Pertanika J. Soc. Sci. & Hum. 26 (3): 1265 - 1281 (2018)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Construction of a Socio-Economic Status (SES) Index in Peninsular Malaysia Using the Factor Analysis Approach

Abdul Rahman, N.* and Abd Naeeim, N. S.

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

ABSTRACT

The state of a country's economy plays an important role in helping it achieve developed nation status. Hence, studies on poverty, inequality and socio-economic matters should use precise measures in order to provide accurate understanding of a country's development. Accurate measures can also provide information on progress already achieved and aid in planning for future improvement and development. This study focusses on constructing a socio-economic status (SES) index and describes a statistical procedure to derive the SES index in a multivariate context for every district in Peninsular Malaysia. Factor analysis is applied to construct the SES index. Data on 19 variables measuring multiple aspects of socioeconomic status such as household amenities, basic facilities, education level and labour force are factorised into three factors that explain 76% of the total variation. The high SES index is situated in Kuala Lumpur, Pulau Pinang, Selangor and some regions in Kedah, Perak, Melaka, Negeri Sembilan and Johor, while the low SES index is situated in areas in Kelantan, Terengganu and some rural areas in Kedah and Pahang. The findings can facilitate the relevant authorities in taking proactive steps to prioritise the development of the relevant areas in order to reduce the socio-economic gap between districts in Peninsular Malaysia.

Keywords: Factor analysis, principal component extraction method, socio-economic status (SES) index

ARTICLE INFO

Article history:
Received: 30 July 2016
Accepted: 25 June 2018
Published: 28 September 2018

E-mail addresses:
nuzlinda@usm.my (Abdul Rahman, N.)
nurulsyafiahabdnaeeim@yahoo.com (Abd Naeeim, N. S.)
* Corresponding author

ISSN: 0128-7702
e-ISSN 2231-8534

INTRODUCTION

A socio-economic status (SES) index is a total measure of the social and economic standing of an individual or area that involves a combination of many variables that capture living standards. An SES index is useful for studying the standard of living of the public and the country's progress. It is also used as an outcome measure to study the relationship between SES and various health behaviours, education and social problems like crime and mortality. Numerous individual and areal measures of SES have been constructed previously, variously termed as socio-economic status index, deprivation index and inequality index. For instance, Australia uses the Socio-Economic Indices for Australia (SEIFA) (Castles, 1994), Jarman index (Jarman, 1983), Townsend index (Townsend, Simpson, & Tibbs, 1985), while the Carstairs index is used in the United Kingdom (Carstairs, 1995; Carstairs & Morris, 1989), New Zealand uses the index of socio-economic deprivation (NZiDep) (Salmond, Crampton, & Sutton, 1998) and Canada uses the Pampalon index and Can-Marg index (Matheson, Dunn, Smith, Moineddin, & Glazier, 2014; Pampalon, Hamel, & Gamache, 2009).

In order to construct an SES index, different researchers have used different numbers of variables. Townsend and Carstairs' index incorporates four variables; three of the variables are similar including unemployment, non-car ownership and household overcrowding, while for the fourth variable, the Carstairs index substitutes non-home ownership with low social class. The NZiDep91 are based on seven variables, which are income, employment, transport, living space, home ownership, qualification and support. Besides that, there are researchers who incorporate more than 10 variables in the construction of an SES index. For instance, Vyas and Kumaranayake (2006) used 26 variables, while Lalloué et al. (2013) used 20 variables for their own index study. The purpose of including various variables is to ensure deeper understanding of socio-economic differentiation multidimensionally, which reflects the pattern of unequal distribution of resources and population. In general, there is no previous study stating the best indicator in measuring SES. The formation of an SES index is based on the objective of the researchers as it is hard to assign a universal measure of SES that would be helpful in all areas. However, most of them have used the demographic and socioeconomic data that were presented in the national census data. Variables such as durable assets (e.g. car, motorcycle, TV), infrastructure and housing characteristics (e.g. electric, source of water, sanitation facility), education level, labour force, marital status, population density and urbanisation are repeatedly used (Holt & Lo, 2008; Howe, Hargreaves, & Huttly, 2008; Krishnan, 2010; Lalloué et al., 2013; Rahman & Zakaria, 2012; Sanusi, 2008; Vyas & Kumaranayake, 2006).

The inclusion of many variables in an SES index may lead to the presence of multicollinearity. A review of the literature shows that recently most researchers have used Factor Analysis (FA) and Principal Component Analysis (PCA) to construct an SES index and, at the same time, to avoid the multicollinearity problem. For example, Earnest et al. (2015); Filmer and Pritchett (2001); Gwatkin, Rutstein, Johnson, Suliman, Wagstaff and Amouzou (2007); Vyas and Kumaranayake (2006) and used the PCA to construct an SES index while Holt and Lo (2008); Krishnan (2010); Sahn and Stifel (2003); Salmond et al. (2006), and Zakaria (2014) used FA to construct their index. Both methods are useful in expressing a set of variables into a smaller number of factors but the approaches are slightly different. DeCoster (1998) pointed out two main difference between FA and PCA. The first was in terms of direction of the influence. FA assumes that the measured responses are based on the underlying factors, while in PCA, the principal components are based on the measured responses. Second, FA assumes that the variance in the measured variables can be decomposed into that accounted for by common factors and by unique factors, while PCA cannot decompose the variance as it is defined as linear combinations of the measurements.

