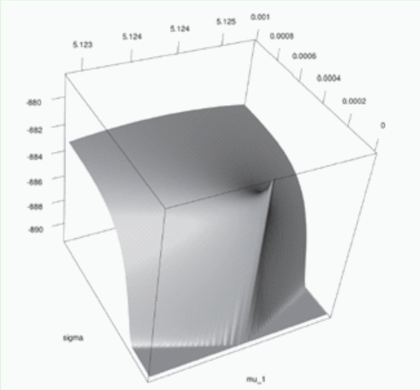
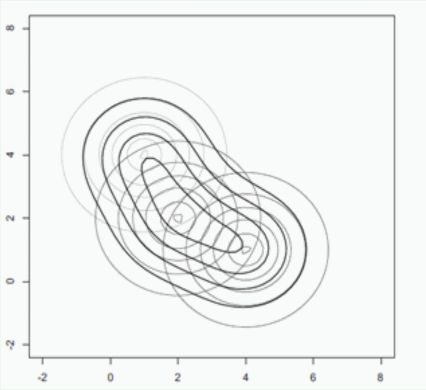




KARIŞIMLI MODEL YAKLAŞIMININ SINIFLANDIRMA AMAÇLI KULLANILMASI

Yıldız Ateş

İKSAD
Publishing House

KARIŞIMLI MODEL YAKLAŞIMININ SINIFLANDIRMA AMAÇLI KULLANILMASI[1]

Yıldız ATEŞ

Editör: Prof. Dr. Abdullah YEŞİLOVA

https://dx.doi.org/10.5281/zenodo.14593708

[1] Bu kitap çalışması yazarın aynı isimle, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı'nda ve Prof. Dr. Abdullah YEŞİLOVA danışmanlığında hazırladığı doktora tezinden Üretilmiştir. Bu çalışma VAN YYÜ Bilimsel Araştırma Projeleri Başkanlığı tarafından FDK-2017-5840 nolu proje olarak desteklenmiştir.

Institution of Economic Development and Social Researches Publications®
(The Licence Number of Publicator: 2014/31220)
TÜRKİYE TR: +90 342 606 06 75
USA: +1 631 685 0 853
E mail: iksadyayinevi@gmail.com
www.iksadyayinevi.com

ISBN: 978-625-378-062-3
Cover Design: İbrahim KAYA
December / 2024
Ankara / Türkiye
Size: 16x24cm

ÖN SÖZ

Bu eserin çalışması sürecinde değerli bilgi ve görüşlerini benimle paylaşan, tez konumun belirlenmesine katkı sunan danışmanın sayın Prof. Dr. Abdullah YEŞİLOVA' ya ve danışmanlığımı yapan, desteğini benden esirgemeyen değerli hocam sayın Prof. Dr. Hayrettin OKUT' a; tez izleme komitesi üyeleri sayın Dr. Öğr. Üyesi Gazel SER' e ve sayın Dr. Öğr. Üyesi Barış KAKİ' ye teşekkür ederim. Eğitim ve öğretim hayatım boyunca maddi ve manevi katkılarını benden esirgemeyen aileme, çalışmalarımda hem bilimsel hem de manevi desteğini esirgemeyen kız kardeşim Öğr. Gör Dr. Güneş AÇIKGÖZ' e ve ayrıca bana destek olan değerli arkadaşım Gökhan ATEŞ' e teşekkür ederim. Doktora sürecinde yardımlarından dolayı Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü' ne teşekkürlerimi sunarım.

Yıldız BORA

,,

İÇİNDEKİLER

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
AIC	Akaiki Bilgi Ölçütü (Akaike Information Criterion)
BIC	Bayesian Bilgi Ölçütü (Bayesian Information Criterion)
CEM	Koşullu Beklenti Maksimizasyonu (Conditional EM)
DIC	Devians Bilgi Ölçütü (Deviance Information Criterion)
EM	Beklenti Maksimizasyonu (Expectation Maximization)
FA	Faktör Analizi (Factor Analysis)
GMM	Gauss Karışımlı Modeller (Gaussian Mixture Models)
GMR	Gauss Karışımlı Regresyon (Gaussian Mixture Regression)
HMM	Hidden Markov Modeller (Hidden Markov Models)
ICL	İntegrali alınmış Tam Olabilirlik (Integrated Completed Likelihood)
LCA	Latent sınıf analizi
LPA	Latent profil analizi
MCMC	Monte Carlo Markov zincirini (Markov Chain Monte Carlo)
ML	Maksimum Olabilirlik (Maximum Likelihood)
SEM	Yapısal eşitlik modellemesi (Structure Equation Model)

Simgeler	**Açıklamalar**
$\hat{\theta}$	Parametre vektörü
L (ψ)	Olasılık fonksiyonu
z_i	Gözlenemeyen değişken
π_k	Karışımlı oran
Y_j	Olasılık yoğunluk fonksiyonu
$f(y_j)$	p-boyutlu rasgele vektör
$f_i(y_j)$	Karışımlı bileşen yoğunlukları
(μ_k)	Normal dağılıma ait ortalama
ψ	Bilinmeyen parametreleri
$\widehat{\psi}$	ψ ’ nin tahmini
ψ^0	ψ ' nin başlangıç değeri
y_i	Gözlem değeri
σ^2	Varyans
$\sum$	Varyans kovaryans matrisi
D_k	Ortogonal matris
A_k	Diyagonal matris
λ_k	Skalar bir değer
h_e	Entropy ölçütü

1. GİRİŞ

İncelenen özelliklerin (değişkenlerin) çok sayıda olması bilimsel çalışmalara olanak tanımaktadır ve ele alınan olaylar genellikle birçok etkinin altındadır. Bununla birlikte, incelenen değişkenlerin çoğu da birbiri ile ilişkilidir. Bu nedenle uygulamalarda, birbiri ile ilişkili çok fazla sayıda değişkenle karşılaşılmaktadır. Geçerli ve güvenilir sonuçların elde edilebilmesi için araştırma kapsamındaki bütün özelliklerin değerlendirmeye alınması gerekmektedir. Bu tip gözlemleri içeren çok değişkenli istatistiksel yöntemlerde amaç, yorumlanması güç, birbiriyle ilişkili çok sayıda değişkenden, en az bilgi kaybı ile bağımsız ve kavramsal açıdan anlamlı az sayıda yeni değişken yapısını ortaya koymaktır. Örneğin, faktör ve kümeleme analizi gibi çok değişkenli istatistiksel analiz tekniklerinde, yorumlanması güç, birbiri ile ilişkili çok sayıda değişkeni kendi içerisinde homojen alt gruplara ayırarak, alt gruplar arası homojenlik sağlanırken, alt gruplar arasındaki heterojenlik açıklanmaya çalışılmaktadır (Wang ve ark., 1998; Muthen ve Muthen, 2002; Yeşilova, 2010; Alpar, 2012).

Sınıflandırma amaçlı olarak; faktör ve kümeleme gibi istatistiksel yöntemler yaygın olarak kullanılmaktadır. Son yıllarda özellikle makine öğrenme esaslı modeller ile karışımlı model (mixture model) yaklaşımları karmaşık veri setlerini sınıflandırmada kullanılmaktadır. Karışımlı model yaklaşımı; ilk olarak veri setini heterojen kabul edip, öncelikle homojen kaç alt gruptan oluşacağını tespit etmektedir. Daha sonra elde edilen her bir alt grup için ayrı parametre tahminleri elde edilir. Bununla birlikte, karışımlı regresyon yaklaşımı ile hem veri seti homojen alt gruplara ayrılmakta hem de her bir alt grup için ayrı regresyon analizi yapılmaktadır. Karışımlı modelin diğer bir amacı da hatalı sınıflandırma olasılığını en aza indirgeyerek, bireyleri ait oldukları gruplara ayırmak ve çekilmiş oldukları kitleleri belirlemektir. Bu yönleriyle karışımlı model yaklaşımı diğer sınıflandırma yöntemlerine göre farklılık göstermektedir (Muthen ve Muthen, 2002; Okut ve ark., 2002; Yeşilova, 2003; Oberski, 2016).

Karışımlı modelde, veri seti tek bir populasyona ait olmayabilir. Başka bir deyişle, veri seti birden fazla populasyona ait heterojen veri kümesi olabilir. Karışımlı modeller; hem sürekli hem de kesikli veriler için kullanılmaktadır. Bunun yanı sıra, karışımlı modelleme hem sınıflandırma hem de regresyon amaçlı da kullanılabilir. Karışımlı regresyon; bağımlı değişkenin binom, Poisson ve normal dağılım gösterdiği durumlarda yaygın olarak uygulanmaktadır.

Normal dağılış gösteren verilere uygulanması normal karışımlı regresyon (normal mixture regression) olarak adlandırılmaktadır. Karışımlı modelde, bir örnekteki gözlemlerin populasyondaki gözlemlenmemiş alt gruplara ait olabileceği ve bu alt grupların belirlenmesi üzerinde durulur. Veri setinin sınırlı sayıda alt gruba sahip olduğu varsayılarak sonlu karışımlı modeller (finite mixture model) kullanılır (Yeşilova ve ark., 2009).

Karışımlı regresyon; sınıflama (alt gruplara ayırma) yapmakla birlikte, kümeleme ve faktör analizlerine göre iki önemli avantajı bulunmaktadır. Bunlardan ilki, her bir gözlemin hangi olasılıkla hangi alt gruba dâhil olabileceğinin belirlenmesidir. Karışımlı modelin ikinci önemli avantajı ise, her bir alt grup için parametre tahminlerinin elde edilmesidir. Çünkü karışımlı model yaklaşımı, veri setini homojen alt gruplara ayırdıktan sonra her bir alt grup için oluşturulan bağımlı değişken ile bağımsız değişkenlere göre çoklu regresyon analizi de yapmaktadır (Yeşilova ve ark., 2016).

Karışımlı modellemede, parametre tahmini için Beklenti Maksimizasyonu (Expectation Maximization=EM) algoritmasını esas alan en çok olabilirlik yöntemi (maximum likelihood=ML) kullanılmaktadır. Sonlu karışım durumuna karşılık gelen olasılık denklemleri analitik olarak çözülemediğinden, nümerik bir yöntem gerekmektedir. Kapalı formlu bir çözüm mevcut değilse, iteratif EM algoritması, olasılık denkleminin çözümü için uygun bir yoldur. Dolayısıyla EM algoritması, karışımlı dağılımların parametre tahminlerinde kullanılabilir. EM yaklaşımı, E ve M aşamalarını kullanarak en çok olabilirlik tahminlerini elde etmektedir. E aşamasında, gözlenmiş veriler üzerinde koşullu beklenen değerler kullanılarak eksik verilerin tahmini yapılmaktadır. Burada eksik veriler, latent sınıflardır. M aşamasında, parametre tahminleri log-olabilirliğin beklenen değerinin maksimize edilmesi ile elde edilmektedir. Karışımlı modelde, doğru sınıflandırma olasılığı için entropy istatistiği, uygun alt grup sayısını belirlemede ise Akaiki ve Bayesian uyum ölçütleri yaygın olarak kullanılmaktadır (Dhanavanthan, 2000; Yeşilova, 2003).

Bu çalışmada; normal karışımlı model yaklaşımı kullanılarak 2013 yılında, Türkiye'deki kişilerin mutluluk düzeylerinin illere göre sınıflandırılmasının yapılması amaçlanmıştır. Böylece her bir mutluluk düzeyi bakımından illerin kendi aralarında gruplandırılması yapılmış ve elde edilen her bir alt grup için mutluluk düzeyleri ortalama ve varyans gibi istatistikler tahmin edilmiştir.

2. KAYNAK BİLDİRİŞLERİ

Karışımlı modellerin gelişimi, on dokuzuncu yüzyıla kadar uzanmaktadır. Karışımlı modeller geniş bir kullanım alanına sahip olduğundan dolayı çok yaygın bir kullanım alanına sahiptir. Karışımlı model yaklaşımı özellikle; biyoloji, tıp, fizik, ekonomi ve pazarlama alanlarında yaygın olarak kullanılmaktadır. Karışımlı modeller, veri setlerini benzerliklerine göre farklı alt gruplara ayırabilen bir olasılık modelidir. Karışımlı modeller, özellikle değişkenler arasındaki ilişkileri tanımlamakda ve çoklu kaynaklardan toplanan verileri modellemede yararlıdır. Karışımlı modeller, kümeleme (clustering) esaslı sınıflandırma yöntemlerine model tabanlı bir yaklaşım getirmektedir. Ayrıca standart istatistiksel teori çerçevesinde hipotez testi ve tahmin yapmaya izin vermektedir. Karışımlı model yaklaşımı kümelemenin dışında birçok farklı probleme uyarlanabilen son derece esnek bir kümeleme algoritması sunmaktadır. Karışımlı modeller, temel populasyonların her birinde gözlemlerin dağılım fonksiyonunun özel bir formunu içeren istatistiksel modellerdir (Wedel ve DeSarbo, 2002; Leisch, 2004; McNicholas ve Murphy, 2008; Ng ve McLachlan, 2014; Anonim, 2016c).

Karışımlı modellerin kullanıldığı ilk analizlerden biri Karl Pearson tarafından 1894’ te yapılmıştır. Pearson tarafından analiz edilen veri seti Naples körfezinden n=1000 yengeç örneğinden alınan baş ve gövde uzunluğu oranı ölçümlerinden oluşmaktaydı. Veri setinin histogram grafiğindeki asimetrinin (çarpıklık) bir populasyonun iki yeni alt türüne ait olabileceği yönünde yorumlamıştır. Pearson bu veri seti için parametre tahmininde momentler yöntemini kullanmıştır (McLachlan ve Peel, 2000).

Populasyon içerisinde yer alan bireylerin farklı bir populasyondan gelme olasılıkları göz önünde bulundurulmalıdır. Bireyleri benzerliklerine göre alt sınıflarda toplayan gözlenemeyen sınıf (latent class) analizi, regresyon analizine göre daha kararlı parametre tahmininde bulunabilmektedir (Dhanavanthan, 2000; Duncan ve ark., 2002).

Son yıllarda istatistik alanındaki en büyük gelişmelerden biri; var olan durumu kullanıp, ileride olması beklenen tahminlerin yapılması yönündedir. Bu açıdan, toplumsal alandaki çalışmalara yönelik verilerin analiz edilmesinde ve veri setinin kendi içerisindeki heterojenliğinin belirlenmesinde gözlenemeyen sınıf analizi kullanılmaktadır. Gözlenemeyen sınıf analizi dâhilinde yapılan tanımlayıcı ve analitik istatistiklerin tümü genel olarak "karışımlı modelleme (mixture models)" olarak adlandırılmaktadır. Bir diğer benzer tanım yapılacak olursa, populasyon içindeki heterojenliği dikkate alan, homojen alt populasyonlar oluşturan ve ilk aşamada sayısı belli olmayan alt sınıflar (latent class) oluşturan istatistiklerden biri, karışımlı model olarak bilinmektedir (Everitt ve Hand, 1981; Titterington ve ark., 1985; Wang ve ark., 1998; Chen ve Kou, 2001; Muthen ve Muthen, 2002; Kayri ve Göktaş, 2005; Okut ve Kayri, 2008).

Karışımlı modelleme, veri setindeki gözlenemeyen heterojenliği (unobserved heteroginety) belirleyerek, örneğin kaç alt gruba ait olduğunu tespit etmekte ve her alt grup için ayrı ayrı parametre tahmini yapmaktır. Böylece bütün gözlenen değişkenler için tek bir parametre tahmini yapmak yerine her alt populasyon için ayrı parametre tahmini yapılmaktadır. O halde her alt grup için parametre tahmin değeri farklı olabilmektedir. Karışımlı modellerde önemli olan alt grup sayısının belirlenmesidir. Genelleştirilmiş olasılık oran testinin bu bağlamda sıklıkla kullanılması önerilmektedir (Wang ve Putterman 1998; Muthen ve Muthen, 2002; Okut ve ark., 2002; Yeşilova, 2003; Guang Sung, 2004; Editorial, 2007; Oberski, 2016).

Karışımlı modellerde amaç;

1) Karışımdaki alt populasyonların sayısı ve büyüklüğünü belirlemektir.
2) Gözlenen değişkenlerin, bu sınıflara düşme olasılıklarını tahmin etmektir.
3) Her bir sınıfın parametresini tahmin etmektedir (Okut ve ark., 2002; Yeşilova, 2003).

Normal karışımlı modellerde en temel amaç, doğru alt grup sayısını belirlemek yani, heterojen bir grupta homojen kaç tane alt grup oluşturulacağına karar vermektir (Guang Sung, 2004)

Doğrusal regresyonda karışımlı modellerin kullanılmasının en büyük avantajlarından biri, her bir gözlemin hangi olasılık ile hangi alt gruba girdiğinin saptanabilmesidir. Karışımlı doğrusal regresyon modellerinde parametre tahmini yapılırken, genellikle EM algoritmasını dikkate alan en çok olabilirlik yöntemi kullanılmaktadır (Fariaa ve Soromenhob, 2010).

Sonlu karışımlı modeller uzun yıllardan beri kullanılmaktadır. Bununla birlikte, mevcut bilgisayar gücündeki artış nedeniyle son on yılda kullanımında ciddi bir artış görülmüştür. Sonlu karışımlı modeller, gözlemlenmemiş heterojenliğin modellenmesinde kullanılmaktadır. Parametre tahmininde en çok olabilirlik yöntemi ile EM algoritması çok sık kullanılmaktadır. 1990' lı yıllarda, sonlu karışımlı modeller, standart doğrusal regresyon modellerinin yanı sıra genelleştirilmiş doğrusal modeller kombine edilerek genişletilmiştir. Belli sayıda bileşen içeren sonlu karışımlı modeller, genellikle bir maksimum olasılık çerçevesinde EM algoritması ile Bayesci bir çerçeve içinde Markov zinciri Monte Carlo (MCMC) örneklemesi ile tahmin edilmektedir (Leisch, 2004; Wengrzik, 2012; Grün, 2017).

Normal karışımlı modellerde (Gaussian mixture models=GMM) uygun model seçimi için bilgi ölçütleri kullanılmaktır. GMM' de alt grup sayısının küçük tutulmasına yardımcı olmak ve farklı alt grup modellerine sahip modeller arasında seçim yapmak için Akaiki Bilgi Kriteri (Akaike Information Criterion, AIC) ve Bayesian Bilgi Kriteri (Bayesian Information Criterion, BIC) ölçütleri önerilmektedir. Her ikisi de, optimize edilmiş, negatif log olasılığını esas alırak daha sonra modeldeki parametrelerin sayısını kullanarak hesaplama yapmaktadır. İyi bir uygulamada, uygun model seçimine karar verirken her iki ölçüte de bakılmaktadır. En küçük AIC veya BIC değerleri, en uygun modeli göstermektedir (McLachlan ve Peel, 2000; Bauer ve Curran, 2003; Marin ve ark., 2005; Lukociene ve Vermunt, 2010).

Normal karışımlı modeller normal dağılım gösteren verilerin sınıflandırılmasında kullanılan popüler bir tekniktir. GMM biyometrik sistemlerde, özellikle konuşmacı tanıma sistemlerinde, temsil etme kabiliyetleri nedeniyle sıklıkla kullanılmaktadır. Ayrıca zaman serilerinin modellemesinde yaygın olarak kullanılmıştır. GMM, Gaussian alt grup yoğunluklarının ağırlıklı bir toplamı olarak gösterilen parametrik olasılık yoğunluk fonksiyonudur (Shental ve ark., 2003; Reynolds, 2008; 2009; Eirola ve Lendasse, 2013).

Gaussian karışımlı modeller, doğrusal sonlu sayıda Gaussian dağılımından oluşmaktadır. Gaussian dağılımı, uygulanabilirliğinin yüksek olması nedeniyle genel olarak kullanılmaktadır. Bir yandan, tutarlı hesaplama özellikleri sayesinde matematiksel açıdan kolay bir analiz sağlar, diğer yandan, özellikle çok sayıda örnek ve bilinmeyen faktör olduğunda merkezi limit teoremine göre yaklaşık olarak hesaplamada daha uygundur. Bu sebeplerden dolayı GMM, gerçekçi problemleri tanımlamak için karışımlı modeli kullanarak Gaussian dağılımının uygulamalarını daha çok genişletmiştir. GMM, makine öğrenmesi problemleri içinde, eğiticisiz öğrenme için yaygın olarak kullanılmaktadır, çünkü veri örüntülerini (alt grupları) açığa çıkararak, benzer davranışlara sahip verileri kümeleyebilmektedir. GMM, sonlu sayıda Gaussian karışımlı yoğunluğuna göre dağıtılacak veriyi değerlendirmeye çalışan sonlu dağılıma bağımlı (parametrik) karışımlı bir modeldir. GMM' yi diğer modelleme yöntemlerinden ayıran en temel özellik, veri setine birkaç üretken Gaussian modelinin doğrusal kombinasyonunun sonucu olarak bakarken, diğer yöntemler ya bütün veriyi tanımlamak için tek bir olasılık dağılımı kullanır ya da bütün veri yapısını görmezden gelip yalnızca eldeki verilerin arasındaki ilişkiyi ortaya çıkarmaya çalışmaktadır. Normal dağılım birçok fiziksel deneyde oldukça tanımlayıcıdır ve karmaşık veri kaynaklarına izin verdiği için doğrusal karışımlı modellerin uygulanabilirliğine katkıda bulunmuştur. Merkezi limit teoreminin uygulandığı durumlara benzer olarak, veri setleri karmaşık ve bilinmeyen faktörler çok olduğunda GMM güçlü bir tanımlama özelliğine sahiptir (Murphy, 2006; Cao, 2010).

Latent sınıf analizi (LCA) ve Latent profil analizi (LPA) gözlenemeyen grupları gözlenen verilerden kurtarmayı amaçlar. Kümeleme analizine benzer ancak açık model tabanlı oldukları için daha esnektir. Latent sınıf analizi gözlenen ve latent değişkenlerin kesikli yapıda olduğu ya da sürekli ölçekleme düzeyinde olup normal dağılımı ve homojenlik gibi varsayımların sağlanamadığı durumlarda kullanmaya uygun güçlü bir alternatiftir. Eğitim, psikoloji gibi araştırmaların büyük bir kısmında, gözlenen değişkenlerden, latent yapılar incelenmektedir. Geleneksel yöntemler kullanıldığında, değişkenin kesikli yapıda olması, normal dağılım göstermemesi, homojen olmaması metodolojik problemlere yol açmaktadır. Latent sınıf modelleri için kullanılan EM algoritmasının formülize edilmesi ve bilgisayar yazılımlarında kullanılmaya başlanmasıyla, son yıllarda latent sınıf analiziyle yapılan çalışma sayısı oldukça artmıştır. Model parametrelerinin en çok olabilirlik tahminleri EM algoritması kullanılarak tahmin edilmektedir (Bailey ve Elkan, 1993; Zio ve ark., 2007; Güngör ve ark., 2015; Oberski, 2016).

Heterojenlik gösteren veri setlerinde karışımlı dağılımların kullanılması oldukça önemlidir. Normal dağılım gösteren veri setlerinin ortak bir varyansa sahip olan sonlu bir karışımı ile (veya çok değişkenli durumda kovaryans matrisi) herhangi bir sürekli dağılıma yaklaştığında, karışımlı modeller heterojenliği modellemek için uygun bir çerçeve sunmaktadır. Örnek olarak; Pribe (1994) log-normal dağılım gösteren n=10.000 büyüklüğündeki bir veri setinin yaklaşık olarak 30 normal karışımlı dağılım ile ifade edilebileceğini gösterilmiştir. Buna karşın, bir

Kernel yoğunluk tahminleyicisi 10.000 normal karışımlı dağılımını kullanmaktadır. Bir karışımlı model, karmaşık verilerin dağılımı için en uygun tahmini elde etmede kullanılmaktadır. Modellenen olaylardan uygun yorumlamalar, karışımlı model yaklaşımı ile yapılabilmaktedir. Bu esneklik karışımlı modellerin yapay sinir ağlarında faydalı bir rol oynamasına olanak tanımaktadır (McLachlan ve Peel, 2000; Constantinopoulos ve ark., 2006).

Latent sınıf dahilinde yapılan tanımlayıcı ve analitik istatistiklerin tümü genel olarak karışımlı modelleme olarak adlandırılmıştır. Bir diğer benzer tanım yapılacak olursa, populasyon içindeki heterojenliği dikkate alan, homojen alt gruplar oluşturan ve ilk aşamada sayısı belli olmayan alt gruplar oluşturan istatistiksel yöntemlerden biri, karışımlı model olarak bilinmektedir (Everitt ve Hand, 1981; Titterington ve ark., 1985; Wang ve ark., 1998; Chen ve Kou, 2001; Muthen ve Muthen, 2002; Okut ve Kayri, 2008). Populasyon içerisinde yer alan bireylerin farklı gruplardan gelme olasılıkları göz önünde bulundurulmalıdır. Bireyleri benzerliklerine göre alt sınıflarda toplayan latent sınıf analizi, regresyona göre daha kararlı parametre tahmininde bulunabilmektedir. Latent sınıf veya karışımlı regresyon analizinin temel amacı, regresyon parametrelerinin gruptan gruba farklılık gösterdiğini ortaya koymaktır. Karışımlı modeller, bireyler arasındaki latent değişkenlerdeki heterojenliği yakalamak için pazar araştırması ve epidemiyolojide yaygın olarak kullanılmaktadır (Dhanavanthan, 2000; Duncan ve ark., 2002; Jeroen, 2004; Okut ve Kayri, 2008; Ng ve McLachlan, 2014).

Normal karışımlı modellerin ana fikri, verilerin ortak yoğunluk için bir dizi GMM oluşturmak ve daha sonra her modelden koşullu yoğunluk fonksiyonları elde edilmektedir. GMM veri madenciliği, örüntü tanıma, makine öğrenimi gibi istatistiksel analizlerde yaygın olarak kullanılmaktadır. Bu tip uygulamalarda parametreler genellikle EM algoritması kullanılarak en çok olabilirlik yöntemi ile tahmin edilmiştir (Guang Sung, 2004; Faria and Soromenho, 2008; 2010; Arı ve ark., 2012).

Uygulamadaki iyi performansı ve arzu edilen analitik özelliklerinin bir sonucu olarak, Gaussian karışımlı regresyon modelleri, istatistik, mühendislik ve diğer alanlarda gittikçe artarak ilgi konusu haline gelmiştir (Bailey ve Elkan, 1993; Halbe ve Aladjem, 2005; Shi ve ark., 2005).

EM algoritması istatistikte en çok kullanılan algoritmalardan biri olup, özellikle eksik veri yapısı olan modellerde etkindir ve ayrıca karışımlı modellerde yaygın bir kullanım alanına sahiptir. Genellikle eksik verinin varlığı, en çok olabilirlik tahminlerin elde edilmesini zorlaştırmıştır. Bu durum karmaşık modellerden daha sık ortaya çıkmaktadır. Özellikle EM, çok sayıda bileşeni olan veya çok boyutlu olan karışımlı modellerle veya olabilirlik fonksiyonunun birçok yerel maksimum değerleri mevcut olduğunda risk altındadır. Nitekim uygun bir karışımlı modelin seçilmesi önemlidir. Özellikle, bir karışımlı modeldeki bileşen sayısının seçimi oldukça büyük etkiye sahiptir (Baudry ve Celeux, 2015).

