

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**ÇOK DEĞİŞKENLİ ANALİZLERDE KULLANILAN UZAKLIK
VE BENZERLİK ÖLÇÜTLERİNİN KULLANIM YERLERİ
ÜZERİNE UYGULAMALI BİR ÇALIŞMA**

Harun Fatih TOMBAK

Hacettepe Üniversitesi
Sağlık Bilimleri Enstitüsü Yönetmeliğinin
Biyoistatistik Programı İçin Öngördüğü
BİLİM UZMANLIĞI TEZİ
olarak hazırlanmıştır.

TEZ DANIŞMANI
Doç. Dr. Reha ALPAR

ANKARA
1996

Sağlık Bilimleri Enstitüsü Müdürlüğüne;

Bu çalışma jürimiz tarafından Biyoistatistik Programında bilim uzmanlığı tezi olarak kabul edilmiştir.

Tez Danışmanı : Doç. Dr. Reha Alpar Hacettepe Üniversitesi

Üye : Doç. Dr. Osman Saraçbaşı Hacettepe Üniversitesi

Üye : Doç. Dr. Ergün Karaağaoğlu Hacettepe Üniversitesi

ONAY :

Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulu'nun kararıyla kabul edilmiştir.

Prof. Dr. A. Atilla Hıncal

Enstitü Müdürü

ÖZET

Bu çalışmada, bazı çok değişkenli analizlerde (kümeleme analizi başta olmak üzere) n tane bireyin uzaklıklarından elde edilen $n \times n$ boyutlu D uzaklıklar matrisini elde etmede kullanılan uzaklık ve benzerlik ölçütleri tanımlanmıştır. Verilerin ölçüm biçimine (aralık (nicel), nitel (kesikli), frekans sayıları ve ikili) uygun uzaklık ölçütleri belirlenmiş ve ilgili sorunun çözümünde hangi uzaklık ölçütünün daha iyi olacağı tablo halinde gösterilmiştir. Kümelenenmiş nesneleri, benzerliklerine göre kümelemede kullanılan uzaklık ve benzerlik ölçütlerinin birbirlerine göre üstünlükleri tartışılmıştır.

Daha sonra uygulama; Afrika, Asya ve Avrupa ülkelerinden toplam 15 ülkenin ölçümle belirtilen 9 değişkeni üzerinde yapılmıştır. Bu ülkelere ait 15×9 boyutlu veri matrisinden yola çıkılarak verinin ölçüm biçimine uygun uzaklık (benzemezlik) ölçütleri yardımı ile uzaklıklar matrisi, benzerlik (yakınlık) ölçütleri yardımı ile de yakınlıklar matrisi bulunmuş ve bu matrisler yardımıyla en yakın komşuluk yöntemi kullanılarak kümeleme analizi yapılmıştır. Kümeleme analizinin sonucu yorumlanarak kümelemede kullanılan uzaklıklar (distance) veya yakınlıklar (proximity) matrisinin diğerlerine göre avantaj ve dezavantajları tartışılmıştır.

Anahtar Kelimeler: uzaklık, yakınlık, benzemezlik, benzerlik, kümeleme analizi.

ABSTRACT

In this study, distance and similarity measures which are used for handling $n \times n$ distance matrix, D were described. Distance matrix, D is generally used in many multivariate analyses. According to the type of data (interval, discrete, frequency counts, binary) suitable distance measures have been indicated and these measures are shown in a table. Superiorities of distance and similarity measures which are used for clustering the unclustered objects were discussed.

Application were carried out for the data which consist of 9 continuous variables of the 15 countries from Afrika, Asia and Europe. By using a 15×9 data matrix, distance and similarity matrices have been constructed and by using the single linkage method, cluster analysis was employed.

Key Words: distance, proximity, dissimilarity, similarity, cluster analysis.

İÇİNDEKİLER

	SAYFA
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
GİRİŞ	1
AMAC	2
I- GENEL BİLGİLER	3
1.1. Ölçüt Seçimi	3
1.2. Uzaklık ve Benzerlikler	3
1.2.1. Benzemezlik (Dissimilarity)-Uzaklık (Distance)	3
1.2.2. Benzerlik (Similarity)-Yakınlık (Proximity)	4
II- YAYGIN OLARAK KULLANILAN UZAKLIK VE BENZERLİK	
ÖLÇÜTLERİ	5
2.1. Uzaklık ve Benzerlik Ölçütlerine İlişkin Genel Bilgiler	5
2.2. Aralık (Interval), Nicel-Sürekli (Quantitative) Veriler ve Bunlara İlişkin	
Uzaklık-Benzerlik Ölçütleri	5
Uzaklık Ölçütleri	5
2.2.1. Ölçekli Öklit Uzaklık Ölçütü (Scale Euclidean Distance Measure)	5
2.2.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)	6
2.2.3. Kare Öklit Uzaklık Ölçütü (Squared Euclidean Distance Measure)	6
2.2.4. Chebychev Uzaklık Ölçütü (Chebychev Distance Measure)	7
2.2.5. Manhattan City-Block Uzaklık Ölçütü (Manhattan City-Block Distance Measure)	7
2.2.6. Minkowski Uzaklık Ölçütü ve Özel Durumlar (Minkowski Distance Measure)	7
2.2.7. Mahalanobis Uzaklık Ölçütü (Mahalanobis Distance Measure)	8
2.2.8. Hotelling T^2 Uzaklık Ölçütü (Hotelling T^2 Distance Measure)	9
2.2.9. Canberra Uzaklık Ölçütü (Canberra Distance Measure)	9
2.2.10. Karl Pearson Uzaklık Ölçütü (Karl Pearson Distance Measure)	10
2.2.11. Penrose Uzaklık Ölçütü (Penrose Distance Measure)	10
Benzerlik Ölçütleri	11
2.2.12. Kosinüs Benzerlik Ölçütü (Cosine Similarity Measure)	11
2.2.13. Pearson İlişki Katsayısı (Pearson Correlation Coefficient)	11

2.3. Nitel-Kesikli (Qualitative) Veriler ve Bunlara İlişkin Uzaklık Ölçütleri	12
Uzaklık Ölçütleri	12
2.3.1. Ölçekli Öklit Uzaklık Ölçütü (Scale Euclidean Distance Measure)	12
2.3.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)	12
2.3.3. Mahalanobis-Benzeri Uzaklık Ölçütü (Mahalanobis-Like Distance Measure)	13
2.3.4. Bhattacharyya Uzaklık Ölçütü (Bhattacharyya Distance Measure)	14
2.4. Frekans Sayıları (Frequency Counts) ve Bunlara İlişkin Uzaklık Ölçütleri	14
Uzaklık Ölçütleri	15
2.4.1. Ki-Kare Uzaklık Ölçütü (Chi-Square Distance Measure)	15
2.4.2. Fi-Kare Uzaklık Ölçütü (Phi-Square Distance Measure)	15
2.5. İkili Veriler (Binary Data) ve Bunlara İlişkin Uzaklık-Benzerlik Ölçütleri	15
Uzaklık Ölçütleri	16
2.5.1. Kare Öklit Uzaklık Ölçütü (Squared Euclidean Distance Measure)	16
2.5.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)	16
2.5.3. Büyüklük Farkları Uzaklık Ölçütü (Size Difference Distance Measure)	18
2.5.4. Biçim Farkları Uzaklık Ölçütü (Pattern Difference Distance Measure)	18
2.5.5. Değişim Uzaklık Ölçütü (Variance Distance Measure)	18
2.5.6. Durum Uzaklık Ölçütü (Shape Distance Measure)	19
2.5.7. Lance ve Williams Uzaklık Ölçütü (Lance & Williams Distance Measure)	19
Benzerlik Ölçütleri	19
2.5.8. Russell ve Rao Benzerlik Ölçütü (Russell & Rao Similarity Measure)	19
2.5.9. Basit Benzerlik Ölçütü (Simple Similarity Measure)	19
2.5.10. Jaccard Benzerlik Ölçütü (Jaccard Similarity Measure)	19
2.5.11. Parçalı Benzerlik Ölçütü (Dice Similarity Measure)	20
2.5.12. Rogers ve Tanimoto Benzerlik Ölçütü (Rogers & Tanimoto Similarity Measure)	20
2.5.13. Sokal ve Sneath Benzerlik Ölçütü-1 (Sokal & Sneath Similarity Measure-1)	20
2.5.14. Sokal ve Sneath Benzerlik Ölçütü-2 (Sokal & Sneath Similarity Measure-2)	20
2.5.15. Sokal ve Sneath Benzerlik Ölçütü-3 (Sokal & Sneath Similarity Measure-3)	21
2.5.16. Kulczynski Benzerlik Ölçütü-1 (Kulczynski Similarity Measure-1)	21
2.5.17. Kulczynski Benzerlik Ölçütü-2 (Kulczynski Similarity Measure-2)	21
2.5.18. Sokal ve Sneath Benzerlik Ölçütü-4 (Sokal & Sneath Similarity Measure-4)	21
2.5.19. Hamann Benzerlik Ölçütü (Hamann Similarity Measure)	22

2.5.20. Goodman ve Kruskal Lamda Benzerlik Ölçütü (Goodman & Kruskal Lambda Similarity Measure)	22
2.5.21. Anderberg'in D Benzerlik Ölçütü (Anderberg's D Similarity Measure)	22
2.5.22. Yule'in Y Benzerlik Ölçütü (Yule's Y Similarity Measure)	23
2.5.23. Yule'in Q Benzerlik Ölçütü (Yule's Q Similarity Measure)	23
2.5.24. Ochiai Benzerlik Ölçütü (Ochiai Similarity Measure)	23
2.5.25. Sokal ve Sneath Benzerlik Ölçütü-5 (Sokal & Sneath Similarity Measure-5)	23
2.5.26. Fi 4-Nokta Korelasyon Benzerlik Ölçütü (Fi 4-Point Correlation Similarity Measure)	23
2.5.27. Yayılım Benzerlik Ölçütü (Dispersion Similarity Measure)	24
Tablo 2.1. Uzaklık ve Benzerlik Ölçütlerinin Ölçüm Biçimlerine Göre Dağılımı	24
2.6. Uzaklık ve Benzerlik Ölçütlerinin Özellikleri	26
2.7. Verilerin Standartlaştırılması	27
III- YÖNTEM VE UYGULAMA	30
3.1. Uygulama Çerçevesi ve Analiz Yöntemi	30
3.2. Bulgular	34
3.2.1. Aralık Veriler ile İlgili Bulgular	34
3.2.2. Nitel Veriler ile İlgili Bulgular	47
3.2.3. Frekans Sayıları ile İlgili Bulgular	51
3.2.4. İkili Veriler ile İlgili Bulgular	53
3.3. Tartışma ve Sonuç	62
KAYNAKLAR	69
EKLER	
Ek 1. Uygulamada Kullanılan Değişken Değerleri ile Bu Değişkenlere İlişkin Toplam, Ortalama, Standart Sapma ve Z-Skoru Değerleri	
Ek 2. Afrika, Asya ve Avrupa Ülkelerinin Kovaryans Matrisleri	
Ek 3. Afrika, Asya ve Avrupa Ülkelerinin Ortak Kovaryans Matrisi ile Bu Matrisin Tersi	
Ek 4. Uygulamada Kullanılan Değişkenlere İlişkin Nitelik ve İkili Veriler	

GİRİŞ

Çok değişkenli analiz $n \times p$ boyutlu X' veri matrisi ile ilgilenmektedir. Bazı analizlerde (kümeleme analizi başta olmak üzere) X' veri matrisi yerine n tane bireyin uzaklıklarından elde edilen $n \times n$ boyutlu D uzaklıklar matrisi kullanılmaktadır (Tatlidil, 1992).

Ayrıca, çok boyutlu ölçeklemede, çok yönlü nesnelerin yakınlık (proximity) matrisiyle verilen verilerinden yararlanılmaktadır.

Uzaklık (distance) ya da benzemezlik (dissimilarity) bilgileri kullanılarak nesnelerin (burada nesne, birden çok değişken içeren bir veri kümesini belirtmektedir) geometrik konumlarının belirlenmesi, şekillendirilmesi uzun zamandır ilgi duyulan bir konudur. Bunların yanı sıra benzerlik (similarity) ölçütleri de uzaklık ya da benzemezlik ölçütleri yerine kullanılabilir.

Kümeleme analizi, faktör analizi ve çok boyutlu ölçekleme gibi çoğu istatistiksel işlemlerin başlangıç noktası, çiftlerin durumlarını içeren değişkenlerden elde edilen uzaklık veya yakınlık matrisleridir. Bu ölçütlerden yola çıkılarak kümeler, faktörler, yapılar ve boyutlar tanımlanmaktadır. Benzerlik ve benzemezlikte birçok farklı ölçüt sürekli olarak kullanılmaktadır. İki değişken arasındaki benzerliğin ölçüsünde en çok kullanılan Pearson İlişki Katsayısıdır. Bu katsayının büyük değere sahip olması iki değişkenin benzerliğini, küçük değere sahip olması ise benzemezliğini göstermektedir.

Benzerlik ölçütleri iki nesne arasındaki yakınlık değerini vermekteyken, benzemezlik ölçütleri ise iki nesne arasındaki uzaklık değerini vermektedir. Büyük benzemezlik değeri iki nesnenin birbirinden çok farklı olduğunu, büyük benzerlik değeri ise bu iki nesnenin birbirine çok yakın olduğunu göstermektedir.

AMAÇ

Ülkemizde, uzaklık ve benzerlik ölçütleri ile ilgili doğrudan bir çalışma yapılmamış ancak, başta kümeleme analizi olmak üzere diğer bazı çok değişkenli yöntemler üzerine yapılan çalışmalarda bu konuya sıkça değinilmiştir. Bu çalışmanın amacı aşağıdaki başlıkları içermektedir;

1. Bazı çok değişkenli analizlerde (kümeleme analizi başta olmak üzere) n tane bireyin uzaklık veya benzerliklerinden elde edilen $n \times n$ boyutlu D uzaklıklar veya yakınlıklar matrisini elde etmede kullanılan uzaklık ve benzerlik ölçütlerini tanımlamak,
2. Verilerin ölçüm biçimine (aralık, nitel, iki durumlu ve frekans sayıları) uygun uzaklık ve benzerlik ölçütlerini belirlemek ve ilgili sorunun çözümünde, hangi uzaklık ve benzerlik ölçütünün daha iyi olacağını tablo halinde göstermek,
3. Kümelennmemiş nesneleri, benzerliklerine göre kümelemede kullanılan uzaklıklar veya yakınlıklar matrisini bulmakta kullanılan uzaklık ve benzerlik ölçütlerinin birbirlerine göre üstünlüklerini tartışmak ve bu konuda yapılacak olan çalışmalara biraz da olsa ışık tutabilmektir.

I- GENEL BİLGİLER

1.1. ÖLÇÜT SEÇİMİ

Benzemezlik veya benzerlik ölçütlerinin kullanılmasına karar verirken, verilerin ölçüm biçimi dikkate alınmaktadır. Çünkü, aralık (interval), nitel (qualitative), frekans sayıları (frequency counts) ve iki durumlu (binary) veriler için geliştirilmiş özel ölçütler vardır. Ayrıca, orijinal verinin bir şekilde standardize edilip edilmeyeceğine ve sonuç uzaklığın standartlaştırılmak istenip istenmeyeceğine karar verilmelidir. Değişik uzaklık ve benzerlik ölçütleri veri özelliklerini farklı biçimde ağırlıklandırır.

Veri içindeki farklılıklar veya benzerlikler üzerinde temellenen ölçütler arasındaki seçim, özel bir uygulama için verideki hangi benzemezlik veya benzerliklerin önemli olduğuna dayandırılır. Örneğin, eğer hayvan kemikleri kümelenecek istenirse, kümelemeye geçmeden önce kemikler arasındaki hangi benzemezlik veya benzerliklerin önemli olduğuna karar verilmelidir. Burada kümeleme, kemik ölçüleri arasındaki gerçek farklılıklara veya boyutlar arasındaki ilişkilerine (benzerliklerine) bakılarak yapılabilir. Fakat, kemik ölçüleri arasındaki gerçek farklılıklara bakılarak yapılan kümelemede sorun; aynı soydan (aileden) gelen hayvanların kemik ölçüleri arasında bile gerçek farklılıkların bulunacağıdır. Ancak, bu farklılıklara rağmen boyutlar arasında aynı ilişkiyi taşıyan kemikler, gerçek farklılıklarına bakılmaksızın benzer olarak değerlendirilme eğilimindedir (Romesburg, 1984).

1.2. UZAKLIK VE BENZERLİKLER

1.2.1. Benzemezlik (Dissimilarity) - Uzaklık (Distance)

Benzemezlik ölçütleri iki nesne arasındaki uzaklık değerini vermektedir. İki nesne arasındaki büyük benzemezlik değeri bu nesnelerin birbirinden çok uzak olduğunu, küçük benzemezlik değeri ise bu nesnelerin birbirine çok yakın olduğunu gösterir.

r ve s nesneleri için uzaklık (benzemezlik) ölçütü p_{rs} olursa, p_{rs} 'nin özellikleri aşağıda verilmiştir.

1. $p_{rs} \geq 0$ tüm r ve s nesneleri için;
2. $p_{rs} = 0$ eğer sadece r ve s tanımlı ise;
3. $p_{rs} = p_{sr}$.

En çok kullanılan benzemezlik (uzaklık) ölçütü Öklit Uzaklık Ölçütü'dür. İki küme arasındaki uzaklığa bakılacaksa, Öklit Uzaklık Ölçütü'nün alternatifi Mahalanobis Uzaklık Ölçütü'dür. Benzemezlik ölçütlerine uzaklık-tipi ölçütler de denir. Benzemezlik (uzaklık) ölçütlerinin en küçük değeri 0, en büyük değeri ise sınırsızdır.

1.2.2. Benzerlik (Similarity) - Yakınlık (Proximity)

Benzerlik ölçütleri iki nesne arasındaki yakınlık değerini vermektedir. İki nesne arasındaki büyük benzerlik değeri bu nesnelerin birbirine çok yakın olduğunu, küçük benzerlik değeri ise bu nesnelerin birbirinden çok uzak olduğunu gösterir.

r ve s nesneleri için yakınlık (benzerlik) ölçütü p_{rs} olursa, p_{rs} 'nin özellikleri aşağıda verilmiştir.

1. $0 \leq p_{rs} \leq 1$ tüm r ve s nesneleri için;
2. $p_{rs} = 1$ eğer sadece r ve s tanımlı ise;
3. $p_{rs} = p_{sr}$.

En çok kullanılan benzerlik ölçütü Pearson İlişki Katsayısıdır. Bu katsayının dağılım aralığı $(-1,1)$ 'dir. Benzerlik (yakınlık) ölçütlerine Q-tipi veya korelasyon-tipi ölçütler de denir.

İki nesnenin seçilen değişkenler bakımından benzerliklerine bakılmak istendiğinde uzaklık ve benzerlik ölçütlerinden herhangi birisi kullanılabilir. Ancak, bu ölçütlerden elde edilen değeri yorumlarken şu kurala dikkat edilmelidir; "Benzerlik ölçütleri kullanılarak elde edilen değer küçük ise iki nesne (seçilen değişkenler bakımından) birbirinden uzak (benzemez), büyük ise birbirine yakındır (benzerdir). Uzaklık (benzemezlik) ölçütleri kullanılarak elde edilen değer küçük ise iki nesne (seçilen değişkenler bakımından) birbirine yakın (benzer), büyük ise birbirinden uzaktır (benzemezdir)." şeklinde yorumlanır.

II- YAYGIN OLARAK KULLANILAN UZAKLIK VE BENZERLİK ÖLÇÜTLERİ

2.1. Uzaklık ve Benzerlik Ölçütlerine İlişkin Genel Bilgiler

Genetik uygulamalar dışında en fazla kullanılan uzaklık ölçütü Kare Öklit Uzaklık Ölçütü'dür. Bazen Kare Öklit Uzaklığı'nın kare kökü, Öklit Uzaklığı olarak da kullanılır. Farklılıkların mutlak değerleri üzerine temellenen uzaklık ölçütü Manhattan City-Block Uzaklığı'dır (bu uzaklığa sadece City-Block Uzaklığı ya da Manhattan Uzaklığı da denmektedir). İki durum olduğunda, bütün değişkenler için değerlerin mutlak farklarının toplamıdır. Farklar kare olmadığı için geniş farklılıklar Kare Öklit Uzaklığı'ndaki gibi ağırlıklandırılmaz. Chebychev Uzaklığı, bütün değişkenlerin farklarının mutlak değerlerinin maksimumu olarak tanımlanmaktadır. Böylece bu uzaklık var olan bilginin çoğunu dikkate almaz. Veri iki durumlu ise özel uzaklık ve benzerlik ölçütleri gereklidir. Bunların çoğu olağanlık tabloları için birlikteliğin bilinen ölçülerine dayanır.

X_i ve X_j gözlem vektörleri arasındaki $d(X_i, X_j) = d_{ij}$ uzaklık değerlerini ifade etmek amacıyla geliştirilmiş pek çok teknik (metrik) bulunmaktadır.

Verinin ölçüm biçimi, benzerlik ve benzemezlik matrislerinin hesaplanmasını belirler. Çoğunlukla kullanılan dört türlü ölçüm biçimi vardır. Bunlar; Aralık (Interval), Nitel (Qualitative), Frekans Sayıları (Frequency Counts) ve İkili (Binary) ölçüdür. Verinin özelliğine göre aşağıdaki seçeneklerden herhangi biri seçilebilir.

2.2. ARALIK (INTERVAL), NİCEL-SÜREKLİ (QUANTITATIVE) VERİLER VE BUNLARA İLİŞKİN UZAKLIK-BENZERLİK ÖLÇÜTLERİ

Boy, ağırlık, baş çevresi vb. nicelik (aralık) değişkenlerle çalıştığımızda, benzemezlik ve benzerlik ölçütü hesaplamalarında aşağıdaki seçeneklerden herhangi biri kullanılabilir;

Uzaklık Ölçütleri:

2.2.1. Ölçekli Öklit Uzaklık Ölçütü (Scale Euclidean Distance Measure)

Değişkenlerin ölçümlerinde, değişik ölçüm birimlerinin kullanılması durumunda (m, kg, kalori v.b.) kullanılır ve formülü aşağıdaki gibidir.

$$d_2(x_i, x_j) = \sqrt{\sum_k^p w_k (x_{ik} - x_{jk})^2} \quad (2.2.1.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir. Formülde, w_k k'inci değişkenin standart sapma değerinin (S_k) ya da dağılım aralığı değerinin (R_k) tersidir. w_k 'nın S_k değerinin tersi olması durumunda elde edilen uzaklığa Karl Pearson uzaklığı da denmektedir. w_k 'nın bazı durumlar için değeri; standart olmadığı zaman $w_k = 1$, standart sapma ile standartlaştırıldığında $w_k = 1/S_k^2$, dağılım aralığı ile standartlaştırıldığında $w_k = 1/R_k^2$ 'dir.

2.2.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)

Uzaklık ölçütleri içinde en çok kullanılanlardan biri Öklit Uzaklık Ölçütü'dür. Ölçekli Öklit Uzaklığı'ndaki $w_k = 1$ olduğunda Öklit Uzaklık Ölçütü hesaplanır. Çok boyutlu uzayda iki nokta (i ve j) arasındaki uzaklık Öklit Uzaklık formülüyle hesaplanabilir. Bu iki nokta çakıştığı anda Öklit Uzaklığı kullanılamaz. Böylece her boyutta i ve j noktaları arasındaki uzaklıkların karesi alınır. Bu kareler boyutlar itibari ile toplanır ve bu toplamın karekökü alınır. Formülü aşağıdaki gibidir;

$$d_2(x_i, x_j) = \sqrt{\sum_k^p (x_{ik} - x_{jk})^2} \quad (2.2.2.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını ve aynı zamanda boyut sayısını ($k = 1, 2, \dots, p$) göstermektedir.

