

# İLLERİN GELİŞMİŞLİK DÜZEYLERİNİ SIRALAMA VE SINIFLAMA BAKIMINDAN VERİ ZARFLAMA ANALİZİ VE ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLERİN KARŞILAŞTIRILMASI ÜZERİNE BİR ÇALIŞMA

H. Hasan ÖRKÜ\*  
Filiz KARDİYEN\*\*

## Öz

Veri Zarflama Analizi (VZA) karar verme birimlerini sıralamada ve sınıflamada geniş kullanıma sahip, oldukça popüler bir analiz olarak karşımıza çıkmaktadır. Çok sayıda girdi ve çıktı değişkeni olması durumunda VZA, çok değişkenli istatistiksel yöntemlere alternatif, parametrik olmayan bir yöntemdir. Bu çalışmada, Türkiye'nin 81 ilinin, gelişmişlik düzeylerini temsil eden 14 sosyo-ekonomik ve demografik değişken bakımından sınıflama ve sıralanmasında VZA ve çok değişkenli istatistiksel yöntemler karşılaştırmalı olarak incelenmiştir. İleri sınıflama amacıyla çok değişkenli istatistiksel yöntemlerden Diskriminant Analizi ve aşamalı olmayan Kümeleme Analizlerinden K-Ortalama Tekniği uygulanmış ve bu yöntemlerin VZA ile ne kadar uyumlu olduklarını tespit etmek için ise Kappa katsayısı kullanılmıştır. İllerin gelişmişlik düzeylerine göre sıralanmasında ise çok değişkenli istatistiksel yöntemlerden Temel Bileşenler Analizi (TBA) kullanılmış ve bu yöntemin VZA ile olan sıralama ilişkisi Spearman sıra korelasyon katsayısı ile test edilmiştir. Analizler sonucunda, hem sınıflama hem de sıralama amacı güdüldüğünde Veri Zarflama Analizi, çok değişkenli istatistiksel yöntemlerle iyi bir uyum sağlamıştır. Bu uyumun genellenebilir ve güvenilir nitelikte olması amacıyla yapılan simülasyon çalışması da benzer sonuçlar vermiştir. VZA çok değişkenli istatistiksel yöntemlerin aksine herhangi bir varsayım gerektirmemesi ve araştırmacının isteğine bağlı olarak modele yeni kısıtlar koyabilme esnekliğinin bulunması nedeniyle de araştırmacılara geniş kullanım imkanı sağlamaktadır. VZA'nın bu avantajlarına ve hem gerçek veri hem de simülasyon çalışması sonuçlarına dayanarak çok değişkenli istatistiksel yöntemlerin yerine kullanılacak bir yöntem olduğu sonucuna varılmıştır.

**Anahtar Sözcükler:** Veri zarflama analizi, çok değişkenli istatistiksel yöntemler, sınıflama, sıralama.

\* Arş. Gör., Gazi Üniversitesi, İstatistik Bölümü, Teknikokullar, ANKARA, hhorcku@gazi.edu.tr

\*\* Arş. Gör., Gazi Üniversitesi, İstatistik Bölümü, Teknikokullar, ANKARA, fkardiyen@hotmail.com

**Abstract****A Study on the Comparison of Data Envelopment Analysis and Multivariate Statistical Methods in Ranking and Classifying Provinces Development Level in Turkey**

Data Envelopment Analysis (DEA), a popular method, has been extensively used for ranking and classifying the decision making units. DEA, a nonparametric technique, is an alternative method to multivariate statistical methods when it is used for the data with multiple inputs and outputs. In this study, DEA's and multivariate statistical methods' performances are compared in ranking and classifying the 81 provinces of Turkey with respect to 14 social-economic and demographic variables. To classify the provinces, Discriminant Analysis, multivariate statistical method, and K-Means Cluster Analysis, nonhierarchical clustering method are used and for determining the correspondence of these methods and DEA Kappa statistics is used. To rank the provinces with respect to their development level, Principal Component Analysis and DEA are used and these methods' ranking relationship is tested by Spearman's rank correlation coefficient. After the applications, for both ranking and classifying. DEA gives similar results with multivariate statistical methods. To make these correspondence, a general and reliable simulation study is done and this study has given similar results. DEA provides researchers a wide usage opportunity since it does not need any assumptions, unlike the multivariate statistical methods and it has a flexibility to add new restrictions to model according to researchers need. Based on DEA's advantages and both real data and simulation study results, it is concluded that DEA can be used instead of multivariate statistical methods.

**Keywords:** Data envelopment analysis, multivariate statistical methods, classification, ranking.

**GİRİŞ**

Ülkelerin, illerin ilçelerin ya da benzer amaçlar için hizmet veren bankalar, üniversiteler ve ticaret firmaları gibi organizasyonel birimlerin sınıflaması ve sıralanması her zaman güncelliğini koruyan bir konudur. Ülkeler gelişmişlik düzeyleri göz önüne alınarak birinci, ikinci, üçüncü dünya ülkeleri olarak sınıflandırılmakta ve sıralanmakta, bu ölçütler bakımından iyi olan ülkeler buldukları coğrafi konumda ve dünyada söz sahibi olabilmektedir.

Ülkemizde illerin ve bölgelerin sınıflaması ve sıralanması çeşitli amaçlarla yapıla gelmektedir. Cumhuriyetimizin kurulmasından bu yana devlet kalkınma planları, gelişmekte olan iller ve bölgeler, kalkınmada öncelik verilecek iller ve bölgeler adı altında devlet çalışmaları mevcuttur. Bu

çalışmalar sayesinde hangi illerin iyi durumda olduğu, hangi illerin gelişmişlik düzeyi bakımından gerilediği ve hangi illerin idari, sosyal ve ekonomik bakımdan durumlarının iyileştirilmesi gerektiği tespit edilebilmektedir.

Ülke, bölge, il hatta ilçelere göre gelişmişlik düzeyleri seçilen tanım ve kapsama göre farklı yöntem ve değişkenler kullanılarak belirlenebilmektedir. Gelişmişlik düzeylerinin belirlenmesinde tek bir değişken kullanılabildiği gibi birden fazla ekonomik ve sosyal göstere birlikte kullanılabilmektedir.

Ülkemizde bölge, il ve ilçelerin gelişmişlik düzeyleri DPT Müsteşarlığı tarafından Temel Bileşenler Analizi Yöntemi ile farklı sosyal ve ekonomik değişkenler ile bileşik endeks oluşturularak tespit edilmektedir. Bu yöntemle belirlenen illerin sosyo-ekonomik gelişmişlik sıralaması kalkınmada öncelikli yörelerin belirlenmesinde bir ölçüt olarak kullanılmaktadır.

Bu çalışmada, iller araştırma birimi olarak ele alınmış, illerin gelişmişlik düzeyleri bakımından çok değişkenli istatistiksel yöntemler ve Veri Zarflama analizi yardımıyla sınıflandırılması ve sıralanması üzerinde durulmuştur. Çalışmanın amacı, illerin sosyal, ekonomik ve demografik değişkenler kullanılarak sınıflama ve sıralanmasında bu analizlerin birbirleriyle uyumunu karşılaştırmalı olarak incelemektir.

Çalışmada önce 81 ile ait gelişmişlik düzeylerini temsil etmek üzere DPT tarafından kullanılan 58 değişken ele alınmış bu değişkenler arasındaki korelasyon yapısı incelenmiştir Çok değişkenli istatistiksel analizlerin uygunsuz sonuçlar vermemesi bakımından, bağımlılık yapısını yok etmek amacıyla sıkı ilişkili olduğu gözlenen değişkenler ve eksik veriye sahip değişkenler araştırmadan çıkartılarak, 14 değişken ile analizler uygulanmıştır. Veriler, DPT'nin yayınladığı 2003 yılına ait illerin ve bölgelerin sosyo-ekonomik gelişmişlik sıralaması araştırmasından alınmıştır (DPT, 2003).

Çalışmada iller aşamalı olmayan kümeleme analizlerinden *K* – Ortalama Tekniği, Diskriminant analizi ve Veri Zarflama Analizi ile sınıflanmıştır. Bu analizlere ait teorik bilgi, uygulama sonuçları ve analizlerin illeri sınıflandırmadaki uyumu ile ilgili sonuçlar Bölüm 2'de yer almaktadır. İllerin sıralanması için ise Temel Bileşenler Analizi ve Veri Zarflama Analizi kullanılmıştır. Bu analizlere ait teorik bilgi, uygulama sonuçları ve analizlerin illeri sıralamadaki uyumu ile ilgili sonuçlar ise Bölüm 3'te yer almaktadır. Bölüm 4'te elde edilen sonuçların güvenilirliğini sorgulamak amacıyla yapılan simülasyon çalışması yer almaktadır. Son bölümde ise çalışmanın sonuçları tartışılmaktadır.

