



FACULTY OF SCIENCE COMPUTER AND MATHEMATICS

APPLIED REGRESSION ANALYSIS (STA602)

PROJECT TITLE:

EFFECTS OF HOUSING PRICE BY USING SPSS

GROUP: N4CS2414T3

PREPARED BY:

NO	GROUP MEMBER	STUDENT ID
1	ALIA NAJWA BINTI MOHAMAD SAID	2020985547
2	FARZANA BINTI ZULKEFLI	2020957837
3	NURUL AMIRA BINTI ROHI'SHAM	2020996921
4	NURUL HAZLIN BINTI ZULKIFLI	2020995029
5	SITI HAJAR BINTI MUHAMMAD JOHAN	2020963691

PREPARED FOR

MADAM ZAITUL ANNA MELISA BINTI MD YASIN

SUBMISSION DATE

18 JANUARY 2021

TABLE OF CONTENT

TABLE OF CONTENTS	ii
LIST OF TABLES AND FIGURES	iv
CHAPTER 1: INTRODUCTION	1
1.1 BACKGROUND OF STUDY	1
1.2 THEORETICAL FRAMEWORK	2
1.3 PROBLEM STATEMENT	2
1.4 RESEARCH OBJECTIVES	3
1.5 RESEARCH HYPOTHESIS	3
1.6 SIGNIFICANCE AND LIMITATION OF STUDY	4
CHAPTER 2: LITERATURE REVIEW	5
2.1 5 REFERENCES AND JOURNAL	5
CHAPTER 3: METHODOLOGY	7
3.1 INTRODUCTION	7
3.2 STATISTICAL METHOD USED IN THE ANALYSIS	7
3.2.1 MULTIPLE LINEAR REGRESSION	7
3.2.2 MODEL ADEQUACY CHECKING	8
CHAPTER 4: ANALYSIS AND RESULTS	10
4.1 DESCRIPTION OF VARIABLES	10
4.2 REGRESSION MODEL	10
4.2.1 ANALYSIS OF VARIANCE (ANOVA)	10
4.2.2 COEFFICIENT OF DETERMINATION, R^2	11
4.2.3 STANDARD ERROR OF THE ESTIMATE	11
4.2.4 SIGNIFICANT VARIABLES	12
4.3 MODEL ADEQUACY CHECKING	13

4.3.1 LINEARITY	13
4.3.2 NORMALITY	14
4.3.3 HOMOSCEDASTICITY	15
4.3.4 INDEPENDENT	16
4.3.5 MULTICOLLINEARITY	16
4.3.6 OUTLIERS	17
4.4 CORRELATION ANALYSIS	18
4.5 MULTIPLE LINEAR REGRESSION	19
4.5.1 COEFFICIENT DETERMINATION, R^2	19
4.5.2 BEST SET OF VARIABLES USING R^2 CRITERION	19
CHAPTER 5: CONCLUSION OF THE STUDY	21
REFERENCES	22
APPENDICES	23

LIST OF TABLES AND FIGURES

FIGURE 1	THEORETICAL FRAMEWORK	2
FIGURE 2	THE SCATTER PLOT OF REGRESSION STANDARDIZED RESIDUAL AGAINST STANDARDIZED PREDICTED VALUE	13
FIGURE 3	NORMAL P-P PLOT OF REGRESSION STANDARDIZED RESIDUAL DEPENDENT VARIABLE: PRICE (RM)	14
FIGURE 4	SCATTERPLOT OF HOMOSCEDASTICITY	15
FIGURE 5	INFLUENTIAL POINTS	17
TABLE 1	DESCRIPTION OF VARIABLES	10
TABLE 2	THE ANALYSIS OF VARIANCE (ANOVA)	10
TABLE 3	COEFFICIENT OF DETERMINATION, R^2	11
TABLE 4	STANDARD ERROR OF THE ESTIMATE	11
TABLE 5	SIGNIFICANT VARIABLES	12
TABLE 6	INDEPENDENT	16
TABLE 7	MULTICOLLINEARITY	16
TABLE 8	CORRELATION ANALYSIS	18
TABLE 9	COEFFICIENT OF DETERMINATION, R^2	19
TABLE 10	BEST SET OF VARIABLES USING R^2 CRITERION	19
TABLE 11	COEFFICIENTS	20

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF STUDY

House is building that works as a living space or home that come in different sorts such as condominium, flat, bungalow and numerous more. Malaysia is one of the major Asian tiger economies, having delighted in solid economic growth surpassing 8% in each year of 1989-1997 (D'Arcy and Keogh, 1999). Malaysia was the primary Asian nation to create recorded property trust as an effective real estate investment. The fast urbanization that has taken place in Malaysia has made the property sector a key portion within the development of Malaysia's financial success. The government recognizes that housing may be a central require for everyone. It too shapes an imperative portion of the urban economy. Housing developments in Malaysia are carried out in terms of low, medium, and high cost houses on both sides, the open division, and the private division (Asiah, 1999).

Over the last few years, quick financial improvement has come about in an expanding request for residential among urban ranges in Malaysia. Checking on the housing cost in Malaysia, the prices have acknowledged significantly whether in major cities or littler town and depending on specific area. Last decade, property market in Malaysia has experienced a noteworthy cost development all through Malaysia, including higher rates. Most individuals are pondering such a tall yearly increment in house costs is completely out of adjust with annual salary increments within the common populace. Progressive governments and approach examiners have recognized the need of satisfactory and reasonable housing as one of the basic issues confronting the nation. House and land costs are spiraling, and indeed middle-class Malaysian are confronting troubles to own a home. Based on the economic theory, house cost developments are inherent within the socioeconomics and financial matters, such as population, GDP, housing finance, RGPT and cost of construction. Other than that, the location, size, and property type also can be the factors that influence the housing price.

1.2 THEORETICAL FRAMEWORK

The size of the house, the numbers of rooms, bathrooms and carparks, the property type and the furnishing of the house influence the housing price. It means that the nearer the location of residential to the town, the higher the price of the house. The bigger the size of the house, the higher the price of the house. The higher the numbers of rooms, bathrooms and carparks will increase the house price. The land area house price is higher compare to built-up house. The price also depending on the furnishing of the house, if the house is unfurnished the price will be decrease.

