KSÜ Tarım ve Doğa Derg 26 (1), 201-209, 2023
KSU J. Agric Nat 26 (1), 201-209, 2023
https://doi.org/10.18016/ksutarimdoga.vi.984229

Use of Multidimensional Scaling Analysis Together With Multivariate Analysis of Variance in Determining Differences Between Groups

Rabia ALBAYRAK DELİALİOĞLU[1], Zahide KOCABAŞ[2]
[1,2]Ankara University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, Ankara, Türkiye
[1]https://orcid.org/0000-0002-1969-4319, [2]https://orcid.org/0000-0002-0310-4062
✉: ralbayrak@ankara.edu.tr

ABSTRACT
Multidimensional Scaling analysis (MDS) and Multivariate analysis of variance (MANOVA) are among the commonly used multivariate statistical methods. While MANOVA is used to evaluate whether there are statistically significant differences between the mean vectors of the experimental groups in terms of more than one independent variable; MDS analysis is used both for dimension reduction and to classify individuals/variables according to their differences. In cases where the relationships between individuals/variables are not known, but the distances between them can be calculated, MDS analysis allows to reveal the relationships between individuals by using these distances, and unlike MANOVA, it does not require any assumptions. In this study, the numerical values produced by the simulation technique were used as the input data, with reference to the real data regarding 5 kinds of pistachios in terms of 13 fatty acids. These data were evaluated with both the MDS analysis and the MANOVA test and the results were interpreted. Considering the convenience in the evaluation of the data, the usability of the MDS analysis as supportive to the MANOVA test and subsequent multiple comparison tests was evaluated.

Animal Science

Research Article

Article History
Received : 18.08.2021
Accepted : 31.01.2022

Keywords
Multidimensional scaling
Stress value
Classification
MANOVA
Multiple comparison test

Gruplar Arası Farklılıkların Belirlenmesinde Çok Boyutlu Ölçekleme Analizinin Çok Değişkenli Varyans Analizi ile Birlikte Kullanımı

ÖZET
Çok Boyutlu Ölçekleme (ÇBÖ) analizi ve Çok değişkenli varyans analizi (MANOVA) yaygın olarak kullanılan çok değişkenli istatistik yöntemler arasında yer almaktadır. MANOVA, birden fazla bağımsız değişken bakımından deney gruplarının ortalama vektörleri arasında istatistik olarak önemli farklılıklar olup olmadığını değerlendirmede kullanılırken; ÇBÖ analizi, hem boyut indirgeme hem de bireyleri/değişkenleri farklılıklarına göre sınıflandırmak için kullanılır. ÇBÖ analizi, bireyler/değişkenler arasındaki ilişkilerin bilinmediği fakat aralarındaki uzaklıkların hesaplanabildiği durumlarda, bu uzaklıklardan yararlanarak bireyler arasındaki ilişkilerin ortaya koyulmasına olanak tanır ve MANOVA'nın aksine herhangi bir varsayım gerektirmez. Bu çalışmada girdi verisi olarak 13 yağ asidi bakımından 5 çeşit antepfıstığına ilişkin gerçek veriler referans alınarak simülasyon tekniği ile üretilen sayısal değerler kullanılmış ve bu veriler hem ÇBÖ analizi hem de MANOVA testi ile değerlendirilmiştir. Her iki test sonucunda da elde edilen bulguların benzer olduğu ve verilerin analizi, yorumlanması ve sonuçların değerlendirilmesindeki kolaylıklar göz önünde bulundurulduğunda, MANOVA testi ve sonrasında yapılacak olan çoklu karşılaştırma testlerini destekleyici olarak ÇBÖ analizinin kullanılabilirliği değerlendirilmiştir.

Zootekni

Araştırma Makalesi

Makale Tarihçesi
Geliş Tarihi : 18.08.2021
Kabul Tarihi : 31.01.2022

Anahtar Kelimeler
Çok boyutlu ölçekleme
Stres değeri
Sınıflama
MANOVA
Çoklu karşılaştırma testi

To Cite : Albayrak Delialioğlu, R., & Kocabaş, Z., (2023). Use of Multidimensional Scaling Analysis together with Multivariate Analysis of Variance in Determining Differences between Groups. *KSU J. Agric Nat* *26* (1), 201-209. https://doi.org/10.18016/ksutarimdoga.vi.984229

Atıf Şekli: Albayrak Delialioğlu, R., & Kocabaş, Z., (2023). Gruplar Arası Farklılıkların Belirlenmesinde Çok Boyutlu Ölçekleme Analizinin Çok Değişkenli Varyans Analizi ile Birlikte Kullanımı. *KSÜ Tarım ve Doğa Derg 26* (1), 201-209. https://doi.org/10.18016/ksutarimdoga.vi.984229

INTRODUCTION

Multivariate analysis methods aim to obtain a general result by considering the relationships between two or more random variables as a whole. However, with the increase in the number of variables studied, the number of dimensions also increases, making it difficult to interpret the results obtained. For this reason, most of the multivariate statistical analysis methods are based on dimension reduction (Jobson, 1992). One of them is Multidimensional Scaling (MDS) analysis.

