

Estudos de caso

Aplicações na Engenharia de Avaliações

11/08/2019

1 Estudos de Caso

Para os estudos de caso foram utilizados os dados disponíveis em HOCHHEIM (2015).

1.1 Duas dimensões

Assim como na regressão linear, é mais fácil a compreensão da regressão quantílica através de exemplos em duas dimensões, e depois generalizar para n dimensões.

Seja primeiramente o caso de dados heteroscedásticos. A figura 1 ilustra a aplicação da regressão quantílica e da regressão linear para este caso. Na figura 1, a reta vermelha é a reta de regressão linear entre as variáveis. A área sombreada em cinza é o intervalo de confiança para a regressão linear @80%. As retas azuis são as retas de regressão quantílica para os quantis 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8 e 0,9.

A regressão quantílica neste caso pode ser usada para demonstrar a não validade dos intervalos de confiança (IC) e predição (IP) para a regressão linear para este tipo de dados: como a variância da população não é constante, mas aumenta com o aumento da área, as retas da regressão quantílica se abrem. Como os intervalos de confiança e predição na inferência clássica são calculados considerando-se que a variância da população é constante, este efeito não se observa no formato do IC.

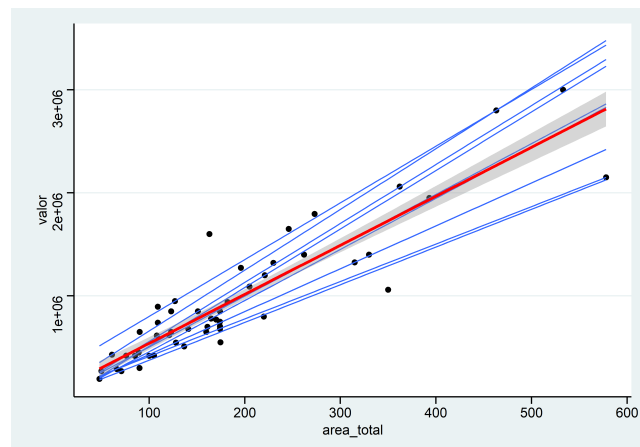


Figura 1: Regressão Linear e Quantílica para dados heteroscedásticos.

Assim como na regressão linear, uma conveniente transformação das variáveis pode ser aplicada para a obtenção da homoscedasticidade. Isto pode ser visto na figura 2, onde as retas para os diferentes quantis obtidas pela regressão quantílica agora são praticamente paralelas entre si, indicando que a heteroscedasticidade foi removida.

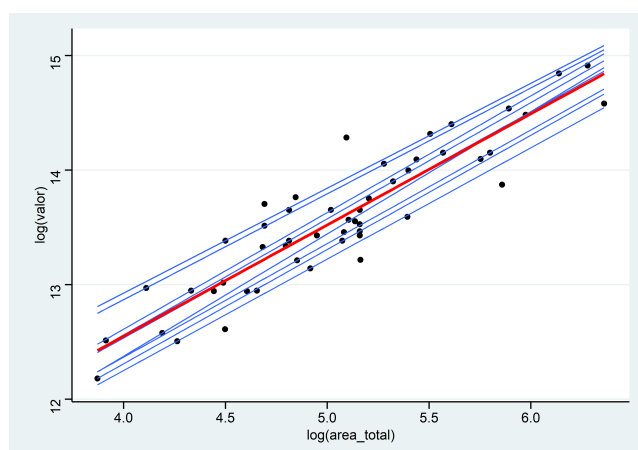


Figura 2: Regressão Linear e Quantílica com dados transformados.

Os coeficientes das retas de regressão quantílica podem ser plotados como na figura 3. Nesta figura, a reta cheia vermelha representa o coeficiente do modelo de regressão linear, enquanto a reta preta pontilhada representa os vários coeficientes da regressão quantílica. As retas vermelhas tracejadas representam o intervalo de confiança de estimação do coeficiente de regressão linear. A área sombreada em cinza representa os intervalos de confiança para os coeficientes da regressão quantílica. Deve-se notar que, entre os quantis aproximados de 0,3 e 0,55, os coeficientes da regressão quantílica não são significamente diferentes, estatisticamente, do coeficiente da regressão linear.

