T.R.
VAN YUZUNCU YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
STATISTICS DEPARTMENT

USING MULTIVARIATE METHODS TO DETERMINE THE MOST IMPORTANT AFFECTING FACTORS FOR STUDENTS' ADMISSION AND THEIR INTERESTS IN THE SPECIALIZATIONS: A SAMPLE OF SALAHADDIN UNIVERSITY

M. Sc. THESIS

PREPARED BY: Mohammed Othman ABDULLAH
SUPERVISOR: Asst. Prof. Dr. Yener ALTUN
CO-SUPERVISOR: Asst. Prof. Dr. Rizgar Maghdid AHMED

VAN-2020

T.R.
VAN YUZUNCU YIL UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
STATISTICS DEPARTMENT

USING MULTIVARIATE METHODS TO DETERMINE THE MOST IMPORTANT AFFECTING FACTORS FOR STUDENTS' ADMISSION AND THEIR INTERESTS IN THE SPECIALIZATIONS: A SAMPLE OF SALAHADDIN UNIVERSITY

M. Sc. THESIS

PREPARED BY: Mohammed Othman ABDULLAH

VAN-2020

ACCEPTANCE AND APPROVAL PAGE

This thesis entitled "**Using Multivariate Methods to Determine The Most Important Affecting Factors for Students' Admission and Their Interests in The Specializations: A Sample Of Salahaddin University**" presented by Mohammed Othman ABDULLAH under supervision of Asst. Prof. Dr. Yener ALTUN in the department of Statistic has been accepted as a M. Sc. thesis according to Legislations of Graduate Higher Education on 23/01/2020 with unanimity of votes members of jury.

Signature

Chair: Prof. Dr. H. Eray ÇELİK

Member: Assoc. Prof. Dr. Ayşe METİN KARAKAŞ

Member: Asst. Prof. Dr. Yener ALTUN

This thesis has been approved by the committee of The Institute of Natural and Applied Science on 31/01/2020 with decision number 2020/8-1

Prof. Dr. Suat ŞENSOY
Director of Institute

THESIS STATEMENT

All information presented in the thesis obtained in the frame of ethical behavior and academic rules. In addition all kinds of information that does not belong to me have been cited appropriately in the thesis prepared by the thesis writing rules.

Signature

Mohammed Othman ABDULLAH

ABSTRACT

USING MULTIVARIATE METHODS TO DETERMINE THE MOST IMPORTANT AFFECTING FACTORS FOR STUDENTS' ADMISSION AND THEIR INTERESTS IN THE SPECIALIZATIONS: A SAMPLE OF SALAHADDIN UNIVERSITY

ABDULLAH, Mohammed Othman
M. Sc. Thesis, Statistics Department
SUPERVISOR: Asst. Prof. Dr. Yener ALTUN
CO-SUPERVISOR: Asst. Prof. Dr. Rizgar Maghdid AHMED
January, 2020, 63Pages

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 the number of the clusters into five clusters. Furthermore, the results of the factor analysis and cluster analysis will be compared to each other. Moreover, this study depends on analysis of 350 questionnaire forms, which is distributed by random stratified sample method on students in the first stage of three different colleges include 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 program has been used in data analysis. Additionally, the results have demonstrated that the 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 the student admission and their interests in choosing specialization. Consequently, the first factor and the first cluster have five influential variables in common, they are V1, V2, V3, V4 and V5 (the system is helpful for student admission in the colleges to get his/her desired professions). Ultimately, the conclusion has shown that there is a kind of approach and similarity between factor analysis and cluster analysis.

Keywords: Admission, Cluster analysis, Factor analysis, Interests, principal component, Specialization.

ÖZET

ÖĞRENCİ KABULÜNDE UZMANLIK ALANLARINI BELİRLEMEK İÇİN ETKİLİ OLAN EN ÖNEMLİ FAKTÖRLERİN BELİRLENMESİNDE ÇOK DEĞİŞKENLİ YÖNTEMLERİN KULLANIMI: SALAHADDİN ÜNİVERSİTESİ ÖRNEĞİ

ABDULLAH, Mohammed Othman
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Dr Öğr Üyesi. Yener ALTUN
İkinci Danışmanı: Dr Öğr Üyesi. Rizgar Maghdid AHMED
Ocak, 2020, 63 Sayfa

Bu tezin temel amacı, öğrenci kabulü için en önemli etkili faktörleri ve uzmanlaşma alanlarına ilgisini çok değişkenli yöntemler kullanarak belirlemektir. Bu nedenle, elde edilen bir dizi faktörü belirleyerek faktör analizi ve kümelerin sayısını beş kümede sınıflandırarak kümeleme analizine odaklanmıştır Ayrıca, faktör analizi ve kümeleme analizinin sonuçları birbiriyle karşılaştırılmıştır. Aynı zamanda, bu çalışma, 2018-2019 akademik yılı için Kuzey Irak'taki Salahaddin Üniversitesi'nin Fen ve Sosyal Bilimler kolejlerini içeren üç farklı kolejin ilk aşamasında öğrenciler üzerinde rastgele tabakalı örnekleme yöntemi ile dağıtılan 350 anket formunun analizine dayanmaktadır. Bu nedenle veri analizinde SPSS yazılım programı kullanılmıştır. Ayrıca sonuçların güvenirliği kabul edildiğini ve faktör analizinde toplam varyans yorumlama oranının% 62.157 olduğu sonucuna ulaşılmıştır. Aynı zamanda, faktör analizi ile kümeleme analizi arasındaki en yaygın değişkenler, öğrenci kabulü ve uzmanlık alanı seçimindeki ilgi alanları için en önemli etkili değişkenler olarak düşünülebilir. Sonuç olarak, ilk faktörün ve ilk kümenin ortak beş etkili değişkeni vardır, bunlar V1, V2, V3, V4 ve V5'tir (sistem, kolejlerde öğrencinin istediği meslekleri almasına yardımcı olur). Böylece bu durum faktör analizi ile küme analizi arasında bir tür yaklaşım ve benzerlik olduğunu göstermiştir.

Anahtar kelimeler: Faktör analizi, Kabul, Kümeleme analizi, Temel bileşen, İlgi alanları, Uzmanlaşma.

ACKNOWLEDGMENT

First and foremost, I would like to thank Allah for helping me in my successful study and I would like to thank my supervisors Asst. Prof. Dr. Yener ALTUN and Asst. Prof. Dr. Rizgar Maghdid AHMED for their supports, guidance, encouragement, and patience. Then, my best regards also to all other members of the Statistic Department, especially to Prof. Dr. H. Eray Çelik. I am also grateful to everyone whom I have learned many forms. And the great characteristics of my parents who they always help me and support me in every field in my life, so words can’t show gratitude to their supports. Furthermore, my evergreen gratefulness goes to my wife for helping me and encouraging me in achieving my best goals. Finally, thanks for everyone who has helped me even by a word.

January 2020

Mohammed Othman ABDULLAH

TABLE OF CONTENTS

Pages

LIST OF TABLES

Table **Pages**

LIST OF FIGURES

Figures **Pages**

SYMBOLS AND ABBREVIATIONS

Some symbols and abbreviations used in this study are presented below, along with Descriptions.

Symbols	Description
%	Percent
$\boldsymbol{\mu}$	Mean of variables
$\boldsymbol{\ell_{ij}}$	Loading of the *i- th* variable on the *j- th* factor
F	Common factor
$\boldsymbol{\mathcal{E}}$	Specific factors
L	Matrix of factor loadings
$\boldsymbol{E}$	Expectation value
$\boldsymbol{\Sigma}$	Covariance matrix
$\boldsymbol{\Psi}$	Pis (Greek)
$\boldsymbol{\sigma^2}$	Variance
$\boldsymbol{\sigma}$	Standard deviation
$\boldsymbol{h_i^2}$	Communalities
$\boldsymbol{r_{jJ}}$	Reliability
$\boldsymbol{R_{j.oth}^2}$	Squared multiple correlation
$\boldsymbol{\lambda}$	Eigenvalu
$\boldsymbol{a}$	Eigenvector
$\boldsymbol{R}$	Correlation matrix
S	Sample covariance matrix

Abbreviations	Description
SMC	Squared Multiple Correlation
Var	Variance
Cov	Covariance

Abbreviations	**Description**
PC	Principal Component
PCA	Principal Component Analysis
FA	Factor Analysis
EFA	Exploratory Factor Analysis
CFA	Confirmatory Factor Analysis
KMO	Kaiser-Meyer-Olkin
V	Variable
SPSS	Statistical package for the social sciences

1. INTRODUCTION

One of the important topics in the life of the students is admission in universities or institutes according to his/her interests. In Iraq and northern Iraq, the students finish the study period for (12) years until he/she reaches his/her future in university or institute. Previously the admission system at the university was based on filling a variety of admission forms in (central admission) department, which links the ministry of education with higher education for the admission of the students in universities and institutes. Nowadays, the ministry of Higher Education in northern Iraq depends on the new system and this system 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 a vital part of any university's running because universities can stay alive with students. The admission system is an influential activity, which helps the universities to survive with the student's contribution. Then, lack of knowledge by the students and poor admission system are serious problems that make fewer students to be admitted in the universities because of slow response of the system, which leads the students to make mistakes. Therefore, students are not able to get her/his desired universities. For these reasons, there must be good educated information about the online admission system; it could be achieved only by choosing the best system for admission (Kumar et al., 2013).

Although the admission system can consider widely to the process when a student gets more interested in reaching higher education until admission in a particular course and university happens (Harman, 1994). The admission process of the university is an entrance process in a university. It generally influences all the resources of the universities its quality restraints (Kaur and Hasija, 2015).

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 is 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 different systems (Zankoline, Credit, Parallel, and Evening study). The student introduces many requests to more than one university by filling a form online in a system, which is specialized for the student admission. The current admission system gives him or her acceptance in one university, so it leads to give many chances for 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 (delay the time to receive results , not taking the interest of students as regarded and lack of trust of student to the system…etc).

The important aim of this topic lies in several aspects: it is related to student's future in his/her life, and there are several problems facing the student in this system such as the chosen department may not be his/her interests or sometimes some students' names might be missed. In this thesis, all the important aspects of this system are studied, processed using advanced statistical method, it includes two types of multivariate analysis they are cluster analysis and factor analysis. In addition, some recommendations are considered as an important effective service to the students to achieve the departments which they are interested.

This study depends on the obtained information through the distribution of questionnaire forms among 350 students of the first stage in three different departments in Salahaddin University for the academic year 2018-2019. Choosing random stratified sample and analyzed data by IBM SPSS Statistics V:25 program to determine the most important affective factors by using factor analysis and cluster analysis then create a new table to compare between them in the results, also presenting many statistical graphs and tables.

The thesis consists of five chapters. The first chapter is about introduction and background information of the new admission system, the second chapter displays literature review, the third chapter contains the theoretical part of factor analysis and cluster analysis, the fourth chapter shows the practical part, its data analysis and results, and the last chapter presents conclusion of the study.

2. LITERATURE REVIEW

The student admission in the universities or institutes is a fateful process; it faces graduate students in high schools. There are many factors, which affected the selection colleges or chosen their desire in the specialization. Moreover, there are many researches about the students admission and their interests in the specialization, each one of them has own way with different explanation to describe the process, then the reasons will be clarified. Finally, the purpose of this chapter is to examine the nearest and related researches to our research.

Chapman (1981) presented a model, which suggests that university choice is affected by a set of students' characteristics. Therefore, they are internal factors and series of external factors choices regarding college selection. Moreover, among the student characteristics can be stated as socio-economic status, aptitude, level of educational aspiration, and high school performance. Then, an external influence includes the influence of significant persons. Thus, the fixed characteristics of the college and the own efforts of the colleges to communicate with students.

Beswick (1989) elucidated that factor study is linked with the choice of student in the colleges and universities. The process shows a selection process of 227 students, which were first year students who have attended one of the colleges in Alberta, the process have been done with a feedback form to identify those factors, which focused on at the period of university selection process. Correlations, means, analyses of variance and qualitative data provided the statistical and descriptive information for interpretation. The result has found that parents, especially mothers are the important and influential characters who impacted the selection process.

Kallio (1995) clarified the factors affecting the decision of graduate students in college choice process, which is built on a 1986 survey of 2,834 admitted students at a major research university, which %38 of the model responded, the importance ratings of factor analysis of 31 college characteristics based on the student decisions. So the results were used to create five scales of importance these results were used to build five scales of importance and favorite, and then examined with other variables in regression form, which the dependent variable was the decision to register or not register at the surveying institution. Ultimately, these models have been influenced the student decisions: residency status, quality, works, financial aid, and the campus social environment.

Mincer-Daszkiewicz (2004) stated that student admission system in Warsaw university is depended on a software application to register students in the university. The program will be revised and improved every year due to the a new system of maturity examinations, which introduced in 2005, it serves the process and utilized by the higher education in Poland. Therefore, the result shows that it was the best model to deliver what is required by the students electronically and closely. Thus, the system, which is remote registration will be broken if students would have to send hard forms of maturity examination at the early phase of the evaluation process. By the way, the maturity examinations are matched and controlled by special commissions appointed.

Khoo et al. (2004) proved that a cluster analysis of LIS students in Singapore and implications for defining areas of specialization. The process has been done via questionnaire survey of students and Master students in Science information, which focuses on the program in Singapore, at the Nanyang Technological University. Besides, a survey of candidates of the Cluster analysis program was fulfilled on the questionnaire information. Then the result proved that the cluster analysis was understood by the students as a useful program of the students' interests and how the interests were ordered.

Biró (2008) assumed that Gale and Shapley as an admission system in Hungary. So, the program is built on the fundamental form and process of Gale and Shapley, which finds this program interesting. In addition, the results found that the process of Gale and Shapley submit an application for the generalized model too, the program produces secure mark-limits, and thus these solutions are either the best or worst stable score-limits for the students.

Braun and Dwenger (2009) pointed at the successful admission process in university, in Germany regional provenance matt. Trying a understandable data the German central clearing house for the process of university admissions in 2006-2007, so the regression model is utilized. In conclusion the successful rates certainly have been showed differently between federal states. Majority of the variation successfully can be clarified by different level grading of the states. By presenting quotas for federal states and preventing struggles between applicants to the level of the, so the relation among level grading of the state and successful rates in the university admission process can be.

Wakil et al. (2014) targeted at a system for Private Universities' admission in Iraq; UHD Case Study, which is called Web Recommender System. Create a new admission process by utilizing a hybrid model of Neural Network, Decision Tree, and Our Proposed Algorithm. Vividly the result determined that the system is better than other variable admission system. Further going, the system can provide a good course for applicants. Accordingly, the system helps high school students to admit to university and choose the best departments to study.

Alotaibi et al. (2016) focused on an admission system in Saudi universities, which is a system approach. Also, the gathered information could be utilized to measure quality, perform analyses and diagnoses, depending on the standards such as curricula and syllabi, and recommend alternatives in decision processes. The results demonstrate that the present admission system in Saudi universities requires improving, because the system is not understandable or not related. Thus, the system needs new approaches related to DSS (such as; the proposed system) to develop the effectiveness of the admission system in the universities.

