

ÖLÇMEDE KÜMELEME ANALİZİ UYGULAMALARI

TÜLİN OTBİÇER

**Seminer Raporu
Olarak Hazırlanmıştır.**

**Ankara
Hacettepe Üniversitesi
Mayıs, 2004**

tulinotbicer@yahoo.com

Bu seminer notlarında,

- *Kümeleme Analizi nedir?*
- *Kümeleme Analizini niye yapıyoruz?*
- *Kümeleme Analizinin yöntemi nedir?*
- *Kümeleme Analizinin iki yöntemi nedir?*
- *Kümeleme Analizinin Diskriminant Analizinden Farkı var mıdır? Varsa farkı nedir?*
- *Yine Kümeleme Analizinin Faktör Analizinden farkı nedir?*
- *Kümeleme Analizinin Ölçme alanında uygulamaları neler olabilir?*

Sorularının cevaplarını bulacaksınız.

Bu notlarda ifade edilenlerin yanı sıra kümeleme analiziyle ilgili oluşabilecek her türlü soru işaretini yaratmak birincil amaç olarak belirlenmiştir.

KÜMELEME ANALİZİ NEDİR?



Düşünelim!

Çiçeklerle dolu bir alanda gezinirken dikkatinizi çeken bir çiçeğin, çiçekler arasında hangi çiçek grubuna ait olduğunu nasıl belirlerdiniz? Etrafınızda gördüğünüz pek çok çiçek bir birine benziyor değil mi? Bu çiçeklerin her birinin özellikleri var. Ağırlığı, çiçek sayısı, yaprak sayısı, uzunluğu, rengi, kokusu gibi pek çok değişken çiçeklere ait özelliklerdir.

İşte bu noktada baktığımız bir çiçeğin hangi tür çiçek grubuna girdiğini belirleme işinde Kümeleme Analizinden yararlanabiliriz.

⇒ Kümeleme analizi, birim ve bu birimlere ait değişkenlerin gruplamaları (sınıflamaları) hakkında kesin bilgilerin bulunmadığı bir ana kitleden (popülasyonlardan) alınan n tane birimin (bireye/objeye) p tane değişkenine ilişkin gözlem sonuçlarını ilgilendir.

Satırları p tane değişkenden ve sütunları n tane birim sayısından oluşan bir iki boyutlu tablo ile kümeleme analizine ilişkin verilerin gösterimi şekildeki gibidir.

	x_1	x_2	x_3	...	x_p
1					
2					
3					
.					
.					
.					
N					

Şekil 1

Satırları p tane değişkenden ve sütunları n tane birim sayısından oluşan bir iki boyutlu tablo ile kümeleme analizine ilişkin verilerin gösterimi şekil 1'deki gibidir.

Matematikte yukarıdaki şekil 1 deki gösterimi matrisler yoluyla ifade ederiz. Örneğin 8 tane bireyin yaş, boy, ağırlık değişkenlerinin ölçümlerine ait X ($n \times p$) şeklinde matris olarak gösterimi yandaki $X_{(8 \times 3)}$ matrisidir.

$$X_{(8 \times 3)} = \begin{bmatrix} \text{Yaş} & \text{Boy} & \text{Ağırlık} \\ 25 & 165 & 55 \\ 32 & 183 & 65 \\ 28 & 175 & 74 \\ 26 & 167 & 60 \\ 24 & 183 & 65 \\ 32 & 174 & 72 \\ 26 & 155 & 80 \\ 29 & 164 & 70 \end{bmatrix}$$

→ **Kümeleme Analizi,**

Yapıları hakkında kesin bilgilerin bulunmadığı bir veri yığını içindeki

Birimleri

Değişkenleri

ya da

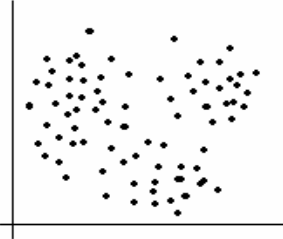
birim ve değişkenleri

birbiri ile benzer olan alt kümelere (grup, sınıf) ayırma yöntemidir.

