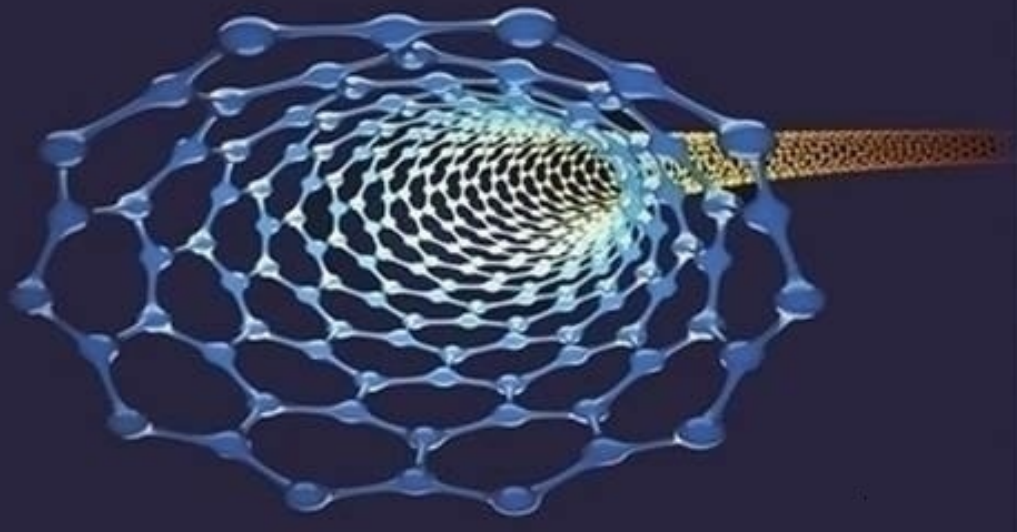
Third International Conference of Mathematics and its Applications

(TICMA 2022)



زانکۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

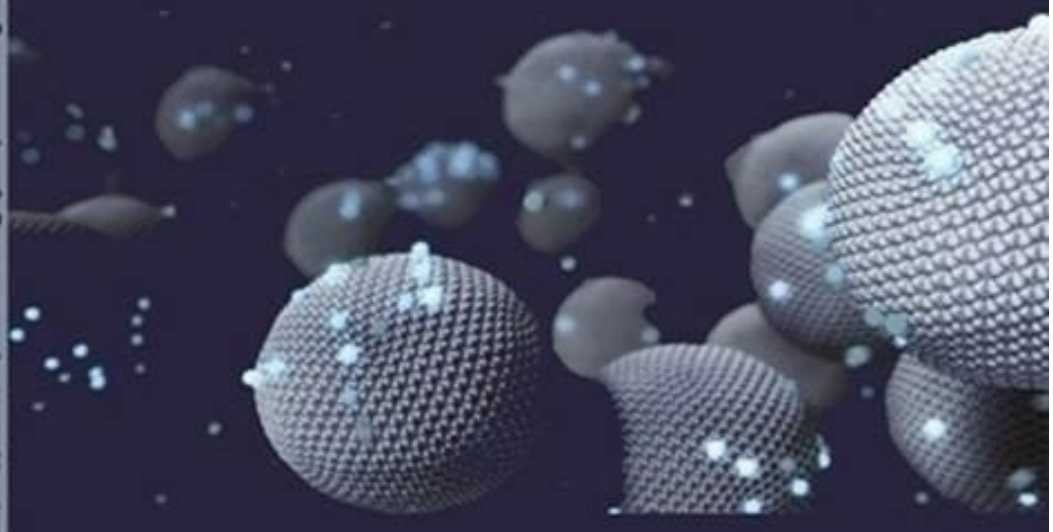
College of science / Mathematics Department

29-31 August 2022

ZANCO Journal of Pure and Applied Sciences

ZANCO

Journal of Pure and Applied Sciences



زانکۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

College of science / Mathematics Department

29-31 August 2022

Proceeding of 3rd International Conference of Mathematics and its Applications TICMA2022
https://conferences.su.edu.krd/su/ticma2022/
DOI: https://doi.org/10.31972/ticma2022

RESEARCH PAPER

Comparison Between Factor Analysis and Cluster Analysis to Determine the Most Important Affecting Factors for Students' Admission and Their Interests in The Specializations: A Sample of Salahaddin University-Erbil

Mohammed O ABDULLAH[1,a], Rizgar M AHMED[2,b,*], Yener ALTUN[3,c]
[1] *STATISTICS DEPARTMENT, INSTITUTE OF NATURAL AND APPLIED SCIENCES, YUZUNCU YIL UNIVERSITY, Van,Turkey.*
[2] *STATISTICS DEPARTMENT, ADMINISTRATOR AND ECONOMICS,, SALAHADDIN UNIVERSITY- ERBIL, Iraq.*
[3] *STATISTICS DEPARTMENT, INSTITUTE OF NATURAL AND APPLIED SCIENCES, YUZUNCU YIL UNIVERSITY, Van, Turkey.*

[a] *muhamad91m91@gmail.com*
[b] *rizgar.ahmed@su.edu.krd*
[c] *yeneraltun@yyu.edu.tr*

A B S T R A C T:

The main goal of this thesis is to determine the most important effective factors for student admission and his/her interests in the specialization by using multivariate methods. Therefore, it focused on using factor analysis by identifying a number of the obtained factors and cluster analysis by classifying them into five clusters. Furthermore, the factor analysis and cluster analysis results will be compared to each other. Moreover, this study depends on the analysis of 350 questionnaire forms, distributed by random stratified sample method on students in the first stage of three different colleges, including Scientific colleges and Humanity colleges of Salahaddin University in Northern Iraq for the academic year 2018-2019.

