Far East Journal of Mathematical Sciences (FJMS)
Volume 77, Number 1, 2013, Pages 41-54
Published Online: June 2013
Available online at http://pphmj.com/journals/fjms.htm
Published by Pushpa Publishing House, Allahabad, INDIA

A STATISTICAL MODEL FOR MEASURING SOCIAL INEQUALITY IN THREE SOUTHERNMOST PROVINCES OF THAILAND

Phattrawan Tongkumchum*, Sunaree Suwanro and Nittaya Tongrong

Department of Mathematics and Computer Science
Faculty of Science and Technology
Prince of Songkla University
Pattani, 94000, Thailand
e-mail: tphattra@bunga.pn.psu.ac.th

Abstract

Measuring social inequalities is essential. However, how to carry out these measurements is a subject of debate. This paper suggests a method for measuring social inequalities using a statistical model of non-participation either at work or at school among youth from different demographic groups. Youth non-participation is defined as not attending school and not employed in the workforce, by gender and religion (Muslim or other) in the sub-districts of Pattani, Yala and Narathiwat Provinces of southern Thailand. Data were obtained from the 2000 Population Census of Thailand. The multiplicative logistics regression model is used to enable identification of geographic areas, where a relatively high level of social inequality existed. The plotting technique used in this paper makes it easy to identify regions that need urgent action to increase social justice because of high levels of inequality.

2010 Mathematics Subject Classification: 62-XX.

Keywords and phrases: demographic factors, school attendance, disparity index.

*Corresponding author

Submitted by K. K. Azad

Received January 23, 2013; Revised April 22, 2013

1. Introduction

There are different views of measuring social inequalities. Recommendation suggests collection of multiple indicators of social position but there is limited empirical knowledge on which indicators to use. Social inequality can be measured by education, income and socioeconomic status. Numerous studies, including Cowell [4], have used incomes for measuring inequality within a society, usually justified by findings that dependent children share the socioeconomic conditions and well-being of their parents (Avramov [1], Figen [5], Buonanno and Montolio [3]). Disadvantage is often generational and Gesemann [6] concluded that education is the key to integration. Phillimore and Goodson [12] used education and employment as indicators of integration within a society. A focus on participation by young people in education and/or employment can help predict likely future levels of integration, and also probable disparity and tensions, in the community.

Although the study by Cowell [4] combined income and purchasing power, there are few other published studies using such indicators. Avramov [1] looked at income, education and housing as composite indicators of disadvantage. This comparison of one's relative condition might be relevant to perceptions and feelings about being “not at work and not at school”. However, debate has focus on the issue of how inequalities are measured. A composite school-employment index could be useful for identifying locations of social inequality in the region studied. This study used a statistical model to examine religion and gender, as well as place of residence in the area, as factors among youth aged 15-17 years in being “not at work and not at school”. The interaction between region and demographic from the model provide a measure of social inequality in the region.

2. Materials and Methods

Measurements used to measure social inequality in health science, social science, and education can be binary. For example, in our study the outcome is binary and the adverse outcome defined as giving the answer “no” to both the questions “attending school” and “employed” on the form for the 2000 Population and Housing Census of Thailand. The determinants were defined as gender, religion (Muslim or non-Muslim) and region (sub-district or aggregated sub-district) of residence.

The logistic regression provides a method for modeling the association between a binary outcome and multiple determinants (Hosmer and Lemshow [8], Kleinbaum and Klein [9]). To allow for possible interactions between region and demographic group, the multiplicative logistic model was fitted to the counts in cells defined by combinations of demographic group and region, and the adequacy of the model was assessed by comparing the residual deviance with the number of degrees of freedom, and also by examining the linearity in the plot of deviance residuals against normal quantiles (Venables and Ripley [15]). The multiplicative logistic model takes the form

$$y_{ij} = \alpha_i + \gamma_i \beta_j. \tag{1}$$

The terms α_i and β_j represent effects associated with region i and demographic group j. In this model, the demographic parameters are scaled to have unit variance as well as zero mean. The parameters γ_i provide a measure of the disparity in the adverse event rate between the different demographic groups in region i. Thus, if region i has $\gamma_i = 0,$ it means that there is no difference in the school adverse event rates between demographic groups in this region, whereas if γ_i is large in magnitude, then there is a high disparity between these groups. We call this measure the *disparity index* for the specified region.

