MODELOS MISTOS PARA CONFECÇÃO DE PLANTAS DE VALORES GENÉRICOS: UMA NOVA ABORDAGEM

Mixed models for development of property value maps: a new approach

Luiz Fernando Palin Droubi
Secretaria de Coordenação e Governança do Patrimônio da União
Superintendência do Patrimônio da União em Santa Catarina
lfpdroubi@gmail.com

Vanessa Espíndola
Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial
nessa.espindola18@gmail.com

Norberto Hochheim
Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial
geap.ufsc@gmail.com

Ana Maria Bencciveni Franzoni
Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial
ana.franzoni@ufsc.br

Resumo:
Este trabalho faz uma breve introdução aos modelos mistos, também conhecidos como modelos hierárquicos, com foco na aplicação para confecção de Planta de Valores Genéricos. A modelagem de efeitos mistos, se bem aplicada, possui uma série de vantagens em relação aos outros métodos usualmente utilizados para tal finalidade. Além da aplicabilidade para confecção de PVGs, a modelagem mista também pode fornecer bons *insights* sobre os efeitos da localidade na formação de preços de imóveis, através da utilização de variáveis que modelem as características dos diversos agrupamentos que compõem a amostra. Isto abre um campo muito interessante para a análise dos possíveis efeitos da realização de investimentos em infraestrutura e/ou intervenções urbanas na valorização dos imóveis. A modelagem mista ainda pode ser eficaz como ferramenta de auxílio ao planejamento urbano e regional, ajudando a orientar a melhor alocação dos recursos disponíveis para os investimentos. Uma aplicação no município de Imbituba/SC, Brasil, é mostrada para ilustrar a aplicação da modelagem.
Palavras-chave: Planta de Valores Genéricos. Modelos mistos. Modelagem hierárquica.

Abstract
This work gives a brief introduction to mixed models, also known as hierarchical models, with a focus on the application for Property Values Mapping. Mixed effects modeling, if well applied, has a number of advantages over other methods usually used for this purpose. In addition to the applicability for making Property Value Maps, mixed modeling can also provide good insights into the effects of locality on the formation of property prices, through the use of variables that model the characteristics of the different groupings that make up the sample. This opens a very interesting field for the analysis of the possible effects of making investments in infrastructure and / or urban interventions in the valuation of properties. Mixed modeling can still be effective as an aid tool for urban and regional planning, helping to guide the best allocation of available resources for investments. An application in the municipality of Imbituba/SC, Brazil, is shown to illustrate the application of the modeling.
Keywords: Property Value Maps. Mixed Models. Hierarchical Modelling.

1. INTRODUÇÃO

Um dos componentes de maior importância do Cadastro Territorial Multifinalitário é o valor das propriedades. Apesar da recente grande evolução e difusão de novos métodos estatísticos e ferramentas para a determinação do valor das propriedades, que permitem previsões de valores a cada dia mais precisos, estes métodos muitas vezes perdem o poder de explicar o mercado, como ocorre em alguns métodos de inteligência artificial amplamente utilizados na comunidade de *Machine Learning,* como as redes neurais artificiais, por exemplo.

Ocorre que, embora a acurácia na previsão de valores seja importante nas avaliações de imóveis, na avaliação em massa para fins tributários muito mais importante é uma boa relação entre os valores previstos em todos a área objeto de estudo.

Um grande problema com alguns métodos de inteligência artificial, como os modelos de redes neurais artificiais, é o sobreajuste (*overfitting*) dos modelos aos dados da amostra, o que pode levar um modelo que se ajusta muito bem aos dados amostrais a fazer previsões não tão boas para os dados fora da amostra. É verdade que para contornar este problema métodos de validação cruzada foram desenvolvidos, porém, deve-se ter em conta que a validação cruzada não é um método estatisticamente consistente[1], a não ser em situações especiais[2] (ver SHAO (1993); RAO e FUNG (2008); MATLOFF 2017, p. 348).

A confecção de Plantas de Valores Genéricos (PVGs) é uma atividade complexa para a qual não há no Brasil, ainda, normatização adequada. A NBR 14653-2 (ABNT, 2011) trata da avaliação de bens imóveis urbanos, porém a norma é muito mais voltada à avaliação de um imóvel em particular do que para a aplicação à avaliação em massa de imóveis para fins tributários. Desta maneira, muitos problemas práticos enfrentados durante a elaboração de PVGs são questionáveis, como a adoção de valores genéricos para agrupamentos onde não foi realizada amostragem, extrapolados de modelos obtidos com dados de outros bairros.

