

BULANIK KÜMELEME ANALİZİ ile İSTATİSTİKİ BÖLGE BİRİMLERİNİN (İBBS) MALİ DEĞİŞKENLERE GÖRE SINIFLANDIRILMASI

Necati Alp ERİLLİ*

ÖZET

Avrupa Birliği; üye ülkelerin ulusal kalkınma düzeylerinin yanı sıra, bölgesel kalkınma politikalarının da geliştirilmesi amacıyla ülkelerin alt bölgelere ayrılarak bölgesel bazda istatistiklerinin tutulması konusunda çalışmalar başlatmıştır. Bu çalışmalar doğrultusunda Türkiye’de; Düzey 1, Düzey 2 ve Düzey 3 olmak üzere 3 alt bölgeye ayrılarak, bölgesel istatistikler toplanmaya ve geliştirilmeye başlanmıştır.

Bulanık Kümeleme yaklaşımı, kümeler birbirinden belirgin bir şekilde ayrılmıyorsa ya da üyeliklerinde bazı birimler küme üyeliğinde kararsızsa uygun bir yöntem olarak ortaya çıkmaktadır. Bulanık Kümeler, kümedeki birimin üyeliği olarak tanımlanan 0 ile 1 arasındaki her birimi belirleyen fonksiyonlardır. Birbirine çok benzeyen birimler aynı kümede yüksek üyelik derecesine göre yer alırlar. Burada amaç; bu göstergeler bakımından aynı özellikleri taşıyan homojen il gruplarının tespit edilmesidir.

Bu çalışmada, bulanık kümeleme analizinde sıkça kullanılan bulanık c-ortalamalar (FCM) yöntemi kullanılarak, İBBS Düzey-2 ve Düzey-3 bölgeleri mali göstergeler yardımıyla gelişmişlik düzeylerine göre sınıflandırılmıştır. Elde edilen sonuçların, birbirleri ile korelasyonu incelenmiş ve yorumlanmıştır.

Anahtar Kelimeler: Bulanık Kümeleme, Düzey-2, Düzey-3, İBBS

* Yrd.Doç.Dr., Cumhuriyet Üniversitesi,İİBF, Ekonometri Bölümü, aerilli@cumhuriyet.edu.tr

CLASSIFICATION OF STATISTICAL AREA UNITS (SCTU) ACCORDING TO FINANCIAL VARIABLES WITH FUZZY CLUSTERING ANALYSIS

ABSTRACT

European Union has started works on keeping regional statistics to develop regional development politics as well as national development levels of member countries by dividing countries into sub regions. In line with these works, Turkey has been divided into three sub regions as Level 1 , Level 2 and Level 3 and regional statistics have begun to be collected and developed.

Fuzzy Clustering approach comes out as a suitable method if the clusters are not separated from each other prominently or if some of the members are indecisive about being a member of the cluster. Fuzzy clusters are functions that determine each unit between 0 and 1 defined as the membership of the unit. Units which are very similar take part in the same cluster according to high membership degree. The purpose here is to determine homogenous city groups that have the same characteristics in terms of these indicators.

This study has used fuzzy c-means (FCM) method which is used frequently in fuzzy clustering analysis and has classified SCTU Level-2 and Level-3 regions through financial indicators. The results have been examined and interpreted in terms of their correlation with each other.

Key Words: *Fuzzy Clustering, Level-2, Level-3, SCTU*

GİRİŞ

Avrupa Birliği için üye ülkelerin ulusal kalkınma düzeylerinin yanı sıra, bölgesel kalkınma politikaları da çok büyük bir öneme sahiptir. AB’nce üye ülkelere, bölgeler arası gelişmişlik farklılıklarını azaltabilmek ve bölgelerin refah düzeylerini geliştirebilmek amacıyla finansal kaynaklar sağlanmaktadır. Bölgesel düzeydeki ekonomik ve sosyal sorunlara çözüm bulmak amacıyla AB, artan bir şekilde çeşitli düzeylerde bölgesel istatistiklere gerek duymaktadır.

Türkiye Avrupa Birliği’ne aday üye bir ülke olarak her alanda olduğu gibi istatistiki alanda da Avrupa Birliği’ne üye ülkelerin yükümlülüklerini yerine getirmeye çalışmaktadır (Taşkan, 2006).

Bu bağlamda aday üye ülkelerce hazırlanması gereken ve İstatistiki Bölge Birimleri Sınıflaması olarak adlandırılan sınıflama 2001 yılı içerisinde TÜİK ve DPT tarafından hazırlanmış olup Bakanlar Kurulu’nun 2002/4720 sayılı Kararı ile 22 Eylül 2002 tarihli Resmi Gazetede yayımlanmıştır.

İBBS AB’nin talepleri doğrultusunda, Ulusal Programa dahil edilmiştir. Bu kapsamda, Türkiye tarafından hazırlanması gereken “AB Müktesebatının Üstlenilmesine İlişkin Türkiye Ulusal Programı”, Bakanlar Kurulu tarafında 13 Mart 2001 tarihinde kabul edilmiş ve 24 Mart 2001 tarihli Resmi Gazete’de yayımlanarak yürürlüğe girmiştir.

