Summary Report
A Regression Model to predict the Market Value of Housing Units in Year 2013

Business question: What predicts 2013 market value of single family housing units?
Linear regression analysis was conducted to identify any relationship between variables within the Housing Affordability Data System dataset regarding factors that predict the current market value of single family housing units for 2013. The conclusion of the analysis was that there are many factors that predict the housing or current market value of single family units for 2013. Factors that have statistical significance and seem to contribute are:

* BEDRMS 2011: number of bedrooms in the housing unit
* LMED 2011: average median income for the area
* FMR 2011: cost or fair market monthly rental rate of the housing unit
* BUILT 2011: the year in which the housing unit was built
* ROOMS 2011: number of rooms in the housing unit
* REGION 2011: location in the United States of the housing unit; however, there is no statistical significance in market value of the housing unit and the housing unit being located in the midwest region of the US as compared to the southern region
* METRO3: location to metropolitan city of the housing unit
* ZINC2 2011: annual household income of the occupant of the housing unit
* AGE 2011: Age of head of household of the housing unit
* VALUE 2011: Market value of housing unit

Regression model listed below:

Ln(VALUE 2013) = β0 + β1AGE2011 + β2BEDRMS2011 + β3Ln(PER_2011) + B4REGION_WEST + B5REGION_MIDWEST + B6REGION_NORTHEAST + β7Ln(LMED_2011) + β8Ln(FMR_2011) + β9Ln(BUILT_2011) + β10Ln(VALUE_2011) + β11Ln(ZINC2_2011) + β12ROOMS_2011 + β13ZADEQ_ADEQUACY + β14ALL_OTHER_METROS + β15Ln(LN_UTILITY_2011) + β16Ln(OTHERCOST_2011)

Interpretation of regression model predictor variables and their coefficients in relation to outcome variable (VALUE 2013):

* Intercept: has no practical managerial value
* β1: a year increase in age for the head of household corresponds to 0.086% increase in market value of housing unit, all other variables held at their current level.
* β2: an increase of one bedroom corresponds to 3.588% decrease in market value of the housing unit, all other variables held at their current level.
* β3: there is no statistical significance for the market value of the housing unit and the number of people in the household, all other variables held at their current level.
* β4: when the housing unit is located in the western region of the US the market value of the housing unit increases by 10.026% as compared to the southern region of the US, all other variables held at their current level.
* β5: there is no statistical significance for the market value of the housing unit and the housing unit being located in the midwest.
* β6: when the housing unit is located in the northeastern region of the US, the market value of the housing unit increases by 3.180% as compared to the southern region of the US all other variables held at their current level.
* β7: each dollar increase in area median income corresponds to a 0.178% increase in market value of the housing unit all other variables held at their current level.
* β8: a 1% change in fair market monthly rent corresponds to a 0.355% increase in market value of the housing unit all other variables held at their current level.
* β9: a 1% change in the year the housing unit was built corresponds to a 3.708% increase in market value of housing unit all other variables held at their current level.
* β10: a 1% change in the current market value of the unit corresponds to a 0.587% change in current market value of the unit, all other variables held at their current level.
* β11: a 1% increase in annual household income is associated with a 0.0468% increase in market value of the housing unit, all other variables held at their current level.
* β12: one additional room in the housing unit corresponds to a 4.969% increase in market value of the housing unit all other variables held at their current level.
* β13: there is no statistical significance for the market value of the housing unit when the housing unit is in moderately inadequate condition or severely inadequate condition compared to adequate condition, all other variables held at their current level.
* β14: when the housing unit is located outside of the "Central City" metropolitan area, the market value of the housing unit increases by 4.256% as compared to the "Central City" metropolitan area, all other variables held at their current level.
* β15: there is no statistical significance for the market value and monthly utility costs for the housing unit, all other variables held at their current level.
* β16: there is no statistical significance for the market value and the sum of other monthly costs for the housing unit, all other variables held at their current level.

