# Modelos mistos na Engenharia de Avaliações

Possibilidades e aplicações

Luiz F. P. Droubi[*] Carlos Augusto Zilli[†] Norberto Hoccheim[‡]

28/08/2020

## 1 Introdução

Na Engenharia de Avaliações, assim como em diversos ramos das ciências sociais, a abordagem padrão para lidar com a heteregoneidade amostral é a modelagem com efeitos fixos. Este tipo de abordagem permite a segregação dos dados da amostra em diferentes agrupamentos, no entanto, ela não…

## 2 Revisão Bibliográfica

A análise de dados multinível é usualmente feita, nos diversos ramos da ciência por meio de modelos de efeitos fixos. No entanto, de acordo com Bell *et al.* (2019, p. 1052), os modelos mistos bem especificados oferecem uma abordagem muito mais completa destes tipos de dados do que a modelagem por efeitos fixos.

Segundo Bell *et al.* (2019, p. 1051), existe uma confusão na literatura a respeito de modelos mistos, no que tange à aglutinação em uma modelagem de diversos tipos de efeitos, fixos e aleatórios, sendo que os modelos mistos mais complexos tem sido chamados, erroneamente, de modelos híbridos.

### 2.1 Modelagem por efeitos fixos

A modelagem por efeitos fixos é frequentemente aplicada na Engenharia de Avaliações, assim como em diversas outras áreas da ciência (BELL et al., 2019, p. 1057), apesar desta terminologia não ser normalmente empregada.

Um modelo de efeitos fixos nada mais é do que um modelo em que a heterogeneidade da amostra é "saneada" através da inclusão de variáveis *dummies*[1] representando cada agrupamento de dados. Após a inclusão destas variáveis saneando a amostra, ou seja, tornando-a homogênea, a estimação é feita pelo métodos dos mínimos quadrados ordinários.

[*]SPU/SC, lfpdroubi@gmail.com
[†]IFSC, carlos.zilli@ifsc.edu.br
[‡]UFSC, hochheim@gmail.com

[1]Neste trabalho as variáveis dicotômicas em grupo serão tratadas simplesmente pelo termo em inglês *dummies*.

A modelagem por efeitos fixos pode ser escrita conforme a equação 1, onde a variância entre os agrupamentos de dados (variância de nível mais alto) é modelada através de variáveis *dummies* $D_j$ (BELL; JONES, 2015, p. 138):

$$y_{ij} = \sum_{j=1}^{j} \beta_{0j} D_j + \beta_1 x_{ij} + \epsilon_{ij} \tag{1}$$

No entanto, uma outra maneira mais conveniente - para a comparação que se pretende de escrever a formulação de efeitos fixos, equivalente à primeira, pode ser vista na equação 2 (BELL et al., 2019, p. 1058):

$$y_{ij} = \beta_1 (x_{ij} - \bar{x}_j) + (\upsilon_i + \epsilon_{ij}) \tag{2}$$

Onde $\upsilon_i$ é um termo discreto para cada agrupamento de dados[2]. 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. $\epsilon_{ij}$ é um termo de erro aleatório com distribuição supostamente normal, média zero e desvio-padrão $\sigma_\epsilon^2$.

Na prática, no entanto, a formulação acima raramente é utilizada, pois perde-se um grau de liberdade na estimação de um intercepto para cada bairro, quando se pode utilizar um nível de referência e estimar apenas a diferença entre os níveis de cada agrupamento em relação a este nível de referência, como ilustrado na equação 3 (BELL et al., 2019, p. 1058):

$$(y_{ij} - \bar{y}_i) = \beta_1 (x_{ij} - \bar{x}_j) + (\epsilon_{ij}) \tag{3}$$

## 2.2 Modelagem por efeitos mistos

Existem diversas maneiras de utilização da modelagem por efeitos mistos. Neste trabalho apresentam-se as diversas possibilidades, desde a abordagem mais simples, que possibilita uma melhor comparação com o modelo de efeitos fixos, muito conhecido na Engenharia de Avaliações, até a abordagem mais complexa, a formulação REWB.

### 2.2.1 Modelo de efeitos mistos simples

A modelagem por efeitos mistos considera, além dos efeitos fixos, um termo de efeitos aleatórios (BELL et al., 2019, p. 1059). Uma das maneiras de escrever a formulação de efeitos mistos pode ser vista na equação 4.

$$y_{ij} = \beta_0 + \beta_1^{RE} x_{ij} + \beta_2 z_j + (\upsilon_j + \epsilon_{ij}) \tag{4}$$

No caso da equação 4, o termo $\upsilon_j$ é um termo estocástico de efeitos aleatórios para os indivíduos, suposto normalmente distribuído e com média zero, ou seja $\upsilon_j \sim N(0, \sigma_\upsilon^2)$ (BELL et al., 2019, p. 1055). A variável $x_{ij}$ é uma variável de nível 1, variante entre os diferentes agrupamentos e a variável $z_j$ é uma variável de nível 2, considerada invariante entre os agrupamentos. Neste tipo de modelagem estimam-se, além dos coeficientes das variáveis de nível

[2] O termo $\upsilon_i$ é o intercepto de cada agrupamento, quando se ajusta um modelo sem intercepto.

mais baixo $\beta_0$ e $\beta_1$, o(s) coeficiente(s) da(s) variável(eis) de segundo nível $\beta_2$ e os parâmetros $\sigma_\upsilon^2$ e $\sigma_\epsilon^2$, ou seja, as variâncias de primeiro e segundo nível, respectivamente.

