

Kardiyolojik Verilerin Doğrusal Olmayan Kanonik Korelasyon Analizi ile İncelenmesi

Mevlüt TÜRE* Necdet SÜT* Turhan KÜRÜM** Gültaç ÖZBAY**

ÖZET

Doğrusal Olmayan Kanonik Korelasyon Analizi (DOKKA) isimsel, sıralı, nümerik ya da bunların karışımı değişkenlerden oluşan iki ya da daha fazla değişken seti arasındaki ilişkileri ortaya koymada yararlanan bir yöntemdir.

*Bu çalışmada Trakya Üniversitesi Tıp Fakültesi Kardiyoloji kliniğinde 2000 yılına kadar koroner anjiyografisi yapılan 526 hastanın koroner arter çapları (1.set - circumflex: CX, sol ön inen damar: LAD, sağ koroner arter: RCA ve sol ana koroner arter: LMCA ($\bar{x} \pm 1*s$) alınarak k:küçük, o:orta, b:büyük), koroner arter çaplarında lezyon olup olmadığı (2.set), cinsiyet, diabetes mellitus (DM), sigara kullanımı ve vücut kitle indeksi (VKİ) (3.set) değişkenleri ele alınarak bu üç setteki değişkenler arasındaki ilişkiler DOKKA ile incelendi.*

Analiz sonucunda iki boyutlu çözümlemede birinci boyutun değişkenler arasındaki ilişkinin %54.7'sini, ikinci boyutun ise %44.3'ünü açıkladığı ve DM, RCA ve LMCA çaplarının LAD ve CX'teki lezyonlarla; sigara, cinsiyet ve RCA'daki lezyonun VKİ, CX ve LAD çapları ile korelasyonlu olduğu bulundu.

Anahtar Kelimeler: Doğrusal olmayan kanonik korelasyon analizi, optimal ölçekleme, koroner arter.

1. GİRİŞ

Doğrusal Kanonik Korelasyon Analizinin (DKKA) temeli 1936 yılında Hotelling tarafından atılmıştır. DOKKA yöntemi üzerine ise ilk kez Carroll (1968) çalışmalarda bulunmuştur. Carroll, tüm setlerin kanonik değişkenleri ve obje skorları arasında kayıp fonksiyonunu minimize ederek çözümler elde etmiştir. Gifi, Van der Burg, De Leeuw yöntemin gelişimine önemli katkılarda bulunmuşlardır.

DKKA iki yada daha fazla değişken setinin birbiriyle ilişkisini belirleyen bir yöntemdir. Bu yöntemde değişken setlerinin doğrusal bileşenleri bulunur ve bileşenler aracılığı ile hesaplanan kanonik değişkenler yardımıyla iki ya da daha fazla set arasındaki korelasyon hesaplanır (Özdamar, 1999; Tatlıdil, 1996). DKKA'nin uygulanabilmesi için 1- değişkenlerin çok değişkenli normal dağılım göstermesi, 2- veri matrislerindeki

* Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı - Edirne

** Trakya Üniversitesi Tıp Fakültesi Kardiyoloji Anabilim Dalı - Edirne

değişken sayılarının en az 20 katı kadar birimden elde edilmiş olması, 3- uç değerlerin bulunmaması, 4- veri matrisinin şişkinlik özelliği göstermemesi (veri matrisinde gereğinden fazla ve problemle ilgili olmayan değişkenlerin yer almaması) ve 5- analiz edilecek değişkenler arasında tam korelasyon ($r_{ij} = 1 ; i \neq j$) bulunmaması (veri matrisinden hesaplanan korelasyon matrisi \mathbf{R} birim değişkenlerden oluştuğunda tersi alınamayacağından bu tür verilere DKKA uygulanamaz) gerekir (Özdamar, 1999). DOKKA ise farklı ölçekli verilerden oluşan değişken setleri arasındaki ilişkileri hem doğrusal hem de eğrisel olarak uç değerlere sahip olmaması dışında herhangi bir varsayım ileri sürmeden inceleyen bir yöntemdir (Van der Burg ve De Leeuw, 1983; SPSS Categories 10.0, 1999).