Uma questão importante na elaboração de PVGs é a falta de homogeneidade do conjunto amostral. Claramente, numa cidade, os dados encontram-se agrupados ou hierarquizados e a autocorrelação espacial tende a estar presente. Diversos métodos foram desenvolvidos para levar em conta a não-homogeneidade da amostra, como a regressão espacial e a regressão geograficamente ponderada, porém é mais comum na Engenharia de Avaliações a aplicação de um modelo de efeitos fixos, agrupando os dados amostrais por bairros e criando variáveis dicotômicas em grupo (*dummies*) para representar o subconjunto a que pertence cada dado amostral. Ocorre que, para uma boa estimação por efeitos fixos, o número de dados em cada agrupamento deve ter um tamanho mínimo necessário para que a estimação nestes agrupamentos seja precisa. A NBR 14653-2 (ABNT, 2011) especifica, então, que um mínimo de 3 a 10 dados de cada agrupamento sejam utilizados para a estimação com tal abordagem.

Outro problema central na elaboração de PVGs consiste na dificuldade para a obtenção do valor da terra em separado do valor da propriedade, como no caso de áreas verticalizadas. Como é difícil a obtenção de elementos amostrais de lotes urbanos em bairros consolidados, a

---

[1] Isto é, a habilidade preditiva do modelo não converge para 1 quando o número de dados da amostra n tende a infinito.

[2] Segundo SHAO (1993), para que a validação cruzada seja um método consistente, deve-se ter $n_v / n \xrightarrow{n\to\infty} 1$, porém não é claro como encontrar n_v (tamanho da partição de testes) de forma que isso ocorra.

alternativa seria estimar o valor da terra nua à partir de imóveis edificados, separando o valor das estruturas do valor da terra. Porém, ainda que os custos de construção para fabricação de um bem imóvel sejam bem conhecidos, o que torna trivial a estimativa do valor necessário para reproduzi-lo, critérios como depreciação e fator de comercialização (vantagem da coisa feita) por vezes são subjetivos e impedem uma melhor estimação do valor da terra nua em imóveis edificados.

É comum, portanto, que na elaboração de PVGs nem todos os bairros da cidade estejam representados no conjunto amostral utilizado para a sua confecção, seja pela completa ausência de dados amostrais seja pelo problema da micronumerosidade, o que é mais comum. Estatisticamente falando, um modelo confeccionado com tal amostra possui um viés, no caso, um viés amostral (viés também pode ser causado por outros fatores, como a não inserção de uma variável explicativa importante no modelo).

Ora, um modelo construído como o descrito no último parágrafo irá efetuar melhores previsões de valores para imóveis constantes dos bairros contemplados na amostra e tenderá a subestimar ou superestimar os valores dos imóveis nos bairros não constantes da amostra, o que é conhecido como viés amostral.

O objetivo geral deste artigo é apresentar uma abordagem diferente, a saber, a aplicação de modelos hierárquicos ou mistos, em que a variância amostral é dividida em duas ou mais componentes, *i.e.* a variância total é dividida entre a variância entre os indivíduos da amostra e a variância entre os diferentes grupos de indivíduos constituintes da amostra. O objetivo específico deste trabalho é mostrar quando a aplicação deste tipo de modelo pode ser interessante na avaliação em massa de imóveis.

2. REVISÃO BIBLIOGRÁFICA

Cichulska e Cellmer (2018) elencaram três diferentes abordagens para a análise de dados com estrutura agrupada utilizando-se mínimos quadrados ordinários:

- A desagregação dos dados, ignorando o agrupamento dos dados em diferentes subconjuntos;
- A agregação dos dados em grupos, gerando diversas unidades de análise;
- A adoção de modelos de efeitos fixos, que consiste em atribuir variáveis *dummies* (variáveis dicotômicas em grupo), o que implica na obtenção de médias ponderadas para os valores dos regressores advindos de todas os grupos de observação.

Segundo Bell, Fairbrother e Jones (2019), existe uma confusão entre diversas áreas de pesquisa sobre as propriedades, capacidades e limitações dos modelos de efeitos fixos e de efeitos aleatórios. Bell, Fairbrother e Jones (2019) fazem, então, um comparativo entre os dois tipos mais famosos de modelagens para dados agrupados: os modelos de efeitos fixos e os modelos mistos, mostrando as diversas formulações deste último tipo de modelos, capazes de modelar os mais diversos efeitos.

Descreve-se a seguir as diversas formas de modelagem para depois compará-las.

