

Türkiye'deki İllerin Ulaştırma Göstergelerine Göre Kümeleme Analizi Yöntemleriyle Sınıflandırılması

Aynur İncekırık^{a, b}, Erkan Altın^c

Özet

Bu çalışmada, Türkiye'deki illerin 2004-2018 yılları arasındaki ulaştırma göstergeleri veri setini oluşturan 40 adet değişken kullanılarak kümeleme analizi yöntemleri uygulanmıştır. Analizde kullanılan veri setinin uygunluğunu sağlamak üzere Öklid uzaklık ve Min-Max standartlaştırma ölçüleri kullanılmıştır. Çalışmada, ulaştırma göstergelerini oluşturan 40 adet değişken dikkate alınarak Türkiye'deki mevcut 81 ilin kümeleme yapısının ortaya konulması amaçlanmıştır. Tek-bağlantı yöntemi, kullanılan hiyerarşik yöntemler arasında kofenetik korelasyon katsayısı değerinin yüksekliği sebebiyle en başarılı yöntem olarak kabul edilerek sonuçları değerlendirilmiştir. Ayrıca hiyerarşik olmayan yöntemlerden olan k-ortalama yöntemini sonuçlarının incelenmesi ve yorumlanması yer almaktadır. Küme sayısı, küme geçerlilik indeksleri yardımı ile beş olarak belirlenmiştir. Elde edilen sonuçlar değerlendirildiğinde, ulaştırma göstergeleri bakımından büyük kent merkezlerine sahip illerde diğer illere göre daha belirgin farkların ortaya çıktığı görülmektedir.

Anahtar Kelimeler

Ulaştırma Göstergeleri
Kümeleme Analizi
Küme Geçerlilik İndeksleri
Kofenetik Korelasyon Katsayısı

Makale Hakkında

Geliş Tarihi: 05.03.2021
Kabul Tarihi: 24.09.2021
Doi: 10.18026/cbayarsos.891649

Classification of Provinces in Turkey by Clustering Analysis Methods According to Transport Indicators

Abstract

In this study, constituting the transport indicators provinces in Turkey between the 2004-2018 year data set of 40 variables using cluster analysis methods were applied. Euclidean distance and Min-Max standardization measures were used to ensure the suitability of the data set used in the analysis. In the study, considering 40 variables constitute the transportation indicators available in Turkey 81 Province cluster is intended to reveal the structure. The single-connection method was accepted as the most successful method among the hierarchical methods used, due to the high value of the cophenetic correlation coefficient, and its results were evaluated. In addition, the analysis and interpretation of the results of the k-means method, which is one of the non-hierarchical methods, is included. The number of clusters was determined as five with the help of cluster validity indexes. Finally, it is seen that there are more distinct differences in transportation indicators in provinces with large city centers compared to other provinces.

Keywords

Transport Indicators
Cluster Analysis
Cluster Validation Indices
Cophenetic Correlation
Coefficient

About Article

Received: 05.03.2021
Accepted: 24.09.2021
Doi: 10.18026/cbayarsos.891649

^a İletişim Yazarı: aynur.incekirik@cbu.edu.tr

^b Dr. Öğr. Üyesi, Manisa Celal Bayar Üniversitesi, İİBF, Ekonometri Bölümü, ORCID: 0000-0002-5029-6036.

^c Yüksek Lisans Öğrencisi, Manisa Celal Bayar Üniversitesi, İİBF, Ekonometri Bölümü, ORCID: 0000-0001-9426-8670.

Giriş

Ulaştırma, herhangi bir canlı veya cansız varlığın bir noktadan başka bir noktaya belli araçlar kullanılarak belli amaçlar doğrultusunda ekonomik bir kavram olarak ele alınabilir nitelikte yer değiştirmesini sağlayan bir hizmettir. Bir objenin yer değiştirme işlemi için ulaştırma faaliyetini gerçekleştirecek unsurlardan faydalanılması şarttır. Ulaştırma büyüklüğü, etki alanı ve ulaşımın ivedilik durumuna göre ulaştırma altyapısı, araç ve hız talebi değişebilir olup İngilizcede “transport”, “transportation” şeklinde ifade edilen ulaştırma, aynı zamanda nakliye, taşımacılık şeklinde de kullanılabilir (Kaya, 2012).

Ulaştırma faaliyet ve sistemleri, tarih boyunca medeniyetlerin yükselişinde (Mısır, Roma ve Çin medeniyetleri gibi), toplumların gelişmesinde (sosyal yapıların yaratılması) ve ulusal savunmada (Roma İmparatorluğu karayolu ağı, Amerikan karayolu ağı gibi) birçok farklı rol oynamıştır (Rodrigue ve diğerleri, 2013). Ulaştırma türleri yeryüzünde hava, kara ve deniz olmak üzere üç tür mekânda faaliyet göstermekte olup bu türlerin her biri kendine has avantaj ve dezavantajlara sahip beş genel ulaştırma şekliyle incelenmektedir:

- a. Karayolu ulaştırması,
- b. Demiryolu, ulaştırması,
- c. Denizyolu ulaştırması,
- d. Havayolu ulaştırması,
- e. Boru hatları ulaştırması,

Bu çalışmada, yukarıda belirtilen ulaşım türleri göz önüne alınarak 2004-2018 yılları arasında ulaşım faaliyetlerini oluşturan 40 adet gösterge kullanılmıştır. Türkiye’deki illerin ulaşım faaliyetleri bakımından benzerliklerini ve gelişim durumlarını ortaya çıkarabilmek amacıyla kümeleme analizi uygulanmıştır. Analiz sonuçlarından elde edilen benzerlik ve farklılıkların sebeplerinin incelenmesi ve yorumlanması, çalışmanın diğer bir amacıdır. Literatürde illerin sınıflandırılması konusunda yapılan çalışmalar mevcuttur. Kümeleme analizi yöntemleri, kendi içlerinde homojen aralarında heterojen küme yapıları oluşturarak araştırmacılara veri setini incelemeyi kolaylaştırıcı ve sağlıklı yorumlama yapılabilme imkânı sağlamayı amaçlamaktadır. Bu analizin ortaya çıkışı 1940’lı yıllara kadar uzanmaktadır ve yöntemin çeşitli veri setleri üzerine uygulanışıyla ilgili başarılı çalışmalar literatürde mevcuttur. Kümeleme analizi yöntemleri genel olarak tıptan ekonomiye kadar birçok alanda kullanılmaktadır.

Literatürde daha önce ekonomi, sağlık, eğitim, kültür ve ziraat gibi alanlarda iller bazında incelenen bazı çalışmalardan farklı olarak, bu çalışmada ulaştırma kavramına özgü faaliyetler ele alınmış ve illerin benzerlik durumları ortaya çıkarılarak gelişimleri incelenmiştir. Ulaştırma üzerine yapılan çalışmada, ulaşımın tüm yönleriyle değerlendirilmesi konusundaki literatür boşluğu dikkate alınarak ulaştırma göstergelerinin kullanılması tercih edilmiştir.

Çakmak v.d. (2005) tarafından yapılan bu çalışmada, 1990 ve 2000 yılı olarak iki dönem şeklinde 73 ilin kültürel açıdan gelişmişliklerini belirlemek amacıyla SPSS 10 programı kullanılarak Ward, tek bağlantı ve k-ortalamlar yöntemleriyle kümeleme analizi uygulanmıştır. Analiz sonucunda, İstanbul ilinin diğer illere göre oldukça farklı bir kültürel konuma sahip olduğu, İzmir ve Ankara’nın genel olarak benzer kültürel durumda

oldukları tespit edilmiştir. Burada bölgesel ve iktisadi gelişmişlik yönünden bu iller arasında görülen farklılığın kültürel açıdan gelişmişliği de etkilediği söylenebilir.

Er'in (2006) çalışmasında 2004 yılında Türkiye'ye gelen yabancıların kullandıkları havayolu, denizyolu, karayolu ve demiryolu taşıt araçları değişkenleri ile 79 ülke ele alınmış ve R programı yardımıyla k-medoid yöntemi kullanılarak kümeleme analizi yapılmıştır. Burada taşıt araçlarına göre birbirine benzer ülkelerin olup olmadığı belirlenmiştir. Analiz sonuçlarına göre Türkiye'ye en çok turist gönderen ülkelerden olan Almanya, Hollanda, Rusya ve İngiltere öne çıkmıştır ve ülkelerin taşıt araçları bakımından kümeleneğinde gelen yabancıların ülkelerinin Türkiye ile olan coğrafik bağlarının taşıt araçlarını kullanımını etkilediği ortaya çıkartılmıştır.

Cengiz ve Öztürk (2012) çalışmasında illerin eğitim performansını belirlemek amacıyla TÜİK verileri esas alınarak Türkiye'nin 81 ili için 8 farklı eğitim göstergesi kullanılmıştır. k-ortalamlar yöntemi SPSS 17 paket programında uygulanmış ve küme sayısı 6 olarak belirlenmiştir. Yapılan analiz sonucunda doğu ve güneydoğu illerinin durumunun diğer illere göre oldukça kötü olduğu ve diğer kümelerden farklı bir konumda yer aldığı tespit edilmiştir. Diğer kümelerde eğitim performansı bakımından coğrafi ve sosyoekonomik sınırlar oluşmamıştır.

Atalay ve Tortum (2013), Türkiye'deki illerin karayolu trafik hareketliliğini analiz etmek amacıyla KGM'den elde edilen yolcu-km, taşıt-km ve ton-km verilerini kullanarak k-ortalamlar yöntemiyle SPSS 20 istatistik programında illeri 6 kümede sınıflandırmıştır. Bu çalışma sonucunda İstanbul ve Ankara en yüksek trafik hareketliliğine sahip iller olarak, Antalya, Bursa, Kocaeli, Konya, İzmir ise bu iki ili takip eden ikinci en yüksek trafik hareketliliğine sahip iller olarak belirlenmiştir.

