto a taxa de decréscimo no valor das casas com o aumento da porcentagem de indivíduos de baixo status socioeconômico é similar independentemente da localização. Os resultados destacam a eficácia dos modelos não lineares em capturar relações complexas entre variáveis socioeconômicas e preços imobiliários.

INTRODUÇÃO

Modelos de regressão polinomiais são amplamente utilizados para ajustar dados, oferecendo maior precisão à medida que se aumenta a ordem do polinômio. No entanto, esses modelos são puramente empíricos e não consideram o mecanismo subjacente dos dados. Em contraste, modelos não lineares, como o de decaimento exponencial, são baseados em teorias mecanicistas, proporcionando interpretações físicas claras para seus parâmetros e frequentemente requerem menos parâmetros do que os modelos polinomiais. Esses modelos não lineares também oferecem previsões mais robustas fora do intervalo observado dos dados. Neste estudo, aplicamos um modelo de decaimento exponencial para explorar a influência do status socioeconômico e da proximidade ao Rio Charles sobre o valor mediano das propriedades residenciais em Boston, usando o conjunto de dados clássico de

Harrison e Rubinfeld (1978). A escolha do modelo não linear busca capturar de forma mais precisa a relação complexa entre essas variáveis.

MATERIAIS E MÉTODOS

Material

Para ilustrar a teoria estudada, foi utilizado o conjunto de dados clássico sobre valores habitacionais de Boston, descrito por Harrison e Rubinfeld (1978), com base no Censo dos EUA de 1970. Cada observação representa um setor do Censo na Área Estatística Metropolitana Padrão de Boston, totalizando 506 observações limitadas a setores com domicílios ocupados por proprietários. O objetivo é prever o valor mediano das casas (em milhares de dólares) ocupadas por proprietários de status socioeconômico mais baixo. As variáveis utilizadas incluem: “medv” (valor mediano das casas ocupadas pelos proprietários em US\$ 1.000), “chas” (variável dummy que indica se o setor limita o Rio Charles, sendo 1 se sim e 0 caso contrário), e “lstat” (percentual da população de status socioeconômico mais baixo). No artigo original, Harrison e Rubinfeld (1978) definem o status mais baixo como a proporção da população que possui metade da proporção de adultos sem educação secundária e metade da proporção de trabalhadores do sexo masculino classificados como operários.

Método

Visando atender ao objetivo do trabalho, é considerado o modelo não linear de decaimento exponencial (1), capaz de descrever a curvatura dos dados apresentados na Figura 1a). Além da curvatura o modelo contempla uma assíntota inferior, ou seja, mesmo aumentando drasticamente a proporção de indivíduos de baixo nível socioeconômico o valor das residências permanecerá estabilizado. O modelo será o descrito da seguinte forma por (Bates e Watts, 1988)

$$y = \theta_1 + (\theta_2 - \theta_1) \cdot \exp(-\theta_3 x) + \varepsilon \quad (1)$$

no qual θ_1 que representa o limite inferior da função, ou seja, o valor para o qual a função se aproxima à medida que x tende ao infinito; θ_2 o valor inicial quando $x = 0$; e θ_3 a taxa de decaimento, que determina a rapidez do decaimento da função. Um valor menor de θ_3 indica um decaimento mais rápido. E ε o termo do erro aleatório seguindo distribuição normal de média zero e variância constante.

O ajuste do modelo será com a função do pacote do ambiente estatístico, R Core Team (2024). Será considerado o algoritmo padrão é um algoritmo de Gauss-Newton.

RESULTADOS E DISCUSSÃO

Ao ajustar um modelo não linear há a necessidade de atribuir valores iniciais aos parâmetros, o que às vezes torna-se uma complicação extra. Dada a funcionalidade dos parâmetros do modelo (1), os parâmetros θ_1 e θ_2 terão os valores extraídos diretamente dos valores observados dispersos na Figura 1a), isto é, $\theta_1^{(0)} = 5$ e $\theta_2^{(0)} = 50$. O valor inicial $\theta_3^{(0)} = 0.1704$ foi obtido com o auxílio da função,

$$y = \theta_1^{(0)} + (\theta_2^{(0)} - \theta_1^{(0)}) \cdot \exp(-\theta_3^{(0)} x) \rightarrow \log\left(\frac{y - \theta_1^{(0)}}{\theta_2^{(0)} - \theta_1^{(0)}}\right) = (-\theta_3^{(0)} x) \rightarrow y^* = \theta_3^* x.$$

