

KANONİK KORELASYON ANALİZİ

Tanım: Hotelling (1936) tarafından geliştirilmiş olup, çok sayıda değişkenden oluşan iki değişken seti arasındaki ilişkileri inceleyen çok değişkenli bir yöntemdir. Başka bir ifadeyle, kanonik korelasyon katsayısı birinde p diğerinde q tane değişken bulunan ($p \geq 2$ ve $q \geq 2$) iki sayısal değişken kümesi arasındaki ilişkinin derecesini belirlemek amacıyla kullanılan bir ilişki katsayısıdır.

Genel Gösterimi:

$$Y_1 + Y_2 + Y_3 + \dots + Y_p = X_1 + X_2 + X_3 + \dots + X_n$$

- Kanonik korelasyon analizi bağımlı ve bağımsız değişken setleri arasındaki ilişkiyi maksimize edecek en uygun yapıyı veya boyutluluğu belirlemeye yardım etmektedir.
- Kanonik korelasyon analizinde iki değişken setinin bağımlı ve bağımsız değişken seti gibi bir ayrıma tutulması zorunlu değildir.
- Kanonik korelasyonda kümelerdeki değişken sayısının eşit olması gerekli değildir.
- Birinci kümedeki değişken sayısı $p=1$ ve ikinci kümedeki değişken sayısı $q=1$ ise kanonik korelasyon katsayısı ile Pearson korelasyon katsayısı aynı sonucu verir.
- Birinci kümedeki değişken sayısı $p=1$ ve ikinci kümedeki değişken sayısı $q>1$ olduğu durumlarda, kanonik korelasyon katsayısı çoklu korelasyon katsayısına (R) eşit olur.

KULLANIM AMAÇLARI:

- Aynı bireyden elde edilen iki değişkenler setinin birbirinden bağımsız olup olmadığının sınanması,
- Setler arası korelasyona en çok katkıda bulunan her iki setteki değişkenlerin saptanması,
- Tahmin ve ölçüt değişkenleri setleri arasındaki korelasyonu en büyük yapan doğrusal kombinasyonların belirlenmesi,
- Bireylerin değişken setlerinin birindeki doğrusal kombinasyon değerlerine bakarak diğer setteki değerlerin tahmin edilmesi.

VARSAYIMLARI:

- Ele alınan özellikler bakımından verilerin çok değişkenli normal dağılım göstermesi,
- Üzerinde durulan özellikler bakımından ölçüm hatasının minimum seviyede olması,
- Ele alınan özellikler arasında çoklu bağlantı (multicollinearity) olmaması,
- Elde edilen sonuçlara güvenilirlik bakımından, örnek genişliğinin mümkün olduğunca büyük olması (5, 10, 20, 40 katı)

KANONİK DEĞİŞKENLER:

$$V_1 = a_{11}Y_1 + a_{12}Y_2 + \dots + a_{1p}Y_p$$

$$W_1 = b_{11}X_1 + b_{12}X_2 + \dots + b_{1k}X_k$$

Yukarıdaki eşitlikler, X ve Y değişkenlerinin doğrusal bir fonksiyonu olan V_1 ve W_1 değişkenlerini vermektedir. V_1 ve W_1 değişkenleri arasındaki korelasyon C_1 olsun. Kanonik korelasyonun amacı; $a_{11}, a_{12}, \dots, a_{1p}$ ve $b_{11}, b_{12}, \dots, b_{1k}$ katsayılarını C_1 değerini maksimum

edecek şekilde tahmin etmektir. V_1 ve W_1 kanonik değişkenleri ve C_1 kanonik korelasyon katsayısını göstermektedir.

