

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOOTEKNİ ANABİLİM DALI

**NORMAL DAĞILIM GÖSTEREN VERİLERİN ANALİZİNDE GAUSSIAN
KARIŞIMLI REGRESYONUN KULLANILMASI**

DOKTORA TEZİ

HAZIRLAYAN: Serdar ABUT
DANIŞMAN: Prof. Dr. Abdullah YEŞİLOVA

VAN – 2019

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOOTEKNİ ANABİLİM DALI

**NORMAL DAĞILIM GÖSTEREN VERİLERİN ANALİZİNDE GAUSSIAN
KARIŞIMLI REGRESYONUN KULLANILMASI**

DOKTORA TEZİ

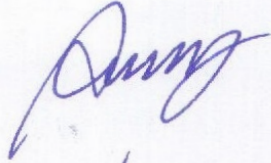
HAZIRLAYAN: Serdar ABUT


Bu çalışma YYÜ Bilimsel Araştırma Projeleri Başkanlığı tarafından FDK-2018-6791
No'lu proje olarak desteklenmiştir

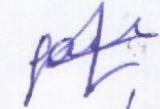
VAN-2019

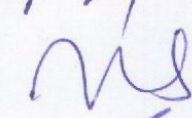
KABUL VE ONAY SAYFASI

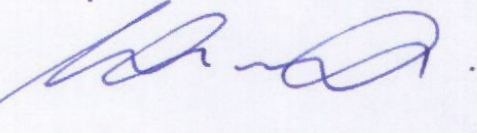
Zootekni Anabilim Dalı'nda Prof. Dr. Abdullah YEŞİLOVA danışmanlığında, SERDAR ABUT tarafından sunulan "Normal Dağılım Gösteren Verilerin Analizinde Gaussian Karışımli Modelin Kullanılması" isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 30.01.2019 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile başarılı bulunmuş ve doktora tezi olarak kabul edilmiştir.

Prof. Dr. Abdullah YEŞİLOVA
Başkan:..... İmza: 

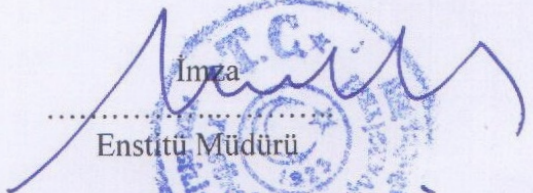
Doç. Dr. Hamit MIRTAGIOĞLU
Üye:..... İmza: 

Doç. Dr. Gazel SER
Üye:..... İmza: 

Dr. Öğr. Üyesi Yılmaz KAYA
Üye:..... İmza: 

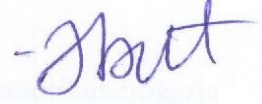
Dr. Öğr. Üyesi. Gürül ZIRHLIOĞLU
Üye:..... İmza: 

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 15.02.2019 tarih ve 2019/14-I sayılı kararı ile onaylanmıştır.

İmza: 
Enstitü Müdürü
Prof. Dr. Bülent SENSÖY
Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.



(İmza)

Serdar ABUT



ÖZET

NORMAL DAĞILIM GÖSTEREN VERİLERİN ANALİZİNDE GAUSSIAN KARIŞIMLI REGRESYONUN KULLANILMASI

ABUT, Serdar
Doktora Tezi, Zootekni Anabilim Dalı
Tez Danışmanı : Prof. Dr. Abdullah YEŞİLOVA
OCAK 2019, 99 sayfa

Gaussian karışimli regresyon modelleri, sınırlı bir latent sınıf kümesi aracılığıyla populasyon heterojenliğini tahmin edebilir. Bu modeller özellikle bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkisindeki farklılıkların tanımlanmasına odaklanmaktadır. Bu tez çalışmasında, solunum fonksiyon testinin (SFT) bir sonucu olan FEV1/FVC gibi bir bağımlı değişken üzerinde karışimli regresyon modeli kurulmuştur. FEV1/FVC, özellikle akciğerin obstrüktif ve restriktif hastalıklarının ayırımında kullanılan en önemli SFT parametresidir. Bağımsız değişkenler olarak, yaş, cinsiyet, boy ve kilo seçilmiştir. AIC ve BIC değerlerine bakılarak, veri setinin 5 alt grubun karışımından oluştuğu saptanmıştır. Bu alt gruplardaki FEV1/FVC ortalama değerleri, ilk alt grupta en düşük, beşinci alt grupta en yüksek olacak şekilde küçükten büyüğe doğru sıralandığı görülmüştür. İlk alt gruptan son alt gruba doğru düzenli şekilde büyüyen FEV1/FVC değerlerine bakılarak, obstrüktif patternden restriktif patterne doğru düzenli ve anlamlı bir geçiş gözlenmiştir. Birinci alt grubun, kilo değişkeni açısından en yüksek, fakat FEV1/FVC değişkeni açısından en düşük değerlere sahip olduğu bulunmuştur. Dolayısıyla, obstrüktif patternin en belirgin olduğu alt grup birinci alt gruptur. Bu sonuç, KOAH ile obezite arasındaki ilişkinin bir göstergesidir. Buna karşın, beşinci alt grubun, kilo değişkeni açısından en düşük, fakat FEV1/FVC değişkeni açısından en yüksek değerlere sahip olması, kilo bağımsız değişkeninin, SFT sonuçlarında oldukça önemli bir rol oynadığını göstermektedir. Bu tez çalışması, heterojen bir yapıya sahip olan verilerin analizinde, karışimli regresyon modellerinin kullanılmasının veriyi anlamlandırmada etkili bir yol olduğunu göstermiştir.

Anahtar kelimeler: Akciğer hastalıkları, Alt grup, Karışimli regresyon, Parametre tahmini, Solunum fonksiyon testi.

ABSTRACT

USE OF GAUSSIAN MIXED REGRESSION FOR ANALYSIS OF NORMALLY DISTRIBUTED DATA

ABUT, Serdar
Ph. D. Thesis, Animal Science
Supervisor : Prof. Dr. Abdullah YESILOVA
January 2019, 99 pages

Gaussian mixture regression models could estimate population heterogeneity through a finite set of latent classes. These models specifically focus on the identification of differences in the effect of independent variables on a dependent variable. In this thesis, a mixed regression model was established on a dependent variable such as FEV1 / FVC which is a result of pulmonary function test (PFT). FEV1 / FVC is the most important PFT parameter used to differentiate obstructive and restrictive diseases of the lung. Age, gender, height and weight were chosen as independent variables. Based on the AIC and BIC values, it was determined that the dataset was a mixture of 5 subgroups. The mean values of FEV1 / FVC in these subgroups were found to be the smallest in the first subgroup and the highest in the fifth subgroup. A regular and meaningful transition from obstructive pattern to restrictive pattern was observed by looking at FEV1 / FVC values growing from the first subgroup to the last subgroup. The first subgroup was found to be the highest in terms of weight variable but the lowest for the FEV1 / FVC variable. Therefore, the subgroup with the most obstructive pattern is the first subgroup. This result is an indication of the relationship between COPD and obesity. On the other hand, the fifth subgroup was the lowest in terms of weight, but the highest for the FEV1 / FVC variable, indicating that the weight independent variable plays an important role when it is filled in the results of the PFT. The results of this thesis showed that the use of mixture regression models in the analysis of data with a heterogeneous structure is an effective way to make sense of the data.

Keywords: Lung diseases, Subgroup, Mixture regression, Parameter estimation, Pulmonary function test.



ÖN SÖZ

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen danışmanım Sayın Prof. Dr. Abdullah YEŞİLOVA' ya teşekkür ederim. Ayrıca tez izleme komitemde yer alan Doç. Dr. Gazel SER ve Dr. Öğr. Üyesi Yılmaz KAYA'ya teşekkürlerimi sunarım.

Ayrıca, istatistiksel analiz sonuçlarının tıp literatürdeki karşılığının belirlenmesi ve yorumlanmasında değerli katkılarından dolayı Van YYÜ Dursun Odabaşı Tıp Merkezi'nde görev yapan Acil Tıp Uzmanı Dr. Fatih DOĞANAY'a teşekkürlerimi sunarım.

Çalışmalarımın her aşamasında desteğini hissettiğim fedakar eşim Kıymet Minas ABUT'a ve bu çalışma nedeniyle kendisine hak ettiği zamanı ayıramadığım oğlum Ali Musa ABUT'a şükranlarımı sunarım.

2019

Serdar ABUT



İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	iii
ÖN SÖZ	v
İÇİNDEKİLER.....	vii
ÇİZELGELER LİSTESİ.....	ix
ŞEKİLLER LİSTESİ.....	xi
SİMGELER VE KISALTMALAR.....	xiii
EKLER DİZİNİ.....	xv
1. GİRİŞ	1
2. KAYNAK BİLDİRİŞLERİ.....	5
3. MATERYAL VE YÖNTEM.....	13
3.1. Materyal.....	13
3.2. Yöntem	13
3.2.1 Tek değişkenli normal dağılım	13
3.2.2 Çok değişkenli normal dağılım.....	16
3.2.3 Karışımli modellerin genel yapısı.....	20
3.2.4 Karışımli modellerin açıklanması.....	22
3.2.5 Eş varyanslı iki normal dağılımın karışımı.....	23
3.2.6 Çok değişkenli karışımli normal dağılım	25
3.3. Veri Asimetrikliğinin Modellenmesi.....	29
3.4. Karışımli Modelin Parametrik Gösterimi	30
3.5. Karışımli Modellerde Parametre Tahmini	31
3.6. Karışımli Modelde Alt Grup Sayısının Test Edilmesi.....	32
3.7. Karışımli Modellerde En Çok Olabilirlik Yöntemi	33
3.7.1 Karışımli modellerde en çok olabilirlik.....	33
3.7.2. Bilgi matrisleri	33
3.7.3. Maksimum olabilirlik tahmininin asimptotik kovaryans matrisi....	34

	Sayfa
3.7.4. Karışımli modeller için EM algoritması	34
3.7.4.1. Doğrudan yaklaşım	34
3.7.4.2. E- aşaması	35
3.7.4.3. M- aşaması	37
3.8. Model Seçiminde Uyum Ölçütleri	37
3.8.1. Log olabilirliğin sapması	38
3.8.2. Akaike bilgi ölçütü	40
3.8.3. Bilgi karmaşıklığı kriteri	41
3.9. Bayesian Tabanlı Uyum Ölçütleri	42
3.9.1. Bayesian yaklaşım	42
3.9.2. Bayesian bilgi kriteri	42
4. BULGULAR	43
5. TARTIŞMA VE SONUÇ	55
KAYNAKLAR	63
EKLER	69
ÖZ GEÇMİŞ	99

ÇİZELGELER LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Normal karışimli yoğunluğun on beş örneği için parametreler	26
Çizelge 4.1. Solunum fonksiyon testi sonuçlarına göre homojen alt gruplar için model bilgi ölçütleri	45
Çizelge 4.2. En yakın sınıf üyelik derecelerine sahip alt gruplar için elde edilen sınıf büyüklükleri ve karışma olasılıkları.....	46
Çizelge 4.3. Alt gruplara ait parametre tahminleri ve standart hataları.....	51
Çizelge 4.4. Değişkenlerin alt gruplara göre ortalamaları.....	52



ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 2.1. 1000 adet karidesin başın gövde uzunluğuna oranlarının histogramı	5
Şekil 3.1. Standart normal yoğunluk grafiği.....	15
Şekil 3.2. Farklı ρ korelasyonlara göre iki değişkenli (bivariate) normal dağılım gösteren X_1 ve X_2 değişkenlerinin olasılık yoğunluğu grafikleri.....	19
Şekil 3.3. İki değişkenli normal dağılımda eksponansiyel terimin elips şeklindeki grafiği.....	20
Şekil 3.4. Varyansları eşit ve σ olan iki adet ortalamaları μ_1 ve μ_2 olan tek tepeli normal bileşenlerin eşit miktarlardaki karışımları	24
Şekil 3.5. Varyansları eşit ve σ olan iki adet ortalamaları μ_1 ve μ_2 olan tek tepeli normal bileşenlerin 0,75 ve 0,25 oranlarındaki karışımları	24
Şekil 3.6. Karışım normal dağılımların grafikleri	27
Şekil 3.7. Karışım normal dağılımların grafikleri (devam).....	28
Şekil 3.8. Lognormal ve iki bileşenli normal karışımın yoğunluk grafiği	29
Şekil 3.9. Karışım dağılımları konusundaki akademik yazıların yıllara göre değişimi.....	32
Şekil 3.10. Rasgele başlangıç değerleri kullanılarak EM algoritması için farklı k değerleri için üretilen parametrelere dayalı her alt grup için asimptotik (%95) elipsoidler grafiği.	38
Şekil 4.1. Boy ve kilo bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşı değişim grafiği	43
Şekil 4.2. Yaş ve boy bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşı değişim grafiği	44
Şekil 4.3. Yaş ve kilo bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşı değişim grafiği	44
Şekil 4.4. SFT sonuçlarına göre farklı alt grup sayılarına göre AIC ve BIC değerleri.	46
Şekil 4.5. Veri setindeki değişkenler arası korelasyon matrisi grafiği.	47
Şekil 4.6. Birinci alt grupta değişkenler arası korelasyon matrisi.....	48

Şekil	Sayfa
Şekil 4.7. İkinci alt grupta değişkenler arası korelasyon matrisi.....	48
Şekil 4.8. Üçüncü alt grupta değişkenler arası korelasyon matrisi.....	49
Şekil 4.9. Dördüncü alt grupta değişkenler arası korelasyon matrisi	49
Şekil 4.10. Beşinci alt grupta değişkenler arası korelasyon matrisi	50
Şekil 4.11. Bağımlı değişkenin beş alt gruba göre dağılım grafiği	52
Şekil 4.12. Gaussian karışimli regresyona göre elde edilen 5 alt grubun olasılık dağılımı yoğunluklarının dağılımı.....	52



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
μ	Örneklem ortalaması
σ^2	Örneklem varyansı
$n(\mu, \sigma^2)$	Normal Dağılım
$f(x \mu, \sigma^2)$	Normal Dağılımın olasılık yoğunluk fonksiyonu
$P(Z \leq z)$	Koşullu Standart normal dağılım fonksiyonu
EZ	Standart normal dağılımın beklenen değeri
π_g	g adet altgrupun karışım oranları
$f_i(y_j)$	Karışımındaki altgrup yoğunlukları
Z_j	Altgrup etiket değerlerini alan kategorik değişken
$f_i(y_i; \theta_i)$	Karışımındaki altgrup yoğunlukları
ψ	Karışımındaki bilinmeyen tüm parametreleri içeren vektör.
Ω	ψ için belirlenmiş parametre uzayı

Kısaltmalar	Açıklama
GKR	Gaussian Karışımlı Regresyon
FMM	Finite Mixture Model
AIC	Akaike Bilgi Ölçütü (Akaike's Information Criteria)
BIC	Bayesian Bilgi Ölçütü (Bayesian Information Criteria)
EM	Expectation-Maximization
ML	Ençok Olabilirlik – Maximum Likelihood
SFT	Solunum Fonksiyon Testi
FVC	Zorlu Vital Kapasite – Force Vital Capacity
FEV	Zorlu ekspiratuvar hacim - Forced Expiratory Volume



EKLER DİZİNİ

EK	Sayfa
EK 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş deęişkenlerine göre oluşturulan modelde, her bir örneęin her alt gruba girme olasılıkları ve atandığı alt gruplar.....	69
EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları.....	74





1. GİRİŞ

İncelenen özelliklerin çok sayıda olması, sorunun kümeleme ve faktör analizi gibi klasik istatistiksel tekniklerle çözümüne olanak tanımamaktadır. Bilimsel çalışmalarda ele alınan olaylar genellikle birçok etkinin altındadır. Bununla birlikte, incelenen değişkenlerin çoğu da birbiri ile ilişkilidir. Bu nedenle uygulamalarda, çok fazla sayıda değişkenle karşılaşmaktadır. Yapılan çalışmalarda geçerli ve güvenilir sonuçların elde edilebilmesi için araştırma kapsamındaki bütün özelliklerin değerlendirmeye alınması gerekmektedir. Söz konusu bu gereksinimden dolayı çok değişkenli veri yapısı ile karşılaşmaktadır. Çok değişkenli istatistiksel yöntemlerde amaç, yorumlanması güç, birbiri ile ilişkili çok sayıda değişkenden, en az bilgi kaybı ile bağımsız ve kavramsal açıdan anlamlı az sayıda yeni değişken yapısını ortaya koymaktır. 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. Bununla birlikte son yıllarda karışımli model yaklaşımı çok değişkenli istatistikte yoğun olarak kullanılmaktadır (Wang ve Putterman, 1998). Karışımli modelleme (mixture model); 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 dahil olabileceğinin hesaplanmasıdır. Karışımli modelin ikinci önemli avantajı ise, her bir alt grup için parametre tahminlerinin elde edilmesidir. Çünkü karışımli modelleme, veri setini homojen alt gruplara ayırdıktan sonra her bir alt grup için oluşturulan bağımlı değişken ile bağımsız değişkenlere göre çoklu regresyon analizi de yapmaktadır. Bağımlı değişkenin normal dağılım göstermesi durumunda, Gaussian karışımli regresyon (Gaussian mixture regression) analizi kullanılmaktadır (Yeşilova ve ark., 2016). Karışımli modelin diğer bir amacı da hatalı sınıflandırma olasılığını en aza indirgeyerek, bireyleri ait oldukları gruplara ayırmak ve çekilmiş oldukları kitleleri belirlemektir. Karışımli modellemede, parametre tahminleri expectation and maximization (EM) algoritması kullanılarak en çok olabilirlik (Maximum Likelihood=ML) yöntemi ile elde edilmektedir (Wang ve Putterman, 1998). Karışımli modellemede, doğru sınıflandırma olasılığı için entropy istatistiği kullanılmaktadır.

Uygun alt grup sayısını belirlemede ise Akaiki ve Bayesian uyum ölçütleri yaygın olarak kullanılmaktadır.

Karışımli modeller, çok sayıda deęişkenin olduęu veri setlerinde, istatistiksel analiz yapılabilmesi için farklı matematiksel yaklaşımlar sunmaktadır. Bu tarz modeller, aşırı esnek yapılarından dolayı, hem teorik hem pratik çalışmalarda ilgi çeken bir konu olmuştur. Karışımli model yaklaşımlarının uygulama potansiyeli artıkça, bu konu ile ilgili çalışmalar yaygınlaşmaktadır. Astronomi, biyoloji, genetik, tıp, psikiyatri, ekonomi, mühendislik, borsa gibi fen ve sosyal bilimler alanlarındaki uygulamalarda, karışımli modeller etkin olarak kullanılmaktadır. Bu uygulamalarda karışımli modeller; kümeleme analizi ve latent sınıf analizi, ayırma analizi, imge analizi, sağ kalım (*survival*) analizi gibi önemli istatistiksel tekniklere destek oluşturmaktadır (McLachlan ve Peel, 2000).

Son yıllarda karışımli modellere olan ilgi, heterojen veri setleri ile gerçekleştirilen çalışmalarla artış göstermiştir. Veri üzerinde grup yapısı mevcut olduęu veya veriyi kümeleme analizindeki gibi alt gruplara ayırma amaçlandığı uygulamalar, karışımli modellerin bilinen ve yaygın olarak kullanıldığı alanlardır. Buna ek olarak, karışımli modeller, esnek yapıda olmaları sayesinde, dağılım şekli bilinmeyen veriler üzerinde model kurmak için kullanılabilecek elverişli araçtır.

Karışımli dağılımların en belirgin faydalarından biri kümeleme analizinde heterojenliğin modellenmesine olanak sağlanmasıdır. Eğer bir grup yapısı var ise, karışımli modellerin hata oranlarını değerlendirilmesinde (duyarlılık/özgünlük – *sensitivity/specificity*) faydalı bir rolü vardır. Herhangi bir sürekli dağılımın, ortak varyanslı (çok deęişkenli olması durumunda kovaryans matrisi) normal yoğunlukların sonlu bir karışımı ile isteęe baęlı olarak yaklaşımı mümkün olduğundan dolayı, karışımli modeller, bilinmeyen dağılım şekillerini modellemek için uygun yarı-parametrik bir çerçeve sağlar. Örneğin Priebe (1994) karışımli model yaklaşımı ve kernel tahminine dayalı bir tahmin teknięi olarak sunduęu, adaptif karışımli adlı çalışmasında, $n = 10.000$ gözlem ile, lognormal yoğunluğunun, yaklaşık 30 adet normalin karışımına iyi derecede yakınsadığını bulmuştur. Bundan farklı olarak, bir kernel yoğunluk tahmincisi 10 000 normalin karışımını kullanır. Bir karışımli model, bileşenlerin uygun seçilmesi kaydıyla oldukça karmaşık dağılımları modelleyebilir. Böylece, tek bir parametrik ailenin gözlemlenen verilerdeki yerel varyasyonlar için tatmin edici bir model sağlayamadığı durumlarda karışımli modeller faydalı olabilir. Karışım bileşenlerine bakarak modellenen olgu hakkında kolayca çıkarımlar yapılabilir

(McLachlan ve Peel, 2000). Karışımli modeller bu esnek yapısı ile yapay sinir ađlarında faydalı bir rol üstlenir. Radyal temelli fonksiyonları kullanarak oluşturulan sinir ađı modelleri için, giriş verisi karışımli model ile modellenir. Yani, temel fonksiyonlar, giriş verilerinden en çok olabilirlik yöntemiyle tahmin edilmiş karışımli modelin bileşenleri olarak alınabilir (McLachlan ve Peel, 2000).

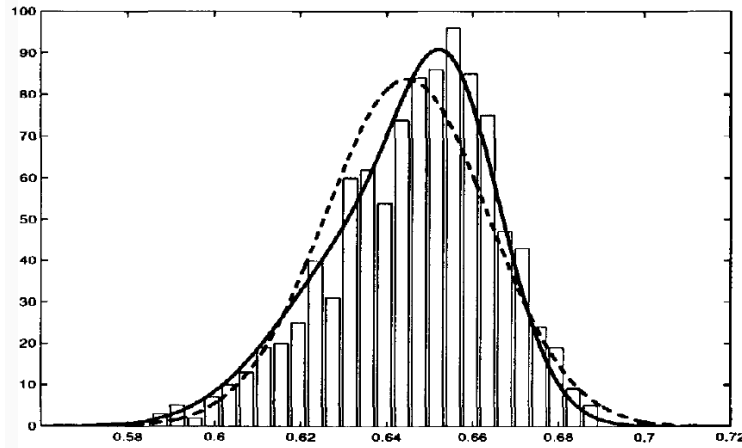
Karışımli modellemede amaç, veri setinin heterojen bir yapı gösterdiği varsayılır. Veri setinin kendi içerisinde homojen alt gruplara ayrılarak söz konusu heterojenlik giderilmektedir. Böylece her bir alt grup içi homojenlik sağlanırken, alt gruplar arası heterojenliğin belirlenmektedir. Karışımli model yaklaşımı bu yönüyle kümeleme ve faktör analizlerine benzemektedir. Ancak karışımli modelin, kümeleme ve faktör analizlerine göre iki önemli avantajı bulunmaktadır (Muthén and Muthén, 2017). Bunlardan ilki, her bir gözlemin (bireyin) hangi olasılıkla hangi alt gruba dahil olduğu hesaplanmasıdır. İkinci avantajı ise her bir alt grup için parametre tahminlerinin elde edilmesidir (Liesch, 2004). Karışımli modellemede veri setinin yapısı ile ilgili herhangi bir sınırlama yoktur. Bağımlı değişkenin normal dağılışı göstermesi durumunda Gaussian karışımli regresyon kullanılmaktadır.

Bu tez çalışmasında; normal dağılım gösteren bağımlı değişkene göre, bağımsız değişkenler esas alındığında, veri setin var olan heterojenliğin, Gaussian karışımli regresyon kullanılarak, homojen kaç alt grupta toplanacağını belirlenmiştir. Alt grup sayısı belirlendikten sonra, her bir alt grup için ayrı Gaussian karışımli regresyon uygulanmıştır. Gaussian karışımli regresyon ile bireylerin hangi alt gruba girdiğinin saptamasında kullanılan entropy ölçütü kullanılarak, hatalı sınıflandırma olasılığının tespit edilmiştir. Gaussian karışımli regresyonun veri setindeki hem heterojenliği tespit etmesi hem de elde edilen her bir alt grup için ayrı regresyon analizi yapmasından dolayı, diğer yöntemlere nazaran daha iyi olduğu saptanmıştır.



2. KAYNAK BİLDİRİŞLERİ

Karışımli modellerin kullanımı ile ilgili ilk önemli çalışma, ünlü biometrisyen Karl Pearson (1894) tarafından yapılmıştır. Pearson (1894) çalışmasında, başka bir çalışmadan (Weldon, 1893) elde edilmiş, iki normal olasılıklı yoğunluk fonksiyonunun karışımı için model uydurmuştur. Bu iki fonksiyonun ortalamaları μ_1 ve μ_2 , varyansları σ_1^2 ve σ_2^2 ve karışım oranları ise π_1 ve π_2 dir. Pearson (1894)'ün analiz ettiği veri seti, $n = 1000$ adet karidesin elde edilen baş uzunluklarının vücut uzunluklarına oranlarını içermektedir. Bu verilerin histogramı, veriye uyan normal dağılım eğrisi ile birlikte gösterilmiştir (Şekil 2.1). Weldon (1893), veriyapısındaki bu asimetriklikten kuşkulunmuş ve aslında bu ölçümlerin alındığı verisetinde bir heterojenlik olduğunu düşünmüştür. Bu konuyu meslektaşı Pearson ile paylaşmış ve Pearson tarafından heterojenliğin kaynağı araştırılmıştır. Pearson (1894)'a göre dağılımdaki asimetriklik verisetindeki karideslerin iki alttürünün mevcut olmasından kaynaklanmıştır. Çalışmasında, moment yaklaşımıyla dokuzuncu dereceden polinom çözümü ile beş parametreye sahip denklemleri çözerek normal heterokedastik karışımli modeli sunmuştur (McLachlan ve Peel, 2000).



Şekil 2.1. 1000 adet karidesin başın gövde uzunluğuna oranlarının histogramı, 1 bileşen (kesikli çizgi) ve 2 bileşen (düz çizgi) normal uyumu (McLachlan ve Peel, 2000).

Kanji G. K. (1985) tarafından uçak kontrol sistemlerinin etkinliği ve güvenilirliğini değerlendirmek ve eğitim pilotlarının bir rüzgar perdesiyle karşılaştıkları çalışmalarında onu doğru tanıma ve uygun tepki sağlanması amacıyla gerçekleştirdiği

çalışmasında, uçakların inişe geçme esnasında karşılaşılabilecek belirli form ve büyüklükteki rüzgar dağılımlarını karışıklı Gaussian yaklaşımını kullanarak modellemiştir.

Yeşilova ve ark. (2010) tarafından, karışım modellemede, Akaike'nin bilgi kriterleri (AIC), Bayesyen bilgi kriterleri (BIC) ve Entropy sınıflandırma kriterleri alt popülasyonların sayısını belirlemek için kullanılmıştır. Alt popülasyonların sayısı belirlendikten sonra, model, her bir gözlemin belirli bir alt popülasyona düşme olasılığını belirlemiştir. Bu çalışmada, ilçelerin meyve özelliklerine göre sınıflandırılması, sekiz ilçeden toplanan ceviz meyvelerine karışım modellemesi uygulanarak elde edilmiştir. AIC, BIC ve entropi kriterlerine göre, beş alt popülasyonlu bir model seçilmiştir. Bu nedenle, her bir ilçenin, incelenen ceviz meyve özelliklerine göre farklı bir popülasyon oluşturmadığı, ancak beş alt popülasyona dağıldığı belirlenmiştir. Dördüncü alt grubun, ceviz gelişimi için en çok arzulanan özelliklere sahip olduğu bulunmuş, ve bu popülasyonu oluşturan ürünlerin Adilcevaz ve Ahlat ilçelerinden elde edilen ürünler olduğu gözlenmiştir.

Yeşilova ve ark. (2016) tarafından, Gauss karışım regresyon modeli kullanılarak Anadolu mandası, ayrık ve sürekli çevresel etkilere göre sınıflandırılmıştır. Gauss karışım modeli olarak hem gruplar arasında hem de gruplar arasında ayrı ayrı regresyon analizi yapılmıştır. Veriler 1455 Anadolu manda laktasyon süt verim kayıtlarından, Türkiye'nin Bitlis ilindeki yedi farklı lokasyonundan elde edilen kayıtlardan elde edilmiştir. Ana yaşı, laktasyon süresi ve yerleri laktasyon süt verimi üzerine çevresel etkiler olarak kabul edilmiştir. Veri seti, laktasyon süt verimindeki çevresel etkilere bağlı olarak Gauss karışım regresyonunda AIC ve BIC açısından üç homojen alt gruba ayrılmıştır. Laktasyon süresi, ana yaşı ve köylerin etkisi laktasyon süt verimi için en yüksek ortalama olan alt grup 1'deki laktasyon süt verimi açısından istatistiksel olarak anlamlı bulunmuştur ($p < 0.01$). Çalışma sonucu, sonuçlar Gauss karışımının regresyonunun, hayvancılıkta çevresel etkileri göz önünde bulundurarak niceliksel özelliklerin sınıflandırılmasında önemli bir araç olduğunu göstermiştir.

Shi ve ark. (2005) tarafından, tekrarlı ölçümlerin olduğu, büyük veri setleri ile kurulacak modellerde ortaya çıkan iki farklı sorunu çalışmalarında çözmeye çalışmışlardır. Bu sorunlar farklı tekrarlar arasındaki sistematik heterojenlikten kaynaklanan ve modelin uygulanmasında yer alan bir kovaryans matrisini tersine çevirme gerekliliğinden doğmaktadır. Çalışmalarında, gerçek verilere uygulanan bir Gaussian karışıklı regresyon modeli oluşturmuşlardır.

Naidu ve Babu (2018) çalışmalarında, yüz, parmak ve ses gibi 3 farklı biyometrik özelliği kullanarak, biyometrik eşleştirme modeli sunmuşlardır. Daha önceki çalışmalara kıyasla (Babu ve Naidu, 2014, 2016, Balaka ve Surendra, 2017) daha yüksek güvenliğin sağlandığı bu modelde, farklı biyometrik özellikleri kombine ederek Gaussian Karışım Modeli yardımıyla olasılık yoğunluk fonksiyonları hesaplanmıştır. Farklı üç biyometrik veriden çıkarılan üç farklı öznelik seti, füzyonlanarak birleştirilmiştir. Çalışmada yüz, parmakizi ve ses gibi verilerin doğada normal dağıldığından bahsedilmiştir.