Gulluce et al. (2016) elucidated Factors Affecting the University Preferences of Students. It is a survey was focused on the students studying in Faculty of Administrative Science and Economics of Kafkas University (2013/2014). 309 of the surveys gained at the departments of economics, administration, and public administration were observed. Accordingly, the result of the factor analysis shows six dimensions consisting of prestige, opportunity, campus; knowledge, location, and economy were gained. More to the point, it was recognized that these factors differ in demographic properties.

Ilgan et al. (2018) proved Factors Affecting University Choice: An evaluation on University Freshman Students. The data was gained by a questionnaire form improved by the researchers and gave to the 630 students admitted to two different universities in Turkey. The study was circled in a quantitative paradigm. The survey depends on the questionnaire, which is called “Factors Affecting University Choice Scale” expressed 59.58 % of total variance with six dimensions, another result which is considered to be significant is that as the student’s university access exam scores increase, and the dimensions that are effective in the university preference have been given less importance.

Ultimately, after focusing on many researches, the most important feature of the research is that there is not a single research which used factor analysis and cluster analysis as a new system for student admissions in the universities and institutions in North of Iraq. Thus, the results show several significant factors, which influence the new system.

3. METHODOLOGY

This chapter includes theoretical part. Primarily, shows methodology of creating questionnaire form and distribution and collection data. Likewise, in this section try to study definitions completely such as, meaning, hypothesis, aim, conditions, and processes of the factor analysis and the cluster analysis by getting benefit from many books, articles, internet, and researches.

For better understanding, the action of students' admission and those problems, which face students in the new system in northern Iraq, with good and rescannable solutions for the preferred system, by using a statistical data that had been analyzed through cluster analysis and factor analysis. Therefore, a statistical Mechanism will be applied to get the data, which explained below:

A set of questions is prepared in questionnaire form, the questions will be prepared according to admission system strategy, which focused on age, gender, specialization, interest, average score, and 39 variables, which related to the student admission system and their interests in the specializations. Then, this questionnaire form is distributed among 350 students of first stage at Salahaddin University-Erbil for the academic year 2018-2019 in 3 different colleges, which are scientific and humanity colleges at 10 different departments. Moreover, Stratified Random Sampling has been used. Thus, they are 251 students in Administrator and Economics College, 74 students in Science College, and 25 students in Education College, as follows:

$$n_0 = \frac{z^2\, pq}{e^2} \tag{3.1}$$

$$n_0 = \frac{(1.96)^2 (0.5)(0.5)}{(0.05)^2}$$

$$n_0 = 385$$

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}} \tag{3.2}$$

$$n = \frac{385}{1 + \frac{(385 - 1)}{2405}} \ , \qquad n = 331 \ , \qquad n \simeq 350$$

Table 3. 1. Stratified random sampling explanation

College	Nh = population size for hth stratum	Formula: $nh = \frac{Nh}{N} * n$	nh = sample size for hth stratum
Administrator and Economics	1725	$n1 = \frac{1725}{2405} * 350$	$n1 = 251$
Science	509	$n2 = \frac{509}{2405} * 350$	$n2 = 74$
Education	171	$n3 = \frac{171}{2405} * 350$	$n3 = 25$
Total	$N = 2405$		$n = 350$

Profoundly, the data is collected manually and entered into (statistical package for the social sciences) IBM SPSS statistics V:25 program for analyzing them. Additionally, codes have been given to the variables; the data was analyzed by using multivariate method, which is focused on factor analysis (principal component method) and cluster analysis (division of variables into five clusters). Besides, some graphs and tables have been produced, which demonstrates the relationship between variables. Then, it facilitates to indicate the problematic points between the new admission system and students' interests in the specializations, which it may lead to find a suitable solution by using both factor analysis and cluster analysis. In addition, the results have been found in groups with related variables. Then, created a new table to compare between factor analysis and cluster analysis in the results also presented many statistical graphs and tables that have been obviously showed in chapter four.

3.1. Factor Analysis

Factor analysis (FA) is a strategy, which is used in Statistics that is applied to a solo set of variables by statistical researchers who interested in finding which variables are related to others and which one is independent in the set of the variables. Therefore, those variables that are in association with other variables but are not correlated with other subsets of variables are linked into factors. In addition, factors are accustomed to reflect to the process, which have built the correlation between variables (Tabachnick et al., 2012). Moreover, factor analysis is a set of method for clarifying the correlations between variables in form of basic entities, which are called factors. Additionally, where the model 'factor analysis' is broadly comprehended to refera collection of nearly associated forms intended for clarifying or creating co-relational structure between the observed random variables (Basilevsky, 2009).

One of the major 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 to random errors. Thus, the assumption is specially well-organized to the forms like psychology, where it is not suited to measure directly the concepts, which one interested in for instance 'intelligence'. Then it is mostly ambiguous to define the concepts (Mardia et al., 1979). Furthermore, the utilization of the factor analysis's methods have been enlarged in many scientific research eras such as psychological, educational, athletic, marketing, behavioral, social, medicine, economics, and geography as a result of the technological development of computers.

The aims of factor analysis are to explain the number of basic impacts underlying a domain of variables, to count the amount of the variable association with the factors and to gain enough instruction about their nature to observe the contribution of each factor with its variables. Likewise, the specific targets of the subject is to give a brief about the patterns of correlations between observed variables, by this way the 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 (Tinsley and Brown, 2000; Härdle and Hlávka, 2007; Singh and Kumar, 2014).

3.1.1. Type of factor analysis

Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are two fundamental kinds of factor analysis, which normally a difference is made between Exploratory and Confirmatory factor analysis, the aim of exploratory analysis is to demonstrate the factor structure for a collection of variables. It mostly includes identification how a lot of the factor loadings. Although majority of the EFA programs permit for the amount of factors to be particularized in advance. Moreover, the variables could not be forced to load just on some specific factors. Therefore, EFA is usually identified to be more of a theory- generating than a theory-testing program.

On other side, Confirmatory Factor Analysis is normally built on a powerful theoretical and empirical organization that permits the scientific researcher to identify accurate factor model in advance. Therefore, the model generally identifies which variables will load on which factors. Likewise, which factors are associated. Furthermore, Confirmatory Factor Analysis is theoretical testing program more than procedure EFA. Practically, the research may include aspects of both Exploratory and Confirmatory analysis. It is beneficial to differentiate between the both techniques in a situational form in which they are utilized (Stevens, 2002; Ellis, 2017; Kenny et al., 2006; Matsunaga, 2010).

Table 3. 2. Comparison between exploratory theory generating and confirmatory theory testing

Exploratory theory generating	Confirmatory theory testing
• Heuristic—weak literature base • Determine the number of factors • Determine whether the factors are correlated or uncorrelated • Variables free to load on all factors	• Strong theory and/or stronng gempirical base • Number of factors fixed a priori • Factors fixed a priori as correlated or uncorrelated • Variables fixed to load on a specific factor or factors

3.1.2. 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 $F_1, F_2, ..., F_m$, which are called common factors, and ρ extra sources of variation such as $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, which are called errors, or sometimes particular factors. Especially, the factor analysis model is

$$X_1 = \mu_1 + \ell_{11} F_1 + \ell_{12} F_2 + \cdots + \ell_{1m} F_m + \varepsilon_1$$

$$X_2 = \mu_2 + \ell_{21} F_1 + \ell_{22} F_2 + \cdots + \ell_{2m} F_m + \varepsilon_2$$

$$X_p = \mu_p + \ell_{p1} F_1 + \ell_{p2} F_2 + \cdots + \ell_{pm} F_m + \varepsilon_p \quad (3.1)$$

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

$$X_{(p\times1)} = \mu_{(p\times1)} + L_{(p\times m)} F_{(m\times1)} + \varepsilon_{(p\times1)} \quad (3.2)$$

X: vector of observable random variables. μ: mean of variables. L: matrix of factor loadings. F: common factors. ε: which are called errors, or from time to time specific factors.

Or in matrix data.

$$\underline{x} = L\underline{F} + \underline{\varepsilon} \quad (3.3)$$

(Johnson and Wichern, 2002; Comrey and Lee, 1992)

3.1.3. These assumptions and the relation ın the orthogonal factor model

It is be assumed that familiar factor variables have zero means and unit variances, Furthermore, the connection among any given common factor and any unique factor is zero.

The coefficient ℓ_{ij}, which is called the loading of the *i*th variable on the *j*th factor, Therefore, the matrix **L** is the matrix of factor loadings. Make a note that the ith particular factor ε_i is correlated just with the ith response Xi. The P deviations $X_1 - \mu_1$, $X_2 - \mu_2$, … , $X_p - \mu_p$ are identified in terms $p + m$ random variables F_1 , F_2 ,..., F_m, ε_1 , ε_2 ,…, ε_p that are unobservable.

With a big number of unobservable quantities, an exact confirmation of the factor from explanation on X_1, X_2 … X_p is desperate. However, with a number of extra assumptions about the random vectors **F** and ε, the model in (3.2) implies clear covariance associations that can be checked.

For further simplification, we assume that

$$E(F) = 0_{(m\times1)}$$

$$\text{cov}(F) = E(FF') = I_{(m\times m)}$$

$$\mathrm{E}(\varepsilon) = 0_{(p\times 1)}$$

$$\mathrm{cov}(\varepsilon) = \mathrm{E}(\varepsilon\varepsilon') = \Psi_{(p\times p)} \tag{3.4}$$

Where Ψ is a diagonal matrix

$$\Psi_{(p\times p)} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$$

And that F and ε are independent, so

$$\mathrm{cov}(\varepsilon, \mathrm{F}) = \mathrm{E}(\varepsilon \mathrm{F}') = 0_{(p\times m)}$$

From the model (3.2) and the assumptions $\mathrm{E(F)} = 0$ to $cov(\varepsilon, F) = 0$, we obtain the following covariance structure for X and F:

$$(\mathrm{X}-\mu)(\mathrm{X}-\mu)' = (\mathrm{LF}+\varepsilon)(\mathrm{LF}+\varepsilon)'$$

$$= (\mathrm{LF}+\varepsilon)((\mathrm{LF})' + \varepsilon')$$

$$= \mathrm{LF}(\mathrm{LF})' + \varepsilon(\mathrm{LF})' + \mathrm{LF}\varepsilon' + \varepsilon\varepsilon'$$

So that

$$\Sigma = \mathrm{cov}\,(\mathrm{X}) = \mathrm{E}(\mathrm{X}-\mu)(\mathrm{X}-\mu)'$$

$$= \mathrm{LE}(\mathrm{FF}')\mathrm{L}' + \mathrm{E}(\varepsilon\mathrm{F}')\mathrm{L}' + \mathrm{LE}(\mathrm{F}\varepsilon') + \mathrm{E}(\varepsilon\varepsilon')$$

$$\mathrm{cov}\,(\mathrm{X}) = \mathrm{LL}' + \Psi \tag{3.5}$$

Or

$$\mathrm{var}\,(\mathrm{X_i}) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i \tag{3.6}$$

$$\mathrm{cov}\,(\mathrm{X_i}, \mathrm{X_k}) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km} \tag{3.7}$$

According to Ψ is a diagonal matrix. Also, by independence,

$$\mathrm{cov}(\varepsilon, \mathrm{F}) = \mathrm{E}(\varepsilon, \mathrm{F}') = 0$$

Also, by the model in(3.2) ,

$$(\mathrm{X}-\mu)\mathrm{F}' = (\mathrm{LF} + \varepsilon)\mathrm{F}' = \mathrm{LFF}' + \varepsilon\mathrm{F}'$$

So

$$\mathrm{cov}(\mathrm{X}, \mathrm{F}) = \mathrm{E}(\mathrm{X}-\mu)\mathrm{F}' = \mathrm{LE}(\mathrm{FF}') + \mathrm{E}(\varepsilon\mathrm{F}') = \mathrm{L} \tag{3.8}$$

Or

$$\mathrm{cov}(\mathrm{X_i}, \mathrm{F_j}) = \ell_{ij} \tag{3.9}$$

The form $\mathrm{X} - \mu = \mathrm{LF} + \varepsilon$ is linear in the familiar factors. If the *P* replies X are actually associated to underlying factors, but the correlation is nonlinear. Furthermore, the covariance structure $= LL' + \Psi$ may not be sufficient. The essential assumption of linearity is natural in the formulation of the usual factor model.

That section of the variance of the *i* th variable involved by the *m* common factors is called the *i* th communality. The section of $var\ (X_i) = \sigma_{ii}$ because of the specific factor is mostly named the uniqueness, or specific variance. Denoting the *i* th communality by h_i^2, we notes that the *i* th communality is the sum of squares of the loadings of the *i* th variable on the *m* common factors.

$$\sigma_{ii}^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 + \psi_i \tag{3.10}$$

$\text{Var}\ (\text{Xi}\ ) = \sigma_{ii}^2$, $\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 = \text{communality}$, $\psi_i = \text{specific variance}$

or

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 \tag{3.11}$$

and

$$\sigma_{ii} = h_i^2 + \psi_i \quad , \qquad i = 1,2,3, \dots, p \tag{3.12}$$

(Johnson and Wichern, 2002; Cattell, 1965; Hassan et al., 2012)

3.1.4. Sample size

Sample size is essential in factor analysis. There are various views and a number of leading rules of thumb are referred in literature. So, from sample to sample, correlation coefficients will be fluctuated, it will be appeared in small samples more than ample samples. Thus, the dependability of factor analysis is reliant on sample size. Besides, the necessity sample size for factor analysis concluding in several rules of thumb, the common rule is to propose, which the researcher merely has 10 to 15 participants for every single variables. Actually, Tabachnick et al. (2007) demonstrate that factor analysis at least needs 300 cases. Furthermore, Comrey and Lee (1992) indicated that the guide to sample sizes: 100 as poor, 200 as fair, 300 as good, 500 as very good, And 1000 or more as excellent (Field, 2005).

3.1.5. Communalities

The communality is the sum of squared factor loadings for a given indicator across all factors. The communality is normally the factor loading squared. Factor analysis utilizes variances to prepare communalities among variables. The variance is equivalent to the square of the factor loadings. In addition, in several methods of factor analysis, the target of pulling out is to eliminate mostly common variance in the first factor as possible. The communality is the variance in the noted variables that are accounted for by a common factor or common variance. So, h^2 denotes the communality, which is the summation of the squared correlations of the variable with the factors. The formula for deriving the communalities is: (Yong and Pearce, 2013)

$$h_i^2 = \sum_{i=1}^{m} \ell_{ij}^2 \qquad j= 1,2,\ldots, p \quad , i = 1.2,\ldots, m \tag{3.13}$$

Where $\boldsymbol{\ell_{ij}}$ equals the loadings for j variables. The values of Communality are positive, which fall between 0 and 1.

$$0 \leq h_j^2 \leq 1$$

Estimating communalities have several different ways. Commonly, it is the broad method that has been used in choosing the communality of each variables, which is related to the squared multiple correlation coefficient (SMC) of the variable with all other variables. (Reyment and Jvreskog, 1996)

$$\mathrm{SMC} = R_{j.oth}^2 = 1 - \frac{1}{r_{jj}}$$

The values of communality, which is located between tow limits such as, the squared multiple correlation can be utilized to put a lower bound on h_j^2 and the reliability of the variable sets an upper bound on h_j^2.

$$R_{j.oth}^2 \leq h_j^2 \leq r_{jJ}$$

Where $R_{j.oth}^2$ is the square of the multiple correlation between variable *j* and the *m* -1 other variables, and where r_{jJ} is the reliability with which variable *j* is measured (Harris, 2001).