Prediction in the 'Hold-out' Data:
The regression model has a R-square of 0.59 since we added the Market Value for year 2011 as an additional 'X' variable, suggesting the linear regression model explains 59% of the market value variation.
Using the coefficients from this regression model and using the set of 'X' variables in the hold out data we make predictions of the Market Value for the 1000 housing units held out. The MAD statistic (Mean Absolute Deviation) for the prediction turns out to be $100,974.19 with the average Market Value is around $251,403.08 .

Please see the worksheet 'Statistical Tests' for various calculations

Descriptive Statistics

AGE 2011	
Mean	55.22127875
Standard Error	0.109974
Median	55
Mode	54
Standard Deviation	15.46885076
Sample Variance	239.2853438
Kurtosis	-0.590702367
Skewness	0.163535941
Range	76
Minimum	17
Maximum	93
Sum	1092553
Count	19785

BEDRMS 2011	
Mean	3.249380844
Standard Error	0.005888297
Median	3
Mode	3
Standard Deviation	0.828242959
Sample Variance	0.685986399
Kurtosis	1.004500518
Skewness	0.460139258
Range	8
Minimum	0
Maximum	8
Sum	64289
Count	19785

LN_BUILT_2011	
Mean	7.585383202
Standard Error	9.52887E-05
Median	7.588323677
Mode	7.575584652
Standard Deviation	0.013403228
Sample Variance	0.000179647
Kurtosis	-0.849976792
Skewness	-0.38962876
Range	0.046827894
Minimum	7.559559496
Maximum	7.60638739
Sum	150076.8066
Count	19785

LN_VALUE_2011	
Mean	12.11047694
Standard Error	0.00532507
Median	12.08390501
Mode	12.20607265
Standard Deviation	0.749019879
Sample Variance	0.561030779
Kurtosis	0.931277436
Skewness	0.092541652
Range	7.293954882
Minimum	8.006367568
Maximum	15.30032245
Sum	239605.7862
Count	19785

REGION_MIDWEST	
Mean	0.2686884
Standard Error	0.00315151
Median	0
Mode	0
Standard Deviation	0.443288705
Sample Variance	0.196504876
Kurtosis	-0.910737423
Skewness	1.043721539
Range	1
Minimum	0
Maximum	1
Sum	5316
Count	19785

REGION_WEST	
Mean	0.197624463
Standard Error	0.00283108
Median	0
Mode	0
Standard Deviation	0.39821734
Sample Variance	0.15857705
Kurtosis	0.306782283
Skewness	1.518799289
Range	1
Minimum	0
Maximum	1
Sum	3910
Count	19785

ZADEQ_ADEQUACY	
Mean	0.02254233
Standard Error	0.001055337
Median	0
Mode	0
Standard Deviation	0.148442875
Sample Variance	0.022035287
Kurtosis	39.39430689
Skewness	6.433531275
Range	1
Minimum	0
Maximum	1
Sum	446
Count	19785

LN_VALUE_2013	
Mean	12.11570431
Standard Error	0.005484881
Median	12.10071213
Mode	11.91839057
Standard Deviation	0.771498698
Sample Variance	0.595210241
Kurtosis	0.881243026
Skewness	0.022273443
Range	5.529429088
Minimum	9.210340372
Maximum	14.73976946
Sum	239709.2098
Count	19785

Descriptive Statistics

LN_PER 2011	
Mean	0.840226356
Standard Error	0.00377795
Median	0.693147181
Mode	0.693147181
Standard Deviation	0.531403295
Sample Variance	0.282389462
Kurtosis	-0.714695205
Skewness	-0.081619987
Range	2.63905733
Minimum	0
Maximum	2.63905733
Sum	16623.87846
Count	19785

LN_LMED_2011	
Mean	11.09115698
Standard Error	0.001208243
Median	11.05888967
Mode	11.04780589
Standard Deviation	0.169950414
Sample Variance	0.028883143
Kurtosis	0.76158575
Skewness	0.555166688
Range	1.323534672
Minimum	10.42525312
Maximum	11.74878779
Sum	219438.5409
Count	19785

LN_FMR_2011	
Mean	7.033712121
Standard Error	0.002212717
Median	6.98471632
Mode	6.873163834
Standard Deviation	0.311238944
Sample Variance	0.09686968
Kurtosis	1.367973905
Skewness	0.591471501
Range	4.980758223
Minimum	6.03787092
Maximum	11.01862914
Sum	139161.9943
Count	19785

