BIBECHANA
ISSN 2091-0762 (Print), 2382-5340 (Online)
Journal homepage: http://nepjol.info/index.php/BIBECHANA
Publisher: Department of Physics, Mahendra Morang A.M. Campus, TU, Biratnagar, Nepal

Dichotomization of quantitative variables in poverty analysis

Krishna Prasad Acahrya[1], Shankar Prasad Khanal[1*], and Devendra Chhetry[1]

[1]Central Department of Statistics, Institute of Science and Technology, Tribhuvan University, Kirtipur, Nepal
Email: drshankarcds@gmail.com

Article Information:
Received: Dec 15, 2021
Accepted: Jan 30, 2022

Keywords:
Dichotomization
Headcount index
Poverty gap index
Square poverty gap index
Vulnerable

ABSTRACT
It has been proposed four schemes of dichotomization for the four household level quantitative variables – area of land holding, geographic accessibility to the nearest market centre, number of children under 15 and number of literate members of working-age – with justification in the selection of threshold value for each variable to dichotomize into disadvantaged and advantaged group of households using the Nepal Living Standard Survey 2010/11 data with 5988 households and 28,670 of their household members. Association of each dichotomized variable with household level poverty status (poor/non-poor) was found highly significant. Finally, the proposed schemes of dichotomization have tested empirically for their ability to differentiate the poor people into two categories - 'more vulnerable' and 'less vulnerable' - by fist estimating the three measures of poverty – head count index, poverty gap index and squared poverty gap index - of each group of population and comparing the estimated measures between the disadvantaged and advantaged group of populations. Statistical analysis has been performed by using IBM SPSS version 20. To a large extent the proposed schemes of dichotomization have found to differentiate the poor people into two groups; for example, the head count index of the disadvantaged group of the number of children under 15 is 3.1 times higher than that of the advantaged group. The results of this paper are expected to be useful to the policy makers and development planners of Nepal for focusing their poverty reduction program towards the more vulnerable group of population as well as academician.

DOI: https://doi.org/10.3126/bibechana.v19i1-2.46407

1. Introduction

Several statistical methods are available for assessing the association of a dichotomous variable with a set of quantitative variables, not necessarily all continuous. When dichotomous variable is treated as dependent variable and the set of quantitative variables is treated as independent variables, the association can be assessed through *logistic regression* which assesses the effect of independent variables on dependent variable simultaneously. When the dichotomous variable is treated as grouping variable and the set of quantitative variables are

treated as test variables, the association can be assessed through *independent samples t-test* which assess the association by comparing the two group means of each quantitative variable. The former method is more rigorous from both theoretical and practical point of view than the later method. However, the use of later method faces the problem of normality assumption which means each test variable has to follow normal distribution within each of the two categories of the grouping variables since the t-test was developed under this normality assumption. Conceptually the normality assumption is hard to justify when a test variable is discrete. When normality assumption fails, instead of independent samples t-test it is a common practice to use Mann-Whitney test. Nonetheless, the Mann-Whitney non parametric test transforms the quantitative variables into their rank orders and the test works on rank ordered data, and the test results are not easily understandable to wider users. The association between a binary variable with a set of quantitative variables having only two levels, the association can also be assessed through biserial correlation. The less frequently used method is to first dichotomize each independent variable using a rationally defined threshold value and then use either logistic regression or use Chi-square test for the dichotomous variable and each dichotomized quantitative variable. Several scholars have discussed the advantages and disadvantages on this less frequently used method [1-3]. Sometimes, dichotomization of quantitative variable is absolutely necessary. For example, dichotomization of per capita consumption expenditure using poverty line as threshold value is absolutely necessary in measuring monetary poverty.

This paper has two-fold objectives. First objective is to dichotomize the four household level quantitative variables by justifying in the selection of threshold value for each variable, and assess the association of the four dichotomized variables with the dichotomous variable - poverty status (poor/non-poor). Second objective is to estimate the three measures of poverty – *head count index*, *poverty gap index*, and *squared poverty gap index* – for all the four dichotomized variables in order to investigate the ability to differentiate poor peoples into 'more vulnerable' and 'less vulnerable' through the estimated measures of poverty.

