Using Fuzzy Clustering Analysis to Classify the Provinces of Iraq Based on Infrastructure Deprivation Levels

ABSTRACT:

Aims: The research aims to use fuzzy clustering analysis to classify the provinces of Iraq based on their levels of infrastructure deprivation, by dependent on several indicators.

Study design: Depending on the infrastructure indicators such as drinking water, sanitation and time it takes to reach the source to control and stabilize electricity and supply it the infrastructure network continues, Triple deprivation level.

Study design: Depending on the infrastructure indicators such as drinking water, sanitation, the time required to reach the source, waste collection, electricity stability and power supply, households' evaluation of infrastructure services, and the level of triple deprivation. For 18 governorates of Iraq

Methodology: Central Agency for Statistics and Information Technology, basic tables, inventory and numbering results, environmental survey for the year 2010.

Results: Including the number of clusters that can be from the silhouette (Fc(U) equals 0.0390 corresponding to cluster four with reduction (Dc(U) equals 0.9036 corresponding to cluster four. Therefore we choose the number of clusters (4) clusters.

Conclusion The fuzzy cluster analysis using the NCSS system (12) with the number of clusters (4) clusters showed that the most deprived governorates are Ninawa - Kirkuk - Anbar - Babylon - Wasit - Salah al-Din and the least deprived governorates are Dohuk - Erbil And Sulaymaniyah.

Keywords: Fuzzy Concept, Fuzzy c-means (FCM), Fuzzy cluster structure viability test

1.1 INTRODUCTION

Clustering or cluster analysis is a typical method for grouping data points (elements) within the context of an undirected classification. A set of data is divided into several subsets or clusters based on the similarity of the elements, where the elements within one cluster have a high degree of similarity while the elements belonging to other and different clusters have a high degree of dissimilarity. The process of classifying elements into clusters is based on the criteria placed on these elements. For example, suppose we have a set of data that is placed in the vector of the elements' properties belonging to the sample space. In that case, the goal is to segment or determine the subsets or clusters of similar elements based on a set of vectors of the elements' properties. Clustering includes several specific steps, including determining the appropriate distance between the elements based on appropriate properties and characteristics, and then the appropriate clustering algorithm must be chosen and applied. The selection and classification of clustering algorithms depends on the following:

- The type of data entering the method.
- Clustering criterion by defining the similarity measure between elements.
- Determining the basic theories and concepts that show what cluster analysis methods depend on (such as fuzzy theory and statistics).

The concept of uncertainty has gone through two stages in its treatment, the first stage is represented by traditional theories and the second stage is represented by modern theories. When the data points (elements) are distributed within well-separated clusters, then the classification process is carried out clearly and without any problem. In most cases, the elements in the data set cannot be divided into well-separated clusters, then there will be arbitrariness in assigning the elements to the specific clusters. For example, if there is an element (x_j) located near the edges of two clusters but slightly close to one of the clusters (s_i), then it is appropriate to determine the weight (u_{ij}) that determines the degree of belonging of

(x_j) to the cluster (s_i). Most often, probabilistic methods such as mixture models ($u_{ij} = p_{ij}$) are used to estimate the probability that an element (x_j) belongs to a cluster (s_i) , but when there is difficulty in determining the appropriate statistical model, it is convenient to provide non-probabilistic cluster methods that provide the same fuzzy properties.

Fuzzy Clustering Methods are based on modern theories, including:

1. Fuzzy Set Theory.

2. Possibilistic Theory.

Fuzzy Cluster Analysis represents a special case of cluster analysis and is characterized by flexibility. Fuzzy cluster analysis allows for overlapping groups, which can be useful when the data has a complex structure or when there are vague or overlapping category boundaries. Additionally, it offers robustness, as fuzzy clustering can be more resilient against outliers and noise in the data by allowing a more gradual transition from one group to another. Moreover, it enhances interpretability by providing a more precise understanding of the data structure, Fuzzy clustering allows for gradual transitions between clusters and enhances interpretability, as it allows for a more detailed representation of the relationships between data points and clusters.

1.2 THE AIM OF RESEARCH:

The research aims to use fuzzy clustering analysis to classify the provinces of Iraq based on their levels of infrastructure deprivation, relying on several indicators such as drinking water, sanitation, the time required to reach the source, waste collection, electricity stability and power supply, households' evaluation of infrastructure services, and the level of triple deprivation.

1.3 THE PROBLEM OF RESEARCH:

The problem of this research arises from the fact that Iraq's 18 provinces are considered to have the same relative importance in terms of infrastructure deprivation. However, this contradicts the practical reality. According to the data under study, the provinces were classified into clusters (or groups) based on a ranking from the most deprived to the least deprived. This aims to draw the attention of policymakers and legislators in the country to prioritize and accelerate the provision of services to the most deprived provinces.