2.1. Modelagem por mínimos quadrados ordinários simples

O Método dos Mínimos Quadrados Ordinários (MQO) simples assume a homogeneidade da amostra. Desta maneira, se utilizado para a análise de dados agrupados, o valor dos coeficientes dos regressores será uma média ponderada dos efeitos de cada elemento, não interpretável. Além disto, os erros-padrões serão estimados com a hipótese de independência das observações, hipótese que não se verifica, o que tem implicâncias sobre a estimação dos intervalos de confiança e nos testes de hipótese para os coeficientes.

$$y_i = \beta_0 + \beta_1^{OLS} x_i + \varepsilon_i \tag{1}$$

A equação 1 ilustra a modelagem pelo método dos mínimos quadrados ordinários simples, onde ε_i é um termo estocástico considerado independente e identicamente distribuído (i.i.d.).

2.2. Modelagem por efeitos fixos

A modelagem por efeitos fixos é aplicada na Engenharia de Avaliações, assim como em diversas outras áreas da ciência, como uma espécide de “padrão ouro” para lidar com a heterogeneidade amostral (BELL; FAIRBROTHER; JONES, 2019, p. 1057). Uma das maneiras de escrever a formulação de efeitos fixos, a mais conveniente para a comparação que se pretende, pode ser vista na equação 2:

$$y_{ij} = \beta_1 (x_{ij} - \bar{x}_i) + (\upsilon_j + \varepsilon_{ij}) \tag{2}$$

Onde υ_j é um termo discreto para cada agrupamento de dados. Os índices *i* e *j* se referem aos dois níveis de análise: *i*, neste caso, representa o nível dos indivíduos e *j* o nível dos agrupamentos. ε_{ij} é um termo de erro aleatório com distribuição supostamente normal, média zero e desvio-padrão σ^2.

Um dos problemas com este tipo de modelagem, como será visto, está no fato de que toda a variância de nível hierárquico mais alto, *i.e.* a variância entre os agrupamentos, é “absorvida” pelas variáveis *dummies*, que consomem, portanto, todos os graus de liberdade neste nível de análise (entre os grupos).

2.3. Modelagem por efeitos mistos

A modelagem por efeitos mistos considera, além dos efeitos fixos, um termo de efeitos aleatórios (BELL; FAIRBROTHER; JONES, 2019, p. 1057). Uma das maneiras de escrever a formulação de efeitos mistos pode ser vista na equação 3:

$$y_{ij} = \beta_0 + \beta_1^{RE} x_{ij} + \beta_2 z_j + (\upsilon_j + \varepsilon_{ij}) \tag{3}$$

No caso da equação 3, o termo υ_j é um termo estocástico de efeitos aleatórios para os interceptos dos agrupamentos, suposto normalmente distribuído, com média zero e variância σ_{υ}^{2}, ou seja $\upsilon_j \sim N(0,\sigma_{\upsilon}^{2})$ (BELL; FAIRBROTHER; JONES, 2019, p. 1055). A variável x_{ij} é uma variável de nível 1, variante entre os indivíduos. Opcionalmente ainda é possível a introdução de variáveis de nível mais alto, como a variável z_j, invariante dentro dos agrupamentos. A diferença de nível entre as variáveis será melhor explicada no item 2.4.

Em suma, para Bell, Fairbrother e Jones (2019, p. 1061), a grande diferença entre a formulação de efeitos fixos e a formulação de efeitos mistos está na maneira como as modelagens tratam os grupamentos de dados, se de maneira discreta (efeitos fixos) ou de maneira aleatória (efeitos aleatórios). Enquanto na modelagem de efeitos fixos são adicionadas variáveis *dummies* discretas para a modelagem da variância entre os diferentes agrupamentos, consumindo assim todos os graus de liberdade *de nível hierárquico mais alto*, na modelagem de efeitos aleatórios é considerado que a diferença entre os dados de diferentes agrupamentos pode ser modelada por uma variável aleatória normal. Desta forma, como apenas um grau de liberdade é utilizado para modelar a variância *entre* os agrupamentos, é possível (e desejável) que esta variância seja modelada, *i.e.* explicada, através da utilização de variáveis de nível hierárquico mais alto.

Segundo os autores citados, ainda, esta visão não é unânime: na econometria, por exemplo, é considerado que a diferença fundamental entre as modelagens de efeitos fixos e efeitos mistos está de fato na hipótese considerada pelos modelos mistos de que não há correlação dos efeitos aleatórios (representados por υ_j) e os regressores (x_{ij}), o que é permitido na modelagem de efeitos fixos (BELL; FAIRBROTHER; JONES, 2019, p. 1060).