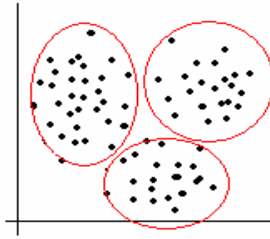
Kümeleme Analizini niye Yapıyoruz?

Ana kitlede var olan yapıyı mümkün olduğunca ortak özellikli (homojen) ve kendi aralarında farklı (heterojen) alt gruplara bölerek birimleri ve değişkenleri gruplamak, sınıflamak, kümelemek için yapıyoruz.

Veri yığını



Kümeleme analizi sonucu veri yığınları

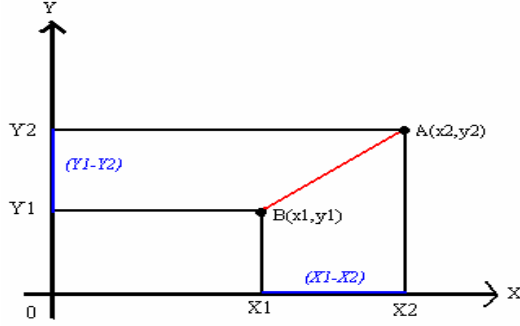


Kümeleme analizi sonucunda karmaşık bir dağılım olan veri yığını (soldaki şekil), kümeleme analizi sonucunda veriler benzer özelliklerine göre gruplanırlar.

YÖNTEM: Kümeleme Analizini Neye Göre Yapıyoruz?

Değişkenler arası **Benzerlik** (similarity) ya da **farklılıklara** (dissimilarity) dayanır. Yani Kümeleme analizinde bir birime ait değişkenlerin birbirleriyle olan uzaklıkları hesaplanır. Kısaca **Uzaklık matrisi** ya da **benzerlik matrisinden** yararlanıyoruz.

Koordinat sisteminde iki nokta arasındaki uzaklığın gösterimi ise;



- Kümeleme Analizinde, birimlerin P değişkenine göre birbirleri arasındaki uzaklıklarını hesaplamak için çok çeşitli ölçü birimleri ileri sürülmüştür.

Uzaklık ya da benzerlik ölçülerini nasıl hesaplıyoruz?

Hesaplama Yöntemleri; Veri matrisinde yer alan değişkenlerin ölçü birimlerine göre farklılık göstermektedir.

Eğer değişkenler (X_p).

* **Oransal ya da aralıklı ölçekle** elde edilmiş ölçümler ise **Uzaklık(distance)** ya da **ilişki (correlation)** türü ölçülerden

* **Sıralama ölçeği ile** edilmiş ölçümler ise **Ki-kare uzaklık ölçüsü** ya da **phi kare uzaklık** türü ölçülerden

* **İkili (binary) ölçekle** elde edilmiş ölçümler ise **öklid, kare öklid, size difference, pattern difference, Lance and williams difference, shape difference** gibi ölçülerden yararlanılır.

Açıklama [T01]: Uzaklık(dist
ance), ilişki (correlation), Ki-
kare uzaklık ölçüsü, phi kare
uzaklık, öklid, kare öklid, size
difference, pattern
difference, Lance and williams
difference, shape difference
ölçülerinin nasıl hesaplandığına
kapsam gereği
değinilmemiştir.