LN_ZINC2_2011	
Mean	10.97168439
Standard Error	0.006846798
Median	11.08214255
Mode	11.5125654
Standard Deviation	0.963064856
Sample Variance	0.927493917
Kurtosis	8.289383131
Skewness	-1.583833585
Range	13.13244637
Minimum	0
Maximum	13.13244637
Sum	217074.7757
Count	19785

ROOMS 2011	
Mean	6.626585797
Standard Error	0.011416061
Median	6
Mode	6
Standard Deviation	1.605773548
Sample Variance	2.578508687
Kurtosis	1.017184133
Skewness	0.768274335
Range	13
Minimum	1
Maximum	14
Sum	131107
Count	19785

REGION_NORTHEAST	
Mean	0.172251706
Standard Error	0.002684565
Median	0
Mode	0
Standard Deviation	0.37760861
Sample Variance	0.142588263
Kurtosis	1.014113978
Skewness	1.73609086
Range	1
Minimum	0
Maximum	1
Sum	3408
Count	19785

ALL_OTHER_METROS	
Mean	0.783219611
Standard Error	0.002929508
Median	1
Mode	1
Standard Deviation	0.412062173
Sample Variance	0.169795234
Kurtosis	-0.109980048
Skewness	-1.374784012
Range	1
Minimum	0
Maximum	1
Sum	15496
Count	19785

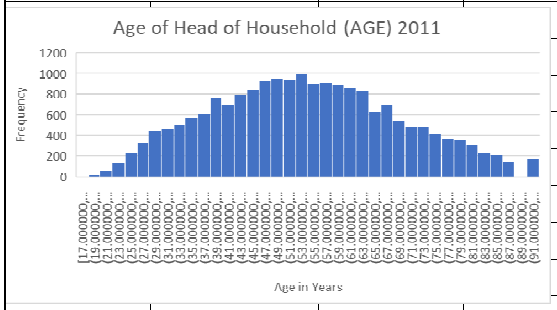
LN_UTILITY_2011	
Mean	5.90253816
Standard Error	0.371531275
Median	5.537334267
Mode	5.438079309
Standard Deviation	52.25927793
Sample Variance	2731.03213
Kurtosis	19782.01436
Skewness	140.6432433
Range	7353.793559
Minimum	2.206441232
Maximum	7356
Sum	116781.7175
Count	19785

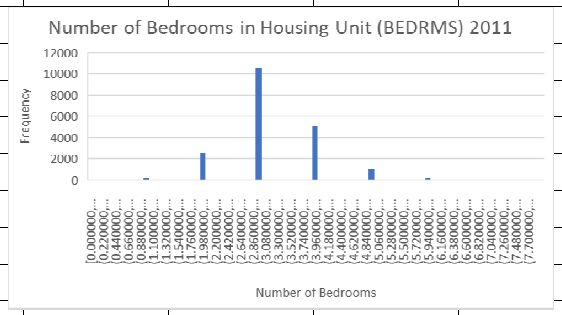
LN_OTHERCOST_2011	
Mean	4.23590979
Standard Error	0.005248702
Median	4.199705078
Mode	3.912023005
Standard Deviation	0.738278081
Sample Variance	0.545054525
Kurtosis	1.06566732
Skewness	0.080699969
Range	7.362116366
Minimum	0
Maximum	7.362116366
Sum	83807.47519
Count	19785

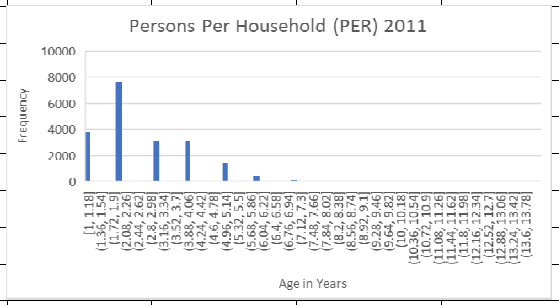
VALUE_2013	
Mean	249478.3927
Standard Error	1866.826742
Median	180000
Mode	150000
Standard Deviation	262586.2861
Sample Variance	68951557674
Kurtosis	36.20571861
Skewness	4.941396009
Range	2510000
Minimum	10000
Maximum	2520000
Sum	4935930000
Count	19785

