Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

WEIGHTING METHODS IN THE CONSTRUCTION OF AREA DEPRIVATION INDICES

S. F. Fam[1,*], N. Ismail[2], A. L. Maukar[3], H. Yanto[4], D. D. Prastyo[5], A. A. Jemain[2] and Z. L. Chuan[6]

[1]Universiti Teknikal Malaysia Melaka, Malaysia
[2]Universiti Kebangsaan Malaysia, Malaysia
[3]President University, JababekaEducatin Park, Indonesia
[4]Fakultas Ekonomi, UniversitasNegeri Semarang, Indonesia
[5]Institut TeknologiSepuluhNopemberKampus ITS-Sukolilo, Surabaya, Indonesia
[6]Fakulti SainsdanTeknologiIndustri, Universiti Malaysia Pahang, Malaysia

Published online: 1 May 2018

ABSTRACT

This study applies and compares several weighted average (WA) methods and Principal Component Analysis (PCA) for the construction of composite area-based deprivation index. The WA methods are based on weights that depend on standard deviation, correlation and data entropy. This paper also proposes three new approaches of WA method by suggesting their respective weights to depend on mean absolute deviation, inter-quartile range and data entropy where the probability is estimated by empirical density function and Gaussian kernel function. The deprivation indices produced by WA methods and PCA are then utilized to rank deprivation level of eighty-one administrative districts in Peninsular Malaysia.

Keywords:deprivation index; weighted average; PCA; entropy; Gaussian kernel function.

Author Correspondence, e-mail: famsoofen@utem.edu.my

doi: http://dx.doi.org/10.4314/jfas.v10i6s.201

1. INTRODUCTION

National wealth should be distributed equitably to all communities, so that it can be enjoyed equally [1-3]. Relevant and accurate indicators should be obtained to provide policymakers with a clearer picture of what is currently happening in the country. Any indicators to be used should consider various important features that provide explanation regarding the level of socioeconomic deprivation of an area. Once the features have been identified, the next step is to figure out how these features can be used to develop a composite measure of deprivation which is a blend of several deprivation indicators.

Deprivation index plays an important role in the fields of epidemiology, demography and public health, besides providing significant and relevant contributions in policymaking [4-6]. As examples, several deprivation indices have been constructed and used in health related studies in United Kingdom, especially in health inequalities since the 1980s [7-9]. The main advantage of using deprivation indices is that they are simple and inexpensive as they are composed of indicators from national censuses and estimated from different types of statistical procedures [10]. The problems encountered in constructing composite indices of deprivation have been highlighted by many researchers. For instants, procedures for selecting suitable indicators were explained[1,4,11] and published in country reports and manual [12-13] importance of normalizing indicators was discussed exclusively [10,14] importance of weights in the construction of deprivation indices were highlighted [15-17] and comparison of commonly used methods for aggregation of a single deprivation measure also can be found [18]. This study applies several existing weighting methods and proposes new weighting methods, besides comparing and emphasizing their differences which resulted in different weights for different methods.

For many researches, the assignment of equal weights to each indicator, as carried out by [7-8] were considered as an arbitrary approach. However, weights can be gauged not only arbitrarily but also by using research results, empirical data, government policies, government expenditures and external consultations [4]. In fact, factor loadings from Factor Analysis (FA) and component loadings from PCA can also be utilized as alternatives to construct composite indices and such examples can be found in several developed countries such as Japan, Australia, France and New Zealand[19-22], and developing country such as Malaysia [2-3,6].

The problem in determining an appropriate weight to each indicator for the construction of composite index has also been studied. Based on literatures, methods for determining weights can be categorized into two general groups, subjective and objective weights. Subjective weights were determined based on expert judgments, consensus and policy relevance, whereas objective weights were based on data driven method determined from mathematical models, and can be further divided into methods of loading and WA. Generally, the total weights for all indicators in loading method is not equal to one, or $\sum_j w_j \neq 1$, indicating that the correct term to be used is loading and not weight. The method has been widely used in multivariate analysis such as PCA and FA, and further examples of constructing deprivation indices using these methods were discussed [19-22]. On the other hand, the total weights for all indicators in WA methods are equal to one, or $\sum_j w_j = 1$. Examples of WA method that assign equal weights to each indicator were applied in [7-8]. For unequal weights, in[23-25] discussed several ways to determine weights. On the other hand, examples of weighting methods based on data entropy were discussed by [26-27] who suggested two different approaches of probability measurement.

