

# Clustering the Economic Status via Partitioning around Medoid and Its Association with Common Non-communicable Diseases

Elaheh Sanjari<sup>1</sup>, MSc; Ali Ahmadi<sup>2</sup>, PhD; Hadi Raeisi Shahraki<sup>2</sup>, PhD

<sup>1</sup>Student Research Committee, Shahrekord University of Medical Sciences, Shahrekord, Iran;

<sup>2</sup>Modeling in Health Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran

## Correspondence:

Hadi Raeisi Shahraki, PhD; Rahmatieh Educational Complex, Shahrekord University of Medical Sciences, Postal code: 88139-44349, Shahrekord, Iran

Tel: +98 38 33320379

Email: raeisi.shahraki\_hadi@yahoo.com

Received: 15 October 2023

Revised: 07 November 2023

Accepted: 01 December 2023

## Abstract

**Background:** During the last decades, the role of economic status and wealth-related variables in relation to the mortality and incidence of a wide range of diseases have received increased attention. This study focused on clustering the economic status of a population-based study using partitioning around the medoid (PAM) and then investigating the association between the obtained economic clusters and the incidence of non-communicable diseases (NCDs).

**Methods:** The present study was based on data from Shahrekord Cohort Study (SCS). This study considered nine NCDs, including cardiac disease, myocardial infarction, diabetes, hypertension, stroke, all types of malignancies, chronic lung disease, depression, and obesity, among 7034 participants aged 35 and 70 from the urban population of Sharekord (IRAN) in 2022. Four quantitative and four qualitative variables were used to cluster the economic status. The *NbClust* package was used to determine the optimal number of clusters, and the K-med package in R software (version 4.2.1) was used for PAM clustering. Descriptive statistics were reported as frequency (%) or median (IQR), and statistical analysis was performed using the Chi square test and Mann-Whitney test in SPSS software (version 19.0).  $P < 0.05$  was considered statistically significant.

**Results:** The estimated optimal number of clusters was two. The first cluster contained individuals with good economic status, while the second cluster contained those with a moderate economic status. The findings indicated that individuals with a good economic status had significantly higher rates of cardiac disease (7.2% versus 5.3%,  $P < 0.001$ ), stroke (1.3% versus 0.6%,  $P < 0.001$ ), diabetes (12.8% versus 9.1%,  $P < 0.001$ ), hypertension (21.6% versus 15.6%,  $P < 0.001$ ), depression ( $P < 0.001$ ), and obesity ( $P = 0.03$ ).

**Conclusion:** The findings of the present study showed that economic status was significantly associated with the majority of NCDs.

Please cite this article as: Sanjari E, Ahmadi A, Raeisi Shahraki H. Clustering the Economic Status via Partitioning around Medoid and Its Association with Common Non-communicable Diseases. Iran J Med Sci. doi: 10.30476/ijms.2023.100513.3275.

**Keywords** • Cluster analysis • Economic status • Non-communicable diseases

## What's Known

- The economic status and wealth-related variables are associated with the occurrence of a wide range of diseases.
- Measuring economic status has always presented critical challenges, and various statistical methods have been used to combine the information on wealth-related variables.

## What's New

- Significant associations were found between the economic status and the majority of non-communicable diseases.
- Participants with good economic status had significantly higher rates of cardiovascular diseases, diabetes, hypertension, and stroke.

## Introduction

During the last decades, the role of economic status and wealth-related variables have received increased attention in relation

to the mortality and occurrence of a wide range of diseases.<sup>1, 2</sup> Nowadays, most health studies consider economic status a potential confounding variable since it might influence numerous aspects of health through a wide range of mechanisms.<sup>3, 4</sup> Nevertheless, evaluating economic status has always presented critical challenges, and various statistical methods have been used to combine the information on wealth-related variables.

The use of principal component analysis as one of the most prevalent methods of dimensional reduction leads to the extraction of several continuous quantitative components called "principal". Nonetheless, this numerical variable lacks a defined unit. It is not interpretable, and cannot incorporate qualitative variables.<sup>5, 6</sup> To address the limitation of qualitative variables incorporation, the multiple correspondence analysis method was proposed to combine the information of variables related to the economy in several components.<sup>7, 8</sup> However, the obtained coordinates are still not interpretable, and also its division into quartiles does not seem rational.

