

Modelos hierárquicos na Engenharia de Avaliações

Para confecção de PVG's e índices de mercado

Luiz F. P. Droubi Carlos Augusto Zilli Norberto Hoccheim

20/08/2020

1 Introdução

2 Revisão Bibliográfica

3 Estudo de Caso

3.1 Criação de dados via simulação

Foram criados 550 dados de lotes, divididos igualmente em 11 bairros, a partir de simulação com o auxílio do software R.

Os dados foram criados conforme parâmetros da tabela 1:

Variável	Tipo	Distribuição	Parâmetros	Obs
Área (A)	Quantitativa	Normal	$\mu = 400, \sigma = 50$	-
Bairro	Qualitativa	-	A a K	-
Áreas Verdes (A_V)	Quantitativa	Uniforme	$\mu = 0,2 \quad a \quad 0,70$	Um valor para cada bairro
β_0	Coeficiente	Discreta	2000	-
v	Termo de erro	Normal	$\mu = 0, \sigma = 150$	-
β_{0j}	Coeficiente	Não definida	$\beta_0 + 3000A_V + v$	-
ϵ	Termo de erro	Normal	$\mu = 0, \sigma = 50$	-
Valor Unitário (VU)	Quantitativa	Não definida	$\beta_{0j} - 3,0A + \epsilon$	-

3.2 Análise exploratória dos dados

Na figura 1 é possível ver os principais gráficos dos dados gerados.

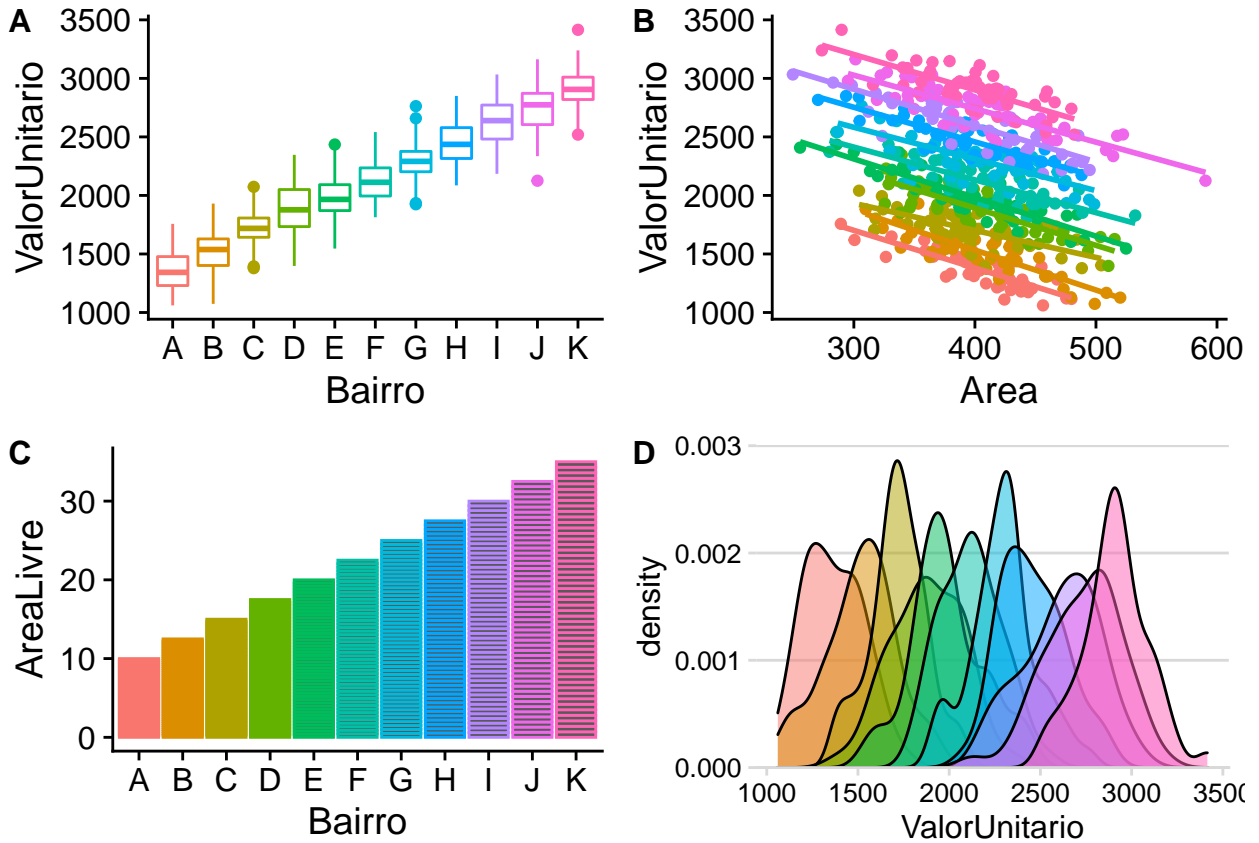


Figura 1: Análise exploratória dos dados.