İBBS’nin hazırlanmasındaki başlıca amaçlar; bölgesel istatistiklerin toplanması ve geliştirilmesi, bölgelerin sosyo-ekonomik analizlerinin yapılması, bölgesel politikaların çerçevesinin belirlenmesi, Avrupa Birliği Bölgesel İstatistik Sistemine uygun, karşılaştırılabilir bir istatistiki veri tabanı oluşturulmasına katkıda bulunacak temel göstergeleri ortaya koymak şeklinde sıralanabilir.

Genel olarak İBBS’ler hazırlanırken iki ana ölçüt göz önüne alınmıştır: Her bir üst düzeyin kendi içinde alt düzey birimlerinin birleştirilerek oluşturulması zorunluluğu ve bölgesel birimlerin birbirleri ile coğrafi sınırlar bakımından bütünlük oluşturacak bir süreklilik içinde olmaları (Taşkan, 2006).

İstatistiki Bölge Birimleri Sınıflaması’na göre; Düzey 1 olarak 12, Düzey 2 olarak 26 ve Düzey 3 (iller) olarak 81 İstatistiki Bölge Birimi tanımlanmıştır.

İBBS oluşturulurken önce illerin Düzey 3 olmasına karar verilmiştir. Bunun üç temel nedeni vardır. Türkiye’de temel idari birimlerin illerden oluşması, istatistiki verilerin çoğunlukla il bazında derlenmesi (Nüfus, tarım, sanayi vb.) ve AB’ye üye ülkelerdeki NUTS 3 birimlerinin nüfus büyüklüğü ve ele alınan birim sayısı ile göz önüne alındığında Türkiye için iller uygun olmasıdır.

Düzy 3'lerin tanımlanmasının ardından Düzy 2'ler belirlenmiştir. NUTS 2 bölgeleri AB yardımlarından en fazla yardımı alacak olan geri kalmış bölgelerin belirlendiği birim olduğundan Düzy 2'lerin belirlenmesi sırasında, ortak sorunlara sahip, sosyoekonomik ve kültürel olarak birbirine yakın ve coğrafi olarak benzer özellikler gösteren illerin gruplandırılmasına özen gösterilmiştir. Tablo.1'de Düzy-2 bölgeleri ve merkez illeri verilmiştir.

Tablo.1 Düzy-2 Bölgeleri

Kod	Düzy 2	Kod	Düzy 2
TR10	İstanbul	TR71	Kırıkkale
TR21	Tekirdağ	TR72	Kayseri
TR22	Balıkesir	TR81	Zonguldak
TR31	İzmir	TR82	Kastamonu
TR32	Aydın	TR83	Samsun
TR33	Manisa	TR90	Trabzon
TR41	Bursa	TRA1	Erzurum
TR42	Kocaeli	TRA2	Ağrı
TR51	Ankara	TRB1	Malatya
TR52	Konya	TRB2	Van
TR61	Antalya	TRC1	Gaziantep
TR62	Adana	TRC2	Şanlıurfa
TR63	Hatay	TRC3	Mardin

Son olarak Düzy 1'lerin oluşturulmasında da bu ölçütler göz önüne alınmış ve Düzy 2 bölgelerinin bir araya getirilerek kendi aralarında gruplandırılması ile Düzy 1'ler hazırlanmıştır.

Tablo.2 Düzy-1 Bölgeleri

Kod	Düzy-1
TR1	İstanbul
TR2	Batı Marmara
TR3	Ege
TR4	Doğu Marmara
TR5	Batı Anadolu
TR6	Akdeniz
TR7	Orta Anadolu
TR8	Batı Karadeniz
TR9	Doğu Karadeniz
TRA	Kuzeydoğu Anadolu
TRB	Ortadoğu Anadolu
TRC	Güneydoğu Anadolu

Sonuç olarak, topluluğun yapısal fonlarına temel teşkil eden, bölgesel istatistiklerin toplanması, geliştirilmesi ve uyumlu hale getirilmesinin yanı sıra bölgelerin sosyoekonomik analizlerinin yapılmasına olanak sağlayan istatistiki bölgeler hazırlanmıştır. Örnek olarak Batı Karadeniz bölgesine ait Düzey 1, Düzey 2 ve Düzey 3 bilgileri Tablo.3’de verilmiştir.

Tablo.3 Batı Karadeniz İBBS

Kod	Düzey-1	Düzey-2	Düzey-3
TR8	Batı Karadeniz		
TR81		Zonguldak	
TR811			Zonguldak
TR812			Karabük
TR813			Bartın
TR82		Kastamonu	
TR821			Kastamonu
TR822			Çankırı
TR823			Sinop
TR83		Samsun	
TR831			Samsun
TR832			Tokat
TR833			Çorum
TR834			Amasya

Bulanık Kümeleme Analizi

Bu yaklaşımda, kümeler birbirinden belirgin bir şekilde ayrılmıyorsa ya da üyeliklerinde bazı birimler küme üyeliğinde kararsızsa uygun bir yöntem olarak ortaya çıkmaktadır. Bulanık Kümeler kümedeki birimin üyeliği olarak tanımlanan 0 ile 1 arasındaki her birimi belirleyen fonksiyonlardır. Birbirine çok benzeyen birimler aynı kümede yüksek üyelik derecesine göre yer alırlar (Erilli, 2009).

