

Doğrusal Olmayan Temel Bileşenler Analizinin Tanıtımı ve Uygulanabilirliği

Yıldırım DEMİR^{1*}, Sıddık KESKİN², Şeyda ÇAVUŞOĞLU³

¹Van Yüzüncü Yıl Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Van, ²Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü, Van, ³Van Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü, Van

¹<https://orcid.org/0000-0002-6350-8122>, ²<https://orcid.org/0000-0001-9355-6558>, ³<https://orcid.org/0000-0001-8797-6687>

✉: ydemir@yyu.edu.tr

ÖZET

Doğrusal olmayan temel bileşenler analizi (NLPCA), aralarında doğrusal ya da doğrusal olmayan ilişki bulunan çok değişkenli veri setlerinde, değişkenler arasındaki ilişkileri inceleyen ve sonuçları sayısal ve görsel olarak gösteren, açıklayıcı bir boyut indirgeme yöntemidir. Bu çalışmada, doğrusal olmayan temel bileşenler analizi (NLPCA) hakkında temel açıklayıcı bilgilerin sunulması ve uygulaması yapılarak kullanılabilirliğinin vurgulanması amaçlanmıştır. Çalışmada 3 biber çeşidine ait 17 sürekli değişken için 270 örnekten elde edilen veriler, Temel bileşenler analizi (PCA) ile değerlendirilmiştir. PCA sonucunda elde edilen 4 temel bileşen ile çeşit, depolama süresi ve uygulama olmak üzere 3 kategorik değişken, NLPCA ile analiz edilmiştir. PCA ile yapılan analizde toplam varyansın yaklaşık %74'lük kısmı, NLPCA ile yapılan analizde ise yaklaşık % 58'lik kısmı açıklanmıştır. Analiz sonucunda; PC1 ile depolama süresi ve çeşit ile PC3 ve PC2 değişkenleri arasında güçlü bir ilişki olduğu gözlenirken, PC4 ve uygulama değişkenleri ile tüm değişkenler arasındaki ilişkilerinin düşük olduğu gözlenmiştir. Sonuç olarak; çok değişkenli veri setlerinde yer alan değişkenler arasındaki doğrusal ve doğrusal olmayan ilişkilerin incelenerek, bu ilişkilerin kolay yorumlanabilir ve basit anlaşılabilir şekilde iki boyutlu uzayda sunulabilmesi açısından; NLPCA'nın tek başına ve/veya diğer çok değişkenli analiz yöntemleri ile birlikte kullanılabilmesi vurgulanmıştır.

Araştırma Makalesi

Makale Tarihi

Geliş Tarihi : 17.07.2020

Kabul Tarihi : 27.08.2020

Anahtar Kelimeler

Boyut indirgeme
Optimal ölçekleme
Özdeğer
SOD
MDA

Introduction and Applicability of Nonlinear Principal Components Analysis

ABSTRACT

Nonlinear principal component analysis (NLPCA) is a descriptive dimension reduction method that examines the relationships between variables and displays the results numerically and visually in multivariate datasets that have a linear or nonlinear relationship between them. In this study, it was aimed to present the basic explanatory information about nonlinear principal components analysis (NLPCA) and to emphasize its usability by performing application. In the study, data obtained from 270 samples for 17 continuous variables concerning 3 pepper varieties were evaluated by Principal components analysis (PCA). With the 4 principal components obtained as a result of PCA, being 3 categorical variables Variety, storage time and Application were analyzed by NLPCA. In the analysis made with PCA, approximately 74% of the total variance was explained and in the analysis made with NLPCA, approximately 58% was explained as well. As a result of the analysis; it was observed that there was a strong relationship between PC1 and storage time and variety, and PC3 and PC2 variables, while the relationship between PC4 and application variables and all variables was low. As a result; by examining the linear and nonlinear relationships between the variables in the multivariate datasets, these relationships intended to be presented in an easily interpreted and easily understandable way in two-dimensional space; it was emphasized that NLPCA can be used alone and/or together with other multivariate analysis methods.

Research Article

Article History

Received : 17.07.2020

Accepted : 27.07.2020

Keywords

Dimension reduction
Optimal scaling
Eigenvalue
SOD
MDA

Atf İçin: Demir Y, Keskin S, Çavuşoğlu S 2021. Doğrusal Olmayan Temel Bileşenler Analizinin Tanıtımı ve Uygulanabilirliği. KSÜ Tarım ve Doğa Derg 24 (2): 442-450. <https://doi.org/10.18016/ksutarimdog.vi.770817>.
To Cite: Demir Y, Keskin S, Çavuşoğlu S 2021. Introduction and Applicability of Nonlinear Principal Components Analysis. KSU J. Agric Nat 24 (2): 442-450. <https://doi.org/10.18016/ksutarimdog.vi.770817>.

GİRİŞ

Doğadaki olayları incelemek ve mevcut veya olası sorunları çözebilmek amacıyla geliştirilen bilimlerden biri de veri analizi bilimidir. Veri analizi biliminde amaç, veri yapısına uygun yöntem ve teoriler kullanılarak sınırlı sayıda gözlem ile üzerinde durulan konuyu belirli bir olasılıkla açıklayabilmek ve gelecekteki araştırmalara ışık tutabilmektir (Keskin, 2002). Doğadaki olaylar, birbiri ile ilişkili olan çok sayıda değişkenin ya da faktörün ayrı ayrı ya da birlikte etkileşimleri sonucu ortaya çıkan karmaşık yapıdaki durumlardır. Bu karmaşık yapıda, ilgili olaya etki eden değişkenlerin belirlenmesi, bunların birbirleriyle olan ilişkilerinin çözülmesi ve oluşum mekanizmalarının basit ve anlaşılır bir şekilde açıklanması kolay olmayabilir. Zira belirli bir zamanda ve belirli sayıda veri ile bu yapı içindeki olayın, doğru ve güvenilir bir şekilde ortaya konulabilmesi, ancak doğru istatistik yöntemlerin kullanılması ile mümkündür (Demir, 2010).