3.3 Ajuste de modelos

Com os dados gerados foram ajustados um modelo de efeitos fixos e três modelos mistos: um modelo misto simples, que praticamente equivale ao modelo de efeitos fixos, um modelo misto com a adição de uma variável de segundo nível e um modelo misto utilizando-se a formulação de Mundlak.

Para o ajuste do modelo de efeitos fixos foram utilizados todos os dados gerados, pois não há como prever valores para bairros não contemplados na amostra em um modelo deste tipo.

Para o ajuste dos modelos mistos foram removidos os 50 dados relativos ao bairro H, que foram reservados para serem utilizados posteriormente para validação dos modelos, mostrando como a previsão de valores em modelos de efeitos mistos pode ser feita para agrupamentos não contemplados na amostra.

3.3.1 Modelo de efeitos fixos

Com os dados gerados, foi elaborado um modelo de efeitos fixos, do tipo:

$$ValorUnitario = \beta_0 + \beta_1 Area + \beta_{2i} Bairro_i + \epsilon$$

onde β_{2i} são os coeficientes das variáveis dicotômicas em grupo ($Bairro_i$).

3.3.2 Modelo misto simples

Também foi elaborado um modelo misto do tipo:

Tabela 2: Valores dos interceptos para cada bairro.

	A	B	C	D	E	F	G	I	J	K
(Intercept)	2.579,7	2.730,8	2.895,8	3.091	3.181,9	3.333	3.512,2	3.796,2	3.940	4.100

Tabela 3: Efeitos randômicos do modelo misto.

grp	vcov	sdcov
Bairro	240.555,13	490,46
Residual	12.105,17	110,02

$$ValorUnitario = \beta_0 + \beta_1 Area + v_i + \epsilon$$

Onde v_i é uma variável aleatória que foi utilizada para modelar os diferentes bairros.

Os efeitos aleatórios do modelo misto simples podem ser visualizados na Figura 2.

Como se pode notar na Figura 2, os valores dos interceptos aleatórios para cada bairro giram em torno de zero, o seu valor médio. Como o Bairro H (com $A_L = 0,55$) foi omitido no ajuste do modelo, não há valores estimados para os efeitos aleatórios para este bairro.

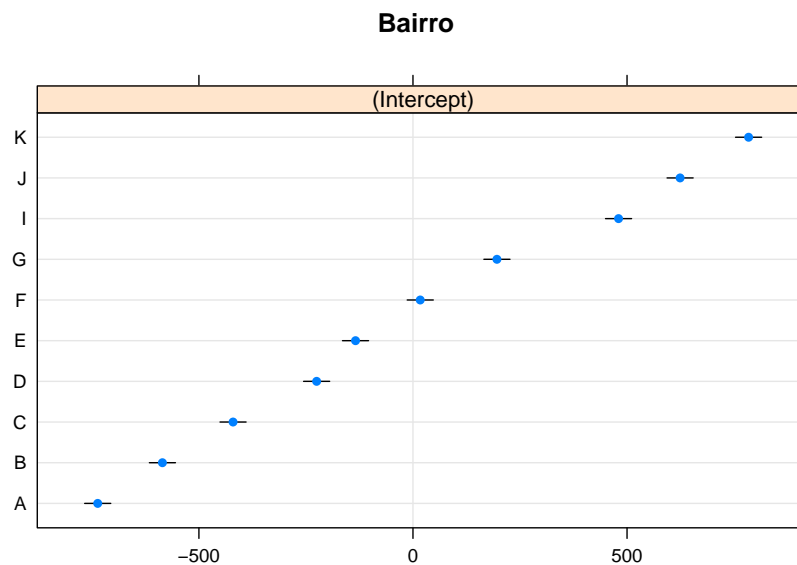


Figura 2: Efeitos aleatórios do modelo.

Os valores de referência para cada bairro podem ser obtidos através da soma do intercepto global do modelo misto simples com os interceptos aleatórios, o que pode ser visto na tabela 2.

