

Poverty Groups Identification and Assessment of Poverty Alleviation Programs in Rural China

Zeng Zhen^{1 3}, Zhu Mengxian^{2 3*}

¹ *Guizhou University of Finance and Economics, Guiyang 550004, China*

² *Huazhong University of Science and Technology Library, Wuhan 430072, China*

³ *Big Data Institute, Wuhan University, Wuhan 430072, China*

*Corresponding author. Email: mxzhu@hust.edu.cn

ABSTRACT

This study analyzed the outcomes of 194 individuals living in rural poverty in Guizhou Province using data obtained from a poverty alleviation information system together with several machine-learning tools. First, four dimensions were abstracted from a multiple factor analysis: family background and work condition, self-development, regional context, and health status and labor capacity. Second, five heterogeneous groups were identified through k-medoids clustering based on the above dimensions. The results showed that the most significant differences among poverty groups were related to health status and household size. The effectiveness of poverty alleviation programs for the different poverty groups was then evaluated. By employing association rule mining, we showed that education assistance is effective in poverty groups in relatively ‘good’ condition but not in others, whereas the ‘Minimum Living Standard’ program failed in all groups, particularly when it was the sole program. Thus, our study indicated that additional targeted alleviation programs should be implemented to address the needs of different poverty groups, especially in terms of education assistance, medical care, and medical prevention programs.

Keywords: *rural China, rural poverty, poverty alleviation programs, association rule, k-medoids clustering, multiple factor analysis*

1. Introduction

In 2015, there were 55 million poor in China’s rural areas [1]. After the Chinese government adjusted the rural poverty standard in 2011, to a per capita net income of 2300 RMB (vs. 1067 RMB in 2007 and 1196 in 2009), the absolute poverty rate in rural areas rose sharply, to about 6% compared to 2010. This upward trend was maintained over the next 2 years. Consequently, poverty remains a major concern of the Chinese government, especially the gap between urban and rural populations and between the east coast vs. the western hinterlands, due to the potential for serious social and political unrest. In response, the current government, led by President Xi Jing Ping, set poverty reduction and rural development as key priorities (Government of China). With its targeted poverty alleviation policy of 2014, followed by a 2016 conference on central poverty alleviation, the government has officially set a goal of lifting China’s 70.17 million poor out of poverty by 2020. This number of people is equivalent to the populations of Canada and Poland combined[2]. While this is an ambition and arduous task, the recent policy differs from previous poverty-reducing policies because of the introduction of targeted poverty alleviation, which emphasizes ‘accurate identification, appropriate project

arrangement and accurate implementation and assessment’ to ensure that the alleviation policy will truly be implemented.

But how can rural poverty be identified and classified? How can suitable poverty alleviation measures be chosen and implemented? And can alleviation effects be objectively evaluated?[3]. Especially at the operational level, the process of poverty identification is mainly based on income, health status and education level, with other issues seldom considered by China’s local rural governments[4]. Hence, income alone cannot be used to classify poverty and to implement associated poverty alleviation measures, as it ignores the importance of multidimensional well-being. Moreover, not only does an income-based approach greatly deviate from the theory of poverty classification, it directly conflicts with the proposition of a targeted poverty alleviation policy[5]. In China and in other developing countries, the causes of rural poverty are heterogeneous, with differences in socio-economic characteristics, asset endowments, and the relationship between assets and income-generating activities[6]. Accordingly, precise, objective, and effective ways to identify the different forms of poverty and to implement suitable poverty alleviation programs are needed. Meanwhile, the Chinese government has invested enormous amounts of money and employed a large workforce to establish several poverty alleviation information management systems(PAMISs) and to upload information regarding rural

poverty[7]. The data in these PAMISs are a precious asset that can be utilized to build an analytical model aimed at identifying poverty-related patterns or revealing concealed problems.

The contribution of this paper lies in its identification of different poverty group using data from the PAMIS together with multiple factor analysis and k-medoids clustering. It takes into account the multiple dimensions of poverty, including health status, education level, living space, family background, work conditions, etc. These findings were further examined using association rule learning to explore the differences in the impact of poverty alleviation programs on the diverse poverty groups. A multidimensional approach to understanding the distinct characteristics of rural poverty groups and an analysis of the effectiveness of various alleviation measures in these groups is meaningful in assessments of current targeted poverty alleviation policies.

2. A MULTIDIMENSIONAL APPROACH TO THE STUDY OF POVERTY AND AN ASSESSMENT OF POVERTY ALLEVIATION PROGRAMS

Poverty, particularly rural poverty, is a complex problem that cannot simply be measured and characterized based on income or expenditures[8]. The case for paying greater attention to non-monetary indicators of the quality of life using multidimensional measures is supported by the inability of traditional approaches to thus far elucidate the true nature of poverty and social exclusion[9]. A better approach may be the multidimensional measurement and modeling of poverty to obtain a more precise and comprehensive view and thereby understand the interrelationships amongst poverty factors[10]. The framework for a multidimensional measurement and classification of poverty is rooted in the work of Sen (1976), who divided the poverty measurement problem into two distinct problems: (1) identifying the poor among the total population and (2) constructing a calculation of poverty using the available information on the poor, which relies on setting a poverty line. The latter considers poverty as being multidimensional rather than solely a function of income level and defines a poverty limit for each individual attribute[11]. Furthermore, the core feature of a multidimensional theory of poverty is understanding poverty as a lived experience and thus defined not just by a low income but also by ill health, insecurity, discomfort, isolation, and lack of agency [12].

In the past few decades, there has been considerable research into the multidimensional facets of poverty and deprivation. This has included adding new dimensions or categories to the evaluation method, with the aim of expanding the traditional concept of economic poverty and thereby obtain a more complete description of rural poverty. Among the multidimensional features of poverty that go beyond the

monetary aspect are:(1) personal reasons that restrict people from pursuing a better life, including age, education level, ill health, physical handicap, sex[8]; (2) social or family support, whether from a family member or a community worker [5],[10],[13],[14]; and (3) a regional or spatial context—for example, the distance of the impoverished community from public service facilities such as the hospital, school, main road—which could influence the ability of the rural poor to find work or obtain emergency medical care[5], [15], [16], [17]. In addition, western China is typically multi-ethnic, and ethnicity may influence the nature of rural poverty, resulting in significant variations in its multiple dimensions[18].

To understand the characteristics of multidimensional poverty, several studies have examined the use of statistical methods, including synthetic index approaches and data reduction techniques[19],[20]. The most well known and widely adopted of the former is the method of Alkire & Foster and the UNDP Human Development Index[8],[21]. In a synthetic index approach, a set of related dimensions, each represented by an array of indicators, is selected and then synthesized into a one-dimensional scale, for which a cut-off value is established. Values higher than the cutoff indicate poverty based on the selected multidimensional factors [22].