Apart from PCA and FA, there are various methods for SES index construction such as correspondence analysis, multiple correspondence analysis, multivariate regression and fuzzy. Cortinovis, Vella and Ndiku (1993) used correspondence analysis to derive an SES measure. However, the analysis can only be used for categorical data both nominal and ordinal and also for continuous data, but the data need to be reorganised into range. Meanwhile, according to Howe et al. (2008), multiple correspondence analyses were analogous to PCA but were for discrete data. Marí-Dell'Olmo et al. (2011); Pornet et al. (2012) used others methods such as the fuzzy set composite indices and also Bayesian factor analysis, respectively.

Due to the limited number of studies related to an SES index construction in Malaysia, this study aimed to construct an SES index based on the geographical and socio-economic factors in Peninsular Malaysia by including variables that are provided in census data. The index was constructed using a method that is widely used, which is factor analysis to extract the latent variables from multiple perspectives, weigh each factor appropriately and calculate a single number index for each district in Peninsular Malaysia. The result of this study, it is hoped, will further assist policy-makers in planning regions optimally and effectively so that all districts will receive an appropriate resource distribution.

MATERIALS AND METHODS

Malaysia is a country on the continent of Asia, specifically in Southeast Asia, that comprises Peninsular Malaysia and the states of Sabah and Sarawak in Borneo. However, this study only covers Peninsular Malaysia, which is to the west of Borneo, due to the equitable development of this region and data availability.

In this study, the administrative district is used as the unit of analysis. In year 2000, Peninsular Malaysia consisted of 82 administrative districts. The demographic and socio-economic data for year 2000 were obtained from the national population census report published by the Department of Statistics Malaysia, while the crime data were obtained from the Royal Malaysia Police. Census data, which are collected every 10 years, are a source of data that contain much information on the number of persons and households together with a wide range of demographic and socio-economic characteristics in Malaysia. They also provide information on the availability of basic amenities. For the purpose of this study, 20 variables related to household amenities, basic facilities, school attendance, education level and labour force were selected from the available data to construct the SES index based on the indicators that were most often used in the literature in addition to several other relevant variables. The 20 variables are as follows:

1. Percentage of households that have a car (car)
2. Percentage of households that have a motorcycle (motorcycle)
3. Percentage of households that have an air-conditioner (air-conditioner)
4. Percentage of households that have a washing machine (washing machine)
5. Percentage of households that have a telephone (telephone)
6. Percentage of households that have a television (television)
7. Percentage of households that have a video (video)
8. Percentage of households that use tap water (water)
9. Percentage of households that have 24-hour electricity supply (electric)
10. Percentage of households that have proper toilet facilities (toilet)
11. Percentage of residents who have tertiary education (tertiary education)
12. Percentage of persons who did not attend school (school attendance)
13. Percentage of married persons who are single mothers (single mother)
14. Percentage of persons who are married (married)
15. Mortality rate per 1000 persons of the population (mortality)
16. Percentage of persons of age 15 to 64 years old who are professional workers (professional)
17. Percentage of persons of age 15 to 64 years old who are engaged in basic work (basic work)
18. Property crime rate per 1000 persons of the population (crime)
19. Percentage of urbanisation (urbanisation)
20. Population density (population density)

In this study, since there is a relatively high number of variables in constructing the SES index, factor analysis was used to create the index with application of SPSS 16.0. The essential purpose of factor analysis is to analyse interrelationships among a large number of variables and to classify variables into their common factors. As a result, the original number of variables can be condensed into a smaller set of dimensions with a minimum loss of information.

The method for constructing the SES Index is explained in several steps given as follows:

Step 1: Creating variables and data screening

The variables used in this study were in the form of percentage and rate to avoid having undue influence on the overall index since different districts had a different number of household and population. Variables should roughly be normally distributed to derive a preferred solution with no outliers for the data related to each variable since factor analysis is sensitive to outliers.

Step 2: Measuring the internal consistency of a scale

Based on Tawakol and Dennick (2011), internal consistency should be determined before a test can be employed for research or examination purposes to ensure validity. Thus, Cronbach's alpha that was developed by Cronbach (1951) was used to measure the internal consistency and it was expressed as a number in the range of 0 to 1. The Cronbach's alpha is defined as:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum s_v^2}{s_{test}^2}\right) \quad [1]$$

Where,

n = Number of items

s_v^2 = Variance of the vth item

s_{test}^2 = Variance of the total score formed by summing all the items

Step 3: Checking of assumptions

Before proceeding to factor analysis, two main tests were used for testing the appropriateness of factor analysis namely, the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test. The KMO test was used to measure sampling adequacy and interrelationships among variables, while Bartlett's test was used to determine whether the variables were correlated or not. Factor analysis is appropriate if the KMO value is greater than 0.5 and if Bartlett's test is significant.

