Cilt/Vol: 22 – Sayı/No: 2

Anadolu Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi

APA	Dikmen, F. C. (2021). TÜRKİYE'DEKİ İLLERİN İYİ OLUŞ VE YAŞAM KALİTESİNİN R KÜMELEME ÇÖZÜMLEMESİYLE İNCELENMESİ. Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 22 (2), 59-72.
DOI	10.53443/anadoluibfd.942544

Araştırma Makalesi
Başvuru Tarihi: 25.05.2021
Kabul Tarihi: 23.06.2021

Research Article
Date Submitted: 25.05.2021
Date Accepted: 23.06.2021

TÜRKİYE'DEKİ İLLERİN İYİ OLUŞ VE YAŞAM KALİTESİNİN R KÜMELEME ÇÖZÜMLEMESİYLE İNCELENMESİ

Doç. Dr. Feyyaz Cengiz Dikmen[1]

ÖZET

Anahtar Kelimeler:

- Kümeleme analizi,
- İyi oluş

İyi oluş ve yaşam kalitesi açısından her ilin kendine özgü özellikleri bulunmaktadır. Bu özgün özelliklere göre illeri sınıflandırmak mümkündür. Bu çalışma, Türkiye İstatistik Kurumu'nun 2015 yılı için yayınladığı yaşam endeksi gösterge değerlerine dayanarak illeri kendi içinde birbirine benzer diğer gruplardan mümkün olduğu ölçüde ayrık gruplara ayırmaya çalışmaktadır. İllerin bu benzerlik ve ayrışmasında 81 ilin konut, çalışma hayatı, gelir ve servet, sağlık, eğitim, çevre, güvenlik, sivil katılım, altyapı hizmetlerine erişim ve sosyal yaşam göstergeleri temel alınmaktadır. Bu bağlamda illerin benzer özelliklerine göre sınıflandırılmasında ve özet bilgi elde edilmesinde kümeleme analizinden yararlanılmıştır. Problemin yapısı açısından önceden küme sayısını belirlemek mümkün olmadığından farklı sayıda küme sayısı ile sınıflandırmalar gerçekleştirilmiştir. Çalışmada en uygun küme sayısını belirlemek için farklı sayıda deneme (k=3, k=4, k=5) yapılmıştır. İlleri iyi oluş ve yaşam kalitesi açısından dört ya da üç kümeye ayırmanın anlamlı olacağı sonucu ortaya çıkmaktadır.

INVESTIGATION OF WELL-BEING AND QUALITY OF LIFE OF THE TURKISH PROVINCES BY CLUSTERING ANALYSIS

Assoc. Prof. Dr. Feyyaz Cengiz Dikmen

ABSTRACT

Keywords:

- Clustering Analysis;
- Well-being

Each province has its own characteristics in terms of well-being and quality of life. It is possible to classify the provinces according to these specific features. This study tries to divide the provinces into some discrete subsets such that provinces in a particular subset sharing similar properties while provinces in a particular subset showing different properties. Data taken into consideration consists of the indicator values of well-being index for 81 provinces, published by Turkey Statistical Institute for the year 2015. In this similarity and discrimination of the 81provinces, clustering is based on housing, working life, income and wealth, health, education, environment, security, civic participation, access to infrastructure services and social life indicator values. In this context, clustering analysis was used to classify the provinces according to their similar characteristics and to obtain summary information. Since it was not possible to determine the number of clusters in advance in terms of the structure of the problem, classifications were carried out with different number of clusters. In the study, different number of trials (k=3, k=4, k=5) were conducted to determine the optimal number of clusters. As a result it will be meaningful to distinguish either three or four clusters in terms of well being and quality of life

[1]İşletme Bölümü, Ağrı İbrahim Çeçen Üniversitesi, fcdikmen@agri.edu.tr , https://orcid.org/0000-0002-4697-0761

1. GİRİŞ

Yaşam kalitesi ve iyi oluş tarihte tartışmalı en eski konulardan birdir. Özellikle de ekonomi bilimi alanında, yaşam kalitesi mutlulukla ilgili akademik çalışmaların ana konusudur. İyi oluş ve yaşam kalitesi 1960 ve 1970'li yıllarda ortaya çıkmış iki önemli kavramdır. Yaşam kalitesinin ölçülmesinde ekonomik göstergelerin tek başına yeterli olmadığı, bunun yanında sosyal göstergelerin de göz önüne alınması gerektiği tartışılmaktadır. Kuşkusuz iyi oluş ve yaşam kalitesinin incelenmesi bölgesel farklılıklara göre ne tür kararlar alınması bağlamında kamu politika yapıcılar açısından genel ve yerel ölçekte oldukça önemlidir. İyi oluş ve yaşam kalitesi açısından her ilin kendine özgü özellikleri bulunmaktadır. Bu özgün özelliklere göre illeri kümelere ayırmak mümkündür. Kamu politika yapıcılarının ülke kaynaklarını iller arasında eşit olarak dağıtmamalarından kaynaklanan iller arasındaki yaşam kalitesi ve iyi oluş farklılıkları sosyal huzursuzluk ve dengesizlikleri de beraberinde getirmektedir. Bu çalışma illerin iyi oluş ve yaşam kalitesi açısından gruplara ayrılması ve böylelikle bundan sonraki ekonomik, sosyal ve kültürel kaynak dağıtımında ve sürdürülebilir gelişmenin sağlanması açısından karar vericiler için yol gösterici olabileceği düşüncesiyle ele alınmıştır. İllerin bu benzerlik ve ayrışmasında 81 ilin konut, çalışma hayatı, gelir ve servet, sağlık, eğitim, çevre, güvenlik, sivil katılım, altyapı hizmetlerine erişim ve sosyal yaşam göstergeleri temel alınmaktadır. Ele alınan göstergeler Avrupa Komisyonunun Eylül 2011 sonuç raporunda iyi oluşa katkı sağlayan faktörlerle uyuşmaktadır (Avrupa Komisyonu Sonuç Raporu,2011). Bu bağlamda illerin benzer özelliklerine göre sınıflandırılmasında ve özet bilgi elde edilmesinde tıp, biyoloji, jeoloji, veterinerlik, spor ve ekonomi gibi disiplinlerde yaygın olarak kullanılan kümeleme analizinden yararlanılmıştır.

Bu çalışma, Türkiye İstatistik Kurumu'nun ilk kez 2015 yılı için yayınladığı yaşam endeksi gösterge değerlerine dayanarak illeri kendi içinde birbirine benzer diğer gruplardan mümkün olduğu ölçüde ayrık gruplara ayırmaya çalışmaktadır. TUİK 2015 yılında yayınladığı bu istatistiği, sonraki yıllarda yayınlamamıştır. Bilimsel yazında yaşam endeksi gösterge değerlerine göre illerin ayrıştırılmasına rastlanılmamıştır. Bu çalışmanın sınırlılıkları göz önünde bulundurulmakla birlikte, bu alanda ilk olduğu düşünülmektedir. Ancak kümeleme analizi kullanılarak Türkiye'deki illerin sağlık göstergelerine göre (Çelik,2013; Tekin,2015), illerin sosyoekonomik göstergelere göre (İlknur,1998; Koç,2001; Dinçer,vd.,2003; Karabulut,vd.,2004; Yılancı, 2010), Avrupa Birliği ülkeleri ve aday ülkelerin sosyoekonomik göstergelere göre (Şahin ve Hamarat, 2002; Sandal,vd.,2005; Turanlı,vd.,2006; Erol, 2013) ve OECD ülkelerinin eğitim göstergelerine göre (Akın ve Eren,2012) sınıflandırıldığı çalışmalar bulunmaktadır.