Diğer kümeleme yöntemlerine benzer olarak Bulanık Kümeleme de uzaklık ölçümlerine dayanır. Bu uzaklık ölçütlerinden hangisinin seçileceği küme yapısına ve kullanılan algoritmaya bağlıdır. Bulanık Kümelemenin kullanışlı bazı özelliklerini şu şekilde sıralayabiliriz:

- i. Yorum açısından kullanışlı olan üyelik değerleri sağlar.
- ii. Uzaklık kullanımı konusunda esneklik.
- iii. Üyelik değerlerinin bazıları bilindiğinde sayısal optimizasyonla birleştirilebilir (Naes ve Mevik, 1999).

Bulanık Kümelemenin klasik kümeleme yöntemlerine göre avantajı, veri hakkında daha detaylı bilgi vermesidir. Diğer taraftan dezavantajları da vardır. Çok sayıdaki birey ve küme durumunda çok fazla çıktı olacağından, özetlemek ve bilgiyi tasnif etmek zordur. Ayrıca bulanık kümeleme algoritmaları genellikle karmaşık yapıdadırlar ve daha çok belirsizlik söz konusu olduğunda kullanılır (Şahinli, 1999).

Bulanık C-Ortalamlar (BCO) Algoritması

Bulanık C-Ortalamlar algoritması, amaç fonksiyonuna dayanan bütün kümeleme tekniklerinin temelini oluşturmaktadır. Bezdek (1974) tarafından geliştirilmiştir. BCO algoritması sonuçlandığında, p boyutlu uzaydaki noktalar küresel bir şekil halini alır. Bu kümelerin yaklaşık olarak aynı boyutta olduğu varsayılır. Her bir kümeyi, küme merkezleri temsil eder ve bunlara prototip denir. Uzaklık ölçüsü olarak veriler ile küme merkezi arasındaki Öklid uzaklığını kullanır.

$$d_{ik} = d(x_i, v_k) = \left[\sum_{t=1}^p (x_{ji} - v_{jk})^2 \right]^{\frac{1}{2}}$$

Burada x_i gözlem değerinin koordinat sistemindeki konumunu, v_k ise küme merkezini simgelemektedir.

Bu tekniğin uygulanabilmesi için küme sayısının ve bireylerin kümeye üyelik derecelerinin önceden bilinmesi gerekmektedir. Bu tür parametrelerin önceden bilinmesi zor olduğundan, bu değerler deneme yanılma yoluyla ya da geliştirilen bazı tekniklerle bulunabilir.

Bu kümeleme yöntemi için kullanılan amaç fonksiyonu şu şekildedir:

$$J(u, v) = \sum_{j=1}^n \sum_{k=1}^c u_{jk}^m \|x_{ji} - v_{jk}\|^2$$

Bu fonksiyon ağırlıklandırılmış en küçük kareler fonksiyonudur. n parametresi gözlem sayısını, c ise küme sayısını gösterir. u_{jk}^m ise k. kümedeki x_j 'nin üyeliği, $J(u, v)$ değeri ise tüm ağırlıklandırılmış hata karelerinin toplamının bir ölçüsüdür (Sintas vd., 1999).

Eğer $J(u, v)$ fonksiyonu c'nin her değeri için minimize edilecek olursa, diğer bir deyişle v_i 'lere göre 1. dereceden türevi alınıp 0'a eşitlenirse BCO Algoritmasının prototipi şu şekilde olacaktır;

$$v_{jk} = \sum_{j=1}^n u_{jk}^m \cdot x_{ik} \quad \frac{1}{\sum_{j=1}^n u_{jk}^m}$$

BCO Algoritması için gerekli adımlar ise şu şekildedir:

Adım 1: Başlangıç değerlerini belirle: Küme sayısı c , bulanıklık indeksi m , işlem bitirme kriteri ε ve üyelik dereceleri matrisi U veya V küme prototiplerini rasgele üretilir.

Adım 2: U küme prototiplerinin rasgele üretildiği varsayılırsa bu değerleri kullanarak üyelik dereceleri matrisini hesaplanır.

$$u_{jk} = \left[\sum_{j=1}^c \left(\frac{d_{ji}}{d_{jk}} \right)^{\frac{2}{m-1}} \right]^{-1}$$

Adım 3: Adım 2 eşitliğine göre U küme prototiplerini güncellenir.

Adım 4: $\|U^{(t)} - U^{(t-1)}\| \leq \varepsilon$ ise iterasyon durdurulur, aksi takdirde Adım 2'ye geri dönlür.

BCO Algoritması uygulandıktan sonra hangi bireyin hangi kümeye gireceğine karar vermek için üyelik dereceleri kullanılır. Her bir bireyin hangi kümeye olan üyeliğinin en büyük olduğuna bakılır ve bu bireyler o kümeye dâhil edilir. Ancak her bir birey diğer kümelere de belli bir üyelik dereceleri ile girebilir.