Para auxiliar na convergência do método de estimação, um modelo geral, sem considerar o efeito do limite do setor com o Rio foi ajustado e obteve-se estimativas $\hat{\theta}_1 = 12.3788$, $\hat{\theta}_2 = 53.8746$ e $\hat{\theta}_3 = 0.1413$. Essas estimativas passaram a ser valores iniciais para o ajuste do modelo contendo a variável categórica *Chas*, com a finalidade de discriminar possíveis variações na distribuição do valor das residências pela localização do setor. A partir do modelo ajustado detectou-se que valores iguais a 50 (\$1000) na linha de corte do valor das residências causavam distúrbios nos resíduos. Esses foram removidos passando o banco de dados a contar com 488 observações, e resíduo bem comportado como apresentado na Figura 1b).

Os modelos ajustados para estimar o valor das casas em setores com limite no Rio Charles e sem limite com o rio são, respectivamente:

$$\hat{y}_{Rio} = 10.7686 + 34.8031 \cdot \exp(-0.1106x) \text{ e } \hat{y}_{NRio} = 17.8862 + 35.18 \cdot \exp(-0.1771x)$$

De acordo com o teste de Bonferroni, comparando os parâmetros do modelo em relação a variável, o parâmetro θ_1 apresentou diferença a 5% de significância (valor-p=0,0049). Porém os parâmetros θ_2 e θ_3 não apresentaram diferença significativas, com valor-p iguais a 0,4501 e 0,3405 respectivamente.

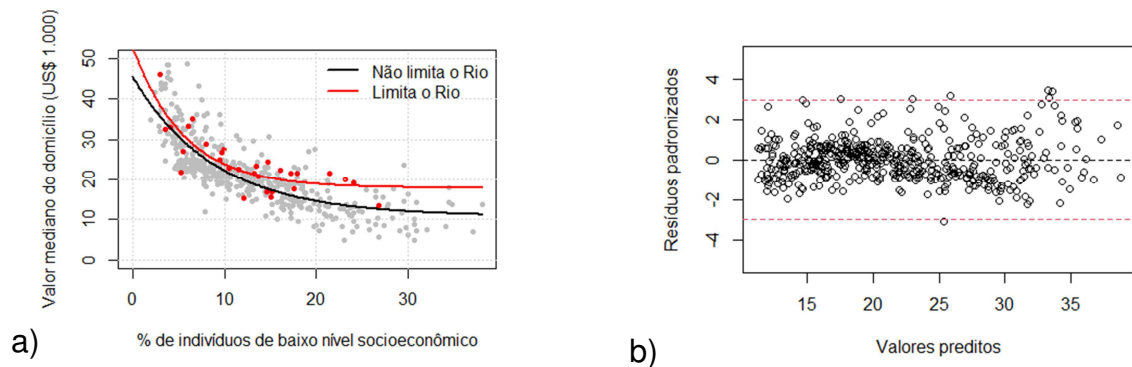


Figura 1 – No gráfico a) dispersão dos valores observados com modelo ajustado; no gráfico b) Resíduos padronizados do modelo ajustado.

CONCLUSÕES

O modelo de decaimento exponencial foi eficaz para analisar a relação entre o valor das casas e a proporção de indivíduos de baixo status socioeconômico na Área Metropolitana de Boston. Os resultados mostram que, embora a proximidade ao Rio Charles eleve o valor das propriedades, a taxa de decréscimo dos preços com o aumento da população de baixo status é semelhante, independentemente da localização. O modelo capturou bem a relação entre fatores socioeconômicos e geográficos nos preços imobiliários.

AGRADECIMENTOS

Agradeço ao Programa PIBIC/CNPq/FA/UEM pelo aprendizado proporcionado pela realização desta pesquisa e pelo apoio financeiro recebido. E ao professor Vanderly Janeiro pela oportunidade de trabalhar e contribuir com esse estudo.

REFERÊNCIAS

BATES, D. M.; WATTS, D. G. **Nonlinear Regression Analysis and Its Applications** 1988 Disponível em: <https://api.semanticscholar.org/CorpusID:62726730>

HARRISON, D.; RUBINFELD, D. L. [Hedonic housing prices and the demand for clean air](#). **Journal of Environmental Economics and Management**, v. 5, n. 1, p. 81–102, 1978.

33° Encontro Anual de Iniciação Científica
13° Encontro Anual de Iniciação Científica Júnior



10 e 11 de Outubro de 2024

R CORE TEAM. [R: A Language and Environment for Statistical Computing.](#)
Vienna, Austria: R Foundation for Statistical Computing, 2024.