Although dividing observations into homogeneous groups using different clustering methods has a long history, the majority of prevalent clustering methods are distance-based and do not support qualitative variables. In the method of partitioning around the medoid (PAM), the basis of clustering is the similarity of the samples to each other. Therefore, it is possible to use all types of variables in the clustering process.<sup>9, 10</sup> In the PAM analysis, the number of optimal clusters is calculated using appropriate statistical indicators, indicating that there is no bias in estimating the number of optimal clusters.<sup>11</sup>

PAM clustering was already being applied in a variety of medical studies. Recently, it was used to cluster patients with knee movement disorders,<sup>12, 13</sup> assigned patients with back pain to the treatment-based homogenous subgroups,<sup>14</sup> clustered COVID-19 medications,<sup>15</sup> and determined the socio-economic status of mothers to assess its impact on children's growth at 60 months.<sup>16</sup>

Considering the importance of the economic condition in health studies and the need to consider its function, the correct estimation of this variable using the most up-to-date statistical methods can be introduced as the first step in determining its role. To the best of our knowledge, only a few studies considered the use of advanced clustering methods to cluster people based on their economic status. In this way, the present study aimed to cluster the economic status of a population-based study via PAM clustering and then investigate the association between the

obtained economic clusters and the occurrence of non-communicable diseases (NCDs).

## Participants and Methods

This study was based on data from Shahrekord Cohort Study (SCS), which was conducted as one of the prospective epidemiological studies in the southwest of Iran with a sample size of 10075 participants in 2022. According to census data, 7034 participants from the urban population, aged 35 and 70, were included in this study.<sup>17</sup> The only exclusion criterion was unwillingness to participate.

The purpose of the research was explained to all participants before the sampling procedure, and written informed consent was obtained from all the participants. This research was approved by the Ethics Committee of Shahrekord University of Medical Sciences (code: IR.SKUMS.REC.1401.191).

This study considered nine NCDs, including cardiac disease, myocardial infarction (MI), diabetes, hypertension, stroke, all types of malignancies, chronic lung disease, depression, and obesity. Diabetes was defined as having FBS levels equal to or higher than 6.99 mmol/L (126 mg/dL) or taking blood glucose-lowering medication. Individuals undergoing treatment for physician-diagnosed diabetes were also considered diabetic. Hypertension was defined as a systolic blood pressure of  $\geq 140$  mmHg or a diastolic blood pressure of  $\geq 90$  mmHg, accompanied by a previous diagnosis of hypertension or being on antihypertensive medication. Individuals with a body mass index (BMI)  $\geq 30$  were classified as obese. Moreover, they had an approved history of disease, based on medical evidence, and the consumption of related drugs regularly based on specialist recommendations for the other diseases.<sup>18, 19</sup> To cluster the economic status, eight variables were utilized, four quantitative and four qualitative. The size of the home refers to the entire surface area, excluding areas that are measured in square meters, including the porch, garden, yard, and even parking or garage. The second quantitative variable was the number of bedrooms in the participant's present residence. The third variable in this study was the total number of travels made by the participants around the world, excluding foreign pilgrimages. The next quantitative variable was the number of participant trips within Iran over the last ten years. This variable comprised pilgrimage or non-pilgrimage travels that were more than 100 Km from the participant's residence. Access to the four household items, namely a computer or laptop, an automobile, a vacuum cleaner,

and a freezer, were considered the qualitative variables.

The dissimilarity of quantitative variables was calculated using Euclidean distance. This index was generated using equation 1 for two observations, x and y.<sup>20</sup>

$$(x,y) = \sqrt{\sum_{i=1}^p (y_i - x_i)^2} \quad (1)$$

The dissimilarity for nominal qualitative variables could be expressed in the following equation:

$$d(i,j) = \frac{p - u}{p} \quad (2)$$

Where p is the total number of variables, and u is the total number of variables that belong to the same category in two observations.<sup>11</sup>

One of the most important steps in the clustering process was to determine the number of clusters, which should be based on the nature of the data and without bias. The NbClust package was used to determine the optimal number of clusters. This package employed the "majority rule" to calculate more than 20 indices to determine the optimal number of clusters.<sup>21</sup>

The partitioning around medoid was developed by Kaufman and Rousseeuw.<sup>22</sup> The term medoid refers to an observation within a cluster that has the lowest dissimilarity to other members of that cluster.<sup>11, 20</sup> To execute PAM clustering, the following steps were implemented:

