

Kanonik Korelasyon Katsayısının Elde Hesaplanması: KSÜ Zootekni Bölümü Öğrenci Tanıma Anketi Örneği

Şeyma KOÇ¹, Ayşe YENİPINAR¹, Ercan EFE¹, Mustafa ŞAHİN¹

¹Kahramanmaraş Sütçü İmam Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Kahramanmaraş
✉: kockutlu1@gmail.com

ÖZET

Bu çalışmada 4 değişkenli/ölçümlü ($p=4$) X bağımlı değişken grubu ve 6 değişkenli/ölçümlü ($q=6$) Y bağımsız değişken grubu kullanılarak bu değişkenler arasındaki kanonik korelasyon katsayıları hesaplanmıştır. Bulunan bu katsayıların, korelasyon matrisi üzerinden bulunan H_1 ve H_2 çözüm matrisinin öz değerlerinin kareköklerine eşit olduğu teorik olarak gösterilmiştir. Ayrıca X değişken kümesi ve Y değişken kümesi arasındaki, bulunan katsayıların önem kontrolleri yapılmıştır. Paket programlarla kısa sürede hesaplanan kanonik katsayıların ve bunların önem kontrollerinin teorik olarak nasıl hesaplandığı, bu hesaplamaların matematiksel alt yapısı ve işlem basamakları da gösterilmiştir.

DOI:10.18016/ksudobil.285930

Makale Tarihçesi

Received : 15.01.2017

Accepted : 15.03.2017

Anahtar Kelimeler

Kanonik Korelasyon Analizi,
Kanonik Değişken,
Çok Değişkenli Analizler

Araştırma Makalesi

Canonical Correlation Coefficient of Determination with Hand: KSU Department of Animal Science Student Recognition Survey Example

ABSTRACT

In this study the canonical correlation coefficients between the X ($p=4$) and Y ($q=6$) variable group is calculated. It is theoretically shown that, these coefficients are equal to the square root of the eigenvalues of matrix solution with the calculated correlation matrix H_1 and H_2 . Importance control between coefficient of X independent variable set and Y dependent variable set is also carried out. How to calculate the canonical coefficients calculated in a short time with package programs and their theoretical calculations of importance controls, mathematical sub-structure and treatments of these calculations steps are shown.

Article History

Geliş : 15.01.2017

Kabul : 15.03.2017

Keywords

Canonical Correlation Analysis,
Canonical Variable,
Multivariate Analysis

Research Article

To Cited : Koç Ş, Yenipınar A, Efe E, Şahin M 2018. Kanonik Korelasyon Katsayısının Elde Hesaplanması: KSÜ Zootekni Bölümü Öğrenci Tanıma Anketi Örneği. KSÜ Tarım ve Doğa Derg 21(1):66-72, DOI:10.18016/ksudobil.285930.

GİRİŞ

Uygulamalı bilim dallarında araştırmacılar çoğu zaman birden çok bağımsız değişkenin yine birden çok bağımlı değişkeni etkilemesi durumuyla karşı karşıya kalırlar. İlgilenilen sonuç çok sayıda değişken tarafından etkilenmekte bu değişkenler kendi aralarında ilişkili olabilmektedir. Bu durumda tek değişkenli analiz yöntemlerini kullanmak doğru yorumlamalar yapma açısından uygun olmayabilir. Bu sebeple çok değişkenli analiz yöntemlerinin kullanımı yaygınlaşmıştır. Ele alınan iki değişkenden hangi değişkenin bağımsız hangisinin bağımlı olduğuna bakılmaksızın, iki değişkenin birlikte artmaları, azalmaları veya ters ilişki göstermeleri birlikte değişim olarak ifade edilmektedir (Çankaya,2005). İstatistikte değişkenler arasındaki birlikte değişimin ölçüsü korelasyon katsayısı ile ortaya konmaktadır (Gürsakal, 1998). Bilindiği gibi istatistikte bir bağımlı

değişken ile bir bağımsız değişken arasındaki doğrusal ilişkinin derecesinin ve yönünün belirlenmesi için hesaplanan ilişki katsayısı Basit Korelasyon Katsayısıdır. Bağımlı değişken sayısı bir bağımsız değişken sayısı birden çok olduğu durumda hesaplanan ilişki katsayısı ise Çoklu Korelasyon katsayısıdır. Bundan farklı olarak $p>1$ $q>1$ sayıda değişken içeren iki veri seti (X ve Y) arasındaki değişkenler arası ilksileri ortaya koymak amacıyla değişkenleri bağımlı-bağımsız ayırımı yapmadan aralarındaki iç ilişkinin ortaya çıkarılması için hesaplanan ilişki katsayısı ise Kanonik Korelasyon Katsayısı (KKK) olarak adlandırılır. Birbirleriyle ilişkili olduğu düşünülen değişken kümelerinin doğrusal bileşenler yardımıyla ne ölçüde açıklanabildiğinin belirlenmesi, değişkenlerin kendi aralarındaki karşılıklı iç bağımlılık yapısı ve bu bağımlılığın önem kontrolleri Kanonik Korelasyon Analizinin (KKA) konusudur. Bu analizde değişken