EM algoritması, bir objenin hangi kümeye ait olduğunu belirlemede kesin mesafe ölçütlerini kullanmak yerine tahminsel ölçütleri kullanmayı tercih etmektedir. Son yıllarda birçok araştırmada kullanılan popüler bir yaklaşım olmuştur. Tam olmayan veri problemlerini çözmek için en çok olabilirlik tahminlerini elde eden iteratif bir algoritmadır. Her tekrarı iki adımda gerçekleşmektedir. Bu adımlar, bekleneni bulma (E Adımı) ve maksimizasyon (M Adımı) olarak adlandırılmaktadır. E-adımında gözlenen verilerin parametrelerine ait kestirimler kullanılarak eksik (missing) veri ile ilgili en iyi olasılıklar tahmin edilirken, M-adımında ise tahmin edilen eksik veri yerine konulup bütün veri üzerinden en çok olabilirlik hesaplanarak parametrelerin yeni tahminleri elde edilmektedir. Karışımlı modellerde başlangıçta veri setinin kaç alt gruba ait olduğu bilinmemektedir. Bu nedenle, EM algoritmasının E aşamasında, bilinmeyen alt populasyon sayısı eksik gözlem kabul edilerek buna yönelik tahminde bulunulurken, M aşamasında ise bu tahminleme maksimum yapılmaktadır (Dhanavanthan, 2000; Yeşilova, 2003; Montuelle ve ark., 2013; Sezgin ve Çelik, 2013; Erten, 2015; Oberski, 2016; Anonim, 2017a).

Karışımlı regresyon, ana heterojenliğini araştırmak için yararlı bir yaklaşım sağlamaktadır. Karışımlı regresyon, gözlenemeyen sınıflar arasında farklılık gösteren etkiyi bulmaya çalışmaktadır. Karışımlı regresyonun model varsayımlarına ve örneklem büyüklüklerine olan ihtiyaca güçlü bir şekilde bağlıdır (Horn ve ark., 2015).

Karışımlı modeller,

- Tek veya birden fazla grup analizi,
- Eksik gözlemlerin tahmin edilmesi,
- Karışık anket çalışmaları,
- Latent değişken etkileşimleri ve en çok olabilirlik kullanarak doğrusal olmayan faktör analizi,
- Rastgele eğimler,

- Bireysel değişkenli gözlem zamanları,
- Doğrusal veya doğrusal olmayan parametre kısıtlamaları,
- Tüm parametreler için en çok olabilirlik tahmini,
- Latent değişkenler arası kategorik analiz gibi durumlarda kullanılmaktadır (Anonim 2016b).

Ding ve Cody (2006) tarafından; eğitim ile ilgili yapılan bir araştırmada karışımlı regresyon modellerinin güçlü ve sınırlı olan yönleri ele alınmıştır. Karışımlı regresyon analizi, verilerdeki latent sınıfların tanımlanabilmesi ve regresyon parametre tahminlerinin her latent sınıf içerisinde farklılaşabilmesi açısından regresyon analizinden daha esnek olduğu belirtilmiştir.

Yeşilova ve Atlıhan (2007), tarafından yapılan çalışmada ise karışımlı Poisson Regresyon modelinin teorik özellikleri incelenerek, Bitki koruma alanında elde edilen bir veri kümesine uygulanmıştır. İlk olarak karışımlı Poisson regresyon modelin, veri kümesinin tek bir populasyondan elde edilmiş şekilde uyumu yapılarak, veri kümesindeki heterojenliği saptanmıştır. Daha sonrasında, söz konusu heterojenliği gidermek için elde edilen alt populasyonlara göre modelin uyumu yapılmış, her alt populasyon için parametre tahmini ve alt populasyona düşen bireylerin oranları elde edilmiştir.

Gross ve ark., (2000) ise; değişmeyen yüz tanıma özelliğini göstermek için bir yaklaşım önerilmiştir. İnsan yüzlerini karakterize etmek için farklı GMM ile model poz varyansını hesaplamak için karışımlı modeller kullanılmıştır. Her bir kişi için optimum karışımlı bileşenleri sayısı, karışım modellerini büyütmek suretiyle eğitim verilerinden otomatik olarak öğrenilmiştir. Sadece yüz kimliği ile etiketlenmiş girdi görüntülerini kullanan bir GMM kontrollü büyümesi için yeni bir algoritma önerilmiştir. Önerilen algoritma, bir toplantı odasında kaydedilen gerçek veriler üzerinde denenmiştir. Deneysel sonuçlar, toplantı salonunda toplanan verilerle elde edilen sonuçların, yeni algoritmanın geleneksel yaklaşımlardan daha iyi performans sergilediğini ortaya koymuştur.

Lindsay ve Basak (1993) çalışmalarında; çok değişkenli normal dağılımlı karışımları tanımlamak için örnek momentleri kullanma olanakları araştırılmış ve simülasyon ve Fisher' ın iris verilerine uygulanmıştır. Belirli bir moment denklem sistemi tasarlanmış ve daha sonra, bazı sınırlamaları içeren gerçek karışımlı dağılımını tanımlayan ve tutarlı tahminler sağlayan bir sistem olduğu gösterilmiştir. Dahası, tahminlerin herhangi bir boyutta hızla hesaplandığı ve olasılık fonksiyonunu en büyük yapan parametre değerlerine yakın olma açısından oldukça verimli olduğu gösterilmiştir.

DeSarbo ve Cron (1988), karışımlı regresyon modelini ML yöntemi ile incelemiş ve çalışmada kullandıkları gözlemlerin farklı sınıflara atanmalarını AIC ile gerçekleştirmiştir. Model, tek değişkenli normal yoğunlukların sonlu bir karışımıdır. Bu yoğunlukların beklentileri bir dizi açıklayıcı değişkenin doğrusal fonksiyonları olarak belirtilmiştir. Model, ticaret fuarı performansını etkileyen faktörlerin analiz edilmesinde ve ticaret fuar performansını değerlendirmede niteliklerinin önemine göre farklılaşan sınıfları araştırmak için kullanılmıştır. Çalışmalarında, 129 pazarlama yöneticisine firmalarının fuar performansını, sekiz performans faktörüne ve genel fuar performansına göre değerlendirmeleri istenmiştir. Bununla birlikte, AIC kullanılarak elde edilen karışımlı regresyon modeli, sırasıyla 59 ve 70 pazarlama yöneticisini içeren iki sınıf elde edilmiştir. Sınıf 1' deki yöneticiler, öncelikle, mevcut müşterilere hizmet vermeyi, kurumsal imajı ve ahlaki gelişmeyi içeren satış dışı faktörler açısından ticaret fuarlarını değerlendirirken, sınıf 2' deki yöneticiler ise, ticaret fuarlarını yeni beklentileri tanımlama, satış beklentileri, yeni ürün tanıtımı, şovlarda satış ve yeni ürün testleri gibi satış faktörleri üzerinde yoğunlaşmıştır.

Yeşilova ve ark. (2009) tarafından yapılan çalışmada, karışımlı model kullanılarak sekiz faklı ilçeden alınan ceviz meyvelerine ait meyve özellikleri bakımından, ilçelerin sınıflandırılması yapılmıştır. 8 farklı ilçe, incelenen ceviz meyve özellikleri bakımından 5 alt populasyonlu modele ayrılmıştır. Böylece her bir ilçenin farklı birer populasyon oluşturmadıkları, ancak beş alt populasyona ayrıldıkları saptanmıştır.

Yeşilova ve ark. (2016) Anadolu mandasının laktasyon süt verimine etkili olan faktörlerin belirlenmesine yönelik olarak, Gaussian karışımlı regresyon modelinin kullanıldığı çalışmada, Bitlis iline bağlı 7 farklı yerleşim yerinde yetiştirilen 1460 anaç Anadolu mandasının 2013 yılına ait laktasyon süt verimine ilişkin kayıtlar kullanılmıştır. Gaussian karışımlı regresyon modelinde AIC ve BIC ölçütlerine göre, Anadolu mandasının laktasyon süt verimi bakımından, yaş, köy ve laktasyon süresi gibi çevre faktörleri göz önünde bulundurulduğunda, 7 ayrı köydeki mandaların homojen 3 alt grup oluşturulabileceği saptanmıştır. Gaussian karışımlı regresyon aracılığıyla kantitatif karakterlerde sınıflandırma yapılırken çevre faktörlerine göre düzeltme yapılabilmesi hayvan ıslahı uygulamalarında önemlidir. Aynı zamanda, kendi içlerinde homojen ve kendi aralarında heterojen olarak oluşturulan her alt grup için ayrı regresyon analizi yapması, diğer çok değişkenli istatistik yöntemlerden ayıran önemli bir özelliktir.

Yeşilova ve ark. (2016) tarafından yapılan çalışmada; Van Gölü sahil şeridinde 2005–2006 yıllarında yürütülmüş ve su içinden örneklenen Notonecta viridis (Hemiptera: Notonectidae) populasyon yoğunluklarının suyun fiziko-kimyasal yapısının etkisi altında farklı örnekleme istasyonlarına göre dağılımları sayıma dayalı olarak elde edilmiştir. Burada, bağımlı değişken örnekleme istasyonlarında elde edilen böcek yoğunlukları, bağımsız değişkenler olarak suyun fiziko-kimyasal özellikleri alınarak karışımlı Poisson regresyon analizi uygulanmıştır. Analiz sonucunda, 20 farklı örnekleme istasyonundan elde edilen böcek yoğunluklarının, bağımsız değişkenler esas alındığında, kendi içlerinde homojen ve kendi aralarında heterojen olmak üzere sekiz alt gruba ayrıldıkları saptanmıştır.

Normal karışımlı modeller için BIC, Laplace metodu kullanılarak bütünleşmiş logaritmik olabilirliğin iki katına yakın bir değer olarak türetilir, ancak genel olarak gerekli düzenlilik koşulları karışımlı modelleri için geçerli değildir. Bununla birlikte, Roeder ve Wasserman (1997); tarafında yapılan çalışmada BIC' nin; karışımlı yoğunluğunun tutarlı bir tahmincisine götürdüğünü göstermiştir ve Keribin (2000) ise; bir karışımlı modelinde bileşen sayısını seçmek için BIC' nin tutarlı olduğunu göstermiştir.

Steele ve Raftery (2009) tarafından yapılan birçok farklı çalışmada, AIC test istatistiği için gerekli bulunan koşulların bozulmasından dolayı karışımlı modeller için alt grup sayısı fazla tahmin edilmiştir. Ayrıca, AIC ve BIC karşılaştırıldığında, AIC' nin modeldeki parametrelerin fazla olması durumunda daha çok karışımlı alt grupların elde edilmesi söz konusu olduğunu ifade etmişlerdir.

Bilinmeyen parametre vektörü olan $\hat{\theta}$ için başlangıç değer seçimi, parametre tahmini üzerinde büyük bir etkiye sahiptir. Çoğu durumda ortalamayı kullanmasına rağmen, bazı durumlarda en büyük mod tercih edilmektedir. Sonlu ortalama (ve sonlu medyan), karışımlı yoğunluğunun çok modlu olması durumunda doğru olmayan sonuçlar vermektedir. EM yönteminin yakınsamasının çok yavaş olabileceği durumlarda, makul başlangıç noktaları seçmek önemlidir. Bazıları verileri standart kümeleme yöntemlerini kullanarak kümelemek ve daha sonra küme oranlarını, ortalamalarını ve kovaryans matrislerini bir başlangıç noktası olarak kullanmayı tavsiye etmektedir. Parametre tahmini için standart hataların hesaplanması ile ilgili olarak en pratik yaklaşım, Quasi-Newton olabilirliğini son EM iteratif değerinden başlatmak ve Hessian tahmini ile ML tahminin örnekleme dağılımına, standart asimptotik normal yaklaşımı için kullanılabilecek yaklaşık bir bilgi matrisi değeri oluşturmaktır (McLachlan ve Peel, 2000; Bashir, 2003; Nasios ve Bors, 2006; Baudry ve Celeux, 2015; Grosseve Srivastava, 2017).

Algoritma tabanlı çalışmalarda başlangıç değerlerinin seçimi büyük önem taşımaktadır. Çünkü bu değerler algoritmanın yakınsama hızını ve küresel maksimum değerini oluşturma kabiliyetini büyük oranda etkilemektedir. Laird (1978) tarafından; başlangıç değerlerini ayarlamak için bir ağ araştırması önerilmiştir. Leroux (1992), tarafından yapılan çalışmada ise başlangıç değeri olarak kullanılan kümeleri oluşturmak için ek bilgi kullanımı önerilmiştir. McLachlan (1988) tarafından yapılan çalışmada; çok değişkenli karışımlar için başlangıç değerlerini seçmek amacıyla temel bileşen analizinin kullanılması önerilmiştir. Farklı bir kümeleme fikri Woodward ve ark. (1984) tarafından anlatılmıştır.

Finch ve ark. (1989) tarafından yapılan çalışmada; iki alt gruplu normal bir karışım için, karışım oranının yalnızca başlangıç değerine sahip olması gerektiği öne sürülmüştür. Diğer kalan parametreler, bu değer üzerinden otomatik olarak tahmin edilebilecektir. Onların fikri, karışma oranı p olduğunda, modelin iki parçaya ayrıldığı ve birincisinin, karışımın birinci bileşeni olarak kabul edilen ilk $[n, p]$ gözlemleri olduğunu ve ikincisinin ise ikinci gözlem grubuna ait olduğunu kabul eden ve geri kalan gözlemleri de içeren iki bölümü içermektedir. Modelin; ilk bölümündeki gözlemlerin ortalaması, karışımın birinci bileşeninin ortalaması için başlangıç değeri olarak kullanılırken, ikinci bölümün ortalaması, ikinci bileşenin ortalaması için başlangıç değeri olarak kullanılmıştır. Atwood ve ark. (1992) tarafından; gruplar halinde verilerin farklı bölümlere dayalı olarak p' nin 5 farklı olası seçimi incelenmiştir.

Bohning ve ark. (1994) çalışmalarında, algoritmanın daha hızlı yakınsaması için iyi başlangıç değerleri ile başlamayı önermiştir. Bu durum, simülasyon çalışmalarıyla doğrulanmıştır. Ancak sonlu normal karışım durumunda, nispeten daha küçük bir başlangıç değişkeni ile başlamak gerekmektedir. Sonlu Poisson karışımlı modellerde ortalamaya karşılık gelen parametre değerinin değişkeni de belirlediği durumlarda, bu tür "iyi ayrılmış" başlangıç bileşenlerini bulmanın kolay olmadığı önerilmişti.

Bir başka doğal tercih, moment yöntemi gibi farklı tahmin yöntemleri ile elde edilen tahminlerle başlamaktır. Furman ve Lindsay (1994a,b) ve Lindsay ve Basak (1993) ise bu başlangıç değerlerini normal karışımlar için göz önünde bulundurmuştur. Fowlkes (1979), tarafından yapılan çalışmada ise normal karışımlılarda, başlangıç değerlerini seçmek için bazı grafiksel yöntemler önerilmiştir. Seidel ve ark. (2000a, b, c) sonlu üslü karışımların başlangıç değerleri için bazı farklı seçenekleri incelenmiştir. Küresel maksimumun elde edilmesini sağlamak için birkaç farklı başlangıç değerinden başlamanın tercih edildiğini göstermiştir. Bohning (1999), birkaç farklı başlangıç değeri bulmak için bir strateji olarak büyük bir parametre alanı üzerinde bir ağ araştırması önerilmiştir (Karlis ve Xekalaki, 2003).

Nasios ve Bors, (2006) çalışmalarında; EM algoritmasının ortaya çıkmasından bu yana, en çok olabilirlik (ML), karışımlı dağılımlarının uyarlanması için en yaygın kullanılan yaklaşım olmuştur. Maksimum olasılık algoritmaları model parametrelerini tahmin etmektedir. Böylece olasılık fonksiyonunu en üst düzeye çıkarmıştır.

En çok olabilirlik tahmininin amacı tutarlı ve asimptotik olarak etkili olan olasılık denkleminin kök dizisini tanımlamak için her ψ parametre düzeyi için bir tahmini esas almaktadır. Böyle bir dizinin, uygun teorik koşullar altında var olduğu bilinmektedir. Tahmin edilen bu kökler parametre alanının içindeki yerel maksimuma karşılık gelir. Bu tutarlı kök dizisi esasen benzersizdir. Genel olarak tahmin modelleri için olasılık genellikle parametre alanının içinde genel bir maksimuma sahiptir. Ardından tipik olarak, istenen asimptotik özelliklere sahip olasılık denkleminin kök dizisi, her bir n’ nin global olarak olasılık fonksiyonunu L (ψ) en üst düzeye çıkaran kök olan ψ ' yi esas alarak sağlanmaktadır. Olabilirlik fonksiyonun maksimize yapılamadığı durumlarda bile ψ parametre vektörü en çok olabilirlik tahmini esas almaktadır. Böyle durumlarda, karışımlı modellerde hesaplanan olasılıklar sonsuz olabilmektedir. Bununla birlikte, bu tip modellerde gerekli koşullar altında; tutarlılık, etkinlik ve asimptotik normal dağılım özelliklerine sahip yerel maksimuma karşılık gelen olasılık denkleminin kök dizileri halen mevcut olabilmektedir (McLachlan ve Peel, 2000; Leisch, 2004; Wengrzik, 2012).

Normal karışımlı modellerde istenen sonuçların elde edilmesi, parametre tahmin yöntemlerine bağlıdır. Bazı çok değişkenli yoğunluklu modellerde oldukça fazla sayıda bileşen (alt grup) olması ve her bileşenin çok sayıda parametreye sahip olması gerektiği göz önüne alındığında, çok yüksek boyutlu tahmin yapabilen yöntemler gerektmektedir. Dahası, uygun olmayan bir tahmin yöntemi, özellikle normal karışımlı modelleri için sorunlu olmaktadır (McLachlan ve ark., 1997; 1998; McLachlan ve Peel, 2000; Verbeek ve ark., 2003; Reynolds, 2009; Rodriguez, 2011; Plasse, 2013).

Normal karışımlı modeller iki nedenden dolayı caziptir. Bunlar;

1. Tek değişkenli ve çok değişkenli normal dağılımlara uyumunun oldukça iyi olması
2. Çok fazla sayıda ve karmaşık alt gruba gerek kalmadan, yalnızca birkaç alt grup ile normal dağılım gösteren verileri modellemesidir (Murphy, 2006; Cao, 2010).

Çok değişkenli normal dağılım karışımlı formda tanımlamak, elde edilecek parametrelerin yorumlanması açısından önemlidir. Her bir gözlem için π_k olasılıklı K muhtemel sonuçlu çok değişkenli bir dağılımdan z_i gözlenemeyen değişkeni elde edilmektedir. Çok terimli değişken olan π_k çok değişkenli normal dağılım alt gruplarının belirlenmesini sağlamaktadır. Bu durumda z_i ve y_i değişkenlerinin modellenmesi aşağıdaki gibi olmaktadır (McLachlan ve Peel, 2000).

$$z_i \approx MN(\pi)$$
$$y_i \approx N(\mu z_i, \sum z_i)$$

Araştırmacılar ve uygulamacılar arasında Karışım modelleme için Bayes yaklaşımları büyük ilgi görmektedir. Bayes yöntemleri sonlu karışımlı modelde alt grup sayısını seçmek için yaygın olarak kullanılmaktadır. Bayes çerçevesindeki tahmin, son zamanlarda geliştirilen Markov zincir Monte Carlo (MCMC) yöntemleri ile posterior dağılımların tahmini için simülasyon tekniği yaygın olarak kullanılmaktadır. Karışımlı modeller için Bayes tahmin ediciler, ön (prior) dağılımlar doğru tanımlandığı sürece iyi sonuçlar vermektedir. Bununla birlikte, karışımlılar için Bayesci yaklaşımın uygulanmasının pratik olmasının nedeni Tanner ve Wong (1987) ile Gelfand ve Smith (1990) tarafından önerilen MCMC yöntemlerinin geliştirilmesiyle olmuştur. Karışımlı modelleme sorununun ortaya çıkmasından neredeyse 100 yıl sonra, Pearson (1894)'nın klasik karışımlı model yaklaşımlarıyla birlikte gerçek bir Bayesci çözümün ortaya çıktığı görülmüştür (Marin ve ark., 2005; Steele ve Rafter, 2009; McLachlan ve Peel, 2000; Bashir, 2003). Bununla birlikte, bayesian parametrik olmayan karışımlı modelleri ise kompleks verileri tanımlamak için önemli bir çerçeve oluşturmaktadır (Dahua, 2013).

Everitt (1984), tarafından yapılan çalışmada iki normal dağılımın üç çeşit karışımında maksimum olasılık tahminini elde etmek için bir çok algoritmayı karşılaştırmıştır. Böylece tahmin sonuçlarını araştırmak için her durumda üç farklı başlangıç değeri kullanılmıştır. Çoğu veri setinde her algoritmayı çeşitli başlangıç değerlerinden gelen olasılık fonksiyonun aynı yerel maksimumuna yakınsamıştır. Bununla birlikte bazı durumlarda farklı sonuçların meydana gelebileceği saptanmıştır. Uygulamada, olasılık durumunu araştıran farklı başlangıç noktaları seçilebilir ve bir maksimumu kabul etmeye veya devam etmeye karar verebilmektedir. Güvenilir başlangıç değerlerinin nasıl bulunacağı sorusu çok değişkenli veri durumunda değerlendirilmiştir. Başlangıçta kümelerin varlığı ve verileri araştırmak için iki boyutlu bir dağılım planı kullanılması önerilmiştir. Bu şekilde elde edilen verilerin görsel olarak kümelenmesi başlangıç değerlerine yansıtılabilmektedir. Bu tür grafiksel değerlendirmelere dayanarak, başlangıçtaki tahmin sözlü olarak yapılabilmektedir. Karlis ve Xekalaki (2003) tarafından; EMalgoritmasının başlangıç değerlerini seçmek için çeşitli yöntemler karşılaştırılmıştır. Bununla birlikte, Finch (1989) tarafından yapılan çalışmada; iki bileşenli Gaussian karışımı için, karışım oranlarının başlangıç değeri olarak verilmesi gerektiğini öne süren bir yaklaşım analiz edilmiştir.

Belirli numune büyüklükleri, başlangıç değerleri ve aralık genişlikleri verildiğinde, önerilen algoritmalar farklı karışımlı modellerine göre karşılaştırılmıştır. Böylece, her bir parametrenin hem doğruluğunu hem de hassasiyetini ölçmek için karşılaştırma kriteri olarak relatif karesel ortalama hata (rRMSE, relative root mean squared error) kullanılmaktadır. Sonlu karışımların olasılık fonksiyonu birkaç maksimuma sahip olabileceğinden, genel maksimuma yakınsama garanti edilememektedir. Aslında, yakınsama hiç garanti edilememektedir ve hatta birçok numune ve yöntemde ıraksama görülmektedir. Dolayısıyla, yakınsama başarısızlık oranı ikinci karşılaştırma kriteri olarak seçilmektedir. Dahası, bu sorunun üstesinden gelmek için, numune sayısının büyüklüğü performansı artırabileceğinden, farklı numune büyüklükleri incelenmiştir (McLachlan ve Peel, 2000).

Yakınsama hatalarını çözmenin diğer bir yolu farklı başlangıç değerleri kullanmaktır. Genel maksimumdan çok farklı değerlerle başlamak, yerel maksimuma yakınsamaya, hatta ıraksamaya neden olabilmektedir. Aynı şekilde, farklı başlangıç değerleri farklı tahminlere yol açabilmektedir. Bu nedenle, uygun bir başlangıç değerinin seçilmesi kaçınılmazdır. Bireysel gözlemli karışımlar için EM algoritmasında kullanılacak başlangıç değerlerini elde etmek için bazı teknikler bulunmuştur ve bunların dördü, gruplanmış gözlemlerde kullanılmak üzere değiştirilmeye çalışılmıştır. Bu değişikliklerin bütün karışımlı modelleri için uygun olmadığı ortaya çıktığından, çok daha fazla sayıda karışımlı modelini kapsayacak şekilde yeni bir strateji geliştirilmiştir (McLachlan ve Peel, 2000; Nasios ve Bors, 2006; Grosseve Srivastava, 2017).

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu çalışmada, Türkiye İstatistik Kurumu tarafından 2013 yılında ilk defa il düzeyinde yapılan, "Yaşam Memnuniyeti Araştırması" sonuçlarına göre illere ait mutluluk kaynağı göstergelerinde 10 değişkene yer verilmiştir. Araştırma verileri, mutluluk kaynağı olan kişiler ve mutluluk kaynağı olan değerlerden oluşmaktadır. Çocuklar, Eş, Baba, Kendisi, ve Torunlar mutluluk kaynağı olan kişileri, Sağlık, Sevgi, Başarı, Para ve İş ise mutluluk kaynağı oluşturan değerleri ifade etmektedir.

Çizelge 3.1. İllere ait mutluluk kaynağı göstergeleri

Mutluluk kaynağı olan kişiler					Mutluluk kaynağı olan değerler				
Çocuklar	Eş	Anne/baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş

TÜİK tarafından yapılan araştırmada, Türkiye Cumhuriyeti sınırları içinde bulunan hanelerde yaşayan 18 ve daha yukarı yaştaki Türkiye Cumhuriyeti vatandaşları ile yabancı uyruklu kişileri kapsamakta olup araştırmada, kurumsal nüfus (üniversite öğrenci yurdu, huzurevi, bakımevi, ceza ve tutukevi, ıslahevi, yetiştirme yurdu, askeri birlik ve kışla vb.) kapsam dışında tutulmaktadır. 2013 yılında araştırmanın örneklem büyüklüğü, İstatistiki Bölge Birimleri Sınıflaması (İBBS) Düzey 3 (81 il) bazında tahmin üretecek şekilde hesaplanmıştır. Araştırma kapsamında 2013 yılında Türkiye genelinde 125.720 Haneye gidilerek, 18 ve daha yukarı yaştaki 196.203 birey ile yüz yüze görüşmeler yapılmıştır.