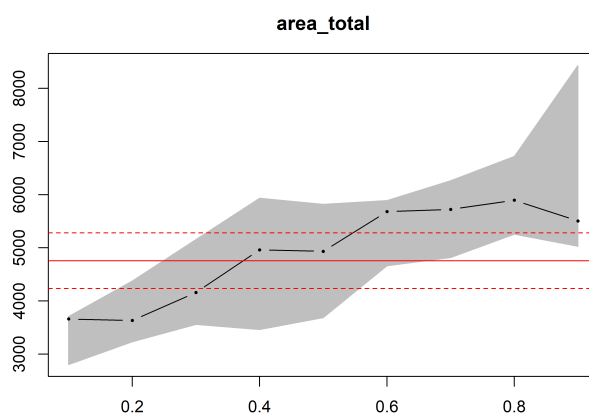


Figura 3: Variação dos coeficientes de regressão quantílica (variáveis originais).

Já para os dados transformados, pode-se notar na figura 4 que para todos os quantis, os coeficientes da regressão quantílica não podem ser considerados estatisticamente diferentes do coeficiente da regressão linear. Também se pode notar nesta figura como o estimador de regressão linear, para uma variável normalmente distribuída e na ausência de heteroscedasticidade, é mais eficiente do que o estimador da regressão quantílica, como a teoria já prevê (ver MATLOFF (2017), 238).

(Zilli, não sei se tu pesquisou isso na revisão bibliográfica, mas acho que se não, era bom colocar! Colocar algo do tipo: as vantagens e desvantagens da regressão quantílica. Apesar da regressão quantílica ser robusta à presença de *outliers*, ela é menos eficiente do que a regressão linear, caso a distribuição da variável estudada seja normal, claro.)

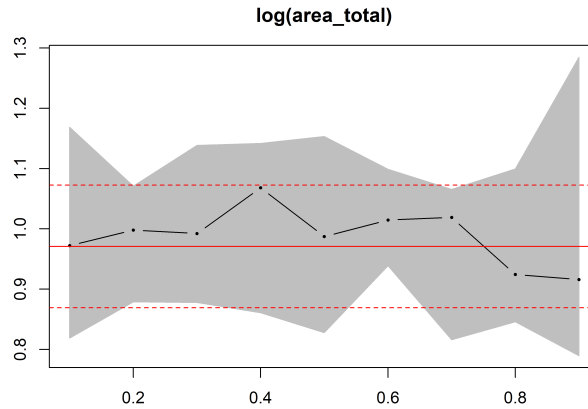


Figura 4: Variação dos coeficientes de regressão quantílica (variáveis transformadas).

1.2 Análise Multivariada

Para os dados obtidos de Hochheim (2015, pp. 22–23) foram ajustados dois modelos, um de regressão linear, com os dados saneados, e outro de regressão quantílica, utilizando-se a totalidade dos dados, para os quantis 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8 e 0,9.

Na figura 5 podem ser vistos os valores dos coeficientes de cada variável para os diferentes quantis. Pode-se perceber, mais uma vez, que o valor dos coeficientes da regressão quantílica não diferem significativamente dos coeficientes da regressão linear (exceção para alguns quantis superiores nas variáveis `area_total` e `padrao`).

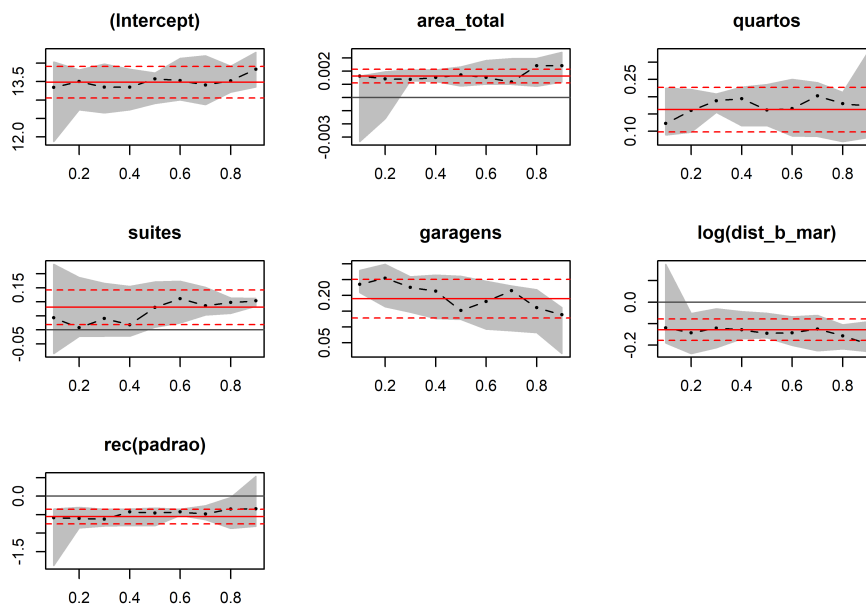


Figura 5: Coeficientes de regressão linear e quantílica. Análise multivariada.

