

CENTRALIZAÇÃO E ESCALONAMENTO DE DADOS AMOSTRAIS: PRÓS, CONTRAS E APLICAÇÃO NA ENGENHARIA DE AVALIAÇÕES

RESUMO

É prática comum na Engenharia de Avaliações a adoção de transformações das variáveis do modelo de regressão linear visando a obtenção de melhor ajuste, verificação das hipóteses da inferência clássica, assim como o enquadramento do modelo segundo os graus de fundamentação previstos na NBR 14.653-2. No entanto, apesar de comum entre os estatísticos, não é comum em Engenharia de Avaliações a prática da centralização e/ou escalonamento de variáveis. Este estudo pretende demonstrar as vantagens da utilização destas transformações, visando atingir, entre outros objetivos, uma melhor interpretabilidade do modelo. Para isto, foram elaborados dois estudos de casos com dados de terrenos, em que se buscou, através da centralização dos dados em relação a um imóvel paradigma, dar interpretação ao intercepto do modelo, assim como facilitar a obtenção de fatores de avaliação de acordo com o modelo de regressão obtido. Os resultados mostram que os dados centralizados possibilitaram uma melhor interpretação do modelo, haja vista que o intercepto deste é aproximadamente o valor do metro quadrado do lote paradigma. Nos dois casos também se observou um ganho no grau de fundamentação do modelo e uma redução da colinearidade entre as variáveis no segundo caso, quando usados dados centralizados.

Palavras-Chave: *Centralização de Variáveis; Escalonamento; Avaliação de Imóveis.*

1 INTRODUÇÃO

Em alguns problemas de engenharia de avaliações, embora algumas variáveis apresentem correlação linear com a variável dependente, outras não seguem necessariamente esse modelo. Nesses casos, a real correlação entre a variável independente e a variável dependente poderia ser obtida através da adição de termos polinomiais ao modelo (ZILLI *et al*, 2018). Além disto, possíveis interações entre as variáveis também devem ser consideradas.

Segundo Vieira e Faia (2014), uma das grandes preocupações quanto à análise do efeito interativo sobre uma variável dependente é a presença da multicolinearidade, tornando difícil distinguir os efeitos diretos da variável independente, da variável moderadora e da variável interativa sobre a variável dependente. É recomendável, sempre que possível, fazer um tratamento dos dados.

Há diversas maneiras de se tratar os dados, entre elas pode-se citar a centralização e o escalonamento de variáveis.

A ideia por trás da centralização de variáveis está em subtrair o valor médio ou um valor significativo constante de uma variável independente. Segundo Afshartous e Preston (2011), é conhecido que a centralização da variável pode aumentar a interpretabilidade dos coeficientes de regressão, bem como reduzir a multicolinearidade entre variáveis preditoras de menor e maior ordem.

Já o escalonamento é uma forma de contornar os problemas relacionados à escala, mantendo a informação estatística dos dados. O procedimento consiste em realizar uma transformação sobre o conjunto original dos dados de modo que cada variável apresente média zero e variância unitária.

A recomendação é que o escalonamento só deva ser feito quando as unidades de medidas das características observadas não forem as mesmas.

O presente trabalho objetiva comparar estudos relacionados à avaliação imobiliária com e sem tratamento de variáveis. Esse tratamento é realizado através da centralização de dados e/ou escalonamento de dados.

2 REVISÃO DA LITERATURA

Quando se desenvolve um modelo de regressão, é recomendado que antes seja feito um tratamento dos dados para se eliminar amostras discrepantes, pontos influenciadores e minimizar ruídos. Há diversas maneiras de se tratar os dados, entre elas pode-se citar a centralização e o escalonamento de variáveis.

Em um cenário paramétrico, como o modelo linear, centralizar e escalonar tem o objetivo de reduzir o erro numérico de arredondamento (MATLOFF, 2017). Na sequência serão abordados a centralização e o escalonamento de variáveis.

2.1 CENTRALIZAÇÃO DE VARIÁVEIS

O processo de centralização de variáveis consiste em subtrair o valor médio ou um valor significativo constante, de uma variável independente da amostra. A Equação 01 mostrada abaixo representa esse processo de centralização:

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (01)$$

Em que o \bar{x}_j é o valor da estimativa para a média da característica j .

A literatura estatística concorda que a centralização de variáveis preditoras pode aumentar o sucesso na interpretação dos coeficientes da regressão, bem como reduzir a multicolinearidade entre variáveis preditoras do modelo.

Segundo Luz (2003), do ponto de vista estatístico, a centralização tem como objetivo prevenir que os pontos mais distantes do centro dos dados tenham maior influência que os mais próximos. Dependendo do tipo dos dados e da sua aplicação, a centralização pode ter efeito positivo, negativo ou neutro.

Em avaliação de imóveis, por exemplo, assim que os dados forem centrados em torno da média, a nova matriz de dados terá valores positivos e negativos e, se estivermos falando da variável área, o novo valor da área média para toda a amostra será igual a zero. Do mesmo modo, se estivermos trabalhando com a variável frente, a centrar essa variável em torno da média, a nova frente média será igual a zero.

A decisão de centrar ou não os dados da variável área é totalmente independente da decisão de centrar ou não os dados da variável frente.

Com o objetivo de ilustrar a centralização em torno da média, a imagem abaixo mostra em A os dados antes da centralização e em B após a centralização.

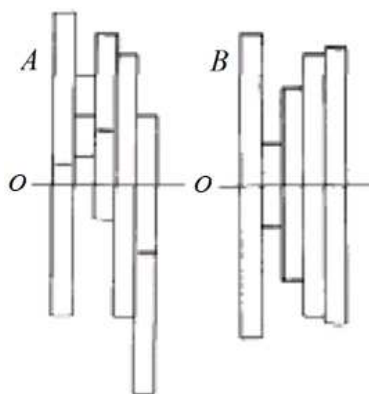


Figura 01 - Dados antes e depois da centralização. Fonte: Luz (2003).