setlerinden birisi bağımsız değişken seti diğeri bağımlı değişken seti olarak tanımlanabilir. Ancak değişken setlerinin bu şekilde tanımlanma zorunluluğu yoktur (Keskin ve Özsoy,2004). Faktör analizi ile birlikte en karmaşık işlem aşamalarını gerektiren teknikler arasında yer almaktadır (Tatlıdil, 1996). Hotelling tarafından 1935 yılında ortaya atılan KKA, maksimum korelasyonlu ve birim varyanslı doğrusal bileşenler elde ederek iki değişken kümesi arasındaki ilişkiyi tanımlamakta ve birçok bağımlı değişken üzerine birçok bağımsız değişkenin etkisini ortaya koymaktadır (Çankaya ve ark., 2009). Değişkenleri bağımlı-bağımsız şeklinde kümelemediği için bu analize literatürde Kümeler Arası İlişki Analizi de denmektedir. Teorik olarak iki değişken seti arasında bağımlı bağımsız değişken ayrımı yapılabiliyorsa, bu durumda kanonik korelasyon analizinin amacı bağımsız değişken setinin bağımlı değişken setini etkileyip etkilemediğini belirlemeye yöneliktir (Hair ve ark. 1998; Sharma, 1996).

Bu çalışmanın amacı, paket programlarla kısa sürede hesaplanan kanonik değişkenlerin, kanonik katsayıların ve bu katsayıların önem kontrollerinin teorik olarak elde nasıl hesaplandığını, matematiksel alt yapının adım adım tüm işlem basamaklarını açıklamak, lisansüstü çalışmalarına yeni başlayan öğrencilere, konunun meraklısı araştırmacılara yol gösterici nitelikte temel prensipleri ortaya koyarak rehberlik etmektir.

MATERYAL ve METOT

Materyal

Çalışmada Kahramanmaraş Sütçü İmam Üniversitesi Ziraat Fakültesi Zootekni Bölümünden seçilen 100 lisans Öğrencisine uygulanan tanıma anketi verileri kullanılmıştır. Ankette kullanılan X_1, X_2, X_3, X_4 ve $Y_1, Y_2, Y_3, Y_4, Y_5, Y_6$ verilerin yapısı Çizelge 1'de verilmiştir. Matris cebri ve hesap makinesi kullanılarak yapılan hesaplamaların doğruluğu SAS istatistik programı yardımıyla kontrol edilmiştir.

Çizelge 1. Değişken isimleri

Değişken no	Değişken İsmi	Değişken no	Değişken İsmi
X1	Kardeş sayısı	Y1	Kaldığı yer ücreti
X2	Aile reisinin geliri	Y2	Sınav giriş puanı
X3	Ek gelir	Y3	Kişisel harcama
X4	Son döneme ait not ortalaması	Y4	Aldığı burs
		Y5	Okula Geliş Süresi
		Y6	Kaçıncı tercihi

Metot

Kanonik korelasyon analizinin ön şartları, amacı aynı zamanda katsayının teorik olarak hesaplanışına dair işlem basamakları verilmiştir. Bu katsayıların önem kontrolleri yapılırken kullanılan test istatistikleri ve bu istatistiklerin serbestlik dereceleri alt başlıklar olarak verilmiştir.

Kanonik Korelasyon Analizinin Ön Şartları ve Amacı

Kanonik korelasyon analizinin ön şartları :

- Her bir değişken kümesinin çok değişkenli normal dağılım göstermesi,
- Değişken kümeleri arasındaki ilişkinin doğrusal olması,
- Her değişken kümesindeki değişkenler arasında çoklu bağlantı veya çoklu birlikte değişim (multicollinearity) bulunmamalıdır (Özdamar, 1999; Tabachnick ve Fidell, 2001).
- Ölçüm hatasının minimum seviyede olması,
- Elde edilen sonuçların güvenilir olması için, örnek genişliğinin mümkün olduğunca büyük olması (Değişken sayısının 5 katı kadar).
- İki grup değişken kümesinde yer alan değişkenlerin eşit sayıda olma zorunluluğu yoktur (Çankaya,2005).
- Değişkenler arasındaki korelasyonu önemli düzeyde etkilemesi nedeni ile veri kümesinde aykırı değerlerin analiz öncesinde saptanarak gerekli düzeltme ya da elimine edilmesi gerekmektedir(Çankaya, 2005).

Bu analizde aynı bireyden elde edilen iki değişken kümesinin istatistiksel olarak birbirinden bağımsız olup olmadığının test edilmesi, kümeler arası korelasyona en fazla katkıda bulunan her iki değişken kümesindeki değişkenlerin belirlenmesi, bağımsız ve bağımlı değişken kümelerine ait değişkenler arasındaki korelasyonu maksimum yapan doğrusal kombinasyonların belirlenmesi, bir değişken kümesinin diğer bir değişken kümesi tarafından ne ölçüde açıklanabildiğinin belirlenmesi, bir kanonik değişkenin dahil olduğu değişkenler kümesinin açıklayıcı gücüne ne ölçüde katkı sağlayabildiğinin belirlenmesi ve farklı kanonik fonksiyonların ilişkilerinin açıklanması ya da bu ilişkiyi tahmin etmedeki nispi (görelî) gücün ne kadar olduğunun belirlenmesi amaçlanır (Çankaya,2005).