MDS analysis is a multivariate method used to classify variables/individuals, which allows modelling nonlinear relationships between variables and evaluating all data types. Unlike other multivariate methods, it does not require assumptions like data type, relationships between variables, and multivariate normal distribution (Yiğit & Mendeş, 2016). MDS analysis evaluates the differences and similarities between data, individuals, variables and even events, yielding graphical results that can be easily interpreted by anyone. Due to these advantages, it has found a wide range of use in practice (Jaworska & Anastasova, 2009).

One of the areas where MDS analysis is widely used is agriculture. Many researchers classify the trial material they are working on with MDS analysis in terms of various properties. In the study conducted by Suarez et al. (2016) on 30 sweet potato varieties; the nutritional composition, mineral and trace element amounts of potatoes were determined and significant differences were found between the varieties in terms of these characteristics. Then, sweet potato varieties were classified by MDS analysis. Yamamoto et al. (2015) improved a image analysis system which can simultaneously assess multiple appearance properties of strawberries, in detail. Then, they tried to reveal the efficiency of the system using clustering, MDS and discriminant analysis. Can et al (2021) evaluated the heavy metal accumulation that occurs as a result of intensive production in fruit and vegetables produced today. For this purpose, the most common fruity vegetables in Kyrgyzstan markets (ten different fruity vegetables including tomato (2), pepper (5), eggplant, cucumber and zucchini) were included in the study and their B, Ca, Cd, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb and Zn contents were measured. Differences between fruity vegetables in terms of measurements were evaluated by Kruskal Wallis test and these fruity vegetables classified by MDS analysis. s a result of the MDS analysis, the fruity vegetables evaluated within the scope of this study were clearly divided into four groups in terms of mineral nutrients and heavy metal contents. In the study by Lopes et al (2017), the results obtained by electrical impedance spectroscopy (EIS) and standard chemical analyses were compared regarding characterization different wine varieties. Hence, impedance parameters and chemical analysis results of 16 Portuguese wines were evaluated using MDS analysis. Consequently, it was observed that the wines could be classified with the impedance data obtained from the EIS, based on the strong correlations found between the electrical measurements of the wine and its chemical properties. This conclusion has been confirmed through MDS-maps.

In this study, numerical values produced by simulation technique, by taking the real values of 13 fatty acids of 5 types of pistachios as reference, were used as input data. These data were evaluated with metric-MDS analysis and the similarity of the results obtained with the MDS method and multivariate analysis of variance (MANOVA) technique was shown.

MATERIAL and METHOD

The data used in this study were produced by simulation technique with the help of Microsoft Power Station Developer Studio and IMSL Library in the FORTRAN PowerStation 4.0 package program, with reference to the mean, standard deviation and correlation structure of the fatty acids (*myristic acid, palmitic acid, palmitoleic acid, margaric acid, margaoleic acid, stearic acid, oleic acid, linoleic acid, linolenic acid, arachidic acid, gadoleic acid, behenic acid and lignoseric acid*) measured in 5 varieties of pistachio (*Siirt, Uzun, Halebi, Kırmızı, Ohadi*) in the study of Çınar (2012). The mean and standard errors for the simulated data were given in Table 1.

As a result of the simulation study, the data produced in terms of 13 fatty acids for a total of 50 pistachios, 10 of each variety, were evaluated using the NCSS 2007 Version 07.1.5.

Multidimensional Scaling Method

In the MDS method, depending on the type of variables, the configuration distances (δ_{ij}) that will represent the original distances (d_{ij}) with the least error are determined and displayed graphically in a less dimensional space. For this, the data coordinates must be converted to graphical representation coordinates with the least error. This criterion, which measures the discrepancy/difference between the original distances and the configuration distances, is called the stress value (Özdamar, 2004).

The MDS method can generally be grouped under two headings, metric and non-metric. If the data to be analysed is obtained in nominal or ordinal scale, non-metric, if it is obtained in interval or ratio scale, metric MDS analysis is used. In the metric method, the solution is done with an approach similar to the principal component analysis, while in the non-metric method, the analysis is made by using the rank numbers of the distances. While performing MDS analysis, various input (similarity or dissimilarity) matrices can be used according to the data structure and purpose of the researcher (Cox & Cox, 2001).

Table 1. Descriptive statistics of simulated data
Çizelge 1. Üretilen verilere ait tanıtıcı istatistikler

