

Avrupa Ülkelerinin İntihar Oranlarına Göre Sınıflandırılması

Selay GİRAY*
Ferda ESİN GÜLEL**

ÖZET

Bu çalışmada intihar oranı ve intihar oranını etkileyen faktörlere açısından 27 Avrupa ülkesinin sınıflandırılması ve benzer yapıda olanlarının saptanması amaçlanmıştır. Analize dahil edilecek değişkenler konu hakkında yapılmış temel eserler esas alınarak belirlenmiştir. Çalışmanın uygulama kısmında söz konusu değişkenlerin en son küresel krizin görüldüğü 2008 yılı verileri kullanılmıştır. Çalışmada yöntem olarak alternatif çok değişkenli analiz tekniklerinden Bulanık Kümeleme Analizi uygun görülmüştür. Metodoloji bölümünde Bulanık Kümeleme Analizinin teorisi kısaca açıklandıktan sonra NCSS 2007 paket programı kullanılarak analiz yapılmıştır. Analiz sonucunda ülkeler 2 sınıfa ayrılmıştır. Bulgular bu sınıflamada Türkiye'nin yeri irdelenerek yorumlanmıştır.

Anahtar Kelimeler: *Avrupa Ülkeleri, İntihar Oranı, Bulanık Kümeleme Analizi*

The Classification of European Countries According to Their Suicide Rates

ABSTRACT

The aim of this study is to classify 27 European countries according to their suicide rates and examine the factors which affect these rates and similarities among countries. The variables used in the application are based on basic studies about this theme. In the application, the data from the last global crisis is used. Fuzzy Cluster Analysis is chosen from among the other alternative multivariate analysis techniques. NCSS 2007 commercial program is used to analyse the data. According to the findings, the countries are classified into 2 clusters. This study briefly interprets findings and takes a closer look at the location of Turkey.

Keywords: *European countries, Suicide Rate, Fuzzy Cluster Analysis*

1. Giriş

Sosyal bilimlerde istatistik yöntemlerin kullanıldığı ilk çalışma olarak nitelendirilen çalışma Emile Durkheim tarafından gerçekleştirilmiş olan, intiharın nedenlerini araştıran bir çalışmadır. Ekonomik krizlerin intihara neden olduğunu öne süren Durkheim, bunun nedeninin toplumsal yapıdaki değişiklik olduğunu belirtir. Toplumda meydana gelen değişiklik bireyin yaşam koşullarını etkiler. İntiharın nedenlerinden birinin bu kargaşa hali olduğu belirtilmektedir¹.

Bu çalışmada en son küresel krizin yaşanmış olduğu 2008 yılı için Avrupa ülkeleri intihar oranları açısından incelenmektedir. İntihar oranları üzerinde etkili olduğu belirlenen değişkenler de analize dahil edilmiş, böylece konu bütüncül bir yaklaşımla ele alınmaya çalışılmıştır.

Avrupa ülkeleri ilgilenilen konuya göre sınıflandırılırken çeşitli çok değişkenli istatistik tekniklerinden yararlanılabilir. Ancak bu çalışmada Bulanık Kümeleme Analizi'nden yararlanılması uygun görülmüştür. Bu analizin, sosyal bilimler konulu

* Yrd.Doç.Dr., *Marmara Üniversitesi İktisat Fakültesi, selaygiray@marmara.edu.tr*

** Yrd.Doç.Dr., *Pamukkale Üniversitesi İktisadi ve İdari Bilimler Fakültesi, fegulel@pau.edu.tr*

¹ http://www.felsefe.gen.tr/emile_durkheim_ve_intihar.asp, Erişim Tarihi:Mayıs 2013.

çalışmalarda daha aydınlatıcı olacağı düşünülmüştür. Bunun sebebi Bulanık Kümeleme Analizi sayesinde analiz kapsamındaki birimlerin kümelerine ait olma olasılıklarının da elde edilebilmesidir. Böylece ülkelerin benzerliklerine göre kümeleneceği sonucunda elde edilen bulgular daha zengin bir şekilde yorumlanabilecektir.

2. Literatür Taraması

İntihar oranı konulu literatür taraması sonucunda ulaşılan çalışmalar incelendiğinde belirli ülkelere odaklanıldığı (Avustralya, Güney Kore, Sovyetler Birliği, vb.) ve belli bir zaman periyodunda makro bazda değişkenler arasındaki ilişkilerin (intihar oranı-işsizlik oranı, intihar oranı-silah sahipliği, intihar oranı-sosyoekonomik faktörler gibi) araştırıldığı görülmüştür. İncelenen eserler arasında ülkelerin intihar oranlarına göre oluşturdukları kümeleri analiz eden bir çalışmaya rastlanmamıştır.

İntihar oranının yanı sıra çalışmanın metodoloji kısmını oluşturan Bulanık Kümeleme Analizi konulu ulusal ve uluslararası literatür taraması da yapılmıştır. Bu tarama sonucunda karşılaşılan eserler aşağıda özetlenmiştir:

Alpaslan vd.², “Bulanık Kümelemede En Uygun Küme Sayısının Yapay Sinir Ağları ve Diskriminant Analizi ile Belirlenmesi” adlı çalışmasında bulanık kümelemede en iyi küme sayısının belirlenmesi için ileri beslemeli yapay sinir ağları ve diskriminant analizi yöntemlerini benzetim ve gerçek veri üzerinde uygulayarak elde edilen sonuçları PC, CE gibi küme geçerlilik indekslerinden elde edilen sonuçlar ile karşılaştırmışlardır. Bu karşılaştırma sonucunda Yapay sinir ağlarına dayalı yaklaşım ve diskriminant analizi sınıflama hatası ile bulanık kümelemede en uygun küme sayısının belirlenebileceği sonucuna ulaşmışlardır.