DOKKA'nin amacı, daha az boyutta değişken setleri arasındaki varyasyonu maksimum yapmaktır. Maksimum boyut sayısı en az değişken içeren settteki değişken sayısı kadardır. Boyut sayısı olabildiğince düşük tutulmaya çalışılırken varyans açıklayıcılığı ise olabildiğince maksimum yapılmaya çalışılır (Van der Burg ve ark. 1994; Van der Burg ve Dijksterhuis, 1996; Van der Burg ve De Leeuw, 1983). DKKA, değişkenlerin oransal ölçekli (ölçülerek elde edilmiş: cm, gr, mm/Hg gibi) olması durumunda değişken setleri arasında doğrusal ilişkileri incelemektedir. Oysa DOKKA isimsel, sıralı, aralıklı ve oransal ölçekli değişkenler arası doğrusal yada doğrusal olmayan ilişkilerden hareketle değişken setleri arasındaki ilişkileri ortaya koymada kullanılan bir yöntemdir.

Günümüzde kategorik yapıya sahip değişkenlerin tıp alanında da kullanılmakta olduğu dikkate alındığında iki yada daha fazla değişken seti arasındaki ilişkiyi hesaplamada doğrusal yöntemler kullanılamamaktadır. Ayrıca değişken setleri arasında eğrisel ilişkiler mevcut ise bu durumda doğrusal ilişkiler gözlenemeyecek ve ilişkisiz bulunacaklardır. Oysa DOKKA değişken setleri arasındaki ilişkiyi hem doğrusal hem de eğrisel olarak incelemektedir (Van der Burg, 1988; Henk ve Kiers, 1989).

Bu çalışmada koroner arter çapları, koroner arter çaplarında lezyon olup olmadığı, cinsiyet, DM, sigara kullanımı ve VKİ değişkenleri arasındaki ilişkilerin DOKKA ile incelenmesi amaçlanmıştır.

2. DOĞRUSAL OLMAYAN KANONİK KORELASYON ANALİZİ

DOKKA'de optimal ölçek uzayı $C(\mathbf{h}_j)$ ($j = 1, 2, \dots, m$), her bir değişken için \mathbf{h}_j orijinal veri vektörüyle belirlenmektedir. DOKKA'nin en küçük kareler formülünde $\mathbf{q}_1, \dots, \mathbf{q}_m$ optimal ölçeklenmiş vektör parametreleri yer alırken, DKKA formülünde $\mathbf{h}_1, \dots, \mathbf{h}_m$ veri vektörleri yer alır. DOKKA'de her bir değişken için ($j = 1, 2, \dots, m$) \mathbf{h}_j veri vektörleri yeniden ölçeklenerek \mathbf{q}_j optimal ölçeklenmiş veri vektörleri bulunur.

DOKKA'de kayıp fonksiyon,

$$\sigma(\mathbf{X}, \mathbf{Q}, \mathbf{A}) = \sum_{j=1}^k \text{tr}(\mathbf{X} - \mathbf{Q}_j \mathbf{A}_j)'(\mathbf{X} - \mathbf{Q}_j \mathbf{A}_j) / nk$$

$X'X = nI$ ve $u'X = 0$ koşulları altında minimize edilerek çözüm sağlanır. Formüldeki A katsayı matrisini, Q dönüştürülmüş veri matrisini, X, (n x p) boyutlu obje skorları matrisini (ortalaması 0, varyansı 1), u n birimden oluşan vektörü ifade etmektedir (Gifi, 1990; Van der Burg, 1988; Van der Burg ve ark., 1994; Catrien ve Van der Burg, 1993; Van der Burg ve De Leeuw, 1987; Van der Burg ve Dijksterhuis, 1989; Van der Burg ve Dijksterhuis, 1993).

DOKKA yönteminde oransal, sıralı ve isimsel olmak üzere 3 tip optimal ölçek düzeyi kullanılmaktadır. DOKKA çıktılarında yer alan uygunluk, kayıp ve özdeğerler optimal ölçeklenmiş verilerin uygunluklarını göstermektedir. Her bir boyut ve set için hesaplanan kayıp değeri; her bir sette yer alan değişkenlerin ağırlıklı kombinasyonlarıyla hesaplanamayan obje skorlarında değişimin oranını göstermektedir ve aşağıdaki formül ile hesaplanmaktadır (De Leeuw, 1984; Van der Burg, 1988; Van der Burg ve De Leeuw, 1987; Van der Burg ve Dijksterhuis, 1989; Van der Burg ve Dijksterhuis, 1993). Her bir boyut için

$$\text{özdeğer} = 1 - \text{kayıp}$$

biçiminde bulunur ve bu değer her bir boyutun açıklayıcılığını gösterir. Özdeğerler toplamı toplam uygunluğu vermektedir (Van der Burg, 1988; Van der Burg et al., 1994; SPSS Categories 10.0, 1999).