Graphs and Charts

Histograms

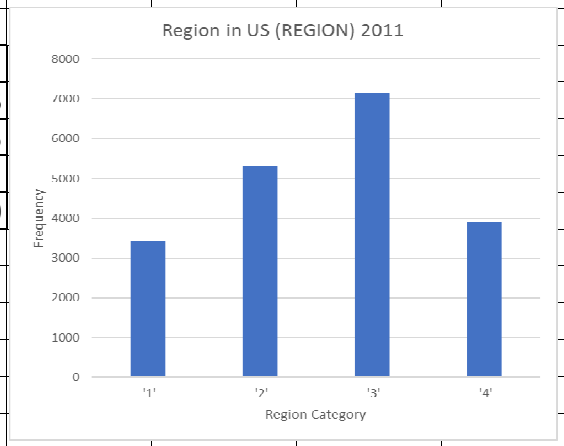




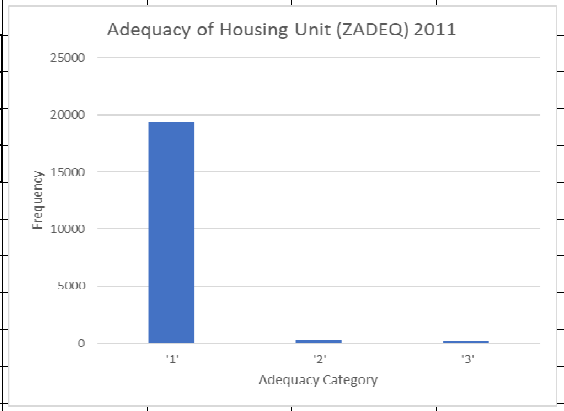


Bar Graphs

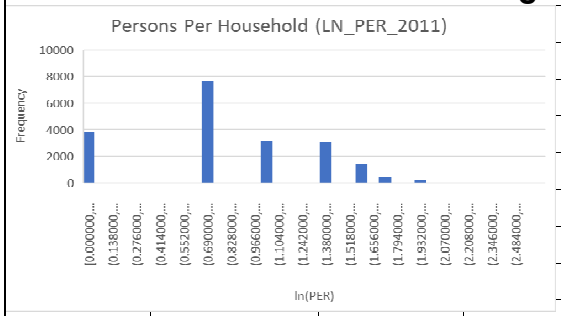
Category	Frequency
'1'	3408
'2'	5316
'3'	7151
'4'	3910

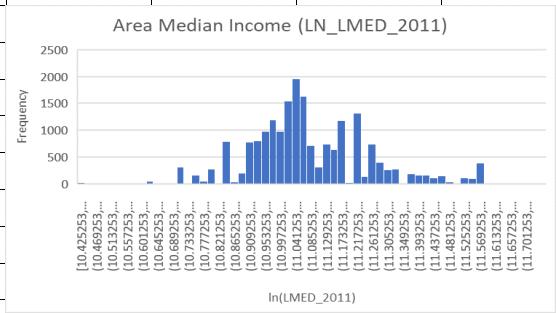


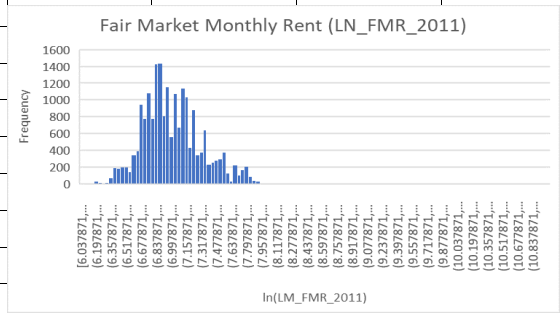
Category	Frequency
'1'	19339
'2'	284
'3'	162



