

Econometria - APS 4

4º B

Pedro Xavier Marino

João Gabriel Cerqueira Gomes

Gabriel Fischel

Gustavo Sansaloni

João Casella

Insper - Instituto de Ensino e Pesquisa

2022

1 - Introdução

1.1 - Pergunta - Tema

“Quais as principais variáveis que afetam o preço de um imóvel no mercado imobiliário da Califórnia, Estados Unidos?”

1.2 - Relevância da Pergunta-Tema

O preço dos imóveis dentro de uma economia é uma das variáveis econômicas mais importantes a ser estudada. Sua variação é determinante em definir quantas pessoas conseguem comprar sua própria residência, quantas pessoas terão de viver de aluguel e quantas pessoas irão ser obrigadas a morar em abrigos ou na rua.

Analizar as variáveis que alteram preços no mercado imobiliário garantem a policymakers a rota de ação necessária para diminuir esses preços de maneira orgânica, caso necessário. Ao invés de congelar preços, legisladores podem influenciar as variáveis que influenciam diretamente o preço em si.

1.3 - Artigo Base

Artigo “What Drives House Price Cycles? International Experience and Policy Issues”

A Grande Recessão trouxe a atenção dos pesquisadores no que afeta os preços das casas e está associado aos ciclos de crédito, como essas variáveis afetam o cenário macroeconômico e a estabilidade financeira. O fracasso generalizado dos macro-modelos anteriores em prever e explicar as grandes oscilações nos preços das casas antes e depois da Crise de 2008 estimulou novas pesquisas. Os antigos modelos, ou seja, anteriores à crise, omitiram relações satisfatórias entre a economia real, os mercados de crédito e os preços dos ativos. Sem um acelerador financeiro, eles efetivamente descartaram a possibilidade de uma crise financeira global. Os modelos pós crise passaram a conter essas relações antes omitidas, o dano causado pela crise motivou essa evolução nos modelos. Este artigo analisa as contribuições recentes para a literatura sobre o que impulsiona os ciclos de preços da habitação, comparando e contrastando com a literatura anterior e abordando novas questões políticas. O foco particular é em estudos empíricos e teóricos que conectam mercados imobiliários com mercados de crédito, questões de estabilidade financeira e macroeconomia. Além disso, analisa como o papel econômico da moradia e dos mercados de crédito associados mudou ao longo do tempo e varia entre os países. A pesquisa enfatiza o papel da habitação como acelerador financeiro, que amplifica e propaga choques para a economia em geral, especialmente se ocorrer uma crise bancária e possivelmente uma crise de dívida soberana relacionada. Os excessos de oferta habitacional e de endividamento, e a consequente desalavancagem, podem deprimir a atividade econômica por períodos consideráveis.

As recentes crises no mercado imobiliário impulsionaram diversas pesquisas sobre o setor e seus ciclos, segundo o artigo base, vários avanços nesse sentido já foram feitos, com o controle de expectativas iracionais, regulamentações, DSGE e modelos empíricos, além de principalmente fatores que influenciam a construção e valorização de imóveis. Segundo o artigo, a modelagem de pesquisas sobre o mercado habitacional internacional possibilita não somente a melhor compreensão de seus ciclos como também ajudarão na criação de políticas que irão reduzir as consequências dos piores momentos desses ciclos, assim proporcionando mercados mais rentáveis, essenciais não apenas para a sobrevivência de empresas do ramo, como também para os que demandam tais serviços. Essas modelagens possuem impactos ao longo de várias partes da economia de um país e de suas políticas, com influências no mercado de trabalho, políticas de mobilidade, de tamanho, regulamentação e taxamento de terras, urbanismo da região e habitações sociais, além de influências indiretas, seja no saneamento, na poluição ou até na inflação.

Esse artigo pode ser encontrado no link: <https://www.aeaweb.org/articles?id=10.1257/jel.20201325>

2 - Análise Empírica

2.1 - Base de Dados

A base de dados decorre do Censo feito no estado americano da Califórnia em 1990 pelo próprio estado, e conta com diversas variáveis que podem alterar os preços dos imóveis do mercado imobiliário californiano. A amostragem conta com 20.640 regiões averiguadas.

Descrição das variáveis:

- *Median House Value*: Variável quantitativa de interesse. Estabelece a mediana do preço de imóveis em uma dada região da Califórnia em dólares.
- *Median Income*: Variável quantitativa regressora. Estabelece a mediana da renda de uma dada região na Califórnia em dólares. Pode ser usado como proxy para a restrição orçamentária da região.
- *Median Age*: Variável quantitativa regressora. Estabelece a mediana de idade de uma dada região da Califórnia em anos. A idade, por limitar a renda, acaba diminuindo o preço dos imóveis, diminuindo-os.
- *Tot Rooms*: Variável quantitativa regressora. Estabelece o total de salas de estar de uma dada região da Califórnia em unidade. Existe com o intuito de relacionar o tamanho das famílias com o preço dos imóveis.
- *Tot Bedrooms*: Variável quantitativa regressora. Estabelece o total de quartos em uma dada região da Califórnia em unidade. Existe com o intuito de relacionar o tamanho das famílias com o preço dos imóveis.
- *Population*: Variável quantitativa regressora. Estabelece a população de uma determinada região. O tamanho da população influencia a demanda e, portanto, o preço dos imóveis.

- *Households*: Variável quantitativa regressora. Estabelece a quantidade de pessoas que partilham da mesma casa em uma dada região da Califórnia. Quanto menos pessoas buscando uma casa, menos demanda.
- *Distance to Coast*: Variável quantitativa regressora. Estabelece a distância de uma dada região até a costa da Califórnia, em quilômetros.
- *Distance to Los Angeles*: Variável quantitativa regressora. Estabelece a distância de uma dada região até a cidade de Los Angeles da Califórnia, em quilômetros.
- *Distance to San Diego*: Variável quantitativa regressora. Estabelece a distância de uma dada região até a cidade de San Diego da Califórnia, em quilômetros.
- *Distance to San Jose*: Variável quantitativa regressora. Estabelece a distância de uma dada região até a cidade de San José da Califórnia, em quilômetros.
- *Distance to San Francisco*: Variável quantitativa regressora. Estabelece a distância de uma dada região até a cidade de San Francisco da Califórnia, em quilômetros.

Dividimos a mediana da renda média das regiões em 5, sendo a faixa inferior (1) as regiões mais pobres e a faixa mais superior (5) as regiões 2 mais ricas, e assim por diante. Os limites de cada classe de renda foram traçados com base nas classificações sociais por renda americanas disponíveis na Forbes. Aqui estão elas:

- *Lower Class*: Variável qualitativa dummy que determina se uma determinada região da Califórnia pertence a faixa 1 das regiões.
- *Middle Class*: Variável qualitativa dummy que determina se uma determinada região da Califórnia pertence a faixa 2 das regiões.
- *Upper Class*: Variável qualitativa dummy que determina se uma determinada região da Califórnia pertence a faixa 3 das regiões.
- *Middle Upper Class*: Variável qualitativa dummy que determina se uma determinada região da Califórnia pertence a faixa 4 das regiões.
- *Lower Upper Class*: Variável qualitativa dummy que determina se uma determinada região da Califórnia pertence a faixa 5 das regiões.

2.2 - Análise Descritiva Univariada

Na intenção de melhor compreender os comportamentos de cada variável individualmente, realizou-se uma análise descritiva univariada, compreendendo as medidas resumo de cada variável, suas assimetrias e curtoses (observando histogramas), e presença de outliers (através de boxplots). As tabelas, histogramas e boxplots utilizados estão disponíveis no tópico 5 - Anexos. Obteve-se o entendimento, na resultante das variáveis, que:

- *Median House Value*: É esperado que a mediana dos preços dessa região seja aproximadamente a média. Ainda assim, o desvio padrão dessas medianas é relativamente alto, demonstrando como a média, provavelmente, não é uma boa estimativa para uma mediana qualquer. Percebe-se que há curtose na cauda da direita, com uma forte concentração em valores mais baixos. Apesar disso, existe uma grande quantidade de imóveis com valores em torno de 500.000 dólares que puxam a média para cima. É teorizado que esses grandes valores nos outliers se concentrem em áreas de grande concentração urbana

como San Francisco ou Los Angeles, mas diminuem em áreas rurais, que compõem a maior parte do estado americano. O grande volume de observações impede uma análise mais aprofundada desses outliers.

- *Median Income*: Quando observado o desvio padrão maior do que 1, entende-se que a população da variável não se condensa ao redor da média. A população possui uma tendência assimétrica à direita, com maior concentração de frequências a valores menores do que a média.
- *Median Age*: Nota-se que a média da variável median age é 28,64, com desvio padrão de 12,59. Ao observar o histograma nota-se que a distribuição do histograma se aproxima da normal, porém sua curtose é menor que a da normal e ainda possui valores aberrantes na cauda da direita.
- *Total Rooms*: Nota-se que a média da variável total Rooms, ou seja, total de salas de estar, é 2635,76 com desvio padrão de 2181,62. Quando se observa o histograma, nota-se que a distribuição do histograma da variável se aproxima da normal, somente com poucos valores aberrantes na cauda da direita, no entanto, não atrapalham muito a análise de seu comportamento.
- *Total Bedrooms*: A média de quartos de uma região californiana é de aproximadamente 538. Sem uma estatística do número médio de casas dessa região, todavia, é impossível dizer se esse número é alto ou baixo. Ainda assim, a relativamente alta dispersão medida por um desvio padrão de 421,25, indica como esses números são muito diferentes entre uma região e outra. Isso é esperado, visto a vasta diferença geográfica entre espaços no enorme estado californiano. Existe uma forte concentração de valores perto da média, até aproximadamente os 2.000 quartos por região. A grande variabilidade decorre de alguns poucos valores que se espalham além dos 2.100 até os 6.000 quartos. Esses valores são outliers.
- *Population*: Nota-se que a média da variável population é 1425 pessoas aproximadamente com desvio padrão de 1132. Quando se observa o histograma nota-se que a distribuição do histograma da variável se aproxima da normal, somente com poucos valores aberrantes na cauda da direita.
- *Households*: Nota-se que a média da variável Households é 499 moradores aproximadamente com desvio padrão de 382. Quando se observa o histograma nota-se que a distribuição do histograma da variável se aproxima da normal, possuindo poucos valores aberrantes na cauda da direita.
- *Distance to Coast*: É possível observar que a média é de 40.509,26 milhas e seu desvio padrão é de 49.140,04. Quando se observa o histograma, é possível notar a existência de valores aberrantes na cauda, gerando assim uma tendência para uma assimetria à direita.
- *Distance to LA*: É possível notar que a média da variável em questão é de 269.422 milhas enquanto seu desvio padrão é de 247.732,40 milhas. Por se tratar de uma das cidades mais importantes da Califórnia, os preços tendem a se alterar também conforme a distância vai mudando. Ao analisar o histograma, não é possível tomar uma conclusão sobre sua assimetria e normalidade, uma vez que graças ao seu alto desvio padrão, as caudas são bem extensas.
- *Distance to San Diego*: É possível ver que a média é de 398.164,90 milhas e seu desvio padrão é de 289.400,60. Levando em consideração que San Diego é a segunda maior cidade do estado da Califórnia, é possível dizer que os preços poderiam sofrer alterações quando comparados a essa distância. Analisando o histograma 10 é possível perceber que, assim como na variável anterior, o desvio padrão da variável é muito alto, fazendo com que a cauda seja extensa impossibilitando perceber a existência de assimetrias ou até mesmo de normalidade.

- *Distance to San Jose*: É possível perceber que a média é de 349.187,60 e seu desvio padrão é de 217.149,90 milhas. Tendo em vista que San Jose é a terceira maior cidade do estado da Califórnia, os valores das residências poderiam sofrer influência dessas distâncias. Porém ao analisar o histograma não é possível tomar conclusões a respeito de sua normalidade ou até de possíveis assimetrias, uma vez que as caudas também são extensas e graças a presença de diversos valores aberrantes.
- *Distance to San Francisco*: Nota-se que a média da variável Distance to San Francisco é 386688,40 milhas com desvio padrão de 250122,20 milhas. O alto desvio padrão demonstra que as regiões analisadas são relativamente alto devido a imensidão do estado da Califórnia em relação a localização da região de San Francisco que é a cidade mais importante economicamente da Califórnia e que por conseguinte tem os maiores preços de residências. Quando se observa o histograma nota-se que a distribuição do histograma da variável não se aproxima da normal, existem diversos valores aberrantes e assimetrias tanto nas caudas da direita quanto na esquerda distorcendo totalmente a análise do histograma. O grande número de valores aberrantes torna essa variável pouco precisa.

2.3 - Análise Descritiva Bivariada

Uma vez analisadas as variáveis individualmente, a pesquisa foi aprofundada, desenvolvendo-se para uma análise descritiva bivariada - que compreende a interpretação dos gráficos de dispersão da variável resposta para com cada regressor; seleção das melhores formas funcionais para se explicar o modelo; entendimento das correlações entre as variáveis e comportamento das medidas resumo da variável resposta para com cada classe da variável categórica Renda_Classes.

Modelo/Var	Lin-Lin	Lin-Log	Log-Lin	Log-Log
Median Income	X			
Median Age	X			
Total Rooms	X			
Total Bedrooms	X			
Population	X			
Household	X			
Distance to Coast			X	
Distance to LA			X	
Distance to San Diego				X
Distance to San José				X
Distance to San Francisco				X

Em conclusão, todas as variáveis que menos as variáveis que descrevem distâncias são bem explicadas linearmente contra a variável resposta por meio do modelo lin-lin. As distâncias, porém, são mal descritas de maneira linear pelo lin-lin, tendo de ser ajustadas pelo log-lin ou log-log, mas nunca ficando realmente lineares, tal qual as outras variáveis, que também não ficam perfeitas, mas ficam melhores.

- Median Income não possui correlação forte ou moderada com nenhum outro regressor. As maiores correlações são com total rooms, que é positiva, e distan Coast, que é negativo.

- Median Age não possui correlação forte com nenhum outro regressor, porém observa-se que apenas distan diego possui correlação positiva. As maiores correlações são todas negativas e são com total rooms, total bedrooms, households e population.
- Total Rooms possui correlação muito alta e positiva com total bedrooms, population e households.
- Total Bedrooms possui correlação muito alta e positiva com total rooms, households e population.
- Population possui relação forte e positiva com total rooms, total bedrooms, households.
- Households possui correlação forte e positiva com total rooms, total bedrooms e population.
- Distan Coast não possui correlação forte ou moderada com nenhuma variável explicativa. Apenas distan Jose e distan Francisco possuem relação positiva com distan Coast, essas duas também são as maiores associações. As outras relações são negativas.
- Distan LA possui relação positiva e forte com Distan Diego.
- Distan Diego possui relação negativa e forte com distan Jose, distan Francisco e distan LA.
- Distan Jose possui relação forte e negativa com distan Diego, distan LA e distan Francisco.
- Distan Francisco possui relação forte e positiva com distan Jose.

Como a base de dados não apresenta variáveis categóricas, foram criadas variáveis dummy separando a Median Income em 5 categorias: “Lower Income”, “Lower Middle Income”, “Middle Income”, “Upper Middle Class” e “Upper Class”.

A fim de observar a tabela resumo da análise bivariada das variáveis categóricas percebe-se que, como esperado, o preço médio das residências das classes “Upper Class” é o mais alto, cerca de 488094,54 dólares e também essa variável dummy tem o menor desvio padrão entre as demais de 57833,74 dólares, portanto, essa classe de renda concentra os valores aberrantes da variável Median House Value. Nota-se claramente que quanto maior a renda das classes sociais maior é a média dos valores das residências e isso influencia os valores locais das residências em determinadas regiões. Além disso, também que a mediana da Upper Class é 500000 enquanto a média é 488094,54, por consequência, indicando ainda uma assimetria à direita. A classe Upper Middle Class apresenta maior desvio padrão, portanto, dispersando os valores dessa classe.

2.4 - Averiguação de Interações entre Regressores

Analisando o efeito marginal de cada um dos regressores em relação às dummies, foi averiguada a necessidade de regressores de interação entre três variáveis quantitativas e as dummies. São elas: Median Age, Population e Household. Os gráficos utilizados nessa análise podem ser conferidos nos anexos. Elas foram devidamente incluídas no modelo. Essa inclusão foi ponderada a partir dos gráficos, de tal forma a afirmar a necessidade de interação quando as retas do regressor (para cada categoria de dummy) não se mostrarem paralelas. Assim, as seguintes interações foram colocadas no modelo final:

INTERAÇÕES	Renda Classes				
	Lower Class	Lower Middle Class	Middle Class	Upper Middle Class	Upper Class
Median Age		X	X	X	X
Tot Rooms					
Tot Bedrooms					
Population		X	X	X	X
Households			X	X	
lDistance to Coast					
lDistance to LA					
lDistance to San Diego					
Numeração Renda_Classes	1	2	3	4	5

2.5 - Proposição do Modelo de Regressão

Como conferido nas análises uni e bivariadas, existe a necessidade de que algumas variáveis entrem no modelo por meio de logaritmo natural, estas são as variáveis quantitativas que incluem distância. Também foi necessário a inclusão de regressores de interação como averiguado na seção 2.4.

Feito então feito um modelo com todas as variáveis e todas as regressões que foram consideradas relevantes. Segue:

$$\begin{aligned}
 \text{Median House Values} = & \beta_0 + \beta_1(\text{Median Income}) + \beta_2(\text{Median Age}) + \beta_3(\text{Total Rooms}) + \beta_4(\text{Total Bedrooms}) + \beta_5(\text{Population}) \\
 & + \beta_6(\text{Households}) + \beta_7[\ln(\text{Distance to Coast})] + \beta_8[\ln(\text{Distance to LA})] + \beta_9[\ln(\text{Distance to San Diego})] \\
 & + \beta_{10}[\ln(\text{Distance to San Jose})] + \beta_{11}[\ln(\text{Distance to San Francisco})] + \beta_{12} * (\text{Median Age} * \text{Lower Class}) + \beta_{13} \\
 & * (\text{Median Age} * \text{Lower Middle Class}) + \beta_{14} * (\text{Middle Class}) + \beta_{15} * (\text{Median Age} * \text{Upper Middle Class}) + \beta_{16} \\
 & * (\text{Median Age} * \text{Upper Class}) + \beta_{17} * (\text{Household} * \text{Lower Class}) + \beta_{18} * (\text{Household} * \text{Lower Middle Class}) + \beta_{19} \\
 & * (\text{Household} * \text{Middle Class}) + \beta_{20} * (\text{Household} * \text{Upper Middle Class}) + \beta_{21} * (\text{Household} * \text{Upper Class}) + \beta_{22} \\
 & * (\text{Population} * \text{Lower Middle Class}) + \beta_{23} * (\text{Population} * \text{Middle Class}) + \beta_{24} * (\text{Population} * \text{Upper Middle Class}) + \beta_{25} \\
 & * (\text{Population} * \text{Upper Class}) + \varepsilon
 \end{aligned}$$

O modelo, todavia, apresentou uma série de variáveis que não passaram no teste-t, e, portanto, não tiveram impacto significativo na variável resposta. Além do mais, esse modelo inicial era heterocedástico. Por meio de um cuidadoso processo de eliminação dessas variáveis, fomos deixados com um modelo final ROBUSTO cujo qual todos os regressores são estatisticamente relevantes. O modelo pode ser considerado a seguir (vide passo a passo no Anexo):

$$\begin{aligned}
 \text{Median House Values} = & -1953138 + 7015,303 * (\text{Median Income}) + 1082,233 * (\text{Median Age}) + 2,80013 * (\text{Tot.Roms}) + 175527,9 \\
 & * [\ln(\text{Distance to Coast})] - 443,5267 * (\text{Median Age} * \text{Lower Middle Class}) - 882,4519 * (\text{Median Age} * \text{Middle Class}) \\
 & - 232,2503 * (\text{Median Age} * \text{Upper Middle Class}) - 409,8461 * (\text{Median Age} * \text{Upper Class}) + 54,20719 \\
 & * (\text{Households} * \text{Middle Class}) + 95,45534 * (\text{Households} * \text{Upper Middle Class}) - 10,35671 \\
 & * (\text{Population} * \text{Lower Middle Class}) - 27,95064 * (\text{Population} * \text{Middle Class}) - 41,5779 * (\text{Population} * \text{Upper Middle Class}) \\
 & + 21,03388 * (\text{Population} * \text{Upper Class})
 \end{aligned}$$

Pelo Anexo, é possível observar pelo teste F de significância global da regressão que o modelo é relevante, pois o p-Valor do teste F dele é aproximadamente 0, portanto, rejeitando a hipótese nula de que todos os betas sejam iguais a zero.

3 - Análise dos Resíduos

3.1 - Heterocedasticidade

Primeiramente, como pode ser averiguado nos anexos, foi conduzido a construção dos gráficos de dispersão entre os resíduos padronizados contra cada um dos regressores do modelo. A maior parte deles existe homogeneamente distribuída ao redor do zero, o que indica possível heterocedasticidade. Ainda assim, quando os regressores que incluem distância são comparados com os resíduos padronizados, é possível perceber uma descrição parabólica dos dados. Por causa disso, foi realizado um teste de White para a regressão, que revelou de fato a heterocedasticidade do modelo.

Em resposta a isso, o modelo de regressão foi refeito de maneira robusta, para que a inferência estatística não seja comprometida pela falta de homocedasticidade dos erros. O modelo robusto pode ser conferido nos anexos.

3.2 - Dependência

Feito isso, foi feito um gráfico de dispersão entre o erro padronizado e a ordem de coleta das observações, para averiguar qualquer tipo de dependência entre os erros e a ordem. Os dados, todavia, se organizaram homogeneamente ao redor do zero, garantido nenhum tipo de dependência.

3.3 - Normalidade

Por fim, foi averiguado a normalidade dos resíduos, via qq-plot e histograma, que indicaram uma possível normalidade dos erros. Ambos podem ser averiguados nos anexos. Ainda assim, nenhum deles pode garantir a normalidade, então foi conduzido também o teste de Jarque-Bera, que garantiu numericamente que os erros são normalmente distribuídos.

4 - Conclusão

4.1 - Implicações do Modelo

O modelo final nos apresenta com algumas interessantes conclusões que podem ser traçadas a sobre o mercado imobiliário da Califórnia, aqui estão algumas delas:

- Um aumento de renda nas regiões de classe alta tem o menor impacto de todos os aumentos de renda entre todas as classes. Um aumento de renda em classes baixas, por exemplo, aumenta o preço do imóvel bem mais que o mesmo aumento nas classes altas.
- O envelhecimento da população tem correlação positiva com os preços dos imóveis. Populações mais velhas pagam mais caro por moradia.
- Um aumento da distância da costa foi ligado a um gigantesco aumento positivo nos preços dos imóveis. Regiões costeiras pagam mais barato por moradia.

- O aumento da população em regiões de renda média alta é o único aumento populacional capaz de causar um aumento no preço dos imóveis. Se a população aumentar em qualquer outra renda, o preço dos imóveis cai.

Essas conclusões podem ser utilizadas tanto pelo mercado imobiliário quanto pelo estado da Califórnia. Sabendo os fenômenos que levarão ao aumento nos preços imobiliários, os corretores podem expandir sua produção para aproveitar o futuro aumento nos preços, e o estado pode antecipar aumentos nos imóveis para preparar políticas de contração monetária ou fiscal para diminuir a inflação que está por vir.

Analizando constantemente não o mercado imobiliário em si, que é extremamente volátil, mas o aumento da população, seu envelhecimento e seu aumento ou diminuição de renda, o estado e as firmas têm uma vantagem preditiva ao saber o que segue estes fenômenos e podem responder a eles com antecedência.

4.2 - Limitações do Modelo

O modelo apresentou um R-quadrado de aproximadamente 92%, sendo, portanto, capaz de explicar grande parte do fenômeno que são os preços do mercado imobiliário do estado da Califórnia. Ainda assim, essa explicação diz respeito à MEDIANA dos preços dos imóveis, que não é tão relevante ou significante como a média desses preços. Um aumento da mediana pode capturar um aumento real na população, mas pode facilmente ser uma péssima representação da realidade no geral, sendo facilmente enviesada.

Além disso, parte das variáveis do modelo foram omitidas por terem apresentado colinearidade perfeita. Elas participam do R-quadrado, mas sua inferência estatística está comprometida, dificultando a análise do modelo.

Por fim, algumas conclusões traçadas estão muito provavelmente erradas. O fato de que um aumento da distância até a costa aprecia é o preço dos imóveis é algo extremamente improvável, ainda mais em um estado cuja costa é composta de praias extremamente populosas. Isso nos prova que os parâmetros estimados do modelo muito possivelmente não estão sendo capazes de replicar os parâmetros estruturais da realidade.

5 - Anexos

Tabela 1 – Medidas-Resumo p/ Variáveis Numéricas

Variável Resposta	Média	Desvio Padrão	Mínimo	Máximo
Median House Value	206855,80	115395,60	14999	500001
Variáveis Explicativas	Média	Desvio Padrão	Mínimo	Máximo
Median Income	3,87	1,90	0,50	150.001
Median Age	28,64	12,59	1	52
Tot Rooms	2.635,76	2.181,62	2	39.320
Tot Bedrooms	537,90	421,25	1	6.445
Population	1.425,48	1.132,46	3	35.682
Households	499,54	382,33	1	6.082
Distance to Coast	40.509,26	49.140,04	120,68	333.804,70
Distance to LA	269.422,00	247.732,40	420,59	1.018.260
Distance to San Diego	398.164,90	289.400,60	484,92	1.196.919
Distance to San José	349.187,60	217.149,90	569,45	836.762,70
Distance to San Francisco	386.688,40	250.122,20	456,14	903.627,70

Tabela 2 – Medidas-Resumo p/ Variável Categórica

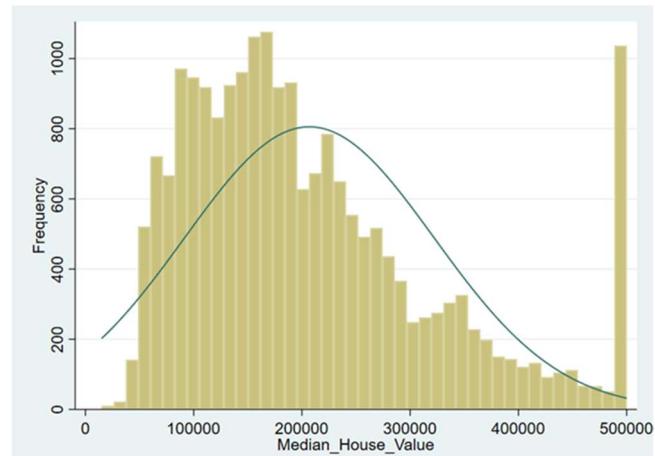
Median House Value Renda Classes	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Lower Class	120.591,29	87.924,76	92.600,00	14.999	500001
Lower Middle Class	124.573,05	69.704,12	106.300	14.999	500001
Middle Class	200.868,43	91.787,23	182.400	14.999	500001
Upper Middle Class	336.189,94	107.684,70	325.000	47.500	500001
Upper Class	488.094,54	57.833,74	500.001	112.500	500001

Tabela 3 – Correlações

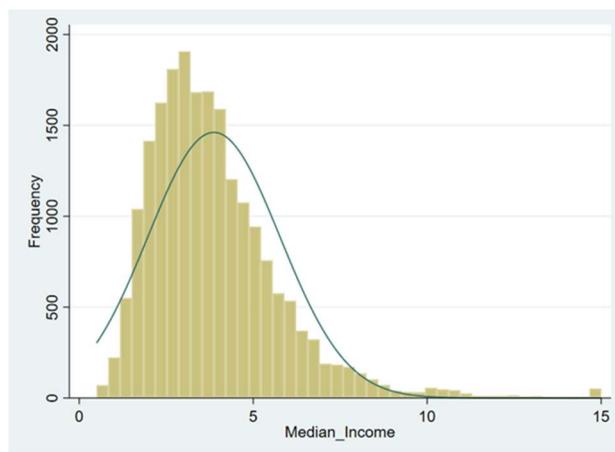
	House Value	Median Income	Median age	Total Rooms	Total Bed	Population	Households	Distan Coast	Distan LA	Dista Diego	Distan Jose	DistatFrancisc
House Value	1											
Median Income	0.6881	1										
Median age	0.1056	-0.1190	1									
Total Rooms	0.1342	0.1980	-0.3613	1								
Total Bed	0.0506	-0.0081	-0.3205	0.9299	1							
Population	-0.0246	0.0048	-0.2962	0.8571	0.8780	1						
Households	0.0658	0.0130	-0.3029	0.9185	0.9798	0.9072	1					
Distan Coast	-0.4694	-0.2434	-0.2266	-0.0015	-0.0223	-0.0403	-0.0620	1				
Distan LA	-0.1307	-0.0654	-0.0314	-0.0198	-0.0558	-0.1104	-0.0621	0.1977	1			
Dista Diego	-0.0925	-0.0553	0.0361	-0.0389	-0.0676	-0.1097	-0.0691	0.2145	0.9515	1		
Distan Jose	-0.0416	-0.0368	-0.0898	0.0319	0.0597	0.0791	0.0479	-0.0775	-0.7940	-0.8876	1	
DistatFrancisc	-0.0306	-0.0224	-0.1014	0.0329	0.0602	0.0886	0.0502	-0.0682	-0.8487	-0.9278	0.9899	1

Histogramas:

