International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

Employment of Cluster Analysis to Study the Classification Variables for the Youth Development in Iraq

Huda Mahdi Ahmed and Amenah Hameed Mahmood

Abstract--- *Cluster analysis is one of the branches of multivariate statistical analysis. It is a set of procedures that seek to classify the sample of observations into two or more dependent categories, depending on the combinations of the variables categories, for discovering common characteristics to organize the views (individuals) and dividing them into groups with the same characteristics. The young people are the cause of the rise of the nations, the secret of their strength, Therefore, the effective role played by the state in the development of societies is to create a mentally and scientifically qualified young people to serve their country and society appropriate solutions. In this study we used two different type of clustering analysis methods, hierarchical, and k-means, and used the classification from two to four groups. So, it was concluded that the Iraqi cities were classified into four groups depending on some variables of youth development in Iraq. Moreover, the cities of the second group are characterized by the highest rate of economic activity, while the cities of the third group have the highest rate of literacy, the cities of the first group have the highest rate of food insecure, while the cities of the fourth group have the highest rate of youth who have the best health.*

Keywords--- *Agglomeration Schedule, K- means, Hierarchical Cluster, Cluster Analysis, Proximity Matrix.*

I. INTRODUCTION

The young people are the essential cause of the rise of the nations, the secret of their energy, their durability and the social necessity of human development. Therefore, the effective role played by the country in the development of societies is to create a mentally and scientifically qualified young people to serve his/her country and society. Hence, for the mentioned reasons the research, analysis and evaluate some variables of young people to develop their ability and using the cluster analysis in the Iraqi cities to know the weakness and strong points experienced by these cities for the variables of youth development and treatment and develop appropriate solutions. The research problem is represented by developing young people in society in general and especially in Iraqi society due to the young people are the most important part in the progress of society and the secret of strength and social necessity in human development, and this is carried out by cluster analysis method. The variables are represented by (economic activity rate, literate rate, food insecure, the percentage of young people that assessed in good health) that are classified depending on the Iraqi cities by using the Hierarchical Cluster Analysis and the K-Means algorithms. The cluster analysis literature is vast and heterogeneous (the yearly Classification Literature Automated Search Service lists many books and hundreds of papers on the topic in each issue).Cluster analysis algorithms draw upon statistics, mathematics and computer science. Cluster analysis was initially used within the disciplines of biology and ecology

Huda Mahdi Ahmed, Inst., Statistic Department, University of Diyala, Diyala, Iraq. E-mail: Huda_ahmed10@yahoo.com
Amenah Hameed Mahmood, Asst. Inst., Ruqayia Primary School for Girls, Ministry of Education. E-mail: Zxww33ky@gmail.com

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

(Sokal & Sneath, 1963).Although this technique has been employed in the social science, it has not gained the same widespread popularity as in the natural sciences. A general interest in cluster analysis increased in the 1960s, resulting in the development of several new algorithms that expand possibilities of analysis. It was during this period that researchers began utilizing various innovative tools in their statistical analysis to uncover underlying structures in data sets. And its algorithms reached a high point recent decades, there has been a gradual incorporation of cluster analysis into other areas, such as the health and social sciences. However, the use of cluster analysis within the field of psychology continues to be infrequent (Borgen & Barnett, 1987). (Youssef Haithem et al in 2011 and 2006) which is applied to the reality of the young development guide in Iraq and classified by cities, including (Economic activity rate, Literacy rate, food insecure, and the percentage of young people that assessed in good health).

II. METHODOLOGY

The analytical description approach were used to carry out this work, and the results which are obtained from the SPSS23.0 are displayed with the next tables and results. The approaches depends on the bag of statistical social sciences. Cluster analysis is applied to the reality of guide young development in Iraq for the ages of 15 years which is classified in 2014, and according to the governorates and included the following:

- Economic activity rate.
- Literacy rate.
- Food insecure.
- Percentage of young people that assessed in good health.

Table (1) shows the variables of the young development directory in Iraq for the year were used, that are classified by different governorates which carried by [3].

Table 1

Cities	Food insecure	Economic activity rate	Literacy rate	The percentage of young people that assessed in good health
Duhok	**24.4**	**31.5**	**79.7**	**98.7**
Ninawa	**19.3**	**38.2**	**79.9**	**98.1**
Alsulaimania	**9.9**	**38**	**90.7**	**98.1**
Kurkuk	**15.9**	**37.9**	**89**	**99**
Erbil	**10.7**	**34.9**	**82.3**	**98.8**
Diayla	**16.5**	**37.3**	**92.4**	**99.1**
Alanbar	**11.3**	**43.6**	**87.6**	**99.2**
Baghdad	**14.8**	**38.4**	**91.9**	**98.9**
Babelon	**14.1**	**40.9**	**87.9**	**98.9**
Karbalaa	**11.7**	**39**	**83.1**	**98.1**
Wasit	**24.2**	**44.2**	**80.5**	**97.6**
Salahaldeen	**12.8**	**39.3**	**81.7**	**98.1**
Alnajaf	**18**	**43.7**	**78**	**97.6**
Alkadisia	**18.9**	**35.5**	**73.3**	**98.5**
Almothana	**12.9**	**38.5**	**74.1**	**99.1**
ThiQar	**21.3**	**35.6**	**75.6**	**98.6**
Maisan	**15.2**	**37.3**	**72.2**	**99**
Albaserah	**27.3**	**37.3**	**85**	**98.5**