Model (1) is non-linear and thus cannot be fitted simply using regression. However, Theil [13] showed that the least squares estimates of the β_j parameters in model (1) are the elements of the eigenvector of the matrix

$Y_c^T Y_c$ corresponding to its largest eigenvalue, where Y_c is the matrix with elements $y_{ij} - \bar{y}_i$ and Y^T denotes the transpose of *Y*. The corresponding least squares estimates of the γ_i parameters are then expressed in terms of the eigenvector components that define the *disparity index*.

$$\gamma_i = \sum_{j=1}^{4} \beta_j (y_{ij} - \bar{y}_i). \tag{2}$$

Since the vector β_j is scaled to have mean 0 and standard deviation 1, it has only two free parameters. If these parameters are regarded as fixed, then model (1) can be fitted using standard linear regression, which provides both estimates and standard errors for the remaining parameters. The total number of parameters is thus 2*n*, where *n* is the number of regions.

Model (1) thus contains a pair of parameters (α_i, γ_i) for each region, where α_i is the proportion of non-participating subjects and γ_i is the disparity index measuring the extent to which different demographic groups have different non-participation rates. This model has been used extensively for mortality forecasting in population science, where it is known as the Lee-Carter model (see, for example, Lee and Carter [10], Booth et al. [2]). In this research y_{ij} is the logarithm of the mortality rate in a population, where the indexes *i* and *j* refer to age group and year, respectively.

To allow for values of the probability of an adverse outcome p_{ij} equal to 0 or 1, a small constant *d* is added to the numerator and denominator before log-transforming. Having estimated the vector β_j in model (1) by least squares we then used logistic regression to estimate the parameters (α_i, γ_i) for each region. Sum contrasts (Venables and Ripley [15], Chapter 6) were used instead of the standard treatment contrasts, so that the standard error for each α_i parameter provides a confidence interval for the difference between the non-participation rates in region *i* and the overall mean non-participation rate in the study area.

The final results are displayed as a scatter plot of the estimated parameters (α_i, γ_i) for the regions. This plot shows the pattern of both non-participation and disparity in the study area. Using the standard errors estimated from the logistic regression model, the regions may then be classified into groups according to whether the confidence intervals for the non-participation rates exceed, contain, or fall below the overall mean, and according to whether the confidence intervals for the disparity indexes exceed zero, contain zero, or fall below zero.

3. Applications

We applied our methods to the population data selected from the 2000 Population and Housing Census of Thailand. Omitted were persons who did not state their age, and persons aged < 15 years and > 17 years, giving a total study sample of 35,022 persons in Pattani, 23,642 persons in Yala and 41,933 persons in Narathiwat. Our analysis using gender, religion (Muslim or non-Muslim) and region (sub-district or aggregated sub-district) of residence are as factors of interest.

Some sub-districts had low populations for either Muslim or non-Muslim residents. To ensure that the statistical analysis was not compromised by such small sample sizes, adjoining sub-districts were combined, where necessary to form larger regions, each with a minimum total population of approximately 1,600 persons for both Muslim and non-Muslim residents. This reduced the number of residence locations from 115 to 34, 58 to 19 and 77 to 25 in Pattani, Yala and Narathiwat, as shown in Figure 1 and Table 1.

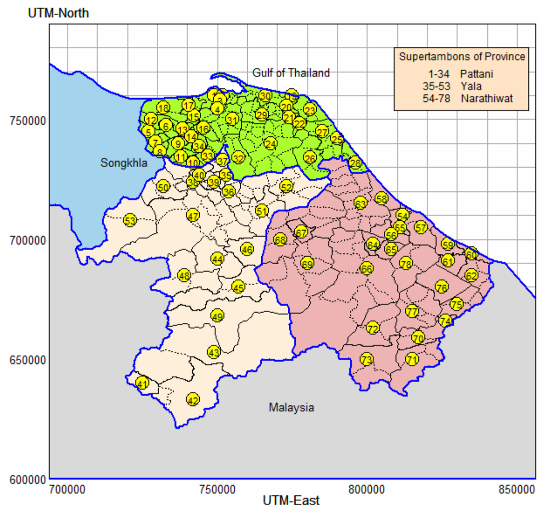

Figure 1. Map of 115, 58 and 77 sub-districts in Pattani Yala and Narathiwat, with combinations to create 34, 19 and 25 statistical regions.

