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

CLASSIFICATION OF PUBLIC AND PRIVATE HOSPITALS IN TERMS OF HEALTH SERVICES QUALITY: A SAMPLE OF ERBIL PROVINCE

M.Sc. THESIS

PREPARED BY: Ayad Faisal ISSA
SUPERVISOR: Asst. Prof. Dr. Yener ALTUN
CO-SUPERVISOR: Asst. Prof. Dr. Rizgar Maghdid AHMED

VAN-2021

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

CLASSIFICATION OF PUBLIC AND PRIVATE HOSPITALS IN TERMS OF HEALTH SERVICES QUALITY: A SAMPLE OF ERBIL PROVINCE

M.Sc. THESIS

PREPARED BY: Ayad Faisal ISSA

VAN-2021

ACCEPTANCE AND APPROVAL PAGE

This thesis entitled "**Classification of Public and Private Hospitals in Terms of Health Services Quality: A Sample Of Erbil Province**" and prepared by Ayad Faisal ISSA under consultation of Asst. Prof. Dr. Yener ALTUN in Department of Statistic, on date of ../01/2021 it has been successful with a unanimous vote by the following jury and it has been recognized as a Master's Thesis, in accordance with Postgraduate Education and training regulation with the relevant provisions.

Signature

Chair: Prof. Dr. H. Eray ÇELİK

Member: Doç. Dr. Hamit MİRTAGİOĞLU

Member: Dr. Öğr. Üyesi Yener ALTUN

The Board of Directors of the Institute of Science: ..../......./............... date and is approved by decision No................

Prof. Dr. Suat ŞENSOY

Director of the 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

Ayad Faisal ISSA

ABSTRACT

CLASSIFICATION OF PUBLIC AND PRIVATE HOSPITALS IN TERMS OF HEALTH SERVICES QUALITY: A SAMPLE OF ERBIL PROVINCE

ISSA, Ayad Faisal
M.Sc. Thesis, Statistics Department
Supervisor: Asst. Prof. Dr. Yener ALTUN
Co supervisor: Asst. Prof. Dr. Rizgar Maghdid AHMED
January 2021, 69 Pages

Health services quality has been the most important criteria of judging. In a health system, public and private sectors provide services and typically have been effective in promoting health services quality of community. The aim of this study is to classification public and private hospitals in Erbil province, Iraq, in terms of health services quality. This study includes a quantitative cross-sectional questionnaire based on the SERVQUAL dimensional model. In 2020, 370 questionnaire forms were distributed to patients by random stratified sampling method in all departments of 11 different hospitals, including public and private hospitals in Erbil. Data were analyzed by IBM – SPSS software version 25 by using factor analysis and discriminant analysis with ROC curve to evaluate the model. In public and private hospital, the factor analysis identified seven factors accounting for 64% of the total variance. It is seen that the correct classification rate of discriminant analysis is 78.1%, public and private hospitals are distinguish from each other. As a result of the analysis, the discriminatory model of the ROC curve criterion and the AUC value of the area under the ROC curve were obtained as 0.863, which means that the diagnostic test carries excellent discrimination. The X1, X7 and X8 variables are the most important and influential variables in terms of patient satisfaction between the factor analysis and discriminant analysis. As a result, it can be said that there is a significant difference between public and private hospitals.

Keywords: Discriminant analysis, Factor analysis, Patient, Private hospital, Public hospital, ROC curves, Service quality.

ÖZET

SAĞLIK HİZMETLERİ KALİTESİ AÇISINDAN KAMU VE ÖZEL HASTANELERİN SINIFLANDIRILMASI: ERBİL İLİ ÖRNEĞİ

ISSA, Ayad Faisal
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Dr. Öğr. Üyesi Yener ALTUN
İkinci Danışmanı: Dr. Öğr. Üyesi Rizgar Maghdid AHMED
Ocak 2021, 69 Sayfa

Sağlık hizmetlerinin kalitesi, değerlendirmenin en önemli kriteri olmuştur. Bir sağlık sisteminde, kamu ve özel sektör hizmet sağlar ve genellikle toplumun sağlık hizmetlerinin kalitesini yükseltmede etkili olmuştur. Bu çalışmanın amacı, Irak'ın Erbil ilinde bulunan kamu ve özel hastaneleri sağlık hizmetleri kalitesi açısından sınıflandırmaktır. Bu çalışma, SERVQUAL boyutsal modele dayalı niceleyici kesitsel bir anket içermektedir. 2020 yılında Erbil ilinde kamu ve özel hastaneler de dâhil olmak üzere 11 farklı hastanenin tüm bölümlerinde 370 anket formu rastgele tabakalı örnekleme yöntemi ile hastalara dağıtılmıştır. Veriler, modeli değerlendirmek için faktör analizi ve ROC eğrisi ile diskriminant analizi kullanılarak IBM – SPSS 25 istatiksel analiz programı ile analiz edilmiştir. Kamu ve özel hastanelerde faktör analizi, toplam varyansın% 64'ünü oluşturan yedi faktör tanımlamıştır. Diskriminant analizinin doğru sınıflandırma oranı % 78.1 olduğu, kamu hastaneleri ile özel hastaneleri birbirinden ayrıldığı görülmektedir. Analiz sonucunda, ROC eğrisi kriterinin ayırt edici modeli ve ROC eğrisi altındaki alanın AUC değeri 0.863 olarak elde edildi, bu da tanısal testin olağanüstü ayrımcılık taşıdığı anlamına gelmektedir. X1, X7 ve X8 değişkenleri faktör analizi ile diskriminant analizi arasında hasta memnuniyeti açısından en önemli ve etkili değişkenlerdir. Sonuç olarak kamu ve özel hastaneler arasında anlamlı bir fark olduğu söylenebilir.

Anahtar kelimeler: Devlet hastanesi, Diskriminant analizi, Faktör analizi, Hasta, Hizmet kalitesi, ROC eğrisi, Özel hastane.

ACKNOWLEDGMENT

Most of all I thanks "almighty Allah" with countless and gratitude, for all this blessing, Who gave me this great opportunity to complete this study and I would like to thank my supervisors Asst. Prof. Dr. Yener ALTUN and Asst. Prof. Dr. Rizgar Maghdid AHMED for their assistances, ideas, and feedbacks during the process of doing this study. Without their guidance’s and support, this thesis couldn’t be completed on time. I'm also very thankful for the head of Statistics Department Prof. Dr. H. Eray Çelik for his continued encouragement, support and providing facilities required for this investigation. I am very grateful and thankful to my father, my mother, my wife and my brothers for their support during my study and an impeccable understanding allowed me to accomplish this thesis. I would also like to thanks for Rizgare teaching hospital, west emergency hospital, Erbil teaching hospital, east emergency hospital, Zheen hospital, CMC hospital, Sardam hospital, Paky hospital, Balsam hospital and Hawler hospital for taking this data that helped me in this thesis. Finally, thanks for everyone who has helped me even by a word.

2021

Ayad Faisal ISSA

TABLE OF CONTENTS

Pages

Pages

Pages

LIST OF TABLES

LIST OF FIGURES

SYMBOLS AND ABBREVIATIONS

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

Symbols	**Description**
$\mathbf{h_i^2}$	Communalities
$\boldsymbol{\ell_{ij}}$	Loading of the *i*- *Th* Variable on The *j*- *Th* Factor
%	Percent
$\boldsymbol{\mu}$	Mean of Variables
$\boldsymbol{P}$	Rho
$\boldsymbol{\Sigma}$	Covariance Matrix Σ
$\boldsymbol{R}$	Correlation Matrix
$\boldsymbol{\lambda}$	Eigenvalues
$\boldsymbol{e}$	Eigenvector
$\boldsymbol{\sigma}$	Standard Deviation
S	Sample Covariance Matrix
$\boldsymbol{V_r}$	Varimax
$\boldsymbol{a}$	Discriminant Coefficient
$\underline{\boldsymbol{a}}$	Vector Form Parameters
$\underline{\boldsymbol{x}}$	Vector Variables.
$\boldsymbol{\partial}$	Partial Differentiontial
$\gimel$	Gimel
$\boldsymbol{\mathcal{L}}$	Laplace Transform
$\boldsymbol{\Lambda}$	Lambda

Abbreviations	**Description**
FA	Factor Analysis
EFA	Exploratory Factor Analys

Abbreviations	**Description**
CFA	Confirmatory Factor Analysis
PCA	Principal Component Analysis
Var	Variance
Cov	Covariance
PSI	Person Separation Index
KMO	Kaiser-Meyer-Olkin
RCV	Receiver Operating Characteristic
Z	Discriminant
X	Variable

LIST OF APPENDIXES

1. INTRODUCTION

1.1. Preface

The most significant factor in the effectiveness and survival of health organizations is the quality of health services, thereby increasing loyalty and customer satisfaction of supplier organizations (Ibarra et al., 2014). In order to be considered customer-oriented institutions, the quality of services, particularly in institutions facing high visitor volumes, should be used as a crucial measure; this has contributed to the satisfaction of service users and the growth of activities of health organizations, contributing to the effectiveness and efficiency of health service organizations (Aghamolaei et al., 2014). In addition, patients are becoming more open to competitive change and more familiar with health care facilities, so the standard of service alone might not be enough to preserve a long-term patient-hospital partnership (Gaur et al., 2011). Quality of service is consumers' perception of how well a service meets or exceeds its expectations, and customers judge it, not organizations. The collaborative design of the service process results in the client's quality evaluation directly after the delivery and success of that service (Douglas and Connor, 2003).

In the healthcare sector, identifying and assessing service quality has been a major challenge, because it requires intangibility, heterogeneity and inseparability. In hospitals, service quality has historically concentrated solely on the quality of medical care, but today it encompasses both medical services and hospitality. The service sector maintains two types of quality, namely the technological aspects and the functional aspects of the services. Professional consistency consists of procedural and diagnostic technical precision. Functional consistency refers to the delivery to patients of health care services. The technological quality of facilities can not necessarily be measured by patients, so functional quality is generally accepted as a determinant of patient satisfaction in hospitals. Since the perceived quality of health care services has a direct effect on health service consumption behavior and behaviors, it is necessary to understand the preferences and perceptions of health service quality patients (Baltussen et al., 2002).

In Iraq, health services are divided into public and private sectors. In the public sector, all types of health services such as medicine, oxygen, all types of rays, medical tests, food, surgery, patient transportation equipment like (Ambulance, wheelchair and patient bed ….) and patient accommodation, car parking, heating or cooling systems and equipment, hospital hygiene and also all hospital employees and medical service staff such as specialist doctors, nurses, administrative staff, servants, etc. all of these services must be provided by the government without any expectation to obtain patient satisfaction.

But in the private sector, all the health services mentioned in the public sector are done through payment by person or company.

1.2. Statement of the Problem

After we felt there is a big difference in the quality of health services between public and private hospitals and neglecting of public hospitals by the people and giving importance to private hospitals because people complain a lot about public hospitals in terms of the quality of health service such as they are not clean and hygienic enough, lack of availability of all types of medicines, medical rays, medical test, defects in heating and cooling systems and adequate patient care by hospital staff …. And that’s why we did this research.

1.3. Aim of the Research

1. Using two advanced statistical ways which are factor analysis and discriminant analysis to reach the most important factors that affect the quality of health care in hospitals.
2. Estimate the distinguish formula between public and private hospitals using discriminant function.
3. Using receiver operating characteristic (ROC) curves criterion as an important measure to evaluate the model.

1.4. Service Quality in Health Sector

In healthcare organizations, service quality and customer satisfaction attract tremendous attention and this topic is taken into account in their strategic method for planning. The hospital's image and profitability are also impacted by the perceptions of patients regarding the services offered by a particular health care organization. (Donabedian, 1980; Williams and Calnan, 1991). And in terms of loyalty and word-of-mouth, it also significantly affects the patient's actions. (Andaleeb, 2001). In addition, healthcare service providers have realized enhanced patient service quality standards to identify the key determinants needed to increase patient satisfaction healthcare services and also help service providers reduce the time and resources involved in handling patient complaints (Pakdil and Harwood, 2005).

1.5. SERVQUAL Method

The SERVQUAL method is a comprehensive measurement tool developed for the purpose of measuring service quality by (Parasuraman et al., 1985). By showing them the difference between expectations before consumption and perceptions after consumption of the item, it guides service suppliers. In the process, questionnaires are used to assess the preferences and perceptions of customers. The fifth distance was based on the SERVQUAL scale and the initial ten dimensions were further consolidated into five service quality dimensions. These five dimensions are described as follows:

1.5.1. Tangibles

It is the physical image of the operation that will be used to assess quality by customers. Tangibles are synonymous with the physical infrastructure, tools and machines used to deliver the service, as well as representations of the services, such as statements, debit and credit cards, transaction speed and performance. Many privileges are found in tangibles, such as external presence, bank counters, overdraft facilities, opening hours, speed and accuracy of transactions (Parasuraman et al., 1994).

1.5.2. Reliability

Reliability is that suppliers deliver the service at the right time and fulfill their obligations, particularly accurate billing, proper record keeping, and delivery to the consumer of the service at the agreed time (Kondasani and Panda, 2015). When service providers fulfill their obligations, consumer loyalty increases and their confidence in the service provider increases as the performance of the supplier constantly improves and continues to meet customer standards (Rad et al., 2010).

1.5.3. Responsiveness

Responsiveness involves the ability or willingness of the service provider to provide a timely service. (Parasuraman et al., 1985). This deals with timeliness, such as providing the customer with fast services, setting up appointments as soon as possible, submitting the transaction slip to the customer immediately so that he/she does not build the wrong impression and quickly calling the customer (Calisir et al., 2014).

1.5.4. Assurance

Assurance refers to understanding, courtesy and the ability of workers to show trust and confidence. Employees should have the requisite expertise and skills to provide their clients with the best service. Courtesy means courtesy, courtesy, friendliness and respect, such as customer/property care. The conveyance of confidence and trust by customers implies trustworthiness, honesty and integrity. This includes keeping the best interests of the client at heart, such as the features of communication workers (Kitapci et al., 2014).

1.5.5. Empathy

Empathy applies to helpful, respecting client demands and offering individualized treatment to clients. The ability to identify client expectations contributes to the aptitude to reply to them, such as recognizing repeat clients and knowledge their particular needs

(Parasuraman et al., 1985). Empathy for vendors and consumer relationships affects client satisfaction positively (Zarei et al., 2012).

2. LITERATURE REVIEW

Mensah et al. (2016), the aim was to compare the relationships in public and private hospitals between patients and medical staff. To determine if there was a disparity between private and public hospitals in diagnostic services and to quantify the difference between private and public hospitals in diagnostic facilities. The thesis used the methodology of ease sampling in the choice of the sample. The key analytical method used in evaluating the data was the independent sampling T-test. The findings were that more quality healthcare was offered by the private hospital than by the public hospital and it was recommended that appropriate diagnostic facilities should be provided for the management of public hospitals and private hospitals and ensure proper maintenance of the existing ones.

Alijanzadeh et al. (2016), the purpose of this research was to compare the quality of health services from the viewpoint of Qazvin residents in both the private and public sectors (Iran). The estimated sample size was 1002. A perception of service quality level questionnaire was the study method. Educated interviewers visiting homes collected data and analyzed it with IBM-SPSS version 22 software and t-test and linear regression. The alpha coefficient of Cronbach was 0.91 and the test-re-test coefficient was 83 percent. The findings demonstrated the value of a tangible area for the satisfaction of people with health care. The public sector, it seems, should pay more attention to this issue.

Amin and Nasharuddin (2013), aims of research study was to analyze the quality of hospital service and its effect on patient satisfaction and the purpose of conduct. In this study, a comfort sampling technique was used. 350 questionnaires were distributed in total and 216 were returned (61.7 per cent response rate). The results indicate that the perceived quality of service and customer satisfaction should be used by managers as exit strategy tools to maximize loyalty among current customers.

Andaleeb (2000), the idea of the study was that the success of hospital facilities depends on the incentive system in which these organizations work. Adequate predictive performance was demonstrated by the model, based on discriminant analysis. Only 300 interviews from Dhaka City alone were planned due to resource and time constraints, and the preliminary nature of this inquiry. The findings showed that many patients derive performance criteria from private hospitals (e.g. the sample of this study shows that a greater proportion of patients were seeking private care).

Taner and Antony (2006), the aim was to investigate the difference in the standard of service between public and private hospitals in Turkey. The SERVQUAL approach has been used to research the level of service offered by them. The study consisted of 200 patients in total. The outcome point for inpatients in private hospitals was that they were more pleased with the standard of care than in public hospitals. The finding also shows that hospital patients are more pleased with physicians, nurses and support staff in private hospitals than those in public hospitals. In conclusion, the outcome indicates that the key determinants of service quality in public hospitals is satisfaction with doctors and fair prices. The study shows that SERVQUAL was trustworthy and relevant in a hospital setting as a standard method for assessing functional service quality.