2. Materials and Methods

The main source of data of this study is NLSS III which provides household level data on socio-economic and demographic variables of 5,988 households and 28,670 household members. The available data on the variable 'household poverty status' was taken as binary variable by assigning code 1 for poor and 0 for non-poor. In this study a household is defined as *poor (non-poor) depending upon the per capita expenditure of the household members falls below (above) the poverty line of NRs 19,261*. The un-weighted and weighted proportions of poor households were correspondingly 18.5% and 20.0%. Similarly, un-weighted and weighted proportions of poor household members were correspondingly 23.4% and 25.2%.

Out of many household level variables that influence the monetary poverty, only the following four quantitative variables are considered in the present study.

1. Area of land holding
2. Geographic accessibility to market center (defined in the present study by time taken in minutes to reach the nearest market irrespective of transport mode)
3. Number of children under 15
4. Human capital (defined by number of literate members of working-age (15 – 64 years)).

The available data on the above four variables were dichotomized. The main reason for dichotomization of each of these variables is to divide the households into two groups:

advantaged and *disadvantaged group* with respect to each variable. The rationale behind such demarcation of households is that the disadvantaged group of households would be in a more difficult position to escape out of poverty than the advantaged group of households. The process of dichotomization, particularly choosing the threshold value for each of the four quantitative variables is rationalized below.

2.1 Dichotomizing households by area of land holding

The available data on land holding is highly skewed (skewness = 5.55) with extremely high measure of kurtosis (excess of kurtosis = 65.15), and considerable number of households had no lands. As a result, analysis based on original data suffers from various problems. Moreover, farm size is not a good determinant of poverty [4]. In this context, analysis based on dichotomizing the quantitative variable is more sensible than analyzing the data as it is. In view of this fact, the threshold value for this quantitative variable was chosen to be 0 which demarcates households into two groups - one group of households each had no land (disadvantaged group) and the other group of households each had land (advantaged group).

2.2 Dichotomizing households by the number of children under 15

Children under 15 are considered as dependent population in the sense that their basic needs have to be fulfilled by their parents. In this context, large number children would be burden to parents. As a result large number of children aggravates poverty [5]. However, small number of children is desirable. The ideal number of children responded by women respondents on an average was 2.1 and by men respondents was 2.3 in 2011 [6]. Based on these results, the threshold value of 2 is used for dichotomizing the quantitative variable. This threshold value demarcates the households into two groups - one group of households each had more than two children (disadvantaged group) and the other group of households each had less than or equal to two children (advantaged group).

2.3 Dichotomizing households by the number of literate working-age members

In the context of Nepal, number of illiterate persons in a household is major disadvantage of the poor households [5]. In view of this fact, the household level quantitative variable "the number of literate members of working-age" was selected and converted it into dichotomous variable by grouping the households into two groups: one group of households each had no literate member of working-age (disadvantaged group) and the other group of households each had at least one literate member of working-age (advantaged group).

2.4 Dichotomizing households by access to nearest market

The available data on access to nearest market center is highly skewed (skewness = 3.46) with high measure of kurtosis (excess of kurtosis = 16.74). The mean and median of the time taken to reach market center in minutes are correspondingly 80.63 and 30.00. The analysis based on dichotomizing the quantitative variable is more sensible rather than analyzing the data as it is. In view of this fact, the threshold value for this quantitative variable was chosen 30 minutes which demarcates households into two groups - one group of households each is required more than 30 minutes to reach market center (disadvantaged group) and the other group of households each is required less than or equal to 30 minutes (advantaged group). The threshold value of 30 minutes is taken because it is a common in Nepal [7].

2.5 Test of association

The association of poverty status with each dichotomized variable is assessed using Chi-square test and effect size of each test is measured by Phi-coefficient whose values range from -1 to 1. Just like the correlation coefficient, a negative value of Phi-coefficient indicates that

when one variable increases, the other decreases and a positive value indicates that when one variable increases, so does the other.

