

Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine bir Araştırma

Öznur İŞÇİ*

ÖZET

İncelenen değişkenleri bağımlı ve bağımsız olmak üzere iki set halinde toplamak mümkün ise, bu setler arasındaki karşılıklı ilişkileri araştırmak amacıyla kullanılan yöntemlerden biriside kanonik korelasyon analizidir. Bu çalışmada Bornova Anadolu Lisesi öğrencilerinin not ortalamaları ile psikolojik skorlar arasındaki ilişki incelenmiştir. Öğrencilerin not ortalamalarını alarak kanonik korelasyon ve açıklanmış varyans oranı istatistiklerini GENSTAT paket programı ile hesaplayarak yorumlamalar yapılmaktadır. Çalışmada kanonik korelasyon analizi, kanonik açıklama analizi ve diğer çok değişkenli yöntemlerle ilişkileri gösterilmiş ve Wollenberg'in ortalama açıklanmış varyans yönteminin 9x9 korelasyon matrisi üzerinde bir uygulaması verilmiştir.

Anahtar Kelimeler : Kanonik korelasyon analizi, açıklanmış varyans oranı istatistikleri, çok değişkenli yöntemler.

1. GİRİŞ

Kanonik Korelasyon Analizi(Canonical Correlation Analysis), $p > 1$ ve $q > 1$ sayıda değişken içeren iki veri seti (x ve y) arasındaki değişkenler arası ilişkileri ortaya koymak amacıyla yararlanılan bir yöntemdir. Basit korelasyon analizi, iki değişken (y_i, x_i $i=1,2,\dots,n$) arasındaki ikili ilişkiyi r_{xy} korelasyon katsayısı aracılığı ile değerlendiren bir yöntemdir. Çoklu korelasyon analizi ise bir bağımlı değişken ile iki ya da daha fazla bağımsız değişken ($y_1, x_1, x_2, \dots, x_p$) arasındaki ilişkileri değerlendiren bir yöntemdir. Kanonik korelasyon ise, iki veya daha fazla değişken içeren değişkenler seti ($x_1, x_2, \dots, x_p; y_1, y_2, \dots, y_q$) arasındaki ilişkiyi doğrusal bileşenler aracılığı ile değerlendiren çok değişkenli bir yöntemdir(Özdamar,1999).

Kanonik açıklama analizinin kanonik korelasyon ve çok değişkenli çoklu regresyon arasında yer aldığı söylenebilir. Çok değişkenli çoklu regresyon bir orijinal uzaydan diğer bir orijinal uzaya tekil bir dönüşüm sağlamaktadır. Halbuki kanonik açıklama analizi bir setin değişkenlerinin tümünü aynı anda dikleştirmekte ve diğer setin orijinal uzayını tahminlemek için kullanılmak üzere dik bir dönüşüm sağlamaktadır.

* Dr., Celal Bayar Üniversitesi, Bilgisayar Araştırma ve Uygulama Merkezi

Kanonik korelasyon analizinde yapısal korelasyonların (bir setin değişkenleri ile o setin faktörleri arasındaki korelasyonlar) zayıf olması durumunda, değişkenlerin faktörler tarafından açıklanmış varyans oranları da oldukça düşük gerçekleşecektir. Buna karşılık bu iki önemsiz faktör arasında yüksek bir kanonik korelasyon bulunabilir. Gerçekte bu güçlükler kanonik korelasyon analizinin yapısal korelasyonlar üzerinde durmamasından kaynaklanmaktadır ve faktörlerin yorumlanmasında güçlükler neden olmaktadır(Wollenberg,1977). Halbuki en yüksek korelasyonlu olarak elde edilen bu faktörlere bir setteki değişkenlerin ne derece katkıda bulunduğu bilinmesi gerçekten önemlidir. Bu da yapısal korelasyonların bilinmesini gerektirir.

2. KANONİK KORELASYON ANALİZİ

Kanonik korelasyon analizi çok değişkenli bir istatistik tekniktir. Kanonik korelasyon analizi gittikçe artan bir uygulama alanına sahip olmakla beraber, bağımlı (kriter) ve bağımsız (tahmin) değişken setleri arasında ilişki araştırıldığı zaman kullanılan tekniklerden biridir(Şen H., Kalyoncu, C.,2001).

Kanonik korelasyon analizi, iki değişken seti arasındaki ilişkilerin incelenmesi ve bunların yorumlanması amacıyla kullanılır. Çok boyutlu sistemdeki değişkenlerin bağımlı ve bağımsız olmak üzere iki sete ayrılabilmesinin mümkün olduğu durumlarda uygulanan yöntemlerden birisi olan kanonik korelasyon analizi, ilk olarak Hotelling(1936) tarafından geliştirilmiştir. Analizin esas amacı, her iki değişken setine ait olmak üzere, aralarında en yüksek korelasyon bulunan doğrusal bileşenlerin elde edilmesidir. Bu en yüksek korelasyona kanonik korelasyon, elde edilen doğrusal bileşenlere ise, kanonik değişkenler adı verilmektedir. Bununla birlikte önemli kanonik korelasyon sayısı birden çok olabilmektedir.

Kanonik korelasyon analizi, rassal değişkenler kümesinin doğrusal fonksiyonları arasındaki en büyük korelasyonları hesaplamaya çalışır ve çok boyutlu ana kümeden çekilmiş iki veya daha çok değişken kümesi arasındaki ilişki ile ilgilenir. Rassal değişkenler kümesinin doğrusal fonksiyonları arasındaki maksimum korelasyonları hesaplamaya çalışır. Kanonik korelasyon ile, rassal değişkenler kümesinin maksimum korelasyonlu ve birim varyanslı doğrusal bileşimi elde edilmek istenir. Kanonik analiz çoklu regresyon analizinin özel bir halidir. Çoklu regresyon analizi bir bağımlı, birden fazla bağımsız değişken arasındaki ilişkiyi araştırmasına karşın kanonik analizde p ve q tane değişken bulunmaktadır.