Bu çalışma Türkiye'deki iyi oluş ve yaşam kalitesine göre illerin sınıflandırılması alanında deneysel çalışmalara katkı yapacağı düşünülmektedir. Çalışma dört bölümden oluşmaktadır. Giriş bölümünden sonra genel olarak kümeleme kavramına yer verilmekte, yöntem olarak bu çalışmada kullanılan K-ortalama kümeleme yöntemi ve uygulama bölümünde de K-ortalama kullanılarak illerin yaşam kalitesine göre sınıflandırılması yapılmakta ve sonuçları tartışılmaktadır.

2. KÜMELEME

İnsanoğlunun en temel becerilerinden biri de çevresinde gördüğü nesneleri benzer özelliklerine göre gruplama ihtiyacını karşılamaktır. İnsanlığın başlangıcından itibaren insanoğlu çevresinde yer alan nesneleri, zehirli, yırtıcı, yenilebilir gibi çeşitli özelliklere göre sınıflandırma düşüncesini her zaman taşımıştır. Sınıflandırma düşüncesi aynı zamanda bilimsel çalışmaların da temel faaliyetlerinden biridir. Örneğin, ünlü düşünür Aristoteles, hayvanlar evreninin türlerini sınıflandırmak için, hayvanları iki ana gruba ayırarak, omurgalılar ve omurgasızları olarak ayıran ayrıntılı bir sistem kurmuştur. Bu sınıflandırmayı daha da ayrıntılandırarak bu iki grubu, üretilme

şekline göre alt bölümlere ayırmıştır. Aristoteles'in ardından Theophrastos , bitkilerin yapısı ve sınıflamasıyla ilgili ilk temel açıklamaları yazmıştır (Everitt, vd., 2011)

Günümüz dünyasında, ekonomik, sosyal ve tıp alanında çok miktarda veri kümeleri (big data) ortaya çıkmaktadır. Kümeleme çözümlemesinin amacı bu yığın veri içerisinden anlamlı olabilecek bilgileri özetleyerek, anlaşılmasını kolaylaştırmak ve yönetebilmektir. Bu tür veri yığınlarının kümeleme ve diğer çok değişkenli çözümleme yöntemleriyle çözümlenmesi bilim alanında veri madenciliği olarak da adlandırılmaktadır. Kümeleme çözümlemesi, pazarlama araştırmalarında, psikiyatri, arkeoloji, astronomi, biyobilişim, genetik ve hava durum tahmini gibi çeşitli alanlarda yaygın olarak kullanılabilen bir çözümleme aracıdır.

Veri kümeleme (clustering) yöntemleri yığın veri içinde var olan gizil yapıyı ortaya çıkarmada çok değişkenli veri kümelerine uygulanan açıklayıcı çok değişkenli çözümleme yöntemlerinden biridir. Kümeleme, veri kümesindeki farklı birimler arasındaki yerleşik özelliklerin benzerlikleri ya da benzeşmezlikleri değerlendirilerek yapılmaktadır. Birimlerin kümelenmesi ortaya çıkartılan bu benzerliklere göre yapılır. Bunun sonucu olarak her küme içinde yer alan birimler birbirine benzer, diğer küme içinde yer alan birimlere benzemez yapılar oluştururlar. Bu kümeler biçim, boyut ve yoğunluk bakımından birbirlerinden farklıdırlar. İdeal bir küme diğer diğer kümelerden ayrık, kendi içinde tekparça noktalar kümesidir (Choudhary, 2016).

Kümeleme çözümlemesi ile ilgili yazında çok sayıda farklı türlere ayrılabilecek kümeleme yöntemi vardır. Kümeleme yöntemlerini beş gruba ayırmak mümkündür: bölümlere ayırma (partitioning methods), hiyerarşik (hieararchical), yoğunluk tabanlı (density based), grid tabanlı (grid based), ve kısıt tabanlı (constraint based) yöntemler (Choudhary, 2016; Sheikholeslami,vd., 2000). Ayrıca kümeleme çözümlemesi başka bir yönüyle ayrık ve bulanık kümeleme, tam ve kısmi kümeleme, tek yönlü ve iki yönlü kümeleme ve hiyerarşik ve bölümlü kümeleme olarak da ayırt edilebilmektedir (Charrad, vd.,2014).

3. K-ORTALAMA KÜMELEME ANALİZİ

Bu çalışmada kullanılan K-ortalama (K-means) bölümlere ayırma kümeleme yöntemleri içinde en yaygın kullanılan yöntemlerden biridir. Bölümlere ayırma yaklaşımında veriler bir kaç ölçüt fonksiyonuna göre sınıflandırılmaktadır. Sınıflandırılmada ölçüt olarak birimlerin çeşitli özelliklere benzerlikleri temel alınır. Benzerliğin ölçülmesinde küme içinde yer alan her bir birimin küme ortalama değerine olan uzaklıkları gözetilir. Genellikle kullanılan uzaklık ölçüsü Euclid uzaklığıdır. K-ortalama yöntemini diğer hiyerarşik kümeleme yöntemlerinden ayıran en önemli farklılık, çözümleme öncesi araştırmacının küme sayısını gelişigüzel ya da mantıksal olarak belirlemesidir. Hangi biçimde olursa olsun küme sayısı seçimi güçsüz olsa da çözüm sonuçları etkilememekte, sadece hesaplama zamanını artırmaktadır. Buna göre, araştırmacı başlangıçta k sayıda küme belirler. Her bir küme için rassal olarak seçilen küme merkezine uzaklığına göre her birim bir kümeye atanır. Bu atamaya göre küme merkezleri yeniden hesaplanır ve birimler bu merkezlere uzaklık ölçülerine göre yeniden atanırlar (Cleff, 2019). Bu süreç hata kareleri ölçüt minimum oluncaya dek devam eder. X, nxp (n= birim sayısı, p=değişken sayısı) boyutlu bir veri kümesi olmak üzere hata kareleri ölçütü (E) :

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - M_i|^2$$

Özetle k-ortalama yöntemi, n sayıda birimi **E** hata terimini minimize edecek şekilde **k** sayıda kümeye bölmektedir. K-ortalama yönteminin uygulanması, en iyi küme sayısına ilişkin bilgi; çözümleme öncesi nicel değişkenlerin standartlaştırılması ve çoklu eşdoğrusallık testinin yapılmasını gerektirmektedir. Ayrıca, K-ortalama kümeleme çözümlemesi, ortalamalara dayandığından veri kümesindeki uç değerlere karşı çok duyarlıdır. Dolayısıyla çözümleme öncesi uç değerlerin saptanması gerekmektedir (Alpar, 2017; Morissette,2013).

Veri kümesindeki birimler arasındaki uzaklıkların hesaplanmasında, değişkenlerin ölçülmesinde kullanılan ölçeklerin –sürekli, kategorik ya da hem sürekli hem de kategorik- farklı olmasına bağlı olarak çeşitli benzersizlik ölçüleri kullanılmaktadır. Geniş anlamda benzersizlik ölçüleri, uzaklık ölçüleri ve korelasyon-türü ölçüler olarak ikiye ayırılmaktadır. Sürekli nicel veriler için yaygın olarak kullanılan uzaklık ölçüleri aşağıda özetlenmektedir (Everitt, vd., 2011).