Step 4: Determining number of factors to be retained and undergoing factor rotation

The factors were extracted using the principal component method and administering varimax rotation with Kaizer normalisation although there were other options. Using the principal component method for extraction allows the number of factors extracted to be defined by the user or to be regulated to the Eigenvalue rule or to be obtained via scree plot. The Eigenvalue is the variance extracted by the factors. Under the Eigenvalue rule, only those factors with an Eigenvalue of 1.0 and above are retained. Meanwhile, for scree plotting, the number of factors to be kept should be equal to the number of factors before the bending point. Besides that, the number of factors can be determined by keeping the factors, which in total, account for about 60-

80% of the variance. Meanwhile, varimax rotation is used to minimise the number of variables that have a high loading on the first factor.

Step 5: Index calculation

The variables were standardised using the following standardisation so that the standardised observed values would range between zero and one.

$$Z_{ik} = \frac{x_{ik} - \min(x_{ik})}{\max(x_{ik}) - \min(x_{ik})} \quad [2]$$

Where, x_{ik} is the actual observed value in district *i* for variable *k* and $\min(x_{ik})$ and $\max(x_{ik})$ are the minimum and maximum observed value for variable *k*, respectively. This standardisation method was chosen as it can show the gap between the observed area and the area that has the lowest observed value. Zero corresponds to the lowest level and one corresponds to the highest level in each set of variables (Sanusi, 2008).

Meanwhile, the index coefficient, $\mathbf{W}_{kf}$, was computed by multiplying the inverse matrix of original variables correlation, $\mathbf{R}_{kk}^{-1}$, with the rotated component matrix, $\mathbf{S}_{kf}$:

$$\mathbf{W}_{kf} = \mathbf{R}_{kk}^{-1}\mathbf{S}_{kf} \quad [3]$$

Then, the standardised observed values were multiplied by the matrix of factor score coefficient, $\mathbf{W}_{kf}$, to obtain the estimated factor indices for each district, $\mathbf{F}_{if}$:

$$\mathbf{F}_{if} = \mathbf{Z}_{ik}\mathbf{W}_{kf} = \sum_{k=1}^{K} Z_{ik}W_{kf} \quad [4]$$

Where, *K* is the number of variables in the measured factor.

Then, based on Fukuda, Nakamura and Takano (2007), the index value for each factor was summed up to get the socio-economic index:

$$\text{SES Index} = \sum_{f=1}^{c} F_{if} \quad [5]$$

Where, *c* is the number of reduced factors.

RESULTS AND DISCUSSION

The results of this analysis are given in Table 1 to Table 10. Table 1 shows the Cronbach's alpha value for reliability checking, while Table 2 and Table 3 show the changes in the value of Cronbach's alpha when items were deleted. Table 1 shows the Cronbach's alpha value when all variables were included i.e. 0.123. The small value was neither good nor moderate; hence, some variables needed to be dropped for factor analysis. The 'Cronbach's Alpha if Items Deleted' column in Table 2 shows that the Cronbach's alpha was 0.865 if variable population density were deleted and only several changes would happen if another item were deleted. Hence, only population density was dropped; this improved the Cronbach's alpha value to 0.865, which is considered acceptable according to Tavakol and Dennick (2011). Statistically, the best reduced set had 19 variables. This conclusion was further supported by examination of the variability of the values of the Cronbach's alpha when individual variables were deleted for the reduced set of 19 variables. Based on Table 3, the restricted values of alpha varied from 0.834 to 0.889, close to the overall value of Cronbach's alpha, suggesting none of those 19 variables were to be deleted and they were internally consistent.

Table 1
Cronbach's alpha value for reliability checking

	Cronbach's alpha
By including all variables	0.123
After deleting variable population density	0.865

Table 2
Cronbach's alpha if items deleted by including all variables

Variable	Cronbach's Alpha if Item Deleted
Car	0.110
Motorcycle	0.136
Air-Conditioner	0.113
Washing machine	0.116
Telephone	0.111
Television	0.121
Video	0.106
Water	0.117
Electricity	0.122
Toilet	0.099
Tertiary education	0.118
School attendance	0.127
Single mother	0.124
Married	0.122
Mortality	0.125
Professional	0.121
Basic work	0.123
Crime	0.115
Population density	0.865
Urbanisation	0.102

Table 3
Cronbach's alpha if item deleted after deleting variable population density

Variable	Cronbach's Alpha if Item Deleted
Car	0.841
Motorcycle	0.889
Air-Conditioner	0.850
Washing machine	0.845
Telephone	0.837
Television	0.857
Video	0.834
Water	0.844
Electricity	0.864
Toilet	0.847
Tertiary education	0.861
School attendance	0.878
Single mother	0.869
Married	0.862
Mortality	0.870
Professional	0.865
Basic work	0.864
Crime	0.859
Urbanisation	0.849

Table 4 shows the result of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. For this study, the KMO value was 0.858, signalling that a factor analysis of the variables could proceed. Furthermore, the results of Bartlett's test showed a significant value of <0.001, a value that was small enough to reject the null hypothesis although it was at the 1% significant level. This implied that the strength of the relationship among the variables was strong and the correlation matrix was not an identity matrix. Hence, these diagnostic procedures indicated the suitability of factor analysis for the data.