Desta forma, deve ser observado que a hipótese de modelar os diversos agrupamentos como uma variável aleatória é razoável apenas quando o número de agrupamentos for grande o suficiente de maneira que a hipótese da normalidade seja verificável. Para poucos agrupamentos, a modelagem por efeitos fixos é mais adequada.

2.4. Modelagem hierárquica

É comum que se encontre na literatura o termo modelagem hierárquica, ou ainda modelagem multinível, que são casos particulares da modelagem de efeitos mistos, onde os dados encontram-se aninhados (BATES, 2015, p. 7). Na modelagem hierárquica duas ou mais equações são apresentadas em separado, uma para cada nível hierárquico, do nível mais baixo até o nível hierárquico mais alto. Na prática, porém, a estimação é feita através de uma única equação, de maneira iterativa (BELL; JONES, 2015, p. 136; JONES; BULLEN, 1994, p. 258). A equação de estimação é formada pela substituição do termo β_{0j} na equação 4 pelo segundo termo da equação 5, de maneira que a equação final será da forma da equação 3 anteriormente apresentada.

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \varepsilon_{ij} \tag{4}$$

$$\beta_{0j} = \beta_0 + \beta_2 z_j + \upsilon_j \tag{5}$$

É comum na literatura se referir ao grupo hierárquico mais baixo (nível 1) como o nível dos indivíduos, o que no caso da Engenharia de Avaliações corresponde à equação composta pelas variáveis hedônicas dos imóveis.

Na Engenharia de Avaliações a hierarquização dos dados se dá em nível geográfico. Desta maneira, pode-se pensar em vários níveis hierárquicos, como bairros (nível 2), macrozonas urbanas (nível 3), cidades (nível 4), regiões (nível 5), *etc*. Desta maneira, as equações de mais altos níveis hierárquicos podem apresentar variáveis que representem estes níveis hierárquicos. Por exemplo, no nível 2 de modelagem, podem ser acrescentadas variáveis qualitativas (qualidade da infraestrutura, por exemplo) ou quantitativas (áreas públicas ou áreas verdes, por exemplo) que representem características dos bairros (e não dos imóveis) que possam contribuir na formação de valor final dos imóveis. É interessante notar, portanto, que diferentemente do que ocorre na modelagem por efeitos fixos, na modelagem hierárquica ou mista o que está a diferenciar os diversos agrupamentos pode ser explicado através destas variáveis de níveis hierárquicos mais altos. Na modelagem por efeitos fixos isto não ocorre, pois as variáveis *dummies* apenas quantificam a diferença média de valor entre os diferentes agrupamentos, sem no entanto explicá-las.

2.5. Coeficientes Aleatórios

Segundo Bell e Jones (2015, p. 133), os modelos mistos podem ser facilmente extendidos, em diversos aspectos. Uma extensão importante ao modelo misto padrão apresentado na equação 3, conhecido como modelo de interceptos aleatórios, é a introdução de outro termo aleatório ($\gamma_j \sim \mathrm{N}(0,\sigma_\gamma^2)$) associado aos coeficientes das variáveis hedônicas, formando assim um modelo de inclinações e interceptos aleatórios.

$$y_{ij} = (\beta_0 + \upsilon_j) + (\beta_1 + \gamma_j)x_{1ij} + \varepsilon_{ij} \tag{6}$$

Para a confecção de uma modelagem análoga a este tipo de modelo misto em um modelo de efeitos fixos seria necessária a introdução de interações entre as variáveis *dummies* e as variáveis hedônicas, prejudicando ainda mais a estimação dos coeficientes.

Já na modelagem de efeitos mistos, apenas mais um grau de liberdade é consumido para estimar a variância da variavel aleatória γ_j.

2.6. Vantagens e desvantagens conhecidas dos modelos de efeito mistos

De acordo com Bell, Fairbrother e Jones (2019, p. 1060), a verificação da necessidade de adição de efeitos aleatórios a um modelo MQO simples, ou seja, a verificação se a amostra pode ser considerada homogênea ou agrupada, pode ser feita através do ajuste de modelos com e sem efeitos aleatórios de níveis mais altos, comparando-os depois de acordo com algum critério estatístico para comparação de modelos, como o critério de Akaike (AIC) ou de Bayes (BIC).

Um problema comum na aplicação de modelos de efeitos fixos está na estimação das variáveis *dummies* para agrupamentos com pequenos números de dados amostrais, especialmente no caso da modelagem de coeficientes separados para cada agrupamento. Por este motivo a NBR 14.653-02 (ABNT, 2011) definiu um critério de micronumerosidade em cada grupo para variáveis *dummies*. De acordo com o Anexo A da referida norma, o número mínimo de dados de cada característica, variando de 3 a 10, a depender do tamanho da amostra, devem ser efetivamente utilizados. De acordo com Bell et al. 2018 (*apud* BELL; FAIRBROTHER; JONES, 2019, p. 1061), um modelo de efeitos aleatórios fornece estimativas mais confiáveis nestes casos.