Thus, the IBM SPSS Statistics V: 25 software programs have been used in data analysis. Additionally, the results have demonstrated that Reliability is accepted, and also in factor analysis, the rate of the total variance interpretation is %62.157. Moreover, the most common variables between the factor analysis and cluster analysis can be considered the most important and influential variables for student admission and their interests in choosing a specialization. Consequently, the first factor and the first cluster have five significant variables in common; they are V1, V2, V3, V4 and V5 (the system is helpful for student admission to colleges to get their desired professions). The second factor and the second cluster have four influential variables in common they are V24, V32, V35 and V37 (the new system may help master's and PhD students to be admitted to colleges and get competitive results by utilizing their accounts). In the fourth factor and the fourth cluster, there is one variable in common, which is V18 (decreasing the number of students admitted in the parallel system by using the graduated students who must not be able to refill admission forms). Ultimately, the conclusion has shown a kind of approach and similarity between factor analysis and cluster analysis.

KEY WORDS: Factor Analysis, Cluster Analysis, Cluster Analysis.
DOI: https://doi.org/10.31972/ticma22.03

1. INTRODUCTION:

One of the important topics in the student's life is Admission to universities or institutes according to their interests. In Iraq and northern Iraq, the students finish the study period for (12) years until they reach their future in university or institute. Previously the admission system at the university was based on filling out various admission forms in (the central Admission) department, which links the ministry of education with higher education for the admission of the students in universities and institutes.

*** Corresponding Author: Rizgar M AHMED[2]**
E-mail: *rizgar.ahmed@su.edu.krd*
Article History:
Received: 01/08/2022
Accepted: 15/09/2022
Published: 07/12/2022

DOI: https://doi.org/10.31972/ticma2022

Nowadays, the Ministry of Higher Education in northern Iraq depends on the new system, which depends on filling the forms of Admission on the internet to extract the results according to the interests with a reasonable proportion. Student admissions are vital to any university's running because universities can stay alive with students. The admission system is an influential activity that helps the universities survive with the student's contribution. Then, lack of knowledge by the students and poor admission system are serious problems that cause fewer students to be admitted to the universities because of the system's slow response, which leads the students to make mistakes. Therefore, students are not able to get to their desired universities. For these reasons, there must be good educated information about the online admission system; it can be achieved only by choosing the best method for Admission [1].

Although the admission system can widely consider the process when a student gets more interested in reaching higher education until Admission in a particular course and university happens [2]. The admission process at the university is an entrance process in a university. It generally influences all the resources of the universities and their quality restraints [3]. The Ministry of Higher Education and Scientific Research in the north of Iraq has introduced a new system for Admission to colleges and institutes in the region for the academic year (2011-2012). Furthermore, this system has a great success since its inception. Also, this system has been used to accept all students in the region from its inception until now.

This research studies the new student admission system for attending colleges and institutes with their interests in the specialization. The new admission system includes four methods (Zankoline, Credit, Parallel, and Evening study). The student introduces many requests to more than one university by filling a form online in a specialized system for student admission. The current admission system gives them acceptance into one university, giving many chances to other students. The proposed system accepts only one request from each student in all universities. On the other hand, it helps the universities have only one place to receive the students' requisitions. The problem is that there are some deficiencies or imperfections in this system, which (include delaying the time to receive results, not taking the interest of students as regarded, and lack of trust of the student in the system…etc.). The vital aim of this topic lies in several aspects: it is related to a student's future in their life, and there are several problems facing the student in this system, such as the chosen department may not be his/her interest or sometimes some students' names might be missed. In this thesis, all the important aspects of this system are studied and processed using an advanced statistical method; it includes two types of multivariate analysis: cluster analysis and factor analysis. In addition, some recommendations are considered an essential effective service to the students to achieve the departments in which they are interested.

2. METHODOLOGY:

Factor analysis (FA) is a strategy used in Statistics that is applied to a solo set of variables by statistical researchers interested in finding which variables are related to others and which one is independent in the set of the variables. Therefore, those variables associated with other variables but not correlated with different subsets of variables are linked into factors. In addition, factors are accustomed to reflecting on the process, which has built the correlation between variables [4]. Moreover, factor analysis is a set of methods for clarifying the correlations between variables in the form of basic entities called factors. Additionally, the model' factor analysis' is broadly comprehended to refera collection of nearly associated forms intended to clarify or create a co-relational structure between the observed random variables [5].

One of the central assumptions of the factor analysis is that it is not normal to observe the factors directly; the variables rely on the factors, also they are reasons for random errors. Thus, the assumption is especially well-organized to the forms like psychology, where it is not suited to measure the concepts which one is interested in directly, for instance, 'intelligence'. Then it is mainly ambiguous to define the concepts [6]. Furthermore, the utilization of factor analysis methods has been enlarged in many scientific research eras such as psychological, educational, athletic, marketing, behavioural, social, medicine, economics, and geography as a result of the technological development of computers. Factor analysis aims to explain the number of basic impacts underlying a domain of variables, count the amount of the variable associated with the factors, and gain enough instruction about their nature to observe the contribution of each factor with its variables.

