Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 9, 2022, page 222-232

Multplie Linear Regression on Population Growth Rate in Malaysia

Muhammad Izzatur Fitri Mohd Fuad, Adina Najwa Kamarudin*
Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia
*Corresponding author: adina.najwa@utm.my

Abstract
Malaysia has seen a substantial urban population increase for the past 90 years. Unstable population growth rate whether too high or too low may affect human needs, basic services, and infrastructure. This study aimed to explore the trends of population growth in Malaysia in past few years and to predict population growth rate in Malaysia by using Multiple Linear Regression (MLR). This study investigates the relationship between population growth rate and the potential independent variables which are crude birth rate, crude death rate, neonatal mortality rate, and maternal mortality rate. The data are obtained from DOSM which is from 1978 to 2018. All statistical analysis in this project were performed by using Microsoft Excel and SPSS software. The study concludes that the crude birth and death rates are important variables to consider when developing a population growth regression model. Crude birth rate has positive relationship with population growth while crude death rate has negative relationship with population growth. The MLR model for this research found to be $\hat{y}$= 2.961+0.079x_1+(-0.541)x_2.
Keywords: Multiple Linear Regression; Population; Crude birth rate; Crude death rate

1. Introduction

Population is defined as a group of individuals of the same species living in the same region and interbreeding [1]. The global average population density is 25 persons per square kilometre and the world's top five most populous countries are China (1.44 billion), India (1.39 billion), United States (333 million), Indonesia (276 million), and Brazil (214 million) [2]. The global human population is growing at a rate of roughly 83 million people per year, or 1.1 percent per year [3]. From 1 billion in 1800 to 7.9 billion in 2020, the world's population has increased dramatically [4]. Outside of the UN, however, increasing number of proposing prediction models on the development of the proposed model has accounted the negative influences of the population increment, in which population would peak before year 2100 [5].

Malaysia is ranked 45th in terms of population among countries [4]. Average annual population growth rate in Malaysia is gradually increasing from 1989 to 2019 [6]. As of 2020, Malaysia's population is expected to expand at a rate of 1.30 percent. This is much lower than the 2.51 percent rate recorded in 2000. This slowing of population growth is projected to continue in the next decades, decreasing population growth until it reaches draw up and then declines [7]. The government's population policy is to achieve a stabilised population of 70 million by 2070, which means that fertility must decrease from 4 to 2 children per woman [9].

Every day, 227,000 people are added to the world, and the United Nations projects that by the end of the century, the human population will have surpassed 11 billion [8]. Demands for water, land, trees, and fossil fuels are increasing as the world's population expands. Therefore, population projection is significant because it assists people, such as the government and scholars, in making future decisions. For example, the population projection result can be used by people that want to estimate the essential human needs.

This research aims to (1) explore the trends of population growth in Malaysia in past few years and (2) predict population growth rate in Malaysia. The data of population in Malaysia from year 1978 to year 2018 are obtained from Department of Statistic Malaysia (DOSM). This study focused on modelling a prediction model population growth rate in Malaysia by using multiple linear regression. This study is expected to be useful for community to solving their global optimization problem especially in Malaysia to stabilize resources and upgrade its facilities.

2. Literature Review

In China, the population size is modelled together with time by using regression and fitting techniques [10]. Qu [10] used simple exponential growth model at first phase then establish the mathematical model by adapting the multiple linear regression to the total number of individuals and potential factors which is fertility rate, mortality rate, urbanization, birth rate, proportion of aged population and sex ratio of population. The main advantage of this proposed model is, it is feasible to predict the future trend of population growth. Then Qu [10] used logistic regression model to connect proportion of ageing population and fertility rate in screening the factors, Qu [10] applied the stepwise regression analysis method [10]. This research creates a prediction model of China's population increase using the stepwise regression approach. This model reflects the characteristics of China's population development and forecasts future population changes, mostly in accordance with its hypothetic circumstances. This regression model may be applied in Malaysia's population growth rate prediction. This research also shows that multiple linear regression model can be done to predict population growth rate because they use the adaptation of it.