Kılıç vd.³, “Bulanık Kümeleme Analizi İle Ülkelerin Turizm İstatistikleri Bakımından Sınıflandırılması” adlı çalışmalarında standartlaştırılmış ve ham veri kümelerini kullanarak 30 ülkeyi 9 değişkene göre bulanık kümeleme yöntemlerini uygulamışlardır. Analiz sonucunda ortalama gölge istatistiği, Dunn katsayıları ve ayırma analizi ile 30 ülkenin 3 kümeye ayrıldığını görmüşlerdir.

Atalay ve Tortum⁴, “Türkiye’deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi” adlı çalışmalarında 1997-2006 yılları arasında Türkiye’deki illerde meydana gelen şehir dışı trafik kazası verilerini kullanarak her il için ölüm ve yaralanma oranlarını hesaplamışlardır. Bu oranlar üzerinde k-ortalamlar ve bulanık c-ortalamlar yöntemlerini uygulayarak kümeleme analizi sonucunda illeri 5 kümeye ayırmışlar, en yüksek ölüm ve yaralanma oranlarına sahip illeri belirlemişlerdir. Daha sonra kullandıkları bu iki yöntem arasında karşılaştırma yapmışlar, bu karşılaştırma sonucunda bulanık c-ortalama yönteminin k-ortalamlar yöntemi kadar doğru ve tutarlı sonuçlar verdiğini gözlemişlerdir.

² Faruk Alpaslan vd., “Bulanık Kümelemede En Uygun Küme Sayısının Yapay Sinir Ağları ve Diskriminant Analizi İle Belirlenmesi”, *Atatürk Üniversitesi İİBF Dergisi*, 10. Ekonometri ve İstatistik Sempozyumu Özel Sayısı, 2011, s. 475-488.

³ İbrahim Kılıç ve Ceylan Özbeyaz, “Bulanık Kümeleme Analizinin Koyun Yetiştiriciliğinde Kullanımı ve Bir Uygulama”, *Kocatepe Veteriner Dergisi*, 3, 2, 2010, s.33.

⁴ Ahmet Atalay ve Ahmet Tortum, “Türkiye’deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi”, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, Cilt 16, Sayı 3, 2010, s. 335-343.

Işık ve Çamurcu⁵, “K-Means ve Aşırı Küresel C-Means Algoritmaları ile Belge Madenciliği” adlı çalışmalarında internette arama yapıldığında bulunan belgeler içinde gerekli bilgileri elde etmede kullanılan veri madenciliği yöntemi olan belge madenciliği üzerinde bulanık kümeleme yöntemlerini uygulamışlardır. Bu amaçla internet belgesi içeren üç ayrı veri setine k-ortalama ve aşırı küresel c-ortalama yöntemlerini uygulayarak kümeleme başarılarını karşılaştırmışlardır. Karşılaştırma sonucunda aşırı küresel c-ortalama yöntemine göre daha iyi sonuçlar verdiğini elde etmişlerdir.

Yılancı⁶, “Bulanık Kümeleme Analizi ile Türkiye’deki İllerin Sosyoekonomik Açından Sınıflandırılması” adlı çalışmada Türkiye’deki 81 ili 11 sosyoekonomik değişken ile bulanık kümeleme yöntemini kullanarak analiz etmiştir. Bu analiz sonucunda elde ettiği sonuçları k-ortalama yöntemine karşılaştırmıştır. Karşılaştırma sonucunda bulanık kümeleme analizinin homojen birimlerin kümelemesinde gücünün azaldığı k-ortalama yönteminin ise önsel olarak belirlenen küme sayısına karşı duyarlı olduğu sonucuna varmıştır.

Işık ve Çamurcu⁷, “K-means, K-medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti” adlı çalışmalarında literatürde bulunan sentetik veri kümeleri üzerinde sözkonusu bulanık kümeleme algoritmalarını uygulamışlar ve bu algoritmaları karşılaştırmışlardır. Bu üç algoritmadan en iyisinin k-medoids algoritması olduğunu ancak bu algoritmanın çok boyutlu verilerde ve çok büyük veri kümelerinde zaman ve bellek yetersizliği gibi ciddi problemlere neden olduğunu bulmuşlardır. Ayrıca rastgele atanan başlangıç küme merkezlerine göre k-means algoritmasındaki sonuçların k-medoids ve bulanık c-means algoritmalarına göre daha çok değişkenlik gösterdiği ve algoritmaların küresel kümeleri bulma eğilimli olduklarını da elde etmişlerdir.

Sönmez ve Er⁸, “Türkiye’de İllere Göre İç Göç Hareketlerinin Modern Kümeleme Teknikleri ile İncelenmesi” adlı çalışmalarında Türkiye’deki 81 il için 2004 yılında toplanan göç verisi üzerinde iç göç hareketleri bakımından birbirine benzeyen illeri belirlemek için bulanık kümeleme ve medoidlere göre kümeleme analizi yapmışlardır. Analiz sonucunda büyük şehirlerden dışarıya olan göçlerin çoğunun beklenilen aksine köylere değil, üç büyük il (İstanbul, İzmir, Ankara) dışında kalan illere yapıldığı sonucuna ulaşmışlardır.

Konu ile ilgili yapılan uluslararası literatür taraması sonucunda göze çarpan başlıca çalışmalar aşağıda özetlenmiştir:

⁵ Meltem Işık ve Ali Yılmaz Çamurcu, “k-Means ve Aşırı Küresel c-Means Algoritmaları ile Belge Madenciliği”, *Marmara Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Cilt 22, 2010, s. 1-18.

⁶ Veli Yılancı, “Bulanık Kümeleme Analizi ile Türkiye’deki İllerin Sosyoekonomik Açısından Sınıflandırılması”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15, 3, 2010, s.453-470. (Yılancı, 2010, s. 456)

⁷ Meltem Işık ve Ali Yılmaz Çamurcu, “K-means, K-medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Vol. 6, No. 11, 2007, s. 31-45.