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

In the hierarchical clustering procedure with SPSS, we can standardize variables in different ways. We can compute standardized scores or divide by just the standard deviation, range, mean, or maximum. This results in all variables contributing more equally to the distance measurement. That's not necessarily always the best strategy, since variability of a measure can provide useful information. If variables are measured on different scales, variables with large values contribute more to the distance measure than variables with small values.

Cluster analysis is one of the branches of multivariate statistical analysis. It is a set of procedures that seek to classify the sample of observations into two or more dependent categories, depending on the combinations of the variables categories, for discovering common characteristics to organize the views (individuals) and dividing them into groups with the same characteristics. Assuming we have the following matrix [5]:

$$Y = \begin{pmatrix} \acute{y}_1 \\ \acute{y}_2 \\ \vdots \\ \acute{y}_n \end{pmatrix} = (y_{(1)} \quad y_{(2)} \quad \cdots \quad y_{(n)}) \ (1)$$

Where

$\acute{y}_i$: $Represents\ a\ vector\ Views$

$y_{(j)}$: $It\ represents\ a\ vector\ variables$

The purpose of the analysis is to discover a particular pattern that regulates the observations shown in matrix (1), whose elements have common properties, enabling us to coordinate these observations in specific groups [8].

2-1 The proximity matrix between objects

It is a similar matrix whose number of rows is equal to the number of columns. Cluster analysis usually starts with the formation of the matrix, whose components represent one of the measures of distance between observations. The idea is to link units that are similar to each other in separate groups. So that the general shape of that matrix can be represented as follows [4]:

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} (2)$$

Where

D: Represent the proximity matrix between objects.

n: Represent the observations.

p: Represent the variables.

The elements of the matrix represent the distance between observations or variables and measured that distance in one of the following formulas:

1- Euclidean distance square

The Euclidean space for variables (x, y) can be written in the following format:

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

$$d^2(x,y) = (x-y)'(x-y) = \sum_{j=1}^{p}\left(x_j - y_j\right)^2 \text{ (3)}$$

The formula (3) can be rewritten as

$$d^2(x,y) = \left(v_x - v_y\right)^2 + P(\overline{x} - \overline{y})^2 + 2v_x v_y\left(1 - r_{xy}\right) \text{ (4)}$$

Where:

$$v_x^2 = \sum_{i-1}^{p}(x_i - \overline{x})^2$$

And

$$\overline{x} = \sum_{j=1}^{p} x_j/p$$

Also

$$r_{xy} = \frac{\sum_{j=1}^{p}\left(x_j - \overline{x}\right)\left(y_j - \overline{y}\right)}{\sqrt{\sum_{j=1}^{p}\left(x_j - \overline{x}\right)^2 \sum_{j=1}^{p}\left(y_j - \overline{y}\right)^2}}$$

2- Euclidean distance

The Euclidean distance for variables (x, y) can be written according to the following formula:

$$d(x,y) = \sqrt{\sum_{j=1}^{p}\left(x_j - y_j\right)^2} \text{ (5)}$$

3- City Block distance

This distance for the variables (x, y) can be written according to the following formula:

$$d(x,y) = \sum_{j=1}^{p}\left|x_j - y_j\right| \text{ (6)}$$

2-2 Agglomeration Schedule

To calculate the distance between samples, where samples are distributed in group, we use the second phase of cluster analysis as following:

2-2-1 Hierarchical cluster analysis

The hierarchical method is used to find efficient groups of samples, It is characterized by the fact that it does not require prior knowledge of the group's number. In this way, the distribution of samples in groups according to two methods:

First method (The divisive technique)

This technique is applied to the assumption there is one set of data that can be divided into partial groups. These partial groups are also divided into smaller subsets and continue until each individual group has its own subset.

