
Araştırma Makalesi / Research Article

Asya Ülkelerinin Beklenen Yaşam Süresi Bakımından Sınıflandırılmasında Etkili Olan Sosyoekonomik Değişkenlerin Kısmi En Küçük Kareler Diskriminant Analizi ile Belirlenmesi

Alime Kıvılcım KOZAN¹, Esra POLAT^{2*}

¹*Çukurova Kalkınma Ajansı, Seyhan, Adana*

²*Hacettepe Üniversitesi, İstatistik Bölümü, Beytepe, 06800, Ankara
(ORCID: 0000-0001-8631-2627) (ORCID: 0000-0001-9271-485X)*

Öz

Her ne kadar son yıllarda yapılan çalışmalar doğumda yaşam beklentisinin (DYB) hemen hemen tüm toplumlarda artmakta olduğunu göstermiş olsa da, ölümlerde ve dolayısıyla DYB'de toplum içinde olduğu gibi toplumlarda da önemli farklılıklar vardır. DYB'deki bu eşitsizliğin köklerinin, farklı sosyal grupların farklı sosyoekonomik arka planlarında olduğuna inanılmaktadır. Yaşam beklentisi, en önemli sağlık sonuçlardan biri ve insani gelişmişliğin önemli bir göstergesi olarak kabul edilmektedir. Bu çalışmanın amacı, bir kaç sosyoekonomik değişkenin beklenen yaşam süresi ile ilişkisini araştırmak ve Asya'daki 51 ülke için DYB'ye etkili olan en önemli değişkenleri belirlemektir. Bu amaçla, Dünya Nüfus Veri Sayfası, 2015'den elde edilen veri kümesi üzerinde Kısmi En Küçük Kareler Diskriminant Analizi uygulanmıştır. Analiz sonucunda Asya ülkelerinin beklenen yaşam sürelerine göre Asya ortalamasının üstünde ve altında olmak üzere iki sınıf olarak sınıflandırılmasında en önemli/etkili değişkenin Bebek Ölüm Oranı olduğu görülmüştür.

Anahtar kelimeler: Beklenen Yaşam Süresi, Sosyoekonomik Değişkenler, Kısmi En Küçük Kareler Diskriminant Analizi.

Determination of Socioeconomic Variables Affecting the Classification of Asian Countries in Terms of Life Expectancy by Partial Least Squares Discriminant Analysis

Abstract

Although studies conducted in recent years have shown that life expectancy at birth (LEB) is increasing in almost all societies, there are significant differences between deaths and, therefore, in LEB, occur among societies as well as within society. It is believed that this inequality in the LEB have its roots in different socioeconomic backgrounds of different social groups. Life expectancy is recognized as one of the most important health outcomes and an important indicator of human development. The aim of this study is to investigate the relationship between several socioeconomic variables and life expectancy and to identify the most important variables that affect the LEB for 51 countries in Asia. For this purpose, Partial Least Squares Discriminant Analysis was applied on the data set obtained from the World Population Data Sheet, 2015. As a result of the analysis, it was found that the most important/effective variable in the classification of Asian countries into two classes as above and below the Asian average according to their life expectancy, is the Infant Mortality Rate.

Keywords: Life Expectancy, Socioeconomic Variables, Partial Least Squares Discriminant Analysis.

*Sorumlu yazar: espolat@hacettepe.edu.tr

Geliş Tarihi: 18.09.2019, Kabul Tarihi: 05.02.2020

1. Giriş

Doğumda beklenen yaşam süresi, insani gelişmişliğin önemli bir göstergesidir. Yaşam beklentisi oranı, evrensel olarak özellikle gelişmiş ülkelerde yaşayan savunmasız topluluklarda, sosyoekonomik risk değişkenlerinden büyük ölçüde etkilenmiştir [1]. Her hükümetin temel hedeflerinden biri, ölüm oranını mümkün olan en düşük düzeye indirerek, nüfusunun yaşam süresini uzatmaktır. Son yıllarda yapılan çalışmalar doğumda yaşam beklentisinin (DYB) neredeyse tüm toplumlarda artmakta olduğunu göstermiş olmasına karşın [2], hem toplum içinde hem de toplumlar arasında önemli ölüm oranı farklılıkları bulunmaktadır. Ölüm oranındaki bu eşitsizliğin köklerinin, farklı sosyal grupların farklı sosyoekonomik geçmişinde olduğuna inanılmaktadır. Çünkü sosyoekonomik değişkenler, sonunda hastalığa ve ölüme yol açan biyolojik süreçleri etkilemektedir. Bu amaçla, Dünya Bankası, Dünya Sağlık Örgütü, diğer uluslararası kuruluşlar ve kalkınma yardımı bağışçıları sosyoekonomik statüler ile sağlık eşitsizliğini azaltmaya çalışarak [3], dezavantajlılar arasındaki ölüm oranlarını azaltmaya çalışmaktadır. Ancak, daha geniş sosyal ve ekonomik eğilimler seviyelerindeki değişiklikler nedeniyle sağlık durumundaki zaman içinde meydana gelen değişiklikleri ve dolayısıyla ölüm oranlarındaki değişiklikleri de tespit etmek için çok fazla çaba gösterilmemiştir [4]. Bu olgu, özellikle sosyoekonomik gelişme ile yaşam beklentisi arasındaki bağlantının doğrudan olduğu algısının ardından ivme kazanmış ve araştırmacıların yaşam beklentisini etkileyen sosyoekonomik değişkenleri belirleme konusundaki ilgilerini uyandırmıştır. Bunun altında yatan mantık, sosyoekonomik ve çevresel değişkenlerin, ölüm oranı üzerinde etkileşimli bir etki yaratmasının yanı sıra, bağımsız bir şekilde ortaya çıkmasıdır [5].

Kısmi En Küçük Kareler (KEKK) yöntemi ilk olarak, istatistik, matematik ve bilgisayar kullanılarak kimyasal verilerin işlenmesini kapsayan kimya alanında bir bilim dalı olan “Kemometri” alanında popüler olmuştur. Kemometri biliminde, KEKK yöntemi yeni bir kestirim tekniği olarak ortaya çıkmış ve Kısmi En Küçük Kareler Regresyonu (KEKKR) çok değişkenli ölçümler arasındaki doğrusal ilişkileri modellemek için yerleşmiş bir araç haline gelmiştir [6]. Kısmi En Küçük Kareler Diskriminant Analizi (KEKKDA) yönteminin temeli KEKKR’ye dayanır. KEKKR’den farkı bağımlı değişkenin kategorik yapıda olmasıdır. KEKK ve KEKKDA yöntemlerinin özellikle son yıllarda Kemometri dışında ekonomi ya da sosyal alanlardaki veri kümelerinde uygulamalarına da rastlanmaktadır. Örneğin; Kemalbay ve Korkmazoglu [7] çeşitli KEKK algoritmalarının ekonometri uygulamalarını ve yorumlamalarını Türkiye için gerçek bir veri kümesi üzerinde R programını kullanarak göstermiştir. Kemalbay ve Korkmazoglu [8], 1981-2010 yılları arasındaki tarihsel veri kümesi üzerinde KEKKR yöntemini uygulayarak Türkiye için yıllık elektrik tüketimini tahmin etmeye çalışmıştır. Ümit ve Bulut [9], 2005-2010 yılları arasındaki veri kümesini kullanarak 2008 küresel krizinden sonra Türkiye’de işsizliği etkileyen değişkenleri KEKKR yöntemiyle incelemiştir. Serrano-Cinca ve Gutiérrez-Nieto [10], 2008 ABD bankacılık krizinin tahmini için KEKKDA’yı kullanmıştır. Sawatsky vd. [11], KEKKR yönteminin sosyal bilimlerde kısıtlı sayıda kullanımdan dolayı örnek olması amacıyla, özellikle adli psikoloji ile ilgili hipotetik bir veri kümesi üzerinde uygulama yapmıştır. Sghaier vd. [12], 2005-2011 dönemine ait finansal bilgilere dayanarak Orta Doğu ve Kuzey Afrika (MENA) bölgesindeki geleneksel ve İslami bankaları ayırmak için KEKKDA yöntemini uygulamıştır. Yoon ve Klasen [13], OECD tarafından üretilen bileşik bir endeks olan Sosyal Kurumlar ve Cinsiyet Endeksi’ni (SIGI), Temel Bileşenler Analizi (TBA) ve KEKK tarafından üretilen ağırlıkları kullanılarak yeniden oluşturmuştur. Polat [14] ise, KEKKDA yönteminin 28 Avrupa Birliği (AB) üyesi ülke ile potansiyel adayları da içeren AB’ye 7 aday ülkeyi önceden bilinen sınıflarına doğru bir şekilde sınıflandırma performansını analiz etmiş ve sınıflandırmada etkili olan ekonomik ve/veya demografik göstergeleri belirlemiştir.