⁸ Harun Sönmez ve Fikret Er, “Türkiye’de İllere Göre İç Göç Hareketlerinin Modern Kümeleme Teknikleri ile İncelenmesi”, *Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, Cilt XX, Sayı 1, 2007, s. 17-32.

Nassar ve Hosny⁹, “Fuzzy Clustering Validity For Contractor Performance Evaluation: Application to UAE Contractors” adlı çalışmalarında 294 projenin veritabanını kullanarak UAE’de çalışan 14 müteahhiti sınıflandırmak için en yaygın kullanılan 5 kümeleme algoritmasını kullanmış ve bu algoritmalar arasında karşılaştırma yapmıştır. Analiz sonucunda nicel ve nitel ölçümlere dayalı olarak 4 küme belirlemişlerdir. Ayrıca çalışılan veri kümesinde en yüksek geçerliliği bulanık c-ortalama algoritmasının verdiğini bulmuşlar, farklı uygulama gruplarındaki müteahhitleri sınıflandırmak için makul ve yansız bir şekilde bu kümeleme algoritmalarının kullanılabilirliğini göstermişlerdir.

Kumpulainen vd.¹⁰, “Analysing 3G Radio Network Performance with Fuzzy Methods” adlı çalışmalarında günümüzde daha karmaşık ve detaylı işleyim verisi sağlayabilen 3G ağlarını incelemişlerdir. Bu amaçla benzer davranış gruplarında kümelenen ağ hücrelerini elde etmek için kanal kalite ölçüm dağılımlarına bulanık hesaplama yöntemlerini uygulamışlardır. Dağılımsal veriden çıkarılan özelliklere dayalı olarak bulanık çıkarsama sistemini tanımlar ve ağ performansı görüntüleme ile bozukluk saptamada bulanık kümelemenin nasıl kullanılabilirliğini göstermişlerdir. Son olarak zaman boyutunun ağ hücre davranışını analiz etmede dikkat çekici bir bakış açısı olduğunu tanıtmışlardır. Çalışma sonunda elde edilen tüm sonuçlar radyo ağ performans tarafından bilgilendirici ve faydalı bulunmuştur.

Maharaj ve D’urso¹¹, “Fuzzy Clustering of Time Series in The Frequency Domain” adlı çalışmalarında geleneksel ve bulanık kümeleme analizlerinin ilişkisiz olan değişkenlere uygulanırken değişken değerleri serisel ilişkili olan zaman serileri sözkonusu olduğunda bulanık kümeleme analizinin nasıl uygulanabileceğini örneklendirmişlerdir.

Izakian ve Abraham¹², “Fuzzy C-Means and Fuzzy Swarm for Fuzzy Clustering Problem” adlı çalışmalarında bulanık c-ortalama ve bulanık parça küme (swarm) optimizasyon algoritmalarına dayalı bu iki algoritmanın iyi yönlerini kullanan hibrid bulanık kümeleme yöntemini önermişlerdir. Bu yöntemi uygulamada bulunan 6 veri kümesine (Fisher’in zambak verisi, cam verisi, Wisconsin göğüs kanseri verisi, şarap verisi, doğum kontrol yöntemi seçim verisi ve sesli harf veri kümesi) uygulamışlardır. Bunun sonucunda önerdikleri hibrid bulanık kümeleme yönteminin etkin olduğu ve umut verici sonuçlar verdiğini bulmuşlardır.

Fang vd.¹³, “A New Look at Quantifying Tobacco Exposure During Pregnancy Using Fuzzy Clustering” adlı çalışmalarında doğum öncesi sigara maruziyetini ölçmek için bulanık kümeleme analizi yapmışlardır. Yenidoğan düzenlilik yetenekleri (neonatal regularity skills) üzerinde doğum öncesi sigara maruziyetinin etkisini değerlendirmek

⁹ Khaled Nassar, Ossama Hosny, “Fuzzy Clustering Validity for Contractor Performance Evaluation: Application to UAE Contractors”, *Automation in Construction*, Vol. 31, 2013, s. 158-168.

¹⁰ Pekka Kumpulainen vd, “Analysing 3G Radio Network Performance with Fuzzy Methods”, *Neurocomputing*, Vol. 107, 2013, s. 49-58.

¹¹ Elizabeth Ann Maharaj, Pierpaolo D’urso, “Fuzzy Clustering of Time Series in the Frequency Domain”, *Information Sciences*, Vol. 181, No. 7, 2011, s. 1187-1211.

¹² Hesam Izakian, Ajith Abraham, “Fuzzy c-Means and Fuzzy Swarm for Fuzzy Clustering Problem”, *Expert Systems with Applications*, Vol. 38, 2011, s. 1835-1838.

¹³ H. Fang vd., “A New Look at Quantifying Tobacco Exposure During Pregnancy Using Fuzzy Clustering”, *Neurotoxicol Teratol*, Vol. 33, No. 1, 2011, s.155-165.

için tasarlanan bir projenin verisi üzerinde analiz yöntemlerini uygulamışlardır. Bu analiz sonucunda maruz kalmamış, hafif maruziyet ve ağır maruziyet olmak üzere 3 küme elde ederek doğum öncesi sigara maruziyetinin yapısını tanımlamışlardır.

Yang vd.¹⁴, “Fuzzy Clustering Algorithms for Mixed Feature Variables” adlı çalışmalarında sembolik ve bulanık karma özellikli veriye göre bulanık kümeleme algoritmalarını göstermişlerdir. Bu veri için düzenlenmiş farklılık ölçüsünü ve bu karma veri türlerine göre bulanık c-ortalama kümeleme algoritmasını göstermişlerdir. Simülasyon verisine uygulama sonrası düzenlenmiş farklılık ölçümünün daha iyi sonuçlar verdiği sonucuna ulaşmışlardır.