2.6 The measures of poverty

In contemporary studies three measures of poverty are used. They were originated from a class of poverty measures P(α) introduced by - Foster-Greer-Thorbecke [8], and expressed as

$$\mathbf{P(\alpha) = \frac{1}{N} \sum_{i=1}^{q} \left(1 - \frac{y_i}{z}\right)^{\alpha}}$$

where α is index ≥ 0, q is the number of poor peoples, N is the total number of individuals, z is the poverty line, y_i is the per capita consumption expenditure and sum of the expression within parentheses is total poverty gap expressed as a proportion of the poverty line. In particular, P(0), P(1) and P(2) correspondingly yield the three measures of poverty – such as head count index, poverty gap index and squared poverty gap index. The measure P(0) is also known as head count ratio, incidence of poverty or poverty rate. It is a simple concept to understand and, therefore, widely used in political debate. However, it does not take into account of how poor the poor are, and this issue is addressed by the measure P(1) which in simple term measures how far away the poor peoples are from the poverty line, consequently P(1) satisfies the Monotonicity Axiom of Amartya Sen [9] and larger the value of P(1) larger the investment and effort would require to alleviate poverty. However, the measure P(1) does not take into account of the inequality in distribution of per capita expenditure among poor, and this issue takes into account by the measure P(2) and consequently P(2) satisfies the Transfer Axiom of Amartya Sen [9]. The two measures – P(1) and P(2) – are difficult concept to understand and, therefore, they are not widely used in political debate but they are useful for policy makers as well as for academicians.

Several developing countries, including Nepal, have been estimating and using the three measures of poverty for monitoring, evaluation and planning program of poverty reduction. Academicians are also using three measures in their academic work [10 - 21].

All the statistical analysis has been performed by using IBM SPSS version 20.

3. Results and Discussions

The results of this study are summarized in three tables where the first table displays the descriptive statistics of quantitative variables for the disadvantaged and advantaged group of households, the second table displays the association between the dichotomized variables with poverty status, and finally the third table provides weighted estimates of the three measures of poverty for the disadvantaged and advantaged group of population where weights are the population weights provided by CBS in the data file.

Table 1 shows that among the total 5,988 households, around 29% of have no land, 48% have at least three children, 26% have no literate persons of working age and 19% have poor access to market. The mean difference between advantaged and disadvantaged is highest in the variable ‘access to market’ and least is in the variable ‘area of land holding’. Skewness is positive in all variables.

Table 2 shows that the percentage of poor households is larger among the disadvantaged group of households than among the advantaged group. Within group difference in percentage of poor is highest in the variable ‘number of children under 15 and least in the variable ‘area of land holding’. All four dichotomized variables were found statistically significant with poverty status. The effect-size for each Chi-square test is positive and it is minimum for the test of association between the dichotomized variable of access to market center and poverty status, and it is maximum for the test of association between the dichotomized variable of the number of children under 15 and poverty status.

Table 1: Descriptive statistics of quantitative variables by group

	% of households	Mean	Std. Dev.	Skewness
Status of land holding:				
With land (0) - AG	71.2	0.36	0.55	5.00
Without land (1) – DG	28.8	0.00	0.000	NA
Number of children under 15:				
At most 2 (0) - AG	52.0	0.96	0.83	0.07
At least 3 (1) - DG	48.0	3.73	1.10	2.50
Number of literate working-age members:				
At least one (0) - AG	73.8	2.23	1.28	1.61
None (1) - DG	26.2	0.00	0.00	NA
Access to nearest market center:				
Having better access (0) - AG	80.7	13.91	11.42	0.12
Having poor access (1) - DG	19.3	152.96	148.86	2.77

Note: Figures within parentheses are coding scheme (0 for advantaged group and 1 for disadvantaged group). The dichotomous variable - poverty status – is coded as follows: 0 for non-poor and 1 for poor. AG = advantaged group and DG = Disadvantaged group NA: Not Applicable.