## I. İLLERİN SINIFLANDIRILMASI

### I. 1. Kümeleme Analizi, K-Ortalama Yöntemi

Kümeleme analizi, doğal grupları kesin belli olmayan birim veya değişkenleri aynı kümenin üyelerinin diğer kümelerin üyelerine göre daha çok benzerlik gösterdiği alt kümelere ayıran yöntemler topluluğudur. Analiz, birimleri  $p$  değişkene göre hesaplanan ve benzerlik ölçüsü olarak kullanılan bazı ölçüler kullanarak homojen gruplara ayırmayı amaçlar (Jajuga, Hans ve Bock, 2002, 113).

Küme, birbirine yakın (benzer) bireylerin oluşturduğu topluluk ya da çok boyutlu uzayda oluşturdukları bulutlar şeklinde ifade edilebilir. Bu anlamda küme kavramı, benzerlik ve uzaklık kavramlarını çağrıştırır. Kümeleme analizinde benzerlik ölçüsü olarak, gözlem vektörleri arasındaki uzaklık değerleri kullanılmaktadır.

Kümeleme Analizi, birimleri  $p$  değişken bakımından özelliklerine göre gruplara ayırmak,  $p$  sayıda değişkeni alt kümelere ayırmak ve ortak faktör yapılarını ortaya koymak, taksonomik sınıflama yapmak gibi çeşitli amaçlara hizmet eder. Değişik işlevleri yerine getirmesi nedeniyle farklı amaçlar için farklı yöntemler uygulanır. Ayrıca değişkenlerin ölçü birimlerinin ve ölçme tekniklerinin farklı olmasından dolayı birimlerin benzerliklerinin ortaya konmasında da değişik ölçüler kullanılır.

Kümeleme Analizinde, verilerin normal dağılımlı olması gerektiği varsayımı bulunmakla beraber normallik varsayımı prensipte kalmakta, uzaklık değerlerinin normalliği yeterli görülmektedir (Tatlıdil, 1996, 329-330).

$K$  – Ortalama Yöntemi, veri kümeleme problemlerinde yaygın kullanılan aşamalı olmayan kümeleme yöntemlerinden biridir. Aşağıdaki adımlar izlenerek birimler kümelere ayrılır:

1. Verilerden elde edilecek ilk  $k$  (küme sayısı) nokta çekirdek nokta olarak alınır. Bu noktaların her birinin  $p$  değişken değerleri birer küme ortalama vektörü olarak kabul edilir. Her bir birimin küme ortalama vektörlerine olan uzaklıkları hesaplanır.

2. Bu  $n - k$  birim en yakın ortalama vektörlü kümeye atanır. Her atamadan sonra, atama yapılan kümenin ortalama vektörü yeniden hesaplanır. Bir sonraki birimin atanacağı küme, bu yeni küme ortalama

vektörüne göre uzaklık hesaplanarak belirlenir. En yüksek benzerliğe sahip birimler bir araya getirilir.

3. Küme içi varyansın minimum ve kümeler arası varyansın maksimum olduğu kümeleme yapısına ulaşıncaya kadar tüm birimler  $k$  kümeye atanmaya devam eder (Özdamar, 1998, 248-249).

## **I.2. Diskriminant Analizi**

Diskriminant Analizi, grup üyeliğini tahmin etmek için kullanılan bir çok değişkenli istatistiksel yöntemdir. Bu analiz, grupları ayırmada ayırma fonksiyonu kullanır. Analiz, matematiksel teknikler kullanarak, grupların birbirinden en iyi şekilde ayırmasını sağlar. Bir diğer ifadeyle, bağımsız değişkenlerin değerlerine bağlı olarak, hangi birimin hangi grupta yer aldığına ne kadar iyi tahmin edildiğini gösterir. Her birimin hangi grupta yer aldığını belirlemek için ayırma skoru hesaplanır. Söz konusu skorlar, bağımsız değişkenlerin doğrusal kombinasyonundan oluşur. Doğrusal kombinasyon, her bir değişkeni belirli bir sabit ile çarpıp, bu çarpımların toplamından oluşur (Akgül ve Çevik, 2003, 401-402).

Fisher'in Doğrusal Diskriminant Fonksiyonu, iki ya da çoklu grupları ayırma problemine yönelik önemli istatistiksel bir yaklaşımdır. Fisher, birimleri gözlenen skorlarına dayanarak uygun gruplara sınıflamak için, gruplar arası kareler toplamının grup içi kareler toplamına oranını maksimize eden doğrusal bir fonksiyon önermiştir (Mardia, Kent ve Bibby, 1979, 318).

Doğrusal ayırma fonksiyonunun optimal olabilmesi ve yanlış sınıflandırma ihtimalini minimuma indirmek amacıyla, değişkenlerin çoklu normal dağılıma sahip olmaları ve bütün gruplar için kovaryans matrislerini eşit olması gerekir.

Analizde ele alınan değişkenler arasında çoklu bağlantı olması uygun olmayan katsayıların oluşmasına sebep olur. Bu nedenle Diskriminant Analizi yapılmadan önce çoklu bağlantı olup olmadığı araştırılmalıdır. Bu bakımdan analiz öncesinde bağımsız değişkenler arasındaki korelasyonlar incelenmeli, 0.70 ve üzerinde korelasyon katsayısına sahip değişken çiftlerinden birisi analizden çıkarılmalı veya iki değişkenin birleştirilmesi düşünülmelidir (Akgül ve Çevik, 2003, 402-403).

Diskriminant fonksiyonu elde edildikten sonra, gözlemler fonksiyonda tekrar yerine konur ve sınıflandırılır. Bu sınıflandırma sonucunda yanlış sınıflandırma oranı elde edilir. Bu yöntem, sınıflama ve bu sınıflamayı

değerlendirmede aynı gözlemlerin kullanılması nedeniyle sapmalıdır. Bir başka yol ise, her zaman mümkün olmamakla beraber veri setini bölmek ve sınıflamayı bu şekilde kontrol etmektir (Lachenbruch, 1975, 7-8).

### I.3. Veri Zarflama Analizi

VZA, girdiler ve çıktılar olarak ifade edilen benzer değişkenlere sahip karar verme birimlerinin görelî etkinliklerini ölçmek amacıyla geliştirilmiş ve çözümde doğrusal programlama tekniklerinden faydalanan parametrik olmayan bir tekniktir. Bu yöntem homojen oldukları varsayılan birimleri kendi aralarında kıyaslar. En iyi birim(ler)i etkinlik sınırı kabul ettikten sonra diğer birimler bu etkin birim(ler)e göre değerlendirilir.

VZA ilk olarak Charnes, Cooper ve Rhodes (1978) tarafından, ürettikleri mal veya hizmet açısından birbirlerine benzer karar verme birimlerinin görelî etkinliklerinin ölçülmesi amacı ile geliştirilmiş olan parametrik olmayan bir etkinlik yöntemidir. Bu yöntemin sahip olduğu özellikleri kısaca özetlersek; her KVB'deki etkinsizlik miktarını ve kaynaklarını tanımlayabilmesi, her bir KVB'nin etkinlik değeri diğerlerine göre hesaplandığından, hesaplanan etkinliklerin görelî etkinlikler olması ve değişkenler üzerinde herhangi bir fonksiyonel varsayım (Normal dağılım gibi) öne sürmemesidir (Cooper, Seiford ve Tone, 2000, 1-4).

Farrell'in fikirlerini geliştiren Charnes, Cooper ve Rhodes tek bir çıktının tek bir girdiye oranlanmasıyla elde edilen etkinlik değerini, çoklu çıktılarının çoklu girdilere oranlanmasına genişletmişlerdir. Bu sayede her bir KVB için yapay bir çıktı ve yapay bir girdi bulunmakta, bu yapay çıktı ve girdiler vasıtasıyla KVB'lerin etkinlik değerleri bulunabilmektedir. Burada ağırlıklar, etkinlik değerlerini 1'den büyük yapmayacak şekilde seçilirler (Bousofiane, Dyson ve Thanassoulis, 1991).