Eggleston et al. (2010), the aim was to compare the efficiency and effectiveness of the operations of public and private hospitals in Guangdong Province, China, focusing on case-mix inequalities and the quality of care of patients. Survey data obtained from Guangdong's 362 government and private hospitals was analyzed. To compare hospital characteristics and to classify variables correlated with simple systemic quality measures and patient results, the model used univariate and multi-variate regression analyses. The findings showed that improvements in the form of ownership alone are unlikely to boost or damage overall performance significantly. System incentives, regardless of the type of hospital ownership, need to be structured to reward desired hospital success and protect vulnerable patients.

Yeşilada and Direktouml (2010), the purpose of this research was to appraise the quality of service given in Northern Cyprus' private and public hospitals and to define the dimensions of quality healthy service that play significant role in patient satisfaction. Data from the same study, consisting of 806 systematically selected individuals over the age of eighteen, were collected in two stages. A three-factor approach, namely: reliability-confidence, empathy and tangibles, was found in factor analysis. The five factor model of the original SERVQUAL does not support this result. In all three dimensions of service quality, gap analysis found that private hospitals had smaller differences than public hospitals. Finally, results of logistic regression showed that while all three dimensions were somewhat influential on patient satisfaction.

Irfan and Ijaz (2011), the purpose of this research was to determine the standard of service rendered by private hospitals in Pakistan, which was simply focused on the experience of the patient. Based on the SERVQUAL model, a questionnaire was established comprising 22 variables representing five dimensions of service quality; empathy, observable, assurance, timeliness and responsiveness. The goal population of this study was the workers employed in the service organizations at officer level and using health facilities from the best private hospitals in the city, Lahore, Pakistan, including consultation and in-patient. For the study, a total of 320 patients were considered. The findings of the study show that the quality of service in private hospitals exceeded the satisfaction of patients, i.e. private hospitals offered better healthcare facilities. Results from the five variables showed that there was a reasonable fit for the measurement model for service quality constructs and the model was accurate and reliable.

Abuosi and Atinga (2013), the purpose of this study was to address two key issues with the authors: to examine the perceptions of the quality of hospital service and expectations of patients using SERVQUAL; and to describe the different definitions used to measure the perceptions of patients. After admission and follow-up appointments, questionnaires were administered to 250 patients. The 22 paired expectation and perception elements of SERVQUAL were adopted. For data analysis, repeated t-measures and factor analysis with Varimax rotation were used. The findings revealed that during medical care, patient standards were not observed. For all variables, perceived service quality was rated lower than expectations. The primary discrepancy was statistically significant between beliefs and expectations. In comparison to the five-factor model of SERVQUAL, the analysis identified four service-quality variables.

Karekar et al. (2015), this research was carried out in order to examine the quality of customer service in two types of hospitals. The SERVQUAL instrument was used for this purpose to assess the patient's view of the quality of service rendered by these hospitals. By serving a questionnaire to the respondent, primary data was obtained. The questionnaire was distributed to respondents who were chosen as a sample and the consequences of the questions were clarified in certain instances by the researcher.

Results have shown that, compared to government hospitals, private hospitals provide their patients with higher quality facilities.

D'Cunha and Suresh (2015), patient satisfaction was the key metric used to measure the standard of treatment. The research was a cross sectional descriptive study. In order to pick a group of 100 patients who were admitted for more than two days, the random sampling methodology was introduced. The standardized questionnaire was used to measure the dimensions of quality and the degree of customer satisfaction. The data was evaluated based on the goals of the analysis in terms of percentage, mean score, standard deviation. Results show that the experience of the patient differed according to different quality measurements. It was found that most patients were pleased with the hospital services, quality physician care, and quality nursing care rendered by the nurses.

Alumran et al. (2020), the purpose of this research was to evaluate of the health services quality from a patient perspective and to compare of the quality of service of private and public hospitals in Saudi Arabia's eastern region. The research has a cross-sectional quantitative design, with a questionnaire based on the dimensional model of SERVQUAL. This was a random sample of 258 patients in Eastern Saudi Arabia, in private and public hospitals. The findings showed that patients perceived a higher level of quality of health care facilities in private hospitals ($t = 3.390$, $p < 0.01$).

3. METHODOLOGY

This study is intended to make classification and evaluation model of the quality health services rendered by public and private hospitals in Erbil Province. For better developing, the action of quality health service problems from public and private hospitals in north Iraq, with good and rescannable solution, we prefer statistical data that had been analyzed through factor and discriminant analysis with ROC curve to evaluate the model. Therefore to get the data, we will apply the mechanism which explained below:

We applied the SERVQUAL instrument, widely used in contemporary research (Sohail, 2003; Mostafa, 2005; Wiesniewski and Wiesniewski, 2005), in the study. The study questionnaire consists of two parts; the first section includes demographic characteristics of the hospitalized patient such as gender, age, marriage, profession, economic status and education, and the second section includes 22 variables related to quality health services. We adopted all the original SERVQUAL questions without modification, the SERVQUAL scale was based on the five dimensions of service quality namely Tangibles, Reliability, Responsiveness, Assurance and Empathy. The perception and expectation variables were measured on a five point Likert scale ranging from 1="strongly disagree" to 5="strongly agree". Then this questionnaire form is distributed randomly among 370 patients in all departments of eleven different hospitals including public and private hospitals in Erbil province in the year 2020.Moreover, stratified random sampling has been used. Thus, they are 188 patients in public hospital and 182 patients in private hospital.

After that 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 analysis) to extract the most important variables which affect patient satisfaction towards public and private hospitals were identified. Using varimax rotation, loadings over 0.5 were used to interpret components and Stepwise discriminant analysis was performed to develop a model based on the subgroup of variables that best distinguished between public and private hospitals and used the ROC curve criterion to evaluate the discriminant model.

The thesis consists of five chapters. The first chapter is about introduction and background information of quality health services, the second chapter displays literature review, the third chapter contains the theoretical part of factor analysis and discriminant Analysis with ROC curve to evaluate the model, the fourth chapter shows the practical part, its data analysis and results, and the last chapter presents conclusion of the study.

3.1. Factor Analysis

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 refer a collection of nearly associated forms intended for clarifying or creating co-relational structure between the observed random variables (Basilevsky, 2009). Factor analysis is a data reduction technique. For instance, factor analysis can be used to identify the underlying components (factors) that explain the correlations among a set of variables. In this way, it is possible to employ a smaller set of measures (the factors) to explain a substantial portion of the total variance that is explained by all the original variables.

There are two major types of factor analysis: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). In exploratory FA, one seeks to describe and summarize data by grouping together variables that are correlated. The variables themselves may or may not have been chosen with potential underlying processes in mind. Exploratory FA is usually performed in the early stages of research, when it provides a tool for consolidating variables and for generating hypotheses about underlying processes. Confirmatory FA is a much more sophisticated technique used in the advanced stages of the research process to test a theory about latent processes. Variables are carefully and specifically chosen to reveal underlying processes (Child, 2006).

In factor analysis we represent the variables $x_1, x_2, \dots, x_p$, as linear combinations of a few random variables $f_1, f_2, \dots, f_m$ (m < p) called factors. The factors are underlying constructs or latent variables that “generate” the x’s. Like the original variables, the factors vary from individual to individual; but unlike the variables, the factors cannot be measured or observed (Shiker, 2012).

The goals of factor analysis are to determine the number of fundamental influences underlying a domain of variables, to quantify the extent to which each variable is associated with the factors, and to obtain information about their nature from observing which factors contribute to performance on which variables (Tinsley and Brown, 2000).

3.1.1. Assessment of the suitability of the data for factor analysis

3.1.1.1. Sample size

A wide range of recommendations regarding sample size in FA have been made. The recommendations are usually stated in terms of either the minimum sample size (N) for a particular analysis or the minimum ratio of N to the number of variables, p, i.e. the number of survey items being subjected to FA (MacCallum et al., 1999). Gorsuch (1983), recommended five subjects per item, with a minimum of 100 subjects, regardless of the number of items. Mundfrom et al. (2005) subsequently provided sample size recommendations for 180 population conditions on the basis of a Monte Carlo study that varied the number of factors, the ratio of variables to factors (an important aspect of over determination) and communalities.

3.1.1.2. Correlation matrix

A correlation matrix should be used in the EFA process displaying the relationships between individual variables. Tabachnick et al. (2007) recommended inspecting the correlation matrix (often termed Factorability of *R*) for correlation coefficients over 0.30. Hair et al. (1995) categorised these loadings using another rule of thumb as ±0.30=minimal, ±0.40=important, and ±.50=practically significant. If no correlations go beyond 0.30, then the researcher should reconsider whether factor analysis is the appropriate statistical method to utilize. In other words a factorability of 0.3 indicates that the factors account for approximately 30% relationship within the data, or in a practical sense, it would indicate that a third of the variables share too much variance, and hence becomes impractical to determine if the variables are correlated with each other or the dependent variable (multicollinearity).

3.1.1.3. Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy/Bartlett's Test of Sphericity

Prior to the extraction of the factors, several tests should be used to assess the suitability of the respondent data for factor analysis. These tests include Kaiser-Meyer-Olkin (KMO): A measure of whether your distribution of values is adequate for conducting factor analysis. The KMO index, in particular, is recommended when the cases to variable ratio are less than 1. The KMO index ranges from 0 to 1, with 0.50 considered suitable for factor analysis (Kaiser, 1974). And Bartlett's Test of Sphericity a measure of the multivariate normality of your set of distributions. It also tests whether the correlation matrix is an identity matrix (factor analysis would be meaningless with an identity matrix). Should be significant ($p<.05$) for factor analysis to be suitable (George and Mallery, 2019).

3.1.2. Communalities

Communality h^2 gives the variance accounted for a particular variable by all the factors. The communality is then computed for all variables (the sums of squares of the rows of the factor loadings). These communalities are compared with the original estimates in the diagonals. Unless the differences are tiny these h^2values are put into the diagonals and principal factors are again computed, the same number of factors being extracted. This cyde is repeated again and again until the h^2 values no longer differ from the diagonals. The factor loadings which were used in the final iteration are regarded as correct (Nunnally, 1978). The formula for deriving the communalities is:

$$h_i^2 = \sum_{i=1}^{m} \ell_{ij}^2 \qquad j = 1, 2, \ldots, p \qquad i = 1, 2, \ldots, m \tag{3.1}$$

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

3.1.3. Factor extraction

Factor extraction involves determining the smallest number of factors that can be used to best represent the inter relation among the set of variables. There are a variety of approaches that can be used to identify (extract) the number of underlying factor or dimensions (Tabachnick and Fidell, 1996).

3.1.3.1. Method of factor extraction

There are numerous ways to extract factors (Principal components analysis (PCA), principal axis factoring (PAF), image factoring, and maximum likelihood, Alpha factoring, unweighted least squares and generalized least squares). When variables have high reliability or where there are 30 or more variables that PCA is the default method in many statistical programs. For this thesis the principal component analysis was used (Thompson, 2004).

3.1.3.1.1. Principal components analysis

The principal components are ordered, with the first component extracting the most variance and the last component the least variance. The solution is mathematically unique and, if all components are retained, exactly reproduces the observed correlation matrix. Further, since the components are orthogonal, their use in other analyses (e.g., as DVs in MANOVA) may greatly facilitate interpretation of results. PCA is the solution of choice for the researcher who is primarily interested in reducing a large number of variables down to a smaller number of components.

The goal of PCA is to extract maximum variance from the data set with each component. The first principal component is the linear combination of observed variables that maximally separates subjects by maximizing the variance of their component scores. The second component is formed from residual correlations; it is the linear combination of observed variables that extracts maximum variability uncorrelated with the first component. Subsequent components also extract maximum variability from residual correlations and are orthogonal to all previously extracted components (Tabachnick et al., 2007). Mathematically, principal components are special linear combinations of the P random variables $X_j = (X_1, X_2, \dots, X_p)$, where principal components depend only on the covariance matrix $\boldsymbol{\Sigma}$ (or the correlation matrix $\boldsymbol{R}$) of $(X_1, X_2, \dots, X_p)$.Let the random vector $\mathrm{X}' = [X_1, X_2, \dots, X_p]$have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P \geq 0$. Consider the linear combinations:

$$Y_1 = \mathrm{a}_1' X = a_{11} X_1 + a_{12} X_2 + \cdots + a_{1p} X_p \, a_{1p} X_p$$

$$Y_2 = \mathrm{a}_2' X = a_{21} X_1 + a_{22} X_2 + \cdots + a_{2p} X_p$$

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

$$Y_P = \mathrm{a}_p' X = a_{p1} X_1 + a_{p2} X_2 + \cdots + a_{pp} X_p \tag{3.2}$$

Or in matrix form.

$\underline{\mathrm{Y}} = \mathrm{A}\, \underline{\mathrm{X}}$

$$\mathrm{Var}\,(Y_i) = \mathrm{a}_i' \Sigma\, \mathrm{a}_i \qquad i = 1, 2, \dots, p \tag{3.3}$$

$$\mathrm{Cov}\,(Y_i, Y_k) = \mathrm{a}_i' \Sigma\, \mathrm{a}_k \qquad i, k = 1, 2, \dots, p \tag{3.4}$$

The principal components are those uncorrelated linear combinations $Y_1, Y_2, \dots, Y_p$.

Whose variances in (3.3) are as large as possible.

The first principal component is the linear combination with maximum variance. That is, it maximizes $\mathrm{Var}(Y_1) = \mathrm{a}_1' \Sigma\, \mathrm{a}_1$. It is clear that $\mathrm{Var}(Y_1) = \mathrm{a}_1' \Sigma\, \mathrm{a}_1$ can be increased by multiplying any a_1 by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length. We therefore define

First principal component = linear combination $\mathrm{a}_1' X$ that maximizes

$$\mathrm{Var}(\mathrm{a}_1'\mathrm{X}) \text{ subject to } \mathrm{a}_1'\mathrm{a}_1 = 1$$

Second principal component = linear combination $\mathrm{a}_1'\mathrm{X}$ that maximizes

$$\mathrm{Var}\,(\mathrm{a}_2'\mathrm{X}) \text{ subject to } \mathrm{a}_2'\mathrm{a}_2 = 1$$

and

$$\mathrm{Cov}\left(\mathrm{a}_1'\mathrm{X}, \mathrm{a}_2'\mathrm{X}\right) = 0\,.$$

At the ith step, ith principal component = linear combination $\mathrm{a}_1' X$ that maximizes

$$\text{Var}\ (a_i' X)\ \text{subject to}\ a_i' a_i = 1$$

and

$$\text{Cov}\left(a_i' X, a_k' X\right) = 0 \quad \text{for} \quad k < i.$$

From result (3.2), let Σ be the covariance matrix associated with the random vector $X' = [X_1, X_2, \dots, X_p]$. Let Σ have the eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P \geq 0$. Then the ith principal component is given by

$$Y_i = e_i' X = e_{i1} X_1 + e_{i2} X_2 + \cdots + e_{ip} X_p, \qquad i = 1, 2, \dots, p \tag{3.5}$$

with these choices,

$$\text{Var}\ (Y_i) = e_i' \Sigma e_i = \lambda_i \qquad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = e_i' \Sigma e_k = 0 \qquad i \neq k. \tag{3.6}$$

The principal components are uncorrelated and have variances equal to the eigenvalues of covariance matrix Σ.

In result (3.3) Let $X' = [X_1\ X_2, \dots, X_P]$have covariance matrix Σ, with eigenvalue-eigenvector pairs$(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P \geq 0$. Let $Y_1 = e_1' X, Y_2 = e_2' X, \dots, Y_p = e_p' X$ be the principal component then

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{J=1}^{P} Var\ (X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_P = \sum_{J=1}^{P} Var\ (Y_i). \tag{3.7}$$

From result (3.4), let $Y_1 = e_1' X, Y_2 = e_2' X, \dots, Y_p = e_p' X$ are the principal components obtained from the covariance matrix Σ, then

$$P\ (Y_i, X_k) = \frac{e_{iK} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i, k = 1, 2, \dots, p. \tag{3.8}$$

The correlation coefficients between the components Y_i and the variables X_k. here $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ are the eigenvalue-eigenvector pairs for Σ (Johnson and Wichern, 2002).

3.1.3.2. Determining the number of factors to be extracted

There are two conventional criteria for determining the number of initial unrotated factors to be extracted. These are the eigenvalue and the scree test criterion.

3.1.3.2.1. Eigenvalue

Only factors with eigenvalues of 1 or greater are considered to be significant; all factors with eigenvalues less than 1 are disregarded. An eigenvalue is a ratio between the common (shared) variance and the specific (unique) variance explained by a specific factor extracted. The rationale for using the eigenvalue criterion is that the amount of common variance explained by an extracted factor should be at least equal to the variance explained by a single variable (unique variance) if that factor is to be retained for interpretation. An eigenvalue greater than 1 indicates that more common variance than unique variance is explained by that factor (Robert Ho, 2006).