A diferença de nível entre as variáveis pode ser melhor compreendida ao se dividir a modelagem em dois níveis (alguns autores se referem a este tipo de modelagem como hierárquica), como pode ser visto nas equações 5 e 6:

$$y_{ij} = \beta_{0j} + \beta_1^{RE} x_{ij} + \epsilon_{ij} \tag{5}$$

$$\beta_{0j} = \beta_0 + \beta_2 z_j + \upsilon_j \tag{6}$$

Segundo Bell e Jones (2015, pp. 135–136), a equação 5 é chamada de parte micro, enquanto a equação 6 é chamada de parte macro da formulação de efeitos mistos, que são estimadas em conjunto ao substituir 6 em 5 para se obter o modelo misto da equação 4.

Em suma, para Bell *et al.* (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 agrupamentos 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 dos diferentes agrupamentos, na modelagem de efeitos aleatórios é considerado que a diferença entre os dados de diferentes agrupamentos pode ser modelada por uma variável aleatória normal. Assim, os modelos de efeitos fixos consomem todos os graus de liberdade do segundo nível hierárquico , matando assim a possibilidade de inferências a respeito da variância em nível mais alto, sendo impossível medir os efeitos de variáveis como $z_j$, na equação 4 (BELL; JONES, 2015, p. 139).

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 $\upsilon_i$) e os regressores ($x_i j$), o que é permitido na modelagem de efeitos fixos (BELL et al., 2019, p. 1060).

Este tipo de modelagem ainda pode ser facilmente estendida para incorporar outros níveis hierárquicos mais altos, assim como variabilidade não apenas para os interceptos, mas também para os coeficientes das variáveis explicativas (BELL et al., 2019, p. 1052).

### 2.2.2 Formulação de Mundlak

Esta formulação consiste da introdução de um termo adicional à parte macro do modelo, que leva em conta a variação entre os grupos (*between effect*), de maneira que a equação 6 torna-se:

$$\beta_{0j} = \beta_0 + \beta_2 z_j + \beta_3 \bar{x}_j + \upsilon_j \tag{7}$$

A combinação das equações 5 e 7 toma a forma da equação 8, que é conhecida na literatura como formulação de Mundlak (BELL; JONES, 2015, p. 1055):

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_3 \bar{x}_j + \beta_2 z_j + (\upsilon_j + \epsilon_{ij}) \tag{8}$$

Segundo Bell *et al.* (BELL et al., 2019, p. 1055), na formulação de Mundlak o efeito contextual, representado na equação 8 pelo termo $\beta_3$ (algumas vezes escrito como $\beta_C$), é de interesse, pois mostra a *diferença* entre os efeitos dentro (*within effect*) e entre (*between effect*) os grupos.

Na prática, segundo Bell *et al.* (BELL et al., 2019, p. 1057), os modelos de Mundlak e os modelos de efeitos fixos estimarão exatamente os mesmos valores para os coeficientes de efeitos dentro dos agrupamentos (*within effects*), representado na formulação acima pelo termo $\beta_1$ (algumas vezes escrito $\beta_W$).

Esta formulação, segundo Bell *et al.* (2015, p. 1056), é particularmente interessante para a análise de dados em seção transversal, pois o coeficiente $\beta_C$ representa o efeito da mundança de grupo de um indivíduo, mantidas as suas características. Na Engenharia de Avaliações, por exemplo, o valor de $\beta_C$ pode representas o quanto um lote-padrão aumentaria (ou diminuiria) de valor se ele pertencesse a um outro bairro ou zona, o que é particularmente interessante para a confecção de planta de valores genéricos.

### 2.2.3 Formulação Within-Between

Uma maneira às vezes mais adequada de escrever a formulação de modelos mistos consiste na separação total dos efeitos dentro dos agrupamentos dos efeitos entre os agrupamentos, o que é conhecido na literatura por formulação *within-between*. Esta formulação é a mais genérica, capaz de modelar diversos efeitos separadamente e é particularmente interessante na análise de dados em painéis ou séries temporais, dada a sua melhor interpretabilidade para estes tipos de dados (BELL; JONES, 2015, p. 143).

Partindo da formulação de Mundlak, os efeitos *within* e *between* podem ter seus efeitos totalmente separados pela divisão do coeficiente $\beta_3$ da equação 8, escrevendo-o explicitamente como uma diferença em relação ao coeficiente $\beta_1$, conforme mostrado pela equação 10

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (\beta_4 - \beta_1)\bar{x}_j + \beta_2 z_j + (\upsilon_j + \epsilon_{it}) \tag{9}$$

Rearranjando conveniente a equação 10, chega-se à formulação REWB, como mostra a equação 11:

$$y_{ij} = \beta_0 + \beta_1(x_{ij} - \bar{x}_j) + \beta_4 \bar{x}_j + \beta_2 z_j + (\upsilon_j + \epsilon_{it}) \tag{10}$$

É possível ainda a construção de um modelo ainda mais genérico que permite não apenas a modelagem de interceptos aleatórios mas também a modelagem de coeficientes aleatórios, como descrito na equação 11:

$$y_{ij} = \mu + \beta_W(x_{ij} - \bar{x}_j) + \beta_B \bar{x}_j + \beta_2 z_j + \upsilon_{j0} + \upsilon_{j1}(x_{ij} - \bar{x}_j) + \epsilon_{ij0} \tag{11}$$

Onde o termo $\upsilon_{j0}$ está relacionado à aleatoriedade do intercepto e o termo $\upsilon_{j1}$ está relacionado à aleatoriedade do coeficiente da variável $x$.