Bu çalışmanın amacı ise, farklı sosyoekonomik değişkenlerin Asya ülkelerinde doğumda yaşam beklentisi üzerindeki etkilerini tespit etmektir. Bu analizdeki verilerin kaynağı, Dünya Nüfus Veri Sayfası, 2015’dir. Doğumda beklenen yaşam ve bu analizde kullanılan diğer tüm dokuz değişkene ilişkin veriler, Dünya Nüfus Veri Sayfasında, 2015’de 51 Asya ülkesi için mevcuttur. 2015 yılında Asya’daki tüm ülkeler için beklenen ortalama yaşam süresi yani Asya ortalaması 72 olduğu için, KEKKDA kullanılarak Asya ülkeleri beklenen yaşam sürelerine göre <72 ve ≥ 72 şeklinde iki gruba sınıflandırılmıştır. Analiz

sonucunda ülkelerin beklenen yaşam sürelerinin belirlenmesinde etkili olan en önemli sosyoekonomik değişken belirlenmiştir.

2. Kısmi En Küçük Kareler Regresyon

KEKKR yönteminde, bağımlı değişken/değişkenler (Y) ile bağımsız değişkenler (X) arasındaki kovaryansı en büyük yapacak şekilde p sayıda bağımsız değişkenden daha az sayıda h tane bileşen elde edilir. Yöntem birden fazla bağımlı değişken durumunda da kullanılabilir. Bir boyut indirgeme aracı da olan KEKKR yöntemi ile elde edilen bileşenler/gizli değişkenler yeni bağımsız değişkenler olarak kullanılır ve bağımlı değişken/değişkenler bu bileşenleri kullanarak modellenir. KEKKR yöntemi değişken sayısı çok fazla olduğunda hatta gözlem sayısından çok fazla sayıda bağımsız değişken olduğunda bile ve bağımsız değişkenler arasında ilişki olduğu yani çoklu bağlantı durumunda bile kullanılabilir ve iyi sonuçlar verir. Yöntemin avantajı, normallik ve bağımsız değişkenlerin ilişkisiz olması gibi herhangi varsayım gerektirmemesidir. Ayrıca bu yöntem, kayıp veri durumunda da kullanılır.

Temel Bileşenler Regresyonu (TBR) ve KEKKR gibi çok değişkenli regresyon yöntemleri, birçok alanda çok geniş bir üne sahiptir. Bu yöntemler özellikle kimya alanında sıklıkla karşılaşılan çok fazla ilişkili bağımsız değişkenin ve nispeten daha az gözlemin olduğu durumlarda kullanılmaktadır. p, bağımsız değişken sayısı ve n, bağımlı değişken sayısı olmak üzere; $p \ll n$. Uygulamada TBR ve KEKKR'nin kullanımında büyük farklılıklar yoktur ve birçok durumda yöntemler benzer kestirim doğrulukları sağlamaktadır. Modeldeki aynı sayıda bileşen için KEKKR Y'deki değişimin daha fazlasını açıklarken, TBR ise X'deki değişimin daha fazlasını açıklar [15]. KEKKR'de özellikle, X ve Y'nin eş zamanlı bir ayrışımını gerçekleştiren bir küme bileşen bulunmaya çalışılmaktadır. Ancak bu bileşenler, X ve Y arasındaki kovaryansı mümkün olduğunca çok açıklama ölçütü altında elde edilir.

KEKKR model parametrelerini hesaplamak için literatürde geliştirilmiş çeşitli algoritmalar mevcuttur. Ancak, en popüler olanı ve klasik algoritma olarak da bilineni “Doğrusal Olmayan Yinelemeli En Küçük Kareler (NIPALS/Non-linear Iterative Partial Least Squares)” algoritmasıdır. Günümüzde de yaygın olarak birçok bilgisayar programlarında KEKK yöntemi içeren analizlerde, NIPALS algoritması kullanılmaktadır.

2.1. KEKKR için İdeal Bileşen Sayısı

İdeal bileşen sayısını belirlemek KEKKR'de çok önemli bir problemdir. İdeal bileşen sayısından daha az bileşen seçmek bilgi kaybına yol açarken, ideal bileşen sayısından daha fazla bileşen seçmek ise zayıf kestirim yeteneğine sahip modellerin elde edilmesine yol açar [16]. Çok fazla ve ilişkili X değişkenleri olduğunda “aşırı uyum” için yani, çok az ya da hiç kestirim gücüne sahip olmayan ancak veriye iyi uyan bir model etme konusunda önemli bir risk doğar. Bu nedenle, her bir KEKK bileşenin kestirim önemini test etmek gerekir [15]. KEKKR'de modeldeki bileşenlerin sayısının belirlenmesi genellikle çapraz geçerlilik (ÇG) içeren ölçütlere dayanmaktadır. ÇG yöntemi, KEKK yöntemi için modelin kestirim yeteneğini en iyi yapan ideal bileşen sayısını seçmede kullanılır [14]. ÇG yönteminde, çalışma kümesinde (training set) yer alan gözlemlerden bazıları bir geçerlilik kümesi (validation set) olarak ayrılır ve geriye kalan gözlemler ile model oluşturulur. Farklı sayıda bileşenler kullanılarak, ayrılan gözlemler için kestirim hataları hesaplanır. Bütün gözlemler için kestirimler elde edilinceye kadar süreç tekrarlanır. En küçük kestirim hata kareler toplamını veren modelin bileşen sayısı modelde kalacak ideal bileşen sayısı olarak seçilir.

Temel olarak, ÇG, veriyi bir dizi gruba bölerek, örneğin k tane gruba, genellikle beş ile dokuz arasında değişen sayıda gruba bölerek ve daha sonra silinen gruplardan biri ile azaltılmış verilerden bir dizi paralel model geliştirilerek gerçekleştirilir. $k=n$ olduğunda yani grup sayısı gözlem sayısına eşit olduğunda yöntem birini dışarıda bırakma çapraz geçerlilik (BDBÇG) yöntemi adını almaktadır. Bir model oluşturulduktan sonra silinen veri için, gerçek ve kestirilen Y değerleri arasındaki farklar hesaplanır. Bu farklılıkların karelerinin toplamı tüm paralel modellerden hesaplanıp toplanarak, PRESS değeri (kestirim artık kareler toplamı) elde edilir [15].

