Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 4, October 2020, pp.1999-2006

The cluster analysis of most important citrus trees in some governorates of Iraq for the year 2019

Hind Waleed Abdulrahman[1] , Lamyaa Mohammed Ali Hameed [2]
[1] Department of Industrial Management, College of Administration and Economics University of Baghdad
[2] Departments of Statistics, College of Administration and Economics University of Baghdad

ABSTRACT

Citrus fruits are one of the consumer agricultural products of the Iraqi citizen. It is rich in vitamins and used in many food industries as well as medicines. Classifying the amount of production of citrus trees according to the producing governorates has been done to find a map that shows the production of citrus trees according to Iraqi governorates. A cluster analysis method was used according to the hierarchical method. The results showed that Najaf and Qadisiyah are the most similar in citrus production, while Saladin and Najaf were the two governorates with the furthest distance in proximity matrix. Diyala governorate was clustered in the first cluster within two, three, four or five of the clusters for classifying Iraqi governorates covered by the research into three groups according to the production of citrus including the first group with the provinces of Najaf, Qadisiyah, Babil and Karbala, while annexation of the second cluster provinces includes Diyala and Wasit. The third cluster has included the provinces of Baghdad and Saladin. The lowest distance has been between the governorates of Najaf and Qadisiyah as compared with the longest distance between the governorates of Saladin and Najaf.

Keywords: Cluster Analysis, Proximity Matrix, Euclidean Distance, Citrus

Corresponding Author:

Hind Waleed Abdulrahman
College of Administration and Economics, Department of Industrial Management
Baghdad University, Iraq
E-mail: hind.w@coadec.uobaghdad.edu.iq

1. Introduction

Citrus fruits are among the important consumer agricultural products of the Iraqi individual. Citrus fruits are spread in different regions of the Iraq's governorates, and the local production may not be sufficient to meet the need. Therefore, the imported product may be resorted, especially from neighboring countries. Fruits grown in Iraq are divided into two parts, dates and fruits. Fruits are divided into three sections, the first one stands for a hard-core fruit such as apricots, peaches, plums, Greengage, and quince. The second one includes apples and Pears, and the third section includes citrus fruits such as mandarin, orange, sweet lemon, sour lemon, pomelo, grapefruit and other types such as dried lime, hybrid lemon, and trifoliate orange. There are five main types of citrus fruits (orange, lemon, sweet lemon, mandarin, bitter orange) that are used as staple foods in the ethnic consumer basket. These types have been addressed as basic variables for research based on the citrus tree production survey for 2019 year. It is a survey within annual plan of the Central Bureau of Statistics. Citrus production is estimated based on the citrus sample designed in 1979. The citrus survey is carried out in the winter season and the fruit survey in the summer season. This is done with the first indicators the number of fruitful trees and the second average productivity of one fruitful tree. These two indicators are provided through the development of preparing Citrus fruit trees under three conditions: trees in the production stage, trees that were not estimated, and trees planted from the previous year. The average tree productivity is determined by the sample data that is obtained by direct inquiry through a visit to the surveyed holders at the time of marketing and is often in the middle of the marketing season. The two reports depend on

