

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



KANONİK KORELASYON ANALİZİ ÜZERİNE BİR
İNCELEME

YÜKSEKLİSANS TEZİ

Danışman
Prof. Dr. Mehmet GÜNGÖR

Hazırlayan
Gözde MENEVŞEOĞLU

Malatya – 2019

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ

KANONİK KORELASYON ANALİZİ ÜZERİNE BİR İNCELEME

YÜKSEK LİSANS TEZİ

Hazırlayan
Gözde MENEVŞEOĞLU

Danışman
Prof. Dr. Mehmet GÜNGÖR

MALATYA 2019

T.C.
İNÖNÜ ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ

**KANONİK KORELASYON ANALİZİ
ÜZERİNE BİR İNCELEME**

YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. Mehmet GÜNGÖR

HAZIRLAYAN
Gözde MENEVŞEOĞLU

Jürimiztarihinde yapılan savunma sınavı sonucunda bu yüksek lisans tezini (oybirliği /oyçokluğu) ile başarılı bulunarak Ekonometri Anabilim dalında yüksek lisans tezi olarak kabul edilmiştir.

- | Jüri Üyelerinin Unvan Ad Soyadı | imzası |
|-----------------------------------|--|
| 1. Prof. Dr. Mehmet GÜNGÖR |  |
| 2. Dr. Öğr. Üyesi Fahrettin ÖZBEY |  |
| 3. Doç. Dr. Yunus BULUT |  |

İnönü Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Kurulunun tarih vesayılı kararıyla bu tezin kabulü onaylanmıştır.

Prof. Dr. Mehmet KUBAT
Sosyal Bilimler Enstitüsü Müdürü

ONUR SÖZÜ

Yüksek Lisans Tezi olarak sunduğum “**KANONİK KORELASYON ANALİZİ ÜZERİNE BİR İNCELEME**” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın tarafımdan yazıldığını ve yararlandığım tüm kaynakların hem metin içinde hem de kaynakçada yönteme uygun biçimde gösterilenlerden oluştuğunu, belirtir ve bunu onurumla doğrularım.

Gözde MENEVŞEOĞLU

İmza

TEŞEKKÜR

Tez çalışmam süresince çalışmalarında bana her türlü yardımı sağlayan ve ilgisini esirgemeyen danışmanım Prof. Dr. Mehmet GÜNGÖR'e sonsuz teşekkürlerimi ve saygılarımı sunarım.

Tez jürisinde yer alan üyelere ve yetişmemde katkıları olan tüm hocalarıma da minnettar olduğumu ifade etmek isterim.

Bütün yaşamım boyunca iyi ve kötü her durumda sonsuz, karşılıksız ve şartsız desteklerini hep yanımda hissettiğim, sevgili eşim, kıymetli babam ve kıymetli anneme, iş hayatımda çok büyük desteğini gördüğüm değerli çalışma arkadaşım Celile ÖZTÜRK'e

Teşekkür ederim.

ÖZET

Günümüzde gelişen teknoloji ve bilişim sistemleri sayesinde çok değişkenli veri setlerinin analizi kolaylaşmıştır. Çok değişkenli istatistiksel yöntemlerden biri olan kanonik korelasyon analizi hem metrik hem de metrik olmayan veriler için kullanılabildiği ve değişken setlerinden birinin bağımlı diğerinin bağımsız olması gibi bir zorunluluğu olmadığı için araştırmacılar tarafından sıklıkla tercih edilmektedir. Bu çalışmada kanonik korelasyon analizi yöntemi ve matematiksel çözüm süreci detaylıca anlatılmıştır. Son olarak, OECD ülkeleri sağlık risk faktörleri ile OECD ülkeleri refah göstergeleri arasındaki ilişki kanonik korelasyon analizi kullanılarak incelenmiştir.

Anahtar Kelimeler: Kanonik Korelasyon Analizi, Çok Değişkenli İstatistiksel Yöntemler, OECD Ülkeleri Sağlık Risk Faktörleri, OECD Ülkeleri Refah Göstergeleri.

ABSTRACT

Nowadays, thanks to developing technology and information systems, the analysis of multivariate data sets has become easier. canonical correlation analysis, which is one of the multivariate statistical methods, is frequently preferred by researchers because it can be used for both metric and non-metric data, and one of the variable sets does not have to be dependent on the other. In this study, canonical correlation analysis method and mathematical solution process are explained in detail. Finally, the relationship between health risk factors in OECD countries and welfare indicators in OECD countries was examined using canonical correlation analysis.

Keywords: Canonical Correlation Analysis, Multivariate Statistical Methods, OECD Countries Health Risk Factors, OECD Countries Welfare Indicators.

ÖZGEÇMİŞ

1985 yılı Tokat doğumluyum. Orta öğrenimimi Malatya Turgut Özal Anadolu Lisesinde tamamladım. 2006 yılında Karadeniz Teknik Üniversitesinde İstatistik ve Bilgisayar Bilimleri Bölümünden mezun oldum.

2007-2012 yılları arasında Yapı ve Kredi Bankasında Portföy Yönetmeni olarak görev yaptım. Ardından 2014 yılında Kamu Personeli Seçme Sınavına girerek 2016 yılında T.C. Sağlık Bakanlığı kadrosuna İstatistikçi olarak atandım. Halen Adana İl Sağlık Müdürlüğünde İstatistikçi olarak görev yapmaktayım.



İÇİNDEKİLER

ONUR SÖZÜ	iii
TEŞEKKÜR	iv
ÖZET	v
ABSTRACT.....	vi
ÖZGEÇMİŞ	vii
İÇİNDEKİLER	viii
TABLolar	x
ŞEKİLLER.....	xi
SİMGELER VE KISALTMALAR	xii
1.GİRİŞ	.1
2. ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ TEKNİKLERİNE GENEL BAKIŞ	4
2.1. Çok Değişkenli İstatistiksel Analiz Teknikleri ile İlgili Temel Kavramlar	4
2.1.1. Bağımlı Çok Değişkenli İstatistik Teknikleri	6
2.1.2. İç Bağımlı Çok Değişkenli İstatistik Teknikleri.....	11
2.2. Çok Değişkenli İstatistiksel Analiz Tekniklerinin Varsayımları.....	12
2.3. Çok Değişkenli İstatistiksel Analiz Tekniklerinin Amaçları	13
2.4. Doğrusallık	14
2.4.1. Doğrusal Olmayan Çok Değişkenli İstatistiksel Analiz Teknikleri.....	14
2.5. Çok Değişkenli İstatistiksel Teknikler İçin Bazı Tanımlar.....	15
2.5.1.Standartlaştırma	15
2.5.2.Normallik	16
2.5.3. Uç Değerler	16
2.5.4. Eşvaryanslılık	17
2.5.5. Evren ve Örneklem	17
2.5.6. İstatistiksel Güç	17
2.5.7. Betimsel ve Yordamsal İstatistikler	17
2.5.8. Varyans Kovaryans Matrisi.....	18
2.5.9. Anlamlılık Testleri	18

2.5.10. Güven Aralığı	18
3.KANONİK KORELASYON ANALİZİ YÖNTEMİ.....	19
3.1. Basit Korelasyon Analizi	19
3.2. Çoklu Korelasyon Analizi	20
3.3. Kanonik Korelasyon Analizi	21
3.3.1. Kanonik Korelasyon Analizinin Avantajları ve Dezavantajları	23
3.3.2. Kanonik Korelasyonun Tarihçesi ve Kullanım Alanları	24
3.3.3. Kanonik Korelasyon Analizinin Amaçları	27
3.3.4. Kanonik Korelasyon Analizinin Varsayımları.....	28
3.3.5. Kanonik Korelasyon Analizi Temel Kavramları.....	29
3.4. Doğrusal Olmayan Kanonik Korelasyon Analizi	31
4.KANONİK KORELASYON ANALİZİNİN MATEMATİKSEL TANIMI	33
4.1. Kanonik Değişkenler ve Kanonik Korelasyonlar	34
4.2. Kanonik Değişkenler ve Kanonik Korelasyonların Hesaplanması.....	34
4.3 Korelasyon Matrisi ile Kanonik Korelasyon Hesaplaması	39
4.4 Kanonik Değişkenlerle Orijinal Değişkenler Arasındaki Korelasyonlar	42
4.5. Kanonik Korelasyon Analizinde Katsayıların Anlamlılık Testi.....	43
4.6. Kanonik Ağırlıklar ve Kanonik Yükler.....	44
4.7. Redundancy (Açıklanabilirlik) İndeksinin Hesaplanması.....	44
4.8. Kanonik Korelasyon Hesaplama Örneği	46
5.UYGULAMA	49
5.1. Veri Toplama.....	49
5.2. Veri Tablosu	49
5.3. Doğrusal Kanonik Korelasyon Analizi Uygulaması	51
5.4. Kanonik Korelasyon Analizi Bulgularının Yorumlanması	52
5.5. Sonuç.....	59
EKLER	66
Ek1: SPSS Kanonik Korelasyon Analizi Output Çıktısı.....	66
KAYNAKÇA	69

TABLÖLAR

Tablo 2.1:	Bağımlı Teknikler Arası İlişkiler Tablosu	7
Tablo 2.2:	ANOVA / MANOVA Tablosu	9
Tablo 4.1:	Örnek Korelasyon Matrisi	40
Tablo 4.2:	Kanonik Korelasyon Hesaplaması Örneği İçin Kullanılan Veri Seti....	46
Tablo 4.3:	Tablo 4.2’ deki Veri Seti İçin Korelasyon Matrisi	47
Tablo 5.1:	Uygulama Veri Setleri	50
Tablo 5.2:	Kanonik Katsayıların Anlamlılık Testi	53
Tablo 5.3:	Uygulama Değişkenleri Arası Korelasyonlar	53
Tablo 5.4:	Kanonik Korelasyon Tablosu	54
Tablo 5.5:	1. Veri Seti Kanonik Korelasyon Katsayıları	55
Tablo 5.6:	2. Veri Seti Kanonik Korelasyon Katsayıları	55
Tablo 5.7:	1. Veri Setine Ait Kanonik Yükler ve Çapraz Kanonik Yükler	56
Tablo 5.8:	2. Veri Setine Ait Kanonik Yükler ve Çapraz Kanonik Yükler	56
Tablo 5.9:	Kanonik Korelasyonlar ve Kanonik Öz Değerler	57
Tablo 5.10:	1. Set Redundancy İndeks Tablosu	58
Tablo 5.11:	2. Set Redundancy İndeks Tablosu	58

ŞEKİLLER

Şekil 2.1	:	Uygun Çok Değişkenli Analiz Tekniğinin Seçim Diyagramı ...	5
Şekil 2.2	:	Doğrusal Olmayan İlişki Grafiği	14
Şekil 2.3	:	Normal Dağılım Grafiği	16
Şekil 3.1	:	Pozitif-Negatif-Nötr İlişki Grafikleri	19
Şekil 3.2	:	Kanonik Korelasyon Analizi Genel Yapısı	22
Şekil 5.1	:	SPSS Ana Ekran Görüntüsü	52



SİMGELER VE KISALTMALAR

ANOVA	Tek Bağımlı Değişkenin Varyans Analizi
DOKKA	Doğrusal Olmayan Kanonik Korelasyon Analizi
KKA	Kanonik Korelasyon Analizi
M	Metrik
MANOVA	Birden Fazla Bağımlı Değişkenin Varyans Analizi
MO	Metrik Olmayan
OECD	İktisadi İşbirliği ve Gelişme Teşkilatı
vb.	Ve benzeri
vd.	Ve diğerleri

1. BÖLÜM

GİRİŞ

Kanonik Korelasyon Analizi (KKA), çok sayıda değişkeni olan iki değişken seti arasındaki ilişkinin incelenmesinde kullanılan çok değişkenli analiz metotlarından birisidir. Bu analiz yönteminde, değişken setlerinden birisinin bağımlı, diğerinin bağımsız olması gibi bir zorunluluk yoktur. Ancak değişken setlerinin birisi bağımlı diğeri bağımsız olarak da analizde ele alınabilir. Her bir sette olan değişkenlerin doğrusal bileşimlerinden yeni değişkenler yaratılmaktadır. Bu yeni değişkenler, kanonik değişken olarak adlandırılmaktadır ve yeni değişkenler arasındaki korelasyonun maksimum olması hedeflenmektedir.

Bu çalışmanın amacı, KKA ile ilgili kavramsal çerçevenin ortaya konularak, ulusal ve uluslararası düzeyde bu konuda yapılmış akademik ve bilimsel çalışmaların irdelenmesidir. Yapılan literatür taramasında KKA' nin sıklıkla kullanıldığı görülmektedir. Çünkü, setler arası korelasyona en çok katkıda bulunan değişkenlerin saptanması akademik çalışmalar için önem arz etmektedir. Bu nedenle yapılacak olan bu çalışmanın hem akademik literatüre hem de ekonometrik alana katkı sağlaması amaçlanmaktadır.

KKA, iki değişken seti arasındaki ilişkileri ortaya koymada güçlü bir yöntem olarak bilinmektedir. Bu analiz yöntemi, iki değişken setinin doğrusal fonksiyonlarını en yüksek korelasyon gösterecek biçiminde bulmaya çalışmaktadır. Bu anlamda KKA' nin önce teorik çerçevesinin irdelenmesi gerekmektedir.

KKA 'nin daha iyi anlaşılması ve kavranması açısından çalışma, 'Giriş', "Çok Değişkenli İstatistiksel Analiz Tekniklerine Genel Bakış", 'Kanonik Korelasyon Analizi Yöntemi', "Kanonik Korelasyon Analizinin Matematiksel Tanımı" ve "Uygulama" ana başlıkları altında beş kısımda incelenmektedir.

Çalışmanın birinci bölümünde, çalışmanın önemi, amacı ve kapsamı hakkında genel bir giriş yapılmıştır. Çalışmanın ikinci bölümünde, çok değişkenli istatistiksel analiz teknikleri genel bir çerçevede verilmiştir. Bununla birlikte bazı faydalı tanımlardan bahsedilmiştir. Çalışmanın üçüncü bölümde, KKA ile ilgili teorik çerçeve tanımlanıp, basit ve çoklu korelasyon analizleri hakkında temel bilgiler verildikten sonra, KKA' nin tarihçesi, temel prensipleri, amaçları ve kullanım alanları anlatılmıştır. Üçüncü bölümün

sonunda da doğrusal olmayan kanonik korelasyon analizi (DOKKA) tanımlanmıştır. Çalışmanın dördüncü bölümde ise, KKA matematiksel çerçevede değerlendirilmiş olup kanonik korelasyon hesaplama aşamaları anlatılmıştır. Çalışmanın uygulama bölümde ise, KKA ile OECD ülkelerinin sağlık risk faktörleri ve refah göstergeleri arasındaki ilişkiyi inceleyen bir araştırmaya yer verilmiştir.

Çalışmanın uygulama bölümünde sağlık risk faktörleri veri setinin seçilme nedeni, bulaşıcı olmayan hastalıkların görülme sıklığının ve ölüm nedenleri içindeki payının arttırmasıdır. Mevcut bilgiler ve araştırmalar bulaşıcı olmayan hastalıkların önlenabilir olduğunu göstermektedir. Son yıllarda gözlenen kalitesiz yaşam tarzı, bulaşıcı olmayan hastalıkların görülme riskini arttırmaktadır.

Bulaşıcı olmayan hastalıkların birçoğu sigara ve alkol kullanımı ile sağlıksız beslenme ve yetersiz aktivite gibi temel risk faktörlerine bağlı gelişmektedir. Bulaşıcı olmayan hastalıklarla mücadele ederken kişilere ve topluma göre ortak bakım sağlanmalıdır.

Küresel Bulaşıcı Olmayan Hastalıklar Raporu'nda da belirtildiği gibi gelişmekte olan ülkelerde bulaşıcı olmayan hastalıklar için sürveyans sistemi ya yoktur ya da çok kısıtlıdır. (T.C. Sağlık Bakanlığı Kütüphanesi-Türkiye Kronik Hastalıklar ve Risk Faktörleri Sıklığı Sunumu)

Türkiye Cumhuriyeti Sağlık Bakanlığı 2018–2022 Stratejik Planında kurum misyonu şu şekilde tanımlanmıştır:

“İnsan merkezli bir yaklaşımla birey ve toplum sağlığının en üst düzeyde korumak ve sağlık sorunlarına zamanında uygun ve etkili çözümleri yüksek hizmet kalitesi ile sunmak.”

2018-2022 Stratejik Planında vurgulanan bu misyon doğrultusunda Sağlık Bakanlığı tarafından belirlenen altı amaçtan biri de sağlığa yönelik çevresel risklerden bireyleri korumak ve sağlıklı hayat tarzını teşvik ederek yaygınlaştırmaktır. Bir başka deyişle, bulaşıcı olmayan hastalıklardan kaynaklanan olumsuzlukların azalması hedeflenmiştir. Tüm Dünyadaki ve ülkemizdeki sağlık kuruluşlarınca önemsenen bireysel korunma ve sağlıklı yaşam tarzının benimsenmesi, sağlığın geliştirilmesinin en önemli adımlarından biridir.

Dünya genelinde ve ülkemizde sağlık risk faktörleri obezite, tütün kullanımı ve alkol kullanımı olarak üç ana başlıkta incelenmektedir. Refah göstergesi başlığı altında

ise birçok gösterge yer almaktadır. Bu çalışmada ise kişi başı gayrisafi yurtiçi hâsıla, bireysel internet kullanımı, yaşam tatmini ve eğitimsel kazanım değişkenleri refah göstergesi veri seti olarak değerlendirilmiştir. Kullanılan iki veri seti de OECD (İktisadi İşbirliği ve Gelişme Teşkilatı) Library kütüphanesinin 2017 yılı raporlarından alınmıştır. Çalışma OECD ülkelerinin sağlık risk faktörleri ile refah göstergeleri arasındaki ilişkinin KKA yardımıyla incelenmesini kapsamaktadır. Çalışmanın sonuç bölümünde ise bu çalışmada elde edilen bulgular özetlenmiştir.



2. BÖLÜM

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ TEKNİKLERİNE GENEL BAKIŞ

Çok değişkenli istatistiksel yöntemler çoklu veri setlerini analiz etmede kullanılan ve kullanım sıklığı gittikçe artan tekniklerdir. Karmaşık araştırma sonuçlarına tek değişkenli analiz yöntemlerinin cevap vermesi oldukça güçtür. Ancak günümüzde çok değişkenli istatistiksel yöntemler oldukça gelişmiştir. Çünkü istatistiksel paket programların içinde hazır olarak bulabildiğimiz bu teknikler sayesinde karmaşık araştırma sonuçlarının üstesinden kolaylıkla gelinebilmektedir.

Bu bölümde, çok değişkenli istatistiksel analiz teknikleri, çeşitleri, varsayımları ve amaçları açıklanmaktadır. Bununla birlikte doğrusallık kavramı incelendikten sonra, doğrusal olmayan istatistiksel teknikler ve bazı faydalı tanımlar verilmektedir. Bu bölüm, çalışmamızın ana konusu olan KKA için bir altyapı özelliği taşımaktadır.

2.1. Çok Değişkenli İstatistiksel Analiz Teknikleri ile İlgili Temel Kavramlar

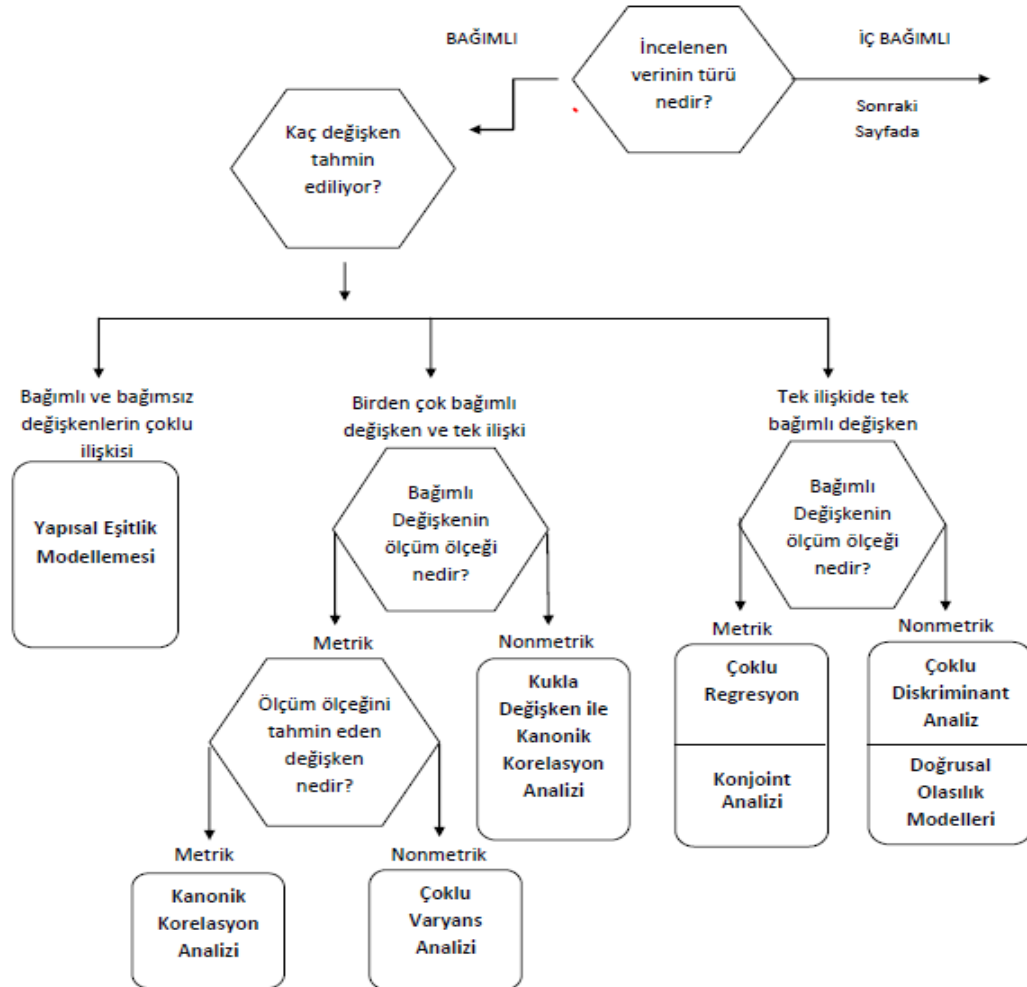
Çok değişkenli analiz yöntemleri ile ilgili çeşitli tanımlar yapılmıştır. Basit tanımı ile, örnek üzerinde ikiden fazla değişkeni eş zamanlı çözümleyen tüm istatistik tekniklerine verilen isimdir (Sheth, 1971:157-162). Çok değişkenli analiz yöntemleri, değişken grupları arasındaki karşılıklı ilişkileri ölçme ve açıklama imkânı veren istatistik yöntemlerdir (Gatty, 1957-162). Çok değişkenli analizlerde aynı anda birden fazla bağımlı ya da bağımsız değişkenle ilgilenilmektedir. İstatistiğin önemli konularından biri olan çok değişkenli teknikler Ekonomi, İktisat, Tıp, Psikoloji, Ziraat, Jeoloji, Kimya, Meteoroloji, Çevre, Sağlık gibi çok çeşitli alanlarda kullanılmaktadır.