İstatistik veri analizinde kullanılan değişkenler arasındaki ilişkiler her zaman doğrusal olmayacağı gibi cevap değişkeni sayısı tek de olmayabilir. Böyle durumlarda değişkenler arasındaki ilişkilerin doğrusal olmayan çok değişkenli istatistik yöntemlerle belirlenmesi daha makuldür. Zira değişkenler arasındaki doğrusal ilişkilerin belirlenmesinde, doğrusal yöntemlerin kullanılabilmesi için bunlarla ilgili bazı varsayımların sağlanması gerekmektedir. Söz konusu varsayımlar sağlanmadığında veya veri yapısına uygun olmayan analiz yöntemleri kullanıldığında, elde edilen sonuçlar ve yorumlamalar yanlış olabilmektedir.

Değişkenler arasındaki ilişkiyi belirlemede kullanılan çok değişkenli istatistik analiz yöntemlerinden birisi de temel bileşenler analizidir (PCA: Principal Component Analysis). Bu analizde değişkenler arasındaki ilişkilerin doğrusal ve değişkenlerin nümerik (aralık veya oran) ölçekte olması gerekmektedir. Bununla birlikte, genellikle araştırmalarda nümerik değişkenlerin yanı sıra, sıralı, kategorik veya kesikli değişkenler de kullanılabilir. Bu iki varsayım sağlanmadığı takdirde PCA hatalı sonuçlar verebilmektedir. Ancak, başta sosyal bilimlerde olmak üzere birçok alanda yapılan çalışmalarda PCA'nın varsayımları sağlanmadığında, alternatif bir yöntem olarak doğrusal olmayan temel bileşenler analizi (NLPCA: Nonlinear principal component analysis) kullanılabilir (Linting ve ark., 2007; Kapucu, 2016).

NLPCA, aralarında doğrusal ya da doğrusal olmayan

ilişki bulunan çok değişkenli veri setlerindeki değişkenler arasındaki ilişkilerin yönünü ve derecesini belirleyen, sonuçları sayısal ve görsel olarak gösteren, açıklayıcı bir boyut indirgeme yöntemidir. NLPCA'da amaç, PCA'da olduğu gibi p boyutlu bir veri setinin k ($k < p$) boyuta indirgenerek değişkenler arasındaki ilişkinin belirlenmesidir (Kramer, 1991; Kapucu, 2016).

Bu bağlamda çalışmada, NLPCA hakkında temel açıklayıcı bilgilerin sunulması ve bir uygulama yapılarak, başta ziraat olmak üzere fen, sağlık ve sosyal bilimlerde birçok alanda kullanılabilirliğinin vurgulanması amaçlanmıştır.

MATERYAL ve METOD

Çalışmada, Akra F1, Melek F1 ve Gökkuşuğu F1 olmak üzere üç biber çeşidinden; Ağırlık kaybı (g), L, a, b, c Hue, Toplam Fenolik, Toplam Antioksidan, O₂, CO₂, N₂, Solunum, Etilen, SOD (Süperoksit dismutaz), CAT (Katalaz), APX (Askorbat peroksidaz) ve MDA (Malondialdehit) olmak üzere 17 adet sürekli değişkene ait elde edilen veriler kullanılmıştır.

Elde edilen toplam 270 adet veri, PCA ile analiz edilmiş ve özdeğerler dikkate alınarak ilk dört temel bileşene ait skor değerleri elde edilmiştir. Daha sonra çeşit, depolama süresi, uygulama ve PCA ile elde edilen ilk dört temel bileşen NLPCA ile analiz edilmiş, elde edilen sonuçlar sayısal ve grafiksel olarak gösterilerek yorumlanmıştır.

İstatistik Analizler

PCA'nın amacı, k sayıda orijinal değişkenin (X_1, X_2, \dots, X_k) doğrusal bileşenlerini elde ederek, temel bileşen olarak adlandırılan daha az sayıda (p) yeni değişkene dönüştürmektir. Böylece PCA çok değişkenli veri setlerinde, etkili bir boyut indirgeme yöntemi olarak kullanılmaktadır. PCA'da maksimum elde edilebilecek temel bileşen sayısı, orijinal değişken sayısına eşittir. Ancak analiz sonuçları, daha az sayıda ve genellikle de ilk iki veya ilk üç temel bileşen dikkate alınarak yorumlanmaktadır. Öyle ki, çoğu zaman, orijinal değişkenlerdeki varyasyonun büyük bir kısmı ilk iki temel bileşenle açıklanabilmektedir. Böylece orijinal değişkenlerin, önemli bir oranda özetlenmesi ve daha kolay yorumlanabilmesi sağlanmaktadır (Jolliffe, 1986; Demir ve ark., 2016). PCA'da uygun temel bileşen sayısını belirlemede; 1'den büyük özdeğere sahip bileşenler ya da toplam varyansın en az %67'sini açıklayabilen temel bileşenler dikkate alınabilir (Özdamar, 2010).

Kategorik iki değişken arasındaki ilişkiyi incelemek amacıyla Ki-kare veya basit uyum analizi

kullanılmaktadır. Ancak değişken sayısı ikiden fazla olduğu durumlarda optimal ölçekleme yöntemlerine başvurulması önerilmektedir (Güç, 2015). Optimal ölçeklemenin temelini ALS (Dalgalı En Küçük Kareler) algoritması oluşturmaktadır.