## 2.3 Considerações sobre a pertinência de cada modelagem

Segundo Bell e Jones (2015, p. 143), um modelo de efeitos fixos pode ser visto como uma forma de modelo de efeitos mistos onde a variância nos níveis mais altos é restringida, de maneira que para BELL; JONES (2015) os modelos de efeitos mistos são muito mais interessantes do que os modelos de efeitos fixos, já que além de propiciarem todas as informações que os modelos de efeitos fixos propriciam, eles ainda tem possibilidade de ir além e fornecer outras informações que os modelos de efeitos fixos não fornecem, como a separação dos efeitos de uma variável entre (efeito *between*) os agrupamentos e dentro (efeito *within*) dos agrupramentos. Para Bell *et al.* (2019, p. 1060), para o efeito dentro dos agrupamentos (*within effect*), os resultados estimados pela formulação REWB ou pela formulação de Mundlak serão rigorosamente os mesmos obtidos pelo modelo de efeitos fixos. No entanto, enquanto os modelos de REWB e de Munlak fornecem informações a respeitos dos efeitos entre os agrupamentos (*between effects*), os modelos de efeitos fixos não são capazes de fornecer qualquer informação a este respeito.

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. Para poucos agrupamentos, a modelagem por efeitos fixos ainda parece ser mais adequada (ver BELL et al., 2019, p. 1071).

Também deve ser observado que a utilização da formulação simples para modelos mistos, como a da equação 4, como muitas vezes se encontra na prática (ver CICHULSKA; CELLMER, 2018), não é tão interessante como a aplicação das modelagens REWB e de Mundlak. No entanto, a aplicação da formulação REWB e de Mundlak só faz sentido se houver efetivamente uma diferença entre os efeitos dentro e entre os agrupamentos. Caso $\beta_{1W} = \beta_{2B}$ (REWB) ou $\beta_{2C} = 0$ (Mundlak), não faz sentido utilizar estas formulações e a formulação simples de efeitos mistos (equação 4) deve ser a adotada (BELL et al., 2019, p. 1058).

### 2.4 Estimação em modelos mistos

Segundo Clark (2019), a estimação em modelos mistos é feita à partir do encolhimento.

## 3 Estudo de Caso

Para o estudo de caso foi utilizada a implementação do pacote **lme4** (BATES et al., 2015) no R, versão 4.0.2 (R CORE TEAM, 2020).

### 3.1 Criação de dados via simulação

Foram criados 500 dados de lotes, divididos igualmente em 10 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 | $\mu = 400 @ 500, \sigma = 50$ | - |
| Bairro | Qualitativa | - | A a J | - |
| Áreas Verdes ($A_V$) | Quantitativa | Uniforme | $\mu = 0,2 \quad a \quad 0,65$ | Um valor para cada bairro |
| $\beta_0$ | Coeficiente | Discreta | 2500 | - |

| Variável | Tipo | Distribuição | Parâmetros | Obs. |
|---|---|---|---|---|
| $\upsilon$ | Termo de erro | Normal | $\mu = 0, \sigma = 50$ | - |
| $\beta_{0j}$ | Coeficiente | Não definida | $\beta_0 + 4000A_{Vj} - 3.5\overline{A}_j + \upsilon_j$ | - |
| $\epsilon$ | Termo de erro | Normal | $\mu = 0, \sigma = 100$ | - |
| 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.