$Minkowski\ Uzaklığı$	$d_\lambda(x_i, x_j) = \sum_{k=1}^{p} \left[\left\|x_{ik} - x_{jk}\right\|^{\lambda}\right]^{\prime} \quad ; \lambda \geq 1$
$City - Block\ Uzaklığı$	$d_1(x_i, x_j) = \sum_{k=1} \left\|x_{ik} - x_{jk}\right\|; \lambda = 1$
$Euclid\ Uzaklığı$	$d_2(x_i, x_j) = \sum_{k=1} \left[\left\|x_{ik} - x_{jk}\right\|^{2}\right]^{\prime} \quad ; \lambda = 2$
$Mahalanobis\ Uzaklığı$	$d(x_i, x_j) = D^2 = (x_i - x_j)^{\prime} S^{-1} (x_i - x_j)$
$Hotteling\ T^2$	$T^2 \frac{n_1 n_2}{n} (\overline{x}_i, \overline{x}_j)^{\prime} S^{-1} (\overline{x}_i, \overline{x}_j)$
$Canberra\ Uzaklığı$	$d(x_i, x_j) = \sum_{k=1}^{p} \left\|x_{ik} - x_{jk}\right\| / \sum_{k=1}^{p} (x_{ik} + x_{jk})$
$Pearson\ Korelasyon$	$\delta_{ij} = (1 - \varphi_{ij})/2$
	$\varphi_{ij} = \sum_{k=1}^{p} w_k (x_{ik} - \bar{x}_i.)\,(x_{jk} - \bar{x}_j.) / [\sum_{k=1}^{p} w_k (x_{ik} - \bar{x}_i.)^2 \sum_{k=1}^{p} w_k (x_{jk} - \bar{x}_j.)^2]^{1/2}$
$Açısal\ Ayrılma$	$\bar{x}_i = \sum_{k=1}^{p} w_k x_{ik} / \sum_{k=1}^{p} w_k$ $\delta_{ij} = (1 - \varphi_{ij})/2$
	$\varphi_{ij} = \sum_{k=1}^{p} w_k x_{ik} x_{jk} / (\sum_{k=1}^{p} w_k x_{ik}^2 \sum_{k=1}^{p} w_k x_{jk}^2)^{1/2}$

K-ortalama kümeleme çözümlemesinde çok sayıda kümeleme algoritması kullanılmaktadır. Bu algoritmalardan en yaygın olarak kullanılanları, Forgy/Lloyd, MacQueen ve Hartigan & Wong algoritmalarıdır. Bu algoritmalardan hangisinin kullanılacağı veri kümesinin boyutuna ve değişken sayısına bağlıdır. Bu nedenle hangi algoritmanın kullanılacağına, her bir algoritma ile elde edilen çözüm sonuçlarının karşılaştırılarak değerlendirilmesi gerekmektedir (Morissette, 2013).

K-ortalama kümeleme çözümlemesinde standart olarak kullanılan algoritma Hartigan & Wong (Hartigan, Wong,1979) algoritmasıdır. Bu algoritmaya göre küme içi toplam değişkenlik, gözlemler ve karşılık gelen ağırlık merkezi arasındaki Euclid uzaklıklarının kareleri toplamı olarak tanımlamaktadır:

$$W(C_k) = \sum_{x_i \in C_k} \begin{matrix} (x_{i-}\mu_k)^2;\ x_i, C_k\ kümesinde \\ yer\ alan\ bir\ gözlem; \mu_k, C_k\ kümesinin \\ ortalamasıdır. \end{matrix}$$

Bu yaklaşıma göre, her gözlemin atandığı küme merkezine uzaklığının kareli toplamı en küçüktür. Küme içi toplam değişkenlik;

$$Küme\ İçi\ Toplam\ Değişkenllik = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \in C_k} (x_{i-} \mu_k)^2$$

Küme içi toplam değişkenlik ölçüsü kümelemenin uygunluğunu ölçmekte ve en küçük olması beklenmektedir.

Seçilen algoritmadan bağımsız olarak kümeleme çözümlemesinde en büyük sorunlardan biri en uygun **k** küme sayısının belirlenmesidir. Uygun küme sayısının belirlenmesi için birkaç deneme yapmak gerekli olabilmektedir. Küme sayısının yaklaşık olarak belirlenmesinde parmak kuralı yaklaşımından yararlanılabilmektedir (Alpar, 2017; Madhulatha, 2012).

$$k \approx \sqrt{n/2}$$

Küme sayısının belirlenmesinde diğer bir yaklaşım Marriott (Marriott,1971) tarafından önerilen hesaplama yöntemidir. **W** grup içi kareler toplam matrisi olmak üzere:

$M = k^2|W|$ eşitliğini sağlayan **M** sayısı küme sayısı olarak alınmaktadır.

Ayrıca küme sayısının belirlenmesinde bilgiye dayalı, Akaike bilgi kriteri (AIC), Bayes bilgi kriteri (BIC), ve Deviance bilgi kriteri (DIC) kullanılmaktadır (Madhulatha, 2012; Akoğul ve Erişoğlu, 2016).

Bir kümeleme algoritmasının sonuçlarını değerlendirme yönergesi, küme geçerliliği (*cluster validity)* terimi olarak adlandırılır. Kümeleme çözümlerinin geçerliğinin araştırılmasında çeşitli yaklaşımlar bulunmaktadır. Bu yaklaşımlardan birincisi, küme analizinin sonuçlarının dışarıdan bilinen sonuçlarla karşılaştırılmasından oluşan dışsal ölçütlere; ikinci yaklaşım, kümeleme analiz sonuçlarının verilere ne kadar iyi uyduğunu değerlendirmek için kümeleme süreci içinden elde edilen bilgileri kullanan içsel ölçütlere; üçüncü yaklaşım, bir kümeleme yapısının diğer kümeleme şemalarıyla karşılaştırılarak değerlendirilmesinden oluşan, aynı algoritma ile ancak farklı parametre değerleri, örneğin küme sayısı ile sonuçlanan göreceli ölçütlere dayanmaktadır (Charrad, vd.,2014). Kümeleme yazınında her bir yaklaşım için kümeleme analizinin sonuçlarını değerlendirmeyi amaçlayan çeşitli göstergeler tanımlanmakta ve önerilmektedir. Kümeleme yazınında, *CH, CCC, Gamma, Gap, Silhouette, Hartigan, Cindex, DB, Ratrowsky, Scott, Mariott, Friedman, Rubin,* Dunn, *Jaccard* gibi çok sayıda değerlendirme göstergesi bulunmaktadır.

Bu göstergelerden biri tamamen yöntemin uygulandığı veri kümesine dayanan Dunn göstergesidir (Dunn, 1974). Dunn göstergesinin amacı, kümenin üyeleri arasında küçük bir değişkenlikle kompakt olan ve küme içi değişkenlikle karşılaştırıldığında farklı kümelerin ortalamalarından yeterince uzakta olduğu yeterince ayrık kümeleri tanımlamaktır. Dunn göstergesi ne kadar büyükse, kümeleme sonuçları o kadar iyidir. Δk, küme içi değişkenliği ve $d(c_i, c_j)$ küme merkezleri arasındaki uzaklık ölçüsü olmak üzere *Dunn* göstergesi, küme içi benzerliğin kümeler arası benzerliğe oranıdır.

$$DI = \min_{j=1\ldots m} \left\{ \min_{j=1\ldots m, i \neq j} \left\{ \frac{d(c_i, c_j)}{\max_{k=1\ldots m} \Delta k} \right\} \right\}$$

Veri kümesi kompakt ve iyi ayrılmış kümeler içermekteyse, kümelerin çapının küçük olması ve kümeler arasındaki mesafenin büyük olması beklenmektedir. Bu nedenle, *Dunn* göstergesi maksimize edilmelidir (Charrad, vd.,2014).