A synthetic index approach is an elegant way to construct multidimensional measures of deprivation and has the advantage of explicitly describing how to aggregate and weight the selected dimension. However, it has also been criticized for sacrificing multidimensionality such that the same score may blur different forms of deprivation[19]. Moreover, if the entropy function does not reflect the information contained in the database, a synthetic approach will be biased[20]. Data reduction, by contrast, is a non-parametric method in which all weights are endogenous to the database and determined by statistical rules [23]. This approach includes clustering and factor analysis techniques, which are better strategies than the synthetic index approach, as they not only assess a full set of items covering different aspect of life but also produce a higher-dimensional profile, especially when data from a database are used [19]. In addition, clustering may facilitate the creation of internally homogeneous but externally heterogeneous poverty groups, which is valuable for comparisons and for identifying both the factors that influence these groups and the appropriate alleviation program [24],[25].

Nevertheless, identifying and classifying poverty based on a multidimensional theory and advanced statistical tools is only the first step. The second and more important one is matching the poverty group with the appropriate alleviation program and then correctly assessing the outcome. Anti-poverty actions constitute a complex process that requires a series of precise policies guided and aided by the government or by non-governmental organizations[26]. A large body of research has emphasized targeted poverty alleviation[27] rather than the allocation of monetary support. In a targeted approach, whether by category (e.g., age, disability or

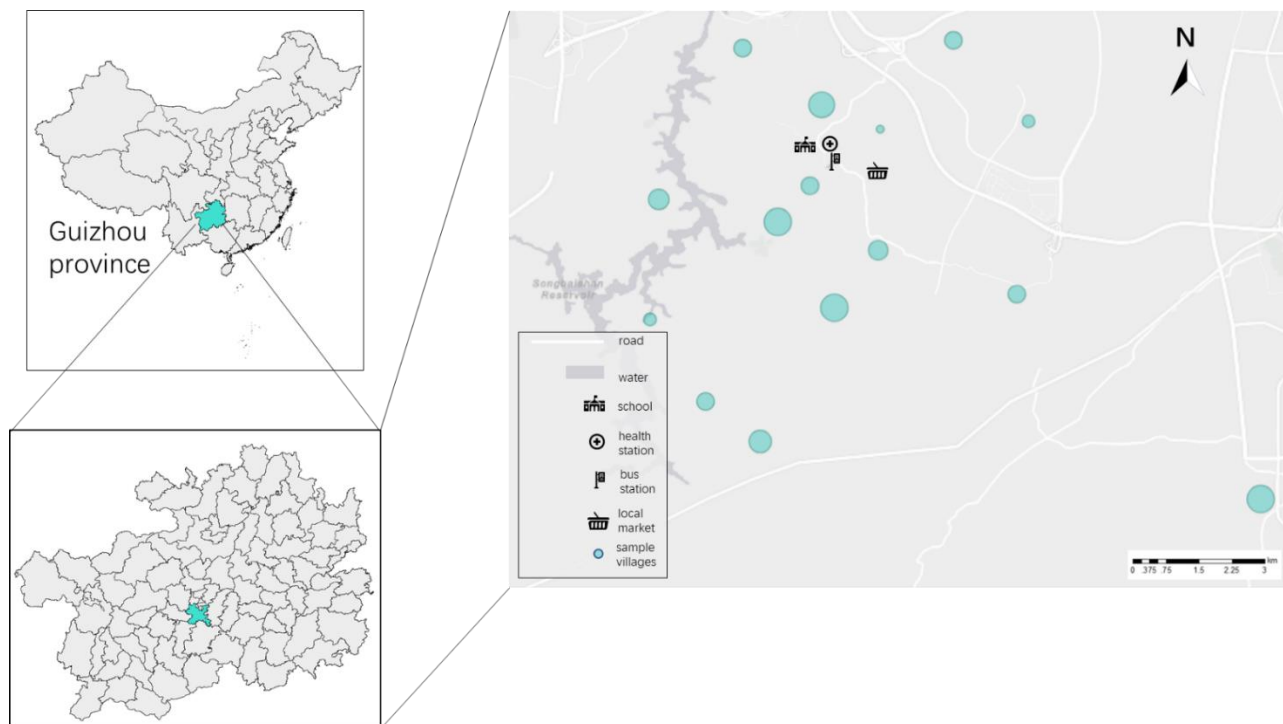


Fig. 1. Location of the sample villages. The size of the circle represents the number of rural poor. (Source: authors)

In Dang Wu county, one-third of the population belongs to ethnic minority. The primary source of local income is agriculture, followed by outside income and tourist income. The PAMIS database primarily stores information on the socio-economic background of the rural poor. The data are collected and uploaded by local workers in the villages. Demographic data include age, sex, income, education level, health status, employment status, ethnicity, family members living in the same household, and the number of family members working elsewhere. The PAMIS also stores address information. Then, with the cooperation of the local government (usually, the village committees), the different levels of rural poverty receive different alleviation measures, from dibao(direct monetary assistance) to skill training. This information is recorded and uploaded to the PAMIS to evaluate the results of the poverty alleviation measures, the PAMIS also records every specific poverty reduction measure for every person and the final poverty alleviation result. If the person's income has increased to above the poverty line by the end of year, the targeted poverty alleviation measure is considered successful.

The availability from the PAMIS of a wide range of non-monetary data covering different aspects of daily life facilitated the adoption of a truly multidimensional approach to our study of deprivation. Together with detailed records on poverty alleviation measures and the outcome of those efforts, the effectiveness of the different alleviation programs could be analyzed.

Several machine learning methods were used to analyze the data, including multiple factor analysis(MFA) and clustering, to establish the optimal heterogeneous subgroups. The relationship between the various poverty-reduction programs and the final effects in the different clusters was then examined using association rule.

4. IDENTIFICATION OF HETEROGENEOUS GROUPS AMONG THE RURAL POOR

4.1 Constructing poverty dimensions based on a multiple factor analysis

In practice, prior to their clustering, high-dimensional data are often transformed into a lower dimensional subspace via a principal component analysis(PCA), which allows coherent patterns to be detected more clearly. Cluster analysis is then applied to this lower dimensional subspace and is able to capture the characteristics of the cluster in a simplified manner [35]. This approach has largely solved the problem of strong correlations between the original variables [36].

However, the original PCA method can be applied only to continuous data whereas the rural poverty variables in our study are of the mixed type. As an alternative method we applied the improved R package of PCAmixdata, which is based on an MFA and is able to handle mixtures of

quantitative and qualitative variables. In the MFA, each group is normalized by dividing all the variables belonging to this group by the first eigenvalue derived from the PCA of this group. A traditional PCA is then applied to all the weighted variables combined. This method was initially developed for groups described only by quantitative variables, but it was later improved to deal simultaneously with groups described by qualitative and/or quantitative variables[37],[38] . To apply the MFA, we split the original dataset into a numerical data matrix and a categorical matrix and then used the PCAmix () function in the PCAmixdata package to obtain a linear combination of observed variables that accounted for the largest variance among the two matrices. An orthogonal VARIMAX rotation was then applied to this solution to simplify the interpretation of the component matrix.

