T.C. Dokuz Eylül Üniversitesi İktisadi ve İdari
Bilimler Fakültesi Dergisi, Cilt:5, Sayı:1-2, Yıl:1990

KANONİK ANALİZLERDE AÇIKLANMIŞ VARYANS ORANLARININ EN YÜKSEK YAPILMASI ÜZERİNE BİR ARAŞTIRMA

Mehmet Cudi OKUR (*)
Kadir ERTAŞ (**)

ÖZET

İncelenen değişkenleri bağımlı ve bağımsız olmak üzere iki set halinde toplamak mümkün ise, bu setler arasındaki karşılıklı ilişkileri araştırmak amacıyla kullanılan yöntemlerden birisi de kanonik korelasyon analizidir.Bu analizin açıklanmış varyans oranları üzerinde durmaması önemli yorumlama güçlükleri getirmektedir. Bu çalışmada, karşılaşılan bu güçlükleri çözmek amacıyla ileri sürülmüş yeni bir metod olan kanonik açıklama analizi tanıtılarak çeşitli yönlerden karşılaştırmalar yapılmaktadır. Metodların diğer bazı çok değişkenli istatistiksel yöntemlerle ilişkileri de ortaya konularak GENSTAT paket programı yardımıyla yazılan bir kanonik açıklama analizi programı tanıtılmaktadır.

1.GİRİŞ

Kanonik korelasyon analizinde yapısal korelasyonların (bir setin değişkenleri ile o setin faktörleri arasındaki korelasyonlar)zayıf olması durumunda, değişkenlerin faktörler tarafından açıklamış varyans oranları da oldukça düşük gerçekleşecektir. Buna karşılık bu iki önemsiz faktör arasında yüksek bir kanonik korelasyon bulunabilir. Örneğin, λ kanonik korelasyonu oldukça yüksek olmasına rağmen, $\underline{v}'\ \underline{y}$ kanonik faktörü y setindeki varyasyonun çok küçük bir kısmını açıklıyor olabilir. Aynı durum $\underline{w}'\ \underline{x}$ faktörü için de geçerli olabilir. Gerçekte bu güçlükler kanonik korelasyon analizinin yapısal korelasyonlar üzerinde durmamasından kaynaklanmaktadır ve faktörlerin yorumlanmasında güçlüklere neden olmaktadır (Wollenberg, 1977). Halbuki en yüksek korelasyonlu olarak elde edilen bu faktörlere bir setteki değişkenlerin ne derece katkıda bulunduğunun bilinmesi gerçekten önemlidir. Bu da yapısal korelasyonların belirlenmesini gerektirir.

(*) Prof. Dr., E.Ü.M.B.F. Bilgisayar Mühendisliği Bölümü

(**) Yrd.Doç. Dr., D.E.Ü.,İ.İ.B.F. Ekonometri Bölümü

2. AÇIKLANMIŞ VARYANS ORANLARININ BELİRLENMESİ

2.1. Yapısal Korelasyonlar ve Açıklama İndeksi

$\underline{x}$ setinin değişkenleri ile bu setin i. kanonik faktörü arasındaki yapısal korelasyonları (faktör yüklerini) içeren h_i (p*1) sütun vektörü,

$$\underline{h}_i = \mathrm{Cov}(\underline{x}, a_i) = R_{11}\,\underline{w}_i \qquad (1)$$

olarak elde edilir. Benzer olarak, $\underline{y}$ seti değişkenleri ile bu setin i.kanonik faktörü arasındaki yapısal korelasyonları içeren $\underline{g}_i$(q*1) sütun vektörü ,

$$\underline{g}_i = \mathrm{Cov}(\underline{y}, b_i) = R_{22}\underline{v}_i \qquad (2)$$

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

$$R_{\underline{x}_i} = \lambda_{1i}(\underline{h}'_i\,\underline{h}_i)/p \qquad (3)$$

$$R_{\underline{y}_i} = \lambda_{1i}(\underline{g}'_i\,\underline{g}_i)/q \qquad (4)$$

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, $\underline{x}$ ve $\underline{y}$ setlerinin toplam açıklanmış varyans oranları sırasıyla,

$$R_{d1} = \sum_i R_{\underline{x}_i} \qquad (5)$$

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

$$R_{d2} = \sum_i R_{\underline{y}_i} \qquad (6)$$

olarak tanımlanır.

(3) ifadesinin birinci kısmını oluşturan λ_{1i} ,bilindiği gibi i.kanonik korelasyonun karesidir İkinci kısmın payında yer alan $\underline{h}'_i \underline{h}_i$ ise,

$$[Cov(\underline{x}, a_i)]' [Cov(\underline{x}, a_i)] = \underline{w}'_i R_{11} R_{11} \underline{w}_i \qquad (7)$$