Dönüştürülmüş değişkenler ile obje skorları arasındaki korelasyonlar bileşen yükleri olarak adlandırılmaktadır. Bileşen yükleri grafiğinde yer alan her bir değişkenin orijinden olan uzaklığı bu değişkenin önemini göstermektedir. İki boyutlu bileşen yükleri grafiği incelenirken grafik, orijinden geçen dikey ve yatay çizgi ile 4 bölgeye ayrılır. Sol-üst bölge ile sağ-alt bölgede yer alan değişkenler 1. boyut olarak; sol-alt bölge ile sağ-üst bölge 2. boyut olarak ele alınır. Bu boyutlardaki değişkenler orijinden geçen dikey çizgilerle bağlanırlar. Çizilen çizgiler ne kadar uzun ve bir doğru şeklinde olursa bu değişkenler arasındaki ilişki de o kadar yüksek olur.

3. DOĞRUSAL OLMAYAN KANONİK KORELASYON ANALİZİNİN KARDİYOLOJİK VERİLERE UYGULANMASI

Trakya Üniversitesi Tıp Fakültesi Kardiyoloji kliniğinde 2000 yılına kadar koroner anjiyografisi yapılan 526 hastanın koroner arter çapları (circumflex: CX, sol ön inen damar: LAD, sağ koroner arter: RCA ve sol ana koroner arter: LMCA ($\bar{x} \pm 1*s$) alınarak k:küçük, o:orta, b:büyük), koroner arter çaplarında lezyon olup olmadığı (var, yok), cinsiyet (E:erkek, K:Kadın), DM (var, yok), sigara içme (e=içiyor, h:içmiyor) ve VKİ (<24.9, 25-40, 40+) değişkenleri ele alındı.

DOKKA'nde birinci sette koroner arter çapları (CX, LAD, RCA ve LMCA), ikinci sette koroner arter çaplarında lezyon olup olmadığı (rca, cx, lad), üçüncü sette cinsiyet, DM, sigara kullanımı ve VKİ değişkenleri ele alındı.

DOKKA'nin uygulanabilmesi için setlerde yer alan değişkenlerin sapan değerlerinin (obje skorları üzerinden elde edilmektedir) olmaması gerekmektedir. Sapan

değerlerin olup olmadığını görmek için obje skorları grafiği çizilir. Şekil-1'de yer alan obje skorları grafiği incelendiğinde veri setlerinde yer alan değişkenlerin sapan değerlerinin olmadığı görülmektedir.

DOKKA'nin iki boyutlu çözümü sonucunda toplam açıklayıcılık %87,2 olarak bulunmuş ve bu açıklayıcılığın %54,7'lik kısmı Boyut-1, %44,3'lük kısmı ise Boyut-2 tarafından açıklanmaktadır. Şekil-2'ye göre; DM, RCA ve LMCA çaplarının lad ve cx'deki lezyonlarla; sigara, cinsiyet ve rca'daki lezyonların VKİ, CX ve LAD çapları ile korelasyonlu olduğu bulunmuştur.

Birbirleri ile ilişkili olan değişkenlerin hangi kategorilerinin ilişkili olduğu ise Şekil-3'te verilmiştir.

Şekil-3 incelendiğinde Boyut-1'de lad ve cx'te lezyon yoksa LMCA ve RCA çaplarının büyük; DM'u bulunmayan, lad ve cx'te lezyon olanlarda LMCA ve RCA çaplarının küçük olduğu bulunmuştur. Boyut-2 incelendiğinde VKİ 40'ın üzerinde olan ve rca'sında lezyon olmayan kadınların sigara içmediği ve CX ile LAD çaplarının büyük; sigara içen, vücut kitle indeksinin 24.9'un altında olan ve rca'da lezyon olan erkeklerin ise CX ve LAD çaplarının küçük olduğu bulunmuştur.