According to the Kaiser-Harris criterion, we retained the first four components with eigenvalues > 1. The cumulative variance of this four-component solution accounted for 63.4%, which is acceptable for the social sciences 36. The eigenvalues for sex were < 1 and thus could not include any of the four components. To capture the main dimensions of each component, variables with a squared loading > 0.3 were assigned to the associated component. In the case of multiple components containing the same variables with simultaneous loading > 0.3, only the larger components were retained, to simplify the interpretation. The loadings of the four components for each observation variable are reported in Table1.

Table 1 Eigenvalues and percentage of the explained inertia of the MFA in an evaluation of rural poverty characteristics

Variable	Dimension1	Dimension2	Dimension3	Dimension4
Nearest city(kilometer)	0.0026	0.0006	0.8209	0.0106
Ethnicity	0.0059	0.8801	0.0013	0.0001
Nearest station(kilometer)	0.0115	0.0023	0.8422	0.0202
Age	0.0147	0.7242	0.0037	0.0058
Public service proximity(kilometer)	0.0198	0.0005	0.7002	0.0063
Sex	0.0335	0.0060	0.0358	0.0046
Education level	0.0683	0.3258	0.0158	0.1102
Outside work (months)	0.4016	0.2263	0.0038	0.0257
Outside work (range)	0.4169	0.2964	0.0277	0.0349
Health status	0.4664	0.0053	0.0318	0.7908
labor capacity	0.5030	0.1347	0.0194	0.7595
Number of family members	0.6456	0.0139	0.0039	0.0046
Number of outside workers in the family	0.7323	0.0057	0.0008	0.0000
Number of working family members	0.7993	0.0016	0.0029	0.0001

The first component consisted of ‘family background and work condition’ which included the highly positive family-related variables of ‘number of family members’ ‘number of family members in the workforce’ and ‘number of outside workers in the family’ . A correlation between ‘health status’ ‘outside work (months)’ and ‘outside work (range)’ was determined (positive loading) and indicated the influence of the family situation on the level of outside work among the rural poor and therefore potentially the family health status. The variables strongly influencing the fourth component, ‘education level’ ‘ethnicity (race)’ and ‘age’ can be regarded as ‘self-development’ component. The third component was clearly influenced by the variables associated with a regional context, with the high squared loading of ‘public service

proximity’ ‘nearest bus station’ and ‘nearest city’ For the final component, ‘health status and labor capacity’ there was a significant squared loading for both ‘health status’ and ‘labor capacity’.

Based on the results of the MFA, the component scores accounting for the various asset dimensions were determined for entry into the cluster analysis. This approach solves the problem of multi-collinearity and thus makes the cluster more effective and precise. Moreover, many cluster solutions are sensitive to mixed-type variables, after conversion of the MFA component into continuous variables, the selection of the cluster solution could be expanded and the different rural poverty groups, characterized by maximum intragroup homogeneity or heterogeneity, identified.

4.2 Identifying poverty clusters based on multidimensions

Cluster analysis allows identifying different poverty groups, characterized by maximum intra-group homogeneity and inter-group heterogeneity. First, the number of clusters be determined based on information from the Scree plot. In moving from four to five clusters, there is a distinct drop in the within-groups sum of squares. After five clusters, this decrease drops off, suggesting that the solution of five clusters is a good fit to the dataset.

Second, the data were clustered through a k-medoids analysis. Many studies have adopted k-mean clustering to classify different poverty group along various dimensions [25],[39]. However, the k-mean is sensitive to outliers and in some cases does not accurately represent the correct center. By contrast, our study used the k medoids algorithm, which is based on the true rather than the mean value of the data points. Thus, the sum of pairwise dissimilarities rather than the sum of the squared Euclidean distances is minimized.

Last, the silhouette value was used to evaluate the clustering quality, as it illustrates the consistency within the data clusters.

Each horizontal line corresponds to an element and the length of the line to the silhouette value, which ranges from -1 to 1 . A highly positive silhouette value indicates that the element is well matched to its own cluster and poorly matched to the neighboring cluster. A negative value indicates that the observation has been placed in the wrong cluster and the cluster solution may have the incorrect number of clusters [40]. In our study, the average silhouette value of five clusters was > 0 , and the silhouette values of only five elements were negative, corresponding to 2.6% of all observations. Thus, the cluster solution was considered acceptable.

Once the cluster result was generated, statistically significant differences in the cluster attributes were determined by analyses of variance (ANOVAs) followed by Games-Howell test post hoc comparisons, which is suitable for pair-wise multiple comparisons between groups in case of unequal or unknown variances between groups and unequal group sizes. The frequencies attribute was analyzed by a chi-square test followed by Bonferroni-adjusted z-tests. The results indicated a significant difference in the attributes among the different clusters. A description of the dimensions of the five clusters is provided in Table 2.

Table 2 Rural poverty: cluster comparisons (Source: authors).

Dimension	Cluster	c1	c2	c3	c4	c5	Total
	Number of cases	49	56	57	8	24	194
		25.26%	28.87%	29.38%	4.12%	12.37%	
Self-development	<i>age (F=53.67, p=0.000***)</i>						
	median	47	68	47	55.5	50.5	52
	mean	42.29 ^{c2,c4}	67.84^{all}	46.91 ^{c2}	53.13 ^{c2}	47.58 ^{c2}	52.12
	SD	13.9	4.09	8.02	8.17	13.03	14.18
	<i>sex (Pearson chi-square=9.77, p=0.054)</i>						
	male	67.30%	71.40%	89.50%	87.50%	83.30%	77.80%
	female	32.70%	28.60%	10.50%	12.50%	16.70%	22.20%
	<i>education level (Pearson chi-square=56.62, p=0.000***)</i>						
	illiteracy	14.3% ^{c2}	57.10%^{all}	7% ^{c2}	0% ^{c2}	16.7% ^{c2}	24.20%
	elementary school	55.10%	28.6 ^{c3,c4,c5}	59.6 ^{c2}	87.5^{c2}	62.6% ^{c2}	51.00%
	middle school	30.6%	12.5%	33.3%	12.5%	16.7%	23.70%