3.1.3.2.2. Scree test

The scree test is derived by plotting the eigenvalues (on the Y axis) against the number of factors in their order of extraction (on the X axis). The initial factors extracted are large factors (with high eigenvalues), followed by smaller factors. Graphically, the plot will show a steep slope between the large factors and the gradual trailing off of the rest of the factors. The point at which the curve first begins to straighten out is considered to indicate the maximum number of factors to extract. That is, those factors above this point of inflection are deemed meaningful, and those below are not. As a general rule, the scree test results in at least one and sometimes two or three more factors being considered significant than does the eigenvalue criterion (Cattel, 1966).

3.1.4. Factor rotation and interpretation

Another consideration when deciding how many factors you will analyse your data is whether a variable might relate to more than one factor. Rotation maximises high item loadings and minimises low item loadings, therefore producing a more interpretable and simplified solution. There are two main approaches to rotation, resulting in either orthogonal (uncorrelated) or oblique (correlated) factor solution. According to orthogonal rotation results in solutions that are easier to interpret and to report, however they do require the researcher to assume that the underlying constructs are independent (not correlated). Oblique approaches allow for the factors to be correlated, however they are more difficult to interpret, describe and report. In practice the two approaches (orthogonal and oblique) often result in very similar solution, particularly when the pattern of correlations among the items is clear.

Within the two broad categories of rotational approaches there are a number of different rotational techniques provided by SPSS (orthogonal: Varimax, Quartimax, Equamax; oblique: Direct Oblimin, Promax). The most commonly used orthogonal approach is the Varimax method (Tabachnick and Fidell, 1996).

3.1.4.1. Varimax rotation

The goal of the varimax method is to obtain factors that have high loadings for a subset of variables on only one factor and zeros for the other factors. This generally leads to easy interpretation since variables are not confounded by factors. To accomplish this, Kaiser (1958), varimax criterion makes the sum of the variances of the loadings in each column of the matrix $\widehat{\Lambda}_{*}$large, subject to the constraint that the communality of each variable is unchanged; or letting $\widehat{\Lambda}_{*} = [\hat{\lambda}_{ij}]$

$$V_j = \sum_{i=1}^{P} (\hat{\lambda}_{ij}^2 - \hat{\lambda}_{.j}^2)^2 / P$$

$$= \left[P \sum_{i=1}^{P} \hat{\lambda}_{ij}^4 - \left(\sum_{i=1}^{P} \hat{\lambda}_{ij}^2 \right)^2 \right] / P^2$$

Where V_j is the variance of the communality of the variables within factor j , and $\hat{\lambda}^2_{.j} = \sum_{i=1}^{P} (\hat{\lambda}^2_{ij} - \hat{\lambda}^2_{.j})^2 / P$ is the average of the squared loadings for factor *j*. Summing V_j over all factors

$$V_r = \sum_{j=1}^{k} V_j = \sum_{j}^{k} \left[\frac{1}{P} \sum_{i=1}^{P} \hat{\lambda}^4_{ij} - \frac{1}{P^2} \left(\sum_{i=1}^{P} \hat{\lambda}^2_{ij} \right)^2 \right] \quad (3.9)$$

Yields the raw varimax criterion (Magnus and Neudecker, 2019).

3.2. Discriminant Analysis

Discriminant or discriminant function analysis is a parametric technique to determine which weightings of quantitative variables or predictors best discriminate between 2 or more than 2 groups of cases and do so better than chance (Cramer, 2003). Moreover, Discriminant analysis is a powerful multivariate statistical method that can classify individual objects or units into groups. Specifically, in discriminant analysis, there are many individual objects and each object has several variables associated with it, that is, $(x_1, and\ x_2, \ldots, x_n)$. The whole population of objects can be partitioned into several groups with the values of variables strongly related to the group the object belongs to. Discriminant analysis is a method that classifies each object into groups depending on the values of the variables of each object (Yang and Trewn, 2004).

There are two major objectives in separation of groups:

1. Description of group separation, in which linear functions of the variables (discriminant functions) are used to describe or elucidate the differences between two or more groups. The goals of descriptive discriminant analysis include identifying the relative contribution of the p variables to separation of the groups and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups.

2. Prediction or allocation of observations to groups, in which linear or quadratic functions of the variables (classification functions) are employed to assign an individual sampling unit to one of the groups. The measured values in the observation vector for an individual or object are evaluated by the classification functions to find the group to which the individual most likely belongs (Mushtak, 2012).

The goal of discriminant analysis is to describe, either graphically (in three or fewer dimension) or algebraically, the differential features of objects (observation) from several known collection (population). We try find “discriminant” whose numerical values are such that the collections are separated as much as possible. And to sort objects (observations) into two or more labeled classes. The emphasis is on deriving a rule that can be used to optimally assign new objects to the labeled classes (Johnson and Wichern, 2002).

3.2.1. The basic points of discriminant analysis

1. One of the multivariate statistical methods used to study the range of sets overlap.
2. Depends on the basis of classification of observations between the sets under research and study.
3. Two or more sets that share one another shall be dealt with in a set of characteristics but separate from each other quantitatively (Talib et al., 2019).

There are also some important points in the method of discriminatory analysis:

1. This is used for the method of discriminate between two or more groups depending on the comparison of independent variable values and can be used for any number of dependent variables and sets.
2. This is an extension method for multiple regression analysis and is therefore an advanced point for the use of statistical methods in analysis.
3. This method is more sensitive and more efficient than the logistic regression system depending on the error classification average. (Rencher, 2003).

3.2.2. Sample size

The investigator needs to determine the sample size needed to perform a discriminant analysis. We suggest that investigators avoid very large sample sizes because statistical tests are more likely to yield significant results for trivial differences under such circumstances. As sample size increases, the variability in the sampling distributions (i.e., the denominator in most correlational statistics) decreases, assuming all other factors are held constant. This will result in a reduction of the overlap between sampling distributions under the null and alternative hypotheses, leading to a greater probability of finding statistical significance (viz., rejecting the null hypothesis). However, small sample sizes are not recommended because the idiosyncrasies in the sample will unduly influence the statistical results, thereby lessening the stability and generalizability of the results (Morrison, 1974; Huberty, 1975).

Descriptive discriminant analysis requires that there are complete data on the discriminator variables and that group membership is known. Those cases that do not have scores on all of the discriminator variables automatically are excluded from the analysis, possibly affecting the adequacy of the data in satisfying sample and group size requirements adversely. Furthermore, cases having missing data may be systematically different from those cases containing complete data, so use of those with incomplete data could result in generalizability problems (Norusis, 1990).

3.2.3. Linear discriminating function of two groups

The function of discrimination is a model that can be explanation the randomly selected sample and placed in two different groups. By this function we can test observation and determine its return to any group (Fawzy, 2004).

$$Z = a_1 x_1 + a_2 x_2 + \cdots + a_P x_P \tag{3.10}$$

a = the discriminant coefficient or weight for that variable

P = the number of predictor variables

$Z = \underline{a}'\underline{x}$

This is the linear function of Fisher.

$\underline{a}$: Vector form parameters and used in the classification Process.

$\underline{x}$: Vector variables.

The maximization is when:

$$\alpha = S^{-1}(Z_1 - Z_2) \tag{3.11}$$

It is necessary to find the vector that produce the Q maximize value as possible.

$$Q = \frac{[\bar{Z}_1 - \bar{Z}_2]^2}{S_Z^2} = \frac{[\bar{Z}_1 - \bar{Z}_2]^2}{\sum_{i=1}^{n_1}[Z_{i1} - \bar{Z}_1]^2 + \sum_{i=2}^{n_2}[Z_{i2} - \bar{Z}_2]^2} \tag{3.12}$$

Q: Square Mahalanobis Distance

$\bar{Z}_1$: Median observations values of the first group

$\bar{Z}_2$: Median observations of the second group

Z_{i1}: The value observations i in the first set

Z_{i2}: The value observations i in the second set

Then find $\bar{Z}_1$ $\bar{Z}_2$

$\bar{Z}_1 = \underline{a}' \, \underline{\bar{x}}_1 = a_1\bar{x}_{11} + a_2\,\bar{x}_{21} + \cdots + a_p\bar{x}_{p1}$

$\bar{Z}_2 = \underline{a}' \, \underline{\bar{x}}_2 = a_1\bar{x}_{22} + a_2\,\bar{x}_{22} + \cdots + a_p\bar{x}_{p2}$ (3.13)

$[\,\bar{Z}_1 - \bar{Z}_2\,]^2 = [\,\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,]^2 = \underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,\underline{a}$ (3.14)

$\sum_{i=1}^{n_1} [\,Z_{i1} - \bar{Z}_1\,]^2 = \underline{a}' [\sum_{i=1}^{n_1} (Z_{i1} - \bar{Z}_1)' (Z_{i1} - \bar{Z}_1)]\,\underline{a} = \underline{a}' (n_1 - 1)\, S_1\, \underline{a}$ (3.15)

$\sum_{i=1}^{n_2} [\,Z_{i2} - \bar{Z}_2\,]^2 = \underline{a}' [\sum_{i=1}^{n_2} \left(Z_{i2} - \bar{Z}_2\right)' (Z_{i2} - \bar{Z}_2)\big]\underline{a} = \underline{a}'(n_2 - 1)\, S_2\, \underline{a}$ (3.16)

S_1, S_2 Matrix of variance and estimated variance within each of the first and second sets.

Find the differences within the groups.

$\sum_{i=1}^{n_1} [\,Z_{i1} - \bar{Z}_1\,]^2 + \sum_{i=1}^{n_2} [\,Z_{i2} - \bar{Z}_2\,]^2 = \underline{a}' (n_1 - 1)\, S_1\underline{a} + \underline{a}' (n_2 - 1)\, S_2\, \underline{a} =$

$(n_1 + n_2 - 2)\, \underline{a}' S\, \underline{a}$ (3.17)

Assuming that S is the variance –covariance matrix within the two sample

$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}$

$S\,(n_1 + n_2 - 2) = (n_1 - 1)S_1 + (n_2 - 1)S_2$

Find the value of Q, we compensate in equation (3.11−3.12)

$Q = \left[\frac{\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,\underline{a}}{(n_1 + n_2 - 2)\, \underline{a}' S\, \underline{a}} \right] = \frac{1}{(n_1 + n_2 - 2)} * \frac{\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,\underline{a}}{\underline{a}' S\, \underline{a}} = \frac{1}{(n_1 + n_2 - 2)}\, Q^*$ (3.18)

Is the ration to be maximize: Q^*

The condition can therefore be settled $\underline{a}' S\, \underline{a} = 1$

That is, the variance within the groups is constant and we maximize the variance between the groups' $\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,\underline{a}$ to be used by the Lagrange multiplier in the function and it becomes as follows:

$\mathcal{L}\,(\underline{a}, \lambda) = \underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)\,\underline{a} - \lambda\,(\underline{a}' S\, \underline{a} - 1)$ (3.19)

$\frac{\partial \mathcal{L}}{\partial \underline{a}} = \underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) = \lambda\, S\, \underline{a}$ (3.20)

After the calculations, we find value.

$$\lambda = \frac{\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \underline{a}}{\underline{a}' S \underline{a}} \tag{3.21}$$

$$\lambda = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) = D^2 \tag{3.22}$$

Mahalanobis distance, a measure of distances between group centers D^2

Where: $\underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) D^2$ (3.23)

$$\therefore \underline{a}' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) = \lambda S \underline{a}$$

$$\therefore \lambda S \underline{a} = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) D^2$$

$$\underline{a} = S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \tag{3.24}$$

The linear discriminating function (Fischer) is in the form

$$Z = \underline{a}' \underline{x} = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{x}$$

After this, the significance of the linear discriminating function is tested. After the extraction of the coefficients, attention is focused on the robust of the linear discriminating function and it distinguish. This can be determined by the analysis of variance,

Sum squares between groups equals: $Between\ \ S.S = \left(\frac{n_1 n_2}{n_1 + n_2}\right) [D^2]^2$

K. The degree of freedom

Sum squares within groups: Within $S.S = D^2$

The degree of freedom $(n_1 + n_2 - k - 1)$

Mean squares between groups: $Mean\ S.S = \frac{\left(\frac{n_1 n_2}{n_1 + n_2}\right)[D^2]^2/\mathrm{k}}{D^2/\,(n_1 + n_2 - \mathrm{k} - 1)}$ (3.25)

The latter ratio follows the distribution of F with freedom degree K and$(n_1 + n_2 - k - 1)$.When this is significant, this reference that the function is correct and that the variables used give the best distinction between the two groups.

To extract the linear discrimination function in the classification of observations belonging to unknown groups, we are the first step is to find the median values of the distinguishing function of the two samples as follows:

$$\bar{Z}_i = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{\bar{x}}_i \qquad i = 1, 2$$

Then find the average mean linear discrimination function

$$\bar{Z}_1 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{\bar{x}}_1$$

$$\bar{Z}_2 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{\bar{x}}_2$$

$$\text{Cut Point (CP)} = \frac{\bar{Z}_1 + \bar{Z}_2}{2} = \frac{(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)}{2} \tag{3.26}$$

To facilitate the classification, the breakpoint can be subtracted from the value given by the discriminant function to be so the cutoff point is zero and is as follows:

$$\therefore W = Z - CP$$

$$\therefore W = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} \underline{x} - \frac{1}{2} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2) \tag{3.27}$$

$$W = \begin{cases} > 0 \text{ Group I} \\ < 0 \text{ Group II} \\ 0 \ \text{ Can not classify} \end{cases}$$

It is the last one represents the value of the discriminative function at the boundary between the two groups, so if W> 0, then the item returns to the first group, but if W <0, then the item returns to the second group, but if W = 0 there is not classify, Since (W) is called the statistic of the classification that it has set (Afifi and Clark, 1984)

Another distinguishing function of Rao is used in the case of two or more groups, since the Fisher function is only for two groups, figure (1) explains the classification steps.

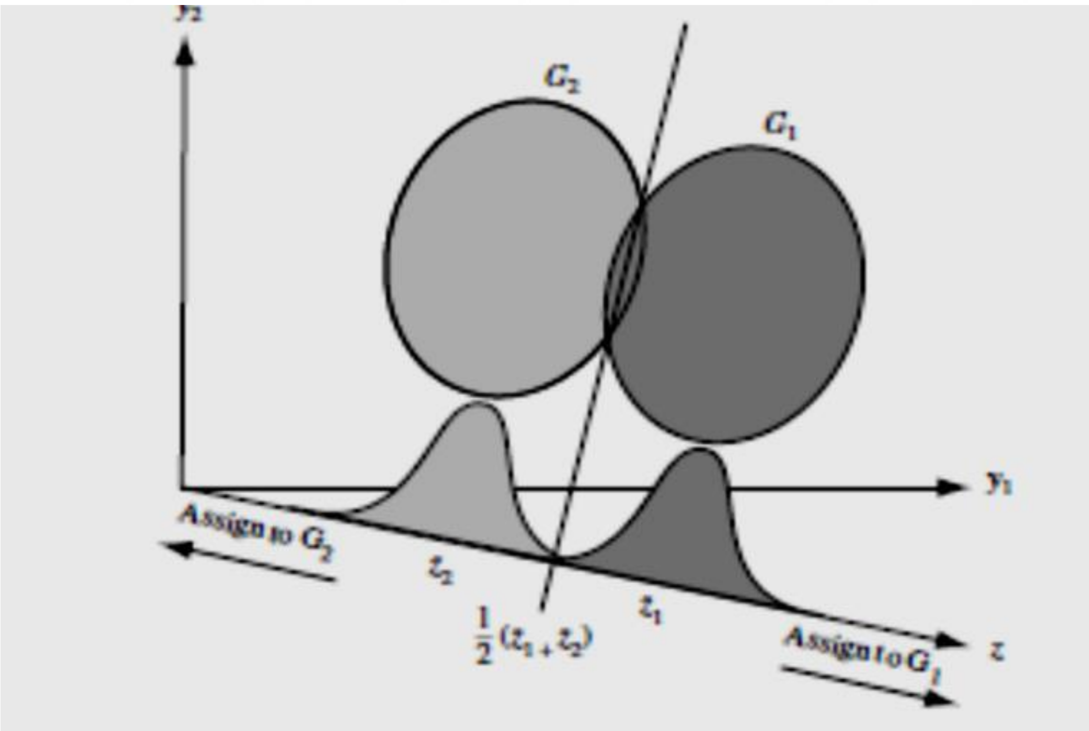

Figure 3.1. Classification steps (Abdel-Karim, 2006; Alkarkhi and Easa, 2008; Talib et al., 2019).