Trianto ve ark (2018) tarafından yapılan çalışmada, otomatik konuşma tanıma amacıyla, karışımli Gaussian model ve gizli markov modelini hibritleştirerek model önermişlerdir. Çalışmalarında, akustik modelleme üzerine bazı derin öğrenme ağ tekniklerini tartışmışlardır.

Bruneau ve ark. (2018) ise, nükleotid dizilimleri üzerinde kümeleme amacı ile geliştirdikleri python paketini sundukları çalışmalarında, veriseti hakkında önsel (apriori) bir bilgiye gerek kalmadan, sınıflandırma yapabilmek amacı ile Gaussian mixture modellerden faydalanmışlardır. Çalışmalarında en iyi alt grup sayısını belirlemek için genellikle çok tercih edilen, Akaiki bilgi kriteri ve Bayesyen bilgi kriterini kullanmışlardır.

Görüntü işleme alanında, görüntü segmentasyonu uygulamalarında sonlu karışım modelleri oldukça faydalı bir araçtır. Fakat, sonlu karışım modelleri bu uygulamalarda pikselleri bağımsız olarak ele alır ve bir pikselin komşulukları ile olan ilişkilerini hesaba katmaz. Bu sınırlılıklar, sonlu karışım modellerini, gürültüye karşı hassaslaştırarak gürültüden fazla etkilenmeleri sorununu doğurur. Zang ve ark. (2013) tarafından yapılan çalışmada, bir ortalama şablon yardımı ile sonlu karışım modelleri ile oluşturulacak araçların gürültüye karşı olan hassasiyetini giderip daha gürbüz bir araç oluşturmayı önermişlerdir. Çalışmalarında, sonlu karışım modellerinin, önsel ve koşullu olasılığın lineer bir kombinasyonu olduğu düşünülmüş, bu iki olasılık, ağırlıklandırılmış aritmetik ortalama şablonu ve ağırlıklandırılmış geometrik ortalama şablonu diye adlandırdıkları, iki ortalama şablonunda hesaplanmıştır. Bu sayede kurdukları modelin, bir pikselin koşullu olasılığının hesaplanmasında, yerel komşu piksellerinin koşullu olasılıkları da hesaba katılarak, gürültü faktörünün görüntü segmentasyonu açısından oluşturduğu problemlerden daha az etkilenen bir model olduğunu öne sürmüşlerdir.

Ji ve ark. (2017) ise, görüntü segmentasyonu amacıyla modifiye edilmiş gürbüz bir Gaussian Karışımli Model (GKM) önerisi sundukları çalışmalarında, Gaussian Karışım Modellerini gürültüye karşı daha sağlam hale getirmek amacıyla yeni bir mekânsal ağırlık faktörü oluşturmuşlardır. Bu faktörü, Zang ve ark. (2013) görüntü segmentasyonunda gürültüye karşı çözüm olarak önerdikleri yöntem ile oluşturmuşlardır.

Beklenti maksimizasyonu (EM) algoritması, Gaussian karışım modellerinin parametrelerini tahmin etmenin en çok kullanılan yollarından biridir. Bununla birlikte, EM algoritması, gerçek karışım bileşenlerinin (*mixture subgroups*) sayısını önceden bilmesi gerekir ve bu algoritmanın performansı başlangıç parametrelerine bağlıdır. Bu iki dezavantajın üstesinden gelmek için Peng ve ark. (2017), değişken bir adım öğrenme algoritması önermişlerdir. Sundukları algoritmanın ayırıcı özelliği, Gaussian karışım bileşenlerinin tam sayısını veya yaklaşık sayısını bulabilmek için değişken bir adım arama stratejisinin önerilmesidir. Bulunan sayı, bhattacharyya katsayısı ölçütünü en üst düzeye çıkararak elde edilmiş; bu da Gaussian karışımli modellerinin gözlemlenen verilere ne kadar yakın olduğunu göstermektedir. Çalışmalarında önerdikleri algoritma ile gerçekleştirilen deneylerde hem parametre öğrenmeyi hem de model seçiminin daha verimli hale getirildiğini savunmuşlardır.

Susan ve Sharma (2017), çalışmalarında herhangi bir kullanıcı girdisi gerektirmeyen ve aynı zamanda yüksek doğruluk sağlayan yeni, denetimsiz, otomatik bir doku kusuru tespiti önermişlerdir. Bu amaca ulaşmak için, doku örneklerinin yerel bölgeleri üzerinde Gaussian kazanımlı düzenlilik indeksleri hesaplanmıştır. Genelleştirilmiş Gaussian karışım modeli birçok bilgisayar görmesi ve görüntü tanıma problemleri için esnek ve uygun bir araç sağlar.

Aydemir (2009) tarafından yapılan çalışmada GMR yönteminin teorik temellerini açıklamaya çalışmıştır. Kovaryans matrisinin oluşturulma şekline göre ve grup ortalama vektörleri arasındaki farka göre doğruluk oranlarının değiştiği, grup gözlem sayılarına göre doğruluk oranlarının değişmediği, büyük gözlem sayılarında GMR'nin yüksek doğruluk oranları verdiğini belirtilmiştir. Aynı çalışmada, GMR'nin parametrik olmayan regresyon modellemede diğer yöntemlerin yerine kullanılabilirdiği önerilmiştir.

İşaret dili hareketlerinin bir insansı robot tarafından taklit yoluyla gerçekleşmesi ve öğrenilmesi alanındaki bir uygulamada, gauss karışımli modeller ile insan hareketleri ve işaret dili hareketlerinin önemli özellikleri çıkarılmıştır. Gaussian

karışımı regresyon kullanılarak hareket genel yörünge formları elde edilip, bu yörünge formları choregraph benzetim programı kullanılarak nao insansı robot üzerinde gerçekleştirilmiş ve insan hareketine yakın bir hareketlerin taklit edildiği belirtilmiştir (Kıvrak, 2013).

Yan ve ark. (2017) tarafından genel karışımı modeller üzerine yapılan çalışmada, genel küme sayısı ve karıştırma katsayıları için gauss karma modelleri için gradyan beklenti maksimizasyon algoritmasının yakınsaklık özelliklerini incelenmiştir. Karıştırma katsayılarına, gerçek merkezler arasındaki minimum ve maksimum eş aralıklı uzaklıklara, boyutsallık ve bileşenlerin sayısına bağlı olarak yakınsama oranını türettiklerini belirtmişlerdir.

Zivkovic (2004) ise, Gaussian karışımı model kullanarak görüntü analizi (*Image Analysis*) alanında geliştirilen bir uygulamada, bilgisayar görmesindeki önemli bir problem olan arkaplan çıkarım işlemi gerçekleştirilmiş, klasik yöntemlere göre daha başarılı olduğu saptanmıştır

Rasmussen (1999) tarafından, Bayesian karışımı modellerdeki bileşen sayısının bulunması problemine, tamamen gibbs örneklemesine dayanan parametreden bağımsız etkili bir markov zinciri kullanılarak sonsuz Gaussian karışımı model önerilmiştir (Rasmussen, 1999).

Williams (1996) tarafından Gaussian prosesleri yardımıyla regresyon modeli önerdikleri çalışmada, MacKay'ın robot problemi diye adlandırılan bir problem ve gerçek dünyadan elde edilen beş farklı problem üzerinde çalışılmıştır. Çalışma sonucunda Gaussian regresyon modellerinin gerçek dünya problemlerine daha uygun olduğu saptanmıştır.

Wang ve arkadaşları (2002) tarafından, sinyal işleme alanındaki bir çalışmada, konuşmacı tanıma, ayırtma problemine Gaussian karışımı modeller ile çözüm bulunmaya çalışılmıştır. Kümeleme yöntemlerinin çözüme ulaştırılması gereken problemlerinin halen var olduğu belirtilmiştir. Koşullu k-means algoritmasının, gerçek verinin dağılımı olmayan küresellik varsayımı ile ve öklit uzaklıkları ile çalıştığı vurgulanmıştır. Wang ve arkadaşlarının bu çalışmasında, Gaussian karışımı modellerin kümelenmesine yön vermek için kovaryans bilgilerini kullanan yeni bir yöntem önerilmiştir. Önerilen bu yöntem, iki bölüme ayrılmıştır. İkili ağaca dayanan veri paylaşım tekniğini kullanarak kovaryans matrislerini elde etmek ve doğrudan kümeleme için kovaryans matrislerinden faydalanmak. Deney sonuçları yardımıyla, bu yöntemin, konuşmacı tanımlamasında hata oranlarının düşürülmesini sağladığı saptanmıştır.

Li ve ark. (2012), erken evre göğüs kanseri teşhisinde iki farklı sınıfa (benign sınıf ve malign sınıf) ayırmak için gauss karışım modeli temelli bir metot önerdikleri çalışmada, göğüs kanseri görüntülerinin Gaussian karışım model tabanlı model ile iki ayrı gruba sınıflandırmasının etkili olduğu belirtmişlerdir (Li, 2012).

Gaussian karışımlı model sınıflandırıcılarını tam olarak tanımlayabilmek için eğitici (unsupervised) sınıflandırıcı kategorisinde değerlendirme yapılabilir. Eğitici sınıflandırıcılarda modele dahil olan örnekler, önceden hangi kategoriye ait olduğu etiketlenmemiştir. Gaussian karışımlı modelleri denetimsiz yapan nokta, sınıflandırıcının eğitimi sırasında gözlemlerin altında yatan olasılık yoğunluğu fonksiyonlarını tahmin etmeye çalışmaktır (Scherrer, 2007). Gaussian karışımlı model sınıflandırıcıları, örüntü sınıflandırmadan (Sanderson ve ark. 2002; Omachi ve ark., 2011; Vaněk ve ark., 2013), metinden bağımsız olarak ses tanıma uygulamalarına (Reynolds ve ark., 1995) kadar birçok alanda kullanılmıştır. Marques ve ark. (1999), yaptıkları çalışmalarında, Gaussian karışımlı model sınıflandırıcılarını sinyal işleme alanında kullanarak sekiz farklı enstrümanı * oluşturduğu sesin otomatik olarak sınıflandırmasını gerçekleştirmişlerdir. Müzikal enstrüman sınıflandırması amacıyla gerçekleştirilen bir çalışmada, 0.2 saniyelik bir ses parçasından, %30 oranında hata ile gayda, klarnet, flüt, klavsen, org, piyano, trombon veya keman gibi sekiz tane enstrümanı ayırt edildiği bildirilmiştir. Ses üzerinden doğrusal örüntü katsayıları, Fourier dönüşüm katsayıları gibi öznelikler çıkarılmış, bu öznelikler üzerinde sınıflandırma amacıyla Gaussian karışımlı model ve destek vektör makineleri olarak iki algoritma önerilmiştir.

Eggink ve Brown (2003), enstrüman sesinden hangi enstrümanın olduğunun tespiti işleminin bilgisayarlara otomatik olarak yaptırılmasında, monfonik (tek enstrümanlı) seslerde Gaussian karışımlı modeller, iyi bir sınıflandırma başarısı vermektedirler. Birden fazla enstrüman sesinin mevcut olduğu polifonik müziklerde enstrüman seslerinin otomatik ayırt edilebilmesi amacıyla gerçekleştirilen bir çalışmada, probleme eksik veri teorisi ile yaklaşım, Gaussian karışımlı modeller ile sınıflandırma yapılmıştır. Çalışmanın sonucunda genelleştirilebilir, olumlu bir performans sağlandığı belirtilmiştir.

Francis ve ark. (2015)'e göre sonlu karışım modelinde bileşenlerin sayısını seçmek zorlu bir iş olduğunu belirtmişler, çalışmalarında, gözlem sırasının seçimi için

* Gayda, klarnet, flüt, klavsen, org, piyano, trombon ve keman

bilgi kriterlerinin davranışını, gözlemlenen olasılık veya bileşen etiketleri dahil olmak üzere tam olasılığa dayalı olarak incelemişlerdir. İncelemeleri sonucunda, tutarlı olduğu gösterilen AIC_{mix} olarak adlandırdıkları yeni bir gözlem olasılığı kriteri önermişlerdir. Gerçek modelde önemli derecede bir sınıflandırma belirsizliği seviyesi olduğunda, tam olasılık kriterleri, gerçek bileşen sayısını asimptotik olarak olduğundan daha küçük göstermektedir. Çalışmalarındaki, simülasyon çalışması ile, tam olasılık kriterinin(complete likelihood criteria) potansiyel olarak yetersiz sonlu-örnek performansını gösterdiğini, AIC_{mix} ve Bayesyen bilgi kriterinin ise, sınıflandırma belirsizliği düzeyine bakılmaksızın güçlü bir performans gösterdiğini öne sürmüşlerdir.

Sanderson ve Paliwal (2002)'de, olabilirlik (*likelihood*) normalizasyon tekniklerinin etkinliğinin araştırıldığı çalışmada, Gaussian karışımli modeli sınıflandırıcı olarak kullanan dört farklı sayısal yüz tanıma sisteminin performansı arttırılmıştır.

Muthén ve Shedden (1999) tarafından, aşırı alkol tüketiminin tahmini üzerine gerçekleştirilen çalışmalarında genişletilmiş bir sonlu karışımli modelin ele alınmıştır. Bağımlı değişkenin, hasta olup olmama şeklinde ifade edilen ikili (binary) bir değişken olduğu, ve belirli zaman aralıklarında ölçüm alınarak tekrarlı ölçümlerden faydalandığı belirtilmiştir. Alkol bağımlılığında yaş gruplarının veya ırk etkeninin veri kümesi içerisinde bir heterojenliğe neden olduğu, bu yüzden genişletilmiş bir karışımli model kullanıldığı belirtilmiştir. Çalışmada, genç erişkinler arasında üç gizli sınıfın ağır alkol kullanım seyrinde, alkol bağımlılığının tahminine yönelik rastgele bir katsayı büyüme modeli analiz edilmiştir.

Khalili ve Chen (2007) tarafından yapılan çalışmada, sonlu karışım regresyon kullanan uygulamalarda, genellikle birçok değişken kullanılır ve bunların bağımlı değişkene katkıları, karışım modelinin bileşenlerine göre farklılık gösterir. Akaike bilgi ölçütü ve bayesian bilgi ölçütü gibi mevcut yöntemlerde, karışımli modeldeki kovarietlerin ve alt grupların sayısı arttıkça hesaplama maliyeti artmaktadır. Sonlu karışım modellerinde değişken seçimi konusunda yapılan çalışmada cezalandırılabilirlik (*penalized likelihood*) yaklaşımı sunulmuştur. Önerilen yöntemde, regresyon katsayılarının hacmine, çeşidine ve karışım yapısına bağlı cezalar geliştirilmiştir. Geliştirilen yöntemin tutarlı olduğunu belirtilen çalışmada, simülasyonlara ek olarak iki gerçek veri üzerinde denemeler gerçekleştirilmiştir.

Dawkins ve ark. (2018)'de, taşımacılık alanında yolcu davranış profillerinin önceden saptanabilmesi amacıyla yapılan çalışmada, Bayesyen karışimli model yaklaşımı önerilmiştir.

Liang ve ark. (2018) e göre, karışimli regresyon, analizi yapılacak veri üzerinde çoklu regresyon modellerini öğrenen karışimli modelleme tasarımlarına başvurmaktadır. Bu çoklu regresyon modellerinin her biri bir alt gruba aittir. Karışimli regresyon, basit regresyon modellerinin heterojenliği yakalayamadığı durumlarda, örneklemin heterojenliğini analiz etme konusunda etkin bir tasarımdır. Sonlu karışimli regresyon kullanılarak sentetik ve gerçek veriler üzerinde örneklem alt kümelerini bulma ve aykırı örneklerin tespit edilmesi sağlanmıştır.

Yüksek hızlı bilgisayar teknolojisinin gelişmesine kadar, birden fazla boyutlu verilerin karışimli modellere uyarlanması tereddüt yaşanmıştır. Çünkü, bu işlem sırasında oluşabilecek olumsuzlukların anlaşılması olacağı düşünülmüştür. Örneğin karışimli olabilirlik fonksiyonunun birden fazla maksimum noktasının olabileceği ve eşit olmayan kovaryans matrislerine sahip normal dağılımlı bileşenlerin olması durumunda, olabilirlik fonksiyonunun sınırlarının olmayacağı düşünülmüştür. Bu hesaplama konularındaki zorlukların anlaşılması ve çözüm yollarının bulunması ile birlikte karışimli modellerde çok değişkenli yaklaşım daha etkin kullanılmaya başlanmıştır. 1960 larda, en çok olabilirlik yaklaşımı ile karışimli modeller geliştirilmiştir. 1977 de EM (Expectation-Maximization) yaklaşımının ortaya konması (Dempster ve ark., 1977) ile heterojen verinin modellenmesi için sonlu karışım dağılımlarının kullanılması konusunda büyük bir ilgi oluşmuştur (McLachlan ve Peel, 2000).

3. MATERYAL VE YÖNTEM

3.1. Materyal

Çalışmada Van Erciş Devlet Hastanesine 2017-2018 yılları içerisinde başvuran 171 hastadan alınmış olan solunum fonksiyon testi sonuçları kullanılmıştır. Hastaların solunum fonksiyon durumlarının belirlenmesinde en çok kullanılan FEV1/FVC değişkeni bağımlı değişken olarak, boy, yaş, kilo ve cinsiyet ise bağımsız değişken olarak değerlendirilmiştir.

3.2. Yöntem

3.2.1 Tek değişkenli normal dağılım

Gaussian dağılım olarak adlandırılan normal dağılım, istatistik biliminin büyük bir kısmına merkezi öneme sahiptir. Böyle olmasının üç nedeni vardır. Birincisi, normal dağılım ve bu dağılımla ilişkili olan dağılımlar, analitik olarak kolay çözülebilir yapıya sahiptir. İkinci neden olarak, normal dağılım çan şekline benzer olup bu simetrik yapısından dolayı, birçok popülasyon modelinde tercih sebebi olmuştur. Çan şekline benzer başka dağılımlar da mevcuttur, fakat normal dağılım gibi analitik olarak kolay çözülebilir yapıya sahip değildirler. Üçüncü neden ise ortalama koşullar altında, geniş örneklem kümelerinin kendileri normal dağılım göstermese bile örneklemelerinin dağılımlarının normal dağılıma yakınsamasını sağlayan merkezi limit teoremidir (Casella ve Berger, 2001a).

Normal dağılımda, genellikle μ ve σ^2 şeklinde gösterilen ve sırasıyla ortalama ve varyans diye isimlendirilen iki parametre vardır. Normal dağılımın olasılık dağılım fonksiyonu, $f(x|\mu, \sigma^2)$ şeklinde gösterilir, ve Eş.(3.1)'deki gibi yazılabilir (Casella ve Berger, 2001a).

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \quad (3.1)$$

Eğer $X \sim n(\mu, \sigma^2)$ ve $Z = (X - \mu) / \sigma$ rastgele değişkeni $n(0,1)$ dağılımına sahipse standart normal olarak adlandırılır ve koşullu standart normal dağılım fonksiyonu $P(Z \leq z)$ Eş.(3.2)'deki gibi hesaplanır (Casella ve Berger, 2001a).

$$\begin{aligned}
 P(Z < z) &= P\left(\frac{X - \mu}{\sigma} \leq z\right) \\
 &= P(X \leq z\sigma + \mu) \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{z\sigma + \mu} e^{-(x-\mu)^2/(2\sigma^2)} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt, \left(t = \frac{x - \mu}{\sigma} \text{ dönüşümü yapılırsa}\right)
 \end{aligned} \tag{3.2}$$

Tüm normal olasılıklar, standart normal terimi ile hesaplanmaya çalışılmaktadır. Ayrıca, beklenen değer, $n(\mu, \sigma^2)$ teriminin, $n(0,1)$ şekline dönüştürülerek hesaplanmaktadır. Standart normal dağılımda beklenen değer Eş.(3.3)'deki gibi hesaplanmaktadır (Casella ve Berger, 2001a).

$$EZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = -\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Big|_{-\infty}^{\infty} = 0 \tag{3.3}$$

Teorem 3.1: X rastgele değişken, a, b ve c sabitler olmak üzere herhangi bir $g_1(x)$ ve $g_2(x)$ fonksiyonları için beklenen değerler için aşağıdaki kurallar geçerlidir (Casella ve Berger, 2001b).

- $E(ag_1(X) + bg_2(X) + c) = aEg_1(X) + bEg_2(X) + c$
- Bütün x değerleri için, eğer $g_1(x) \geq 0$ sağlanıyorsa $Eg_1(X) \geq 0$ dir.
- Bütün x değerleri için, eğer $g_1(x) \geq g_2(x)$ sağlanıyorsa $Eg_1(X) \geq Eg_2(X)$ dir.
- Bütün x değerleri için, eğer $a \leq g_1(x) \leq b$ sağlanıyorsa $a \leq Eg_1(X) \leq b$ dir.

Teorem 3.1'e ve Eş.(3.3)'e göre beklenen değer

$$EX = E(\mu + \sigma Z) = \mu + \sigma EZ = \mu. \tag{3.4}$$

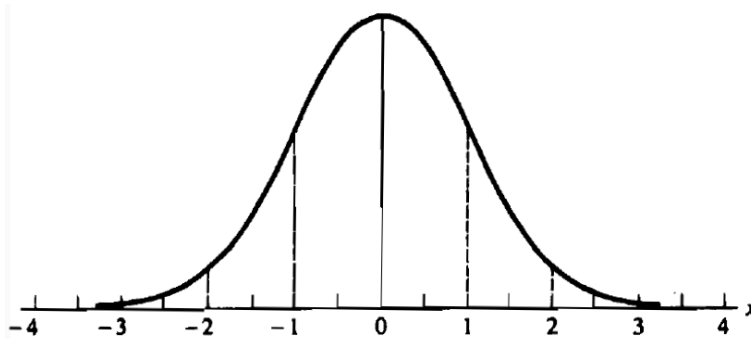
şeklinde bulunmuş olur. Aynı şekilde $VarZ = 1$ ve $VarX = \sigma^2$ eşitlikleri elde edilir. Eş. 3.1'deki ifadenin tüm reel sayılar için integrali alındığında Eş.(3.5)'de aşağıdaki gibi elde edilir (Casella ve Berger, 2001a).

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dz = 1 \quad (3.5)$$

Dikkat edildiği takdirde integralin, sıfırın etrafında simetrik olduğu, $(-\infty, 0)$ arasındaki integral ile $(0, \infty)$ arasındaki integralin birbirine eşit olduğu görülmektedir. Doğru bir matematiksel analiz ile Eş.(3.1)'deki normal olasılık yoğunluk fonksiyonunun maksimum değeri, $x = \mu$ sağlandığında elde edilir. Ayrıca $X \sim n(\mu, \sigma^2)$ ve $Z \sim n(0, 1)$ olmak kaydıyla, ortalamanın 1,2 ve 3 standart sapma uzağındaki olasılık değerleri tablolardan veya bilgisayar yardımıyla hesaplanarak elde edilmiş ve Eş.(3.6)'de gösterilmiştir (Casella ve Berger, 2001a).

$$\begin{aligned} P(|X - \mu| \leq \sigma) &= P(|Z| \leq 1) = 0.6826 \\ P(|X - \mu| \leq 2\sigma) &= P(|Z| \leq 2) = 0.9544 \\ P(|X - \mu| \leq 3\sigma) &= P(|Z| \leq 3) = 0.9974 \end{aligned} \quad (3.6)$$

Bu değerler genellikle virgülden sonra iki hane olmak üzere sırasıyla 0.68, 0.95 ve 0.99 olarak kullanılmaktadır. Bu sayılan özellikler çerçevesinde olasılık yoğunluk fonksiyonu gösterilmiştir (Şekil 3.1).



Şekil 3.1. Standart normal yoğunluk grafiği (Casella ve Berger, 2001a).

3.2.2 Çok deęişkenli normal dağılım

d – boyutlu çok deęişkenli normal dağılımın olasılık yoğunluk fonksiyonu Eş.3.7’de gösterilmiştir.

$$y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} e^{\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right)} \quad (3.7)$$

Eşitlik (3.7) te Σ : varyans kovaryans matrisi, $|\Sigma|$; varyans kovaryans matrisinin determinantını göstermektedir. Bu dağılımda da en yüksek deęer, tek deęişkenli normal dağılımda görüldüğü gibi, x vektörünün ortalama μ vektörüne eşit veya en yakın olduğu yerde görülmektedir. Eş.(3.7)’deki d deęişkeni 2’ye eşit olduğunda oluşan dağılım, iki deęişkenli (bivariate) normal dağılım olarak adlandırılır ve 3 boyutlu uzayda çan şekline benzer bir görünüme sahip olur. Çok deęişkenli dağılımın kısa gösterimi tek deęişkenli dağılıma benzer olarak;

$$X \sim N(\mu, \Sigma) \quad (3.8)$$

şeklinde gösterilir. Eş.(3.8)’e göre, “ X deęişken vektörü, μ ortalama vektörüne ve Σ varyans-kovaryans matrisine sahip çok deęişkenli normal dağılıma uygundur” denir. Çok deęişkenli normal dağılımın Eş.(3.7)’deki üssel formundaki terim kuadratik formdadır (Eş.(3.9)).

$$(x - \mu)' \Sigma^{-1} (x - \mu) \quad (3.9)$$

Eş.(3.9)’deki kuadratik forma, x rastgele deęişkeni ile μ ortalama vektörü arasındaki Mahalonobis uzaklığı denir. Eęer deęişkenler birbirleri ile korelasyon göstermiyorlarsa varyans-kovaryans matrisi diyagonaldir denir ve bu yapı Eş.(3.10)’de gösterildiği gibidir.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{pmatrix} \quad (3.10)$$

Varyans-kovaryans matrisinin diyagonal olması durumunda çok deęişkenli normal dağılım Eş.(3.11)’deki gibi gösterilebilir.

$$y = f(x, \mu, \Sigma) = \prod_{j=1}^p \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{\left(-\frac{1}{2\sigma_j^2}(x_j - \mu_j)^2\right)} \quad (3.11)$$

Eş.(3.11)'de verilen \prod operatörü, sigma toplama operatörü gibi davranır fakat $j=1$ den p ye kadar tüm terimleri çarparak sonuç elde edilir. \prod çarpım operatörünün yanındaki terim, x rasgele değişkenin j indisine sahip olması farkıyla tek değişkenli normal dağılım fonksiyonuna çok benzemektedir. Bu durumda, rasgele vektörün elemanları olan X_1, X_2, \dots, X_p bağımsız rasgele değişkenler olarak adlandırılır.

Eş.(3.12)'de gösterildiği gibi çok değişkenli normal rasgele değişkenin öğelerinin doğrusal kombinasyonları dikkate alınabilir.

$$Y = \sum_{j=1}^p c_j X_j = c'X \quad (3.12)$$

Doğrusal bir kombinasyonu tanımlamak için X_j rastgele değişkenlerinin ilişkisiz olmasına gerek yoktur. X rastgele değişkeninin Eş.(3.8)'de gösterildiği gibi çok değişkenli normal dağılım gösterdiği varsayıldığında, Y değişkeni, ortalaması Eş.(3.13)'de gösterildiği gibi, ve Eş.(3.14)'de gösterildiği gibi bir varyansa sahip şekilde normal dağılım gösterir.

$$c'\mu = \sum_{j=1}^p c_j \mu_j \quad (3.13)$$

$$c'\Sigma c = \sum_{j=1}^p \sum_{k=1}^p c_j c_k \sigma_{jk} \quad (3.14)$$

Özetle, Y değişkeni ortalaması $c'\mu$ ve varyansı $c'\Sigma c$ olan normal dağılım gösterir (Eş.(3.15)).

$$Y \sim N(c'\mu, c'\Sigma c) \quad (3.15)$$

Ortalama vektörü μ , varyans-kovaryans matrisi Σ olan çok değişkenli normal dağılım gösteren veriler için şu önermeler geçerlidir:

- Her bir değişken tek değişkenli normal dağılım gösterir. Böylece, çok değişkenli normallik değerlendirilirken, her bir değişken için tek değişkenli normal dağılıma uygunluk testlerine başvurulabilir.
- Tüm veri çok değişkenli normal dağılım gösteriyorsa, her hangi bir değişken seti de çok değişkenli normal dağılım gösterecektir.

- Değişkenlerin herhangi doğrusal bir kombinasyonu tek değişkenli normal dağılım gösterir.

Çok değişkenli normal dağılımın daha iyi anlaşılabilmesi için iki değişkenli (bivariate) normal dağılımın incelenmesi gerekmektedir. Ortalama vektör bileşenleri μ_1 ve μ_2 olan, varyans-kovaryans matrisi Eş.(3.16)'de gösterildiği gibi olan X_1 ve X_2 gibi iki rastgele değişkenin,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right] \quad (3.16)$$

şeklinde iki değişkenli normal dağılım gösterdiği varsayıldığında; eşitlikte gösterilen matrisin diagonal elemanları varyansları, geriye kalan diyagonal olmayan elemanları ise kovaryansları temsil etmektedir. Kovaryans, iki değişkenin korelasyonu ile standart sapmalarının çarpımlarına eşittir. Varyans-kovaryans matrisinin determinanı iki varyansın ve korelasyonun karesinin 1 den farkının çarpımlarına eşittir (Eş.(3.17)).

$$|\Sigma| = \sigma_1^2 \sigma_2^2 (1 - \rho^2) \quad (3.17)$$

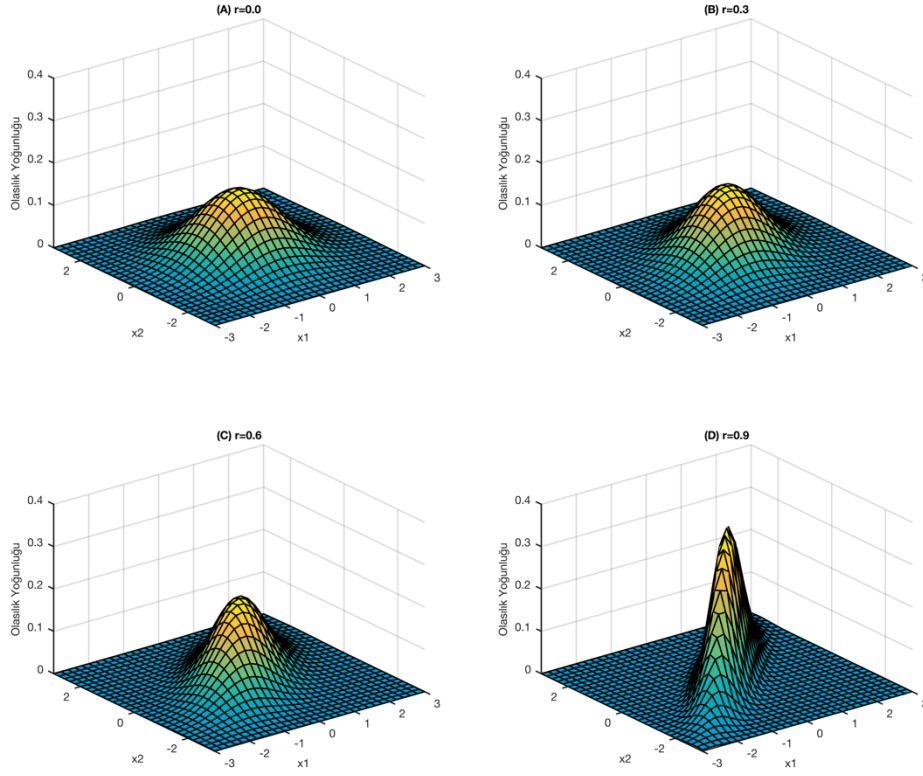
Varyans-kovaryans matrisinin tersi şu şekilde hesaplanır;

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \quad (3.18)$$

Varyans-kovaryans matrisinin determinantını ve tersini, Eş.(3.7)'de yerine konularak gerekli sadeleştirmeler yapıldığında, iki değişkenli normal dağılım yoğunluk fonksiyonu;

$$\phi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \quad (3.19)$$

şeklinde hesaplanır. Çeşitli ρ korelasyonlara göre iki değişkenli normal dağılım gösteren değişkenlerin grafikleri Şekil 3.2'de gösterilmiştir. Şekil 3.2.a, korelasyonun (ρ) sıfıra eşit olduğu durumu göstermektedir. Bu özel dağılım çeşidi *dairesel normal dağılım* olarak adlandırılmaktadır. Bu çeşit dağılımda, 3 boyutta tam olarak simetrik bir çan eğrisine ulaşılmaktadır. ρ değeri arttıkça, çan şeklindeki eğri, 45 derecelik çizgi boyunca düzleşir. Yani, ρ değerleri sırasıyla 0.3, 0.6 ve 0.9 değerlerini aldıkça eğrinin -3 ve +3'e doğru uzandığı görülmekte ve dik yönde düzleşmektedir (Şekil 3.2.b,c,d).