olduğundan, $\underline{x}$ seti değişkenleri ile yine $\underline{x}$ setinin i. kanonik faktörü arasındaki korelasyonların kareleri toplamıdır. Diğer taraftan, λ^2_{1i} ifadesi-bir setin i. kanonik faktörünün diğer setin karşılık gelen kanonik faktörü tarafından açıklanmış varyans oranıdır. Bu anlamda, $\underline{w}'_i R_{11} R_{11} \underline{w}_i$ ifadesinin de $\underline{x}$ seti değişkenlerinin yine $\underline{x}$ setinin i. kanonik faktörü tarafından açıklanan varyans oranı olduğu ifade edilebilir. Böylece, açıklama indeksi formülünün iki ayrı açıklanmış varyans ifadesinden meydana geldiği görülüyor. Kanonik korelasyon analizinin, yalnızca birinci kısmı oluşturan λ^2_{1i} açıklanmış varyansını maksimum yaptığına dikkat edilmelidir. Bilindiği gibi kanonik korelasyon analizinde λ_{1i} ve λ_{2i} kanonik korelasyonları eşit idiler. Açıklama indeksinin farklı bir yönü de simetrik olmayışıdır. Örneğin, λ_1 kanonik korelasyonu verildiğinde elde edilicek $R_{\underline{x}_i}$ ve $R_{\underline{y}_i}$ açıklama indeksleri çoğunlukla eşit olmayacaklardır. Bu durumun nedenini anlamak gerçekten önemlidir. Aslında her iki ifadenin de birinci kısımları eşittir Farklılık ikinci kısımdan doğmaktadır. $\underline{x}$ setinin a_1 faktörü tarafından açıklanan varyansı ile $\underline{y}$ setinin b_1 faktörü tarafından açıklanan varyansı aynı değildir.Örneğin; b_1, $\underline{y}$ setinin sonuncu ana bileşenine oldukça yakın bir faktör yani bu setin önemsiz bir faktörü olduğu halde, a_1, $\underline{x}$ setinin birinci anabileşenine oldukça yakın bir faktör yani bu setin önemli bir faktörü olabilir. Böyle bir durumda $\underline{x}$ seti değişkenlerinin a_1 faktörü tarafından açıklanan varyansı, $\underline{y}$ seti değişkenlerinin b_1 faktörü tarafından açıklanan vaaryansından büyük olacak ve $R_{\underline{x}_i} > R_{\underline{y}_i}$ gerçekleşecektir. Diğer bir ifade ile, a_1 faktörü tahminleyici değişkenler setindeki varyasyonun önemli bir kısmı açıklıyorsa, yüksek bir $R_{\underline{x}_i}$ oranı ile karşılaşılacaktır.

(3) nolu ifade daha açık olarak,

$$R_{\underline{x}_i} = \frac{1}{p} \lambda_{1i} \underline{w}'_i R_{11} \underline{w}_i \lambda_{1i} \qquad (8)$$

şeklinde yazılabilir. Diğer taraftan,

$$R_{12} \underline{v}_i = \lambda_{1i} R_{11} \underline{w}_i \qquad (9)$$

eşitliği de geçerlidir. Bu son ifade dikkate alınarak (8) eşitliği,

$$R_{\underline{x}_i} = \frac{1}{p} \underline{v}'_i R_{21} R_{12} \underline{v}_i \qquad (10)$$

şeklinde gösterilebilir. Benzer şekilde,

$$R_{\underline{y}_i} = \frac{1}{p} \underline{w}'_i R_{12} R_{21} \underline{w}_i \qquad (11)$$

yazılabilir. Diğer taraftan,

$$\mathrm{Cov}(\underline{x}, b_i) = \mathrm{Cov}(\underline{x}, \underline{v}'_i \underline{y}) = R_{12} \underline{v}_i \qquad (12)$$

ve benzer şekilde,

$$\mathrm{Cov}(\underline{y}, a_i) = \mathrm{Cov}(\underline{y}, \underline{w}'_i \underline{x}) = R_{21} \underline{w}_i \qquad (13)$$

y a z ı l a b i l i r .

(12) ifadesine göre, $R_{12}\underline{v}_i$(p*1) sütun vektörünün elemanları.$\underline{x}$ s ti değişkenleri ile $\underline{y}$ setinin i. kanonik faktörü arasındaki korelasyonlara karşılık gelmektedir. Buna göre,

$$[Cov\ (\underline{x}, b_i)]\ '\ [Cov\ (\underline{x}, b_i)] = \underline{v}'_i\ R_{21}\ R_{12}\ \underline{v}_i \qquad (14)$$

skalar ifadesi bu korelasyonların kareleri toplamı olacaktır. Böylece 10 nolu ifade ile verilen $R_{\underline{x}_i}$ ifdesinin payı; $\underline{x}$ seti değişkenlerinin,$\underline{y}$ setinin i.kanonik faktörü tarafından açıklanmış varyansıdır. Benzer bir yorumlama $R_{\underline{y}_i}$ açıklama indeksi için de yapılabilir.Bu durumda açıklama indeksinin, bir setin değişkenlerinin diğer setin verilen bir kanonik faktörü tarafından açıklanan varyans oranı olduğu (10) ve (11) ifadelerinde görülebilmektedir.

2.2. Açıklama İndeksinin Maksimum Yapılması Sorunu

Bir önceki bölümde yapılan açıklamalar dikkate alındığında,açıklama indeksini maksimum yapan faktörler bir anlamda tahminleme gücü optimum olan fattörlerdir. Bu nedenle, daha çok tahminleme amacına dayanan araştırmalarda bu indeksi maksimum yapan faktörlerin kullanılması kanonik faktörlerin kullanılmasından daha uygundur. Örneğin bir pazarlama yöneticisi tüketicinin marka tercihlerini (y), bireysel psikolojik skorlarla (x) tahminleyebilmek için, $\underline{y}$ seti değişkenlerinin, $\underline{w}'\underline{x}$ faktörü tarafından açıklanan varyansını maksimum yapacak bir $\underline{w}$ vektörü araştıracaktır (De Sarbo, 1981).

Açıklama indeksini maksimum yapan yöntem ilk olarak Wollenberg (1977) tarafından geliştirilmiştir. Yöntemin esas amacı, $R_{\underline{y}}$ indeksini maksimum yapa. birim varyanslı $a = \underline{w}'\ \underline{x}$ ve $R_{\underline{x}}$ indeksini maksimum yapan, birim varyanslı $b=\underline{v}'\underline{y}$ doğrusal bileşenlerinin belirlenmesidir. Burada a ve b değişkenleri 'en yüksek varyans açıklayıcı değişkenler' olarak tanımlanır. Bu işi yapan analiz ise 'kanonik açıklama analizi' olarak adlandırılır. (10) ve (11).ifadeleri dikkate alındığında, $R_{\underline{x}}$ ve $R_{\underline{y}}$ indekslerini maksimum yapmak için, bu ifadelerin payındaki açıklanmış varyansları maksimum yapmak yeterlidir. R_y yi maksimum yapmak için $\underline{w}'R_{12}R_{21}\underline{w}$ ifadesi,

$$Var\ (a) = Var\ (\underline{w}\ '\underline{x}) = \underline{w}'\ R_{11}\ \underline{w} = 1$$

birim varyans kısıtlaması altında maksimum yapılacaktır. Kısıt için

bir çarpan kullanarak lagrange çarpanları yöntemiyle amaç fonksiyonu,

$$z_1 = \underline{w}' R_{12} R_{21} \underline{w} - \mu_1[\underline{w}' R_{11} \underline{w} - 1] \qquad (15)$$