3.2. Yöntem

3.2.1 Karışımlı modellerin genel yapısı

3.2.1.1. Karışımlı modeller için temel gösterimler

$Y_1,...,Y_n$, n boyutlu rasgele bir örneklemi göstersin ve Y_j (j=1,2,…,n) ise $\Re^p$ üzerinde olasılık yoğunluk fonksiyonu $f(y_j)$ olan bir p-boyutlu rasgele vektör olsun. Y_j, incelenen olayın bazı özelliklerinin j'inci durumunda yapılan *p* ölçümlerine karşılık gelen rasgele değişkenleri içermekte olup, $Y=(Y_1^T,...,Y_n^T)^T$ biçiminde yazılabilir. Burada üst simge olarak verilen T vektörün transpozunu göstermektedir. Y'yi tüm örneklemleri temsil eden kitle için kullanılmaktadır. Başka bir ifadeyle, Y, $\Re^p$ 'deki bir n-nokta sayısıdır. Bu nedenle, söz konusu kitleden çekilen rasgele bir vektörün gerçekleştirilmesini küçük y harfi gösterilir. Örneğin, $y=(y_1^T,...,y_n^T)^T$ gözlemlenen rasgele bir örneği göstersin, burada y_j, rasgele vektör Y_j 'nin gözlemlenen değeridir. Burada özellik Y_j vektörünü sürekli bir rasgele vektör olarak almamıza rağmen, Y_j sayıma dayalı bir değişken olarak kabul edildiğinde ve birbirinden bağımsız olduğu varsayıldığında, $f(y_j)$ bir olsılık fonksiyonu olur. Böylece y_j' nin karışımlı fonksiyonu formu;

$$f(y_j)=\sum_{i=1}^{k}\pi_i f_i(y_j),\; j=1.2,...,n \tag{1}$$

biçiminde verilebilir. Burada $f_i(y_j)$ bir olasılık yoğunluk fonksiyonunu, k alt grup sayısını ve π_i toplamı bir olan negatif olmayan bir niceliği göstermekte olup aşağıdaki gibi yazılabilir.

$$0 \le \pi_i \le 1 \quad (i = 1,\ldots,\mathrm{k}) \text{ ve } \sum_{i=1}^{k} \pi_i = 1$$

Burada $\pi_1,\ldots,\pi_k$ karışma olasılığını (Mixing probability) veya ağırlıkları, $f_1(y_j),\ldots,f_k(y_j)$ fonksiyonları birer olasılık yoğunluk fonksiyonunu ifade eder ve dolayısıyla eşitlik 1'in bir yoğunluk fonksiyonun tanımladığı açıktır. Ayrıca eşitlik 1' de verilen $f_i(y_j)$, karışımlı bileşen yoğunlukları olarak adlandırılır. Eşitlik 1' de *k*, karışımlı olasılık yoğunluk fonksiyonun alt grup sayısını (bileşen sayısını) ifade eder. Bununla birlikte Eşitlik 1 yoğunluğuna k-alt gruplu sonlu karışımlı yoğunluğu ve karşılık gelen dağılım fonksiyonu olan $f(y_j)$' ye bir k alt gruplu sonlu karışımlı dağılımı olarak ifade edilir. Karışım modelinin bu formülasyonunda, alt grup sayısı k sabit olarak kabul edilir (McLachlan ve Peel, 2000; Wedel ve DeSarbo, 2002; Barber, 2012; Anonim 2017b).

3.2.2. Sonlu karışımlı modeller

Bir sonlu karışımlı model iki veya daha fazla olasılık yoğunluk fonksiyonunun kombinasyonudur. Karışımlı modeller, özel olasılık yoğunluk fonksiyonlarının özelliklerini birleştirerek, herhangi bir rasgele dağılıma yaklaşım mümkündür. Sonlu karışımlı modeller, karmaşık verileri modellemek için güçlü ve esnek bir araçtır. Sonlu karışımlı modelini formüle etmek için, N büyüklüğünde bir örneğin alındığını varsayalım. Her denekte, K değişkenleri $y_n = (y_{nk}, n = 1,..., N; k = 1,..., K)$ ölçülür. Bu deneklerin, K bilinmeyen alt grupların (sınıfların) karışımı olan bir kitleden bilinmeyen $\pi_1, \ldots, \pi_K$ oranlarda ortaya çıktığı kabul edilir. Belli bir deneğin hangi alt gruptan geldiği önceden bilinmemektedir. y_{nk} ' nın k sınıfından geldiği göz önüne alındığında, y_n ölçümlerinin vektörünün dağılım fonksiyonu $f_k(y_n|\theta_k)$ genel formu ile temsil edilir. Burada, θ_k, k alt grubu için bütün bilinmeyen parametrelerin vektörünü belirtir. Örneğin, her alt gruptaki y_{nk} ' nın bağımsız normal dağılımlı olması durumunda θ_t, K alt gruplarının her birindeki normal dağılıma ait ortalama (μ_k)ve varyansı σ_k^2 içermektedir. Karışımlı dağılımların, koşullu dağılım formu aşağıdaki gibi yazılabilir.

$$f(y_n|\varphi) = \sum_{k}^{K} \pi_k f_k(y_n|\theta_k), \tag{2}$$

Burada $\varphi = (\pi, \theta)$ modelin tüm parametrelerini göstermektedir. Eşitlik 2'de görüldüğü gibi; verilen k' nın koşullu olasılığının ile k 'nın olasılığı ile çarpımı ve bunun k' nın olası tüm değerleri için elde edilerek toplanması gerekmektedir.

Koşullu yoğunluk fonksiyonu olan $f_k(y_n|\theta_k)$; normal, Poisson ve binom dağılım fonksiyonlarının yanı sıra negatif binom, üstel, gamma ve benzeri diğer iyi bilinen dağılım fonksiyonları da dâhil olmak üzere birçok formunda yazılmaktadır. Daha yaygın şekilde kullanılan bu dağılımların tümü, üstel dağılım ailesinin belirli üyelerini temsil eder. Bu aile hem nümerik hem de kategorik dağılımları kapsayan genel bir yapıdır(Wedel ve DeSarbo, 2002; Anonim, 2016a).

$Z, j = 1,..., k$ değerleri sırasıyla $\pi_1, \ldots, \pi_k$ olasılıklarını alan kategorik rasgele bir değişken olsun ve Y'nin koşullu yoğunluğunun $Z_j = i$ olarak verilen $f_i(y_j)$ $(i = 1, \ldots, \mathrm{k})$ olduğunu varsayalım. Y_j ' nin koşulsuz yoğunluğu olan marjinal yoğunluk fonksiyonu $f(y_j)$ ile verilir. Z_j, k boyutlu bileşen etiket vektörü olmak üzere i. elemanı $Z_j, Z_{ij} = (Z_j)_i$ olup, j. rasgele değişken vektörü Y_j karışımlı modelin i. bileşeninden elde edilmiş ise 1 diğer durumlarda ise 0 değerini alır. Z_j, olasılıkları $\pi_1, \ldots, \pi_k$ olan bir k kategorilerinin üzerine çok terimli (multinominal) bir dağılıma göre dağıtılır; Yani,

$$pr\{Z_j \approx z_j\} = \pi_1^{z_{1J}} \pi_2^{z_{2J}} \ldots \pi_g^{z_{gJ}}$$

(3)

biçiminde yazılabilir ve

$\pi = (\pi_1, \ldots, \pi_k)^T$ olduğu durumda şöyle yazarız;

$$Z_J \sim Multino\min al_k(1, \pi)$$

(McLachlan ve Peel, 2000).

3.2.3. Sonlu karışımlı modeller ile sonsuz karışımlı modeller arasındaki farklılıklar

Sonlu karışımlı modeller, veri kümesinde sınırlı sayıda alt grubun olduğunu varsaymaktadır. Bununla birlikte, sonsuz karışımlı modellemede ise, veri setinin sonsuz alt gruptan oluştuğu varsayılmaktadır. Başka bir değişle k alt grup sayısı olmak üzere, sonlu karışımlı model aşağıdaki gibi yazılabilir (McLachlan ve Peel, 2000).

$$f(y|\varphi,\pi) = \sum_{k}^{K} \pi_k f(y|\varphi_k)$$

Eşitlikte,

$$\pi_k > 0 \text{ ve } \sum_{k=1}^{K} \pi_k = 1$$

olmaktadır. Sonsuz karışımlı modellemede, alt grup sayısı sınırlı sayıda olmadığı için değişim aralığı;

$$\sum_{k=1}^{\infty} \pi_k = 1$$

olmaktadır. Bu durumda sonsuz karışımlı model aşağıdaki gibi yazılabilir.

$$f(y|\varphi,\pi) = \sum_{k}^{\infty} \pi_k f(y|\varphi_k)$$

3.2.4. Karışımlı modelde en çok olabilirlik yöntemi ve EM algoritmasının genel yapısı

Gözlem değerleri Y_j d-boyutlu bir rasgele vektör ile gösterilsin. Y_j rasgele vektörün olasılık fonksiyonu (ya da olasılık yoğunluk fonksiyonu) $f(y_j,\psi)$ olsun. Burada, ψ bilinmeyen parametreleri göstersin. Bu durumda en çok olabilirlik yöntemi kullanılarak ψ 'nin tahmini olan $\hat{\psi}$ elde edilebilir. Bunun için ilk olarak $Lf(y_j,\psi)$ olabilirlik fonksiyonu yazılır ve daha sonra aşağıda verilen kısmi türev işlemleri kullanılarak bilinmeyen parametreler tahmin edilir.

$$\partial L(\psi)/\partial\psi = 0,$$

veya aynı şekilde;

$$\partial \log L(\psi)/\partial\psi = 0,$$

Burada;

$$L(\psi) = \prod_{j=1}^{n} f(y_j;\psi)$$

Bağımsız veri $y_1,...,y_n$ varsayımı altında oluşturulacak olasılık fonksiyonunu gösterir.

Doğrusal ya da doğrusal olmayan modellar için karışımlı model yaklaşımının kullanılması mümkündür. Karışımlı modellemede parametre tahmini için yaygın olarak kullanılan en çok olabilirlik tahminlerinin hesaplanmasında, birçok araştırıcı tek güvenilir yolunun

EM algoritması kullanması gerektiğini savunmaktadırlar. EM algoritması, bir "E-aşaması" ve bir de "M aşaması" (maksimizasyon) olmak üzere iki aşamadan oluşan iteratif bir prosedürdür. Model parametreleri göz önüne alındığında, E-aşamasında gözlenmemiş alt grupların belirlenmesi ve M aşamasında ise tam log olabilirlik fonksiyonunun maksimize edilmesi ile olası parametre tahminleri elde edilmektedir.
EM algoritması için log olabilirlik aşağıdaki gibi yazılabilir.

$$\log(L_c(H,\psi)) = \sum_{i=1}^{n}\sum_{k=1}^{K} h_{ik}(\log(\pi_i) + \log\varphi(y_i|\psi))$$

Eşitlikte; K alt grup sayısını, h_{ik} hangi gözlemin hangi alt gruba gittiğini göstermektedir. Eğer i gözlemi alt grup k'ya ait ise $h_{ik}=1$ dir. ψ tüm karışımlı parametrelerini içeren bir vektördür. $\pi,\psi \; k=1,...;K$

EM algoritması ψ' nin başlangıç değeri ψ^0' dan başlar ve sonraki aşamalar aşağıdaki gibi gerçekleşmektedir.

E aşaması: Gözlemlenmemiş h_{ik} değerlerine göre tam veri log olasılığını bekleneni alınarak ve parametreler için ψ^0 başlangıç değerleri kullanılarak log olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\mathrm{E}\left[\log(L_c(H,\psi))\right] = \sum_{i=1}^{n}\sum_{k=1}^{K} \mathrm{E}\left[h_{ik}\middle|\psi^0\right] \left(\log(\pi_i) + \log\varphi(y_i|\psi)\right)$$

Burada;

$$\mathrm{E}\left[h_{ik}\middle|\psi^0\right] = \tau_k(y_i|\psi^o)$$

$$\tau_k(y_i \mid \psi^0) = \pi_k^0 \varphi(y_i \mid \psi_k^0) \,/\, \sum_{j=1}^{K} \pi_j^0 \varphi(y_i \mid \psi_j^0)$$

olur.

M aşaması: karışımlı parametreleri $\hat{\psi}$ in yeni tahminlerini oluşturmak için tekrardan hesaplanan log olabilirlik olasılığının beklenen değeri maksimize edilir. Sonuçta, log olabilirlik fonksiyonun maksimize edilmesi ile ilgili parametrelerin (ψ) tahmini yapılmaktadır (McLachlan ve Peel, 2000).

$$\max_{\psi} Q(\psi \mid y) = \sum_{i=1}^{n} \sum_{k=1}^{K} \tau_k(y_i \mid \psi^0)(\log(\pi_i) + \log \varphi(y_i \mid \psi_k))$$

3.2.4.1. EM algoritmasının başlatılması

$\hat{\theta}$ parametre vektörü için başlangıç değer seçimi EM algoritması için belirleyici niteliktedir. Karışımlı parametrelerini tahmin ederken EM algoritması için başlangıç değerleri için farklı stratejiler önerilmiştir ve bunlar laten GOLD, Mplus ve R gibi programlarda Karışımlı yazılımlar mevcuttu (Bashir, 2003). Parametrelerin başlangıç değerleri için bazı yaklaşımlar aşağıdaki gibi verilebilir.

3.2.4.2. Rastgele seçim

EM algoritması yakınsanıncaya kadar parametreler için en yüksek olabilirliği sağlayan farklı rasgele başlangıç değerleri kullanılır. Laten GOLD ve Mplus gibi istatistik yazılım programları başlangıç değerleri için otomatik olarak başlangıç değer ataması yapmaktadır. Eğer EM algoritması için yakınsama sorunu yaşanırsa, araştırıcı başlangıç değerlerini kendisi belirleyebilir. Bazı durumlarda başlangıç değeri için sıfır değer ataması da yapılmaktadır (Baudry ve Celeux, 2015).

3.2.4.3. Küçük EM

Bu başlangıç değer atama yöntemi EM algoritmasının kısa zamanda yakınsaması için büyük bir sayı kullanması esasına dayanmaktadır. EM algoritmasının kısa zamanda sonuçlanmasında yakınsama beklenmeden, birkaç iterasyon sonrasında algoritmanın durdurulması amaçlanmaktadır. Daha sonra, EM algoritmasından kısa zamanda elde edilen en büyük olabilirliği sağlayan parametre değerinden EM değeri elde edilir (Baudry ve Celeux, 2015).

3.2.4.4. Yapısal eşitlik modellemesi

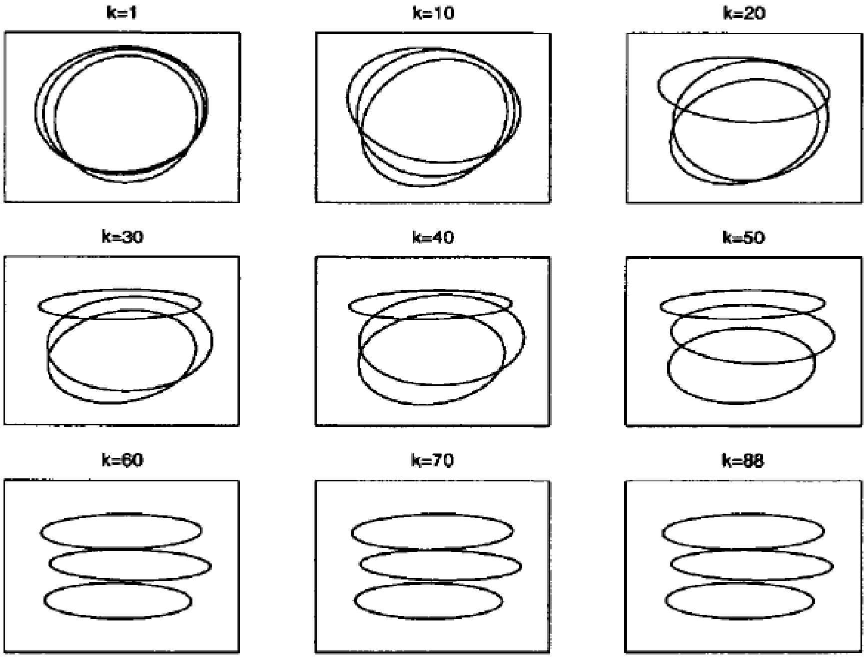
Yapısal Eşitlik Modellemesi (Structure Equation Model SEM); EM algoritmasının E ve M aşamaları arasında birleştirici bir stokastik algoritmadır. Bu algoritma bilinmeyen $z_i \quad i = 1, 2, \ldots, n$ bileşenlerini (karışımlı modelde alt gruplara ait olan), bunlara ait koşullu dağılımı kullanarak rasgele elde edilmesini sağlayan bir birleştirme işlemidir. SEM algoritması tek yinelemeli Monte Carlo EM algoritmasının aynısıdır. Bu algoritma, değişkenin başlangıç dağılımından bağımsız olan tek bir sabit dağılıma sahip Markov zinciri üretir. SEM; SEM algoritmasının uzun süre çalışması sonucunda elde edilen en geniş olabilirliği veren parametre değeri ile EM algoritmasının başlamasını içermektedir (Baudry ve Celeux, 2015).

3.2.4.5. Koşullu EM

Koşullu EM (Conditional EM=CEM), EM'nin E ve M aşamaları arasındaki sınıflandırma adımını içeren bir algoritmadır. Bu sınıflandırma adımı, geçerli π_{ik} $\left(k = 1, 2, \ldots, K\right)$ koşullu olasılık fonksiyonunu maksimize eden karışımlı bileşene (alt grup) her y_i gözlem değerini atama sürecinden oluşmaktadır. Bu algoritma, az sayıda yenileme ile yakınsama sağlamaktadır (Baudry ve Celeux, 2015).

3.2.5. Karışımlı model seçme ölçütleri

Karışımlı modellerde, her karışma bileşeni, gizli bir alt grup değişkeni gibi düşünülebileceğinden uygun alt grup sayısının seçimi ayrı bir problem olarak ele alınmaktadır. Alt grup sayısını belirlerken K için tahmini model parametrelerinin değerleri alınır. Bunun için en çok olabilirlik tahminlerini hesaplamak için EM algoritması kullanılır. Farklı alt grup sayısı gösterimleri Şekil 3.1' de verilmiştir.



Şekil 3.1. Rasgele başlangıç değerleri kullanılarak EM algoritması için farklı k değerleri için üretilen parametrelere dayalı her alt grup için asimptotik (% 95) elipsoidler grafiği (McLachlan ve Peel, 2000).

Asimptotik olarak doğru dağılımın Kullback-Leibler mesafesini, yani doğru model olma durumu, Bayesci çerçeve içerisinde bir ölçüt türetilmiştir. Gözlem sayısı sonsuza giderken, ortak olabilirliğin logaritmasının bir asimptotik yaklaşımı olup aşağıdaki gibi yazılabilir.

$$f_k(y) = \int f(y;\theta_k)\pi(\theta_k)d\theta_k$$

π , θ_k için bir ön (prior) dağılım olmak üzere

$$BIC(\mathrm{K}) = LogL(\theta_k) - \frac{D_K}{2}\log n, \qquad (4)$$

Eşitlik 4' de; D_k , θ_k 'nın boyutunu gösterir.

Karışımlı modellemenin amacı model tabanlı kümeleme ise tümleşik tam olabilirlik (Integrated Completed Likelihood=ICL) kriteri BIC' a göre tercih edilebilir. ICL, tam ortak olabilirlik logaritmasının asimptotik BIC benzeri bir yaklaşımıdır. ICL' nin genel formu aşağıdaki gibi yazılabilir.

$$ICL(K) = BIC(K) + ENT(K)\,, \tag{5}$$

Eşitlik 5' te,

$$ENT(K) = -\sum_{i=}^{n} \sum_{k=1}^{K} \hat{\pi}_{ik} \log \hat{\pi}_{ik}$$

olarak yazılabilir. Burada, $ENT(K)$ entropy değerini göstermektedir (Baudry ve Celeux, 2015).

3.2.5.1. Bayesian bilgi ölçütü

Bayesian Bilgi Kriteri (Bayesian Information Criterion, BIC), karışımlı modeller için yaygın olarak kullanılan olabilirliğe dayalı bir yaklaşım sağlamaktadır. Roeder ve Wasserman (1997) tarafından oluşturulan karışımlı modeller için kullanılmış ve özellikle veri setini alt gruplara ayırmada etkin bir ölçüttür. BIC' nin genel formu aşağıdaki gibi yazılabilir.

$$BIC = -\log L + d\log(n) \tag{6}$$

$$BIC = 2p(y \mid \hat{\tau}, G) - d\log(n),$$

Burada d karışımlı modelindeki parametrelerin sayısını göstermektedir (Steele ve Raftery, 2009).

3.2.5.2. Devians bilgi ölçütü

Devians Bilgi Ölçütü (Deviance Information Criterion=DIC), modeldeki gerçek serbest parametre sayısı yerine etkili model parametrelerinin sayısının kullanıldığı AIC benzeri bir olabilirlik esaslı ölçüttür. Model karşılaştırması için DIC' nin belirttiği bir amaç, seçilen modelin tahmini hatasını en aza indirgemektir. Aşağıdaki formda yazılabilir.

$$DIC(G) = -2\log\mathrm{L} + 2p_d \tag{7}$$

Eşitlik 7' de; p_d aşağıdaki gibi yazılabilir

$$p_D = E_{K|y}(\log p(y|\hat{\theta})) - \log p(y|\tau\hat{\theta})$$

Burada; $\hat{\theta}$, verilerin sonsal dağılımına (genellikle bir sonsal ortalama, medyan veya mod) göre yeterli bir tahmin edicidir (Steele ve Raftery, 2009).

3.2.5.3. Akaiki bilgi ölçütü

Akaiki Bilgi Ölçütü (Akaike Information Criterion= AIC), alt grup sayısını belirlemek için kullanılan bilgi ölçütlerinden en iyi bilineni Akaike bilgi ölçütüdür. AIC' nin genel formu aşağıdaki gibi yazılabilir.

$$AIC = -\log L + 2d \tag{8}$$

En küçük AIC değerine sahip model, gerçek modele olan Kullback-Leibler mesafesinde asimptotik olarak en yakın olmalıdır (Steele ve Raftery, 2009).

3.2.6. Normal karışımlı dağılımlar

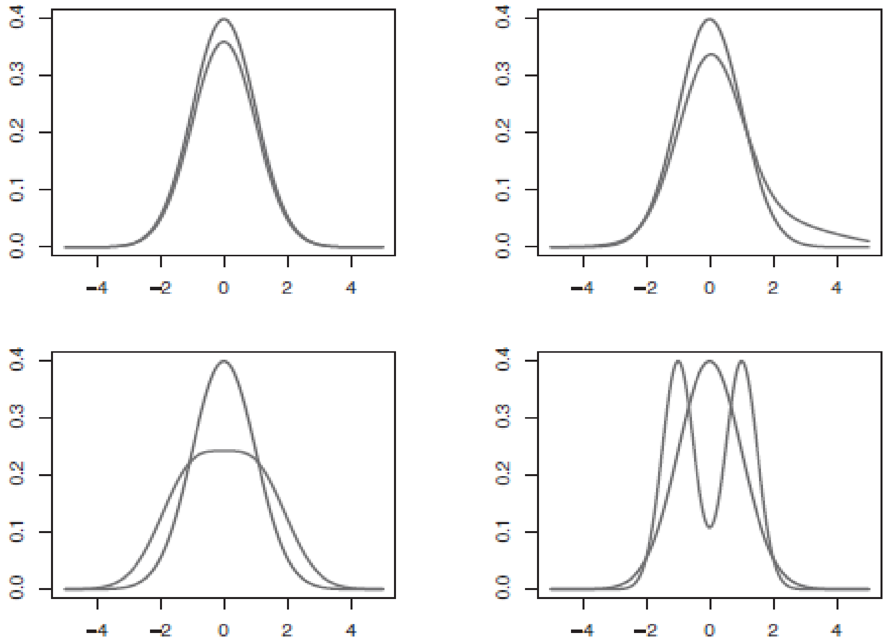
3.2.6.1. Tek değişkenli normal karışımlı dağılım

X rasgele değişkeni ortalaması μ ve varyans σ^2 olan bir normal dağılım göstermekte ve $x \sim N(\mu,\sigma^2)$ şeklinde gösterilmektedir. Gözlem değerlerinin K alt gruplu ortak dağılım fonksiyona sahip olsun. X rasgele değişkenin karışımlı tek değişkenli normal yoğunluk fonksiyonu $f(x|\pi,\mu,\sigma^2)$ şeklinde olup, aşağıdaki gibi yazılabilir.

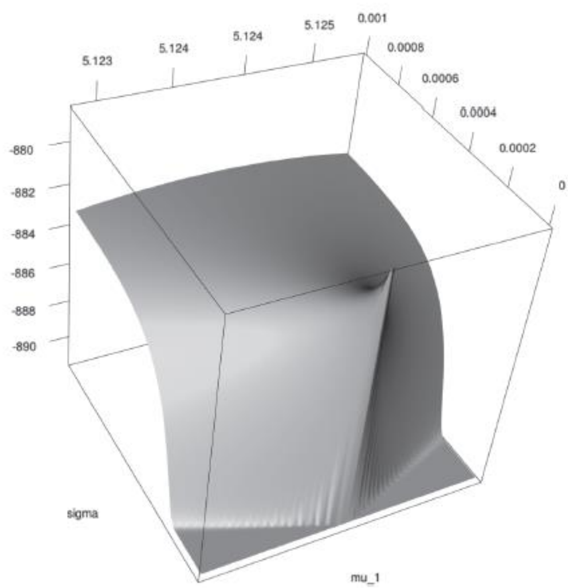
$$f(x|\pi,\mu,\sigma^2) = \prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k f_k(x_i|\mu_k,\sigma^2{}_k)$$

$$=\prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k \frac{1}{\sqrt{2\pi\sigma^2{}_k}}\exp\{\frac{-1}{2}\left(\frac{(x_i-\mu_k)^2}{\sigma_k}\right)\}$$

Eşitlikte $\sigma > 0, \pi_k > 0$ ve $\sum_{k=1}^{K}\pi_k = 1$ olmaktadır. π_k, x_i değişkenine ilişkin k' ıncı alt gruptaki karışma olasılığını göstermektedir.



Şekil 3.2. Tek Değişkenli Normal Karışımlılar (McLachlan ve Peel, 2000).



Şekil 3.3. Tek Değişkenli Normal Karışımlı Dağılımın Olabilirlik Fonksiyonunun Grafiği (McLachlan ve Peel, 2000).

3.2.6.2. İki değişkenli normal karışımlı dağılım

$X = (x_1,\ x_2)'$ Değişkenlerinden oluşan bir rasgele vektör olsun. $-\infty < x < +\infty$ olmak üzere;

Çok değişkenli dağılım gösteren $X = (x_1,\ x_2)'$ gözlem değerlerinin K alt gruplu ortak dağılım fonksiyona sahip olsun. $X = (x_1,\ x_2)'$ rasgele değişkenlerinin karışımlı çok değişkenli normal yoğunluk fonksiyonu $f(x|\pi,,\mu,\Sigma)$ şeklinde olup, aşağıdaki gibi yazılabilir.

$$f(x\left|\pi,\mu,\Sigma\right.) = \prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k f_k(x_i\left|\mu_k,\Sigma_k\right.)$$

$$\prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k\left|\Sigma_k\right|^{-1/2}\exp\{-\tfrac{1}{2}(x_i-\mu_k)'\Sigma_k^{-1}(x_i-\mu_k)\}$$

Eşitlikte,

$$\mu' = (\mu_1, \mu_2)' \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

$$i = 1, 2, \cdots, p, \qquad p = 2$$

olduğundan;

$\sigma_i > 0 \ \ ve \ \ |\sigma_{12}| < \sigma_1\sigma_2$'dir.

