ISSN. 0972 - 8406
The NEHU Journal, Vol XVI, No. 1, January - June 2018, pp. 19-30

The Dilemma of Classification of Income Levels in Social Research

Jenny Jami*

Abstract

Income is a widely used parameter in social researches. Most community studies generally collect information on the household income for better understanding of the socio-economic status of the people. Income is usually classified as low, middle and high where thresholds were selected depending on the population under study. Thus, classification of income levels is mostly arbitrary. The present paper attempts to examine the discrepancies in results that may arise due to differential classification of income levels. For the purpose of the study, analysis of a fieldwork-based data set of 708 individuals (20 – 70 years) from 470 households based on different income classifications and various socio-demographic, behavioral and nutritional indicators were used. It is observed that varying classification of income levels may not influence the interpretation of results in case of discrete or continuous variables, however, it is found to skew the direction of interpretation of results to a considerable extent in categorical variables.

Keywords: Income, middle class, categorical variables, continuous variables

Introduction

Income is one of the most important indicators of socio-economic status of an individual or a population. Given its widespread importance, it is one of the most frequently used

* Jenny Jami, Department of Anthropology, North-Eastern Hill University, Shillong.
Email: jennyjami77@gmail.com

parameters in many studies, especially in the humanities and social science researches. On an International scale, the World Bank (Fantom & Serajuddin, 2016) classifies countries as low income, lower middle income, upper middle income, and high income based on the gross national income (GNI). However, for smaller scale studies of income distribution such as region-based or community-based studies, the household income is generally considered for categorizing income levels where the median values often provide generalized cut-off points (Noss, 2013; United Nations, 2011; Desai *et al.,* 2010; McNeil, 1998). While it may be convenient to classify the income levels as low and high through the median value, such a classification is rendered impractical with the emergence of the middle class. Defining the middle class is crucial not only in itself but also because it sets the thresholds for the high and low income groups. However, it is often difficult to decide on the identity of the middle class (Elwell, 2014; Meyer & Birdsall, 2012). As Elwell (2014) explains, "how far the middle class stretches above and below the median is the question" (p. 4). These factors add to the knowledge that the actual dynamics involved in income classification are complicated where the criteria employed for income classification is seen to differ across and within countries, states or regions. In most cases, income levels are often found to be arbitrarily classified based on means, medians, percentiles, quartiles or quintiles (United Nations, 2011). However, there is no international framework for micro-level household income statistics (United Nations, 2011). It is also further observed that the rationale for setting threshold levels for income is often unclear (Elwell, 2014; Veit-Wilson, 1998). This is a serious limitation as such classification of income levels often renders comparability in income data difficult or may also even lead to inaccurate reporting of results (United Nations, 2011). With these considerations, the present paper seeks to examine discrepancies in results that may arise due to differential classification of income levels.

Methods:

Income is often found to be associated with socio-demographic factors such as gender, age, region, marital status, family size, education and occupation (Bobbitt-Zeher, 2007; Jagsi *et al.*, 2013; Chetty *et al.*, 2016; Pradhan *et al.*, 2000; Madalozzo, 2008; Chen & Yang, 2016; Blanden & Gregg, 2004). It is also found to be associated with behavioral patterns such as physical activity, drinking and smoking (Wang *et al*., 2018; Kari *et al.,* 2015; Kim and So, 2014; Lemstra *et al.*, 2009; Auld, 2005) as well as with health and nutritional status of populations (Abraham *et al.*, 2017; Godoy *et al.*, 2005; Kennedy *et al.*, 1998). Thus, for the purpose of the present paper, a cross-sectional study was conducted among the Lotha Naga adults in Wokha Town of Nagaland and analysis of the urban-based data set of 708 adult individuals (20 – 70 years) from 470 households based on different income classifications and various socio-demographic, behavioral and nutritional indicators were used. Data on these variables were collected directly from the participants through structured interviews. Socio-demographic data included gender, age groups, place of birth, marital status, family size, education, income and occupation. Gender was classified as male and female. Age is classified into five groups with equal class intervals. Place of birth was categorized as urban and rural. Marital status included unmarried, married or divorced/separated/widowed (DSW) categories. Family size was classified as small (≥4), medium (5-6) or large (≥7) based on the total number of family members. Education was classified as illiterate, primary, secondary, higher secondary, undergraduate and above (InSCED, 2014). Occupation was arbitrarily divided into eight categories – Government employee, non-government employee, self-employed, agriculturist or labourer, student, homemaker, retired service employees and unemployed. Behavioral data included levels of physical activity (IPAQ, 2005), alcohol and tobacco use as well as aspects of health and morbidity (NSSO, 2004). Anthropometric measurements such as weight, height, waist circumference, hip circumference, biceps, triceps, sub-scapular and supra-iliac were measured directly from

the individuals using standard techniques (Weiner and Lourie, 1981; Centers for Disease Control and Prevention, 2012; WHO, 2011) and accordingly, nutritional indicators such as body mass index, fat mass index, fat free mass index, conicity index, waist-hip ratio and waist-height ratio were calculated.

