

# KÜMELEME ANALİZİ (Cluster Analysis)

Kümeleme analizi bir veri setinin farklı gruplar içerip içermediğini belirlemek ve eğer içeriyorsa bu grupları tespit etmek için kullanılan çok değişkenli istatistiksel bir yöntemdir.

Çok boyutlu uzayda verilerin özetlenmesi ve tanımlanmasında yol gösterici bir araştırma yöntemi olan kümeleme analizi; heterojen olan farklı gruplardaki gözlem yapılarını ya da homojen olan benzer gruplardaki gözlemleri uygun yöntemlerle gruplamayı sağlayan bir yöntemdir.

Diğer çok değişkenli istatistiksel yöntemlerde önemli bir yer tutan normallik, doğrusallık ve homojenlik varsayımları bu yöntemde prensipte kalmakta ve uzaklık değerlerinin normalliği yeterli görülmektedir (Çelik ve ark., 2004).

Kümeleme analizinin genel amacı, gruplanmamış verileri benzerliklerine göre sınıflandırmak (gruplamak) ve araştırmacıya uygun işe yarar özetleyici bilgiler elde etmede yardımcı olmaktadır (Tatlıdil, 1996).

**Kümeleme analizi gözlemler arası kümelendirme, değişkenler arası sınıflandırma ya da gözlemlerin ve değişkenlerin bir arada sınıflandırılmasını amaçlar.**

**Kesikli değişkenlerinde olduğu anket yapılarında kümeleme yöntemlerinin uygulanmasında bir sınırlama olmaması Kümeleme Analizinin kullanılmasını kolaylaştırır.**

Kümeleme analizinde değişkenler gösterdikleri özellikler kullanılarak karşılaştırılır. Çünkü Kümeleme Analizi sadece nesnelere tanımlanan özelliklerini kapsar.

Kümeleme Analizi ile araştırmacının belirlediği kriterlere göre tüm gözlemleri kümelendirerek veriler azaltılabilir veya özet bilgiler veren üst gruplar oluşturulabilir (Kalaycı ve ark., 2005).

Kümeleme analizi aşağıda belirtilen amaçlar için yapılabilir (Özdamar, 2002).

- n sayıda birimi, nesneyi, oluşumu, p değişkene göre saptanan özelliklerine göre olabildiğince kendi içinde türdeş (homojen) ve kendi aralarında farklı (heterojen) alt gruplara ayırmak
- p sayıda değişkeni, n sayıda birimde saptanan değerlere göre ortak özellikleri açıkladığı varsayılan alt kümelere ayırmak ve ortak faktör yapıları ortaya koymak.
- Hem birimleri hem de değişkenleri birlikte ele alarak ortak n birimi p değişkene göre ortak özellikli alt kümelere ayırmak.
- Birimleri p değişkene göre saptanan değerlere göre, izledikleri biyolojik ve tipolojik sınıflamayı ortaya koymak ( taksonomik sınıflandırma yapmak ).

Kümeleme analizi çok değişkenli varyans analizi, lojistik regresyon analizi, çok boyutlu ölçekleme gibi diğer çok değişkenli analizlerle de sıkı ilişkisi olan bir tekniktir.

## Kümeleme Analizi İle Diğer Bazı Çok Değişkenli Yöntemler Arasındaki İlişki

- Kümeleme Analizinin Ayırma (Diskriminant) Analizinde farkı, **Ayırma Analizinde** gruplar önceden belirlenirken **Kümeleme Analizinde** bu belirleme analiz sonucunda elde edilmektedir.
- **Kümeleme Analizi Faktör Analizindeki gibi değişkenler bağımlı-bağımsız diye ikiye ayrılmaz.** Her iki yöntemde de birey yada nesnelere aralarındaki benzerlikler itibarıyla bir araya getirilir.
- Yakınlık matrisleri ve onların görüntülenmesini sağlayan Çok Boyutlu Ölçekleme ile bu özelliklere sahip olan kümeleme analizi arasındaki temel farklılık, **Çok Boyutlu Ölçekleme yakınlıkların uzaysal görüntülenmesini sağlarken, Kümeleme Analizi yakınlıkların ağaç (dendrogram) biçiminde görüntülenmesini sağlar.** Kümeleme Analizinde küçük benzemezliklerden bir anlam çıkarılabilmekte ancak geniş benzemezlikleri yorumlamak güçtür. **Çok Boyutlu Ölçekleme Analizinde ise geniş benzemezlikleri değerlendirme veya anlam çıkarma özelliğine sahiptir.**

## **Kümeleme analizi kullanıldığı genel amaçların dışında aşağıdaki özel amaçlar için de kullanılabilir.**

- Gerçek tiplerin (cinslerin-ırkların) belirtilmesi
  - Model uydurmanın kolaylaştırılması
  - Gruplar için ön tahmin
  - Hipotezlerin testi
  - Veri yapısının netleştirilmesi
  - Veri indirgemesi (veriler yerine kümelerin değerlendirilmesi)
  - Aykırı değerlerin (outliers) bulunması
- olarak özetlenebilir.

(Tatlídil, 1996).



Sürekli deęişkenler içeren veri matrisinde birimler arasındaki uzaklıklar için Öklid uzaklığı ya da karesel Öklid uzaklığı, deęişkenler arasındaki uzaklığın belirlenmesinde ise Pearson Korelasyon uzaklığının kullanılması gerekir.

**Kümeleme aşamasında ilk önce uzaklıklar matrisi elde edilir. Uzaklık ölçüleri doğrudan birim ya da deęişkenlerin kümelenmesinde kullanılabileceęi gibi birim ya da deęişkenler arasındaki benzerlik –farklılıkları hesaplamada da kullanılabilir.**

Veri matrisinde yer alan  $n$  birimin  $p$  deęişkene göre uzaklık matrisi  $D$ , deęişkenler arasındaki benzerlik ilişkisi matrisi  $R$  ile gösterilir. Birimlerin birbirlerine olan benzerlik düzeyleri Benzerlik matrisi gösterilir. Benzerlik matrisinin elemanları  $D$  matrisinin elemanlarına göre belirlenir.

## KÜMELEME ANALİZİNDE KULLANILAN BENZERLİK VE UZAKLIK ÖLÇÜTLERİ

Birimler hakkında düzenli bilgiler verilmesi için birimleri gruplamak ya da sınıflandırmak gerekmektedir. Sınıflandırma çoğu özelliği yakınlık veya benzerlik gösteren birimleri gruplandırmaktır. İki değişkene göre incelenen ve özellikleri farklı olan birimler, değişken sayısı arttırıldığında birbirlerine benzedikleri ve çoğu benzer özelliklerinden dolayı bir sınıfta yer alabilirler.

**Benzerlik gösteren birimleri bir arada toplamak ve bu grubun ortak özelliklerini tanımlamak ve incelemek, değişkenler arasındaki neden-sonuç ilişkilerini açıklamak için daha kolay olacaktır.**

Kümeleme analizinde "**Benzerlik**" en temel kavramdır. Analizde ele alınan yöntemlerin tümü benzerlik ölçütlerine dayanır. Bireyler arasındaki benzerliğin ölçülmesinin farklı yolları bulunmaktadır. **Benzerliği ölçmenin en önemli iki yolu, korelasyona dayalı ölçüler ve uzaklık ölçüleridir** (Junhagen, 2000).