Table 1. Name of statistical regions

1. Sabarang+Anoru	2. Bana	3. ChabangTiko+Talubo	4. TanyongLulo+Puyut
5. KhokPho	6. Makrut+ BangKro	7. ChangHaitok+ PaBon	8. SaiKhao
9. NaPradu	10. PakLo	11. ThungPhala	12. ThaRuea
13. Naket	14. KhuanNori	15. Tuyong+BangTawa	16. KoPo+Donrak
17. KholoTanyong+Yabi	18. BoThong+ThaKamcham	19. Panare	20. ThaKham+Bannok
21. Don	22. Khuan+ThaNam	23. KhokKrabue+PhoMing	24. Mayo+Puku
25. Taluban	26. ThungKhla+PlongHoi	27. Paen+ManangDalam	28. SaiThong+DonSai
29. Takae+Talo	30. TaloKaPo+Charang	31. Yarang+Wat	32. Krado+MoMawi
33. MaeLan+MuangTia	34. PaRai	35. Sateng	36. Budi+SatengNok
37. Tase+Yopo	38. Lidon+Yala	39. ThaSap+NaTham	40. LamMai
41. Betong	42. Yarom+ThanNamThip	43. TanoMaero+Aiyoeweng	44. BannangSata+Bacho
45. BanRae+KhueanBanLang	46. TalingChan	47. KrongPinang+Purong	48. ThanTo
49. MaeWat	50. Yaha	51. KayuBoKo+Kalupang	52. ThaThong+Kero
53. Kabang+Bala	54. BangNak	55. LamPhu	56. ManangTayo+BanPo
57. Kaluwo+KalueoNuea	58. KhokKhian	59. PhraiWan	60. Chehe+SalaMai
61. BangKhunThong	62. Phron+NaNak	63. Bacho+KayoMati	64. TanyongMas
65. TanyongLimo	66. Bo-ngo+Chanae	67. Ruso	68. Riang+KhokSato
69. Kalong+Sako	70. Waeng+Erawan	71. Kholo+MaeDong	72. Sukhirin+Kia
73. PhuhhaoThong+Mamong	74. Su-ngiKolok	75. Pasemat+Puyo	76. Su-ngaiPadi
77. Riko+SaKo	78. Chuap+Bukit		

The logistic regression models were fitted to the non-participation rates separately in each province. The model for Pattani gave a residual deviance of 136.63 with 68 degrees of freedom, and the residual plot indicated a clear departure from linearity. The least squares estimates of β_j were 1.042 for Muslim males, 0.653 for Muslim females, –0.687 for non-Muslim males and –1.008 for non-Muslim females. The model for Yala gave a residual deviance of 101.31 with 38 degrees of freedom, and the residual plot indicated a good fit. The least squares estimates of β_j were 0.824 for Muslim males, 0.904 for Muslim females, –0.795 for non-Muslim males and –0.933 for non-Muslim females. The model for Narathiwat gave a residual deviance of 86.69 with 50 degrees of freedom, and the residual plot indicated an acceptable fit. The least squares estimates of β_j were 1.052 for Muslim males, 0.595 for Muslim females, –0.520 for non-Muslim males and –1.127 for non-Muslim females. Thus, a positive disparity index for a region indicates that Muslim youths in the region had a higher non-participation percentage than the non-Muslim youths.

Figures 2, 3 and 4 show plots of observed counts and percentages against fitted values as well as residuals against normal quantiles in its upper panels, with a scatter plot of the non-participation rates versus disparity indexes for the regions in the lower panel.

