Predicting the Philippine Population by 2045 using Least Squares Fitting

Gimel David F. Velasco
Department of Mathematics and Computer Science
University of the Philippines
gfvelasco@up.edu.ph

ABSTRACT

This paper seeks to predict the population in the Philippines by the year 2045 using the data presented in the Philippine Statistics Authority. The method of Curve Fitting or Least Squares Fitting is used to generate a function that represents the trend of the population in the Philippines.

1. INTRODUCTION

The Philippine Statistics Authority published an article in the PSA Website on July 28, 2014 discussing about their predicted population in the Philippines by the year 2045. The articles says that the Philippines would have a population of 142,095,100 by the year 2045. This includes both male and female of all ages in the whole country. This study in now replicated by using the datasets from the PSA. The datasets consists of the recorded and predicted population in the whole Philippines from 2010 to 2020. From these data, three different functions were generated by using the Curve Fitting or Least Squares Fitting Method.

2. METHODOLOGY

The Curve Fitting or Least Squares Fitting Method is used to generate the functions that will be able to predict the population of the Philippines. In this study, the year 2045 is used. The kind of Curve Fitting Methods used in this study are the following: Linear Curve Fit, Quadratic Curve Fit and Exponential Curve Fit. Thus, generating three different functions for predicting population size. The methods are further explained in the next subsections.

2.1 Linear Curve Fitting

The Linear Curve Fitting which is also known as Least-Squares Line is used so that a straight line would be fitted a set of data points by solving the coefficients a and b in the function:

$$f(x) = ax + b \tag{1}$$

The coefficients A and B are then computed by solving a system of equations that can be represented as matrices of size 2x2. That is

$$AX = B \tag{2}$$

where the matrix X contains a and b. The coefficients then are plugged into the function in equation 1.

2.2 Quadratic Curve Fitting

The Quadratic Curve Fitting which is a Polynomial Curve Fitting with a degree of 2. The same procedure to that of Linear Curve Fitting is followed for generating the function for the Quadratic Curve Fitting. Only that the function has three unknown coefficients and the matrix has the size of 3x3. The function form for the Quadratic Curve Fitting is

$$f(x) = a_2 x^2 + a_1 x + a_0 \tag{3}$$

After solving the coeffients in the matrix, it is then plugged into the function above.

2.3 Exponential Curve Fitting

The Exponential Curve Fitting which also follows a similar procedure to that of the Linear Curve Fitting. Only that this procedure has the function form for the Exponential Curve Fitting is

$$f(x) = Ae^{Bx} \tag{4}$$

,where A = exp(a) and B = b. The variables a and b are solved by analyzing a 2x2 matrix similar as that in equation 2 only that the elements of each matrix is different and the matrix X contains a and b. After solving for the values of a and b, the value of A and B will be solved and then plugged into the equation stated above.

2.4 Error Analysis

The error of each generated function is then calculated based on the actual data. The error equation is represented below

$$err = \Sigma_{i=1}^{n}(y_i - f(x_i))^2 \tag{5}$$

where n is the number of data points. This is the equation that explains why the Least Square Fitting got its term. The expression inside the parenthesis is simply the vertical distance of the actual data point to the point in the equation. The distance then is squared.

2.5 Predicting Population

Now for predicting the Philippines population by the year 2045, a value is plugged in to the variable x in all the three different functions. Note that for the function f(x), the year 2010 is the initial year of the dataset. Thus, makes 2010 represented as x = 0 in the function. Therefore, this implies that the year 2045 is x = 35. The value x = 35 then in plugged into all the three generated functions. After all the values of f(x=35) is calculated, the results were then compared with the prediction of PSA of a Philippine population of 142,095,100 by 2045. And for calculating the error between the population yielded in the functions and the predicted population of PSA, the following equation is used:

$$error_{relative} = \left| \frac{P_{psa} - P_{f(x=35)}}{P_{psa}} \right| \qquad (6)$$

3. RESULTS & DISCUSSION

All the functions and their corresponding err value is shown below along with the predicted Philippine population by 2045.

3.1 Generated Functions

The results of performing all three of the methods of curve fitting are discussed below. For the Linear Curve Fit, the function below is generated.

$$f(x) = 1681286.363636x + 93147568.181818 \qquad (7)$$

with err = 596099545.454397. Now for the Quadratic Curve Fit, the function below is generated.

$$f(x) = -833.449883x^2 + 1689620.862471x + 93135066.433566 \qquad (8)$$

with err = 99533.799536. For the Exponential Curve Fit, the following function is generated.

$$f(x) = 93353740.628710e^{0.016565x} \qquad (9)$$

err = 186835129329.435610 Note that the err value is not represented as an error percentage nor a truncation error. It is the summation of all the vertical distances between the curve fit function and their respective data point.

3.2 Comparing Predictions

The Predicted Population by 2045 is the ff: Using Linear Curve Fit:

$$P_{2045} = 151992590.909091 \qquad (10)$$

with error = 0.069654

Using Quadratic Curve Fit:

$$P_{2045} = 151250820.512820 \qquad (11)$$

with error = 0.064434

Using Exponential Curve Fit:

$$P_{2045} = 166694669.899408 \qquad (12)$$

with error = 0.173120

So looking at the relative error of each yielded population prediction, both the Linear and Quadratic Curve Fit yielded a low error compared to PSA's predicted Philippine population by 2045. The Exponential Curve Fit though had a big error. If we analyze the three Curve Fit predictions, the Quadratic Curve Fit provided the closest prediction. The downside of selecting a quadratic function is that if we were to predict the Philippine population in a much further year, we will yield a value that is decreasing since this quadratic function has a parabolic graph facing downward. So using a quadratic curve fit is not advisable to be used for much further predictions. Meanwhile, both the Linear Curve Fit and Exponential Curve Fit are the much more fit kind of functions to follow the trend of the Philippine population since both of them follows an increasing population size through time.

4. CONCLUSION

The functions have generated a population prediction that is very close to PSA's population prediction by 2045. So therefore based on the Least Squares Fitting Method and PSA's prediction, the Philippine Population by the year 2045 is around 140-160 Million. This also is most likely to happen due how close the predicted populations are to each other. For further studies, it is recommended to have a much larger dataset covering the population of the Philippines in the past years before 2010. Also, since the Philippine population has an increasing trend, a function that has an increasing trend is also recommendable. Such examples are the Linear and Exponential Curve Fit.

5. REFERENCES

[1] L. S. Bersales, "A 142 Million Philippine Population by 2045?", July 28, 2014. Available: https://psa.gov.ph/content/142-million-philippine-population-2045. [Accessed May 21, 2016].

[2] J. H. Matthews and K. D. Fink, "Numerical Methods Using MATLAB, 3rd ed.", pp. 252-275,

[3] E. W. Weisstein "Least Squares Fitting–Exponential" Available: http://mathw

orld.wolfram.com LeastSquaresFittingExponential.html. [Accessed May 21, 2016]