Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1077162

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 5 - Issue 2: 71-80 / April 2022

MULTIDIMENSIONAL SCALING ANALYSIS AND AN APPLICATION

Gülşah KEKLİK[1]*, Mustafa ŞAHİN[2]

[1]*Cukurova University, Faculty of Agriculture, Department of Animal Science, 01250, Adana, Turkey*
[2]*Kahramanmaras University, Faculty of Agriculture, Department of Agricultural Biotechnology, 46050, Kahramanmaras, Turkey*

Abstract: The aim of this study is to determine the factors affecting student's failure. For this, a survey study was carried out. In the questionnaire applied to people of all ages and occupations living in Turkey between 2013 and 2015, inferences were made on the answers of 4183 people, 2928 of whom were men and 1255 were women, regarding the relationship between 26 variables that are closely related to the success factor. In this study, non-metric multidimensional scaling technique was used considering the Euclidean distance. This is because the data is not quantitative and smaller size solutions can be obtained. As a result of the analysis, Kruskal's stress value was observed as 0.00 and no significant difference was observed between the variables. Therefore, the observation of stress values in the range of (0-0.025) clearly reveals that there is a "complete" agreement between the variables. Another one words, the measure of the actual figure compatibility chart obtained S=0.00 and shows full compliance. In this case, we can say that the results obtained adequately reflect the data set we have.

Keywords: Multidimensional scaling analysis, Metric multidimensional scaling, Non-metric multidimensional scaling, Stress value

***Corresponding author:** Cukurova University, Faculty of Agriculture, Department of Animal Science, 01250, Adana, Turkey
E mail: gulsahkeklk@gmail.com (G. KEKLİK)
Gülşah KEKLİK https://orcid.org/0000-0002-1775-2773
Mustafa ŞAHİN https://orcid.org/0000-0003-3622-4543

Received: February 22, 2022
Accepted: March 14, 2022
Published: April 01, 2022
Cite as: Keklik G, Şahin M. 2022. Multidimensional scaling analysis and an application. BSJ Eng Sci, 5(2): 71-80.

1. Introduction

Multidimensional scaling analysis (MDS) is an analysis method that allows the positions of objects to be represented as points in multidimensional space using distances or change information. MDS analysis; it is one of the multivariate statistical analysis methods used in the analysis of some data such as personal preference, attitude, belief (Kurtulus, 1998). The purpose of MDS analysis is to reveal the structure of objects (by using distance values) with a small size, close to their original shape (Tatlidil, 2002). For this reason, the MDS can be used as a size reduction method. As well as being a dimension reduction technique, the MDS is also a helpful technique for examining the dependency structure of the data and establishing hypothesis tests. This analysis is used when univariate statistical analyzes cannot answer the problems. Our main purpose when performing MDS analysis is to classify objects, reduce their size, and simplify problems. The objectives of multivariate analysis methods are to summarize, interpret and use the results of scientific studies and research that can be expressed in numbers. In the MDS analysis, new coordinates are formed to represent the values in the n×n dimensional distance matrix. There are too many algorithms in the MDS analysis, and these algorithms are generally evaluated in two groups as metric and non-metric algorithms. The main purpose of the algorithms is to try to minimize the stress value. MDS found widespread use with the development of computer infrastructures in the 19th and 20th centuries.

The first studies on multidimensional scaling were Young and Household's (1938) study, which showed the applicability of the model proposed by Richardson (1938) for dimension reduction, with the study of Young and Household (1938) that it could be represented in the two-dimensional coordinate system by preserving the distances between units in the Euclidean distance matrix. Two important stages have taken place in the development of the MDS. The first stage is metric approximations. The second stage to be described is the discovery of the non-metric approach to MDS analysis by Shepard (1962) at the Bell Telephone Laboratory after a ten-year hiatus. This approach is also known as the 'Kruskal-Shepard' approach. Kruskal (1964) brought some conceptual innovations to Shepard's (1962) approach. Along with the innovations made by Kruskal (1964) for non-metric MDS analysis, the use of non-metric MDS analysis has also become widespread.

Between 1960 and 1980, many applications were made regarding metric and non-metric MDS analysis, and a long way was covered. The approaches of multivariate data analysis such as regression, cluster analysis, factor analysis, to the interpretation of dimensions, and studies on multidimensional scaling continued in 1980 and later. Young and Householder (1938) with Richardson (1938), size reduction which is considered one of the art multidimensional scaling analysis, multidimensional data is a data visualization technique that enables the display graphically a less extent phrase they have. They have been widely used in different disciplines, as they allow the relationships between units or variables in the data to be seen graphically.

Calis (1995), in his study, found that 8 automobile brands. The consumer perception map was obtained by multidimensional scaling analysis. Findikkaya (1995), in his study, examined the perceptions and similarities of five national newspapers by the readers. Hall (2001) classified the speech samples of a male and a female speaker by using the MDS analysis based on the distance matrix, which consists of perceptual distances. Dogan (2003), the factors affecting growth in lambs were taken into account in four different conditions defined according to gender and birth type, and it was aimed to find the similarities of growth in combinations of these two factors. According to the results obtained in this study, it was found that birth type was prominent in growth in Morkaraman lambs and gender was more important in Akkaraman lambs.