2.1 Kanonik Değişkenlerin Elde Edilmesi

$q < p$ olmak üzere, $x(p+q) \times 1$ şans vektörünün $x_1(q \times 1)$ ve $x_2(p \times 1)$ şeklinde iki alt sete ayrıldığı varsayalım. Bu alt setler,

$$x_1 = (x_1, x_2, \dots, x_q)^T$$

ve

$$x_2 = (x_{q+1}, x_{q+2}, \dots, x_p)^T$$

şeklinde gösterilebilir. Burada x_1 seti açıklayıcı değişkenler seti ve x_2 seti ise bağımlı değişkenler seti adı verilmektedir. Bunun sonucu olarak, x şans vektörünün ortalaması ve varyans-kovaryans matrisi de,

$$\mu_x = \begin{pmatrix} \mu_{x_1} \\ \mu_{x_2} \end{pmatrix}, \quad \text{Cov}(x) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (1)$$

şeklinde parçalanabilecektir. Böylece,

$$\Sigma_{11} = \text{Cov}(x_1), \quad \Sigma_{12} = \text{Cov}(x_1, x_2) = \Sigma_{21}^T, \quad \Sigma_{22} = \text{Cov}(x_2)$$

olduğu açıktır. x_1 setine ait doğrusal bileşen $a = w^T x_1$ ve x_2 setine ait doğrusal bileşen ise $b = v^T x_2$ olsun. Görüldüğü gibi, a ve b yapay değişkenleri orijinal değişkenlerin doğrusal kombinasyonları olarak ifade edilmektedir. a ve b yapay değişkenleri için,

$$\text{Var}(a) = w^T \Sigma_{11} w, \quad \text{Var}(b) = v^T \Sigma_{22} v, \quad \text{Cov}(a, b) = \text{Cov}(b, a)^T = w^T \Sigma_{12} v$$

ifadeleri geçerlidir. Buna dayanarak a ve b doğrusal bileşenleri arasındaki korelasyon,

$$\rho_{a,b} = \frac{w^T \Sigma_{12} v}{\left[(w^T \Sigma_{11} w) (v^T \Sigma_{22} v) \right]^{1/2}} \quad (2)$$

olacaktır. $\text{Var}(a) = w^T \Sigma_{11} w$ ve $\text{Var}(b) = v^T \Sigma_{22} v$ şeklinde birim varyans kısıtlamaları altında, $\rho_{a,b}$ korelasyonu en yüksek yapan w ve v vektörlerinin elde edilmesi kanonik analizin çözüm aşamasını oluşturmaktadır. Kısıtlar için iki lagrange çarpanı kullanılarak amaç fonksiyonu şu şekilde verilebilir:

Amaç fonksiyonunun çözümü araştıracıyı,

$$Z = w^T \Sigma_{12} v - 1/2 \lambda_1 (w^T \Sigma_{11} w - 1) - 1/2 \lambda_2 (v^T \Sigma_{22} v - 1) \quad (3)$$

$$(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \lambda_1^2 I) w = 0 \quad (4)$$

ve

$$(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \lambda_2^2 I) v = 0 \quad (5)$$

özdeğer-özvektör problemlerine götürmektedir. (4) ve (5) ifadelerinden elde edilecek özdeğerler (λ_1^2, λ_2^2) birbirlerine ve aynı zamanda da kanonik korelasyonun karesine eşittirler. Burada, $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ ve $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ matrisleri pozitif tanımlı matrisler olsunlar ve rankı küçük olanın rankı k olsun. Böylece, yukarıda verilen (4) ve (5) ifadelerinden elde edilecek özdeğerler için,

$$\lambda_{11}^2 > \lambda_{12}^2 > \dots > \lambda_{1k}^2 > 0$$

yazılabilir ve $(w_1, v_1), (w_2, v_2), \dots, (w_k, v_k)$ ise bu özdeğerlere karşılık gelen özvektörlerdir.

2.2 Kanonik Değişkenlerin Kovaryans Yapısı

Σ_{11} ve Σ_{22} matrislerinin tekil olmadıkları varsayılın ve $K(q \times q)$ matrisi,

$$K = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \quad (6)$$

şeklinde tanımlansın. Buna dayalı olarak da N_1 ve N_2 matrisleri,

$$N_1 = KK^T \quad N_2 = K^TK \quad (7)$$

olsun. Böylece,

$$M_1 = \Sigma_{11}^{-1/2} N_1 \Sigma_{11}^{1/2} \quad M_2 = \Sigma_{22}^{-1/2} N_2 \Sigma_{22}^{1/2} \quad (8)$$

matrisleri için,

$$M_1 = \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad (9)$$

$$M_2 = \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad (10)$$

yazılabilecektir. (9) ve (10) ifadelerinden görüldüğü gibi M_1 ve M_2 matrisleri kanonik korelasyon analizinde özdeğer-özvektör ayrışımı uygulanan matrisler olmaktadır.

3. AÇIKLANMIŞ VARYANS ORANLARI

İki değişken seti arasındaki ilişkilerin incelenmesi amaçlandığında, kanonik korelasyon istatistiğine alternatif bir istatistik Stewart ve Love(1968) tarafından ileri sürülen açıklanmış varyans oranı(redundancy index) istatistiğidir. Bilindiği gibi kanonik korelasyon analizinde, aralarında en yüksek korelasyona sahip doğrusal bileşenler elde edilmekte, diğer bir ifadeyle a_i ve b_i gibi iki yapay değişken arasındaki korelasyon en yüksek yapılmakta fakat elde edilen bu yapay değişkenlerin kendi setlerindeki değişimin ne kadarını açıklayabildikleri üzerinde hiç durulmamaktadır.