Como se pode perceber, quase não houve diferença entre os valores estimados em cada modelo.

A única informação a mais que se pode extrair do modelo de efeitos mistos é a componente de variância devido à localidade, separada da variância ao nível dos imóveis, o que pode ser visto na tabela 3.

Pode-se notar que a variância devido à localidade é relevante para o modelo, haja vista que a variância devido à localidade é maior do que a variância devido às características dos imóveis.

3.3.3 Modelo misto com variáveis de segundo nível

Finalmente, foi elaborado um modelo misto do tipo:

$$ValorUnitario = \beta_0 + \beta_1 Area + \beta_2 A_V + v_i + \epsilon$$

Onde A_V é uma variável de nível 2 que representa a porcentagem de áreas verdes em cada bairro, em relação à área total.

3.3.4 Modelo misto com formulação de Mundlak

Finalmente, foi ajustado um modelo com a formulação de Mundlak. Este modelo foi elaborado de acordo com a seguinte formulação:

$$ValorUnitario = \beta_0 + \beta_1 Area + \beta_{1C} \overline{Area} + \beta_2 A_V + v_i + \epsilon$$

3.4 Resultados

A tabela 4 mostra as estatísticas básicas dos diversos modelos mistos (colunas (2), (3) e (4)) comparados aos modelo de efeitos fixos (coluna (1)).

Como pode ser vistos nesta tabela, o valor do intercepto global (a dita grande média) foi melhor estimada (comparando-se com o valor simulado) pelo modelo misto com a variável de segundo nível A_L , obtendo um resultado de R\$ 1.987,28. Isto ocorre porque o efeito da variável relevante A_V não inclusa no modelo anterior havia sido absorvida por este coeficiente.

Nos modelos onde houve a inclusão da variável Área Livre (A_L), seu coeficiente foi bem estimado: o valor da influência das áreas verdes, simulado como aumento R\$ 30,00/m² a cada ponto percentual a mais de áreas verdes no bairro do imóvel, foi precisamente estimado.

Para o modelo de Mundlak, a estimação do coeficiente contextual (β_{1C}) foi insignificante. Isto era esperado, dado que a variável Área foi simulada da mesma maneira para todos os bairros. Em outras palavras, a simulação foi feita como se os imóveis em todos os bairros tivessem a mesma distribuição (normal) com mesma média e desvio-padrão. Isto dificilmente ocorrerá na prática da elaboração de PVG's, mas é interessante notar a flexibilidade deste tipo de formulação: quando não existe na realidade o efeito esperado pela formulação, o coeficiente resultará insignificante. Diz-se em estatística que o modelo degenerou, ou seja, o modelo com a formulação de Mundlak se degenerou para uma formulação mais simples. Basta remover este termo da modelagem para obter-se o modelo mais correto para o caso. Portanto, na prática, deve-se partir da formulação mais complexa, no caso a de Mundlak, e observar se os resultados obtidos são significantes. Caso positivo, mantém-se o modelo. Caso contrário, descarta-se o termo insignificante.

Por último, porém não menos relevante, percebe-se que este modelo tem critérios de informação de Akaike (AIC) e de Bayes (BIC) melhores que os dois modelos iniciais.

Tabela 4: Comparação dos modelos de efeitos fixos e efeitos mistos.

	Dependent variable:			
	ValorUnitario			
	OLS		linear mixed-effects	
	(1)	(2)	(3)	(4)
Constant		3.316,04 (160,12)***	1.987,28 (43,75)***	1.858,21 (597,46)***
Area	-2,99 (0,09)***	-3,00 (0,10)***	-3,00 (0,10)***	-3,00 (0,10)***
BairroA	2.577,39 (40,67)***			
BairroB	2.728,65 (40,75)***			
BairroC	2.893,84 (39,59)***			
BairroD	3.089,23 (40,57)***			
BairroE	3.180,19 (40,11)***			
BairroF	3.331,46 (40,63)***			
BairroG	3.510,80 (40,71)***			
BairroH	3.655,00 (40,22)***			
BairroI	3.795,18 (39,72)***			
BairroJ	3.939,05 (40,43)***			
BairroK	4.099,23 (39,98)***			
Area_between				0,32 (1,48)
AreaLivre			3.018,16 (38,65)***	3.021,14 (43,42)***
Observations	550	500	500	500
R ²	1,00			
Log Likelihood		-3.094,33	-3.055,32	-3.054,02
Akaike Inf. Crit.	6.734,21	6.196,66	6.120,65	6.120,04
Bayesian Inf. Crit.	6.790,23	6.213,51	6.141,72	6.145,33
Residual Std. Error	108,90 (df = 538)			
F Statistic	18.776,45*** (df = 12; 538)			