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

Second method (Agglomerative technique)

This technique is applied on the assumption that each term describes its own subset. Then similar partial groups are grouped into more comprehensive subsets, and this process is repeated several times until we reach a single subset of all data. To calculate the distance between partial groups, there are several important methods to do that [7]:

1- Single linkage (Nearest neighbor)

By using this method, we can find the smallest distance for each pair of groups and merge them together, according to the following formula:

$$D(A,B) = min\{d(y_i, y_j)\} \quad (7)$$

2-Complete linkage (Farthest neighbor)

By using this method, we can find the largest distance for each pair of groups and merge them together, according to the following formula:

$$D(A,B) = max\{d(y_i, y_j)\} \quad (8)$$

3-Average linkage

This method is used to merge two groups, depending on the distance between a points from the first group to the point from the second group according to the following formula [9]:

$$D(A,B) = \frac{1}{n_A n_B}\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} d(y_i, y_j) \quad (9)$$

4-Centroid

This method summarizes the calculation of the general average by combining the average multiplication of each group by the number of its items and dividing it by the total number of the items, as follows:

$$D(A,B) = d(\overline{y}_A, \overline{y}_B) \quad (10)$$

Where

$$\overline{y}_A = \sum_{i=1}^{n_A} y_i / n_A$$

And

$$\overline{y}_{AB} = \frac{n_A \overline{y}_A + n_B \overline{y}_B}{n_A + n_B}$$

5-Ward's methods

This method is based on the merger between the totals on the square of the distance inside the aggregates, which can be found to link the two groups like (A, B) as follows:

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \overline{y}_A)' (y_i - \overline{y}_A)$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \overline{y}_B)' (y_i - \overline{y}_B)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \overline{y}_{AB})' (y_i - \overline{y}_{AB})$$

Where

n_A: Represents the number of items in the first set.

n_B: Represents the number of items in the second set.

And

$$n_{AB} = n_A + n_B$$

Also

$$\overline{y}_{AB} = (n_A \overline{y}_A + n_B \overline{y}_B)/(n_A + n_B)$$

And $SSE_A, SSE_B, and\ SSE_{AB}$ Represents the sum of the squares within the group (A, B, AB) respectively, and that this method is based primarily on the merger of the two groups (A, B) that have the smallest SSE that define as follows:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

There are other ways such as the link between the groups and the link within the groups and linking depending on the broker and other.

2-2-2 K-means cluster analysis

This method requires prior knowledge of the number of groups so that the sets can be distributed in a flat way. In this way, the sets are distributed to groups on a given basis as a first step. The averages of these groups are then calculated and then redistributed and redistributed based on group averages as a second step. At the end of this step the group averages are calculated, and this process is repeated until the Sets are set so that the group does not change any individual [4].

III. RESULTS AND DISCUSSIONS

The statistical bag was used is SPSS 22.0, for extracting the results of cluster analysis and according to the following two methods, Hierarchical method, and K-means cluster analysis. These ways need to get many information, and that will be as follows:

3-1 Proximaty Matrix

To obtain the proximity between the objects for the matrix, the aggregation method was used (Linking groups) depending on the square distance of the Euclidean space. In the conversion values option, we obtain the proximity

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

between the objects for the matrix and notation on the bottom of the list on 1, 0 and the values of the standard is within the range. So, we can obtain the matrix depend on the distance of the Euclidean space as follows:

Table 2: Proximity Matrix

Case	Absolute Squared Euclidean Distance																	
	1:Duhok	2:Ninawa	3:Alsulaimania	4:Kurkuk	5:Erbil	6:Diayla	7:Alanbar	8:Baghdad	9:Babelon	10:Karbala	11:Wasit	12:Salahaldeen	13:Alnajaf	14:Alkadisia	15:Almothana	16:ThiQar	17:Maisan	18:Albaserah
1:Duhok	.000	.505	1.394	.740	.712	.872	1.725	.980	1.079	1.050	1.474	.972	1.538	.315	.880	.181	.661	.321
2:Ninawa	.505	.000	.578	.558	.517	.804	1.010	.670	.541	.220	.401	.155	.300	.215	.609	.198	.522	.343
3:Alsulaimania	1.394	.578	.000	.442	.426	.545	.697	.334	.380	.158	1.266	.237	.911	1.111	1.097	1.121	1.251	1.145
4:Kurkuk	.740	.558	.442	.000	.271	.036	.292	.030	.073	.467	1.416	.491	1.285	.767	.580	.632	.696	.568
5:Erbil	.712	.517	.426	.271	.000	.432	.602	.361	.342	.300	1.709	.327	1.264	.458	.296	.500	.368	.999
6:Diayla	.872	.804	.545	.036	.432	.000	.396	.033	.165	.697	1.717	.741	1.648	1.074	.872	.883	1.009	.660
7:Alanbar	1.725	1.010	.697	.292	.602	.396	.000	.289	.106	.654	1.675	.680	1.374	1.290	.620	1.221	.893	1.300
8:Baghdad	.980	.670	.334	.030	.361	.033	.289	.000	.080	.474	1.479	.523	1.342	1.018	.804	.874	.963	.703
9:Babelon	1.079	.541	.380	.073	.342	.165	.106	.080	.000	.348	1.199	.366	.999	.842	.523	.751	.692	.739
10:Karbalaa	1.050	.220	.158	.467	.300	.697	.654	.474	.348	.000	.798	.009	.429	.545	.595	.612	.666	.893
11:Wasit	1.474	.401	1.266	1.416	1.709	1.717	1.675	1.479	1.199	.798	.000	.679	.144	1.006	1.602	.936	1.497	.693
12:Salahaldeen	.972	.155	.237	.491	.327	.741	.680	.523	.366	.009	.679	.000	.341	.448	.536	.512	.581	.808
13:Alnajaf	1.538	.300	.911	1.285	1.264	1.648	1.374	1.342	.999	.429	.144	.341	.000	.790	1.170	.847	1.128	.976
14:Alkadisia	.315	.215	1.111	.767	.458	1.074	1.290	1.018	.842	.545	1.006	.448	.790	.000	.317	.036	.166	.589
15:Almothana	.880	.609	1.097	.580	.296	.872	.620	.804	.523	.595	1.602	.536	1.170	.317	.000	.388	.039	1.126
16:ThiQar	.181	.198	1.121	.632	.500	.883	1.221	.874	.751	.612	.936	.512	.847	.036	.388	.000	.232	.357
17:Maisan	.661	.522	1.251	.696	.368	1.009	.893	.963	.692	.666	1.497	.581	1.128	.166	.039	.232	.000	.983
18:Albaserah	.321	.343	1.145	.568	.999	.660	1.300	.703	.739	.893	.693	.808	.976	.589	1.126	.357	.983	.000

