

Analysis of Socio-Economic Similarities of the Cities in East Anatolia Region

¹H. Eray Çelik, ²Sinan Saraçlı and ³Sanem Şehribanoğlu

¹Department of Statistics, Faculty of Science, Yüzüncü Yıl University, Van, Turkey

²Department of Statistics, Afyon Kocatepe University Science-Literature Faculty, Afyonkarahisar

³Yüzüncü Yıl University, Ozalp Junior Technical School, Van, Turkey

Abstract: The purpose of this study is to determine the similarities and dissimilarities of the 14 city in East Anatolia Region, via socio-economic variables. Hierarchical cluster analysis using Ward's technique in the study and K - the average technique is used to cluster them non-hierarchically. As a result of the study, these cities are included in meaningful clusters for their similarity and dissimilarity.

Key words: East Anatolia Region • Cluster Analysis • Socio-Economic Development

INTRODUCTION

This is a well known fact that socio-economic development, both in terms of time and a number of differences in development levels and counties have changed over time [1]. Socio-economic development of provinces in inequality contains a number of levels. Important parameters affecting the level of development of the provinces are the geographical, functional and social inequality. In terms of socio-economic development levels between regions and provinces since a difference of yesterday, today and tomorrow will continue to exist as a permanent phenomenon. It is an inevitable necessity that balancing this unequal situation between the provinces on the basis of different factors as well as the economic development. The growth of the individuals' income is not a sufficient factor to identify the economic growth, exclusively. In parallel, as stated in the report titled the ranking of the socio-economic development of provinces and regions, to focus on the income element that was only measured by the Gross Domestic Product (GDP) is insufficient at the identification and evaluation situation of the provinces that they came [2,3]. It is required that using a data set of the large number of selected variables from economic, social and cultural fields, GDP variable is also included in, to compare the level of sophistication and rankings of the provinces. This situation is stated in the relevant text, it is required that the adoption of a qualitative perspective on socio-economic development instead of focusing on a quantitative growth, [2-8]. As emphasized in the summary

above, the economic growth needed to explain the social sophistication relatively.

East Anatolia Region has the highest ratio at the inequality between regions and provinces, with South Eastern Anatolia Region. The biggest reasons for this, economic resources, income distribution and equality of opportunity are unbalanced. In the regional cities, despite the reforms at education and health services, the existing deficiency, lack of residential estate, water, energy and infrastructure, environmental pollution, noise, traffic and migration are the basic results of this inequality.

There is a growth-based perspective due to stable and competitive economic conditions in the region is presented to perform, after the causes of backwardness of the region are described in detail within the framework of SPO - Eastern Anatolia Project Master Plan that have the finding of the status of the region. This report was prepared to determine the comprehensive activities that will accelerate the development of the Eastern Anatolia Region [9]. Also, the remarkable results that are revealed from main reasons of the regional backwardness are presented in summary below in this report.

Eastern Anatolia Region is a predominantly agricultural region. In this report, as one of the main reasons for the emergence of regional inequality are shown to remain predominantly agricultural economy. The low agricultural productivity in a predominantly agricultural region can be described as the cause for the lowness of the real economic activity and per capita income. One of the other important problems of the region is the high natural increase in population. High population

growth is also the basic cause of inadequate education and health services infrastructure. Because of the climatic and geographical factors, partially the shortness of market opportunities, different agricultural products that have high value-added and high labor-intensive and provide employment opportunities can't be cultivated widely. Finally, the industry could not be developed because of the geographical location of the East Anatolia Region that requires high cost to access the national and international markets and the gates of imported raw materials [9].

Oktay (2001) was established in his work that is performed by the clustering of provinces on the basis of sectoral per capita GDP within the scope of Eastern Anatolia Project by using different methods under the hierarchical clustering approach work: the provinces of Hakkari, Agri and Bingol are the weakest, the provinces of Mus, Bitlis, Ardahan and Iğdir are weak, the provinces of Van, Kars, Bayburt and Erzurum are representing the average in the region, the provinces of Tunceli and Erzincan Gumushane are over the regional average, finally the provinces of Malatya and Elazığ are the highest cities with regard to per capita GDP amount [4].

