Population Distribution Pattern of 76 Provinces in Thailand: Application of Factor Analysis

Binita K. Paudel[1,2], Apiradee Lim[1] & Nurin Dureh[1]

[1] Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani, Thailand

[2] Nepal Institute of Health Science, Jorpati, Kathmandu, Nepal

Correspondence: Apiradee Lim, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Muang, Pattani, 94000, Thailand. Tel: 66-081-957-7625. E-mail: lapirade@bunga.pn.psu.ac.th; api_45@hotmail.com

Abstract

Thailand is in the demographic transition phase. The shape of population pyramid is shifting from stationary to contracting pattern. Age-sex distribution may vary by province. This study explores and describes the population distribution pattern of 76 provinces in Thailand using data from 2000 Thai population census. Factor analysis, a multivariate statistical method, was used to cluster provinces, based on pattern of age-sex distribution of the population. The study found three distinct patterns of population distribution in Thailand. Twenty-seven southern and northeastern region provinces, mainly bordering Myanmar, Cambodia or Malaysia, share the classical pattern of population distribution. The majority of central region provinces, and also Phuket from the south share a similar population distribution pattern which peaks at the young age group. So too, most of the northern region provinces share another pattern that dips at the young age group. In conclusion, this study found that population distribution is not symmetrical across Thailand. The factor model approximated well this variation and clustered the provinces in three patterns. The method applied in this study is straightforward and can be used in future demographic studies.

Keywords: factor analysis, multivariate, pattern, population distribution, province, Thailand

1. Introduction

Population is the function of key demographic variables that are fertility, mortality and migration. Population Reference Bureau defines population distribution as the patterns of settlement and dispersal of a population. It is an undeniable fact that developed and developing countries have different types of population distribution patterns (Cohen, 2003). Developing countries mainly have the classical pattern in which the number of children is high and the skew is towards young ages, whereas in developed countries the skew tends to be towards older ages (Abbasi-Shavazi, 2011). African countries, affected from HIV/AIDS epidemic, have dip among young age group (Zimmer, 2009). According to world population data sheet (2012), such differences may exist within a country. Even in United States, population change and age-sex structure varies widely within states.

In Thailand, fertility was high until 1970. Then it moved into a decline phase. After 1990 it was in a low fertility phase for 6-7 years. Now it is in the phase of 'below replacement' (Prasartkul, Patama, & Varachai, 2011). Mortality is also in a decreasing trend; infant mortality is decreasing at a slow pace. According to the CIA the World Factbook 2012, the current net migration rate is zero but the internal migration within the country, from rural province to urban and industrial province, always remained substantial (Thailand Migration Report, 2011; Guest et al., 1994). Similarly, the HIV epidemic in Thailand is yet another important event that affects the population distribution. It was estimated that the number of deaths from AIDS before the year 2000 was 550 000 (Surasiengsunk et al., 1998). AIDS was the leading cause of male deaths and the second leading cause of female deaths amounting to 16.5% and 6.3 % respectively, of total deaths in 1999 (Porapakkham et al., 2010). The provinces in Northern Thailand, adjacent to Myanmar and Laos, were greatly affected by this epidemic (Surasiengsunk et al., 1998).

As already mentioned in the first paragraph, the age - sex structure may vary within the country. There are 5 different regions and 76 provinces in Thailand. Each province is geographically different. Also, types of people, their religion, culture belief and health related practices are different. One can predict many differences in the pattern of population distribution in the different provinces in Thailand. One also can predict that some of these provinces may follow the same pattern. In order to find the evidence of the above predictions, this study applied a statistical method called "factor analysis". Factor analysis was invented more than 100 years ago by psychologist Charles Spearman (1904) and has been used, since, mostly in psychological studies. This method also has been applied widely in different areas, for example, the assessment of water quality (Liu, Kao, & Kuo, 2003), in ecological data (Rittibon, Tongkumchum, & Karntanut, 2012). There is a scope of using factor analysis in demographic research. Carey (1966) had also applied this method to interpret the population and housing pattern.

This study aimed to explore and describe the different patterns of population distribution in each province of Thailand by applying factor analysis. Evidence based information on patterns of population distribution is very important to plan and implement programs related to fertility, mortality and migration. Allocation of the different resources also depends on the population distribution in Thailand. After comparing the pattern of changes with the recent findings of the 2010 census, the researchers trust that the findings will provide fresh insight. Understanding the many details on the pattern of distribution will be helpful in future projections of population growth or decline in Thailand.