Günümüzde araştırmacılar elde ettikleri verilerle belli hesaplamaları kolaylıkla yapabilmektedirler. Ancak asıl önemli olan elimizdeki veriler için güvenilir ölçmeleri kullanabilmek, uygun testleri seçebilmek ve analiz sonucunu doğru yorumlayabilmektir. Bu nedenle sınıflara ayrılmış olan çok değişkenli tekniklerden, veri setimize en uygun olan analizi seçmek doğru sonuca ulaşabilmek adına büyük önem taşımaktadır. Sınıflandırma yapılırken, ölçeğin türü, değişken sayısı ve değişkenler arasındaki bağımlı

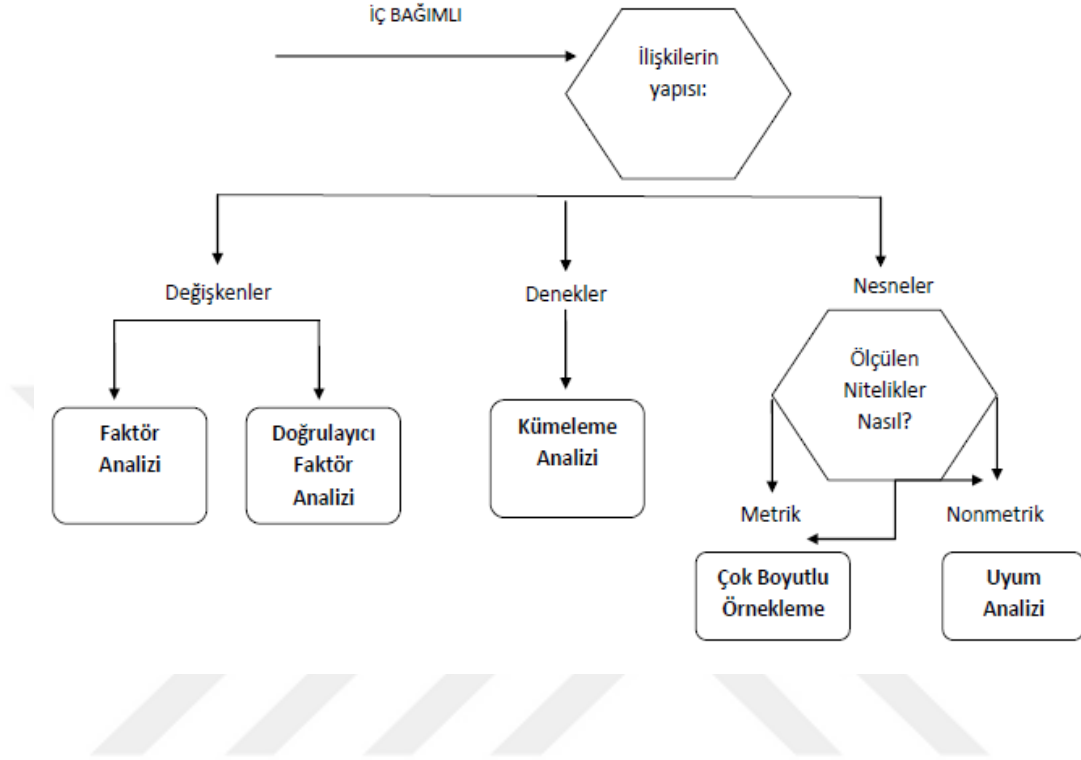
bağımlı değişken ayrımının yapıp yapılmaması kullanılan en yaygın ölçütlerdendir. Değişkenlerin iki ayrı veri seti olarak ele alındığı ve bu iki set arasındaki ilişkinin araştırıldığı tekniklere **bağımlı teknikler** adı verilmektedir. Bağımlı teknikler, bağımlı-bağımsız değişken sayısı ve bağımlı-bağımsız değişken ölçeğine (metrik veya metrik olmayan) göre sınıflandırılabilir. Değişkenler arasında bağımlı bağımsız değişken ayrımı yapmayan istatistik tekniklere ise **iç bağımlı teknikler** adı verilmektedir. İç bağımlı tekniklerin amacı, değişkenler arası ilişkilerin niçin ve nasıl gerçekleştiğini belirlemektir (Albayrak, 2006 :4-5).

Çok değişkenli veri setleri için uygun çok değişkenli analiz tekniğinin seçilmesine yardımcı olacak diyagram Şekil 2.1 'de yer almaktadır.

Şekil 2.1: Uygun Çok Değişkenli Tekniğin Seçilmesi (Hair J.F. vd. Multivariate Data Analysis Seventh Edition, 2014:11-12)



Şekil 2.1 (devamı) : Uygun Çok Değişkenli Tekniğin Seçilmesi (Hair J.F. vd. Multivariate Data Analysis Seventh Edition, 2014:11-12)



2.1.1. Bağımlı Çok Değişkenli İstatistik Teknikleri

Belirtilen tanım doğrultusunda KKA, Yapısal Eşitlik Modeli, Çok Değişkenli Varyans Analizi, Çoklu Regresyon Analizi, Lojistik Regresyon Analizi, Çoklu Diskriminant Analizi ve Konjoint Analizi bağımlı model sınıflamasında yer almaktadır. KKA' nin daha net anlaşılması için bağımlı teknikler sınıfındaki diğer analizlerin de incelenmesi gerekmektedir. Tablo 2.1 'de bağımlı teknikler arası ilişkiler, ölçütler esas alınarak verilmiştir.

Tablo 2.1: Bağımlı Teknikler Arası İlişkiler (Hair vd. , 1998:19)

1	KKA	$b_{i1}Y_{i1} + b_{i2}Y_{i2} + \dots + b_{ip}Y_{ip} \Leftrightarrow a_{i1}X_{i1} + a_{i2}X_{i2} + \dots + a_{iq}X_{iq}$ (M, MO) r_{ci} (M,MO)
2	MANOVA	$b_1Y_1 + b_2Y_2 + \dots + b_pY_p = a_1X_1 + a_2X_2 + \dots + a_qX_q$ (M) (MO)
3	Çift Yönlü ANOVA	$Y = a_1X_1 + a_2X_2 + \dots + a_qX_q$ (M) (MO)
4	Çoklu Diskriminant Analizi	$Y = a_1X_1 + a_2X_2 + \dots + a_qX_q$ (MO) (M)
5	Lojistik Regresyon Analizi	$Y = Xa_1X_1 + a_2X_2 + \dots + a_qX_q$ (MO) (M , MO)
6	Çoklu Regresyon Analizi	$Y = a_1X_1 + a_2X_2 + \dots + a_qX_q$ (M) (M,MO)
7	Konjoint Analizi	$Y = a_1X_1 + a_2X_2 + \dots + a_qX_q$ (M, MO) (MO)
8	Yapısal Eşitlik Modeli	$Y_1 = a_{11}X_{11} + a_{12}X_{12} + \dots + a_{1q}X_{1q}$ $Y_2 = a_{21}X_{21} + a_{22}X_{22} + \dots + a_{2q}X_{2q}$ $\vdots$ $Y_p = a_{p1}X_{p1} + a_{p2}X_{p2} + \dots + a_{pq}X_{pq}$ (M) (M, MO)
NOT: ‘M= Metrik’ ve ‘MO= Metrik Olmayan’ ifadelerinin kısaltmalarıdır.		

- **KKA:** Çok Değişkenli İstatistiksel Yöntemlerden biri olan KKA bağımlı teknikler sınıfında yer almaktadır. Çalışmamızın ana metodu olan KKA üçüncü ve dördüncü bölümde detaylıca anlatılacaktır. KKA' ni basit bir tanımla ifade etmek gerekirse iki veri seti arasındaki ilişkinin analiz edilmesidir denilebilir. Her iki veri setinin ölçüldüğü analizde, bu setlerin birbirleri ile nasıl bir ilişki içerisinde olduğunu açıklayan bir istatistiksel tekniktir.

- **Varyans Analizi:** Varyans analizi normal dağılımlı bir veri setinde iki veya daha fazla bağımsız ortalama arasında anlamlı bir fark olup olmadığını tespit etmekte kullanılmaktadır. Varyans analizinde analiz tipi, veri setinin özelliklerine ve analiz edilen bağımlı ve bağımsız değişkenlerin sayısına göre değişmektedir. Analiz edilen tek bağımlı değişken olması durumunda ANOVA, birden fazla bağımlı değişken olması durumunda ise, MANOVA kullanılmaktadır.

- **Çift Yönlü ANOVA:** Tek yönlü ANOVA varyansın en basit analizidir. Tek yönlü ANOVA tek değişkenli tekniklere girmektedir. Bu nedenle Çift yönlü ANOVA' ya değinmek daha doğru olacaktır. ANOVA 'nın iki temel varsayımı vardır. Bu varsayımlara göre, her grup normal dağılımdan gelmektedir ve grupların varyansları eşittir. İki veya daha fazla bağımsız değişkenin, bir bağımlı değişkenin analizinde ise çift yönlü ANOVA kullanılmaktadır.

- **Tek ve Çift Yönlü MANOVA:** Tek yönlü MANOVA analizi bağımsız bir değişkenin çoklu bağımlı değişkenler üzerindeki etkisini incelemektedir. Varsayımları ANOVA' nın varsayımları ile aynıdır. Çift yönlü MANOVA' da ise iki ya da daha fazla bağımsız değişkenin, birden fazla bağımlı değişken üzerindeki etkisi incelenmektedir. Bağımlı değişkenler sürekli verilerden oluşurken, bağımsız değişkenler birden fazla düzeyden oluşmaktadır. İki yönlü ANOVA ve tek yönlü MANOVA karışımı gibidir.

Tablo 2.2 : ANOVA / MANOVA Tablosu

ANOVA / MANOVA TABLOSU		Bağımsız Değişken Sayısı	
		Bir	Birden Fazla
Bağımlı Değişken Sayısı	Bir	Tek Yönlü ANOVA	Çift Yönlü ANOVA
	Birden Fazla	Tek Yönlü MANOVA	Çift Yönlü MANOVA

- **Diskriminant Analizi:** Ayırma analizi olarak da bilinen, kategorik bağımlı değişkenler ve metrik bağımsız değişkenler arasındaki ilişkileri tahmin etmeyi hedefleyen çok değişkenli istatistiksel tekniklerden birisidir. Grup ortalamaları arasındaki farklılığı maksimize edecek doğrusal kombinasyonları bulma işlemidir (Alpar (2013:723). İki gruplu diskriminant analizi birimlerin çok sayıda değişkene göre iki ana kütleli birbirinden ayırma problemi üzerinde durmaktadır. Bu analizde ana kütle grupları önceden belirlendikten sonra, bu iki ana kütle ile ilgili özellikler ölçülmektedir. Diskriminant analizi mevcut ya da gelecekte elde edilecek birimleri iki gruptan birine atamak için bir ölçüt (diskriminant fonksiyonu) geliştirir (Albayrak, 2006:309). Bu analize örnek vermek gerekirse arıların kanat uzunluğu, bacak uzunluğu, vücut büyüklüğü vb. özelliklerinin elde edilmesi ile ırk özelliği belli olan sınıflardan hangisine ait olduğunun belirlenmesi söylenebilir.
- **Çoklu Regresyon Analizi:** Bağımlı ve bağımsız değişkenler arası ilişkiyi matematiksel modellerle açıklayarak bulmayı amaçlayan çok değişkenli bir istatistiksel yöntemdir (Alpar, (2013:415)). Çoklu Regresyon birkaç bağımsız değişkeni kullanarak bir bağımlı değişkenin puanlarını kestirmede kullanılır (Tabachnick, Fidell, 2015: 18). Örneğin voleybol oynama becerisi ile beden eğitimi süresi, boy uzunluğu, vücut esnekliği, yaş gibi birkaç değişken arasındaki bağın büyüklüğünün ölçülmesi. Bulunan regresyon denklemi değişik amaçlarda kullanılabilir.
- **Lojistik Regresyon Analizi:** Bağımlı ve bağımsız değişken ayrımının yapıldığı çok değişkenli bir modelde, bağımlı değişken bir değişken ölçek değişkeni olduğunda, en küçük kareler tekniğiyle elde edilen tahminler yetersiz kalmaktadır.

Dolayısıyla, öngörülen varyanslar minimum olmadığından ve ayırıştırma işleminde ayırmacı analiz varsayımları sağlanmadığında lojistik regresyon analizi kullanılmaktadır.

Lojistik regresyon analizi özellikle bağımlı değişkene verilen tepkilerin dağılımının bir veya birden fazla bağımsız değişken ile doğrusal olmayan ilişki göstermesinin beklendiği durumlarda kullanışlıdır. Lojistik regresyon araştırmacıya yordayıcı değişkenlerin değerlerinin kombinasyonu ile gruplardan birinde üye olma olasılığını değerlendirme imkânı verir (Tabachnick, Fidell, 2015: 24, 439).

- **Konjoint Analizi:** Bu ifade ilk olarak 1974 yılında Richard M. Johnson tarafından kullanılmıştır. Konjoint analize, açıklayıcı olabilecek, çağrışım yapılabilecek Türkçe karşılık verilmek istenirse “İlişkilerin Analizi” veya “İlişkilendirme Analizi” denilebilir (Sönmez, 2006:186). Konjoint Analizini diğer pek çok istatistiksel analiz tekniğinden ayıran özelliği, nitelikleri nicel olarak karşılaştırılabilir imkânı sunmasıdır. Bu analiz sayesinde, sayısal olarak ifade edilmeyen özellikler (renk gibi), sayısal olarak ifade edilebilen verilere dönüştürülür. Belli bir ürün ya da hizmet için belirlenen değişkenler ve bu değişkenlerin düzeyleri arasındaki ilişki ile değişkenlerin önem düzeyleri sayısal olarak ifade edilebilen verilere dönüştürülür.
- **Yapısal Eşitlik Modeli:** Yapısal eşitlik modeli, faktör analizi, KKA ve çoklu regresyon analizlerini birleştirir. Faktör analizi gibi bazı verileri örtük bazıları direk ölçülebilen olabilir. KKA gibi birçok bağımlı bağımsız değişken barındırabilir. Çoklu regresyondaki gibi amaç birçok değişken arasındaki ilişkinin incelenmesi olabilir. Bu teknik, sunulan modelin veri ile kabul edilebilir bir uyum sunup sunmadığını ve her bir bağımsız değişkenin bağımlı değişkene katkısını değerlendirir. Alternatif modeller arası karşılaştırmalar ve gruplar arası farklılıkların değerlendirilmesi de bu analizde mümkündür (Tabachnick, Fidell, 2015: 26).

2.1.2. İç Bağımlı Çok Değişkenli İstatistik Teknikleri

Analize konu olan değişkenler için bağımlı ve bağımsız gibi bir ayrım yapılmadan eş zamanlı analiz edildiği tekniklere iç bağımlı teknikler adı verilmektedir. Faktör analizi, kümeleme analizi, uyum analizi ve temel bileşenler analizi iç bağımlı tekniklere verilebilecek örnek analizlerdir (Albayrak, 2006 ; 3).

- **Faktör Analizi:** Faktör analizi, özellikle çok karmaşık ve çok boyutlu ilişki analizi ile karşılaşılması durumunda kullanılabilecek bir yöntemdir. Aynı yapıyı ölçen çok sayıda değişkenden önemli sayıda tanımlanabilir değişken elde etmek için kullanılan çok değişkenli bir istatistiktir. Çok sayıda değişkenden oluşan yapıyı tanımlamamızı sağlar. Değişkenlerin daha anlamlı, kolay, anlaşılır ve özet biçimde yorumlanmasına yardımcı olur (Kleinbaum vd.,1994:601-602). Faktör analizinde, yüksek değişkenli bir dizi kümesi birleştirilerek, faktör adı verilen genel değişkenler (faktörler) oluşturmak mümkündür.

Bu analizin temel amaçları;

- Değişken arası ilişkileri en iyi açıklayan az sayıdaki faktör sayısını belirlemek,
- Değişkenler arası ilişkinin kökenini belirlemek,
- Çok sayıdaki karmaşık verilerin arkasında yatan gerçek nedenleri ve gizli boyutları ortaya çıkarmak,
- Yapıyı değişkenler arasındaki ilişkilerde ortaya koymak,
- En az bilgi kaybı ile analizi gerçekleştirmek,

olarak sayılabilir.

- **Kümeleme Analizi:** Verilerin homojen alt sınıflara ayrılmasına kümeleme analizi denilmektedir. Uzaklık-Benzerlik ölçüleri yardımı ile birbirine benzer gözlemlerin kümelenmesidir (Alpar, 2013:386). Birincil amaç, bireylerin veya nesnelerin temel özelliklerini göz önünde bulundurarak gruplandırma yapmaktır. Yani kümeleme analizi, yapıları hakkında kesin bilgilerin bulunmadığı bir veri yığını içindeki birimleri, değişkenleri ya da birim ve değişkenleri birbiri ile

benzer olan alt kümelere (grup, sınıf) ayırma yöntemidir. Kümeleme analizi küme ayırma işleminin ne kadar uygun yapıldığını analiz eder (Kılıç, 2016; 3).

- **Uyum Analizi:** Uyum Analizine sıklıkla iki kategorik değişkenin kategorileri arasındaki ilişkileri çözümlemek amacıyla başvurulur. İki'den fazla kategorik değişken arasındaki ilişkilerin eşanlı olarak çözümlenmesinde de kullanılmaktadır (Alpar, 2013:365). Uyum analizinin ana amacı karmaşık yapıdaki bir veri matrisini önemli bir veri kaybı yaratmayacak şekilde daha basit yapıdaki yeni bir veri matrisiyle ortaya koymaktır.
- **Temel Bileşen Analizi:** Temel bileşen analizi genellikle değişkenler arasındaki bağımlılık yapısını ortadan kaldırmak veya boyut yapısını azaltmak ve çok sayıda değişken kullanmak için kullanılmaktadır. Gözlemler arasında bir endeks oluşturmak için bu analiz tekniği tercih edilmektedir. Faktör analizinden tek farkı dönüşüm işleminin uygulanmadığı, sadece çoklu bağlantının boyutunun küçültülmesi işleminin yapılmasıdır.

2.2. Çok Değişkenli İstatistiksel Analiz Tekniklerinin Varsayımları

Çok değişkenli istatistik tekniklerinin hepsi bazı varsayımlar gerektirmektedir. Bu varsayımlar bazı durumlarda araştırmacılar için büyük bir zorluk oluşturabilmektedir. Ancak analizin güvenliği de bu varsayımlara bağlıdır (Sharma, 1996:374). Çok değişkenli istatistiksel teknikler için gerekli varsayımlar şunlardır (Albayrak, 2006: 49):

- **Çoklu Normal Dağılım:** Çok değişkenli istatistik yöntemlerin çoğu, örneklerin çok değişkenli normal dağılımlı ana küleden geldiğini varsaymaktadır. Bu varsayım, bazı yorumları kolaylaştırmanın ötesinde dağılım teorisi açısından gereklidir (Tatlídil, 1996:53).
- **Sabit Varyans:** Tüm gruplar için kovaryans matrislerinin eşit olduğu varsayımdır.
- **Çoklu Doğrusal Bağlantı:** Bağımsız değişkenler arasında anlamlı doğrusal ilişkinin olmadığı varsayımdır.

- **Doğrusallık:** Bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusal olması varsayımdır.
- **Otokorelasyonun Olmaması:** Değişken birim değerlerinin birbirinden bağımsız olması varsayımdır.

Yukarıdaki varsayımlardan sapma istatistik testlerin gücünü ve anlamlılığını etkilemektedir.