Klinger v.d. (2013) tarafından yapılan çalışmada, Almanya'da bulunan 44 şehir için ulaşım karakteristiklerini belirlemek amacıyla SPSS istatistik programından faydalanılarak 23 gösterge ile faktör analizi yapılmış, Ward ve k-ortalamlar yöntemleri kullanılarak kümeleme analizleri uygulanmış ve 6 küme tespit edilmiştir. Analiz sonucunda elde edilen kümeler, bisiklet şehirleri kümesi, transit metropoller kümesi, otomobil odaklı şehirler kümesi, çok türlü ulaşım potansiyeli olan yürümeye uygun şehirler kümesi ve transit şehirler kümesidir.

Cabral ve Ramos (2014) çalışmalarında Brezilya'nın 2009 yılındaki 17 liman bölgesinin performans durumunu değerlendirmek amacıyla, 9 farklı değişken ele alınmışlar ve R programı yardımıyla hiyerarşik kümeleme yöntemlerini uygulamışlardır. Analiz sonucunda 3 seviyede sınıflama yapılarak en yüksek performansa sahip limanın Sao Paulo bölgesi Tecon terminalindeki Santos liman bölgesi olduğu ortaya çıkarılmıştır.

Yılmaz ve Uzgören (2014) yaptıkları çalışmada Türkiye'deki illerin bankacılık faaliyetleri bakımından performanslarını incelemek amacıyla 9 adet değişken belirleyerek Minitab 14 ve SPSS programları ile k-ortalamlar yöntemi uygulamışlardır. Analiz sonuçlarına göre İstanbul ve Ankara şehirleri özel sektörün ve kamu sektörünün finans merkezleri olmaları nedeniyle farklı birer grup oluşturdukları gözlemlenmiştir.

Alonso v.d. (2015) tarafından yapılan çalışmada, 2010 yılında 23 Avrupa şehrinin kentsel ulaşım durumunu incelemek amacıyla SPSS programında Ward ve k-ortalamlar yöntemi kullanılarak kümeleme analizi uygulanmıştır. Uygulama sonucunda çevresel etkinlik bakımından sürdürülebilir şehirler, sosyal açıdan sürdürülebilirliği en yüksek şehirler,

ekonomik sürdürülebilirliği olan rekabetçi yapıdaki şehirler ve en az sürdürülebilir küçük nüfuslu şehir grupları tespit edilerek tüm şehirler dört farklı şekilde sınıflandırılmıştır.

Haustein ve Nielsen (2016) tarafından yapılan çalışmada Avrupa'nın ulaşım karakteristiğini belirlemek amacıyla 28 AB ülkesinde 2014 yılı 15 yaş üstündeki yaklaşık 28000 kişiyle görüşme yapılarak bu kişilerin kullandıkları ulaşım araç ve sistemleriyle ilgili sorular sorulmuştur. Bu görüşmeler sonucu toplanan veriler SPSS 21.0 programında k-ortalamlar yöntemiyle analiz edilmiştir. Bu analiz sonucunda, konforu ön planda tutan ulaşım aracı kullanıcılarının yoğunlukta olduğu, bisiklet kullanımının yoğun olduğu ve tüm seyahat imkânlarında çevreci bilincin yüksek olduğu ülkeler ayrı kümelerde sınıflandırılmıştır.

Nezerenko v.d. (2017) ulaştırma alanında yapılabilecek ortak projeler için ulaştırmada benzer karakteristiklere sahip ülkeleri tespit edip sınıflandırmak amacıyla Baltık denizine komşu 9 ülkede 2004-2011 dönemi arasındaki kara, deniz ve demiryolu verileri alınarak SPSS programında Ward yöntemi ile kümeleme analizi yapılmıştır. Analiz sonuçlarına göre özellikle Almanya ve Polonya'nın tüm ulaştırma türleri dikkate alındığında diğer ülkelere göre ciddi farklılıklarının olduğu tespit edilmiştir.

Karaoğlan (2018) çalışmasında sosyoekonomik açıdan benzerlik gösteren illeri belirlemek amacıyla TÜİK' ten 55 değişkenden oluşan 2002, 2008 ve 2013 yılı verileri alınarak Türkiye'deki 81 il için SPSS programında kümeleme analizi uygulanmıştır. Analiz sonucunda özellikle iktisadi gelişmişliği benzer olan illerin sosyoekonomik açıdan da benzer oldukları ortaya çıkarılmıştır. Ayrıca İstanbul ilinin dikkat çekici bir şekilde diğer illerden ayrı olarak tek bir kümede yer aldığı görülmektedir.

Bulut (2019), Türkiye'de yaşam endeksine göre illerin sınıflandırılması amacıyla yaptığı çalışmada TÜİK tarafından oluşturulan 11 adet göstereyi alarak EM ve k-ortalamlar kümeleme algoritmalarını kullanmıştır. Analiz sonucunda özellikle doğu ve güneydoğu illerinin diğer illere göre aynı sınıfta ve birbirlerine benzer durumda olduğu gözlenmiştir. Çalışmada istatistik alanında yaygın olarak kullanılan R programından yararlanılmıştır.

Çelik v.d. (2019) tarafından yapılan çalışmada Türkiye'de 2016 yılında 81 ile ait 11 tarım değişkeni kullanılarak benzer illerin sınıflandırılması amaçlanmıştır. SPSS 23 programında k-ortalamlar yöntemi ile kümeleme analizi yapılmıştır. Yapılan analiz sonucunda tarım göstergelerine göre birbirine en çok benzeyen illerin Rize ve Yalova, en az benzeyen illerin ise Konya ve Adana olduğu görülmüştür.

Gitto ve Mancuso (2019) çalışmalarında havaalanlarının marka değerlerini belirlemek amacıyla 2016-2018 yılları arasındaki sosyal medya verileri ile 46 ülkeden 118 havaalanını ele alarak R programında k-ortalamlar yöntemiyle kümeleme analizi yapmışlardır. Analiz sonucunda elde edilen 5 küme içinde kuzey Amerika havaalanları kümesi çevresel özelliği ile, Avustralya ve İngiltere havaalanları kümesi ise engelli dostu özelliği ile ön plana çıkmışlardır.

Wang ve Pham (2020) Vietnam'da 2015 ve 2016 yılları arasında uluslararası niteliğe sahip 28 havaalanını bazı performans göstergelerine göre Ward yöntemi ile sınıflandırmıştır. Analiz sonucunda oluşturulan kümeler ile havaalanlarının performans değerlendirmeleri yapılarak tespit edilen eksikliklerin tamamlanması ve iyileştirme çalışmalarının yapılması yönünde önerilerde bulunulmuştur.

İncekırık v.d. (2021) çalışmasında Avustralya Sağlık ve Refah Enstitüsü kayıtlarındaki 57 kanser türü için erkeklerde ve kadınlarda yaş gruplarına göre 1982-2016 yılları arasındaki veri setlerine kümeleme analizi yöntemleri uygulanmış; ayrıca uygun kümeleme yönteminin ve küme sayısının belirlenmesinde sırasıyla kofenetik korelasyon katsayılarından ve 26 adet küme geçerlilik indeksinden yararlanılmıştır. Kanser gruplarının en iyi şekilde ayrılmasına yol açan 3 farklı yaş grubu karakteristiği ile oluşturulan 4 farklı veri setinde kanser türlerinin cinsiyete göre belirlenen yaş gruplarındaki dağılımı gözlemlenmiş ve kanserlerin ilgili yaş gruplarında ortaya çıkış frekanslarına bağlı olarak göstermiş oldukları kümelenme eğilimleri araştırılmıştır.

Yöntem

Çalışmada Türkiye'deki illere ait 2004-2018 yılları arasındaki ulaşım ile ilgili bazı kurumlarında bulunan karayolu, denizyolu, demiryolu, havayolu ve boru hattı verilerinden faydalanılmıştır. Ulaştırma faaliyetlerine bağlı göstergelerin yer aldığı bu veriler Türkiye İstatistik Kurumu (T.Ü.İ.K.), Türkiye Cumhuriyeti Devlet Demiryolları (T.C.D.D.), Boru Hatları ile Petrol Taşıma Anonim Şirketi (BOTAŞ), Deniz Ticareti Genel Müdürlüğü, Devlet Hava Meydanları İşletmesi (D.H.M.İ.), Karayolları Genel Müdürlüğü (K.G.M.) kurumlarının internet ortamlarından alınmıştır. Elde edilen veriler kullanılarak uygulanan klasik kümeleme analizi yöntemleri için beş adet küme sayısı belirlenerek 81 ilin sınıflandırılması yapılmıştır. Birçok kümeleme analizi yöntemi arasında bu çalışma için tek bağlantı yöntemi, tam bağlantı yöntemi, ortalama bağlantı yöntemi, Ward yöntemi, medyan yöntemi ve ağırlık merkezi yöntemiyle k-ortalamlar yönteminden faydalanılmıştır. İlerleyen aşamalarda hiyerarşik yöntemler arasında kümeleme analizi yöntemlerinin veride uygulama başarısını ölçen kofenetik korelasyon katsayısı değerleri karşılaştırıldıktan sonra ortaya çıkan en başarılı yöntem olarak tek bağlantı yöntemi sonuçlarıyla hiyerarşik olmayan yöntemler arasında en sık kullanılan popüler yöntem olan k-ortalamlar yönteminin sonuçları çalışmada esas alınmıştır. Uygulanan kümeleme analizi yöntemlerinin tümünde, açık kaynak kodlu R-3.5.1 programı ile factoextra, readxl, scales, NbClust, moments, scatterplot3d, psych, cluster ve stats paketleri kullanılmıştır.