Albert Gifi (1990) tarafından, optimal ölçeklemeye dayalı doğrusal olmayan çok değişkenli istatistik analiz yöntemleri ayrıntılı bir şekilde ele alınmıştır. Bu nedenle bu yöntemler Gifi yöntemleri olarak da adlandırılmakta ve bu yöntemlerin temelini kategorik verilere dayanan gösterge (G) matrisi oluşturmaktadır. Gösterge matrisi ile kategorik değişkenler sayısallaştırıldıktan sonra bilinmeyen nesne ve kategori skorlarını içeren kayıp fonksiyonu oluşturulmaktadır. Kayıp fonksiyonu, en küçük değerini alana kadar iterasyon sürdürülmekte ve bu işlem optimal ölçekleme süreci olarak adlandırılmaktadır. Gifi yöntemleri temel olarak, değişkenler çoklu olarak ele alındığında oluşabilecek bilgi kaybını (açıklanmayan varyansı veya kayıp fonksiyonunu) minimize etmeyi amaçlamaktadır (Mair ve De Leeuw, 2010).

Doğrusal olmayan çok değişkenli analiz yöntemlerinde kategorik değişkenler çeşitli dönüşümlerle sayısallaştırılmakta ve kayıp fonksiyonu ile çok değişkenli analiz yöntemlerinin uygulanmasına imkân sağlanmaktadır. Böylece minimum kayıpla gerçek uzaydaki ilişkiler daha düşük boyutlu bir uzayda gösterilmektedir (Karaman, 2019).

Çok sayıda değişken arasındaki ilişkinin etkili bir şekilde yorumlanması çoğu kez zordur. Bu durumda orijinal değişkenler, orijinal değişkenlerde bulunan bilgilerin çoğunu temsil eden daha küçük ilişkisiz bileşenler kümesine indirgenmektedir. Böylece, çok sayıda değişken yerine birkaç bileşenle sonuçlar yorumlanmaktadır. Bu bağlamda, kategorik temel bileşenler analizi (CATPCA) olarak da bilinen NLPCA, boyut indirgemekte ve kategorik değişkenleri sayısallaştırmaktadır (quantifies).

NLPCA, geometrik olarak ele alındığında gerçek uzay yerine daha düşük boyutlu bir uzayda değişken ve kategorileri grafiksel olarak göstermeyi amaçlamaktadır. Diğer bir ifade ile NLPCA, sayısal dönüşümler yaparak X 'in nesne skorlarını bulmayı ve Y_j 'nin bir dizisini çeşitli yollarla kısıtlayarak minimize etmeyi amaçlamaktadır (Michailidis ve De Leeuw, 1998; Karaman, 2019). Bu amaç doğrultusunda $j \in J$ ise $c=p$ ile ve $j \notin J$ ise $c=1$ ile Eşitlik 1 minimize edilmektedir (Demir, 2010).

$$\sigma(X; Y) = n_w^{-1} \sum_j c^{-1} \text{tr} \left((X - G_j Y_j)' M_j W (X - G_j Y_j) \right) \quad (1)$$

$$j = 1, \dots, m$$

Kayıp fonksiyonu olarak bilinen Eşitlik 1'de, X , $n \times p$ boyutlu nesne skorları;

Y , değişkenlerin çoklu ölçekleme seviyesindeki

sentroid koordinatları (Y_j) ve çoklu nominal olmayan ölçekleme seviyesindeki vektör koordinatları (y_j, a_j) toplamı;

n_w , $\sum_{i=1}^n w_i$ eşitliği ile ağırlıklandırılmış gözlem sayısı; G_j , elemanları $i = 1, \dots, n$ ve $r = 1, \dots, k_j$ olan $n \times k_j$ boyutlu j . değişkeninin ikili gösterge matrisi (i . gözlem j . değişkenin r . kategorisinde ise $g_{(j)ir} = 1$; i . gözlem, j . değişkenin r . kategorisinde değilse $g_{(j)ir} = 0$ olur.) (Gifi, 1990; Güç, 2015);

Y_j , $k_j \times p$ boyutlu sentroid koordinatları;

M_j , diyagonal elemanları $m_{(j)ii}$ olan $n \times n$ diyagonal matris (i . gözlem eksik ve j . değişken pasif olduğu zaman veya i . gözlem, j . değişkenin r . kategorisinde olduğu ve r . kategori sadece tamamlayıcı gözlemler tarafından kullanıldığı zaman $m_{(j)ii} = 0$ ve diğer durumlarda ise $m_{(j)ii} = v_j$ olur.);

W , diyagonal elemanları w_j olan $n \times n$ boyutlu diyagonal matrisi;

m , analiz edilen değişken sayısı ve

J , hangi değişkenlerin çoklu nominal ölçeklendirme seviyesine sahip olduğunu belirleyen indeks değeridir.

Ayrıca, Eşitlik 1'in çözümü için gerekli olan diğer notasyonlar ise;

n , toplam analiz edilen tamamlayıcı gözlem sayısı; p , boyut sayısı; y_j , k_j boyutlu çoklu nominal olmayan ölçekleme seviyesindeki değişkenlere ait kategori sayısı; w_i , i . gözlemin ağırlığı (i . gözlem ağırlıklandırılmamışsa $w_i = 1$ ve i . gözlem tamamlayıcı gözlemse $w_i = 0$ olur.) ve k_j , j . değişkenin kategorilerinin sayısı şeklinde açıklanabilir.