	high school	0.00%	2.00%	0.00%	0.00%	4.00%	1.00%
	<i>Ethnicity (Pearson chi-square=35.74, p=0.000^{***})</i>						
	Han	100% ^{c2,c3,c5}	83.9% ^{c1,c3}	54.4% ^{c1,c2}	87.50%	83.3% ^{c1}	79.40%
	minority	0.00%	16.10%	45.60%	12.50%	16.70% ^{c1}	20.60%
	<i>pension insecure condition (Pearson chi-square=162.6, p=0.000=0.000^{***})</i>						
	yes	0.00% ^{c2,c4,c5}	100.00% ^{all}	1,8% ^{c2,c4,c5}	25.00% ^{c1,c2,c3}	25.00% ^{c1,c2,c3}	33.50%
	no	100.00%	0.00%	98.20%	75.00%	75.00%	66.50%
	<i>number of family members (F=11.66, p=0.000^{***})</i>						
	Median	3	3	3	3.5	1	3
	mean	2.65 ^{c3,c5}	2.88 ^{c5}	3.46 ^{c1,c5}	3.13	1.33 ^{c1,c2,c3}	2.81
	SD	1.42	1.38	1.24	1.64	0.7	1.43
	<i>number of working family members (F=18.25, p=0.000^{***})</i>						
	Median	2	2	2	1.5	0	2
	mean	1.63 ^{c3,c5}	1.59 ^{c3,c5}	2.54 ^{c1,c2,c5}	1.5	0.17 ^{c1,c2,c3}	1.7
	SD	1.3	1.22	1.15	1.31	0.48	1.35
	<i>number of outside workers in the family (F=14.23, p=0.000^{***})</i>						
	Median	1	1	1	1	0	1
	mean	1.16 ^{c5}	0.84 ^{c3,c5}	1.47 ^{c2,c5}	1	0.08 ^{c1,c2,c3}	1.02
	SD	0.94	0.8	0.78	0.76	0.28	0.89
	<i>health status (Pearson chi-square=258.24, p=0.000^{***})</i>						
	heath	79.60% ^{c3,c4,c5}	85.70% ^{c3,c4,c5}	100.00% ^{all}	0.00% ^{c1,c2,c3}	16.70% ^{c1,c2,c3}	76.30%
	disability	18.40% ^{c3}	10.70% ^{c5}	0.00% ^{c1,c5}	0.00% ^{c5}	83.30% ^{c1,c2,c3,c4}	18.00%
	cancer	0.00%	0.02%	0.00%	0.00%	0.00%	0.50%
	tuberculosis	0.00%	0.00%	0.00%	12.50%	0.00%	0.50%
	cerebral infarction	0.02%	0.00%	0.00%	12.50%	0.00%	0.50%
	hyperthyroidism	0.00%	0.00%	0.00%	12.50%	0.00%	0.50%
Family background and work situation							
Health status and labor capacity							

	other chronic disease	1.98% ^{c4}	3.58% ^{c4}	0.00% ^{c4}	62.50% ^{all}	0.00% ^{c4}	3.70%
<i>labor capacity (Pearson chi-square=217.30, p=0.000^{***})</i>							
	skilled labor	0.00%	0.00%	1.80%	0.00%	0.00%	0.50%
	common labor	67.30% ^{all}	30.40% ^{c1,c3}	98.20% ^{all}	0.00% ^{c1,c3}	8.30% ^{c1,c3}	55.70%
	no labor	32.70% ^{all}	69.60% ^{c1,c3}	0.00% ^{all}	100.00% ^{c1,c3}	91.70% ^{c1,c3}	43.80%
<i>public service distance, kilometer (F=51.44, p=0.000^{***})</i>							
	median	5.6	7.8	15.8	11.8	19.1	15.8
	mean	6.48 ^{c2,c3,c5}	10.85 ^{c1,c3,c5}	15.62 ^{c1,c2,c5}	14.08	20.52 ^{c1,c2,c3}	12.48
	SD	2.23	5.96	1.35	9.43	5.91	6.34
<i>city distance, kilometer (F=39.02, p=0.000^{***})</i>							
	median	11.5	11.5	12.4	10.8	15.05	12.4
	mean	10.46 ^{c2,c3,c5}	11.48 ^{c1,c3,c5}	12.38 ^{c1,c2,c5}	11.04	14.25 ^{c1,c2,c3}	11.81
	SD	1.58	1.39	0.12	2	1.64	1.73
<i>market distance, kilometer (F=50.49, p=0.000^{***})</i>							
	median	1.6	2.6	5.1	3.85	6.5	5.1
	mean	2.02 ^{c2,c3,c5}	3.5 ^{c1,c3,c5}	5.04 ^{c1,c2,c5}	4.61	6.82 ^{c1,c2,c3}	4.03
	SD	0.82	2.01	0.46	3.21	2.01	2.15
<i>Station Distance, kilometer (F=62.30, p=0.000^{***})</i>							
	Median	12.6	12.6	15.8	12.3	16.1	15.8
	mean	12.02 ^{c2,c3,c5}	13.71 ^{c1,c3,c5}	15.74 ^{c1,c2,c4}	12.39 ^{c3}	16.21 ^{c1,c2,c4}	14.14
	SD	1.27	1.96	0.42	2.64	1.22	2.14

Note: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. (1) Statistically-significant differences at $p < 0.001$ between the means or the frequency attribute of the clusters, (2) The superscripts letters(c1,c2,c3,c4,c5) indicate statistically-significant differences between clusters.

Rural poverty in clusters 1 and 3 referred to a relatively 'good' condition, as the average age in these two clusters was younger than that in any other cluster (42.29 years in cluster 1 and 46.91 in cluster 3 vs. an average 52.12 in all clusters). The health status of these two poverty clusters was also relatively good, especially with no disability or disease identified within cluster 3.

Furthermore, in terms of self-development ability, the education level in both clusters is was higher than in the other poverty clusters. Given the comparatively good education background and good health condition in these two clusters, the work situation was relatively positive. For example, outside work experience occurred only in people in these two clusters. Nevertheless, cluster 3 and cluster 1 differed significantly in terms of the regional context and the ethnicity

ratio. Compared to cluster 3, the distance from a registered road, public service (school, hospital, market), and the city was closer in cluster 1, which implies more local job opportunities and thus a more stable life. Thus, while the poor in cluster 3 were in better health and had more family members as well as more working household members, the number of family members working outside was higher in cluster 1 than in cluster 3. Among the surveyed poor themselves, 37.5% of the rural poor belonging to cluster 1 worked outside the local village, compared to 19.5% in cluster 3. Further, the average outside work period in cluster 1 was five times longer than in cluster 3. In terms of ethnicity, cluster 1 consisted entirely of Han whereas almost half of the population in cluster 3 belonged to the Miao people. This difference may explain the larger household size in cluster 3, since minorities in China are not subject to the one-child policy. Thus, overall, cluster 3 could be considered as a ‘healthy and large family’ cluster and cluster 1 as a ‘young and conveniently located’ cluster.

Cluster 4 largely comprised rural poor who were ‘chronic diseases, no labor.’ These individuals were in very poor health and all of them had at least one disease, particularly the chronic diseases, ranging from rheumatoid arthritis to cerebral infarction. Accordingly, people in this cluster were unable to work or to leave the village to seek a better life.