Table 4
KMO measure of sampling adequacy and Bartlett's Test of Sphericity

KMO Measure of Sampling Adequacy	Bartlett's Test of Sphericity		
	Chi-Square	**Df**	**Sig.**
0.858	**1787.99**	**171**	**0.000**

In the initial stage of factor analysis, the pattern of intercorrelations between the studied variables was examined through a correlation matrix. Variables that were influenced by the same factor were highly correlated compared to variables from different factors. Based on the correlation matrix of the 19 variables (refer Appendix), the variables car, air-conditioner, washing machine, video, water, toilet and school attendance were highly correlated to each other, while the variable motorcycle was highly correlated to the variables tertiary education and professional.

Table 5 shows the Eigenvalue and the percentage of total variation in line with the number of factors. The table shows that at least two factors are necessary for a minimum of 60% of total variation and at most, four factors can be retained as four factors gave an Eigenvalue of more than 1. However, four was not chosen

Table 5
Total variance derived from principal component analysis extraction method

Factor	Initial Eigenvalues		
	Eigenvalue	% of Variance	Cumulative Variance %
1	9.868	51.936	51.936
2	2.845	14.973	66.909
3	1.809	9.523	76.432
4	1.012	5.324	81.756
5	0.652	3.434	85.127
6	0.593	3.123	88.249
7	0.456	2.400	90.649
8	0.423	2.225	92.874
9	0.264	1.389	94.263
10	0.247	1.299	95.562
11	0.199	1.046	96.608
12	0.169	0.888	97.496
13	0.116	0.610	98.106
14	0.103	0.544	98.651
15	0.080	0.422	99.073
16	0.058	0.304	99.377
17	0.056	0.292	99.669
18	0.035	0.184	99.854
19	0.028	0.146	100.000

because the variables that were represented in the four factors were relatively not meaningful and difficult for naming. For this reason, the number of factors for extraction was set at three. The Eigenvalues of the extracted factors were 9.868, 2.845 and 1.809, respectively, giving 76.432% of the total variance.

Once the factor extraction was done, it was important to check the communalities. The communalities represented the proportion of variance in the original variables that was accounted for by the factor solution after the extraction process. If the communalities are low, the extracted factors account for only a small part of the variance and more factors may be retained. Table 6 shows that the communality value for all the variables are high, implying that these particular variables were well reflected via the extracted factors; hence, the three factors were reliable. For instance, it can be said that 91% of the variance for the variable car can be explained by retaining factors after extraction.

Table 6
Communalities

Variable	Extraction
Car	**0.910**
Motorcycle	**0.711**
Air-Conditioner	**0.815**
Washing machine	**0.755**
Telephone	**0.859**
Television	**0.864**
Video	**0.926**
Water	**0.722**
Electricity	**0.771**
Toilet	**0.872**
Tertiary education	**0.809**
School attendance	**0.821**
Single mother	**0.825**
Married	**0.754**
Mortality	**0.591**
Professional	**0.896**
Basic work	**0.410**
Crime	**0.661**
Urbanisation	**0.550**

The total variance explained by factor and the factor matrix after varimax rotation are presented in Table 7 and Table 8, respectively. Comparing the extraction results shows that the percentage of variance of the first factor with rotation was smaller compared to without rotation. Meanwhile, based on the factor matrix, the number of variables that had high loading in the first factor decreased after rotation. These outputs indicated that varimax rotation altered the pattern of factor loading by minimising the number of variables that had high loading on the first factor and maximising the contrast between the factors. After rotation, factor 1 accounted for 37.53% of the variability, factor 2 accounted for 27.40% of the variability and factor 3 accounted for 11.49% of the variability, which finally explained 76.43% of the total variance in all 19 variables. The result inferred that the three factors could explain over 75% of the information contained in the original variables.

Focussing on the rotated factor matrix in Table 8, it can be seen that the factor loading in the matrix represented the correlation of the original variables with the factor,

Table 7
Total variance explained before and after varimax rotation

Factor	Extraction Sums of Squared Loading			Rotation of Squared Loading		
	Eigenvalue	% of Variance	Cumulative Variance %	Eigenvalue	% of Variance	Cumulative Variance %
1	9.868	51.936	51.936	7.132	37.534	37.534
2	2.845	14.973	66.909	5.207	27.404	64.938
3	1.809	9.523	76.432	2.184	11.494	76.432

Table 8
Factor matrix

Variable	Factor Matrix			Rotated Factor Matrix		
	1	2	3	1	2	3
Car	**0.947**	-0.116	-0.004	**0.692**	0.646	-0.113
Motorcycle	-0.493	**0.672**	-0.118	-0.014	**-0.772**	0.340
Air-Conditioner	**0.889**	-0.118	0.064	0.616	**0.659**	-0.035
Washing machine	**0.822**	0.281	0.004	**0.793**	0.329	0.132
Telephone	**0.847**	0.376	0.007	**0.870**	0.271	0.167
Television	**0.643**	0.591	0.306	**0.724**	0.155	0.562
Video	**0.938**	0.057	-0.199	**0.836**	0.440	-0.182
Water	**0.712**	0.412	-0.215	**0.847**	0.071	0.016
Electricity	0.364	0.335	**0.722**	0.277	0.312	**0.773**
Toilet	**0.913**	-0.062	-0.179	**0.746**	0.514	-0.226
Tertiary education	**0.720**	-0.434	0.317	0.243	**0.866**	-0.007
School attendance	**-0.902**	-0.067	0.043	**-0.757**	-0.497	0.026
Single mother	-0.574	0.332	**0.630**	-0.470	-0.289	**0.722**
Married	**0.678**	0.363	-0.398	**0.853**	0.000	-0.161
Mortality	-0.364	**0.593**	0.350	-0.073	-0.467	**0.606**
Professional	**0.584**	-0.506	0.542	0.030	**0.935**	0.145
Basic work	0.263	**0.583**	-0.059	**0.532**	-0.235	0.268
Crime	**0.721**	-0.286	0.255	0.348	**0.735**	0.010
Urbanisation	**0.740**	-0.047	0.010	**0.568**	0.473	-0.067

