

T. C.
ESKİŞEHİR OSMANGAZİ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

ESNEK AYIRMA ANALİZİ VE BİR UYGULAMA

DOKTORA TEZİ

Ahmet ÖZTÜRK

TEZ YÖNETİCİSİ
Prof. Dr. Kazım ÖZDAMAR

EYLÜL-2006

KABUL VE ONAY SAYFASI

Ahmet ÖZTÜRK'ün Doktora Tezi olarak hazırladığı “**Esnek Ayırma Analizi ve Bir Uygulama**” başlıklı bu çalışma, jürimizce Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliği'nin ilgili maddesi uyarınca değerlendirilerek “**KABUL**” edilmiştir.

10.10.2006

Prof. Dr. Kazım ÖZDAMAR
JÜRİ BAŞKANI

Doç. Dr. İsmet DOĞAN
ÜYE

Yrd. Doç. Dr. K. Setenay ÖNER
ÜYE

Yrd. Doç. Dr. Fezan Şahin MUTLU
ÜYE

Yrd. Doç. Dr. Canan DEMİRÜSTÜ
ÜYE

Osmangazi Üniversitesi Sağlık Bilimler Enstitüsü Yönetim Kurulu'nun
12.10.2006 Tarih ve 678/2147... sayılı kararıyla onaylanmıştır.

.....
Prof. Dr. Ferruh YÜCEL
Enstitü Müdürü

İÇİNDEKİLER

İÇİNDEKİLER.....	i
ÖZET.....	iii
SUMMARY.....	iv
ÇİZELGE DİZİNİ.....	v
1. GİRİŞ VE AMAÇ.....	1
2. GENEL BİLGİLER.....	5
2.1. Doğrusal Ayırma Analizi (DAA, Fisher Linear Discriminant Analysis)	5
2.1.1. Çokdeğişkenli Doğrusal Ayırma Analizi	7
2.2. Karesel Ayırma Analizi (KAA, Quadratic Discriminant Analysis) ...	14
2.3. Esnek Ayırma Analizleri (EAA, Flexible Discriminant Analyses) ...	16
2.3.1. Esnek Ayırma Analizi (FDA, Flexible Discriminant Analysis)..	17
2.3.2. Cezalandırılmış Ayırma Analizi (PDA, Penalized Discriminant Analysis.	20
2.3.3. Karma Ayırma Analizi (MDA, Mixture Discriminant Analysis).	21
2.3.4. MARS ve BRUTO Modelleri	22

2.4. MARS: Multivariate Adaptive Regression Splines	23
2.4.1. MARS Algoritmaları	25
2.4.1.1. Yinelemeli Bölünme Regresyonu (Recursive Partitioning Regression EAA/MARS degree=1).....	25
2.5. BRUTO: Uyumlu Eklemeli Modelleme (Adaptive Additive Modelling).....	28
3. GEREÇ VE YÖNTEMLER.....	30
3.1. Veri Türetimi (Simülasyon).....	30
3.2. Veri Analizi Yöntemleri	35
4. BULGULAR.....	36
5. TARTIŞMA VE SONUÇ.....	67
KAYNAKLAR DİZİNİ.....	71
ÖZGEÇMİŞ.....	75

ÖZET

Bu araştırmada parametrik (DAA ve KAA) ve parametrik olmayan çokdeğişkenli ayırma yöntemi Esnek Ayırma Analizinin uygulama modelleri; EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerinin performansları farklı birim, parametre, varyans ve kovaryans yapılarına göre türetilmiş veri yapılarında karşılaştırılmıştır.

Yöntemlerin etkinliklerini karşılaştırmak için, 4 bağımsız değişken içeren farklı ortalama vektörleri ve varyans-kovaryans matrislerine sahip çokdeğişkenli normal dağılım ile normal dağılım göstermeyen dağılımdan veriler türetilmiştir. 1000 tekrarlı simülasyon çalışmaları yapılarak türetilen veri setleri analizlerde kullanılmıştır.

Elde edilen bulgular incelendiğinde, veri setlerindeki gruplar kendi içinde homojen yapıda ve aralarında heterojen yapıda ise tüm yöntemlerin daha iyi performans gösterdiği belirlendi. Grup merkezlerinin birbirine yakın olduğu durumlarda eğer varyanslar küçük ise yöntemler sınıfları yüksek DSO değerleriyle birbirlerinden ayırdığı saptandı. Ancak birimlerin özellikleri birbirlerine karıştığında KAA, DAA'ya göre daha etkin DSO değerleri verdiği ve böylece sınıfları ayırma performansının daha iyi olduğu sonucuna varıldı. EAA yöntemlerinin ise genellikle farklı koşullarda DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri verdiği belirlendi.

Sonuç olarak, normal dağılmayan ve heterojen veri yapılarında parametrik olmayan EAA yöntemlerinin birbirlerine alternatif olarak kullanılması uygundur. Normal dağılan veri yapılarında ise koşulların tam yerine geldiği ve homojen yapılarda DAA'nın, aksi durumda ise yani heterojen yapılarda KAA'nın uygulanması gerektiği belirlendi.

Anahtar Kelimeler: Esnek Ayırma Analizi, Ayırma Analizi, MARS, BRUTO, Düzleştirme, Spline

SUMMARY

In this study, parametric (LDA and QDA) and nonparametric (FDA/MARS degree=1, FDA/MARS degree=2 and FDA/BRUTO) multivariate discriminant methods were compared under different conditions include sample sizes, parameters, variance and covariance structures by using generated data.

To compare the performance of the methods, the data used in the analyses were generated from multivariate normal and nonnormal distributions. There are four independent variables, varied mean vectors and variance covariance matrices in these data sets. 1000 replicated simulation studies were applied in the analyses.

According to the findings, it was determined that all the methods demonstrated good performances in the data which were homogeneous within groups and heterogeneous between groups. When the mean vectors of groups were similar and variances were small, it was determined that the whole methods showed high accurate classification rates. However, when there was overlapping in the data, QDA gave more efficient performance than LDA. Methods of FDA gave more accurate classification performance than LDA, but gave less accurate classification performance than QDA.

Consequently, it is appropriate that the usage of nonparametric FDA methods in the nonnormal and heterogeneous data sets. On the contrary in the normal and homogeneous data sets LDA should be used, but in the normal and heterogeneous data sets QDA is the superior method.

Keywords: Flexible Discriminant Analysis, Discriminant Analysis, MARS, BRUTO, Smoothing, Spline.

ÇİZELGE DİZİNİ

3.1.	Grup sayısı iki ($g=2$) olacak şekilde $V_1, V_2, \dots, V_{12}$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan $\mathbf{X}_i$ veri matrislerinin ortalama vektörleri	32
3.2.	Grup sayısı üç ($g=3$) olacak şekilde $V_1, V_2, \dots, V_{12}$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan $\mathbf{X}_i$ veri matrislerinin ortalama vektörleri	33
3.3.	Grup sayısı dört ($g=4$) olacak şekilde $V_1, V_2, \dots, V_{12}$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan $\mathbf{X}_i$ veri matrislerinin ortalama vektörleri	33
3.4.	Grup sayısı iki ($g=2$) (SET 14) ve grup sayısı üç ($g=3$) olacak şekilde V_1, V_2, V_3, V_7, V_8 ve V_9 kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan $\mathbf{X}_i$ veri matrisleri	34
4.1.	Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları sıfır olan (V_1, V_2, V_3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki $n=50, 100, 250, 500$ ve 1000 gözlemi DSO değerleri	38
4.2.	Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.25 olan (V_4, V_5, V_6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki $n=50, 100, 250, 500$ ve 1000 gözlemi DSO değerleri	40
4.3.	Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.75 olan (V_7, V_8, V_9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki $n=50, 100, 250, 500$ ve 1000 gözlemi DSO değerleri	42
4.4.	Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.90 olan (V_{10}, V_{11}, V_{12} kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki $n=50, 100, 250, 500$ ve 1000 gözlemi DSO değerleri	45
4.5.	Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları sıfır olan (V_1, V_2, V_3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki $n=50, 100, 250, 500$ ve 1000 gözlemi DSO değerleri	48

4.6.	Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.25 olan (V4, V5, V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	50
4.7.	Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.75 olan (V7, V8, V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	53
4.8.	Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.90 olan (V10, V11, V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	56
4.9.	Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları sıfır olan (V1, V2, V3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	59
4.10.	Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.25 olan (V4, V5, V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	61
4.11.	Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.75 olan (V7, V8, V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	62
4.12.	Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyans bir, üç ve on olan ve kovaryansları 0.90 olan (V10, V11, V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri	64
4.13.	Ayırma Analizi Yöntemleri'nin, iki (SET 14) ve üç (SET 15) veri matrisi birleştirmeli ve her biri 12 alt gruplu normal dağılıma uygunluk göstermeyen her alt sette DSO değerleri	66

1- GİRİŞ ve AMAÇ

Bilimin temel yöntemlerinden biri de karmaşık durumları olabildiğince basit bir forma indirgeyerek toplum ya da alt toplum dinamikleri ile ilgili geçerli, güvenilir ve tekrar edilebilir bilimsel bilgilere ulaşmaktır (35). Toplum içinde var olan doğal ya da gizli grupları (alt toplum, subpopulation) ortaya çıkarmak tıbbi sorunların çözümünde büyük önem taşımaktadır. Özellikle belirtileri birbirlerine benzeyen ancak değişkenlerin değişimlerinde küçük farklılıkların olduğu durumlarda tanı ve tedaviye ulaşmak oldukça güçleşmektedir. Tanı ve tedavide etkin ayırımlar ve tutarlı, yüksek güvenilirlikte kararlar üretmek için toplumları ya da toplum içindeki alt grupları sınıflama ve gruplama yöntemleri ile doğru belirlemek gerekmektedir (2, 18, 23, 24, 29).

Çok değişkenli istatistiksel yöntemler değişik amaçlarla uygulanmaktadır. Bunlar; veri indirgeme, kümeleme ve sınıflama, ölçekleme, çok değişkenli hipotezlerin test edilmesi biçiminde özetlenebilir. Çok değişkenli istatistiksel analiz yöntemlerinin büyük bir bölümü “çok değişkenli normal dağılım gösterdiği” varsayımını kullanır (7, 11, 13-17, 21, 23).

Normal dağılım; hem gerçek dünyada çoğu olaylara ilişkin popülasyon dağılımlarının normal olması bakımından, hem de çok değişkenli istatistiklerin örneklem dağılımlarının normal olması bakımından özel öneme sahiptir. Çok değişkenli problemlerin incelenmesinde normal dağılım varsayımını kullanmak alışılmış ve yaygın biçimde kullanılan bir yaklaşım olmasına karşın bazı oluşumların incelenmesinde sorunlar yaratmaktadır. Bu nedenle parametrik çok değişkenli istatistiksel yöntemler yanında normal dağılım varsayımı gerektirmeyen parametrik olmayan çok değişkenli istatistiksel yöntemler geliştirilmiştir (1, 7, 13-17, 21, 23, 31).

Tıp alanındaki problemlere uygulanması yeni olan bu yöntemlerin karar mekanizmalarındaki etkinliğinin ayrıntılı biçimde incelenmesini gerektirmektedir (1, 13-14, 21, 32-34).

Çok değişkenli yöntemlerin önemli avantajlarından biri de belirli özellikler yönünden birimleri ya da alt grupları sınıflama ve kümelemedir. Özellikle homojen alt grupları tanımlanmış ve belirlenen bu amaç ya da alt gruplara atama sağlık bilimlerinde prototip gruplar tanımlama bakımından büyük önem taşımaktadır. Özellikle iyi tanımlanmış hastalık gruplarını belirlemek ve yeni hasta bireylerin bu gruplara doğru olarak en az hata ile atanmasını sağlamak tanı ve tedavide önemli katkılar sağlamaktadır (3-8, 18-20).

Parametrik çok değişkenli istatistiksel yöntemler çokdeğişkenli normal dağılım varsayımı kullanmayı gerektirdiğinden her türlü tıbbi probleme uygulanamamaktadır. Bu nedenle; çok az ya da hemen hemen hiçbir varsayım gerektirmeyen parametrik olmayan çokdeğişkenli istatistiksel yöntemlerle tıbbi problemlerin çözümü yaygınlık kazanmıştır (1, 2, 11, 18, 23).

P sayıda değişkene göre toplumları ya da toplum içindeki homojen altgrupları (subgroups, subpopulation) belirlemek için çoğunlukla parametrik çokdeğişkenli ayırma analizleri olan Doğrusal Ayırma Analizi (DAA) ya da Karesel Ayırma Analizinden (KAA) (Linear Discriminant Analysis/Quadratic Discriminant Analysis) yararlanılmıştır. Parametrik Ayırma analizleri çok sayıda varsayım gerektiren bir çokdeğişkenli yöntem olması nedeniyle her türlü tıbbi problemin sınıflandırma/gruplandırma işlemlerinde uygulanamamaktadır. Ayırma analizi'nin en önemli iki koşulu olan Çokdeğişkenli Normal Dağılım varsayımı ve Grup Kovaryans Matrislerinin benzerliği varsayımları çoğu tıbbi problemde yerine getirilememekte ve Parametrik Ayırma Analizlerinin uygulanabileceği durumları sınırlamaktadır. Bu nedenle daha esnek ve varsayımları daha sınırlı Parametrik olmayan Ayırma analizlerine gereksinim duyulmaktadır. Özellikle tıbbi problemlerde gözlenen değişkenlerden bazılarının sıralı, aralıklı ya da skor değişkenler olması parametrik yöntemlerin uygulanmasını kısıtlamaktadır. Bu değişkenlerden elde edilen verilerin çokdeğişkenli normal dağılım varsayımına uygun olmamalarından dolayı toplumların/alttoplumların birbirlerinden ayrılmasında sorunlar yaşanmaktadır.

Bu gibi durumlarda Parametrik olmayan ayırma yöntemlerinden yararlanılması gerekmektedir. Parametrik olmayan ayırma yöntemleri genel olarak Esnek Ayırma Analizi (Flexible Discriminant Analysis) olarak anılmaktadır. Esnek Ayırma Analizi (EAA), veri matrisinde yer alan değişkenlerin İsimsel, Sıralı, Aralıklı, Skor ve Oransal ölçekli olmaları halinde bile ayırma ve gruplama işlemlerine uygulanabilmektedir (18, 23, 30-35). EAA kendi içinde yer alan ve sınıflama işlevini yerine getiren model yada prosedür adı verilebilecek yöntemlerden oluşmaktadır. Bunlardan en yaygın kullanıma sahip olanları Parametrik olmayan regresyon (Nonparametric Regression), Optimal Skorlara dayalı Regresyon (Nonparametric Regression wit Optimal Scoring), Esnek Ayırma Analizi (Flexible Discriminant Analysis), Cezalandırılmış Ayırma Analizi (Penalized Discriminant Analysis), Karma Ayırma Analizi (Mixture Discriminant Analysis), Çokdeğişkenli Uyarlanmış Regresyon Splinleri (Multivariate Adaptive Regression Splines, MARS), BRUTO yöntemi (BRUTO Models) vb. İsimlerle anılan yöntemleri içermektedir (1-7, 11-14, 30-34).

Her bir yöntemin ya da modelin uygulanabileceği özel durumlar olabilmektedir. Bazı yöntemler bazı koşullar yerine gelmiş olsa bile veri setindeki bağımsız değişken sayısının değişimine, birim sayısının değişimine, grup sayısının değişimine, değişken ölçeklerinin değişimine bağlı olarak performanslarında değişimler meydana gelmektedir. Bu nedenle Esnek Ayırma Analizi'nin karma veri yapıları içeren tıbbi problemlerde uygulanabilirliğinin araştırılması gerekmektedir (1, 11, 33, 34).

Esnek ayırma analizinin veri seti üzerindeki bazı koşulların varlığında dağılım varsayımları gerçekleştiğinde ve dağılım varsayımının gerçekleşmediği ya da göz önüne alınmadığında, veri sayısının büyüklüğüne ve değişken sayısına göre performansının değerlendirilmesi ve yöntemin uygulanabilirliğini belirlemek açısından önem taşımaktadır (6, 27, 31, 33, 34).

Literatürde Esnek Ayırma Analizi'nin tıbbi problemlere uygulanışı ile ilgili sınırlı sayıda yayın bulunmakta, ancak bu yayınlarda veri sayısının

değişimine, değişken sayısının değişimine ve dağılım varsayımlarının göz önüne alınıp alınmamasına bağlı olarak performansın incelendiği kapsamlı bir yayına rastlanamamıştır. Bu açıdan, bu araştırmada belirli koşulların Doğrusal ya da Karesel Ayırma Analizi yöntemlerine uygun olan veri yapılarında Esnek Ayırma Analizi'nin performansını değerlendirmek önem taşımaktadır. Ayrıca Doğrusal ya da Karesel Ayırma Analizi'nin uygulanamadığı tıbbi problemlerde ayırma ve gruplama işlemlerinin esnek ayırma analizi yöntemlerinden hangisi ile yapılmasının uygun olacağının belirlenmesi gerekmektedir (13-17).

Bu araştırma “sınıflama ve gruplama” amacıyla kullanılan parametrik olmayan çokdeğişkenli istatistiksel yöntem olan Esnek Ayırma Analizi (Flexible Discriminant Analysis) yöntemi içinde yer alan uygulama modelleri EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerini tanıtmak ve benzetim çalışmaları ile bu yöntemleri klasik yöntemler olan Doğrusal ve Karesel ayırma analizleri ile karşılaştırarak kullanımını göstermek amaçlanmıştır.

Bu çalışmada;

- Parametrik ve Parametrik Olmayan Çokdeğişkenli “Ayırma / Sınıflama ve Gruplama” İstatistiksel Yöntemlerin karşılaştırmalı tanıtımını yapmak,
- EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerini açıklamak,
- DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO modellerinin hangi tür veri setlerinde doğru sınıflandırma oranlarına sahip olduklarını belirlemek,
- Esnek Ayırma Analizi modellerinin heterojen yapıdaki tıbbi veri setlerine uygulanabilirliğini test etmek ve Doğru sınıflama oranlarına göre tercih edilebilecek yöntem ya da yöntemleri belirlemek

amaçlanmıştır.

2. GENEL BİLGİLER

Normal dağılım varsayımının geçerli olmadığı çokdeğişkenli veri setlerinde ayırma işlemlerinin yapılabilmesi için temelde üç farklı yaklaşım ele alınmaktadır. Bu yaklaşımlar Doğrusal Ayırma Analizinin (DAA) farklı varsayımlar altında birer genellemesi olarak ele alınmaktadır. Başka bir anlatım ile, Esnek Ayırma Analizi (EAA) yaklaşımları DAA'nin skorlama teknikleri kullanılarak birer genellemeleri olarak ele alınabilir. Bu yöntemler uygulamada; Esnek Ayırma Analizi (EAA), Cezalandırılmış Ayırma Analizi (CAA, Penalized Discriminant Analysis) ve Karma Ayırma Analizi (MDA, Mixture Discriminant Analysis) olarak ele alınmaktadır. Bu yaklaşımlar kısaca aşağıda açıklanmıştır (6, 15, 16, 22).

2.1. DOĞRUSAL AYIRMA ANALİZİ (DAA, FISHER LINEAR DISCRIMINANT ANALYSIS)

Doğrusal Ayırma Analizi (DAA, Discriminant Analysis, Fisher Linear Discriminant Function Analysis) X veri setindeki değişkenlerin iki ve daha fazla gerçek gruba ayrılmasını sağlayan, birimlerin p tane özelliğini ele alarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına optimal düzeyde atanmalarını sağlayacak fonksiyonlar türeten bir yöntemdir (12,18, 23).

DAA aracılığıyla elde edilen diskriminant (ayırma) fonksiyonları, tahmin değişkenlerinin doğrusal bileşenlerinden oluşur. Ayırma fonksiyonları gruplararası farklılığa etki eden tahmin değişkenlerinin hangileri olduğunu ortaya çıkarır. Gruplararası farklılığa etki eden bu değişkenlere de diskriminant (ayırma) değişkenler adı verilir. Ayırma analizinin bir diğer işlevi ise, gruplardan herhangi birine ait olan fakat hangi gruptan geldiği bilinmeyen bir birimin ait olduğu grubu en az hata ile atanmasını sağlamaktır (18, 24).

Bu analiz yardımı ile alt grupları belirleyen p tane değişken yerine, bu değişkenlerin doğrusal birleşimi olan yeni fonksiyonlar tanımlanarak, alt

gruplar arasındaki uzaklığın maksimum yapılması sağlanır. Yeni fonksiyonlardan önemli olanlarının seçilmesi, p boyutlu sistem yerine daha az boyutlu bir sistemin elde edilmesini ve alt grupların bu sistemde incelenebilmesini sağlar. Böylece alt grupların tanımlanması ve yorumlanması mümkün olur (12,18, 26).

Bazı araştırmacılar ayırma analizinde ayırma fonksiyonu katsayılarının hesaplanmasında başvurulan yöntemlere göre Ayırma Analizi isminin başına getirilen ek sözcüklere göre Kanonik Ayırma Analizi, (Canonical Discriminant Analysis), En büyük Benzerlik Ayırma Analizi (Maximum Likelihood Discriminant Analysis), Bayes Ayırma Analizi (Bayesian Discriminant Analysis) gibi isimlerle anmayı uygun bulmaktadırlar (23).