Pode-se verificar que as posições relativas das variáveis preditoras da amostra não foram alteradas. Nota-se apenas a mudança da origem do novo sistema de coordenadas para o centróide dos dados.

No estudo de caso apresentado no presente trabalho optou-se por realizar a centralização em torno das preditoras do imóvel paradigma.

2.2 ESCALONAMENTO DE VARIÁVEIS

Normalmente as variáveis de determinada amostra são observadas em escalas diferentes entre si, e neste caso, é conveniente realizar um processo chamado de escalonamento de variáveis. O tipo mais comum de escalonamento é o de variância (*Variance Scaling*), muitas vezes chamado de padronização.

Segundo Luz (2003), uma maneira de resolver os problemas de escala, mantendo a informação estatística dos dados, é realizar uma transformação sobre o conjunto original dos dados de modo que cada variável apresente média zero e variância igual a um (autoescalonamento). Esta transformação expressa cada observação como o número de desvios padrões da média.

O escalonamento pode ser realizado pela equação abaixo, que torna a média zero e a variância unitária.

$$x_{ij(esc)} = \frac{x_{ij} - \bar{x}_j}{s(x_j)} \quad (02)$$

Nessa equação, \bar{x}_j e $s(x_j)$ são, respectivamente, o valor da estimativa para a média e o desvio padrão da característica j .

Conforme Vicini (2005), para que seja possível escalonar as variáveis, é necessário ter-se uma matriz de dados com p variáveis ($j = 1, 2, \dots, p$) e n objetos ($i = 1, 2, \dots, n$). Sendo que, na matriz de dados, o valor do i -ésimo objeto e j -ésima variável será denotado por x_{ij} , no qual o valor escalonado será representado por $x_{ij(esc)}$, onde as variáveis escalonadas terão média 0 e variância constante 1.

Nesse sentido, ao se realizar o escalonamento, se está ajustando o conjunto de dados de modo a igualar a variância de cada variável e, desta forma, igualando a influência que cada variável tem sobre o conjunto de dados.

O objetivo do escalonamento dos dados originais é expressar, segundo Luz (2003), cada observação em termos de variações inerentes ao sistema.

Em seguida é exemplificada a importância desse pré-tratamento dos dados analisando o comportamento de algumas variáveis preditoras do mercado imobiliário.

Caso se considere, por exemplo, para o estudo de caso mostrado na sequência, que a variável área para os dados pesquisados apresenta área mínima de 360 m² e área máxima de 1200 m², tem-se uma amplitude de 840 m². Já a variável frente, por exemplo, apresenta frente mínima de 10 metros e frente máxima de 22 metros, cuja amplitude é de 12 metros.

Nesse caso, é simples perceber que uma diferença de 6 metros na variável frente entre duas amostras corresponde a uma variação de 50% em relação à amplitude. Entretanto, uma variação desta mesma ordem de grandeza da variável área é praticamente desprezível, na ordem de ~ 0,7%.

Importante mencionar também que o valor numérico entre as variáveis difere, de maneira que ao se realizar uma comparação direta entre elas, seríamos levados a uma ponderação maior das variáveis com maior valor numérico.

Diferentemente de centrar os dados, escaloná-los irá alterar as posições relativas de todos os dados, entretanto não irá modificar a posição do centróide do conjunto de dados, como se pode acompanhar pela figura 02 - C abaixo:

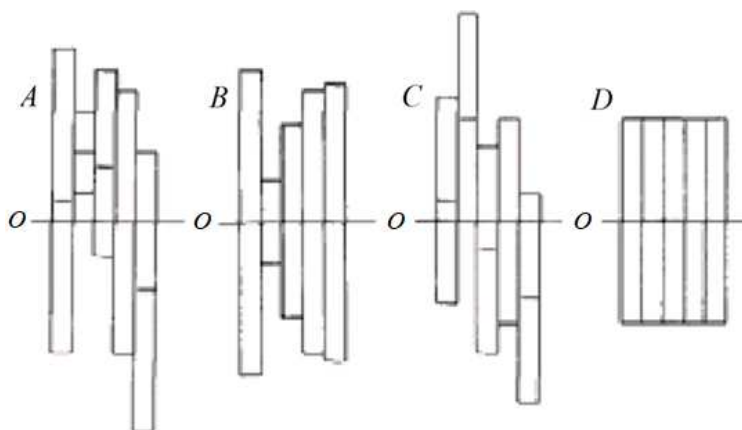


Figura 02 - Dados em diversas situações de tratamento. Fonte: Luz (2003).

Na figura 02 pode-se observar que os dados de cada variável em estudo estão representados por um retângulo de variância, na posição vertical e seu centro.

Em A da figura 02 observa-se os dados sem nenhuma espécie de tratamento, ao passo que em B vê-se os dados após ser efetuada somente a centralização em torno da média. Já em C percebe-se a realização somente do escalonamento (todos os retângulos passam a ter o mesmo comprimento). Por fim, em D, observa-se que os dados passaram por centralização e por escalonamento.

3 ESTUDOS DE CASOS

3.1 LOTEAMENTO COM TOPOGRAFIA VARIÁVEL

Neste estudo de caso foi inicialmente ajustado um modelo sem quaisquer transformação de variáveis, porém com a adição de termos polinomiais para a modelagem da variável inclinação, visando averiguar possíveis diferentes comportamentos em relação à inclinação positiva (aclive) e negativa (declive), o que não seria possível apenas com o termo linear. Foram então adicionados os termos quadrático e cúbico da variável. No entanto, neste primeiro modelo o termo quadrático não apresentou significância, segundo os critérios da NBR14.653-02.