Dura et al. (2004) examined the development level of Turkey against the European Union in terms of human capital. The variables in this study were used together with the MDS analysis and cluster analysis, and it was seen that the 26 countries examined formed 5 different groups, Turkey was in a group alone and did not share the same level of development with any EU country. In their study, they demonstrated the advantages of the proposed method on a real data set. Kacar and Azkan (2005) provided the grouping of the species by multidimensional scaling analysis performed according to the morphological characteristics of different hypericum species collected from various parts of Turkey in 2001 and 2002, and observed that the species formed different groups in the two-dimensional graphical representation as a result of the analysis through the Euclidean distance function.

Simsek (2006) used clustering, multidimensional scaling, confirmatory and explanatory factor analyzes in his study and the multidimensional anger scale consisting of 47 items was used. According to the results of the analysis applied to a total of 542 people in Hacettepe University student dormitories, the number of suitable dimensions in terms of stress value is 4, and the variables in the dimensions differed from other analyzes. Aydin and Baskir (2013), in their study, found similarities in terms of 2012 capital adequacy in 44 banks. In this study, it is aimed to determine the banks that are similar or different in terms of 2012 capital adequacy ratios in the Turkish banking sector operating at an international level. Banks that were structurally similar in the MDS analysis were also found in the same cluster in the cluster analysis and it was considered important for international banks to have a minimum level of capital against risk situations.

Students are subjected to various exams and evaluations throughout their education life. While some of these students are able to fulfill their duties and responsibilities, some of them cannot fulfill their homework due to some environmental, social and psychological reasons. Within the scope of the purpose of this study, the reasons mentioned here were approached with the dimension reduction method and stress statistics values, appropriate dimension numbers and coordinate values of dimensions were calculated over real values.

2. Material and Methods

This study was carried out with the help of a questionnaire in order to observe and measure the reasons for student's failure. The questionnaire applied to people of all ages and occupations living in Turkey between 2013 and 2015, inferences based on various statistical values were made on the answers of 4183 people, 2928 men and 1255 women, regarding the relationship between 26 variables that are closely related to the success factor and all the variables and the factors affecting these variables were determined on the basis of problems that may arise as a result of exchanging ideas with competent people. The questions in this survey study were asked to people between the ages of 18-50 and it was aimed to get answers from all age groups. In this study, non-metric scaling technique was used considering the Euclidean distance. This is because the data is not quantitative, smaller size solutions can be obtained.

2.1. Research Group

The questionnaire was created by the researchers mentioned in this study, and some of these research questions are as follows:

1. Is the stress factor effective in the student's failure?
2. Does working in any job have an effect on the student's failure?

The answers were taken on a 5-point scale expressed as "Very little: 1, Little: 2, Moderate: 3, Much: 4, No effect: 5".

2.2. Measuring and Scale

Measurement is the process of observation and recording. The process of comparing a set of experimental observations of the magnitudes of a variable with the set of numbers to measure that magnitude and ensuring that each quantity matches a number in the set of numbers is called measurement. With another definition, measurement is to give numbers to objects and events in the most general sense in accordance with some rules. It is the process of distinguishing between objects and events at different levels. It is symbols or numbers or data in general that show the distinction made. Measurement, that is, the degree to which an object or individuals have a certain feature may vary from object to object, person to person, situation to situation within the same individual or object from time to time. This change means measuring variables. This measurement process is used to explain the current situation in the variables of the research or the difference between any two different times or situations.

2.3. Similarity Measures for MDS Analysis

MPI analysis has been developed to obtain a graphical view by using similarity, distance or difference information between objects. In this graphical view, the distance function is used when calculating the distances between objects. Each distance measure is a metric. Distance function (equation 1),

$$d: p^{*}p \rightarrow R^{+}(x,y) \rightarrow d(x,y) \quad (1)$$

It is a positive definite function of the form and provides the following four properties (equation 2, 3, 4 and 5):

$$1.\ \forall_{\underline{X},\underline{Y}} \in R^{p} \text{ for } d(\underline{X},\underline{Y}) \geq 0 \quad (2)$$

$$2.\ \forall_{\underline{X},\underline{Y}} \in R^{p} \text{ for } d(\underline{X},\underline{Y}) = 0 \leftrightarrow \underline{X} = \underline{Y} \quad (3)$$

$$3.\ \forall_{\underline{X},\underline{Y}} \in R^{p} \text{ for } d(\underline{X},\underline{Y}) = d(\underline{Y},\underline{X}) \quad (4)$$

$$4.\ \forall_{\underline{X},\underline{Y},\underline{Z}} \in R^{p} \text{ for } [d(\underline{X},\underline{Z}) \leq d(\underline{X},\underline{Y}) + d(\underline{Y},\underline{Z})] \quad (5)$$

MDS analysis treats the distance matrix as the difference matrix. If the data are obtained as intermittent or proportional scale, the dissimilarity values are calculated as Euclidean distance, Quadratic Euclidean distance, Chebyshev distance, City-block distance and Minkowski distance.

