INFO589_91FA23
APPLIED STATS FOR BUSINESS ANALYTICS

PREPARED BY:
GROUP J

FARZANA ZANNAT
SYLVIA ISLAM PROGYA
ABHIMELECH TYAGI
NAFIZ IMTIAZ

23 September 2023

Table of Contents

1.Introduction

Welcome to Montclair & Millburn Real Estate, where we provide you with extensive insights into the ever-changing Montclair and Millburn, New Jersey real estate markets. To provide educated and data-driven advice to our customers, we have conducted a thorough examination of local real estate market trends. Our research is based on a solid dataset that includes data from 93 property sets in these two diverse but related areas. We want to provide a helpful peek into the dynamics of the local housing market through this vast dataset, revealing the complicated linkages that influence property values and transactions in particular places. At Montclair & Millburn Real Estate, we recognize the importance of an accurate valuation methodology in making informed investment decisions. Furthermore, our dedication to market knowledge goes beyond value. We will get significant insights into larger industry trends and effective sales methods by attentively reviewing the results of this regression analysis.

2.Background

Multiple regression is a statistical modelling approach that is used to investigate the associations between a dependent variable and several independent variables. In the context, it appears we want to design a multiple regression model which has 93 sets to understand how numerous independent factors impact the dependent variable, which is most likely the "PRICE" of a property (k).

Dependent Variable (PRICE): "PRICE(k)" denotes the dependent variable in this case. It is the variable that we wish to explain.

Independent Variable: ASSESSVAL(k): This variable may reflect the property's assessed value. Assessed values are frequently utilized in property taxation and can be used to anticipate the market value of a property.

ROOMS: The number of rooms in a home can have a considerable impact on its pricing. Larger residences with more rooms tend to be more expensive.

TAXAMOUNT(k): The amount of property taxes is another crucial indicator. Higher tax bills may suggest that the property is in a more attractive location or has a higher worth.

AGE: The age of the property might also have an impact on its pricing. Due to depreciation, older properties may attract lower prices, whilst newer properties may fetch greater prices.

STYLECODE: This variable appears to be a code or category for the prop's style or kind.

STYLE_Label: This variable is expected to include human-readable labels relating to the style codes. It may be used to analyze and visualize data. This is Categorical Variable which is needed to change into numerical before analyzing.

TOWN: The town in which a property is located can have a substantial influence on its pricing. Property prices in attractive or wealthy towns tend to be higher and this is Categorical Variable too and we need to make it numerical and continuous.

The goal of creating a multivariate regression model using these variables is to better understand the link between these independent factors and the price of the property. The model can assist in answering questions such as: Which variables have the greatest influence on property prices? What effect do changes in these factors have on property prices? What is the model's total predictive power? In conclusion, a multiple regression model based on these variables might provide important insight into the factors driving property values in a certain location. It may be a useful tool for real estate professionals, policymakers, and investors to make educated property appraisal and investment decisions.

3.Simple Regression Modeling

Our goal in this analysis is to make a model that shows a relationship between the price of the house sold (in thousand dollars) and the assessed value of the house (in thousands of dollars). In this data set, we have gotten 93 observations from different houses in Montclair and Millburn. To make a visualization, we have created a scatter plot with the price of the house sold (PRICE (K)) in the x-axis and the assessed value of the house (ASSESSVAL (k)) in the y-axis. This plot shows that there is a clear positive correlation between the assessed value of the house and the price of the house in which it is sold, which indicates that houses that has bigger assessed value are sold at a high price. Here, we performed a simple linear regression analysis where the price of the house is the dependent variable on the Y-axis, and the assessed value of the house on the X-axis.

- We have shown the scatter plot and Excel regression output in *Appendix A, B and C.*
- The sample regression equation is: y = 1.0263 * (Price (K)) + 389.24
- The Y-intercept of 389.24 represents the assessed value (in thousand) for a hypothetical price of a house at $0 (if X=0). The slope of 1.0263 indicates the expected $1,026.3 increase in the assessed value for each additional $1000 of the price of the house, assuming all other factors are equal.
- R-squared of 0.7809 means that approximately 78.09% of the variability in assessed house value can be explained by the variation in the sale price of the house. The remaining 28.91% is due to the other factors.
- The standard error of 235.916 thousand dollars is the standard deviation of the plots or the typical amount that the observed values deviate from the regression line that we have got. This standard error seems appropriate given the price range of the house.
- The residual plots in Appendix C confirm that the assumptions of linearity, normality, and equal variance are satisfied, indicating that the simple linear regression model that we did is appropriate.
- The F-test p-value of 9.45E-32 shows that the overall fit is statistically significant (at the 5% level), meaning that a significant linear relationship exists between assessed value of the house and the price of the house at which it is sold. The p-value is significantly below the 0.05 significance level, and this indicates strong evidence of a significant linear relationship.
- The 95% confidence interval for the slope (0.9131 to 1.1395) suggests that for an increase of $1000 in the house price, the average assessed value of the house is expected to change between $913.1 and $1139.5 with 95% confidence.
- For a house sold at $500 thousand dollars, the predicted assessed value of the house is:
- y = 1.0263*(500) + 389.24 = $902.39 thousand dollars
- The simple linear regression analysis indicates that a statistically significant positive relationship between the price of a house sold and assessed value of the house in Montclair and Millburn. On average, an additional thousand dollars is associated with an extra $102.63 thousand dollars of sales price. This model appears appropriate based on the residual analysis and explains on more than half of the observed variation in home prices.