Eğer x_1 ve x_2 bağımsız rasgele değişkenler ise iki değişkenli normal dağılım değişkeni X_{2x1};

$x' = (x_1, x_2)$ $Z \sim \mathrm{N}(0,1)$ Olan Z değerlerinden yararlanılarak $x' = \sigma_1 \cdot Z + \mu_1, \sigma_2 \cdot Z + \mu_2$'den elde edilir. İki değişkenli normal dağılımda $\sum_{2x2}$ boyutlu olduğundan

$$\left|\sum\right| = \sigma_1^2\sigma_2^2 - \sigma_{12}\sigma_{12} \Rightarrow \sigma_{12} = g_{12}\sqrt{\sigma_{11}\sigma_{22}} \quad ise$$

$$\left|\sum\right| = \sigma_1^2\sigma_2^2 - g_{12}\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2\left(1 - g_{12}^2\right)$$

$$\sum{}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2}\left(1 - g_{12}^2\right)\cdot\begin{bmatrix} \sigma_2^2 & -g\sigma_1\sigma_2 \\ -g\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}'dir.$$

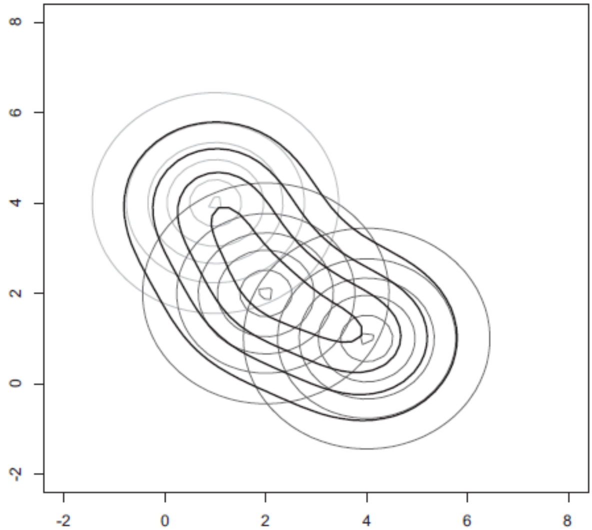
İki değişkenli normal dağılıma ait bazı özellikler;

1) x_1 *ve* x_2 Değişkenleri birlikte iki değişkenli normal dağılım gösteriyorlarsa, bu değişkenler tek tek de $x_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ $i = 1,2$ normal dağılım gösterir.

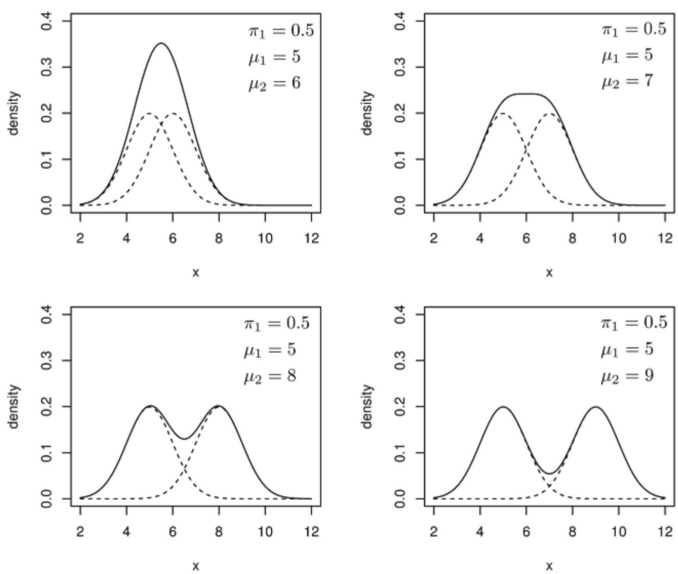
2) $|g| < 1$ Olmak şartıyla X_1'e ait şartlı dağılım $\mu = \mu_1 + g(\sigma_1/\sigma_2)(x_2 - \mu_2)$ olan ve $\sigma^2 = \sigma_1^2(1 - g_{12}^2)$ olan bir normal dağılım gösterir.

3) x_1 *ve* x_2 Değişkenleri ancak $g_{12} = 0$ durumunda bağımsızdır.

4) $\sum$ Pozitif tanımlı bir matrisdir. $\left|\sum\right| \neq 0$ *ve* $\sum_{pxp}$ Matrisinin rankı $r = p'dir.$



Şekil 3.4. İki Değişkenli Normal Karışımlılar (McLachlan ve Peel, 2000).



Şekil 3.5. $\sigma_1^2 = \sigma_2^2 = 1$ ve farklı ortalamalar (μ_1, μ_2) için iki değişkenli normal dağılım grafikleri (McLachlan ve Peel, 2000).

3.2.6.3. Çok değişkenli normal karışımlı dağılım

Çok değişkenli dağılım gösteren $x_1,...x_n$ gözlem değerlerinin K alt gruplu ortak dağılım fonksiyona sahip olsun. Çok değişkenli normal dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibi verilebilir.

$$\int\left(x;\mu,\sum\ \right)=\frac{1}{(2\pi)^{1/2}\sigma}\cdot e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$x_1,...x_n$ rasgele değişkenlerinin karışımlı çok değişkenli normal yoğunluk fonksiyonu $f(x|\pi,,\mu,\Sigma)$ şeklinde olup, aşağıdaki gibi yazılabilir.

$$f(x\left|\pi,\mu,\Sigma\right.)=\prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k f_k(x_i\left|\mu_k,\Sigma_k\right.)$$

$$=\prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k\left|\Sigma_k\right|^{-1/2}\exp\{-1/2\,(x_i-\mu_k)'\Sigma_k^{-1}(x_i-\mu_k)\}$$

p değişkeni normal dağılımın olasılık yoğunluk fonksiyonundaki $\mu(px1)$ boyutlu bir vektördür. p adet değişkene ait ortalamalardan oluşur. $\sum$ *ise* pxp boyutlu simetrik bir kare matristir.

$$\mu=\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_p\end{bmatrix}\qquad\qquad\sum=\begin{bmatrix}\sigma_{11} & \sigma_{12}\cdots\sigma_{1P}\\ \sigma_{21} & \sigma_{22}\ \ddots\ \vdots\\ \vdots & \\ \sigma_{P1}\ldots\ldots\ldots\ldots\sigma_{pp} & \end{bmatrix}$$

Bu matrisin köşegen elemanları p adet değişkene ait varyanslardan oluşur.

$i:1\cdots p$

$j:1\cdots p$ $\quad i \neq j \quad için \quad \sigma_{ij}$ ise i. ve j. değişkenler arasındaki kovaryanslardan oluşur (McLachlan ve Peel, 2000).

3.2.7. Normal karışımlı model

Çok değişkenli dağılım gösteren $x_1,...x_n$ gözlem değerlerinin K alt gruplu ortak dağılım fonksiyona sahip olsun. $x_1,...x_n$ rasgele değişkenlerinin karışımlı çok değişkenli normal yoğunluk fonksiyonu $f(x|\pi,,\mu,\Sigma)$ şeklinde olup, aşağıdaki gibi yazılabilir.

$$f(x|\pi,\mu,\Sigma) = \prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k f_k(x_i|\mu_k,\Sigma_k)$$

$$=\prod_{i=1}^{n}\sum_{k=1}^{K}\pi_k|\Sigma_k|^{-1/2}\exp\{-1/2(x_i-\mu_k)'\Sigma_k^{-1}(x_i-\mu_k)\} \qquad \textbf{(9)}$$

Eşitlik 9' da; π_k, x_i değişkenine ilişkin k'ıncı alt gruptaki karışma olasılığını göstermektedir. Burada k;

$(0<\pi_k<1,\sum_k \pi_k=1)$ ve $k=1,\ldots,K$

aralığında değişmektedir. μ ve $\sum$ parametreleri sırasıyla ortalama vektörü ile varyans kovaryans matrisini göstermektedir.

Eşitlik 9 kullanılarak en çok olabilirlik parametre tahminlerini doğrudan elde etmek oldukça zordur. Bunun için EM algoritması kullanılmaktadır. Her bir x değişkenini gösteren rasgele değişkeninin k'ıncı alt gruba dahil olup olmadığını tanımlamak için aşağıdaki eşitlik yazılabilir.

$$\zeta_{ik} = \begin{Bmatrix} 1, \; gözlem \quad k\text{'}ıncı \quad altgruba \quad ait \\ 0, \qquad diğer \quad durumda \end{Bmatrix}$$

$\hat{\mu}_k, \hat{\pi}_k$ ve $\hat{\sum}_k$ en çok olabilirlik parametreleri

$$\hat{\mu}_k = \frac{\sum_{i \in E_k} x_i}{n_k}, \qquad \hat{\pi}_k = \frac{n_k}{n}, \quad \hat{\Sigma}_k = \frac{\sum_{i \in E_k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)'}{n_k}$$

biçiminde yazılabilir. Burada $E_k = \{i : \gamma_i = k\}$ ve n_k, E_k'daki elementlerin sayısını göstermektedir. Başlangıçta, ζ_{ik} değerleri bilinmemekte birlikte EM algoritması tahmin sürecinde elde edilmektedir. Eşitlik 9 için Tam olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\ell(\vartheta; x_1, x_2, ..., x_n) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{K} \zeta_{ik} \left\{ \log \left| \sum{}_k \right| + (x_i - \mu_k)' \sum_k^{-1} (x_i - \mu_k) \right\}$$

$$+ \sum_{i=1}^{n} \sum_{k=1}^{K} \zeta_{ik} \log \pi_k - \frac{pn}{2} \log 2\pi \qquad (10)$$

Eşitlik 10' da p $x_1,...x_n$ rasgele değişkenleri için π_{ik} i' inci gözlemin k' ıncı alt grupta olma olasılığı olup, EM algoritmasının E aşamasında aşağıdaki gibi tahmin edilmektedir.

$$\pi_{ik} = E(\zeta_{ik} | x_i)$$

E aşamasında s' inci iterasyondaki π_{ik} değeri,

$$\pi_{ik}^{(s)} = \frac{\pi_k^{(s-1)} \phi(x_i; \mu_k^{(s-1)}, \Sigma_k^{(s-1)})}{\sum_{k'=1}^{K} \pi_{k'}^{(s-1)} \phi(x_i; \mu_{k'}^{(s-1)}, \Sigma_{k'}^{(s-1)})}$$

biçiminde tahmin edilmektedir. EM algoritmasının M aşamasında, s' inci iterasyonda tahmin edilen $\pi_k^{(s)}$, $\mu_k^{(s)}$ ve $\Sigma_k^{(s)}$ sırasıyla aşağıdaki gibi tahmin edilmektedir.

$$\pi_k^{(s)} = \frac{1}{n} \sum_{i=1}^{n} \pi_{ik}^{(s)}$$

$$\mu_k^{(s)} = \frac{\sum_{i=1}^{n} \pi_{ik}^{(s)} x_i}{\sum_{i=1}^{n} \pi_{ik}^{(s)}}$$

$$\Sigma_k^{(s)} = \frac{\sum_{i=1}^{n} \pi_{ik}^{(s)} (x_i - \mu_k^{(s)})(x_i - \mu_k^{(s)})'}{\sum_{i=1}^{n} \pi_{ik}^{(s)}}$$

Yakarıda verilen EM algoritmasının E ve M aşamalarında, yakınsama sağlanıncaya kadar yapılan iteratif işlemlerde, parametrelerin başlangıç değerlerinin seçimi önemlidir. Özellikle Σ_k' nin (varyans kovaryans) tahmin edilmesinde yakınsama konusunda sorun çıkabiliyor. Σ_k' nın tahmin edilmesinde;

1. Karışımlı modelde alt grup sayısı ve sınıflama yapıldığında $\sum_k$ elipsoidal bir şekil almalı ve μ_k etrafında yoğunlaşmalıdır.

2. K alt gruplarında homojen kovaryans matrisi olmalıdır. $\sum_k$ kovaryans matrisinin özelliği k' ıncı alt grup için geometrik özellikler sağladığından dolayı, karışımlı modelde her bir alt grup için $\sum_k$ tahmini aşağıdaki gibi yazılabilir.

$$\sum_k = \lambda_k D_k A_k D_k' \tag{11}$$

Eşitlik 11' de; D_k, $\sum_k$ 'nın öz vektörlerinin ortogonal matrisini, A_k bir diyagonal matris olup, her bir elementi $\sum_k$' nın öz vektörleri ile orantılıdır. λ_k ise skalar bir değerdir. D_k değeri $\sum_k$' nın temel bileşenlerinin yönünü, $\sum_k$ yoğunluk fonksiyonun şeklini ve λ_k ise elipsoidin büyüklüğünü belirlemektedir (Baudry ve Celeux, 2015).

3.2.8. Entropy sınıflandırma ölçütü

Karışımlı model kullanılarak elde edilen alt grupların ne derece doğru sınıflandırıldığını saptamak için yaygın olarak kullanılan ölçüt entropy istatistiğidir. Genel gösterimi çalışmalarda farklılık gösterebilir. Normal karışımlı modellemede kullanılan entropy ölçütü h_e ile gösterilmekte olup, entropy istatistiğinin hesaplama aşamaları aşağıda verilmiştir.

$$h_e = -\int_{-\infty}^{\infty} p(x)\ln(p(x))dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi\sigma}} \left[e^{-(x-\mu)^2/2\sigma^2} + e^{-(x+\mu)^2/2\sigma^2} \right] \ln\left(\frac{1}{\sqrt{2\pi\sigma}} e^{-(x^2+\mu^2)/2\sigma^2} \cosh(\mu x/\sigma^2) \right) dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi\sigma}} \left[e^{-(x-\mu)^2/2\sigma^2} + e^{-(x+\mu)^2/2\sigma^2} \right] \left[\ln(\frac{1}{\sqrt{2\pi\sigma}}) - \frac{(x^2+\mu^2)}{2\sigma^2} + \ln\left(\cosh(\mu x/\sigma^2)\right) \right] dx$$

$$= \ln(\sqrt{2\pi\sigma}) + \frac{\mu^2}{2\sigma^2} + \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} \frac{x^2}{2\sqrt{2\pi\sigma}} (e^{-(x-\mu)^2/2\sigma^2} + e^{-(x+\mu)^2/2\sigma^2}) dx$$

$$- \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{(x^2+\mu^2)/2\sigma^2} \cosh(\mu x/\sigma^2) \ln(\cosh(\mu x/\sigma^2)) dx$$

$$= \ln(\sqrt{2\pi\sigma}) + \frac{\mu^2}{2\sigma^2} + \frac{1}{2\sigma^2}(\sigma^2 + \mu^2)$$

$$- \frac{1}{\sqrt{2\pi\sigma}} e^{-\mu^2/2\sigma^2} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} \cosh(\mu x / \sigma^2) \ln(\cosh(\mu x / \sigma^2)) dx$$

(12)

Eşitlik 12' de, $y = \mu x / \sigma^2$ dönüşümü uygulandığında ve gerekli integral işlemleri aşağıdaki gibi olur.

$$h_e(X) = \ln(\sqrt{2\pi\sigma}) + \frac{\mu^2}{\sigma^2} + \frac{1}{2} - \frac{1}{\sqrt{2\pi\sigma}} e^{-\mu^2/2\sigma^2} (\sigma/\mu) \int_0^{\infty} e^{-\sigma^2 y^2/2\mu^2} \cosh(y) \ln(\cosh(y)) dy$$

$$h_e(X) = \frac{1}{2}\ln(2\pi e\sigma^2) + \frac{\mu^2}{\sigma^2} - \frac{2}{\sqrt{2\pi}} e^{-\mu^2/2\sigma^2} (\sigma/\mu) \int_0^{\infty} e^{-\sigma^2 y^2/2\mu^2} \cosh(y) \ln(\cosh(y)) dy$$

$\alpha = \mu/\sigma$ olsun. O halde $h_e(X)$ aşağıdaki gibi yazılabilir.

$$h_e(X) = \frac{1}{2}\ln(2\pi e\sigma^2) + \alpha^2 - \frac{2}{\sqrt{2\pi\sigma}} e^{-\alpha^2/2} (\sigma/\mu) \int_0^{\infty} e^{-y^2/2\alpha^2} \cosh(y) \ln(\cosh(y)) dy \qquad (13)$$

Eşitlik 13' de ilk terim Gaussian dağılımında entropy olarak tanımlanır. $\mu = 0$ (ve bu nedenle $\alpha = 0$) olduğunda dağılım Gaussian dağılımına dönüşür ve entropy eşitliğin ilk terimi olur. Eşitlik 13' ün çözümünde analitik yaklaşımlar yetersiz kalmaktadır. Bu nedenle aşağıda tanımlanan analitik sınırlar bu integralin çözümü için kullanılabilir.

$$y - \ln 2 \leq \ln(\cosh(y)) \leq y \forall y \geq 0$$

Böylece, integral terimi için üst sınırı aşağıdaki gibi yazılabilir.

$$I = \frac{2}{\sqrt{2\pi\sigma}} e^{-\alpha^2/2} \int_0^\infty e^{-y^2/2\alpha^2} \cosh(y) \ln(\cosh(y)) dy$$

$$\leq \frac{2}{\sqrt{2\pi\sigma}} e^{-\alpha^2/2} \int_0^\infty e^{-y^2/2\alpha^2} \cosh(y) \ln(\cosh(y) dy$$

$$= \frac{2}{\sqrt{2\pi\sigma}} e^{-\alpha^2/2} \left[\frac{\alpha^2}{2} \sqrt{2\alpha^2\pi} e^{\alpha^2/2} erf(\alpha/\sqrt{2}) + \alpha^2 \right]$$

$$= \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2}$$

Benzer şekilde alt sınır için,

$$I \geq \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2} - \frac{2}{\sqrt{2\pi}\alpha} e^{-\alpha^2/2} \ln 2 \int_0^\infty e^{-y^2/2\alpha^2} \cosh(y) dy$$

$$= \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2} - \frac{2}{\sqrt{2\pi}\alpha} e^{-\alpha^2/2} \ln 2 \left[\frac{1}{2} \sqrt{2\alpha^2\pi} e^{\alpha^2/2} \right]$$

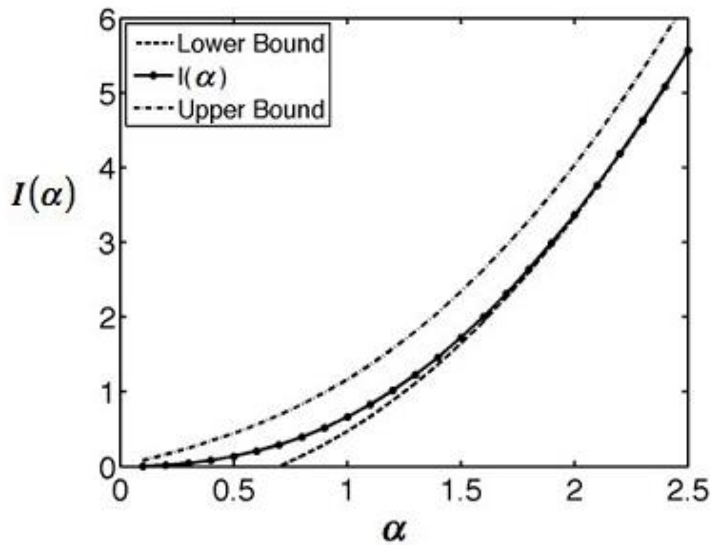
$$= \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2} - \ln 2$$

I' daki integral değeri her zaman sıfırdan büyük eşit olduğu için $I \geq 0$ olduğu bilinmektedir. Sonuçta entropy aşağıdaki gibi yazılabilir.

$$h_e(X) = \frac{1}{2}\ln(2\pi e\sigma^2) + \alpha^2 - I$$

$$\max(0; \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2} - \ln 2) \le I \le \alpha^2 erf(\alpha/\sqrt{2}) + \sqrt{2/\pi}\alpha e^{-\alpha^2/2}$$

Hepsi için $\alpha = \mu/\sigma \ge 0$ olur. *I'* nın değişimi Şekil 3.6' da verilmiştir.



Şekil 3.6 $I(\alpha)$ ve α için alt ve üst sınırlar.

Şekil 3.6' da *I* fonksiyonun analitik alt ve üst sınırları verilmiştir. α büyüdüğünde *I* hızla alt sınıra yaklaşır (Michalowicz ve ark., 2008).

$$h_e(X) = \frac{1}{2}\ln(2\pi e\sigma^2) + (\alpha^2 - I)$$

3.2.9. Genelleştirilmiş doğrusal modellerde karışımlı yaklaşım

Genelleştirilmiş doğrusal modeller bir bağlantı fonksiyonu (link function) kullanarak ($h(.)$) şansa bağlı değişken (Y bağımlı değişkeni) ve sistematik bileşen ($X^T\beta$ doğrusal tahminleyicisi) ilişkisini sağlamaktadır (McCullagh and Nelder, 1989). Özellikle bu ilişki,

$$E(Y) = \mu = h(X^T\beta).$$

biçiminde yazılabilir. Burada Y rasgele değişkeni üstel dağılım ailesine ait olduğu varsayılır. Bu üstel dağılım ailesi formu aşağıdaki gibi yazılabilir.

$$q(y;\theta,\phi) = \exp\{y(\theta) - b(\theta))/\alpha(\phi) + c(y,\phi)\},$$

Eşitlikte $a(.)$, $b(.)$ ve $c(.)$ ilgili (belirtilen) fonksiyonlardır. Bilinen doğrusal regresyon modeli (bağımlı değişken Y normal dağılım gösterdiğinde) en temel GLM olup, $h(.)$ bağlantı fonksiyonu birim (1) değerine eşit olur (identity link).

k alt gruplu Genelleştirilmiş doğrusal modellerde karışımlı model aşağıdaki gibi yazılabilir.

$$f(y_{i;} x_i, \psi) = \sum_{j=1}^{k} \lambda_j \exp\{g(y_i, x_i; \beta_{j,} \kappa_J)\},$$

Eşitlikte, $g(y_i, x_i; \beta_{j,} \kappa_J)$ fonksiyonu aşağıdaki gibi yazılabilir.

$$g(y_i, x_i; \beta_{j,} \kappa_J) = \kappa_j^{-1}(y_i h(x_i^T \beta_J) - b[h(x_i^T \beta_J)]) + c(y_i; \kappa_J).$$

Burada, κ_J yayılım parametresini (dispersion parameter) göstermektedir. $b(.)$, $c(.)$ and $h(.)$ fonksiyonlarının tamamının bilindiği varsıyılır.

3.2.10. EM algoritması

$\psi = (\beta_1^T,..., \beta_k^T, \lambda_1,..., \lambda_{k-1})^T$ ile gösterilsin. Bu durumda gözlenmiş veriler için log-olabilirlik ve tam log olabilirlik fonksiyonları sırasıyla aşağıdaki gibi yazılabilir.

$$\ell_o(\psi) = \sum_{i=1}^{n} \log \sum_{j=1}^{k} \lambda_j \exp\left\{g(y_i, x_i; \beta_{j,} \kappa_J)\right\}$$

ve

$$\ell c(\psi) = \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{i,j} \log\left[\lambda_j \exp\left\{g(y_i, x_i; \beta_{j,} \kappa_J)\right\}\right]$$

$t = 0,1,...,$ iterasyonun E aşaması boyunca, beklenen tam log-olabilirlik aşağıdaki gibi hesaplanır.

$$\wp(\psi;\psi^{(t)}) = \sum_{j=1}^{k} \wp(\psi;\psi^{(t)})$$

$$\wp_j(\psi) = \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{i,j}^{(t)} \log\left[\lambda_j \exp\left\{g(y_i, x_i; \beta_{j,} \kappa_J)\right\}\right]$$

ve

$$Z_{i,j}^{(t)} = \frac{\lambda_j^{(t)} \exp\{g(y_i, x_i; \beta_{j,}\kappa_J)\}}{\sum_{l=1}^{k} \lambda_l^{(t)} \exp\{g(y_i, x_i; \beta_{l,}\kappa_l)\}} = \left[1 + \sum_{l \neq j}^{k} \frac{\lambda_l^{(t)}}{\lambda_j^{(t)}} \exp\{g(y_i, x_i; \beta_{j,}\kappa_J) - g(y_i, x_i; \beta_{j,}\kappa_J)\}\right]^{-1}$$

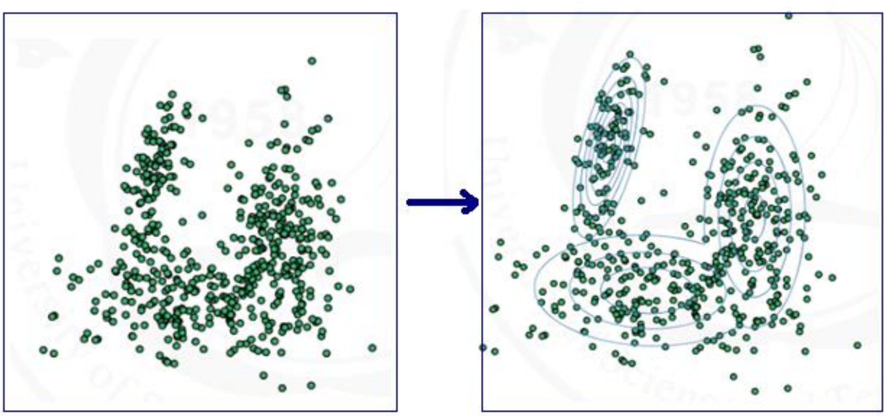
M aşamasında, ψ 'ye göre $\wp(\psi;\psi^{(t)}$ fonksiyonu maksimize edilir ($\psi^{(t+1)}$ güncellemesi yapılarak). Bu maksimize işlemi $\ell_0(\psi^{(t+1)}) - \ell_0(\psi^{(t)}) < \varepsilon = 10^{-8}$ koşulu sağlanıncaya kadar devam eder.

$$\lambda_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} z_{i,j}^{(t)}$$

ve

$$\beta_j^{(t+1)} = \arg\max \wp_j(\psi;\psi^{(t+1)})$$

Bir karmaşık veri setinin normal karışımlı model kullanılarak sınıflandırılması aşağıdaki şekilde verilmiştir.



Şekil 3.7. Normal Karışımlı model kullanılarak sınıflandırma.

4. BULGULAR

Bu çalışmada, öncelikle illere göre çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi mutluluk düzeyi ölçütleri kullanılarak, illerin homojen kaç alt grupta toplandığının normal karışımlı model ile ortaya konması amaçlanmıştır.Bunu belirlemek için AIC, BIC ve entropy gibi uyum ölçütleri kullanılmıştır. Bu ölçütlerden AIC ve BIC en iyi modelin elde edilmesini sağlarken, entropy ölçütü ise olası her alt gruptaki bireylerin doğru sınıflandırma olasılıklarını vermektedir. Bu uyum ölçütlerine ilişkin sonuçlar Çizelge 4.1' de verilmiştir.

Çizelge 4.1' de verilen sekiz farklı alt gruplu normal karışımlı modellere ilişkin AIC ve BIC istatistik değerleri; en yüksek 1 alt gruplu modelde sırasıyla 3432.330 ve 3480.219; en küçük ise 6 alt gruplu modelde sırasıyla 3041.249 ve 3221.833 olarak elde edilmişlerdir.