# Uzaklık Ölçütleri

**Minkowski Uzaklığı:** En genel kullanılan uzaklıktır. Farklı  $\lambda$  değerleri için artıp azalmasına göre ağırlıkları değişir.  $\lambda \geq 1$  için bu uzaklık, aşağıdaki gibi bulunur.

$$d_{\lambda}(x_i, x_j) = \left[ \sum_{k=1}^p |x_{ik} - x_{jk}|^{\lambda} \right]^{\frac{1}{\lambda}}$$



**Öklit (Euclidean) Uzaklığı:** Minkowski uzaklığının  $\lambda=2$  için özel durumudur. En çok kullanılan uzaklık ölçütleri arasındadır. **Objeler birbirinden farklı ölçüm birimleriyle ölçülmüş değişkenlere göre kümelenecekse, değişkenler önce standart değerlere dönüştürülmeli ve sonra öklit uzaklıkları hesaplanmalıdır** (Oktay, 2001). n-boyutlu bir kitlede iki nokta arasındaki öklit uzaklığı aşağıdaki gibi hesaplanır.

$$d(i, j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad \begin{array}{l} i=1,2,\dots,n \\ j=1,2,\dots,n \end{array}$$

**Karesel Öklit Uzaklığı:**

$$d(i, j) = \sum_{k=1}^p (X_{ik} - X_{jk})^2 \quad \begin{array}{l} n: \text{birim sayısı} \\ p: \text{değişken sayısı} \end{array}$$

**Örnek.** Üç birim ve iki değişkenden oluşan X veri matrisi aşağıdaki gibidir. Öklit ve karesel öklit uzaklığı matrislerini bulunuz?

$$X = \begin{bmatrix} 4 & 7 \\ 5 & 9 \\ 3 & 6 \end{bmatrix}$$

$$d(i, j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2}$$

1. ve 2. birim arasındaki uzaklık:  $d(1,2) = \sqrt{(4-5)^2 + (7-9)^2} = 2,236$

1. ve 3. birim arasındaki uzaklık:  $d(1,3) = \sqrt{(4-3)^2 + (7-6)^2} = 1,414$

2. ve 3. birim arasındaki uzaklık:  $d(2,3) = \sqrt{(5-3)^2 + (9-6)^2} = 3,606$

$$D_{\ddot{o}} = \begin{bmatrix} 0 & 2,236 & 1,414 \\ & 0 & 3,606 \\ & & 0 \end{bmatrix}$$

Karesel öklid uzaklığı:

$$d(i, j) = \sum_{k=1}^p (X_{ik} - X_{jk})^2$$

$$d(1,2) = (4 - 5)^2 + (7 - 9)^2 = 5$$

$$d(1,3) = (4 - 3)^2 + (7 - 6)^2 = 2$$

$$d(2,3) = (5 - 3)^2 + (9 - 6)^2 = 13$$

$$D_{\text{ö}}^2 = \begin{bmatrix} 0 & 5 & 2 \\ & 0 & 13 \\ & & 0 \end{bmatrix}$$

**Benzerlik Matrisi:** Öklid uzaklık matrisi kullanılarak benzerlik matrisi elde dlebilir.

$$X = \begin{bmatrix} 4 & 7 \\ 5 & 9 \\ 3 & 6 \end{bmatrix} \quad D_{\ddot{o}} = \begin{bmatrix} 0 & 2,236 & 1,414 \\ & 0 & 3,606 \\ & & 0 \end{bmatrix}$$

$$\text{Benz.}\ddot{o} = \left[ 1 - \frac{d(i, j)}{\text{Max}d(i, j)} \right] * 100$$

$$\text{Benz.}\ddot{o} = \begin{bmatrix} 100 & 37,99 & 60,79 \\ & 100 & 0 \\ & & 100 \end{bmatrix}$$

## Değişkenlerin Standardizasyonu (Dönüşüm)

Kümeleme analizinde değişkenlerden birisinin ortalaması ve varyansı çok büyük olduğunda diğer değişkenleri etkilemektedir. Aşırı değerlere sahip değişkenler de kümeleme üzerinde olumsuz etki yapmaktadır. Böyle durumlarda verilerin standardize edilmesi gerekir. Verilerin standardize yada belirli aralıklara dönüştürülmesi yöntemlerinden bazıları şunlardır:

### 1. Z skorlarına dönüştürme

Oransal ya da aralıklı ölçekle elde edilen ve normal dağılışı gösteren verilere uygulanan yöntemdir.

$$Z = \frac{X_i - \bar{X}}{S}$$



## Değişkenlerin Standardizasyonu (Dönüşüm)

### 2. $-1 \leq x \leq +1$ aralığına dönüştürme

Heterojen yapıda değerlerin ve aşırı uçlarda değerlerin yer aldığı durumlarda tercih edilir. Değerler arasında artı ve eksi değer olması gerekir.

$$X_i = \frac{X_i}{X_{\max}}$$

### 3. $0 \leq x \leq +1$ aralığına dönüştürme

Heterojen yapıda değerlerin ve aşırı uçlarda değerlerin yer aldığı durumlarda tercih edilir. Değerler pozitif olmalıdır.

$$X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} = \frac{X_i - X_{\min}}{\text{Range}}$$

## Değişkenlerin Standardizasyonu (Dönüşüm)

### 4. Maksimum değer 1 olan dönüştürme.

Veriler pozitif olmalıdır.

$$X_i = \frac{X_i}{X_{\max}}$$

### 5. Ortalaması 1 olan dönüştürme

$$X_i = \frac{X_i}{\bar{X}}$$

### 6. Standart sapması 1 olan dönüştürme.

$$X_i = \frac{X_i}{S}$$

**Kümeleme yöntemleri izledikleri yaklaşımlara göre iki temel gruba ayrılırlar. Bunlar, aşamalı (hiyerarşik) kümeleme yöntemleri ve aşamalı olmayan kümeleme yöntemleridir.**

Hiyerarşik kümeleme yaklaşımında ağaç diyagramından (dendogram) faydalanılır. Kümeleme sürecinin başlangıcında her bir birey bir kümedir (ağacın dalları). Süreç sonunda ise tüm bireyler bir kümede toplanır (ağacın gövdesi). Süreç aşağıdaki gibidir (Tatlıdil, 1996).

- i.  $n$  tane birey  $n$  tane küme olmak üzere işleme başlanır.
- ii. En yakın iki küme (dij farklılıkları en küçük olan) birleştirilir.
- iii. Küme sayısı bir indirilerek yinelenmiş uzaklıklar matrisi bulunur.
- iv. 2 ve 3 nolu adımlar  $n-1$  kez tekrarlanır.

## I. Hiyerarşik (Aşamalı) Kümeleme Tekniđi

Aşamalı kümeleme tekniđi, birimlerin (deđişkenlerin, gözlemlerin) benzerliklerini dikkate alarak belirli düzeylerde (küme uzaklık ölçüleri) birbiri ile birleştirmeyi amaçlayan tekniklerdir. Hiyerarşik kümeleme Birleştirici ve Ayırıcı Hiyerarşik Teknikler olmak üzere ikiye ayrılır (Özdamar, 2002).

## Hiyerarşik (Aşamalı) Kümeleme Yöntemleri

### 1. Birleştirici Aşamalı (Agglomerative) Kümeleme Yöntemleri

Başlangıçta tüm birimlerin ayrı birer küme oluşturduğunu kabul ederek  $n$  birimi aşamalı olarak  $n, n-1, n-2, \dots, 2, 1$  kümeye yerleştirmeyi amaçlar. Birimlerin birbirleri ile hangi aşamada ve hangi benzerlik düzeyinde ortak özelliklere sahip kümeler oluşturduğunu gösterir ve genelde sık tercih edilen bir yöntemdir.

**Bu yöntemde her birim başlangıçta tek başına birer küme kabul edilir. Daha sonra birbirleri ile yüksek derecede benzerlik gösteren iki birim bir küme oluşturur. Daha sonra bu kümeye benzerlik gösteren diğer birimler eklenerek birimlerin tümü bir kümede toplanacak biçimde birbirleri ile birleştirilirler ve bunlar dendrogram adlı ağaç grafikleri ile gösterilirler.**



## **Hiyerarşik (Ařamalı) Kúmeleme Yöntemleri**

### **2. Ayırıcı (Divise) Ařamalı Kúmeleme Yöntemleri**

Başlangıçta tüm birimlerin bir küme oluşturduğunu kabul ederek birimleri sırasıyla  $1, 2, \dots, n-1, n$  kümeye ayırmayı amaçlayan bir yaklaşımdır.

**Ward Yöntemi**: Kümeleme yayılımının küçültülmesine dayanan **Ward yöntemi kümeler içi varyansı en küçük yapmayı amaçlar**. Aşamalı kümeleme yöntemlerinden biri olan **Ward yöntemi birimlerin/değişkenlerin kümelenmesinde, varyansı minimuma indiren ve optimum küme sayısını tahmin eden bir yöntemdir**. “En küçük varyans” yöntemi olarak da adlandırılan ward yönteminin en önemli özelliği birey sayısı çok olduğunda ( $n > 50$ ), diğer aşamalı kümeleme yöntemlerine göre daha iyi sonuçlar vermesidir. (Cryer et al, 2001; Çelik ve ark, 2004).