The poverty condition in cluster 5 was relatively close to that of cluster 4, as both had health issues. Thus, these two clusters can be both considered as ‘poor condition’. Only 16.67% of the poor in cluster 5 could be considered as healthy. However, while the health status of cluster 5 was better than that of cluster 4, 8.33% of the people in cluster 5 were able to work compared to 0% in cluster 4. This does not necessarily imply that life was easier for people in cluster 5 than in cluster 4, for the following reasons. Firstly, family human resources were higher in cluster 4, with an average number of 3.13 people per family vs. only 1.33 in cluster 5, which was the lowest among

all clusters. Secondly, household labor, which is essential for supporting unhealthy family members, was just 0.17 per family in cluster 5, which was also the lowest in all clusters. Further, the regional context was more ‘poverty-friendly’ in cluster 4 than in cluster 5, as public services, including a hospital, were nearer. Access to public services is particularly important for the elderly who are living in poverty and who are in poor health. Thus, cluster 5 comprised the ‘isolated and sick’ poverty group.

Cluster 2 comprised the ‘elderly, with little education’ group and the average age of cluster 2 was significantly older ($M=67.14$) than that of other clusters. However, age was not the sole problem in cluster 2, as the illiteracy rate in this cluster was significantly higher than in other clusters. The combination of old age and little education was an important constraint preventing people in clusters 2 from finding work and earning a livelihood. Nevertheless, 85.71% of the poor in this cluster were considered as healthy, and their health status was better than that of the poor in cluster 4 and, especially, in cluster 5. An evaluation of the results of the self-assessment of the reason for poverty, over half of the poor in cluster 2 reported a lack of skills as the cause of their poverty, which indicated that these individuals were still interested in learning new skills in order to find work. Thus, it may be that in rural China, older people without serious disease are still interested in working in order to earn a living.

Finally, to provide an overall illustration of the different poverty groups, we normalized the sum of the component scores of five groups for four dimensions and thereby established a 5-point scale (the regional dimension score was calculated as 5 minus the original number of points). We then compared the four poverty dimensions and the overall condition of the different poverty groups (Fig. 2). Note that the overall condition of cluster 3 and cluster 1 was much better than that of cluster 5 and cluster 4, while cluster 2 had an intermediate condition.

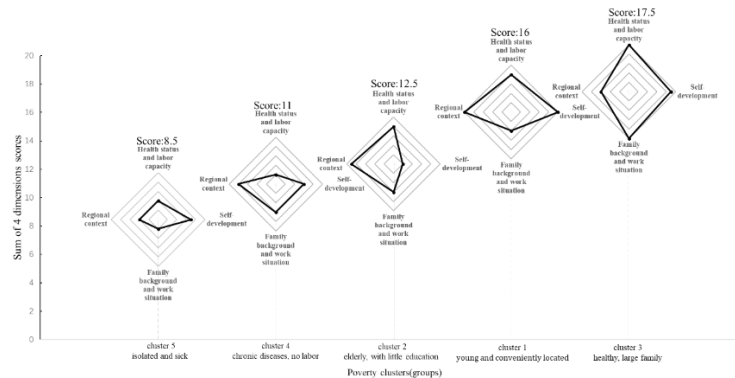


Fig. 2. The -5-point score based on health status and labor capacity, regional context, family background and work situation, as well as self-development dimensions. (Source: authors)

4.3 Comparison of the results of poverty reduction among the different groups

The final goal of targeted poverty alleviation is the elimination of poverty but determining whether this goal has been met is complicated. Currently, the main computing method to evaluate poverty reduction results is poverty reducing rate (PRR), which based on the annual income of the poor after targeted poverty alleviation divided by 6000 yuan. A result > 100%, indicates that poverty was successfully reduced.

To achieve its poverty reduction goal, the local government can employ various poverty-reduction methods based on the

Table 3 Comparison of the poverty reduction rate in the different poverty groups

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Mean	1.29 ^{c3, c5}	1.14 ^{c5}	1.67 ^{c2, c4, c5}	0.93 ^{c3}	0.84 ^{c1, c2, c3}
SD	0.72	0.43	0.86	0.33	0.29

Note: The superscripts letters (c1, c2,c3,c4,c5) indicate statistically-significant differences between clusters.

As expected, cluster 3(healthy, large family) had the highest poverty reduction rate (M=1.6698, SD=0.11). The difference was significant and the rate higher than that of every other cluster except cluster 1, comprising the young and conveniently located (M=1.2873, SD=0.72). Thus, clusters 1 and 3 were the groups in which poverty reduction was the most successful.

Poverty reduction result in cluster 2, comprising the ‘old, with a low level of education,’ could also be considered as successful based on the government standard (M=1.13, SD=0.43), given that the average rate was > 100%. The rate was still much lower than that of cluster 3 but much higher than that of cluster 5, which had the lowest reduction rate.

5. ASSESEMENT OF POVERTY ALLEVIATION PROGRAMS BASED ON ASSOCIATION RULE MINING

The clustering and ANOVA results revealed differences between the poverty groups and the influence of the final poverty reduction outcome. However, an evaluation of which poverty alleviation program is truly effective for the different cluster is more important and more meaningful, as the results may help to identify the best policy for the particular poverty group and thereby conserve limited resources.

As one of several broadly used machine learning technologies, association rule mining(ARM) is well suited to the evaluation of poverty alleviation, as it may identify an association between the different Poverty reduction program(PRP) and policy results. A rule generated by association learning has the form LHS (left hand side) => RHS (right hand side). In our study, the LHS represented the combination of the different clusters and different PRPs, and the RHS the different poverty

poverty background. These methods can be classified in five main categories: education assistance, medical insurance, farm livestock breeding projects, minimum living standard, and skill training. Each category can be further divided into subcategories according to the exact situation. Local governments tend to use more than one method in evaluating poverty alleviation and commonly include the family condition.

To determine the relationship between the different cluster characteristics and the results of poverty reduction, a one-way ANOVA was conducted. The results revealed statistically significant differences in the mean poverty reduction rate between clusters (Table 3).

Lastly, in cluster 4 (‘chronic diseases, no labor’) and cluster 5 (‘isolated and sick’), poverty reduction was unsuccessful, as the mean reduction rate in both was < 100%. However, the post-hoc tests indicated that cluster 4 did not significantly differ from any of the other clusters except cluster 3, whereas cluster 5 was significant different and the poverty reduction rate lower than in other clusters except cluster 4.

Thus, based on a comprehensive analysis and a description of the different clusters, our results suggest that the reduction rate was related to the characteristics of the cluster, which further supported our poverty clustering results.

reduction rates. According to the rule, the RHS is likely to occur whenever the LHS occurs. In the process of the association rule, support, confidence, and lift are the three best-known measures to select rules of interest from the set of all possible rules. Support is an indication of how frequently the item set appears in the dataset, and confidence is an indication of how often the rule is found to be true. In its basic structure, every association rule meeting the minimum support and confidence level is extracted (Ihibuchi et al., 2007). In our study, the minimum support and confidence were 0.5% and 50%, respectively. Lift represents the importance of a rule, the value of an association rule lift is the ratio of the actual confidence of the rule vs. the expected confidence of the rule. A lift value > 1 indicates that the LHS and RHS coincide more often than expected, in which case the occurrence of the rule body has a positive effect on the occurrence of the rule head. A lift value near 1 indicates that the occurrence of the rule body has almost no effect on the occurrence of the rule head.