Kanonik Korelasyon Katsayısının Hesaplanışı

Kanonik korelasyon analizinde, X ve Y değişken kümeleri arasındaki doğrusal ilişkiler araştırılır. Bu yöntem, bir kaç doğrusal bileşen oluşturulması ile ilgili olup, bu doğrusal bileşenler yoluyla iki veya daha fazla küme arasındaki ilişkileri en iyi şekilde ortaya koymaktadır. Bu doğrusal bileşenler "**kanonik değişkenler**" ve aynı şekilde kanonik değişkenlerin benzer çiftleri arasındaki ilişkiler de "**kanonik korelasyonlar**" olarak adlandırılır.

Çalışmada 4 değişkenli/ölçümlü ($p=4$) X değişken

grubu ve 6 değişkenli/ölçümlü ($q=6$) Y değişken grubu kullanılarak bunların doğrusal bileşenleri kullanılarak U ve V yeni kanonik denklemler ve R korelasyon matrisi bulunur. Korelasyon matrisinin gösterimi

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \text{ şeklinde olup} \quad (1)$$

R_{11} = p tane değişkenin kendi aralarında korelasyon matrisi

R_{22} = q tane değişkenin kendi aralarında korelasyon matrisi

R_{12} = p ve q tane değişkenin aralarındaki korelasyon matrisi

R_{21} = R_{12} matrisinin devrik matrisi olarak ifade edilir. R korelasyon matrisinin H_1 ve H_2 çözüm matrisleri ise Eşitlik 2'de verilmiştir.

$$\begin{aligned} H_1 &= R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} \\ H_2 &= R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} \end{aligned} \quad (2)$$

Kanonik katsayıları hesaplayabilmek için Eşitlik 3'de verilen karakteristik denklemlerin determinantlarının herhangi birinden özdeğerler (λ) hesaplanır.

$$\begin{aligned} |R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} - \lambda I| &= 0 \\ |R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} - \lambda I| &= 0 \end{aligned} \quad (3)$$

Genel olarak; eğer bir kümede p tane değişken diğer kümede ise q tane değişken söz konusu ise, olabilecek kanonik değişken ve kanonik korelasyon sayısı min (p,q), iki kümedeki en küçük değişken sayısı kadar olacaktır. Örneğin, birinci kümede p=4, ikinci kümede q=6 tane değişken söz konusu ise min (4, 6) den 4 tane kanonik değişken ve kanonik korelasyon katsayısı elde edilir. Karakteristik denklemin kökleri olan değerlerin kanonik korelasyon ile $r_i = \sqrt{\lambda_i}$ şeklinde ilişkisi vardır. Çalışmada 100 tane örnek biriminden p=4 olan X değişken kümesi ve q=6 olan Y değişken kümesi ile bunların doğrusal kombinasyonları yardımıyla U ve V kanonik değişkenleri hesaplanıp bu değişkenler arasındaki r_i kanonik korelasyonlar hesaplanmış, bu katsayının karakteristik matrisin özdeğerlerinin kareköküne eşit olduğu gösterilmiştir.

Kanonik Korelasyon Katsayısının Önem Testi

Bulunan kanonik değişken çiftlerinden kaç tanesinin önemli olduğu, bir başka deyişle, değişken grupları arasındaki ilişkinin (kovaryansın) kaç tanesi ile büyük oranda açıklanabileceğine karar verilmelidir (Çemrek, 2012). Bunun için uygulamada genel yaklaşım şöyledir. Önce tüm kanonik korelasyonlar birlikte test edilir. Bu test sonucu anlamlı bulunmazsa yani H_0 hipotezinin red edilmesi durumunda en büyük olan kanonik korelasyon katsayısı hipotezden çıkarılarak işlemler H_0 hipotezi kabul edilinceye kadar

tekrarlanır. İşlem sırasında çıkarılan her kanonik korelasyon katsayısı için değişken sayıları 1 eksilti olarak serbestlik derecesi hesaplanır. Kanonik korelasyon analizi sonucu elde edilen kanonik korelasyon katsayılarının kontrolü için en yaygın olarak kullanılan test yöntemi Bartlett tarafından 1941 yılında geliştirilen Wilk's lamda ya da Bartlett test istatistiğidir (Bartlett, 1941; Thompson, 1985).

$$\begin{aligned} H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_p = 0 \\ H_1: \text{En az bir } \rho_i \neq 0 \end{aligned}$$

Test istatistik değeri,

$$\Lambda = \prod_{i=1}^p (1 - \lambda_i) \quad (4)$$

$$\chi^2 = -[(n-1) - (p+q+1)/2] \log_e (\prod_{i=1}^p (1 - \lambda_i^2)) \quad (5)$$