1.4 THE IMPORTANCE OF RESEARCH:

The importance of the fuzzy clustering method comes in analyzing the data of deprivation from basic infrastructure in Iraq for several variables using fuzzy clustering. This work has the potential to guide policymakers in prioritizing resources and interventions to reduce disparities between Provinces. By using different clustering methods, the fuzzy clustering method is more flexible and capable of dealing with diversity and complexity in data, as it allows separating data into groups or clusters based on similarity and close relationships between them, which allows planners and policymakers to prioritise the most deprived governorates and then the least to achieve social justice in a way that serves and satisfies everyone .

2.1 PREFACE:

The significance of cluster analysis lies in its ability to classify elements or observations by grouping multiple and homogeneous data. The main objective of this analysis is to gather elements of a homogeneous group into different sections based on specific indicators under study or research. Fuzzy cluster analysis, as a special case of cluster analysis, is particularly popular in the field of fuzzy sets and systems and other areas. The clustering of digital data is the basis for several classification and system-modelling algorithms. Clustering aims to identify natural clusters within a large data set, which helps produce a concise representation of the system's behaviour.

2.2 FUZZY CONCEPT:[2] [7][9][11][16]

The fuzzy concept was proposed by Zadeh in (1965) which aims to determine the degree of belonging of each element within the fuzzy set. We often find in our real life many uses towards data analysis, especially those related to determining the nature of the data form and how to group it into entities or clusters so that this entity is homogeneous in terms of its points, as the grouping of these points depending on the hard-clustering method, meaning that the point belongs to the cluster. The grouping of these points is consistent with the theoretical structure of classical sets. If we assume that x_i is an element belonging to set A, which is a partial set of the comprehensive high is a non-empty set, then set A is called the set space that contains all the elements defined within set C, meaning that: $UA(x) = \begin{cases} 1 & if\ x\ \in\ A \\ 0 & x\ \notin\ A \end{cases}$

$U_A(x)$: cd

However, in reality, points can belong to two or more clusters, and this problem occurs within clustering methodologies, as there can be a point that uniquely shares with all clusters through a certain membership degree. Therefore, in such a case, fuzzy clustering is used. It is also worth noting that multivariate data suffers from interference in the boundaries of groups (clusters) that hinder the possibility of obtaining single and completely separated cluster groups to determine their patterns. We can clarify the difference between normal clustering and fuzzy clustering through Figure (2) and Figure (1).

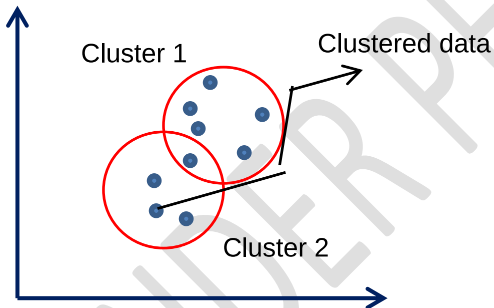

Figure (1) shows non-separate clusters

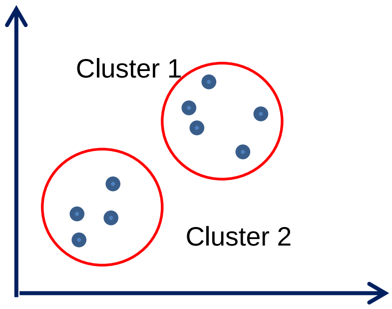

Figure (2) shows completely separated clusters

2.3 CLUSTER ANALYSIS:

Cluster analysis or clustering is the task of grouping a set of items in such a way that items in the same group (called a cluster) are more similar to each other than to those in other groups. It is an exploratory data mining task, and a popular technique is known as cluster analysis. There is no specific definition of a cluster, which is one of the reasons why there are so many clustering algorithms. Again, different algorithms can be given for each clustering model because the cluster concept is different.

2.4 FUZZY CLUSTERING ANALYSIS USING THE METHOD C-MEANS : [12] [13][4]

The fuzzy C-means method represents an unsupervised classification method belonging to the clustering partition class, as it was derived from the non-hierarchical sharp C-means method, where Dunn expanded the HCM method by extending the concept of regular element partitioning to fuzzy partitioning by allowing data elements to belong to all clusters with membership degrees falling within the interval [0,1].

In fuzzy clustering methods, membership is distributed among all possible clusters so that the value of m_{iK} can range from 0 to 1, provided that the sum of these values is equal to 1. This property, known as fuzziness in cluster classification, gives the important advantage of not imposing a definite membership of every object in a specific cluster. However, it has the disadvantage of requiring the interpretation of a larger amount of information. In contrast, partitioned clustering methods such as K-Means and Medoid classify an object into only one cluster. For example, if we assume that there are K clusters, a set of variables m_{i1}, m_{i2}, ... and m_{iK} are defined that represent the probability of object i being classified into cluster k. In this case, one of the values is 1 and the rest is 0, which means that the object belongs to only one cluster.
The need to develop fuzzy clustering comes to understand data cases that contain noise or lack of clarity in belonging to a specific group, such as the two-variable data set whose values are plotted in the example below.