According to the Program in 2011 in frame of the Ninth Development Plan 2007-2013, it is decided to continue the program, that intends to reduce the imbalance and instability of the milk and meat markets particularly and initiates the support to the establishment of large enterprises engaged in beef breed of animal husbandry in East Anatolia Region in 2010, in 2011. In addition, preparations for the Eastern Anatolia Project (EAP) Prioritizing and Update Study have begun. In order to minimize the element of inequality between provinces at a minimum level, it is aimed to take some precautions under a series of sub-section at the period 2011-2013 Medium-Term Program (MTP). The main objective of the MTP has defined as, in line with the ultimate goal of increasing the welfare of our country's growth, stabilizing the growth, increasing employment, improving the public balance and to ensure price stability [10]. According to the MTP, the public investment aims to reduce disparities between regions and to intensify the economic and social infrastructure projects particularly in Southeastern Anatolia Project, Eastern Anatolia Region and the Konya Plain Project. Under the perspective of reducing regional disparities, reducing regional disparities, developing the minimum living standards in rural areas and to increase the competition between the regions are adopted as the basic objectives [10,11]. These steps that are taken to reduce regional inequality, will contribute to the development of the region in a long term.

The main purpose of the study, determine the level of socio-economic development of the provinces of Eastern Anatolia Region and in accordance with this determination to reveal clusters of provinces with similar characteristics.

Research Methodology: The main purpose of the study, to determine the basic similarities of the provinces of Eastern Anatolia Region in this region because it is within the scope of the geographic partitioning of the 14 provinces are included in the entire research process. The new provinces in the region are also within the scope of this research.

Research Variables: This research has based on the variables classified under 10 main headings in order to determine the similarities and differences of the provinces of the region. The variables were assessed under the headings: demographic indicators, employment indicators, education indicators, health indicators, industry indicators, agricultural indicators, financial indicators and the other indicators of welfare. Data have been compiled by using official statistics of TurkStat, the Ministry of Education, OIZ Information System, the Banks Association, the Ministry of Finance and the Ministry of Health. The data were quite sure that the current data in this study. However, the most current data for the indicators of employment 12 years old and over population "for the employed population by economic activity and sex-TURKSTAT" were obtained by the year 2000. The variables that are used in the study area are given in Table 1 following.

Cluster Analysis: Cluster Analysis is a collection of methods that help to separate the groupings in the data matrix and units, or unit of the variables and variables that are similar to each other sub-clusters (group, class) that have unknown natural groupings. Cluster analysis is used to divide the units to homogeneous groups by utilize some of the dimensions that are calculated on the basis of similarities or differences between variables. Cluster analysis have similar features with discriminant and exploratory factor analysis, because of the dimensionality reduction and the observation units have the goals of the collection performed by for the same factors. Cluster analysis, similar to the objects in the cluster, the clusters will be different between clusters. If the clustering process is successful, when there is a geometric drawing objects will be very close to each other within the cluster, the clusters will be far away from each other [12-14].