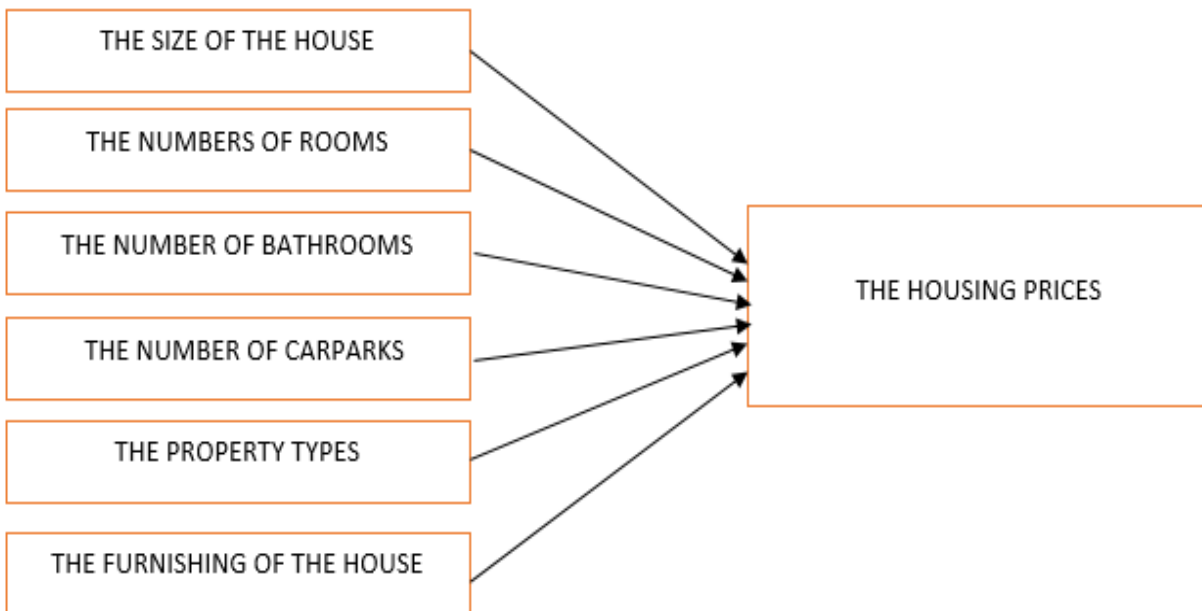


Figure 1 Theoretical Framework

1.3 PROBLEM STATEMENT

In 21st century Australia, Generation Y experiencing difficulties when it comes to buying their first house. Sydney is one of the most expensive cities in the world and because of the high of cost living. The same thing is happening in Malaysia, people face difficulties to buy their first house because the house purchase rate is very expensive and house prices are rising from year to year. Kuala Lumpur is one of the most expensive

cities to buy house because Kuala Lumpur is the capital of Malaysia. Everyone nowadays experiencing difficulties to buy their house because the younger generation today is struggling to find job and find the cost of living . Paired with the younger generation's low to medium starting income, this creates unfair opportunities for them in a housing market dominated by older, richer generations and property investors. In Malaysia, the rate of increase in house price is even higher than the rate of salary increase.

As we all know, house prices are set according to the type of house itself. There is a lot of type house such as condominium, apartment , semi-D and a lot more. Before 2009, house prices in Malaysia were still affordable but since 2009, house price have started to increase by 8% a year. One of the reasons why house prices have risen sharply since 2009 is because of the easy access to housing loans from banks. Another reason why house is very expensive nowadays because the refusal of developers to lower the price of expensive house they build because they are confident, they can sell at this expensive price to others. Many houses are expensive to sell and sold out. But at the same time, it also encourages many people to buy a home beyond their financial means.

1.4 RESEARCH OBJECTIVES

This study has listed research objectives related to the topic:

- i. To identify the relationship of the housing price between the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur.
- ii. To determine which independent variable that most affected the housing price between the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur.
- iii. To determine the best set of variables that should be used to predict the housing price between independent variables.

1.5 RESEARCH HYPOTHESIS

There is some research hypothesis stated for this study:

- i. There is a significant relationship of the housing price between the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur.
- ii. There is most affected the housing price between the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur.
- iii. There is the best set of variables that should be used to predict the housing price between independent variables.

1.6 SIGNIFICANCE AND LIMITATION OF STUDY

Generally, with the estimation of 32.7 million current populations of Malaysia in 2020, there is a lot of demand on property from the Malaysian citizens. In 2012, in the list of “The World’s Hottest Real Estate Markets” by Knight Frank, Malaysia’s housing market was ranked at 9th (Rajeshni, 2012). From this study, it will help the investors to invest properly and carefully to the property in Malaysia since there is many factors that will affect the property’s prices. Other than that, from this study we can predict the prices of any kind of property based on the factors such as strategic locations, rooms, car parking and many more. According to Karsten (2007) and Tu and Goldfinch (1996), the investors will find a property which is easiest to reach the places that they want to go. Seeing that the crimes in Malaysia is dangerous, the investors gladly to invest more if the neighborhood of the property is safer and secured properly.

For this study, the main limitation is that it is only focused on property in Kuala Lumpur, Malaysia so it is difficult to predict what the investors want in another states since the expenses in Kuala Lumpur is a bit pricey than other. Furthermore, the investment in the property might be bias since it is also based on the preference of the investors. In addition, by using the secondary data for this study the results may not be truly accurate since there might be an error in the data. According to Husain, Rahman & Ibrahim (2011), with lack of government support and policy of housing so that it carries on with the speculations that been made. Besides that, 36% of investors choose a new property rather than invest in secondary market which is about 17% since the developers always offer a new property (PropertyGuru, 2016).

CHAPTER 2

LITERATURE REVIEW

From the article entitled “Listed Property Trusts in Malaysia” by Graeme Newell, Ting Kien Hwa and Peter Acheampong said that Malaysia has now overcome and recovered from the Asian economic crisis of 1997-1998, the listed property trust sector has not kept pace with the improvements in the overall Malaysian economy over 1999-2000. Malaysia is one of the major Asian tiger economies and was the first Asian country which is listed as an effective property trust development as well as an effective indirect property investment vehicle in 1989. In September 1989, the first listed property trust in Malaysia was established (Arab Malaysian First Property Trust). Some regulatory and operational differences exist for listed property trusts in Malaysia. These local regulatory and operational differences highlight the structural factors that have contributed to the limited success of property trusts in Malaysia. These changes have been well received by property trust managers, enabling listed property trusts to acquire real estate more easily. Strategy are articulated that would contribute to the ongoing development of the property trust sector for Malaysian Listed Property Trusts section.