Kümeleme Analizi Yöntemleri

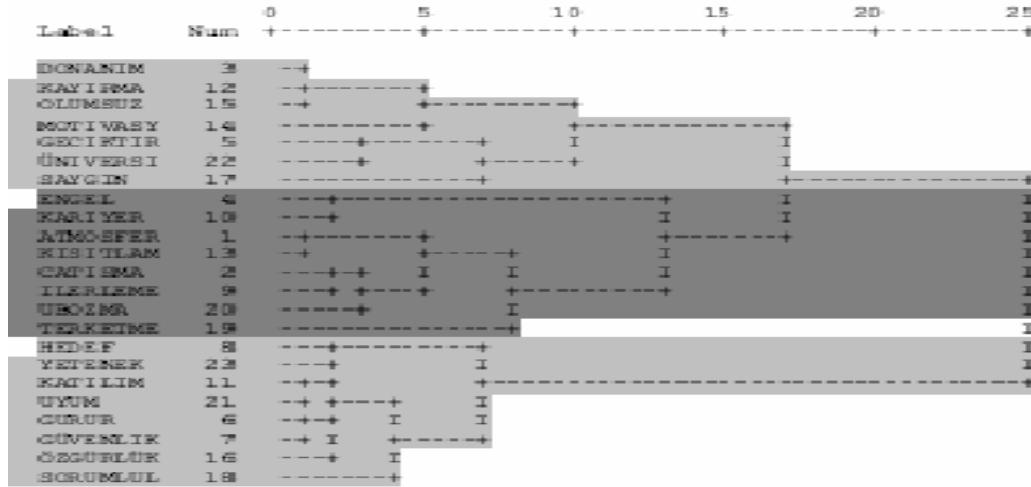
1- Aşamalı Kümeleme Yöntemi (Hierarchical cluster Analysis):

Bu yöntemde, kümeler ardarda birleştirilir ve bir grup değeri ile bir kez birleştirildikten sonra, devam eden adımlarda bir daha ayrılmaz. Bu yöntemde ele alınan değişkenler Hiyerarşik bir yapı oluştururlar. Hiyerarşik kümeleme yöntemlerinde küme sayısına görsel olarak karar verilir. Bunun için de genellikle *dendogram* olarak bilinen *ağaç diyagramı* kullanılır.

Yapılan bir çalışmaya göre üniversitelerdeki öğretim elemanlarının akademik kariyer tutumlarını değerlendirmek ve tutum davranışlarını gruplandırmak amacı için 50 tane devlet üniversitesinden 3512 öğretim üyesine 44 tutum maddesinden oluşan ölçek uygulanmıştır. Öğretim elemanlarının tutumlarını ölçmek için Yirmi üç (23) ayrı boyut, değer belirlenmiş. Bu çalışma sonunda toplanan verilere Kümeleme analizi uygulanmış ve kümeleme analizi sonucunda oluşan *dendogram* (*ağaç diyagramı*) şu şekildedir.

Açıklama [TO2]: Merak edenler için seminer notları sonuna uygulanan ölçek eklenmiştir.

Tutum Değişkenleri İçin Ağaç Diyagramı



Ağaç diyagramında da anlaşılacağı gibi 3 tane temel grup oluşmuştur.

Açıklama [TO3]: Kümeleme analizine *değişkenler* girmiştir.

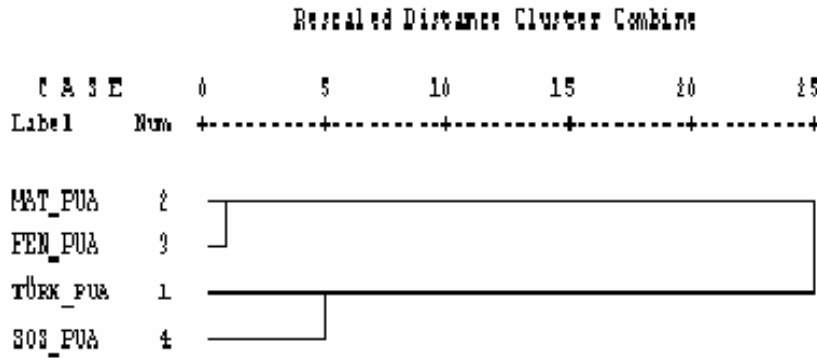
Birinci grubu: Hedef, yetenek, katılım, uyum, gurur, güvenlik, özgürlük ve sorumluluk değerleri oluştururken bireylerin kariyerleriyle ilgili *kişisel* değerler kümesidir.

İkinci grupta genel olarak *çift kariyerli eşlere* (aynı iş kolunda çalışan eşler) yönelik değer kümelerinin yer aldığı görülmektedir. Bu grubu oluşturan değerler; engel, kariyer, atmosfer, kısıtlama, çatışma, ilerleme, bozma, terk etme dir.