Table 1: Variables used in the study

Year	Variable	Source	Unit
Demographical Indicators			
2009	Total Population	TÜRKSTAT	Person
2009	Urbanization Rate	TÜRKSTAT	Per cent
2008-2009	Annual Average Population Growth Rate	TÜRKSTAT	Thousand
2009	Population Density	TÜRKSTAT	Person/km ²
2009	Average Household Size	TÜRKSTAT	Person
Employment Indicators			
2000	Agricultural Sectors Ratio Of Employees In Total Employment	TÜRKSTAT	Per cent
2000	Industrial Sectors Ratio Of Employees In Total Employment	TÜRKSTAT	Per cent
2000	Commercial Sectors Ratio Of Employees In Total Employment	TÜRKSTAT	Per cent
2000	Financial Sectors Ratio Of Employees In Total Employment	TÜRKSTAT	Per cent
2000	Salaried Employees Ratio In Total Employment	TÜRKSTAT	Per cent
2000	Women In Paid Employees Ratio In Total Employment	TÜRKSTAT	Per cent
Education Indicators			
2010	Literacy Rate of Population	TÜRKSTAT	Per cent
2010	Literacy Rate of Female Population In Female Population	TÜRKSTAT	Per cent
2010	University Graduates Rate In 23+ Population	TÜRKSTAT	Per cent
2010	School Enrollment Rate of Primary Schools	MOE	Per cent
2010	School Enrollment Rate of High Schools	MOE	Per cent
2010	School Enrollment Rate of Vocational and Technical High Schools	MOE	Per cent
Health Indicators			
2010	Infant Mortality Rate	TÜRKSTAT	Thousand
2009	Number of Doctors Per Ten Thousand People	Ministry of Health	Person
2009	Number of Dentists Per Ten Thousand People	Ministry of Health	Person
2009	Number of Pharmacy Per Ten Thousand People	Ministry of Health	Number
2009	Number of Hospital Beds Ten Thousand People	Ministry of Health	Number
Industrial Indicators			
2010	Number of Parcels In The Organized Industrial Zone	OIZ Information System	Number
2009	Number of Offices of manufacturing Industry	TÜRKSTAT	Number
2009	Board Power Capacity in Manufacturing Industry	TÜRKSTAT	Horse Power
2009	Per Capita Consumption of Electricity in Manufacturing Industry	TÜRKSTAT	Kws
2009	Per Capita Value Added in Manufacturing Industry	TÜRKSTAT	TL
Agricultural Indicators			
2009	Per Capita Value of Crop Production	TÜRKSTAT	TL
2009	Per Capita Value of Livestock	TÜRKSTAT	TL
2009	Per Capita Value of Animal Products	TÜRKSTAT	TL
Economical Indicators			
2006	Gross Value Added Per Capita	TÜRKSTAT	TL
2009	Number of Bank Branches	Banks Association	Number
2009	Bank Deposits Per Capita	Banks Association	Thousand TL
2009	Share In The Total Bank Deposits	Banks Association	Thousand TL
2009	Share In The Total Bank Deposits and Loans	Banks Association	Thousand TL
2009	Amount of Agricultural Credit Per Capita Rural Population	Banks Association	Thousand TL
2009	Amount of Loans for Industrial, Commercial and Tourism Per Capita	Banks Association	Thousand TL
2009	Amount of Exports Per Capita	TÜRKSTAT	US Dollar
2009	Amount of Imports Per Capita	TÜRKSTAT	US Dollar
Infrastructure Indicators			
2009	Population Rate Have Drinking Water Service	TÜRKSTAT	Per cent
2011	Asphalt Road Rate of State and Provincial Roads	KGM	Per cent
Constructional Indicators			
2008	Toilet Found in Apartment Flat Rate	TÜRKSTAT	Per cent
2008	Bathroom Found in Apartment Flat Rate	TÜRKSTAT	Per cent
2008	Kitchen Found in Apartment Flat Rate	TÜRKSTAT	Per cent
2008	Piped Water System Found in Apartment Flat Rate	TÜRKSTAT	Per cent
Other Welfare Indicators			
2009	Number of Private Cars Per Ten Thousand People	TÜRKSTAT	Number
2009	Number of Motor Vehicles Per Ten Thousand People	TÜRKSTAT	Number
2008	Electricity Consumption Per Capita	TÜRKSTAT	Mws
2010	Population Rate Have Green Card	Ministry of Health	Per cent

The phases of the cluster analysis; capturing data matrix by obtaining p number variables of the n number units from main solids regarding the absence of accurate information about the natural groupings of the unit or variables, determining the similarity matrix of variables or units with an appropriate measure of similarity that shows their similarities or differences of the units or variables, separating the units or variables to the convenient number of clusters according to similarity matrix with the help of appropriate clustering method and finally interpretation of the clusters that are derived from the separation.

Cluster analysis is discussed under two headings: hierarchical and non-hierarchical clustering. Hierarchical cluster techniques are the process of consecutive defragmentation of the clusters. Hierarchical clustering techniques are performed with successive assemblies or consecutive divisions. Hierarchical clustering techniques provide to obtain the hierarchical decomposition of the units by using the distance values from each other of the data set's units. First, the most similar groups are combined and then these groups which are combined at the first stage are merged with each other according to their similarities. As a result of the reduction of similarities, all sub-groups are collected in a single cluster. In the hierarchical clustering phase, the clusters are shown by using the tree diagram, also known as dendrogram. Hierarchical clustering techniques are single linkage, complete linkage, average link and Ward's techniques. Non-hierarchical clustering technique is used, differently from hierarchical clustering technique, when it comes to the presence of large data sets. It is aimed to collect observation units in K units of the cluster more than variables with this technique. K-average technique is the most commonly used technique of the hierarchical clustering techniques. The purpose of K-average technique is having the largest similarities in the cluster and the smallest similarities between the clusters in the resulting clusters [15].

