



Gri Kümeleme Analizi ile Türkiye'deki Şehirlerin Sosyo-Ekonomik Verilere Göre Sınıflandırılması

Kübra KARADAŞ

Sivas Cumhuriyet Üniversitesi, Ekonometri Bölümü
kubrakaradas45@gmail.com
Orcid No: 0000-0003-4858-313X

Necati Alp ERİLLİ

Sivas Cumhuriyet Üniversitesi, Ekonometri Bölümü / Prof. Dr.
aerilli@cumhuriyet.edu.tr
Orcid No: 0000-0001-6948-0880

Özet

Sınıflandırma, herhangi bir konu hakkında gözlemlenen verilerin yapılarını inceleyerek verileri birbirlerine benzer gruplara, içeriklere ve belirlediğimiz bazı özelliklere göre gruplandırma işlemi olarak tanımlanabilir. Böylece birbirlerine benzeyen gruplarda yer alan gözlemlerin araştırma amaçlarına göre değerlendirilmesi ve yorumlanması kolaylaşacaktır. Kümeleme analizi sınıflandırma çalışmalarında en sık kullanılan yöntemlerden biridir. Kümeleme analizi doğal grupları bulunmayan verileri grup veya kümelerin sınırlı bir sayısına ayırmayı amaçlayan çok değişkenli istatistiksel bir yöntemdir. Gri İlişkisel Kümeleme yönteminde kümelerin belirli bir kurala göre gruplanmış nesnelere oluştuğu için kümeler homojenliğe sahip olması ve yeniden hesaplama yapılmaya gerek duymadan nesnelere kendi içerisinde ayrımını yapabilen bir yöntemdir. Ayrıca küme sayısı analiz öncesinde değil, kümeleme gerçekleştirildikten sonra belirlenebileceği için yaygın kullanılan diğer kümeleme algoritmalarına göre alternatif bir yaklaşım sunabilmektedir.

Bu çalışmada Türkiye'deki illerin sosyo-ekonomik veriler yardımıyla gri kümeleme analizi ile sınıflandırılması amaçlanmıştır. Türkiye İstatistik Kurumu'ndan alınmış veriler ile öncelikle gri ilişkisel katsayı matrisleri oluşturulmuş ve sonrasında kümeleme analizi yapılarak illerin sosyo-ekonomik verilere göre dağılımı belirlenmiştir. Analizler eğitim, tarım, istihdam gibi alt başlıklarda verilen veri kategorileri için de ayrı ayrı yapılarak Türkiye'nin sosyo-ekonomik haritası belirlenmeye çalışılmıştır.

Anahtar sözcükler: Gri Kümeleme Analizi, Sosyo-Ekonomik Veri, Sınıflandırma, Gri İlişkisel Analiz

Corresponding Author / Sorumlu Yazar: 1- Necati Alp ERİLLİ, Sivas Cumhuriyet Üniversitesi.

Atıf / Citation: KARADAŞ K., ERİLLİ N. A. (2023). Gri Kümeleme Analizi ile Türkiye'deki Şehirlerin Sosyo-Ekonomik Verilere Göre Sınıflandırılması. *İstatistik Araştırma Dergisi*, 13 (2), 60-74.

Bu çalışma Prof. Dr. N. Alp Erilli danışmanlığında Kübra Karadaş tarafından yazılan ve Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri bölümü tarafından onaylanan "Gri Kümeleme Analizi ve Bir Uygulama" isimli yüksek lisans tezinden türetilmiştir.

Classification of Provinces in Türkiye According to Socio-Economic Data with Gray Cluster Analysis

Abstract

Classification can be defined as the process of examining the structures of the data observed on any subject and grouping the data according to similar groups, contents and some characteristics we have determined. Thus, it will be easier to evaluate and interpret the observations in similar groups according to the research objectives. Cluster analysis is one of the most frequently used methods in classification studies. Cluster analysis is a multivariate statistical method that aims to classify data without natural groups into a limited number of groups or clusters. In the Gray Relational Clustering method, since the clusters consist of objects grouped according to a certain rule, the clusters have homogeneity and it is a method that can distinguish objects within itself without the need for recalculation. In addition, since the number of clusters can be determined after clustering is performed, not before the analysis, it can offer an alternative approach to other commonly used clustering algorithms.

In this study, it is aimed to classify the provinces in Türkiye by using gray clustering analysis with the help of socio-economic data. Gray relational coefficient matrices were first created with the data obtained from the Turkish Statistical Institute and then clustering analysis was performed to determine the distribution of provinces according to socio-economic data. Analyses were also conducted separately for the data categories given under sub-headings such as education, agriculture and employment, and with this it is aimed to determine the socio-economic map of Türkiye.

Keywords: Gray Clustering Analysis, Socio-Economic Data, Classification, Gray Relational Analysis

1. Giriş

Bireyler, kurumlar veya organizasyonlar kendi ihtiyaçları doğrultusunda çeşitli amaç veya hedeflerini gerçekleştirmek için karar vermek durumundadırlar. Kendilerine maksimum fayda elde etmek için birçok alternatif karşısında en iyi seçimi belirlemek çoğu zaman kolay olamamaktadır. Bu seçimleri bazı alternatif tekniklerden yararlanarak yapmak hem zaman hem de maliyet açısından tasarruf sağlanmasına yardımcı olmaktadır. Bu teknikler genel olarak karar verme ve sınıflama olarak iki genel başlık altında toplanabilir. Karar verme ve sınıflandırma neredeyse benzer faaliyetlerdir. Her iki durumda da bir problem hakkındaki bilgi ve veriler bir çözüm aramak için kullanılır. Karar kuralları ve sınıflandırma algoritmaları bilgi çıkarımını gerçekleştirir. Kullanıcı tanımlı sınıflar bağlamında veriler dönüştürülür, sınıf etiketleriyle eşleştirilir. Bilgi, bir dizi soru bağlamında ilgili verilerdir (Mulder, 1985). Karar vermede eylemler genellikle gerçek dünyada gerçekleşen sorunlardır ve elde edilen çözümler direkt sonuca etki edebilmektedir. Sınıflandırma çalışmalarında eylemler ise genellikle bir gözlemin bir gruba dahil edilmesi veya isimlendirilmesi olarak karşımıza çıkmaktadır ve sonuçlar bazen gerçek dünyada teori düzeyinde kalabilmektedirler.

Genel bir tanım vermek gerekirse sınıflandırma, herhangi bir konu hakkında gözlemlenen verilerin yapılarını inceleyerek verileri birbirlerine benzer gruplara, içeriklere ve belirlenen bazı özelliklere göre gruplandırma işlemi olarak tanımlanabilir. Sonuç olarak birbirlerine benzeyen gruplarda yer alan gözlemlerin araştırma amaçlarına göre değerlendirilmesi ve yorumlanması kolaylaşacaktır. Karar verme, canlıların çeşitli alternatifler ve imkanlar arasında amaçlarına uygun olanlarını seçmek olarak tanımlanabilir. Karar verme süreci, karar vericinin değerlerine, tercihlerine ve inançlarına dayanan bir akıl yürütme sürecidir. Her karar verme süreci bir seçimden oluşur ve bireyin bazen lehine bazen de aleyhine sonuçlar verebilir.