Verilerin Toplanması ve Analizi

Mevcut çalışmada, öncelikle ulaşım türlerinin çerçevesi çizilerek; karayolu, denizyolu, demiryolu, havayolu ve boru hatları ana başlıkları içindeki göstergelerden Türkiye'deki 81 ile ait veriler elde edilerek 81x40 boyutlu bir veri seti oluşturulmuştur. Daha sonra, oluşturulan bu veri setiyle ilgili bazı tanımlayıcı istatistikler elde edilmiştir. Verilerdeki kümelenme yapısını görsel olarak inceleyebilmek amacıyla ana bileşenler analizi kullanılarak ana bileşen skorları elde edilmiş ve bu ana bileşen skorları ile iki ve üç boyutlu grafikler oluşturulmuştur.

Yöntem bölümünde adı geçen kurumlardan alınan verilere ait değişkenler Tablo 1'de gösterilmiştir. Bu göstergelere ait bazı tanımlayıcı istatistikler de Tablo 2'de verilmiştir.

Tablo 1. Kullanılan Değişkenler

V1	İniş-Kalkış Yapan Uçak / Türk	V21	Toplam İhracat Yük Miktarı
V2	İniş-Kalkış Yapan Uçak / Yabancı	V22	Otoyol Taşıt-Km Değeri
V3	İniş-Kalkış Yapan Uçak / Diğer	V23	Devlet Yolu Taşıt-Km Değeri
V4	Yolcu Sayısı / İç Hat Gelen	V24	Otoyol Yolcu-Km Değeri
V5	Yolcu Sayısı / İç Hat Giden	V25	Devlet Yolu-Yolcu-Km-Değeri
V6	Yolcu Sayısı / Dış Hat Gelen	V26	Otoyol Yük-Km Değeri
V7	Yolcu Sayısı / Dış Hat Giden	V27	Devlet Yolu-Ton-Km Değeri
V8	Taşınan Yük / İç Hat Gelen	V28	Yol Uzunlukları: Otoyol (Km)
V9	Taşınan Yük / İç Hat Giden	V29	Yol Uzunlukları: İl ve Devlet Yolu (Km)
V10	Taşınan Yük / Dış Hat Gelen	V30	Yol Uzunlukları: Köy Yolu (Km)
V11	Taşınan Yük / Dış Hat Giden	V31	Kara Taşıtları Sayısı: Otobüs
V12	Havalimanı Sayısı	V32	Kara Taşıtları Sayısı: Minibüs
V13	Sahillerde Mevcut Liman Başkanlığı Sayısı	V33	Kara Taşıtları Sayısı: Otomobil
V14	Toplam İhracat Konteyner Miktarı (Teu)	V34	Kara Taşıtları Sayısı: Kamyonet
V15	Toplam İthalat Konteyner Miktarı (Teu)	V35	Kara Taşıtları Sayısı: Kamyon
V16	Toplam Kabotaj Konteyner Miktarı (Teu)	V36	Kara Taşıtları Sayısı: Motosiklet
V17	Toplam Transit Konteyner Miktarı (Teu)	V37	Kara Taşıtları Sayısı: Özel Amaçlı Taşıtlar
V18	Toplam İthalat Yük Miktarı	V38	Kara Taşıtları Sayısı: Traktör
V19	Toplam Kabotaj Yük Miktarı	V39	Yol Uzunlukları: Demir Yolu (Km)
V20	Toplam Transit Yük Miktarı	V40	Boru Hattı Sayısı (Petrol, Doğalgaz, Maden)

Veri setinde, 2004-2018 yılları arasında ulaştırma göstergelerinin beklenen değerleri alınmıştır. Uzaklık ölçüsü olarak Öklid uzaklık ölçüsü (1) kullanılmış ve farklı ölçü birimleri barındıran göstergelerin bir arada bulunmasından ötürü 0-1 arası Min-Max yöntemi (2) kullanılarak standartlaştırma yapılmıştır (Alpar, 2017).

$$d(x_i, x_k) = \sqrt{\sum_{j=0}^n w_j (x_{ij} - x_{kj})^2} \quad (1)$$

$$Z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

Kümeleme Analizi, araştırmacıların objeleri (gözlemleri) bir dizi özelliklere göre sınıflayarak birtakım kümeler elde edilmesini sağlamaktadır. Bu sınıflama işlemi sonunda elde edilen kümelerde, kümelerin sahip olduğu gözlemlerde yüksek homojenlik, bu kümelerin arasında ise yüksek heterojenlikte grupların elde edilmesi beklenmektedir (Hair ve diğerleri, 2014).

Klasik anlayışa dayalı kümeleme analizi yöntemleri genel olarak hiyerarşik (aşamalı) ve hiyerarşik (aşamalı) olmayan yöntemler olarak iki temel gruba ayrılmaktadır. Çalışmada, hiyerarşik yöntemlerden tek bağlantı, tam bağlantı, ortalama bağlantı, Ward yöntemi, medyan yöntemi ve ağırlık merkezi yöntemi, hiyerarşik olmayan yöntemlerden ise k-

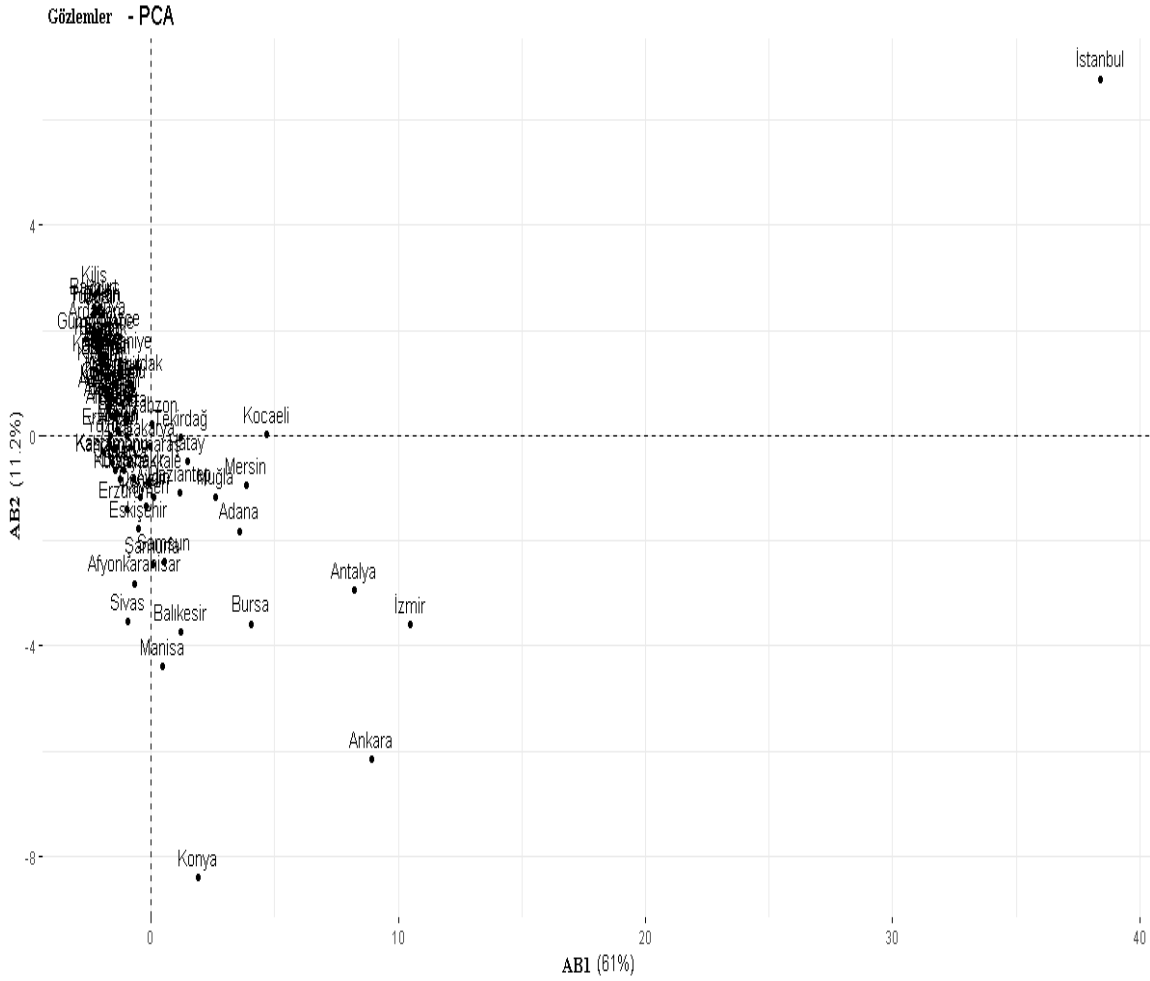
Türkiye'deki İllerin Ulaştırma Göstergelerine Göre Kümeleme Analizi Yöntemleriyle Sınıflandırılması

ortalamalar yöntemi kullanılmıştır. Kullanılan veri seti ile ilgili bazı tanımlayıcı istatistikler Tablo 2'de verilmiştir.