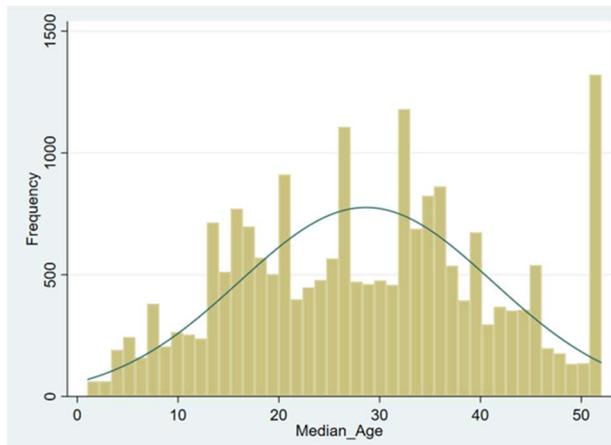
Histograma 1 – Median House Value



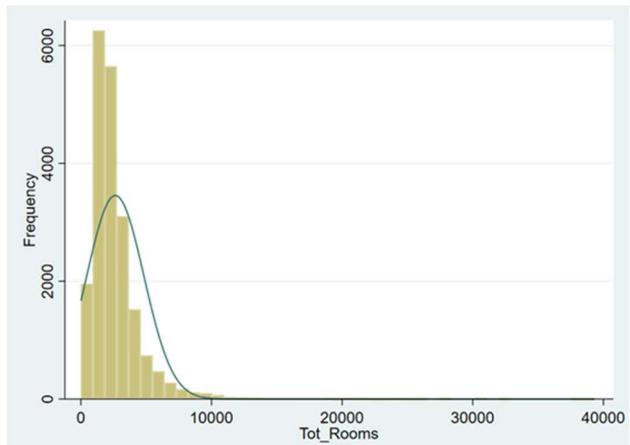
Histograma 2 – Median Income



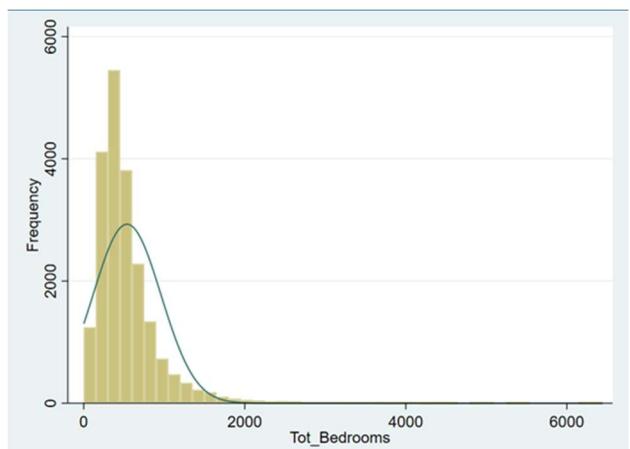
Histograma 3 – Median Age



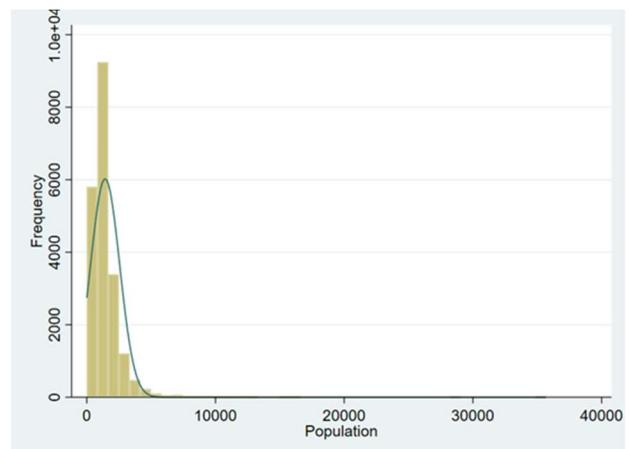
Histograma 4 – Total rooms



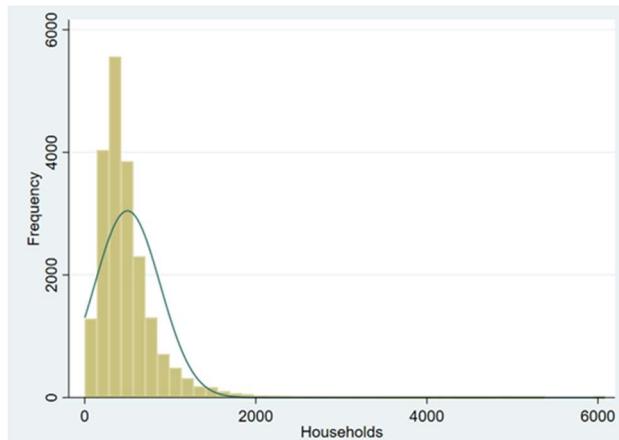
Histograma 5 – Total Bedrooms



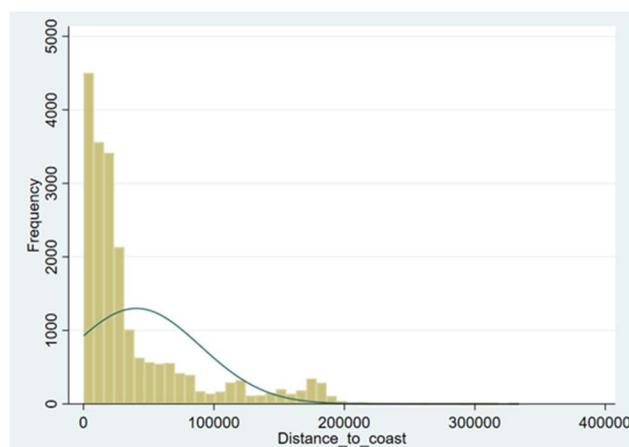
Histograma 6 – Population



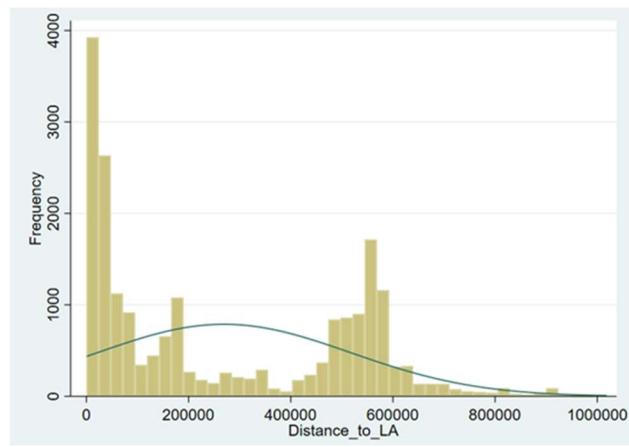
Histograma 7 – Households



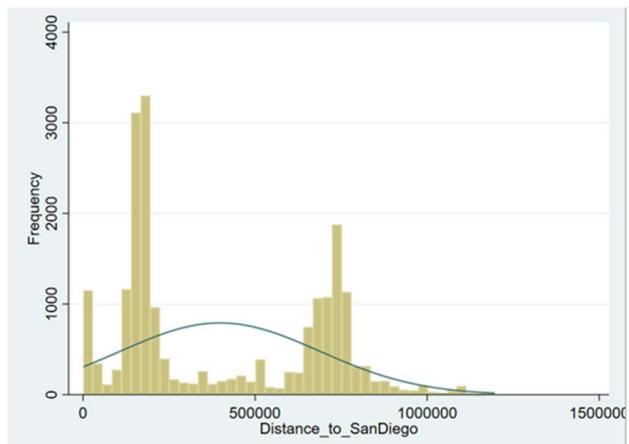
Histogramma 8 – Distance to Coast



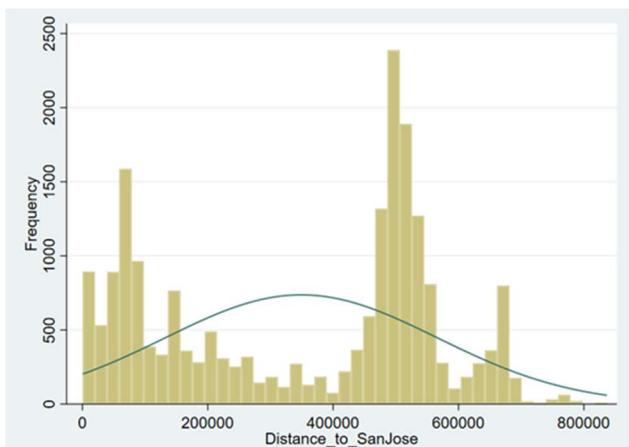
Histogramma 9 – Distance to LA



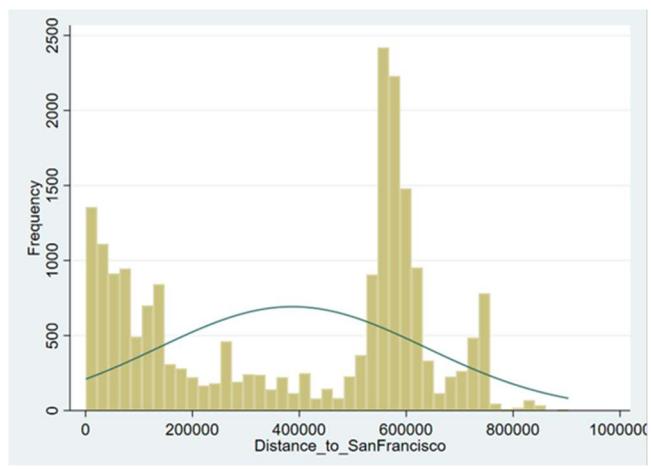
Histogramma 10 – Distance to San Diego



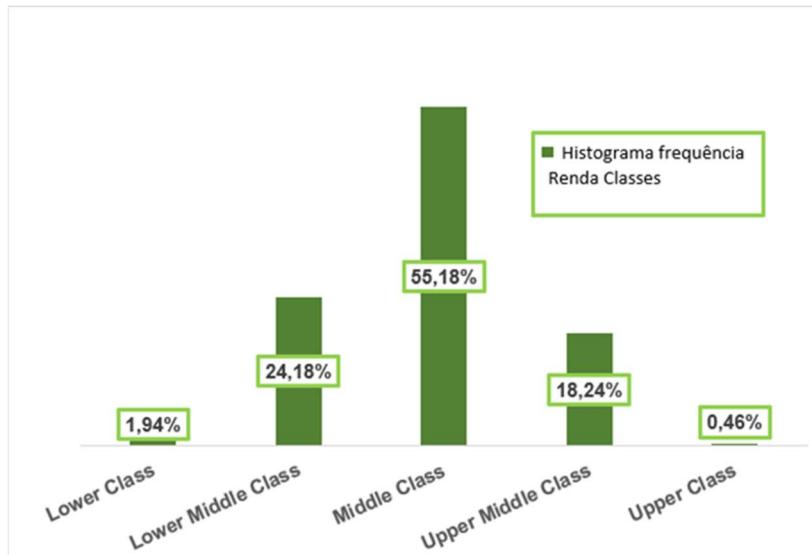
Histograma 11 – Distance to San José



Histograma 12 – Distance to San Francisco

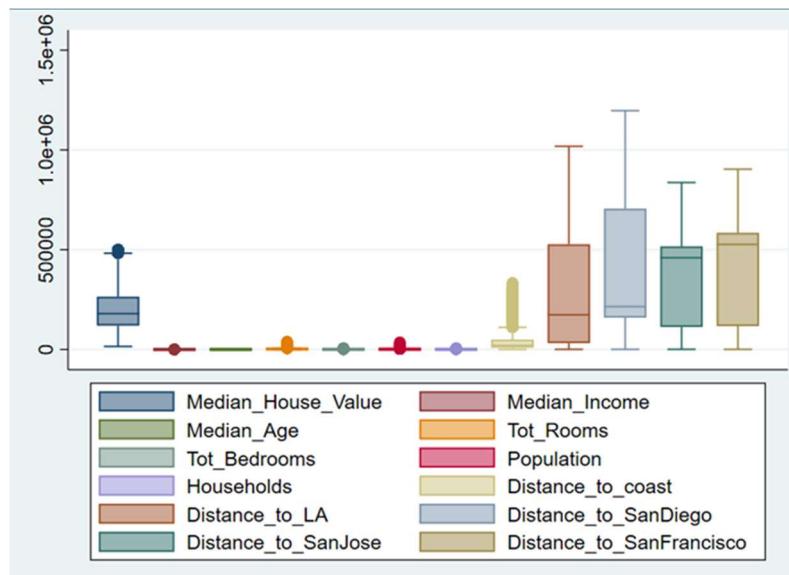


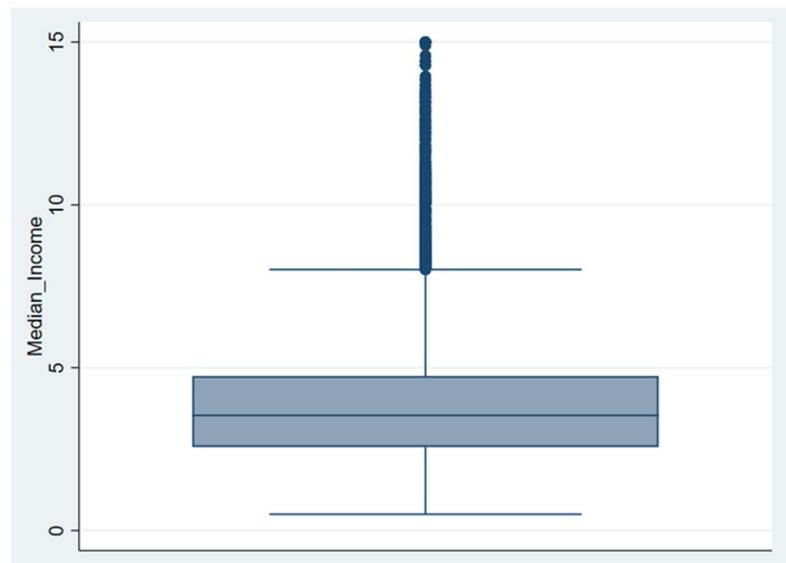
Histograma 13 – Renda Classes (categórica)



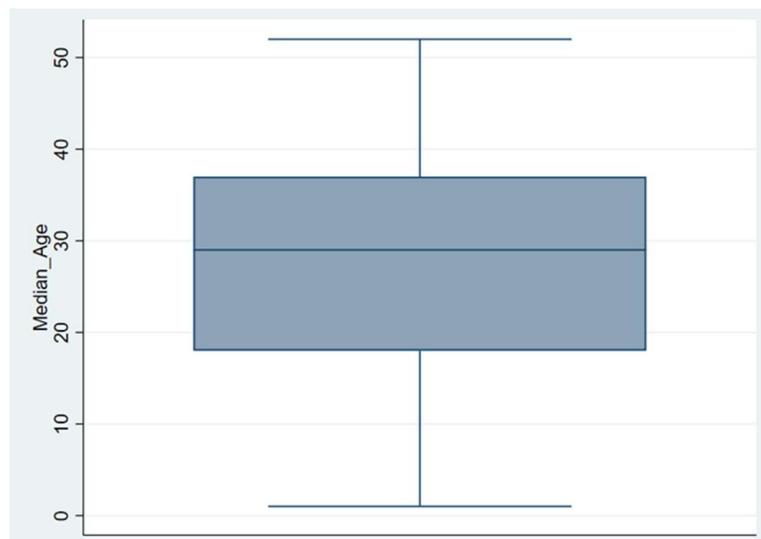
Boxplot:

Boxplot 1 – Todas as variáveis

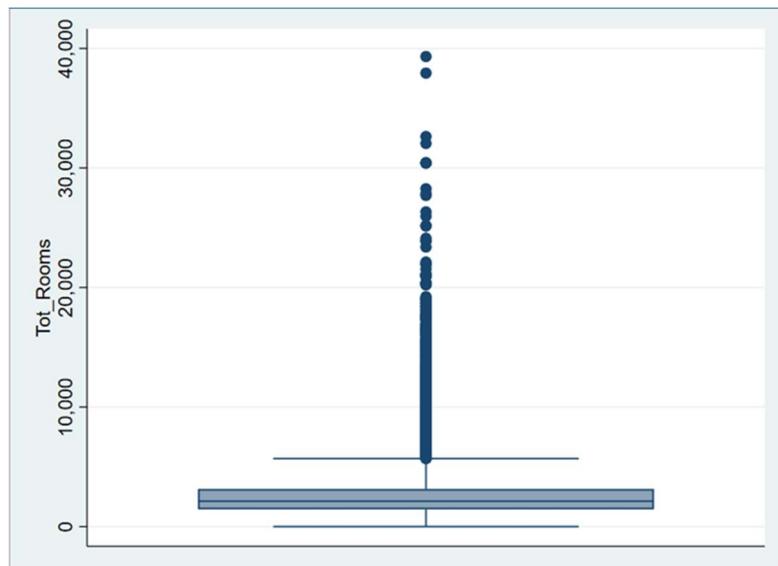




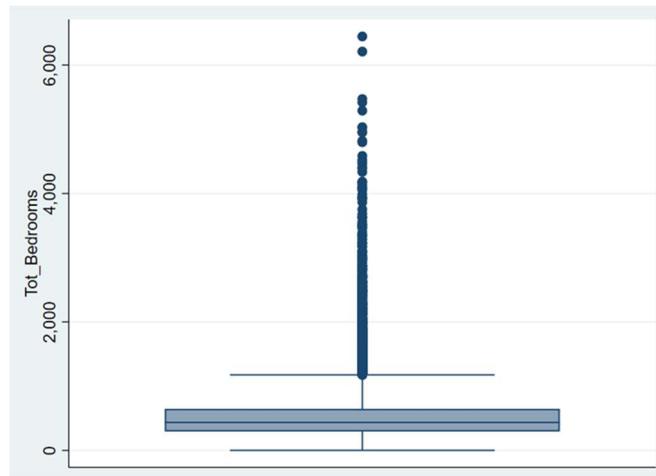
BoxPlot2 – Median Age



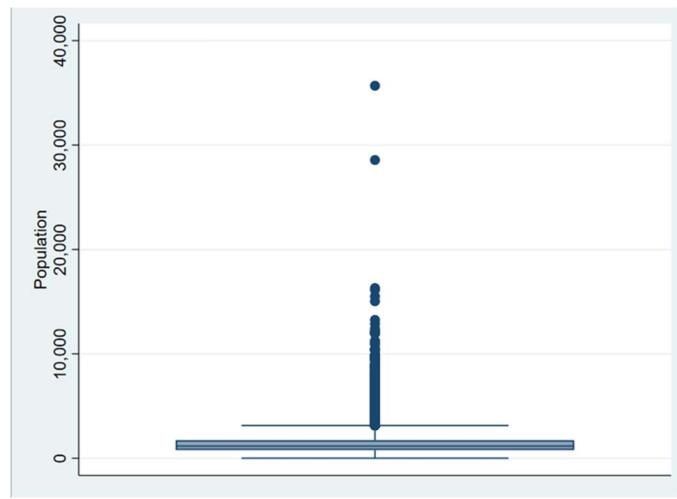
BoxPlot 3- Total Rooms



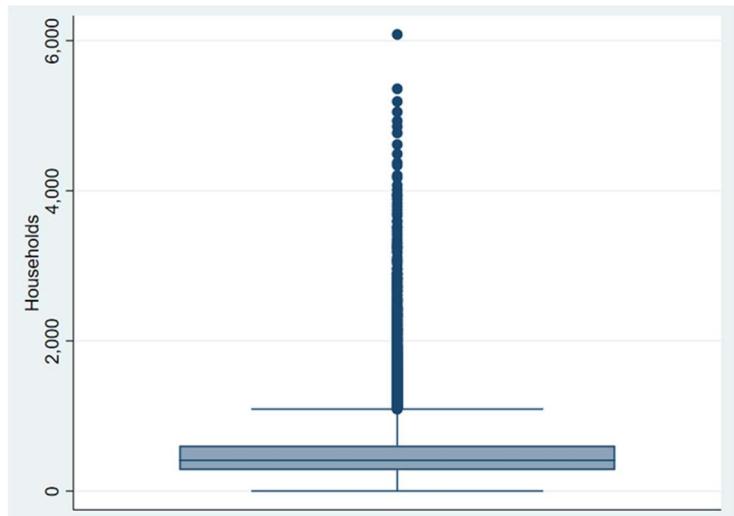
BoxPlot 4 – Tot Bedrooms



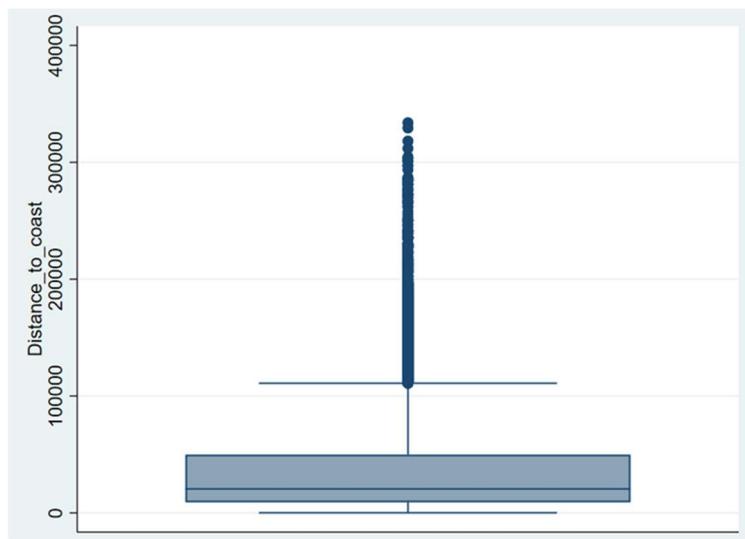
BoxPlot 5 –Population



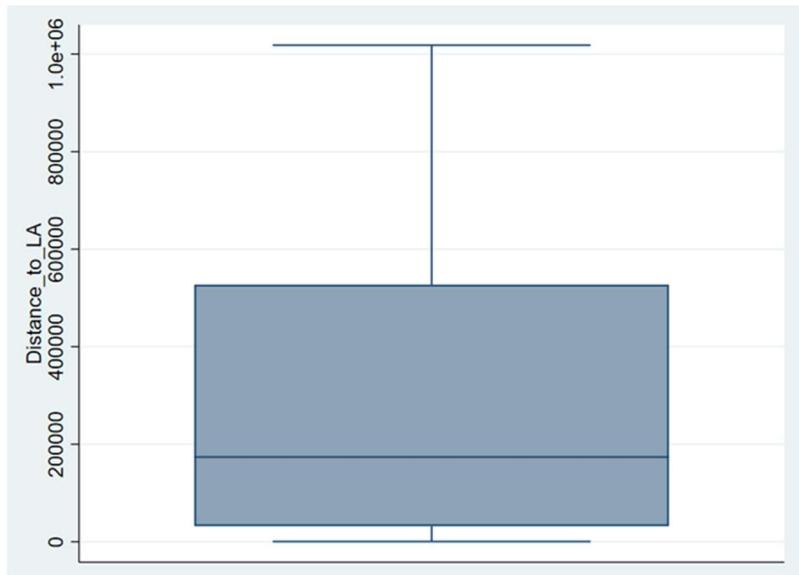
BoxPlot 6 – Households



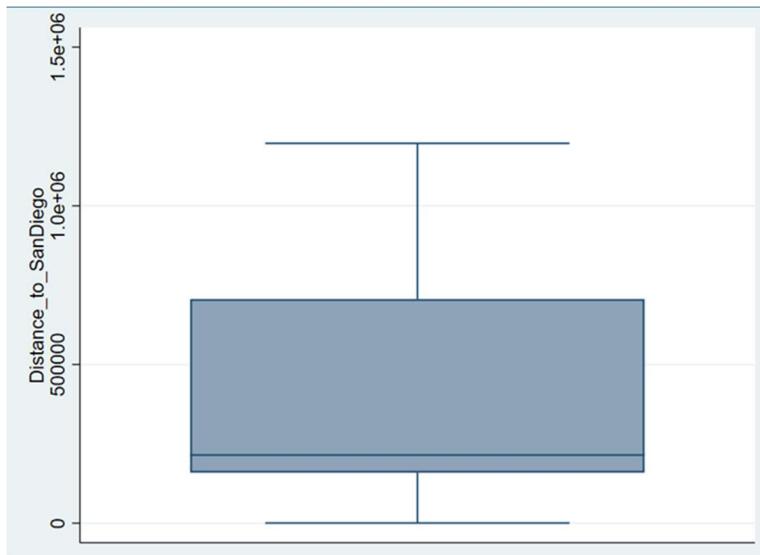
BoxPlot 7 – Distance to Coast



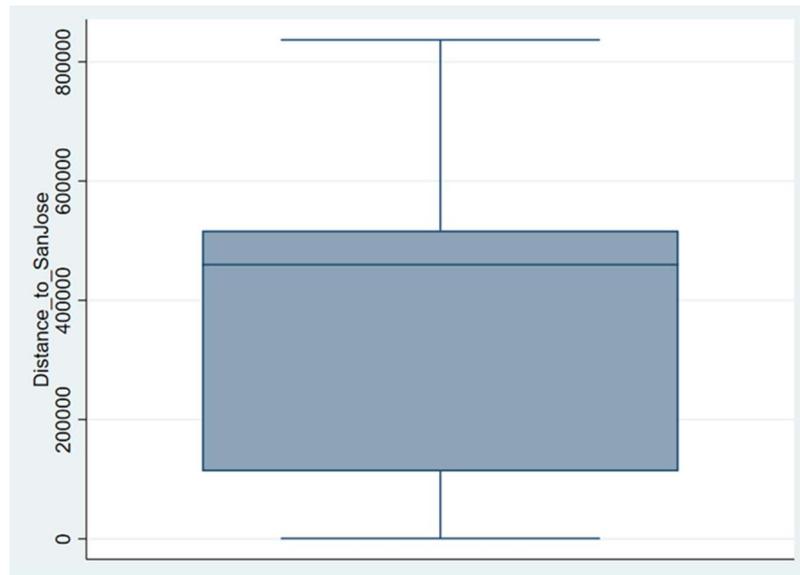
BoxPlot 8 – Distance to LA



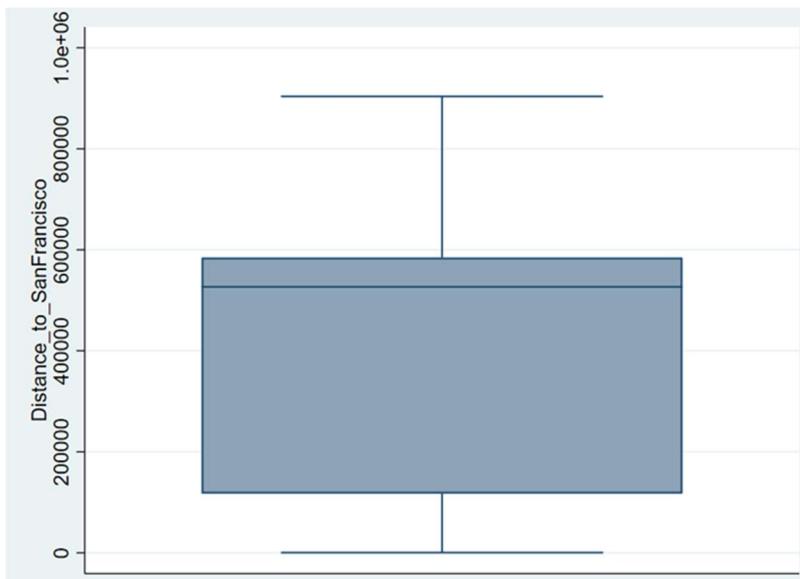
BoxPlot 9 – Distance to San Diego



BoxPlot 10 – Distance to San Jose



BoxPlot 11 – Distance to San Francisco



Gráficos de Dispersão Lin – Lin:

Gráfico de Dispersão (lin – lin) 1 – Median Income

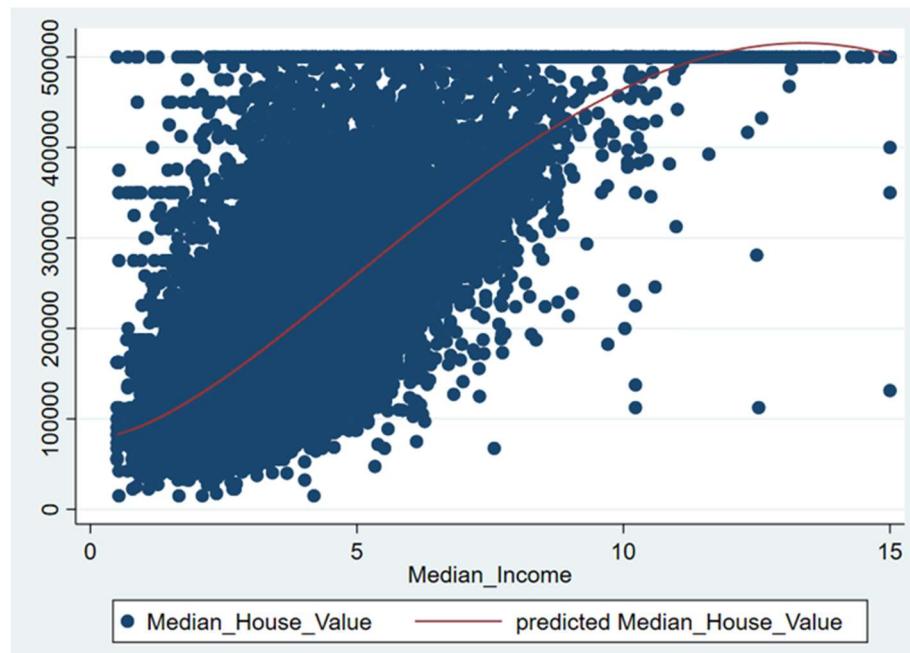


Gráfico de Dispersão (lin – lin) 2 – Median Age

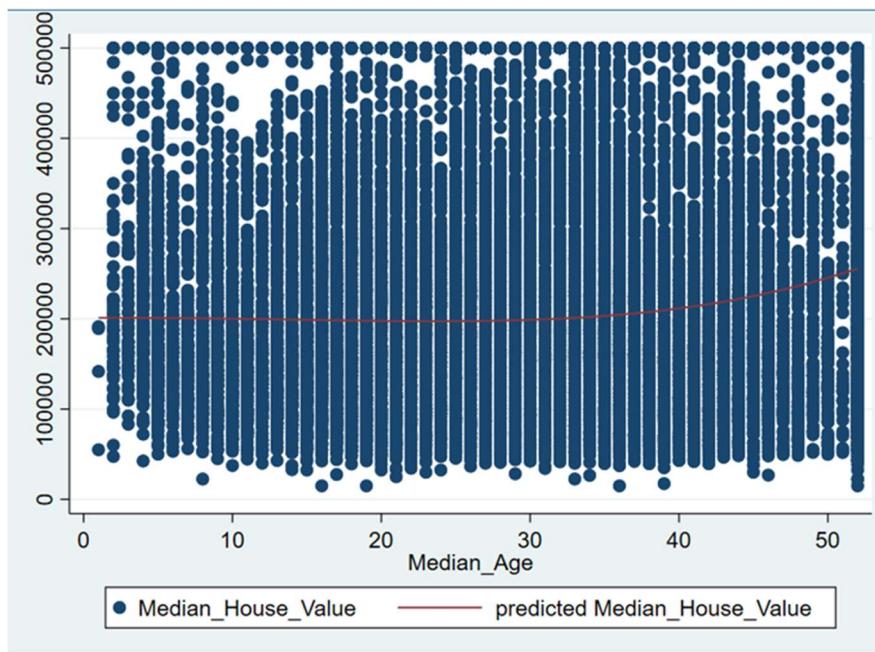


Gráfico de Dispersão (lin – lin) 3 – Total Rooms

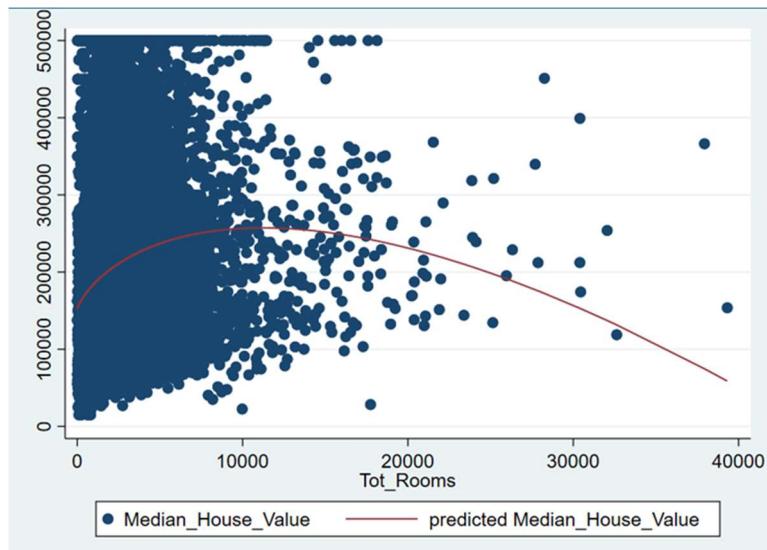


Gráfico de Dispersão (lin – lin) 4 – Total Bedrooms

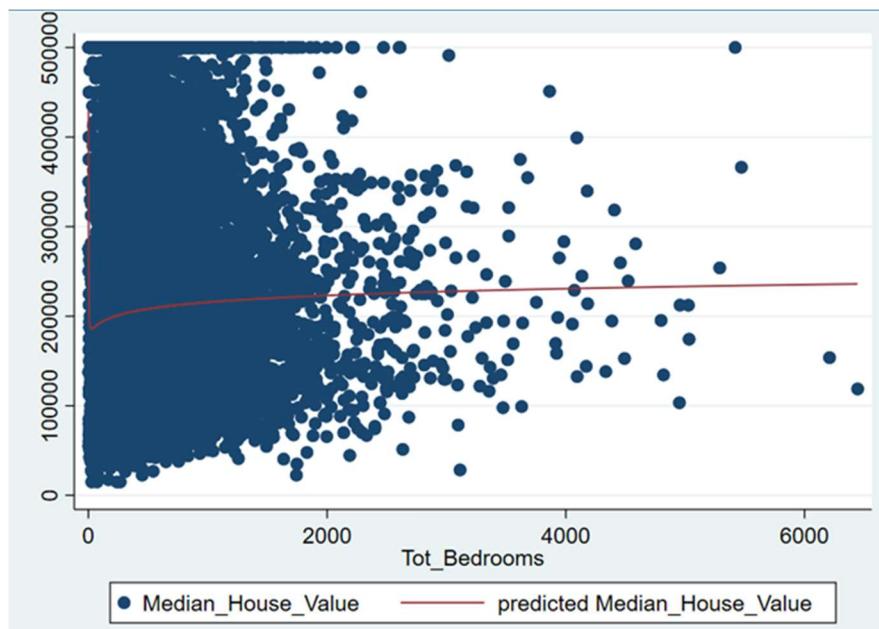


Gráfico de Dispersão (lin – lin) 5 – Population

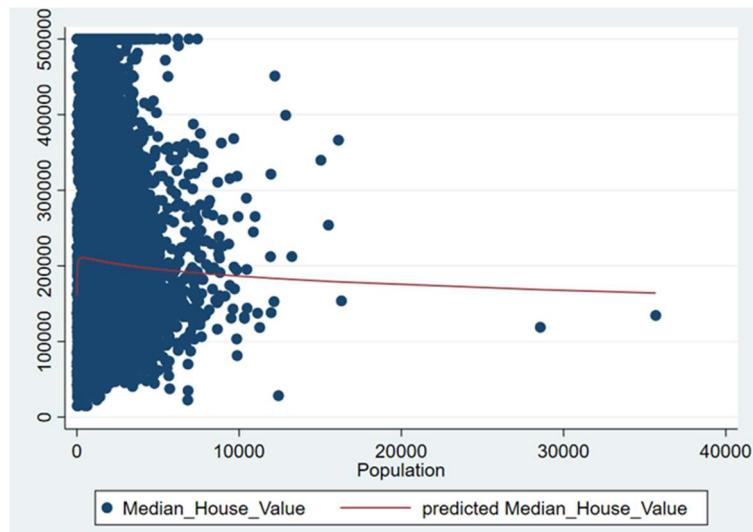


Gráfico de Dispersão (lin – lin) 6 – Household

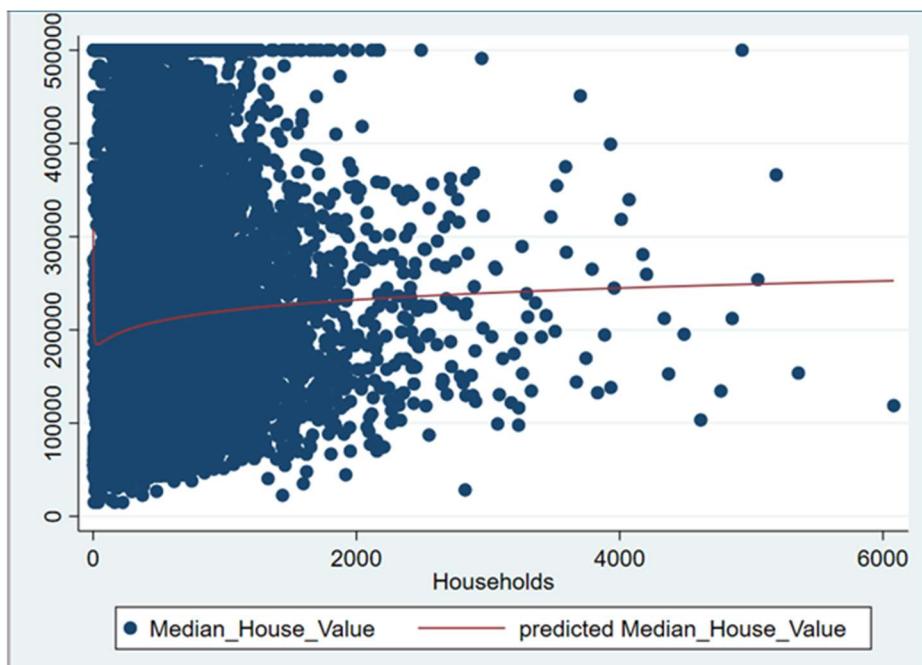


Gráfico de Dispersão (lin – lin) 7 – Distance to Coast

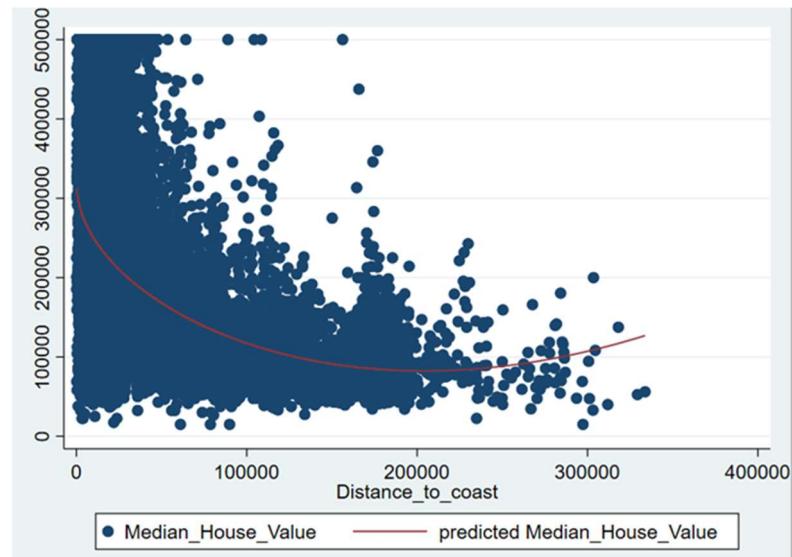


Gráfico de Dispersão (lin – lin) 8 – Distance to LA

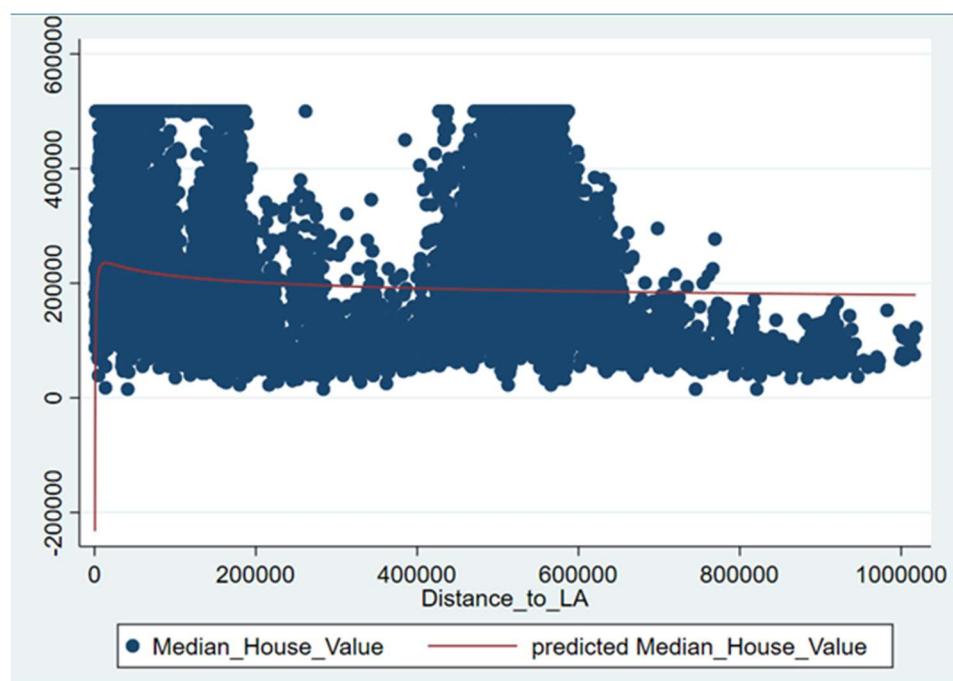


Gráfico de Dispersão (lin – lin) 9 – Distance to San Diego

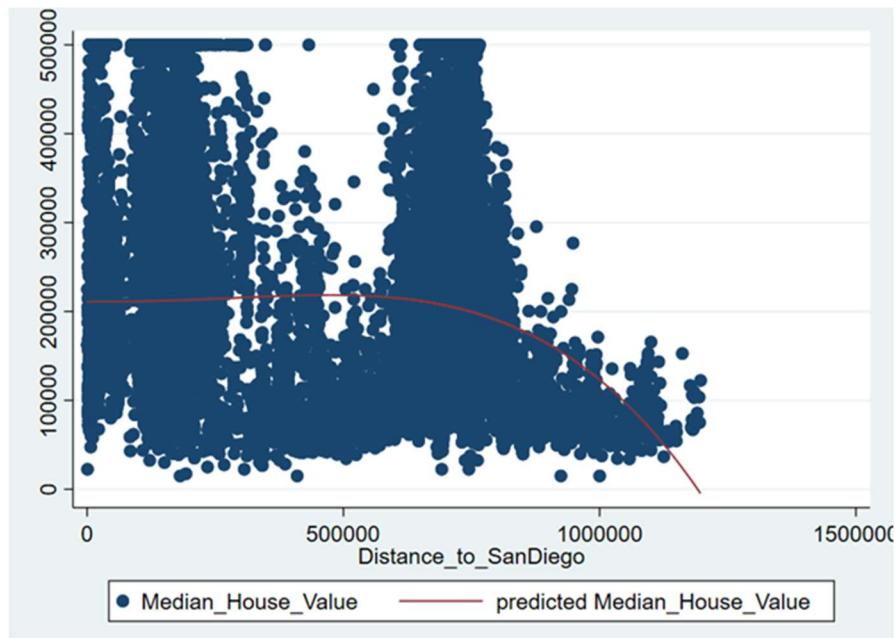


Gráfico de Dispersão (lin – lin) 10 – Distance to San José

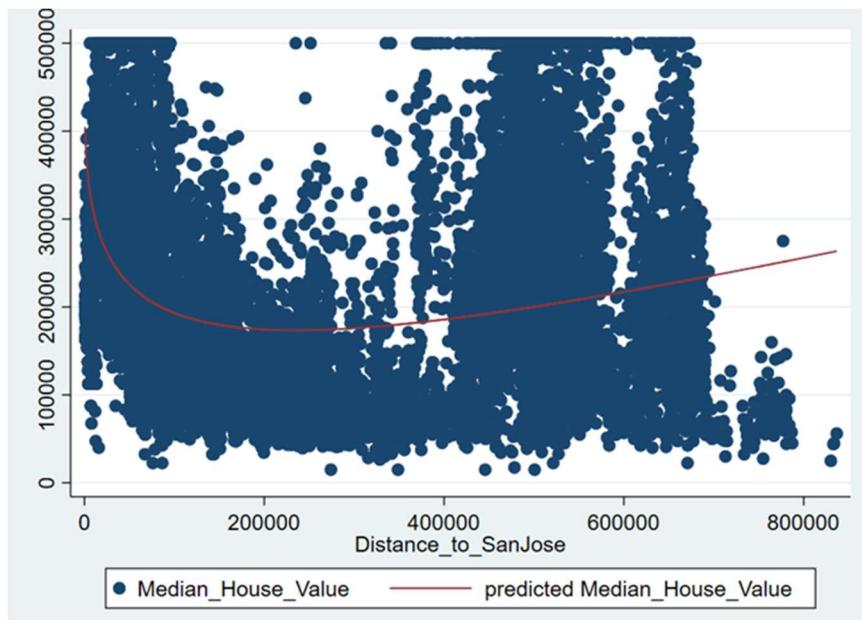
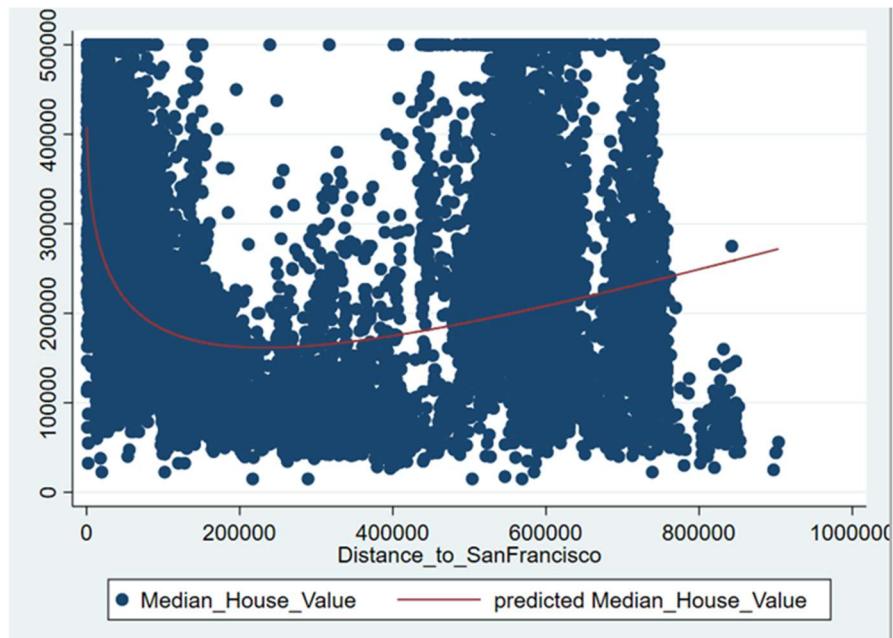


Gráfico de Dispersão (lin – lin) 11 – Distance to San Francisco



Gráficos de Dispersão Log – Lin:

Gráfico de Dispersão (log – lin) 1 – Median Income

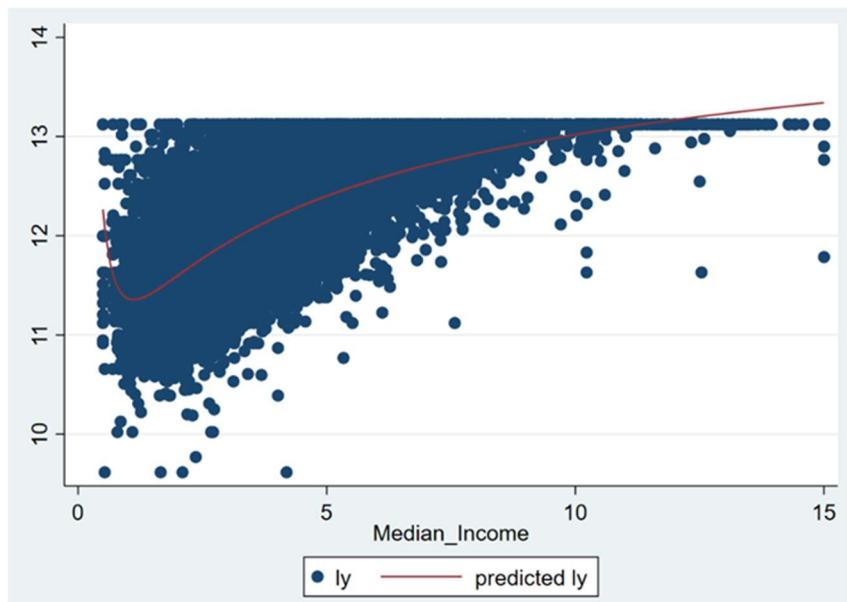


Gráfico de Dispersão (log – lin) 2 – Median Age

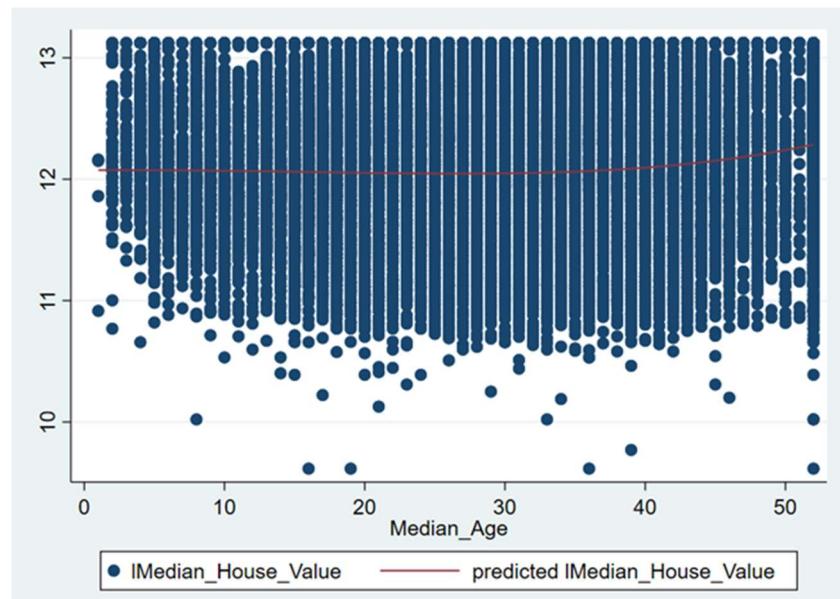


Gráfico de Dispersão (log – lin) 3 – Total Rooms

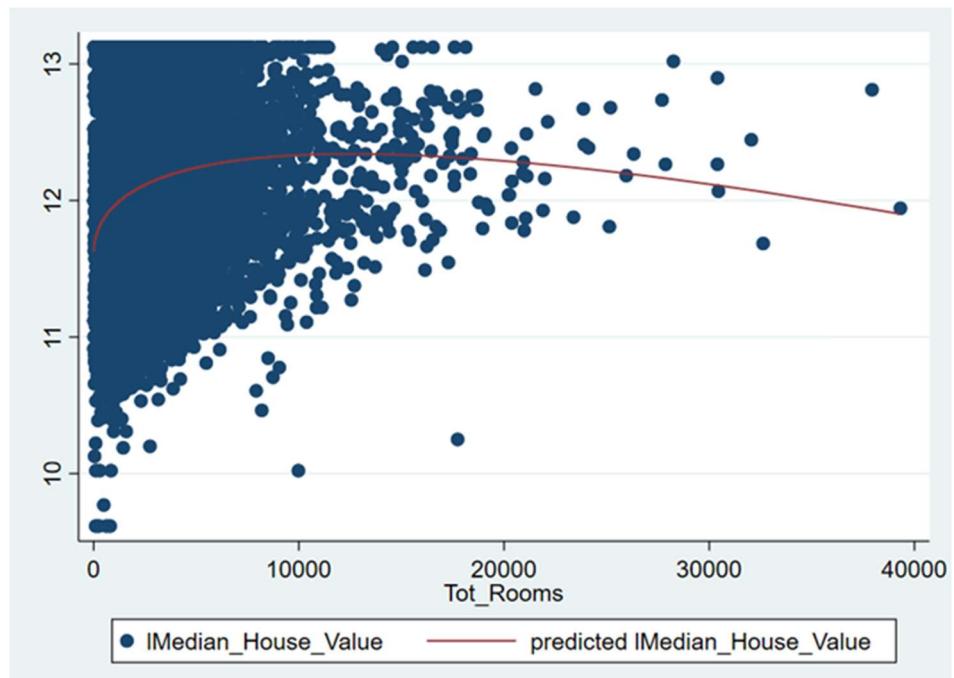


Gráfico de Dispersão (log – lin) 4 – Total Bedrooms

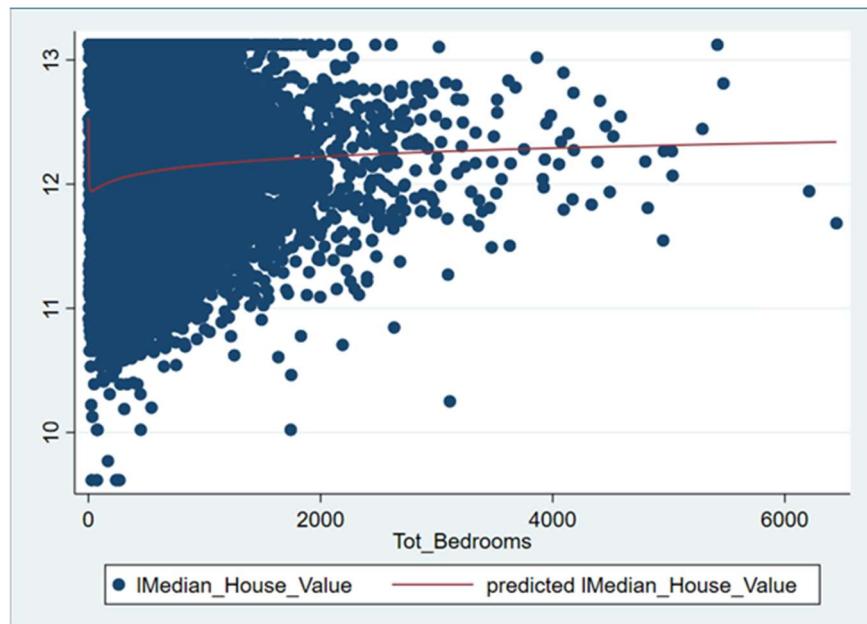


Gráfico de Dispersão (log – lin) 5 – Population

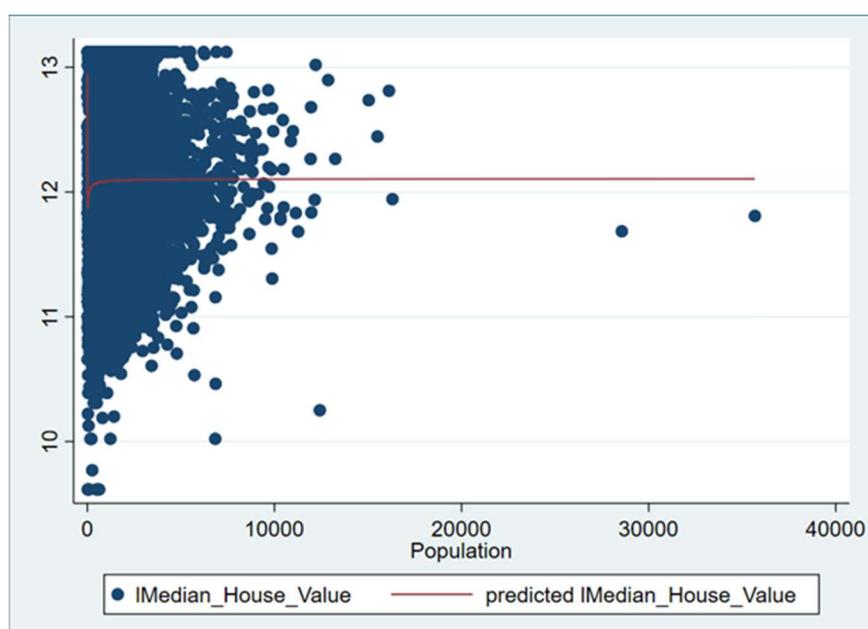


Gráfico de Dispersão (log – lin) 6 – Household

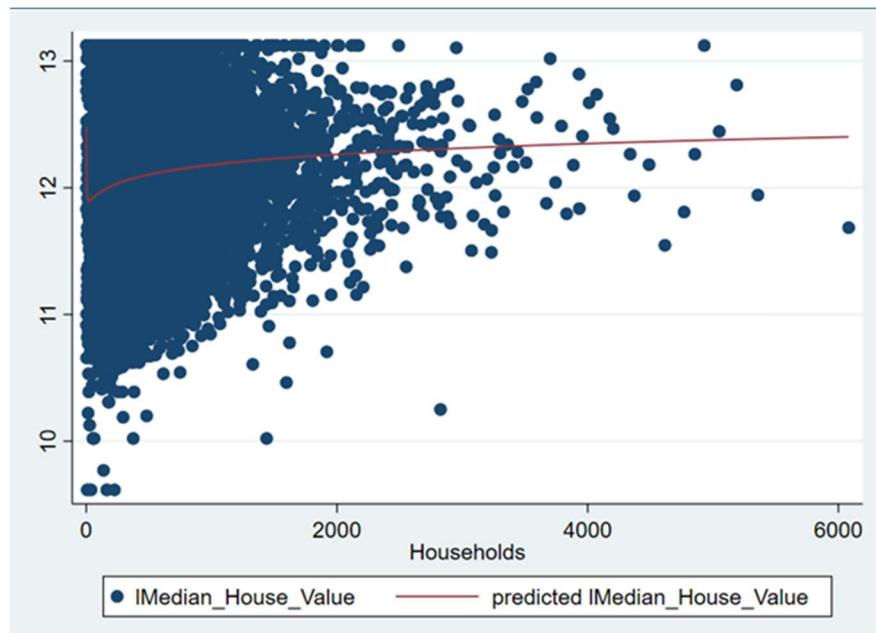


Gráfico de Dispersão (log – lin) 7 – Distance to Coast

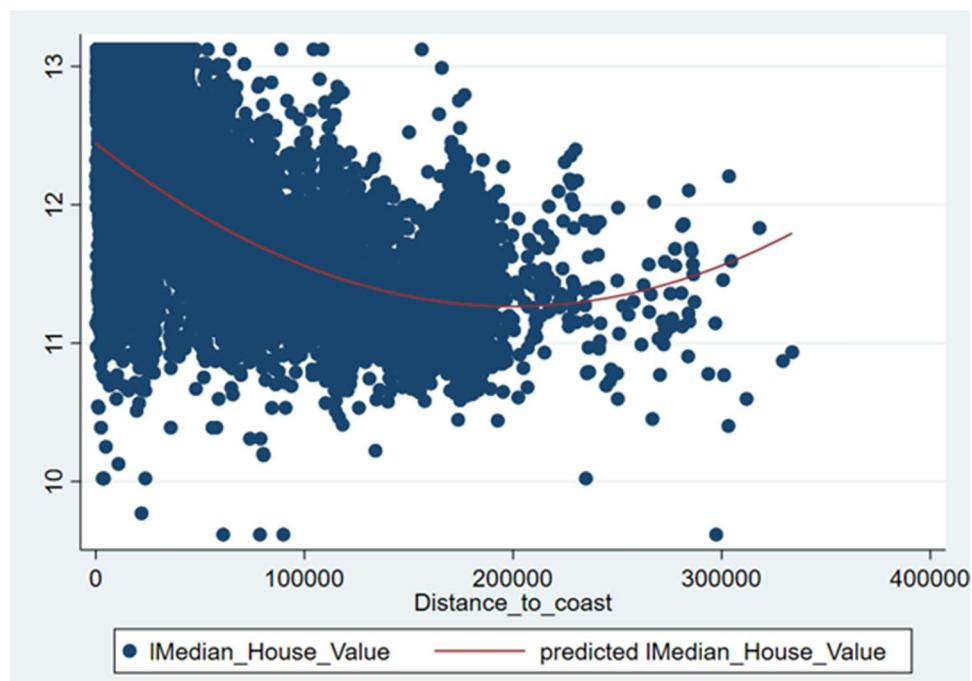


Gráfico de Dispersão (log – lin) 8 – Distance to LA

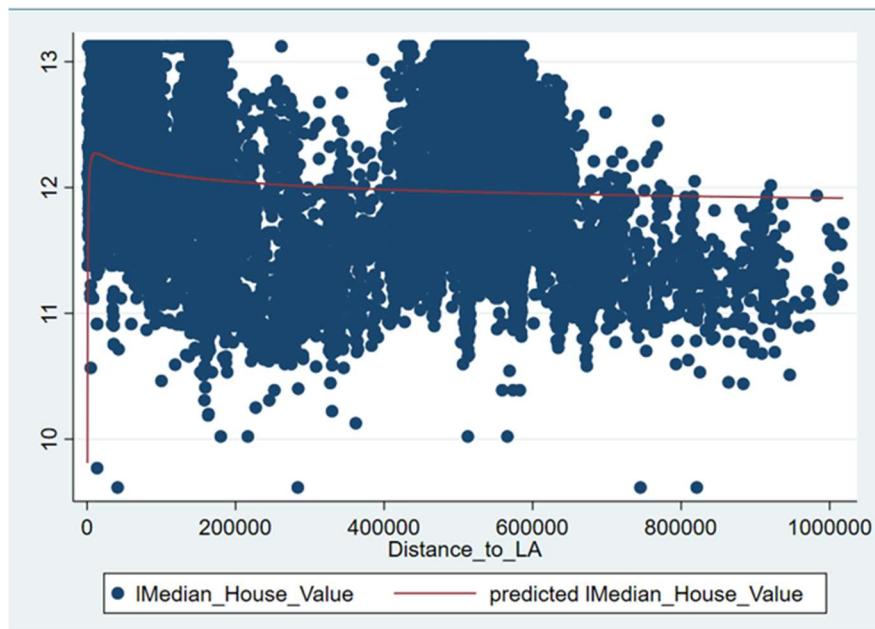


Gráfico de Dispersão (log – lin) 9 – Distance to San Diego

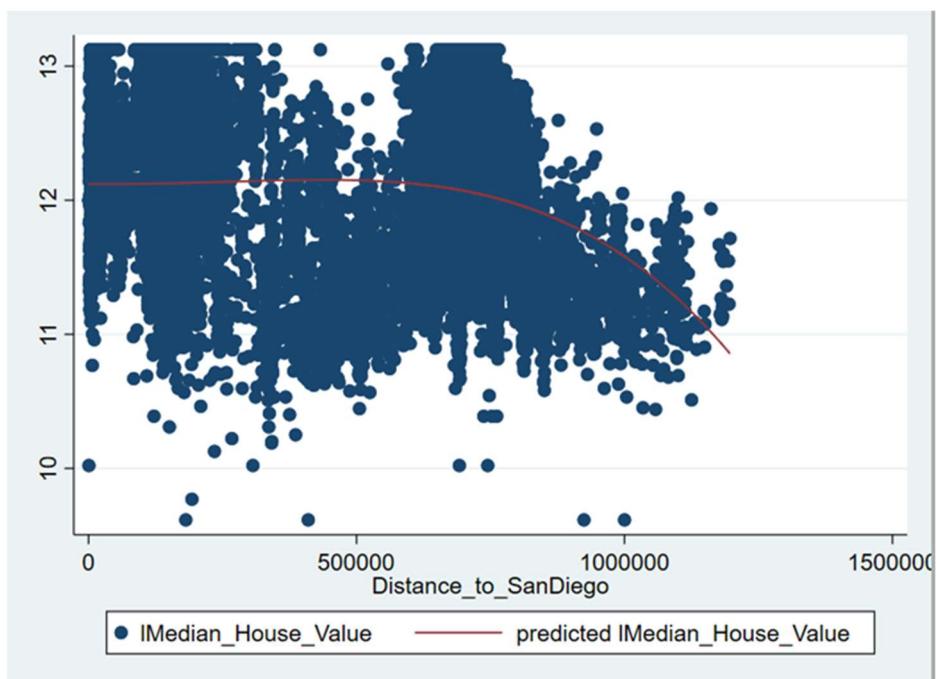


Gráfico de Dispersão (log – lin) 10 – Distance to San José

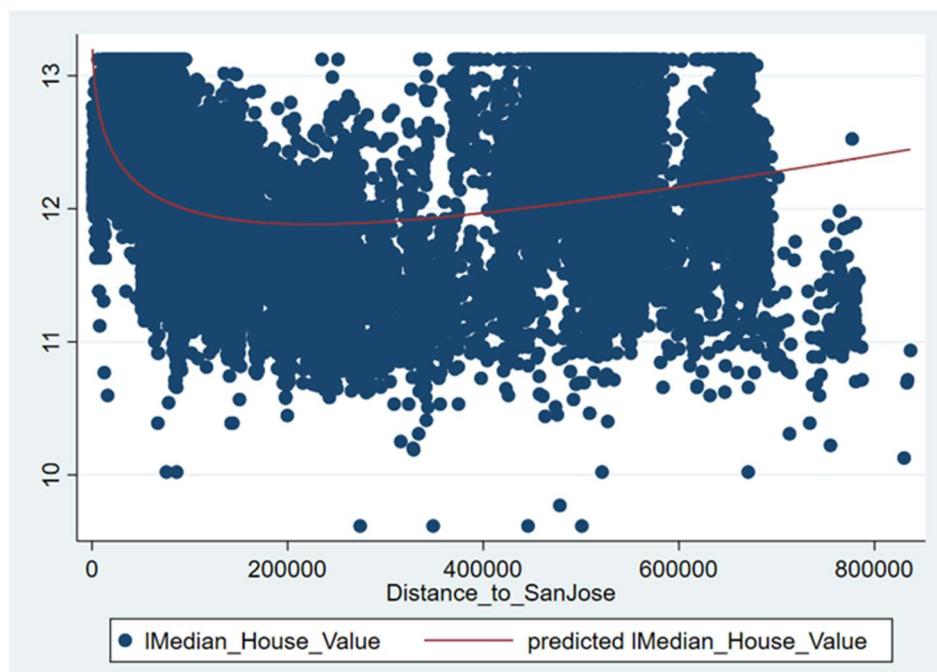
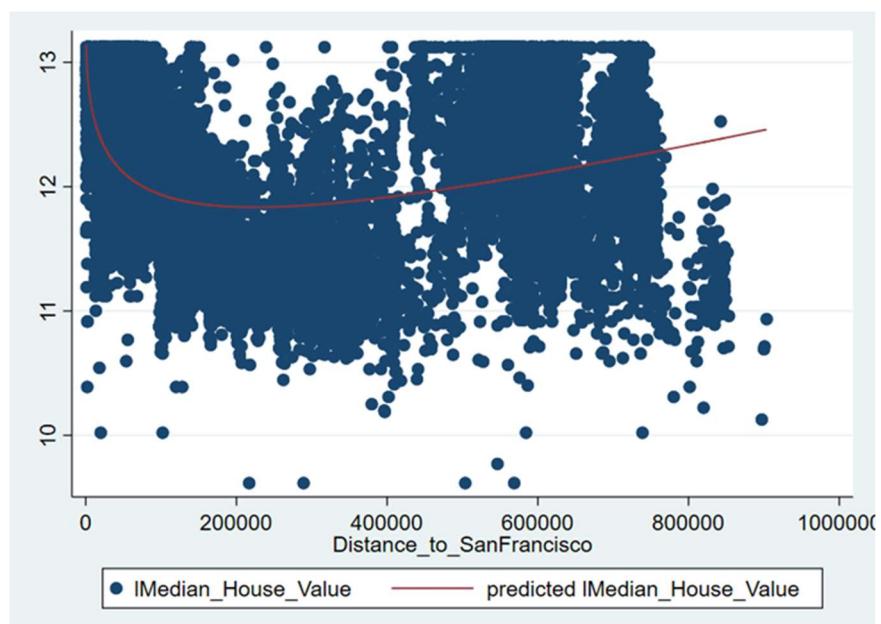


Gráfico de Dispersão (log – lin) 11 – Distance to San Francisco



Gráficos de Dispersão Lin – Log:

Gráfico de Dispersão (lin – log) 1 – Median Income

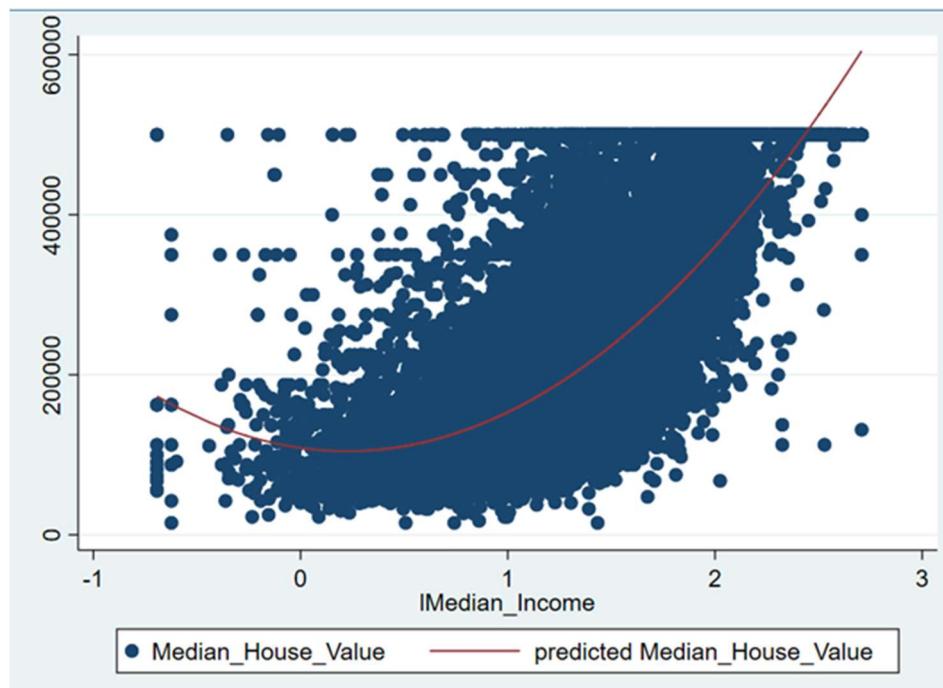


Gráfico de Dispersão (lin – log) 2 – Median Age

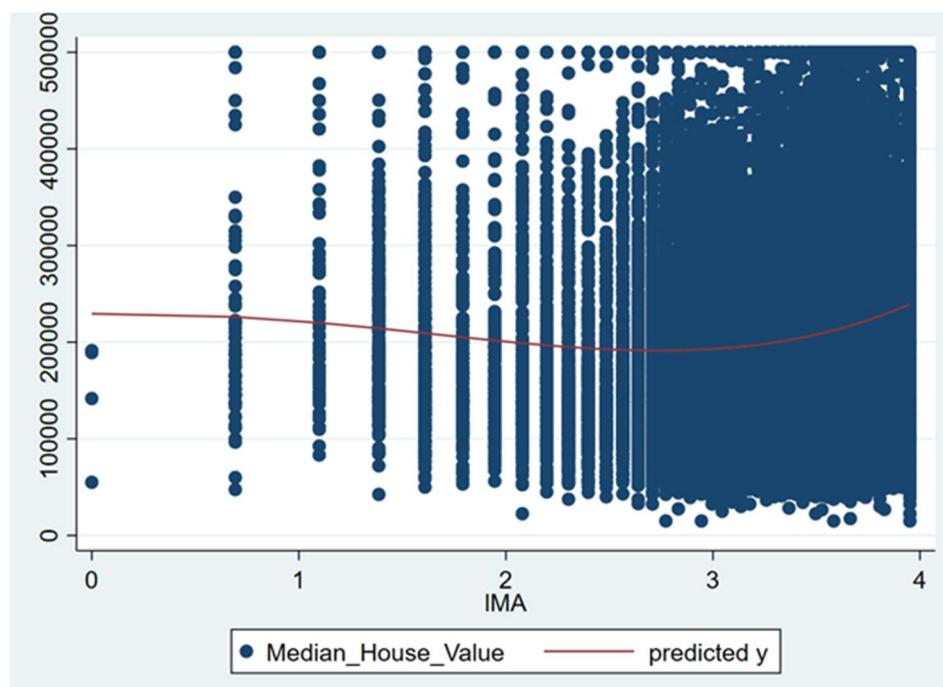


Gráfico de Dispersão (lin – log) 3 – Total Rooms

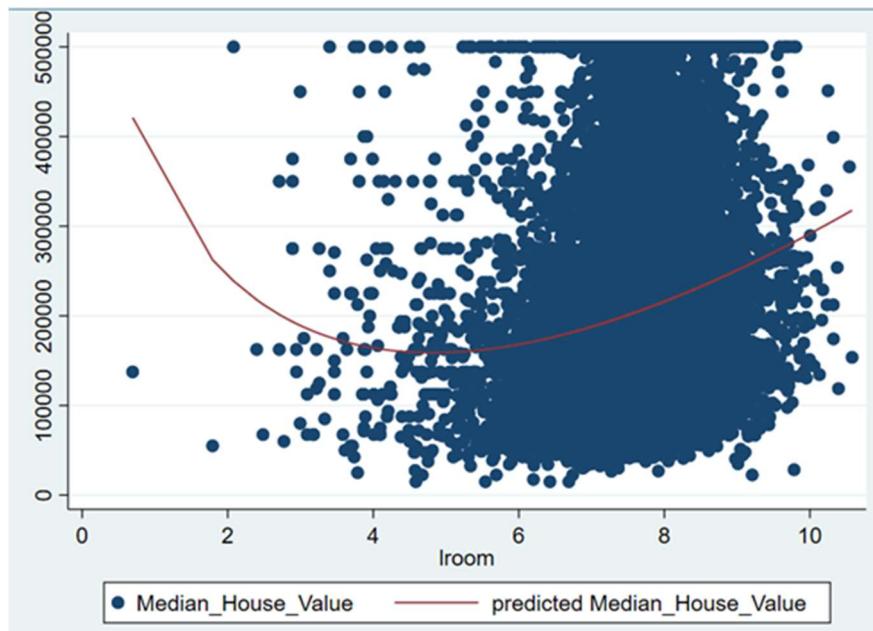


Gráfico de Dispersão (lin – log) 4 – Total Bedrooms

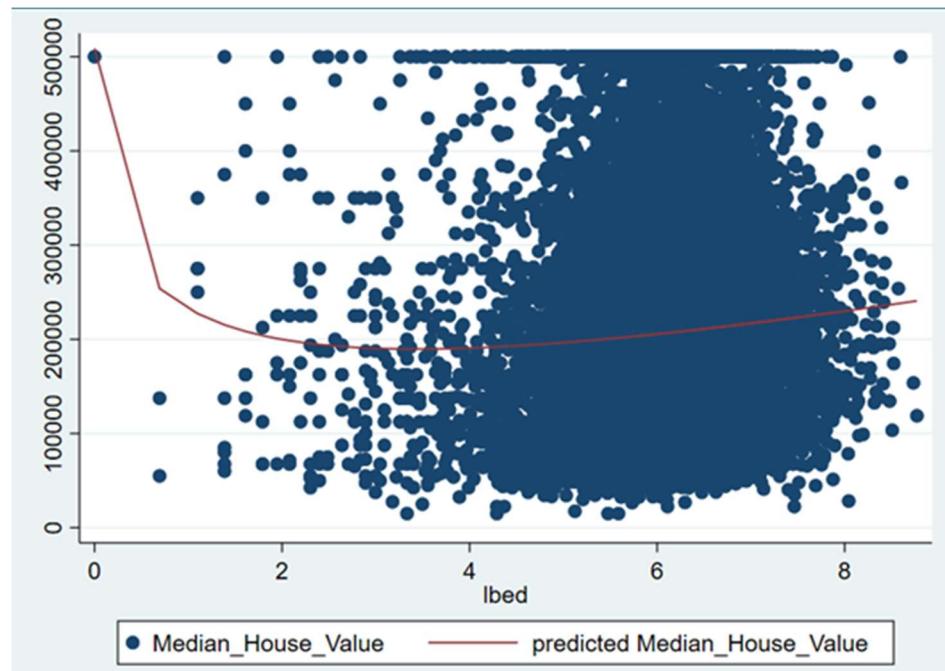


Gráfico de Dispersão (lin – log) 5 – Population

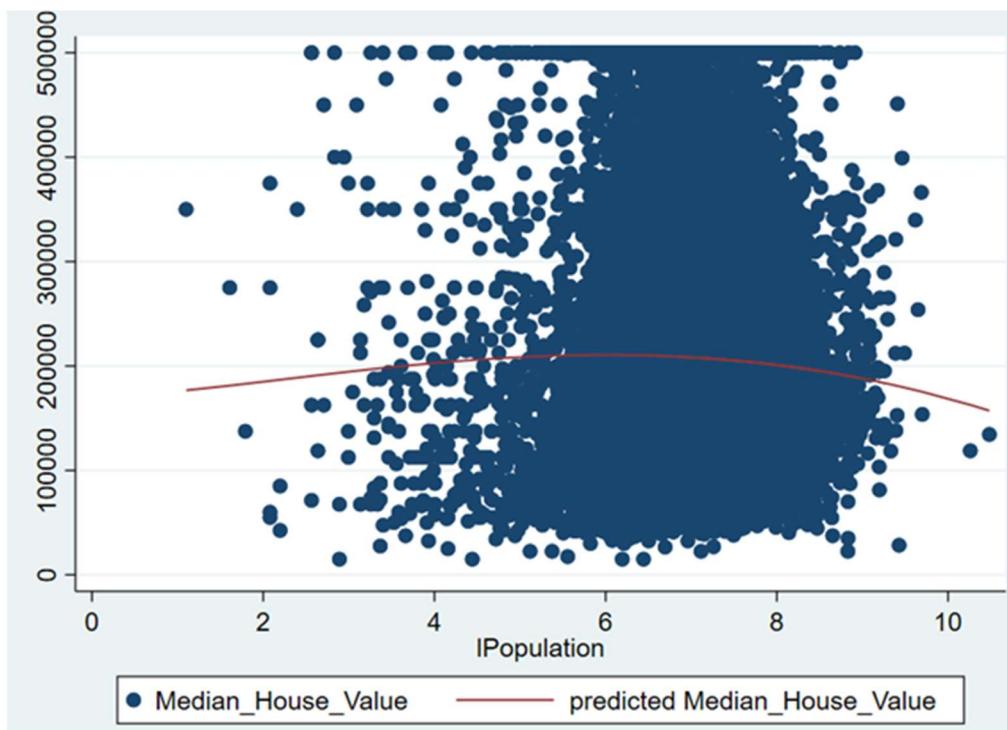


Gráfico de Dispersão (lin – log) 6 – Household

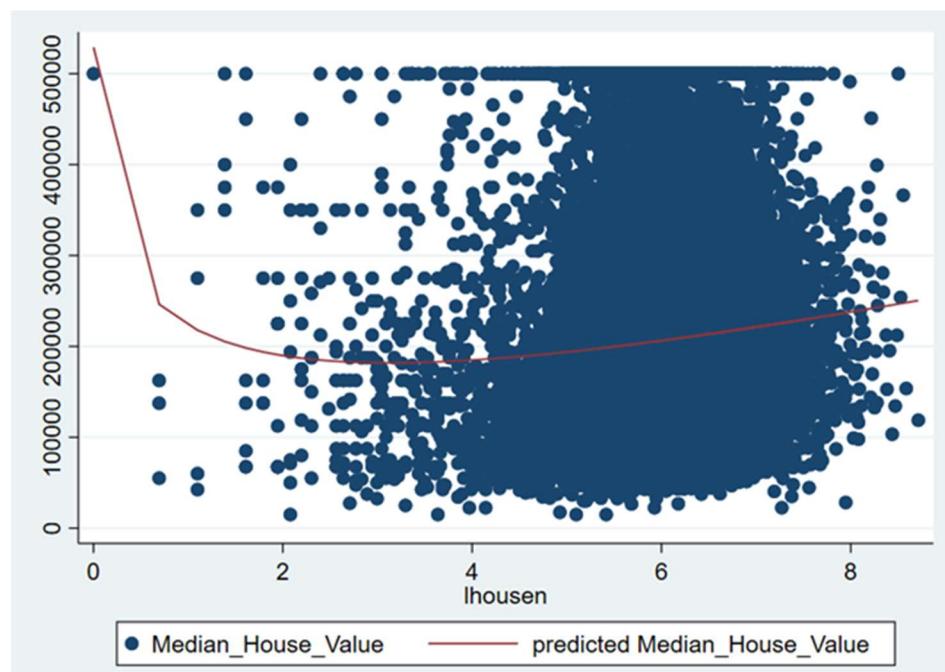


Gráfico de Dispersão (lin – log) 7 – Distance to Coast

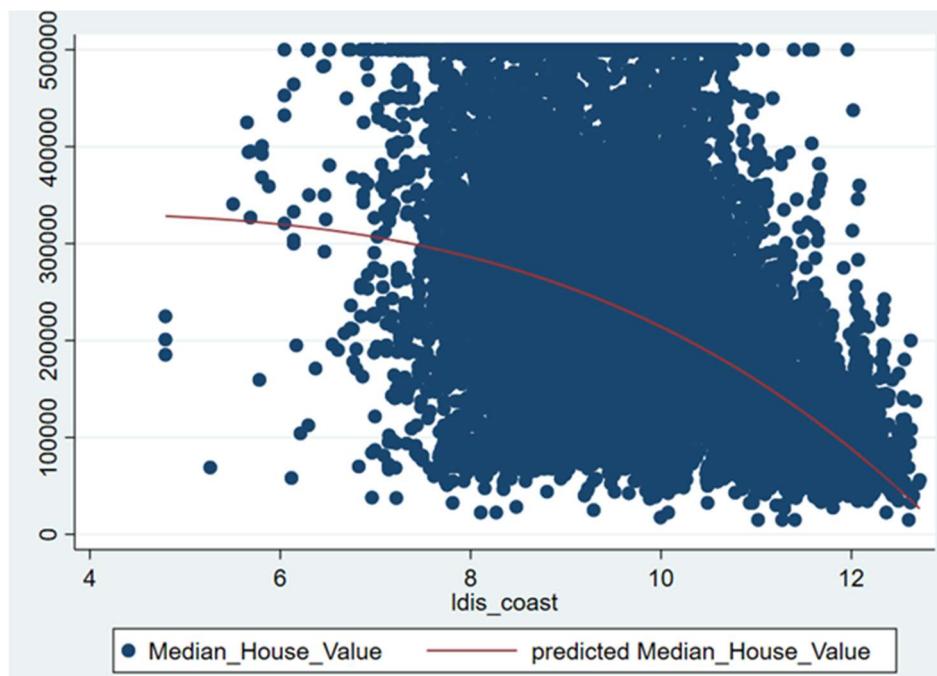


Gráfico de Dispersão (lin – log) 8 – Distance to LA

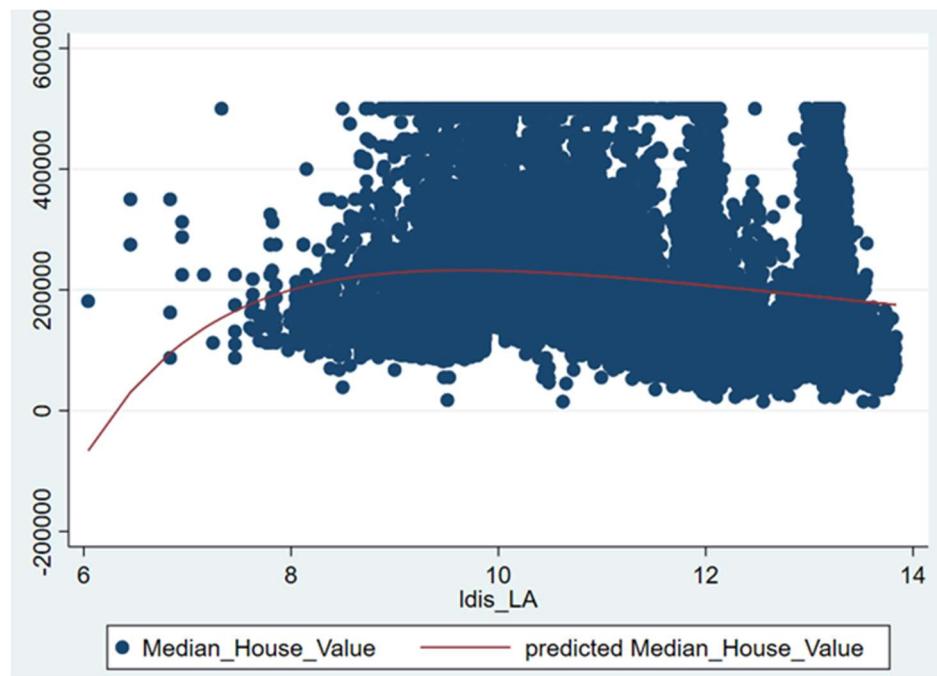


Gráfico de Dispersão (lin – log) 9 – Distance to San Diego

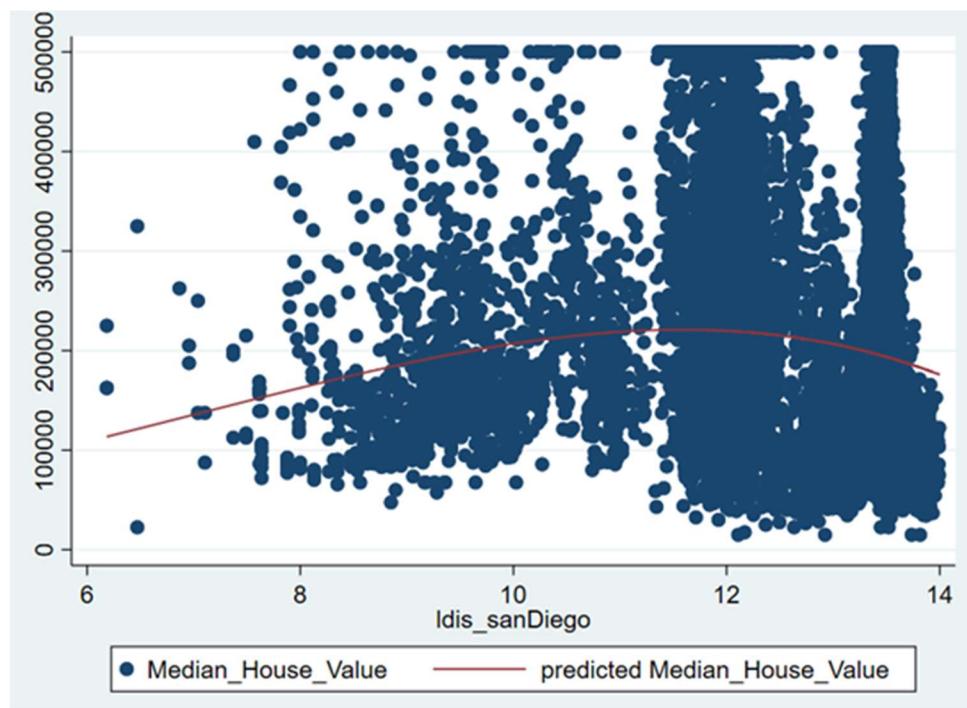


Gráfico de Dispersão (lin – log) 10 – Distance to San José

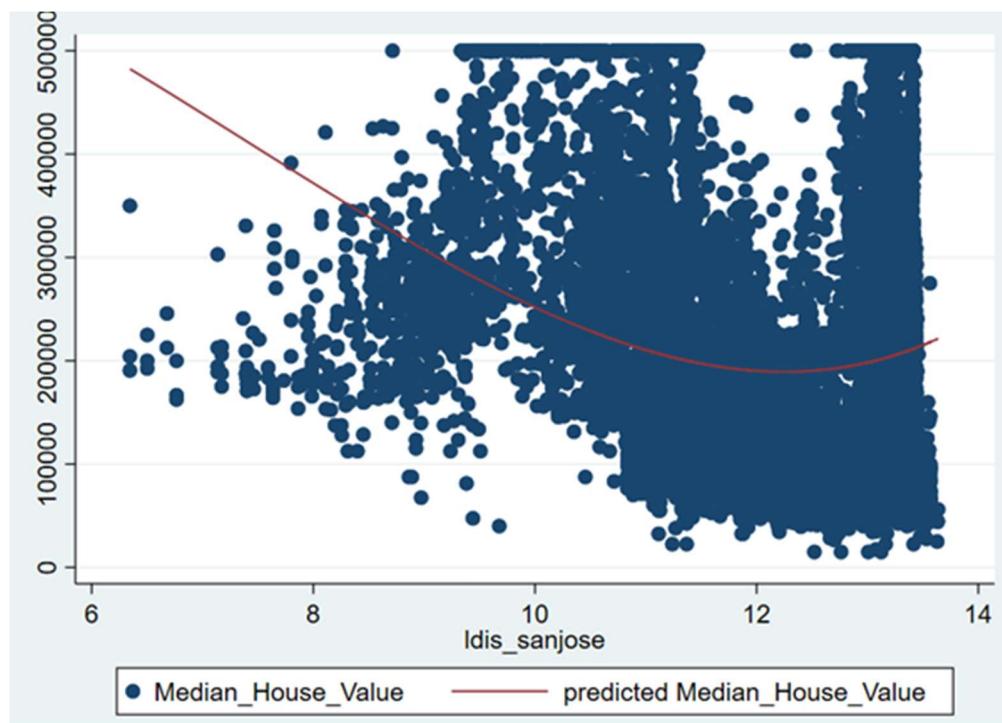
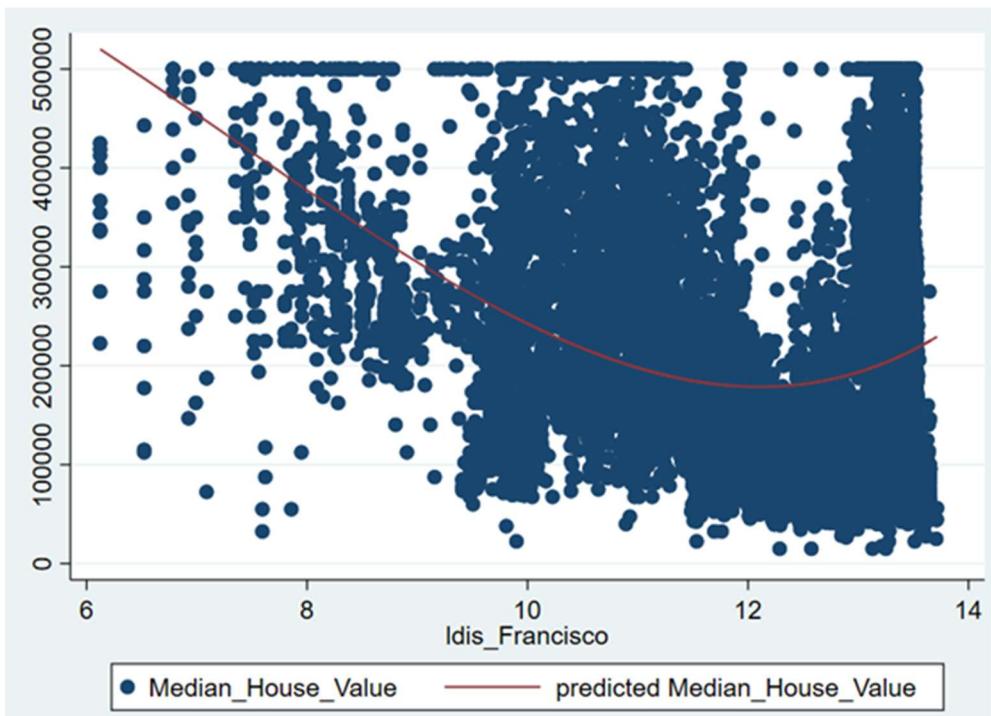


Gráfico de Dispersão (lin – log) 11 – Distance to San Francisco



Gráficos de Dispersão Log – Log:

Gráfico de Dispersão (log – log) 1 – Median Income

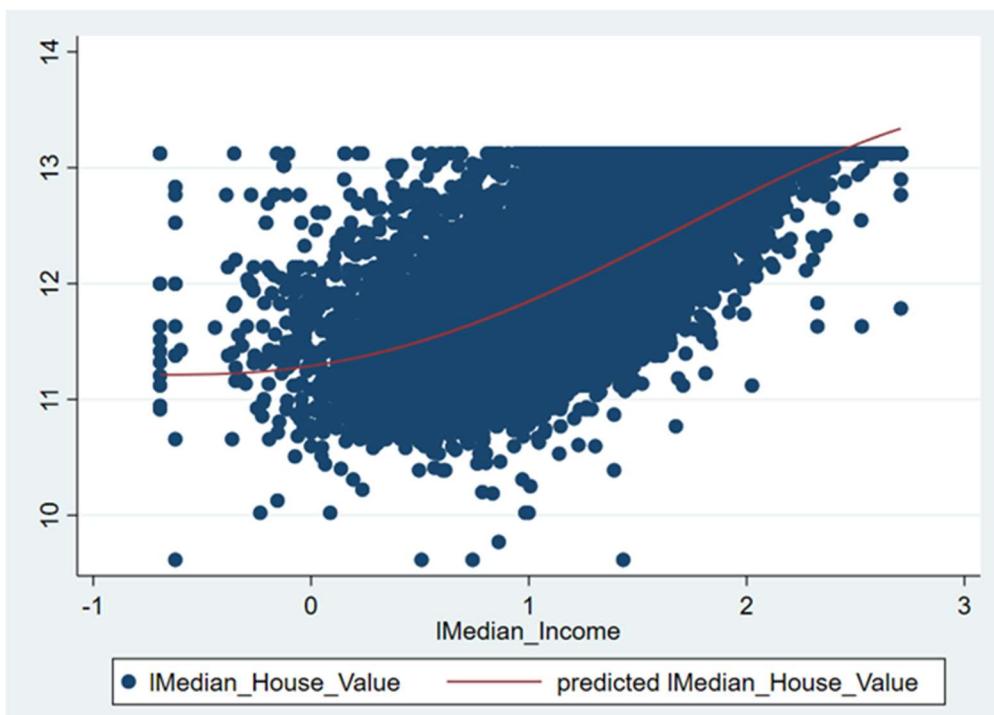


Gráfico de Dispersão (log – log) 2 – Median Age

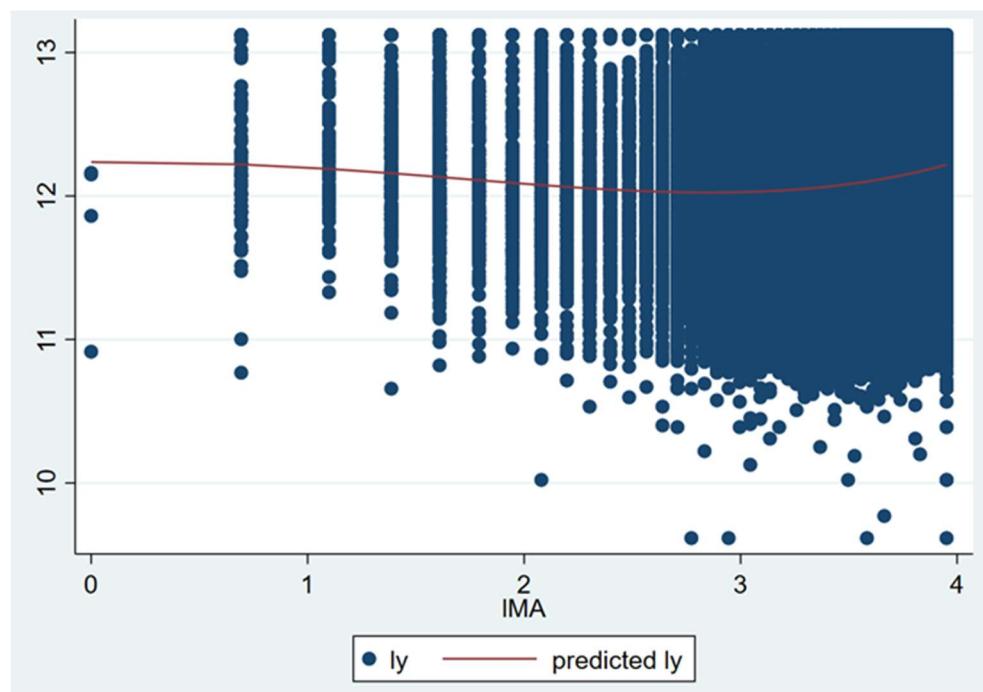


Gráfico de Dispersão (log – log) 3 – Total Rooms

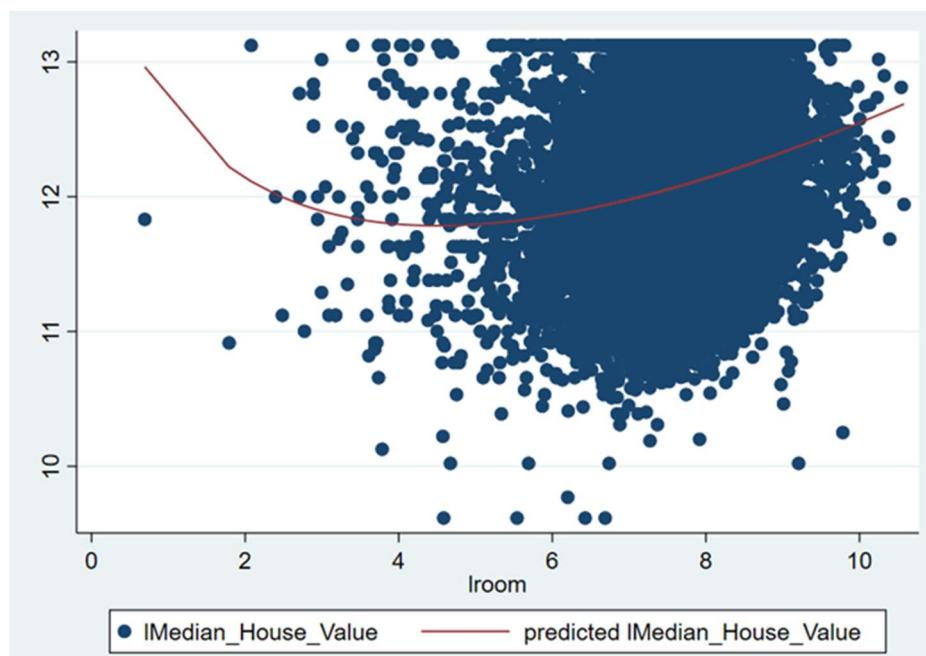


Gráfico de Dispersão (log – log) 4 – Total Bedrooms

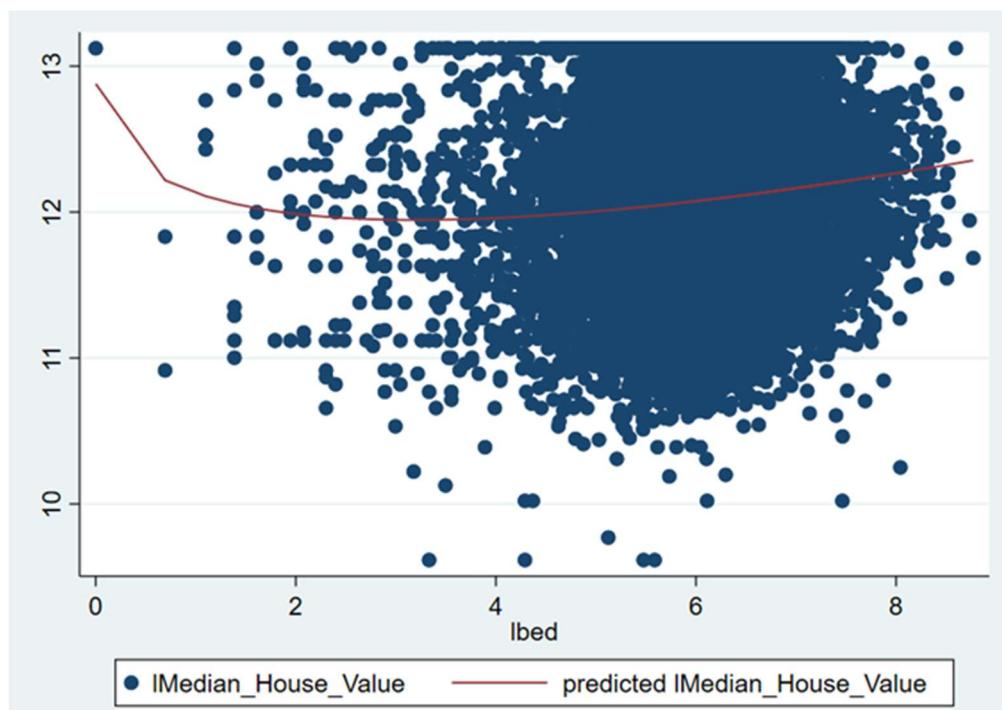


Gráfico de Dispersão (log – log) 5 – Population

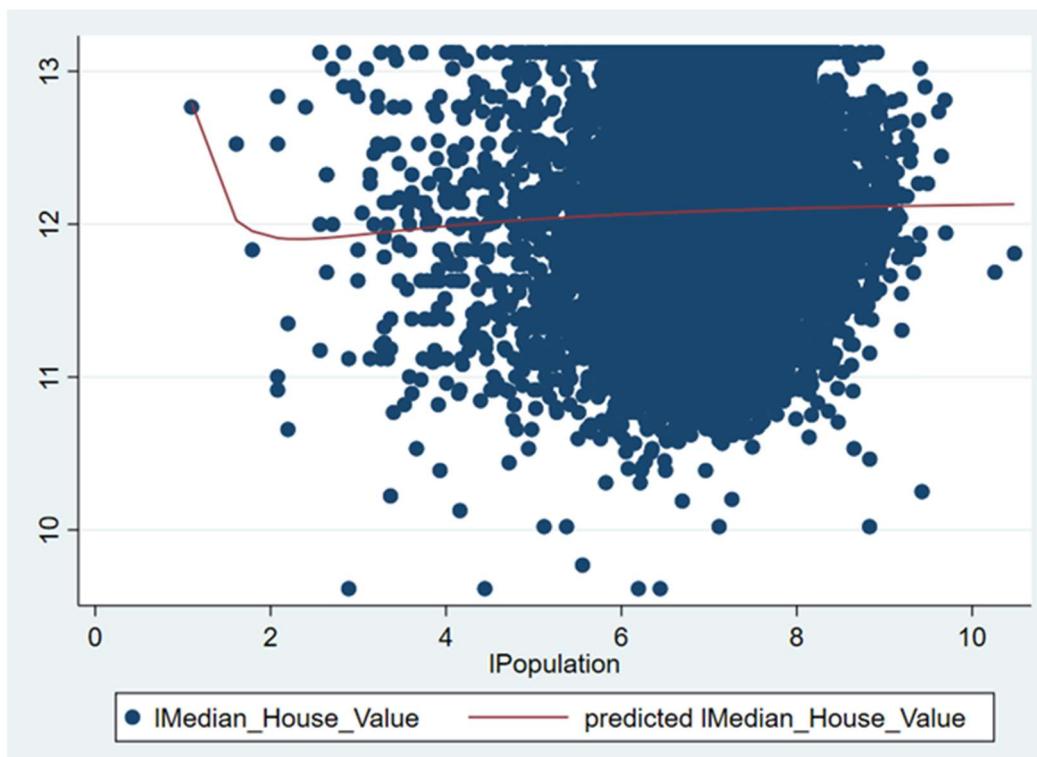


Gráfico de Dispersão (log – log) 6 – Households

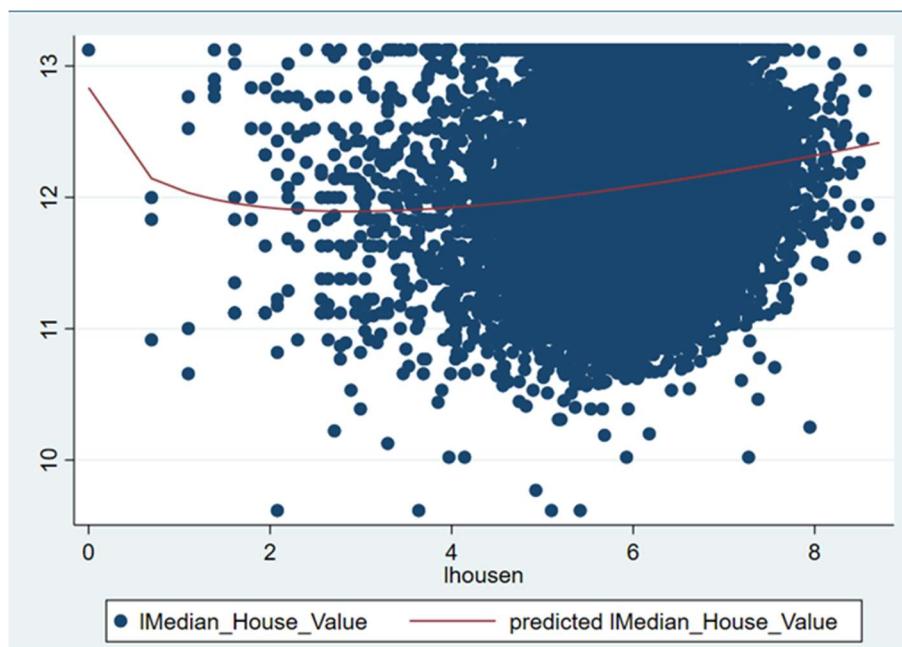


Gráfico de Dispersão (log – log) 7 – Distance to Coast

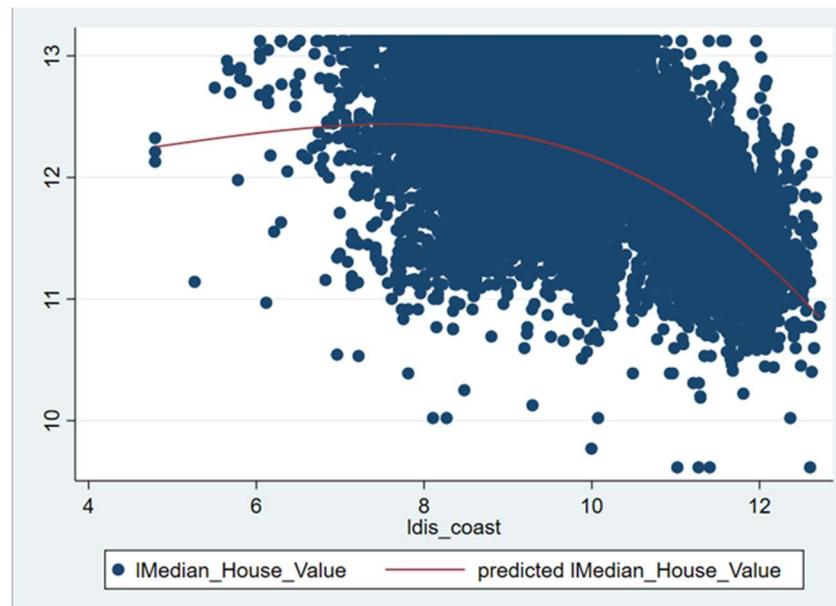


Gráfico de Dispersão (log – log) 8 – Distance to LA

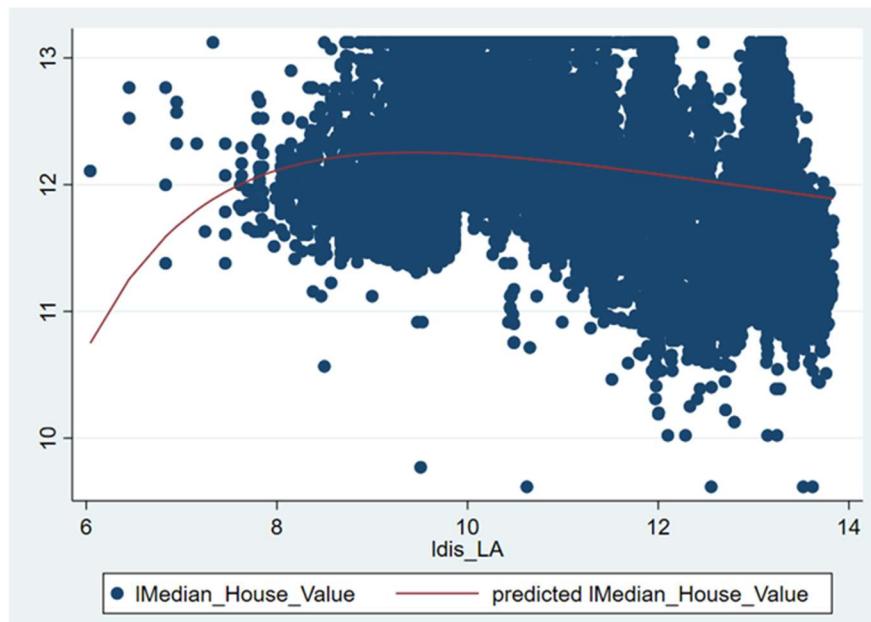


Gráfico de Dispersão (log – log) 9 – Distance to San Diego

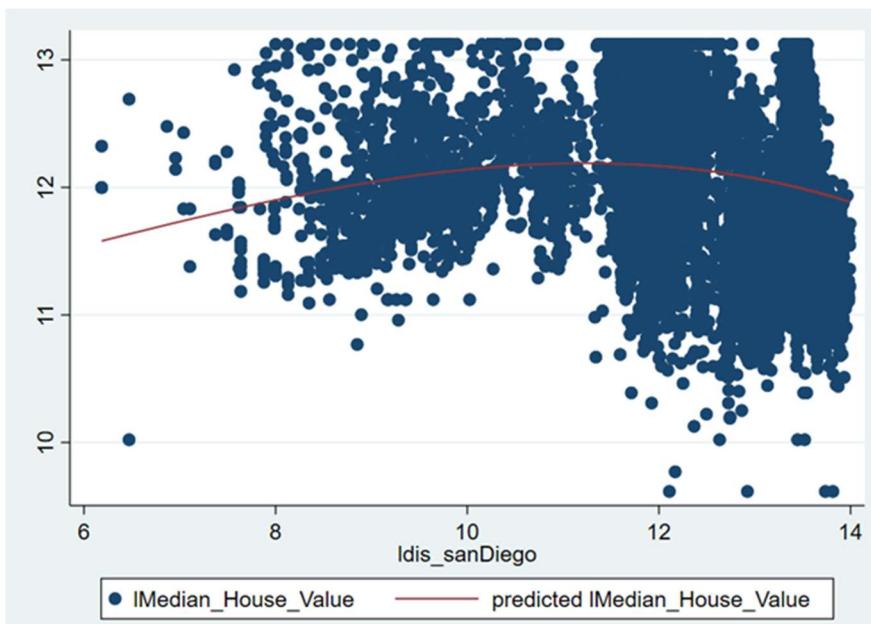


Gráfico de Dispersão (log – log) 10 – Distance to San Jose

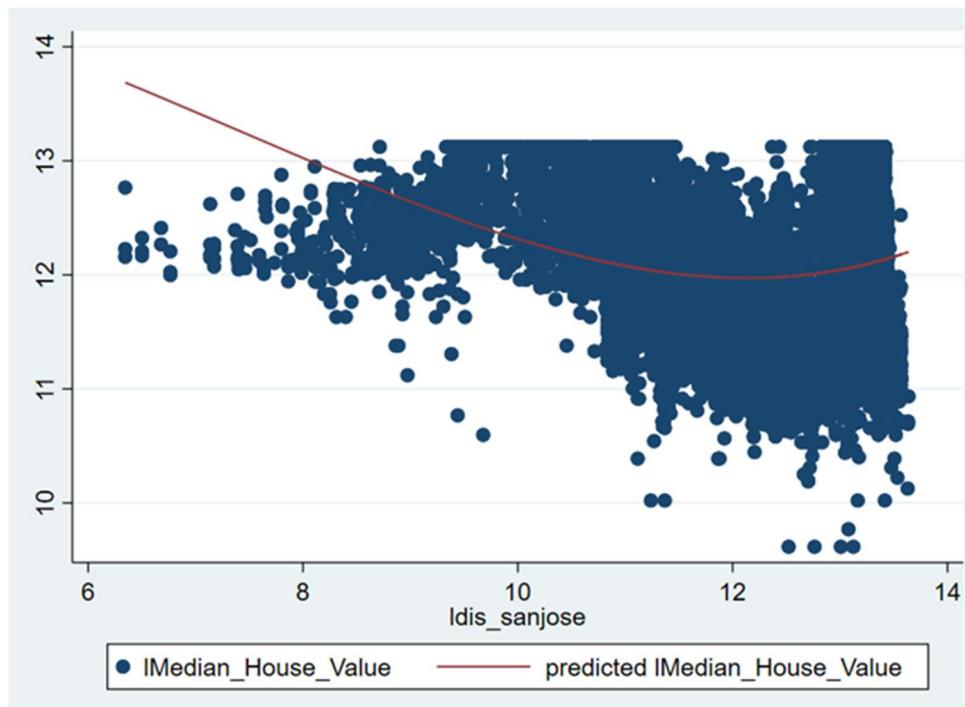
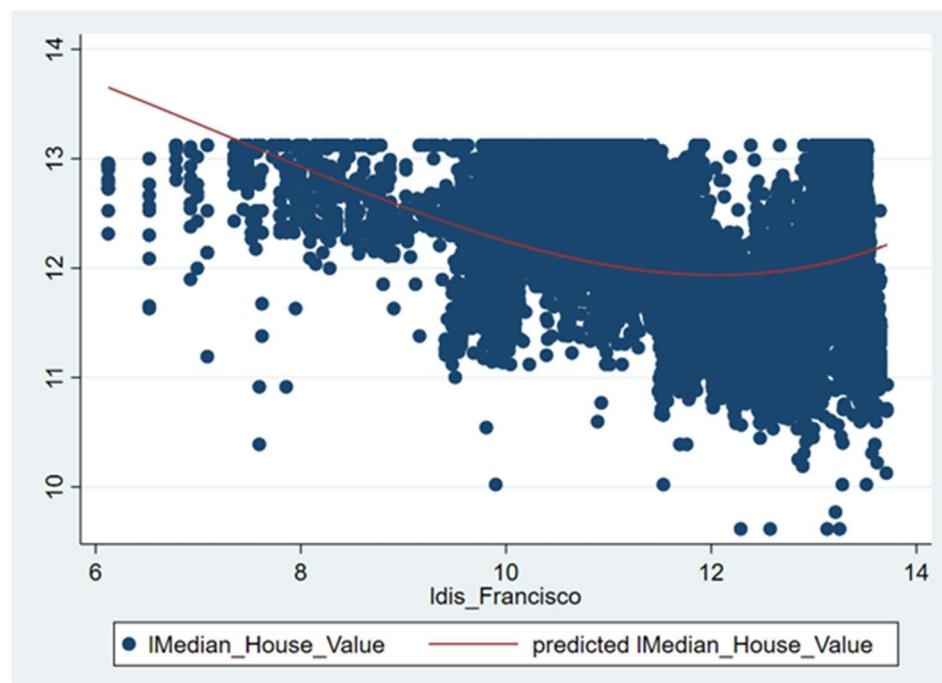


Gráfico de Dispersão (log – log) 11 – Distance to San Francisco



Formas Funcionais das Variáveis com a base de dados livres de Outliers

Gráfico de Dispersão sem valores aberrantes (Lin-Lin) 12 – Median Income

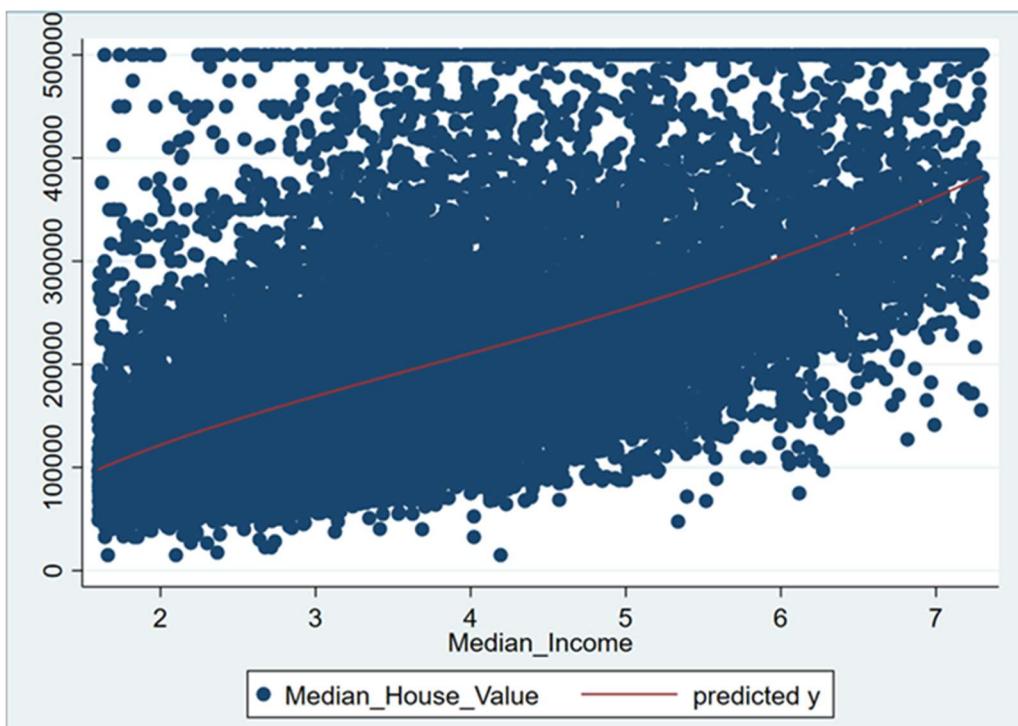
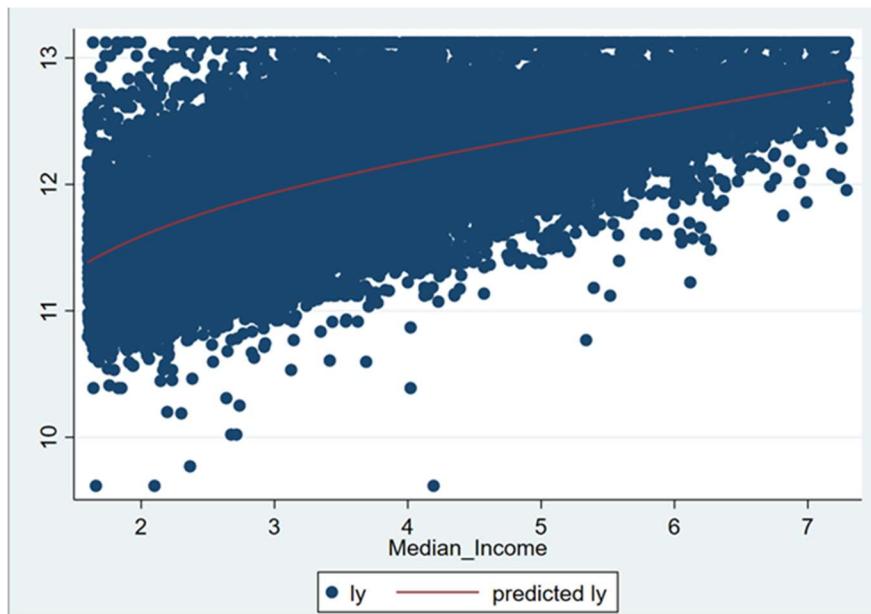
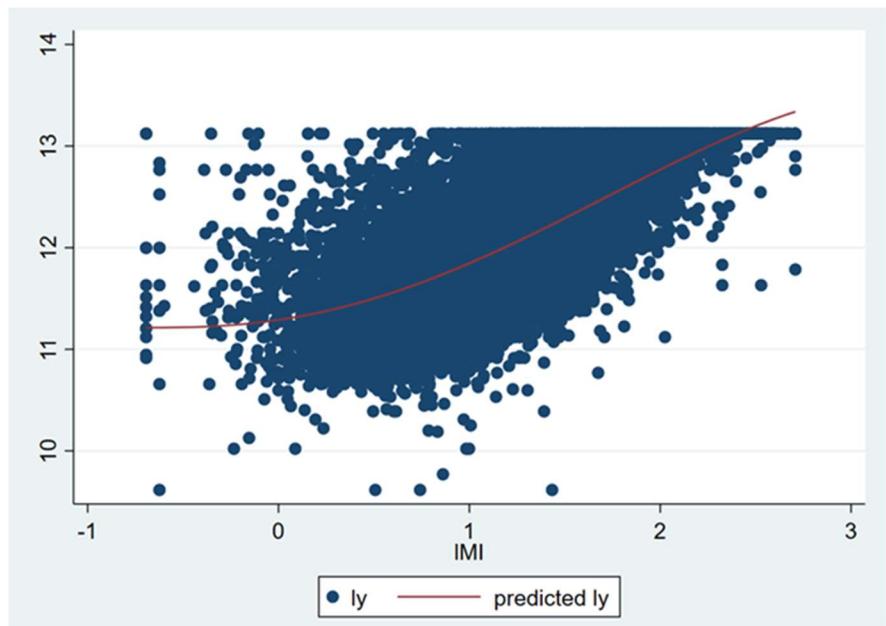


Gráfico de Dispersão sem valores aberrantes (Log-Lin) 13 – Median Income



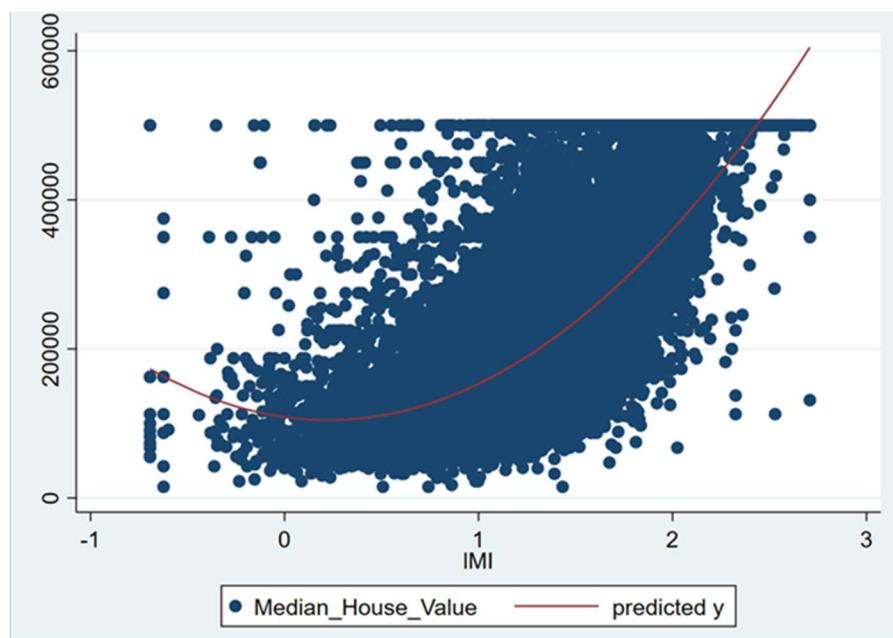
Log – Log

Gráfico de Dispersão sem valores aberrantes



Lin – Log

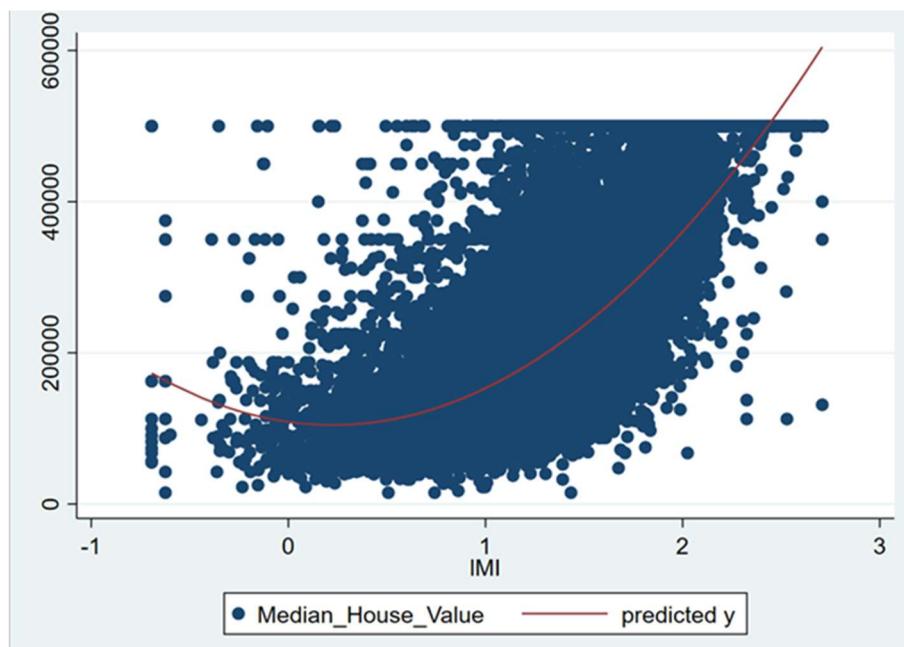
Gráfico de Dispersão sem valores aberrantes



