International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 6 800 – 805

Development of a New Model for Population Prediction in Anambra State, Nigeria

*Okeke, Ogochukwu Clementina, **Dr. Onyesolu, Moses Okechukwu, ***Ndigwe, Chinwe F.
* (Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University,
Uli, Anambra State, Nigeria
Email: ogoookeke@yahoo.com)
**(Department of Computer Science, Nnamdi Azikiwe University,
Awka, Anambra State, Nigeria
Email: mo.onyesolu@unizik.edu.ng)
*** (Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University,
Uli, Anambra State, Nigeria
Email: franchy_dex@yahoo.com)

Abstract:- This paper was motivated by the desire to develop a new model for National Population Commission in Anambra state for better prediction. Population demographic analysis are calculated with specified formula using non-free software, this makes it difficult for researchers to access data, Census also Census figures are calculated without any mining of patterns or trends of existing demographic information. These are the gaps this research paper covers.The system is a desktop application based on states and behaviours of objects. It was approached with the Object-Oriented Analysis and Design Methodology and implemented with Visual Basic. Net. The developed model accepts given probabilistic demographic information to make prediction. Data-mining warehouse was also developed to make population distribution decision, find hidden and relationship using SQLite database. This desktop application requires only data as input and not formulae. The system brings a multi-line user input interface which allows users to input five (5) or more data input (observations). These observations are employed by the new model to predict future values of demographic information. The software uses voice instruction to tell users how use the software.The resultof this paper is functional predictive demographic information in Anambra State.

Keywords: *Census, model, Prediction*

*****

Introduction

Census data is often not critically analyzed to bring out some of the hidden and important attributes of the population in Anambra State. This is usually due to non-availability of the required tools for carrying out such analysis. This paper makes use of a new modelto extract hidden information from large census data warehouse. The new model uses five data input to make predictions (Okeke, Onyesolu and Ndigwe, 2017). Data-mining is all about trends and patterns of existing information. The model predicts and display output in tabular manner. National Population Commission have a lot of repository of data that needs to be extracted for effective planning, this model extracts the hidden demographic information which are used in policy making(Okeke, Onyesolu and Ndigwe, 2017). The Information on population growth is critical among decision makers, planners and local land users who rely on such information to assess the physical extent, characteristics, and consequences of past and future population development. Demographic research recently uses quantitative methods for analyzing changes in population growth without consideration of the spatial aspect of population distribution (Branislav, Nikola, Milan and Mileva, 2011). The problem of population flows, the phenomena and processes by which it is determined, is often not sufficiently dealt with in traditional demographic analysis, especially on the strategic level, and in most cases, it is reduced to the simple statistical analysis (Branislav, Nikola, Milan and Mileva, 2011). The links and relations that exist between certain population structures and natural-geographic characteristics of a region are not researched in such analyses (Branislav, Nikola, Milan and Mileva, 2011).Moreover, population demographic analyses are not always accurate because of the prediction tool used.Some people in some areas will be counted while others will not be counted. Some people will be employed while others will not be employed and some states are more marginalized than others. Why must you vote where you register only, all these questions will be answered only when accurate and standard census process is taken using data-mining model. Demographic characteristics of population were collected in carrying out this research.

Databases today can range in size up to the terabytes more than 1,000,000,000,000 bytes of data. Within these masses of data lies hidden information of strategic importance. But when there are so many trees, how do you draw meaningful conclusions about the forest? The newest answer is data mining, which is being used both to increase revenues and to reduce costs (Yang, Akers, Klose and Barcelon, 2008). The potential returns are enormous. Innovative organizations worldwide are already using data mining to locate and appeal to higher-value customers, to reconfigure

their product offerings, to increase sales, and to minimize losses due to error or fraud.

Data-mining is the process of discovering previously unknown, actionable and profitable information from large consolidated databases and using it to support tactical and strategic decisions (Michael and Gordon, 2012). It is also extraction of hidden predictive information from large databases; it is a powerful new technology with great potential to help companies, industries, institutions, government etc. focus on the most important information in their data warehouses (Kurt, 2012). Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems (Acm Sigkdd, 2006). The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use (Michael and David, 2007).