Kim vd.¹⁵, “Fuzzy Clustering of Categorical Data Using Fuzzy Centroids” adlı çalışmalarında kategorik veri kümelemesi için klasik bulanık k-modlar algoritmasını orijinal algoritmada kullanılan sert tipte merkezilerin yerine bulanık merkezli kategorik veri kümelemeyi göstererek genişletmiştir. Bu yöntemi, uygulama alanında bulunan soya, kredi ve hayvanat bahçesi verisi üzerine uygulamış ve k-modlar ile bulanık k-modlar yöntemleriyle karşılaştırmıştır. Karşılaştırma sonucunda önerilen yöntemin önemli derecede iyi sonuçlar verdiğini elde etmiştir.

Görüldüğü gibi ulusal ve uluslararası literatür taraması sonucunda, Avrupa ülkelerinin intihar istatistikleri açısından benzerliklerine göre sınıflandırılmalarını inceleyen bir çalışmaya rastlanmamıştır.

3. Çalışmanın Amacı ve Kapsamı

Çalışmanın amacı 2008 yılı için intihar oranları ve bunun belirleyicisi olduğu saptanan değişkenler açısından Avrupa ülkelerini benzerliklerine göre sınıflandırmaktır.

Çalışmada 2008 yılının seçilmiş olmasının sebebi, küresel krizlerle intihar oranları arasında anlamlı ilişkinin var olduğunu belirten çeşitli çalışmalara rastlanmıştır.

Çalışma kapsamında Avrupa’daki 27 ülkenin 2008 yılı verileri OECD, United Nations, EuroStat ve Dünya Sağlık Örgütü veri tabanlarından elde edilmiştir. Analize dahil edilen ülkelerin listesi aşağıda verilmiştir:

Tablo 1. Analize Dahil Edilen Ülkelerin Listesi

Almanya	İtalya
Avusturya	İzlanda
Bulgaristan	Letonya
Çek Cumhuriyeti	Litvanya
Estonya	Lüksemburg
Finlandiya	Macaristan
Fransa	Norveç
Hırvatistan	Polonya
Hollanda	Romanya

¹⁴ Miin- Shen Yang, Pei-Yuan Hwang, De-Hua Chen, “Fuzzy Clustering Algorithms for Mixed Feature Variables”, *Fuzzy Sets and Systems*, Vol. 141, No. 2, 2004, s. 301-317.

¹⁵ Kim Dae-Won, H. Lee Kwang, Doheon Lee, “Fuzzy Clustering of Categorical Data Using Fuzzy Centroids”, *Pattern Recognition Letters*, Vol. 25, No. 11, 2004, s. 1263-1271.

İngiltere	Slovakya
İrlanda	Slovenya
İspanya	Türkiye
İsveç	Yunanistan
İsviçre	

Uygulamada intihar oranı verisi ve bu veriyi açıklayan değişkenler kullanılarak ülkeler sınıflandırılmaya çalışılmıştır. İntihar oranı üzerinde etkisi olduğu düşünülen değişkenler Fransız sosyolog Durkheim¹⁶'n çalışması esas alınarak seçilmiştir. Bu değişkenlerin listesi aşağıda verilmiştir:

Tablo 2 Analiz Kapsamındaki Değişkenlerin İsimleri

Değişken	Tanım
İO	İntihar oranı (100.000 kişi başına)
İ	Uzun dönem işsizlik oranı (%)
AKO	Kişi başına alkol kullanım oranı (15+)
DO	Doğum oranı (1.000 kişi başına)
BO	Boşanma oranı (1.000 kişi başına)
EO	Evlenme oranı (1.000 kişi başına)
İKO	İşgücü katılım oranı (15+, kadın)
LogGSYİH	Kişi başına Gayri Safi Yurtiçi Hasıla (\$)
E	Enflasyon (tüketici fiyatı) (%), yıllık

Ülkeler benzerliklerine göre sınıflandırılırken oluşacak yapı (yani gruplar) önceden bilinmediğinden Diskriminant Analizi ya da Lojistik Regresyon Analizi gibi tekniklerin kullanılması uygun görülmemiş, Kümeleme Analizi'nden yararlanılmıştır. Kümeleme Analizi bilindiği gibi verileri benzerliklerine göre kümelere ayıran çok değişkenli bir tekniktir. Uygulama öncesinde oluşacak gruplar hakkında bir ön bilgi olmadığında rahatlıkla kullanılabilir. Ayrıca Kümeleme Analizi'ne dahil edilen değişkenlerin ilişki olmaması varsayımı aranmamaktadır. Bu açıdan da çalışmanın amacı doğrultusunda kullanılabilir en uygun teknik olduğu düşünülmüştür. Klasik Kümeleme Analizi sonucunda birimlerin ait oldukları kümeler belirlenir. Ancak bu birimlerin ilgili kümeye aidiyetini gösteren herhangi bir ağırlık katsayısı ya da olasılık değeri elde edilmemektedir. Çalışmada istenen, ülkeleri intihar konusunda benzerliklerine göre kümelemektir. Bu gerçekleştirilirken atamasında zorluk çekilen ülkelerin yani bir anlamda aykırı gözlemlerin belirlenmesi, yorumlamalar açısından daha sağlıklı olacaktır. Bu nedenle çalışmada modern kümeleme tekniklerinden Bulanık Kümeleme Analizi'nin uygulanması daha doğru bulunmuştur.

4. Metodoloji - Bulanık Kümeleme Analizi

Çok değişkenli analiz tekniklerinden Kümeleme Analizi'nde amaç gözlemlerin (birimlerin) benzerliklerine göre homojen kümelere ayrılmasıdır. Bu amaç doğrultusunda çeşitli uzaklık ölçülerinden yararlanılmakta ve kümeler oluşturulmaktadır. Analiz kapsamındaki her bir birimin yer aldığı küme bellidir. Ancak

¹⁶ E. Durkheim, "Suicide", New York: Free Press, 1951(Original work published 1897).

bilindiği gibi sosyal bilimler konulu çalışmalarda kesin yorumlar yapmak olanaksızdır. Herhangi bir birimin atandığı küme önemli olmakla birlikte bu kümeye hangi olasılıkla atandığını da bilmek daha yararlı olacak ve daha sağlıklı yorumlara ulaşmayı sağlayacaktır. Bu bağlamda Kümeleme Analizi'ne kıyasla daha uygun olan tekniğin Bulanık Kümeleme Analizi olduğu söylenebilir.