2.2.3. Kare Öklit Uzaklık Ölçütü (Squared Euclidean Distance Measure)

Öklit Uzaklık Ölçütü'nün karesi olup, formülü aşağıdaki gibidir;

$$d_2(x_i, x_j) = \sum_k^p (x_{ik} - x_{jk})^2 \quad (2.2.3.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.2.4. Chebychev Uzaklık Ölçütü (Chebychev Distance Measure)

Chebychev Uzaklık Ölçütü, farkların mutlak değerlerinin maksimumu olarak tanımlanır ve formülü aşağıdaki gibidir;

$$d_{\infty}(x_i, x_j) = \max_k^p |x_{ik} - x_{jk}| \quad (2.2.4.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.2.5. Manhattan City-Block Uzaklık Ölçütü (Manhattan City-Block Distance Measure)

Bu metrik (ölçüt), n boyutlu iki nokta arasındaki farkı ölçer. Farkların mutlak değerlerinin toplamı olarak tanımlanır ve formülü aşağıdaki gibidir;

$$d_1(x_i, x_j) = \sum_k^p |x_{ik} - x_{jk}| \quad (2.2.5.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.2.6. Minkowski Uzaklık Ölçütü ve Özel Durumlar (Minkowski Distance Measure)

i'ninci değişken üzerindeki j'ninci veri birimi tarafından gerçekleştirilen skor X_{ij} ve j'ninci veri birimi için skor vektörü $X_i = (X_{i1}, \dots, X_{ip})$ olsun. O zaman j ve k veri birimleri arasındaki Minkowski Metriği (Ölçütü) " $\lambda > 1$ " olduğu yerlerde,

$$d_{\lambda}(x_i, x_j) = \left\{ \sum_k^p |x_{ik} - x_{jk}|^{\lambda} \right\}^{1/\lambda} \quad (2.2.6.1)$$

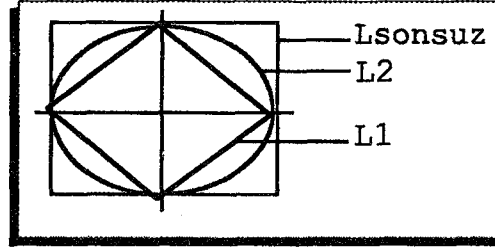
biçimindedir. Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir.

Birçok farklı metrik fonksiyonu, λ 'nın çeşitli değerleri seçilerek elde edilebilir.

" $\lambda = 1$ " için Manhattan City-Block Uzaklığı veya 1. Metriği (Ölçütü) olur. Bu metrik (ölçüt) n boyutlu iki nokta arasındaki farkı ölçer. Bilinen Öklit Uzaklığı " $\lambda = 2$ " metriği (ölçütü) alınarak elde edilir. Chebychev Metriği (Ölçütü) ise λ 'nın sınırsız artışına bağlı olarak $d_{\lambda}(x_i, x_k)$ 'nin limiti olarak elde edilir ve bazen 1-sonsuz metriği olarak adlandırılır.

$d_\lambda(x_i, x_j) = 1$ olduğunda noktanın oluşturduğu küme, birim küre veya metriğin birim küresi olarak adlandırılır. İki boyutlu bakıldığında l_1 , l_2 ve l_∞ metriği (ölçütü) aşağıdaki şekilde gösterilmiştir.

Şekil 2.2.6.1. Genel Bir Metrik (Ölçüt) İçin Birim Küre



$1 < \lambda < 2$ l_λ metriği için birim küre l_1 ve l_2 metrikleri arasında kalan dışbükey eğridir. $2 < \lambda < \infty$ olduğu durumda, verilen bir λ değeri için birim küre, l_2 ve l_∞ metrikleri arasında kalan dışbükey eğridir. Birim küre bir noktadan diğer bir noktaya birim (sabit) uzaklıkta olan tüm noktaların birlikte oluşturduğu yüzeydir.

Bütün olası metriklerden en çok dikkat edilmesi gereken Öklit veya l_2 Metriği'dir.

2.2.7. Mahalanobis Uzaklık Ölçütü (Mahalanobis Distance Measure)

Mahalanobis Uzaklığı'nda x_i ve x_j noktaları arasındaki uzaklıkların kareleri alınır. Mahalanobis Uzaklığına genelleştirilmiş Öklit Uzaklığı da denmektedir. İki küme arasındaki uzaklığı hesaplar ve formülü,

$$d_2(x_i, x_j) = \sqrt{\sum_k^p (x_{ik} - x_{jk})' S^{-1} (x_{ik} - x_{jk})} \quad (2.2.7.1)$$

veya

$$d_2(x_i, x_j) = \sqrt{\sum_k^p (\mu_{ik} - \mu_{jk})' S^{-1} (\mu_{ik} - \mu_{jk})} \quad (2.2.7.2)$$

ile tanımlanır. (2.2.7.1)'de x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini, S^{-1} ; toplanmış varyans-kovaryans matrisinin tersini ve p ise değişken sayısını göstermektedir. (2.2.7.2)'de μ_{ik} ; i'inci kümenin k'inci değişken ortalamasını, μ_{jk} ; j'inci kümenin k'inci değişken ortalamasını, S^{-1} ; toplanmış varyans-kovaryans matrisinin tersini ve p ise değişken sayısını göstermektedir.

Mahalanobis Uzaklık Ölçütü dikkönüşümlerden etkilenmez. Kümeler arasındaki uzaklık fonksiyonları ve bunların sürekli, çok değişkenli normal ve eşit kovaryans matrisli, çok değişkenli normal dağılımlardaki metrik özellikleri Duran ve Odell'in çalışmasında verilmiştir (Duran & Odell, 1974).

Sonuç olarak, çok değişkenli normal populasyonlarla ilgilendiğimizde, ortalama değerlerdeki farklılıkları ölçmek için Mahalanobis uygun bir uzaklık ölçütüdür. Ama, daha çok ölçümler faktör yapısına sahipse kullanılması önerilir. Mahalanobis Uzaklık Ölçütü'nün aşağıdaki gibi ilgi çekici özellikleri vardır;

a) Lineer dönüşümlerde değişmez,

b) Rao (1954)'nın gösterdiği gibi, faktör yapısı olduğu kabul edildiğinde, ölçümlerin toplanması ya da seçilmesi durumunda sabittir (Ryzin, 1973).

Mahalanobis Uzaklık Ölçütü'nde kovaryans matrisi kullanıldığı için değişkenler arasındaki korelasyonlar dikkate alınmaktadır. Mahalanobis Uzaklık Ölçütü değişkenler arasında ilişki varsa kullanılır. Değişkenler arasında ilişki yoksa bunun yerine standart değişkenlerden hesaplanan Öklit Uzaklık Ölçütü kullanılır (Everitt, 1974).

2.2.8. Hotelling T^2 Uzaklık Ölçütü (*Hotelling T^2 Distance Measure*)

İki grup ya da kümenin ortalama vektörlerinin karşılaştırılmasında kullanılan Hotelling T^2 değeri de yine bir uzaklık ölçütüdür. Formülü aşağıdaki gibidir;

$$T^2 = \left(\frac{n_1 n_2}{n}\right) \sum_k^p (\mu_{1k} - \mu_{2k})' S^{-1} (\mu_{1k} - \mu_{2k}) \quad (2.2.8.1)$$

Burada, μ_{1k} ; i'inci kümenin (grupun) k'inci değişken ortalamasını, μ_{2k} ; j'inci kümenin (grupun) k'inci değişken ortalamasını, n_1 ; 1. kümedeki (grup) nesne sayısını, n_2 ; 2. kümedeki nesne sayısını, n ; 1. ve 2. kümedeki toplam nesne sayısını, S^{-1} ; toplanmış varyans-kovaryans matrisinin tersini ve p ise değişken sayısını göstermektedir.

2.2.9. Canberra Uzaklık Ölçütü (*Canberra Distance Measure*)

Farkların mutlak değerlerinin toplamlara bölünmesiyle elde edilen değerlerin toplamıdır ve formülü aşağıdaki gibidir;

$$d(x_i, x_j) = \sum_k^p \frac{|x_{ik} - x_{jk}|}{(x_{ik} + x_{jk})} \quad (2.2.9.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.2.10. Karl Pearson Uzaklık Ölçütü (Karl Pearson Distance Measure)

Ölçekli Öklit Uzaklığı'nda, (2.2.1.1) formülündeki $w_k = 1/S_k^2$ olduğunda, Karl Pearson Uzaklık Ölçütü aşağıdaki formülle hesaplanır;

$$d_2(x_i, x_j) = \sqrt{\sum_k^p \frac{1}{S_k^2} (x_{ik} - x_{jk})^2} \quad (2.2.10.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini, S_k^2 ; k'inci değişkenin varyansını ve p ise değişken sayısını göstermektedir.

Gruplar arası uzaklıklarda ise Karl Pearson Benzemezlik Katsayısı aşağıdaki formülle hesaplanır;

$$d_2(\mu_r, \mu_s) = \sqrt{\frac{1}{p} \sum_k^p \frac{(\mu_{rk} - \mu_{sk})^2}{\left(\frac{S_{rk}^2}{n_r}\right) + \left(\frac{S_{sk}^2}{n_s}\right)}} \quad (2.2.10.2)$$

Burada, μ_{rk} ; r'inci nesnenin k'inci değişken ortalamasını, μ_{sk} ; s'inci nesnenin k'inci değişken ortalamasını, S_{rk}^2 ; r'inci nesnenin k'inci değişken varyansını, S_{sk}^2 ; s'inci nesnenin k'inci değişken varyansını, n_r ; r'inci grubun (kümenin) nesne sayısını, n_s ; s'inci grubun nesne sayısını ve p ise değişken sayısını göstermektedir.

2.2.11. Penrose Uzaklık Ölçütü (Penrose Distance Measure)

Birden fazla ölçüsü olan çok değişkenli kümeler arasındaki uzaklığın hesaplanmasında kullanılır. Ancak, bu uzaklık ölçütünün hesaplanabilmesi için kümelerin ortalama, varyans ve ortak kovaryanslarının bilinmesi gerekmektedir.

Çok değişkenli kümelerdeki p tane değişken $x_1, x_2, \dots, x_p$ olsun. x_k değişkeninin i-inci kümedeki ortalaması μ_{ik} , ve varyansı tüm kümeler için V_k 'dir. Penrose (1953)'de i ve j kümeleri arasındaki uzaklığı aşağıdaki şekilde tanımlamıştır.

$$P_{ij} = \sum_k^p \frac{(\mu_{ik} - \mu_{jk})^2}{pV_k} \quad (2.2.11.1)$$

Burada, μ_{ik} ; i'inci kümenin k'inci değişken ortalamasını, μ_{jk} ; j'inci kümenin k'inci değişken ortalamasını, V_k ; k'inci değişkenin (tüm kümeler için) toplanmış varyansını ve p ise değişken sayısını göstermektedir.

Penrose Uzaklık Ölçütü'nde değişkenler arasındaki korelasyonlar dikkate alınmamaktadır. Mahalanobis Uzaklık Ölçütü kovaryansları kullandığı için Penrose Uzaklık Ölçütü'nden daha iyi sonuç vermektedir. Ancak, kovaryanslar bilinmiyorsa Penrose, Mahalanobis'in yerine kullanılabilir (Manly, 1990).

Benzerlik Ölçütleri:

2.2.12. Kosinüs Benzerlik Ölçütü (Cosine Similarity Measure)

Bu bir benzerlik ölçütüdür, vektör değerleri için formülü aşağıdaki gibidir;

$$\text{Benzerlik}(x_i, x_j) = \frac{\sum_{i,j}^p (x_i x_j)}{\sqrt{(\sum_i^p x_i^2)(\sum_j^p x_j^2)}} \quad (2.2.12.1)$$

Burada, x_i ; i'inci nesnenin x değişken değerini, x_j ; j'inci nesnenin x değişken değerini ve p ise değişken sayısını göstermektedir.

2.2.13. Pearson İlişki Katsayısı (Pearson Correlation Coefficient)

Bu metrik (ölçüt), 1-p'yi kullanır. p_{ij} ; i ve j objeleri arasındaki Pearson çarpım-momentinin korelasyonudur. Sürekli veya aralık ölçeklerde kullanılır ve bir benzerlik ölçütüdür. Vektörler arası korelasyon için formülü aşağıdaki gibidir;

$$\text{Benzerlik}(x_i, x_j) = \frac{\sum_{i,j}^p (Zx_i Zx_j)}{N-1} \quad (2.2.13.1)$$

Burada, Zx_i ; i'inci nesnenin x değişkeninin z-skoru değerini, Zx_j ; j'inci nesnenin x değişkeninin z-skoru değerini, p; değişken sayısını ve N ise toplam nesne sayısını göstermektedir.

2.3. NİTEL-KESİKLİ (QUALITATIVE) VERİLER VE BUNLARA İLİŞKİN UZAKLIK ÖLÇÜTLERİ

Pratikte değişkenlerin tümünün sürekli (aralık veya oransal ölçekli) olması çoğu kez mümkün değildir. Bazı değişkenlerin kesikli (ordinal ölçekli) olması durumlarıyla karşılaşılabilir. Böyle durumlarda yukarıda verilen formüller doğrudan kullanılamaz onların yerine uzaklıklar aşağıdaki gibi belirlenir.

$$w_k = \begin{cases} 1 & ; \text{ nicel veriler için,} \\ 1 & ; \text{ nitel veriler için,} \end{cases} \frac{1}{R_k}$$

k'inci değişkenin dağılım aralığı (R_k)

Nitel veriler için uzaklık ölçütleri aşağıdaki formüllerden yararlanılarak bulunur.

Uzaklık Ölçütleri:

2.3.1. Ölçekli Öklit Uzaklık Ölçütü (Scale Euclidean Distance Measure)

Nitel veriler için Ölçekli Öklit Uzaklığı aşağıdaki formülle hesaplanmaktadır.

$$d_2(x_i, x_j) = \sqrt{\sum_k^p \frac{1}{R_k^2} (x_{ik} - x_{jk})^2} \quad (2.3.1.1)$$

Burada, x_{ik} ; i'inci nesnenin k'inci değişken değerini, x_{jk} ; j'inci nesnenin k'inci değişken değerini, R_k ; k'inci değişkenin dağılım aralığını ve p ise değişken sayısını göstermektedir.

Nicel veriler için geliştirilmiş formüllerin tümüne w_k değeri de eklenerek uzaklıklar yaklaşık olarak hesaplanacağı gibi Bhattacharyya Uzaklığı başta olmak üzere özel amaçlı bazı metriklerden de yararlanılabilmektedir (Anderberg,1973), (Everitt, 1974).

2.3.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)

P kategorisine göre sıralanmış veriler düşünelim. n boyutta herbiri için; $r = 1, 2, \dots, g$; $(x_{r1}, \dots, x_{rp}) = x_r$ olduğunda her kategori n büyüklüğündedir. Örneğin, x_r insanların kan grubunu A, AB, B ve 0 gibi, her g ülkesinde gösterdiğinde formül aşağıdaki gibi olur.

$$\sum_i^p x_{ri} = 1, (r = 1, \dots, g). \quad (2.3.2.1)$$

Bu toplumların uzaklık ölçütleri için çeşitli alternatiflerimiz vardır. x_r noktaları hiperdüzlemde $x_1 + x_2 + \dots + x_p = 1$ ve R^p pozitif görünümde olduğunda Öklit Uzaklık formülü aşağıdaki gibi olur.

$$d_{rs} = \sqrt{\sum_i^p (x_{ri} - x_{si})^2} \quad (2.3.2.2)$$

Burada, x_{ri} , r 'inci nesnenin i 'inci değişken değerini, x_{si} , s 'inci nesnenin i 'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.3.3. Mahalanobis-Benzeri Uzaklık Ölçütü (Mahalanobis-Like Distance Measure)

Stokastik varyasyon modeli kabul edilmediğinde, nitel verilerin uzaklık ölçüm oranları Mahalanobis-Benzeri Uzaklık Ölçütü ile verilebilir.

$r = 1, 2, \dots, g$ x_r , n_r^p oranlarını gösterdiğinde a parametre dağılımıda $a = (a_1, \dots, a_p)'$ (bu her r için geçerlidir) olur. Bundan dolayı x_r ortalama anlamında kullanılır ve kovaryans matriside, $\Sigma_r = n_r^{-1} \Sigma$, $\Sigma = (\sigma_{ij})$ 'dir. Buna göre de;

$$\sigma_{ij} = \begin{cases} a_i(1-a_i), & i = j, \\ -a_i a_j, & i \neq j; \end{cases}$$

bundan dolayı $\Sigma = \text{diag}(a) - aa'$ olur.

x_r hiperdüzlemde olduğundan Σ tektir. Bunun doğruluğunu görmek için a , g -ters Σ -ters için kullanılır.

$$\Sigma^{-} = \text{diag}(a_1^{-1}, \dots, a_p^{-1}) \quad (2.3.3.1)$$

x_r ve x_s noktaları arasındaki Mahalanobis Uzaklığı'nı, $n_r n_s / (n_r + n_s)$ olursa aşağıdaki şekilde açıklayabiliriz (Mardia, Kent & Bibby, 1979).

$$\sum_i^p \frac{(x_{ri} - x_{si})^2}{a_i} \quad (2.3.3.2)$$

Yukarıdaki formülden Mahalanobis-Benzeri Uzaklık Ölçütü'nü aşağıdaki şekilde tanımlarız.

$$D_{1;rs} = \sqrt{2 \sum_i^p \frac{(x_{ri} - x_{si})^2}{|x_{ri} + x_{si}|}} \quad (2.3.3.3)$$

Burada, x_{ri} ; r'inci nesnenin i'inci değişken değerini, x_{si} ; s'inci nesnenin i'inci değişken değerini ve p ise değişken sayısını göstermektedir.

2.3.4. Bhattacharyya Uzaklık Ölçütü (Bhattacharyya Distance Measure)

v_r vektörleri K^p uzayında birim kürenin merkezi üzerinde yer alan noktalardan oluşsun;

$$v_k = (x_{r1}^{1/2}, \dots, x_{rp}^{1/2})', (r = 1, \dots, g). \quad (2.3.4.1)$$

v_g ve v_s noktaları arasındaki en büyük açı, B_{rs} açısıdır.

$$\cos B_{rs} = \sum_i^p v_{ri} v_{si} = \sum_i^p \sqrt{x_{ri} x_{si}} \quad (2.3.4.2)$$

Buradan v_r ve v_s arasındaki uzaklık aşağıdaki formülle hesaplanır.

$$D_{2;rs} = \sqrt{\sum_i^p (x_{ri}^{1/2} - x_{si}^{1/2})^2} \quad (2.3.4.3)$$

Burada, x_{ri} ; r'inci nesnenin i'inci değişken değerini, x_{si} ; s'inci nesnenin i'inci değişken değerini ve p ise değişken sayısını göstermektedir.

$D_{2;rs}$ Bhattacharyya Uzaklığı olarak tanımlanmaktadır (bazı zamanlar B_{rs} olarak da adlandırılır). İki ölçüt arasındaki bağlantı aşağıdaki gibidir;

$$D_{2;rs}^2 = 4 \sin^2 \frac{1}{2} B_{rs} \quad (2.3.4.4)$$

2.4. FREKANS SAYILARI (FREQUENCY COUNTS) VE BUNLARA İLİŞKİN UZAKLIK ÖLÇÜTLERİ

Bu tip (nitel) veriler için aşağıdaki seçeneklerden herhangi biri kullanılarak uzaklık (benzemezlilik) ölçütleri hesaplanmaktadır.

Uzaklık Ölçütleri:

2.4.1. Ki-Kare Uzaklık Ölçütü (Chi-Square Distance Measure)

Bu aynı zamanda bir sayaçtır. Temelde ki-kare testinden kaynaklanmaktadır. İki olayın frekans sayılarıyla ilgilenmektedir. Bu ölçütün değeri, uzaklıkları hesaplanabilen değerlerin toplam frekanslarına bağlıdır. x_i ve x_j değişkenleri veya durumları için uzaklıklar aşağıdaki formülle hesaplanmaktadır;

$$\text{Uzaklık}(x_i, x_j) = \sqrt{\sum_i^p \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum_j^p \frac{(x_j - E(x_j))^2}{E(x_j)}} \quad (2.4.1.1)$$

2.4.2. Fi-Kare Uzaklık Ölçütü (Phi-Square Distance Measure)

Ki-kare uzaklık ölçütünün normalize edilmiş şeklidir. Normalleştirme, uzaklıkları hesaplanabilen x_i ve x_j değerlerinin frekanslarına bağlı değildir. x_i ve x_j değişkenleri veya durumları için uzaklıklar aşağıdaki formülle hesaplanmaktadır;

$$\text{Uzaklık}(x_i, x_j) = \sqrt{\frac{\sum_i^p \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum_j^p \frac{(x_j - E(x_j))^2}{E(x_j)}}{N}} \quad (2.4.2.1)$$

Yukarıdaki (2.4.1.1) ve (2.4.2.1) formüllerinde, x_i ; i'inci nesne x değişkeninin gözlem değerini, $E(x_i)$; i'inci nesne x değişkeninin beklenen değerini, x_j ; j'inci nesne x değişkeninin gözlem değerini, $E(x_j)$; j'inci nesne x değişkeninin beklenen değerini, p; değişken sayısını ve N ise i ve j nesnelerinin değişken değerlerinin toplamını göstermektedir.

2.5. İKİLİ VERİLER (BINARY DATA) VE BUNLARA İLİŞKİN UZAKLIK-BENZERLİK ÖLÇÜTLERİ

Bu tip veriler için pek çok uzaklık ve benzerlik ölçütleri hesaplanmaktadır. Birey veya değişken çiftleri olduğunda özel uzaklık ve benzerlik ölçütlerine gereksinim vardır. Durum veya değişken çiftleri için kullanılan bu ölçütlerde, önce 2x2 boyutunda olağanlık tablosu oluşturulup, bu tablo üzerinden uzaklık ve benzerlik ölçütleri hesaplanmaktadır. Bu değerler, değişkenler veya durumlar arasındaki var veya yok çetelemesinden elde

edilmektedir. İkili ölçümde var için 1, yok için ise 0 değeri verilir (ayrıca 0 ve 1 yerine farklı tam sayı değerlerde kullanılabilir). İki durum için aşağıdaki olağanlık tablosu oluşturulur;

Tablo 2.5.1. Olağanlık Tablosu

		Durum 1		
		Var	Yok	
Durum 2	Var	a	b	(a + b)
	Yok	c	d	(c + d)
		(a + c)	(b + d)	p

ve bu tablodan, aşağıdaki uzaklık ve benzerlik ölçütlerinden herhangi biri ile uzaklık ya da benzerlik değerleri hesaplanabilir.

Aşağıda verilen formüllerin tümünde geçerli olmak üzere; a; birlikte var olarak uyuşan çiftlerin sayısını, b ve c; uyuşmayan çiftlerin sayısını, d; birlikte yok olarak uyuşan çiftlerin sayısını ve p ise değişken sayısını göstermektedir. Burada, $p = a + b + c + d$ 'dir.

Uzaklık Ölçütleri:

2.5.1. Kare Öklit Uzaklık Ölçütü (Squared Euclidean Distance Measure)

İkili veride Kare Uzaklık Ölçütü'nün minimum değeri 0, üst değeri ise sınırsızdır ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık (X, Y)} = b + c \quad (2.5.1.1)$$

2.5.2. Öklit Uzaklık Ölçütü (Euclidean Distance Measure)

İkili veride Öklit Uzaklık Ölçütü'nün minimum değeri 0, üst değeri ise sınırsızdır ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık(X, Y)} = \sqrt{b + c} \quad (2.5.2.1)$$

Elemanlar anlamlı p boyutlu ölçeklerle temsil edilmedikleri zaman, eleman çiftleri çoğunlukla kesin karakteristiklerin varlığı veya yokluğu temelinde karşılaştırılırlar.

Benzer elemanlar, benzer olmayan elemanlardan daha fazla ortak karakteristiğe sahiptir. Karakteristik olduğu zaman 1, olmadığı zaman 0 değerini alan ikili kodlar üretilir. Karakteristikliğin varlığı veya yokluğu matematiksel olarak açıklanabilir. Örneğin, $p = 6$ için ikili değerleri i ve k olarak alalım. Bu elemanlar için skor (score) değişkeni aşağıdaki gibi düzenlenebilir.