In this paper, we apply several WA methods discussed by [24] in producing social development indices for international comparison. In addition, we introduce three new weighting methods which can also be used for constructing DI. Our suggested method for the estimation of entropy probability using empirical density function is more compatible with the definition of entropy as suggested by [28]. The empirical density function can be obtained by using a non-parametric method discussed in [29].To test the compatibility and suitability of our new methods, the methods are compared with PCA method which is a widely used method for constructing composite indices. Finally, the deprivation levels of eighty-one administrative districts (ADs) in Peninsular Malaysia are ranked based on the DIs produced from all methods considered in this study.

The data for developing socioeconomic DI were obtained from combined resources of census, administrative registration, vital statistics, insurance and safety data of eighty-one ADs in Peninsular Malaysia in year 2000, obtained from Department of Statistics Malaysia and Insurance Services Malaysia (ISM).

Based on literatures regarding socioeconomic deprivation and availability of data, twenty-three indicators were initially considered. The indicators were relevant to the concept of deprivation in the Malaysian context, besides being compatible and applicable over time. Pearson correlations were initially calculated to examine correlations between the twenty-three indicators. Low correlated indicators (r<0.3) were dropped and high correlated indicators were retained. If the indicators showed extremely high correlations (more than 0.95), indicators with relatively lower correlations were dropped to avoid multicollinearity. After retaining highly correlative indicators and checking for multicollinearity, thirteen indicators were selected. The thirteen indicators and their respective domains are shown in Table 1.

2. METHODOLOGY

2.1. Principle Component Analysis (PCA)

PCA is a mathematical model that linearly combines a large set of correlated variables into a substantially smaller set of uncorrelated variables called principal components, which represents a large proportion of information of the original data set. In this study, the DI derived from PCA was defined as the first component of PCA, accounting about 70.08% of the overall variation and correlating strongly with deprivation. The DI of the ith AD, $i = 1, 2, ..., n$ which was independent of all other principal components, can be calculated using a linear combination of thirteen indicators:

$$y_{1,i} = \sum_{j=1}^{13} w_{1,j} z_{ij},$$

where $w_{1,j}$ denotes the component loading for the *j*th indicator from the first PCA component and z_{ij} the *j*th standardized indicator for the *i*th AD calculated from

$$z_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}, \quad \overline{x}_j = \frac{\sum_i x_{ij}}{n} \text{ and } s_j^2 = \frac{\sum_i (x_{ij} - \overline{x}_j)^2}{n-1}.$$

Table 1. Thirteen indicators for DI

Indicator	Definition
	Family, economic and material deprivation domain
x_1	Total dependency ratio (100 population aged 15-64 years)
x_2	Percentage of private household without a car
x_3	Percentage of private household without a radio
x_4	Percentage of private household without a television
x_5	Percentage of private household without a fixed telephone line
x_6	Percentage of private household without a refrigerator
x_7	Percentage of private household without a washing machine
Basic utilities deprivation domain	
x_8	Percentage of private household without treated piped water
x_9	Percentage of private household without a toilet with flush system
	Employment deprivation domain
x_{10}	Percentage of unemployed population aged 15-64 years who are economically active
	Education deprivation domain
x_{11}	Percentage of population without education
x_{12}	Percentage of population aged 10 years and above without receiving SPM certificate
	Health deprivation domain
x_{13}	Relative risk of infant mortality

2.2. WA Method

Simple additive (SA) model has been widely used for aggregating indicators in the construction of composite indices. Before aggregation, the indicators of SA model are required to be standardized into unit-free indicators. In this study, we divide each indicator with its maximum value, $z_{ij} = x_{ij} / \max_i(x_{ij})$, so that the standardized indicators range from zero to one, $0 \le z_{ij} \le 1$. By using SA model in WA method, the composite DI for the *i*th AD can be defined as:

$$y_i = \sum_{j=1}^{13} w_j z_{ij}$$

where $0 \le w_j \le 1$ and $\sum_j w_j = 1$.

2.2.1. Standard Deviation (SD) Method

If we want to assign higher weights for indicators with larger standard deviation, the appropriate weight is:

$$w_j = \frac{s_j}{\sum_i s_i} ,$$

where $s_j^2 = \frac{\sum_i (z_{ij} - \overline{z}_j)^2}{n-1}$ and $\overline{z}_j = \frac{\sum_i z_{ij}}{n}$.