- n gözlemden oluşan q tane bağımsız değişken (X_q) ile p tane bağımlı değişkenden (Y_p) türetilen doğrusal bileşen çiftlerine kanonik değişken adı verilir ve bunlar sırasıyla V ve W ile gösterilir.
- Kanonik değişken çiftleri arasındaki maksimum korelasyona ise kanonik korelasyon adı verilir ve C ile gösterilir.
- Kanonik korelasyon katsayıları ile basit korelasyon katsayıları benzer özelliklere sahiptir. Fakat basit korelasyonun değeri -1 ile 1 arasında değişirken kanonik korelasyon 0 ile 1 arasında değişir.
- Birinci kanonik değişkenler olan (V_1, W_1) arasındaki korelasyon C_1 ile gösterilir ve birinci kanonik korelasyon olarak adlandırılır.
- Birinci kanonik değişkenler bulunduğundan sonra bunlardan bağımsız olarak sırayla $(V_2, W_2) \dots (V_n, W_n)$ kanonik değişkenleri ve $(C_2 \dots C_n)$ kanonik korelasyonları bulunur.
- Yani n sayıda kanonik değişken seti birbirinden bağımsız ve bu değişken setleri arasındaki Kanonik korelasyonlar maksimum olacak şekilde tanımlanmaktadır.
- Bağımlı (Y) ve bağımsız (X) setlerin değişken sayılarından (k ve q) hangisi daha az ise o sayıda kanonik değişken elde edilir.

1- KANONİK KORELASYON NASIL HESAPLANIR?

Kanonik korelasyonun hesaplanma süreci; bağımlı ve bağımsız değişkenler kümesinden oluşturulan matrislerin çarpımlarından oluşan yeni matrisin öz değerlerinin ve öz vektörlerinin bulunması sürecidir.

1. Her veri setinin kendi içerisindeki korelasyon katsayıları hesaplanır ve R_{XX} ve R_{YY} ile gösterilir
2. Setler arasındaki korelasyon katsayıları hesaplanır ve R_{XY} ve R_{YX} ile gösterilir
3. Setler arasındaki korelasyon katsayıları kullanılarak $m \times n$ boyutunda bir matris oluşturulur.
4. $\begin{matrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{matrix}$ Matrisi elde edilir.
5. Oluşturulan bu matrisin öz değerleri ve öz vektörleri bulunur.
6. Kanonik korelasyonlar korelasyon matrisi yardımı ile $|R^{-1}_{YY} R_{YX} R^{-1}_{XX} R_{XY} - V| = 0$ eşitliğinin çözülmesi ile elde edilir. Buradan elde edilecek denklemin kökleri (V_j) sistemin özdeğerleridir. Bir matrisin öz değerlerini (köklerini) bulmak için yapılan bir çözüm, matristeki varyansın yeniden dağıtılması işlemidir.

2- KANONİK KORELASYONUN ANLAMLILIĞI

Bulunacak kanonik değişken çiftlerinden hangisi ya da hangilerinin anlamlı olduğu belirlenmeye çalışılır. Dikkat edilmesi gereken iki nokta bulunmaktadır.

1. Gözlem sayısı fazla olduğunda, kuvvet yönünden zayıf kanonik korelasyon katsayılarının istatistiksel açıdan anlamlı olarak bulunması söz konusu olabilmektedir. Böyle durumlarda istatistiksel anlamlılıktan çok, uygulama anlamlılığı dikkate alınarak yorum yapılması uygun olacaktır.
2. Benzer şekilde, kuvvetli kanonik korelasyonların varlığı iki değişken kümesi arasındaki ilişki katsayılarının da yüksek olacağı anlamına gelmemektedir, çünkü

kanonik korelasyon yardımıyla, her iki kümedeki değişkenlerin doğrusal bileşenleri arasındaki korelasyonlar en büyüklenmeye çalışılmaktadır.

Kanonik korelasyonların anlamlılığı için geliştirilmiş farklı yaklaşımlar vardır. Bunlardan birisi de Barlett testidir. Bu test ki-kare dağılımına dayanır. Hesaplanma süreci;

1. Hipotezler kurulur, H_0 : kanonik korelasyonlar önemsizdir, H_1 : kanonik korelasyonlar önemlidir (en az bir)
2. Anlamlılık düzeyi belirlenir
3. Test istatistiği elde edilir. Bu amaçla Wilks Lambda eşitliği elde edilir
4. İstatistiksel karar verilir.

3- KANONİK YÜKLERİN /ÖZ VEKTÖRLERİN HESAPLANMASI

Kanonik yükler, orijinal değişkenleri birleştirmek için kullanılacak ağırlıklar olup kanonik ağırlıklar olarak da bilinirler.