Tablo 2.5.2.1. İki Eleman İçin İkili Değişkenler

	Değişken						
	1	2	3	4	5	6	(j)
Eleman i	1	0	0	1	1	0	
Eleman k	1	1	0	1	0	0	

Bu durumda iki adet 0-0, iki adet 1-1 uyuşması ve iki adet uyuşmazlık vardır. x_{ij} 'yi i 'ninci eleman üzerinde j 'ninci ikili değişken (1 veya 0) olarak, x_{kj} 'yi k 'inci eleman üzerindeki j 'ninci değişken değeri olarak kabul edelim. Sonuç olarak,

$$\Sigma(x_{ik} - x_{kj})^2 = \begin{cases} 0 ; & x_{ik} = x_{kj} = 1 \text{ veya } x_{ik} = x_{kj} = 0 \\ 1 ; & x_{ik} \neq x_{kj} \end{cases}$$

elde edilir. Öklit uzaklık farkının karesi; $\Sigma(x_{ik} - x_{kj})^2$, ($k = 1, 2, \dots, p$)'dir. Bu da uyuşmazlıkların sayısını verir. Buradan elde edilecek büyük fark daha fazla uyuşmazlığa karşılık gelir. Yani benzer olmayan elemanların sayısının daha fazla olduğunu gösterir. Yukarıdaki şeklin i ve k elemanları arasındaki kare farkı,

$$\begin{aligned} \Sigma(x_{ik} - x_{kj})^2 &= (1 - 1)^2 + (0 - 1)^2 + (0 - 0)^2 + (1 - 1)^2 + (1 - 0)^2 + (0 - 0)^2 \\ &= 1 + 1 = 2 \text{ olur.} \end{aligned}$$

Bu örnek için 2x2 boyutlu olağanlık tablosu aşağıda verilmiştir.

Tablo 2.5.2.2. İki Eleman İçin Olağanlık Tablosu

		Eleman i		
		1	0	
Eleman k	1	2	1	3
	0	1	2	3
		3	3	6

Yukarıdaki tablodan Öklit Uzaklık değeri,

$$\text{Uzaklık}(X, Y) = \sqrt{b+c} = \sqrt{1+1} = \sqrt{2} = \underline{1.414} \text{ olarak bulunur.}$$

Sonuç olarak Öklit Uzaklık Ölçütü, değişkenler uzayının ortogonal olduğunu veya başka bir deyişle, değişkenlerin birbirinden bağımsız olduğunu varsayar. Teorik olarak Öklit Uzaklık Ölçütü'nün uygulanabilmesi için verilerin en azından aralık ölçekle ölçülmüş olması gerekir (Frank & Green, 1968).

2.5.3. Büyüklük Farkları Uzaklık Ölçütü (Size Difference Distance Measure)

Minimum değeri 0, üst değeri sınırsızdır ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık}(X, Y) = \frac{(b-c)^2}{(a+b+c+d)^2} \quad (2.5.3.1)$$

2.5.4. Biçim Farkları Uzaklık Ölçütü (Pattern Difference Distance Measure)

Dağılım aralığı 0'dan 1'e kadardır ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık}(X, Y) = \frac{bc}{(a+b+c+d)^2} \quad (2.5.4.1)$$

2.5.5. Değişim Uzaklık Ölçütü (Variance Distance Measure)

Minimum değeri 0, üst değeri sınırsızdır ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık}(X, Y) = \frac{b+c}{4(a+b+c+d)} \quad (2.5.5.1)$$

2.5.6. Durum Uzaklık Ölçütü (Shape Distance Measure)

Minimum değeri 0, üst değerinde sınır yok ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık}(X, Y) = \frac{(a+b+c+d)(b+c)-(b-c)^2}{(a+b+c+d)^2} \quad (2.5.6.1)$$

2.5.7. Lance ve Williams Uzaklık Ölçütü (Lance & Williams Distance Measure)

İkili nonmetrik benzemezlik ölçütüdür. Bu ölçütün sınırları 0 ile 1'dir ve aşağıdaki formülle hesaplanır;

$$\text{Uzaklık}(X, Y) = \frac{b+c}{(2a+b+c)} \quad (2.5.7.1)$$

Benzerlik Ölçütleri:

2.5.8. Russell ve Rao Benzerlik Ölçütü (Russell & Rao Similarity Measure)

Bu benzerlik katsayısı sadece var olarak uyuşan çiftlerin toplam içerisindeki payını vermektedir. Burada yok olarak uyuşanlar dikkate alınmamaktadır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik}(X, Y) = \frac{a}{a+b+c+d} \quad (2.5.8.1)$$

2.5.9. Basit Benzerlik Ölçütü (Simple Similarity Measure)

Bu ölçüt, toplam özellikler içerisinde birlikte var ve yok olarak uyuşan çiftlerin oranıdır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik}(X, Y) = \frac{a+d}{a+b+c+d} \quad (2.5.9.1)$$

2.5.10. Jaccard Benzerlik Ölçütü (Jaccard Similarity Measure)

Bu ölçüt benzerlik oranı olarakta bilinir. Birlikte yok olarak uyuşan çiftler pay ve paydada dikkate alınmaz. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a}{a+b+c} \quad (2.5.10.1)$$

2.5.11. Parçalı Benzerlik Ölçütü (Dice Similarity Measure)

Bu ölçüt Czekanowski veya Sorensen ölçütü olarakta bilinir. Birlikte var olarak uyuşanlara iki kat ağırlık verilir. Burada birlikte yok olarak uyuşanların sayısı dikkate alınmamaktadır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{2a}{2a+b+c} \quad (2.5.11.1)$$

2.5.12. Rogers ve Tanimoto Benzerlik Ölçütü (Rogers & Tanimoto Similarity Measure)

Bu ölçütte birlikte var ve yok olarak uyuşanlar pay ve paydada dikkate alınmaktadır. Paydada uyuşmazlıkların iki katı alınmaktadır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a+d}{a+d+2(b+c)} \quad (2.5.12.1)$$

2.5.13. Sokal ve Sneath Benzerlik Ölçütü-1 (Sokal & Sneath Similarity Measure-1)

Bu benzerlik katsayıları ölçütünde, birlikte var ve yok olarak uyuşan çiftlerin pay ve paydada iki katı alınmaktadır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{2(a+d)}{2(a+d)+b+c} \quad (2.5.13.1)$$

2.5.14. Sokal ve Sneath Benzerlik Ölçütü-2 (Sokal & Sneath Similarity Measure-2)

Bu benzerlik katsayıları ölçütünde birlikte yok olan çiftler pay ve paydada dikkate alınmamaktadır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a}{a+2(b+c)} \quad (2.5.14.1)$$

2.5.15. Sokal ve Sneath Benzerlik Ölçütü-3 (Sokal & Sneath Similarity Measure-3)

Bu benzerlik ölçütünde tüm uyuşan çiftler paydaya dahil edilmemektedir. Böylece uyuşan ve uyuşmayan çiftler eşit ağırlığa sahiptirler. Bu ölçütün minimum değeri 0, üst sınırı yok ve uyuşmazlıklar olmadığına ($b=0$ ve $c=0$) olmadığına tanımsızdır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a+d}{b+c} \quad (2.5.15.1)$$

2.5.16. Kulczynski Benzerlik Ölçütü-1 (Kulczynski Similarity Measure-1)

Bu benzerlik ölçütünde tüm uyuşan çiftlerin sayısı paydada, yok olarak uyuşan çiftlerin sayısında pay ve paydada dikkate alınmamaktadır. Bu ölçütün minimum değeri 0, üst sınırı yok ve uyuşmazlıklar olduğunda ($b=0$ ve $c=0$) tanımsızdır. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a}{b+c} \quad (2.5.16.1)$$

2.5.17. Kulczynski Benzerlik Ölçütü-2 (Kulczynski Similarity Measure-2)

Bu ölçüt, Tablo 2.5.1'de belirtilen $[P(D_1 \text{ var} / D_2 \text{ var}), P(D_2 \text{ var} / D_1 \text{ var})]$ koşullu olasılığının ortalamasıdır. Bu ölçütün dağılım aralığı (0,1)'dir. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a/(a+b)+a/(a+c)}{2} \quad (2.5.17.1)$$

2.5.18. Sokal ve Sneath Benzerlik Ölçütü-4 (Sokal & Sneath Similarity Measure-4)

Bu ölçüt, Tablo 2.5.1'de belirtilen $[P(D_1 \text{ var} / D_2 \text{ var}), P(D_2 \text{ var} / D_1 \text{ var})]$ koşullu olasılığıdır. Dağılım aralığı (0,1)'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{a/(a+b)+a/(a+c)+d/(b+d)+d/(c+d)}{4} \quad (2.5.18.1)$$

2.5.19. Hamann Benzerlik Ölçütü (Hamann Similarity Measure)

Bu ölçütte bir karakteristik bakımından uyuşan çiftler (var veya yok olarak) pay ve paydada dikkate alınmaktadır. Böylece uyuşan çiftler eşit ağırlıklandırılmaktadır. Dağılım aralığı -1 ile 1 arasında değişmektedir. Ayrıca Sokal ve Sneath-1 ile Rogers ve Tanimoto ölçütlerinin birleşimidir. Formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{(a+d)-(b+c)}{a+b+c+d} \quad (2.5.19.1)$$

2.5.20. Goodman ve Kruskal Lamda Benzerlik Ölçütü (Goodman & Kruskal Lambda Similarity Measure)

Bu ölçütte önceden bilinen olasılıkla belirlenen durum karakterleri bir sorunda (var veya yok) olarak biliniyorsa diğer durum içinde aynı şekilde geçerlidir. Dağılım aralığı (0,1)'dir ve formülü aşağıdaki gibidir;

$$\begin{aligned} \text{Benzerlik (X, Y)} = & \frac{\max(a,b)+\max(c,d)+\max(a,c)+\max(b,d)}{2(a+b+c+d)} \\ & - \left(\frac{\max(a+c,b+d)+\max(a+b,c+d)}{2(a+b+c+d)} \right) \end{aligned} \quad (2.5.20.1)$$

2.5.21. Anderberg'in D Benzerlik Ölçütü (Anderberg's D Similarity Measure)

Bu ölçütte bir gözde bir karakteristik (var veya yok) olarak ortaya çıkıyorsa diğer gözde de aynı şekilde ortaya çıkmaktadır. Dağılım aralığı (0,1)'dir ve formülü aşağıdaki gibidir;

$$\begin{aligned} \text{Benzerlik (X, Y)} = & \frac{\max(a,b)+\max(c,d)+\max(a,c)+\max(b,d)}{2(a+b+c+d)} \\ & + \frac{\max(a+c,b+d)+\max(a+b,c+d)}{2(a+b+c+d)} \end{aligned} \quad (2.5.21.1)$$

2.5.22. Yule'in Y Benzerlik Ölçütü (Yule's Y Similarity Measure)

Yule'in Y katsayısı olarak bilinen bu ölçütte, bu katsayayı bulabilmek için Tablo 2.5.1'deki olağanlık tablosu oluşturulmaktadır. Dağılım aralığı $(-1,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik } (X, Y) = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \quad (2.5.22.1)$$

2.5.23. Yule'in Q Benzerlik Ölçütü (Yule's Q Similarity Measure)

Bu ölçüt, Tablo 2.5.1'deki olağanlık tablosuna dayanan Goodman ve Kruskal'ın sıralı ölçüm için gama katsayısı olarak tanımlanabilir. Dağılım aralığı $(-1,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik } (X, Y) = \frac{ad-bc}{ad+bc} \quad (2.5.23.1)$$

2.5.24. Ochiai Benzerlik Ölçütü (Ochiai Similarity Measure)

Kosinüsün ikili formudur. Dağılım aralığı $(0,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik } (X, Y) = \sqrt{\left(\frac{a}{a+b}\right) \left(\frac{a}{a+c}\right)} \quad (2.5.24.1)$$

2.5.25. Sokal ve Sneath Benzerlik Ölçütü-5 (Sokal & Sneath Similarity Measure-5)

Bu ölçütün dağılım aralığı $(0,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik } (X, Y) = \frac{ad}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (2.5.25.1)$$

2.5.26. Fi 4-Nokta Korelasyon Benzerlik Ölçütü (Phi 4-Point Correlation Similarity Measure)

Bu ölçüt, Pearson İlişki katsayısının ikili formudur. Dağılım aralığı $(0,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (2.5.26.1)$$

2.5.27. Yayılım Benzerlik Ölçütü (Dispersion Similarity Measure)

Bu ölçütün dağılım aralığı $(-1,1)$ 'dir ve formülü aşağıdaki gibidir;

$$\text{Benzerlik (X, Y)} = \frac{ad-bc}{(a+b+c+d)^2} \quad (2.5.27.1)$$

Yukarıda verdiğimiz uzaklık ve benzerlik ölçütlerinin ölçüm biçimlerine göre dağılımı Tablo 2.1'de özet olarak verilmiştir.

TABLO 2.1. UZAKLIK VE BENZERLİK ÖLÇÜTLERİNİN ÖLÇÜM BİÇİMLERİNE GÖRE DAĞILIMI

ARALIK VERİLER	
UZAKLIK	BENZERLİK
- Ölçekli Öklit Uzaklık Ölçütü	- Kosinüs Benzerlik Ölçütü
- Öklit Uzaklık Ölçütü	- Pearson İlişki Katsayısı
- Kare Öklit Uzaklık Ölçütü	
- Chebychev Uzaklık Ölçütü	
- Manhattan City-Block Uzaklık Ölçütü	
- Minkowski Uzaklık Ölçütü	
- Mahalanobis Uzaklık Ölçütü	
- Hotelling T^2 Uzaklık Ölçütü	
- Canberra Uzaklık Ölçütü	
- Karl Pearson Uzaklık Ölçütü	
- Penrose Uzaklık Ölçütü	

NİTEL VERİLER

UZAKLIK**BENZERLİK**

- Ölçekli Öklit Uzaklık Ölçütü
 - Öklit Uzaklık Ölçütü
 - Mahalanobis-Benzeri Uzaklık Ölçütü
 - Bhattacharyya Uzaklık Ölçütü
-

FREKANS SAYILARI

UZAKLIK**BENZERLİK**

- Ki-Kare Uzaklık Ölçütü
 - Fi-Kare Uzaklık Ölçütü
-

İKİLİ VERİLER

UZAKLIK**BENZERLİK**

- | | |
|------------------------------------|---------------------------------------|
| - Kare Öklit Uzaklık Ölçütü | - Russell ve Rao Benzerlik Ölçütü |
| - Öklit Uzaklık Ölçütü | - Basit Benzerlik Ölçütü |
| - Büyüklük Farkları Uzaklık Ölçütü | - Jaccard Benzerlik Ölçütü |
| - Biçim Farkları Uzaklık Ölçütü | - Parçalı Benzerlik Ölçütü |
| - Değişim Uzaklık Ölçütü | - Rogers ve Tanimoto Benzerlik Ölçütü |
| - Durum Uzaklık Ölçütü | - Sokal ve Sneath Benzerlik Ölçütü-1 |
| - Lance ve Williams Uzaklık Ölçütü | - Sokal ve Sneath Benzerlik Ölçütü-2 |
| | - Sokal ve Sneath Benzerlik Ölçütü-3 |
| | - Kulczynski Benzerlik Ölçütü-1 |
| | - Kulczynski Benzerlik Ölçütü-2 |
| | - Sokal ve Sneath Benzerlik Ölçütü-4 |
| | - Hamann Benzerlik Ölçütü |

- Goodman ve Kruskal Lamda Benzerlik Ölçütü
- Anderberg'in D Benzerlik Ölçütü
- Yule'in Y Benzerlik Ölçütü
- Yule'in Q Benzerlik Ölçütü
- Ochiai Benzerlik Ölçütü
- Sokal ve Sneath Benzerlik Ölçütü-5
- Fi 4-Nokta Korelasyon Benzerlik Ölçütü
- Yayılım Benzerlik Ölçütü

2.6. UZAKLIK VE BENZERLİK ÖLÇÜTLERİNİN ÖZELLİKLERİ

Yukarıda verilen $d(x_i, x_j)$ uzaklık değerlerinin bazı özellikleri taşımaları gerekir. Bunlar kısaca şöyle sıralanabilir;

- I- $d(x_i, x_j) = d(x_j, x_i)$ simetri özelliği,
- II- $d(x_i, x_j) \geq 0$ negatif olmama özelliği,
- III- $d(x_i, x_j) = 0$; $i = j$ iken tanım özelliği,
- IV- $d(x_i, x_j) \leq d(x_i, x_k) + d(x_k, x_j)$ üçgen eşitsizliği özelliği, bu özellik metrik özelliğini belirler.

Johnson (1967), beşinci bir özellik tanımlayarak kümeleme yöntemlerinden aşamasıralı (en yakın komşuluk) yapının bu özelliğe bağlı olduğunu belirtmiştir, bu da ultrametrik özelliğidir.

$$V- d(x_i, x_j) \leq \max \{d(x_i, x_k), d(x_k, x_j)\} \text{ ultrametrik özelliği.}$$

Bazı uygulamalar için I-III özellikleri uzaklık fonksiyonu için yeterlidir. Ancak, genel anlamda I-V özellikleri sağlandığında uzaklık fonksiyonu olarak kabul edilir. I. özellik her zaman doğru olmayabilir; sosyolojideki uygulamalarda simetrik özelliği her zaman sağlanmayabilir. Bazı örnekler için IV. özellik sağlanmadığında, IV. özellik yerine daha kuvvetli koşullar içeren V. ultrametrik eşitsizlikler, uzaklık fonksiyonunu sağlar.

P ve Q farklılaştıkça $d(P, Q)$ fonksiyonunda "benzemezlik" ya da "ayrılma" özelliklerini içerir. Bu nedenle III. ve IV. metrik özellikleri sağlanılmadığında $d(P, Q)$ benzemezlik katsayısı olarak tanımlanır.

Herhangi bir uzaklık ölçütüyle hesaplanan ve uzaklık değerlerinden oluşan D uzaklık matrisi $n \times n$ boyutlu olup, $n(n-1)/2$ tane uzaklık değerinden oluşmaktadır. Örneğin, 5 farklı cins hayvanın çene kemik yapılarından yararlanılarak genetik aileleri belirlenmek istensin. Burada $n = 5$, uzaklık matrisi 5×5 boyutlu ve $5(5-1)/2 = 10$ tane uzaklık değerinden oluşur.

2.7. VERİLERİN STANDARTLAŞTIRILMASI

Durum ya da değişken değerlerini standartlaştırmak için farklı derecelendirme (scale)'den yararlanılmaktadır.

İki durum arasındaki uzaklık veya benzerliği görebilmek için pek çok yöntem vardır. Ancak, hesaplamaya geçmeden önce değişkenlerin ölçüm biçiminin yeniden gözden geçirilmesine gereksinim olup olmadığına karar verilmesi gereklidir. Değişkenlerin değişik ölçüleri olduğunda örneğin, fiyatlar ve kaloriler gibi ve bunlar standart olmadığında, herhangi bir uzaklık ölçütü öncelikle, geniş birimler içinde ölçülen değişkenlerden ileri gelen bir hatayı yansıtacaktır.

Herhangi bir veri grubu için uzaklık veya benzerlik ölçütünü hesaplamaya geçmeden önce, değişkenlere ilişkin değerler ilgili değişkenin standart sapmasına, dağılım aralığına, ortalamasına veya maksimum değerine bölünerek standartlaştırılabilir (bkz. Romesburg, 1984 veya Anderberg, 1973). Bunların dışında en yaygın kullanılan standartlaştırma yöntemi; ortalaması 0 ve standart sapması 1 olan z-skora dönüştürülerek yapılır. Dönüştürülmüş veri temelinde, çok çeşitli uzaklık ve benzerlik ölçütlerini hesaplamak mümkündür.

Z-skora dönüştürerek standartlaştırma, her bir değer ortalamadan çıkartılıp, çıkan değerın standart sapmaya bölünmesiyle yapılmaktadır. Formülü aşağıdaki gibidir;

$$Z = \frac{x_i - \mu}{\sigma} \quad (2.7.1)$$

Burada, x_i ; i'inci değişkenin değerini, μ ; x değerlerinin ortalamasını ve σ ise x değerlerinin standart sapmasını göstermektedir.

Dağılım aralığı (-1, 1) için standartlaştırma, her bir değer elde edilen dağılım aralığına bölünerek yapılmaktadır. Aralık sıfırsa bütün değerler sıfırdır. Örneğin, beş ülkenin, 5 yaşından küçüklerde ölüm oranı ile doğumda yaşam beklentisi verileri Tablo 2.7.1'de verilmiştir.

Tablo 2.7.1. Beş Ülkenin 5 Yaşından Küçüklerde Ölüm Oranı ile Doğumda Yaşam Beklentisi (1993)

Ülkeler	5 Yaşından Küçüklerde Ölüm Oranı (binde)	Doğumda Yaşam Beklentisi (yıl)
Moritanya	202	48
Bhutan	197	49
Nijerya	191	53
Zaire	187	52
Uganda	185	42
Dağılım Aralığı	$(202-185) = 17$	$(53-42) = 11$

Her bir değeri dağılım aralığına bölerek aşağıdaki standart değerleri elde ederiz.

$$202/17 = 11.88, 197/17 = 11.59, 191/17 = 11.24, 187/17 = 11.00, 185/17 = 10.88$$

$$48/11 = 4.36, 49/11 = 4.45, 53/11 = 4.82, 52/11 = 4.73, 42/11 = 3.82$$

Bu değerlerin dağılım aralığı;

$$(11.88-10.88) = 1.00 \text{ ve } (4.82-3.82) = 1.00'dır.$$

Dağılım aralığı (0, 1) için standartlaştırma, her bir değerden en küçük değer çıkartılıp, elde edilen değerlerin dağılım aralığına bölünmesiyle yapılmaktadır. Aralık sıfırsa bütün değerler 0,50 alınmaktadır. Örneğin, Tablo 2.7.1'deki veriler bu yöntemle;

$$(202-185)/17 = 1.00, (197-185)/17 = 0.71, (191-185)/17 = 0.35, (187-185)/17 = 0.12, (185-185)/17 = 0$$

$$(48-42)/11 = 0.55, (49-42)/11 = 0.64, (53-42)/11 = 1.00, (52-42)/11 = 0.91,$$

$$(42-42)/11 = 0$$

şekilde standartlaştırılır. Bu değerlerin dağılım aralığı;

$(1.00-0) = 1.00$ ve $(1.00-0) = 1.00$ 'dir.

Dağılım aralığı $(-1, 1)$ ve dağılım aralığı $(0, 1)$ için standartlaştırılan değerlerin dağılım aralığı 1 olmakta ve dağılım aralığı $(0, 1)$ 'le standartlaştırılan değerlerin maksimum değeri 1, minimum değeri ise 0 olmaktadır.

Standartlaştırmanın gerekliliği için şu örneği verebiliriz. Farklı otomobilleri birçok özellikleri açısından kümelemek istediğimizde, bu özellikler açısından otomobiller arasındaki uzaklık veya benzerlik değerlerine ihtiyaç duyulur. Bu amaçla her bir otoyola ait; galon başına yaptığı mil, motor kapasitesi, ağırlığı, silindir sayısı ve beygirgücü özellikleri dikkate alınmış olsun. Bu değişkenlerin değerleri skala üzerinde farklı ölçü birimleriyle ölçülmekte ve değişik aralıkta değişmektedir. Çünkü, silindir sayısı 12'ye kadar değişirken, ağırlığı 2000 kg'a kadar çıkabilmektedir. Bu nedenle hesaplanmak istenen uzaklık ölçütü, değişken değerleri arasındaki farklılıklardan etkilenebilir. Örneğin, tüm değişkenlerin birbirinden farkına göre elde edilen bir uzaklık matrisi için büyük değerli değişkenlerle, küçük değerli değişkenler arasında büyük bir fark olacaktır.