Çoklu nominal değişkenlerde bir boyut için açıklanan varyans Eşitlik 2 ile hesaplanmaktadır.

$$VAF1_s = n_w^{-1} \sum_{j \in J} v_j \text{tr}(Y_j' D_j Y_j) \quad s = 1, \dots, p \quad (2)$$

Eşitlik 2'de v_j , değişken ağırlığıdır (j . değişkenin ağırlığı belirlenmemişse $v_j = 1$ olur). D_j , $k_j \times k_j$ boyutlu diyagonal matrisi olmak üzere ağırlıklandırılmış tek değişkenli marjinal değerler içerdiğinden G_j 'nin ağırlıklandırılmış sütun toplamları Eşitlik 3 ile hesaplanmaktadır.

$$D_j = G_j' W G_j \quad (3)$$

Çoklu nominal olmayan değişkenlerde bir boyut için açıklanan varyans ise Eşitlik 4 ile hesaplanmaktadır.

$$VAF2_s = \sum_{j \in J} v_j a_{js}^2 \quad s = 1, \dots, p \quad (4)$$

Eşitlik 4'te a_j , p . boyutta çoklu nominal olmayan ölçekleme seviyesindeki değişkenlere ait bileşen yükleridir (Demir, 2010).

Boyutlara ait özdeğerler, toplam özdeğer olarak ifade edilen toplam açıklanan varyans ve NLPCA için vektör koordinatları sırasıyla Eşitlik 5, 6, ve 7 ile

hesaplanmaktadır.

$$\sqrt{\lambda_s} = VAF1_s + VAF2_s \quad s = 1, \dots, p \quad (5)$$

Eşitlik 5'te λ_s , Λ 'nin s. diyagonal elemanıdır.

$$tr(\sqrt{\Lambda}) = p^{-1} \sum_s VAF1_s + \sum_s VAF2_s \quad s = 1, \dots, p \quad (6)$$

$$VAF_{js} = v_j a_{js}^2 \quad s = 1, \dots, p \quad \text{ve} \quad j \notin J \quad (7)$$

Optimal ölçekleme sonrası elde edilen değişkenler için korelasyon matrisi Eşitlik 8 ile hesaplanmaktadır.

$$R = n_w^{-1} Q'WQ \quad (8)$$

Analizde çoklu nominal olmayan değişkenler varsa, kayıp gözlem yoksa ya da pasif olarak belirlenmişse Eşitlik 8 kullanılmakta ve $q_j = G_j y_j$ ile korelasyon matrisi hesaplanmaktadır. R'nin ilk p özdeğeri $\sqrt{\Lambda}$ e eşittir.

Analizde çoklu nominal değişkenler varsa p korelasyon matrisleri,

$$R_s = n_w^{-1} Q'_s W Q_s \quad s = 1, \dots, p \quad (9)$$

Eşitlik 9 ile hesaplanmaktadır. Eşitlik 9'da q_{js} , çoklu nominal olmayan değişkenler için $G_j y_j$ olarak ve çoklu nominal değişkenler için ise $\frac{G_j Y_{js}}{\sqrt{Y'_{js} D_j Y_{js}}}$ olarak

hesaplanmaktadır.

R_s matrisinin 1'nci özdeğeri genellikle daha yüksek olup $\sqrt{\lambda_s}$ 'e eşittir. $\sqrt{\Lambda}$ 'nin daha düşük değerleri genellikle R_s 'nin 2'nci ve daha sonraki özdeğerlerine aittir. Özdeğerlerin hesaplanmasında; R matrisinin tekil değer ayrışımı için j değişkeni tamamlayıcı değişkense, önce R matrisinden ilk sütun ve j. satır çıkarılır, sonra R_{ij} , $\sqrt{v_i v_j}$ ile çarpılır (IBM SPSS, 2011).

BULGULAR ve TARTIŞMA

Çalışmada, ele alınan değişkenlere ait Pearson korelasyon katsayıları Çizelge 1'de verilmiştir.

Çizelge 1. Orijinal değişkenler arası korelasyonlar
Table 1. Correlations among the original variables

	A. Kaybı	L	a	b	c	Hue	To. Fe	To. An	O ₂	CO ₂	N ₂	Solunum	Etilen	SOD	CAT	APX
L	-0.20															
a	0.11	0.38														
b	0.05	0.58*	-0.15													
c	0.02	0.42	-0.31	0.97**												
Hue	-0.14	-0.71*	-0.75*	-0.52*	-0.37											
To. Fe	0.69*	0.09	0.28	0.10	0.03	-0.31										
To. An	0.55*	0.01	0.25	0.08	0.04	-0.27	0.90**									
O ₂	-0.31	0.09	-0.28	0.38	0.42	0.00	-0.20	-0.08								
CO ₂	0.31	-0.14	0.27	-0.40	-0.43	0.02	0.13	0.01	-0.95**							
N ₂	0.21	-0.08	0.25	-0.33	-0.38	0.00	0.14	0.03	-0.86*	0.79*						
Solunum	0.34	-0.36	-0.48	0.27	0.36	0.23	-0.08	-0.06	0.05	-0.01	-0.10					
Etilen	0.18	-0.14	-0.25	0.20	0.25	0.07	-0.06	-0.07	-0.08	0.10	0.02	0.55*				
SOD	0.40	0.05	0.31	-0.17	-0.23	-0.16	0.41	0.25	-0.48	0.41	0.42	-0.11	-0.05			
CAT	0.12	0.05	0.10	-0.01	-0.03	-0.09	0.07	0.03	-0.27	0.28	0.23	0.06	0.07	0.16		
APX	0.01	0.07	0.37	-0.35	-0.41	-0.10	-0.07	-0.22	-0.47	0.46	0.43	-0.25	-0.09	0.42	0.21	
MDA	0.70*	-0.45	-0.08	-0.13	-0.11	0.15	0.37	0.26	-0.28	0.32	0.18	0.37	0.22	0.24	0.14	0.03