Ward tekniği bir kümenin ortasına düşen gözlemin, aynı kümenin içinde bulunan gözlemlerden ortalama uzaklığı esas alır. Toplam sapma karelerinden yararlanır (Nakip, 2003).

**m ve j kümeleri arasındaki uzaklık**

$$d_{mj} = \frac{(N_j + N_k)d_{kj} + (N_j + N_l)d_{lj} - N_j d_{kl}}{(N_j + N_m)}$$

## II. Hiyerarşik Olmayan Kümeleme Tekniđi

**Küme sayısı konusunda ön bilgi var ise ya da arařtırmacı anlamlı olacak küme sayısına karar vermiřse bu durumda çok uzun zaman alan hiyerarşik teknikler yerine hiyerarşik olmayan teknikler tercih edilir.**

**Hiyerarjik olmayan kümeleme yönteminde önce küme sayısı belirlenir. Bu arařtırmacının ön bilgisine ve tecrübesine dayanılarak yapılır.** Analizde işlemlerin tekrarlanma sayısı (iteration numbers) ve yakınsama kriteri (convergence criterion) girilmelidir.

Tekrarlanma sayısı genelde 10 yakınsama kriteri de 0-1 arasında ve küçük olması önerilir. Oran azaldıkça gözlemlerin kümelere atanması daha güvenilir olmaktadır. Küme merkezleri ve deđişkenlerin her kümedeki ortalaması ANOVA ile bulunur.

## Hiyerarşik Olmayan Yöntemler

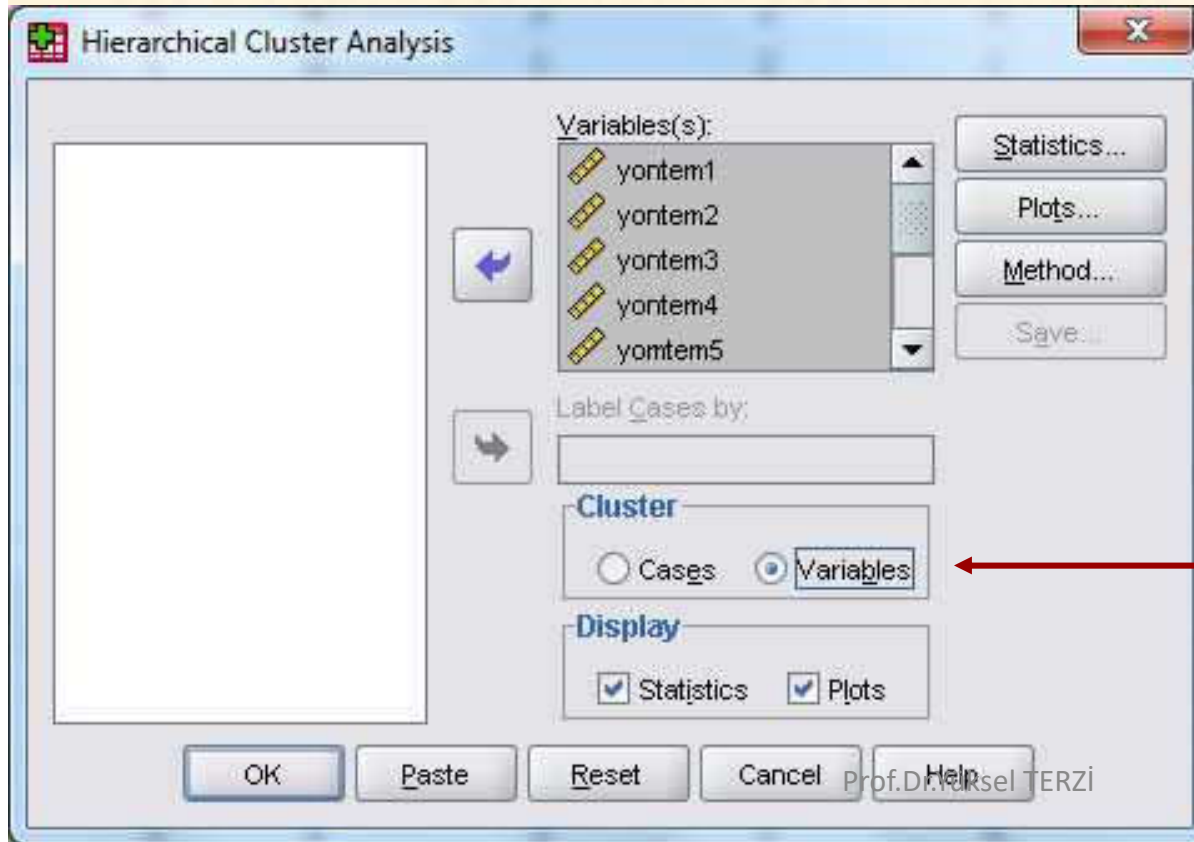
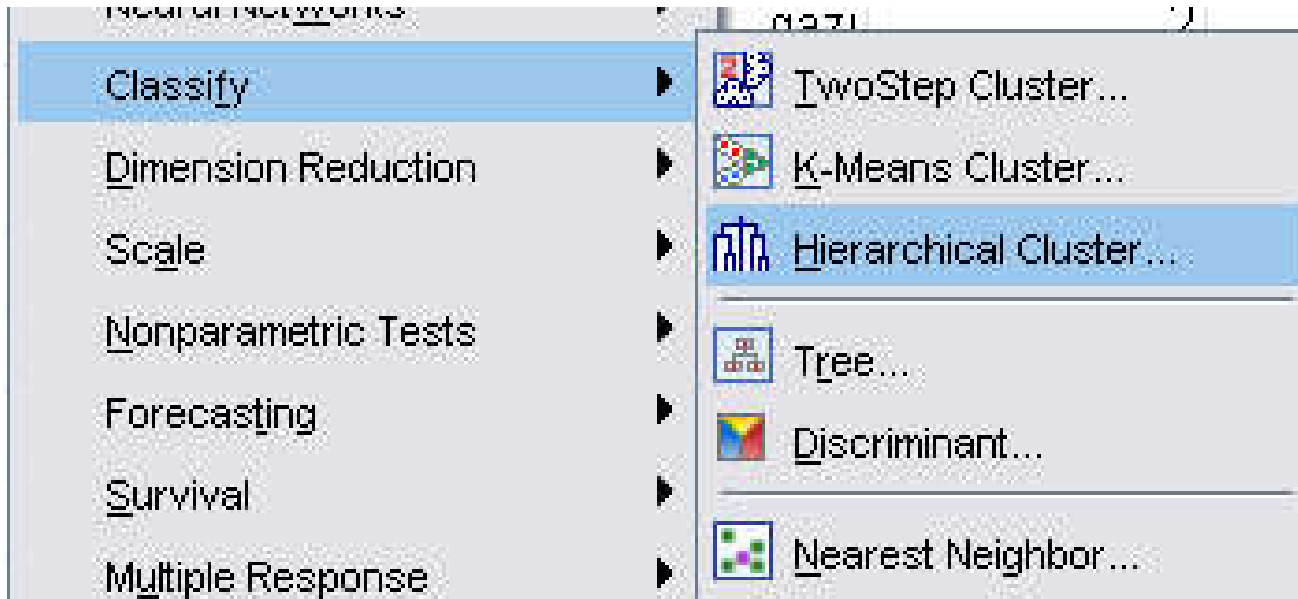
1. K-Ortalama (K-Means) Yöntemi
2. Metoid Parçalama Yöntemi
3. Yiğilma Yöntemi
4. Bulanık (Fuzzy) Kümeleme Yöntemi

Hiyararşik olmayan teknikler içinde en sık kullanılanı K-Ortalamalar (K-Means) tekniğidir (Dinçer, 1992).