Fatty Acid *Yağ asidi*	Siirt (n=10)	Uzun (n=10)	Halebi (n=10)	Kırmızı (n=10)	Ohadi (n=10)
Myristic	0.07±0.001	0.10±0.003	0.09±0.001	0.10±0.001	0.08±0.004
Palmitic	7.61±0.034	8.67±0.031	8.81±0.023	8.81±0.027	8.52±0.034
Palmitoleic	0.53±0.006	0.69±0.018	0.77±0.006	0.65±0.011	0.65±0.006
Margaric	0.04±0.001	0.04±0.001	0.04±0.002	0.05±0.007	0.06±0.004
Margaoleic	0.08±0.002	0.06±0.002	0.08±0.010	0.08±0.004	0.06±0.001
Stearic	1.96±0.008	1.77±0.002	1.82±0.012	2.11±0.012	1.26±0.009
Oleic	69.78±0.149	64.08±0.077	71.95±0.193	71.05±0.123	58.21±0.095
Linoleic	18.48±0.044	23.07±0.048	14.94±0.088	15.65±0.055	29.98±0.028
Linolenic	0.34±0.006	0.35±0.006	0.34±0.012	0.36±0.019	0.34±0.004
Arachidic	0.17±0.003	0.14±0.003	0.16±0.004	0.20±0.005	0.12±0.011
Gadoleic	0.65±0.003	0.52±0.013	0.43±0.006	0.47±0.006	0.55±0.018
Behenic	0.19±0.015	0.18±0.014	0.19±0.009	0.25±0.017	0.23±0.018
Lignoseric	0.13±0.006	0.16±0.002	0.1±0.006	0.17±0.037	0.22±0.002

In MDS analysis, the distances (d_{ij}) between the i^{th} and j^{th} individuals/variables in the data set are calculated and these distances are represented in a geometric space (Euclidean space, etc.). In the obtained p-dimensional Euclidean space, the relationship between the original distances (d_{ij}) and the configuration distances (δ_{ij}) can be graphically represented by the Shepard diagram. Shepard diagram is a scatterplot with observed distances on the Y-axis and configuration distances on the X-axis. By looking at the Shepard diagram and the pseudo-R^2 statistics, which is an index similar to the determination coefficient in regression analysis, the goodness of fit between the observed distances and the configuration distances can be observed (Shepard, 1962; Özdamar, 2004; Yiğit, 2007; Gündüz, 2011; Mair et al., 2016).

Metric-multidimensional scaling method

The first foundation of this method, known as metric or classical MDS, was laid by Young and Householder (1938) in the 1930s. Later, the Psychometrics group at Princeton University, which included Messick-Albelson (1956) and Torgerson (1952), conducted studies on this subject. Torgerson first demonstrated the applicability of the metric-MDS method for interval and ratio data in a paper he published in 1952 (Young and Hamer, 1987).

Metric-MDS analysis uses nxn dimensional proximity matrix as input data. The objective is to find estimated distances (configuration distances) which are approximately equal to the observed ones, in the distance matrix (D) for k-dimensional space. Computation of the coordinates to represent individuals is possible by finding the eigenvectors of the B-matrix from which the coordinates of the distance matrix will be obtained. Therefore, in the metric-MDS method, the B-matrix must be obtained first. For this purpose, the following steps are followed (Tatlıdil, 2002; Sığırlı et al., 2006; Alpar, 2013).

1. Firstly, the data is standardized using the appropriate method. Later, nxn-dimensional distance matrix (D) is created by using the distance measure (usually Euclidean distance is preferred in practice) that we determine in accordance with the structure of the data.
2. D-matrix is not a positive semi-definite matrix because its diagonal elements are '0'. However, by using this matrix positive definite B-matrix can be obtained. For this purpose, the matrix A must first be obtained primarily. The nxn dimensional A-matrix is obtained by Equation 1.

$$A = (a_{ij}) = \left(-\frac{1}{2} d_{ij}^2 \right) \quad (1)$$

d_{ij}: elements of the D-matrix (observed distances)

3. Using the A-matrix, the B-matrix (Equation 2.), which is a symmetrical matrix that can be divided into diagonal elements and column vectors, is created.

$$B = -\frac{1}{2}\left[I_n - \frac{1}{n} i_n i'_n \right] D^2 \left[I_n - \frac{1}{n} i_n i'_n \right] \quad (2)$$

I_n: Identity matrix with nxn dimensional, i_n: Unit vector with nx1 dimensional, D^2: The matrix obtained by squaring the elements of the matrix D.

4. The eigenvalues and eigenvectors of the B-matrix are calculated. Since the B-matrix is a positive semi-definite matrix, the number of positive eigenvalues is equal to the number of dimensions of the distance matrix (matrix D). The B-matrix is expressed as $B = V\Lambda V'$ (V: matrix of eigenvectors of the B-matrix, Λ: matrix whose diagonal elements are eigenvalues).

5. After finding the eigenvalues and eigenvectors of the B-matrix, the graphical coordinates are found with the help of $\sqrt{\lambda_i} v_i$. (λ_i: The eigenvalue of the B-matrix calculated for the i^{th} dimension, v_i: Eigenvectors corresponding to the i^{th} dimension of matrix B.

6. It is necessary to decide in how many dimensional space data matrix will be represented. Therefore, MDS solutions are obtained for each dimension. Then, the goodness of fit of each analysis to the real distance matrix, that is, the stress value, is calculated and it is decided which analysis will be applied. In practice, 2 or 3 dimensions are generally preferred for easy interpretation.