*(p<0.05), **(p<0.01); A. Kaybı: Ağırlık Kaybı, To. Fe: Toplam Fenolik, To. An: Toplam Antioksidan, SOD: Süperoksit dismutaz, CAT: Katalaz, APX: Askorbat peroksidaz, MDA: Malondialdehit,

Çizelge 1 incelendiğinde, ağırlık kaybı ile toplam fenolik, toplam antioksidan ve MDA arasında istatistik olarak önemli (p<0.05) pozitif ilişkinin olduğu görülmektedir. Ayrıca, b ile L ve b ile c arasında pozitif; Hue ile L, a ve b arasında negatif; toplam fenolik ile toplam antioksidan arasında pozitif; O₂ ile CO₂ ve N₂ arasında negatif; CO₂ ile N₂ arasında ve solunum ile etilen arasında istatistik olarak önemli pozitif korelasyonlar bulunmuştur. Buna karşılık, SOD, CAT ve APX değişkenlerinin kendi aralarında ve diğer değişkenlerle olan korelasyonları önemli

bulunmamıştır.

Uygulanan temel bileşenler analizi sonucunda, ilk dört temel bileşene ait sonuçlar Çizelge 2'de verilmiştir. Çizelge 2' de görüldüğü üzere; toplam varyansın %27.024'lük kısmı ilk temel bileşenle, %45.879'lük kısmı ilk iki temel bileşenle, %63.129'lük kısmı ilk üç temel bileşenle ve %74.071'lik kısmı ise ilk dört temel bileşenle açıklanmıştır. Böylece onyedinci orijinal değişken, toplam varyansta meydana gelen yaklaşık % 25'lik kayıpla dört (temel) bileşene indirgenmiştir. Birikimli varyans oranı %67'den büyük olduğu

durumda orijinal veri setini açıklamak amacıyla p değişken yerine ilk k temel bileşen kullanılabilir (Özdamar, 2010). Bu nedenle çalışmanın sonraki

aşamalarında ilk dört temel bileşen (%74.071) alınarak işlemler yürütülmüştür.

Çizelge 2. İlk dört temel bileşen için analizi sonuçları

Table 2. Analysis results for the first four principal components

Bileşen (Component)	Özdeğer (Eigenvalue)	Açıklanan varyans % (% of explained variance)	Birikimli varyans % (Cumulative variance %)
1	4.594	27.024	27.024
2	3.205	18.855	45.879
3	2.932	17.250	63.129
4	1.860	10.941	74.071

Çizelge 3. İlk dört temel bileşene ait yükler

Table 3. Loadings of the first four principal components

	PC1	%	PC2	%	PC3	%	PC4	%
A. Kaybı	0.471	6.106	0.430	7.460	0.642	11.349	-0.125	2.464
L	-0.217	2.818	0.638	11.074	-0.540	9.542	0.314	6.215
a	0.450	5.841	0.484	8.411	-0.542	9.582	-0.093	1.846
b	-0.580	7.522	0.641	11.124	0.157	2.772	0.415	8.215
c	-0.642	8.332	0.514	8.917	0.260	4.603	0.403	7.974
Hue	-0.016	0.205	-0.840	14.592	0.357	6.307	-0.200	3.956
Top. Fen.	0.389	5.043	0.688	11.948	0.324	5.734	-0.415	8.218
Top. An.	0.254	3.294	0.630	10.941	0.316	5.587	-0.517	10.218
O ₂	-0.875	11.354	0.050	0.869	0.001	0.013	-0.320	6.324
CO ₂	0.848	10.999	-0.107	1.861	0.023	0.401	0.343	6.795
N ₂	0.787	10.204	-0.074	1.289	-0.071	1.257	0.316	6.257
Solunum	-0.174	2.254	-0.108	1.877	0.773	13.673	0.367	7.258
Etilen	-0.058	0.754	-0.041	0.705	0.527	9.320	0.514	10.169
SOD	0.642	8.323	0.267	4.636	0.005	0.096	0.013	0.263
CAT	0.298	3.867	0.095	1.649	0.039	0.686	0.361	7.144
APX	0.580	7.524	-0.113	1.968	-0.390	6.886	0.258	5.112
MDA	0.428	5.557	0.039	0.683	0.689	12.186	-0.080	1.580

Orijinal değişkenler ile temel bileşenler arasındaki korelasyonları ifade eden temel bileşen yükleri Çizelge 3'te verilmiştir. Çizelge 3'te; ağırlık kaybı, a, b, c, O₂, CO₂, N₂, SOD, APX ve MDA değişkenleri ile ilk temel bileşen arasında yüksek korelasyon bulunmuştur. Ancak bunlar içerisinde de özellikle O₂, CO₂ ve N₂'nin önemli olduğu gözlenmiştir. Bu değişkenlerden b, c ve O₂ ile PC1 arasında pozitif korelasyon bulunurken, kalan diğer 7 değişken ile PC1 arasında ise negatif korelasyon bulunmuştur. İkinci temel bileşen; ağırlık kaybı, L, a, b, c, Hue, toplam fenolik ve toplam antioksidan değişkenleri ile tanımlanmaktadır. Bu değişkenlerden Hue ile PC2 arasında negatif korelasyon, diğer değişkenlerle PC2 arasında ise pozitif korelasyon gözlenmiştir. Üçüncü temel bileşen ağırlık kaybı, L, a, solunum, etilen ve MDA değişkenlerinden oluşurken, dördüncü temel bileşen ise toplam antioksidan, etilen, toplam fenolik, b ve c değişkenlerinden oluşmaktadır.

On yedi orijinal değişkenden indirgenen dört temel bileşen ve üç faktör (Çeşit, depolama süresi ve uygulama) arasındaki ilişki yapısı, doğrusal olmayan temel bileşenler analizi (NLPCA) ile incelenerek, değişken kategorilerinin iki boyutlu uzaydaki konfigürasyonu Şekil 1'de gösterilmiştir.