3.1.6. Factor extraction method

The factor extraction method permits us to guess factor loadings and correlations between factors. The selection of each method will rely on the researcher's aim; the method needs the performance of the distributional assumptions. Besides, several methods utilize to estimate the common factor model such as (Principal components, Principal factors, Image factoring, maximum likelihood factoring, Alpha factoring, Unweighted least squares, and generalized least squares). For this thesis the principal component analysis was used (Tabachnick et al., 2012; Brown, 2006; Williams et al., 2010).

3.1.6.1. Principal components

Principal components are uncorrelated, linear combinations of the original variables. Therefore, they are provided for a preservative partition of the total variance. So, Principal Component's analysis is available to be utilized to can be used to construct composite variables, which are orthogonal to one another. Besides, the principal components are suitable to summarize many of the variance in a big collection of variables in association with a small number of components, thus the components are not mostly interpretable (Stevens, 2002).

The aim of PCA is to pull out most variance from the data collection with each component. In addition, the primer main component is the linear combination of noted variables the mostly splits subjects by getting the most out of the variance of their component scores. Moreover, the second component is shaped from remaining correlations; additionally, it is the linear combination of observed variables, which pulls out most variability uncorrelated with the primer component. Going further, Subsequent components also pull out most variability from residual correlations and are orthogonal to all earlier pull out components. Therefore, the principal components are structured with the primer component pulling out the maximum variance and the final component the smallest amount variance. Moreover, the solution is mathematically exclusive and, if all components are preserved, accurately reproduces the observed correlation matrix. By the way, PCA is the solution of selection for the researcher who is mainly interested in reducing a big number of variables down to a minor number of components. Also, PCA is essential as an first step in FA where it shows a large deal about most number and nature of factors (Tabachnick et al., 2012; Krishnan, 2011).

3.1.6.2.The characteristic equation

The principal components analysis targeted on estimating the correlation matrix and it can be done via discovering the characteristic equation of the matrix. it needs two collections of values.

1- The characteristic vectors of the matrix, also is named latent vectors or eigenvectors, is a vector normally is a column or row of numbers in a matrix. And its symbol is ($\boldsymbol{a}$).
2- Characteristic roots, also named latent roots or eigenvalues, so its symbol is ($\boldsymbol{\lambda}$) (Kline, 1994).

3.1.6.3. Definition of principal components

Mathematically, principal components are special linear combinations of the p random variables X_j =(X_1, X_2... X_p), where PCs depend only on the covariance matrix Σ (or the correlation matrix$\boldsymbol{R}$) of (X_1, X_2... X_p). Let the random vector $X' = [X1, X2, \ldots, Xp]$ have the covariance matrix Σ with eigenvalues $\boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \cdots \geq \boldsymbol{\lambda}_p \geq 0$. Consider the linear combinations:

$$pc_1 = a'_1X = a_{11}\, X_1 + a_{12}\, X_2 + \cdots + a_{1p}\, X_p$$

$$pc_2 = a'_2X = a_{21}\, X_1 + a_{22}\, X_2 + \cdots + a_{2p}\, X_p$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$pc_j = a'_jX = a_{j1}\, X_1 + a_{j2}\, X_2 + \cdots + a_{jp}\, X_p \tag{3.14}$$

$$pc_j = \sum_{k=1}^{p} a_{jk}\, X_k \qquad j, k = 1,2, \ldots, p \tag{3.15}$$

or, in matrix form,

$$\underline{\mathrm{pc}} = \underline{\mathrm{x}}\mathrm{A} \tag{3.16}$$

In this discuss a number of the properties of principal components. The sample variance of each of the other principal components is equivalent to its related eigenvalue:

$$\mathrm{var}\left(\mathrm{pc_j}\right) = \mathrm{a'_j\,S\,a_j} = \boldsymbol{\lambda}_\mathrm{j} \qquad \mathrm{j} = 1,2,\dots,\mathrm{p}$$

Covariance between two components is equal to zero:

$$\mathrm{cov}\left(\mathrm{pc_j},\mathrm{pc_k}\right) = \mathrm{a'_j\,S\,a_k} = 0 \quad \text{for any} \quad \mathrm{j} \neq \mathrm{k}$$

Covariance between variables and each of principal components is:

$$\mathrm{cov}\left(\mathrm{pc_j},\mathrm{X_k}\right) = \mathrm{a'_k}\boldsymbol{\lambda}_\mathrm{j}\mathrm{a_j} = \boldsymbol{\lambda}_\mathrm{j}\mathrm{a_{jk}}$$

$$\mathrm{var}\left(\mathrm{pc_j}\right) = \boldsymbol{\lambda}_\mathrm{j} \quad \text{and} \quad \mathrm{var}(\mathrm{X_k}) = \mathrm{S}$$

Then

$$\mathrm{cov}\left(\mathrm{pc_j},\mathrm{X_k}\right) = \boldsymbol{\lambda}_\mathrm{j}\mathrm{S}$$

That is by using *S* matrix:

$$\mathrm{r}\left(\mathrm{pc_j},\mathrm{X_k}\right) = \frac{\mathrm{a_{jk}}\,\sqrt{\boldsymbol{\lambda}_\mathrm{j}}}{\sqrt{\mathrm{S}}} \qquad \mathrm{i,k} = 1,2,\dots,\mathrm{p} \tag{3.17}$$

By using R matrix:

$$\mathrm{r}\left(\mathrm{pc_j},\mathrm{X_k}\right) = \mathrm{a_{jk}}\,\sqrt{\boldsymbol{\lambda}_\mathrm{j}} \qquad \mathrm{i,k} = 1,2,\dots,\mathrm{p} \tag{3.18}$$

The total sample variance of the components is equal to the total sample variance of the variables. By using (*S* and *R*)

$$\mathrm{tr(S)} = \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{var(x)} = \sum_{\mathrm{l}=1}^{\mathrm{p}} \lambda_\mathrm{j} \tag{3.19}$$

By using R matrix

$$\mathrm{tr(R)} = \sum_{\mathrm{l}=1}^{\mathrm{p}} \boldsymbol{\lambda}_\mathrm{j} = \mathrm{p} \tag{3.20}$$

(Rencher, 1998; Harris, 2001; Johnson and Wichern, 2002)

3.1.7. Rotation of factors

Rotation of factors is a process, which the solution is created extra interpretable without altering underlying properties. Therefore, rotation is normally utilized after pulling out to make the most of the high correlations among factors and variables and reduce low ones. Thus, the consequences of factor extraction, unaccompanied by rotation, are possible to be tough to understand in spite of which technique of extraction is utilized. Subsequent to extraction, rotation is utilized to develop the interpretability and scientific utility of the solution. So it is not utilized to develop the quality of the algebraic fit between the observed and reproduced correlation matrices because of all orthogonally rotated solutions are algebraically equal to one another to the solution before rotation. Fundamentally, factor rotation has three dissimilar approaches, which are familiarly utilized in factor analysis: (a) graphical rotation, (b) analytic rotation, and (c) rotation to a target matrix. Subsequently, there are two common kinds of rotation: (Fabrigar et al., 1999; Beavers et al., 2013)

3.1.8. Orthogonal rotation

If rotation is orthogonal, it means all factors are uncorrelated with each other. So, a loading matrix is created, which is a matrix of correlations among observed variables and factors. Further going, the range of the loadings reflect the degree of relationship among observed variables and factors. ultimately, Orthogonal rotation has three methods (Varimax, quartimax, and Equamax):

3.1.9. Varimax rotation

Digging deep down to the Varimax rotation, which is a variance-maximizing process. Therefore, the Varimax way of orthogonal rotation was projected by Kaiser (1958). Its underlying principle is to supply axes with a small number loadings and as a big number of approximate zero loadings as potential. It is skilled by a maximum repeated of a quaternary function of the loadings. The aim of Varimax rotation is to make simpler by maximizing the variance of the loadings inside factors across variables. Thus, Varimax is simply the majority usually utilized of all the rotations available:

$$V = \frac{1}{P}\sum_{j=1}^{m}\left[\sum_{i=1}^{P}\boldsymbol{\ell}_{\mathbf{ij}}^{*\,\mathbf{4}} - \frac{\left(\sum_{i=1}^{P}\boldsymbol{\ell}_{\mathbf{ij}}^{*\,\mathbf{2}}\right)^{2}}{p}\right] \tag{3.21}$$

3.1.10. Oblique rotation

If rotation is oblique, it means the factors are correlated within themselves, and then a number of additional matrices are created. Furthermore, the factor correlation matrix involves the correlations between the factors. So, in oblique rotation, the loading matrix be the pattern matrix. When the values are squared in the pattern matrix are presented the exclusive involvement of the variables but it is not included segments of variance, which come overlap among correlated factors. On the other hand, if oblique rotation is utilized, the structure matrix (C), which is the correlations among both variables and factors. Therefore, the correlations evaluate the exclusive association among the variables and factors in the pattern matrix, plus the connection among the variable and the overlapping variance between the factors. some forms of oblique rotation have improved. (e.g., promax, quartamin, orthooblique) (Tabachnick et al., 2012; Marcoulides and Hershberger, 2012; Johnson and Wichern, 2002; Mardia et al., 1979; Härdle and Hlávka, 2007; Brown, 2006; Forina et al. 1989).

3.1.11. Number of significant factors

The number of factors has a number of criteria for selecting *m*:

1- Percent of Variance: Choose *m* to be the number of factors needed to achieve a preferred percentage, which is 80%, of the total variance tr(S) or tr(R).

2- Average Eigenvalue: select m equivalent to the number of eigenvalues superior than the common eigenvalue. For ***R*** the standard is 1; but for ***S*** it is $\sum_{j=1}^{p}\frac{\boldsymbol{\lambda}_j}{p}$

3- Scree Graph: it utilizes the scree check built on a plot of the eigenvalues of *S* or *R*. if the graph downs piercingly, which is followed by a direct line associated with a smaller slope. Thus, selecting m is equivalent to the number of eigenvalues before the direct line starts.

4- Significance Tests: is a test of the hypothesis, which is m is true number of factors, $H_0: \Sigma = LL^{'} + \boldsymbol{\Psi}$ where L is $p{\times}m$.

3.2. Cluster Analysis

Cluster analysis is a multivariate statistical method, which is a technique of categorization utilizing an evaluate of similarity or distance that has given for an random couple of objects. Thus, objects divided into groups, which are similarity among pair objects inside a group ought to be a big number and similarity among two groups ought to be small; by the way, a distance measure inside a group ought to be a small and distance among groups should be a big number (Miyamoto, 1990). 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 (Zakharov, 2016).

The target of a cluster analysis is to division an arranged data set or objects into clusters or its term such as (subsets, groups, classes). The mentioned division ought to have the coming properties: A- Homogeneity inside the clusters. It is a data, which is belong to the same cluster ought to be similarly possible. B- Heterogeneity among clusters means data, which belong to dissimilar cluster that ought to be differently possible (Höppner et al., 1999). 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 a number of groups that shapes a partition, or a family of partitions, of the collection objects (Miyamoto, 1990).

The methods of cluster analysis applied broadly to the collection in several eras, like medicine, psychiatry, sociology, criminology, anthropology, archaeology, geology, geography, remote sensing, market research, economics, and engineering.(Rencher and Christensen, 2002).

Cluster analysis could be defined in several ways, with depending on the objective of clustering. Usually, one of its definition 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 (Abonyi and Feil, 2007; Burns and Burns, 2008). Methods of cluster analysis can be divided into two categories of hierarchical clustering and nonhierarchical clustering.

3.2.1. Hierarchical clustering

Hierarchical techniques are between the long-established techniques of cluster analysis. Moreover, hierarchical clustering consists of successive aggregation or partition of the observations and their separations. Concluding form this sort of procedures, there is a form of tree structure, which is considered as a dendrogram (Wierzchoń and Kłopotek, 2018). 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 (Miyamoto, 1990; Everitt et al., 2011).

The agglomerative methods begin from the collection of observations, which is behaved as a divided cluster. In addition, clusters are combined in accordance with the lessening quantity of similarity (or the rising amount of dissimilarity) until one cluster is recognized (Wierzchoń and Kłopotek, 2018). Hierarchical classifications shaped by one of the cluster methods such as the agglomerative or divisive techniques, it may be may be characterized by a two-dimensional diagram, which is recognized a dendrogram that elucidates the fusions or separations , which created at each phase of the analysis.(Miyamoto, 1990). In a dendrogram, the resemblance of variables and groups of variables are possible to be recognized.

3.2.2. Dendrogram

The hierarchical structure is mostly demonstrated by dual-dimensional diagram; so it is named a tree diagram or dendrogram. Therefore, the dendrogram is a graphical demonstration of results of hierarchical cluster analysis. Besides, it emerges in the shape of tree as plot, where each pace of hierarchical clustering is demonstrated as a fusion of couple branches of the tree solution into a single one. By the way, the branches demonstrate clusters obtained on each pace of hierarchical clustering. (Meena et al., 2017).

Going further, 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 shows the order of the (n-1) links; the fork demonstrates 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 (Mardia et al., 1979). Additionally, the tree is mostly shown upside down. By the way, the root of the tree is located at the bottom and the branches are located at the top. On the other hand, when the trees are created by a computer, it is mostly suitable to print them out and the tree is on its side with the branches on the left (Chatfield and Collins, 1980; Everitt et al., 2011).

3.2.3. Nonhierarchical clustering

Nonhierarchical clustering has as point to begin, which is specification of the number clusters. Sometimes it is identified the number and the objects are assigned into clusters. So, it is a couple-phase process. Firstly, the cluster seed is particularized. Thus, it is a begin point, which can be illustrated by the researcher either a systematic or random choice. Further going, observations are assigned in accordance with its similarities to the pre-defined seed (Figueiredo Filho et al., 2014). 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 particularized in developed way or verified as part of the clustering procedure, because the dissimilarities of a matrix does not have to be determined and the fundamental data does not be stored throughout the computer run.

Nonhierarchical techniques begin from either (1) a first division of items into groups or (2) a primary collection of seed points that will shape the nuclei of clusters. God selections for beginning configurations ought to be liberated of obvious biases. One of the paths to begin is to choose randomly seed points from among the items into first group (Johnson and Wichern, 2002). 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 big 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 (Figueiredo Filho et al., 2014).

3.2.4. Select a clustering algorithm

There are some agglomerative procedures and they can be illustrious. By the way, they identify the distance from a recently formed cluster to a clear object, or to other clusters in the solution, so the most famous agglomerative clustering procedures involve the following:

Single linkage: It is a distance among a couple of clusters is equal to the distance among two nearest elements belonging to dissimilar cluster, by the way, In order to discover the optimum solution to the task, including the technique, which is specified the algorithms are used, referring to the smallest amount on both sides of tree.

Complete linkage: It is distance among a couple clusters, which is equivalent to the distance among a couple of extreme objects, belonging to dissimilar clusters. Therefore, it is a technique is mostly suitable, when the actual objects shapes separated well and compact clusters.