4- KANONİK SKORLARIN ELDE EDİLMESİ

Elde edilen kümesel ilişkiyi saçılım grafiği ile gösterebilmek için kanonik skorların elde edilmesi gerekir. Kanonik skorlar, z ile standartlaştırılmış değişkenlerin ilgili kümeye ilişkin kanonik yüklerle çarpılmasıyla elde edilir.

5- FAKTÖR YAPILARININ ELDE EDİLMESİ

Kanonik skorlar ile orijinal değişkenler arasındaki korelasyon katsayılarının hesaplanmasıdır. Bulunan korelasyon katsayılarına kanonik faktör yükleri de denir ve faktör analizindeki faktör yükleri gibi yorumlanırlar.

6- AÇIKLANAN VARYANS VE REDUNDANCY ÖLÇÜLERİ

Kanonik korelasyon katsayıları iki kümedeki değişkenlerin ağırlıklı toplamaları arasındaki korelasyon hakkında bilgi verirken, her bir kanonik korelasyon katsayısının değişkenlerdeki değişimin (varyansın) ne kadarını açıkladığı konusunda bilgi vermez. Kanonik faktör yüklerinin incelenmesiyle, istenen kanonik korelasyon için her bir kümedeki açıklanan varyans oranları elde edilebilmektedir.

Açıklanan varyans, ilgili kümedeki kanonik faktör yüklerinin kareleri toplamının ilgili kümedeki değişken sayısına bölünmesi ile elde edilir. Elde edilen karelerin ortalaması ilgili kanonik yüklere ilişkin veri kümesindeki değişkenliğin ortalama ne oranda açıkladığı konusunda bilgi verir.

Bağımsız değişken grubu için elde edilen kanonik yükler tarafından bağımlı gruptaki varyansın ne kadarının açıklanabildiği (ya da tersi) konusu Redundancy ölçüleri ile verilir.

Kümelere ilişkin açıklanan varyansların, sistemin özdeğerleri (ya da kanonik korelasyonların kareleri) ile çarpımı redundancy ölçüsünü verir.

KANONİK KORELASYONUN SPSS İLE ÇÖZÜMÜ

SPSS' te kanonik korelasyon makro dosyası aracılığı ile çözülebilmektedir.

- 1- SPSS'te analizlerde kullanılacak veri dosyası oluşturulur ve bir klasöre kaydedilir.

- 2- SPSS programının kurulu olduğu dizin altında bulunan "Canonical correlation.sps" dosyası veri dosyasının olduğu klasöre kopyalanır
- 3- SPSS programından SYNTAX editörü açılır. Buraya aşağıdaki kod yazılır

```
CD "C:\Veri Dosyalarının Bulunduğu Klasörün Yolu".  
INCLUDE "Canonical correlation.sps".  
GET FILE="üzerinde işlem yapılacak veri dosyasının adi.sav".  
CANCORR SET1= X1 X2 X3 'setler ve değişkenler tanımlanıyor'  
/SET2= Y1 Y2 Y3 Y4 Y5.
```

Örnek: Veriler İstanbul Üniversitesi İstanbul Tıp Fakültesi Hastanesine şişmanlık şikâyeti ile başvuran hastalar arasından tesadüfi olarak elde edilmiştir. Boy (cm, X1), beden ağırlığı (kg, X2), uyluk kemiğinin çevresinin uzunluğu (cm, Y1) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (cm, Y2)'nu göstermektedir. Birinci veri seti boy (X1) ve beden ağırlığı (X2), ikinci veri seti uyluk kemiğinin çevresinin uzunluğu (Y1) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (Y2) den oluşmaktadır. Veri seti 1 ve veri seti 2 arasında anlamlı bir ilişki var mıdır?

X1	X2	Y1	Y2
175,25	75,57	109,47	73,91
180,45	144,78	126,49	71,63
176,60	118,74	131,83	93,98
178,50	127,64	137,92	78,99
165,75	81,91	107,19	78,49
190,50	137,79	136,91	60,20
185,25	172,09	148,59	70,10
183,30	161,29	132,33	77,72
193,50	135,26	126,75	58,93
182,56	122,56	135,89	62,99
174,55	161,29	143,76	76,20
183,60	172,72	144,02	71,88
164,90	74,29	118,11	58,42
162,35	113,03	112,27	72,64
159,51	81,28	108,46	54,10
195,56	151,76	138,18	76,45
165,35	143,51	140,46	65,28
199,40	161,29	148,84	62,48
158,50	93,98	122,43	68,83
165,95	133,99	129,54	69,85

Çözüm:

- 1- SPSS'te makro dosyasının çalıştırılması
CD "C:\kanonik".
INCLUDE "Canonical correlation.sps".
GET FILE="obezite.sav".
CANCORR SET1= X1 X2 / SET2= Y1 Y2.