Kıtalar, ülkeler veya bölgesel çalışmalarda gelişmenin ve kalkınmanın en önemli göstergelerinden biri sosyo-ekonomik gelişmişlik verileri gösterilebilir. Birçok yerde olduğu gibi Türkiye'de de il veya bölgesel bazda kullanılan bu veriler ile gelişmişliğin düzeyi belirlenebilmektedir. Türkiye'de yapılacak olan kamu veya özel yatırımlar ve bunlara ayrılacak paylar, bu veriler esas alınarak hazırlanmaktadır. Türkiye'de gelişmenin mekân üzerinde dağılımı iller çevresinde yoğunlaşma tarzındadır. Bu dağılımın kapsamı ve çevre iller üzerinde etkisi, merkezlerdeki iktisadi ve sosyal faaliyetlerin yoğunluğu ile de doğru orantılıdır. Mekânsal gelişme eğilimleri bölgesel ve bütünsel gelişmelere de altyapı oluşturmaktadır (Erilli, 2009). Yine de yapılacak birçok çalışmadan elde edilen sonuçların doğru ve tarafsız bir şekilde değerlendirilmesi, bölgesel gelişmişlik farklarının azaltılması konusunda önemli bir yer tutmaktadır.

Literatürde sosyo-ekonomik veriler yardımıyla yapılan sınıflama çalışmaları her geçen gün artış göstermektedir. Farklı istatistiksel yöntemler ile elde edilen farklı sınıflama sonuçları ile araştırmacılar, Türkiye'nin sosyo-ekonomik haritasına katkı sağlamaktadırlar. Filiz (2005) çalışmasında Türkiye'deki illeri k-ortalama tekniği ile sınıflandırmış ve 16 sosyo-ekonomik değişken ile 7 küme elde edilmiştir. Albayrak (2005) çalışmasında Türkiye'deki illerin sosyo-ekonomik gelişmişlik düzeylerini faktör analiziyle araştırmış ve 8 faktör elde etmiştir. Kaygısız vd. (2005) çalışmalarında Path analizi ve Kümeleme analizi birlikte kullanılmış ve sonuç olarak Türkiye'de iller 5 kümeye ayrılmıştır. Erilli vd. (2009) çalışmalarında Türkiye'deki iller, bulanık kümeleme analizi ile 58 değişken kullanılarak sınıflandırılmış ve sonuç olarak 5 küme bulunmuştur. Çelik (2013) çalışmasında k-ortalamalar yöntemi kullanarak Türkiye'deki illeri sağlık değişkenlerine göre sınıflandırmış ve elde edilen sonuçları 7, 10 ve 15'li küme bölünmesi için yorumlamıştır. Tekin (2015), sağlık göstergeleri yardımıyla Türkiye'deki illeri Ward yöntemi ile gruplandırmış ve sonuçta 5, 7 ve 11'li küme bölünmeleri elde edilmiştir. Zorlutuna ve Erilli (2018) çalışmalarında 2002, 2008 ve 2013 yıllarına ait 3 farklı dönem sosyo-ekonomik veriler kullanarak Bulanık C-ortalamalar yöntemi ile Türkiye'deki illeri sınıflandırmışlardır. Çalışmada her 3 dönem için iller ve yıllar arası korelasyon değerleri hesaplanarak yorumlanmıştır. Kandemir (2018) çalışmasında konaklama verileri kullanarak Türkiye'deki illeri Bulanık kümeleme yöntemi ile sınıflandırmıştır. Yerli ve yabancı turizm için ayrı ayrı yapılan analizlerde illerin 2 kümede toplandıkları belirlenmiştir. Bulut (2019) çalışmasında Türkiye'deki illeri memnuniyet endeks verilerine göre modele dayalı EM algoritması ve k-ortalamalar yöntemi kullanarak kümelere ayırmıştır. Analizlerin sonucunda EM algoritması ile 2 küme, k-ortalamalar yöntemi ile 5 küme elde edilmiştir. Birkalan ve Bay (2022) çalışmalarında Türkiye İBBS Düzey-2 bölgelerini bölgesel asgari ücret varsayımı altında, 2015-2020 yılları için Ward kümeleme yöntemi ile analiz etmişler ve bazı yıllar için 3, bazı yıllar için de 4 küme bulmuşlardır. Tüm bu çalışmalarda kullanılan farklı sınıflama yöntemleri ve farklı sayıda değişkenler ile elde edilen küme sayılarının değişkenlik göstermesi beklenen bir sonuçtur. Önemli olan kullanılan değişkenlerin, yapılan sınıflama çalışmalarındaki yapıyı ne derece etkilediğinin belirlenmesi ve kullanılan istatistiksel yöntemlerin sonuçlar üzerindeki etkilerinin yorumlanabilmesidir.

Bu çalışmada 2019 yılına ait sosyo-ekonomik veriler kullanılarak Türkiye'nin sosyo-ekonomik haritası belirlenmeye çalışılmıştır. Demografik, Eğitim, Nüfus ve Göç, Tarım ve Hayvancılık, Sağlık ve Finans başlıkları altında hazırlanan bu veriler ile oluşturulacak sosyo-ekonomi haritaları, Çok Kriterli Karar Verme (ÇKKV) ve sınıflama yöntemlerinin ortak yönlerini barındıran Gri kümeleme analizi ile gerçekleştirilmiştir.

2. Materyal ve Metot

2.1. Kümeleme Analizi

Kümeleme analizi sınıflandırma çalışmalarında en sık kullanılan yöntemlerden biridir. Kümeleme analizi sınıfları (doğal grupları) bulunmayan verileri grup veya kümelerin sınırlı bir sayısına ayırmayı amaçlayan çok değişkenli istatistiksel bir yöntem olarak tanımlanabilir. Ayırma, aynı gruptaki gözlemlerin birbirine benzer iken farklı gruplardaki gözlemlerin birbirlerinden farklı olacak şekilde yapılandırılması işlemleridir. Böylece gruplar içi homojen, gruplar arası ise heterojen yapılar elde edilebilecektir. Kümeleme analizi ilk kez Tryon tarafından 1939 yılında kullanılmış ve 1960'lı yıllardan sonra kullanımı yaygınlaşmıştır (Anderberg 1973). Kümeleme analizi, iş ve bilim dünyasında sık kullanılan çok değişkenli istatistiksel yöntemlerinden belki de en önemlisidir. Kümeleme Analizi incelenen bir çalışmada gözlemler arasındaki benzerliklerine veya benzemezliklerine göre kümeler oluşturularak sınıflandırma yapmayı, birimlerin benzer özelliklerini belirlemeyi, bu sınıflar ile ilgili genel düzenlemeler yapmayı ve genel yargılarda bulunmayı sağlayan bir yöntem olarak tanımlanabilir. Yöntemin amacı grup oluşturmamış gözlemleri benzerliklerine (benzemezliklerine) göre sınıflandırmak ve araştırma yapanlara uygun, genellenebilir özet bilgiler elde etmelerinde yardımcı olmaktır (Tatlıdil, 2002). Kümeleme Analizi, birbirine çok benzeyen gözlemleri aynı küme içinde gruplandırır ve bunun sonucunda oluşan kümeler kendi içlerinde yüksek homojen ve kümeler arası düşük heterojen değerler taşımaktadırlar (Kalaycı, 2005). Kümeleme analizindeki genel amaçların dışında bazı özel amaçları da bulunmaktadır. Bunlar; örneklemelerin bağlı olduğu gerçek tiplerin belirlenmesi, modelleştirmenin kolaylaştırılması, gruplar için önsel tahminler, önceden belirlenen hipotezlerin test edilmesi, veri yapısının belirginleştirilmesi, veri (boyut) azaltma ve uç (aykırı) değerlerin bulunması olarak özetlenebilir (Tatlıdil, 2002). Kümeleme analizinde kullanılmak üzere bireylerin veya nesnelerin aralarındaki benzerlik için uzaklık ölçüleri, korelasyon ölçüleri veya katsayıları karşılaştırma ölçütleri kullanılmaktadır. Uzaklık ölçüleri bireyleri sınıflandırmak için kullanılırken, korelasyon ölçüleri ise değişkenleri sınıflandırmada kullanılır (Erilli, 2009).