Infelizmente, é prática entre os Engenheiros de Avaliações descartar do modelo os termos que não possuem a significância desejada. Isto, porém, é controverso. Entre os estatísticos, é sabido que, se o objetivo principal do modelo é o de predição, ou seja, obter boas estimativas, o ideal é manter as variáveis que possuem p-valores até um *cutoff* mais alto, entre 0,25 e 0,35 (MATLOFF 2017, 349).

Efetuiu-se, então, a padronização da variável, ou seja, ela foi transformada de maneira que sua média se igualou a zero e seu desvio-padrão se igualou a 1, obtendo-se então um segundo modelo, com todos os termos significantes. Além disto, as variáveis frente e profundidade foram centralizadas em relação às

dimensões do lote paradigma, isto é, as variáveis frente e profundidade sofreram a subtração de 15 e 30m, respectivamente.

Pode-se mostrar que ambos os modelos são equivalentes, isto é, produzem as mesmas estimativas. No entanto, de acordo com a NBR 14.653-02, apenas o segundo modelo possui grau de fundamentação. As estatísticas básicas dos dois modelos podem ser vistas na tabela 02.

3.1.1 Dados

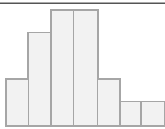
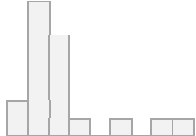


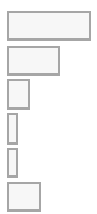
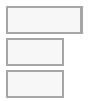
Para este estudo de caso em específico foram utilizados vinte dados de mercado de terrenos, conforme mostrado na tabela 03.

Quadro Resumo dos Dados:

Loteamento

Dimensões 20 x 11

Duplicates: 0

N	Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
1	valor [numeric]	Mean (sd) : 25415.1 (8065.6) min < med < max: 12500 < 25000 < 44122 IQR (CV) : 10167.7 (0.3)	16 distinct values		0 (0%)
2	area [numeric]	Mean (sd) : 562.2 (214) min < med < max: 360 < 493 < 1200 IQR (CV) : 113.8 (0.4)	16 distinct values		0 (0%)
3	tipo [factor]	1. venda 2. oferta	11 (55.0%) 9 (45.0%)		0 (0%)
4	frente [numeric]	Mean (sd) : 15.8 (3.2) min < med < max: 10 < 15 < 22 IQR (CV) : 4.2 (0.2)	10 : 1 (5.0%) 12 : 2 (10.0%) 13 : 2 (10.0%) 14 : 1 (5.0%) 15 : 5 (25.0%) 16 : 1 (5.0%) 17 : 2 (10.0%) 18 : 2 (10.0%) 20 : 3 (15.0%) 22 : 1 (5.0%)		0 (0%)
5	profundidade [numeric]	Mean (sd) : 39 (11.3) min < med < max: 30 < 35 < 60 IQR (CV) : 12.5 (0.3)	30 : 8 (40.0%) 35 : 5 (25.0%) 40 : 2 (10.0%) 50 : 1 (5.0%) 55 : 1 (5.0%) 60 : 3 (15.0%)		0 (0%)
6	topografia [factor]	1. plano 2. aclave 3. declive	8 (40.0%) 6 (30.0%) 6 (30.0%)		0 (0%)

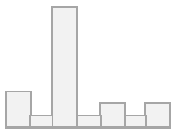
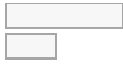
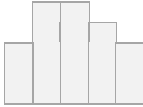
7	inclinacao [numeric]	Mean (sd) : 0 (0.1) min < med < max: -0.2 < 0 < 0.2 IQR (CV) : 0.1 (9.5)	11 distinct values		0 (0%)
8	pedologia [factor]	1. seco 2. pantanoso	14 (70.0%) 6 (30.0%)		0 (0%)
9	VU [numeric]	Mean (sd) : 45.2 (13.3) min < med < max: 23.8 < 47.2 < 67 IQR (CV) : 21 (0.3)	20 distinct values		0 (0%)

Tabela 01 - Estatísticas descritivas da amostra.

3.1.2 Modelo Inicial

De posse de todos os dados, sem qualquer transformação exceto a da variável inclinação, por construção, foi primeiramente ajustado um primeiro modelo, apenas para o saneamento da amostra, em que foram utilizadas as variáveis frente e profundidade em detrimento da variável área.

Os gráficos diagnósticos deste primeiro modelo podem ser vistos na Figura 3. Como se pode ver nesta figura, os pontos 7 e 19 encontram-se bem afastados da média e foram excluídos do modelo final.

Para esse estudo de caso, considerou-se o paradigma da região como sendo um terreno plano e seco, com 15 m de frente e 30 m de profundidade.

3.1.3 Modelo Final

Uma vez obtido o modelo final saneado, então, foi ajustado outro modelo, onde se adotou a centralização das variáveis frente e profundidade, de acordo com o lote paradigma. Já a variável inclinação, por possuir os termos quadrático e cúbico, com vias de reduzir a multicolinearidade, foi centralizada e escalonada, de maneira que a nova variável inclinação tem média zero e desvio padrão igual a 1.