Transformed Variables: all natural log

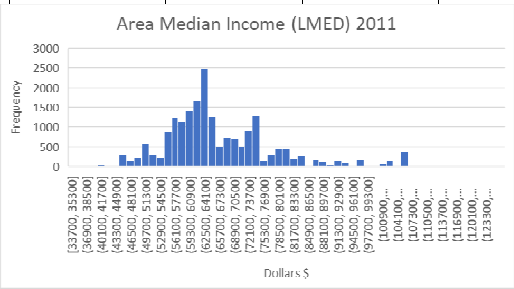


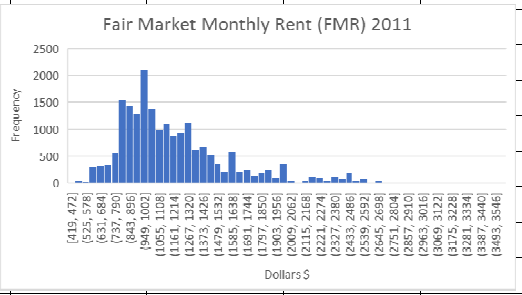


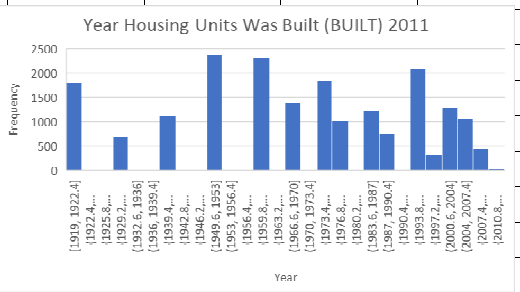


Graphs and Charts

Histograms

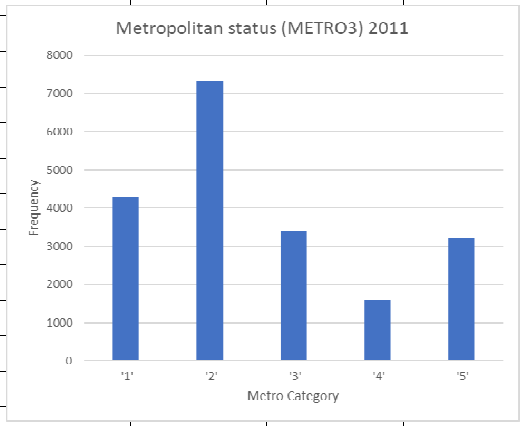




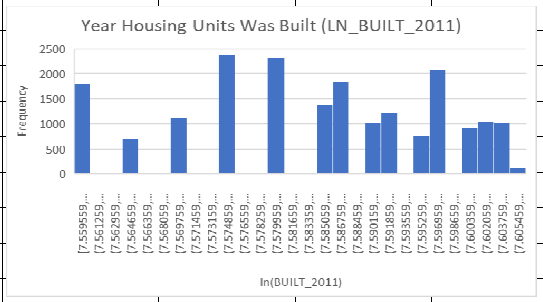


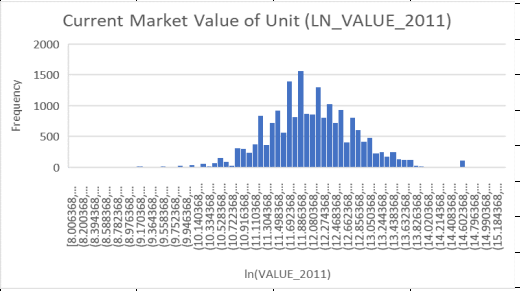
Bar Graphs

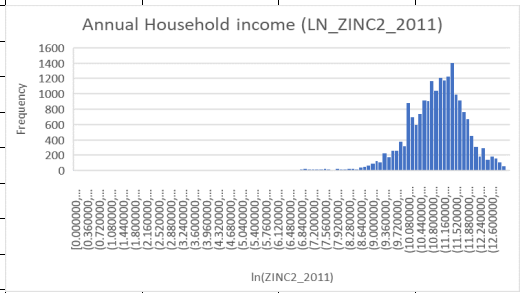
Category	Frequency
'1'	4289
'2'	7325
'3'	3375
'4'	1581
'5'	3215