Average linkage: It is a distance among a couple of clusters is equivalent the standard distance among whole couple of objects, which belong to the both considered clusters.

Centroid method: In this approach, the statistical center of each cluster is computed first. Moreover, the distance among the couple clusters equals the distance among the two centroids.

Ward's method: In this technique, the amount of squares of distances among objects and the center of the cluster, to which the objects belong, is reduced. The technique, although considered to be too effectual, tends to shape clusters, which have the same low cardinalities (Wierzchoń and Kłopotek, 2018).

4. RESULTS AND DISCUSSION

This part includes an applied study depends on the information, which obtained through the distribution of questionnaire forms among students of the first stage in three different colleges of Salahaddin University for the academic year 2018-2019. Random stratified sample has been followed. For data analysis, SPSS program is used to determine the most important affective factors for students' admission and their interests in the specializations by using multivariate method, which focused on factor analysis (principal component method) and cluster analysis. Then, creating a new table to compare between each other in the results, also presents many statistical graphs and tables.

4.1. Profiles Of Respondents

Descriptive Statistics that is discussed in this work is the bar chart for eight variables. Thus, the height or length of the bar indicates the measured value or frequency, graph 4.1. to 4.8. illustrate the demographic respondent profiles of the students according to variables such as gender, age, college, admission method, the form was filled by, the department that I am studying in was due to, and Do I like the department that I am studying in.

The first figure below shows the gender for 350 students that are includes 122 males in the rate of 34.9% and 228 females in the rate of 65.1%.

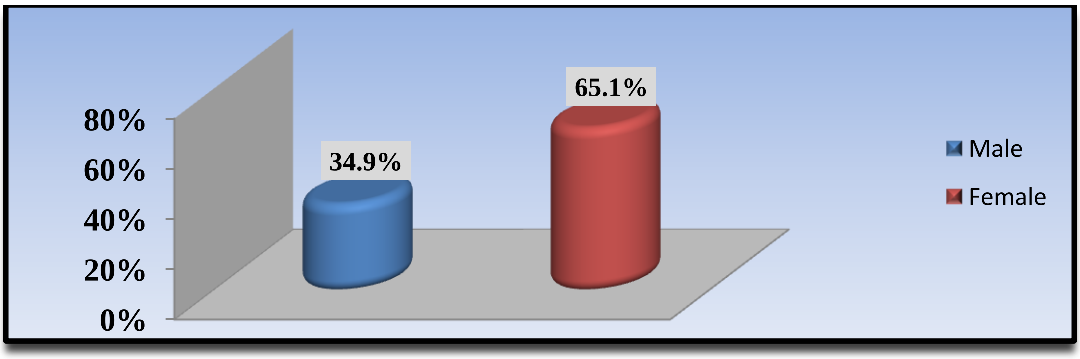

Figure 4. 1. Student's Gender.

Figure 4.2 shows age for the students, it is started in 17 years old, (65.14%) 228 students aged 17-19 years old, (33.71%) 118 students aged 20-22 years old, (.57%) 2 students aged 23-25 years old, and (.57%) 2 students aged equal or more than 26 years old.

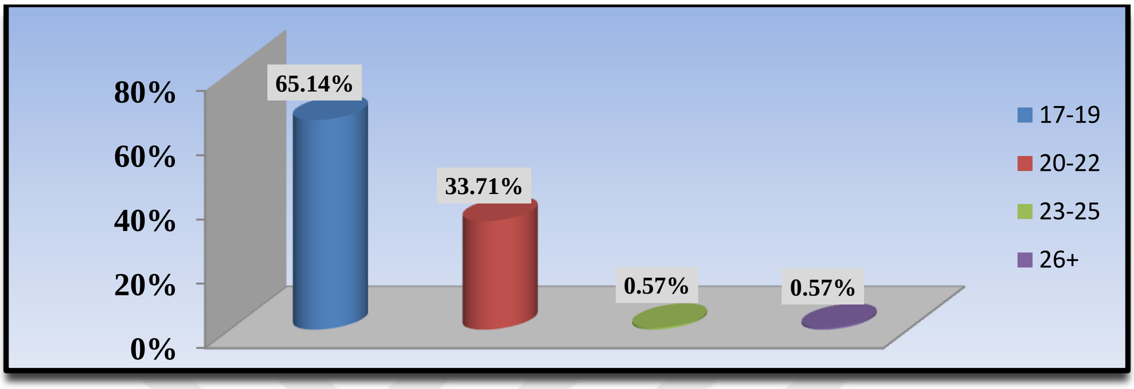

Figure 4. 2. Student's Age.

Figure 4.3. Shows the contribution of the students in the colleges that (71.7%) 251 students in Administrator and Economics college, (21.1%) 74 students in Science college, and (7.1%) 25 students in Shaqlawa Education college.

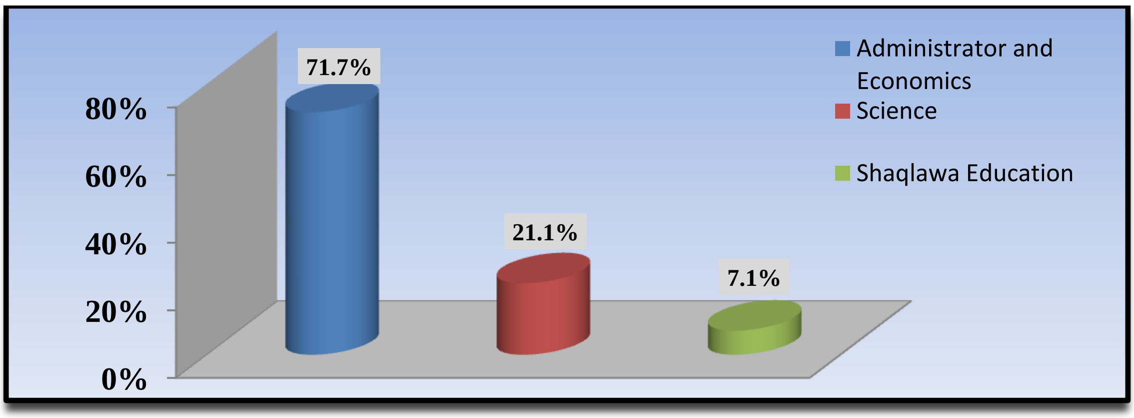

Figure 4. 3. Contribution of the students in the colleges.

Figure 4.4. shows the student departments, which is divided into ten different departments, (3.4%)12 students at Business Administration, (24.3%)85 students in Statistic, (5.4%) 19 students in Economics, (19.4%)68 students in Accounting, (6.9%)24 students in Financial and Banking, (12.3%)43 students in Tourism Organization

Administration, (13.7%)48 students in Mathematic, (7.4%)26 students in Chemistry, (4.3%)15 students in Arabic language, and (2.9%) 10 students in Physics.

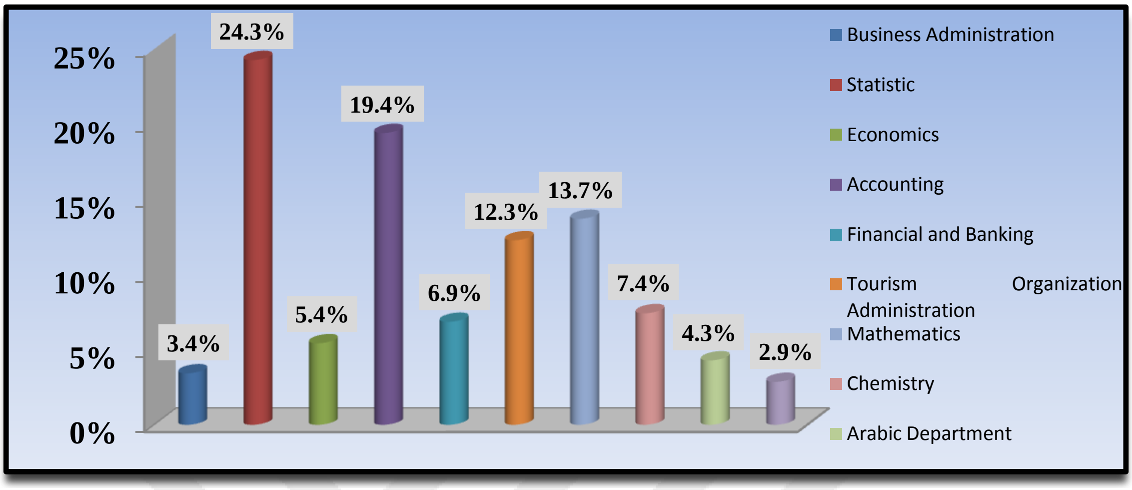

Figure 4. 4. Student departments.

Figure 4.5. Shows admission method, it includes three systems, (9.1%) 32 students filled form at Zancoline system, (84.9%) 297 students filled form at credit system, and (6.0%) 21 students filled form at parallel system.

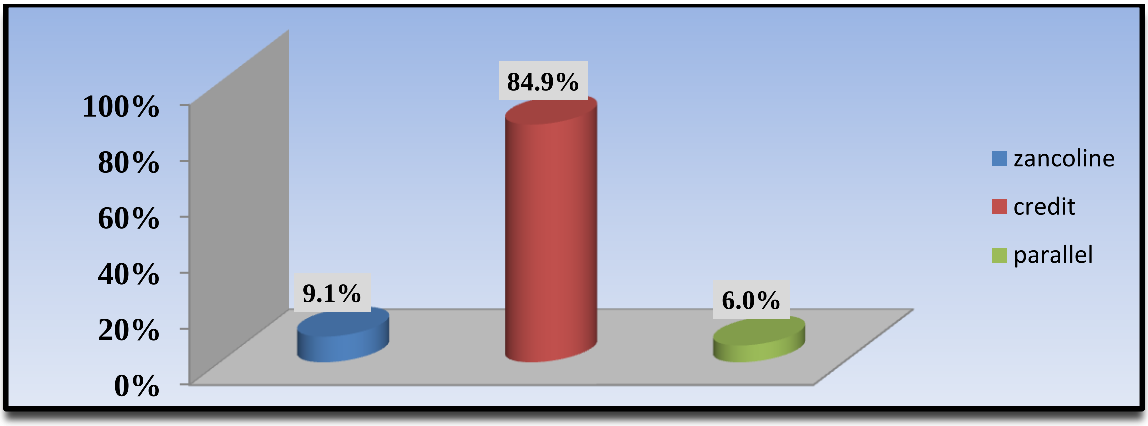

Figure 4. 5. Admission methods.

Figure 6.4. shows (The form was filled by), who does fill the admission form for the students, which are (26.3%) 92 students filled form by his/herself, (32.3%) 113 students filled form by a relative, and 145 (41.4%) students filled form by student centers.

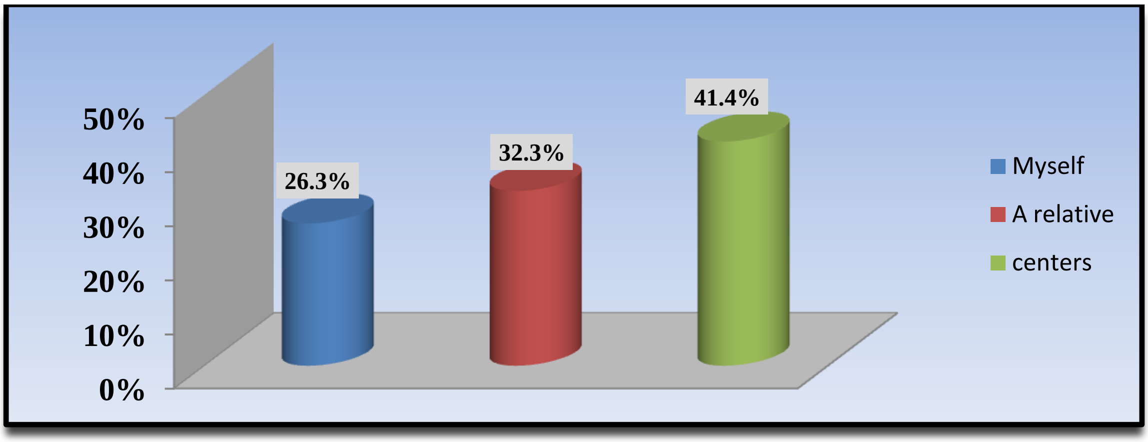

Figure 4. 6. The filled forms.

Figure 4.7. shows the department that the students studying in was due to, (25.7%) 90 students are studying in their departments that they were wanted, (6.0%) 21 students are studying in their department were recommended by their parents or relatives, and (68.3%) 239 students are studying in their department were admitted by the new system.

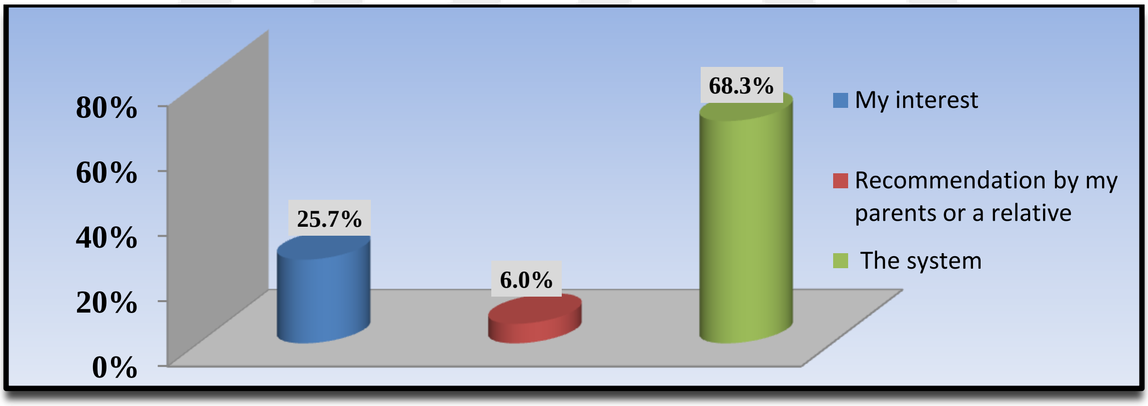

Figure 4. 7. The departments that have been chosen for students.

Figure 4.8. shows the variable (Do I like the department that I am studying in), which are (32.0%) 112 students have answered Yes, (26.3%) 92 students have answered No, and (41.7) 146 students have answered to some extent.

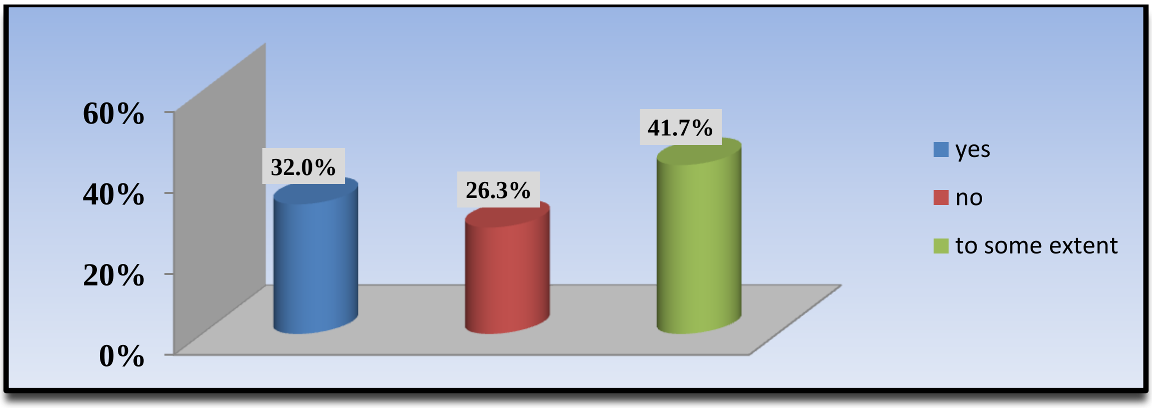

Figure 4. 8. The Students' desired departments.