the sample commensurate with the size of the community, which is the number of orchards in the district, and the sample chosen using the method of Systematic Random Sample. The sample is updated according to agricultural surveys and censuses. The sample was designed according to the fruit tree census of 1978. It was updated according to the fruit tree survey of 1989 dependent on the growth rates of the preparation fruit and citrus trees according to the age classification based on a scientific committee formed in 2006 and agricultural census data in 2001. The methods of statistically multivariate complex have wide uses in various fields such as natural sciences and social sciences. The basic idea of cluster analysis is based on knowing the similarities between the variables and obtaining a measure of similarity. Then, the variables are classified according to the degree of similarity in this research based on the method of cluster analysis along with a principle of similarity and difference between the citrus producing governorates and finding a measure of the distance between them. An important basic issue of agricultural planning in general and citrus production in particular is the accurate assessment of production, the possibility of predicting it for the coming seasons, and knowing the pattern of spread, density, similarities and differences between the governorates producing them. This study aims to classify the governorates of Iraq according to the amount of citrus trees production using cluster analysis methods. The annual report for 2019 year issued by the agricultural statistics directorate of the central statistical organization for citrus production was adopted [1]. The research included eight Iraqi governorates which are Diyala, Baghdad, Babil, Karbala, Saladin, Wasit, Najaf, and Qadisiyah. As for the search variables, which are the types of citrus studied including oranges, lemons, sweet lemons, mandarin, and bitter orange. There are several articles in this area, the most important of it, starting with Kadhim [2], performed a statistical analysis using the cluster analysis method for the governorates of Iraq using 20 variables for the lifestyle. It was found through clustering that the Basra governorate pattern did not interact with any other governorate, and this reflects the nature and lifestyle in that governorate, where aspects of life are characterized by economic and social diversification more than other governorates. Mustafa [3], classified the Arab Maghreb countries into groups with common characteristics for economic integration using cluster analysis to know the most homogenous countries in spending trends for national income. Namiq [4], applied the method of cluster analysis for classifying expenditures for the Iraqi community, urban and rural, relying on commodity groups to know the consumption behavior of the Iraqi family. The analysis represented the emergence of three main clusters, the first cluster comprising six main variables in the spending. Rashid and Mahdi [5], analyzed the reality of education in some governorates of Iraq using 25 variables from the level of services provided, applying hierarchical and non-hierarchical cluster analysis methods. They concluded that Baghdad Governorate was the best in providing education services. Ahmed [6], completed his research using the analysis of cluster differentiation to solve the problem of distributing food commodities in Cairo. Abdul Latif and Abdullah [7], used cluster analysis to assess livestock problems in Iraq based on data for the year 2001. It is the year in which a comprehensive agricultural and livestock census took place. Iraqi governorates were classified into four groups according to livestock problems. Wiecek [8], used cluster analysis to classify greenhouse gas emissions agriculture in OECD countries using the Wald method. He was able to classify the diameters into six clusters, each grouping comprising a number of countries. Feng et al. [9], used a hierarchical cluster analysis to classify different citrus fruits as a polynomial method adopted with spectroscopy in addition to analyzing the main compounds in determining peaks and identifying the different types of eight types of cultivated and imported citrus fruits in China. Ahmed [10], categorized the Syrian governorates according to the variables of consumption expenditures for the family using cluster analysis. He concluded that there are three different clusters that Damascus Governorate was among the first cluster including the governorates with high spending. Kadhim and Muhammad [11], categorized Iraqi governorates that suffer from deprivation by relying on data for the year 2009 using cluster analysis, relying on eight different variables of deprivation and was able to divide the provinces into three different clusters according to the variables of the deprivation guide. Arora and Jain [12], presented a research using cluster analysis on a soybean data set that contains 35 types of soybean diseases. The data has been classified into four clusters, each cluster has a group of pests and plant diseases. Shahid [13], accomplished a research through which he categorized the variables of the economic sectors in Algeria for the period 1969-2013 relying on nineteen variables. He managed to divide the economic sectors into three clusters in which the hydrocarbon sector was distinguished from the rest of the sectors, which confirms the dependence of the Algerian economy on this sector by a large percentage. Ibrahim [14], classified the governorates of Iraq using cluster analysis using the nearest neighbor method, the individual link and the K-means method according to the health indicators for the year 2010. He concluded that Baghdad governorate is the best in providing services. Abu Assaf et al. [15], used the cluster analysis to compare the indices of consumer prices in Syria according to the governorates,