Transformed Variables: all natural log

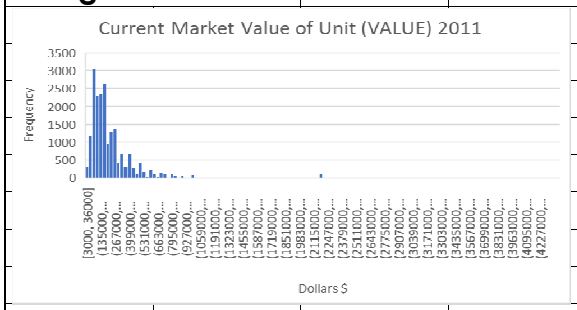


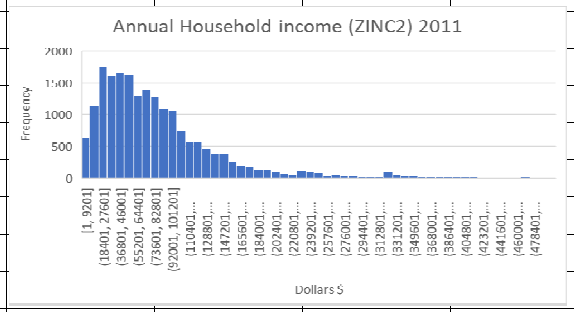


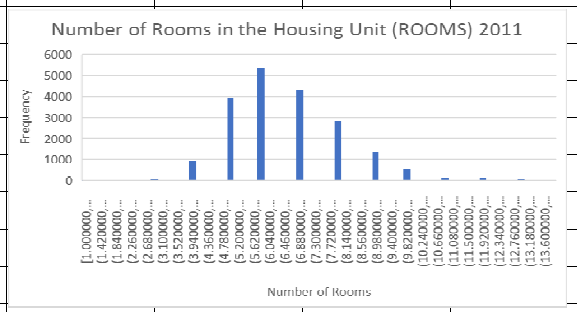


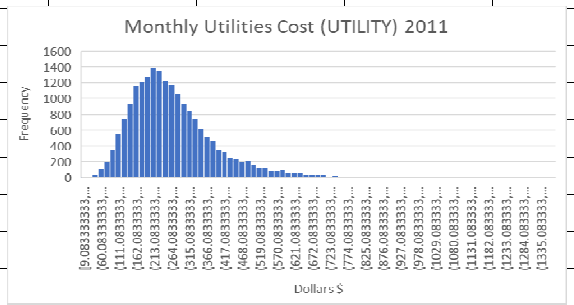
Graphs and Charts

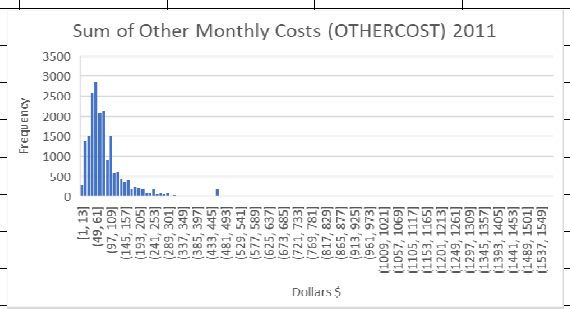
Histograms

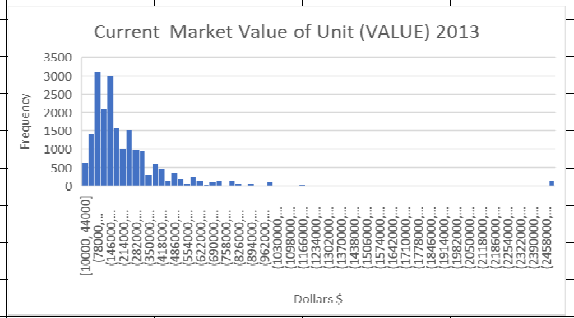




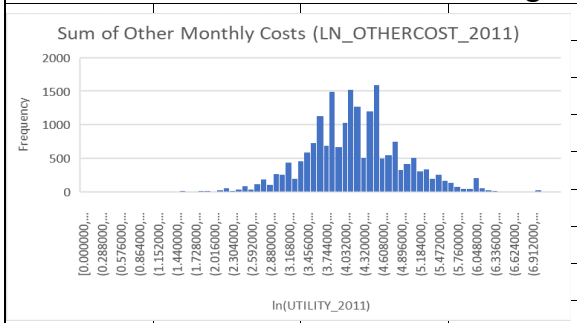


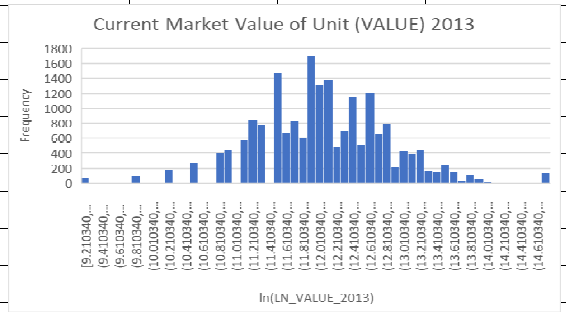






Transformed Variables: all natural log





SUMMARY OUTPUT

Ln(VALUE 2013) = β0 + β1AGE + β2BEDRMS + β3Ln(LN_PER_2011)
+ B4REGION_WEST + B5REGION_MIDWEST + B6REGION_NORTHEAST
+ β7Ln(LN_LMED_2011) + β8Ln(LN_FMR_2011) + β9Ln(LN_BUILT_2011)
+ β10Ln(LN_VALUE_2011) + β11Ln(LN_ZINC2_2011) + β12ROOMS_2011
+ β13ZADEQ_ADEQUACY + β14ALL_OTHER_METROS + β15Ln(LN_UTILITY_2011)
+ β16Ln(LN_OTHERCOST_2011)

Regression Statistics	
Multiple R	0.773180316
R Square	0.597807802
Adjusted R Square	0.597482272
Standard Error	0.489471832
Observations	19785

Note: "ALL_OTHER_METROS" is a dummy variable = 1 when METRO3 variable is = '1', otherwise it is =0

Note: "REGION_MIDWEST", "REGION_NORTHEAST" and "REGION_WEST" are dummy variables for the categorical variable REGION

Note: "ZADEQ_ADEQUACY" is a dummy variable =1 when the categorical variable ZADEQ is ="2" OR "3", otherwise it is =0

Note: This regression also takes a logarithmic transformation of many of the 'X' variables

Note: This regression also uses the natural logarithm of the 2011 Market Value as one of the 'X' variables

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	7039.569112	439.9730695	1836.414387	0
Residual	19768	4736.070301	0.239582674		
Total	19784	11775.63941			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-28.46876578	2.208516882	-12.89044517	7.24018E-38	-32.79764438	-24.13988718	-32.79764438	-24.13988718