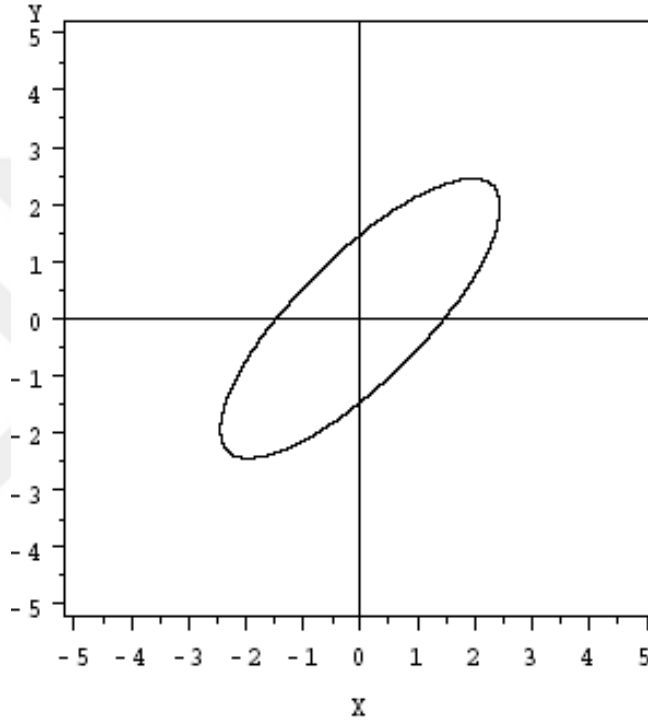


Şekil 3.2. Farklı ρ korelasyonlara göre iki değişkenli (bivariate) normal dağılım gösteren X_1 ve X_2 değişkenlerinin olasılık yoğunluğu grafikleri. a) $\rho = 0.0$ b) $\rho = 0.3$ c) $\rho = 0.6$ d) $\rho = 0.9$.

Eş.(3.7)'de ifade edilmiş çok değişkenli normal yoğunluk fonksiyonunun eksponansiyel terimi Eş.(3.9)'de Mahalanobis uzaklığı olarak tanımlanmıştır. Eş.(3.9)'de gösterilen $(x - \mu)' \Sigma^{-1} (x - \mu)$ ifadesi, merkezi μ olan hiper-elips denklemdir. İki değişkenli normal dağılımda, Mahalanobis uzaklığını gösteren elipsin grafiği Şekil 3.3'de gösterilmiştir. $(x - \mu)' \Sigma^{-1} (x - \mu)$ terimi hakkında önemli olarak şu özelliklerden bahsedilebilir:

- c herhangi bir sabit değer olmak üzere, $(x - \mu)' \Sigma^{-1} (x - \mu) = c$ eşitliğini sağlayacak tüm x ler için, $f(x)$ yoğunlukları eşittir ve aynı olasılığa sahiptirler.

- $(x - \mu)' \Sigma^{-1} (x - \mu)$ değeri arttıkça, yoğunluk fonksiyonu değeri azalır. $(x - \mu)' \Sigma^{-1} (x - \mu)$ değerinin artması, x ve μ arasındaki farkın (uzaklığın) artması anlamına gelmektedir.
- $d^2 = (x - \mu)' \Sigma^{-1} (x - \mu)$ değişkeni, p (değişken sayısı) serbestlik dereceli χ^2 dağılımı göstermektedir.



Şekil 3.3. İki değişkenli normal dağılımda eksponansiyel terimin elips şeklindeki grafiği.

3.2.3 Karışımli modellerin genel yapısı

Y_1, \dots, Y_n ifadesi, \mathfrak{R}^p üzerindeki olasılık yoğunluk fonksiyonu $f(y_j)$ olan, n büyüklüğünde ve bu ifadedeki Y_j ler p boyutunda olup, rastgele bir örnekleme ifade etmektedir. Uygulamada, Y_j , çalışmadaki özelliklerin j inci gözleminden elde edilen p kadar ölçüme karşılık gelen rasgele değişkenleri içermektedir. Üstel T ifadesi, transpoz işlemini ifade etmek üzere tüm örneklem $Y = (Y_1^T, \dots, Y_n^T)^T$ şeklinde ifade edilir. Burada Y , \mathfrak{R}^p içindeki noktaların n boyutlusudur. Mümkün merteye, rastgele

bir vektörün gerçekleştirilmesi ilgili küçük harfle gösterilmektedir. Örneğin, Y_j rastgele vektörünün gözlenen değeri y_j olmak üzere, $y = (y_1^T, \dots, y_n^T)^T$, gözlenen rastgele bir örnekleme temsil eder (McLachlan ve Peel, 2000).

Her ne kadar, Y_j özellik vektörünü sürekli bir rastgele vektör gibi alsak ta, Y_j sayımla elde edilmiş ölçümlerden oluştuğu kabul edilen $f(y_j)$ yoğunluklu kesikli bir yapıdadır. $f(y_j)$ yoğunluğu Eş.(3.20)'deki gibi gösterilir (McLachlan ve Peel, 2000).

$$f(y_i) = \sum_{i=1}^g \pi_i f_i(y_j), \quad (3.20)$$

burada $f_i(y_j)$ yoğunlukları, π_i ise negatif olmayıp Eş.(3.22)'deki gibi toplamları 1'e eşit olan büyüklükleri(oranları) göstermektedir (McLachlan ve Peel, 2000).

$$0 \leq \pi_i \leq 1, \quad (i = 1, \dots, g) \quad (3.21)$$

ve

$$\sum_{i=1}^g \pi_i = 1 \quad (3.22)$$

π_1, \dots, π_g büyüklükleri, karışma oranları (mixing probability) veya ağırlıklar olarak ifade edilir. $f_i(y_j)$, karışımdaki alt grup yoğunlukları olarak adlandırılır. Eş.(3.20)'deki ifade g alt gruplu sonlu karışım yoğunluğu olarak ifade edilebilir. İlgili dağılım fonksiyonu $F(y_i)$ şeklinde gösterilip, g alt gruplu sonlu karışım dağılımı diye adlandırılabilir. Karışımların sonlu dağılımları, genel olarak karışımli modeller olarak ifade edilmiştir (McLachlan ve Peel, 2000).

Karışımli modellerin bu formüllerindeki g alt grup sayısı, sabit şeklinde nitelendirilir. Fakat birçok uygulamada, doğal olarak g alt grup sayısı bilinmemektedir. Alt grup sayısı, alt grup yoğunlukları için belirlenen parametreler ve karışım oranları ile beraber üzerinde çalışılan verilerden elde edilmektedir (McLachlan ve Peel, 2000).

3.2.4 Karışımli modellerin açıklanması

g – alt gruplu $f(y_j)$ karışımli yoğunluğa sahip Y_j rastgele vektörü Eş.(3.20)'de gösterilmiştir. Z_j kategorik rasgele değişken olmak üzere, sırasıyla π_1, \dots, π_g olasılıklara sahip $1, \dots, g$ arası değerler almaktadır. Y_j 'nin koşullu yoğunluğu, $Z_j = i$ olacak şekilde bir $f_i(y_j)$ ($i=1, \dots, g$) olarak varsayılmaktadır. Y_j 'nin koşulsuz yoğunluğu (marjinal yoğunluk) ise $f(y_j)$ şeklinde verilmektedir. Bu bağlamda, Z_j , Y_j özellik vektörünün alt grup etiketlerini tutan bir değişken olarak düşünülebilir. İlerleyen aşamalarda, Z_j kategorik değişken yerine g boyutlu alt grup etiket vektörü olarak gösterilen, i ninci elemanı $Z_{ij} = (Z_j)_i$ şeklinde gösterilen, karışımdeki Y_j kaynaklı alt grupların i ye eşit olup olmamasına göre 1 veya 0 değerini alan Z_j ile çalışmak daha uygun olacaktır ($i=1, \dots, g$). Böylece, Z_j , olasılıkları π_1, \dots, π_g olan g sayıdaki kategoriye içeren Eş.(3.23)'deki gibi multinominal dağılım gösterir (McLachlan ve Peel, 2000).

$$pr\{Z_j = z_j\} = \pi_1^{z_{1j}} \pi_2^{z_{2j}} \dots \pi_g^{z_{gj}} \quad (3.23)$$

Eş.(3.23)'de ifade edilen multinominal dağılım, $\pi = (\pi_1, \dots, \pi_g)^T$ olmak üzere kısaca Eş.(3.24)'deki gibi gösterilir (McLachlan ve Peel, 2000).

$$Z_j \sim Mult_g(1, \pi) \quad (3.24)$$

Eş.(3.20)'de sunulan karışımli modeldeki Y_j , π_1, \dots, π_g oranlarına sahip g alt grubu içeren $G = G_1, \dots, G_g$ popülasyonundan çekilmiştir. Eğer, $i=1, \dots, g$ için G_i grubundaki Y_j nin yoğunluğu $f_i(y_j)$ şeklinde verilmişse, Y_j Eş.(3.20)'deki durumun g alt gruplu karışım formudur denilir. Bu durumda, karışımın g alt grubu, G_1, \dots, G_g harici grupları içinden oluşmuştur (McLachlan ve Peel, 2000).

Uygulamalarda, heterojenliğin kaynağı örnek olarak yaş, cinsiyet, tür, coğrafi köken gibi faktörler olabilmektedir. Örneğin G popülasyonu, üyelerinin belirgin bir hastalığa sahip olup olmamasına göre, iki alt popülasyon olan G_1 ve G_2 popülasyonlarını içermektedir. Bu problemi çözüme ulaştırabilmek için, popülasyondaki üyelerin rastgele seçilmiş örnekleme üzerinden elde edilen ölçümler

üzerindeki bazı özellik vektörü temelinde hastalığın görülme sıklığının(yani π_1 karışma oranının) tahmin edilmesi gerekmektedir. Örneğin bir çalışmada (Do ve McLachlan, 1984) Malezyalı farelerden toplanan $p = 4$ ölçüme bakılarak, karışimli modele en iyi uyum gösteren grup sayısı $g = 7$ bulunmuş ve farelerin aslında 7 farklı türden oluştuğu saptanmıştır (McLachlan ve Peel, 2000).

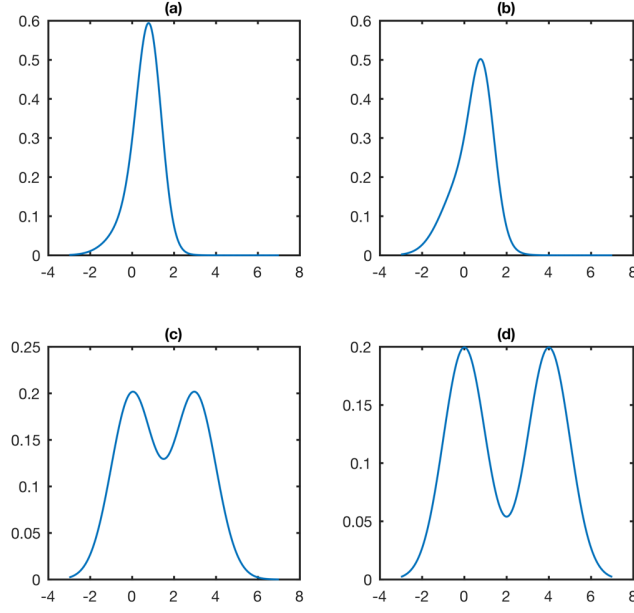
3.2.5 Eş varyanslı iki normal dağılımın karışımı

Tek değişkenli normal dağılımların karışımlarının anlaşılması için öncelikle ortak σ^2 varyansına sahip, ortalamaları μ_1 ve μ_2 , karışım oranları ise π_1 ve π_2 olan iki adet normal dağılımın karışımının Eş.(3.25)'deki gibi ifade edilmesi gerekmektedir (McLachlan ve Peel, 2000).

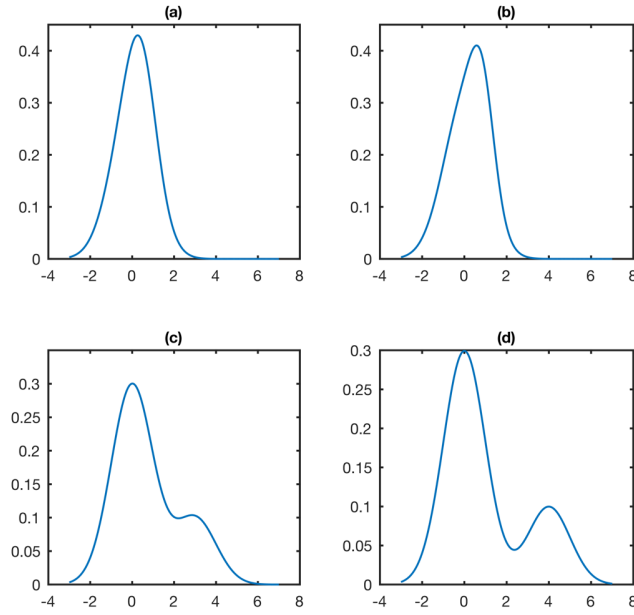
$$f(y_j) = \pi_1 \phi(y_j; \mu_1, \sigma^2) + \pi_2 \phi(y_j; \mu_2, \sigma^2) \quad (3.25)$$

Buradaki $\phi(y_j; \mu_1, \sigma^2)$ ifadesi ortalaması μ_1 , varyansı σ^2 olan Eş.(3.1)'de gösterilen normal dağılımın olasılık yoğunluk fonksiyonunu $f(x|\mu, \sigma^2)$ göstermektedir. Bundan sonraki kısımlarda, normal dağılımın olasılık yoğunluk fonksiyonu, daha sade bir gösterim olması açısından $\phi(y_j; \mu_1, \sigma^2)$ şeklinde gösterilecektir.

Eğer iki bileşenin normal yoğunlukları birbirlerinden ayrık ise $f(y_j)$ karışım yoğunluğu, birbirine benzeyen iki normal yoğunluğun bileşimi şeklinde gözükür ki buna iki tepeli (bimodal) yoğunluk denir. Bu durumu göstermek amacı ile normal karışimli yoğunluğun $\mu_1 = 0, \mu_2 = \Delta, \sigma^2 = 1$ olmak ve $\pi_1 = \pi_2 = 0.5$ eşit oranlarda olmak kaydı ile çeşitli Δ değerlerine göre elde edilmiş grafik çizimleri MATLAB yazılımı ile oluşturulmuştur (Şekil 3.4). Şekil 3.4'de görüldüğü gibi Δ değeri arttıkça $f(y_i)$ karışimli yoğunluğu, tek tepeli halden iki tepeli hale dönüşmektedir.



Şekil 3.4. Varyansları eşit ve $\sigma^2 = 1$ olan iki adet ortalamaları $\mu_1 = 0$ $\mu_2 = \Delta$ olan tek tepeli normal bileşenlerin eşit miktarlardaki karışımları a) $\Delta = 1$; b) $\Delta = 2$; c) $\Delta = 3$; d) $\Delta = 4$.



Şekil 3.5. Varyansları eşit ve $\sigma^2 = 1$ olan iki adet ortalamaları $\mu_1 = 0$ $\mu_2 = \Delta$ olan tek tepeli normal bileşenlerin 0,75 ve 0,25 oranlarındaki karışımları a) $\Delta = 1$; b) $\Delta = 2$; c) $\Delta = 3$; d) $\Delta = 4$.

Burada, ortak varyansa sahip sadece $g = 2$ bileşene sahip normal karışımlar ele alınmıştır. İleriki kısımlarda, iki bileşenden daha fazla bileşene sahip olan ve eşit olmayan varyanslara sahip olan normal karışımların esnek yapılarına değinilmiştir.

3.2.6 Çok değişkenli karışimli normal dağılım

Eisenberger (1964) çalışmasında, değişken σ_1^2 ve σ_2^2 varyanslara sahip iki normal dağılımın karışımlarının şekillerinin iki tepeli (bimodal) ya da tek tepeli (unimodal) olma durumlarını incelemiştir. Eş.(3.25)'de gösterilen iki normal dağılımın karışımlarını gösteren yoğunluk fonksiyonu için, Eş.(3.26)'de gösterilen şart sağlandığında tek tepeli olduğu söylenir. Benzer şekilde, Eş.(3.25)'de gösterilen karışimli yoğunluk fonksiyonu için, Eş.(3.27)'deki şart sağlandığında, bu fonksiyonun şeklinin iki tepeli olduğu söylenir (Eisenberger, 1964).

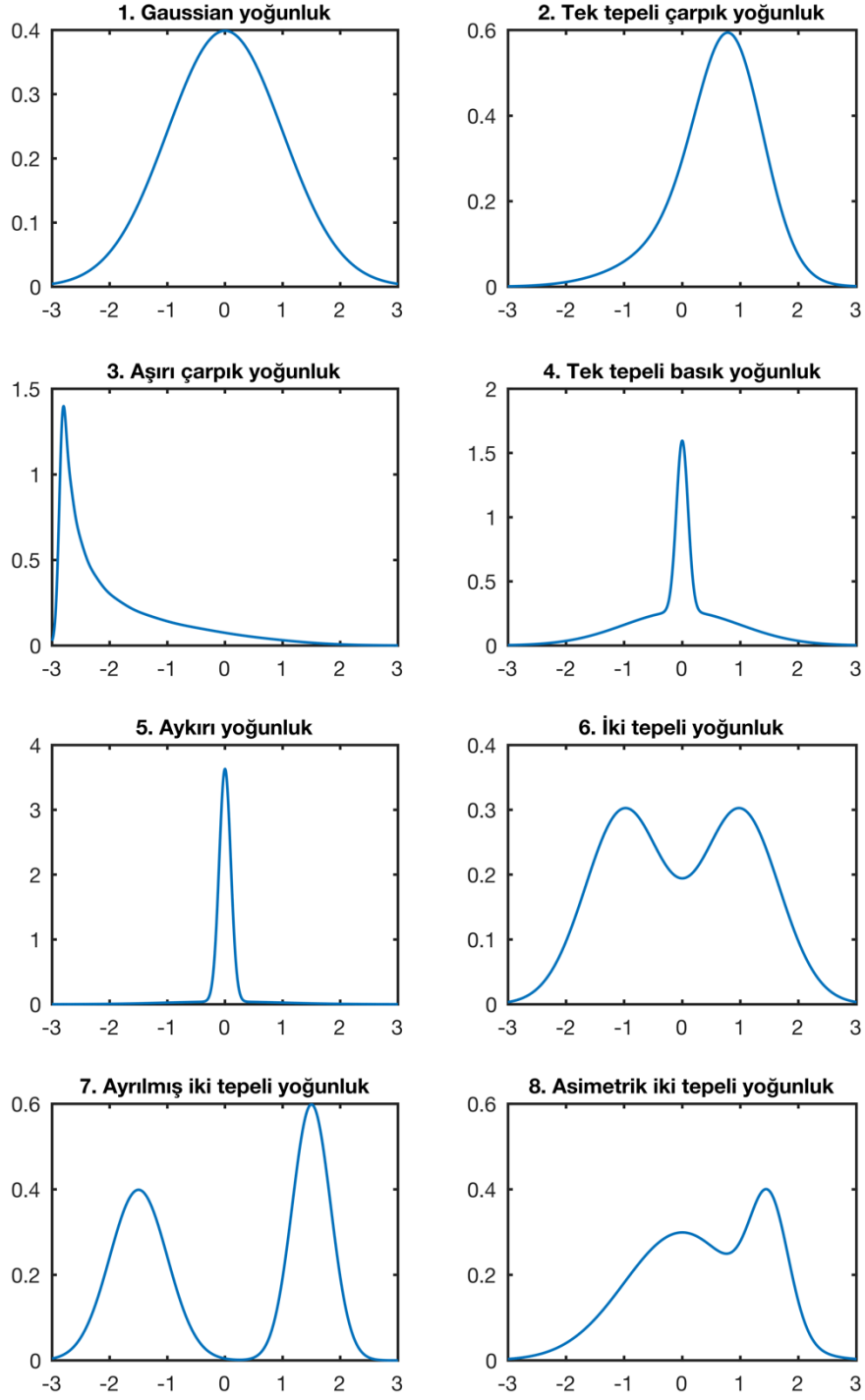
$$(\mu_2 - \mu_1)^2 < \frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2 + \sigma_2^2)} \quad (3.26)$$

$$(\mu_2 - \mu_1)^2 > \frac{8\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (3.27)$$

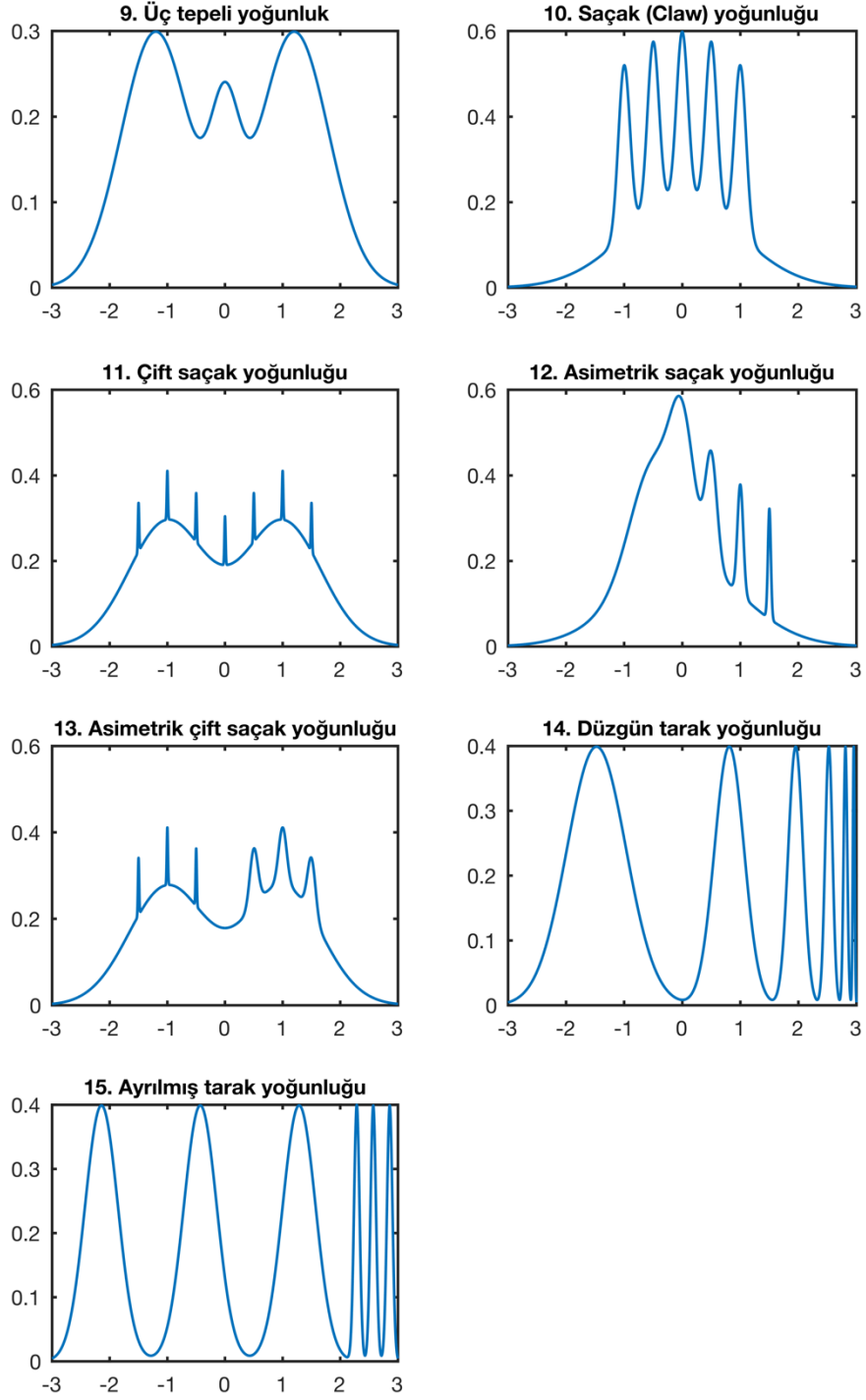
Marron ve Wand (1992) ye göre, \mathcal{S}^- alt gruplu karışımlar esnek yapılara sahiptir ve çalışmalarında geniş bir çeşitlilikte farklı karışım yoğunluklarını analitik olarak incelenmiştir. Çok geniş olan normal karışimli dağılım ailesini göstermek için, Çizelge 3.1.'de verilen on beş tane tek değişkenli normal karışimli dağılım sunulmuştur (Marron ve Wand, 1992). Çizelge 3.1 deki formüller kullanılarak MATLAB programı yardımı ile normal karışimli dağılımların grafikleri oluşturularak Şekil 3.6'de gösterilmiştir.

Çizelge 3.1. Normal karışımli yoğunluğun on beş örneği için parametreler.

Yoğunluk	$f(y)$ yoğunluk fonksiyonu
1. Gaussian	$N(0,1)$
2. Tek tepeli çarpık	$\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}N\left(\frac{13}{15}, \left(\frac{5}{9}\right)^2\right)$
3. Aşırı çarpık	$\sum_{i=0}^7 \frac{1}{8}N\left(3\left(\left(\frac{2}{3}\right)^i - 1\right), \left(\frac{2}{3}\right)^{2i}\right)$
4. Tek tepeli basık	$\frac{2}{3}N(0,1) + \frac{1}{3}N\left(0, \left(\frac{1}{10}\right)^2\right)$
5. Aykırı	$\frac{1}{10}N(0,1) + \frac{9}{10}N\left(0, \left(\frac{1}{10}\right)^2\right)$
6. İki tepeli	$\frac{1}{2}N\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{1}{2}N\left(1, \left(\frac{2}{3}\right)^2\right)$
7. Ayrılmış iki tepeli	$\frac{1}{2}N\left(-\frac{3}{2}, \left(\frac{1}{2}\right)^2\right) + \frac{1}{2}N\left(\frac{3}{2}, \left(\frac{1}{2}\right)^2\right)$
8. Asimetrik iki tepeli	$\frac{3}{4}N(0,1) + \frac{1}{4}N\left(\frac{3}{2}, \left(\frac{1}{3}\right)^2\right)$
9. Üç tepeli	$\frac{9}{20}N\left(-\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{9}{20}N\left(\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{1}{10}N\left(0, \left(\frac{1}{4}\right)^2\right)$
10. Saçak (Claw)	$\frac{1}{2}N(0,1) + \sum_{i=0}^4 \frac{1}{10}N\left(\frac{i}{2} - 1, \left(\frac{1}{10}\right)^2\right)$
11. Çift saçak	$\frac{49}{100}N\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{49}{100}N\left(1, \left(\frac{2}{3}\right)^2\right) + \sum_{i=0}^6 \frac{1}{350}N\left(\frac{i-3}{2}, \left(\frac{1}{100}\right)^2\right)$
12. Asimetrik saçak	$\frac{1}{2}N(0,1) + \sum_{i=-2}^2 \frac{2^{1-i}}{31}N\left(i + \frac{1}{2}, \left(\frac{2^{-i}}{10}\right)^2\right)$
13. Asimetrik çift saçak	$\sum_{i=0}^1 \frac{46}{100}N\left(2i - 1, \left(\frac{2}{3}\right)^2\right) + \sum_{i=1}^3 \frac{1}{300}N\left(-i/2, \left(\frac{1}{100}\right)^2\right) + \sum_{i=1}^3 \frac{7}{300}N\left(i/2, \left(\frac{7}{100}\right)^2\right)$
14. Düzgün tarak	$\sum_{i=0}^5 \frac{2^{5-i}}{63}N\left(\left(\frac{65 - 96\left(\frac{1}{2}\right)^i}{21}\right), \left(\frac{32}{63} \cdot \frac{1}{2^i}\right)^2\right)$
15. Ayrılmış tarak	$\sum_{i=0}^2 \frac{2}{7}N\left(\frac{12i-15}{7}, \left(\frac{2}{7}\right)^2\right) + \sum_{i=8}^{10} \frac{1}{21}N\left(\frac{2i}{7}, \left(\frac{1}{21}\right)^2\right)$



Şekil 3.6. Karışım normal dağılımların grafikleri.



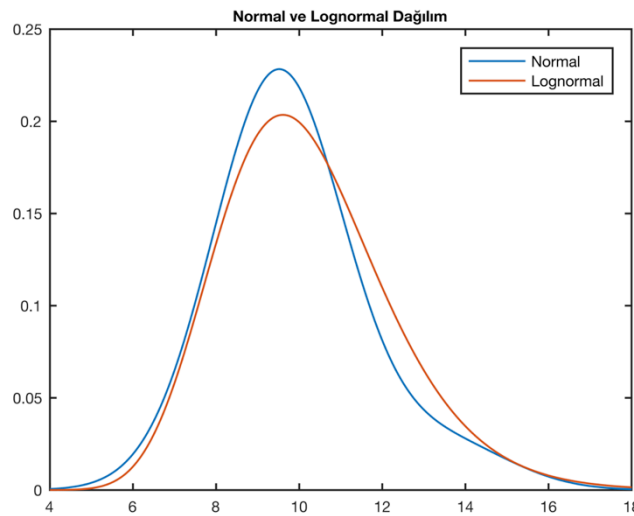
Şekil 3.7. Karışıklı normal dağılımların grafikleri (devam).