2. Methods

2.1 Data Source

The population data were retrieved from the website of National Statistical Office, Ministry of Information and Communication Technology, Thailand. The original data table contains 76 rows (provinces) and 36 columns (by 5 years interval age group and sex).

2.2 Statistical Methods

Factor analysis, a multivariate statistical method, was used to cluster the provinces, which were based on age-sex structure of the population. Since our objective was to cluster the provinces, they have to be considered as variables or outcomes so the original data table needs to be transposed. In this case there will be 76 columns (as outcomes) and 36 rows (as subjects) in the new table. Doing the factor analysis, the data table needs to have many more rows than columns. Therefore, we have to add more rows by extending the population into single-year age groups. For this purpose, the natural cubic spline was applied (McNeil, Odton & Ueranantasun, 2011). This method interpolated the data for single-year ages up to 105 of which 90 for each sex were used to create a data table with 76 columns and 180 rows for factor analysis. So the new data matrix has 76 columns corresponding to the provinces and 90 x 2 = 180 rows corresponding to a single year age population of males and females. We constructed 76*76 covariance matrix by using Spearman rank-order correlation method for handling non-linear relationships between province variables.

2.2.1 Factor Model and Factor Loadings

If y_{ij} is the outcome in row i and column j of the r x c matrix data array, the factor model is formulated as

$$y_{ij} = \mu_j + \sum_{k=1}^{p} \lambda_j^{(k)} f_i^{(k)} \tag{1}$$

Where, the p column vectors $f^{(k)}$ in this model are called common factors and the p row vectors $\lambda^{(k)}$ are called their loadings. The factor loadings obtained from factor analysis were used to determine the correlation between provinces and common factors.

2.2.2 Extraction of Factors and Factor Rotation

Factor analysis, in this study, used maximum likelihood method to extract the appropriate number of factors (Costello & Osborne, 2005). Three factors were remained based on the non-significant p-value (p>0.05) of Chi-squared test. To obtain a clearer pattern and interpretable result, the provisional factors were transformed in order to find the new factors that are easier to interpret. The factor rotation can be orthogonal or oblique. The varimax, quartimax, and equamax are commonly available orthogonal methods of rotation; direct oblimin, quartimin, and promax are oblique. The fit of the model is unchanged by rotating the factors (Johnson & Wichern, 1988). The only desirable element in selection of type of rotation is that the factor loading should be either close to zero or very different from zero, so that the result will be clear and interpretable (Manly, 1994). In this study, "Promax" rotation provided the clearer pattern.

The factor model also gives the "uniquenesses" corresponding to each province, for which values close to 1 provide evidence that they cannot be associated with any factor, and thus should be omitted from the factor model. In this study, the uniqueness ranges 0.005 to 0.060 so no evidence emerged to omit any province from the factor model. Therefore, all the variables (provinces) were included in the 3 factor model. The factor loading higher than 0.57 were considered as a significant level of a pure factor. For the variables (provinces) distributed into two or more factors, factor loadings between 0.33 - 0.56 were considered to indicate mixed factors. Data management and analysis was done by using R: A language and environment for statistical computing (R Core Team, 2013)

3. Results

The results of this study are organized in three different headings: loadings from factor analysis, interpretation of factors and regional variation in the population distribution patterns.

Table 1. Loadings obtained from three factor model for Thai province gender-age population at the 2000 census