2.3. Çok Değişkenli İstatistiksel Analiz Tekniklerinin Amaçları

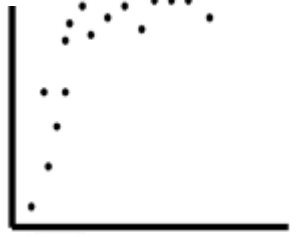
Çok değişkenli istatistiksel analizler, problemin niteliğine göre problemi analiz etmek ve çevredeki iç ve dış faktörleri göz önüne almak için geliştirilmiş bir dizi yöntemdir. Çok değişkenli istatistiklerin uygulanmasındaki temel amaçlar dört sınıfta kategorize edilmektedir (Akyıldız, 2006: 3).

- **Veri Azaltma:** Veri azaltmanın amacı, veri yapısını, p sayısını içeren ve bunlarla ilişkili olmayan veri setinin varyasyonunu tanımlayan daha az sayıda değişken ($k < p$) ile açıklamaktır.
- **Kümeleme ve Sınıflandırma:** Kümeleme ve sınıflandırmadaki amaç, bilinmeyen popülasyon özelliklerine sahip yapılar hakkında prototip kümelerinin (gruplar, sınıflar) belirlenmesine yardımcı olmak ve önceden belirlenmiş gruplara yeni birimler atamaktır.
- **Ölçeklendirme:** Ölçeklendirmenin amacı, birçok değişkenli p boyutlu ölçümlerden daha az değişken kullanarak birimleri göstermektir ve birimlerin birbirleriyle olan benzerlik ve farklılıklarını $k < p$ boyut skalasında incelemek hedeflenmektedir.
- **Çok Değişkenli Hipotezlerin Test Edilmesi:** Çok değişkenli analiz yöntemleri, K toplumunun çok değişkenli ortalama vektörünün eşitliği/farkı üzerine kurulacak hipotezleri test etmek için kullanılmaktadır.

2.4. Doğrusallık

Doğrusallık, matematiksel bir ilişkinin veya işlevin özelliğidir; bu, grafik olarak düz bir çizgi olarak gösterilebileceği anlamına gelmektedir. Bir $f(x)$ işlevi, x değişkeninin bağımsız bir fonksiyonunun doğrusal bir fonksiyonudur. Doğrusallık, en temel ve en çok kullanılan öngörücü analizdir. Regresyon tahminleri verileri tanımlamak ve bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi açıklamak için kullanılmaktadır. Regresyon analizinin merkezinde, dağılım grafiği boyunca tek bir çizgiyi yerleştirme görevi söz konusudur. Bir bağımlı ve bir bağımsız değişkeni olan en basit form, $y = a + b * x$ formülü ile tanımlanmaktadır (Kang vd. 2003: 974).

Şekil 2.2: Doğrusal Olmayan İlişki Grafiği



Pai vd. (2015:15467)'e göre doğrusallık, ölçme araçlarının işletme kapsamındaki önyargısıdır. Başka bir deyişle, ölçülen değer ile farklı aralıklar için standart değer arasındaki fark olarak ifade edilmektedir.

2.4.1. Doğrusal Olmayan Çok Değişkenli Analiz Teknikleri

Doğrusal olmayan çok değişkenli analiz tekniklerinin başında, doğrusal olmayan kanonik korelasyon analizi (DOKKA) ve doğrusal olmayan regresyon analizi gelmektedir.

- **DOKKA:** DOKKA 1980'li yıllarda Gifi ve de Leeuw tarafından ortaya atılan bir tekniktir (de Leeuw, 1988:4). Araştırma verilerinin, kategorik olduğu durumlarda sıklıkla kullanılmaktadır. Diğer doğrusal olmayan çok değişkenli

istatistiksel yöntemleri gibi varsayımları olmaması ve istatistik paket programlarında yer alması nedeniyle uygulamaya kolaylığı sağlayarak günümüzde oldukça tercih edilmektedir.

- **Doğrusal Olmayan Regresyon Analizi:** Sürekli değişkenlerinin dağılımlarının normal olmadığı, çarpıklık gösterdiği durumlarda, verilerin kesikli ya da kategorik olduğu durumlarda doğrusal istatistiki yöntemlerin kullanılması önemli iki sorunlara neden olmaktadır. Bu sorunlar çoğu değişkenlerin sıfır olmasından dolayı dağılımın sağa çarpık olması ve kuramsal olarak mümkün olmayan negatif parametre temin edilmesidir (Frome ve ark., 1973; Cox, 1983; SAS, 2005). Saymaya dayalı elde edilen veri analizlerde normal dağılımı sağlamak güç olmaktadır. Bu gibi durumlarda tercih edilen bir diğer doğrusal olmayan çok değişkenli istatistiksel model doğrusal olmayan regresyon analizidir.

2.5. Çok Değişkenli İstatistiksel Teknikler İçin Bazı Tanımlar

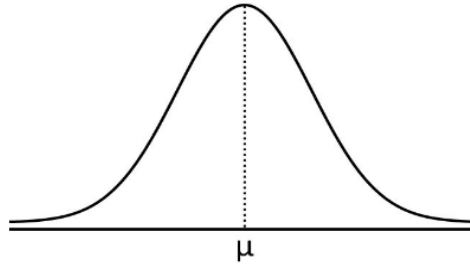
2.5.1. Standartlaştırma: Çok değişkenli analizde çoğunlukla birimleri birinden farklı olan veriler yer alır. Bazı durumlarda verilerin birbiri ile aynı olması analiz için daha uygun olmaktadır. Bu standartlaştırmayı yapmak için birçok standartlaştırma yönteminden yardım alınabilir. Z standartlaştırması en yaygın kullanılanıdır.

Z standartlaştırması dışında kullanılan standartlaştırmaları şöyle sayabiliriz (Alpar, 2013;97:98):

- Dağılım aralığı 1 olacak şekilde standartlaştırma,
- Dağılım aralığı (-1,+1) olacak şekilde standartlaştırma,
- Dağılım aralığı (0,+1) olacak şekilde standartlaştırma,
- En büyük değer 1 olacak şekilde standartlaştırma,
- Aritmetik ortalama 1 olacak şekilde standartlaştırma,
- Standart sapma 1 olacak şekilde standartlaştırma,
- Ortalama 50 standart sapma 10 olacak şekilde standartlaştırma.

2.5.2. Normallik: Çoğu çok değişkenli istatistiksel yöntemler veri matrislerinin normal dağıldığı varsayımına dayanmaktadır. Normal dağılım, Gauss dağılımı olarak da bilinmektedir. Standart normal dağılım bir veri setinde ortalamanın 0, varyansın ise 1 olduğu durumda sağlanmaktadır.

Şekil 2.3: Normal Dağılım Grafiği



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{1}{2}\right)\left[\frac{x-\mu}{\sigma}\right]^2} \quad 2.1$$

Normal Dağılım, 2.1' de verilen olasılık yoğunluk fonksiyonuna sahiptir. Bu fonksiyonun ortalaması μ ile ± 1 standart sapma ($\pm 1\sigma$) aralığında gözlemlerin % 68,26'sı, ± 2 standart sapma ($\pm 2\sigma$) aralığında gözlemlerin % 95,44'ü, ± 3 standart sapma ($\pm 3\sigma$) aralığında gözlemlerin yaklaşık % 100'ü bulunur (Alpar, 2013;97:98).

Verilerin normal dağılıma uygun olup olmadığını ortaya koymak amacıyla çeşitli normallik testlerinden yararlanmak mümkündür. Bu testler arasında en bilinenleri Ki-Kare, Kolmogorov - Smirnov, ve Shapiro – Wilk normallik testleridir (<http://mustafaotrar.net/> Erişim tarihi: 21.05.2019).

2.5.3. Uç Değerler: Elde ettiğimiz araştırma veri setlerinde yer alan bazı gözlemler kimi zaman genel eğilimin çok dışında yer alabilir. Diğer gözlemlerden aşırı farklı olan bu gözlemlere uç değerler (aşırı gözlemler) denir. Bu değerlerin kaynağı veri giriş hatası, veri kaydetme hatası, hatalı kodlama gibi insan kaynaklı olabildiği gibi analiz sonucu elde edilmiş doğru veriler de olabilir. Anormal bulgulara sebep olan bu verilerin analiz dışında bırakılması daha uygun görülmektedir. Eğer uç değerlerin örnekleme temsil ettiği düşünülüyor ve uç değerler analiz dışı bırakılmak istenmiyorsa kullanılacak doğru

istatistiki tekniklerle farklılaşmaya gidilebilir ([https:// olcemediginseyiyonetemezsin.wordpress.com/](https://olcemediginseyiyonetemezsin.wordpress.com/) Erişim tarihi:21.05.2019).

2.5.4. Eşvaryanslılık: Varyansın regresyon doğrusu boyunca açıklayıcı (tahminleyen) değişkenin (X) tüm değerleri için aynı olması varsayımı olarak açıklanmaktadır. (<http://istatistiknotlarim.blogspot.com/> Erişim tarihi: 21.05.2019)

2.5.5. Evren ve Örneklem: Örneklem evreni açıklayabilmek için, evrenin özelliklerini taşıyan küçük parçalardır. Evren hakkında genelleme yapabilmek için aldığımız örneklem yansıma süreci sonunda seçilirler. Bunun nedeni de ait olduğu evreni doğru yansıması isteğidir. Fakat gerçek hayatta bunun tersi olur ve evrenler sıklıkla örneklem aracılığı ile tanımlanırlar. (Tabachnick, Fidell, 2015: 7).

2.5.6. İstatistiksel Güç: Çok değişkenli istatistiksel tekniklerin varsayımlarından bahsederken, bu varsayımlardan sapmanın istatistik gücünü olumsuz etkileyeceğinden bahsetmiştik. Bu başlık altında istatistiksel güç tanımını detaylandıralım.

İstatistiksel güç gerçekte var olan etkilerin veri analizi sonucunda istatistiksel manidarlık ortaya koyma olasılığına işaret etmektedir. Örneğin YDS ile ortaöğretim İngilizce ders not ortalaması arasındaki anlamlı ilişki araştırılıyor olsun. Gerçekte de aralarında anlamlı bir ilişki var olduğunu varsayalım. Yeterli örneklem büyüklüğüne ulaşılmadan yapılan bir analizde bağımlı değişken üzerinde var olan etki analizde gerçekte olduğu kadar güçlü görünmeyebilir. Güç ile ilgili en kritik nokta araştırmacının örneklem büyüklüğüne karar verdiği aşamadır. Örneklem büyüklüğüne karar verilirken beklenen etki büyüklüğü, etkinin ölçülmesinde beklenen değişkenliği, istenen alfa değeri ve istenen güç önemli 4 parametredir (Tabachnick, Fidell, 2015: 11).

2.5.7. Betimsel ve Yordamsal İstatistikler: Betimsel istatistik gözlemleri ve verileri tanımlayıcı indekslere dönüştüren, çok sayıdaki gözlemi özetlemek, organize etmek ya da azaltmak için kullanılan istatistiklerdir.

Yordamsal istatistik ise evren ve örneklem arasındaki benzerliklerin kestiriminde sonuç çıkarmak ve tahmin yapmak için kullanılan bir istatistiktir. Birçok araştırma

örneklemden yola çıkarak evrenle ilgili sonuca varmak amacındadır. Yordamsal istatistik evrenle ilgili sonuçların ifade edilmesinde kullanılır.

Günümüzdeki araştırmalarda ise genellikle veri setimizi hem betimlemek hem yordamlarda bulunmak isteriz. Ancak yordama işleminde betimlemeden daha fazla varsayım mevcuttur. Hatta birçok çok değişkenli istatistiksel analiz varsayımı yordamsal istatistikler içindir (Tabachnick, Fidell, 2015:8).

2.5.8. Varyans – Kovaryans Matrisi: Varyans-Kovaryans Matrisi kare ve simetriktir. Bu matriste köşegen üzerinde her bir değişkenin varyans değeri bulunurken, köşegen dışındaki değerler değişken çiftleri arasındaki kovaryansı belirtir (Tabachnick, Fidell, 2015:14).

2.5.9. Anlamlılık Testleri: Araştırma sonucu elde edilen sonuçların araştırmacı açısından kayda değer olup olmadığının test edilmesidir. Bir başka deyişle bulunan sonucun tesadüfî olup olmadığının sorgulanmasıdır. İngilizce olasılık anlamına gelen “probability” sözcüğünden gelen “p” gösterimi ile simgelenmektedir. P değeri için 0,05 değeri genel kabul görmüş bir değer sayılmaktadır. Ancak yapılan çalışmanın hassaslığına göre p değeri değişebilmektedir. Yapılan hatanın ciddi sonuçları olabileceği durumlarda anlamlılık düzeyi %1 ya da %0,01 alınabilir. P değeri $p < 0,05$ çıkması yokluk hipotezinin reddedilmesi için güçlü kanıtların olduğunu göstermektedir (<http://biistatistik.com>/Erişim tarihi: 29.05.2019).

2.5.10. Güven Aralığı: Güven aralığı ortalamaların dağılımı ile ilgili bilgiler sunarak p değeri hakkında yorum yapmamıza olanak sağlar. Bir anakütle parametre değerinin tek bir sayı ile tahmininin yapılacağına, bu parametre değerini kapsayabilecek iki (alt ve üst sınır) sayıdan oluşan bir aralık bulunur. Böylece güven aralıkları bir kestirimin ne kadar güvenilir olduğunu gösterir.

Bulunan aralığın ana kütle parametresini ne kadar temsil ettiği güven aralığı ile belirlenir. Bir güven aralığı genel olarak %99 güven aralığı veya % 95 güven aralığı veya %90 güven aralığı şeklinde ifade edilir. Bir güven aralığının iki uç (alt ve üst) noktası güven sınırları olarak anılır. İstenilen güven düzeyini yükseltmek, güven aralığının genişlemesine yol açar. Bir araştırma için, güven düzeyi gittikçe büyütülürse, çıkartılan güven aralığı da gittikçe genişleyecektir

3. BÖLÜM

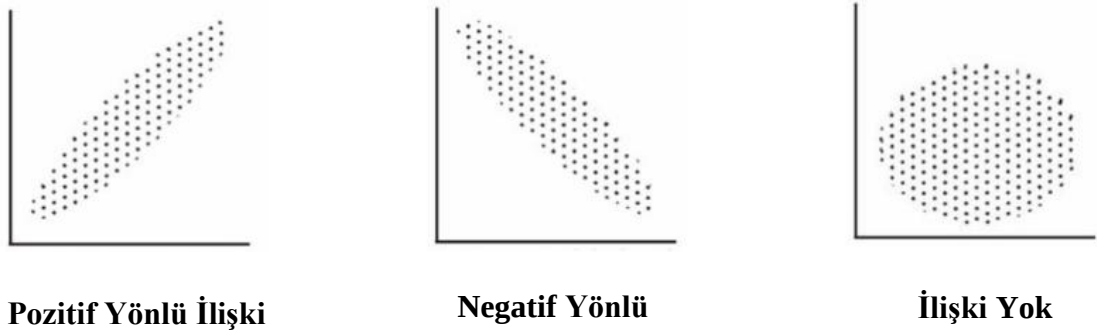
KANONİK KORELASYON ANALİZİ YÖNTEMİ

İkinci bölümde çok değişkenli istatistiksel yöntemlerin içinde tanımlanan KKA, bu bölümde detaylıca anlatılmıştır. KKA' nden bahsetmeden önce, basit korelasyon ve çoklu korelasyon kavramları tanımlanmıştır. Daha sonra ise, KKA' nin kavramsal çerçevesi, özellikleri, varsayımları, amaçları ve kullanım alanları ile DOKKA incelemiştir.

3.1. Basit Korelasyon Analizi

İstatistikte bilinen en basit ilişki korelasyondur. X ve Y ile tanımlanan iki tesadüfi değişken arasındaki ilişkidir. İki rastgele değişken arasındaki bu ilişkilerin derecesinin ölçülmesine **basit korelasyon** denir. Bu analiz temel olarak nicel değişkenler arasındaki düz çizgi [doğrusal] bir ilişki olduğu varsayımına dayanmaktadır. İkili değişkenler için birleşme ölçütlerine benzer şekilde değişkenler ve onun yönü arasındaki ilişkinin “gücünü” veya “derecesini” ölçmektedir. Bir korelasyon analizinin sonucu, -1 ile +1 değerleri arasında değişen bir korelasyon katsayısıdır. +1 korelasyon katsayısı, iki değişkenin, pozitif [doğrusal] yönde kusursuz ilişkisini; -1 korelasyon katsayısı ise iki değişkenin zıt [doğrusal] yönde kusursuz ilişkisini gösterir. Sıfır korelasyon katsayısı çalışılan iki değişken arasında doğrusal bir ilişki olmadığını göstermektedir (Gogtay ve Thatte, 2017: 78).

Şekil 3.1: Pozitif-Negatif-Nötr İlişki Grafikleri



Söz konusu ilişkinin normal dağılım göstermesi durumunda değişkenler arası ilişki Pearson tarafından önerilen korelasyon katsayısı formülü ile hesaplanır. Eşitlik 3.1’de Pearson korelasyon katsayısı denklemi yer almaktadır.

$$r_{xy} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad 3.1$$

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Kısmi korelasyon, bağımsız değişkenlerden birinin diğer değişkenlerin sabit kalması durumunda bağımlı değişkenle ilişkisi anlamına gelmektedir. Kısmi korelasyon formülü, eşitlik 3.2 de verilmiştir. y değişkeni bağımlı değişken, x ile z değişkenleri ise bağımsız değişkenlerdir. $r_{xy,z}$, z bağımsız değişkeni sabit tutulurken, x ile y değişkenleri arasındaki kısmi korelasyon katsayısını ifade etmektedir.

$$r_{xy,z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad 3.2$$

Değişkenler metrik ve normal dağılımlı değil ise Spearman ve Kendall türü parametrik olmayan ilişki ölçün formüller kullanılır (Albayrak, 2006:467).

3.2. Çoklu Korelasyon Analizi

Değişken sayısının ikiden fazla ise ve değişkenlerden biri ile diğer kalan değişkenler arasındaki korelasyon araştırılıyor ise hesaplanan korelasyon katsayısına **çoklu korelasyon katsayısı** denilmektedir. Çoklu korelasyon, basit doğrusal korelasyonun bir uzantısıdır. İki veya daha fazla bağımsız değişkenin değerini temel alarak bağımlı bir değişkenin değerini tahmin edilmek istendiğinde çoklu korelasyon kullanılmaktadır. Çoklu korelasyon, modelin genel uyumunu (açıklanan varyansı) ve her

bir tahmincinin açıklanan toplam varyansa nispi katkısının belirlenmesini sağlamaktadır (Zaid, 2015: 17).

Değişken sayısının ikiden fazla olduğu, bu değişkenlerden sadece biri bağımlı diğerlerinin bağımsız olduğu durumlarda (Y_1, X_1, X_2 gibi) çoklu korelasyon hesaplanmaktadır. Çoklu korelasyona ait denklem eşitlik 3.3' te verilmiştir.

$$R_{y,x} = \sqrt{\frac{\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}{\sigma_{11}}}$$

3.3

$$R_{y,x} = \sqrt{\frac{b' X' y - \frac{1}{n} (\sum y_i)^2}{\sum y_i^2 - \frac{1}{n} (\sum y_i)^2}}$$

3.3. Kanonik Korelasyon Analizi

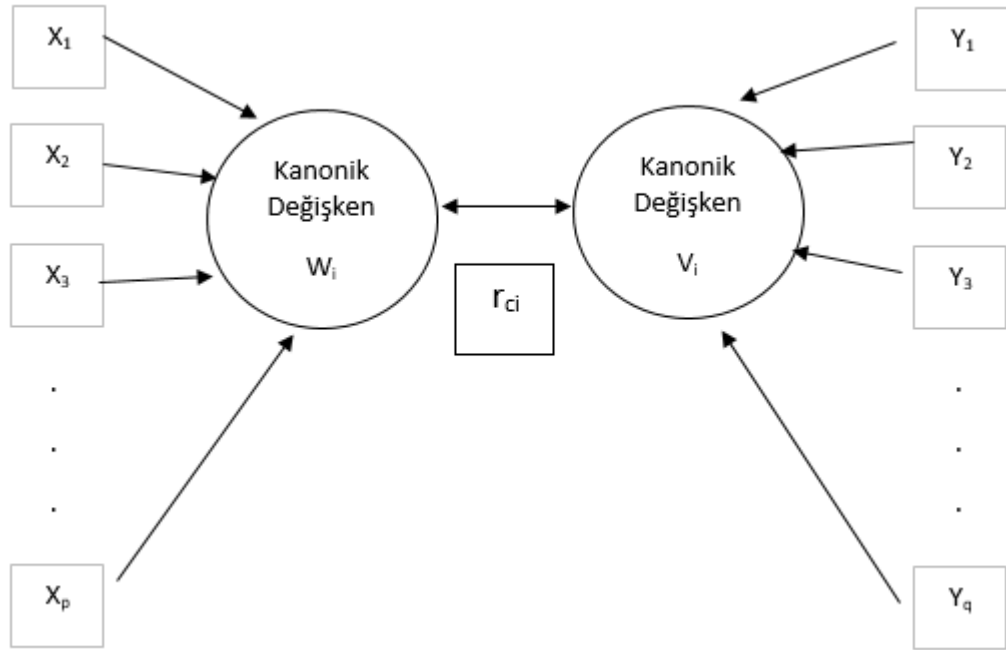
Her veri seti en az iki değişkenden oluştuğunda, iki değişken kümesi arasındaki ilişkilerin incelenmesine KKA denilmektedir. Başka bir tanıma göre, KKA bir çift değişkenli veriden ortak özellikleri çıkarmak için kullanılan bir tekniktir. KKA, korelasyon katsayısının maksimuma çıkarıldığı bir çift çok değişkenli doğrusal bir dönüşümdür. KKA iki değişken grubu arasında bağımlı ve bağımsız bir ilişki zorunluluğu barındırmaz. Ancak bir değişken grubunu bağımlı, diğer değişken grubunu bağımsız olarak düşünmek faydalı olabilir. KKA' de veri setlerindeki değişken sayısı eşit olmak zorunda değildir. Analizde iki değişken seti arasında yer alan değişkenlerin doğrusal kompozitleri ile yeni değişkenler üretilir. Bu yeni değişkenler arasındaki korelasyonun maksimum ve birim varyanslı olması amaçlanır.