DOI: https://doi.org/10.31972/ticma2022

Likewise, the specific target of the subject is to describe the patterns of correlations between observed variables briefly; this way, a large number of the observed variables will be reduced to a smaller number of factors. The final aim which to recognize the observed reasons that illuminate the data variation. Moreover, to reach this goal, the relation of the factors and original variables should be checked, then give them an explanation in the framework of how the data were generated. Lastly, the aim generally is to demonstrate a few important common factors [7,8,9].

3. Type of Factor Analysis:

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are two fundamental kinds of factor analysis. Usually, a difference is made between Exploratory and Confirmatory factor analysis. The exploratory analysis aims to demonstrate the factor structure for a collection of variables. It mostly includes the identification how a lot of the factor loadings. However, most EFA programs permit the number of factors to be particularized in advance. Moreover, the variables could not be forced to load just on some specific characteristics. Therefore, EFA is usually identified to be more of a theory-generating than a theory-testing program. Conversely, Confirmatory Factor Analysis is normally built on a powerful theoretical and empirical organization that permits the scientific researcher to identify accurate factor models in advance. Therefore, the model generally identifies which variables will load on which factors. Likewise, whichh factors are associated. Furthermore, Confirmatory Factor Analysis is an academic testing program more than a procedure EFA. Practically, the research may include aspects of both Exploratory and Confirmatory analysis. It is beneficial to differentiate between both techniques in their situational form[10,11,12,13].

4. The Orthogonal Factor Model:

The observable random vectors X with ρ components have mean μ and covariance matrix Σ. Moreover, the factor model postulates, which X is linearly reliant on a small number of unobservable random variables such as F1, F2,…, Fm, which are called common factors, and ρ extra sources of variation such as ε1, ε2, …, εp, which are called errors, or sometimes certain factors. Significantly, the factor analysis model is

$$\mathbf{X_1} = \boldsymbol{\mu_1} + \boldsymbol{\ell_{11}}\,\mathbf{F_1} + \boldsymbol{\ell_{12}}\,\mathbf{F_2} + \cdots + \boldsymbol{\ell_{1m}}\,\mathbf{F_m} + \varepsilon_1 \qquad (1)$$

$$\mathbf{X_2} = \boldsymbol{\mu_2} + \boldsymbol{\ell_{21}}\,\mathbf{F_1} + \boldsymbol{\ell_{22}}\,\mathbf{F_2} + \cdots + \boldsymbol{\ell_{2m}}\,\mathbf{F_m} + \varepsilon_2 \qquad (2)$$

$$\mathbf{X_p} = \boldsymbol{\mu_p} + \boldsymbol{\ell_{p1}}\,\mathbf{F_1} + \boldsymbol{\ell_{p2}}\,\mathbf{F_2} + \cdots + \boldsymbol{\ell_{pm}}\,\mathbf{F_m} + \varepsilon_p \qquad (3)$$

The form for the P variables can be joined in the solo matrix expression; the public factor form can be penned as

$$\mathbf{X_{(p\times 1)}} = \boldsymbol{\mu_{(p\times 1)}} + \mathbf{L_{(p\times m)}}\,\mathbf{F_{(m\times 1)}} + \varepsilon_{(p\times 1)} \qquad (4)$$

X: vector of observable random variables. μ: mean of variables. L: matrix of factor loadings. F: common factors. ε: errors or specific factors from time to time.
Or in matrix data.

$$\mathbf{X} = \mathbf{L}\underline{\mathbf{F}} + \underline{\varepsilon} \qquad (5)$$

[14][15].

5. Cluster Analysis:

Cluster analysis is a multivariate statistical method, a technique of categorization utilizing an evaluation of similarity or distance given for a random couple of objects. Thus, objects divided into groups are similar

DOI: https://doi.org/10.31972/ticma2022

among pair objects inside a group, which ought to be a big number, and similarity among two groups ought to be small. By the way, the distance measured inside a group should be small, and the distance among groups should be a big number [16]. The goal of cluster analysis is to find a structure in an arranged data collection-clusters. Furthermore, objects inward a cluster ought to be similar to one another and be dissimilar from the objects of other clusters [17]. The target of cluster analysis is to the division an arranged data set or objects into clusters or its term such as (subsets, groups, and classes). The mentioned division should have the following properties: A- Homogeneity inside the clusters. Data belonging to the same cluster ought to be similarly possible. B- Heterogeneity among clusters means data belonging to dissimilar clusters that ought to be differently possible [18]. Therefore, put into a system of cluster analysis, which is a collection of objects for categorization and similarity (or distance) among a couple of objects; output from cluster analysis means several groups that shape a partition, or a family of partitions, of the collection objects[16].

The cluster analysis methods were applied broadly to the collection in several eras, like medicine, psychiatry, sociology, criminology, anthropology, archaeology, geology, geography, remote sensing, market research, economics, and engineering.[19]. Depending on the clustering objective, cluster analysis could be defined in several ways. Usually, one of its definitions is one may agree that a cluster is a collection of objects, which are more similar to one another than to participants of other clusters. Subsequently, it also could be elucidated as set of homogenous observation [20][21]. Methods of cluster analysis can be divided into two categories of hierarchical clustering and nonhierarchical clustering.