months, and commodity totals for the year 2014 and compare them to 2010 year. It was concluded that ten governorates out of 14 governorates in one cluster, which indicates the homogeneity of numbers between the provinces of this cluster. Hanaish [16], used cluster analysis to classify drinking sources in Egypt by studying the common characteristics of each group and determining the basic characteristics of it. The results showed that the classification of mineral elements in drinking water sources has three groups (clusters) for both winter and summer seasons. Li et al. [17], used cluster analysis to analyze agricultural production data in ASEAN countries to obtain an image that reflects the agricultural production of these countries. The results indicated that agricultural development is uneven among these countries. Etumnu and Gray [18], used cluster analysis to classify farmers based on success strategies for management priorities. The results indicated that two groups out of five groups considered costs to be more important than other departments. Ashour [19], ranked the Iraqi governorates according to the level of health indicators for the 2017 year for the citizens. The study included (13) governorates and (25) variables of health indicators. The cluster analysis was used in a hierarchical and non-hierarchical way, and the results indicated that Baghdad governorate is the best in providing services to citizens.

2. Cluster analysis

One of the classification methods used is the bulk grouping of elements in monolithic groups and different from other groups. There are two types of cluster analysis [20]:

2.1. Hierarchical cluster analysis

What distinguishes this analysis is that it does not require prior knowledge of the number of clusters on which cases will be classified, and it is one of the preferred methods in cluster analysis. The vocabulary is grouped in clusters depending on the amount of similarity between the vocabulary under study. The vocabulary cluster is progressively upward from weakest to strongest with m number of different clusters of hierarchy clustering depending on the similarity matrix or kinship to measure similarity between the elements and data values in the variables must be converted to standard values if measured in different units. The first two types include division according to the cases and the second division according to the variables. In this division, the absolute values of Pearson correlation coefficient are relied upon:

- A proximities matrix are a symmetric matrix that expresses the convergence distances or spacing between each pair of vocabulary and there are different measures for it, that the common measure is the Euclidean distance or the square of the traditional distance
- The Euclidean distance square between two vectors is calculated as follows:

$$d(x,y) = \sqrt{(x-y)'(x-y)} \quad (1)$$

2.1.1. Clustering methods

- **Single linkage clustering method**

It is called the "Neighbor Nearest" method, and it depends on considering the two elements that are most similar between the elements in the shape of the cluster core, then the rest of the units are added in series to the nucleus and according to the degree of similarity with the elements of the nucleus.

- **Complete linkage clustering method**

It is a first reverse method that depends on the less similar elements, as the candidate to enter the cluster is the distance between it and any of the elements of the cluster is the largest distance.

- **Ward method**

This method is based on the lowest information loss for the cluster function depending on the sum of the squares of errors for the k cluster, since the sum of the squares of errors at the beginning of clustering is equal to zero and after linking the two clusters will increase.

2.2. Non - hierarchical clustering method

2.2.1. K- means method

It has the following steps:

A. Determining the number of clusters.
B. Initially determining the average values of centroids.
C. Calculating the distances between each data pair and the centers of the averages.

D. The average cluster is considered to be the closest meeting point to the cluster data.
E. Recalculating the averages of the cluster centers.
F. Repeating the steps until a situation is reached in which all points are as close as possible to the average of the cluster centers.

3. Results and discussion

The application of the cluster analysis method to the governorates (variables), from the outputs of the SPSS program was obtained as in Table 1 .The traditional distance between the variables was used and then the distribution of the variables in groups using the hierarchical method.

Table 1. Proximity Matrix between governorates producing citrus trees

Case	Euclidean Distance							
	Diyala	Baghdad	Babylon	Karbala	Saladin	Wasit	Najaf	Qadisiyah
Diyala	.000	12420.877	23856.798	23089.332	25673.939	11880.579	25776.112	25760.707
Baghdad	12420.877	.000	36130.455	35333.455	14178.778	23539.414	38052.794	38038.968
Babylon	23856.798	36130.455	.000	902.545	49429.405	13163.384	1929.093	1917.053
Karbala	23089.332	35333.455	902.545	.000	48684.937	12305.144	2753.158	2744.869
Saladin	25673.939	14178.778	49429.405	48684.937	.000	37331.812	51340.285	51322.973
Wasit	11880.579	23539.414	13163.384	12305.144	37331.812	.000	15048.545	15042.624
Najaf	25776.112	38052.794	1929.093	2753.158	51340.285	15048.545	.000	39.560
Qadisiyah	25760.707	38038.968	1917.053	2744.869	51322.973	15042.624	39.560	.000