Optimal ölçeklemede, değişkenlerin yapısı göz önüne alınmadan değişkenlerin doğrudan analiz edilmesi sağlıklı sonuçlar vermeyebilir. Bu nedenle, değişkenlerin yapıları göz önüne alınarak bazı dönüşümler yapılmalı ve değişkenler optimal ölçekleme için uygun bir şekilde dönüştürmelidir. Bu bağlamda, nümerik bir değişken kategorik bir değişkene dönüştürülerek doğrusal olmayan yöntemlerle daha küçük boyutlu çözümler elde edilmektedir (Meulman ve Heiser, 2011, Güç, 2015). PCA ile elde edilen temel bileşenlerin skor değerleri, nümerik olarak elde edilmiştir. Ancak, NLPCA için kullanılan ilk dört temel bileşenin skorları, 3 gruplu kategorik değişkene dönüştürülmüştür.

Kategorilerin boyutlara olan katkısının ve ayırma gücünün artmasına paralel olarak, boyutlara ait katsayı değerleri de artmaktadır. Diğer bir ifade ile herhangi bir kategorinin boyutlarda almış olduğu değerlerinin orijinden uzaklaşması, adı geçen kategorinin boyutu belirlemede etkisinin daha yüksek olduğunu göstermektedir. Buna göre, birinci boyutta 0.913 ile PC2'nin "0.613-3.044" kategorisi en yüksek pozitif değeri alırken, -1.119 ile depolama süresinin 0. gün kategorisi en yüksek negatif değeri almıştır. Böylece bu iki kategorinin, birinci boyuta göre birbiri

varyans açıklama oranı da buna bağlı olarak düşmektedir. Dolayısıyla, NLPCA'da kullanılan diğer temel bileşen değişkenlerine göre PC4 değişkenine ait varyans açıklama oranının düşük olması, bu

değişkenin diğer değişkenler ile olan ilişkisini de etkilemektedir. Zira PC4 değişkeni, diğer değişkenler ile düşük ilişkili bulunmuştur.

Çizelge 4. Optimal ölçekleme sonrası değişkenler arasındaki korelasyonlar
Table 4. Correlations the among variables after optimal scaling

	Çeşit (Variety)	Uygulama (Application)	Depolama süresi (Storage time)	PC1	PC2	PC3	PC4
Çeşit	1.000						
Uygulama	0.000	1.000					
Depolama Süresi	0.000	0.000	1.000				
PC1	0.085	-0.052	0.858	1.000			
PC2	0.546	0.148	0.220	0.192	1.000		
PC3	-0.736	0.060	0.221	0.036	-0.132	1.000	
PC4	-0.260	-0.104	-0.097	-0.155	-0.066	0.032	1.000
Boyut (Dimension)	1	2	3	4	5	6	7
Özdeğer (Eigenvalue)	2.121	1.911	1.112	0.941	0.676	0.148	0.091

Çizelge 4'te NLPCA ile analiz edilen değişkenler arasındaki korelasyonlar verilmiştir. Çizelge 4'te en yüksek korelasyonun 0.858 değeri ile PC1 ve depolama süresi değişkenleri arasında olduğu görülmektedir. Ayrıca çeşit ile PC3 (-0.736) ve PC2 (0.546) değişkenleri arasındaki korelasyonun da yüksek olduğu görülmektedir. Ancak çeşit, uygulama ve depolama süresi değişkenleri arasında herhangi bir korelasyon belirlenmemiştir. En yüksek üç korelasyon değeri göz önüne alındığında, depolama süresi ile PC1 ve çeşit ile PC2 değişkenleri arasındaki korelasyonlar pozitifken, çeşit ile PC3 arasındaki korelasyon negatiftir.

Özdeğerler, varyans açıklama oranının değişken sayısı ile çarpımına ($\lambda = VAF * m$) eşittir (Meulman ve ark., 2004). Buna göre birinci boyut için özdeğer (%30.303*7) 2.121, ikinci boyut için (%27.298*7) 1.911 ve her iki boyutun toplamı için de 4.032 olarak bulunmuştur.

Şekil 2'de iki boyutlu uzayda değişkenlere ait bileşen yükleri konumlarının yanı sıra vektör koordinatları, bileşen yük koordinatları ve ilk iki boyuta ait varyans açıklama oranları verilmiştir.

Şekil 2 incelendiğinde, birinci boyuta göre çeşit (0.764), PC2 (0.677), PC1 (0.639), depolama süresi (0.564) ve uygulama (0.070) değişkenleri pozitif yüklü değişkenleri, PC3 (-0.457) ve PC4 de (-0.373) negatif yüklü değişkenleri oluşturmaktadır. İkinci boyuta göre depolama süresi (0.772), PC3 (0.708), PC1 (0.672) ve PC4 (0.009) pozitif yüklü değişkenleri, çeşit (-0.588), PC2 (-0.129) ve uygulama (-0.014) negatif yüklü değişkenleri oluşturmaktadır. Ayrıca, PC1 ve depolama süresi arasında yüksek pozitif ilişki görülürken, bu iki değişkenin diğer değişkenler ile yok denecek kadar düşük ilişkili olduğu söylenebilir. Bunun yanı sıra çeşit, PC2 ve PC3 arasında da bir ilişkinin olduğu görülmektedir. Çeşit ile PC3

arasındaki ilişki, çeşit ile PC2 arasındaki ilişkiden daha güçlüdür. Ancak, çeşit ile PC3 arasında negatif ilişki mevcutken, çeşit ile PC2 arasında da pozitif ilişki mevcuttur. Kısmen de olsa PC4 ile çeşit arasında bir ilişki görülmesine rağmen, genel olarak PC4 ile uygulama değişkenlerinin diğer değişkenlerle ilişkisinin çok düşük olduğu görülmektedir.