2.2. Gri Sistem Teorisi

Çok Kriterli Karar Verme (ÇKKV) Teknikleri, sayıca fazla olan kriterlere göre seçeneklerin avantaj ve dezavantajlarını değerlendirebilen analitik yöntemler topluluğu olarak özetlenebilir. ÇKKV, çok sayıda ve aynı anda uygulanan kriterler arasında en iyi ve en doğru alternatifi seçilmesini sağlayan yöntemdir (Hamurcu ve Eren, 2015). Başka bir ifade ile ÇKKV teknikleri ile karar vericiler farklı özelliğe sahip seçenekleri birden çok kriterlere göre değerlendirerek sıralayabilirler. Bu yöntemlerden biri de Gri Sistem Teorisidir.

Gri sistem teorisi (GST) ilk defa Deng (1982) çalışması ile tanıtılmıştır. GST küçük örnekleme sahip ve yeterli bilgi içermeyen problemlerin çözümünde kullanılmak üzere önerilen bir yöntemdir (Liu ve Forrest, 2010). GST, belirsizliğin sayısallaştırılmasında kullanılan bir yöntem olarak da tanımlanmaktadır (Köse vd, 2013). Son yıllarda dünya genelinde çok kriterli çalışmalarda sıklıkla görülmeye başlanan Gri sistem teorisi istatistik, işletme, mühendislik gibi birçok alanda farklı konu başlıkları altında hızla yayılmıştır. Bulanık matematik yöntemi kullanılarak incelenen tüm nesnelere belirgin olmayan sınırları ve kesin bir içselliğe sahiptir. Örneğin "sıcak çorba" kavramı bulanık bir kavramdır. Zira birçok kişi sıcak çorba denilince aslında neyin anlaşılması gerektiğini bilir fakat belirgin bir sıcaklık kavramını tanımlamak zordur. Çünkü "Sıcak çorba" dediğimiz kavram çok geniş anlamları ifade eder. Gri sistem teorisinde ise incelenen nesnelere kesin sınır ve kesin olmayan bir içselliğe sahiptir. Örneğin, "Türkiye Merkez Bankası, 2023 yılsonu için dolar kur aralığını 28 ile 30 TL arasında kontrol etmeyi planlıyor" cümlesindeki "28 ile 30 TL" aralığı parasal büyüklük ile ilgili kesin bir sınırı belirtirken, tanımlanan aralığın bilinen (veya hesaplanan) değeri hakkında herhangi bir netlik içeren bilgiye sahip olamamız gri bir kavram olarak tanımlanabilir.

Problem çözümlerinde kullanılan matematiksel veya istatistiksel sistemler, gerçek hayatın küçük bir bölümünün modellenmesine benzer. Sistem modelleri karmaşık yapıları, farklı güvenilirlik düzeyleri veya belirsizlik noktaları arasındaki yapılar ile değerlendirilir. Ortaya çıkarılan matematiksel modeller ne kadar ayrıntı içerse de üzerinde araştırma yapılan gerçeği tam olarak yansıtamamaktadırlar. Çünkü gerçek dünya gerek işleyişi gerekse kavramları ile karmaşık bir yapı içermektedir. Bu karmaşıklıkların temel sebebi kavramlar üzerindeki belirsizlik, kesinlikten yoksunluk veya kararsızlıklardan kaynaklanmaktadır. Bilim insanları bu tarz sorunların yaşandığı problemlere her geçen gün farklı analizler ile yaklaşarak gerçeği farklı boyutlarda aramaya çalışmaktadırlar. Bu yöntemlerden biri de gri sistem teorisi ile kümeleme analizinin birleşimi olana Gri Kümeleme Analizi yaklaşımıdır.

2.3. Gri Kümeleme Analizi

Gri kümeleme analizinin temeli Deng (1982) tarafından geliştirilen Gri Sistem Teorisine (GST) dayanmaktadır. GST küçük örnekleri ve sınırlı bilgileri içeren problemlerin incelenmesine olanak sağlayan bir teoridir (Liu ve Forrest, 2010). Kümeleme; nesnelere ayrılmasına göre gri kümeleme analizi ve gri beyaz ağırlık fonksiyonu kümeleme analizi olarak ayrılabilir. Gri beyaz ağırlık fonksiyonu kümeleme analizi, bir gözlemin belirlenen farklı kategorilere ait olup olmadığını kontrol etmek için kullanılır. Gri kümeleme analizi ise temel olarak karmaşık sistemleri basitleştirilmiş hale getirmek için benzer faktörlerde kullanılır (Ke vd, 2012).

Gri kümeleme analizi; gözlemleri sınıflandırmak veya önceden belirlenen bir kümeye gözlemleri dahil etmek için gri sayıların beyazlaştırma ağırlık fonksiyonu kullanır. Gri kümeleme analizi birbirlerine benzeyen unsurların sınıflandırılması yoluyla karmaşık sistemlerin basitleştirilmesi için kullanılmaktadır (Liu ve Forrest, 2010).

Beyazlaştırma ağırlık fonksiyonlarını kullanan gri kümeleme yöntemi esas olarak gözlem nesnelere önceden belirlenmiş sınıflara ait olup olmadıklarını test etmek için uygulanır. Gri Kümeleme Yöntemi gözlem nesnelere tanımlanabilir gruplarla sınıflandırmak için geliştirilmiş ve gri insidans matrisleri veya beyazlaştırma ağırlık fonksiyonları aracılığıyla gerçekleştirilebilir (Liu ve Forrest, 2010). Gri kümeleme analizinin avantajları kümeleme nesnesinin örnek dağılımını özel bir gereksinim olmaksızın ele alması ve basit bir hesaplama süreci aracılığıyla daha yararlı kümeleme sonuçları elde edilebilmesidir (Ke vd, 2012). Gri kümeleme analizindeki, değerlendirilen nesnelere, fonksiyon değerine ve gri kategori değerine göre alt bölümlere ayrılır. Temel adımları aşağıda verildiği gibidir (Lin, 2015).

1. Gri kümeleme analizinin m göstergesi ve n tane gri kategorisi olmak üzere;
2. Gri kümeleme analizi işlevleri δ_{ij} ; m göstergenin ve n gri kategoriye ait olacaktır.

3. Gerçek değere veya tahmini değere göre gri kümeleme analizinin değerleri $n_j^k = \frac{\lambda_j^k}{\sum_{j=1}^m \lambda_j^k} u_i^j(P)$ dir. P

değerlendirmesinin altındaki nesnelere elde edilir ve R matrisi elde edilir.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pm} \end{bmatrix} \quad (1)$$

4. Gri kümeleme analizinin değerine göre matris δ^t , t nesnesinin yaklaşık m göstergesi ve n gri kategorisi elde edilir.

$$\delta^t = \begin{bmatrix} \delta_{11}^t & \delta_{12}^t & \cdots & \delta_{1n}^t \\ \delta_{21}^t & \delta_{22}^t & \cdots & \delta_{2n}^t \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{m1}^t & \delta_{m2}^t & \cdots & \delta_{mn}^t \end{bmatrix} \quad (2)$$

5. Değer γ_m^t t nesnesinin i -inci gri kategorisinin kapsamlı bir gri kümeleme analizi fonksiyonu elde edilir. T nesnesinin gri kategori fonksiyon değerine göre onaylanır.

$$\gamma_j^t = \frac{1}{m} \sum_{i=1}^m \delta_{ij}^t \quad (3)$$

Gri kümeleme adımları şu şekilde sıralanır (Chang ve Yeh, 2005):

Adım 1: Geçici bir küme tanımlanır.