Outro problema comum nos modelos de efeitos fixos consiste no consumo de muitos graus de liberdade para estimação das *dummies.* De acordo com Tukey (*apud* Matloff, 2017, p.344) e outro trabalho posterior de Portnoy (1984), via de regra, deve existir, no máximo, um número de $\sqrt{n}$ regressores no modelo, afim de se evitar sobreajuste. Nos modelos mistos, apenas um grau de liberdade é consumido, para a estimação da variância dos grupos.

É também possível com os modelos de efeitos mistos, através do uso de variáveis de mais alto nível hierárquico, inferir o impacto destas variáveis sobre a formação de valor, o que não é possível com os modelos de efeitos fixos, haja vista que nestes, como já explanado, toda a variabilidade em nível mais alto é absorvida pelas *dummies*. Assim, seria possível, por exemplo, estimar o quanto o investimento em infraestrutura básica poderia aumentar o valor dos imóveis em um determinado bairro ou região, ou ainda a criação de mais parques ou áreas verdes.

No entanto, para que isto seja possível é necessário que não haja omissão de variáveis importantes no segundo nível hierárquico dos dados. A omissão de variáveis importantes no segundo nível de modelagem afeta a estimação dos coeficientes dos efeitos entre os agrupamentos (*between or contextual effects*) no primeiro nível e nas outras variáveis do segundo nível (BELL; FAIRBROTHER; JONES, 2019, p. 1060).

A hipótese da normalidade dos efeitos aleatórios nem sempre será verificada, o que pode levar alguns analistas a abandonar esta modelagem em detrimento da modelagem de efeitos fixos, dado que esta última não faz esta hipótese. Apesar de realmente ser um problema, segundo Bell, Fairbrother e Jones (2019, p. 1070) a violação desta hipótese não impacta muito fortemente as estimativas da parte fixa do modelo, nem os erros-padrões (os coeficientes dos regressores e os seus erros padrão são não-viesados mesmo na ausência da normalidade para vários tipos de modelos testados), mas apenas as estimativas dos efeitos aleatórios em si e, mesmo assim, para variáveis dependentes contínuas (como é o caso na Engenharia de Avaliações) não foram encontradas evidências da presença de viés nos modelos mesmo com a simulações com efeitos aleatórios afastados na normalidade.

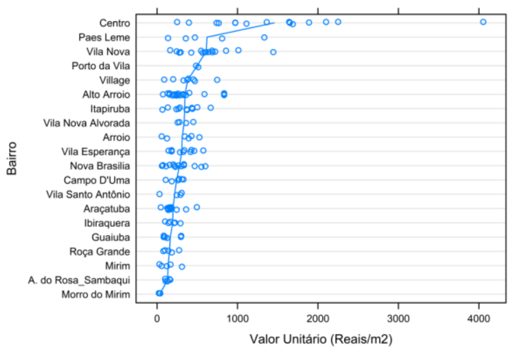
3. ESTUDO DE CASO

Foi elaborado um estudo de caso no município de Imbituba, localizado ao sul de Santa Catarina. O município abrange uma área territorial de 182 km² e fica distante aproximadamente 90 km da capital Florianópolis. Segundo o plano diretor do município todo o território é considerado urbano, logo o IPTU incide sobre toda a cidade. O município é dividido em cinco zonas de planejamento.

Foram coletados 170 dados amostrais, dos quais foram efetivamente utilizados apenas 166. Quatro dados foram retirados como *outliers*. Os dados estão distribuídos por todas as regiões do município. Foram levantadas informações acerca desses bairros para a composição das variáveis de nível dois ou variáveis de grupo. O município é composto por 30 bairros, entretanto existem dados de terrenos coletados apenas para 20 bairros.

A Figura 1 mostra a distribuição da variável *VU* em cada bairro, mostrando-se uma tendência de aumento de preços unitários dos terrenos na área mais central do município, como é o caso dos bairros do Centro, Vila Nova e Paes Leme.

Figura 1 - Distribuição da variável VU por bairro.



