



A MULTIDIMENSIONAL LOOK ON THE GROWTH OF POPULATION IN ASIA THROUGH DATA MINING

Ma. Victoria D. Naboya
Leyte Normal University
Tacloban City

Article DOI: <https://doi.org/10.36713/epra3963>

ABSTRACT

The world population is at a critical turning point. Its increasing population is eating away the earth itself wherein its impact has been sufficient to make permanent changes in the environment. Asia is the largest continent in the world, both in terms of area and population that was basically the reason why this study was conducted. The main purpose of this study is to determine what causes the growth of population in it. There are many factors which affect the growth of human population in Asia. These include the geographic, demographic and socio-economic factors. This study employs the exploratory data analysis or data mining which is a statistical procedure for exploring data sets and for formulating theory on the multidimensional look on the growth of population in Asia. The study revealed that population growth in Asia was largely affected by these factors specifically its land area, fertility rate and population literacy of the country.

KEYWORDS: Asia, data mining, demography, geography, literacy, population, population growth.

INTRODUCTION

The continent of Asia is the world's largest and most populous continent with over 4 billion people. Asia contains around 30% of the world's land area and 60% of the world's population (World Population Statistics, 2013). It is located primarily in the eastern and northern hemispheres and covers 8.7% of the Earth's total surface area and comprises 30% of its land area. With approximately 4.3 billion people, it makes up 60% of the world's current human population.

Asia's growth rate is very high for the modern era and has quadrupled during the last 100 years. Asia's wealth of natural resources, ecological variability, and biological diversity put it in a great position to support such a high growth rate (Realuyo, 2000). Asia's population growth and quickly growing economic development efforts are threatening the region's rich and limited resources through their expansion and intensification of agriculture, uncontrolled growth of industrialization, destruction of natural habitats, and urban sprawl. Asian populations will continue to grow for many decades to come which will lead to increasing pressure on the region's natural resources (Lee, 2008).

With this scenario, this study is conducted to identify the factor that possibly affects its population growth. There are various geographic, demographic and socio-economic factors affecting the population growth that were identified from the various study all over the world but this is more focused on what factors really are applicable in Asia.

OBJECTIVE

This study is geared towards the exploration of statistical procedures for data mining of the variables that contribute to the growth of population in Asia given the various data sets from its data bank.

METHODOLOGY

This paper employs the exploratory data analysis or data mining which is a statistical procedure for exploring data sets and for formulating theory on the multidimensional look on the growth of population in Asia. This study used the Association Rule discovery to determine the occurrences of the various item sets. The variables included in the study are the land area in km², climate, birth rate, fertility rate, mortality rate, net migration rate, and population literacy.

We look at these variables that would relate to the growth of population among the forty-eight (48) countries in Asia namely:

X₁: Land Area in km² X₅: Mortality rate
X₂: Climate X₆: Net Migration rate
X₃: Birth rate X₇: Population literacy
X₄: Fertility rate

The profile among Asian countries is shown in the table below.