BCO Algoritmasının sonucu başlangıçta rasgele üretilen değerlere oldukça bağlıdır. Bu yüzden rasgelelikten kaynaklanan problemleri ortadan kaldırmak için çeşitli algoritmalar geliştirilmiştir ve geliştirilmeye devam edilmektedir.

BCO, küme merkezlerini ve her veri noktası için üyelik derecelerini iterasyon yöntemi ile günceller ve küme merkezlerini veri seti içinde olması gereken yere taşır.

Küme merkezlerinin ilk yerleri, başlangıçta değeri rasgele atanan U matrisi kullanılarak oluşturulduğu için, BCO optimal sonuca yaklaşmayı garanti etmeyecektir (Sintas vd., 1999).

Performans; merkezlerin başlangıç yerlerine bağlıdır. Daha güçlü bir yaklaşım için aşağıda tanımlanan iki yol vardır.

- i. Tüm merkezleri tanımlamak için bir algoritma kullanmak.
- ii. BCO'yı farklı başlangıç merkezleri ile tekrarlı olarak çalıştırmak (Yeniden Başlama Stratejisi).

Bulanık Kümeleme Geçerlilik İndeksleri

Kümeleme Analizi, benzer nesnelere aynı gruplara yerleştirmeyi amaçlamaktadır. Birçok kümeleme algoritması küme sayısının önceden bilinmesini gerektirir. Gerçek verilere dayalı çalışmalarda; araştırmacının küme sayısı hakkında ön bilgisinin olmaması, bulunan küme sayısının gerçek küme sayısından az ya da çok olup olmadığının bilinmemesine yol açmaktadır. Optimal küme sayısının belirlenme işlemlerine genel olarak Küme Geçerliliği (Cluster Validity) adı verilmektedir. Böylece kümeleme işlemleri yapıldıktan sonra bulunan küme sayısının doğruluğunu tespit edilebilmektedir (Erilli, 2009).

Veriler iki boyutlu uzayda olduğunda küme sonuçlarını görsel olarak yorumlayarak küme sayısına karar verilebilmektedir. Ama uzaydaki boyut sayısı arttıkça görsellik zorlaşmakta ve geçerlilik indekslerine ihtiyaç duyulmaktadır.

Sonuç olarak, kümeleme değeri ve en uygun kümeleme planlaması için iki kriterden bahsedilebilir.

1.Yoğunluk: Küme elemanlarının birbirlerine yakınlıklarını ölçer. Buna en iyi örnek olarak varyansı verebiliriz.

2.Ayrılma: İki kümenin birbirlerinden ne kadar ayrıldıklarını gösterir. İki farklı küme arasındaki mesafeyi ölçer.

Literatürde birçok bulanık kümeleme analizi geçerlilik indeksi kullanılmaktadır (Bezdek, 1974 ve 1981; Rezaee v.d., 1998; Kwon, 1998; Xie ve Beni, 1991). Veri yapısına veya değişken sayısına bağlı olarak uygun kümeleme geçerlilik analizi kullanılmaktadır. Bu çalışmada Yapay Sinir Ağlarına Dayalı Küme Geçerlilik İndeksi kullanılmıştır.

YSA'na Dayalı Küme Geçerlilik İndeksi (YSA)

Bu yöntem Erilli v.d. (2011) tarafından önerilmiştir. Yöntem, en uygun küme sayısını bulabilmek için yapay sinir ağlarını kullanmaktadır. Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım olmadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir (Öztemel, 2006). Geleneksel programlama yöntemleri ile gerçekleştirilmesi oldukça zor yöntemlerdir. Geliştirilmiş güçlü bilgisayarlar ve programlar sayesinde, yapay sinir ağları yeni bir bilim dalı olarak kabul görmektedir.

Bu yöntemde öncelikle, veriye uygun olabilecek en düşük ve en yüksek küme sayısına karar verilir. Belirleyeceğimiz en uygun küme sayısı bu aralıkta olacaktır. En uygun küme sayısı c_{opt} , en düşük küme sayısı c_{min} ve en yüksek küme sayısı

c_{maks} ise; $c_{min} \leq c_{opt} \leq c_{maks}$ olacaktır. Daha sonra, girdisi veri matrisi ve hedef değeri bulanık kümeleme sonucunda her bir verinin atandığı küme numarası olacak şekilde ileri beslemeli yapay sinir ağları mümkün küme sayılarının her biri için uygulanır. Her bir küme sayısı için, çeşitli gizli tabaka birim sayılarına göre yapay sinir ağlarından elde edilen HKOK (Hata kareler ortalaması karekök değeri) değerlerinin medyanı hesaplanır. Her bir küme sayısı için elde edilen medyan değerlerinin veya sınıflama hatası grafiği çizilerek, ilk sıçramanın olduğu (HKOK medyan değerinin ilk aşırı büyüdüğü) küme sayısından bir önceki değer en uygun küme sayısı olarak belirlenir.