Tablo 2. Kullanılan Veri Seti ile Hesaplanan Bazı Tanımlayıcı İstatistikler

Değişkenler	Ortalama	S. Sapma	Uzaklık	Değişkenler	Ortalama	S. Sapma	Uzaklık
V1	8.936,34	41.902,38	362.624,7	V21	591.843,5	4.348.708	38.733.467
V2	1.897,03	9.572,05	61.724,33	V22	197.221,9	633.567,5	5.001.385
V3	1.672,53	3.971,549	25.317,67	V23	778.652,5	691.379	2.979.701
V4	377.746,9	1.395.694	11.462.625	V24	629.647,5	1.968.889	15.428.644
V5	376.822,8	1.391.248	11.418.987	V25	2.073.452	1.881.367	8.418.310
V6	362.507,1	2.004.288	15.732.527	V26	553.325,8	1.568.221	11.763.577
V7	366.499,4	2.032.342	15.979.628	V27	1.858.332	1.562.660	7.770.214
V8	3.722,67	14.265,15	118.408,2	V28	104.911,2	259.291,2	2.062.190
V9	3.722,94	14.303,54	118.698,4	V29	779,36	452,1067	2.917,8
V10	9.542,78	62.665,75	547.864	V30	4.984,484	8.231,503	68.826,73
V11	10.772,17	71.671,38	628.794,5	V31	26,26	54	243,13
V12	0,60	0,56	2,07	V32	2.514,80	5.983,50	48.963,93
V13	0,62	1,16	4,33	V33	3.176,39	1.681,75	9.100,93
V14	32.779,03	137.667,7	982.687,9	V34	31.677,55	65.651,27	525.908,7
V15	33.297,19	143.629,8	1.073.160	V35	9.287,723	16.336,03	126.644,3
V16	4.658,42	18.357,16	141.052,2	V36	29.919,86	46.920,42	218.377,9
V17	7.707,10	55.139,14	484.274,4	V37	493,78	927,07	7.154,2
V18	1.018.212	2.919.289	17.142.944	V38	18.573,68	16.272,84	72.480,13
V19	2.156.514	6.448.652	38.521.913	V39	116,90	127,86	551,6
V20	541.919,5	1.513.222	8.270.949	V40	1,09	0,50	2

Çalışmanın yüksek doğrulukla tamamlanabilmesi için veri seti yapısının görüntülenmesi oldukça önemlidir. Kümeleme yapısının görüntülenebilmesi amacıyla çalışmada grafiklerden yoğun şekilde faydalanılmıştır. Burada ilk başvuru yöntem, gözlemlerin konumlarının görüntülenebilmesi amacıyla iki boyutlu ve üç boyutlu saçılım grafiklerinin oluşturulmasıdır. İki veya üç değişkenli verilerin bu yöntemle veri yapılarının tam doğrulukta görüntülenebilmesi mümkün olsa da daha fazla değişkenin söz konusu olduğu veri setlerinde ana bileşenleri kullanarak belli doğrulukta bu grafikleri oluşturabilmek mümkün olmaktadır. Nitekim, bu çalışmadaki 40 adet değişkene sahip veri setinde ana bileşenler analizi uygulanmıştır. Sonuç olarak, en büyük iki bileşenden veri setinin toplam varyansının %72'sini, üçüncü bileşenle birlikte toplam varyansının %79'unu açıklayabilecek şekilde iki ve üç boyutlu saçılım grafikleri elde edilmiştir. Bunlar Şekil 1 ve Şekil 2'de incelenebilir.

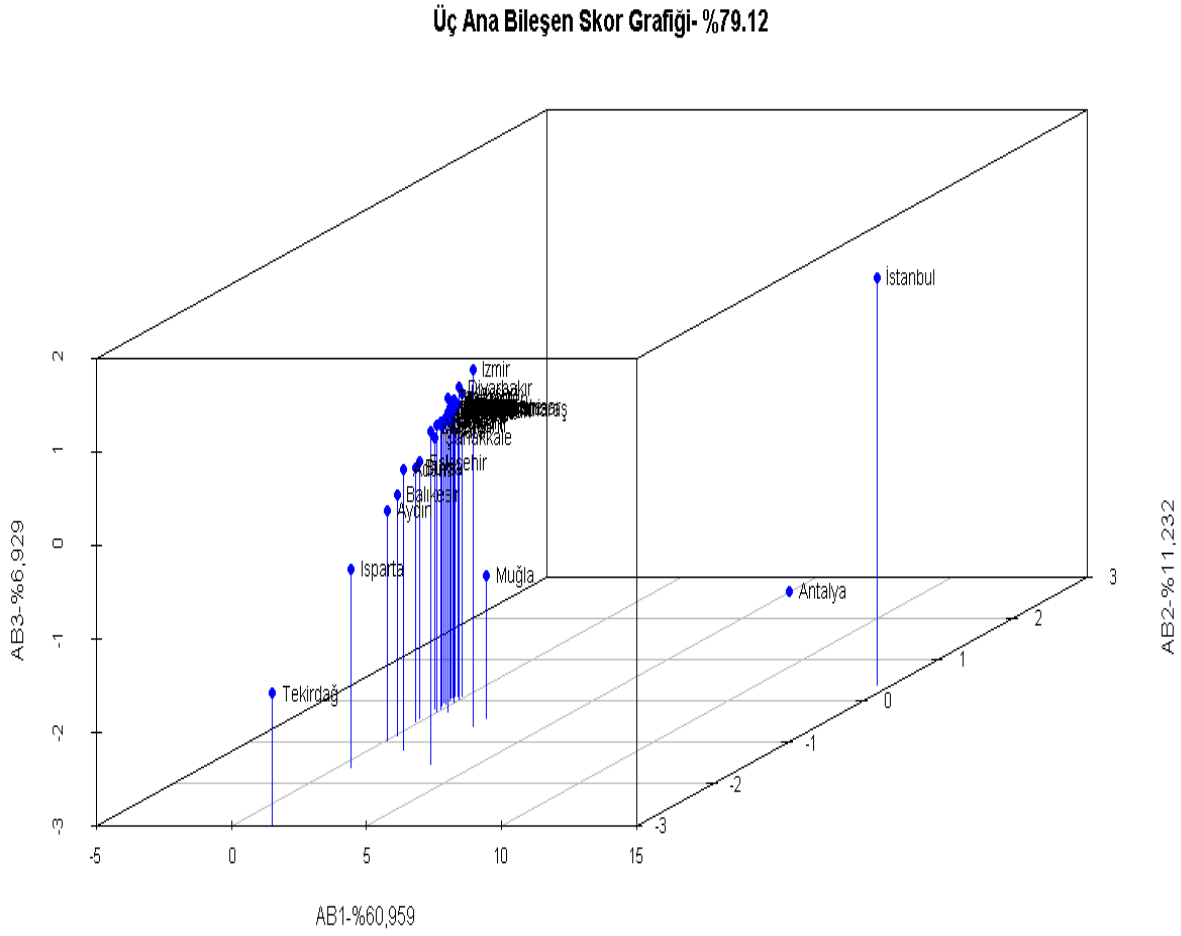


Şekil 1. İki boyutlu ana bileşen skor grafiği

Bu grafiklerde illerin iki boyutlu ve üç boyutlu ana bileşen uzayındaki saçılımı incelendiğinde, İstanbul, Ankara, İzmir, Antalya gibi illerin diğer illerden uzak ve düzensiz bir konumda aykırı gözlem gibi davrandıkları ve gözle görülebilecek kadar keskin şekilde ayrılan büyük kümeler olmadığı gözlemlenmektedir. Tablo 3'te, veri setine ait karakteristikler için özdeğer ve önem değerleri verilmiştir. Ana bileşen skorları yardımıyla iller verisi incelendiğinde Tablo 3'ten de ilk iki bileşenin toplam varyansın %72,191'i, ilk 3 bileşenin toplam varyansın %79,12'sini temsil ettiği bu tabloda net olarak görülebilir.

Tablo 3. Veri Setine Ait Karakteristikler İçin Özdeğer ve Önem Değerleri

Ana Bileşen	Özdeğer	Açıklanan Varyans Yüzdesi	Kümülatif Toplam
1	24,384	60,959	60,959
2	4,493	11,232	72,191
3	2,771	6,929	79,12



Şekil 2. Üç boyutlu ana bileşen skor grafiği

Şekil 3'te, İllerin birbiriyle uzaklık ilişkisini gösteren Öklid uzaklık matrisi ısı grafiği gösterilmiştir. Bu grafikte, diyagonal bölgelerdeki koyu mavi renk minimum uzaklık ilişkisi olan 0 değerine (kendisi) eşittir. Şekil 3 dikkatli bir biçimde incelendiğinde, ilk satır ve sütundaki İstanbul başta olmak üzere birçok büyükşehir gözlemlerinin illerle ilişkisinin kırmızı ve kırmızıya yakın renklerle ifade edildiği, diğer illerin ise maviye yakın renklerle ifade edildiği görülmektedir. Bundan dolayı, İstanbul başta olmak üzere diğer büyükşehirlerle diğer illerin aralarında daha az bir benzerliğe sahip olduğu sonucuna varılabilir.

korelasyon katsayı değeri ne kadar yüksek çıkarsa o yonteme ait kümeleme analizi sonuçlarının o kadar iyi olduğu kabul edilir.

Aşağıdaki formül uygulandığında elde edilecek değer kofenetik korelasyon katsayısıdır (3). Dendrogram birleşim değeri yüksekliklerinin t , uzaklık değerlerinin x olarak gösterildiği bir veri grubunda x ve t tekil simgeler $x(i,j)$ ve $t(i,j)$ değerlerinin genel ortalamalarını ifade etmektedir (Saraçlı ve diğerleri, 2013).

$$c = \frac{\sum_{i<j}(x(i,j) - x)(t(i,j) - t)}{\sqrt{[\sum_{i<j}(x(i,j) - x)^2][\sum_{i<j}(t(i,j) - t)^2]}} \quad (3)$$

Tablo 4. Hiyerarşik Kümeleme Yöntemlerindeki Kofenetik Korelasyon Katsayıları

Yöntem	Kofenetik Korelasyon Katsayısı
Tek Bağlantı Yöntemi	<u>0.968526</u>
Tam Bağlantı Yöntemi	0.917049
Ortalama Bağlantı Yöntemi	0.965494
Ağırlık Merkezi Yöntemi	0.964410
Ward Bağlantı Yöntemi	0.525156
Medyan Yöntemi	0.955254

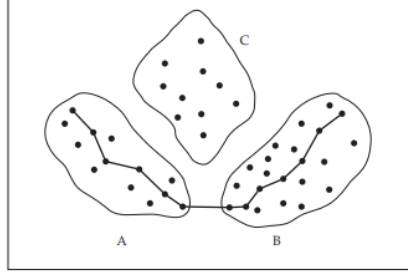
Tablo 4'te gösterildiği gibi hiyerarşik yöntemler arasında en yüksek kofenetik korelasyon katsayısı değeri 0.968526 ile tek bağlantı yöntemine ait çıkmıştır. Bu katsayı sonucuna göre, tek bağlantı yönteminin bu çalışmaya ait veri seti için en başarılı hiyerarşik kümeleme analizi yöntemi olduğu söylenebilir. Bu sebeple, çalışma sonuçlarının değerlendirilmesi ve yorumlanması aşamasında bu yöntemin bulguları ve yapısal olarak kullanışlı olan k-ortalama yöntemlerinin bulguları ağırlıklı olarak dikkate alınmıştır.