2.2.2. Correlation (CORR) Method

If the correlations between indicators are significant and we want to take into account the size of correlations between indicators, the appropriate weight is:

$$w_j = \frac{r_j}{\sum_j r_j},$$

where $r_j = \sum_k |r_{jk}|$ and $r_{jk} = \sum_i \left(\frac{z_{ij} - \overline{z}_j}{s_j} \right) \left(\frac{z_{ik} - \overline{z}_k}{s_k} \right)$. For this method, r_{jk} is the correlation between the jth and the kth indicators.

2.2.3. Mean Absolute Deviation (MAD) Method

We can also propose higher weights for indicators with larger deviation from the mean, which can be called MAD method:

$$w_j = \frac{\sum_i \left| z_{ij} - \overline{z}_j \right|}{\sum_k \sum_i \left| z_{ik} - \overline{z}_k \right|} .$$

2.2.4. Inter Quartile Range (IQR) Method

We can also propose higher weights for indicators with larger range of specified quartiles, say, between the 75th and the 25th quartiles. The method can be called IQR:

$$w_j = \frac{\left| Q_{0.75,j} - Q_{0.25,j} \right|}{\sum_j \left| Q_{0.75,j} - Q_{0.25,j} \right|} ,$$

where $Q_{0.75,j}$ and $Q_{0.25,j}$ are the third and first quartile of the jth indicator respectively.

2.2.5. Entropy Hwang & Yoon (ENHY) and Entropy Chen & He (ENCH) Methods

As an alternative, we can also assign the weight of WA method to depend on data entropy.Entropy method was introduced by[28] in describing information transfer. Entropy is a measure of uncertainty associated with a random variable, or a measure of average information one is missing when one does not know the value of the random variable. If there are n eventswith p_i probability ofevents $i = 1, 2, ..., n,$ the entropy is:

$$En = -\lambda \sum_i p_i \log(p_i),$$

where λ is afactorthat standardized theentropyandtakes the value betweenzero andone. In constructing composite index, the entropy of the jth indicator can be equated as:

$$En_j = -\lambda \sum_i p_{ij} \log(p_{ij}),$$

where p_{ij} is the probability of event z_{ij}. In [21-22] respectively proposed:

$$p_{ij} = \frac{z_{ij}}{\sum_i z_{ij}}$$

and

$$p_{ij} = \frac{|z_{ij} - \bar{z}_j|}{\sum_i |z_{ij} - \bar{z}_j|},$$

where $\bar{z}_j = \frac{\sum_i z_{ij}}{n}$. If each observation has equal chance of being selected, or $p_{ij} = n^{-1}$, the entropy is maximum and the standardized factor is $\lambda = \frac{1}{\ln(n)}$.

2.2.6. Entropy Kernel (ENK) Method

In this study, we suggest the empirical density function of a continuous z_{ij} for the estimation of entropy probability. Our proposed approach provides a more precise description of probability and is compatible with the definition of entropy, which is constructed in the environment of probability. Let z_{ij} be the observation values of n random sample from population with density function $f_j(z_{ij})$. Using a non-parametric approach suggested by[24], the estimate of $f_j(z_{ij})$ is:

$$\hat{f}_j\left(z_{ij}\right)=\hat{f}_{ij}=\frac{1}{nh}\sum_{l=1}^{n}k\left(\frac{z_{ij}-z_{lj}}{h}\right)$$

where the bandwidth is $h=1.06sn^{-1/5}$ and $k(.)$ is Gaussian kernel function. Further discussions on kernels and approximations of bandwidth can be found in.[24]. The probability of the *j*th indicator can be calculated as:

$$\hat{p}_{ij}=\frac{\hat{f}_{ij}}{\sum_i \hat{f}_{ij}}$$

so that the entropy is:

$$En_j=-\frac{1}{\ln(n)}\sum_i \hat{p}_{ij}\ln(\hat{p}_{ij}).$$

The weight for the *j*th indicator, which is determined by standardizing the degree of certainties, is:

$$w_j=\frac{1-En_j}{\sum_j(1-En_j)}.$$

The weight of entropy indicates that indicators with higher degree of uncertainties are given less priority. In this study, the Gaussian kernel was used to estimate the empirical density function.

3. RESULTS AND DISCUSSION

Table 2 displayed the weights produced by WA methods and PCA. The results indicate that different methods give different weight priority. The IQR and MAD methods stress more on households without flush toilets and populations without education, the PCA gives more weights for households without telephone lines and refrigerators, the correlation method assigns higher weights for households without telephone lines and populations with high dependency ratios, and the standard deviation method has higher priority for households without flush toilets and treated piped water. As for the entropy, the ENHY method gives more weights to households without treated piped water and flush toilets, the ENCH method to populations with high dependency ratios and households without treated piped water, and the Gaussian kernel method to households without treated piped water and televisions.