Based on a journal entitled “House Price and Affordability in Housing in Malaysia” written by Zainal Abidin Hashim, research shows that income can influence the house price movement. As the proof, this research makes the comparison of the housing price in two states which is Selangor and Kelantan related to the affordability issues as to income from 1995 until 2006. By that, the house price naturally higher in the state of Selangor on the average 35 percent to 53 percent higher compared to Kelantan. However, the total units of house sold also higher in the terms of the high in income. Median household income for Selangor, the most develop states in Malaysia, demonstrated an average annual growth rate of 5.1 percent in 1998. As for Kelantan, the least developed state, experienced an average annual growth rate of 4.6 percent but with a much lower median income. Other than that, an overall view of the property market from time to time in terms of type, location and features gives a biased and misleading price evaluation.

In the journal entitled “Factors Affecting the Purchase Decision of Investors in the Residential Property”, the researcher using a questionnaire for the method of data collection that consists of 200 questions but the response rate only 52.5%. For the sample of the data, the researcher takes the respondents that has make more than one transaction on property for the past three years from Greater Kuala Lumpur (GKL), Malaysia. According to Tan & Phang (2014), GKL is one of the nation’s gross national product which the area coverage mostly in Kuala Lumpur

and Selangor with 37%. In Malaysia, it is difficult to new investors or investors who had a family to own a property since the property price in Malaysia is highly pricey (Chin, 2013). Based on the calculation that had been made in this journal, the p-value for location, finance and structure is 0.026, 0.011 and 0.024 respectively with the 5% of significant. As a result, these three factors which is location of the property, financing and structure of the property affect the investors for making the decisions to invest on the property.

Furthermore, in the journal entitled, "An Analysis of the Factors Affecting House Prices in Malaysia", the researchers explain on house price in Malaysia which is how Malaysia remains the issues and the problem of housing price makes a significant concern to the government. In this study, the researchers explain these issues using Multiple Regression Method that covers within 23 years from 1988 to 2010. The median house price in Malaysia are at 4.4 times the median annual income in the year 2014 which places Malaysian housing market at an "unaffordable" level (Khazanah Research Institute, 2015). On top of that, this makes the issues regarding house price in Malaysia that can attract attention with various reasoning attached to the volatile pricing also because of various factors such as affecting housing price, strong income, including interest rates also growth the credit in country. In this research, we point that the government can look more benefits to projects including providing better public transportation network covering a wider area. It can conclude that from this research, 95.77% adjusted R-squared resulted in study, federal government play big roles to control house prices from escalating also the government should including state.

The paper of Factors Affecting the Price of Housing in Malaysia (Ong, 2013) analyses observationally whether the increasing trend in the Malaysia housing price is related to changes within the gross domestic product (GDP), population, inflations rate, costs of development, interest rate and real property gains tax (RPGT). The investigate conducted by Tse (1999) states that the fluctuation within the housing price has significantly impact on the financial conditions of the population and society. Besides, the demand for housing is increasing in the market. So, the housing price is expected to rise due to the imbalance between buyers and dealers. Subsequently, when there are more buyers than dealers, the housing price will increase. This could cost a self-fulfilling speculative price bubble (Levin and Wright, 1997). In Malaysia, lack of reasonable choices to own property forces many Malaysia into the rental market and informal settlement, which expanding due to population development and urbanization. In the real economy, there are lot more factors that affect housing prices, but we cannot deny the facts that one of the significant factors might be population growth.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

In this chapter are explain the procedure that will conduct in the data analysis. Every method that can use be determine in this chapter in generally, details and step of the procedure. Each method that used stated the reason to make sure the data analysis is appropriate. There are several methods that will conduct in this study for example model adequacy checking and multiple linear regression. To find the relationship between each independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur we use correlation analysis while to check whether the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house most affected the housing prices in Kuala Lumpur we determine the analysis with use the multiple linear regression also we use in this study to determine the best set of variables.

3.2 STATISTICAL METHOD USED IN THE ANALYSIS

3.2.1 Multiple Linear Regression

Multiple linear regression also known as multiple regression is an expansion of simple linear regression. It is utilized when we need to predict the value of a variable based on the value of two or more other factors. The variable we need to predict is called the dependent variable while the variables we are utilizing to predict the value of dependent variable are called independent variable. Multi linear regression is used broadly in econometrics and money related induction.

Linear regression can as it were be utilized when one has two continuous variables which is independent variables and dependent variables. The independent variable in the parameter that is used to calculate the dependent variable or result. A multiple regression model expands to a few explanatory variables. We use multiple linear regression analysis where it considered a regression model in which the dependent variable, Y is affected by more than one independent variable(s). For this study, the Housing Price (Y) may be influenced by the number of rooms (X_1), the size of the house (X_2), the number of car

parks (X_3), the property types (X_4), the number of bathrooms (X_5) and the furnishing of the House (X_6). To use this analysis, five assumptions need to be fulfilled.

- Linearity assumption where the relationships between the independent variables (X_1, X_2, X_3, X_4, X_5 , and X_6) and therefore the dependent variable (Y) should be linear.
- Normality assumption where the errors should be normally distributed.
- Homoscedasticity assumption where the error variance should be constant.
- Independence assumption where the errors related to one observation are not correlated with the errors of the other observation.
- Model specification where the model should be properly specified (including all relevant variables and excluding irrelevant variables).