Basit korelasyon (-1,+1) arası değer alırken, kanonik korelasyon (0,+1) arası değer almaktadır. Kanonik korelasyon değerinin 1 olması iki set arası mükemmel doğrusal ilişkiyi göstermektedir. Kanonik korelasyon değerinin 0,9 - 1 arasında olması setler arası çok yüksek bir ilişkiyi gösterirken, değerin 0 - 0,25 arasında olması zayıf bir ilişkiyi ifade etmektedir.

Kanonik korelasyonu anlamamanın en iyi yolu çoklu regresyonu düşünmektir. Çoklu regresyondan farklı olarak bağımlı değişken sayısı birden fazladır. İki veri seti arasındaki ilişkiyi en yüksek yapacak doğrusal bileşenler oluşturulur. Ancak çoklu korelasyonda bağımlı tek değişken bulunduğundan, bağımsız değişkenlerin sadece bir kombinasyonunun oluşturulması yeterlidir. KKA 'nde ise birden fazla bağımlı değişken bulunduğu için değişkenlerin birden fazla farklı kombinasyonu oluşturulmaktadır. Genelde kombinasyonların ilk ikisi ya da üçü yorumlamaya değerdir (Tabachnick, Fidell, 2015: 571).

KKA genel olarak aşağıdaki gibi ifade edilmektedir:

Şekil 3.2: KKA' nin Genel Yapısı



Bağımlı yöntemlerin çoğu, genel bir yöntem olarak KKA' nin özel bir şekli olarak yorumlanabilir:

- KKA' nde tek bir bağımlı değişken söz konusu ise KKA çoklu regresyon analizine dönüşmektedir.
- ANOVA ve diskriminant analizi çoklu regresyon analizinin özel formları olarak sayılabilmektedir.

- MANOVA ve çoklu diskriminant analizi de kanonik korelasyonun özel şeklidir (Albayrak, 2006:467).
- Temel bileşenler analizi ile oldukça benzeyen KKA' nin temel farkı, temel bileşenler analizinde incelenen tüm değişkenlerin tek bir küme içinden geldiğinin düşünülerek hesaplamının yapılmasıdır.
- KKA setleri tek değişken olarak belirlemesi açısından faktör analizine de benzemektedir (Karagöz, 2016: 920).

3.3.1. Kanonik Korelasyon Analizinin Avantajları ve Dezavantajları

KKA, teori denemeleri, keşif çalışmaları ve açıklayıcı/tahmin edici araştırmalarda kullanılan, son derece esnek bir analiz yöntemidir. Kanonik çözümlerin yorumlanma kolaylığı temel formülasyondaki dönüşüm ve diğer gelişmelerdeki son yenilikler sayesinde artmıştır.

Cooley ve Lohnes tarafından bu konu ile ilgili ortaya atılan bir görüş analiz ile ilgili önemli bir özet niteliği taşımaktadır. Bu görüşe göre;

“Kanonik korelasyon modeli ilk bakışta, iki ölçüm değerleri arasındaki ilişkinin açıklanmasında kullanılan son derece karmaşık bir model gibi görünmekte, analitik gerçekte ise bilimsel genellemenin zor problemlerinde en basit modeldir”.

Özel bir doğrusal ilişki biçimi olan KKA kullanımının önemli avantajlarından bir diğeri de kanonik ağırlıklar ve kanonik yüklerin test istatistiklerinin bulunmamasından kaynaklanan sorunların çözülmüş olmasıdır. Bireysel ölçüt ve tahmin ağırlıkları ile çapraz yükler için standart hataları, bilgi matrisinin tersi alınmak suretiyle tahmin edilebilir. Bu ağırlık ve yüklerin istatistiksel önemi, kritik oranın hesaplanması parametre tahminlerinin, tahmin edilmiş standart hatasına bölünmesi suretiyle ortaya konabilir (Özçomak ve Demirci, 2010:261-274).

KKA iki değişken grubu arasında bağımlı ve bağımsız bir ilişki zorunluluğu barındırmadığından kümeler arasındaki ilişkiyi iki yönlü analiz edebilmektedir. Eğer X ve Y değişkenleri ters çevrilmiş olsaydı sonuçlar aynı olurdu. Teknik olarak, KKA 'ın özü, her bir çift arasındaki korelasyonu en üst düzeye çıkarmak için öngörücü ve kriter değişkenlerinin doğrusal kombinasyonlarının çiftlerini oluşturmaktır.

Değişkenler arasında doğrusal olmayan bir ilişki varsa, KKA her zaman yararlı özellikleri ortaya koymaz (Akaho, 2006: 2). KKA bazı dezavantajları da barındırmaktadır. Örneğin KKA, altta kalan belirgin olmayan yapıların ortaya çıkarılması konusundaki yetersiz kalabilmektedir. Fornell (1979) tarafından ortaya atılan endişe ise, KKA' nde verilerin oldukça küçük varyanslardan etkilenebileceğidir (Özçomak S. Vd ,2010:264).

KKA' nde bağımlı bağımsız değişken kavramının olmaması nedeniyle analizde tam bir regresyon denklemi oluşturulmamaktadır. KKA bir boyut indirme yöntemi olarak da ele alınabilir. Ancak bu durumda da kavramsal anlamlılık aşamasında sorunlar ortaya çıkacaktır (Khattree ve Naik, 2010:77).

3.3.2. Kanonik Korelasyon Analizinin Tarihçesi ve Kullanım Alanları

KKA ilk defa Amerikalı matematiksel istatistikçi Hotelling tarafından 1936 yılında bir grup içerisinde alınan iki değişken seti arasındaki ilişkileri inceleyen bir model olarak ortaya atılmıştır. Kanonik korelasyonun ortaya çıkmasının ardından Bartlett (1941), kanonik korelasyonun istatistiksel önemini ortaya koymak için Hotelling'in geliştirdiği önem testi ile Fisher ve Hsu' nun 1939 yılında geliştirdikleri önem testleri arasındaki mantıksal ilişkiyi açıklamaya çalışmıştır (Çankaya, Çukurova Üniversitesi, Türkiye, 2005:6).

Gelişmiş bir ilişki analizi olan KKA' nin, ekonomi, tıp, meteoroloji, kemometri, biyoloji ve nöroloji, doğal dil işleme, konuşma işleme, bilgisayarlı görme ve multimedya sinyal işleme dahil olmak üzere birçok alanda kullanım yelpazesi bulunmaktadır (Andrew,vd., 2013: 1247).

KKA üzerine Türkiye'de yayımlanan birçok makale bulunmaktadır. Yayımlanmış makalelerden bazılarına aşağıda yer verilmiştir:

Sıddık Keskin ve Abdullah Nuri Özsoy 2004 yılında KKA üzerine örnek bir uygulama yayımlamıştır.

Girginer vd., (2007) çalışmalarında, öğrencilerin istatistik dersine yönelik tutumları ile istatistik dersindeki başarı arasındaki ilişkiyi doğrusal olmayan kanonik korelasyon analizi ile incelemiştir.

İlter Ünlükaplan (2009) Avrupa Birliği Üyesi Ülkelerde İktisadi Kalkınma, Rekabetçilik ve İnovasyon İlişkileri arasındaki yüksek dereceli ilişkiyi KKA ile belirlemiştir.

Fatma Lorcü ve Bilge Acar Bolat (2009) yaşlara göre ölüm oranları ile soysa-ekonomik göstergeler arasındaki ilişkiyi KKA ile irdelemişlerdir.

Emir ve Kılıç'ın 2011'deki çalışmalarında ise müşterilerin alınan hizmete yönelik memnuniyet düzeyleri ile bazı demografik özellikleri arasında anlamlı ilişkiler olduğunu belirtmiştir.

Tatlı vd. , (2011) yaptıkları çalışmada, memnuniyet düzeyleri ile cinsiyet arasında anlamlı ilişki aramışlardır.

Suphi Özçomak ve arkadaşları çeşitli iklim ve ürün verileri arasındaki ilişkiyi KKA ve veri zarflama analizi yöntemleri ile ortaya koymuşlardır.

Ayfer Sayın ve arkadaşları (2012) Aşamalı Dersler Arasındaki İlişkileri Kanonik Korelasyon Tekniğiyle incelemiştir.

Filiz ve Kolukısaoğlu' nun 2012'de yaptığı çalışmada ise, KKA ile müşterilerin demografik özellikleri ile hizmet işletmesinden (otel) aldıkları hizmetten duydukları memnuniyet düzeyleri arasındaki ilişki incelenmiştir.

Fatih Çekmek (2012) , Türkiye'deki İllerin Gelir ve Refah Düzeyi Değişkenleri Arasındaki İlişkiyi Kanonik Korelasyon Analizi ile irdelemiştir.

KKA kullanılarak Türkiye'de hazırlanan bazı yüksek lisans çalışmaları aşağıda yer almaktadır ([https:// tez.yok.gov.tr/ UlusalTezMerkezi/](https://tez.yok.gov.tr/UlusalTezMerkezi/) 08.05.2019):

M. Hakkı Özel (1984), hazırladığı Yüksek Lisans tezinde ekonomik kalkınma eğitim arasındaki ilişkiyi KKA ile incelemiştir.

Öznur İşçi 1995 yılında, “Kanonik Korelasyon ve Açıklanmış Varyans Oranı İstatistiklerini Optimum Yapan Faktörlerin Bornova Anadolu Lisesi Öğrencilerinin Başarı Düzeylerinin Tahminlenmesinde Kullanılması Üzerine Bir Araştırma” başlıklı tezini hazırlamıştır.

Ali Murat Tatar (1999), yüksek lisans tezinde France X Akkaraman melezi erkek kuzularında süt emme dönemindeki canlı ağırlık ve vücut ölçüleri arasındaki ilişkinin kanonik korelasyon metodu ile araştırılmasını konu almıştır.

İlkay Barıtcı (2001), Kilis keçilerinin doğumdan itibaren belli dönemlerde vücut ölçüleri arasındaki ilişkiyi KKA ile incelemiştir.

Hakan Çınar (2002), Avrupa Birliği ülkeleri ile Türkiye'nin sosyal ve ekonomik göstergeleri arasındaki ilişkilerin analizini KKA ile yapmıştır.

Soner Çankaya (2005) doktora tezinde KKA ve KKA' nin hayvancılıkta kullanımını konu almıştır.

Uzgören tarafından 2007'de yapılan çalışmada, öğrencilerin mevcut eğitimden memnuniyetleri ile bireysel özellikleri arasındaki ilişkiyi analiz etmiştir.

Tektaş vd., (2009) öğrencilerin üniversitenin sağladığı akademik hizmetlere ilişkin beklenti ve memnuniyet düzeylerini belirleme amacıyla yaptıkları çalışmada anlamlı farklılıklar ortaya çıkmıştır.

Giray (2011) çalışmasında, bireylerin mutluluk algısı ile demografik özellikleri arasında doğrusal olmayan KKA ile anlamlı ilişkiler tespit etmiştir.

Kemal Gürkan Toker (2013) KKA ile Sistem Tanıma konulu yüksek lisans tezini sunmuştur.

Kübra Soğancı (2017), yüksek lisans tezinde Orta Kızılırmak vadisinden toplanan yerel kuru fasulye (*Phaseolus vulgaris* L.) genotiplerine ait karakterler arası özellikleri KKA yoluyla ortaya koymuştur.

Uluslararası çalışmalardan da bazı örnekler vermek konunun kullanım alanının genişliğini ifade edebilmek açısından önem taşımaktadır. Aşağıda bazı uluslararası çalışma örnekleri yer almaktadır (Çankaya, Çukurova Üniversitesi, Türkiye, 2005:11:14):

Bradshaw ve ark.(2000), 64 adet kedinin bireysel ve koloni halinde iken yiyecek tercihlerindeki farklılıkları aramışlardır.

Larson ve ark. (2000), kuzey Caroline'nin Duck kıyı bölgesinden 11 yıl boyunca ölçülen dalga hareketleri ile sahil görüntüsü arasındaki ilişkiyi belirlemek amacıyla KKA'ni kullanmışlardır.

Vainionpaa ve ark. (2000), farklı patates kültürlerinin çeşitli yapı ve doyumsal özelliklerini içeren veri kümesinde, kaliteyi oluşturan faktörleri belirlemek ve kalite özellikleri ile üretim faktörleri arasındaki ilişkiyi incelemek için temel bileşenler analizi ile KKA' ni kullanmışlardır

Greenwood ve ark. (2002), Sert Peynirin Özel Duyusal Tekstür Özellikleri ve Ön Proteolitik Örnekler isimli çalışmalarında, altı ay bekletilmiş sert peynirler ile daha kısa süre bekletilmiş sert peynirlerden alınan sertlik, esneklik, nem, ufalanabilirlik gibi tekstür özellikleri arasındaki ilişkiyi KKA ile belirlemişlerdir.

Görüldüğü gibi KKA, birçok iki ve çok değişkenli parametrik istatistik yöntemi için genel bir model olarak yaygın bir şekilde göz önünde tutulmaktadır. Eş zamanlı olarak çoklu ölçüt ve çoklu tahmin değerini kontrol altında tutabilme kapasitesinden dolayı KKA, özellikle davranış bilimlerinde kullanımı açısından çok cazip hale gelmiştir (Bagozzi, Fornell, ve Larcker, 1981).

3.3.3. Kanonik Korelasyon Analizinin Amaçları

KKA' nin amaçları aşağıdaki gibi özetlenebilir. (Kandemir, 2018: 2031; Kalkan ve Özden, 2017: 13; Çankaya, Çukurova Üniversitesi, Türkiye, 2005:24:25).

- Bağımsız ve bağımlı değişken kümelerine ait değişkenler arasındaki korelasyonu maksimum yapan doğrusal kombinasyonların belirlemek,
- İki değişken grubunun birbirinden istatistiksel olarak bağımsız olup olmadığını test etmek,
- Değişken kümeler arasındaki korelasyona en çok katkıda bulunan iki kümedeki değişkenlerin belirlemek,
- Değişken kümelerinin değişkenleri arasında maksimum korelasyon yapan doğrusal kombinasyon değerini tahmin etmek,
- Kanonik fonksiyonların, değişken kümeler arasındaki ilişkileri açıklama veya öngörmedeki oransal etkilerini belirlemek,
- İki veri kümesindeki ilişkiye en çok katkıda bulunan değişkenleri, iki kümedeki katkı sırasını ve değişkenlerin miktarını belirlemek,
- Bir değişkenin değişken kümesinin açıklayıcı gücüne ne ölçüde dâhil olduğunu belirlemek,
- Bir değişkenin, dâhil edilmediği değişkenler dizisinin doğrusal bileşiminin açıklamasına veya tahminine ne ölçüde katkıda bulunduğunu belirlemek,

- Her bir değişkenin çıkarılan kanonik fonksiyonlara nasıl katkıda bulunduğunu değerlendirerek iki değişken grubu arasındaki ilişkinin niteliğini açıklamaktır.

3.3.4. Kanonik Korelasyon Analizinin Varsayımları

KKA, rastgele değişken grubunun doğrusal fonksiyonları arasındaki maksimum korelasyonu bulmaya çalışmaktadır. KKA' nde, iki veya daha fazla küme arasındaki ilişki, bağımlı ve bağımsız değişken kümeleri arasındaki lineer ilişkiler aranırken oluşan birkaç lineer bileşen (kanonik bileşenler) yardımıyla incelenmektedir. Bu kapsamda, KKA genel olarak dört farklı temel varsayımı vardır. Kanonik analiz öncesi bu şartların olması gerekmektedir. Bu varsayımlar şunlardır (Kandemir, 2018: 2030; Keskin ve Özsoy, 2004: 58; Çankaya, Çukurova Üniversitesi, Türkiye, 2005:25:26):

- Değişkenler kümeleri arasında ilişki doğrusal olması,
- Çok değişkenli normal verinin dağılım özellikleri göstermesi,
- Özellikleri arasında ölçüm hatası asgari düzeyde olması ve aykırı değerlere gerekli düzenlemeler analiz öncesinde yapılması,
- Değişkenler arasında çoklu bağlantı olmaması,
- Stevens'a göre (1986) güvenilirlik açısından, kanonik korelasyon katsayısının 0,70-1 arasında tahmin edilmesi halinde örnek büyüklüğünün 50, katsayının 0,50-0,70 olması durumunda 100, daha az olması halinde ise örnek büyüklüğünün 200 olması uygun olacaktır.

KKA' nin varsayımları istatistiksel olarak tanımlanan ve aşağıda yer alan bazı kuram/teorilere dayanmaktadır. Bunlar (Türkegün, 2015: 9-10):

- **Doğrusallık:** Değişken çiftler ve değişken kümeleri arasındaki ilişki doğrusal olmalıdır. KKA, değişken kümelerdeki değişkenler arasındaki doğrusal ilişkiyi maksimuma çıkarmaktadır. Değişkenler arasında doğrusal olmayan bir ilişki varsa KKA hesaplanamamaktadır.

- **Çok Değişkenli Normal Dağılım:** Korelasyon katsayılarının hipotez testi için, numunenin çizildiği evrenin çok değişkenli normal dağılım varsayımı olmasını gerektirmektedir.
- **Sabit Varyans:** Sabit varyans durumunun olması halinde KKA' ndeki değişkenlerin iyi şekilde tanımlandığı varsayılmaktadır.
- **Çoklu Bağlantı Sorunu:** Değişkenler arasında çoklu bağlantı olmaması gerekmektedir.
- **Gözlem Sayısı:** Gözlem sayısının KKA' nde fazla olması gerekmektedir ve önerilen gözlem sayısı toplam değişkenin 20 katı kadar olmalıdır.
- **Aykırı Değerler:** Yetersiz veri olması durumunda, veri setinde bir aykırı değerlerin olması gerçeği, korelasyonun yanlış hesaplanmasına neden olmaktadır. Bu nedenle, veri setinde aykırı değerler olmamalıdır. Çünkü aykırı değerler kanonik analizi önemli ölçüde etkilemektedir.

3.3.5. Kanonik Korelasyon Analizi Temel Kavramları

Özçomak, vd. (2012: 114), KKA' nin formüsel gösterimini açık bir şekilde ifade etmektedir ve bu gösterim aşağıda yer almaktadır.

$$\begin{array}{ll}
 V_1 = b_{11}y_{11} + \dots + b_{1q}y_{1q} & W_1 = a_{11}x_{11} + \dots + a_{1p}x_{1p} \\
 \cdot & \cdot \\
 \cdot & \cdot \\
 \cdot & \cdot \\
 V_i = b_{i1}y_{i1} + \dots + b_{iq}y_{iq} & W_i = a_{i1}x_{i1} + \dots + a_{ip}x_{ip}
 \end{array} \quad 3.1$$

Formüldeki değişkenlerin açık olarak tanımları ise aşağıdaki şekildedir:

y_{ij} = ölçüt değişkenler ($1 \leq j \leq q$)

x_{ik} = tahmin değişkenler ($1 \leq k \leq p$)

i = değişken çifti sayısı (p ve q'nun en küçükleri)

q = ölçüt değişken sayısı,

p = tahmin değişken sayısı,

V_i = ölçüt değişkenin i.'inci çifti için kanonik rastlantı değişkeni,

W_i = tahmin değişkenin i.'inci çifti için kanonik rastlantı değişkeni,

b_{ij} = ölçüt değişkenin i.'inci çiftindeki j'inci değişkenin kanonik ağırlığı ($1 \leq j \leq q$)

a_{ik} = tahmin değişkenin i.'inci çiftindeki k'inci değişkenin kanonik ağırlığı ($1 \leq k \leq p$) olarak ifade edilmektedir.

Kanonik korelasyonun kavramları aşağıdaki gibi açıklanabilir. (Türkegün, 2015: 11-12).

- **Kanonik Değişken:** Lineer bileşenlerde iki değişken kümesi ifade edilerek oluşturulan özet değişkenlerdir. Eşitlik 3.1' de yer alan V_i ve W_i değişkenlerini ifade etmektedir.
- **Kanonik Korelasyon Katsayısı:** İki değişken grubu arasındaki değişkenlerin doğrusal kombinasyonları, bu kombinasyonlar arasındaki ilişkinin katsayılarıdır ve $r_{v,w}$ ile gösterilir.
- **Kanonik Ağırlıklar:** Orijinal değişkenlerden her birinin kanonik değişkene yaptığı katkının derecesine kanonik ağırlık denilmektedir. Eşitlik 3.1'de yer alan a_{ip} ve b_{iq} katsayıları kanonik ağırlıkları ifade etmektedir.
- **Kanonik Yükler:** Orijinal değişkenler ile kanonik değişkenler arasındaki basit korelasyon katsayılarına kanonik yük denilmektedir. Orijinal değişkenlerin karşı kümenin kanonik değişkenler arasındaki basit korelasyona ise kanonik çapraz yük denilmektedir.
- **Kanonik Kök:** Kanonik korelasyon katsayısının karesidir.
- **Açıklanan Varyans:** Her değişken kümesindeki kanonik yüklerin ilgilenilen orijinal değişkenlerdeki değişkenliği açıklayabildiği yüzdesini (%) belirtmektedir.
- **Gereksizlik İndeksi:** Katsayılar veri setlerinde hesaplanan varyansın oranını vermektedir. Gereksizlik belirleme katsayıları, uygunluk katsayılarının, kanonik korelasyon katsayılarının karelerinin (öz değerler) çarpılması ile elde edilir. Her bir kümedeki her bir kanonik değişken için elde edilen gereksizlik indeks katsayılarının toplamı, toplam gereksizlik indeksini verir. Toplam gereksizlik indeks katsayısı ise ilgili kümedeki değişkenlerin yüzde (%) oranının ne kadarının diğer değişkenler kümesindeki kanonik değişkenler

tarafından açıklandığını belirtir (Turan, Trakya Üniversitesi, Türkiye, 2019: 7).