Üçüncü grupta yer alan değer kümelerine baktığımızda ise, yine genel olarak bireyin **örgüte** yönelik değer kümeleri olduğu görülmekte. Bir diğer ifade ile, akademisyenlerin çalıştıkları kurum ile ilgili değer yargılarının oluşturduğu kümelerdir. Grubu oluşturan değerler ise donanım, kayırma, olumsuzluk, motivasyon, geciktirme, üniversite, saygınlık tır.

Bir başka örnek ise 200 öğrencinin Fen bilgisi, sosyal bilgiler, matematik ve Türkçe test başarı puanları değişkenlerinden oluşan bir veri setine kümeleme analizi uygulandığında ortaya çıkan ağaç diyagramı ise Dendrogram using Average Linkage (Between Groups)

Açıklama [T04]: Yine kümeleme analizine birimler değil değişkenler analiz edilmiştir.



2- Aşamalı olmayan Kümeleme Yöntemi (Nonhierarchical Analysis):

n tane birimden k tane küme oluştururuz, bunun içinde ön görülen bir küme sayısı vardır. Küme sayısı bellidir. Örneğin bir önceki örnekte bulunan 200 öğrencinin 4 alt test başarı puanı değişkenleri baz alarak bu yöntemle küme sayısı 4 olacak şekilde kümeleme analizini yaptığımızda oluşan kümelerin veri sayısı ve verilerin toplam test başarı puanları ortalamalarına ait tablo:

Açıklama [T05]: Kümeleme analizine değişkenler değil birimler (kişiler) girmiştir.

küme sayısı	N	ortalama
1	37	58,48
2	67	29,23
3	19	78,89
4	77	39,89
Toplam	200	43,47

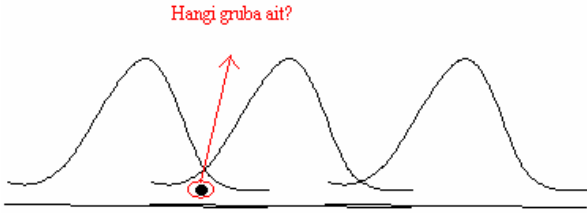
Veri setinde hangi bireylerin hangi kümeye ait olduğunu bilebilmekteyiz. Yani veri setindeki 10.kişi hangi kümede, 20.kişi hangi kümede bunu saptanabiliyor. Yandaki tabloda kümeleme analizi sonucunda 1 kümeyi oluşturan 37 kişinin kimler olduğu belli olmaktadır.



Diğer İstatistiksel yöntemlerle Kümeleme Analizinin İlişkisi Var mı?

Ayırma (Diskriminant) Analizinden Farkı Nedir?

Ayırma analizi, bir ayırma fonksiyonu kullanır. Ayırma fonksiyonunu, örnekleme yeni girecek birimler için tekrar tekrar kullanabilirken, kümeleme analizinde sonradan gelen ya da örneklemden çıkarılan bir birim için analizi tekrarlamak gerekmektedir, çünkü birimler/değişkenler arasındaki yakınlık-uzaklıklar söz konusu olduğundan hesaplar değişecektir.



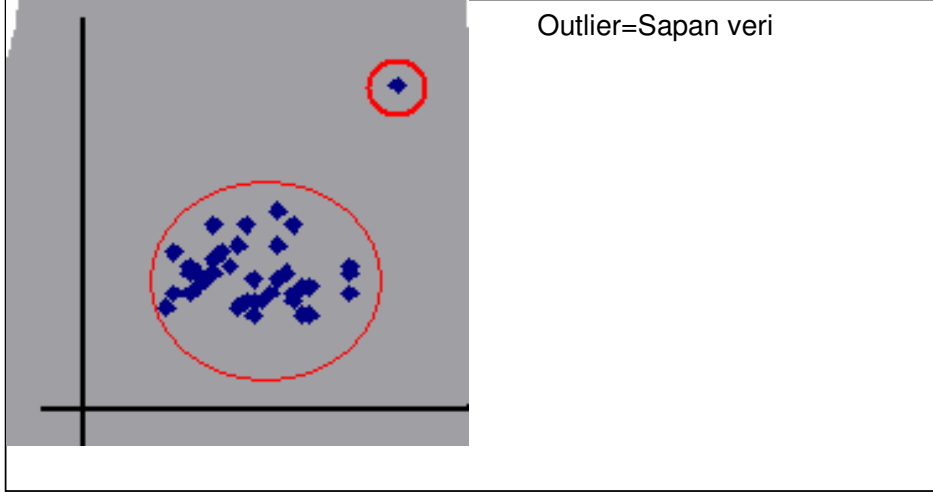
Faktör Analizinden Farkı Nedir ya da Aralarında Bir İlişki Var mı?