According to the study Growth Sensitivity of the Nigerian Population and a Prediction for the Future [11], it was based on a systematic data collecting, analyzing, and interpreting in a research and calculations made using MATLAB's Logistic Growth mathematical model. Ekakitie [11] focus on the Thomas R. Malthus [17] population growth model's anticipated and sensitivity growth model, which is used to simulate the population of Nigeria from 2015 to 2125. The carrying capacity, sensitivity, and significant coefficients that influence population increase in Nigeria have been determined. The carrying capacity of the Nigerian population is 808719320.62, with coefficients and of 0.03 and 3.71×10^{-11}, respectively. Nigeria's population growth rate, according to this estimate, is 3% each year. Nigeria's population is expected to reach 802,430,000 by 2125, with a population sensitivity of 189,900. Then from 2115 to 2215, the population rises significantly before stabilising. As a result, population growth after 2215 becomes far more constant which means it is approaching zero [11].

Fertility has the greatest effect on population growth because of its multiplier effect the children born today will have children in the future, and so on. In predicting population size, the long-term population size is determined by both the eventual fertility level and the transition path. The two elements interact as well: the lower the predicted final fertility level, the more important the rate of fertility decrease becomes in terms of projected population size [12]. The majority of mortality projections are based on estimating life expectancy at birth. Mortality projections must include how the distribution of mortality across different age and gender groups may change over time. Population growth and age structure are affected differently by changes in mortality at different ages. When baby and child mortality rates fall, for example, a higher percentage of babies will live to adulthood and have their own children, contributing to future growth. Because the surviving is already passed reproductive age, mortality declines among the elderly have a more immediate impact on population growth [13].

The COVID-19 pandemic has plainly wreaked havoc on the global economy and social order. Many governments have imposed temporary limitations on personal mobility, crossing state and international borders, social gatherings, employment attendance, and which retail businesses are allowed to remain open to slow the spread of the disease. To some extent, most of the the world's population has been affected [14]. Forecasting the population is particularly dangerous at this time, as even the short-term future of the epidemic and its demographic effects is a source of enormous uncertainty. Like the Spanish flu a century ago, it's possible that COVID-19 will spread in waves [15]. Government decisions on border re-opening will have a significant impact on international migration trends, but they are difficult to predict at this time. They conducted a study in Australia to determine the influence of COVID- 19-related changes in underlying demographic characteristics on the number and size of the elderly Australian population. COVID-19 would reduce the entire population by 522,000 (shorter scenario) to 1.97 million individuals by 2041, according to the researchers (Longer scenario) [16]. Furthermore, the pandemic may modify our specified age profiles of fertility, mortality, immigration, and emigration, which are based on recent years of data. Considering all of this uncertainty, the situation will need to be continuously monitored, and population predictions will need to be adjusted on a frequent basis.

3. Methodology

3.1. Description of the Data and Method

The data for this research is extracted from Department of Statistic Malaysia (DOSM). The yearly data of Malaysia's Population Statistics which covered the period of year 1978 until year 2018. The data consists of total population by age category, average annual population growth rate, crude birth rate, crude death rate, neonatal mortality rate and maternal mortality rate. The data are analyzed using Microsoft Excel and SPSS software. Population growth rate is the dependent variable for this research, while the others variable is independent variables. The population growth rate in Malaysia will be predicted by using Multiple Linear Regression model.

3.2. Multiple Linear Regression

Regression analysis is a type of predictive model that estimates the relationship between two or more variables [18]. The relationship can be either a straight line (linear regression), a polynomial curve (polynomial regression), or a non-linear relationship (non-linear regression). A scatter plot of the target and predictor variables (simplest and most popular way). The strength of the relationship can describe the direction of the relationship whether the independent variables have negative or positive relationship with dependent variable. More specifically, regression analysis explains how the typical value of the dependent variable changes when any independent variable is changed while the other independent variables remain constant. One of the most widely used predictive modelling techniques is linear regression.

The multiple linear regression model is defined as the following:

$$y\, y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon \qquad (1)$$

where:
y is the predicted value of the dependent variable
β_0 is the y-intercept (value of y when all other parameters are set to 0)
$\beta_1 x_1$ is the regression coefficient (β_1) of the first independent variable (x_1)
$\beta_n x_n$ is the regression coefficient of the last independent variable
ϵ is model error (how much variation there is in our estimate of y)

3.3. Matrix Notation

Matrix notation is used to simplify the presentation of calculations that are performed in the multiple linear regression. An r × c matrix is a rectangular array of elements with r rows and c columns. The least squares approach is used to estimate the regression coefficients in multiple linear regression analysis. The regression coefficients represent the unrelated contributions of the independent variables to predicting the dependent variable. The computations to find the least squares estimators, β turned out to be fairly difficult. As a result, to make calculations easier, multiple linear regression can be represented in matrix notation. The model is in the form $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$ and when written in matrix notation we get