1. Median Age

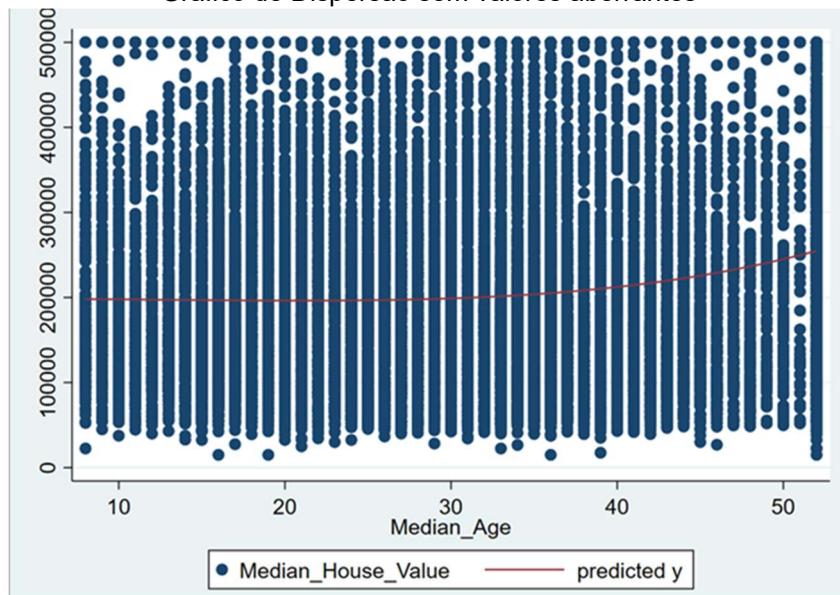
Lin – Lin

Gráfico de Dispersão



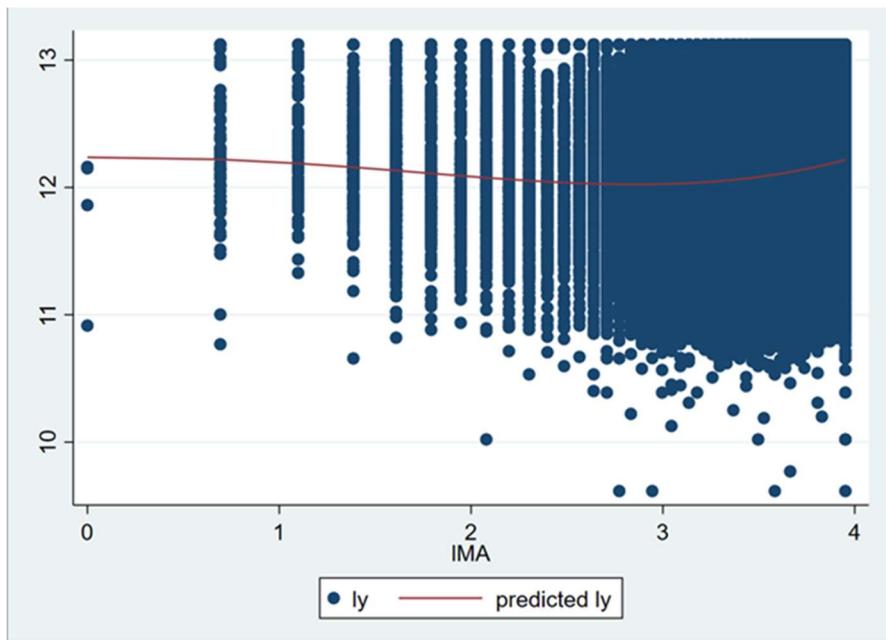
Log – Lin

Gráfico de Dispersão sem valores aberrantes



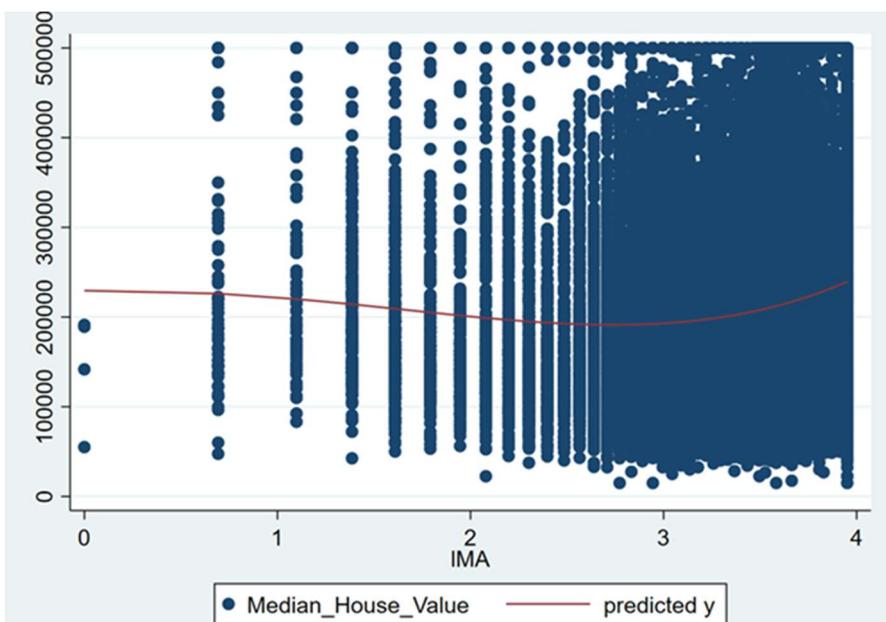
Log – Log

Gráfico de Dispersão sem valores aberrantes



Lin – Log

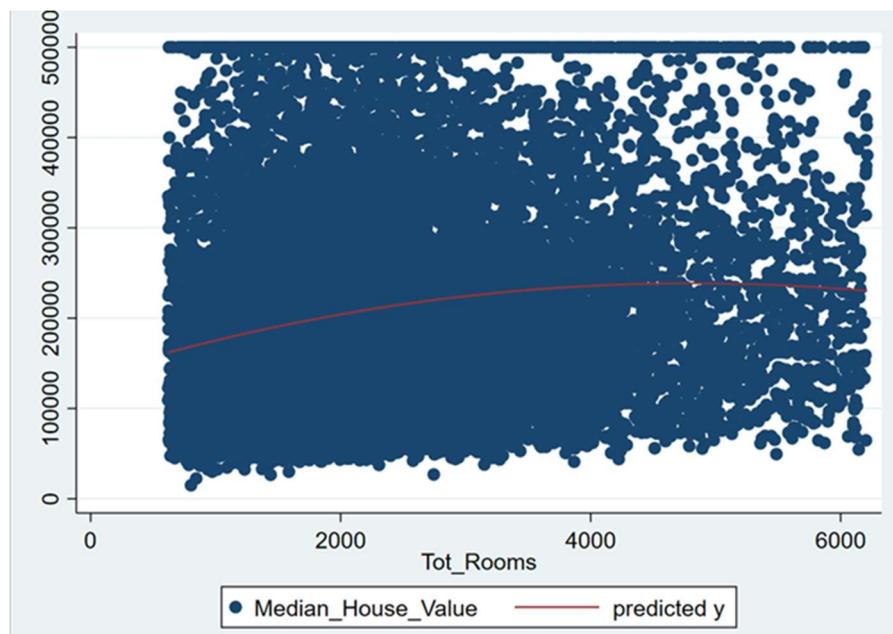
Gráfico de Dispersão



2. Total Rooms

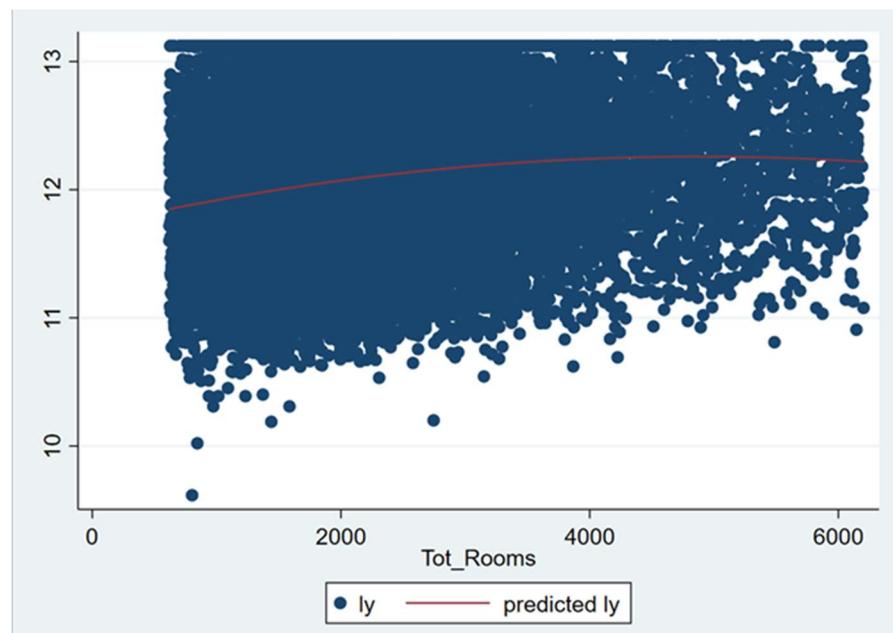
Lin – Lin

Gráfico de Dispersão



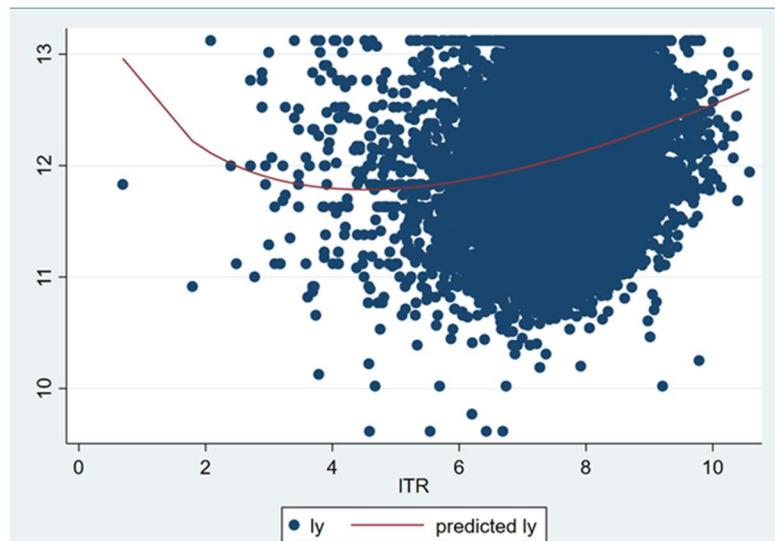
Log – Lin

Gráfico de Dispersão



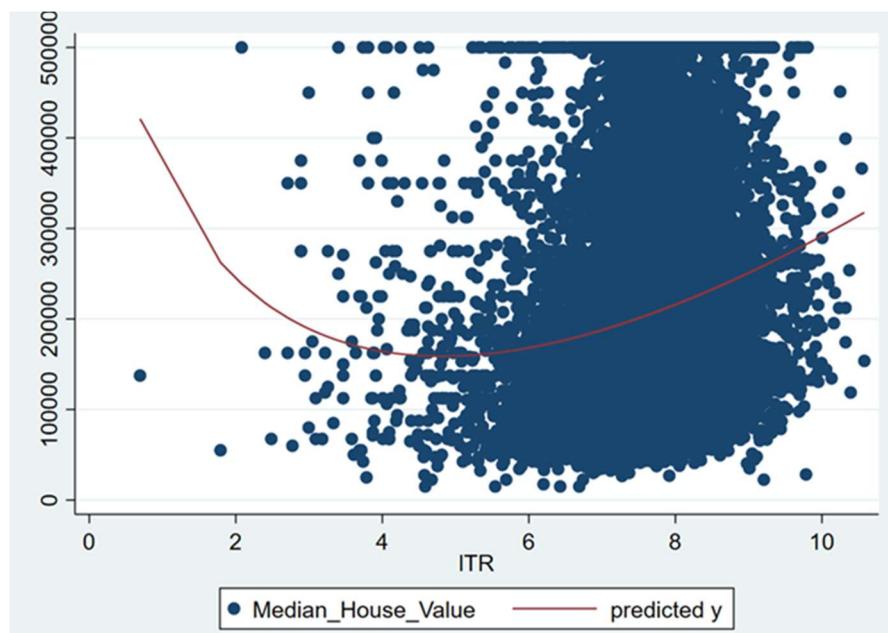
Log – Log

Gráfico de Dispersão



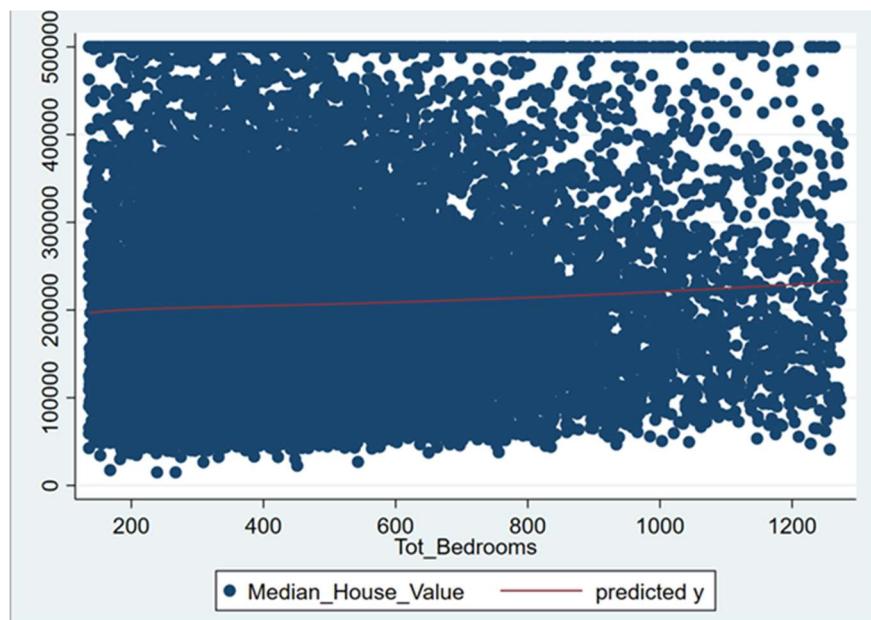
Lin – Log

Gráfico de Dispersão



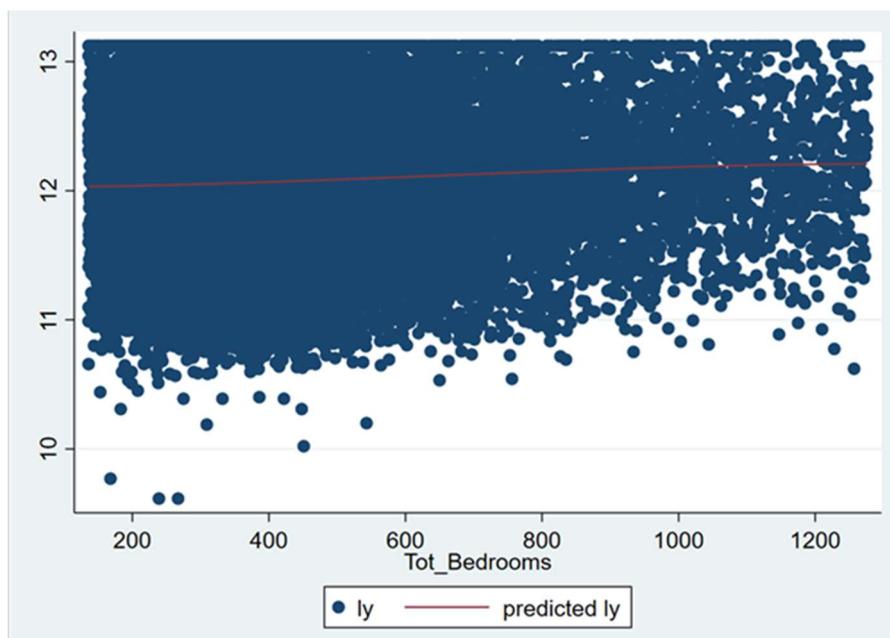
3. Total Bedrooms

Lin – Lin



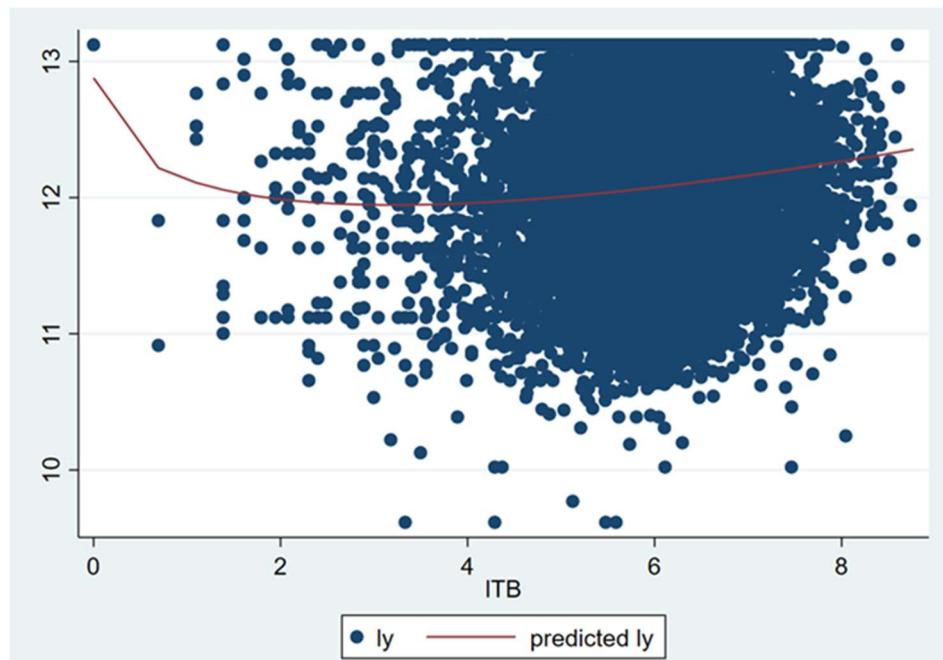
Log – Lin

Gráfico de Dispersão



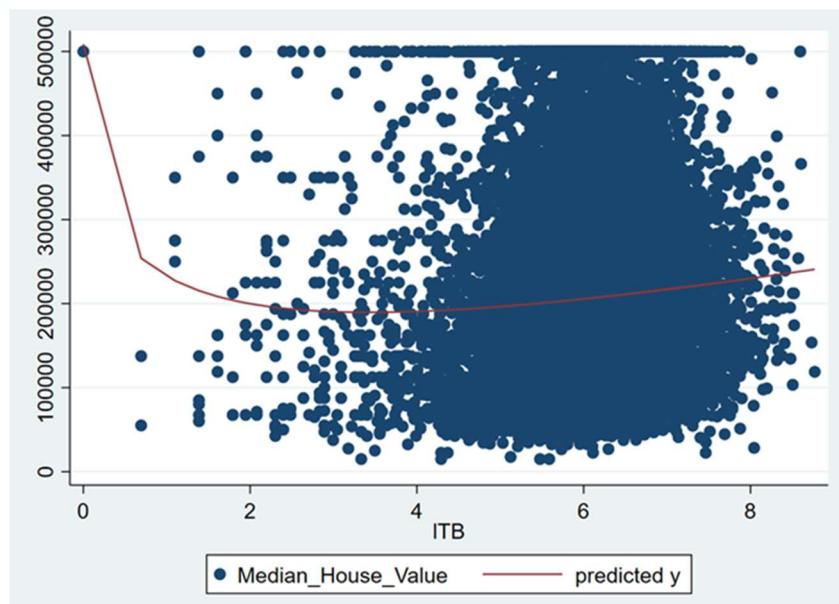
Log – Log

Gráfico de Dispersão



Lin – Log

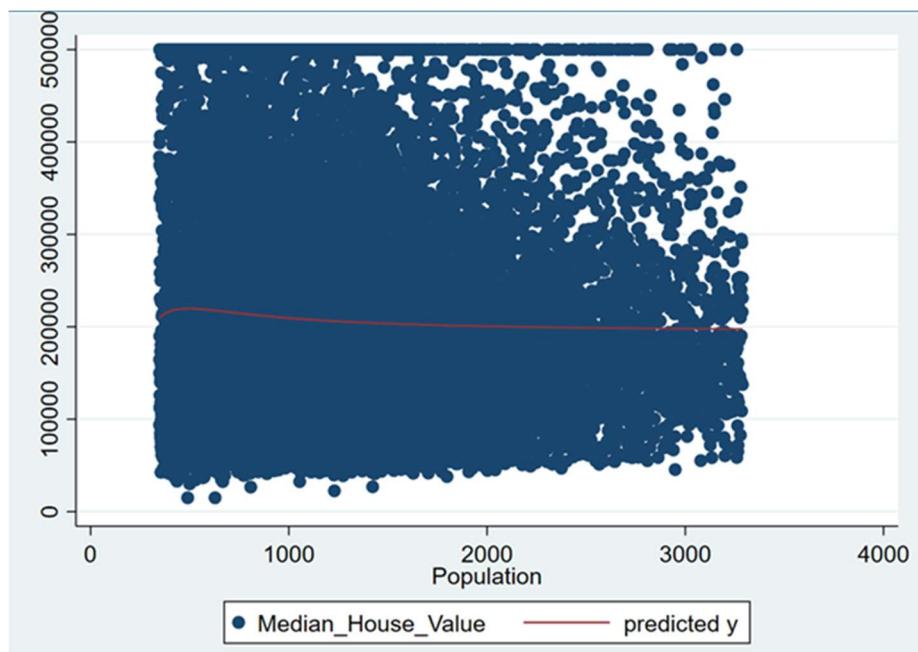
Gráfico de Dispersão



4. Population

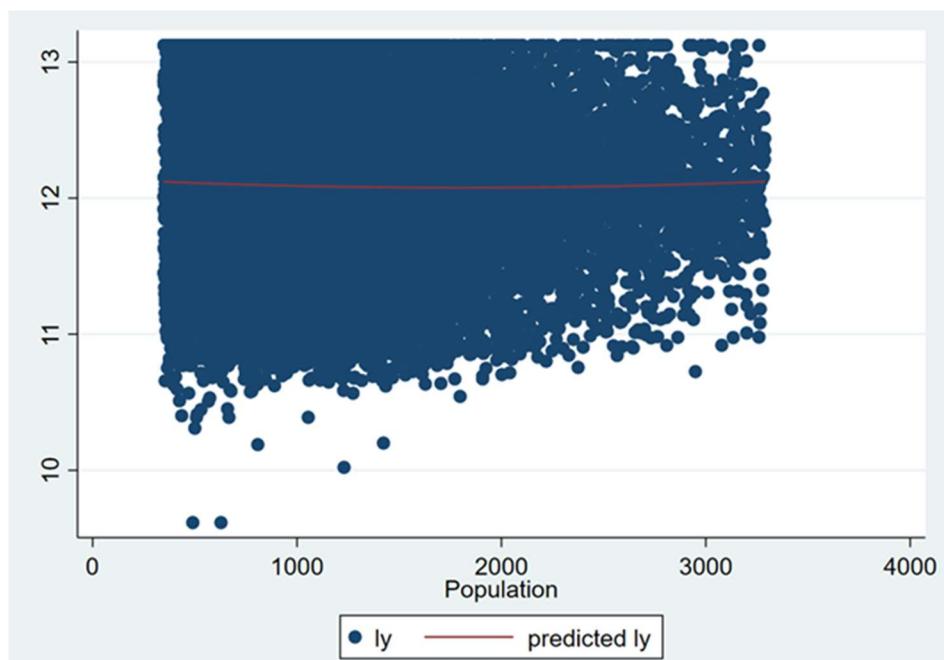
Lin – Lin

Gráfico de Dispersão



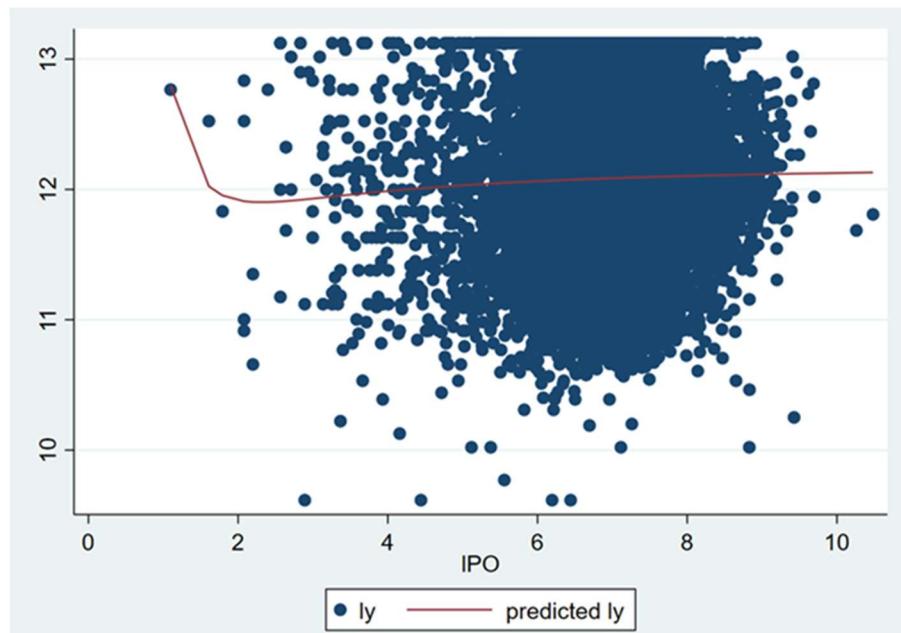
Log – Lin

Gráfico de Dispersão



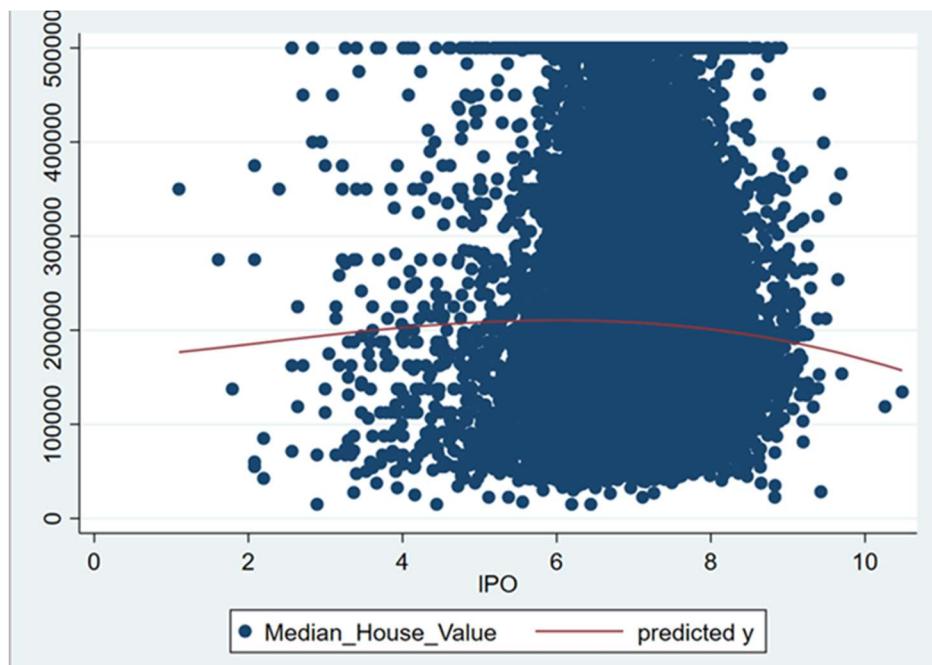
Log – Log

Gráfico de Dispersão



Lin – Log

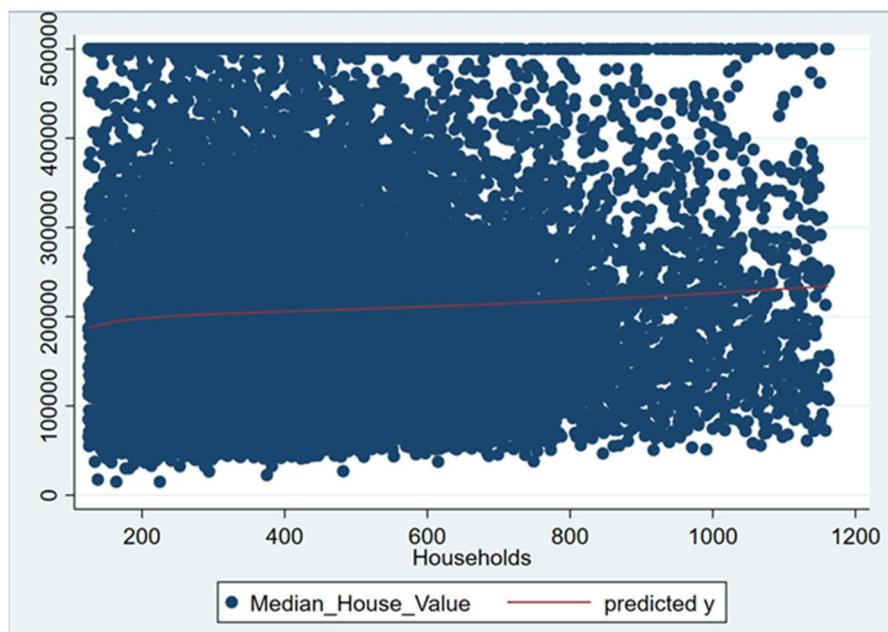
Gráfico de Dispersão



5. Household

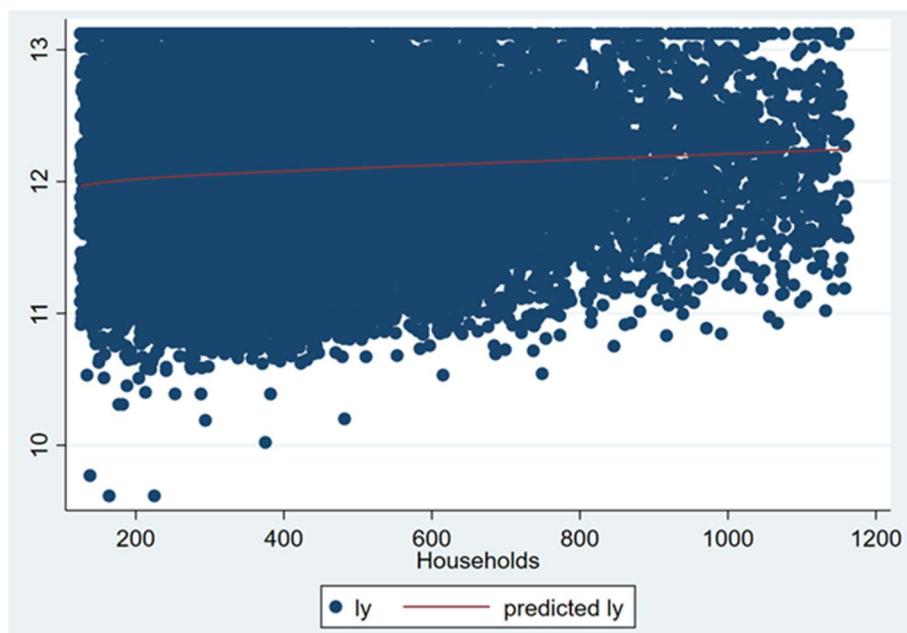
Lin – Lin

Gráfico de Dispersão



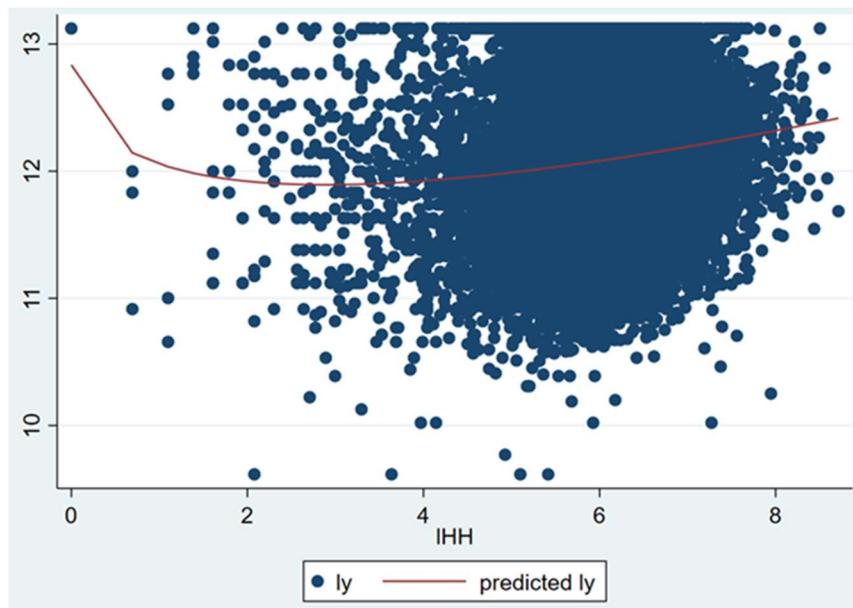
Log – Lin

Gráfico de Dispersão



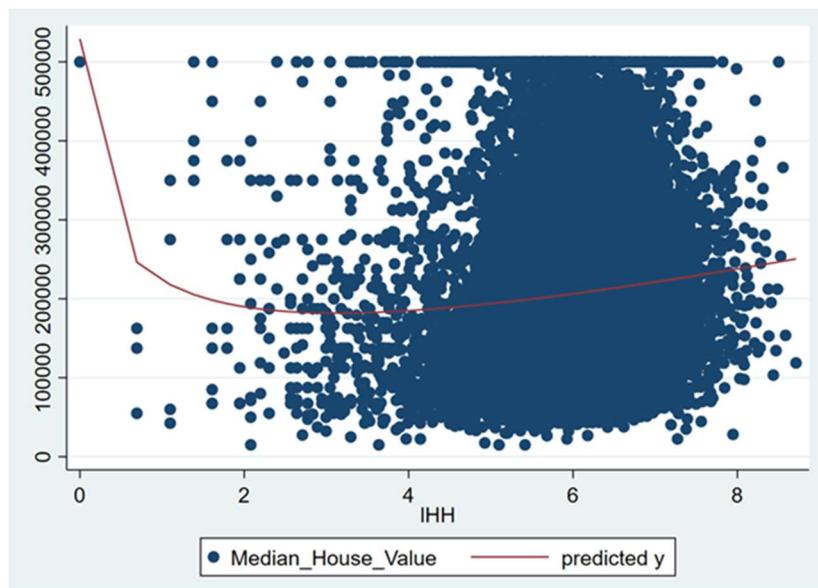
Log– Log

Gráfico de Dispersão



Lin – Log

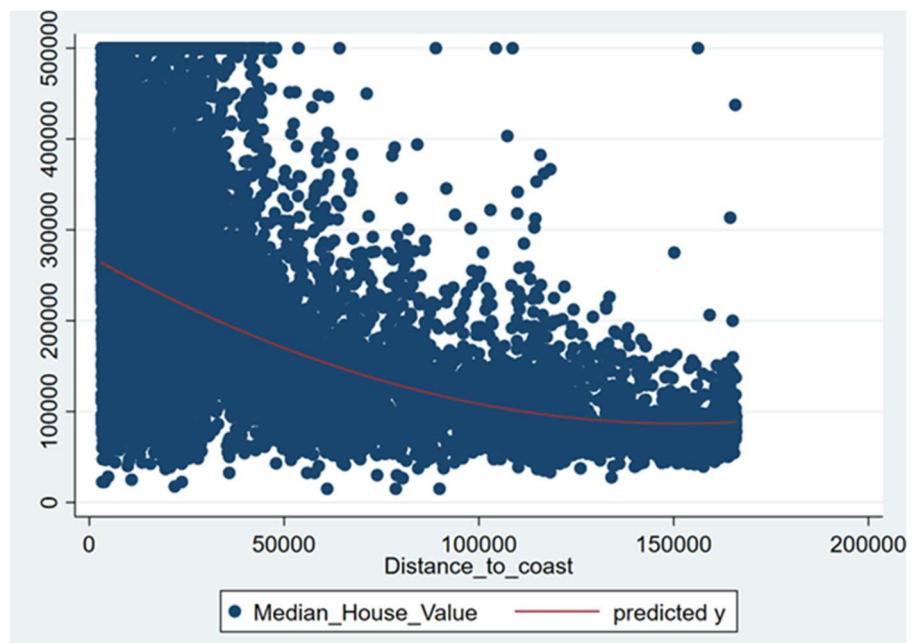
Gráfico de Dispersão



6. Distance to Coast

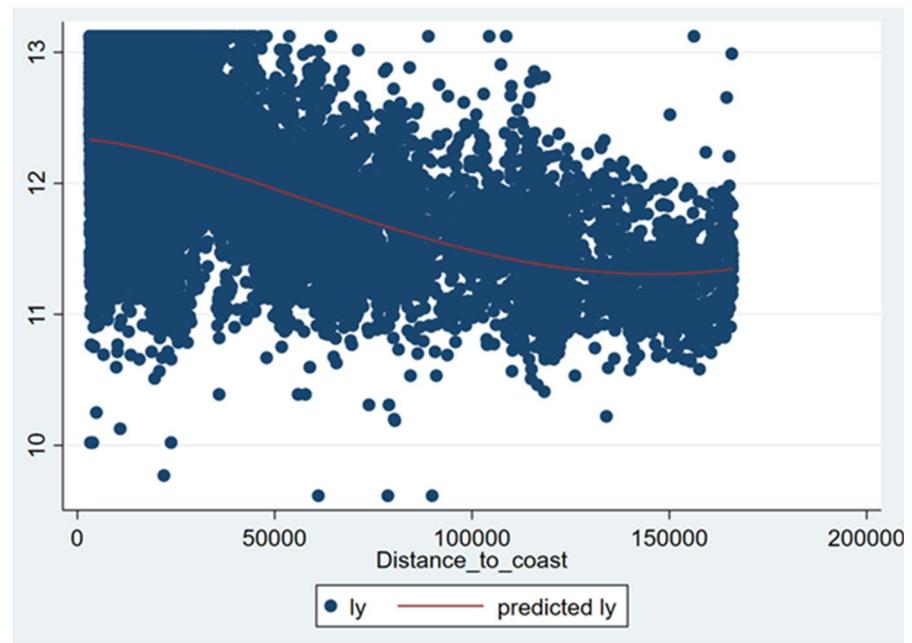
Lin – Lin

Gráfico de Dispersão



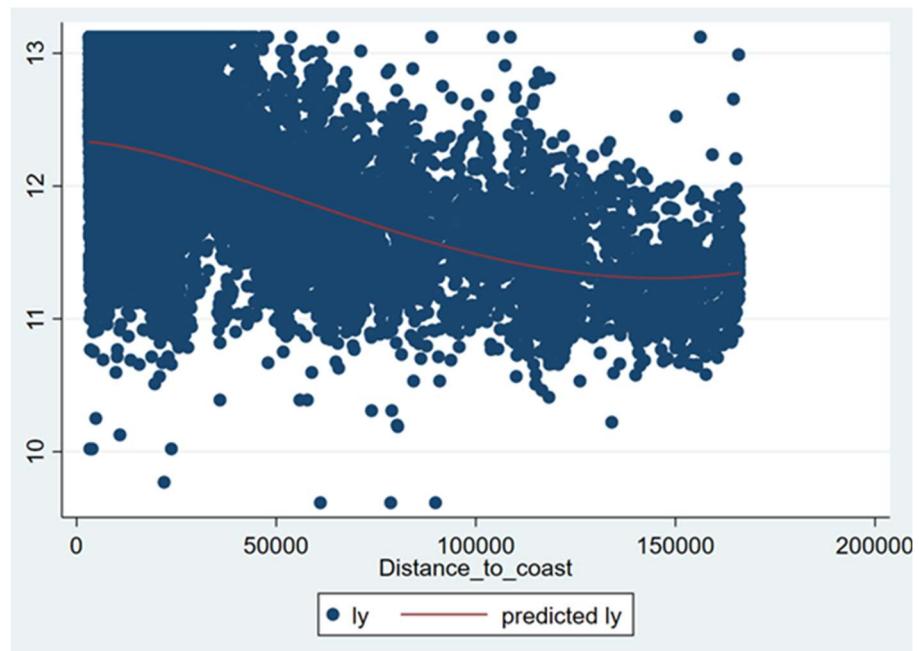
Log – Lin

Gráfico de Dispersão



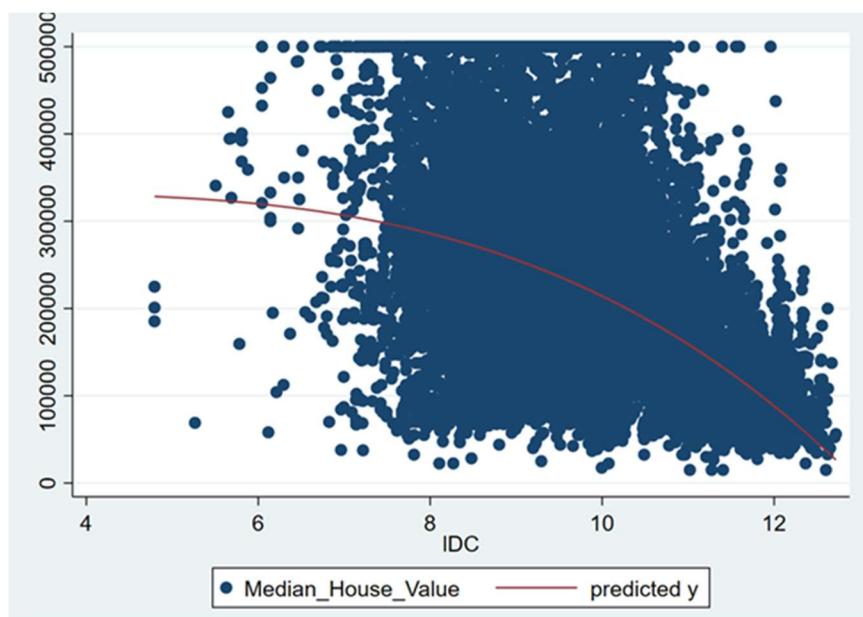
Log – Log

Gráfico de Dispersão



Lin – Log

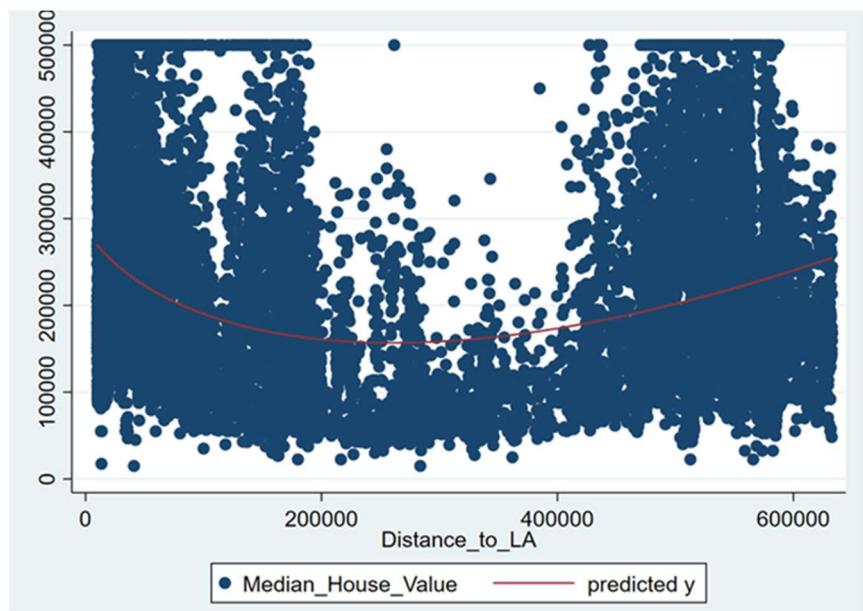
Gráfico de Dispersão



7. Distance to LA

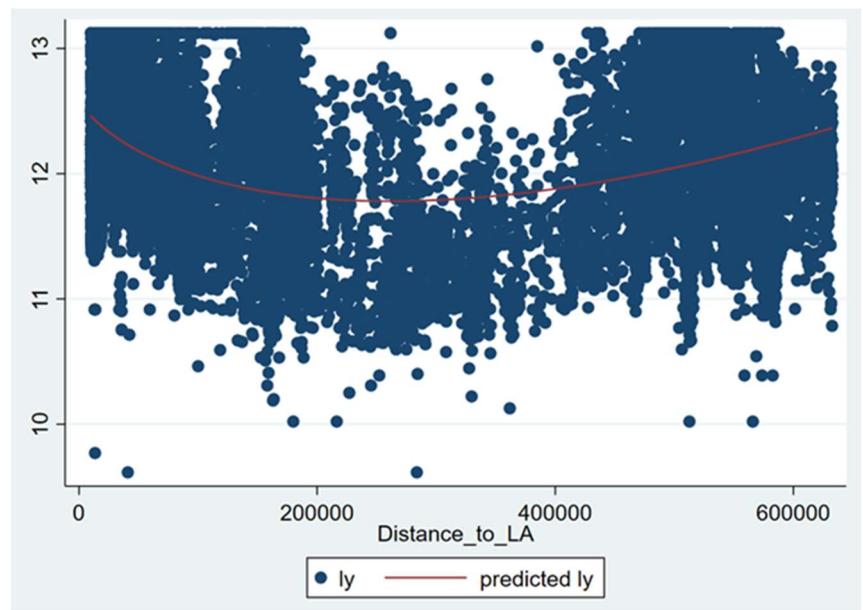
Lin – Lin

Gráfico de Dispersão



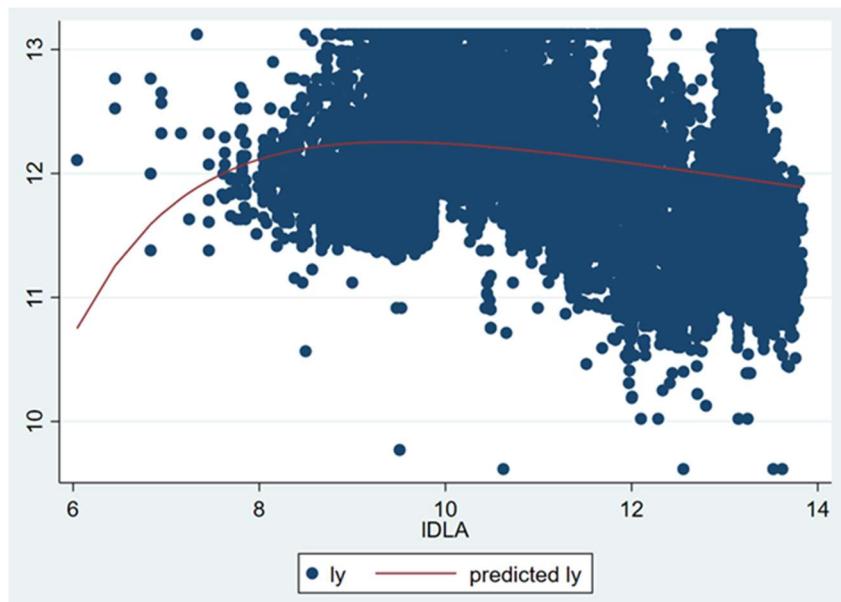
Log – Lin

Gráfico de Dispersão



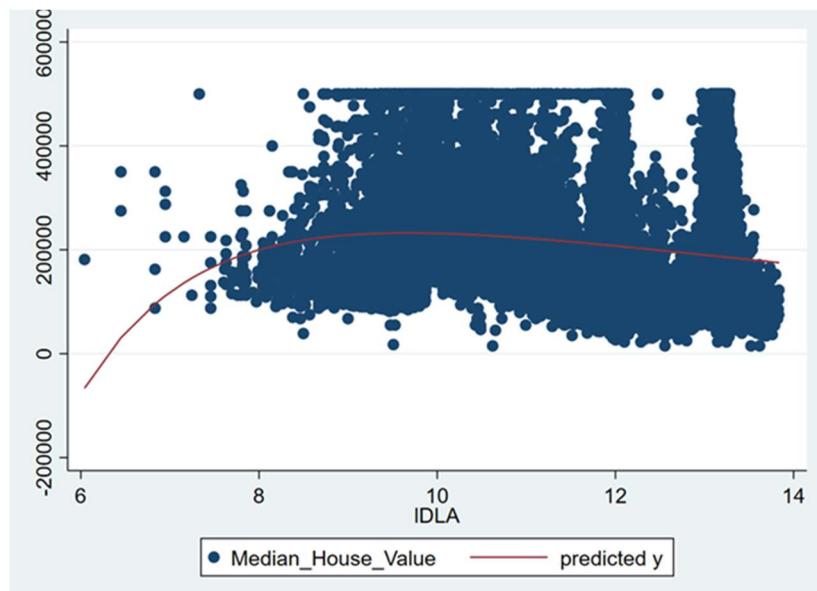
Log – Log

Gráfico de Dispersão



Lin – Log

Gráfico de Dispersão



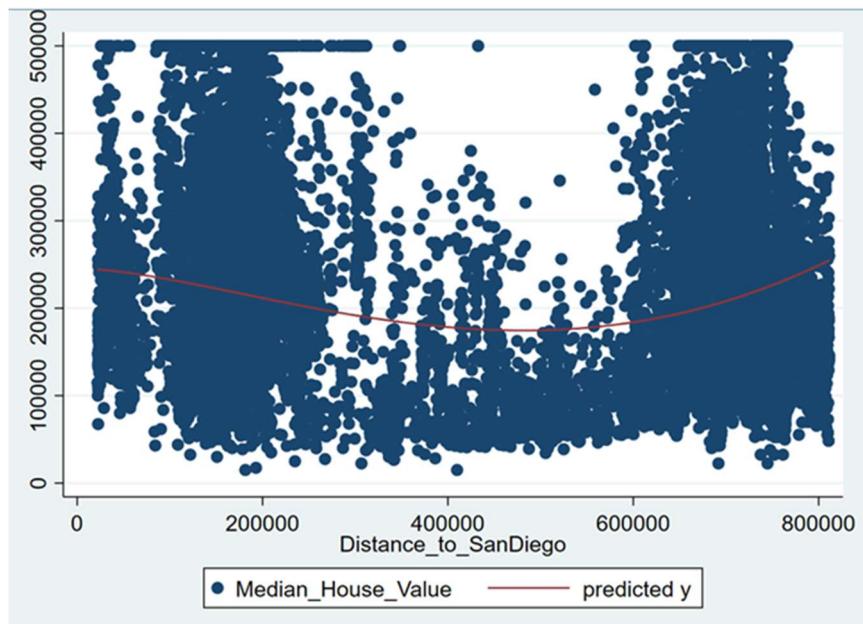
Melhor Forma:

Mais próximo de linearidade é o Log – Lin.