Findings: It was confirmed by our simple linear regression model that there was a significant and positive correlation, and for every $1000 in assessed value, we were seeing a sale price increase by $1026.3 which is a 78.09 of the variability in assessed house values. Our findings substantiate our initial hypothesis and provides our homeowners, investors and buyers with a tool to estimate sale prices based on their assessed values

4.Multiple Regression Modeling

4.1.a. Scatterplot matrix (use JMP)

	PRICE(k)	ASSESSVAL(k)	ROOMS	TAXAMOUNT(k)	AGE	TOWN	STYLECODE
PRICE(k)	1.0000	0.8837	0.6620	0.7739	-0.2270	-0.3892	0.0734
ASSESSVAL(k)	0.8837	1.0000	0.5869	0.6362	-0.3524	-0.6261	0.0566
ROOMS	0.6620	0.5869	1.0000	0.6995	0.1145	-0.1250	-0.0217
TAXAMOUNT(k)	0.7739	0.6362	0.6995	1.0000	-0.0151	-0.0667	0.0610
AGE	-0.2270	-0.3524	0.1145	-0.0151	1.0000	0.3933	0.0484
TOWN	-0.3892	-0.6261	-0.1250	-0.0667	0.3933	1.0000	0.0256
STYLECODE	0.0734	0.0566	-0.0217	0.0610	0.0484	0.0256	1.0000

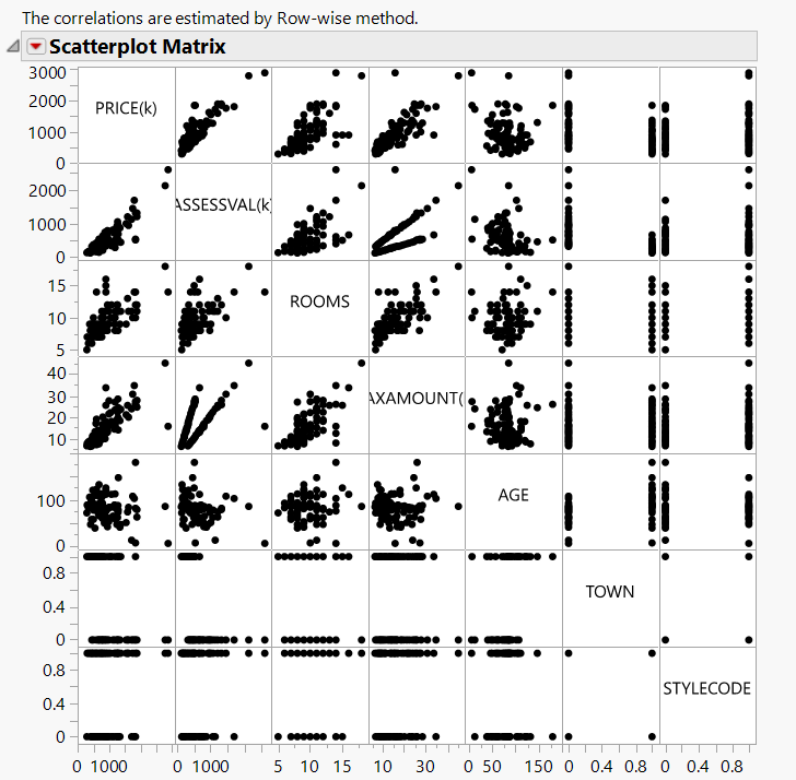


Table 1: Scatterplot Matrix

For the ease of simplifying the report, we have assigned the value 0 to “Millburn” and 1 to "Montclair” in the "Town" column. Similarly, in the "STYLE_Label" column, we assigned the value 1 to "Colonial" and 0 to all other styles ("Bi-Level," "Cape Cod," "Custom Home," "Ranch," "Split Level," "Tudor," "Victorian") to streamline the analysis.

Based on the scatterplot, it is evident that the independent variable, which is PRICE(k), exhibits a strong linear correlation with the dependent variable ASSESSVAL(k), ROOMS, and TAXAMOUNT(k). However, there is no observable correlation between PRICE(k) and AGE.

4.1.b. Preliminary multiple regression model

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	19930532	3321755	89.7671
Error	86	3182357	37004	Prob > F
C. Total	92	23112890		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	37.344941	105.5978	0.35	0.7245	.
ASSESSVAL(k)	0.8585942	0.102372	8.39	<.0001*	4.8524437
ROOMS	10.935397	13.57631	0.81	0.4228	2.4358488
TAXAMOUNT(k)	18.360568	4.556036	4.03	0.0001*	2.8912587
AGE	-0.158656	0.877033	-0.18	0.8569	1.422113
STYLECODE	14.058248	41.62099	0.34	0.7364	1.032925
TOWN	100.93072	62.22125	1.62	0.1084	2.4299524

Table 2: Multiple Regression Model

Based on the regression model generated by JPM, the significant value is less than 0.0001 which is less than α(0.05) means that the preliminary model is statistically significant. As per the model the y intercept is 37.34 and with this we can create the regression line as below:

$y= \beta 0 + \beta 1.X1+ \beta 2.X2+\ldots +\beta n.Xn$

4.1.c. Collinearity Assessment with VIF

All the independent (predictor variable) do not show any significant associations with one another, according to the regression model we built in table 2. This inference is made because the VIF for each variable is less than 5 hence this model does not have any collinearity problems.