3.4. Doğrusal Olmayan Kanonik Korelasyon Analizi

Doğrusal Olmayan Kanonik Korelasyon Analizi (DOKKA) ilk olarak 1981'de Gifi tarafından bulunmuştur. Analiz, iki veya daha fazla değişken kümesini incelemektedir ve kümelerin birbirine nasıl görüldüğünü araştırmaktadır. DOKKA, farklı ölçüm seviyelerine veya değişkenlerin doğrusallığına sahip olabilecek değişkenlerin dağılımı hakkında herhangi bir varsayımda bulunmamaktadır. Bu da DOKKA' nın birçok alanda kullanılabilmesine olanak tanımıştır. DOKKA, değişken kategorik değişken kümeleriyle ilgili problemler için tasarlanmıştır. Analiz edilecek veri setinde kategorik veriler bulunduğu KKA yerine (DOKKA) kullanılmaktadır (Bülbül Ş., Giray S., 2011:116). DOKKA' da çoğunlukla veriler nominal ve ordinaldir. Sayısal değişkenlere ek olarak, kategorik değişkenlerin aynı anda analize dâhil edilmesi ve analize dâhil edilen değişkenlerin grafiksel olarak iki boyutlu haritalarda gösterilmesi, analizin tercih edilme sebepleridir.

DOKKA' nin temel özelliği, kullanıcının değişkenlerin ölçek seviyelerini seçebilmesidir. Analiz kategorik verilere uygulanabildiğinden, sayısal değişken kategorize edilip analize dâhil edilebilmektedir. DOKKA, bilgisayar programlarında da kullanılabilir ve bu durum DOKKA' yı son yıllarda popüler hale getirmiştir. Bu gelişim, değişkenlerin sıralayıcı, sınıflandırıcı veya sayısal yöntemler olarak kullanımını ve değişken gruplar oluşturarak gruplar arası ilişkilerin analizinde kolaylık sağlamaktadır.

DOKKA ve diğer teknikler arasındaki ilişki ise aşağıdaki gibidir (Filiz ve Kolkısaoğlu, 2012: 60).

- Her kümede yalnızca bir değişken varsa ve tüm değişkenler sınıflandırıcı bir ölçüm seviyesine sahipse, DOKKA çoklu yanıt analizine dönüştürülmektedir.
- Her küme yalnızca bir değişken içeriyorsa ve değişkenlerin ölçüm seviyeleri karıştırılırsa, DOKKA doğrusal olmayan bir temel bileşen analizine dönüştürülmektedir.
- Her kümede yalnızca bir değişken varsa ve tüm değişkenler sayısal, DOKKA temel bileşen analizi haline gelmektedir.

- Değişkenler iki küme halinde ve tüm değişkenler sayısal ise, DOKKA doğrusal KKA' ine dönüştürülmektedir.

DOKKA ikiden fazla değişken grubunun analizine izin vermektedir. Örneğin, birden fazla bağımsız küme ve bir bağımlı kümeden meydana gelebilir. Böylece değişken kümeler arasındaki doğrusal olmayan ilişkiler analiz edilebilmektedir.

DOKKA' nde, değişken kümeleri arasındaki korelasyonları maksimize etmek yerine, kümeler, nesne puanları tarafından tanımlanan bilinmeyen bir uzlaşma setiyle karşılaştırılmaktadır. Dolayısıyla, DOKKA yöntemi, kategorik değişken kümeleri ile veri kümesinin yapısal benzerlikleri arasındaki ilişkilerin incelenmesine olanak sağlamaktadır (Yazıcı vd. 2010: 504).

4. BÖLÜM

KANONİK KORELASYON ANALİZİNİN MATEMATİKSEL TANIMI

İstatistiksel olarak ifade edilebilecek en basit ilişki, X ve Y ile gösterilen iki rastgele değişken arasındaki korelasyondur (Kandemir , 2018: 2027). Çok boyutlu bir ana kitleden çizilen iki veya daha fazla bağımlı ve bağımsız değişken olması hâlinde, ilişkiyi incelemek için KKA kullanılmaktadır.

Hotelling tarafından 1936 'da önerilen KKA, iki değişken kümesi için temel vektörleri bulma problemi olarak görülmektedir ve bu değişken setleri arasında maksimum korelasyonlu ve lineer kombinasyon çiftleri olmalıdır. KKA, n gözlemden oluşan p tane bağımsız değişken (X) ile q tane bağımlı değişkenin (Y) doğrusal bileşenlerinden türetilen kanonik değişken çiftleri arasındaki maksimum korelasyonu veren doğrusal bileşen çiftlerinin araştırılması esasına dayanan, çok değişkenli bir yöntemdir (Kalaycı,2009; Özdamar, 1999).

İki değişken kümesinden, değişken sayısı az olan kümedeki değişken sayısı kadar kanonik korelasyon değeri üretilir. $m = \min (p,q)$ olarak ifade edildiğinde, m sayıdaki her kanonik değişken çifti arasındaki korelasyon katsayısı maksimum olacak şekilde hesaplanır. Kanonik değişken çiftleri arasındaki kanonik korelasyonlar birbirinden bağımsız hesaplanmaktadır.

X ve Y değişkenleri için oluşturulan W_i ve V_i kanonik değişkenlerinin denklemleri eşitlik 4.1' de gösterilmektedir (Hotelling, 1992).

$$\begin{array}{ll} V_1 = b_{11}y_{i1} + \dots + b_{1q}y_{iq} & W_1 = a_{11}x_{i1} + \dots + a_{1p}x_{ip} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ V_i = b_{i1}y_{i1} + \dots + b_{iq}y_{iq} & W_i = a_{i1}x_{i1} + \dots + a_{ip}x_{ip} \end{array} \quad 4.1$$

“m” sayıdaki kanonik korelasyon seti;

$[(V_1, W_1), [(V_2, W_2), [(V_3, W_3), \dots [(V_m, W_m)]]$ birbirinden bağımsızdır.

$$\text{Korelasyon } (V_i, V_j) = 0 \text{ ve } i \neq j$$

$$\text{Korelasyon } (W_i, W_j) = 0 \text{ ve } i \neq j$$

$$\text{Korelasyon } (V_i, W_j) = 0 \text{ ve } i \neq j \quad 4.2$$

4.1. Kanonik Değişkenler ve Kanonik Korelasyonlar

p ve q, 1'den büyük olmak üzere, birinci değişken kümesinde p tane ve ikinci değişken setinde q tane değişken olduğu kabul edilsin. Değişkenlerin lineer kombinasyonlarından oluşan yeni değişken çiftlerine (W_i, V_j) **kanonik değişkenler** denilmektedir. Oluşturulan kanonik değişken çiftleri arasındaki korelasyona ise **kanonik korelasyon** denilmekte ve r ile gösterilmektedir. Kanonik korelasyon katsayısı 0 ile 1 arasında değer almaktadır.

m sayıda kanonik korelasyon çifti hesaplanır ve en büyük değere sahip kanonik korelasyon r_1 olarak tabir edilir. W_1 ve V_1 değişkenleri arasındaki korelasyon r_1 olarak tabir edildiğinde KKA, $a_{11}, a_{12}, \dots, a_{1p}$ ve $b_{11}, b_{12}, \dots, b_{1q}$ katsayılarını r_1 değerini maksimum kılacak şekilde tahmin eder. Ardından r_1 hesaplama dışı bırakılarak, ikinci en büyük kanonik korelasyona sahip değer r_2 olarak tabir edilir. W_2 ve V_2 değişkenleri arasındaki korelasyon r_2 olarak tabir edildiğinde KKA, $a_{21}, a_{22}, \dots, a_{2p}$ ve $b_{21}, b_{22}, \dots, b_{2q}$ katsayılarını r_2 değerini maksimum kılacak şekilde tahmin eder. Kanonik korelasyonlar büyükten küçüğe sıralanır ve bu işlem m. kanonik korelasyon değerine ait kanonik değişkenler hesaplanana kadar devam eder.

4.2. Kanonik Değişkenler ve Kanonik Korelasyonların Hesaplanması

X kümesine ait veri seti p ve Y kümesine ait veri seti de q değişken barındırmaktadır. Her bir değişkenin içinde n tane veri yer almaktadır. Bu matrisler eşitlik 4.3' deki gibi gösterilmektedir.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_q \end{bmatrix} \quad 4.3$$

X matrisi $\bar{X}$ ortalama vektörüne sahiptir. Y matrisinin $\bar{Y}$ ortalama vektörüne sahiptir. C_{11} , C_{22} , C_{12} kovaryans matrislerini ifade etmektedir.

$$C_{11} = \begin{bmatrix} \sigma_{x_1x_1} & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_p} \\ \sigma_{x_2x_1} & \sigma_{x_2x_2} & \cdots & \sigma_{x_2x_p} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{x_px_1} & \sigma_{x_px_2} & \cdots & \sigma_{x_px_p} \end{bmatrix}$$

$$C_{22} = \begin{bmatrix} \sigma_{y_1y_1} & \sigma_{y_1y_2} & \cdots & \sigma_{y_1y_q} \\ \sigma_{y_2y_1} & \sigma_{y_2y_2} & \cdots & \sigma_{y_2y_q} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{y_qy_1} & \sigma_{y_qy_2} & \cdots & \sigma_{y_qy_q} \end{bmatrix}$$

$$C_{12} = \begin{bmatrix} \sigma_{x_1y_1} & \sigma_{x_1y_2} & \cdots & \sigma_{x_1y_q} \\ \sigma_{x_2y_1} & \sigma_{x_2y_2} & \cdots & \sigma_{x_2y_q} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{x_py_1} & \sigma_{x_py_2} & \cdots & \sigma_{x_py_q} \end{bmatrix}$$

4.4

Değişkenlerin doğrusal bileşiminden oluşan kanonik değişkenler, W ve V ile; kanonik değişkenler arasındaki kanonik korelasyon ise r ile gösterilmektedir.

$$W_i = a_{11}x_{i1} + \cdots \cdots + a_{ip}x_{ip} \quad V_i = b_{11}y_{i1} + \cdots \cdots + b_{1q}y_{iq}$$

$$r_i = Kor (W_i, V_i) \quad i=1,2,\dots,m \quad 4.5$$

İki veri setini de aşağıdaki matris gösterimi ile ifade etmek de mümkündür.

$$W = a'_i X \quad V_i = b'_i Y \quad 4.6$$

KKA veri setlerinin doğrusal bileşenleri olan $W_i = a'_i X$ ile $V_i = b'_i Y$ arasındaki korelasyonu maksimum yapacak a ve b matrislerini bulma problemi olduğu belirtilmiştir. Kanonik korelasyon hesaplama formülü, eşitlik 4.7' de verilmiştir.

$$\begin{aligned} Var(W) &= a' C_{11} a \\ Var(V) &= b' C_{22} b \\ Kov(W, V) &= a' C_{12} b \\ r_{W,V} &= \frac{Kov(W,V)}{\sqrt{Var(W) Var(V)}} = \frac{a' C_{12} b}{\sqrt{a' C_{11} a \quad b' C_{22} b}} \end{aligned} \quad 4.7$$

W ve V kanonik değişkenlerini maksimum yapan a ve b katsayılarını bulmak için W ve V kanonik değişkenleri arasında yer alan birim varyansa sahip kanonik değişken çifti bulunmalıdır.

$$\begin{aligned} Var(W) &= a' C_{11} a = 1 \\ Var(V) &= b' C_{22} b = 1 \end{aligned} \quad 4.8$$

Kanonik korelasyon denklemi olan eşitlik 4.7' de W ve V 'nin varyans değerleri yerine 1 yazıldığında kanonik korelasyon denklemi eşitlik 4.9' daki formuna kavuşacaktır. En büyük korelasyonun hesaplandığı kanonik değişken çifti 1. kanonik değişken çifti olarak adlandırılır ve bu ilk korelasyon katsayısı eşitlik 4.9' da görüldüğü gibi W ile V arasındaki kovaryans değerine eşit olacaktır.

$$r_{V,W} = Kov(W, V) = a' C_{12} b \quad 4.9$$

Bu şekilde Kanonik değişkenlerin maksimum yapılması kısıtlı bir maksimizasyon problemi haline gelmektedir. Bulunmak istenen **Max $a' C_{12} b$** maksimizasyon problemi **$a' C_{11} a = b' C_{22} b = 1$** kısıtı altında çözümlenmelidir. Bu kısıtlar altında katsayıların eniyileme problemi olarak dikkate alınıp çözülmesi için **γ_1 ve γ_2** langrange çarpanları olmak üzere bir langrange fonksiyonu oluşturulur.

$$L = a' C_{12} b - 0,5 \gamma_1 (a' C_{11} a - 1) - 0,5 \gamma_2 (b' C_{22} b - 1) \quad 4.10$$

Lagrange fonksiyonunun a, b, γ_1, γ_2 değişkenlerine göre kısmi türev alınarak sıfıra eşitlenmesi ile aşağıdaki denklemler elde edilmektedir.

$$\begin{aligned}\frac{\partial L}{\partial a} &= C_{12}b - \lambda_1 C_{11}a = 0 \\ \frac{\partial L}{\partial b} &= C_{21}a - \lambda_2 C_{22}b = 0 \\ \frac{\partial L}{\partial \gamma_1} &= a' C_{11} a - 1 = 0 \\ \frac{\partial L}{\partial \gamma_2} &= b' C_{22} b - 1 = 0\end{aligned}$$

4.11

Eşitlik 4.11’de yer alan ilk denklem a' ile ikinci denklem ise b' ile çarpıldığında aşağıda yer alan denklemler elde edilir.

$$\begin{aligned}a' C_{12}b - \gamma_1 (a' C_{11}a) &= 0 \\ b' C_{21}a - \gamma_2 (b' C_{22}b) &= 0\end{aligned}$$

4.12

Eşitlik 4.11’ in son iki denkleminde görüldüğü üzere $a' C_{11} a$ ve $b' C_{22} b$ değerleri 1’dir. Bu değerler denklem 4.12’ de yerine konulduğunda eşitlik 4.13’ deki formüller elde edilmektedir.

$$\begin{aligned}\gamma_1 &= a' C_{12}b \\ \gamma_2 &= b' C_{21}a\end{aligned}$$

4.13

Böylece denklem 4.13’den 4.14 eşitliği tespit edilmiş olmaktadır.

$$\gamma_1 = \gamma_2 = a' C_{12}b$$

4.14

Yukarıda verilen formüllere göre eşitlik 4.11’ deki ilk iki denklem aşağıdaki gibi yazılabilir.

$$\begin{bmatrix} -r C_{11} & C_{12} \\ C_{21} & -r C_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

4.15

Denklem 4.15' te sıfırdan farklı a ve b değerleri bulunabilmesi için söz konusu matrisin sıfır olması (tekil olması) gerekmektedir. Bu da matrisin determinantının sıfır olması durumunda mümkün olmaktadır. Bu nedenle, determinanı sıfır yapacak r değerinin bulunması gerekmektedir.

$$\begin{vmatrix} -r C_{11} & C_{12} \\ C_{21} & -r C_{22} \end{vmatrix} = 0 \quad 4.16$$

Eşitlik 4.16'daki determinantın çözümü aşamasında aşağıdaki ifadelerden biri tercih edilmelidir.

$$\begin{aligned} &|-r^2 C_{11} C_{22} + C_{12} C_{21}| = 0 \\ &|-r^2 C_{11} + C_{12} C_{22}^{-1} C_{21}| = 0 \text{ veya } |-r^2 C_{22} + C_{12} C_{11}^{-1} C_{21}| = 0 \\ &|C_{11}^{-1} C_{12} C_{22}^{-1} C_{21} - r^2 I| = 0 \text{ veya } |C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} - r^2 I| = 0 \end{aligned} \quad 4.17$$

Eşitlik 4.17 denklemlerinden elde edilen r^2 değeri eşitlik 4.18'de yer alan karakteristik denklemlerde yerine konduğunda a ve b vektörleri belirlenir.

$$\begin{aligned} (-r^2 C_{11} + C_{12} C_{22}^{-1} C_{21}) a &= (C_{11}^{-1} C_{12} C_{22}^{-1} C_{21} - r^2 I) a = 0 \\ (-r^2 C_{22} + C_{21} C_{11}^{-1} C_{12}) b &= (C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} - r^2 I) b = 0 \end{aligned} \quad 4.18$$

$$b_i = \frac{1}{r_i} C_{22}^{-1} C_{21} a_i \quad 4.19$$

Görüldüğü gibi kanonik korelasyon analizi yöntemi eşitlik 4.20' deki iki ayrı öz değer problemine indirgenmiştir.

$$\begin{aligned} C_{11}^{-1} C_{12} C_{22}^{-1} C_{21} a &= r^2 a \\ C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} b &= r^2 b \end{aligned} \quad 4.20$$

Denklem 4.20'de görüldüğü üzere problemin çözümü aynı öz değere fakat farklı öz vektörlere sahiptir. Elde edilen öz değerler elde edilen kanonik korelasyonun karesini oluştururken bu öz değerlere karşılık gelen öz vektörler de kanonik korelasyon katsayılarını vermektedirler. Denklem 4.20' de r değeri yerine konulduğu zaman a ve b

kanonik katsayıları elde edilecektir. İlk önce en büyük kanonik korelasyon yani r değeri denklemde yerine konulur ve (W_1, V_1) birinci kanonik korelasyon çifti elde edilir. Bu yerine koyma işlemi büyükten küçüğe, değişken sayısı az olan veri setindeki değişken sayısı kadar devam eder (Turan, Trakya Üniversitesi, Türkiye, 2019: 11).

Eşitlik 4.20’ de yer alan formül genelleştirilmiş öz değer - öz vektör problemi olarak aşağıdaki şekilde de yazılabilmektedir.

$$\begin{bmatrix} 0 & C_{12} \\ C_{21} & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = r \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad 4.21$$

r_1 değeri en büyük öz değer olarak ele alındığı için ilk değişken çifti de en yüksek korelasyona sahip olacak şekilde seçilecektir. İkinci değişken çifti ise ilk r değeri hesaplama dışı bırakılarak elde edildiği için en yüksek ikinci ilişkiyi ortaya koymaktadır. Bu işlem tüm kanonik korelasyon değerleri bulunana kadar devam etmektedir.

Ancak elde edilen p tane kanonik korelasyon değişken çifti de birbirinden bağımsız olmak durumunda olduğu unutulmamalıdır.

$$Korelasyon (V_i, V_j) = 0 \text{ ve } i \neq j$$

$$Korelasyon (W_i, W_j) = 0 \text{ ve } i \neq j$$

$$Korelasyon (W_i, V_j) = 0 \text{ ve } i \neq j$$

4.3. Korelasyon Matrisi İle Kanonik Korelasyon Hesaplaması

Veri setinde yer alan değişkenlerin ölçü birimlerinin birbirlerinden farklı olması durumunda ya değişkenler standardize edilmelidir ya da korelasyon matrisi ile analizin yapılması gerekmektedir. Korelasyon matrisinde değişkenler arası ilişkiler de göz önüne sunulmaktadır. Korelasyon matrisi sayesinde değişkenler arası ilişkilerin yönü ve büyüklüğü daha net ortaya çıkmakta olup, analiz daha iyi yorumlanabilir hale getirmektedir. Bu nedenle analizlerde kolaylık sağlamaktadır (Çankaya, Çukurova Üniversitesi, Türkiye, 2005:35).

Korelasyon matrisi ile yapılan hesaplama, iki deęişken kümesinden oluşturulan matrislerin çarpımlarından oluşan yeni matrisin öz deęerlerinin ve öz vektörlerinin bulunma sürecidir (Alpar, 2013;823). Dört bağımlı deęişken ve üç bağımsız deęişken örneęi ile hazırlanan örnek korelasyon matrisi gösterimi aşağıda yer almaktadır:

Tablo 4.1: Örnek Korelasyon Matrisi

Deęişkenler	X ₁	X ₂	X ₃	Y ₁	Y ₂	Y ₃	Y ₄
X ₁	1	r ₁₂	r ₁₃	r ₁₄	r ₁₅	r ₁₆	r ₁₇
X ₂	r ₂₁	1	r ₂₃	r ₂₄	r ₂₅	r ₂₆	r ₂₇
X ₃	r ₃₁	r ₃₂	1	r ₃₄	r ₃₅	r ₃₆	r ₃₇
Y ₁	r ₄₁	r ₄₂	r ₄₃	1	r ₄₅	r ₄₆	r ₄₇
Y ₂	r ₅₁	r ₅₂	r ₅₃	r ₅₄	1	r ₅₆	r ₅₇
Y ₃	r ₆₁	r ₆₂	r ₆₃	r ₆₄	r ₆₅	1	r ₆₇
Y ₄	r ₇₁	r ₇₂	r ₇₃	r ₇₄	r ₇₅	r ₇₆	1

Kanonik korelasyonun hesaplanabilmesi için öncelikle tüm deęişkenler arasındaki korelasyon katsayıları hesaplanmalıdır. Bu korelasyon katsayıları yardımıyla çözüm gerçekleştirilecektir.

$$R = \begin{bmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{bmatrix}$$

R_{11} = Bağımsız deęişkenler arası korelasyon matrisi

R_{22} = Bağımsız deęişkenler arası korelasyon matrisi

R_{12} ve R_{21} = Bağımlı ve bağımsız deęişkenler arası ikili korelasyon matrisi (R_{21}, R_{12} 'nin devrik matrisidir.)