Table (2), represented that the smallest distance that is (0.03) that was between cities Bagdad and Kirkuk that depend on the indicators of young development in Iraq.

3-2 The agglomeration steps

The following table shows the steps of aggregation for the cities according to the changing of young development in Iraq.

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

Table 3: The result of agglomeration

Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	4	8	.030	0	0	2
2	4	6	.034	1	0	5
3	14	16	.036	0	0	6
4	10	12	.072	0	0	7
5	4	9	.106	2	0	8
6	14	17	.133	3	0	13
7	2	10	.219	0	4	10
8	4	7	.271	5	0	14
9	13	15	.295	0	0	13
10	2	5	.329	7	0	12
11	1	18	.340	0	0	15
12	2	3	.365	10	0	14
13	13	14	.427	9	6	15
14	2	4	.492	12	8	16
15	1	13	.692	11	13	16
16	1	2	.760	15	14	17
17	1	11	1.232	16	0	0

Note from the above table that the city number (4) Kirkuk has been linked to the city number (8) Baghdad being the square of the distance between them and the smallest possible number of (o.o3) then step 2 will be the next step to be moved to it, where we can note that from the second step has been connecting group (4), which was formed in the first step, was assigned to the number (6), which is Diyala city. The distance between them and measured in the square of the Euclidean distance reached (0.034). Then, move to step (5), which is connected to the group (4), which is the second step in the city number (9), that will be done now move to step (8) of connecting group (4), which is in step (5), with city number (7) Anbar province, and so on.

3-3 Groups membership:

The following table shows the cities membership according to each group and according to the variables related to the Young Development Index in Iraq, after the groups were divided groups from 2 to 4 groups.

Table 4: Distribute sets as members of groups-Cluster membership

Case	4 Clusters	3 Clusters	2 Clusters
1:Duhok	**1**	**1**	**1**
2:Ninawa	**2**	**2**	**1**
3:Alsulaimania	**2**	**2**	**1**
4:Kurkuk	**2**	**2**	**1**
5:Erbil	**3**	**1**	**1**
6:Diayla	**2**	**2**	**1**
7:Alanbar	**2**	**2**	**1**
8:Baghdad	**2**	**2**	**1**
9:Babelon	**2**	**2**	**1**
10:Karbalaa	**2**	**2**	**1**
11:Wasit	**4**	**3**	**2**
12:Salahaldeen	**2**	**2**	**1**
13:Alnajaf	**4**	**3**	**2**
14:Alkadisia	**1**	**1**	**1**
15:Almothana	**3**	**1**	**1**
16:ThiQar	**1**	**1**	**1**
17:Maisan	**3**	**1**	**1**
18:Albaserah	**1**	**1**	**1**

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

We can note from the table (4) the cities of (Najaf, Qadisiyah, Muthanna, Dhi Qar and Maysan) have been linked to Group 4. The cities of (Dohuk and Basra) have been connected to Group 1, Wasit city, Group 3, the cities in Group 2 and in the classification into four groups, and in the case of classification into three groups Wasit was connected to Group 3 and linking the two (Dohuk, Najaf, Qadisiyah, Muthanna, Dhi Qar, Maysan, Basra) It was linked to group 2 and in the case of classification into two groups, the city (Wasit) was connected to group 2 and the rest of the cities it has been linked to Group 1.