Data-mining tools predict future trends and behaviors, allowing business to make proactive, knowledge –driven decisions. The automated, prospective analyses offered by data-mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems (Kurt, 2012).The most important aspect in data mining is the quality of data because it influences the quality of the result. Data-mining tools can answer business questions that traditionally were time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations (Kurt, 2012).Most companies already collect and refine massive quantities of data. Data-mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data-mining tools can analyze massive databases to deliver answers to questions.

Census is the procedure of systematically acquiring and recording information about the members of a given population (Baffour, King and Valente, 2013). It is regular occurring and official count of a particular population (Baffour, King and Valente, 2013). A population and housing census is of great relevance to the economics, political and socio-cultural planning of a country (United Nation, 2008). Reliable and detailed data on the size, structure, distribution and socio-economic and demographic characteristics of a country's population is required for planning, policy intervention and monitoring of development goals (United Nation, 2008). Population projection refers to the estimates of population in near future (David, 2013). Almost every country carries out census to collect different features of each geographic region and to have count on population in each region. Collection, analysis and presentation of these data are called Statistics (David, 2013). These censual data can be used to make future predictions which could help planning commission to make plans and focus on areas which require special attention (Swati, Nitin, SakshiandSaumya, 2015). Government mostly relies on centralized and official sets of population forecasts based on which capital facilities are planned. If the value is underestimated, it will make system inadequate for the purpose intended; similarly if the value is overestimated, the system will become costly. Changes in the population of the city over the years occur, and the system should be designed taking into account these changes in the population at the end of the design period. Within the masses of information in the census database lays hidden information of strategic importance. Data-mining is a key element in finding the particular pattern and relationship that can help governments, organizations and businesses (Kurt, 2012). Data-mining find those patterns and relationships using sophisticated data analysis tool and technique to build models. The new data-mining model predicts attributes of the population like total population in each local government, employment, unemployment, death rate, birth rate, number of males, number of females, marital status.

Problem Statement; Census demographic analyses are calculated with formulae that uses wide range constant making it a poor predictor. The objective of this paper is develop a new model that uses narrow range constant for prediction and outputs a better result of future values.

Methodology

Object Oriented Analysis and Design Methodology was used in the analysis. Objects in this work are coefficient model, time coefficient and prediction model. The objects are organized into data preparation object and prediction object. Data preparation object was initiated first and prediction object carries out the prediction.

The new model: The new model was adopted from multiple regressions.

The new model is $\hat{\beta} = (T'T + \ln KI)^{-1}T^{'}Y$is adopted from multiple regression $\hat{\beta} = (T'T)^{-1}T^{'}Y$. Where T is a matrix,$T^{'}$ is matrix with rows and columns, Y is a centered n-vector,T'T is a scalar that represents sum of squares of the element (Scalar).

The above model yields a 3 x 3 matrix that contains

$$T^{'}T = \begin{bmatrix} n & \sum t & \sum t^2 \\ \sum t & \sum t^2 & \sum t^3 \\ \sum t^2 & \sum t^3 & \sum t^4 \end{bmatrix} \quad 1$$

 ISSN: 2321-8169
800 – 805

Where n is the number of years observed, $\sum t$ is the summation of years, $\sum t^2$ is the summation of years squared, $\sum t^3$ is summation of years in order three and $\sum t^4$ is thesummation of years in order four. The number of years to be observed by the model must be five i.e. n = 5 based on the algorithm developed.

$$T^{'}Y = \begin{bmatrix} \sum p \\ \sum tp \\ \sum t^2 p \end{bmatrix} \quad 2$$

$\sum p$ is the summation of the total population of the years observed
$\sum tp$ is the summation of the product of time and population
$\sum t^2 p$ is the summation of the product of time squared and population

K is a constant that is used in the model which ranges from 1.0 to 2.70 i.e. $1 \geq K \leq 2.70$. After simulation, it was found the optimum K to be 1.0001
K= 1.0001, then lnK which is the natural logarithm of K is 0.000099995

Then lnKI

$$\begin{bmatrix} 0.000099995 & 0 & 0 \\ 0 & 0.000099995 & 0 \\ 0 & 0 & 0.000099995 \end{bmatrix} \quad 3$$

The calculations were done to obtain

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad 4$$

$$P_t = \beta_0 + \beta_1 t + \beta_2 t^2 \quad 5$$

The demographic information which was predicted with this model includes: number of registered births, number of registered deaths, total population of each local government areas in Anambra State, marital status, gender, employment and unemployment.