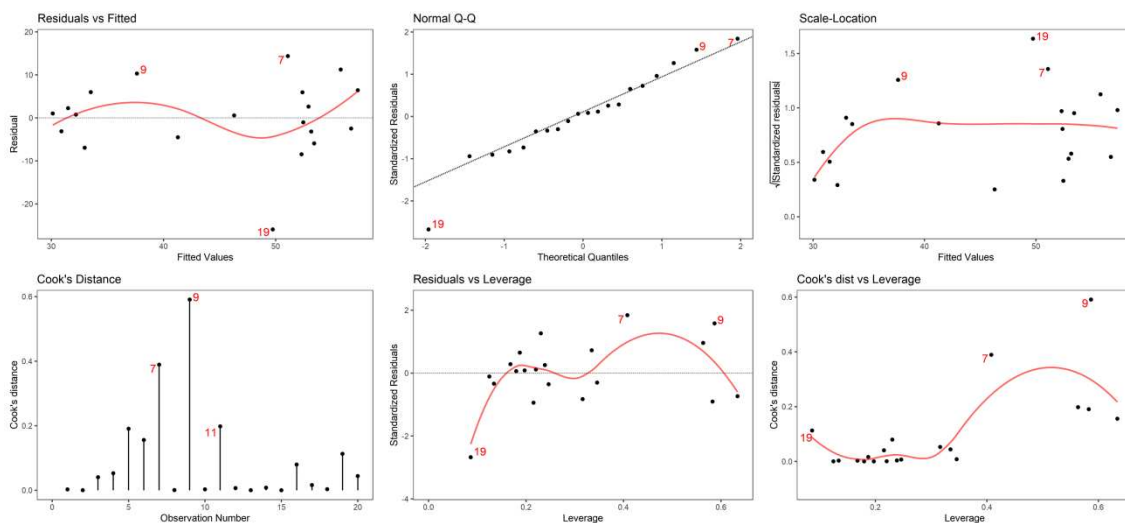


Figura 03 - Modelo com todos os dados.

Os dois modelos são correspondentes entre si, produzem as mesmas estimativas, porém apenas o modelo com as variáveis centralizadas e escalonadas conforme explicitado possui Grau I de especificação pela NBR 14.653-02 (2011), conforme se pode notar na Tabela 02.

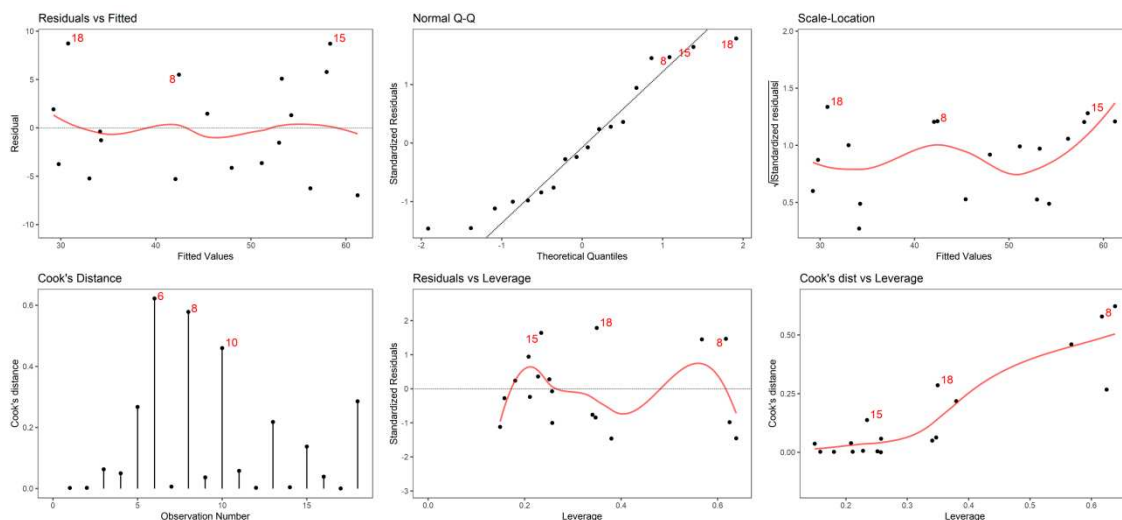


Figura 04 - Modelo saneado.

A Tabela 03 mostra os dados amostrais, com o acréscimo dos valores ajustados. A Figura 05 mostra o gráfico do poder de predição do modelo.

Dependent Variable:		
	VU	
	(1)	(2)
Constant	44,813 (9,368) t = 4,784 p = 0,0005***	54,265 (2,584) t = 21,000 p = 0,000***
frente	0,993 (0,508) t = 1,954 p = 0,075**	0,997 (0,508) t = 1,960 p = 0,074**
profundidade	-0,179 (0,146) t = -1,229 p = 0,243*	-0,178 (0,146) t = -1,220 p = 0,246*
I(inclinacao2)	-173,083 (170,248) t = -1,017 p = 0,330	-1,617 (1,295) t = -1,249 p = 0,236*
I(inclinacao3)	-1,071,862 (771,291) t = -1,390 p = 0,190**	-0,889 (0,533) t = -1,667 p = 0,122**
pedologiapantanoso	-21,201 (3,727) t = -5,689 p = 0,0002***	-21,111 (3,701) t = -5,703 p = 0,0001***
R2	0,825	0,825
Adjusted R2	0,752	0,752
Residual Std. Error (df = 12)	6,054	6,061
F Statistic (df = 5; 12)	11,323***	11,291***
Note:	*p<0,3; **p<0,2; ***p<0,1	

Tabela 02 - Comparação dos modelos sem (01) e com (02) centralização e escalonamento.

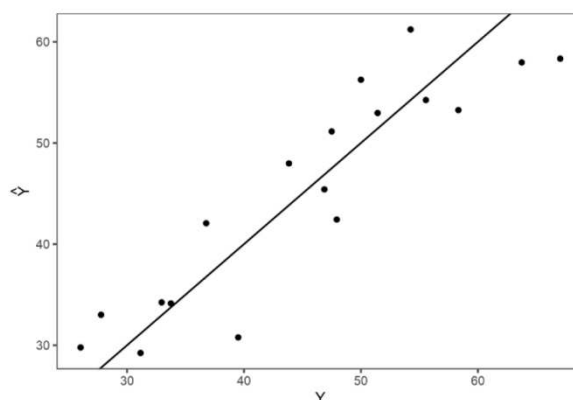


Figura 05 - Poder de predição do modelo.