Literatürde modelde kalacak ideal bileşen sayısının belirlenmesinde çeşitli ölçütler kullanılmaktadır. Bu ölçütlerden biri Eş. (1)'deki Q_{cum}^2 indeksidir. Bu çalışmada bu ölçüt kullanıldığı için, bu bölümde detaylı olarak incelenecektir. Bu indeks ilk h bileşenin modelin kestirim kalitesine katkısını ölçer. Q_{cum}^2 indeks hesabında yer alan ve Eş. (2)'de gösterilen PRESS ise modelin geçerlilik ve kestirimdeki başarısını ölçen bir ölçüttür [17].

$$Q_{cum(h)}^2 = 1 - \sum_{j=1}^h \frac{PRESS_j}{RESS_{j-1}} \quad (1)$$

$$PRESS_j = \sum_{i=1}^n (y_{j,i} - \hat{y}_{j,-i})^2 \quad (2)$$

PRESS istatistiği ise, h bileşen içeren model için kestirim kareler toplamıdır ve Eş. (2)'de gösterildiği üzere ÇG yöntemini kullanarak hesaplanır. Burada $\hat{y}_{j,-i}$; regresyon parametreleri tahmin edilirken i. gözlemin modelden çıkarıldığı j bileşenli modelden elde edilen i. gözlem için kestirilen y değerini gösterir [17]. PRESS değeri ne kadar küçük ise, modelin kestirim yeteneğinin de daha iyi olacağı kabul edilir. PRESS değerini, BDBÇG yöntemini kullanarak tahmin etmek çok yaygındır. BDBÇG, her seferde bir gözlemi model oluşumundan çıkararak ve bir kez kestirerek çapraz geçerliği yerine getirmektedir. RESS ise h-1 tane bileşen içeren modelin artık kareler toplamıdır. j=1 için

$$RESS_0 = \sum_{i=1}^n (y_i - \bar{y})^2 = n - 1 \text{ olur [6, 17, 18, 19].}$$

Q^2 indeksi gözlemin hangi sınıfa ait olduğunu gösteren bilinen y ve kestirilen kategorik değişken \hat{y} arasındaki hatanın değerlendirilmesine dayanan bir ölçüttür. Q^2 'nin birikimli değeri Q_{cum}^2 , belirli sayıdaki bileşenli modelin global uyum iyiliğinin ve kestirimdeki yeteneğinin bir göstergesidir. Q_{cum}^2 indeksinin en büyüğünü bulmaya yönelik çaba, en kararlı modeli bulmaya eşdeğerdir. En büyük Q_{cum}^2 indeksini veren “minimum bileşen sayısı”, yani ilave bileşenin daha fazla katkı sağlamadığı durum, ideal bileşen sayısını verir. Modelde kalacak ideal bileşen sayısına karar verirken, Q_{cum}^2 indeksinin yanı sıra R^2X_{cum} ve R^2Y_{cum} indeksleri de dikkate alınır. R^2X_{cum} indeksi bağımsız değişkenler ve ilk h bileşen arasındaki belirtme katsayılarının toplamına eşittir. Bu nedenle, modelin bağımsız değişkenlerine ilişkin ilk h bileşenin açıklama gücünün bir ölçütüdür. R^2Y_{cum} indeksi bağımlı değişkenler ve ilk h bileşen arasındaki belirtme katsayılarının toplamına eşittir. Bu nedenle, modelin bağımlı değişkenlerine ilişkin ilk h bileşenin açıklama gücünün bir ölçütüdür. [6, 20-22]. Bu iki indekse ilişkin formüller Eş. (3) ve Eş. (4)'de verilmiştir. Eş. (3)'deki t_i $n \times i$ boyutlu bileşen vektörü, p_i ise ilgili bileşen vektörüne ilişkin yük vektörüdür. Eş. (4)'deki $\hat{y}_{h,i}$, h bileşenli model için y_i gözleminin kestirim değeridir [20].

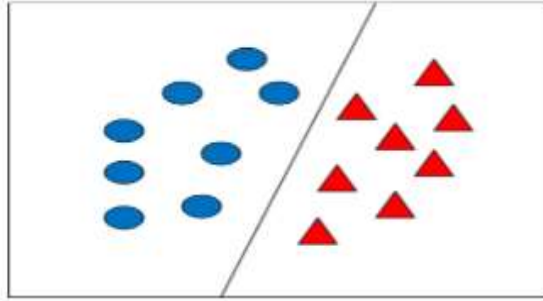
$$RX_h^2 = \frac{\sum_{i=1}^h \|t_i\|^2 \|p_i\|^2}{(n-1)p} \quad (3)$$

$$RY_h^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_{h,i})}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

3. Kısmi En Küçük Kareler Diskriminant Analizi

KEKKDA bağımlı değişkeninin kategorik olduğu, istatistiksel birimlerin sınıf üyeliklerini açıklayan klasik bir KEKKR'den oluşur. Son yirmi yıl boyunca, KEKKDA yüksek boyutlu veri kümelerini çeşitli amaçlar için modellemede büyük başarı göstermiştir; gıda analizinde ürün doğrulaması, tıbbi tanıda hastalıkların sınıflandırılması ve adli bilimlerde kanıt analizi gibi pek çok alanda uygulamaları mevcuttur. Bahsedilen bilim dallarının yanı sıra kemometrik alanında popüler bir istatistiksel yöntemdir ve omik veri analizinde kullanılması önerilen çok değişkenli bir boyut indirgeme aracıdır. Kemometrik ve omik veri kümeleri büyük hacimli, çok sayıda değişken barındıran, gürültülü ve kayıp veri içeren veri kümeleri olarak tanımlanır. KEKKDA'nın algılanan avantajlarından biri, yüksek derecede ilişkili ve gürültülü verileri analiz etme kabiliyetine sahip olmasıdır. Ayrıca, KEKKDA modelinin kestirim doğruluğunun yanı sıra skorlar ve yük grafikleri gibi çeşitli faydalı istatistikler sağladığı bilinmektedir. KEKKDA'da TBA'nın aksine gözlemlerin hangi sınıfa ait olduğu önceden bilinir. KEKKDA sınıf etiketlerini göz önünde bulundurarak boyut indirgemeyi gerçekleştirmesi bakımından TBA'nın denetimli/güdümlü bir biçimi olarak düşünülebilir [22, 23, 24]. Barker ve Rayens [25] KEKKDA'yı klasik diskriminant analizi ile karşılaştırmış ve KEKKDA'nın klasik yöntemle karşı değişken seçimi ve hata azaltma gibi avantajları olduğunu vurgulamıştır.

KEKKDA, doğrusal iki sınıflı bir sınıflandırıcı olarak kabul edilebilir. Bu açıdan bakıldığında KEKKDA'nın bir amacı da, Şekil 1'de gösterildiği gibi uzayı iki bölgeye bölen düz bir çizgi bulmaktır. Sınıflandırmada kullanılan birçok algoritmanın amacı da bu diskriminatörü/ayırıcıyı/karar fonksiyonunu bulmaktır. Beklendiği üzere ikiden fazla değişken olduğunda çok boyutlu bir uzayda bir hiperdüzlem ile temsil edilir [26].