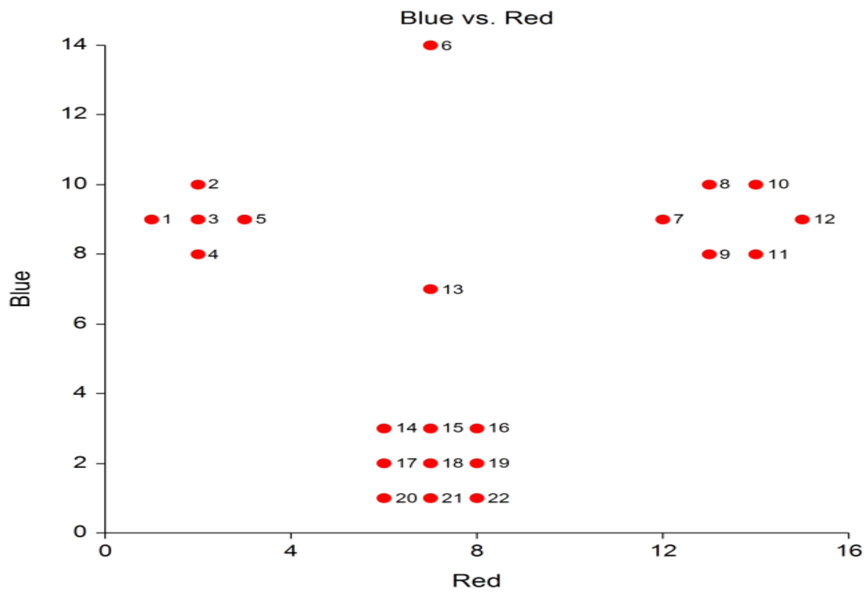

Fig 3- Fuzzy Clustering Analysis Using Method C-Means

From the above figure we can see that the data contains three clearly ordered clusters and contains two outliers (6 and 13). A conventional clustering algorithm, which seeks to identify three clusters, would result in these two points being included in certain clusters. This can result in distortions in the final solution.
However, fuzzy clustering will assign a probability of about 0.33 to each cluster, indicating equal membership, indicating that these two points are anomalous. When dealing with only two variables, the data can be plotted to identify clusters. However, most studies include more than two variables, making plotting impossible. Therefore, methods such as fuzzy clustering must be used to deal with potential anomalies. This method is a clustering of data such that each data point belongs to a cluster based on a degree that determines its membership. The FCM (Fuzzy Cluster-Based Clustering) algorithm begins with an initial estimate of the cluster centers, which reflect the average location of each cluster. The FCM assigns a membership degree to each data point for each cluster. By repeatedly updating the cluster centers and membership degrees, the algorithm moves the cluster centers toward optimal locations within the data set. This iteration is based on minimizing an objective function, which reflects the distance between any given data point and the cluster centre, taking into account the membership degree of that point.
To aggregate data, a set of data points x_i is selected such that their sum contains N rows. The number of columns for each data point is equal to the dimensions of the

$$x_j = [xj_1 xj_2 \cdots xj_n]^{\top}, \quad 1 \le j \le N \qquad (2.1)$$

The FCM algorithm calculates cluster centers. c_i This matrix contains one row for each cluster centre where the number of columns in it matches the number of columns in the data points x_i.

$$c_i = [c_{i1} c_{i2} \cdots c_{in}]^{\top}, \quad 1 \le i \le C \qquad (2.2)$$

The FCM algorithm works to minimize the following objective function:

$$J_m = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} D_{ij}^{2} \qquad (2.3)$$

Where:

- m: It is the power of the fuzzy partition matrix to control the degree of fuzzy overlap, with m > 1. Fuzzy overlap refers to how fuzzy the boundaries between groups are, i.e. the number of data points that have significant membership in more than one group. To determine the power of the fuzzy partition matrix
- D_{ij} : is the distance from data point j to the i-th cluster.
- m_{ij} : It is the membership degree of data point j in the ith set. For a given data point, the sum of the membership values of all sets is one.

Three types of FCM sets support the FCM function. These methods differ based on the distance metric used to compute D_{il}. To define the FCM algorithm [12].