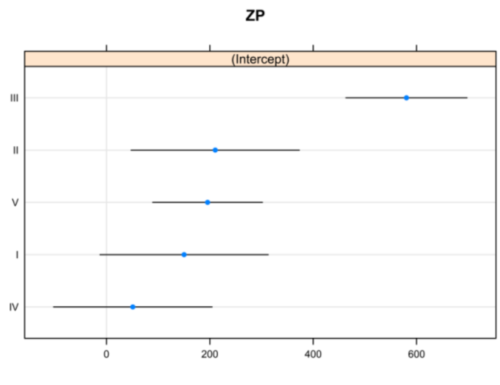
Fonte: Os autores (2020).

3.1 Estudo sobre zonas de planejamento

Inicialmente foi realizado um estudo sobre as zonas de planejamento, em que se constatou que as zonas de planejamento são I, II, IV e V são muito parecidas entre si em termos de valor dos imóveis, enquanto a zona III é significativamente diferente das demais.

Para melhor averiguar a diferença entre as zonas de planejamento, foi ajustado um modelo misto agrupando os bairros entre as zonas, além das variáveis de efeitos fixos de primeiro nível Área do imóvel (A), numérica e contínua; Pavimentação (PAV), variável dicotômica que mede o quanto o fato de uma rua estar pavimentada aumenta o valor unitário de um imóvel e; Posição na Quadra (PQ) variável dicotômica que representa o efeito da posição na quadra (esquina ou meio de quadra) no valor unitário dos imóveis. A Figura 2 mostra os efeitos aleatórios deste modelo para as diversas zonas de planjamento:

Figura 2 - Efeitos aleatórios do modelo misto com zonas de planjamento



Fonte: Os autores (2020).

Pode-se notar na Figura 2 que a diferença de valores unitários entre algumas zonas de planjamento não parecem ser relevantes: apenas a zona III parece estar deslocada das demais.

Para uma melhor comparação entre as zonas, então, foi utilizado o método das múltiplas comparações de médias de Tukey (*Tukey Honest Differences*). Os resultados obtidos podem ser visualizados na Tabela 1, em que a coluna 1 apresenta as zonas analisadas duas a duas, a segunda coluna apresenta a diferença média de valores entre elas, as colunas 3 e 4 os limites inferior e superior destes intervalos de confiança das diferenças médias (@95%) e, finalmente, a coluna 5 representa a significância da diferenca entre as médias:

É difícil dizer se estas diferenças podem ser utilizadas para dividir a cidade entre zonas rural e urbana: existem outras diferenças entre os imóveis que não apenas os seus valores unitários. Porém, a existência de uma zona de planejamento simultaneamente diferente de todas as demais (zona III) e, por sua vez, a existência de zonas de planejamento parecidas entre si em termos de valor (zonas I, II, IV e V), podem ser um bom indicativo de que as últimas devam pertencer a um contexto diferente.

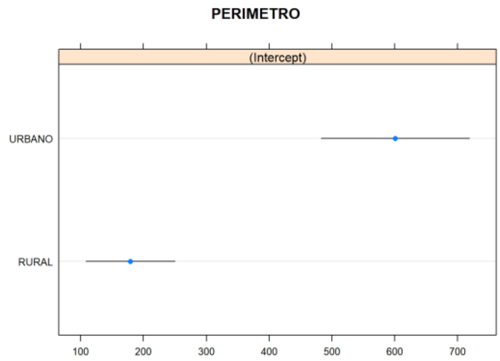
Tabela 1- Método das múltiplas comparações de médias de Tukey.

Zonas	**DIFF**	**L.I.**	**L.S.**	**P-VALOR**
II-I	75,54	-260,75	411,85	0,97
III-I	443,07	151,82	734,32	**<0,001**
IV-I	-83,37	-409,43	242,69	0,95
V-I	47,79	-234,32	329,90	0,99
III-II	367,52	76,28	658,77	**0,006**
IV-II	-158,92	-484,93	167,14	0,66
V-II	-27,75	-309,87	254,36	0,99
IV-III	-526,44	-805,80	-247,09	**<0,001**
V-III	-395,28	-621,80	-168,76	**<0,001**
V-IV	131,16	-138,66	400,98	0,67

Fonte: Os autores (2020)

Foi ajustado então, primeiramente, uma nova variável dicotômica, englobando as zonas I, II, IV e V, chamando-a perímetro "rural" e a zona III como perímetro "urbano". Posteriormente, foi ajustado um modelo com essa variável, em que se obteve uma diferença significante de nível entre as os dois perímetros, o que pode ser visualizado na Figura 3.

Figura 3 - Diferença de nível entre os perímetros "rural" e "urbano"



Fonte: Os autores (2020).