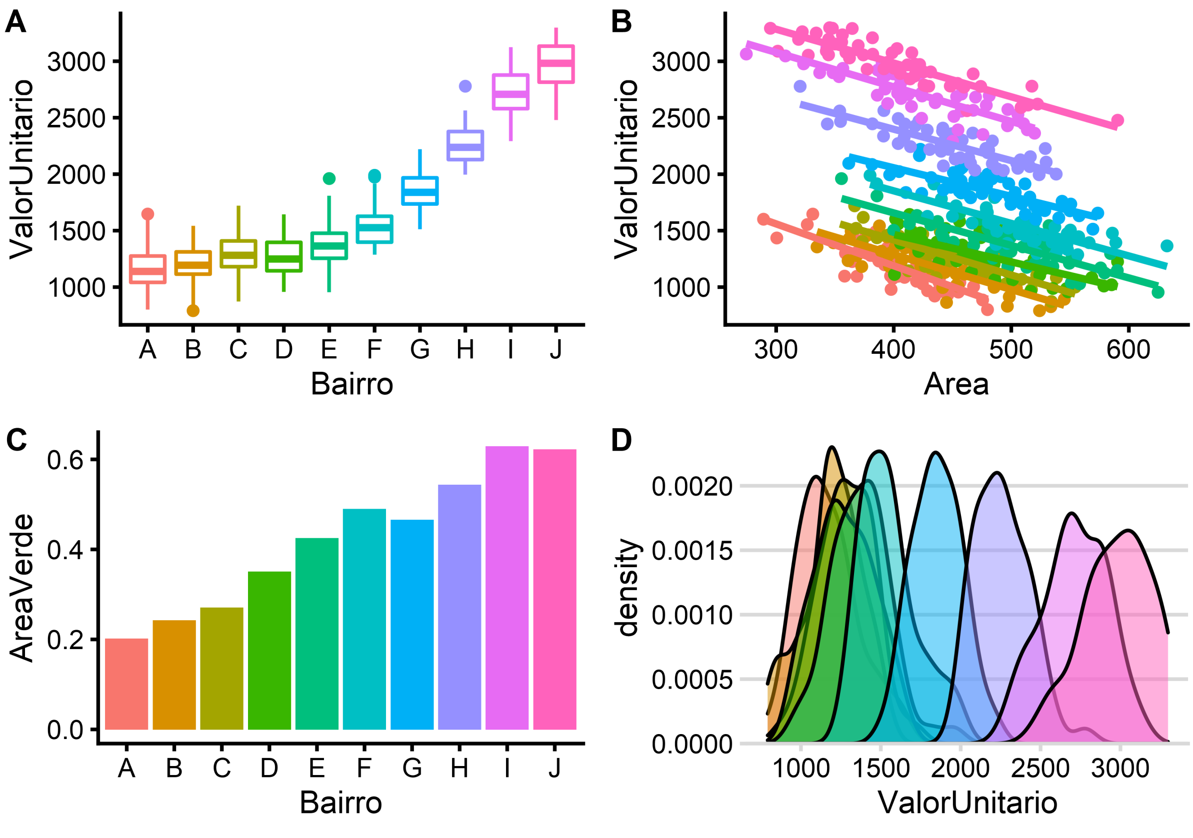


Figura 1: Análise exploratória dos dados.

## 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.

#### 3.3.1 Modelo de efeitos fixos

Com os dados gerados, foi elaborado um modelo de efeitos fixos sem intercepto, apenas para que fique claro o valor do intercepto aleatório de cada bairro. Para dar interpretação a estes interceptos aleatórios, a variável $Area$ foi centralizada em relação à área de um lote-padrão, considerado de 400 $m^2$[3]

O modelo é descrito pela equação 12:

$$ValorUnitario = \beta_1(Area - 400) + \beta_{2i}Bairro_i + \epsilon \tag{12}$$

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

#### 3.3.2 Modelos mistos

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.

**3.3.2.1 Modelo misto simples** Foi elaborado um modelo misto simples, sem separação de efeitos entre e dentro dos agrupamentos, de acordo com a equação 13:

$$ValorUnitario = \beta_0 + \beta_1(Area - 400) + v_i + \epsilon \tag{13}$$

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

**3.3.2.2 Modelo misto com variável de segundo nível** Foi elaborado um modelo misto simples, porém com a presença de variáveis de segundo nível hierárquico, como demonstrado na equação 14

$$ValorUnitario = \beta_0 + \beta_1(Area - 400) + \beta_2 A_V + v_i + \epsilon \tag{14}$$

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] A centralização de variáveis aqui presente não pretende nenhuma separação entre efeitos *within* e *between*, mas apenas possibilitar uma interpretação para os valores do coeficientes dos interceptos para cada bairro. Ver Droubi *et al.* (2019) para mais detalhes sobre este tipo de centralização.

#### 3.3.2.3 Modelo misto com formulação de Mundlak

Finalmente, foi ajusta um modelo com a formulação de Mundlak. Este modelo foi elaborado de acordo com a formulação exibida na equação 15:

$$ValorUnitario = \beta_0 + \beta_1 Area + \beta_{1C}\overline{Area_i} + \beta_2 A_V + \upsilon_i + \epsilon \qquad (15)$$

## 3.4 Resultados

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

Tabela 2: Comparacão dos modelos de efeitos fixos e efeitos mistos.

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | ValorUnitario | | | |
| | *OLS* | | *linear mixed-effects* | |
| | (1) | (2) | (3) | (4) |
| Intercepto | | 1.861,70 (214,20)*** | 259,79 (204,01)* | 3.428,45 (371,57)*** |
| (Area - 400) | −2,92 (0,10)*** | −2,94 (0,10)*** | −2,95 (0,10)*** | |
| Bairro A | 1.185,02 (15,39)*** | | | |
| Bairro B | 1.283,46 (15,67)*** | | | |
| Bairro C | 1.409,95 (15,92)*** | | | |
| Bairro D | 1.505,32 (17,16)*** | | | |
| Bairro E | 1.661,76 (18,08)*** | | | |
| Bairro F | 1.857,19 (18,40)*** | | | |
| Bairro G | 2.089,09 (17,23)*** | | | |
| Bairro H | 2.405,77 (16,11)*** | | | |
| Bairro I | 2.770,98 (15,49)*** | | | |
| Bairro J | 2.984,74 (15,38)*** | | | |
| Area | | | | −2,94 (0,10)*** |
| Area (contexto) | | | | −4,45 (0,81)*** |
| Area Verde | | | 3.901,65 (467,16)*** | 3.940,15 (205,04)*** |
| Observations | 500 | 450 | 450 | 450 |
| Log Likelihood | | −2.781,72 | −2.765,12 | −2.758,14 |
| Akaike Inf. Crit. | 6.120,87 | 5.571,44 | 5.540,23 | 5.528,27 |
| Bayesian Inf. Crit. | 6.171,44 | 5.587,87 | 5.560,78 | 5.552,93 |
| *Note:* | | | | *p<0,3; **p<0,2; ***p<0,1 |