- 1- The number of clusters (k) was determined.
- 2- k observations were randomly selected as the medoid of the clusters.
- 3- The dissimilarity matrix was calculated for all two by two observations.
- 4- Each observation was assigned to the cluster that had the least dissimilarity to its medoid.
- 5- The medoid for each cluster was updated.
- 6- The dissimilarity of each observation to the medoid of all existing clusters was recalculated. If the dissimilarity of each observation to the medoid of its cluster was less than the dissimilarity of that observation to the medoid

of other clusters, the algorithm terminated; otherwise, we would return to the fourth step.<sup>11</sup>

Descriptive statistics were presented as frequencies and percentages or median (IQR). The statistical analysis was performed using the Chi square test and Mann-Whitney test in the SPSS software, version 19.0 (IBM Corp., Armonk, NY: IBM Corp.). For the clustering aim, NbClust, K-med, and cluster packages in R software, version 4.2.1 (R Core Team, Vienna, Austria) were used. P<0.05 was considered statistically significant.

## Results

This study included 7034 participants, with 3515 (50%) men and 3519 (50%) women. The mean age of the participants was 49.36±9.27, ranging from 35 to 70. About 1476 (21%) of the participants were illiterate, 1806 (25.7%) had less than a diploma, 2086 (29.7%) had diplomas, and 1666 (23.7%) had a Bachelor's degree or above. In this study, the median and interquartile range of the participants' house size and number of rooms were 120 (100-150) and 2 (2-3), respectively. In addition, the median and interquartile range of the participants' international travels during their lifetime was 0 (0-1), and the median and interquartile range of the participants' national travels in the last 10 years was 27 (9-61). 5204 (74%) of the participants in this research had access to a computer, and 5978 (85%) owned an automobile. The majority of the participants, 6953 (98.8%) owned a vacuum cleaner, and 6939 (98.6%) had a freezer.

The majority rule in the NbClust package predicted that the optimal number of clusters was two. The silhouette index for the first cluster was 0.26, for the second cluster was 0.17, and the suggested model had an average silhouette of 0.212.

The majority in the second cluster was men 2015 (52.1%), while the majority in the first cluster was women 1581 (52.9%), which was statistically significant (P<0.001). As shown in table 1, the second cluster had more illiterates than the first cluster (21.8% versus 19.9%).

**Table 1:** Comparison of demographic variables in two clusters

Variables	Subgroups (N)	Cluster 1 n (%)	Cluster 2 n (%)	P value*
Sex	Male (3515)	1410 (47.1)	2105 (52.1)	<0.001
	Female (3519)	1581 (52.9)	1938 (47.9)	
Education	Illiterate (1476)	596 (19.9)	880 (21.8)	0.09
	High school (1806)	772 (25.8)	1034 (25.6)	
	Diploma (2086)	877 (29.3)	1209 (29.9)	
	Bachelor and above (1666)	746 (24.9)	920 (22.8)	

\*Chi square test, P<0.05 was considered statistically significant.

However, there was no statistically significant association ( $P=0.09$ ). Besides, the median was significantly higher in the first cluster than in the second one (51.23 versus 47.98,  $P<0.001$ ).

Figure 1 compares the four quantitative variables between the two groups. The first cluster had a significantly larger median housing size (160) than the second cluster (100,  $P<0.001$ ). The median number of bedrooms in the first cluster was three, while it was two in the second cluster ( $P<0.001$ ). As indicated in figure 2, the first cluster had significantly greater proportions of assets, including a computer (78.4% versus 70.7%), an automobile (88.6% versus 82.3%), and a freezer (99.0% versus 98.4%,  $P<0.001$ ). As a result, those in the first cluster had good economic status, while those in the second had a moderate one.

The findings showed that the proportion of

cardiac disease, stroke, diabetes, hypertension, depression, and obesity was significantly higher among the participants with good economic status. Instead, the proportion of MI, cancers, and chronic lung disease had no significant difference between individuals with good and moderate economic status (table 2).