4.2. Reliability Test

The Reliability is one of the most important and fundamental countenance in the evaluation of any measurement instrument or tool for a good research, for an exploratory or pilot study(Namdeo and Rout, 2016). George and Mallery (2016) 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 (Mohajan, 2017). Accordingly, in this study the reliability test is conducted by the means of Chronbach's alpha by using SPSS software before applying (factor and cluster) analysis, it has summarized the results in the following table.

Table 4. 1. Reliability Statistics

Cronbach's Alpha	N of Items
.70	39

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

4.3. Descriptive Statistics

Descriptive statistics for 39 variables are summarized in table (4.2.) For all variables the Mean values, standard deviation, count and percent for likert scale (strongly disagree, disagree, not with both, agree and strongly agree) shows below.

Table 4. 2. Descriptive statistics

	Strongly disagree		Disagree		Neutral		Agree		Strongly agree	
Variables	Count	%	Count	%	Count	%	Count	%	Count	%
V1	99	28.3%	83	23.7%	54	15.4%	56	16.0%	58	16.6%
V2	82	23.4%	119	34.0%	62	17.7%	38	10.9%	49	14.0%
V3	111	31.7%	115	32.9%	32	9.1%	31	8.9%	61	17.4%
V4	80	22.9%	105	30.0%	71	20.3%	41	11.7%	53	15.1%
V5	91	26.0%	103	29.4%	37	10.6%	50	14.3%	69	19.7%
V6	76	21.7%	67	19.1%	59	16.9%	58	16.6%	90	25.7%
V7	53	15.1%	40	11.4%	59	16.9%	65	18.6%	133	38.0%
V8	60	17.1%	48	13.7%	36	10.3%	63	18.0%	143	40.9%
V9	53	15.1%	44	12.6%	52	14.9%	88	25.1%	113	32.3%
V10	58	16.6%	28	8.0%	62	17.7%	74	21.1%	128	36.6%
V11	72	20.6%	47	13.4%	62	17.7%	42	12.0%	127	36.3%
V12	83	23.7%	62	17.7%	49	14.0%	61	17.4%	95	27.1%
V13	64	18.3%	57	16.3%	54	15.4%	50	14.3%	125	35.7%
V14	83	23.7%	50	14.3%	56	16.0%	59	16.9%	102	29.1%
V15	46	13.1%	39	11.1%	89	25.4%	57	16.3%	119	34.0%
V16	45	12.9%	32	9.1%	73	20.9%	57	16.3%	143	40.9%
V17	42	12.0%	24	6.9%	61	17.4%	61	17.4%	162	46.3%
V18	87	24.9%	60	17.1%	78	22.3%	34	9.7%	91	26.0%
V19	124	35.4%	73	20.9%	37	10.6%	47	13.4%	69	19.7%
V20	78	22.3%	86	24.6%	45	12.9%	70	20.0%	71	20.3%
V21	49	14.0%	31	8.9%	36	10.3%	84	24.0%	150	42.9%
V22	67	19.1%	34	9.7%	32	9.1%	51	14.6%	166	47.4%
V23	97	27.7%	38	10.9%	58	16.6%	42	12.0%	115	32.9%
V24	43	12.3%	22	6.3%	100	28.6%	49	14.0%	136	38.9%
V25	31	8.9%	17	4.9%	26	7.4%	54	15.4%	222	63.4%
V26	59	16.9%	31	8.9%	43	12.3%	52	14.9%	164	47.0%
V27	130	37.1%	41	11.7%	37	10.6%	46	13.1%	96	27.4%
V28	74	21.1%	36	10.3%	69	19.7%	56	16.0%	115	32.9%
V29	92	26.3%	54	15.4%	74	21.1%	33	9.4%	97	27.7%
V30	38	10.9%	23	6.6%	75	21.4%	59	16.9%	155	44.3%
V31	37	10.6%	18	5.1%	36	10.3%	59	16.9%	200	57.1%
V32	45	12.9%	31	8.9%	39	11.1%	54	15.4%	181	51.7%
V33	86	24.6%	29	8.3%	112	32.0%	43	12.3%	80	22.9%
V34	35	10.0%	20	5.7%	69	19.7%	58	16.6%	168	48.0%
V35	31	8.9%	15	4.3%	57	16.3%	62	17.7%	185	52.9%
V36	67	19.1%	23	6.6%	61	17.4%	68	19.4%	131	37.4%
V37	30	8.6%	16	4.6%	39	11.1%	51	14.6%	214	61.1%
V38	50	14.3%	24	6.9%	92	26.3%	59	16.9%	125	35.7%
V39	29	8.3%	34	9.7%	58	16.6%	56	16.0%	173	49.4%

Table 4.2. Descriptive statistics(continued)

variables	mean	median	S.D	Rate of agreement(%)
V1	2.69	2	1.449	53.771
V2	2.58	2	1.332	51.600
V3	2.47	2	1.453	49.486
V4	2.66	2	1.352	53.257
V5	2.72	2	1.482	54.457
V6	3.05	3	1.503	61.086
V7	3.53	4	1.467	70.571
V8	3.52	4	1.542	70.343
V9	3.47	4	1.435	69.371
V10	3.53	4	1.463	70.629
V11	3.30	3	1.564	66.000
V12	3.07	3	1.545	61.314
V13	3.33	3.50	1.538	66.571
V14	3.13	3	1.554	62.686
V15	3.47	4	1.395	69.371
V16	3.63	4	1.418	72.629
V17	3.79	4	1.398	75.829
V18	2.95	3	1.519	58.971
V19	2.61	2	1.551	52.229
V20	2.91	3	1.465	58.286
V21	3.73	4	1.442	74.571
V22	3.61	4	1.592	72.286
V23	3.11	3	1.627	62.286
V24	3.61	4	1.372	72.171
V25	4.20	5	1.291	83.943
V26	3.66	4	1.537	73.238
V27	2.82	3	1.675	56.400
V28	3.29	3	1.531	65.829
V29	2.97	3	1.554	59.371
V30	3.77	4	1.360	75.429
V31	4.05	5	1.354	80.971
V32	3.84	5	1.456	76.857
V33	3.01	3	1.452	60.114
V34	3.87	4	1.339	77.371
V35	4.01	5	1.290	80.286
V36	3.49	4	1.512	69.886
V37	4.15	5	1.288	83.029
V38	3.53	4	1.401	70.571
V39	3.89	4	1.337	77.714
Total	3.36	4	1.457	67.199

Through a table (4.2.) notes generaly that all the answers to the questionnaire questions have agreement on all paragraph variables, where the rate of the total agreement is (%67.199), its average agreement reached (3.36), and its standard deviation (1.457). Besides, the variable 25 (It is better if counting of the average of grade 10 and 11 be optional) has the highest reached agreement ratio, which is %83.943. Then, the variable 37 (The changes and guides must be given to all of the high schools before starting a new academic year) has a high agreement in the rate of %83.029, these two variables are more important for the students.

4.4. Factor Analysis

Factor analysis has been used to construct the new affective factors for the student admission and their interests in the specialization. Moreover, the goal of factor analysis is to reduce the redundancy among the variables by using a smaller number of factors. Principal component methods are used to analyze the correlation matrix to show the significance of each variable on the basis of the relationship between the variables.

A correlation matrix should be used in the process of factor analysis it is displaying the correlation or relationships between a single variable and every other variables in the investigation. So, the table appendix1 is a Correlation table has presented the correlation matrix for 39 variables.

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 that there is a strong association between all or most of the questions, not only among a few of them. 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 4. 3. 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 enough questions are predicted by each factor. Therefore, for these data, the value is 0.713, which is acceptable and is good. Furthermore, there is a test , which is colled Bartlett test of sphericity ought to be significant (i.e., a significance value of less than .05); it is meant that the questions are correlated extremely adequate to give a logical basis for factor analysis. In addition, in this case (p-value is less than .05, demonstrating that the correlation matrix is significantly unlike from an identity matrix), which shows that the data is suitable for factor analysis.

Table 4. 4. 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		

Extraction Method: Principal Component Analysis.

The communalities table shows the initial commonalities before rotation, table 4.4. is table of communalities, which shows how much of the variance in the variables has been accounted for the extracted factors.

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

The Total Variance Explained table shows how the variance is divided among the 39 possible factors.

It is noted that the tables has shown the Eigen value of the fifteen and sixteen Components, which is less than one, then the first sixteen Components depend on the interpretation variance ratio. In the extraction window has not based on Eigen value, it is based on fixed number of the factors that divided into sixteen factors, in terms of the number of variable's worth of each one of the variance explanations. So, first Components shows almost, Not: that 7.818% of the variance is explained by the first component after rotation, (as much variance as in five variables). 6.683% of the variance is explained by the second component, 3.849% of the variance is explained by the third component, 3.828 % of the variance is explained by the fourth component, 3.789% of the variance is explained by the fifth component, 3.707% of the variance is explained by the sixth component, 3.672% of the variance is explained by the seventh component, 3.559% of the variance is explained by the eighth component, 3.405% of the variance is explained by the ninth component, 3.299% of the variance is explained by the tenth component, 3.285% of the variance is explained by the eleventh component, 3.161% of the variance is explained by the twelfth component, 3.127% of the variance is explained by the thirteenth component, 3.120% of the variance is explained by the fourteenth component, 2.967% of the variance is explained by the fifteenth component, and 2.888% of the variance is explained by the sixteenth component. Thus, when the sixteen components depend on the interpretation variance ratio, the component explains less information than a single variable would have explained. The following figure illustrates this.

Table 4. 5. Total variance

	Initial Eigen values			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
17	.916	2.348	64.505			
18	.900	2.308	66.813			
19	.887	2.275	69.088			
20	.839	2.150	71.238			
21	.808	2.073	73.311			
22	.784	2.011	75.322			
23	.774	1.985	77.308			
24	.736	1.886	79.194			
25	.710	1.820	81.014			
26	.698	1.790	82.804			
27	.683	1.752	84.556			
28	.662	1.698	86.254			
29	.615	1.578	87.832			
30	.579	1.486	89.318			
31	.567	1.454	90.772			
32	.530	1.358	92.130			
33	.521	1.337	93.467			
34	.520	1.333	94.800			
35	.477	1.224	96.023			
36	.446	1.143	97.167			
37	.397	1.018	98.185			
38	.377	.966	99.151			
39	.331	.849	100.000			

Extraction method: Principal component analysis.

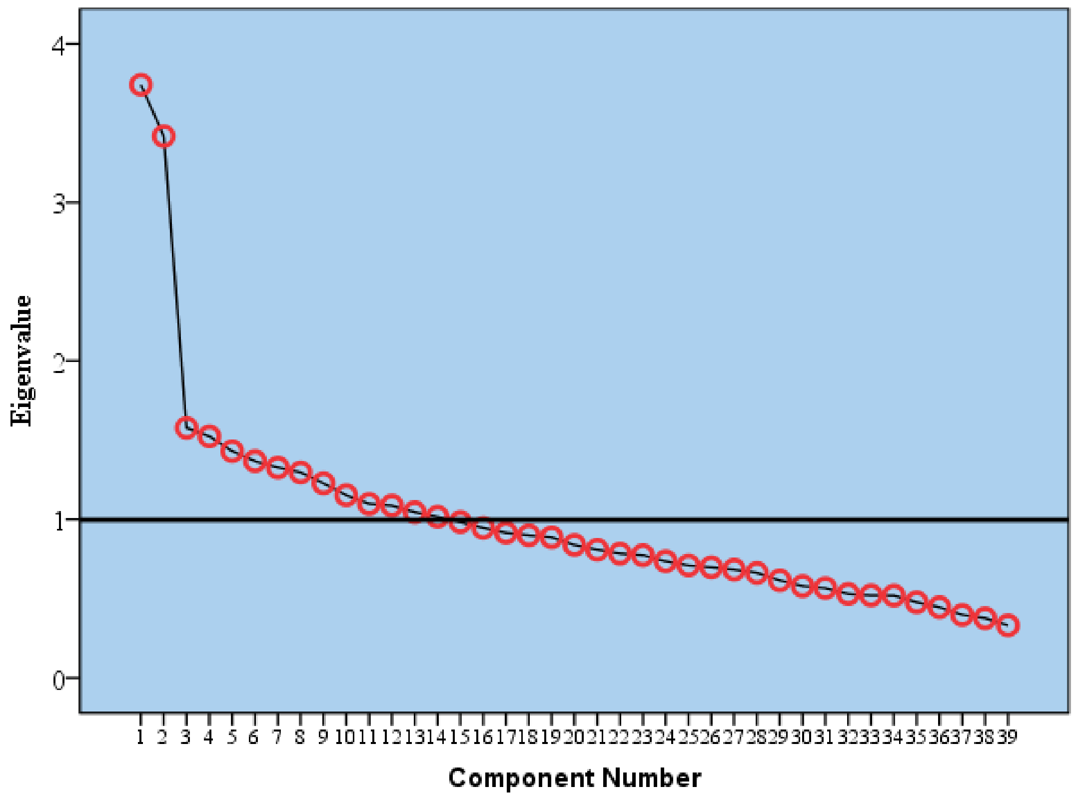

Figure 4. 9. Scree plot.

The Scree plot demonstrates, which after the first sixteen components are different with the Eigen values decline (the curve flattens) and they are less than (0.946) or dependence 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 were able to be reduced to sixteen components. It notes that the Scree plot flattens out after the sixteen components.

Cumulative Percent of variance between variables are accounted by each component, before and after rotation 62.157% from the variance is accounted by the first sixteen component. So, those components 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, and each component explains more than one variable expect component 9, component 12, component 13, component14, component 15 and component 16 explain one variable, this is a condition, which is named simple structure. So, it is the aim of rotation. Actually, it is not usually gained. in the Rotated Component of factor loadings matrix, there is one thing to look for, which is an extent to which simple structure is gained. It summarizes the results in the following table:

Table 4. 6. Rotated Component Matrix[a]

	Components															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
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											
V38						.659										
V25						.528										
V28						.507										
V36							.749									
V39							.544									
V26								.675								
V11								.668								
V14									.724							
V29										.755						
V30										.508						
V15											.638					
V33											.604					
V20												.810				
V23													.778			
V12														.760		
V9															.835	
V13																.81

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 21 iterations.