Klasik kümeleme yöntemleri her bir birim için kesin kararlar alarak bir kümeye atarlarken Bulanık Kümeleme Analizi'nde her bir birimin oluşan kümelere ait olma derecesini gösteren ağırlık katsayıları hesaplanmaktadır¹⁷.

Bu yöntemde bir birim için ağırlık katsayılarının toplamı her zaman 1 olmak üzere, bir birimin bir kümede olma olasılığı tüm olası kümeler arasında 0 ile 1 arasında değişmektedir.¹⁸ Sonuç olarak birim, ait olma olasılığının en yüksek olduğu kümeye atanmaktadır¹⁹. Birimlerin ağırlık katsayılarının toplam değeri için üst limit, analizdeki toplam birim sayısına eşittir.

Bulanık Kümeleme Analizi'nin en yaygın kullanılan algoritması Kaufman ve Rousseeuw tarafından geliştirilen Fanny algoritmasıdır. Uygulanabilmesi için bütün değişkenlerin ölçek düzeyleri en azından eşit aralıklı ölçek olmalıdır. Algoritmanın amaç fonksiyonu aşağıda yer almaktadır²⁰.

$$C = \sum_{v=1}^K \frac{\sum_{i,j=1}^n u_{iv}^2 u_{jv}^2 d(i,j)}{2 \sum_{j=1}^n u_{jv}^2}$$

Amaç fonksiyonunun enküçüklenmesi hedeflenmekte olup fonksiyonda geçen simgelerin ne anlama geldikleri aşağıda yer verilmiştir²¹.

$d(i,j)$ = i. ve j. birim arasındaki uzaklık

K = toplam küme sayısı

u_{iv} = i. birimin v kümesine olan bilinmeyen üyeliği ($u_{iv} \geq 0$ ve $\sum_{v=1}^K u_{iv} = 1$)

u_{jv} = j. birimin v kümesine olan bilinmeyen üyeliği

n = Toplam birim sayısı

Küme üyelikleri negatif olamaz ve bir birim için kümelere ait olma katsayıları toplamı 1'e eşit çıkmalıdır. Bu kısıtlar aşağıdaki şekilde ifade edilir:

$u_{iv} \geq 0$ ve $\sum_{v=1}^K u_{iv} = 1$; $i=1, \dots, n$ ve $v=1, \dots, K$.

Amaç fonksiyonu bu kısıtlar altında yinelemeli bir algoritma ile enküçüklenerek katsayılar matrisine ($U_{n \times K}$) ulaşılır.²²

¹⁷ An Introduction to Cluster Analysis and Data Mining,

http://www-users.cs.umn.edu/~han/dmclass/cluster_survey_10_02_00.pdf, Erişim Tarihi: Mart 2013.

¹⁸ Veli Yılcı, "Bulanık Kümeleme Analizi ile Türkiye'deki İllerin Sosyoekonomik Açından Sınıflandırılması", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15, 3, 2010, s.453-470. (Yılcı, 2010, s. 456)

¹⁹ Vicenç Torra, "Fuzzy c Means for Fuzzy Hierarchical Clustering", *FUZZ-IEEE*, Reno, Nevada, 2005, s.1-6.

²⁰ Leonard Kaufman ve Peter.J. Rousseeuw, *Finding Groups Data: An Introduction to Cluster Analysis*, John Wiley and Sons Inc, New York, 1990, s.50.

²¹ Yılcı, *a.g.e*, s. 456.

²² Bülent Tütmez ve A. Erhan Tercan, "Bulanık Modelleme Yaklaşımının Tenör Kestiriminde Kullanılması", *Madencilik*, 45, 2, 2006, s.39-47.

Her birimin tüm kümelerde eşit üyelik katsayısına sahip olması durumunda kümeleme tamamen bulanık olarak isimlendirilir. Kümelemenin ne derece bulanık olduğu hakkında fikir sahibi olabilmek için Dunn'ın Parçalama Katsayısı hesaplanmalıdır. Aşağıda verilen bu katsayı üyelik değerlerinin kareler toplamına dayanmakta ve tüm birimlerin tüm kümeler üzerindeki üyelik değerlerini dikkate almaktadır.

$$F(u) = \frac{\sum_{i=1}^n \sum_{v=1}^K u_{iv}^2}{n}$$

Dunn'ın Parçalama Katsayısı tamamen bulanık kümeleme durumunda (u_{iv} değerlerinin $1/k$ 'ya eşit olduğu durumda) $1/k$ değerini alır. Kümelemenin çok net (katı) olduğu durumda ise (her bir gözlem / birime ait u_{iv} değerlerinin $(v-1)$ tanesinin 0, 1 tanesinin 1 olduğu durum: kesin küme durumları) 1'e eşit olur.

Dunn'ın Parçalama Katsayısı minimum değeri 0, maksimum değeri 1 olacak şekilde normalleştirilebilir. Böylece katsayı küme sayısından bağımsız hale gelecek şekilde standartlaştırılmış olur. Standartlaştırma sonrası Bulanıksızlık İndeksi adını alan Dunn Katsayısının değerinin 0 olması tamamen bulanıklığı, 1 olması ise güçlü kümelenmeyi göstermektedir.²³ Söz edilen Bulanıksızlık İndeksi aşağıda verilmiştir:

$$F_c(U) = \frac{k \cdot F(u) - 1}{k - 1}$$

Bu katsayıya alternatif olarak Kaufman Ayrıştırma Katsayısı'ndan da yararlanılabilmektedir. Kaufman Katsayısı ve Normalleştirilmiş Kaufman Katsayısı formülleri aşağıda sırasıyla yer almaktadır:

$$D(U) = \frac{1}{n} \sum_{v=1}^K \sum_{i=1}^n (h_{iv} - u_{iv})^2$$

$$D_c(U) = \frac{D(U)}{1 - \left(\frac{1}{K}\right)}$$

Hesaplanan bu katsayının 0'a yakın olması güçlü kümelenmeyi göstermektedir.