3.2.4. Testing the significance of the linear discriminant function

When the distinction between two groups is to be tested, the following Hypothesis can be tested:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

The test statistic used in the case of the distinction between two groups, (Hoteling -T^2) and its formula as follows

$$T^2 = \frac{n_1 n_2}{n_1+n_2} D^2$$

Where D^2represents (Mahalanobis Distance) and its formula is as follows

$$D^2 = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

$$T^2 = \frac{n_1 n_2}{n_1+n_2} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S^{-1} (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)$$

The (F) test is used, and its formula is as follows:

$$F = \frac{(n_1+n_2 - k-1)}{(n_1+n_2 - 2)k} * T^2 \qquad (3.28)$$

With the degree of freedom$(K, n_1 + n_2 - k - 1)$, we reject H_0 at a significant level F_α If:

$F_{cal} > F_\alpha, (K, n_1 + n_2 - k - 1)$

We accept H_1 and this indicates that the mean of the groups is not equal and that there are significant differences between the two groups. This means that the linear discriminant function is highly distinguishable.

The Wilks-Criteria can also be used according to the following equation.

$$\Lambda = \frac{|w|}{|T|} = \frac{|w|}{|w+B|} \qquad (3.29)$$

W: Matrix of variance –covariance in groups

T: Matrix of variance and total variance of groups

B: Matrix of variance –covariance between groups

The value of Λ is between zero and one, so if it is close to or equal to one, this indicates that the mean of the groups are equal and thus there is no distinction between the groups, and this means that the computed discriminant function is unsuccessful. But if its value is close to zero, then that indicates the strength of discrimination. (Pyryt, 2004).

3.2.5. Probability of classification error

There are two types of Probability error rating, and they are:

1- The probability of the P_{12}classification error is the possibility of classifying the individual into the second group, while it originally belongs to the first group.

2- The probability of the P_{21}classification error and the probability of classifying the individual into the first group, while it originally belongs to the second group, and thus the rating probability will be as follows:

$$P_{21} = P_{12} = \Phi\left(-\frac{D}{2}\right) \tag{3.30}$$

Where Φ represents the normal distribution function

D: is the root of the Mahalanobis scale D^2 (Hardle and Hlavka, 2007).

3.3. Receiver Operating Characteristic (ROC) Curves

Receiver operating characteristic (ROC) curve is the plot that depicts the trade-off between the sensitivity or true positive rate (TPR) on y-axis and (1-specificity) or false positive rate (FPR) on x-axis across a series of cut-off points when the diagnostic test is continuous or on ordinal scale (minimum 5 categories) (Kumar and Indrayan, 2011). Is an effective method of evaluating the quality or performance of diagnostic tests, and is widely used in radiology to evaluate the performance of many radiological tests. Receiver operating characteristic (ROC) analysis is a tool used to describe the discrimination accuracy of a diagnostic test or prediction model. While sensitivity and specificity are the basic metrics of accuracy, they have many limitations when characterizing test accuracy, particularly when comparing the accuracies of competing tests (Obuchowski and Bullen, 2018). ROC analysis is commonly used for characterizing the accuracy of medical imaging techniques (e.g. mammograms for detecting breast cancer, low-dose Computed Tomography to detect lung cancer), as well as non-imaging diagnostic tests (fasting glucose test for detecting diabetes, exercise stress test for detecting coronary artery disease), and prediction/risk scores (e.g. Framingham risk score to predict risk of cardiac events) in a variety of settings including screening, diagnosis, staging, prognosis, and treatment (Shiraishi et al, 2009).

3.3.1. Advantages of the ROC curve

ROC curve has following advantages compared with single value of sensitivity and specificity at a particular cut-off.

1. The ROC curve displays all possible cut-off points, and one can read the optimal cut-off for correctly identifying diseased or non-diseased subjects as per the procedure given later.

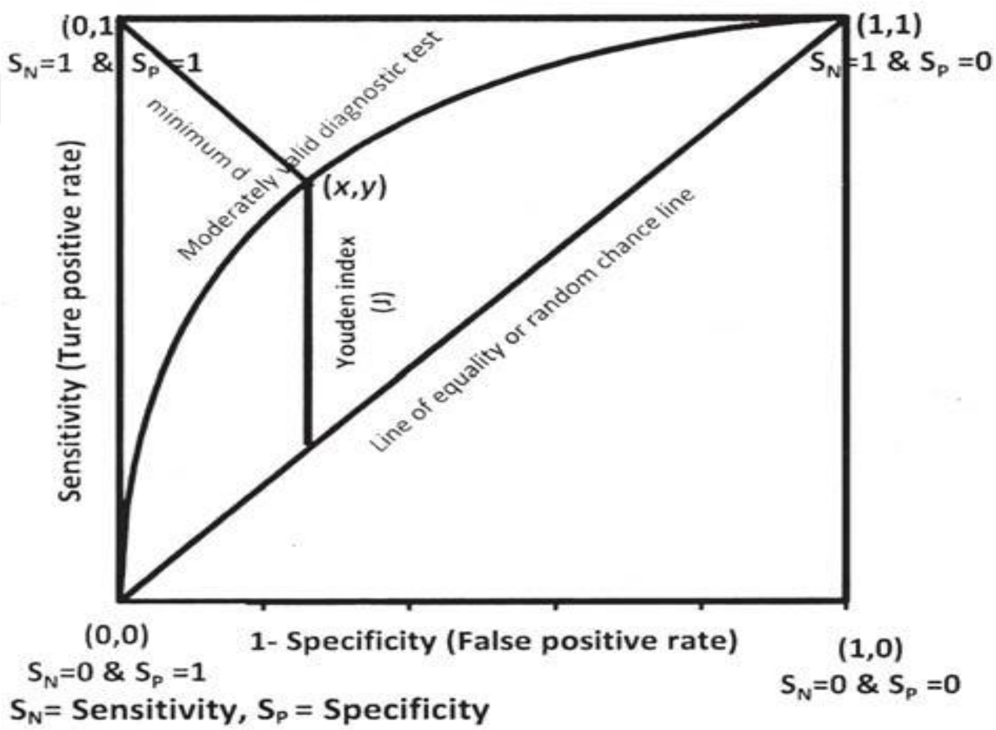

Figure 3.2. ROC curve and its components (Kumar and Indrayan, 2011).

2. The ROC curve is independent of prevalence of disease since it is based on sensitivity and specificity which are known to be independent of prevalence of disease (Hanley and McNeil, 1982).
3. Two or more diagnostic tests can be visually compared simultaneously in one figure.
4. Sometimes sensitivity is more important than specificity or vice versa, ROC curve helps in finding the required value of sensitivity at fixed value of specificity.

5. Useful summary of measures can be obtained for determining the validity of diagnostic test such as AUC and partial area under the curve.
6. Empirical area under the ROC curve (explained later) is invariant with respect to the addition or subtraction of a constant or transformation like log or square root. Log or square root transformation condition is not applicable for binormal ROC curve. Binormal ROC is also shortly explained. (Campbell, 1994).

3.3.2. The confusion matrix

The classification matrix is a statistical indicator of the suitability of the model and thus its compatibility with the data. Where it works on classification of binary events by using the confusion matrix, which shows the actual versus predicted affiliation of each group (Soderstorm and Leitner, 1997). The use of this analysis is based on the assumption that if the model expects to classify the cases correctly based on a criterion, it gives proof that the model matches the observed data. The general format of the classification table (3.1) is:

Table 3.1. Confusion matrix

Classification		Observations	
		Positive	Negative
Prediction	Positive	True positive (TP) (sensitivity)	False positive (FP)
	Negative	False negative (FN)	True negative (TN) (specificity)

We can obtain from this matrix the following rating accuracy criteria:

1-Accuracy

Accuracy is the measure of how good our model is. It is expected to be closer to 1, if our model is performing well.

$$\text{Accuracy} = \frac{TP+TN}{N} \tag{3.31}$$

Whereas:-

TN: The number of samples classified as negative (does not have the characteristic) is actually negative.

TP: The number of samples classified as positive (possessing the characteristic) is in fact positive.

N: Total number of samples.

2-Sensitivity

Also called true positive (TP) a measurement of the proportion of positives that are correctly predicted given it is truly positive,

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{3.32}$$

Where FN the number of samples classified as negative is actually positive, and the model becomes more sensitive if it can identify the largest number of positive cases.

3-Specificity

Also called true negative (TN) a measure of the proportion of predicted negative, given it is truly negative). It is quite obvious that a better classifier is expected to have both higher sensitivity and specificity. Note that specificity is 1 – FP.

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{3.33}$$

Where FP the number of samples classified as positive is actually negative, and the model becomes more specific if it can determine the largest number of negative cases.

4-Error rate

$$\text{Error rate} = \frac{FP+FN}{N} \tag{3.34}$$

5-Positive predictive value (PPV)

Precision is defined as how many selected items are relevant. That is, how many of the predicted ones are actually correctly predicted (Ferrer and Wang, 1999).

$$\text{(PPV)} = \frac{TP}{TP+FP} \qquad (3.35)$$

6-Negative predictive value (NPV)

$$\text{(NPV)} = \frac{TN}{TN+FN} \qquad (3.36)$$

3.3.3. Area under the ROC Curve (AUC)

Is a measure of the overall performance of a diagnostic test and is interpreted as the average value of sensitivity for all possible values of specificity (Zhou et al., 2009).

If we choose a discriminating cut-off value for the predictive variable to be less than the lowest value observed, we generate the (0, 0) point in the ROC space. As we increase the discriminating cut-off value to include more and more data points, we generate a series of points within the ROC space that can be connected by a curve. A discriminating cut-off value greater than the highest value observed generates the (1, 1) point. The diagonal line connecting the (0, 0) point and the (1, 1) point indicates test predictions no better than random guesses. The further a point in the ROC space is above the diagonal line, the better the predictive value of the test.

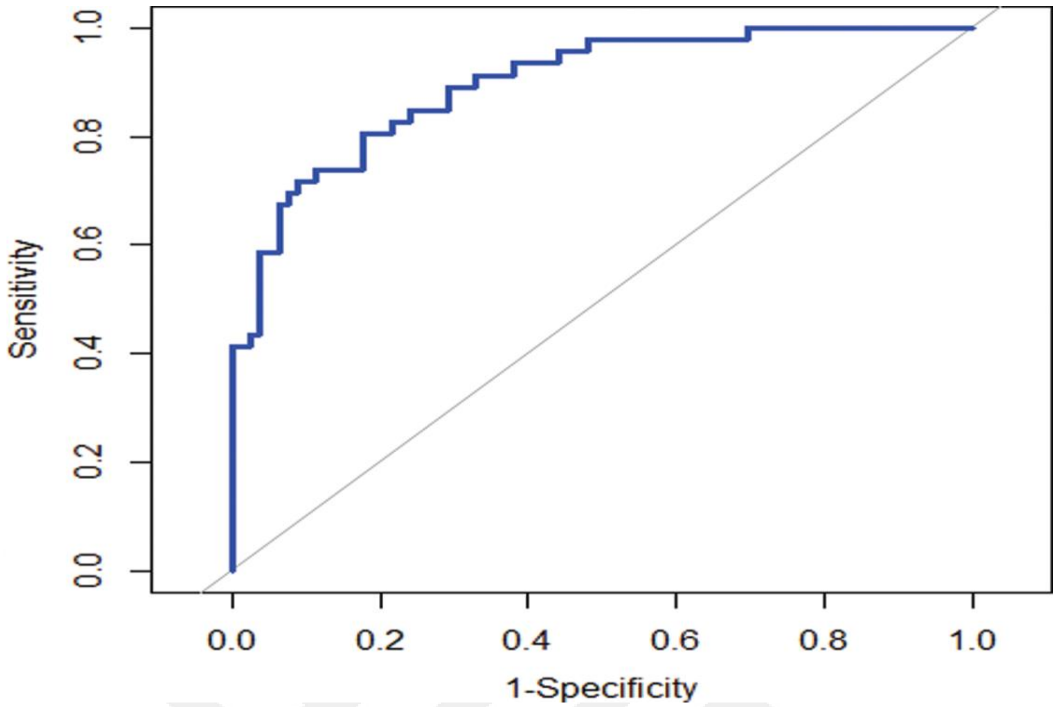

Figure 3.3. Hypothetical ROC curve demonstrating the tradeoff between sensitivity and specificity (Yang and Berdine, 2017).

Particularly, sensitivity and specificity are inversely related, i.e., as the sensitivity increases, the specificity decreases, and vice versa. For example, if we use a lower hemoglobin cut-off value, more non-anemic patients will be considered as normal, and thus the true negative rate is higher (i.e., higher specificity); meanwhile, fewer anemia patients will be considered as having the disease, and thus the proportion of true positive is lower (i.e., lower sensitivity). Similarly, if we use a higher cut-off value, then we will have lower specificity and higher sensitivity.

The AUC (also known as the c-statistic) can be used to evaluate the diagnostic ability of a test to discriminate the true disea se status of a patient. In general, the rule of thumb for interpreting AUC value is, if AUC = 0.5 there is no discrimination, $0.6 \geq AUC > 0.5$ = Poor discrimination, $0.7 \geq AUC > 0.6$ = Acceptable discrimination, $0.8 \geq AUC > 0.7$ = Excellent discrimination, $AUC > 0.9$ = Outstanding discrimination.

In addition, AUC can also be used to find the optimal cut-off value for a specific test, as well as compare the performance between two or more alternative tests (Yang and Berdine, 2017).

4. RESULTS AND DISCUSSION

This study depended on questionnaire forms and was distributed randomly among 370 patients in all departments of eleven different hospitals including public and private hospitals in Erbil city in the year 2020.Moreover, stratified random sampling has been used and analyzed data by IBM SPSS Statistics V: 25 program. In the first stage, descriptive statistics reflecting the mean and standard deviation for each service quality structure were used to improve understanding of the public and private hospital classification of quality health care for patients against each service quality dimension. At the second stage, factor analysis was performed to extract the most important variable for public and private hospitals. At the third stage discriminant analysis was performed to distinguish between public and private hospital .At the fourth stage ROC curve criterion was performed to evaluating the discriminant model.

4.1. Demographic Profile of Patients

To demographic profile of the patients, we used descriptive statistics which are focused on frequency for eight variables such us (gender, age, marital status, occupation, economic condition …etc.). After that we used pie chat for each variables to measure value or frequency, graph 4.1 to 4.8 illustrate the demographic respondent profile of the patient.

4.1.1. Gender of patients

Figure 4.1 shows the demographic profile of the patients. A total of 370 respondents by gender represented that, (41.9%) 155 of the patients were male and (58.1%) 215 were female.

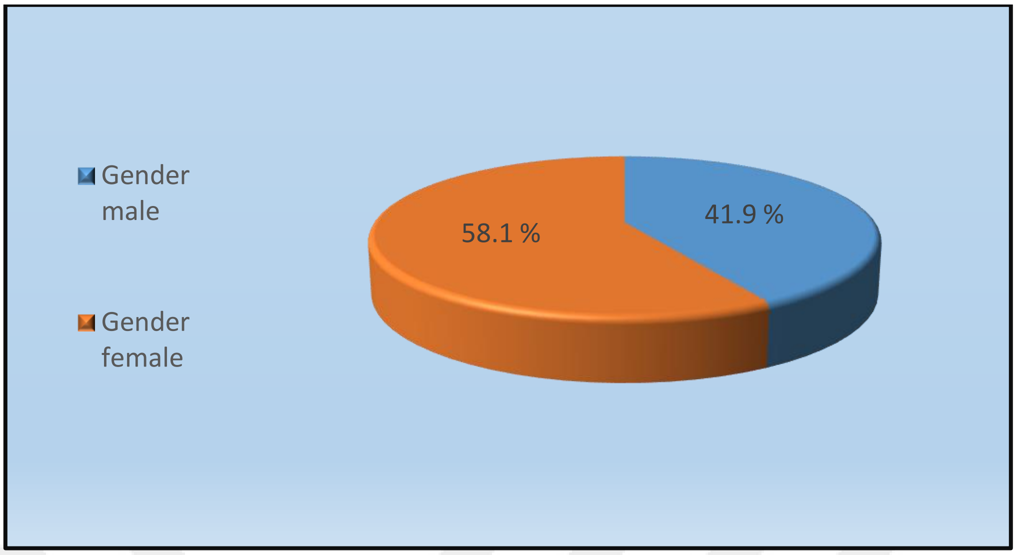

Figure 4.1. Gender of patients.

4.1.2. Age of patients

Figure 4.2 shows the age patient that (18.1%)67 of the patient's age were under 25 years old, (26.8%)99 of the patient's age were 25–39 years old, (20%)74 of the patient's age were 40-54 years old and (35.1%)130 were equal or more than 55 years old.