Computed from NLSS-III (2010/11)

Table 2: Association of dichotomized variables with poverty status

	% of poor households Within group	Chi-square & p-value	Phi-coefficient
Status of land holding:			
With land	15.0	114.9	0.14
Without land	27.0	(p<0.001)	
Number of children under 15:			
At most 2	11.0	653.0	0.33
At least 3	40.0	(p<0.001)	
Number of literate working-age members:			
At least one	16.0	142.0	0.15
None	31.0	(p<0.001)	
Access to nearest market center:			

Having better access Having poor access	11.0 26.0	206.7 (p<0.001)	0.08

Computed from NLSS-III (2010/11)

Table 3 shows that the scheme of dichotomization for each quantitative variable is able to differentiate poor peoples as 'more vulnerable' and 'less vulnerable' very distinctly according to each of the three measures of poverty. Comparison of the estimated three measures of poverty with the corresponding estimate of the national level, each estimate of each advantaged group of population is below the estimate of the national level, on the contrary each estimate of each disadvantaged group of population is above the estimate of the national level.

Table 3: Three measures of poverty for eight groups of household population

Variables	Head Count Index (P(0))×100	Poverty Gap Index P(1)×100	Square Poverty Gap Index P(2)×100
Status of land holding:			
With land (LV)	21.4	4.5	1.4
Without land (MV)	32.9	7.5	2.6
Number of children under 15:			
At most 2 (LV)	13.5	2.4	0.7
At least 3 (MV)	41.4	9.6	3.3
Number of literate persons of working age:			
At least one (LV)	21.5	4.3	1.4
None (MV)	41.7	10.4	3.9
Access to nearest market center:			
Having better access (LV)	16.3	3.3	0.9
Having poor access (MV)	32.1	7.1	2.5
National level of estimates	25.2	5.4	1.8

Note: LV = Less Vulnerable and MV = More Vulnerable.

Computed from NLSS-III (2010/11)

The extent of differentiations of poor people by the scheme of dichotomization varies across the dichotomized variables. For instance, the ratio of the head count index of the more vulnerable group to that of the less vulnerable group of population is highest for the dichotomized variable of 'the number of children under 15' and the ratio is 3.1 showing that the head count index of the more vulnerable group is 3.1 times higher than that of the less vulnerable group of population. Such ratios for the poverty gap index and the squared poverty gap index of the same dichotomized variable are correspondingly 4.0 and 4.7. Whereas the ratio of the head count index of the more vulnerable group to that of the less vulnerable group of

population is lowest for the dichotomized variable of 'the status of land holding' and the ratio is 1.5 and such ratios for the poverty gap index and the squared poverty gap index are correspondingly 1.7 and 1.9.

4. Conclusions

Dichotomization of a quantitative variable with the aid of a reasonable threshold value in poverty analysis is a useful strategy because this study to a large extent succeeded to show that such dichotomization scheme differentiates the poor people into two groups 'more vulnerable' and 'less vulnerable', so that the policy makers and development planners focus their poverty reduction program towards the more vulnerable people. Among the more vulnerable household populations of those households having three or more children under 15 is the most vulnerable. A baffling issue 'why the poor households tend to have large number children?' has to be resolved because for the reduction of the number of children from the most vulnerable group of households. The currently available data fail to resolve this issue. The impact of outmigration of literate and young population for employment as well as settlement in abroad was seen in the dichotomized variable of the quantitative variable 'the number of literate persons of working-age group'.

Conflict of Interest

Authors declare that there is not any conflict of interest.

Acknowledgments

Authors would like to thank Central Bureau of Statistics (CBS) for providing NLSS III 2010/11 data, and the department research committee of Central Department of Statistics, Tribhuvan University, Nepal for their critical comments and suggestions for this study. University Grant Commission (UGC) of Nepal is also acknowledged for providing Ph.D. fellowship as this is the part of Ph.D. research work.

References

[1]J. E. Hunter & F. L. Schmidt. Dichotomization of continuous variables: The implications for meta-analysis. J. Appl. Psychol. 75(3) (1990) 334–349.