Note: in table (4.6.) that the analysis has sorted the (39) variables with size and loading more than absolute value (±0.5) (V1 to V39) into (16) somewhat overlapping groups of variables. Then, the variables are sorted because of the variable that has the highest loading (not considering whether the correlation is positive or negative). Actually, every variable has some loading from every component, but we requested for 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 component, which explains the largest variance ratio containing five variables they are (V3: It helps the students completely be accepted in college and the specialization that he/she likes.V2: The students can trust this system. V1: I prefer the system. V4: It has a good effect on the students' 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 Admission of students in master and PhD degrees must be done by using the new system.V35: I suggest that daily the students be aware of the results when the competition is done in the system by using his/her account. V37: The changes and guides must be given to all of the high schools before starting a new academic year.V32: I like all students be accepted at the same time.), with loadings (.667, .608, .527, .504). The third component contains two variables respectively (V22: My low marks didn't let me to 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 parallel system the graduated students must not be able to refill the admission forms. V16: I like the number of students in parallel system be decreased), with loadings (.716, .624). The fifth component contains two variables respectively (V7: It doesn't account for economic status of the student. V10: Employing the geographic area dependent method affects choosing colleges and specializations), with loadings (.724, .620). The sixth component contains three variables they are (V38: The number of the colleges and departments that accept only students of literary part is few. V25: It is better if counting of the average of grade 10 and 11 be optional. V28: I prefer the students that have one remained subject to be accepted 'transit'), with loadings (.569, .528, .507). the seventh component consists of (variable 36 and variable 39), the eighth component consists of (variable 26 and variable 11), the ninth component contains the (variable 14), the tenth component consists of (variable 29 and variable 30), the eleventh component consists of (variable 15 and variable 33), the twelfth component consists of (variable 20), the thirteenth component consists of (variable 23), the fourteenth component consists of (variable 12), the fifteenth component consists of (variable 9), the sixteenth component consists of (variable 13), the loadings resulting from an orthogonal rotation are correlation coefficients between each variable and the factor, so they range from −1.0 through 0 to + 1.0.

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 on the Rotated Factor Matrix will be low (blank or less than .50).

4.5. Cluster Analysis

Cluster analysis has been applied in this study. It is one of the most important statistical methods that are used to classified variables into homogeneous groups. Further, it depends on the differences and similarities between the data. Using average linkage (between groups) and rescaled distance cluster combine to find the distance matrix and relationship between clusters; so dividing the number of clusters 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 4. 7. 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 was classified into five clusters of the thirty nine variables are Analyzed by cluster analyses, the first cluster includes ten variables in the rate of 25.656%, the second cluster includes 22 variables for a large percentage 56.410%, the third cluster includes 4 variables in the rate of 10.256%, the fourth cluster includes 2 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. The Figure 4.10. displays dendrogram by using the average ward linkage will be chosen the number of clusters into five discrete clusters, which based on 39 variables, where shows the variables and distance between them.

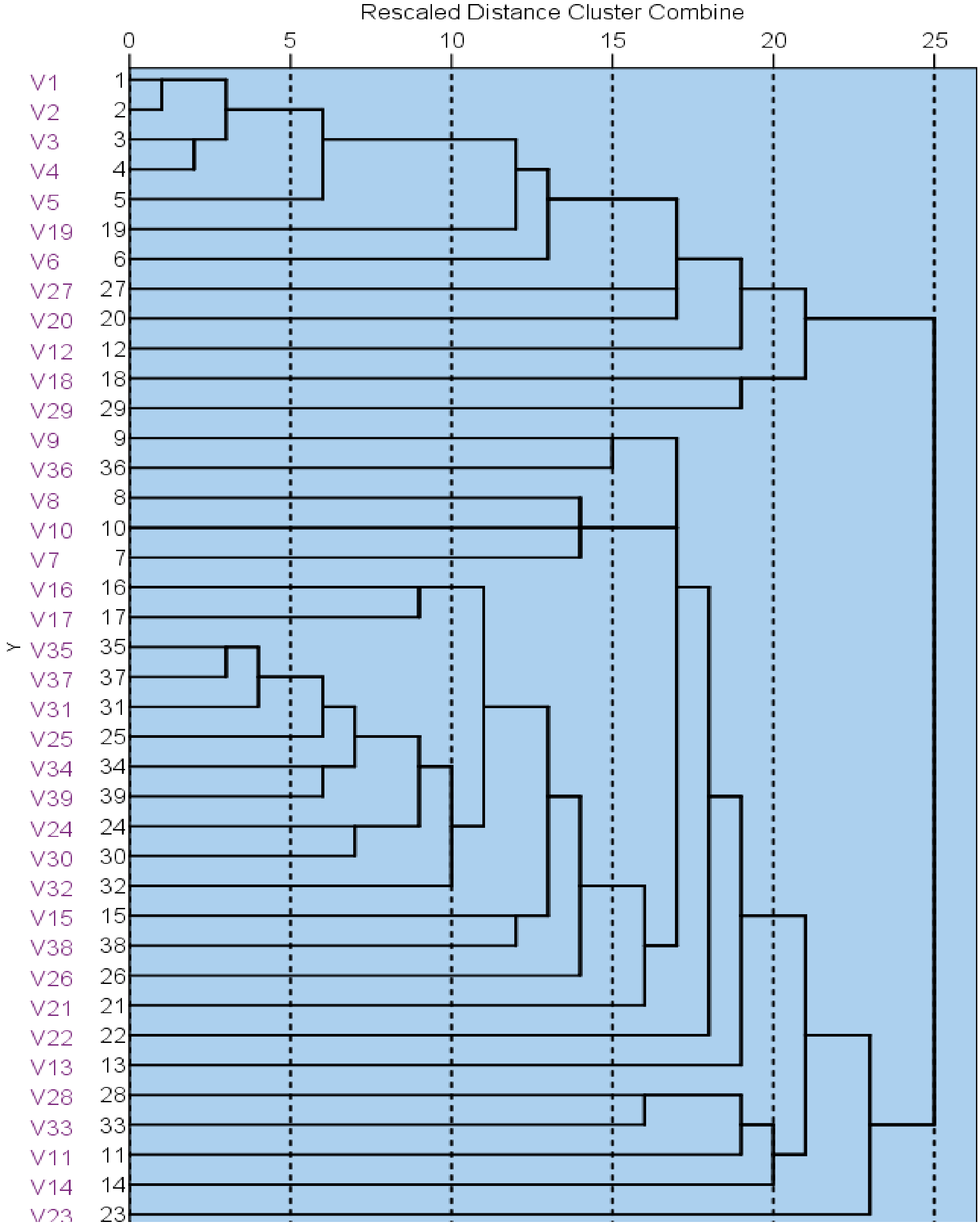

Figure 4. 10. Dendrogram using average linkage (between Groups).

4.6. Comparison between Results of the Factor Analysis and the Cluster Analysis

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

Table 4. 8. 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 completely to be accepted in colleges and the specializations that he/she likes. V4: It has a good effect on the students' future. V5: It facilitates for the students to choose the department that he/she wants. The second factor and the second cluster have four variables in common, which are V24: I prefer the Admission of students in master 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 his/her account). V37: The changes and guides must be given to all of the high schools before starting a new academic year. And the fourth row has one variable in common, which is V18: In parallel system, the graduated students must not be allowed to refill admission forms.

5. CONCLUSIONS

The purpose of this thesis is to find the most important factors affecting student admission and their interests in the specialization by using multivariate method, the main conclusions can be summarized based on the results has been obtained.

The answers to the questionnaire items have agreement on all variable paragraphs where the rate of total agreements are %67.199. The variable 25th (It is better if counting of the average grade of 10th and 11th be optional) has the highest agreement ratio has reached %83.943 and variable 37th (The changes and guides must be given to all of the 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: first component is more influential, it represents the five variables (the system is a helpful for student's admission in the 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 and PhD students to be admitted at the colleges and get the competitive results by utilizing their accounts)and the total effect of this component is % 6.683. Moreover, the third component (the student's desired professions need a higher mark to be admitted in the departments), the total effect of this component is % 3.849. The fourth component (decreasing the number of student's admission in parallel system by using the graduated students must not be able to refill admission forms) its rate interpretation is %3.828 and the fifth component (geographic area and economic status), its rate interpretation is %3.789.

The Cluster analysis presents the numbers of the clusters have been used in the analysis are five clusters. Therefore, the 39 variables have been obtained were divided into these five clusters. Therefore, the first cluster (the system is a helpful for student's admission in the college to get their desired professions) involves 10 variables in the rate of %25.656. The second cluster (the new system may help the master and PhD students to be admitted at the colleges and get the competitive results by utilizing their accounts) includes 22 variables of a large percentage %56.410. The third cluster (the chance of the admission will be the same in both private and government universities with filling a form that suitable for student's ability) involves four variables in the rate %10.256. In addition, the fourth cluster (in parallel system the graduated and employee students are not able to refill the admission form), which includes two variables in the rate of %5.128. The last cluster includes one variable by the rate %2.564, which is (I like the order and the choices will be decreased to 25 choices).

There is a big rapprochement and similarity between factor analysis and cluster analysis, in the way that 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 important and influential variables for the student's admission and their interests in the specializations. Consequently, in the first, five variables are in common: V1, V2, V3, V4, V5, in the second, four variables are in common: V24, V32, V35, V37 and in the third, one is not in common, but in the fourth, there is one variable in common, which is the V18.

REFERENCES

Alotaibi, A., Ayesh, A., Hall, R., 2016. Managing admission in Saudi universities: A system approach. ***International Journal of Information and Education Technology,*** **6**(4): 314-321.

Abonyi, J., Feil, B., 2007. ***Cluster Analysis For Data Mining and System Identification***. Springer science and business media.

Burns, R., Burns, R., 2008. ***Business Research Methods and Statistics Using SPSS.*** London, Thousand oaks sage publications.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., Esquivel, S. L., 2013. Practical considerations for using exploratory factor analysis in educational research. ***Practical Assessment Research and Evaluation,*** **18**(6) :1-13.

Brown, T. A., 2006. ***Confirmatory Factor Analysis For Applied Research.*** Guilford Publications.

Basilevsky, A. T., 2009. ***Statistical factor analysis and related methods: Theory and Applications.*** Wiley-Interscience Publication.

Braun, S., Dwenger, N., 2009. Success in the university admission process in Germany: regional provenance matters. ***Higher Education,*** **58**(1) : 71-80.

Biró, P., 2008. Student Admissions in Hungary As Gale and Shapley Envisaged. ***University of Glasgow technical report tr,*** **29**:1-7.

Beswick, R. L., 1989. ***A Study of Factors Associated With Student Choice in The University Selection Process.*** Doctoral dissertation, Lethbridge, AB: University of Lethbridge, faculty of education.

Cattell, R. B., 1965. A biometrics invited paper. Factor analysis: An introduction to essentials I. The purpose and underlying models. ***Biometrics,*** **21**(2) : 405-435.

Comrey, A., & Lee, H. (1992). ***A First Course in Factor Analysis. 2nd edn.*** Lawrence earlbaum associates. *Publishers: Hillsdale, new jersey*.

Chapman, D. W., 1981. A model of student college choice. ***The Journal of Higher Education,*** **52**(5): 490-505.

Chatfield, C., Collins, A. J., 1980. ***Introduction To Multivariate Analysis.*** First edition , originally published by chapman and hall, Boca raton London New York Washington, D .C.

Everitt, B. S., Landau, S., Leese, M., Stahl, D., 2011. ***Cluster Analysis. 5th Edition***. Wiley series in probability and statistics.

Figueiredo Filho, D. B., Da Rocha, E. C., Da Silva Júnior, J. A., Paranhos, R., Da Silva, M. B., Duarte, B. S. F., 2014. Cluster analysis for political scientists. ***Applied Mathematics,*** **5**(15): 2408-2415.

Field, A., 2005. ***Discovering Statistic Using SPSS: and Sex Drugs and Rock 'n' Roll***. 2nd edn. Sage publication, London.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., Strahan, E. J., 1999. Evaluating the use of exploratory factor analysis in psychological research. ***Psychological Methods,*** **4**(3): 272-299.

Forina, M., Armanino, C., Lanteri, S., Leardi, R., 1989. Methods of varimax rotation in factor analysis with applications in clinical and food chemistry. ***Journal of Chemometrics*, 3(S1)**:115-125.

Gulluce, A. C., Yilmaz, T., Kaygin, E., 2016. Factors affecting the university preferences of students. ***American Journal of Industrial and Business Management*, 6**(03): 357-372.

George, D., Mallery, P., 2016. ***IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference***. 4th edn, Routledge .

Höppner., F., Klawonn, F., Kruse, R., Runkler, T., 1999. ***Fuzzy Cluster Analysis: Methods For Classification, Data Analysis and İmage Recognition***. John Wiley and Sons, LTD.

Harris, R. J., 2001. ***A Primer of Multivariate Statistics***. Third edition, lawrence erlbaum assocıates, publıshers.

Härdle, W. K., Hlávka, Z., 2007. ***Multivariate Statistics: Exercises and Solutions***. Springer

Harman, G., 1994. Student selection and admission to higher education: policies and practices in the asian region. ***Higher Education*, 27**(3): 313-339.

Hassan, S., Ismail, N., Jaafar, W. Y. W., Ghazali, K., Budin, K., Gabda, D., Samad, A. S. A., 2012. Using factor analysis on survey study of factors affecting students' learning styles. ***International Journal of Applied Mathematics and Informatics*, 1**(6): 33-40.

Ellis, J. L., 2017. Factor analysis and item analysis. ***https://www.applyingstatisticsinbehaviouralresearch.com/documenten/factor_analysis_and_item_analysis_version_11_.pdfhttps://docplayer.net/12782167-Chapter-9-cluster-analysis.html.***

İlgan, A., Ataman, O., UğurlU, F., Yurdunkulu, A., 2018. Factors affecting university choice: a study on university freshman students. ***The Journal of Buca Faculty of Education*,** (46):199-216.

Johnson, R. A., Wichern, D. W., 2002. ***Applied multivariate statistical analysis.*** 5th Edition, Upper Saddle River, N J: Prentice hall.

Kumar, R., Desale, P., Puri, C. G., 2013. Web based admission system. ***International Journal of Emerging Technology and Advanced Engineering*, 3**(7): 634-639.

Kaur, A., Hasija, S., 2015. A conceptual model of admission system and performance evaluation for a university. ***International Journal of Computer Applications*, 125**(4): 29-33.

Khoo, C. S., Higgins, S. E., Foo, S., Lim, S. P., 2004. A cluster analysis of LIS students in Singapore and implications for defining areas of specialization. ***Journal Of Education For Library and İnformation Science*, 45**(1): 36-57.

Kallio, R. E., 1995. Factors influencing the college choice decisions of graduate students. ***Research in Higher Education*, 36**(1): 109-124.

Kline, P., 1994. ***An Easy Guide To Factor Analysis*.** Routledge.

Kenny, D. A., Kashy, D. A., Cook, W. L., 2006. ***Methodology in The Social Sciences*** David A. Kenny, series editor, Guılford press.

Krishnan, V., 2011. ***A Comparison Of Principal Components Analysis and Factor Analysis For Uncovering The Early Development İnstrument (EDI) Domains.*** Community-university partnership (CUP), faculty of extension, university of Alberta, edmonton, Alberta, Canada.

Mincer-Daszkiewicz, J., 2004. ***Student Admission System For Warsaw University***. In eunıs’2004, the 10th International conference of european university information systems.