From this Proximity Matrix, we note that the closest distance between the governorates was between Najaf and Qadisiyah according to the citrus tree production variable and Euclidean distance scale as it reached to 39.560. It is also noticed that the farthest distance between the governorates was between Saladin and Najaf as it reached to 51340.285.

3.1. Agglomeration schedule

We note from Table 2 that the Qadisiyah and Najaf governorates have a link between them because the traditional distance between them has reached the smallest possible amount which is 39.560. The third step is in which Najaf governorate resulting from the first step was linked with the Babel 3. Then, it moves to the sixth stage and the province of Babel 3 resulting from the third step is linked with the item of Diyala 1. Then, the moving is made to the seventh step in which the Diyala 1 resulting from the sixth step is linked with the item of Baghdad 2, and so on. The rest of the groups it is noticeable that the adjacent geographical location has an effect on the formation of the cluster, as we find the Qadisiyah governorate clustering with Najaf, and Karbala governorate with Babil governorate. This reflects the similarity of soil, climate, and agricultural conditions that affected citrus production.

Table 2. Agglomeration schedule for citrus trees

Stage	Cluster combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	Najaf 7	Qadisiyah 8	39.560	0	0	3
2	Babylon 3	Karbala 4	902.545	0	0	3
3	Babylon 3	Najaf 7	2336.043	Baghdad 2	Diyala 1	6
4	Diyala 1	Wasit 6	11880.579	0	0	6
5	Baghdad 2	Saladin 5	14178.778	0	0	7
6	Diyala 1	Babylon3	19255.331	Karbala 4	Babylon 3	7
7	Diyala 1	Baghdad 2	37274.943	Wasit 6	Saladin 5	0

3.2. Icicle plots

The following diagram represents the bonding process using Icicle plots; each governorate is represented by a rectangle:

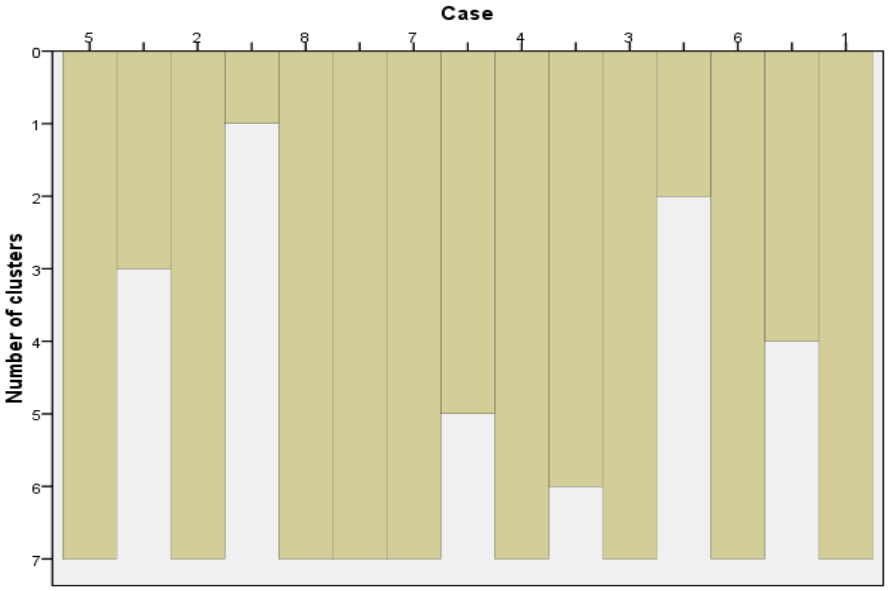

Figure 1. The Icicle plots of citrus trees according to the producing governorates