1) Minkowski distance: Let r>0 and $X_i, X_k \in R^p \quad i,k = 1,2,\ldots,$ n any two observation vectors. In this case (equation 6);

$$d(\underline{X}_i,\underline{X}_k) = \left[\sum_{j=1}^{p} |x_{ji} - x_{jk}|^{r}\right]^{\frac{1}{r}} \quad (6)$$

is called the Minkowski distance between these two observation vectors.

2) City-block distance: What will be obtained when r=1 in equation (equation 7 and 8);

$$d(\underline{X}_i,\underline{X}_k) = \left[\sum_{j=1}^{p} |x_{ji} - x_{jk}|^{r}\right]^{1/r}, \quad (7)$$

$$d(\underline{X}_i,\underline{X}_k) = \left[\sum_{j=1}^{p} |x_{ji} - x_{jk}|\right] \quad (8)$$

is called the City-block distance between the observation vector $\underline{X}_i$ and $\underline{X}_k$.

3) Euclidean distance: What will be obtained when r = 2 in equation 9;

$$d(\underline{X}_i,\underline{X}_k) = \left[\sum_{j=1}^{p} |x_{ji} - x_{jk}|^{2}\right]^{\frac{1}{2}} \quad (9)$$

its relation is called the Euclidean distance between the observation vectors $\underline{X}_i$ and $\underline{X}_k$ (Figure 1 and Figure 2).

4) When the equation 10 is squared;

$$d^{2}(\underline{X}_i,\underline{X}_k) = \left[\sum_{j=1}^{p} |x_{ji} - x_{jk}|^{2}\right] \quad (10)$$

is called the Quadratic Euclidean distance between the observation vectors $\underline{X}_i$ and $\underline{X}_k$.

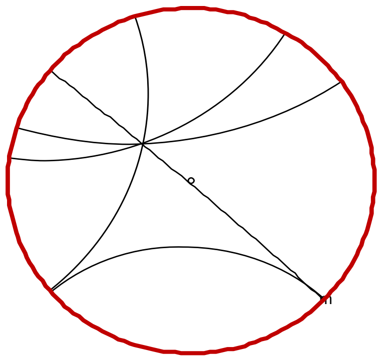

Figure 1. Henri Poincare's disc model.

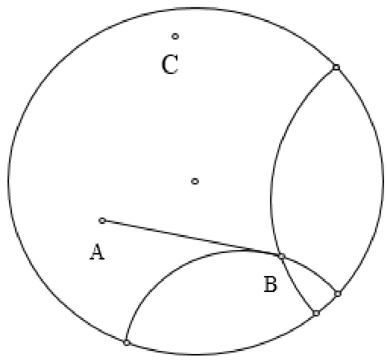

Figure 2. Tangent drawn to lines at the intersection point of lines.

5) Chebyshev distance: The absolute maximum of the difference (distance) between the observation vector $\underline{X}_i$ and $\underline{X}_k$ (equation 11);

$$d(\underline{X}_i,\underline{X}_k) = \max|x_{ji} - x_{jk}| \quad (11)$$

is called the Chebyshev distance. In a data set, the ratio of data being K standard deviations away from the mean is 1-1/K. K is a positive number greater than one. For K=2 and K=3 values; at least 3/4 (75%) of the data are two standard deviations away from the mean, at least 8/9 (89%) are three standard deviations away from the mean.

Although the Euclidean distance function is the most widely used distance function in MDS, there should be difference matrices suitable for the data shape. If the data is binary scaled, it is calculated using one of the following formats: Euclidean distance, square Euclidean distance, sample difference, variance, or lance-williams distance (Ozdamar, 2004). These distances are also called the similarity coefficient. Similarity coefficients, which are a measure of the relationship between any two individuals, take values in the range of [0, 1]. These coefficients are; it is defined differently depending on whether the data is binary, qualitative and quantitative. In this respect, it is useful to examine similarity measures according to data types. The presence of a variable is indicated by 1 or (+), and its absence is indicated by 0 or (–). If the similarity coefficient between any individuals i and k is denoted by d_{ik}, Table 1, called the 2x2 joining table, is used to calculate d_{ik}. Here p is $a+b+c+d=p$ to represent the number of variables. Various similarity coefficients have been proposed for this type of data. Some of these are as given in Table 2.

Table 1. 2×2 join table

	1/+	0/-
1/+	a	b
0/-	c	d

Table 2. Some similarity coefficients for binary and qualitative data

Name	Equation
Euclidean distance	$\sqrt{b+c}$
Quadratic Euclidean distance	$b+c$
Sample difference	$\frac{b+c}{p^2}$
Variance	$\frac{b+c}{4p}$
Lance and Williams	$\frac{b+c}{2a+b+c}$

2.4. Purpose of the MDS method

The MDS is theoretically applied to amorphous data. If the data cannot be researched based on a certain theory, the person conducting the research can apply this method. The data used for MDS is the dissimilarities between pairs of objects. The main purpose of the MDS is to represent these dissimilarities in a lower dimensional space by fitting the distances between points as closely as possible to the dissimilarities.

This method is used by psychologists to determine the psychological dimension in the data in order to discover the psychological structure. Psychologists use this technique to evaluate the way individuals speak and personality structure, while market researchers use this technique to compare consumers' products. This method can be used to easily observe the relationships between objects graphically.