Göstergelerden bir diğeri de veri kümesi için daha önce bilinen çözüm ile yeni çözümün karşılaştırılmasına dayanan *Jaccard* göstergesidir. Bu gösterge genellikle, verilerin sınıflandırıldığı önceki güvenilir bir sınıflandırma olduğunda kullanılmaktadır. Bulunan çözüm ile önceki sınıflandırmanın arasındaki benzerliği, doğru sınıflandırmanın bir yüzdesi olarak hesaplamaktadır. Özetle, kesişim boyutunun (her iki çözümde aynı kümede bulunan durumlar) birleşimin boyutuna (her iki veri kümesindeki tüm durumlar) bölünmesiyle hesaplanmaktadır (Morissette,2013).

4. İLLERİN K-ORTALAMA İLE YAŞAM KALİTESİNE GÖRE SINIFLANDIRILMASI

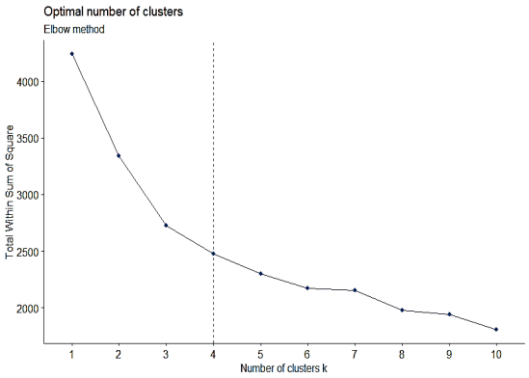
Yaşam kalitesinin illere göre sınıflandırılmasında, veri kümesi olarak Türkiye İstatistik Kurumu'nun ilk kez 2015 yılında 81 il için yayınladığı yaşam gösterge endeksi değerleri kullanılmaktadır. Kümeleme çözümlemesinde *SPSS 20.0* paket yazılımı ile birlikte *R V3.4.3* yazılımından yararlanılmaktadır.

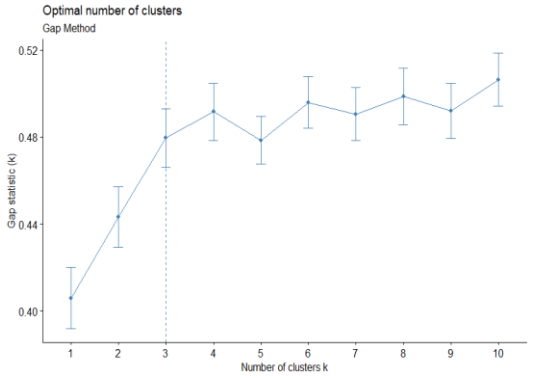
K-ortalama kümeleme çözümlemesinde araştırmaya başlamadan önce **k** küme sayısının belirlenmesi gerekmektedir. Bu çalışmada küme sayısının belirlenmesinde R yazılımında *factoextra* paketinden yararlanılmaktadır. Bu paket içinde yer alan *fviz_nbclust*() işlevi en uygun küme sayısını belirlenmesinde uygun bir çözüm sağlamaktadır. Bu işlev kullanılarak, en uygun küme sayısının belirlenmesinde üç farklı yaklaşımın, *elbow*, *silhoutte* ve *gap* (Tibshirani, vd., 2001) istatistiği, sonuçlarından yararlanılmaktadır. Çözüme başlamadan, veri kümesindeki değişkenler z-ölçülerine dönüştürülmelidir.

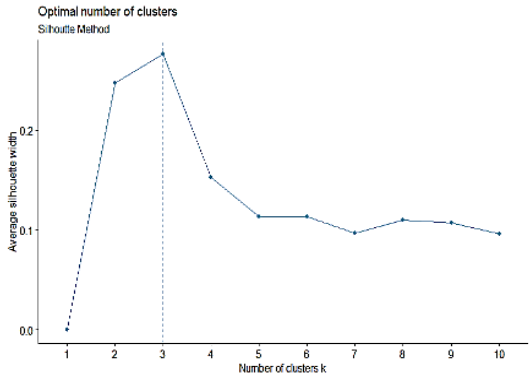
K-ortalama yöntemi, küme içindeki toplam değişkenliğin en küçük olduğu kümeleri tanımlamaya çalışmaktadır. Dolayısıyla, toplam WSS (küme içi kareler toplamı) kümelenmenin kompaktlığını ölçmekte ve mümkün olduğu kadar küçük olması beklenmektedir. *Elbow* yöntemi, toplam WSS'ye küme sayısının bir fonksiyonu olarak bakmakta ve bu yaklaşıma göre küme sayısı öyle belirlenmelidir ki, başka bir kümenin eklenmesi toplam WSS'yi daha iyi geliştirmesin. Bir başka yaklaşımla, *Elbow ölçütü*, bir küme sayısı belirlendiğinde, başka bir küme eklemenin ilişkiyi açıklayacak yeterli bilgi eklememesi gerektiğini ifade etmektedir. Başka bir deyişle, kümeler tarafından açıklanan değişkenlik yüzdesi küme sayısına göre grafiğe dökülürse, ilk kümeler çok daha fazla bilgi ekleyecek, ancak küme sayısı arttıkça belli bir noktada (büküm noktası) küme eklemenin marjinal faydası azalacaktır (Madhulatha, 2012).

Aşağıda verilen grafiklerde, *Elbow, Gap ve Sikhoutte* yöntemleri ile en uygun küme sayısı belirlenmektedir. Grafikler kümeler arasındaki değişkenliği göstermektedir. Küme sayısı artıkça değişkenlik azalmaktadır. Elbow yönteminin çıktısı olan grafikten anlaşılabileceği üzere **k=4** olduğunda bir büküm (elbow) oluşmaktadır. Bu büküm dördüncü kümeden sonra eklenecek yeni bir kümenin değişkenlikte ek bir bilgi sağlamayacağını göstermektedir.

Elbow yaklaşımına göre 4, *Gap* ve *Silhoutte* yaklaşımına göre en uygun küme sayısı 3 olarak belirlenmektedir. Küme sayısının belirlenmesinde SPSS paketinin K-Means Cluster Analysis komutu kullanılarak **k=2,3,4** alınarak deneme yapılmış ve en uygun küme sayısı dört olarak belirlenmiştir.



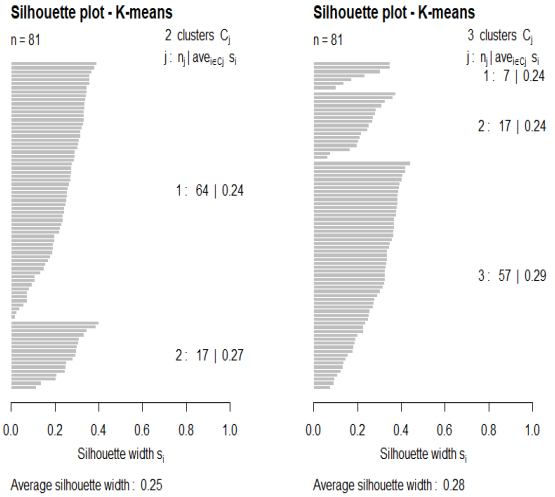




Kümeleme algoritmalarının en büyük problemlerinden biri de herhangi bir küme olmasa da küme oluşturabilmesidir. Bu bağlamda kümeleme çözümlemesinden sonra sonuçların geçerliliğinin araştırılması gerekmektedir. Kümeleme yazınında kümeleme sonuçlarının değerlendirilmesi için çok sayıda geçerlilik ölçüsü geliştirilmiştir. Bu ölçüler genellikle dört grupta toplanmaktadır. Kümelemede kullanılan kümeleme algoritmasında başlangıç küme sayısını değiştirerek en uygun küme sayısını belirleyerek küme yapısını değerlendiren *göreli kümeleme geçerliliği* yaklaşımı. Yapılan küme çözümlemesini daha önceden yapılan araştırma sonucunu kullanarak karşılaştıran *dışsal geçerlilik* yaklaşımı. Dışsal bir bilgiye başvurmadan veri kümesinden üretilen bilgilere dayanarak geçerliliği araştıran *içsel geçerlilik* yaklaşımı. Bir diğeri de içsel geçerlilik yaklaşımının bir uzantısı olan *kümeleme denge geçerliliği* yaklaşımıdır.