3.3. Veri Asimetrikliğinin Modellenmesi

Buraya kadar anlatılan kısımlarda, normal karışımı dağılımların, asimetrik dağılıma sahip verilerin dağılımlarının modellenmesinde faydalı bir rolü olduğu gösterilmiştir. Çarpık verilerin modellenmesinde kullanılabilecek diğer bir yol, verideki çarpıklığın giderilmesi veya azaltılması amacıyla uygulanacak ilk dönüşümü uygulamaktır. Bu görevi gerçekleştirmek için log dönüşümü genellikle faydalı bir yöntemdir. Bilindiği üzere, iki tek değişkenli eş varyanslı normal karışımın parametreleri, log normal dağılımın görünümüne yakın olacak yoğunlukta seçilebilir. Y_j rasgele değişkeni logaritmik ölçekte, normal dağılıma sahip, ortalaması μ ve varyansı σ^2 ise, yoğunluk fonksiyonu Eş.(3.28)'deki gibi gösterilir (McLachlan ve Peel, 2000).

$$\frac{\sqrt{2\pi}}{y_j\sigma} \exp\left\{-\frac{1}{2} \cdot \frac{(\log y_j - \mu)^2}{\sigma^2}\right\} \quad (3.28)$$

Titterington ve ark. (1985) ise, Eş.(3.25)'de gösterilen, parametreleri $\pi_1 = 0.9$, $\mu_1 = 9.5$, $\mu_2 = 13.5$, $\sigma^2 = 2.5$ olan iki alt gruba sahip normal karışım yoğunluğu, ve Eş.(3.28)'de gösterilmiş $\mu = \log(10)$, $\sigma^2 = 2.5$ parametrelerine sahip olan log normal yoğunluğunu incelemişlerdir (Şekil 3.8).



Şekil 3.8. Lognormal ve iki bileşenli normal karışımın yoğunluk grafiği.

İki alt gruba sahip karışım normal dağılım ile log normal dağılımın birbirine bu kadar benzemesi, uygulamada birbirlerinden ayırt edilmesinin zor olduğu anlamına gelmektedir. Altta yatan dağılıma tatmin edici şekilde yakınsayacak bir dağılım bulunmak istenen analizlerde bu tarz zorluklarla karşılaşmaktadır. Normal karışım model ve log normal modelden hangisinin seçilmesi gerektiği ilgi çekici bir konu olmuştur (Swales, 1985; Schork ve ark., 1990; Kasahara ve Shimotsu, 2014, de Bragança Pereira ve de Bragança Pereira.2016; Wallace ve ark., 2018)

3.4. Karışım Modelin Parametrik Gösterimi

Bir çok uygulamada, $f_i(y_j)$ alt grup yoğunluklarının bazı parametrik ailelere ait oldukları belirtilmiştir. Bu durumda, bilinmeyen parametrelerin karışımındaki i ninci alt grup yoğunluğu için kabul edilen formdaki vektörü θ_i olmak üzere, $f_i(y_j)$ alt grup yoğunlukları, $f_i(y_j; \theta_i)$ şeklinde belirtilir. $f(y_j)$ karışım yoğunluğu Eş.(3.29)'deki gibi ifade edilebilir.

$$f(y_j; \psi) = \sum_{i=1}^g \pi_i f_i(y_j; \theta_i) \quad (3.29)$$

Eş.(3.29)'deki ψ vektörü, karışımındaki bilinmeyen tüm parametreleri içerir ve Eş.(3.30)'deki gibi gösterilir.

$$\psi = (\pi_1, \pi_2, \dots, \pi_{g-1}, \xi^T)^T \quad (3.30)$$

Eş.(3.30)'deki ξ vektörü, $\theta_1, \theta_2, \dots, \theta_g$ vektörleri içindeki tüm parametreleri içerir. ψ için belirlenmiş parametre uzayı Ω ile gösterilmektedir. π_i karışım oranlarının toplamı 1 e eşit olacağı için bunlardan biri gereksizdir. Bu yüzden, Eş.(3.30)'de g ninci karışım oranını gösterecek olan π_g ifadesi keyfi olarak ihmal edilmiştir.

Parametrik bir karışımı tanımlamak için yukarıdaki eşitlikler kullanıldığında, Kanji (1985)'nin uçak inişlerinde karşılaşılan rüzgar makaslarının dağılımlarının modellenmesinde önerdiği gibi ortak bir μ ortalamaya sahip Laplace bileşeni ve tek değişkenli normal bileşenin karışımı ele alınmaktadır. Bu modelde rüzgar makaslarının Y_j karışım yoğunluğu;

$$\psi = (\pi_1, \xi^T)^T \text{ olduğunda ve } \xi = (\mu, \sigma^2, \kappa)^T \text{ iken} \quad (3.31)$$

$$f(y_j; \psi) = \pi_1 \phi(y_j; \mu, \sigma^2) + \pi_2 (2\kappa)^{-1} \exp(-|y_j - \mu|/\kappa)$$

şeklinde gösterilmektedir. Çoğu zaman alt grup yoğunlukları, aynı parametrik aileye ait olarak belirtilir. θ için parametre uzayı Θ şeklinde ifade edildiğinde,

$$\{f(y_j, \theta) : \theta \in \Theta\} \quad (3.32)$$

olarak gösterilen parametrik ailenin genel üyesi, $f(\cdot, \theta)$ şeklinde ifade edilir ve $f(y_j; \psi)$ karışım yoğunluğu;

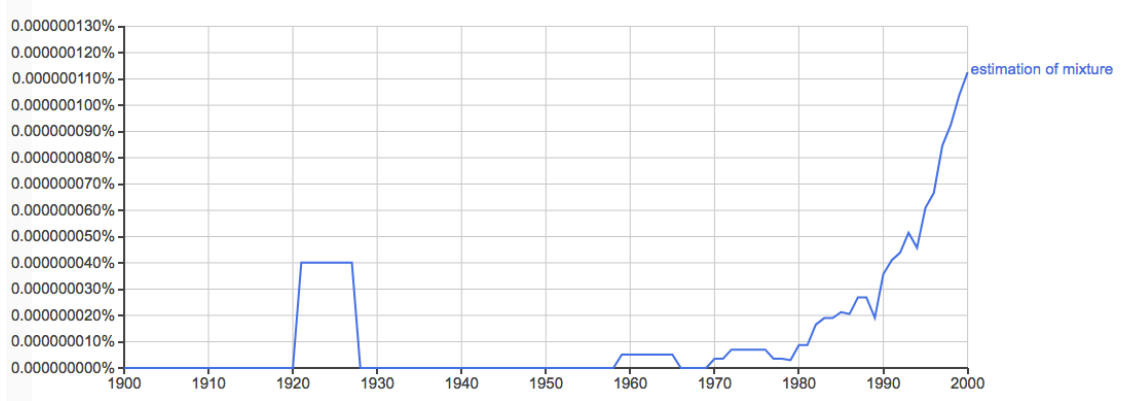
$$f(y_j; \psi) = \sum_{i=1}^g \pi_i f(y_j; \theta_i) \quad (3.33)$$

şeklinde gösterilir. Genel bir gösterim olarak f sembolü, yoğunluk fonksiyonunu belirtmek amacıyla kullanılmaktadır. Örneğin $f(y_i; \psi)$ ifadesi karışım yoğunluğunu göstermek için kullanılmıştır. $f(y_i; \theta_i)$ ifadesi ise Eş.(3.31) ailesinin altındaki i ninci alt grup yoğunluğunu göstermektedir.

3.5. Karışımli Modellerde Parametre Tahmini

Karışımli dağılımları tahmin edebilmek için çeşitli yaklaşımlar ortaya konmuştur. Bu yaklaşımlar, grafiksel metotları, momentler, en yakın mesafe, en çok olabilirlik (Maximum Likelihood - ML) ve bayesian gibi metotları içermektedir. Karışımli için tahmin metodolojileri üzerine bu kadar geniş bir literatür olmasının ana nedeni olarak, parametre tahmini için açık formüllerin tipik olarak mevcut olmamasıdır (Titterington, 1997). Örneğin, karışma oranları ve alt grup ortalama ve varyans/kovaryansları için en çok olabilirlik tahmini (MLE-Maximum Likelihood Estimation), normal karışımli söz konusu olduğunda kapalı bir formda yazılıp gösterilemez. Bu olabilirlik tahminleri iteratif olarak hesaplanır. Literatür bildirişi bölümünde de değindiğimiz Pearson'un çalışmasında (1894), iki tek değişkenli normal dağılım karışımının parametre tahmini yapılırken momentler metodu kullanılmıştır. Son yıllarda artan bir şekilde (Şekil 3.9), normal karışımli parametre tahminleri üzerine

çeşitli çalışmalar gerçekleştirilmiştir (Lindsay ve Başak, 1993; Furman ve Lindsay, 1994; Withers 1996; Ramos-López ve ark., 2018; Orellana ve ark., 2018).



Şekil 3.9. Karışım dağılımları konusundaki akademik yazıların yıllara göre değişimi.

3.5.1. Karışımli Modelde Alt Grup Sayısının Test Edilmesi

Karışımli modellerin bazı uygulamalarında, karışımın içinde bulunan g alt grup sayısı hakkında yeterli bir ön bilginin bulunmadığı durumlarla karşılaşmaktadır. Bu duruma örnek olarak, alt grupların, normal dağılım gösteren harici gruplara karşılık gelmesi verilebilir. Bununla birlikte, birçok durumda, alt grup sayısı, alt grup yoğunluklarındaki parametrelerle birlikte verilerden çıkarılmalıdır. Örneğin, McLaren ve arkadaşları (1991), iki bileşenli log normal karışım yoğunluğu kullanarak, önceden anemi hastalığına yakalanmış kişilerin akyuvar hücre hacimleri üzerindeki bazı dağılımları açıklamaya çalışmışlardır. Sağlıklı kişilerdeki akyuvar hacim dağılımları sadece bir log normal dağılım ile modellenebiliyorken, anemisi tam olarak iyileşmemiş hastalarda akyuvar hacimleri demir tedavisi bitiminde iki log normal dağılımın karışımı şeklinde görülmüştür. McLaren ve arkadaşlarının çalışmasında, istatistiksel testler yapılarak, demir tedavisine ne zaman son verilebileceği konusunda hekimlere erken teşhis imkanı sunulabileceği sonucuna ulaşılmıştır.

3.5.2. Karışımli Modellerde En Çok Olabilirlik Yöntemi

EM algoritması geliştirilinceye kadar geçen sürede, en çok olabilirlik, karışımli dağılımların modellenmesinde sıkça kullanılmıştır.

3.7.1 Karışımli modellerde en çok olabilirlik

En çok olabilirlik yaklaşımı ile d -boyutundaki Ψ parametre vektörünün $f(y_i, \Psi)$ dağılımına uyduğu varsayıldığında, Y_j şansa bağlı vektörü çalışmadaki j inci olguyu göstermek şartıyla, $\hat{\Psi}$ tahmini, düzenli duruma uygun çözüm olarak Eş.(3.34)'de gösterilen en çok olabilirliği sağlar.

$$\partial L(\Psi) / \partial(\Psi) = 0 \quad \text{veya} \quad \partial \log L(\Psi) / \partial(\Psi) = 0 \quad (3.34)$$

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

$L(\Psi)$, Ψ şeklindeki bağımsız olduğu varsayılan y_1, \dots, y_n değişkenlerin en çok olabilirlik fonksiyonudur.

Daha kısa olarak ifade edilecek olunursa, en çok olabilirlik tahmin işleminin amacı, her bir n için tahmin değerlerini saptamaktır (yukarıda $\hat{\Psi}$ şeklinde gösterilmiştir). Böylelikle en çok olabilirlik tahmini, en çok olabilirlik eşitliğinin kök dizisini tutarlı ve asimptotik olarak etkili bir şekilde sağlar. Böyle bir dizi, uygun düzenlilik koşulları sağlandığında bilinir. Olasılığın 1 e tamamlanma eğiliminde olması ile, parametre uzayı içinde bu kökler yerel maksimuma yakınsar. Bu köklerin tutarlı sıralaması aslında benzersizdir.

3.5.2. Bilgi matrisleri

Gözlenen değerler $y = (y_1^T, \dots, y_n^T)^T$ olmak üzere log olabilirlik fonksiyonunun $S(y; \Psi)$ gradyan vektörü (istatistik skoru);

$$S(y; \Psi) = \partial \log L(\Psi) / \partial(\Psi) \quad (3.35)$$

şeklinde gösterilir. Eş.(3.35)'de gösterilen gradyan matrisi kullanılarak, Ψ parametre vektörü hakkındaki $I(\Psi)$ beklenen bilgi matrisi;

$$I(\Psi) = E_{\Psi} \{S(Y; \Psi)S^T(Y; \Psi)\}, \quad (3.36)$$

şeklinde ifade edilebilir. Gözlenen bilgi matrisi diye de tanımlanan Log olabilirlik fonksiyonunun Hessian matrisi;

$$I(\Psi; y) = -\partial^2 \log L(\Psi) / \partial \Psi \partial \Psi^T \quad (3.37)$$

şeklinde gösterilir. $I(\Psi; y)$ Eş.(3.37)'de gösterildiği gibi olduğunda, düzenli şartlar altında $I(\Psi)$ beklenen bilgi matrisi;

$$I(\Psi) = E_{\Psi} \{I(\Psi; y)\} \quad (3.38)$$

şeklinde ifade edilir.

3.5.3. Maksimum olabilirlik tahmininin asimptotik kovaryans matrisi

Maksimum olabilirlik tahmininin asimptotik kovaryans matrisi, beklenen bilgi matrisi $I(\Psi)$ nin tersine eşittir ve $I(\hat{\Psi})$ ye yakınsar. $(A)_{rs}$ standart gösterim olarak, A matrisinin r inci satır, s inci sütununu gösterdiği göz önünde bulundurularak, $\hat{\Psi}_r = (\hat{\Psi})_r$ nin standart hatası;

$$SE(\hat{\Psi}_r) \approx \left(\Gamma^{-1}(\hat{\Psi}) \right)_{rr}^{1/2} \quad (r = 1, \dots, d) \quad (3.39)$$

şeklinde ifade edilir.

3.5.4. Karışımli modeller için EM algoritması

3.5.4.1. Doğrudan yaklaşım

ξ vektörü, önsel olarak bilinen $\theta_1, \dots, \theta_g$ içindeki tüm parametreleri içermek kaydıyla, karışımli modeldeki bilinmeyen tüm parametreleri kapsayan Ψ vektörü;

$$\Psi = (\pi_1, \dots, \pi_{g-1}, \xi^T)^T \quad (3.40)$$

şeklinde gösterilir.

Gözlenen rastgele örneklem; $y = (y_1^T, \dots, y_n^T)^T$ şeklinde olup, üzerinde maksimum olabilirlik için EM algoritması uygulanacak karışımli modelin parametrik gösterimi;

$$f(y_j; \Psi) = \sum_{i=1}^g \pi_i f_i(y_j, \theta_i) \quad (3.41)$$

şeklinde ifade edilir.

Ψ vektörü için log olabilirlik, gözlenen veri yardımıyla ;

$$\log L(\Psi) = \sum_{j=1}^n \log f(y_j; \Psi) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^g \pi_i f_i(y_j, \theta_i) \right\} \quad (3.42)$$

şeklinde oluşturulabilir.

Olabilirlik eşitliğinin;

$$\partial \log L(\Psi) / \partial \Psi = 0 \quad (3.43)$$

şeklinde ifade edildiği düşünüldüğünde, Eş.(3.42)'deki ifadenin Ψ vektörüne göre türevinin alınıp sıfıra eşitlenmesi gerekmektedir. EM algoritması bu çözümü iteratif metotlarla çözmeyi önermektedir.

3.5.4.2. E- aşaması

z_1, \dots, z_n ilişkili alt grup etiket vektörü mevcut olmadığından, $y = (y_1^T, \dots, y_n^T)^T$ gözlenen veri vektörü tam olmayan bir görünümüdür. Burada, her bir y_j 'in, Eşitlik (3.41) de gösterilen karışımli modelin alt gruplarından biri olan i 'nci alt gruba dahil olup olmamasına göre $z_{ij} = (z_j)_i$ 1 veya 0 değerini alarak, g – boyutlu vektör olan z_j vektörü oluşacaktır ($i = 1, \dots, g; j = 1, \dots, n$). Böylelikle

$$z = (z_1^T, \dots, z_n^T)^T \quad (3.44)$$

olmak kaydıyla, tam veri (*complete-data*)

$$y_c = (y^T, z^T)^T \quad (3.45)$$

şeklinde ifade edilir. z_1, \dots, z_n alt grup etiket vektörleri, Z_1, \dots, Z_n rastgele vektörlerinin gerçekleşen değerleri olarak alınmıştır. Tam veri için log olabilirlik;

$$\log L_c(\Psi) = \sum_{i=1}^g \sum_{j=1}^n z_{ij} \{ \log \pi_i + \log f_i(y_i; \theta_i) \} \quad (3.46)$$

şeklinde ifade edilebilir. EM algoritması, z_{ij} değişkenini eksik veri (*missing data*) olarak ele almaktadır. EM algoritması, E aşaması (*Expectation: Beklenti*) ve M aşaması (*Maximization: Maksimizasyon*) olarak iki adımdan oluşmaktadır. Probleme, gözlenen y verisi üzerinde $\log L_c(\Psi)$ tam veri log olabilirliğin koşullu beklenti değeri olarak hesaplanan z_j şeklindeki gözlemlenemeyen verilerin eklenmesi, E aşaması ile sağlanmaktadır. Ψ değişkeninin başlangıç değeri $\Psi^{(0)}$ şeklinde gösterilmektedir. EM algoritmasının ilk iterasyonunda, yani E aşamasında, verilmiş olan y gözlenen değerleri için, Ψ yerine $\Psi^{(0)}$ kullanarak, $\log L_c(\Psi)$ 'nin koşullu beklentisi;

$$Q(\Psi; \Psi^{(0)}) = E_{\Psi^{(0)}} \{ \log L_c(\Psi) | y \} \quad (3.47)$$

şeklinde hesaplanır. Beklenti operatörü olarak gösterilen E 'nin altındisi olan $\Psi^{(0)}$, bu beklentinin Ψ için $\Psi^{(0)}$ kullanarak hesaplanacağını ifade etmektedir. $k+1$ inci iterasyondaki E aşaması için hesaplanması gereken $Q(\Psi; \Psi^{(k)})$ ifadesinde, $\Psi^{(k)}$ değeri, Ψ 'nin k 'ncü EM iterasyonu sonucundaki değerini gösterir. Gözlemlenemeyen z_{ij} ile bağlantılı $\log L_c(\Psi)$ 'nin $k+1$ inci EM iterasyonu sonucundaki koşullu beklenti değeri Z_{ij} ,

$$E_{\Psi^{(k)}}(Z_{ij} | y) = pr_{\Psi^{(k)}} \{ Z_{ij} = 1 | y \} = \tau_i(y_j; \Psi^{(k)}) \quad (3.48)$$

şeklinde ifade edilir. $i=1, \dots, g$ ve $j=1, \dots, n$ olmak üzere, örneklemin j 'inci üyesini gösteren y_j gözlenen değerinin i ninci alt gruba ait olma sonsal(*posterior*) olasılığı;

$$\tau_i(y_j; \Psi^{(k)}) = \frac{\pi_i^{(k)} f_i(y_j; \theta_i^{(k)})}{f(y_j; \Psi^{(k)})} = \frac{\pi_i^{(k)} f_i(y_j; \theta_i^{(k)})}{\sum_{h=1}^g \pi_h^{(k)} f_h(y_j; \theta_h^{(k)})} \quad (3.49)$$

şeklinde gösterilir. Verilen bir y vektörü için, Eş.(3.48)'i, Eş.(3.46)'de kullanarak, koşullu beklenti fonksiyonu;

$$Q(\Psi; \Psi^{(k)}) = \sum_{i=1}^g \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) \{ \log \pi_i + \log f_i(y_i; \theta_i) \} \quad (3.50)$$

şeklinde hesaplanır.

3.5.4.3. M- aşaması

M aşaması, Ω parametre uzayındaki Ψ vektörüne göre, $Q(\Psi; \Psi^{(k)})$ 'nin k+1 inci iterasyonu sonucunda global maksimumlarını bulmayı içermektedir. Sonlu karışım modelleri için, π_i karışım oranlarının güncellenmiş $\pi_i^{(k+1)}$ tahminleri, alt gruplardaki bilinmeyen parametreleri içeren ξ parametre vektörünün güncellenmiş ξ_i tahminlerinden bağımsız olarak hesaplanmaktadır.

Eğer, z_{ij} gözlemlenebiliyorsa, π_i nin en çok olabilirlik tahmini

$$\hat{\pi}_i = \sum_{j=1}^n z_{ij} / n \quad (i = 1, \dots, g). \quad (3.51)$$

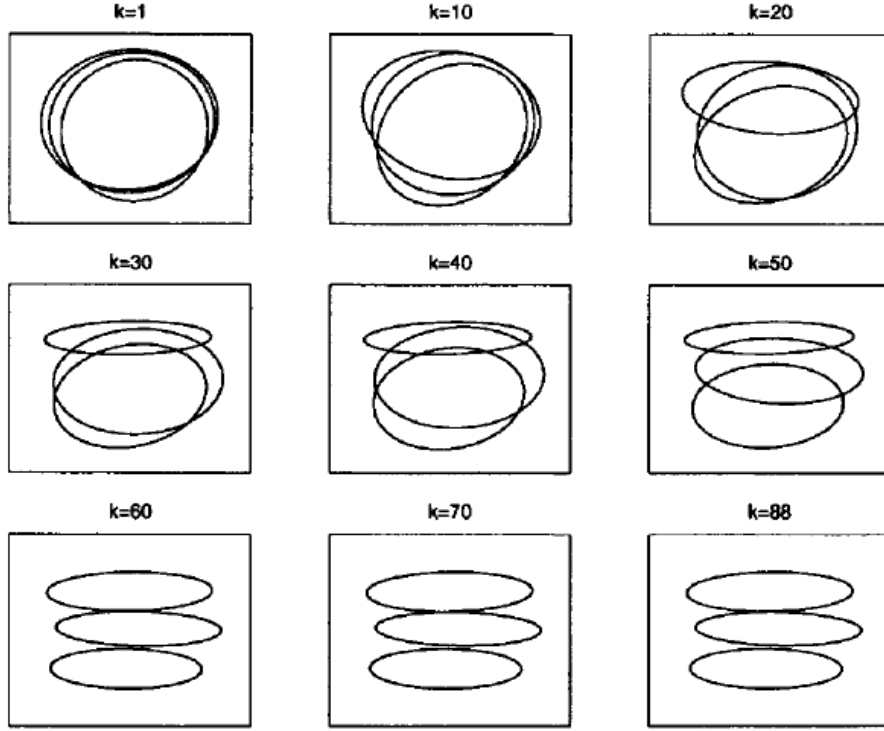
şeklinde ifade edilebilir. E- aşamasındaki her bir z_{ij} 'nin, o anki koşullu beklentisi olan $\tau_i(y_j; \Psi^{(k)})$ ile yer değiştirmesi gibi, Eş.(3.51)'de verilen z_{ij} yerine $\tau_i(y_j; \Psi^{(k)})$ konularak, π_i 'nin güncel tahmini;

$$\pi_i^{(k+1)} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) / n \quad (i = 1, \dots, g). \quad (3.52)$$

şeklinde hesaplanır.

3.5.5. Model Seçiminde Uyum Ölçütleri

Karışımli modellerde, her karışma bileşeni, gizli bir alt grup değişkeni gibi düşünülebileceğinden uygun alt sayısının seçimi ayrı bir problem olarak ele alınmaktadır. Alt grup sayısını belirlerken, tahmini model parametrelerinin değerleri alınır. Bu amaçla, en çok olabilirlik tahminlerini hesaplamak için EM algoritması kullanılır. Farklı alt grup sayısı gösterimleri Şekil 3.10'da verilmiştir.



Şekil 3.10. 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).

3.5.6. Log olabilirliğin sapması

Model seçiminde, oluşturulan (fitted) modelin gerçek verilerine ulaşılması amacıyla Kullback-Leibler (1951) yaklaşımı kullanılabilir. $f(\omega)$; gerçek yoğunluğu ifade ederse, $f(\omega)$ 'nin $f(\omega; \hat{\Psi})$ tahminine göre Kullback-Leibler değeri;

$$I\{f(\omega); f(\omega; \hat{\Psi})\} = \int f(\omega) \log f(\omega) d\omega - \int f(\omega) \log f(\omega; \hat{\Psi}) d\omega \quad (3.53)$$

şeklinde ifade edilir. Bu ifade, $f(\omega; \hat{\Psi})$ 'ye göre $f(\omega)$ 'nin sapmasının bir ölçüsü olarak verilmektedir. Burada amaç, Eş.(3.53)'deki Kullback-Leibler ifadesini küçük yapmaya çalışmaktır. Eş.(3.53)'de verilen denklemin sağ tarafındaki ilk terim modele bağlı değildir, yalnızca ikinci terim modele ilişkin bir terim olup, F gerçek dağılımı göstermek ve $y = \left(y_1^T, \dots, y_n^T \right)^T$ gözlenen verileri içermek şartıyla, bu ikinci terim;

$$\begin{aligned} \eta(y; F) &= \int f(\omega) \log f(\omega; \hat{\Psi}) d\omega \\ &= \int \log f(\omega; \hat{\Psi}) dF(\omega) \end{aligned} \quad (3.54)$$

şeklinde genişletilebilir. $\eta(y; F)$ 'nin basit bir tahmini, Eş.(3.54)'teki ifadede F yerine \hat{F} deneysel dağılım fonksiyonu bırakılarak;

$$\begin{aligned}\eta(y; \hat{F}_n) &= \frac{1}{n} \sum_{j=1}^n \log f(y_j; \hat{\Psi}) \\ &= \frac{1}{n} \log L(\hat{\Psi})\end{aligned}\quad (3.55)$$

şeklinde ifade edilir ve her bir $y_j (j=1, \dots, n)$ gözleminde ağırlıkça $1/n$ kadar bulunur.

Buradan beklenen log yoğunluğunun aşırı tahmini olarak;

$$\int \log f(\omega) dF(\omega) \quad (3.56)$$

elde edilir. Genellikle \hat{F}_n deneysel dağılım fonksiyonu, gerçekteki F dağılımına, $F_{\hat{\Psi}}$ ile ifade edilen oluşturulmuş dağılım fonksiyonunun yakınsadığından daha fazla yakınsar. Bağımsız Y_1, \dots, Y_n 'nin ortak dağılım fonksiyonu olarak F kullanıldığında beklenti değeri E_F olmak şartıyla, Eş.(3.56)'nın bir tahmin edicisi olarak $\eta(y; \hat{F}_n)$ 'nin sapması;

$$\begin{aligned}b(F) &= E_F \left\{ \eta(Y; \hat{F}_n) - \eta(Y; F) \right\} \\ &= E_F \left\{ \frac{1}{n} \sum_{j=1}^n \log f(y_j; \hat{\Psi}) - \int \log f(\omega; \hat{\Psi}) dF(\omega) \right\}\end{aligned}\quad (3.57)$$

olarak ifade edilir.

$b(F)$ sapma teriminin uygun tahmini yapılarak, sapması düzeltilmiş log-olabilirlik;

$$\log L(\hat{\Psi}) - b(F) \quad (3.58)$$

şeklinde ifade edilir ve bu fonksiyon, model seçiminde bilgi ölçütü olarak kullanılabilir. Burada amaç, seçilen modelde mevcut bileşen sayısını en üst düzeye çıkarmak (Eş. (3.58)) ve böylece Kullback-Leibler bilgisini (Eş. (3.53)) en aza indirmektir.

Literatürde, bu şekilde oluşturulan bilgi ölçütleri, genellikle bu farkın negatif değerinin iki katı olacak şekilde, yani;

$$-2 \log L(\hat{\Psi}) + 2C \quad (3.59)$$

şeklinde gösterilir. Eş.(3.59)'in sağ tarafındaki ilk terim uyum eksikliğini (*lack of fit*) ölçmekte, ikinci terimdeki C ise modelin karmaşıklığını ölçen ceza (*penalty*) terimini göstermektedir. Buradaki amaç, kriteri en aza indirgeyen bir modelin seçilebilmesidir.

3.5.7. Akaike bilgi ölçütü

Akaike çalışmalarında (1973, 1974) $b(F)$ 'nin asimptotik olarak d 'ye eşit olduğunu göstermiştir; burada d , modeldeki toplam parametre sayısına eşittir. Akaike bilgi ölçütü (Akaike's Information Criterion - AIC),

$$-2\log L(\hat{\Psi}) + 2d \quad (3.60)$$

değerini en aza indirebilen modeli seçer. Akaike'den sonra, AIC'nin bir karışım içindeki alt grup sayısını seçmek için kullanımı konusunda, yeni yaklaşımlar içeren çalışmalar yapılmıştır (Bozdoğan ve Sclove, 1984; Sclove, 1987).

Konishi ve Kitagawa (1996), gerçek yoğunluğun $f(y_j)$, öngörülen parametrik aileye ait olmadığı ve parametre vektörünün zorunlu olarak en çok olabilirliğe göre tahmin edilmediği durumlarda, ilgili asimptotik sapmayı türetmişlerdir. Bununla birlikte bu asimptotik açılımların $b(F)$ için geçerliliği, normal asimptotik teorisi için gereken tekdüzelik (*regularity*) koşullarına bağlıdır. Bu şekliyle, Titterington ve ark. (1985)'da değinilen LRTS(Likelihood Ratio Test Statistic)'nin boş dağılımı geçerli kılınmış olur. Bu durumda, LRTS, bir karışım modelindeki alt grup sayısı üzerinde yapılan testler için işlevsiz hale gelmektedir.

Buna rağmen, AIC kriteri, karışımli modellerin düzenini değerlendirmek için sıklıkla kullanılmaktadır. Koehler ve Murphee (1988), AIC'nin düzeninin tutarsız olduğunu ve modellere uyma eğiliminde olduğunu gözlemlemiştir. Karışım bağlamında, AIC'nin doğru sayıda alt grubu, olduğundan fazla hesaplama eğiliminde olduğu bildirilmiştir. (Soromenho, 1993; Celeux ve Soromenho, 1996).