Valor	Área	Tipo	Frente	Profun.	Topo	Inclinação	Pedologia	VU	VU*
25.000,00	450	venda	15	30	plano	0,00	seco	55,56	54,25
30.000,00	525	oferta	15	35	active	0,05	seco	51,43	52,97
28.500,00	650	venda	13	50	declive	-0,08	seco	43,85	47,98
29.500,00	1.020	oferta	17	60	plano	0,00	pantanososo	26,03	29,78
19.000,00	360	oferta	12	30	declive	-0,15	seco	47,50	51,14
44.122,04	1.200	venda	20	60	active	0,18	seco	36,77	42,06
40.000,00	550	oferta	10	55	declive	-0,10	seco	65,45	44,03
18.000,00	520	oferta	13	40	declive	-0,02	pantanososo	31,15	29,23
21.570,77	450	venda	15	30	active	0,18	seco	47,94	42,43
23.000,00	414	oferta	18	30	active	0,07	seco	50,00	56,26
25.500,00	400	venda	20	35	declive	-0,14	seco	63,75	57,97
12.500,00	450	venda	15	30	declive	-0,02	pantanososo	27,78	33,01
19.609,79	595	venda	17	35	plano	0,00	pantanososo	32,96	34,24
30.500,00	506	oferta	22	30	plano	0,00	seco	54,25	61,22
25.000,00	480	oferta	12	40	active	0,12	seco	46,88	45,41
29.500,00	440	venda	20	35	plano	0,00	seco	67,05	58,34
24.500,00	420	venda	14	30	plano	0,00	seco	58,33	53,25
18.000,00	480	oferta	16	30	plano	0,00	pantanososo	33,75	34,13
12.500,00	525	venda	15	35	active	0,10	seco	23,81	50,84
32.000,00	810	venda	18	60	plano	0,00	pantanososo	39,51	30,78

Tabela 03 - Dados do modelo com valores ajustados.

3.1.4 Coerência do Modelo

O modelo é coerente, conforme se pode notar nas estimativas abaixo:

Segundo o modelo, o lote paradigma vale R\$ 54,25/m², o que é muito próximo do intercepto do modelo final (R\$ 54,27/m²). Apenas para efeito de comparação, o mesmo lote paradigma avaliado de acordo com o modelo original, representado por (1) na tabela 02, vale R\$ 54,33/m².

Um lote com as mesmas características do lote paradigma, porém com 5 m a mais de frente, segundo o modelo, vale R\$ 59,23/m² ($\approx 54,27 + 0,997 \times 5m$). No modelo original, vale R\$ 59,30/m².

Um lote com as mesmas características do lote paradigma, porém com 45 m de profundidade, segundo o modelo, vale R\$ 51,57/m² ($\approx 54,27 - 0,178 \times 15m$). No modelo original, vale R\$ 51,64/m².

Um lote com as mesmas características do lote paradigma, porém com 20 m de frente e 45 m de profundidade, segundo o modelo, vale R\$ 56,56/m². No modelo original, vale R\$ 56,61/m².

Um lote com as mesmas características do lote paradigma, porém com declive de 10%, segundo o modelo, vale R\$ 53,47/m². No modelo original, vale R\$ 53,67/m² ($\approx 54,27 - 173,08 \times (-0,10)^2 - 1.071,86 \times (-0,10)^3 = 54,27 - 1,73 + 1,07$).

Um lote com as mesmas características do lote paradigma, porém com aclave de 10%, segundo o modelo, vale R\$ 51,73/m². No modelo original, vale R\$ 51,53/m² ($\approx 54,27 - 173,08 \times (0,10)^2 - 1.071,86 \times (0,10)^3 = 54,27 - 1,73 - 1,07$).

Finalmente, um lote com as mesmas características do lote paradigma, porém em terreno pantanoso, segundo o modelo, vale R\$ 33,14/m² ($\approx 54,27 - 21,11$). No modelo original, vale R\$ 33,13/m².

A Tabela 04 mostrada abaixo resume a influência de cada variável na composição do valor unitário total:

Terreno	V_0	Frente	Profundidade	Inclinação	Pedologia Pantanoso	VU
Paradigma	54,27	0	0	0	0	54,27
(1) +05 m frente	54,27	4,98	0	0	0	59,25
(2) +15 m fundo	54,27	0	-2,68	0	0	51,59
(3) = (1) + (2)	54,27	4,98	-2,68	0	0	56,60
(4) 10% aclave	54,27	0	0	-2,80	0	51,47
(5) 10% declive	54,27	0	0	-0,66	0	53,61
(6) Pantanoso	54,27	0	0	0	-21,11	33,16

Tabela 04 - Influência de cada variável na composição do valor unitário.

Pode-se então, baseado nestes modelos, adaptar, para posteriormente serem utilizados através do método de avaliação por fatores, o seguinte:

a. Fator frente:

$$C_f = 1 + \frac{0,997}{54,265} * (F_p - F_r)$$

b. Fator profundidade:

$$C_p = 1 - \frac{0,178}{54,265} * (P_e - P_r)$$

c. Fator inclinação:

$$C_t = 1 - \frac{173,1 \cdot (i_r)^2 + 1071,9 \cdot (i_r)^3}{54,265}$$

d. Fator pedologia pantanoso:

$$C_{pp} = 1 - \frac{21,11}{54,265} = 0,61$$

Terreno	C_f	C_p	C_t	C_{pp}	$C_{aval.total}$
Paradigma	1,00	1,00	1,00	1,00	1,00
(1) +05 m frente	1,09	1,00	1,00	1,00	1,09
(2) +15 m fundo	1,00	0,95	1,00	1,00	0,95
(3) = (1) + (2)	1,09	0,95	1,00	1,00	1,04
(4) 10% aclave	1,00	1,00	0,95	1,00	0,95
(5) 10% declive	1,00	1,00	0,99	1,00	0,99
(6) Pantanoso	1,00	1,00	1,00	0,61	0,61

Tabela 05 - Tabela de fatores.