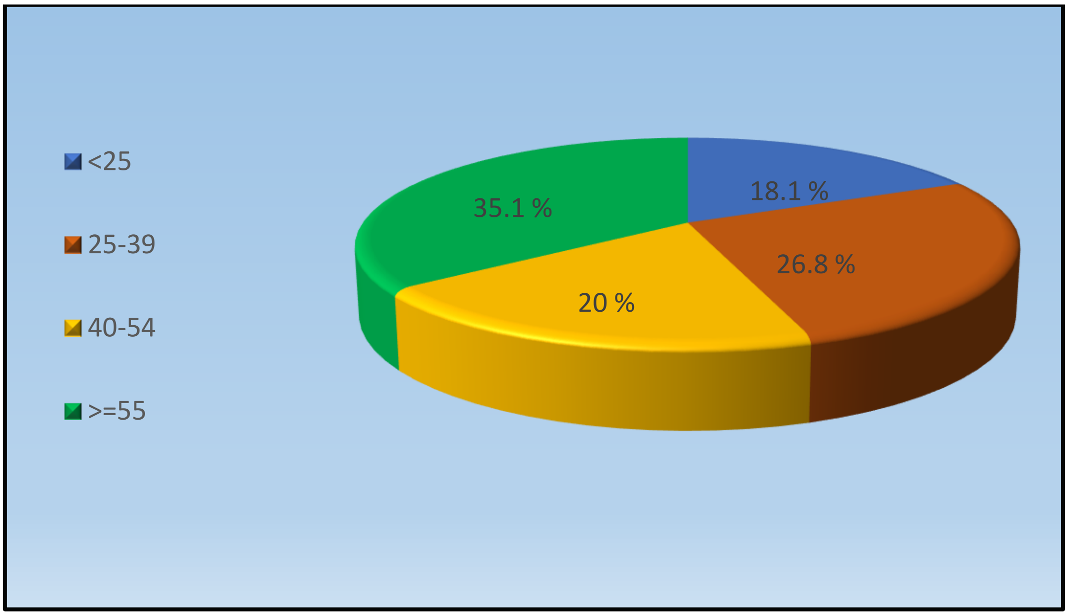

Figure 4.2. Age of patients.

4.1.3. Marital status of patients

Figure 4.3 shows marital status for the patients that (70.5%) 261 of the patients were married,(15.9%) 59 of the patients were single,(0.8%) 3 of the patients were divorce and (12.7%)47 were another status.

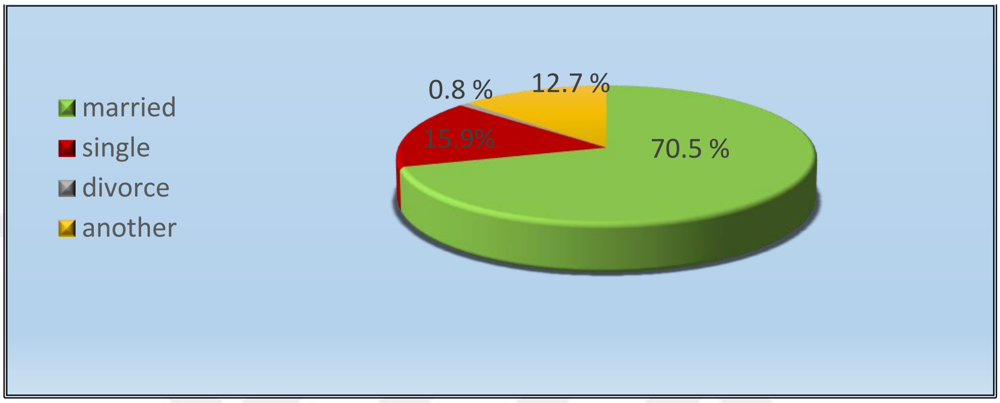

Figure 4.3. Marital status of patients.

4.1.4. Occupation of patients

Figure 4.4 shows the output of occupation represented that (14.1%) 52 of the patients were employee, (13.8%) 51 patients were retired, (37.8%) 140 patients were wife home, (17.3%) 64 patients were worker, (7.6%) 28 patients were impotence and (9.5%) 35 were students.

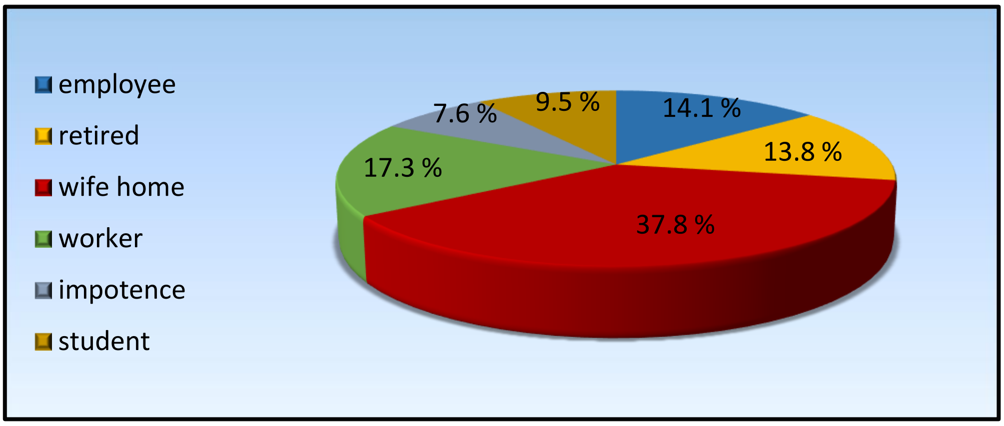

Figure 4.4. Occupation of patients.

4.1.5. Economic condition of respondents

Figure 4.5 shows economic condition for the patients that (24.6%) 91 of the patients had bad economic, (46.2%) 171 patients had middle economic, (28.1%)104 patients had good economic and(1.1%) 4 patients had very good economic.

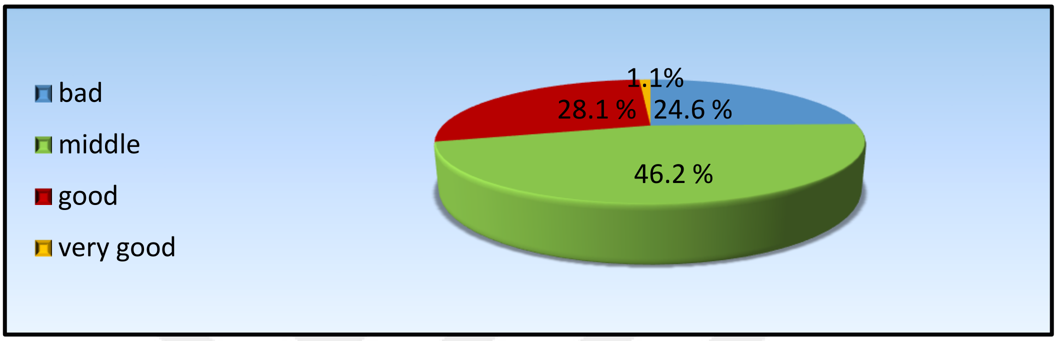

Figure 4.5. Economic condition of respondents.

4.1.6. Education of respondents

Figure 4.6 shows education for the patients that (36.8%) 136 of the patient's illiterate, (14.3%) 53 patients could just read and write, (10%) 37 patients had primary degree, (13.5%) 50 patients had intermediate degree, (7.6%) 28 patients had secondary degree, (7.6%) 28 patients had institute degree, (9.2%) 34 patients had bachelor degree and(1.1%) 4 patients had higher certificate.

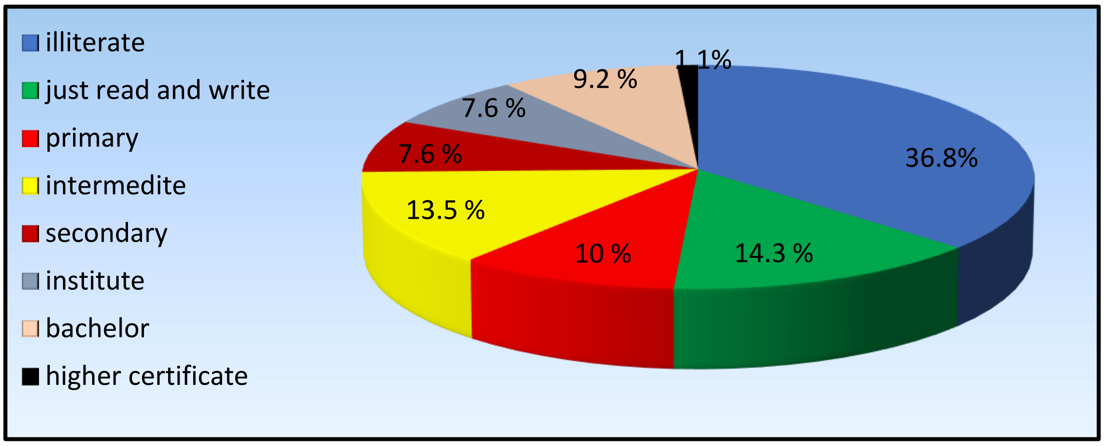

Figure 4.6. Education of respondents.

4.1.7. Respondents of hospital type

Figure 4.7 shows the contribution of the patients in the hospital type. A total of 370 patients that (50.8%) 188 of the patients in public hospital and (49.2%) 182 of the patients in private hospital.

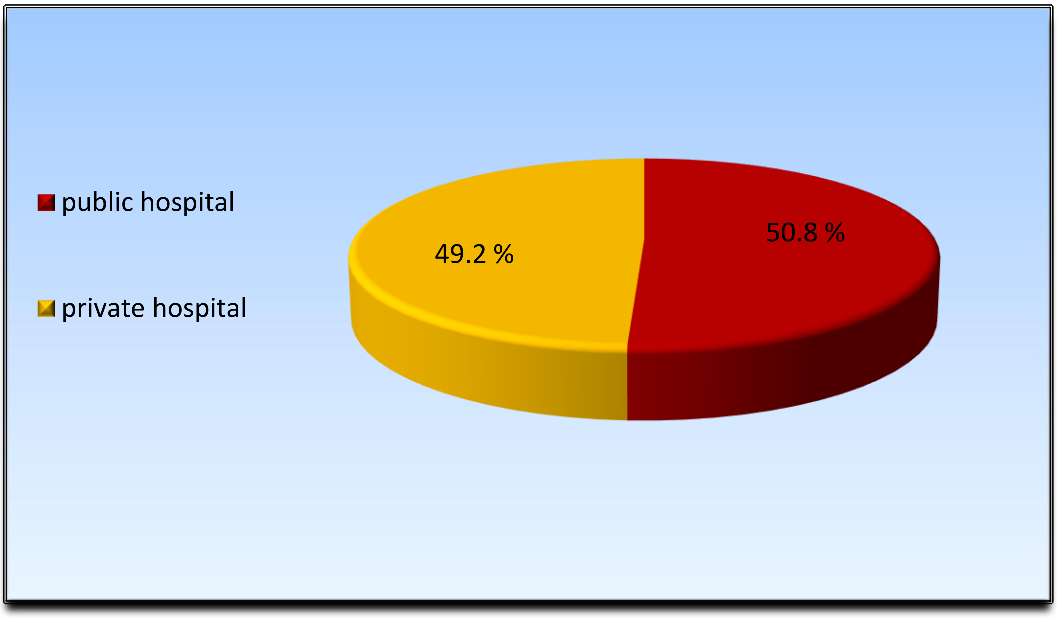

Figure 4.7. Contribution of the patients in the hospitals.

4.1.8. Respondents divided by hospitals

Figure 4.8 shows patients that are divided by eleven hospitals. A total of 370 patients that (14.6%) 54 patients in Rizgare Teaching Hospital, (9.5%) 35 patients in West Emergency Hospital, (13.2%) 49 patients in Erbil Teaching Hospital, (13.5%) 50 patients in East Emergency Hospital, (10.5%) 39 patients in Zheen Hospital, (10.5%) 39 patients in CMC Hospital, (8.1%) 30 patients in Sardam Hospital, (8.1%) 30 patients in Paky Hospital, (3.8%) 14 patients in Zanko Hospital, (4.1%) 15 patients in Balsam Hospital, (4.1%) 15 patients in Hawler Hospital.

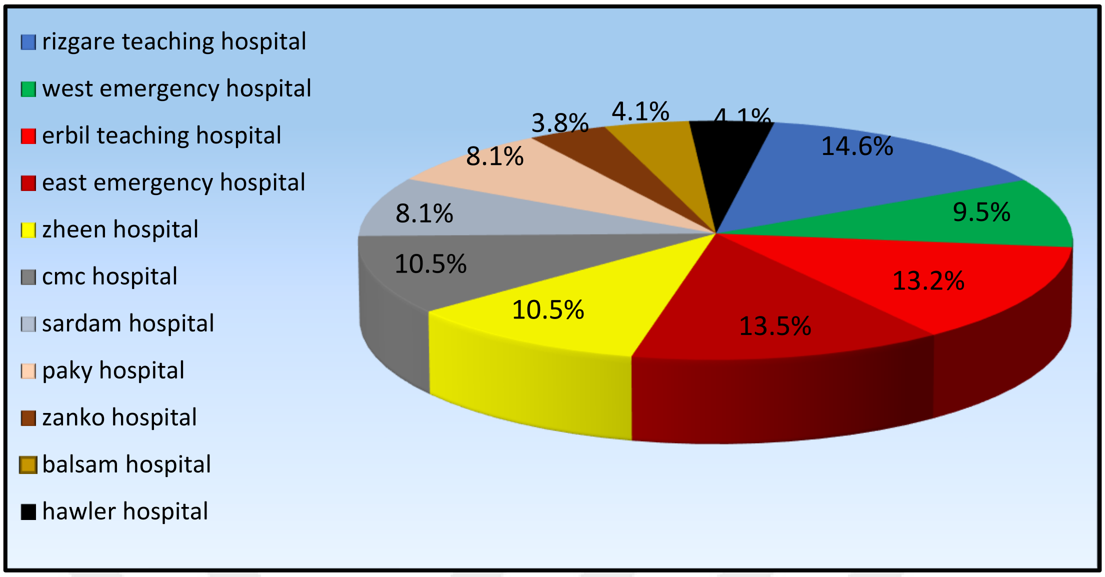

Figure 4.8. Respondents divided by hospitals.

4.2. Reliability Test

Reliability is the degree to which the questionnaire, the evaluation, the assessment or any measurement technique generates the same results in a repeat test. In short, the stability or accuracy of scores over time or through raters (Bolarinwa, 2015). Reliability was tested using Cronbach's alpha (α), the Rasch version of Cronbach's α, except that it is measured from the person's logit scale estimates. (Andrich and Marais, 2019). It is suggested that $\alpha \geq 0.90$ = excellent, $0.90 > \alpha \geq 0.80$ = good, $0.8 > \alpha \geq 0.7$ = acceptable, $0.7 > \alpha \geq 0.6$ = questionable, $0.6 > \alpha \geq 0.5$ = poor, and $\alpha < 0.5$ = unacceptable (Hansen and Kjaersgaard, 2020). The reliability test for our data as shown in the following table:

Table 4.1. Reliability test

Cronbach's alpha	N of Items
0.788	22

Table 4.1 shows that total number of items is 22 items and Cronbach’s alpha is 0.788 .Therefore, the data is acceptable to analyze.

4.3. Descriptive Statistics

Table 4.2 shows the descriptive statistics for all variables (quality health service) the mean and standard deviation, count and percent for Likert scale (strongly disagree, disagree, neutral, agree, strongly agree).

Table 4.2. Descriptive Statistics

Var	strongly disagree		disagree		neutral		agree		strongly agree			
	Count	%	Count	%	Count	%	Count	%	Count	%	Mean	S. D
X1	0	0.0	9	2.4	27	7.3	218	58.9	116	31.4	4.19	0.669
X2	2	0.5	13	3.5	29	7.8	261	70.5	65	17.6	4.01	0.667
X3	0	0.0	0	0.0	15	4.1	211	57.0	144	38.9	4.35	0.556
X4	0	0.0	15	4.1	136	36.8	174	47.0	45	12.2	3.67	0.739
X5	2	0.5	21	5.7	50	13.5	153	41.4	144	38.9	4.12	0.887
X6	0	0.0	7	1.9	22	5.9	211	57.0	130	35.1	4.25	0.650
X7	0	0.0	1	0.3	4	1.1	292	78.9	73	19.7	4.18	0.432
X8	0	0.0	3	0.8	14	3.8	238	64.3	115	31.1	4.26	0.562
X9	0	0.0	2	0.5	4	1.1	259	70.0	105	28.4	4.26	0.498
X10	0	0.0	1	0.3	6	1.6	227	61.4	136	36.8	4.35	0.525
X11	0	0.0	2	0.5	12	3.2	242	65.4	114	30.8	4.26	0.541
X12	0	0.0	0	0.0	9	2.4	300	81.1	61	16.5	4.14	0.412
X13	0	0.0	1	0.3	6	1.6	122	33.0	241	65.1	4.63	0.532
X14	1	0.3	4	1.1	19	5.1	248	67.0	98	26.5	4.18	0.592
X15	0	0.0	4	1.1	6	1.6	290	78.4	70	18.9	4.15	0.476
X16	0	0.0	54	14.6	37	10.0	238	64.3	41	11.1	3.72	0.847
X17	1	0.3	8	2.2	25	6.8	230	62.2	106	28.6	4.17	0.662
X18	0	0.0	6	1.6	19	5.1	216	58.4	129	34.9	4.26	0.629
X19	0	0.0	6	1.6	19	5.1	307	83.0	38	10.3	4.02	0.468
X20	0	0.0	8	2.2	23	6.2	253	68.4	86	23.2	4.13	0.605
X21	1	0.3	6	1.6	31	8.4	294	79.5	38	10.3	3.98	0.525
X22	0	0.0	19	5.1	40	10.8	258	69.7	53	14.3	3.93	0.673

4.4. Factor Analysis

Factors analysis used to describe the basic components that clarify the relationships between a numbers of variables. Furthermore, the study of the principal components is used to derive full variation from the data set for each component, reducing a large number of variables to a smaller number of components. (Tabachnick and Fidell, 2007).