Kanonik korelasyonun belirlenmesinde H_1 veya H_2 matrisleri kullanılır.

$$H_1 = R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} \text{ veya } H_2 = R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} \quad (4.22)$$

Kanonik korelasyon analizi daha sonra denklem (4.22)'deki matrislere ilişkin öz deęer ve öz vektörleri çözerek devam edecek ve sonuçlandırılacaktır. Öz vektörler hesaplanmadan önce denklem 4.23' den öz deęerlerin hesaplanması gerekmektedir.

$$|R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} - \lambda_i| = 0 \text{ veya } |R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} - \lambda_i| = 0 \quad (4.23)$$

Bir matrisin öz değerlerinin çözümlenmesi, matristeki varyansın birçok tek tek değişken yerine birkaç kompozit değişken içinde yeniden dağıtılması sürecidir. Her bir öz değere denk gelen öz vektör, orijinal değişkenlerden kompozit değişkeni oluşturmak için kanonik katsayılarla dönüştürülür. Amaç orijinal değişkendeki varyansı az sayıda kanonik değişken çiftine dağıtmaktır (Tabachnick, Fidell, 2015: 578).

Kanonik korelasyon çiftine ait her bir öz değer kanonik korelasyonun karesine eşittir. Başka bir deyişle her bir kanonik değişken için hesaplanan öz değerin karekökü alındığında kanonik korelasyon değeri elde edilmektedir (Tabachnick, Fidell, 2015: 578). Kanonik korelasyon ve öz değer arasındaki ilişki denklem 4.24'de yer almaktadır.

$$\lambda_i = r_{wv}^2 \quad 4.24$$

X değişken kümesine ait kanonik katsayıların belirlenmesinde H_1 matrisinden yararlanılır. X değişken setine ait en büyük öz değer için (λ_1) standardize edilmemiş öz vektör elemanı e_1 , eşitlik 4.25 ile hesaplanabilmektedir.

$$(H_1 - \lambda_1 I) e_1 = 0 \quad 4.25$$

Y değişken kümesine ait kanonik katsayıların belirlenmesinde H_2 matrisinden yararlanılır. Y değişken setine ait en büyük öz değer için (λ_1) standardize edilmemiş öz vektör elemanı e_1 , eşitlik 4.25 ile hesaplanabilmektedir.

$$(H_2 - \lambda_1 I) e_1 = 0 \quad 4.26$$

X ve Y kümeleri için standardize edilmiş kanonik katsayılar sırasıyla denklem 4.27 ve denklem 4.28 ile hesaplanabilmektedir (Özdamar, 1999).

$$a_i = \frac{e_1}{\sqrt{\text{Var}(W_i)}} \quad 4.27$$

$$b_i = \frac{e_1}{\sqrt{\text{Var}(V_i)}} \quad 4.28$$

Buradan elde edilen öz vektörler orijinal değişkende oluşan bir birimlik standart artışa karşı kanonik değişkenlerde meydana gelen değişimi göstermektedir. Tüm öz değerler için aynı formül kullanılarak öz vektörleri hesaplanmaktadır.

4.4. Kanonik Değişkenlerle Orijinal Değişkenler Arasındaki Korelasyonlar

Örnekleme genişliği küçük veya değişkenler arasında çoklu bağıntı varsa standardize edilmiş kanonik katsayılar yerine kanonik değişkenle o kümede yer alan orijinal değişkenler arasındaki korelasyon katsayılarının kullanılması önerilmektedir (Çankaya, Çukurova Üniversitesi, Türkiye, 2005:38). Kanonik değişkenlerle orijinal değişkenler arasındaki korelasyonlar kanonik değişkene herhangi bir orijinal değişkenin ne ölçüde katkı sağladığını göstermek için kullanılır.

W_i kanonik değişkeninin kendi veri setindeki (X) orijinal değişkenler arasındaki ilişki:

$$Kor (W_i, X_i) = \frac{Kor (W_i, X_i)}{[Köş (Var (W_i))] [Köş (Var (X_i))]} = \frac{a'_i C_{11}}{\sqrt{[Köş (C_{11})]}}$$

W_i kanonik değişkeninin karşı veri setindeki (Y) orijinal değişkenler arasındaki ilişki:

$$Kor (W_i, Y_i) = \frac{a'_i C_{12}}{\sqrt{[Köş (C_{22})]}}$$

V_i kanonik değişkeninin kendi veri setindeki (Y) orijinal değişkenler arasındaki ilişki:

$$Kor (V_i, X_i) = \frac{b'_i C_{21}}{\sqrt{[Köş (C_{11})]}}$$

V_i kanonik değişkeninin karşı veri setindeki (X) orijinal değişkenler arasındaki ilişki :

$$Kor (V_i, Y_i) = \frac{b'_i C_{22}}{\sqrt{[Köş (C_{22})]}}$$

şeklindedir.

4.5. Kanonik Korelasyon Analizinde Katsayıların Anlamlılık Testi

Kanonik değişkenleri bulmadan önce kanonik korelasyonların anlamlılığı üzerine hipotez testlerinin yapılması uygundur (Alpar, 2013:824). Başka bir deyişle kanonik değişken çiftlerinin, değişken grupları arasındaki kovaryansın kaç tanesi ile büyük oranda açıklanabildiğine karar verilmelidir. Hangi kanonik çiftlerin önemli olduğunu ve hangi kanonik katsayıların anlamlı olduğunu belirlemek için farklı testler mevcuttur (Kandemir, 2018: 2027). Hipotelerin test edilmesi için Wilks' Lambda, Pillai's Trace, Lawley-Hotelling Trace, Roy's Largest Root ve Barlett yöntemleri en yaygın kullanılan yöntemlerdir (Ünlükaptan, 2009:243).

$$H_0: r_1 = r_2 = \dots = r_i = 0$$

$$H_A: \text{En az bir } r_i \neq 0$$

Wilks' Lambda Test istatistiği değeri:

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

$$\chi^2_{\text{hes}} = - \left[(n - 1) - \frac{1}{2}(p + q + 1) \right] \ln \lambda \quad 4.29$$

Eşitlik (4.29)' da verilen test istatistiği pq serbestlik dereceli χ^2 tablo değeri ile karşılaştırılmalıdır. $r_i^2 >$ kritik değer ise, H_0 hipotezi red edilir. Bu birinci kanonik korelasyonun anlamlı olduğunu göstermektedir. Birinci korelasyon saf dışı bırakılarak işleme devam edilir. Diğer kanonik değişkenler için işlem yenilenir.

Her adımda kanonik korelasyon büyükten küçüğe sıralanır ve maksimum değer kaldırılır. Aynı işlemler $i+1$ ' inci kanonik korelasyon için H_0 hipotezi kabul edilinceye kadar yani r_i^2 değeri önemsiz çıkana kadar devam eder (Kandemir, 2018: 2030).

Kanonik katsayıların anlamlılığı test edildikten sonra yorumlama aşamasına geçilebilir. Kanonik değişkenlerden anlamlı olanların yorumu yapılmalıdır. Yorum yapılırken standartlaştırılmış katsayılar kullanılır. Standartlaştırılmış kanonik katsayılar ilgili değişkenin kanonik değişkenlerinin tanımlanmasındaki standart ağırlıklarını göstermektedir. Örnek genişliğinin çok küçük ya da değişkenler arasında çoklu bağlantının olması durumunda, kanonik katsayılar durağan olmayabilir. Bu durumda

standartlaştırılmış kanonik katsayılar yerine, kanonik yükler olarak ifade edilen kanonik değişken ile o kümede yer alan orijinal değişkenler arasındaki basit korelasyon katsayılarının kullanılması önerilmektedir (Albayrak, 2006:485).

4.6. Kanonik Ağırlıklar ve Kanonik Yükler

Orijinal değişkenlerden her birinin kanonik değişkene yaptığı katkının derecesine **kanonik ağırlık** denilmektedir. Daha büyük değere sahip orijinal değişkenler, kanonik değişkene daha fazla katkı sağlamaktadırlar. Kanonik ağırlıklar KKA' ni yorumlamak için kullanılmaktadır. Kanonik ağırlıkların yorumlanması, orijinal değişkenlerin her birinin kanonik değişkene katkı derecesini göstermektedir ve regresyon denklemindeki beta katsayılarının yorumlanmasına benzemektedir. Bundan dolayı, orantılı olarak daha fazla ağırlığa sahip olan orijinal değişkenler, kanonik değişkene daha fazla katkıda bulunmaktadır.

Orijinal değişkenler ile kanonik değişkenler arasındaki basit korelasyon katsayılarına ise **kanonik yük** denilmektedir. Bir kümedeki orijinal değişkenler ile diğer kümenin kanonik değişkenleri arasındaki basit doğrusal korelasyon katsayıları ise **kanonik çapraz yük** olarak ifade edilmektedir (Kalkan ve Özden, 2017: 6).

Kanonik ağırlıklar orijinal değişkenlerin kanonik analize nasıl katkıda bulunduğunu göstermediğinden, orijinal değişkenlerin katkısı **standart kanonik korelasyon katsayıları** kullanılarak belirlenmektedir.

4.7. Redundancy (Açıklanabilirlik) İndeksinin Hesaplanması

Büyük örneklerde; küçük kanonik korelasyon katsayıları istatistiksel olarak anlamlı olabilirken, X ve Y değişken kümeleri arasındaki büyük kanonik korelasyon katsayıları bu kümeler arasında güçlü bir korelasyon olduğunu göstermeyebilir. Benzer şekilde büyük örneklerdeki zayıf kanonik korelasyonlar olduğunda dahi anlamlı sonuçlar elde edilebilmektedir. Çünkü kanonik korelasyon, X ve Y değişkenlerinin rasyonelliğini en üst düzeye çıkarmaktadır. X ve Y değişken kümelerinin herhangi birinde varyasyonun diğeri tarafından açıklanan kısmı belirtilmemektedir. Bundan dolayı, Steward ve Love (1968) tarafından öngörülen redundancy endeksi hesaplanmaktadır. Bu indeks, kümelerden birindeki varyasyonun diğeri ile açıklanan bölümünü göstermektedir.

Redundancy indeksi (RI) her kanonik korelasyon için hesaplanabilmektedir (Keskin ve Özsoy, 2004: 59). Ancak, genellikle ilk kanonik korelasyon dikkate alındığı için sadece ilki için hesaplanır.

Kanonik değişkenler ile orijinal değişkenler arasındaki korelasyonlar incelendiğinde istenen kanonik korelasyon için her bir kümedeki açıklanan varyans elde edilir. Kanonik faktör yüklerinin kareleri her bir değişkene karşılık gelen varyans oranını yansıtır. Karelerin ortalaması ise ilgili kanonik yüklere ilişkin veri kümesindeki değişkenliğin ortalama ne oranda açıkladığı konusunda bilgi verir (Alpar, 2011:792).

Bağımlı değişken tarafından açıklanan yüksek kanonik korelasyon ve yüksek derecede ortak varyans, yüksek bir redundancy indeksi vermektedir. Her kanonik fonksiyon tarafından açıklanabilen paylaşılan varyansın miktarını bulmak için iki bileşenin çarpılmasıyla elde edilen redundancy indeksi (varyasyonun karesel kanonik korelasyon ile çarpımıyla paylaşılan varyans) matematiksel olarak aşağıdaki gibidir (Takma, vd. 2017: 95).

$$OV(Y/V_i) = \sum_{i=1}^p \frac{LY_{ij}^2}{p}$$

$$RI_{W_i/V_i} = OV(Y/V_i) \times r_{wv}^2$$

4.30

Eşitlik 4.30' da $OV(Y/V_i)$; kanonik değişken tarafından Y değişken setindeki açıklanan ortalama varyansı, LY_{ij} ; i. kanonik değişken üzerinde Y değişkenlerinin j. kanonik ağırlığını göstermektedir. r_{wv}^2 W ve V kanonik değişkenleri arasındaki ortak varyansı verdiği bilindiğine göre RI_{W_i/V_i} , ortalama varyans ile ortak varyansın çarpımına eşit olmaktadır (Albayrak, 2006 :487).

Toplam açıklanabilirlik ise bir kümedeki varyasyonun, diğer veri seti ile toplam açıklanabilen kısmıdır. Y değişken kümesindeki toplam redundancy indeksi ($TRI_{Y/X}$) X değişkenlerinin Y değişkeninde açıkladığı toplam varyansı vermektedir. $TRI_{Y/X}$ formülü, eşitlik 4.31' de gösterilmektedir.

$$TRI_{Y/X} = \sum_{i=1}^p RI_{W_i/V_i} \quad 4.31$$

4.8. Kanonik Korelasyon Hesaplama Örneği

Kanonik Korelasyon Hesaplaması Örneği için Tablo 4.2’ deki veri seti kullanılmıştır. Örnek veri setinde sekiz kişi aylık geliri, eğitim durumu ve geri dönüşüm bilinci, çevre bilinci açısından nitelendirilmiştir. Nitelendirilen puanların ortalamaları alınmıştır. Analizin amacı, kişinin sosyal konumu belirleyen göstergeleri ile çevre ve geri dönüşüm bilinci göstergeleri arasında bir ilişki olup olmadığının belirlenmesidir. Bu örneğin hesaplaması elle ya da bilgisayar ile yapılabilir.

Tablo 4.2: Kanonik Korelasyon Hesaplama Örneği İçin Kullanılan Veri Seti

Sıra	X		Y	
	Aylık Gelir (TL)	Eğitim Durumu	Geri Dönüşüm Bilinci	Çevre Bilinci
1	2000	2	2	3
2	2000	1	0	2
3	3000	3	3	3
4	3500	4	3	2
5	8000	5	4	4
6	5000	5	3	3
7	5500	6	3	4
8	4000	7	2	3
Not: Yüksek puanlar yüksek niteliği göstermektedir.				

1.Adım: Kanonik korelasyon analizinde birinci adım korelasyon matrislerinin üretilmesidir.

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

R_{11} : Bağımsız değişkenler arasındaki korelasyonlar

R_{22} : Bağımlı değişkenler arasındaki korelasyonlar

R_{12} : Bağımsız ve bağımlı değişkenler arası korelasyonlar

R_{21} : Bağımlı ve bağımsız değişkenler arası korelasyonlar

Tablo 4.3: Tablo 4.2’ teki Veri Seti İçin Korelasyon Matrisi

	Gelir /TL	Eğitim Durumu	Geri Dönüşüm Bilinci	Çevre Bilinci
Gelir /TL	1	0,66	0,74	0,75
Eğitim Durumu	0,66	1	0,56	0,59
Geri Dönüşüm Bilinci	0,74	0,56	1	0,63
Çevre Bilinci	0,75	0,59	0,63	1

$$H_1 = R_{11}^{-1}R_{12} R_{22}^{-1}R_{21} \quad veya \quad H_2 = R_{22}^{-1}R_{21} R_{11}^{-1}R_{12} \quad 4.32$$

Denklem 4.32’ deki formüllerle önce öz değerler (kanonik korelasyon katsayısının karesi), ardından ise kanonik fonksiyonların katsayılarını verecek olan öz vektörler hesaplanır.

2.Adım: Kanonik korelasyonu değerini elde edebilmek için hesaplanması gereken determinantlar denklem 4.33’ de verilmiştir. Oluşturulan matrisler formülde yerine yazılarak öz değerler hesaplanır.

$$|R_{11}^{-1}R_{12} R_{22}^{-1}R_{21} - \lambda_i I| = 0 \quad veya \quad |R_{22}^{-1}R_{21} R_{11}^{-1}R_{12} - \lambda_i I| = 0 \quad 4.33$$

$$H1 = \begin{bmatrix} 1,77 & -1,16 \\ -1,16 & 1,77 \end{bmatrix} \begin{bmatrix} 0,74 & 0,75 \\ 0,56 & 0,59 \end{bmatrix} \begin{bmatrix} 1,66 & -1,05 \\ -1,05 & 1,66 \end{bmatrix} \begin{bmatrix} 0,74 & 0,56 \\ 0,75 & 0,59 \end{bmatrix}$$

$$H2 = \begin{bmatrix} 1,66 & -1,05 \\ -1,05 & 1,66 \end{bmatrix} \begin{bmatrix} 0,74 & 0,56 \\ 0,75 & 0,59 \end{bmatrix} \begin{bmatrix} 1,77 & -1,16 \\ -1,16 & 1,77 \end{bmatrix} \begin{bmatrix} 0,74 & 0,75 \\ 0,56 & 0,59 \end{bmatrix}$$

$$H_1 - \lambda I = 0 \quad veya \quad H_2 - \lambda I = 0 \quad 4.34$$

Hesaplamalar sonucunda elde edilen öz değerler denklem 4.35’ de verilmiştir.

$$\lambda_1 = 0,692 \quad \lambda_2 = 0,00012 \quad 4.35$$

3.Adım: Doğrusal bileşenli katsayılar olan a_i ve b_i vektörleri, denklem 4.36 ile hesaplanır.

$$\begin{aligned}(R_{11}^{-1}R_{12}R_{22}^{-1}R_{21} - \lambda I). a &= 0 \\ (R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} - \lambda I). b &= 0\end{aligned}\tag{4.36}$$

$$a = \begin{bmatrix} 0,890 & -0,988 \\ -0,156 & 1,321 \end{bmatrix} \quad b = \begin{bmatrix} -0,541 & 1,172 \\ -0,566 & 1,160 \end{bmatrix}$$

$$W_1 = 0,890 x_1 + -0,988x_2$$

$$W_2 = -0,156x_1 + 1,321x_2$$

$$V_1 = -0,541y_1 + 1,172y_2$$

$$V_2 = -0,566y_1 + 1,160y_2$$

5. BÖLÜM

UYGULAMA

Bu bölümde-Türkiye'nin kurucu üye sıfatı yer aldığı İktisadi İşbirliği ve Gelişme Teşkilatı (OECD), demokratik yapılara ve piyasa ekonomisine sahip 36 ülkenin sorunlarını çözmek ve fırsatlardan faydalanmak üzere bir araya geldiği bir örgüttür. Bu örgütte yer alan ülkelerin gelişmişlik düzeyleri geniş bir yelpazede yer aldığından sağlık risk faktörleri ile ilişkisinin daha net görülebileceği öngörülmektedir.

5.1. Veri Toplama

OECD daha iyi yaşamlar oluşturmayı amaçlamıştır. Dünyanın yakından takip ettiği bu örgütten temin edilecek veriler doğru analiz sonuçlarına ulaşmak adına önem arz etmektedir.

OECD iLibrary internet sitesinden alınan raporda, sağlık risk faktörleri başlığı altında obezite, tütün kullanımı ve alkol kullanımı verileri yer almaktadır. 2017 yılına ait OECD ülkelerinin sağlık risk faktörleri verileri resmi sayfasından indirilerek edinilmiştir.

Refah ve gelişmişlik veri seti altında birçok gösterge mevcuttur. Analizi anlamlı kılacağı düşünülen kişi başı gayrisafi yurtiçi hâsıla, bireysel internet kullanımı, yaşam tatmini ve eğitimsel kazanımlar verileri yine OECD resmi sayfasından indirilmiştir.

Açıklanmamış kayıp veriler ise SPSS programı yardımıyla grup ortalamaları alınarak düzenlenmiştir.

5.2. Veri Tablosu

Tablo 5.1' de yer alan tabloda 2017 yılındaki OECD ülkeleri sağlık risk faktörleri veri seti ile refah göstergeleri veri seti yer almaktadır.