4.1.d. Stepwise Regression Approaches

(i) Froward Method:

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.05

Prob to Leave 0.05

Direction: Forward

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

Training Rows 93

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
3324519	90	192.19547	0.8562	0.8530	2.8417772	3	1247.411	1257.087

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	155.227055	1	0	0.000	1
☐	☑	ASSESSVAL(k)	0.76355582	1	5946229	160.974	9.4e-22
☐	☐	ROOMS	0	1	36415.77	0.986	0.32349
☐	☑	TAXAMOUNT(k)	23.8156762	1	1740209	47.110	8.3e-10
☐	☐	AGE	0	1	2870.879	0.077	0.78216
☐	☐	STYLECODE	0	1	4866.496	0.130	0.7188
☐	☐	TOWN	0	1	115782.8	3.211	0.07652

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	ASSESSVAL(k)	Entered	0.0000	18048162	0.7809	47.869	2	1284.38	1291.71	○
2	TAXAMOUNT(k)	Entered	0.0000	1740209	0.8562	2.8418	3	1247.41	1257.09	◉

Table 3: Stepwise Regression Model (Forward Method)

Based on the Forward method in Table 3, the assessed value (ASSESSVAL(k)) and taxes paid in the prior year (TAXAMOUNT(k)) are the most important independent factors (predictors) to determine the house price (dependent variable).

(ii) Backward Method:

Stepwise Fit for PRICE(k)

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.05

Prob to Leave 0.05

Direction: Backward

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

Training Rows 93

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
3324519	90	192.19547	0.8562	0.8530	2.8417772	3	1247.411	1257.087

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	155.227055	1	0	0.000	1
☐	☑	ASSESSVAL(k)	0.76355582	1	5946229	160.974	9.4e-22
☐	☐	ROOMS	0	1	36415.77	0.986	0.32349
☐	☑	TAXAMOUNT(k)	23.8156762	1	1740209	47.110	8.3e-10
☐	☐	AGE	0	1	2870.879	0.077	0.78216
☐	☐	STYLECODE	0	1	4866.496	0.130	0.7188
☐	☐	TOWN	0	1	115782.8	3.211	0.07652

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	All	Entered	.	.	0.8623	7	7	1252.61	1271.15	○
2	AGE	Removed	0.8569	1210.963	0.8623	5.0327	6	1250.25	1266.66	○
3	STYLECODE	Removed	0.7477	3811.831	0.8621	3.1357	5	1248.02	1262.23	○
4	ROOMS	Removed	0.4446	21356.08	0.8612	1.7129	4	1246.35	1258.32	○
5	TOWN	Removed	0.0765	115782.8	0.8562	2.8418	3	1247.41	1257.09	◉

Table 4: Stepwise Regression Model (Backward Method)

Based on the Backward method in Table 4, the assessed value (ASSESSVAL(k)) and taxes paid in the prior year (TAXAMOUNT(k)) are the most important independent factors (predictors) to determine the house price (dependent variable).

(iii) Mixed Method:

Stepwise Fit for PRICE(k)

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.05

Prob to Leave 0.05

Direction: Mixed

Enter All | Make Model | Remove All | Run Model

Go | Stop | Step

Training Rows 93

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
3324519	90	192.19547	0.8562	0.8530	2.8417772	3	1247.411	1257.087

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	155.227055	1	0	0.000	1
☐	☑	ASSESSVAL(k)	0.76355582	1	5946229	160.974	9.4e-22
☐	☐	ROOMS	0	1	36415.77	0.986	0.32349
☐	☑	TAXAMOUNT(k)	23.8156762	1	1740209	47.110	8.3e-10
☐	☐	AGE	0	1	2870.879	0.077	0.78216
☐	☐	STYLECODE	0	1	4866.496	0.130	0.7188
☐	☐	TOWN	0	1	115782.8	3.211	0.07652

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	ASSESSVAL(k)	Entered	0.0000	18048162	0.7809	47.869	2	1284.38	1291.71	○
2	TAXAMOUNT(k)	Entered	0.0000	1740209	0.8562	2.8418	3	1247.41	1257.09	◉

Table 5: Stepwise Regression Model (Mixed Method)

Based on the Forward method in Table 5, the assessed value (ASSESSVAL(k)) and taxes paid in the prior year (TAXAMOUNT(k)) are the most important independent factors (predictors) to determine the house price (dependent variable).

Determination of independent variables based on r2 from Forward, Backward and Mixed Method which should be included in our regression model:

According to the tables 3, 4, and 5 show that the Adjusted R square value of the across all different approaches is 0.8530. Therefore, for this model, any of the methods will work. According to the adjusted R square value, we can state that this model can account for 85% of the variation in house prices by varying assessed values (ASSESSVAL(k)) and taxes paid in the previous year (TAXAMOUNT(k)).

