Measuring spatial distribution of education inequality

Siti Nurliza Samsudin[1]

[1] Malaysian Bureau of Labour Statistics (MBLS), Department of Statistics Malaysia (DOSM)

Abstract:
Education inequality is where access to good education is unequally based on circumstances such as ethnic group, geographical location, income group or other factors. The objective of this paper is to compute indicators of education inequality using the Population and Housing Census data and list of schools under the Ministry of Education. In particular, the mean years of schooling and Gini coefficient of education attainment were calculated. The paper will also investigate the factors affecting the mean years of schooling, including its relationship with distance to the nearest high performing school through ordinary least squares regression and geographic weighted regression.

Keywords: Education inequality, Gini coefficient, Spatial data

1. Introduction:

Based on a 2018 report by the UNICEF, education inequality is when a child performs worse than others due to circumstances beyond their control. In other words, it is where access to good education is unequally based on circumstances such as ethnic group, geographical location, income group or other factors.

In Malaysia, many types of schools exist to cater the requirements of the diverse society across different states, religion and objectives. This occur at every stage of education i.e. primary, secondary and tertiary. Even among government schools, such diversity occurs. Hence there exist schools which are rewarded based on merits of the students, including Cluster Schools, High Performing Schools and others. As these schools continue to maintain its status as a high performing school, the students have access to various opportunities beyond a typical national school. Upon graduation from secondary schools, it is also possible that the students have access to better tertiary education, and finally perhaps better job prospects.

2. Methodology:

A Gini index will be calculated based on information on education attainment in 1991, 2000 and 2010 to measure education inequality in Malaysia. This is based on calculations suggested in a paper by the World Bank (Thomas, Wang, & Fan, 2001). This paper calculated based on population aged 15 and above, however, according to the UNESCO Institute for Statistics Methodology for Estimation of Mean Years of Schooling, the population covered are aged 25 and above. Hence, taking into account the latest recommended methodology, the population covered for the compilation of education inequality indicators in this paper are those aged 25 and above.

First, the Gini index is calculated using a formula as follows:

$$E_L = \frac{1}{\mu} \sum_{i=2}^{n} \sum_{j=1}^{i=1} p_i |y_i - y_j| p_j$$

Where,
E_L is the education Gini based on educational attainment distribution,
μ is the average years of schooling of the population,

p_i and p_j are the proportions of populations with certain levels of schooling,
y_i and y_j are the years of schooling at different attainment levels,
n is the number of levels/categories; in this paper $n = 6$.

As in the methodology used by Thomas, Wang, & Fan (2001), the average years of schooling (AYS) is calculated as follows:

$$\mu = AYS = \sum_{i=1}^{n} p_i y_i$$

In relation to the census conducted in Malaysia, the years of schooling will be calculated based on the variable highest level of education attained. Then, upon calculations of the AYS and the Gini Index, the AYS will be calculated at the enumeration block level using the census data. Following this, education inequality will be analysed based on granular geographical location, as well as based on comparison between data in 1991 and in 2010. This will use the spatial data based on the census in 1991, 2000 and 2010 which are at enumeration block level.

Based on a research conducted by the Ministry of Health in 2013 to map spatial distributions of health clinics for public and sectors in Malaysia (Hazrin, et al., 2013), this paper will attempt to measure geographical regression between AYS and several key variables of socioeconomic status including population density, distance to the nearest high-performance school and percentage of skilled employed persons for each enumeration block.

3. Findings and Discussion

Computing the Gini coefficient of education attainment

According to the Handbook on Measuring Equity in Education by United Nations Educational, Scientific and Cultural Organization (UNESCO), a method of measuring education inequality is computing the Gini coefficient of education attainment, that is a univariate model based on the distribution of an educational variable. Similar to the Gini coefficient based on income, the coefficient varies between 0 and 1, where a value equals to 0 means that the population has perfect equality and a coefficient being 1 means that there is maximal inequality.