Table 1. Profile of the Countries in Asia

	COUNTRIES	REGION	AREA (km ²)	CLIMATE CODE	BIRTH RATE	FERTILITY RATE	MORTALITY RATE	NET MIGRATION	POPULATION LITERACY
1	Russia	1	17098242	1	11	2	13	0	99.7
2	Bangladesh	2	147570	2	25	3	51	-0.02	57.7
3	Bhutan	2	38394	4	20	2	44	0	52.8
4	India	2	3287263	2	22	3	48	-0.05	74.4
5	Maldives	2	298	2	15	2	27	-12.67	99.0
6	Nepal	2	147181	4	23	2	45	3.71	66.0
7	Pakistan	2	796095	3	28	3	63	-1.69	55.0
8	Sri Lanka	2	65610	2	16	2	10	-1.54	91.2
9	Afghanistan	3	647500	3	45	6	149	-1.83	28.1
10	Armenia	3	29743	1	13	1	19	-5.88	99.6
11	Azerbaijan	3	86600	1	18	2	51	0	99.8
12	Bahrain	3	760	3	17	2	10	13.6	94.6
13	Georgia	3	69700	2	11	1	15	-3.25	99.7
14	Iran	3	1648195	3	17	2	42	-0.08	85.0
15	Iraq	3	438317	3	30	4	42	0	80.2
16	Israel	3	20770	3	20	3	4	1.68	97.1
16	Jordan	3	89342	3	20	3	16	17.22	93.4
18	Kuwait	3	17818	1	22	3	8	-1.11	94.0
19	Lebanon	3	10400	3	17	2	16	83.82	89.6
20	Oman	3	309500	3	35	3	15	-0.45	81.4
21	Qatar	3	11586	3	16	2	12	27.35	96.3
22	Saudi Arabia	3	2149690	3	29	3	16	-0.59	86.6
23	Syria	3	185180	3	26	3	16	-113.51	79.6
24	Turkey	3	783562	4	19	2	24	0.46	95.3
25	United Arab Emirates	3	83600	3	16	2	12	13.58	77.9
26	Yemen	3	527968	3	42	4	55	2.61	63.9
27	Kazakhstan	4	2724900	1	17	3	24	0.42	99.5
28	Kyrgyzstan	4	199951	1	23	3	29	-6.16	98.7
29	Tajikistan	4	143100	4	27	3	39	-1.17	99.7
30	Turkmenistan	4	488100	1	20	2	42	-1.95	98.8
31	Uzbekistan	4	447400	1	18	2	22	-2.46	99.3
32	China	5	9596961	4	14	2	16	-0.32	95.1
33	Japan	5	377915	4	8	1	3	0	99.0
34	Mongolia	5	1564116	1	21	2	37	-0.85	97.4
35	North Korea	5	120538	1	15	2	4	-0.04	100.0
36	South Korea	5	100210	4	9	1	27	0	97.9
37	Brunei	6	5765	2	18	2	12	2.47	95.4
38	Burma	6	676578	4	17	2	49	-0.3	92.7
39	Cambodia	6	181035	2	26	3	55	-0.32	73.9
40	East Timor	6	14874	2	26	5	38	0	58.6
41	Indonesia	6	1904569	2	19	2	28	-1.18	92.8
42	Laos	6	236800	2	34	3	59	-1.1	73.0
43	Malaysia	6	329847	2	22	3	15	-0.34	93.1
44	Philippines	6	300000	2	26	3	19	-1.23	95.4
45	Singapore	6	697	2	9	1	2	14.55	95.9
46	Taiwan	6	36193	2	9	1	5	0.9	98.3
47	Thailand	6	513120	2	13	2	16	0	93.5
48	Vietnam	6	331212	4	16	2	21	-0.32	94.0
MEAN		102051.9	2.45	20.417	2.4375	28.854	0.4575	87.081	

Source: CIA World Fact book

The converted data should be determined to present the derived association rules. These values should be determined to establish the association rules that if it is more than or equal the mean, the item set is assigned to 1 and if below or less than, 0 is being assigned to the item set meaning that it has no association to the various variables given.

Given the set of different countries in Asia, we find rules that will predict the occurrence of an item based on the occurrences of other items among the variables that possibly affect the growth of Asia. The implications of these item set occurrences means co-occurrence not causality which shows that we just count the frequency of the appearance of the item set of variables within a particular country in Asia.

In determining the frequent item set which is a collection of one or more items, we determine the support count. The support count is determined by the frequency of an item set. Support is determined based on the fraction of countries in Asia that contains an item set or variables which possible affect its population growth. The frequent item set is an item set whose support is greater than or equal to a minsup threshold. The goal here is to find all rules having support \geq minsup threshold and confidence \geq minconf threshold.

In this study the formula used to determine the support is:

$$s = \frac{\text{frequency of variable in the itemset}}{\text{number of Asian countries}}$$

From the data, Brute-force approach is used where we all possible association rules exist and then compute the support and confidence for each rule. If the value is lower than 25, prune rules will then be applied since it fails the minsup and minconf thresholds of the association rules. The pruning means that those item sets identified if they are less than 25, we eliminate those variables which do not have bearing on the population growth of Asia.