$C = \{c_1, c_2, \dots, c_q\}$ şeklinde geçici bir küme tanımlanır. Geçici küme $C = X$ şeklinde düşünülmüş bir veri setidir.

Bu durumda $q = m$ ve $c_i = x_i$ olacaktır. Geçici küme, öğrenme sürecinde bir veri kümesi olarak kullanılır ve tam bir öğrenme yinelemesinden sonra C 'deki öğeler güncellenir.

Adım 2: Eşiğin başlangıç değeri ayarlanır.

Başlangıç eşik değeri ε olarak $[0, 1]$ aralığında olmak üzere bir gerçek sayı seçilmelidir. İlk eşik değeri, nihai kümeleme sonuçlarını (yani kümelerin sayısı ve küme merkezlerinin konumu) güçlü bir şekilde etkileyecektir (Bu daha sonra tanımlanacaktır).

Adım 3: Bir öğrenme süreci başlatılır.

$X_r = X_c = C$ olmak üzere ve $i = 1$ olarak tanımlanır.

Adım 4: Gri ilişkisel ilişki dereceleri hesaplanır.

Bu yineleme için $x_{ri} \in X_r$ ilgili referansı eğitim modeli yapın. Ardından, referans x_{ri} ile X_c 'deki tüm örüntüler

arasında gri ilişkisel örüntü dereceleri; $v(x_{ri}, x_{cj}), J = 1, 2, \dots, q$ ile gösterilen $v(x_{ri}, x_{cj}) = \left(\frac{d_{\max} - d_{ij}}{d_{\max} - d_{\min}} \right)^\xi$ ile

belirlenebilir. ($\xi \in (0, \infty)$ ayırt edici bir katsayıdır).

Adım 5: Etkin modeli güncellenir.

Karşılaştırmalı bir model x_{cj} için; gri ilişkisel örüntü derecesi $v(x_{ri}, x_{cj}) \geq \varepsilon$ 'yi karşılıyorsa; o zaman C 'de karşılık

gelen öğeye anlamlı bir örüntü denir. Bu anlamlı örüntüler arasında, aktif bir örüntü olan C_l , kendisine karşılık gelen karşılaştırma örüntüsü en büyük gri ilişkisel örüntü derecesi olan $v(x_{ri}, c_l) = \max_{j \in \{1, 2, \dots, q\}} v(x_{ri}, x_{cj})$ eşitliğine

sahip olacak şekilde tanımlanır. Böylece aktif olan örüntü C_l şu şekilde güncellenir: $c_l = \frac{\sum_{k=1}^N w_k c_k^*}{\sum_{k=1}^N w_k}$. Burada w_k ,

önemli c_k^* modelinin ağırlıklandırma faktörüdür. Genellikle tüm k değerleri için $w_k = 1/N$ alınır. Başka bir seçenek ise, $w_k = v(x_{ri}, c_k^*)$ eşitliğinin aktif örüntünün yeni konumunu daha büyük ilişki dereceli örüntülere daha

yakın hale getirmesine izin vermektir. Birden fazla aktif örüntü varsa, hepsi aynı anda $c_l = \frac{\sum_{k=1}^N w_k c_k^*}{\sum_{k=1}^N w_k}$ ile

güncellenir.

Eğer her eğitim verisi $i=q$ şeklinde tanımlanırsa Adım 6'ya gidilir. Aksi takdirde, l 'yi 1 artırıp ardından Adım 4'ten devam edilir.

Adım 6: Elde edilen sonuçlar kontrol edilir.

Güncellemeden sonra geçici küme aynı kalırsa, yani C 'nin yakınsaması sağlanırsa, öğrenme sürecini durdurulur ve küme merkezlerini belirlemek için Adım 8'e gidilir. Aksi takdirde, bir sonraki adıma geçilir.

Adım 7: Elde edilen sonuca göre eşik değeri artırılır.

Eşik değeri (ε) gerekirse artırılır. Böylece başka bir öğrenme yinelemesi için güncellenmiş C kümesiyle 3. adıma gidilir. ε eşik, $\varepsilon = 1$ olana kadar doğrusal veya üstel olarak arttırılabilir.

Adım 8: Kümeleri ve küme merkezleri belirlenir.

C 'nin yakınsaması sağlandıktan sonra, C 'deki bazı elemanlar eşit olabilir. Son güncellenen C kümesini, aynı alt kümedeki tüm öğelerin aynı olduğu birkaç ayrı alt kümeye bölün. Alt kümelerin sayısı verideki küme sayısıdır. C 'nin bir alt kümesindeki öğeler için, orijinal X veri kümesindeki karşılık gelen kalıpları bir küme olarak bölünebilir ve küme merkezi bu alt kümedeki öğelerden biri olacaktır.

Yüksek ilişkisel dereceli kalıpları bir küme olarak gruplamak için, referansa oldukça benzeyen N anlamlı kalıp seçilmeli ve ardından 5. Adımda aktif kalıbın yeni konumu olarak bunların ortalamasını alınmalıdır. Adım 5 ayrıca, düşük ilişkisel dereceli kalıpların eğitildikten sonra değişmediğini ortaya koymaktadır. Adım 7'deki eşik artırılması, kendi kendini organize eden özellik haritalarındaki mahalle boyutunu küçültme fikrine çok benzer (Kohonen, 1995). Eşik arttıkça, bu önemli modeller yakınsak bir duruma doğru hareket ederler. Nihai sonuçlar, istenen küme merkezleridir. Tüm prosedür, yalnızca veri kümesinin altında yatan yapıya dayandığından, önerilen yöntem denetimsiz bir kümeleme yaklaşımı olarak tanımlanabilir.

Adım 5'te, anlamlı modellerin seçiminin eşik değere bağlı olduğu görülmektedir. Aslında, Adım 2'de seçtiğimiz eşik başlangıç değeri, yöntemimizin nihai sınıflandırma sonuçlarını güçlü bir şekilde etkilemektedir. Genel olarak konuşursak, farklı başlangıç eşikleri, veri kümeleme probleminde farklı sonuçlar verebilir. Başlangıç eşik parametresinin duyarlılığı farklı örneklerde farklılık gösterilecektir. Yukarıda önerilen gri kümeleme algoritması, sabit bir başlangıç eşik için bir veri kümesini bölmek için kullanılır. $[0, 1]$ içinde farklı başlangıç eşikleri için sonuçlar almak için gri kümeleme algoritmasında fazladan bir prosedür kullanılabilir. Bu sonuçlar arasında, uygun bir kümeleme sonucu bulmamıza yardımcı olacak bir kritere ihtiyacımız olacaktır. Bir performans endeksi şu şekilde tanımlanmıştır (Jain ve Dubes, 1988):

$$PI = \frac{\sum_{d=1}^h m_d}{\sum_{d=1}^h m_d s_d} \quad (4)$$

Burada h küme sayısını, $m_d, d = 1, 2, \dots, h$ d-nci kümeyle ait örüntü sayısını ve s_d ise aşağıda verildiği gibi tanımlanmıştır:

$$s_d = \frac{\min_{j, j \neq d} \|o_d - o_j\|^2}{\sqrt{\sum_{t=1}^{m_d} \|x_{dt} - o_d\| / m_d}} \quad (5)$$

Burada o_d , d-nci kümenin küme merkezini ve x_{dt} ise d-nci kümeyle ait örüntülerini ifade etmektedir.