Review of Related Works

Manoj and Madhu (2012) applied a time series ARIMA model for predicting sugarcane production in India; their output was to forecast sugar cane production in future years. Md and Syed (2012) used cohort-component method in forecasting future levels of fertility, mortality, sex composition, migration and other parameters and their output was to forecast factors that cause population change.David, Alan and Bob (2010) applied Hamilton – Perry Method to forecast population using data of two recent censuses and the result showed how population increased in years to come.Kayode and Jimoh(2011) used regression analysis with time series data to periodically forecast stock market prices stock market prices and have been proven complement to other numeric forecasting method and the result was to predict stock market prices in future and financial institution.

Felix, Maren and Miroslaw (2013) proposed method for online failure prediction to find out causes of flight failure on the air. Uncertainty in measuring forecast was measured by David and Jeff (2013) using two approaches namely; projections based on alternative scenario and statistical forecast intervals (model based and empirical based).Hamilton-Perry method, which is a regression based approach, was used on four states(US) and nine test points to determine the factors affecting forecast accuracy. Stanley (2013) proposed a hybrid model for population projection based on age, sex and education. Firstly, Cohort-component method was used to predict population by sex and age group. Then, this two dimensional projection was converted to three dimensional projection by adding level of education. It was used to determine area that were educated and of a particular sex. David (2013) used rescaled version of Mean Absolute Percent Error (MAPE) for evaluating cross-sectional ,sub-national forecasts .He suggested that rescaled version should be used rather than MAPE as it preserves the useful information about error structure when substantial outliers exists. Furthermore, Spatial Prediction of Human Population Change by Branislav, Nikola, Milan and Mileva (2011) was carried out using multiple regression, geographically weighted regression and regression kriging. Their forecast showed how human population changed with factors like good health, infrastructure and social amenities.

ISSN: 2321-8169
800 – 805

Results

Full Table View — Number of registered birth

Print

Local Gov	Year	Population
Aguata	2020	9643.16
Anambra E...	2020	7246.76
Anambra ...	2020	9812.07
Aniocha	2020	5412.47
Awka North	2020	3410.71
Awka South	2020	18705.11
Ayamelum	2020	7711
Dunukofia	2020	12197.15
Ekwusigo	2020	10789
Idemili North	2020	32032.15
Idemili So...	2020	21286.97
Ihiala	2020	9868.7
Njikoka	2020	17566.29
Nnewi North	2020	11047.36
Nnewi South	2020	17792.29
Ogbaru	2020	7564.09
Onitsha N...	2020	23791.74
Onitsha So...	2020	14422.49
Orumba N...	2020	11510.59
Orumba So...	2020	5599.18
Oyi	2020	7386.29

Full Table View — Number of registered death

Print

Local Gov	Year	Population
Aguata	2020	190.81
Aniocha	2020	239.62
Awka South	2020	472.47
Awka North	2020	311.04
Ayamelum	2020	87.13
Dunukofia	2020	114.64
Ekwusigo	2020	620.14
Idemili North	2020	5235.75
Idemili So...	2020	1370.99
Ihiala	2020	764.15
Njikoka	2020	677.17
Nnewi North	2020	439.65
Nnewi South	2020	63.75
Ogbaru	2020	33.25
Onitsha N...	2020	392.47
Onitsha So...	2020	139.97
Orumba N...	2020	435.15
Orumba So...	2020	319.28
Oyi	2020	125.53
Anambra E...	2020	856.58
Anambra ...	2020	877.89