şeklinde dir. Eşitlik 5'te verilen test istatistiği pq serbestlik dereceli χ^2 tablo değeri ile karşılaştırılmalıdır. H_0 yokluk hipotezi reddedilirse, ilk kanonik korelasyon katsayısı λ_{11} çıkarılır ve kalan p-1 kanonik korelasyon katsayısı için test istatistiği ve formüller

$$\Lambda^1 = \prod_{i=1}^p (1 - \lambda_i),$$

$$\chi^2 = -[(n-1) - (p+q+1)/2] \log_e \Lambda^1$$

şeklinde hesaplanır. Bu durumda yeni istatistiğin serbestlik derecesi (p-1)(q-1) dir. H_0 hipotezi reddedilir ise değeri en büyük olan katsayı hipotezden çıkarılır ve işlemlere H_0 hipotezi kabul edilinceye kadar devam edilir. Genel olarak s tane kanonik korelasyon katsayısı çıkarıldıktan sonra durum;

$$\Lambda^s = \prod_{i=s+1}^p (1 - \lambda_i) \quad s=1,2,\dots,p \quad (6)$$

ve

$$\chi^2 = -[(n-1) - (p+q+1)/2] \log_e \Lambda^s \quad (7)$$

şeklinde dir. Eşitlik 7 de verilen s' inci χ^2 nin serbestlik derecesi (p-s) (q-s) dir. Bu işlemler önemsiz χ_{hesap}^2 değerine kadar devam eder. Ayrıca Wilk's Lambda katsayısı sıfıra yaklaştıkça, H_0 hipotezinin reddedileceği (kanonik korelasyon katsayısının anlamlı olduğu), χ^2 değeri ile korelasyon katsayılarının sıfırdan farklı (anlamlı) olacağı söylenebilir (Kaya, 2008).

BULGULAR ve TARTIŞMA

Bu çalışmada K.Maraş Sütçü İmam Üniversitesi Ziraat Fakültesi Zootečni Bölümünde okuyan rastgele seçilmiş 100 öğrenciye uygulanan tanıma anketi sonucunda bulunan özet istatistikler Çizelge 2'de verilmiştir. Çalışmada kullanılan iki değişken grubunun birbirleriyle ilişkisini veren korelasyonlar Çizelge 3'de verilmiştir.

Değişken grupları arasındaki korelasyon matrisi

tablosuna bakarak birkaç yorum yapılmak istenirse örneğin Y_1 (Kaldığı yer ücreti) ve Y_3 (Kişisel harcama) değişkenleri arasındaki katsayı 0.55 dir. Bu katsayıya bakarak öğrencinin kaldığı yer ücreti ile kişisel harcaması arasında pozitif yönlü %55 lik bir ilişki olduğu söylenebilir. Benzer şekilde X_3 (Ek gelir) ile Y_4 (Aldığı burs) arasında negatif yönlü bir ilişki söz konusudur. Ek gelir arttıkça öğrencinin aldığı burs azalmaktadır. X_1 kardeş sayısı ile Y_4 alınan burs arasındaki ilişki negatif yönlü ve çok düşüktür. X_2 (aile reisinin geliri) değişkeni ile Y_3 (Kişisel harcama) değişkeni arasında pozitif yönlü %45 lik bir ilişki olduğu söylenebilir. X_3 (Ek gelir) değişkeninin, Y_3

(Kişisel harcama) değişkeni ile arasındaki bulunan ilişki pozitif yönlü ve %34 iken, Y_6 tercih sıralaması değişkeni ile ilişkisi negatif yönlü ve zayıf bir ilişkidir(%7). Y_1 kaldığı yerin ücreti ve Y_5 okula geliş süresi arasında %41 lik negatif bir ilişki vardır. Bulunan katsayıların değişkenlerin birbirleri arasındaki ilişkiyi tanımlamada tatmin edici düzeyde olmadığı görülse de çalışmada amaç deneysel bulguya erişim değildir. Paket programların sonuç/çıktı ekranlarında bu katsayılara erişen araştırmacı aşağıdaki matematiksel işlem basamaklarını yaparak kanonik korelasyon katsayısına ulaşır.

Çizelge 2. Özet istatistikler

Değişken	Ortalama	Standart Sapma	Değişken	Ortalama	Standart Sapma
X1	3.7900	1.7712	Y1	245.0000	183.3333
X2	2183.0000	1196.6329	Y2	269.0900	37.4708
X3	111.2000	282.0891	Y3	492.0000	260.7990
X4	2.2922	0.5652	Y4	217.1300	156.7776
			Y5	31.3800	18.6497
			Y6	6.9900	6.5881

Çizelge 3. X ve Y Değişken grubu arasındaki korelasyon matrisi

	X ₁	X ₂	X ₃	X ₄	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆
X ₁	1.0000	0.0212	-0.0482	0.0964	0.0170	-0.1151	-0.0999	-0.0026	0.0034	0.1513
X ₂	.	1.0000	0.3869	0.1085	0.3023	0.2041	0.4537	-0.3116	-0.0875	0.1022
X ₃	.	.	1.0000	0.0357	0.3984	0.1116	0.3442	-0.3314	-0.1243	-0.0727
X ₄	.	.	.	1.0000	0.0418	0.2328	0.0222	-0.2198	-0.0886	-0.1344
Y ₁	1.0000	0.1744	0.5569	0.1126	-0.4119	0.1534
Y ₂	1.0000	0.2804	-0.1956	-0.0702	-0.0229
Y ₃	1.0000	-0.1397	-0.2259	0.1951
Y ₄	1.0000	-0.0643	0.1012
Y ₅	1.0000	-0.0328
Y ₆	1.0000