Yukarıda belirtilen nedenlerden dolayı, veri özelliğine uygun bir standartlaştırma tekniğiyle standartlaştırılır.

III- YÖNTEM VE UYGULAMA

3.1. UYGULAMA ÇERÇEVESİ VE ANALİZ YÖNTEMİ

UNICEF'in 1995 yılı "Dünya Çocuklarının Durumu" adlı yayınından, Afrika, Asya ve Avrupa ülkelerinden 5'er adet olmak üzere toplam 15 ülke olasılıksız örnekleme yöntemlerinden gelişigüzel (haphazard) yöntemiyle seçilmiştir. Bu ülkelerin, 1993 yılına ait ölçümle belirtilen 9 değişkene ilişkin orijinal veriler ve bu verilere ait; toplam, ortalama ve standart sapma değerler ile z-skoruna dönüştürülmüş değerlere ilişkin veri matrisleri Ek 1'de verilmiştir. 15x9 boyutlu veri matrisinden yola çıkılarak uzaklık ölçütleri yardımı ile uzaklıklar matrisi ve benzerlik ölçütleri yardımı ile de yakınlıklar matrisi hesaplanmıştır. Bu uygulamada kümeleme analizi yapıldığından aşağıda kısaca değinilmiştir.

Uzaklık kümeleme yöntemleriyle yakın ilişkilidir. Kümeleme analizinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak (gruplamak) ve araştırmacıya; uygun, işe yarar özetleyici bilgiler elde etmede yardımcı olmaktır (Manly, 1990).

Küme, birbirine yakın (benzer) bireylerin çok boyutlu uzayda oluşturdukları bulutlar benzetmesi ile ifade edilebilir. Bu ifadelerden de anlaşılacağı gibi küme kavramı, benzerlik ve uzaklık gibi kavramları çağrıştırmaktadır.

Kümeleme analizinde ilk aşama, veri giriş aşamasıdır. Verilerin kümelemeye uygun biçimde girilmesi ile ilgili olan bu aşamada uzaklıklar matrisi elde edilir. İkinci aşama, kümeleme tekniğinin seçilmesi ve uygulanmasıdır. Son aşama ise sonuç aşaması olup bu aşamada, sonuçların duyarlılığının ve anlamlılığının tartışması yapılır. Sonuçların uygun olmaması durumunda (değişkenlerin uygun olmaması ve / veya küme sayısının doğru belirlenmemiş olması nedeniyle) tekrar ikinci aşamaya dönülerek sonuçlar uygun olana kadar işlemlere devam edilir. Sonuçların uygun olup olmamasına; seçilen değişkenler açısından gerçekte birbirine benzemeyen (uzak) iki nesnenin (ülke, bölge, ... v.b.) aynı kümede yer alması verilebilir. Bunun nedeni, uygun uzaklık veya benzerlik ölçütünün kullanılmaması olabileceği gibi yanlış kümeleme tekniğinin seçilmesi de

olabilir. Çünkü uzaklık ya da benzerlik matrisinin yenilenmesi aşamasıralı kümeleme yöntemlerinin türünü belirler.

Bundan sonra z-skoruna dönüştürülerek standartlaştırılan 15x9 boyutlu veri matrisinden yola çıkılarak, aralık veriler için kullanılan uzaklık ve benzerlik ölçütleriyle (her uzaklık ve benzerlik ölçütüyle ayrı ayrı olmak üzere) uzaklıklar ve yakınlıklar matrisi bulunmuş ve bu matrisler yardımıyla, hiyerarşik kümeleme yöntemlerinden en yakın komşuluk-tek bağlantı tekniği (nearest neighbour-single linkage-SLINK) ile (her uzaklıklar ve yakınlıklar matrisiyle ayrı ayrı olmak üzere) kümeleme analizi yapılmıştır. Bu yöntemle elde edilen dendrogramların uzaklık ya da yakınlık değerleri SPSS tarafından 0-25 arasında yeniden ölçeklendiğinden uygulamalarda aynı şekilde sunulmuştur.

Daha sonra orijinal veriler niteliğe çevrilerek nitel veriler için kullanılan uzaklık ölçütleriyle (her uzaklık ölçütüyle ayrı ayrı olmak üzere) uzaklıklar matrisi elde edilmiş ve bu matrisler yardımıyla, yukarıdaki yöntemle (her uzaklıklar matrisiyle ayrı ayrı olmak üzere) kümeleme analizi yapılmıştır. Nitelik veriler Ek 4'de verilmiştir.

Aynı orijinal verilerden Toplam Doğurganlık Oranı (TDO) dikkate alınmadan frekans sayılarına uygun uzaklık ölçütleriyle (her uzaklık ölçütüyle ayrı ayrı olmak üzere) uzaklıklar matrisi bulunmuş ve aynı yöntemle (her uzaklıklar matrisiyle ayrı ayrı olmak üzere) kümeleme analizi yapılmıştır.

Son olarak da aynı veriler değişken değeri iyi olarak kabul ediliyorsa 1, kötü olarak kabul ediliyorsa ise 0 değeri verilerek iki durumlu veri haline getirilmiş ve bu tip veriler için uygun olan uzaklık ve benzerlik ölçütleriyle (her uzaklık ve benzerlik ölçütüyle ayrı ayrı olmak üzere) uzaklıklar ve yakınlıklar matrisi bulunarak aynı yöntemle (her uzaklıklar ve yakınlıklar matrisiyle ayrı ayrı olmak üzere) kümeleme analizi yapılmıştır. İkili veriler Ek 4'de verilmiştir.

Bu uygulamada, uzaklık ve yakınlık matrisleri ile istatistiklerin bulunmasında, kümeleme analizlerinin yapılmasında SPSS for MS WINDOWS Release 5.0, Systat 5.0 istatistik paket programlarından yararlanılmıştır. Bu programlarla bulunamayan uzaklık

ve yakınlık matrisleri ise Excel 5.0 for WINDOWS paket programından yararlanılarak bulunmuştur.

İncelenen 9 Değişkene İlişkin Özet Bilgiler Aşağıda Verilmiştir

<u>Değişken Adı</u>	<u>Kaynak</u>	<u>Yılı</u>
Beş Yaşından Küçüklerde Ölüm Oranı (BYKÖO) (Binde)	UNICEF	1993
Bebek Ölüm Oranı (BÖO) (Binde)	UNICEF	1993
Doğumda Yaşam Beklentisi (DYB) (Yıl)	UNICEF	1993
Tam Olarak Bağışıklananların Yüzdesi (1 Yaşındaki Çocuklarda)		
BCG Bağışıklanma Yüzdesi	UNICEF	1990-93
DBT Bağışıklanma Yüzdesi	UNICEF	1990-93
POLİO Bağışıklanma Yüzdesi	UNICEF	1990-93
KIZAMIK Bağışıklanma Yüzdesi	UNICEF	1990-93
Toplam Doğurganlık Oranı (TDO)	UNICEF	1993
Kentsel Nüfus Oranı (KNO) (Yüzde)	UNICEF	1993

Beş Yaşından Küçüklerde Ölüm Oranı (BYKÖO): Her 1000 canlı doğumda beş yaşından küçük çocuklarda görülen ölüm sayısıdır. Daha somut olarak söylenirse bu, doğumdan başlayarak tam 5 yaşa kadar olan süre içindeki ölüm olasılığıdır. Bu dönem ölümlerinin boyutları bebek ölümlerinden daha az olmakla birlikte, çocuk sağlık düzeyi düşük olan ülkelerde yüksektir. Az gelişmiş ülkelerde yüksek, gelişmiş ülkelerde düşük çıkması beklenir.

Bebek Ölüm Oranı (BÖO): Her bin canlı doğuma karşı kaç bebeğin daha bir yaşına girmeden (0-364 günlük iken) öldüğünü gösterir. Bir yıl içinde, bir yaşına girmeden ölen bebeklerin o yılda canlı doğan bebek sayısına bölünmesiyle elde edilir. Az gelişmiş ülkelerde yüksek, gelişmiş ülkelerde düşük çıkması beklenir.

Doğumda Yaşam Beklentisi (DYB): Belirli bir yılda doğan kişinin ortalama kaç yıl yaşayacağını gösteren ölçüdür. Doğumları sırasında içinde bulundukları nüfus kesiti

açısından geçerli olan ölüm riskleri değişmediği takdirde, yeni doğmuş bebeklerin yaşamları beklenen süredir. Az gelişmiş ülkelerde sağlık düzeyinin düşük olması nedeniyle düşük, gelişmiş ülkelerde ise yüksek çıkması beklenir.

BCG: Günümüzde yaygın olarak kullanılan verem aşısına bağışıklanma yüzdesidir. Bu aşı, canlı fakat etken olmayan basillerle hazırlanır. Basiller veremli bir sığırdan alınır ve bunlardan, yıllar boyunca yeni basiller üretilir. Bu aşuya bağışıklanma; az gelişmiş ülkelerde düşük, gelişmiş ülkelerde ise yüksek çıkması beklenir.

DBT: Difteri, boğmaca ve tetanoz aşısına bağışıklanma yüzdesidir. Bu aşuya bağışıklanma; az gelişmiş ülkelerde düşük, gelişmiş ülkelerde ise yüksek çıkması beklenir.

POLİO: Çocuk felci aşısına bağışıklanma yüzdesidir. Bu aşuya bağışıklanma; az gelişmiş ülkelerde düşük, gelişmiş ülkelerde ise yüksek çıkması beklenir.

KIZAMIK: Kızamık aşısına bağışıklanma yüzdesidir. Bu aşuya bağışıklanma; az gelişmiş ülkelerde düşük, gelişmiş ülkelerde ise yüksek çıkması beklenir.

Toplam Doğurganlık Oranı (TDO): Bu oran doğurgan çağıdaki (15-49 yaş) kadın nüfusun yaş yapısından etkilenmeden doğurganlık düzeyi hakkında kesin bilgi veren bir ölçüdür. Doğurgan çağa giren bir kadının bu çağ sonuna kadar ortalama kaç canlı doğum yapacağını gösterir. Bu oranın az gelişmiş ülkelerde yüksek, gelişmiş ülkelerde düşük çıkması beklenir.

Kentsel Nüfus Oranı (KNO): En son nüfus sayımında kullanılan ulusal tanıma göre kentsel sayılan bölgelerde yaşayanların yüzdesidir. Toplam nüfusun kentsel sayılan bölgelerdeki nüfusa bölünmesiyle bulunur. Bu oranın az gelişmiş ülkelerde düşük, gelişmiş ülkelerde yüksek çıkması beklenir.

3.2. BULGULAR

Ölçüm biçimine uygun uzaklık ve benzerlik ölçütleri yardımı ile uzaklık ve yakınlık matrisleri bulunmuş ve bu matrislerdeki bir değerin (elemanın) bulunuşu formülde ilgili değerler yerine konularak gösterilmiştir. Uzaklık ve yakınlık matrislerinden yola çıkılarak hiyerarşik kümeleme yöntemlerinden tek bağlantı tekniği (Single Linkage) tekniği kullanılarak Ağaç Diyagramı (Dendrogram) çizilmiştir. Uzaklık matrislerinin başlığı Benzemezlik Katsayı Matrisi (Dissimilarity Coefficient Matrix) ve yakınlık matrislerinin başlığı ise Benzerlik Katsayı Matrisi (Similarity Coefficient Matrix) olarak yazılmıştır. Bütün matrisler ve dendrogramlar ölçüm biçimlerine göre aşağıda verilmiştir.

3.2.1. Aralık Veriler ile İlgili Bulgular

Bu tip veriler için uzaklık ve yakınlık matrislerini hesaplamaya geçmeden önce veriler z-skoruna dönüştürülerek standartlaştırılmıştır. Standartlaştırılmış veriler Ek 1'de verilmiştir. Aralık verilerde Ölçekli Öklit Uzaklık Ölçütü ve Karl Pearson Uzaklık Ölçütü burada hesaplanmamıştır. Ayrıca, Canberra Uzaklık Ölçütü, Kosinüs Benzerlik Ölçütü ve Pearson İlişki Katsayısı orijinal verilerden hesaplanmıştır.

Euclidean Dissimilarity Coefficient Matrix

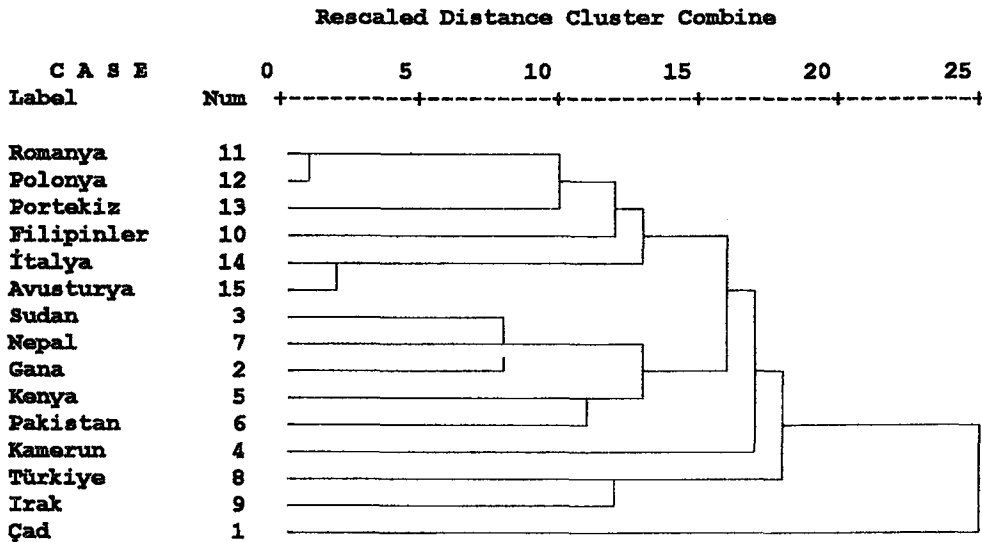
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	3,1074				
Sudan	3,3369	1,2992			
Kamerun	2,5601	2,2052	1,9280		
Kenya	6,1498	3,1864	2,9396	4,4206	
Pakistan	5,0782	2,0086	2,2712	3,6838	1,4556
Nepal	4,2909	1,7537	1,2790	3,0854	2,2097
Türkiye	5,9126	3,5660	3,6755	3,8824	3,1504
Irak	6,4609	3,8845	4,0223	4,3983	2,8476
Filipinler	6,8625	4,0279	3,8868	4,8157	1,8838
Romanya	8,0418	5,3345	5,2304	5,9131	3,2923
Polonya	8,4254	5,7926	5,6765	6,2229	3,8052
Portekiz	8,4133	5,7927	5,5252	6,2609	3,6697
İtalya	8,3227	5,9878	5,9217	6,1195	4,5612
Avusturya	8,0602	5,6359	5,5166	5,8317	4,0735

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,6600				
Türkiye	2,8750	3,6496			
Irak	2,8115	4,0055	1,5391		
Filipinler	2,5325	3,3890	2,0431	1,9743	
Romanya	3,9685	4,7791	2,7125	2,7269	1,5330
Polonya	4,4644	5,3076	2,8649	2,8137	1,9987
Portekiz	4,4901	4,9788	3,4493	3,7086	2,1781
İtalya	5,0450	5,7260	3,2842	3,4859	3,1014
Avusturya	4,6400	5,2637	3,0590	3,2730	2,5754

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,7145			
Portekiz	1,4211	1,6066		
İtalya	2,1966	2,1262	2,8457	
Avusturya	1,6546	1,7099	2,1716	,7954

$$\begin{aligned}
 (2.2.2.1) \text{ formülünden; } d(X_{\text{Çad}}, X_{\text{Gana}}) &= [(1,946 - 1,3724)^2 + (1,7095 - 1,2332)^2 \\
 &+ (-1,6557 - (-0,7993))^2 + (-1,622 - (-0,2969))^2 + (-2,3152 - (-0,9601))^2 \\
 &+ (-2,3154 - (-0,993))^2 + (-2,0247 - (-0,6987))^2 + (0,867 - 0,867)^2 + (-0,5505 - (-0,5505))^2]^{1/2} \\
 &= [0,329 + 0,2269 + 0,7334 + 1,7559 + 1,8363 + 1,7487 + 1,7583 + 0 + 0]^{1/2} \approx 3,1074
 \end{aligned}$$

Dendrogram using Single Linkage



Squared Euclidean Dissimilarity Coefficient Matrix

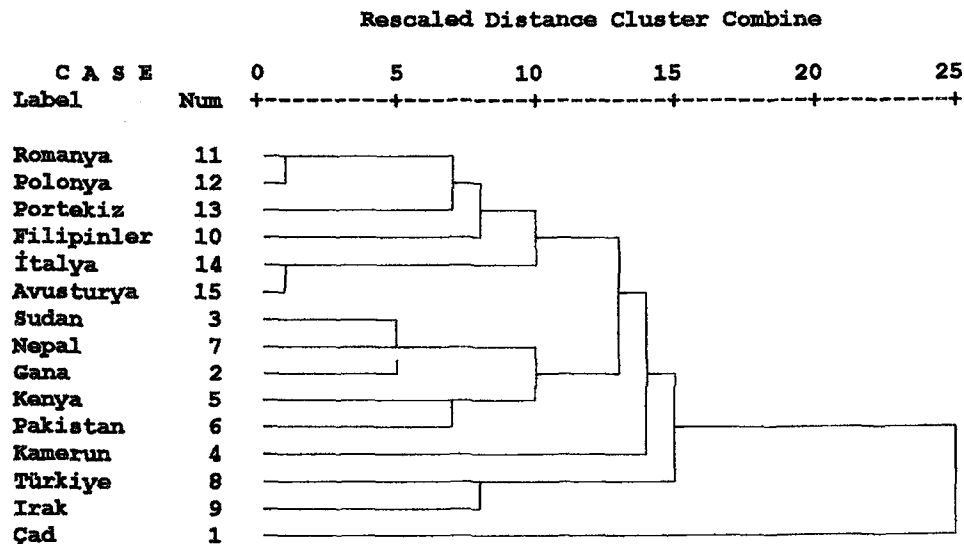
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	9,6559				
Sudan	11,1346	1,6880			
Kamerun	6,5539	4,8630	3,7174		
Kenya	37,8195	10,1531	8,6412	19,5418	
Pakistan	25,7882	4,0345	5,1585	13,5703	2,1187
Nepal	18,4121	3,0754	1,6360	9,5200	4,8829
Türkiye	34,9592	12,7163	13,5093	15,0733	9,9247
Irak	41,7436	15,0893	16,1793	19,3452	8,1090
Filipinler	47,0946	16,2239	15,1076	23,1912	3,5487
Romanya	64,6704	28,4567	27,3566	34,9642	10,8393
Polonya	70,9871	33,5542	32,2231	38,7241	14,4797
Portekiz	70,7831	33,5558	30,5284	39,1990	13,4669
İtalya	69,2676	35,8543	35,0669	37,4481	20,8044
Avusturya	64,9668	31,7636	30,4326	34,0087	16,5932

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	2,7555				
Türkiye	8,2659	13,3194			
Irak	7,9047	16,0442	2,3688		
Filipinler	6,4134	11,4851	4,1741	3,8979	
Romanya	15,7488	22,8395	7,3577	7,4360	2,3501
Polonya	19,9311	28,1701	8,2075	7,9171	3,9949
Portekiz	20,1606	24,7882	11,8977	13,7536	4,7441
İtalya	25,4519	32,7869	10,7863	12,1516	9,6188
Avusturya	21,5300	27,7062	9,3575	10,7124	6,6326

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,5106			
Portekiz	2,0195	2,5811		
İtalya	4,8250	4,5207	8,0978	
Avusturya	2,7377	2,9238	4,7160	,6327

$$\begin{aligned}
 (2.2.3.1) \text{ formülünden; } d(X_{\text{Çad}}, X_{\text{Gana}}) &= (1,946 - 1,3724)^2 + (1,7095 - 1,2332)^2 \\
 &+ (-1,6557 - (-0,7993))^2 + (-1,622 - (-0,2969))^2 + (-2,3152 - (-0,9601))^2 \\
 &+ (-2,3154 - (-0,993))^2 + (-2,0247 - (-0,6987))^2 + (0,867 - 0,867)^2 + (-0,5505 - (-0,5505))^2 \\
 &= 0,329 + 0,2269 + 0,7334 + 1,7559 + 1,8363 + 1,7487 + 1,7583 + 0 + 0 \approx 9,6559
 \end{aligned}$$

Dendrogram using Single Linkage



Chebychev Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	1,7388				
Sudan	1,4779	,6880			
Kamerun	1,4818	1,4007	1,0254		
Kenya	2,9464	1,4779	1,6422	2,6082	
Pakistan	2,5599	1,0501	1,2558	2,2218	,8997
Nepal	1,9835	1,1873	,5937	1,6191	1,0626
Türkiye	2,5669	1,6731	2,2667	1,7890	2,1588
Irak	2,6835	2,1048	2,6985	2,0532	2,5906
Filipinler	2,9558	1,7686	1,6255	2,3667	1,1593
Romanya	3,2908	2,2466	1,9658	2,8014	2,0666
Polonya	3,3058	2,4696	2,1409	2,6949	2,0666
Portekiz	3,4220	2,5334	2,4621	2,8232	2,3690
İtalya	3,3058	2,5652	2,6762	2,6565	2,4698
Avusturya	3,1548	2,5812	2,5691	2,7048	2,3690

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,1334				
Türkiye	1,7270	2,8604			
Irak	2,1588	3,2922	1,1593		
Filipinler	1,3232	1,7270	1,3041	1,5651	
Romanya	2,0162	2,2667	1,7388	1,8146	,9073
Polonya	2,1700	2,6985	1,4973	1,8146	,9715
Portekiz	2,3186	2,2480	1,6731	2,1170	1,2097
İtalya	2,4194	3,0763	1,5939	2,2178	1,5827
Avusturya	2,3288	2,5366	1,6422	2,1170	1,2097

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,4318			
Portekiz	1,0794	1,5112		
İtalya	1,7538	1,9677	2,0960	
Avusturya	1,3260	1,5399	1,6682	,5397

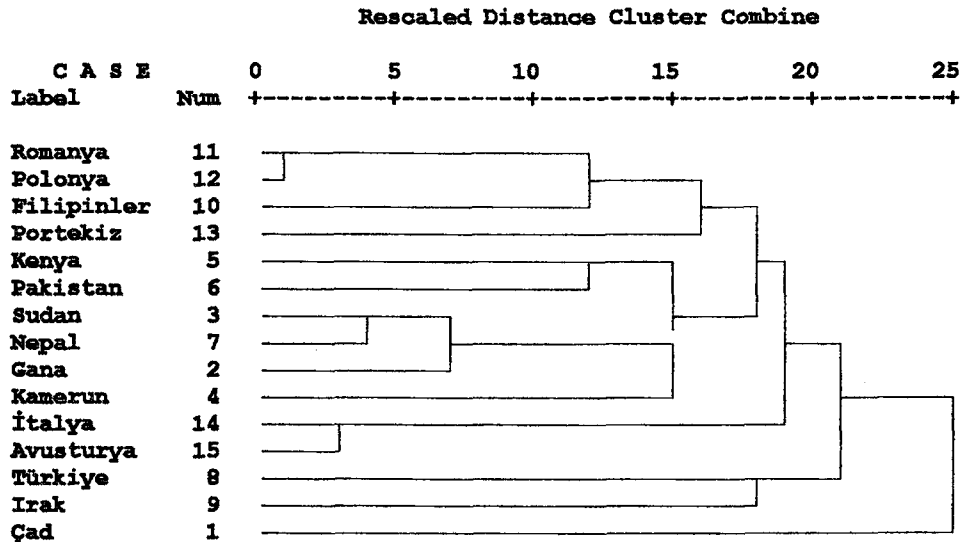
(2.2.4.1) formülünden; $d(X_{\text{çad}}, X_{\text{gana}}) = \max[|1,946-1,3724|, |1,7095-1,2332|$