Name	ID	Factor 1	Factor 2	Factor 3	Name	ID	Factor 1	Factor 2	Factor 3
Pattani	94	**0.989**			Petchabun	67	**0.434**	0.125	**0.496**
Narathiwat	96	**0.989**			Nakhon Nayok	26	**0.382**	0.209	**0.458**
Mae Hong Son	58	**0.982**			Nong Khai	23	**0.362**	0.252	**0.437**
Yala	95	**0.977**			Loei	42	**0.343**	0.247	**0.468**
Tak	63	**0.974**			Bangkok	10	-0.112	**0.960**	0.151
Ubon Ratchathani	34	**0.973**			Samut Prakan	11		**0.940**	0.124
Buri Ram	31	**0.972**			Samut Sakhon	74		**0.927**	
Krabi	81	**0.950**	0.101		PathumThani	13		**0.926**	
Si Sa Ket	33	**0.946**		0.160	Chon Buri	20		**0.905**	
Satun	91	**0.939**	0.114		Ayutthaya	14	0.122	**0.872**	
Surin	32	**0.938**		0.162	Phuket	83	0.109	**0.868**	
Nakhon Phanom	48	**0.925**			Nakhon Pathom	73		**0.863**	
Roi Et	45	**0.902**		0.114	Rayong	21	0.161	**0.852**	
Amnat Charoen	37	**0.897**		0.171	Nontha Buri	12	-0.168	**0.849**	0.329
Mukdahan	49	**0.895**		0.103	Chachoengsao	24	0.250	**0.798**	
Yasothon	35	**0.885**		0.224	Saraburi	19		**0.759**	0.196
Nakhon Si Thamarat	80	**0.885**			Ratchaburi	70	0.154	**0.666**	0.219
Trang	92	**0.862**		0.129	Prachuap	77	0.284	**0.652**	0.102
Nong Bua Lam Phu	39	**0.858**	0.176		Phetchaburi	76	0.219	**0.586**	0.238
Phatthalung	93	**0.822**		0.162	Chanthaburi	22		**0.614**	0.398
Sa Kaeo	27	**0.820**	-0.108	0.312	Samut Songkharam	75		**0.588**	0.401
Phangnga	82	**0.773**		0.227	Chaing Mai	50	-0.168	**0.467**	**0.704**
Maha Sarakham	44	**0.702**	0.128	0.214	Lopburi	16	0.162	**0.372**	**0.504**
Kalasin	46	**0.702**	0.157	0.183	Lamphun	51	-0.201	**0.355**	**0.842**
Nakhon Ratchasima	30	**0.696**	0.108	0.240	Phrae	54			**0.993**
Kanchanaburi	71	**0.684**	0.192	0.162	Chai Nat	18			**0.934**
Khon Kaen	40	**0.573**	0.311	0.167	Phayao	56			**0.923**
Sakon Nakhon	47	**0.754**	**0.339**		Lampang	52		0.119	**0.922**
Nong Khai	43	**0.736**	**0.418**	-0.137	Sing Buri	17			**0.916**
Udon Thani	41	**0.669**	**0.463**	-0.107	Uttaradit	53		0.188	**0.865**
Songkhla	90	**0.616**	**0.509**	-0.113	Phichit	66	0.319	-0.170	**0.856**
Surat Thani	84	**0.524**	**0.489**		Chaing Rai	57		0.107	**0.823**
Ranong	85	**0.515**	**0.612**	-0.113	Sukhothai	64	0.120	0.104	**0.809**
Chumphon	86	**0.473**	**0.360**	0.219	Ang Thong	15	0.215		**0.763**
Prachin Buri	25	**0.352**	**0.555**	0.128	Nan	55	0.223		**0.742**
Chiyaphum	36	**0.635**		**0.372**	Nakhon Sawan	60	0.222	0.106	**0.718**
Kamphaeng Phet	62	**0.619**		**0.500**	Phitsanulok	65	0.158	0.243	**0.646**
Uthai Thani	61	**0.470**		**0.633**	Suphan Buri	72	0.266	0.175	**0.606**
					SS loadings		26.07	14.22	14.06
					Proportion Variance		0.34	0.19	0.19
					Cumulative Variance		0.34	0.53	0.72

Note: - Cutoff value is >0.33, single loading >0.33 is considered as pure factors and others are associated with mixed factor; Loadings below 0.1 are not shown and loadings that exceed 0.33 are in bold font

3.1 Loadings from Factor Analysis

Table 1 presents the factor loading of the 76 provinces of Thailand. This factor loading reflects the correlation between province variable and common factors. Three Factor model was best fitted (p>0.762) with the data and it explains 72% of the total provinces variation. Unique differences range from 0.005 to 0.060. Based on the cutoff value, the study found that 27, 15 and 14 provinces correlated purely to factor 1, 2 and 3 respectively. Twenty provinces were found to correlate with two factors. In this study, three different factors represent the 3 different patterns of population distribution which will be described in detail below.