şeklinde yazılabilir. $\underline{w}$ ya göre kısmi türev alınıp sıfıra eşitlenirse,

$$\frac{\partial z_1}{\partial \underline{w}} = 2R_{12} R_{21} \underline{w} = 2\mu_1 R_{11} \underline{w} = \underline{0} \qquad (16)$$

Buradan da,

$$R_{12} R_{21} \underline{w} - \mu_1 R_{11} \underline{w} = \underline{0} \qquad (17)$$

elde edilir. Eşitlik düzenlenirse,

$$(R^{-1}_{11} R_{12} R_{21} - \mu_1 1)\underline{w} = \underline{0} \qquad (18)$$

sonucuna varılır. Bu homojen eşitliğin $\underline{w}=\underline{0}$ dan farklı çözümünün olabilmesi için μ_1 in $R^{-1}_{11}R_{12}R_{21}$ matrisinin bir özdeğeri, $\underline{w}$ nun da bu özdeğere kaşılık gelen özvektör olması gerekir. $R_{11}^{-1}R_{12}R_{21}$ matrisinin tam ranklı olduğu ve özdeğerlerinin birbirinden farklı oldukları varsayılarak,

$$\mu_{11} > \mu_{12} > \cdots > \mu_{1p} > 0 \qquad (19)$$

yazılabilecektir. (17) ifadesi gözönüne alınırsa,

$$\mu_1 R_{11} \underline{w} = R_{12} R_{21} \underline{w} \qquad (20)$$

yazılabilir. Bu ifade soldan $\underline{w}'$ ile çarpılarak,

$$\mu_1 \underline{w}' R_{11} \underline{w} = \underline{w}' R_{12} R_{21} \underline{w} \qquad (21)$$

elde edilir. Yukarıda verilen birim varyans kısıtlaması gereği,

$$\mu_1 = \underline{w}' R_{12} R_{21} \underline{w} \qquad (22)$$

bulunur. Bu sonuca göre μ_{1i} özdeğeri (aynı zamanda lagrange çarpanı), $\underline{y}$ seti değişkenlerinin $\underline{x}$ setinin i. en yüksek varyans açıklayıcı değişkeni tarafından açıklanan varyansına eşittir. Diğer bir ifade ile, μ_{1i} özdeğeri, $\underline{y}$ setinin i. açıklama indeksi $R_{\underline{y}_i}$ nin q katı olarak yorumlanabilir. Böylece $\underline{y}$ setinin toplam açıklama indeksini gösteren (6) ifadesi,

$$R_{d2} = \frac{1}{q} \sum_i \mu_{1i} \qquad (23)$$

şeklinde de yazılabilecektir.

Diğer taraftan $R_{\underline{X}}$ indeksini maksimum yapmak için, $\underline{v}' R_{21} R_{12} \underline{v}$ ifadesi,

$$\text{Var (b)} = \text{Var} (\underline{v}' \underline{y}) = \underline{v}' R_{22} \underline{v} = 1 \qquad (24)$$

kısıtlaması altında maksimum yapılacaktır. Kısıt için bir çarpan kullanarak lagrange çarpanları yöntemiyle yukarıdaki yol izlenerek,

$$\mu_2 = \underline{v}' R_{21} R_{12} \underline{v} \qquad (25)$$

elde edilir. Bu son ifa leye göre μ_{2i} özdeğeri $\underline{x}$ seti değişkenlerinin $\underline{y}$ setinin en yüksek varyans açıklayıcı değişkeni tarafından açıklanan varyansıdır. Diğer bir ifade ile μ_{2i} özdeğeri, $\underline{x}$ setinin açıklama indeksi $R_{\underline{x}_i}$ nin p katı olarak yorumlanabilir. Böylece (5) ifadesi ile verilen $\underline{x}$ setinin toplam açıklama indeksi,

$$R_{d1} = \frac{1}{p} \sum_i \mu_{2i} \qquad i = 1,2, \ldots, q \qquad (26)$$

şeklinde ifade edilebilecektir.

Kanonik korelasyon analizinde olduğu gibi, kanonik açıklama analizinde de aynı değişken setinden elde edilecek doğrusal bileşenler (en yüksek varyans açıklayıcı değişkenler) arasında korelasyon yoktur ve bu aranılan bir özellikti'

2.3. Diğer Çok Değişkenli Yöntemlerle İlişkiler

Daha önce verilen tanımdan farklı olarak sadece bir tane y değişkeni olduğu varsayılsın. Bu durumda X (N*p) yine aynı N gözlemden elde edilmiş veri matrisi , $\underline{B}$(p*1) populasyon kısmi regresyon katsayıları vektörü, $\underline{v}$(n*1) hata vektörü olmak üzere çoklu regresyon modeli,

$$\underline{y} = X\,\underline{B} + \underline{v} \qquad (27)$$

şeklinde ifade edilebilir. Regresyon denklemleri, bağımsız değişkenlerin bağımlı değişkeni tahmin bakımından optimum kombinasyonlarıdır. Çoklu korelasyon $R_{y.x}$, elde edilen bu optimum kombinasyon ile bağımlı değişken arasındaki ilişkinin bir ölçüsüdür. Çoklu korelasyonun karesi $R^2_{y.x}$ ise bağımsız değişkenler tarafından açıklanabilen varyasyonun genel varyasyona oranıdır.Bu durumda ilk p değişken bağımsız, (p+1). değişken bağımlı olmak üzere örnek korelasyon matrisi R,

$$R = \left[\begin{array}{c:c} R_{11} & R_{12} \\ \hdashline R_{21} & 1 \end{array}\right] \qquad (28)$$

şeklinde parça'anacak ve kısmi regresyon katsayılarının tahmini,

$$\underline{\beta} = (X'X)^{-1}X'\underline{Y}$$

$$= R^{-1}_{11}R_{12} \qquad (29)$$

olacaktır. Çoklu korelasyon katsayısının ve karesinin tahminleri

$$R_{y.x} = \rho\,(Y, X\,\underline{\beta}) = (R_{21}R^{-1}_{11}R_{12})^{1/2} \qquad (30)$$

ve, $R^2_{y.x} = R_{21}R^{-1}_{11}R_{12}$ (31)

şeklinde ifade edilecektir.