Bu durum en yüksek korelasyonlu olarak elde edilen kanonik değişkenlerin yorumlanmasında araştırmacıya önemli güçlükler getirmektedir [Van Den Wollenberg, A.L.(1977) ve De Sarbo, W.S.(1981)]. Çünkü x_1 ve x_2 değişken setlerinden elde edilen en yüksek korelasyonlu a_1 ve b_1 doğrusal bileşenleri, kendi setlerindeki değişimin çok küçük bir kısmını açıklıyor olabilmektedir. İki set arasındaki karşılıklı ilişkilerin yorumlanmasının ve değerlendirilmesinin kendi setlerini temsil gücü bu derece zayıf olan doğrusal bileşenlere dayandırılmasının uygun olmayacağı açıktır.

Bu sorunun çözümü için, Wollenberg(1977), kendi setlerindeki varyasyonun mümkün olan en yüksek kısmını açıklayabilme özelliğine sahip (elde edildikleri setleri temsil gücü bakımından en güçlü faktörleri veren) doğrusal bileşenlerin elde edilmesi ile ilgili bir yöntem geliştirilmiştir.

3.1 Yapısal Korelasyonlar ve Açıklanmış Varyans İndeksi

x_1 setinin değişkenleri ile bu setin i . kanonik faktörü arasındaki yapısal korelasyonları (faktör yüklerini) içeren $h_i(p \times 1)$ sütun vektörü,

$$\begin{aligned} h_i &= \text{Cov}(x_1, c_i) = E\{[x_1 - E(x_1)][t_i^T x_1 - E(t_i^T x_1)]^T\} \\ &= E\{[x_1 - \mu_{x1}][x_1 - \mu_{x1}]^T t_i\} = \sum_{11} t_i = R_{11} t_i \end{aligned} \quad (11)$$

olarak elde edilir. Benzer olarak, y seti değişkenleri ile bu setin i . kanonik faktörü arasındaki yapısal korelasyonları içeren $g_i(q \times 1)$ sütun vektörü,

$$\begin{aligned} g_i &= \text{Cov}(x_2, d_i) = E\{[x_2 - E(x_2)][s_i^T x_2 - E(s_i^T x_2)]^T\} \\ &= E\{[x_2 - \mu_{x2}][x_2 - \mu_{x2}]^T s_i\} = \sum_{22} s_i = R_{22} s_i \end{aligned} \quad (12)$$

şeklinde elde edilir. Stewart ve Love(1968), bu yapısal korelasyonlardan hareket ederek "açıklama indeksi" (redundancy index) olarak adlandırılan bir açıklanmış varyans indeksi tanımlamışlardır. Bu indeks, bir setin değişkenlerinin diğer setin verilen bir faktörü tarafından açıklanan varyans olarak tanımlanır. i . kanonik değişken çifti dikkate alınarak, x_1 ve x_2 setlerine ait açıklama indeksleri,

$$R_{x1i} = \lambda^2_{1i} (h_i^T h_i) / q \quad (13)$$

$$R_{x2i} = \lambda^2_{2i} (g_i^T g_i) / p \quad (14)$$

şeklinde tanımlanırlar (Wollenberg, 1977). Ayrıca Stewart ve Love (1968), bir sete ait k tane açıklama indeksini toplayarak o setin bütün kanonik faktörleri tarafından açıklanabilen toplam varyans oranını bir indeks olarak ifade etmişlerdir. Buna göre, x_1 ve x_2 setlerinin toplam açıklanmış varyans oranları sırasıyla,

$$R_{d1} = \sum R_{x1i} \quad (15)$$

$$i = 1, 2, \dots, k$$

$$R_{d2} = \sum R_{x2i} \quad (16)$$

olarak tanımlanır.

3.2 Açıklama İndeksinin En Yüksek Yapılması Sorunu

Açıklama indeksini maksimum yapan faktörler bir anlamda tahminleme gücü optimum olan faktörler olacağından, daha çok tahminleme amacına dayanan araştırmalarda açıklama indeksini en yüksek yapan faktörlerin kullanılması kanonik faktörlerin kullanılmasından daha uygundur.

Örneğin bir pazarlama yöneticisi tüketici marka tercihleri(y), bireysel psikolojik skorlarla(x) tahminleyebilmek için, y seti değişkenlerinin, $w \times x$ faktörü tarafından açıklanan varyansını maksimum yapacak bir w vektörü araştıracaktır (De Sarbo, 1981).

Van Den Wollenberg(1977), kanonik korelasyon istatistiği yerine, açıklama indeksi istatistiğini en yüksek yapan bir yöntem geliştirmiştir. Yöntemin esas amacı R_{x1i} ve R_{x2i}

istatistiklerini en yüksek yapan birim varyanslı $c=t^T x_1$ ve $d=s^T x_2$ doğrusal bileşenlerin belirlenmesidir. Burada, c ve d değişkenleri “en yüksek varyans açıklayıcı değişkenler” olarak tanımlanır. Bu işi yapan analize ise, “kanonik açıklama analizi” denilmektedir.

$R_{x_{1i}}$ ve $R_{x_{2i}}$ indekslerini en yüksek yapmak için, bu ifadelerin payındaki açıklanmış varyansları en yüksek yapmak yeterli olacaktır. Daha önce verilen birim varyans kısıtlamaları altında, kısıtlı optimizasyon probleminin amaç fonksiyonları,

$$\begin{aligned} Z_1 &= t^T R_{12} R_{21} t - \mu_1 (t^T R_{11} t - 1) \\ \text{ve} \\ Z_2 &= s^T R_{21} R_{12} s - \mu_2 (s^T R_{22} s - 1) \end{aligned} \quad (17)$$