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \qquad (2)$$

The vector of least square estimators is determined as the initial step in multiple linear regression analysis. The parameter estimators, $\hat{\beta}$ can be estimated by minimizing the normal equation. Where y is in the form

$$yy = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon \tag{3}$$

To find least square estimates for $\beta_0, \beta_1, \ldots, \beta_n$ in a MLR model, we must solve a set of simultaneous equations, called normal equations as the following.

$$x^T x^n \hat{\beta} = x^T y \tag{4}$$

$$\sum y_i = n\widehat{\beta_0} + \sum x_{i1}\widehat{\beta_1} + \cdots + \sum x_{i\rho}\widehat{\beta_\rho} \tag{5}$$

$$\sum x_{i\rho}\, y_i = \sum x_{i\rho}\widehat{\beta_0} + \sum x_{i1}\widehat{\beta_1} + \cdots + \sum x^2{}_{i\rho}\widehat{\beta_\rho} \tag{6}$$

Thus, the least squares estimators can be determined by solving the $\hat{\beta}$ as following

$$\hat{\beta} = (x^T x)^{-1} x^T y \tag{7}$$

3.4. Relationship among Variables

A correlation matrix is the correlation between all possible pairings of variables in a matrix. It is a useful tool for quickly summarizing a large dataset and to investigate the direction of relationship among variables whether they are negatively related or positively related, a scatter plot can help visualize its type and direction. However, to determine the strength of the relationship, a correlation matrix provides insight with estimates of Pearson's correlation coefficient. A correlation matrix also can visualize trends in the data. The models have a number of independent variables. The correlation matrix determines the correlation coefficients between the independent variables in multivariate linear regression in a model.

3.5. Assumptions of Multiple Linear Regression

To interpret a model and its limitations, it is important to understand the underlying assumptions of the method and how these affect the treatment of the data and modelling choices made. The assumptions of MLR model are:

- Linearity between independent variables and dependent variables
- Absence of multicollinearity problem
- Constant variance
- Normality of the residuals

3.5.1. Linearity between independent variables and dependent variables

The main assumption of the MLR is to model the linear relationship between dependent variable and independent variables, thus each independent variables will be checked whether the linearity is significant with dependent variables. To check the linearity, visually we used simple scatter plot between dependent variable and each independent variable. Making scatterplots and visually inspecting them for linearity is the best technique to verify for linear correlations. If the scatterplot shows a non-linear relationship, the analyst must perform a non-linear regression or transform the data to have a linear relationship.

3.5.2. Absence of Multicollinearity Problem

Multicollinearity problem exist when the independent variables (explanatory variables) are highly interrelated. When independent variables exhibit multicollinearity, determining the precise variable that contributes to the variance in the dependent variable becomes difficult. To check whether multicollinearity occurs, a correlation matrix of all independent variables is produced as discussed in section 3.4. The other method to check multicollinearity is by using the Variance Inflation Factor (VIF) for each independent variable. Higher correlation exists between the vars. If VIF value is more than 10, the independent variables are said to have serious multicollinearity problem. However, acceptance range is subject to requirements and constraints.

3.5.3. Constant Variance

A constant variance is another important assumption has to be checked before build a MLR model.
To check this assumption, a plot of predicted value versus standardized residual can be plotted and if the points are distributed evenly, the variance is constant. however, if an obvious pattern is observed in which residuals are increasing with higher predicted values, homoscedasticity assumption is violated. This diagnostic plot can be produced using, Microsoft excel or SPSS.

3.5.4. Normality of The Residuals

Normality is the assumption that the underlying residuals are normally distributed. The null hypothesis can be rejected and it can be determined that the residuals are not from a normal distribution if the test p-value is less than the predetermined significance level. Only with tiny sample numbers does the normalcy assumption come into question.