The results, which were computed based on the 1991 and 2010 census, found that the Gini coefficient had improved over the span of 20 years. The average years of schooling of the population aged 25 and above grew from 6.1 in 1991 to 9.1 in 2010. Consequently, the Gini coefficient depicted improvement from 0.425 in 1991 to 0.300 in 2010. By state, it was found that the lowest average years of schooling is in Sabah (1991: 4.6 years; 2010: 7.0 years), whereas the highest is in W.P. Kuala Lumpur during the year 1991 (8.2 years) and in W.P. Putrajaya during the year 2010 (13.7 years). Similarly, for the measurement of inequality, it was found that it is most unequal in Sabah and most equal in W.P Kuala Lumpur (1991) and W.P. Putrajaya (2010). Analysing based on demographic characteristics, it was found that male had higher average years of schooling in both 1991 and 2010. By age group, the most unequal was the older population aged 65 and above with an average year of schooling of 1.6 years in 1991 and 3.7 years in 2010. This may have shown that access to education had been better as the generation progresses. In terms of ethnic group, 'Other Bumiputera' had the lowest average years of schooling in 1991, with Gini coefficient of 0.628. In 2010, however, non-citizens recorded the lowest instead, with average years of schooling of 5.8 years and Gini coefficient of 0.514. By area, it was recorded that rural areas had higher inequality and lower average years of schooling, both in 1991 and in 2010.

Figure 2: Mean years of schooling, Malaysia, 1991 & 2010

1991   2010
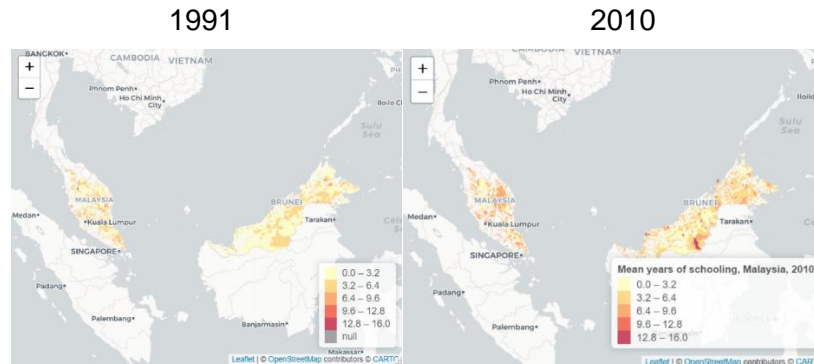
The heat map of average years of schooling illustrates that the areas with higher values were mainly concentrated in high population density areas. Thus, it may be possible that with economic development and job opportunity in a specific area, the population are more accessible to higher education.

Measuring spatial distribution of mean years of schooling

To measure the spatial distribution of education inequality, the indicator used is the computed mean years of schooling, as above. The mean years of schooling was computed for each enumeration block. As high-performing schools only existed since year 2008, this analysis is only conducted for the census data in 2010. There would be two model conducted onto the data i.e. an ordinary least squares method and a geographically weighted regression (GWR) model.

The ordinary least squares model was conducted using Queen contiguity weights via the software GeoDa. The dependent variable was average years of schooling (a_2010_), whereas the covariates are enumeration block area in squared metres (are_mtr), population (JUM_JAN), distance to the nearest high-performing school (skk_dst) and percentage of skilled workers (skilled). Based on the results, it was found that the model had an R-squared of 0.411, meaning that about 40 percent of the enumeration block-level data fit into the model. In this study, it can be deduced that the covariates have p values less than 0.05, hence the probability to reject the null hypothesis of these covariates contribute to the mean years of schooling for each enumeration block is less than 0.05.