In metric and non-metric MDS, the estimated display (configuration) distances are calculated by choosing the appropriate method according to the distances of the data. PAV (pool-adjacent violator) algorithm can also be used according to the relationship between distances and rankings by using the Shepard algorithm iteratively for non-metric MDS. After the estimated display distances are calculated, a difference matrix is created in the light of these values. Stress statistics are calculated by comparing objects in lower dimensional space through the difference matrix. The ratios developed by Kruskal-Shepard are used to interpret the stress values that show the level of compliance according to the value ranges related to this statistic. The criteria and comments given by Shepard (1962) and Kruskal (1964), regarding stress values are given in the Table 3.

In order to determine the number of dimensions as a result of dimension reduction, the stress statistic is checked and the stress statistic converges to a certain value as a result of iterations, and the number of dimensions belonging to the converged value is selected. As the number of dimensions increases, the difficulty of display will increase, so in practice, two or three dimensions are usually chosen. According to the number of dimensions, the coordinate values of each dimension of the objects are calculated. By looking at these values, an answer can be found to the question of which size and which objects are more dominant.

Table 3. Stress value ranges

Stress Value	Compliance Level
0-0.025	Perfect fit
0.026-0.05	Very good fit (perfect)
0.06-0.10	Good fit
0.11-0.20	Medium fit
over 0.20	Low fit (poor)

One of the important problems encountered in the MDS analysis is to determine the number of dimensions. The number of dimensions indicates the number of coordinate axes. When determining the number of dimensions, attention is paid to determining the appropriate number of dimensions rather than the correct number of dimensions. Since the difficulty of display will increase as the number of dimensions' increases, the graphical arrangement obtained by choosing two or three dimensions in practice is ensured to be understandable and interpretable. Also; the stress value is used when deciding whether the number of dimensions is appropriate. If the stress value is high, there is a large dissonance, and if it is low, there is a low discordance. It is also possible to see this in a graph called Scree Plot, which shows the number of dimensions versus the stress value. The scree plot graph is in the form of an elbow, and the number of dimensions reaching the extreme point of the elbow is preferred (Harman, 1967). Dimensions are named after choosing the number of dimensions in dimensioning in MDS analysis.

2.5. Multidimensional Scaling Techniques

MDS Analysis; according to the type of data, it is divided into two groups as the metric MDS technique and the non-metric MDS technique. Metric MDS technique based on quantitative and metric distances; on the other hand, non-metric MDS technique is applied to sequential and categorical data. The non-metric MDS requires fewer assumptions than the metric MDS and is the most preferred method in analysis. Non-metric at the MDS metric to MDS can be formed by a smaller size solution (Ozdamar, 2004).

2.6. Non-Metric Multidimensional Scaling Technique

Non-metric scaling is applied when orders of magnitude are used instead of numerical values of d_{ij} distances and are ordinal numbers of distance values. In the non-metric approach, the difference measures matrix is taken instead of the distance matrix D. Since an analytical solution is not possible in the general algorithm, the stress value is minimized with an iterative approach. These algorithm steps are as follows:

In the first step; all the elements of the D differences matrix (except the diagonal elements) are sorted (equation 12).

$$d_{i_1 j_1} < d_{i_2 j_2} < \cdots < d_{i_m j_m};\ m = \frac{n.(n-1)}{2} \tag{12}$$

d^*_{ij} values that are monotonically associated with d_{ij} are defined (equation 13).

$$d_{ij} < d_{uv} \rightarrow d^*_{ij} \leq d_{uv} \tag{13}$$

In the second step; the stress value is calculated, which helps us to find the difference between the real shape in multidimensional space (p-dimensional) and the figure predicted in reduced-dimensional (r-dimensional) space. Stress of $\hat{X}$ (equation 14);

$$S(\hat{X}) = \left(\frac{\sum_{i<k}(d_{ik}{}^* - \hat{d}_{ik})^2}{\sum_{i<k}\hat{d}_{ik}^2}\right)^{1/2} \tag{14}$$

measured by the equation.

In the third step; the shape that has the smallest stress value for each r dimension is called the best shape for the r dimension. This is the smallest stress value (equation 15);

$$S_r = minS(\hat{X}) \tag{15}$$

it's like in the equation. S_r is a decreasing function of r.

In the last step; in order to determine the appropriate number of dimensions, this process is continued until the smallest stress value is obtained by calculating $S_1, S_2, \ldots, S_r$ values (Sığırlı et al., 2006).