VZA'da etkinlik ölçümünde kullanılan çeşitli modeller vardır ve bu modeller etkinliğin, çıktılarının ağırlıklı toplamının girdilerin ağırlıklı toplamına oranı olarak ölçüldüğü oran modelinden türetilir (Cooper, Seiford ve Tone, 2000, 22-27)

$x_{ij}$  ( $i = 1, 2, \dots, m$ ) ve  $y_{rj}$  ( $r = 1, 2, \dots, s$ ) olarak  $j$ . örneğin (birimin)  $i$ . girdisini ve  $r$ . çıktısını tanımlamak üzere oran formu aşağıdaki gibi tanımlanabilir:

$$\begin{aligned} \max h_j &= \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \\ \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} &\leq 1 \quad j=1,2, \dots, n \\ u_r, v_i &\geq 0 \end{aligned} \quad (1)$$

Burada  $u_r$ ,  $r$ . çıktıya atanan ağırlığı,  $v_i$  ise  $i$ . girdiye atanan ağırlığı göstermektedir. Etkinliği ölçülen KVB'nin ağırlıklı girdi toplamı  $(\sum_{i=1}^m v_i x_{io})$  1'e eşitlenir ise CCR modeli olarak bilinen temel etkinlik modeli elde edilmiş olur.

$$\begin{aligned} \max h_j &= \sum_{r=1}^s u_r y_{rj} \\ \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0 \quad j=1,2, \dots, n \\ \sum_{i=1}^m v_i x_{io} &= 1 \\ u_r, v_i &\geq 0 \quad i=1,2, \dots, m ; r=1,2, \dots, s \end{aligned} \quad (2)$$

Bu model ile KVB'lerin etkinliği ölçülürken modelin her KVB için yani  $n$  defa çözülmesi gerekmektedir. Optimal amaç fonksiyonu ilgili KVB'nin etkinlik skorunu vermektedir. Etkinlik skoru  $h_j = 1$ 'e eşit olan her KVB etkin olarak değerlendirilir. Etkinlik skoru 1'in altında olan her birim de etkin olmayan olarak değerlendirilecektir.

VZA' da değişkenlerin girdi ve çıktı olarak ayrılması gerekir. Değişkenlerin, girdi ve çıktı olarak ayrılması birim üzerindeki etkilerine bağlıdır. Retzlaff ve Roberts (1996), girdi ve çıktı değişkenleri yerine birimler üzerinde pozitif ve negatif etkili değişkenler kavramının kullanılmasının daha doğru olacağını belirtmiştir. Retzlaff ve Roberts (1996) artışı birimin daha iyi olarak değerlendirilmesini sağlayan değişkenlerin *pozitif etkili*, tersine düşüşü birimin daha iyi olarak değerlendirilmesini sağlayan değişkenlerin ise *negatif etkili* olarak alınmasını önermiştir.

CCR modeli üretim fonksiyonunun ölçeğe göre sabit getirili olduğunu varsaymaktadır. BCC modeli, Banker, Charnes ve Cooper (1984) tarafından önerilmiştir. BCC modeli ölçeğe göre değişen getiri varsayımı altında karar verme birimlerinin etkinliğini ölçmektedir. Banker (Banker, Charnes ve Cooper,

1984), Farrell (1957) tarafından tanımlanan ve CCR modeli ile bulunan teknik etkinliğin, ölçek etkinliği ile karışmış olduğunu belirlemiş, teknik etkinliğin ölçek etkinliği ve saf teknik etkinlik olarak ayrılması gerektiğini göstermiştir. Bu nedenle ölçeğe göre değişen getiri varsayımı altında BCC modeli ile saf teknik etkinlik bulunabilmektedir. BCC modeli aşağıda verildiği gibi ifade edilebilir:

$$\begin{aligned} \max w_o &= \sum_{r=1}^s \mu_r y_{ro} + u_o \\ \sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + u_o &\leq 0 \quad j = 1, 2, \dots, n \\ \sum_{i=1}^m v_i x_{io} &= 1 \\ \mu, v &\geq 0, i = 1, 2, \dots, m; r = 1, 2, \dots, s, u_o \text{ serbest} \end{aligned} \quad (3)$$

Burada  $u_o$  ölçeğe göre getiri durumunu ölçmek için kullanılmaktadır.

#### I.4. Kappa Katsayısı

Kappa katsayısı, iki farklı gözlemcinin, bir birimler örneğini sınıflandırmaları durumunda, gözlemci çiftinin kategori yargıları arasındaki uyumu ölçen katsayıdır. Tipik olarak dört gözlü tablolarda değerlendirilir. Değerlendirme yorumları her zaman hasta/sağlam şeklinde iki sonuçlu olmayabilir, buna rağmen tablo zorunlu olarak bir kare tablodur. (Agresti, 1990, 366-367).

İki gözlemcinin veya tanı yönteminin değerlendirmeleri arasındaki uyumu ifade eden kappa katsayısı aşağıdaki gibi hesaplanır:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (4)$$

(4) eşitliğinde  $P(A)$  gözlemci kararları arasında gözlenen uyumunu ifade eder, gözlemcilerin hem fikir oldukları durum sayısının oranıdır. Çapraz hücre değerleri toplamının (her iki yargının da aynı olduğu hücreler) toplam içindeki oranı bulunarak hesaplanır.  $P(E)$  ise gözlemci kararları arasında tablo verilerine göre beklenen uyumu ifade eder.



$$\sum_{i=1}^I (\text{satır toplamı})_i * (\text{sütun toplamı})_i / N^2, (I : \text{satır sayısı} = \text{sütun sayısı})$$
 şeklinde hesaplanır.

Kappa katsayısı -1 ile +1 arasında değişir, 1'e yaklaşması tam bir uyumun göstergesidir. 0 tam bir uyumsuzluğu, -1 ise ters uyum olduğunu gösterir. 0.7 'den büyük değerlerden başlayarak yargılar arasında uyum olduğundan söz edilebilir. Kappa, yorumlanabilir olması ve farklı sonuçların karşılaştırılmasına olanak sağladığı için geniş kullanımı olan, kabul görmüş bir istatistiktir (Carletta, 1996).

### **I.5. İllerin Çok Değişkenli İstatistiksel Yöntemlerle ve Veri Zarflama Analizi ile Sınıflandırılması**

CCR, BCC Modelleri, Diskriminant ve Kümeleme Analizini birimleri sınıflama bakımından karşılaştırmayı amaçlayan bu bölümde, toplam 81 il, her dört yöntemle gelişmişlik düzeyleri bakımından etkin ve etkin olmayan şeklinde iki kümeye ayrılmak üzere ele alınmıştır. İllerin değerlendirilmesinde gelişmişlik durumlarını yansıtan on dört sosyo-ekonomik ve demografik değişken kullanılmıştır (DPT, 2003). Bu değişkenler aşağıdaki gibidir.

- $y_1$  : Şehirleşme oranı
- $y_2$  : Ücretli çalışanların toplam istihdama oranı
- $y_3$  : İşverenlerin toplam istihdama oranı
- $y_4$  : Okur-yazar nüfus oranı
- $y_5$  : 1 – bebek ölüm oranı
- $y_6$  : On bin kişiye düşen hekim sayısı
- $y_7$  : On bin kişiye düşen hastane yatağı sayısı
- $y_8$  : Fert başına gayri safi yurt içi hasıla
- $y_9$  : Banka şube sayısı
- $y_{10}$  : Fert başına banka mevduatı
- $y_{11}$  : Kırsal yerleşmelerde asfalt yol oranı
- $y_{12}$  : Yeterli içme suyu götürülen nüfus oranı
- $y_{13}$  : Devlet ve il yolları asfalt yol oranı
- $y_{14}$  : On bin kişiye düşen motorlu kara taşıtı sayısı

Uygulamanın birinci aşamasında, tüm değişkenlerin birimler üzerinde pozitif etkili oldukları düşünüldüklerinden çıktı değişkeni olarak ele alınmış ve Veri Zarflama Analizinin iki ayrı modeli olan CCR ve BCC Modelleri ile yukarıdaki değişkenler bakımından illerin etkinlik değerleri hesaplanmıştır. İller etkin olan (etkinlik değeri 1 olan) ve etkin olmayan (etkinlik değeri 1'den küçük olan) şeklinde iki kümeye ayrılmıştır. Bu modellerin çözümlemesinde WINQSB paket programı kullanılmıştır.

Kümeleme ve Diskriminant Analizinde, değişkenler girdi ve çıktı değişkenleri olarak ayrılmaksızın toplam 14 değişkenli veriye  $K$  – Ortalama yöntemi ve Diskriminant Analizi SPSS-13 paket programı kullanılarak uygulanmıştır. Diskriminant Analizi için normallik ve kovaryans matrislerinin eşitliği varsayımlarının sağlandığı test edilmiş ve Fisher'in Diskriminant Fonksiyonu kullanılmıştır. Diskriminant Analizi uygulanırken illerin gerçek grupları şöyle belirlenmiştir: DPT'nin gelişmişlik sıralaması araştırması sonucu elde edilen birinci ve ikinci derece gelişmiş iller etkin iller grubuna, diğer iller ise etkin olmayan iller grubuna sınıflandırılarak analize başlanmıştır.