Income levels were classified based on commonly used summary measures of income level i.e., median, percentiles, quartiles and quintiles (UN, 2011) of the per capita monthly income of the households. For the analysis, these were arbitrarily classified as:

1. Income levels (Median): Income was classified into two categories - low and high, using the median as the threshold value.

2. Income levels (Percentiles): Based on percentiles, income was categorized into three groups i.e., low (below 50th percentile), middle (50th – 75th percentile) and high (above 75th percentile).

3. Income levels (Quartiles): Income was classified into four income groups viz., Low, Lower Middle, Upper Middle and High Income Group based on quartiles of the per capita monthly income of the households.

4. Income levels (Quintiles): In this case, distribution of per capita household monthly income was based on quintiles. Accordingly, income was classified as poor (lowest quintile), low income (second), middle income (third), upper-middle income (fourth) and high income (highest).

Pearson's chi-square tests of independence were performed to examine the association between the different income levels with each of the socio-demographic and behavioral variables and One-way Analysis of Variance (ANOVA) was performed separately for the anthropometric measurements and income levels to examine whether each of these variables differed based on their income levels. SPSS software was used for all computational purposes.

Results:

The income distribution of respondents based on different classification methods are shown in **Table 1**. It was observed that when income was classified through the median values, about 54.24% of the participants were grouped under the low income group (LIG) while 45.76% were classified under the high income group (HIG). When classification was based on percentiles, it was observed that income levels under low, middle and high income groups were 45.76%, 29.38%, 24.86% respectively. Based on quartiles, it is observed that the percentage of participants were higher for upper middle income group (29.38%) followed by lower middle group (24.72%), high income group (24.86%) and low income group (21.05%). The percentage of participants under first, second, third, fourth and fifth quintiles were found to be 17.80%, 16.67%, 23.45%, 21.75% and 20.34% respectively. A comparison of the results on relationships between income levels and few selected variables (nominal or ordinal scale) based on different income classifications are shown in **Table 2.** It is observed that family size, education and occupation showed highly significant χ^2 values at all income levels which indicate that if income is significantly very highly related to a variable ($p < 0.0001$), the results would mostly yield similar interpretations irrespective of the income categorization. Conversely, it was found that no significant χ^2 values were seen in marital status, alcohol and tobacco use, hospitalization and treatment behavior for all the income levels indicating that it may be possible to screen variables that are truly not associated with income using any of the classification. Interpretation of results gets tricky when variables are found to be associated at the significance levels (α) of 0.05 or 0.01. In the present analysis, it is observed that it showed uniform interpretations of results for place of birth, however, varying results are seen in some of the other variables depending on the method of income classification. For instance, if the study considers income levels based on quartiles or quintiles, it can be interpreted that gender, age groups and levels of physical activity are found to be significantly related to income.

On the contrary, one would end up interpreting that there is no statistically significant relationship between income and gender, age groups or levels of physical activity if income levels are based on median or percentiles. Alternatively, it can be interpreted that there is a relation between between income and type of hospital if income was based on quartiles, median and percentiles but not for quintiles. Self-reported morbidity was significantly related to income when classification was on the basis of quartiles, percentiles and quintiles but not for median. Income was found to be associated with treatment from government sources only when classification was based on quartiles. **Table 3** shows a comparison of the results on relationships between income levels and the selected variables (interval or ratio scale) based on different income classifications. It is observed that when the test variables are discrete or continuous, the results mostly yielded similar interpretations irrespective of the income categorization. Results that produced different interpretation are seen only for the variables- height and triceps which were found to be significantly related to income based on quartiles and quintiles but not for median and percentiles.

From the above comparisons, it can be observed that degree of association varies with change in the method of income classification. Also, while it is observed that varying classification of income levels may not influence the interpretation of results in case of discrete or continuous variables, however, it is found to skew the direction of interpretation of results to a considerable extent in nominal or ordinal data. This observation is especially crucial as categorical variables are mostly used in social researches. Though the probability of type 1 or type 2 errors cannot be ruled out in any of the classifications, the chances of making these errors might be reduced by identifying the correct operational classification.