2.7. Metric Multidimensional Scaling Technique

In metric scaling, the similarities or differences between the observation values obtained from units or objects are expressed with distance values (equation 16).

$$d_{ij} + d_{ik} \geq d_{jk}(\forall_{i,j,k}\ için) \tag{16}$$

The B matrix, which is given in the form and used as the 'metric inequality' and consists of length values, is called the “distance matrix” (Tatlidil, 2002). Both the D matrix and a $B = XX'$ 'format using the X data matrix for n units and p variables. Relationships between D matrix and B matrix (equation 17);

$$d_{ij}^2 - \frac{\sum_{i=1}^{n} d_{ik}{}^2}{n} - \frac{\sum_{k=1}^{n} d_{ik}{}^2}{n} + \frac{\sum_{i=1}^{n}\sum_{k=1}^{m} d_{ik}^2}{n^2} = -2\sum_{j=1}^{m} x_{ij}x_{kj} \quad k, i = 1,2, \ldots, n \tag{17}$$

It is expressed by the equation. Here is the row index, k is the column index, n is the number of units and m is the number of dimensions. In this case, to represent the I_{nxn} dimensional unit matrix and the n dimensional unit vector in L, the B matrix with symmetrical and semi-positive definition (equation 18);

$$B = -\frac{1}{2}\left[I - \frac{1}{n}LL'\right] D^2 \left[I - \frac{1}{n}LL'\right] \tag{18}$$

found using the equation. Since the singular value decomposition of matrix B can be performed with the matrix V, whose columns are the eigenvectors of the matrix B, and the matrix Λ, which is the diagonal matrix of the non-negative eigenvalues of B, in the form B=VΛV′,it can be obtained from matrix B to matrix X. For this (equation 19);

$$B = V\Lambda V' = V\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}V' = XX' \tag{19}$$

equality is used. The columns of the matrix X consist of $(\sqrt{\lambda_j})e_j$ values. Here, λ_j values are the eigenvalues of matrix B, and e_j values are eigenvectors of matrix B. Since the eigenvalues are ordered in descending order, it is aimed to have a smaller representation by determining the required number of maximum $r(r \leq m)$ eigenvalues (Sığırlı et al., 2006). In determining the appropriate number of dimensions, a criterion that is also used in principal component analysis and based only on the eigenvalues of the B matrix can be used (Tatlidil, 2002). This criterion (equation 20);

$$\frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{n}|\lambda_i|} \geq \frac{2}{3} \tag{20}$$

is an inequality. MDS Analysis includes many methods. Although these methods have little differences in terms of application, they are similar to the steps applied in the classical MDS method. The classical MDS method can be summarized in six steps:

1. Selecting an appropriate standardized method depending on the data type and obtaining the transformed data. This step is a step that will be applied if necessary.
2. Calculation of the distance matrix.
3. Deciding how many dimensional space to express n units with 3rd dimensional data matrix. In addition, determining the compatibility of the solutions obtained for each dimension with the original distance matrix (stress measure), deciding which solution will be used and determining which size is the appropriate solution.
4. The regression of estimated display

(configuration) distances with respect to data distances is calculated according to the data type. Estimated display distances are determined through the determined regression equation and these estimated distances are called “differences”.

5. The stress value, which measures the fit between the display distance and the estimated distance, is calculated.
6. Coordinate values of the units are determined according to the m dimension. These coordinates are represented in an m-dimensional space, expressing the position of each unit relative to the other unit. The desired solution is a solution in three dimensions or less. Thus, a more easily traceable graphical view of the units can be obtained.

MDS Analysis does not require any probability distribution assumptions regarding the data. This analysis allows to determine the distances (configuration) that will find the distances between the objects calculated depending on the type of the variables with the least error, with the help of any function (linear, polynomial, monotonic) (Oguzlar, 1995). In the data set and the units of distance measure between to be represented by the MDS in a geometrical space of this distance (e.g., Euclidean in space) used to display. In an m-dimensional Euclidean space, and the distance between points (equation 21),

$$\delta_{ik} = \sqrt{\sum_{a=1}^{m}(x_{ia} - x_{ja})^2} \quad (21)$$

is in the form. The relationship between the configuration distances d_{ik} and the observed distances δ_{ik} is found with the help of a suitable transform $d_{ik} = f(\delta_{ik})$. This relationship can be a linear relationship to be represented by the function $f(\delta_{ik})$=a+bδ_{ik}. Here a and b are the coefficients. We can show this relationship graphically with the Shepard diagram. This graph, on the other hand, is determined by both linear and non-linear forms. The Shepard diagram is drawn according to the display distances and determines which model fits the data better. It creates a scatter chart with the distances observed in the Shepard diagram on the y-axis and the difference (disparities) values on the x-axis (Ozdamar, 2004). The graphical representation of the distances between the objects obtained from the data distance matrix in a less dimensional space in the MDS Analysis is called “graphical representation”. To create the graphical representation, the data coordinates must be converted to graphical representation coordinates with the least error. Among n objects (equation 22);

$$\frac{n(n-1)}{2} \quad (22)$$

distance is calculated. According to these distances, a representation coordinate system very close to the original distances is created in order to obtain the most appropriate geometric representation. The measure that determines the correspondence between the original and display distances is called the “stress measure”. Measure of stress (equation 23);

$$S(\hat{X}) = min(\frac{\sum_{i<k}(d_{ik}-\hat{d}_{ik})^2}{\sum_{i<k}\hat{d}_{ik}^{\,2}})^{1/2} \quad (23)$$

is found by the equation where $\hat{d}_{ik}$ is the estimation of d_{ik} configuration distance. In the interpretation of the stress measure, the tolerance ratios developed by Kruskal-Shepard given in Table 5 were used.