Analizler sonucunda illerin sınıflandığı kümeler Tablo 1'de verilmiştir. Analiz sonuçları incelendiğinde, Adana, Ankara, Antalya, Balıkesir, Bolu, Bursa, Edirne, İstanbul, İzmir, Kırklareli, Kocaeli, Muğla, Tekirdağ ve Yalova illerinin her dört yöntem tarafından da etkin iller olarak sınıflandırıldığı görülebilir. Bu iller DPT'nin illerin ve bölgelerin gelişmişlik sıralaması araştırmasında belirlenen, birinci ve ikinci derece gelişmiş iller grubunda yer almaktadır. Adıyaman, Afyon, Ağrı, Amasya, Artvin, Bingöl, Bitlis, Çankırı, Çorum, Diyarbakır, Elazığ, Erzincan, Erzurum, Giresun, Gümüşhane, Hakkari, Isparta, Kars, Kastamonu, Kırşehir, Konya, Kütahya, Malatya, K.Maraş, Mardin, Muş, Niğde, Ordu, Rize, Samsun, Siirt, Sinop, Sivas, Tokat, Tunceli, Urfa, Uşak, Van, Yozgat, Aksaray, Bayburt, Kırıkkale, Batman, Şırnak, Bartın, Ardahan, Iğdır, Karabük, Kilis, Osmaniye, Düzce illeri ise her dört yöntem tarafından etkin olmayan iller olarak sınıflandırılmışlardır. Bu iller DPT'nin illerin ve bölgelerin gelişmişlik sıralaması araştırmasında belirlenen, üçüncü, dördüncü ve beşinci derece gelişmiş iller grubunda yer almaktadır. Toplam 66 il bütün yöntemler tarafından aynı gelişmişlik düzeyinde değerlendirilmiştir. Diğer bir deyişle, tüm yöntemler %81.5 oranında aynı sınıflandırma kararını vermişlerdir.

**Tablo-1: Dört Farklı Yöntem İle İllerin Sınıflandırılması**

Şehirler	CCR	BCC	Küme	Disk. A.	Şehirler	CCR	BCC	Küme	Disk. A.
Adana	1	1	1	1	Konya	0.907	0.961	0	0
Adıyaman	0.624	0.691	0	0	Kütahya	0.788	0.815	0	0
Afyon	0.735	0.783	0	0	Malatya	0.813	0.902	0	0
Ağrı	0.421	0.501	0	0	Manisa	0.821	0.900	1	1
Amasya	0.692	0.729	0	0	K.Maraş	0.791	0.864	0	0
Ankara	1	1	1	1	Mardin	0.591	0.667	0	0
Antalya	1	1	1	1	Muğla	1	1	1	1
Artvin	0.744	0.763	0	0	Muş	0.409	0.483	0	0
Aydın	0.899	1	1	1	Nevşehir	0.871	0.947	1	0
Balıkesir	1	1	1	1	Niğde	0.687	0.786	0	0
Bilecik	0.813	0.854	1	1	Ordu	0.718	0.764	0	0
Bingöl	0.431	0.581	0	0	Rize	0.847	0.913	0	0
Bitlis	0.499	0.602	0	0	Sakarya	0.819	0.868	1	1
Bolu	1	1	1	1	Samsun	0.790	0.806	0	0
Burdur	0.925	1	1	0	Siirt	0.378	0.456	0	0
Bursa	1	1	1	1	Sinop	0.587	0.659	0	0
Çanakkale	0.976	1	1	1	Sivas	0.617	0.711	0	0
Çankırı	0.812	0.835	0	0	Tekirdağ	1	1	1	1
Çorum	0.691	0.708	0	0	Tokat	0.710	0.758	0	0
Denizli	0.912	1	1	1	Trabzon	0.991	1	0	0
Diyarbakır	0.502	0.564	0	0	Tunceli	0.716	0.854	0	0
Edirne	1	1	1	1	Ş.Urfa	0.515	0.681	0	0
Elazığ	0.803	0.815	0	0	Uşak	0.717	0.782	0	0
Erzincan	0.759	0.862	0	0	Van	0.337	0.419	0	0
Erzurum	0.388	0.427	0	0	Yozgat	0.784	0.807	0	0
Eskişehir	0.893	1	1	1	Zonguldak	0.768	0.842	1	1
Gaziantep	0.818	0.864	0	1	Aksaray	0.698	0.738	0	0
Giresun	0.678	0.801	0	0	Bayburt	0.715	0.769	0	0
Gümüşhane	0.814	0.829	0	0	Karaman	0.704	0.718	1	0
Hakkari	0.443	0.504	0	0	Kırıkkale	0.841	0.907	0	0
Hatay	0.831	0.845	0	1	Batman	0.538	0.610	0	0
Isparta	0.867	0.912	0	0	Şırnak	0.567	0.687	0	0
Mersin	0.866	0.894	1	1	Bartın	0.789	0.796	0	0
İstanbul	1	1	1	1	Ardahan	0.397	0.437	0	0
İzmir	1	1	1	1	İğdır	0.564	0.636	0	0
Kars	0.427	0.462	0	0	Yalova	1	1	1	1
Kastamonu	0.614	0.699	0	0	Karabük	0.884	0.897	0	0
Kayseri	0.913	1	0	1	Kilis	0.534	0.615	0	0
Kırklareli	1	1	1	1	Osmaniye	0.883	0.931	0	0
Kırşehir	0.914	0.950	0	0	Düzce	0.736	0.744	0	0
Kocaeli	1	1	1	1					

Veri Zarflama Analizi kapsamında ele alınan CCR ve BCC modelleri ile Kümeleme ve Diskriminant Analizleri'nin illeri sınıflama bakımından ne kadar uyumlu olduklarının tespitinde Bölüm 2.4'de incelenen Kappa istatistiği kullanılmıştır.

CCR ve Kümeleme Analizi'nin illeri sınıflama kararlarındaki uyumun derecesini belirlemek üzere çapraz tablo aşağıdaki gibi olur:

**Tablo-2: CCR ve Kümeleme Analizi Uyumu**  
*Kümeleme Analizi*

		<b>0</b>	<b>1</b>	<b>Toplam</b>
<b>CCR</b>	<b>0</b>	55	12	67
	<b>1</b>	0	14	14
	<b>Toplam</b>	55	26	81

Analizlerin sınıflama kararları arasında gözlenen uyumunu ifade eden  $P(A)$  oranı ve analizlerin sınıflama kararları arasında tablo verilerine göre beklenen uyumu ifade eden  $P(E)$  oranı aşağıdaki gibi hesaplanır.

$$P(A) = \frac{(55+14)}{81} = 0.85 \quad , \quad P(E) = \frac{(67*55)}{81^2} + \frac{(26*14)}{81^2} = 0.62$$

Kappa katsayısı ise

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} = \frac{0.85 - 0.62}{1 - 0.62} = 0.61$$

olarak hesaplanır. Bu sonuç iki analiz değerlendirmeleri arasında çok iyi bir uyum olmadığını söylebileceğini gösterir.  $P(A)$  değerine dikkat edilecek olursa, her iki yöntemin aynı sınıflandırma oranı 0.85 olup, bu ise oldukça iyi bir orandır.

BCC ve Kümeleme Analizi'nin illeri sınıflama kararlarındaki uyumun derecesini belirlemek üzere çapraz tablo aşağıdaki gibi olur:

**Tablo-3: BCC ve Kümeleme Analizi Uyumu**  
*Kümeleme Analizi*

	<b>0</b>	<b>1</b>	<b>Toplam</b>
<b>BCC</b>			
<b>0</b>	53	7	60
<b>1</b>	2	19	21
<b>Toplam</b>	55	26	81

CCR ile Kümeleme Analizi için yapılan işlemlere benzer olarak, aşağıdaki hesaplamalar yapılmıştır.

$$P(A) = \frac{(53+19)}{81} = 0.89 \quad , \quad P(E) = \frac{(60*55)}{81^2} + \frac{(21*26)}{81^2} = 0.59$$

Kappa katsayısı ise

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} = \frac{0.89 - 0.59}{1 - 0.59} = 0.73$$

olarak hesaplanır. Bu sonuç iki analizin değerlendirmeleri arasında iyi bir uyum olduğunun söylenebileceğini gösterir.  $P(A)$  değerine dikkat edilecek olursa, her iki yöntemin aynı sınıflandırma oranı 0.89 olup, bu oldukça iyi bir orandır.