### Discussion

The present study found that those with a higher socioeconomic class were more likely to suffer from NCDs than others, which was consistent with previous research. Mtintsilana and colleagues found that in South Africa and Kenya, good socio-economic status was associated with an increased risk of NCDs, and that even after controlling for smoking and alcohol consumption, socio-economic status was found

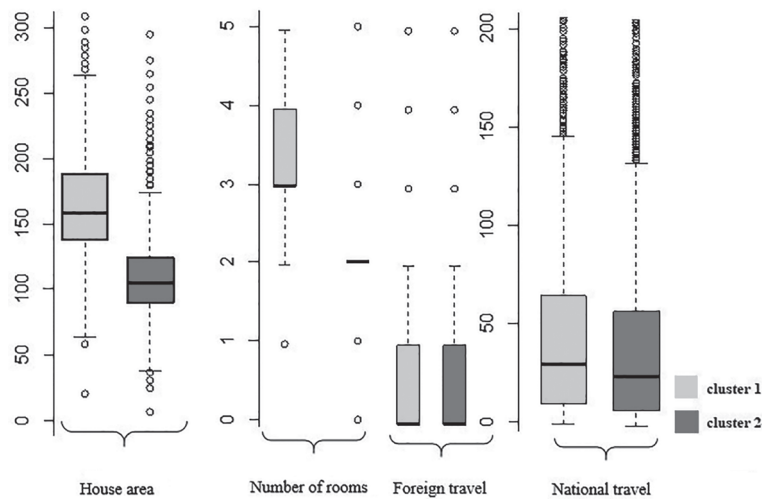


Figure 1: The box plots show the spread of quantitative economic variables in each cluster.

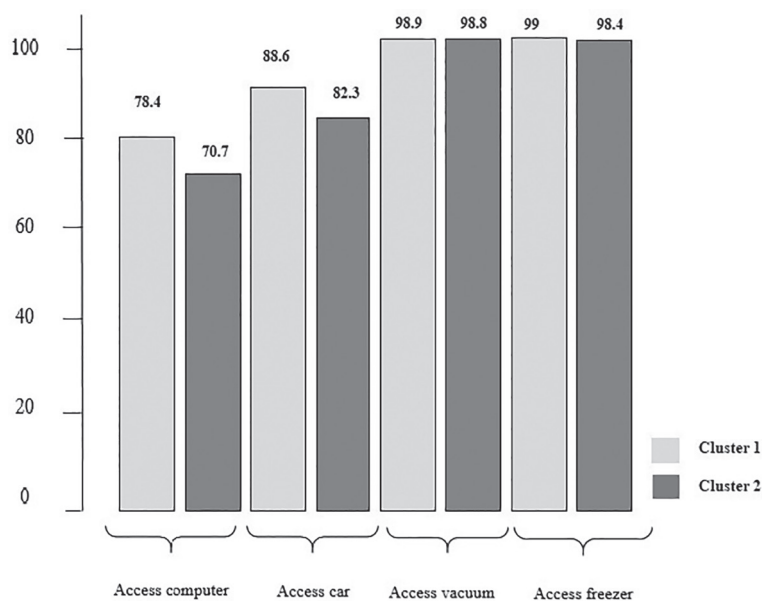


Figure 2: The bar charts show the relative frequency of qualitative economic variables in each cluster.

**Table 2:** Comparison of economic status and the incidence of common non-communicable diseases

Variables	Subgroups	Cluster 1 (good economic status) n (%)	Cluster 2 (moderate economic status) n (%)	P value
Has cardiac disease	Yes (432)	216 (7.2)	216 (5.3)	<0.001 <sup>†</sup>
	No (6602)	2775 (92.8)	3827 (94.7)	
Has myocardial infarction	Yes (111)	52 (1.7)	59 (1.5)	0.38 <sup>#</sup>
	No (6923)	2939 (98.3)	3984 (98.5)	
Has stroke	Yes (65)	39 (1.3)	26 (0.6)	<0.001 <sup>#</sup>
	No (6969)	2952 (98.7)	4017 (99.4)	
Has diabetes	Yes (751)	382 (12.8)	369 (9.1)	<0.001 <sup>†</sup>
	No (6283)	2609 (87.2)	3674 (90.9)	
Has hypertension	Yes (1275)	645 (21.6)	630 (15.6)	<0.001 <sup>†</sup>
	No (5759)	2346 (78.4)	3413 (84.4)	
Has cancer	Yes (62)	28 (0.9)	34 (0.8)	0.70 <sup>#</sup>
	No (6972)	2963 (99.1)	4009 (99.2)	
Has chronic lung disease	Yes (343)	162 (5.4)	181 (4.5)	0.07 <sup>#</sup>
	No (6691)	2829 (94.6)	3862 (95.5)	
Has depression	Yes (1272)	601 (20.1)	671 (16.6)	<0.001 <sup>†</sup>
	No (5762)	2390 (79.9)	3372 (83.4)	
Obesity	Yes (1960)	878 (29.7)	1082 (27.4)	0.03 <sup>†</sup>
	No (4955)	2082 (70.3)	2873 (72.6)	