Table 2. Weights derived from PCA and WA methods

Methods		x_1	x_2	x_3	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{11}	x_{12}	x_{13}
PCA	Eigen vector	0.3029	0.2952	0.2774	0.2495	0.3068	0.3034	0.2927	0.2457	0.2619	0.2564	0.2949	0.2256	0.2783
WA	SD	0.0648	0.0679	0.0685	0.0607	0.0816	0.0778	0.0792	0.1029	0.1116	0.0900	0.0889	0.0498	0.0563
	IQD	0.0529	0.0775	0.0742	0.0560	0.0788	0.0946	0.0878	0.0468	0.1397	0.0920	0.0978	0.0491	0.0529
	MAD	0.0612	0.0896	0.0634	0.0466	0.0825	0.0685	0.0962	0.0293	0.1519	0.1027	0.1070	0.0478	0.0532
	CORR	0.0844	0.0822	0.0769	0.0694	0.0852	0.0843	0.0814	0.0730	0.0774	0.0714	0.0821	0.0634	0.0690
	ENHY	0.0189	0.0199	0.0298	0.0475	0.0398	0.0502	0.0334	0.5311	0.0926	0.0563	0.0446	0.0116	0.0242
	ENCH	0.0946	0.0850	0.0735	0.0774	0.0604	0.0652	0.0666	0.0892	0.0728	0.0828	0.0790	0.0737	0.0797
	ENK	0.0657	0.0360	0.0918	0.1074	0.0783	0.0838	0.0561	0.2151	0.0194	0.0562	0.0509	0.0787	0.0604

The weights derived from all methods were applied to produce DI of eighty-one ADs in Peninsular Malaysia, the ADs were then ranked from the least deprived to the most deprived, and the ranked ADs were classified into four quartiles where each quartile has an approximately equal number of ADs (twenty ADs per quartile). Fig. 1provides the choropleth maps of ADs in Peninsular Malaysiawith their associated quartiles for all methods. The lighter shades indicate areas with relatively lower level of deprivation and the darker shades indicate areas with relatively higher level of deprivation. In general, the most deprived ADs for all methods were located in the northeast of Peninsular Malaysia, as indicated by the darkest shade while the least deprived ADs were mainly located in the west coast as indicated by the lightest shade.