Full Table View — Married

Print

Local Gov	Year	Population
Aguata	2020	1762.09
Anambra E...	2020	362.83
Anambra ...	2020	93.43
Aniocha	2020	5686.33
Awka North	2020	240.46
Awka South	2020	1258.34
Ayamelum	2020	1441.99
Dunukofia	2020	3909.05
Ekwusigo	2020	1330.25
Idemili So...	2020	699.35
Njikoka	2020	850.51
Nnewi North	2020	1703.25
Nnewi South	2020	1916.12
Ogbaru	2020	2123.18
Onitsha N...	2020	1192.41
Onitsha So...	2020	1638.74
Orumba N...	2020	7633.89
Orumba So...	2020	9212.9
Oyi	2020	1029.13
Idemili North	2020	7211.73
Ihiala	2020	1178.76

Full Table View — Not Married

Print

Local Gov	Year	Population
Aguata	2020	3827.21
Anambra E...	2020	945.9
Anambra ...	2020	3378.95
Aniocha	2020	8220.33
Awka North	2020	5328.69
Awka South	2020	6719.49
Ayamelum	2020	1643.58
Dunukofia	2020	1555.96
Ekwusigo	2020	5478.57
Idemili North	2020	8922.92
Idemili So...	2020	3774.69
Ihiala	2020	2569.63
Njikoka	2020	3008.95
Nnewi North	2020	5093.19
Nnewi South	2020	2075.13
Onitsha N...	2020	1058.02
Ogbaru	2020	6240.92
Onitsha So...	2020	3341.78
Orumba N...	2020	1502.11
Orumba So...	2020	2466.44
Oyi	2020	1565.52

 ISSN: 2321-8169
800 – 805

Full Table View — Divorced

Print

Local Gov	Year	Population
Aguata	2020	239.75
Anambra E...	2020	97.15
Anambra ...	2020	417.54
Aniocha	2020	287.19
Awka North	2020	10.38
Awka South	2020	932.37
Ayamelum	2020	233.47
Dunukofia	2020	120.77
Ekwusigo	2020	364.26
Idemili North	2020	1961.87
Idemili So...	2020	681.82
Ihiala	2020	663.51
Njikoka	2020	239.83
Nnewi North	2020	647.56
Nnewi South	2020	181.04
Ogbaru	2020	76.5
Onitsha N...	2020	390.2
Onitsha So...	2020	136.17
Orumba N...	2020	127.9
Orumba So...	2020	118.21
Oyi	2020	125.32

Full Table View — Employed

Print

Local Gov	Year	Population
Aguata	2020	13994.44
Anambra E...	2020	8981.45
Anambra ...	2020	14079.62
Aniocha	2020	29358.27
Awka North	2020	10127.76
Awka South	2020	18105.59
Ayamelum	2020	12646.62
Dunukofia	2020	2637.4
Ekwusigo	2020	16722.7
Idemili North	2020	24543.18
Idemili So...	2020	14050.38
Ihiala	2020	27317.01
Njikoka	2020	17016.5
Nnewi North	2020	12998
Nnewi South	2020	7974.2
Ogbaru	2020	12898.7
Onitsha N...	2020	22774.63
Onitsha So...	2020	23495.82
Orumba N...	2020	5609.86
Orumba So...	2020	21000.59
Oyi	2020	17372.33

Full Table View — Unemployed

Print

Local Gov	Year	Population
Aguata	2020	28951.44
Anambra E...	2020	15818.2
Anambra ...	2020	29634.6
Aniocha	2020	33281.13
Awka North	2020	29120.29
Awka South	2020	24987.35
Ayamelum	2020	21390.72
Dunukofia	2020	37147.09
Ekwusigo	2020	24892.98
Idemili North	2020	34020.58
Idemili So...	2020	29518.9
Ihiala	2020	34986.63
Njikoka	2020	20812.58
Nnewi North	2020	27970.77
Nnewi South	2020	27144.65
Ogbaru	2020	26451.31
Onitsha N...	2020	29872.38
Onitsha So...	2020	29984.94
Orumba N...	2020	31111.59
Orumba So...	2020	19303.44
Oyi	2020	32097.46