CCR ve Diskriminant Analizi'nin illeri sınıflama kararlarındaki uyumun derecesini belirlemek üzere çapraz tablo aşağıdaki gibi olur:

**Tablo-4: CCR ve Diskriminant Analizi Uyumu**  
*Diskriminant Analizi*

	<b>0</b>	<b>1</b>	<b>Toplam</b>
<b>CCR</b>			
<b>0</b>	55	12	67
<b>1</b>	0	14	14
<b>Toplam</b>	55	26	81

$$P(A) = \frac{(55+14)}{81} = 0.85 \quad , \quad P(E) = \frac{(67*55)}{81^2} + \frac{(26*14)}{81^2} = 0.62$$

Kappa katsayısı ise

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} = \frac{0.85 - 0.62}{1 - 0.62} = 0.61$$

olarak hesaplanır. Bu sonuç iki analizin değerlendirmeleri arasında çok iyi bir uyum olmadığının söylenebileceğini gösterir.  $P(A)$  değerine dikkat edilecek olursa, her iki yöntemin aynı sınıflandırma oranı 0.85 olup, bu ise oldukça iyi bir orandır.

BCC ve Diskriminant Analizi'nin illeri sınıflama kararlarındaki uyumun derecesini belirlemek üzere çapraz tablo aşağıdaki gibi olur:

**Tablo-5: BCC ve Diskriminant Analizi Uyumu**  
*Diskriminant Analizi*

	0	1	Toplam
0	53	7	60
1	2	19	21
Toplam	55	26	81

$$P(A) = \frac{(53+19)}{81} = 0.89 \quad , \quad P(E) = \frac{(60*55)}{81^2} + \frac{(21*26)}{81^2} = 0.59$$

Kappa katsayısı ise

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} = \frac{0.89 - 0.59}{1 - 0.59} = 0.73$$

olarak hesaplanır. Bu sonuç iki analizin değerlendirmeleri arasında iyi bir uyum olduğunun söylenebileceğini gösterir.  $P(A)$  değerine dikkat edilecek olursa, her iki yöntemin aynı sınıflandırma oranı 0.89 olup, bu oldukça iyi bir orandır.

Uyum sonuçlarına bakıldığında, Kappa istatistikleri bakımından Veri Zarflama Analizlerinden BCC Modelinin, Kümeleme ve Diskriminant Analizi ile daha uyumlu olduğu gözlenmiştir. Ortak sınıflandırma yüzdelerine bakıldığında ise, CCR Modelinin de Çok Değişkenli İstatistiksel Yöntemlerle birimleri aynı sınıflandırma oranının oldukça yüksek olduğu (0.85) görünmekle beraber, BCC Modeli yüzdeler bakımından da diğer yöntemlerle daha fazla uyum (0.89) göstermektedir.

## II. İLLERİN SIRALANMASI

### II.1. Süper Etkinlik VZA Modelleri

VZA'da karar verme birimleri yapılan çözümleme sonunda bulunan etkinlik skorlarına göre sıralanmaktadır. En yüksek etkinlik skoruna sahip karar verme birimi birinci sırada yer alırken, en düşük etkinlik skoruna sahip KVB son sırada yer almaktadır. Fakat, VZA'da etkin bulunan karar verme birimlerinin tamamına "1" etkinlik skoru atanması, etkin birimlerin kendi aralarında bir sıralama yapılmasına imkan vermemektedir. Sadece etkin olmayan karar verme birimlerinin sıralanmasında kullanılabilen VZA'nın bu dezavantajını yok etmek için çeşitli yöntemler geliştirilmiştir (Adler, Friedman, Sinuany ve Stern, 2002). Etkin karar birimlerin sıralanması için geliştirilen yöntemlerden en çok kullanılanı Andersen ve Petersen (1993) tarafından geliştirilen süper etkinlik modelidir. Bu yöntemdeki temel fikir, incelenen karar verme birimini tüm diğer karar verme birimlerinin doğrusal kombinasyonları ile karşılaştırmaktır. Elde edilen süper etkinlik skorunun değeri en yüksek olan karar verme birimi birinci sırada yer alacaktır. Diğer karar verme birimleri de süper etkinlik skor değerine göre büyükten küçüğe sıralanacaktır. İncelenen karar verme birimi için süper etkinlik modeli, CCR Modeli için,

$$\begin{aligned} \max h_j &= \sum_{r=1}^s u_r y_{rj_o} \\ \sum_{i=1}^m v_i x_{ij_o} &= 1 \\ \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0, \quad j=1, 2, \dots, m \quad j \neq o \\ u_r, v_i &\geq 0 \quad \text{tüm } r \text{ ve } i \text{ ler için} \end{aligned} \quad (5)$$

olarak ifade edilir. Burada  $o$ ; incelenen karar verme birimini göstermekte ve  $j \neq o$  ise süper etkinlik modelinin temelinde yatan, incelenen karar verme biriminin kısıt kümesinden çıkarılmasını simgelemektedir. Benzeri olarak BCC modeli için de süper etkinlik modeli oluşturulabilir:

$$\begin{aligned}
\max w_o &= \sum_{r=1}^s \mu_r y_{ro} + u_o \\
\sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + u_o &\leq 0 \quad j=1, 2, \dots, n ; j \neq o \\
\sum_{i=1}^m v_i x_{io} &= 1 \\
\mu, v \geq 0, i=1, 2, \dots, m ; r=1, 2, \dots, s, u_o &\text{ serbest}
\end{aligned} \tag{6}$$

## II.2. Temel Bileşenler Analizi

TBA, temel amacı aralarında korelasyon bulunan çok sayıda değişkenden oluşan veri setinin boyutunu indirgemek olan istatistiki bir yöntemdir. Yöntem, boyut indirgemeyi, veri setini yeni bir değişken setine (temel bileşenler) dönüştürerek gerçekleştirir.  $p$  sayıda değişken tarafından açıklanan ilişki yapısı, aralarında korelasyon bulunmayan ve orijinal değişken sayısından daha az sayıda orijinal değişkenlerin doğrusal bileşenleri olan temel bileşenler ile açıklanmaktadır.  $\mathbf{X}$  veri matrisinde yer alan  $p$  değişkenin doğrusal bileşenlerini bulmak için kovaryans ya da korelasyon matrisinin öz değerleri ve birim öz vektörleri kullanılır.  $\mathbf{X}'$  [ . . . . . ] rasgele vektörünün kovaryans matrisi  $\mathbf{\Sigma}$ , korelasyon matrisi  $\mathbf{\rho}$  ve korelasyon matrisinin öz değerleri  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  ve birbirine dik özvektörler  $l_1, l_2, \dots, l_p$  olsun. Değişkenlere ilişkin doğrusal bileşenler,  $PC_i = l_i' \mathbf{X}$ , ( $i=1, 2, \dots, p$ ) şeklinde hesaplanır.  $k$ . temel bileşenin toplam varyansı açıklama oranı,  $\frac{\lambda_k}{\lambda_1 + \dots + \lambda_p}$  olarak tanımlanmaktadır (Jolliffe, 2002, 1-6).

Temel Bileşenler Analizi boyut indirgeme amacıyla kullanılabilirdiği gibi, birimleri hesaplanan skorlarına göre sıralama işlemini de gerçekleştirebilmektedir. Her hangi bir veri matrisi için TBA sıralama süreci aşağıdaki gibi yürütülür (Zhu, 1998, Premachandra, 2002) :

*1.Adım:* Örneklemeye ilişkin korelasyon matrisi ( $R$ ) bulunur.

*2.Adım:* Örneklemeye ilişkin korelasyon matrisinin öz değer ve öz vektörleri hesaplanır. Bunun için,  $I_p$ ,  $p \times p$  lik birim matris olmak üzere,



$$|R - \lambda I_p| = 0$$

denklemi çözülerek,  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p$  olmak üzere  $p$  tane öz değer ( $\sum_{k=1}^p \hat{\lambda}_k = p$ ) ve her bir öz değere karşılık  $p$  tane  $(\hat{l}_1, \hat{l}_2, \dots, \hat{l}_p)$  öz vektör elde edilir.

3.Adım: Temel bileşenler hesaplanır. Her bir temel bileşen  $PC_k = \sqrt{\hat{\lambda}_k} \hat{l}_k$  ( $k = 1, \dots, p$ ) şeklinde elde edilir.

4.Adım:  $\sum_{k=1}^m \hat{\lambda}_k / p > 0.90$  olacak şekilde  $m$  tane temel bileşen alınır.

5.Adım:  $t = \sum_{k=1}^m w_k PC_k$ ; 4. adımda seçilen  $m$  tane temel bileşenin açıklama oranları ile ağırlıklandırılarak bir doğrusal bağıntı elde edilmesini sağlar.

6.Adım:  $Z$  matrisi standardize veri matrisi olmak üzere, temel bileşenler skorları  $PC_{Skor} = Z \times t$  yardımı ile hesaplanır ve bu skorların büyüklüklerine göre birimler sıralanır.