As mentioned above, multiple linear regression is an expansion of linear regression, but rather than relating one dependent result variable Y to at least one independent X , one tries to clarify the result value Y because the weighted entirely of impacts from multiple independent variables X_1, X_2 and then forth. In equation form:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_k X_{i,k} + \epsilon_i$$

where,

- β_0 is that the intercept of the line on the y-axis
- β_1, β_2 , and β_k are the slope of the relations between y and X_1, X_2 , and X_i
- k is that the number of predictor variables, which means we have $k+1$ regression parameter (the β coefficients)
- the subscript i refers to the i^{th} individual or unit within the population (i simply denotes which x-variable it is)
- ϵ_i have a normal distribution with mean 0 and constant variance (random error term)

3.2.2 MODEL ADEQUACY CHECKING

The fitting of the linear regression model, estimation of parameters testing of hypothesis properties of the estimator is based on the following major assumptions:

- i. The relationship between the study variable and explanatory variables is linear, at least approximately
- ii. The error term has zero mean

- iii. The error term has a constant variance
- iv. The errors are uncorrelated
- v. The errors are normally distributed

The validity of these assumptions is needed for the results to be meaningful. If these assumptions are violated, the result can be incorrect and may have serious consequences. If these departures are small, the result may not be changed significantly. But if the deviations are large, the model obtained may become unstable in the sense that a different sample could lead to an entirely different model with opposite conclusions. So, such underlying assumptions must be verified before attempting to regression modeling. To check the violation of regression assumptions are based on the study of the model residual which is:

- a. Linear relationship between the dependent and independent variable.
To determine if there is a linear relationship between the dependent and independent variables, we use the scatter plot matrix or the partial plots.
- b. Normality of the error terms.
Are the error terms normally distributed? We can examine this by using the following plots; the P-Plot of the residuals, boxplot of residual, histogram or residual, plot of residual against the independent variable and plot of residuals against the predicted values.
- c. Homoscedasticity
To examine the error variance constant by using the plot of residuals against the predicted (fitted) values.
- d. Multicollinearity
To examine the independent highly correlated we use the correlation matrix which shows the correlation coefficients of all pairs of data. Any pair of independent variables with a correlation coefficients value greater than 0.8 is said to be highly correlated.
- e. Presence of outliers
Outliers are extreme values which can affects the coefficients of the regression line. This will also be done in more detail later.

CHAPTER 4

ANALYSIS AND RESULTS

4.1 DESCRIPTION OF VARIABLES

The Property Listings in Kuala Lumpur dataset is obtained from the website of Kaggle.com. The secondary data include the housing price (Y), the numbers of rooms (X_1), the size of the house (X_2), carparks (X_3), the property type (X_4), bathrooms (X_5), and the furnishing of the house (X_6) influence the housing price.

Variables Name	Description
Rooms, X_1	The numbers of rooms (integer)
Size, X_2	The size of the house (unit: sq. ft)
Car parks, X_3	The number of carparks (integer)
Property type, X_4	The property type
Bathrooms, X_5	The number of bathrooms (integer)
Furnishing, X_6	The furnishing of the house
Price, Y	The housing price in Kuala Lumpur (in '000)

Table 1 Description of Variables

4.2 REGRESSION MODEL

4.2.1 Analysis of Variance (ANOVA)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.569E+14	6	5.949E+13	75.804	.000 ^b
	Residual	1.122E+14	143	7.848E+11		
	Total	4.692E+14	149			
a. Dependent Variable: Price (RM)						
b. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms						

Table 2 The Analysis of Variance (ANOVA)

The test statistic of the ANOVA table for regression shows that there is a significant on the result of p-value obtained. Since the p-value = 0.000 which is less than $\alpha=0.05$ with F-test value of 75.804, therefore we reject H_0 . Reject H_0 implies that the model is a good fit model.

4.2.2 Coefficient of Determination, R^2

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.872 ^a	.761	.751	885890.570	.761	75.804	6	143	.000	2.128

a. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms
b. Dependent Variable: Price (RM)

Table 3 Coefficient of Determination, R^2

75.1% of the total variation in housing price is explained by the number of rooms, the number of bathrooms, the number of car park, property type, the size of property and the furnishing. While, the balance, 24.9% is explained by the other factors.

4.2.3 Standard Error of the Estimate

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.872 ^a	.761	.751	885890.570	.761	75.804	6	143	.000	2.128

a. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms
b. Dependent Variable: Price (RM)

Table 4 Standard Error of the Estimate

The standard error of estimate with value 885890.570 is greater than 10% of the mean of the dependent variable ($\bar{y} = 1886142.609$) which indicates that the model cannot be considered as good fit model.

4.2.4 Significant Variables

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1098417.986	421592.199		2.605	.010		
	Rooms	-344687.803	101926.227	-.300	-3.382	.001	.213	4.699
	Bathrooms	223929.985	106616.359	.187	2.100	.037	.210	4.763
	Car Parks	-310834.081	122261.409	-.190	-2.542	.012	.298	3.355
	Property Type	12190.956	248152.850	.003	.049	.961	.632	1.582
	Size	1111.767	76.575	1.041	14.519	.000	.325	3.073
	Furnishing	-78760.727	159927.275	-.021	-.492	.623	.896	1.116

a. Dependent Variable: Price (RM)

Table 5 Significant Variables

The result shows that the two independent variables (numbers of rooms (X_1), bathrooms (X_2)) are significant respectively with p-value < 0.001 at $\alpha=0.05$. The other variables (carparks (X_3), property type (X_4), size of the house (X_5), and the furnishing of the house (X_6)) are not significant respectively since p-value > 0.001 at $\alpha=0.05$.

Initial Model Equation:

$$\hat{Y} = 1098417.986 - 344687.803X_1 + 223929.985X_2 - 310834.081X_3 + 12190.956X_4 + 1111.767X_5 - 78760.727X_6$$