Full Table View — Total Population

Print

Local Gov	Year	Population
Aguata	2020	515023.44
Anambra E...	2020	103991.19
Anambra ...	2020	195807.28
Aniocha	2020	360120.53
Awka North	2020	105085.73
Awka South	2020	234562.1
Ayamelum	2020	154136.72
Dunukofia	2020	122367.83
Ekwusigo	2020	160074.77
Idemili North	2020	499832.14
Idemili So...	2020	209674.36
Ihiala	2020	322389.04
Njikoka	2020	130953.01
Nnewi North	2020	217326.85
Nnewi South	2020	264656.46
Ogbaru	2020	344521.68
Onitsha N...	2020	217495.11
Onitsha So...	2020	242865.51
Orumba N...	2020	228836.96
Orumba So...	2020	166439.81
Oyi	2020	197324.99

Conclusion

A new model that gives better prediction with some demographic information in Anambra State was developed. Problem of under estimation and over estimation have been taken control of since the model uses narrow range constant in making prediction and does not give optimal value of unknown parameters. It only considers values that are useful in explaining dependent variable.

REFERENCES

[1] Acm Sigkdd (2006). Machine Learning Algorithm and Data-mining. Cambridge UniversityPress. London.

[2] Baffour. B, King. T and Valente, P. (2013).The modern Census : Evolution, Examples and Evaluation. International Statistical Review 81(3):407-425.

[3] Branislave,B, Nikola, K, Milan,K, and Mileva, S (2011). Spatial Prediction of Human Population Change.AGILE.

[4] David, A. S. (2013). Measuring Uncertainty In Population Forecasts: A New Approach. UNECE.
[5] David, A. S, and Jeff, T. (2013). Measuring Uncertainty in Population Forecasts: A new
[6] Approach. UNECE Work Session on Demographic Projections organized in Cooperation with Istat (Rome, Italy).
[7] David, A. S, Alan, S and Bob, S. (2010). Forecasting the Population of Census tracts by age,and sex. Springer Population Research and Policy Review, Volume 29, Issue 1, pp 47-63.
[8] Felix, S, Maren, L, and Miroslaw, M. (2013). A Survey of Online Failure Prediction Methods. ACM Journal Name, Pages 1–68.
[9] Kayode, J and Jimoh, M (2011).Used regression analysis with times series to periodically forecast stock. Advances in Information Mining Volume 4,Issue 1, pp57-66
[10] Kurt, T. (2012). Introduction to Data Mining, Wiley: Asia.
[11] Manoj, K and Madhu, A. (2012). International Journal of Studies in Business and Economics. Pg 81- 94.
[12] Md. M and Syed, S. H. (2012). Population Forecasts for Bangladesh, Using a Bayesian Methodology, Journal of Health, Population and Nutrition (JHPN) , 30(4): 456–463.
[13] Micheal, B and Gordon, L (2012).Data mining Techniques.3rd Edition. Wiley: United States.
[14] Micheal, R. B and David .J. H (2007).Intelligent Data Analysis.2nd Edition. Springer: New York.
[15] Stanley, K. S. (2013). Further thoughts on simplicity and Complexity in Population
[16] Projection Models. International Journal of Forecasting. Elsevier.
[17] Swati, J. Nitin. M, Sakshi. B, and Saumya, C. (2015). A Review Paper on Forecasting of
[18] Demographic Features Using Statistical and Data - Mining Methods. International Journal of Science and Research (IJSR).ISSN (Online): 2319-7064.
[19] Tomlinson, (2011).Thinking about GIS: Geographical information system planning for managers: ESRI Press: New Delhi.
[20] United Nations (2008). Principles and Recommendation for Population and Housing CensusStatistical Papers. Series M No 67/ Rev.2.p8.
[21] Yang, C. Akers, V and Barcelon, D (2008).Application of Data-mining.ESRI Press: New Delhi.