şeklinde verilir. (17) nolu ifadenin t ve s’ye göre kısmi türevleri alınıp sıfıra eşitlenirse,

$$\begin{aligned} (R_{11}^{-1} R_{12} R_{21} - \mu_1 I) t &= 0 \\ \text{ve} \\ (R_{22}^{-1} R_{21} R_{12} - \mu_2 I) s &= 0 \end{aligned} \quad (18)$$

ifadeleri elde edilir. (18) ifadesi ile verilen ilk homojen eşitliğin $t=0$ dan farklı çözümünün olabilmesi için μ_1 ’in $R_{11}^{-1} R_{12} R_{21}$ matrisinin bir özdeğeri, t’nin de bu özdeğere karşılık gelen özvektör olması gerekir. $R_{11}^{-1} R_{12} R_{21}$ matrisinin tam ranklı olduğu ve özdeğerlerinin birbirlerinden farklı oldukları varsayıldığında,

$$\mu_{11} > \mu_{22} > \dots > \mu_{1p} > 0$$

yazılabilecektir. Daha önce verilen birim varyans kısıtlaması gereği,

$$\mu_{1i} = t_i^T R_{12} R_{21} t_i \quad (19)$$

elde edilir. Bu sonuca göre, x_2 seti değişkenlerinin, x_1 setinin i.inci en yüksek varyans açıklayıcı değişkeni tarafından açıklanan varyansı μ_{1i} özdeğerine eşit olmaktadır. μ_{1i} özdeğeri x_2 setinin i.inci açıklama indeksi $R_{x_{2i}}$ ’nin q katı olarak yorumlanabilir. Bu durumda x_2 setinin toplam açıklama indeksini gösteren (16) ifadesi,

$$R_{d_2} = 1/q \sum \mu_{1i} \quad i = 1, 2, \dots, p \quad (20)$$

şeklinde yazılabilecektir. Diğer taraftan, $R_{x_{1i}}$ indeksini en yüksek yapmak için, (18) ifadesi ile verilen ikinci homojen eşitlik, μ_{2i} ve s_i için çözüldüğünde ve ilgili basitleştirmeler yapıldığında,

$$\mu_{2i} = s_i^T R_{21} R_{12} s_i \quad (21)$$

elde edilir. Bu son eşitliğe göre, μ_{2i} özdeğeri, x_1 setinin i.inci açıklama indeksi $R_{x_{1i}}$ ’nin p katı olarak yorumlanabilir. Bu sonuca dayanarak, x_1 setinin toplam açıklama indeksini gösteren (15) ifadesi aşağıda verildiği gibidir,

$$R_{d_1} = 1/p \sum \mu_{2i} \quad i = 1, 2, \dots, q \quad (20)$$

4. UYGULAMA

Bu bölümde, kanonik korelasyon analizi ve kanonik açıklama analizi daha önce üzerinde durulan istatistiksel özellikler bakımından karşılaştırılabilmek amacıyla, Tablo 1'deki korelasyon matrisi esas alınarak her iki analizin uygulaması yapılmıştır. Bornova Anadolu Lisesi öğrencilerinin fen dersleri ve sosyal derslerindeki başarı düzeylerinin, bağımsız değişkenler setini oluşturan psikolojik, sosyal ve çevre ile ilgili skorlara dayalı olarak tahminlenmesi amacıyla Lise 2'de okuyan 6 sınıftaki 216 öğrenciye anket uygulanmıştır. Bazı öğrencilerin derste bulunmaması ya da anket formundaki eksik bilgiler dolayısıyla 150 öğrencinin formu değerlendirilmiştir. Lise öğrencilerinin başarı düzeylerini etkileyebileceği düşünülen değişkenler ve 3 döneme ilişkin her dersten aldıkları notlar dikkate alınmıştır. (İşçi, 1995). Bağımsız değişkenler seti (x) aşağıdakileri kapsar:

- x₁= Babanın mesleği
- x₂= Annenin mesleği
- x₃= Arkadaşlık ilişkileri
- x₄= Okulu kendi isteği ile seçip seçmediği?
- x₅= Anne ve baba ayrı mı?
- x₆= Özel ders alıyor mu?
- x₇= Kimin yanında kalarak okuyor?
- x₈= Okulun bulunduğu şehrin kültürel yapısını benimsiyor mu?
- x₉= Ailede geçimsizlik ya da huzursuzluk ortamı var mı?

Bağımlı değişkenler seti (y) aşağıdakileri kapsar:

- y₁= Matematik
- y₂= Geometri, Analitik Geometri
- y₃= Fen Bilimleri(Fizik, Kimya, Biyoloji)
- y₄= Coğrafya
- y₅= Tarih
- y₆= Yabancı Dil(İngilizce, Almanca)
- y₇= Türk Dili Edebiyatı
- y₈= Din
- y₉= Beden

Bu anket ile öğrencilerin başarılarını etkileyebileceği düşünülen faktörler ve psikografik cevaplar alınmıştır. Öğrenciler 1. soruda anne ve babasının mesleğini işaretlemişlerdir. 2. soruda da arkadaşlarıyla olan ilişkilerinde; çok kötü(-2), kötü(-1), arkadaşlık benim için önemli değil(0), normal(1), iyi(2), çok iyi(3) olarak derecelendirilmiştir. 3,4,5 ve 8. sorular da evet(1), hayır(2) olarak belirlenmiştir. 6. soruda kimin yanında kalarak okuyorsunuz sorusuna ailemle(1), öğrenci yurdunda(2), arkadaşlarımla aynı evde kalıyorum(3); 7.soruda okulunuzun bulunduğu şehrin kültürel yapısını ne derece benimsiyorsunuz sorusuna hiç benimsemiyorum(0), çok az benimsiyorum(1), orta derece benimsiyorum(2), tam benimsiyorum(3) olarak dikkate alınmıştır. Öğrencilerin 3 dönemde aldıkları ortak ya da seçmeli dersler birleştirilerek her derse ilişkin tek bir not hesaplanmıştır. Örneğin matematik 1. dönem 4, 2. dönem 3, 3. dönem 5 olsun, matematik notu 4 olarak değerlendirilmiştir. Diğer ders notları da benzer şekilde hesaplanmıştır. Tablo 1'de kanonik korelasyon analizi ve kanonik açıklama analizi için korelasyon matrisi görülmektedir. Tablo 1'de verilen korelasyon matrisi, ilk 9 değişken bağımsız değişkenler setini, diğer 9 değişken bağımlı değişkenler setini meydana getirecek şekilde parçalanmıştır.