4.3 MODEL ADEQUACY CHECKING

4.3.1 Linearity

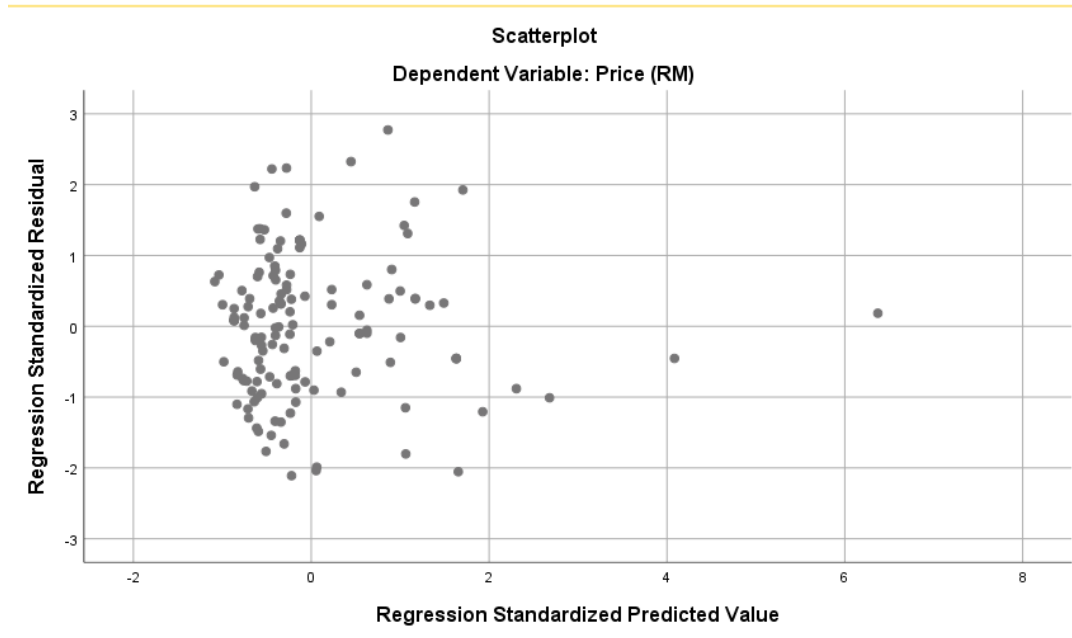


Figure 2 The Scatter plot of Regression Standardized Residual against Standardized Predicted Value

Figure 2 shows the scatterplot between regression standardized predicted value versus regression standardized residual. Based on this scatterplot, there is not any curvilinear pattern which can be indicated that the data is linear. Hence, there is linear relationship between housing price (Y) and the others independent variables; the numbers of rooms (X_1), the size of the house (X_2), carparks (X_3), the property type (X_4), bathrooms (X_5), and the furnishing of the house (X_6) in Kuala Lumpur

4.3.2 Normality

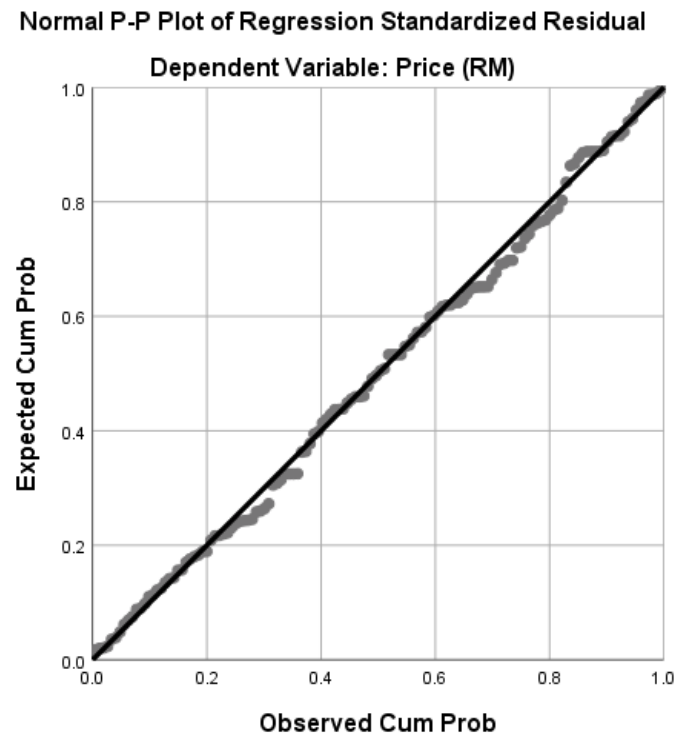


Figure 3 Normal P-P Plot of Regression Standardized Residual Dependent Variable: Price (RM)

Based on figure 3 above, the P-P plot of the residuals shown a cumulative probability plot of standardized residuals. That's means the graph shows that the errors are normally distributed since most of the points lies approximately near on 45-degree line or straight line. The closer it lies to the diagonal line, the closer to normal the residuals are distributed.

4.3.3 Homoscedasticity

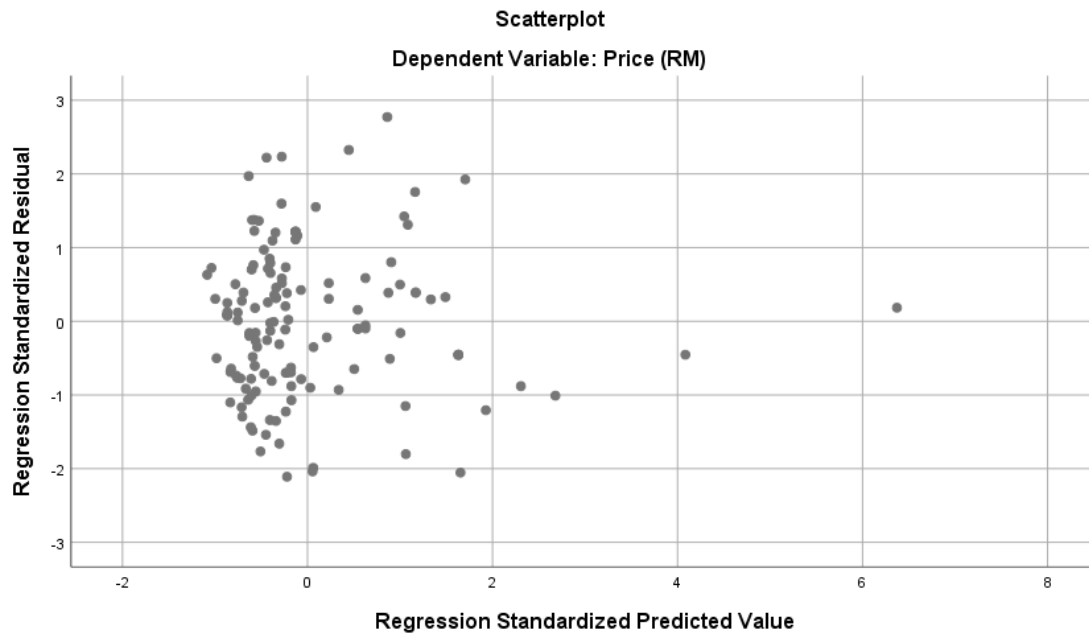


Figure 4 Scatterplot of Homoscedasticity

Figure 4 shows the scatter plot of regression standardized predicted value versus regression standardized residual. It shows that most of the points are lies between 3 and -3. Therefore, the residuals have constant variance and the homoscedasticity assumption is satisfied.

4.3.4 Independent

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.931 ^a	.867	.865	467141.229	.867	438.230	2	135	.000	2.008
a. Predictors: (Constant), Size, Rooms										
b. Dependent Variable: PriceRM										

Table 6 Independent

The Durbin-Watson value is 2.008 which is lies between 1.5 and 2.5 indicates that there is no serial correlation exist. So that, the error terms are independent.