Chart 1 FCM Algorithm

Description	Distance Measure Value	FCM Algorithm
Distances are computed using the Euclidean distance metric, which assumes a spherical shape for all clusters.[2]	"Euclidean"	Classical FCM
Distances are calculated using the Mahalanobis distance criterion, where the variance of the group is weighted by the group membership values. This method is useful for detecting non-spherical groups.[5]	"Mahalanobis"	Gustafson-Kessel FCM
Calculate distances using an exponential distance measure based on fuzzy maximum likelihood estimation. This method is useful for detecting asymmetric clusters in variable cluster densities and unequal numbers of data points in each cluster.[8]	"fmle"	Gath-Geva FCM

The FCM clustering algorithm executes the following steps [12]:

1. We randomly assign the initial group centers. Initialize the cluster membership values μij randomly.
2. Randomly determine the group affiliation values μij
3. Find the center of the group using the following formula:

$$c_i = \frac{\sum_{j=1}^{N}\mu_{ij}^{m} x}{\sum_{j=1}^{N}\mu_{ij}^{m}}, \quad 1 \le i \le C \qquad (2.4)$$

4 . Based on the complexity algorithm, the distance to each data point is measured. Using the Euclidean distance FCM

- Calculate the distance using the Euclidean FCM.
- Calculate the distance using the Gustafson-Kessel method.
- Calculate the distance using the Gath-Geva method.

5. Update the membership values for each data point using the following formula.

$$Subscript\ \mu_{ij} = \frac{1}{\sum_{k=1}^{N}\left(\frac{D_{ij}}{D_{ik}}\right)^{2m-1}}, \quad 1 \le i \le C, \quad 1 \le j \le N \qquad (2.5)$$

6. Calculate the objective function J_m
7. To satisfy one of the following conditions: The objective function Jm becomes equal to or less than the specified minimum. Or all iterations in the algorithm implementation are executed. We repeat the steps from step 3 to step 6.

2.2.1 CALCULATING THE DISTANCE USING EUCLIDEAN FCM: [12]

The classical FCM algorithm calculates the Euclidean distance from each data point to each cluster centre, as shown in the following equation.

$$\boldsymbol{D_{ij}} = \sqrt{(\boldsymbol{x_j} - \boldsymbol{c_i})^{\top}(\boldsymbol{x_j} - \boldsymbol{c_i})}, \quad \boldsymbol{1 \le i \le C}, \quad \boldsymbol{1 \le j \le N} \qquad (\mathbf{2.6})$$

2.2. CALCULATING THE DISTANCE USING GUSTAFSON-KESSEL [GUSTAFSON-KESSEL DISTANCE : [10][12]

The GK method is somewhat similar to the FCM method, in which the clusters are spherical (spherical cluster) and the cluster is not allowed to change its shape based on the data. Therefore, Gustafson-Kessel proposed a method that is an extension of the FCM method so that the centre can be adapted to OBE and ellipsoidal shapes (hyperbolic shapes) in order to adapt to different OBEs. The covariance matrix (covariance matrix) was used to extract ellipsoidal crops for clusters.

- Compute the covariance matrices Fi for each cluster centre.

$$F_i = \frac{\sum_{j=1}^{N} \mu_{ij}^{m} (x_j - c_i)(x_j - c_i)^{\top}}{S_i}, \quad 1 \le i \le C \qquad (2.7)$$

Where $S_i = \sum_{j=1}^{N} \mu_{ij}^{m}$

The Mahalanobis distance from each data point to each cluster is then calculated using the covariance matrices.

$$the\ the\ the\ D_{ij} = \sqrt{(x_j - c_i)^{\top}[det(F_i)^{1/N} F_i^{-1}](x_j - c_i)}, \quad 1 \le i \le C, \quad 1 \le j \le N\ ) \qquad (2.8)$$

2.2.3 CALCULATING THE DISTANCE USING GATH-GEVA (GG): [6] [12]

The Gath-Geva (GG) FCM algorithm calculates as follows:

- Compute the covariance matrices F_i for each cluster centre.

Equals$F_i = \frac{\sum_{j=1}^{N} \mu_{ij}^{m} (x_j - c_i)(x_j - c_i)^{\top}}{S_i}, \quad 1 \le i \le C \qquad (2.9)$

$$S_i = \sum_{j=1}^{N} \mu_{ij}^{m}$$

Then, calculate the prior probability for selecting each cluster.

$$P_i = \frac{S_i}{\sum_{i=1}^{C} S_i}, \quad 1 \le i \le C \qquad (2.10)$$

The distance from each data point to each cluster is calculated using the following exponential distance metric:

$$D_{ij} = A_i \cdot exp(0.5 \sum_{j=1}^{N} (x_j - c_i)^{\top} F_i^{-1} (x_j - c_i)), \quad 1 \le i \le C, \quad 1 \le j \le N \quad (2.11)$$