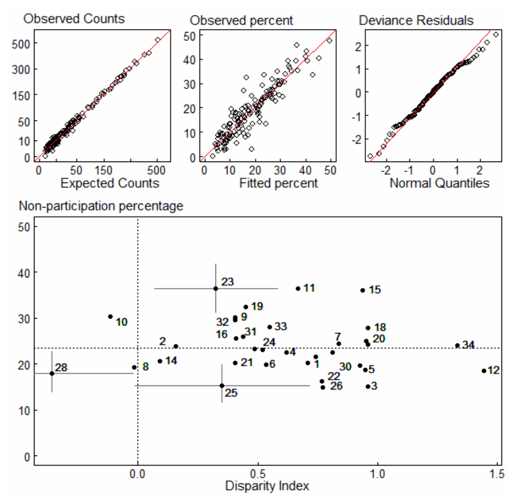

Figure 2. Results from multiplicative logistic model of youth non-participation rates in Pattani province.

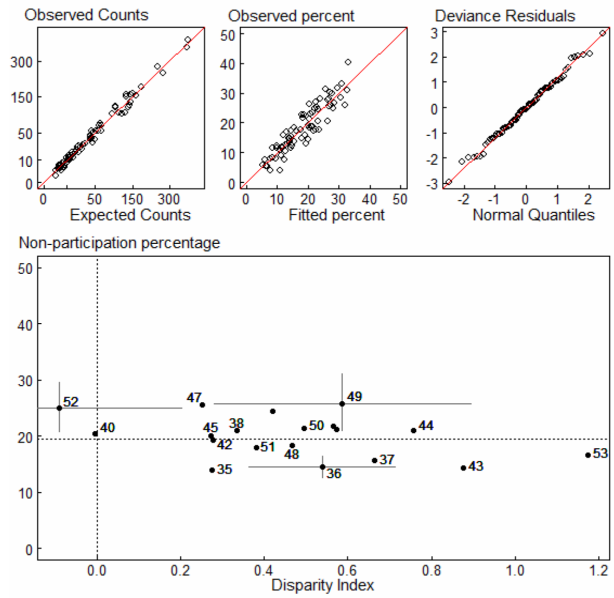

Figure 3. Results from multiplicative logistic model of youth non-participation rates in Yala province.

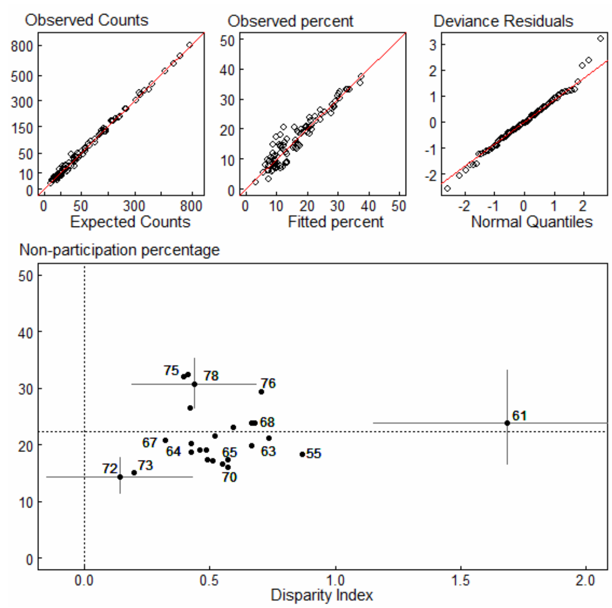

Figure 4. Results from multiplicative logistic model of youth non-participation rates in Narathiwat province.

The lower panel shows a scatter plot of the non-participation rates and the disparity indexes for 34 statistical regions in Pattani, 19 statistical regions in Yala and 25 statistical regions in Narathiwat. A score of zero for the disparity index would mean that, for the particular region, all four gender-religion groups would have the same non-participation levels, a positive value indicates greater non-participation rates for Muslim youth, and a negative value indicates greater non-participation rates for non-Muslims. On the vertical axis, a placement near the dotted line would indicate that, for that particular region, the overall level of non-participation would be close to the overall average for all regions. Regions with points above or below the dotted line have non-participation rates that are higher or lower, respectively, than average.

The plot also shows confidence intervals that can be used to make valid statistical conclusions, the vertical line segments denote a 95% confidence interval for the non-participation rates (when compared with the overall mean), whereas the horizontal line segments denote a 95% confidence interval for the disparity indexes (compared to zero disparity). To reduce clutter the confidence intervals are shown only for selected regions.