DAA'nin varsayımları şu şekilde özetlenebilir (18, 23):

1. X veri matrisi çok değişkenli normal dağılım göstermelidir.
2. Değişkenlerin varyans ve kovaryansları homojen olmalıdır. X matrisinde yer alan değişkenler ortak kovaryans matrisine sahip çok değişkenli toplumdaki çekilmiş örnek olmalıdır (Homogeneity of variances/covariances).
3. Değişkenlerin ortalamaları ve varyansları arasında bir korelasyon bulunmamalıdır.
4. Değişkenler arasında çoklu bağımlılık bulunmamalıdır.
5. X matrisi gereğinden fazla değişken ve gereksiz değişken içermemeli, g grubun birbirlerinden ayrılmasını sağlayacak kadar doğru ve gerekli değişkenleri içermelidir. X matrisi grupların birbirlerinden ayrılmasında rol oynamayacak gereksiz değişkenler içermemelidir.

Bu varsayımlar ANOVA ve MANOVA uygulanabilmesi için de gerekli olan varsayımlardır.

2.1.1. Çokdeğişkenli Doğrusal Ayırma Analizi

A (π_1) ve B (π_2) , ..., G(π_g) isimli G tane toplum olsun. Bu toplumlardan n birimlik p tane birbirleri ile ilişkili gözlem yapılsın. $\mathbf{X}$ veri matrisinde gözlemler toplumdan topluma az ya da çok farklılıklar gösterir. $\mathbf{X}$ matrisi, π_1 toplumundan alınan örnekler için $\mathbf{X}_1$ gözlem matrisi, π_2 toplumundan alınan örnekler için $\mathbf{X}_2$ gözlem matrisi ve π_g toplumundan alınan örnekler için $\mathbf{X}_g$ gözlem matrisi elde edilir. Örneğin iki çokdeğişkenli normal dağılımdan elde edilen g=2 için örnek gözlem matrisleri aşağıdaki gibi yazılır (23).

π_1 populasyonundan alınan örnekler için $\mathbf{X}_1$ gözlem matrisi,

$$\underset{(p \times n_1)}{\mathbf{X}_1} = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1n_1} \end{bmatrix} \text{ şeklinde,}$$

π_2 populasyonundan alınan örnekler için $\mathbf{X}_2$ gözlem matrisi ise,

$$\underset{(p \times n_2)}{\mathbf{X}_2} = \begin{bmatrix} x_{21}, x_{22}, \dots, x_{2n_2} \end{bmatrix} \text{ şeklinde yazılır.}$$

Her iki toplum için olasılık yoğunluk fonksiyonu $f_1(\mathbf{X})$ ve $f_2(\mathbf{X})$ 'dir.

Çok değişkenli $\mathbf{X}$ gözlem matrisi π_1 ve π_2 toplumlarına ilişkin gözlemleri içeren $\mathbf{X}_1$ ve $\mathbf{X}_2$ gözlem matrislerinin birleştirilmesi ile elde edilir.

Bu veri matrislerinden örnek ortalama vektörleri ve kovaryans matrislerini aşağıdaki gibi hesaplanır.

$$\underset{(p \times 1)}{\bar{\mathbf{x}}_1} = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1j}; \quad \underset{(p \times p)}{\mathbf{s}_1} = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)(x_{1j} - \bar{x}_1)'$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j}; \quad s_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)(x_{2j} - \bar{x}_2)'$$

İncelenen populasyonların aynı kovaryans matrise (Σ) sahip oldukları kabul edilerek, örnek kovaryans matrisleri S_1 ve S_2 'nin birleşimi S_p (pooled variance-covariance matrix) aşağıdaki gibi hesaplanır.

$$S_{\text{pooled}} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}$$

Ortak kovaryans matrisi, $\hat{\Sigma} = E(X - \bar{x}_i)(X - \bar{x}_i)'$ biçiminde de hesaplanabilir.

X gözlem matrisi kullanılarak elde edilen çokdeğişkenli gözlem matrisi tek değişkenli y değerlerine dönüştürülür. Bu y değerleri X gözlem matrisinin doğrusal bileşenleridir.

Doğrusal Ayırma analizinde her bir grup için birer ayırma fonksiyonu aşağıdaki gibi hesaplanır.

$$Y_i = b_{0i} + b_{1i}X_1 + b_{2i}X_2 + \dots + b_{pi}X_p \quad i=1,2 \text{ grup sayısı}$$

Bu fonksiyonda b_{0i} sabit değeri (constant), b_{ij} ise doğrusal bileşenleri belirtmektedir. Doğrusal bileşenlere **kanonik değişkenler** adı da verilmektedir.

Her bir grup için değişkenlere ilişkin doğrusal bileşenler, değişkenlerin ayırma fonksiyonundaki etkinliklerini, belirleyiciliklerini göstermektedir.

Doğrusal bileşenler, $b_{ij} = S^{-1}(\bar{x}_i) \quad i=1,2,g ; j=1,2,p$ biçiminde hesaplanır. b_i katsayılarına grup ayırma fonksiyonu katsayıları ya da kanonik

değişkenler adı verilir. Bazen katsayılar ölçeklendirilerek (scaled) ya da normalize edilerek kullanılmaktadır. Normalizasyonun temel amacı katsayıları genel katsayılar içinde ağırlıklandırarak elemanların kolay yorumlanması sağlanır.

İki tür normalizasyon yaklaşımı vardır.

$$1- \mathbf{b}^* = \frac{\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{b}}} \quad \text{ve} \quad 2- \mathbf{b}^* = \frac{\mathbf{b}}{\mathbf{b}_1} .$$

Burada $\mathbf{b}_1$ en büyük doğrusal bileşendir.

Her iki normalizasyon yaklaşımında da katsayılardan birinin diğerlerine göre önemli olarak ağırlığını artırmak ve diğerlerine göre daha çok önemli değişkenleri modelde etkin olarak görmek amaçlanmaktadır.

Gruplar arasındaki farklılığı maksimize edecek bir ayırma fonksiyonu aracılığı ile grupları birbirinden ayırmak mümkün olacaktır. Bu nedenle ortak bir ayırma fonksiyonu belirlenir. i ve j grupları arasındaki ayırma fonksiyonu;

$$\mathbf{Y} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{X}_1 + \mathbf{b}_2\mathbf{X}_2 + \dots + \mathbf{b}_p\mathbf{X}_p \text{ şeklinde yazılır.}$$

Bu fonksiyondaki $\mathbf{b}_i$ doğrusal bileşenleri de ortalama fark vektörü aracılığı ile aşağıdaki gibi bulunur.

$$\mathbf{b}_i = \mathbf{S}^{-1}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j) \quad i = 1, 2, \dots, g$$

Sabit değer olan $\mathbf{b}_0$ katsayısı ise;

$$\mathbf{b}_0 = -(\mathbf{1} / 2) \bar{\mathbf{x}}' \mathbf{S}^{-1} \bar{\mathbf{x}} \text{ biçiminde hesaplanır.}$$

Her gruba ilişkin ayırma fonksiyonları aşağıdaki gibi hesaplanır:

S ortak kovaryans matrisi ve $\bar{x}_i$ i. grup ortalama vektörü olmak üzere her bir grubun b_i katsayılar vektörü hesaplanır.

$$b_{ij} = S^{-1} \bar{x}_i \quad i = 1, 2, \dots, g \quad j = 1, 2, \dots, p$$

Sabit değer ise ;

$$b_{i0} = -(1/2) \bar{x}_i' S^{-1} \bar{x}_i \quad \text{biçiminde hesaplanır.}$$

Gruplara göre belirlenen sabit ve kanonik katsayılar değişken değerleri ile çarpılarak doğrusal ayırma fonksiyonları (kanonik discriminant functions) hesaplanır.

1. grup için ayırma fonksiyonu;

$Y_1 = b_{10} + b_{11}X_{11} + b_{12}X_{12} + \dots + b_{1p}X_{1p}$ eşitliğinde değerler yerine konularak belirlenir.

2. grup için ayırma fonksiyonu;

$Y_2 = b_{20} + b_{21}X_{21} + b_{22}X_{22} + \dots + b_{2p}X_{2p}$ eşitliğinde değerler yerine konularak belirlenir.

g. grup için ayırma fonksiyonu;

$Y_g = b_{g0} + b_{g1}X_{g1} + b_{g2}X_{g2} + \dots + b_{gp}X_{gp}$ eşitliğinde değerler yerine konularak belirlenir.

p değişkenli g grup arasındaki karesel uzaklığı veren ve Mahalanobis tarafından ileri sürülen D^2 uzaklık aşağıdaki gibi hesaplanır.

$$D_{ij}^2 = (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_i - \bar{x}_j)$$

D^2 uzaklığının i ve j gruplarını birbirinden ayırmada etkin rol oynayıp oynamadığı test edilebilir. Bu amaçla Hotelling T^2 yaklaşımı ile D^2 'nin önemliliği test edilebilir.

$$T^2 = \left(\frac{n_1 * n_2}{n_1 + n_2} \right) D^2$$

T^2 'nin önemliliğinin belirlenmesi için F yaklaşımından yararlanılır.

$$F = \frac{(n_1 + n_2 - p - 1)}{p(n_1 + n_2 - 2)} T^2$$

F'in önemliliği, p, (n_1+n_2-p-1) serbestlik dereceli F dağılımının kritik değerleri kullanılarak belirlenir [$F(\alpha, p, (n_1+n_2-p-1))$].

Bir birimin gerçekte π_1 toplumuna ait olduğu halde π_2 olarak ayrılmasının koşullu olasılığı (18 24, 26);

$$P(2 | 1) = P(X \in R_2 | \pi_1) = \int_{R_2} f_1(x) dx$$

şeklinde hesaplanır. Aynı şekilde bir birimin gerçekte π_2 'ye ait olduğu halde π_1 olarak ayrılmasının koşullu olasılığı;

$$P(1 | 2) = P(X \in R_1 | \pi_2) = \int_{R_1} f_2(x) dx$$

şeklinde hesaplanır. Burada R_1 , birimleri π_1 toplumu olarak ayrılan değişken setini, R_2 ise birimleri π_2 toplumu olarak ayrılan değişken setini göstermektedir.

Toplamları ayırmada ön olasılık (prior probability) değerleri biliniyorsa yukarıdaki eşitlikler aşağıdaki şekilde ifade edilirler.

$$P(1|1) = P(X \in R_1 | \pi_1)p_1 = \int_{R_1} f_1(x) dx p_1$$

$$P(2|2) = P(X \in R_2 | \pi_2)p_2 = \int_{R_2} f_2(x) dx p_2$$

$$P(2|1) = P(X \in R_2 | \pi_1)p_1 = \int_{R_2} f_1(x) dx p_1$$

$$P(1|2) = P(X \in R_1 | \pi_2)p_2 = \int_{R_1} f_2(x) dx p_2$$

Burada $p_1+p_2=1$ olmak zorundadır. Birimleri yanlış ayırmanın ya da sınıflamanın maliyeti ise $c(1|2)$ ve $c(2|1)$ şeklinde gösterilir. Buradan beklenen yanlış sınıflandırma maliyeti ise (ECM, Expected Cost of Misclassification)

$$ECM = c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$$

şeklinde hesaplanır.

Ayırma analizinde sınıflandırma kuralının temelinde ECM 'değerinin minimize edilmesi önemlidir. R_1 ve R_2 alanları aşağıdaki şekilde ifade edilir ise ECM minimize edilir.

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right),$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right)$$

Eğer yanlış sınıflandırma maliyetlerini ve ön olasılıkları eşit olarak verildiğinde aşağıdaki şekilde ifadeler elde edilir.

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1, \quad R_2 : \frac{f_1(x)}{f_2(x)} < 1$$

Ayırma analizinde kullanılan bir diğer kriter ise beklenen yanlış sınıflandırma maliyetini elimine edip Toplam Yanlış Sınıflandırma olasılığını (TPM, Total Probability of Misclassification) bulmaktır ve olasılık aşağıdaki gibi hesaplanır.

$$TPM = p_1 \int_{R_2} f_1(x) dx + p_2 \int_{R_1} f_2(x) dx$$

Yanlış sınıflandırma maliyetleri eşit olduğu zaman TPM'nin minimize edilmesi ECM' nin minimize edilmesi ile denktir. Eşit maliyetler söz konusu olduğunda yeni gözlenen ve X_0 veri setine sahip bir birimin atanması Bayes kuralına göre hesaplanan posterior olasılıklara göre yapılır ve aşağıdaki şekilde hesaplanır.

$$P(\pi_1 | x_0) = \frac{p_1 f_1(x_0)}{p_1 f_1(x_0) + p_2 f_2(x_0)}$$

$$P(\pi_2 | x_0) = \frac{p_2 f_2(x_0)}{p_1 f_1(x_0) + p_2 f_2(x_0)}$$

Yukarıda görüldüğü gibi her iki eşitlikte paydalar aynı olduğundan yeni gözlemin atanması aşağıdaki şekilde yapılır.

Eğer $P(\pi_1 | x_0) > P(\pi_2 | x_0)$ yeni gözlem birimi birinci topluma, tersi söz konusu ise ikinci topluma atanır (12,18,23,26).

$$R_2 : \frac{f_1(x)}{f_2(x)} = \exp \left[-\frac{1}{2}(x-\mu_1)' \Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)' \Sigma^{-1}(x-\mu_2) \right] < \left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right)$$

Burada gerekli işlemler yapıldığında yeni gözlenen $\mathbf{X}_0$ veri setine sahip bir birimin hangi topluma atanacağı aşağıdaki şekilde belirlenir.

Eğer;

$$(\mu_1 - \mu_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2}(\mu_1 - \mu_2)' \Sigma^{-1}(\mu_1 + \mu_2) \geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

ise yeni gözlenen birim birinci topluma, tersi söz konusu olduğunda ikinci topluma atanır.

2.2. KARESEL AYIRMA ANALİZİ (KAA, QUADRATIC DISCRIMINANT ANALYSIS)

Verilerin normal dağıldığı ancak, grupların varyans-kovaryans matrislerinin birbirlerinden farklı olduğu durumda veri matrislerinin ayırma analizinde Karesele Ayırma Analizi (KAA, Quadratic Discriminant Analysis) uygulanır (18, 23, 24, 26).

KAA'da katsayıların hesaplanmasında ortak kovaryans matrisi yerine ($\mathbf{S}$) grupların kovaryans matrislerinin farkları alınır. Örneğin iki grup için KAA'da ayarlanılan kovaryans matrisi ($\mathbf{S}_1 - \mathbf{S}_2$) biçiminde hesaplanır.

Burada $\Sigma_1 \neq \Sigma_2$ olduğundan minimum ECM alanları ise aşağıdaki gibi gösterilir.

$$R_1 : -\frac{1}{2} \mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1})\mathbf{x} - k \geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

$$R_2 : -\frac{1}{2} \mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1})\mathbf{x} - k < \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

Burada;

$$k = \frac{1}{2} \ln \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$$

şeklinde hesaplanır.

Burada gerekli işlemler yapıldığında yeni gözlenen X_0 veri setine sahip bir birimin hangi topluma atanacağı aşağıdaki şekilde belirlenir.

Eğer;

$$-\frac{1}{2} x_0' (\Sigma_1^{-1} - \Sigma_2^{-1}) x_0 + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1}) x_0 - k \geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

ise yeni gözlenen birim birinci topluma, tersi söz konusu olduğunda ikinci topluma atanır.

Genelde toplum parametreleri bilinmediğinden yukarıdaki atama fonksiyonunda parametre değerleri yerine bunların tahminçileri kullanılır. Bu durumda yukarıdaki eşitlik aşağıdaki şekilde ifade edilir.

$$-\frac{1}{2} x_0' (S_1^{-1} - S_2^{-1}) x_0 + (\bar{x}_1' S_1^{-1} - \bar{x}_2' S_2^{-1}) x_0 - k \geq \ln \left[\left(\frac{c(1|2)}{c(2|1)} \right) \left(\frac{p_2}{p_1} \right) \right]$$

$$k = \frac{1}{2} \ln \left(\frac{|S_1|}{|S_2|} \right) + \frac{1}{2} (\bar{x}_1' S_1^{-1} \bar{x}_1 - \bar{x}_2' S_2^{-1} \bar{x}_2)$$

Aynı şekilde TPM kuralını uyguladığımızda atamaya aşağıdaki şekilde karar verilir.

$p_1 f_1(x_0) > p_2 f_2(x_0)$ ise yeni gözlenen birim birinci topluma, tersi söz konusu ise ikinci topluma atanır. Bu eşitlik aşağıdaki şekilde yazılırsa aynı durum yine geçerli olmaktadır.

$$\ln[p_1 f_1(x_0)] > \ln[p_2 f_2(x_0)]$$

Gerekli işlemler yapılarak ayırma fonksiyonları aşağıdaki şekilde elde edilir.

$$d_i^Q(x) = -\frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) + \ln p_i \quad i=1,2$$

Yukarıdaki ayırma fonksiyonunda, parametre tahmincileri yerine konularak aşağıdaki eşitlik elde edilir.

$$d_i^Q(x) = -\frac{1}{2} \ln|S_i| - \frac{1}{2} (x - \bar{x}_i)' S_i^{-1} (x - \bar{x}_i) + \ln p_i \quad i=1,2$$

Bu ayırma fonksiyonları ile değişkenlerin fonksiyondaki katsayıları elde edilerek yeni gözlenen X_0 veri setine sahip bir birimin hangi topluma atanacağı belirlenir. $d_1(x_0) > d_2(x_0)$ ise yeni gözlenen birim birinci topluma tersi söz konusu ise ikinci topluma atanır (12, 18, 19, 23-26).

2.3. ESNEK AYIRMA ANALİZLERİ (EAA, FLEXIBLE DISCRIMINANT ANALYSES)

Çokdeğişkenli ancak Çokdeğişkenli Normal dağılıma uymayan toplumlardan alınan verilerin sınıflandırılmasında yararlanılan parametrik olmayan yöntemleri içine alan analize verilen isimdir. Bu sınıfta yer alan yöntemler, parametrik çokdeğişkenli ayırma yöntemlerinden farklı olarak ya çok az ya da hemen hemen hiçbir varsayım gerektirmeyen yöntemlerdir (6, 11, 15).

Bu yöntemler içinde Esnek ayırma Analizi, Penalized Ayırma Analizi ve Karma Ayırma Analizi yöntemleri yer almaktadır (13-15).

Parametrik olmayan ayırma yöntemleri içinde MARS (Multivariate Adaptive Regression Splines) ve BRUTO modelleri ayırma işlevlerini yerine getirmede güçlü yöntemler olarak ileri sürülmüştür (14-16).

2.3.1- Esnek Ayırma Analizi (FDA, Flexible Discriminant Analysis)

DAA yöntemi ile hesaplanan ayırma fonksiyonlarının katsayıları, gerçek gözlemlerin birer doğrusal bileşeni olan doğrusal bir eşitlik ile regresyon formunda ifadesi olarak ele alınabilir. DAA yaklaşımı parametrik olmayan vektör uzayına genişletilmiş gizli bir doğrusal regresyon yöntemi olarak ele alınır ve yeniden düzenlenirse bu yöneme Esnek Ayırma Analizi adı verilmektedir.

Ele alınan problemin K sınıflı bir problem olduğu göz önüne alındığında N birimden elde edilen gözlemsel verinin p boyutlu bir uzayda konuşlandığı varsayalım. Bu durumda K sınıf için Y Gösterge Cevap Matrisi (*indicator response matrix* Y (K=5)) aşağıdaki gibi yazılabilir (6, 9,15, 16).

$$\begin{array}{c}
 \\
 g_1 = 2 \\
 g_2 = 1 \\
 g_3 = 1 \\
 g_4 = 5 \\
 g_5 = 4 \\
 \vdots \\
 g_N = 3
 \end{array}
 \begin{pmatrix}
 C_1 & C_2 & C_3 & C_4 & C_5 \\
 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 1 & 0 \\
 \vdots & & & & \\
 0 & 0 & 1 & 0 & 0
 \end{pmatrix}$$

X bağımsız değişkenler matrisine göre Y gösterge matrisinin çokdeğişkenli doğrusal regresyonu her bir veri noktasında elde edilebilir. Yeni

test noktasının ait olduđu sınıf en büyük tahmin değere (largest fitted value) eşit olacaktır.

Her bir sınıfın skor değerlerini belirten yeni bir fonksiyon aşağıdaki gibi yazılabilir.

$$\theta : 1, \dots, K \rightarrow \mathbb{R}$$

Ayrıca $\mathbf{X}$ 'e göre doğrusal regresyon yaklaşımı ile optimal olarak elde edilebilen θ değerine gereksinim vardır.

β ve θ değerleri aşağıdaki gibi tahmin edilebilir.

$$\min_{\beta, \theta} \sum_{i=1}^N (\theta(g_i) - x_i^T \beta)^2$$

Bu minimizasyon değeri sınıflar arasında tek boyutlu bir ayırma (*one-dimensional separation*) olanağı sağlar.

DAA yaklaşımını parametrik olmayan problemlere uygulayabilmek amacıyla DAA eşitliklerinin aşağıdaki gibi yeniden düzenlenmesi yapılabilir.

Genel olarak sınıf sayısı daha az sayıda $L \leq K - 1$ sınıf oluşturacak biçimde yazılabilir. $\theta_1, \theta_2, \dots, \theta_L$ bağımsız skor değerlerini işaret edecek biçimde

$$\eta_l(X) = X^T \beta_l, l = 1, \dots, L$$

gibi doğrusal bir form elde edilebilir.