Bu indeks, bir kümedeki kalıpların merkeze yakın olması ve küme merkezlerinin birbirinden uzakta ayrılması gerektiği anlamında veri kümelemenin amacını yansıtır. s_d performans indeksi ne kadar büyük olursa, kümeleme sonucunun o kadar iyi olduğunu görebiliriz (Wong vd, 2001).

Şimdi, sonucun performans indeksini (PI) maksimize etmesi anlamında optimal bir sonuç belirlemek için gri kümeleme algoritmasının son adımını verilebilir.

Adım 9: En uygun sonuç belirlenir.

Adım 8’de farklı başlangıç eşikleri kullanılarak elde edilen her sonuç için performans indeksi (*PI*) hesaplanmalıdır. Optimal kümeleme sonucu, performans indeksini maksimize eden sonuçtur. *PI* yerine, Adım 9’da farklı performans indeksleri de kullanılabilir. Örneğin, küme merkezi ile kümedeki örüntüler arasındaki mesafelerin ortalaması, ele alınan tek amaç ise, alternatif performans indeksini en aza indiren sonucu aramak için Adım 9 değiştirilmelidir.

3. Uygulama

Bu çalışmada, Türkiye’deki şehirler 2019 yılına ait Türkiye sosyo-ekonomik verileri kullanılarak gri kümeleme analizi yöntemi sınıflandırılmaya çalışılmıştır. Çalışmada kullanılan 2019 yılı verileri Türkiye İstatistik Kurumu’nun internet adresinden alınmış ve derlenmiştir (URL-1).

2019 yılı verileri ile ilk olarak gri ilişkisel katsayı matrisleri oluşturulmuş ve sonrasında gri kümeleme analizi yapılarak illerin sosyo-ekonomik verilere göre dağılımı belirlenmiştir. Analizler MATLAB 2009, SPSS.25 ve Microsoft Excel paket programları ile yapılmıştır. Çalışmada Sosyo-Ekonomik verilerin tümü kullanılarak gri kümeleme analizi sonuçları elde edildikten sonra sırasıyla Demografik, Eğitim, Sağlık, Tarım, Nüfus ve Finans alt başlıkları için de ayrı ayrı gri kümeleme analizi uygulanmış ve Türkiye’nin sosyo-ekonomik haritası belirlenmeye çalışılmıştır.

Gri kümeleme analizi hesaplamalarında uygun küme sayılarının belirlenmesi için geçerli bir yöntem bulunmamaktadır. Bu yüzden uygun küme sayılarının belirlenmesinde klasik kümeleme yöntemlerinde sıklıkla kullanılan Silhouette ve bulanık kümeleme yöntemlerinde kullanılan Xie-Beni küme geçerlilik yöntemleri beraber kullanılmıştır (Rousseuw, 1987; Xie ve Beni, 1991). Her 2 yöntemine göre de aynı küme sayısının ulaşıldığı küme sayıları optimal küme sayısı olarak alınmıştır. Analizlerde kullanılan sosyo-demografik verileri oluşturan başlıklar ve bileşenleri Tablo 1’de verilmiştir.

Tablo 1. Analizlerde kullanılan 2019 yılı Türkiye Sosyo-ekonomik değişkenleri

| <i>Demografik</i> | <i>Eğitim</i> |
|---|--|
| Bebek Ölüm Hızı (binde) | İlkokul okullaşma oranı 2012 ve sonrası (kadın)-net (%) |
| 5 yaş altı ölüm hızı (binde) | İlkokul okullaşma oranı 2012 ve sonrası-net (%) |
| Toplam doğurganlık hızı (çocuk sayısı) | İlkokul /Öğretmen başına düşen öğrenci sayısı |
| Kaba evlenme hızı (binde) | İlköğretim okullaşma oranı 2012 ve sonrası (kadın)-net (%) |
| Kaba ölüm hızı (binde) | İlköğretim okullaşma oranı 2012 ve sonrası-net (%) |
| Kaba doğum hızı (binde) | Okuma yazma bilen oranı (%) |
| Kaba boşanma hızı (binde) | Ortaokul /Öğretmen başına düşen öğrenci sayısı |
| Doğuştan beklenen yaşam süreleri (yıl) | İlk ve ortaokul /Derslik başına düşen öğrenci sayısı |
| Nüfusa göre boşanma oranları | |
| <i>Nüfus ve Göç</i> | <i>Tarım-Hayvancılık</i> |
| Net göç hızı (binde) | Küçükbaş hayvan sayısı (baş) |
| Nüfus yoğunluğu (kilometrekareye düşen kişi sayısı) | Tahıllar ve diğer bitkisel ürünlerin üretim miktarı (ton) |
| Ortalama hane halkı büyüklüğü | Büyükbaş hayvan sayısı (baş) |
| Yıllık nüfus artış hızı (binde) | |
| Türkiye'den yurt dışına giden göç | |
| <i>Sağlık</i> | <i>Finans</i> |
| Yüzbin kişi başına toplam hastane yatak sayısı | Toplam ithalat (bin \$) |
| Bin kişi başına düşen toplam hekim sayısı | Toplam ihracat (bin \$) |
| | Kişi başına GSYH (TL) |

Kaynak: www.tuik.gov.tr (Erişim Tarihi: 10.11.2022)

Yukarıda da belirtildiği üzere Silhouette ve Xie-Beni küme geçerlilik yöntemlerinin aynı uygun sonuçları verdiği küme sayıları optimal küme sayısı olarak alınmış ve gri kümeleme analizi bu küme sayılarına göre yapılmıştır. Silhouette kriterine göre toplamda en büyük Silhouette değerini veren küme sayısı uygun küme sayısı olarak alınır (Rousseuw, 1987). Buna karşılık Xie-Beni indeksinde en küçük indeks değerini veren küme sayısı uygun küme sayısı olarak alınır (Balasko vd., 2005). Değişken grupları için hesaplanan Silhouette ve Xie-Beni indeks değerleri Ek.1'de verilmiştir. EK.1'de verilen tablo sonuçlarına göre sadece Eğitim ve Tarım-Hayvancılık değişken gruplarında bu 2 endeks değeri istenen durumda çıkmamış, diğerlerinde istenen durumlarda çıkmıştır (Silhouette indeksinin en büyük olduğu yerde Xie-Beni indeksinin en düşük çıkması). Eğitim ve Tarım-Hayvancılık değişkenlerinde Silhouette indekslerinin istenen duruma oldukça yakın çıkmalarından dolayı, minimum Xie-Beni indeks değerinin olduğu küme sayısı optimal sayı olarak belirlenmiştir. Bulunan optimal küme sayıları Tablo 2'de verilmiştir.

Tablo 2. Değişken gruplarına göre optimal küme sayıları

| Değişken Grubu | Küme Sayısı |
|-------------------|-------------|
| Sosyo-Ekonomik | 7 |
| Demografik | 8 |
| Eğitim | 8 |
| Sağlık | 3 |
| Tarım-Hayvancılık | 7 |
| Nüfus-Göç | 8 |
| Finans | 10 |

Tablo 2'de verilen küme sayılarına uygun Gri kümeleme analizi sonuçları Türkiye haritası üzerinde her değişken için ayrı ayrı gösterilmiştir. EK.2'de ilgili değişken gruplarına göre yapılan kümeleme sonuçlarından elde edilen kümelerin ilgili değişkenlere göre tanımlayıcı istatistikleri verilmiştir. Şekil 1'de Demografik değişkenlere göre illerin kümelere dağılımı gösterilmiştir.