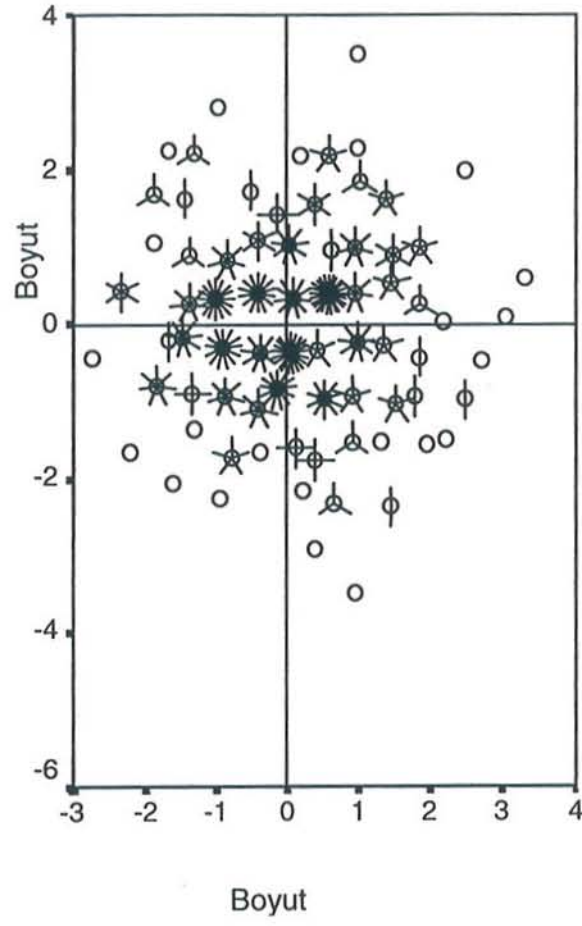
4. TARTIŞMA VE SONUÇ

Nwasokwa ve arkadaşları (1996), koroner anjiyografi için başvuran 884 hasta üzerine yaptıkları çalışmada, koroner arter çaplarının küçük ve büyük olması üzerinde LMA için koroner lezyonların risk faktörü olduğunu; RCA için koroner lezyonların, cinsiyetin ve DM'un; LAD için koroner lezyonların, yaş, cinsiyet ve DM'un; CX için yaş ve DM'un risk faktörü olduğunu lojistik regresyon analizi uygulayarak belirlemişlerdir.

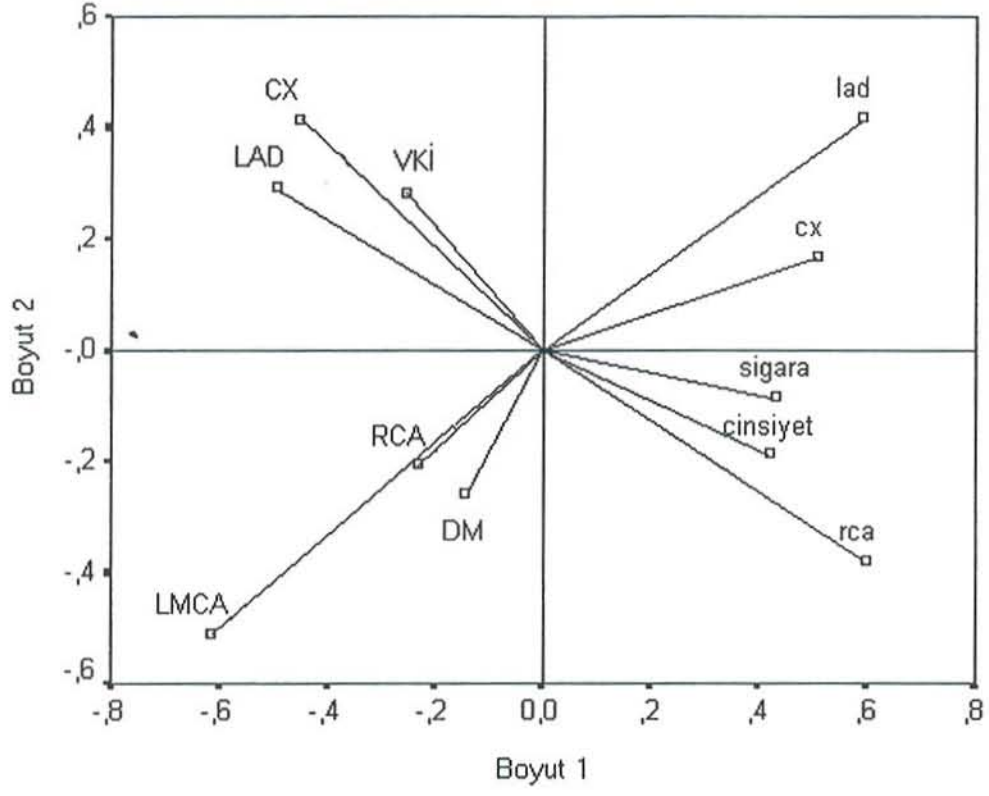
Bu çalışmada koroner arter çapları ile ilgili verilere DOKKA uygulanması sonucunda iki boyutun toplam açıklayıcılığı %87,2 olarak bulundu. Boyut-1'e göre; DM, RCA ve LMCA çaplarının lad ve cx'deki lezyonlarla; Boyut-2'ye göre ise sigara, cinsiyet ve rca'daki lezyonun VKİ, CX ve LAD çapları ile korelasyonlu olduğu bulunmuştur (Şekil-2).