Şekil 1. İki değişkenle oluşturulmuş iki sınıflı doğrusal diskriminatör

KEKKDA, niceliksel bağımsız değişkenleri kullanarak KEKK algoritması ile gözlemlerin hangi sınıfa ait olduğunu belirler. NIPALS algoritmasına yalnızca bir bağımlı değişkenin bulunduğu durumda ($m=1$) 'PLS1' ve birkaç bağımlı değişkenin bulunduğu durumda 'PLS2' denir. KEKKDA, klasik diskriminant analizinin uygulanmadığı birçok durumda uygulanabilir. Örneğin, gözlem sayısı düşük ve bağımsız değişken sayısı yüksek olduğunda kullanılması uygundur. Kayıp gözlemlerin olduğu durumda eldeki veriye uygulanabilir, bağımsız değişkenler arasında çoklu bağlantının yüksek olduğu durumlarda tercih edilebilir [14, 27, 28].

KEKKDA'da bağımlı değişken kategori sayısı kadar model elde edilebilir. İlgilenilen özellik neyse onu gösteren Y matrisi ya da vektörü eğer iki kategori/sınıf var ise 0 ya da 1 olarak kodlanır ($G=2$). İki den fazla sınıf olduğunda 0 ve 1 kodlamaları yaparak ya da her bir sütununun bir sınıfı temsil ettiği $n \times G$ boyutlu bir matris oluşturularak PLS2 algoritmasını kullanarak kategori sayısı kadar model oluşturulabilir [23].

KEKKDA yöntemi KEKKR yönteminden elde edilir. KEKKDA’da Y bağımlı değişken matrisinin kesikli değerlere sahip olduğu varsayılır. Klasik Çok Değişkenli Regresyon yönteminde model Eş. (5)’deki gibidir. p , bağımsız değişken sayısı; m , bağımlı değişken sayısı ve n , gözlem sayısını gösterebilir. Burada X , $n \times p$ boyutlu bağımsız değişkenler matrisi, B ise $p \times m$ boyutlu regresyon katsayılar matrisi, F $n \times m$ boyutlu hata matrisi ve Y $n \times m$ boyutlu bağımlı değişken matrisidir. Bu yöntemde, en küçük kareler ile B katsayılar matrisi Eş. (6)’daki gibi elde edilir [15].

$$Y = XB + F \quad (5)$$

$$B = (X'X)^{-1} X'Y \quad (6)$$

Birçok durumda problem $X'X$ matrisinin tekil olmasıdır. Örneğin, veride çoklubağlantı sorunu olduğunda ya da bağımsız değişken sayısı gözlem sayısından fazla olduğunda, $p \gg n$ karşılaşılan bir durumdur. KEKKR ve KEKKDA’nın her ikisi de X veri matrisini h tane dik skor T ’ye ($n \times h$) ve yükler matrisi P ’ye ($p \times h$) ve bağımlı değişken matrisi Y ’yi h tane dik skor T ’ye ($n \times h$) ve yükler matrisi Q ’ya ($m \times h$) ayırır. E ve F ise, sırasıyla X veri matrisi ve Y bağımlı değişken matrisine ilişkin $n \times p$ ve $n \times m$ boyutlu hata matrisleridir. Buna göre, KEKKDA’da Eş. (7) ve Eş. (8)’de gösterildiği üzere iki temel model vardır [15, 29-31]:

$$X = TP' + E \quad (7)$$

$$Y = TQ' + F \quad (8)$$

$p \times h$ boyutlu ağırlık matrisi W tanımlanırsa, skor/bileşen/gizli değişken matrisi Eş.(9)’daki gibi yazılır. Bu eşitlik KEKKDA modelinde yerine koyulduğunda Eş. (10) elde edilir. Bu denklemdeki $p \times m$ boyutlu KEKKDA katsayılar matrisi B ise Eş. (11)’deki gibi elde edilebilir [15, 29, 31]:

$$T = XW(P'W)^{-1} \quad (9)$$

$$Y = XW(P'W)^{-1} Q' + F \quad (10)$$

$$\hat{B} = W(P'W)^{-1} Q' \quad (11)$$

Buna göre Y ’nin bilinmeyen örneklem değeri $\hat{Y} = X\hat{B}$ ile yani $\hat{Y} = XW(P'W)^{-1} Q'$ şeklinde kestirilebilir [15].

KEKK algoritmasının ayırt edici sınıflandırma problemleriyle başa çıkabilmesi için, gözlemin ait olduğu sınıfla ilgili bilginin Y bağımlı değişkeninde kodlanması gerekir. Daha sonra, kestirim modelini elde etmek için X matrisinden Y değişkeni kestirilir [32]. Bunun için Y , gözlem sayısı kadar satıra (n) ve sınıf sayısı kadar sütuna (G) sahip olan bir “kukla” ikili matris olarak tanımlanır. Y ’deki her satır; karşılık gelen gözlemin ait olduğu sınıf bilgisini kodlayan bir vektördür, ancak her sütun belirli bir sınıfla ilişkilidir. Örneğin sınıf sayısı G olarak düşünüldüğünde ilk sütun 1. sınıfla, ikinci sütun 2. sınıfla ilişkilidir ve bu şekilde G . sınıfa kadar devam eder. Bu nedenle, belirli bir gözleme karşılık gelen satır vektörü, ait olduğu sınıfa ait sütun da 1 diğer tüm sütunlarda 0 değerini içerecektir. Örneğin, üç kategorili bir durumda, ikinci sınıfa ait bir gözlem $y_i = [0 \ 1 \ 0]$ vektörü ile temsil edilir. Daha sonra, X veri matrisi ve yapay Y arasında bir KEKKR modeli hesaplanır ve elde edilen regresyon katsayılar matrisi yeni gözlemler için bağımlı değişken değerini kestirmek için kullanılır. Bağımlı değişken kategorik bilgilerle ilişkilendirildiğinden,

gözlemlerin sınıflandırılması kestirilen bağımlı değişken değerlerine (\hat{Y}_{yeni}) dayanır. Ancak bu y değerleri ikili değil, gerçek değerlidir. Literatürde bu \hat{Y}_{yeni} değerlerine dayalı olarak KEKKDA'da nasıl sınıflandırılma yapılacağını belirlemek için farklı yaklaşımlar önerilmiştir.

KEKKDA'da gözlemleri ait oldukları sınıfa atamak için kullanılan yöntemlerden biri hesaplanan diskriminant fonksiyonlarını kullanarak atama yapmaktır. Bu çalışmada kullanılan XLSTAT programı Microsoft Excel kullanarak analizleri yapan ve sonuçları veren bir programdır. Bu program, bu yöntemi kullanarak gözlemleri ait oldukları sınıflara atamaktadır. Bu diskriminant fonksiyonlar doğrusaldır. Gözlem, en yüksek değerli sınıflandırma fonksiyonunu içeren kategoriye atanır. G, Y bağımlı değişkeninin kategori sayısı olsun. Her bir kategori için a_1, \dots, a_G modelin denklemi Eş. (12)'de olduğu gibi elde edilir [14, 28].

$$F(y_i, a_G) = b_0 + \sum_{j=1}^p b_j x_{ij} \quad (12)$$

a_G bağımlı nitel değişkenin kategorisidir. b_0, a_G ile ilişkili modelin kesim noktası, p bağımsız değişkenlerin sayısını ve b_i aynı modelin katsayılarıdır. Eğer, $G^* = \arg \max_G F(y_i, a_G)$ sağlanır ise, i. gözlem, G. sınıfına atanır [14, 28].