### II.3. Spearman'ın Sıra Korelasyon Katsayısı

Spearman'ın sıra korelasyon testi aşağıdaki gibi tanımlanabilir (Gamgam, 1998, 210-218): Rassal olarak seçilen  $n$  çaplı örnekten  $X$  ve  $Y$  değişkenlerinin değerleri saptanmış olsun. Örnek birimlerinin  $X$  değişkeni bakımından aldıkları değerlerine büyüklük sıra sayıları verilir ve bu büyüklük sıra sayıları  $R(x_i)$  ile gösterilir. Benzer şekilde örnek birimlerinin  $Y$  değişkeni bakımından aldıkları değerlerine de sıra sayıları verilir ve bu büyüklük sıra sayıları da  $R(y_i)$  ile gösterilir. Bu durumda  $R(x_i)$  ve  $R(y_i)$  değerlerine bağlı olarak sıra korelasyon katsayısı olan  $r_s$  istatistiği aşağıdaki gibi tanımlanabilir.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (R_{(x_i)} - R_{(y_i)})^2 \quad (7)$$

(7)'deki test istatistiği kullanılarak verilen bir  $\alpha$  anlamlılık düzeyinde,  $X$  ve  $Y$  değişkenleri arasında ilişki olmadığını iddia eden  $H_0$  hipotezi,  $X$  ve  $Y$  değişkenleri arasında ilişki olduğunu (aynı veya ters yönlü) iddia eden  $H_1$  hipotezine karşı test edilir.

#### II.4. İllerin Temel Bileşenler Analizi ve Veri Zarflama Analizi ile Sıralanması

Bu bölümde toplam 81 ilin gelişmişlik düzeyleri bakımından sıralanması, VZA'nın süper etkinlik CCR-BCC modelleri ve sıralama amaçlı Temel Bileşenler Analizi kullanılarak yapıldı. Bu analizlerde kullanılan değişkenler sınıflama amaçlı analizlerde kullanılan değişkenler ile aynıdır. Süper etkinlik CCR-BCC modellerinin analizi WINQSB paket programı kullanılarak, Temel Bileşenler Analizi ise MATLAB 7.0 programı kullanılarak gerçekleştirildi.

Tablo 6'da 81 ile ait süper etkinlik CCR-BCC modelleri ve Temel Bileşenler Analizi ile elde edilen sıralama skor değerleri ve illerin gelişmişlik sıralamaları yer almaktadır.

Tablo 6 incelendiğinde, 3 yöntemin de İstanbul, Ankara, İzmir, Kocaeli, Muğla, Antalya, Bolu, Bursa ve Yalova gibi gelişmiş illere yaklaşık olarak aynı ve 81 il içersinde ilk sıralama değerlerini verdikleri görülebilir. Bu illerin dışında, her üç yöntemin Adana, Afyon, Ağrı, Balıkesir, Bingöl, Burdur, Çanakkale, Diyarbakır, Elazığ, Gazi Antep, Hakkari, Isparta, Kars, Muş, Nevşehir, Ordu, Sakarya, Siirt, Sivas, Tokat, Ş.Urfa, Van, Aksaray, Batman, Şırak, Bartın, Ardahan, Iğdır, Kilis illerine de yaklaşık sıralama değerleri verdikleri gözlenmektedir. Tablonun geneli için bu sıralama değerleri arasındaki ilişkiyi test etmek istediğimizde ise Spearman'ın sıra korelasyonu kullanılabilir.

**Tablo-6: Süper Etkinlik ve PCA Skor Değerleri ve Sıralama Sonuçları**

İller	CCR		BCC		PCA		İller	CCR		BCC		PCA	
	Skor	Sıra	Skor	Sıra	Skor	Sıra		Skor	Sıra	Skor	Sıra	Skor	Sıra
Adana	1.72	11	10.07	10	9.70	12	Konya	0.907	21	0.961	22	-0.01	33
Adıyaman	0.624	61	0.691	63	-10.67	70	Kütahya	0.788	43	0.815	44	-0.04	34
Afyon	0.735	49	0.783	50	-3.80	48	Malatya	0.813	37	0.902	29	-6.06	54
Ağrı	0.421	76	0.501	75	-14.41	80	Manisa	0.821	32	0.900	30	7.44	18
Amasya	0.692	57	0.729	58	-2.28	41	K.Maraş	0.791	40	0.864	35	-5.89	53
Ankara	18.89	2	28.45	2	42.46	2	Mardin	0.591	64	0.667	66	-11.51	73
Antalya	8.28	7	15.67	7	11.71	7	Muğla	12.78	5	19.01	6	17.99	6
Artvin	0.744	47	0.763	54	0.24	31	Muş	0.409	77	0.483	76	-14.73	81
Aydın	0.899	22	7.18	13	8.07	16	Nevşehir	0.871	26	0.947	24	2.90	29
Balıkesir	1.64	12	6.95	14	9.08	15	Niğde	0.687	59	0.786	49	-1.84	39
Bilecik	0.813	36	0.854	37	4.03	24	Ordu	0.718	50	0.764	53	-8.31	59
Bingöl	0.431	74	0.581	72	-12.61	77	Rize	0.847	29	0.913	26	-1.42	37
Bitlis	0.499	72	0.602	71	-13.76	79	Sakarya	0.819	33	0.868	33	3.51	27
Bolu	10.09	6	20.08	5	21.93	5	Samsun	0.790	41	0.806	46	0.11	32
Burdur	0.925	17	5.47	16	7.45	17	Siirt	0.378	80	0.456	78	-11.61	74
Bursa	7.19	8	14.28	8	10.94	8	Sinop	0.587	65	0.659	67	-3.22	44
Çanakkale	0.976	16	5.54	15	7.26	19	Sivas	0.617	62	0.711	60	-6.12	55
Çankırı	0.812	38	0.835	41	-8.43	60	Tekirdağ	1.08	14	2.17	19	3.75	26
Çorum	0.691	58	0.708	61	-1.92	40	Tokat	0.710	54	0.758	55	-6.96	57
Denizli	0.912	20	8.28	12	9.42	14	Trabzon	0.991	15	1.18	20	-3.63	47
Diyarbakır	0.502	71	0.564	73	-9.89	68	Tunceli	0.716	52	0.854	38	-5.65	52
Edirne	1.16	13	4.92	17	7.14	20	Ş.Urfa	0.515	70	0.681	65	-9.84	66
Elazığ	0.803	39	0.815	43	-3.21	43	Uşak	0.717	51	0.782	51	5.02	22
Erzincan	0.759	46	0.862	36	-6.47	56	Van	0.337	81	0.419	81	-11.90	75
Erzurum	0.388	79	0.427	80	-9.86	67	Yozgat	0.784	44	0.807	45	-10.36	69
Eskişehir	0.893	23	12.18	9	10.93	9	Zonguldak	0.768	45	0.842	40	10.88	10
Gaziantep	0.818	34	0.864	34	-0.50	36	Aksaray	0.698	56	0.738	57	-4.66	49
Giresun	0.678	60	0.801	47	-4.79	50	Bayburt	0.715	53	0.769	52	-9.18	64
Gümüşhane	0.814	35	0.829	42	-9.17	63	Karaman	0.704	55	0.718	59	4.38	23
Hakkari	0.443	73	0.504	74	-13.62	78	Kırıkkale	0.841	30	0.907	28	-1.43	38
Hatay	0.831	31	0.845	39	3.89	25	Batman	0.538	68	0.610	70	-11.21	72
Isparta	0.867	27	0.912	27	0.99	30	Şırnak	0.567	66	0.687	64	-9.28	65
Mersin	0.866	28	0.894	32	6.92	21	Bartın	0.789	42	0.796	48	-3.25	45
İstanbul	21.17	1	32.68	1	45.46	1	Ardahan	0.397	78	0.437	79	-12.01	76
İzmir	13.19	4	28.17	3	22.19	4	Iğdır	0.564	67	0.636	68	-8.59	61
Kars	0.427	75	0.462	77	-10.90	71	Yalova	2.87	9	9.82	11	9.66	13
Kastamonu	0.614	63	0.699	62	-0.06	35	Karabük	0.884	24	0.897	31	-3.49	46
Kayseri	0.913	19	3.87	18	3.02	28	Kilis	0.534	69	0.615	69	-5.27	51
Kırklareli	2.18	10	1.08	21	10.23	11	Osmaniye	0.883	25	0.931	25	-8.79	62
Kırşehir	0.914	18	0.950	23	-2.56	42	Düzce	0.736	48	0.744	56	-8.01	58
Kocaeli	14.28	3	25.37	4	25.05	3							

Temel Bileşenler Analizinin sıralamasının süper etkinlik CCR ve BCC modelleri ile illerin sıralama değerleri arasında fark olup olmadığını test etmede kullanılan Spearman'ın sıra korelasyon testi sonuçları aşağıda verilmiştir.