Şekil 1. Demografik değişkenlere göre illerin kümelere dağılımı

Şekil 1'de verilen sonuçlara göre Tokat, Zonguldak, Isparta gibi küçük illerin İstanbul, Ankara, İzmir ile aynı kümede yer alması, Karabük'ün tek başına bir kümeyi oluşturması ve Bartın ile Giresun'un aynı kümede yer alması beklenenden farklı sonuçlar olarak karşımıza çıkmaktadır. Şekil 2'de Eğitim değişkenlerine göre illerin kümelere dağılımı gösterilmiştir.



Şekil 2. Eğitim değişkenlerine göre illerin kümelere dağılımı

Şekil 2’de verilen sonuçlara göre Hakkâri ve Tokat’ın aynı kümede yer alması, Gümüşhane’nin ise tek bir küme oluşturması farklı bir sonuç olarak ele alınabilir. İstanbul, İzmir, Ankara gibi illerin aynı kümede yer alması beklenen bir durumdur. 7. ve 8. kümelerdeki illerin eğitim olanaklarının diğer kümelerden daha az olduğu düşünülebilir. Şekil 3’te Sağlık değişkenlerine göre illerin kümelere dağılımı gösterilmiştir.



Şekil 3. Sağlık değişkenlerine göre illerin kümelere dağılımı

Şekil 3’te verilen sonuçların en dikkat çekici olanı İstanbul’un gerek nüfus gerekse demografik olarak benzediği birçok il ile aynı kümede yer alması olarak gösterilebilir. Şekil 4’te Tarım ve Hayvancılık değişkenlerine göre illerin kümelere dağılımı gösterilmiştir.



Şekil 4. Tarım ve Hayvancılık değişkenlerine göre illerin kümelere dağılımı

Şekil 4’te verilen sonuçlarına göre Konya tahıl üretimindeki lider konumuyla tek bir kümede yer aldığı söylenebilir. 4. küme genel olarak tarım alanı kısıtlı illerden oluşmuştur. Van ilinin tek kümede olması

hayvancılığın ildeki önemi ile açıklanabilir. Şekil 5’te Nüfus ve Göç değişkenlerine göre illerin kümelere dağılımı gösterilmiştir.



Şekil 5. Nüfus ve Göç değişkenlerine göre illerin kümelere dağılımı

Şekil 5’te verilen sonuçlarına göre İstanbul’un tek bir kümeyi oluşturmasının nedeni en fazla göç alması ve buradaki iş olanaklarının fazla olması etkili bir faktördür. Çanakkale ve Giresun illerinin de tek başlarına bir küme oluşturdukları görülmektedir. Şekil 6’da Finans değişkenlerine göre illerin kümelere dağılımı gösterilmiştir.



Şekil 6. Finans değişkenlerine göre illerin kümelere dağılımı

Şekil 6’da verilen sonuçlarına göre İstanbul, Ankara ve Gaziantep illerinin ticaret ve sanayi kaynaklı ayrı ayrı tek bir kümeyi oluşturduğu söylenebilir. Genellikle sanayi ve ticaretin az olduğu illerin 9. kümede yoğunlaştığı görülmektedir. Son olarak 30 değişkenden oluşan Sosyo-Ekonomik veriler yardımıyla elde edilen Gri Kümeleme sonuçları Şekil 7’de verilmiştir.



Şekil 7. Sosyo-Ekonomik verilere göre illerin kümelere dağılımı

Şekil 7’de verilen sonuçlarına göre Tüm veriler değerlendirildiğinde Ankara, İzmir ve İstanbul illerinin aynı kümede yer aldığı görülmektedir. Bartın ve Giresun illerinin aynı kümede yer alması ve Karabük ilinin tek bir

kümeyi oluşturması farklı sonuçlar olarak söylenebilir. 3. küme genel olarak kırsal nüfusun çok olduğu iller olarak karşımıza çıkmaktadır. Sosyo-Ekonomik dağılımının beklenildiği gibi homojen şehirlerden oluşmadığı görülmektedir. Bazı benzer yapıdaki şehirlerin aynı kümede yer almalarına karşın genel olarak bazı kümelerde küme içi farklılıklar göze çarpmaktadır.

4. Sonuç ve Tartışma

Bu çalışmada gri kümeleme analizi yardımıyla Türkiye'deki iller 2019 yılına ait sosyo-ekonomik verileri kullanılarak kümelere ayrılmıştır. Sosyo-ekonomik veri başlığı altında 30 değişken ile yapılan analizlerin yanı sıra sosyo-ekonomik başlığı altındaki sırasıyla demografik, eğitim, sağlık, tarım, nüfus ve finans alt başlıkları için de ayrı ayrı gri kümeleme analizi uygulanmış ve yorumlanmıştır.

Sonuçlara genel olarak bakıldığında Gri kümeleme analizi ile belirlenen küme sayılarının Tarım-Hayvancılık alt kategorisi hariç 7 ve 7'den büyük oldukları görülmektedir. Bu sonuçların hepsinde de ayrıca en az 1 şehrin tek başına küme oluşturdukları görülmektedir. Genelde İstanbul ilinin gerek nüfus yapısı gerekse ekonomik yapısı dolayısıyla tek başına küme oluşturduğu çalışmalara nazaran bu çalışmada Karabük, Gümüşhane, Van, Konya gibi illerin farklı kategorilerde tek başlarına küme oluşturdukları görülmektedir. Eski dönemlerde Devlet Planlama Teşkilatı, günümüzde Kalkınma Ajanslarının yaptığı Türkiye'deki illerin sınıflandırma çalışmalarında 5 veya 6 küme yapısı görülürken bu çalışmada hesaplanan 7 veya üzeri küme oluşumlarının gri kümeleme analizinde kullanılan gri ilişkisel matris hesaplamalarının etkisi olduğu düşünülmektedir. Özellikle tüm değişkenlerin kullanıldığı Sosyo-Ekonomik verilerin değerlendirilmesinde oluşan küme yapılarının bir bütünlük göstermemesi ve hemen her bölgeden benzer sosyo-kültürel ve büyüklükteki şehirlerin farklı kümelerde yer alması gri kümeleme analizinin farkı olarak görülmektedir. Kümeleme, diskriminant analizi gibi sınıflama çalışmalarında genellikle değişken başına 20 gözlemin olması önerilir. Bunun daha azı nispeten belirsizlik sorununa yol açabilir. Bu çalışmada kullanılan 30 değişkene sadece 81 gözlem kullanılmıştır. Bu ise kümeleme analizinde belirsiz sonuçların oluşmasına sebep olabileceği düşünülmektedir. Gri kümeleme analizi ile yapılan sınıflandırmaların, belirsizlik altında karar vermede yardımcı olduğu düşünülmektedir.

Bu çalışmada Gri Kümeleme analizi ile elde edilen küme sayılarının Türkiye sosyo-ekonomik verileri ile yapılan benzer çalışmalara göre daha farklı sonuçlar elde edildiği görülmüştür. Gri kümeleme analizinde küme sayısının belirlenmesi için henüz belirgin yöntemlerin bulunmaması ve gözlem sayısı-değişken sayısı farkının az olmasının bu farklılıklarda etkisi olduğu düşünülmektedir. Gri kümeleme analizinin etkinliğinin test edilmesi için literatürde sıklıkla kullanılan veriler ile benzer çalışmaların yapılarak gerek yöntem hakkında gerekse küme sayısının belirlenmesi hakkında daha detaylı bilgiler elde edilebilecektir.