Tablo 5.1 : Uygulama Veri Setleri

Ülke	X1	X2	X3	Y1	Y2	Y3	Y4
	Yüzde	Yüzde	Yüzde	Kişi Başı / Dolar	Yüzde	Endeks	Yüzde
	Obezite	Tütün	Alkol	GSYİH	Bireysel İnt. Kull.	Yaşam Tatmini	Eğitimsel Kazanım
Avustralya	63,40	12,40	9,70	49.743,70	87,46	7,30	81,00
Avusturya	46,70	24,30	11,40	52.708,40	87,82	7,10	85,00
Belçika	51,00		10,40	47.646,30	89,93	6,90	77,00
Kanada	53,10	12,50	8,20	46.359,90	87,60	7,40	91,00
Şili	48,30	24,50	7,90	23.666,60	84,57	6,50	65,00
Çekya	55,40	19,60	11,70	36.959,60	87,21	6,70	94,00
Danimarka	51,00	16,90	9,40	50.724,00	97,85	7,60	81,00
Estonya	51,70	21,30	9,90	31.634,00	89,98	5,70	89,00
Finlandiya	56,00	15,00	8,40	45.257,30	94,66	7,60	88,00
Fransa	46,10	22,40	11,70	42.784,90	89,34	6,50	78,00
Almanya	50,70	18,80	10,90	50.822,10	93,40	7,00	87,00
Yunanistan	56,20	27,30	6,50	27.789,30	73,36	5,40	73,00
Macaristan	54,50	25,80	10,80	28.327,80	79,50	5,60	84,00
İzlanda	57,60	9,40	7,70	53.982,30	99,17	7,50	77,00
İrlanda	62,00	18,00	11,20	76.770,80	84,35	7,00	82,00
İsrail	50,90	17,20	2,60	38.435,10	74,10	7,20	87,00
İtalya	45,20	20,00	7,10	39.792,20	76,52	6,00	61,00
Japonya	25,40	18,30	7,20	43.903,30	96,20	5,90	
Kore	34,50	18,40	8,70	38.275,00	96,06	5,90	88,00
Letonya	52,00	24,10	11,20	27.474,60	84,65	5,90	88,00
Litvanya		20,40	13,20	31.765,20	80,50	5,90	93,00
Lüksemburg	48,00	16,00	11,30	104.202,90	96,56	6,90	77,00
Meksika	72,50	7,60	4,40	19.139,70	67,59	6,50	38,00
Hollanda	47,60	18,00	8,30	53.100,30	95,23	7,40	78,00
Yeni Zelanda	66,60	13,80	8,80	40.916,90	85,00	7,30	79,00
Norveç	46,00	11,00	6,00	61.255,50	97,83	7,60	82,00
Polonya	53,30	22,70	10,40	29.290,60	79,26	6,10	92,00
Portekiz	52,80	16,80	10,70	32.145,40	75,35	5,40	48,00
Slovakya	54,20	22,90	9,90	32.137,40	83,25	6,20	91,00
Slovenya	55,60	18,90	10,50	34.816,00	80,99	5,90	88,00
İspanya	51,50	23,00	8,60	38.127,00	86,61	6,30	59,00
İsveç	48,80	10,90	7,20	50.483,40	92,99	7,30	83,00
İsviçre			9,30	65.331,50	94,52	7,50	88,00
Türkiye	53,90	26,50	1,30	28.241,80	71,94	5,50	39,00
Birleşik Krallık	61,40	16,10	9,50	44.019,10	94,92	6,80	81,00
ABD	65,10	11,80	8,80	59.531,70	78,00	6,90	91,00

Veriler 10.06.2019 Tarihinde OECD Stat. İnternet Sitesinden Alınmıştır.

X1: Obezite Yaygınlığının Nüfus İçerisindeki Yüzdesi

X2: Tütün Kullanımının Nüfus İçerisindeki Yüzdesi

X3: Alkol Kullanımının Nüfus içindeki Yüzdesi

Y1: Gayrisafi Yurtiçi Hâsılanın kişi başına düzen dolar karşılığı

Y2: Bireysel İnternet Kullanımının Nüfus İçerisindeki Yüzdesi

Y3: Yaşam Tatmini Endeksi

Y4: Eğitimsel Kazanımlar Yüzdesi

5.3. Doğrusal Kanonik Korelasyon Analizi Uygulaması

Analiz SPSS22 paket programı kullanılarak gerçekleştirilmiştir. SPSS 22’de kanonik korelasyona ait hazır bir araç çubuğu olmadığı için analiz “*syntex*” yazılarak çalıştırılmıştır.

File → Yeni → Syntex adımları seçilerek Syntex formu açılır. Açılan formda SPSS program dosyasına girilerek *Samples* dosyası içindeki kanonik korelasyon kod dokümanı seçilir. Kod formda açılacaktır. Formdaki tüm satırlar seçilerek play butonuna basılır. Bu aşamadan sonra kanonik korelasyon paket programımızda çalıştırılabilecektir. Ardından aşağıdaki kod başka bir syntex formuna yazılarak çalıştırılır.

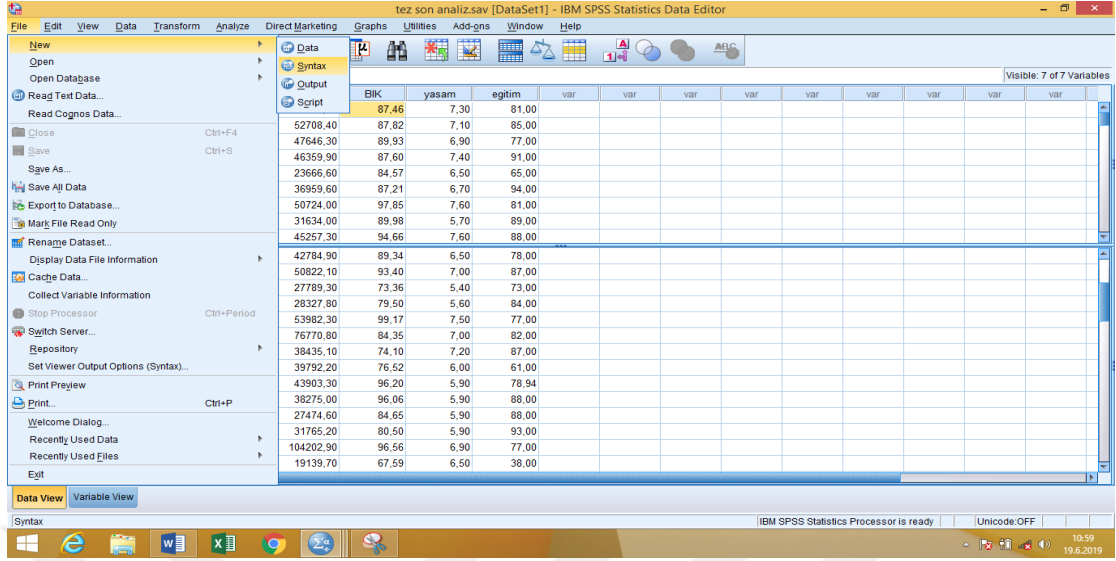
```
cancorr set1=obezite tutun alkol  
/set2=GSYIH BIK yasam egitim.
```

Açılan Output ekranında analize ait tablolar görülebilmektedir. Kanonik Korelasyon Analiz Bulguları başlığı altında elde ettiğimiz sonuçlar incelenecek ve yorumlanacaktır.

Veriler SPSS22 paket programına ilk üç değişken birinci küme, son dört değişken ise ikinci küme olacak şekilde giriş yapılmıştır.

Analiz amaçları arasında sağlık risk faktörleri ile refah göstergelerinin arasında anlamlı bir ilişki olup olmadığını anlamak, ilişki var ise bu ilişkiye en çok katkı sağlayan değişkenleri belirlemek, bir değişkenin kendi kümesine ve karşı kümeye ne derecede etki ettiğini görmek, bir değişken grubunun ne kadarının diğer grubun yüzde kaçını açıkladığını tahmin etmek yer almaktadır.

Şekil 5.1 : SPSS 22 Ana Ekran Görüntüsü



	BIK	yasam	egitim	var	var	var	var	var	var	var	var	var
52708.40	87.46	7.30	81.00									
47646.30	89.93	6.90	77.00									
46359.90	87.60	7.40	91.00									
23666.60	84.57	6.50	65.00									
36959.60	87.21	6.70	94.00									
50724.00	97.85	7.60	81.00									
31634.00	89.98	5.70	89.00									
45267.30	94.66	7.60	88.00									
42784.90	89.34	6.50	78.00									
50622.10	93.40	7.00	87.00									
27789.30	73.36	5.40	73.00									
28327.80	79.50	5.60	84.00									
53982.30	99.17	7.50	77.00									
76770.80	84.35	7.00	82.00									
38435.10	74.10	7.20	87.00									
39792.20	76.52	6.00	61.00									
43903.30	96.20	5.90	78.94									
38275.00	96.06	5.90	88.00									
27474.60	84.65	5.90	88.00									
31765.20	80.50	5.90	93.00									
104202.90	96.56	6.90	77.00									
19139.70	67.59	6.50	38.00									

5.4. Kanonik Korelasyon Analiz Bulgularının Yorumlanması

Ek1’de SPSS analiz raporunun *Output* ekran çıktısı yer almaktadır. Bu bölümde ise *Output* ekranından elde ettiğimiz veriler düzenlenerek birleştirilmiş ve yoruma hazır hale getirilmiştir.

Uygulamamızın X veri setinin değişken sayısı 3, Y veri setinin değişken sayısı 4 olmasından dolayı, $\min(p,q) = 3$ olduğu görülmektedir. Bu nedenle üç tane kanonik korelasyon hesaplanacaktır.

Analizi yorumlama başlamadan önce hesaplanan bu üç kanonik korelasyona ait anlamlılık testinin sonucuna bakılmalıdır. SPSS analiz çıktısında Wilks Lambda istatistiği tablosuna (Tablo 5.2) baktığımızda ilk iki kanonik değişkenin Sig. değerinin 0,05’in altında olduğu görülmüştür. Başka bir deyişle ilk iki kanonik korelasyon anlamlıdır ve yorumlamaya değerdir. İlk iki kanonik korelasyonun anlamlı çıkması nedeniyle bundan sonraki analiz bulguları bu iki kanonik korelasyon için yorumlanacaktır.

Tablo 5.2: Kanonik Katsayıların Anlamlılık Testi

	Wilk's	Chi-SQ	DF	Sig.
1	0,169	55,137	12,000	0,000
2	0,489	22,185	6,000	0,001
3	0,837	5,534	2,000	0,063

- **Kanonik Korelasyon**

Tablo 5.3' de değişkenler arası ilişkiyi gösteren kanonik korelasyon matrisi oluşturulmuştur. Kanonik korelasyon matrisi, birinci veri setinin değişkenleri arasındaki, ikinci veri setinin değişkenleri arasındaki ve birinci veri seti ile ikinci veri seti değişkenleri arasındaki dört korelasyon matrisinin çarpımı olduğu önceki bölümde belirtilmişti.

Tablo 5.3: Uygulama Değişkenleri Arası Korelasyonlar

Değişken	X ₁	X ₂	X ₃	Y ₁	Y ₂	Y ₃	Y ₄
X ₁	1,000	-0,310	-0,140	0,800	0,420	0,190	0,140
X ₂	-0,310	1,000	0,180	0,380	0,270	0,630	0,008
X ₃	-0,140	0,180	1,000	0,210	0,300	0,040	0,520
Y ₁	0,800	0,380	0,210	1,000	0,560	0,590	0,250
Y ₂	0,420	0,270	0,300	0,560	1,000	0,560	0,440
Y ₃	0,190	0,630	0,040	0,590	0,560	1,000	0,300
Y ₄	0,140	0,008	0,520	0,250	0,440	0,300	1,000

X ve Y değişken setlerinin veri türleri birbirinden farklı olduğu için KKA korelasyon matrisi üzerinden hesaplanmıştır.

$$R = \begin{bmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{bmatrix}$$

Bilindiği gibi oluşturulan korelasyon matrisi üzerinden öz değerler ;

$$H_1 = R_{11}^{-1} R_{12} R_{22}^{-1} R_{21}$$

$$H_2 = R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} \text{ formülleri ile hesaplanmaktadır.}$$

$|R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} - \lambda_i| = 0$ matrisinin öz değerleri hesaplanıp, öz değerlerin karekökü alındığında kanonik korelasyon değerleri elde edilmektedir.

$$\sqrt{\lambda_i} = r_i$$

Tablo 5.4: Kanonik Korelasyon Tablosu

1	0,809
2	0,645
3	0,404

Tablo 5.4’ de görüldüğü üzere en yüksek değer 0,809 ile birinci kanonik korelasyona aittir. İkinci kanonik korelasyon değeri 0,645’ tir. Ve test istatistiği sonucunda anlamlı çıkmayan üçüncü kanonik korelasyon değeri ise 0,404’tür.

- **Kanonik Katsayılar**

Öz değerler hesaplaması ile kanonik korelasyon bulunduktan sonra öz vektörlerin hesaplanmasına geçilmelidir. Teorik bölümde belirtildiği üzere ham verilerden oluşturulan öz vektör matrisi ve standartlaştırılmış öz vektör matrisi hesaplanabilmektedir. 5.1’ deki denklemler standartlaştırılmamış öz vektör matrisinin hesaplama formüllerini gösterirken; 5.2’ deki denklemler standartlaştırılmış öz vektör formüllerini ifade etmektedir. Veri ilişkilerinin daha net görülmesi için standartlaştırılmış öz vektör matrisi hesaplanacaktır.

$$\begin{aligned} (H_1 - \lambda_1 I) e_1 &= 0 \\ (H_2 - \lambda_1 I) e_1 &= 0 \end{aligned} \tag{5.1}$$

$$a_i = \frac{e_i}{\sqrt{\text{Var}(W_i)}}$$

$$b_i = \frac{f_i}{\sqrt{\text{Var}(V_i)}} \quad 5.2$$

Hesaplamalar sonucu oluşan birinci sete ait standartlaştırılmış öz vektör matrisi Tablo 5.5’ de görülmektedir. Aynı zamanda bu değerler kanonik korelasyon denklemlerinin katsayılarını oluşturmaktadırlar.

Tablo 5.5: 1. Veri Seti Kanonik Korelasyon Katsayıları

SET 1	W ₁	W ₂
X ₁	0,759	-0,264
X ₂	0,059	-1,068
X ₃	0,67	0,177

$$W_1 = 0,759 X_1 (\text{Obezite}) + 0,059 X_2 (\text{Tütün K.}) + 0,670 X_3 (\text{Alkol K.})$$

$$W_2 = -0,264 X_1 (\text{Obezite}) + -1,068 X_2 (\text{Tütün K.}) + 0,177 X_3 (\text{Alkol K.})$$

İkinci sete ait standartlaştırılmış öz vektör matrisi ise Tablo 5.6’ da gösterilmektedir.

Tablo 5.6: 2. Veri Seti Kanonik Korelasyon Katsayıları

SET 2	V ₁	V ₂
Y ₁	-0,222	-0,157
Y ₂	-0,862	0,276
Y ₃	0,893	-0,765
Y ₄	-0,393	-0,18

$$V_1 = -0,222 Y_1(\text{GSYİH}) + -0,862 Y_2(\text{BİK}) + 0,893 Y_3 (\text{Yaşam Tatmini}) + -0,393 Y_4 (\text{Eğitimsel Kazanımlar})$$

$$V_2 = -0,157 Y_1 (\text{GSYİH}) + 0,276 Y_2 (\text{BİK}) + -0,765 Y_3 (\text{Yaşam Tatmini}) + -0,180 Y_4 (\text{Eğitimsel Kazanımlar})$$

Standart kanonik korelasyon katsayıları orijinal değişkenlerin kanonik korelasyona yaptığı katkıyı göstermektedirler.

- **Kanonik Yükler**

Tablo 5.7: 1. Veri Setine Ait Kanonik Yükler ve Çapraz Kanonik Yükler

Değişken	W ₁	W ₂	V ₁	V ₂
X ₁	0,750	0,065	0,600	0,041
X ₂	-0,300	-0,954	-0,240	-0,615
X ₃	-0,670	-0,012	-0,540	-0,008

Anlamli sonuç veren ilk iki kanonik korelasyonun kanonik ağırlıklarını birinci set için incelediğimizde karşımıza Tablo 5.7’ deki tablo çıkmaktadır. X₁ (obezite) orijinal değişkeninin, kendi setindeki W₁ kanonik değişkeni arasındaki korelasyon değeri 0,75 olduğu görülmüştür. Aynı şekilde X₃ (alkol k.) ile W₁ arasındaki korelasyon değeri hesaplandığında -0,67 çıktığı görülmektedir. Birinci veri setinin ilk kanonik değişkeni olan W₁ kanonik değişkenine en büyük katkıyı sırası ile X₁ ve X₃ sağlamaktadır. X veri setinin anlamlı ikinci kanonik değişkeni olan W₂ kanonik değişkenine en büyük katkıyı 0,954 korelasyon değeri ile X₂ sağlamaktadır.

X veri seti için çapraz yükleri incelediğimizde ise; karşı (ikinci) veri setinin birinci kanonik değişkeni olan V₁ (GSYİH) kanonik değişkenine en büyük katkıyı sırasıyla 0,606 ve 0,542 korelasyon değerleri ile X₁ ve X₃ sağlamaktadır. Karşı veri setinin ikinci kanonik değişkeni olan V₂ ye ise en büyük katkıyı X₂ (tütün k.) değişkeni sağlamıştır.

Tablo 5.8: 2. Veri Setine Ait Kanonik Yükler ve Çapraz Kanonik Yükler

Değişken	V ₁	V ₂	W ₁	W ₂
Y ₁	-0,280	0,714	-0,225	0,460
Y ₂	-0,663	0,709	-0,537	0,457
Y ₃	0,164	0,956	0,133	0,616
Y ₄	-0,561	0,213	-0,454	0,137

Tablo 5.8’ de ikinci veri setinin kendi kanonik değişkenleri ve karşı kanonik değişkenleri arasındaki ilişki tablosu görülmektedir. İkinci veri setinin V₁ kanonik değişkeni ile en yüksek korelasyona sahip değişken Y₂ (bireysel internet kullanımı) değişkenidir. İkinci sete ait kanonik değişkenler ile orijinal değişkenler arasındaki ilişki incelendiğinde Y₄ (eğitimsel kazanımlar) değişkeni dışındaki tüm değişkenlerle arasında yüksek korelasyon olduğunu söylenebilmektedir. Diğer veri setinin birinci kanonik değişkeni olan W₁ ile en yüksek korelasyona Y₁ (GSYİH) ve Y₂ (bireysel internet

kullanımı) arasında aynı derecede ilişki bulunmaktadır. W_1 'e en az katkı sağlayan değişken ise 0,137 korelasyon değeri ile Y_4 (eğitimsel kazanımlar) değişkenidir.

- **Kanonik Korelasyon ve Kanonik Öz Değer**

Tablo 5.9: Kanonik Korelasyonlar ve Kanonik Öz Değerler

Kanonik Korelasyon Katsayısı (r)		Kanonik Öz Değer Açıklanan Varyans (r^2)
1	0,809	0,654
2	0,645	0,429
3	0,404	0,163

Tablo 5.9' da SPSS analizinde bulunan öz değerler yer almaktadır. Bulunan öz değerlerin karekökü alındığında ise kanonik korelasyon katsayısı elde edilmektedir.

$$\lambda_i = r_{ci}^2$$

Önceki bölümde belirtildiği üzere bu kanonik korelasyonlardan ilk ikisi anlamlıdır. Birinci kanonik korelasyon 0,809 değeri ile iki veri setinin arasında yüksek dereceli bir ilişki olduğunu göstermektedir. İkinci kanonik korelasyon birinciye göre kısmen düşük olmasına rağmen iki veri seti arasında var olan ilişkiyi orta dereceli olarak ifade etmektedir.

Kanonik öz değer, kanonik korelasyonun açıkladığı varyansı göstermektedir. Birinci kanonik değişken en yüksek varyans açıklama değerine sahiptir.

Kanonik korelasyonların kareleri değişkenler arası paylaşılan varyans tahminini buldurmaktadır. Ancak bu değer bağımlı ve bağımsız veri setlerinin doğrusal bileşenlerinin paylaşılan varyansını ifade ettiğinden yetersiz kalmaktadır. Bu belirsizliği ortadan kaldırmak için paylaşılan varyans hesaplamasında gereksizlik ölçüsünün (redundancy analiz) kullanılması önerilmektedir (Orhunbilge,2010).

- **Redundancy (Gereksizlik) Analiz**

Gereksizlik indeksi, i. kanonik değişkenin açıklanan varyans oranı ile i. kanonik korelasyon katsayısının karesinin (i. öz değer) çarpımından elde edilmektedir (Tabachnick ve Fidell, 2007).

Tablo 5.10: 1.Set Redundancy Indeks Tablosu

Set1 (X)	Kendi Setinde Açıkladığı Varyans		Karşı Sette Açıkladığı Varyans	
	CVI-1	0,366	CV2-1	0,240
	CVI-2	0,305	CV2-2	0,127
	CVI-3	0,329	CV2-3	0,054

Tablo 5.10’ da birinci veri setine (X) ait gereksizlik analiz tablosu yer almaktadır. Birinci veri setindeki varyansı açıklama oranı en yüksek olan değişken W_1 kanonik değişkenidir. 1. Setin kendi setindeki varyansları açıklama oranları tüm kanonik değişkenler için birbirlerine oldukça yakın çıkmıştır.

Birinci veri setindeki kanonik değişkenlerin karşı veri setindeki varyansı ne ölçüde açıkladığı yine Tablo 5.10’ de verilmiştir. Birinci veri setine ait W_1 kanonik değişkeni karşı veri setinin varyansını 0,240 oranında açıklamaktadır. İkinci kümeyle ait üç kanonik değişken, birinci setin varyansının %42 sini açıklamaktadır.

Tablo 5.11: 2.Set Redundancy Indeks Tablosu

Set2 (Y)	Kendi Setinde Açıkladığı Varyans		Karşı Sette Açıkladığı Varyans	
	CV2-1	0,215	CVI-1	0,141
	CV2-2	0,493	CVI-2	0,208
	CV2-3	0,158	CVI-3	0,026

Tablo 5.11’ te ikinci veri setine ait olan kanonik değişkenlerin kendi setindeki ve karşı setteki varyansı açıklama oranları görülmektedir. V_2 kanonik değişkeni kendi setinde 0,493 değer ile en yüksek varyansı açıklamaktadır. Yine V_2 değişkeni karşı setteki varyansı en yüksek açıklayan kanonik değişkendir. Ancak açıklama oranı 0,141 değeri

ile oldukça düşüktür. Birinci kümeye ait üç kanonik değişken, ikinci setin varyansının % 37,5 unu açıklamaktadır.

5.5. Sonuç

- I. Birinci SPSS çizelgesinde, birinci veri seti olan sağlık riskleri kümesinin kendi içindeki korelasyonları yer almaktadır. Obezite, tütün kullanımı ve alkol kullanımı arasında yüksek bir korelasyonun olmadığı görülmektedir. Tüm korelasyon değerleri 0,50'nin altındadır. Obezite ile tütün kullanımı arasında 0,31 değerinde korelasyon bulunurken; obezite ile alkol arasında 0,01 gibi düşük bir korelasyon olduğu görülmektedir. Tütün ve alkol kullanımı arasında ise 0,18 değerinde korelasyon tespit edilmiştir.