ranging from -1 to +1. Positive values indicate that the variables had a proportional relationship with the factor, while negative values indicate that the variables had an inverse relationship with the factor. A high absolute value of the loading means that the factor contributes more to the variable. The values given in bold give an indication as to which factor the variable belongs. Factor 1 consisted of the variables car, washing machine, telephone, television, video, water, toilet, school attendance, married person, basic work and urbanisation, while Factor 2 consisted of motorcycle, air-conditioner, tertiary education, professional worker and property crime. The last factor, Factor 3, consisted of electricity, single mother and mortality. Factor 1 may represent the factor for basic household needs, Factor 2 may represent middle-class households and the third factor may represent social factors.

Table 9
Factor score coefficient matrix, W_{kf}

Variable	Factor		
	1	**2**	**3**
Car	0.056	**0.082**	**-0.030**
Motorcycle	**0.100**	-0.205	**0.098**
Air-Conditioner	**0.034**	0.105	**0.012**
Washing machine	0.116	**-0.009**	**0.056**
Telephone	0.140	**-0.036**	**0.065**
Television	0.108	**-0.015**	**0.252**
Video	0.122	**-0.009**	**-0.087**
Water	0.172	**-0.107**	**-0.024**
Electricity	**-0.023**	**0.120**	0.387
Toilet	0.093	**0.024**	**-0.098**
Tertiary education	**-0.080**	0.228	**0.060**
School attendance	-0.089	**-0.034**	**0.004**
Single mother	**-0.090**	**0.046**	0.344
Married	0.192	**-0.145**	**-0.116**
Mortality	**0.029**	**-0.080**	0.255
Professional	**-0.143**	0.295	**0.149**
Basic work	0.139	**-0.130**	**0.085**
Crime	**-0.038**	0.173	**0.052**
Urbanisation	0.054	**0.052**	**-0.017**

Next, as discussed in the methodology presented above, factor indices were obtained by multiplying the inverse matrix of coefficient correlations with a factor score coefficient matrix as presented in Table 9. The bold numbers represent the score coefficient according to their respective factor. Since the number of factors extracted was three, the factor indices calculated for each district was separated into three parts. The indices for the first factor, F_{i1}, second factor, F_{i2}, and third factor, F_{i3} were calculated as follows:

$$F_{i1} = \sum_{k=1}^{11} Z_{ik} W_{k1},$$

$$F_{i2} = \sum_{k=1}^{5} Z_{ik} W_{k2},$$

$$F_{i3} = \sum_{k=1}^{3} Z_{ik} W_{k3},$$

For instance, the indices for Factor 1, Factor 2 and Factor 3 in the Federal Territory of Kuala Lumpur were calculated as:

$$\begin{aligned} F_{kl1} &= (0.805 \times 0.056) + (0.789 \times 0.116) \\ &+ (0.776 \times 0.140) + (0.838 \times 0.122) \\ &+ (0.997 \times 0.172) + (0.981 \times 0.093) \\ &+ (0.063 \times -0.089) + (0.684 \times 0.192) \\ &+ (0.350 \times 0.139) + (1.000 \times 0.054) \\ &\approx 0.918 \end{aligned}$$

$$\begin{aligned} F_{kl2} &= (0.000 \times -0.205) + (0.771 \times 0.105) \\ &+ (0.673 \times 0.228) + (0.816 \times 0.295) \\ &+ (1.000 \times 0.173) \approx 0.648 \end{aligned}$$

$$\begin{aligned} F_{kl3} &= (0.990 \times 0.387) + \\ &(0.3480 \times 0.344) + \\ &(0.1781 \times 0.2551) \approx 0.548 \end{aligned}$$

Finally, the overall factor index for a particular district was obtained by summing up the index value from each factor. For instance, the overall factor index for Kuala Lumpur was computed as below:

$$F_{kl} = F_{kl1} + F_{kl2} + F_{kl3}$$
$$= 0.918 + 0.648 + 0.548 = 2.114$$

The complete index is listed in Table 10 from the highest SES index to the lowest index. The larger the index value, the higher the SES index for the area. The top five districts with a high SES index were districts located in the Federal Territory of Kuala Lumpur, Pulau Pinang and Selangor. The highest SES index was located in a metropolitan area, Kuala Lumpur, which also the capital of Malaysia. Besides that, the high SES areas were also situated in the administration areas, which were also the capital cities of the states i.e. Georgetown in Timur Laut (Pulau Pinang), Shah Alam in Petaling (Selangor), Bandaraya Melaka in Melaka Tengah (Melaka), Seremban (Negeri Sembilan), Ipoh in

Kinta (Perak), Alor Setar in Kota Setar (Kedah) and Johor Bahru (Johor). The main economic drivers in these areas are manufacturing and services. These activities contribute to a wide scope of better job opportunities both in the public and private sectors. The overall growth in employment indirectly results in higher purchasing power to own material facilities and also basic utilities. Besides that, the high SES index in these areas was also due to good development of infrastructure and facilities such as public transport and highways (transportation), schools, colleges and universities (education), hospitals (health) and many more.