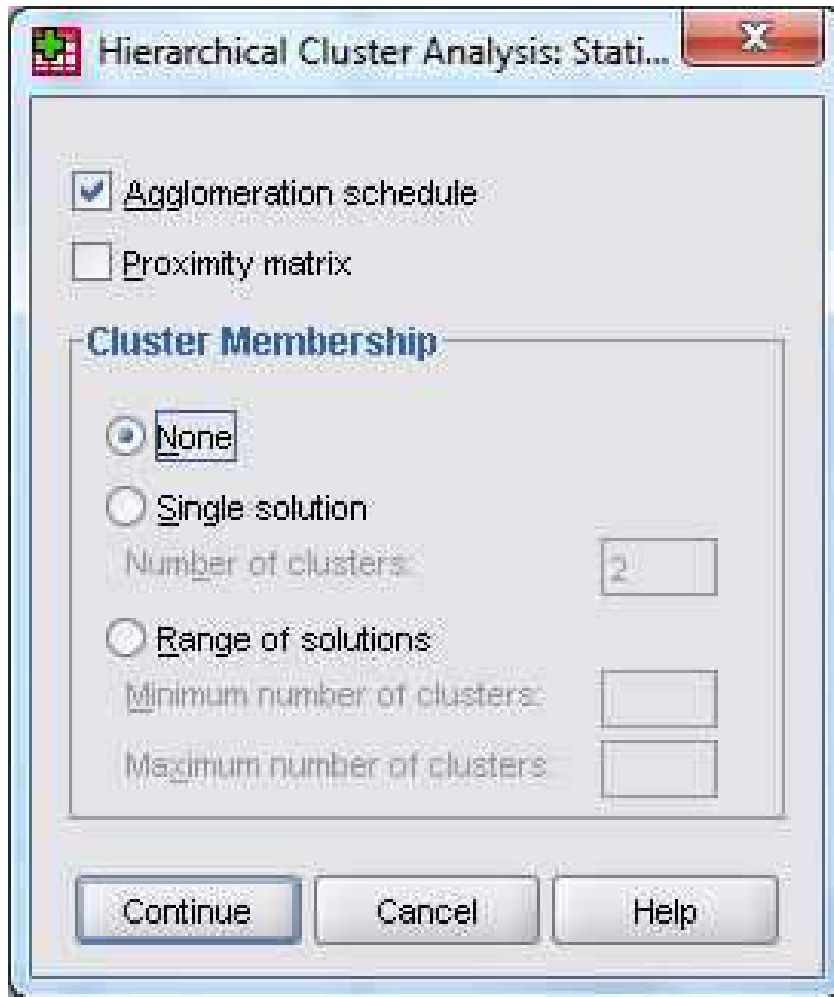
**Örnek:** Öğrencilere yöntemsel yeterlilik testinde 10 soru sorulmuş ve cevaplar 1-5 arası likert ölçekli olarak en olumsuzdan en olumluya doğru alınmıştır. Bu soru setindeki optimum küme sayısını aşamalı (hiyerarşik) kümeleme analizi ile bulalım.

yontem1	yontem2	yontem3	yontem4	yontem5	yontem6	yontem7	yontem8	yontem9	yontem10
2	3	4	2	5	3	2	2	4	5
3	4	4	3	3	3	4	3	3	3
4	4	4	4	3	3	3	2	2	3
4	2	4	4	2	4	4	2	2	3
3	3	2	2	3	1	3	1	2	1
2	2	3	3	3	2	1	1	1	2
2	2	2	2	3	3	1	1	2	2
2	3	3	2	1	2	3	3	1	1
2	5	4	5	5	4	3	3	4	5
1	1	2	2	2	1	2	1	1	2
2	2	2	1	2	1	1	1	3	2
1	2	2	2	2	1	2	1	2	3
3	4	3	2	4	3	4	2	3	3
2	1	3	2	3	2	1	1	2	2
3	3	2	3	3	3	2	2	1	1
1	2	1	1	2	2	2	2	3	1
1	2	2	2	1	1	3	1	3	1
1	2	1	1	1	3	1	1	1	3
3	4	4	5	3	3	3	3	1	5
4	4	5	4	4	3	3	3	3	4

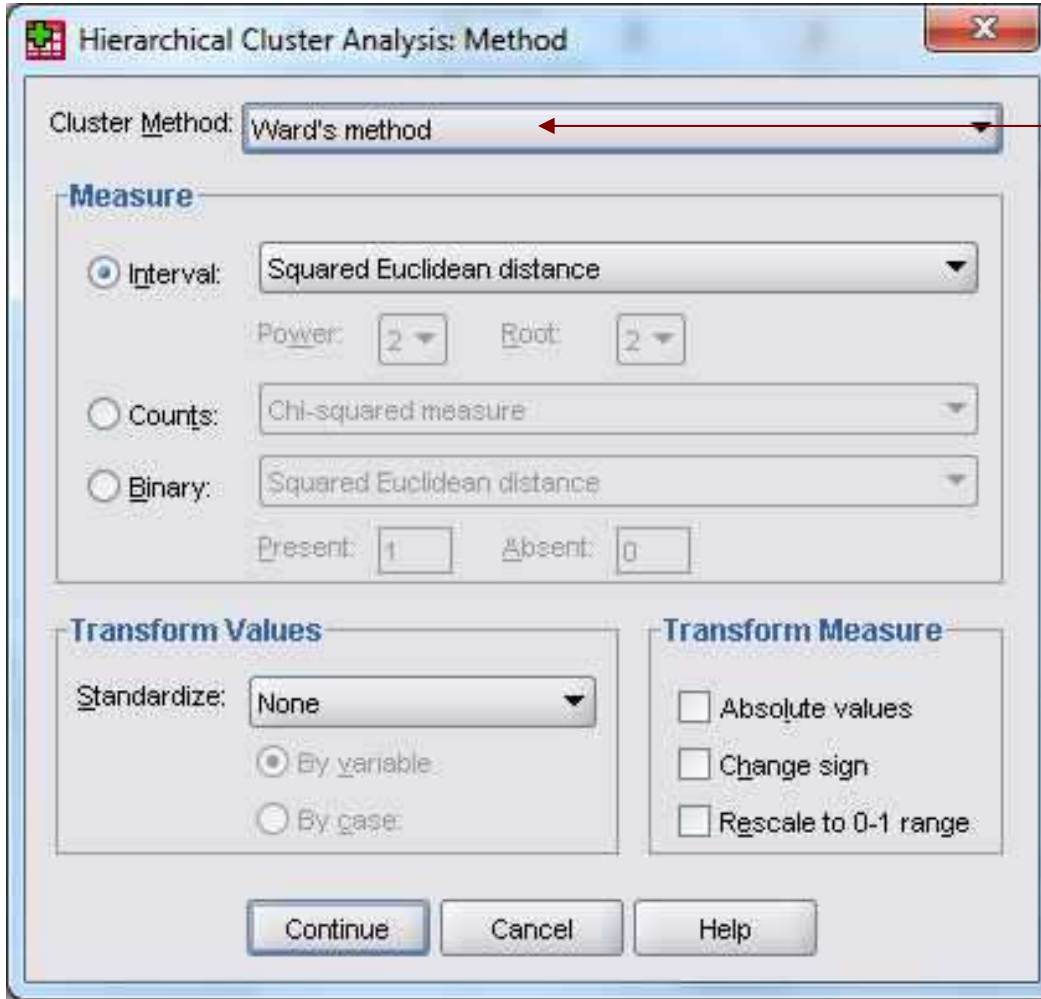




Cluster kısmına analize dahil edilecek olan bireyler (cases) değil, ölçek maddeleri olduğundan Variables işaretlenir.



**Cluster Membership:** Eğer kümelemeyi programın kendisi belirleyecekse None, sabit bir küme sayısı belirlenecekse Single solution, belli bir aralıkta küme sayısı belirlenecekse Range of solutions seçeneği seçilir.



Aşamalı kümeleme yöntemlerinden biri olan, birimlerin/değişkenlerin kümelenmesinde varyansı minimuma indiren ve optimum küme sayısını tahmin eden Ward yöntemi cluster Method olarak seçilir.

Uzaklık ölçüsü olarak karesel öklit uzaklığı seçilir.

## Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	2	3	139,500	0	0	3
2	6	7	294,000	0	0	4
3	1	2	451,833	0	1	5
4	5	6	633,333	0	2	6
5	1	4	840,250	3	0	6
6	1	5	1120,857	5	4	9
7	8	10	1431,357	0	0	8
8	8	9	1782,190	7	0	9
9	1	8	2317,600	6	8	0

**Stage:** Kümeleme analizinin ilk aşaması olup ölçeğin 9 ( $p-1=10-1$ ) kümeden oluştuğunu belirtir.

**Cluster Combined:** Birleştirilmiş kümeler.

**Coefficients:** Katsayılar

**Stage Cluster First Appears:** Kümelerin ilk görüldüğü aşamalar.

**Next Stage:** Sonraki aşama

## Cluster Membership

Case	2 Clusters
yontem1	1
yontem2	1
yontem3	1
yontem4	1
yontem5	1
yontem6	1
yontem7	1
yontem8	2
yontem9	2
yontem10	2

## Cluster Membership(Küme üyelikleri):

Bu tabloda ölçek maddelerinin hangi kümelerde yer aldığı belirtilir.

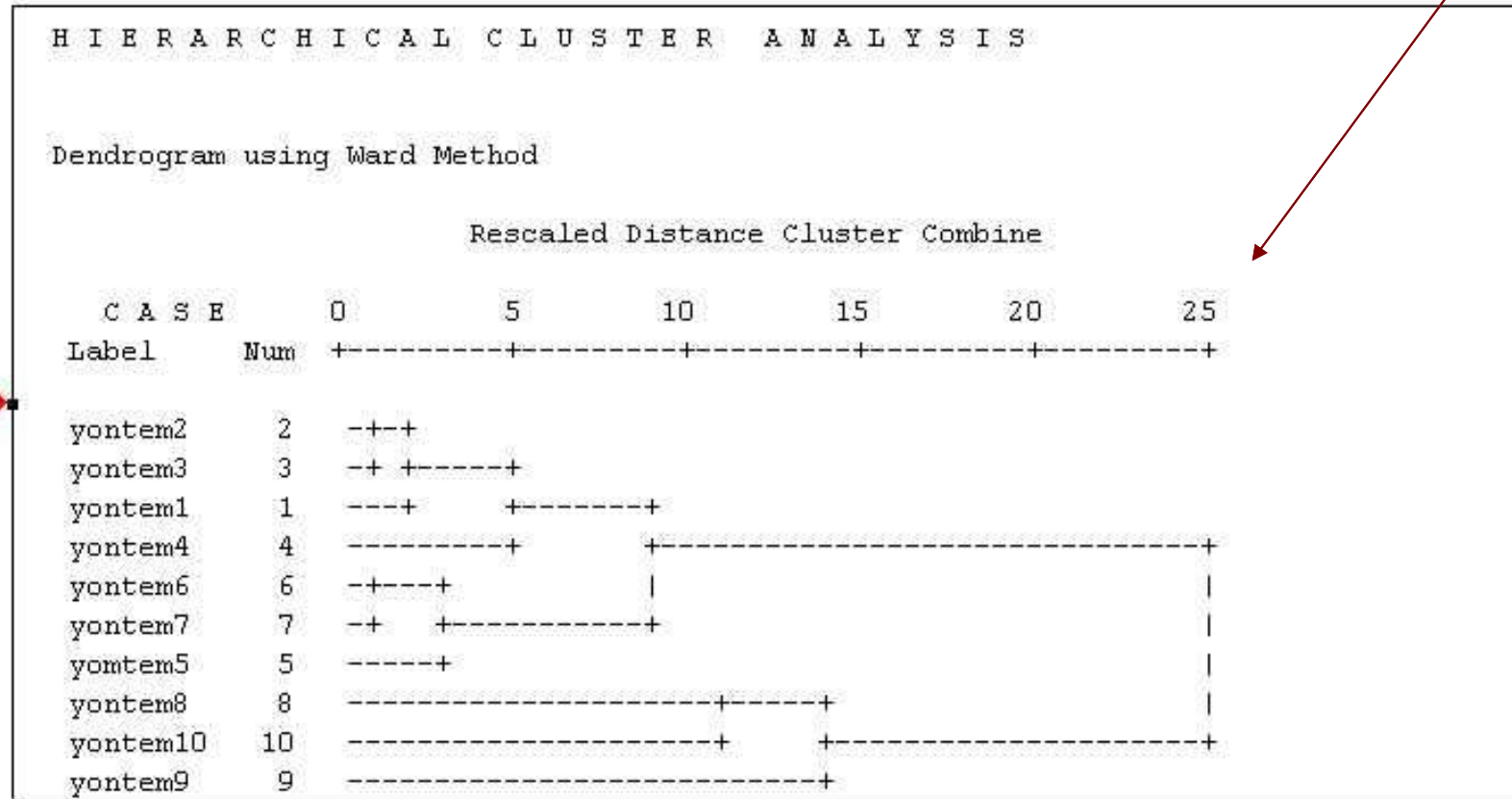
I.Küme 7 maddeden oluşur. Bunlar: Yöntem 1,2,3,4,5,6,7

II.Küme 3 maddeden oluşur. Bunlar: :Yöntem 8,9,10



## Dendrogram

Max uzaklık mik.



I.Küme :Yöntem 2,3,1,4,6,7,5

II.Küme:Yöntem 8,10,9

**Örnek:** Bir galeri sahibi müşterilerinin profillerini, müşterilerin meslekleri dolayısıyla gelir durumları ile otomobillere bakış açıları arasında bir fark olup olmadığını belirlemek istemiştir. Galeriye gelen ve rasgele seçilen 20 kişiye ait veriler aşağıdaki gibidir. Bu kişilere 7'li likert ölçeği ile hazırlanmış bir anket uygulanmıştır.