For a rule, $LHS \rightarrow RHS$

$$Support(LHS \rightarrow RHS) = P(LHS, RHS)$$

Because the association rule works on factor type data, in our study, we first divided the RHS data (poverty reduction) into successful (> 100%) and unsuccessful (< 100%) datasets, each of which was further split into two equal-size sub-datasets: very poor (58.3–69.7%), poor (69.7–98.1%), good (100–143%), and very good (143–339%). The data of the 194 poor individuals in the five clusters were included in the PRPs. Comparisons of the observed association rule between the different clusters with both the PRPs and the corresponding poverty reduction rate allowed us to draw the follow conclusions.

Within the ‘very good’ level (Table 4), several methods could be considered as effective in cluster 3, based on high support, confidence, and lift levels. These methods were: combination of ‘high school or university assistance’ and ‘collective breeding by cooperative and take annual dividend’, combination of ‘basic education assistance’ and ‘breeding and sell by poverty independently’, and ‘collective breeding by cooperative and take annual dividend’, ‘breeding and sell by poverty independently’. However, for the other clusters, the number of effective programs was limited, and only ‘breeding and sell by poverty independently’, or a ‘collective breeding by cooperative and take annual dividend’ was related to a high rate of poverty reduction. Specifically, given the many similarities between clusters 1 and 3, it was of interest to analyze the reason(s) for the differences in the effective PRPs. While this may have been due to the limited amount of data, it was more likely linked to the different conditions represented by the two clusters. The poor in cluster 3 were in good health and had the potential of extensive family support (given the significantly higher family number in cluster 3 than in cluster1), but a large family may also be a double-edged

$$Confidence(LHS \rightarrow RHS) = P(RHS | LHS) = P(LHS, RHS) / P(LHS)$$

$$Lift(LHS \rightarrow RHS) = P(RHS | LHS) / P(RHS) = Confidence(LHS \rightarrow RHS) / P(RHS)$$

sword. On the one hand, it implies more family support and a larger work force, which is important for agricultural work and enables more opportunities to earn non-farm income. Nevertheless, a larger family number means a larger education burden, particularly since the poor in cluster 3 were middle aged and had the highest proportion of ethnic minorities (who are not subject to the one-child policy). Thus, the poor in cluster 3 tended to have more school age children. This may explain why education assistance combined with other alleviation programs was successful in cluster 3. By contrast, since the health status of the poor was not as good in cluster 1 as in cluster 3 and given the higher disability rate in cluster 1, people in this cluster tended to stay on their farms and breed livestock on their own, because it allowed them more freedom and the avoidance of corruption which may cause by local government workers.

Although both cluster 4 and cluster 5 were characterized by sickness and an inability to work away from the village, the most effective PRPs differed. For cluster 4, ‘breeding and sell by poverty independently’ was more likely to increase income, whereas in cluster 5 the ‘collective breeding by cooperative and take annual dividend’ was the better solution, as evidenced by comparatively high confidence and lift levels (confidence=100%, lift=3). A larger family may play an essential role in this case, as farm work is labor-intensive and the poor in clusters 4 and 5 were unlikely to manage on their own but to rely on the assistance of family members. Thus, farm work, including livestock breeding, was more difficult for cluster 5, with an average of only 1.32 family members, and these individuals would therefore benefit from a dividend. By contrast, the poor in cluster 4 could rely on help from family members in performing farm activities.

Table 4 ARM for ‘very good’ poverty reduction results (143–339%)

Left-hand side	Right-hand side	Support	Confidence	Lift
Poverty reduction program=high school or university education assistance + collective breeding by cooperative and take annual dividend Cluster=3	poverty reduction rate = (1.43,3.39]	1.56%	100.00%	3.00
Poverty reduction program= collective breeding by cooperative and take annual dividend Cluster=5	poverty reduction rate = (1.43,3.39]	1.52%	100.00%	3.00
Poverty reduction program= collective breeding by cooperative and take annual dividend	poverty reduction rate = (1.43,3.39]	8.85%	77.27%	2.30

Cluster=3				
Poverty reduction program=breeding and sell by poverty independently Cluster=3	poverty reduction rate = (1.43,3.39]	6.25%	66.67%	2.00
Poverty reduction program=collective Breeding, annual dividend Cluster=2	{poverty reduction rate = (1.43,3.39]}	3.13%	60.00%	1.80
Poverty reduction program=basic education assistance + breeding and sell by poverty independently Cluster=3	{poverty reduction rate = (1.43,3.39]}	1.56%	60.00%	1.80
Poverty reduction program=breeding and sell by poverty independently Cluster=1	poverty reduction rate = (1.43,3.39]	6.77%	58.15%	1.44
Poverty reduction program=breeding and sell by poverty independently Cluster=4	poverty reduction rate = (1.43,3.39]	1.52%	50.00%	1.50

The association rule results for a ‘good’ level (Table 5) indicated that there were no remarkable differences between the ‘very good’ level for cluster 1 and cluster 4 and therefore that ‘breeding and sell by poverty independently’ may prompt a final reduction in the poverty rate. However, in cluster 2, skill training and medical assistance were more likely to eradicate poverty, as, in contrast to the other clusters, the confidence level was very high (confidence=100%, lift =3). The poor in cluster 2 were elderly but still had ambitions to work and were in relatively good health (especially compared to the poor in clusters 4 and 5). In Dang Wu county, skill training is mainly focused on agricultural skills. This is appropriate because rural China has a large number of elderly and, as the young leave, the older generation has become an

important part of the rural labor force. Furthermore, medical assistance also had a positive effect on poverty reduction in cluster 2, which was the oldest cluster. While the health problems in clusters 4 and 5 were much more serious, an association rule emerged between medical assistance and a ‘very good’ or ‘good’ reduction of the poverty level. This may have been due to the insufficient provision of medical assistance for people in these clusters by local government workers. As the assessment of medical needs and the implementation of medical assistance are complicated, labor-intensive, and time-consuming, among the various poverty alleviation measures local government workers may opt for shortcuts, such as direct assistance in the form of money or livestock.