$$R_{11} = \begin{bmatrix} 1 & 0.0212 & -0.0482 & 0.0964 \\ 0.0212 & 1 & 0.3869 & 0.1015 \\ -0.0482 & 0.3869 & 1 & 0.0357 \\ 0.0964 & 0.1015 & 0.0350 & 1 \end{bmatrix}$$

$$R_{22} = \begin{bmatrix} 1 & 0.1744 & 0.569 & 0.1126 & -0.4119 & 0.1534 \\ 0.1744 & 1 & 0.2804 & -0.1956 & -0.0702 & -0.0229 \\ 0.5569 & 0.2804 & 1 & -0.1397 & -0.2259 & 0.1951 \\ 0.1126 & -0.1956 & -0.1397 & 1 & -0.0643 & 0.1012 \\ -0.4119 & -0.0702 & -0.2259 & -0.0643 & 1 & -0.0328 \\ 0.1534 & -0.0229 & 0.1951 & 0.1012 & -0.0328 & 1 \end{bmatrix}$$

$$R_{12} = \begin{bmatrix} 0.0170 & -0.1151 & -0.0999 & -0.0026 & 0.0034 & 0.1513 \\ 0.3023 & 0.2041 & 0.4537 & -0.3116 & -0.0875 & 0.1022 \\ 0.3984 & 0.1116 & 0.3442 & -0.3314 & -0.1243 & -0.0727 \\ 0.0418 & 0.2328 & 0.0222 & -0.2198 & -0.0886 & -0.1344 \end{bmatrix}$$

$$R_{21} = \begin{bmatrix} 0.0170 & 0.3023 & 0.3984 & 0.0418 \\ -0.1151 & 0.2041 & 0.1116 & 0.2328 \\ -0.0999 & 0.4537 & 0.3442 & 0.0222 \\ -0.0026 & -0.3116 & -0.3314 & -0.2198 \\ 0.0034 & 0.0875 & -0.1243 & -0.0886 \\ 0.1513 & 0.1022 & -0.0727 & -0.1344 \end{bmatrix}$$

$H_1=R_{11}^{-1} R_{12}R_{22}^{-1}R_{21}$ matrisinin ve $H_2=R_{22}^{-1} R_{21}R_{11}^{-1} R_{12}$ bulunmuştur. H_1 matrisi aşağıdaki hesaplamalar ile değeri MATLAB R2012a paket programı kullanılarak

$$H_1 = \begin{bmatrix} 1 & 0.0212 & -0.0482 & 0.0964 \\ 0.0212 & 1 & 0.3869 & 0.1015 \\ -0.0482 & 0.3869 & 1 & 0.0357 \\ 0.0964 & 0.1015 & 0.0350 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0170 & -0.1151 & -0.0999 & -0.0026 & 0.0034 & 0.1513 \\ 0.3023 & 0.2041 & 0.4537 & -0.3116 & -0.0875 & 0.1022 \\ 0.3984 & 0.1116 & 0.3442 & -0.3314 & -0.1243 & -0.0727 \\ 0.0418 & 0.2328 & 0.0222 & -0.2198 & -0.0886 & -0.1344 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0.1744 & 0.5569 & 0.1126 & -0.4119 & 0.1534 \\ 0.1744 & 1 & 0.2804 & -0.1956 & -0.0702 & -0.0229 \\ 0.5569 & 0.2804 & 1 & -0.1397 & -0.2259 & 0.1951 \\ 0.1126 & -0.1956 & -0.1397 & 1 & -0.0643 & 0.1012 \\ -0.4119 & -0.0702 & -0.2259 & -0.0643 & 1 & -0.0328 \\ 0.1534 & -0.0229 & 0.1951 & 0.1012 & -0.0328 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0170 & 0.3023 & 0.3984 & 0.0418 \\ -0.1151 & 0.2041 & 0.1116 & 0.2328 \\ -0.0999 & 0.4537 & 0.3442 & 0.0222 \\ -0.0026 & -0.3116 & -0.3314 & -0.2198 \\ 0.0034 & 0.0875 & -0.1243 & -0.0886 \\ 0.1513 & 0.1022 & -0.0727 & -0.1344 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 0.0615 & -0.0256 & -0.0112 & -0.0355 \\ -0.0225 & 0.2152 & 0.1525 & 0.0328 \\ -0.0008 & 0.1751 & 0.2613 & 0.0792 \\ -0.0316 & 0.0474 & 0.0728 & 0.1092 \end{bmatrix} \text{ şeklinde bulunur.}$$