From Figure 1, we find that both of Najaf 7 and Qadisiyah 8 in a colored rectangle in first step. Then, Babylon 3 and Karbala 4 are joined . It is possible to join according to the diagram to get a better description using Average Linkage (Between Groups) Rescaled Distance Cluster Combine. Through the scheme of the cluster tree for Iraqi governorates according to citrus production, the tree includes measurements that extend to 25 units of measure, as the length of the line indicates an increase in the difference between the provinces. The nodes in the tree represent the merging of two or four provinces. We find at the top of the tree a process of cluster between two provinces It is Najaf and Qadisiyah and we found another cluster that includes Babylon, and Karbala and through the drawing. We found that the degree of similarity in these two clusters is as high as possible as the span extending between them does not exceed 3 units of measure. It is a very strong degree of similarity which indicates that the production of citrus in these provinces is very close. The third cluster includes the provinces of Diyala and Wasit, we find that the measurement between them is greater than the first two clusters, as the distance reaches approximately with 8 units of measure. It means that the degree of similarity between the production of citrus fruits is less than the degree of similarity of the first and second cluster, while the fourth cluster reaches the unit of measurement to about 13. It is the highest degree, which confirms that the similarity here is the least possible and includes the provinces of Baghdad and Saladin. Items and other details were distributed as members of groups as in Tables 3-6.

Table 3. Cluster membership

Case	5 Clusters	4 Clusters	3 Clusters	2 Clusters
1 Diyala	1	1	1	1
2 Baghdad	2	2	2	2
3 Babylon	3	3	3	1
4 Karbala	3	3	3	1
5 Saladin	4	4	2	2
6 Wasit	5	1	1	1
7 Najaf	3	3	3	1
8 Qadisiyah	3	3	3	1

Table 4. Initial cluster centers

	Cluster	
	1	2
Orange	134.00	51302.00
sour Lemon	51.00	1294.00
sweet Lemon	.00	691.00
Mandarin	.00	920.00
bitter orange	59.00	3905.00

Table 5. Iteration history

Iteration	Change in Cluster Centers	
	1	2
1	3283.166	9893.494
2	547.194	2473.373
3	91.199	618.343
4	15.200	154.586
5	2.533	38.646
6	.422	9.662
7	.070	2.415
8	.012	.604
9	.002	.151
10	.000	.038

Table 6. Final Cluster Centers

	Cluster	
	1	2
orange	3811.20	38148.00
sour lemon	394.80	1068.00
sweet lemon	45.80	402.33
mandarin	135.40	1163.33
bitter orange	1423.40	4793.67

We note that the iterations have stopped because the maximum number of iterations has been achieved and the convergence has not occurred. The maximum change of absolute coordinates for any center is .038 with current iteration of 10. The minimum distance between the initial centers is 51340.285.

4. Conclusions

The Iraqi governorates under study were classified into three groups according to the different types of citrus. The first group has been considered for Najaf, Qadisiyah, Babel and Karbala governorates, while the second group has been considered for Diyala and Wasit governorates. Lastly, the third group has been considered for Baghdad and Saladin governorates. The consequences have shown that Najaf and Qadisiyah are the most comparable in term of citrus production. The smallest distance was between the governorates of Najaf and Qadisiyah. As for the longest distance, it was between the governorates of Saladin and Najaf.

References

[1] Ministry of Planning, Central Statistical Organization, Agricultural Statistics Directorate, "Citrus trees production for the year 2019, 2019.

[2] F. M. Kazim, "Statistical analysis of the Millennium Development Goals using the method of factor and cluster analysis". Higher Diploma Research in Applied Statistics - Higher Institute for Training and Statistical Research, 2006.