Bu durumda ASR (Average Squared Rediduals) enküçüklenecek biçimde aşağıdaki gibi yazılabilir.

$$ASR = \frac{1}{N} \sum_{l=1}^L \left[\sum_{i=1}^N (\theta_l(g_i) - x_i^T \beta_l)^2 \right]$$

Deneme setindeki sıfır ortalama ve birim varyans yapılarına ilişkin önemsiz çözümlerden kaçınmak için θ üzerine bazı kısıtlamalar getirilebilir.

β_l parametrelerinin, Fisher DAA ile elde edilen v_l Fisher ayırma koordinatlarına bir sabit değer ilavesi ile elde edilecek değer ile eşit olduğu gösterilmiştir (15). ASR çözümlerinden D^2 (Mahalanobis) uzaklıkları hesaplanabilir.

DAA içinde η_l uzayındaki sınıfları izleyerek adımsal doğrusal regresyon analizleri yapılabilir. Nonparametrik sınıflandırmaları yapabilmek amacıyla; DAA'da elde edilecek doğrusal regresyon denklemleri ile bazı çok daha fazla esnek nonparametrik denklemlerle yer değiştirerek (örneğin splin'ler ya da kernel'ler) DAA sınıflandırma prosedürlerini genişletmeye izin verecek yeni formüller elde edilebilir.

Bu çok karmaşık denklemlerin kullanımı, sınıflar arasındaki ayırımların daha kolay yapılabileceği çok yüksek boyutlu uzaylarda verilerin iz düşümünü (projeksiyon) elde etmeyi sağlar.

MARS, Splin'ler olarak isimlendirilecek olan bu modeller, bazı düzenlemelere izin vererek sınıfların üst üste gelmelerini önlemeyi sağlar.

MARS modeline göre ASR değeri aşağıdaki gibi elde edilebilir.

$$ASR = \frac{1}{N} \sum_{l=1}^L \left[\sum_{i=1}^N (\theta_l(g_i) - \eta_l(x_i))^2 + \lambda J(\eta_l) \right]$$

şeklinde gösterilmektedir.

Bu formülde J smooting spline'lar, eklemeli spline'lar ve lower-order anova spline modeller gibi doğrusal olmayan regresyonun bazı formları için düzenleyici bir ayraçtır. Aşağıdaki algoritma yardımı ile η_l regresyon denklemleri ve θ_l optimal skorları hesaplanabilir.

1. Y tahmin değerlerini elde etmek için X'e göre Y değerleri parametrik olmayan regresyon denklemleri elde edilir. S_λ , $\hat{Y} = S_\lambda Y$ denkleminin doğrusal operatörüdür. $\eta^*(x)$ ise model olarak belirlenen regresyon fonksiyonları vektörüdür.

2. Optimal skorlar $Y^t \hat{Y}$ 'nin Θ özvektörler matrisinin hesaplanması ile elde edilir. Bu vektörler aşağıdaki gibi normalize edilir.

$$\Theta^T \left(\frac{1}{N} Y^T Y \right) \Theta = I$$

Optimal skorlar $\eta(x) = \Theta_\eta^T * (x)$ eşitliği kullanılarak model güncellenir.

2.3.2- Cezalandırılmış Ayırma Analizi (PDA, Penalized Discriminant Analysis)

Cezalandırılmış Ayırma Analizi, DAA yaklaşımının net olarak belirlenmemiş/ kabaca ortaya konmuş koordinatlar içinde sınırlandırılmış olan ve temelde bir regresyon yöntemi olarak ele alınan bir yaklaşımdır. Bu yaklaşımda katsayıların belirli sınırlandırmalar ile (penalized) düzleştirilmesi (smoothing) yapılarak uzayda uygun formların oluşturulması sağlanır. EAA ile genişletilmiş temel set genellikle o kadar geniş ki, düzgün hale getirmek gerekmektedir (6, 13-16).

EAA, DAA'da gerçekleştirilen regresyonun $h(x)$ temel fonksiyonlarının daha geniş açıklayıcı değişkenler seti üzerinden genişletilmiş şeklidir ve ASR,

$$ASR(\{\theta_l, \beta_l\}_{l=1}^L) = \frac{1}{N} \sum_{l=1}^L \left[\sum_{i=1}^N (\theta_l(g_i) - h^T(x_i)\beta_l)^2 + \lambda \beta_l^T \Omega \beta_l \right]$$

şeklinde tanımlanır. Burada uygulanan işlem, açıklayıcı değişkenlerin, temel $h(x)$ fonksiyonu aracılığıyla genişletilmesi ve Cezalandırılmış Mahalanobis uzaklığı ile DAA'nın uygulanmasıdır.

Mahalanobis uzaklığı,

$$D(x, \mu) = (h(x) - h(\mu))^T (\Sigma_W + \lambda \Omega)^{-1} (h(x) - h(\mu))$$

olarak ifade edilir. Optimal indirgenmiş boyut ise,

$$\max_u \frac{u^T \Sigma_{bet} u}{u^T (\Sigma_w + \Omega) u}$$

şeklinde ifade edilir.

2.3.3- Karma Ayırma Analizi (MDA, Mixture Discriminant Analysis)

Karma Ayırma analizi her bir sınıfın Gaussian mixture yaklaşımı içinde yer almasına izin veren çok sınıflı prototip sınıfların oluşturulmasını sağlayan bir yaklaşımdır. MDA yaklaşımında her bir sınıf Gaussian mixture yaklaşımı içinde gösterilerek çoklu sınıf prototipleri belirlenir. Bu üçüncü durumda ise farklı merkezler (centroids) ile iki ya da daha fazla Gaussian mixture yaklaşımı tarafından her bir sınıfın modellenmesidir, fakat her parçanın (compenent) aynı kovaryans matrisinde sınıflar içinde ve sınıflar arasında her bileşenin Gaussian mixture yaklaşımı olmak şartı aranır. Bu genişleme Mixture Discriminant Analysis ya da MDA olarak adlandırılır (6, 7, 14, 16, 17).

Karma ayırma analizi DAA'nın çok gruplu prototiplere uygulanmasıdır. Gruplarda yer alan gözlemlerin Gaussian mixture yaklaşımı içinde yer aldığını varsayan bir yöntemdir ve açıklayıcı değişkelere göre bir gözlemin k. gruba ait olma olasılığı

$$P(X|G = k) = \sum_{r=1}^{R_k} \pi_{kr} \phi(X; \mu_{rk}, \Sigma)$$

şeklinde ifade edilir. Burada R_k , k. grup için prototip sayısını, π_{kr} karma oranını, Σ ise her bir Gaussian mixture yaklaşımı dağılan gözlem için varyans yapısını göstermektedir ve her bir Gaussian mixture yaklaşımı için tekdir. Karma parametrelerin en çok olabilirlik tahmincileri EM algoritması ile elde edilir.

Bu analizde kullanılan Σ 'nin her bir Gaussian mixture yaklaşımı için tek olma varsayımına göre; DAA'nın L boyutlu bir uzaydan indirgenmiş daha küçük boyutlu bir uzaydaki uygulaması, karma ayırma analizinden elde edilen en çok olabilirlik tahmincileri ile benzer sonuçlar verir. Bu özelliğin karma modellerine uygulanmasında bir kez daha EM algoritması kullanılır.

2.3.4- MARS VE BRUTO Modelleri

Burada daha çok uyumlu paratmetrik olmayan regresyonlar için iki strateji üzerinde durulacaktır. Bunlar MARS ve BRUTO dur. MARS, etkileşimleri hiyerarşik bir tarzda tutabilen/yakalayabilen/belirleyebilen bir uyumlu regresyon tekniğidir. BRUTO ise düzleştirilmiş splin'ler (smoothing splines) kullanarak, $\alpha + \sum_1^p f_j(X_j)$ formunun eklemeli (additive) modelini tahmin etmek için uyumlu bir metottur (6, 10, 15, 16).

2.4. MARS: Multivariate Adaptive Regression Splines

MARS (Multivariate Adaptive Regression Splines) tekniği ilk olarak 1991 yılında Jerome Friedman tarafından ortaya atılmıştır. Mars adımsal bir regresyon yöntemidir ve regresyon setinin performansını geliştirmek için tekrarlamalı ayırma metodunun ve adımsal doğrusal regresyonun genelleştirilmiş bir hali olarak düşünülebilir.

MARS, bağımlı ve bağımsız değişkenler arasındaki doğrusal olmayan ilişkileri doğrusal yapıya dönüştürmek için uygun transformasyonlar kullanan ve bağımsız değişkenler arasındaki interaksionları belirlemeyerek sınıflama işlevleri yerine getirebilen bir yöntemdir. MARS modelleri bağımlı ve bağımsız değişkenler arasında oluşturulan fonksiyonel ilişkiye ait hiçbir varsayım içermeyen parametrik olmayan bir yöntemdir (1, 11, 13-15).

MARS modelleri aracılığı bağımsız değişkenlerin farklı aralıklarına karşılık gelen dönüştürülmüş tek değişkenli ya da çok değişkenli interaksion terimlerinden meydana gelen **temel fonksiyonları** kullanarak esnek regresyon modeli oluşturulur. Model oluşturulurken regresyon modellerindeki hem ileri doğru hem de geriye doğru ilerleme algoritmalarına benzer algoritmalar kullanılır. Ayrıca parçalı temel fonksiyonlar ve bu fonksiyonların kombinasyonlarını kullanılarak Mars modelleri elde edilebilir.

Çok büyük bir model bulununcaya kadar yani modelin karmaşıklığı maksimum seviyeye ulaşincaya kadar eklenen temel fonksiyonlarla model büyütülür. Eklenen temel fonksiyonlarla model daha karmaşık ve daha esnek bir model haline gelir. Temel fonksiyon sayısı maksimum seviyeye ulaşincaya kadar bu adım sürer.

$$\text{MARS Modeli genel olarak, } Y = \beta_0 + \sum_{k=1}^K \beta_k B_k(X) \text{ biçiminde yazılır.}$$

Temel fonksiyon modeldeki bağımsız değişkenlerin interaksiyonlarını ve doğrusal olmayan dönüşümlerini hesaba katan bir fonksiyondur. Temel fonksiyonlar oluşturulurken tüm değişken ve değişkenlerin kombinasyonları tek tek ele alınır. Bu durumda, her bir bağımsız değişken ve bu değişkenlerin birbirleri ile olan kombinasyonlarının tanımlı bulunduğu aralıklardaki tüm olası değerler birer düğüm olarak düşünülüp, mümkün olan tüm olası temel fonksiyonlar belirlenir.

Temel fonksiyon ise aşağıdaki gibi tanımlanır.

$$B_k(X) = \prod_{j=1}^{J_k} [s_{kj}(x_{wkj} - t_{kj})]_+$$

J_k interaksiyon derecesini göstermektedir. $[.]_+ = \max[0, .]$, $s_{kj} : \in [\pm 1]$, t_{kj} : düğüm değeri ve x_{wkj} : bağımsız değişken değerini göstermektedir.

Düğüm Değerinin Elde Edilmesi

Bağımlı ve bağımsız değişken arasındaki ilişkiler doğrusal veya doğrusal olmayan şekilde olabilir. Aynı bağımsız değişken üzerinde, ilişkinin şeklinin değiştiği bağımsız değişken değerine düğüm değeri denir.

Bağımsız değişken değerinin doğrunun eğimini değiştirmeyen en son değeri, düğüm değeri olarak alınır. Bu düğüm değerlerinde hata kareler toplamı en küçük değerini alır.

2.Adım:Geriye Doğru Adım

Ardışık iki düğüm değerini birleştirerek çizilen doğrunun eğimi, modelde β ile gösterilen regresyon katsayısıdır. Bu haliyle Mars modeli parçalı regresyon modeline benzer. Oluşturulan her temel fonksiyonda bağımsız değişkenden seçilen uygun düğüm değerleri bağımlı değişken ile bağımsız

değişken arasında monoton transformasyonlar yapmadan aralarındaki ilişkiyi doğrusal (Smoothing) hale getirirler. Bu adıma budama adımı denir. Biaslar ve varyansların optimal bir dengesi bulununcaya kadar temel fonksiyonlar modelden çıkarılır. Oluşturulan maksimum model budanarak yani önemli bağımsız değişkenler ve bu değişkenlerin interaksiyonları belirlenerek, hata kareler toplamı minimum olan optimum model oluşturulur. Budama algoritması olarak en yaygın olarak genelleştirilmiş hata kareler ortalaması (GCV) kullanılır. Çalışmalarda MARS'ın uygulanabileceği örnek büyüklüğünün $50 \leq N \leq 1000$ arasında ve değişken sayısının $3 \leq p \leq 20$ arasında olması istenir (11).

2.4.1 MARS Algoritmaları

EAA/MARS yöntemi, uygulamalarında iki farklı algoritma yaygın olarak kullanılmaktadır. Bu algoritmalar, literatürde yaygın adı ile degree=1 ve degree=2 olarak isimlendirilmektedir. Bu algoritmalar aşağıda kısaca açıklanmıştır.

2.4.1.1. Yinelemeli Bölünme Regresyonu (Recursive Partitioning Regression EAA/MARS degree=1)

Yinelemeli Bölünme Regresyonu modeli (EAA/MARS degree=1) olarak isimlendirilir. Yinelemeli bölünme özellikle basit parçalı (piecewise) sabit tahmini kullanıldığında güçlü bir yöntemdir. Fonksiyonlar düşük bölge boyutunu (dimensionality) kullanma özelliğine sahiptir. Bağımlı değişken üzerindeki küçük etkiye sahip olan değişkenlerin bölünme için kullanılması uygun değildir. Bu bağımsız değişken alt küme seçiminde artış göstermektedir (11).

Yinelemeli bölünme regresyonun diğer bir özelliği etkileşim etkilerinin marjinal sonuçlarıdır. Parçalı sabit tahminleri kullanan yinelemeli bölünme modelleri çok basit olduğundan ve ikili ağaç (binary tree) ile temsil edilebildiği için oldukça kolay yorumlanabilir niteliktedir. Aynı zamanda hesaplama oldukça hızlıdır.

Yinelemeli bölünme, çok değişkenli fonksiyon tahmini için en uyumlu metod olmasına rağmen onun etkinliğini kısıtlayan değer kaybedebilmektedir. Bunlardan en önemlisi, tahmin fonksiyonun alt bölge sınırlarında sürekli olmayışıdır. Bu özellikle doğru temeli oluşturan fonksiyon sürekli olduğunda tahmin doğruluğunu sınırlandırır. Yalnızca fonksiyonun sürekliliğini genellikle de tahminin doğruluğunu artırmak yeterlidir (11).

EAA/MARS degree=1 modeli recursive partitioning model olarak aşağıdaki algoritma ile tahminlenir.

Algoritma 1: (recursive partitioning)

```

 $B_1(x) \leftarrow 1$ 
For  $M = 2$  to  $M_{max}$  do :  $lof^* \leftarrow \infty$ 
  For  $m = 1$  to  $M - 1$  do :
    For  $v = 1$  to  $n$  do :
      For  $t \in \{x_{vj} \mid B_m(x_j) > 0\}$ 
         $g \leftarrow \sum_{i \neq m} a_i B_i(x) + a_m B_m(x) H[+(x_v - t)] + a_M B_M(x) H[-(x_v - t)]$ 
         $lof \leftarrow \min_{a_1, \dots, a_M} LOF(g)$ 
        if  $lof < lof^*$  then  $lof^* \leftarrow lof$ ;  $m^* \leftarrow m$ ;  $v^* \leftarrow v$ ;  $t^* \leftarrow t$  end if
      end for
    end for
  end for
   $B_M(x) \leftarrow B_{m^*}(x) H[-(x_{v^*} - t^*)]$ 
   $B_{m^*}(x) \leftarrow B_{m^*}(x) H[+(x_{v^*} - t^*)]$ 
end for
end algorithm

```

EAA/MARS degree=2 modeli forward stepwise modeli olarak bilinir ve aşağıdaki algoritma ile tahminler yapılır.

Algoritma 2 (MARS-forward stepwise)

```

 $B_1(x) \leftarrow 1; \leftarrow 2$ 
Loop until  $M > M_{max} : lof^* \leftarrow \infty$ 
For  $m = 1$  to  $M - 1$  do :
  For  $v \notin \{v(k, m) | 1 \leq k \leq K_m\}$ 
    For  $t \in \{x_{vj} | B_m(x_j) > 0\}$ 
       $g \leftarrow \sum_{i=1}^{M-1} a_i B_i(x) + a_M B_m(x) [+(x_v - t)] + a_{M+1} B_m(x) [-(x_v - t)] +$ 
       $lof \leftarrow \min_{a_1, \dots, a_{M+1}} LOF(g)$ 
      if  $lof < lof^* \leftarrow lof; m^* \leftarrow m; v^* \leftarrow v; t^* \leftarrow t$  end if
    end for
  end for
end for
 $B_M(x) \leftarrow B_{m^*}(x) [+(x_{v^*} - t^*)] +$ 
 $B_{M+1}(x) \leftarrow B_{m^*}(x) [-(x_{v^*} - t^*)] +$ 
 $M \leftarrow M + 2$ 
end loop
end algorithm

```

EAA/MARS degree=3 modeli backward stepwise modeli olarak bilinir ve aşağıdaki algoritma ile tahminler yapılır.

Algoritma 3 (MARS- backwards stepwise)

```

 $J^* = \{1, 2, \dots, M_{max}\}; K^* \leftarrow J^*$ 
 $lof^* \leftarrow \min_{\{a_j | j \in J^*\}} LOF\left(\sum_{j \in J^*} a_j B_j(x)\right)$ 
For  $M = M_{max}$  to 2 do :  $b \leftarrow \infty; L \leftarrow K^*$ 
  For  $m = 2$  to  $M$  do :  $K \leftarrow L - \{m\}$ 
     $lof \leftarrow \min_{\{a_k | k \in K\}} LOF\left(\sum_{k \in K} a_k B_k(x)\right)$ 
    if  $lof < b$  then  $b \leftarrow lof; K^* \leftarrow K$  end if
    if  $lof < lof^*$  then  $lof^* \leftarrow lof; J^* \leftarrow K$  end if
  end for
end for
end Algorithm

```

2.5. BRUTO: UYUMLU EKLEMELİ MODELLEME (ADAPTIVE ADDITIVE MODELLING)

Burada göz önüne alınan model eklemelidir ve MARS modelinden daha kısıtlayıcıdır. Diğer taraftan çok sayıda açıklayıcı değişkenlerin olduğu durumlarda daha etkin olabilen düzleştirici tarzdaki açıklayıcı değişkenlerle ilişkili bir yöntemdir (10, 11, 15, 16).

BRUTO fonksiyon tahmin modeli basit biçimde aşağıdaki gibi yazılabilir.

$$\sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j^T h_j(x_{ij}) \right)^2 + \sum_{j=1}^p \lambda_j \beta_j^T \Omega_j \beta_j$$

Burada sabit λ_j için çözümler aşağıdaki tahmin denklem seti ile bulunabilir.

$$f_j = S_j(\lambda_j) \left\{ y_i - \sum_{k \neq j} f_k(x_{ik}) \right\}$$

Burada S_j , j'inci açıklayıcı değişkene karşılık gelen düzleştirmeyi sağlayan düzleştirme spline operatörlerini gösterir. Bu denklemler, budama (backfitting) olarak da bilinen yinelemeli Gauss-Seidel algoritması aracılığı ile tahmin edilirler.

p değişkenli bir model için p-terimli eklemeli tüm p düzleştirme parametrelili λ_j 'ler GCV kriterine göre optimize edilebilir. Hastie bu yaklaşıma BRUTO modeli adı vermiş ve eklemeli bir model için GCV'yi aşağıdaki gibi belirlemiştir (15, 16).

$$GCV(\lambda) = \frac{\sum_{i=1}^N \left\{ y_i - \sum_j \hat{f}_{j,\lambda_j}(x_{ij}) \right\}^2}{N \left\{ 1 - \text{tr}S(\lambda) / N \right\}^2}$$

Burada λ_j ekleme operatörüdür. BRUTO, GCV türü istatistikleri aşağıdaki eşitlikteki gibi minimize etmeyi amaçlar.

$$GCV^B(c, \lambda) = \frac{\sum_{i=1}^N \left\{ y_i - \sum_j \hat{f}_{j, \lambda_j}(x_{ij}) \right\}^2}{N \left(1 - \left[1 + c \cdot \sum_1^p df_j \right] / N \right)^2}$$

Burada $df_i = trS_j(\lambda_j) - 1$, j'inci düzleştirme teriminde kullanılan df 'iye yaklaşık olan bir ölçüdür.

Böylece $GCV^B(1, \lambda)$, sadece $O(N)$ hesaplamalarını gerektiren $1 + \sum_j \{trS_j(\lambda_j) - 1\}$ tarafından $trS(\lambda)$ 'ya yakınsar. $GCV^B(1, \lambda)$ 'i j'inci terime ilişkin bir formda aşağıdaki gibi yazılabilir:

$$GCV^B(1, \lambda) = \frac{\sum_{i=1}^N \left\{ r_i^{(j)} - f_{j, \lambda_j}(x_{ij}) \right\}^2}{N \left(1 - \left[df^{(j)} + trS_j(\lambda_j) \right] / N \right)^2}$$

Burada $r^{(j)} = y_i - \sum_{k \neq j} \hat{f}_k(x_{ij})$, modelde diğer tüm ilişkileri sabitleyen kısmi artıktır.

GCV kriteri yardımıyla optimum düzleştirilmiş modeller bulunabilirler. Böylece BRUTO algoritması sınıflandırma için düzleştirilmiş en uygun indirgenmiş modeli seçerek tahminler yapılmasına izin verir. Diğer GCV kriterleri ile budama prosedürlerine benzer biçimde tüm model genellenir (10, 11, 13-16).