$, |-1,6557-(-0,7993)|, |-1,622-(-0,2969)|, |-2,3152-(-0,9601)|$

$, |-2,3154-(-0,993)|, |-2,0247-(-0,6987)|, |0,867-0,867|, |-0,5505-(-0,5505)|]$

$= [|0,5736|, |0,4763|, |-0,8564|, |-1,3251|, |-1,3551|, |-1,3224|, |-1,326|, 0,0] \stackrel{N}{=} 1,7388$

Dendrogram using Single Linkage



Manhattan City-Block Dissimilarity Coefficient Matrix

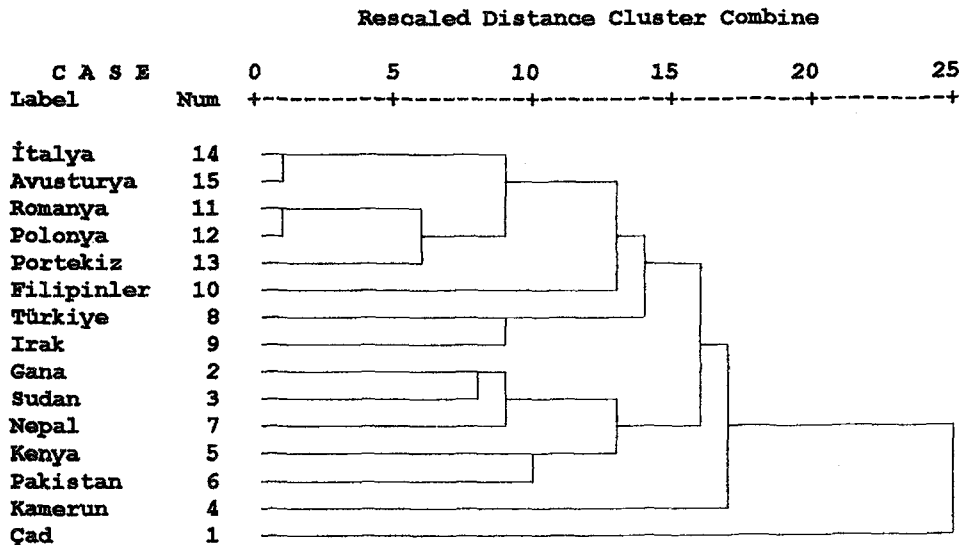
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	7,6485				
Sudan	9,0159	3,1787			
Kamerun	6,6830	5,5407	5,0500		
Kenya	16,2228	8,5742	7,4228	10,6049	
Pakistan	12,6380	4,9895	5,9410	9,0552	3,5847
Nepal	11,6975	4,4772	3,4705	7,7898	6,4325
Türkiye	17,2154	10,2431	9,4877	10,5324	6,7777
Irak	18,1576	10,5091	10,4299	11,4746	5,2200
Filipinler	19,1940	11,5454	11,4662	12,5110	4,7281
Romanya	23,3460	15,6975	15,6183	16,6630	8,3971
Polonya	24,6853	17,0368	16,9576	18,0023	9,8330
Portekiz	23,6497	16,0011	15,9220	17,8302	8,9906
İtalya	24,2687	16,6202	16,5410	17,5857	11,5442
Avusturya	23,5084	15,8599	15,7807	16,8254	9,9284

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	4,2627				
Türkiye	7,2054	8,8586			
Irak	6,6020	9,1372	3,4749		
Filipinler	6,8655	9,8712	5,1776	4,3807	
Romanya	11,0175	14,0232	7,3180	7,2393	4,1520
Polonya	12,3568	15,3625	7,7937	7,7150	5,4913
Portekiz	11,3212	14,3269	9,7804	9,7017	5,5351
İtalya	13,7368	15,7159	9,1066	9,1949	8,2401
Avusturya	12,1210	14,1856	8,1384	8,6586	6,6243

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	1,8223			
Portekiz	3,4484	2,8786		
İtalya	4,9524	3,7511	4,8110	
Avusturya	3,7827	3,4685	3,7383	1,7443

$$\begin{aligned}
 (2.2.5.1) \text{ formülünden; } d(X_{\text{Çad}}, X_{\text{Gana}}) &= |1,946 - 1,3724| + |1,7095 - 1,2332| \\
 &+ |-1,6557 - (-0,7993)| + |-1,622 - (-0,2969)| + |-2,3152 - (-0,9601)| \\
 &+ |-2,3154 - (-0,993)| + |-2,0247 - (-0,6987)| + |0,867 - 0,867| + |-0,5505 - (-0,5505)| \\
 &= 0,5736 + 0,4763 + 0,8564 + 1,3251 + 1,3551 + 1,3224 + 1,326 + 0 + 0 \stackrel{N}{=} 7,6485
 \end{aligned}$$

Dendrogram using Single Linkage



Power(3;3) (Minkowski(3)) Dissimilarity Coefficient Matrix

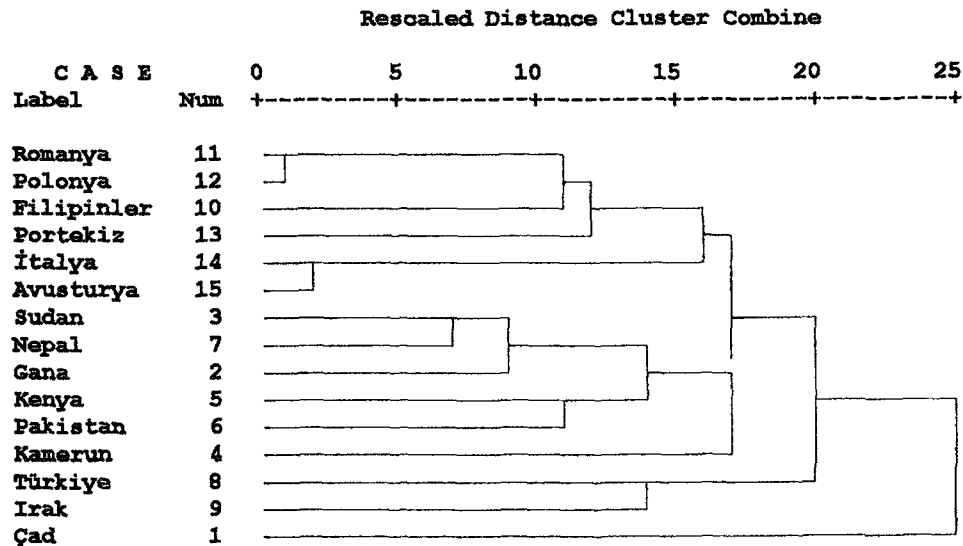
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	2,3701				
Sudan	2,4401	1,0003			
Kamerun	1,9546	1,7017	1,4552		
Kenya	4,5750	2,3452	2,2429	3,4562	
Pakistan	3,8618	1,5390	1,7014	2,8759	1,1324
Nepal	3,1622	1,3798	,9349	2,3482	1,5747
Türkiye	4,2120	2,5474	2,8101	2,8509	2,5825
Irak	4,6198	2,8643	3,1285	3,2663	2,6365
Filipinler	4,9718	2,8910	2,7361	3,6386	1,4769
Romanya	5,7085	3,7618	3,6427	4,2969	2,5534
Polonya	5,9515	4,0878	3,9523	4,4499	2,8894
Portekiz	5,9823	4,1747	3,9413	4,4817	2,8870
İtalya	5,8946	4,3026	4,2510	4,3776	3,4827
Avusturya	5,7089	4,0825	3,9494	4,1714	3,1409

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,3045				
Türkiye	2,2161	3,0533			
Irak	2,3432	3,4415	1,2924		
Filipinler	1,8904	2,4231	1,6177	1,6893	
Romanya	2,9342	3,3880	2,0758	2,1131	1,1474
Polonya	3,2762	3,8100	2,1310	2,1444	1,4738
Portekiz	3,4190	3,5763	2,5044	2,8649	1,6653
İtalya	3,7306	4,2311	2,4047	2,6724	2,3146
Avusturya	3,5105	3,8748	2,3136	2,5154	1,9190

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,5389			
Portekiz	1,1685	1,5222		
İtalya	1,8698	1,9915	2,5185	
Avusturya	1,4075	1,5646	1,9238	,6505

$$\begin{aligned}
 (2.2.6.1) \text{ formülünden } \lambda=3 \text{ için; } d(X_{\text{Çad}}, X_{\text{Gana}}) &= [(1,946 - 1,3724)^3 + (1,7095 - 1,2332)^3 \\
 &+ (-1,6557 - (-0,7993))^3 + (-1,622 - (-0,2969))^3 + (-2,3152 - (-0,9601))^3 \\
 &+ (-2,3154 - (-0,993))^3 + (-2,0247 - (-0,6987))^3 + (0,867 - 0,867)^3 + (-0,5505 - (-0,5505))^3]^{1/3} \\
 &= [0,1887 + 0,1081 + 0,6281 + 2,3267 + 2,4884 + 2,3125 + 2,3315 + 0 + 0]^{1/3} \approx 2,3701
 \end{aligned}$$

Dendrogram using Single Linkage



Power(4;4) (Minkowski(4)) Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	2,0976				
Sudan	2,0998	,8870			
Kamerun	1,7475	1,5356	1,2830		
Kenya	3,9865	2,0266	1,9872	3,0939	
Pakistan	3,4012	1,3616	1,4948	2,5816	1,0230
Nepal	2,7458	1,2660	,8072	2,0702	1,3446
Türkiye	3,5918	2,1713	2,5202	2,4694	2,3773
Irak	3,9262	2,4972	2,8621	2,8354	2,6004
Filipinler	4,2656	2,4718	2,3094	3,2027	1,3403
Romanya	4,8374	3,1792	3,0462	3,7087	2,3068
Polonya	5,0291	3,4579	3,3044	3,8019	2,5637
Portekiz	5,0560	3,5658	3,3547	3,8183	2,6213
İtalya	4,9878	3,6682	3,6279	3,7327	3,0916
Avusturya	4,8310	3,5130	3,3708	3,5623	2,7959

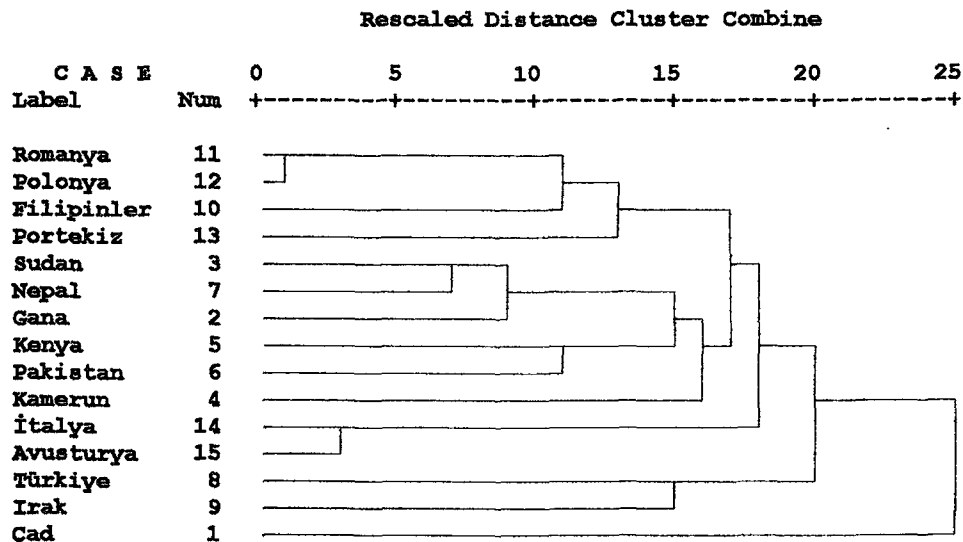
Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,1998				
Türkiye	1,9837	2,9186			
Irak	2,2215	3,3284	1,2197		
Filipinler	1,6647	2,0814	1,4819	1,6138	
Romanya	2,5708	2,8836	1,8813	1,9313	1,0173
Polonya	2,8454	3,2790	1,8727	1,9463	1,2819
Portekiz	3,0210	3,0660	2,1646	2,5843	1,4785
İtalya	3,2451	3,7000	2,0871	2,4129	2,0220
Avusturya	3,0920	3,3610	2,0508	2,2805	1,6682

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,4791			
Portekiz	1,1068	1,5127		
İtalya	1,7905	1,9719	2,3791	
Avusturya	1,3508	1,5442	1,8238	,5997

(2.2.6.1) formülünden lamda=4 için; $d(X_{\text{Çad}}, X_{\text{Gana}}) = [(1,946 - 1,3724)^4 + (1,7095 - 1,2332)^4 + (-1,6557 - (-0,7993))^4 + (-1,622 - (-0,2969))^4 + (-2,3152 - (-0,9601))^4 + (-2,3154 - (-0,993))^4 + (-2,0247 - (-0,6987))^4 + (0,867 - 0,867)^4 + (-0,5505 - (-0,5505))^4]^{1/4}$

$= [0,1083 + 0,0515 + 0,5379 + 3,0831 + 3,372 + 3,0581 + 3,0915 + 0 + 0]^{1/4} \approx 2,0976$

Dendrogram using Single Linkage



Canberra Dissimilarity Coefficient Matrix (Orijinal Verilerden Hesaplanmıştır.)

Case	Çad	Gana	Sudan	Kamerun	Kenya	Pakistan	Nepal
Gana	2,1889						
Sudan	2,6035	0,6672					
Kamerun	1,9808	1,3314	1,2688				
Kenya	3,5390	1,6861	1,3437	2,1764			
Pakistan	2,8736	0,9232	1,1094	1,9086	0,7828		
Nepal	3,1438	1,1286	0,8549	1,9637	1,3517	0,9377	
Türkiye	3,7740	1,9598	1,7965	2,1455	1,1603	1,3264	1,7557
Irak	3,8507	1,9852	1,9311	2,2745	0,8897	1,2441	1,7823
Filipinler	4,0729	2,2781	2,2679	2,5603	1,0376	1,4594	2,1251
Romanya	4,9788	3,2804	3,3317	3,6475	2,1233	2,5057	3,1640
Polonya	5,3008	3,6550	3,7603	4,1046	2,6355	2,9153	3,5627
Portekiz	5,2243	3,5745	3,7286	4,2685	2,6451	2,8429	3,5892
İtalya	5,4623	3,7393	3,8788	4,2852	3,1945	3,3531	3,8221
Avusturya	5,3945	3,6629	3,8137	4,2111	2,9477	3,0936	3,6035

Case	Türkiye	Irak	Filipinler	Romanya	Polonya	Portekiz	İtalya
Irak	0,6770						
Filipinler	1,0147	0,8258					
Romanya	1,8334	1,8001	1,2419				
Polonya	2,2091	2,2143	1,8369	0,7621			
Portekiz	2,7823	2,7710	2,1237	1,4159	0,8816		
İtalya	2,7633	2,7856	2,6567	1,8159	1,2126	1,0245	
Avusturya	2,5824	2,6812	2,3952	1,6212	1,1540	0,8625	0,3794

(2.2.9.1) formülünden; $d(X_{\text{Çad}}, X_{\text{Gana}}) = [|206-170| / (206+170)] + [|121-103| / (121+103)] +$
 $[|48-56| / (48+56)] + \dots + [|35-35| / (35+35)] = (0,0957 + 0,0804 + 0,0769 + 0,3462 + 0,5738$
 $+ 0,5667 + 0,4493 + 0 + 0 = \underline{2,1889}$

Cosine Similarity Coefficient Matrix (Orişinal verilerden hesaplanmıřtır.)

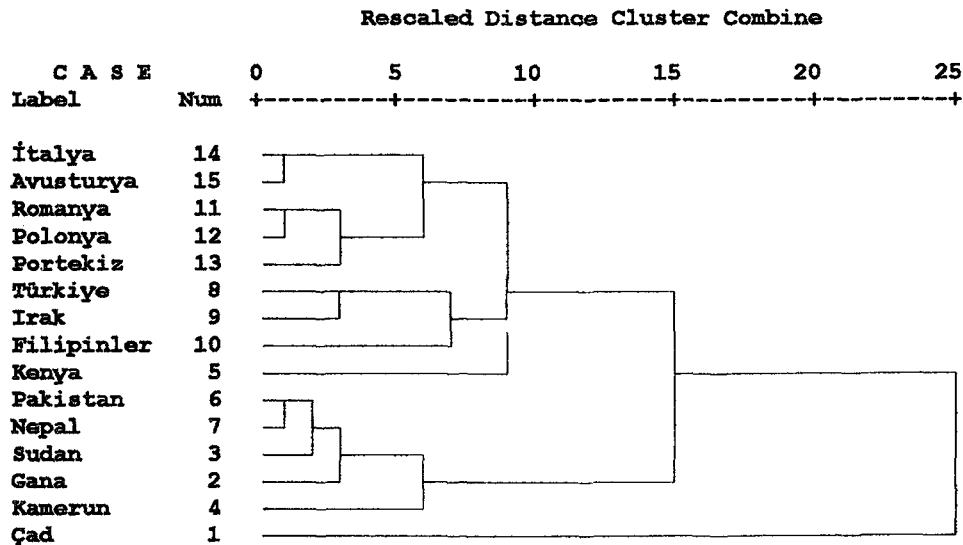
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,9481				
Sudan	,9028	,9918			
Kamerun	,9355	,9867	,9799		
Kenya	,6878	,8825	,9324	,8734	
Pakistan	,8451	,9705	,9913	,9587	,9684
Nepal	,8638	,9757	,9936	,9555	,9540
Türkiye	,7108	,8821	,9236	,9069	,9681
Irak	,6392	,8362	,8855	,8629	,9663
Filipinler	,5651	,7943	,8589	,8052	,9800
Romanya	,4057	,6684	,7477	,6907	,9272
Polonya	,3310	,6031	,6882	,6340	,8891
Portekiz	,3022	,5831	,6751	,6034	,8869
İtalya	,3066	,5708	,6535	,6140	,8541
Avusturya	,3056	,5763	,6611	,6144	,8667

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,9954				
Türkiye	,9552	,9287			
Irak	,9292	,8947	,9926		
Filipinler	,9126	,8837	,9692	,9844	
Romanya	,8185	,7784	,9185	,9537	,9809
Polonya	,7652	,7196	,8903	,9323	,9606
Portekiz	,7526	,7154	,8637	,9052	,9534
İtalya	,7278	,6795	,8610	,9052	,9261
Avusturya	,7369	,6905	,8640	,9098	,9380

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,9952			
Portekiz	,9883	,9912		
İtalya	,9704	,9761	,9564	
Avusturya	,9819	,9864	,9741	,9969

$$\begin{aligned}
 (2.2.12.1) \text{ formülünden; Benzerlik } (X_{\text{çad}}, X_{\text{gana}}) &= [(206 \cdot 170) + (121 \cdot 103) + (48 \cdot 56) + (34 \cdot 70) \\
 &+ (13 \cdot 48) + (13 \cdot 47) + (19 \cdot 50) + (5,9 \cdot 5,9) + (35 \cdot 35)] / [((206)^2 + (121)^2 + (48)^2 + \dots + (35)^2) \cdot \\
 &\cdot ((170)^2 + (103)^2 + (56)^2 + \dots + (35)^2)]^{1/2} = (55995,81) / (62495,81 \cdot 55817,81)^{1/2} \\
 &= 55995,81 / 59062,5029 = 0,9481
 \end{aligned}$$

Dendrogram using Single Linkage



Correlation Similarity Coefficient Matrix (Orijinal verilerden hesaplanmıştır.)

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,9509				
Sudan	,8811	,9810			
Kamerun	,9560	,9605	,9225		
Kenya	,2779	,5597	,6961	,4304	
Pakistan	,7624	,9226	,9709	,8335	,8280
Nepal	,7807	,9251	,9744	,8266	,7962
Türkiye	,3543	,5532	,6452	,5744	,7872
Irak	,0624	,3043	,4131	,3116	,7735
Filipinler	-,1370	,1634	,3245	,0879	,8764
Romanya	-,4693	-,1933	-,0356	-,2357	,6510
Polonya	-,5775	-,3270	-,1754	-,3458	,5278
Portekiz	-,5582	-,2923	-,1229	-,3418	,5899
İtalya	-,5615	-,3464	-,2147	-,3228	,4355
Avusturya	-,5759	-,3428	-,2020	-,3368	,4794

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,9873				
Türkiye	,7428	,6647			
Irak	,5625	,4547	,9320		
Filipinler	,5079	,4387	,7839	,9011	
Romanya	,1595	,0787	,5821	,8041	,9268
Polonya	,0150	-,0690	,5096	,7529	,8650
Portekiz	,0640	,0079	,4711	,6939	,8825
İtalya	-,0513	-,1344	,4395	,6786	,7521
Avusturya	-,0294	-,1086	,4399	,6903	,7975

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,9862			
Portekiz	,9760	,9751		
İtalya	,9076	,9229	,8648	
Avusturya	,9467	,9561	,9192	,9902

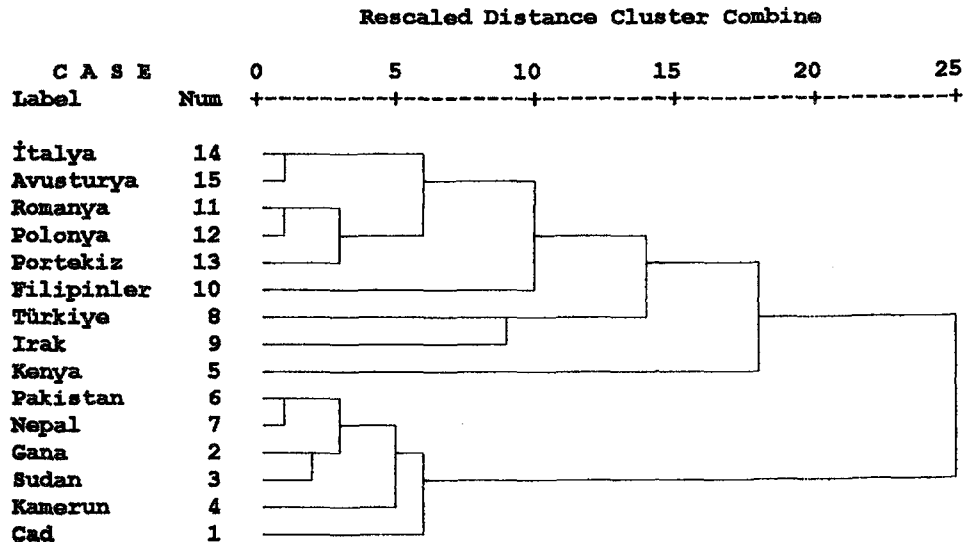
(2.2.13.1) formülünden; Benzerlik ($X_{\text{çad}}, X_{\text{gana}}$) = $(1,946 \cdot 1,3724 + 1,7095 \cdot 1,2332$

$+ (-1,6557 \cdot (-0,7993)) + (-1,622) \cdot (-0,2969) + (-2,3152) \cdot (-0,9601) + (-2,3154) \cdot (-0,993)$

$+ (-2,0247) \cdot (-0,6987) + 0,867 \cdot 0,867 + (-0,5505) \cdot (-0,5505) / 15 - 1 = (2,6707 + 2,1082$

$+ 1,3234 + 0,4816 + 2,2228 + 2,2992 + 1,4147 + 0,7517 + 0,3031 = 13,5754 / 14 = 0,9509$

Dendrogram using Single Linkage



Afrika, Asya ve Avrupa kıtaları birer küme kabul edilerek, kümeler arasındaki uzaklıklar Mahalanobis, Hotelling T^2 , Karl Pearson (Gruplar Arasında) ve Penrose uzaklık ölçütleriyle hesaplanarak uzaklık matrisleri aşağıda verilmiştir.