Deve-se notar, primeiramente, que os valores estimados pelo modelo de efeitos fixos para os interceptos de cada bairro (coluna 1 da tabela 2) são praticamente os mesmos valores obtidos pela estimação do modelo misto simples, descritos na tabela 3, onde os valores de referência para cada bairro foram obtidos através da soma do intercepto global do modelo misto simples com os interceptos aleatórios do modelo misto simples, que 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 ajusto do modelo, não há valores estimados para os efeitos aleatórios para este bairro.

Tabela 3: Valores dos interceptos para cada bairro.

| A | B | C | D | E | F | G | I | J |
|---|---|---|---|---|---|---|---|---|
| 1.185,5 | 1.284,3 | 1.410,9 | 1.506,9 | 1.663,5 | 1.859 | 2.090,3 | 2.770,8 | 2.984,1 |

Tabela 4: Efeitos randômicos do modelo misto.

| grp | vcov | sdcor |
|---|---|---|
| Bairro | 412.441,97 | 642,22 |
| Residual | 12.123,29 | 110,11 |

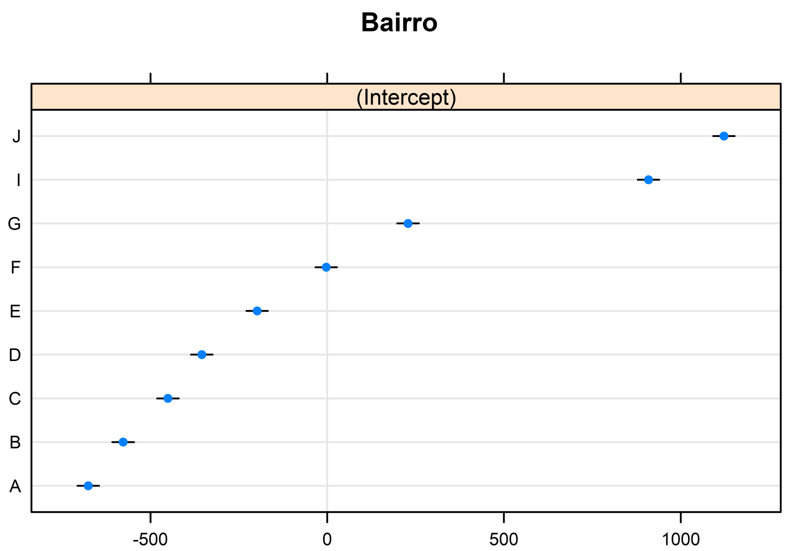


Figura 2: Efeitos aleatórios do modelo.

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 4.

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.

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 ($\beta_{1C}$) 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. Aqui pptou-se, porém, por apresentar desta maneira para se facilitar a compreensão dos modelos.

Deve-se notar a flexibilidade deste tipo de formulação: quando não existe na realidade o efeito esperado pela formulação, o coeficiente resultará insignificante. Diz-se em estatís-

Tabela 5: Análise de Variância.

| | npar | AIC | BIC | logLik | deviance | Chisq | Df | Pr(>Chisq) |
|---|---|---|---|---|---|---|---|---|
| fit_lmer | 4 | 5.581,24 | 5.597,67 | -2.786,62 | 5.573,24 | NA | NA | NA |
| fit_lmer2 | 5 | 5.561,69 | 5.582,24 | -2.775,85 | 5.551,69 | 21,54 | 1 | 0 |
| mundlak | 6 | 5.547,46 | 5.572,12 | -2.767,73 | 5.535,46 | 16,23 | 1 | 0 |

tica que o modelo degenera, ou seja, o modelo com a formulação de Mundlak se degenera 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, mantem-se o modelo. Caso contrário, descarta-se o termo insignificante.

Essa mesma degeneração ocorre com os modelos de efeitos mistos se não houver de fato variabilidade entre os agrupamentos. Fosse o caso da localização por bairros não afetar na formação final de preços, o valor estimado para o desvio-padrão do efeito aleatório $\upsilon$ seria igual a zero, *i. e.* $\hat{\sigma}_\upsilon = 0$ (BATES, 2010, pp. 10–11) situação em que o modelo misto degenera para um modelo de regressão linear ordinária.

Outra maneira de se testar a pertinência da formulação de Mundlak seria através da Análise de Variância. A tabela 5 faz a comparação entre o modelo de efeitos mistos sem variáveis de segundo nível ( primeira linha), com a variável de segundo nível $A_V$ (segunda linha) e com a formulação de Mundlak (terceira linha). Percebe-se que é significante a melhora advinda da adição de um novo parâmetro no segundo modelo, porém não é significante a adoção de um novo parâmetro pela formulação de Mundlakm, o que se nota nos p-valores constantes da última coluna.

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.