Where : $A_i = \frac{\sqrt{det(F_i)}}{P_i}$

2.2.4 FUZZY CLUSTER STRUCTURE VIABILITY TEST: [14][15]

After performing the clustering algorithms and forming the partial groups, whether sharp or fuzzy, choosing the validity of the clustering results and the extent to which they represent the target data set is another goal that must be taken care of, as most of the clustering methodologies are based on the principle of initial guessing of the clustering models and the number of clusters that can be formed from the data. Therefore, testing the validity of the cluster structure aims to verify the integrity of the structure assumed by the clustering method. This helps to evaluate the relationships between the cluster patterns that have been formed. Therefore, some measures were used to test the validity of the fuzzy cluster structure according to the following formula:

1- PARTITION COEFFICIENT SCALE : [5]

The partition coefficient scale was proposed to test the validity of fuzzy clustering by Bezdek (1974) based on the development in the field of fuzzy sets and fuzzy clustering in which the degree of belonging of each element to the clusters is determined to lay the theoretical foundation for this scale, and was developed by (Trauwaert) in (1988) to be compatible with the method of clustering averages (FCM), as this indicator measures the amount of overlap between clusters, and is defined according to the following formula:

$$\boldsymbol{the\ the\ F(U) = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N} M_{ik}^{2}} \qquad \mathbf{(2.12)}$$

The main goal of this measure is to maximize the distance between clusters (inter-cluster) and minimize the distance between cluster elements (intra-cluster). Therefore, large values of (D) measures indicate the presence of good clusters, and thus the number of clusters that maximize the measure (D) is considered the optimal number of clusters. The value of the coefficient ranges from 1/K to 1. Where the value is equal to 1/K when all memberships are equal to 1/K. The value is one when the value of one membership for each object is one and the remainder is zero. The Dan partition coefficient can vary from 0 (completely blurry) to 1 (hard cluster). It can be written in the following formula:

$$Fc(U) = \frac{F(U) - \left(\frac{1}{K}\right)}{1 - \left(\frac{1}{K}\right)} \qquad (2.13)$$

2- ANOTHER PARTITION COEFFICIENT,AS MENTIONED IN KAUFMAN (1990)

[13]: it is given by the following formula :

$$D(U) = \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}(\, h_{IK} - m_{ik})^2 \qquad (2.14)$$

This coefficient ranges from 0 (crisp clusters) to 1−1/K1 (completely fuzzy). The normalized version of this equation is given by the following formula:

$$Dc(U) = \frac{D(U)}{1-)1/K)} \qquad (2.15)$$

Both $F_C(U)$ and$D_c(U)$ together provide a good indicator for the optimal number of clusters. You should choose K such that $F_C(U)$ is large $D_c(U)$ is small.

3. **AVERAGE DISTANCE:**
   This is the average value of the difference. It should be noted that this value has been measured as a percentage of the maximum distance in the difference matrix to improve readability..
4. **AVERAGE SILHOUETTE:**
   It is the average value of the shadow indices for all rows (Medoid) that is used to support the process of finding the appropriate number of clusters. This is done by determining the number that maximizes this value, which contributes to improving the distribution of data within the clusters.

3.PRACTICAL PART

The field of infrastructure is one of the most critical areas where deprivation rates in this sector are evident, varying by province. According to the Central Bureau of Statistics and Information Technology, the basic tables, census, and enumeration results, and the Environmental Survey for the year 2010 [1], Table 1 shows the deprivation rates for infrastructure indicators by province for the year 2010. The deprivation rates of households in Iraq from infrastructure services at the national level reached 58.9%, which is very high. More importantly, the deprivation rate in some provinces reaches up to 95.5% of households in Muthanna and 89% in Diyala, Anbar, and Wasit. The lowest deprivation rate in this field was found in the northern provinces, with 31.7% in Erbil.

As for deprivation according to the indicators under study, the percentage of households deprived of sanitation services in Iraq is 86%, with the rate reaching 99% in some provinces, such as Diyala, 95% in Nineveh, and 97% in Muthanna.

Regarding the deprivation rate for the drinking water indicator, 47% of households in Iraq are deprived of this service, and the rate reaches 73% in Muthanna. The lowest deprivation rates were found in the northern provinces (Duhok, Erbil, and Sulaymaniyah).

Households in Iraq also suffer from deprivation in the waste collection indicator, with 44% of households lacking this service nationwide. The rate rises to 63% of households in Nineveh and 59% in Al-Qadisiya, while it drops to 9% in the three northern provinces.

Similarly, the deprivation rates for electricity as a primary source or its stability are significant. The proportion of households deprived of stable electricity supply at the national level is 69%. The rates increased to 95% of households in Karbala and 93% in Al-Qadisiya, while the lowest deprivation rate in this indicator was 23% in Basra, followed by Muthanna at 35%.