Kümelerin ortalama vektörleri;

	141,4		95,8		14,4
	86,6		70,8		11,8
	54,2		62,2		74,0
	60,2		78,4		95,6
$\mu_{\text{Afrika}} =$	46,0	$\mu_{\text{Asya}} =$	77,4	$\mu_{\text{Avrupa}} =$	95,0
	45,8		77,6		94,2
	45,4		74,4		79,2
	5,9		4,9		1,7
	32,6		46,4		56,6

Kümelerin Ortalama fark vektörleri;

	45,6		127,0		81,4
	15,8		74,8		59,0
	-8,0		-19,8		-11,8
	-18,2		-35,4		-17,2
$\mu_{\text{Afrika}} - \mu_{\text{Asya}} =$	-31,4	$\mu_{\text{Afrika}} - \mu_{\text{Avrupa}} =$	-49,0	$\mu_{\text{Asya}} - \mu_{\text{Avrupa}} =$	-17,6
	-31,8		-48,4		-16,6
	-29,0		-33,8		-4,8
	1,0		4,2		3,2
	-13,8		-24,0		-10,2

Mahalanobis Dissimilarity Coefficient Matrix

	Afrika	Asya	Avrupa
Afrika	0,0000	Simetrik	
Asya	5,6237	0,0000	
Avrupa	8,1084	7,6291	0,0000

(2.2.7.2) formülünden; Burada kullanılan S^{-1} Ek 3'de verilmiştir.

$$d_2 (X_{Afrika}, X_{Asya}) = \{[(141,4-95,8) (86,6-70,8) (54,2-62,2) \dots (32,6-46,4)] \times S^{-1} \times \begin{matrix} 141,4-95,8 \\ 86,6-70,8 \\ 54,2-62,2 \\ . \\ . \\ . \\ 32,6-46,4 \end{matrix} \}^{1/2}$$

$$= \{[1,8305 -3,0808 0,5135 -0,4722 1,1970 -0,8630 0,1525 6,2646 -0,0468] \times \begin{matrix} 45,6 \\ 15,8 \\ -8,0 \\ -18,2 \\ -31,4 \\ -31,8 \\ -29,0 \\ 1,0 \\ -13,8 \end{matrix} \}^{1/2}$$

$$= \{83,4708 + (-48,6766) + (-4,108) + 8,594 + (-37,5858) + 27,4434 + (-4,4225) + 6,2646 + 0,6458\}^{1/2}$$

$$= \{31,6257\}^{1/2} = \underline{5,6237}$$

Hotelling T^2 Dissimilarity Coefficient Matrix

	Afrika	Asya	Avrupa
Afrika	0,0000	Simetrik	
Asya	79,0643	0,0000	
Avrupa	164,3645	145,5073	0,0000

(2.2.8.1) formülünden; $T^2 = (5 \times 5 / 10) \times [d_2 (X_{Afrika}, X_{Asya})]^2 = (2,5 \times 31,6257) = \underline{79,0643}$

Karl Pearson Uzaklık Ölçütü'nü hesaplayabilmek için kümelerdeki değişkenlerin varyans vektörlerine ihtiyaç vardır. Varyans vektörleri aşağıda verilmiştir.

	2154,8		1210,7		73,8
	610,8		456,2		45,2
	18,2		30,7		8,5
	590,7		118,8		7,3
$S^2_{Afrika} =$	702	$S^2_{Asya} =$	81,8	$S^2_{Avrupa} =$	11,0
	701,2		87,3		13,2
	455,3		112,8		508,7
	0,033		1,395		0,14
	59,3		604,3		175,3

Karl Pearson Dissimilarity Coefficient Matrix (Kümeler Arası)

	Afrika	Asya	Avrupa
Afrika	0,0000	Simetrik	
Asya	2,0633	0,0000	
Avrupa	8,9644	4,1446	0,0000

(2.2.10.2) formülünden; $d_2(\mu_{Afrika}, \mu_{Asya}) = \{1/9 \times [((45,6)^2/(2154,8/5+1210,7/5)) + ((15,8)^2/(610,8/5+456,2/5)) + \dots + ((-13,8)^2/(59,3/5+604,3/5))]\}^{1/2} = [38,3167/9]^{1/2} = [4,2574]^{1/2} = \underline{2,0633}$

Penrose Dissimilarity Coefficient Matrix

	Afrika	Asya	Avrupa
Afrika	0,0000	Simetrik	
Asya	2,1844	0,0000	
Avrupa	12,4108	5,1001	0,0000

Burada kullanılan varyanslar Ek 3'deki kovaryans matrisinden alınmıştır.

(2.2.11.1) formülünden; $P_{Afrika,Asya} = [(45,6)^2/9 \times 1146,433 + (15,8)^2/9 \times 370,733 + (-8,0)^2/9 \times 19,133 + \dots + (-13,8)^2/9 \times 279,633] = [0,2015 + 0,0748 + 0,3717 + 0,154 + 0,4135 + 0,4205 + 0,2603 + 0,2124 + 0,0757] = \underline{2,1844}$

3.2.2. Nitel Veriler ile İlgili Bulgular

Scale Euclidean Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya	Pakistan	Nepal
Gana	0,7071						
Sudan	0,9373	0,6153					
Kamerun	0,7314	0,7314	0,6153				
Kenya	1,7095	1,1390	0,9158	1,3275			
Pakistan	1,3706	0,7928	0,8498	1,1741	0,6809		
Nepal	1,0130	0,5255	0,4770	0,8810	0,8313	0,6706	
Türkiye	1,5654	1,2552	1,2635	1,1750	1,1535	1,0270	1,2078
Irak	2,0153	1,5203	1,5873	1,6175	1,1196	1,0633	1,5139
Filipinler	2,1200	1,5794	1,5694	1,6998	0,9312	1,1263	1,3966
Romanya	2,4749	1,9685	1,9490	2,0664	1,3155	1,4601	1,7906
Polonya	2,7333	2,2296	2,2163	2,2817	1,5456	1,7387	2,0783
Portekiz	2,7641	2,2672	2,2223	2,3184	1,5542	1,8089	2,0655
İtalya	2,6734	2,2410	2,2946	2,2928	1,7655	1,8785	2,1430
Avusturya	2,6103	2,1653	2,1701	2,2189	1,6004	1,7876	2,0092

Case	Türkiye	Irak	Filipinler	Romanya	Polonya	Portekiz	İtalya
Irak	0,8418						
Filipinler	0,9493	0,8408					
Romanya	1,1411	1,0691	0,5255				
Polonya	1,2985	1,0270	0,7431	0,5255			
Portekiz	1,3846	1,2915	0,7491	0,4528	0,3887		
İtalya	1,3152	1,1879	1,0923	1,1009	0,8448	1,0035	
Avusturya	1,2722	1,2338	0,9272	0,9373	0,7762	0,8207	0,3333

(2.3.1.1) formülünden; $d_2(X_{\text{Çad}}, X_{\text{Gana}}) = \{[(6 - 6)^2/(5)^2] + [(5 - 4)^2/(4)^2] + [(1 - 1)^2/(3)^2] + \dots + [(2 - 2)^2/(3)^2]\}^{1/2} = [0 + 0,0625 + 0 + 0,25 + 0,0625 + 0,0625 + 0,0625 + 0 + 0]^{1/2}$
 $= [0,50]^{1/2} = \underline{0,7071}$

Euclidean Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya	Pakistan	Nepal
Gana	2,8284						
Sudan	3,7417	2,4495					
Kamerun	3,1623	3,1623	2,4495				
Kenya	6,8557	4,5826	3,6056	5,1962			
Pakistan	5,4772	3,1623	3,1623	4,6904	2,6458		
Nepal	4,0000	2,0000	2,0000	3,4641	3,3166	2,4495	
Türkiye	6,3246	5,0990	4,8990	4,4721	4,5826	4,4721	4,4721
Irak	7,8740	5,8310	5,6569	6,0000	3,6056	4,0000	5,2915
Filipinler	8,7178	6,6332	6,4807	6,7823	4,1231	5,0990	5,6569
Romanya	10,0000	8,0000	7,8740	8,1240	5,5678	6,3246	7,0711
Polonya	11,1136	9,1652	8,9443	9,0554	6,4031	7,4833	8,2462
Portekiz	11,4018	9,4868	9,2736	9,3808	6,8557	8,0000	8,4853
İtalya	10,9087	9,2195	9,2195	9,1104	7,2111	8,0623	8,4261
Avusturya	10,7703	9,0554	8,9443	8,9443	6,8557	7,8740	8,1240

Case	Türkiye	Irak	Filipinler	Romanya	Polonya	Portekiz	İtalya
Irak	3,4641						
Filipinler	3,7417	3,1623					
Romanya	4,4721	4,2426	2,0000				
Polonya	5,2915	4,4721	2,8284	2,0000			
Portekiz	5,6569	5,4772	3,1623	2,0000	1,4142		
İtalya	5,3852	5,3852	4,1231	4,1231	3,3166	3,6056	
Avusturya	5,2915	5,4772	3,7417	3,7417	3,1623	3,1623	1,0000

(2.3.2.2) formülünden; $d_{\text{Çad,Gana}} = [(6 - 6)^2 + (5 - 4)^2 + (1 - 1)^2 + (1 - 3)^2 + (1 - 2)^2 + \dots +$

$$(2 - 2)^2]^{\frac{1}{2}} = [0 + 1 + 0 + 4 + 1 + 1 + 1 + 0 + 0]^{\frac{1}{2}} = (8)^{\frac{1}{2}} = \underline{2,8284}$$

Mahalanobis-Like Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya	Pakistan	Nepal
Gana	2,0548						
Sudan	2,4557	1,4547					
Kamerun	1,7701	1,9371	1,6745				
Kenya	4,2055	2,6459	2,1544	3,2937			
Pakistan	3,5288	1,8572	2,0095	2,9478	1,5836		
Nepal	2,7816	1,2839	1,1258	2,3611	1,9134	1,5769	
Türkiye	3,8297	2,8669	2,8435	2,8435	2,5894	2,3279	2,7150
Irak	4,7228	3,3094	3,3864	3,7264	2,1802	2,1954	3,1818
Filipinler	5,0700	3,6515	3,6279	4,0415	2,2154	2,6643	3,2367
Romanya	5,6871	4,3567	4,3260	4,6884	2,9627	3,3118	3,9497
Polonya	6,3546	5,1018	5,0643	5,2897	3,6423	4,1366	4,7620
Portekiz	6,5378	5,3283	5,2400	5,5084	3,8828	4,4558	4,9097
İtalya	6,2277	5,2108	5,2744	5,3949	4,1725	4,5373	4,9464
Avusturya	6,2029	5,1205	5,1205	5,3077	3,9761	4,4333	4,7819

Case	Türkiye	Irak	Filipinler	Romanya	Polonya	Portekiz	İtalya
Irak	1,8949						
Filipinler	2,0840	1,7192					
Romanya	2,4424	2,2928	1,0631				
Polonya	3,0789	2,5881	1,7566	1,3984			
Portekiz	3,3881	3,2416	2,0702	1,4907	1,0328		
İtalya	3,2810	3,2071	2,5943	2,4753	1,8772	1,9760	
Avusturya	3,2372	3,2514	2,4077	2,2790	1,7995	1,7238	0,5345

(2.3.3.3) formülünden; $D1_{\text{Çad,Gana}} = \{2 \times [(6 - 6)^2/|6 + 6| + (5 - 4)^2/|5 + 4| + (1 - 1)^2/|1 + 1| + \dots + (2 - 2)^2/|2 + 2|]\}^{1/2} = \{2 \times (0 + 0,1111 + 0 + 1 + 0,3333 + 0,3333 + 0,3333 + 0 + 0)\}^{1/2}$
 $= [2 \times (2,1110)]^{1/2} = [4,2220]^{1/2} = \underline{2,0548}$

Bhattacharyya Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya	Pakistan	Nepal
Gana	1,0518						
Sudan	1,2519	0,7325					
Kamerun	0,8490	0,9906	0,8468				
Kenya	2,1914	1,3390	1,0897	1,7129			
Pakistan	1,8412	0,9379	1,0158	1,5247	0,8014		
Nepal	1,4325	0,6474	0,5651	1,2132	0,9656	0,7958	
Türkiye	1,9756	1,4573	1,4541	1,4541	1,3189	1,1772	1,3876
Irak	2,4636	1,6847	1,7429	1,9334	1,1359	1,0193	1,6406
Filipinler	2,6580	1,8677	1,8526	2,1091	1,1216	1,3541	1,6495
Romanya	3,0010	2,2447	2,2289	2,4492	1,5150	1,6943	2,0300
Polonya	3,3936	2,6754	2,6366	2,7783	1,8825	2,1595	2,4853
Portekiz	3,5026	2,8123	2,7598	2,9104	2,0515	2,3646	2,5816
İtalya	3,3549	2,7526	2,7874	2,8527	2,2100	2,4045	2,6111
Avusturya	3,3186	2,7083	2,7029	2,8100	2,1023	2,3536	2,5207

Case	Türkiye	Irak	Filipinler	Romanya	Polonya	Portekiz	İtalya
Irak	0,9558						
Filipinler	1,0593	0,8683					
Romanya	1,2381	1,1656	0,5332				
Polonya	1,5760	1,3227	0,8855	0,7070			
Portekiz	1,7377	1,6950	1,0574	0,7553	0,5221		
İtalya	1,6845	1,6785	1,3223	1,2606	0,9585	1,0093	
Avusturya	1,6630	1,6998	1,2274	1,1606	0,9203	0,8812	0,2679

(2.3.4.3) formülünden; $D2_{\text{Çad,Gana}} = \{[(6)^{\frac{1}{2}} - (6)^{\frac{1}{2}}]^2 + [(5)^{\frac{1}{2}} - (4)^{\frac{1}{2}}]^2 + [(1)^{\frac{1}{2}} - (1)^{\frac{1}{2}}]^2 + \dots +$

$$[(2)^{\frac{1}{2}} - (2)^{\frac{1}{2}}]^2\}^{\frac{1}{2}} = [(2,4495 - 2,4495)^2 + (2,2361 - 2)^2 + \dots + (1,4142 - 1,4142)^2]^{\frac{1}{2}}$$

$$= [0 + 0,0557 + 0 + 0,5359 + 0,1716 + 0,1716 + 0,1716 + 0 + 0]^{\frac{1}{2}} = \underline{1,0518}$$

3.2.3. Frekans Sayıları ile İlgili Bulgular

Chi-Square Dissimilarity Coefficient Matrix

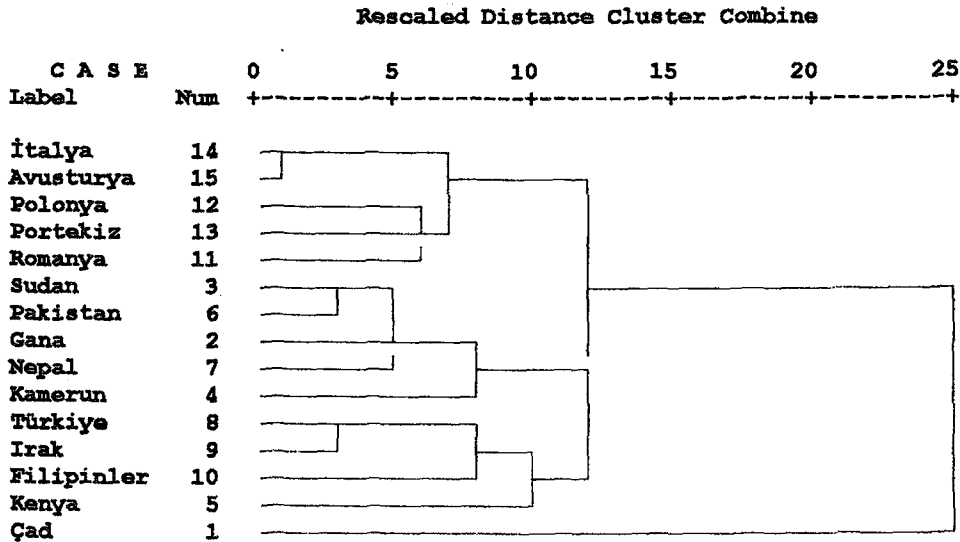
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	8,0073				
Sudan	9,8094	2,4280			
Kamerun	7,6757	3,2308	4,0269		
Kenya	15,1937	8,1951	5,7316	8,3549	
Pakistan	11,8837	4,3376	2,1320	5,3680	4,2277
Nepal	11,5260	4,6477	2,5829	6,3796	5,1767
Türkiye	14,4666	8,2430	6,4701	6,7216	5,0330
Irak	15,9591	9,7366	7,9129	8,2399	5,2307
Filipinler	17,2376	10,7511	8,5500	9,8846	3,8976
Romanya	20,2308	14,3157	12,2993	13,1683	7,8789
Polonya	21,7460	16,1909	14,3181	14,9789	10,2433
Portekiz	21,9110	16,3383	14,3688	15,4983	10,1771
İtalya	21,7069	16,6419	14,9963	15,2885	11,5753
Avusturya	21,5483	16,3267	14,6633	15,1455	11,0763

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	2,7529				
Türkiye	5,5089	7,3519			
Irak	6,8056	8,6035	2,1091		
Filipinler	7,1982	8,5852	4,4143	3,2451	
Romanya	11,1583	12,3679	7,8564	6,2291	4,4193
Polonya	13,2666	14,4241	9,8544	8,3163	6,9546
Portekiz	13,3567	14,1623	10,8474	9,5791	7,5360
İtalya	14,1756	15,2725	10,9578	9,6259	8,8372
Avusturya	13,7633	14,8522	10,7463	9,3800	8,3173

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	2,8632			
Portekiz	4,1601	2,8474		
İtalya	5,5338	4,0408	5,3333	
Avusturya	4,8175	3,0971	4,0925	1,4237

(2.4.1.1) formülünden; Uzaklık ($X_{\text{Çad}}, X_{\text{Gana}}$) = $\left[\left((206-172,2)^2 / 172,2 + (121-102,6)^2 / 102,6 + (48-47,6)^2 / 47,6 + \dots + (35-32)^2 / 32 \right) + \left((170-203,8)^2 / 203,8 + (103-121,4)^2 / 121,4 + (56-56,4)^2 / 56,4 + \dots + (35-38)^2 / 38 \right) \right]^{1/2} = [34,1252 + 29,3349]^{1/2} \approx 8,0073$

Dendrogram using Single Linkage



Phi-Square Dissimilarity Coefficient Matrix

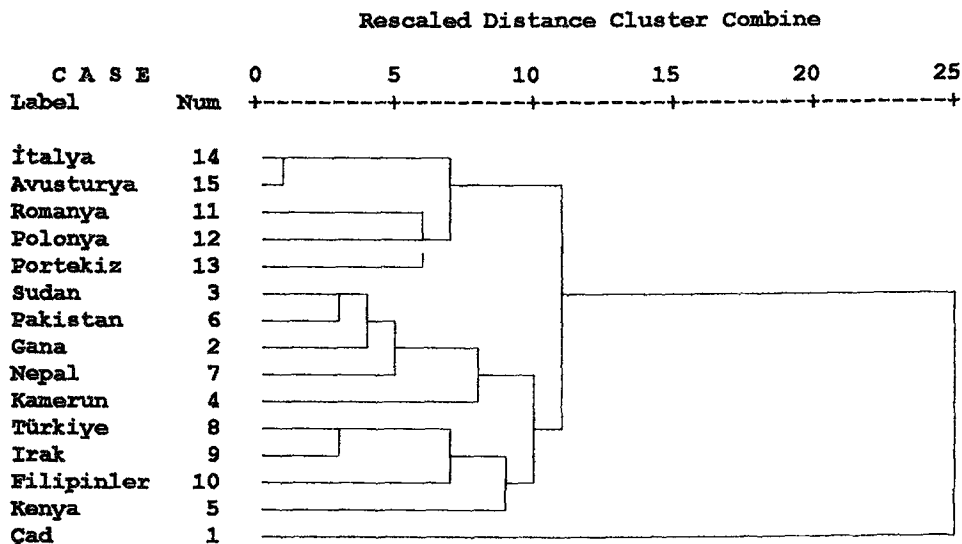
Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,2450				
Sudan	,3130	,0742			
Kamerun	,2542	,1021	,1331		
Kenya	,4634	,2410	,1752	,2642	
Pakistan	,3551	,1247	,0636	,1653	,1216
Nepal	,3584	,1386	,0802	,2050	,1545
Türkiye	,4427	,2422	,1976	,2123	,1480
Irak	,4854	,2845	,2402	,2586	,1530
Filipinler	,5302	,3174	,2625	,3140	,1152
Romanya	,6255	,4247	,3796	,4206	,2340
Polonya	,6750	,4821	,4436	,4804	,3053
Portekiz	,6939	,4956	,4542	,5079	,3090
İtalya	,6895	,5062	,4754	,5027	,3524
Avusturya	,6894	,4998	,4682	,5018	,3394

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,0803				
Türkiye	,1584	,2193			
Irak	,1946	,2552	,0616		
Filipinler	,2079	,2573	,1303	,0953	
Romanya	,3237	,3726	,2331	,1838	,1318
Polonya	,3862	,4361	,2934	,2462	,2081
Portekiz	,3958	,4364	,3290	,2888	,2297
İtalya	,4211	,4720	,3333	,2910	,2702
Avusturya	,4114	,4621	,3290	,2854	,2559

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0861			
Portekiz	,1275	,0876		
İtalya	,1701	,1246	,1678	
Avusturya	,1490	,0962	,1297	,0452

$$\begin{aligned}
 (2.4.2.1) \text{ formülünden; Uzaklık } (X_{\text{çad}}, X_{\text{gana}}) &= [(((206-172,2)^2/172,2 + (121-102,6)^2/102,6 \\
 &+ (48-47,6)^2/47,6 + \dots + (35-32)^2/32) + ((170-203,8)^2/203,8 + (103-121,4)^2/121,4 \\
 &+ (56-56,4)^2/56,4 + \dots + (35-38)^2/38)) / 1068]^{1/2} = [(34,1252 + 29,3349) / 1068]^{1/2} \\
 &= [63,4601 / 1068]^{1/2} \approx 0,2450
 \end{aligned}$$

Dendrogram using Single Linkage



3.2.4. İkili Veriler ile İlgili Bulgular

Orijinal veriler ikili verilere aşağıdaki şekilde dönüştürülmüştür;

BYKÖO $\leq 85 \rightarrow 1$ $\geq 86 \rightarrow 0$	POLİO $\geq 80 \rightarrow 1$ $\leq 79 \rightarrow 0$
BÖO $\leq 50 \rightarrow 1$ $\geq 51 \rightarrow 0$	KIZAMIK $\geq 80 \rightarrow 1$ $\leq 79 \rightarrow 0$
DYB $\geq 65 \rightarrow 1$ $\leq 64 \rightarrow 0$	TDO $\leq 3 \rightarrow 1$ $\geq 4 \rightarrow 0$
BCG $\geq 80 \rightarrow 1$ $\leq 79 \rightarrow 0$	KNO $\geq 60 \rightarrow 1$ $\leq 59 \rightarrow 0$
DBT $\geq 80 \rightarrow 1$ $\leq 79 \rightarrow 0$	

Bu dönüştürmeden sonra ülkelerin ikiye bölünmüş gruplar halinde 2x2 boyutlu olağanlık tabloları oluşturulur. Bu tablolardan uzaklık ölçütleri yardımı ile benzemezlik katsayı matrisleri, benzerlik ölçütleri yardımı ile benzerlik katsayı matrisleri hesaplanır. Çad ile diğer ülkelerin olağanlık tabloları aşağıda verilmiştir.

		Çad	
		1	0
	1	0	0
Gana	0	0	9

		Çad	
		1	0
	1	0	0
Sudan	0	0	9

		Çad	
		1	0
	1	0	0
Kamerun	0	0	9

		Çad	
		1	0
	1	0	3
Kenya	0	0	6

		Çad	
		1	0
	1	0	1
Pakistan	0	0	8

		Çad	
		1	0
	1	0	0
Nepal	0	0	9

		Çad	
		1	0
1	0	3	
Türkiye	0	0	6

		Çad	
		1	0
1	0	6	
Irak	0	0	3

		Çad	
		1	0
1	0	7	
Filipinler	0	0	2

		Çad	
		1	0
1	0	8	
Romanya	0	0	1

		Çad	
		1	0
1	0	9	
Polonya	0	0	0

		Çad	
		1	0
1	0	8	
Portekiz	0	0	1

		Çad	
		1	0
1	0	8	
İtalya	0	0	1

		Çad	
		1	0
1	0	8	
Avusturya	0	0	1

Olağanlık tablolarından hesaplanan benzemezlik katsayı matrisleri aşağıda verilmiştir. Ancak, benzerlik ölçütleri yardımı ile hesaplanan benzerlik katsayı matrislerinin birçok elemanı belirsiz olduğundan bu matrisler aşağıda verilmemiştir.