İçsel geçerliliğinin araştırılmasında kullanılan Silehouette yaklaşımına göre ortalama siluet genişliği s_i (- tüm veri setinin üzerindeki s_i'nin ortalaması -) grup sayısının seçilmesi için daha yapısal bir ölçüt sağlamak üzere en büyük yapılmalıdır. Kaufman, L. ve Rousseeuw, P. J.'a göre (Kaufman ve Rousseeuw,1990) makul bir sınıflandırmanın 0.5'in üzerinde bir siluet genişliğiyle nitelendirilebileceğini düşünmekte ve küçük bir siluet genişliğinin, örneğin 0.2'nin altındaki bir ortalama genişliğin, önemli bir küme yapısının eksikliği olarak yorumlanması gerektiğine işaret etmektedirler (Everitt,vd.,2011) . Ortalama siluet genişliği ±1 arasında değişen, -1'e yaklaştıkça zayıf bir kümelemeye, +1'e yaklaştıkça güçlü bir kümeleme sonucuna işaret eden bir ölçüdür. Siluet genişliği 0 ya da 0'yakın olduğunda ilgili gözlem iki kümenin arasında yer aldığını, siluet değeri negatif olan gözlemler ise yanlış bir kümede yer aldığını göstermektedir. Siluet genişliği şöyle hesaplanmaktadır: $s_i = (b_i - a_i)/\max(a_i, b_{i)}$. Burada a_i, i. gözlemin, aynı küme içinde yer alan gözlemlerle olan ortalama benzersizlik ölçüsüdür. b_i, i. gözlemin yer almadığı tüm kümelerdeki (C) gözlemlere göre hesaplanan benzersizlik ölçüsünün d(i,C) en küçük değeridir, $b_i = min_C d(i, C)$ (Kassambara, 2017).

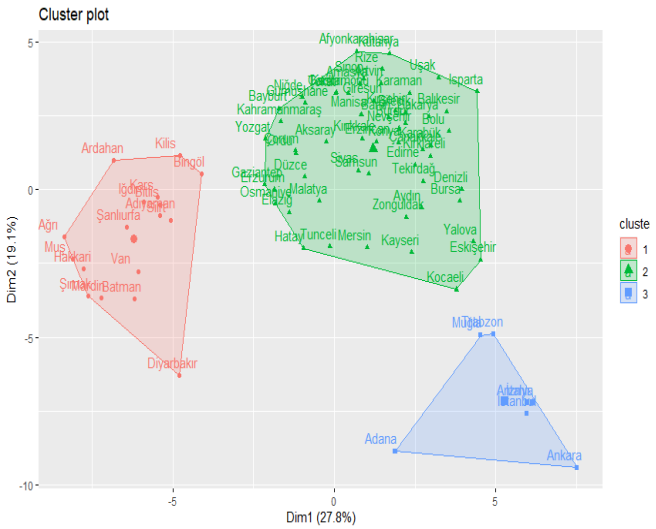
Şekil 1. İçsel Geçerlilik Göstergesi (Siluet Genişliği)



Silehouette yaklaşımı ile elde edilen sonuçlara göre küme sayısı üç alındığında siluet genişliği artmaktadır. Bu sunuca göre en makul küme sayısının üç olması gerekmektedir. Küme sayısı dört alındığında siluet genişliği 0.2'nin altına düşmektedir. Bu sonuçlara göre bu araştırmada yaşam göstergelerine göre illerin sınıflandırılmasında en uygun küme sayısı üç alınmaktadır. Küme sayısının üç ve tüm yaşam endeksi göstergelerinin alınarak yapılan çözümleme sonucunda birinci kümede 7, ikincide 17 ve üçüncüde 57 il kümelenmektedir.

R yazılımında *kmeans()* fonksiyonu kullanılarak elde edilen kümeleme sonuçları ***Tablo1'de*** ve ***Şeki*** 1'de görülmektedir. *kmeans()* fonksiyonu kümeleme işlemini varsayımsal olarak *Hartigan-Wong* algoritmasını kullanarak gerçekleştirmektedir. Yöntemin uygulanışında *Lloyd*, *Forgy*, *MacQueen* algoritmaları seçenek olarak kullanılabilmektedir. *Lloyd* ve *Forgy* algoritmaları ile yapılan kümelemede yakınsama sağlanamamıştır. **Tablo 1**'de verilen kümeler *Hartigan-Wong* algoritması kullanılarak ve küme sayısının üç alınarak elde edilen sonuçlarıdır.

Şekil 2. İllerin Yaşam Gösterge değerlerine göre kümelenmesi (k=3)



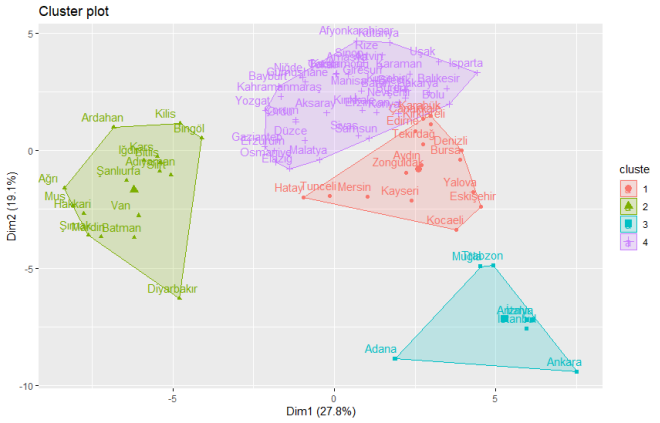
Tablo 1. İllerin Yaşam Gösterge değerlerine göre sınıflandırılması

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

Şekil 2'ye bakıldığında 1. kümede yer alan illerin diğer illerden oldukça farklılık gösterdiği gibi kendi içlerinde de açık bir farklılık olduğu görülmektedir. Ankara, aynı kümede yer alan illerle karşılaştırıldığında yaşam kalitesi açısından oldukça ileri olduğu; Antalya ve İstanbul arasında pek bir farklılık olmadığı, ancak bu kümede yer alan Adana'nı diğerlerinden oldukça farklı olduğu anlaşılmaktadır. Ayrıca bu kümedeki iller sosyoekonomik açıdan gelişmiş illerdir. 2. kümede yer alan Diyarbakır aynı kümedeki diğer illerden farklılık göstermektedir. Ayrıca 2. kümede yer alan illerin Doğu ve Güney Doğu illerinden olması şaşırtıcı olmaması gerekmektedir. Bu illeri sosyoekonomik açıdan az gelişmiş iller olarak sınıflandırmak mümkündür. Dolayısıyla kamu politika yapıcılarının bu kümede yer alan illere daha fazla kaynak aktarması gerektiği açıkça görülmektedir. Bu kümelemede şaşırtıcı olan sonuç, Aydın, Balıkesir, Bursa, Eskişehir, Gaziantep,

Kocaeli vb., sanayi ve ticaretin gelişmiş olduğu illerin görece daha az gelişmiş illerle birlikte yer almakta olmasıdır. Şekil 1'e dikkatle bakıldığında söz konusu illerin 3. kümede yer alan diğer illerden farklı oldukları görülmektedir. Bu farklılaşmayı da gözeterek kümeleme çözümlemesi küme sayısı dört alınarak yeniden yapılmıştır. Aşağıdaki Şekil 3'de, küme sayısı dört alındığında bu farklılaşmanın öne çıktığı açıkça görülmektedir.