3- GEREÇ VE YÖNTEMLER

3.1- Veri Türetimi (simülasyon)

Araştırmada türetilmiş verilerden yararlanılmıştır. Veri türetimi belirli koşullar göz önüne alınarak bu koşullara uygun biçimde R 2.3.1 istatistik paket programı (The R Project for Statistical Computing) kullanılarak yapılmıştır (7, 21, 25).

Veri türetimi için 1 bağımlı değişken ve 4 bağımsız değişken içeren ortalama vektörleri ve kovaryans matrisleri farklı olan çokdeğişkenli normal dağılımdan X_i veri matrisleri türetilerek yapılmıştır (7, 21, 23).

Analiz edilecek veri setinde grup sayısı (g) g=2, 3 ve 4 olacak şekilde X veri matrisleri oluşturulmuştur. Her bir türetime ilişkin veri setlerini elde etmek için grup sayısı dikkate alınarak farklı parametrelili X_i ($i=1, g$) veri matrisleri alt alta birleştirildi. Örneğin g=2 için türetilen $X1$ ve $X2$ veri matrisleri g=3 için türetilen $X1$, $X2$ ve $X3$ veri matrisleri alt alta birleştirildi. Bağımlı değişken y vektörü elemanları ise her grup için 1, 2; 1, 2, 3 ve 1, 2, 3, 4 rakamlarından oluşacak biçimde birleştirilen veri matrislerinin sıralarına eşit sayıda belirlendi. Örneğin 2 gruplu y bağımlı değişken vektörü $y' = (1, 1, \dots, 1, 2, 2, \dots, 2)$ biçiminde, 3 gruplu bağımlı değişken kodlarını içeren y vektörü $y' = (1, 1, \dots, 1, 2, 2, \dots, 2, 3, 3, \dots, 3)$ biçiminde belirlendi.

Türetimlerde farklı grup ortalama vektörleri kullanıldı. Veri setlerinin oluşturulmasında yararlanılan grup ortalama vektörleri aşağıdaki gibi belirlendi.

$$\begin{array}{llll} \mu_1 = (0,0,0,0) & \mu_2 = (0.5,0.5,0.5,0.5) & \mu_3 = (1,1,1,1) & \mu_4 = (2,2,2,2) \\ \mu_5 = (0.5,1,1.5,2) & \mu_6 = (0.5,1,2,3) & \mu_7 = (2.5,3,3.5,4) & \mu_8 = (3,3,3,3) \\ \mu_9 = (4,4,4,4) & \mu_{10} = (4.5,5,5.5,6) & \mu_{11} = (6,6,6,6) & \end{array}$$

Yine türetimlerde farklı varyans ve kovaryans düzeylerinin yöntemlerin sınıflama/gruplandırma performansına etkilerini incelemek için farklı kovaryans matrisleri düzenlendi. Bu kovaryans matrislerinde varyansların 1, 3 ve 10 olduğu fakat kovaryansların 0, 0.25, 0.75 ve 0.90 olduğu durumlar belirlendi.

Türetimde yararlanılan kovaryans matrisleri aşağıdaki gibi belirlendi.

$$\begin{aligned}
 V1 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & V2 &= \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} & V3 &= \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} \\
 V4 &= \begin{bmatrix} 1 & .25 & .25 & .25 \\ .25 & 1 & .25 & .25 \\ .25 & .25 & 1 & .25 \\ .25 & .25 & .25 & 1 \end{bmatrix} & V5 &= \begin{bmatrix} 3 & .25 & .25 & .25 \\ .25 & 3 & .25 & .25 \\ .25 & .25 & 3 & .25 \\ .25 & .25 & .25 & 3 \end{bmatrix} & V6 &= \begin{bmatrix} 10 & .25 & .25 & .25 \\ .25 & 10 & .25 & .25 \\ .25 & .25 & 10 & .25 \\ .25 & .25 & .25 & 10 \end{bmatrix} \\
 V7 &= \begin{bmatrix} 1 & .75 & .75 & .75 \\ .75 & 1 & .75 & .75 \\ .75 & .75 & 1 & .75 \\ .75 & .75 & .75 & 1 \end{bmatrix} & V8 &= \begin{bmatrix} 3 & .75 & .75 & .75 \\ .75 & 3 & .75 & .75 \\ .75 & .75 & 3 & .75 \\ .75 & .75 & .75 & 3 \end{bmatrix} & V9 &= \begin{bmatrix} 10 & .75 & .75 & .75 \\ .75 & 10 & .75 & .75 \\ .75 & .75 & 10 & .75 \\ .75 & .75 & .75 & 10 \end{bmatrix} \\
 V10 &= \begin{bmatrix} 1 & .90 & .90 & .90 \\ .90 & 1 & .90 & .90 \\ .90 & .90 & 1 & .90 \\ .90 & .90 & .90 & 1 \end{bmatrix} & V11 &= \begin{bmatrix} 3 & .90 & .90 & .90 \\ .90 & 3 & .90 & .90 \\ .90 & .90 & 3 & .90 \\ .90 & .90 & .90 & 3 \end{bmatrix} & V12 &= \begin{bmatrix} 10 & .90 & .90 & .90 \\ .90 & 10 & .90 & .90 \\ .90 & .90 & 10 & .90 \\ .90 & .90 & .90 & 10 \end{bmatrix}
 \end{aligned}$$

Bu ortalama vektörleri ve grupların farklı kovaryans matrisleri kullanılarak yapılan türetimlerde her bir grupta eşit olmak üzere n=50, 100, 250, 500 ve 1000 birimlik türetimler yapıldı. Grup sayısına göre farklı ortalama vektörü ve kovaryans matrislerine göre türetilmiş X_i veri matrisleri birleştirilerek veri setleri (SET) belirlendi.

Veri türetiminde X matrisindeki değişken sayısının dört ile ve en fazla türetim sayısının gruplarda 1000 ile sınırlandırılması, türetim ve analizde geçen bilgisayar kullanım zamanını (uygun zaman ile) kısıtlamak için yapılmıştır. Ön deneme çalışmalarında yapılan türetimlerde, grup sayısının dört, değişken sayısının beş ve türetim sayısının 1000 alınması durumunda bir döngülük analiz (türetim ve analiz) programında (1 Gb. Anabellek ve Pentium IV işlemcili bir masa üstü bilgisayarda) beş saat gibi uzun bir zaman aldığı gözlenmiştir.

Araştırmamızda değişik birim sayısı, ortalama vektörü ve kovaryans matrisleri kullanarak 158 veri seti üzerinde çalışma planlandığından bu kısıtlamalara gidilmesi uygun bulunmuştur. Değişik kovaryans matrisleri için tekrarlanan türetim ve analizler için çok uzun bilgisayar kullanım zamanı gerektiği için bu sınırlamalara gereksinim duyulmuştur.

MARS analizinin doğru tahminleme yapabilmesi için birim sayısının elli olması istenmektedir (11). Bu çalışmada söz konusu ayırma analizlerinin güvenilir sonuçlar vermesi için, N/p oranı en düşük 12.5 olarak alınmıştır. Bu oranın beş veya daha büyük olduğu durumlarda çokdeğişkenli tekniklerden güvenilir sonuçlar elde edilebilmektedir. Sınıflandırma metodlarında, doğru sınıflandırma olasılığını gerçek değerine ulaşabilmek için bu oranın büyük olması önerilmektedir (28, 29).

Türetilen veriler aracılığı ile grup sayısı $g=2,3$ ve 4 için X matrislerinin oluşturulmasında, birleştirme sonucu elde edilen X_i veri matrislerinin türetiminde yararlanılan ortalama vektörleri ve oluşturulan veri setleri aşağıda Çizelge 3.1, Çizelge 3.2 ve Çizelge 3.3’de verilmiştir.

Çizelge 3.1- Grup sayısı iki ($g=2$) olacak şekilde $V_1, V_2, \dots, V_{12}$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan X_i veri matrislerinin ortalama vektörleri

Veri seti	$g=1$ için türetimde kullanılan Ortalama vektörleri	$g=2$ için türetimde kullanılan Ortalama vektörü	Farklı varyans ve kovaryansları içeren kovaryans matrisleri
SET 1	$\mu_1 = (0,0,0,0)$	$\mu_2 = (0.5,0.5,0.5,0.5)$	$V_1, V_2, \dots, V_{12}$
SET 2	$\mu_1 = (0,0,0,0)$	$\mu_3 = (1,1,1,1)$	$V_1, V_2, \dots, V_{12}$
SET 3	$\mu_1 = (0,0,0,0)$	$\mu_4 = (2,2,2,2)$	$V_1, V_2, \dots, V_{12}$
SET 4	$\mu_1 = (0,0,0,0)$	$\mu_8 = (3,3,3,3)$	$V_1, V_2, \dots, V_{12}$
SET 5	$\mu_1 = (0,0,0,0)$	$\mu_6 = (0.5,1,2,3)$	$V_1, V_2, \dots, V_{12}$

Örneğin Set 1’de yer alan birim sayılarına göre oluşan on iki farklı setin elde edilmesinde $n=50$ için aşağıdaki yol izlendi. Her bir grupta elli birimin

bulunduğu ($n_i=50$) Set 1'in birinci alt seti olan Set1 için $X1 \sim Np(\mu_1, V1)$ ve $X2 \sim Np(\mu_2, V1)$ parametrelerine göre veri matrisleri türetildi. $X1$ ve $X2$ veri matrisleri alt alta birleştirildi. y grup kodları vektörü için ise ilk elli değer kod=1 ve son elli değer kod=2 olacak şekilde $y'=(1,1, ...,1,2,2, ...,2)$ biçiminde belirlendi. Bu işlem Set 1 için on iki kez tekrarlandı.

Çizelge 3.2- Grup sayısı üç ($g=3$) olacak şekilde $V1, V2, ..., V12$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan X_i veri matrislerinin ortalama vektörleri

Veri seti	$g=1$ için ortalama vektörü	$g=2$ için ortalama vektörü	$g=3$ için ortalama vektörü	kovaryans matrisleri
SET 6	$\mu_1 = (0,0,0,0)$	$\mu_2 = (0.5,0.5,0.5,0.5)$	$\mu_3 = (1,1,1,1)$	$V1, V2, ..., V12$
SET 7	$\mu_1 = (0,0,0,0)$	$\mu_3 = (1,1,1,1)$	$\mu_4 = (2,2,2,2)$	$V1, V2, ..., V12$
SET 8	$\mu_1 = (0,0,0,0)$	$\mu_4 = (2,2,2,2)$	$\mu_9 = (4,4,4,4)$	$V1, V2, ..., V12$
SET 9	$\mu_1 = (0,0,0,0)$	$\mu_5 = (0.5,1,1.5,2)$	$\mu_7 = (2.5,3,3.5,4)$	$V1, V2, ..., V12$

Çizelge 3.3- Grup sayısı dört ($g=4$) olacak şekilde $V1, V2, ..., V12$ kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan X_i veri matrislerinin ortalama vektörleri

Veri seti	$g=1$ için ortalama vektörü	$g=2$ için ortalama vektörü	$g=3$ için ortalama vektörü	$g=4$ için ortalama vektörü	Kovaryans matrisleri
SET 10	$\mu_1 = (0,0,0,0)$	$\mu_2 = (0.5,0.5,0.5,0.5)$	$\mu_3 = (1,1,1,1)$	$\mu_4 = (2,2,2,2)$	$V1, V2, ..., V12$
SET 11	$\mu_1 = (0,0,0,0)$	$\mu_3 = (1,1,1,1)$	$\mu_4 = (2,2,2,2)$	$\mu_8 = (3,3,3,3)$	$V1, V2, ..., V12$
SET 12	$\mu_1 = (0,0,0,0)$	$\mu_4 = (2,2,2,2)$	$\mu_9 = (4,4,4,4)$	$\mu_{11} = (6,6,6,6)$	$V1, V2, ..., V12$
SET 13	$\mu_1 = (0,0,0,0)$	$\mu_5 = (0.5,1,1.5,2)$	$\mu_7 = (2.5,3,3.5,4)$	$\mu_{10} = (4.5,5,5.5,6)$	$V1, V2, ..., V12$

Farklı ortalama vektörleri ve kovaryans matrisleri kullanılarak temel olarak farklı yapılarda çokdeğişkenli normal dağılım gösteren bağımsız değişken veri matrislerini içeren on üç Set oluşturuldu. Her bir temel set içinde ise kovaryans matrisi tiplerine göre farklı on iki alt set oluşturuldu. Böylece analiz için yüz elli altı veri seti türetilmiş matrislerin birleştirilmesi ile elde edildi.

Setler oluşturulduğunda her bir setteki birim sayıları birleştirilen türetim matrislerindeki X_i eleman sayılarının sayısına eşit oldu. Örneğin $g=2$ için $n_1=50$ ve $n_2=50$ birim sayılı X_1 ve X_2 matrisleri birleştirilerek Set 1'deki eleman sayısı $N=n_1+n_2=100$ olarak elde edildi. Dört gruplu $g=4$ ve $n_1=1000$, $n_2=1000$, $n_3=1000$ ve $n_4=1000$ birimlik türetimlere sahip olan X_1 , X_2 , X_3 ve X_4 matrisleri birleştirilerek Set 13'de $N=n_1+n_2+n_3+n_4=4000$ veri toplanmış oldu. Böylece Set olarak isimlendirilen her analiz veri setinde çok sayıda birime ilişkin homojen yapıda gruplardan oluşmuş veri setleri, kovaryans matrisinin seçimine bağlı olarak ortalamaları yakın ancak kovaryans matrisleri heterojen olan üst üste binmiş karmaşık sınıfları içeren veri setleri belirlendi. Böylece analizlerde ayırma yöntemlerinin homojen ve farklı konumlardaki sınıfları içeren veri setlerinin ayırımı ile heterojen yapıda ve üst üste çakışan sınıfları içeren veri setlerinde yöntemlerin ayırma performansları hesaplandı.

Yöntemlerin ayırma performanslarını belirlerken, y grup kodları matrisinin grup elemanları gerçek sınıf olarak alındı ve yöntemlerin çalıştırılması sonucunda elde edilen sınıflamalar ile karşılaştırıldı ve doğru sınıflama oranları hesaplandı.

Ayrıca çokdeğişkenli normal dağılım göstermeyen iki veri seti daha türetili, Set 14 (1 alt set) ve Set 15 (1 alt set) olarak analize alınacak bu setlerinde kullanılan ortalama vektörleri ve kovaryans matrisleri Çizelge 4'deki gibi belirlendi.

Çizelge 3.4- Grup sayısı iki ($g=2$) (SET 14) ve grup sayısı üç ($g=3$) olacak şekilde V1,V2, V3, V7, V8 ve V9 kovaryans matrisleri yardımı ile veri seti oluşumunda kullanılan X_i veri matrisleri

Veri seti	$g=1$ için ortalama vektörü	$g=2$ için ortalama vektörü	$g=3$ için ortalama vektörü	Kovaryans matrisleri
SET 14	$\mu_4 = (2, 2, 2, 2)$	$\mu_8 = (3, 3, 3, 3)$		V1, V2, V3,
SET 15	$\mu_3 = (1, 1, 1, 1)$	$\mu_4 = (2, 2, 2, 2)$	$\mu_8 = (3, 3, 3, 3)$	V7,V8,V9

Araştırmamızda böylece 158 (156 çokdeğişkenli normal dağılım gösteren veri setleri + 2 normal dağılım göstermeyen çokdeğişkenli veri setleri) veri seti incelendi. Çok değişkenli normal dağılıma sahip 156 veri setinin her bir grubunda eşit olmak üzere $n=50$, 100, 250, 500 ve 1000 birimlik kombinasyonlarda ($156 \cdot 5 = 780$) toplam 780 farklı

özellikte alt veri setleri türetili ve ayrıca çokdeğişkenli normal dağılım göstermeyen iki veri seti de her bir grupta $n=100$ birimlik ($2*3=6$) toplam 6 farklı özellikte alt veri setleri türetilerek toplam 786 koşulda/kombinasyonda döngü oluşturularak analizleri yapıldı.

Doğru sınıflandırma oranları; birimlerin türetildiği gruplar gerçek gruplar varsayılarak, birimlerin gerçekte ait olduğu grup ile model aracılığı ile atandığı grubun benzer olduğu durumlara ilişkin birim sayısının (n_{DSO}) gruplardaki toplam birim sayısına (N) bölünmesi ile elde edilmiştir.

3.2- Veri Analizi Yöntemleri

Oluşturulan 158 türetilmiş veri setinin her biri Bölüm 2'de ayrıntılı biçimde açıklanan Doğrusal Ayırma Analizi (DAA), Karesel Ayırma Analizi (KAA), Esnek Ayırma Analizinin (EAA) yaygın kullanımı olan ve algoritmaları açık biçimde yayınlanmış olan EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerine göre analiz edildi. Burada kullanılan degree sözcüğü modelin algoritmik yapısını yansıtmak için kullanıldı. Degree sözcüğü algoritma_1 ve algoritma_2 olarak ele alınan MARS modeli uygulamalarını belirtmektedir. Kaynaklarda yaygın olarak algoritma yerine modeli belirtmek için degree sözcüğü kullanıldığı için isim değişikliğine gidilmeksizin bu sözcüğün kullanılması uygun bulundu.

Analizlerde R paket programı kullanıldı. Verilerin analizinde DAA ve KAA'yı uygulamak için R paket programının MASS modülünden yararlanıldı. EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemleri ile verilerin analizi için ise R paket programı MDA (Mixture Discriminant Analysis, Karma Ayırma Analizi) modülünden yararlanıldı (7, 11, 21, 25).

R programının türetim ve analiz işlemlerinde kullanılabilmesi için her bir türetim ve analiz için R programlama dilinde macro programlar (paket) yazıldı.

R programı ve paketleri internet üzerinden <http://www.r-project.org> adresinden ücretsiz olarak ulaşılan bir program olduğu için R'nin son versiyonu R 2.3.1 sürümü internetten indirilerek kullanıldı. Pakette yer almayan bazı macrolar tarafımızdan yazıldı ve programa entegre edilerek çalıştırıldı, sonuçlar alındı (23).

4- BULGULAR

Ayırma Analizi Yöntemleri'nin (DAA, KAA, EAA/MARS degree=1, EAA/MARD degree=2 ve EAA/BRUTO), değişik koşullara göre türetilen çokdeğişkenli normal dağılımdan türetilen Set 1 ile Set 13 arasındaki her bir Set'te yer alan 12 alt setteki DSO değerleri ilk 12 çizelgede (Çizelge 4.1-Çizelge 4.12) verilmiştir.

Bu çizelgelerde yer alan her bir set oluşturulan koşullara göre yöntemlerin verdikleri doğru sınıflandırma oranlarına (DSO) bakılarak performans değerlendirmeleri yapılmıştır. Çizelgelerde **X** matrisinin birim sayısına göre ve kovaryans matrislerinin içeriklerine göre belirlenen DSO değerleri (minimum DSO-Maksimum DSO) biçiminde değer aralıkları biçiminde ele alınarak irdelenmiştir.

Çizelge 4.1'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı sıfır olduğu durumda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin değişik koşullara göre SET'lerde hesaplanan DSO değerleri varyanslara göre aşağıdaki gibi hesaplanmıştır.

Set 1'deki varyansı 1 olan koşulda DAA (0.60-0.69) değer aralığında, KAA (0.62-0.70), EAA/MARS degree=1 (0.62-0.77), EAA/MARS degree=2 (0.62-0.74) ve EAA/BRUTO (0.59-0.75) değer aralığında birbirlerine yakın DSO değerleri vermektedir. Varyans 3 ve 10 olan koşullar için DAA en düşük performansa sahip iken diğer yöntemler birbirlerine yakın ancak DAA'dan çok yüksek olmayan performans göstermiştir. Varyans 1 iken her bir alt grupta N=100 den büyük gruplarda 5 yöntemin DSO değerlerinde artış olduğu görülmektedir. Varyans 3 ve 10 olduğunda N sayısındaki değişim DSO değerlerini etkilememiştir.

Set 2'de varyansın 1 olduğu koşulda 5 yöntem de birbirlerine yakın DSO değerleri vermiştir. Set 1'de olduğu gibi Set 2'de de varyansın 3 ve 10 olduğu durumda DAA'nın diğer yöntemlerden daha düşük bir performansa sahip olduğu görülmüştür.

Set 3'de varyansın 1 olduğu koşulda 5 yöntem de birbirlerine yakın DSO değerleri vermiş ve varyansın 3 ve 10 olduğu durumda DAA diğer yöntemlerden daha düşük bir performans göstermiştir.

Set 4'de varyansın 1 olduğu koşulda 5 yöntem de birbirlerine yakın DSO değerleri vermiştir. Bu sette, varyansın 3 ve 10 olduğu durumda DAA diğer yöntemlerden daha düşük bir performans göstermiştir.

Set 5'de varyansın 1 olduğu koşulda 5 yöntem de birbirlerine yakın DSO değerleri verirken, varyansın 3 ve 10 olduğu durumda DAA diğer yöntemlerden daha düşük bir performans göstermiştir.

Sonuç olarak, varyans 1 iken her bir alt grupta N=100 değerinin üzerine çıktığında bütün yöntemlerin performansında artış olduğu görülmektedir. Ayrıca bu 5 veri setinde, varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği bulunmuştur.