X1: Otomobillerle ilgilenmekten zevk alırım

X2: Otomobil almak bütçemi zorlar

X3: Günümüzde otomobil sahibi olmak bir gerekliliktir

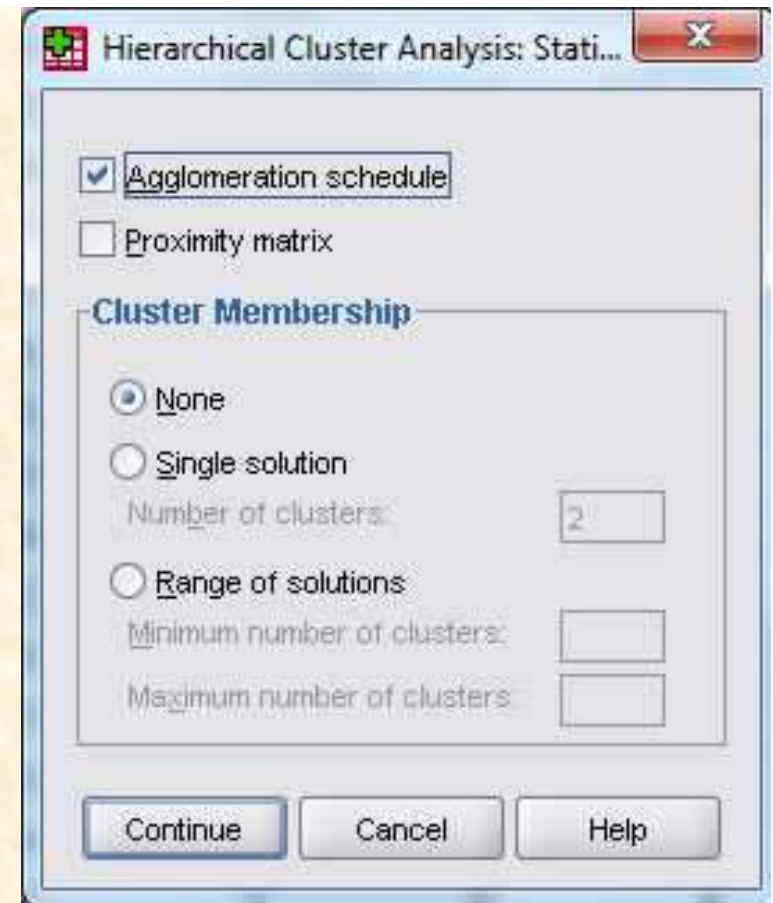
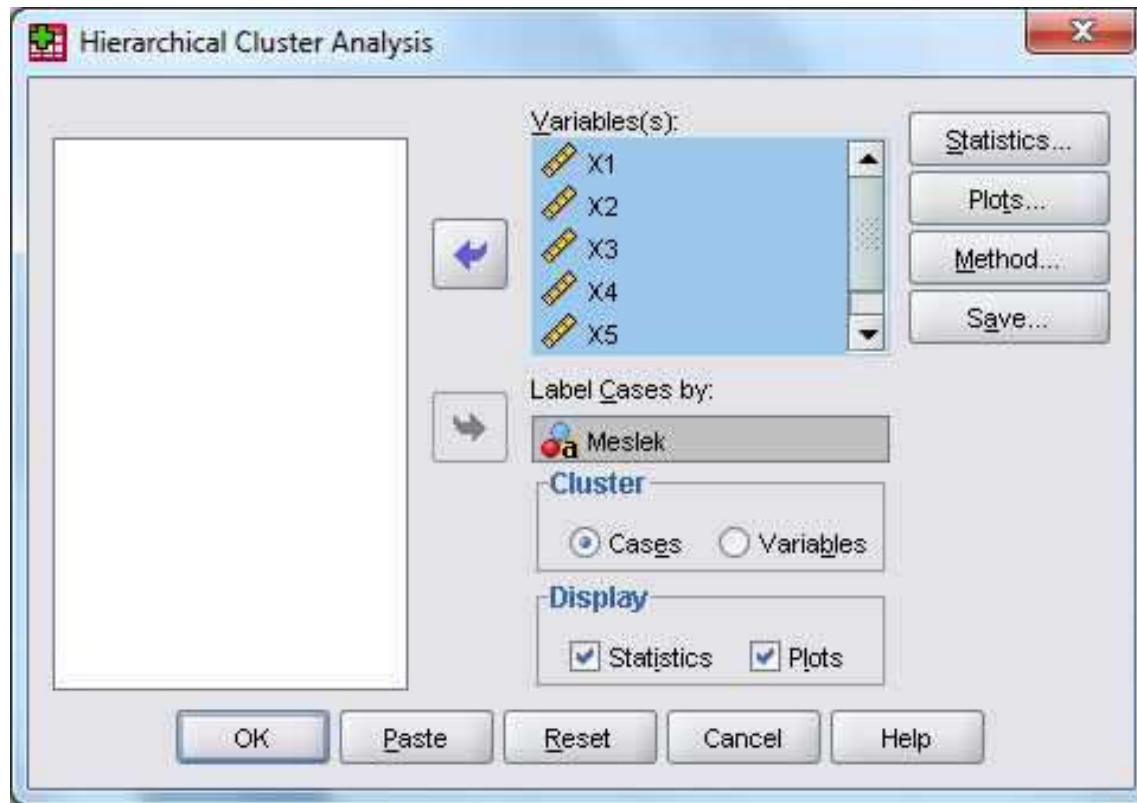
X4: Otomobil alırken öncelikle fiyatına dikkat ederim

X5: Otomobillerin özelliklerini bilmem

X6: Otomobilimi değiştirmekten hoşlanmam

**Verileri (cases)** Hiyerarşik Kümeleme Analizi ile ve hiyerarşik olmayan Kümeleme analizi ile çözüünüz?

	Meslek	X1	X2	X3	X4	X5	X6
1	Doktor	6	4	7	3	2	3
2	Emekli	2	3	1	4	5	4
3	Yatırım	7	2	6	4	1	3
4	Öğretmen	4	6	4	5	3	6
5	Memur	1	3	2	2	6	4
6	Mühendis	6	4	6	3	3	4
7	Öğrenci	5	3	6	3	3	4
8	Doktor	7	3	7	4	1	4
9	Evhanımı	2	4	3	3	6	3
10	Polis	3	5	3	6	4	6
11	İşçi	1	3	2	3	5	3
12	Tüccar	5	4	5	4	2	4
13	Muhasebe	2	2	1	5	4	4
14	Öğrenci	4	6	4	6	4	7
15	Avukat	6	5	4	2	1	4
16	İşçi	3	5	4	6	4	7
17	Mimar	4	4	7	2	2	5
18	Öğretmen	3	7	2	6	4	3
19	Memur	4	6	3	7	2	7
20	Hemşire	2	3	2	4	7	2



**Hierarchical Cluster Analysis: Plots**

Dendrogram

**Icicle**

All clusters

Specified range of clusters

Start cluster:

Stop cluster:

By:

None

**Orientation**

Vertical

Horizontal

**Hierarchical Cluster Analysis: Method**

Cluster Method:

**Measure**

Interval:

Power:  Root:

Counts:

Binary:

Present:  Absent:

**Transform Values**

Standardize:

By variable

By case

**Transform Measure**

Absolute values

Change sign

Rescale to 0-1 range

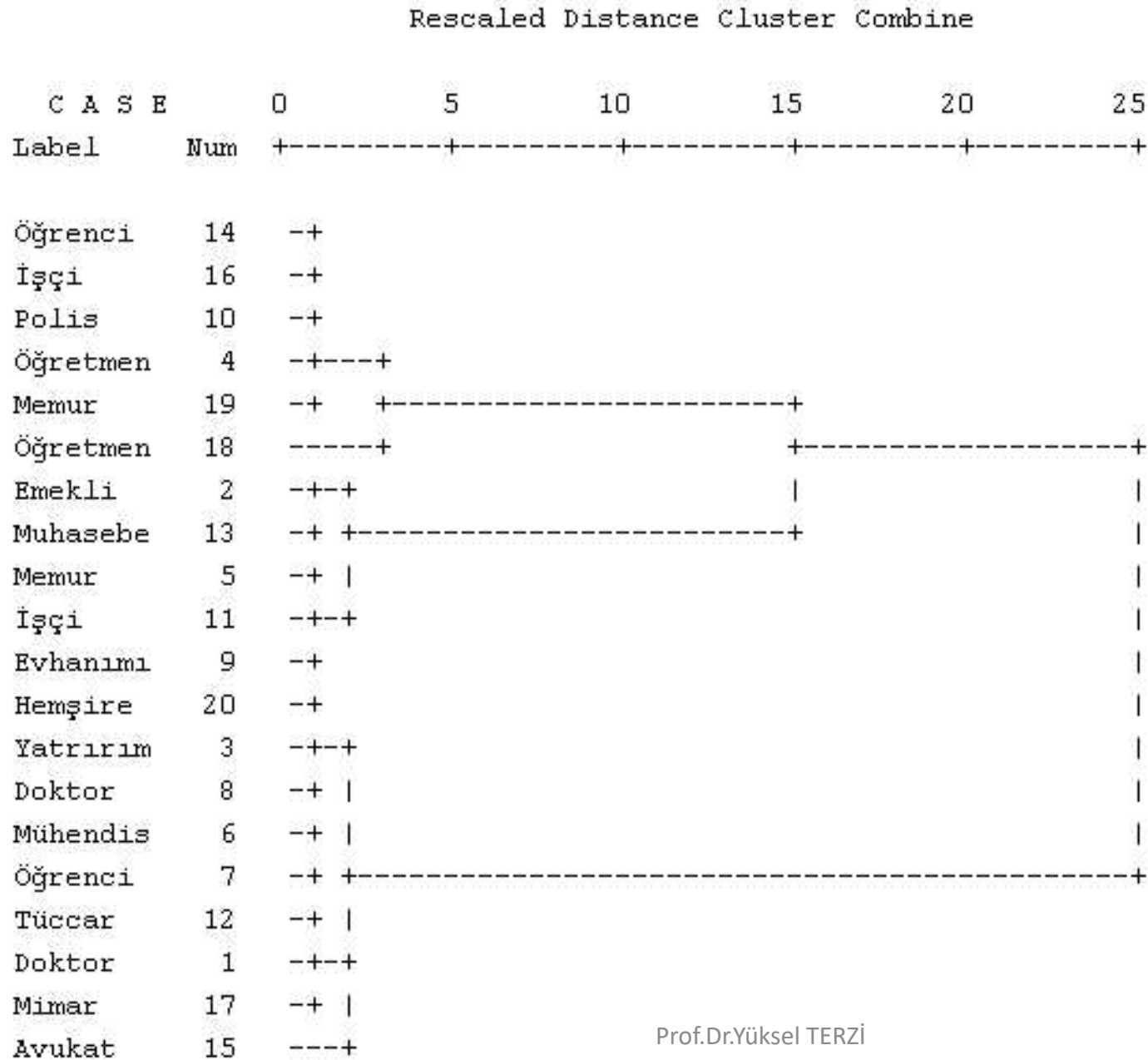


# Ward Linkage

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	14	16	1,000	0	0	6
2	6	7	2,000	0	0	7
3	2	13	3,500	0	0	15
4	5	11	5,000	0	0	11
5	3	8	6,500	0	0	16
6	10	14	8,167	0	1	9
7	6	12	10,500	2	0	10
8	9	20	13,000	0	0	11
9	4	10	15,583	0	6	12
10	1	6	18,500	0	7	13
11	5	9	23,000	4	8	15
12	4	19	27,750	9	0	17
13	1	17	33,100	10	0	14
14	1	15	41,333	13	0	16
15	2	5	51,833	3	11	18
16	1	3	64,500	14	5	19
17	4	18	79,667	12	0	18
18	2	4	172,667	15	17	19
19	1	2	328,667	16	18	0

## Dendrogram using Ward Method



Gözlemlerin 3 grupta kümelendiği görülmektedir.

### I.Küme:

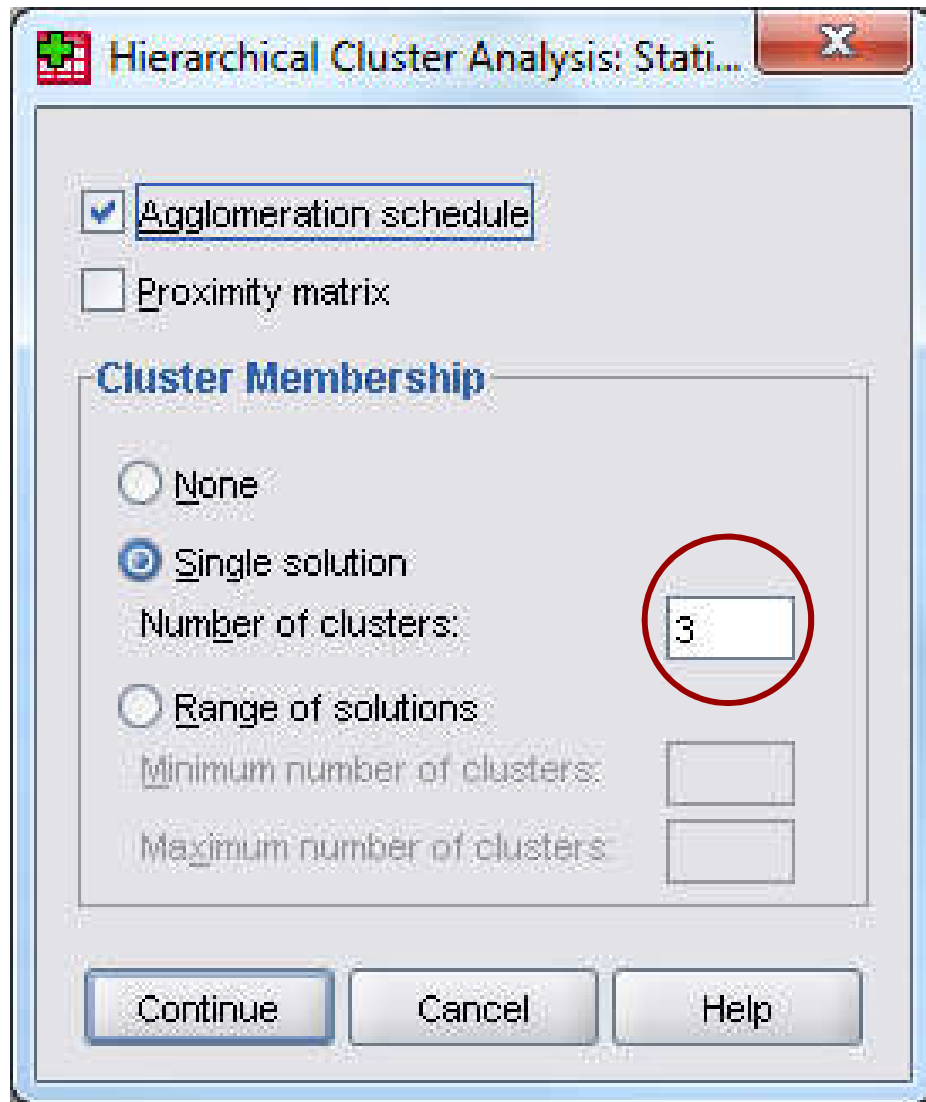
6 gözlemden oluşmaktadır.  
14,16,10,4,19,18

### II.Küme:

6 gözlemden oluşmaktadır.  
2,13,5,11,9,20

### III.Küme:

8 gözlemden oluşmaktadır.  
3,8,6,7,12,1,17,15



### Cluster Membership

Case	3 Clusters
1:Doktor	1
2:Emekli	2
3:Yatırım	1
4:Öğretmen	3
5:Memur	2
6:Mühendis	1
7:Öğrenci	1
8:Doktor	1
9:Evhanımı	2
10:Polis	3
11:İşçi	2
12:Tüccar	1
13:Muhasebe	2
14:Öğrenci	3
15:Avukat	1
16:İşçi	3
17:Mimar	1
18:Öğretmen	3
19:Memur	3
20:Hemşire	2



## Hierarchical Cluster Analysis: Save



### Cluster Membership

None

Single solution

Number of clusters:

3

Range of solutions

Minimum number of clusters:

Maximum number of clusters:

Continue

Cancel

Meslek	X1	X2	X3	X4	X5	X6	CLU3_1
Doktor	6	4	7	3	2	3	1
Emekli	2	3	1	4	5	4	2
Yatırım	7	2	6	4	1	3	1
Öğretmen	4	6	4	5	3	6	3
Memur	1	3	2	2	6	4	2
Mühendis	6	4	6	3	3	4	1
Öğrenci	5	3	6	3	3	4	1
Doktor	7	3	7	4	1	4	1
Evhanımı	2	4	3	3	6	3	2
Polis	3	5	3	6	4	6	3
İşçi	1	3	2	3	5	3	2
Tüccar	5	4	5	4	2	4	1
Muhasebe	2	2	1	5	4	4	2
Öğrenci	4	6	4	6	4	7	3
Avukat	6	5	4	2	1	4	1
İşçi	3	5	4	6	4	7	3
Mimar	4	4	7	2	2	5	1
Öğretmen	3	7	2	6	4	3	3
Memur	4	6	3	7	2	7	3
Hemşire	2	3	2	4	7	2	2