Koroner arter çapı küçük olan hastalarda büyük olanlara kıyasla erkek cinsiyette, sigara kullanımı, DM ve VKİ düşüklüğünün daha fazla olduğu, ayrıca lad, cx ve rca lezyonlarının var olduğu görülmüştür. Bu faktörlerden VKİ artışının LAD ve CX'in çaplarını artırıcı etki gösterdiği bulunmuştur (Şekil-3).

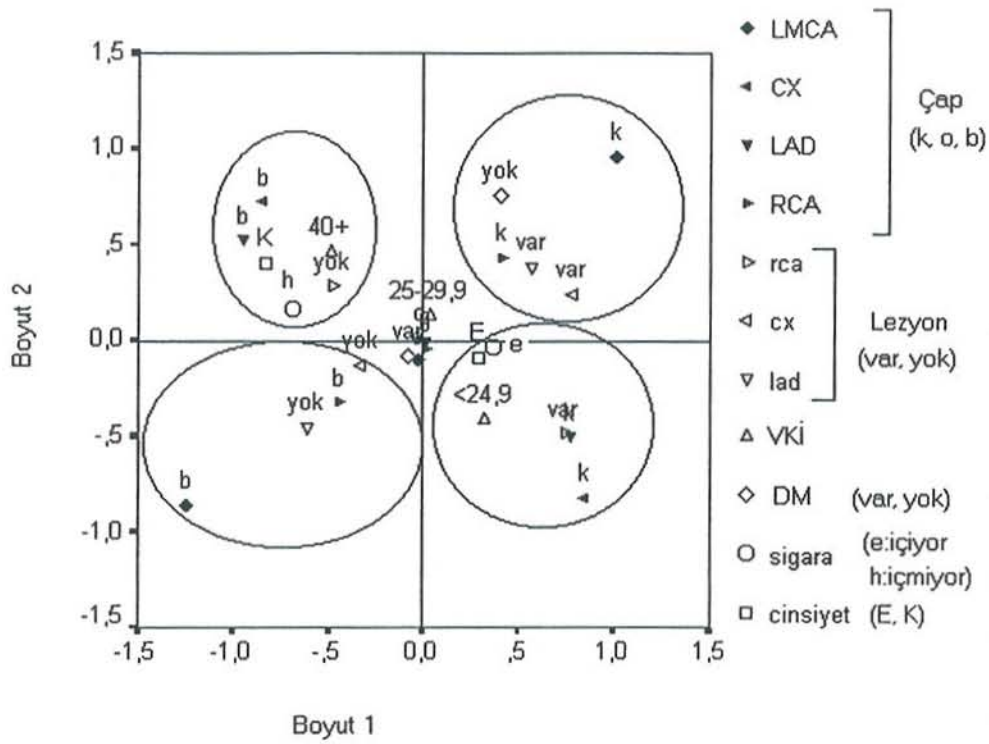
Sonuç olarak, DOKKA isimsel, sıralı, aralıklı, oransal ölçekli yada bunların karışımı değişkenlerden oluşan veri setleri arasındaki ilişkilerin belirlenmesinde uygun bir yöntemdir.