Meanwhile, the bottom five districts were situated in Kelantan and Pahang. This was followed by a few districts in Terengganu and Kedah. Generally, the low SES areas were far from developed areas or in other words, they are situated in rural areas. They were mostly located in the agricultural, fisheries, hilly and protected areas. In most cases, agriculture and fisheries were the main socio-economic resources for the people in the low SES areas.

It is to be noted that there are a limited number of studies related to SES index construction in Malaysia; nevertheless, the findings of this study are comparable to those of Fam, Ismail and Jemain (2017). However, the difference between both studies was in variable selection as well as the method used in constructing the index. Fam et al. (2017) constructed the General Index of Deprivation (GID), which focussed on deprivation criteria and detecting disadvantages. However, our study focussed on the socioeconomic development index. Based on their result, the majority of deprived areas were located in Kelantan, Terengganu, Kedah, Pahang and Perlis, while the majority of the affluent areas were located in the west coast of Peninsular Malaysia, comprising metropolitan and urban areas. The affluent areas in Peninsular Malaysia were found to be Kuala Lumpur, Petaling Jaya, Johor Bahru, Melaka Tengah and all districts in Pulau Pinang. The comparison analysis between these two studies implied that there was no significant difference between the results obtained from our study and those obtained by Fam et al. (2017). The high GID index indicated a low SES index, while the low GID index indicated a high SES index. However, there was a slight difference in the ranking of our index with the GID index, which may have been due to the difference between the variables and the method used in constructing the index.

Table 10
The SES index value for the districts in Peninsular Malaysia

District	SES Index value	District	SES Index value
Kuala Lumpur	2.1145	Kerian	1.2767
Timur Laut	2.0139	Kota Bharu	1.2757
Petaling	1.9721	Jelebu	1.2575
Gombak	1.7699	Bentong	1.2544
Ulu Langat	1.7387	Hulu Selangor	1.2502
Kinta	1.7040	Temerloh	1.2239
Melaka Tengah	1.6633	Dungun	1.2188
Seremban	1.6535	Batang Padang	1.2178
Kota Setar	1.6467	Kota Tinggi	1.2129
Barat Daya	1.5912	Kemaman	1.2122
Johor Bahru	1.5658	Kuala Selangor	1.2032
Klang	1.5629	Langkawi	1.1908
Seberang Perai Tengah	1.5439	Raub	1.1785
Kuala Muda	1.5155	Maran	1.1503
Kuala Pilah	1.5145	Mersing	1.1341
Seberang Perai Utara	1.4840	Yan	1.1302
Rembau	1.4747	Baling	1.0945
Alor Gajah	1.4694	Marang	1.0838
Kuantan	1.4624	Jempol	1.0465
Batu Pahat	1.4500	Pendang	1.0171
Muar	1.4176	Jerantut	1.0058
Seberang Perai Selatan	1.4137	Machang	0.9714
Jasin	1.4085	Cameron Highlands	0.9443
Manjung	1.4035	Besut	0.9300
Kuala Kangsar	1.3987	Hulu Terengganu	0.9040
Hilir Perak	1.3975	Tumpat	0.8921
Kulim	1.3902	Ulu Perak	0.8832
Segamat	1.3820	Pasir Mas	0.8479
Port Dickson	1.3816	Pekan	0.8314
Tampin	1.3811	Bera	0.8285
Larut Matang	1.3789	Tanah Merah	0.8167

Table 10 *(continue)*

District	SES Index value	District	SES Index value
Pontian	1.3757	Padang Terap	0.8150
Perak Tengah	1.3732	Pasir Puteh	0.7851
Kluang	1.3726	Sik	0.7826
Kuala Terengganu	1.3523	Lipis	0.7715
Bandar Baharu	1.3406	Setiu	0.7508
Sepang	1.3263	Bachok	0.7238
Kangar	1.3179	Rompin	0.6658
Sabak Bernam	1.2849	Jeli	0.6525
Kubang Pasu	1.2841	Kuala Krai	0.6067
Kuala Langat	1.2784	Gua Musang	0.0965

CONCLUSION

Most of the districts with a high SES were located in states with a high percentage of urban population such as the Federal Territory of Kuala Lumpur, Pulau Pinang, Selangor and Melaka. However, there were also a few districts with a high SES index that were also the capital cities of states. The districts located in the undeveloped states were mostly classified as areas with a low SES index.