Tablo 1: Orijinal Değişkenlerin Korelasyon Matrisi (R)

R₁₁

R₁₂

1.0000																					
0.2500	1.0000																				
0.1200	0.1900	1.0000																			
-0.1800	-0.0200	-0.3200	1.0000																		
0.0900	0.0900	0.2200	-0.2700	1.0000																	
-0.0800	-0.0200	-0.0700	0.1100	-0.1600	1.0000																
-0.1400	0.0200	-0.0900	0.2500	-0.0100	0.0100	1.0000															
0.2500	0.2300	0.2400	-0.3800	0.2600	-0.0200	-0.0200	1.0000														
0.1200	0.1600	0.2500	-0.1700	0.0300	0.0000	-0.0100	0.2000	1.0000													
-0.0700	0.0600	0.0500	-0.0500	0.0600	-0.0100	-0.0500	-0.1000	0.1400	1.0000												
0.0900	0.0700	-0.0600	-0.0100	0.0700	0.0800	0.0600	-0.1000	0.1200	0.6400	1.0000											
0.1000	0.1200	-0.0100	0.0500	0.1700	-0.0100	0.0000	-0.1300	0.0600	0.7600	0.7800	1.0000										
0.0200	-0.0700	-0.0500	0.0200	0.0100	0.0500	0.0200	0.0100	0.1400	0.5500	0.5200	0.5900	1.0000									
0.0300	0.0100	0.0600	0.0500	0.0100	0.0700	0.0000	-0.0500	0.0400	0.6200	0.5400	0.6700	0.6100	1.0000								
0.1900	0.0700	0.1900	-0.0400	0.1100	0.0100	0.0400	-0.0700	0.1000	0.5800	0.5700	0.7300	0.5100	0.6200	1.0000							
0.0400	0.0700	0.1800	-0.0400	0.1700	-0.0100	-0.0800	-0.0700	0.0900	0.6400	0.5200	0.6600	0.4200	0.6500	0.6200	1.0000						
-0.0200	0.0200	0.0500	-0.0100	0.1400	0.1100	-0.0100	-0.0600	0.0200	0.5100	0.5200	0.5500	0.5000	0.6300	0.5500	0.5900	1.0000					
0.0800	0.0400	0.0000	0.0200	-0.0100	0.0900	-0.1400	-0.0700	0.0900	0.2100	0.1400	0.1700	0.1800	0.1700	0.0500	0.2100	0.1700	1.0000				

R₂₁

R₂₂

Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine bir Araştırma

Kanonik korelasyon analizinin x ve y setleri ile ilgi sonuçlar sırasıyla, Tablo 2 ve 3'te verilmiştir.

Tablo 2: x Setine Ait Kanonik Korelasyon Analizi Sonuçları

Kök no	λ_i	λ_i^2	H_i	$R_{xi} = \lambda_i^2 H_i$	R_{xi} / R_{d1}
1	0.4690	0.2199	0.0946	0.0208	0.2845
2	0.4116	0.1694	0.0993	0.0168	0.2298
3	0.3620	0.1310	0.1169	0.0153	0.2063
4	0.3012	0.0907	0.0820	0.0074	0.1012
5	0.2096	0.0439	0.0822	0.0034	0.0465
6	0.1990	0.0396	0.1079	0.0043	0.0588
7	0.1880	0.0353	0.1187	0.0042	0.0574
8	0.0790	0.0062	0.1458	0.0009	0.0123
9	0.0138	0.0002	0.1526	0.0000	0.0000

Toplam **1.0000** **$R_{d1} = 0.0731$**

Tablo 3: y Setine Ait Kanonik Korelasyon Analizi Sonuçları

Kök no	λ_i	λ_i^2	G_i	$R_{yi} = \lambda_i^2 G_i$	R_{yi} / R_{d2}
1	0.4690	0.2199	0.1193	0.0262	0.3819
2	0.4116	0.1694	0.0399	0.0068	0.0991
3	0.3620	0.1310	0.0554	0.0073	0.1064
4	0.3012	0.0907	0.1781	0.0162	0.2361
5	0.2096	0.0439	0.0931	0.0041	0.0598
6	0.1990	0.0396	0.0857	0.0034	0.0496
7	0.1880	0.0353	0.0740	0.0026	0.0379
8	0.0790	0.0062	0.3170	0.0020	0.0291
9	0.0138	0.0002	0.1526	0.0000	0.0000

Toplam **1.0000** **$R_{d2} = 0.0686$**

Tablolarda yer alan $H_i = h_i^T h_i / p$ değeri a_i faktörlerinin x setinde açıkladıkları varyans oranını, $G_i = g_i^T g_i / q$ değeri b_i faktörlerinin y setinde açıkladıkları varyans oranını ifade etmektedir. λ_i ise i.inci kanonik korelasyondur. Tablo 2 ve Tablo 3 incelendiğinde x ve y setlerinin aralarında en yüksek korelasyona sahip en önemli faktörlerin, kendi setlerindeki değişimin sırasıyla ancak 0.0946 ve 0.1193'lük kısımlarını açıklayabildikleri görülmektedir. Bu, birinci kanonik faktör çiftinin bile açıklanan varyans anlamında kendi setlerini temsil etme gücünün çok zayıf olduklarını göstermek bakımından oldukça anlamlıdır. x setinin birinci faktörü, y setindeki değişimin 0.0262'sini açıklayabilmektedir ve bu değer de y setinde açıklanması mümkün olan varyans oranının 0.3819'una karşılık gelmektedir. y setinin birinci

faktörü, x setindeki değişimin ancak 0.0208'ini açıklayabilmektedir ve bu değer de x setinde açıklanması mümkün olan varyans oranının yalnızca 0.2845'idir. y setinin en önemsiz faktörlerinden biri olarak dikkate alınabilecek olan $v_8(0.3170)$ doğrusal bileşeni, bu sete ait tüm doğrusal bileşenler içerisinde en yüksek varyans açıklama yüzdesine sahip olan bileşendir. Benzer şekilde, x setinin en önemsiz faktörlerinden birisi olarak dikkate alınabilecek $w_9(0.1526)$ doğrusal bileşeni, bu sete ait tüm doğrusal bileşenler içerisinde en yüksek varyans açıklama yüzdesine sahip olan bileşendir.