<sup>†</sup>Chi square test, <sup>#</sup>Fisher's Exact Test; P<0.05 was considered statistically significant.

to be a significant risk factor for the development of NCDs. There was a positive association between socioeconomic position and the prevalence of NCDs, which could be attributed to fast urbanization, epidemiological transmission, as well as their influence on lifestyle factors. In addition, as in the present study, NCDs such as obesity, diabetes, and hypertension were more prevalent in rich groups.<sup>23</sup>

The findings of the present study revealed that those with higher levels of education were more likely to develop NCDs than other groups. Reddy and colleagues found that those with a university education were more likely to suffer from NCDs. One reason for this was that this group of people was more concerned with their studies and might not have been screened for NCDs.<sup>24</sup> Furthermore, according to the findings of the present study, individuals with a higher economic status might have greater access to facilities and health care, increasing the likelihood of identifying their diseases.

The current study also showed that obesity was more prevalent among those with a high economic status. According to Marthias and colleagues, obesity was associated with several NCDs, including cardiovascular disease, hypertension, stroke, arthritis, and high blood cholesterol. In support of the present study, it can be stated that groups of people with higher income, economic status, and education had better health literacy and access to healthcare staff. Thus, the likelihood of contracting a non-communicable disease was higher than that of groups with lower social status.<sup>25</sup>

Lotfi and colleagues found that cardiovascular diseases were less likely in those with a high socioeconomic status, which contradicted our findings. This disparity could be explained by case-control research with a different sample size than ours.<sup>26</sup>

Kundu conducted a study on socio-economic inequalities in the burden of communicable and NCDs among older persons in India. They found that the prevalence of NCDs differed from the anthropometric status defined by BMI.<sup>27</sup> Obesity was also a major factor in explaining the disparities in the prevalence of NCDs. Obesity was associated with an elevated risk of several serious NCDs, including type 2 diabetes, cardiovascular diseases, stroke, asthma, and various malignancies.<sup>28</sup>

Although this study intended to employ clustering based on a wide range of economic characteristics, only eight variables were evaluated due to the convergence issues, and lack of access to data on individual's income and expenses. Despite these limitations, the present study utilized a large sample of participants from a population-based study, as well as an advanced statistical method that incorporated mixed-type data sets.

## Conclusion

A better understanding of the association between economic status and NCDs at the individual level could assist health policymakers in developing preventive measures and health promotion programs. The findings of the



present study showed that economic status was significantly associated with the majority of NCDs, which might be due to the effect of lifestyle.

### Acknowledgment

The present study was extracted from the MSc thesis of Elaheh Sanjari, a master's student in biostatistics. This article was supported by Shahrekord University of Medical Sciences (code: 6654). We appreciate the Department of Biostatistics and Epidemiology and the Health Modeling Research Center for their invaluable cooperation and assistance.

### Authors' Contribution

E.S: Study design and drafting; A.A: Study design, data acquisition, and reviewing the manuscript; H.RSh: Study concept, study design, and drafting; All authors read and approved the final manuscript and provided an agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### Conflict of Interest

Hadi Raeisi Shahraki, as the Editorial Board Member, was not involved in any stage of handling this manuscript. A team of independent experts was formed by the Editorial Board to review the article without his knowledge.