Tek bağlantı yönteminde kümeleme analizi sonuçları, bir dendrogram veya ağaç diyagramı şeklinde grafiksel olarak gösterilebilir. Ağaç diyagramındaki dallar kümeleri temsil etmektedir. Bu dallar, bir uzaklık (veya benzerlik) ekseni boyunca belirli konumlarda oluşan düğümlerle birleşmektedir. Tek bağlantı yöntemini uygularken ilk olarak, elde edilen uzaklık matrisindeki en küçük $D = \{dik\}$ bulunmalı ve (UV) kümesini elde etmek için U ve V'ye karşılık gelen objeler birleştirilmelidir. (UV) ifadesi ile diğer herhangi bir W kümesi arasındaki uzaklıklar aşağıdaki şekilde hesaplanır (Johnson ve Wichern, 2007):

$$d_{(uv)w} = \min\{d_{uw}, d_{vw}\} \quad (4)$$

Bu yöntem, kullanılan veri setine göre avantajlı olduğu kadar dezavantajlı da olabilmektedir. Şekil 4'te bunun bir örneği görülebilir. Şekilde üç küme (A, B ve C) birleştirilecektir. Her kümede yalnızca en yakın noktalara odaklanan tek bağlantı yöntemi, kümelerin en uç noktalarındaki kısa mesafeleri nedeniyle A ve B kümelerini birbirine bağlayacaktır. A ve B

kümelerinin birleştirilmesi, C kümesini çevreleyen yeni bir küme oluşturur. Bununla birlikte, kümelerin homojenliği için çabalıncaksa, C kümesinin A veya B kümesi ile birleşmesi daha makuldür. Bu şekil, tek bağlantı yönteminin temel dezavantajını ortaya çıkarmaktadır (Grimm ve Yarnold, 1995).



Berzer Olmayan A ve B Kümelerinin Tek Bağlantıda Birleşiminin Örneği

Şekil 4. Tek bağlantı yöntemi

k-ortalamlar terimi ilk defa Macquenn tarafından 1967 yılında tanımlanıp kullanılmasına rağmen bu yöntemin uygulanma fikri 1956 yılına kadar uzanmaktadır. Zaman içinde en popüler kümeleme analizi algoritmalarından birisi haline gelmiştir. Hiyerarşik olmayan kümeleme analizi kavramında ilk akla gelen yöntemlerden birisi haline gelmiş ve neredeyse tüm kümeleme analizi paket yazılımlarında da yerini almıştır. Fonksiyonu ise aşağıdaki gibidir (Macqueen, 1967):

$$\arg \min x_1, \dots, x_k = \sum_{i=0}^n \min_{j=1,2,\dots,k} \|z_i - x_j\|^2 \quad (5)$$

k-ortalamlar yönteminin amacı, diğer kümeleme analizi yöntemlerindeki gibi grup içi homojenliği ve gruplar arası heterojenliği en büyük yapmaktır (Bulut, 2018). Verileri kullanıcı tarafından belirlenen k sayıda kümeye bölerek çalışır ve ardından bazı sayısal ölçütler yerine getirilinceye kadar gözlemleri yinelemeli olarak kümelere yeniden atar. k-ortalamlar ölçütü, kümelerdeki gözlem uzaklığını en aza indirmeye ve kümeler arasındaki uzaklığı en yükseğe çıkarma ile ilgili bir amacı belirler. (Hair ve diğerleri, 2014).

Araştırmacıların küme sayısını belirlemek için kullandığı belli bir yöntem olmamakla birlikte çok sayıda yöntem mevcuttur. Bu yöntemler arasında küme geçerlilik indeksleri de bulunmaktadır. Küme geçerlilik indeksleri içsel (internal) indeksler ve dışsal (external) indeksler olmak üzere iki ana kategoride sınıflandırılabilir. Bu indeksler arasındaki temel fark, küme geçerliliği için harici bilgilerin kullanılıp kullanılmadığıdır (Aggarwal ve Reddy, 2014).

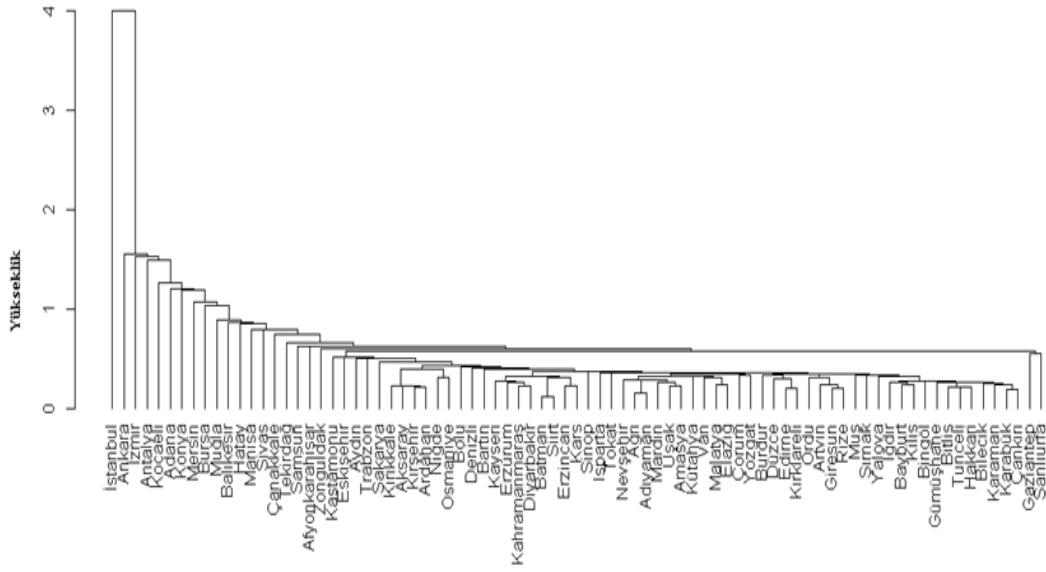
Bu çalışmada, küme sayısının belirlenebilmesi için küme geçerlilik indeksleri ile oluşturulan iki ve üç boyutlu grafiklerden yararlanılmıştır. R 3.5.1 programında kullanılabilen NbClust paketinde yer alan 13 adet içsel geçerlilik indeksi sonuçları Tablo 5'te gösterilmiştir. Şekil 1 ve Şekil 2'deki gözlemlerin dağılışı dikkate alınarak indekslerin alt limiti için k=5, üst limiti için k=12 alınmıştır. Tabloda yer alan indeks değerleri incelendiğinde, uygun küme sayısının çoğunlukla 5'i işaret ettiği görülmektedir.

Tablo 5. Tek Bağlantı ve k-ortalamlar Yönteminde Uygun Küme Sayısının Belirlenmesi

Tek Bağlantı Yöntemi			k-ortalamlar Yöntemi		
	Uygun Küme Sayısı	Kritik Değer		Uygun Küme Sayısı	Kritik Değer
<i>Cal. Harb.</i>	5	18,3491	<i>Cal. Harb.</i>	5	28,0122
CCC	5	2,3175	CCC	11	10,5674
Cindex	7	0,3458	Cindex	8	0,2064
<i>DaviesBoul.</i>	5	0,2945	DaviesBoul.	11	1,1178
<i>Silhouette</i>	5	0,5652	Silhouette	11	0,2175
<i>Duda</i>	5	1,016	<i>Duda</i>	5	0,9796
<i>PseudoT²</i>	5	-0,1204	<i>PseudoT²</i>	5	0,6234
<i>Beale</i>	5	-0,0449	<i>Beale</i>	5	0,5625
<i>Ratkowsky</i>	5	0,341	<i>Ratkowsky</i>	5	0,3595
<i>PtBiserial</i>	5	0,7411	PtBiserial	6	0,3636
<i>McClain</i>	5	0,0335	McClain	6	1,2673
Dunn	8	0,5984	Dunn	10	0,1791
<i>SDindex</i>	5	1,3961	SDindex	11	9,3149

Bulgular

Beş adet küme için tek bağlantı yöntemi sonucunda elde edilen bulgular Şekil 5, Tablo 5 ve Şekil 6'da gösterilmiştir.



Şekil 5. Tek bağlantı yöntemi için kümeleme analizi dendrogramı

Şekil 5 dikkate alındığında beş küme için gözlemler dendrogram çerçevesinde ayrıldığında Tablo 5'te yer alan sonuçlar elde edilmiştir.