Households' evaluation of these services at the national level shows a deprivation rate of 84.7%. The highest deprivation rate was recorded in Muthanna, where it reached 95.5%, followed by Al-Anbar with 89.5%. The lowest deprivation rate was found in the northern provinces, at 31.7%.

TABLE 1: DEPRIVATION RATES FOR INFRASTRUCTURE INDICATORS BY PROVINCE FOR THE YEAR 2010

INFRASTRUCTURE	HOUSEHOLD ASSESSMENT OF INFRASTRUCTURE SERVICES	ELECTRICITY STABILITY	ELECTRIC POWER	TIME TAKEN TO REACH THE SOURCE	WASTE COLLECTION	SANITATION	DRINKING	GOVERNORATE	N	TRIPLE DEPRIVATION LEVEL
69.9	79.5	63	79	82	63	95	28	Ninawa	1	1.39
63.2	79.7	62	64	93	42	84	18	Kirkuk	2	1.26
59.6	89.2	58	4	88	51	99	28	Diyala	3	1.19
67.6	89.5	79	63	97	36	91	18	Anbar	4	1.35
54.9	80.7	87	61	54	32	34	18	Baghdad	5	1.09
66.6	85.2	91	21	96	45	98	30	Babil	6	1.33
55.9	74.7	95	9	75	23	73	42	Karbala	7	1.11
63.5	89.5	90	21	76	51	89	28	Wasit	8	1.27
58.8	85	76	93	95	30	91	23	Salahuddin	9	1.17
68.3	75.7	88	63	57	31	87	77	Najaf	10	1.36
65.1	87.2	93	4	87	59	85	41	Al-Qadisiya	**11**	1.30
61.2	95.5	35	2	100	26	97	73	Muthanna	**12**	1.22
60.7	87.2	70	2	65	55	79	67	The-Qir	**13**	1.21
63.5	83	77	11	98	42	40	94	Maysan	**14**	1.27
52	86.5	23	2	99	30	25	99	Basra	**15**	1.04
56	54	38	78	96	9	95	22	Duhok	**16**	1.12
39.8	31.7	62	19	64	9	85	8	Erbil	**17**	0.79
33.7	41	51	7	97	9	17	14	Sulaymaniyah	**18**	0.67

58.9	84.7	69	32	87	44	86	47	Iraq		1.17

Source [1]: Central Bureau of Statistics and Information Technology, Basic Tables and Results of Census and Enumeration, Environmental Survey for the Year 2010.

To conduct the statistical analysis using the NCSS (12) system to classify the provinces according to the indicators studied in Table 1, the following results illustrate this.

3.1 Descriptive Statistics: Below is Table (2) which shows the descriptive statistics.

Table(2) Descriptive statistics for the deprivation indicators in the field of infrastructure in Iraq.

Variable	Count	Mean	Standard Deviation	Standard Error	Lower 95% CLMean	Upper 95% CL Mean
C2	18	40.44444	28.57647	6.735538	26.2337	54.65519
C3	18	75.77778	26.89972	6.340324	62.40086	89.15469
C4	18	35.72222	16.78517	3.956302	27.37515	44.06929
C5	18	84.38889	15.46586	3.645338	76.6979	92.07988
C6	18	33.5	32.59962	17.28859	7.683805	49.71141
C7	18	68.77778	21.53945	5.076896	58.06646	79.48909
C8	18	77.48889	17.44935	4.112852	68.81153	86.16624
C9	18	58.90556	9.497645	2.238616	54.18249	63.62862
C10	18	1.174444	0.1902698	0.04484703	1.079826	1.269063

3.2 Results of the Fuzzy Clustering Analysis: Below are the results of the fuzzy clustering analysis.

3.2.1 Binary Clustering: Below is the table for the classification of provinces into two clusters.

Table(3) Classification of Provinces by Fuzzy Clustering Analysis into Two Clusters.

Cluster - 2	**Cluster – 1**	T
Bagdad	Ninawa	1
Karbala	Kirkuk	2
Muthanna	Diyala	3
The-Qir	Anbar	4
Basra	Babil	5
Duhok	Wasit	6
Erbil	Salahuddin	7
Sulaymaniyah	Najaf	8
	Al-Qadisiya	9
	Maysan	10

And below is the membership matrix section for the two clusters.