The first step we used Kaiser-Meyer-Olkin (KMO) and Bartlett's test to know the data is suitable for factor analysis. The KMO index ranges from 0 to 1, with 0.50 considered to be sufficient for factor analysis. The Bartlett Test Sphericity should be probability of value ($p<.05$) in order for the factor analysis to be sufficient.

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

Table 4.3. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.748
Bartlett's Test of Sphericity	Approx. Chi-Square	168.749
	Df	171
	Sig	0.000

Table 4.3 shows that the Kaiser-Meyer-Olkin (KMO) measure is 0.748, Greater than 0.5, it indicates that sample is good for sampling. And Bartlett's test is significant at 0.000, less than 0.05 than the correlation matrix is significant, which indicates that the data is sufficient for factor analysis.

Table 4.4. Communalities

variable	Initial	Extraction
X1	1	0.548
X2	1	0.663
X3	1	0.531
X4	1	0.665
X5	1	0.616
X6	1	0.577
X7	1	0.652
X8	1	0.828
X9	1	0.557
X10	1	0.556
X11	1	0.838
X12	1	0.686
X13	1	0.661
X14	1	0.644
X15	1	0.628
X16	1	0.610
X18	1	0.593
X19	1	0.791
X20	1	0.600

The segment Communalities presents before rotation the communality of each component, the communality of a variable is 1 for all the variables, the entire initial communalities are higher than 0.50, which is good. We remove variables (X17, X21, X22) because their variance smaller than 0.5 and don't have effect on the analysis, Table 4.4 above shows how much variation the extracted factors accounted for in the variables.

The total variance explained section presents how many components (factor) to extract. We used the criterion of retaining only factor with eigenvalue of 1 or greater, in Table 4.5 shows the first seven components eigenvalue greater than 1 will be retained for rotation. These seven components account for (12.892%, 10.892%, 9.513%, 8.879%, 8.660%, 7.766% and 5.841) of the total variance. These seven components explain a total 64.449 per cent of the variance. It can also be inferred these seven components have influenced people's satisfaction from public and private hospitals.

Table 4.5. Total variance

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.081	21.481	21.481	2.450	12.892	12.892
2	1.724	9.076	30.557	2.070	10.897	23.789
3	1.597	8.405	38.962	1.808	9.513	33.302
4	1.409	7.415	46.377	1.687	8.879	42.181
5	1.277	6.721	53.098	1.645	8.660	50.841
6	1.132	5.956	59.055	1.476	7.766	58.608
7	1.025	5.394	64.449	1.110	5.841	64.449
8	0.835	4.394	68.843			
9	0.780	4.104	72.947			
10	0.728	3.830	76.776			
11	0.684	3.601	80.378			
12	0.617	3.245	83.623			
13	0.602	3.169	86.792			
14	0.558	2.939	89.731			
15	0.507	2.669	92.400			
16	0.457	2.406	94.806			
17	0.437	2.299	97.105			
18	0.401	2.108	99.213			
19	0.149	0.787	100.000			

Extraction Method: Principal Component Analysis.

4.4.1. Scree test

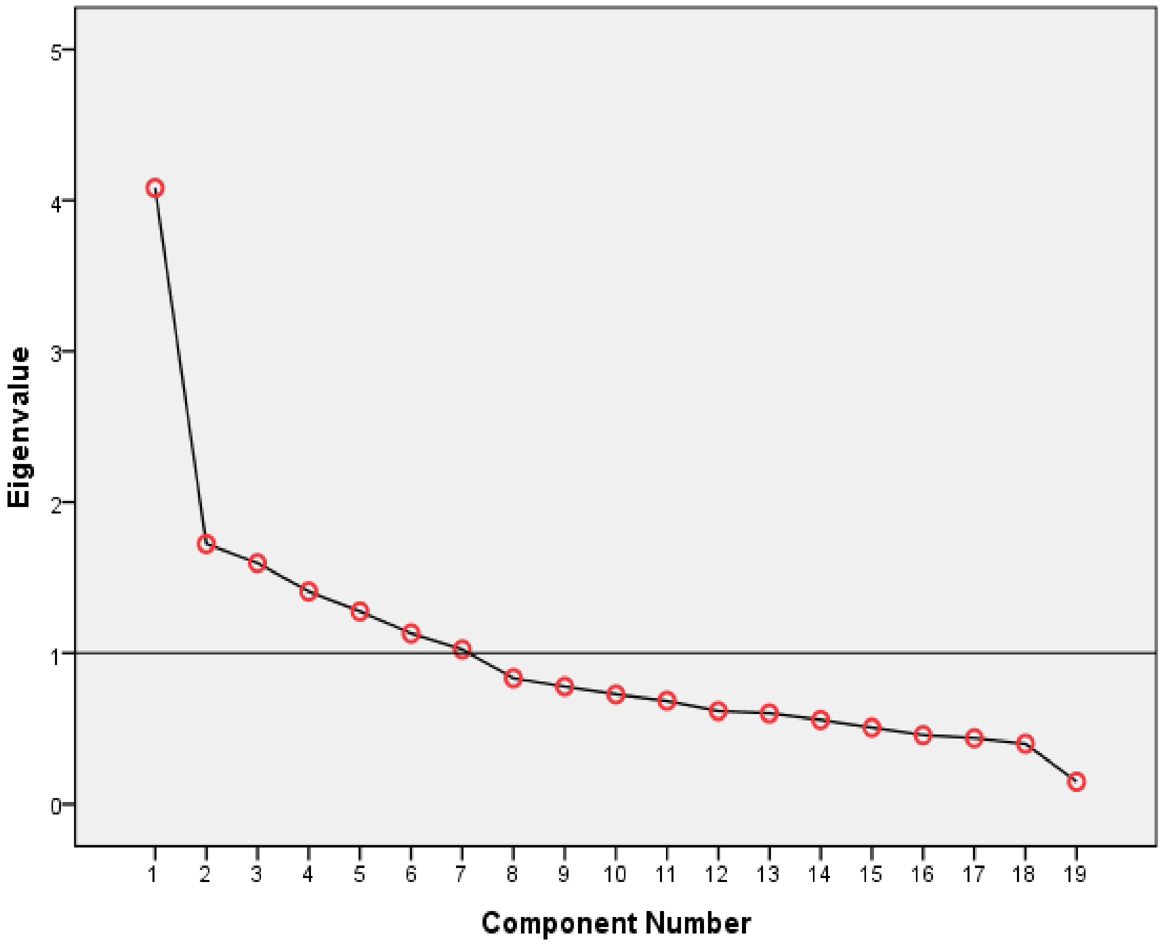

Figure 4.9. Scree plot.

The scree test is another way of identifying the number of valuable components for public and private hospitals. In Figure 4.9 we can look at the change in shape of the plot there is quite clear break between second and third components. Component 1 explain much of the variance than the other components. After that there is also another little break after seventh component, explains the first seven components the eigenvalue above 1 explains much of the variance than the other components.

After varimax rotation, the rotated component matrix shows the seven components. To determine what these components represent, it would be necessary to consider the variable loaded on each of the seven components. We used absolute value (±0.5) to identify high loadings. Table 4.6 shows that the first component is more important than the second component and so on. Therefore, the need for the hour is for public and private hospitals to reflect on these variables in order to attract patients.

Table 4.6. Rotated component matrix

component	% of Variance	Factor Interpretation	Variables included in the factors	Loading
C1	12.895	Tangibility and Regularity of Services	The hospital staff provide prompt services.	0.899
			Providing timely services.	0.896
			Modern equipment in the hospital.	0.695
C2	10.897	Assurance of Services	Patients feel safe in the hospital.	0.769
			Hospital staff demeanor instils confidence.	0.749
			The hospital staff communicate to patients about service provision.	0.698
C3	9.513	Responsiveness	Medical staff willingness to help patients.	0.795
			The hospital staff get things done the first time.	0.771
C4	8.879	Reliability	The hospital staff provide promised services.	0.767
			The hospital staff insist on error free records.	0.615
C5	8.660	Empathy	Hospital attractiveness.	0.788
			Hospital staff give patients individual attention.	0.743
C6	7.766	flexibility service	Hospital staff are always courteous towards Patients.	0.751
			Hospital staff give patients personal attention.	0.598
			Medical staff never too busy to respond to patients request.	-0.526
C7	5.841	Time and materials	Hospital has convenient opening hours.	0.840
			Attractiveness of medical materials.	-0.565

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

4.5. Discriminant Analysis

Before determining the discriminatory functions of the two groups (public and private), it is needed to do some tests that like the significance of the linear discriminating function and test the significance of the variables in the discriminatory function.

4.5.1. Test the significance of the linear discriminating function

When we want to discriminate between the two groups (public and private) is to be measured and the composition of statistically significant discriminatory functions is statistically significant, the two groups are equal to the mean is determined by the following hypothesis.

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Using the test tester (Wilks' Lambda) to test the above hypothesis.

$$\Lambda = \frac{SS_{withi\_group}}{SS_{total}}$$

Table 4.7. The significant test of the discriminatory function

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	0.624	171.183	9	0.000

Table 4.7 shown that the probability values (sig = 0.000) smaller than 0.05, we reject the null hypothesis and accept the alternative hypothesis, and that there is a significant difference between the two groups. This means that the linear discriminant function distinguishes between the two groups (public) and (private).

4.5.2. Test the significance of the variables in the discriminatory function

The significance of all variables is tested to find out the importance of each variable in the function, and its effect on the analysis of the results using the F test.

Table 4.8. Testing the significance of F for each variable in the discriminant function

Variable	Wilks' Lambda	F	df1	df2	Sig.
X1	0.971	11.008	1	368	0.001
X2	0.928	28.537	1	368	0.000
X4	1.000	0.005	1	368	0.942
X5	0.993	2.563	1	368	0.110
X7	0.981	7.070	1	368	0.008
X8	1.000	0.102	1	368	0.749
X9	0.961	15.129	1	368	0.000
X10	0.996	1.432	1	368	0.232
X11	1.000	0.119	1	368	0.731
X12	0.995	1.986	1	368	0.160
X13	0.998	0.813	1	368	0.368
X14	0.996	1.324	1	368	0.251
X15	0.991	3.437	1	368	0.065
X16	0.862	58.982	1	368	0.000
X18	0.999	0.213	1	368	0.645
X19	0.992	2.832	1	368	0.093
X20	0.816	82.759	1	368	0.000

Table 4.8 shows that the variables (X1, X2, X7, X9, X16, X20.) their probability values (sig ≤ 0.05) are highly significant and have a great effect in terms of differentiating between the two groups, while the remainder of the variables have no significant effect.

4.5.3. Estimation of linear discriminatory function

We used stepwise method for the most important 17 variables that we extracted at factor analysis from public and private hospitals and then we extract the most important 9 variable in the two groups, the results are shown in the following table:

Table 4.9. Canonical discriminant function coefficients

Variable	Function 1
X1	0.486
X2	0.601
X5	-0.285
X7	-0.852
X8	-0.417
X9	0.476
X16	0.654
X18	-0.340
X20	1.112
Constant	-5.532

Unstandardized coefficients

The relative value of independent variables in the prediction of dependent variables is shown in the Table 4.9 by canonical discriminant function coefficients. And unique predictor factors that are relevant in distinguishing between public and private hospitals. Coefficients with higher absolute value have a higher discriminating ability for the variable grouping. The discriminant function is:

$Z = -5.532 + 0.486X1 + 0.601X2 – 0.285X5 – 0.852X7 – 0.417X8 + 0.476X9 + 0.654X16 – 0.340X18 + 1.112X20.$

4.5.4. Differential function according to group means (cutoff point)

It can be observed that the linear discriminative function is assessed on the basis of the group averages as shown in the following table:

Table 4.10. Function at group centroids

Type hospital	Function 1
Public hospital	-0.761
Private hospital	0.786

To classify any term depending on the function (W), we substitute the values of the variables related to this term in the equation. If W> 0, then it returns to the first group (public), but if W <0, then the item returns to the second group (private), but if W = 0 there is not classify, that is:

CP = $\frac{\bar{Z}_1 + \bar{Z}_2}{2}$

$\therefore$ W = Z – CP

4.5.5. Calculate the probability of correct classification

Table 4.11 shows the results of the probability of correct classification of hospitals. In public hospital 139 were classified correctly out of 188 cases at 73.9%, and in the private hospital 150 were classified correctly out of the 182 cases at 82.4%. And that the correct classification rate for the model was 78.1% out of 370 cases, and this means that the probability of classification error is 21.9%.

Table 4.11. Classification results

			Predicted Group Membership		
		Type hospital	Public hospital	Private hospital	Total
Original	Count	Public hospital	**139**	49	188
		Private hospital	32	**150**	182
	%	Public hospital	73.9	26.1	100
		Private hospital	17.6	82.4	100

a.78.1 % of original grouped cases correctly classified.

4.6. Receiver Operating Characteristic (ROC) Curves

The receiver operating characteristic (ROC) curves indicates the ability of a marker or diagnostic test to discriminant between two groups of subjects. Sensitivity and specificity are the basic metrics of diagnostic test accuracy; however, they rely on the cut-off point used to identify "positive" and "negative" test results. Sensitivity and specificity change as the cut point changes, and the plot of a test's sensitivity versus its false-positive rate for all possible cut points. The figure ROC curve for our models is:

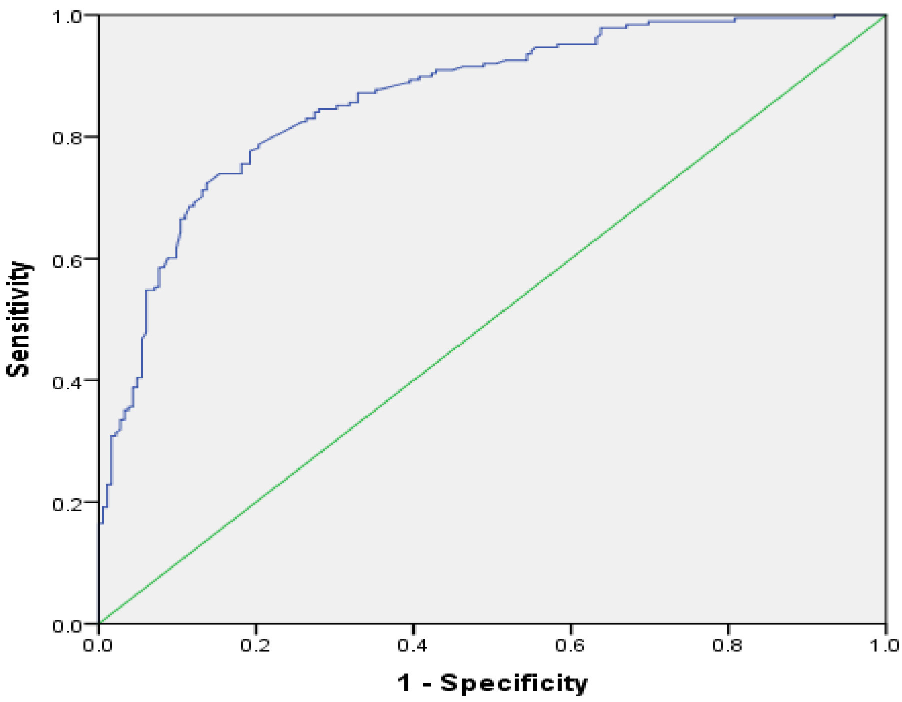

Figure 4.10. Receiver's operating characteristic curve demonstrating strong discriminatory strength.