Tablo 4: Kanonik Korelasyon Analizinde x Setine Ait Kanonik Faktörler

	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈	a ₉
x ₁	0.4781	0.6731	-0.5572	-0.0851	-0.1996	-0.1393	0.0223	0.2753	-0.1314
x ₂	0.3749	-0.3304	0.1156	0.2061	0.3366	-0.1092	0.5797	0.0288	0.6216
x ₃	0.1037	0.6630	0.7307	-0.3486	-0.0107	0.1591	0.0125	-0.3085	0.0393
x ₄	0.2085	-0.1135	-0.1439	-0.6016	-0.6931	-0.0002	-0.0231	-0.6716	0.1984
x ₅	0.6904	-0.3889	0.0375	0.0726	-0.0557	0.2244	-0.6890	-0.0124	0.0005
x ₆	0.0320	0.1652	-0.2332	0.2254	0.0310	0.9379	0.0367	-0.0249	0.1543
x ₇	-0.0583	0.3442	-0.3384	0.0316	0.7821	-0.1572	-0.2941	-0.3644	0.0546
x ₈	-0.5942	-0.0916	-0.1460	-0.5950	-0.2387	-0.0108	-0.3128	0.2286	0.6245
x ₉	-0.1501	0.0672	-0.0674	0.7676	-0.4665	-0.2446	-0.2656	-0.3292	0.1922

Tablo 5: Kanonik Korelasyon Analizinde y Setine Ait Kanonik Faktörler

	b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	b ₇	b ₈	b ₉
y ₁	-0.6774	-0.3088	0.8970	0.8054	0.3056	-0.3082	0.4278	-0.3014	-0.5130
y ₂	-0.2756	0.4419	-0.7054	0.8805	0.5164	0.2459	0.0872	0.2010	0.8801
y ₃	1.4593	-1.2296	-0.5097	-0.6724	-0.2602	-0.2428	0.0362	-0.2548	-0.3899
y ₄	-0.5321	-0.0027	-0.3683	0.1100	-0.6986	-0.2547	-0.9343	-0.0817	-0.1711
y ₅	-0.3802	0.2892	-0.2155	-0.6307	-0.0626	0.3600	0.8660	-0.9609	0.4331
y ₆	0.2234	1.3818	0.1741	0.1365	0.2582	-0.1971	-0.0249	0.0743	-0.6294
y ₇	0.2348	0.0725	0.7158	-0.0021	-0.4970	-0.0708	-0.6440	0.2945	1.0632
y ₈	0.1843	-0.4606	0.1268	-0.0133	0.4147	1.0343	-0.3940	0.1213	-0.5638
y ₉	0.1925	0.2093	-0.1156	0.2879	-0.6399	0.2948	0.5112	0.3349	-0.2228

Kanonik korelasyon analizi uygulaması sonucu elde edilen bağımsız ve bağımlı değişkenler setine ait kanonik değişkenler(kanonik bileşenler) sırasıyla Tablo 4 ve Tablo 5'de verilmiştir. Aralarında en yüksek korelasyona sahip birinci kanonik değişken çifti(w_1, v_1) dikkate alınsın. w_1 kanonik faktörün yorumuna, kendi setine ait değişkenler arasında en yüksek katkıyı sağlayan değişkenlerin sırasıyla x_5 (anne ve babanın ayrı olup olmaması), x_8 (okulun bulunduğu çevrenin kültürel yapısına uyum derecesi), x_1 (babanın mesleği) ve x_2 (annenin mesleği) değişkenleri olduğu görülmektedir. y setine ait en önemli faktör olan v_1 faktörü dikkate alındığında ise, bu doğrusal bileşenin yorumuna en yüksek katkıda bulunan değişkenlerin sırasıyla y_3 (fen bilimleri), y_1 (matematik), y_4 (coğrafya), y_5 (tarih) dersleri olduğu

görülmektedir. x seti, öğrencinin başarısını etkileyen psikolojik, sosyal ve ekonomik değişkenler seti ve y seti ise öğrencinin fen ve sosyal bilimlerle ilgili derslerindeki başarı durumunu yansıtan değişkenler olduğuna göre, öğrencinin içinde bulunduğu psikolojik, sosyal ve ekonomik yapıyı en fazla belirleyen bu değişkenlerin anne ve babanın ayrı olup olmaması, okulun bulunduğu çevrenin kültürel yapısına uyum derecesi, babanın mesleği ve annenin mesleği olduğu ifade edilebilir.

Tablo 4 ve Tablo 5'deki bilgiler dikkate alındığında, hangi değişkenlerin her iki setin en önemli faktörleri olan birinci faktör çiftine veya daha sonra gelen faktör çiftlerinin yorumuna en çok katkıda buldukları görülebilir. Kanonik korelasyon analizinde bağımlı ve bağımsız değişkenler setine ait kanonik faktörlerin tümü için sonuçlar mevcut ise de ilk faktörün yorumu yapılmıştır.