UYGULAMA

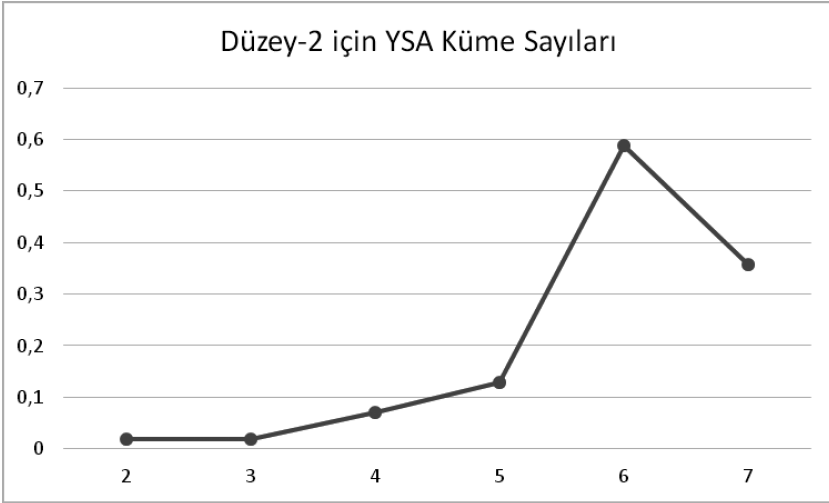
26 Bölge-2 ve 81 Bölge-3 verileri 15 adet mali değişkene göre bulanık kümeleme analizi ile sınıflandırılmıştır. Mali değişkenler; Devlet Planlama Teşkilatı tarafından en son 2003'de açıklanan sosyo-ekonomik verilerin mali değişkenler başlığı altında yer alan 15 değişkenden oluşmaktadır. Uygulamada kullanılan veriler Tablo.4'de verilmiştir.

Tablo.4 2003 Sosyo-Ekonomik Veriler: Mali Değişkenler

Gayri Safi Yurt İçi Hasıla İçindeki Payı
Fert Başına Gayri Safi Yurt İçi Hasıla
Banka Şube Sayısı
Fert Başına Banka Mevduatı
Toplam Banka Mevduatı İçindeki Payı
Toplam Banka Kredileri İçindeki Payı
Kırsal Nüfus Başına Tarımsal Kredi Miktarı
Fert Başına Sınai, Ticari Ve Turizm Kredileri Miktarı
Fert Başına Belediye Giderleri
Fert Başına Genel Bütçe Gelirleri
Fert Başına Gelir Ve Kurumlar Vergisi Miktarı
Fert Başına Kamu Yatırımları Miktarı
Fert Başına Teşvik Belgeli Yatırım Tutarı
Fert Başına İhracat Miktarı
Fert Başına İthalat Miktarı

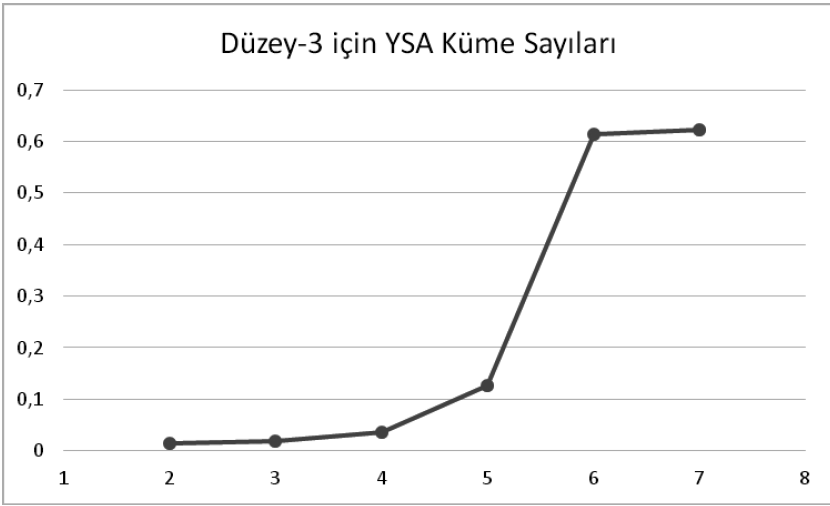
Uygulama, MATLAB2009b paket programında kod yazılarak yapılmıştır. Bulanık kümeleme analizinde ilk önce küme sayısı belirlenmelidir. Bu çalışmada, küme geçerlilik indekslerinden YSA küme geçerlilik indeks değerleri kullanılmıştır. Küme sayılarını belirlemek için, verilere 2’den 10’a kadar ayrı ayrı bulanık kümeleme analizi uygulanmış ve bölgelerin üye oldukları kümeler belirlenmiştir.

İlk olarak Düzey-2 için YSA küme geçerlilik sayıları hesaplanmıştır. Şekil.1’de verilen, YSA’ndan elde edilen HKOK değerlerine baktığımızda, ilk önemli sıçramanın 5. kümede olduğunu görmekteyiz. Böylece Düzey-2 için küme sayısının 5 olması gerektiğinin sonucuna ulaşılmıştır.