Benzer işlemler yardımıyla H_2 matrisi ise ;

$$H_2 = \begin{bmatrix} 1 & 0.1744 & 0.5569 & 0.1126 & -0.4119 & 0.1534 \\ 0.1744 & 1 & 0.2804 & -0.1956 & -0.0702 & -0.0229 \\ 0.5569 & 0.2804 & 1 & -0.1397 & -0.2259 & 0.1951 \\ 0.1126 & -0.1956 & -0.1397 & 1 & -0.0643 & 0.1012 \\ -0.4119 & -0.0702 & -0.2259 & -0.0643 & 1 & -0.0328 \\ 0.1534 & -0.0229 & 0.1951 & 0.1012 & -0.0328 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0170 & 0.3023 & 0.3984 & 0.0418 \\ -0.1151 & 0.2041 & 0.1116 & 0.2328 \\ -0.0999 & 0.4537 & 0.3442 & 0.0222 \\ -0.0026 & -0.3116 & -0.3314 & -0.2198 \\ 0.0034 & 0.0875 & -0.1243 & -0.0886 \\ 0.1513 & 0.1022 & -0.0727 & -0.1344 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0.0212 & -0.0482 & 0.0964 \\ 0.0212 & 1 & 0.3869 & 0.1015 \\ -0.0482 & 0.3869 & 1 & 0.0357 \\ 0.0964 & 0.1015 & 0.0350 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0170 & -0.1151 & -0.0999 & -0.0026 & 0.0034 & 0.1513 \\ 0.3023 & 0.2041 & 0.4537 & -0.3116 & -0.0875 & 0.1022 \\ 0.3984 & 0.1116 & 0.3442 & -0.3314 & -0.1243 & -0.0727 \\ 0.0418 & 0.2328 & 0.0222 & -0.2198 & -0.0886 & -0.1344 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 0.1738 & 0.0379 & 0.1307 & -0.1485 & -0.0558 & -0.0174 \\ -0.0158 & 0.0634 & 0.0078 & -0.0306 & -0.0117 & -0.0342 \\ 0.0836 & 0.0607 & 0.1603 & -0.0710 & -0.0167 & 0.0250 \\ -0.1727 & -0.0926 & -0.1663 & 0.1815 & 0.0645 & 0.0179 \\ 0.0203 & -0.0129 & 0.0230 & -0.0014 & 0.0010 & 0.0153 \\ -0.0287 & -0.0436 & -0.0208 & 0.0425 & 0.0209 & 0.0672 \end{bmatrix} \text{ şeklinde hesaplanır.}$$

$|H_1 - \lambda I| = 0$ denklemi ile öz değerler (λ_{1i} değerleri) sırasıyla

$$\begin{vmatrix} 0.0615 - \lambda & -0.0256 & -0.0112 & -0.0355 \\ -0.0225 & 0.2152 - \lambda & 0.1525 & 0.0328 \\ -0.0008 & 0.1751 & 0.2613 - \lambda & 0.0792 \\ -0.0316 & 0.0474 & 0.0728 & 0.1092 - \lambda \end{vmatrix} = 0$$

$= 0.4275, \lambda_{12} = 0.1065, \lambda_{13} = 0.0819, \lambda_{14} = 0.0313$ şeklinde hesaplanır.

λ_{11}

$|H_2 - \lambda I| = 0$ denklemi ile öz değerler (λ_{2i} değerleri) sırasıyla

$$\begin{vmatrix} 0.1738 - \lambda & 0.0379 & 0.1307 & -0.1485 & -0.0558 & -0.0174 \\ -0.0158 & 0.0634 - \lambda & 0.0078 & -0.0306 & -0.0117 & -0.0342 \\ 0.0836 & 0.0607 & 0.1603 - \lambda & -0.0710 & -0.0167 & 0.0250 \\ -0.1727 & -0.0926 & -0.1663 & 0.1815 - \lambda & 0.0645 & 0.0179 \\ 0.0203 & -0.0129 & 0.0230 & -0.0014 & 0.0010 - \lambda & 0.0153 \\ -0.0287 & -0.0436 & -0.0208 & 0.0425 & 0.0209 & 0.0672 - \lambda \end{vmatrix} = 0$$

$\lambda_{21} = 0.4275, \lambda_{22} = 0.1065, \lambda_{23} = 0.0819, \lambda_{24} = 0.0314, \lambda_{25} = 0.0000, \lambda_{26} = 0.0000$ şeklinde hesaplanır.

SAS istatistiksel paket programı yardımıyla bulunan kanonik korelasyon katsayıları ve output sayfası Çizelge 4'te; elde hesaplanarak bulunan özdeğerlerin karekökleri Çizelge 5 'te verilmiştir.

Çalışmada, X değişken kümesinde 4, Y değişken kümesinde 6 adet değişken sayısı vardır. Değişken sayısı düşük olanın sayısı kadar yani 4 adet kanonik değişken çifti elde edilmiş ve bu kanonik değişken çiftleri arasından 4 adet kanonik korelasyon katsayısı hesaplanmıştır. Hesaplanan kanonik korelasyon katsayılarına ait istatistik bulgular Çizelge 6'te verilmiştir.