Nas Figuras 3 e 4 podem ser vistos os gráficos de densidades para os parâmetros estimados pelos modelos de efeitos mistos simples e com variável de segundo nível, respectivamente. Nota-se que a ausência da variável de segundo nível levou o modelo de efeitos mistos simples a uma má estimação da distribuição da variável $\sigma_\upsilon$ (no gráfico $\sigma_1$), ficando esta longe da normalidade. Com o acréscimo da variável de segundo nível esta variável ficou com magnitude bem menor, ou seja, a introdução da variável de segundo nível reduziu a variação não-explicada pelo modelo.

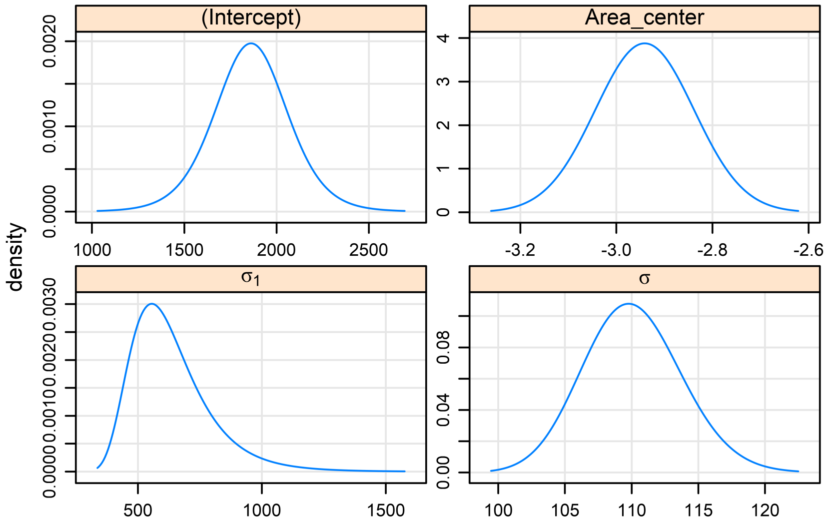


Figura 3: Densidades dos parâmetros do modelo de efeitos mistos simples.

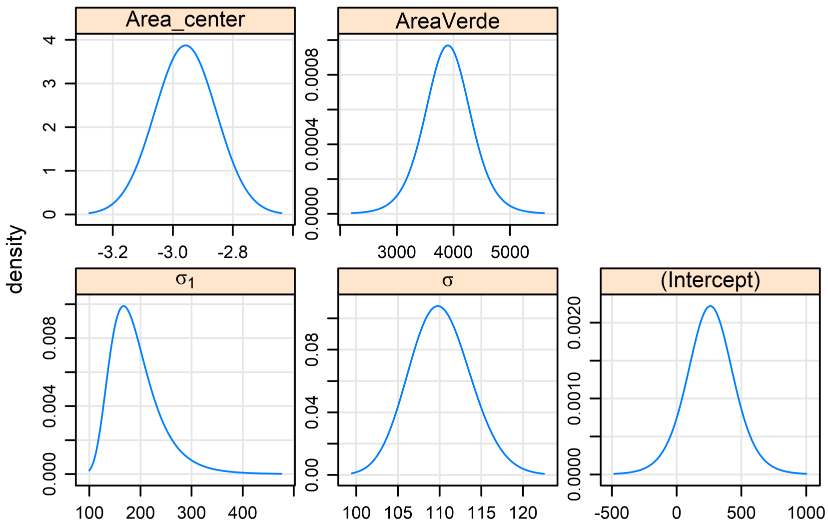


Figura 4: Densidades dos parâmetros do modelo de efeitos mistos com variável de segundo nível.

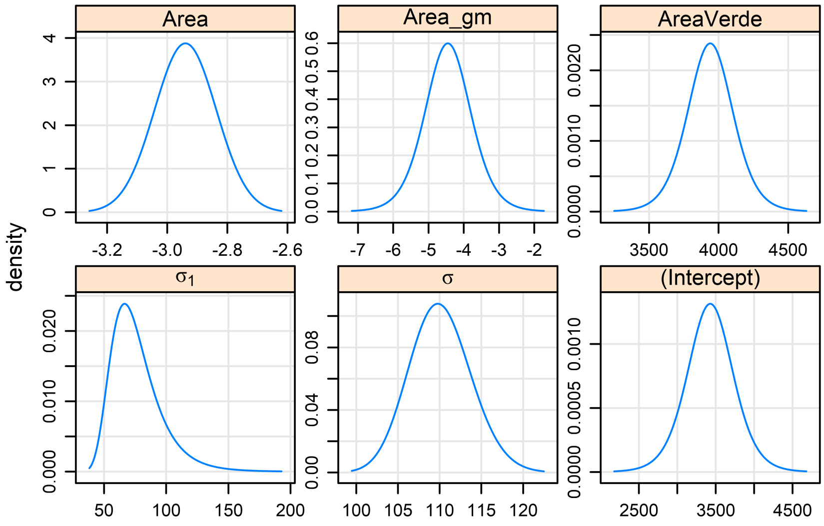


Figura 5: Densidades dos parâmetros do modelo de Mundlak.

### 3.5 Previsão de Valores

Para ilustrar como os modelos mistos podem ser utilizados no contexto de predição, foram elaboradas previsões no bairro H, que havia sido propositalmente excluído no ajuste dos modelos mistos, com os diversos modelos apresentados.