Böylece AIC ve BIC istatistiklerine göre, veri seti söz konusu özellikler bakımından homojen 6 alt gruptan oluşmaktadır. AIC ve BIC uyum istatistikleri 6 alt gruplu modele kadar azaldığı ve daha sonrasında büyüdüğü saptanmıştır. 6 alt gruplu modelden sonraki uyum ölçütleri büyüdüğünden dolayı 8 alt gruplu modelden sonraki uyum ölçütlerine yer verilmemiştir. Bu sonuçlara göre illere Göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kullanılarak Türkiye'deki illerin kendi içlerinde homojen ve kendi aralarında ise heterojen olmak üzere 6 alt gruba ayrılmıştır. Entropy doğru sınıflandırma ölçütü bakımında en yüksek doğru sınıflandırma oranı 6 alt gruplu modelde % 96.8 olarak elde edilmiştir. Bununla birlikte en düşük entropy doğru sınıflandırma oranı 3 alt gruplu modelde % 89.3 olarak elde edilmiştir.

Çizelge 4.1. İllere göre mutluluk düzeyi için homojen alt grupların belirlenmesinde kullanılan uyum ölçütleri

Alt gruplar (modeller)	Uygun model seçme ölçütleri		
	AIC	BIC	Entropy(%)
Bir alt gruplu model	3432.330	3480.219	-
İki alt gruplu model	3249.219	3323.466	0.939
Üç alt gruplu model	3169.397	3269.964	0.893
Dört alt gruplu model	3100.808	3227.714	0.923
Beş alt gruplu model	3074.378	3227.623	0.927
Altı alt gruplu model	3041.249	3221.833	0.968
Yedi alt gruplu model	3046.960	3247.883	0.946
Sekiz alt gruplu model	3049.233	3281.495	0.956

Bu çalışmada kullanılan değişkenlere ilişkin kovaryans ve korelasyon matrisler ekler dizininde Çizelge 6.7 ve Çizelge 6.8’ de verilmiştir

Normal karışımlı modelleme kullanılarak elde edilen ve en iyi model olan 6 alt gruplu modeldeki her bir alt gruba giren bireylerin sayısı ve karışma olasılıkları çizelge 4.2’ de verilmiştir. Bu alt gruplara giren birey sayıları ve oranları sırasıyla 12 (% 14.815), 28 (34.568), 29 (35.802), 3 (% 3.704), 8 (% 9.877) ve 1 (% 1.235) olarak elde edilmiştir. En çok birey alt grup 3’ te ve en az birey ise alt grup 6’ da olduğu saptanmıştır.

Çizelge 4.2. İllere göre mutluluk düzeyi bakımından altı alt gruplu modelde bireylerin alt gruplara göre dağılımı

Alt gruplar	Alt gruplara giren birey sayıları ve oranları	
	N	Oran (%)
Alt grup 1	12	0.14815
Alt grup 2	28	0.34568
Alt grup 3	29	0.35802
Alt grup 4	3	0.03704
Alt grup 5	8	0.09877
Alt grup 6	1	0.01235

Normal karışımlı modelleme kullanılarak elde edilen 6 alt gruplu modelde her bir alt grup için entropy doğru sınıflandırma oranları Çizelge 4.3' de verilmiştir. Elde edilen alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için entropy doğru sınıflandırma oranları sırasıyla % 97.7, % 91.6, % 98.8, % 100, % 94.6 ve % 100 olarak elde edilmiştir. Her bir alt grup için elde edilen entropy doğru sınıflandırma oranları oldukça yüksek elde edilmiştir.

Çizelge 4.3. Altı alt gruplu modelde her bir alt grup için tahmin edilen entropy doğru sınıflandırma oranları

	Alt gruplar					
	Alt grup 1	Alt grup 2	Alt grup 3	Alt grup 4	Alt grup5	Alt grup6
Alt grup 1	0.977	0.000	0.023	0.000	0.000	0.000
Alt grup 2	0.000	0.916	0.084	0.000	0.000	0.000
Alt grup 3	0.000	0.016	0.980	0.000	0.004	0.000
Alt grup 4	0.000	0.000	0.000	1.000	0.000	0.000
Alt grup 5	0.000	0.000	0.054	0.000	0.946	0.000
Alt grup 6	0.000	0.000	0.000	0.000	0.000	1.000

İllere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kullanılarak elde edilen 6 alt gruplu modelde, normal karışımlı modelleme kullanılarak tahmin edilen ortalama değerler Çizelge 4.4’ de verilmiştir. İllere göre mutluluk kaynağı olarak çocuk diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 10.634, 11.672, 11.880, 10.533, 14.210 ve 12.300 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak eş diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 8.595, 4.821, 4.904, 16.796, 6.526 ve 7.400 olarak elde edilmiştir. İllere Göre mutluluk kaynağı olarak anne/baba diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 4.763, 2.206, 2.538, 5.233, 2.597 ve 8.900 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak kendisi diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 2.160, 1.734, 2.631, 2.000, 2.560 ve 3.800 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak torun diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 0.802, 1.905, 1.637, 0.633, 2.438 ve 0.400 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak sağlık diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 59.930, 75.775, 68.672, 56.002, 61.572 ve 52.700 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak sevgi diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 21.737, 11.870, 14.800, 24.732, 17.859 ve 13.900 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak başarı diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 7.760, 5.832, 8.471, 7.934, 10.669 ve 10.700 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak para diyenlerin; alt grup alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 3.909, 3.369, 4.252, 5.299, 5.695 ve 6.800 olarak elde edilmiştir. İllere göre mutluluk kaynağı olarak iş diyenlerin; alt grup 1, alt grup 2, alt grup 3, alt grup 4, alt grup 5 ve alt grup 6 için tahmin edilen ortalama değerleri sırasıyla 3.788, 1.982, 2.426, 4.267, 2.571 ve 12.000 olarak elde edilmiştir.

Çizelge 4.4. İllere göre mutluluk düzeyi bakımından alt gruplara ilişkin normal karışımlı modelleme kullanılarak elde edilen ortalama değerler

Alt gruplar	Normal karışımlı modelleme kullanılarak elde edilen ortalama değerler									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Alt grup 1	10.634	8.595	4.763	2.160	0.802	59.930	21.737	7.760	3.909	3.788
Alt grup 2	11.672	4.821	2.206	1.734	1.905	75.775	11.870	5.832	3.369	1.982
Alt grup 3	11.880	4.904	2.538	2.631	1.637	68.672	14.800	8.471	4.252	2.426
Alt grup 4	10.533	16.796	5.233	2.000	0.633	56.002	24.732	7.934	5.299	4.267
Alt grup 5	14.210	6.526	2.597	2.560	2.438	61.572	17.859	10.669	5.695	2.571
Alt grup 6	12.300	7.400	8.900	3.800	0.400	52.700	13.900	10.700	6.800	12.000

İllere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kullanılarak elde edilen 6 alt gruplu modelde, Her bir alt gruba dâhil olan illerin dağılımı Çizelge 4.5' de verilmiştir. En fazla il alt grup 2 ve 3' e sırasıyla 28 ve 29 olarak dağılmışken, en az dağılım ise alt grup 6' da sadece Mardin ili olmuştur.

Çizelge 4.5. İllerin alt gruplara göre dağılımı

Alt gruplar	İller
Alt grup 1	Ardahan, Batman, Diyarbakır, Erzurum, Erzincan, Hakkâri, Iğdır, Kars, Siirt, Van, Şanlıurfa ve Şırnak
Alt grup 2	Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir
Alt grup 3	Adana, Aksaray, Ankara, Artvin, Aydın, Bartın, Bingöl, Bursa, Denizli, Elazığ, Eskişehir, Gaziantep, Gümüşhane, Hatay, Isparta, Karabük, Kilis, Konya, Kırklareli, Malatya, Mersin, Muğla, Osmaniye, Samsun, Tekirdağ, Trabzon, Tunceli, Zonguldak ve İstanbul
Alt grup 4	Ağrı, Bitlis ve Muş
Alt grup 5	Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat
Alt grup 6	Mardin

Normal karışımlı modelleme kullanılarak elde edilen 6 alt gruplu modelde alt grup 1 için illere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.6' da ve güven aralıkları Çizelge 4.7' de verilmiştir. Alt grup 1' de çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 10.634, 8.595, 4.763, 2.160, 0.802, 59.930, 21.737, 7.760, 3.909 ve 3.788 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1' deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Bununla birlikte, alt grup 1 için elde edilen varyans istatistik değerleri ekler dizininde Çizelge 6.1' de, kovaryans değerleri ise Çizelge 6.9' da verilmiştir.

Çizelge 4.6. Alt grup 1 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	10.634	1.204	8.835	0.000
Eş	8.595	0.690	12.453	0.000
Anne/Baba	4.763	0.475	10.026	0.000
Kendisi	2.160	0.236	9.164	0.000
Torunlar	0.802	0.108	7.456	0.000
Sağlık	59.930	1.056	56.753	0.000
Sevgi	21.737	1.612	13.482	0.000
Başarı	7.760	0.522	14.876	0.000
Para	3.909	0.260	15.031	0.000
İş	3.788	0.282	13.434	0.000

Çizelge 4.7. Alt grup 1 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	7.534	8.275	8.654	10.634	12.614	12.993	13.734
Eş	6.817	7.242	7.460	8.595	9.731	9.948	10.373
Anne/Baba	3.539	3.832	3.981	4.763	5.544	5.694	5.986
Kendisi	1.553	1.698	1.773	2.160	2.548	2.623	2.768
Torunlar	0.525	0.591	0.625	0.802	0.979	1.013	1.080
Sağlık	57.210	57.860	58.193	59.930	61.667	62.000	62.650
Sevgi	17.584	18.577	19.085	21.737	24.389	24.897	25.890
Başarı	6.416	6.737	6.902	7.760	8.618	8.782	9.103
Para	3.239	3.399	3.481	3.909	4.336	4.418	4.578
İş	3.062	3.235	3.324	3.788	4.252	4.340	4.514

Normal karışımlı modelleme kullanılarak elde edilen alt grup 2 için çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenlerine göre elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.8' de ve güven aralıkları Çizelge 4.9' de verilmiştir. Alt grup 2' de çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 11.672, 4.821, 2.206, 1.734, 1.905, 75.775, 11.870, 5.832, 3.369 ve 1.982 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1'deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Bunun yanı sıra, alt grup 2 için elde edilen varyans değerleri ekler dizininde Çizelge 6.2' de, kovaryans değerleri ise Çizelge 6.10' da verilmiştir.

Çizelge 4.8. Alt grup 2 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	11.672	0.865	13.499	0.000
Eş	4.821	0.392	12.308	0.000
Anne/Baba	2.206	0.392	5.624	0.000
Kendisi	1.734	0.178	9.758	0.000
Torunlar	1.905	0.207	9.210	0.000
Sağlık	75.775	1.338	56.638	0.000
Sevgi	11.870	0.605	19.632	0.000
Başarı	5.832	0.672	8.673	0.000
Para	3.369	0.221	15.266	0.000
İş	1.982	0.295	6.721	0.000

Çizelge 4.9. Alt grup 2 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	9.444	9.977	10.249	11.672	13.094	13.366	13.899
Eş	3.812	4.053	4.177	4.821	5.465	5.589	5.830
Anne/Baba	1.196	1.437	1.561	2.206	2.851	2.975	3.216
Kendisi	1.277	1.386	1.442	1.734	2.027	2.083	2.192
Torunlar	1.372	1.499	1.564	1.905	2.245	2.310	2.437
Sağlık	72.329	73.153	73.575	75.775	77.976	78.398	79.222
Sevgi	10.313	10.685	10.876	11.870	12.865	13.055	13.428
Başarı	4.100	4.514	4.726	5.832	6.938	7.150	7.564
Para	2.800	2.936	3.006	3.369	3.732	3.801	3.937
İş	1.223	1.404	1.497	1.982	2.468	2.560	2.742

Normal karışımlı modelleme kullanılarak elde edilen alt grup 3 için çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenlerine göre elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.10’ da ve güven aralıkları Çizelge 4.11’ de verilmiştir. Alt grup 3’ de çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 11.880, 4.904, 2.538, 2.631, 1.637, 68.672, 14.800, 8.471, 4.252 ve 2.426 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1’ deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Bununla birlikte, alt grup 3 için elde edilen varyans istatistik değerleri ekler dizininde Çizelge 6.3’ de, kovaryans değerleri ise Çizelge 6.10’ da verilmiştir.

Çizelge 4.10. Alt grup 3 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	11.880	0.499	23.795	0.000
Eş	4.904	0.259	18.939	0.000
Anne/Baba	2.538	0.292	8.689	0.000
Kendisi	2.631	0.237	11.106	0.000
Torunlar	1.637	0.175	9.366	0.000
Sağlık	68.672	1.300	52.828	0.000
Sevgi	14.800	0.510	29.034	0.000
Başarı	8.471	0.404	20.990	0.000
Para	4.252	0.466	9.118	0.000
İş	2.426	0.145	16.707	0.000

Çizelge 4.11. Alt grup 3 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	10.594	10.902	11.059	11.880	12.701	12.859	13.166
Eş	4.237	4.397	4.478	4.904	5.330	5.412	5.571
Anne/Baba	1.786	1.965	2.057	2.538	3.018	3.110	3.290
Kendisi	2.020	2.166	2.241	2.631	3.020	3.095	3.241
Torunlar	1.187	1.295	1.350	1.637	1.925	1.980	2.088
Sağlık	65.324	66.124	66.534	68.672	70.811	71.220	72.021
Sevgi	13.487	13.801	13.962	14.800	15.639	15.800	16.113
Başarı	7.431	7.680	7.807	8.471	9.135	9.262	9.510
Para	3.051	3.338	3.485	4.252	5.019	5.166	5.453
İş	2.052	2.141	2.187	2.426	2.664	2.710	2.800

Normal karışımlı modelleme kullanılarak elde edilen alt grup 4 için çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenlerine göre elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.12' de ve güven aralıkları Çizelge 4.13'de verilmiştir. Alt grup 4'de çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 10.533, 16.796, 5.233, 2.000, 0.633, 56.002, 24.732, 7.934, 5.299 ve 4.267 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1' deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Bununla birlikte, alt grup 4 için elde edilen varyans istatistik değerleri ekler dizininde Çizelge 6.4' de, kovaryans değerleri ise Çizelge 6.11' de verilmiştir.

Çizelge 4.12. Alt grup 4 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	10.533	1.007	10.455	0.000
Eş	16.796	0.577	29.103	0.000
Anne/Baba	5.233	0.687	7.618	0.000
Kendisi	2.000	0.170	11.762	0.000
Torunlar	0.633	0.072	8.786	0.000
Sağlık	56.002	1.389	40.308	0.000
Sevgi	24.732	1.212	20.401	0.000
Başarı	7.934	0.694	11.427	0.000
Para	5.299	0.295	17.980	0.000
İş	4.267	0.699	6.100	0.000

Çizelge 4.13. Alt grup 4 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	7.938	8.559	8.876	10.533	12.190	12.508	13.128
Eş	15.310	15.665	15.847	16.796	17.746	17.928	18.283
Anne/Baba	3.463	3.886	4.103	5.233	6.363	6.579	7.002
Kendisi	1.562	1.666	1.720	2.000	2.279	2.333	2.438
Torunlar	0.447	0.492	0.515	0.633	0.752	0.774	0.819
Sağlık	52.424	53.279	53.717	56.002	58.288	58.725	59.581
Sevgi	21.609	22.356	22.737	24.732	26.726	27.108	27.854
Başarı	6.145	6.573	6.791	7.934	9.076	9.294	9.722
Para	4.540	4.721	4.814	5.299	5.784	5.877	6.058
İş	2.465	2.896	3.116	4.267	5.417	5.637	6.068

Normal karışımlı modelleme kullanılarak elde edilen alt grup 5 için çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenlerine göre elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.14’ de ve güven aralıkları Çizelge 4.15’ de verilmiştir. Alt grup 5’de çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 14.210, 6.526, 2.597, 2.560, 2.438, 61.572, 17.859, 10.669, 5.695 ve 2.571 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1’ deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Alt grup 4 için elde edilen varyans istatistik değerleri ekler dizininde Çizelge 6.5’de, kovaryans değerleri ise Çizelge 6.12’ de verilmiştir.

Çizelge 4.14. Alt grup 5 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	14.210	0.960	14.805	0.000
Eş	6.526	0.678	9.626	0.000
Anne/Baba	2.597	0.276	9.417	0.000
Kendisi	2.560	0.292	8.772	0.000
Torunlar	2.438	0.174	13.978	0.000
Sağlık	61.572	1.989	30.948	0.000
Sevgi	17.859	1.613	11.074	0.000
Başarı	10.669	0.600	17.780	0.000
Para	5.695	0.376	15.165	0.000
İş	2.571	0.385	6.687	0.000

Çizelge 4.15. Alt grup 5 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	11.738	12.329	12.631	14.210	15.789	16.091	16.682
Eş	4.779	5.197	5.410	6.526	7.641	7.854	8.272
Anne/Baba	1.887	2.056	2.143	2.597	3.051	3.138	3.307
Kendisi	1.808	1.988	2.080	2.560	3.040	3.132	3.312
Torunlar	1.989	2.096	2.151	2.438	2.725	2.780	2.887
Sağlık	56.447	57.672	58.299	61.572	64.844	65.471	66.696
Sevgi	13.705	14.698	15.206	17.859	20.511	21.019	22.012
Başarı	9.123	9.493	9.682	10.669	11.656	11.845	12.214
Para	4.727	4.959	5.077	5.695	6.312	6.431	6.662
İş	1.581	1.817	1.939	2.571	3.204	3.325	3.562

Normal karışımlı modelleme kullanılarak elde edilen alt grup 6 için çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenlerine göre elde edilen parametre tahmin değerleri ve standart hataları Çizelge 4.16' da ve güven aralıkları Çizelge 4.17' de verilmiştir. Alt grup 6' da çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için elde edilen tahmin değerleri sırasıyla 12.300, 7.400, 8.900, 3.800, 0.400, 52.700, 13.900, 10.700, 6.800 ve 12.000 olarak elde edilmiştir. Söz konusu tüm değişkenlerin alt grup 1' deki etkileri istatistiksel olarak önemli bulunmuştur ($p<0.01$). Bununla birlikte, alt grup 6 için elde edilen varyans istatistik değerleri ekler dizininde Çizelge 6.6' da, kovaryans değerleri ise Çizelge 6.14' de verilmiştir.

Çizelge 4.16. Alt grup 6 için elde edilen parametre tahmin değerleri standart hatalar

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	12.300	1.678	7.330	0.000
Eş	7.400	0.976	7.582	0.000
Anne/Baba	8.900	0.829	10.736	0.000
Kendisi	3.800	0.932	4.077	0.000
Torunlar	0.400	0.078	5.128	0.000
Sağlık	52.700	1.835	28.719	0.000
Sevgi	13.900	1.051	13.224	0.000
Başarı	10.700	1.178	9.083	0.000
Para	6.800	0.981	6.932	0.000
İş	12.000	1.421	8.445	0.000

Çizelge 4.17. Alt grup 6 için elde edilen parametre tahmin değerlerine ilişkin güven aralıkları

Değişkenler	Alt sınır(%.5)	Alt sınır(%2.5)	Alt sınır (%5)	Tahmin	Üst sınır (%5)	Alt sınır (%2.5)	Alt sınır (%.5)
Çocuklar	12.300	12.300	12.300	12.300	12.300	12.300	12.300
Eş	7.400	7.400	7.400	7.400	7.400	7.400	7.400
Anne/Baba	8.900	8.900	8.900	8.900	8.900	8.900	8.900
Kendisi	3.800	3.800	3.800	3.800	3.800	3.800	3.800
Torunlar	0.400	0.400	0.400	0.400	0.400	0.400	0.400
Sağlık	52.700	52.700	52.700	52.700	52.700	52.700	52.700
Sevgi	13.900	13.900	13.900	13.900	13.900	13.900	13.900
Başarı	10.700	10.700	10.700	10.700	10.700	10.700	10.700
Para	6.800	6.800	6.800	6.800	6.800	6.800	6.800
İş	12.000	12.000	12.000	12.000	12.000	12.000	12.000

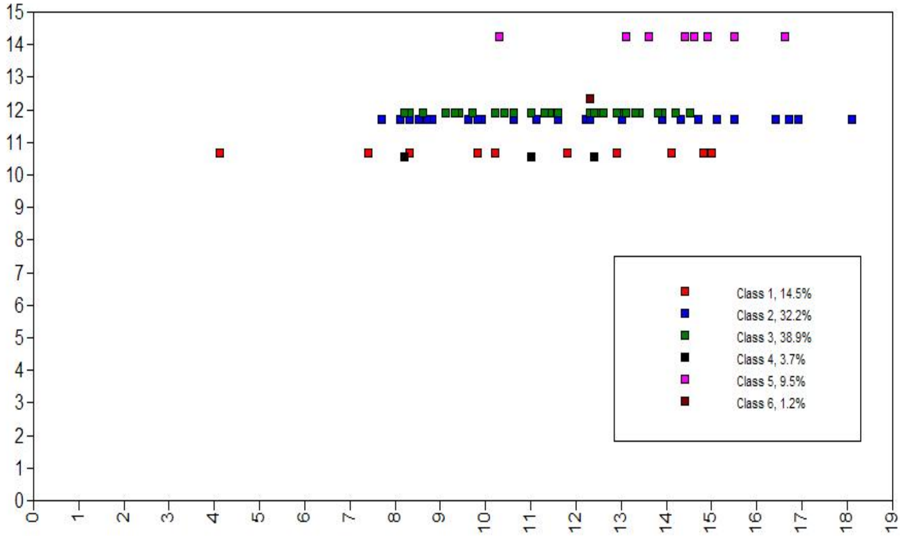
En iyi model olarak seçilen 6 alt gruplu modelde her bir alt grupta çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş değişkenleri için kullanılan başlangıç değerleri Çizelge 4.17' de verilmiştir.

Çizelge 4.18. Alt gruplar için kullanılan başlangıç değerleri

Alt gruplar	Değişkenler									
	Çocuk	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Alt grup 1	6.313	0.155	-0.110	0.583	0.159	54.271	6.846	3.512	1.709	-0.413
Alt grup 2	9.061	3.096	1.414	1.420	0.892	61.313	11.177	5.623	2.917	1.134
Alt grup 3	11.809	6.037	2.937	2.258	1.626	68.356	15.507	7.733	4.125	2.680
Alt grup 4	11.809	6.037	2.937	2.258	1.626	68.356	15.507	7.733	4.125	2.680
Alt grup 5	14.556	8.978	4.460	3.096	2.360	75.398	19.838	9.844	5.332	4.227
Alt grup 6	17.304	11.920	5.984	3.933	3.093	82.440	24.168	11.955	6.540	5.774

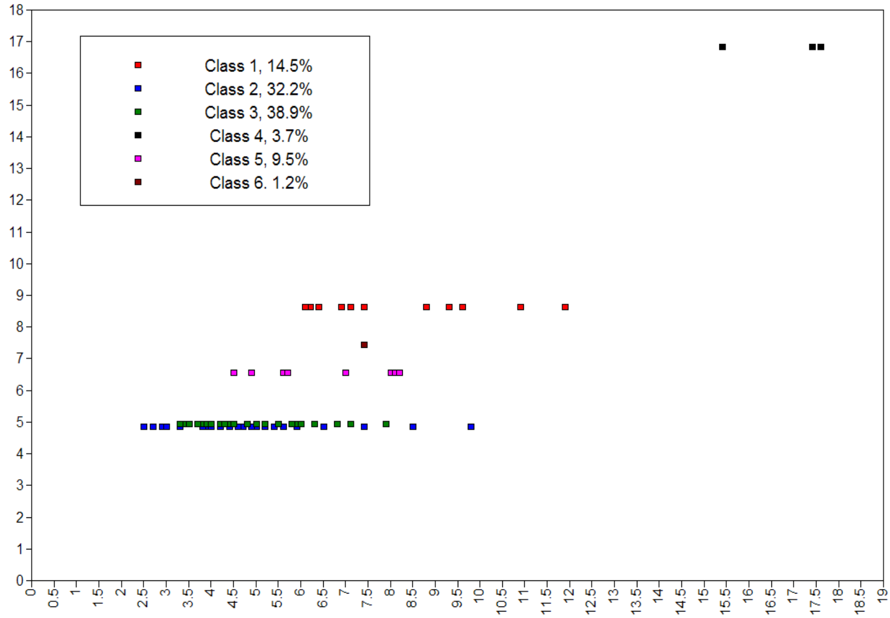
İllere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunlar, sağlık, sevgi, başarı, para ve iş gibi değişkenlerin her bir alt gruptaki tahmin edilen ortalama değerleri Şekil 1-Şekil 10 arası verilmiştir.

Şekil 1' de mutluluk kaynağı olarak çocuklar; en yüksek ortalama ile alt grup 5' de (class 5), en düşük ortalama ile alt grup 4 (class 4) ve alt grup 1'de (class 1) elde edilmiştir. Bu bağlamda, mutluluk kaynağı olarak çocukları gören iller alt grup 5'de yer almış olup ve sırasıyla; Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat olarak yazılabilir. Buna karşın mutluluk kaynağı olarak en az çocukları en gören iller alt grup 1 ve alt grup 4' de yer almış olup ve sırasıyla; Ardahan, Batman, Diyarbakır, Erzurum, Erzincan, Hakkâri, Iğdır, Kars, Siirt, Van, Şanlıurfa ve Şırnak, Ağrı, Bitlis ve Muş olarak söylenebilir.



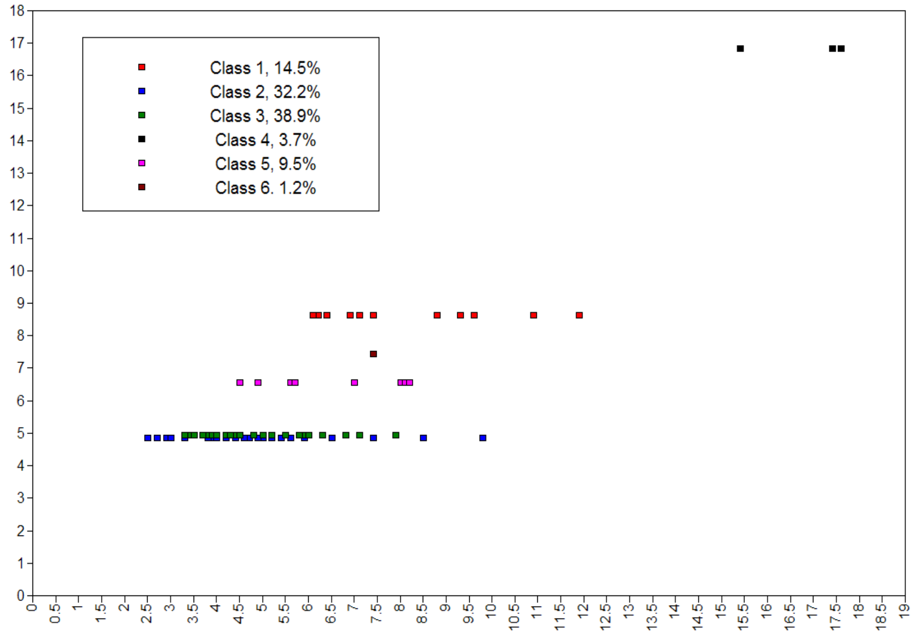
Şekil 4.1. Mutluluk kaynağı olarak çocuk değişkenin alt gruplara dağılımı.