CCR süper etkinlik sıralaması ve TBA sıralaması arasında ilişki olmadığını iddia eden  $H_0$  hipotezine ilişkin test istatistiğinin değeri bu örnek için  $(r_s)_{CCR-TBA} = 1 - \frac{6(14402)}{81(81^2 - 1)} = 0.837$  olarak hesaplanmış, bu test istatistiğine ilişkin  $p$  değeri

$p < 0.000001$  olduğundan sıralama değerleri arasında aynı yönlü ilişki olduğunu iddia eden  $H_1$  hipotezine karşı reddedilmiştir. Benzer olarak, BCC süper etkinlik sıralaması ve TBA sıralaması arasında ilişki olmadığını iddia eden  $H_0$  hipotezine ilişkin test istatistiğinin değeri için

$(r_s)_{BCC-TBA} = 1 - \frac{6(14124)}{81(81^2 - 1)} = 0.840$  olarak hesaplanmış, bu test istatistiğine

ilişkin  $p$  değeri  $p < 0.000001$  olduğundan sıralama değerleri arasında aynı yönlü ilişki olduğunu iddia eden  $H_1$  hipotezine karşı reddedilmiştir.

Bu veri kümesi için yapılan hesaplamalarda, TBA ve CCR-BCC kullanılarak elde edilen sıralama değerleri arasında aynı yönlü ilişki bulunmuş, bu üç analiz KVB'leri sıralamada birbirleri yerine kullanılabilir sonucuna varılmıştır.

### III. SİMÜLASYON ÇALIŞMASI

Bu çalışmadaki ana amaç, illerin sınıflandırılması ve sıralanmasını çok değişkenli istatistiksel yöntemler ve herhangi bir varsayıma gerek duymayan Veri Zarflama Analizi ile incelemek ve karar verme birimlerini sınıflandırma ve sıralama ihtiyaçlarına yeni bir açılım sunmaktır. Bu amaçla yapılan çalışmaya ek olarak, bu bölümde sınıflandırma yöntemleri (Kümeleme Analizi ve BCC modeli, Diskriminant Analizi ve BCC modeli) arasındaki uyumun ve sıralama yöntemleri (sıralama amaçlı Temel Bileşenler Analizi ve Süper etkinlik CCR-BCC modelleri) arasındaki ilişkinin farklı değişken sayıları ve karar verme birimi sayıları için incelendiği bir simülasyon çalışması tasarlanmıştır. Bu sayede, elde edilen sonuçların genellenebilir nitelikte olması hedeflenmiştir.

Gerçek veri çalışmasında Kümeleme Analizi ve BCC modeli, Diskriminant Analizi ve BCC modeli arasında gözlenen uyumun, farklı KVB sayıları ve farklı girdi-çıkıtı değişkeni sayıları için de elde edilip edilemeyeceğini görmek amacıyla yapılan simülasyon çalışmasında, KVB' leri değerlendirmek için kullanılan girdi ve çıktı değişkenlerine ait gözlem değerleri Çok Değişkenli Normal Dağılımdan rasgele seçilmiştir. Çalışma,  $n = 30, 50, 80$  ve  $100$  örneklem büyüklükleri için farklı sayıda girdi-çıkıtı değişkeni kullanılarak, her durum için 1000 defa tekrarlanmıştır. Her bir durum için Kümeleme Analizi ve BCC, Diskriminant Analizi ve BCC tarafından yapılan sınıflandırmalar arasındaki uyum illere ait gerçek veri setinde olduğu gibi Kappa katsayısı ile incelenmiştir. Simülasyon çalışması MATLAB 7 programı ile yapılmıştır. Elde edilen sonuçlar Tablo-7' de özetlenmiştir.

**Tablo-7: Sınıflandırma Yöntemlerinin Uyumuna İlişkin Simülasyon Sonuçları**

$n$	girdi	çıkıtı	$\mathcal{K}_{küme-BCC}$	$\mathcal{K}_{disk-BCC}$	$n$	girdi	çıkıtı	$\mathcal{K}_{küme-BCC}$	$\mathcal{K}_{disk-BCC}$
30	2	2	0.771	0.800	80	5	5	0.733	0.748
	2	4	0.748	0.778		7	5	0.701	0.724
	3	2	0.751	0.781		6	8	0.682	0.718
	3	3	0.749	0.774		6	6	0.707	0.740
	4	3	0.731	0.760		8	7	0.671	0.700
50	3	3	0.737	0.769	100	6	6	0.713	0.738
	4	4	0.721	0.730		8	7	0.668	0.692
	4	5	0.718	0.734		7	9	0.672	0.721
	6	4	0.713	0.722		10	9	0.654	0.695
	5	5	0.715	0.739		10	10	0.663	0.708

$\mathcal{K}_{küme-BCC}$  sütununda, her bir durum için her 1000 tekrarda Kümeleme Analizi ve BCC modeli tarafından yapılan sınıflamanın uyumunu gösteren Kappa katsayılarının ortalaması yer almaktadır. Örneğin, 30 tane KVB ( $n = 30$ ), 2 tane girdi ve 2 tane çıktı olduğunda yapılan 1000 denemede, Kümeleme Analizi ve BCC modeli için elde edilen ortalama uyum istatistiği  $\mathcal{K}_{küme-BCC} = 0.771$  olarak hesaplanmıştır. Benzer olarak  $\mathcal{K}_{disk-BCC} = 0.800$  değeri de Diskriminant Analizi ve BCC modeli arasındaki uyumu ölçen Kappa istatistiğinin 1000 denemedeği ortalama değeridir. Tablo 7 incelendiğinde, karar verme birimlerinin ve değişkenlerin sayısı küçük olduğunda çok değişkenli sınıflandırma yöntemleri ile BCC modeli arasında yüksek sayılabilecek bir uyum mevcuttur. Karar verme birimlerinin ve değişkenlerin sayısı arttıkça bu uyum gözle görülür bir şekilde azalmaktadır. Ayrıca, ele alınan her durumda

Diskriminant Analizinin BCC modeli ile olan uyumu Kümeleme Analizinin BCC modeli ile olan uyumundan açık bir şekilde üstündür. Buradan, uygulandığında herhangi bir varsayıma gerek duymayan doğrusal programlama tabanlı VZA'nın BCC modeli makul büyüklükteki (ele alınan durumlar için  $n < 80$  ve girdi sayısı + çıktı sayısı  $< 12$ ) karar verme birimi ve değişken sayılarında Kümeleme ve Diskriminant Analizi yöntemlerine alternatif olduğu söylenilebilir.

Benzer şekilde sıralama amaçlı Temel Bileşenler Analizi ve süper etkinlik CCR-BCC modelleri için de simülasyon çalışması tasarlanmıştır. Çalışmada seçilen karar verme birimlerinin ve değişkenlerin sayısı sınıflama amaçlı yapılan simülasyon çalışmasındaki ile aynıdır. Her bir durum için sıralama amaçlı uygulanan Temel Bileşenler Analizi ve süper etkinlik CCR-BCC modelleri tarafından hesaplanan sıralama değerleri arasında ilişki olup olmadığı Spearman'ın sıra korelasyonu katsayısı ile test edilmiştir. Elde edilen sonuçlar Tablo 8' de özetlenmiştir.

**Tablo-8: Sıralama Yöntemleri Arasındaki İlişkiye İlişkin Simülasyon Sonuçları**

$n$	girdi	çıkı.	$r_{hs1}$	$r_{hs2}$
30	2	2	900	895
	2	4	912	920
	3	2	899	911
	3	3	905	887
	4	3	891	908
50	3	3	886	884
	4	4	925	901
	4	5	910	927
	6	4	847	863
	5	5	895	888
50	5	5	908	910
	7	5	903	924
	6	8	918	884
	6	6	875	859
	8	7	864	876
100	6	6	851	900
	8	7	884	878
	7	9	886	891
	10	9	852	849
	10	10	852	861

*rhs1* ve *rhs2* sütunlarında, sırasıyla her bir durum için her 1000 tekrarda süper etkinlik CCR modeli ile Temel Bileşenler Analizi ve süper etkinlik BCC modeli ile Temel Bileşenler Analizi sıra sayıları arasında ilişki olmadığı hipotezini reddetme sayısı verilmiştir. Örneğin, 30 tane KVB ( $n = 30$ ), 2 tane girdi ve 2 tane çıktı olduğunda yapılan 1000 denemenin 900 ( $rhs1 = 900$ ) tanesi için süper etkinlik CCR modeli ile Temel Bileşenler Analizi tarafından hesaplanan sıra sayıları arasında ilişki olmadığı hipotezi reddedilmiş, bu sıra sayıları arasında aynı yönlü ilişki olduğu sonucuna varılmıştır. Tablo 8 bir bütün halinde incelendiğinde, karar verme birimlerinin sıralanmasında Temel Bileşenler Analizinin hem süper etkinlik CCR modeli ile hem de süper etkinlik BCC modeli ile ele alınan tüm durumlarda, yüksek oranlarda aynı yönlü ilişkili olduğu gözlenmiştir. Buradan, süper etkinlik CCR-BCC modellerinin karar verme birimlerinin sıralanmasında, Temel Bileşenler Analizi'ne alternatif olarak kullanılabilceği söylenebilir.