[3] N. Mustafa ,"Using some methods of cluster analysis in classification with practical application", journal of Techniques, Vol. 20, No. 2, Administrative Technical College , 2007.

[4] F.N. Namiq, "cluster analysis method for classifying expenditures on basic goods and services according to the environmental level (urban and rural) for the years 1971-2007" Journal of Baghdad College for Economic Sciences, University, No. 25, 2010.

[5] A. A. Rashid, and N. N. Mahdi, "Analysis of the reality of education in Iraq using cluster analysis methods (a comparative study)" Al-Qadisiyah Journal of Administrative Sciences - Volume 13 - No. 2, 2011.

[6] M. A. Ahmad, "Using goal programming to choose variables in analyzing cluster discriminant by applying it to the problem of distributing subsidized food commodities" PhD in Statistics - Faculty of Economics and Political Science - Cairo University, 2012.

[7] H.F. Abdul Latif , and M. M. Abdullah, "Using cluster analysis to assess some of the livestock problems in Iraq" the scientific journal of Cihan University - Sulaimanyia Vol. 1 Issue (3),2013.

[8] A.K. Wiecek,"The use of cluster analysis in the classification of similarities in variables associated with agricultural greenhouse gases emissions in OECD countries" WIES I ROLNICTWO, Vol. 1, No.158, 2013.

[9] X. Feng, et al. "Rapid Classification of Citrus Fruits Based on Raman Spectroscopy and Pattern" Food Sci. Technol. Res., Vol.19, No. 6, pp.1077 – 1084, 2013.

[10] T. Ahmad "Classifying the Syrian governorates according to the consumption expenditures of the family using cluster analysis" Tishreen University Journal for Research and Scientific Studies - Series of Economic and Legal Sciences, Vol. 37, No. 2, 2015.

[11] I. J. Kazim, and A. S. Muhammad, "classification and evaluation of evidence of deprivation in Iraq 2009 by using cluster analysis ", Journal of Economic and Administrative Sciences, Vol. 21, No. 82, pp. 391-411, 2015.

[12] A. Arora, and R. Jain, "Clustering-case studies in agriculture", Agrotech Publishing Academy, 2016.

[13] E.Shahid, "Analytical Statistical Studies of the Contribution of Economic Sectors to Algerian Macroeconomic Policy during the period (1969-2013), Master Degree in econometrics - Faculty of Economic Sciences, Commercial Sciences and Management Sciences - Department of Economic Sciences, 2016.

[14] O. S. Ibrahim "Using health indicators for 2010 to classify the governorates of Iraq using cluster analysis" Tikrit Journal of Pure Sciences, Vol. 21, No.4, pp.149-158, 2016.

[15] S. M. Abu Assaf, et al., "An analytical study of consumer price indices in Syria according to the methodology of cluster analysis" Syrian Journal of Agricultural Research, Vol. 4, No.2, pp. 31-51, 2017.

[16] I. S. Hanash "Using some methods of cluster analysis in classification with practical application on some drinking sources in the city of Egypt" 2017 https://www.researchgate.net/publication/316279402.

[17] L. LI, et al." Cluster Analysis of Agricultural Production in ASEAN Free Trade Area" , 3rd International Conference on Social Science and Management (ICSSM) ISBN: 978-1- 60595-445-5, 2017.

[18] C. Etumna, and, A. Gray, "A clustering approach to understanding farmers success strategies" Selected paper prepared for presentation at the 2018 & Applied Economics Association Annual Meeting ,Washington, D.C. 5-7 August. 2018.
[19] W. A. Ashour, "Classifying the Iraqi governorates healthily using cluster analysis for year 2016 ", Journal of Natural, Life and Applied Sciences, Vol.3, No. 30, pp. 121-135 , 2019.
[20] H. Christian, et al., "Handbook of Cluster Analysis" Chapman &Hall/CRC Handbook of Modern Statistical Methods, CRC Press, 2015.