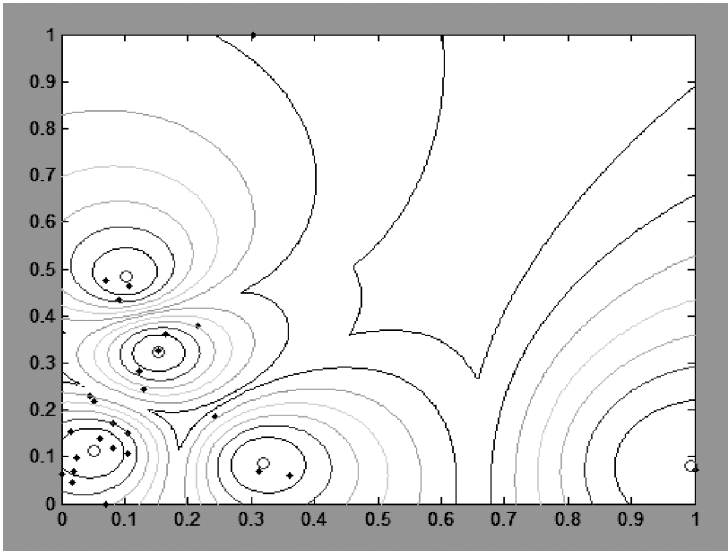
Şekil.1 Düzey-2 için YSA Küme Sayıları

Benzer şekilde Düzey-3 için YSA küme geçerlilik sayıları hesaplanmıştır. Şekil.2’de verilen, YSA’ndan elde edilen HKOK değerlerine baktığımızda, ilk önemli sıçramanın 5. kümede olduğunu görmekteyiz. Böylece Düzey-3 için de küme sayısının 5 olması gerektiğinin sonucuna ulaşılmıştır.



Şekil.2 Düzey-3 için YSA Küme Sayıları

Küme sayıları belirlendikten sonra, mali değişkenlere göre Düzey-2 ve Düzey-3 değişkenleri, ayrı ayrı bulanık kümeleme analizi ile sınıflandırılmıştır. İlk önce Düzey-2 için Bulanık kümeleme analizi yapılmıştır. Düzey-2 için hesaplanan 5 küme dağılım grafiği Şekil. 3’de ve küme sonuçları ise Tablo.5’de verilmiştir.



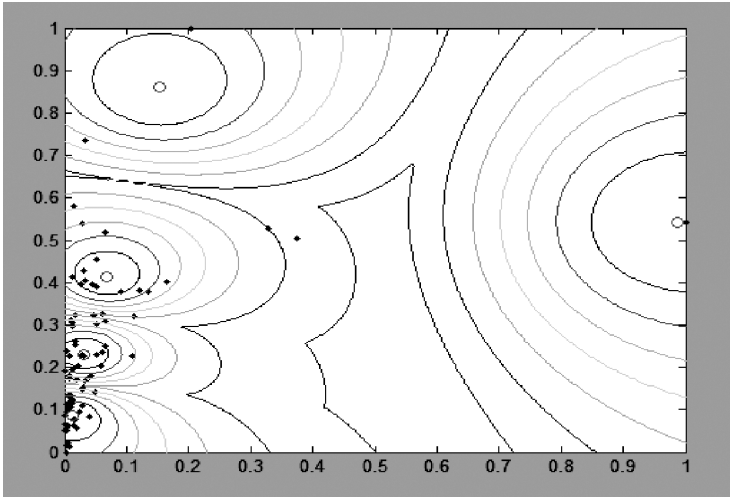
Şekil.3 Düzey-2 için Küme Dağılım Grafiği

Tablo.5 Düzey-2 için bulanık kümeleme sonuçları

1	2	3	4	5
TR10 (İstanbul)	TR31 (İzmir)	TR32 (Aydın, Denizli, Muğla)	TR21 (Tekirdağ, Edirne, Kırklareli)	TR22 (Balıkesir, Çanakkale)
	TR51 (Ankara)	TR33 (Manisa, Afyon, Kütahya, Uşak)	TR42 (Kocaeli, Sakarya, Düzce, Bolu, Yalova)	TR52 (Konya, Karaman)
	TR62 (Adana, Mersin)	TR41 (Bursa, Eskişehir, Bilecik)	TR71 (Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir)	TR81 (Zonguldak, Karabük, Bartın)
		TR61 (Antalya, Isparta, Burdur)	TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane)	TR63 (Hatay, K.Maraş, Osmaniye)
		TR83 (Samsun, Tokat, Çorum, Amasya)		TR72 (Kayseri, Sivas, Yozgat)
				TR82 (Kastamonu, Çankırı, Sinop)
				TRA1 (Erzurum, Erzincan, Bayburt)
				TRA2 (Ağrı, Kars, Iğdır, Ardahan)
				TRB1 (Malatya, Elazığ, Bingöl, Tunceli)
				TRB2 (Van, Muş, Bitlis, Hakkari)
				TRC1 (Gaziantep, Adıyaman, Kilis)
				TRC2 (Şanlıurfa, Diyarbakır)
				TRC3 (Mardin, Batman, Şırnak, Siirt)

Analiz sonuçlarına göre TR10 (İstanbul) bölgesi tek başına birinci kümeyi oluştururken, TR31 (İzmir), TR51 (Ankara) ve TR62 (Adana, Mersin) bölgeleri de ikinci kümeyi oluşturmuşlardır. Bu iki küme gelişmiş şehirlerimizin oluşturdukları bölgelerdir ve sonuçlar beklenildiği gibi bulunmuştur. Düzey-2 bölge verileri; bölgeyi oluşturan şehirlerden oluştuğu için Kayseri, Konya veya Gaziantep gibi büyükşehirler, bölgeyi oluşturan diğer şehirler daha küçük olmalarına ve az gelişmiş olmalarından dolayı ancak beşinci kümede yer almışlardır. Doğu ve güneydoğu illerimizden oluşan bölgelerin tamamı, beşinci kümede yer almışlardır.