3.2 Interpretation of Three Factors from the Model
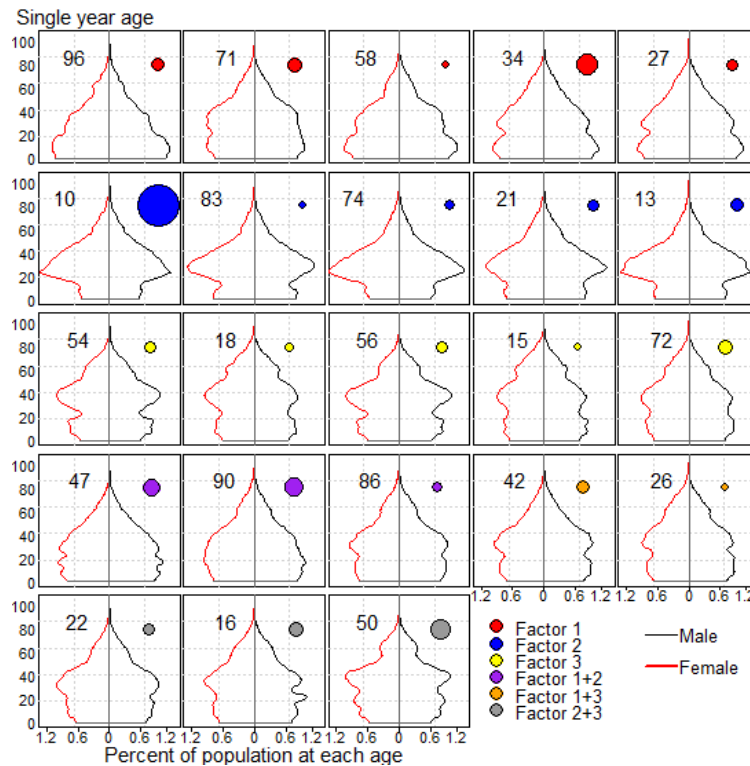
Figure 1. Spline-smoothed single year age and sex distribution from selected Thai provinces at 2000 Census
Note: The area of the bubble in each province denotes its population size
The integers at the top left of each graph indicates the province ID, full name is in Table 1

Figure 1 presents the Spline-smoothed population distribution by single year age and sex from selected Thai provinces based on the 2000 Census. Five representative graphs of the provinces associated purely with three factors and eight representative graphs of the provinces associated with mixed factors were selected and arranged in rows. The first factor, at the first row of the graph, shows a traditional pattern of population distribution characterized by a young age structure. In this pattern, each new cohort is larger, so the shape is like an exponential decay. The second factor, at the second row of the graph, shows the rapid declined of the population then leveled-off in the recent years. The third factor, at the third row of the graph, shows the fluctuating or complicated trends in the population distribution. This pattern includes the population dip between ages 20 and 40. It is interesting to note that the short decline at the beginning of the graphs shows that the fertility was starting to decrease yearly in some provinces, whereas the provinces associated with the second population distribution pattern had already leveled off followed by a rapid decline.

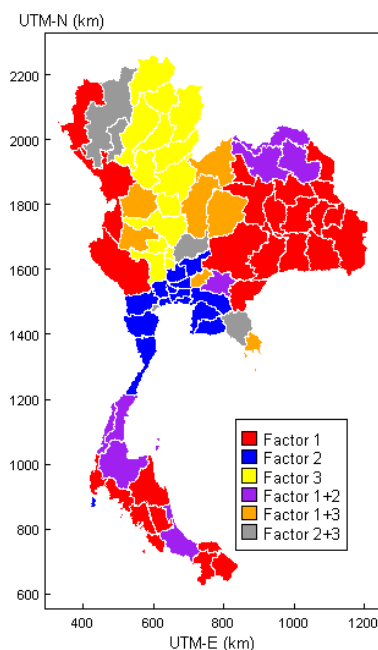*3.3 Regional Variation in the Population Distribution Pattern*

Figure 2. Thematic map of Thailand shows the regional variation in population distribution pattern in each province presented in colors from factor analysis of gender-age population at the 2000 census

Figure 2 shows the regional variation in population distribution pattern; denoted in six different colors from factor analysis. The majority of the provinces in the south and northeastern regions, mainly in the border provinces of Myanmar, Cambodia and Malaysia, shown in red color, have the same dominant traditional pattern of population distribution. The Bangkok and surrounding provinces share a similar pattern of distribution which was peaked and then stabilized, shown in blue color. The central plain, north of Bangkok contains the provinces with the third pattern, with substantial dips in the populations aged between 20 and 40.