3.6. Analysis of Variance

Analysis of Variance (ANOVA) provides the analysis of the variance in the model. Regression statistics provides numerical information on the variation and how well the model explains the variation for the given data/observation. Meanwhile the residual output provides the difference between the actual value and its predicted value by the regression model for each data point. The null hypothesis states that a population parameter (such as the mean, the standard deviation, and so on) is equal to a hypothesized value. The null hypothesis states that all coefficients in the model are equal to zero. In other words, none of the predictor variables have a statistically significant relationship with the dependent variable. The alternative hypothesis states that not every coefficient is simultaneously equal to zero.

$$\boldsymbol{H_0} : \beta_1 = \beta_2 = \ldots = \beta_4 = 0$$
$$\boldsymbol{H_1} : \beta_1 = \beta_2 = \ldots = \beta_4 \neq 0$$

4. Results and discussion

4.1. Scatter Plot

Each independent variable is investigated whether there is a linear relationship with the dependent variable by plotting scatter plots as in figure 4.1 - 4.4. Scatter plot for each independent variable is presented in Figure below. Based on the plots, only crude birth rate is observed to have positive linear relationship as most of the points are lie on a straight line. As the crude birth rate increase, the population growth will increase. The remaining plots indicate negative relationship exist between crude death rate, maternal mortality rate, and neonatal mortality rate with the dependent variable, population growth. This implies, increasing crude birth will increase the population growth rate while if the crude death rate increase, then the population growth rate will decrease. The strength of the relationship is various for different independent variables and will be investigated using correlation matrix in the next subsection.

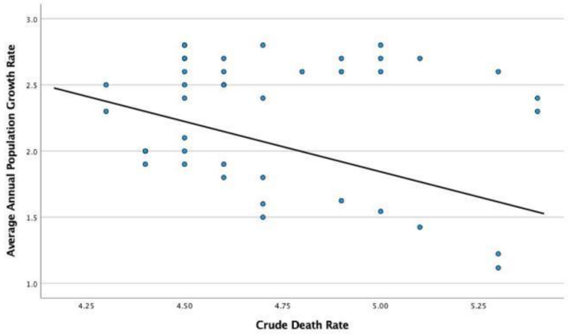

Figure 4.1 Simple scatter of population growth rate by crude death rate

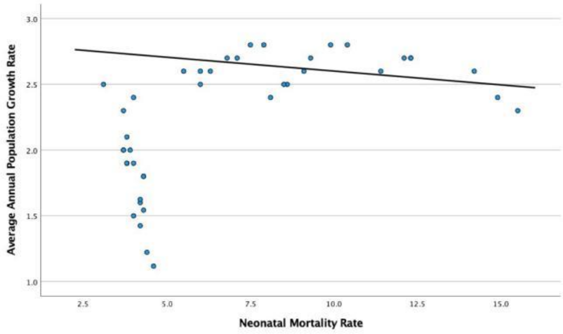

Figure 4.3 Simple scatter of population growth rate by neonatal mortality rate

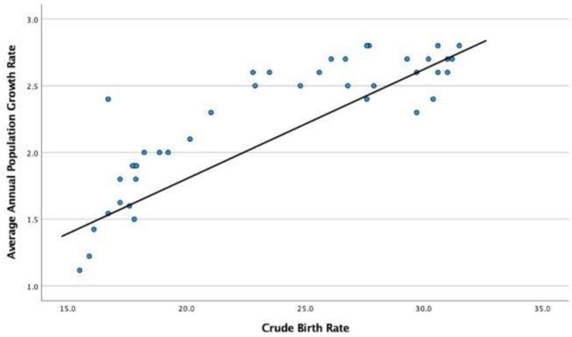

Figure 4.2 Simple scatter of population growth rate by crude birth rate

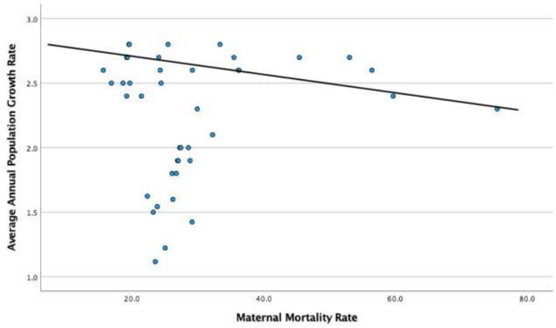

Figure 4.4 Simple scatter of population growth rate by maternal mortality rate

4.2. Correlation Matrix

From Table 4.2, crude birth rate has strong positive relationship with neonatal mortality rate with $\rho = 0.877$, crude death rate has moderate positive relationship with maternal mortality rate with $\rho = 0.641$ and neonatal mortality rate has moderate relationship with maternal mortality rate with $\rho = 0.683$. These indicate there exist multicollinearity problem in the data because of high Pearson correlation coefficient ($\geq \pm 0.6$) thus few variables will be excluded from the modelling. Those with higher Pearson correlation estimates towards dependent variable are prioritised to stay for modelling since they have higher power in explaining the variation in population growth. The revised correlation matrix is produced and shown in Table 4.3.