3.5.8. Bilgi karmaşıklığı kriteri

Bozdoğan, AIC'nin performansını iyileştirmek için bilgi karmaşıklığı (*Informational Complexity-ICOMP*) kriterini önermiştir (Bozdoğan, 1990; Bozdoğan, 1993). Bu kriter, modeldeki parametre sayısı d olmak kaydıyla, Eş.(3.61) ile aşağıdaki gibi ifade edilebilir.

$$ICOMP(g) = -2 \log L(\hat{\Psi}) + C_1 + C_2 \quad (3.61)$$

Eş.(3.61)'deki C_1 ;

$$C_1 = d \log \{d^{-1} tr I^{-1}(\hat{\Psi})\} \quad (3.62)$$

şeklinde ifade edilir ve C_2 ise;

$$C_1 = d \log \{d^{-1} tr I^{-1}(\hat{\Psi})\} \quad (3.63)$$

şeklinde gösterilebilir. Bu kriterde, yeni bir parametre oluşturma (*reparameterization*) sırasında modelin değişken olması, sorun olarak görülen bir konudur. C_1 ve C_2 ifadelerindeki, örnekleme beklenen değeri olan $I(\hat{\Psi})$ 'yi, tek bir gözlemden beklenen değeri ile değiştirebiliriz. Bu işlem, $d \log n$ 'in iki terimden de çıkarılması ile sonuçlanacaktır. Bu kriteri pratikte uygulayabilmek için, beklenen bilgi matrisi olan $I(\hat{\Psi})$ 'nin tahmin edilmesi gerekmektedir. Bozdoğan (1993), bu eşitliği bir karışım model içerisinde uygulamanın zor olduğunu düşünerek, sınıflandırılmış bir örneklem için (tahmini) beklenen bilgi matrisini kullanmak yerine bu matrise yaklaşma yönünde çalışmıştır. g -alt gruplu normal karışımli model için, $d = gp + \frac{1}{2}gp(p+1)$ olmak şartıyla, C_1 ve C_2 'nin tahmin edilebilmesinde, aşağıdaki iki denklem kullanılabilir.

$$C_1 = d \log \left[d^{-1} \sum_{i=1}^g \left\{ \pi_i^{-1} tr \hat{\Sigma}_i + \frac{1}{2} tr(\hat{\Sigma}_i \hat{\Sigma}_i) + \frac{1}{2} (tr \hat{\Sigma}_i)^2 + \sum_{v=1}^p (\hat{\Sigma}_i)_{vv}^2 \right\} \right] \quad (3.64)$$

ve

$$C_2 = (p+2) \sum_{i=1}^g \log |\hat{\Sigma}_i| - p \sum_{i=1}^g \log(n \hat{\pi}_i) + gp \log(2n)$$

Normal karışımli modeller ile gerçekleştirdikleri simülasyonlarda, Celeux ve Soromenho (1996) tarafından, alt grupların kovaryans matrislerinin çok farklı olduğu durumlarda, ICOMP'nin, karışımli model alt grup sayısını, AIC'den daha belirgin bir

şekilde, olduğundan daha fazla tahmin etme(*overestimate*) eğiliminde olabileceği belirtilmiştir.

3.5.9. Bayesian Tabanlı Uyum Ölçütleri

3.5.10. Bayesian yaklaşım

Model seçimi için bir Bayesian çerçevesi içerisinde türetilmiş bazı kriterler mevcuttur. Fakat bu yaklaşımlar, Bayesian olmayan bir çerçevede de uygulanabilir. Bu nedenle bu yaklaşımlar, Bayesian olan veya başka bir bakış açısıyla ele alınan karışımli modellerde, alt grup sayısının seçiminde tercih edilmektedir. Bu başlık altında, yalnızca Bayesian çerçevede uygulanan bazı kriterler açıklanmıştır.

Bayesian temelli bilgi kriterleri konusunda ilk olarak, Schwarz (1978) tarafından, Bayesian bilgi kriterinin (BIC) kullanımında entegre olabilirliğe göre bir yaklaşım sunulmuştur. Bu yaklaşımın genel teorik gerekçeleri, bileşen sayısının ortaya çıkarılması için, frekanslara dayalı bir yaklaşımla, veri setini benzer düzenliliğin olduğu gruplara ayırma esasına dayanır.

3.5.11. Bayesian bilgi kriteri

$$\log p(y) = \log L(\hat{\Psi}) + \log p(\hat{\Psi}) - \frac{1}{2} \log |I(\hat{\Psi}; y)| + \frac{1}{2} d \log(2\pi) \quad (3.65)$$

Schwarz (1978)'in önerdiği, Eş.(3.65)'de $O(1)$ teriminin ihmal edilmesi ile;

$$|I(\hat{\Psi}; y)| = O(n^d) \quad (3.66)$$

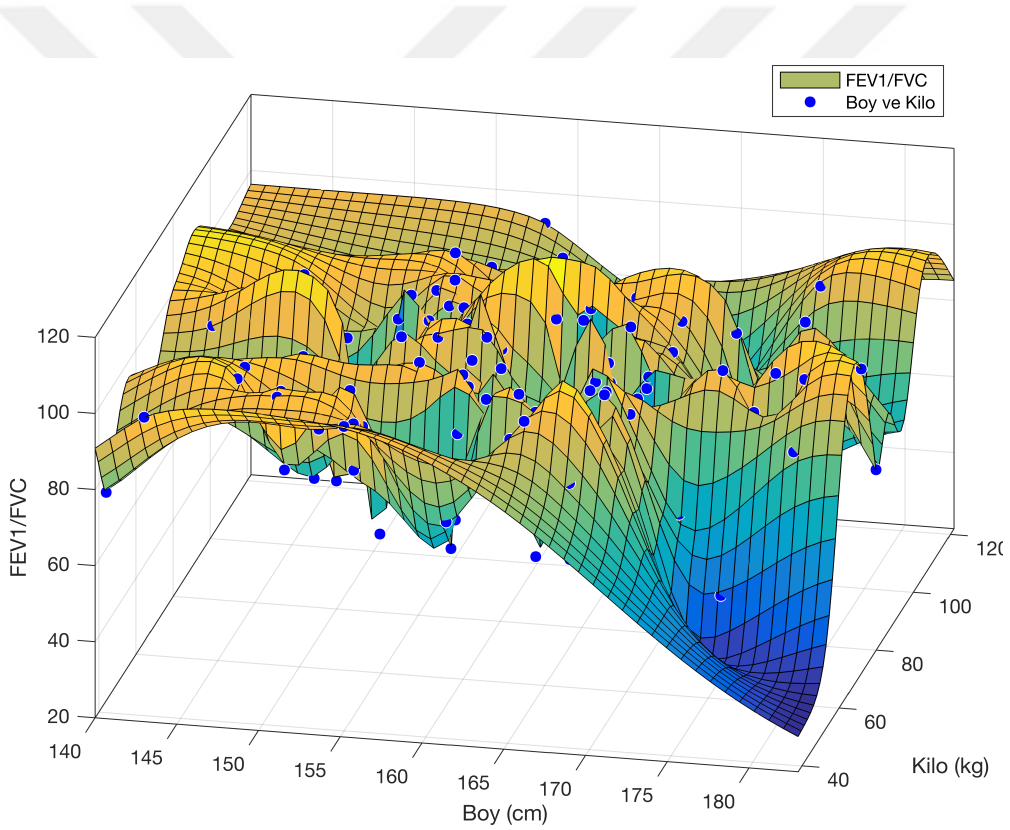
olmak şartıyla, karışımli modelde, g alt grupların sayısı için, negatif log olabilirlik cezası(*penalty*) iki ile çarpım olacak şekilde, model seçiminde en küçük olarak elde edilecek Bayesian bilgi kriteri (BIC);

$$-2 \log L(\Psi) + d \log n \quad (3.67)$$

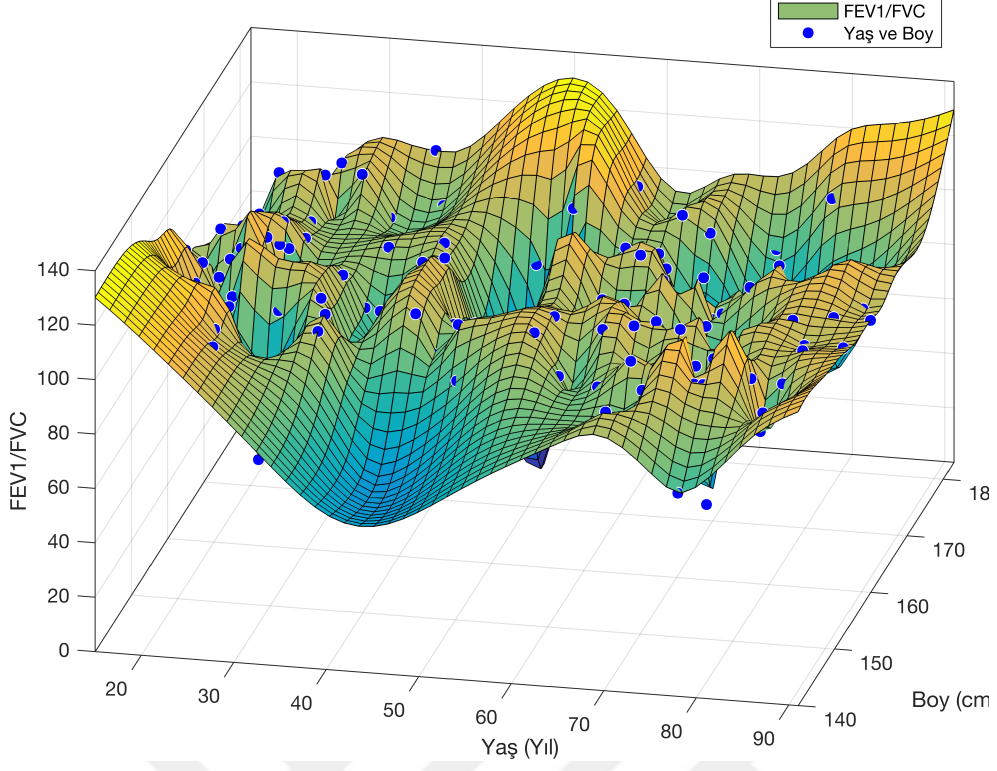
şeklinde hesaplanır.

4. BULGULAR

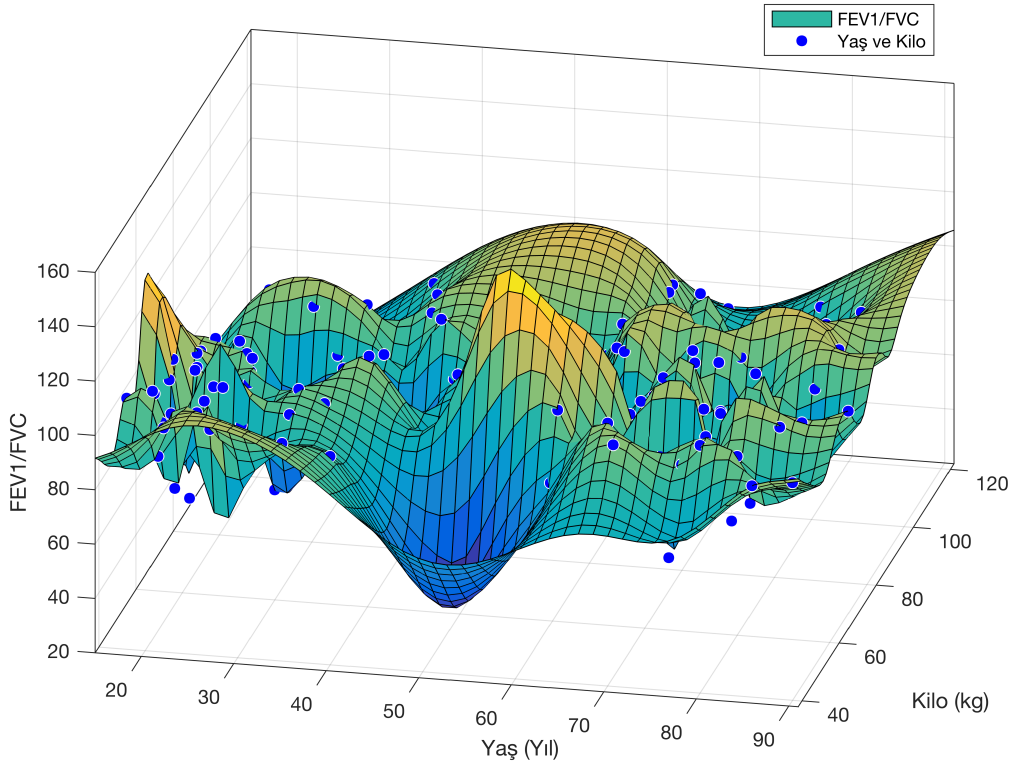
Bu çalışmada, akciğer kapasitesinin bir göstergesi olan FEV1/FVC oranının solunum fonksiyon testi (SFT) üzerine hastaların boy, kilo, yaş ve cinsiyetin etkilerini modellemek için Gaussian karışımli regresyon yöntemi kullanılmıştır. İlk olarak, bağımsız değişkenlerin; boy – kilo, yaş – boy ve yaş – kilo değişken ikililerine göre FEV1/FVC değerleri üzerinde uygulanan biharmonik enterpolasyonla oluşturulan yüzey grafikleri sırasıyla Şekil 4.1, Şekil 4.2 ve Şekil 4.3'te gösterilmiştir. Şekillere bakıldığında SFT sonuçlarının, bağımsız değişkenler karşısında heterojen bir yapı gösterdiği görülmektedir.



Şekil 4.1. Boy ve kilo bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşın değişim grafiği.



Şekil 4.2. Yaş ve boy bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşın değişim grafiği.



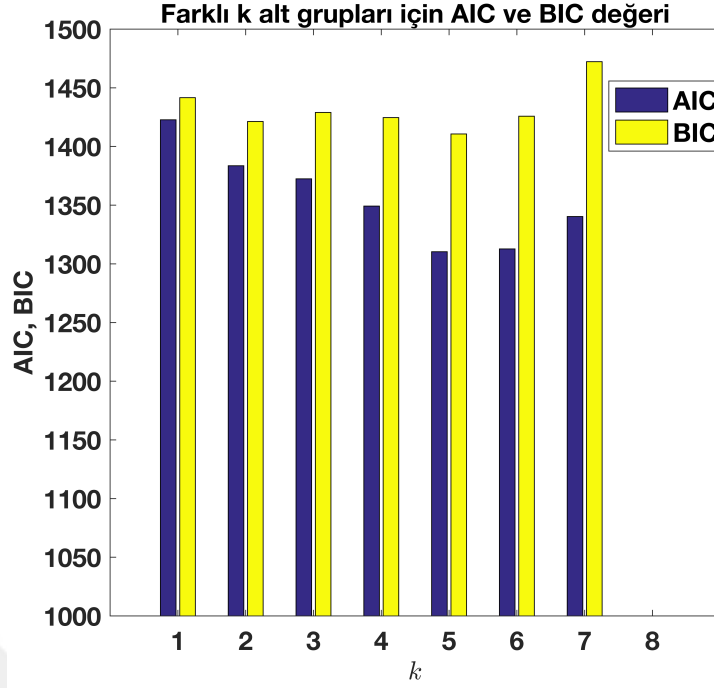
Şekil 4.3. Yaş ve kilo bağımsız değişkenlerinin FEV1/FVC bağımlı değişkenine karşın değişim grafiği.

Çalışmada, SFT sonuçlarından FEV1/FVC bağımlı değişkeni üzerine kilo, boy, yaş ve cinsiyet değişkenlerinin etkilerine bağlı olarak veri setinin nasıl bir heterojen yapı sergilediğini saptamak için Gaussian karışımli regresyon uygulaması sonucu, veri setinin kendi içerisinde homojen olmak üzere kaç alt gruba ayrıldığını belirlemek için kullanılan AIC ve BIC uyum ölçütleri ile entropy doğru sınıflandırma oranlarına ilişkin istatistikler Çizelge 4.1’de verilmiştir.

Çizelge 4.1’de, 7 alt gruplu modele kadar Gaussian karışımli modele ilişkin AIC ve BIC uyum ölçütleri ile entropy doğru sınıflandırma oranları verilmiştir. En yüksek AIC ve BIC uyum ölçütleri 1 alt gruplu modelde sırasıyla 1422.759 ve 1441.609; en küçük ise beş alt gruplu modelde sırasıyla 1310.366 ve 1491.616 olarak elde edilmişlerdir. Böylece, AIC ve BIC bilgi ölçütlerine göre veri setinin kilo, boy, yaş ve cinsiyet özellikleri bakımından kendi içlerinde homojen 5 alt gruptan oluşmaktadır. AIC ve BIC ait değerlerin, 5 alt gruplu modele kadar azaldığı daha sonra ise arttığı gözlenmiştir (Şekil 4.4, Çizelge 4.1). Bu nedenle 5 alt gruplu modelden sonra 6 ve 7 alt gruplu modellerden sonra AIC ve BIC değerlerine yer verilmemiştir. Bu sonuçlara göre, hastalar, SFT sonuçlarının bağımsız değişkenleri olarak boy, yaş, kilo ve cinsiyet değişkenlerine göre homojen 5 alt gruba ayrılmışlardır. Sınıflandırma kalitesinin göstergesi olarak entropi değeri en yüksek 5 alt gruplu modelde %92.5 olarak bulunmuştur. Bununla birlikte en düşük entropy oranı 2 alt gruplu modelde %84.5 olarak elde edilmiştir.

Çizelge 4.1. Solunum fonksiyon testi sonuçlarına göre homojen alt gruplar için model bilgi ölçütleri.

Alt gruplar	Model bilgi ölçütleri		
	AIC	BIC	Entropi (%)
Bir alt gruplu model	1422.759	1441.609	87.0
İki alt gruplu model	1383.531	1421.231	84.5
Üç alt gruplu model	1372.400	1428.950	89.3
Dört alt gruplu model	1349.211	1424.611	90.8
Beş alt gruplu model	1310.366	1410.616	92.5
Altı alt gruplu model	1312.757	1425.857	91.9
Yedi alt gruplu model	1340.366	1472.316	86.2



Şekil 4.4. SFT sonuçlarına göre farklı alt grup sayılarına göre AIC ve BIC değerleri.

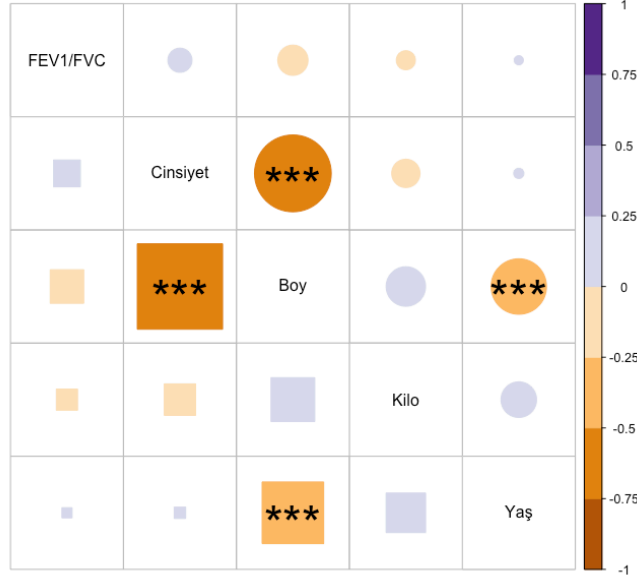
Normal karışımli regresyon modeli kullanılarak elde edilen 5 alt gruplu modelde, her bir alt gruba giren bireylerin sayısı ve karışma olasılıkları (mixing probability) sırasıyla 11(%6.43), 18(%10.53), 49(%28.65), 15(%8.77) ve 78(%45,61) olarak elde edilmiştir. En fazla bireyin alt grup 5 te, en az bireyin ise alt grup 1 de olduğu belirlenmiştir (Çizelge 4.2).

Çizelge 4.2. En yakın sınıf üyelik derecelerine sahip alt gruplar için elde edilen sınıf büyüklükleri ve karışma olasılıkları.

Altgruplar	N	Oran
1	11	0.0643
2	18	0.1053
3	49	0.2865
4	15	0.0877
5	78	0.4561

Çalışmada kullanılan veri setindeki; FEV1/FVC, cinsiyet, boy, kilo ve yaş gibi değişkenler arasındaki korelasyon matrisi Şekil 4.5'te verilmiştir. Boy ve yaş değişkenleri arasında doğrusal ve negatif yönlü bir ilişki saptanmıştır. Bu ilişkinin

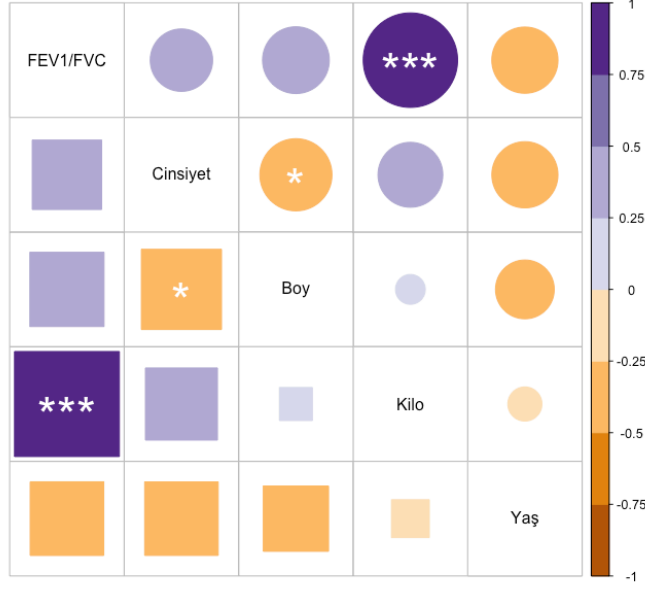
derecesi ise -0.297 olup, istatistiksel olarak önemli bulunmuştur ($p < 0.001$). Dolayısıyla, örneklemden deneklerin yaşları arttıkça boylarının kısa olduğu söylenebilir. Bağımlı FEV1/FVC değişkeninin kilo ve boy ile negatif korelasyon gösterdiği, yaş ve cinsiyet değişkeninde kadın kategorisi ile pozitif korelasyon gösterdiği saptanmış ve aradaki bu ilişkiler istatistiksel olarak önemli bulunmamıştır ($p > 0.05$) (Şekil 4.5).



Şekil 4.5. Veri setindeki değişkenler arası korelasyon matrisi grafiği.

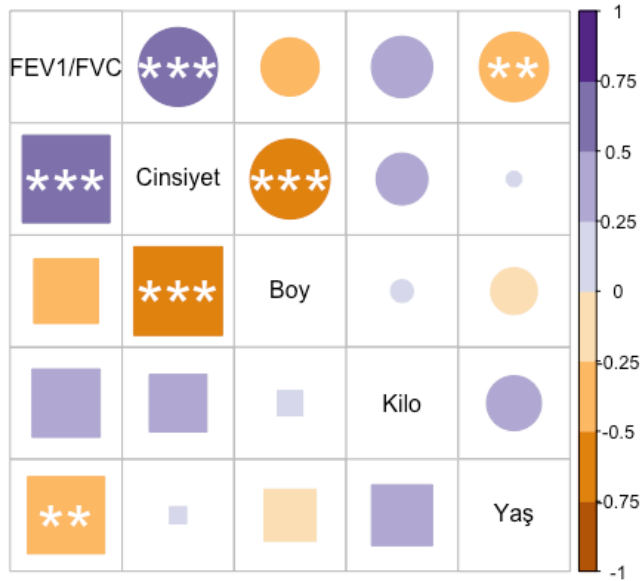
Gaussian karışımli regresyon sonucunda, en iyi model olarak seçilen 5 alt gruplu modelde; bağımlı ve bağımsız değişkenler arasındaki korelasyonlar sırasıyla birinci alt grup için Şekil 4.6, ikinci alt grup için Şekil 4.7, üçüncü alt grup için Şekil 4.8, dördüncü alt grup için Şekil 4.9 ve son olarak beşinci alt grup için Şekil 4.10'da verilmiştir.

Birinci alt grubun değişkenleri arasındaki korelasyon matrisi Şekil 4.6'da gösterilmiştir. Birinci alt gruptaki korelasyon matrisinde en çok göze çarpan ilişki, bağımlı değişken (FEV1/FVC) ile kilo değişkeni arasında bulunan pozitif yönlü ve istatistiksel olarak anlamlı bulunan ilişki olduğu saptanmıştır ($r = -0.840$, $p < 0.001$).



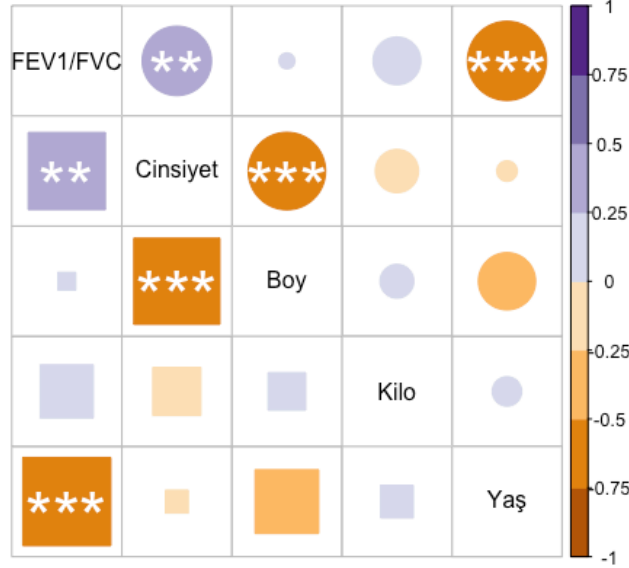
Şekil 4.6. Birinci alt grupta değişkenler arası korelasyon matrisi.

İkinci alt grup için bağımlı ve bağımsız değişkenler arasındaki korelasyonlar Şekil 4.7’ de gösterilmiştir. Bu grupta, yaş değişkeni ile bağımlı değişken arasında negatif bir ilişki olup, bu ilişki istatistiksel olarak anlamlı bulunmuştur ($r=-0.472$, $p<0.01$). Bununla birlikte, FEV1/FVC ile boy ve kilo arasındaki ilişki katsayıları sırasıyla 0.331 ve 0.369 olup, her iki ilişki düzeyi de istatistiksel olarak önemli bulunmamıştır ($p>0.05$).



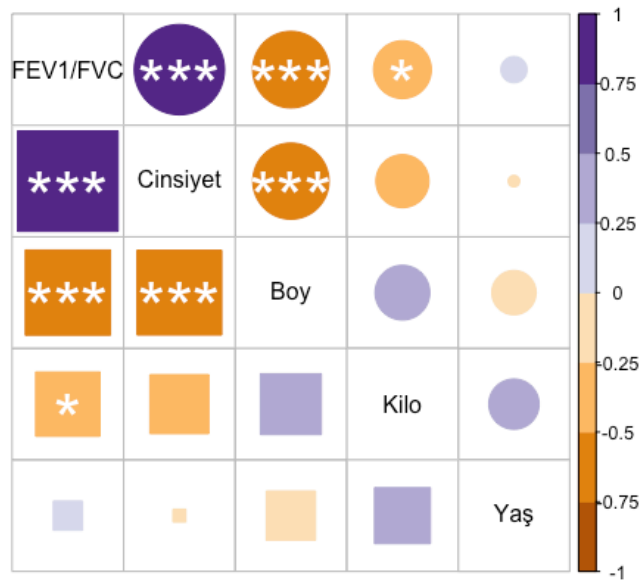
Şekil 4.7. İkinci alt grupta değişkenler arası korelasyon matrisi.

Üçüncü alt gruptaki bağımlı ve bağımsız değişkenler arasındaki korelasyonun grafiği Şekil 4.8’ de gösterilmiştir. FEV1/FVC ile boy ve kilo arasındaki ilişki katsayıları sırasıyla 0.026 ($p>0.05$) ve 0.369 ($p<0.05$) olarak elde edilmiştir.



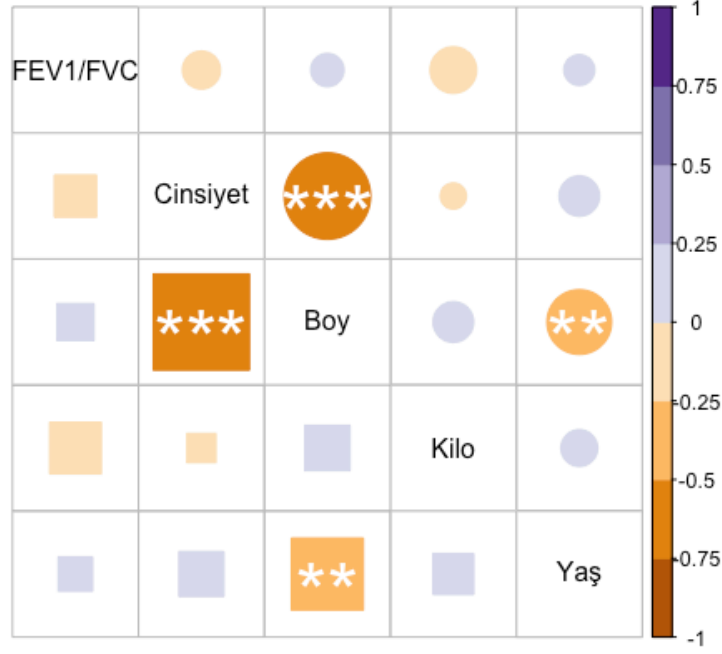
Şekil 4.8. Üçüncü alt grupta değişkenler arası korelasyon matrisi.

Dördüncü alt grup için boy, kilo, yaş ve FEV1/FVC değişkenleri arasındaki korelasyonlar matrisi Şekil 4.9’da gösterilmiştir. Dördüncü alt grupta, diğer gruplarda görülen durumlardan farklı olarak Boy değişkeni ile FEV1/FVC arasındaki doğrusal ilişki istatistiksel olarak çok önemli bulunmuştur ($r=0.583$, $p<0.001$).



Şekil 4.9. Dördüncü alt grupta değişkenler arası korelasyon matrisi.

Beşinci alt grup için oluşan korelasyon matrisi Şekil 4.10'da gösterilmiştir. Bu alt grupta, FEV1/FVC bağımlı değişkeni ile tüm bağımsız değişkenler arasındaki ilişki katsayıları istatistiksel olarak önemsiz bulunmuştur ($p>0.05$).



Şekil 4.10. Beşinci alt grupta değişkenler arası korelasyon matrisi.