Çizelge 4.2'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.25 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 1'deki varyansı 1 olan koşulda DAA (0.65-0.66) değer aralığında DSO değerleri verirken, KAA (0.65-0.68), EAA/MARS degree=1 (0.65-0.68), EAA/MARS degree=2 (0.65-0.68) ve EAA/BRUTO (0.64-0.65) yöntemleri birbirlerine yakın DSO değerleri vermektedir. Varyans 3 ve 10 olma durumunda ise DAA (0.60-0.65) en düşük performansa sahip iken diğer yöntemlerin DSO değerleri (0.80-0.95) birbirlerine yakın değerler olarak bulunmuştur.

Set 2'de varyansın 1 olduğu koşulda DAA (0.71-0.78) değer aralığında DSO değerleri verirken, KAA (0.78-0.95), EAA/MARS degree=1 (0.78-0.93), EAA/MARS degree=2 (0.78-0.94) ve EAA/BRUTO (0.78-0.92) yöntemleri birbirlerine yakın DSO değerleri vermiştir. Set 1'de olduğu gibi Set 2'de de varyansın 3 ve 10 olduğu durumda DAA'nın (0.70-0.75) diğer yöntemlerden (0.78-0.96) daha düşük bir performansa sahip olduğu görülmüştür.

Çizelge 4.1- Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları sıfır olan (V1,V2,V3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3
SET1	100	0.603	0.671	0.661	0.622	0.818	0.946	0.615	0.806	0.924	0.617	0.809	0.928	0.586	0.785	0.918
	200	0.698	0.662	0.645	0.709	0.805	0.940	0.765	0.805	0.924	0.735	0.802	0.930	0.745	0.921	0.921
	500	0.696	0.656	0.635	0.700	0.798	0.939	0.702	0.796	0.920	0.702	0.796	0.928	0.696	0.796	0.922
	1000	0.693	0.654	0.630	0.695	0.796	0.938	0.696	0.792	0.926	0.698	0.792	0.922	0.692	0.794	0.919
	2000	0.692	0.653	0.629	0.694	0.795	0.938	0.695	0.789	0.915	0.696	0.790	0.923	0.693	0.793	0.918
SET2	100	0.851	0.789	0.746	0.858	0.867	0.952	0.860	0.859	0.930	0.869	0.867	0.938	0.852	0.842	0.925
	200	0.845	0.784	0.738	0.848	0.858	0.947	0.853	0.853	0.930	0.858	0.860	0.938	0.846	0.846	0.928
	500	0.842	0.781	0.735	0.843	0.851	0.944	0.846	0.846	0.926	0.848	0.852	0.934	0.844	0.846	0.926
	1000	0.842	0.780	0.734	0.843	0.850	0.943	0.841	0.842	0.921	0.846	0.847	0.931	0.842	0.844	0.920
	2000	0.842	0.780	0.733	0.843	0.849	0.944	0.843	0.842	0.921	0.843	0.845	0.931	0.843	0.844	0.923
SET3	100	0.979	0.930	0.862	0.981	0.955	0.967	0.982	0.947	0.950	0.984	0.961	0.961	0.981	0.940	0.944
	200	0.978	0.929	0.858	0.979	0.953	0.965	0.980	0.944	0.947	0.984	0.955	0.958	0.979	0.940	0.945
	500	0.978	0.928	0.858	0.978	0.950	0.963	0.976	0.940	0.942	0.978	0.950	0.954	0.980	0.940	0.942
	1000	0.977	0.926	0.857	0.978	0.948	0.962	0.975	0.935	0.939	0.976	0.946	0.952	0.976	0.936	0.940
	2000	0.977	0.926	0.857	0.977	0.948	0.962	0.975	0.935	0.939	0.975	0.946	0.953	0.976	0.936	0.940
SET4	100	0.999	0.862	0.918	0.999	0.967	0.983	0.999	0.950	0.969	0.999	0.961	0.980	0.999	0.943	0.963
	200	0.999	0.979	0.916	0.999	0.990	0.982	0.999	0.986	0.967	0.999	0.991	0.977	0.999	0.997	0.965
	500	0.999	0.979	0.914	0.999	0.990	0.981	0.999	0.960	0.960	0.999	0.999	0.974	0.998	0.984	0.960
	1000	0.999	0.978	0.914	0.999	0.989	0.980	0.998	0.984	0.959	0.998	0.989	0.972	0.998	0.984	0.961
	2000	0.999	0.979	0.914	0.999	0.990	0.981	0.999	0.983	0.959	0.998	0.988	0.971	0.998	0.984	0.961
SET5	100	0.972	0.921	0.852	0.975	0.947	0.965	0.968	0.936	0.970	0.978	0.949	0.958	0.965	0.929	0.928
	200	0.972	0.919	0.850	0.972	0.945	0.963	0.965	0.935	0.946	0.975	0.945	0.956	0.964	0.930	0.943
	500	0.971	0.918	0.848	0.972	0.943	0.961	0.962	0.928	0.942	0.970	0.940	0.952	0.960	0.928	0.940
	1000	0.971	0.916	0.848	0.971	0.941	0.960	0.959	0.927	0.938	0.968	0.937	0.950	0.960	0.925	0.939
	2000	0.971	0.917	0.848	0.971	0.941	0.960	0.960	0.926	0.936	0.967	0.936	0.949	0.958	0.925	0.938

Set 3'de varyansın 1 olduğu koşulda DAA (0.78-0.94) değer aralığında DSO değerleri verirken, KAA (0.78-0.94), EAA/MARS degree=1 (0.78-0.95), EAA/MARS degree=2 (0.79-0.96) ve EAA/BRUTO (0.78-0.94) yöntemleri birbirlerine yakın DSO değerleri vermiştir ve varyansın 3 ve 10 olduğu koşulda DAA (0.70-0.90) diğer yöntemlerden (0.70-0.96) daha düşük bir performans göstermiştir.

Set 4'de varyansın 1 olduğu koşulda DAA (0.98-1.00) değer aralığında DSO değerleri verirken, KAA (0.98-1.00), EAA/MARS degree=1 (0.98-1.00), EAA/MARS degree=2 (0.99-1.00) ve EAA/BRUTO (0.98-1.00) yöntemleri birbirlerine yakın DSO değerleri vermiştir. Bu sette, varyansın 3 olduğu durumda DAA (0.96-0.97) DSO değerleri, KAA (0.96-0.97), EAA/MARS degree=1 (0.98-1.00), EAA/MARS degree=2 (0.97-1.00) ve EAA/BRUTO (0.96-0.99) DSO değerleri vermiştir. DAA diğer yöntemlerden daha düşük performans göstermiştir. Aynı zamanda varyansın 10 olduğu durumda DAA (0.91-0.92) DSO değer aralıkları, KAA (0.98-0.99), EAA/MARS degree=1 (0.96-0.97), EAA/MARS degree=2 (0.97-0.98) ve EAA/BRUTO (0.99-0.99) aralığında DSO değerleri vermiştir. Bu koşulda da DAA, diğer yöntemlerden daha düşük performans göstermiştir.

Set 5'de varyansın 1 olduğu koşulda DAA (0.83-0.96) değer aralığında DSO değerleri verirken, KAA (0.95-0.96), EAA/MARS degree=1 (0.93-0.96), EAA/MARS degree=2 (0.94-0.96) ve EAA/BRUTO (0.93-0.96) değer aralığında ve birbirlerine yakın DSO değerleri vermiştir. Bu sette, varyansın 3 olduğu durumda DAA (0.89-0.90) değer aralığında DSO değerleri verirken, KAA (0.93-0.94), EAA/MARS degree=1 (0.91-0.93), EAA/MARS degree=2 (0.92-0.94) ve EAA/BRUTO (0.91-0.92) aralığında DSO değerleri vermiştir. DAA yine bu koşulda da diğer yöntemlere göre daha düşük performans göstermiştir. Aynı zamanda varyansın 10 olduğu durumda DAA (0.83-0.84) değer aralığında DSO değerleri verirken, KAA (0.96-0.97), EAA/MARS degree=1 (0.93-0.94), EAA/MARS degree=2 (0.94-0.95) ve EAA/BRUTO (0.93-0.94) değer aralığında DSO değerleri vermiştir. DAA, diğer yöntemlerden daha düşük performans göstermiştir.

Sonuç olarak, bu 5 veri setinde varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği belirlenmiştir.

Çizelge 4.2- Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.25 olan (V4,V5,V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6
SET1	100	0.651	0.650	0.650	0.680	0.830	0.950	0.675	0.820	0.930	0.687	0.810	0.930	0.643	0.780	0.910
	200	0.660	0.635	0.627	0.670	0.816	0.948	0.660	0.797	0.922	0.670	0.811	0.929	0.650	0.790	0.919
	500	0.650	0.630	0.612	0.656	0.806	0.944	0.656	0.786	0.914	0.660	0.800	0.924	0.650	0.786	0.916
	1000	0.648	0.627	0.610	0.652	0.806	0.944	0.653	0.783	0.911	0.656	0.796	0.923	0.648	0.785	0.914
	2000	0.648	0.626	0.606	0.649	0.803	0.943	0.650	0.779	0.909	0.651	0.790	0.919	0.648	0.782	0.912
SET2	100	0.780	0.750	0.720	0.790	0.860	0.960	0.790	0.830	0.930	0.810	0.850	0.930	0.780	0.820	0.920
	200	0.708	0.742	0.708	0.952	0.850	0.952	0.926	0.832	0.925	0.934	0.848	0.934	0.923	0.823	0.923
	500	0.776	0.738	0.702	0.780	0.846	0.950	0.782	0.826	0.918	0.786	0.840	0.932	0.780	0.824	0.922
	1000	0.776	0.739	0.700	0.778	0.846	0.948	0.779	0.823	0.916	0.780	0.836	0.928	0.778	0.824	0.919
	2000	0.775	0.737	0.699	0.776	0.844	0.948	0.776	0.821	0.914	0.777	0.832	0.926	0.776	0.823	0.917
SET3	100	0.940	0.900	0.840	0.940	0.940	0.970	0.950	0.930	0.940	0.960	0.940	0.960	0.940	0.920	0.940
	200	0.937	0.895	0.836	0.938	0.933	0.966	0.941	0.918	0.941	0.947	0.933	0.952	0.939	0.913	0.939
	500	0.776	0.738	0.702	0.780	0.846	0.950	0.782	0.826	0.918	0.786	0.840	0.932	0.780	0.824	0.922
	1000	0.935	0.892	0.832	0.936	0.928	0.963	0.935	0.910	0.932	0.937	0.923	0.946	0.934	0.908	0.934
	2000	0.935	0.892	0.832	0.935	0.929	0.963	0.934	0.908	0.931	0.935	0.921	0.945	0.934	0.908	0.933
SET4	100	1.000	0.970	0.910	1.000	0.980	0.980	1.000	0.970	0.970	1.000	0.980	0.980	1.000	0.970	0.960
	200	0.989	0.965	0.907	0.989	0.980	0.981	0.990	0.974	0.961	0.993	0.981	0.976	0.990	0.945	0.959
	500	0.988	0.964	0.904	0.988	0.978	0.982	0.988	0.970	0.956	0.988	0.978	0.966	0.988	0.970	0.956
	1000	0.988	0.964	0.905	0.988	0.979	0.980	0.988	0.968	0.954	0.989	0.987	0.965	0.988	0.988	0.955
	2000	0.988	0.964	0.904	0.988	0.978	0.979	0.987	0.967	0.953	0.988	0.975	0.964	0.987	0.982	0.954
SET5	100	0.960	0.890	0.840	0.960	0.940	0.960	0.960	0.930	0.940	0.960	0.940	0.950	0.960	0.920	0.940
	200	0.954	0.901	0.834	0.954	0.935	0.966	0.950	0.920	0.941	0.958	0.933	0.953	0.949	0.914	0.939
	500	0.952	0.898	0.834	0.952	0.932	0.962	0.946	0.914	0.936	0.952	0.928	0.950	0.948	0.914	0.938
	1000	0.952	0.897	0.833	0.952	0.932	0.963	0.946	0.912	0.933	0.952	0.926	0.946	0.945	0.910	0.935
	2000	0.832	0.897	0.832	0.962	0.931	0.962	0.931	0.911	0.931	0.944	0.924	0.944	0.933	0.910	0.933

Çizelge 4.3'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.75 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 1'deki varyansın 1 olduğu koşulda DAA (0.61-0.64) değer aralığında DSO değerleri verirken, KAA (0.61-0.66), EAA/MARS degree=1 (0.61-0.65), EAA/MARS degree=2 (0.62-0.64) ve EAA/BRUTO (0.61-0.64) DSO değerleri vermiştir. Bu koşulda beş yöntemde birbirlerine yakın DSO değerleri vermektedir. Bu sette, varyansın 3 olduğu durumda DAA (0.60-0.63) değer aralığında DSO değerleri verirken, KAA (0.90-0.91), EAA/MARS degree=1 (0.75-0.77), EAA/MARS degree=2 (0.83-0.86) ve EAA/BRUTO (0.75-0.81) aralığında DSO değerleri vermiştir. DAA diğer dört yöntemden daha düşük performans göstermiştir. Aynı zamanda varyansın 10 olduğu durumda DAA (0.59-0.64) yöntemi değer aralığında DSO değerleri verirken, KAA (0.97-0.98), EAA/MARS degree=1 (0.88-0.89), EAA/MARS degree=2 (0.91-0.93) ve EAA/BRUTO (0.89-0.90) değer aralığında DSO değerleri vermiştir. DAA, diğer yöntemlerden daha düşük performans göstermiştir.

Set 2'de varyansın 1 olduğu koşulda DAA (0.71-0.72), KAA (0.71-0.74), EAA/MARS degree=1 (0.71-0.73), EAA/MARS degree=2 (0.71-0.74) ve EAA/BRUTO (0.71-0.72) aralıklarında ve birbirlerine yakın DSO değerleri vermiştir. Bu sette, varyansın 3 olduğu durumda DAA (0.69-0.70), KAA (0.91-0.92), EAA/MARS degree=1 (0.78-0.80), EAA/MARS degree=2 (0.85-0.88) ve EAA/BRUTO (0.77-0.78) aralıklarında DSO değerleri vermiştir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Aynı zamanda varyansın 10 olduğu durumda DAA (0.66-0.69) aralığında, KAA (0.98-0.98), EAA/MARS degree=1 (0.89-0.90), EAA/MARS degree=2 (0.91-0.94) ve EAA/BRUTO (0.90-0.90) aralıklarında DSO değerleri vermiştir. Bu duruma göre DAA yöntemi diğer yöntemlerden daha düşük performans göstermiştir. KAA tüm yöntemlerden yüksek sınıflandırma olasılıklarına sahip olarak belirlenmiştir. EAA yöntemleri DAA'dan yüksek ancak KAA'dan düşük değerler vermiştir.

Set 3'de varyansın 1 olduğu koşulda DAA (0.87-0.88) aralığında, KAA (0.87-0.88), EAA/MARS degree=1 (0.87-0.88), EAA/MARS degree=2 (0.87-0.89) ve EAA/BRUTO (0.87-0.88) değer aralıklarında ve birbirlerine yakın DSO değerleri vermiştir.

Çizelge 4.3- Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.75 olan (V7,V8,V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9
SET1	100	0.644	0.632	0.643	0.666	0.911	0.982	0.636	0.762	0.897	0.638	0.861	0.933	0.636	0.810	0.894
	200	0.630	0.615	0.615	0.644	0.905	0.980	0.650	0.766	0.899	0.630	0.858	0.929	0.628	0.753	0.897
	500	0.614	0.605	0.595	0.624	0.901	0.978	0.618	0.758	0.890	0.622	0.846	0.919	0.614	0.752	0.895
	1000	0.612	0.601	0.590	0.616	0.900	0.978	0.615	0.752	0.885	0.617	0.840	0.913	0.612	0.750	0.893
	2000	0.610	0.599	0.586	0.612	0.899	0.978	0.612	0.749	0.883	0.614	0.834	0.910	0.610	0.751	0.893
SET2	100	0.722	0.706	0.693	0.743	0.921	0.983	0.731	0.795	0.905	0.740	0.882	0.939	0.720	0.777	0.903
	200	0.717	0.698	0.674	0.727	0.916	0.982	0.722	0.797	0.905	0.729	0.876	0.933	0.717	0.784	0.902
	500	0.713	0.692	0.667	0.717	0.913	0.980	0.716	0.789	0.895	0.721	0.866	0.923	0.713	0.783	0.900
	1000	0.712	0.692	0.665	0.714	0.911	0.980	0.714	0.783	0.892	0.716	0.860	0.917	0.712	0.782	0.899
	2000	0.712	0.691	0.664	0.713	0.911	0.980	0.713	0.779	0.891	0.714	0.856	0.913	0.712	0.783	0.898
SET3	100	0.875	0.845	0.803	0.880	0.952	0.987	0.882	0.888	0.929	0.896	0.930	0.955	0.877	0.870	0.923
	200	0.872	0.840	0.798	0.875	0.948	0.986	0.877	0.880	0.924	0.886	0.921	0.948	0.874	0.870	0.921
	500	0.868	0.839	0.794	0.869	0.946	0.985	0.870	0.875	0.915	0.874	0.915	0.939	0.869	0.869	0.917
	1000	0.868	0.839	0.793	0.868	0.945	0.984	0.869	0.869	0.913	0.871	0.911	0.933	0.868	0.867	0.916
	2000	0.867	0.838	0.792	0.867	0.944	0.985	0.868	0.866	0.910	0.869	0.909	0.929	0.867	0.867	0.914
SET4	100	0.886	0.934	0.886	0.991	0.978	0.991	0.953	0.952	0.953	0.973	0.973	0.973	0.946	0.934	0.946
	200	0.955	0.931	0.882	0.956	0.975	0.991	0.958	0.945	0.946	0.964	0.967	0.965	0.957	0.963	0.945
	500	0.952	0.929	0.880	0.953	0.974	0.990	0.954	0.940	0.939	0.957	0.960	0.957	0.954	0.939	0.940
	1000	0.952	0.929	0.879	0.952	0.974	0.990	0.953	0.938	0.936	0.954	0.958	0.954	0.954	0.939	0.938
	2000	0.952	0.929	0.879	0.953	0.974	0.990	0.953	0.937	0.930	0.953	0.956	0.952	0.953	0.937	0.937
SET5	100	0.986	0.900	0.823	0.986	0.965	0.988	0.979	0.922	0.928	0.987	0.952	0.957	0.981	0.915	0.922
	200	0.984	0.895	0.819	0.984	0.962	0.987	0.975	0.916	0.927	0.984	0.945	0.952	0.976	0.910	0.923
	500	0.983	0.893	0.816	0.984	0.959	0.985	0.972	0.910	0.917	0.980	0.937	0.944	0.972	0.907	0.919
	1000	0.984	0.893	0.815	0.984	0.959	0.985	0.971	0.909	0.915	0.978	0.934	0.921	0.974	0.906	0.918
	2000	0.983	0.893	0.814	0.983	0.959	0.985	0.970	0.907	0.912	0.978	0.932	0.937	0.968	0.906	0.917

Bu sette, varyansın 3 olduđu durumda DAA (0.84-0.85) aralığında, KAA (0.94-0.95), EAA/MARS degree=1 (0.87-0.89), EAA/MARS degree=2 (0.91-0.93) ve EAA/BRUTO (0.87-0.88) aralığında DSO deęerleri vermiřtir. KAA bu kořulda da en ylıksek doęru sınıflandırma performansı gstermiřtir. Varyansın 10 olduđu durumda DAA (0.79-0.80) deęer aralığında, KAA (0.98-0.99), EAA/MARS degree=1 (0.91-0.93), EAA/MARS degree = 2 (0.93-0.96) ve EAA/BRUTO (0.91-0.92) aralığında DSO deęerleri vermiřtir. DAA bu kořulda da dięer yntemlere gcre en dűřyk performansa sahip olarak belirginleřmiřtir. EAA yntemleri KAA'dan dűřyk ancak KAA'nın deęerlerine yakın DSO deęerleri vermiřlerdir.

Set 4'de varyansın 1 olduđu kořulda DAA (0.89-0.95) deęer aralığında, KAA (0.95-0.99), EAA/MARS degree=1 (0.95-0.96), EAA/MARS degree=2 (0.95-0.97) ve EAA/BRUTO (0.94-0.95) deęer aralığında DSO deęerleri vermiřlerdir. Tcm yntemlerde DSO 0.90 deęerinin izerinde gczlenmiřtir. Ancak KAA ve EAA modelleri 0.95'den daha ylıksek DSO deęerlerine sahiptir. Bu sette, varyansın 3 olduđu durumda DAA (0.93-0.93) aralığında, KAA (0.97-0.98), EAA/MARS degree=1 (0.94-0.95), EAA/MARS degree=2 (0.96-0.97) ve EAA/BRUTO (0.94-0.96) deęer aralığında DSO deęerleri vermiřtir. DAA yntemi dięer yntemlerden daha dűřyk performans gstermiřtir. Aynı zamanda varyansın 10 olduđu durumda DAA (0.88-0.89) aralığında, KAA (0.99-0.99), EAA/MARS degree=1 (0.93-0.95), EAA/MARS degree=2 (0.95-0.97) ve EAA/BRUTO (0.94-0.95) deęer aralığında DSO deęerleri vermiřtir. DAA dięer yntemlerden daha dűřyk performans gstermiřtir ve KAA en iyi performansı gstermiřtir.