AGE 2011	0.000868058	0.000257736	3.368013524	0.000758563	0.000362874	0.001373242	0.000362874	0.001373242
BEDRMS 2011	-0.035885372	0.007429493	-4.830123779	1.3747E-06	-0.050447802	-0.021322941	-0.050447802	-0.021322941
LN_PER 2011	-0.001230281	0.007836278	-0.15699808	0.875247965	-0.016590043	0.014129482	-0.016590043	0.014129482
REGION_WEST	0.100258313	0.011159758	8.983914661	2.84251E-19	0.07838425	0.122132376	0.07838425	0.122132376
REGION_MIDWEST	-0.002677998	0.010223408	-0.26194769	0.793364501	-0.022716737	0.017360741	-0.022716737	0.017360741
REGION_NORTHEAST	0.031800174	0.011871696	2.678654602	0.007397974	0.008530652	0.055069696	0.008530652	0.055069696
LN_LMED_2011	0.178080715	0.034186527	5.209090524	1.91672E-07	0.11107225	0.24508918	0.11107225	0.24508918
LN_FMR_2011	0.355400737	0.023762164	14.95658136	2.65904E-50	0.3088249	0.401976575	0.3088249	0.401976575
LN_BUILT_2011	3.708232994	0.287226705	12.91047428	5.59513E-38	3.145244525	4.271221463	3.145244525	4.271221463
LN_VALUE_2011	0.586730767	0.006650353	88.22550142	0	0.573695515	0.599766018	0.573695515	0.599766018
LN_ZINC2_2011	0.046791803	0.004296218	10.89139485	1.51704E-27	0.038370856	0.055212751	0.038370856	0.055212751
ROOMS 2011	0.049687272	0.003431601	14.47932762	2.85776E-47	0.042961046	0.056413498	0.042961046	0.056413498
ZADEQ_ADEQUACY	-0.036620853	0.023662464	-1.547634845	0.121726245	-0.08300127	0.009759563	-0.08300127	0.009759563
ALL_OTHER_METROS	0.042565793	0.008668048	4.910654818	9.14971E-07	0.02557569	0.059555896	0.02557569	0.059555896
LN_UTILITY_2011	3.84903E-05	6.6615E-05	0.577801714	0.563404579	-9.20808E-05	0.000169061	-9.20808E-05	0.000169061
LN_OTHERCOST_2011	0.010707253	0.00561791	1.905913896	0.056675867	-0.000304323	0.021718829	-0.000304323	0.021718829