Çizelge 4. SAS program çıktısı

The CANCORR Procedure			
Canonical Correlation Analysis			
Canonical correlation	Adjusted Canonical correlation	Standard Error	Squared Canonical correlation
0.653780	0.617005	0.057546	0.427428
0.326290	0.157733	0.089804	0.106465
0.286009	.	0.092282	0.081801
0.177102	.	0.097351	0.031365

Çizelge 5. SAS sonuçları ile elle hesaplamaların kıyaslanması

No	SAS İle Elde Edilen Kanonik Korelasyon Katsayıları	Elde Hesaplanan Kanonik Korelasyon Katsayıları (Özdeğerlerin Karekökleri)
1	0.653780	$0.4275^{1/2}=0.6538$
2	0.326290	$0.1065^{1/2}=0.3263$
3	0.286009	$0.0819^{1/2}=0.2861$
4	0.177102	$0.0313^{1/2}=0.1772$

Çizelge 6. Kanonik değişkenlerle ilgili istatistik bulgular

Kanonik değişkenler	F değeri	Serbestlik derecesi	P değeri
U_1V_1	3.33	24	<.0001
U_2V_2	1.46	15	0.1223
U_3V_3	1.39	8	0.2041
U_4V_4	1.00	3	0.3948

Bulunan bu kanonik korelasyon katsayısının önem kontrolü için aşağıdaki hipotezler kurulur.

$$H_0: \rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$$

$$H_1: \text{En az bir } \rho_i \neq 0$$

Test istatistik değeri,

$$\Lambda = \prod_{i=1}^4 (1 - \lambda_i)$$

$$\chi^2 = -[(n-1) - (p+q+1)/2] \log_e \left(\prod_{i=1}^4 (1 - r_i^2) \right)$$

$$= - \left[(100-1) - \frac{4+6+1}{2} \right] \log_e [(1 - 0.6538^2) (1 - 0.3263^2) (1 - 0.2860^2) (1 - 0.1784^2)]$$

$$= -(99 - 5.5) \ln[(0.5725)(0.8932)(0.9182)(0.9681)]$$

$$= -93.5 * -0.7884 = 73.71 = \chi^2_{hesap}$$

$$= \chi^2_{p*q,\alpha} = \chi^2_{24,0.05} = 36.415$$

$\chi^2_{hesap} > \chi^2_{24,0.05}$ olduğundan H_0 reddedilir. İlk kanonik korelasyon çıkarılarak kalan korelasyonlarla hipotez testine devam edilir. Bulunan kıkare değeri (p-1)(q-1) serbestlik dereceli kıkare tablo değeri ile karşılaştırılır.

$$H_0: \rho_2 = \rho_3 = \rho_4 = 0$$

$$H_1: \rho_2 \neq \rho_3 \neq \rho_4$$

Test istatistik değeri,

$$\chi^2 = - \left[(100-1) - \frac{4+6+1}{2} \right] \log_e [(1 - 0.3263^2) (1 - 0.2860^2) (1 - 0.1784^2)]$$

$$= -(99 - 5.5) \ln[(0.8932)(0.9182)(0.9681)] = 21.57$$

$$= \chi^2_{hesap} = 21.57$$

$$\chi^2_{tablo} = \chi^2_{(4-1)(6-1),0.05} = 24.99$$

$\chi^2_{hesap} < \chi^2_{15,0.05}$ olduğundan H_0 hipotezi kabul edilir.

Bu durumda birinci kanonik korelasyon katsayısının önemli olduğuna diğer üç kanonik korelasyon katsayısının önemsiz olduğuna karar verilir. İki değişken kümesi arasındaki ilişkinin birinci kanonik katsayı ve kanonik değişken vektörü ile açıklanacağı söylenebilir. Değişken kümeleri arasındaki yorumlar ilk kanonik katsayı üzerinden yapılmalıdır. İlk kanonik korelasyon katsayısının önemli bulunmasından dolayı U_1 ve V_1 kanonik değişkenlerin ve bu değişkenler ile orijinal değişkenler arasındaki korelasyona bakılması ve hesaplanan kanonik değişkenlerin dahil oldukları ve olmadıkları değişken kümelerinde görülen toplam varyasyonun ne kadarını izah edebildiğinin açıklanması için Çizelge 6, 7, 8 ,9 oluşturulmuştur.

Kanonik değişkenler ile kendi kümeleri içerisinde yer alan orijinal değişkenler arasındaki kanonik yükler Çizelge 7 ve Çizelge 8'de verilmiştir.

Çizelge 7. U_i ve kendi orijinal değişkenleri X_i arasındaki kanonik yükler

	U_1	U_2	U_3	U_4
X_1	-0.0751	0.3703	-0.5288	0.7600
X_2	0.7720	0.3509	0.4640	0.2560
X_3	0.8441	0.0471	-0.4370	-0.3070
X_4	0.2981	-0.7788	0.0226	0.5514

Çizelge 7 ye göre U_1 değişkeninin oluşmasında en büyük katkıyı işaretlere bakılmaksızın en yüksek kanonik yüke sahip olan X_3 (Ek gelir) değişkeni yapmıştır. Buna göre katsayılar (yükler) incelendiğinde U_1 değişkenine katkısı sırasıyla X_2 , X_4 ve X_1 değişkenleri sağlamıştır

Çizelge 8. V_i ve kendi orijinal değişkenleri Y_i

arasındaki kanonik yükler

	V ₁	V ₂	V ₃	V ₄
Y ₁	0.6336	0.2451	-0.2816	-0.1795
Y ₂	0.3605	-0.5110	0.5044	0.3019
Y ₃	0.7038	0.3701	0.5071	-0.2479
Y ₄	-0.6391	0.2183	0.0327	-0.4095
Y ₅	-0.2198	0.1461	0.0827	-0.0934
Y ₆	-0.0571	0.7266	0.1493	0.6188

Çizelge 8 e göre V₁ değişkenine en büyük katkıyı işaretlere bakılmaksızın en yüksek kanonik yüke sahip olan Y₃ (Kişisel harcama) değişkeni yapmıştır. En düşük katkıyı ise Y₆ değişkeni yapmıştır.