4. Bulgular ve Tartışma

Bu uygulamada, Asya ülkelerindeki beklenen yaşam süreleri verisi kullanılarak sosyoekonomik değişkenlerden hangisi ya da hangilerinin ülkelerin beklenen yaşam sürelerinin belirlenmesinde etkin olduğu bulunmaya çalışılmıştır. Ayrıca KEKKDA'nın sınıflandırma performansının analiz edilmesi amaçlanmıştır. Veri kümesi 2015 yılı için hazırlanan Dünya Nüfusu Veri Sayfası'ndan elde edilmiştir. İncelenen yıl için Asya genelinde beklenen yaşam süresi 72 yıldır. Bu nedenle, çalışmada "<72 yaş BY" ve "≥72 yaş BY" iki kategori olarak elde edilmiştir. Çalışmada kullanılan ve beklenen yaşam süresini belirlemede etkili olduğu düşünülen değişkenler [1] ile [5] çalışmalarından yola çıkılarak belirlenmiştir. Bu değişkenler aşağıda açıklanmıştır.

Değişkenler:

BY: Beklenen Yaşam Süresi (Yaş)

KY: Kentsel Yüzde (%)

MD: Modern Doğum Kontrolü Kullanan Evli Kadınlar (%)

KBG: Kişi Başına Gayri Safi Milli Gelir (\$)

NY: Nüfus Yoğunluğu (Tarıma Elverişli Arazinin km²'si Başına Düşen Kişi Sayısı)

BÖO: Bebek Ölüm Oranı (1000 Canlı Doğumdaki Ölü Bebek Sayısı)

TDO: Toplam Doğurganlık Oranı (Ömrü Boyunca Bir Kadının Doğurduğu Ortalama Çocuk Sayısı)

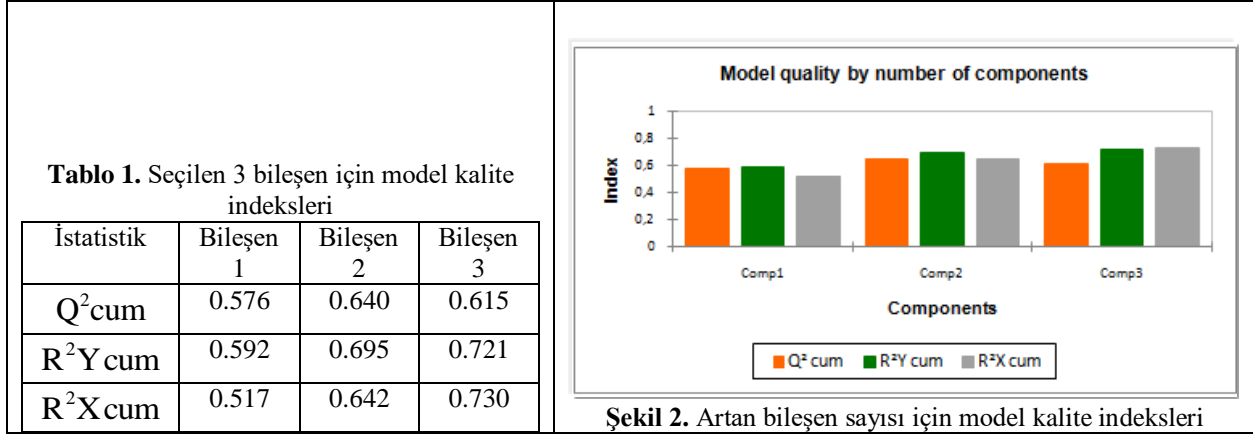
BN: Bağımlı Nüfus (15 yaşından küçük ve 65 yaşından büyük nüfus toplamı, %)

NBD: 1000 Nüfus Başına Doğumlar (Doğum Sayısı)

NBÖ: 1000 Nüfus Başına Ölümler (Ölüm Sayısı)

Veride kayıp gözlemler olduğu için KEKKDA yöntemi sınıflandırma için tercih edilir. Kayıp gözlemler, NIPALS algoritmasını kullanarak XLSTAT programında tahmin edilir. Daha sonra, yukarıda söz edilen üç indeks kullanılarak modelde kalacak ideal bileşen sayısı seçilir. Tablo 1'de kalite indeksleri verilmektedir. Burada kalite, bileşenlerin indekslere birikimli katkısına karşılık gelir. KEKK yöntemi otomatik olarak modelde kalacak bileşen sayısını üç olarak seçmiştir. Üç bileşenli model için Q²cum, R²Ycum ve R²Xcum değerleri sırasıyla 0.615, 0.721 ve 0.730 olarak elde edilir. Buna göre, KEKK kullanılarak elde edilen üç bileşenin hem X'leri hem de Y'yi iyi özetlediği söylenebilir.

Şekil 2’de gösterildiği üzere bir çubuk grafiği de elde edilmiştir. Bu grafikten, modeldeki bileşen sayısı arttıkça bu üç indeksin değişimi görülmektedir. Şekil 2 incelendiğinde, bileşen sayısı arttıkça R^2Y_{cum} ve R^2X_{cum} indekslerinin mutlaka arttığı ancak Q^2_{cum} indeksi için bu durumun geçerli olmadığı görülmektedir. Q^2_{cum} en iyi kestirim modelini elde etmek için kullanılır ve 1’e yakın olması istenir. Q^2_{cum} indeksi üç bileşenden sonra azalma eğiliminde olduğundan ve üç bileşenli model hem X’leri hem de Y’yi iyi özetlediğinden, model için en uygun bileşen sayısı 3 olarak seçilmiştir. Modeldeki ideal bileşen sayısı seçildikten sonra üç bileşenli model için sonuçlar yorumlanır.



Tablo 2'deki standartlaştırılmış katsayılar (beta katsayıları) tablosu, modeldeki değişkenlerin göreceli ağırlığının karşılaştırılmasını sağlar. KEKKR’de güven aralıklarını hesaplarken klasik en küçük kareler regresyonunda kullanılan normallik hipotezlerini temel alan klasik formüller geçerli değildir. Güven aralıklarını tahmin ederken bir bootsrap yöntemi kullanılır. Tablo 2’de verilen standartlaştırılmış katsayılar için güven aralığı 0 değerini içeriyorsa, modeldeki değişkeninin katsayısı anlamlı değildir. Bu sonuçlara göre, Bebek Ölüm Oranı’nın Asya ülkelerinin yaşam beklentilerini belirleyen en önemli değişken olduğu görülmektedir. Bu değişkenin modele olan katkısı istatistiksel olarak anlamlı olduğu için ve mutlak değerce en büyük katsayıya da sahip olduğundan, Bebek Ölüm Oranı modele katkısı en fazla olan ve tek değişkendir. Modeldeki diğer değişkenlerin, istatistiksel olarak modele anlamlı bir katkısı bulunamamıştır.

Tablo 2. Modelin standart katsayıları.

Değişken	Katsayı	Standart Hata	Alt Sınır (%95)	Üst Sınır (%95)
KY	-0.267	0.193	-0.655	0.121
MD	0.087	0.139	-0.192	0.367
KBG	0.050	0.110	-0.171	0.272
NY	-0.002	0.118	-0.238	0.235
BÖÖ	0.521	0.169	0.181	0.861
TDO	0.057	0.158	-0.261	0.375
BN	-0.012	0.138	-0.289	0.264
NBD	0.216	0.127	-0.038	0.470
NBÖ	-0.043	0.173	-0.389	0.304

Kısmi En Küçük Kareler Diskriminant Analizi Özel Sonuçları Tablo 3 ve Tablo 5’de görülür:

- **Sınıflandırma Fonksiyonları:** Çeşitli bağımsız değişkenlere ilişkin değerleri kullanarak, bu fonksiyon bir gözlemin hangi sınıfa atanacağını belirlemek için kullanılır. Bu fonksiyonlar

doğrusaldır. $F(y_i, a_G) = b_0 + \sum_{j=1}^p b_j x_{ij}$ denkleminde gösterildiği üzere, gözlem en yüksek sınıflandırma fonksiyon (F) değerinin olduğu sınıfına atanır.