Mardia, K. V., Kent, J. T., Bibby, J. M., 1979. ***Multivariate Analysis .*** Academic press inc. London, LTD.

Marcoulides, G. A., Hershberger, S. L., 2012. ***Multivariate Statistical Methods: A First Course.*** Psychology press.

Miyamoto, S., 1990. ***Fuzzy sets in information retrieval and cluster analysis.*** Springer science and business media.

Meena, L. K., Sen, C., Kushwaha, S., 2017. Cluster analysis to form similarity for major selected crops in rajasthan, India. ***Int. J. Curr. Microbiol. App. Sci,*** **6**(4): 2673-2682.

Mohajan, H. K., 2017. Two criteria for good measurements in research: Validity and reliability. ***Annals of Spiru Haret University. Economic Series,*** **17**(3): 58-82.

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.

Namdeo, S. K., 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.

Rencher, A. C., 1998. ***Multivariate Statistical İnference and Applications.*** John Wiley and Sons, Inc.

Rencher, A. C., Christensen, W. F., 2002. ***Methods of Multivariate Analysis.*** Second edition, a John Wiley and Sons. INC.: Toronto, on, Canada.

Reyment, R. A., Jvreskog, K. G., 1996. ***Applied Factor Analysis in The Natural Sciences.*** Cambridge university press.

Singh, G., Kumar, M., 2014. Exploratory factor analysis of service quality dimensions for higher educational institutes: A students perspective. ***Global Journal of Management and Business Research,*** **14** :39-48.

Stevens, J. P., 2002. ***Applied Multivariate Statistics For The Social Sciences***. Fourth edition, lawrence erlbaum assocıates.

Tabachnick, B. G., Fidell, L. S., Ullman, J. B., 2012. ***Using Multivariate Statistics*** (sixth edition). Pearson .

Tinsley, H. E., Brown, S. D., 2000. ***Handbook Of Applied Multivariate Statistics and Mathematical Modeling.*** Elsevier science and technology books.

Tabachnick, B. G., Fidell, L. S., Ullman, J. B., 2007. ***Using multivariate statistics***.5th Edition, Pearson education, Inc.

Wakil, K., Akram, B., Kamal, N., Safi, A., 2014. Web recommender system for private universities' admission in Iraq: UHD case study. ***International Journal of e-Education, e-Business, e-Management and e-Learning,*** **4**(5):329-340.

Wierzchoń, S. T., Kłopotek, M., 2018. ***Modern algorithms of cluster analysis.*** (p. 421). Berlin, Germany:Springer International Publishing.

Williams, B., Onsman, A., Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. ***Australasian Journal of Paramedicine,*** **8**(3):1-13.

Yong, A. G., Pearce, S., 2013. A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. ***Tutorials in Quantitative Methods For Psychology,*** **9**(2): 79-94.

Zakharov, K., 2016. Application of k-means clustering in psychological studies. ***Tutorials in Quantitative Methods for Psychology,*** **12**(2): 87-100.

APPENDIX

APPENDIX1. Correlation matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
V1	1																			
V2	.487	1																		
V3	.400	.482	1																	
V4	.364	.359	.448	1																
V5	.324	.347	.440	.338	1															
V6	.170	.287	.252	.243	.294	1														
V7	.008	-.105	.024	.038	.051	.022	1													
V8	-.082	-.084	-.094	-.088	-.082	-.028	.160	1												
V9	-.011	-.089	-.080	-.078	.005	-.059	.076	.101	1											
V10	-.048	-.093	-.053	-.082	.013	-.012	.151	.199	.087	1										
V11	.085	.106	.084	.075	.076	.031	.104	.021	-.011	.024	1									
V12	.100	.124	.194	.079	.074	.159	.044	.022	-.074	-.084	.040	1								
V13	-.035	.035	-.090	-.124	-.055	-.016	.020	.116	.079	.030	.054	-.042	1							
V14	.120	.015	-.007	.024	.078	.063	-.006	-.011	.069	.040	.085	-.085	.052	1						
V15	.037	-.062	-.019	-.131	-.014	-.052	.006	.099	.021	.062	.026	-.032	.070	.112	1					
V16	-.089	-.057	-.006	-.024	.027	-.057	.057	.192	.063	.083	.000	-.010	-.013	-.022	.156	1				
V17	-.042	.034	.016	.025	.027	.014	.016	.044	.080	.050	.006	.004	.037	-.016	.088	.231	1			
V18	-.016	.070	.071	.020	.012	-.013	.048	.096	-.016	-.074	.020	.054	.007	-.053	.077	.210	.111	1		
V19	.269	.148	.263	.197	.287	.133	.056	-.198	-.050	-.029	.103	.106	-.052	.095	.020	-.029	-.015	.036	1	
V20	.041	.086	.053	-.001	.076	.043	-.042	-.076	.002	.114	-.070	-.057	.034	.068	-.005	-.067	.027	-.051	.157	1
V21	.230	.225	.161	.218	.205	.192	-.017	-.045	-.064	-.042	-.023	.121	-.060	.085	-.048	-.032	.003	.057	.247	.083
V22	.052	.105	-.007	-.034	-.048	.048	.048	.199	.036	.098	-.006	-.025	.101	-.051	.115	.148	.143	.070	-.156	.012
V23	-.037	-.042	.012	.034	-.058	.011	-.005	.105	.042	.040	-.017	-.029	.086	-.025	.054	.038	.007	.008	-.030	.106
V24	.057	.109	.126	.019	.085	.077	.036	.071	.015	-.038	.031	.028	.107	.041	.193	.039	.234	.040	.055	-.010
V25	-.083	-.010	.028	-.171	-.030	-.039	-.049	.071	.031	.083	.052	-.011	-.019	-.049	.041	.232	.083	.104	-.046	-.059
V26	.003	.039	.006	-.027	.002	-.030	.027	.120	-.032	.048	.153	.040	.021	.011	.039	.053	.128	-.020	.043	-.013
V27	.251	.198	.187	.162	.199	.210	-.044	.036	.030	-.041	.070	.049	-.063	.152	.011	-.018	.038	.112	.166	.102
V28	.052	.044	.047	-.048	.080	.031	.041	.047	.025	-.035	.049	.074	.053	-.014	.027	-.018	-.018	.048	-.024	-.089
V29	-.003	-.038	-.019	.043	-.026	-.010	.045	.053	.094	-.039	-.116	-.023	.027	-.022	-.016	.014	-.015	.085	.072	.038
V30	.033	-.036	-.082	-.045	-.008	.035	.102	.095	-.053	.144	.015	-.002	.094	.004	.076	.118	.092	.095	-.066	-.008
V31	-.027	-.028	-.015	.000	.076	-.014	.087	.093	-.016	.128	.039	.070	.026	.048	.186	.106	.123	.000	-.053	-.007
V32	-.004	-.034	.046	.002	.010	.005	-.011	-.013	.034	.057	-.054	.038	.082	-.061	.104	.051	.160	-.033	.060	.086
V33	.058	.056	.041	.008	.106	.024	-.062	.074	.001	.035	.141	.050	.065	.118	.136	.025	.035	-.017	.049	.054
V34	.019	.009	.036	.014	.047	.081	.100	.061	-.038	.201	.060	.010	.050	-.016	.095	.191	.211	-.002	.002	.011
V35	-.038	.056	.050	-.055	-.010	.040	-.011	.120	-.001	.021	-.011	.050	.060	-.013	.116	.182	.205	.088	.085	.018
V36	-.036	-.086	-.065	-.129	-.046	-.044	.077	-.033	.137	.129	-.079	-.026	.114	-.041	.038	.039	.074	.129	-.057	-.038
V37	.026	.036	-.055	-.085	.008	.055	.095	.100	-.035	.171	-.075	-.002	.089	.001	.146	.168	.158	.065	-.013	-.013
V38	-.046	-.020	-.071	-.045	-.022	.001	-.008	.183	-.078	.088	-.023	.008	.045	.000	.128	.081	.040	.030	.087	-.011
V39	.040	-.042	-.029	-.014	-.062	-.057	-.012	.074	-.040	.017	-.003	.017	.069	.033	.073	.198	.173	.089	-.025	-.027

APPENDIX1. Correlation matrix(Continued)

	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	V34	V35	V36	V37	V38	V39
21	1																		
22	.083	1																	
23	-.008	.047	1																
24	.094	.151	.074	1															
25	.022	.164	.091	.200	1														
26	-.004	.061	.072	.146	.192	1													
27	.130	.082	.030	.051	-.013	-.019	1												
28	.011	-.011	-.060	.082	.176	.175	-.024	1											
29	-.023	.027	.079	-.034	.032	-.037	.031	.044	1										
30	-.016	.137	.066	.242	.201	.229	-.047	.060	.131	1									
31	.076	.063	.154	.167	.260	.240	-.067	.124	-.052	.250	1								
32	-.014	.145	.144	.243	.115	.180	-.136	.172	.014	.122	.296	1							
33	-.013	.007	.068	.120	.015	.126	.124	.146	.017	.104	.181	.161	1						
34	.111	.158	.130	.175	.154	.125	-.022	.063	-.093	.122	.228	.176	.077	1					
35	.070	.034	.120	.258	.233	.194	-.073	.109	.000	.184	.307	.268	.132	.192	1				
36	-.050	.090	.028	.046	.067	.078	-.122	.131	.002	.197	.133	.125	-.009	.114	.126	1			
37	.074	.160	.048	.179	.167	.128	-.050	.022	-.114	.166	.285	.212	.061	.261	.316	.187	1		
38	-.032	-.006	.147	.079	.179	.046	.044	.174	.054	.110	.270	.101	.099	.128	.212	.103	.096	1	
39	.051	.184	.113	.110	.172	.184	-.040	.132	-.115	.085	.244	.147	.043	.245	.239	.218	.213	.147	1

APPENDIX2. SPSS Statistic Syntaxes

Syntax1. Descriptive Statistic

```
* Encoding: UTF-8.
DATASET ACTIVATE DataSet1.
* Custom Tables.
CTABLES
/VLABELS VARIABLES=Gender Age College Department Admission_ method Filled_ form Includes Interest
  DISPLAY=LABEL
/TABLE Gender [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Age [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + College [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Department [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Admission-method [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Filled-form [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Includes [COUNT F40.0, TABLEPCT.COUNT PCT40.1] + Interest [COUNT F40.0, TABLEPCT.COUNTPCT40.1]
/CATEGORIES VARIABLES=Gender College Department Admission_method Filled_form Includes Interest
ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=Age ORDER=A KEY=VALUE EMPTY=EXCLUDE.
```

Syntax2. Factor Analysis

```
/VARIABLES V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39
 /MISSING LISTWISE
```

```
/ANALYSIS V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39
/PRINT INITIAL CORRELATION SIG KMO EXTRACTION ROTATION
  /FORMAT SORT BLANK (0.5)
  /CRITERIA FACTORS (16) ITERATE (25)
  /EXTRACTION PC
  /CRITERIA ITERATE (25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
DATASET DECLARE D0.804157593851612.
PROXIMITIES   V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39
  /MATRIX OUT (D0.804157593851612)
  /VIEW=VARIABLE
  /MEASURE=SEUCLID
  /PRINT NONE
  /STANDARDIZE=VARIABLE NONE.
```

Syntax3. Cluster Analysis

```
/MATRIX IN (D0.804157593851612)
  /METHOD BAVERAGE
  /PRINT SCHEDULE
  /PRINT DISTANCE
  /PLOT DENDROGRAM VICICLE.
Dataset Close D0.804157593851612.
```

APPENDIX3. Questionnaire Form

Mohammed Othman ABDULLAH
VAN YÜZÜNCÜ YIL UNIVERSITY
DEPARTMENT OF STATISTIC

Questionnaire Form

(Using Multivariate Methods to Determine the Most Important Affecting Factors for Students' Admission and Their Interests in the Specializations: A Sample of Salahaddin University)

Introduction:

Dear student,

This form is composed of many questions and parts of a master scientific research about the new system for admitting students and their interests in the specializations. Answering in a proper and suitable way will benefit the scientific researches.

General information:

1- Gender Male ☐ Female ☐

2- Age ☐

3- College ☐

4- Department ☐

5- Admission method

Zanko-line ☐ Credit ☐ Parallel ☐

6- The form was filled by:.

Myself ☐ a relative ☐ Centers ☐

7- The department that I am studying in was due to:

☐ My interest.

☐ Recommendation by my parents or a relative.

☐ The system.

8- Do I like the department that I am st

Yes ☐ No ☐ xtent ☐

	The parts below belong to the new system of admitting students.	Strongly disagree	disagree	Not with both	Agree	Strongly agree
1	I prefer the system.					
2	The students can trust this system.					
3	It helps the students completely be accepted in college and the specialization that he /she like.					

4	It has a good effect on the students' future.					
5	It facilitates for the students to choose the department that he/she wants.					
6	It is a reason to achieve justice and to get rid of administrative corruption.					
7	It doesn't account for economic status of the student.					
8	It made me impossible be able to study in the department and specialization I desired.					
9	Getting results will be late.					
10	Employing the geographic area dependent method affects choosing colleges and specialization.					
11	The chance of admission students in private universities must be the same as government universities.					
12	The private centers of filling forms are very helpful for correctly filing student's forms.					
13	The high schools fill the form for their students.					
14	The student is able to work on the system and fill their forms.					
15	In credit system, only the average of related subjects must be counted by % 100 proportions on a condition that the student just passed the other subjects.					
16	I like the number of students in parallel system be decreased.					
17	In parallel system the competition of scientific competence must be done.					
18	In parallel system the graduated students must not be able to refill admission forms.					
19	The department and specialization that I am studying in is %100 my interest.					
20	My parents or a relative had effect on choosing my specialization.					
21	The department and specialization that I am studying in enables me to be successful.					
22	My low marks did not let me to study in the department and specialization I desired.					

23	I like the order and the choices will be decreased to 25 choices.					
24	I prefer the Admission of students in master and PhD must be done by using the new system.					
25	It is better if counting of the average of grade 10th and 11th be optional.					
26	It is better if the age condition be omitted (if the student is at any age be able to fill the form.					
27	Differentiating between the students who passed in the first term and second term is beneficial for the students.					
28	I prefer the students that have remained only one subject be accepted (transit).					
29	I prefer that in government universities will not accept employee students.					
30	I prefer that handicapped students fill the form by the new system.					
31	It is better to give another chance to the students who have applied to universities twice, but have not been admitted.					
32	I like all students be accepted at the same time.					
33	I prefer applying to universities and admitting students in two different semesters.					
34	I prefer that the new system and registration of colleges be Linked so as to help students register their names.					
35	I suggest that daily the students be aware of the results when the competition is done in the system (by using his/her account).					
36	The slow of internet has an effect on the new system.					
37	The changes and guides must be given to all of the high schools before starting a new academic year.					
38	The number of the colleges and departments that accept only students of literary part is few.					
39	The process of(Improving average) is a good point.					