Stress Value

The stress value is used to determine whether the number of dimensions obtained as a result of the MDS analysis is appropriate. The stress value is the sum of the deviations of the points from the regression line in the Shepard diagram and is calculated to determine the correspondence between the observed distances and the configuration distances. The stress value is, in a way, a statistic similar to the correlation coefficient. However, it does not measure the goodness of fit, but the badness of fit. The Stress (STandardized REsidual Sum of Squares) value is a measure of the difference between the multidimensional (p-dimensional) real model and the model estimated in reduced (k-dimensional) space. In other words, it measures the discordance between the observed distances and the configuration distances and is calculated as in Equation 3 (Borg and Groenen, 2005).

$$Stress = \sqrt{\frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(d_{ij}-\delta_{ij})^2}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}d_{ij}^2}} = \sqrt{\frac{\sum(d_{ij}-\delta_{ij})^2}{\sum_{i<j}d_{ij}^2}} \qquad (3)$$

d_{ij}: observed distances between i^{th} and j^{th} points, δ_{ij}: Estimated configuration distances between the i^{th} and j^{th} points as a result of the c^{th} iteration.

According to the magnitude of the stress value obtained as a result of the MDS analysis, the fitness of the configuration distances with the observed distances can be classified as in Table 2. The main purpose in both metric and non-metric approach; is to minimize the stress value, which is an indicator of the discordance between the observed and the configuration distances.

In MDS analysis, it is desired that the stress value be as close to zero as possible. A stress value of exactly zero indicates perfect fit, while a stress value equal to 1 indicates complete incompatibility (Borg and Groenen, 2005; Borg et al., 2013).

Table 2. Classification of stress value
Çizelge 2. Stres-değerinin sınıflandırılması

Stress value *Stres değeri*	Goodness of fit *Uyum iyiliği*
0.20 ≤ stress	Poor
0.10 ≤ stress < 0.20	Fair
0.05≤ stress < 0.10	Good
0.025≤ stress < 0.05	Excellent
stress < 0.025	Perfect

In the MDS analysis, increasing the number of dimensions decreases the stress value. However, it is necessary to establish a balance between the stress value and the number of dimensions in order to interpret the obtained dimensions and express the results easily (Cox and Cox, 2001; Alpar, 2013).

Shepard Diagram and Pseudo-R^2 Statistic

The diagram showing how well the obtained MDS model fits the data and the compatibility of observed distances and configuration distances is called Shepard diagram. With this diagram, the linearity of the fit can be examined. If the fit is good, the points are located on or around the 45^0 line. In the Shepard diagram, the distances observed on the Y-axis and the configuration distances on the X-axis are located (Shepard, 1962).

How well the configuration distances adapt to the observed distances is measured by the degree of linear relationship between the two features in question, that is, the square of the correlation coefficient (R^2). The pseudo-R^2 statistic calculated in the MDS analysis is an index similar to the coefficient of determination in the regression analysis. The pseudo-R^2 statistic gives a measure of how much of the sum of the squares of the mean-corrected dissimilarity values can be explained by using the number of dimensions determined as a result of the MDS analysis. Pseudo-R^2 statistic can be calculated using the Equation 4 (Cox and Cox, 2001; Alpar, 2013).

$$Pseudo\text{-}R^2 = 1 - \frac{\sum_{i=1}^{n}\left(d_{ij}-\delta_{ij}\right)^2}{\sum_{i=1}^{n}\left(d_{ij}-\bar{d}\right)^2} \qquad (4)$$

d_{ij} : Observed distance or proximity values, δ_{ij} : Configuration distances, $\bar{d}$: Average of observed distances.

In order for the number of dimensions obtained as a result of the MDS analysis to be sufficient, the pseudo-R^2 statistic must be greater than 0.80 (in some sources, it is 0.60). Thus, it is understood that the obtained configuration distances are in good agreement with the observed distances (Gevrekçi et al., 2011; Alpar, 2013).

Multivariate Analysis of Variance (MANOVA)

In most biological events, the effects of factors on more than one variable is curious. For this purpose, MANOVA test is widely used and applied by following the steps below:

1. The experimental units should be chosen randomly from the population, the observation values should be independent from each other, the data should be continuous and show multivariate normal distribution, the number of experimental units should be more than the number of variables, and the variance-covariance matrix should be homogeneous.

2. After determination the control hypothesis (H_0 : The differences between the mean vectors of the groups in terms of the studied features are not statistically significant) and the alternative hypothesis (H1: The difference between the mean vectors of at least two groups in terms of the studied features is statistically significant), the test statistic is calculated.