Küme sayısı belirlemede Bulanıksızlık İndeksi $F_c(U)$ değerinin mümkün olduğunca büyük olması $D_c(U)$ değerinin ise mümkün olduğunca küçük olması istenmektedir.²⁴ Bu katsayılar literatürde "Küme Geçerlilik İndeksleri" olarak da yer almaktadır.²⁵

Yukarıda bahsi geçen katsayıların yanısıra birimlerin ne kadar iyi kümelendiğinin belirlenmesi veya kümelerin kararlılık yapısı için Gölge İstatistiği (*SC: Silhouette Coefficient*) kullanılmaktadır. Gölge istatistiği, i. birimin kendi kümesi içindeki diğer birimlerle farklılığını kendisine en yakın komşu kümedeki birimlerin farklılığı ile karşılaştırmaktadır. Hesaplanan istatistik değeri -1 ile +1 arasında olup +1'e yakın olması durumunda i.birimin doğru sınıflandırıldığı yorumu yapılır. Bu istatistik

²³ Yılcı, a.g.e, s.457.

²⁴ <http://ncss.npengine.netdna-cdn.com/wp-content/uploads/2012/09/NCSSUG4.pdf>, s.448, Erişim Tarihi: Mart 2013

²⁵ Faruk Alpaslan vd., "Bulanık Kümelemede En Uygun Küme Sayısının Yapay Sinir Ağları ve Diskriminant Analizi İle Belirlenmesi", *Atatürk Üniversitesi İİBF Dergisi*, 10. Ekonometri ve İstatistik Sempozyumu Özel Sayısı, 2011, s. 475-488.

değerlerinin ortalaması Ortalama Gölge İstatistiği adını almaktadır. Bir veri setindeki kümeleme yapısının ‘uygun’ olarak nitelendirilebilmesi için bu ortalama değer 0,50 olması beklenmektedir²⁶. Ortalama Gölge İstatistiği değerinin en yüksek elde edildiği küme sayısı analiz için en uygun küme sayısı olacaktır²⁷.

Bulanık Kümeleme Analizi’nde küme sayısının belirlenmesinde en çok tercih edilen yöntem, $k=2,3,4,\dots$ için analiz yapıldıktan sonra Normalleştirilmiş Dunn Katsayılarını, Normalleştirilmiş Kaufman Katsayılarını ve Ortalama Gölge İstatistiği değerlerini kontrol etmektir. Bunların sonrasında belirlenmiş küme üyelikleri ile Diskriminant Analizi de uygulanarak en yüksek doğru sınıflama oranına sahip küme yapısını (dolayısıyla sayısını) belirlemektir²⁸. Ancak Diskriminant Analizi uygulanmadan önce analizin gerektirdiği Normallik ve homojenlik varsayımlarının geçerliliği kontrol edilmelidir. Bu aşamada Diskriminant Analizi’nin yanısıra Yapay Sinir Ağları yönteminden de yararlanılabilir. Son olarak analiz kapsamındaki değişkenlerin hangilerinin kümelerin oluşumunda etkili olduğunu belirlemek amacıyla MANOVA uygulanabilir²⁹. Burada da öncelikle analiz sonuçlarının güvenilirliği için sağlanması gereken varsayımların geçerliliği kontrol edilmelidir. Varsayımların sağlanmadığı tespit edilirse parametrik olmayan testler uygulanarak Kümeleme Analizi’nde ANOVA tablosunun parametrik olmayan alternatifi oluşturulabilir.

5. Bulgular

Çalışmada NCSS 2007 paket programından yararlanılmıştır.

İlk aşamada $k = 2,3,4,5$ için analiz sonuçları ve sırasıyla Dunn Katsayısı, Bulanıksızlık İndeksi(standartlaştırılmış Dunn Katsayısı), Kaufman Katsayısı, standartlaştırılmış Kaufman Katsayısı aşağıdaki gibi elde edilmiştir:

Tablo 3 Uygun Küme Sayısının Belirlenmesi

Küme Sayısı	Ortalama Uzaklık	Ortalama Silouette	$F(U)$	$F_c(U)$	$F_e(U)$	$D_c(U)$
2	38,639599	0,467350	,6232	,2465	,1550	,3100
3	25,253013	0,282865	,4557	,1836	,3035	,4553
4	18,567488	0,278631	,3974	,1965	,3446	,4594
5	14,569476	0,282349	,3546	,1933	,3893	,4866

Tablo 3’te verildiği gibi en yüksek $F_c(U)$ değerini ve en düşük $D_c(U)$ değerini veren küme sayısının 2 olduğu görülmektedir.

Analiz sonucunda oluşan iki kümeye ilişkin betimleyici istatistik değerleri aşağıdaki tabloda yer almaktadır:

²⁶ İbrahim Kılıç ve Ceylan Özbeyaz, “Bulanık Kümeleme Analizinin Koyun Yetiştiriciliğinde Kullanımı ve Bir Uygulama”, *Kocatepe Veteriner Dergisi*, 3, 2, 2010, s.33.

²⁷ Mehmet Şahin, Bahattin Hamarat, “G10-Avrupa Birliği ve OECD Ülkelerinin Sosyoekonomik Benzerliklerinin Fuzzy Kümeleme Analizi ile Belirlenmesi”, *ODTÜ Uluslararası Ekonomi Kongresi VI*, Ankara, 2002, s.8.

²⁸ İbrahim Kılıç, Oktay Emir, Gonca Kılıç, “Bulanık Kümeleme Analizi İle Ülkelerin Turizm İstatistikleri Bakımından Sınıflandırılması”, *İstatistikçiler Dergisi* (4), 2011, s.34.