Também foram utilizados o modelo de efeitos fixos e o modelo misto com a variável de segundo nível $A_V$ para a previsão de valores de lote-padrão para os diversos bairros, inclusive para o bairro H, não utilizado para a confecção dos modelos mistos.

#### 3.5.1 Previsão de dados no bairro H

Na Figura 6 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.

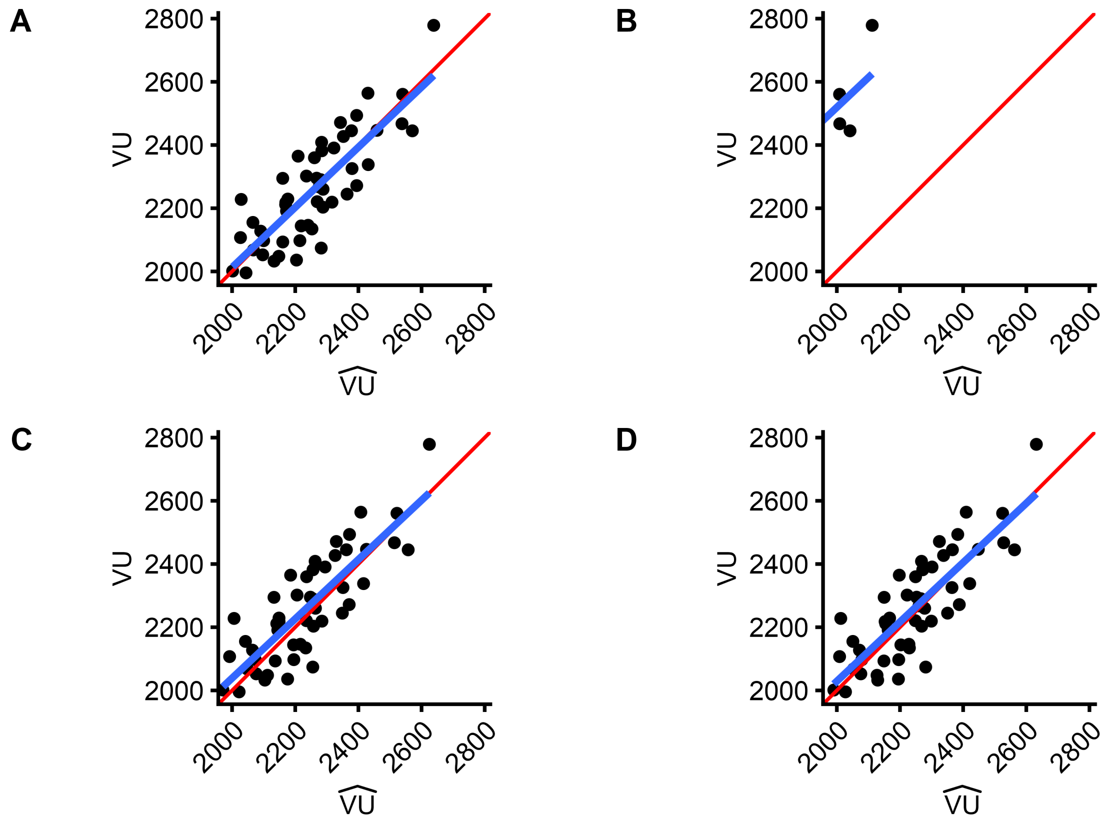


Figura 6: Gráficos de poder de predição para cada modelo.

#### 3.5.2 Previsão de valores para lotes-padrão

Assim como a previsão de valores para os dados simulados para o bairro H, conforme se mostrou no item anterior, é possível prever valores para um lote-padrão nos diferentes bairros.

A tabela abaixo mostra os valores previsto pelos modelos para um lote-padrão de 400 $m^2$ nos diversos bairros:

Tabela 6: Previsões para valores de lote-padrão nos diferentes bairros.

| Bairro | Modelo de efeitos fixos | Modelo misto com variável de segundo nível | Modelo Mundlak |
|---|---|---|---|
| A | 1.185,02 | 1.184,39 | 1.186,85 |
| B | 1.283,46 | 1.283,94 | 1.284,12 |
| C | 1.409,95 | 1.410,66 | 1.408,79 |
| D | 1.505,32 | 1.508,33 | 1.506,38 |
| E | 1.661,76 | 1.666,10 | 1.664,58 |
| F | 1.857,19 | 1.862,03 | 1.861,09 |
| G | 2.089,09 | 2.091,38 | 2.086,18 |
| H | 2.405,77 | 2.383,37 | 2.391,40 |
| I | 2.770,98 | 2.771,21 | 2.774,04 |
| J | 2.984,74 | 2.983,12 | 2.982,68 |

### 3.5.3 Intervalos de predição

No caso dos modelos mistos, os intervalos de predição são calculados separadamente para cada efeito. Na tabela 7 podem ser vistos os intervalos de predição (@80%) para o lote-padrão no bairro G. A primeira linha mostra o intervalo total de predição combinado para os dois efeitos. A segunda linha mostra o intervalo de predição para o efeito aleatório (Bairro) e a terceira linha mostra o intervalo de predição para o efeito fixo.