4.1.e. Comment on the Consistency of Findings

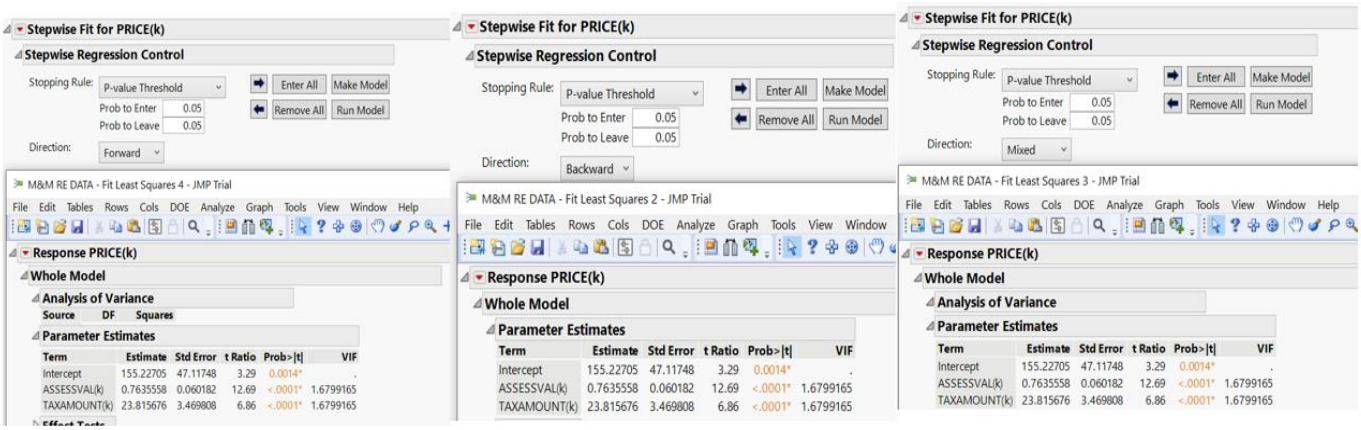


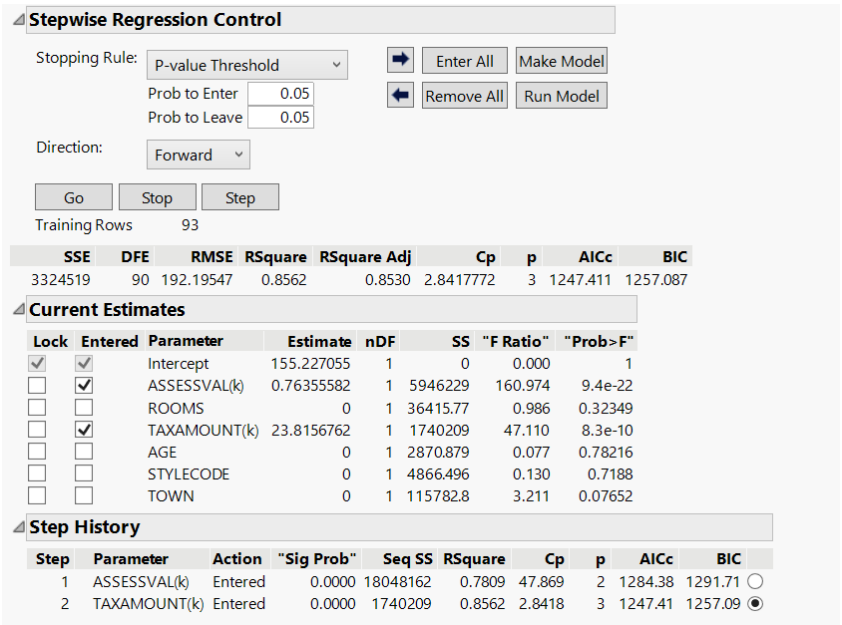
Table 6. VIF determination from Forward, Backward and Mixed Method

With a consistent VIF value of 1.699 observed across various approaches, it can be concluded that this model maintains its stability and consistency across these different approaches.

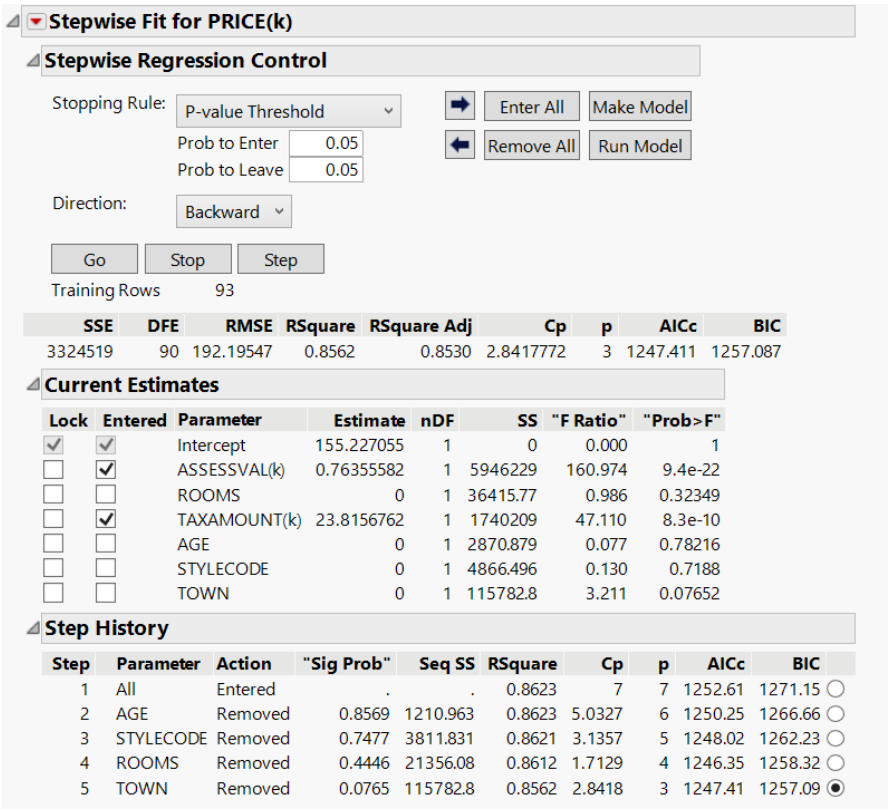
4.1.f. Screenshots of Outputs

The outputs from Step 4.1.d above into our report are shown as (1), (2), and (3) in the screenshots below:

(1).



(2).