Overall, this study provides an idea of how to construct an SES index for areal data using appropriate statistical methods. The results obtained in this study are beneficial and can be used by the authorities to draft proactive action, especially in the development of low SES areas in order to reduce the gap in development between districts in Peninsular Malaysia and to plan further national progress. As different districts have different SES indices, the specific information yielded by this study can advise on the amount of resources to be allocated for different areas for homogenous development between the districts. In addition, the results of the index construction can be used as input data or explanatory variables for further study involving spatial regression analysis in a large variety of contexts like public health studies, social epidemiology, environment assessment and urban and social planning.

ACKNOWLEDGEMENT

This research was fully supported by a Short-Term Grant offered by Universiti Sains Malaysia (304/PMATH/6313154).

REFERENCES

Bland, J., M., & Altman, D., G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, *314*(86), 572–575.

Carstairs, V. (1995). Deprivation indices: Their interpretation and use in relation to health. *Journal Epidemiology Community Health*, *49*(Suppl. 2), S3–S8.

Carstairs, V., & Morris, R. (1989). Deprivation: Explaining differences in mortality between Scotland and England and Wales. *BMJ*, *299,* 886–889.

Castles, I. (1994). 1991 census socio-economic indexes for areas. *Information paper.* Canberra: Australian Bureau of Statistics.

Cortinovis, I., Vella, V., & Ndiku, J. (1993). Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Social Science & Medicine*, *36*(8), 1087–1097.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

DeCoster, J. (1998). Overview of factor analysis. Retrieved March 1, 2016, from http://www.stat-help.com/notes.html.

Earnest, A., Ong, M. E. H., Shahidah, N., Chan, A., Wah, W., & Thumboo, J. (2015). Derivation of indices of socioeconomic status for health services research in Asia. *Preventive Medicine Reports*, *2*, 326–332.

Fam, S. F., Ismail, N., & Jemain, A. A. (2017). Geographical and socio-economic analysis in Peninsular Malaysia. *The Social Sciences, 12*(9), 1695–1704.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effect without expenditure data – Or tears: An application to educational enrolments in states of India. *Demography, 38*(1), 115–132.

Fukuda, Y., Nakamura, K., & Takano, T. (2007). Higher mortality in areas of lower socioeconomic position measured by a single index of deprivation in Japan. *Public Health*, *121*(3), 163–73.

Gwatkin, D. R., Rutstein, S., Johnson, K., Suliman, E., Wagstaff, A., & Amouzou, A. (2007). Socio-economic differences in health, nutrition, and population within developing countries. *Washington, DC: World Bank*, 287.

Holt, J., & Lo, C. (2008). The geography of mortality in the Atlanta metropolitan area. *Computers, Environment and Urban Systems*, *32*(2), 149–164.

Howe, L. D., Hargreaves, J. R., & Huttly, S., R. (2008). Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging Themes in Epidemiology*, *5*(1), 3–16.

Jarman, B. (1983). Identification of underprivileged areas. *British Medical Journal*, *286*, 1705–1708.

Krishnan, V. (2010). Constructing an area-based socioeconomic status index: A principal component analysis approach. *Early Childhood Council Annual Conference 2010,* 2010, May 7, Christchurch, New Zealand.

Lalloué, B., Monnez, J., Padilla, C., Kihal, W., Le Meur, N., Zmirou-Navier, D., & Deguen, S. (2013). A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. *International Journal for Equity in Health*, *12*(1), 21.

Marí-Dell'Olmo, M., Martínez-Beneito, M. A., Borrell, C., Zurriaga, O., Nolasco, A., & Domínguez-Berjón, M. F. (2011). Bayesian factor analysis to calculate a deprivation index and its uncertainty. *Epidemiology*, *22*(3), 356–364.

Matheson, F. I., Dunn, J. R., Smith, K. L. W., Moineddin, R., & Glazier, R. H. (2014). Development of the Canadian marginalization index: A new tool for the study of inequality. *Canadian Journal of Public Health, 103*, 12–16.

Nelder, R., & Maconachie, M. (1997). *Deprivation scores at neighbourhood level within Plymouth City*. Dartington, Devon, England: South & West Devon Health Authority .

Pampalon, R., Hamel, D., & Gamache, P. (2009). A comparison of individual and area- based socioeconomic data for monitoring social inequalities in health. *Health Reports*, *20*(3), 85–94.

Pornet, C., Delpierre, C., Dejardin, O., Grosclaude, P., Launay, L., Guittet, L., … Launoy, G.

(2012). Construction of an adaptable European transnational ecological deprivation index: The French version. *Journal of Epidemiol Community Health*, *66*(11), 982–989.

Rahman, N. A., & Zakaria, S. (2012). The household-based socio-economic index for every district in Peninsular Malaysia. *International Journal of Mathematical, Computational, Natural and Physical Engineering*, *6*(10), 184–191.

Sahn, D., & Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Rev Income Wealth*, *49*(4), 463–489.

Salmond, C., Crampton, P., King, P., & Waldegrave, C. (2006). NZiDep: A New Zealand index of socioeconomic deprivation for individuals. *Social Science & Medicine*, *62*(6), 1474–1485.

Salmond, C., Crampton, P., & Sutton, F. (1998). NZDep91: A New Zealand index of deprivation. *Australian and New Zealand Journal of Public Health*, *22*(7), 835–837.