EXTENDED SUMMARY IN TURKISH
(GENİŞLETİLMİŞ TÜRKÇE ÖZET)

ÖĞRENCİ KABULÜNDE UZMANLIK ALANLARINI BELİRLEMEK İÇİN ETKİLİ OLAN EN ÖNEMLİ FAKTÖRLERİN BELİRLENMESİNDE ÇOK DEĞİŞKENLİ YÖNTEMLERİN KULLANIMI: SALAHADDİN ÜNİVERSİTESİ ÖRNEĞİ

ABDULLAH, Mohammed Othman
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Dr Öğr Üyesi. Yener ALTUN
İkinci Danışmanı: Dr Öğr Üyesi. Rizgar Maghdid AHMED
Ocak, 2020, 63 Sayfa

Bu tezin temel amacı, öğrenci kabulü için en önemli etkili faktörleri ve uzmanlaşma alanlarına ilgisini çok değişkenli yöntemler kullanarak belirlemektir. Bu nedenle, elde edilen bir dizi faktörü belirleyerek faktör analizi ve kümelerin sayısını beş kümede sınıflandırarak kümeleme analizine odaklanmıştır Ayrıca, faktör analizi ve kümeleme analizinin sonuçları birbiriyle karşılaştırılmıştır. Aynı zamanda, bu çalışma, 2018-2019 akademik yılı için Kuzey Irak'taki Salahaddin Üniversitesi'nin Fen ve Sosyal Bilimler kolejlerini içeren üç farklı kolejin ilk aşamasında öğrenciler üzerinde rastgele tabakalı örnekleme yöntemi ile dağıtılan 350 anket formunun analizine dayanmaktadır. Bu nedenle veri analizinde SPSS yazılım programı kullanılmıştır. Ayrıca sonuçların güvenirliği kabul edildiğini ve faktör analizinde toplam varyans yorumlama oranının% 62.157 olduğu sonucuna ulaşılmıştır. Aynı zamanda, faktör analizi ile kümeleme analizi arasındaki en yaygın değişkenler, öğrenci kabulü ve uzmanlık alanı seçimindeki ilgi alanları için en önemli etkili değişkenler olarak düşünülebilir. Sonuç olarak, ilk faktörün ve ilk kümenin ortak beş etkili değişkeni vardır, bunlar V1, V2, V3, V4 ve V5'tir (sistem, kolejlerde öğrencinin istediği meslekleri almasına yardımcı olur). Böylece bu durum faktör analizi ile küme analizi arasında bir tür yaklaşım ve benzerlik olduğunu göstermiştir.

Anahtar kelimeler: Faktör analizi, Kabul, İlgi alanları, Kümeleme analizi, Temel bileşen, Uzmanlaşma.

1. GİRİŞ

Öğrencilerin hayatındaki önemli konulardan biri, üniversitelere veya enstitülere ilgi alanlarına göre kabul edilmesidir. Irak ve Kuzey Irak'ta öğrenciler, eğitim dönemini üniversitede veya enstitüde 12 yıllık bir süre içinde bitirmektedir. Daha önce üniversitedeki kabul sistemi (merkezi kabul) bölümünde, eğitim bakanlığını üniversitelerde ve enstitülerdeki öğrencilerin kabulü için yükseköğrenimle ilişkilendiren çeşitli giriş formlarının doldurulmasına dayanmaktaydı. Günümüzde Kuzey Irak'taki Yükseköğretim Bakanlığı, yeni sisteme bağlıdır ve bu sistem, sonuçları makul bir oranda ilgilerine göre ortaya koymak için internete kabul formlarını doldurmaya dayanmaktadır.

Üniversitelerde veya enstitülerde öğrenci kabulü daha çok şansa bağlı bir süreçtir ve bu süreç lise mezunu öğrencilerin kaderi olarak algılanabilir. Kolejlerin seçimini etkileyen veya uzmanlıktaki isteklerini seçen birçok faktör vardır. Bu nedenle, Irak'ın kuzeyindeki Yükseköğretim ve Bilimsel Araştırma Bakanlığı, akademik yıl (2011-2012) için bölgedeki kolej ve enstitülere kabul için yeni bir sistem başlatmıştır. Aynı zamanda, bu sistem kurulduğu günden bu yana büyük bir başarıya sahiptir. Ayrıca, bu sistem ayrım yapmaksızın bölgedeki tüm öğrencileri, kuruluşundan bugüne kadar kabul etmek için kullanılmaktadır.

Bu araştırma, kolej ve enstitülere uzmanlık alanlarıyla ilgilenen yeni öğrenci kabul sistemini incelemektedir; yeni kabul sistemi dört farklı sistem içermektedir (Zankoline, Credit, Parallel ve Evening study). Öğrenci, öğrenci kabulü için uzmanlaşmış bir sistemde çevrimiçi bir form doldurarak birden fazla üniversiteye birçok istekte bulunur. Mevcut kabul sistemi öğrenciye bir üniversite tarafından kabul edilmesini sağlar, bu nedenle diğer öğrenciler için olumlu sonuçlara sebep olmaktadır. Önerilen sistem tüm üniversitelerdeki her öğrenciden sadece bir istek kabul etmektedir. Öte yandan, üniversitelerin öğrencilerin taleplerini sadece tek bir merkezden almaktadır.

Bu çalışmanın temel amacı çeşitli yönlerle araştırma sağlamak: öğrencinin hayatındaki geleceğiyle bağlantılı olarak bu sistemde öğrencinin karşı karşıya kaldığı ve uzmanlık alanlarındaki ilgilerini kabul etmek için karşılaştığı çeşitli sorunlar vardır. Bu tezde, bu sistemin tüm önemli yönleri incelenmiş, ileri istatistiksel yöntem kullanılarak işlenmiş, kümeleme analizi ve faktör analizi olmak üzere iki tür çok değişkenli analiz içermektedir. Ayrıca, bazı öneriler öğrencilerin ilgilendikleri bölümlere ulaşmaları için önemli ve etkili bir hizmet olarak kabul edilmektedir. Sorun, bu sistemde bazı eksiklikler veya kusurların (sonuç alma sürecinin uzun sürmesi, öğrencilerin ilgi görmemesi ve öğrencilerin sisteme olan güveninin azalması vb.) olmasıdır.

2. MATERYAL VE YÖNTEM

Kuzey Irak'taki yeni sistemde öğrencilerin yüzleştiği problemler ve öğrenci tercihlerinin daha iyi anlaşılması için, tercih edilen sistemler dikkate alınarak mantıklı ve doğru çözümler ile kümeleme ve faktör analizi ile incelenen bir istatiksel veri kullanılmıştır. Bu nedenle, aşağıda açıklanan verileri elde etmek için istatistiksel bir mekanizma uygulanacaktır.

Anket formunda uzmanlık alanlarına göre; ilgi alanları, yaş, cinsiyet, uzmanlık ve ortalama puan dikkate alınarak 39 değişkene odaklanan kabul sistemi stratejisine göre sorular hazırlanmıştır. Daha sonra bu anket formu, 2018-2019 akademik yılı için Salahaddin Üniversitesi-Erbil'de Fen ve Sosyal Bilimler kolejleri dahil 3 farklı kolejde 10 farklı bölümdeki 350 öğrenciye uygulanmıştır. Ayrıca Tabakalı Rastgele Örnekleme kullanılmıştır. Anketimize; Yönetici ve Ekonomi Koleji'nde 251, Fen Koleji'nde 74, Eğitim Koleji'nde 25 öğrenci katılım sağlamıştır. Veriler uygun şekilde toplandıktan sonra istatiksel bir programlama olan SPSS programına aktarılmıştır. Ayrıca değişkenlere kodlar verilmiştir; veriler faktör analizi (ana bileşen yöntemi) ve kümeleme analizi (değişkenlerin beş kümeye bölünmesi) üzerine odaklanan çok değişkenli yöntem kullanılarak incelenmiştir. Bunun yanı sıra, değişkenler arasındaki ilişkiyi gösteren bazı grafikler ve tablolar hazırlanmıştır.

3. BULGULAR

3.1. Faktör Analizi

Faktör analizi, öğrenci kabulü için yeni etkileyici faktörleri ve uzmanlaşmaya olan ilgilerini oluşturmak için kullanılmıştır. Aynı zamanda, faktör analizinin amacı, daha az sayıda faktör kullanarak değişkenler arasındaki çokluğu azaltmaktır. Değişkenler arasındaki ilişki temelinde her değişkenin önemini göstermek için korelasyon matrisini analiz etmede temel bileşen yöntemleri kullanılır.

Not: Tablo 4.6.’ da Döndürülmüş Bileşen Matrisi kullanılarak aşağıda yer alan sonuçlar elde edilmiştir.

Tabloda analizin boyut ve yükleme ile birlikte mutlak değerden (± 0,5) (V1 - V39) daha fazla değişken olan (16) değişkenleri ile sıraladığı gösterilmiştir. Daha sonra değişkenler, en yüksek yüklemeye sahip olan değişken nedeniyle sıralanır (korelasyonun pozitif veya negatif olup olmadığı dikkate alınmaz). Aslında, her değişkenin her bileşenden bazı yüklemeleri vardır, ancak | .50 | çıkıştan hariç tutulmalıdır, bu nedenle düşük yüklemelerin olduğu boşluklar vardır. (| .50 |, işaret dikkate alınmaksızın mutlak değer veya değer anlamına gelir).

3.2. Kümeleme Analizi

Bu çalışmada kümeleme analizi uygulanmıştır. Değişkenleri homojen gruplara ayırmak için kullanılan en önemli istatistiksel yöntemlerden biride kümeleme analizidir. Ayrıca, kullanılan kümeleme analizi veriler arasındaki farklılıklara ve benzerliklere bağlıdır. Ortalama bağlantı (gruplar arasında) ve yeniden ölçeklendirilmiş mesafe kümesini kullanarak uzaklık matrisini ve kümeler arasındaki ilişkiyi bulmak için birleştirme kullanılır; böylece ayrıştırma stilini kullanarak kümelerin sayısını beş ayrı kümeye bölmüş oluruz. Sonuçlar, Tablo 4.7.’de gösterildiği gibi homojenlik derecesi farklı gruplar arasında güçlü veya zayıftır.

4. SONUÇLAR

Faktör analizi, öğrencilerin kabulünü ve uzmanlığa ilgilerini etkileyen on altı bileşen üretmiştir. İlk beş bileşen sonuçlarda temeldir: ilk bileşen daha etkilidir, beş değişkeni temsil eder (sistem, öğrencinin üniversiteye kabul edilmesinde istenen meslekleri almasında yardımcı olur) ve bu bileşenin toplam etkisi % 7.818'dir. Ayrıca, yeni sistemin ikinci (yüksek lisans ve doktora öğrencilerinin kolejlere kabul edilmelerine ve hesaplarını kullanarak rekabetçi sonuçlar elde etmelerine yardımcı olur) bileşenin toplam etkisi % 6.683'tür. Aynı zamanda, üçüncü (öğrencinin istediği mesleklerin bölümlerde kabul edilmek için daha yüksek bir not alması gerekir) bileşenin toplam etkisi % 3.849'dur. Dördüncü bileşenin (mezun olan öğrencileri kullanarak öğrencinin paralel sisteme giriş sayısını azaltmak kabul kayıtlarını dolduramayacaktır) oran yorumlaması % 3.828 ve beşinci bileşenin (coğrafi alan ve ekonomik durum) oran yorumlaması % 3,789 dür.

Kümeleme analizinde kullanılan kümelerin sayısı beş kümeden ibarettir. Böylece elde edilen 39 değişken bu beş kümeye ayrılmıştır. Bu nedenle, ilk küme (sistem, öğrencilerin üniversiteye kabul edilmeleri için istedikleri meslekleri almaları için yararlıdır) % 25.656 oranında 10 değişken içerir. İkinci küme (yeni sistem, yüksek lisans ve doktora öğrencilerinin kolejlere kabul edilmelerine ve hesaplarını kullanarak rekabetçi sonuçlar elde etmelerine yardımcı olabilir) büyük bir oranın % 56.410'u oranında 22 değişken içerir. Üçüncü küme (kabul şansı, hem özel hem de devlet üniversitelerinde, öğrencinin yeteneğine uygun bir form doldurmakla ilgili) % 10.256 oranında dört değişken içerir. Ayrıca, dördüncü küme (paralel sistemde mezun olan ve çalışan öğrenciler kabul formunu dolduramazlar) % 5.128 oranında iki değişken içerir. Son küme % 2.564 oranında bir değişken içerir (düzen beğenilmiş ve seçenekler 25 seçeneğe düşürülecek).

Faktör analizi ile küme analizi arasında, ilk kümenin ve bileşenin ortak olarak aynı değişkenlere sahip olması şeklinde büyük bir yakınlaşma ve benzerlik vardır. Bu durum değişkenindeki küme ve bileşen analizlerinin çalışmada sınıflandırıldığını doğrular. Ayrıca sonuçlar, faktör analizi ile küme analizi arasındaki en karşılıklı değişkenlerin, öğrencinin kabulü ve uzmanlık alanlarındaki ilgileri için önemli ve etkili değişkenler olarak değerlendirilebileceğini göstermektedir. Sonuç olarak, ilkinde beş değişken ortaktır: V1, V2, V3, V4, V5, ikincisinde dört değişken ortaktır: V24, V32, V35, V37 ve üçüncüsünde bir değişken ortak değil, ancak dördüncüsünde sadece ortak bir değişken var o da V18’dir.

CURRICULUM VITAE

Mohammed Othman ABDULLAH was born in 1991, in Erbil province, Iraq. He completed primary, secondary and high school in Shaqlawa district. He graduated from Statistic Department in Salahaddin University - College of Administration and Economic in 2013. He started his Postgraduate study at the department of Statıstıc, Institute of Natural Applied Sciences at Van Yuzuncu Yil University in Turkey on February 2018.

UNIVERSITY OF VAN YUZUNCU YIL
THE ISTITUTE OF NATURAL AND APPLIED SCIENCES
THESIS ORIGINALITY REPORT

Date: 29/01/2020

Thesis Title: Using Multivariate Methods to Determine the Most Important Affecting Factors for Students' Admission and Their Interests in the Specializations: A Sample of Salahaddin University

The title of the mentioned thesis, above having total 63 pages with cover page, introduction, main parts and conclusion, has been checked for originality by Turnitin computer program on the date of 29/01/2020 and its detected similar rate was % 16 according to the following specified filtering

Originality report rules:

- Excluding the Cover page,
- Excluding the Thanks,
- Excluding the Contents,
- Excluding the Symbols and Abbreviations,
- Excluding the Materials and Methods
- Excluding the Bibliography,
- Excluding the Citations,
- Excluding the publications obtained from the thesis,
- Excluding the text parts less than 7 words (Limit match size to 7 words)

I read the Thesis Originality Report Guidelines of Yuzuncu Yil University for Obtaining and Using Similarity Rate for the thesis, and I declare the accuracy of the information I have given above and my thesis does not contain any plagiarism; otherwise I accept legal responsibility for any dispute arising in situations which are likely to be detected.

Sincerely yours,

29/01/2020

Name and Surname: Mohammed Othman Abdullah

Student ID#:17910002173

Science: Statistic

Program: Statistic

Statute: M. Sc. ☒ Ph.D. ☐

APPROVAL OF SUPERVISOR
SUITABLE

Asst. Prof. Dr. Yener ALTUN

APPROVAL OF THE INSTITUTE
SUITABLE