Jiho Shin
Regression Analysis
VEE Student Project
Summer 2010

Introduction:

My wife and I recently purchased a single family home in a town of Bergen County, NJ. Based on our wish list criteria, it had taken us about two years to find a home in which we both could appreciate. Having gone through the recent purchase, I wanted to use regression analysis to determine if the criteria that mattered to us most were also a driving factor in the sales price of the homes within this area.

Data:

All data was collected from the following two websites:

http://tax1.co.monmouth.nj.us/cgi-bin/prc6.cgi?menu=index&ms_user=glou&passwd=data&district=0801&mode=11

http://www.anateisenberg.com/content/article.html?id=1202390

The first website was used to extract data for sales price, the size of a home, the year in which it was built, the size of land and the taxes on the home. The second website was used to extract data on the number of bedrooms and bathrooms that the house offered.

To keep real estate market trends and the location from becoming a factor that drives the price, I have kept my analysis to 2011 home sales in one specific town of Bergen County.

Variables:

Y = Sales price of home
X1 = Number of bedrooms in home
X2 = Number of bathrooms in home
X3 = Number of half-baths in home
X4 = Square feet of home
X5 = Year in which home was built
X6 = Acreage of land on which home is built
X7 = Taxes on home

Seven Variable Equation:

$$Y = 6334.9\ X1 + 3926.6\ X2 - 5321.8\ X3 + 50.9\ X4 - 735.7\ X5 - 238616.3\ X6 + 44.2\ X7 + 1382037.0$$

Regression Statistics	
Multiple R	0.957084738
R Square	0.916011196
Adjusted R Square	0.911984336
Standard Error	196676.5329
Observations	154

	Coefficients	Standard Error	t Stat	P-value
Intercept	1382037.01	1272516.629	1.086065972	0.279239724
Rooms	6334.915655	24916.32831	0.254247559	0.799662042
Bathroom	3926.599757	27789.70154	0.141296939	0.887830141
Half Bath	-5321.791047	28421.16439	-0.187247467	0.851726572
Sq. Ft.	50.8636328	38.99656577	1.304310567	0.194180158
Yr. Built	-735.6900577	659.7666373	-1.115076174	0.266649875
Acreage	-238616.2815	100204.6612	-2.381289239	0.018541707
Taxes	44.15123986	3.581057934	12.32910516	2.19619E-24

The model provides a high R Square and Adjusted R Square of 0.916 and 0.912 which suggests that the model is good predicator of the sales price of a home. Although, based on the high P-values, the analysis implies that the number of rooms and/or bathrooms is not a driving factor on the price of the home. I suspect that this is the case since the number of rooms and bathrooms can always be increased (provided the size of the home is large enough), where-as other factors cannot be changed such as the size of the land.

We will remove the variables representing rooms and bathrooms for our next analysis.

Four Variable Equation:

$$Y = 57.8\ X4 - 692.4\ X5 - 238764.1\ X6 + 44.1\ X7 + 1312354.2$$

Regression Statistics	
Multiple R	0.957029771
R Square	0.915905982
Adjusted R Square	0.913648425
Standard Error	194808.405
Observations	154

	Coefficients	Standard Error	t Stat	P-value

Intercept	1312354.205	1235624.709	1.062097735	0.289909742
Sq. Ft.	57.83181659	32.10345913	1.801420101	0.073658723
Yr. Built	-692.4185297	638.8353752	-1.083876311	0.280170393
Acreage	-238764.0536	99182.22222	-2.407327122	0.017293129
Taxes	44.09899745	3.48392914	12.65783421	2.14843E-25

This new model also provides for a high R Square and Adjusted R Square value (immaterial reduction from previous model). Although the year in which the house was built has a high p-value suggesting that it is also not a driving factor in the sales price of a home. I suspect that this is the case since provided the house is in relatively good condition, the appearance of the home can always be improved upon.

We will remove the variable representing the year in which a home was built for our next analysis.

Three Variable Equation:

$Y = 57.6\ X4 - 233404.8\ X6 + 43.3\ X7 - 26279.8$

Regression Statistics	
Multiple R	0.956683304
R Square	0.915242944
Adjusted R Square	0.913547803
Standard Error	194921.8725
Observations	154

	Coefficients	Standard Error	t Stat	P-value
Intercept	-26279.78103	37928.56772	-0.69287565	0.489458948
Sq. Ft.	57.62470296	32.12158897	1.793955555	0.074834546
Acreage	-233404.7759	99116.59239	-2.35485069	0.019825302
Taxes	43.26952877	3.400819043	12.72326702	1.28997E-25

This model provides for a high R Square and Adjusted R Square value and low P-values for all listed independent variables. I suspect that this model could be a good predicator of the sales price of a home.

Although I am curious to see what the model would show provided the variable representing square feet were excluded. As previously mentioned certain criteria of a home such as the size of the land cannot be changed. This would also almost always be the case for the size of the home except for those elect rich. Comparatively to other towns in the vicinity, this particular town does have a larger group of rich homeowners.

Two Variable Equation:

Y = -202236.6 X6 + 48.2 X7 + 18365.4

Regression Statistics	
Multiple R	0.955732424
R Square	0.913424467
Adjusted R Square	0.912277771
Standard Error	196348.4094
Observations	154

	Coefficients	Standard Error	t Stat	P-value
Intercept	18365.38007	28831.86869	0.636981954	0.525101134
Acreage	-202236.5874	98296.13484	-2.057421563	0.041365731
Taxes	48.19178673	2.023964204	23.81059242	5.49746E-53

This model still provides for a high R Square and Adjusted R Square value and low P-values for all listed independent variables. It is my assumption that this model is more representative of the rich group of people that live in the town. Acreage is something that they cannot typically buy more of for a single home and all forms of taxes (both real estate and other) are something that I believe the rich really work to avoid within the confines of the law.

Conclusion:

According to the regression analysis, it would appear that the following model would be the best predicator for the sales price of a home for the average of home owners in this town:

Y = 57.6 X4– 233404.8 X6 + 43.3 X7 – 26279.8

Y = Sales price of home
X4 = Square feet of home
X6 = Acreage of land on which home is built
X7 = Taxes on home

It is also representative of the more significant criteria that were used in choosing and purchasing our home.