Şekil 4.2' de mutluluk kaynağı olarak eş; en yüksek ortalama ile alt grup 4' de (class 4), en düşük ortalama ile alt grup 2 (class 2) ve alt grup 3' de (class 3) elde edilmiştir. Bu bağlamda, eşi mutluluk kaynağı olarak gören iller alt grup 4' de yer almış olup ve sırasıyla; Ağrı, Bitlis ve Muş olarak saptanmıştır. Buna karşın mutluluk kaynağı olarak en az eşleri gören iller alt grup 2 ve alt grup 3' de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir, Adana, Aksaray, Ankara, Artvin, Aydın, Bartın, Bingöl, Bursa, Denizli, Elazığ, Eskişehir, Gaziantep, Gümüşhane, Hatay, Isparta, Karabük, Kilis, Konya, Kırklareli, Malatya, Mersin, Muğla, Osmaniye, Samsun, Tekirdağ, Trabzon, Tunceli, Zonguldak ve İstanbul olmuştur.



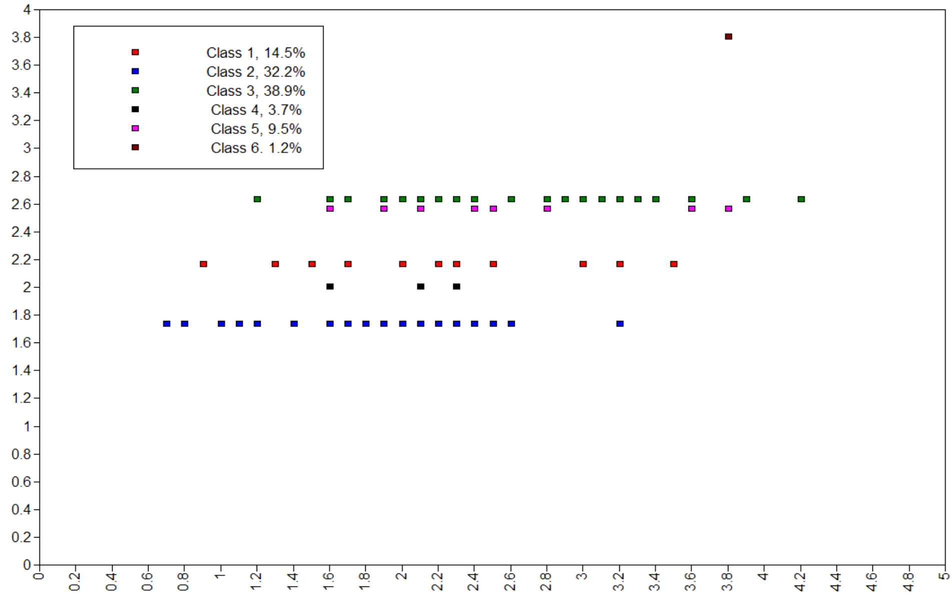
Şekil 4.2. Mutluluk kaynağı olarak eş değişkenin alt gruplara dağılımı.

Şekil 4.2' de mutluluk kaynağı olarak anne/baba; en yüksek ortalama ile alt grup 4' de (class 4), en düşük ortalama ile alt grup 3 (class 3) ve alt grup 2' de (class 2) elde edilmiştir. Anne ve babayı mutluluk kaynağı olarak gören iller alt grup 4' de yer almış olup ve sırasıyla; Ağrı, Bitlis ve Muş'tan oluşmaktadır. Buna karşın mutluluk kaynağı olarak en az eşleri gören iller alt grup 3 ve alt grup 2' de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum, İzmir, Adana, Aksaray, Ankara, Artvin, Aydın, Bartın, Bingöl, Bursa, Denizli, Elazığ, Eskişehir, Gaziantep, Gümüşhane, Hatay, Isparta, Karabük, Kilis, Konya, Kırklareli, Malatya, Mersin, Muğla, Osmaniye, Samsun, Tekirdağ, Trabzon, Tunceli, Zonguldak ve İstanbul olarak saptanmıştır.



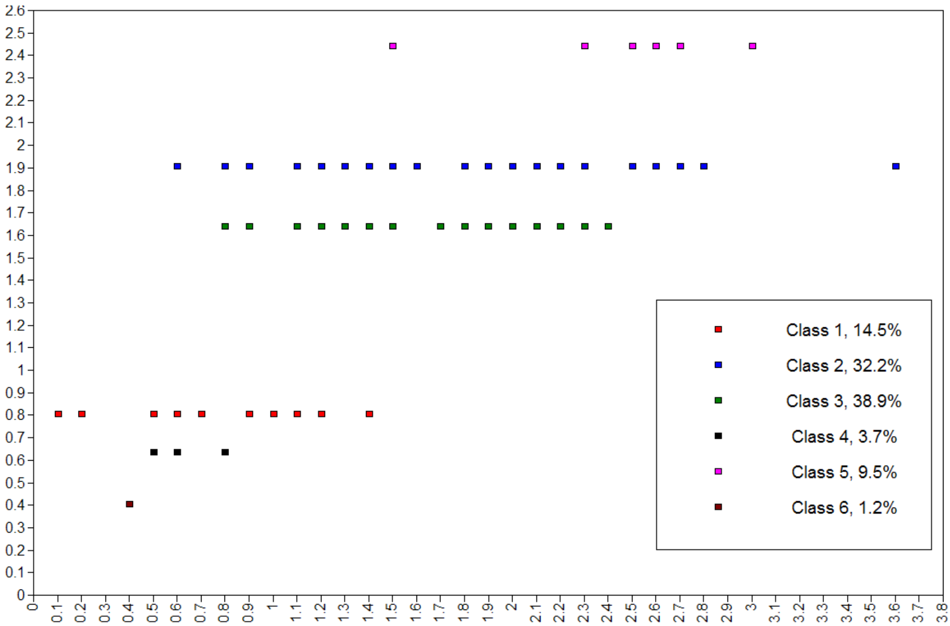
Şekil 4.3. Mutluluk kaynağı olarak anne/baba değişkenin alt gruplara dağılımı.

Şekil 4.3' de mutluluk kaynağı olarak kendisi; en yüksek ortalama ile alt grup 6' da (class 6), en düşük ortalama ile alt grup 2' de (class 2) elde edilmiştir. Böylece, kendisini mutluluk kaynağı olarak gören iller alt grup 6' da Mardin ili yer almıştır. Buna karşın mutluluk kaynağı olarak en az kendisini gören iller alt grup 2'de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir'den oluşmakt



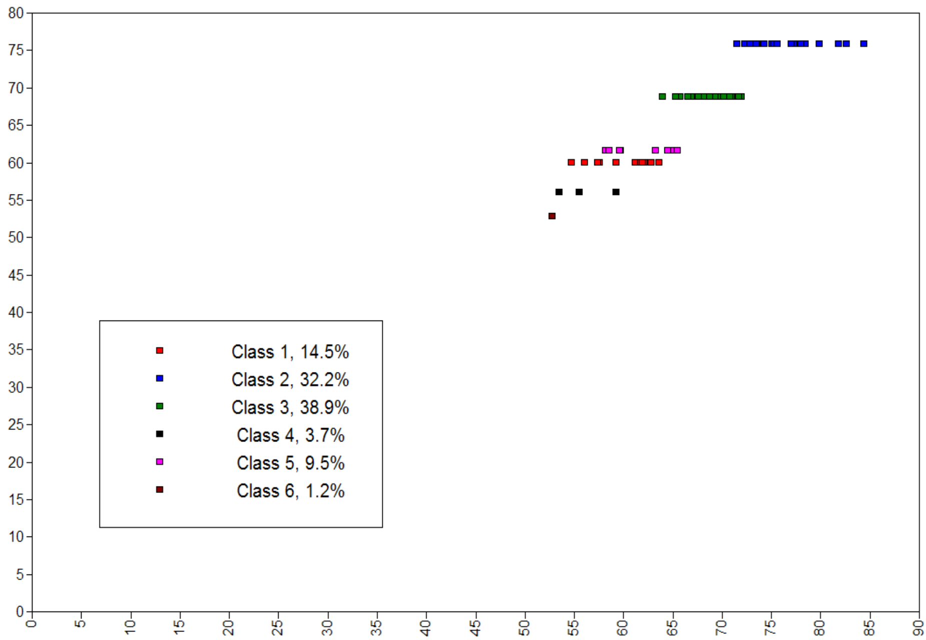
Şekil 4.4. Mutluluk kaynağı olarak kendisi değişkenin alt gruplara dağılımı.

Şekil 4.4' de mutluluk kaynağı olarak torun; en yüksek ortalama ile alt grup 5' de (class 5), en düşük ortalama ile alt grup 6' da (class 6) elde edilmiştir. Bu bağlamda, torunları mutluluk kaynağı olarak gören iller alt grup 5' de yer almış olup ve sırasıyla; Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat illerinden oluşmaktadır. Buna karşın mutluluk kaynağı olarak en az torunları gören il alt grup 6' da Mardin olduğu saptanmıştır.



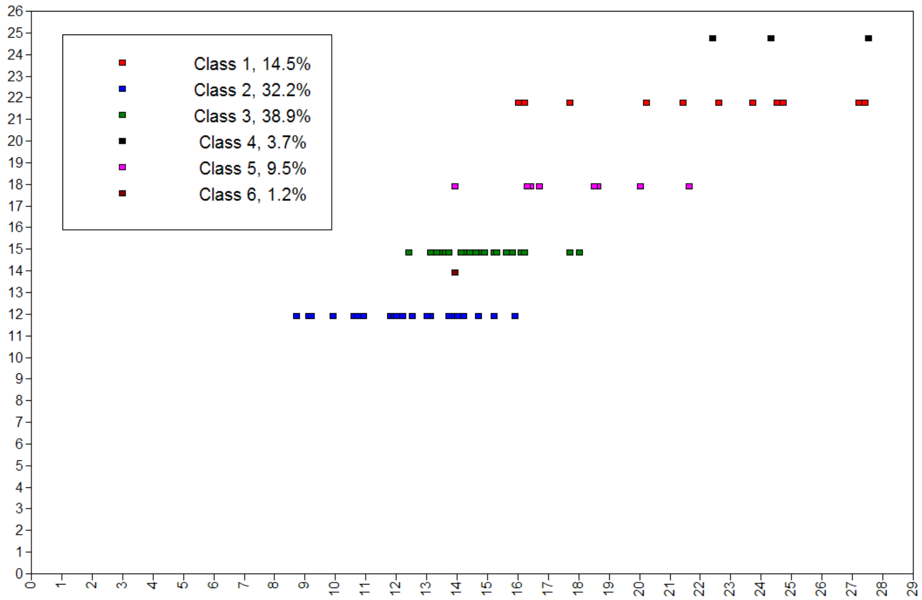
Şekil 4.5. Mutluluk kaynağı olarak torun değişkenin alt gruplara dağılımı.

Şekil 4.5' de mutluluk kaynağı olarak sağlık; en yüksek ortalama ile alt grup 2' de (class 2), en düşük ortalama ile alt grup 6' da (class 6) elde edilmiştir. Buna göre, sağlığı mutluluk kaynağı olarak gören iller alt grup 2' de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olarak belirlenmiştir. Buna karşın mutluluk kaynağı olarak en az sağlığı gören il alt grup 6' da Mardin olmuştur.



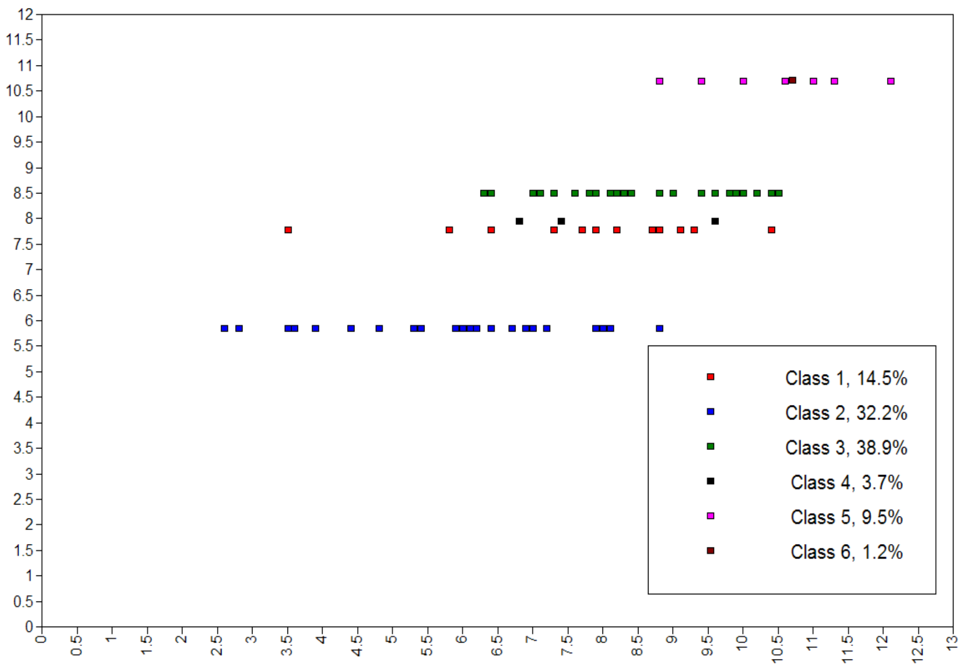
Şekil 4.6. Mutluluk kaynağı olarak sağlık değişkenin alt gruplara dağılımı.

Şekil 4.6' da mutluluk kaynağı olarak sevgi; en yüksek ortalama ile alt grup 4' de (class 4), en düşük ortalama ile alt grup 2' de (class 2) elde edilmiştir. Bu bağlamda, sevgiyi mutluluk kaynağı olarak gören iller alt grup 4'de yer almış olup ve sırasıyla; Ağrı, Bitlis ve Muş olarak saptanmıştır. Buna karşın mutluluk kaynağı olarak sevgiyi en az gören iller alt grup 2' de yer almış olup, sırasıyla Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olarak belirlenmiştir.



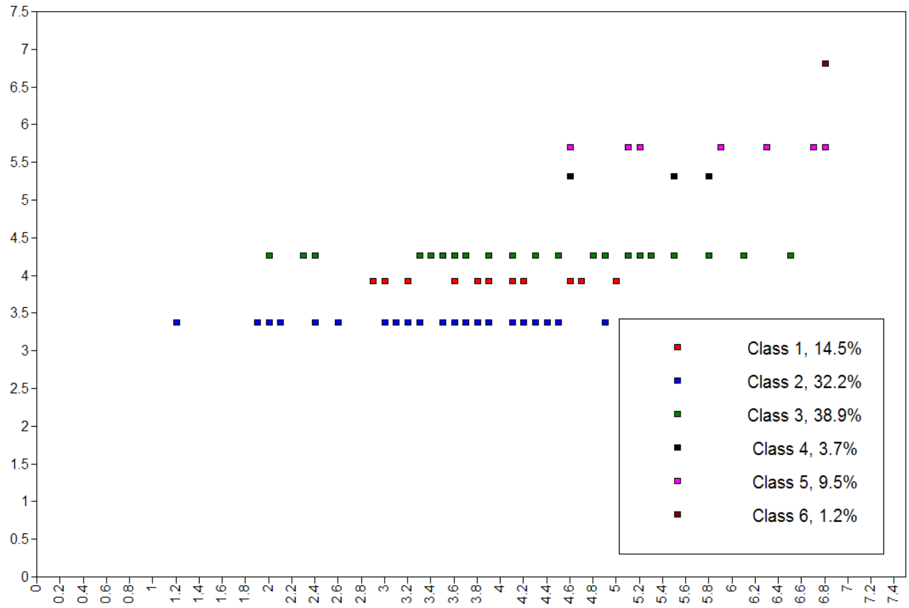
Şekil 4.7. Mutluluk kaynağı olarak sevgi değişkenin alt gruplara dağılımı.

Şekil 4.7' de mutluluk kaynağı olarak başarı; en yüksek ortalama ile alt grup 5' de (class 5), en düşük ortalama ile alt grup 2' da (class 2) elde edilmiştir. Bu bağlamda, başarıyı mutluluk kaynağı olarak gören iller alt grup 5' de yer almış olup ve sırasıyla; Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat olarak belirlenmiştir. Buna karşın mutluluk kaynağı olarak başarıyı en az gören iller alt grup 2' de yer almış olup, sırasıyla Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olduğu saptanmıştır.



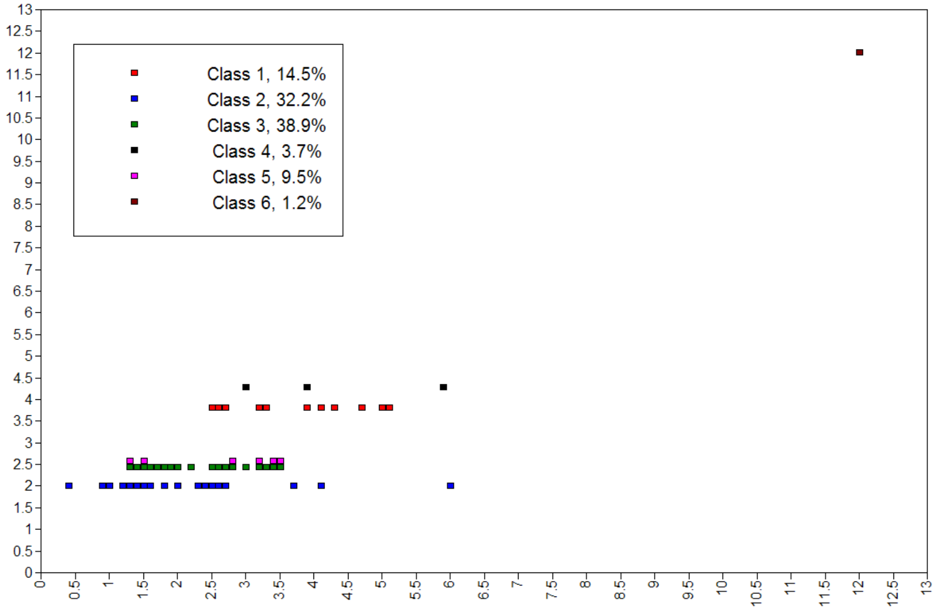
Şekil 4.8. Mutluluk kaynağı olarak başarı değişkenin alt gruplara dağılımı.

Şekil 4.8' de mutluluk kaynağı olarak para; en yüksek ortalama ile alt grup 6' da (class 6), en düşük ortalama ile alt grup 2' de (class 2) elde edilmiştir. Bu bağlamda, parayı mutluluk kaynağı olarak gören il alt grup 6' da Mardin olmuştur. Buna karşın mutluluk kaynağı olarak en az para diyen iller alt grup 2' de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olarak belirlenmiştir.



Şekil 4.9. Mutluluk kaynağı olarak para değişkenin alt gruplara dağılımı.

Şekil 4.9 ’da mutluluk kaynağı olarak iş; en yüksek ortalama ile alt grup 6’ da (class 6), en düşük ortalama ile alt grup 2’ de (class 2) elde edilmiştir. Bu sonuçlara göre, işi mutluluk kaynağı olarak gören il alt grup 6’ da Mardin olmuştur. Bununla birlikte, mutluluk kaynağı olarak en az iş diyen iller alt grup 2’ de yer almış olup ve sırasıyla; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olarak belirlenmiştir.



Şekil 4.10. Mutluluk kaynağı olarak iş değişkenin alt gruplara dağılımı.

5. TARTIŞMA VE SONUÇ

Bu çalışmada, 2013 yılında, illere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kullanılarak Türkiye'deki illerin mutluluk haritası elde edilmiştir. Başka bir ifadeyle, normal karışımlı model kullanılarak Türkiye'deki illerin kendi içlerinde homojen ve aralarında heterojen olmak üzere 6 alt grup oluşturulacağı saptanmıştır.

Çok değişkenli istatistiksel yöntemlerde en önemli amaç, karmaşık bir veri setinin kavramsal olarak yorumlanmasını kolay bir formatta sunmaktır. Bu nedenle son yıllarda buna yönelik sınıflandırma amaçlı istatistiksel yöntemler üzerinde durulmaktadır. Karışımlı model yaklaşımı bu yönü ile diğer yöntemlere nazaran istatistiksel yorumlamada önemli avantajlar sağlamaktadır. Bunlar; gözlemlerin doğru sınıflandırma olasılığını sağlamak, elde edilen her alt grup için ayrı parametre tahminini yapmak, sınıflandırmadaki hata payını en aza indirmektir **(**Yeşilova ve ark., 2009**)**.

Karışımlı model yaklaşımında heterojen yapı gösteren veri setinin homojen kaç alt gruba ayrılacağını belirlemek için AIC ve BIC istatistikleri yaygın olarak kullanılmaktadır. Genel olarak en küçük AIC ve BIC değerlerine sahip model en iyi model olarak seçilmektedir (Wang ve Putterman, 1998; Wang ve ark., 1998; Stele ve Raftery, 2009). Bu çalışmada, en küçük AIC ve BIC istatistikleri 6 alt gruplu modelden elde edilmiştir. Bu nedenle tüm parametre tahminleri bu modele göre yorumlanmıştır. Böylece, Türkiye' de illere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kendi içlerinde homojen ve kendi aralarında heterojen olmak üzere 6 alt gruba ayrılmıştır. Bunun yanı sıra, karışımlı modellemede, veri setinin dağılacağı uygun modeli saptamak için entropy doğru sınıflandırma ölçütü yaygın olarak kullanılmaktadır (Banavent ve ark., 2009). Bu çalışmada, en yüksek entropy ölçütü 6 alt gruplu modelde %96.8 olarak elde edilmiştir (Çizelge 4.1). Veri setinin dağıldığı en iyi model olan 6 alt gruplu modelde her bir alt grup için entropy doğru sınıflandırma oranları 0.916-1.00 arasında oldukça yüksek oranda olmuştur (Çizelge 4.3). Alt grup 4 ve alt grup 6' da gözlemlerin doğru sınıflandırma oranları %100 olarak saptanmıştır. Çizelge 4.3' de, en düşük entropy sınıflandırma oranı % 91.6 ile alt grup 2' de olmuştur. Bu gruptaki gözlemlerin % 8.4' ü alt grup 3' de sınıflandırılmıştır.

İllere göre mutluluk kaynağı olarak; çocuklar, eş, anne/baba, kendisi, torunları, sağlık, sevgi, başarı, para ve iş gibi değişkenler kullanılarak en iyi model olarak seçilen 6 alt gruplu modelde her bir alt grupta olası değişkenler için elde edilen ortalama tahmin değerleri alt gruplar arası farklılık göstermiştir (Çizelge 4.4). Karışımlı modellemeyi kümeleme analizi gibi diğer çok değişkenli analizlerden ayıran en önemli avantajlarından biri, elde edilen alt gruplar için parametre tahminlerinin elde edilmesidir (Yeşilova, 2003; Guang Sung, 2004; Di Zio ve ark., 2007; Cao, 2010; Anonim, 2016a). Karışımlı modelleme sınıflandırma amaçlı olarak kullanıldığı gibi, hem regresyon analizi hem de sınıflandırma için birlikte kullanılmaktadır (Muthen ve Muthen, 2002; Shi ve ark., 2005; Ng ve McLachlan, 2014). Örneğin normal karışımlı regresyon, karışımlı Poisson regresyonu ve karışımlı lojistik regresyon gibi modeller kullanılmaktadır (Wang ve Putterman, 1998; Wang ve ark., 1998; Reynolds, 2008; Rodriguez, 2011; Montuelle ve ark., 2013; Plasse, 2013).

Çizelge 4.4'e bakıldığında, mutluluk kaynağı olarak çocuklar için en yüksek ortalama değeri alt grup 5 en düşük ortalama değeri ise alt grup 4'de elde edilmiştir. Çocukları mutluluk kaynağı olarak gören illere bakıldığında; Doğu Anadolu, Güneydoğu Anadolu ve Karadeniz bölgelerinde (Ordu ili hariç) hiçbir ilin yer almadığı saptanmıştır. Bu bölgelerde doğurganlığın fazla olduğu söylenebilir. Dolayısıyla çocuk sayısı arttıkça çocuk önemini yitirmektedir. Kağıtçıbaşı (1991), tarafından yapılan çalışmada, çocuğun değeri araştırmasında Türkiye' de çocuğun ekonomik değerinin de ön plana çıktığı belirtilmiştir. Burada ilginç olan, anne-babanın çocuğa atfettiği bu iki değerin sosyal değişme içinde birbirine zıt yönde değişmesidir. Özellikle, yaşanılan yörenin gelişmişlik düzeyi, aile geliri, eğitim ve özellikle kadının eğitim düzeyi yükseldikçe, kırdan kente hareketlilik arttıkça ve nihayet ailede mevcut olan çocuk sayısı azaldıkça, çocuğun genel ekonomik değeri önem kaybetmektedir. Aynı gelişmelerle çocuğun psikolojik değeri de artmaktadır. Ayrıca çocuğun sevgi sağlayıcı ve aileyi tamamlayıcı işlevi önem kazanmaktadır. Geleneksel yapı içinde çocuk, sağladığı maddi yarar bakımından değer taşırken, gelişme süreci ile çocuğun psikolojik değeri artmaktadır. Çünkü çocuğun faydacı (ekonomik) değeri çocuk sayısıyla bağlantılı 'birikimli' bir değerdir (daha çok çocuk maddi katkı demektir). Oysa psikolojik değer ilave çocukta artmamakta, hatta azalabilmektedir. Az sayıda çocuk da çok sayıda çocuğun sağladığı mutluluğu verebilir.

Çizelge 4.4' de, eş değişkeni için en yüksek ortalama değeri alt grup 4' de en düşük ortalama değeri ise alt grup 2' de elde edilmiştir. Eşi mutluluk kaynağı olarak gören iller diğer illere nazaran oldukça yüksek bir ortalama değer (16.797) ile Ağrı, Bitlis ve Muş olmuştur. Eşi mutluluk kaynağı olarak görmeyen illerin dağılımı ülke geneli diğer iller olup, Doğu Anadolu ve Güneydoğu Anadolu illerinde yalnızca Bingöl ili yer almıştır. S. Parin' e göre (2017, sözlü görüşme) sosyoekonomik düzeyi düşük olan muhafazakâr, evlilik dışı ilişkinin az olduğu ayrıca boşanma oranının düşük olduğu yerlerde eş mutluluk kaynağı olarak görülmektedir.