Tablo 5. Tek Bağlantı Yöntemine Göre Belirlenen Kümeler

Kümelere	İller
1.Küme	Adana, Afyonkarahisar, Aksaray, Amasya, Artvin, Aydın, Balıkesir, Bartın, Bayburt, Bilecik, Bolu, Burdur, Bursa, Çanakkale, Çankırı, Çorum, Denizli, Düzce, Edirne, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Hatay, Isparta, Kahramanmaraş, Karabük, Karaman, Kastamonu, Kayseri, Kırıkkale, Kırklareli, Kırşehir, Kocaeli, Konya, Kütahya, Malatya, Manisa, Mersin, Muğla, Nevşehir, Niğde, Ordu, Rize, Sakarya, Samsun, Sinop, Sivas, Tekirdağ, Tokat, Trabzon, Tunceli, Uşak, Yalova, Yozgat, Zonguldak, Adıyaman, Ağrı, Ardahan, Batman, Bingöl, Bitlis, Diyarbakır, Hakkari, Iğdır, Kars, Kilis, Mardin, Muş, Osmaniye, Siirt, Şanlıurfa, Şırnak, Van
2.Küme	İzmir
3.Küme	Antalya
4.Küme	Ankara
5.Küme	İstanbul

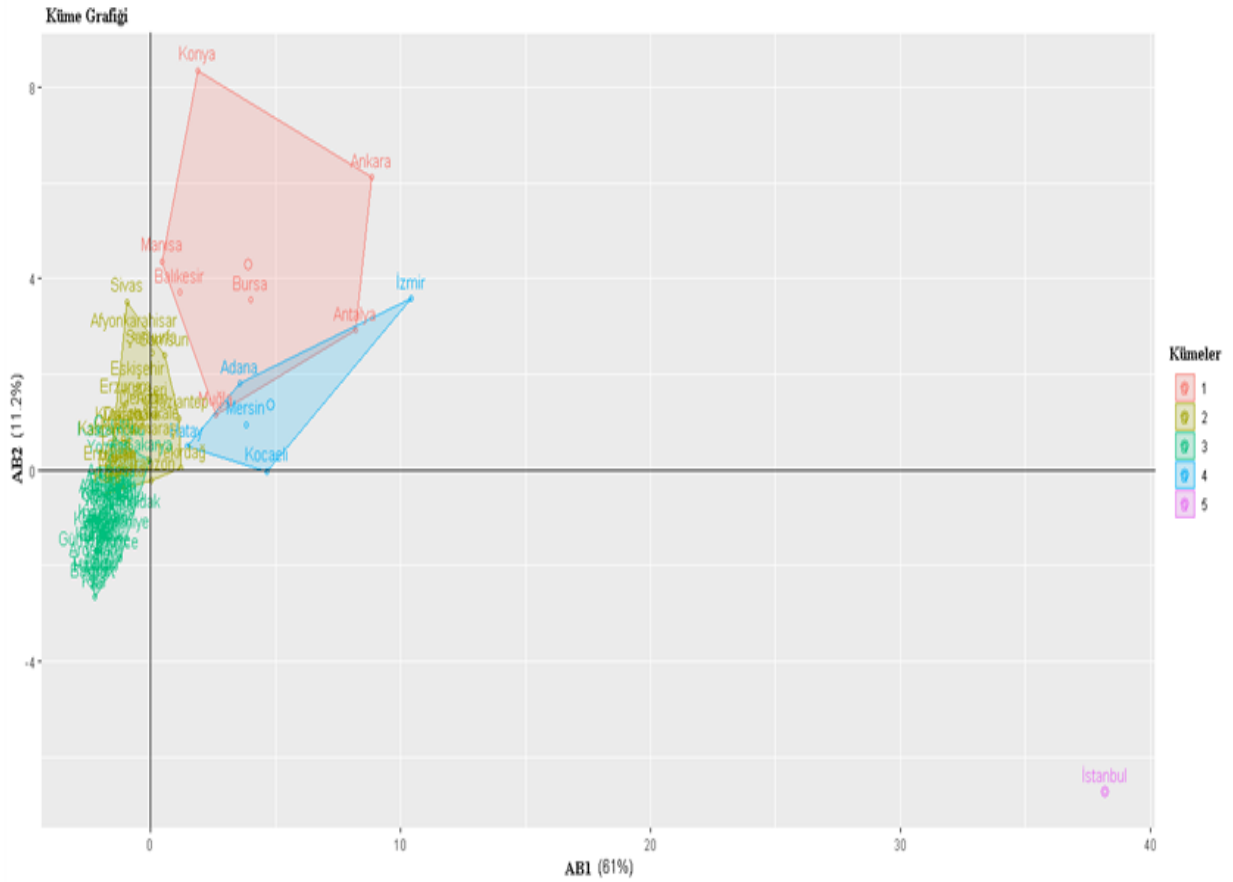
**Şekil 6.** Tek bağlantı yöntemi sonuçlarına göre Türkiye il haritası

Tablo 5'te görüldüğü gibi tek bağlantı yöntemi kullanılarak kümeleme analizi uygulandığında, belirlenen 5 adet kümede büyükşehir unvanına sahip 4 ilin birer kümeyi sahiplendiği gözlemlenmekte ve diğer illerin birinci kümede toplandığı dendrogram sonucu ortaya çıkmaktadır. Aykırı gözlemlerden olan İstanbul, Ankara, İzmir ve Antalya diğer dört

Türkiye'deki İllerin Ulaştırma Göstergelerine Göre Kümeleme Analizi Yöntemleriyle Sınıflandırılması

kümeye tek bir üye olacak şekilde yerleşmiştir. Tek bağlantı yönteminin kofenetik korelasyon katsayısındaki değerinin (0.9685) diğer hiyerarşik kümeleme algoritmalarına göre en yüksek seviyede olması sebebiyle veri setini en başarılı kümeleyen yöntem olduğu söylenebilir. 1. Küme'deki iller dışındaki illerin bu şekilde kümelenmesi Türkiye'nin en büyük beş ilinden 4'ünün yine söz konusu iller olması ve dolayısıyla ulaşım faaliyetlerinin bu duruma yansımalarıyla açıklanabilir. Ayrıca, tek bağlantı yönteminin yapısının da bu duruma yol açtığı söylenebilir. Şekil 6, Tablo 5'te yer alan bilgilerin daha net görülebilmesi açısından Türkiye haritası üzerinde tüm illerin bulunduğu kümeleri göstermektedir.

k=5 olarak belirlendiği k-ortalamlar yöntemi bulguları ise Tablo 6, Şekil 7 ve Şekil 8'de gösterilmektedir.



Şekil 7. k-ortalamlar yöntemi küme grafiği (k=5)

Şekil 1'de yer alan ilk iki ana bileşen skorları grafiğinin görsel olarak incelenmesi ve geçerlilik indekslerinin göz önünde bulundurulması sonucu k değerinin 5 olarak alınıp uygulandığı bu yöntem sonucunda Tablo 6'da görüldüğü gibi 1. Küme için Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa, Muğla illeri olmak üzere 8 il yer almaktadır. Bu illerin, ulaşım faaliyetleri ve sistemleri bakımından oldukça işlek ve gelişmiş bir ulaşım ağına sahip oldukları söylenebilir. Bununla birlikte, bu kümedeki bazı illerin önemli ulaşım göstergeleri çok zayıf veya olmamasına rağmen yine de sahip oldukları diğer ulaşım göstergelerinin etkisiyle kendilerine özel farklı ve gelişmiş konumda bu sınıfa dâhil oldukları ifade edilebilir.

Tablo 6. k-ortalamalar Yöntemine Göre Belirlenen Kümeler

Kümeler	İller
1.Küme	Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa, Muğla
2.Küme	Adana, İzmir, Kocaeli, Hatay, Mersin
3.Küme	Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak, Van
4.Küme	Adıyaman, Ağrı, Aksaray, Amasya, Ardahan, Artvin, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çankırı, Çorum, Düzce, Edirne, Giresun, Gümüşhane, Hakkari, Iğdır, Karabük, Karaman, Kastamonu, Kırıkkale, Kırklareli, Kırşehir, Kilis, Muş, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Siirt, Sinop, Şırnak, Tunceli, Yalova, Yozgat, Zonguldak
5.Küme	İstanbul



Şekil 8. k-ortalamalar yöntemi sonuçlarına göre Türkiye il haritası (k=5)

2. Küme’de yer alan illerin, sanayisi ve limanları gelişmiş olan ve ithalat ve ihracat faaliyetlerinin odak noktalarına sahip kıyı illeri oldukları görülmektedir. 3. Küme’de ise birinci, ikinci ve beşinci kümeler kadar yoğun ve gelişmiş bir ulaşım faaliyeti görülüyor olsa da bu gruptaki diğer illerin de zayıf olmayan ve gelişmeye müsait iller olduğu sonucu çıkartılabilir. Bu kümede yer alan illerin, önemli bir demiryolu hattı üzerinde bulunması veya önemli bir liman kenti olma özelliği göstermesi dikkat çekicidir. Örneğin, Doğu Ekspresi ve bu ekspresin olduğu bazı önemli boru hatlarının geçtiği iller bu küme içinde yer

almaktadır. Önemli otoyollara ve demiryolu ağına sahip İç Ege ve Aydın illeri de bu küme içinde yer almıştır. Ayrıca, önemli liman bölgeleri olan Samsun, Trabzon, Çanakkale, Tekirdağ da bu kümededir. 3. Küme'de, Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak, Van illeri yer almaktadır. 4. Küme'deki illerin ise genellikle içinde bulunduğu coğrafi sebepler ve dezavantajlı konumları nedeniyle yeterli ulaşım faaliyet ve sistemlerine sahip olmadığı söylenebilir. Bu kümenin dahil olduğu iller genellikle coğrafi şartların zorladığı Karadeniz Bölgesi, Doğu Anadolu, Kuzey Marmara ve Toros Dağları'nın yüksek olduğu noktalarda toplanmışlardır. Batı Karadeniz ve Ankara'nın doğusundaki bazı illerin bu kümeye katılması genel görüntüyü bozsa da bu gruptaki illerin ulaşım faaliyetlerinin zayıflığına işaret ettiğini göstermektedir. Tablo 6'da, 4. Küme'nin gözlemlerini oluşturan iller ise Adıyaman, Ağrı, Aksaray, Amasya, Ardahan, Artvin, Bartın, Batman, Bayburt, Bilecik, Bingöl, Bitlis, Bolu, Burdur, Çankırı, Çorum, Düzce, Edirne, Giresun, Gümüşhane, Hakkari, Iğdır, Karabük, Karaman, Kastamonu, Kırıkkale, Kırklareli, Kırşehir, Kilis, Muş, Nevşehir, Niğde, Ordu, Osmaniye, Rize, Sakarya, Siirt, Sinop, Şırnak, Tunceli, Yalova, Yozgat, Zonguldak'tır.