Table 5 ARM for ‘good’ poverty reduction results (100–143%)

Left-hand side	Right-hand side	Support	Confidence	Lift
Poverty reduction program=medical assistance Cluster=2	poverty reduction rate = (1,1.43]	0.52%	100.00%	3.00
Poverty reduction program=high school or university education assistance + breeding and sell by poverty independently Cluster=3	poverty reduction rate = (1,1.44]	0.52%	100.00%	3.00
Poverty reduction program=agriculture skill training Cluster=2	poverty reduction rate = (1,1.45]	1.56%	100.00%	3.00

Poverty reduction program= collective breeding by cooperative and take annual dividend Cluster=4	poverty reduction rate = (1,1.46]	0.52%	100.00%	3.00
Poverty reduction program=basic education assistance + collective breeding by cooperative and take annual dividend Cluster=3	poverty reduction rate = (1,1.47]	1.04%	66.67%	2.00
Poverty reduction program=breeding and sell by poverty independently Cluster=2	poverty reduction rate = (1,1.48]	6.77%	61.90%	1.86
Poverty reduction program=breeding and sell by poverty independently Cluster=4	poverty reduction rate = (1,1.49]	0.52%	58.10%	1.50

Two important results were obtained in the analyses of poverty alleviation at the ‘very poor’ (Table 6) and ‘poor’(Table 7) levels. Firstly, regardless of the cluster, the invalid solutions for individually applied programs indicated that a combined poverty reduction approach was more appropriate for most people living in poverty. Secondly, a minimum living standard (MLS) and education assistance were strongly associated with ‘poor’ and ‘very poor’ levels of poverty reduction, respectively.

In terms of the MLS, it may be that people who receive an MLS are among the most helpless from the view of local government, which lacks a suitable solution other than giving money, in this case a worse start results in a bad end. This was true of all clusters, except clusters 1 and 3, whose self-development-was much better. Therefore, giving money to the poor is far from sufficient and solutions beyond the existing ones are needed, especially in the long term.

For the education assistance program, the ARM results for ‘very good’ and ‘good’ poverty reduction showed that the combination of education assistance and other poverty alleviation methods was effective, but only in cluster 3, the

cluster with the best starting conditions. Thus, education assistance is not ineffective, but it is far from sufficient and it had a notable impact only on cluster 3, but even in this cluster it had to be complemented by other poverty alleviation methods.

Basic governmental education assistance for primary or middle school is 300 yuan (48 dollars) per month, for high school 2660 yuan (422 dollars) per year, and for university 4830 yuan (767 dollars) per year. Nevertheless, this is not enough. Although primary education is free, in practice various fees are levied: for uniforms, for lunch. And, for parents who work in the city they must pay ‘temporary schooling fees’, or they may choose to have their children live with grandparents and enroll in a local school (where the education is likely to be of lower quality). For a poor family, the cost for non-compulsory education (high school or university) may be prohibitive, as tuition is not free, and children must leave home to study. By not working near home or finding outside work, these children will need to have their families pay their living expenses. Moreover, these families lose an important income source.

Table 6 ARM for the ‘poor’ poverty reduction results (69.7–98.1%)

Left-hand side	Right-hand side	Support	Confidence	Lift
Poverty reduction program =high school or university education assistance Cluster=2	poverty reduction rate = (0.697,0.981]	1.52%	100.00%	6.19
Poverty reduction program=basic education assistance Cluster=5	poverty reduction rate = (0.697,0.982]	1.52%	100.00%	6.19
Poverty reduction program=basic education assistance Cluster=3	poverty reduction rate = (0.697,0.983]	1.04%	100.00%	6.19

Poverty reduction program=minimum living standard Cluster=3	poverty reduction rate = (0.697,0.984]	1.52%	50.00%	3.10
Poverty reduction program=minimum living standard Cluster=1	poverty reduction rate = (0.697,0.984]	4.69%	52.86%	2.65
Poverty reduction program=minimum living standard Cluster=4	poverty reduction rate = (0.697,0.984]	1.04%	50.00%	2.48

Table 7 ARM for ‘very poor’ poverty reduction results (58.3–69.7%)

Left-hand side	Right-hand side	Support	Confidence	Lift
Poverty reduction program=high school or university education assistance+ Minimum Living Standard Cluster=2	poverty reduction rate = (0.583,0.697]	1.52%	100.00%	6.19
Poverty reduction program=Minimum Living Standard Cluster=4	poverty reduction rate = (0.583,0.698]	1.56%	60.00%	3.72
Poverty reduction program=Minimum Living Standard Cluster=5	poverty reduction rate = (0.583,0.699]	5.73%	50.00%	3.10
Poverty reduction program=Minimum Living Standard Cluster=3	poverty reduction rate =(0.583,0.700]}}	5.73%	50.00%	3.10

6. CONCLUSIONS

Our analysis thus supports the implementation of a targeted alleviation policy and different pathways to reach different poverty groups.

(1) It is important that the government single out ‘poor condition’ type poverty (cluster 4 and cluster 5). This is especially the case for cluster 5, in which family support is lacking, health problems are serious, and public service is far away and thus difficult to access. The government should devote considerable attention to this most vulnerable poverty group. For these individuals, medical support is crucial and should include regular medical visits or social support from NGOs.

(2) Special educational assistance targeting those not included in ‘good condition poverty’ groups, i.e., cluster 2, cluster 4, and cluster 5, is needed, such as increased monetary support

from the government to compensate families who will lose a potential working member of the family. Moreover, based on the results of the ARM for these ‘poor condition’ clusters, education assistance should be provided along with other forms of alleviation, rather than as a single solution.

(3) Except for the two most vulnerable clusters (4 and 5), the poor in clusters 1, 2, and 3 are likely to benefit from targeted skill training programs. For cluster 2, based on the ARM result, current skill training (agriculture training) is effective and should be continued. In cluster 3, comprising healthy individuals with a large family as well as the best educational background, good health, and ample family help, advanced skill training, not limited to agriculture, such as training in domestic service, driving, or construction, may be worthwhile, as it would allow these individuals to engage in off-farm skilled employment, either in the city or in the east coast provinces. For cluster 1, the situation is more complex. While the overall condition of people in cluster 1 is good, the disability rate is high, which distinguishes this cluster from

cluster 3. Work outside the village to work as heavy physical labor the is therefore much more difficult. However, given the convenient regional context of cluster 1, with the poor living especially near the market, training programs could focus on assistance in business training or e-business training to encourage startups of local agribusiness or other small business.

(4) Current alleviation programs aimed at ‘good condition’ poverty (clusters 3 and 1) are comparatively effective, but they profit from the good health status and younger age of the poor in these groups (42.29 years old in cluster 1 ,46.91 in cluster 3). As the poor in both clusters age and thus become vulnerable to health problems, poverty could weaken the respective individuals. Particularly for cluster 1, the only significant difference between it and cluster 4 was health status, chronic disease potential, and age. Accordingly, government-funded medical prevention programs will be advantageous not only now but also in the future, as they will help to preserve the comparatively good health of people in cluster 3 and cluster 1 and thereby avoid their later suffering from chronic disease.

Appendix A. Supplementary data

Supplementary data related to this article can be found at: <http://dx.doi.org/10.17632/45xz623d85.1#file-9e6938c0-404e-4b4e-a6cf-64dd30879f02>

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (71964007).