The area under the ROC curve (AUC) is generally accepted as an indicator of the discriminatory power of a diagnostic test. Table 4.12 shows that the probability of value (sig $\leq$ 0.05), there is a significant different from the random area and AUC = 0.863 means that the diagnostic test is excellent discrimination.

Table 4.12. Area under ROC curve

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.863	0.019	0.000	0.827	0.900

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

4.7. Comparison Between Results of the Factor and Discriminant Analysis

In factor analysis we selected the first three most important components explains much of the variance than the other components, containing 8 variables and stepwise discriminant analysis extract the most important 9 variables from both type of hospitals. A table was then developed to compare their findings. Thus the relation is shown in the following table.

Table 4.13. Comparison between results of the factor and discriminant analysis.

Factor analysis variables	Discriminant analysis variables
X1, X7, X8, X10, X11, X12, X14, X15	X1, X2, X5, X7, X8, X9, X16, X18, X20

5. CONCLUSIONS

The mean of this study is classification and evaluation model of the quality health services between public and private hospitals and extract the most important variables to effect patient to choose which hospitals, by using factor and discriminant analysis with ROC curve, based on the results obtained.

In public and private hospitals the factor analysis identified seven components accounting for 64% of the total variance, the first component is most influential which explain the largest variance ratio containing three variables (tangibility and regularity of services) and the per cent of variance to this component is 12.892%. The second component containing three variables (assurance of services) and the per cent of variance to this component is 10.897%. The third component containing two variables (responsiveness services) and per cent of variance to this component is 9.513%. The fourth component containing two variables (reliability service) and the per cent of variance to this component is 8.879%. The fifth component containing two variables (empathy service) and per cent of variance to this component is 8.660%. The sixth component containing three variables (flexibility service) its rate interpretation is 7.766%. The seventh component containing two variables (time and materials) its rate interpretation is 5.841%.

Discriminant analysis showed that the key variable associated with the selection of a hospital, we used stepwise method for the most important 17 variables that we extracted at factor analysis from public and private hospitals and then we extract the most important 9 variable in the two groups, and the correct classification rate for the discriminant analysis is 78.1% distinguish between public and private hospitals. We used ROC curve criterion to evaluate the discriminant model and area under the roc curve. AUC was 0.863 means that the diagnostic test is excellent discrimination.

A convergence and similarity exists between factor analysis and discriminant analysis, factor analysis from public and private hospitals we selected the first three most important components explains much of the variance than the other components, containing 8 variables and stepwise discriminant analysis extract the most important 9 variables from both type of hospitals. Additionally, the results show that the variables (X1: Modern equipment in the hospital, X7: The hospital staff get things done the first time and X8: Providing timely services) are the most important and influential variables between the factor analysis and discriminant analysis for the patient satisfaction. Therefore, the government should take much care of these three variables to increase the quality of health care in both types of private and public hospitals.

REFERENCES

Abdel-Karim, N. D., 2006. ***Using Statistical Discriminatory Methods for Diagnosing Some Heart Diseases***. Master of science, to college management and economics, Mustansiriya university.

Abuosi, A. A., Atinga, R. A., 2013. Service quality in healthcare institutions: establishing the gaps for policy action. ***International Journal of Health Care Quality Assurance,*** **26**(5): 481-492.

Afifi, A. A., V. Clark, 1984. ***Computer Aided Multivariate Analysis***. Lifetime learning publications, Belmont, California, U .S .A.

Aghamolaei, T., Eftekhaari, T. E., Rafati, S., Kahnouji, K., Ahangari, S., Shahrzad, M. E., Hoseini, S. H., 2014. Service quality assessment of a referral hospital in southern Iran with SERVQUAL technique: Patients' perspective. ***BMC Health Services Research,*** **14**(1): 1-5.

Alijanzadeh, M., Zare, S. A. M., Rajaee, R., Fard, S. M. A. M., Asefzadeh, S., Alijanzadeh, M., Gholami, S., 2016. Comparison quality of health services between public and private providers: the Iranian people's perspective. ***Electronic Physician,*** **8**(9): 2935-2941.

Alkarkhi, A. F., Easa, A. M., 2008. Comparing discriminant analysis and logistic regression model as a statistical assessment tools of arsenic and heavy metal contents in cockles. ***J Sustainable Dev.,*** **1**(2): 102-106.

Alumran, A., Almutawa, H., Alzain, Z., Althumairi, A., Khalid, N., 2020. Comparing public and private hospitals' service quality**.** ***Journal of Public Health,*** 13: 1-7

Amin, M., Nasharuddin, S. Z., 2013. Hospital service quality and its effects on patient satisfaction and behavioural intention. Clinical governance: ***An International Journal,*** **18**(3): 238-254.

Andaleeb, S. S., 2000. Public and private hospitals in Bangladesh: service quality and predictors of hospital choice. ***Health Policy and Planning,*** **15**(1): 95-102.

Andaleeb, S. S., 2001. Service quality perceptions and patient satisfaction: a study of hospitals in a developing country. ***Social Science and Medicine,*** **52**(9): 1359-1370.

Andrich, D., Marais, I., 2019. ***A Course in Rasch Measurement Theory.*** D. Andrich y I. Marais (Coords.), Measuring in the educational, social and health sciences, 41-53.

Baltussen, R. M. P. M., Yé, Y., Haddad, S., Sauerborn, R. S., 2002. Perceived quality of care of primary health care services in Burkina Faso. ***Health Policy and Planning,*** **17**(1): 42-48.

Basilevsky, A. T., 2009. ***Statistical Factor Analysis and Related Methods***: ***Theory and Applications***. John Wiley Sons.

Bolarinwa, O. A., 2015. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. ***Nigerian Postgraduate Medical Journal,*** **22**(4): 195-201.

Campbell, G., 1994. Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. ***Statistics in Medicine,*** **13**(5-7): 499-508.

Cattell, R. B., 1966. The scree test for the number of factors. ***Multivariate Behavioral Research,*** **1**(2): 245-276.

Child, D., 2006. ***The essentials of factor analysis***. (3rd ed.). New York, NY: Continuum international publishing group.

Calisir, F., Bayraktaroglu, A. E., Gumussoy, C. A., Kaya, B., 2014. Effects of service quality dimensions including usability on perceived overall quality, customer satisfaction, and return intention in different hospital types. ***International Journal of Advanced Operations Management,*** **6**(4): 309-323.

Cramer, D., 2003. ***Advanced Quantitative Data Analysis***. McGraw- hill education (UK).

D'Cunha, S., Suresh, S., 2015. The measurement of service quality in healthcare: a study in a selected hospital. ***International Journal of Health Sciences and Research,*** **5**(7): 333-345.

Donabedian, A., 1980. ***The Definition of Quality and Approaches to its Assessment***. Health administration press. Ann Arbor, Michigan.

Douglas, L., Connor, R., 2003. Attitudes to service quality–the expectation gap. ***Nutrition, Food Science,*** **33**(4): 165-172.

Eggleston, K., Lu, M., Li, C., Wang, J., Yang, Z., Zhang, J., Quan, H., 2010. Comparing public and private hospitals in China: evidence from Guangdong. ***BMC Health Services Research***, **10**(1): 1-11.

Fawzy, S. O., 2004. ***The Use of the Discriminatory Function in Qualitative Control with the Application of Births of Children Infants*** (Doctoral dissertation, M. Sc. Thesis, university of Mosul, Mosul, Iraq).

Ferrer, A. J. A., Wang, L., 1999. ***Comparing the Classification Accuracy among Nonparametric, Parametric Discriminant Analysis and Logistic Regression Methods***. Paper presented at the annual meeting of the American educational research association, Montreal, Canada.

Gaur, S. S., Xu, Y., Quazi, A., Nandi, S., 2011. Relational impact of service providers' interaction behavior in healthcare. Managing service quality***: An International Journal***, **21**(1): 67-87.

George, D., Mallery, P., 2019***. IBM SPSS Statistics 26 Step by Step: A simple Guide and Reference.*** Routledge.

Gorsuch, R. L., 1983. ***Factor Analysis*** (2nd Ed). Hillsdae, NJ: Erlbaum.

Hair, J. F., Anderson, R. E., Tatham, R. L., William, C., 1995. ***Black: Multivariate Data Analysis with Readings.*** **4**. Auflage, Englewood Cliffs.

Hanley, J. A., McNeil, B. J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. ***Radiology,*** **143**(1): 29-36.

Hansen, T., Kjaersgaard, A., 2020***. Item Analysis of the Eating Assessment Tool (EAT-10) by the Rasch Model: a Secondary Analysis of Cross-Sectional Survey Data Obtained Among Community-Dwelling Elders***. Health and quality of life outcomes, 18, 1-14.

Hardle, W., and Hlavka, Z., 2007. ***Multivariate Statistics***. Springer science and business media, LLC.

Ho, R., 2006. ***Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS***. CRC press.

Huberty, C. J., 1975. Discriminant analysis. ***Review of Educational Research,*** **45**(4): 543-598.

Ibarra, L., Casas, E., Partida, A., 2014***. SERVQUAL Method Applied to Agencia Fiscal Del Estado de Sonora: An Analysis about Service Quality***. The 2nd international conference on strategic innovative marketing, México.

Irfan, S. M., Ijaz, A., 2011. Comparison of service quality between private and public hospitals: empirical evidences from Pakistan. ***Journal of Quality and Technology Management,*** **7**(1): 1-22.

Johnson, R. A., Wichern, D. W., 2002. ***Applied Multivariate Statistical Analysis*** Upper Saddle River, NJ: Prentice hall.

Kaiser, H. F., 1974. Little Jiffy, Mark IV. ***Educational and Psychological Measurement,*** **34**(1): 111-117.

Kaiser, H. F., 1958. The varimax criterion for analytic rotation in factor analysis. ***Psychometrika,*** **23**(3): 187-200.

Karekar, P., Tiwari, A., Agrawal, S., 2015. Comparison of service quality between private and government hospitals: empirical evidences from Yavatmal City, Maharashtra. ***International Journal of Advance Research in Computer Science and Management Studies,*** **3**(6): 39-43.

Kitapci, O., Akdogan, C., Dortyol, İ. T., 2014. The impact of service quality dimensions on patient satisfaction, repurchase intentions and word-of-mouth communication in the public healthcare industry. ***Procedia-Social and Behavioral Sciences,*** **148**, 161-169.

Kondasani, R. K. R., Panda, R. K., 2015. Customer perceived service quality, satisfaction and loyalty in Indian private healthcare. ***International Journal of Health Care Quality Assurance,*** **28**(5): 452-467.

Kumar, R., Indrayan, A., 2011. Receiver operating characteristic (ROC) curve for medical researchers. ***Indian Pediatrics,*** **48**(4): 277-287.

MacCallum, R. C., Widaman, K. F., Zhang, S., Hong, S., 1999. Sample size in factor analysis. ***Psychological Methods,*** **4**(1): 84-99

Magnus, J. R., Neudecker, H., 2019. ***Matrix Differential Calculus with Applications in Statistics and Econometrics.*** John Wiley, Sons.

Mensah, F., Asirifi, M. K., Sherif, M., 2016. Comparative analysis of patient satisfaction between private and public hospital. ***American Based Research Journal,*** **5**(10): 2304-7151.

Morrison, D., 1974. ***Discriminant Analysis***. Handbook of marketing research, 2, 443-457.

Mostafa, M. M., 2005. An empirical study of patients expectations and satisfactions in Egyptian hospitals, ***International Journal of Health Care Quality Assurance,*** **18**(7): 516-532.

Mundfrom, D. J., Shaw, D. G., Ke, T. L., 2005. Minimum sample size recommendations for conducting factor analyses. ***International Journal of Testing,*** **5**(2): 159-168.

Norusis, M. J., 1990. ***Discriminant Analysis: Procedure Discriminant***. SPSS Advanced statistics user’s guide, Chicago, SPSS Inc, 1-42.

Nunnally, J. C., 1978. ***Psychometrie Theory*** (2" d ed.). New York: McGraw- hill.

Obuchowski, N. A., Bullen, J. A., 2018. ***Receiver Operating Characteristic (ROC) Curves: Review of Methods with Applications in Diagnostic Medicine***. Physics in medicine and biology, 63(7), 07TR01.

Pakdil, F., Harwood, T. N., 2005. Patient satisfaction in a preoperative assessment clinic: an analysis using SERVQUAL dimensions. ***Total Quality Management and Business Excellence,*** **16**(1): 15-30.

Parasuraman, A., Zeithaml, V. A., Berry, L. L., 1985. A conceptual model of service quality and its implications for future research. ***Journal of Marketing,*** **49**(4): 41-50.

Parasuraman, A., Zeithaml, V. A., Berry, L. L., 1994. Reassessment of expectations as a comparison standard in measuring service quality: implications for further research. ***Journal of Marketing,*** **58**(1): 111-124.

Pyryt, M. C., 2004. Pegnato revisited: using discriminant analysis to identify gifted children. ***Psychology Science,*** **46**(3): 342-347.

Rad, N. F., Som, A. P. M., Zainuddin, Y., 2010. Service quality and patients' satisfaction in medical tourism. ***World Applied Sciences Journal,*** **10**(1): 24-30.

Rencher, A. C., 2003. ***Methods of Multivariate Analysis***. John Wiley, Sons.

Shiker, M. A., 2012. Multivariate statistical analysis. ***British Journal of Science,*** **6**(1): 55-66.

Shiraishi, J., Pesce, L. L., Metz, C. E., Doi, K., 2009. Experimental design and data analysis in receiver operating characteristic studies: lessons learned from reports in radiology from 1997 to 2006. ***Radiology,*** **253**(3): 822-830.

Soderstrom, R. and Leitner، W., 1997. ***The Effects of Base Rate, Selection Ratio ،Sample Size ، and Reliability of Predictors on Predictive Efficiency Indicies Associated with Logistic Regression Models***. Paper R11 presented at the annual meeting of the mid- western educational research association (Chicago, IL, October 15-18, 1997).

Sohail, M. S., 2003. Service quality in hospitals: more favorable than you might think, ***Managing Service Quality,*** **13** (3): 197-206.

Tabachnick, B. G., Fidell, L. S., 1996. ***Using Multivariate Statistics.*** Harper Collins. New York.

Tabachnick, B. G., Fidell, L. S., Ullman, J. B., 2007. ***Using Multivariate Statistics,*** Boston, MA: Pearson.

Talib, H. R., Abd, A. R., Shimal, A. H., 2019. Using discriminatory analysis to determine the factors affected eye. ***Diyala Journal for Pure Science,*** **15**(03): 125-144.

Taner, T., Antony, J., 2006. Comparing public and private hospital care service quality in Turkey. ***Leadership in Health Services,*** **19**(1): 1366-0756.

Thompson, B., 2004. ***Exploratory and Confirmatory Factor Analysis***. American psychological association.

Tinsley, H. E., Brown, S. D., 2000. ***Handbook of Applied Multivariate Statistics and Mathematical Modeling.*** Academic press.

Wiesniewski, M., Wiesniewski, H., 2005. Measuring service in a hospital colposcopy clinic, ***International Journal of Health Care Quality Assurance,*** **18** (3): 217-228.

Williams, S. G., Calnan, M., 1991. Key determinants of consumer satisfaction with general practice. ***Journal of Family Practice,*** **8**: 237-242.

Yang, K., Trewn, J., 2004**.** ***Multivariate Statistical Methods in Quality Management***. McGraw hill professional.

Yang, S., Berdine, G., 2017. The receiver operating characteristic (ROC) curve. ***The Southwest Respiratory and Critical Care Chronicles,*** **5**(19): 34-36.

Yeşilada, F., Direktouml, E., 2010. Health care service quality: A comparison of public and private hospitals**.** ***African Journal of Business Management,*** **4**(6): 962-971.

Zarei, A., Arab, M., Froushani, A. R., Rashidian, A., Tabatabaei, S. M. G., 2012. Service quality of private hospitals: The Iranian patients' perspective. **https://doi.org/10.1186/1472-6963-12-31.**

Zhou, X. H., McClish, D. K., Obuchowski, N. A., 2009. ***Statistical Methods in Diagnostic Medicine***. John Wiley, Sons.