Hierarchical and non-hierarchical clustering techniques were used in this study. All the techniques, that are used as hierarchical clustering technique, are applied in an established combinations according to the account approaches of different distance matrix, it was decided to anticipate the results of Ward's technique. In short, determining the distance matrix was used in squared Euclidean distance, Ward's technique was used in the clustering of provinces. Clustering of units in hierarchical clustering are carried out by using the

distances of the units from each other or similarities with each other. The squared Euclidean distance is the most widely used criterion to determine distances. Measure of Euclidean is a measurement that specifies distance between variables or units without affected by measurement units.

Ward's technique of merging the two groups is based on the basis of minimizing the "information loss". In other words, Ward's technique aims to be a minimum of variability within clusters. This technique is often resulting with an increase in the loss of information at the error squares criterion (ESS). First, for the k unit cluster ESS_k is determined as the sum of squares of deviations of the cluster average (center) of each item. If K is the number of cluster, ESS is expressed as the sum of ESS_k . Or for each stage;

In the form of $ESS = ESS_1 + ESS_2 + \dots + ESS_k$, a combination of all possible cluster pair is determined and ultimately this process continues until the two clusters that lead to a small increase the ESS. In the first stage, each cluster with a single substance combination in question. If N is the number of items, it would be $ESS_k = 0, k=1,2,3,\dots,N$ and so will the $ESS = 0$. ESS value is obtained as follows with the combination of N unit material in a single group in other aspects [13].

$$ESS = \sum_{j=1}^n (x_j - \bar{x})' (x_j - \bar{x})$$

Here, X_j is the average of multiple measurements that are associated with j. substance and \bar{x} is the average of all substances. Ward's technique resulting clusters can be represented by a dendrogram. The horizontal axis shows the values of ESS began joints [13]. The statistical meaningfulnesses of the variables were obtained with simplex analysis of variance (ANOVA) by using the K-average method from non-hierarchical clustering techniques.

RESULTS OF ANALYSIS

The results that regard hierarchical clustering analysis technique are obtained by using Ward's technique. In Ward's technique resulting dendrogram analysis, at the first stage the provinces of Mus and Agri are combined and formed a cluster, then there is a new cluster with the addition of the provinces of Kars and Igdir, this cluster becomes a new cluster by adding the cluster formed by provinces of Bingol and Bitlis.

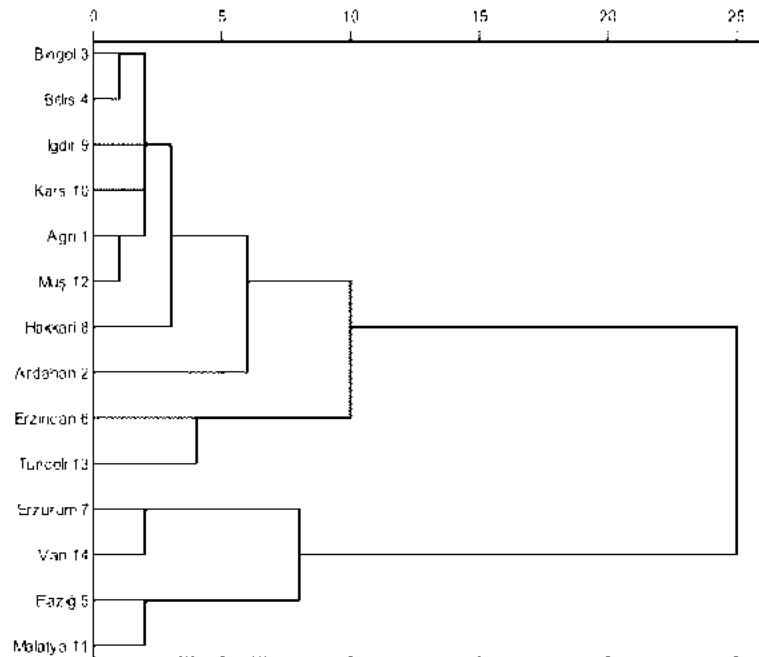


Fig. 1: Dendrogram via Ward's Technique

Table 2: Results of Hiyerarchical Cluster Analysis

Cluster No.	Cluster Memberships
1. Cluster	Bingol, Bitlis, Igdir, Kars, Agri, Mus, Hakkari, Ardahan
2. Cluster	Erzincan, Tunceli
3. Cluster	Erzurum, Van
4. Cluster	Elazig, Malatya