6. Hierarchical Clustering:

Hierarchical techniques are between the long-established methods of cluster analysis. Moreover, hierarchical clustering consists of successive aggregation or partition of the observations and their separations. Concluding from this sort of procedure, there is a form of a tree structure, considered a dendrogram [22]. Techniques of hierarchical cluster analysis are divided to a couple of classes; (agglomerative techniques - a succeeding pooling of divisions of the collection of objects and divisive techniques- succeeding partitions of the collection of objects). Hierarchical cluster analysis or agglomerative hierarchical cluster analysis is a technique to generate a family of categorization of a limited collection of objects built on a measure of resemblance identified on a couple of objects. Therefore, the method applies to different fields in natural and social sciences [16,23]. The agglomerative methods begin with the collection of observations, which is a divided cluster. In addition, groups are combined in accordance with the lessening quantity of similarity (or the rising amount of dissimilarity) until one cluster is recognized [22]. Hierarchical classifications shaped by cluster methods, such as the agglomerative or divisive techniques, may be characterized by a two-dimensional diagram, which is recognized as a dendrogram that elucidates the fusions or separations created at each phase of the analysis.[16]. A dendrogram can recognize the resemblance of variables and groups of variables.

7. Dendrogram:

A dual-dimensional diagram mainly demonstrates the hierarchical structure, so it is named a tree diagram or dendrogram. Therefore, the dendrogram is a graphical demonstration of the results of hierarchical cluster analysis. Besides, it emerges in the shape of a tree as plot, where each pace of hierarchical clustering is demonstrated as a fusion of a couple of branches of the tree solution into a single one. By the way, the branches demonstrate clusters obtained at each pace of hierarchical clustering. [24]. A dendrogram is a tree diagram in which the (X) axis demonstrates the objects while the lower *Y* axis displays distances. Besides, the tree branches show the order of the (n-1) links; the fork explains the primer link. In addition, the second fork shows the second link continually until all link together at the trunk. Thus, the dendrogram could be utilized to build a new distance matrix among the objects [6]. Additionally, the tree is mainly shown upside down. By the way, the tree's root is located at the bottom, and the branches are at the top. On the other hand, when a computer creates trees, it is most suitable to print them out, and the tree is on its side with the branches on the left [25,23].

DOI: https://doi.org/10.31972/ticma2022

8. Non-Hierarchical Clustering:

Non-hierarchical clustering has a point to begin with: the specification of the number of clusters. Sometimes it identifies the number, and the objects are assigned to groups. So, it is a couple-phase process. Firstly, the cluster seed is particularized. Thus, it is a beginning point, which can be illustrated by the researcher as either a systematic or random choice. Furthermore, observations are assigned according to their similarities to the pre-defined seed [26]. In addition, nonhierarchical clustering methods are shaped to group items rather than variables into a set of K clusters, so the number of clusters, which is named *K*, may either be particularized in a developed way or verified as part of the clustering procedure because the dissimilarities of a matrix do not have to be determined and the 16 fundamental data does not be stored throughout the computer run. Nonhierarchical techniques begin from either (1) the first division of items into groups or (2) a primary collection of seed points that will shape the nuclei of clusters. God's selections for beginning configurations ought to be liberated of apparent biases. One of the paths to begin is to choose seed points randomly from among the items in the first group [14]. Besides, the nonhierarchical group of algorithms is the K-means. The K-means works by separating the data into a pre-specified number and systematically assigning observations to the clusters. Nonhierarchical techniques can be applied to a large number of data collections. The K-means technique is apt for big samples (n > 1000) since it does not figure the closeness matrix among all cases [26].

9. RESULTS AND DISCUSSIONS:

To better understand the action of students' Admission and those problems facing students in the new system in northern Iraq, with excellent and scannable solutions for the preferred system, by using statistical data that had been analyzed through cluster analysis and factor analysis. Therefore, a statistical Mechanism will be applied to get the data. A set of questions is prepared in questionnaire form; the questions will be prepared according to the admission system strategy, which focuses on age, gender, specialization, interest, average score, and 39 variables related to the student admission system and their interests in the fields. Then, this questionnaire form is distributed among 350 students of the first stage at Salahaddin University-Erbil for the academic year 2018-2019 in 3 scientific colleges and humanity colleges at ten different departments. Moreover, Stratified Random Sampling has been used. Thus, they are 251 students in Administrator and Economics College, 74 in Science College, and 25 in Education College. Profoundly, the data is collected manually and entered into (statistical package for the social sciences) IBM SPSS statistics V: 25 programs for analyzing them.

10. Reliability Test:

Reliability is one of the most important and fundamental countenances in evaluating any measurement instrument or tool for good research, for an exploratory or pilot study[27]. George and Mallery had provided the rules of thumb e. i. if the value of alpha is >0.9 = Excellent, >0.8 = Good, >0.7 = Acceptable, >0.6 = Questionable, >0.5 =Poor, and <0.5 = Unacceptable[28]. Accordingly, in this study, the reliability test is conducted by means of Chronbach's alpha using SPSS software before applying (factor and cluster) analysis; it has summarized the results i' the following table.

TABLE 1. Reliability Statistics

Cronbach's Alpha	N of Items
.70	39

The reliability analysis shows that Cronbach's Alpha is 0.70 for 39 items. Therefore, the data is good to analyze.