The following figure shows the ice sheets of the classification of these cities:

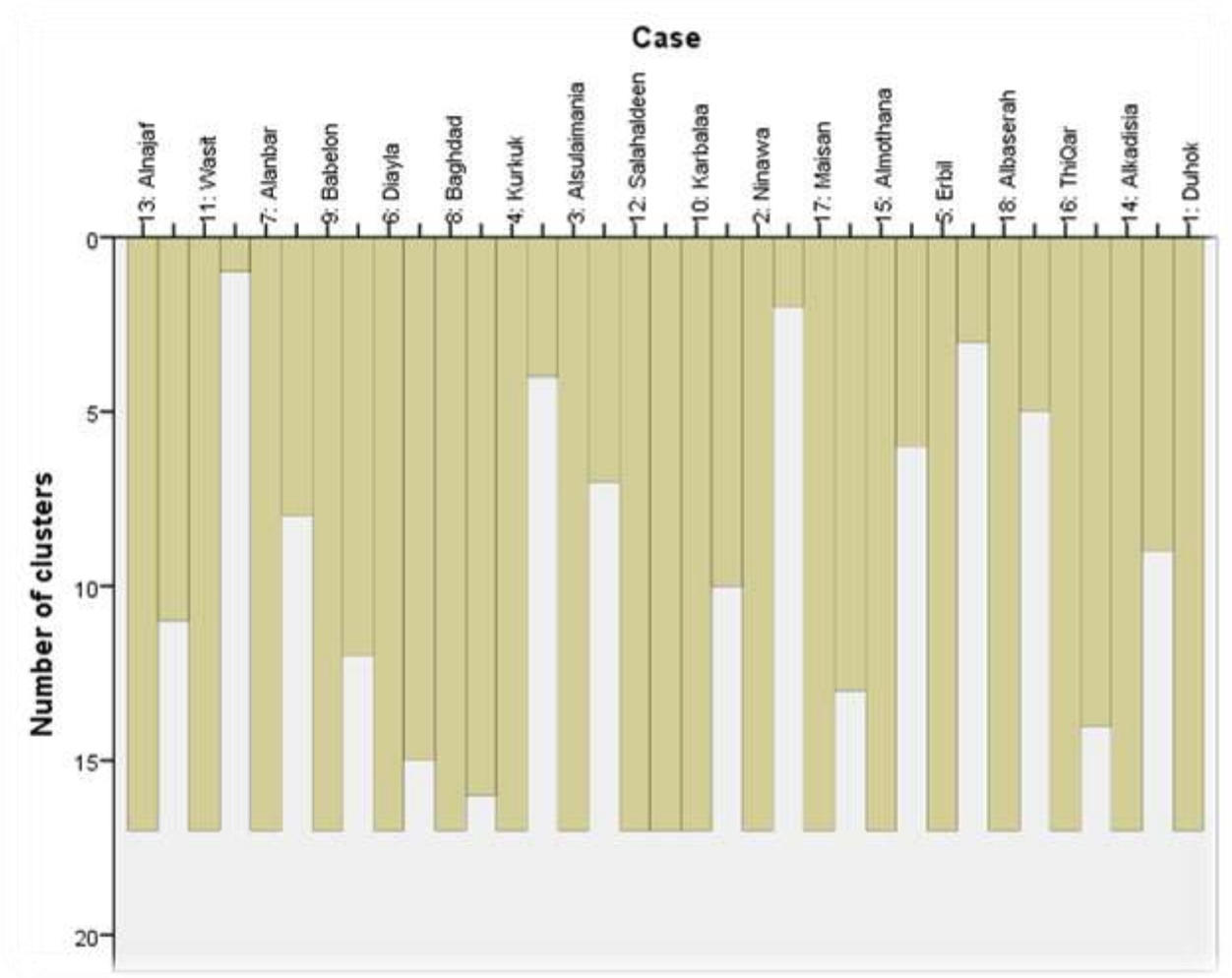

Figure 1: Ice sheet

Note from Figure (1) that both the cities of Kirkuk and Baghdad have been included as the square of the distance between them is the smallest of them, which is (0.03), and on this basis, the cities are combined and merged into groups.

3-4 Cluster analysis using Hierarchical method of variables

In this analysis, some variables of young development in Iraq are used to express variables. Where the collection of the variables of the research and not the cities to determine which of these variables are associated with the matrix

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

in the form of groups where the method of linking between groups. The standard adopted in the composition of the approximate matrix of the variables is the simple correlation coefficient as shown in the following table:

Table 5: Approximate matrix for the variables (young development)

Proximity Matrix				
	Matrix File Input			
Case	Economic_activity_rate	Literacy_rate	Food_insecure	The_percentage_of_young_people_that_assessed_in_good_health
Economic_activity_rate	.000	.394	1.409	.611
Literacy_rate	.394	.000	1.498	.518
Food_insecure	1.409	1.498	.000	1.810
The_percentage_of_young_people_that_assessed_in_good_health	.611	.518	1.810	.000

It is noted from the table (5) that the highest correlation coefficient was (1.810) among the percentage of young people who were evaluated as healthy and deprived of food security. The lowest correlation coefficient was between the literacy rate and economic activity rate reached 0.394.