3. Results

In this study, the factors affecting student’s failure were examined. A survey study was conducted on 4183 people. 26 variables were used. Using more than 20 variables makes the analysis lazy and ineffective. In this application, the relations between the variables given, stress values, correlation results and graphics are given.

In Table 4, the difference between the variables was found to be very close to 0, according to the results of the analysis made considering the repetition history (0.00). Therefore, when the factors affecting education are examined, it is clearly seen that the variables are similar to each other. The stress value is 0.00, which indicates this similarity. In addition, while the similarity of the punishment stress value is seen as a result of iterations, the dissimilarity of the punishment values attracts attention. In addition, the iteration ends when the stress value is closest to 0.

In Table 5, stress Value-I=0.00, which means that the difference between the analyzed variables was not observed. In other words, it is said that the compatibility of the examined variables is perfect. Young stress value=0.00 and this value supports the observation result.

In Table 6, the final row coordinates were determined by using the dimension reduction method, which is the basis of the MLS analysis, and the coordinate values in the 1st dimension were in harmony with each other. The 9.29 value in the 4th row of the 2nd dimension caused the difference and got the highest coordinate value.

As evaluated in Table 7, the findings that emerged as a result of the classification of the factors with the same sign in both dimensions; gender, marital status, number of children, the effect of physical disability, low level of perception, the effect of communication skills and the effect of financial opportunities.

Table 4. Repeat history

Iteration	PSV	D	SV	PV
0	1.28		0.31	5.38
5000[a]	0.00	0.00	0.00	2.05

[a]Maximum number of repetitions (MAXITER) exceeded.
PSV= punishment stress value, D= difference, SV= stress value, PV= penalty value.

Table 5. Precautions table

Repeats		5000
Final Function Value		0.00
Function Value Parts	Stress Part	0.00
	Penalty Section	2.05
The Evil of Conformity	Normalized Stress	0.00
	Kruskal's Stress Value-I	0.00
	Kruskal's Stress Value-II	0.00
	Young's S-Stress Value-I	0.00
	Young's S-Stress Value-II	0.00
Goodness of Conformity	Calculated Distribution	1.00
	Calculated Variance	1.00
	Recovered Preference Orders	0.82
	Spearman Rank Correlation Coefficient	0.64
	Kendall Rank Correlation Coefficient	0.56
Variation Coefficients	Variation Affinities	1.12
	Transformed Variation Affinities	1.17
	Variation Affinities	1.20
Degeneration Indices	Sum of Squares DeSarbo's ScramBled Indexes	1.38
	Shepard's Rough Uncommon Index	0.16

Table 6. Final row coordinates

	Size	
	1st	2nd
1	2.53	1.17
2	2.53	1.17
3	2.53	1.17
4	2.25	9.29
5	2.53	1.17

Table 7. Coordinate values of each variable in dimensions (final column coordinates) for general comparison in two-dimensional MDS

	Size	
	1st	2nd
Age	-45.16	0.11
Gender	5.28	4.62
Marital status	2.56	5.57
Number of children	2.56	5.57
Education information	1.75	-3.17
Monthly income	6.15	-1.36
Note effect	0.90	-1.61
Teacher influence	1.75	-3.17
Family influence	1.75	-3.17
Diet effect	0.45	4.41
Health effect	1.75	-3.17
Work habits	1.75	-3.17
Stress effect	1.75	-3.17
Work effect	1.75	-3.17
The effect of early school initiation	0.45	4.41
Seasonal variation	0.45	4.41
The effect of repetition	1.75	-3.17
Sporting activity effect	3.24	-3.19
Parental separation effect	1.75	-3.17
Bodily influence	0.90	-1.61
Social influence	1.75	-3.17
Low level of perception	0.90	-1.61
The impact of course activities	0.45	4.41
The effect of communication skills	0.90	-1.61
Material effects	-0.90	-1.61
Siblings influence	1.75	-3.17

The findings that emerged as a result of the classification of factors with opposite signs in both dimensions; age, education information, monthly income, teacher effect, family status, nutrition effect, health effect, work habits effect, stress effect, effect of working in any job, effect of starting school early, effect of seasonal change, effect of course repetition, sportive activity effect, the effect of parental separation, the effect of social environment, the effect of participation in course activities and the effect of the number of siblings. As can be seen in the scattering diagram in Figure 3, it is observed that the coordinates of the other answers are close to each other, except for the "many" answer. In the representation of the coordinates of the answers in two-dimensional space, it is observed that the answer "many" is far from the other answers and differs from the others.

The Scatter Diagram showing the distribution of observational distances and differences is shown in Figure 3 and Figure 4. It is clear that there is a linear relationship between the differences in shape and the distances between the variables. It reveals that the considered (estimated) distances agree with the real values and a suitable solution can be found through linear modeling. In Figure 4, it is observed that the coordinates of the factors other than age are close to each other. In the representation of the coordinates of the factors affecting the success of the variables in two-dimensional space, the age coordinates are far from other factors and differ.