Set 5'de varyansın 1 olduđu kořulda DAA (0.98-0.99) aralığında, KAA (0.98-0.99), EAA/MARS degree=1 (0.97-0.98), EAA/MARS degree=2 (0.98-0.99) ve EAA/BRUTO (0.97-0.98) deęer aralıklarında ve birbirlerine yakın DSO deęerleri vermiřtir. Bu sette, varyansın 3 olduđu durumda DAA (0.89-0.90) aralığında, KAA (0.95-0.96), EAA/MARS degree=1 (0.91-0.92), EAA/MARS degree=2 (0.93-0.95) ve EAA/BRUTO (0.91-0.92) aralıklarında DSO deęerleri vermiřtir. DAA, dięer yntemlerden daha dűřyk performans gstermiřtir ve en iyi performansı KAA gstermiřtir. Varyansın 10 olduđu durumda DAA (0.81-0.82) aralığında, KAA (0.98-0.99), EAA/MARS degree=1 (0.91-0.93), EAA/MARS degree=2 (0.93-0.96) ve EAA/BRUTO (0.92-0.92) aralıklarında DSO deęerleri vermiřtir. DAA, dięer yntemlerden daha dűřyk performans gstermiřtir ve yine en iyi performansı KAA gstermiřtir.

Sonuç olarak, bu 5 veri setinde varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği ve KAA'nın ise en iyi sınıflandırma performansını gösterdiği belirlenmiştir.

Çizelge 4.4'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.90 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 1'deki varyansı 1 olan koşulda DAA (0.60-0.64) aralığında, KAA (0.61-0.69), EAA/MARS degree=1 (0.61-0.65), EAA/MARS degree=2 (0.61-0.64) ve EAA/BRUTO (0.59-0.64) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Bu sette, varyansın 3 olduğu durumda DAA (0.60-0.62) aralığında, KAA (0.71-0.98), EAA/MARS degree=1 (0.74-0.84), EAA/MARS degree=2 (0.85-0.96) ve EAA/BRUTO (0.72-0.84) aralıklarında DSO değerleri vermişlerdir. DAA, diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda DAA (0.59-0.73) aralığında, KAA (0.99-1.00), EAA/MARS degree=1 (0.88-0.93), EAA/MARS degree=2 (0.91-0.96) ve EAA/BRUTO (0.89-0.94) değer aralıklarında DSO değerleri vermişlerdir. DAA, diğer yöntemlerden daha düşük performans ve buna karşın KAA en yüksek performansı göstermiştir.

Set 2'de varyansın 1 olduğu koşulda DAA (0.70-0.73) aralığında, KAA (0.70-0.74), EAA/MARS degree=1 (0.70-0.73), EAA/MARS degree=2 (0.70-0.74) ve EAA/BRUTO (0.70-0.73) değer aralıklarında DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.68-0.70) aralığında, KAA (0.96-0.97), EAA/MARS degree=1 (0.77-0.80), EAA/MARS degree=2 (0.87-0.90) ve EAA/BRUTO (0.77-0.77) değer aralıklarında DSO değerleri vermişlerdir. DAA, diğer yöntemlerden daha düşük performans gösterirken, en yüksek performansı KAA yöntemi göstermiştir. Aynı zamanda varyansın 10 olduğu durumda DAA (0.66-0.68) aralığında, KAA (0.99-0.99), EAA/MARS degree=1 (0.88-0.90), EAA/MARS degree=2 (0.91-0.94) ve EAA/BRUTO (0.89-0.90) değer aralıklarında DSO değerleri vermişlerdir. DAA diğer yöntemlerden daha düşük performans gösterirken, yine en yüksek performansı KAA yöntemi göstermiştir.

Çizelge 4.4- Ayırma Analizi Yöntemleri'nin, iki alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.90 olan (V10,V11,V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
2 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12
SET1	100	0.635	0.623	0.646	0.687	0.953	0.994	0.652	0.745	0.893	0.637	0.882	0.939	0.640	0.718	0.889
	200	0.596	0.611	0.612	0.627	0.958	0.993	0.605	0.758	0.893	0.603	0.879	0.930	0.588	0.742	0.890
	500	0.608	0.600	0.624	0.616	0.956	0.996	0.612	0.750	0.932	0.614	0.864	0.964	0.606	0.736	0.944
	1000	0.606	0.596	0.585	0.610	0.956	0.993	0.608	0.746	0.879	0.610	0.856	0.911	0.606	0.739	0.890
	2000	0.604	0.595	0.733	0.606	0.955	0.992	0.605	0.743	0.876	0.607	0.852	0.908	0.604	0.740	0.890
SET2	100	0.732	0.699	0.682	0.744	0.966	0.993	0.729	0.796	0.900	0.738	0.900	0.942	0.733	0.774	0.895
	200	0.706	0.689	0.669	0.716	0.962	0.994	0.711	0.789	0.899	0.716	0.891	0.934	0.706	0.774	0.895
	500	0.700	0.684	0.662	0.704	0.960	0.992	0.702	0.780	0.890	0.706	0.878	0.922	0.700	0.770	0.892
	1000	0.699	0.683	0.658	0.701	0.960	0.993	0.701	0.774	0.885	0.703	0.874	0.916	0.700	0.770	0.894
	2000	0.700	0.682	0.657	0.701	0.959	0.993	0.701	0.770	0.883	0.703	0.870	0.911	0.701	0.772	0.895
SET3	100	0.858	0.838	0.797	0.864	0.978	0.995	0.865	0.887	0.925	0.878	0.941	0.958	0.859	0.868	0.918
	200	0.854	0.830	0.790	0.858	0.975	0.995	0.859	0.876	0.918	0.868	0.930	0.949	0.855	0.861	0.916
	500	0.852	0.826	0.784	0.852	0.972	0.996	0.854	0.864	0.910	0.858	0.918	0.938	0.852	0.858	0.912
	1000	0.852	0.826	0.783	0.852	0.972	0.994	0.853	0.860	0.905	0.855	0.913	0.932	0.852	0.858	0.910
	2000	0.852	0.826	0.783	0.852	0.972	0.994	0.852	0.858	0.903	0.853	0.910	0.928	0.852	0.858	0.909
SET4	100	0.946	0.930	0.880	0.947	0.989	0.996	0.949	0.950	0.950	0.962	0.977	0.974	0.948	0.942	0.942
	200	0.874	0.922	0.874	0.997	0.987	0.997	0.942	0.940	0.942	0.965	0.968	0.965	0.940	0.947	0.940
	500	0.944	0.918	0.872	0.944	0.986	0.996	0.944	0.934	0.940	0.946	0.960	0.956	0.944	0.932	0.936
	1000	0.941	0.919	0.871	0.942	0.985	0.996	0.942	0.931	0.930	0.943	0.957	0.952	0.941	0.931	0.933
	2000	0.871	0.918	0.871	0.996	0.985	0.996	0.929	0.929	0.929	0.950	0.955	0.950	0.932	0.930	0.932
SET5	100	0.998	0.908	0.824	0.998	0.987	0.995	0.998	0.931	0.926	0.998	0.966	0.963	0.998	0.924	0.918
	200	0.999	0.903	0.815	0.999	0.984	0.995	0.998	0.923	0.922	0.999	0.957	0.953	0.999	0.917	0.917
	500	1.000	0.902	0.902	1.000	0.982	0.982	0.996	0.920	0.920	0.996	0.950	0.950	1.000	0.916	0.916
	1000	1.000	0.899	0.811	1.000	0.983	0.994	0.997	0.915	0.908	0.997	0.944	0.940	0.998	0.913	0.913
	2000	0.999	0.899	0.811	0.999	0.982	0.995	0.997	0.914	0.908	0.997	0.943	0.937	0.998	0.914	0.914

Set 3'de varyansın 1 olduğu koşulda DAA (0.85-0.86) aralığında, KAA (0.85-0.86), EAA/MARS degree=1 (0.85-0.87), EAA/MARS degree=2 (0.85-0.89) ve EAA/BRUTO (0.85-0.86) değer aralıklarında DSO değerleri vermişlerdir. Bütün yöntemler birbirlerine yakın DSO değerleri vermiştir. Varyansın 3 olduğu durumda DAA (0.83-0.84) aralığında, KAA (0.97-0.98), EAA/MARS degree=1 (0.86-0.88), EAA/MARS degree=2 (0.91-0.94) ve EAA/BRUTO (0.86-0.87) değer aralıklarında DSO değerleri vermişlerdir. DAA diğer yöntemlerden daha düşük performans gösterirken, en yüksek performansı KAA yöntemi göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.78-0.80) aralığında, KAA (0.99-1.00), EAA/MARS degree=1 (0.90-0.93), EAA/MARS degree=2 (0.93-0.96) ve EAA/BRUTO (0.91-0.92) değer aralıklarında DSO değerleri vermişlerdir. DAA diğer yöntemlerden daha düşük performans gösterirken, yine en yüksek performansı KAA yöntemi göstermiştir.

Set 4'de varyansın 1 olduğu koşulda DAA (0.87-0.95) aralığında, KAA (0.94-1.00), EAA/MARS degree=1 (0.93-0.95), EAA/MARS degree=2 (0.95-0.97) ve EAA/BRUTO (0.93-0.95) değer aralıklarında DSO değerleri vermişlerdir. Bütün yöntemleri birbirlerine yakın DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.92-0.93) aralığında KAA (0.96-0.99), EAA/MARS degree=1 (0.93-0.95), EAA/MARS degree=2 (0.96-0.97) ve EAA/BRUTO (0.93-0.95) değer aralıklarında DSO değerleri vermişlerdir. Diğer yöntemlerden daha düşük performans gösterirken, en yüksek performansı KAA ve daha sonra EAA/MARS degree=2 yöntemi göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.87-0.88) aralığında KAA (1.00-1.00), EAA/MARS degree=1 (0.93-0.94), EAA/MARS degree=2 (0.95-1.00) ve EAA/BRUTO (0.94-1.00) değer aralıklarında DSO değerleri vermişlerdir. Diğer yöntemlerden daha düşük performans göstermiştir ve KAA hatasız tahminleme yaparak en mükemmel performansı göstermiştir.

Set 5'de varyansın 1 olduğu koşulda DAA (1.00-1.00) aralığında, KAA (1.00-1.00), EAA/MARS degree=1 (1.00-1.00), EAA/MARS degree=2 (1.00-1.00) ve EAA/BRUTO (1.00-1.00) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler hatasız tahminleme yapmışlardır. Varyansın 3 olduğu koşulda DAA (0.90-0.91) aralığında, KAA (0.98-0.99), EAA/MARS degree=1 (0.91-0.93), EAA/MARS degree=2 (0.94-0.97) ve EAA/BRUTO (0.91-0.92) değer aralıklarında DSO değerleri vermişlerdir. DAA, diğer yöntemlerden düşük performans göstermiştir ve en iyi performansı KAA göstermiştir. Varyansın 10 olduğu koşulda ise, DAA (0.81-0.90) aralığında, KAA (0.98-1.00),

EAA/MARS degree=1 (0.91-0.93), EAA/MARS degree=2 (0.94-0.96) ve EAA/BRUTO (0.91-0.92) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir ve yine en iyi performansı KAA göstermiştir.

Sonuç olarak; bağımlı değişken 2 grulu ve kovaryans 0.90 olduğunda, bu 5 veri setinde varyansın 1 olduğu koşulda bütün yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği ve KAA'nın ise en iyi sınıflandırma performansını gösterdiği belirlenmiştir.

Çizelge 4.5'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 6'da varyansı 1 olan koşulda DAA (0.59-0.60) aralığında, KAA (0.59-0.63), EAA/MARS degree=1 (0.59-0.60), EAA/MARS degree=2 (0.59-0.60) ve EAA/BRUTO (0.59-0.61) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyansın 3 olduğu durumda DAA (0.51-0.54) aralığında, KAA (0.61-0.66), EAA/MARS degree=1 (0.60-0.62), EAA/MARS degree=2 (0.60-0.62) ve EAA/BRUTO (0.59-0.62) değer aralıklarında DSO değerleri vermişlerdir. DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.47-0.51) aralığında, KAA (0.67-0.72), EAA/MARS degree=1 (0.65-0.69), EAA/MARS degree=2 (0.66-0.70) ve EAA/BRUTO (0.66-0.69) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Set 7'de varyansın 1 olduğu koşulda DAA (0.79-0.80) aralığında, KAA (0.79-0.81), EAA/MARS degree=1 (0.78-0.80), EAA/MARS degree=2 (0.79-0.80) ve EAA/BRUTO (0.78-0.80) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.67-0.68) aralığında, KAA (0.71-0.74), EAA/MARS degree=1 (0.70-0.71), EAA/MARS degree=2 (0.70-0.72) ve EAA/BRUTO (0.69-0.71) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda DAA (0.57-0.60) aralığında, KAA (0.71-0.75), EAA/MARS

Çizelge 4.5 - Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0 olan (V1,V2,V3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
3 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3
SET6	150	0.606	0.540	0.511	0.633	0.655	0.725	0.603	0.609	0.692	0.597	0.599	0.689	0.606	0.589	0.689
	300	0.599	0.526	0.491	0.611	0.631	0.701	0.602	0.616	0.689	0.601	0.616	0.699	0.600	0.617	0.694
	750	0.592	0.517	0.476	0.597	0.615	0.680	0.593	0.608	0.671	0.595	0.612	0.683	0.595	0.613	0.680
	1500	0.591	0.513	0.469	0.593	0.609	0.675	0.589	0.602	0.659	0.593	0.607	0.671	0.590	0.608	0.670
	3000	0.590	0.512	0.465	0.592	0.607	0.671	0.586	0.599	0.653	0.590	0.602	0.663	0.588	0.605	0.662
SET7	150	0.798	0.677	0.598	0.810	0.744	0.752	0.797	0.712	0.723	0.801	0.719	0.725	0.801	0.694	0.722
	300	0.793	0.672	0.584	0.799	0.728	0.733	0.792	0.709	0.720	0.800	0.719	0.730	0.794	0.702	0.724
	750	0.791	0.668	0.577	0.792	0.719	0.721	0.783	0.703	0.707	0.795	0.709	0.720	0.788	0.707	0.713
	1500	0.789	0.666	0.573	0.789	0.715	0.716	0.781	0.699	0.699	0.789	0.705	0.711	0.784	0.703	0.707
	3000	0.789	0.666	0.573	0.790	0.713	0.713	0.781	0.696	0.696	0.786	0.702	0.701	0.783	0.700	0.701
SET8	150	0.972	0.876	0.739	0.975	0.897	0.821	0.973	0.883	0.799	0.979	0.897	0.809	0.975	0.884	0.793
	300	0.971	0.872	0.733	0.972	0.890	0.808	0.970	0.879	0.794	0.975	0.890	0.802	0.972	0.879	0.798
	750	0.969	0.869	0.731	0.971	0.887	0.804	0.967	0.873	0.787	0.971	0.883	0.795	0.969	0.875	0.792
	1500	0.971	0.869	0.970	0.971	0.885	0.971	0.965	0.871	0.965	0.968	0.879	0.968	0.967	0.873	0.968
	3000	0.970	0.868	0.729	0.970	0.883	0.800	0.965	0.869	0.802	0.965	0.844	0.789	0.967	0.872	0.785
SET9	150	0.932	0.825	0.698	0.937	0.863	0.814	0.930	0.835	0.791	0.941	0.853	0.796	0.933	0.831	0.787
	300	0.931	0.822	0.693	0.933	0.854	0.802	0.927	0.836	0.786	0.935	0.849	0.795	0.931	0.832	0.790
	750	0.929	0.821	0.688	0.931	0.849	0.793	0.924	0.832	0.777	0.928	0.843	0.789	0.927	0.833	0.784
	1500	0.929	0.819	0.686	0.929	0.847	0.792	0.921	0.829	0.774	0.926	0.838	0.784	0.924	0.831	0.779
	3000	0.928	0.818	0.686	0.929	0.846	0.791	0.921	0.827	0.773	0.924	0.835	0.782	0.923	0.830	0.777

degree=1 (0.70-0.72), EAA/MARS degree=2 (0.70-0.73) ve EAA/BRUTO (0.70-0.72) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Set 8'de varyansın 1 olduğu koşulda DAA (0.97-0.97) aralığında, KAA (0.97-0.98), EAA/MARS degree=1 (0.96-0.97), EAA/MARS degree=2 (0.96-0.98) ve EAA/BRUTO (0.96-0.98) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermiştir. Varyansın 3 olduğu durumda DAA (0.87-0.88) aralığında, KAA (0.88-0.90), EAA/MARS degree=1 (0.87-0.88), EAA/MARS degree=2 (0.84-0.90) ve EAA/BRUTO (0.87-0.88) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.73-0.97) aralığında KAA (0.80-0.97), EAA/MARS degree=1 (0.79-0.97), EAA/MARS degree=2 (0.79-0.97) ve EAA/BRUTO (0.79-0.97) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Set 9'da varyansın 1 olduğu koşulda DAA (0.93-0.93) aralığında, KAA (0.93-0.94), EAA/MARS degree=1 (0.92-0.93), EAA/MARS degree=2 (0.92-0.94) ve EAA/BRUTO (0.92-0.93) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.82-0.83) aralığında KAA (0.85-0.86), EAA/MARS degree=1 (0.83-0.84), EAA/MARS degree=2 (0.84-0.85) ve EAA/BRUTO (0.83-0.83) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.67-0.70) aralığında, KAA (0.80-0.81), EAA/MARS degree=1 (0.79-0.80), EAA/MARS degree=2 (0.78-0.80) ve EAA/BRUTO (0.78-0.79) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Sonuç olarak; 3 gruplu ve kovaryansı 0 olan 4 veri setinde, varyansın 1 olduğu koşulda tüm yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği belirlenmiştir.

Çizelge 4.6'da görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.25 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Çizelge 4.6 - Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve kovaryansı 0.25 olan (V4,V5,V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
3 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6
SET6	150	0.451	0.410	0.501	0.479	0.514	0.728	0.444	0.463	0.690	0.439	0.457	0.683	0.439	0.447	0.683
	300	0.541	0.504	0.479	0.559	0.631	0.702	0.540	0.601	0.683	0.540	0.604	0.694	0.539	0.602	0.688
	750	0.535	0.492	0.463	0.541	0.613	0.684	0.536	0.595	0.664	0.539	0.604	0.679	0.536	0.600	0.676
	1500	0.532	0.489	0.455	0.536	0.607	0.677	0.532	0.588	0.653	0.535	0.597	0.667	0.532	0.594	0.664
	3000	0.531	0.486	0.450	0.533	0.605	0.673	0.530	0.585	0.648	0.532	0.592	0.658	0.530	0.590	0.657
SET7	150	0.713	0.639	0.501	0.730	0.729	0.728	0.713	0.679	0.690	0.715	0.689	0.683	0.713	0.658	0.683
	300	0.706	0.631	0.561	0.715	0.710	0.734	0.708	0.680	0.713	0.714	0.695	0.725	0.708	0.673	0.718
	750	0.768	0.627	0.551	0.773	0.700	0.720	0.763	0.673	0.699	0.775	0.688	0.713	0.765	0.677	0.708
	1500	0.701	0.623	0.548	0.703	0.697	0.716	0.699	0.669	0.691	0.705	0.681	0.703	0.699	0.675	0.700
	3000	0.546	0.623	0.546	0.714	0.695	0.714	0.689	0.667	0.689	0.700	0.678	0.700	0.695	0.672	0.695
SET8	150	0.918	0.835	0.719	0.923	0.867	0.818	0.923	0.845	0.788	0.930	0.865	0.798	0.923	0.843	0.785
	300	0.916	0.832	0.712	0.918	0.860	0.805	0.918	0.844	0.784	0.924	0.857	0.793	0.919	0.842	0.788
	750	0.711	0.829	0.711	0.800	0.855	0.800	0.777	0.837	0.777	0.784	0.848	0.784	0.783	0.839	0.783
	1500	0.914	0.828	0.709	0.915	0.853	0.797	0.910	0.835	0.773	0.914	0.843	0.780	0.913	0.837	0.778
	3000	0.913	0.828	0.708	0.913	0.852	0.796	0.908	0.799	0.770	0.912	0.841	0.777	0.911	0.835	0.775
SET9	150	0.789	0.789	0.789	0.841	0.843	0.841	0.805	0.805	0.805	0.829	0.831	0.829	0.797	0.798	0.797
	300	0.875	0.785	0.672	0.879	0.832	0.800	0.874	0.807	0.778	0.882	0.825	0.788	0.875	0.801	0.782
	750	0.913	0.780	0.667	0.915	0.825	0.793	0.908	0.801	0.769	0.913	0.816	0.780	0.911	0.803	0.775
	1500	0.873	0.779	0.664	0.874	0.823	0.791	0.869	0.798	0.765	0.873	0.811	0.775	0.870	0.800	0.771
	3000	0.872	0.778	0.663	0.872	0.822	0.790	0.867	0.796	0.763	0.870	0.810	0.772	0.868	0.799	0.768

Set 6'da varyansı 1 olan koşulda DAA (0.45-0.54) aralığında, KAA (0.48-0.56), EAA/MARS degree=1 (0.44-0.54), EAA/MARS degree=2 (0.44-0.54) ve EAA/BRUTO (0.44-0.54) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Ayrıca varyans 1 koşulunda her bir alt grupta n=50'nin üzerine çıktığında 5 yöntemin DSO değerlerinde artış olduğu görülmektedir. Varyansın 3 olduğu durumda DAA (0.41-0.50) aralığında, KAA (0.51-0.63), EAA/MARS degree=1 (0.46-0.60), EAA/MARS degree=2 (0.44-0.60) ve EAA/BRUTO (0.45-0.60) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.45-0.50) aralığında, KAA (0.67-0.73), EAA/MARS degree=1 (0.65-0.69), EAA/MARS degree=2 (0.66-0.69) ve EAA/BRUTO (0.66-0.69) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Set 7'de varyansın 1 olduğu koşulda DAA (0.55-0.77) aralığında, KAA (0.70-0.77), EAA/MARS degree=1 (0.69-0.76), EAA/MARS degree=2 (0.70-0.78) ve EAA/BRUTO (0.70-0.77) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 3 olduğu durumda DAA (0.62-0.64) aralığında, KAA (0.70-0.73), EAA/MARS degree=1 (0.67-0.68), EAA/MARS degree=2 (0.68-0.70) ve EAA/BRUTO (0.66-0.68) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA diğer yöntemlerden daha düşük performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.55-0.56) aralığında KAA (0.71-0.73), EAA/MARS degree=1 (0.69-0.71), EAA/MARS degree=2 (0.68-0.73) ve EAA/BRUTO (0.68-0.72) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da DAA diğer yöntemlerden daha düşük performans göstermiştir.