## SONUÇ

Bölge, il ve ilçelerin gelişmişlik düzeylerinin sınıflama ve sıralamaları Temel Bileşenler Analizi ve Kümeleme Analizi gibi çok değişkenli istatistiksel yöntemler ile incelenmektedir. Bu tür yöntemlerin sağlıklı sonuçlar vermesi varsayımlarının sağlanmasına bağlıdır.

Bu çalışmada illerin sınıflama ve sıralanmasında bu çok değişkenli istatistiksel yöntemlerle birlikte bu yöntemlere alternatif olarak, son yıllarda kullanım alanı gittikçe artan ve kullanımında herhangi bir varsayıma gerek duymayan doğrusal programlama tabanlı VZA yöntemi kullanılmış ve bu yöntemlerin sıralama ve sınıflamadaki performansları karşılaştırmalı olarak incelenmiştir.

Sınıflandırma işlemi için, birimlerin yakınlıklarını sınıflandırmaya temel alan çok değişkenli istatistiksel yöntemler olan K-Ortalama Yöntemi ile Diskriminant Analizi ve günümüzde kullanımı yaygınlaşmış, birimlerin çıktı / girdi değişken oranlarını kullanarak sınıflandırma yapan Veri Zarflama Analizi'nin iki değişik formu olan CCR ve BCC Modelleri kullanılmıştır.

Analiz sonuçları değerlendirildiğinde, Kümeleme Analizi'nin gelişmişlik düzeyine göre sınıfları belirlemede, çalışmada kullanılan Veri Zarflama Analizlerinden BCC Modeli ile 0.73 oranında uyumlu oldukları gözlenmiştir. Bu sonuç, analizler arasında sınıflama bakımından iyi bir uyum olduğunu göstermenin yanı sıra, bu iki yöntem illeri 0.89 oranında aynı gruplara sınıflandırmıştır. Benzer olarak, Diskriminant Analizi de BCC Modeli ile 0.73 oranında uyumlu bulunmuş ve bu iki yöntem illeri 0.89 oranında aynı gruplara

sınıflandırmıştır. CCR Modelinin ise bu iki çok değişkenli istatistiksel yöntem ile uyumlu olmadığı gözlenmiştir.

Sıralama işlemi için ise; birimleri hesaplanan skor değerlerine göre sıralayan Temel Bileşenler Analizi ile yine Veri Zarflama Analizi'nin iki değişik formu olan CCR ve BCC Modelleri kullanılmıştır.

Analiz sonuçları değerlendirildiğinde, gelişmişlik düzeylerine göre illeri sıralamada Temel Bileşenler Analizi ile BCC ve CCR modellerinin sıra sayıları arasındaki Spearman'ın korelasyon katsayısı sırasıyla 0.840 ve 0.837 olarak bulunmuştur. Bu katsayılar, yöntemler arasında aynı yönlü, yüksek ilişki olduğunu göstermektedir.

İllerin gelişmişlik düzeylerini sınıflama ve sıralama bakımından belirlemede, illeri çok değişkenli istatistiksel yöntemler ile veri zarflama analizinin iki farklı modeli olan CCR ve BCC modellerini karşılaştırmak için yapılan bu gerçek veri çalışmasının sonuçlarını sağlam bir temele oturtmak ve bu sonuçların güvenilirliğini sağlamak üzere simülasyon çalışması yapılmıştır. Simülasyon çalışmasında, farklı KVB sayıları ve farklı girdi-çıkıtı değişkeni sayıları için deney her bir durumda 1000 kez, toplamda 20000 kez tekrar edilmiştir.

Simülasyon çalışması sonucunda, KVB'leri sınıflamak için kullanılan Diskriminant Analizi ve Kümeleme Analizi ile BCC Modeli arasında yüksek bir uyum olduğu saptanmıştır. Ayrıca, KVB ve değişken sayısı arttıkça bu uyumun gözle görülür bir şekilde azaldığı gözlenmiştir. Elde edilen bir diğer önemli sonuç ise, KVB'leri sıralamak amacıyla kullanılan Temel Bileşenler Analizi ve CCR ve BCC Modelleri arasında ise ele alınan her durumda aynı yönlü yüksek ilişki olduğudur. Bu sonuçlar gerçek veri çalışmasının sonuçlarını destekler niteliktedir.

Simülasyon uygulamasıyla da desteklenen bu çalışmada, son zamanlarda geniş kullanıma sahip, doğrusal programlama tabanlı Veri Zarflama Analizi'nin, varsayım gerektirmemesi ve araştırmacıya sunduğu kullanım esnekliği nedeniyle ele aldığımız çok değişkenli istatistiksel yöntemlere alternatif bir yöntem olarak kullanılabilmesi sonucuna varılmıştır.

## KAYNAKLAR

Adler, N., L., Friedman, Z. Sinuany-Stern (2002) "Review of Ranking Methods in the Data Envelopment Analysis Context", **European Journal of Operational Research**, 140: 249–265.

Agresti A. (1990) **Categorical Data Analysis**, New York: Wiley.



- Akgül, A., O. Çevik (2003) **İstatistiksel Analiz Teknikleri: SPSS'te İşletim Yönetimi Uygulamaları**, Ankara: Emek Ofset Ltd.Şti.
- Andersen, P., N. Petersen (1993) "A Procedure for Ranking Efficient Units in Data Envelopment Analysis", **Management Science**, 39(10), 1261-1264.
- Anderson, T.W. (1984) **An Introduction to Multivariate Analysis**, New York: Wiley.
- Banker, R.D., A. Charnes, W.W. Cooper (1984) "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", **Management Science**, 30(19), 1078-1092.
- Boussofiene, A., R.G. Dyson, E. Thanassoulis (1991) "Applied Data Envelopment Analysis", **European Journal of Operational Research**, 52, 1-15.
- Carletta, J. (1996) "Assessing Agreement on Classification Tasks: the Kappa Statistic", **Computational Linguistics**, 22(2), 249-254.
- Charnes A., W.W Cooper, E. Rhodes (1978) "The Efficiency of Decision Making Units", **European Journal of Operational Research**, 2, 429-444.
- Cooper, W.W., L.M. Seiford, K. Tone (2000) **Data Envelopment Analysis**, Boston: Kluwer Academic Publishers.
- DPT (2003) **İllerin ve Bölgelerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması**, Ankara: DPT Yayınları.
- Farrell, M.J. (1957) "The Measurement of Productivity Efficiency", **Journal of Royal Statistical Society**, Serie A, CXX, 253-287.
- Gamgam, H. (1998) **Parametrik Olmayan İstatistiksel Teknikler**, Ankara: Gazi Üniversitesi Yayınları.
- Jajuga, K., A.S. Hans, H. Bock (2002) **Classification, Clustering and Data Analysis**, New York: Springer.
- Jolliffe, I.T. (2002) **Principal Component Analysis**, New York: Springer.
- Lachenburch, P.A. (1975) **Discriminant Analysis**, New York: Hafner Press.
- Mardia, K., J. Kent, J. Biby (1979) **Multivariate Analysis**, New York: Academic Press.
- Özdamar, K. (1998) **Paket Programlar ile İstatistiksel Veri Analizi II: Çokdeğişkenli İstatistiksel Yöntemler**, Eskişehir: Kaan Kitabevi.

- Premachandra, I.M. (2001) "A Note On DEA vs Principal Component Analysis: An Improvement to Joe Zhu's Approach", **European Journal of Operational Research**, 132: 553-560.
- Retzlaff-Roberts, D.L. (1996) "Relating Discriminant Analysis and Data Envelopment Analysis to One Another", **European Journal of Operational Research**, 23: 311-322.
- Tatlıdil, H. (1996) **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Ankara: Cem Web Ofset.
- Zhu, J. (1998) "Data Envelopment Analysis vs Principal Component Analysis: An Illustrative Study of Economic Performance of Chinese Cities", **European Journal of Operational Research**, 111, 50-61.