Sosyal değişkenler açısından değerlendirildiğinde illerin mutluluk düzeyleri ile boşanma oranları arasında negatif korelasyon tespit edilmiştir. Buna göre, boşanma oranı azaldıkça mutluluk düzeyi artmaktadır (Akgiş, 2015). Türkiye' de geleneksel yaşam tarzına sahip erkeklerin daha çok söz sahibi olduğu ve feodal yapının hakim olduğu bölgelerde, eş mutluluk kaynağı olarak görülmüştür. Bu nedenle, çalışmamızda kullanılan veri setinde cinsiyet faktörü ele alınmadığından Doğu Anadolu ve Güneydoğu Anadolu bölgelerinde eşin mutluluk kaynağı olarak görüldüğü belirlenmiştir.

Anne ve babayı mutluluk kaynağı gören iller sıralamasında Mardin başta olmak üzere diğer Doğu ve Güneydoğu illerin çoğu yer almıştır (Çizelge 4.4). Diğer bölgelerdeki iller ise genellikle anne ve babayı mutluluk kaynağı olarak daha az görmektedirler. Dinsel ve kültürel anlamda anne-babayı önemseyen ve aile bağlarının güçlü olduğu yerlerde anne-baba mutluluk kaynağı olarak görülmektedir (S. Parin, sözlü görüşme).

Mutluluk kaynağı olarak kendisini görenler için Çizelge 4.4' de bakıldığında, en yüksek ortalama değer 3.800 ile Mardin ili olmuştur. Sonrasında ise Mardin ilini Ağrı, Bitlis ve Muş illeri izlemiştir. Doğu Anadolu ve Güneydoğu Anadolu bölgeleri bu açıdan daha çok ön plana çıkmıştır Geleneksel toplumsal yapı içerisinde birey her yönden kuşatılmıştır. Birey kendisini önemsediği ölçüde bu durumdan çıkar. Az gelişmiş toplumsal yapılarda bu durum daha çok baskındır (S. Parin, sözlü görüşme).

Mutluluk kaynağı olarak torunlarını görenler, en yüksek ortalama (2.438) ile Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova ve Yozgat illeri olmuştur. Bununla birlikte torunları mutluluk kaynağı olarak daha az gören iller, başta Mardin olmak üzere Ağrı, Bitlis ve Muş olarak belirlenmiştir. Çocuk sayısının az olduğu yerlerde torunlara daha çok önem verilmektedir. Kalabalık ailelerde sevgi yakınlar ile paylaşıldığında torun daha az değer görür. Hane büyüklüğü artıkça torun mutluluk kaynağı olarak görülmemektedir (S. Parin, sözlü görüşme).

Sağlık durumunu mutluluk kaynağı olarak görenlerin ortalamasının en yüksek olduğu alt grup 75.775 ortalama ile 2. alt grup olmuştur. Bu gruptaki illerin ise Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olduğu saptanmıştır. Mutluluk kaynağı olarak sağlığı en az görenler alt grup 6 ve alt grup 4 olmuştur. Bu alt gruptaki iller ise başta Mardin olmak üzere Ağrı, Bitlis ve Bingöl olarak sıralanmıştır. Genellikle Doğu Anadolu ve Güneydoğu Anadolu bölgeleri sağlık durumunu mutluluk kaynağı olarak pek görmemektedir.

Sağlık insanların hayatlarını etkileyen en değerli durumdur. Ayrıca sağlık, bireylerin yaşam koşullarını belirleyen temel bir faktördür. Bu nedenle sağlık ve mutluluk birbirlerini olumlu etkilemektedir (Akar, 2014). Maslow' a göre fizyolojik ihtiyaçlar ilk basamağı oluşturur.

Çünkü insanın temel fizyolojik gereksinimleri doyurulmadan üst düzeydeki ihtiyaçlara gereksinim sağlanmayacaktır. Örneğin aç ve susuz bir insan bu ihtiyacı giderilmeden sağlık ihtiyacını karşılamayı düşünmeyecektir (S. Parin, sözlü görüşme).

Mutluluk kaynağı olarak sevgiyi en yüksek görenler alt grup 4 ve alt grup 1'de sırasıyla 24.732 ve 21.737 olarak sıralanmışlardır. Bu alt gruptaki iller; Ağrı, Bitlis, Muş, Ardahan, Batman, Diyarbakır, Erzurum, Erzincan, Hakkâri, Iğdır, Kars, Siirt, Van, Şanlıurfa ve Şırnak şeklinde belirlenmiştir. Sevgiyi mutluluk kaynağı olarak alan illerin genellikle Doğu Anadolu ve Güneydoğu Anadolu bölgelerinde oldukları görülmektedir. Mutluluk kaynağı olarak sevgiyi en az görenler ise 11.870 ortalama ile Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir illerinin olduğu alt grup 2' de gözlenmiştir. Daha geleneksel olan ve kentleşme oranı düşük olan illerde sevgi daha çok ön plana çıkmaktadır.

Başarıyı mutluluk kaynağı olarak en yüksek görenler 10.700 ve 10.699 ortalamaları ile alt grup 6 ve alt grup 5 olmuşlardır. Bu alt gruptaki illerin sırasıyla; Mardin, Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat oldukları saptanmıştır. Başarı değişkeni bakımından en düşük olan alt grup 5.832 ortalama ile alt grup 2 olmuştur. Bu gruptaki iller; Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir olarak belirlenmiştir. Başarının mutluluk kaynağı olarak görülmesi bakımından illerin, belli bölgelerde yoğunlaşmadıkları görülmektedir. S. Parin' e göre (2017, sözlü görüşme) bireyselliğin olduğu yerlerde başarı değer görür. Her başarı bireyin kendi hanesine yazıldığı için başarı önem kazanır. Bireyselleşmenin başarıda büyük etkisi olduğu söylenebilir.

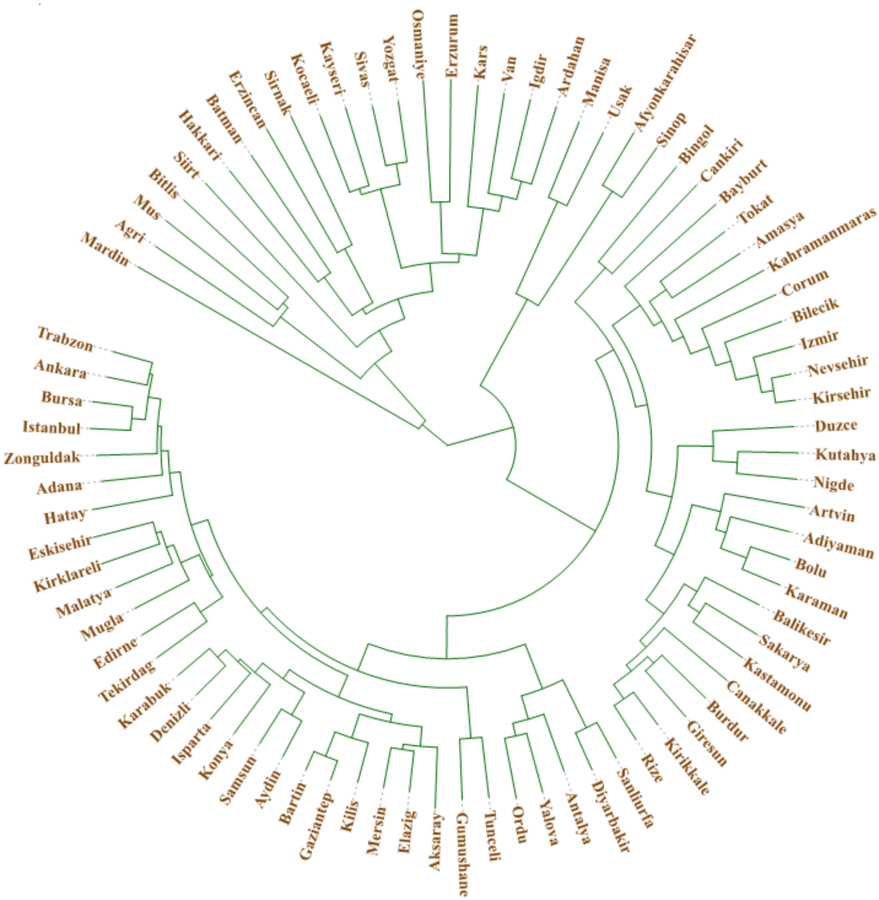
Mutluluk kaynağı olarak parayı en yüksek görenlerin alt grup 6' da 6.800 ortalama ile Mardin ili olduğu belirlenmiştir. Bunu 5.695 ortalama ile alt grup 5 takip etmiştir. Alt grup 5'teki illerin ise; Antalya, Edirne, Kayseri, Kocaeli, Ordu, Sivas, Yalova, Yozgat oldukları saptanmıştır. Parayı mutluluk kaynağı olarak daha az görenlerin 3.369 ortalama ile Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir illerini içeren alt grup 2 olduğu görülmüştür. Başarı mutluluk kaynağında olduğu gibi para değişkeninde de illerin, belli bölgelerde yoğunlaşmadıkları görülmektedir. Sanayinin çok gelişmediği yerlerde paranın daha çok önemli olduğu söylenebilir. Yaşam standartları hakkında fikir veren en önemli gösterge kişi başına düşen milli gelirdir. Bu nedenle ilk çalışmalar kişi başı GSYİH (Gayri Safi Yurt İçi Hâsıla) ve mutluluk arasındaki ilişkiyi incelemiş ve kişi başı gelir düzeyi arttıkça mutluluk düzeyinin de arttığı görülmüştür. Buna göre kişi başı gelirin yüksek olduğu bölgelerde yaşam memnuniyeti (diğer bölgelere kıyasla) daha yüksek çıkmaktadır (Veenhoven ve Dumludağ, 2015).

Para değişkenine benzer olarak, mutluluk kaynağı olarak işi en yüksek görenlerin yine alt grup 6' da 12.000 ortalama ile Mardin ilinin olduğu belirlenmiştir. İşi mutluluk kaynağı olarak daha az görenlerin 1.982 ortalama ile Adıyaman, Afyonkarahisar, Amasya, Balıkesir, Bayburt, Bilecik, Bolu, Burdur, Düzce, Giresun, Kahramanmaraş, Karaman, Kastamonu, Kütahya, Kırıkkale, Manisa, Kırşehir, Nevşehir, Niğde, Rize, Sakarya, Sinop, Tokat, Uşak, Çanakkale, Çankırı, Çorum ve İzmir illerini içeren alt grup 2' de olduğu görülmektedir. Başarı ve para mutluluk kaynaklarında olduğu gibi para değişkeninde de illerin, belli bölgelerde yoğunlaşmadıkları görünmektedir. Akgiş (2015), yaptığı çalışmaya göre sektörel anlamda gelişmemiş, ilerlememiş veya ilerlemesi yavaş olan şehirler işi mutluluk kaynağı olarak görmektedir. Bu illerde kişiler kendi iş alanını oluşturma çabası içerisindedir. İllerin net göç hızları ile mutluluk oranları arasında negatif korelasyon olduğu belirlenmiştir. Buna göre göç hızı arttıkça mutluluk oranları düşmekte, göç hızı azaldıkça mutluluk oranları artmaktadır.

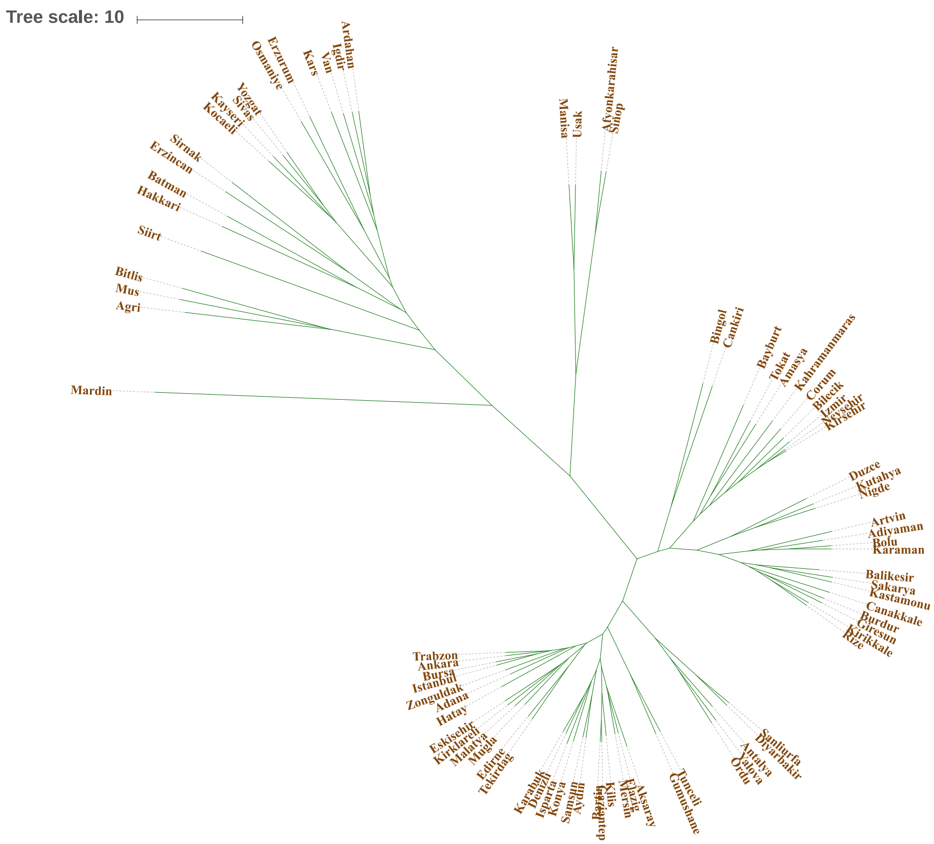
Selim (2008), yaptığı çalışmada kentsel kesimde yaşayan insanlar güçlü olmanın bireysel mutluluğu getirdiğini, kırsal kesimde yaşayan insanlardan 2 kat daha fazla düşünmektedir. İş sahibi olmanın mutluluk getirdiğine inananlar için bu oran biraz daha azdır. Para ve iş kategorilerinde eğitimin önemli bir rolü olduğu açıktır. Bireylerin işteki durumu incelediğinde düzenli ücret ile çalışanlar gücün mutluluk getirdiği inancını herhangi bir işi olmayanlara göre daha az taşırlar. Ayrıca, işveren olarak çalışan bireyler işsizlere göre yaklaşık olarak 2 kat daha fazla başarının mutluluk getirdiğini belirtmişlerdir. Yevmiyeli olarak çalışanlar ise paranın mutluluk getirdiğine işsizlerle karşılaştırıldığında 3 kat daha fazla inanmaktadır. Gelecekten umudu az olan bireyler başarının mutluluk getirmediğine, iş sahibi olma ve paranın ise daha fazla mutlu eden değerler olduğuna inanmaktadırlar.

Çalışmada, veri setine ayrıca kümeleme analizi de uygulandı. Kümeleme analizi sonucunda, mutluluk kaynağı olarak alınan söz konusu değişkenlerin dağılımı Şekil 5.2 ve Şekil 5.3 ve Şekil 6.1' de verilmiştir. Şekil 5.2, Şekil 5.3 ITOL (Interactive Tree Of Life) programı; Şekil6.1 ise SPSS programı kullanılarak elde edilmiştir. Çizelge 4.5' de normal karışımlı modelleme kullanılarak elde edilen illerin dağılımı ile Şekil 5.1, Şekil 5.2 ve Şekil 6.1 (eklerde) sonuçları bir kümeleme analizindeki sonuçlar ile kısmen benzerlik göstermektedir. Örneğin normal karışımlı modellemede; 4. alt grup, 6. alt grup ve 1. alt gruptaki illerin dağılımı ile kümeleme analizindeki alt gruplar benzerlik göstermiştir. Bununla birlikte, normal karışımlı modeldeki diğer alt grupların, kümeleme analizindeki karşılığının çok belirgin olmadığı saptanmıştır. Kümeleme analizi sonuçlarına bakıldığında, mutluluk düzeyi bakımından bazı illeri sınıflandırmak pek mümkün görünmemektedir. Başka bir deyişle bazı iller bazında çok dağınık bir sınıflama yapıldığı söylenebilir. Bu açıdan bakıldığında; mutluluk düzeylerine göre illerin sınıflandırılmasında, normal karışımlı modelin daha açıklayıcı bilgiler verdiği saptanmıştır (Şekil 5.3).

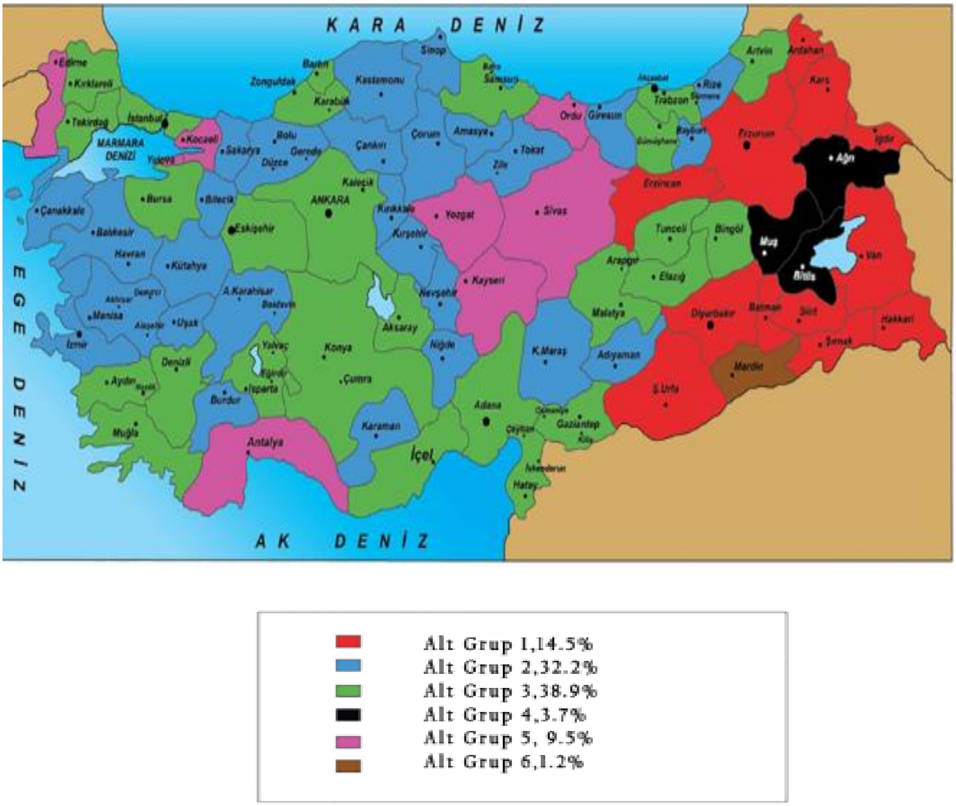
Sonuç olarak; sürekli verilerin sınıflandırılmasında, karışımlı modellerin daha iyi sonuçlar verdiği saptanmıştır. Karışımlı modelleme; veri setini kendi içerisinde homojen alt gruplara ayırmakla birlikte, elde edilen her bir alt grup için parametre tahminlerini vermektedir. Bu yönü ile gerek faktör analizi gerekse kümeleme analizinden farklılık göstermektedir.



Şekil 5.1. Multululuk değişkenlerine göre kümeleme analizi sonucu illerin dağılımı.



Şekil 5.2. Multululuk değişkenlerine göre kümeleme analizi sonucu illerin dağılımı.



Şekil 5.3. Türkiye’deki illerin kendi içlerinde homojen alt grup haritası.

KAYNAKLAR

Akar, S., 2014. Türkiye'de daha iyi yaşam endeksi: OECD ülkeleri ile karşılaştırma. *Journal of Life Economics,* 1: 1-12.

Akgiş, Ö., 2015. Bir refah göstergesi olarak Türkiye'de mutluluğun mekânsal dağılışı. *Türk Coğrafya Dergisi,* 65: 69-76.

Alpar, R., 2017. *Çok değişkenli istatistiksel yöntemler,* Detay Yayıncılık, Yay. No. 5 Ankara.

Anonim, 2016a. Chapter 2. Finite mixture models. *http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000003441/02_chapter2.pdf.* Erişim tarihi: 10.12.2016.

Anonim, 2016b. Chapter 10. Examples: Multilevel mixture modeling. *https://www.statmodel.com/download/usersguide/Chapter10.pdf.* Erişim tarihi: 07.12.2016.

Anonim, 2016c. Mixture models. *http://www.bayesserver.com/docs/techniques/mixture-models.* Erişim tarihi: 14.12.2016.

Anonim, 2017a. Note Set 4: The EM algorithm for gaussian mixtures probabilistic learning: theory and algorithms, CS 274A. *http://www.ics.uci.edu/~smyth/courses/cs274/notes/notes5b.* Erişim tarihi: 27.07.2017

Anonim, 2017b. Chapter 20: Mixture models. *http://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch20.pdf.* Erişim tarihi: 27.07.2017.

Arı, Ç., Aksoy, S., Arıkan, O., 2012. Maximum likelihood estimation of gaussian mixture models using stochastic search. *Pattern Recognition,* 45: 2804–2816

Bailey, T. L., Elkan, C., 1993. Fitting a mixture model by expectation maximization to discover motifs ın biopolymers. *UCSD Technical Report*, CS94-351.

Banavent, A. P., Ruiz, F. E., Saez, J. M., 2009. Learning Gaussian mixture models with entropy-based criteria. *IEEE Transactions on Neural Networks,* 20 (11).

Barber, D., 2012. *Machine Learning Concepts. Bayesian reasoning and machine learning,* Cambridge University Press. University College London. 305-321.

Bashir, S., 2003. *High Breakdown Mixture Discriminant Analysis,* (doktora tezi) Guelph University, Canada.

Baudry, J. P., Celeux, G., 2015. EM for mixture – initialization reguires special care. *Statistics and Computing,* 25 (4): 713-726.

Bauer, D. J., Curran, P. J., 2003. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8: 338-363.

Cao, M., 2010. Practice on classification using Gaussian mixture model. *Course Project Report for COMP-135.* Computer Science, Tufts University, Medford, USA

Chen, Z., Kou, L., 2001. A note on the estimation of the multinominal logit model with random effects. *The American Statisticion*, 55 (2): 89-95.

Constantinopoulos, C., Michalis, K., Titsias, K. M., Likas, A., Member, S., 2006. Bayesian feature and model selection for Gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (6): 1013-1018.

Dahua, L., 2013. Online Learning of nonparametric mixture models via sequential variational approximation. *Advances in Neural Information Processing Systems,* 26: 395-403.

DeSarbo, W. S., Cron, W. L., 1988. A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification* 5: 249-282.

Dhanavanthan, P., 2000. Estimation of the parameters of compound ıntervened poisson distribution. *Biometrical Journal*, 42 (3): 315-320.

Di Zio, M., Guarnera, U., Rocci, R., 2007. A mixture of mixture models for a classification problem: the unity measure error. *Computational Statistics & Data Analysis*, 51: 2573–2585.

Ding, C., 2006. Using regression mixture analysis in educational research. *Practical Assessment Research & Evaluation*, 11 (11): 1-11.

Duncan, T. E., Susan, S. C., Strycker, L. A., Okut, H., 2002. Chapter addendum to an iıntroduction to latent variable growth curve modeling: concepts, issues, and applications. *Growth Mixture Modeling of Adolescent Alcohol Use Data.* Oregon Research Institute, Oregon.

Editorial, 2007. Advances in mixture models. *Computational Statistics & Data Analysis,* 51: 5205–5210.

Eirola, E., Lendasse, A., 2013. Gaussian mixture models for time series modelling, forecasting, and ınterpolation. *Advances in Intelligent Data Analysis XII*, 162-173.

Erten, H., 2015. *Veri Madenciliği Teknikleri İle Organ Nakli İçin Uygun Donör Oranının Hesaplanması* (yüksek lisans tezi). Gazi Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Ankara.

Everitt, B. S., Hand, D. J., 1981. *Finite mixture distributions*. London: Chapman and Hall. 143.

Faria S., Soromenho, G., 2008. Comparison of mixture and classification maximum likelihood approaches in poisson regression models. *Springer,* 1-8.

Faria, S., Soromenhob, G., 2010. Fitting mixtures of linear regressions. *Journal of Statistical Computation and Simulation,* 80 (2): 201–225.

Gross, R., Yang, J., Waibel, A., 2000. Growing gaussian mixture models for pose ınvariant face recognition. *IEEE,* 1088-1091.

Grosse R., Srivastava, N. Lecture 16: Mixture models. *http://www.cs.toronto.edu/~rgrosse/csc321/mixture_models.pdf*. Erişim tarihi: 27.07.2017

Grün, B. Bayes Mix: An R package for Bayesian mixture modeling. *http://ifas.jku.at/gruen/BayesMix/bayesmix-intro.pdf*. Erişim tarihi: 04.08.2017

Guang Sung, H., 2004. *Gaussian Mixture Regression and Classification* (doctoral thesis). Rice University, Houston, Texas.

Güngör, D., Korkmaz, M., Sazak, S.H., 2015. Örtük sınıf analiziyle yapılan ölçme eşdeğerli çalışmalarda model seçimi. *H. U. Journal of Education,* 30 (1): 90-105.

Halbe, Z., Aladjem, M., 2005. Model based mixture discriminant analysis-an experimantal study, *Science Direct*, 38 (3): 437-440.

Jeroen, K., 2004. Latent class regression analysis. *Workshop at the 24th Biennial Conference of SMABS*, 17-22 July 2004, Jena University.

Kağıtçıbaşı, Ç., 1991. Türkiye'de Aile Kültürü. *Çağdaş Kültürümüz: Olgular-Sorunlar*. Çağdaş Yaşamı Destekleme Derneği Yayınları 2, Cem Yayınevi, 49-57.

Karlis, D., Xekalaki, E., 2003. Choosing initial values for the EM algorithm for finite mixtures. *Computational Statistics & Data Analysis,* 41: 577 – 590.

Kayri, M., Göktaş, İ., 2005. Beden eğitimi yetenek sınavına katılan adayların başarılarını etkileyen değişkenlerin incelenmesinde karışımlı model tekniğinin kullanımı. *Hacettepe Journal of Sport Sciences,* 16 (4): 185-199.

Lakshmanan, V., Kain, J.S., 2010. A gaussian mixture model approach to forecast verification. *Weather and Forecasting,* 25: 908–920.

Leisch, F., 2004. FlexMix: a general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software,* 11 (8): 1-18.

Lindsay, B. G., Basak, P., 1993. Multivariate normal mixtures: a fast consistent method of moments. *Journal of American Statical Association*, 88 (422): 468-476.

Lukociene, O., Vermunt, J. K., 2010. Advances in Data Analysis, Data Handling and Business Intelligence, Studies in Classification, Data Analysis, and Knowledge Organization. Chap. *Determining the Number of Components in Mixture Models for Hierarchical Data*. (Editör: A. Fink et al.,). Springer-Verlag Berlin Heidelberg, 241-249.

Marin, J. M., Mengersen, K., Robert, C. P., 2005. Bayesian modelling and ınference on mixtures of distributions. *Handbook of Statistics,* 25: 459-507.

McLachlan, G.J., Peel, D., and Prado, P., 1997. Clustering via normal mixture models. *Proceedings of the American Statistical Association.* Anaheim, August 1997. Alexandria, Virginia: American Statistical Association, 98-103.