3-5 Agglomeration Schedule

To collect the variables, we used the Hierarchical method as shown in the following table:

Table 6: Agglomeration Schedule

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	1	2	.394	0	0	2
2	1	4	.564	1	0	3
3	1	3	1.572	2	0	0

Note from Table (6) that the variable economic activity rate that was formed in the first step was combined with a variable deprived of food security, followed by the transition to the second step, which is represented those deprived of food security with the proportion of young people who have been evaluated in good health, thus the linkage is based on the value of the transactions.

3-6 Cluster membership for variables

After obtained aproximaty matrix and the step of Agglomeration ,distribute variables as members of groups has be evaluated by using the range of solution(2,4) from the order statistic in SPSS program as following table:

Table 7: Distribute variables as members of groups

Cluster Membership		
Case	3 Clusters	2 Clusters
Economic_activity_rate	1	1
Literacy_rate	1	1
Food_insecure	2	2
The_percentage_of_young_people_that_assessed_in_good_health	3	1

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

Notes from table(7) variable of Economic_activity_rate including in the first collection when classify to three collections, and also in the first collection when classifying in to tow collections whether variable of Literacy_rate including in the first collection when classifying to three collections and also include the first collection when classify in to tow collections ,the variable of Food_insecure put in the second collection when classifying three collections and in the second collection when classify tow collection,also the variable of The_percentage_of_young_people_that_assessed_in_good_health include in the third collectin when classifying in to three collectins and include the first set when classifying in to toe sets finally when classifying in to four sets all of variables includes in forth collections. The ice sheets explain that.

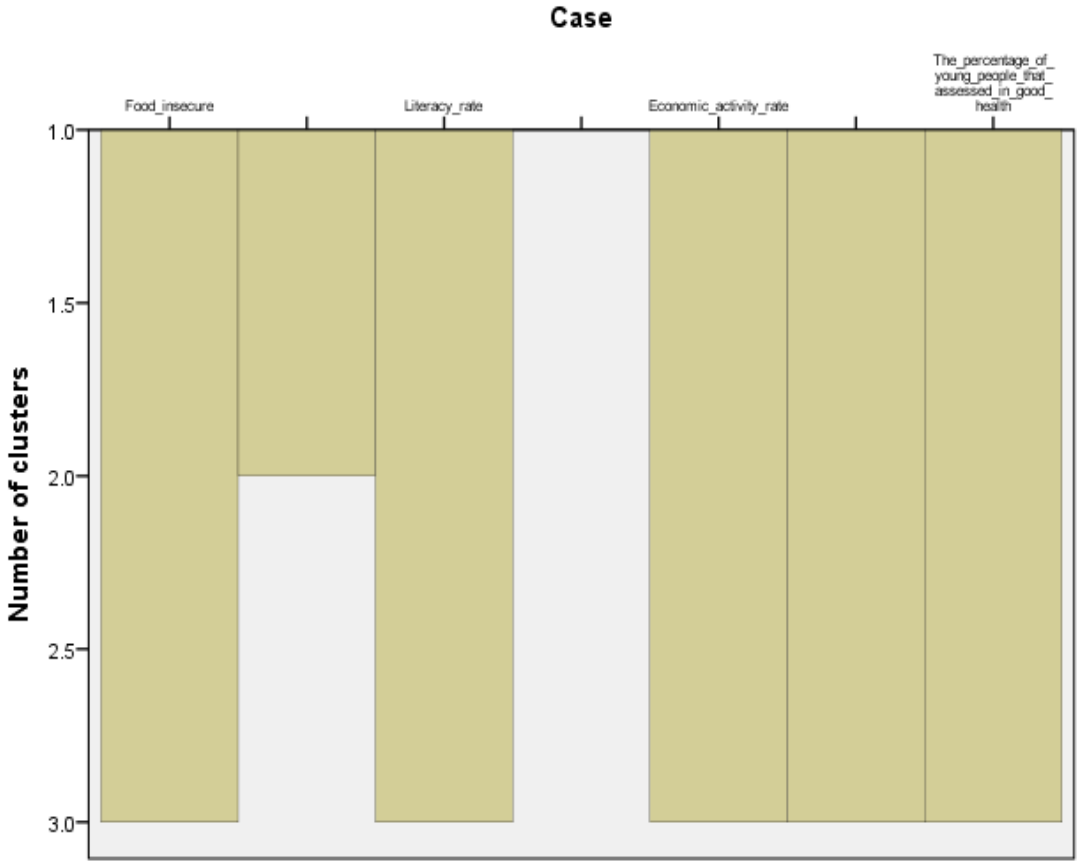

Figure 2: Ice sheets to distribute variables as members of groups

3-7 Results of analytic by using Method of K-Means

For applied this Method must find the following steps:

3-7-1 Cluster membership and distance for each member from the center of group for youth development problems in Iraqi cities

For classifying the cities to the special group for it's and measured the far of each member from the group which included it,we obtained these results

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

1-Cluster membership for youth development problems of Iraqi cities

Table 8: Cluster membership

Cluster Membership			
Case Number	**Iraqi_Cities**	**Cluster**	**Distance**
1	**Duhok**	**1**	**1.159**
2	**Ninawa**	**2**	**1.381**
3	**Alsulaimania**	**3**	**1.442**
4	**Kurkuk**	**3**	**.931**
5	**Erbil**	**4**	**1.257**
6	**Diayla**	**3**	**1.386**
7	**Alanbar**	**3**	**1.738**
8	**Baghdad**	**3**	**.844**
9	**Babelon**	**3**	**.681**
10	**Karbalaa**	**3**	**1.409**
11	**Wasit**	**2**	**1.055**
12	**Salahaldeen**	**3**	**1.504**
13	**Alnajaf**	**2**	**.817**
14	**Alkadisia**	**1**	**1.146**
15	**Almothana**	**4**	**.650**
16	**ThiQar**	**1**	**.580**
17	**Maisan**	**4**	**.776**
18	**Albaserah**	**1**	**1.520**

Note from above table,each of cities(Erbil, Almothana, Maisan) lies in the fourth group, and Erbil city is the farther member from the center of this group because it have the distance (1.257),also that the cities of third group are(Alsulaimania, Kurkuk, Diayla, Alanbar, Baghdad, Babelon, Karbalaa, Salahaldeen)which the farthest member for this group is Babelon which have the distance (0.681),and the cities of second group are(Ninawa, Wasit, Alnajaf) the farest member from this group is Alnajaf which has the distance(0.817).Finally the first group cities are (Duhok, Alkadisia, ThiQar, Albaserah)which have the farest city from the center of this group is (0.580).

2-Cluster membership for variables

The variables divided into four groups as follows:

3-7-2 The distance of the Initial Cluster Centers

The following table shows the **Initial Cluster Centers** of the group status as shown in the following table:

Table 9: Initial Cluster Centers

Initial Cluster Centers				
	Cluster			
	1	2	3	4
Zscore(Economic_activity_rate)	-2.15423-	1.79808	-.13140-	1.64247
Zscore(Literacy_rate)	-.43361-	-.30947-	1.27324	-.69739-
Zscore(Food_insecure)	1.52605	1.48681	-1.31895-	.27033
Zscore(The_percentage_of_young_people_that_assessed_in_good_health)	.07389	-2.36437-	-1.25607-	.96052

Note from the table (9) that the cities of the second group suffer from the problem of economic activity rate more than the rest of the cities, equivalent to (1.79808) of Standard deviation, while the third group cities suffer from the problem of literacy rate, as for the problem of those deprived of security food, the cities of the first group were

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

suffering from this problem with the equivalent of (1.52605) standard deviations, while the percentage of young people who were healthy in the fourth group was (0.96052) of the standard deviation value.

3-7-3 standard deviation of the groups for the youth development problems in Iraqi Cities

Table 10: Standard deviation of the groups

Final Cluster Centers				
	Cluster			
	1	2	3	4
Zscore(Economic_activity_rate)	-1.07279-	.86446	-.06527-	.77110
Zscore(Literacy_rate)	-.63532-	-.35602-	.75730	-.70127-
Zscore(Food_insecure)	1.24646	1.00611	-.65184-	-.44583-
Zscore(The_percentage_of_young_people_that_assessed_in_good_health)	.07389	-1.81022-	.01847	.79428

Table (10) shows that the highest average for the variable (economic activity rate) was in the governorates of the second group where it attain (0.86446) of the standard deviation. The literacy rate reached the highest average in the governorates of the third group (0.75730) of the standard deviation, and for the variable of the food insecure, it reached the highest average (1.24664) of the standard deviation in the governorates of the fourth group.

3-7-4 Distances between Final Cluster Centers

In order to measure the distance between the group centers of the governorates in order to know the extent of convergence or divergence of these groups from each other, the following was found:

1- Distance between group centers for youth development problems:

Table 11 shows the distance between group centers for youth development problems, note from table (11) below, that the average of the third group is a far beyond from the cities of first group which is 2.761 . While the average of the cities of the first group is closer to the fourth group cities which is (2.225).