11. Factor Analysis:

Factor analysis has been used to construct the new affective factors for student admission and their interests in the specialization. Moreover, factor analysis aims to reduce the redundancy among the variables by using

https://conferences.su.edu.krd/su/ticma2022/
DOI: https://doi.org/10.31972/ticma2022

a smaller number of factors. Principal component methods are used to analyze the correlation matrix to show the significance of each variable based on the relationship between the variables. A correlation matrix should be used in the process of factor analysis; it displays the correlation or relationships between a single variable and every other variable in the investigation.

The first step of the factor analysis is to measure the adequacy of the data, Kaiser-Meyer-Olkin (KMO), which measures the appropriateness of the data for the factor analysis; the value of (KMO) is greater than 0.5, the more appropriate data for factor analysis. Additionally, the (KMO) is the partial correlation between the questions to ensure a strong association between all or most of the questions, not only among a few. Bartlett's test of Sphericity will be used to test the strength of these correlations. The null hypothesis of this test is that there are no correlations between the questions. Therefore, the factor analysis requires the rejection of this hypothesis to make the data suitable for this analysis.

KMO and Bartlett's Test for our data summarized the results in the following table:

TABLE 2. KMO and bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.713
Bartlett's Test of Sphericity	Approx. Chi-Square	1985.241
	Df	741
	Sig.	.000

The Kaiser-Meyer-Olkin (KMO) measure should be greater than 0.50 and is inadequate if less than 0.50. The KMO test tells us whether or not each factor predicts enough questions. Therefore, for these data, the value is 0.713, which is acceptable and reasonable. Furthermore, there is a test called the Bartlett test of Sphericity that ought to be significant (i.e., a significance value of less than .05); it means that the questions are correlated adequately to give a logical basis for factor analysis. In addition, the data is suitable for factor analysis in this case (the p-value is less than .05, demonstrating that the correlation matrix is significantly unlike an identity matrix).

TABLE 3. Communalities

Variables	**Extraction**	**Variables**	**Extraction**
V1	.604	V21	.568
V2	.658	V22	.609
V3	.651	V23	.673
V4	.664	V24	.611
V5	.553	V25	.722
V6	.492	V26	.575
V7	.647	V27	.591
V8	.624	V28	.634
V9	.757	V29	.643
V10	.675	V30	.710
V11	.714	V31	.552
V12	.624	V32	.609
V13	.722	V33	.581
V14	.643	V34	.441
V15	.648	V35	.501
V16	.582	V36	.656
V17	.538	V37	.519
V18	.638	V38	.609
V19	.647	V39	.666
V20	.688		

The communalities table shows the initial commonalities before rotation. Table 3. Is a table of communalities, which shows how much of the Variance in the variables has been accounted for by the extracted factors.

Note: that the initial communalities are higher than .50, which is good.

DOI: https://doi.org/10.31972/ticma2022

The Total Variance Explained table shows how the Variance is divided among the 39 possible factors. It is noted that the tables have shown the Eigenvalue of the fifteen and sixteen Components, which is less than one. The first sixteen Components depend on the interpretation variance ratio. The extraction window is not based on Eigenvalue; it is based on a fixed number of the factors divided into sixteen factors, in terms of the variable's worth of each one of the variance explanations. So, the first Components show almost. Note: that the firs' component explains 7.818% of the Variance after rotation (as much Variance as in five variables). The seco'd component explains 6.683% of the Variance, 3.849% of the Variance is explained by the third component, 3.828 % of the Variance is defined by the fourth component, the fifth component explains 3.789% of the Variance,

Thus, when the sixteen components depend on the interpretation variance ratio, the component explains less information than a single variable would have defined. The following figure illustrates this.

TABLE 4. Total Variance

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	3.742	9.595	9.595	3.049	7.818	7.818
2	3.418	8.764	18.359	2.606	6.683	14.501
3	1.578	4.045	22.405	1.501	3.849	18.351
4	1.524	3.908	26.313	1.493	3.828	22.178
5	1.430	3.668	29.981	1.478	3.789	25.967
6	1.367	3.506	33.487	1.446	3.707	29.674
7	1.327	3.403	36.890	1.432	3.672	33.346
8	1.296	3.324	40.214	1.388	3.559	36.906
9	1.228	3.149	43.363	1.328	3.405	40.310
10	1.153	2.955	46.318	1.286	3.299	43.609
11	1.098	2.815	49.134	1.281	3.285	46.894
12	1.088	2.789	51.922	1.233	3.161	50.055
13	1.045	2.680	54.603	1.220	3.127	53.182
14	1.017	2.607	57.210	1.217	3.120	56.302
15	.983	2.521	59.731	1.157	2.967	59.269
16	.946	2.426	62.157	1.126	2.888	62.157

The Scree plot demonstrates that after the first sixteen components are different with the Eigenvalues decline (the curve flattens) and they are less than (0.946) or dependent on the interpretation variance ratio decline (the curve flattens). This again supports a sixteen-component solution. Moreover, it notes that both the Scree plot and the sixteen values support the conclusion that these (39) variables could be reduced to sixteen components. It states that the Scree plot flattens out after the sixteen components.Each component accounts for the cumulative Percent of Variance between variables. Before and after rotation, 62.157% of the Variance is accounted for by the first sixteen components.