3.2 Estudo com dados agrupados em perímetros

Foram utilizados, então, 166 dos 170 dados pesquisados para o ajuste de vários tipos de modelos: (1) um modelo de regressão linear ordinária, onde ignorou-se a heterogeneidade da amostra; (2) um modelo de efeitos fixos, onde os dados foram agrupados em bairros, ignorando-se em que perímetros estejam localizados; (3) um modelo de efeitos mistos considerando-se os dados agrupados em bairros e estes agrupados em perímetros diferentes; (4) um modelo misto análogo ao anterior, com a adição de variáveis de níveis mais altos e; (5) um modelo análogo ao anterior, com a adição de inclinações aleatórias.

A Tabela 3 apresenta os valores estimados para os coeficientes, junto com os valores da estatística t e seus p-valores codificados, além de diversas estatísticas para cada modelo. Os coeficientes das dummies (modelo 2) foram omitidos por questões de espaço, mas todas apresentaram significância estatística. Nos modelos foram utilizadas novamente as variáveis A, PAV e PQ, citadas acima, além de variáveis de segundo (ACESSO_BR) nível hierárquico, que modela características de acesso de cada bairro.

Nota-se que o modelo de efeitos fixos (coluna 2) apresentou um grau de ajuste muito alto, o que é indicativo de sobreajuste do modelo[3]. Já com os modelos mistos foram obtidos modelos bem ajustados com a utilização de muito menos graus de liberdade, já que foram necessários apenas o consumo de 6 a 8 graus de liberdade para a estimação destes modelos.

Tabela 2 - Comparacão dos modelos de efeitos mistos.

Nº Param.	AIC	BIC	loglLik	Desvio	χ^2	Graus Liberdade	$P(>\chi^2)$
6	300,82	319,49	-144,41	288,82	-	NA	NA
7	298,96	320,74	-142,48	284,96	3,86	1	0,05
8	371,57	396,46	-177,78	355,57	0,00	1	1,00

Fonte: Os autores (2020)

[3] Isto deve ser atribuído ao fato de que existem vinte diferentes agrupamentos na amostra. Para uma estimação satisfatória com este número de coeficientes, seria necessário um número muito maior de dados do que o disponível.

Na tabela 2 apresenta-se a comparação dos modelos mistos (modelos 3 a 5 da tabela 3) através da análise da variância: o baixo valor do p-valor na segunda linha da tabela (última coluna) mostra que o segundo modelo (com variável de segundo nível hierárquico) mostrou-se relevante, porém o terceiro modelo (com inclinações aleatórias) não.

A utilização de modelos de efeitos mistos possibilitou, então, a introdução de variáveis de nível hierárquico mais alto.

Tabela 3 - Comparacão dos modelos de regressão linear ordinária e efeitos mistos.

Regressores	*Variável Dependente: log(VU)*				
	Regressão Linear Ordinária		***Efeitos Mistos***		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
log(A)	0,75	-0,34	-0,36	-0,36	-0,30
	t = 31,19***	t = -8,04***	t = -8,87***	t = -8,90**	t = -7,39***
PAV	0,75	0,43	0,46	0,43	0,43
	t = 3,13***	t = 4,41***	t = 4,90***	t = 4,58***	t = 4,44***
PQ	0,44	0,17	0,18	0,21	0,17
	t = 1,59**	t = 1,59**	t = 1,79***	t = 2,03***	t = 1,66***
ACESSO_BR	-	-	-	0,01	0,17
	-	-	-	t = 2,08***	t = 7,88**
Observations	166	166	166	166	166
R^2	0,93	0,99	-	-	-
R^2 ajustado	0,93	0,99	-	-	-
Log Likelihood	-	-	-144,41	-142,48	-177,78
AIC	611,11	264,62	300,82	298,96	371,57
BIC	623,56	339,31	319,49	320,74	396,46
E. P. Resíduos	1,50 (df = 163)	0,50 (df = 143)	-	-	-
Nota:	** p<0,3;* ** p<0,2;*** p<0,1				

Fonte: Os autores (2020)

4. CONSIDERAÇÕES FINAIS

A modelagem de efeitos mistos é um tipo de modelagem atualmente negligenciado na Engenharia de Avaliações, assim como em muitas outras áreas da ciência. No dia a dia da avaliação de imóveis é improvável que este tipo de modelagem venha a ganhar alguma relevância num futuro próximo. No entanto, dado o avanço da importância que os mercados imobiliários vem ganhando na economia mundial nos últimos anos - num mundo em que estão escassas boas oportunidades de investimento, é provável que o setor imobiliário venha a ter algum protagonismo na superação dos trágicos efeitos econômicos da pandemia de COVID-19 - é importante que novos métodos sejam investigados com fins de que haja uma melhor regulação dos mercados imobiliários, que possibilitem um melhor planejamento dos investimentos, especialmente os investimentos públicos necessários para o bom funcionamento do setor, e que canalize os recursos para locais em que estes são mais importantes. Com a modelagem por efeitos mistos é possível fazer diagnósticos mais precisos do solo urbano e os investimentos necessários para a valorização dos imóveis.