REFERENCES

- [1]. The World Bank In China., 2017.
- [2]. Li Y, Su B. Realizing targeted poverty alleviation in China: People’s voices, implementation challenges and policy implications. *CHINA AGR ECON REV.* 2016;8(3):443-54.
- [3]. Wang G, Wei X. Analysis on the mechanism of targeted poverty alleviation in China. *Rural Economy.* 2015 2015-09-25(09):46-50.
- [4]. Ting Z, Yang Y, Ling Z. targeted poverty alleviation: technology, theory and change. *Guizhou Social Sciences.* 2015 2015-08-01(08):156-62.
- [5]. Hui C, Quan-hong Z. Study on the Accurate Identification of Poverty and Accurate Poverty Alleviation Based on Multidimensional Poverty Measurement with the Mountain Area of Northern Guangdong as an Example. *Journal of Guangdong University of Finance & Economics.* 2016 2016-06-08(03):64-71.
- [6]. Walelign SZ, Pouliot M, Larsen HO, Smithhall C. Combining Household Income and Asset Data to Identify Livelihood Strategies and Their Dynamics. *J DEV STUD.* 2017(53):1-19.
- [7]. Xie Z. The problem of implement of big data application in the targeted elevation. *Journal of Fujian Party School.* 2017(8):64-71.
- [8]. Alkire S, Foster J. Counting and multidimensional poverty measurement. *J PUBLIC ECON.* 2011 2011-01-01;95(7):476-87.
- [9] Nolan B, Whelan CT. Resources, Deprivation, and Poverty. *Oup Catalogue.* 1996.
- [10] Ravallion M. Growth, inequality and poverty: looking beyond averages. *WORLD DEV.* 2001;29(11):1803-15.
- [11] .Bourguignon F, Fields G. Discontinuous losses from poverty, generalized P α measures, and optimal transfers to the poor. *J PUBLIC ECON.* 1997;63(2):155-75.
- [12]. Alkire S, Foster J. Counting and multidimensional poverty measurement. *J PUBLIC ECON.* 2012;95(7):476-87.
- [13]. Acevedo-Garcia D, McArdle N, Hardy EF, et al. The Child Opportunity Index: Improving Collaboration Between Community Development And Public Health. *HEALTH AFFAIR.* 2014 2014-11-01;33(11):1948-57.
- [14]. Böhnke P. Are the poor socially integrated? The link between poverty and social support in different welfare regimes. *J EUR SOC POLICY.* 2008;18(2):133-50.
- [15]. Amarasinghe U, Samad M, Anputhas M. Spatial clustering of rural poverty and food insecurity in Sri Lanka. *FOOD POLICY.* 2005 2005-01-01;30(5):493-509.
- [16]. Hentschel J, Lanjouw JO, Lanjouw P, Poggi J. Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador. *The World Bank Economic Review.* 2000 2000-01-01;14(1):147-65.

- [17]. Xu D, Zhang J, Rasul G, et al. Household Livelihood Strategies and Dependence on Agriculture in the Mountainous Settlements in the Three Gorges Reservoir Area, China. *Sustainability*. 2015;7(5):4850.
- [18]. Gradín C. Rural Poverty and Ethnicity in China. *Research on Economic Inequality*. 2015; 23:221-47.
- [19]. Lucchini M, Butti C, Della Bella S, Lisi A. The application of a topological clustering technique to capture forms and dynamics of deprivation in contemporary Switzerland. *QUAL QUANT*. 2017:1-22.
- [20]. Asselin L. Multidimensional poverty theory. MIMAP Training Session on Multidimensional Poverty. Québec, Canada, 2002.
- [21]. McGillivray M, White H. Measuring development? The UNDP's human development index. *J INT DEV*. 2010;5(2):183-92.
- [22]. Alkire S, Foster J. Understandings and misunderstandings of multidimensional poverty measurement. *J ECON INEQUAL*. 2011;9(2):289-314.
- [23]. Nwachukwu TE, Odigie P. Multidimensional Poverty in Cameroon: Determinants and Spatial Distribution. *International Journal of Social Science Studies*. 2011;Vol. 3,(No. 1).
- [24]. Caruso G, Sosa-Escudero W, Svarc M. Deprivation and the Dimensionality of Welfare: A Variable-Selection Cluster-Analysis Approach. *Review of Income & Wealth*. 2014;61(4):702-22.
- [25]. Ansoms A, McKay A. A quantitative analysis of poverty and livelihood profiles: The case of rural Rwanda. *FOOD POLICY*. 2010;35(6):584-98.
- [26]. Hulme D, Shepherd A. Conceptualizing Chronic Poverty. *WORLD DEV*. 2003;31(3):403-23.
- [27]. Byerlee D. Targeting poverty alleviation in priority setting for agricultural research. *FOOD POLICY*. 2000;25(4):429-45.
- [28]. Dutrey AP. Successful Targeting? Reporting Efficiency and Costs in Targeted Poverty Alleviation Programmes: United Nations Research Institute for Social Development; 2007.
- [29]. Lavallée E, Olivier A, Pasquier-Doumer L, Robilliard AS. Poverty alleviation policy targeting: a review of experiences in developing countries. *Economics Papers from University Paris Dauphine*. 2010.
- [30]. Alderman H. Do local officials know something we don't? Decentralization of targeted transfers in Albania. *J PUBLIC ECON*. 2004;83(3):375-404.
- [31]. Meiyan, Wang. Emerging Urban Poverty and Effects of the Dibao Program on Alleviating Poverty in China. *CHINA WORLD ECON*. 2007;15(2):74-88.
- [32]. Zhi-hong Z, Fu-sheng Z. Evaluation of the rural poverty fund's efficiency of the state-supported poverty-stricken. *Journal of Hunan Agricultural University (Social Sciences)*. 2013 2013-10-28(05):1-6.
- [33]. Park A, Wang S. Community-based development and poverty alleviation: An evaluation of China's poor village investment program. *J PUBLIC ECON*. 2010;94(9-10):790-9.
- [34]. Alkire S, Seth S. Multidimensional Poverty Reduction in India between 1999 and 2006: Where and How? *WORLD DEV*. 2015;72:93-108.
- [35]. Zha H, He X, Ding C, Simon H, Gu M. Spectral relaxation for K-means clustering. *International Conference on Neural Information Processing Systems: Natural and Synthetic*, 2001:1057-64.
- [36]. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate data analysis (4th ed.): with readings.*, 1995.
- [37]. Abdi H, Williams LJ, Valentin D. Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews Computational Statistics*. 2013;5(2):149-79.
- [38]. Escofier B, Pages J. Multiple factor analysis (AFMULT package) : Multiway data analysis: software and applications. *COMPUT STAT DATA AN*. 1994.
- [39]. Luzzi GF, Flückiger Y, Weber S. A Cluster Analysis of Multidimensional Poverty in Switzerland. *Ssrn Electronic Journal*. 2006;261(4):63-79.
- [40]. Lletí R, Ortiz MC, Sarabia LA, Sánchez MS. Selecting variables for k -means cluster analysis by using a genetic algorithm that optimises the silhouettes. *ANAL CHIM ACTA*. 2004;515(1):87-100.