Using these confidence intervals the regions may be classified into nine groups according to whether (a) the confidence interval for the non-participation percentages is above the mean, crosses the mean, or is below the mean, and (b) whether the confidence interval for disparity index is above zero, crosses zero, or is below zero. This classification thus gives rise to nine possible clusters as shown in Table 2.

Table 2. Possible clusters with numbers of region

Clusters	Numbers of region		
	Pattani	Yala	Narathiwat
1: Higher than average non-participation, Muslim disadvantage	7	1	5
2: Higher than average non-participation, no evidence of disparity	1	2	0
3: Higher than average non-participation, Muslim advantage	0	0	0
4: Average non-participation, Muslim disadvantage	17	10	12
5: Average non-participation, no evidence of disparity	4	3	1
6: Average non-participation, Muslim advantage	0	0	0
7: Lower than average non-participation, Muslim disadvantage	3	3	5
8: Lower than average non-participation, no evidence of disparity	1	0	2
9: Lower than average non-participation, Muslim advantage	1	0	0

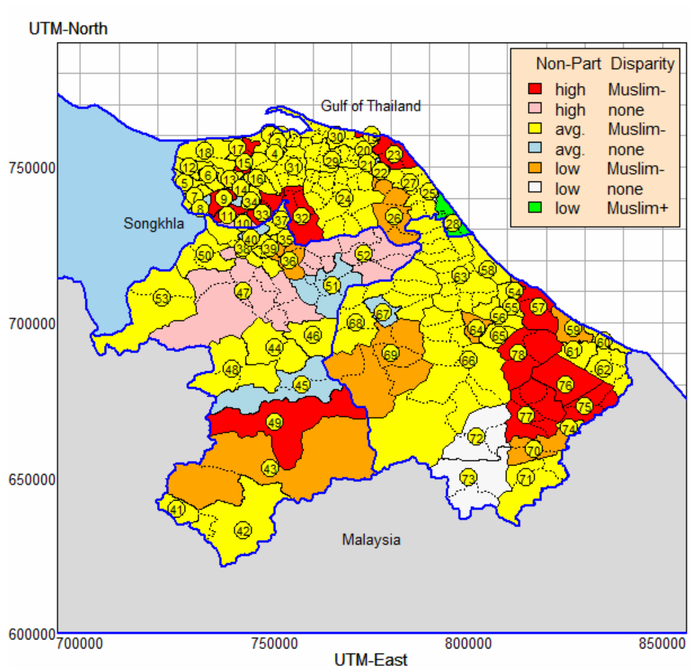

Figure 5. Map from multiplicative model of youth non-participation in Pattani Yala and Narathiwat.

Figure 5 shows corresponding thematic maps, where disparity is labeled as “Muslim-” for Muslim disadvantage and “Muslim+” for Muslim advantage.

In Pattani a high disparity occurred in NaPradu (9), ThungPhala (11), Tuyong+Bangtawa (15), Panare (19), KhokKrabue+PhoMing (23), Krado+MoMawi (32) and MaeLan+MungTia (33) with the Muslims having a higher than average non-participation, positive disparity. In PakLo (10) there is higher than average non-participation but no evidence of disparity. In contrast, the Muslims in the MaiKaen (28) have a lower than average non-participation, negative disparity.

In Yala a high disparity occurred in MeaWat (49) with the Muslim having a higher than average non-participation, positive disparity. In Krong Pinang+Purong (47) and ThaThong+Kero (52) there is higher than average non-participation but no evidence of disparity.

In Narathiwat a high disparity occurred in KaLuWo+KaLuWoNuea (57), Pasemat+Puyo (75), Su-ngaiPadi (76), Riko+SaKo (77) and Chuap+Bukit (78) with the Muslim having a higher than average non-participation, positive disparity.

4. Conclusion and Discussion

This study used statistical model that is well known in other areas (Lee and Carter [10], Booth et al. [2]) applying to social science data in order to measure social inequality in a small area. It was found that the method can be used to identify the area with social inequality exist.