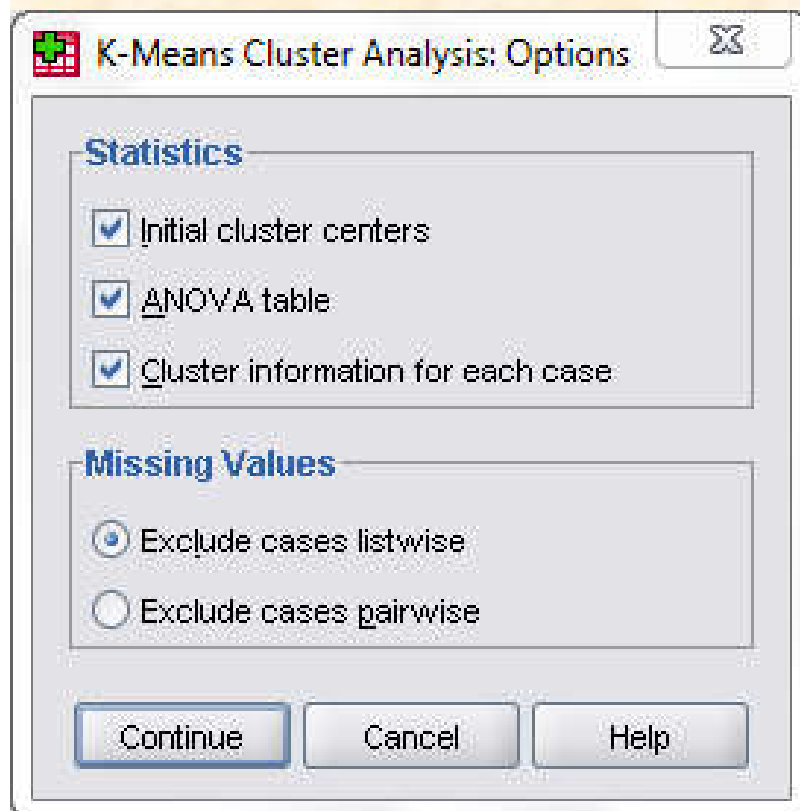
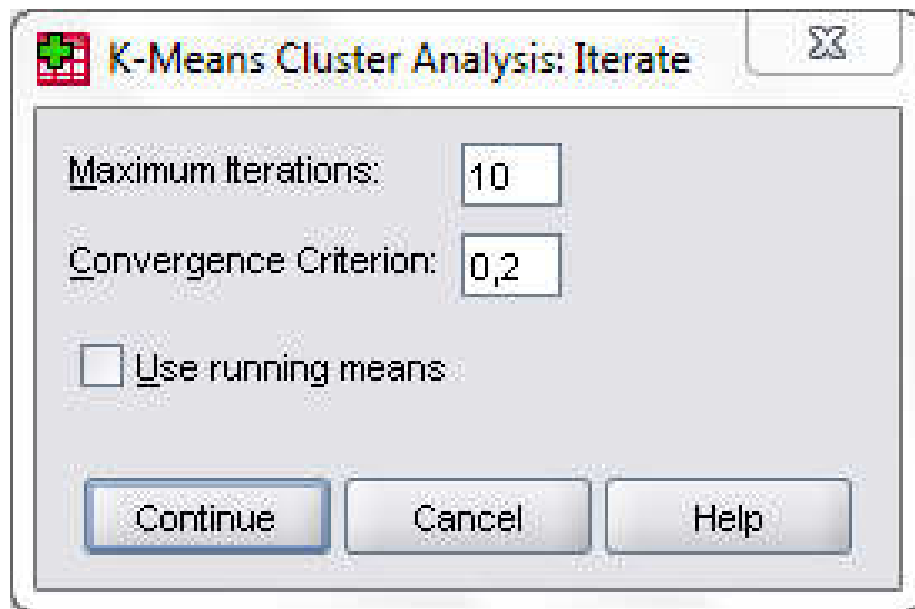
## b) Hiyerarşik Olmayan Kümeleme Analizi

The image shows the SPSS software interface. On the left, the 'Classify' menu is open, showing options: 'TwoStep Cluster...', 'K-Means Cluster...' (highlighted), 'Hierarchical Cluster...', 'Tree...', 'Discriminant...', and 'Nearest Neighbor...'. On the right, the 'K-Means Cluster Analysis' dialog box is displayed. The 'Variables:' list contains X1, X2, X3, X4, X5, and X6. The 'Label Cases by:' field is set to 'Meslek'. The 'Number of Clusters:' is set to 3. The 'Method' section has 'Iterate and classify' selected. The 'Cluster Centers' section has 'Read initial:' and 'Write final:' options, with 'External data file' and 'New dataset' selected. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. A red arrow points from the text 'Küme sayısı önceden girilir.' to the 'Number of Clusters:' field.

**Küme sayısı önceden girilir.**

Prof.Dr.Yüksel TERZİ





	Meslek	QCL_1	QCL_2
1	Doktor	3	1,41421
2	Emekli	2	1,32288
3	Yatırım	3	2,54951
4	Öğretmen	1	1,40436
5	Memur	2	1,84842
6	Mühendis	3	1,22474
7	Öğrenci	3	1,50000
8	Doktor	3	2,12132
9	Evhanımı	2	1,75594
10	Polis	1	1,14261
11	İşçi	2	1,04083
12	Tüccar	3	1,58114
13	Muhasebe	2	2,59808
14	Öğrenci	1	1,40436
15	Avukat	3	2,82843
16	İşçi	1	1,62447
17	Mimar	3	2,59808
18	Öğretmen	1	3,55512
19	Memur	1	2,15381
20	Hemşire	2	2,10159

**QCL\_1:** Her gözlemin hangi kümede olduğunu gösterir.

**QCL\_2:** Her gözlemin küme merkezinde uzaklığını gösterir.

### Initial Cluster Centers

	Cluster		
	1	2	3
X1	4	2	7
X2	6	3	2
X3	3	2	6
X4	7	4	4
X5	2	7	1
X6	7	2	3

### Initial Cluster Centers (İlk küme merkezleri):

Değişkenlerin 3 olarak belirlenen kümelerdeki merkezleri

### Iteration History<sup>a</sup>

Iteration	Change in Cluster Centers		
	1	2	3
1	2,154	2,102	2,550
2	,000	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 2. The minimum distance between initial centers is 7,746.

### Initial Cluster Centers (Tekrarlanma Tablosu):

2 iterasyonda 3 küme oluşmuştur.

### Cluster Membership

Case Number	Meslek	Cluster	Distance
1	Doktor	3	1,414
2	Emekli	2	1,323
3	Yatırım	3	2,550
4	Öğretmen	1	1,404
5	Memur	2	1,848
6	Mühendis	3	1,225
7	Öğrenci	3	1,500
8	Doktor	3	2,121
9	Evhanımı	2	1,756
10	Polis	1	1,143
11	İşçi	2	1,041
12	Tüccar	3	1,581
13	Muhasebe	2	2,598
14	Öğrenci	1	1,404
15	Avukat	3	2,828
16	İşçi	1	1,624
17	Mimar	3	2,598
18	Öğretmen	1	3,555
19	Memur	1	2,154
20	Hemşire	2	2,102

### Cluster Membership (Küme Üyeliği Tablosu):

Gözlemlerin hangi kümelerde olduğu ve uzaklık değerleri verilmiştir. Aynı kümede olanlara bakarak bu kümelere ortak bir isim verilebilir.

### Final Cluster Centers

	Cluster		
	1	2	3
X1	3,50	1,67	5,75
X2	5,83	3,00	3,63
X3	3,33	1,83	6,00
X4	6,00	3,50	3,13
X5	3,50	5,50	1,88
X6	6,00	3,33	3,88

### Final Cluster Centers (Son Küme Merkezleri):

6 değişkenin 3 kümedeki ortalamalarını gösterir. Örneğin otomobillerle ilgilenmekten (X1) en çok zevk alan küme 3. kümedir (5,75).

### Distances between Final Cluster Centers

Cluster	1	2	3
1		5,568	5,698
2	5,568		6,928
3	5,698	6,928	

### Distances between final cluster centers (Son Kümeleme Merkezleri Arasındaki Uzaklıklar):

Bu tablodan 1.-2. ve 1-3. kümenin birbirine yakın olduğunu, 2.-3. kümenin de birbirinden en uzak olduğunu gösterir. Böylece 1. küme, 2. ve 3. kümenin ortasında yer almalıdır.



### Number of Cases in each Cluster

Cluster	1	6,000
	2	6,000
	3	8,000
Valid		20,000
Missing		,000

Number of cases in each cluster (Her kümeye isabet eden gözlem sayısı):

## ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
X1	29,108	2	,608	17	47,888	,000
X2	13,546	2	,630	17	21,505	,000
X3	31,392	2	,833	17	37,670	,000
X4	15,713	2	,728	17	21,585	,000
X5	22,538	2	,816	17	27,614	,000
X6	12,171	2	1,071	17	11,363	,001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

**ANOVA:** ANOVA tablosu deęişkenlerin kümeler itibariyle farklılığının öğrenilmesi amacıyla kullanılır. Deęişkenler kümeler itibariyle farklı çıkmıştır. Çünkü kümeleme analizi ile kümeler arası fark maksimum olmuştur. Böylece kümelerdeki gözlemlerin dağılımı tesadüfi değildir.