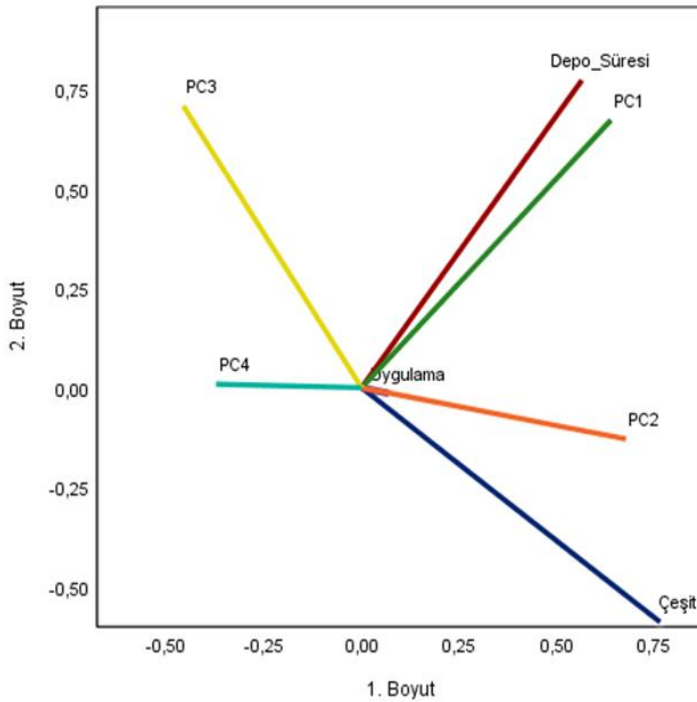
Birinci boyut tarafından toplam varyansın %30.303'ü, ikinci boyut tarafından ise toplam varyansın %27.298'i açıklanmıştır. Böylece her iki boyut tarafından toplam varyansın %57.601'i açıklanmıştır.

Böylece analize dâhil edilen yirmi değişkenden; depolama süresinin artması ile ağırlık kaybı, a, CO₂, N₂, SOD, APX ve MDA'nın artabileceği, b, c ve O₂'nin ise azalabileceği söylenebilir. Ayrıca geriye kalan 9 değişkenin depolama süresinden çok az etkilendiği söylenebilir. Diğer yandan; ağırlık kaybı, L, a, b, c, Hue, toplam fenolik, toplam antioksidan, solunum, etilen ve MDA değerlerinin de çeşide göre değiştiği ifade edilebilir.

Orijine yakın olması nedeniyle, özellikle uygulama değişkenine ait etkinin çok düşük olduğu ve diğer değişkenlerle ilişkili olmadığı gözlemlenmiştir. Buna ek olarak, NLPCA'de kullanılan diğer temel bileşen değişkenlerine göre PC4 değişkenine ait varyans açıklama oranının düşük olması, bu değişkenin diğer değişkenler ile olan ilişkisinin de düşük olması ile ilişkilendirilebilir.

Birçok çalışmada genellikle çok değişkenli veri setleri ile çalışılmakta ve mümkün oldukça çok az bilgi kaybı ile boyut indirgeme yapılmaktadır. Genel olarak, bu tür değişkenleri minimize etmede PCA'nın uygun bir yöntem olduğu ifade edilmektedir (Jolliffe, 1986).

Ancak bu analiz yönteminin uygulanabilmesi için değişkenler arasındaki ilişkinin doğrusal ve değişkenlerin nümerik ölçekte olması gibi iki önemli



Değişkenler (Variables)	Boyutlar (Dimensions)		Toplam (Total)	Boyutlar (Dimensions)	
	1	2		1	2
Çeşit (Variety)	0.584	0.346	0.930	0.764	-0.588
Uygulama (Application)	0.005	0.000	0.005	0.070	-0.014
Depolama süresi (Storage time)	0.318	0.596	0.914	0.564	0.772
PC1	0.408	0.451	0.859	0.639	0.672
PC2	0.458	0.017	0.475	0.677	-0.129
PC3	0.209	0.501	0.709	-0.457	0.708
PC4	0.139	0.000	0.139	-0.373	0.009
Özdeğer (Active Total)	2.121	1.911	4.032		
Açıklanan Varyans % (% of Variance)	30.303	27.298	57.601		

Şekil 2. İki boyutlu uzayda değişkenlere ait bileşen yükleri
Figure 2. Component loads of variables in two-dimensional space

varsayımın sağlanması gerekmektedir (Jolliffe, 1986). Birçok çalışmada bu varsayımlar sağlanamamaktadır. Bu durumda, PCA yönteminin etkinliğinde azalma olabileceğinden, alternatif bir yöntem olarak, PCA ile aynı amacı taşıyan ve PCA'nın doğrusal olmayan bir uzantısı olan NLPCA önerilebilir. NLPCA'nın temelini oluşturan optimal ölçekleme yaklaşımı, değişkenlerin farklı düzeylerde ölçeklendirilmesine izin vermektedir. Bu sayede kategorik değişkenler, istenilen boyutta en uygun şekilde ölçeklenmekte ve böylece, değişkenler arasındaki doğrusal ilişkilerin yanı sıra doğrusal olmayan ilişkiler de modellenabilmektedir (Meulman ve Heiser, 2011; Mori ve ark., 2016). Böylece, NLPCA doğrusal olmayan ilişki yapılarında bu ilişkileri de yakalayabildiğinden daha iyi sonuçlar vermektedir (Kramer, 1991). Ayrıca normallik ve doğrusallık gibi varsayımlara gerek duymaması, PCA'nın aksine nominal ve ordinal ölçekteki değişkenlerle yapılan analizlerde iyi sonuçlar vermesi, nümerik olmayan verilerin analizinde yöntemin daha az kısıtlayıcı ve dolayısıyla daha yüksek varyans açıklama oranına sahip olması NLPCA'nın avantajlarıdır. Bunların aksine, sürekli değişkenlerin kategorize edilerek analiz edilmesi, böylece belli bir miktarda bilgi kaybına yol açması ve herhangi bir istatistik önemlilik testinin olmaması bu yöntemin dezavantajları olarak görülebilir.