Kanonik açıklama analizinde $R^{-1}_{11}R_{12}R_{21}$ matris çarpımı dıkkatle incelenirse, çoklu korelasyon analizi ile bazı ilişkiler olduğu görülebilir. $\underline{y}$ setinin yalnızca tek bir değişkenden meydana geldiğinin varsay lc ğı durumda $R^{-1}_{11}R_{12}(p*1)$ sütun vektörü bu değişkenin çoklu regresyonunda$\underline{\beta}$ ağırlıklar (regresyon katsayıları) vektörüdür. Kanonik açıklama analizinde ise, $R^2_{y.\underline{x}} = \underline{\beta}^T R_{12}$ işlemi yapılarak doğrudan doğruya çoklu korelasyonun karesi elde edilmemekte, $\underline{\beta} R_{21} = R^{-1}_{11}R_{12}R_{21}$ işlemi yap 'a ak bir matris elde edilmektedir. Bu matriste her bir köşegen eleman, y değişkeninin $\underline{x}$ değişkenleri ile çoklu regresyonunda karşılık gelen bağımsız değişkenin $\underline{\beta}_i$ ağırlığı ve bu bağımsız değişken ile y arasındaki korelasyonun çarpımına eşittir. Bu durumda matrisin izi, çoklu regresyonun karesi olan $R^2_{y.\underline{x}}$ ye karşılık gelecektir. Bu ise, $\underline{y}$ değişkenin x seti değişkenleri tarafından açıklanabilmiş varyans oranıdır ve (21) ifadesinde görüldüğü gibi $R^{-1}_{11}R_{12}R_{21}$ matrisinin tek özdeğeri olan μ ye eşit olacaktır.Bu durumda μ aynı zamanda y değişkenin açıklama indeksine karşılık gelecektir. Tek bir y değişkeni için, $R_{12} = r_{12} = r_{21}$ notasyonu kullanılarak, (18) ifadesi ile verilen homojen eşitlik,

$$(R^{-1}_{11}r_{12}r_{21} - \mu I)\underline{w} = \underline{0} \qquad (32)$$

veya

$$(\underline{\beta} r_{21} - \mu I)\underline{w} = \underline{0} \qquad (33)$$

şeklinde yazılabilir. μ yerine $R^2_{y.\underline{x}}$, $\underline{w}$ yerine $\underline{\beta}$ konulduğunda,

$$(\underline{\beta} r_{21} \beta - \mu \underline{\beta}) = \underline{0} \qquad (34)$$

$$\underline{\beta}(r_{21}R^{-1}_{11}r_{12} - \underline{\beta} R^2_{y.x}) = \underline{0} \qquad (35)$$

bulunur. (31) ifadesi dikkate alınırsa,

$$\underline{\beta}(R^2_{y.x} - R^2_{y.x}) = \underline{0} \qquad (36)$$

yazılabilir. Son ifade homojen eşitliğin aynen korunduğunu

göstermektedir. Böylece tek bir y değişkeni durumunda μ nün çoklu korelasyonun karesine, $\underline{w}$ nun da $\underline{\beta}$ ağırlıklar vektörüne karşılık geldiği ifade edilebilir. Bu açıklamalar 1'den fazla y değişkenine genellenebilir. Bu durumda yukarıdaki matrisin j. köşegen elemanı, j. x değişkenin y değişkenleri için tanımlanacak çoklu korelasyona katkısını ifade edecektir.

Değişkenler arasında bağımlı bağımsız ayırımı olmadığında uygulanan çok değişkenli yöntemlerden biri de, çok boyutlu uzayda ortogonal bir döndürme işlemi olan anabileşenler analizidir. Kanonik açıklama analizinin uygulandığı değişkenlerin aynı sete ait oldukları varsayılırsa (18) ifadesi,

$$(R^{-1}_{11}R_{11}R_{11} \quad \mu\, I)\,\underline{w} = \underline{0} \qquad (37)$$

veya

$$(R_{11} - \mu\, I)\,\underline{w} = \underline{0} \qquad (38)$$

şeklinde gösterilebilir. Bu ise anabileşenler analizinin özdeğer eşitliğidir.

Benzer bir durum kanonik korelasyon analizi için düşünüldüğünde, bu analizin özdeğer eşitliği,

$$(R^{-1}_{11}\,R_{11}\,R^{-1}_{11}\,R_{11} - \lambda^2_1, I)\,\underline{w} = \underline{0} \qquad (39)$$

veya

$$(I - \lambda^2_1\ , I\ )\underline{w} = \underline{0} \qquad (40)$$

şeklini alacaktır ve kanonik değişkenler bu yapıdan elde edileceklerdir. Bu durumda bütün kanonik değişken çiftlerinin tam korelasyonlu olacakları açıktır. Bu ise analizin amacına uygun değildir. Bu nedenle anabileşenler analizi kanonik açıklama analizinin özel bir durumu olarak yorumlanabiliceği halde, kanonik korelasyon analizi için böyle bir durum söz konusu değildir.

2.4. Karşılıklı Ortogonal Değişkenlerin Elde Edilmesi

İki ayrı setten elde edilen en yüksek varyans açıklayıcı değişkenlerin p+q boyutlu korelasyon matrisi,

$$Q = \left[\begin{array}{c|c} Q_{11} & Q_{12} \\ \hline Q_{21} & Q_{22} \end{array}\right] \qquad (41)$$

şeklinde

şeklinde parçalandığında, aynı setten elde edilen doğrusal bileşenler kendi aralarında korelasyonsuz olduklarından değişkenlerin set içi korelasyon matrisleri için,

$$Q_{11} = I \text{ ve } Q_{22} = I \qquad (42)$$

ifadeleri yazılabilecektir.