So, those components that are rotated are easier to interpret. Thus, the rotation makes it as much as possible; the different variables are explained or predicted by different underlying components. So, it is the aim of rotation. It is not usually gained. In the Rotated Component of the factor loadings matrix, there is one thing to look for: the extent to which simple structure is achieved. It summarizes the results in the following table:

DOI: https://doi.org/10.31972/ticma2022

TABLE 5. Rotated Component Matrix[a]

	Components				
	1	2	3	4	5
V3	.771				
V2	.728				
V1	.707				
V4	.667				
V5	.655				
V24		.667			
V35		.608			
V37		.527			
V32		.504			
V22			.713		
V19			-.616		
V18				.716	
V16				.624	
V7					.724
V10					.620

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a]

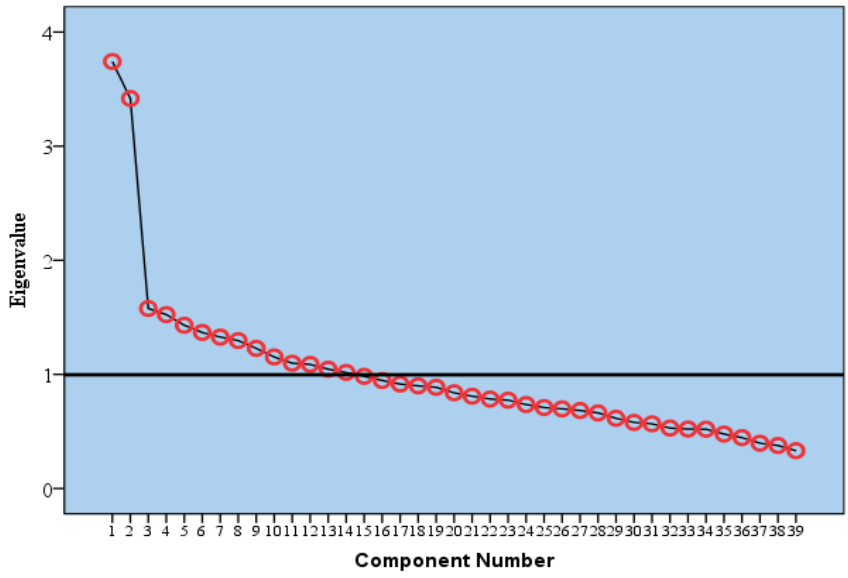


FIGURE 1. Scree plot.

12. Rotation converged in 21 iterations.

The variables are sorted in the table (5.) because the variable has the highest loading (not considering whether the correlation is positive or negative). Every variable has some loading from every component, but we requested loadings less than |.50| to be excluded from the output, so there are blanks where low loadings exist. (|.50| means the absolute value or value without considering the sign).

The first five components are the most fundamental; they have been chosen. Thus, The first component explains the largest variance ratio containing five variables they are (V3: It helps the students be accepted in college and the specialization they like.V2: The students can trust this system. V1: I prefer the system. V4: It has a good effect on the student's future. V5: It facilitates for the students to choose the department that he/she wants.), with loadings (.771, .728, .707, .667, .655). The second component contains four variables respectively (V24: I prefer the Permission of students in master and PhD degrees to be done by using the new system.V35: I suggest daily the students be aware of the results when the competition is done in the system by using their account. V37: The changes and guides must be given to all high schools before starting a new academic year.V32: I like all students to be accepted at the same time.), with loadings (.667, .608, .527, .504).

DOI: https://doi.org/10.31972/ticma2022

The third component contains two variables respectively (V22: My low marks didn't let me study in the department and specialization I desired. V19: The department and specialization that I am studying in is %100 my interest), with loadings (.713, -.616). The fourth component contains two variables respectively (V18: In the parallel system, the graduated students must not be able to refill the admission forms. V16: I like the number of students in the parallel system to be decreased), with loadings (.716, .624). The fifth component contains two variables respectively (V7: It doesn't account for the economic status of the student. V10: Employing the geographic area dependent method affects choosing colleges and specializations), with loadings (.724, .620). Every variable has a loading or a weight within every factor, but in a 'clean' factor analysis, mostly the loadings are not selected. Therefore, we have drawn that the Rotated Factor Matrix will be low (blank or less than .50).

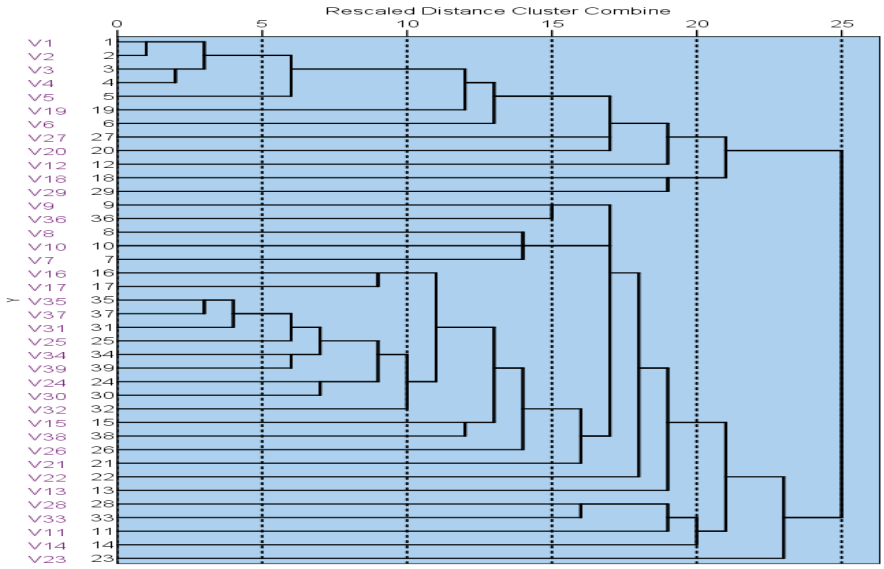
13. Cluster Analysis:

Cluster analysis has been applied in this study. It is one of the most important statistical methods used to classify variables into homogeneous groups. Further, it depends on the differences and similarities between the data. Using average linkage (between groups) and rescaled distance cluster combines to find the distance matrix and relationship between clusters, dividing the number of sets into five discrete clusters using the fragmentation style. The degree of homogeneity is strong or weak between different groups, as the results are shown below:

TABLE 6. Cluster analysis for our data

5 clusters	Membership	Variables
1	10	V1,V2,V3,V4,V5,V6,V12,V19,V20,V27
2	22	V7,V8,V9,V10,V13,V15,V16,V17,V21,V22,V24,V25 ,V26,V30,V31,V32,V34,V35,V36,V37,V38,V39
3	4	V11,V14,V28,V33
4	2	V18, V29
5	1	V23

Note: the variables were classified into five clusters of the thirty-nine variables are Analyzed by cluster analyses, the first cluster includes ten variables at the rate of 25.656%, the second cluster consists of 22 variables for a large percentage 56.410%, the third cluster contains four variables in the rate of 10.256%, the fourth cluster includes two variables in the rate of 5.128, and the last cluster includes one variable by the rate 2.564. A common way to show the cluster analysis is a dendrogram. Figure 2. Displays the dendrogram by using the average ward linkage. The number of clusters will be chosen into five discrete clusters based on 39 variables, showing the variables and the distance between them.



FIUGRE 2. Dendrogram using average linkage (between Groups).

DOI: https://doi.org/10.31972/ticma2022

14. Comparison Between Results of The Factor Analysis And The Cluster Analysis:

After applying factor and cluster analysis to the variables, a table has been created to compare their results. Thus, the following table illustrates the comparison.

TABLE 7. Comparison between results of the factor analysis and cluster analysis

Factor analysis		**Cluster analysis**	
Factor	Variables	Cluster	Variables
1	V1,V2,V3,V4,V5	1	V1,V2,V3,V4,V5,V6,V12,V19,V20,V27
2	V24 ,V32 ,V35,V37	2	V7,V8,V9,V10,V13,V15,V16,V17,V21,V22,V24,V25,V26,V30,V31,V32,V34,V35,V36,V37,V38,V39
3	V22, V19	3	V11,V14,V28,V33
4	V16,V18	4	V18,V29
5	V7,V10	5	V23

Note**:** the first factor and the first cluster have five variables in common, which are V1: I prefer the system. V2: The students can trust this system. V3: It helps the students be accepted in colleges and the specializations they like. V4: It has a good effect on the student's future. V5: It facilitates the students to choose the department they want. The second factor and the second cluster have four variables in common, which are V24: I prefer the admission of students with master's and PhD degrees; it must be done by using the new system. V32: I like all students to be accepted at the same time. V35: I suggest that the students be aware daily of the results when the competition is done in the system (by using their account). V37: The changes and guides must be given to all high schools before starting a new academic year. And the fourth row has one variable in common, which is V18: In a parallel system, the graduated students must not be allowed to refill admission forms.

15. CONCLUSIONS

The questionnaire items' answers agree on all variable paragraphs where the rates of total agreements are %67.199. The variable 25th (It is better if counting the average grade of 10th and 11th is optional) has the highest agreement ratio, reaching %83.943. Variable 37th (The changes and guides must be given to all high schools before starting a new academic year) has the second high agreement in the rate of %83.029; these two variables are more important for students.

Factor analysis has produced sixteen components affecting the students' Admission and their interests in the specialization. The first five components are fundamental in the results: the first component is more influential, representing the five variables (the system is helpful for students' Admission to college to get their desired professions), and the total effect of this component is %7.818. Furthermore, the second component (the new system may help the master's and PhD students to be admitted to the colleges and get competitive results by utilizing their accounts) and the total effect of this component is % 6.683. Moreover, in the third component (the student's desired professions need a higher mark to be admitted to the departments), the total effect of this component is % 3.849. The fourth component (decreasing the number of students' Admission in a parallel system by using the graduated students must not be able to refill admission forms) its rate interpretation is %3.828. The fifth component (geographic area and economic status), its rate interpretation is %3.789.

The Cluster analysis presents the number of clusters used to analyze five clusters. Therefore, the 39 variables that have been obtained were divided into these five clusters. Thus, the first cluster (the system is helpful for students' Admission to college to get their desired professions) involves ten variables with a rate of %25.656. The second cluster (the new system may help the master's and PhD students to be admitted to the colleges and get competitive results by utilizing their accounts) includes 22 variables with a significant percentage %56.410. The third cluster (the chance of Admission will be the same in private and government universities with filling a form suitable for the student's ability) involves four variables in the rate %10.256.

In addition, the fourth cluster (in a parallel system, the graduated and employee students cannot refill the admission form) includes two variables in the rate of %5.128. The last cluster contains one variable by the rate %2.564, which is (I like the order, and the choices will be decreased to 25 options).