NLPCA, PCA ile aynı problemlere uygulanabileceği gibi birlikte de kullanılabilir. Nitekim bu çalışmada ziraat alanında kalite özelliklerinden olan 17 değişken alınarak, PCA ile boyut indirgeme yapılmış ve

yaklaşık %74 açıklanabilen varyans oranı ile 17 değişken 4 temel bileşene (değişkene) indirgenebilmiştir. Daha sonra, 4 temel bileşenin; uygulama, depolama süresi ve çeşit ile ilişkisi NLPCA ile incelenmiştir. NLPCA sonucunda da % 54 açıklanabilen varyans oranı ile 7 değişken 2 boyuta indirgenmiştir. Böylece, kategorik ve sürekli değişkenlerden oluşan toplam 20 değişken arasındaki doğrusal ve/veya doğrusal olmayan ilişkiler, iki boyutlu uzaya indirgenerek, kolay anlaşılabilir ve yorumlanabilir bir şekilde görsel olarak sunulmuştur.

SONUÇ

Sonuç olarak, farklı birçok bilim dalında doğrusal ve doğrusal olmayan ilişkileri içeren ve farklı değişken türlerinden oluşan çok değişkenli veri setleri bulunmaktadır. Bu veri setlerinde yer alan değişkenler arasındaki doğrusal ve doğrusal olmayan ilişkilerin incelenerek, bu ilişkilerin kolay yorumlanabilir ve basit anlaşılabilir şekilde iki boyutlu uzayda sunulabilmesi açısından; doğrusal olmayan bir boyut indirgeme yöntemi olan NLPCA'nın tek başına ve/veya diğer çok değişkenli analiz yöntemleri ile birlikte kullanılabilirliği önerilebilir.

Araştırmacıların Katkı Oranı Beyan Özeti

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

KAYNAKLAR

- Demir C 2010. Doğrusal Olmayan Temel Bileşenler Analizi ve Sağlık Alanında Uygulaması. Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Yüksek Lisans Tezi, 55 sy.
- Demir Y, Esenbuğa N, Bilgin ÖC 2016. İvesi Koyunlarının Et Kalitesini Değerlendirmede Temel Bileşenler Analizinin (PCA) Kullanılması. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi 20(3): 536-541.
- Gifi A 1990. Nonlinear Multivariate Analysis. John Wiley & Sons, New York, 579 sy.
- Güç K 2015. Türkiye’de Resmi Kurumlara Duyulan Güvenin Kategorik Regresyon ve Lojistik Regresyon Analizi İle İncelenmesi. Gazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 75 sy.
- IBM SPSS 2011. IBM SPSS Statistics 20 Algorithms. IBM, Inc., New York, 1069 sy.
- Jolliffe IT 1986. Principal Component Analysis. Springer-Verlag, New York, 487 sy.
- Kapucu T 2016. The Effect of Computer Assisted Instruction On Eight Grade Students’ Permutation-Combination-Probability Achievement And Attitudes Towards Computer Assisted Instruction. Middle East Technical University The Graduate School of Natural And Applied Sciences Statistics Department, Master Thesis, 161 sy.
- Karaman E 2019. Optimal Ölçekleme Teknikleri ve Bir Uygulama. İstanbul Üniversitesi Sosyal Bilimleri Enstitüsü İşletme Anabilim Dalı, Doktora Tezi, 108 sy.
- Keskin S 2002. Varyansların Homojenliğini Test Etmede Kullanılan Bazı Yöntemlerin I. Tip Hata ve Testin Gücü Bakımından İrdelenmesi. Ankara Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı, Doktora Tezi, 210 sy.
- Kramer MA 1991. Nonlinear Principal Component Analysis Using Auto-Associative Neural Networks. AIChE Journal 37(2): 233-43.
- Linting M, Meulman JJ, Groenen PJF, Van der Kooij AJ 2007. Nonlinear Principal Components Analysis: Introduction and Application. Psychological Methods 12(3): 336-358.
- Mair P, De Leeuw J 2010. A General Framework for Multivariate Analysis with Optimal Scaling: The R Package Aspect, Journal of Statistical Software 32(9): 1-23.
- Meulman JJ, Heiser WJ 2011. *IBM SPSS Categories 20*, SPSS Inc., USA, 313 sy.
- Meulman JJ, Van der Kooij AJ, Heiser WJ 2004. Principal Components Analysis with Nonlinear Optimal Scaling Transformations for Ordinal And Nominal Data. (The Sage Handbook of Quantitative Methodology for the Social Sciences, UK: Ed. Kaplan D) 49-70.
- Michailidis G, De Leeuw J 1998. The Gifi System of Descriptive Multivariate Analysis. Statistical Science 13(4): 307-336.
- Mori Y, Kuroda M, Makino N 2016. Nonlinear Principal Component Analysis and Its Applications. Springer Nature, Singapore, 80 sy.
- Özdamar K 2010. Paket Programlar ile İstatistiksel Veri Analizi-2 (Çok Değişkenli Analizler). Kaan Kitapevi, Eskişehir, 506 sy.