Set 8'de varyansın 1 olduğu koşulda DAA (0.71-0.92) aralığında, KAA (0.80-0.92), EAA/MARS degree=1 (0.78-0.92), EAA/MARS degree=2 (0.78-0.93) ve EAA/BRUTO (0.78-0.92) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermiştir. Varyansın 3 olduğu durumda DAA (0.83-0.84), KAA (0.85-0.87), EAA/MARS degree=1 (0.80-0.85), EAA/MARS degree=2 (0.84-0.87) ve EAA/BRUTO (0.84-0.84) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da tüm yöntemler birbirlerine yakın DSO

değerleri vermiştir. Varyansın 10 olduğu durumda ise, DAA (0.71-0.72) aralığında DSO değerleri verirken, KAA (0.80-0.82), EAA/MARS degree=1 (0.77-0.79), EAA/MARS degree=2 (0.78-0.80) ve EAA/BRUTO (0.78-0.79) değer aralıklarında DSO değerleri vermişlerdir. Bu DSO değerlerine göre ortalama vektörleri farklı uzaklıklarda ancak varyansların büyük olması nedeniyle üst üste binmiş sınıfları içeren bu sette DAA yöntemi diğer yöntemlere göre daha düşük performans göstermiştir. KAA en yüksek doğruluk derecesinde sınıflama yaparken EAA yöntemleri KAA'ya yakın ancak ondan daha düşük DSO değerleri vermişlerdir.

Set 9'da varyansın 1 olduğu koşulda DAA (0.79-0.91) aralığında, KAA (0.87-0.92), EAA/MARS degree=1 (0.81-0.91), EAA/MARS degree=2 (0.83-0.91) ve EAA/BRUTO (0.80-0.91) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermiştir. Varyansın 3 olduğu durumda DAA (0.78-0.79) aralığında, KAA (0.82-0.84), EAA/MARS degree=1 (0.80-0.81), EAA/MARS degree=2 (0.81-0.83) ve EAA/BRUTO (0.80-0.80) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda da tüm yöntemler birbirlerine yakın DSO değerleri vermişlerdir. Varyansın 10 olduğu durumda ise, DAA yöntemi (0.66-0.79) aralığında DSO değerleri verirken, KAA (0.79-0.84), EAA/MARS degree=1 (0.76-0.81), EAA/MARS degree=2 (0.77-0.83) ve EAA/BRUTO (0.77-0.80) değer aralıklarında DSO değerleri vermişlerdir. Bu DSO değerlerine göre ortalama vektörleri farklı uzaklıklarda ancak varyansların büyük olması nedeniyle üst üste binmiş sınıfları içeren bu sette DAA yöntemi diğer yöntemlere göre daha düşük performans göstermiştir. KAA en yüksek doğruluk derecesinde sınıflama yaparken EAA yöntemleri KAA'ya yakın ancak ondan daha düşük DSO değerleri vermiştir.

Sonuç olarak; 3 gruplu ve kovaryansı 0.25 olan 4 veri setinde, varyansın 1 olduğu koşulda tüm yöntemler birbirlerine yakın performans gösterirken, varyansın 3 ve 10 olduğu koşulda DAA'nın diğer yöntemlerden daha düşük bir performans gösterdiği bulunmuştur.

Çizelge 4.7'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.75 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Çizelge 4.7- Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.75 olan (V7,V8,V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
3 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9
SET6	150	0.508	.499	0.498	0.543	0.749	0.750	0.495	0.673	0.673	0.485	0.696	0.677	0.497	0.671	0.671
	300	0.475	0.512	0.475	0.724	0.541	0.725	0.497	0.501	0.663	0.681	0.494	0.681	0.505	0.492	0.672
	750	0.484	0.456	0.452	0.493	0.467	0.705	0.483	0.456	0.644	0.484	0.455	0.665	0.483	0.455	0.656
	1500	0.482	0.452	0.440	0.487	0.458	0.698	0.482	0.454	0.631	0.482	0.455	0.653	0.482	0.453	0.645
	3000	0.481	0.459	0.434	0.483	0.656	0.694	0.480	0.555	0.622	0.481	0.605	0.643	0.480	0.559	0.637
SET7	150	0.629	0.592	0.557	0.651	0.751	0.769	0.627	0.633	0.701	0.627	0.679	0.706	0.616	0.617	0.700
	300	0.712	0.583	0.535	0.634	0.735	0.749	0.623	0.638	0.693	0.626	0.687	0.708	0.622	0.631	0.699
	750	0.619	0.577	0.525	0.623	0.725	0.737	0.620	0.632	0.676	0.621	0.683	0.695	0.617	0.633	0.685
	1500	0.616	0.574	0.519	0.618	0.722	0.732	0.617	0.627	0.669	0.619	0.679	0.685	0.615	0.629	0.677
	3000	0.616	0.573	0.518	0.617	0.720	0.729	0.615	0.626	0.666	0.616	0.675	0.679	0.614	0.626	0.672
SET8	150	0.829	0.775	0.685	0.840	0.854	0.823	0.836	0.795	0.767	0.847	0.829	0.782	0.832	0.787	0.765
	300	0.825	0.769	0.679	0.830	0.844	0.812	0.828	0.791	0.763	0.835	0.819	0.777	0.826	0.785	0.867
	750	0.824	0.765	0.675	0.824	0.839	0.805	0.824	0.785	0.753	0.828	0.812	0.765	0.824	0.783	0.760
	1500	0.823	0.765	0.673	0.823	0.837	0.802	0.823	0.715	0.748	0.825	0.808	0.760	0.824	0.781	0.753
	3000	0.823	0.764	0.672	0.823	0.836	0.801	0.823	0.779	0.747	0.823	0.804	0.757	0.824	0.780	0.750
SET9	150	0.856	0.750	0.655	0.864	0.847	0.819	0.854	0.777	0.761	0.863	0.812	0.775	0.853	0.761	0.759
	300	0.852	0.742	0.644	0.856	0.835	0.810	0.848	0.770	0.756	0.855	0.805	0.771	0.847	0.761	0.761
	750	0.851	0.740	0.637	0.852	0.832	0.804	0.844	0.767	0.747	0.848	0.797	0.761	0.845	0.764	0.752
	1500	0.848	0.738	0.635	0.849	0.828	0.800	0.841	0.763	0.741	0.845	0.791	0.755	0.841	0.761	0.745
	3000	0.848	0.804	0.691	0.848	0.855	0.969	0.841	0.821	0.854	0.844	0.829	0.890	0.841	0.818	0.863

Set 6'da varyansı 1 olan koşulda DAA (0.48-0.51), KAA (0.48-0.72), EAA/MARS degree=1 (0.48-0.50), EAA/MARS degree=2 (0.48-0.68) ve EAA/BRUTO (0.48-0.51) değer aralıklarında DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.45-0.51), KAA (0.46-0.75), EAA/MARS degree=1 (0.45-0.67), EAA/MARS degree=2 (0.46-0.70) ve EAA/BRUTO (0.45-0.67) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler çok geniş bir aralıkta doğruluk oranlarına sahiptir. Ancak yöntemler düşük ve birbirlerine benzer türdeş olmayan performans düzeylerine sahiptir. Varyansın 10 olduğu durumda ise, DAA (0.43-0.50) aralığında, KAA (0.69-0.75), EAA/MARS degree=1 (0.62-0.67), EAA/MARS degree=2 (0.64-0.68) ve EAA/BRUTO (0.64-0.67) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA bu koşulda da diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Set 7'de de Set 6'ya benzer sonuçlar elde edilmiştir. Kovaryans matrislerinde homojenitenin kaybolması ile orantılı olarak KAA yöntemi diğer yöntemlere göre daha yüksek oranda sınıflandırma oranları vermiştir. Ancak heterojen veri yapıları söz konusu olduğunda EAA yöntemleri DAA yöntemine göre önemli düzeyde yüksek DSO değerleri vermişlerdir.

Set 8'de de grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise başka bir anlatım ile sınıflar üst üste binmemiş ise yöntemler sınıfları yüksek DSO değeriyle birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Set 9'da da Set 8'e benzer sonuçlar elde edilmiştir. Grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise başka bir anlatım ile sınıflar üst üste binmemiş ise yöntemler sınıfları yüksek doğruluk oranında birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Sonuç olarak; 3 gruplu ve kovaryansı 0.75 olan 4 veri setinde, grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük

ise yöntemler sınıfları yüksek DSO değerleriyle birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) ve varyans arttığında KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.8'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.90 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 6'da varyansı 1 olan koşulda DAA (0.47-0.50) aralığında, KAA (0.47-0.53), EAA/MARS degree=1 (0.47-0.51), EAA/MARS degree=2 (0.42-0.50) ve EAA/BRUTO (0.47-0.51) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyansın 3 olduğu durumda DAA (0.46-0.51), KAA (0.70-0.74), EAA/MARS degree=1 (0.55-0.56), EAA/MARS degree=2 (0.62-0.63) ve EAA/BRUTO (0.54-0.55) aralığında DSO değerleri vermişlerdir. Bu koşulda KAA diğer yöntemlere göre yüksek performans değerlerine sahiptir. Varyansın 10 olduğu durumda ise, DAA (0.43-0.50), KAA (0.70-0.76), EAA/MARS degree=1 (0.62-0.67), EAA/MARS degree=2 (0.64-0.68) ve EAA/BRUTO (0.63-0.67) değer aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Set 7'de de Set 6'ya benzer sonuçlar elde edilmiştir. Kovaryans matrislerinde homojenitenin kaybolması ile orantılı olarak KAA yöntemi diğer yöntemlere göre daha yüksek oranda DSO değerleri vermiştir. Ancak heterojen veri yapıları söz konusu olduğunda EAA yöntemleri DAA yöntemine göre önemli düzeyde yüksek DSO değerleri vermişlerdir..

Set 8 için sonuçlar incelendiğinde, gruplar homojen yapıda ve aralarında heterojen yapıda ise yöntemler daha iyi performans göstermektedir. Grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise yöntemler sınıfları yüksek DSO değerleriyle birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) yine KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.8- Ayırma Analizi Yöntemleri'nin, üç alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.90 olan (V10,V11,V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
3 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12
SET6	150	0.503	0.505	0.500	0.532	0.738	0.756	0.507	0.553	0.666	0.421	0.616	0.673	0.507	0.535	0.665
	300	0.486	0.481	0.474	0.526	0.718	0.733	0.484	0.558	0.658	0.504	0.631	0.678	0.482	0.557	0.665
	750	0.477	0.465	0.448	0.487	0.701	0.714	0.476	0.555	0.636	0.476	0.628	0.661	0.476	0.556	0.649
	1500	0.473	0.458	0.439	0.478	0.695	0.707	0.472	0.549	0.627	0.473	0.623	0.649	0.466	0.551	0.639
	3000	0.471	0.455	0.432	0.474	0.694	0.703	0.470	0.545	0.620	0.471	0.618	0.641	0.471	0.549	0.632
SET7	150	0.613	0.588	0.552	0.640	0.779	0.776	0.614	0.627	0.694	0.627	0.687	0.705	0.615	0.616	0.694
	300	0.605	0.573	0.531	0.619	0.762	0.755	0.607	0.626	0.685	0.608	0.692	0.703	0.605	0.619	0.691
	750	0.600	0.565	0.517	0.606	0.752	0.743	0.603	0.621	0.669	0.603	0.687	0.688	0.599	0.617	0.677
	1500	0.599	0.563	0.514	0.602	0.749	0.739	0.599	0.617	0.663	0.601	0.684	0.681	0.598	0.617	0.671
	3000	0.432	0.563	0.502	0.600	0.748	0.734	0.598	0.615	0.684	0.476	0.681	0.693	0.597	0.615	0.691
SET8	150	0.811	0.759	0.680	0.821	0.863	0.826	0.815	0.780	0.763	0.827	0.870	0.782	0.811	0.772	0.759
	300	0.806	0.754	0.671	0.811	0.854	0.814	0.807	0.779	0.755	0.815	0.815	0.773	0.803	0.770	0.761
	750	0.802	0.752	0.665	0.804	0.850	0.808	0.803	0.773	0.744	0.807	0.805	0.760	0.801	0.769	0.749
	1500	0.802	0.750	0.664	0.803	0.849	0.807	0.802	0.770	0.741	0.803	0.802	0.756	0.801	0.768	0.747
	3000	0.802	0.750	0.664	0.802	0.848	0.806	0.802	0.767	0.740	0.803	0.799	0.754	0.801	0.767	0.744
SET9	150	0.680	0.746	0.649	0.826	0.862	0.826	0.885	0.767	0.755	0.893	0.815	0.771	0.885	0.755	0.753
	300	0.883	0.740	0.639	0.887	0.853	0.816	0.880	0.766	0.750	0.887	0.808	0.769	0.880	0.758	0.754
	750	0.882	0.736	0.632	0.884	0.848	0.808	0.877	0.761	0.737	0.880	0.795	0.757	0.877	0.757	0.743
	1500	0.881	0.751	0.630	0.881	0.872	0.806	0.875	0.799	0.733	0.877	0.821	0.751	0.875	0.799	0.739
	3000	0.881	0.881	0.880	0.881	0.881	0.881	0.875	0.875	0.875	0.876	0.876	0.876	0.875	0.875	0.875

Set 9'da da Set 8'e benzer sonuçlar elde edilmiştir. Varyanslar küçük ise yöntemler sınıfları yüksek DSO değerlerinde birbirlerinden ayırmaktadırlar. Aksi durumda KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Sonuç olarak; 3 gruplu ve kovaryansı 0.90 olan 4 veri setinde, yine varyansların olduğu durumlarda yöntemlerin sınıfları yüksek DSO değerleriyle ayırdığı belirlenmiştir. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) ve varyans arttığında KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.9'da görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 10'da varyansı 1 olan koşulda DAA (0.42-0.63) aralığında, KAA (0.57-0.65), EAA/MARS degree=1 (0.53-0.62), EAA/MARS degree=2 (0.57-0.62) ve EAA/BRUTO (0.57-0.63) aralıklarında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyansın 3 olduğu durumda DAA (0.49-0.52) aralığında , KAA (0.57-0.61), EAA/MARS degree=1 (0.55-0.57), EAA/MARS degree=2 (0.55-0.57) ve EAA/BRUTO (0.55-0.57) aralıklarında DSO değerleri vermişlerdir. Bu koşulda KAA diğer yöntemlere göre yüksek performans göstermiştir. Varyansın 10 olduğu durumda ise, DAA (0.41-0.45) aralığında, KAA (0.57-0.62), EAA/MARS degree=1 (0.55-0.58), EAA/MARS degree=2 (0.56-0.58) ve EAA/BRUTO (0.56-0.59) aralıklarında DSO değerleri vermişlerdir. Bu koşulda DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA bu koşulda da diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Set 11'de de Set 10'a benzer sonuçlar elde edilmiştir. Kovaryans matrislerinde homojenitenin kaybolması ile orantılı olarak KAA yöntemi diğer yöntemlere göre daha yüksek DSO değerleri vermiştir ve DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür.

Set 12 için sonuçlar incelendiğinde, varyansı 1 ve varyans 3 olan koşulda yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyans 10 olan koşulda DAA'nın performansı diğer yöntemlere göre düşüktür. Diğer 4 yöntem benzer sonuçlar vermiş ve birbirlerine üstünlük sağlamamışlardır.

Set 13'de varyans 1 olan koşulda tüm yöntemler, genelde 0.90'ın üzerinde DSO değerleri vermişlerdir. Yöntemler birbirlerine üstünlük sağlamamışlardır. Varyans 3 olan koşulda yöntemlerin genelde DSO değerleri 0.80 civarındadır. 0.01-0.02 oranında da olsa KAA yöntemi diğer yöntemlere göre daha iyi performans göstermiştir. Varyans 10 olan koşulda DAA'nın performansı diğer yöntemlere göre düşüktür. KAA ise diğer yöntemlere göre daha yüksek performans göstermiştir.

Sonuç olarak; 4 gruplu ve kovaryansı 0 olan 4 veri setinde, KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.10'da görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.25 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 10'da varyansı 1 olan koşulda DAA (0.54-0.61) aralığında, KAA (0.54-0.62), EAA/MARS degree=1 (0.54-0.60), EAA/MARS degree=2 (0.54-0.61) ve EAA/BRUTO (0.54-0.61) aralıklarında DSO değerleri vermişlerdir. Varyansın 3 olduğu durumda DAA (0.47-0.49) aralığında, KAA (0.55-0.60), EAA/MARS degree=1 (0.33-0.54), EAA/MARS degree=2 (0.52-0.55) ve EAA/BRUTO (0.51-0.54) aralıklarında DSO değerleri vermişlerdir. Bu koşulda KAA, diğer yöntemlere göre yüksek performans değerlerine sahiptir. Varyansın 10 olduğu durumda ise, DAA (0.40-0.44) aralığında, KAA (0.57-0.62), EAA/MARS degree=1 (0.55-0.57), EAA/MARS degree=2 (0.56-0.57) ve EAA/BRUTO (0.56-0.58) aralıklarında DSO değerleri vermişlerdir. Bu koşula göre DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA bu koşulda da diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Set 11'de varyans 1 olan koşulda yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyans 3 ve varyans 10 olan koşullarda DAA'nın performansı diğer yöntemlere göre en düşüktür.

Çizelge 4.9 - Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0 olan (V1,V2,V3 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
4 GRUPLU VERİ SETLERİ	$N=n_1+n_2+n_3+n_4$	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3	V1	V2	V3
SET10	200	0.628	0.519	0.453	0.652	0.612	0.617	0.622	0.562	0.575	0.614	0.550	0.556	0.634	0.544	0.564
	400	0.620	0.506	0.434	0.633	0.588	0.594	0.620	0.565	0.578	0.621	0.565	0.578	0.622	0.562	0.587
	1000	0.616	0.498	0.421	0.621	0.573	0.576	0.611	0.559	0.562	0.620	0.560	0.572	0.612	0.567	0.574
	2000	0.415	0.495	0.415	0.570	0.568	0.569	0.555	0.552	0.555	0.565	0.557	0.565	0.566	0.561	0.566
	4000	0.613	0.494	0.411	0.615	0.566	0.566	0.529	0.550	0.551	0.611	0.556	0.559	0.606	0.556	0.559
SET11	200	0.518	0.624	0.518	0.645	0.681	0.645	0.605	0.637	0.606	0.597	0.635	0.597	0.610	0.633	0.611
	400	0.768	0.616	0.504	0.774	0.662	0.620	0.764	0.636	0.604	0.773	0.641	0.607	0.767	0.635	0.613
	1000	0.764	0.611	0.496	0.767	0.650	0.607	0.755	0.628	0.593	0.769	0.639	0.603	0.760	0.634	0.602
	2000	0.493	0.610	0.493	0.601	0.646	0.601	0.586	0.627	0.586	0.595	0.636	0.595	0.594	0.632	0.594
	4000	0.763	0.609	0.492	0.764	0.644	0.599	0.752	0.624	0.583	0.761	0.633	0.591	0.755	0.629	0.589
SET12	200	0.969	0.847	0.969	0.972	0.867	0.972	0.969	0.846	0.972	0.967	0.858	0.967	0.972	0.854	0.972
	400	0.768	0.843	0.671	0.774	0.849	0.732	0.764	0.843	0.712	0.773	0.855	0.717	0.767	0.848	0.720
	1000	0.966	0.840	0.668	0.967	0.853	0.724	0.962	0.836	0.704	0.961	0.846	0.710	0.966	0.842	0.713
	2000	0.840	0.840	0.666	0.852	0.852	0.720	0.834	0.835	0.700	0.842	0.842	0.705	0.840	0.840	0.708
	4000	0.966	0.839	0.666	0.966	0.851	0.719	0.960	0.833	0.698	0.960	0.840	0.703	0.963	0.837	0.705
SET13	200	0.938	0.809	0.647	0.943	0.841	0.742	0.936	0.810	0.707	0.938	0.824	0.709	0.942	0.815	0.710
	400	0.936	0.805	0.641	0.939	0.831	0.727	0.933	0.811	0.706	0.935	0.823	0.711	0.938	0.812	0.666
	1000	0.802	0.802	0.635	0.825	0.825	0.717	0.806	0.806	0.696	0.815	0.815	0.703	0.810	0.810	0.705
	2000	0.935	0.802	0.634	0.934	0.824	0.715	0.929	0.804	0.692	0.929	0.811	0.700	0.931	0.808	0.701
	4000	0.935	0.801	0.633	0.935	0.823	0.713	0.928	0.803	0.690	0.929	0.808	0.697	0.929	0.806	0.697

Set 12'de varyans 1 olan koşulda tüm yöntemler, genellikle 0.90 civarında DSO değerleri vermişlerdir. Varyans 3 olan koşulda yöntemler birbirlerine yakın DSO değerleri vermektedir. 0.01-0.02 oranında da olsa KAA yöntemi diğer yöntemlere göre daha iyi performans göstermiştir. Varyans 10 olan koşulda DAA'nın performansı diğer yöntemlere göre çok düşüktür. KAA ise diğer yöntemlere göre daha yüksek performans göstermiştir.