Kanonik değişken çiftleri ile diğer değişken kümesinde yer alan orijinal değişkenlere ait kanonik yükler Çizelge 9 ve Çizelge 10'da verilmiştir.

Çizelge 9. V_i ve X_i değişken kümesi arasındaki kanonik yükler

	V ₁	V ₂	V ₃	V ₄
X ₁	-0.0491	0.1208	-0.1512	0.1346
X ₂	0.5047	0.1145	0.1327	0.0453
X ₃	0.5519	0.0154	-0.1250	-0.0544
X ₄	0.1949	-0.2541	0.0065	0.0977

Çizelge 10. U_i ve Y_i değişken kümesi arasındaki kanonik yükler

	U ₁	U ₂	U ₃	U ₄
Y ₁	0.4142	0.0800	-0.0805	-0.03185
Y ₂	0.2357	-0.1677	0.1443	0.0535
Y ₃	0.4601	0.1208	0.1450	-0.0439
Y ₄	-0.4178	0.0712	0.0099	-0.0725
Y ₅	-0.1437	0.0477	0.0237	-0.0165
Y ₆	-0.0374	0.2371	0.0427	0.1096

SONUÇ

İstatistik analizlerin temeli uzun hesaplamalar gerektiren zahmetli matematiksel işlemlere dayanır. Bunun için paket programların kullanımı adeta zaruret haline gelmiştir. Ancak istatistik paket programların kullanımı ilk bakışta çok büyük bir kolaylık olarak görünse de beraberinde bazı sorunları da getirir. Araştırmacının paket programdan elde ettiği sonucun temel matematik yapısını ve elde teorik hesaplamalarını bilmemesi verilerin yorumlanması ile ilgili ciddi sorunlar oluşturur. Bir paket programın doğru çalıştığından emin olmak için basit bir veri seti ile elde hesaplamalar yapıp programın çıktı sayfasındaki sonuçlarla karşılaştırılması ortaya çıkabilecek hataları minimize eder. Dolayısıyla daha doğru sonuçlara ulaşılır. Bu amaçla bu çalışmada yeni

başlayan araştırmacılara, lisansüstü öğrencilerine ve konunun meraklılarına kanonik korelasyon katsayılarının ve bu katsayıların önem kontrollerinin hangi aşamalardan geçerek elde edildiği gösterilmiştir. Bu tür çalışmaların çoğalması, ülkemizde yaygın kullanılan paket programların çıktı sayfalarının araştırmacı tarafından daha sağlıklı yorumlanabilmesi açısından son derece önemli olduğu düşünülmektedir.

KAYNAKLAR

- Bartlett MS 1941. The Statistical Significance of Canonical Correlations. *Biometrika*. Vol. 32: 29-37.
- Çankaya S 2005. Kanonik Korelasyon Analizi ve Hayvancılıkta Kullanımı, Cukurova Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni, Adana, Doktora Tezi, 145s.
- Çankaya S, Altop A, Oflaz M, Erener G 2009. Karayaka Toklularında Kesim Önceki ve Kesim Sonrası Ölçülen Bazı Özellikler Arasındaki İlişkinin Tahmini için Kanonik Korelasyon Analizi, *Anadolu Tarım Bilimleri Dergisi*, 24 (1): 61-66.
- Çemrek F 2012. Türkiye'deki İllerin Gelir ve Refah Düzeyi Değişkenleri Arasındaki İlişkinin Kanonik Korelasyon Analizi ile İncelenmesi. *Gazi Üniversitesi İİBF Dergisi*, 7(2):197- 215.
- Gürsakal N 1998. Bilgisayar Uygulamalı İstatistik II. Marmara Kitabevi, Bursa, 517s.
- Hair JF, Anderson RE, Babin BJ, Black WC 1998. *Multivariate Data Analysis*, Pearson Education Limited 2014, Harlow, 729s.
- Keskin S, Özsoy AN 2004. Kanonik Korelasyon Analizi ve Bir Uygulaması. *Tarım Bilimleri Dergisi*, 10(1):67-71.
- Kaya L 2008. Birden Fazla Değişken İçeren Setler Arasındaki İlişkinin Kanonik Korelasyon Analizi İle Belirlenmesi. Yüksek Lisans Tezi, Harran Üniversitesi 2008.
- Özdamar K 1999. Paket Programlarla İstatistiksel Veri Analizi. Kaan Kitabevi, Eskişehir, 474s
- Sharma S 1996. *Applied Multivariate Techniques*. John Wiley&Sons Inc. , Canada, 493s.
- Tabachnick BG, Fidell LS 2001. *Using Multivariate Statistics*. Pearson Education Limited , Harlow, 1049s.
- Tatlıdil H 1996. Uygulamalı Çok Değişkenli İstatistiksel Analiz. Cem Web Ofset Ltd. Şti, Ankara, 424s.
- Thompson B 1985. *Canonical Correlation Analysis*. Sage Publication Ltd. , London, 69s.