Table 11: Distances between Final Cluster Centers

Cluster	**1**	**2**	**3**	**4**
1		**2.756**	**2.761**	**2.225**
2	**2.756**		**2.764**	**3.253**
3	**2.761**	**2.764**		**2.065**
4	**2.225**	**3.253**	**2.065**	

2- (Table of variance analysis ANOVA)

The one-way variance analysis table was calculated for the variables of youth development problems in Iraq as follows:

Table 12: ANOVA

Z score (The percentage of young people That assessed in goodhealth	Cluster		Error		F	Sig.
	Mean Square	Df	Mean Square	df		
	4.257	3	.302	14	14.090	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

Note from the table above that the average problems of youth development in Iraq for the variable percentage of young people who have evaluated the health of the least differences between the groups and equivalent to 14.09 are equal for all variables.

IV. CONCLUSIONS AND RECOMMENDATIONS

The mentioned results in the research, the Iraqi governorates for the youth group variables were classified into four groups using cluster analysis classified as follows. First group included all governorates except Wasit governorate, second group(**Wasit**, **Alnajaf, Ninawa**) Third group is Wasit Governor only, and the last and fourth group: the governorates (Dohuk, Basrah). The governorates of the first group suffer from the problem of a lack in the rate of economic activity among the youth. The governorates of the second group suffer from the problem of the health sector, represented by the ratio of young people who were assessed healthy. The governorates of the fourth group suffer from the problem of those disadvantaged of food security.

The need to provide the provisioning card items, especially for the provinces that suffer from the problem of lack of economic activity rate. Support the local market and provide the basic foodstuffs necessary to provide food security. The importance of supporting the agricultural sector as it plays a major role in filling the shortfall in the food aspect. Providing health centers and hospitals in governorates that suffer from the problem of the health sector. Providing health staff with doctors and nurses, which is important in addressing the problem of the health sector. Providing health supplies, including medicines, medical devices, etc. The provision of educational institutions in the governorates of the third group, including the construction of secondary schools, institutes and other educational institutions that contribute to solving the problem of low literacy rate and contribute to the dissemination of awareness and culture among the groups of society, especially the youth. Provision of teaching staff for the governorates of the third group.

REFERENCES

[1] Statistical Software SPSS22.
[2] United Nations Development Program, Human Development Report 1990.
[3] National Report of the Human Development Situation in Iraq 2014.
[4] Okasha, Mahmoud Khaled (2002).” Using SPSS in Data Analysis (In Arabic)”, Al-Azhar University - , Gaza.
[5] Dr. Ahmed Ahmed Abou Bakr "Development of financial analysis in quantitative methods to predict financial crises in life insurance companies (applied to the Egyptian insurance market) Faculty of Commerce - Beni Suef University – Egypt).
[6] Youssef, MD Haitham Yacoub. Pharaoh, Hesham. Mohammed, Eng. Jafar Kassem, "Using Cluster Analysis to Evaluate the Dimensions of the Human Development Index in Iraq 2006", *Diyala Journal of Human Scientific Research*, vol. 49, 2011.
[7] Al-Shamarti, Hamed Saad Nour. Al Tamimi, LLC. Suhad Ali "Using Cluster Analysis to Evaluate Environmental and Sustainable Development Indicators in the Governorates of Iraq for the Years 2007-2011" *Journal of Management and Economics - 35th Year*, No. 99, 2012.
[8] Taha, Hudhayfah Hazem. Hussein, Mohammed Hussein, "Using cluster analysis to classify the quality of groundwater in the wells of Baasheqa area in Nineveh province" *Iraqi Journal of Statistical Sciences* 2012.
[9] National Human Development Report 2014.
[10] Johson, R.A., Wichern, D.W.1988 "Applied multivariate"**2nd**, ed. Prentice hall, Englewoon clffs, New Jersy.

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020

[11] Alvin C. Rencher 2002"Method of Multivariate analysis"$\mathbf{2^{nd}}$, ed. Wiley series in prob. And statistics, ch.14.
[12] W. Hardle, L. Simar, "Applied multivariate statistical analysis"2007",$\mathbf{2^{nd}}$, ed. Berlin and Lourain-la-Neuve.
[13] Wolfgang Hardel2007"Maltivariate statistics" Barlin and Prag.
[14] Tan, Steinbach, Karpatne, Kumar 2018, " Introduction to Data Mining, 2nd Edition".
[15] Battaglia, O. R, Di Paola, B. and Fazio, C.(20160 A New Approach to Investegate Students' Behavior by Using Cluster Analysis as an Unsupervised Methodology in the Field of Education, Applied Mathematics,7,1649-1673.

DOI: 10.37200/IJPR/V24I5/PR2020146
Received: 08 Mar 2020 | Revised: 26 Mar 2020 | Accepted: 04 Apr 2020