There is a significant rapprochement and similarity between factor analysis and cluster analysis because the first cluster and component have the same variables in common. It confirms that both cluster and component analyses in the situational variable are classified in the study. Additionally, the results show that the most mutual variables between the factor analysis and cluster analysis can be considered the essential and influential variables for the student's Admission and their interests in their specializations. Consequently, in the first, five variables are in common: V1, V2, V3, V4, and V5. In the second, four variables are in common: V24, V32, V35, and V37. In the third, one is not in common, but in the fourth, one variable is in common: the V18: In a parallel system, the graduated students must not be allowed to refill admission forms.

ACKNOWLEDGMENTS

The authors would like to thank the yuzuncu yil university and salahaddin university.

REFERENCES

[1] R. Kumar, P. Desale, C. G. Puri., 2013. International Journal of Emerging Technology and Advanced Engineering **3,** 634-639.
[2] Harman, G., 1994. Student selection and admission to higher education: Policies and practices in the Asian region. *Higher Education*, 27(3),313-339.
[3] Kaur, A. and Hasija, S., 2015. A conceptual model of admission system and performance evaluation for a university. *International Journal of Computer Applications*, *125*(4).
[4] Tabachnick, B.G., Fidell, L.S. and Ullman, J.B., 2007. *Using multivariate statistics* (Vol. 5, pp. 481-498). Boston, MA: pearson.
[5] Tabachnick, B.G., Fidell, L.S. and Ullman, J.B., 2007. *Using multivariate statistics* (Vol. 5, pp. 481-498). Boston, MA: pearson.
[6] Basilevsky, A.T., 2009. *Statistical factor analysis and related methods: theory and applications*. John Wiley & Sons.
[7] Sánchez, J., 1982. MARDIA, KV, JT KENT, JM BIBBY: Multivariate Analysis. Academic Press, London-New York-Toronto-Sydney-San Francisco 1979. xv, 518.
[8] Tinsley, H.E. and Brown, S.D. eds., 2000. *Handbook of applied multivariate statistics and mathematical modeling*. Academic press.
[9] Härdle, W.K. and Hlávka, Z., 2015. *Multivariate statistics: exercises and solutions*. Springer.
[10] G. Singh, M. Kumar., 2014. *Global Journal of Management and Business Research* **14,** 39-48.
[11] J. P. Stevens., 2009. "Eploratory and Confirmatory Factor Analysis" in *Applied Multivariate Statistics For The Social Sciences*. (Fifth edition)(Taylor, New York),329-381.
[12] Ellis, J.L., 2017. Factor analysis and item analysis. *Applying Statistics in Behavioural Research*,11-59.
[13] Kenny, D.A., Kashy, D.A. and Cook, W.L., 2006. Dyadic Data Analysis (Methodology in the Social Sciences) The Guilford Press. *New York, NY, USA*.
[14] Matsunaga, M., 2010. How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, 3(1),97-110.
[15] Johnson, R.A. and Wichern, D.W., 2002. Prentice hall Englewood Cliffs. *Applied multivariate statistical analysis. 5th ed. New Jersy: Prentice hall Englewood Cliffs*.
[16] Comrey, A.L. and Lee, H.B., 2013. *A first course in factor analysis*. Psychology press.
[17] Miyamoto, S., 2012. *Fuzzy sets in information retrieval and cluster analysis* (Vol. 4). Springer Science & Business Media.
[18] Zakharov, K., 2016. Application of k-means clustering in psychological studies. *Tutorials in Quantitative Methods for Psychology*, *12*(2), pp.87-100.
[19] Hoppner, F., Klawonn, F., Kruse, R. and Rankler, T., 1999. Fuzzy Cluster Analysis: Methods for Classification Data Analysis and Image Recognition John Wiley & Sons Inc. *Bew York*.
[20] Rencher, A.C. and Christensen, W.F., 2002. Méthods of multivariate analysis. a john wiley & sons. *Inc. Publication*, *727*.
[21] Abonyi, J. and Feil, B., 2007. *Cluster analysis for data mining and system identification*. Springer Science & Business Media.
[22] Burns, R.P. and Burns, R., 2008. *Business research methods and statistics using SPSS*. Sage.
[23] Wierzchoń, S.T. and Kłopotek, M.A., 2018. *Modern algorithms of cluster analysis* (Vol. 34). Springer International Publishing.
[24] Everitt, B.S., Landau, S., Leese, M. and Stahl, D., 2011. Cluster analysis: Wiley series in probability and statistics. *Southern Gate, Chichester, West SussexUnited Kingdom: John Wiley & Sons*.

DOI: https://doi.org/10.31972/ticma2022

[25] Meena, L.K., Sen, C. and Kushwaha, S., 2017. Cluster analysis to form similarity for major selected crops in rajasthan, India. *International Journal of Current Microbiology and Applied Sciences*, 6(4),2673-2682.
[26] C. Chatfield, A. J. Collins.,1980 "Cluster Analysis" in *Introduction To Multivariate Analysis* (First edition)(chapman and hall, Boca raton),212-229
[27] Figueiredo Filho, D.B., da Rocha, E.C., da Silva Júnior, J.A., Paranhos, R., da Silva, M.B. and Duarte, B.S.F., 2014. Cluster analysis for political scientists. *Applied Mathematics*, *2014*.
[28] Namdeo, S.K. and Rout, S.D., 2016. Calculating and interpreting Cronbach's alpha using Rosenberg assessment scale on paediatrician's attitude and perception on self esteem. *International Journal of Community Medicine and Public Health*, 3(6),1371-1374.
[29] Mohajan, H., 2017. Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University Economics Series*, (4).