Na tabela 1 podem ser vistos os coeficientes e estatísticas básicas dos modelos de regressão linear e de regressão à mediana (quantil 0,5).

1.2.1 Estimativas

É interessante comparar as estimativas obtidas com os modelos de regressão linear, com dados saneados, e o modelo de regressão à mediana, com a totalidade dos dados. Por um lado, o modelo de regressão linear tende a ser mais preciso para a estimação da média, como prevê a teoria. Por outro lado, com mais dados, o modelo de regressão à mediana pode tornar-se mais eficiente.

Deve-se levar em conta que as estimativas com o modelo de regressão linear aqui apresentadas são para a mediana da distribuição lognormal.

Pelo modelo de regressão linear, o valor da estimativa central encontrado foi de R\$961.660,64, com intervalo de confiança entre R\$ 924.768,13 e R\$ 1.000.024,94. A amplitude do intervalo de confiança foi de 7,83%.

Já pelo modelo de regressão quantílica, o valor da estimativa central encontrado foi de R\$946.467,87, com intervalo de confiança entre R\$ 886.472,34 e R\$ 1.010.523,85. A amplitude do intervalo de confiança foi de 13,10%.

O modelo de regressão linear mostrou-se, portanto, mais eficiente do que o modelo de regressão a mediana, apesar no menor número de dados.

Os limites inferior e superior do intervalo de predição @80% para o modelo de regressão linear são, respectivamente: R\$ 802.017,63 e R\$ 1.153.080,88.

Tabela 1: Comparação entre os modelos de regressão linear e regressão à mediana.

	<i>Dependent variable:</i>	
	log(valor)	
	<i>OLS</i>	<i>quantile regression</i>
	(1)	(2)
area_total	0.001 (0.001, 0.002) t = 5.113 p = 0.00001***	0.002 (0.001, 0.003) t = 2.300 p = 0.027**
quartos	0.164 (0.118, 0.209) t = 4.626 p = 0.00004***	0.162 (0.107, 0.217) t = 3.788 p = 0.0005***
suites	0.061 (0.018, 0.104) t = 1.810 p = 0.078*	0.080 (0.020, 0.139) t = 1.712 p = 0.095*
garagens	0.209 (0.166, 0.252) t = 6.247 p = 0.00000***	0.152 (0.075, 0.230) t = 2.520 p = 0.016**
log(dist_b_mar)	-0.141 (-0.176, -0.106) t = -5.174 p = 0.00001***	-0.146 (-0.210, -0.081) t = -2.904 p = 0.006***
rec(padrao)	-0.563 (-0.697, -0.428) t = -5.360 p = 0.00001***	-0.459 (-0.650, -0.267) t = -3.070 p = 0.004***
Constant	13.564 (13.268, 13.859) t = 58.847 p = 0.000***	13.574 (13.100, 14.047) t = 36.732 p = 0.000***
Observations	48	50
R ²	0.956	
Adjusted R ²	0.950	
Residual Std. Error	0.136 (df = 41)	
F Statistic	148.921*** (df = 6; 41)	

Note:

5 *p<0.1; **p<0.05; ***p<0.01

Para o modelo de regressão quantílica, o intervalo de predição não faz qualquer sentido. No entanto, é possível estimar os valores diretamente para os quantis 0,1 e 0,9 da população. Neste caso, os valores encontrados foram, respectivamente: R\$ 810.629,32 e R\$ 1.186.954,14.

Podem ainda ser calculados os intervalos de confiança @80% para as estimativas dos quantis 0,1 e 0,9.

Os limites inferior e superior do IC para o quantil 0,1 são, respectivamente: R\$ 781.253,06 e R\$ 841.110,17.

Os limites inferior e superior do IC para o quantil 0,9 são, respectivamente: R\$ 1.116.547,53 e R\$ 1.261.800,41.

Referências

HOCHHEIM, N. **Engenharia de avaliações - módulo básico**. Florianópolis: IBAPE - SC, 2015.

MATLOFF, N. **From linear models to machine learning: Regression and classification, with R examples**. Chapman & Hall, 2017.