8. Distance to San Diego

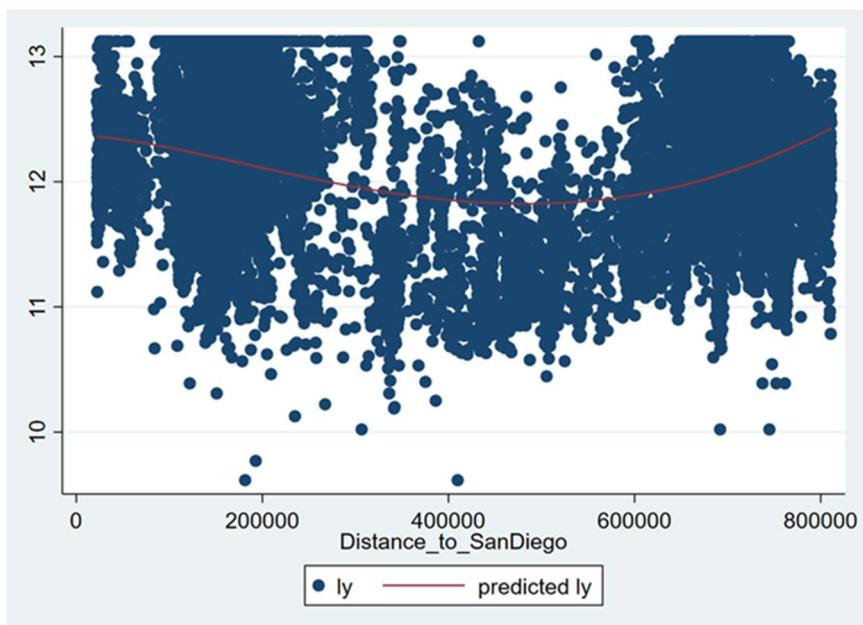
Lin – Lin

Gráfico de Dispersão



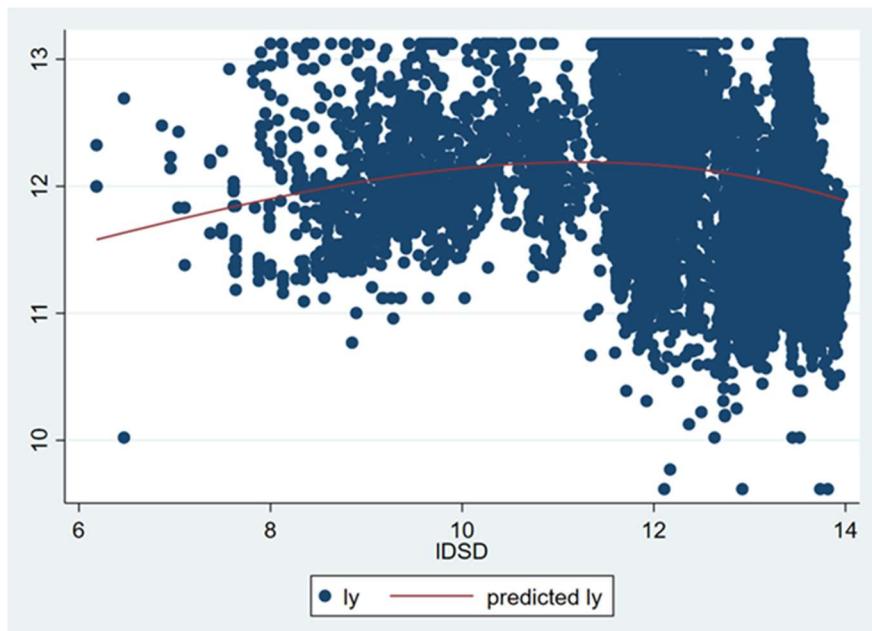
Log – Lin

Gráfico de Dispersão



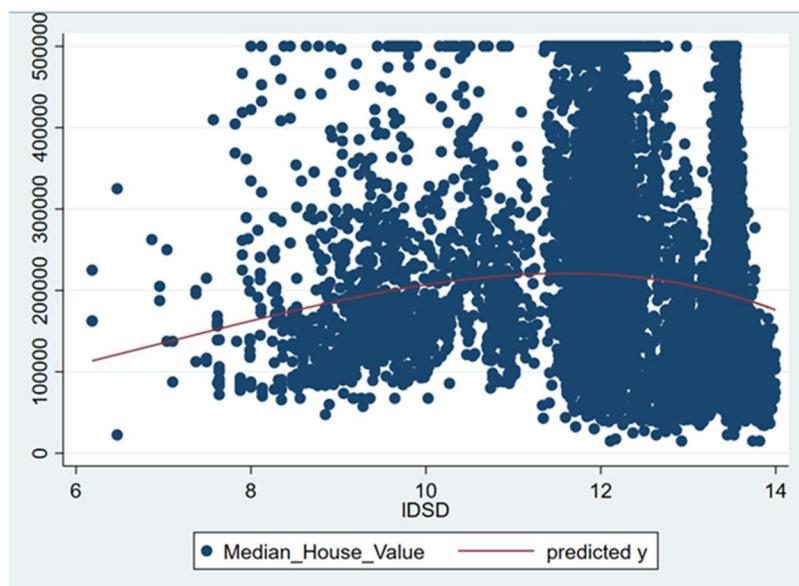
Log – Log

Gráfico de Dispersão



Lin – Log

Gráfico de Dispersão



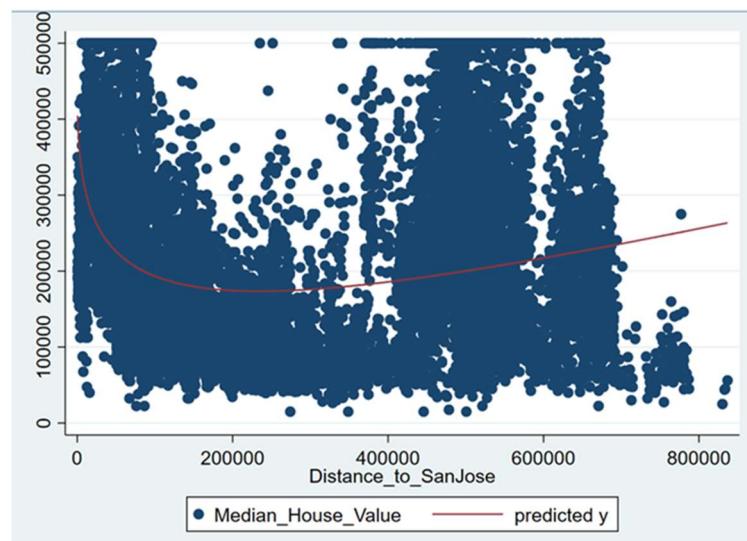
Melhor Forma:

Mais próximo de linearidade é Log – Log.

9. Distance to San José

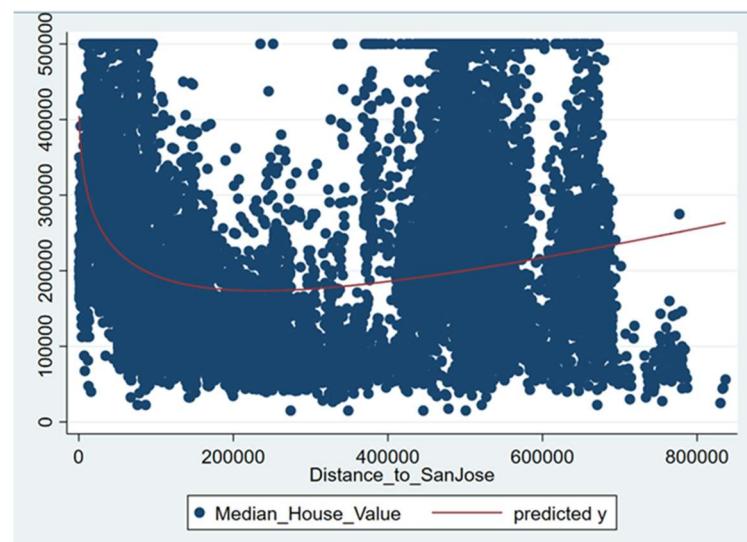
Lin – Lin

Gráfico de Dispersão



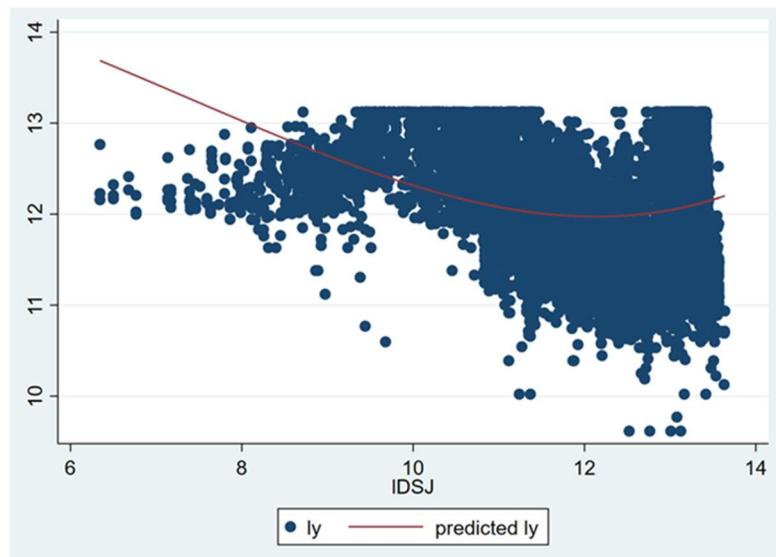
Log – Lin

Gráfico de Dispersão



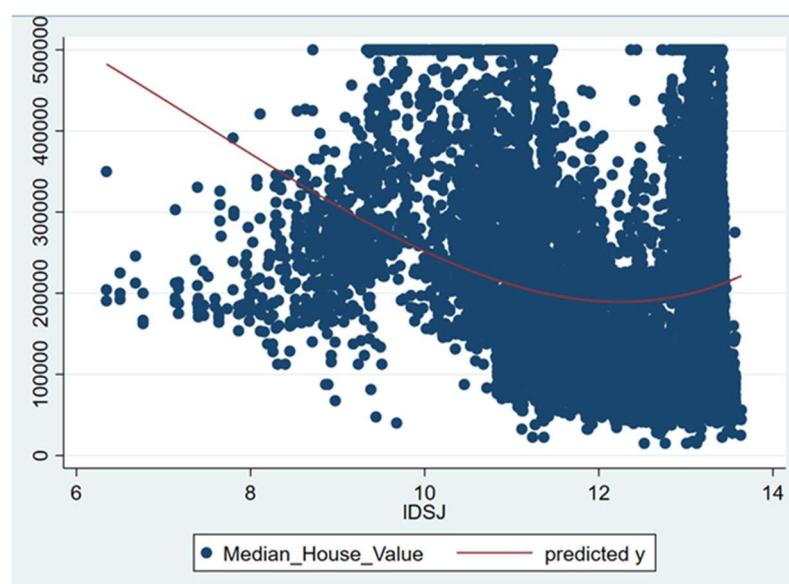
Log – Log

Gráfico de Dispersão



Lin – Log

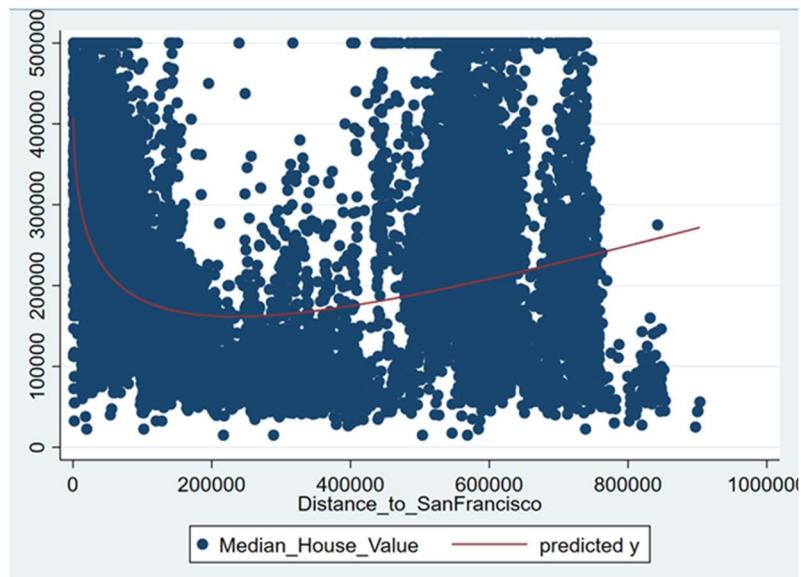
Gráfico de Dispersão



10. Distance to San Francisco

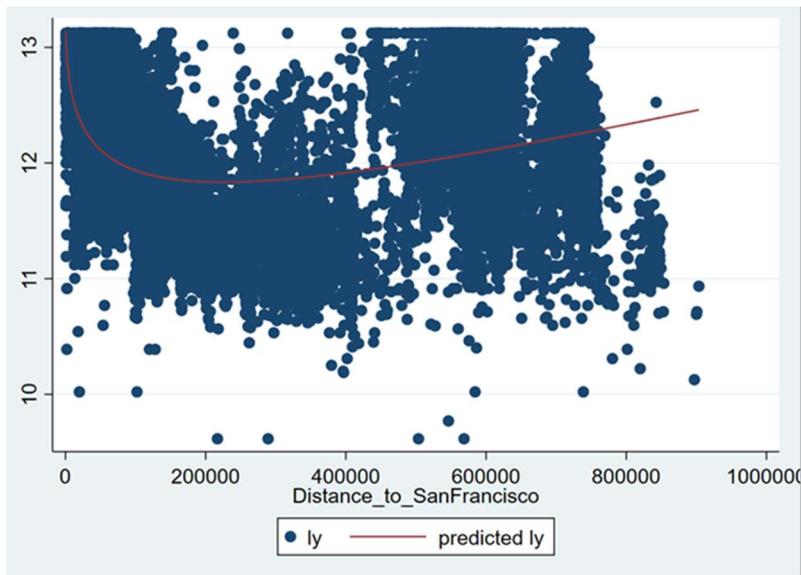
Lin – Lin

Gráfico de Dispersão



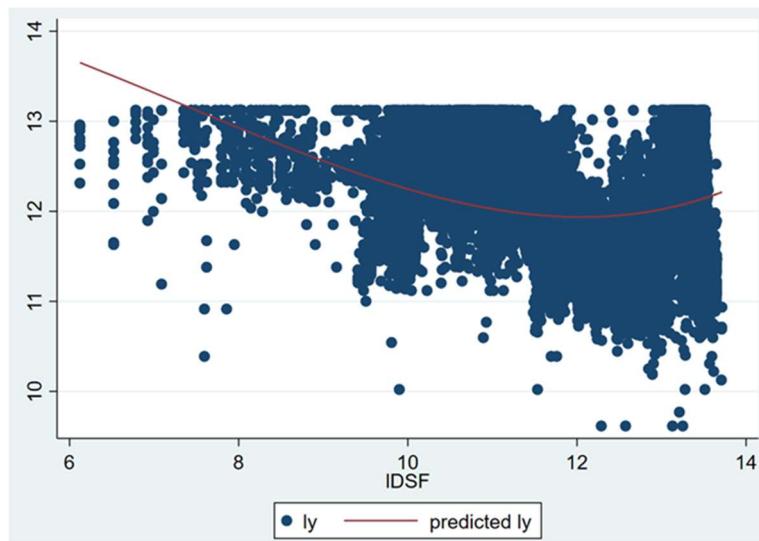
Log – Lin

Gráfico de Dispersão



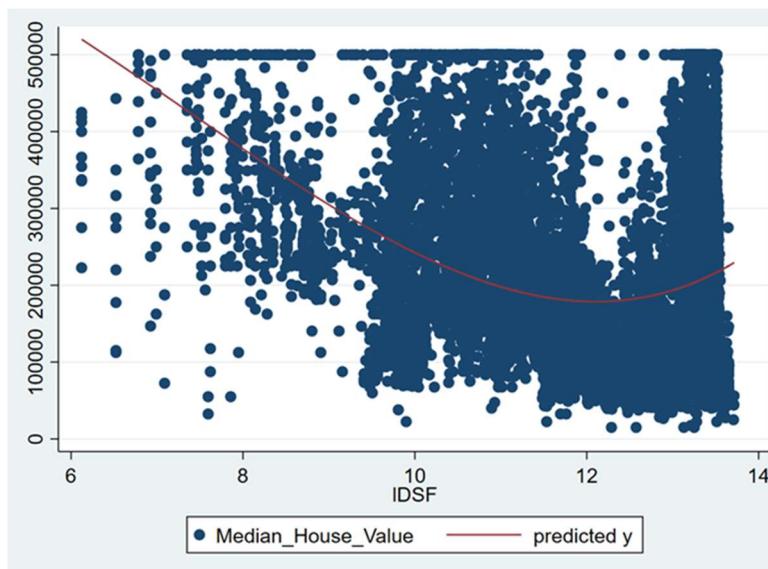
Log – Log

Gráfico de Dispersão



Lin – Log

Gráfico de Dispersão



PASSO A PASSO RETIRANDO CADA VARIÁVEL NÃO RELEVANTE DO MODELO:

Primeiro, rodamos a regressão com todas a variáveis relevantes:

Median_House_Value		Robust Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Median_Income	7015.303	359.8365	19.50	0.000	6309.995	7720.611
Median_Age	1082.233	97.61787	11.09	0.000	890.8946	1273.572
Tot_Rooms	2.80013	.5229135	5.35	0.000	1.775178	3.825082
Tot_Bedrooms	6.009917	4.490801	1.34	0.181	-2.792408	14.81224
Population	9.250438	4.764998	1.94	0.052	-.0893349	18.59021
Households	-29.23188	13.87746	-2.11	0.035	-56.43279	-2.030965
lDistance_to_coast	175527.9	1047.712	167.53	0.000	173474.3	177581.5
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-443.5267	96.68468	-4.59	0.000	-633.0363	-254.017
maUP	-882.4519	96.26021	-9.17	0.000	-1071.13	-693.7743
maLMC	-232.3503	107.2743	-2.17	0.030	-442.6165	-22.08413
maUMC	-409.8461	169.5894	-2.42	0.016	-742.2547	-77.43756
hhMC	20.15373	13.45981	1.50	0.134	-6.228561	46.53603
hhUP	54.20719	13.63981	3.97	0.000	27.47209	80.94229
hhLMC	95.45534	17.08036	5.59	0.000	61.97649	128.9342
hhUMC	-28.13176	53.93826	-0.52	0.602	-133.855	77.5915
ppMC	-10.35671	4.827063	-2.15	0.032	-19.81814	-.8952889
ppUP	-27.95064	4.920185	-5.68	0.000	-37.59459	-18.30669
ppLMC	-41.5779	6.065009	-6.86	0.000	-53.4658	-29.69001
ppUMC	21.03388	20.47778	1.03	0.304	-19.10418	61.17194
_cons	-1953138	11759.16	-166.09	0.000	-1976187	-1930089

Retiramos a interação entre Household e Upper Middle Class e rodamos novamente:

Median_House_Value		Robust Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Median_Income	7018.442	359.5422	19.52	0.000	6313.711	7723.174
Median_Age	1082.739	97.57668	11.10	0.000	891.4812	1273.997
Tot_Rooms	2.79674	.5226585	5.35	0.000	1.772288	3.821192
Tot_Bedrooms	6.028609	4.489782	1.34	0.179	-2.771718	14.82894
Population	9.382365	4.71739	1.99	0.047	.1359073	18.62882
Households	-29.6779	13.69667	-2.17	0.030	-56.52446	-2.831343
lDistance_to_coast	175526.6	1047.644	167.54	0.000	173473.1	177580
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-443.9779	96.651	-4.59	0.000	-633.4215	-254.5343
maUP	-883.0072	96.21389	-9.18	0.000	-1071.594	-694.4204
maLMC	-233.1254	107.2034	-2.17	0.030	-443.2526	-22.99823
maUMC	-408.7671	168.6049	-2.42	0.015	-739.2461	-78.28822
hhMC	20.59314	13.27878	1.55	0.121	-5.434318	46.62059
hhUP	54.64658	13.46097	4.06	0.000	28.26201	81.03115
hhLMC	95.89677	16.93888	5.66	0.000	62.69524	129.0983
ppMC	-10.48818	4.780225	-2.19	0.028	-19.8578	-1.118561
ppUP	-28.08162	4.87453	-5.76	0.000	-37.63608	-18.52716
ppLMC	-41.70917	6.028555	-6.92	0.000	-53.52562	-29.89273
ppUMC	11.14422	4.533446	2.46	0.014	2.258304	20.03013
_cons	-1953132	11758.74	-166.10	0.000	-1976180	-1930084

Retiramos Tot Bedrooms e rodamos novamente:

Median_House_Value	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Median_Income	6899.46	343.9946	20.06	0.000	6225.203	7573.716
Median_Age	1075.257	97.20358	11.06	0.000	884.73	1265.783
Tot_Rooms	3.266861	.4154483	7.86	0.000	2.45255	4.081172
Population	9.212716	4.724362	1.95	0.051	-.047408	18.47284
Households	-24.38671	13.18689	-1.85	0.064	-50.23405	1.460625
lDistance_to_coast	175666.7	1036.554	169.47	0.000	173635	177698.4
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-442.5461	96.67108	-4.58	0.000	-632.0291	-253.0632
maUP	-879.2302	96.10691	-9.15	0.000	-1067.607	-690.8531
maLMC	-224.3958	107.0862	-2.10	0.036	-434.2931	-14.49838
maUMC	-381.8265	167.0013	-2.29	0.022	-709.1623	-54.4908
hhMC	20.21545	13.32377	1.52	0.129	-5.900198	46.33109
hhUP	54.14774	13.51717	4.01	0.000	27.65302	80.64247
hhLMC	95.28595	17.10811	5.57	0.000	61.75271	128.8192
ppMC	-10.45956	4.788972	-2.18	0.029	-19.84632	-1.072796
ppUP	-28.18711	4.879193	-5.78	0.000	-37.75071	-18.6235
pplMC	-41.99047	6.030618	-6.96	0.000	-53.81096	-30.16999
ppUMC	10.52795	4.534878	2.32	0.020	1.639227	19.41667
_cons	-1954269	11688.82	-167.19	0.000	-1977180	-1931358

Retiramos a interação entre Household e Middle Class e rodamos novamente:

Median_House_Value	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Median_Income	6908.095	343.8985	20.09	0.000	6234.027	7582.164
Median_Age	1058.73	96.60741	10.96	0.000	869.3716	1248.088
Tot_Rooms	3.291407	.4158098	7.92	0.000	2.476387	4.106427
Population	3.455185	2.658097	1.30	0.194	-1.754895	8.665265
Households	-5.220128	3.075655	-1.70	0.090	-11.24865	.8083994
lDistance_to_coast	175645.1	1036.536	169.45	0.000	173613.4	177676.8
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-423.5978	95.89123	-4.42	0.000	-611.5522	-235.6434
maUP	-862.1676	95.37289	-9.04	0.000	-1049.106	-675.2293
maLMC	-207.7067	106.5149	-1.95	0.051	-416.4843	1.070899
maUMC	-364.81	166.2877	-2.19	0.028	-690.7471	-38.87291
hhUP	34.89908	4.163423	8.38	0.000	26.73844	43.05972
hhLMC	76.01744	11.35318	6.70	0.000	53.76431	98.27056
ppMC	-4.438346	2.567075	-1.73	0.084	-9.470016	.5933235
ppUP	-22.44347	2.946254	-7.62	0.000	-28.21836	-16.66858
ppLMC	-36.253	4.634429	-7.82	0.000	-45.33685	-27.16915
ppUMC	9.451446	4.474201	2.11	0.035	.6816588	18.22123
_cons	-1954059	11689.18	-167.17	0.000	-1976971	-1931147

Retiramos Population e rodamos novamente:

Median_House_Value	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Median_Income	6899.996	344.0906	20.05	0.000	6225.551	7574.441
Median_Age	1133.05	70.23943	16.13	0.000	995.375	1270.725
Tot_Rooms	3.275627	.4148007	7.90	0.000	2.462584	4.088669
Households	-3.729086	2.947663	-1.27	0.206	-9.506737	2.048566
lDistance_to_coast	175646.6	1036.503	169.46	0.000	173615	177678.2
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-501.6548	67.46037	-7.44	0.000	-633.8825	-369.4272
maUP	-937.0653	68.33053	-13.71	0.000	-1070.999	-803.132
maLMC	-282.0305	83.73986	-3.37	0.001	-446.1672	-117.8937
maUMC	-437.7972	152.8339	-2.86	0.004	-737.3638	-138.2306
hhUP	33.46747	4.101817	8.16	0.000	25.42758	41.50735
hhLMC	74.59913	11.32728	6.59	0.000	52.39676	96.80149
ppMC	-1.380352	.7532411	-1.83	0.067	-2.856764	.0960606
ppUP	-18.98441	1.285444	-14.77	0.000	-21.50398	-16.46484
pplMC	-32.78742	3.785162	-8.66	0.000	-48.20664	-25.36821
ppUMC	12.44169	3.741055	3.33	0.001	5.108925	19.77445
_cons	-1954022	11686.03	-167.21	0.000	-1976928	-1931117

Retiramos Households e rodamos novamente:

Median_House_Value	Robust					[95% Conf. Interval]
	Coef.	Std. Err.	t	P> t		
Median_Income	6994.381	329.3958	21.23	0.000	6348.739	7640.023
Median_Age	1117.294	69.1193	16.16	0.000	981.8146	1252.773
Tot_Rooms	3.02574	.3328502	9.09	0.000	2.373327	3.678152
lDistance_to_coast	175487	1022.675	171.60	0.000	173482.5	177491.6
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-485.9124	65.59306	-7.41	0.000	-614.48	-357.3448
maUP	-918.4551	66.8165	-13.75	0.000	-1049.421	-787.4895
maLMC	-267.5021	82.81603	-3.23	0.001	-429.828	-105.1761
maUMC	-435.9499	152.5999	-2.86	0.004	-735.0578	-136.842
hhUP	30.95225	3.649275	8.48	0.000	23.79938	38.10512
hhLMC	72.21473	11.02509	6.55	0.000	50.60468	93.82478
ppMC	-2.08587	.5011306	-4.16	0.000	-3.068125	-1.103614
ppUP	-18.92039	1.280028	-14.78	0.000	-21.42934	-16.41143
ppLMC	-32.68236	3.767284	-8.68	0.000	-40.06654	-25.29819
ppUMC	11.71391	3.685425	3.18	0.001	4.490188	18.93764
_cons	-1952587	11584	-168.56	0.000	-1975293	-1929882

ESSE FOI CONSIDERADO NOSSO MODELO FINAL

Também é possível observar pelo teste F de significância global da regressão que o modelo é relevante, pois o p-Valor do teste F dele é aproximadamente 0, portanto, rejeitando a hipótese nula de que todos os betas devem ser iguais a zero.

```

Number of obs      = 20,640
F(14, 20625)      = 9226.02
Prob > F          = 0.0000
R-squared          = 0.9235
Root MSE           = 31918

```

Gráficos de Dispersão dos Resíduos:

Gráfico de Dispersão Resíduos 1 - Interação Median Age e Renda Classes1

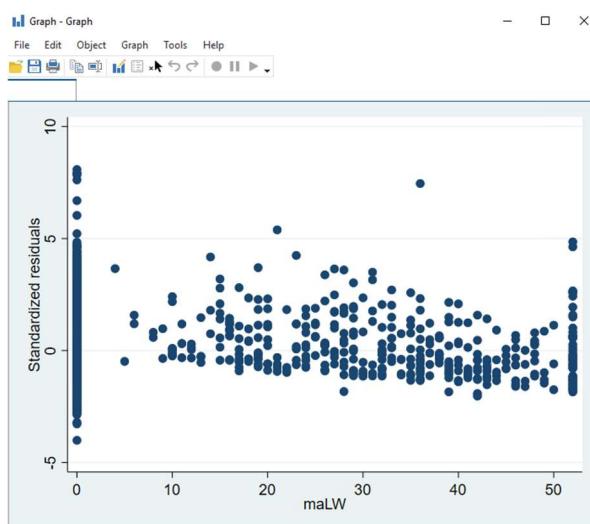


Gráfico de Dispersão Resíduos 2 - Interação Median Age e Renda Classes2

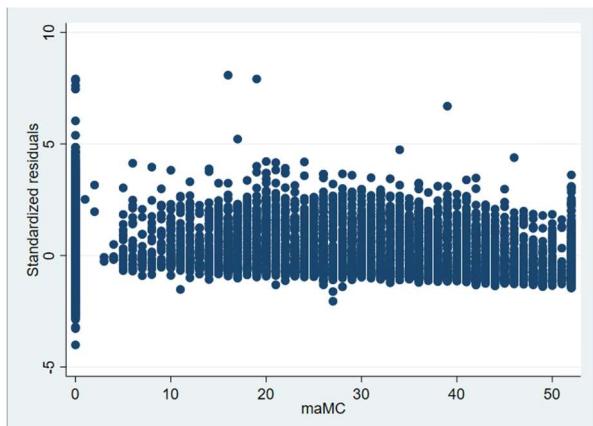


Gráfico de Dispersão Resíduos 3 - Interação Median Age e Renda Classes3

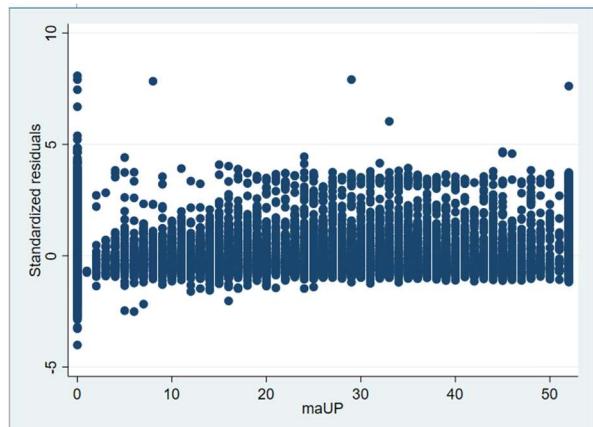


Gráfico de Dispersão Resíduos 4 - Interação Median Age e Renda Classes4

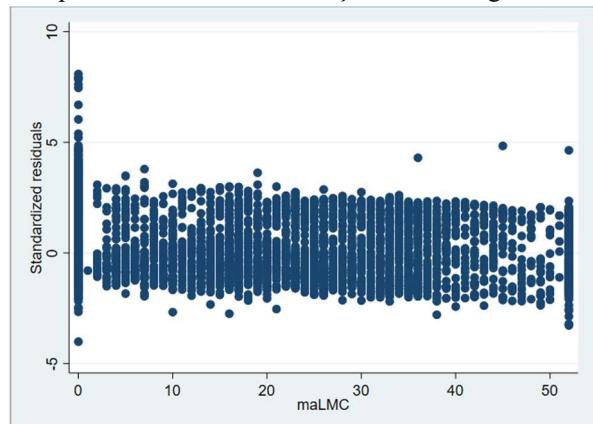


Gráfico de Dispersão Resíduos 5 - Interação Median Age e Renda Classes5

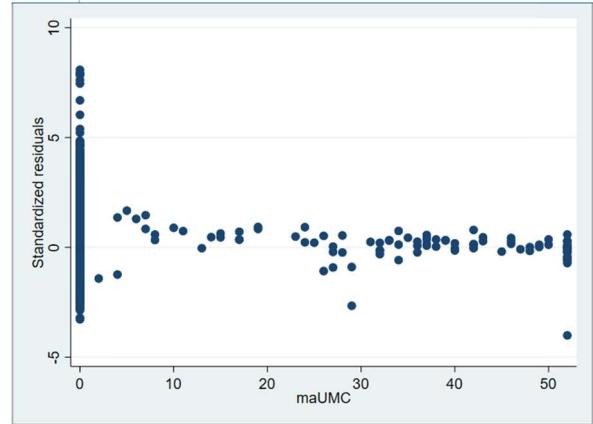


Gráfico de Dispersão Resíduos 6 - Interação Household e Renda Classes1

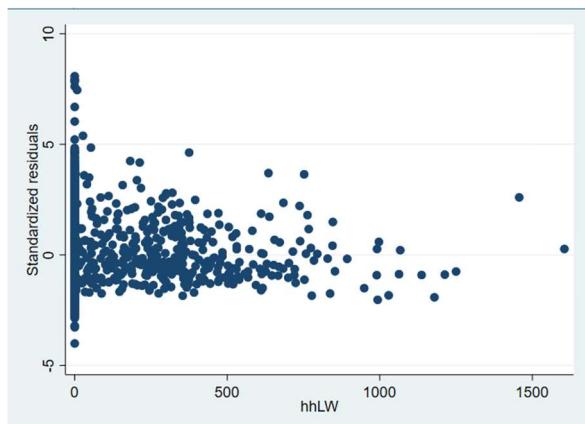


Gráfico de Dispersão Resíduos 7 - Interação Household e Renda Classes2

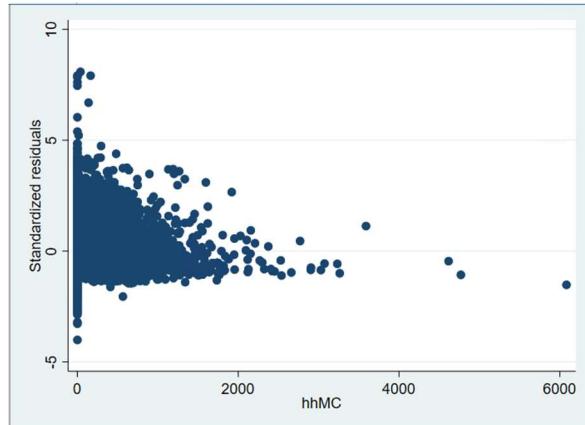


Gráfico de Dispersão Resíduos 8 - Interação Household e Renda Classes3

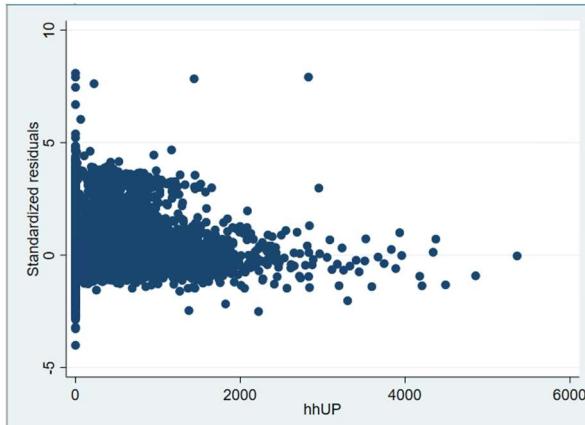


Gráfico de Dispersão Resíduos 9 - Interação Household e Renda Classes4

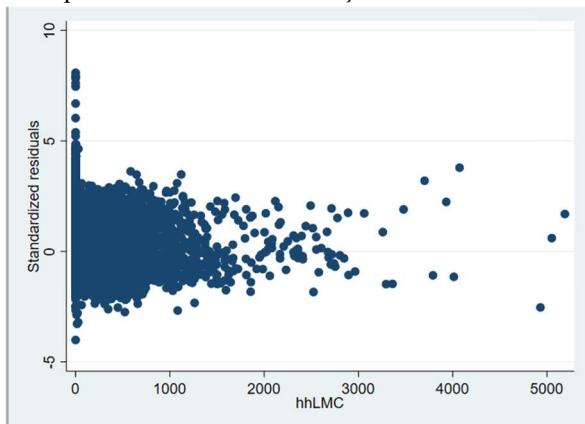


Gráfico de Dispersão Resíduos 10 - Interação Household e Renda Classes5

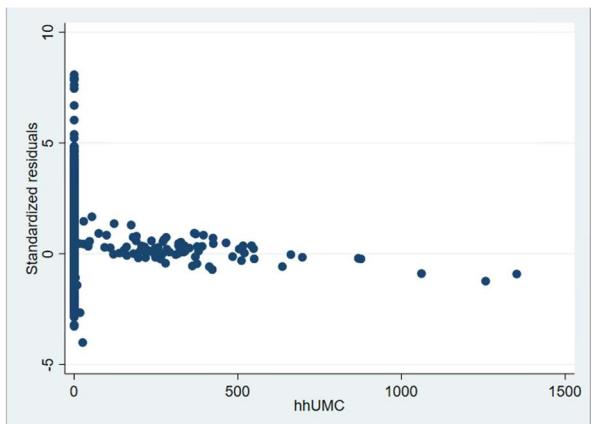


Gráfico de Dispersão Resíduos 11 - Interação Population e Renda Classes1

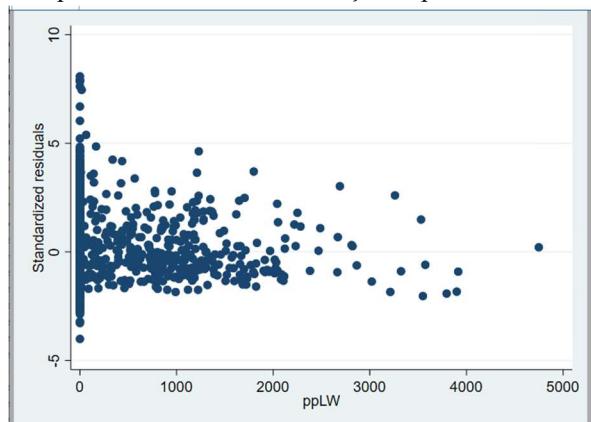


Gráfico de Dispersão Resíduos 12 - Interação Population e Renda Classes2

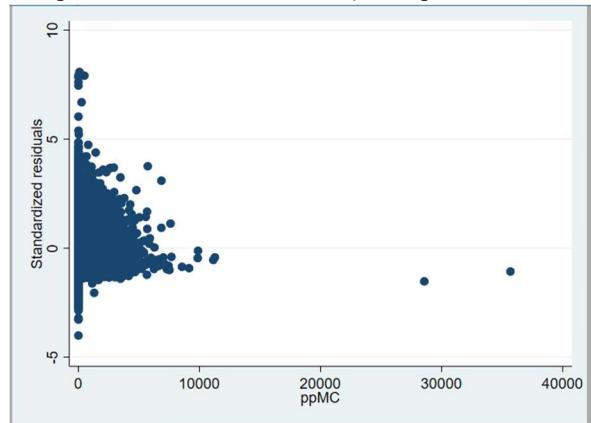


Gráfico de Dispersão Resíduos 13 - Interação Population e Renda Classes3

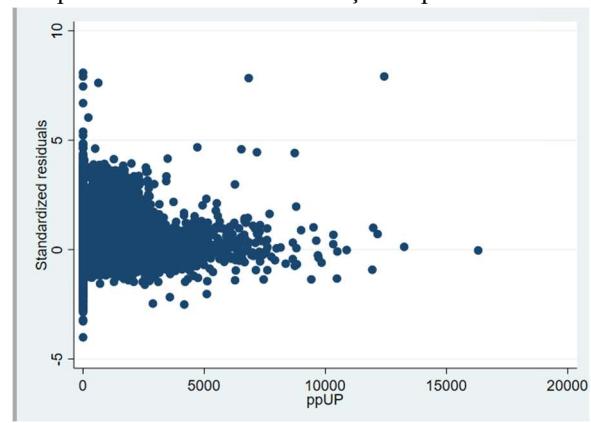


Gráfico de Dispersão Resíduos 14 - Interação Population e Renda Classes4

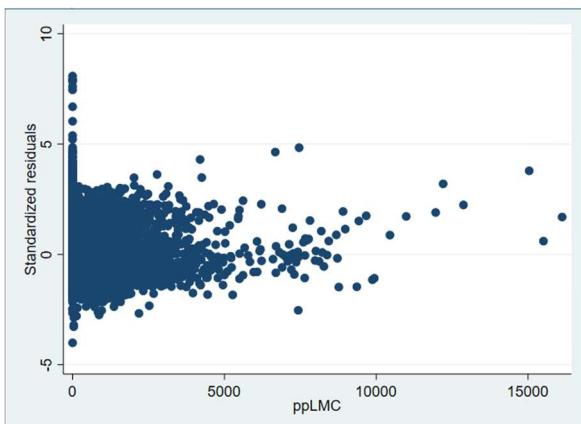


Gráfico de Dispersão Resíduos 15 - Interação Population e Renda Classes5

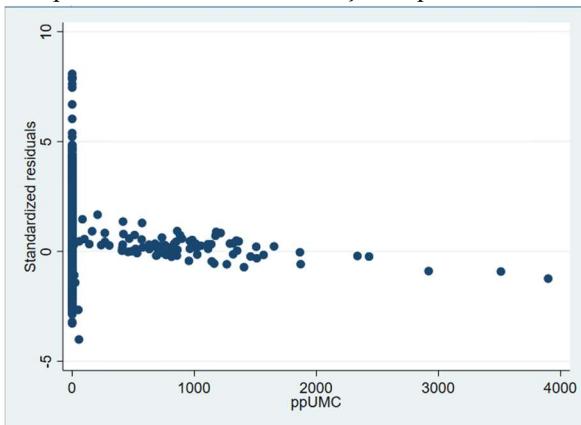


Gráfico de Dispersão Resíduos 16 - Indiv

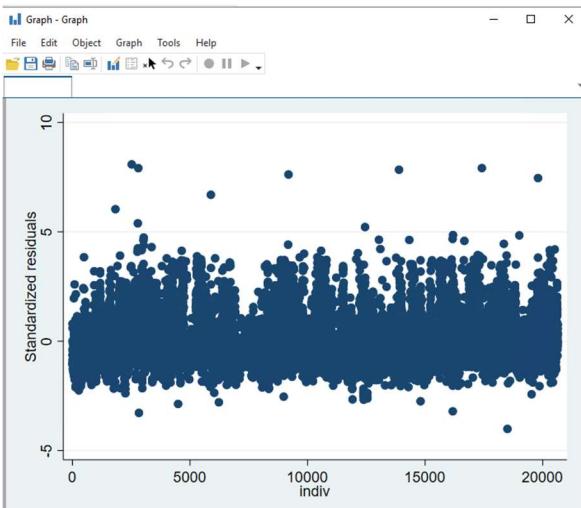


Gráfico de Dispersão Resíduos 17 - Median Income

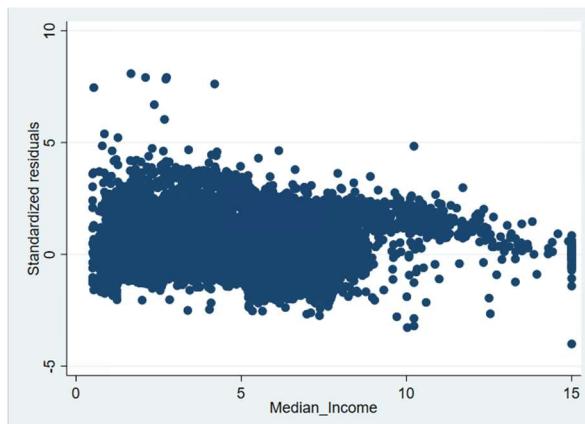


Gráfico de Dispersão Resíduos 18 - Median Income

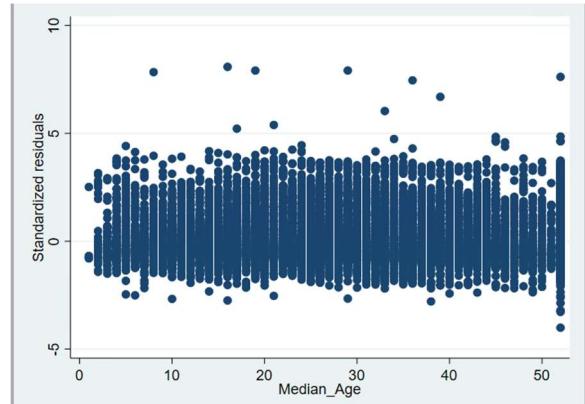


Gráfico de Dispersão Resíduos 19 - Tot Rooms

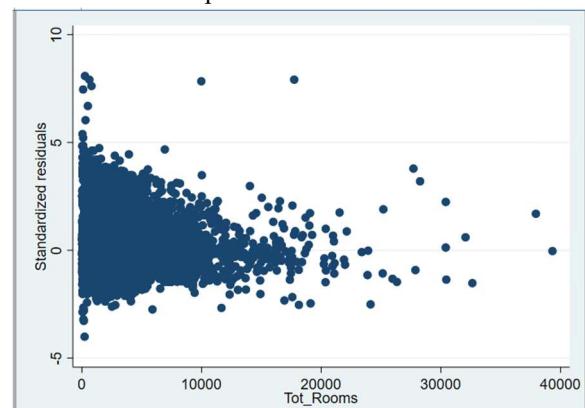


Gráfico de Dispersão Resíduos 20 - Tot Bedrooms

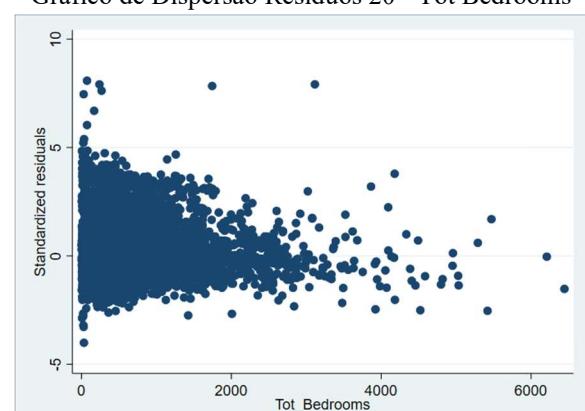


Gráfico de Dispersão Resíduos 21 - Population

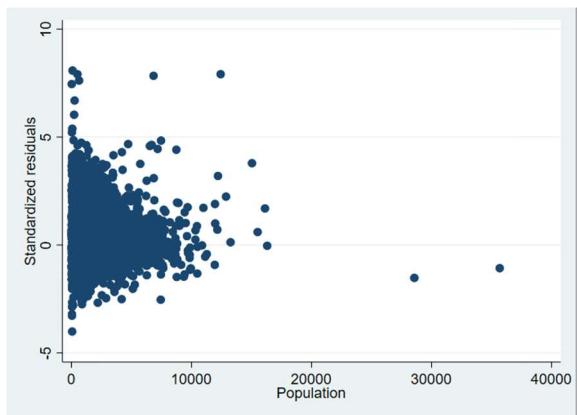


Gráfico de Dispersão Resíduos 22 - Households

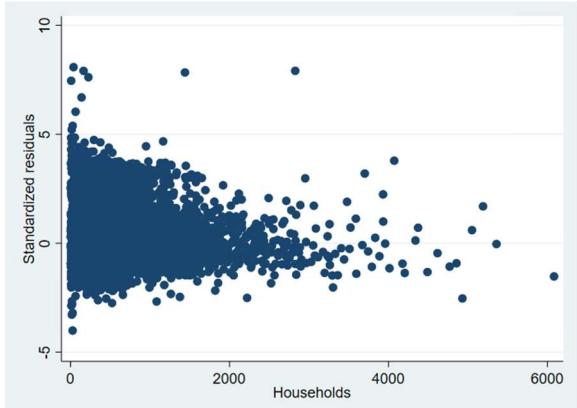


Gráfico de Dispersão Resíduos 23 - Distance to Coast

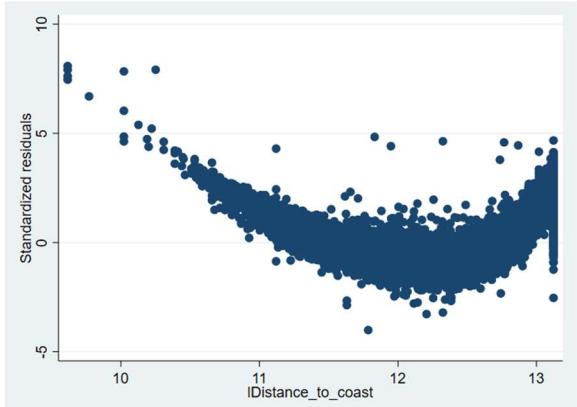


Gráfico de Dispersão Resíduos 24 - Distance to LA

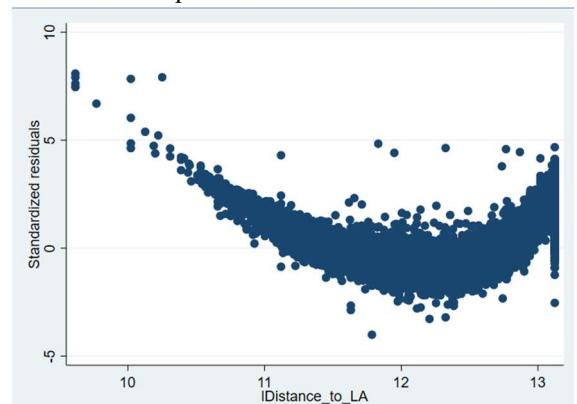


Gráfico de Dispersão Resíduos 25 - Distance to San Diego

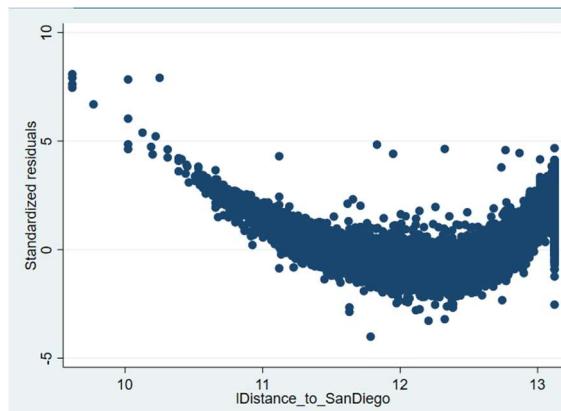


Gráfico de Dispersão Resíduos 26 - Distance to San Jose

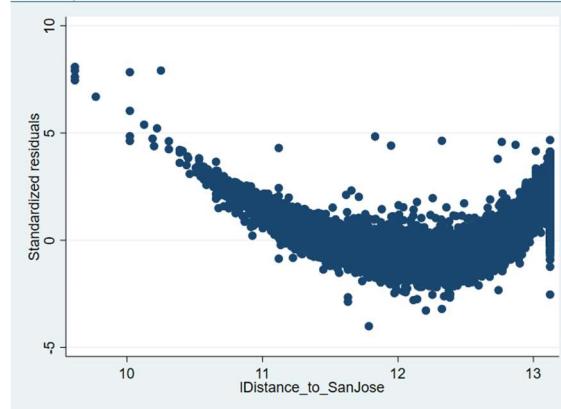
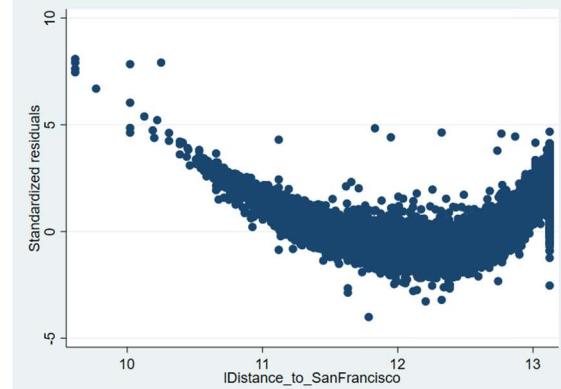


Gráfico de Dispersão Resíduos 27 - Distance to San Francisco



TESTE JARQUE BERA:

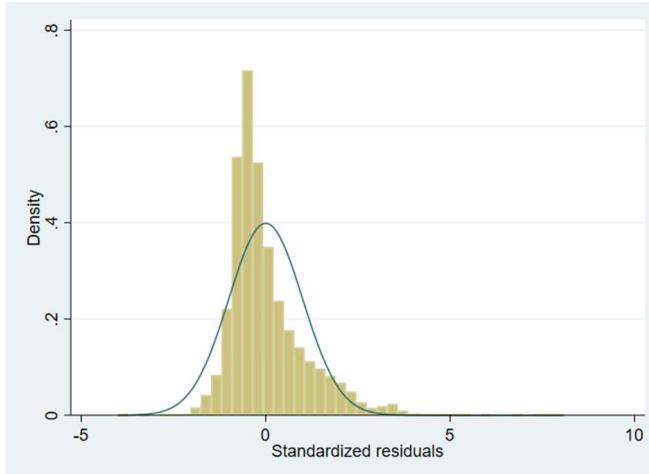
```
. jb resid_pad1
Jarque-Bera normality test: 1.4e+04 Chi(2)      0
Jarque-Bera test for Ho: normality:
```

```
. summ resid_pad1, detail
                                         Standardized residuals

```

	Percentiles	Smallest		
1%	-1.624499	-4.007752		
5%	-1.139986	-3.274153		
10%	-.9234754	-3.202056	Obs	20,640
25%	-.6393256	-2.866019	Sum of wgt.	20,640
50%	-.2771424		Mean	.0000282
75%	.4116405	7.83924	Largest	
90%	1.440462	7.912257	Std. dev.	1.000402
95%	2.034078	7.913677	Variance	1.000804
99%	3.349136	8.081939	Skewness	1.401979
			Kurtosis	5.961616

Histograma 14 - Resíduo Padronizado



WHITE TEST:

```
. whitetst
```

```
White's general test statistic : 13167.82 Chi-sq(119) P-value =      0
```

ERROS PADRÕES ROBUSTOS:

```
. reg Median_House_Value Median_Income Median_Age Tot_Rooms Tot_Bedrooms Population Households lDistance_to_coast lDi
> stance_to_LA lDistance_to_SanDiego lDistance_to_SanJose lDistance_to_SanFrancisco maMC maUP maLMC maUMC hhMC hhUP h
> hLMC hhUMC ppMC ppUP ppLMC ppUMC, rob
note: lDistance_to_LA omitted because of collinearity.
note: lDistance_to_SanDiego omitted because of collinearity.
note: lDistance_to_SanJose omitted because of collinearity.
note: lDistance_to_SanFrancisco omitted because of collinearity.
```

Linear regression	Number of obs	=	20,640
	F(19, 20620)	=	6819.50
	Prob > F	=	0.0000
	R-squared	=	0.9236
	Root MSE	=	31913

Median_House_Value	Robust					
	Coefficient	std. err.	t	P> t	[95% conf. interval]	
Median_Income	7015.303	359.8365	19.50	0.000	6309.995	7720.611
Median_Age	1082.233	97.61787	11.09	0.000	890.8946	1273.572
Tot_Rooms	2.80013	.5229135	5.35	0.000	1.775178	3.825082
Tot_Bedrooms	6.009917	4.490801	1.34	0.181	-2.792408	14.81224
Population	9.250438	4.764998	1.94	0.052	-.0893349	18.59021
Households	-29.23188	13.87746	-2.11	0.035	-56.43279	-2.030965
lDistance_to_coast	175527.9	1047.712	167.53	0.000	173474.3	177581.5
lDistance_to_LA	0	(omitted)				
lDistance_to_SanDiego	0	(omitted)				
lDistance_to_SanJose	0	(omitted)				
lDistance_to_SanFrancisco	0	(omitted)				
maMC	-443.5267	96.68468	-4.59	0.000	-633.0363	-254.017
maUP	-882.4519	96.26021	-9.17	0.000	-1071.13	-693.7743
maLMC	-232.3503	107.2743	-2.17	0.030	-442.6165	-22.08413
maUMC	-409.8461	169.5894	-2.42	0.016	-742.2547	-77.43756
hhMC	20.15373	13.45981	1.50	0.134	-6.228561	46.53603
hhUP	54.20719	13.63981	3.97	0.000	27.47209	80.94229
hhLMC	95.45534	17.08036	5.59	0.000	61.97649	128.9342
hhUMC	-28.13176	53.93826	-0.52	0.602	-133.855	77.5915
ppMC	-10.35671	4.827063	-2.15	0.032	-19.81814	-.8952889
ppUP	-27.95064	4.920185	-5.68	0.000	-37.59459	-18.30669
ppLMC	-41.5779	6.065009	-6.86	0.000	-53.4658	-29.69001
ppUMC	21.03388	20.47778	1.03	0.304	-19.10418	61.17194
_cons	-1953138	11759.16	-166.09	0.000	-1976187	-1930089

Gráfico QQ PLOT 1

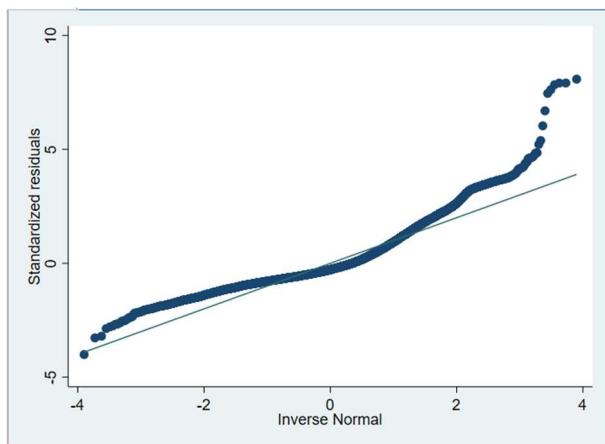


Gráfico de Interação 1 - Median Age

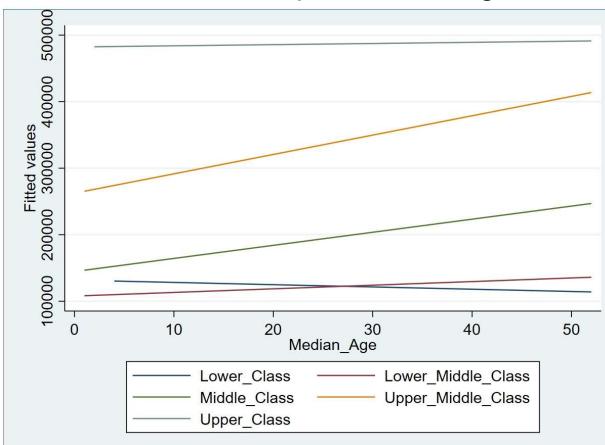


Gráfico de Interação 2 - Tot. Rooms

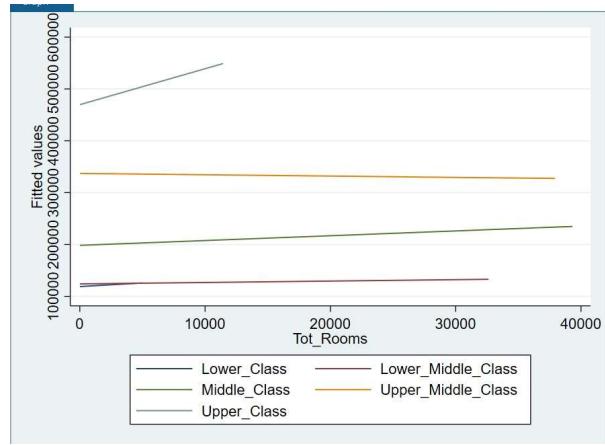


Gráfico de Interação 3 - Tot. Bedrooms

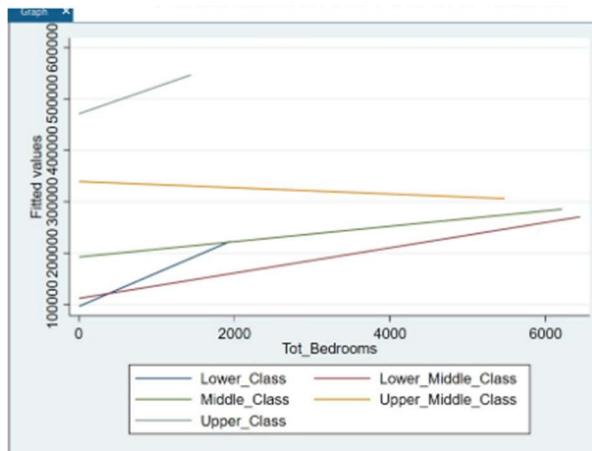


Gráfico de Interação 4 - Population

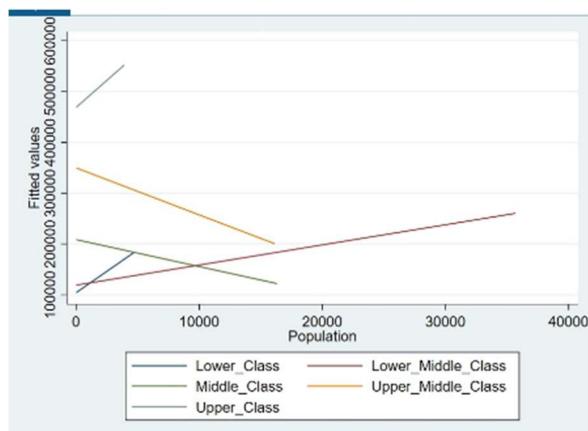


Gráfico de Interação 5 - Households

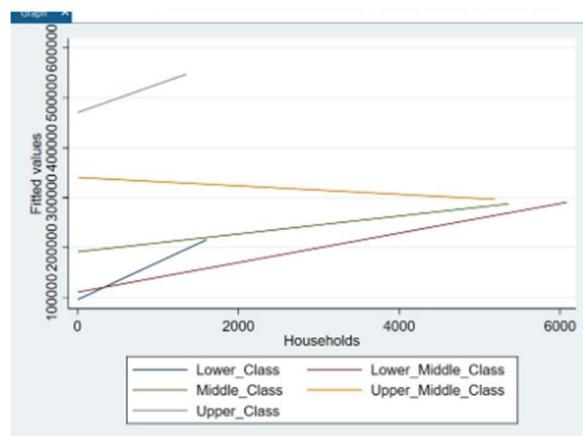


Gráfico de Interação 6 - Distance to Coast

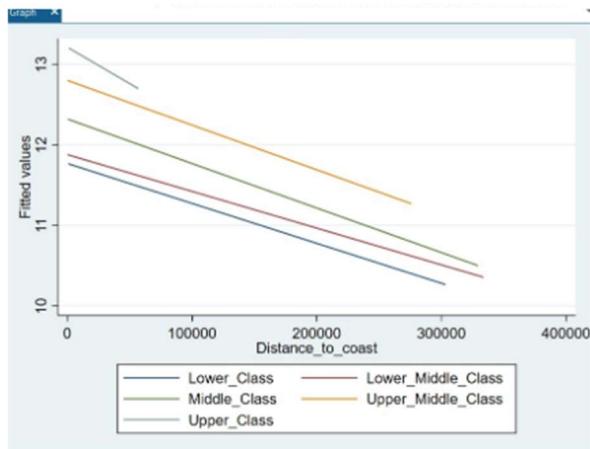


Gráfico de Interação 7 - Distance to LA

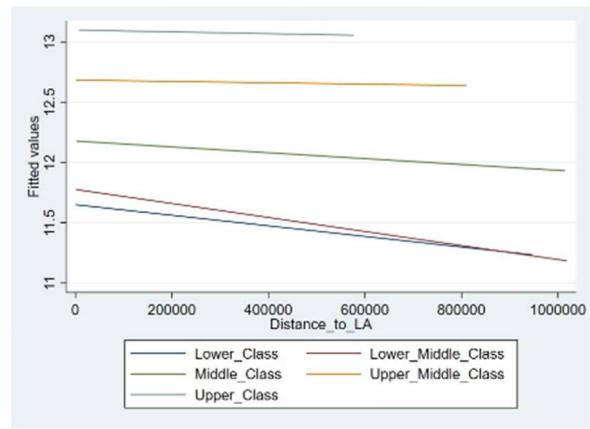


Gráfico de Interação 8 - Distance to San Diego

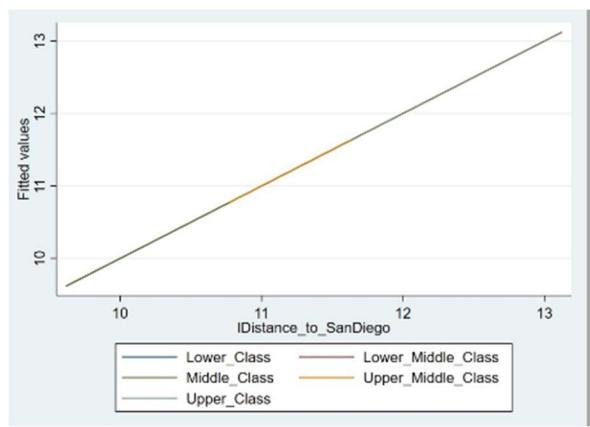


Gráfico de Interação 9 - Distance to San Jose

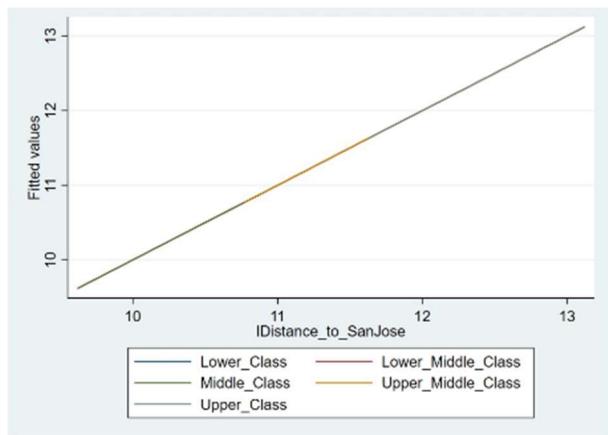


Gráfico de Interação 10 - Distance to San Francisco