Next, the GWR model was conducted on the 2010 data with an adaptive kernel and bandwidth method AICc using ArcGIS. In general, as compared to the previous model which measures at the global level, the GWR measures the fitness of the model on a local level. Prior to the GWR modelling, an analysis of average nearest neighbour was conducted onto the points of the high-performing schools. It was found that given the z-score of -28.67, there is a less than 1 per cent likelihood that this clustered pattern could be the result of random chance. Hence, the points are most likely clustered. The GWR model was found to have an R-squared of 0.600, meaning that 60 percent of the enumeration block-level data fit into the model. The AICc value of this model was also found to be lower than that of the ordinary least squares regression, suggesting that the GWR is a better predictor.

Mapping the local R squared values, it was found that many areas have R squared values above 0.5, thus concluding that the mean years of schooling was dependent on the variables used to a high extent. However, at certain areas, particularly at the south of the Peninsular Malaysia i.e. Johor and Melaka, as well as certain

areas in Sabah and Sarawak, the model seems to have poorly performed. In the case of underfitting, it may be possible to include factors of income and other potential variables which dictate a socioeconomic status of a household. Whereas in the case of overfitting, it may be possible that it is because the distance to the nearest high-performance school did not contribute to the mean years of schooling in the area, especially given that this study did not regard residential schools into the picture.

4. Conclusion

This study attempted to measure education inequality with respect to the equality of condition; that is "educational opportunities must be the same for everyone in the population regardless of their different circumstances" (UNESCO, 2018) through the computation of the Gini coefficients, as well as impartiality: "educational opportunities should be distributed equally by gender, ethnicity, religion, language, location, wealth, disability, and other characteristics" (UNESCO, 2018) through the spatial analysis. Through the study, it was found that education inequality improved over the years from 1991 to 2010, although through the PISA results in 2009, the disparity was still obvious. In the latter, the study illustrated evidence of this variable correlating with location, in particular areas with higher degrees of urbanisation. The study also aimed to investigate relationship between distances to the nearest high-performing schools, although there had not been clear evidence that there exists any relationship.

In any case, the entire study showed evidence of education inequality. Mincer earnings functions do confirm that there is positive relation between schooling and income (Thomas, Wang, & Fan, 2001). In Malaysia, the Household Income and Expenditure Report (DOSM, 2020a), the Gini coefficient of income also improved from 45.1 per cent in 1992/1993 to 44.1 per cent in 2009. Recent data, however, reported worsen income inequality in 2019 as compared to in 2016. This may also mean that education inequality had also worsen. On the other hand, the Government had also shown more effort to target high-need schools, rather than high-performing schools, as shown in the Malaysia Education Blueprint (MEB) 2013-2015 Annual Report 2018. With the insight on spatial distribution of education inequality through this paper, it is hoped that this will potentially assist policymakers to design education policies catering at the locality level. Possibly, this will ensure that the right amount of resources is channelled accurately to increase education opportunities.

5.   References:

DOSM. (2020a). *Household Income and Expenditure Report 2019.* Department of Statistics, Malaysia.

Hazrin, H., Fadhli, Y., Tahir, A., Safurah, J., Kamaliah, M., & Noraini, M. (2013). Spatial patterns of health clinic in Malaysia. *Health Vol.5 No.12*.

Marrero, G. A., & Rodri´guez, J. G. (2012). Inequality of Opportunity in Europe. *Review of Income and Wealth*, 598.

Ministry of Education Malaysia. (2018, November 27). *Senarai Sekolah Kluster Kecemerlangan (SKK)*. Retrieved from Data.gov.my: https://www.data.gov.my/data/ms_MY/dataset/senarai-sekolah-kluster-kecemerlangan-skk

Ministry of Education Malaysia. (2020, February 18). *Senarai sekolah documents*. Retrieved from Ministry of Education Malaysia Portal: https://www.moe.gov.my/en/muat-turun/laporan-dan-statistik/senarai-sekolah

Thomas, V., Wang, Y., & Fan, X. (2001). Measuring Education Inequality: Gini Coefficients of Education. *Policy Research Working Paper 2525*.