4. Discussion

4.1 Application of Factor Analysis

Factor analysis, a multivariate statistical method, is well established and has been used widely (Costello & Osborne, 2005) not only for analyzing data from psychology, but also with data from different field, including water quality assessment (Liu, Kao, & Kuo, 2003; Lueangthuwapranit, Sampantarak, & Wongsai, 2011) and classification of the species of different birds (Rittibon, Tongkumchum, & Karntanut, 2012). The method is mainly used to group and describe the set of variables for further analysis. The factor analysis was also used as an intermediate analysis to determine a social-demographic poverty index (Latifa, Aswatini, & Romdiati, 2008). In this study, the factor analysis was used to model the variation in age-sex structure of the population. The factors were interpreted as patterns and further used as a basis for clustering provinces. Since this method clustered provinces very well, it is believed that this method will have potential wide application in the field of other scientific studies including population studies.

4.2 Patterns of Distribution

The single year age population distribution was found non-symmetrical in 76 provinces of Thailand. This study found three distinct patterns of population distribution. Each pattern has different characteristics of the distribution. The research brought up three principal issues on population distribution and related demographic factors; namely, fertility, mortality and migration.

The general fertility of Thailand has started to decline since the late 1960s. Now it is at the level of 'below replacement' (Prasartkul et al., 2011). The first issue, raised in this study, is why the majority of the bordering provinces in Southern and Eastern and Northern regions have a classical pattern of population distribution. The population in these provinces has recently begun to decrease, whereas the young age population is still high. This type of population distribution pattern is also called "a young age population structure" and will

probably experience further population growth (Abbasi-Shavazi, 2011). This finding is also supported by the findings of a year 2000 census related to the average size of the household in Thailand. The Northeast and the South had the largest average size of household. In the Thailand Census-2000, looking at all provinces, Pattani had the largest size of household (4.8 persons), followed by Narathiwat (4.6 persons). The question here is: why are the statistics of these provinces so different and not undergoing similar changes to the rest of the country of Thailand?

Another issue raised from this study is the peak and steep declining pattern. Most of the provinces of central Thailand, including Bangkok, follow the same pattern of population distribution. This might be because of the fast declining fertility rate and of the slow declining infant mortality rate. Fertility in Thailand has declined steadily over the last few decades. At around 1996 it fell below replacement level (Prasartkul et al., 2011). The reason behind the rise in the young adult population may also be due to internal migration from northern regions of the country, which now displays a gap in the population (25-35 years old), to these provinces where for the same age groups there is a peak in population distribution. Guest et al. (1994) found that migration is highly selective among young adults, females and highly educated adults, and more likely to occur in urban areas or in rural to urban areas. However, after 1990, internal migration rates have steadily declined compared to past statistics (Huguet, Chamratrithirong, & Richter, 2011), the population of municipal areas reached 41.1%; this was only 31.1% in 2000 and 29.4% in 1990 (National Statistical Office Thailand-2011). Thailand is considered to be an industrialized country (NESDB-2010).

The third issue is related to missing the young age population that shows by the dipping portion of the graph as a third pattern. It is interesting to note that such a type of dip is not usual and normal. However, this study is unable to explain the causes behind this issue, based on the literature. Anyone can assume that this can be because of mortality and migration functions of the population. Some part of the gap might be because of the internal migration (Guest et al., 1994) as discussed above and some part might be because of the HIV epidemic (Weniger et al., 1991). Most provinces in the Northern Region, adjacent to Myanmar and Laos have this pattern were affected by this epidemic (Surasiengsunk et al., 1998). AIDS was the first leading cause of death among males and the second leading cause of death among females; this accounted the 16.5% and 6.3 % respectively of total deaths in 1999.

5. Conclusions

While the population distribution in Thailand in 2000 was found to be varied a lot between provinces, the pattern in each province was approximated well by using a factor model. It is suggested that this detail and empirical information on population distribution, showing its changing patterns together, are very useful data for the planning and implementation of programs related to fertility, mortality and migration. Despite the fact that the data of this study is from the 2000 census, this study could be valuable in the understanding of past, present and possibly future population distribution trends. The method applied in this study, factor analysis, is straightforward and can be widely used in future demographic studies.

Acknowledgements

References

Abbasi-Shavazi, M. J. (2001). *Recent changes and the future of fertility in Iran* (PDF document). Retrieved from http://www.un.org/esa/population/publications/completingfertility/2RevisedABBASIpaper.PDF

Carey, G. W. (1966). The regional interpretation of Manhattan population and housing patterns through factor analysis. *Geographical Review, 56*, 551-569.