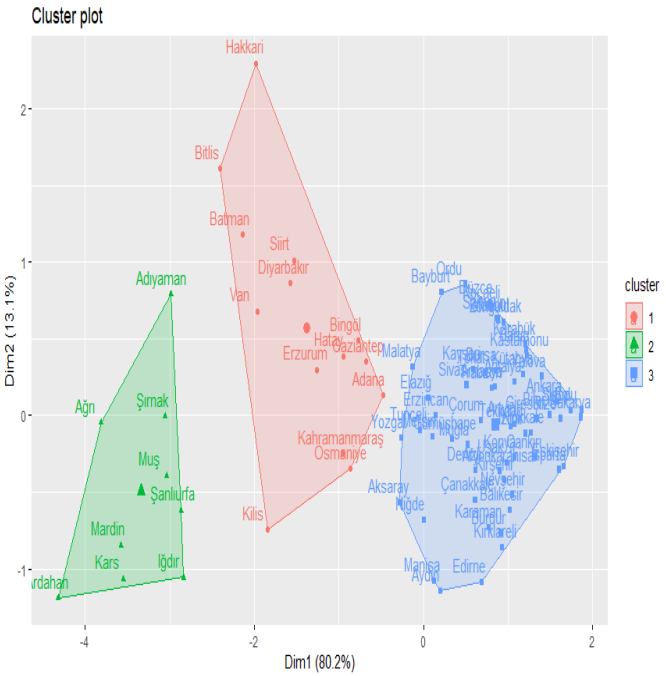
Şekil 3. İllerin Yaşam Gösterge değerlerine göre kümelenmesi (k=4)



Yaşam endeksi göstergelerine göre (konut, çalışma hayatı, gelir ve servet, sağlık, eğitim, çevre, güvenlik, sivil katılım, altyapı hizmetlerine erişim ve sosyal yaşam göstergeleri) daha ayrıntılı bir çözümleme yapılarak genel ve yerel ölçekte kamu politika yapıcılarına yol göstermesi açısından her bir boyutun ayrı ele alınması gerekliliği ortaya çıkmaktadır.

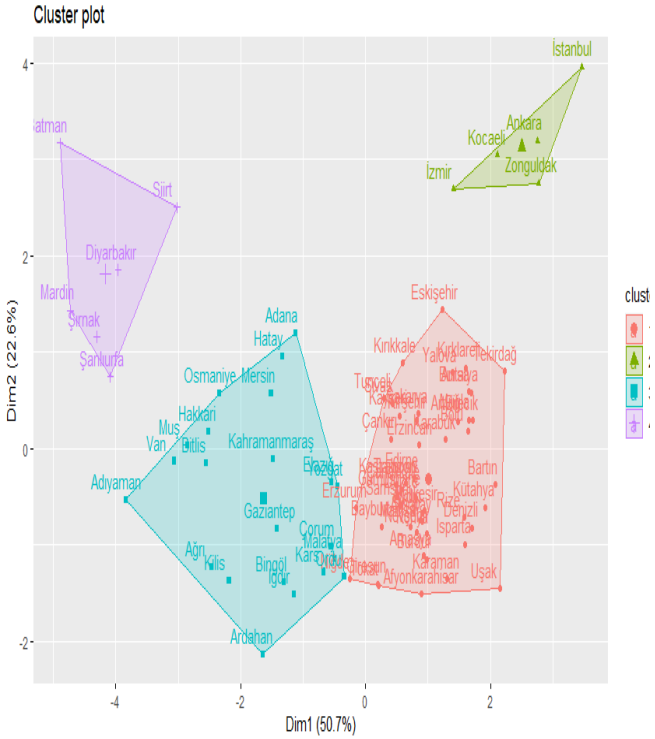
Kişi başına düşen oda sayısı, tuvalet varlığı oranı ve konutun kalitesinde sorun yaşayanların oranlarını içeren özellikleri itibariyle yaşam kalitesinde önemli bir yeri olan konut boyutuna göre yapılan kümeleme sonucuna (siluet genişliği 0.51) göre illeri en uygun üç kümede toplamak gerektiği belirlenmiştir. Küme sayısın artırılması durumunda siluet genişliği, dolayısıyla kümeleme çözümlemesinin iç geçerliği azalmaktadır. Şekil 4'de illerin konutların üç farklı alt boyutuna göre kümelenmesi verilmektedir. Konut özelliklerine göre elde sonuçlara göre, doğu ve güneydoğu bölgelerinde bulunan Ardahan, Ağrı, Adıyaman, Şırnak, Muş, Şanlıurfa, Iğdır, Mardin ve Kars illerinde oda sayısı, tuvalet varlığı ve konut kalitesi açısından sorun yaşandığı; kamu ve yerel politikacıların bu illere daha fazla konut desteği vermesi gerektiği ortadadır.

Şekil 4. İllerin Konut boyutuna göre kümelenmesi (k=3)



Yaşam endeksi göstergelerinin bir diğer boyutu çalışma hayatı (istihdam oranı, işsizlik oranı, ortalama günlük kazanç, işinden memnuniyet oranı, temel ihtiyaçlarını karşılanmayanların oranı, vb.) boyutudur. Çalışma hayatı boyutuna ilişkin kümelemede, k=3 için siluet genişliği 0.39, k=4 için 0.35 bulunmuştur. Küme sayısının dört alınarak daha ayrıntılı bir tablo ortaya konacağı düşünülerek çözümlemede küme sayısı dört alınmaktadır. Çözümleme sonucunda 1. kümede 48, 2. kümede 5, 3. kümede 22 ve 4. kümede 6 il yer almaktadır.

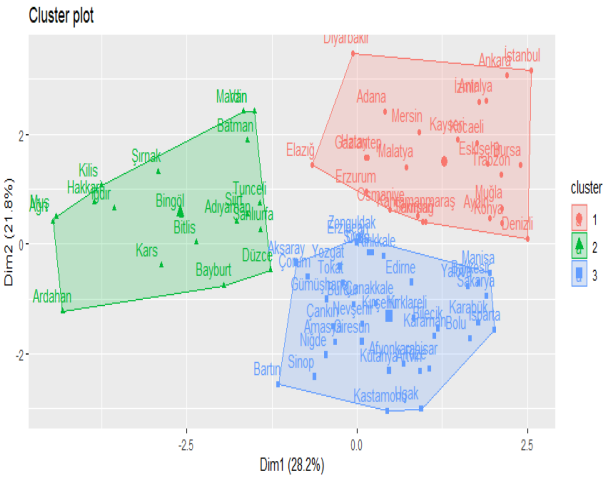
Şekil 5. İllerin Çalışma Hayatı boyutuna göre kümelenmesi (k=4)



Şekil 5 illerin çalışma hayatı boyutuna göre kümelenmesini göstermektedir. 2. kümede yer alan illerin (İzmir, Zonguldak, Ankara, Kocaeli ve İstanbul) çalışma hayatı boyutunda diğer illerden oldukça farklı bir konumda olduğu görülmektedir. 4. kümede yer alan ve çalışma hayatı açısından göreceli sorunlu olan illerin çoğunluğunun güney doğu Anadolu bölgesi illeri olması da dikkat çekmektedir. Bu illerde sanayi gelişmişlik düzeyinin az, tarımsal üretimin yoğun ve aşirete dayalı sosyal bir hayatın varlığı bu sonucu doğru kılmaktadır. 3. kümede yer alan illerin bir kısmı da de yine aynı sosyo-kültürel yapıya sahip ve az sanayi gelişmişliğine sahip olmasına karşın başka faktörlerin etkisiyle olsa gerek bu kümenin içinde yer almaktadırlar.