Sonuç ve Öneriler

Kümeleme analiziyle Türkiye'deki illerin ulaştırma faaliyetlerini oluşturan göstergeler kullanılarak kümeleme yapısını belirlemenin ne derece kullanışlı olabileceği bu çalışmayla ortaya konulmaya çalışılmıştır. Bir bölgedeki ulaştırma faaliyetlerine bağlı göstergelerin sosyal, kültürel, ekonomik, askeri alanlarla doğrudan ilişkisi bulunmaktadır. Modern ve yaygın ulaşım sistemine sahip bir bölgede, söz konusu bu alanlarda da olumlu bir tablonun beraberinde gelmesi kaçınılmazdır.

Bu çalışmada, Türkiye'deki illerin durumunun ulaşım faaliyetlerine bağlı göstergelerine göre yorumlanarak doğru şekilde değerlendirilmesi amaçlanmıştır. Bu amaçla, Türkiye'deki 81 ilin 2004-2018 yılları arasındaki çeşitli kamu kurumlarında bulunan kırk adet karayolu, denizyolu, demiryolu, havayolu ve boru hattı verilerine ait ulaştırma göstergelerinden faydalanılmıştır. 81x40 boyutunda bir veri seti elde edilmiştir. Kümeleme yapısı grafiklerle ortaya konulduktan sonra hiyerarşik kümeleme analizi yöntemlerinden altısı ve hiyerarşik olmayan kümeleme yöntemlerinden ikisi kullanılmıştır. Hiyerarşik yöntemlerde, kümeleme analizi yöntemlerindeki başarı düzeyini ölçmek amacıyla kofenetik korelasyon katsayıları incelenmiştir. Bu yöntemler arasında en yüksek kofenetik korelasyon katsayısına sahip kümeleme yöntemi 0.96'dan fazla bir değer ile tek bağlantı yöntemi olmuştur. Bu da elde edilen kümeleme sonuçlarının hiyerarşik yöntemler arasında en başarılı yöntemin tek bağlantı yöntemi olduğunu göstermektedir. Tek bağlantı yöntemi sonuçlarına göre, İstanbul, Ankara, İzmir ve Antalya illeri sapan gözlem şeklinde ayrı kümeler olarak öne çıkmıştır. Diğer iller de başka bir kümede toplanmıştır. Bu sonuçların oluşması, bahsi geçen illerin diğer illere göre ulaştırma göstergeleri bakımından bariz farklılıklarının olduğunu işaret etmektedir. Bununla birlikte, hiyerarşik yöntemler arasındaki tek bağlantı yönteminin sapan gözlemlere karşı yapısı gereği hassas olmasının da bu sonucun ortaya çıkardığı söylenebilir.

Hiyerarşik olmayan yöntemler, çok sayıda gözlem ve değişken içerikli büyük verilerde rahatlıkla ve hızlı biçimde analiz edebilmesi ve aykırı değerlere karşı daha az hassas olması

nedeniyle araştırmacılar açısından daha uygulanabilir yöntemlerdir; ancak, küme sayısının ayrıca belirlenmesinin gerekliliği ve ön değerlendirme amaçlı bir analiz olarak kullanılmaması gibi dezavantajları bulunmaktadır. Hiyerarşik olmayan kümeleme analizinde literatürde en sık ve yaygın şekilde uygulanan yöntemlerden birisi k-ortalama yöntemidir. Bu yöntemin çalışmadaki sonuçlarına göre İstanbul ili, diğer yöntemlerde olduğu gibi tek kümede yer almıştır. İhracat-ithalat faaliyetlerinin yoğun olarak yapıldığı, gelişmiş liman ve tesislere sahip sanayileşmiş durumdaki Adana, Hatay, İzmir, Kocaeli ve Mersin illeri bir kümede, yine belli bir gelişmişlikte sanayi kollarına sahip Ankara, Antalya, Balıkesir, Bursa, Konya, Manisa ve Muğla illerinin ayrı bir kümede benzer gözlemler olarak diğer illere göre farklı bir konumda olduğu görülmektedir. Bu durum, illerdeki sanayileşme ile ulaşım faaliyetleri arasındaki güçlü ilişkinin bir işareti olarak yorumlanabilir.

Kullanılan veri seti tüm kümeleme analizi yöntemleriyle birlikte değerlendirildiğinde mutlak bir şekilde İstanbul ili, sahip olduğu devasa ulaştırma göstergeleriyle tek başına özel bir konumda ayrı bir küme olarak öne çıkmaktadır. Bu durum, diğer illere göre olağanüstü büyüklükteki ulaşım ağı ve alt yapı yoğunluğundan kaynaklandığı ve diğer illerle arasında büyük bir uçurum olduğu şeklinde açıklanabilir. Bunun yanında, uygulanan analizler sonucunda sıkça öne çıkan iller Marmara Bölgesi'nde Balıkesir, Bursa, Kocaeli illeri, Ege Bölgesi'nde İzmir, Manisa, Muğla illeri, Akdeniz Bölgesi'nde Antalya ili, İç Anadolu Bölgesi'nde Ankara ve Konya illeri, Karadeniz Bölgesi'nde ise Samsun ilidir. Doğu ve Güneydoğu Anadolu bölgelerinde sapan gözlem özelliği gösteren il çıkmasa da özellikle coğrafi özelliklerin zorlu seviyelerde olduğu Hakkâri, Şırnak, Bitlis, Iğdır gibi iller ulaşım göstergeleri açısından oldukça benzer bir görüntü çizmektedir. Doğu Karadeniz bölgesi için de aynı durum söz konusudur. Genellikle bu bölgelerdeki benzer coğrafi şartlar ve yer şekillerinin bu sonuca etki ettiği söylenebilir. Bir diğer önemli husus ise Ege, Akdeniz ve Marmara'ya kıyısı olan illerin uygulanan analiz çıktılarında diğer illere göre avantajlı bir konuma erişebildiği, ancak Karadeniz'e kıyısı olan illerde denizin sağladığı avantajların bu illere tam olarak yansımadağı gözlemlenmektedir. Bu bölgede yalnızca Samsun ve Trabzon illeri, uygun ulaştırma sistemlerine sahip olmasının da etkisiyle bölgede bir adım öne çıkmış görünmektedir.

Literatürde iller bazında yapılmış birçok çalışma bulguları karşılaştırıldığında, mevcut çalışmada elde edilenlerden çok farklı olmadığı görülmüştür. Birçok Doğu ve Güneydoğu Anadolu illeri ile Doğu Karadeniz illeri aynı ortak kümelerde toplanmıştır. Bu iller Türkiye'nin genel anlamda biraz daha geri kalmış illeri olarak bilindiği için ulaşım konusunda gelişime ihtiyaç duyduğu fikri ağır basmaktadır. İstanbul ili başta olmak üzere özellikle Türkiye'nin en büyük kent merkezlerini barındıran Ankara, İzmir, Bursa, Antalya illeri ulaştırma göstergeleri baz alındığında sapan gözlem özelliklerine sahip olduğu ve diğer tüm illerden farklı olacak şekilde ayrı kümeler oluşturma eğiliminde oldukları görülebilmektedir. Sanayi, ticaret ve finans merkezi konumunda yer alan illerin genelde aynı kümelerde sınıflandıkları, ulaştırma göstergelerinin bu durumla bir bağlantısının olduğu fikrini vermektedir. Elde edilen sonuçlarda, mevcut illere ait özel konum ve yer şekillerinin de yakından ilişkili olduğu değerlendirilmektedir. Örneğin, Konya ilinin coğrafi yapısının düz olması ve büyük nüfuslu bölgelere yakınlığı bu sonuçlara yansıdağı söylenebilir.

Çalışmada ön planda yer alan bazı sonuçlar değerlendirildiğinde, hiyerarşik kümeleme analizi yöntemleri arasında en yüksek kofenetik korelasyon katsayısına sahip tek bağlantı yöntemi sonucunda, çok istenen bir durum olmasa da İstanbul, Ankara, İzmir ve Antalya

illeri bir objelik kümeler oluşturmuştur. Mevcut illerin Türkiye'de beş büyük il arasında olması ve tek bağlantı yönteminin sapan gözlemlere karşı olan duyarlılığının da bunda etkili olduğu söylenebilir. Fakat, Türkiye'nin en büyük beş kent merkezi arasında bulunan Bursa'nın diğer 76 ile göre bir farklılık oluşturamaması dikkati çekmektedir. Çünkü, Bursa ilinin ulaşım verileri göz önüne alındığında tıpkı Antalya, İzmir, Ankara illeri gibi büyükşehir ili olarak bir farklılık oluşturması beklenmektedir. Bursa ili, gerek nüfusu, gerek sanayisi, gerek ulaşım verileri ve altyapısıyla bu duruma uygun görünmektedir.