Correlations for Set-1			
	Obezite	tutun	alkol
obezite	1,0000	-0,3099	-0,0136
tutun	-0,3099	1,0000	0,1799
alkol	-0,0136	0,1799	1,0000

SPSS Çizelgesi 1

- II. İkinci SPSS çizelgesine baktığımızda, ikinci veri seti olan refah göstergeleri kümesinin kendi içindeki korelasyonları yer almaktadır. Kişi başına düşen gayrisafi yurtiçi hâsıla, bireysel internet kullanımı, yaşam tatmini ve eğitimsel kazanımlar arasında yüksek bir korelasyonun olmadığı görülmektedir. Örneğin GSYIH ile Bireysel internet kullanımı arasında 0,56 değerinde korelasyon bulunmuşken GSYIH ile yaşam endeksi arasında grubun en yüksek korelasyonu olan 0,59 değeri tespit edilmiştir.

Correlations for Set-2:				
	GSYIH	BIK	yasam	egitim
GSYIH	1,0000	0,5588	0,5868	0,2517
BIK	0,5588	1,0000	0,5567	0,4439
yasam	0,5868	0,5567	1,0000	0,3024
egitim	0,2517	0,4439	0,3024	1,0000

SPSS Çizelgesi 2

- III. İki veri seti arasındaki (X ve Y) korelasyonlar üçüncü çizelgede yer almaktadır. Tablo incelendiğinde en yüksek korelasyonun GSYİH ile obezite arasında tespit edildiği görülmektedir. En düşük korelasyon ise 0,008 değeri ile eğitimsel kazanımlar ile tütün kullanımı arasındadır.

Correlations Between Set-1 and Set-2:

	GSYİH	BIK	yasam	egitim
obezite	-0,0794	-0,4225	0,1896	-0,1413
tutun	-0,3762	-0,2742	-0,6303	-0,0080
alkol	0,2131	0,2978	-0,0393	0,5167

SPSS Çizelgesi 3

- IV. Çizelge 4' de ise elde edilen kanonik korelasyonları değerleri gösterilmektedir. Üç tane kanonik korelasyon hesaplanmıştır. Kanonik korelasyon değerlerine baktığımızda ilk iki değer 0,50' nin üzerinde olduğu görülmekte idi. Hemen altında yer alan Çizelge 5' de ise bu kanonik korelasyon değerlerinin önemlilik test çıktıları yer almaktadır.

$H_0 = \text{Tüm kanonik korelasyonlar sıfıra eşittir.}$

$H_A = \text{Tüm kanonik korelasyonlar sıfırdan farklıdır.}$

Burada hipotez ilk iki kanonik korelasyon için reddedilir. Yani ilk iki kanonik korelasyon sıfırdan farklıdır ve önemlidir.

Çizelge 5' teki önem değerleri (Sig. Sütunu) göz önüne alındığında ilk iki kanonik korelasyonun önem değerinin 0,05' ten küçük olduğu tespit edilmiştir. İki tabloyu eş zamanlı kontrol ettiğimizde ilk 2 kanonik korelasyonun anlamlı olduğu anlaşılmaktadır. İlerleyen tablolarda ilk iki kanonik korelasyon verileri dikkate alınacaktır.

Canonical Correlations:

1	0,809
2	0,645
3	0,404

SPSS Çizelgesi 4

Test that remaining correlations are zero:

	Wilk's	Chi-SQ	DF	Sig.
1	0,169	55,137	12,000	0,000
2	0,489	22,185	6,000	0,001
3	0,837	5,534	2,000	0,063

SPSS Çizelgesi 5

V. Çizelge 6' da birinci veri seti olan sağlık risk faktörleri değişkenlerinin kanonik katsayıları analiz çıktısı görülmektedir. Bu katsayılardan kanonik korelasyon denklemi oluşmaktadır.

Standardized Canonical Coefficients for Set-1:

	1	2	3
obezite	0,759	-0,264	0,680
tutun	0,059	-1,068	0,037
alkol	-0,670	0,177	0,745

SPSS Çizelgesi 6

$$W1 = 0,759 X_1 + 0,059 X_2 + 0,670 X_3$$

$$W2 = -0,264 X_1 + -1,068 X_2 + 0,177 X_3$$

Yukarıdaki denklemler birinci sete ait kanonik korelasyon denklemleri gösterilmektedir.

VI. 7. Çizelgede ikinci veri seti olan refah göstergeleri değişkenlerinin kanonik katsayıları analiz çıktısı görülmektedir.

Standardized Canonical Coefficients for Set-2:

	1	2	3
GSYIH	-0,222	0,157	0,399
BIK	-0,862	0,276	-0,948
yasam	0,893	-0,765	0,188
egitim	-0,393	-0,180	0,977

SPSS Çizelgesi 7

$$V1 = -0,222 Y_1 + -0,862 Y_2 + 0,893 Y_3 + -0,393 Y_4$$

$$V2 = -0,157 Y_1 + 0,276 Y_2 + -0,765 Y_3 + -0,180 Y_4$$

Yukarıdaki denklemler ikinci set olan refah göstergelerine ait kanonik korelasyon denklemleri ve kanonik korelasyon katsayıları yer almaktadır.

- VII.** 8. Çizelgede, birinci veri seti ile kanonik yükleri arasındaki (W ile X arasında) korelasyon verilmiştir. Yükler tablosuna bakıldığında hem obezite hem de alkol kullanımı değişkenleri W_1 kanonik değişkeni üzerinde etkili olmaktadır. W_2 kanonik değişkeni üzerindeki en büyük etkiyi tütün kullanımı değişkeni yaratmaktadır.

Canonical Loadings for Set-1:

	1	2	3
obezite	,750	,064	,659
tütün	-,296	-,954	-,040
alkol	-,670	-,012	,742

SPSS Çizelgesi 8

- VIII.** Çizelge 9' da birinci veri seti olan sağlık riskleri için çapraz yükler (V ile X değişkenleri arasındaki) verilmiştir. Çapraz yükler tablosuna bakıldığında hem obezite hem de alkol kullanımı değişkenleri, V_1 kanonik değişkeni üzerinde etkili olmaktadır. V_2 kanonik değişkeni üzerindeki en büyük etkiyi tütün kullanımı değişkeni yaratmaktadır.

Cross Loadings for Set-1:

	1	2	3
obezite	0,606	0,041	0,266
tütün	-0,240	-0,615	-0,016
alkol	-0,542	-0,008	0,300

SPSS Çizelgesi 9

- IX.** Çizelge 10' da ikinci veri seti ile kanonik yükleri arasındaki (V ile Y arasında) korelasyon verilmiştir. Yükler tablosuna bakıldığında bireysel internet kullanımı ve eğitimsel kazanımlar değişkenleri V_1 kanonik değişkeni üzerinde etkili olmaktadır.

V₂ kanonik değişkeni üzerindeki en büyük etkiyi yaşam tatmini değişkeni yaratmaktadır. GSYİH ve bireysel internet kullanımı verileri de V₂ kanonik değişkeni üzerinde oldukça etkili olduğu görülmektedir.

Canonical Loadings for Set-2:

	1	2	3
GSYİH	-0,279	0,714	0,225
BIK	-0,663	0,709	-0,187
yasam	0,164	0,956	0,190
egitim	-0,561	0,213	0,714

SPSS Çizelgesi 10

- X. 11. Çizelgede ikinci veri seti olan refah göstergeleri için çapraz yükler (W ile Y değişkenleri arasındaki) verilmiştir. Çapraz yükler tablosuna bireysel internet kullanımı ve eğitimsel kazanımlar değişkenleri W₁ kanonik değişkeni üzerinde ortalama bir etkiye sahiptir.

Cross Loadings for Set-2:

	1	2	3
GSYİH	-0,225	0,460	0,091
BIK	-0,537	0,457	-0,076
yasam	0,133	0,616	0,077
egitim	-0,454	0,137	0,289

SPSS Çizelgesi 11

- XI. 12. çizelgede redundancy analiz tablosu yer almaktadır. Birinci veri setindeki varyansı açıklama oranı en yüksek olan değişken V₁ kanonik değişkenidir. Tüm kanonik değişken varyansları açıklama oranları birbirlerine oldukça yakın çıkmıştır.

Redundancy Analysis: Proportion of Variance of Set-1 Explained by

Its Own Can. Var. Prop Var

CV1-1	0,366
CV1-2	0,305
CV1-3	0,329

SPSS Çizelgesi 12

- XII.** 13. Çizelgede ise birinci veri setindeki varyansı diğer veri setine ait kanonik değişkenlerin ne derecede açıkladığı gösterilmektedir. En büyük değere V_1 değişkeni sahiptir.

Proportion of Variance of Set-1 Explained by Opposite

Can.Var.Prop Var:

CV2-1	0,240
CV2-2	0,127
CV2-3	0,054

SPSS Çizelgesi 13

- XIII.** Çizelge 14' de ikinci veri setindeki varyansı açıklama oranı en yüksek olan değişken V_2 kanonik değişkenidir.

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

Prop Var

CV2-1	0,215
CV2-2	0,493
CV2-3	0,158

SPSS Çizelgesi 14

- XIV.** 15. çizelgede ise ikinci veri setindeki varyansı diğer veri setine ait kanonik değişkenlerin ne derecede açıkladığı gösterilmektedir. İkinci veri setindeki varyansı açıklama oranı en yüksek olan değişken W_2 kanonik değişkenidir

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

Prop Var

CV1-1	0,141
CV1-2	0,205
CV1-3	0,026

SPSS Çizelgesi 15

V_1 , W_1 ve V_2 , W_2 kanonik değişkenleri önemlilik testinde anlamlı çıkmış olmalarına karşın dördünün de varyans açıklama oranları düşük olduğu görülmüştür.

Birinci set sağlık risk faktörlerini, ikinci set ise ülkelerin gelişmişlik göstergelerini temsil etmekte idi. Hesaplanan üç kanonik korelasyonun iki tanesi önem kontrolü testinde anlamlı çıkmıştır. Birinci kanonik korelasyon değeri 0,809, ikinci kanonik korelasyon değeri ise 0,645 olarak tespit edilmiştir. Bu kanonik korelasyon değerleri iki veri seti

arasında bir ilişki olduğunu göstermesine rağmen, ikinci setin (ülke gelişmişlik göstergeleri) birinci seti % 42 oranında açıkladığı görülmüştür.

Sonuç olarak, OECD ülkeleri sağlık risk faktörleri ile refah göstergeleri arasında anlamlı bir ilişki var olduğu görülmüştür. Gelişmişlik arttıkça sağlık risk faktörlerine yönelimin azalması ve alternatif çözümlerin bulunması beklenirken sonuç beklentinin tersine refah arttıkça, sağlık risk faktörlerine maruz kalma sıklığının arttığı görülmüştür.



EKLER

Ek 1 : SPSS Analizi Output Ekran Görüntüsü

Run MATRIX procedure:

Correlations for Set-1

	obezite	tutun	alkol
obezite	1,0000	-,3099	-,0136
tutun	-,3099	1,0000	,1799
alkol	-,0136	,1799	1,0000

Correlations for Set-2

	GSYIH	BIK	yasam	egitim
GSYIH	1,0000	,5588	,5868	,2517
BIK	,5588	1,0000	,5567	,4439
yasam	,5868	,5567	1,0000	,3024
egitim	,2517	,4439	,3024	1,0000

Correlations Between Set-1 and Set-2

	GSYIH	BIK	yasam	egitim
obezite	-,0794	-,4225	,1896	-,1413
tutun	-,3762	-,2742	-,6303	-,0080
alkol	,2131	,2978	-,0393	,5167

Canonical Correlations

1	,809
2	,645
3	,404

Test that remaining correlations are zero:

	Wilk's	Chi-SQ	DF	Sig.
1	,169	55,137	12,000	,000
2	,489	22,185	6,000	,001
3	,837	5,534	2,000	,063

Standardized Canonical Coefficients for Set-1

	1	2	3
obezite	,759	-,264	,680
tutun	,059	-1,068	,037
alkol	-,670	,177	,745

Raw Canonical Coefficients for Set-1

	1	2	3
obezite	,091	-,032	,081
tutun	,012	-,214	,007
alkol	-,266	,070	,295

Standardized Canonical Coefficients for Set-2

	1	2	3
GSYIH	-,222	,157	,399
BIK	-,862	,276	-,948
yasam	,893	,765	,188
egitim	-,393	-,180	,977

Raw Canonical Coefficientsfor Set-2

	1	2	3
GSYIH	,000	,000	,000
BIK	-,102	,033	-,112
yasam	1,250	1,071	,263
egitim	-,028	-,013	,069

Canonical Loadings for Set-1

	1	2	3
obezite	,750	,064	,659
tutun	-,296	-,954	-,040
alkol	-,670	-,012	,742

Cross Loadings for Set-1

	1	2	3
obezite	,606	,041	,266
tutun	-,240	-,615	-,016
alkol	-,542	-,008	,300

Canonical Loadings for Set-2

	1	2	3
GSYIH	-,279	,714	,225
BIK	-,663	,709	-,187
yasam	,164	,956	,190
egitim	-,561	,213	,714

Cross Loadings for Set-2

	1	2	3
GSYIH	-,225	,460	,091
BIK	-,537	,457	-,076
yasam	,133	,616	,077
egitim	-,454	,137	,289

Redundancy Analysis:

Proportion of Variance of Set-1 ExplainedbyItsOwn Can. Var.

Prop Var

CV1-1	,366
CV1-2	,305
CV1-3	,329

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

Prop Var

CV2-1 ,240

CV2-2 ,127

CV2-3 ,054

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

Prop Var

CV2-1 ,215

CV2-2 ,493

CV2-3 ,158

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

Prop Var

CV1-1 ,141

CV1-2 ,205

CV1-3 ,026

KAYNAKÇA

Akaho, S. (2006). A Kernel Method for Canonical Correlation Analysis arXiv preprintcs /0609071.1-7.

Akyıldız, M. (2006). Çok Değişkenli İstatistiklerin Kullanımı. <<http://www.istatistik.gen.tr>>, (Erişim Tarihi: 30.03.2019).

Albayrak A.S., (2006), Uygulamalı Çok Değişkenli İstatistik Teknikleri, (1. Baskı) , Asil Yayın Dağıtım, Ankara

Alpar, R, (2013) Çok Değişkenli İstatistiksel Yöntemler, (4.Baskı), Detay Yayıncılık, Ankara

Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013, February). Deep canonical correlation analysis. In International conference on machine learning (pp. 1247-1255).

Arora, R., Marinov, T. V., Mianjy, P., & Srebro, N. (2017). Stochastic approximation for canonical correlation analysis. In Advances in Neural Information Processing Systems (pp. 4775-4784).

Bi İstatistik <<http://biistatistik.com> /> , (Erişim Tarihi: 29.05.2019)

Bülbül Ş. ve Giray S. , Sosyodemografik Özellikler ile Mutluluk Algısı Arasındaki İlişki Yapısının Analizi”, Ege Akademik Bakış, 11/2011 (Özel Sayı) (113-123)

Çankaya S., (2005), Kanonik Korelasyon Analizi ve Hayvancılıkta Kullanımı, Çukurova Üniversitesi, Zootekni Anabilim Dalı, Adana, Türkiye

Çınaroğlu S. (2017) İlaç Harcamalarının Sağlık Sonuçları ile İlişkisi: Bir Kanonik Korelasyon Analizi Uygulaması, Ankara Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , 35 (2) s(23-47)

Demirbağ Halil Anıl <<https://olcemediginseyiyonetemezsin.wordpress.com/>> , (Erişim Tarihi: 21.05.2019).

Filiz, Z.,& Kolukısaoglu, S. (2012). Doğrusal Olmayan Kanonik Korelasyon Analizi ve Bir Uygulama. Uluslararası Yönetim İktisat ve İşletme Dergisi, 8(16), 59-74.

Gogtay, N. J.,&Thatte, U. M. (2017). Principles of correlation analysis. Journal of the association of physicians of India, 65, 78-81.

Hotelling, H. (1992). Relations between two sets of variates. In Breakthroughs in statistics (pp. 162-190). Springer, New York, NY.

İstatistik Notlarım , <<http://istatistiknotlarim.blogspot.com>> , (Erişim Tarihi: 21.05.2019)

Kalaycı, Ş. (2009) Uygulamalı Çok Değişkenli İstatistik Teknikleri. Ankara: Asil Yayıncılık

Kalkan, S. B.,& Özden, Ü. H. (2017). Dünya üniversitelerinin itibarını etkileyen değişkenlerin kanonik korelasyon analizi ile belirlenmesi, Sosyal Bilimler Araştırma Dergisi, 6(2), 11-19.

Kandemir, A. Ş. (2018). Çalışma Hayatı ve Sosyal Yaşam Arasındaki İlişkinin Kanonik Korelasyon Analizi İle İncelenmesi. İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 7(3), 2025-2038

Kang, S.,Choi, B., & Kim, B. (2003). Linearity analysis of CMOS for RF application. IEEE transactions on microwave the oryand techniques, 51(3), 972-977.

Karagöz, Y. (2016). SPSS ve AMOS 23 Uygulamalı İstatistiksel Analizler, Nobel Yayınevi, Ankara.

Keskin, S., & Özsoy, A. N. (2004). Kanonik korelasyon analizi ve bir uygulaması. Tarım Bilimleri Dergisi, 10(1): 57-71

Kurtuluş, B. (2008). Zaman Serisi Analizlerinde Doğrusal Olmayan Modeller ve Türkiye Ekonomisine Uygulaması. Doktora Tezi. İstanbul Üniversitesi. 1-199.

Moreira, P. S., Santos, N. C., Sousa, N., & Costa, P. S. (2015). The use of canonical correlation analysis to assess the relationship between executive functioning and verbal memory in older adults. Gerontology and geriatric medicine, 1, 2333721415602820. 1-8.

OECD iLibrary<<https://www.oecd-ilibrary.org/statistics>> (Erişim Tarihi: 10.06.2019)

Orhunbilge, N. (2010). Çok Değişkenli İstatistik Yöntemler, İstanbul Üniversitesi Basım ve Yayınevi Müdürlüğü, İstanbul.

Otrar Mustafa <[http:// mustafaotrar.net](http://mustafaotrar.net)>(Erişim Tarihi: 21.05.2019)

ÖZDAMAR, K., 1999. Paket Programlar ile İstatistiksel Veri Analizi-2. Kaan Kitabevi. Eskişehir. 502s.

Özçomak, M., Gündüz, M., Demirci, A., & Yakut, E. (2012). Çeşitli İklim Ve Ürün Verileri Arasındaki İlişkinin Kanonik Korelasyon Analizi Ve Veri Zarflama Analizi Yöntemleri İle İncelenmesi. Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 26(1), 111-131.

Pai, F. Y., Yeh, T. M., & Hung, Y. H. (2015). Analysis on accuracy of bias, linearity and stability of measurement system in balls crew processes by simulation. Sustainability, 7(11), 15464-15486

Şanlı, S. (2018). A Canonical Correlation Approach In Determining Growth & Development And Social Inclusion Linkages. Academic Review of Humanities and Social Sciences, 1(3), 130-146.

Tabachnick B.G, Fidell L.S., (2015) Çok Değişkenli İstatistiklerin Kullanımı (6. Baskı) Nobel Yayıncılık, Ankara.

Takma Ç., Gevrekçi Y., Özsoy A.N., & Çevik, M. (2017). Canonical Correlation Analysis on Egg Production Traits of Quails. SDU Journal of the Faculty of Agriculture / SDÜ Ziraat Fakültesi Dergisi, 12(1): 92-99.

Turan C., Sosyo – Demografik Özellikler Ve Akademik Başarıyı Etkileyen Örgütsel Faktörler Arasındaki İlişkinin Doğrusal Olmayan Kanonik Korelasyon Analizi İle İncelenmesi: Trakya Üniversitesi Meslek Yüksekokulu Öğrencileri Örneği, Ekonometri Anabilim Dalı, Edirne, Türkiye

Türkegün. M. (2015). Çoklu Bağlantı Sorununun Kanonik Korelasyon Belirteçleri Üzerine Etkisi.Yüksek Lisans Tezi. Mersin Üniversitesi, Mersin. 1-131.

Ünlükaptan İ. (2009), “Avrupa Birliği Üyesi Ülkelerde İktisadi Kalkınma, Rekabetçilik ve İnovasyon İlişkilerinin Kanonik Korelasyon Analizi ile Belirlenmesi”, Maliye Dergisi, Sayı 157, ss.235-250

Yüksek Öğretim Başkanlığı Tez Merkezi < [https:// tez.yok.gov.tr/ UlusalTezMerkezi/](https://tez.yok.gov.tr/UlusalTezMerkezi/)> Erişim Tarihi: 08.05.2019):

Yazıcı, A. C., Ögüş, E., Ankaralı, H., & Gürbüz, F. (2010). An application of nonlinear canonical correlation analysis on medical data. Turkish Journal of Medical Sciences, 40(3), 503-510.

Zaid, M. A. (2015). Correlation and Regression Analysis. <[http:// www. oicstatcom. Org / file /TEXTBOOK-CORRELATION-AND-REGRESSION-ANALYSIS-EGYPT-EN.pdf](http://www.oicstatcom.Org/file/TEXTBOOK-CORRELATION-AND-REGRESSION-ANALYSIS-EGYPT-EN.pdf)>, (Erişim Tarihi: 30.03.2019).