Chart 2- Membership matrix section for the two clusters

Membership Matrix Section: Distance Type-Euclidean Scale Type-Standard Deviation

Row	**Cluster**	**Prob in1**	**Prob in 2**
1			
1	0.5067	0.4933	
2	1	0.5068	0.4932
3	1	0.5022	0.4978
4	1	0.5072	0.4928
5	2	0.4986	0.5014

6	1	0.5079	0.4921
7	2	0.4991	0.5009
8	1	0.5069	0.4931
9	1	0.5047	0.4953
10	1	0.5010	0.4990
11	1	0.5037	0.4963
12	2	0.4982	0.5018
13	2	0.4996	0.5004
14	1	0.5001	0.4999
15	2	0.4969	0.5031
16	2	0.4976	0.5024
17	2	0.4969	0.5031
18	2	0.4968	0.5032

3.2.2 Ternary Clustering: Below is the table for the classification of provinces into three clusters.

Table (4) Classification of Provinces by Fuzzy Clustering Analysis into Three Clusters.

Cluster -3	Cluster -2	Cluster -1	T
Bagdad	The-Qir	Ninawa	**1**
Karbala	Maysan	Kirkuk	**2**
Muthanna		Diyala	**3**
Basra		Anbar	**4**
Duhok		Babel	**5**
Erbil		Wasit	**6**
Sulaymaniyah		Salahuddin	**7**
		Najaf	**8**
		Al-Qadisiya	**9**

And below is the membership matrix section for the three clusters.

Chart 3- Membership matrix section for the three clusters

Membership Matrix Section : Distance Type -Euclidean Scale Type-Standard Deviation

Row	Cluster	Prob in1	Prob in2	Prob in 3
1	1	0.3446	0.3374	0.3180
2	1	0.3449	0.3368	0.3183
3	1	0.3370	0.3361	0.3269
4	1	0.3454	0.3373	0.3172
5	3	0.3310	0.3323	0.3367
6	1	0.3466	0.3387	0.3147
7	3	0.3319	0.3336	0.3345
8	1	0.3449	0.3385	0.3166
9	1	0.3412	0.3356	0.3232
10	1	0.3350	0.3344	0.3305
11	1	0.3396	0.3372	0.3232
12	3	0.3304	0.3327	0.3369
13	2	0.3326	0.3343	0.3331
14	2	0.3336	0.3342	0.3321
15	3	0.3281	0.3313	0.3406
16	3	0.3293	0.3311	0.3396
17	3	0.3281	0.3308	0.3411
18	3	0.3279	0.3308	0.3414

3.2.3 Quadruple Clustering: Below is the table for the classification of provinces into four clusters.

Table(5) Classification of Provinces by Fuzzy Clustering Analysis into Four Clusters.

Cluster – 4	Cluster – 3	Cluster- 2	Cluster – 1	T
Duhok	Diyala		Ninawa	**1**
Erbil	Bagdad		Kirkuk	**2**
Sulaymaniyah	Karbala		Anbar	**3**
	Najaf		Babel	**4**
	Al-Qadisiya		Wasit	**5**
	Muthanna		Salahuddin	**6**
	The-Qir			**7**
	Maysan			**8**
	Basra			**9**

And below is the membership matrix section for the four clusters

Chart 4- Membership matrix section for the four clusters

Membership Matrix Section : Distance Type-Euclidean Scale Type-Standard Deviation

Row	Cluster	Prob in1	Prob in2	Prob in3	Prob in 4
1	1	0.3136	0.2962	0.2941	0.0961
2	1	0.3184	0.2928	0.2899	0.0989
3	3	0.2935	0.3015	0.3024	0.1026
4	1	0.3177	0.2936	0.2909	0.0977
5	3	0.2632	0.2649	0.2650	0.2069
6	1	0.3177	0.3024	0.3004	0.0795
7	3	0.2819	0.2911	0.2921	0.1348
8	1	0.3127	0.3052	0.3039	0.0782
9	1	0.3093	0.2905	0.2883	0.1119
10	3	0.2858	0.2881	0.2882	0.1379
11	3	0.2972	0.3047	0.3055	0.0926
12	3	0.2739	0.2826	0.2837	0.1598
13	3	0.2807	0.2942	0.2958	0.1293
14	3	0.2856	0.2934	0.2943	0.1267
15	3	0.2483	0.2544	0.2552	0.2421
16	4	0.2443	0.2437	0.2436	0.2684
17	4	0.1709	0.1719	0.1720	0.4852
18	4	0.1561	0.1575	0.1577	0.5286

4.2.3 The Overall Section for the Number of Clusters

Summary Section :

Number Clusters	Average Distance	Average Silhouette	F(U)	Fc(U)	D(U)	Dc(U)
2	5.496042	0.209564	0.5000	0.0001	0.4930	0.9860
3	3.664202	0.069050	0.3335	0.0002	0.6811	1.0216
4	2.767273	-1.000000	0.2792	0.0390	0.6777	0.9036

Since the number of clusters that maximizes the average silhouette $F_c(U)$equals 0.0390 corresponds to the fourth cluster, and minimizing $D_c(U)$equals 0.9036 also corresponds to the fourth cluster, we choose the number of clusters to be four (4 clusters).