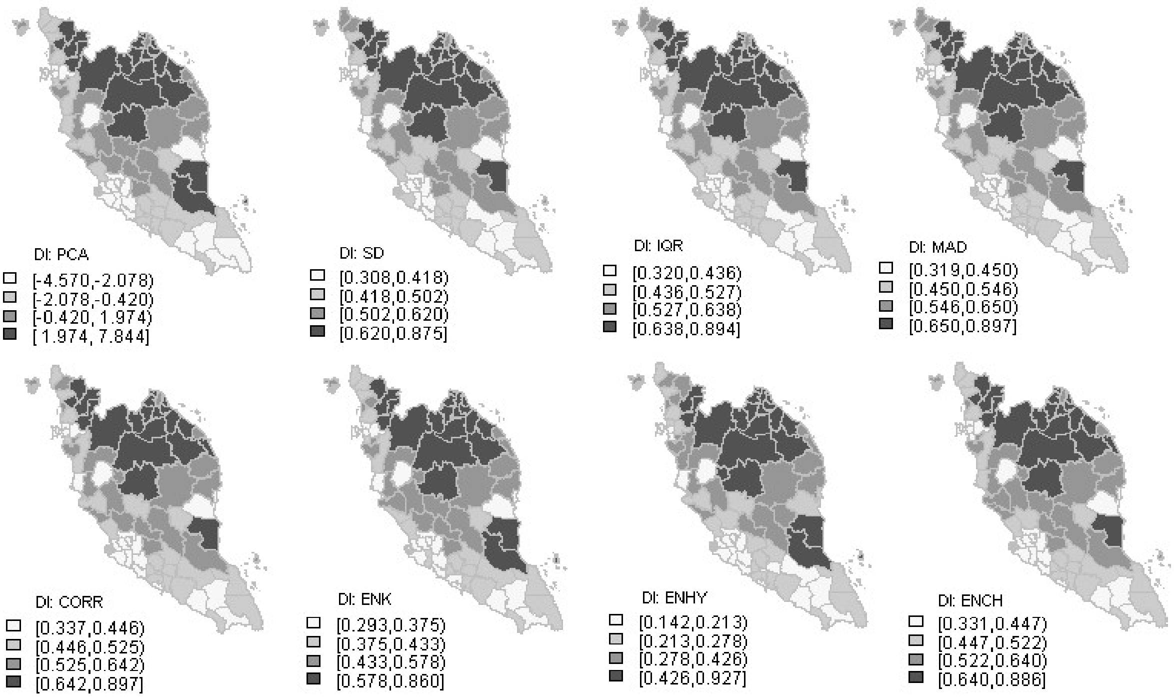

Fig.1. Spatial distribution of deprivation index (DI) according to quartiles for all methods (PCA-principal component analysis, SD-standard deviation, IDR-inter quarter range, MAD-mean absolute deviation, CORR-correlation, ENK-entropy Kennel, ENHY-entropy Hwang and Yoon and ENCH-entropy Chen & He)

The results of Spearman correlation test in Table 3 show extremely high correlations ($r>0.91$) between all indices, indicating that the DI of all methods produce similar geographical distribution and can be used as alternatives to produce composite DI or other indices.

Undeniably, DI is closely related to health inequality. Different selections of indicators or different techniques and methods of normalization, standardization, weighting and aggregation will provide different impacts on the resulted indices. Accordingly, in this study, we are focusing on producing weights from a variety of weighting techniques and comparing the resulted indices based on such weights. Thirteen indicators from various dimensions of deprivation were combined in eight different types of weighting method, in order to explain the distribution of geographical deprivation indices across eighty-one ADs in Peninsular Malaysia in year 2000. The SD and the proposed MAD methods are more suitable for data with no extremes or outliers, the suggested IQR method is more appropriate if we are interested in the 50% or other arbitrary percentages of the central distribution and the CORR method is more proper if we want to account for the strength between indicators. As for the entropy method, since we are dealing with weights based on uncertainties, the estimation of entropy probability using empirical distribution function and Gaussian kernel function which is more compatible with the definition of entropy is proposed in this study. Finally, the PCA assigns weights by accounting the maximal variability of data. The overall concept carried out in this paper is related to the dispersion of data, i.e. the higher the dispersion, the higher the weights and vice versa. For the entropy method, higher weights are assigned to indicators with lower uncertainties.

Table 3. Spearman correlation test

	R_PCA	R_SD	R_IQR	R_MAD	R_CORR	R_ENHY	R_ENCH	R_ENK
R_PCA	1							
R_SD	0.9957	1						
R_IQR	0.9953	0.9980	1					
R_MAD	0.9921	0.9956	0.9992	1				
R_CORR	0.9988	0.9989	0.9979	0.9950	1			
R_ENHY	0.9196	0.9395	0.9171	0.9074	0.9299	1		
R_ENCH	0.9981	0.9992	0.9972	0.9947	0.9995	0.9347	1	
R_ENK	0.9822	0.9846	0.9730	0.9642	0.9844	0.9702	0.9848	1

4. CONCLUSION

To the best of our knowledge, studies on data-driven weighting methods have not been carried out in the construction of DI. Even though the methods of SD, CORR, ENHY and ENCH have been widely used in multi-criteria decision making and construction of development indices, the methods have not been applied in the construction of DI. In addition, our three proposed methods, the MAD, IQR and ENK have been proven to be applicable for the construction of area-based composite socioeconomic index. Therefore, the methods discussed and proposed in this study can be used as alternatives to produce composite DI or other indices. In addition, a variety of other models can be created by replacing the weight of WA method with other statistical indicators or by replacing the probability of entropy with other statistical functions, or by replacing the empirical density function of entropy probability with other kernel functions.

5. ACKNOWLEDGEMENTS

The authors would like to thank Department of Statistics Malaysia for providing the data to make this study possible. The first author is grateful to Public Service Department of Malaysia and Department of Statistics, Malaysia for their sponsorship and support for this research. Authors also would like to indicate their appreciation UniversitiTeknikal Malaysia Melaka for the financial and technical support.