Com a valorização dos imóveis é natural que se aumente a arrecadação e isto possibilita a realização de novos investimentos, criando um círculo virtuoso que pode vir a ser o motor da economia nos anos pós-COVID-19.

A necessidade de uma normatização da atividade de confecção de PVGs pode ser suprida através da adoção de um papel central para os modelos mistos. A modelagem por efeitos mistos, especialmente através da formulação de Mundlak, se encaixa perfeitamente nas necessidades da elaboração de uma PVG. Se existem métodos talvez mais precisos para a previsão de valores, estes métodos sofrem de uma série de desvantagens quando comparados aos modelos mistos: muitos são verdadeiras caixas-pretas, deixando de lado a explicação da formação do valor, em troca de uma possível melhor previsão de valores, o que sempre ocorre para dados amostrais e não são garantia de melhor previsão para a população de imóveis em geral, o que é questionável, mesmo na presença de métodos de validação cruzada; outros são explícitos, porém não explicam a formação de valor advinda da localização, como ocorre com o método padrão para a confecção de PVGs, a modelagem por efeitos fixos.

Com a recente implantação de diversos observatórios de valores imobiliários no Brasil, o que foi possível e necessário nos últimos anos devido ao aquecimento do mercado imobiliário nacional, cada dia mais dados estarão disponíveis para análise. O tratamento destes dados requer a adoção de novos métodos. A confecção de índices de preços de imóveis mais confiáveis é uma necessidade no Brasil. Considera-se que a modelagem por efeitos mistos deva ser investigada em trabalhos futuros, visando a confecção de um ou mais índices de preços nacionais, regionais ou municipais, dada a aplicabilidade destes modelos para a análise de dados em painel.

No estudo de caso pode-se ver que é possível e desejável a utilização de modelos mistos na engenharia de avaliações, especialmente na confecção de PVG's, pois assim é possível tanto o agrupamento dos dados em diversos níveis de análise, assim como a modelagem destes níveis mais altos, através da inclusão de variáveis nestes níveis hierárquicos, o que a modelagem de efeitos fixos impossibilita ao consumir todos os graus de liberdade *de nível mais alto.* Além disto, apesar de não ter se mostrado relevante no estudo de caso apresentado, uma outra característica interessante dos modelos mistos é a fácil extensão da modelagem para a adição de inclinações aleatórias, cuja estimação com modelos fixos tende a ser pobre, devido ao alto consumo de graus de liberdade com aquele modelo.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653**: Avaliação de bens parte 2: Imóveis urbanos. Rio de Janeiro, 2011. 24 p.

BATES, D. et al. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1-48, out. 2015.

BELL, A.; FAIRBROTHER, M.; JONES, K. Fixed and random effects models: Making an informed choice. **Quality and Quantity**, v. 53, p. 1051–1074, 2019. DOI: https://doi.org/10.1007/s11135-018-0802-x.

BELL, A.; JONES, K. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. **Political Science Research and Methods**, v. 3, n. 1, p. 133-153, 2015. DOI: https://doi.org/10.1017/psrm.2014.7.

CICHULSKA, A.; CELLMER, R. Analysis of Prices in the housing market using mixed models. **Real Estate Management and Valuation**, v. 26, n. 4, p. 102-111, 2018.

JONES, K.; BULLEN, N. Contextual Models of Urban House Prices: A Comparison of Fixed and Random-Coefficient Models Developed by Expansion. **Economic Geography**, v. 70, n. 3, p. 252-272, 1994.

MATLOFF, N. **Statiscal Regression and Classification: From Linear Models to Machine Learning.** 1. ed. Boca Raton, FL: CRC Press, 2017.

PORTNOY, S. Asymptotic behavior of *m*-estimators of *p* regression parameters when p^2/n is large. **Ann. Statist.** 12, 4, p. 1298-1309. 1984

RAO, R. B.; FUNG, G.; ROSALES, R. On the dangers of cross-validation. An experimental evaluation. In: 2008 SIAM International Conference on Data Mining. **Proceedings…**. p. 588-596, 2008.

SHAO, J. Linear model selection by cross-validation. **Journal of the American Statistical Association**, v. 88, p. 486-494. 1993.