Kaynaklar

- Albayrak, A. S. (2005). Türkiye'deki illerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi, *ZKÜ Sosyal Bilimler Dergisi*, Cilt:1 Sayı:1.
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*, Academic Press, New York.
- Balasko, B., Abonyi, J., Feil, B. (2005). *Fuzzy clustering and data analysis toolbox*. Department of Process Engineering, University of Veszprem, Veszprem.
- Birkalan, A. Ö., Bay, H. (2022). Bölgesel Asgari Ücret Varsayımı Altında Bölgelerin Kümeleme Analizi Yöntemleriyle Tespiti. *Maliye Dergisi*, Temmuz-Aralık; 183: 24-48.
- Bulut, H. (2019). Türkiye'deki illerin yaşam endekslerine göre kümelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23 (1), 74-82.
- Chang, K. C., Yeh, M. F. (2005). Grey relational analysis-based approach for data clustering, *IEE Proceedings - Vision Image and Signal Processing*, 152 (2), 165-172.
- Çelik, Ş. (2013). Kümeleme analizi ile sağlık göstergelerine göre Türkiye'deki illerin sınıflandırılması. *Doğuş Üniversitesi Dergisi*, 14 (2), 175-194.
- Deng, J. L. (1982). Control Problems of Grey Systems, *Systems & Control Letters*, 1 (5), 288-294.
- Erilli, N.A. (2009). *Kümeleme Analizine Bulanık Yaklaşım Algoritmaları ve Uygulamalar*, Yayınlanmamış Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Samsun.
- Erilli, N. A., Tunç, T., Yüksel, Ö., Yolcu, U. (2009). İllerin Sosyoekonomik Verilere Dayanarak Bulanık Kümeleme Analizi ile Sınıflandırılması, *e-Journal of New World Sciences Academy*, 4(1), 1-11.
- Filiz, Z. (2005). İllerin sosyo-ekonomik düzeylerine göre gruplandırılmasında farklı yaklaşımlar, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(1), 77-100.
- Hamurcu, M., Tamer, E. (2015). Ankara Büyükşehir Belediyesi'nde çok ölçütlü karar verme yöntemi ile monoray güzergâh seçimi, *Transist*, 8 s. 410-419.
- Jain, A.K., Dubes, R.C. (1988). *Algorithms for clustering data*, Prentice Hall, New Jersey.
- Kalaycı, Ş. (2005). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Asil Yayıncılık, Ankara.
- Kandemir, A. Ş. (2018). Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Journal of Travel and Hospitality Management*, 15(3), 657-668.
- Kaygısız, Z., Saraçlı, S., Dokuzlar, K., (2005). İllerin Gelişmişlik Düzeyini Etkileyen Faktörlerin Path Analizi ve Kümeleme Analizi İle İncelenmesi, *VII. Uluslararası Ekonometri ve İstatistik Sempozyumu*, İ. Ü., Ekonometri Bölümü, İstanbul, 26-27 Mayıs.
- Ke, L., Xiaoliub, S., Zhongfua, T., Wenyanb, G. (2012). Grey Clustering Analysis Method for Overseas Energy Project Investment Risk Decision, *Systems Engineering Procedia* 3 (3), 55-62.
- Kohonen, T. (1995). *Self-organizing maps*, Springer-Verlag, New York.
- Köse, E., Aplaç H. S., Kabak, M. (2013). Personel Seçimi için Gri Sistem Teori Tabanlı Bütünleşik Bir Yaklaşım, *Ege Akademik Bakış*, 13(4), 461.
- Lin, Z. (2015). A Multi-attribute Evaluation Model of the Development Competitiveness of Tourism cities based on Grey Clustering Analysis, *International Journal of Hybrid Information Technology*, 8(10), 331-338.
- Liu, S., Forrest J. Y. L. (2010). *Grey systems: Theory and applications*. Springer-Verlag Berlin Heidelberg.
- Mulder, N.J. (1985). Decision Making and Classification. *Photogrammetria*, 40. 95-116.

Citation / Atıf: KARADAŞ K., ERİLLİ N. A. (2023). Gri Kümeleme Analizi ile Türkiye'deki Şehirlerin Sosyo-Ekonomik Verilere Göre Sınıflandırılması. *İstatistik Araştırma Dergisi*, 13 (2), 60-74.

Rousseuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*. 20: 53- 65.

Tatlıdil, H. (2002). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Akademi Matbaası, Ankara.

Tekin, B. (2015). Temel Sağlık Göstergeleri Açısından Türkiye'deki İllerin Gruplandırılması: Bir Kümeleme Analizi Uygulaması. *Çankırı Karatekin Üniversitesi İİBF Dergisi*, 5(2), 389-416.

URL-1: www.tuik.gov.tr (Erişim tarihi: 10.10.2022)

Wong, C. C., Chen, C. C., Su, M. C. (2001). A novel algorithm for data clustering, *Pattern Recognition*, 34(2), 425-442.

Xie, L., Beni, G. (1991). A Validity Measure for Fuzzy Clustering, *IEEE Trans. On Pattern Analysis and Machine Int.* 13(4), 841-846.

Zorlutuna, Ş., Erilli, N. A. (2018). Sosyo-Ekonomik Verilere göre İllerin Bulanık C-Ortalamalar Yöntemi ile Sınıflandırılması: 2002-2008-2013 Dönemleri Karşılaştırması. *İktisadi Yenilik Dergisi*, 5 (2), 13-31.

Ek 1. Değişken gruplarına göre Silhouette ve Xie-Beni indeks değerlerinin farklı küme sayılarına göre değerleri

| Küme Sayısı | Sosyo-Ekonomik | | Demografik | | Eğitim | | Sağlık | |
|-------------|----------------|-------------|---------------|-------------|---------------|-------------|---------------|-------------|
| | Silh. | XB | Silh. | XB | Silh. | XB | Silh. | XB |
| 2 | 0,6552 | 9,87 | 0,7165 | 10,41 | 0,6933 | 21,45 | 0,7362 | 8,25 |
| 3 | 0,6475 | 10,01 | 0,7453 | 14,78 | 0,6984 | 18,75 | 0,7481 | 6,99 |
| 4 | 0,7142 | 8,88 | 0,7586 | 8,61 | 0,7045 | 12,36 | 0,7259 | 11,67 |
| 5 | 0,6968 | 9,85 | 0,7438 | 10,28 | 0,7068 | 11,57 | 0,7148 | 7,82 |
| 6 | 0,6940 | 6,35 | 0,7781 | 3,79 | 0,7166 | 13,4 | 0,7236 | 7,4 |
| 7 | 0,7304 | 4,38 | 0,7912 | 4,89 | 0,7257 | 8,9 | 0,7369 | 8,52 |
| 8 | 0,7161 | 5,88 | 0,7986 | 2,99 | 0,7235 | 6,01 | 0,7389 | 9,91 |
| 9 | 0,6832 | 7,91 | 0,7864 | 5,08 | 0,7135 | 6,94 | 0,7301 | 13,34 |
| 10 | 0,6936 | 10,09 | 0,7784 | 3,87 | 0,7194 | 7,99 | 0,7291 | 12,78 |