Comments:

Given the widespread intricacies of varying socio-economic conditions such as the concentration of wealth, the cost of living,

economic disparities, etc. , it is difficult to decide on a standard method of classifying income levels that can be well fitted for general application, especially in field-work based social science researches. Despite these challenges, it would be very beneficial if some simple standards are set for classification of income across different settings. One such standard that probably can be used is the Gini coefficient[1] which is one of the most commonly used summary indicator of income dispersion. Given the assumptions underlying the Gini coefficient, it is postulated that the smaller the Gini coefficient the more equal the distribution of income (United Nations, 2011). Is it possible then, that this may be used to decide on the number of income categories assuming that if the Gini coefficient is less, fewer income categories are needed to bring out the differences in the population; in which case, can division of the population as low and high, depending on the median value be justified or if the income disparity is high, would more income divisions be needed to understand the extent to which it renders its influence? Whether such theoretical assumptions may lead to a robust standard is uncertain, however, these considerations may be able to provide us with a rationale that can be used to divide income categories. Nonetheless, the objective of this paper is not to suggest methods of income classification but to bring out discrepancies in results that may arise due to differential classification of income levels and to highlight the need to formulate acceptable criteria for income classification that may be statistically viable for interpreting results.

References

Abraham, J., Navaneetha, N., Johns, F., Aiyappan, R., Mili, M., Shibu, P., & Mathew, E., 2017. 'Nutritional status of older adults in a community in Pathanamthitta district of Kerala', *International Journal of Research in Medical Sciences*, 6(1): 210-215.

Auld, M.C., 2005. 'Smoking, drinking, and income', Journal of Human Resources, 40(2):.505-518.

[1] Gini coefficient for the present population is 0.377

Blanden, J., & Gregg, P. 2004. 'Family income and educational attainment: a review of approaches and evidence for Britain', *Oxford Review of Economic Policy*, 20(2): 245-63.

Bobbitt-Zeher, D., 2007. 'The gender income gap and the role of education', *Sociology of education*, 80(1): 1-22.

Centers for Disease Control and Prevention. 2012. National Health and Nutrition Examination Survey (NHANES) anthropometry procedures manual: 2011. CDC, USA.

Chen J, & Yang H. 'Geographical mobility, income, life satisfaction and family size preferences: an empirical study on rural households in Shaanxi and Henan Provinces in China', Social indicators research, 129(1): 277-90.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A. and Cutler, D. 2016. 'The association between income and life expectancy in the United States, 2001-2014', Jama, 315(16): 1750-1766.

Desai, S.B., Dubey, A., Joshi, B.L., Sen, M., Shariff, A., & Vanneman, R. 2010. 'Human development in India.' New York: Oxford University.

Elwell, C. K. 2015. 'The distribution of household income and the middle class', Current Politics and Economics of the United States, Canada and Mexico, 17(4), 657.

Fantom, N. J., & Serajuddin, U. 2016. The World Bank's classification of countries by income. http://documents.worldbank.org/curated/en/408581467988942234/pdf/WPS7528.pdf, accessed on 25/10/2017

Godoy, R., Byron, E., Reyes-García, V., Vadez, V., Leonard, W.R., Apaza, L., Huanca, T., Pérez, E., & Wilkie, D. 2005. 'Income inequality and adult nutritional status: anthropometric evidence from a pre-industrial society in the Bolivian Amazon', *Social Science & Medicine,* 61(5): 907-919.

InSCED. 2014. Indian Standard Classification of Education. Ministry of Human Resource Development, Department of Higher Education, New Delhi.

IPAQ Research Committee. 2005. 'Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)–short and long forms'.

Jagsi, R., Griffith, M.K.A., Stewart, A., Sambuco, M.D., DeCastro, M.R. and Ubel, P.A. 2013. 'Gender differences in salary in a recent cohort of early-career physician-researchers', Academic medicine*: journal of the Association of American Medical Colleges,* 88(11).

Kari, J.T., Pehkonen, J., Hirvensalo, M., Yang, X., Hutri-Kähönen, N., Raitakari, O.T., & Tammelin, T.H., 2015. 'Income and physical activity among

adults: evidence from self-reported and pedometer-based physical activity measurements', PloS one, 10(8): e0135651.

Kennedy, B.P., Kawachi, I., Glass, R. and Prothrow-Stith, D., 1998. 'Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis', Bmj, 317(7163): 917-921.

Kim, I.G. & So, W.Y., 2014. 'The relationship between household income and physical activity in Korea', *Journal of physical therapy science*, 26(12): 1887-1889.

Lemstra, Mark, Johan Mackenbach, Cory Neudorf, Ushasri Nannapaneni, and Anton Kunst. 2009. 'Daily smoking in Saskatoon: the independent effect of income and cultural status', Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique, 51-54.

Madalozzo R. 2008. 'An analysis of income differentials by marital status', Estudos Econômicos (São Paulo), 38(2): 267-92.

McNeil, J. 1998. 'Changes in median household income: 1969 to 1996', US Department of Commerce, Economics and Statistics Administration, Bureau of the Census.

Meyer, C., & Birdsall, N. 2012. 'New estimates of India's middle class,' CGD Note, Center for Global Development, Washington, DC.