Stepwise Fit for PRICE(k)

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.05

Prob to Leave 0.05

Direction: Backward

Enter All | Make Model | Remove All | Run Model | Go | Stop | Step

Training Rows 93

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
3324519	90	192.19547	0.8562	0.8530	2.8417772	3	1247.411	1257.087

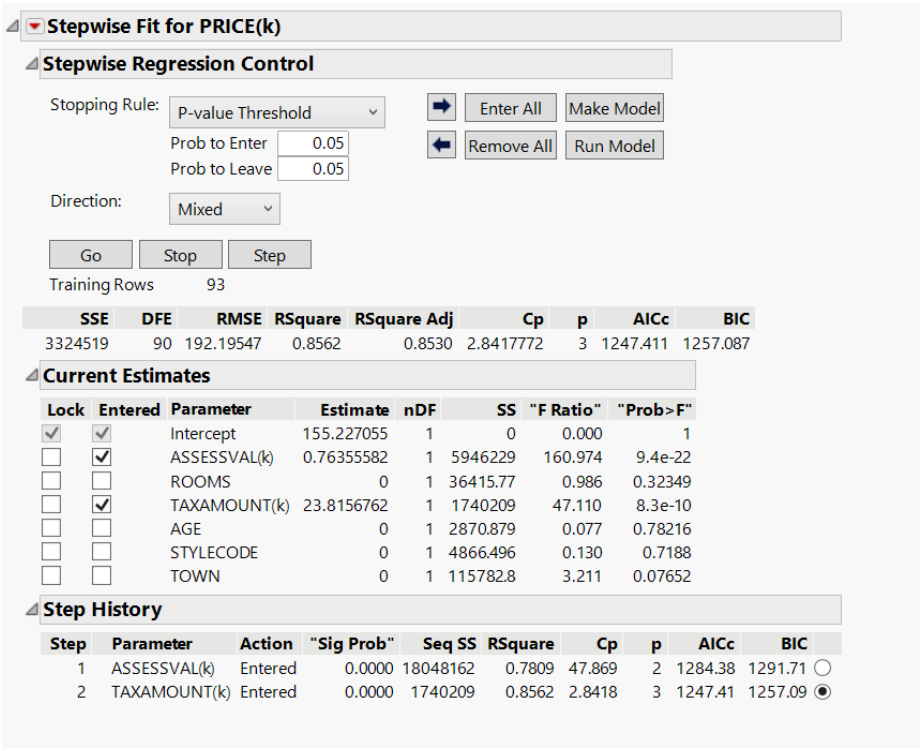
Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	155.227055	1	0	0.000	1
☐	☑	ASSESSVAL(k)	0.76355582	1	5946229	160.974	9.4e-22
☐	☐	ROOMS	0	1	36415.77	0.986	0.32349
☐	☑	TAXAMOUNT(k)	23.8156762	1	1740209	47.110	8.3e-10
☐	☐	AGE	0	1	2870.879	0.077	0.78216
☐	☐	STYLECODE	0	1	4866.496	0.130	0.7188
☐	☐	TOWN	0	1	115782.8	3.211	0.07652

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC
1	All	Entered	.	.	0.8623	7	7	1252.61	1271.15
2	AGE	Removed	0.8569	1210.963	0.8623	5.0327	6	1250.25	1266.66
3	STYLECODE	Removed	0.7477	3811.831	0.8621	3.1357	5	1248.02	1262.23
4	ROOMS	Removed	0.4446	21356.08	0.8612	1.7129	4	1246.35	1258.32
5	TOWN	Removed	0.0765	115782.8	0.8562	2.8418	3	1247.41	1257.09

(3).



Stepwise Fit for PRICE(k)

Stepwise Regression Control

Stopping Rule: P-value Threshold

Prob to Enter 0.05

Prob to Leave 0.05

Direction: Mixed

Enter All | Make Model | Remove All | Run Model | Go | Stop | Step

Training Rows 93

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
3324519	90	192.19547	0.8562	0.8530	2.8417772	3	1247.411	1257.087

Current Estimates

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	155.227055	1	0	0.000	1
☐	☑	ASSESSVAL(k)	0.76355582	1	5946229	160.974	9.4e-22
☐	☐	ROOMS	0	1	36415.77	0.986	0.32349
☐	☑	TAXAMOUNT(k)	23.8156762	1	1740209	47.110	8.3e-10
☐	☐	AGE	0	1	2870.879	0.077	0.78216
☐	☐	STYLECODE	0	1	4866.496	0.130	0.7188
☐	☐	TOWN	0	1	115782.8	3.211	0.07652

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC
1	ASSESSVAL(k)	Entered	0.0000	18048162	0.7809	47.869	2	1284.38	1291.71
2	TAXAMOUNT(k)	Entered	0.0000	1740209	0.8562	2.8418	3	1247.41	1257.09

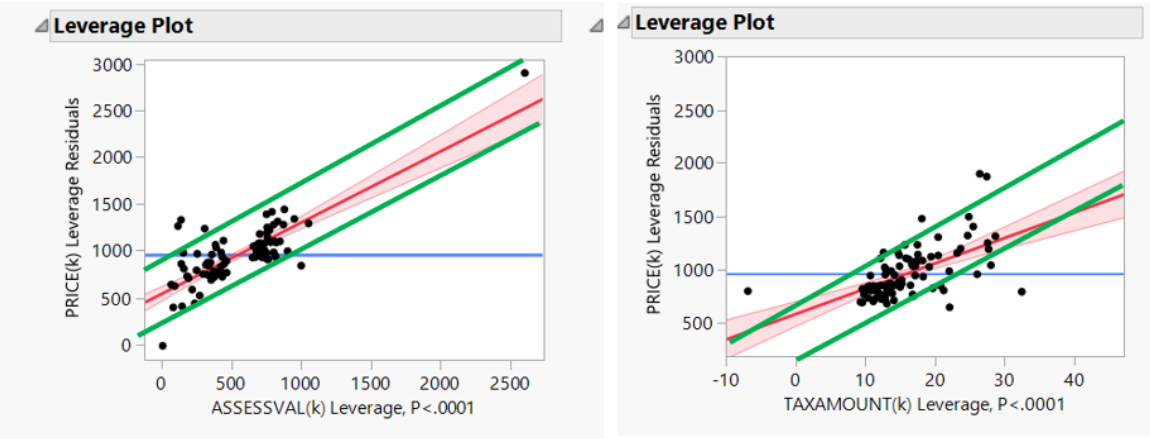
4.1.g. Chosen Multiple Regression Model

From the procedures carried out in 4.1.d , it is evident that when predicting house value, the two primary predictors are assessed value (ASSESSVAL(k)) and taxes paid in the prior year (TAXAMOUNT(k)). While all approaches yield consistent results, we have opted for the Backward Approach. This choice is motivated by the limitations associated with the forward approach, including computational complexity and the lack of control over variable order, among others. Similarly, the mixed approach is ruled out due to its lengthiness and time-consuming nature.