Ek 1. Değişken gruplarına göre Silhouette ve Xie-Beni indeks değerlerinin farklı küme sayılarına göre değerleri (Devam)

| Küme Sayısı | Tarım-Hayvancılık | | Nüfus-Göç | | Finans | |
|-------------|-------------------|-------------|---------------|-------------|---------------|-------------|
| | Silh. | XB | Silh. | XB | Silh. | XB |
| 2 | 0,6833 | 16,87 | 0,6418 | 15,76 | 0,7095 | 12,54 |
| 3 | 0,6912 | 14,5 | 0,6568 | 10,54 | 0,7066 | 9,89 |
| 4 | 0,6988 | 12,31 | 0,6591 | 11,86 | 0,7032 | 12,36 |
| 5 | 0,7064 | 8,61 | 0,6601 | 6,37 | 0,6981 | 10,01 |
| 6 | 0,7099 | 5,91 | 0,6734 | 8,56 | 0,6988 | 8,67 |
| 7 | 0,7092 | 3,22 | 0,6781 | 3,78 | 0,6881 | 4,99 |
| 8 | 0,7051 | 6,45 | 0,6824 | 2,73 | 0,7011 | 5,98 |
| 9 | 0,6984 | 6,75 | 0,6818 | 4,02 | 0,7182 | 3,21 |
| 10 | 0,6831 | 8,88 | 0,6799 | 6,48 | 0,7194 | 3,01 |

Ek 2. Kümeleme analizi sonuçlarına göre ortaya çıkan kümelerin ilgili değişkenlere göre tanımlayıcı istatistikleri

| | Değişkenler | 1.Küme | 2.Küme | 3.Küme | 4.Küme | 5.Küme | 6.Küme | 7.Küme | 8.Küme | 9.Küme | 10.Küme |
|--|---|---|---------|---------|---------|-----------|---------|---------|---------|---------|----------|
| Demografik | Bebek Ölüm Hızı (binde) | 4,75 | 11,4462 | 7,43333 | 8,55789 | 6,24 | 3 | 10,1667 | 14,2286 | | |
| | 5 yaş altı ölüm hızı (binde) | 6,3 | 14,0385 | 8,88 | 10,7789 | 7,7866667 | 3 | 12,1889 | 17,1286 | | |
| | Toplam doğurganlık hızı (çocuk sayısı) | 1,455 | 2,52308 | 1,54533 | 1,91421 | 1,6153333 | 1,41 | 1,85556 | 2,57429 | | |
| | Kaba evlenme hızı (binde) | 5,625 | 6,91692 | 6,108 | 6,69842 | 6,0426667 | 5,74 | 6,14 | 6,71714 | | |
| | Kaba ölüm hızı (binde) | 8,2 | 4,53077 | 6,92 | 5,85789 | 7,0266667 | 7,1 | 5,77778 | 4,24286 | | |
| | Kaba doğum hızı (binde) | 9,45 | 19,3615 | 10,9333 | 14,1895 | 11,313333 | 9,4 | 13,8778 | 19,9429 | | |
| | Kaba boşanma hızı (binde) | 1,755 | 1,06769 | 2,07533 | 1,68579 | 1,78 | 1,9 | 1,19444 | 0,98286 | | |
| | Doğuşta beklenen yaşam süreleri (yıl) | 78,8 | 77,8308 | 78,34 | 78,2316 | 78,486667 | 78,4 | 78,5556 | 77,8286 | | |
| | Nüfusa göre boşanma oranları | 572,18 | 1406,9 | 520,928 | 681,014 | 609,76938 | 526,394 | 1065,24 | 1845,27 | | |
| | Eğitim | İlkokul okullaşma oranı 2012 ve sonrası (kadın)-net (%) | 92,472 | 89,765 | 94,542 | 93,207 | 91,272 | 93,939 | 86,780 | 68,410 | |
| İlkokul okullaşma oranı 2012 ve sonrası-net (%) | | 92,677 | 90,010 | 94,560 | 93,416 | 91,712 | 93,975 | 87,215 | 68,200 | | |
| İlkokul /Öğretmen başına düşen öğrenci sayısı | | 15,556 | 15,500 | 15,381 | 14,778 | 15,286 | 15,571 | 14,500 | 13,000 | | |
| İlköğretim okullaşma oranı 2012 ve sonrası (kadın) | | 97,243 | 94,423 | 97,941 | 97,439 | 97,847 | 97,642 | 92,325 | 70,330 | | |
| İlköğretim okullaşma oranı 2012 ve sonrası-net (%) | | 97,264 | 94,325 | 97,823 | 97,349 | 97,870 | 97,532 | 92,435 | 70,180 | | |
| Okuma yazma bilen oranı (%) | | 95,864 | 95,573 | 97,788 | 96,604 | 94,062 | 97,201 | 96,710 | 96,330 | | |
| Ortaokul /Öğretmen başına düşen öğrenci sayısı | | 13,222 | 12,500 | 13,524 | 12,667 | 13,357 | 12,857 | 12,500 | 9,000 | | |
| İlk ve ortaokul /Derslik başına düşen öğrenci sayısı | | 21,111 | 18,250 | 20,524 | 17,667 | 20,429 | 20,619 | 23,000 | 13,000 | | |
| Sağlık | Yüzbin kişi başına toplam hastane yatak sayısı | 386,25 | 232,371 | 307,524 | | | | | | | |
| | Bin kişi başına düşen toplam hekim sayısı | 3 | 1 | 2 | | | | | | | |
| Tarım ve Hayvancılık | Küçükbaş hayvan sayısı (baş) | 1895245 | 403085 | 1143401 | 110349 | 2708012 | 2459960 | 738916 | | | |
| | Tahullar ve diğer bitkisel ürünlerin üretim miktarı (ton) | 2315467 | 1224438 | 1776096 | 438517 | 1707764 | 1,6E+07 | 2131025 | | | |
| | Büyükbaş hayvan sayısı (baş) | 417158 | 214520 | 245221 | 111866 | 180693 | 927082 | 295044 | | | |
| Nüfus-Göç | Net göç hızı (binde) | -55,08 | 8,75364 | -12,455 | 4,93 | -101,31 | -19,579 | -5,5544 | 0,92429 | | |
| | Nüfus yoğunluğu (kilometrekareye düşen kişi sayısı) | 65,63 | 75,6227 | 57,08 | 2986,77 | 54,58 | 104,13 | 87,6389 | 132,997 | | |
| | Ortalama hanehalkı büyüklüğü | 2,76 | 3,46091 | 3,30273 | 3,33 | 2,63 | 3,27824 | 3,69667 | 3,51143 | | |
| | Yıllık nüfus artış hızı (binde) | -12,22 | 3,95545 | -16,018 | 29,53 | 2,76 | 5,38176 | 4,63667 | 8,50333 | | |
| | Türkiye'den yurt dışına giden göç | 1071 | 3863,91 | 1155,18 | 140223 | 2014 | 3338,29 | 1243,06 | 2506,9 | | |
| Finans | Toplam ithalat (bin \$) | 756186 | 258384 | 3208794 | 1310107 | 7854288,7 | 40513,9 | 2397621 | 5006595 | 1,1E+07 | 1,05E+08 |
| | Toplam ihracat (bin \$) | 916187 | 390852 | 4054981 | 1590084 | 10122356 | 123260 | 2187391 | 7418005 | 8313060 | 85636418 |
| | Kişi başına GSYH (TL) | 45092 | 41131,6 | 40828 | 43899,6 | 66913 | 34230,6 | 43144 | 37633 | 71027 | 86798 |