[2]R. C. MacCallum, S. Zhang, K. J. Preache, & D. D. Rucker. On the practice of dichotomization of quantitative variables.Psychol. Methods **7**(1) (2002).

[3]S. P. Nelson, V. Ramakrishnan, P. J. Nietert, D. L. Kaman, P. S. Ramos, &, B. J. Wolf. An evaluation of common methods for dichotomization of continuous variables to discriminate disease status. Commun. Stat. Theory and Methods 46(21) (2017) 10823-10834.

[4]D. Chhetry. Understanding rural poverty in Nepal. In Asia and Pacific forum on poverty: Reforming policies and institutions for poverty reduction 1 (2001) 293-314.

[5]D. Chhetry. Comparative analysis of socio-demographic and economic characteristics of the poor and non-poor group of households. Nepal Population J. (2000) 101-112.

[6]Ministry of Health. Nepal Demographic Health Survey. Ministry of Health, New ERA Kathmandu, Nepal, the DHS Program, ICF, Rockville, Maryland,USA.(2011).

[7]Central Bureau of Statistics. Nepal Living Standard Survey (2010/11). Summary Report, Central Bureau of Statistics. National Planning Commission Secretariat, Government of Nepal (2011).

[8]J. Foster, J. Greer, & E.Thorebecke. A Class of decomposable measure of poverty, Econometrica 52(1984).

[9]A. Sen. Poverty: an ordinal approach to measurement. Econometrica 44 (2) (1976) 219 – 231.

[10] S. A. Yusuf, A. O.Adesanoye, and D. O.Awotide. Assessment of poverty among urban farmers in Ibadan Metropolis, Nigeria. J. Hum. Ecol. 24(3) (2008) 201-207.

[11]J. I. Onu, and Z. Abayomi. An analysis of poverty among households in Yola Metropolis

of Adamawa State, Nigeria. J. Soc. Sciences 20(1) (2009) 43-48.
[12]I. S. Chaudhry and S. Malik. The impact of socioeconomic and demographic variables on poverty: A village Study. Lahore J. Economics 14(1) (2009) 39-68.
[13]D. Akerele, and S. A. Adewuyi. Analysis of poverty profiles and socioeconomic determinants of welfare among urban households of Ekiti State, Nigeria. Curr. Res. J. Soc. Sci. 3(1) (2011) 1-7.
[14]R. A. Sanusi, T. S. Owagbemi, and M. Suleiman. Determinants of poverty among farm households in Ikorodu local government area of Lagos State, Nigeria. IJAFS 4 (2013) 538-552.
[15]S. O.Akinbode, A. O. Salami, and O. T. Ojo. Impact of micro finance on poverty status of small scale crop farming households in Southwest Nigeria. Am. J. Econ. 3(6) (2013) 322-329.
[16]L. A. Salami, and K. Atiman. An Analytical study of determinants of poverty level among households in Adamawa North District, Nigeria. Mediterranean J. Social Sci. 4(16) (2013) 73-73.
[17]A. Adetayo. Analysis of farm households poverty status in Ogun states, Nigeria. Asian Econo. Financ. Rev. 4(3) (2014) 325-340.
[18]R. S.Margwa, J. I. Onu, J. N. Jalo, and B. Dire. Analysis of poverty level among rural households in Mubi region of Adamawa State, Nigeria. J. Sci. Res. Stud. 2(1) (2015) 29-35.
[19]R.Sarania. Analysis of poverty profile and determinants of welfare among rural households: A Case study of Udalguri District, Assam. Int. J. Humanities & Social Studies (IJHSSS) 1(4) (2015) 138-144.
[20] K. P. Acharya. Analysis of poverty profile by type of house of households in Nepal. Management Dynamics 22(1) (2019) 7-10. DOI: https://doi.org/10.3126/md.v22i1.30231
[21]P.Uprety. Measures, distribution and decomposition of poverty: An empirical analysis in Nepal. Nepalese J. Stat. 4 (2020) 1-16.