- **Önsel ve Sonsal Sınıflandırma ve Skorlar:** Bu tablo, bağımlı değişken tarafından tanımlanan her bir gözlemin ait olduğu sınıfı gösterir. Gözlemlerin hangi sınıfa ait olduğu bağımlı değişkenin her bir kategorisi için hesaplanan üyelik olasılıkları ve sınıflandırma fonksiyon skorları ile belirlenir.
- **Tahmin Örnekleme için Hata Matrisi:** Hata matrisi, önsel ve sonsal sınıflandırmaların tümünden elde edilip, tüm doğru sınıflandırılan gözlemlerin yüzdesini gösterir. Hata matrisi, gözlemlerin yeniden sınıflandırılma özetini verir ve toplam gözlem sayısı üzerinden iyi sınıflandırılmış gözlemlerin yüzdesini görülmelerini sağlar.

KEKKDA'nın Asya ülkelerini beklenen yaşam süreleri bakımından ayırt etme konusundaki performansı, Tablo 3'de gösterildiği gibi hata matrisi ile değerlendirilir. Hata matrisi, bir sınıflandırma işleminin tahmini sonuçlarının özetidir. Her sınıfa ait doğru ve yanlış tahminlerin özetini sunar. Tablo 3'deki, % 96.08'lik doğru sınıflandırma yüzdesinden de görüldüğü üzere 51 Asya ülkesinden 49'u doğru bir şekilde sınıflandırılmıştır. Bu sonuç, Tablo 5'te verilen Kestirilen (BY) sütunu incelenerek de kolayca doğrulanabilir. Suriye ve Moğolistan hariç tüm ülkelerin önceden tanımlanmış sınıflarına (<72 yaş BY veya ≥72 yaş BY) doğru bir şekilde sınıflandırıldığı görülmektedir. Bu iki ülke de “<72 yaş BY” sınıfındayken analiz sonucunda “≥72 yaş BY” sınıfına atanmıştır. Ancak bu iki ülke için elde edilen $P(<72 \text{ yaş BY})$ ve $P(\geq 72 \text{ yaş BY})$ olasılıklarına baktığımızda çok büyük bir fark olmadığını da söyleyebiliriz. Görüldüğü üzere, KEKKDA çok yüksek bir doğru sınıflandırma yüzdesi vermiştir.

Tablo 3. KEKKDA'nın hata matrisi

den/ye karşı	<72 yaş BY	≥72 yaş BY	Toplam	% Doğruluk
<72 yaş BY	19	2	21	% 90.48
≥72 yaş BY	0	30	30	% 100.00
Toplam	19	32	51	%96.08

Bir sınıflandırma modelinin test edilmesi sonucu dört farklı çıktı elde edilir. Bu çıktılar [33, 34];

- True Positive (TP): Doğru pozitif tahminlerin sayısı.
- False Positive (FP): Yanlış pozitif tahminlerin sayısı.
- True Negative (TN): Doğru negatif tahminlerin sayısı.
- False Negative (FN): Yanlış negatif tahminlerin sayısı

Bu çıktılardan yola çıkılarak sınıflandırma sonuçları aşağıdaki gibi de sunulabilir. Bu sonuçlardan yararlanarak KEKKDA modelinin sınıflandırma performansını daha sağlıklı bir şekilde ölçmek için duyarlılık ve özgüllük gibi ölçütler hesaplanır.

	Tahmin Edilen Değer		
	<72 yaş BY (Pozitif)	≥72 yaş BY (Negatif)	
Gerçek Değer	<72 yaş BY (Pozitif)	TP=19	FN=2
	≥72 yaş BY (Negatif)	FP=0	TN=30

Toplam Doğruluk: Doğru tahminlerin sayısının, veri kümesindeki tüm gözlemlerin sayısına bölümünden elde edilir. $\text{Doğruluk} = \frac{TP + TN}{TP + FP + TN + FN} = \frac{49}{51} = 0.96$. Modelin doğruluğu, gözlemlerin ait olduğu sınıfa doğru atanma oranıdır. Bu ölçüte ilişkin sonuçlar Tablo 4'te tekrar sunulmuştur.

Duyarlılık: Doğru pozitif tahminlerin sayısının, veri kümesindeki tüm gerçek pozitif gözlemlerin sayısına bölümünden elde edilir. $Duyarlılık = \frac{TP}{TP + FN} = \frac{19}{21} = 0.90$ ile “<72 yaş BY” sınıfının duyarlılığı hesaplanmış olur. Bir sınıfın duyarlılığı, modelin gözlemi ait olduğu sınıfa doğru atama yeteneğini gösterir. Örneğin “<72 yaş BY” sınıfına ait gözlemlerden hiçbiri “≥72 yaş BY” sınıfına atanmamış olsaydı FN=0 olurdu ve “<72 yaş BY” sınıfı için duyarlılık 1’e eşit çıkardı. Ancak, çalışmamızda FN=2 olduğu için duyarlılık 1’den küçük çıkmıştır.

Özgüllük: Doğru negatif tahminlerin sayısının, veri kümesindeki tüm gerçek negatif gözlemlerin sayısına bölümünden elde edilir. $Özgüllük = \frac{TN}{TN + FP} = \frac{30}{30} = 1$ ile “<72 yaş BY” sınıfının özgüllüğü hesaplanmış olur. Özgüllük, geliştirilen modelin diğer tüm sınıfların gözlemlerini reddetme yeteneğini gösterir. Örneğin “≥72 yaş BY” sınıfına ait gözlemlerden hiçbiri “<72 yaş BY” sınıfına atanmadığı için FP=0 olur ve “<72 yaş BY” sınıfı için özgüllük 1’e eşit çıkar.

Sadece iki sınıflı model durumunda sınıfların duyarlılığı ve özgüllüğü simetriktir. Örneğin, bir sınıfın duyarlılığı diğer sınıfın özgüllüğüne eşittir ve bu ifadenin tam tersi de geçerlidir. Sınıf doğruluk, duyarlılık ve özgüllük değerleri 0 ile 1 arasında değer alır ve herhangi bir sınıfa atanamayan gözlemler hesaplamalara dahil edilmemektedir [33, 34].

Tablo 4’te KEKKDA modeline ilişkin sınıflandırma performansını değerlendirmek için ilgili ölçütler sunulmuştur.

Tablo 4. Sınıflandırma performansı değerlendirme ölçütleri

	Duyarlılık	Özgüllük	Doğruluk
<72 yaş BY	0.90	1	0.90
≥72 yaş BY	1	0.90	1

Duyarlılık ve özgüllük doğru sınıflandırma değerlerini gösterdiklerinden, sınıflandırma analizi için önemli istatistikler olup bu değerlerin olabildiğince 1’e yakın olması istenmektedir. Tablo 4’e bakıldığında bu değerlerin yüksek olduğu görülmektedir. Bu durum, kurulan KEKKDA modelinin sınıflandırmadaki başarısını göstermektedir.