Bebek ölüm hızı, doğuştan beklenen yaşam süresi gibi en temel yaşam göstergesi boyutunun incelenmesi ayrıca bir önemli konudur. İnsani gelişmişlik göstergesinin temel bileşenlerinden biri de sağlıktır. Sağlık halen doğumda beklenen yaşam süresi ile ölçülmektedir. Bu beklentiye bağlı olarak çalışmada illerin sağlık boyutuna göre de değerlendirilmesi gerektiği ortadadır. Sağlık boyutunu etkileyen bebek ölüm hızı, doğuştan beklenen yaşam süresi, sağlığından memnuniyet, kamunun sağladığı sağlık hizmetlerinden memnuniyet yanında çevre sağlığı ile ilgili hava kirliliği, orman alanı, atık hizmeti, gürültü kirliliği ve belediyenin temizlik hizmetlerinden memnuniyet düzeyleri; şebeke suyu erişim ve kanalizasyon hizmetlerine erişimi de katarak bir bütün olarak değerlendirmenin daha doğru bir yaklaşım olacağı düşüncesiyle çözümleme yapılmıştır. Küme sayısının belirlenmesinde önceki çözümlemelerde olduğu siluet genişliği göz önünde tutulmuştur. Küme sayısının üç, dört ve beş alınmasıyla yeterli bir siluet genişliğine ulaşılamadığı gözlenmiş ve son olarak küme sayısı iki alınarak çözümleme yapılmıştır. Ancak, küme sayısının iki ya da üç alınması siluet genişliğinde fazla bir farklılık yaratamadığından illeri üç kümede toplamak daha ayrıntılı bilgi sağlığı sonucuna varılmaktadır. Şekil 6'da kümeleme sonuçlarının görselliği verilmektedir.

Şekil 6. İllerin Sağlık Boyutuna göre Kümelenmesi (k=3)

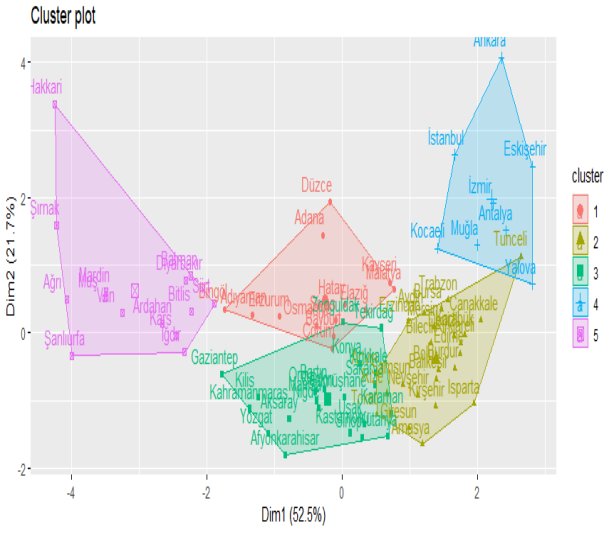


Sağlık boyutu tüm alt boyutları ile birlikte değerlendirildiğinde sağlık hizmeti açısından en az gelişmiş illerin (ikinci küme) yine doğu ve güney doğu illerinden oluştuğu görülmektedir. Batı illerinden sadece Düzce ve Karadeniz'den Bayburt

bu kümeye katılmaktadır. Sağlık kalitesi açısından en kötü iller Ağrı, Muş ve Ardahan olarak görülmektedir. Sanayileşmenin ve eğitim kalitesinin yüksek olduğu 1. kümede yer alan illerde yaşam kalitesinin de daha iyi olduğu görülmektedir. 3. Kümede yer alan illerin de genellikle orta Anadolu ve Karadeniz bölge illeri olduğu saptanabilir.

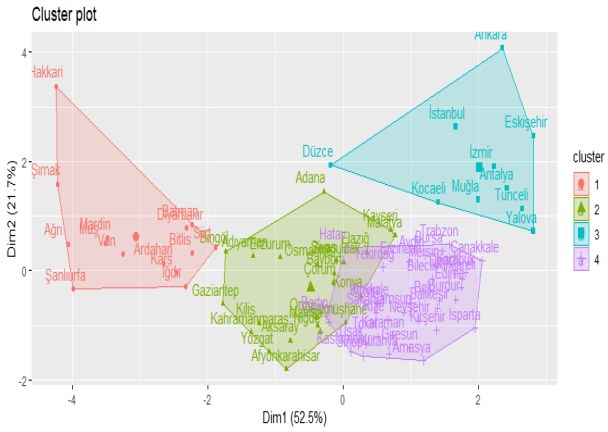
Yaşam kalitesinin diğer önemli bir boyutu da eğitimdir. İllerin eğitim kalitesinin ölçülmesinde okullaşma oranı, TEOG yerleştirmesine esas puan, YGS puanı, fakülte ve yüksekokul mezunu, kamu eğitim hizmetlerinden memnuniyet ve günümüz koşullarında eğitimi etkileyen önemli bir unsur olan internet aboneliği boyutları da katılarak bir başka çözümleme daha yapılmıştır. Siluet genişliği küme sayısı 4 ve 5 için (0.26, 0.25) yüksek çıkmaktadır. Eğitim açısından da bakıldığında önceki kümeleme sonuçlarından farklı sonuçlar elde edilememektedir. Önceki çözümleme sonuçlarında olduğu gibi doğu ve güney doğu illeri yine eğitim açısından da kötü düzeylere sahip iller olmaktadır.

Şekil 7. İllerin Eğitim Açısından Kümelenmesi (K=5)



Bu kümelenmede en ilginç il Tunceli olmaktadır. 2. kümede yer almasına karşın eğitim açısından en gelişmiş 1. kümede yer alan illerin arasına karışmaktadır. Beklendiği gibi 5. kümede yer alan doğu ve güney illeri eğitim kalitesi açısından en düşük iller arasında sınıflandırılmaktadır. Bir farklı bakış açısı vermesi bakımından illerin dört kümeye göre çözümlemesi Şekil 8'de verilmektedir.

Şekil 8. İllerin Eğitim Açısından Kümelenmesi (k=4)



Dörtlü ve beşli kümeleme görüldüğü şekilde sonuçları değiştirmemektedir. Beşli kümelemede 1., 2., ve 3. kümede yer alan iller yine birbiri içine geçen iki farklı kümeye ayrılmaktadır.