Table 3: Results of non-hierarchical Cluster Analysis

Cluster No.	Cluster Memberships
1. Cluster	Agri, Ardahan, Bitlis, Bingol, Hakkari, Igdir, Kars, Mus, Tunceli
2. Cluster	Erzurum, Van
3. Cluster	Elazig, Erzincan, Malatya

One of the two clusters that will constitute at the last stage is formed by adding the cluster that is combined by provinces of Hakkari firstly, then Ardahan and finally Erzincan and Tunceli, respectively. On the other hand, the new cluster formed by the merger of the cluster formed by the provinces of Elazig and Malatya with the provinces of Van and Erzurum, is the other one of two clusters that will consist at the final stage. All the provinces are collected in a cluster with the merger of the two sets and analysis is resulted. Consisting of sets are shown as a summary in Table 2.

Table 3 shows the results of non-hierarchical cluster analysis. When the cluster memberships of the provinces were examined according to the results of K-average technique, the provinces with more socio-economic

similarities were in common clusters. The results obtained are consistent with Erkan (2001) study too. Accordingly, provinces of Elazig, Malatya and Erzincan represented the cluster of the provinces that are above the regional average as of the entire set of variables. The provinces of Erzurum and Van, represent the average of the region, were represented under the same cluster because of their similar features. The other 9 provinces are below the regional average in all of the provinces.

When the differences between provinces as of the variables examined, the employment indicators ($p < 0.01$) including the agricultural sectors in total employment rate of workers, the industrial sectors in total employment rate of workers, the trade sectors in total employment rate of workers, the financial sectors in total employment rate of

workers, the wage earners in total employment rate of workers and the working women in total employment rate of workers variables are identified as the most influential variables that describe the differences between the provinces. The provinces of Agri, Ardahan, Bingöl, Bitlis, Hakkari, Iğdir and Mus were below the regional average in terms of employment indicators. The provinces of Malatya and Elazığ had better employment indicators were represented above the regional average.

The industry indicators ($p < 0.01$) are the other indicator that leads to the emergence of a statistically significant difference between the provinces. The provinces of Malatya, Erzurum, Elazığ and Van were much higher than the regional average in terms of the variables as number of establishments in manufacturing industry, Manufacturing Industry Board Power Capacity Amount, Per Capita Consumption of Electricity in Manufacturing Industry, Per Capita Value Added in Manufacturing Industry. The provinces of Hakkari, Tunceli and Ardahan were located well below the regional average in terms of indicators of industry.

It is determined that there is statistically significant difference between the provinces in terms of the as Gross Value Added Per Capita, number of bank branches, bank deposits per capita, share in total bank deposits, share in total bank deposits and loans, amount of agricultural credit per capita rural population, amount of loans for industrial, commercial and tourism per capita, amount of exports per capita, amount of imports per capita ($p < 0.01$).

It is indicated that there is a significant difference between the averages of the provinces in terms of other welfare indicators as well as the financial indicators.

According to the result of the ANOVA, it is emerged that there are differences between the average of provinces in terms of the amount of imports per capita variable, which is used for comparison between the provinces, although there isn't any difference in terms of the amount of exports per capita.

Featured respect for these variables, the differences between the clusters defined by the K-average technique are caused by the employment indicators, industry indicators, financial indicators and other indicators between the cluster No. 3, that is above the regional average and the cluster No. 1 is well below the regional average.

One of the other important variable is the construction indicators on the basis of the region. However, it is determined as the averages between the clusters didn't show any difference because of this indicator. On the other hand, it is occurred that the

provinces were well below the average in Turkey in terms of the construction indicators when the provinces are examined in terms of these variables. For instance, the rate of housing do not have piped water systems is over 20% in the provinces of Iğdir and Agri, also the rate of the residential apartment that doesn't have the toilet installation is over 50% in the same provinces and the provinces of Kars and Ardahan have similar features.