Kanonikkorelasyon analizinden farklı olarak, kanonik açıklama analizinde $Cov(a_i , b_j) = 0$ durumu sağlanamadığı için $Q_{12} = Q'_{21}$ setler arası korelasyon matrisi köşegen matris değildir. Halbuki bazı uygulamalarda, en yüksek varyans açıklayıcı değişkenlerin karşılıklı ortogonal olmaları aranılan bir özelliktir. Bu özellik en yüksek varyans açıklayıcı değişkenler üzerinde kanonik korelasyon analizi uygulaması yapılarak gerçekleştirilebilir. Kanonik korelasyon analizinin özdeğer eşitliklerine benzer şekilde burada,

$$(Q^{-1}_{11} Q_{12} Q^{-1}_{22} Q_{21} - c^2 I)\, \underline{t} = \underline{0} \qquad (43)$$

ve

$$(Q^{-1}_{22} Q_{21} Q^{-1}_{11} Q_{12} - d^2 I)\, \underline{s} = \underline{0} \qquad (44)$$

Özdeğer eşitlikleri yazılabilir. Q_{11} ve Q_{22} birim matris olduklarından bu eşitlikler,

$$(Q_{12} Q_{21} - c^2 I)\, \underline{t} = \underline{0} \qquad (45)$$

ve

$$(Q_{21} Q_{12} - d^2 I)\, \underline{s} = \underline{0} \qquad (46)$$

şeklinde yazılabilir. Buradan $\underline{t}$ ve $\underline{s}$ vektörleri için,

$$\underline{t} = \frac{1}{c} Q_{12}\, \underline{s} \qquad (47)$$

$$\underline{s} = \frac{1}{c} Q_{21}\, \underline{t}$$

şeklinde yazılabilir. (45) ve (46) nolu ifadeler ile verilen eşitliklerin birisinin çözümü problemin çözümü için yeterlidir. Bu yolla elde edilecek yeni en yüksek varyans açıklayıcı değişkenler için Cov (m_i , n_i) = 0 durumu geçerli olacaktır. Böylece T ve S, kolonları sırasıyla $\underline{t}_i$ ve $\underline{s}_i$ lerden meydana gelen ortogonal rotasyon matrisleri olmak üzere yeni değişkenlerin setler arası korelasyon matrisi olan,

$$R(\underline{m}, \underline{n}) = T' Q_{12} S \qquad (48)$$

matrisi köşegen matris olacaktır. Ayrıca A, elemanları c_i ler (kanonik korelasyonlar) olan köşegen matris olmak üzere,

$$S = A^{-1} Q_{21} T \qquad (49)$$

ilişkisi yazılabilir. Burada T ve S matrisleri, elemanları yeni ve eski en yüksek varyans açıklayıcı değişkenler arasındaki kosinüs ve sinüslere karşılık gelen ortogonal rotasyon matrisleri olarak yorumlanabilir.

3. UYGULAMA

Bu bölümde, kanonik korelasyon analizi ve kanonik açıklama analizini daha önce üzerinde durulan istatistiksel özellikler bakımından karşılaştırabilmek amacıyla, belirli bir korelasyon matrisi esas alınarak her iki analizin uygulaması yapılmıştır. Tablo 1.de verilen korelasyon matrisi, ilk dört değişken bağımsız değişkenler setini, diğer değişkenler bağımlı değişkenler setini meydana getirecek şekilde parçalanmıştır. Kanonik korelasyon analızının $\underline{x}$ ve $\underline{y}$ setleri ile ilgili sonuçları sırasıyla Tablo. 2 ve 3 de, faktör yükleri Tablo. 4 de, kanonik değişkenlerin setlerarası korelasyon matrisi Tablo. 5 de verilmiştir.

Tablo.1: Orijinal değişkenlerin korelasyon matrisi

1.000							
.160	1.000						
.900	.14(	1.000					
.140	.900	.160	1.000				
.776	.146	.682	.140	1.000			
.054	.77(	-.057	.680	.160	1.000		
.696	.227	.775	.232	.900	.139	1.000	
.030	.698	-.032	.774	.140	.889	.156	1.000

Tablo.2: $\underline{x}$ setine ait kanonik korelasyon analizi sonuçları

Kök no	λ_i	λ_i^2	H_i	$R_{\underline{x}_i} = \lambda_i^2 H_i$	$R_{\underline{x}_i} / R_{dl}$
1	.9005	.8109	.0363	.0294	.0464
2	.8332	.6942	.2255	.1565	.2475
3	.8088	.6542	.2050	.1341	.2191
4	.7652	.5855	.5333	.3122	.4938

Tablolarda yer alan $H_i=\underline{h}_i'\underline{h}_i/p$ değeri a_i faktörlerinin $\underline{x}$ setinde açıkladıkları varyans oranını, $G_i=\underline{g}_i'\underline{g}_i/q$ değeri b_i faktörlerinin $\underline{y}$ setinde açıkladıkları varyans oranını ifade etmektedir. λ_i ise i. kanonik korelasyondur.

Tablo.3 : y setine ait kanonik korelasyon analizi sonuçları

Kök no	λ_i	λ_i^2	G_i	$R_{y_i}=\lambda_i^2 G_i$	R_{y_i}/R_{d2}
1	.9005	.8109	.0301	.0244	.0386
2	.8332	.6942	.2430	.1687	.2669
3	.8088	.6542	.1932	.1264	.2000
4	.7652	.5855	.5338	.3125	.4945