EXTENDED SUMMARY IN TURKISH
(GENİŞLETİLMİŞ TÜRKÇE ÖZET)

SAĞLIK HİZMETLERİ KALİTESİ AÇISINDAN KAMU VE ÖZEL HASTANELERİN SINIFLANDIRILMASI: ERBİL İLİ ÖRNEĞİ

ISSA, Ayad Faisal
Yüksek Lisans Tezi, İstatistik Anabilim Dalı
Tez Danışmanı: Dr Öğr Üyesi. Yener ALTUN
İkinci Danışmanı: Dr Öğr Üyesi. Rizgar Maghdid AHMED
Ocak 2021, 69 Sayfa

ÖZ

Bu çalışmada sağlık hizmetlerinin kalitesi, değerlendirmenin en önemli kriteri olmuştur. Kamu ve özelde hizmet sektörü, sağlık sisteminde toplumun sağlık hizmetlerinin kalitesini yükseltmede etkili olmuştur. Bu çalışmanın amacı, Irak'ın Erbil ilindeki kamu ve özel hastaneleri sağlık hizmetleri kalitesi açısından sınıflandırmaktır. Bu çalışma, SERVQUAL boyutsal modele dayalı niceleyici kesitsel bir anket içermektedir. 2020 yılında Erbil ilinde kamu ve özel hastaneler de dâhil olmak üzere 11 farklı hastanenin tüm bölümlerinde 370 anket formu hastalara rastgele dağıtıldı. Elde edilen veriler, modeli değerlendirmek için faktör analizi ve ROC eğrisi ile diskriminant analizi kullanılarak IBM – SPSS 25 istatiksel analiz programı ile analiz edilmiştir. Faktör analizi, toplam varyansın% 64'ünü oluşturan yedi faktör tanımlamıştır. Diskriminant analizinin doğru sınıflandırma oranı % 78.1 olduğu, kamu hastaneleri ile özel hastaneleri birbirinden ayrılıdğı görülmektedir. Analiz sonucunda, ROC eğrisi kriterinin ayırt edici modeli ve ROC eğrisi altındaki alanın AUC değeri 0.863 olarak elde edildi. Bu durum tanısal testin olağanüstü ayrımcılık taşıdığı anlamına gelmektedir. X1, X7 ve X8 değişkenleri faktör analizi ile diskriminant analizi arasında hasta memnuniyeti açısından en önemli ve etkili değişkenler olarak belirlenmiştir. Sonuç olarak analiz sonuçları kamu ve özel hastaneler arasında anlamlı bir fark olduğunu ortaya koymaktadır.

Anahtar kelimeler: Devlet hastanesi, Diskriminant analizi, Faktör analizi, Hasta, Hizmet kalitesi, ROC eğrisi, Özel hastane.

1. GİRİŞ

Sağlık hizmetlerinin kalitesi, sağlık kuruluşlarının başarısında ve sürdürülebilirliğinde en önemli faktördür. Dolayısıyla hizmet veren kuruluşun müşteri memnuniyetini arttırmada önemli bir etkendir (Ibarra ve ark., 2014). Özellikle yüksek hasta hacmiyle karşı karşıya kalan kurumlarda hizmet kalitesi, müşteri odaklı kurumlar olarak görülmek için çok önemli bir ölçü olarak kullanılabilir. Bu durum, hizmet alanların memnuniyetine, sağlık hizmeti organizasyonlarının etkililiğine ve verimliliğine yol açmaktadır (Aghamolaei ve ark., 2014). Ayrıca hastalar rekabette değişime daha açık hale geliyor ve sağlık tesislerine daha aşina hale geliyor ve bu nedenle hizmet standardı tek başına hastalar ve hastane arasında uzun vadeli bir ortaklığı sürdürmek için yeterli olmayabilir (Gaur ve ark., 2011). Hizmet kalitesi, müşterilerin bir hizmetin beklentilerini ne kadar iyi karşıladığına veya aştığına dair algısıdır ve kuruluşlar tarafından değil müşteriler tarafından değerlendirilir. Hizmet sürecinin etkileşimli doğası, tüketicilerin bu hizmetin sağlanması ve performansından hemen sonra kaliteyi değerlendirmesiyle sonuçlanır (Douglas ve Connor, 2003).

1.1. Problem cümlesi

Kamu hastaneleri ile özel hastaneler arasında sağlık hizmetleri kalitesi olarak bir farklılık söz konusudur. Halkın daha çok özel hastaneleri tercih etmesinin temel sebepleri, hizmet kalitesi, temizlik ve hijyenik, ilaç teminatı, kullanılan cihaz kalitesi, laboratuvar kalitesi, hasta bakımı ve sistemsel hız şeklinde sıralanabilir. Bu çalışma ile bu problemlerin daha detaylı sorgulanması söz konusudur.

1.2. Araştırmanın Amacı

Bu çalışmanın amacı ilk olarak kamu ve özel hastanelerinde sağlık hizmetinin kalitesini etkileyen en önemli faktörlere ulaşmak için faktör analizi ve diskriminant analizi olmak üzere iki gelişmiş istatistiksel yol kullanmaktır. Daha sonra, Irak Erbil'deki kamu ve özel hastaneler arasında sağlık hizmetleri kalitesi açısından en önemli değişkenleri sınıflandırma ve değerlendirme modelini ortaya çıkarmaktır. Ayrıca bu çalışma, 2020 yılında Erbil ilindeki kamu ve özel hastaneler de dâhil olmak üzere 11 farklı hastanenin tüm bölümlerinde 370 anket formu rastgele tabakalı örnekleme yöntemi ile hastalara dağıtılarak uygun bir istatiksel analiz programı ile değerlendirilmiştir.

2. MATERYAL VE YÖNTEM

Çalışmada çağdaş araştırmalarda yaygın olarak kullanılan SERVQUAL araçları uygulanmıştır (Sohail, 2003; Mostafa, 2005; Wiesniewski ve Wiesniewski, 2005). Çalışma anketi iki bölümden oluşmaktadır; birinci bölümde hastanede yatan hastanın cinsiyet, yaş, evlilik, meslek, ekonomik durum ve eğitim gibi demografik özellikleri, ikinci bölümde kaliteli sağlık hizmetlerine ilişkin 22 değişken yer almaktadir. Algı ve beklenti değişkenleri, 1 = "kesinlikle katılmıyorum" ile 5 = "kesinlikle katılıyorum" arasında değişen beşli Likert ölçeğinde ölçülmüştür. Daha sonra bu anket formu 2020 yılında Erbil ilindeki kamu ve özel hastaneler de dâhil olmak üzere 11 farklı hastanenin tüm bölümlerinde 370 hastaya rastgele dağıtılmış, ayrıca tabakalı rastgele örnekleme kullanılmıştır. Anket çalışmasında devlet hastanelerinde 188, özel hastanelerde ise 182 hasta yer almaktadır.

Elde edilen anket sonuçları manuel olarak IBM SPSS 25 istatistik programına aktarılıp analizler yapılmıştır. Değişkenler kodlandıktan sonra kamu ve özel hastanelere yönelik hasta memnuniyetini etkileyen en önemli değişkenleri elde etmek için faktör analizi (temel bileşen analizi) yöntemi kullanılarak analizler yapılmıştır. Bileşenleri yorumlamak için Varimax rotasyonu kullanılarak 0.5'in üzerindeki yükler analize dâhil edilip yorumlanmıştır. Kamu hastaneleri ile özel hastaneler arasında en iyi ayrım yapan değişkenler alt grubuna dayalı bir model geliştirmek için aşamalı diskriminant analizi ve diskriminant modelini değerlendirmek için ise ROC eğrisi kriteri kullanılmıştır.

3. BULGULAR

Bu çalışma, 2020 yılında Erbil ilindeki kamu ve özel hastaneler de dâhil olmak üzere 11 farklı hastanenin tüm bölümlerinde 370 hastaya rastgele dağıtılan bir anket çalışmasına dayanmaktadır. İlk aşamada, hastalara kamu ve özel hastane tarafından kaliteli sağlık hizmetinde sınıflandırma konusundaki anlayışı hizmet kalitesi boyutlarına göre ortalamasını ve standart sapmasını temsil eden tanımlayıcı istatistikler kullanılmıştır. İkinci aşamada, kamu ve özel hastaneler için en önemli değişkeni ortaya çıkarmak için faktör analizi yapılmıştır. Üçüncü aşamada kamu ve özel hastaneyi birbirinden ayırmak için diskriminant analizi, dördüncü aşamada ise modelin değerlendirilmesinde ROC eğrisi kriteri uygulanmıştır.

Faktör analizi, faktör olarak adlandırılan temel varlıklar biçiminde değişkenler arasındaki korelasyonları açıklığa kavuşturmak için bir dizi yöntemdir. Ayrıca model faktör analizinin geniş bir şekilde anlaşıldığı durumlarda, gözlemlenen rastgele değişkenler arasındaki ortak ilişkisel yapıyı açıklığa kavuşturmayı veya oluşturmayı amaçlayan neredeyse ilişkili formların bir koleksiyonunu ifade eder (Basilevsky, 2009).

Kamu ve özel hastanelerinde faktör analizi, toplam varyansın % 64'ünü yedi bileşenle açıklamaktadır. Üç değişken içeren ilk bileşen "hizmetlerin somutluğu ve düzenliliği" % 12.892 varyans açıklama oranıyla en önemli faktör olarak elde edilmiştir.

Diskriminant veya diskriminant fonksiyon analizi, nicel değişkenlerin hangi ağırlıklarının 2 veya 2'den fazla vaka grubunu en iyi şekilde ayırt ettiğini ve bunu şanstan daha iyi yaptığını belirleyen parametrik bir tekniktir (Cramer, 2003). Ayrıca, diskriminant analizi, tek tek nesneleri veya birimleri gruplara ayırabilen güçlü, çok değişkenli bir istatistiksel yöntemdir. Spesifik olarak, diskriminant analizinde $x_{1,}x_2, \dots, x_n$ gibi birçok ayrı nesne vardır ve her nesnenin kendisiyle ilişkili birkaç değişkeni vardır. Tüm nesne popülasyonu, nesnenin ait olduğu grupla güçlü bir şekilde ilişkili değişkenlerin değerleri ile birkaç gruba bölünebilir. Diskriminant analizi, her nesnenin değişkenlerinin değerlerine bağlı olarak her nesneyi gruplara ayıran bir yöntemdir (Kai Yang ve Jayant Trewn, 2004).

Diskriminant analizi, hastane seçimi ile ilgili anahtar değişkenin, kamu ve özel hastanelerden faktör analizinde çıkardığımız en önemli 17 değişken için aşamalı yöntem kullandığını ve ardından iki gruptaki en önemli 9 değişkeni çıkardığımızı göstermiştir. Diskriminant analizi için doğru sınıflandırma oranı % 78.1 olup, kamu ve özel hastaneler arasında ayrım yapmaktadır.

Çalışmada ROC eğrisi kriteri, diskriminant modeli ve ROC eğrisi altındaki alanı değerlendirmek için kullanıldı. Yapılan analizler sonucunda ROC eğrisi altındaki alanın

AUC değeri 0.863 olarak elde edilmiş ve bu durum ise teşhis testinin olağanüstü ayrımcılık taşıdığı anlamına gelir.

4. SONUÇLAR

Yapılan analizler sonucunda, kamu ve özel hastanelerde faktör analizi, toplam varyansın % 64'ünü oluşturan yedi bileşen ile tanımlamıştır; üç değişken içeren ilk bileşen “hizmetlerin somutluğu ve düzenliliği”, % 12.892 varyans açıklama oranıyla en önemli faktör olarak elde edilmiştir. Üç değişken içeren “hizmetlerin güvencesi” ikinci bileşen ve bu bileşenin varyans yüzdesi % 10.897'dir. İki değişken içeren “yanıt verme hizmetleri” üçüncü bileşen ve bu bileşene ait varyans yüzdesi % 9.513'tür. İki değişken içeren “güvenilirlik hizmeti” dördüncü bileşen ve bu bileşene ait varyans yüzdesi % 8.879'dur. İki değişken içeren “empati hizmeti” beşinci bileşen ve bu bileşene ait varyans yüzdesi % 8.660'tır. Üç değişken içeren “esneklik hizmeti” altıncı bileşen, oranı % 7.766'dır. İki değişken içeren “zaman ve malzeme hizmeti” yedinci bileşen, oranı % 5.841'dir.

Diskriminant analizi, hastane seçimiyle ilişkili anahtar değişkenin, kamu ve özel hastanelerden faktör analizinde çıkardığımız en önemli 17 değişken için aşamalı yöntem kullandığını ve ardından iki gruptaki en önemli 9 değişkeni çıkardığımızı göstermiştir. Diskriminant analizi için doğru sınıflandırma oranı % 78.1 olup, kamu ve özel hastaneler arasında ayrım yapmaktadır. Diskriminant modeli ve ROC eğrisinin altındaki alanı değerlendirmek için ROC eğrisi kriteri kullanılarak ROC eğrisi altındaki alanın AUC değeri 0.863 olarak elde edilmiştir. Bu durum teşhis testinin olağanüstü ayrımcılık taşıdığı anlamına gelir.

Sonuç olarak, faktör analizi ile diskriminant analizi arasında üç değişkende önemli bir yakınsama ve benzerlik bulunmuştur. Faktör analizi ile belirlenen en önemli ilk üç bileşenle 8, diskriminan analizi ile de 9 önemli değişken belirlenmiştir. Ek olarak, hasta memnuniyeti açısından sonuçlar, değişkenlerin (X1: hastanedeki modern ekipman, X7: Hastane personeli işleri ilk seferde yaptırır ve X8: zamanında hizmet sağlama) faktör analizi ile diskriminant analizi arasındaki en önemli ve etkili değişkenler olduğunu göstermektedir. Bu nedenle, öneri olarak devlet hem kamu hastaneleri hem de özel hastanelerde sağlık hizmeti kalitesini artırmak için bu üç değişkene çok dikkat etmelidir.

APPENDIX

Questionnaire Form

Ayad Faisal ISSA
VAN YÜZÜNCÜ YIL UNIVERSITY
DEPARTMENT OF STATISTIC

Questionnaire Form

Classification and Evaluation of Public and Private Hospitals In Iraq In Terms of Quality Health Services Using Multivariate Data Analysis With ROC Curve: A Sample of Erbil Province.

The following form is consist of a few parts and questions about a scientific MA research entitled (Classification and Evaluation of Public and Private Hospitals In Terms of Quality Health Services) Your answers to the whole questions correctly is very important to the research.

Name hospital ☐

Type hospital ☐

General information

1-Gender Male ☐ Female ☐

2-Age < 25 ☐ 25-39 ☐ 40-54 ☐ ≥ 55 ☐

3-Marital status married ☐ Single ☐ Divorce ☐ Another ☐

4-Occupation ☐

5-Economy Bad ☐ Middle ☐ Good ☐ Very good ☐

6-Education Illiterate ☐ Just read and write ☐ Primary ☐ Intermediate ☐ Secondary ☐ Institute ☐ Bachelor ☐ Higher certificate ☐

N	Variables	Strongly disagree	disagree	neutral	agree	Strongly agree
Tangibles						
1	Modern equipment in the hospital.					
2	Hospital attractiveness.					
3	Medical staff appearance.					
4	Attractiveness of medical materials.					
Reliability						
5	The hospital staff provide promised services.					
6	Hospital staff are interested in solving patient's problem.					
7	The hospital staff get things done the first time.					
8	Providing timely services.					
9	The hospital staff insist on error free records.					
Responsiveness						
10	The hospital staff communicate to patients about service provision.					
11	The hospital staff provide prompt services.					
12	Medical staff willingness to help patients.					
13	Medical staff never too busy to respond to patients request.					
Assurance						
14	Hospital staff demeanour instils confidence.					
15	Patients feel safe in the hospital.					
16	Hospital staff are always courteous towards Patients.					
17	Medical staff are knowledgeable enough.					

N	Variables	Strongly disagree	disagree	neutral	agree	Strongly agree
Empathy						
18	Hospital staff give patients individual attention.					
19	Hospital has convenient opening hours.					
20	Hospital staff give patients personal attention.					
21	Staff have patients interest at heart.					
22	Medical staff understand clearly patients' specific needs.					

CURRICULUM VITAE

Ayad Faisal ISSA was born in 1986, in Duhok province, Iraq. He completed primary, secondary and high school in Bardarash district. He graduated from Statistic Department in Salahaddin University - College of Administration and Economic in 2012. He started his Postgraduate study at the Department of Statistic, Institute of Natural Applied Sciences at Van Yuzuncu Yıl University in Turkey on February 2019.

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

Date: 29/01/2021

Thesis Title: **CLASSIFICATION OF PUBLIC AND PRIVATE HOSPITALS IN TERMS OF HEALTH SERVICES QUALITY: A SAMPLE OF ERBIL PROVINCE**

The title of the mentioned thesis, above having total 69 pages with cover page, introduction, main parts and conclusion, has been checked for originality by turnitin computer program on the date of 29/01/2021 and its detected similar rate was 17 % 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 Van 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,

29/01/2021
Ayad FaisaL ISSA

Name and Surname: Ayad FaisaL ISSA
Student ID#: 18910001399
Science: Statistic
Program: Statistic
Statute: M. Sc. X Ph.D. ☐

APPROVAL OF SUPERVISOR
SUITABLE

Asst. Prof. Dr. Yener ALTUN

APPROVAL OF THE INSTITUTE
SUITABLE

Prof. Dr. Suat SENSOY