4.3.5 Multicollinearity

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	968015.606	101844.207		9.505	.000		
	Rooms	-351338.379	35588.216	-.423	-9.872	.000	.539	1.855
	Size	1090.536	40.101	1.165	27.195	.000	.539	1.855

a. Dependent Variable: Price (RM)

Table 7 Multicollinearity

All collinearity tolerance values are greater than 0.1, there is no multicollinearity problem. Statistics VIF values for all variables are less than 10, this can be said that there is no multicollinearity problem exist. The variables rooms and size are independent among themselves.

4.3.6 Outliers

i. Extreme of Y observations

There is observation that have value greater than 3 and less than -3. It is an observation 4, 9, 23, 32, 34, 64, 75. 94, 95, 112, 141, 130,

ii. Extreme of X observations

Leverage formula $\frac{2p}{n} = \frac{2(6)}{150} = 0.08$. Based on this value, if there is value greater than 0.08, it is indicated as extreme of X. In this research, the observation 4, 8, 25, 64, 86, 90, 95, 99, 112, 117, 134, are 137 are said to be the extreme of X observations.

iii. Influential points.

	SRE_4	COO_4	LEV_4	SRE_5	COO_5	LEV_5	SRE_6	COO_6	LEV_6	var	var
1	-.47915	.00090	.01978	-.43219	.00074	.02010	-.37039	.00056	.02050		
2	-2.04714	.02375	.03117	-2.10873	.02521	.03113	-2.28267	.02971	.03129		
3	.06857	.00003	.03692	.09358	.00006	.03692	.17118	.00019	.03718		
4	-2.15732	.08123	.10189	-2.23660	.08732	.10184	-2.31371	.09355	.10190		
5	-.53312	.00235	.04771	-.59170	.00290	.04780	-.64329	.00343	.04777		
6	.85743	.00929	.07428	.88031	.00979	.07423	.84461	.00906	.07458		
7	-.99685	.00379	.01900	-.99950	.00382	.01904	-1.03345	.00409	.01902		
8	-.27584	.00105	.08114	-.33091	.00152	.08126	-.28697	.00114	.08143		
9	2.62222	.17568	.14472	2.75647	.19433	.14481	2.93696	.22069	.14480		
10	.85863	.00336	.02392	.92494	.00391	.02399	1.01103	.00469	.02402		
11	-.26351	.00072	.06049	-.18693	.00036	.06108	-.11342	.00014	.06144		
12	.39239	.00071	.02421	.44210	.00090	.02428	.50137	.00116	.02430		
13	-.11875	.00009	.03361	-.06024	.00002	.03392	.01514	.00000	.03423		
14	-1.05357	.00726	.03681	-1.04874	.00722	.03692	-1.01259	.00682	.03739		
15	-1.36291	.01063	.03152	-1.34983	.01052	.03181	-1.35104	.01063	.03208		
16	-.37997	.00080	.03012	-.39699	.00087	.03007	-.35699	.00071	.03025		
17	-1.51728	.01767	.04399	-1.49714	.01738	.04443	-1.57809	.01931	.04438		
18	.66329	.00382	.05028	.62722	.00343	.05054	.70072	.00429	.05058		
19	-1.60700	.01490	.03182	-1.60229	.01494	.03211	-1.61585	.01533	.03238		
20	1.73602	.01197	.02005	1.88574	.01449	.02070	2.07487	.01784	.02109		
21	-.71779	.00176	.01637	-.75548	.00195	.01634	-.76436	.00200	.01635		
22	.04709	.00003	.07626	.06059	.00005	.07622	.00669	.00000	.07636		
23	1.01150	.00752	.04196	1.05502	.00819	.04192	1.03955	.00800	.04218		

Figure 5 Influential points

Cook's Distance D_i , measures how much the residuals of all cases would change if a case were excluded from the calculation of the regression coefficients. A large Cook's D indicates that excluding a case from computation of the regression statistics, changes the coefficients substantially. Hence, in this study, the observation 9 is the highest of Cook's D which is 0.22069 and should be remove from the data set.

4.4 CORRELATION ANALYSIS

Correlations				
		Price (RM)	Rooms	Size
Pearson Correlation	Price (RM)	1.000	.368	.878
	Rooms	.368	1.000	.679
	Size	.878	.679	1.000
Sig. (1-tailed)	Price (RM)	.	.000	.000
	Rooms	.000	.	.000
	Size	.000	.000	.
N	Price (RM)	138	138	138
	Rooms	138	138	138
	Size	138	138	138

Table 8 Correlation analysis

The result of Pearson Correlation shows that there is a moderate positive linear relationship between numbers of rooms (X_1) and housing price (Y) with correlation coefficient of 0.368. While, the size of the house (X_5) and housing price (Y) shows that there is strong positive linear relationship with R-values 0.878. thus, it can be seen from the results that size of the house (X_5) is the most correlated factor to house price (Y).

4.5 MULTIPLE LINEAR REGRESSION

4.5.1 Coefficient determination, R^2

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	.931 ^a	.867	.865	467141.229	.867	438.230	2	135	.000	2.008
a. Predictors: (Constant), Size, Rooms										
b. Dependent Variable: PriceRM										

Table 9 Coefficient of Determination, R^2

From the table 9 above, the result for coefficient of determination, R^2 , 86.7% of total variation of housing prices(Y) are explained by the sizes of house(X_2) and number of rooms(X_1). The other 13.3% are explained by other factors. Hence, the independent variables, sizes of house(X_2) and number of rooms(X_1) only contribute 86.7% to the dependent variable, housing prices(Y).

4.5.2 BEST SET OF VARIABLES USING R^2 CRITERION

MODEL	VARIABLE (S)	R^2
1	X_1	0.135
2	X_2	0.770
3	X_1, X_2	0.867

Table 10 Best Set of Variables using R^2 Criterion

Y : Housing prices

X_1 : Number of rooms

X_2 : Sizes of house

For the R^2 value in the table above, the best set of variables that should be used to predict the housing price is X_1 and X_2 since the value of R^2 is the highest which is 0.867 compared to the other set of variables X_1 only and X_2 only. This means that it has the greatest contribution in predicting the house price. Therefore, 86.7% of total variation of housing prices(Y) are explained by sizes of house(X_2) and number of rooms(X_1).