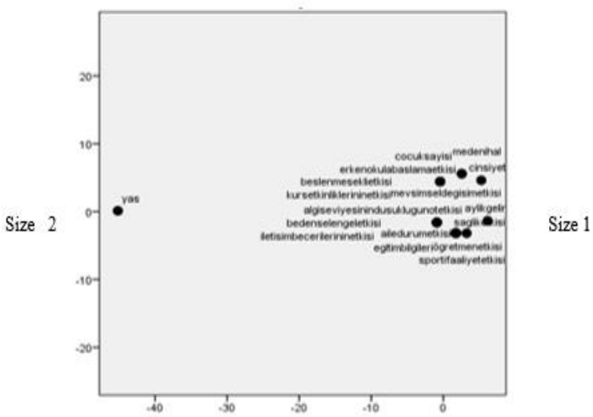

Figure 3. Scatter diagram for responses.

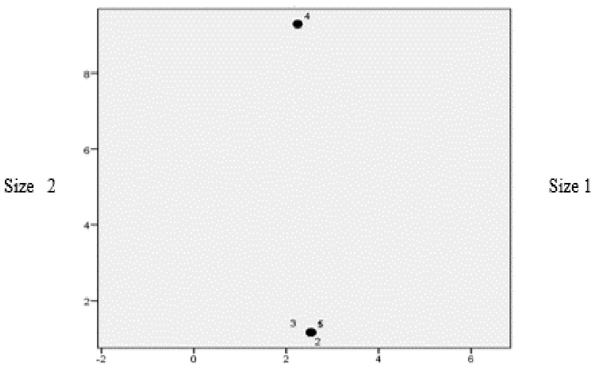

Figure 4. Common scattering diagram

4. Discussion

Multidimensional scaling used to express the interrelationships of objects or units in a less dimensional space; it can be applied on some types of data with the help of ordinal, evenly spaced and proportional scale and it has a wide usage area. In stress value applications, Kruskal's Shepard Diagram is used. Algorithms used in multidimensional scaling analysis compare the similarities of objects or units in pairs, triplets or multiple groups. The necessity of eliminating the differences between the factors in terms of the variables discussed in the study and bringing all the variables to the same level is clearly seen. In this respect, the plans to be prepared have been prepared in a way that will eliminate this difference between the factors affecting success. If other indicators are added to the indicators discussed here, it is possible that changes will occur in the analysis results.

In this study, a questionnaire study was conducted to determine the factors affecting student’s failure. In the questionnaire applied to people of all ages and occupations living in Turkey between 2013 and 2015, inferences were made based on various statistical values on the answers of 4183 people, 2928 of whom were men and 1255 were women, regarding the relationship between 26 variables that are closely related to the success factor. The study was carried out on groups of which most of them were university graduates. In the analysis of the study, non-metric scaling technique was used considering the Euclidean distance. During the study analysis; survey stress values, two-dimensional scatter diagrams (according to the responses and the state of the variables within each other) were created. Row and column coordinates were determined by considering iteration (repeat) histories. Inferences were made by determining the dimensions of the variables in the coordinates.

5. Conclusion

As a result of successive iterations in Table 4, the difference matrix was found to be 0.00, and the stress value in the same iteration was observed to support this result. The sections discussed show parallelism with the work of Yigit (2007). The fact that some cities are visibly separated from each other while others are gathered together supports the results in this study and shows similarities. Considering the measures table in Table 5, the Kruskal Stress value was observed and the difference was found to be very close to 0 (0.00). This value is an indicator of high cohesion and is in line with the studies of Davison (1983), Costa et al. (2005), de Leeuw (2000) and Yigit (2007).

The common scattering diagram in Figure 4 supports the values in Table 4 and Table 5, and it is appropriate to say that all the factors affecting education are similar except for a few factors. This result is also consistent with the analysis of another study by Alan (2008). In addition, the fact that the correlation coefficient was close to 1 supported the similarity between the variables.

As the number of dimensions' increases, the difficulty of display will increase, so in practice, two or three dimensions are usually chosen. In this study, the data were analyzed in two dimensions. Coordinate values of each dimension of the objects were calculated according to the number of dimensions. By looking at these values, it was stated which objects were more dominant in which dimension. The value that is independent of the others is the "age" factor as seen in the general scatter diagram; It is clear that there is a "many" answer as seen in the diagram regarding the answers. Compared to other analysis methods, it can be stated that the MLS analysis is the most appropriate method in determining the distances/closeness and dissimilarities/similarities between variables and gives relatively better results than other methods.

As a result, Kruskal's Stress Value was observed as 0.00 in this study, and no significant difference was observed between the variables. Young's S Stress Value = 0.00, which supported this situation. Therefore, the observation of the stress values in the range (0-0.025) clearly reveals that there is a "complete" agreement between the variables. In their study, Karaçam and Tolan (2014) investigated the prevalence of cigarette, alcohol and other addictive substances use among 830 university students who were educated at Ege University and applied to the outpatient clinics of Health, Culture and Sports Department Health Branch, other than psychiatry, between 2006 and 2008, and the prevalence of this situation in some socioeconomic conditions. He applied multidimensional scaling method while explaining the relationship with demographic characteristics and perceptions of social life and mood. Smoking and alcohol use were located far from academic achievement and religious belief, and were found to be consistent with previous findings. This is compatible with the risk factors stated for smoking, alcohol and other addictive substances. However, negative features that can be called destructive behaviors and emotions are associated with addictive substance use and they are located close to each other. Therefore, considering the approach and attitude in analyzing the subject with the distance model, it is clear that it is similar to this study.