İkinci olarak Düzey-3 için Bulanık kümeleme analizi yapılmıştır. Düzey-3 için hesaplanan 5 küme dağılım grafiği Şekil. 4’de ve küme sonuçları ise Tablo.6’da verilmiştir.



Şekil.4 Düzey-3 için Küme Dağılım Grafiği

Tablo.6 Düzey-3 için bulanık kümeleme sonuçları

1	2	3	4	5
İSTANBUL	ADANA	BOLU	AMASYA	ADIYAMAN
	ANKARA	KOCAELİ	ARTVİN	AFYON
	ANTALYA		BALIKESİR	AĞRI
	AYDIN		BURDUR	AKSARAY
	BİLECİK		ÇORUM	ARDAHAN
	BURSA		DENİZLİ	BARTIN
	ÇANAKKALE		ELAZIĞ	BATMAN
	EDİRNE		GAZİANTEP	BAYBURT
	ESKİŞEHİR		GİRESUN	BİNGÖL
	İZMİR		HATAY	BİTLİS
	KIRIKKALE		ISPARTA	ÇANKIRI
	KIRKLARELİ		K.MARAŞ	DİYARBAKIR
	MANİSA		KARABÜK	DÜZCE
	MUĞLA		KARAMAN	ERZİNCAN
	SAKARYA		KASTAMONU	ERZURUM
	TEKİRDAĞ		KAYSERİ	GÜMÜŞHANE
	YALOVA		KIRŞEHİR	HAKKARİ
	ZONGULDAK		KİLİS	İĞDIR
			KONYA	KARS
			KÜTAHYA	MARDİN
			MALATYA	MERSİN
			NEVŞEHİR	MUŞ
			NİĞDE	ORDU
			RİZE	OSMANIYE
			SAMSUN	SİİRT
			SİNOP	SİVAS
			TRABZON	ŞANLIURFA
			TUNCELİ	ŞIRNAK
			UŞAK	TOKAT
				VAN
				YOZGAT

Analiz sonuçlarına göre İstanbul tek başına birinci kümeyi oluşturmuştur. Düzey-2 bölge olarak da tek başına bir küme oluşturan İstanbul şehri, mali değişkenlere göre, diğer şehirlerden ve bölgelerden tek başına ayrılmaktadır. İzmir, Ankara, Bursa, Mersin gibi büyükşehirlerimiz ikinci kümede yer almışlardır. Bu kümede beklenmeyen sonuç olarak Aydın, Bilecik, Kırklareli, Zonguldak illerini gösterebiliriz.

2003 DPT sosyo-ekonomik veriler (57 değişken ile) ile yapılan bulanık kümeleme analizi sonuçlarına göre Aydın 3. kümede, Kırklareli ve Bilecik 4. kümede ve Zonguldak 5. Kümede yer almışlardır (Erilli, 2009). Birbirine benzer yapıdaki veriler ile analiz yapıldığında sonuçlardan bazılarının beklentilerden farklı olduğu görülmektedir.

Bolu ve Kocaeli şehirlerinin 3. kümeyi oluşturması, Gaziantep ilinin 4. kümede yer alması ve Erzurum ilinin de 5. kümede yer alması da beklenmeyen diğer sonuçlar olarak yorumlanabilir.

Uygulamanın son kısmında ise, Düzey-2 ve Düzey-3 bulanık kümeleme analizi sonuçlarına göre, aynı kümede yer alan illerin ilişki analizi yapılmıştır. Buna göre Düzey-2 ve Düzey-3 sonuçları arasındaki korelasyon katsayısı $r=0,56$ bulunmuştur. Bu değer 0,01 önem seviyesinde anlamlı çıkmıştır ($p=0,000$). Düzey-2 ve Düzey-3 bulanık kümeleme analizi sonucunda ilişki katsayısının çok yüksek olmasa da anlamlı çıkması, illerin dağılımının düzeye göre büyük farklılık göstermediğini belirtmektedir.

Tartışma ve Sonuç

İstatistik bilimi her geçen gün önemi daha da artırmaktadır. Hemen her alanda nitelikli ve doğru istatistiki bilgiler toplamak, ülkelerin de birincil amaçları arasına girmiştir. Avrupa Birliği; üye ülkelerin ulusal kalkınma düzeylerinin yanı sıra, bölgesel kalkınma politikalarına da büyük bir önem vermektedir. Böylece birincil amaç; bölgeler arası gelişmişlik farklılıklarını azaltabilmek ve bölgelerin refah düzeylerini geliştirebilmektir. Bölgesel düzeydeki ekonomik ve sosyal sorunlara çözüm bulmak amacıyla AB, çeşitli düzeylerde bölgesel istatistiklere gerek duymaktadır.