Hiyerarşik olmayan kümeleme yöntemleri, yapıları gereği özellikle sapan değerlerin çok fazla olduğu verilerde daha avantajlıdır. Bu yöntemlerden ilk ön plana çıkan yöntem k-ortalamlar yöntemidir. Bu yöntemin çalışmadaki sonuçları değerlendirildiğinde, belirlenen beş küme içinde Türkiye'nin en büyük nüfusa sahip ili olan İstanbul, ekonomik, tarihi ve sosyokültürel açıdan farklı konumda olduğundan dolayı tek başına bir küme oluşturmuştur. Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa ve Muğla illeri de ayrı bir kümede yer almışlardır. Bu kümede, gelişmiş ulaşım tesislerinin bulunması ve yoğun ulaşım faaliyetleri icra edilmesinin yanında bu faaliyetlerin varlığının bir işareti olarak gelişmiş sanayi ve turizme dönük tesisler bulunmaktadır. Adana, İzmir, Kocaeli, Hatay ve Mersin illerinin oluşturduğu kümede, yoğun ithalat ve ihracat faaliyetleri gerçekleştirilmektedir. Bir diğer küme ise Afyonkarahisar, Aydın, Çanakkale, Denizli, Diyarbakır, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Isparta, Kahramanmaraş, Kars, Kayseri, Kütahya, Malatya, Mardin, Samsun, Sivas, Şanlıurfa, Tekirdağ, Tokat, Trabzon, Uşak ve Van illerinin oluşturduğu kümedir. Sözü edilen küme incelendiğinde, birçok önemli ulaşım göstergeleri olduğu göze çarpmaktadır. Doğu Karadeniz Bölgesi altındaki illerden, Doğu Ekspresi'nin yer aldığı demiryolu hattı, ayrıca yurt dışına çıkan önemli boru hatları ve demiryolu hatları geçmektedir. Karadeniz Bölgesi'ndeki Samsun ve Trabzon illeri, Marmara Bölgesi'ndeki Edirne ve Çanakkale illeri önemli liman geçişlerinin icra edildiği illerdir. Bu kümede yer alan İç Ege illeri de yoğunlaşmış demiryolu ve karayolu ağına ev sahipliği yapmaktadır. Bunların dışında kalan kümede ise ön planda bir ulaşım göstergesinin göze çarpmadığı (Hakkari, Ardahan, Ağrı vs.) büyük yüz ölçümüne sahip olmayan iller yer almaktadır.

Analiz sonuçları göz önüne alındığında, bilhassa k-ortalamlar yönteminin 2020 yılının başında ortaya çıkan Covid-19 pandemisinin Türkiye'deki başlangıç yayılımına yakın benzerlikte sınıflama sonuçları verdiği gözlemlenmiştir. Dolayısıyla, bu yöntem kullanılarak bir virüs yayılım simülasyonunun ulaşım sistemlerinin analiz edilmesiyle yapılabileceği fikri ortaya atılabilir. Bu benzetimin Covid-19 veya başka pandemilerde özellikle ilk yayılış safhalarında başarılı sonuçlar gösterebileceği düşünülmektedir.

İstanbul'dan farklı bir noktaya yatırım yapmak isteyen bir işletmeci mantığı ile çalışma sonuçları değerlendirildiğinde, örneğin yurt dışına kolayca bağlantı kurulabilecek konumda bir tesis yatırımı için k-ortalamlar yöntemindeki sonuçlara göre yurt dışına yoğun ulaşım faaliyetlerinin göze çarptığı Adana, İzmir, Kocaeli, Hatay ve Mersin illerinden oluşan küme değerlendirilebilir. Ya da sanayi ve turizm tesislerinin yoğun şekilde toplandığı bunun sebebinin de ulaşımına bağlı olduğu söylenebilen Antalya, Ankara, Bursa, Balıkesir, Konya, Manisa ve Muğla illerinin toplandığı küme titizlikle değerlendirilebilir. Ayrıca, k-ortalamlar yönteminde ortaya çıkan ve önemli ulaşım faaliyetleri ve tesislerinin bulunduğu 3. Kümede yer alan iller de bu sonuçlara göre ele alınabilir.

Araştırmacıların, bu çalışmanın sonuçlarını dikkate alarak, hiyerarşik yöntemlerde tekil küme oluşturan illeri, ya da diğer illerden farklılaşması beklendiği halde bu beklentiyi

karşılama illerde bunun altında yatan olası farklı sebepleri araştırabilecekleri söylenebilir. Hiyerarşik yöntemlerde dendrogram görüntülerinin incelenmesiyle, bu çalışmayla ilgili farklı araştırmalar yapmak ve sonuçlara ulaşmak mümkün olabilir. Farklı küme sayıları kullanılarak yapılacak araştırmalar da bu seçeneklerden biridir. Örneğin, Ward yöntemi ile olası küme sayısını tek başına 12 olarak belirleyerek bu konuda farklı bir çalışma yapılabilir. Benzer şekilde, çalışmada karşılaşılan ağırlık merkezi ve medyan yöntemlerinde meydana gelen tersine dönme durumları da ayrı bir araştırma konusu olabilir.

Çalışmada elde edilen bulguların özellikle politika yapıcılar ve karar vericiler için illerin ulaştırma türlerini geliştirmeye yönelik gerçekleştirilecek faaliyetleri planlamada önemli bilgiler içerdiği söylenebilir.

Teşekkür ve Bilgilendirme

Bu makale Erkan ALTIN'ın yüksek lisans tezinden üretilmiştir.

Kaynakça

- Aggarwal, C. C., & Reddy, C. K. (2014). *Data Clustering: Algorithms and Applications*. Londra: Chapman ve Hall/CRC Data Mining and Knowledge Discovery Series.
- Alonso, A., Monzón, A., & Cascajo, R. (2015). Comparative Analysis of Passenger Transport Sustainability in European Cities. *Ecological Indicators*, 48, 578-592.
- Alpar, R., (2017). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*. Ankara: Detay Yayıncılık.
- Atalay, A. & Tortum, A. (2013). Trafik Hareketliliğine Göre Türkiye'de İllerin Kümelenmesi. 5. *Karayolu Trafik Güvenliği Sempozyumu ve Sergisi*, 204.
- Bulut, H., (2018). *R Uygulamaları ile Çok Değişkenli İstatistiksel Yöntemler*. Nobel Akademik Yayıncılık, Ankara, Türkiye.
- Bulut, H., (2019). Türkiye'deki İllerin Yaşam Endekslerine Göre Kümelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23 (1), 74-82.
- Cabral, A., & De Sousa Ramos, F. (2014). Cluster Analysis of the Competitiveness of ContainerPorts in Brazil. *Transportation Research Part A: Policy and Practice*, 69, 423-431.
- Carvalho, P. R., Munita, C. S., & Lapolli, A. L. (2019). Validity Studies Among Hierarchical Methods of Cluster Analysis Using Cophenetic Correlation Coefficient. *Brazilian Journal of Radiation Sciences*, 7(2A).
- Cengiz, D., & Öztürk, F. (2012). Türkiye'de İllerin Eğitim Düzeylerine Göre Kümeleme Analizi ile İncelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(1), 69-84.
- Çakmak, Z., Uzgören, N., & Keçek, G. (2005). Kümeleme Analizi Teknikleri ile İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerin İncelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (12).
- Çelik, Ş., Şengül, T., Şengül, Ö., & İnci, H. (2019). Türkiye'de İllere Göre Hayvansal ve Bitkisel Ürünlerin Kümeleme Analizi ile İncelenmesi. *Journal of Awareness*, 3 (5), 385-398.
- Er, F. (2006). 2004 Yılında Türkiye'ye Gelen Yabancıların Taşıt Araçlarına Göre Dağılımının İstatistiksel Analizi. *Anatolia: Turizm Araştırmaları Dergisi*, 17(2), Güz: 184-188.

- Gitto, S., & Mancuso, P. (2019). Brand Perceptions of Airports Using Social Networks. *Journal of Air Transport Management*, 75, 153-163.
- Grimm, L. G., & Yarnold, P. R. (1995). *Reading and Understanding Multivariate Statistics*. USA: American Psychological Association.
- Hair, J. F. Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Multivariate Data Analysis*. USA: Pearson.
- Haustein, S., & Nielsen, T. A. S. (2016). European Mobility Cultures: A Survey Based Cluster Analysis Across 28 European Countries. *Journal of Transport Geography*, 54, 173-180.
- İncekırık, A., Güneri, Ö. İ., & Durmuş, B. (2021). Classification of Cancer Types by Cluster Analysis Methods. *Alphanumeric Journal*, 9(1), 125-142.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis 6th Edition*. Upper Saddle River, NJ: Prentice Hall.
- Karaoğlan, Y. (2018). *Farklı Kümeleme Analizi Yöntemleri ile Türkiye'deki İllerin Sosyoekonomik Özelliklerine Göre Sınıflandırılması ve Karşılaştırılması*. (Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi, Sivas).
- Kaya, E. (2012). Ulaştırma Kavramı ve Önemi. *Ulaştırma Sistemleri*, 2-23. Eskişehir: Anadolu Üniversitesi
- Klinger, T., Kenworthy, J. R., & Lanzendorf, M. (2013). Dimensions of Urban Mobility Cultures—A Comparison of German Cities. *Journal of Transport Geography*, 31, 18-29.
- Lessig, V. P. (1972). Comparing Cluster Analyses with Cophenetic Correlation. *Journal Of Marketing Research*, 9(1), 82-84.
- Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, Pp. 281-297).
- Nezerenko, O., Koppel, O., & Tuisk, T. (2017). Cluster Approach in Organization of Transportation in the Baltic Sea Region. *Transport*, 32(2), 167-179.
- Rodrigue, J. P., Comtois, C., & Slack, B. (2013). *Transportation, Economy and Society. The Geography of Transport Systems, 3rd Edn*. New York: Routledge.
- Saraçlı, S., Doğan, N., & Doğan, İ. (2013). Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation. *Journal of Inequalities and Applications*, 2013(1), 1-8.
- Seber, G. A. (2009). *Multivariate Observations*. USA: John Wiley ve Sons, Inc.
- Sokal, R. R., & Rohlf, F. J. (1962). *The Comparison of Dendrograms by Objective Methods*. *Taxon*, 11(2), 33-40.
- Wang, T. C., & Pham, Y. T. H. (2020). An Application of Cluster Analysis Method to Determine Vietnam Airlines' Ground Handling Service Quality Benchmarks. *Journal of Advanced Transportation*, 2020.
- Yılmaz, Z., & Uzgören, E. (2014). Türkiye'de İllerin Temel Bankacılık Faaliyetleri Yönünden Kümeleme Analizi Yöntemiyle Sınıflandırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 535-554.