Tablo.2 ve 3 incelendiğinde, $\underline{x}$ ve $\underline{y}$ setlerinin aralarında .9005 lik kanonik korelasyona sahip birinci faktörlerinin kendi setlerindeki varyasyonun sırasıyla ancak 0363 ve .0301 lik kısımlarını açıklayabildikleri görülmektedir. Bu, birinci kanonik faktör çiftinin açıklanan varyans anlamında setlerinin en önemsiz faktörleri olduklarını göstermesi bakımından oldukça anlamlıdır, Her iki setin aralarında en zayıf kanonik ilişkiye sahip sonuncu faktörleri ise, elde edildikleri setlerdeki varyasyonun yaklaşık % 53,5 lik kısmını açıklamaktadırlar. Bu ise, sonuncu kanonik faktör çiftinin, elde edildikleri setleri temsil gücü bakımından en önemli faktörler olabileceklerini göstermesi açısından dikkat çekicidir. $\underline{x}$ setinin birinci faktörü, $\underline{y}$ setindeki varyasyonun ancak 0.0244 ünü açıklayabilmektedir ve bu değer $\underline{y}$ setinde açıklanabilecek varyans oranının ancak 0.0386 sına karşılık gelir $\underline{y}$ setinin birinci faktörü $\underline{x}$ setindeki varyasyonun ancak 0.0294 ünü açıklayabilmektedir ve bu değer de $\underline{x}$ setinde açıklanması mümkün olan varyans oranının yalnızca 0.0465 idir. Bu durum ise, ilk kanonik faktör çiftinin karşı setlerdeki varyasyonu açıklama gücü bakımından oldukça zayıf olduğunu göstermektedir. Sonuncu faktör çiftinin ise bu anlamda en yüksek güce sahip olduğu görülüyor. Tablo.4 de yer alan ilk faktörler üzerindeki yüklerin küçük, buna karşılık son faktörler üzerindeki yüklerin yüksek olması bu sonuçların esas nedeni olarak yorumlanabilir.

Tablo.4: Kanonik korelasyon analizinde faktör yükleri

	a1	a2	a3	a4	b1	b2	b3	b4
x1	-.1237	.2725	-.4085	-.8623	-.1114	.2271	-.3304	-.6598
x2	.1711	-.7108	.3359	-.5939	.1541	-.5923	.2716	-.4544
x3	.2946	.3949	-.3249	-.8073	.2653	.3290	-.2627	-.6177
x4	.1170	-.4080	.6594	-.6205	.1054	-.3400	.5333	-.4847
y1	-.1363	.1987	-.2835	-.6833	-.1514	.2385	-.3505	-.8930
y2	-.0767	-.6565	.2730	-.3885	-.0852	-.7881	.3375	-.5077
y3	.2551	.2161	-.2001	-.6806	.2833	.2594	-.2474	-.8895
y4	-.8095	-.3966	.5572	-.4113	-.0993	-.4760	.6890	-.5375

Tablo.5 : Kanonik değişkenlerin setlerarası korelasyon matrisi

$$Q_{12}= \begin{bmatrix} .9005 & & & \\ .0000 & .8332 & & \\ .0000 & .0000 & .8088 & \\ .0000 & .0000 & .0000 & .7652 \end{bmatrix}$$

Kanonik açıklama analizinin sonuçları Tablo.6,7,8 ve 9 da verilmiştir.

Tablo.6: $\underline{x}$ setine ait kanonik açıklama analizi sonuçları

Faktör no	H_i	μ_{2i}	$R_{\underline{x}_i} = \mu_{2i}/p$	$R_{\underline{x}_i}/R_{dl}$
1	.5137	1.3114	.3279	.5185
2	.4092	1.0711	.2678	.4236
3	.0494	.0868	.0217	.0343
4	.0277	.0598	.0149	.0236
	1.0000		.6323	

Tablo.7: $\underline{y}$ setine ait kanonik açıklama analizi sonuçları

Faktör no	G_i	μ_{1i}	$R_{\underline{y}_i}=\mu_{1i}/p$	$R_{\underline{y}_i}/R_{d2}$
1	.5255	1.3118	.3279	.5188
2	.3912	1.0614	.2654	.4199
3	.0494	.0892	.0223	.0353
4	.0340	.0654	.0164	.0259
	1.0000		.6320	

Tablo. 6 ve 7 deki açıklanan varyans oranları incelenirse, kanonik açıklama analizinin faktörleri elde edildikleri setleri temsil etme gücü bakımından optimum bir şekilde sıraladığı görülmektedir. Kanonik korelasyon analizi sonuçlarıyla karşılaştırma yapılırsa, kanonik açıklama analizinde $\underline{x}$ setinin ilk iki faktörü $\underline{y}$ setinde açıklanabilecek toplam varyasyonun .9387 sini açıkladığı halde bu değer kanonik korelasyon analizinde .3055 düzeyinde kalmaktadır. $\underline{y}$ setinin ilk iki en yüksek varyans açıklayıcı faktörü $\underline{x}$ setinde açıklanabilecek toplam varyasyonun .9421 ini açıkladığı halde, bu değer kanonik korelasyon analizinde .2490 da kalmaktadır. Bu durum, kanonik korelasyon analizinin aksine, kanonik açıklama analizinde her iki sete ait tahminleme gücünün setlerin ilk iki faktöründe toplandığını göstermektedir. Örneğin her iki analizde setlerin sonuncu faktörleri değerlendirme dışı bırakılırsa analizler arasındaki farklar kolayca görülebilir. Tablo. 9 da verilen yükler incelenirse, ilk sıradaki faktörlerden sonrakilere doğru gittikçe yüklerin azaldığı görülür. Bu ise yukarıda açıklanan sonuçların nedeni olarak yorumlanabilir.

Tablo. 8 : En yüksek varyans açıklayıcı değişkenlerin karşılıklı or

togonal olmadıklarını gösteren setler arası korelasyon matrisi

$$Q_{12}= \begin{bmatrix} -.7408 & -.1500 & .1613 & .1012 \\ -.1372 & .7911 & -.0999 & -.0619 \\ -.1477 & -.1038 & -.8394 & .0251 \\ -.1145 & .0858 & .0259 & .8485 \end{bmatrix}$$

Tablo.9 : Kanonik Açıklama analizinde faktör yükleri

	a1	a2	a3	a4	b1	b2	b3	b4
x1	-.5529	-.8035	.1990	-.0954	.4795	-.5816	-.1774	-.0922
x2	-.8827	.3552	-.0705	.2996	.6498	.4464	-.1109	.1447
x3	-.4588	-.8628	-.1907	.0937	.4971	-.5859	.1747	.0913
x4	-.8714	.3474	-.3418	-.0557	.6419	.4363	.1102	-.1483
y1	-.4569	-.5966	.1672	-.1121	.6832	-.6765	-.2312	-.1489
y2	-.6752	.4123	.1280	.1421	.7155	.6023	-.3534	.0209
y3	-.4516	-.6096	-.1651	.1111	.7743	-.6127	.1338	.0843
y4	-.6657	.4049	-.1326	-.1426	.7236	.6074	-.0348	-.3261

Uygulamada kullanılan kanonik korelasyon analizi ve kanonik açıklama analizi progr mları GENSTAT paket programı yardımıyla yapılmış ve sırasıyla Ek-I ve Ek-II de verilmiştir.