Şekil 1. Obje Skorları Grafiği



Şekil 2. Bileşen Yükleri Grafiği



Şekil 3. Kategori Açıklayıcılıkları Grafiği

KAYNAKLAR

- CATRIEN, C.J.H. & VAN DER BURG, E. (1993), *Longitudinal K-Sets Analysis Using Lagged Variables*, QUESTIO, 17, 327-338.
- GIFI, A. (1990), *Nonlinear Multivariate Analysis*, Chichester, John Wiley & Sons.
- HENK, A., KIERS L. (1989), *Three-Way Methods For The Analysis of Qualitative and Quantitative Two Way Data*, Leiden, DSWO PRESS.
- DE LEEUW, J. (1984), *Canonical Analysis of Categorical Data*, Leiden, DSWO PRESS.
- NWASOKWA, O.N., WEISS, M., GLADSTONE, C., BODENHEIMER, MM. (1996), *Effect of coronary artery size on the prevalence of atherosclerosis*. Am J Cardiol, 78, 741-746.
- ÖZDAMAR, K. (1999), *Paket Programlar ile İstatistiksel Veri Analizi-2*, Eskişehir, Kaan Kitabevi.
- SPSS INC., (1999), *SPSS Categories 10.0 Chicago*. SPSS Inc.
- TATLIDİL, H. (1996), *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Ankara, Cem Web Ofset.

- VAN DER BURG, E. (1988), *Nonlinear Canonical Correlation and Some Related Techniques*, Leiden, DSWO PRESS.
- VAN DER BURG, E., DE LEEUW, J. (1987), *Nonlinear Canonical Correlation Analysis with k Sets of Variables*, Eric Document, Research Report 87-8.
- VANDERBURG, E., DE LEEUW, J., DIJKSTERHUIS, G. (1994), *OVERALS - Nonlinear Canonical Correlation With K Sets of Variables*, *Computational Statistics & Data Analysis*, 18 (1), 141-163.
- VANDERBURG, E, DIJKSTERHUIS, G., (1996), *Generalised Canonical Analysis of Individual Sensory Profiles and Instrumental Data*, *Multivariate Analysis of Data in Sensory Science*, Elsevier Science B.V, 221-258.
- VAN DER BURG, E, DE LEEUW, J., (1983), *Nonlinear Canonical Correlation*, *British Journal of Mathematical and Statistical Psychology*, 36, 54-80.
- VAN DER BURG, E, DIJKSTERHUIS, G., (1989), *Nonlinear Canonical Analysis of Multiway Data: Multiway Data Analysis*, 147-158.
- VAN DER BURG, E, DIJKSTERHUIS, G., (1993), *Nonlinear Generalised Canonical Analysis: Introduction and Application from Sensory Research*, *Advances in Longitudinal and Multivariate Analysis in the Behavioral Sciences*, ITS, 193-203.

Investigation of Cardiological Data By Nonlinear Canonical Correlation Analysis

ABSTARCT

Nonlinear Canonical Correlation Analysis (NLCCA) is a method to reveal relations between two or more sets contains nominal, ordinal, numerical or mixed variables.

In this study 526 patients who were coronary angiographies were studied at Department of Cardiology, Faculty of Medicine at Trakya University from 1998 to 2000. Coronary artery diameters were taken up in the first set (circumflex: CX, Left Anterior Descending: LAD, Right Coronary Artery: RCA and Left Main Coronary Artery: LMCA), lesion in the coronary artery diameters were taken up in the second set and sex, diabetes mellitus (DM), smoking and body mass index (BMI) variables were taken up in the third set. The relations between these variables were studied with NLCCA.

Result of analysis with two-dimensional solution, the first dimension explained 54.7% and the second dimension explained

44.3% of relations between the sets of variables. The significant correlations were found between lesions in the LAD and CX with diameters of DM, RCA, and LMCA, and also between smoking, sex, lesion in the RCA and BMI, CX, LAD diameters.

Key Words: *Nonlinear canonical correlation analysis, optimal scaling, coronary artery.*