### References

- 1 Bodenmann P, Favrat B, Wolff H, Guessous I, Panese F, Herzig L, et al. Screening primary-care patients forgoing health care for economic reasons. *PLoS One*. 2014;9:e94006. doi: 10.1371/journal.pone.0094006. PubMed PMID: 24699726; PubMed Central PMCID: PMC3974836.
- 2 Mode NA, Evans MK, Zonderman AB. Race, Neighborhood Economic Status, Income Inequality and Mortality. *PLoS One*. 2016;11:e0154535. doi: 10.1371/journal.pone.0154535. PubMed PMID: 27171406; PubMed Central PMCID: PMC4865101.
- 3 He P, Luo Y, Hu X, Gong R, Wen X, Zheng X. Association of socioeconomic status with hearing loss in Chinese working-aged adults: A population-based study. *PLoS One*. 2018;13:e0195227. doi: 10.1371/journal.pone.0195227. PubMed PMID: 29596478; PubMed Central PMCID: PMC5875885.
- 4 Kish JK, Yu M, Percy-Laurry A, Altekruse SF. Racial and ethnic disparities in cancer survival by neighborhood socioeconomic status in Surveillance, Epidemiology, and End Results (SEER) Registries. *J Natl Cancer Inst Monogr*. 2014;2014:236-43. doi: 10.1093/jncimonographs/lgu020. PubMed PMID: 25417237; PubMed Central PMCID: PMC4841168.
- 5 Agyekum AK, Adde KS, Aboagye RG, Salihi T, Seidu AA, Ahinkorah BO. Unmet need for contraception and its associated factors among women in Papua New Guinea: analysis from the demographic and health survey. *Reprod Health*. 2022;19:113. doi: 10.1186/s12978-022-01417-7. PubMed PMID: 35527266; PubMed Central PMCID: PMC9080214.
- 6 Najafi F, Soltani S, Karami Matin B, Kazemi Karyani A, Rezaei S, Soofi M, et al. Socioeconomic - related inequalities in overweight and obesity: findings from the PERSIAN cohort study. *BMC Public Health*. 2020;20:214. doi: 10.1186/s12889-020-8322-8. PubMed PMID: 32046684; PubMed Central PMCID: PMC7014739.
- 7 Jooste S, Mabaso M, Taylor M, North A, Shean Y, Simbayi LC. Socio-economic differences in the uptake of HIV testing and associated factors in South Africa. *BMC Public Health*. 2021;21:1591. doi: 10.1186/s12889-021-11583-1. PubMed PMID: 34445996; PubMed Central PMCID: PMC8390264.
- 8 Omondi I, Odiere MR, Rawago F, Mwinzi PN, Campbell C, Musuva R. Socioeconomic determinants of *Schistosoma mansoni* infection using multiple correspondence analysis among rural western Kenyan communities: Evidence from a household-based study. *PLoS One*. 2021;16:e0253041. doi: 10.1371/journal.pone.0253041. PubMed PMID: 34161354; PubMed Central PMCID: PMC8221481.
- 9 Eyler L, Hubbard A, Juillard C. Assessment of economic status in trauma registries: A new algorithm for generating population-specific clustering-based models of economic status for time-constrained low-resource settings. *Int J Med Inform*. 2016;94:49-58. doi: 10.1016/j.ijmedinf.2016.05.004. PubMed PMID: 27573311.
- 10 Webster TJ. Malaysian economic development, leading industries and industrial clusters. *The Singapore Economic Review*. 2014;59:1450044. doi: 10.1142/S0217590814500441.
- 11 Kassambara A. Practical guide to cluster analysis in R: Unsupervised machine