IV.SONUÇ VE TARTIŞMA

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 faktör çifti en güvenilir faktör çiftleri olarak değerlendirilmekte ve yorumlamalar esasta bu çiftlere dayandırılmaktadır.

Bu araştırmada, bazı durumlarda bu yorumlamanın tamamen geçersiz olabileceği teorik yönleriyle 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 ayrıntılı olarak incelenmiştir. Ayrıca elde edilen en yüksek varyans açıklayıcı faktörler üzerinde tekrar kanonik korelasyon analizi uygulanarak karşılıklı ortogonal faktörler elde edilmiştir. Analizler arasındaki farkların açık bir şekilde gösterilmesini amaçlayan uygulamada belirli yapıdaki bir korelasyon matrisi esas alınarak her iki analiz uygulanmış ve sonuçların bir değerlendirilmesi Bölüm.2 de verilmiştir.

A STUDY OF CANONICAL REDUNDANCY ANALYSIS

In bivariate and multiple correlation analyses, a squared correlation coefficent can be inte preted as the propotion of criterion variance explained by the optimal linear combination of predictors. But this is not the case for the canonical correlation since it is the maximum correlation bet-

ween linear functions (canonical factors) of the two sets of variables. As canonical correlation does not take account of structural correlations, a relatively large canonical correlation may be obtained between the two linear functions, while the variance of variables accounted for by these functions is very low. This can lead the researcher to difficulties in interpretation.

To deal with these interpretation problems, redundancy analysis can be used. In this method, linear functions are determined in such a way that Stewart and Love's redundancy index is maximized.

In this study, redundancy analysis and its relatonships with some other multivariate statistical methods have been presented in detail. Wollenberg's rotational method has been introduced and an application on an 8x8 correlation matrix has been provided. A computer program for performing redundancy analysis by means of GENSTAT statistical package has also been given.

KAYNAKÇA

Anderson, T.W. (1958); An Introduction to Multivariate Statistical Analysis, John Wiley, New York.

Barnett, V.D., Lewis, T. (1963); A Study of the Relation between G.C.E. and Degree Results. J. Roy. Statist. Soc. A, 126, 187-216

Cooley, W.W., Lohnes, P.R. (1971); Multivariate Data Analysis.John Wiley, Canada.

DeSarbo, W.S. (1981); Canonical/ Redundancy Factoring Analysis.Psychometrika, 46, 307-329.

Green, P.E (1978); Analyzing Multivariate Data. Hinsdale, III.Hold, Rinehard and Winston.

Hotelling, H. (1936); Relations between Two Sets of Variates. Biometrika, 28, 321-377.

Kendall, M.G., Stuart, A. (1973); The Advanced Theory of Statis tics. Vol. 3. Charles Griifin and Co. Ltd., London.

Mardia, K.V., Kert, J.T., Bibby, J.M. (1979); Multivariate Analysis Academic Press Inc. London.

Okur, M.C. (1979); Jacknife Yönteminin Çoklu Korelasyon Tahminlerinde Kullanılabilirliği Üzerine Bir Araştırma. Doçentlik Tezi, E.Ü. Elektronik Hesap Bilimleri Enstitüsü Bornova, İzmir.

Öztürk, A., Okur, M.C., Yanbastı, A.G. (1979); Faktör Analizi ve Bunun Psikiyatrideki Bir Uygulaması. E.Ü. Elektronik Hesap Bilimleri Dergisi, 3,2.

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 Al ternative for Canonical Correlation Analysis, Psychometrika, 42, 207-219.

EK -1 :

```
' REFE' CANONIC
' SCALAR' PP = 2 : QQ = 2
' VARIATE' V(1...4) I 215
' SET' VARP = V (1...2)
' SET ' VARQ = V (3...4)
' READ/P' VARP, VARQ I F , 5.3,5.3,5.3,5.3,/
' POINTER' P=VARP : Q=VARQ : PQ=VARP,VARQ
' DSSP' SSP 1 VARP, VARQ
' SSP' SSP
' CALCULATE' R = CORMAT (SSP)
' MATRIX ' C, CC I Q,P : WIT I Q,P : CT I P,Q
' MATRIX ' W1 I P, Q : VIT I Q,Q
' MATRIX ' CT I P,Q : FYY I Q,Q
' MATRIX ' FXX I P,Q : FXY I P,Q : FXYT I Q,P
' MATRIX ' V1 I Q,Q : FYXT I Q,Q : KORXX I Q,Q
' MATRIX ' KO I Q,Q : KORXY I Q,Q : KOI I Q,P
' SYMMAT ' R I PQ
' SYMMAT ' D, DI I P
' SYMMAT ' E, E1, A I Q
' DIAGMAT ' RDX, RDY I Q
' DIAGMAT ' RRI , RR, R1 I Q
' SCALAR ' T
' CALCULATE ' C, D, E = SUBMAT (R)
' CALCULATE ' D1 = INV (D)
' CALCULATE ' E1 = INV(E)
' CALCULATE ' A = RSYMRI (C ; D1)
' LRV ' A, E ; V1 , RR1, T
' CALCULATE ' RR = SQRT (RR1)
' CALCULATE / ZDZ = ZERO' R1 = (RR / RR) / RR
' CALCULATE' CC = PDT (C ; D1)
          :      W1 = IPDT (CC ; V1)
          :      W1 = PDT W1 ; R1)
          :      CT = TRANS (C)
          :      WIT = TRANS (W1)
          :      VIT = TRANS (V1)
          :      FYY = PDT (E ; V1)
          :      FYX = PDT (C ; W1)
          :      FYXT = TRANS (FYX)
          :      RDY = PDT (FYXT ; FYX)
          :      RDY = RDY / QQ
          :      FXX = PDT (D ; W1)
          :      FXY = PDT (CT ; V1)
```

```
        :    KO = PDT (WIT ; CT)
        :    KORXY = PDT ( KO ; V1)
        :    KO1 = PDT (WIT : D)
        :    KORXX = PDT (KO1 ; W1)
' PRINT' R,D,E,C,RR,V1, W1,T,FYY,FYX,RDY, FXX,FXY,RDX,
KORXY,KORXX I 12.4
' RUN'
```