Amaç 1: Çok sayıdaki değişken setini daha az sayıda faktörlere indirgemektir. (data reduction)

Amaç 2: Değişkenler arasındaki var olan bağımlılığı kaldırıp faktörler arasında bağımsızlığı sağlar. Bu amaca yönelik olarak da Kümeleme Analizi yöntemiyle bağımsız hale gelen faktörleri değişken olarak kullanmak mümkündür. Bu durumda, kümeleme analizi, faktör analizinden bir araç olarak faydalanır.

Kümeleme Analizinin Ölçme Alanında Uygulamaları Nasıl Olabilir?

Araştırmanın amacı bağlamında, kümeleme analizi yöntemiyle örnekleme oluşturan birimlerin homojen olup olmadığı sınırlanabilir. Mesela, bir ölçme aracı, 100 kişilik bir gruba uygulanmış ve bireyler üzerinde kümeleme analizi yapılmıştır. Kümeleme analizi sonucunda 99 birey bir küme oluştururken 1 birey (sapan veri) başka bir küme oluşturmuştur.



Eğer yapacağımız araştırma için toplayacağımız verilerde bir benzerlik, homojenlik arıyorsak bunun testi için kümeleme analizi yapılabilir. Örnekteki durum üzerine yorum yapacak olursak, toplanan verilere dahil edilen sapan veri (100 kişi) ile yapılacak istatistikler ve bunun sonuçlarından doğan yorumlar ile sapan veri dahil edilmeden (99 kişi) yapılacak istatistiksel yorumlar kuşkusuz tutarsız olacaktır.



Bir okulda rehber öğretmen ya da ölçme ve değerlendirme uzmanı olduğunuzu düşünün. Okunuzda etkili öğretim için öğrencileri derslerdeki başarılarına göre gruplara ayırıp bu gruplara göre bir öğretim programı uygulayacaksınız. Bunun için okunuzdaki öğrencilerin derslerdeki başarılarını değişken olarak baz alıp bu öğrencileri kümeleme analizi yoluyla gruplara ayırabilirsiniz.

Öğrenciler	Matematik	Türkçe	Sos. Bil.	Fen Bil.
1	M1	T1	S1	F1
2	M2	T2	S2	F2
3				
.				
.				
.				
n	Mn	Tn	Sn	Fn

N sayıda öğrenciden oluşan bir grup. Bu öğrencilerin de Matematik, Türkçe, Sos. Bil. ve Fen Bil. Ders başarı puanları ölçülmüş olsun. Buna göre öğrencileri kümeleme analizi yöntemiyle derslerdeki performansları bakımından farklı şubeler oluşturmada kullanabiliriz.



Lise 2. sınıf öğrencilerinin alan seçiminde belirsizliklerini ortadan kaldırmak için onların derslerdeki başarılarını baz alarak yapacağınız bir kümeleme analizi yardımıyla onlara hangi alana yönlendirebileceğinizin öngörüsünde bulunabilirsiniz.



Tıpta; hastalıkların sınıflandırılmasında



Psikiyatride; paranoya, şizofreni gibi semptomların sınıflandırılmasında



Eğitimde; kültürel ve eğitsel alt yapılara göre eğitim programları geliştirmede



Psikolojide; toplum ya da alt toplumların davranış, bilgi ve tutum kalıplarının ortaya konmasında kümeleme analizinden yararlanılabilir.