3. The most commonly used test statistics for hypothesis control are; Wilks' lambda, Hotelling's trace, Pillai's trace, and Roy's largest roots statistics. The Pillai's Trace test statistic used in this study is obtained as in Equation (5).

$$T = \sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i} \tag{5}$$

In this equation, the λ_i values are the eigenvalues of the BW^{-1} matrix product (B: sum of squares matrix between groups, W: sum of squares matrix within groups). Calculated T-value is then converted to the F_T value, showing the F-distribution with '$s(2m+s+1)$' and '$(s(2 +s+1))$' degrees of freedom, using Equation (6).

$$F_T = \frac{2n+p+1}{2m+p+1} \times \frac{T}{p-T} \tag{6}$$

n: The number of observation in each group, p : the number of variable, $m = \frac{|p-(k-1)|-1}{2}$,

$s = \min(k-1, p)$, k : the number of mean vectors,

$$\tilde{n} = \frac{N-p-k-1}{2}$$

4. If the control hypothesis is rejected as a result of the hypothesis test, it is determined that the differences between the mean vectors of which groups are statistically significant with the appropriate multiple comparison test. For this purpose, commonly used test are: simultaneous confidence interval method, Bonferroni confidence interval method and Mahalanobis distance. However, in terms of convenience, the most common application is to perform ANOVA test for each variable separately, although it ignores the relationships between variables (Al-Abdullatif et al 2019). In addition to these, Discriminant Analysis is also widely used for this purpose (Al-Abdullatif, 2020). The Mahalanobis distance used in this study can be calculated with Equation (7).

$$D_{ij}^2 = (\mu_i - \mu_j)' \; S^{-1} (\mu_i - \mu_j) \tag{7}$$

μ_i : i. grubun ortalama vektörü, μ_j : j. grubun ortalama vektörü, S^{-1} :ise gruplar içi varyans-kovaryans matrisinin tersi

The calculated D^2 values are then converted to F-values using Equation (8).

$$F = \frac{n_i n_j (n_i + n_j - p - 1)}{p(n_i + n_j)(n_i + n_j - 2)} D^2 \tag{8}$$

Finally, the F-value obtained by Equation (8) is compared with the F-table value with '(p)' and '(ni+nj-p-1)' degrees of freedom, and it is determined that the differences between the mean vectors of which groups are statistically significant (Jobson, 1992, Alpar 2013).

RESULTS and DISCUSSION

In most of the studies, the assumptions of the MANOVA technique may not be fulfilled. In addition, in many studies, the number of variables may be higher than the number of experimental units. These and similar situations make it impossible to use the MANOVA technique. Even if all the assumptions are fulfilled, the multiple comparison tests done after MANOVA test for factorial designs are quite complex. For this purpose, the usability of MDS analysis which was not need any assumptions was evaluated in this study, and the results of MANOVA and MDS analysis were compared.

The most important issue to be considered while performing MDS analysis is to determine the number of dimensions. Although it is desired that the stress value be as low as possible and the pseudo-R^2 statistic be as high as possible in theory, the determination of more than 3 dimensions, in practice, makes it difficult to evaluate the results of the study. For this reason, the balance between the number of dimensions and obtaining interpretable results should be well established and the number of dimensions should be determined as 2 or 3 maximum. However, in some studies, it is unavoidable to take more than 3 dimensions. In such cases, since it will not be possible for the researcher to show all dimensions on the same map, she/he can interpret the results obtained by creating different 2-dimensional maps for binary combinations of dimensions (Buja et al., 2008; Dumanoğlu et al., 2018).

The eigenvalues obtained as a result of the metric-MDS analysis for the fatty acids studied in pistachios are given in Table 3.

Table 3. Eigenvalues obtained as a result of metric-MDS analysis
Çizelge 3. Metrik-MDS analizi sonucunda elde edilen özdeğerler

Number of dimension *Boyut sayısı*	Eigenvalue *Özdeğer*	Individual % *Bireysel %*	Cumulative % *Birikimli %*
1	2894.78	99.10	99.10
2	20.66	0.71	99.81
3	3.61	0.12	99.94
...	....	...	...

As seen in Table 3, the eigenvalue of the first dimension is 2894.78, which explains 99.1% of the total variation. As it can be easily estimated when looking at the eigenvalues given in Table 4, as a result of the metric-MDS analysis made for the data it was decided that it would be sufficient to draw a map by considering only the first dimension. This situation can be seen more clearly when looking at the stress values given in Table 4.

Table 4. Stress values and pseudo-R^2 statistics
Çizelge 4. Stres-değeri ve yalancı-R^2 istatistiği

Dimension *Boyut*	Stress value *Stres değeri*	Pseudo-R^2 statistic *Yalancı-R^2 istatistiği*
1	0.023	99.870
2	0.007	99.990
3	0.003	100.000
…	…	…

Considering the stress values given in Table 4, it can be seen that the accordance between the observed distances and the configuration distances is "perfect" even for one dimension. Because the stress value calculated for the 1st dimension is below 2.5% (Table 2). Also, Pseudo-R^2 statistic, calculated as 99.87%, expresses the power of the model created for metric-MDS analysis to explain the data. Both the stress value and the pseudo-R^2 statistics show that only one dimension (the first dimension) is sufficient to classify the pistachios in the study in terms of the fatty acids studied. In other words, it is concluded that 99.87% of the variation in the observed distances can be explained by the configuration distances calculated using the first dimension obtained as a result of the metric-MDS analysis.

The relationship between the observed and the configuration distances calculated for the first dimension can be shown with the Shepard diagram as in Figure 1. When Figure 1 is evaluated, the linear relationship between the observed distances and the configuration distances can easily be seen.