Tablo 5. Önsel ve sonsal sınıflandırma ve skorları (değişken BY)

Ülkeler	BY	Kestirilen (BY)	F(<72yaşBY)	F(≥72yaşBY)	P(<72yaşBY)	P(≥72yaşBY)
Ermenistan	≥72 yaş BY	≥72 yaş BY	0.019	0.981	0.277	0.723
Azerbaycan	≥72 yaş BY	≥72 yaş BY	0.184	0.816	0.347	0.653
Bahreyn	≥72 yaş BY	≥72 yaş BY	-0.049	1.049	0.250	0.750
Kıbrıs	≥72 yaş BY	≥72 yaş BY	0.004	0.996	0.271	0.729
Gürcistan	≥72 yaş BY	≥72 yaş BY	0.077	0.923	0.300	0.700
Irak	<72 yaş BY	<72 yaş BY	0.735	0.265	0.615	0.385
İsrail	≥72 yaş BY	≥72 yaş BY	0.041	0.959	0.285	0.715
Ürdün	≥72 yaş BY	≥72 yaş BY	0.344	0.656	0.423	0.577
Kuveyt	≥72 yaş BY	≥72 yaş BY	0.055	0.945	0.291	0.709
Lübnan	≥72 yaş BY	≥72 yaş BY	-0.038	1.038	0.254	0.746
Umman	≥72 yaş BY	≥72 yaş BY	0.170	0.830	0.341	0.659
Filistin Toprakları	≥72 yaş BY	≥72 yaş BY	0.448	0.552	0.474	0.526
Katar	≥72 yaş BY	≥72 yaş BY	-0.007	1.007	0.266	0.734
Suudi Arabistan	≥72 yaş BY	≥72 yaş BY	0.269	0.731	0.386	0.614
Suriye	<72 yaş BY	≥72 yaş BY	0.382	0.618	0.441	0.559
Türkiye	≥72 yaş BY	≥72 yaş BY	0.131	0.869	0.323	0.677
Birleşik Arap Emirlikleri	≥72 yaş BY	≥72 yaş BY	0.011	0.989	0.273	0.727

Yemen	<72 yaş BY	<72 yaş BY	0.998	0.002	0.730	0.270
Kazakistan	<72 yaş BY	<72 yaş BY	0.566	0.434	0.533	0.467
Kırgızistan	<72 yaş BY	<72 yaş BY	0.676	0.324	0.587	0.413
Tacikistan	<72 yaş BY	<72 yaş BY	0.974	0.026	0.721	0.279
Türkmenistan	<72 yaş BY	<72 yaş BY	0.770	0.230	0.632	0.368
Özbekistan	<72 yaş BY	<72 yaş BY	0.821	0.179	0.655	0.345
Afganistan	<72 yaş BY	<72 yaş BY	1.454	-0.454	0.871	0.129
Bangladeş	<72 yaş BY	<72 yaş BY	0.813	0.187	0.652	0.348
Bhutan	<72 yaş BY	<72 yaş BY	0.844	0.156	0.665	0.335
Hindistan	<72 yaş BY	<72 yaş BY	0.813	0.187	0.652	0.348
İran	≥72 yaş BY	≥72 yaş BY	0.266	0.734	0.385	0.615
Maldivler	≥72 yaş BY	≥72 yaş BY	0.314	0.686	0.408	0.592
Nepal	<72 yaş BY	<72 yaş BY	0.777	0.223	0.635	0.365
Pakistan	<72 yaş BY	<72 yaş BY	1.261	-0.261	0.821	0.179
Sri Lanka	≥72 yaş BY	≥72 yaş BY	0.429	0.571	0.465	0.535
Brunei	≥72 yaş BY	≥72 yaş BY	0.082	0.918	0.302	0.698
Kamboçya	<72 yaş BY	<72 yaş BY	0.725	0.275	0.611	0.389
Endonezya	<72 yaş BY	<72 yaş BY	0.598	0.402	0.549	0.451
Laos	<72 yaş BY	<72 yaş BY	1.231	-0.231	0.812	0.188
Malezya	≥72 yaş BY	≥72 yaş BY	0.056	0.944	0.292	0.708
Myanmar	<72 yaş BY	<72 yaş BY	1.009	-0.009	0.735	0.265
Filipinler	<72 yaş BY	<72 yaş BY	0.528	0.472	0.514	0.486
Singapur	≥72 yaş BY	≥72 yaş BY	-0.168	1.168	0.208	0.792
Tayland	≥72 yaş BY	≥72 yaş BY	0.229	0.771	0.368	0.632
Timor Leste	<72 yaş BY	<72 yaş BY	1.086	-0.086	0.764	0.236
Vietnam	≥72 yaş BY	≥72 yaş BY	0.430	0.570	0.465	0.535
Çin	≥72 yaş BY	≥72 yaş BY	0.241	0.759	0.373	0.627
Çin, Hong Kong Özel Yönetim Bölgesi	≥72 yaş BY	≥72 yaş BY	-0.169	1.169	0.208	0.792
Çin, Makao. Özel Yönetim Bölgesi	≥72 yaş BY	≥72 yaş BY	0.056	0.944	0.292	0.708
Japonya	≥72 yaş BY	≥72 yaş BY	-0.273	1.273	0.176	0.824
Kuzey Kore	≥72 yaş BY	≥72 yaş BY	0.364	0.636	0.432	0.568
Güney Kore	≥72 yaş BY	≥72 yaş BY	-0.076	1.076	0.240	0.760
Moğolistan	<72 yaş BY	≥72 yaş BY	0.491	0.509	0.496	0.504
Tayvan	≥72 yaş BY	≥72 yaş BY	0.007	0.993	0.272	0.728

5. Sonuçlar

Bu çalışmada, klasik diskriminant analizinin uygulanması için gereken varsayımları gerektirmeyen ve kimya biliminde yapılan uygulamalarda yaygın bir şekilde kullanılan ancak son yıllarda ekonomi ve sosyal bilim alanlarında da tercih edilmeye başlayan KEKKDA yöntemi ile Asya ülkelerine ilişkin veriler kullanılarak bir sınıflandırma çalışması yapılmıştır. KEKKDA'nın sonucu olarak, 3 bileşenli model ile 51 Asya ülkesinin % 96.08'nin yani 49 tanesinin beklenen yaşam süreleri bakımından doğru sınıflandırıldığı görülmüştür. Sadece Suriye ve Moğolistan ülkeleri ait oldukları sınıfa düşmemiştir. Yöntem yüksek bir sınıflandırma başarısı göstermiştir.

Asya ülkelerinin beklenen yaşam sürelerine göre sınıflandırılmasında tek ve en önemli/etkili değişkenin ise Bebek Ölüm Oranı olduğu görülmüştür. Buna göre, Suriye'de son birkaç yıldır süren savaşın ve göçün etkisiyle ülkede bu değişken için sağlıklı verilerin toplanmasındaki engellerin varlığı ve Moğolistan'da ise son yıllarda bebek ölüm oranlarında yaşanan düşüşlerin, bu iki ülkenin beklenen yaşam süresi bakımından yüksek olan gruba düşmesine yol açtığını düşündürmektedir. Asya ülkelerindeki

ortalama beklenen yaşam süresini uzatmak için politikacıların özellikle bebek ölüm oranlarını düşürmeye yönelik politikalar geliştirmeleri, yatırımları sağlık alanına yönlendirmeleri önemlidir.