Noss, A. 2013. 'Household Income: 2012', United States Census Bureau, US Department of Commerce 12 (2).

National Sample Survey Organization. 2004. Morbidity and Health Care. NSSO, New Delhi.

Pradhan, B.K., Roy, P.K., Saluja, M.R. and Venkatram, S., 2000. 'Rural-urban disparities: Income distribution, expenditure pattern and social sector', Economic and Political Weekly: 2527-2539.

United Nations. Canberra Group on Household Income Statistics, 2011. Canberra Group Handbook on Household Income Statistics. UN.

Viet-Wilson, J. 1998. Setting Adequacy Standards, How Governments Define minimum income standards. The Policy Press, Bristol.

Wang, Q., Shen, J.J., Sotero, M., Li, C.A., Hou, Z. 2018. 'Income, occupation and education: Are they related to smoking behaviors in China?', PloS one, 13(2):e0192571.

Weiner, J. S., & Lourie, J. A. 1981. *Practical human biology*. Academic Press, London.

World Health Organization. 2011. Waist circumference and waist-hip ratio: Report of a WHO expert consultation. Geneva: WHO.

Table 1: Income distribution of respondents based on different classification methods

Income levels	Respondents					
	Male (n=354)		Female (n=354)		Total (n=708)	
	No.	%	No.	%	No.	%
Income levels (Medians)						
Low Income Group	189	53.39	195	55.08	384	54.24
High Income Group	165	46.61	159	44.92	324	45.76
Income levels (Percentiles)						
Low Income Group	165	46.61	159	44.92	324	45.76
Middle Income Group	104	29.38	104	29.38	208	29.38
High Income Group	85	24.01	91	25.71	176	24.86
Income levels (Quartiles)						
Low Income Group	63	17.80	86	24.29	149	21.05
Lower Middle Income Group	102	28.81	73	20.62	175	24.72
Upper Middle Income Group	104	29.38	104	29.38	208	29.38
High Income Group	85	24.01	91	25.71	176	24.86
Income levels (Quintiles)						
Lowest	53	14.97	73	20.62	126	17.80
Second	65	18.36	53	14.97	118	16.67
Third	96	27.12	70	19.77	166	23.45
Fourth	69	19.49	85	24.01	154	21.75
Highest	71	20.06	73	20.62	144	20.34

Table 2: Comparison of Chi-square values based on different income groups

Nominal/Ordinal Variables	χ^2 values			
	Income levels (Median)	Income levels (Percentiles)	Income levels (Quartiles)	Income levels (Quintiles)
Gender	0.205	0.316	8.561*	10.157*
Age Groups	5.846	12.409	26.081**	31.237**
Place of birth	7.920**	8.360*	10.583**	14.007**
Marital Status	0.928	2.658	7.893	11.792
Family Size	1.406	33.894****	40.144****	44.356****
Education	38.802****	50.538****	56.680****	62.710****
Occupation	56.306****	73.464****	106.8****	123.7****
Levels of physical activity	5.102	6.567	13.171*	15.442*
Alcohol Use	0.718	1.065	2.488	3.061
Tobacco (smoking)	2.862	2.895	2.910	4.675
Tobacco (chewing)	3.897	2.895	3.992	4.185
Hospitalization	0.665	0.891	0.994	0.694
Type of hospital	9.684**	12.567**	13.424*	14.939
Self-reported morbidity	0.130	9.763**	9.815*	10.803*
Treatment on medical advice	0.570	2.297	5.790	4.828
Treatment from government sources	0.083	2.289	9.838*	8.763

*p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001

Table 3: Comparison of ANOVA F- Statistics based on different income groups

Discrete/ Continuous Variables	ANOVA F- statistics			
	Income levels (Median)	Income levels (Percentiles)	Income levels (Quartiles)	Income levels (Quintiles)
Weight	7.459**	7.078***	5.591***	4.586***
Height	1.823	1.655	4.085**	4.086**
Waist Circumference	9.035**	8.008****	5.697***	4.558***
Hip Circumference	9.069**	8.491****	5.837***	4.727***
Biceps	14.056****	11.074****	7.576****	5.353****
Triceps	3.104	2.309	3.322*	3.259**
Sub-scapular	11.438***	7.228***	5.223***	4.740***
Supra-iliac	16.394****	10.969****	7.304****	6.283****
Sum of skinfolds	13.044****	8.872****	6.262****	5.418****
Body Mass Index	5.420*	5.044**	3.358*	2.986*
Fat Mass Index	9.306**	7.175***	5.010**	4.500***
Fat Free Mass Index	0.115	0.165	0.530	0.575
Conicity Index	6.054**	5.091**	3.652**	2.972*
Waist-Hip Ratio	3.340	2.497	2.045	1.839
Waist-Height Ratio	5.715*	4.995**	3.329*	3.038*
*p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001				

*****