We examined religion and gender, as well as place of residence in the province, as factors among youth aged 15-17 years in being “not at work and not at school” using the 2000 Population and Housing Census data of the three southernmost provinces of Thailand. Our main finding was that social inequality among demographic groups existed in some regions. The social inequality was based on the multiplicative logistic model of the adverse outcome of “not at work and not at school”. An acceptable social inequality indicator is still an issue for research in social science as well as in public health area (Veenhoven [14], Næss et al. [11], and Geyer et al. [7]).

The strength of the study is that we applied our method to the data comprises nearly the whole target population in the three provinces. It is a population based study of youth aged 15-17 years in the year 2000. Learning from data in the past provides useful information for long term education plan for youth in the future. The 2000 census data was collected before the escalation of unrest but the non-participation and disparity might have contributed to the escalation of unrest. There is a growing concern about youth contributed to the situation. This suggests an issue for further research to investigate.

The high rates of not work or not study in some areas of Pattani, Yala and Narathiwat, as well as areas of high levels of disparity between religion and gender groups possibly reflect that the social inequality exist in the past. This study raise questions on social inequality whether it still exists.

There is potential value in using this paper’s plotting technique to identify locations, districts or provinces in need of urgent action to increase social justice, not just because of high levels of non-participation but also because of high levels of inequality.

Acknowledgements

We are grateful to Emeritus Professor Don McNeil for his suggestions. Thanks are also due to the National Statistical Office of Thailand for generously providing the data. This study is partly funded by Higher Education Commission of Thailand.

References

[1] D. Avramov, People, Demography and Social Exclusion, Population Studies No. 37, Council of Europe Publishing, Brussels, 2002.

[2] H. Booth, J. Maindonald and L. Smith, Applying Lee-Carter under conditions of variable mortality decline, Population Studies 56(3) (2002), 325-336.

[3] P. Buonanno and D. Montolio, Identifying the socio-economic and demographic determinants of crime across Spanish provinces, International Review of Law and Economics 28 (2008), 89-97.

[4] F. Cowell, Measuring Inequality in LSE Perspectives in Economic Analysis. Oxford University Press, 2008. Available online at: http://darp.lse.ac.uk/papersdb/Cowell_measuringinequality3.pdf.

[5] U. Figen, Social participation of Turkish and Arabic immigrants in the neighborhood: case study of Moabit West, Berlin, Journal of Identity and Migration Studies 1(2) (2007), 44-66.

[6] F. Gesemann, The integration of young Muslims in Germany - education and training as key areas for social integration. Available online at: http://www.migrationeducation.org/22.1.html?&rid=82&cHash=6af0a6874a. 2007.

[7] S. Geyer, R. Perter and J. Siegrist, Socioeconomic differences in children's and adolescents hospital admissions in Germany: a report based on health insurance data on selected diagnostic, J. Epidemiol. Community Health 56 (2002), 109-114.

[8] W. D. Hosmer and S. Lemshow, Applied Logistic Regression, 2nd ed., New York, 2000.

[9] D. G. Kleinbaum and M. Klein, Logistic Regression: A Self-learning Text, 2nd ed., Springer-Verlag, New York, 2002.

[10] R. D. Lee and L. R. Carter, Modeling and forecasting US mortality, Journal of American Statistical Association 87(419) (1992), 659-671.

[11] Øyvind Næss, B. Claussen and S. D. Thelle, Four indicators of socioeconomic position: relative ranking across causes of death, Scand. J. Public Health 33 (2005), 215-221.

[12] J. Phillimore and L. Goodson, Making a place in the global city: the relevance of indicators of integration, Journal of Refugee Studies 21(3) (2008), 305-325.

[13] H. Theil, Linear Algebra and Matrix Methods in Econometrics, North-Holland Publishing Company, Amsterdam, 1983.

[14] R. Veenhoven, Return of inequality in modern society? Trend in dispersion of life-satisfaction in EU-nations 1973-2001, Paper presented at the 15th International Conference of Sociology Brisbane, Australia, July 2002, 2002.

[15] W. Venables and B. Ripley, Statistics Complements to Modern Applied Statistics with S, 4th ed., New York, 2002.