According to the stress value obtained as a result of the metric-MDS analysis, the number of sufficient dimensions was determined as 1. The classification map created for one dimension as a result of metric-MDS analysis using simulated data for various fatty acids for *'Siirt'*, *'Uzun'*, *'Halebi'*, *'Kırmızı'* and *'Ohadi'* cultivars is given in Figure 2.

Especially *Ohadi* (41-50), *Uzun* (11-20) and cultivars differed from the others and from each other in terms of fatty acids studied (Figure 2). *Halebi* (21-30) and *Kırmızı* (31-40) varieties are located close to each other on the map. Although the *Siirt* (1-10) cultivar is located relatively close to the *Kırmızı*, it is clustered separately from all other cultivars on the map. However, upon careful examination, it can be seen that these two cultivars do not mix completely and that the *Kırmızı* variety is positioned slightly lower than the *Halebi*. When the Figure 2 is considered in general, it is seen that pistachios are ranked from the lowest to the highest in terms of the fatty acid contents as; *Ohadi-Uzun-Siirt-Kırmızı-Halebi*. In summary, as a result of the MDS analysis, it can be said that the richest pistachio cultivars in terms of 13 fatty acids studied are Kırmızı and Halabi.

The investigator may wonder whether the observed differences between the mean vectors of cultivars for the 13 fatty acids studied are statistically significant or not. For this purpose, multivariate analysis of variance (MANOVA) technique is used. As a result of the MANOVA test to determine whether there is a statistically significant difference between the mean vectors of pistachio cultivars in terms of fatty acids studied; Pillai's trace test statistic and F-value were respectively calculated as 3.747 and 41.024, and the control hypothesis, which indicate there is not statistically significant difference between mean vectors, was rejected (p<0.01). Afterward, Mahalanobis distance was calculated in order to determine which varieties' mean vector statistically significantly differentiated from the others. The results of multiple comparison were given in Table 5.