From Table (4), we conclude that the provinces with the highest levels of deprivation, gradually decreasing to the least deprived in terms of infrastructure, are as follows:

3	2	1	Cluster number
Less deprived	Average Deprivation	More deprived	Degree of deprivation

Cluster 1 includes the most deprived provinces, which are: Ninawaa, Kirkuk, Anbar, Babil, Wasit, Salahuddin, Najaf, and Al-Qadisiya.
Cluster 3 includes the least deprived provinces, which are: Duhok, Erbil, and Sulaymaniyah.

From Table (5), we conclude that the provinces with the highest levels of deprivation, gradually decreasing to the least deprived in terms of infrastructure, are as follows:

4	3	2	1	Cluster number
Less deprived	Average deprivation	-	More deprived	Degree of deprivation

Cluster 1 includes the most deprived provinces, which are: Ninawa, Kirkuk, Anbar, Babil, Wasit, and Salahuddin.

Cluster 4 includes the least deprived provinces: Duhok, Erbil, and Sulaymaniyah.
3. The number of clusters that can be chosen is four (4).

4.Conclusions and Recommendations
4.1 Conclusions

The fuzzy cluster analysis using the NCSS system (12) with the number of clusters (4) clusters showed that the most deprived governorates are Ninawa - Kirkuk - Anbar - Babylon - Wasit - Salah al-Din and the least deprived governorates are Dohuk - Erbil And Sulaymaniyah

4.2 Recommendations

1. We suggest studying fuzzy clustering analysis in the presence of outliers.
2. We recommend comparing fuzzy clustering analysis with discriminant (or classification) analysis, hierarchical clustering, and non-hierarchical clustering.
3. We suggest prioritizing the most deprived provinces, followed by other provinces according to their ranking in terms of importance.
4. Allocating part of the state's resources to the most deprived governorates to rebuild them. These governorates are Ninawa, Kirkuk, Anbar, Babylon, Wasit, Salah al-Din, Najaf, and Qadisiya, due to the political circumstances that Iraq has gone through.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References

1 .Central Agency for Statistics and Information Technology. (2010). Basic tables and results of inventory and numbering, environmental survey for the year 2010 .

2 .Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4757-0450-1

3 .Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems 2, no. (3), 267–278. https://doi.org/10.3233/IFS-1994-2306

4 .Dunn, J. C. (1973). A fuzzy relative of the ISODATA process. 'Journal of Cybernetics, 3', 32–57.

5 .Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. 'Journal of Cybernetics, 4', 95–104.

6 .Gath, I., & Geva, A. B. (1989). Unsupervised optimal fuzzy clustering. 'IEEE Transactions on Pattern Analysis and Machine Intelligence, 11'(7), 773–780. https://doi.org/10.1109/34.192473

7 .Gustafson, E. E., & Kessel, W. C. (1979). Fuzzy clustering with a fuzzy covariance matrix. In 'Proceedings of the IEEE Conference on Decision and Control' (pp. 761–766). San Diego, CA.

8.Hana Řezanková , (2014) , Cluster Analysis of Economic Data , STATISTIKA , 94(1

9.Höppner, F. (Ed.). (1999). 'Fuzzy cluster analysis: Methods for classification, data analysis, and image recognition' (pp. 49–53). Chichester, New York: J. Wiley.

10.https://www.mathworks.com/help/fuzzy/fuzzy-clustering.html

11.KRUSE, R., DÖRING, C., LESOT, M.-J. (2007) , Fundamentals of Fuzzy Clustering. In OLIVEIRA, J. V., PEDRYCZ, W., eds. Advances in Fuzzy Clustering and Its Applications. .Chichester: John Wiley & Sons, pp. 3–30.

12.Malhotra, V. K., Kaur, H., & Alam, M. A. (2014). An analysis of fuzzy clustering methods. .'International Journal of Computer Applications, 94'(19). New Delhi, India. (ISSN: 0975-8887)

13.NCSS Statistical Software. Fuzzy clustering. In 'Chapter 448'. Retrieved from NCSS.com

14.Rajkuma, K.V. & Others (2019) Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset, International Journal of Electrical and Computer Engineering (IJECE)Vol.9, No.4, pp. 2760~2770ISSN: 2088-8708, . DOI: 10.11591/ijece.v9i4.pp2760-2770

15.Thomas R. & others (1999) Fuzzy cluster analysis: Methods for Classification, Data Analysis and Image Recognition, New York, John Wiley & Sons

16. Yager, R. R., & Filev, D. P. (1994). Generation of fuzzy rules by mountain clustering. 'Journal of Intelligent and Fuzzy Systems 2 no.), 209–219. https://doi.org/10.3233/IFS-1994-2306

17.Zadah, L. A. (1965). Fuzzy sets. 'Information and Control, 8', 338