HYPOTHESES

1. The growth of population in Asia is affected by its geographic characteristics.
2. Demographic profile of the countries affects its population growth.
3. Socio-economic characteristics of the Asian countries influence its population growth.



4. The growth of population in Asia is affected by its geographic, demographic and socio-economic factors.

RESULTS AND DISCUSSION

To determine the items sets and its support, the mean was determined for each variable found in Table 3 affecting the countries in Asia to explore the possibility of determining the association rule of what cause the growth in population.

Table 2. Mean of the Variables from the Asian countries

VARIABLES	MEAN
Land Area (km ²)	102051.9
Climate (code)	2.45
Birth rate	20.417
Fertility rate	2.4375
Mortality rate	28.854
Net Migration rate	0.4575
Population Literacy	87.081

From the mean shown in Table 2, this was used to discover the Association rules namely:

- X₁: Put a 1 if land area is more than 102051.9 km²
- X₂: Put a 1 if its climate exceeds 2.45
- X₃: Put a 1 if its birth rate exceeds 20.417
- X₄: Put a 1 if its fertility rate exceeds 2.4375
- X₅: Put a 1 if its mortality rate exceeds 28.854
- X₆: Put a 1 if its net migrations rate exceeds 0.4575
- X₇: Put a 1 if its population literacy exceeds 87.081

From these, the association rules were shown in Table 3.

Table 3. Converted Data from the Derived Association Rules

	X ₁ Land Area	X ₂ Climate	X ₃ Birth rate	X ₄ Fertility rate	X ₅ Mortality rate	X ₆ Net Migration rate	X ₇ Population literacy	ITEM
1	1	0	0	1	0	0	1	land area, fertility rate, population literacy
2	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate
3	0	1	1	1	1	0	0	climate, birth rate, fertility rate, mortality rate
4	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate
5	0	1	0	1	0	0	1	climate, fertility rate, , population literacy
6	1	1	1	1	1	1	0	land area, climate, birth rate, fertility rate, mortality rate, net migration rate
7	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate
8	0	1	0	1	0	0	1	climate, fertility rate, population literacy
9	1	0	1	1	1	0	0	land area, birth rate, fertility rate, mortality rate
10	0	0	0	0	0	0	1	population literacy
11	0	1	0	1	1	0	1	climate, fertility rate, mortality rate, population literacy
12	0	1	0	1	0	1	1	climate, fertility rate, net migration rate, population literacy
13	0	1	0	0	0	0	1	climate, population literacy
14	1	1	0	1	1	0	0	land area, climate, fertility rate, mortality rate
15	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate
16	0	1	1	1	0	1	1	climate, birth rate, fertility rate, net migration rate, population literacy
17	0	0	1	1	0	1	1	birth rate, fertility rate, net migration rate, population literacy
18	0	1	1	1	0	0	1	climate, birth rate, fertility rate, population literacy
19	0	1	0	1	0	1	1	climate, fertility rate, net migration rate, population literacy
20	1	1	1	1	0	0	0	land area, climate, birth rate, fertility rate