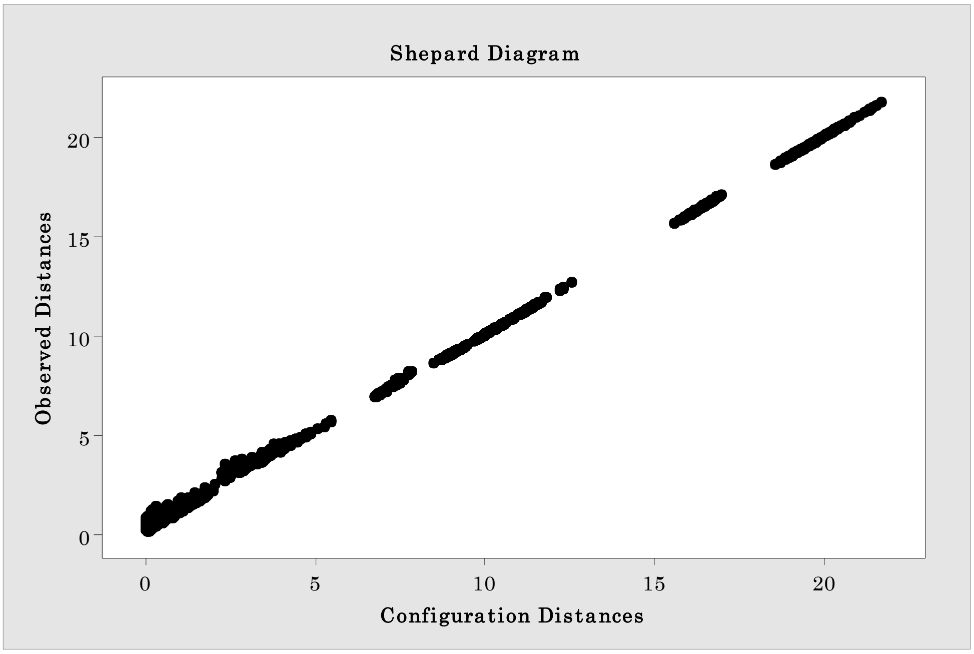

Figure 1. Shepard diagram
Şekil 1. Shepard diyagramı

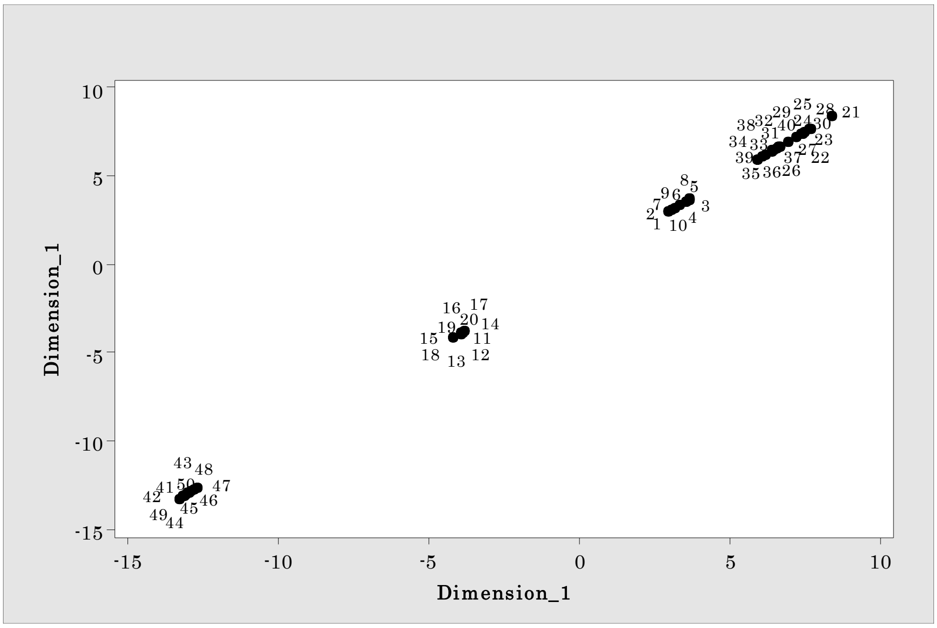

Figure 2. MDS map created to classify pistachios (Codes of pistachios according to varieties: *Siirt*; 1-10 coded, *Uzun*; 11-20 coded, *Halebi*; 21-30 coded, *Kırmızı*; 31-40 coded, *Ohadi*; 41-50 coded)
Şekil 2. Antepfistıklarının sınıflandırılması amacıyla oluşturulan MDS haritası (Antepfistığının çeşitlerine göre kodları: Siirt; 1-10, Uzun; 11-20, Halebi; 21-30, Kırmızı; 31-40, Ohadi; 41-50)

When Table 5 is examined, only the differences between the mean vectors of *Halebi –Kırmızı* and *Siirt –Kırmızı* cultivars were not found to be statistically significant. All other differences were statistically significant (p<0.05). Therefore, it can be considered that the *Kırmızı* variety of pistachios is a transitional form between *Siirt* and *Halabi* varieties in terms of 13 fatty acids studied. To sum up, it is seen that the pistachio cultivars with statistically higher values than the others in terms of 13 fatty acids studied are Kırmızı and Halabi (Table 5). This inference are also consistent with the MDS analysis.

Table 5. Multiple comparison test results using Mahalanobis distance
Çizelge 5. Mahalanobis uzaklığı kullanılarak yapılan çoklu karşılaştırma testi sonuçları

Varieties *Çeşit*	Results *Sonuçlar*
Siirt (1-10)	B
Uzun (11-20)	C
Halebi (21-30)	A
Kırmızı (31-40)	AB
Ohadi (41-50)	D

* The difference between the mean vectors of cultivars that do not have a common letter is statistically significant ($p<0.05$).

CONCLUSION

The classification results of metric-MDS analysis and the results of MANOVA and its multiple comparison test were evaluated. After the evaluation, although the results of the two tests were not exactly the same, it was observed that they were quite similar. Despite of the MDS analysis has located the Siirt variety slightly apart on the map, it is noteworthy that the results obtained with the Mahalanobis distance almost completely overlap with the MDS analysis. Nevertheless, it is much easier for the researcher to both perform the MDS analysis and interpret the results of the graphical representation compared to the MANOVA analysis. In addition, statistical package programs unfortunately do not include multiple comparison tests to determine which groups' mean vectors are statistically significant if the H0 hypothesis is rejected after the MANOVA test. Determining these by the researcher is quite time consuming and complex. MDS analysis and MANOVA post-hoc test results can be evaluated as supportive and alternative to each other (Kızıl and Aydoğan 2014). Because of these advantages, MDS analysis can be considered as an alternative to the MANOVA test and its multiple comparison tests performed afterwards. In addition, if it was wondered which varieties were different from each other in terms of the data obtained in the nominal or ordinal scale, not the metric measurements, the MANOVA test could not be used because the assumptions were not met. In such a case, varieties can be easily classified by MDS analysis, which does not require any assumptions. For this reason, MDS analysis is an advantageous method in that it provides the researcher with easily interpretable preliminary information about the differences between the means by classifying similar groups (Zech et al 2011).

ACKNOWLEDGEMENT

This article was produced from the first author's PhD thesis.

Author's Contributions

The contribution of the authors are equal.

Statement of Conflict of Interest

All the authors declare that they have no conflict of interest.

REFERENCES

Al-Abdullatif F 2020. Discriminant Function Analysis Versus Univariate ANOVA as Post Hoc Procedures Following Significant MANOVA Test: A Monte Carlo Study. Ohio University, 289 pp.

Al-Abdullatif F, Al-Abdullatif M, Brooks G 2019. MANOVA Post Hoc Techniques Used in Published Articles: A Systematic Review. *General Linear Model Journal 45*(1), 4-11

Alpar R 2013. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 4. Baskı. Detay Anatolia Akademik Yayıncılık, Ankara, 886 sy.

Borg I, Groenen PJF 2005. Modern Multidimensional Scaling: Theory and Application. 2nd Ed., Springer-Verlag, New York, 614 pp.