Table 4.1 Correlation Matrix (Four variables)

		Average Annual Population Growth Rate	Crude Birth Rate	Crude Death Rate	Neonatal Mortality Rate	Maternal Mortality Rate
Average Annual Population Growth Rate	Pearson Correlation	1	.859**	-.191	.594**	.134
	Sig. (2-tailed)		<.001	.232	<.001	.403
	N	41	41	41	41	41
Crude Birth Rate	Pearson Correlation	.859**	1	.170	.877**	.380*
	Sig. (2-tailed)	<.001		.288	<.001	.014
	N	41	41	41	41	41
Crude Death Rate	Pearson Correlation	-.191	.170	1	.550**	.641**
	Sig. (2-tailed)	.232	.288		<.001	<.001
	N	41	41	41	41	41
Neonatal Mortality Rate	Pearson Correlation	.594**	.877**	.550**	1	.683**
	Sig. (2-tailed)	<.001	<.001	<.001		<.001
	N	41	41	41	41	41
Maternal Mortality Rate	Pearson Correlation	.134	.380*	.641**	.683**	1
	Sig. (2-tailed)	.403	.014	<.001	<.001	
	N	41	41	41	41	41

After drop two variables which are neonatal mortality rate and maternal mortality rate, the correlation matrix is as the Table 4.2 below.

Table 4.2 Correlation Matrix (Two variables)

		Average Annual Population Growth Rate	Crude Birth Rate	Crude Death Rate
Average Annual Population Growth Rate	Pearson Correlation	1	.859**	-.191
	Sig. (2-tailed)		<.001	.232
	N	41	41	41
Crude Birth Rate	Pearson Correlation	.859**	1	.170
	Sig. (2-tailed)	<.001		.288
	N	41	41	41
Crude Death Rate	Pearson Correlation	-.191	.170	1
	Sig. (2-tailed)	.232	.288	
	N	41	41	41

Based on the table above all independent variables have significant relationship with y with low Pearson correlation coefficient ($\leq \pm 0.6$), hence we can conclude that we have overcome multicollinearity problem. By default, SPSS always creates a full correlation matrix. Each correlation appears twice: above and below the main diagonal. The correlations on the main diagonal are the correlations between each variable and itself which is why they are all 1. Based on table above, the result shows that crude birth rate and crude death rate have the weak correlation, $\rho = 0.170$. It's based on $n = 41$ years and its 2-tailed significance, $p = 0.288$. The more human born, the more human death, but the effect is very small. At 5% significance level the relationship exist between these 2 variables is not significant.

4.3. Variance Inflation Factor (VIF)

The VIF values for each of the predictor variables are as the Table 4.3 below:
Crude Birth Rate: 15.006
Crude Death Rate: 3.713
Neonatal Mortality Rate: 24.824
Maternal Mortality Rate: 3.126

Table 4.3 VIF value (Four variables)

Coefficients[a]

Model		Collinearity Statistics Tolerance	VIF
1	Crude Birth Rate	.067	15.006
	Crude Death Rate	.269	3.713
	Neonatal Mortality Rate	.040	24.824
	Maternal Mortality Rate	.320	3.126

a. Dependent Variable: Average Annual Population Growth Rate

As from the result above, two variables have a value greater than 10 which indicates potentially severe correlation between given predictor variables and other predictor variable in the model. In this case, the coefficient estimates and p-values in the regression output are likely unreliable. Hence, we can conclude the presence of multicollinearity in the data. So, we decide to drop a few highly correlated features to remove multicollinearity in the data.

After drop two variable which is neonatal mortality rate and maternal mortality rate, the new VIF values for each of the predictor variables are as the Table 4.4 below.

Table 4.4 VIF value (Two variables)

Coefficients[a]

Model		Collinearity Statistics Tolerance	VIF
1	Crude Birth Rate	.971	1.030
	Crude Death Rate	.971	1.030

a. Dependent Variable: Average Annual Population Growth Rate

As from the result above, we can see that none of the VIF values for the predictor variable that we choose are greater than 10 which is both VIF value is 1.030. This indicates that multicollinearity problem is absence in the regression model.