Sanusi, Y. (2008). Application of human development index to measurement of deprivations among urban households in Minna, Nigeria. *Habitat International*, *32,* 384–398.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2,* 53–55.

Townsend, P., Simpson, D., & Tibbs, N. (1985). Inequalities in health in the city of Bristol: A preliminary review of statistical evidence. *International Journal of Health Service*, *15,* 637–63.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*, *21*(6), 459–468.

Zakaria, S. (2014). *Analisis reruang bagi kes -kes jenayah di Semenanjung Malaysia* [Spatial analysis of crime cases in Peninsular Malaysia] (Doctoral thesis), Universiti Sains Malaysia, Malaysia.

APPENDIX

Supplementary Table 1
Correlation matrix

Car	1.000	-0.573	0.854	0.771	0.762	0.562	0.906	0.549	0.327	0.901	0.699	-0.832	-0.595	0.595	-0.408	0.601	0.173	0.643	0.647
Motorcycle	-0.573	1.000	-0.447	-0.165	-0.211	0.100	-0.457	-0.009	-0.002	-0.555	-0.627	0.373	0.357	-0.097	0.373	-0.637	0.165	-0.591	-0.358
Air-conditioner	**0.854**	-0.447	1.000	0.644	0.669	0.500	0.798	0.519	0.323	0.775	0.728	-0.740	-0.537	0.606	-0.331	0.623	0.203	0.735	0.641
Washing machine	**0.771**	-0.165	**0.644**	1.000	0.806	0.753	0.743	0.734	0.373	0.669	0.451	-0.820	-0.396	0.521	-0.232	0.358	0.277	0.443	0.551
Telephone	**0.762**	-0.211	**0.669**	**0.806**	1.000	0.789	0.836	0.746	0.377	0.772	0.420	-0.783	-0.290	0.730	-0.062	0.333	0.324	0.465	0.568
Television	0.562	0.100	0.500	**0.753**	**0.789**	1.000	0.583	0.624	0.686	0.499	0.275	-0.627	0.001	0.479	0.093	0.246	0.335	0.281	0.368
Video	**0.906**	-0.457	**0.798**	**0.743**	**0.836**	0.583	1.000	0.669	0.209	0.940	0.549	-0.858	-0.624	0.736	-0.323	0.398	0.292	0.596	0.673
Water	0.549	-0.009	0.519	**0.734**	**0.746**	**0.624**	**0.669**	1.000	0.207	0.627	0.314	-0.697	-0.366	0.653	-0.119	0.098	0.303	0.367	0.578
Electric	0.327	-0.002	0.323	0.373	0.377	**0.686**	0.209	0.207	1.000	0.182	0.304	-0.297	0.246	0.088	0.165	0.380	0.224	0.287	0.276
Toilet	**0.901**	-0.555	**0.775**	**0.669**	**0.772**	0.499	**0.940**	**0.627**	0.182	1.000	0.566	-0.816	-0.613	0.679	-0.362	0.452	0.184	0.610	0.650
Tertiary Education	**0.699**	**-0.627**	**0.728**	0.451	0.420	0.275	0.549	0.314	0.304	0.566	1.000	-0.604	-0.405	0.249	-0.376	0.854	0.025	0.702	0.496
School Attendance	**-0.832**	0.373	**-0.740**	**-0.820**	**-0.783**	**-0.627**	**-0.858**	**-0.697**	-0.297	**-0.816**	**-0.604**	1.000	0.505	-0.542	0.349	-0.474	-0.246	-0.553	-0.634
Single mother	-0.595	0.357	-0.537	-0.396	-0.290	0.001	**-0.624**	-0.366	0.246	**-0.613**	-0.405	0.505	1.000	-0.530	0.647	-0.183	-0.035	-0.317	-0.367
Married	0.595	-0.097	**0.606**	0.521	**0.730**	0.479	**0.736**	**0.653**	0.088	**0.679**	0.249	-0.542	-0.530	1.000	-0.043	0.031	0.455	0.342	0.407
Mortality	-0.408	0.373	-0.331	-0.232	-0.062	0.093	-0.323	-0.119	0.165	-0.362	-0.376	0.349	**0.647**	-0.043	1.000	-0.294	0.307	-0.253	-0.327
Professional	**0.601**	**-0.637**	**0.623**	0.358	0.333	0.246	0.398	0.098	0.380	0.452	**0.854**	-0.474	-0.183	0.031	-0.294	1.000	-0.184	0.648	0.400
Basic work	0.173	0.165	0.203	0.277	0.324	0.335	0.292	0.303	0.224	0.184	0.025	-0.246	-0.035	0.455	0.307	-0.184	1.000	0.153	0.193
Crime	**0.643**	-0.591	**0.735**	0.443	0.465	0.281	0.596	0.367	0.287	**0.610**	**0.702**	-0.553	-0.317	0.342	-0.253	**0.648**	0.153	1.000	0.620
Urbanization	**0.647**	-0.358	**0.641**	0.551	0.568	0.368	**0.673**	0.578	0.276	**0.650**	0.496	**-0.634**	-0.367	0.407	-0.327	0.400	0.193	**0.620**	1.000