²⁹ Özkan Görgülü, “Bulanık Mantık (Fuzzy Logic) Teorisi ve Tarımda Kullanım Olanakları Üzerine Bir Araştırma”, *Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü*, Yayınlanmamış Doktora Tezi, Hatay, 2007

Tablo 4 Kümelere İlişkin Medoid Değerleri

Değişken	Küme 1	Küme 2
İntihar Oranı	12,6	14,1
Uzun Dönem İşsizlik Oranı	50,2	24,2
Alkol Kullanım Oranı	12,1	12,4
Doğum Oranı	11,5	9,3
Boşanma Oranı	3,003	2,363
Evlenme Oranı	5,032	4,225
İşgücüne katılım (Kadın)	49,5	53,4
GSYİH	20728,85	49679,42
Enflasyon	6,351	3,216

Tablo 4'te görüldüğü gibi intihar oranı, kadınların işgücüne katılım oranı ve GSYİH Küme 1'de daha düşük iken uzun dönem işsizlik oranı, doğum oranı, boşanma oranı, evlenme oranı ve enflasyon daha yüksektir. Alkol kullanım oranları açısından ise iki kümenin benzer olduğu söylenebilir.

Sonuçların yorumlanabilirliği için Silhouette Amount değerleri incelenmiş, hiç birinin negatif olmadığı ve çoğunun 0,5'ten büyük olduğu görüldüğünden sonuçların yorumlanmasında herhangi bir sakınca olmadığına karar verilmiştir.

Kümelere yer alan ülkeler tablosu ve ülkelerin ilgili kümeye üyelik değerlerini gösteren katsayılar matrisi aşağıda yer almaktadır.

Tablo 5 Ülkelerin Küme Üyelik Katsayıları Matrisi

Küme 1	Küme 2
Bulgaristan	Avusturya Estonya
Hırvatistan	Finlandiya İzlanda
Çek Cumhuriyeti	İrlanda Letonya
Fransa	Litvanya Lüksemburg
Almanya	Hollanda Norveç
Yunanistan	Polonya İspanya
Macaristan	İsviçre İsveç
İtalya	İngiltere Türkiye
Romanya	
Slovakya	
Slovenya	

Tablo 5 incelendiğinde Türkiye'nin İngiltere, Avusturya, Lüksemburg, Hollanda ve İsveç gibi gelişmiş ülkelerle aynı kümede yer aldığı görülmektedir. Bu kümede intihar oranı diğer kümeye göre nispeten yüksek, uzun dönem işsizlik oranı ve enflasyon ise neredeyse yarı oranda düşük olarak gözlenmiştir. Yine bu kümede diğer kümeye nazaran doğum oranı, boşanma oranı ve evlenme oranının düşük olması kümede yer alan ülke profilleri gözönüne alındığında dikkat çekicidir. Kadınların işgücüne katılım oranı ve GSYİH ise bu kümede daha yüksek olarak elde edilmiştir. Türkiye'nin intihar oranı ve intihar oranını etkileyen faktörler açısından gelişmiş ülkelerle aynı yapıda bulunduğu söylenebilir.

Ülkelerin bu kümelere atanma olasılıkları ise aşağıdaki tabloda belirtilmiştir:

Tablo 6 Küme Üyelikleri Olasılık Matrisi

Ülkeler	Küme	Küme 1'de yer alma olasılığı	Küme 2'de yer alma olasılığı
Avusturya	2	0,1574	0,8426
Bulgaristan	1	0,8494	0,1506
Hırvatistan	1	0,7433	0,2567
Çek Cumhuriyeti	1	0,8551	0,1449
Estonya	2	0,3029	0,6971
Finlandiya	2	0,2166	0,7834
Fransa	1	0,5587	0,4413
Almanya	1	0,8454	0,1546
Yunanistan	1	0,7604	0,2396
Macaristan	1	0,7017	0,2983
İzlanda	2	0,2954	0,7046
İrlanda	2	0,1820	0,8180
İtalya	1	0,7834	0,2166
Letonya	2	0,2519	0,7481
Litvanya	2	0,3468	0,6532
Lüksemburg	2	0,3571	0,6429
Hollanda	2	0,4428	0,5572
Norveç	2	0,2840	0,7160
Polonya	2	0,2214	0,7786
Romanya	1	0,7203	0,2797
Slovakya	1	0,7248	0,2752
Slovenya	1	0,6907	0,3093
İspanya	2	0,2005	0,7995
İsveç	2	0,2340	0,7660
İsviçre	2	0,4205	0,5795
Türkiye	2	0,2958	0,7042
İngiltere	2	0,2030	0,7970

Tablo 6'da verilen küme üyelik olasılık değerleri incelendiğinde Fransa ve Hollanda'nın nispeten iki kümenin kesişiminde kalmış ülkeler olduğu söylenebilir.

6. Sonuç

Amacı Avrupa ülkelerinin intihar oranları ve bu değişkeni etkileyen faktörler açısından kümelenmesini incelemek ve Türkiye'nin yerini irdelemek olan bu çalışmada, bulanık kümeleme analizi uygulamasıyla 27 ülke 9 değişkene göre sınıflara ayrılmıştır. Sınıf (küme) sayısı metodolojide açıklanan çeşitli ölçütler yardımıyla 2 olarak belirlenmiştir. Oluşan kümelerin betimleyici istatistik değerleri yardımıyla küme profilleri belirlenmiştir.

Elde edilen bulgular incelendiğinde Türkiye'nin Avusturya, Estonya, Finlandiya, İzlanda, İrlanda, Letonya, Litvanya, Lüksemburg, Hollanda, Norveç, Polonya, İspanya, İsveç, İsviçre ve İngiltere ülkeleriyle aynı kümede yer aldığı görülmüştür.

Bunların yanısıra bulanık kümeleme analizi yardımıyla elde edilen küme üyelik değerleri incelenmiş ve iki kümenin kesişiminde kalan iki ülke olduğu (Fransa ve Hollanda) tespit edilmiştir.