EK-2:

```
' REFE' REDUNDANCY
' SCALAR' PP=2 : QQ = 2
' VARIATE' V(1....4) I 215
' SET ' VARP = V (1 .....2)
' SET ' VARQ = V (3....4)
' READ/P ' VARP, VARQ I F , 5.3,5.3,5.3,5.3,/
' POINTER' P = VARP : Q = VARQ : PQ= VARP, VARQ
' DSSP' SSP I VARP , VARQ
' MATRIX' C I Q,P : B2 I p,Q
' MATRIX' U, UT, V, I Q,Q : A2 I Q,P
' MATRIX ' U1 , UIT, W, WT I P,P : FXY I P,Q
' MATRIX ' KKUT I Q,Q : CORXX I P,P
' MATRIX ' B11, BB1, A3, A4 I Q,Q
' MATRIX ' C11, CC1, B3, B4 I P,P : FXYIT I Q,P
' MATRIX' CT I P,Q : RE1 I P,Q : RE2 I Q,P
' MATRIX ' REDI 1 P,P : RED2 1 Q,Q
' MATRIX' FYYI 1 Q,Q : FXY1 : P,Q
' MATRIX' FXX 1 P,P : FYX1 1 P,Q : FYXT I P,Q
' MATRIX' CO I P,Q : CORXY I P,Q : CO1 1 P,P
' SYMMAT ' D 1 P
' SYMMAT ' D1 I P
' SYMMAT ' E,A I Q
' SYMMAT ' E1 I Q
' SYMMAT ' B I Q
' SYMMAT ' B1 I P
' SYMMAT ' AA I Q
' SYMMAT ' BB I P
' SYMMAT ' RDX I Q
' SYMMAT ' RDY I P
' DIAGMAT ' K,KK,KI,KI1 1 Q
' DIAGMAT ' R1, RDXX 1 Q
' DIAGMAT ' K1 , KK1,K1I, K1I1 I P
' DIAGMAT ' RR1, RDyy I P
```

```
' SYMMAT' R  1  PQ
' SSP '  SSP
' CALCULATE'  R = CORMAT (SSP)
' CALCULATE ' C, D, E = SUBMAT (R)
' CALCULATE ' D1 = INV (D)
' CALCULATE ' E1 = INV (E)
' CALCULATE' CT = TRANS (C)
' CALCULATE ' RE 1= PDT (D1 ; CT)
' CALCULATE ' RED1 = PDT (RE1 ; C)
' CALCULATE ' RE 2 = PDT (E1 ; C)
' CALCULATE ' RED2 = PDT (RE2 ; CT)
' EQUATE' B = E
' CALCULATE ' A = PDT (C ; CT)
' LRV' B ; U, K,T
' CALCULATE' KK = SQRT (K)
' CALCULATE ' UT = TRANS (U)
' CALCULATE ' KKUT = PDT (KK ; UT)
' CALCULATE / ZDZ = ZERA' KI = ( K/K) / K
' CALCULATE ' KI1 = SQRT ( KI)
' CALCULATE ' BB1 = PDT (KI  1; UT)
              :          B1 I = PDT (KI 1 ; UT)
              :           A2  = PDT (B1 I ; C)
              :           A3  = PDT (A2 ; CT)
              :           A4 = PDT (A3 ; U)
              :           AA = PDT (A4 ; KI 1)
' LRV ' AA ; V, R1, TT
' CALCULATE' V= PDT (BB1 ; V)
' EQUATE '  E= B
' CALCULATE' FYY1 = PDT (E ; V)
              :        FXY1 = PDT (CT ; V)
              :        FXYIT = TRANS (FXY1)
              :        RDX = PDT (FXY1)
              :        RDX = RDX / PP
              :        RDXX = R1 / PP
' EQUATE'   B1 = D
' LRV' B1 ; U1, K1, T1
' CALCULATE' U1T = TRANS (U1)
              :     KK1= SQRT (K1)
' CALCULATE / ZDZ = ZERO' K1I = (K1 / K1) /K1
' CALCULATE' K1I1 = SQRT (K1I)
              :  CC1 = PDT (U1 ; K1II)
              :  C1I = PDT (K1I1 ; U1T)
              :   B2 = PDT (C1I ; CT)
              :   B3 = PDT (B2 ; C)
```

```
            :     B4 = PDT (B3 ; U1)
            :     BB = PDT (B4 ; K1II)
'LRV' BB ;  W, RR1 , TT1
'CALCULATE'  W = PDT (CC1)
'CALCULATE' FXX= PDT (D  ; W)
            :     FYX = PDT (D   ; W)
            :     FYXT = TRANS (FYX)
            :     RDY = PDT (FYX)
            :     RDY = RDY / QQ
            :     RDYY = RR1 / QQ
            :     WT = TRANS (W)
            :     CO = PDT (WT ; CT)
            :     CORXY = PDT (CO  ; V)
            :     CO1 = PDT (WT ; D)
            :     CORXX = PDT (CO1 ; W)
 ' PRINT / S' R, C, D, E, RED1,RED2, B, K, U, T, AA, R1, V, TT,
FYY1, FXY1, RDX, RDXX, BB, W, RR1, TT1, FXX, FYX, RDY,
RDYY, CORXY, CORXX   I   12, 4
'  RUN'
```