Coefficients ^a								
Model	Unstandardized Coefficients			Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error		Beta			Tolerance	VIF
1	(Constant)	968015.606	101844.207		9.505	.000		
	Rooms	-351338.379	35588.216	-.423	-9.872	.000	.539	1.855
	Size	1090.536	40.101	1.165	27.195	.000	.539	1.855

a. Dependent Variable: Price (RM)

Table 11 Coefficients

Final Model Equation:

$$\hat{Y} = 968015.606 - 351338.379X_1 + 1090.536X_2$$

CHAPTER 5

CONCLUSION OF THE STUDY

Some of the study, testing of some empirical data to prove the theory. In the end of the study, it can be learned that some of variables tested in this study to achieve the objectives which is to determine which independent variable that most affected the housing prices in Kuala Lumpur also to determine the best set of variables that should be used to predict the housing price. Hence, this study will identify few factors that affected the housing price. Furthermore, to analyse method in this analysis using multiple linear regression to achieve the objectives of the research. Before furthering into the multiple linear regression analysis, the data must be fulfilling the assumption of the method. According to the results obtained, all assumptions of multiple linear regression has been achieved which may proceed to the method of analysis to determine either significantly or not.

Therefore, first objectives which is to identify the relationship of the housing price between the independent variables; the number of rooms, the size of the house, the number of car park, the property type, the number of bathrooms and the furnishing of the house in Kuala Lumpur. The Pearson Correlation from this analysis has been carried out where numbers of rooms (X_1) and housing price (Y) with correlation coefficient of 0.368. While, the size of the house (X_2) and housing price (Y) shows that there is strong positive linear relationship with R-values 0.878. Thus, it can be seen from the results that size of the house (X_2) is the most correlated factor to house price (Y). Next, for the second objectives to determine which independent variable that most affected the housing price in Kuala Lumpur. From the analysis it can be conclude that the number of rooms dan the size of house is significant to dependent variable which is the housing price.

As a conclusion, to determine the best set of variables that should be used to predict the housing price, we can conclude that from the output, the best set of variables to predict the housing price is X_1 and X_2 since the value of R^2 is the highest which is 0.867 compared to the other set of variables X_1 only and X_2 only. This means that it has the greatest contribution in predicting the house price. Therefore, 86.7% of total variation of housing prices(Y) are explained by sizes of house(X_2) and number of rooms(X_1). However, Final Model Equation: $\hat{Y} = 968015.606 - 351338.379X_1 + 1090.536X_2$

REFERENCES

- Bejarano, pedro gerardo prieto. (2013). No Title消費者型官能評. 日本畜産学会報, 84(2001), 487–492. Retrieved from <http://ir.obihiro.ac.jp/dspace/handle/10322/3933>
- Islam, S. Z., Mariun, N., Hizam, H., Othman, M. L., Radzi, M. A. M., Hanif, M., & Abidin, I. Z. (2012). Communication for Distributed Renewable Generations (DRGs): A review on the penetration to Smart Grids (SGs). *PECon 2012 - 2012 IEEE International Conference on Power and Energy*, 78, 870–875. <https://doi.org/10.1109/PECon.2012.6450339>
- Newell, G. (2009). Listed Property Trusts in Australia. *Global Trends in Real Estate Finance*, (September), 46–63. <https://doi.org/10.1002/9781444315301.ch4>
- Ong, T. S. (2013). Factors Affecting the Price of Housing in Malaysia. *Journal of Emerging Issues in Economics, Finance and Banking*.
- Sean, S. L., & Hong, T. T. (2014). Factors Affecting the Purchase Decision of Investors in the Residential Property Market in Malaysia. *Journal of Surveying, Construction & Property*, 5(2), 1–13. <https://doi.org/10.22452/jscp.vol5no2.4>

APPENDICES

Analysis of Variance (ANOVA)

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.569E+14	6	5.949E+13	75.804	.000 ^b
	Residual	1.122E+14	143	7.848E+11		
	Total	4.692E+14	149			

a. Dependent Variable: Price (RM)

b. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms

Coefficient of Determination, R^2

Model Summary ^b											
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson	
1	.872 ^a	.761	.751	885890.570	.761	75.804	6	143	.000	2.128	

a. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms

b. Dependent Variable: Price (RM)

Standard Error of the Estimate

Model Summary ^b											
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson	
1	.872 ^a	.761	.751	885890.570	.761	75.804	6	143	.000	2.128	

a. Predictors: (Constant), Furnishing, Bathrooms, Property Type, Car Parks, Size, Rooms

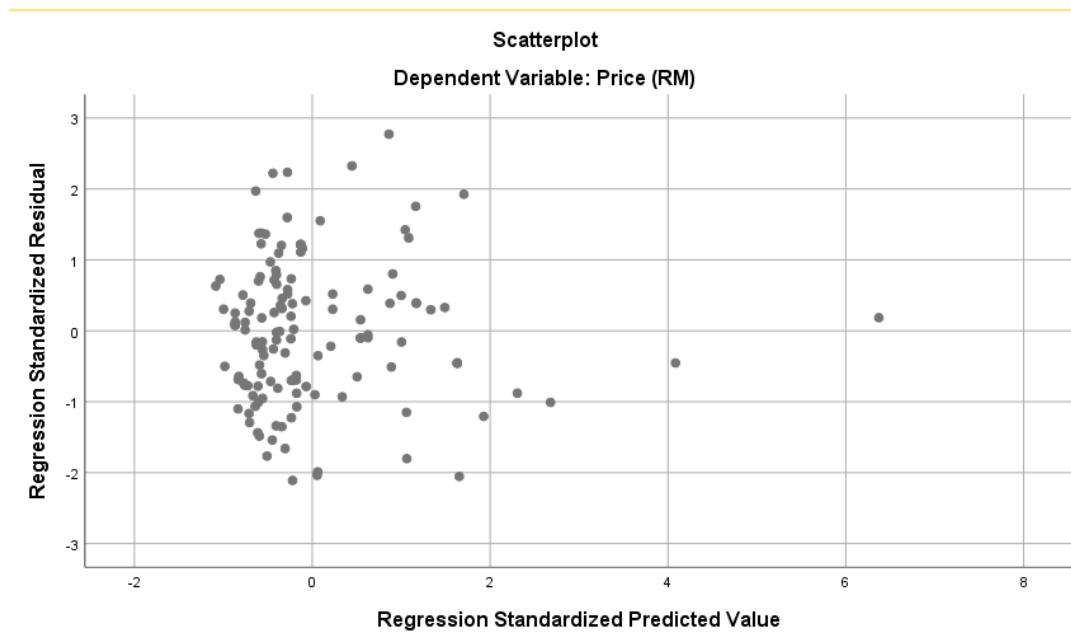
b. Dependent Variable: Price (RM)