Tablo 6: x Setine Ait Kanonik Açıklama Analizi Sonuçları

Kök no	h_i	μ_{2i}	$R_{xi} = \mu_{2i}/p$	R_{xi} / R_{d1}
1	0.0873	0.2004	0.0223	0.3042
2	0.1020	0.1655	0.0184	0.2510
3	0.0905	0.1231	0.0137	0.1869
4	0.1030	0.0626	0.0069	0.0941
5	0.1098	0.0396	0.0044	0.0600
6	0.1052	0.0352	0.0039	0.0532
7	0.0826	0.0272	0.0030	0.0409
8	0.1662	0.0067	0.0007	0.0095
9	0.1534	0.0002	0.0000	0.0000

Toplam 1.0000 $R_{d1} = 0.0733$

Tablo 7: y Setine Ait Kanonik Açıklama Analizi Sonuçları

Kök no	g_i	μ_{1i}	$R_{yi} = \mu_{1i}/p$	R_{yi} / R_{d2}
1	0.1218	0.3749	0.0417	0.6079
2	0.0245	0.0753	0.0084	0.1224
3	0.0453	0.0575	0.0064	0.0933
4	0.2062	0.0403	0.0045	0.0655
5	0.0782	0.0277	0.0031	0.0452
6	0.0912	0.0238	0.0026	0.0379
7	0.0545	0.0131	0.0014	0.2040
8	0.3405	0.0041	0.0005	0.0073
9	0.0378	0.0001	0.0000	0.0000

Toplam 1.0000 $R_{d2} = 0.0686$

Tablo 6 ve Tablo 7'deki açıklanan varyans oranları incelenirse, kanonik açıklama analizinin faktörleri elde edildikleri setleri temsil gücü bakımından optimum bir şekilde sıralandığı görülmektedir. Tablo 6 dikkate alındığında, y setine ait tüm en yüksek varyans açıklayıcı faktörlerin, x setinin toplam değişimin genellikle %7.3'ünü açıklayabildiği görülmektedir. Bu değer, öğrencilerin fen ve sosyal derslerindeki başarı düzeylerine dayalı faktörlerin, bu öğrencilerin aile yapısı ve sosyal yapıları ile ilgili değişkenler setinde açıklayabilmesi mümkün olan en yüksek değişim yüzdesidir. Tablo 7 dikkate alındığında ise, x setine ait tüm en yüksek varyans açıklayıcı faktörlerin, y setindeki toplam değişimin % 6.9'unu açıklayabildiği görülmektedir.

Kanonik korelasyon analizi sonuçlarıyla karşılaştırma yapılırsa, kanonik açıklama analizinde x setinin ilk iki faktörü y setinde açıklanabilecek toplam varyasyonun 0.7303'ünü açıkladığı halde bu değer kanonik korelasyon analizinde 0.4810 düzeyinde kalmaktadır. y setinin ilk iki en yüksek varyans açıklayıcı faktörü x setinde açıklanabilecek toplam varyasyonun 0.5552'sini açıkladığı halde, bu değer kanonik korelasyon analizinde 0.5143'da kalmaktadır. Bu durum, kanonik korelasyon analizinin aksine, kanonik açıklama analizinde her iki sete ait tahminleme gücünün setlerin ilk iki faktöründe toplandığını göstermektedir.

Tablo 8: Kanonik Açıklama Analizinde x Setine Ait Kanonik Faktörler

	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	c ₈	c ₉
x ₁	-0.3357	0.0615	0.8771	-0.1457	0.3083	-0.2378	-0.0791	0.2677	0.1254
x ₂	-0.2320	-0.0167	-0.2960	0.5051	0.2650	-0.3679	0.3867	-0.1610	-0.6134
x ₃	-0.1563	0.9461	-0.0034	-0.4243	-0.1870	0.0676	0.0855	-0.2832	-0.0374
x ₄	-0.0003	0.0985	0.0832	-0.2910	0.4886	0.1829	-0.7364	-0.6758	-0.1943
x ₅	-0.6160	-0.0027	-0.1703	0.3674	0.0766	0.5579	-0.4859	0.2574	0.0013
x ₆	-0.2777	-0.2150	0.1084	-0.3195	0.2123	0.6497	0.5564	0.0209	-0.1596
x ₇	0.0208	0.0212	0.6130	0.4543	-0.6023	0.1924	0.2043	-0.2627	-0.0513
x ₈	0.7886	0.0192	0.918	-0.2221	0.0467	0.2590	-0.3839	0.2081	-0.6305
x ₉	-0.4176	-0.5758	-0.0985	-0.4143	-0.4932	-0.2187	-0.2960	-0.0989	-0.1958

Tablo 9: Kanonik Açıklama Analizinde y Setine Ait Kanonik Faktörler

	d ₁	d ₂	d ₃	d ₄	d ₅	d ₆	d ₇	d ₈	d ₉
y ₁	-0.3769	0.6868	0.8908	-0.5799	-0.0301	0.3310	-0.7588	0.2151	0.5067
y ₂	-0.2892	0.3123	-0.8176	-0.6815	-0.1704	-0.4652	-0.5286	-0.3338	-0.8623
y ₃	0.5326	-1.9912	0.2356	0.1280	-0.1916	0.5198	0.2939	0.2855	0.3727
y ₄	-0.4448	0.2735	-0.1779	-0.2009	-0.6403	0.0141	1.0252	0.2407	0.1699
y ₅	-0.4576	0.0492	-0.4264	0.5847	0.8790	0.2420	-0.3756	0.8609	-0.4667
y ₆	0.8270	0.8818	-0.6804	0.2211	-0.1429	-0.0176	-0.3118	-0.0395	0.6510
y ₇	0.5882	0.3447	0.5680	-0.0127	-0.2059	-0.0469	0.7035	-0.1486	-1.0523
y ₈	-0.0701	-0.3468	0.4085	-0.0066	0.2811	-1.1327	0.0232	-0.1340	0.5650
y ₉	0.1809	0.0668	-0.2956	-0.4409	0.7028	0.2595	0.2712	-0.3024	0.2267

Kanonik açıklama analizi sonucu elde edilen bağımlı ve bağımsız değişkenler setine ait kanonik değişkenler sırasıyla Tablo 8 ve Tablo 9'da verilmiştir. Aralarında en yüksek korelasyona sahip birinci kanonik değişken çifti (t_1, s_1) dikkate alınsın. t_1 kanonik faktörünün yorumuna kendi setine ait değişkenler arasında en yüksek katkıyı sağlayan değişkenlerin sırasıyla, x_8 (okulun bulunduğu çevrenin kültürel yapısına uyum derecesi), x_5 (anne ve babanın ayrı olup olmaması), x_9 (ailede geçimsizlik ya da huzursuzluk ortamının olup olmadığı) ve x_1 (babanın mesleği) değişkenleri olduğu görülmektedir. y setine ait en önemli faktör olan s_1 faktörü dikkate alındığında ise, bu doğrusal bileşenin yorumuna en yüksek katkıda bulunan değişkenlerin sırasıyla y_6 (yabancı dil), y_7 (edebiyat), y_3 (fen bilimleri) ve y_5 (tarih) dersleri olduğu görülmektedir.