2- Set-1'deki değişkenler arası korelasyon katsayıları

Correlations for Set-1	X ₁	X ₂	R _{XX}
X ₁	1,0000	0,6484	
X ₂	0,6484	1,0000	

Yorum: X₁ ve X₂ değişkenleri arasında orta düzeyde bir ilişki vardır.

3- Set-2'deki değişkenler arası korelasyon katsayıları

Correlations for Set-2	Y ₁	Y ₂	R _{YY}
Y ₁	1,0000	0,0846	
Y ₂	0,0846	1,0000	

Yorum: Y₁ ve Y₂ değişkenleri arasında yüksek düzeyde bir ilişki vardır.

4- Set-1 ve Set-2 arasındaki korelasyonlar

Correlations Between Set-1 and Set-2	Y ₁	Y ₂	R _{XY,Y} x
X ₁	0,6307	0,0040	
X ₂	0,8781	0,1424	

Yorum: Tablo iki veri setindeki değişkenlerin birbirleriyle olan korelasyonlarını vermektedir.

X₁ ile Y₁ arasında orta düzeyde bir ilişki vardır.

X₁ ve Y₂ arasında çok zayıf bir ilişki vardır.

X₂ ile Y₁ arasında yüksek düzeyde bir ilişki vardır.

X₂ ile Y₂ arasında çok zayıf bir ilişki vardır.

5- Kanonik değişkenlerin hesaplanması

Örneğimizde bir veri setinde en az iki değişken olduğu için 2 adet kanonik değişken çifti tanımlamamız gerekmektedir. Bunlar V₁, W₁ ve V₂, W₂ kanonik değişken çiftleridir. Kanonik değişken çiftlerini hesaplayabilmek için SPSS programından yararlanılmaktadır. SPSS programı her veri seti için ham kanonik katsayıları vermektedir

Raw Canonical Coefficients for Set-1			Raw Canonical Coefficients for Set-2		
	1	2		1	2
X1	-,009	-,106	Y1	-,075	-,011
X2	-,029	,029	Y2	-,007	,108

Set-1 için ilk kanonik değişken $V_1 = -0,009X_1 - 0,029X_2$

Set-2 için ilk kanonik değişken $W_1 = -0,075Y_1 - 0,007Y_2$

İlk kanonik değişken çiftimiz (V₁ ve W₂) oluşturduk ancak diskriminant analizinde olduğu gibi, kanonik eşitliklerdeki katsayılar doğrudan birbirleri ile karşılaştırılamamaktadır. Sadece katsayıların oranları birbiriyle karşılaştırılabilmektedir. Bu katsayıların normalleştirilmesi (standardize edilmesi) gerekmektedir. V₁ eşitliğinin katsayıları 0,031 (katsayıların kareleri toplamının karekökü) ile, W₁ eşitliğinin katsayıları 0,076 değerine oranlandıktan sonra yorumlanabilir.

$V_1 = - (0,009/0,031)X_1 - (0,029/0,031)X_2$ işlemi tamamlanırsa

$V_1 = -0,290X_1 - 0,935X_2$ elde edilir.

$W_1 = -(0,075/0,076)Y_1 - (0,007/0,076)Y_2$

$W_1 = -986Y_1 - 0,092Y_2$

Hesaplanan bu birinci kanonik değişken çiftleri arasındaki basit korelasyon sayısı %88,4'tür.

İkinci kanonik değişken çiftimiz;

$V_2 = -0,106X_1 + 0,029X_2$ (ham kanonik katsayılar) **$V_2 = -0,972X_1 + 0,266X_2$**

$W_2 = -0,011Y_1 + 0,108Y_2$ (ham kanonik katsayılar) **$W_2 = -0,100Y_1 + 0,990Y_2$**

Hesaplanan ikinci kanonik değişken çiftleri arasındaki kanonik korelasyon katsayısı %12,9'dur.