McLachlan, G., Peel, D., 2000. *Finite Mixture Models*. New York, John Wiley&Sons, Inc, 419.

McNicholas, P. D., Murphy, T. B. 2008. Parsimonious gaussian mixture models. *Statistics and Computing,* 8 (3): 285–296.

Montuelle, L., Pennec, E. L., Cohen, S. X., 2013. Gaussian mixture regression model with logistic weights, a penalized maximum likelihood approach. *Research Report, 8281,* 37.

Michalowicz, J. V., Nichols J. M., Bucholtz, F., 2008. Calculation of differential entropy for a mixed gaussian distribution. *Entropy,* 10: 200-206.

Murphy, K. P., 2006. Mixture Models. *http://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall06/reading/mixtureModels.pdf*. Erişim tarihi: 14.12.2016

Muthen, L. K., Muthen, B., 2002. *Mplus: User's Guide*. Los Angeles: CA, Muthen Muthen.

Nasios, N., Bors, A.G., 2006. Variational Learning for Gaussian Mixture Models. *IEEE Transactions On Systems, Man, And Cybernetıcs—Part B: Cybernetics,* 36 (4): 849-862.

Ng, S. K., McLachlan, G., 2014. Mixture of regression models with latent variables and sparse coefficient parameters. *Proceedings of 21st International Conference on Computational Statistics*. 19- 22 August 2014, Hague, Netherlands. 223-231.

Oberski, D., 2016. Mixture Models: Latent Profile and Latent Class Analysis, Part IV. *Modern Statistical Methods for HCI (*Editor: J. Robertson, M. Kaptein). Springer International Publishing, 275-287.

Okut, H., Duncan, E. T., Duncan, C. S., Strycker, A. L., 2002. Latent variable mixture modelling: analyzing mixture and the structural portion of model. *Joint Statististical Meetings (JSM).* 11-15 August 2002, New York City.

Okut, H., Kayri, M., 2008. Özel yetenek sınavındaki başarıya ilişkin risk analizinin karışımlı lojistik regresyon modeli ile incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi,* 35: 227-239.

Öztürk A., 2006. *Esnek Ayırma Analizi ve Bir Uygulama* (doktora tezi). Eskişehir Osmangazi Üniversitesi, Eskişehir.

Parin, S., 2017, sözlü görüşme, Edebiyat Fakültesi, Sosyoloji Bölümü.

Plasse, J. H., 2013. *The EM Algorithm in Multivariate Gaussian Mixture Models Using Anderson Acceleration.* (A Project Report), Faculty of The Worcester Polytechnic Institute.

Reynolds, D., 2009. Gaussian mixture models. *Encyclopedia of Biometrics*, 13-20.

Reynolds, D., 2008. Gaussian mixture models, Encyclopedia of Biometric Recognition, Springer,*https://ll.mit.edu/mission/cybersec/publications/publication-files/full_papers/0802_Reynolds_Biometrics-GMM.pdf.* Erişim tarihi: 10.12.2016

Rodriguez, M. A., 2011. *Implementation of Gaussian mixture models in net technology for automatic speech recognition.* (yüksek lisans tezi) University of Science and Technology in Krakow.

Rossi, P. E., 1995. *Bayesian Non- and Semi-parametric Methods and Applications.* Princeton University Press. ISBN 978-0-691-14532-7.

Selim, S., 2008. Türkiye'de bireysel mutluluk kaynağı olan değerler üzerine bir analiz: multinomial logit model, *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, 17(3): 345–358.

Sezgin, E., Çelik, Y., 2013. Veri madenciliğinde kayıp veriler için kullanılan yöntemlerin karşılaştırılması. *XV. Akademik Bilişim Konferansı,* 23-25 Ocak 2013, Antalya.

Shental, N., Bar-Hillel, A., Hertz, T., Weinshall, D., 2003. Computing Gaussian mixture models with em using equivalence constraints. *Technical Report-91904*. Hebrew University of Jerusalem, Israel.

Shi, J. Q., Murray–Smith, R., Titterington, D. M., 2005. Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, 15: 31–41.

Steele, J. R., Raftery, E. A., 2009. Performance of bayesian model selection criteria for Gaussian mixture models. *Technical Report, 559.* Department of Statistics University of Washington. 11.

Titterington, D. M., Smith, A. F. M., Markov, U. E., 1985. *Statistical Analyses of Finite Mixture Distributions.* Chichester, U. K., John Willey and Sons. 243.

Van Horn, M. L., Jaki, T., Masyn, K., Howe, G., Feaster, D. J., Lamont, A.E., George, M.R.W., Kim, M., 2015. Evaluating differential effects using regression interactions and regression mixture models. *Educ Psychol Mea,* 75 (4): 677-714.

Veenhoven, R., Dumludağ, D., 2015. İktisat ve mutluluk. *İktisat ve Toplum Dergisi*, 58: 46-51.

Verbeek, J. J., Vlassis, N., Kröse, B., 2003. Efficient greedy learning of gaussian mixture models. *Neural Computation*, 15 (2), 469-485.

Wang, P., Putterman, M. L., 1998. Mixed logistic regression models. *Journal of Agriculture, Biological and Environmental Statistics,* 3 (2):175-200.

Wang, P., Cockburn, I. M., Putterman, M. L., 1998. Analyses of latent data-mixed poisson regression model approach. *Journal of Business and Economic Statistics*, 16 (1), 27-41.

Wedel, M., DeSarbo, W. S., 2002. Applied Latent Class Analysis, Chap 13. *Mixture Regression Models* (Editors: Jacques A. Hagenaars, Allan L. McCutcheon) Cambridge University Press, 366- 382.

Wengrzik, J., 2012. *Parameter Estimation for Mixture Models Given Grouped Data.* (Doktora Tezi) Mathematik und Informatik, Universität, Bremen.

Yeşilova, A., 2003. *Biyolojik Çalışmalardan Elde Edilen Kategorik Verilere Karışık Poisson Regresyon Analizinin Uygulanması* (doktora tezi basılmamış). YYÜ, Fen Bilimleri Enstitüsü, Van.

Yeşilova, A., Atlıhan, R., 2007. Farklı sıcaklıkların *Scymnus subvillosus*'un bıraktığı yumurta sayıları üzerine etkilerinin karışımlı poisson regresyon ile analiz edilmesi. *Tarım Bilimleri Dergisi (J. Agric. Sci.),* 17 (2): 73-79.

Yeşilova, A., Yılmaz, A., Sers, D., Kaki, B., 2016. Modeling with Gaussian mixture regression for lactation milk yield in anatolian buffaloes. *Indian J. Anim. Res.,* 50 (6): 989-994.

Yeşilova, A., Özrenk, K., Kaki, B., Almalı, M. N., Balta, F. 2010. Locational classification of walnut (Juglans Regia L.) genotypes collected from lake van basin by using mixture modeling. *African Journal of Agricultural Res.,* 5 (12): 1509-1514.

6. EKLER

6.1. Sonlu Karışımlı Modeller İçin Bayesian Yaklaşım

K alt gruplu bir sonlu karışımlı model dağılımı aşağıdaki gibi yazılabilir. $f(x \mid \mathrm{K}, \pi, \theta) = \sum_{k=1}^{K} \rho_i f(x \mid \theta_i)$ (6.1)

Eşitlik 6.1' de $f(x \mid \theta_i)$ ilgili değişken için olasılık fonksiyonu; π , $\pi = (\pi_1, \ldots \pi_k)$ her bir alt grup için karışma olasılıkları vektörünü göstermektedir. Burada, tüm olası alt gruplar için karışma olasılıkların toplamı 1 değerine eşittir ($\sum_{k=1}^{K} \pi_i = 1$).

Bayesian yaklaşımı, Monte Carlo Markov zincirini kullanarak (MCMC) karışımlı modelde latent grupların oluşturulmasında ve parametre tahmininde oldukça kolaylık sağlamaktadır. k boyutlu bir sonlu Gaussian karışımlı yoğunluk aşağıdaki gibi tanımlanabilir.

$$f(x \mid k, \rho, \theta) = \sum_{i=1}^{k} \rho_i f_N(x \mid \mu_i, \phi_i) \tag{6.2}$$

Eşitlik 6,2' de, θ normal dağılım parametre vektörünü $\theta = (\mu_1, \phi_1, ..., \mu_k, \phi_k)$; $f_N(x \mid \mu_i, \phi_i)$, μ_i ortalamalı ve ϕ_i kesinlik (precision) parametreli normal dağılım fonksiyonu göstermektedir.

$x = (x_1, \ldots x_n)$ n boyutlu gözlem vektörünün yoğunluk fonksiyonu aşağıdaki gibi yazılabilir.

$$f(x \mid \pi, \theta) = \sum_{k=1}^{K} \pi_i f(x \mid \theta_i)$$

Eşitlikteki model parametreleri için olabilirlik fonksiyonu,

$$I(\rho,\theta|x)=\prod_{j=1}^{n}\sum_{i=1}^{k}\rho_i f(x_j|\theta_i) \tag{6.3}$$

biçiminde verilebilir. Eşitlik 6,3' te verilen olabilirlik fonksiyonunu basitleştirmek için z_j latent değişkenini aşağıdaki gibi tanımlanabilir.

$$X_J|Z_j=i\sim f(x|\theta_i) \qquad P(Z_j=i)=\pi_i$$

Burada, $x=(x_1,...x_n)$ veri setinin her bir örneği için $z=(z_1,\dots z_n)$ eksik veri setinin olduğu varsayılır. Bu eksik veri seti, tüm gözlemler için oluşturulur ve gözlemlerin hangi alt gruba girdiğini göstermektedir. Z değişkeni kullanılarak, olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$I(\pi,\theta|x,z)=\prod_{j=1}^{n}\pi_{z_j}f(x_j|\theta_{z_j})$$

$$=\prod_{k=1}^{K}\pi_i^{ni}\left[\prod_{j:z_j=k}f(x_j|\theta_k)\right]$$

(6.4)

Eşitlik 6.4' te,

$$n_k=\neq\{z_j=k\} \qquad \sum n_k=n$$

olmaktadır. Daha sonra x_j gözleminin k'ıncı alt grupta bulunmasına ilişkin posterior olasılığı aşağıdaki gibi hesaplanır.

$$P(z_j = k \left| x_j, \pi, \theta\right.) = \frac{\pi_i f(x_j \left|\theta_k\right.)}{\sum_{k=1}^{K} \pi_i f(x_j \left|\theta_k\right.)}$$

Gaussian karışımlı dağılım için olabilirlik fonksiyonu,

$$I(\pi, \mu, \varphi \left| x, z\right.) \alpha \prod_{k=1}^{K} (\pi_i \varphi)^{n_k} \exp(-\frac{\varphi}{2} \sum_{j: z_j = k} (x_j - \mu_k)^2)$$

Bu eşitlikte, $n_i =\neq \{z_j = i\}$ olmaktadır. Buradan, x_j gözleminin k' ıncı alt grupta bulunmasına ilişkin posterior olasılığı aşağıdaki gibi hesaplanır.

$$P(z_j = k \left| x_j, \pi, \mu, \theta\right.) = \frac{\rho_k \theta_k \exp\left\{-\frac{\theta_k}{2}(x_j - \mu_k)^2\right\}}{\sum_{k=1}^{K} \rho_k \theta_i \exp\left\{-\frac{\theta_k}{2}(x_j - \mu_k)^2\right\}}$$

Model parametreleri $\left(\pi, \mu, \phi\right)$ için aşağıdaki conjugate ön (prior) dağılımlar esas alınmaktadır.

Prior	Posterior
$\rho \sim \mathrm{D}(\delta_1, ..., \delta_k)$	$\rho \left| x, z\right. \sim \mathrm{D}(\delta^*{}_1, ..., \delta_{*k})$
$\phi_k \sim \mathrm{G}(a/2, b/2)$	$\phi_k \left| \mathrm{x, z}\right. \sim \mathrm{G}(a^*/2, b^*/2)$
$\mu_k \left| \phi_k\right. \sim \mathrm{N}\left(\mathrm{m_k}, \frac{1}{\alpha_i \phi_k}\right)$	$\mu_i \left| x, z, \phi_k\right. \sim \mathrm{N}\left(\mathrm{m}^*{}_k, \frac{1}{\alpha^*{}_i \phi_k}\right)$

Yukarıda tanımlanan conjugate ön (prior) dağılımlar tablosunda;

$\delta^{*}{}_{1} = \delta_{k} + n_{k}$ ve $a^{*}{}_{k} = a + n_{k}$

$b^{*}{}_{k} = b + \sum_{j:z_j=k} (x_j - \mu_k)^2$ ve $\alpha^{*}{}_{k} = \alpha_k + n_k$

$$m^{*}{}_{k} = \frac{\alpha_k m_k + n_k \overline{x}_k}{\alpha_k + n_k} \qquad \overline{x}_k = \frac{1}{n_k} \sum_{j:z_j=k} x_j$$

$\mu_1 < ... < \mu_k$ varsayıldığında, $D(\delta_1,...\delta_k)$ bir Drichlet dağılımı göstermekte olup aşağıdaki gibi yazılabilir.

$$f(\pi_1,\ldots,\pi_k) \alpha \prod_{k=1}^{K} \pi_k^{\delta_k - 1}$$

$(\delta_1,\ldots\delta_k)$ için prior seçimi için genellikle düzgün (uniform) dağılımı esas alınmaktadır. Bunun için başlangıç değerleri olarak 1 sabit değeri,

$$(\delta_1,...\delta_k) = (1,...1)$$

biçiminde atanmaktadır. Karışma olasılığı olan π_k için prior seçimi aşağıdaki gibi yapılmaktadır.

$$\pi_1 = \eta_1$$

$$\pi_k = (1-\eta_1)\ldots(1-\eta_{k-1})\eta_k$$

$$\eta_k \sim \mathrm{B}(1, K-k+1)$$

Karışımlı Bayesian analizi için MCMC algoritmasının aşamaları aşağıdaki gibi yazılabilir.

1. $\eta^{(0)}, \mu^{(0)}$ and $\phi^{(0)}$ başlangıç değerleri ile parametre tahmini başlanır.
2. $z^{(j+1)} \sim z \left| x, \rho^{(j)}, \mu^{(j)}, \phi^{(j)} \right.$ kullanılarak yeni Z değeri elde edilir.
3. $\eta^{(j+1)} \sim \eta \left| x, z^{(j+1)} \right.$ kullanılarak yeni η değeri elde edilir.
4. $\phi_k^{(j+1)} \sim \phi_k \left| x, z^{(j+1)} \right.$ kullanılarak yeni ϕ_k değeri elde edilir.
5. $\mu_k^{(j+1)} \sim \mu_k \left| x, z^{(j+1)}, \phi_k^{(j+1)} \right.$ kullanılarak yeni μ_k değeri elde edilir.
6. $\mu^{(j+1)}$'nin sıralanması ve $\rho^{(j+1)} y \phi^{(j+1)}$, nin yeniden düzenlenmesi aşaması
7. $j = j+1$ ve 2'inci aşamaya gidilmesi. Uygun parametre değerleri elde edilinceye kadar 2-7 aşamaları tekrarlanır (Marin ve ark., 2009).

6.2. Faktör Analizi İçin Karışımlı Model Yaklaşımı

Yorumlanması güç, birbiri ile ilişkili çok sayıda değişkenden, en az bilgi kaybı ile bağımsız ve kavramsal açıdan anlamlı az sayıda yeni değişken bulmayı amaçlayan birçok değişkenli analiz yöntemidir. Faktör analizinin iki temel amacı vardır.

a. Değişken sayısını azaltmak (boyut indirgemek)

b. Değişkenler arasındaki ilişkilerin yapısını araştırmak, yani değişkenleri kavramsal olarak sınıflamaktır (Alpar, 2012).

$Y_1,\ldots,Y_n$ p boyutlu ve n büyüklüğünde bir rasgele vektör olsun. Her bir $\boldsymbol{Y}_j$ gözlem değeri için faktör analizi modeli aşağıdaki gibi yazılabilir.

$$Y_j = \boldsymbol{\mu} + BU_j + e_j,\ j = 1,\ldots,\mathrm{n}) \tag{6.5}$$

Eşitlik 6.5' de U_j bir q boyutlu latent ya da gözlenemeyen değişkenler vektörü olarak adlandırılmakta olup, $N(0,\mathrm{I}_q)$ ile normal dağılım göstermektedir. B, pxq boyutlu faktör yükü matrisi olarak adlandırılmaktadır. e_j hata terimi ise $N\left(0,\mathrm{D}\right)$ ile normal dağılım göstermektedir. D varyansı,

$$D = diag(\sigma_1^2,\ldots,\sigma_p^2)$$

biçiminde hesaplanmaktadır.

Karışımlı Faktör analizi, normal karışımlı bir modelin yüksek boyutlu verilere uymasını sağlar. Genel bir doğrusal olmayan yaklaşım, gözlemlenemeyen faktörler olan $\boldsymbol{u}_j$ verildiğinde, tüm gözlem vektörünün dağılımı $\boldsymbol{Y}_j$ için doğrusal alt modellerin sınırlı bir karışımını varsayarak elde edilebilir. Yani, gözlemin $\boldsymbol{Y}_j$ dağılımının aşağıdaki gibi karışımlı model kullanılarak $\pi_k(\mathrm{k} = 1,\ldots,\mathrm{K})$ olasılıkları ile modellenebileceğini varsayarak bir yerel boyut azaltma yöntemi sağlayabiliriz.

$$Y_j = \boldsymbol{\mu}_k + B_k U_{kj} + e_{kj} \tag{6.6}$$

$j = 1,...,n$ gözlem değerleri için, faktörler $U_{k1},\ldots,U_{kn}$, birbirinden bağımsız $N(0, \mathrm{I}_q)$ ile normal dağılım, e_{ij}' birbirinden bağımsız olarak $N(0, D_k)$ ile normal dağılım göstermektedir. D_k, bir diyagonal matristir $(\mathrm{k}=1,\ldots,\mathrm{K})$. Karışımlı oranları π_k olup toplamları 1 değerine eşittir. Böylece, koşulsuz olarak, her bir Y_j gözlemin yoğunluğu; $\pi_1,\ldots,\pi_K$ karışma oranları ile K normal yoğunluğunun karışımı şeklinde aşağıdaki gibi yazılabilir.

$$f(y_j;\psi) = \sum_{k=1}^{K} \pi_k \phi(y_j;\mu_k,\Sigma_k) \tag{6.7}$$

Eşitlik 6.7'de, Σ_k her bir alt grup için kovaryans matrisi olup aşağıdaki gibi yazılır.

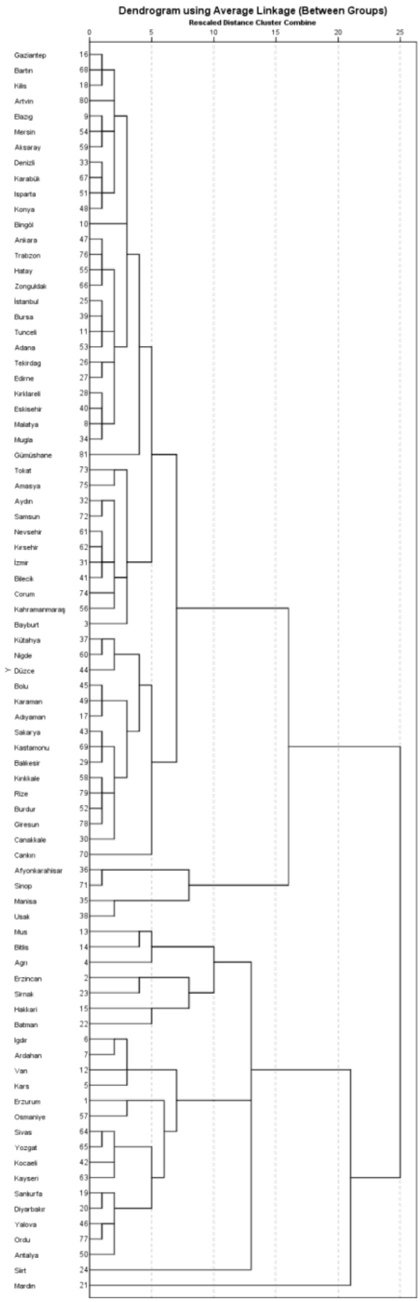
$$\Sigma_k = B_k B_k^T + D_k \quad (\mathrm{k}=1,\ldots,\mathrm{K})$$

ψ ; $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$ 'yı yerine koyarak, $\pi_k (\mathrm{k}=1,\ldots,\mathrm{K}-1)$ karışımlı oranlarıyla birlikte μ_k, $\mathbf{B}_k$ ve D_k parametrelerini içeren parametre vektörüdür. Eşitlik 6.7' i kullanılarak, özellikle sınırlı verilerin bulunduğu durumlarda, normal alt grupların karışımlılarıyla yüksek boyutlu verilerin modellenmesinde de yararlıdır. Kısıtlanmamış Σ_i kovaryans matrisleri olan normal alt gruplardan oluşan bir karışımın uyumu, her bir $\Sigma_k (\mathrm{k}=1,\ldots,\mathrm{K})$ için $\frac{1}{2} p(p+1)$ parametresi bulunur.

Bu, karışımlı modelindeki k alt grupların sayısı arttıkça, toplam parametre sayısı, n örnek boyutuna göre çok hızlı bir şekilde artabilir ve bu da aşırı uyuma neden olur (McLachlan ve Peel, 2000).

Karışımlı faktör analizinde model uyumu (parametre tahminleri) AECM (expectation-conditional maximization) algoritması kullanılarak yapılmaktadır. AECM algoritması Meng ve Dyk (1997) tarafından geliştirilmiş olup karışımlı faktör analizinde kullanılmaya başlanmıştır. AECM için log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\log L\left(\psi\right) = \sum_{J=1}^{N} \log\left\{\sum_{k=1}^{K} \pi_k \phi(y_j; \mu_k, \Sigma_k)\right\}$$



Şekil 6.1. Multululuk değişkenlerine göre kümeleme analizi sonucu illerin dağılımı.

6.3. Normal karışımlı model sonucunda elde edilen alt gruplara ilişkin varyans tahmin değerleri

Çizelge 6.1. Alt grup 1 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

Çizelge 6.2. Alt grup 2 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

Çizelge 6.3. Alt grup 3 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

Çizelge 6.4. Alt grup 4 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

Çizelge 6.5. Alt grup 5 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

Çizelge 6.6. Alt grup 6 için elde edilen varyans değerleri

Değişkenler	Tahmin	Standart hata (S.E)	Tahmin/S.E	P-değeri
Çocuklar	6.640	0.979	6.784	0.000
Eş	2.285	0.415	5.503	0.000
Anne/Baba	0.930	0.275	3.384	0.001
Kendisi	0.509	0.089	5.714	0.000
Torunlar	0.291	0.076	3.805	0.000
Sağlık	7.880	1.825	4.317	0.000
Sevgi	4.736	1.445	3.278	0.001
Başarı	2.097	0.560	3.743	0.000
Para	0.870	0.167	5.199	0.000
İş	0.836	0.217	3.847	0.000

6.4. Normal karışımlı model sonucunda elde edilen değişkenlere ilişkin kovaryans matrisi değerleri

Çizelge 6.7. Değişkenler arasındaki kovaryans matrisi

	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	7.551									
Eş	-0.566	8.651								
Anne/Baba	-0.163	2.447	2.321							
Kendisi	0.239	0.095	0.087	0.701						
Torunlar	0.725	-0.795	-0.582	0.024	0.538					
Sağlık	0.221	-12.205	-6.664	-1.639	1.924	49.596				
Sevgi	-1.201	8.026	2.985	0.464	-1.244	-25.592	18.753			
Başarı	1.084	0.609	0.915	0.609	-0.017	-9.577	2.924	4.456		
Para	0.426	0.798	0.323	0.341	0.091	-4.005	0.968	0.998	1.459	
İş	-0.128	1.717	1.481	0.199	-0.475	-6.382	2.119	0.795	0.536	2.392

6.5. Normal karışımlı model sonucunda elde edilen alt değişkenlere ilişkin korelasyon matrisi değerleri

Çizelge 6.8. Değişkenler arasındaki korelasyon matrisi

	Model Estimated Correlations									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	1.000									
Eş	-0.070	1.000								
Anne/Baba	-0.039	0.546	1.000							
Kendisi	0.104	0.038	0.069	1.000						
Torunlar	0.359	-0.368	-0.521	0.038	1.000					
Sağlık	0.011	-0.589	-0.621	-0.278	0.372	1.000				
Sevgi	-0.101	0.630	0.452	0.128	-0.392	-0.839	1.000			
Başarı	0.187	0.098	0.285	0.344	-0.011	-0.644	0.320	1.000		
Para	0.128	0.225	0.176	0.338	0.103	-0.471	0.185	0.391	1.000	
İş	-0.030	0.377	0.628	0.154	-0.418	-0.586	0.316	0.243	0.287	1.000

6.6. Normal karışımlı model sonucunda elde edilen alt gruplara ilişkin kovaryans matrisi değerleri

Çizelge 6.9. Alt grup 1 için elde edilen kovaryans matris değerleri

Class 1	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

Çizelge 6.10. Alt grup 2 için elde edilen kovaryans matris değerleri

Class 2	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

Çizelge 6.11. Alt grup 3 için elde edilen kovaryans matris değerleri

Class 3	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

Çizelge 6.12. Alt grup 4 için elde edilen kovaryans matris değerleri

Class 4	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

Çizelge 6.13. Alt grup 5 için elde edilen kovaryans matris değerleri

Class 5	Model Estimated covariance									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

Çizelge 6.14. Alt grup 6 için elde edilen kovaryans matris değerleri

Class 6	Model Estimated covariances									
	Çocuklar	Eş	Anne/Baba	Kendisi	Torunlar	Sağlık	Sevgi	Başarı	Para	İş
Çocuklar	6.640									
Eş	0.000	2.285								
Anne/Baba	0.000	0.000	0.930							
Kendisi	0.000	0.000	0.000	0.509						
Torunlar	0.000	0.000	0.000	0.000	0.291					
Sağlık	0.000	0.000	0.000	0.000	0.000	7.880				
Sevgi	0.000	0.000	0.000	0.000	0.000	0.000	4.736			
Başarı	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.097		
Para	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.870	
İş	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.836

ÖZ GEÇMİŞ

1980 yılında Van' da doğdu. İlk, orta ve lise öğrenimini Van' da tamamladıktan sonra 2000 yılında Van Yüzüncü Yıl Üniversitesi Fen – Edebiyat Fakültesi Matematik Bölümü 'nü kazandı. 2004 yılında bu bölümden mezun oldu. 2004 yılında Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalında yüksek lisans öğrenimine başladı. 2006 yılında yüksek lisansı bitirdi. 2011 yılında Formasyon öğrenimini tamamladı. 2004 ve 2013 yılları arasında çeşitli dershane ve özel okullarda görev yaptı. 2013 yılında Hakkari Üniversitesi Çölemerik MYO' da öğretim görevlisi olarak göreve başladı. 2014 yılında Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zootekni/ Biyometri ve Genetik Anabilim Dalında doktora öğrenimine başladı.

İKSAD
Publishing House

ISBN: 978-625-378-062-3