É interessante notar que, para os termos polinomiais da variável inclinação, para o estabelecimento dos fatores, foram utilizados os coeficientes do primeiro modelo. Isto só foi possível porque a média da variável inclinação já era muito próxima de zero. Desta maneira, a diferença entre a adoção dos coeficientes do segundo modelo e os coeficientes do primeiro modelo, na elaboração dos fatores, é irrisória. No entanto, a facilidade de obtenção dos fatores com o primeiro modelo é muito maior, e o resultado muito mais elegante, razões da adoção, então, dos coeficientes do primeiro modelo.

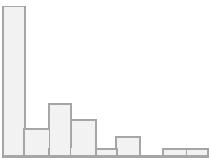
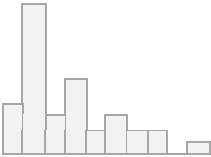
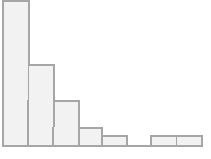
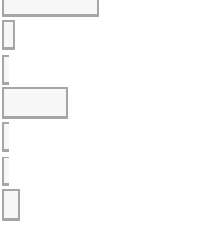

3.2 LOTEAMENTO COM DIFERENTES REGRAS DE INCORPORAÇÃO

3.2.1 Dados

Para este estudo de caso foram aplicados dados de mercado do bairro Jurerê Internacional (anexo A), coletados em Florianópolis no mês de julho de 2017.

Neste caso, o lote paradigma é um terreno com 450 m² com 15 m de testada. As variáveis AREA e TESTADA foram centralizadas nestes valores. Já a variável DIST_MAR foi centralizada a 500m do mar, valor próximo da média amostral.

Além das variáveis numéricas citadas, foi utilizada a variável PAVIMENTOS (dicotômica), que representa o número de pavimentos viável estabelecido para o local do referido imóvel (2 ou 4).

N	Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
1	VALOR TOTAL [numeric]	Mean (sd) : 1585771.4 (1017285.9) min < med < max: 650000 < 1100000 < 4800000 IQR (CV) : 1180000 (0.6)	25 distinct values		0 (0%)
2	VU [numeric]	Mean (sd) : 2606.9 (1182.2) min < med < max: 1359.4 < 2111.1 < 6000 IQR (CV) : 1904.1 (0.5)	30 distinct values		0 (0%)
3	AREA [numeric]	Mean (sd) : 585.8 (177.2) min < med < max: 450 < 582 < 1200 IQR (CV) : 195.5 (0.3)	16 distinct values		0 (0%)
4	TESTADA [numeric]	Mean (sd) : 18.5 (4.5) min < med < max: 15 < 17 < 30 IQR (CV) : 5 (0.2)	15.00 : 16 (45.7%) 17.00 : 2 (5.7%) 18.00 : 1 (2.9%) 20.00 : 11 (31.4%) 21.70 : 1 (2.9%) 25.00 : 1 (2.9%) 30.00 : 3 (8.6%)		0 (0%)
5	ESQUINA [numeric]	Min : 0 Mean : 0.3 Max : 1	0 : 26 (74.3%) 1 : 9 (25.7%)		0 (0%)

6	FRENTES [numeric]	Mean (sd) : 1.3 (0.5) min < med < max: 1 < 1 < 3 IQR (CV) : 0.5 (0.4)	1 : 26 (74.3%) 2 : 8 (22.9%) 3 : 1 (2.9%)		0 (0%)
7	DIST_MAR [numeric]	Mean (sd) : 484.9 (306.6) min < med < max: 33 < 442 < 1094 IQR (CV) : 532.5 (0.6)	34 distinct values		0 (0%)
8	ZONEA MENTO [character]	1. AMC-4.5 2. ARM-4.5 3. ARP-2.4	2 (5.7%) 1 (2.9%) 32 (91.4%)		0 (0%)
9	RUA_PAVI MENTADA [factor]	1. 2 2. 4	32 (91.4%) 3 (8.6%)		0 (0%)
10	CONDOMI NIO_FECH ADO [numeric]	Min : 0 Mean : 0 Max : 1	0 : 34 (97.1%) 1 : 1 (2.9%)		0 (0%)
11	PAVIMEN TO [character]	1. SIM	35 (100.0%)		0 (0%)
12	TOPOGRA FIA [character]	1. PLANO	35 (100.0%)		0 (0%)

Tabela 06 - Estatísticas descritivas da amostra (Jurerê Internacional).

3.2.2 Modelos

As principais estatísticas dos modelos com e sem centralização podem ser vistas na tabela 07. Mais uma vez, os dois modelos se equivalem. Desta maneira, novamente conseguiu-se dar interpretação ao intercepto do modelo (vide tabela 07). Outra vantagem da centralização é a redução da multicolinearidade do modelo. No primeiro modelo, sem centralização, as variáveis AREA e TESTADA apresentam maior grau de colinearidade do que no modelo com as variáveis centralizadas, usando o critério do fator de inflacionamento da variância.

Dependent variable:		
	VU	
	(1)	(2)
Constant	1.644,436 (686,599) t = 2,395 p = 0,024***	2.065,372 (118,867) t = 17,375 p = 0,000***
I (AREA2)	-0,001 (0,001) t = -1,308 p = 0,202*	-0,002 (0,001) t = -1,687 p = 0,103**
TESTADA	111,239 (45,684) t = 2,435 p = 0,022***	100,510 (32,130) t = 3,128 p = 0,004***
DIST_MAR	-2,059 (0,289) t = -7,131 p = 0,00000***	-2,088 (0,283) t = -7,373 p = 0,00000***

PAVIMENTOS4	1.891,938 (296,639) t = 6,378 p = 0,00000***	1.912,312 (291,943) t = 6,550 p = 0,00000***
Observations	34	34
R2	0,833	0,839
Adjusted R2	0,810	0,817
Residual Std. Error (df = 29)	453,253	445,106
F Statistic (df = 4; 29)	36,151***	37,754***
Note:	*p<0,3; **p<0,2; ***p<0,1	

Tabela 07 - Comparação dos modelos sem (01) e com (02) centralização e escalonamento. Jurerê Internacional (2017).