Türkiye Avrupa Birliği'ne aday üye bir ülke olarak her alanda olduğu gibi istatistiki alanda da Avrupa Birliği'ne üye ülkelerin yükümlülüklerini yerine getirmeye çalışmaktadır. Bu bağlamda aday üye ülkelere hazırlanması gereken ve İstatistiki Bölge Birimleri Sınıflaması olarak adlandırılan sınıflama 2002 tarihinde Resmi Gazetede yayımlanarak yürürlüğe girmiştir.

Bu çalışmada Düzey-2 ve Düzey-3 bölgeleri, bulanık kümeleme analizi ile

sınıflandırılmıştır. 2003 sosyo-ekonomik veriler: mali değişkenlere göre yapılan sınıflandırma ile Düzey-2 ve Düzey-3 bölgelerinin kümelerine göre dağılımı analiz edilmiştir. Bazı illerin, bilinen sonuçlardan farklı kümelerde yer aldıkları görülmüştür. Değişken sayısının artması ve verilerin değerlerinin birbirlerine çok yakın olması, klasik kümeleme sonuçlarından farklı sonuçlar bulunmasında etkili olmuştur.

Literatürdeki sosyo-ekonomik veriler ile yapılan sınıflama çalışmaları bulunsa da İBBS düzeylerine göre yapılan çalışmalar yenidir. Bu çalışmanın bu alandaki bazı eksiklikleri de gidermesi amaçlanmaktadır.

Son dönemlerde popülerliğini artıran Bulanık Kümeleme ve BCO algoritması, özellikle veri sayısı ve bunlara ait değişken sayıları arttığında daha iyi sonuçlar vermektedir. Verileri sınıflarken karar vermede zorlandığımızda veya sınıflar arası belirgin farklar varsa Bulanık Kümeleme Analizi oldukça etkili sonuçlar verdiği görülmüştür. Sınıflandırma yaparken karışık veri yapıları karşısında diğer kümeleme yöntemlerine göre daha net çözümler üretilebilmektedir.

Kaynaklar

- Bezdek J.C., (1981). *Pattern recognition with fuzzy objective function algorithms*. NY: plenum press.
- Bezdek J. C., (1974). Cluster validity with fuzzy sets. *J. Cybernetics*, 3, 58-73.
- Erilli N.A., Tunç T., Öner Y., Yolcu U, (2009). İllerin Sosyo-Ekonomik Verilere Dayanarak Bulanık Kümeleme Analizi ile Sınıflandırılması. E-Journal of New World Sciences Academy, v.4, n.1.
- Erilli N.A., Yolcu U., Eğrioğlu E., Aladağ Ç.H., Öner Y., (2011). Determining the Most Proper Number of Cluster in Fuzzy Clustering by Artificial Neural Networks. *Expert Systems with Applications*, 38, 2248-2252.
- Erilli, N.A. (2009). Kümeleme Analizine Bulanık Yaklaşım Algoritmaları ve Uygulamaları, 19 Mayıs Üniv., Fen Bilimleri Enst., Yayınlanmamış Yüksek Lisans Tezi, Samsun. Bilimleri Enst., Samsun.
- Kwon S.H., (1998). Cluster Validity Index For Fuzzy Clustering, *Elec. Letters*, 34(22), pp 2176-2178.
- Naes T., Mevik T.H., (1999). The Flexibility of Fuzzy Clustering Illustrated By Examples, *Journal Of Chemo Metrics*.
- Öztemel E., (2006). *Yapay Sinir Ağları*. Papatya Yayıncılık, İstanbul.
- Rezaee M.R., Lelieveldt B.P.F., Reiber J.H.C., (1998). A New Cluster Validity Index for the FCM, *Pattern Recognition Lett.*, 19 p. 237-246.
- Sintas A.F., CAdenas J.M., Martin F., (1999). Membership functions in the Fuzzy c-Means Algorithm, *Fuzzy Sets and Systems* 101.
- Şahinli F., (1999). Kümeleme Analizine Fuzzy Set Teorisi Yaklaşımı, Gazi Üniversitesi, Yayınlanmamış Yüksek Lisans Tezi, Ankara.
- Şengül Ü., Eslemian S., Eren M., (2013). Türkiye’de İstatistiki Bölge Birimleri Sınıflamasına Göre Düzey 2 Bölgelerinin Ekonomik Etkinliklerinin VZA Yöntemi ile Belirlenmesi ve Tobit Model Uygulaması. *Yönetim Bilimleri Dergisi*, Cilt: 11, Sayı: 21, ss. 75-99.
- Taşkan P., (2006). İstatistiki Bölge Birimleri Sınıflandırması (İBBS). tuikapp.tuik.gov.tr/DIESS/FileUpload/yayinlar/5.iBBS.ppt (Erişim Tarihi: 12.03.2014)
- Xie L., Beni G., (1991). A Validity Measure for Fuzzy Clustering, *IEEE Trans. On Pattern Analysis and Machine Int.* 13(4),pp 841-846.