6- Kanonik Korelasyon katsayıları

Canonical Correlations	
1	,884
2	,129

7- Kanonik korelasyon katsayılarının anlamlılığı

Test that remaining correlations are zero:

	Wilk's	Chi-SQ	DF	Sig.
1	,216	25,321	4,000	0,000
2	,983	,276	1,000	0,600

Yapılan Barlett testi sonucuna göre birinci kanonik korelasyon katsayısı istatistiki açıdan anlamlı iken ikinci kanonik korelasyon katsayısı anlamsız bulunmuştur. Bu durumda sadece birinci kanonik korelasyon katsayısı yorumlanacaktır.

8- Kanonik Yüklerin Hesaplanması (kanonik faktör skoru)

Canonical Loadings for Set-1

	1	2
X1	-,709	-,706
X2	-,997	,082

Canonical Loadings for Set-2

	1	2
Y1	-,998	-,066
Y2	-,150	,989

Kanonik değişkenlerin yorumlanmasında değişkenlerle kanonik değişkenler arasındaki basit korelasyon katsayıları kullanılmaktadır ve bu katsayılar **kanonik yükler/ağırlıklar** adı verilmektedir.

X_1 ve X_2 bağımsız değişkenlerinin V_1 kanonik değişkeni ile olan korelasyonları sırasıyla -%70,9 ve -%99,7; V_2 kanonik değişkeni ile olan korelasyonları ise -%70,6 ve %8,2'dir.

Benzer şekilde, Y_1 ve Y_2 değişkenlerinin W_1 kanonik değişkeni ile olan korelasyonları sırasıyla -%99,8 ve -%15; -%6,6 ve %98,9 olduğu görülmektedir.

Bu sonuçlar V_1 , V_2 , W_1 ve W_2 kanonik değişkenlerinin tanımlanmasında en önemli değişkenlerin sırasıyla X_2 (-%99,7), X_1 (-%70,9), Y_1 (-%99,8) ve Y_2 (%98,9)'dir.

Örneğimizde, ilk kanonik korelasyon katsayısı anlamlı olduğu için sadece ilk kanonik değişken çiftinin (V_1 ve W_1) yorumlanması gerekmektedir.

V_1 kanonik değişkeninin tanımlanmasında en önemli değişkenin X_2 olması ve “beden ağırlığı”nı ifade etmesi sebebiyle V_1 kanonik değişkeni de yorumlamada “beden ağırlığı” olarak adlandırılması uygundur.

$V_1 \rightarrow X_2 \rightarrow$ Beden ağırlığı

$W_1 \rightarrow Y_1 \rightarrow$ Uyluk kemiğinin çevresinin uzunluğu

Bağımsız değişkenler kümesindeki X_2 değişkeninin ilgili kanonik yükü (faktör skoru) daha fazla açıkladığı görülürken, bağımlı değişkenler kümesindeki Y_1 değişkeninin ilgili kanonik yükü daha fazla açıkladığı görülmektedir.

9- Redundancy analizleri

Proportion of Variance of Set-1 Explained by Its Own Can. Var.

	Prop Var
CV1-1	,748
CV1-2	,252

Proportion of Variance of Set-1 Explained by Opposite Can.Var.

	Prop Var
CV2-1	,584
CV2-2	,004

Proportion of Variance of Set-2 Explained by Its Own Can. Var.

	Prop Var
CV2-1	,509
CV2-2	,491

Proportion of Variance of Set-2 Explained by Opposite Can. Var.

	Prop Var
CV1-1	,397
CV1-2	,008

Redundancy (gereksizlik ölçüsü) değişken setlerinden herhangi birisinin diğerinin varyansını hangi düzeyde açıkladığını belirleyen bir ölçüdür. Gereksizlik ölçüsü her bir kanonik korelasyon için hesaplanabilmektedir. Birinci kanonik korelasyon için Y değişkenlerindeki varyansın yaklaşık olarak %39,8'i X değişkenleri tarafından açıklandığı anlaşılmaktadır. Bu değere göre, birinci kanonik korelasyon katsayısının düşük düzeyde pratik anlamlılığa sahip olduğu söylenebilir.