MODELO (1)				
##	I (AREA^2)	TESTADA	DIST_MAR	PAVIMENTOS
##	6.242149	6.723442	1.219102	1.171601
MODELO (2)				
##	I (AREA^2)	TESTADA	DIST_MAR	PAVIMENTOS
##	3.121131	3.448637	1.216489	1.176720

Tabela 08 - fatores de inflacionamento da variância.

Para este caso em específico, apesar dos altos valores dos fatores, a multicolineariedade não era um problema concreto, mas é significativa a diminuição da inflação da variância com o modelo centralizado. Isto pode ser importante em outras situações. No gráfico da figura 06 abaixo se pode verificar o poder de predição do modelo de Jurerê Internacional.

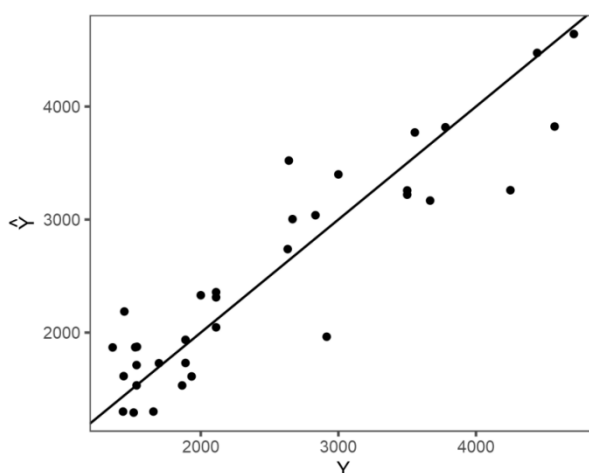


Figura 06 - Poder de predição do modelo (Jurerê Internacional).

Analogamente ao primeiro caso, com o auxílio do modelo foram ajustados os seguintes fatores:

a. Fator Área:

$$C_A = 1 - \frac{2,014 \cdot 10^{-3}}{2.065,37} * (A - A_r)^2$$

b. Fator Testada:

$$C_t = 1 + \frac{100,51}{2.065,37} * (F_p - F_r)$$

c. Fator Distância ao Mar:

$$C_{dm} = 1 - \frac{2,09}{2.065,37} * (D - D_r)$$

d. Fator Pavimentos:

$$C_p = 1 + \frac{1.912,31}{2.065,37} = 1,925$$

A tabela 08 mostra a avaliação pelos fatores obtidos de alguns imóveis. Nesta tabela pode ser observada a coerência do modelo.

Terreno	C_A	C_t	C_{dm}	C_p	$C_{aval.total}$
(1) Paradigma	1,00	1,00	1,00	1,00	1,00
(2) +100m2 área	0,99	1,00	1,00	1,00	0,99
(3) +5m frente	1,00	1,24	1,00	1,00	1,24
(4) =(2)+(3)	0,99	1,24	1,00	1,00	1,23
(5) 33m do mar	1,00	1,00	1,47	1,00	1,47
(6) viab. 4 pav.	1,00	1,00	1,00	1,925	1,925

Tabela 09 - Tabela de fatores (Jurerê Internacional).

4 CONSIDERAÇÕES FINAIS

O modelo com os dados centralizados possibilitou uma melhor interpretação do modelo, haja vista que o intercepto do modelo é aproximadamente o valor do metro quadrado do lote paradigma.

A centralização e escalonamento da variável inclinação no primeiro estudo de caso possibilitou o enquadramento do modelo no Grau I de fundamentação da NBR 14.653-02.

O cálculo das estimativas através do método científico (modelo de regressão linear), possibilitou o correto entendimento da influência da variável inclinação que, neste caso, teve comportamento diverso ao do esperado normalmente. Presume-se que a variável inclinação deprecie menos quando em aclave do que em situação de declive, porém, através da análise do modelo de regressão linear, o comportamento foi o contrário do esperado. Cabe salientar que, caso o coeficiente do termo cúbico tivesse resultado positivo, a variável inclinação teria comportamento similar ao comportamento esperado em situações mais corriqueiras.

Com respeito ao modelo de Jurerê Internacional, mostrou-se que o método também se aplica em dados que apresentam outras características, sendo facilmente aplicável na elaboração de plantas de valores genéricos e no cálculo dos coeficientes de avaliação usados tradicionalmente pelas prefeituras para cálculo do valor base para cobrança do Imposto Predial e Territorial Urbano (IPTU).

Observou-se também, neste estudo de caso, uma redução do grau de colinearidade entre as variáveis. Nos dois casos estudados houve um ganho no grau

de fundamentação do modelo pelo aumento da significação de alguns regressores após a centralização dos dados.

Por fim, não se pode deixar de mencionar a dificuldade para a estimativa de valores e obtenção dos fatores ao se utilizar os coeficientes do modelo com dados centralizados e escalonados, como no primeiro estudo de caso, onde optou-se pelo uso dos coeficientes originais para a variável inclinação, por uma melhor interpretabilidade do efeito desta variável.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2:** Avaliação de Bens. Parte 2: Imóveis Urbanos. Rio de Janeiro, 2011.

AFSHARTOUS, David e PRESTON Richard. **Key results of interaction models with centering.** Journal of Statistics Education v. 19 n. 3, 2011.

LUZ, Elaine Rocha. **Predição de propriedades de gasolinas usando espectroscopia FTIR e regressão por mínimos quadrados parciais.** Dissertação (Mestrado em Química) - Programa de Pós-Graduação em Química, PUCRJ, Rio de Janeiro, 2003.