Cohen, J. E. (2003). Human Population: The Next Half Century. *Science, 302*(5648), 1172-1175. http://dx.doi.org/10.1126/science.1088665

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research and Evaluation, 10*(7). Retrieved from http://pareonline.net/pdf/v10n7.pdf

Guest, P., Chamratrithirong, A., Archavanitkul, K., Piriyathamwong, N., & Richter, K. (1994). Internal migration in Thailand. *Asian and Pacific Migration Journal, 3*(4), 531-545.

Huguet, J. W., Chamratrithirong, A., & Richter, K. (2011). Thailand Migration Profile. In J. W. Huguet, & A. Chamratrithirong (Eds.), *Thailand Migration Report-2011* (pp. 9-15). Bangkok: International Organization for Migration,Thailand.

Johnson, R. A., & Wichern, D. W. (1988). *Applied multivariate statistical analysis* (2nd ed.). New Jersey: Prentice Hall International, Inc.

Latifa, A., Aswatini, & Romdiati, H. (2008). Population and Social Demographic Poverty: a Case Study in the Border Areas of East Kalimantan and North Sulawesi. *Journal of Indonesian Social Sciences and Humanities, 1*, 175-191.

Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *The Science of The Total Environment, 313*, 77-89.

Lueangthuwapranit, C., Sampantarak, U., & Wongsai, S. (2011). Distribution and Abundance of Phytoplankton: Influence of Salinity and turbidity Gradients in the Na Thap River, Songkla Province, Thailand. *Journal of Coastal Research*, *27*(3), 585-594. http://dx.doi.org/10.2112/JCOASTRES-D-10-00123.1

Manly, B. F. J. (1994). *Multivariate statistical methods* (2nd ed.). London: Chapman & Hall.

McNeil, N., Odton, P., & Ueranantasun, A. (2011). Spline interpolation of demographic data revisited. *Songklanakarin Journal of Science and Technology*, *33*(1), 117-120.

National Statistical Office. (2012). *Advance Results: Population and Housing Census 2000*. National Statistical Office, Thailand.

National Statistical Office. *Data: Population and Housing Census 2000*. National Statistical Office,Thailand. Retrieved from http://web.nso.go.th/census/poph/cen_poph.htm

Nelson, K. E. (1998). The demographic impact of the HIV epidemic in Thailand (Editorial). *AIDS, 12*, 813-814.

Porapakkham, Y., Rao, C., Pattaraarchachai, J., Polprasert, W., Vos, T., Adair, T., & Lopej, A. D. (2010). Estimated causes of death in Thailand, 2005: implications for health policy. *Population Health Metrics, 8*, 14. http://dx.doi.org/10.1186/1478-7954-8-14

Prasartkul, P., Patama, V., & Varachai, T. (2011). Fertility Transition and Its Impact. In G. Jones (Ed.), *Impact of demographic change in Thailand* (pp. 17-23). Bangkok: United Nations Population Fund.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rittiboon, K.,Tongkumchum, P., & Karntanut, W. (2012). Bird Assemblages in the Thale Noi Non-hunting Area, Southern Thailand. *Journal of Sustainable Science and Management*, *7*(1), 42-48.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-293. Retrieved from http://psychclassics.yorku.ca/Spearman/chap1-4.htm

Surasiengsunk, S., Kiranandana, S., Wongboonsin, K., Garnett, G. P., Anderson, R., & Griensven, G. (1998). Demographic impact of the HIV epidemic in Thailand. *AIDS*, *12*(7), 775-784.

International Organization for Migration. (2011). *Thailand Migration Report*. International Organization for Migration, Bangkok, Thailand.

Weniger, B. G., Limpakarnjanarat, K., Ungchusak, K., Thanprasertsuk, S., Choopanya, K., Vanichseni, S., … Wasi, C. (1991). The epidemiology of HIV infection and AIDS in Thailand. *AIDS, 5*(2), 71-85.

World population data sheet 2012. *World population highlights. Population Reference Bureau.* Retrieved from http://www.prb.org/pdf12/2012-population-data-sheet_eng.pdf

Zimmer, Z. (2009). Household composition among elders in Sub-Saharan Africa in the context of HIV/AIDS. *Journal of Marraiage and Family*, *71*(4), 1086-1099. http://dx.doi.org/10.1111/j.1741-3737.2009.00654.x