Note:

*p<0,3; **p<0,2; ***p<0,1

3.5 Previsão de Valores

Na Figura 3 podem ser vistos os gráficos de poder de predição para o modelo de efeitos fixos (A), para o modelo misto simples (B), para o modelo misto com a variável de segundo nível (C) e para o modelo misto com formulação de Mundlak (D). Como pode ser visto, todos os modelos possuem poderes de predição praticamente equivalentes, com exceção do modelo misto simples, onde a previsão de valores não pode ser feita com precisão já que, como no modelo de efeitos fixos, este modelo não tem parâmetros para prever valores em bairros não contemplados na amostra. Para efetuar as previsões no bairro H, então, o modelo considerou para a variável aleatória v o valor zero, ou seja, o valor esperado da variável, o que levou a previsões incorretas em relação aos valores simulados para aquele bairro.

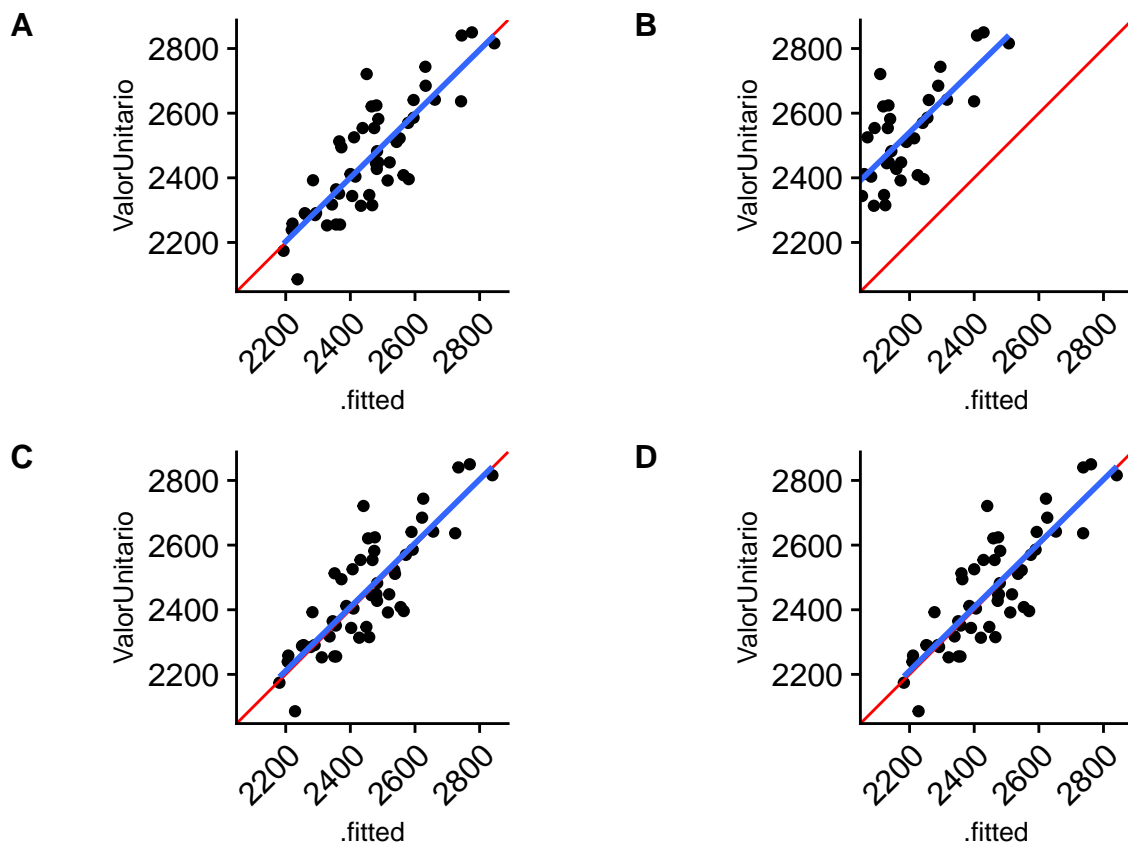


Figura 3: Gráficos de poder de predição para cada modelo.