21	0	1	0	1	0	1	1	climate, fertility rate, net migration rate, population literacy
22	1	1	1	1	0	0	0	land area, climate, birth rate, fertility rate
23	1	1	1	1	0	0	0	land area, climate, birth rate, fertility rate
24	1	1	0	1	0	1	1	land area, climate, fertility rate, net migration rate, population literacy
25	0	1	0	1	0	1	0	climate, fertility rate, net migration rate
26	1	0	1	1	1	1	0	land area, birth rate, fertility rate, mortality rate, net migration rate
27	1	0	0	1	0	0	1	land area fertility rate, population literacy
28	1	1	1	1	1	0	1	land area, climate, birth rate, fertility rate, mortality rate, population literacy
29	1	0	1	1	1	0	1	land area, birth rate, fertility rate, mortality rate, population literacy
30	1	0	1	1	1	0	1	land area, birth rate, fertility rate, mortality rate, population literacy
31	1	1	0	1	0	0	1	land area, climate, fertility rate, population literacy
32	1	1	0	1	0	0	1	land area, climate, fertility rate, population rate, population literacy
33	1	0	0	0	0	0	1	land area, population literacy
34	1	0	1	1	1	0	1	land area, birth rate, fertility rate, mortality rate, population literacy
35	1	1	0	1	0	0	1	land area, climate, fertility rate, population literacy
36	0	1	0	0	0	0	1	climate, population literacy
37	0	1	0	1	0	1	1	climate, fertility rate, net migration rate, population literacy
38	1	1	0	1	1	0	1	land area, climate, fertility rate, mortality rate, population literacy
39	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate,
40	0	1	1	1	1	0	0	Climate, birth rate, fertility rate, mortality rate
41	1	1	0	1	1	0	1	land area, climate, fertility rate, mortality rate, population literacy
42	1	1	1	1	1	0	0	land area, climate, birth rate, fertility rate, mortality rate
43	1	1	1	1	0	0	1	land area, climate, birth rate ,fertility rate, population literacy
44	1	1	1	1	0	0	1	land area, climate, birth rate, fertility rate, population literacy
45	0	1	0	0	0	1	1	climate, net migration rate, population literacy
46	0	1	0	0	0	1	1	climate, net migration rate, population literacy
47	1	1	0	1	0	0	1	land area, climate, fertility rate, population literacy
48	1	1	0	1	0	0	1	land area, climate, fertility rate population literacy

From the association rules derived, we now determine the frequency of the occurrences of the item set. Since there are seven (7) variables identified, there are seven probable item sets that can be determined. The minimum support was determined by looking at the

total number of countries in Asia. There are 48 countries in Asia therefore the minimum support is 25. Since it is 50% + 1. Below is the result of the occurrences of the item sets based on the minimum support of the cases.



Table 4. Item 1 Set

Variable	Support	Decision
Land Area	30	Accept
Climate	38	Accept
Birth rate	22	Eliminate since it is less than 25 minimum support.
Fertility rate	42	Accept
Mortality rate	19	Eliminate since it is less than 25 minimum support.
Net Migration rate	11	Eliminate since it is less than 25 minimum support.
Population literacy	32	Accept

Table 4 shows the variables that were explored that shows bearing to the growth of population in Asia.

Those variables accepted were the land area, climate, fertility rate and population literacy.

Table 5. Item 2 Set

Variable	Support	Decision
Fertility rate, Population literacy	22	Accept

Table 5 shows the relationship between two variables in the growth of population that was accepted. The rest were not shown on the table since they are less than the minimum support for such factor to be accepted. The same thing happened to the Item 3, Item 4, Item 5, Item 6 and Item 7 sets were not anymore shown since all of them were eliminated since the association rule should be more than or equal to the minimum support.

The study then accepts the hypothesis that the growth of population in Asia is affected by its geographic, demographic and socio-economic factors. Land area is under the geographic factor, then fertility rate under the demographic characteristics and the population literacy as a component of the country's socio-economic factor. Therefore, from the exploration of possible associated which has bearing to the growth of population in Asia, there were three variables associated to the growth of population in Asia namely, land area, fertility rate, and population literacy. From this study, the better is the population's literacy; the lesser is the growth of population as seen from the association. Likewise, it was also observed that the greater the land area with high fertility rate of the country, the higher would be the growth of the population.

CONCLUSION

From this study, the growth of population in Asia is attributed to its land area, fertility rate and the population literacy.

LITERARURE CITED

1. Realuyo, Z.M., & Del Pan-Ilarde, I. (2000). *Basic Concepts in Earth Science*. JMC Press, Inc.
2. Sergio J. Lee & Myrna Ano. 2008. *Environmental Science: The Economy of Nature and Ecology of Man*. C & E Publishing House Inc.
3. *World Population Statistics, 2013* Retrieved October 24, 2014 at <http://www.worldpopulationstatistics.com/asia-population-2013/>