4.1.h. Multiple Regression Assumptions

Regression Assumptions Test:

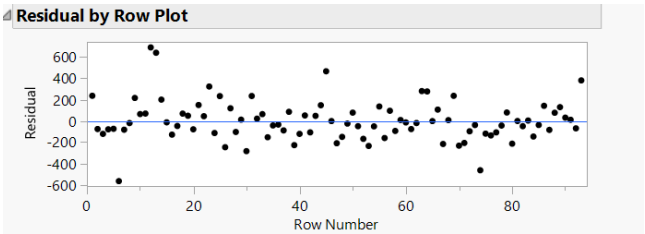
(i) Linearity Test:



Graph A: Relationships between ASSESSVAL(k) Residuals and TAXAMOUNT(k) & residuals

Linearity test determine the relationship between the independent variables and the dependent variable should be linear. The graph A shows that the linearity condition is met because the residuals of the price per ASSESSVAL(k) and TAXAMOUNT(k) are contained within two parallel lines.

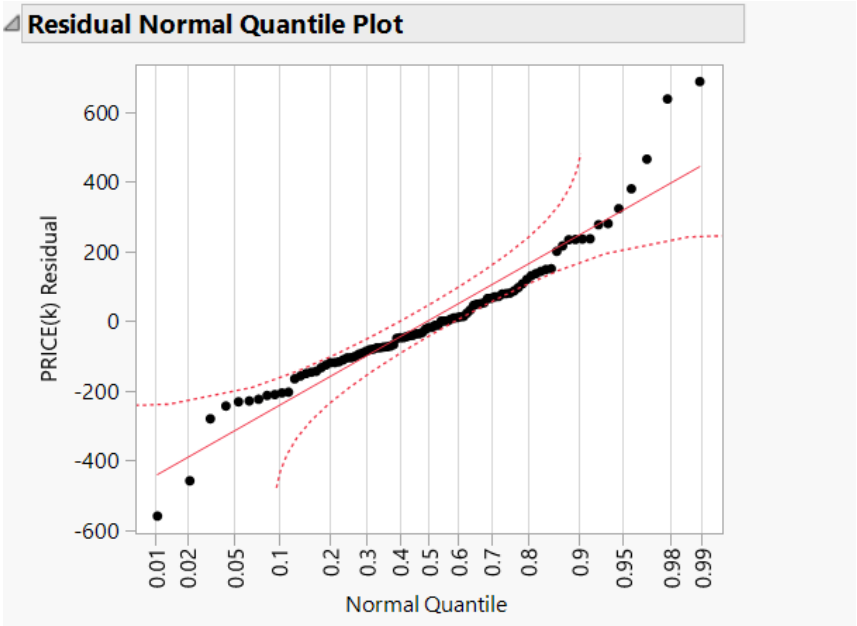
(ii) Independence Test:



Graph B: Residual by Row Plot

The independence test shows cyclical pattern. This test is typically significant when dealing with time series data. However, given that the dataset in question is not a time series dataset hence it is reasonable to assume that the independence assumption is met.

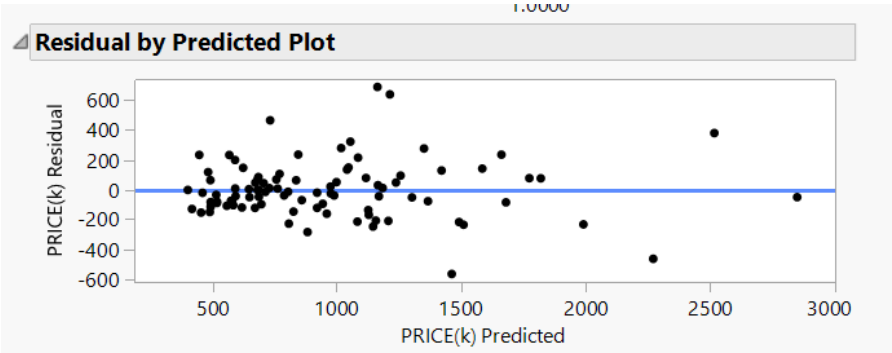
(iii) Normality Test:



Graph C: Normal Probability Test

Since the residuals of price shows straight line hence it can be concluded that the normality is satisfied.

(iv) Equal variance Test:



Graph D: Residual Variation Plot

According to the graph D , the variance of the residuals is consistent across all fitted values hence Equal Variance assumptions is satisfied.

4.1.i. Significance of the Overall Fitted Model Assessment

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	19788371	9894185	267.8513
Error	90	3324519	36939	Prob > F
C. Total	92	23112890		<.0001*

The importance of the model is shown by a significance level less than 0.0001. Because significance F is less than α(0.05), we can state that we are 95% certain that this model is helpful. There is a positive linear relationship between house price and assessed value and tax amount this indicates that the price of the house varies with change of assessed value and tax paid in the previous year.