Veri setinin dağıldığı en iyi model olan 5 alt gruplu modelde, parametre tahminleri Çizelge 4.3'te verilmiştir. Birinci alt grupta FEV1/FVC bağımlı değişkeni üzerine kilo bağımsız değişkeninin etkisi istatistiksel olarak önemli ($p<0.01$); boy ve yaş bağımsız değişkenlerin etkileri ise istatistiksel olarak önemsiz oldukları belirlenmiştir ($p>0.05$). İkinci alt grupta, FEV1/FVC bağımlı değişkeni üzerine kilo ve yaş bağımsız değişkenlerin etkileri istatistiksel olarak önemli ($p<0.01$); boy bağımsız değişkeninin etkisi istatistiksel olarak önemsiz olduğu saptanmıştır ($p>0.05$). Üçüncü alt grupta, FEV1/FVC bağımlı değişkeni üzerine boy bağımsız değişkeninin katsayısı istatistiksel olarak önemsiz bulunmuşken ($p>0.05$); buna karşın yaş ($p<0.05$) ve kilo ($p<0.01$) bağımsız değişkenlerin etkileri ise istatistiksel olarak önemli oldukları saptanmıştır. Dördüncü alt grupta FEV1/FVC bağımlı değişkeni üzerinde kilo, boy ve yaş bağımsız değişkenlerin etkileri istatistiksel olarak önemsiz oldukları belirlenmiştir ($p>0.05$). Son olarak beşinci alt grupta, FEV1/FVC bağımlı değişkeni üzerine kilo

bağımsız değişkeninin etkisi istatistiksel olarak önemli ($p < 0.01$); boy ve yaş bağımsız değişkenlerin etkileri ise istatistiksel olarak önemsiz oldukları belirlenmiştir ($p > 0.05$).

Çizelge 4.3. Alt gruplara ait parametre tahminleri ve standart hataları.

Alt gr.	Kesim Noktası β_0	FEV1/FVC ortalaması	Cinsiyet	Boy (cm)	Kilo (kg)	Yaş
1	-101.056	50.071	8.153 (7.746)	0.626 (0.552)	0.445 (0.129)**	-0.006 (0.179)
2	101.930	67.034	3.984 (3.300)	-0.295 (0.166)	0.226 (0.053)**	-0.226 (0.035)**
3	42.738	82.156	9.144 (4.571)*	0.137 (0.230)	0.173 (0.053)**	-0.188 (0.074)*
4	95.608	93.592	7.740 (2.207)*	-0.074 (0.107)	-0.036 (0.026)	0.019 (0.025)
5	95.067	98.466	-0.467 (0.558)	-0.037 (0.028)	-0.040 (0.018)*	0.019 (0.010)

Gaussian karışımli regresyon parametreleri olarak kullanılan, FEV1/FVC bağımlı değişkeni ve cinsiyet, boy, kilo ve yaş bağımsız değişkenlerin alt gruplara elde edilen ortalama değerleri Çizelge 4.4'te verilmiştir.

Çizelge 4.4'de, FEV1/FVC bağımlı değişkenin; en yüksek ortalaması beşinci alt grupta (98.466), daha sonra sırasıyla dördüncü alt grupta (93.592), üçüncü alt grup (82.156), ikinci alt grupta (67.034) ve birinci alt grupta (50.071) olarak elde edilmiştir. Boy bağımsız değişkeni açısından alt gruplar sırasıyla; 164.180, 164.800, 163.816, 162.667 ve 162.798 olarak elde edilmiştir.

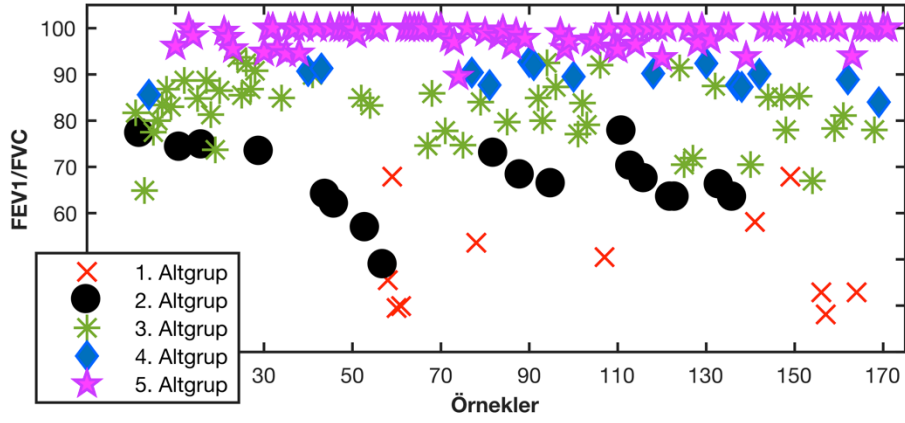
Kilo bağımsız değişkeni 5 alt popülasyona dağılımı sırasıyla; 78.862, 74.775, 70.099, 75.295 ve 71.341 olarak elde edilmiştir. 5 ayrı alt grubun yaş ortalamalarına bakıldığında, en düşük yaş ortalaması 42.548 ile ikinci alt grupta gözlenirken, en yüksek 55.521 ile dördüncü alt gruptur. Yaş değişkeni bakımından; üçüncü, dördüncü ve beşinci alt gruplar benzerlik göstermişlerdir.

FEV1/FVC bağımlı değişkenin 5 alt gruba göre dağılımı Şekil 4.11'de verilmiştir. FEV1/FVC bağımlı değişkenin alt gruplara dağılımı sırasıyla; 50.071, 67.034, 82.156, 93.592 ve 98.466 olarak elde edilmiştir. FEV1/FVC bağımlı değişkenin alt gruplara göre gittikçe büyüdüğü belirlenmiştir.

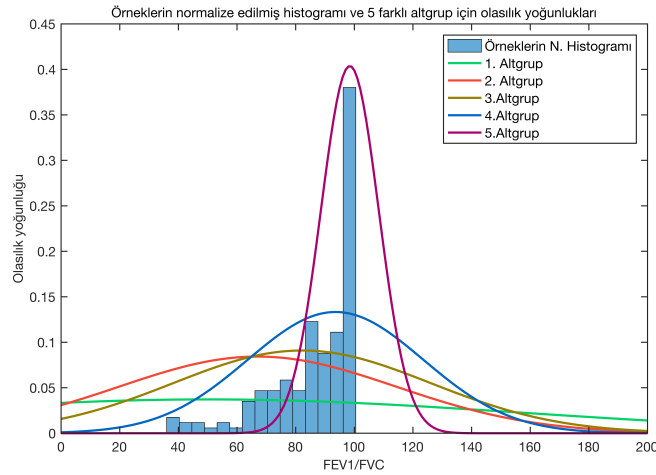
Çizelge 4.4.'te her bir gruba dahil olan bireylerin cinsiyet dağılımları sayı ve yüzde (%) olarak gösterilmiştir. Birinci alt gruba dahil olan bireylerin %36.4'ü erkek, %63.6'sı ise kadındır. Yüzde olarak dördüncü grup hariç, diğer tüm gruplarda kadınlar %60'ın üzerinde bir dağılım göstermiştir. Dördüncü grubun tamamını erkekler oluşturmuştur.

Çizelge 4.4. Değişkenlerin alt gruplara göre ortalamaları.

Alt gruplar	FEV1/FVC ortalaması	Cinsiyet (Sayı, %)		Boy (cm)	Kilo (kg)	Yaş
		Erkek	Kadın			
1	50.071 (3.238)	4 (%36.4)	7 (%63.6)	164.180 (2.717)	78.862 (4.540)	47.643 (6.859)
2	67.034 (1.725)	7 (%38.9)	11 (%61.1)	164.800 (1.806)	74.775 (3.713)	42.548 (5.036)
3	82.156 (1.329)	17 (%34.7)	32 (%65.3)	163.816 (1.322)	70.099 (2.167)	54.843 (3.009)
4	93.592 (1.041)	15 (%100)	0 (%0)	162.667 (1.610)	75.295 (3.181)	55.521 (4.084)
5	98.466 (0.303)	30 (%38.5)	48 (%61.5)	162.798 (1.229)	71.341 (1.970)	53.221 (3.014)



Şekil 4.11. Bağımlı değişkenin beş alt gruba göre dağılım grafiği.



Şekil 4.12. Gaussian karışımı regresyona göre elde edilen 5 alt grubun olasılık dağılımı yoğunluklarının dağılımı.

Elde edilen karışimli regresyon parametrelerine göre oluşturulan 5 farklı alt grubun normal dağılım yoğunluk fonksiyonları Şekil 4.12’de verilmiştir. En yakın sınıf üyelik (olasılık) derecelerine sahip alt gruplar için elde edilen sınıf büyüklükleri ile beraber dikkate alındığında, beşinci alt grubun karışım oranının %45.61 olarak en yüksek karışım oranına sahip olduğu saptanmıştır.





5. TARTIŞMA VE SONUÇ

Şekil 4.1, Şekil 4.2 ve Şekil 4.3'te verilen boy ile kilo, yaş ile boy ve yaş ile kilo değişken ikililerine göre FEV1/FVC değerleri üzerinde uygulanan biharmonik enterpolasyonla oluşturulan yüzey grafiklerinde, heterojen bir yapı ortaya çıktığı saptanmıştır. Bu durumda, karışımli model yaklaşımı kullanılarak heterojen yapı sergileyen veri seti için kendi içinde homojen alt gruplar oluşturularak, söz konusu heterojenlik giderilmektedir (Verbeke ve Lesaffre, 1996; Lubke ve ark., 2005). Bilindiği gibi çok değişkenli istatistiksel analizlerde, büyük veri setini, homojen alt gruplara ayırarak, bir nevi boyut indirgeme amaçlanmaktadır (Liland, 2011).

Solunum fonksiyon testi, obstrüktif akciğer hastalıklarında (astım, KOAH gibi), restriktif akciğer hastalıklarında (parankim hastalıkları, solunum kaslarının hastalıkları, toraks duvarı hastalıkları, akciğer ödemi, konjestif kalp yetersizliği gibi) ve toraks dışı solunum yolu darlıklarında (trakea obstrüksiyonu, ses teli paralizisi gibi) tanısaldır (Miller ve ark., 2005).

FEV1/FVC değeri, solunum sistemi hastalıklarının tanısında kullanılan, SFT (solunum fonksiyon testi) içinde ölçülen bir parametredir. FEV1/FVC, özellikle akciğerin obstrüktif ve restriktif hastalıklarının ayırımında kullanılan en önemli SFT parametresidir (Minelli, 1992). Obstrüktif akciğer hastalıklarında FEV1/FVC değerinin %70'in altına düşmesi beklenirken, restriktif akciğer hastalıklarında ise normal seyretmesi ya da %75'in üzerine çıkması beklenir (Sahebjami ve Gartside, 1996; Eisner ve ark., 2007; Nathell ve ark., 2007; Swanney, 2008). Koah tanısı için FEV1/FVC değerinin %70'in altında olması beklenmektedir (Eisner ve ark., 2007). Obstrüktif bozukluklarda akciğerden atılan maksimum hava akımında, maksimum hacimle orantısız bir azalma görülmektedir. Bu durum soluk verme sırasında hava yollarında daralmayı temsil eder ve FEV1/FVC oranının kaynak değerlerin %70'in altında olması ile tanımlanır (Nathell, 2007).

Karışımli model yaklaşımında, heterojen bir yapı gösteren bir veri setinin, kendi içinde homojen ve aralarında heterojen olmak üzere kaç alt gruplu modelden oluşacağına karar vermede AIC ve BIC model uygun istatistikleri yaygın olarak kullanılmaktadır (Wang ve Liu, 2006; Peny, 2012; Xie ve ark., 2010). En küçük AIC ve BIC değerlerine sahip alt gruplu model, veri setinin en iyi dağıldığı model olarak kabul edilmektedir (McDonald, 1989). Bununla birlikte, veri setinin en iyi dağıldığı modeli

belirlemede, entropy doğru sınıflandırma olasılığı da yaygın olarak kullanılmaktadır. Entropy ölçütü hangi alt gruplu modelde daha yüksek elde edilirse, en iyi modelin o olduğuna karar verilmektedir (Liddle, 2007). Bu çalışmada, AIC ve BIC değerlerinin, 5 alt gruplu modele kadar azaldığı ve bundan sonra da arttığı saptanmıştır. Böylece, en küçük istatistiklere sahip AIC ve BIC ve yüksek doğru sınıflandırma oranına sahip entropy değeri 5 alt gruplu modelde elde edilmiştir. Veri setindeki heterojenlik yüksek olduğunda, genellikle 1 alt gruplu modelin bilgi ölçütleri diğer alt gruplu modellere nazaran çok daha yüksek olmaktadır. Bu çalışmada, bir alt gruplu modelde AIC ve BIC değerleri sırasıyla 1422.759 ve 1441.609 olarak elde edilmişken, bu model uyum ölçütlerinin en küçük olarak elde edildiği 5 alt gruplu modelde ise sırasıyla 1310.366 ve 1410.616 olarak elde edilmiştir (Çizelge 4.1 ve Şekil 4.4). Yani, AIC ve BIC değerlerinde çok büyük değişim olmamakla beraber, bir azalma saptanmıştır. Entropi doğru sınıflandırma istatistiği, olası tüm alt gruplu modellerde %84.5 ve %92.5 arası değişmiştir. Tüm alt gruplu modeller dikkate alındığında, yüksek bir entropi sınıflandırma olasılığının elde edildiği söylenebilir. En uygun model seçilen 5 alt gruplu modelde, entropi ölçütü %92.5 olarak elde edilmiştir. Bağımlı değişkene göre bağımsız değişkenler dikkate alındığında, her bir hastanın ait olduğu grup olasılığı ne kadar yüksek olursa, genel entropi değeri de o kadar yüksek olmaktadır. Çünkü entropi istatistiği tüm hastaların bireysel olarak, ait oldukları grup olasılıkları kullanılarak hesaplanmaktadır (Liddle, 2007).

171 hastanın 5 alt gruba göre dağılımına bakıldığında, en çok bireyin dahil olduğu grup, 78 (%45.61) birey ile beşinci alt grup olmuştur. Bu alt grupta, FEV1/FVC bağımlı değişkenin 98.446 ile en yüksek ortalama değerine sahip grup olmuştur. Bireylerin en az dağıldığı grubun, 11 birey (%6.43) ile birinci alt grup olduğu saptanmıştır (Çizelge 4.2 ve Çizelge 4.3). Bağımlı değişken olarak alınan FEV1/FVC'nin beş alt grup bazında ortalama değerinin değişimi Şekil 4.11'de verilmiştir. Alt grup sayısı 1'den 5'e doğru gittiğinde, ortalama FEV1/FVC değerinin de büyüdüğü saptanmıştır. Alt grup 1'den 5'e doğru sıralı olarak FEV1/FVC değerleri incelendiğinde, obstrüktif paternden restriktif paterne doğru düzenli ve anlamlı bir geçiş gözlenmektedir.

FEV1/FVC bağımlı değişkeni üzerine etkili olan cinsiyet, yaş, boy ve kilo bağımsız değişkenlerine ait Gaussian karışimli regresyon kullanılarak elde edilen sonuçlar Çizelge 4.3'de verilmiştir. FEV1/FVC ortalamasının 50.071 ile en düşük olduğu birinci alt grupta, yalnızca kilonun etkisi istatistiksel olarak önemli bulunmuştur

($p < 0.01$). Kilo bir birim arttığında FEV1/FVC değeri 0.445 kat arttığı saptanmıştır. Bununla birlikte cinsiyet, yaş ve boyun FEV1/FVC üzerindeki etkileri istatistiksel olarak önemsiz bulunmuştur ($p > 0.05$). Bu alt grup, kilo bağımsız değişkenin ortalamasının 78.862 kg ile en yüksek olduğu grup olmuştur. Birinci alt grup, kilo ortalaması (78.862) olarak en yüksek ortalamaya sahip grup olmakla birlikte, FEV1/FVC değeri ortalamasının en düşük olduğu (50.071), yani obstrüktif paternin en belirgin olduğu grup olarak öne çıkmaktadır. Obstrüktif paterne sahip kişilerde, FEV1/FVC değerinin kilo ile ters orantılı olduğu sonucuna ulaşılmaktadır. Obezite ile Kronik Obstrüktif Akciğer Hastalığı arasındaki sebep sonuç ilişkisi henüz net belirlenmemiş olmakla birlikte, Amerika Birleşik Devletleri'nde 2007 yılında yapılan bir çalışmada, KOAH'lı hastaların obezite sıklığının, genel obezite sıklığına göre yaklaşık 2 kat daha fazla olduğu gösterilmiş olup, (%54-%20) bu oran oldukça yüksektir (Eisner ve ark., 2007).

Birinci alt gruptan sonraki en yüksek ortalama FEV1/FVC 67.034 ile ikinci alt grupta elde edilmiştir. Bu alt grupta kilo ve yaş değişkenlerinin FEV1/FVC üzerine olan etkileri istatistiksel olarak önemli bulunmuştur ($p < 0.01$). Yaş bir birim arttığında FEV1/FVC değeri -0.226 kat azaldığı, kilo birim arttığında ise FEV1/FVC değerinin 0.226 kat arttığı saptanmıştır. İkinci alt grup popülasyonu, FEV1/FVC değeri için kritik eşiği oluşturan %70 değerine yakın, alt ve üst değerlere sahip bireylerden oluştuğu, obstrüktif, normal ve restriktif paterne sahip bireylerden oluşan, FEV1/FVC değeri açısından karma bir gruptur. Bu gruptaki bireylerden obstrüktif ve restriktif paterne sahip olanlarda, hastalığın hafif aşamada olduğu söylenebilir.

FEV1/FVC ortalama değerinin 98.466 ile en yüksek olduğu beşinci alt grupta, bağımsız değişkenler içinde sadece kilonun etkisinin FEV1/FVC üzerine olan etkisi istatistiksel olarak önemli bulunmuştur ($p < 0.01$). Bu alt grupta kilo bir birim arttığında FEV1/FVC değerinin de -0.04 kat azaldığı saptanmıştır. Kilo bağımsız değişkenin ortalamasının 71.341 kg ile en düşük olduğu grup, beşinci alt grup olmuştur. Burada, ortalama kilo değeri düştükçe, FEV1/FVC ortalama değerinin de arttığı söylenebilir.

Çizelge 4.3'de verilen beş alt gruba bakıldığında, alt grup 1'den alt grup 5'e doğru gidildiğinde FEV1/FVC ortalama değerinin de büyüdüğü belirlenmiştir. Tüm alt gruplara bakıldığında (dördüncü alt grup hariç), bağımsız değişkenler içinde yalnızca kilonun etkisi istatistiksel olarak önemli bulunmuştur. FEV1/FVC ortalama değeri büyüdüğünde, kilonun etkisinin katsayı olarak azaldığı saptanmıştır. Örneğin FEV1/FVC ortalama değerinin en yüksek olduğu dördüncü (93.592) ve beşinci (98.466)

alt gruplarda kilonun değeri katsayı (-0.04 ile -0.036) olarak eksi işaretli olduğu belirlenmiştir. Solunumsal olarak normal kişilerde veya hafif hastalık belirtilerinin olduğu durumlarda, yüksek kilonun obstrüktif paternedeki hastalıklara yatkınlığı arttırdığı sonucuna ulaşılabilir. Kilo değişkeninin katsayısının büyüklüğü, büyükten küçüğe doğru, birinci, ikinci ve üçüncü alt grup şeklinde sıralandığı göze çarpmaktadır. Tüm alt grupların obstrüktif paternden restriktif paterne doğru sıralandığı göz önünde bulundurulduğunda, obstrüktif paterne sahip kişilerde kilonun, FEV1/FVC değerini düşürerek hastalığın şiddetini arttırdığı, restriktif paterne doğru gittikçe bu etkinin azaldığı, ve restriktif paternin belirgin olduğu hastalarda ise (dördüncü ve beşinci alt grup) kilonun arttıkça FEV1/FVC'yi düşürerek hastalığın şiddetini azalttığı sonucuna ulaşılmaktadır. Bu bulgulardan hareketle, obezitenin obstrüktif akciğer hastalıkları ile yakın ilişkisi sonucuna ulaşılması dikkat çekicidir.

Yaş bağımsız değişkenin FEV1/FVC üzerine olan etkisinin alt gruplara göre dağılımına bakıldığında (Çizelge 4.3), FEV1/FVC ortalama değerinin düşük olduğu alt gruplarda yaş değişkeninin de katsayısının eksi işaretli olduğu saptanmıştır. Bununla birlikte, FEV1/FVC ortalama değerinin yüksek olduğu alt gruplarda yaş değişkeninin de katsayısının işaretinin artıya döndüğü belirlenmiştir. Başka bir değişle, FEV1/FVC ortalama değeri büyüdükçe, yaş bağımsız değişkenin katsayı işareti eksiden artıya dönüşmektedir. FEV1/FVC ortalama değeri büyüdükçe, yaş bağımsız değişkenin katsayı olarak değeri de pozitif yönde büyümektedir.

Çizelge 4.4'e bakıldığında, yaş ortalama değeri büyüdükçe, FEV1/FVC ortalama değerinin de büyüdüğü saptanmıştır. Buradan, yaşı büyük olan hastaların FEV1/FVC değerinin de büyüdüğü söylenebilir. Obstrüktif paterne sahip hastalığın başlangıç aşamasındaki hastaların yığıldığı grubu ikinci alt grup olarak kabul ettiğimizde, bu aşamadaki hastalarda yaş arttıkça FEV1/FVC değeri diğer gruplara nazaran daha belirgin şekilde düşmekte ve bu alt grupta hastalığın şiddetinin yaş artışı ile ciddi oranda arttığı sonucuna ulaşılmaktadır. Amerika Birleşik Devletlerinde 1998 ile 2009 yılları arasında veri toplanarak yapılan bir çalışmada, KOAH prevalansının yaşamın büyük bölümünde hem erkek hem de kadınlar için yaşla birlikte arttığı sonucuna ulaşılmıştır (Akinbami, 2011). Normal yaşlanma ile solunum sisteminde oluşan fizyolojik değişikliklerin; akciğerin elastikiyetinde önemli bir azalma, göğüs duvarında sertleşme ve solunum kaslarında güç kaybı ile karakterize olduğu, yapılan çalışmalarda gösterilmiştir (Turner ve ark., 1968; Knudson ve ark., 1977). Bu fizyolojik

değişiklikler; FEV1/FVC değerinin yaşlı populasyonda daha düşük olmasından sorumlu tutulmuştur (Swanney ve ark., 2008).

Boy bağımsız değişkenin FEV1/FVC bağımlı değişken üzerine olan etkisi tüm alt gruplarda önemsiz bulunmuştur. Boy bağımsız değişkeni katsayı işareti olarak, dağınık bir görünüm sergilemiştir. Çizelge 4.3'e bakıldığında, FEV1/FVC ortalama değeri büyüdükçe boy bağımsız değişkenin katsayı değerinin küçüldüğü ve etkisinin de negatif yönlü olduğu saptanmıştır. Başka bir deyişle boyu kısa olan hastaların FEV1/FVC değerinin daha yüksek olduğu söylenebilir.

Cinsiyet bağımsız değişkenine bakıldığında, FEV1/FVC bağımlı değişken üzerine olan etkisi yalnızca üçüncü ve dördüncü alt grupta istatistiksel olarak önemli bulunmuştur. Beşinci alt grup hariç, kadınların erkeklerden daha çok FEV1/FVC ortalama değerinin artırdığı saptanmıştır (Çizelge 4.3). Solunumsal olarak normal kişilerde veya hafif hastalık belirtilerinin olduğu durumlarda, Erkek cinsiyetin obstrüktif paternde olan hastalıklara, kadın cinsiyetin ise restriktif paterne sahip hastalıklara daha meyilli olduğu sonucuna ulaşılmaktadır. Üçüncü alt grup, cinsiyet değişkeninin katsayı olarak en etkili olduğu grup olarak öne çıkmaktadır. Kadın cinsiyetin restriktif paterne sahip hastalıklara daha meyilli olduğu sonucuna ulaşılmaktadır. KOAH'ta cinsiyet dimorfizmi ilk olarak 1987'de çalışılmıştır (Burrows ve ark, 1987). 2015 yılında yapılan bir çalışmada, restriktif paterne sahip en önemli tanı olan, diffüz interstisyel akciğer hastalığı tanısı alan olgular retrospektif olarak değerlendirilmiş olup, vakaların %7,97'sinin erkek, %92,02'sinin ise kadın olduğu görülmüştür. Erkeklerde akciğerlerde obstrüktif akciğer hastalığı bulgusu olan amfizematöz değişikliklerin görülme sıklığının daha fazla olduğu sonucuna ulaşılmıştır. Erkek cinsiyette solunum fonksiyonlarının daha düşük (obstrüktif paternde) olduğu, obstrüktif paterne bağlı mortalitenin ise daha yüksek olduğu görülmüştür (Aşuk ve ark., 2015). Beşinci alt grupta durum değişmiştir. Bu alt grup FEV1/FVC ortalama değerinin en yüksek olduğu gruptur.

Veri seti bir bütün olarak alındığında, FEV1/FVC bağımlı değişkenin, bağımsız değişkenler ile olan korelasyon katsayıları genellikle küçük ve istatistiksel olarak önemsiz bulunmuşlardır (Şekil 4.5). Gaussian karışımı regresyon uygulandıktan sonra en iyi model olarak seçilen beş alt gruplu modelde, FEV1/FVC bağımlı değişkenin, bağımsız değişkenler ile olan korelasyon katsayılarının daha yüksek olduğu ve bazılarının da istatistiksel olarak önemli çıktığı saptanmıştır (Şekil 4.6, Şekil 4.7, Şekil 4.8, Şekil 4.9 ve Şekil 4.10). Böylece, veri setindeki heterojenlik düzeyinin, FEV1/FVC

bağımlı değişken ile bağımsız değişkenler arasındaki ilişki düzeyinin büyüklüğünü ve önem düzeyini ne derece etkilediği de görülmüştür.

Karışımli model yaklaşımı, son zamanlarda çok değişkenli verilerin analizinde yaygın olarak kullanılmaktadır (Kim ve ark., 2006; Davison ve Mastropietro, 2009; Martin ve Tokdar, 2011; Hui ve ark., 2015; Wong ve ark., 2015; Zhao ve Ma, 2018). Karışımli modeller hem sınıflandırma amaçlı hem de regresyon analizi amaçlı olarak kullanılmaktadır. Karışımli modeller, regresyon analizi ile birlikte sınıflandırmayı da yapan, nadir istatistiksel analiz yöntemlerinden biridir. Bu çalışmada bağımlı değişkenin normal dağılım göstermesi durumunda Gaussian karışımli regresyon analizi kullanılmıştır. Öncelikle, veri setinin heterojen bir yapı sergilediği varsayılarak, veri setinin kendi içlerinde homojen ve aralarında heterojen olmak üzere kaç alt grupta toplanacağı AIC, BIC istatistikleri kullanılarak belirlenmiştir. Bu alt grup sayısı, FEV1/FVC bağımlı değişkeni üzerine cinsiyet, boy, yaş ve kilo bağımsız değişkenleri dikkate alınarak belirlenmiştir. Böylece veri setinin beş alt grupta dağılabileceği saptanmıştır. Alt grup sayısı belirlendikten sonra her bir alt grup için ayrı parametre tahminleri elde edilmiştir (Çizelge 4.3 ve Çizelge 4.4). Karışımli modeller kullanılarak, hem alt grupları oluşturmada hem de elde edilen her bir alt grup için ayrı parametre tahminleri elde edilerek, çok daha detaylı sonuçlara ulaşmak mümkündür.

Sonuç olarak, hastaların FEV1/FVC değerleri üzerine etkili olabilecek cinsiyet, yaş, boy ve kilo bağımsız değişkenleri dikkate alınarak, Gaussian karışımli regresyon analizi sonucunda hem uygun alt grup sayısı hem de her bir alt grup için parametre tahminleri elde edilmiştir. İncelenen populasyonun farklı alt gruplardan oluştuğu sezgisel olarak bilinmektedir. Fakat bu alt grupların sayısı ve bu alt gruplarda hastaların hangi değişkenden ne kadar etkileneceğinin istatistiksel olarak tahmin edilmesi gerekmektedir. Bu noktada oluşan her alt gruba ait regresyon parametreleri ile değişkenlerin bağımlı değişken üzerindeki etkisi de Gaussian karışımli regresyon ile belirlenebilmektedir. Tıp alanında her zaman homojen veri seti ile çalışmanın mümkün olmayacağı düşünüldüğünde, Gaussian karışımli regresyonun önemli avantajlar sağlayabileceği düşünülmektedir.

İleriki çalışmalarda, Bayesian ya da yapay sinir ağları yaklaşımları kullanılarak karışımli regresyon analizinin yapılması ve elde edilen sonuçların Gaussian karışımli regresyon ile karşılaştırılması, uygun parametre tahminlerinin elde edilmesi açısından önemli olacaktır. Yine Gaussian karışımli regresyonun genelleştirme(*generalisation*)

yeteneđi ile performans ölçütlerinin ortaya konması için mukayeseli çalışmaların yapılmasının literatüre katkı sağlayabileceđi düşünölmektedir.





KAYNAKLAR

- Akinbami, L. J., Liu, X., 2011. Chronic obstructive pulmonary disease among adults aged 18 and over in the United States, *NCHS Data Brief*, (63):1-8.
- Aşuk Yaşar, Z., Unalmış, D., Gulru, P., Melih, B., Gulcan, U., Tibet, G., ... ,Talay, F. Diffüz interstisyel akciğer hastalığı tanısı alan olgularımızın retrospektif Değerlendirilmesi. *Abant Tıp Dergisi*, 4(4): 355-359.
- Bozdoğan, H. ,1990. On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, 19(1): 221-278.
- Bozdoğan, H. ,1993. Choosing the number of subgroup clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix, Chap. 5. *Information and Classification*, Springer, Berlin, Heidelberg, 460.
- Bozdoğan, H., Sclove, S. L. ,1984. Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistical Mathematics*, 36(1): 163-180.
- Bruneau, M., Mottet, T., Moulin, S., Kerbirou, M., Chouly, F., Chretien, S., Guyeux, C., 2018. A clustering package for nucleotide sequences using Laplacian Eigenmaps and Gaussian Mixture Model. *Computers in Biology and Medicine*, 93: 66-74.
- Burrows, B., Bloom, J. W., Traver, G. A., Cline, M. G., 1987. The course and prognosis of different forms of chronic airways obstruction in a sample from the general population. *New England Journal of Medicine*, 317(21): 1309-1314.
- Casella, G., Berger, R. R. 2001a. *Statistical Inference*. Duxbury Press, USA,102.
- Casella, G., Berger, R. R. 2001b. *Statistical Inference*. Duxbury Press,USA,57.
- Celeux, G. G. Soromenho, 1996. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2): 195-212.
- Davison, A. and D. Mastropietro,2009. Saddlepoint approximation for mixture models. *Biometrika*, 96(2): 479-486.
- Dawkins, L., Williamson, D., Barr, S., Lampkin, S., 2018. Influencing transport behaviour: a Bayesian modelling approach for segmentation of social surveys, *Journal of Transport Geography*, 20: 91-103.
- de Bragança Pereira, B., de Bragança Pereira, C. A., 2016. Preliminaries. In *Model Choice in Nonnested Families*. Springer, Berlin, Heidelberg.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*: 1-38.
- Do, K., McLachlan, G. J., 1984. Estimation of mixing proportions: a case study. *Applied Statistics*, 134-140.
- Eggink, J., Brown, G. J., 2003. A missing feature approach to instrument identification in polyphonic music. *2003 IEEE International Conference*, 5:V-553.
- Eisenberger, I., 1964. Genesis of bimodal distributions. *Technometrics*, 6(4): 357-363.
- Eisner, M. D., Blanc, P. D., Sidney, S., Yelin, E. H., Lathon, P. V., Katz, P. P., ... Iribarren, C., 2007. Body composition and functional limitation in COPD. *Respiratory Research*, 8(1): 7.