Bu araştırmada yapılan her iki uygulama için kullanılan bilgisayar programları Ege Üniversitesi Bilgisayar Mühendisliği bölümündeki GENSTAT istatistiksel paket programından yararlanılarak hazırlanmıştır. Uygulamada kullanılan kanonik korelasyon analizi ve kanonik açıklama analizi programlarının hazırlanmasında Okur, M.C ve Ertaş K.(1990)'daki programlardan faydalanılmıştır.

5. SONUÇ

Analizler arasındaki farkların açık bir biçimde gösterilmesini amaçlayan uygulamada, Bornova Anadolu Lisesi öğrencilerinin psikolojik, sosyal ve ekonomik faktörlere dayalı olarak başarı düzeylerinin belirlenmesi amacıyla elde edilen korelasyon matrisi esas alınarak her iki analizin uygulaması yapılmıştır. Bu çalışmada, bağımlı ve bağımsız değişken setleri için olmak üzere toplam 18 değişken bakımından 150 öğrenci üzerinde ölçümler yapılmıştır.

Kanonik korelasyon analizi uygulamalarında gözlem sayısının çok olması analiz sonuçlarının güvenilirliğini arttırmaktadır. Kanonik korelasyonların önemliliği Wilk's Lambda, Hotelling-Lawley, Pillai Trace, Roy's maksimum kök gibi test istatistikleri ile hesaplanabilir.

Çalışmamızda çok değişkenli istatistiksel analiz yöntemlerinden birisi olan kanonik korelasyon analizi ve kanonik açıklama analizi yöntemlerinin GENSTAT paket programında çözümü ele alınmıştır. Anket verileri düzenlendikten sonra GENSTAT paket programında veriler girilmiştir. Kanonik korelasyon analizi ve kanonik açıklama analizi yöntemlerinin tüm aşamaları gerçekleştirildikten sonra sonuçlar alınmıştır.

Bazı uygulamalarda kanonik korelasyonların iki değişken seti arasındaki karşılıklı ilişkiler hakkında önemli fikirler vereceği açıktır. Özellikle aralarında yüksek ilişki bulunan vektörlere ihtiyaç duyulan araştırmalarda bu analizi uygulamak yerinde olacaktır. Bununla beraber uygulamalarda birkaç kanonik vektör çifti en güvenilir faktör çiftleri olarak değerlendirilmekte ve yorumlamalar esasta bu çiftlere dayandırılmaktadır. Bazı durumlarda bu yorumlamanın tamamen geçersiz olabileceği ortaya konulmaya çalışılmış ve sözü edilen yorumlama güçlüklerini ortadan kaldırmak amacıyla Wollenberg(1977) tarafından gerçekleştirilen ve elde edildikleri setleri temsil etme bakımından en güçlü faktörleri veren kanonik açıklama analizi incelenmiştir. Analizler arasındaki farkların açık bir şekilde gösterilmesini amaçlayan uygulamada korelasyon matrisi esas alınarak her iki analiz uygulanmış ve sonuçların bir değerlendirmesi verilmiştir.

KAYNAKLAR

- DESARBO, W.S. (1981), *Canonical/ Redundancy Factoring Analysis*. Psychometrica,46,307-329.
- HOTELLING, H. (1936), *Relations Between Two Sets of Variates*. Biometrika, 28,321-377.
- İŞÇİ, Ö.(1995), *Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerini Optimum Yapan Faktörlerin Bornova Anadolu Lisesi Öğrencilerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine Bir Araştırma*, DEÜ Sosyal Bilimler Enstitüsü, Yüksek lisans Tezi.
- OKUR, M. CUDİ, ERTAŞ, K.,(1990), *Kanonik Analizlerde Açıklanmış Varyans Oranlarının Maksimum Yapılması Üzerine Bir Araştırma*, Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt:5, Sayı:1-2,s:23-46.
- ÖZDAMAR, K.(1999), *Paket Programlar İle İstatistiksel Veri Analizi (Çok Değişkenli Analizler)-2*, Kaan Kitabevi, 385-427.
- STEWART, D.K., LOVE, V.A(1968), *A General Canonical Correlation Index*, Psychological Bulletin, 70,160-163.
- VAN DEN WOLLENBERG, A.L.(1977), *Redundancy Analysis:As an Alternative for Canonical Correlation Analysis*, Psychometrika, 42, 207-219.
- ŞEN, H., KALYONCU, C.(2001), *Eskişehir Osmangazi Üniversitesi Tıp Fakültesindeki Beslenme Bozukluğu (Malnütrisyon Sıklığı) İle İlgili Araştırmanın Kanonik Korelasyon Analizi İle Çözümlemesi*, Çukurova Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü V. Ulusal Ekonometri ve İstatistik Sempozyumu 19-22 Eylül 2001.

A Research on Using Canonical Correlation and Redundancy Analysis Statistics to Estimate Accomplishment Levels

ABSTRACT

If it is possible to collect examined variables into two sets as dependent and independent, canonical correlation analysis is one of the methods to search the relationship between these sets. In this study the relationship between Bornova Anatolian High School Students' average grades and psychological scores is examined. Comments are given by taking the students' average grades and calculating the canonical correlation and the redundancy statistics by using GENSTAT program. In the study the relationship among correlation analysis, redundancy analysis and the other multivariate statistical methods is shown and the application of Wollenberg's redundancy method on 9x9 correlation matrix is given.

Key Words : *Canonical correlation analysis, redundancy analysis, multivariate statistical methods.*