4.1.j. Significance of Each Predictor Variable Assessment

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	155.22705	47.11748	3.29	0.0014*	61.619951	248.83416
ASSESSVAL(k)	0.7635558	0.060182	12.69	<.0001*	0.6439947	0.8831169
TAXAMOUNT(k)	23.815676	3.469808	6.86	<.0001*	16.922297	30.709055

Effect Tests

The significance value for ASSESSVAL(k) and TAXAMOUNT(k)is less than 0.0001. In both cases we observe that it is lower than a(0.05) hence the both independent variables are significant.

We can say, we are 95% confident houses prices will increase between 0.64 and 0.88 in every 1 dollar changed in assessed value.

Additionally, we have a 95% confidence level that property prices will fluctuate by 1 dollar for every dollar of taxes paid in the previous year, between 16.92 and 30.71.

4.2.a. Meaning of the Y Intercept and Slopes of The Model

The Y-intercept represents the predicted value of the dependent variable which is PRICE(k) and all independent variables (ASSESSVAL(k) and TAXAMOUNT(k)) X are set to zero. According to our model, it means that estimated house price depends on change in ASSESSVAL(k) and TAXAMOUNT(k).

The slopes (coefficients) for each independent variable represent the change in the dependent variable (Y) associated with a one-unit change in the corresponding independent variable while holding all other variables constant. The unit of measurement for the change in Y depends on the units of the independent variable.

For Example: if the ASSESSVAL(k) value is 100 and TAXAMOUNT(k) is 300 then the house price will be

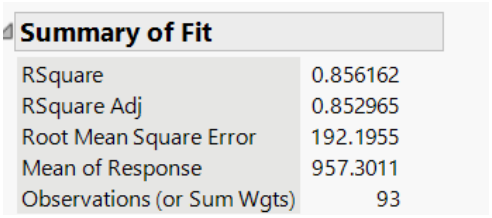
$Y= \beta 0 + \beta 1x1+ \beta 2x2$

House price = 155.22+(0.76 x 100) + (23.82 x 300)

= 7377.22 K

Above house price will be changed if any changes in ASSESSVAL(k) or TAXAMOUNT(k)

4.2.b. Meaning of the Coefficient of Multiple Determination r2

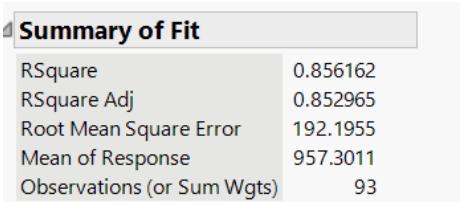


Summary of Fit	
RSquare	0.856162
RSquare Adj	0.852965
Root Mean Square Error	192.1955
Mean of Response	957.3011
Observations (or Sum Wgts)	93

An R-squared value between 0 and 1 represents the proportion of the variance in the dependent variable that is accounted for by the independent variables in the model.

According to our model $R^2 = 0.86$ (or 86%), it means that 86% of the variation in the house price can be explained by the predictors in the model, and the remaining 20% is due to unexplained factors or random variation.

4.2.c. Meaning of the Standard Error of the Estimate SYX



Summary of Fit	
RSquare	0.856162
RSquare Adj	0.852965
Root Mean Square Error	192.1955
Mean of Response	957.3011
Observations (or Sum Wgts)	93

Standard error of estimate represents the average distance that the observed values fall from the regression line. Our model predicts a standard error of 192K. Our dependent variable, which is the range of house price, is between 289K and 2895K, hence the standard error is not very high.

4.2.d. Determination the 95% Confidence Interval Estimate

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	155.22705	47.11748	3.29	0.0014*	61.619951	248.83416
ASSESSVAL(k)	0.7635558	0.060182	12.69	<.0001*	0.6439947	0.8831169
TAXAMOUNT(k)	23.815676	3.469808	6.86	<.0001*	16.922297	30.709055

We are 95% confident that the true coefficient of ASSESSVAL(k) is between (0.66, 0.88) and the true coefficient of TAXAMOUNT(k) is between (16.92, 30.71).

Since the coefficient range of both independent variables in our model is positive, the expectation is that the price of a house will rise if ASSESSVAL(k) or TAXAMOUNT(k) increase by one unit while all other variables remain constant. Or to put it another way, there is a favorable link. The projected growth in the dependent variable is greater the more the independent variable increases within this range.

We need to consider the confidence interval estimate of coefficients because we are working with samples but not population while expecting the outcome to work on population. The confidence interval is crucial since we will obtain different coefficient estimate values each time, we use a different sample.

4.2.e. Prediction of $\hat{y}$ Value

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	155.22705	47.11748	3.29	0.0014*	61.619951	248.83416
ASSESSVAL(k)	0.7635558	0.060182	12.69	<.0001*	0.6439947	0.8831169
TAXAMOUNT(k)	23.815676	3.469808	6.86	<.0001*	16.922297	30.709055

Assuming the ASSESSVAL(k) is 660.8 and the TAXAMOUNT(k) is 13.3

$\hat{y}$ = 155.22+(0.76 x 660.8) + (23.82 x 13.3)

