Modelos mistos

Estudo de Caso

Luiz F. P. Droubi

20/08/2020

# Estudo de Caso

## Criação de dados via simulação

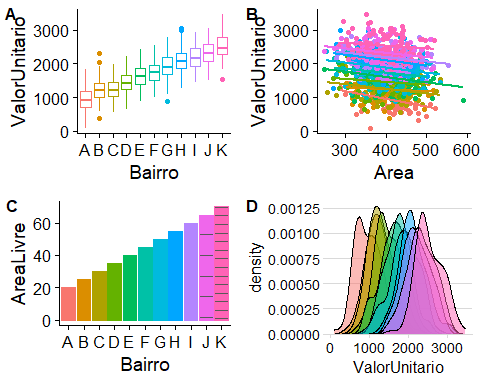
Foram criados 1100 dados a partir de simulação com o auxílio do software **R**.

Os dados foram criados conforme parâmetros da tabela 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variável | Tipo | Distribuição | Parâmetros | Obs |
| Área () | Quantitativa | Normal |  | - |
| Bairro | Qualitativa | - | A a K | - |
| Áreas Verdes () | Quantitativa | Uniforme |  | Um valor para cada bairro |
|  | Coeficiente | Discreta | 1000 | - |
|  | Termo de erro | Normal |  | - |
|  | Coeficiente | Não definida |  | - |
|  | Termo de erro | Normal |  | - |
| Valor Unitário () | Quantitativa | Não definida |  | - |

## Análise exploratória dos dados

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

 ## Ajuste de modelos

Com 90% dos dados gerados foram ajustados um modelo de efeitos fixos e dois modelos mistos. Para o ajuste do modelo foram removidos os dados relativos ao bairro F, que foram reservados para serem utilizados posteriormente para validação dos modelos.

### Modelo de efeitos fixos

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

onde são os coeficientes das variáveis dicotômicas em grupo ().

### Modelo de efeitos mistos

Também foi elaborado um modelo misto do tipo:

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

### Resultados

As estatísticas básicas dos coeficientes podem ser observadas na tabela .

Como pode ser vistos nesta tabela, o valor do intercepto global (a dita grande média) do modelo de efeitos mistos é de R$ 2.282,61.

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

Como se pode notar na Figura , os valores dos interceptos aleatórios para cada bairro giram em torno de zero, o seu valor médio.

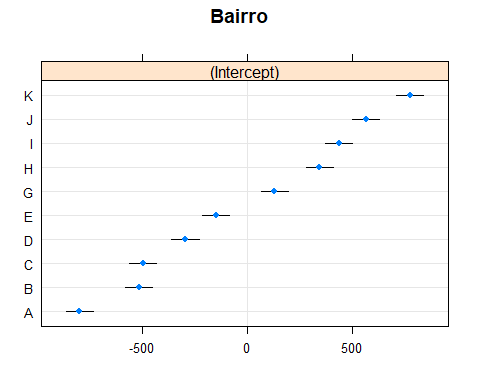
Os valores de referência para cada bairro podem ser obtidos através da soma do intercepto global do modelo misto com os interceptos aleatórios, o que pode ser visto na tabela .

# Comparacão dos modelos de efeitos fixos e efeitos mistos.

Dependent variable:   
 ---------------------------------------------  
 ValorUnitario   
 OLS linear   
 mixed-effects   
 (1) (2)

|  |
| --- |
| Constant 2.282,6 (178,1)\*\*\* |
| Area -1,3 (0,2)\*\*\* -1,3 (0,2)\*\*\* |
| BairroA 1.476,2 (84,7)\*\*\* |
| BairroB 1.761,6 (83,4)\*\*\* |
| BairroC 1.779,6 (84,0)\*\*\* |
| BairroD 1.981,7 (84,2)\*\*\* |
| BairroE 2.128,7 (83,4)\*\*\* |
| BairroF 2.255,9 (83,4)\*\*\* |
| BairroG 2.408,4 (82,0)\*\*\* |
| BairroH 2.624,1 (83,8)\*\*\* |
| BairroI 2.717,0 (83,8)\*\*\* |
| BairroJ 2.846,1 (83,8)\*\*\* |
| BairroK 3.056,3 (83,7)\*\*\* |

Observations 1.100 1.000  
Adjusted R2 1,0  
Log Likelihood -7.245,6  
Akaike Inf. Crit. 15.888,0 14.499,1  
Bayesian Inf. Crit. 15.953,0 14.518,7  
Residual Std. Error 329,2 (df = 1088)  
F Statistic 2.790,8\*\*\* (df = 12; 1088)  
================================================================= Note: *p<0,3;* ***p<0,2;*** p<0,1



Efeitos aleatórios do modelo.

Valores dos interceptos para cada bairro.

A

B

C

D

E

G

H

I

J

K

(Intercept)

1.484,4

1.768,5

1.786,4

1.987,6

2.134

2.412,4

2.627,3

2.719,8

2.848,3

3.057,6

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

A ínica 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 .

Efeitos randômicos do modelo misto.

grp

vcov

sdcor

Bairro

251.490,8

501,49

Residual

108.963,2

330,10

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.

### Modelo com variáveis de segundo nível

Finalmente, foi elaborado um modelo misto do tipo:

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

A tabela mostra as estatísticas básicas deste modelo comparado aos dois modelos previamente estudados.

# Comparacão dos modelos de efeitos fixos e efeitos mistos.

Dependent variable:   
 ---------------------------------------------------------------  
 ValorUnitario   
 OLS linear   
 mixed-effects   
 (1) (2) (3)

|  |
| --- |
| Constant 2.282,6 (178,1)\*\*\* 925,2 (90,9)\*\*\* |
| Area -1,3 (0,2)\*\*\* -1,3 (0,2)\*\*\* -1,3 (0,2)\*\*\* |
| BairroA 1.476,2 (84,7)\*\*\* |
| BairroB 1.761,6 (83,4)\*\*\* |
| BairroC 1.779,6 (84,0)\*\*\* |
| BairroD 1.981,7 (84,2)\*\*\* |
| BairroE 2.128,7 (83,4)\*\*\* |
| BairroF 2.255,9 (83,4)\*\*\* |
| BairroG 2.408,4 (82,0)\*\*\* |
| BairroH 2.624,1 (83,8)\*\*\* |
| BairroI 2.717,0 (83,8)\*\*\* |
| BairroJ 2.846,1 (83,8)\*\*\* |
| BairroK 3.056,3 (83,7)\*\*\* |
| AreaLivre 3.020,8 (86,2)\*\*\* |

Observations 1.100 1.000 1.000  
Adjusted R2 1,0  
Log Likelihood -7.245,6 -7.212,2  
Akaike Inf. Crit. 15.888,0 14.499,1 14.434,4  
Bayesian Inf. Crit. 15.953,0 14.518,7 14.458,9  
Residual Std. Error 329,2 (df = 1088)  
F Statistic 2.790,8\*\*\* (df = 12; 1088)  
=================================================================================== Note: *p<0,3;* ***p<0,2;*** p<0,1

Pode-se notar que este último modelo estimou precisamente os valores como simulados. Primeiramente, deve-se notar que o valor da influência das áreas verdes, simulado como aumento R$ 3.000,00/m2 a cada ponto percentual a mais de áreas verdes no bairro do imóvel, foi precisamente estimado.

Também deve-se notar que o valor de , o intercepto reflete muito melhor o valor atribuído a este coeficiente na simulação dos dados (1.000,00). Isto ocorre porque o efeito da variável relevante não inclusa no modelo anterior havia sido absorvida por este coeficiente.

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.

### Mundlak

## Previsão de Valores