- Francis K.C. Hui, David I. Warton, Scott D. Foster; Order selection in finite mixture models: complete or observed likelihood information criteria?, *Biometrika*, **102**(3): 724-730.
- Furman, W. D., Lindsay, B. G., 1994. Testing for the number of subgroups in a mixture of normal distributions using moment estimators. *Computational Statistics & Data Analysis*, **17**(5): 473-492.
- Hui, F. K., Warton, D. I., Foster, S. D., 2015. Order selection in finite mixture models: complete or observed likelihood information criteria?. *Biometrika*, **102**(3): 724-730.
- Ji, Z., Huang, Y., Xia, Y., Zheng, Y., 2017. A robust modified Gaussian mixture model with rough set for image segmentation. *Neurocomputing*, **266**: 550-565.
- Kanji G. K., 1985. A mixture model for wind shear data, *Journal of Applied Statistics*, **12**(1): 49-58.
- Kasahara, H., Shimotsu, K., 2014. Non-parametric identification and estimation of the number of subgroups in multivariate mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**(1):97-111.
- Khalili, A., Chen, J., 2007. Variable selection in finite mixture of regression models. *Journal of the American Statistical Association*, **102**(479): 1025-1038.
- Kim, S., Tadesse, M. G., Vannucci, M., 2006. Variable selection in clustering via Dirichlet process mixture models. *Biometrika*, **93**(4): 877-893.
- Knudson, R. J., Clark, D. F., Kennedy, T. C., Knudson, D. E., 1977. Effect of aging alone on mechanical properties of the normal adult human lung. *Journal of Applied Physiology*, **43**(6): 1054-1062.
- Koehler, A. B., Murphree, E. S., 1988. A comparison of the Akaike and Schwarz criteria for selecting model order. *Applied Statistics*:187-195.
- Konishi, S., Kitagawa, G., 1996. Generalised information criteria in model selection. *Biometrika*, **83**(4): 875-890.
- Li Z., Shin S., Jeon S. I., Son S. H., Pack J. K., 2012. A new histogram-based breast cancer image classifier using Gaussian mixture model. *In Proceedings of the 2012 ACM Research in Applied Computation Symposium (RACS '12)*. New York, USA.
- Liang, J., Chen, K., Lin, M., Zhang, C., Wang, F., 2018. Robust finite mixture regression for heterogeneous targets. *Data Mining and Knowledge Discovery*: 1-52.
- Liddle, A. R., 2007. Information criteria for astrophysical model selection. *Monthly Notices of the Royal Astronomical Society: Letters*, **377**(1): L74-L78.
- 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.
- Liland, K. H., 2011. Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis. *TrAC Trends in Analytical Chemistry*, **30**(6): 827-841.
- Lindsay, B. G., Basak, P., 1993. Multivariate normal mixtures: a fast consistent method of moments. *Journal of the American Statistical Association*, **88**(422): 468-476.
- Lubke, G. H., Muthén, B., 2005. Investigating population heterogeneity with factor mixture models. *Psychological methods*, **10**(1): 21.
- Martin, R., Tokdar, S. T., 2011. Semiparametric inference in mixture models with predictive recursion marginal likelihood. *Biometrika*, **98**(3): 567-582.
- M.S.P. Babu, B.R. Naidu, 2014. A novel framework for JPEG image compression using baseline coding with parallel process, *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research*, 1–7.

- M.S.P. Babu, B.R. Naidu, 2016. Development of a biometric authentication system based on HAAR transformation and score level fusion, *Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science*, 1092–1099.
- Marques, J., Moreno, P. J., 1999. A study of musical instrument classification using Gaussian mixture models and support vector machines. *Cambridge Research Laboratory Technical Report Series CRL*, 4.
- McDonald, R. P., 1989. An index of goodness-of-fit based on noncentrality. *Journal of classification*, 6(1): 97-103.
- McLachlan, G., Peel, D., 2000. General Introduction. *Finite Mixture Models*, John Wiley & Sons, New York, 1-37.
- McLaren, C. E., Wagstaff, M., Brittenham, G. M., Jacobs, A., 1991. Detection of two-subgroup mixtures of lognormal distributions in grouped, doubly truncated data: Analysis of red blood cell volume distributions. *Biometrics*, 607-622.
- Miller, M. R., Hankinson, J. A. T. S., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Jensen, R., 2005. Standardisation of spirometry. *European Respiratory Journal*, 26(2): 319-338.
- Minelli, R., 1992. *Appunti dalle lezioni di fisiologia umana*. La Goliardica Pavese, ISBN: 8878300403, Italy, 170.
- Muthén, B. O., Muthén, L. K., Asparouhov, T., 2017. Regression and mediation analysis using Mplus. *Los Angeles, CA: Muthén & Muthén*.
- Muthén, B., Shedden, K., 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2): 463-469.
- Naidu, B. R., Babu, M. S. P., 2018. Biometric authentication Data with three traits using compression technique, HOG, GMM and Fusion technique. *Data in Brief*, 18: 1976-1986.
- Nathell, L., Nathell, M., Malmberg, P., Larsson, K., 2007. COPD diagnosis related to different guidelines and spirometry techniques. *Respiratory Research*, 8(1): 89.
- Nguyen, T. M., Wu, Q. J., Zhang, H., 2014. Bounded generalized Gaussian mixture model. *Pattern Recognition*, 47(9): 3132-3142.
- Omachi, M., Omachi, S., 2011. Pattern recognition with Gaussian mixture models of marginal distributions. *IEICE TRANSACTIONS on Information and Systems*, 94(2), 317-324.
- Orellana, R., Carvajal, R., Agüero, J. C., 2018. Maximum Likelihood Infinite Mixture Distribution Estimation Utilizing Finite Gaussian Mixtures. *IFAC-PapersOnLine*, 51(15): 706-711.
- Pearson, K., 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185: 71-110.
- Peng, W., Fang, Y., Zhan, R., 2017. A variable step learning algorithm for Gaussian mixture models based on the Bhattacharyya coefficient and correlation coefficient criterion. *Neurocomputing*, 239: 28-38.
- Penny, W. D., 2012. Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage*, 59(1): 319-330.
- Priebe, C., 1994. Adaptive Mixtures. *Journal of the American Statistical Association*, 89(427): 796-806.
- R.N. Balaka, P.B. Surendra, 2017. A novel biometric authentication system with score level fusion, *Ann. Data Science*.4(3): 383 – 404.
- Ramos-López, D., Masegosa, A. R., Salmerón, A., Rumí, R., Langseth, H., Nielsen, T. D., Madsen, A. L., 2018. Scalable importance sampling estimation of Gaussian

- mixture posteriors in Bayesian networks. *International Journal of Approximate Reasoning*, **100**: 115 – 134
- Rasmussen, C. E., 1999. The infinite Gaussian mixture model. *In Proceedings of the 12th International Conference on Neural Information Processing Systems (NIPS'99)*, MIT Press, Cambridge, MA, USA, 554-560.
- Sahebajami, H., Gartside, P. S., 1996. Pulmonary function in obese subjects with a normal FEV1/FVC ratio. *Chest*, **110**(6): 1425-1429.
- Sanderson, C., Paliwal, K. K., 2002. Likelihood normalization for face authentication in variable recording conditions. *Proc. International Conf. Image Processing*, IEEE, 301 – 304.
- Scherrer, Bertrand, 2007. Gaussian mixture model classifiers. *Lecture Notes, February*.
- Schork, N. J., Weder, A. B., Schork, M. A., Rao, D. C., 1990. On the asymmetry of biological frequency distributions. *Genetic Epidemiology*, **7**(6): 427-446.
- Sclove, S. L., 1987. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, **52**(3): 333-343.
- Shi, J. Q., Murray-Smith, R., Titterington, D. M., 2005. Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, **15**(1): 31-41.
- Soromenho, G. I. L. D. A., 1994. Comparing approaches for testing the number of subgroups in a finite mixture model. *Computational Statistics*, **9**(1): 65-78.
- Susan, S., Sharma, M., 2017. Automatic texture defect detection using Gaussian mixture entropy modeling. *Neurocomputing*, **239**: 232-237.
- Swales, J.D., 1985. Platt versus Pickering. *An Episode in Recent Medical History*, Keynes Press (British Medical Association), London, 8: vii – 155.
- Swanney, M. P., Ruppel, G., Enright, P. L., Pedersen, O. F., Crapo, R. O., Miller, M. R., Stocks, J., 2008. Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. *Thorax*, **63**(12): 1046 – 1051.
- Titterington, D. M., Smith, A. F., Makov, U. E., 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley, 48.
- Titterington, D. M., 1997. Mixture distributions (update). *Encyclopedia of Statistical Sciences*, 399 – 407.
- Trianto, R., Tai, T. C., Wang, J. C., 2018. Fast-LSTM acoustic model for distant speech recognition. *2018 IEEE International Conference on Consumer Electronics (ICCE)*, 1-4.
- Turner, J. M., Mead, J., Wohl, M. E., 1968. Elasticity of human lungs in relation to age. *Journal of Applied Physiology*, **25**(6), 664-671.
- Vaněk, Jan, Lukáš Machlica, Josef Psutka, 2013. Estimation of Single-Gaussian and Gaussian mixture models for pattern recognition. *Iberoamerican Congress on Pattern Recognition*. Springer, Berlin, Heidelberg.
- Verbeke, G., Lesaffre, E., 1996. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**(433): 217-221.
- Wallace, M. L., Buysse, D. J., Germain, A., Hall, M. H., Iyengar, S., 2018. Variable selection for skewed model-based clustering: Application to the identification of novel sleep phenotypes. *Journal of the American Statistical Association*, **113**(521): 95-110.
- Wang, P., Puterman, M. L., 1998. Mixed logistic regression models. *Journal of Agricultural, Biological, and Environmental Statistics*, 175-200.
- Wang, Y., Liu, Q., 2006. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research*, **77**(2): 220-225.

- Wang Z, Liu Y, Ding P., Bo X., 2002. Covariance-Tied Clustering Method In Speaker, Identification. *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI '02)*. IEEE Computer Society, Washington, DC, USA, 81.
- Weldon, W. F. R., 1894. II. On certain correlated variations in *Carcinus mænas*. *Proceedings of the Royal Society of London*, **54**(326-330): 318-329.
- Williams, C. K. I. C. E. Rasmussen, 1996. Gaussian Processes for Regression, *NIPS 8*, MIT Press.
- Withers, C. S., 1996. Moment estimates for mixtures of several distributions with different means or scales. *Communications in Statistics-Theory and Methods*, **25**(8), 1799-1824.
- Wong, W. K., Chen, R. B., Huang, C. C., Wang, W., 2015. A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLoS one*, **10**(6):3-23.
- Xie, W., Lewis, P. O., Fan, Y., Kuo, L., Chen, M. H., 2010. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, **60**(2): 150-160.
- Yan B., Yin M., Sarkar P, 2017. Statistical Convergence Analysis of Gradient EM on General Gaussian Mixture Models, *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 6940-6950.
- Yeşilova, A., Kaki, B., Almali, 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 Research*, **5**(12): 1509-1514.
- Yesilova, A., Yilmaz, A., Ser, G., Kaki, B., 2016. Modeling with Gaussian mixture regression for lactation milk yield in Anatolian buffaloes. *Indian Journal of Animal Research*, **50**(6): 989-994.
- Zhang, H., Wu, Q. J., Nguyen, T. M., 2013. Incorporating mean template into finite mixture model for image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, **24**(2): 328-335.
- Zhao, J., Ma, Y., 2018. Optimal pseudolikelihood estimation in the analysis of multivariate missing data with nonignorable nonresponse. *Biometrika*, **105**(2): 479-486.
- Zivkovic, Z., 2004. Improved Adaptive Gaussian Mixture Model for Background Subtraction, *ICPR 2004*, 2: 28-31.



EKLER

EK 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş değişkenlerine göre oluşturulan modelde, her bir örneğin her alt gruba girme olasılıkları ve atandığı alt gruplar.

FEV1/FVC	Cinsiyet	Boy	Kilo	Yaş	P(1)	P(2)	P(3)	P(4)	P(3)	Alt Gr.
81.700	2	157.000	82.000	45.000	0.000	0.024	0.976	0.000	0.000	3
77.200	2	166.000	76.000	34.000	0.000	0.998	0.002	0.000	0.000	2
64.900	1	160.000	56.000	71.000	0.000	0.013	0.987	0.000	0.000	3
85.600	1	180.000	80.000	62.000	0.000	0.000	0.140	0.859	0.000	4
77.500	2	154.000	54.000	61.000	0.000	0.000	1.000	0.000	0.000	3
80.000	1	175.000	75.000	52.000	0.000	0.000	0.987	0.013	0.000	3
83.400	2	151.000	55.000	74.000	0.000	0.000	1.000	0.000	0.000	3
86.800	2	160.000	57.000	27.000	0.000	0.000	0.998	0.001	0.001	3
83.000	1	178.000	69.000	26.000	0.000	0.000	0.857	0.143	0.000	3
96.100	1	157.000	87.000	43.000	0.000	0.000	0.000	0.048	0.952	5
74.200	2	164.000	84.000	37.000	0.000	1.000	0.000	0.000	0.000	2
88.700	1	180.000	77.000	21.000	0.000	0.000	0.500	0.494	0.006	3
100.000	2	160.000	65.000	18.000	0.000	0.000	0.007	0.302	0.691	5
98.300	1	178.000	70.000	20.000	0.000	0.000	0.000	0.002	0.998	5
84.800	1	180.000	62.000	23.000	0.000	0.000	0.690	0.310	0.000	3
74.700	2	165.000	61.000	17.000	0.000	1.000	0.000	0.000	0.000	2
88.700	2	164.000	54.000	19.000	0.000	0.000	0.979	0.012	0.009	3
81.300	2	153.000	83.000	49.000	0.000	0.036	0.964	0.000	0.000	3
73.700	1	166.000	73.000	66.000	0.000	0.000	1.000	0.000	0.000	3
86.500	1	180.000	84.000	37.000	0.000	0.000	0.533	0.467	0.000	3
99.300	1	180.000	68.000	26.000	0.000	0.000	0.000	0.001	0.999	5
98.000	2	162.000	55.000	18.000	0.000	0.000	0.009	0.330	0.662	5
95.500	2	156.000	88.000	48.000	0.000	0.000	0.041	0.321	0.638	5
93.700	2	168.000	69.000	24.000	0.000	0.000	0.446	0.258	0.296	3
85.500	2	155.000	50.000	38.000	0.000	0.000	1.000	0.000	0.000	3
92.400	2	170.000	66.000	26.000	0.000	0.000	0.642	0.197	0.161	3
86.800	2	160.000	79.000	46.000	0.000	0.000	0.996	0.002	0.002	3
90.800	2	165.000	58.000	32.000	0.000	0.000	0.824	0.096	0.081	3
73.300	2	160.000	96.000	34.000	0.000	1.000	0.000	0.000	0.000	2
94.800	2	159.000	63.000	32.000	0.000	0.000	0.050	0.349	0.601	5
100.000	1	183.000	72.000	35.000	0.000	0.000	0.000	0.000	1.000	5
100.000	2	170.000	62.000	21.000	0.000	0.000	0.009	0.241	0.750	5
95.600	2	153.000	44.000	34.000	0.000	0.000	0.000	0.324	0.676	5

EK. 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş değişkenlerine göre oluşturulan modelde, her bir örneğin her alt gruba girme olasılıkları ve atandığı alt gruplar (devam).

FEV1/FVC	Cinsiyet	Boy	Kilo	Yaş	P(1)	P(2)	P(3)	P(4)	P(3)	Alt Gr.
84.900	1	178.000	84.000	32.000	0.000	0.000	0.690	0.310	0.000	3
94.500	2	170.000	90.000	41.000	0.000	0.000	0.340	0.279	0.381	5
100.000	2	155.000	71.000	38.000	0.000	0.000	0.000	0.340	0.660	5
100.000	1	178.000	86.000	29.000	0.000	0.000	0.000	0.000	1.000	5
94.600	2	160.000	63.000	18.000	0.000	0.000	0.240	0.286	0.474	5
100.000	1	160.000	73.000	21.000	0.000	0.000	0.000	0.001	0.999	5
90.700	1	160.000	81.000	22.000	0.000	0.000	0.091	0.775	0.134	4
89.900	2	174.000	72.000	25.000	0.000	0.000	0.843	0.112	0.045	3
100.000	2	150.000	62.000	76.000	0.000	0.000	0.000	0.353	0.647	5
91.300	1	170.000	71.000	65.000	0.000	0.000	0.000	0.924	0.076	4
64.000	2	158.000	68.000	76.000	0.000	1.000	0.000	0.000	0.000	2
100.000	2	155.000	40.000	90.000	0.000	0.000	0.000	0.336	0.664	5
61.900	2	162.000	70.000	47.000	0.000	1.000	0.000	0.000	0.000	2
100.000	2	165.000	81.000	71.000	0.000	0.000	0.000	0.286	0.714	5
100.000	2	160.000	65.000	91.000	0.000	0.000	0.000	0.324	0.676	5
100.000	2	150.000	48.000	69.000	0.000	0.000	0.000	0.346	0.654	5
100.000	2	156.000	91.000	67.000	0.000	0.000	0.000	0.340	0.660	5
98.700	1	178.000	70.000	25.000	0.000	0.000	0.000	0.001	0.999	5
84.900	2	165.000	60.000	72.000	0.000	0.000	1.000	0.000	0.000	3
56.800	1	167.000	56.000	25.000	0.000	1.000	0.000	0.000	0.000	2
83.300	2	160.000	60.000	67.000	0.000	0.000	1.000	0.000	0.000	3
100.000	1	155.000	79.000	64.000	0.000	0.000	0.000	0.003	0.997	5
100.000	1	160.000	88.000	62.000	0.000	0.000	0.000	0.002	0.998	5
48.800	1	176.000	58.000	73.000	0.088	0.912	0.000	0.000	0.000	2
45.500	1	168.000	68.000	68.000	0.995	0.005	0.000	0.000	0.000	1
67.900	2	175.000	96.000	18.000	0.952	0.048	0.000	0.000	0.000	1
39.500	1	174.000	56.000	51.000	1.000	0.000	0.000	0.000	0.000	1
39.900	1	169.000	69.000	53.000	1.000	0.000	0.000	0.000	0.000	1
100.000	1	168.000	61.000	78.000	0.000	0.000	0.000	0.003	0.997	5
100.000	1	160.000	60.000	87.000	0.000	0.000	0.000	0.006	0.994	5
100.000	1	168.000	80.000	63.000	0.000	0.000	0.000	0.001	0.999	5
100.000	2	147.000	53.000	78.000	0.000	0.000	0.000	0.351	0.649	5
100.000	2	169.000	57.000	24.000	0.000	0.000	0.002	0.256	0.742	5
74.600	1	169.000	58.000	61.000	0.000	0.000	1.000	0.000	0.000	3
85.900	2	153.000	90.000	60.000	0.000	0.000	0.998	0.000	0.001	3
100.000	1	175.000	68.000	53.000	0.000	0.000	0.000	0.001	0.999	5
100.000	2	150.000	48.000	24.000	0.000	0.000	0.000	0.356	0.644	5

EK. 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş değişkenlerine göre oluşturulan modelde, her bir örneğin her alt gruba girme olasılıkları ve atandığı alt gruplar (devam).

FEV1/FVC	Cinsiyet	Boy	Kilo	Yaş	P(1)	P(2)	P(3)	P(4)	P(3)	Alt Gr.
77.800	2	150.000	52.000	83.000	0.000	0.000	1.000	0.000	0.000	3
97.800	2	165.000	82.000	76.000	0.000	0.000	0.000	0.337	0.663	5
97.000	2	150.000	50.000	82.000	0.000	0.000	0.000	0.304	0.696	5
89.600	2	153.000	71.000	76.000	0.000	0.000	0.174	0.224	0.602	5
74.700	2	140.000	44.000	78.000	0.000	0.000	1.000	0.000	0.000	3
100.000	1	167.000	54.000	18.000	0.000	0.000	0.000	0.001	0.999	5
90.400	1	177.000	67.000	67.000	0.000	0.000	0.000	0.979	0.021	4
53.600	2	150.000	80.000	41.000	1.000	0.000	0.000	0.000	0.000	1
84.000	2	168.000	47.000	20.000	0.000	0.000	1.000	0.000	0.000	3
99.000	2	154.000	78.000	57.000	0.000	0.000	0.000	0.343	0.657	5
87.700	1	168.000	100.000	66.000	0.000	0.000	0.047	0.947	0.006	4
72.900	2	165.000	120.000	67.000	0.357	0.643	0.000	0.000	0.000	2
98.200	2	167.000	92.000	23.000	0.000	0.000	0.275	0.208	0.517	5
99.600	2	168.000	68.000	20.000	0.000	0.000	0.028	0.254	0.718	5
79.700	1	180.000	89.000	73.000	0.000	0.000	0.984	0.016	0.000	3
96.200	1	177.000	103.000	64.000	0.000	0.000	0.000	0.013	0.987	5
98.900	1	180.000	78.000	58.000	0.000	0.000	0.000	0.001	0.999	5
68.200	1	160.000	51.000	21.000	0.000	0.990	0.010	0.000	0.000	2
97.800	1	165.000	49.000	21.000	0.000	0.000	0.000	0.016	0.984	5
92.700	1	168.000	85.000	54.000	0.000	0.000	0.000	0.562	0.438	4
92.000	1	164.000	50.000	25.000	0.000	0.000	0.000	0.856	0.144	4
84.900	1	158.000	71.000	33.000	0.000	0.000	0.521	0.479	0.000	3
80.000	2	148.000	80.000	70.000	0.000	0.000	1.000	0.000	0.000	3
92.500	2	168.000	66.000	25.000	0.000	0.000	0.623	0.195	0.182	3
66.300	2	165.000	82.000	36.000	0.002	0.998	0.000	0.000	0.000	2
87.300	2	165.000	80.000	73.000	0.000	0.000	0.990	0.006	0.004	3
99.000	1	157.000	73.000	58.000	0.000	0.000	0.000	0.008	0.992	5
95.600	2	165.000	55.000	15.000	0.000	0.000	0.126	0.346	0.528	5
97.400	2	163.000	84.000	89.000	0.000	0.000	0.000	0.347	0.653	5
89.500	1	160.000	106.000	38.000	0.000	0.000	0.315	0.618	0.067	4
77.100	1	172.000	86.000	59.000	0.000	0.000	0.999	0.001	0.000	3
83.800	2	165.000	62.000	27.000	0.000	0.000	1.000	0.000	0.000	3
79.100	1	169.000	79.000	75.000	0.000	0.000	0.993	0.007	0.000	3
96.900	2	155.000	45.000	21.000	0.000	0.000	0.001	0.338	0.661	5
97.700	1	155.000	90.000	34.000	0.000	0.000	0.000	0.011	0.989	5
92.000	2	158.000	122.000	49.000	0.000	0.000	0.444	0.189	0.367	3
50.500	2	164.000	64.000	20.000	1.000	0.000	0.000	0.000	0.000	1

EK. 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş değişkenlerine göre oluşturulan modelde, her bir örneğin her alt gruba girme olasılıkları ve atandığı alt gruplar (devam).

FEV1/FVC	Cinsiyet	Boy	Kilo	Yaş	P(1)	P(2)	P(3)	P(4)	P(3)	Alt Gr.
100.000	2	155.000	100.000	68.000	0.000	0.000	0.000	0.351	0.649	5
95.900	2	143.000	75.000	69.000	0.000	0.000	0.000	0.256	0.744	5
95.400	2	157.000	102.000	85.000	0.000	0.000	0.001	0.335	0.664	5
77.700	2	155.000	86.000	38.000	0.000	0.999	0.001	0.000	0.000	2
100.000	2	164.000	55.000	67.000	0.000	0.000	0.000	0.307	0.693	5
70.100	2	160.000	62.000	20.000	0.000	1.000	0.000	0.000	0.000	2
96.500	1	165.000	90.000	41.000	0.000	0.000	0.000	0.019	0.981	5
100.000	2	154.000	52.000	24.000	0.000	0.000	0.000	0.340	0.660	5
67.400	1	183.000	80.000	19.000	0.002	0.998	0.000	0.000	0.000	2
100.000	2	164.000	93.000	41.000	0.000	0.000	0.022	0.268	0.710	5
90.200	1	170.000	85.000	87.000	0.000	0.000	0.000	0.971	0.029	4
100.000	2	154.000	85.000	86.000	0.000	0.000	0.000	0.349	0.651	5
93.600	1	170.000	72.000	35.000	0.000	0.000	0.000	0.361	0.639	5
100.000	2	162.000	63.000	21.000	0.000	0.000	0.003	0.292	0.704	5
63.400	1	163.000	76.000	62.000	0.000	0.999	0.001	0.000	0.000	2
63.400	2	150.000	68.000	77.000	0.000	1.000	0.000	0.000	0.000	2
91.400	2	170.000	62.000	35.000	0.000	0.000	0.738	0.159	0.102	3
70.500	1	170.000	68.000	70.000	0.000	0.000	1.000	0.000	0.000	3
100.000	2	157.000	63.000	69.000	0.000	0.000	0.000	0.331	0.669	5
71.900	2	160.000	62.000	77.000	0.000	0.001	0.999	0.000	0.000	3
96.800	2	161.000	105.000	64.000	0.000	0.000	0.021	0.329	0.650	5
100.000	1	175.000	68.000	53.000	0.000	0.000	0.000	0.001	0.999	5
92.400	1	155.000	50.000	76.000	0.000	0.000	0.000	0.896	0.104	4
97.500	2	158.000	80.000	74.000	0.000	0.000	0.000	0.339	0.661	5
87.500	2	151.000	104.000	65.000	0.000	0.000	0.983	0.003	0.014	3
66.100	1	173.000	62.000	18.000	0.000	1.000	0.000	0.000	0.000	2
100.000	2	147.000	57.000	25.000	0.000	0.000	0.000	0.377	0.623	5
100.000	1	150.000	50.000	78.000	0.000	0.000	0.000	0.015	0.985	5
63.400	1	174.000	92.000	65.000	0.067	0.933	0.000	0.000	0.000	2
87.700	1	168.000	100.000	66.000	0.000	0.000	0.047	0.947	0.006	4
87.300	1	168.000	110.000	79.000	0.000	0.000	0.037	0.958	0.005	4
93.800	2	151.000	77.000	85.000	0.000	0.000	0.000	0.275	0.725	5
70.500	1	170.000	68.000	70.000	0.000	0.000	1.000	0.000	0.000	3
58.100	1	169.000	102.000	76.000	0.582	0.418	0.000	0.000	0.000	1
90.100	1	173.000	93.000	63.000	0.000	0.000	0.005	0.941	0.054	4
100.000	1	165.000	85.000	43.000	0.000	0.000	0.000	0.001	0.999	5
85.100	2	158.000	90.000	77.000	0.000	0.000	1.000	0.000	0.000	3

EK. 1. FEV1/FVC, Cinsiyet, Boy, Kilo ve Yaş değişkenlerine göre oluşturulan modelde, her bir örneğin her alt gruba girme olasılıkları ve atandığı alt gruplar (devam).