= 974.234 K USD

5.Summary of Findings

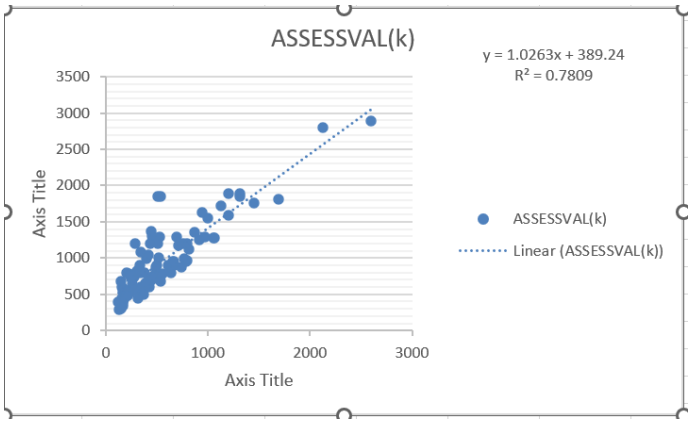
In our analysis, we have employed simple linear regression in the previous project, and in this project, we have used multiple regression approaches to thoroughly examine the dynamics of the real estate market. Through simple linear regression, we have established a noteworthy finding—a significant and positive correlation exists between assessed value and sale price, where every $1000 increase in assessed value results in a corresponding $1026.3 increase in sale price. This relationship explains a substantial 78.09% of the variability in assessed house values, providing a valuable tool for homeowners, investors, and buyers to estimate sale prices based on assessed values.

In contrast, our multiple regression analysis has revealed that the two primary predictors for house value prediction are assessed value (ASSESSVAL(k)) and taxes paid in the prior year (TAXAMOUNT(k)). For our multiple regression approach, we have chosen the Backward Approach instead of the other approaches, prioritizing its efficiency and control over variable order while maintaining accuracy, thus allowing us to gain deeper insights into the intricate relationships that drive house values in the real estate market. With a 95% confidence level, we have estimated that the true coefficient of ASSESSVAL(k) falls within the range of 0.66 to 0.88, and for TAXAMOUNT(k), it lies between 16.92 and 30.71. Notably, both coefficients are positive, indicating that as ASSESSVAL(k) or TAXAMOUNT(k) increase by one unit while holding all other variables constant, it leads to a corresponding increase in house prices. This suggests a favorable link between these independent variables and the dependent variable, which is the house price. Moreover, the magnitude of the projected increase in the dependent variable is greater when the independent variables move within this confidence interval range. It's important to emphasize the significance of considering confidence intervals when dealing with sample data and expecting the results to apply to the larger population. This is crucial because coefficient estimates can vary when different samples are used. As an illustrative example, if we assume ASSESSVAL(k) is 660.8 and TAXAMOUNT(k) is 13.3, our model predicts a house price of $974.234 thousand USD based on the given coefficients and input values. By embracing the stepwise backward approach, we acknowledge the inherent variability in our data and ensure the generalizability of our findings to the wider population, making this method the most reliable and comprehensive choice for informed decision-making in the real estate market.

Appendix: A Summary Statistics

PRICE (K)		ASSESSVAL (K)
957.3010753	**Average**	553.4827957
501.2255585	**STD D**	431.5488316
289	**Minimum**	120
2895	**Maximum**	2600
2606	**Range**	2480

Appendix: B Scatter Plot and Regression Output



Summary Output

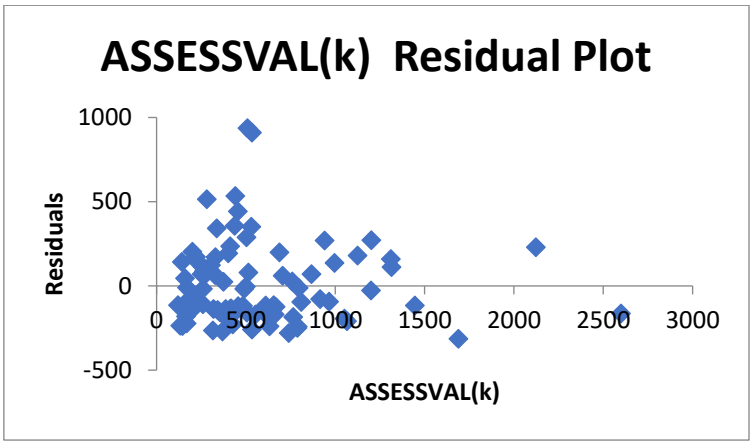
Regression Statistics	
Multiple R	0.883668
R Square	0.78087
Adjusted R Square	0.778462
Standard Error	235.916
Observations	93

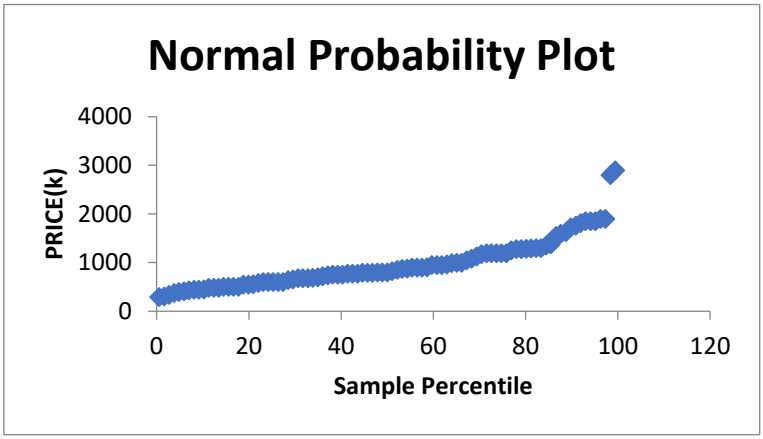
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18048162	18048162	324.2786	9.45E-32
Residual	91	5064728	55656.35		
Total	92	23112890			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	389.2377	39.9196	9.750541	8.48E-16	309.9424	468.5331	309.9424	468.5331
ASSESSVAL(k)	1.026343	0.056995	18.00774	9.45E-32	0.913131	1.139556	0.913131	1.139556

Appendix C: Residual Plots





The residual plots indicate that the conditions of linearity, normality, and equal variance are met, indicating that the basic linear regression model may be used.