Pairwise Correlation across all "X" variables in "Data for Estimation"																
	AGE 2011	*BEDRMS 2011*	*LN_PER 2011*	*REGION_WEST*	*REGION_MIDWEST*	*REGION_NORTHEAST*	*LN_LMED_2011*	*LN_FMR_2011*	*LN_BUILT_2011*	*LN_VALUE_2011*	*LN_ZINC2_2011*	*ROOMS 2011*	*ZADEQ_ADEQUACY*	*ALL_OTHER_METROS*	*LN_UTILITY_2011*	*LN_OTHERCOST_2011*
AGE 2011	1															
BEDRMS 2011	-0.113842141	1														
LN_PER 2011	-0.420884458	0.33144339	1													
REGION_WEST	-0.008215396	0.04535066	0.04396629	1												
REGION_MIDWEST	-0.018194589	-0.043188474	-0.020250119	-0.300818637	1											
REGION_NORTHEAST	0.022895832	-0.006608494	0.026182655	-0.226393592	-0.276506819	1										
LN_LMED_2011	-0.027009782	0.103166236	0.075704554	0.147587016	-0.091923158	0.390408332	1									
LN_FMR_2011	-0.053912095	0.46794254	0.217975391	0.425747734	-0.409588455	0.200657125	0.655984605	1								
LN_BUILT_2011	-0.151937581	0.15270929	0.114667856	0.068100776	-0.109826379	-0.191525044	-0.119031562	0.040950555	1							
LN_VALUE_2011	-0.01105979	0.34400851	0.153504992	0.253062638	-0.214230034	0.150377505	0.383143014	0.531501081	0.192223042	1						
LN_ZINC2_2011	-0.3069871	0.253593506	0.365731734	0.058837645	-0.028705184	0.045374549	0.16277149	0.225575388	0.153736244	0.369294806	1					
ROOMS 2011	-0.060209079	0.739940818	0.270393142	0.004351461	-0.014622942	0.036311417	0.140736976	0.342761774	0.129914906	0.41570461	0.31486811	1				
ZADEQ_ADEQUACY	0.020544438	-0.054359895	-0.018120514	-0.030047922	-0.025990007	0.010979435	-0.037823366	-0.045170812	-0.079025468	-0.095459756	-0.079685212	-0.052685844	1			
ALL_OTHER_METROS	0.025685303	0.023932706	0.005127435	-0.100847849	0.017268425	0.069448633	-0.068087086	-0.092601875	0.129021247	0.030824477	0.023349116	0.034026312	-0.009350927	1		
LN_UTILITY_2011	-0.01192511	0.000765193	-0.008334368	-0.00405662	-0.005840874	-0.001466795	-0.003163499	-0.001257905	0.002903875	0.005193316	0.0085949	0.004953757	-0.001227025	0.003361652	1	
LN_OTHERCOST_2011	-0.000372894	0.190489216	0.076954858	0.073909306	-0.15244744	0.010407759	0.147274902	0.297328915	0.212681702	0.502186043	0.237762222	0.250972854	-0.053872861	0.013917845	0.013469525	1