Yazarların Katkısı

Yazarların makaleye katkısı eşittir.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada, araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- [1] Ali M., Ali M. 2015. Discriminant Analysis of Socioeconomic Factors of Life Expectancy at Birth in Asia. *Science International (Lahore)*, 27 (5): 3971-3975.
- [2] Dowd K., Blake D., Cairns A.J.G. 2010. Facing up to Uncertain Life Expectancy: The Longevity Fan Charts, *Demography*, 47(1): 67-78.
- [3] Wagstaff A. 2000. Socioeconomic Inequalities in Child Mortality: Comparisons Across Nine Developing Countries. *Bulletin of the World Health Organization*, 78: 19-29.
- [4] Sastry N. 2004. Trends in Socioeconomic Inequalities in Mortality in Developing Countries: The Case of Child Survival in Sao Paulo, Brazil, *Demography*, 41(3): 443-464.
- [5] Sufian A.J.M. 2013. Life Expectancy and Its Socioeconomic Determinants-A Discriminant Analysis of National Level Data. *International Journal of Humanities and Social Science, Special Issue 3 (12)*: 303-312.
- [6] Polat E. 2009. Kısmi En Küçük Kareler Regresyonu. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, 170s, Ankara.
- [7] Kemalbay G., Korkmazoglu Ö.B. 2012. Econometrics Application of Partial Least Squares Regression: An Endogeneous Growth Model for Turkey. *Procedia-Social and Behavioral Sciences*, 62: 906-910.
- [8] Kemalbay G., Korkmazoglu Ö.B. 2012. Effects of Multicollinearity on Electricity Consumption Forecasting using Partial Least Squares Regression. *Procedia-Social and Behavioral Sciences*, 62: 1150-1154.
- [9] Ümit A.Ö., Bulut E. 2013. Türkiye’de İşsizliği Etkileyen Faktörlerin Kısmi En Küçük Kareler Regresyon Yöntemi ile Analizi: 2005-2010 Dönemi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 37: 131-142.
- [10] Serrano-Cinca C., Gutiérrez-Nieto B. 2013. Partial Least Square Discriminant Analysis for Bankruptcy Prediction. *Decision Support Systems*, 54: 1245-1255.
- [11] Sawatsky M.L., Clyde M., Mee F. 2015. Partial Least Squares Regression in the Social Sciences. *The Quantitative Methods for Psychology*, 11 (2): 52-62.
- [12] Sghaier A., Jabeur S.B., Bannour B. 2018. Using Partial Least Square Discriminant Analysis to Distinguish Between Islamic and Conventional Banks in the MENA Region. *Rev Financ Econ.*, 36: 133-148.
- [13] Yoon J., Klasen S. 2018. An Application of Partial Least Squares to the Construction of the Social Institutions and Gender Index (SIGI) and the Corruption Perception Index (CPI). *Soc Indic Res*, 138: 61-88.
- [14] Polat E. 2018. Determination of the Effective Economic and/or Demographic Indicators in Classification of European Union Member and Candidate Countries Using Partial Least Squares Discriminant Analysis. *Journal of Data Science*, 16 (1): 79-92.

- [15] Fordellone M., Bellincontro A., Mencarelli F. 2018. Partial Least Squares Discriminant Analysis: A Dimensionality Reduction Method to Classify Hyperspectral Data. <https://arxiv.org/pdf/1806.09347.pdf> (Erişim Tarihi: 07.10.2019)
- [16] Wiklund S., Nilsson D., Eriksson L., Sjöström M., Wold S., Faber K. 2007. A Randomization Test for PLS Component Selection. *Journal of Chemometrics*, 21: 427-439.
- [17] Pérez-Enciso M., Tenenhaus M. 2003. Prediction of Clinical Outcome with Microarray Data: A Partial Least Squares Discriminant Analysis (PLS-DA) Approach. *Human Genetics*, 112: 581-592.
- [18] Polat E., Gunay S. 2009. Kısmi En Küçük Kareler ve Bir Uygulama. VI. İstatistik Günleri Sempozyumu Bildiriler Kitabı, 437-444.
- [19] Tenenhaus M. 1998. *La Régression PLS Théorie et Pratique*. Editions Technip. 27, Rue Ginoux 75737, Paris Cedex 1.
- [20] Goyal M.K., Ojha C.S.P. 2010. Application of PLS-Regression as Downscaling Tool for Pichola Lake Basin in India. *International Journal of Geosciences*, 1: 51-57.
- [21] Ibrahim M.A.M. 2009. Comparison Between Different Procedures to Determine the Relative Importance of the Lifetime Performance Traits in Predicting Breeding Values of Holstein Cows. *Egyptian Journal of Animal Production*, 46 (2): 93-102.
- [22] Rohman A., Lumakso F.A., Riyanto S. 2016. Use of Partial Least Square Discriminant Analysis Combined with Mid Infrared Spectroscopy for Avocado Oil Authentication. *Research Journal of Medicinal Plants*, 10 (2): 175-180.
- [23] Almeida M.R., Correa D.N., Rocha W.F.C., Scafi F.J.O. 2013. Discrimination Between Authentic and Counterfeit Banknotes Using Ramanspectroscopy and PLS-DA with Uncertainty Estimation. *Microchemical Journal*, 109: 170-177.
- [24] Ruiz-Perez D., Narasimha G. 2018. So you think you can PLS-DA? IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 18-20 October, Las Vegas, NV, USA. <https://www.biorxiv.org/content/biorxiv/early/2018/01/15/207225.full.pdf> (Erişim Tarihi: 07.10.2019)
- [25] Barker M., Rayens W.S. 2003. Partial Least Squares for Discrimination. *Journal of Chemometrics*, 17: 166-173.
- [26] Brereton R.G., Lloyd G.R. 2014. Partial Least Squares Discriminant Analysis: Taking the Magic Away. *Journal of Chemometrics*, 28: 213-225.
- [27] Partial Least Squares Discriminant Analysis PLSDA Tutorial. 2018. https://help.xlstat.com/customer/en/portal/articles/2062368-partial-least-squares-discriminant-analysis-plsda-tutorial?b_id=9283 (Erişim Tarihi: 18.09.2019)
- [28] Xlstat, 2018. Paris, France. <https://help.xlstat.com/customer/en/portal/articles/2178395-download-the-xlstat-help-documentation> (Erişim Tarihi: 18.09.2019)
- [29] Polat E., Gunay S. 2015. The Comparison of Partial Least Squares Regression, Principal Component Regression and Ridge Regression with Multiple Linear Regression for Predicting PM10 Concentration Level Based on Meteorological Parameters. *Journal of Data Science*, 13 (2): 663-692.
- [30] Rosipal R., Krämer N. 2006. Overview and Recent Advances in Partial Least Squares, in Subspace, Latent Structure and Feature Selection. Edited by Saunders C., Grobelnik M., Gunn S. & Taylor J.S., Springer: Berlin, 34-51.
- [31] Wold S., Sjöström M., Eriksson L. 2001. PLS-Regression: A Basic Tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58: 109-130.
- [32] Sjöström M., Wold S., Söderström B. 1986. PLS Discriminant Plots, in *Pattern Recognition in Practice*. Ed: Gelsema E.S., Kanal L.N. (Amsterdam: Elsevier), 461-470.
- [33] Ballabio D., Consonni V. 2013. Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA, *Analytical Methods*, 5: 3790-3798.
- [34] Aliakbarzadeh G., Parastar H., Sereshti H. 2016. Classification of Gas Chromatographic Fingerprints of Saffron Using Partial Least Squares Discriminant Analysis Together with Different Variable Selection Methods. *Chemometrics and Intelligent Laboratory Systems*, 158: 165-173.