- learning. Sthda; 2017
- 12 Farzadaghi M, Razeghi M, Sobhani S, Raeisi-Shahraki H, Alipour Haghighi M, Farzadaghi M, et al. Knee impairments: Comparison between new clinical classification by cluster analysis and movement system impairment model. *J Bodyw Mov Ther.* 2022;30:210-20. doi: 10.1016/j.jbmt.2022.02.003. PubMed PMID: 35500973.
  - 13 Reza Farzadaghi M, Razeghi M, Sobhani S, Raeisi Shahraki H, Motealleh A. A New Clustering Method for Knee Movement Impairments using Partitioning Around Medoids Model. *Iran J Med Sci.* 2020;45:451-62. doi: 10.30476/ijms.2019.82033. PubMed PMID: 33281262; PubMed Central PMCID: PMCPMC7707633.
  - 14 Shokri E, Razeghi M, Raeisi Shahraki H, Jalli R, Motealleh A. The Use of Cluster Analysis by Partitioning around Medoids (PAM) to Examine the Heterogeneity of Patients with Low Back Pain within Subgroups of the Treatment Based Classification System. *J Biomed Phys Eng.* 2023;13:89-98. doi: 10.31661/jbpe.v0i0.2001-1047. PubMed PMID: 36818010; PubMed Central PMCID: PMCPMC9923237.
  - 15 Alqurneh A, Mustapha A, Sharef NM. A partitioning-based approach for clustering COVID-19 drugs and co-medication for safe use. *International Journal of Integrated Engineering.* 2020;12:224-32. doi: 10.30880/ijie.2020.12.05.028.
  - 16 Maharlouei N, Sarkarinejad A, Shahraki HR, Rezaianzadeh A, Lankarani KB. Socio-economic Status and Child Developmental Delay: A Prospective Cohort Study. *Shiraz E-Medical Journal.* 2021;22. doi: 10.5812/semj.100166.
  - 17 Zarean E, Looha MA, Amini P, Ahmadi A, Dugue PA. Sleep characteristics of middle-aged adults with non-alcoholic fatty liver disease: findings from the Shahrekord PERSIAN cohort study. *BMC Public Health.* 2023;23:312. doi: 10.1186/s12889-023-15251-4. PubMed PMID: 36774488; PubMed Central PMCID: PMCPMC9922458.
  - 18 Ahmadi A, Shirani M, Khaledifar A, Hashemzadeh M, Solati K, Kheiri S, et al. Non-communicable diseases in the southwest of Iran: profile and baseline data from the Shahrekord PERSIAN Cohort Study. *BMC Public Health.* 2021;21:2275. doi: 10.1186/s12889-021-12326-y. PubMed PMID: 34903205; PubMed Central PMCID: PMCPMC8670056.
  - 19 Ahmadi A, Tajji F, Shahraki HR. Comparing health service reception in individuals with and without non-communicable diseases before and during the COVID-19 pandemic: Shahrekord cohort study. *Iranian Journal of Endocrinology and Metabolism.* 2023;24:92-100. Persian.
  - 20 Maechler M. Finding groups in data: Cluster analysis extended Rousseeuw et al. R package version. 2019;2:242-8.
  - 21 Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software.* 2014;61:1-36. doi: 10.18637/jss.v061.i06.
  - 22 Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New Jersey: John Wiley & Sons; 2009.
  - 23 Mtintsilana A, Craig A, Mapanga W, Dlamini SN, Norris SA. Association between socio-economic status and non-communicable disease risk in young adults from Kenya, South Africa, and the United Kingdom. *Sci Rep.* 2023;13:728. doi: 10.1038/s41598-023-28013-4. PubMed PMID: 36639432; PubMed Central PMCID: PMCPMC9839722.
  - 24 Reddy MM, Zaman K, Yadav R, Yadav P, Kumar K, Kant R. Prevalence, Associated Factors, and Health Expenditures of Noncommunicable Disease Multimorbidity-Findings From Gorakhpur Health and Demographic Surveillance System. *Front Public Health.* 2022;10:842561. doi: 10.3389/fpubh.2022.842561. PubMed PMID: 35462842; PubMed Central PMCID: PMCPMC9019118.
  - 25 Marthias T, Anindya K, Ng N, McPake B, Atun R, Arfyanto H, et al. Impact of non-communicable disease multimorbidity on health service use, catastrophic health expenditure and productivity loss in Indonesia: a population-based panel data analysis study. *BMJ Open.* 2021;11:e041870. doi: 10.1136/bmjopen-2020-041870. PubMed PMID: 33597135; PubMed Central PMCID: PMCPMC7893673.
  - 26 Lotfi MH, Amiri F, Forouzannia SK, Fallahzadeh H, Shekari H. The Association Between Socio-Economic Factors and Coronary Artery Disease in Yazd Province: A Case-Control Study. *Journal of Community Health Research.* 2014;3:168-76. Persian.
  - 27 Kundu J, Chakraborty R. Socio-economic inequalities in burden of communicable and non-communicable diseases among older adults in India: Evidence from Longitudinal Ageing Study in India, 2017-18. *PLoS One.* 2023;18:e0283385. doi: 10.1371/journal.pone.0283385. PubMed PMID: 36996071; PubMed Central PMCID: PMCPMC10062644.

28 Biswas T, Islam MS, Linton N, Rawal LB. Socio-Economic Inequality of Chronic Non-Communicable Diseases in Bangladesh. PLoS One. 2016;11:e0167140. doi:

10.1371/journal.pone.0167140. PubMed PMID: 27902760; PubMed Central PMCID: PMC5130253.