Tabela 7: Previsão de valores com o modelos de efeitos mistos simples.

| Efeitos | Valor Central | Limite Superior | Limite Inferior | Observações |
|---|---|---|---|---|
| combined | 2.107,83 | 2.399,25 | 1.774,01 | 1 |
| Bairro | 228,07 | 365,87 | 88,94 | 1 |
| fixed | 1.875,53 | 2.182,62 | 1.557,22 | 1 |

A tabela 8 mostra o intervalo de predição para o lote-padrão no bairro H.

Tabela 8: Previsão de valores com o modelos de efeitos mistos com variável de segundo nível.

| Efeitos | Valor Central | Limite Superior | Limite Inferior | Observações |
|---|---|---|---|---|
| combined | 2.383,37 | 2.571,42 | 2.189,59 | 1 |
| Bairro | -6,43 | 149,31 | -136,54 | 1 |
| fixed | 2.380,55 | 2.571,13 | 2.195,37 | 1 |

A tabela 9 mostra o intervalo de predição para o lote-padrão no bairro H, obtido com o modelo de Mundlak.

Tabela 9: Previsão de valores com o modelo de Mundlak.

| Efeitos | Valor Central | Limite Superior | Limite Inferior | Observações |
|---|---|---|---|---|
| combined | 2.391,40 | 2.543,85 | 2.240,62 | 1 |
| Bairro | 2,13 | 136,18 | -148,36 | 1 |
| fixed | 2.392,34 | 2.543,80 | 2.241,66 | 1 |

Para efeitos de comparação, a tabela 10 mostra o intervalo de predição para o bairro H calculado com o modelo de efeitos fixos.

Tabela 10: Intervalo de Predição com o modelo de efeitos fixos

| Efeitos | Valor Central | Limite Superior | Limite Inferior |
|---|---|---|---|
| Fixo | 2.405,77 | 2.546,85 | 2.264,69 |

Nota-se que os intervalos de predição do modelo de efeitos fixos apresentados na tabela 9 e o intervalo de predição total combinado (linha 1) do modelo de efeitos mistos com variável de segundo nível praticamente se equivalem. Deve-se lembrar, porém, que para o modelo de efeitos fixos foram utilizados 10% mais dados e que o modelo de efeitos mistos não utilizou qualquer dado amostral provenienente do bairro H.

# 4 Conclusão

A aplicação da modelagem mista ou hierárquica na Engenharia de Avaliações pode ser feita das mais diversas maneiras, desde a aplicação em avaliações de precisão, até a avaliação em massa para fins tributários, assim como para confecção de índices de preços de imóveis.

Neste trabalho foi mostrado como a Engenharia de Avaliações pode se valer da modelagem hierárquica ou mista para a confecção de PVG's, com a utilização de modelos com interceptos aleatórios, especialmente para estimação de valores para lotes-padrão em agrupamentos não presentes na amostra, através da utilização de variáveis de segundo nível que *expliquem* a variabilidade entre os bairros ou outros agrupamentos. Tais modelos são mais complexos e ao mesmo tempo elegantes, dividindo a variabilidade em diversos níveis, deixando claro ao analista de onde advém a variabilidade dos preços.

Embora a modelagem hierárquica seja considerada mais elegante do que a modelagem de efeitos fixos, deve-se ter em conta que a elaboração de modelos mistos sem variáveis de segundo nível, como é comum encontrar na literatura, não é tão interessante e quase nada agrega a uma melhor explicação do fenômeno estudado. Deve até haver uma melhora na estimação com os modelos mistos caso os dados de alguns agrupamentos estejam em número reduzido, mas o ideal é utilizar as formulações mais complexas da modelagem hierárquica de maneira a explorar ao máximo este tipo de modelagem.

Na análise de dados em seção transversal, como na elaboração de avaliações de precisão ou na elaboração de PVG's, deve ser utilizada, preferencialmente, a formulação de Mundlak, enquanto para dados em painéis, como na confecção de índices de preços de imóveis, deve ser preferencialmente utilizada a formulação REWB.

Na modelagem hierárquica ainda é possível incorporar outras hipóteses úteis, além dos interceptos aleatórios, como também coeficientes aleatórios, o que deve ser tema de outro trabalho.

Outra possibilidade é a modelagem em mais níveis hierárquicos. Não apenas os imóveis podem ser agrupados em bairros, mas também os bairros podem, por sua vez, ser agrupados em macrozonas urbanas, assim como estas podem ser agrupadas em cidades, as cidades em regiões e assim por diante. A execução de modelos tão complexos com efeitos fixos é praticamente inviável.

Outra possibilidade é a modelagem dos dados ao longo do tempo, o que possibilita a sua utilização para a confecção de índices de preços de imóveis.

# 5 Sugestões para trabalhos futuros

# Referências


BATES, D.; MÄCHLER, M.; BOLKER, B.; WALKER, S. Fitting linear mixed-effects models using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1–48, 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.

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. Cambridge University Press.

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.

CLARK, M. Shrinkage in mixed effects models. **Michael Clark**, 2019. Disponível em: <https://m-clark.github.io/posts/2019-05-14-shrinkage-in-mixed-models/>..

DROUBI, L. F. P.; ZILLI, C. A.; HOCHHEIM, N. Centralização e escalonamento de dados amostrais: Prós, contras e aplicação na engenharia de avaliações. In: XX Congresso Brasileiro de Avaliações e Perícias. **Anais...**, 2019. Florianópolis: COBREAP.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2020.