FEV1/FVC	Cinsiyet	Boy	Kilo	Yaş	P(1)	P(2)	P(3)	P(4)	P(3)	Alt Gr.
100.000	2	157.000	63.000	69.000	0.000	0.000	0.000	0.331	0.669	5
100.000	1	175.000	68.000	53.000	0.000	0.000	0.000	0.001	0.999	5
85.100	2	158.000	90.000	77.000	0.000	0.000	1.000	0.000	0.000	3
78.000	2	168.000	60.000	77.000	0.000	0.000	1.000	0.000	0.000	3
67.900	2	175.000	96.000	18.000	0.952	0.048	0.000	0.000	0.000	1
98.700	1	178.000	70.000	25.000	0.000	0.000	0.000	0.001	0.999	5
85.300	2	162.000	59.000	63.000	0.000	0.000	1.000	0.000	0.000	3
100.000	2	156.000	91.000	67.000	0.000	0.000	0.000	0.340	0.660	5
100.000	1	169.000	58.000	61.000	0.000	0.000	0.000	0.002	0.998	5
67.000	1	170.000	75.000	84.000	0.000	0.000	1.000	0.000	0.000	3
100.000	1	168.000	58.000	85.000	0.000	0.000	0.000	0.004	0.996	5
42.900	2	153.000	77.000	76.000	1.000	0.000	0.000	0.000	0.000	1
38.100	2	157.000	80.000	26.000	1.000	0.000	0.000	0.000	0.000	1
100.000	2	147.000	89.000	77.000	0.000	0.000	0.000	0.392	0.608	5
78.300	1	175.000	73.000	49.000	0.000	0.000	0.998	0.002	0.000	3
100.000	2	143.000	38.000	86.000	0.000	0.000	0.000	0.333	0.667	5
81.100	2	160.000	64.000	68.000	0.000	0.000	1.000	0.000	0.000	3
88.900	1	155.000	75.000	68.000	0.000	0.000	0.000	0.985	0.015	4
94.000	2	148.000	79.000	71.000	0.000	0.000	0.000	0.252	0.748	5
42.900	2	153.000	77.000	76.000	1.000	0.000	0.000	0.000	0.000	1
100.000	2	160.000	65.000	91.000	0.000	0.000	0.000	0.324	0.676	5
100.000	2	156.000	77.000	72.000	0.000	0.000	0.000	0.337	0.663	5
100.000	1	180.000	75.000	79.000	0.000	0.000	0.000	0.001	0.999	5
78.000	2	168.000	60.000	77.000	0.000	0.000	1.000	0.000	0.000	3
84.000	1	164.000	73.000	53.000	0.000	0.000	0.351	0.649	0.000	4
100.000	1	168.000	58.000	85.000	0.000	0.000	0.000	0.004	0.996	5
100.000	2	159.000	93.000	83.000	0.000	0.000	0.000	0.323	0.677	5

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

MODEL FIT INFORMATION
 Number of Free Parameters 30
 Loglikelihood -628.183

H0 Value -628.183
 H0 Scaling Correction Factor 1.2531
 for MLR

Information Criteria
 Akaike (AIC) 1316.366
 Bayesian (BIC) 1410.616
 Sample-Size Adjusted BIC 1315.623
 ($n^* = (n + 2) / 24$)

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
 BASED ON THE ESTIMATED MODEL

Latent Classes		
1	10.99909	0.06432
2	18.06106	0.10562
3	46.57835	0.27239
4	32.06723	0.18753
5	63.29427	0.37014

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
 BASED ON ESTIMATED POSTERIOR PROBABILITIES

Latent Classes		
1	10.99909	0.06432
2	18.06106	0.10562
3	46.57835	0.27239
4	32.06723	0.18753
5	63.29427	0.37014

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
 BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions Latent Classes		
1	11	0.06433
2	18	0.10526
3	49	0.28655
4	15	0.08772
5	78	0.45614

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

CLASSIFICATION QUALITY					
Entropy	0.808				
Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)					
by Latent Class (Column)	1	2	3	4	5
1	0.953	0.047	0.000	0.000	0.000
2	0.029	0.970	0.001	0.000	0.000
3	0.000	0.002	0.902	0.071	0.026
4	0.000	0.000	0.069	0.858	0.073
5	0.000	0.000	0.017	0.202	0.781
Classification Probabilities for the Most Likely Latent Class Membership (Column)					
by Latent Class (Row)	1	2	3	4	5
1	0.953	0.047	0.000	0.000	0.000
2	0.029	0.967	0.004	0.000	0.000
3	0.000	0.000	0.949	0.022	0.029
4	0.000	0.000	0.108	0.401	0.490
5	0.000	0.000	0.020	0.017	0.963
Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)					
by Latent Class (Row)	1	2	3	4	5
1	13.767	10.758	0.000	0.000	0.000
2	10.264	13.782	8.325	0.000	0.000
3	-10.274	-4.545	3.489	-0.267	0.000
4	-13.103	-13.103	-1.511	-0.201	0.000
5	-13.777	-13.777	-3.869	-4.015	0.000
MODEL RESULTS					
Latent Class 1	Estimate	S.E.	Est./S.E.	Two-Tailed	P-Value
DEP					
CINS	8.153	7.746	1.053	0.293	
BOY	0.626	0.552	1.133	0.257	
KILO	0.445	0.129	3.461	0.001	
YAS	-0.006	0.179	-0.032	0.975	
Intercepts					
DEP	-101.056	103.021	-0.981	0.327	
Residual Variances					
DEP	9.264	2.050	4.519	0.000	

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

Categorical Latent Variables		Means		Lower 2.5%		Lower 5%		Estimate		Upper 5%		Upper 2.5%		Upper .5%	
C#1		-1.750	0.347	-5.039		0.000		0.000		20.896	23.336	28.106			
C#2		-1.254	0.288	-4.350		0.000		0.000		1.535	1.709	2.049			
C#3		-0.307	0.235	-1.303		0.192		0.192		0.656	0.697	0.776			
C#4		-0.680	0.490	-1.388		0.165		0.165		0.288	0.345	0.455			
QUALITY OF NUMERICAL RESULTS								0.227E-06							
Condition Number for the Information Matrix (ratio of smallest to largest eigenvalue)															
CONFIDENCE INTERVALS OF MODEL RESULTS															
Latent Class 1		Lower .5%		Lower 2.5%		Lower 5%		Estimate		Upper 5%		Upper 2.5%		Upper .5%	
DEP	ON														
CINS		-11.799	-7.029	-4.589		8.153		8.153		20.896	23.336	28.106			
BOY		-0.797	-0.457	-0.283		0.626		0.626		1.535	1.709	2.049			
KILO		0.114	0.193	0.233		0.445		0.445		0.656	0.697	0.776			
YAS		-0.466	-0.356	-0.300		-0.006		-0.006		0.288	0.345	0.455			
Intercepts															
DEP		-366.419	-302.978	-270.526		-101.056		-101.056		68.414	100.866	164.307			
Residual Variances															
DEP		3.984	5.246	5.892		9.264		9.264		12.636	13.282	14.544			
Latent Class 2															
DEP	ON														
CINS		-4.517	-2.485	-1.445		3.984		3.984		9.413	10.452	12.484			
BOY		-0.722	-0.620	-0.567		-0.295		-0.295		-0.022	0.030	0.133			
KILO		0.089	0.122	0.138		0.226		0.226		0.313	0.330	0.363			
YAS		-0.315	-0.294	-0.283		-0.226		-0.226		-0.168	-0.157	-0.136			
Intercepts															
DEP		21.518	40.742	50.576		101.930		101.930		153.283	163.117	182.341			
Residual Variances															
DEP		3.984	5.246	5.892		9.264		9.264		12.636	13.282	14.544			
Latent Class 3															
DEP	ON														
CINS		-2.631	0.184	1.624		9.144		9.144		16.664	18.104	20.919			
BOY		-0.454	-0.313	-0.241		0.137		0.137		0.515	0.587	0.729			
KILO		0.035	0.068	0.085		0.173		0.173		0.260	0.277	0.310			
YAS		-0.378	-0.333	-0.310		-0.188		-0.188		-0.1067	-0.044	0.002			
Intercepts															
DEP		-78.119	-49.226	-34.446		42.738		42.738		119.921	134.701	163.595			

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

	Residuals for Covariances	
DEP	DEP	
	<u>14.254</u>	
	ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 2	
	Model Estimated Means	
	DEP	
	<u>67.034</u>	
	Residuals for Means	
	DEP	
	<u>0.000</u>	
	Model Estimated Covariances	
	DEP	
	<u>47.341</u>	
DEP	Residuals for Covariances	
	DEP	
	<u>4.206</u>	
	ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 3	
	Model Estimated Means	
	DEP	
	<u>82.156</u>	
	Residuals for Means	
	DEP	
	<u>0.000</u>	
	Model Estimated Covariances	
	DEP	
	<u>43.870</u>	
DEP	Residuals for Covariances	
	DEP	
	<u>2.553</u>	

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 4
Model Estimated Means
DEP

93.592
Residuals for Means
DEP

0.000
Model Estimated Covariances
DEP

DEP 29.931
Residuals for Covariances
DEP

DEP -2.525
ESTIMATED MODEL AND RESIDUALS (OBSERVED - ESTIMATED) FOR CLASS 5
Model Estimated Means
DEP

98.466
Residuals for Means
DEP

0.000
Model Estimated Covariances
DEP

DEP 9.881
Residuals for Covariances
DEP

DEP -4.277

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

TECHNICAL 3 OUTPUT
ESTIMATED COVARIANCE MATRIX FOR PARAMETER ESTIMATES

	1	2	3	4	5
1	0.106134D+05				
2	-0.770602D+03	0.600019D+02			
3	-0.568306D+02	0.411639D+01	0.305112D+00		
4	0.121337D+02	-0.939152D+00	-0.659572D-01	0.165175D-01	
5	-0.180690D+02	0.130487D+01	0.962698D-01	-0.202044D-01	0.319446D-01
6	-0.365011D+02	0.254228D+01	0.187870D+00	-0.425121D-01	0.936710D-01
7	0.205282D+03	-0.146334D+02	-0.112194D+01	0.220807D+00	-0.271959D+00
8	-0.936214D+01	0.654375D+00	0.512363D-01	-0.874218D-02	0.114088D-01
9	-0.135787D+01	0.967886D-01	0.740340D-02	-0.149688D-02	0.190536D-02
10	0.716765D+00	-0.500854D-01	-0.385095D-02	0.838403D-03	-0.129613D-02
11	-0.572324D+00	0.399850D-01	0.307126D-02	-0.653438D-03	0.101453D-02
12	-0.181072D+03	0.114059D+02	0.949599D+00	-0.208641D+00	0.460780D+00
13	0.191726D+02	-0.121481D+01	-0.100543D+00	0.221123D-01	-0.484974D-01
14	0.905734D+00	-0.572201D-01	-0.474636D-02	0.104381D-02	-0.230930D-02
15	-0.267671D+00	0.176868D-01	0.138944D-02	-0.309310D-03	0.690734D-03
16	0.274578D+00	-0.174174D-01	-0.143633D-02	0.315822D-03	-0.703907D-03
17	-0.964636D+02	0.647438D+01	0.500103D+00	-0.111896D+00	0.246715D+00
18	0.115207D+02	-0.763836D+00	-0.598796D-01	0.133411D-01	-0.293837D-01
19	0.455959D+00	-0.308964D-01	-0.235981D-02	0.529400D-03	-0.116626D-02
20	0.550784D-01	-0.361981D-02	-0.285913D-03	0.640222D-04	-0.143370D-03
21	0.694323D-02	-0.642284D-04	-0.420516D-04	0.712116D-05	-0.156018D-04
22	0.888894D+00	-0.540596D-01	-0.475667D-02	0.989274D-03	-0.202098D-02
23	-0.158675D+01	0.114579D+00	0.809170D-02	-0.186226D-02	0.414229D-02
24	0.167394D-01	-0.123275D-02	-0.846871D-04	0.198652D-04	-0.449040D-04
25	0.188591D-01	-0.133689D-02	-0.968154D-04	0.219568D-04	-0.482101D-04
26	-0.161915D-01	0.118377D-02	0.824298D-04	-0.190378D-04	0.420813D-04
27	-0.434983D+01	0.280584D+00	0.236741D-01	-0.471955D-02	0.716950D-02
28	0.336677D+01	-0.226780D+00	-0.180072D-01	0.373075D-02	-0.628752D-02
29	0.135867D-01	-0.568499D-02	0.926585D-05	0.298396D-04	-0.876265D-04
30	0.250738D+01	-0.167939D+00	-0.129993D-01	0.290883D-02	-0.642656D-02

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

	6	7	8	9	10
6	0.420225D+01				
7	0.585340D+01	0.974564D+03			
8	-0.101657D+00	-0.937303D+02	0.108920D+02		
9	-0.344231D-01	-0.512697D+01	0.488095D+00	0.274948D-01	
10	-0.848911D-02	0.346690D+00	-0.514910D-01	-0.276771D-02	0.282509D-02
11	0.105587D-01	-0.258235D+00	0.114097D-01	0.132109D-02	-0.522027D-03
12	0.516541D+02	0.229487D+03	-0.124904D+02	-0.125306D+01	-0.483655D-01
13	-0.523833D+01	-0.222545D+02	0.122250D+01	0.121258D+00	0.601079D-02
14	-0.254718D+00	-0.111681D+01	0.600539D-01	0.610680D-02	0.231639D-03
15	0.562246D-01	0.182676D+00	-0.761912D-02	-0.102381D-02	-0.480792D-04
16	-0.758004D-01	-0.330237D+00	0.171088D-01	0.181486D-02	0.535292D-04
17	0.173426D+02	0.544780D+02	-0.264391D+01	-0.300804D+00	-0.249236D-01
18	-0.230231D+01	-0.762810D+01	0.379236D+00	0.419663D-01	0.320188D-02
19	-0.743216D-01	-0.203145D+00	0.929631D-02	0.112714D-02	0.119399D-03
20	-0.123656D-01	-0.618281D-01	0.333103D-02	0.339715D-03	0.633701D-05
21	-0.113403D-01	-0.648155D-01	0.377717D-02	0.350577D-03	0.543715D-05
22	-0.259187D+00	0.973062D+00	-0.791022D-01	-0.532502D-02	0.119794D-02
23	0.894327D-01	0.344841D-01	0.751326D-02	-0.346627D-03	-0.156075D-03
24	-0.543838D-03	-0.105123D-01	0.639342D-03	0.590605D-04	-0.370917D-05
25	-0.152503D-02	0.707965D-02	-0.629264D-03	-0.368079D-04	0.630479D-05
26	0.504159D-03	-0.257506D-02	0.239562D-03	0.124533D-04	-0.206855D-05
27	0.979088D-01	-0.396644D+00	0.267109D-01	0.208599D-02	0.230259D-03
28	-0.201386D-02	0.355335D+00	-0.244765D-01	-0.176507D-02	-0.306941D-03
29	0.112308D+00	0.453376D+00	-0.246020D-01	-0.243957D-02	-0.140045D-03
30	-0.462306D+00	-0.161134D+01	0.783736D-01	0.891680D-02	0.561128D-03

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

	11	12	13	14	15
11	1.000				
12	0.029	1.000			
13	-0.036	-0.972	1.000		
14	-0.029	-0.998	0.964	1.000	
15	0.035	0.683	-0.692	-0.717	1.000
16	-0.023	-0.922	0.871	0.919	-0.724
17	0.039	0.531	-0.537	-0.532	0.415
18	-0.043	-0.670	0.678	0.670	-0.525
19	-0.035	-0.364	0.371	0.366	-0.290
20	-0.012	-0.533	0.524	0.530	-0.406
21	-0.001	-0.497	0.486	0.492	-0.353
22	-0.004	0.069	-0.070	-0.068	0.045
23	0.019	0.064	-0.056	-0.064	0.040
24	-0.002	-0.116	0.119	0.116	-0.076
25	-0.010	0.098	-0.099	-0.096	0.044
26	0.013	-0.029	0.019	0.029	-0.018
27	0.020	-0.097	0.100	0.096	-0.054
28	-0.025	-0.098	0.105	0.097	-0.077
29	0.005	0.335	-0.319	-0.334	0.267
30	-0.040	-0.669	0.676	0.667	-0.551

	16	17	18	19	20
16	1.000				
17	-0.479	1.000			
18	0.610	-0.904	1.000		
19	0.319	-0.966	0.784	1.000	
20	0.515	-0.007	0.313	-0.224	1.000
21	0.485	-0.137	0.273	-0.057	0.469
22	-0.058	0.034	-0.080	-0.006	-0.136
23	-0.066	0.090	0.041	-0.116	0.135
24	0.097	-0.055	0.107	0.020	0.136
25	-0.076	0.000	-0.044	0.024	-0.010
26	0.045	-0.051	-0.008	0.058	-0.058
27	0.085	-0.042	0.053	0.030	0.032
28	0.087	-0.071	0.079	0.060	0.033

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

29	-0.342	0.087	-0.171	0.002	-0.281
30	0.637	-0.374	0.479	0.256	0.380
ESTIMATED CORRELATION MATRIX FOR PARAMETER ESTIMATES					
	21	22	23	24	25
21	1.000				
22	-0.019	1.000			
23	-0.057	-0.675	1.000		
24	0.053	-0.955	0.601	1.000	
25	-0.121	0.058	-0.165	-0.304	1.000
26	0.097	-0.331	0.142	0.181	0.132
27	0.045	0.004	0.028	-0.009	-0.008
28	0.023	0.008	0.000	-0.008	0.006
29	-0.305	0.035	0.057	-0.063	0.069
30	0.345	-0.010	-0.081	0.045	-0.030
ESTIMATED CORRELATION MATRIX FOR PARAMETER ESTIMATES					
	26	27	28	29	30
26	1.000				
27	0.013	1.000			
28	-0.008	0.162	1.000		
29	-0.029	0.225	0.301	1.000	
30	-0.015	0.265	0.336	0.041	1.000
TECHNICAL 4 OUTPUT					
ESTIMATES DERIVED FROM THE MODEL FOR CLASS 1					
ESTIMATED MEANS FOR THE LATENT VARIABLES					
DEP		CINS	BOY	KILO	YAS
50.071	1.660	164.180	78.862	47.643	
S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES					
DEP		CINS	BOY	KILO	YAS
3.238	0.143	2.717	4.540	6.859	
EST./S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES					
DEP		CINS	BOY	KILO	YAS
15.462	11.629	60.422	17.372	6.946	

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

TWO-TAILED P-VALUE FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP		BOY	YAS
	0.000	0.000	0.000
ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP		BOY	YAS
107.334			
CINS	0.224		
BOY	-2.090	81.210	
KILO	2.822	11.253	226.666
YAS	-4.437	-66.602	-37.267
S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP		BOY	YAS
43.300			
CINS	0.096		
BOY	1.433	34.630	
KILO	2.312	41.050	96.655
YAS	3.513	64.993	103.878
EST./S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP		BOY	YAS
2.479			
CINS	2.345		
BOY	-1.459	2.345	
KILO	1.220	0.274	2.345
YAS	-1.205	-1.025	-0.359
TWO-TAILED P-VALUE FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP		BOY	YAS
0.013			
CINS	0.019		
BOY	0.145	0.019	
KILO	0.031	0.784	0.019
YAS	0.228	0.305	0.720

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	1.000			
CINS	0.367	1.000		
BOY	0.420	-0.490	1.000	
KILO	0.840	0.396	0.083	1.000
YAS	-0.413	-0.412	-0.325	-0.109
S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	0.000			
CINS	0.291	0.000		
BOY	0.293	0.229	0.000	
KILO	0.102	0.254	0.299	0.000
YAS	0.277	0.250	0.270	0.298
EST./S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	999.000			
CINS	1.259	999.000		
BOY	1.431	-2.137	999.000	
KILO	8.223	1.556	0.277	999.000
YAS	-1.493	-1.645	-1.205	-0.365
TWO-TAILED P-VALUE FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	0.000			
CINS	0.208	0.000		
BOY	0.153	0.033	0.000	
KILO	0.000	0.120	0.782	0.000
YAS	0.136	0.100	0.228	0.715

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

ESTIMATES DERIVED FROM THE MODEL FOR CLASS 2				
ESTIMATED MEANS FOR THE LATENT VARIABLES				
DEP	67.034	1.598	164.800	42.548
			74.775	
S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES				
DEP	1.725	0.115	1.806	5.036
			3.713	
EST./S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES				
DEP	38.850	13.846	91.274	8.450
			20.140	
TWO-TAILED P-VALUE FOR ESTIMATED MEANS FOR THE LATENT VARIABLES				
DEP	0.000	0.000	0.000	0.000
			0.000	
ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
DEP	47.341			
CINS	2.057	0.240		
BOY	-17.479	-2.354	58.879	
KILO	40.075	2.035	6.017	248.962
YAS	-69.534	0.237	-34.962	99.673
				457.974
S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
DEP	16.555			
CINS	0.979	0.080		
BOY	12.616	1.044	19.593	
KILO	27.907	1.883	28.524	82.847
YAS	38.337	2.470	39.506	82.843
				152.400
EST./S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
DEP	2.860			
CINS	2.102	3.005		
BOY	-1.386	-2.254	3.005	

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

KILO	1.436	1.081	0.211	3.005	
YAS	-1.814	0.096	-0.885	1.203	3.005
	TWO-TAILED P-VALUE FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
DEP		CINS	BOY	KILO	YAS
DEP	0.004				
CINS	0.036	0.003			
BOY	0.166	0.024	0.003		
KILO	0.151	0.280	0.833	0.003	0.003
YAS	0.070	0.923	0.376	0.229	0.003
	ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
DEP		CINS	BOY	KILO	YAS
DEP	1.000				
CINS	0.610	1.000			
BOY	-0.331	-0.626	1.000		
KILO	0.369	0.263	0.050	1.000	
YAS	-0.472	0.023	-0.213	0.295	1.000
	S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
DEP		CINS	BOY	KILO	YAS
DEP	0.000				
CINS	0.153	0.000			
BOY	0.203	0.143	0.000		
KILO	0.206	0.219	0.235	0.000	
YAS	0.181	0.235	0.225	0.215	0.000
	EST./S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
DEP		CINS	BOY	KILO	YAS
DEP	999.000				
CINS	3.993	999.000			
BOY	-1.631	-4.368	999.000		
KILO	1.788	1.201	0.212	999.000	
YAS	-2.603	0.096	-0.948	1.374	999.000
	TWO-TAILED P-VALUE FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
DEP		CINS	BOY	KILO	YAS
DEP	0.000				
CINS	0.000	0.000			

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

BOY	0.103	0.000	0.000			
KILO	0.074	0.230	0.832		0.000	
YAS	0.009	0.923	0.343		0.169	0.000
ESTIMATES DERIVED FROM THE MODEL FOR CLASS 3						
ESTIMATED MEANS FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	82.156	1.661	163.816	70.099	54.843	
S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	1.329	0.069	1.322	2.167	3.009	
EST./S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	61.819	23.951	123.947	32.345	18.227	
TWO-TAILED P-VALUE FOR ESTIMATED MEANS FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	0.000	0.000	0.000	0.000	0.000	
ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	43.870					
CINS	1.540	0.224				
BOY	1.544	-2.614	81.362			
KILO	22.679	-1.330	15.426	218.769		
YAS	-87.111	-0.420	-61.782	26.628	421.673	
S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES						
DEP		CINS	BOY		KILO	YAS
	10.288					
CINS	0.680	0.046				
BOY	8.879	0.734	16.860			
KILO	17.890	1.044	19.679	45.332		
YAS	27.523	1.426	28.610	44.674	87.377	

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

	DEP	4.264						
	CINS	2.263	4.826					
	BOY	0.174	-3.564	4.826				
	KILO	1.268	-1.274	0.784	4.826			
	YAS	-3.165	-0.295	-2.159	0.596	4.826		
	DEP	0.000						
	CINS	0.024	0.000					
	BOY	0.862	0.000	0.000				
	KILO	0.205	0.203	0.433	0.000			
	YAS	0.002	0.768	0.031	0.551	0.000		
	DEP	1.000						
	CINS	0.491	1.000					
	BOY	0.026	-0.612	1.000				
	KILO	0.232	-0.190	0.116	1.000			
	YAS	-0.640	-0.043	-0.334	0.088	1.000		
	DEP	0.000						
	CINS	0.180	0.000					
	BOY	0.149	0.092	0.000				
	KILO	0.172	0.141	0.145	0.000			
	YAS	0.122	0.146	0.130	0.145	0.000		
	DEP	999.000						
	CINS	2.723	999.000					
	BOY	0.174	-6.686	999.000				
	KILO	1.345	-1.345	0.800	999.000			
	YAS	-5.266	-0.295	-2.561	0.603	999.000		

EST./S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES

ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES

S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES

TWO-TAILED P-VALUE FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

TWO-TAILED P-VALUE FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES			
DEP	0.000		
CINS	0.006		
BOY	0.862		
KILO	0.179	0.000	
YAS	0.000	0.547	0.000
ESTIMATES DERIVED FROM THE MODEL FOR CLASS 4			
ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
93.592	1.512	162.667	75.295
S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
1.041	0.088	1.610	3.181
EST./S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
89.881	17.134	101.065	23.669
TWO-TAILED P-VALUE FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
0.000	0.000	0.000	0.000
ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
29.931			
CINS	2.217		
BOY	-29.082	83.073	
KILO	-32.738	48.694	324.501
YAS	8.548	-40.826	104.483
S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES			
DEP	CINS	BOY	KILO
7.017			
CINS	0.062		
BOY	10.392	20.746	
			534.873
			YAS
			YAS

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

KILO	17.182	1.651	30.242	81.040	
YAS	24.549	2.042	37.916	75.849	133.578
EST./S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES					
DEP			BOY	KILO	YAS
DEP	4.266				
CINS	3.393	4.004			
BOY	-2.799	-2.831	4.004		
KILO	-1.905	-1.523	1.610	4.004	
YAS	0.348	-0.072	-1.077	1.378	4.004
TWO-TAILED P-VALUE FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES					
DEP			BOY	KILO	YAS
DEP	0.000				
CINS	0.001	0.000			
BOY	0.005	0.005	0.000		
KILO	0.057	0.128	0.107	0.000	
YAS	0.728	0.943	0.282	0.168	0.000
ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES					
DEP			BOY	KILO	YAS
DEP	1.000				
CINS	0.811	1.000			
BOY	-0.583	-0.577	1.000		
KILO	-0.332	-0.279	0.297	1.000	
YAS	0.068	-0.013	-0.194	0.251	1.000
S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES					
DEP			BOY	KILO	YAS
DEP	0.000				
CINS	0.080	0.000			
BOY	0.132	0.118	0.000		
KILO	0.155	0.163	0.161	0.000	
YAS	0.192	0.177	0.170	0.165	0.000

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

EST./S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES			
DEP	999.000		
CINS	10.102	999.000	
BOY	-4.412	-4.904	999.000
KILO	-2.150	-1.715	1.841
YAS	0.352	-0.072	-1.140
TWO-TAILED P-VALUE FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES			
DEP			
CINS			999.000
BOY			1.515
KILO			999.000
YAS			999.000
ESTIMATES DERIVED FROM THE MODEL FOR CLASS 5			
ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP			
CINS			
BOY			
KILO			0.000
YAS			0.130
			0.000
ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP			
CINS			
BOY			
KILO			71.341
YAS			53.221
S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP			
CINS			
BOY			
KILO			1.970
YAS			3.014
EST./S.E. FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP			
CINS			
BOY			
KILO			36.208
YAS			17.655
TWO-TAILED P-VALUE FOR ESTIMATED MEANS FOR THE LATENT VARIABLES			
DEP			
CINS			
BOY			
KILO			0.000
YAS			0.000

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	9.881			
CINS	-0.177	0.250		
BOY	2.645	-2.850	95.660	
KILO	-8.273	-0.412	19.565	245.720
YAS	5.565	1.518	-76.390	39.548
S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	2.081			
CINS	0.125	0.044		
BOY	2.339	0.711	17.004	
KILO	4.778	0.986	19.427	43.679
YAS	6.266	1.518	31.007	47.513
EST./S.E. FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	4.747			
CINS	-1.409	5.626		
BOY	1.131	-4.008	5.626	
KILO	-1.732	-0.417	1.007	5.626
YAS	0.888	1.000	-2.464	0.832
TWO-TAILED P-VALUE FOR ESTIMATED COVARIANCE MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	0.000			
CINS	0.159	0.000		
BOY	0.258	0.000	0.000	
KILO	0.083	0.676	0.314	0.000
YAS	0.374	0.317	0.014	0.405
				0.000

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	1.000			
CINS	-0.112	1.000		
BOY	0.086	-0.583	1.000	
KILO	-0.168	-0.053	0.128	1.000
YAS	0.074	0.127	-0.326	0.105
S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	0.000			
CINS	0.079	0.000		
BOY	0.076	0.083	0.000	
KILO	0.092	0.125	0.124	0.000
YAS	0.083	0.124	0.112	0.124
EST./S.E. FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	999.000			
CINS	-1.420	999.000		
BOY	1.139	-7.031	999.000	
KILO	-1.827	-0.419	1.032	999.000
YAS	0.895	1.024	-2.898	0.846
TWO-TAILED P-VALUE FOR ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES				
	DEP	CINS	BOY	YAS
DEP	0.000			
CINS	0.155	0.000		
BOY	0.255	0.000	0.000	
KILO	0.068	0.675	0.302	0.000
YAS	0.371	0.306	0.004	0.397

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

TECHNICAL 7 OUTPUT						
SAMPLE STATISTICS WEIGHTED BY ESTIMATED CLASS PROBABILITIES FOR CLASS 1						
Means						
	DEP	CINS	BOY	KILO	YAS	
	50.071	1.660	164.180	78.862	47.643	
Covariances						
	DEP	CINS	BOY	KILO	YAS	
DEP	121.588					
CINS	1.800	0.224				
BOY	39.176	-2.090	81.210			
KILO	131.095	2.822	11.253	226.666		
YAS	-97.369	-4.437	-66.602	-37.267	517.493	
SAMPLE STATISTICS WEIGHTED BY ESTIMATED CLASS PROBABILITIES FOR CLASS 2						
Means						
	DEP	CINS	BOY	KILO	YAS	
DEP	67.034	1.598	164.800	74.775	42.548	
Covariances						
	DEP	CINS	BOY	KILO	YAS	
DEP	51.546					
CINS	2.057	0.240				
BOY	-17.479	-2.354	58.879			
KILO	40.075	2.035	6.017	248.962		
YAS	-69.534	0.237	-34.962	99.673	457.974	
SAMPLE STATISTICS WEIGHTED BY ESTIMATED CLASS PROBABILITIES FOR CLASS 3						
Means						
	DEP	CINS	BOY	KILO	YAS	
DEP	82.156	1.661	163.816	70.099	54.843	
Covariances						
	DEP	CINS	BOY	KILO	YAS	
DEP	46.423					
CINS	1.540	0.224				
BOY	1.544	-2.614	81.362			
KILO	22.679	-1.330	15.426	218.769		

EK 2. MPLUS yazılımı ile oluşturulan modelin sonuçları (devam).

YAS	-87.111	-0.420	-61.782	26.628	421.673
SAMPLE STATISTICS WEIGHTED BY ESTIMATED CLASS PROBABILITIES FOR CLASS 4					
Means					
DEP		CINS	BOY	KILO	YAS
	93.592	1.512	162.667	75.295	55.521
Covariances					
DEP		CINS	BOY	KILO	YAS
	27.407				
CINS	2.217	0.250			
BOY	-29.082	-2.630	83.073		
KILO	-32.738	-2.515	48.694	324.501	
YAS	8.548	-0.147	-40.826	104.483	534.873
SAMPLE STATISTICS WEIGHTED BY ESTIMATED CLASS PROBABILITIES FOR CLASS 5					
Means					
DEP		CINS	BOY	KILO	YAS
	98.466	1.517	162.798	71.341	53.221
Covariances					
DEP		CINS	BOY	KILO	YAS
	5.604				
CINS	-0.177	0.250			
BOY	2.645	-2.850	95.660		
KILO	-8.273	-0.412	19.565	245.720	
YAS	5.565	1.518	-76.390	39.548	575.141

ÖZ GEÇMİŞ

Serdar ABUT 1984 yılında Van'da doğdu. İlkokulu, Van Vali Mithat Bey İlkokulu'nda, ortaokul ve liseyi Van Gölü Anadolu Lisesi'nde okudu. 2001 yılında yerleştiği Selçuk Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü'nde Lisans eğitimini 2008 yılında tamamladı. Yüksek lisansını Fırat Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilimdalı Donanım Bilim Dalı'nda tamamladı. 2015 yılında Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilimdalı Biyometri – Genetik Bilim Dalı'nda doktora eğitimine başladı. Daha önce Van Bölge İdare Mahkemesi (2008) ve Siirt Üniversitesi Eğitim Fakültesi'nde (2008-2015) görev yaptı. Evli ve bir çocuk babasıdır.

T.C
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU

Tarih: 13/02/2019

Tez Başlığı / Konusu:

Normal Dağılım Gösteren Verilerin Analizinde Gaussian Karışımli Regresyonun Kullanılması

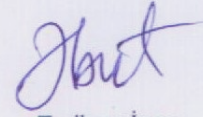
Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam 99 sayfalık kısmına ilişkin, 13/02/2019 tarihinde şahsım/tez danışmam tarafından TURNITIN intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinalite raporuna göre, tezimin benzerlik oranı % 8 (sekiz) dir.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimedenden daha az örtüşme içeren metin kısımları hariç (Limit inatch size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinalite Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.


Tarih ve İmza
13.02.2019

Adı Soyadı: Serdar ABUT

Öğrenci No: 159101056

Anabilim Dalı: Zootekni

Programı: Biyometri – Genetik

Statüsü: Y. Lisans

Doktora

DANIŞMAN ONAYI
UYGUNDUR

(Prof. Dr. Abdulllah YEŞİLOVA)

ENSTİTÜ ONAYI
UYGUNDUR

(Unvan, Ad Soyad, İmza)
Prof. Dr. Suat SENSOY
Enstitü Müdürü