SONUÇ

Bu çalışmada K-ortalama kümeleme çözümlemesi kullanılarak, Türkiye'deki iller yaşam gösterge değerleri açısından gruplandırılmaya çalışılmıştır. Tüm yaşam göstergeleri ele alındığında elde edilen sonuçlar illerin üç ya da dört grupta toplanabileceğini göstermektedir. İktisadi açıdan bakıldığında 3. kümede yer alan illerin ekonomik açıdan ileri ve 2. kümede yer alan illerin de ekonomik açıdan geri kalmış iller olduğu, 2. ve 4. kümede yer alan illerin ise nispeten ekonomik olarak birbirine benzer iller olduğu görülmektedir. Yaşam göstergelerinin konut, sağlık, eğitim ve çalışma hayatı gibi farklı boyutlarına bakıldığında, hemen hemen her boyutta doğu ve güney doğu illerinin en alt düzeyde kaldıkları gözlemlenmektedir. Buna karşın batı ve batı bölgelerine yakın bölgelerde yer alan illerin genel yaşam göstergeleri açısından aralarında farklılık da olsa nispeten doğu ve güneydoğu bölgelerinde olan illere göre daha iyi oldukları

gözlemlenmektedir. Özet olarak bu çözümleme sonuçlarına göre, kamu ve yerel karar vericilerin kümeler arasındaki bu farklılıkları giderecek politikalar geliştirmeleri gerektiği ortadadır. Sağlık, eğitim, konut, çalışma hayatı gibi insani gelişme endeksini etkileyen bu göstergelerin gelişiminde merkezi yöntemin kamu kaynaklarının dağıtımında daha adil olması gerektiğini göstermektedir.

ARAŞTIRMACILARIN KATKI ORANI BEYANI VE ÇIKAR ÇATIŞMASI BİLDİRİMİ

Araştırmacılar herhangi bir çıkar çatışması bildirmemiştir.

Araştırmacılar makaleye ortak olarak katkıda bulunmuşlardır.

KAYNAKÇA

Akın,H.B., Eren, Ö., (2012). *OECD Ülkelerinin Eğitim Göstergelerinin Kümeleme Analizi Ve Çok Boyutlu Ölçekleme Analizi İle Karşılaştırmalı Analizi,* **Öneri.**C.10.S.37., 175-181.

Akoğul, S., Erişoğlu, M., (2016). *A Comparison of Information Criteria in Clustering Based on Mixture of Multivariate Normal Distributions,* **Math. Comput. Appl.,** 21, 34; doi:10.3390/mca21030034

Alpar, R., (2017). **Uygulamalı Çok Değişkenli İstatistiksel Yöntemler,** Detay Yayıncılık, Ankara

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., (2014). *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set",* **Journal of Statistical Software,** Volume 61, Issue 6.

Choudhary, A.,(2016). *Survey on K-Means and Its Variants,* **International Journal of Innovative Research in Computer and Communication Engineering,** Vol. 4, Issue 1

Cleff, T., (2019). **Applies Statistics and Multivariate Data Analysis for Business and Economics A Modern Approach Using SPSS,** Stata and Excel, Springer, Switzerland

Çelik, Ş., (2013). *Kümeleme Analizi İle Sağlık Göstergelerine Göre Türkiye'deki İllerin Sınıflandırılması,* **Doğuş Üniversitesi Dergisi,** 14 (2) 2013, 175-194.

Dinçer,B., Özaslan, M., Kavasoğlu, T.,(2003). *İllerin Sosyo-Ekonomik Gelişmişlik Sıralaması Araştırması,* **DPT,** Bölgesel Gelişme ve Yapısal Uyum Genel Müdürlüğü.

Dunn, J. C., (1974). *Well-Separated Clusters and Optimal Fuzzy Partitions,* **Journal of Cybernetics**, Volume 4, Issue 1

Erol, E.(2013). *Türkiye Ve Avrupa Birliği Üyesi Ülkelerin Sosyo - Ekonomik Gelişmişlik Düzeylerinin Karşılaştırmalı Analizi.* **Sosyal ve Beşeri Bilimler Dergisi,** 5 (1) , 198-208.

European Comision, (2011). **Eurobarometer Qualitative Studies,** Well-Being Aggregate Report

Everitt,B.S., Landau,S., Leese, M., Stahl, D., (2011). **Cluster Analysis,** John Wiley & Sons, Ltd.

Hartigan, J. A., Wong, M. A, (1979). *Algorithm AS 136: A K-Means Clustering Algorithm",* **Journal of the Royal Statistical Society.** Series C (Applied Statistics) , 1979, Vol. 28, No. 1 (1979), pp. 100-108

İlknur, Ö., (1998). *İlçelerin Sosyo-Ekonomik Gelişmişlik Sıralaması ve Gruplandırılmasına İlişkin Bir Çalışma,* **Hazine Dergisi,** S.11, 41-61.

Karabulut, M., Gürbüz, M., Sandal, E.K., (2004). *Hiyerarşik Kluster(Küme) Tekniği Kullanılarak Türkiye'de İllerin Sosyo-Ekonomik Benzerliklerinin Analizi,* **Coğrafi Bilimler Dergisi,** C.2, S.2, 71-85.

Kassambara, A., 2017. **Practical Guide to Cluster Analysis in R Unsupervised machine Learning,** STHDA (http://www.sthda.com)

Kaufman, L. and Rousseeuw, P. J., (1990). **Finding Groups in Data. An Introduction to Cluster Analysis.** John Wiley & Sons, Inc., New York.

Koç, S., (2001). *Türkiye'de İllerin Sosyo-Ekonomik Özelliklere Göre Sınıflandırılması,* **5.Ulusal Ekonometri ve İstatistik Sempozyumu,** Çukurova Üniversitesi, Adana.

Madhulatha, T.S., (2012). *An Overvıew On Clusterıng Methods,* **Journal of Engineering,** Vol. 2(4) pp: 719-725

Marriott, F. H. C., (1971). *Practical Problems in a Method of Cluster Analysis,* **Biometrics,** Vol. 27, No. 3, pp. 501-514.

Morissette, L., Chartier, S., (2013). **The K-Means Clustering Technique: General Considerations And İmplementation in Mathematica,** Vol.9(1),15-24.

Morissette,L., Chartier, S., (2013). **The k-means clustering technique: General considerations and implementation in Mathematica, Tutorials in Quantitative Methods for Psychology 2013,** Vol. 9(1), p. 15-24.

Sandal, E. K., Karabulut, M., (2005). *Sosyo-Ekonomik Kriterler Bakımından Türkiye'nin Konumu ve Avrupa Birliği,* **Fırat Üniversitesi Sosyal Bilimler Dergisi,** C.15, S.1, 1-14.

Sheikholeslami C, Chatterjee S, Zhang A., (2000). *WaveCluster: A Multi-Resolution Clustering Approach for Very Large Spatial Database*. **The International Journal on Very Large Data Bases,** 8(3–4), 289–304.

Şahin, M., Hamarat, B., (2002). *Avrupa Birliği ve OECD Ülkelerinin Sosyo_Ekonomik Benzerliklerinin Fuzzy Kümeleme Analizi ile Belirlenmesi,* **ODTÜ Uluslararası Ekonomi Kongresi,** VI, Ankara, 1-19.

Tekin, B., (2015). *Temel Sağlık Göstergeleri Açısından Türkiye'deki İllerin Gruplandırılması: Bir Kümeleme Analizi Uygulaması,* **Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,** Cilt 5, Sayı 2, ss.389-416

Tibshirani, R, Walther, G., Hastie, T., (2001). *Estimating The Number Of Clustersin A Data Set Via Gap Statistic,* **Royal Statistical Society,** 63, Part 2, pp.411-423

Turanlı, M., Özden, Ü. H., Türedi, S., (2006). *Avrupa Birliği'ne Aday Ve Üye Ülkelerin Ekonomik Benzerliklerinin Kümeleme Analiziyle İncelenmesi.* **İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi**, Yıl:5 Sayı:9 Bahar 2006/1 s.95-108

Yılancı, V., (2010). *Bulanık Kümeleme Analizi İle Türkiye'deki İllerin Sosyoekonomik Açıdan Sınıflandırılması,* **Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,** C.15, S.3 s.453-470.