Borg I, Groenen PJF, Mair P 2013. Applied Multidimensional Scaling. Springer-Verlag Berlin Heidelberg, 113 pp.

Buja A, Swayne DF, Littman ML, Dean N, Hofmann H, Chen L 2008. Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics, 17*(2), 444-472.

Can H, Ozyigit II, Can M, Hocaoglu Ozyigit A, Yalcin IE 2021. Multidimensional Scaling of the Mineral Nutrient Status and Health Risk Assessment of Commonly Consumed Fruity Vegetables Marketed in Kyrgyzstan. *Biol Trace Elem Research* 1-15.

Cox TF, Cox MAA 2001. Multidimensional Scaling. 2nd Ed., Chapman and Hall/CRC, 328 pp.

Çınar B 2012. Türk Antepfıstığı Çeşitlerinin Vitamin, Mineral Madde, Yağ ve Yağ Asitleri Bileşimi Üzerinde Araştırmalar. Ankara Üniversitesi Fen Bilimleri Enstitüsü Bahçe Bitkileri Anabilim Dalı, Yüksek Lisans Tezi, 184 sy.

Dumanoglu H, Aygun A, Delialioglu RA, Erdogan V, Serdar U, Kalkisim O, Bastas K, Kocabas Z 2018. Analyses of Fruit Attributes by Multidimensional Scaling Method of Apple Genetic Resources from Coastal Zone of North Eastern Anatolia. *Scientia Horticulturae 240*, 147-154.

Gevrekçi Y, Ataç FE, Takma Ç, Akbaş Y, Taşkın T 2011. Koyunculuk Açısından Batı Anadolu İllerinin Sınıflandırılması. *Kafkas Universitesi Veteriner Fakültesi Dergisi, 17*(5), 755-760.

Gündüz S 2011. Uzaklık Fonksiyonlarının Çok Boyutlu Ölçekleme Algoritmalarındaki Etkinliğinin İncelenmesi ve Uygulamalar. Çukurova University, Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 124 sy.

Jaworska N, Anastasova AC 2009. A Review of Multidimensional Scaling (MDS) and Its Utility in Various Psychological Domains. *Tutorials in Quantitative Methods for Psychology 5*(1),1-10.

Jobson JD 1992. Applied multivariate data analysis: Volume II: Categorical and Multivariate Methods. Springer-Verlag New York, Inc. pp.731.

Kızıl SH, Aydoğan M 2014. Evaluation of major cattle breeds in Turkey for slaughter and carcass traits using MANOVA and multidimensional scaling technique. *Erciyes University Faculty of Veterinary Medicine 11*(1),15-22.

Lopes AM, Machado JT, Ramalho E 2017. On the fractional-order modeling of wine. *European Food Research and Technology 243*(6),921-929.

Mair P, Borg I, Rusch T 2016. Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate behavioral research, 51*(6),772-789.

Messick SJ, Abelson RP 1956. The Additive Constant Problem in Multidimensional Scaling. *Psychometrika 21*(1), 1-15.

Özdamar K 2004. Paket Programlar ile İstatistiksel Veri Analizi (Çok değişkenli Analizler). Kaan Kitabevi, Eskişehir, 528 sy.

Shepard RG 1962. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. *Psychometrika 27*, 125-140.

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 İncelenmesi. *İnönü Üniversitesi Tıp Fakültesi Dergisi 13*(2),81-85.

Suarez MH, Hernandez AIM, Galdon BR, Rodriguez LH, Cabrera CEM, Mesa DR, Rodriguez EMR, Romero CD 2016. Application of Multidimensional Scaling Technique to Differentiate Sweet Potato (Ipomoea batatas (L.) Lam) Cultivars According to Their Chemical Composition. *Journal of Food Composition and Analysis 46*, 43-49.

Tatlıdil H 2002. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Ziraat Matbaacılık, Ankara 424 sy.

Torgerson WS 1952. Multidimensional Scaling: I. Theory and Method. *Psychometrika 17*(17), 401-419.

Yamamoto K, Ninomiya S, Kimura Y, Hashimoto A, Yoshioka Y, Kameoka T 2015. Strawberry cultivar identification and quality evaluation on the basis of multiple fruit appearance features. *Computers and Electronics in Agriculture 110*, 233-240.

Yiğit E 2007. Çok Boyutlu Ölçekleme Yöntemlerinin İncelenmesi ve Bir Uygulama. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 60 sy.

Yiğit S, Mendeş M 2016. Usage of Multidimensional Scaling Technique for Evaluating Performance of Multivariate Normality Tests. *British Journal of Applied Science and Technology 16*(1),1-8.

Young FW, Hamer RM 1987. Multidimensional Scaling: History, Theory, and Applications. Lawrence Erlbaum Associates, Hillsdale 336 pp.

Young G, Householder AS 1938. Discussion of a Set of Points in terms of Their Mutual Distances. *Psychometrika 3*, 19-22.

Zech H, Echtermeyer C, Wöhlbrand L, Blasius B, Rabus R 2011. Biological versus technical variability in 2-D DIGE experiments with environmental bacteria. *Proteomics, 11*(16),3380-3389.