MANLY, B. F. J. **Multivariate statistical methods:** a primer. London: Chapman and Hall, 1986.

MATLOFF, N. **Statistical regression and classification:** From Linear Models to Machine Learning. Boca Raton, Florida: Chapman & Hall, 2017

VICINI, Lorena. **Análise multivariada:** da teoria à prática. Santa Maria: Editora UFSM, 2005.

VIEIRA, Valter Afonso e FAIA, Valter da Silva. **Efeitos moderadores duplos e triplos na análise de regressão.** In XXXVIII Encontro da AMPAD, 2014.

ZILLI, C.A.; DROUBI, L.F.P.; HOCHHEIM, N. **Regressão polinomial e redes neurais artificiais: um estudo de caso.** In VIII Simpósio da SOBREA, 2018.

ANEXO A – DADOS DE MERCADO DE JURERÊ INTERNACIONAL

NOME	VALOR TOTAL	VU	AREA	TESTADA	ESQUINA	N_DE FRENTES	DIST-MAR	ZONEAMENTO	PAVIMENTOS	CONDOMINIO	PAVIMENTO	TOPOGRAFIA
JUR_01	2100000,00	3500,00	600,00	20,00	0	1	166,00	ARP-2.4	2	0	SIM	PLANO
JUR_02	690000,00	1533,33	450,00	15,00	0	1	755,00	ARP-2.4	2	0	SIM	PLANO
JUR_03	850000,00	1888,89	450,00	15,00	0	1	660,00	ARP-2.4	2	0	SIM	PLANO
JUR_04	4800000,00	4571,43	1050,00	30,00	0	1	33,00	ARP-2.4	2	0	SIM	PLANO
JUR_05	900000,00	2000,00	450,00	15,00	0	1	373,00	ARP-2.4	2	0	SIM	PLANO
JUR_06	880000,00	1512,03	582,00	20,00	1	2	1094,00	ARP-2.4	2	0	SIM	PLANO
JUR_07	950000,00	2111,11	450,00	15,00	0	1	360,00	ARP-2.4	2	0	SIM	PLANO
JUR_08	3200000,00	2666,67	1200,00	30,00	1	3	230,00	ARP-2.4	2	0	SIM	PLANO
JUR_09	850000,00	1888,89	450,00	15,00	0	1	562,00	ARP-2.4	2	0	SIM	PLANO
JUR_10	1100000,00	1864,41	590,00	20,00	0	1	977,00	ARP-2.4	2	0	SIM	PLANO
JUR_11	2550000,00	4250,00	600,00	20,00	0	1	147,00	ARP-2.4	2	0	SIM	PLANO
JUR_12	1700000,00	2833,33	600,00	20,00	0	1	253,00	ARP-2.4	2	0	SIM	PLANO
JUR_13	950000,00	1439,39	660,00	20,00	1	2	914,00	ARP-2.4	2	0	SIM	PLANO
JUR_14	2200000,00	3666,67	600,00	20,00	1	2	191,00	ARP-2.4	2	0	SIM	PLANO
JUR_15	1980000,00	2640,00	750,00	25,00	1	2	197,00	ARP-2.4	2	0	SIM	PLANO
JUR_16	650000,00	1434,88	453,00	15,00	0	1	866,00	ARP-2.4	2	0	SIM	PLANO
JUR_17	750000,00	1655,63	453,00	15,00	0	1	866,00	ARP-2.4	2	0	SIM	PLANO
JUR_18	950000,00	2111,11	450,00	15,00	0	1	382,00	ARP-2.4	2	0	SIM	PLANO
JUR_19	780000,00	1523,44	512,00	17,00	1	2	686,00	ARP-2.4	2	0	SIM	PLANO
JUR_20	870000,00	1933,33	450,00	15,00	1	2	717,00	ARP-2.4	2	0	SIM	PLANO
JUR_21	1300000,00	2631,58	494,00	17,00	1	2	272,00	ARP-2.4	2	0	SIM	PLANO
JUR_22	1700000,00	3777,78	450,00	15,00	0	1	577,00	ARM-4.5	4	0	SIM	PLANO
JUR_23	650000,00	1444,44	450,00	15,00	0	1	442,00	ARP-2.4	2	0	SIM	PLANO
JUR_24	2000000,00	4444,44	450,00	15,00	0	1	262,00	AMC-4.5	4	0	SIM	PLANO
JUR_25	690000,00	1533,33	450,00	15,00	0	1	669,00	ARP-2.4	2	0	SIM	PLANO
JUR_26	2120000,00	4711,11	450,00	15,00	0	1	182,00	AMC-4.5	4	0	SIM	PLANO
JUR_27	4200000,00	6000,00	700,00	15,00	0	1	54,00	ARP-2.4	2	0	SIM	PLANO
JUR_28	2100000,00	3500,00	600,00	20,00	0	1	148,00	ARP-2.4	2	0	SIM	PLANO
JUR_29	1950000,00	2914,80	669,00	18,00	0	1	647,00	ARP-2.4	2	1	SIM	PLANO
JUR_30	1272000,00	1696,00	750,00	20,00	0	1	815,00	ARP-2.4	2	0	SIM	PLANO
JUR_31	950000,00	2111,11	450,00	15,00	0	1	509,00	ARP-2.4	2	0	SIM	PLANO
JUR_32	1800000,00	3000,00	600,00	20,00	1	2	80,00	ARP-2.4	2	0	SIM	PLANO
JUR_33	1000000,00	1536,10	651,00	21,70	0	1	875,00	ARP-2.4	2	0	SIM	PLANO
JUR_34	3200000,00	3555,56	900,00	30,00	0	1	210,00	ARP-2.4	2	0	SIM	PLANO
JUR_35	870000,00	1359,38	640,00	20,00	0	1	800,00	ARP-2.4	2	0	SIM	PLANO