4.4. Constant Variance

Before conducting the regression analysis, the assumption of the regression must all be checked. The Figure 4.5 below shows the residuals are scattered randomly about zero in no particular pattern with roughly constant variance at every level of the fitted values.

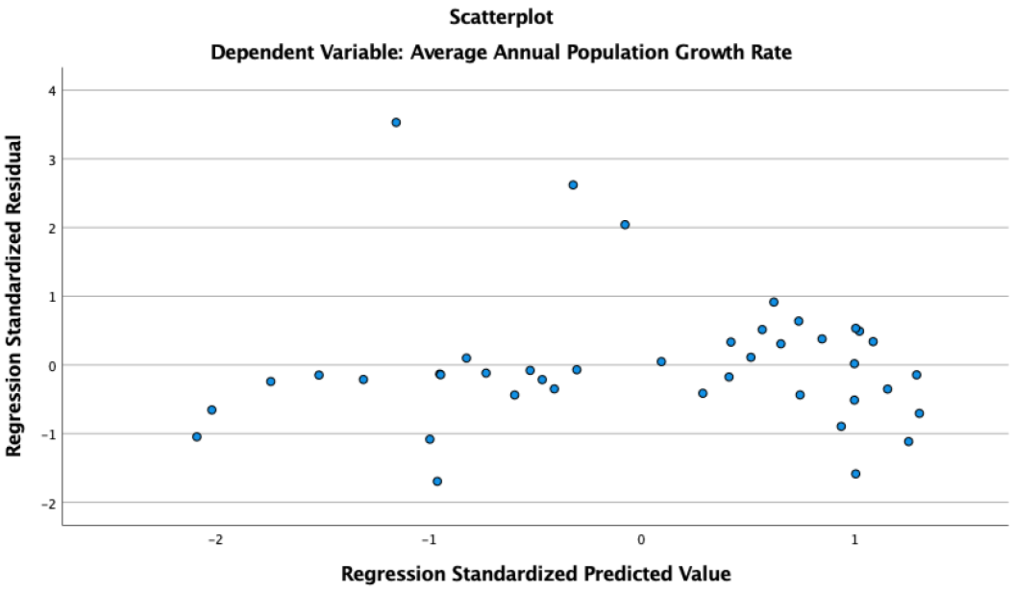

Figure 4.5 Residual scatter plot

4.5. Multiple Linear Regression

Table 4.5 below provide R and R^2. The R value represents the simple correlation and is 0.925 which indicates a high degree of correlation. The R^2 value indicates how much of the total variation in the dependent variable can be explained by the independent variables. In this case, 85.5% can be explained, which is very large.

Table 4.5 Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.925[a]	.855	.848	.1888

a. Predictors: (Constant), Crude Death Rate, Crude Birth Rate

Table 4.6 below indicates the statistical significance of the regression model that was run. Here, p-value < 0.001, which is less than 0.05, and this indicates that overall, the regression model statistically significant to predict the outcome variable. In other words, it is a good fit for the data to predict dependent variable.

Table 4.6 ANOVA Table

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.008	2	4.004	112.356	<.001[b]
	Residual	1.354	38	.036		
	Total	9.362	40			

a. Dependent Variable: Average Annual Population Growth Rate

b. Predictors: (Constant), Crude Death Rate, Crude Birth Rate

The Coefficients Table 4.7 below provides us with the necessary information to predict average annual population growth rate from crude birth rate and crude death rate, as well as determine whether crude birth rate and crude death rate contributes statistically significant to the model. The significance values for both independent variables are less than 0.05. Thus, both of it are significant variables to include in the regression model.

Table 4.7 Coefficients Table

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	2.961	.459		6.451	<.001
	Crude Birth Rate	.079	.005	.918	14.667	<.001
	Crude Death Rate	-.541	.098	-.347	-5.544	<.001

a. Dependent Variable: Average Annual Population Growth Rate

From table above, the parameter estimators are:

$\widehat{\beta_0}$ = 2.961
$\widehat{\beta_1}$ = 0.079
$\widehat{\beta_2}$ = *-0.541*

Thus, the final multiple linear regression can be written as the following:

$$\hat{y} = 2.961 + 0.079x_1 + (-0.541)x_2.$$

Thus, we can say that the overall mean population growth rate is 2.961(unit). Then with one-unit increase in crude birth rate, the population growth rate increases by *0.079* $(\widehat{\beta_1})$ while with one-unit increase in crude death rate, the population growth rate decreases by *0*.541 $(\widehat{\beta_2})$ when crude birth rate is constant.