Sonuç olarak bu çalışmada en son küresel krizin yaşanmış olduğu 2008 yılı için 27 Avrupa ülkesi intihar oranları ve intihar oranları üzerinde etkili olan değişkenler açısından bulanık kümeleme analizi kullanılarak sınıflara ayrılmış ve Türkiye'nin yeri irdelenmiştir.

Kaynaklar

- ALPASLAN Faruk vd., “Bulanık Kümelemede En Uygun Küme Sayısının Yapay Sinir Ağları ve Diskriminant Analizi İle Belirlenmesi”, *Atatürk Üniversitesi İİBF Dergisi*, 10. Ekonometri ve İstatistik Sempozyumu Özel Sayısı, 2011, s. 475-488.
- ATALAY Ahmet, TORTUM Ahmet, “Türkiye’deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi”, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, Cilt 16, Sayı 3, 2010, s. 335-343.
- DAE-WON Kim, LEE Kwang H., LEE Doheon, “Fuzzy Clustering of Categorical Data Using Fuzzy Centroids”, *Pattern Recognition Letters*, Vol. 25, No. 11, 2004, s. 1263-1271. +
- FANG H. vd., “A New Look at Quantifying Tobacco Exposure During Pregnancy Using Fuzzy Clustering”, *Neurotoxicol Teratol*, Vol. 33, No. 1, 2011, s.155-165. +
- GÖRGÜLÜ Özkan, “Bulanık Mantık (Fuzzy Logic) Teorisi ve Tarımda Kullanım Olanakları Üzerine Bir Araştırma”, *Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü*, Yayınlanmamış Doktora Tezi, Hatay, 2007.
- İŞİK Meltem ve ÇAMURCU Ali Yılmaz, “K-means, K-medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Vol. 6, No. 11, 2007, s. 31-45.
- İŞİK Meltem, ÇAMURCU Ali Yılmaz, “k-Means ve Aşırı Küresel c-Means Algoritmaları ile Belge Madenciliği”, *Marmara Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Cilt 22, 2010, s. 1-18.
- IZAKIAN Hesam, ABRAHAM Ajith, “Fuzzy c-Means and Fuzzy Swarm for Fuzzy Clustering Problem”, *Expert Systems with Applications*, Vol. 38, 2011, s. 1835-1838.
- KAUFMAN Leonard, ROUSSEEUW Peter.J., *Finding Groups Data: An Introduction to Cluster Analysis*, John Wiley and Sons Inc., New York, 1990.
- KILIÇ İbrahim, EMİR Oktay, KILIÇ Gonca, “Bulanık Kümeleme Analizi İle Ülkelerin Turizm İstatistikleri Bakımından Sınıflandırılması”, *İstatistikçiler Dergisi*, Cilt 4, 2011, s. 31-38.
- KILIÇ İbrahim, ÖZBEYAZ Ceylan, “Bulanık Kümeleme Analizinin Koyun Yetiştiriciliğinde Kullanımı ve Bir Uygulama”, *Kocatepe Veteriner Dergisi*, Cilt 3, No 2, 2010, s. 31-37.
- KUMPULAINEN Pekka vd, “Analysing 3G Radio Network Performance with Fuzzy Methods”, *Neurocomputing*, Vol. 107, 2013, s. 49-58.
- MAHARAJ Elizabeth Ann, D’URSO Pierpaolo, “Fuzzy Clustering of Time Series in the Frequency Domain”, *Information Sciences*, Vol. 181, No. 7, 2011, s. 1187-1211.

- NASSAR Khaled, HOSNY Ossama, “Fuzzy Clustering Validity for Contractor Performance Evaluation: Application to UAE Contractors”, *Automation in Construction*, Vol. 31, 2013, s. 158-168.
- SÖNMEZ Harun, ER Fikret, “Türkiye’de İllere Göre İç Göç Hareketlerinin Modern Kümeleme Teknikleri ile İncelenmesi”, *Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, Cilt XX, Sayı 1, 2007, s. 17-32.
- ŞAHİN Mehmet, HAMARAT Bahattin, “G10-Avrupa Birliği ve OECD Ülkelerinin Sosyoekonomik Benzerliklerinin Fuzzy Kümeleme Analizi ile Belirlenmesi”, *ODTÜ Uluslararası Ekonomi Kongresi VI*, Ankara, 2002, s. 1-20. +
- TORRA Vicenç, “Fuzzy c Means for Fuzzy Hierarchical Clustering”, *The 14th IEEE International Conference on Fuzzy Systems*, Reno, Nevada, 2005, s. 1-6.
- TÜTMEZ Bülent, TERCAN A. Erhan, “Bulanık Modelleme Yaklaşımının Tenör Kestiriminde Kullanılması”, *Madençilik*, Cilt 45, Sayı 2, 2006, s. 39-47. +
- YANG Miin- Shen, HWANG Pei-Yuan, CHEN De-Hua, “Fuzzy Clustering Algorithms for Mixed Feature Variables”, *Fuzzy Sets and Systems*, Vol. 141, No. 2, 2004, s. 301-317.
- YILANCI Veli, “Bulanık Kümeleme Analizi ile Türkiye’deki İllerin Sosyoekonomik Açidan Sınıflandırılması”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Cilt 15, No 3, 2010, s. 453-470.
- URL-1 http://www.felsefe.gen.tr/emile_durkheim_ve_intihar.asp
- URL-2 <http://stats.oecd.org/index.aspx>
- URL-3 <http://www.un.org/en/>
- URL-4 http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database
- URL-5 <http://www.who.int/research/en/>
- URL-6 An Introduction to Cluster Analysis and Data Mining,
http://www-users.cs.umn.edu/~han/dmclass/cluster_survey_10_02_00.pdf
- URL-7 <http://ncss.wpengine.netdna-cdn.com/wp-content/uploads/2012/09/NCSSUG4.pdf>