Significant Variables

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1098417.986	421592.199		2.605	.010		
	Rooms	-344687.803	101926.227	-.300	-3.382	.001	.213	4.699
	Bathrooms	223929.985	106616.359	.187	2.100	.037	.210	4.763
	Car Parks	-310834.081	122261.409	-.190	-2.542	.012	.298	3.355
	Property Type	12190.956	248152.850	.003	.049	.961	.632	1.582
	Size	1111.767	76.575	1.041	14.519	.000	.325	3.073
	Furnishing	-78760.727	159927.275	-.021	-.492	.623	.896	1.116

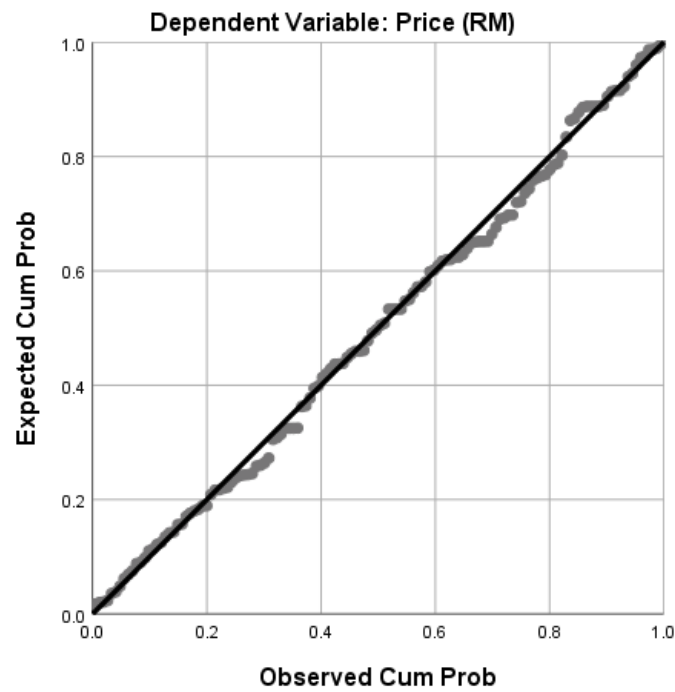
a. Dependent Variable: Price (RM)

Linearity

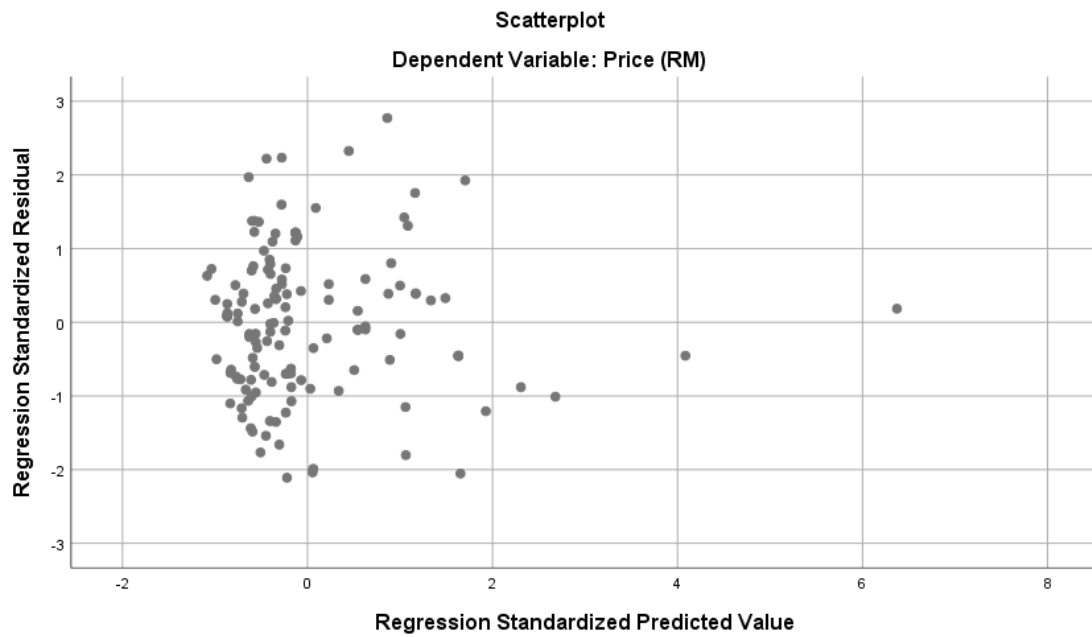


Normality

Normal P-P Plot of Regression Standardized Residual



Homoscedasticity



Independent

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	.931 ^a	.867	.865	467141.229	.867	438.230	2	135	.000	2.008

a. Predictors: (Constant), Size, Rooms
b. Dependent Variable: PriceRM

Multicollinearity

Coefficients ^a								
		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
Model		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	968015.606	101844.207		9.505	.000		
	Rooms	-351338.379	35588.216	-.423	-9.872	.000	.539	1.855
	Size	1090.536	40.101	1.165	27.195	.000	.539	1.855

a. Dependent Variable: Price (RM)

Correlation Analysis

Correlations				
		Price (RM)	Rooms	Size
Pearson Correlation	Price (RM)	1.000	.368	.878
	Rooms	.368	1.000	.679
	Size	.878	.679	1.000
Sig. (1-tailed)	Price (RM)	.	.000	.000
	Rooms	.000	.	.000
	Size	.000	.000	.
N	Price (RM)	138	138	138
	Rooms	138	138	138
	Size	138	138	138

Coefficient determination, R^2

Model Summary ^b										
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change	Durbin-Watson
						F Change	df1	df2		
1	.931 ^a	.867	.865	467141.229	.867	438.230	2	135	.000	2.008

a. Predictors: (Constant), Size, Rooms
b. Dependent Variable: PriceRM

Coefficient

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	968015.606	101844.207		9.505	.000		
	Rooms	-351338.379	35588.216	-.423	-9.872	.000	.539	1.855
	Size	1090.536	40.101	1.165	27.195	.000	.539	1.855

a. Dependent Variable: Price (RM)