Detayları paylaşmak için iletişim:

tulinotbicer@yahoo.com

www.groups.yahoo.com/group/egitimde_olcme

Kaynakça

<http://idari.cu.edu.tr/sempozyum/bil20.htm>

<http://idari.cu.edu.tr/sempozyum/bil29.htm>

<http://iktisat.uludag.edu.tr/dergi/8/ayse/ayse2.htm>

<http://trochim.human.cornell.edu/tutorial/flynn/cluster.htm>

<http://www.angelfire.com/ia/selcukkoc/ka.html>

http://www.biyostatistik.com/egit/not6_1.htm

http://www.frekans.com.tr/html/7tr_analizler.asp

<http://www.psychstat.smsu.edu/multibook/mlt04m.html>

<http://www.trafik.gov.tr/icerik/bildiriler/C5-61.doc>

ÖZDAMAR, Kazım., **Paket Programlar ile İstatistiksel Veri Analizi**, Kaan Kitapevi, Eskişehir, 1999

TATLIDİL, Hüseyin., **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Cem Web Ofset Ltd.Şti., Ankara, 1996

Akademisyenlerin Çalışma Yaşamı ve Kariyer Sorunları, Proje No: 99/29

Öğretim elemanlarının tutumlarına ilişkin yargı cümleleri ve bu yargı cümlelerini tanımlayan anahtar kelimeler aşağıda verilmiştir.

1. Seçmiş olduğum meslek bana gurur veriyor (GURUR)
2. Yeteneklerimin ne olduğunu biliyorum (YETENEK)
3. Hedeflerime ulaşmak için neler yapmam gerektiğini biliyorum (HEDEF)
4. Kararlarımın farklılık yaratacağı bir iş yerinde sorumluluk düzeyine ulaşmak isterim (SORUMLULUK)
5. Bir iş yerinin, iş garantisi, sosyal ve ekonomik faydalar için iyi bir emeklilik vb. vasıtalarla güvenlik sağlayacak olması benim için çok önemli (GÜVENLİK)
6. Kendi işimi ve ilerleyebileceğim mevki seçmek bana bağımsızlık ve özgürlük kazandırıyor (ÖZGÜRLÜK)
7. Kendi kişilik yapımla, seçtiğim mesleğim ve kariyerim arasında uyum var (UYUM)
8. Öğretim üyeliği mesleği toplumda saygın bir yere sahip (SAYGIN)
9. Üniversitem, bana gerekli mesleki bilgi ve deneyimi kazandırmak için hiçbir şey yapmıyor (ÜNİVERSİ)
10. Üniversitede alınacak tüm kararlara her statüden katılımın sağlanması, (yönetmelik demokrasi) iş doyumunu artırır (KATILIM)
11. Üniversite-sanayi işbirliğinin gerçekleşmesi için üniversitem, sanayinin yararlanabileceği teknik donanıma sahip değil (DONANIM)
12. Benim kariyerim, her şeyden önce gelir (KARİYER)
13. Evlilik, akademik kariyerde ilerlemeye engeldir (ENGEL)
14. Kadının iş ve aile yükümlülükleri, akademik kariyer ilerlemesini geciktirmektedir (GECİKTİR)
15. Eşimle aynı kurumda çalışıyor olmak özgürlüklerimi kısıtlar (KISITLAM)
16. Aynı kurumda çalışan eşler, kurumun atmosferini bozar (ATMOSFER)
17. Aynı kurumda çalışan eşlerin iş ve aile çatışmaları daha fazla olur (ÇATIŞMA)
18. Aynı kurumda çalışan eşler, diğer akademisyenlerden daha hızlı kariyer ilerleme fırsatına sahiptir (İLERLEME)
19. Aynı kurumda çalışan eşlerden birinin, kurum içindeki diğer çalışanlarla çatışması, eşinin konumunu olumsuz olarak etkiler (OLUMSUZ)
20. Aynı kurumda çalışan eşler, özel durumlarda birbirlerini kayırırlar (KAYIRMA)
21. Aynı kurumda çalışan eşlerin boşanması karşısında eşlerden birinin işinden ayrılıp kariyerini terk etmesi gerekir (TERKETME)
22. Aynı kurumda çalışan eşlerin boşanması, iş tatminini ve motivasyonunu azaltır (MOTİVASY)
23. Çalışan eşlerin kariyer problemi, eşler arasındaki uyumu bozar (UBOZMA)