Binary Squared Euclidean Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	3,0000	3,0000	3,0000	3,0000	
Pakistan	1,0000	1,0000	1,0000	1,0000	2,0000
Nepal	,0000	,0000	,0000	,0000	3,0000
Türkiye	3,0000	3,0000	3,0000	3,0000	6,0000
Irak	6,0000	6,0000	6,0000	6,0000	5,0000
Filipinler	7,0000	7,0000	7,0000	7,0000	4,0000
Romanya	8,0000	8,0000	8,0000	8,0000	5,0000
Polonya	9,0000	9,0000	9,0000	9,0000	6,0000
Portekiz	8,0000	8,0000	8,0000	8,0000	5,0000
İtalya	8,0000	8,0000	8,0000	8,0000	5,0000
Avusturya	8,0000	8,0000	8,0000	8,0000	5,0000

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,0000				
Türkiye	4,0000	3,0000			
Irak	7,0000	6,0000	3,0000		
Filipinler	6,0000	7,0000	6,0000	3,0000	
Romanya	7,0000	8,0000	7,0000	4,0000	1,0000
Polonya	8,0000	9,0000	6,0000	3,0000	2,0000
Portekiz	7,0000	8,0000	7,0000	4,0000	1,0000
İtalya	7,0000	8,0000	5,0000	4,0000	3,0000
Avusturya	7,0000	8,0000	5,0000	4,0000	3,0000

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	1,0000			
Portekiz	,0000	1,0000		
İtalya	2,0000	1,0000	2,0000	
Avusturya	2,0000	1,0000	2,0000	,0000

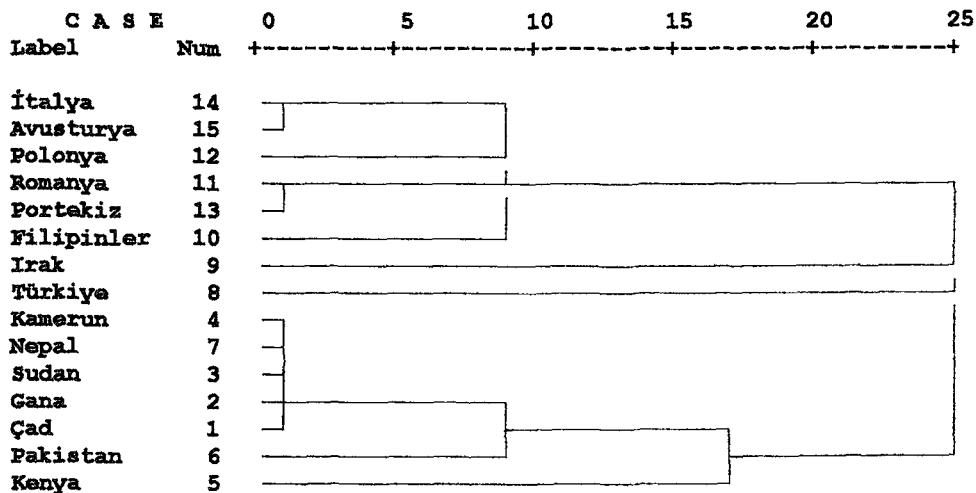
Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

	Çad		
	1 0		
	--- ---		
1	0 3		
Kenya			
0	0 6		

Buradan, Uzaklık ($X_{\text{çad}}, X_{\text{kenya}}$) = $b+c=3+0=3$

Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



Binary Euclidean Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	1,7321	1,7321	1,7321	1,7321	
Pakistan	1,0000	1,0000	1,0000	1,0000	1,4142
Nepal	,0000	,0000	,0000	,0000	1,7321
Türkiye	1,7321	1,7321	1,7321	1,7321	2,4495
Irak	2,4495	2,4495	2,4495	2,4495	2,2361
Filipinler	2,6458	2,6458	2,6458	2,6458	2,0000
Romanya	2,8284	2,8284	2,8284	2,8284	2,2361
Polonya	3,0000	3,0000	3,0000	3,0000	2,4495
Portekiz	2,8284	2,8284	2,8284	2,8284	2,2361
İtalya	2,8284	2,8284	2,8284	2,8284	2,2361
Avusturya	2,8284	2,8284	2,8284	2,8284	2,2361

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,0000				
Türkiye	2,0000	1,7321			
Irak	2,6458	2,4495	1,7321		
Filipinler	2,4495	2,6458	2,4495	1,7321	
Romanya	2,6458	2,8284	2,6458	2,0000	1,0000
Polonya	2,8284	3,0000	2,4495	1,7321	1,4142
Portekiz	2,6458	2,8284	2,6458	2,0000	1,0000
İtalya	2,6458	2,8284	2,2361	2,0000	1,7321
Avusturya	2,6458	2,8284	2,2361	2,0000	1,7321

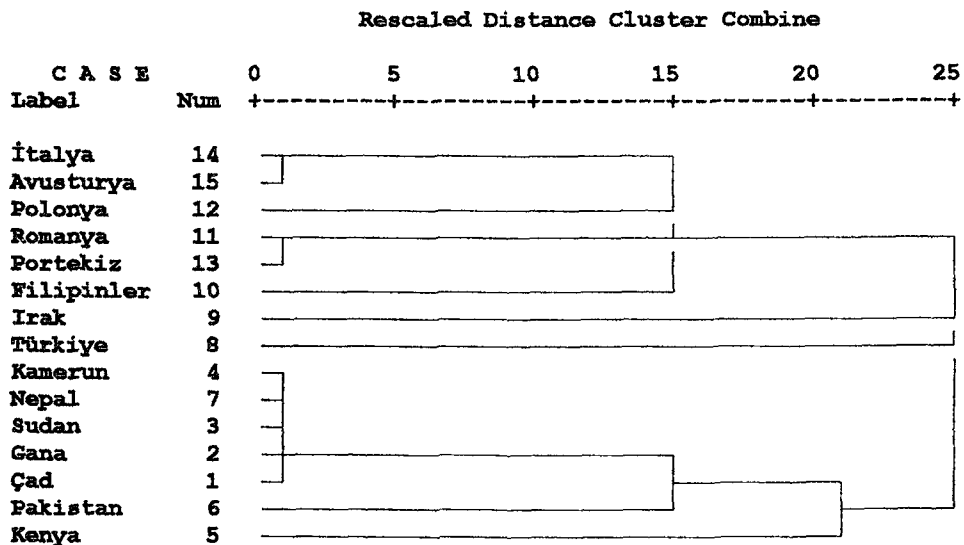
Case	Romanya	Polonya	Portekiz	İtalya
Polonya	1,0000			
Portekiz	,0000	1,0000		
İtalya	1,4142	1,0000	1,4142	
Avusturya	1,4142	1,0000	1,4142	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

	Çad		
	1 0		
	--- ---		
1	0 3		
Kenya			
0	0 6		

Buradan, Uzaklık ($X_{\text{Çad}}, X_{\text{Kenya}}$) = $(b+c)^{\frac{1}{2}} = (3+0)^{\frac{1}{2}} = 1,7321$

Dendrogram using Single Linkage



Size Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	,1111	,1111	,1111	,1111	
Pakistan	,0123	,0123	,0123	,0123	,0494
Nepal	,0000	,0000	,0000	,0000	,1111
Türkiye	,1111	,1111	,1111	,1111	,0000
Irak	,4444	,4444	,4444	,4444	,1111
Filipinler	,6049	,6049	,6049	,6049	,1975
Romanya	,7901	,7901	,7901	,7901	,3086
Polonya	1,0000	1,0000	1,0000	1,0000	,4444
Portekiz	,7901	,7901	,7901	,7901	,3086
İtalya	,7901	,7901	,7901	,7901	,3086
Avusturya	,7901	,7901	,7901	,7901	,3086

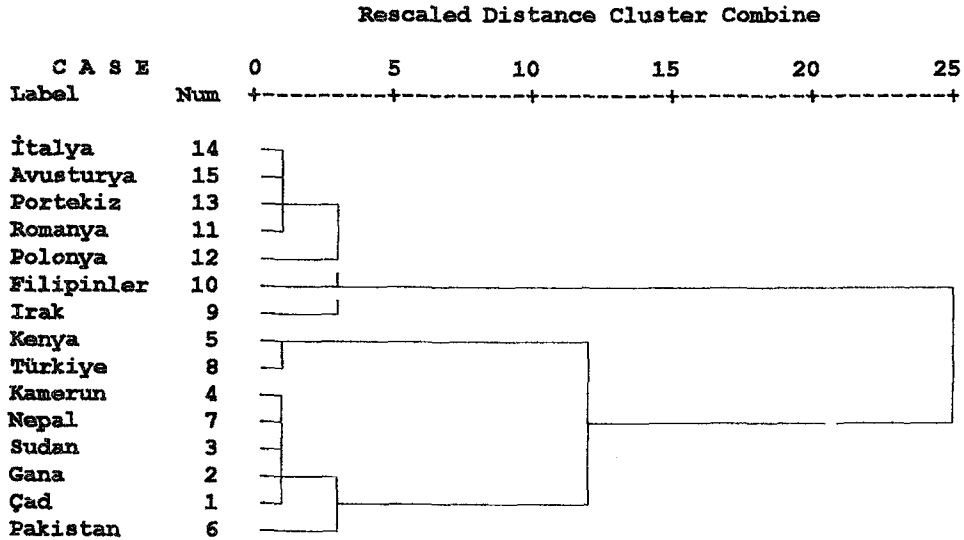
Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,0123				
Türkiye	,0494	,1111			
Irak	,3086	,4444	,1111		
Filipinler	,4444	,6049	,1975	,0123	
Romanya	,6049	,7901	,3086	,0494	,0123
Polonya	,7901	1,0000	,4444	,1111	,0494
Portekiz	,6049	,7901	,3086	,0494	,0123
İtalya	,6049	,7901	,3086	,0494	,0123
Avusturya	,6049	,7901	,3086	,0494	,0123

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0123			
Portekiz	,0000	,0123		
İtalya	,0000	,0123	,0000	
Avusturya	,0000	,0123	,0000	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

	Çad		
	1	0	Buradan, Uzaklık ($X_{\text{çad}}, X_{\text{kenya}}$) = $(b-c)^2 / (a+b+c+d)^2$
	1	0	3
Kenya	0	0	6
			= $(3-0)^2 / (0+3+0+6)^2 = 9/81$
			= 0,1111

Dendrogram using Single Linkage



Pattern Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	,0000	,0000	,0000	,0000	
Pakistan	,0000	,0000	,0000	,0000	,0000
Nepal	,0000	,0000	,0000	,0000	,0000
Türkiye	,0000	,0000	,0000	,0000	,1111
Irak	,0000	,0000	,0000	,0000	,0494
Filipinler	,0000	,0000	,0000	,0000	,0000
Romanya	,0000	,0000	,0000	,0000	,0000
Polonya	,0000	,0000	,0000	,0000	,0000
Portekiz	,0000	,0000	,0000	,0000	,0000
İtalya	,0000	,0000	,0000	,0000	,0000
Avusturya	,0000	,0000	,0000	,0000	,0000

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,0000				
Türkiye	,0370	,0000			
Irak	,0741	,0000	,0000		
Filipinler	,0000	,0000	,0617	,0247	
Romanya	,0000	,0000	,0741	,0370	,0000
Polonya	,0000	,0000	,0000	,0000	,0000
Portekiz	,0000	,0000	,0741	,0370	,0000
İtalya	,0000	,0000	,0000	,0370	,0247
Avusturya	,0000	,0000	,0000	,0370	,0247

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0000			
Portekiz	,0000	,0000		
İtalya	,0123	,0000	,0123	
Avusturya	,0123	,0000	,0123	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

	Pakistan		
	1 0	Buradan, Uzaklık (Xpakistan,Xtürkiye)=(b*c) / (a+b+c+d) ²	
	1 0 3	= (3*1) / (0+3+1+5) ² =3/81	
Türkiye	0 1 5		=0,0370

Dendrogram using Single Linkage

		Rescaled Distance Cluster Combine					
C A S E		0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+					
İtalya	14						
Avusturya	15						
Polonya	12						
Portekiz	13						
Romanya	11						
Filipinler	10						
Nepal	7						
Türkiye	8						
Irak	9						
Pakistan	6						
Kenya	5						
Kamerun	4						
Sudan	3						
Gana	2						
Çad	1						

Variance Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	,0833	,0833	,0833	,0833	
Pakistan	,0278	,0278	,0278	,0278	,0556
Nepal	,0000	,0000	,0000	,0000	,0833
Türkiye	,0833	,0833	,0833	,0833	,1667
Irak	,1667	,1667	,1667	,1667	,1389
Filipinler	,1944	,1944	,1944	,1944	,1111
Romanya	,2222	,2222	,2222	,2222	,1389
Polonya	,2500	,2500	,2500	,2500	,1667
Portekiz	,2222	,2222	,2222	,2222	,1389
İtalya	,2222	,2222	,2222	,2222	,1389
Avusturya	,2222	,2222	,2222	,2222	,1389

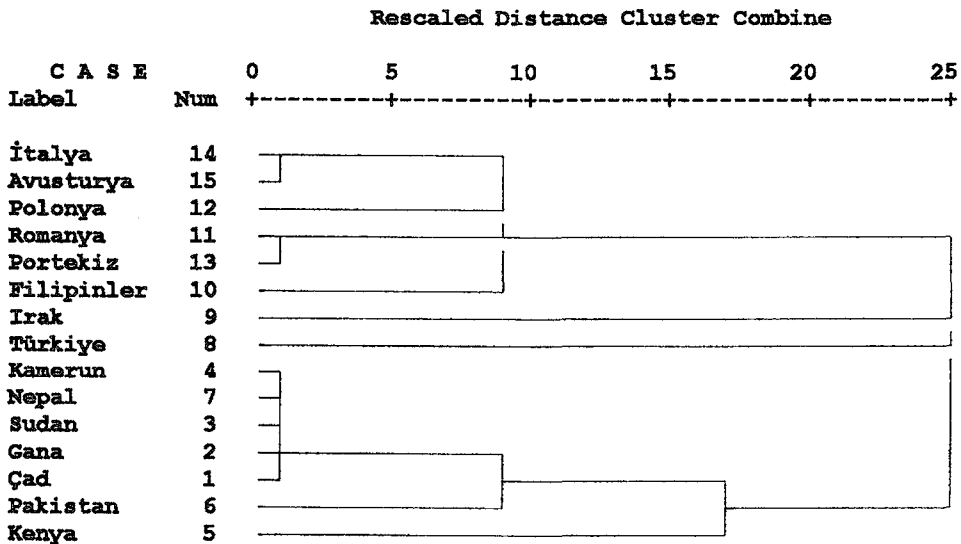
Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,0278				
Türkiye	,1111	,0833			
Irak	,1944	,1667	,0833		
Filipinler	,1667	,1944	,1667	,0833	
Romanya	,1944	,2222	,1944	,1111	,0278
Polonya	,2222	,2500	,1667	,0833	,0556
Portekiz	,1944	,2222	,1944	,1111	,0278
İtalya	,1944	,2222	,1389	,1111	,0833
Avusturya	,1944	,2222	,1389	,1111	,0833

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0278			
Portekiz	,0000	,0278		
İtalya	,0556	,0278	,0556	
Avusturya	,0556	,0278	,0556	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

Çad			
	1	0	Buradan, Uzaklık (X _{Çad} , X _{Kenya}) = (b+c) / 4 * (a+b+c+d)
	1	0	
Kenya	0	3	= (3+0) / 4 * (0+3+0+6) = 3/36
	0	6	
			=0,0833

Dendrogram using Single Linkage



Shape Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	,0000				
Sudan	,0000	,0000			
Kamerun	,0000	,0000	,0000		
Kenya	,2222	,2222	,2222	,2222	
Pakistan	,0988	,0988	,0988	,0988	,1728
Nepal	,0000	,0000	,0000	,0000	,2222
Türkiye	,2222	,2222	,2222	,2222	,6667
Irak	,2222	,2222	,2222	,2222	,4444
Filipinler	,1728	,1728	,1728	,1728	,2469
Romanya	,0988	,0988	,0988	,0988	,2469
Polonya	,0000	,0000	,0000	,0000	,2222
Portekiz	,0988	,0988	,0988	,0988	,2469
İtalya	,0988	,0988	,0988	,0988	,2469
Avusturya	,0988	,0988	,0988	,0988	,2469

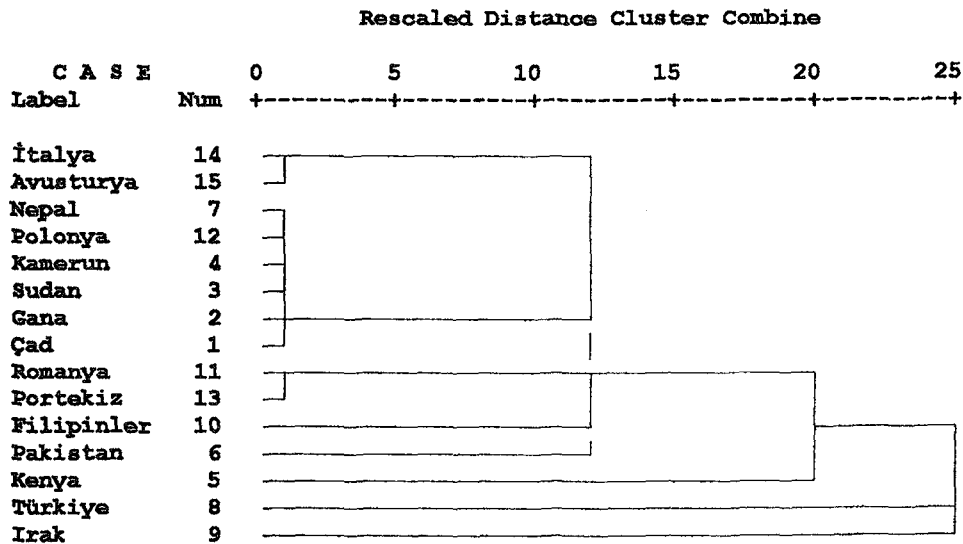
Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	,0988				
Türkiye	,3951	,2222			
Irak	,4691	,2222	,2222		
Filipinler	,2222	,1728	,4691	,3210	
Romanya	,1728	,0988	,4691	,3951	,0988
Polonya	,0988	,0000	,2222	,2222	,1728
Portekiz	,1728	,0988	,4691	,3951	,0988
İtalya	,1728	,0988	,2469	,3951	,3210
Avusturya	,1728	,0988	,2469	,3951	,3210

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0988			
Portekiz	,0000	,0988		
İtalya	,2222	,0988	,2222	
Avusturya	,2222	,0988	,2222	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

	Çad		
	1	0	Buradan, Uzaklık ($X_{\text{Çad}}, X_{\text{Kenya}}$) = $(a+b+c+d) * (b+c) - (b-c)^2 / (a+b+c+d)^2$
1	0	3	$= (0+3+0+6) * (3+0) - (3-0)^2 / (0+3+0+6)^2$
Kenya	0	0	$= (27-9) / 81 = 18/81 = 0,2222$

Dendrogram using Single Linkage



Binary Lance and Williams Dissimilarity Coefficient Matrix

Case	Çad	Gana	Sudan	Kamerun	Kenya
Gana	.	.	.	.	.
Sudan	.	.	.	.	.
Kamerun	.	.	.	.	.
Kenya	1,0000	1,0000	1,0000	1,0000	.
Pakistan	1,0000	1,0000	1,0000	1,0000	,5000
Nepal	.	.	.	.	1,0000
Türkiye	1,0000	1,0000	1,0000	1,0000	1,0000
Irak	1,0000	1,0000	1,0000	1,0000	,5556
Filipinler	1,0000	1,0000	1,0000	1,0000	,4000
Romanya	1,0000	1,0000	1,0000	1,0000	,4545
Polonya	1,0000	1,0000	1,0000	1,0000	,5000
Portekiz	1,0000	1,0000	1,0000	1,0000	,4545
İtalya	1,0000	1,0000	1,0000	1,0000	,4545
Avusturya	1,0000	1,0000	1,0000	1,0000	,4545

Case	Pakistan	Nepal	Türkiye	Irak	Filipinler
Nepal	1,0000	.	.	.	.
Türkiye	1,0000	1,0000	.	.	.
Irak	1,0000	1,0000	,3333	.	.
Filipinler	,7500	1,0000	,6000	,2308	.
Romanya	,7778	1,0000	,6364	,2857	,0667
Polonya	,8000	1,0000	,5000	,2000	,1250
Portekiz	,7778	1,0000	,6364	,2857	,0667
İtalya	,7778	1,0000	,4545	,2857	,2000
Avusturya	,7778	1,0000	,4545	,2857	,2000

Case	Romanya	Polonya	Portekiz	İtalya
Polonya	,0588	.	.	.
Portekiz	,0000	,0588	.	.
İtalya	,1250	,0588	,1250	.
Avusturya	,1250	,0588	,1250	,0000

Çad ve Kenya için olağanlık tablosu ve formülle gösterimi aşağıda verilmiştir.

		Çad	
		1 0	Buradan, Uzaklık (X _{Çad} , X _{Kenya}) = (b+c) / (2a+b+c)
		1 0 3	= (3+0) / (2*0+3+0) = 3/3 = 1
Kenya		0 0 6	

3.3. TARTIŞMA VE SONUÇ

15 ülkenin ölçümle belirtilen 9 değişkeni üzerinde uzaklık (dissimilarity) ve yakınlık (similarity) matrisleri hesaplanmıştır. Bu matrislerden dendrogramlar (ağaç diyagramları) çizilmiştir. Bu bulguların ölçüm biçimlerine göre yorumları aşağıda verilmiştir.