In this study, some factors affecting the reasons for student's failure were observed and measured, and the questionnaire could be answered without keeping the age range in the next study. For example, research can be done so that the age range is limited to middle school or high school students. In future studies, a research can be developed on teacher's or prospective teachers' perceptions of success, prestige and earnings. In addition, a new survey study can be conducted on the roles and concepts that teachers undertake in increasing the success of students, and the stress values of the data obtained can be determined. It can be suggested to reevaluate the prominent factors by examining the variations of the variables with each other.

Author Contributions

All authors have equal contribution and the authors reviewed and approved the manuscript.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgements

This study is based on a master's thesis titled "Multidimensional Scaling Analysis and an Application". It was completed in 2017 by Gülşah KEKLİK in consultation of Mustafa ŞAHİN.

References

Alan S. 2008. Çok boyutlu ölçekleme yöntemi ve bir uygulama (Yayınlanmamış yüksek lisans tezi), Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, Turkey.

Aydın D, Başkır M. 2016. Bankaların 2012 yılı sermaye yeterlilik rasyolarına göre kümeleme analizi ve çok boyutlu ölçekleme sonucu sınıflandırılma yapıları. Banka Sigort Araş Derg, 1(5): 29-47.

Çalış AF. 1995. Tüketici marka algılamaları ve satın alma davranışı: Binek otomobil örneği, (Basılmamış Yüksek Lisans Tezi), Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü), İzmir, Turkey.

Davison ML. 1983. Introduction to multidimensional scaling and its applications. App Psychol Measur, 7(4): 373-379.

de Leeuw J. 2000. Multidimensional scaling. UCLA: Department of Statistics, UCLA, US, pp. 13.

Doğan İ. 2003. Kuzularda büyümenin çok boyutlu ölçekleme yöntemi ile değerlendirilmesi. Uludağ Üniv Vet Fak Derg, 22(1): 33-37.

Dura C, Atik H, Türker O. 2004. Beseri sermaye açısından Türkiye'nin Avrupa Birliği karşısındaki kalkınma seviyesi. 3. Ulusal Bilgi, Ekonomi ve Yönetim Kongresi Bildiriler Kitabı, 24-26 Kasım 2004, Eskişehir, Turkey, pp.13-20.

Fındıkkaya A. 1995. Çok boyutlu ölçekleme analizi ve bir uygulama denemesi. (Yüksek Lisans Tezi), Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı İstatistik Bilim Dalı, Bursa, Turkey, pp. 74.

Hall JL. 2001. Application of multidimensional scaling to subjective evaluation of coded speech. J Acoustical Society of America, 110: 2167.

Harman HH. 1967. Modern factor analysis, 3rd ed. University of Chicago Press, Chicago, US, pp. 508.

Kaçar O, Azkan N. 2005. Türkiye'de bulunan bazı Hypericum L. (Hyperiaceae) türleri üzerinde sayısal taksonomik bir çalışma. Fırat Üniv Fen Müh Bil Derg, 17(2): 424-434.

Karaçam Ö, Totan T. 2014. Ege üniversitesi öğrencileri arasında bağımlılık yapıcı madde kullanımının çok boyutlu ölçekleme yöntemiyle incelenmesi. Anatolian J Psychiatry, 15(2): 16-23.

Kruskal JB. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29(1): 1-27.

Kurtuluş K. 1998. Pazarlama araştırmaları. Avcıol Basın-Yayın, İstanbul, Turkey, pp. 715.

Oğuzlar A. 1995. Çok boyutlu ölçekleme ve kümeleme analizi arasındaki ilişkiler. Uludağ Üniv İİBF Derg, 18(2).

Özdamar K. 2004. Paket programları ile istatistiksel veri analizi. Kaan Kitabevi, Eskişehir, Turkey, pp. 518.

Richardson MW. 1938. Multidimensional psychophysics.

Psychol Bul, 35: 659-660.
Shepard RN. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function I. Psychometrika, 27(3): 219-246.
Sığırlı D, Ediz B, Cangür Ş, Ercan İ, Kan İ. 2006. Türkiye ve Avrupa Birliği'ne üye ülkelerin sağlık düzeyi ölçütlerinin çok boyutlu ölçekleme analizi ile incelenmesi. J Turgut Ozal Medical Center, 13(2): 81-85.
Şimşek D. 2006. Kümeleme analizi, çok boyutlu ölçekleme, doğrulayıcı ve açıklayıcı faktör analizi ile elde edilen yapı geçerliği kanıtlarının karşılaştırılması (Yayınlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, Turkey.
Tatlıdil H. 2002. Uygulamalı çok değişkenli istatistiksel analiz. Ankara Ziraat Matbacılık, Ankara, Turkey, pp. 279.
Yiğit E. 2007. Çok boyutlu ölçekleme yöntemlerinin incelenmesi ve bir uygulama. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Yayınlanmamış Yüksek Lisans Tezi, Samsun, Turkey.
Young G, Householder AS. 1938. Discussion of a set of points in terms of their mutual distances. Psychometrika 3: 19-22.