RESULTS

The similarities / differences of the provinces in Eastern Anatolia Region are determined in terms of different indicators according to hierarchical and non-hierarchical clustering techniques considering the geographic definition without regard to statistical region. Ward's technique was used as hierarchical clustering techniques, K - the average technique was used as a non-hierarchical clustering technique.

As a result of the analysis, 14 provinces in the Eastern Anatolia Region took part in the same clusters because of their similarity to each other respect for 49 variables according to the geographic definition. The provinces of Malatya, Elazığ, Erzincan, Erzurum and Van are determined as the region's most developed provinces. It is an inevitable necessity that the projects and development plan facing the region should provide that the remaining provinces below the average of the region need to take greater share of promote and investments to correct this inequality in the region. The current unequal development between the cities of the region continue to maintain that its presence was detected in terms of other welfare indicators used in the study, employment indicators, industrial and financial indicators. In addition, it is an undeniable reality for the region that the whole regional provinces approximated to the average of region in terms of infrastructure, education, health and construction indicators and conducted policy, investment, projects and incentives are effective to eliminate the inequality significantly.

REFERENCES

1. Albayrak, S., 2005. Türkiye'de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi, Zonguldak Karaelmas Üniversitesi, Sosyal Bilimler Dergisi, 1(1): 153-177.
2. DPT, 2003. İllerin ve Bölgelerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması, Yayın Ankara. pp: 2671.

3. DPT, 2004. İlçelerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması, Ankara.
4. Oktay, E., 2001. Doğu Anadolu Projesi Kapsamındaki İllerin GSYİH Değerlerine Göre Hiyerarşik Kümeleme Analiziyle Gruplanması, Atatürk Üniversitesi, İktisadi ve İdari Bilimler Dergisi, Cilt. 15(3-4): 329-345.
5. Albayrak, S., 2004. Türkiye’de İllerin Sosyoekonomik Gelişmişlik Düzeylerini Belirleyen Hipotetik Yapıların Faktör Analizi ile İncelenmesi, İşletme İktisadi Enstitüsü Yönetim Dergisi, 48: 53-79.
6. Karabulut, M., M. Gürbüz, ve and E.K. Sandal, 2004. Hiyerarşik Kluster (Küme) Tekniği Kullanılarak Türkiye’de İllerin Sosyo-Ekonomik Benzerliklerinin Analizi, Coğrafi Bilimler Dergisi, 2(2): 71-85.
7. Koç, S., 2001. İllerin Sosyo-Ekonomik Özelliklere Göre Sınıflandırılması, Çukurova Üniversitesi 5. Ulusal Ekonometri ve İstatistik Sempozyumu (Bildiri), Çukurova Üniversitesi, Adana.
8. İlalmaz, Ö. ve and M.S. Temurlenk, 2005. Türkiye’deki İstatistik bölgelerin Kişi Başına Düşen Gelir Açısından Hiyerarşik ve Hiyerarşik Olmayan Kümeleme Analizi İle Değerlendirilmesi: 1965-2001, Atatürk Üniversitesi, İktisadi ve İdari Bilimler Dergisi, 19(2): 75-92.
9. DPT, 2000. Doğu Anadolu Projesi Ana Planı Mevcut Durum ve Analizi, Ankara.
10. DPT, 2010. Orta Vadeli Program (2011-2013), Ankara.
11. DPT, 2010. Dokuzuncu Kalkınma Planı, 2011 Programı, Ankara.
12. Tatlıdil, H., 1992. Uygulamalı Çok Değişkenli İstatistiksel Analiz, Akademi Matbaası, Ankara.
13. Johnson, R.A. and W. Wichern, 1998. Applied Multivariate Statistical Analysis, Prentice Hall, Inc., New Jersey, pp: 816.
14. Hair, Jr. J.F., R.E. Anderson, R.L. Tatham, W.C. Black, 1998. Multivariate Data Analysis, Prentice-Hall, Upper Saddle River, New Jersey.
15. Demiralay, M. ve. and Çamurcu, 2005. Cure, Agnes, ve k-means Algoritmalarındaki Kümeleme Yeteneklerinin Karşılaştırılması, İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü Dergisi, 4(8): 1-18.