6. REFERENCES

[1] Fam SF, Noriszura I, Abdul A J. Geographical and socio-economic analysis in Peninsular Malaysia. The Social Sciences, 2017, 12(9):1695-1704

[2] Fam S F, Ismail N, Jemain A A. Where are Peninsular Malaysia's most deprived areas? International Statistical Institute, South East Asia Regional Statistics Conference, 2014, pp. 1-10

[3] Fam SF, Jemain AA, Ibrahim K. Spatial analysis of socioeconomic deprivation in Peninsular Malaysia. International Journal of Arts and Sciences, 2011, 4(17):241-255

[4] Noble M, Barnes H, Wright G, Roberts B. Small area indices of multiple deprivation in South Africa. Social Indicators Research, 2010, 95(2):281-297

[5] Carstairs V. Deprivation indices: Their interpretation and use in relation to health. Journal of Epidemiology and Community Health, 1995, 49(Suppl 2):S3-8

[6] Fam SF, Jemain AA, Ibrahim K. The association between material deprivation and relative risk of infant mortality in Peninsular Malaysia. International Journal of Arts and Sciences, 2011, 4:257-269

[7] Townsend P., Phillimore P., Beattie A. Health and deprivation: Inequalities and the north. London: Croom Helm, 1988

[8] Carstairs V, Morris R. Deprivation: Explaining differences in mortality between Scotland and England and Wales. British Medical Journal, 1989, 299(6704):886-889

[9] Jarman B. Identification of underprivileged areas. British Medical Journal, 1983,

287(6385):1705-1709

[10] Testi A, Ivaldi E. Material versus social deprivation and health: A case study of an urban area. European Journal of Health Economics, 2009, 10(3):323-328

[11] Noble MI, Lloyd MY, Wright GE, Dibben CH, Smith GE. Developing deprivation measures for Northern Ireland. Journal of the Statistical and Social Inquiry Society of Ireland, 2001, 31:1-35

[12] McConnachie A, Weir C. Evaluation of statistical techniques in the Scottish index of multiple deprivation. Glasgow: University of Glasgow, 2005

[13] Noble M, Babita M, Barnes H, Dibben C, Magasela W, Noble S, Ntshongwana P, Phillips H, Rama S, Roberts B, Wright G. The provincial indices of multiple deprivation for South Africa 2001. Technical report, England: University of Oxford, 2006

[14] Gilthorpe MS. The importance of normalisation in the construction of deprivation indices. Journal of Epidemiology and Community Health, 1995, 49(Suppl 2):S45-50

[15] Gordon D. Census based deprivation indices: Their weighting and validation. Journal of Epidemiology and Community Health, 1995, 49(Suppl 2):S39-44

[16] Saunders J. Weighted Census-based deprivation indices: Their use in small areas. Journal of Public Health, 1998, 20(3):253-260

[17] Bell N, Schuurman N, Hayes MV. Using GIS-based methods of multicriteria analysis to construct socio-economic deprivation indices. International Journal of Health Geographics, 2007, 6(17):1-19

[18] Folwell K. Single measures of deprivation. Journal of Epidemiology and Community Health, 1995, 49(Suppl 2):S51-56

[19] Fukuda Y, Nakamura K, Takano T. Higher mortality in areas of lower socioeconomic position measured by a single index of deprivation in Japan. Public Health, 2007, 121(3):163-173

[20] Baum S. Measuring socio-economic outcomes in Sydney: An analysis of census data using a general deprivation index. Australasian Journal of Regional Studies, 2004, 10(1):105-131

[21] Rey G, Jougla E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997–2001: Variations with spatial scale, degree

of urbanicity, age, gender and cause of death. BMC Public Health, 2009, 9(1):1-12

[22] Exeter DJ, Zhao J, Crengle S, Lee A, Browne M. The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand. Plos One, 2017, 12(8):1-19

[23] Kantiray A. On the measurement of certain aspects of social development. Social Indicators Research, 1989, 21(1):35-92

[24] Ray AK. Measurement of social development: An international comparison. Social Indicators Research, 2008, 86(1):1-46

[25]Ray AK. India's social development in a decade of reforms: 1990–91/1999–2000. Social Indicators Research, 2008, 87(3):409-425

[26] Tzeng G. H., Huang J. J. Multiple attribute decision making: Methods and applications. Florida: CRC Press, 2011

[27] Chen JJ, He Z. Using analytic hierarchy process and fuzzy set theory to rate and rank the disability. Fuzzy Sets and Systems, 1997, 88(1):1-22

[28] Shannon C. E., Weaver W. The mathematical theory of communication. Urbana: University of Illinois Press, 1949

[29] Silverman B. W. Density estimate for statistics and data analysis.Abingdon: Routledge, 2018

How to cite this article:
Fam S F, Ismail N, Maukar A L, Yanto H, Prastyo D D, Jemain A A, Chuan Z L. Weighting methods in the construction of area deprivation indices. J. Fundam. Appl. Sci., 2018, *10(6S), 2655-2668.*