5. Conclusion

This research aimed to predict and explore the trends of population growth in Malaysia. In general, Malaysia population increase from 1978 to 1993 and then slightly decrease until 2018. Few studies have discussed potential factors affect population growth rate. Due to limitation of the data, only 4 independent variables that have been used for this research

Among all the factors, few have multicollinearity which are neonatal mortality rate and maternal mortality rate. The variables that are significant to explain population growth rate are only crude birth rate and death rate with low correlation 0.170 with each other. Both variables are significantly contributed to the model with p-value less than 5%. There exist positive relationship between annual population growth rate, y and crude birth rate, x_1 and, negative relationship between annual population growth rate,y and crude death rate, x_2. The result of MLR model for this research is ŷ = 2.961 + 0.079x_1 + (-0.541)x_2. From the final model, the iteration of crude birth and death rate for future years can be used to predict population growth rate. The study can conclude that crude birth and death rate are essential factors to be considered when designing regression model for population growth.

Selecting the best regression model is both a science and an art. Statistical approach can steer us in the correct path, ultimately, we need to incorporate other considerations. It is recommended to investigate what other researchers have done related to prediction in population growth rate and use that information to build the prediction model. We need to develop the understanding of the main variables and their relationships before commencing the regression analysis. Without data mining, it is easier to acquire the correct data and describe the best regression model following suggestions for other studies by building on the results of others. One main point is the theoretical implications should not be discounted only based on statistical data. After fitting the model, the researcher needs to check whether it matches with the theory and make any necessary revisions. Finally, no single metric can determine which model is the best. Statistical approaches are incapable of comprehending the underlying process or subject-matter.

Acknowledgement
The content is solely the responsibility of the author and does not necessarily reflect the official views of the university. The researcher would like to thank all people who have supported the research.

References

[1] Momoh, O. Population. Fundamental Analysis, 2021.
[2] Max Roser, H. R. and Ortiz-Ospina, E. World Population Growth. Our World in Data, 2013.
[3] Nations, U. Department of Economic and Social Affairs, Population Division (2017). World population prospects: the 2017 revision, key findings and advance tables. In: Working Paper No. ESA/P/WP/248. 2017.
[4] World Population Clock Retrieved from https://www.worldometers.info/world- population/.
[5] Roser, M. Future population growth. Our world in data, 2013.
[6] Population Distribution and Basic Demographic Characteristic Report 2010, 2011.
[7] Malaysia Population 2022 Retrieved from https://worldpopulationreview.com/countries/malaysia-population, 2022.
[8] Tackling the Population Problem Retrieved from https://worldpopulationreview.com/countries/malaysia-population.
[9] Arshat, H. and Peng, T. N. An overview of the population dynamics in Malaysia. Malaysian journal of reproductive health: a publication of the Reproductive Research Centre of the National Population and Family Development Board, Malaysia, 1988. 6(1): 23–46.
[10] Qu, J., Cui, Y., Yang, A., Peng, Y. and Zhang, H. The Study for the Prediction Model of China Population Growth. J. Comput., 2011. 6(10): 2076–2083.
[11] Ekakitie, O. and Ekereke, L. Growth Sensitivity of the Nigerian Population and a Prediction for the Future. International Journal of Applied Science and Mathematical Theory, 2019. 5(3).
[12] O'Neill, B. C., Scherbov, S. and Lutz, W. The long-term effect of the timing of fertility decline on population size. Population and Development Review, 1999. 25(4): 749–756.
[13] Lutz, W. and Qiang, R. Determinants of human population growth. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 2002. 357(1425): 1197–1210.
[14] of Migration, I. O. IOM COVID-19 Impact On Key Locations Of Internal Mobility Weekly Analysis 24 June 2020. 2020.
[15] Martini, M., Gazzaniga, V., Bragazzi, N. L. and Barberis, I. The Spanish Influenza Pandemic: a lesson from history 100 years after 1918. Journal of preventive medicine and hygiene, 2019. 60(1): E64.
[16] Wilson, T., Temple, J. and Charles-Edwards, E. Will the COVID-19 pandemic affect population ageing in Australia? Journal of Population Research, 2021: 1–15.
[17] Malthus, T. R., Winch, D., & James, P. (1992). Malthus:'An Essay on the Principle of Population'. Cambridge university press.
[18] Panchotia, R. Predictive Modelling Using Linear Regression. 2020.