Aralık (Interval) Veriler

- ♦ Standart verilere Öklit ve Kare Öklit Uzaklık Ölçütleri uygulandığında, iki ölçütte de birbirine en yakın (benzer) iki ülke Öklit'te 0,7145 ve Kare Öklit'te ise 0,5106 değerleriyle Romanya ile Polonya, en uzak iki ülke ise Öklit'te 8,4254 ve Kare Öklit'te ise 70,9871 değerleriyle Çad ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.1. Öklit ve Kare Öklit Uzaklık Ölçütleri Kümeleme Sonucu

Küme No	Ülkeler
I	(Romanya, Polonya, Portekiz, Filipinler, İtalya, Avusturya)
II	(Sudan, Nepal, Gana, Kenya, Pakistan)
III	(Kamerun, Türkiye, Irak, Çad)

- ♦ Standart verilere Chebychev Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0,4318 değeriyle Romanya ile Polonya, en uzak iki ülke ise 3,4220 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.2. Chebychev Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(Romanya, Polonya, Filipinler, Portekiz)
II	(Kenya, Pakistan, Sudan, Nepal, Gana, Kamerun, İtalya, Avusturya)
III	(Türkiye, Irak, Çad)

- ♦ Standart verilere Manhattan City-Block Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 1,7443 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 24,6853 değeriyle Çad ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.3. Manhattan City-Block Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Romanya, Polonya, Portekiz, Filipinler, Türkiye, Irak)
II	(Gana, Sudan, Nepal, Kenya, Pakistan)
III	(Kamerun, Çad)

- ♦ Standart verilere Minkowski Uzaklık Ölçütü $\lambda = 3$ için uygulandığında, birbirine en yakın iki ülke 0,5389 değeriyle Romanya ile Polonya, en uzak iki ülke ise 5,9823 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.4. Minkowski Uzaklık Ölçütü Kümeleme Sonucu ($\lambda = 3$)

Küme No	Ülkeler
I	(Romanya, Polonya, Filipinler, Portekiz, İtalya, Avusturya)
II	(Sudan, Nepal, Gana, Kenya, Pakistan, Kamerun)
III	(Türkiye, Irak, Çad)

- ♦ Standart verilere Minkowski Uzaklık Ölçütü $\lambda = 4$ için uygulandığında, birbirine en yakın iki ülke 0,4791 değeriyle Romanya ile Polonya, en uzak iki ülke ise 5,0560 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.5. Minkowski Uzaklık Ölçütü Kümeleme Sonucu ($\lambda = 4$)

Küme No	Ülkeler
I	(Romanya, Polonya, Filipinler, Portekiz)
II	(Sudan, Nepal, Gana, Kenya, Pakistan, Kamerun)
III	(İtalya, Avusturya)
IV	(Türkiye, Irak, Çad)

- ♦ Orijinal verilere Canberra Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0,3794 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 5,4623 değeriyle Çad ile İtalya'dır.
- ♦ Orijinal verilere Kosinüs Benzerlik Ölçütü uygulandığında, birbirine en yakın iki ülke 0,9969 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 0,3022 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.6. Kosinüs Benzerlik Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Romanya, Polonya, Portekiz)
II	(Türkiye, Irak, Filipinler, Kenya)
III	(Pakistan, Nepal, Sudan, Gana, Kamerun, Çad)

- ♦ Orijinal verilere Pearson İlişki Katsayısı Benzerlik Ölçütü uygulandığında, birbirine en yakın iki ülke 0,9902 değeriyle İtalya ile Avusturya, en uzak iki ülke ise -0,5775 değeriyle Çad ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.7. Pearson Benzerlik Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Romanya, Polonya, Portekiz, Filipinler)
II	(Türkiye, Irak, Kenya)
III	(Pakistan, Nepal, Gana, Sudan, Kamerun, Çad)

Aralık verilerde kullanılan uzaklık ve benzerlik ölçütlerinden elde edilen sonuçlara bakıldığında, ülkeler genelde 3 kümede toplanmaktadır. Bu kümelerde Filipinler Asya ülkesi olmasına rağmen Avrupa ülkeleri içerisinde yer almaktadır. Avrupa ülkeleri genelde aynı küme içerisinde yer almakta, ancak Afrika ve Asya ülkeleri birbirine karışmaktadır. Bu tip verilerde, Öklit, Kare Öklit, Manhattan City-Block, $\lambda = 3$ için Minkowski uzaklık ölçütleri ile Kosinüs ve Pearson İlişki Katsayısı benzerlik ölçütleri

kullanıldığı zaman iyi sonuç alınmaktadır. Bu nedenle bu tip verilerde bu ölçütlerden herhangi birisi kullanılabilir.

♦ Afrika, Asya ve Avrupa ülkelerini 3 kümede toplayarak kümeler arasındaki uzaklığa, orijinal veriler üzerinden; Mahalanobis Uzaklık Ölçütü ile bakıldığında, birbirine en yakın iki kıta 5,6237 değeriyle Afrika ile Asya, en uzak iki kıta ise 8,1084 değeriyle Afrika ile Avrupa'dır.

Aynı kümeler arasındaki uzaklığa Hotelling T² Uzaklık Ölçütü ile bakıldığında, en yakın iki kıta 79,0723 değeriyle Afrika ile Asya, en uzak iki kıta ise 164,3645 değeriyle Afrika ile Avrupa'dır.

Aynı kümeler arasındaki uzaklığa Karl Pearson (Kümeler Arası) Uzaklık Ölçütü ile bakıldığında, en yakın iki kıta 2,0633 değeriyle Afrika ile Asya, en uzak iki kıta ise 8,9644 değeriyle Afrika ile Avrupa'dır.

Aynı kümeler arasındaki uzaklığa Penrose Uzaklık Ölçütü ile bakıldığında, en yakın iki kıta 2,1844 değeriyle Afrika ile Asya, en uzak iki kıta ise 12,4108 değeriyle Afrika ile Avrupa'dır.

Kümeler arasındaki uzaklığa bakan bütün uzaklık ölçütlerinde, seçilen değişkenler bakımından Afrika ile Asya birbirine yakın (benzer), Afrika ile Avrupa ise birbirinden uzak (benzemez) sonucu çıkmıştır. Aralık veriler için kümeler arasındaki uzaklığa bakılacak ise yukarıdaki 4 ölçütten herhangi birisi kullanılabilir.

Nitel (Qualitative) Veriler

Bu veriler, ölçümle belirtilen değişken değerlerinden yapay olarak elde edilmiştir. Bu tip verilere uygun uzaklık ölçütlerinden uzaklık matrisleri hesaplanmıştır. Bu verilere;

- ♦ Ölçekli Öklit Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0,3333 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 2,7641 değeriyle Çad ile Portekiz'dir.
- ♦ Öklit Uzaklık Ölçütü uygulandığında, en yakın iki ülke 1,0000 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 11,4018 değeriyle Çad ile Portekiz'dir.
- ♦ Mahalanobis-Benzeri Uzaklık Ölçütü uygulandığında, en yakın iki ülke 0,5345 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 6,5378 değeriyle Çad ile Portekiz'dir.

- ♦ Bhattacharyya Uzaklık Ölçütü uygulandığında, en yakın iki ülke 0,2679 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 3,5026 değeriyle Çad ile Portekiz'dir.

Nitel verilerde kullanılan uzaklık ölçütlerinin hepsi aynı sonucu vermiştir. Ancak, bu tip veriler için özel olarak geliştirilmiş olan Bhattacharyya Uzaklık Ölçütü'nün kullanılması tavsiye edilmektedir.

Frekans Sayıları

Bu veriler, ölçümle belirtilen verilerden Toplam Doğurganlık Oranı (TDO) değişken değerleri çıkartılarak elde edilmiştir. Bu verilere;

- ♦ Ki-Kare Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 1,4237 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 21,9110 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.8. Ki-Kare Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Polonya, Portekiz, Romanya)
II	(Sudan, Pakistan, Gana, Nepal, Kamerun)
III	(Türkiye, Irak, Filipinler, Kenya, Çad)

- ♦ Fi-Kare Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0,0452 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 0,6939 değeriyle Çad ile Portekiz'dir. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.9. Fi-Kare Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Romanya, Polonya, Portekiz)
II	(Sudan, Pakistan, Gana, Nepal, Kamerun)
III	(Türkiye, Irak, Filipinler, Kenya, Çad)

Frekans sayılarında kullanılan uzaklık ölçütlerinde sonuçlar birbirinin aynı çıktığından bu tip verilerde, iki ölçütten herhangi birisi kullanılabilir.

İkili Veriler

Bu veriler, aynı verilerden iyi durum için 1 değeri, kötü durum için 0 değeri verilmek suretiyle yapay olarak üretilmiştir. Bu verilere;

- ♦ Kare Öklit ve Öklit Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0 değeriyle İtalya ile Avusturya, en uzak ülkeler ise Kare Öklit'te 9 ve Öklit'te ise 3 değerleriyle Çad ile Polonya, Gana ile Polonya, Sudan ile Polonya, Kamerun ile Polonya ve Nepal ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.10. Kare Öklit ve Öklit Uzaklık Ölçütleri Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Polonya, Romanya, Portekiz, Filipinler)
II	(Irak, Türkiye)
III	(Kamerun, Nepal, Sudan, Gana, Çad, Pakistan, Kenya)

- ♦ Büyüklük Farkları Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0 değeriyle İtalya ile Avusturya, en uzak ülkeler ise 1 değeriyle Çad ile Polonya, Gana ile Polonya, Sudan ile Polonya, Kamerun ile Polonya ve Nepal ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.11. Büyüklük Farkları Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Portekiz, Romanya, Polonya, Filipinler, Irak)
II	(Kenya, Türkiye)
III	(Kamerun, Nepal, Sudan, Gana, Çad, Pakistan)

- ♦ Biçim Farkları Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0 değeriyle İtalya ile Avusturya, en uzak iki ülke ise 0,1111 değeriyle Kenya ile Türkiye'dir. Dendrogramdan aşağıdaki kümenin oluştuğu gözlenmektedir.

Tablo 3.3.12. Biçim Farkları Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Polonya, Portekiz, Romanya, Filipinler, Nepal, Türkiye, Irak, Pakistan, Kenya, Kamerun, Sudan, Gana, Çad)

- ♦ Değişim Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0 değeriyle İtalya ile Avusturya, en uzak ülkeler ise 0,2500 değeriyle Çad ile Polonya, Gana ile Polonya, Sudan ile Polonya, Kamerun ile Polonya ve Nepal ile Polonya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.13. Değişim Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya, Polonya, Romanya, Portekiz, Filipinler)
II	(Irak, Türkiye)
III	(Kamerun, Nepal, Sudan, Gana, Çad, Pakistan, Kenya)

- ♦ Durum Uzaklık Ölçütü uygulandığında, birbirine en yakın iki ülke 0 değeriyle İtalya ile Avusturya, en uzak ülkeler ise 0,6667 değeriyle Türkiye ile Kenya'dır. Dendrogramdan aşağıdaki kümelerin oluştuğu gözlenmektedir.

Tablo 3.3.14. Durum Uzaklık Ölçütü Kümeleme Sonucu

Küme No	Ülkeler
I	(İtalya, Avusturya)
II	(Nepal, Polonya, Kamerun, Sudan, Gana, Çad)
III	(Romanya, Portekiz, Filipinler)
IV	(Pakistan, Kenya, Türkiye, Irak)

Bu tip verilerde kullanılan uzaklık ölçütlerinin tümünde birbirine en çok benzeyen iki ülke İtalya ile Avusturya olmasına rağmen en uzak (benzemeyen) ülkeler değişiklik göstermektedir. Kare Öklit, Öklit, Durum Farkları ve Dağılım uzaklık ölçütleri iyi bir sonuç vermiştir. Yani ikili verilerde yukarıda saydığımız 4 ölçütten herhangi birisi kullanılabilir.

İkili verilerde kullanılan diğer ölçütlerden elde edilen matris elemanlarında belirsizlikler (0/0 veya sayı/0 gibi) olduğundan dendrogramlar çizilememiş ve aynı zamanda matris elemanlarında birden fazla en küçük ve en büyük değer bulunduğundan, bu veriler için bir sonuç çıkmamıştır.

KAYNAKLAR

- 1- Manly, Bryan F.J. **Multivariate Statistical Methods. A Primer**, Department of Mathematics and Statistics University of Otago, New Zealand, 1990.
- 2- Mardia, K.V., Kent, J.T. and Bibby, J.M. **Multivariate Analysis**. Academic Press Inc. (London), 1979.
- 3- Zupan, Jure. **Clustering of Large Data Sets**. Research Studies Press, p: 2-11, 1982.
- 4- Jobson, J.D. **Applied Multivariate Data Analysis**. Volume II: Categorical and Multivariate Methods, USA, 1992.
- 5- Tatlıdil, Hüseyin. **Çok Değişkenli İstatistiksel Analiz**. Ankara, 1992.
- 6- Anderberg, Michael R. **Cluster Analysis for Application**. London, 1973.
- 7- Everitt, Brian. **Cluster Analysis**. Heineman Educational, London, 1974.
- 8- Mardia, K.V. **Mahalanobis Distance and Angles In "Multivariate Analysis-IV"**. Krishnaiah, P.R. (ed.), North Holland, Amsterdam, 1979.
- 9- Ryzin, J.Van. **Classification and Clustering**. Publication No. 37 of the Mathematics Research Center the University of Wisconsin at Madison, Academic Press, Inc. London, 1973.
- 10- Hartigan, John A. **Clustering Algorithms**. Department of Statistics Yale University, 1975.
- 11- Duran, Benjamin S. and Odell, Patrick L. **Cluster Analysis**. Springer-Verlag Berlin, Heidelberg New York, 1974.
- 12- Jambu, M. and Lebeaux, M-O. **Cluster Analysis and Data Analysis**. North-Holland Publishing Company, 1983.
- 13- Atamer, Belgin. **Kümeleme Analizi ve Kümeleme Analizinin İlaç Sektörüne Uygulanması**. Yayınlanmamış Yüksek Lisans Tezi, İstanbul, 1992.
- 14- **SPSS for WINDOWS: Professional Statistics**, Printed in the United States of America. SPSS Inc., 1993.
- 15- **Dünya Çocuklarının Durumu**. UNICEF Türkiye Temsilciliği, UNICEF House, Ankara, 1995.
- 16- Sümbüloğlu, Kadir. **Sağlık Alanına Özel İstatistiksel Yöntemler**. Ankara, 1994.

EK 1: UYGULAMADA KULLANILAN DEĞİŞKEN DEĞERLERİ İLE BU DEĞİŞKENLERE İLİŞKİN TOPLAM, ORTALAMA, STANDART SAPMA VE Z-SKORU DEĞERLERİ

KITA	ÜLKE	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
	Çad	206	121	48	34	13	13	19	5,9	35
	Gana	170	103	56	70	48	47	50	5,9	35
AFRİKA	Sudan	128	77	52	61	51	51	49	6,0	24
	Kamerun	113	71	56	41	33	33	33	5,7	43
	Kenya	90	61	59	95	85	85	76	6,2	26
	Pakistan	137	95	59	87	74	74	71	6,1	34
	Nepal	128	90	54	73	64	64	59	5,4	13
ASYA	Türkiye	84	67	67	63	79	79	74	3,4	66
	Irak	71	57	66	79	82	82	81	5,7	74
	Filipinler	59	45	65	90	88	89	87	3,9	45
	Romanya	29	23	70	99	98	92	91	2,1	55
	Polonya	15	13	72	94	98	98	96	2,1	63
AVRUPA	Portekiz	11	9	75	92	94	93	99	1,5	35
	İtalya	9	7	77	96	95	98	50	1,3	70
	Avusturya	8	7	76	97	90	90	60	1,5	60
Toplam		1258	846	952	1171	1092	1088	995	62,7	678
Ortalama		83,87	56,40	63,47	78,07	72,80	72,53	66,33	4,18	45,20
Standart Sapma		62,76	37,79	9,34	27,17	25,83	25,71	23,38	1,98	18,53
Z-SKORU DEĞERLERİ										
	Çad	1,9460	1,7095	-1,6557	-1,6220	-2,3152	-2,3154	-2,0247	0,8670	-0,5505
	Gana	1,3724	1,2332	-0,7993	-0,2969	-0,9601	-0,9930	-0,6987	0,8670	-0,5505
AFRİKA	Sudan	0,7032	0,5451	-1,2275	-0,6282	-0,8440	-0,8375	-0,7414	0,9174	-1,1442
	Kamerun	0,4642	0,3864	-0,7993	-1,3644	-1,5409	-1,5375	-1,4259	0,7662	-0,1187
	Kenya	0,0977	0,1217	-0,4781	0,6233	0,4723	0,4849	0,4135	1,0182	-1,0362
	Pakistan	0,8466	1,0215	-0,4781	0,3288	0,0465	0,0570	0,1996	0,9678	-0,6045
	Nepal	0,7032	0,8892	-1,0134	-0,1865	-0,3407	-0,3319	-0,3137	0,6149	-1,7378
ASYA	Türkiye	0,0021	0,2805	0,3782	-0,5546	0,2400	0,2515	0,3279	-0,3932	1,1226
	Irak	-0,2050	0,0159	0,2712	0,0344	0,3562	0,3682	0,6274	0,7662	1,5543
	Filipinler	-0,3962	-0,3017	0,1641	0,4393	0,5885	0,6404	0,8840	-0,1411	-0,0108
	Romanya	-0,8742	-0,8839	0,6994	0,7705	0,9756	0,7571	1,0551	-1,0484	0,5289
	Polonya	-1,0973	-1,1485	0,9135	0,5865	0,9756	0,9904	1,2690	-1,0484	0,9607
AVRUPA	Portekiz	-1,1610	-1,2543	1,2346	0,5129	0,8208	0,7960	1,3973	-1,3509	-0,5505
	İtalya	-1,1929	-1,3073	1,4487	0,6601	0,8595	0,9904	-0,6987	-1,4517	1,3385
	Avusturya	-1,2088	-1,3073	1,3417	0,6969	0,6659	0,6793	-0,2709	-1,3509	0,7988

EK 2: AFRIKA, ASYA VE AVRUPA ÜLKELERİNİN KOVARYANS MATRİSLERİ

KOVARYANS MATRİSİ (AFRİKA)									
	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
BYKÖO	2154,800	1144,700	-154,350	-666,600	-944,250	-951,400	-710,700	-2,770	95,700
BÖO	1144,700	610,800	-78,400	-334,900	-486,500	-490,600	-364,300	-1,380	52,800
DYB	-154,350	-78,400	18,200	77,700	90,250	89,800	72,150	0,215	-1,150
BCG	-666,600	-334,900	77,700	590,700	623,750	621,300	510,650	3,590	-118,900
DBT	-944,250	-486,500	90,250	623,750	702,000	701,500	563,250	3,700	-127,500
POLİO	-951,400	-490,600	89,800	621,300	701,500	701,200	701,200	3,710	-128,100
KIZAMIK	-710,700	-364,300	72,150	510,650	563,250	562,100	455,300	3,005	-103,550
TDO	-2,770	-1,380	0,215	3,590	3,700	3,710	3,005	0,033	-1,230
KNO	95,700	52,800	-1,150	-118,900	-127,500	-128,100	-103,550	-1,230	59,300
KOVARYANS MATRİSİ (ASYA)									
	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
BYKÖO	1210,700	737,950	-162,450	-19,900	-273,650	-282,850	-314,650	25,050	-612,650
BÖO	737,950	456,200	-94,450	-36,150	-170,650	-177,100	-198,150	14,775	-340,150
DYB	-162,450	-94,450	30,700	-5,600	43,900	44,600	48,900	-3,725	127,150
BCG	-19,900	-36,150	-5,600	118,800	36,050	38,950	52,550	4,900	-56,950
DBT	-273,650	-170,650	43,900	36,050	81,800	84,450	95,300	-5,025	158,300
POLİO	-282,850	-177,100	44,600	38,950	84,450	87,300	98,450	-5,275	157,950
KIZAMIK	-314,650	-198,150	48,900	52,550	95,300	98,450	112,800	-4,625	178,300
TDO	25,050	14,775	-3,725	4,900	-5,025	-5,275	-4,625	1,395	-9,375
KNO	-612,650	-340,150	127,150	-56,950	158,300	157,950	178,300	-9,375	604,300
KOVARYANS MATRİSİ (AVRUPA)									
	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
BYKÖO	73,800	57,600	-23,000	12,450	20,250	-4,850	98,900	2,550	-10,050
BÖO	57,600	45,200	-18,500	9,400	16,000	-3,700	82,300	2,100	-7,600
DYB	-23,000	-18,500	8,500	-2,500	-7,250	0,750	-46,750	-1,050	4,750
BCG	12,450	9,400	-2,500	7,300	0,500	-3,400	-24,150	0,250	18,050
DBT	20,250	16,000	-7,250	0,500	11,000	6,750	40,500	0,900	4,750
POLİO	-4,850	-3,700	0,750	-3,400	6,750	13,200	-4,050	0,050	22,600
KIZAMIK	98,900	82,300	-46,750	-24,150	40,500	-4,050	508,700	5,750	-198,900
TDO	2,550	2,100	-1,050	0,250	0,900	0,050	5,750	0,140	0,050
KNO	-10,050	-7,600	4,750	18,050	4,750	22,600	-198,900	0,050	175,300

**EK 3: AFRIKA, ASYA VE AVRUPA ÜLKELERİNİN ORTAK
KOVARYANS MATRİSİ İLE BU MATRİSİN TERSİ**

KOVARYANS MATRİSİ									
	BYKÖÖ	BÖÖ	DYB	BCG	DBT	POLİÖ	KIZAMİK	TDO	KNO
BYKÖÖ	1146,433	646,750	-113,267	-224,683	-399,217	-413,033	-308,817	8,277	-175,667
BÖÖ	646,750	370,733	-63,783	-120,550	-213,717	-223,800	-160,050	5,165	-98,317
DYB	-113,267	-63,783	19,133	23,200	42,300	45,050	24,767	-1,520	43,583
BCG	-224,683	-120,550	23,200	238,933	220,100	218,950	179,683	2,913	-52,600
DBT	-399,217	-213,717	42,300	220,100	264,933	264,233	233,017	-0,142	11,850
POLİÖ	-413,033	-223,800	45,050	218,950	264,233	267,233	265,200	-0,505	17,483
KIZAMİK	-308,817	-160,050	24,767	179,683	233,017	218,833	358,933	1,377	-41,383
TDO	8,277	5,165	-1,520	2,913	-0,142	-0,505	1,377	0,523	-3,518
KNO	-175,667	-98,317	43,583	-52,600	11,850	17,483	-41,383	-3,518	279,633
KOVARYANS MATRİSİNİN TERSİ									
	BYKÖÖ	BÖÖ	DYB	BCG	DBT	POLİÖ	KIZAMİK	TDO	KNO
BYKÖÖ	0,1595	-0,2504	0,0717	-0,0618	0,1326	-0,0620	0,0090	0,4144	-0,0058
BÖÖ	-0,2428	0,3916	-0,0871	0,0829	-0,1897	0,0996	-0,0171	-0,6106	0,0059
DYB	0,0534	-0,0653	0,3308	-0,0366	0,0549	-0,0610	0,0152	0,6378	-0,0361
BCG	-0,0884	0,1300	-0,0732	0,0816	-0,1241	0,0352	0,0042	-0,4638	0,0148
DBT	0,0824	-0,1353	0,0680	-0,0045	0,1805	-0,2120	0,0482	0,0098	0,0056
POLİÖ	0,0357	-0,0372	-0,0131	-0,0907	-0,0102	0,1638	-0,0578	0,4533	-0,0183
KIZAMİK	-0,0090	0,0140	-0,0077	0,0068	-0,0277	0,0176	0,0037	-0,0442	0,0018
TDO	0,5308	-0,8289	0,8268	-0,4216	0,6122	-0,2522	-0,0040	5,6715	-0,1055
KNO	-0,0105	0,0147	-0,0437	0,0130	-0,0187	0,0104	0,0001	-0,1049	0,0103

EK 4: UYGULAMADA KULLANILAN DEĞİŞKENLERE İLİŞKİN NİTELİK VE İKİLİ VERİLER

NİTELİK VERİLER										
KITA	ÜLKE	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
	Çad	6	5	1	1	1	1	1	5	2
	Gana	6	4	1	3	2	2	2	5	2
AFRİKA	Sudan	5	3	1	2	3	2	2	6	1
	Kamerun	4	3	1	1	2	1	2	5	2
	Kenya	4	2	2	4	4	4	3	6	1
	Pakistan	5	4	2	4	4	3	3	6	2
	Nepal	5	4	1	3	3	3	2	5	1
ASYA	Türkiye	3	3	3	2	4	3	3	3	3
	Irak	3	2	3	4	4	4	4	5	4
	Filipinler	2	2	3	5	4	4	4	3	2
	Romanya	2	2	4	5	5	4	5	2	2
	Polonya	1	1	4	5	5	5	5	2	3
AVRUPA	Portekiz	1	1	4	5	5	5	5	1	2
	İtalya	1	1	4	5	5	5	2	1	4
	Avusturya	1	1	4	5	5	5	2	1	3
İKİLİ VERİLER										
KITA	ÜLKE	BYKÖO	BÖO	DYB	BCG	DBT	POLİO	KIZAMIK	TDO	KNO
	Çad	0	0	0	0	0	0	0	0	0
	Gana	0	0	0	0	0	0	0	0	0
AFRİKA	Sudan	0	0	0	0	0	0	0	0	0
	Kamerun	0	0	0	0	0	0	0	0	0
	Kenya	0	0	0	1	1	1	0	0	0
	Pakistan	0	0	0	1	0	0	0	0	0
	Nepal	0	0	0	0	0	0	0	0	0
ASYA	Türkiye	1	0	1	0	0	0	0	0	1
	Irak	1	0	1	0	1	1	1	0	1
	Filipinler	1	1	1	1	1	1	1	0	0
	Romanya	1	1	1	1	1	1	1	1	0
	Polonya	1	1	1	1	1	1	1	1	1
AVRUPA	Portekiz	1	1	1	1	1	1	1	1	0
	İtalya	1	1	1	1	1	1	0	1	1
	Avusturya	1	1	1	1	1	1	0	1	1