Set 13'de varyans 1 olan koşulda tüm yöntemler, genellikle 0.88 civarında DSO değerleri vermişlerdir. Yöntemler birbirlerine göre üstünlük sağlamamışlardır. Varyans 3 olan koşulda DAA'nın performansı diğer yöntemlere göre çok düşüktür. KAA ise diğer yöntemlere göre daha iyi performans göstermiştir. Varyans 10 olan koşulda DAA'nın performansı diğer yöntemlere göre çok düşüktür. 0.02-0.03 oranında da olsa KAA yöntemi diğer yöntemlere göre daha iyi performans göstermiştir.

Sonuç olarak; 4 gruplu ve kovaryansı 0.25 olan 4 veri setinde, yine varyansların küçük olduğu durumlarda yöntemlerin sınıfları yüksek DSO değerleriyle ayırdığı belirlenmiştir. Varyans arttığında KAA, DAA'ya göre daha etkin DSO değerleri vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.11'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.75 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 10'da varyansı 1 olan koşulda DAA (0.47-0.49) aralığında, KAA (0.47-0.52), EAA/MARS degree=1 (0.47-0.48), EAA/MARS degree=2 (0.47-0.47) ve EAA/BRUTO (0.46-0.49) aralığında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyansın 3 olduğu durumda DAA (0.43-0.47), KAA (0.58-0.62), EAA/MARS degree=1 (0.49-0.50), EAA/MARS degree=2 (0.41-0.54) ve EAA/BRUTO (0.48-0.51) aralığında DSO değerleri vermişlerdir. Bu koşulda KAA, diğer yöntemlere göre yüksek performansa sahiptir. Varyansın 10 olduğu durumda ise, DAA (0.38-0.44), KAA (0.58-0.63), EAA/MARS degree=1 (0.52-0.55), EAA/MARS degree=2 (0.53-0.56) ve EAA/BRUTO (0.53-0.56) aralığında DSO değerleri vermişlerdir. Bu koşulda DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA bu koşulda da diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Çizelge 4.10- Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.25 olan (V4,V5,V6 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
4 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃ +n ₄	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6	V4	V5	V6
SET10	200	0.555	0.491	0.443	0.584	0.597	0.620	0.544	0.526	0.573	0.538	0.520	0.555	0.552	0.511	0.576
	400	0.545	0.478	0.423	0.560	0.575	0.593	0.545	0.539	0.571	0.542	0.544	0.571	0.548	0.541	0.580
	1000	0.539	0.469	0.408	0.546	0.561	0.575	0.539	0.536	0.554	0.543	0.545	0.566	0.538	0.544	0.568
	2000	0.537	0.466	0.402	0.541	0.556	0.569	0.535	0.531	0.548	0.540	0.540	0.560	0.535	0.538	0.559
	4000	0.613	0.474	0.411	0.615	0.547	0.566	0.604	0.329	0.551	0.611	0.537	0.559	0.606	0.537	0.559
SET11	200	0.676	0.583	0.500	0.696	0.658	0.643	0.670	0.599	0.597	0.669	0.603	0.587	0.678	0.593	0.601
	400	0.669	0.574	0.485	0.679	0.639	0.619	0.672	0.604	0.598	0.674	0.614	0.601	0.672	0.601	0.608
	1000	0.665	0.569	0.475	0.669	0.627	0.605	0.665	0.598	0.584	0.671	0.612	0.594	0.669	0.604	0.595
	2000	0.664	0.567	0.472	0.666	0.623	0.599	0.661	0.594	0.578	0.667	0.606	0.588	0.662	0.601	0.585
	4000	0.763	0.583	0.492	0.763	0.618	0.599	0.752	0.599	0.583	0.761	0.608	0.591	0.755	0.603	0.589
SET12	200	0.908	0.803	0.659	0.915	0.832	0.741	0.908	0.805	0.702	0.912	0.818	0.704	0.914	0.810	0.707
	400	0.905	0.800	0.652	0.908	0.823	0.726	0.904	0.804	0.701	0.903	0.817	0.705	0.910	0.808	0.710
	1000	0.903	0.797	0.648	0.904	0.818	0.718	0.899	0.796	0.692	0.898	0.809	0.699	0.904	0.802	0.702
	2000	0.903	0.796	0.647	0.903	0.816	0.714	0.898	0.794	0.688	0.896	0.804	0.694	0.902	0.799	0.697
	4000	0.806	0.806	0.665	0.818	0.818	0.719	0.802	0.802	0.697	0.809	0.809	0.702	0.806	0.806	0.704
SET13	200	0.877	0.770	0.629	0.885	0.814	0.737	0.876	0.774	0.694	0.883	0.792	0.699	0.882	0.795	0.701
	400	0.874	0.763	0.621	0.878	0.801	0.723	0.873	0.776	0.694	0.876	0.789	0.700	0.878	0.804	0.704
	1000	0.873	0.760	0.615	0.875	0.795	0.712	0.868	0.770	0.684	0.871	0.782	0.693	0.872	0.774	0.695
	2000	0.872	0.759	0.613	0.873	0.793	0.709	0.866	0.769	0.681	0.869	0.777	0.688	0.870	0.773	0.691
	4000	0.935	0.769	0.633	0.935	0.791	0.713	0.928	0.773	0.690	0.928	0.778	0.698	0.929	0.776	0.697

Çizelge 4.11- Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.75 olan (V7,V8,V9 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
4 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃ +n ₄	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9	V7	V8	V9
SET10	200	0.487	0.465	0.435	0.521	0.622	0.630	0.476	0.484	0.551	0.470	0.505	0.534	0.486	0.480	0.554
	400	0.477	0.449	0.414	0.495	0.601	0.605	0.473	0.500	0.553	0.470	0.538	0.555	0.478	0.510	0.563
	1000	0.470	0.437	0.395	0.479	0.585	0.587	0.469	0.503	0.546	0.469	0.541	0.549	0.470	0.510	0.548
	2000	0.467	0.432	0.387	0.472	0.580	0.581	0.467	0.500	0.527	0.467	0.537	0.543	0.466	0.503	0.539
	4000	0.466	0.429	0.382	0.468	0.577	0.577	0.465	0.496	0.522	0.465	0.407	0.536	0.464	0.498	0.533
SET11	200	0.584	0.534	0.479	0.610	0.665	0.651	0.581	0.551	0.577	0.578	0.580	0.570	0.584	0.545	0.777
	400	0.576	0.525	0.462	0.590	0.645	0.628	0.578	0.560	0.578	0.575	0.590	0.583	0.575	0.560	0.588
	1000	0.570	0.451	0.451	0.576	0.614	0.614	0.572	0.663	0.563	0.572	0.575	0.675	0.568	0.574	0.574
	2000	0.568	0.515	0.447	0.571	0.628	0.608	0.569	0.551	0.558	0.570	0.585	0.567	0.567	0.555	0.566
	4000	0.567	0.514	0.445	0.568	0.625	0.605	0.567	0.548	0.554	0.568	0.580	0.563	0.566	0.551	0.560
SET12	200	0.808	0.738	0.629	0.820	0.803	0.739	0.810	0.745	0.677	0.819	0.770	0.689	0.804	0.749	0.683
	400	0.803	0.734	0.620	0.809	0.792	0.723	0.805	0.745	0.678	0.810	0.768	0.688	0.803	0.748	0.688
	1000	0.802	0.729	0.616	0.804	0.785	0.715	0.801	0.739	0.768	0.803	0.756	0.678	0.802	0.741	0.676
	2000	0.801	0.729	0.613	0.802	0.783	0.711	0.800	0.738	0.664	0.801	0.752	0.673	0.802	0.740	0.671
	4000	0.800	0.728	0.613	0.801	0.782	0.710	0.799	0.736	0.661	0.800	0.749	0.670	0.801	0.738	0.668
SET13	200	0.827	0.719	0.604	0.837	0.798	0.739	0.827	0.731	0.673	0.835	0.757	0.682	0.823	0.731	0.680
	400	0.824	0.712	0.594	0.829	0.786	0.723	0.820	0.730	0.670	0.825	0.750	0.683	0.820	0.730	0.685
	1000	0.820	0.708	0.587	0.822	0.778	0.712	0.819	0.727	0.662	0.819	0.745	0.672	0.818	0.726	0.670
	2000	0.820	0.708	0.586	0.821	0.777	0.711	0.817	0.725	0.659	0.817	0.740	0.668	0.816	0.726	0.666
	4000	0.820	0.707	0.613	0.821	0.775	0.710	0.817	0.723	0.662	0.816	0.738	0.670	0.815	0.723	0.668

Set 11'de varyans 1 olan koşulda yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyans 3 ve varyans 10 olan koşullarda DAA en düşük performansı göstermiş ve KAA diğer yöntemlere göre daha yüksek performans göstermiştir.

Set 12'de varyans 1 olan koşulda bütün yöntemler 0.80 civarında DSO değerleri vererek, birbirlerine yakın performans göstermişlerdir. Varyans 3 ve varyans 10 olan koşullarda DAA en düşük performansı göstermiş ve KAA diğer yöntemlere göre daha yüksek performans göstermiştir.

Set 13'de varyans 1 olan koşulda bütün yöntemler 0.82 civarında DSO değerleri vererek, birbirlerine yakın performans göstermişlerdir. Varyans 3 ve varyans 10 olan koşullarda DAA en düşük performansı göstermiş ve KAA diğer yöntemlere göre daha yüksek performans göstermiştir.

Sonuç olarak; 4 gruplu ve kovaryansı 0.75 olan 4 veri setinde, yine varyansların olduğu durumlarda yöntemlerin sınıfları yüksek doğruluk oranında ayırdığı belirlenmiştir. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) ve varyans arttığında KAA, DAA'ya göre daha etkin doğru sınıflandırma sonuçları vermektedir. EAA yöntemleri ise DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri vermektedir.

Çizelge 4.12'de görüleceği üzere 5 ayırma analizi yönteminin kovaryansı 0.90 koşulunda aşağıda belirtilen sonuçlara ulaşılmıştır. Beş ayırma yönteminin performans düzeyleri (DSO değerleri) aşağıdaki gibi belirlenmiştir.

Set 10'da varyansı 1 olan koşulda DAA (0.45-0.48) aralığında, KAA (0.46-0.51), EAA/MARS degree=1 (0.45-0.47), EAA/MARS degree=2 (0.45-0.46) ve EAA/BRUTO (0.45-0.48) aralığında DSO değerleri vermişlerdir. Bu koşulda tüm yöntemler birbirlerine yakın DSO değerleri vermektedir. Varyansın 3 olduğu durumda DAA (0.42-0.46), KAA (0.60-0.65), EAA/MARS degree=1 (0.48-0.49), EAA/MARS degree=2 (0.51-0.55) ve EAA/BRUTO (0.48-0.50) aralığında DSO değerleri vermişlerdir. Bu koşulda, KAA diğer yöntemlere göre yüksek performans değerlerine sahiptir. koşulda DAA'nın performansı yüksek varyans düzeylerinde diğer yöntemlere göre düşüktür. KAA bu koşulda da diğer yöntemlere göre yüksek performans değerlerine sahiptir.

Çizelge 4.12- Ayırma Analizi Yöntemleri'nin, dört alt gruplu bağımlı değişken ve varyansı bir, üç ve on olan ve kovaryansları 0.90 olan (V10,V11,V12 kovaryans matrisleri) bağımsız değişkenlere sahip her alt gruptaki n=50, 100, 250, 500 ve 1000 gözlemi DSO değerleri

Ayırma Analizi Yöntemleri		DAA			KAA			EAA/MARS degree=1			EAA/MARS degree=2			EAA/BRUTO		
4 GRUPLU VERİ SETLERİ	N=n ₁ +n ₂ +n ₃ +n ₄	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12	V10	V11	V12
SET10	200	0.476	0.462	0.438	0.510	0.647	0.634	0.466	0.476	0.544	0.460	0.511	0.530	0.477	0.476	0.548
	400	0.463	0.444	0.413	0.482	0.625	0.610	0.460	0.490	0.545	0.455	0.540	0.550	0.463	0.495	0.558
	1000	0.457	0.431	0.394	0.467	0.610	0.594	0.455	0.493	0.531	0.455	0.548	0.545	0.456	0.497	0.544
	2000	0.455	0.425	0.385	0.460	0.604	0.587	0.453	0.491	0.522	0.453	0.544	0.539	0.453	0.493	0.534
	4000	0.453	0.421	0.379	0.455	0.600	0.583	0.451	0.487	0.517	0.451	0.538	0.531	0.450	0.488	0.528
SET11	200	0.567	0.529	0.476	0.595	0.682	0.656	0.564	0.544	0.572	0.564	0.581	0.569	0.567	0.540	0.580
	400	0.557	0.514	0.458	0.572	0.661	0.632	0.558	0.548	0.573	0.560	0.590	0.578	0.553	0.545	0.583
	1000	0.551	0.507	0.446	0.557	0.650	0.618	0.551	0.546	0.559	0.552	0.589	0.570	0.549	0.546	0.569
	2000	0.550	0.505	0.441	0.553	0.646	0.612	0.550	0.540	0.552	0.551	0.583	0.563	0.547	0.543	0.559
	4000	0.548	0.503	0.439	0.550	0.643	0.609	0.548	0.537	0.548	0.549	0.579	0.557	0.546	0.539	0.554
SET12	200	0.784	0.724	0.619	0.797	0.807	0.740	0.786	0.731	0.669	0.794	0.762	0.682	0.780	0.734	0.672
	400	0.780	0.717	0.612	0.787	0.795	0.725	0.783	0.730	0.670	0.785	0.758	0.683	0.773	0.730	0.680
	1000	0.778	0.717	0.607	0.781	0.789	0.716	0.778	0.727	0.661	0.782	0.747	0.673	0.776	0.728	0.671
	2000	0.777	0.712	0.605	0.779	0.787	0.713	0.777	0.724	0.657	0.779	0.743	0.618	0.776	0.725	0.665
	4000	0.776	0.712	0.604	0.777	0.786	0.711	0.776	0.722	0.654	0.777	0.739	0.664	0.776	0.724	0.661
SET13	200	0.710	0.710	0.598	0.805	0.805	0.739	0.724	0.724	0.666	0.754	0.754	0.676	0.722	0.722	0.673
	400	0.838	0.704	0.588	0.843	0.795	0.724	0.838	0.723	0.663	0.843	0.750	0.678	0.838	0.720	0.675
	1000	0.837	0.700	0.582	0.837	0.788	0.716	0.835	0.721	0.655	0.838	0.741	0.668	0.834	0.719	0.664
	2000	0.836	0.699	0.579	0.837	0.785	0.713	0.834	0.717	0.652	0.835	0.737	0.663	0.834	0.717	0.659
	4000	0.836	0.698	0.578	0.836	0.784	0.711	0.833	0.715	0.650	0.833	0.734	0.659	0.848	0.714	0.655

Set 10'da Varyansın 10 olduđu durumda ise, DAA (0.38-0.44), KAA (0.58-0.63), EAA/MARS degree=1 (0.52-0.55), EAA/MARS degree=2 (0.53-0.55) ve EAA/BRUTO (0.53-0.56) aralığında DSO deęerleri vermiřlerdir. Bu Set 11'de varyans 1 olan kořulda yontemler birbirlerine yakın DSO deęerleri vermektedir. Varyans 3 ve varyans 10 olan kořullarda DAA en dűřűk performansı gűstermiř ve KAA dięer yontemlere gűre daha yűksek performans gűstermiřtir.

Set 12'de varyans 1 olan kořulda bűtűn yontemler 0.78 civarında DSO deęerleri vererek, birbirlerine yakın performans gűstermiřlerdir. Varyans 3 ve varyans 10 olan kořullarda KAA dięer yontemlere gűre daha yűksek performans gűsterirken, DAA en dűřűk performansı gűstermiřtir.

Set 13'de varyans 1 olan kořulda bűtűn yontemler genellikle 0.80'in űzerinde DSO deęerleri vererek, birbirlerine yakın performans gűstermiřlerdir. Varyans 3 ve varyans 10 olan kořullarda DAA en dűřűk performansı gűstermiř ve KAA dięer yontemlere gűre daha yűksek performans gűstermiřtir. EAA yontemleri ise benzer sonuęlar vermiřtir.

Sonuę olarak; 4 grulu ve kovaryansı 0.90 olan 4 veri setinde, yine varyansların kűűk olduđu durumlarda yontemlerin sınıfları yűksek doęruluk oranında ayırdıęı belirlenmiřtir. Ancak birimlerin űzellikleri birbirlerine karıřtıęında ve varyans arttıęında KAA, DAA'ya gűre daha etkin DSO deęerleri vermektedir. EAA yontemleri ise DAA'ya gűre yűksek ancak KAA'dan daha dűřűk performans deęerleri vermektedir.

Normal daęılım gűstermeyen daęılımdan tűretilmiř g=2 ve g=3 grulu sınırlı kořullara gűre elde edilmiř veri setlerinde ayırma yontemlerinin performanslarına iliřkin bulgular izelge 4.13'de verilmiřtir.

Çizelge 4.13 - Ayırma Analizi Yöntemleri'nin, iki (SET 14) ve üç (SET 15) veri matrisi birleştirmeli ve her biri 3 alt gruplu normal dağılıma uygunluk göstermeyen her alt sette DSO değerleri

Veri Setleri	Grup	N	Varyans	DAA	KAA	EAA/MARS Degree=1	EAA/MARS Degree= 2	EAA/BRUTO
SET14	g=2	200	V1	0.712	0.722	0.730	0.735	0.715
			V2	0.676	0.698	0.705	0.710	0.677
			V3	0.655	0.681	0.682	0.692	0.655
SET15	g=3	300	V7	0.888	0.895	0.893	0.903	0.893
			V8	0.603	0.654	0.627	0.647	0.607
			V9	0.815	0.904	0.850	0.863	0.853

Çizelge 4.13'de görüleceği üzere 5 ayırma analizi yönteminin iki (Set 14) ve üç (Set 15) gruplu veri matrisi birleştirmeli ve her biri 12 alt gruplu normal dağılıma uygunluk göstermeyen her alt sette DSO değerleri normal dağılımdan türetilmiş verilere göre düşük DSO değerleri vermiştir. DAA ve KAA normal dağılmayan verilere uygulanan yöntemler olmamasına rağmen bir fikir vermesi için analizlere alınmıştır. Ancak normalden farklı dağılan ve kovaryansın arttığı durumlarda her beş yöntem de 0.85 dolaylarında ve birbirlerinden farklı olmayan DSO değerleri vermiştir. KAA varyans heterojenliğine duyarlılığı sürdürerek 0.90 dolayında DSO değeri vererek en yüksek performansa sahip yöntem olarak ortaya çıkmıştır. Ancak veri matrisinin normal dağılmadığını göz önüne almak gerekmektedir. EAA modelleri olarak ele alınan yöntemler hem g=2 ve g=3 için hemde varyans kovaryans yapılarına göre birbirlerine benzer sonuçlar vermiştir. Ancak BRUTO, her koşulda MARS modellerinden biraz daha düşük performans değerlerine sahiptir.

5- TARTIŞMA VE SONUÇ

Sınıflama, Gruplama ve Atama yaklaşımları tıbbi problemlerin çözümünde önemli sorunlar yaratmaktadır. Veri yapısı, tipi ve birim sayıları sınıflandırma işlemlerinde farklı yaklaşımların ve modellerin kullanılmasını gerektirmektedir. Verilere uyan geçerli, güvenilir ve doğruluk derecesi yüksek sonuçların elde edilmesinde uygun yöntemin ve uygun modelin seçimi büyük önem taşımaktadır (9-11, 22, 33).

Ancak bu güne değin çalışılan yöntem ve modellerden sınıflamayı %100'e yakın oranda belirleyen bir yaklaşım henüz bulunmamaktadır. Yöntemlerin verilerin ortalama vektörlerine ve kovaryans matrislerine bağlı olarak çok değişik yelpazede yapısal özellik göstermeleri yöntem/modellerin etkinlikleri üzerine olumsuz etkilerde bulunmaktadır. Her yöntemin uygulanabilmesi için ideal koşullar belirlenmiş olmasına karşın gözlemsel veri setlerinde bu koşulların tümü gerçekleşmemektedir. Bu durum yöntemler içinden veri setinin analizi için en uygun olanını seçme sorunlarını gündeme getirmektedir. Tıbbi problemlerde bu durumların gözlenmesi, tıbbi problemlerin çözümünde uygun yöntem/model seçimlerinin daha titiz yapılmasını gerektirmiştir. Bu nedenlerle benzer problemlerin çözümü için çok sayıda yöntem/model geliştirilmiştir. Bu modellerin problem tiplerine, verilerin taşıdığı koşullara en uygun olanını seçmek gerekmektedir. Bu amaçla çok sayıda araştırma yapılmış ve yöntemler/yöntem içi modeller değişik koşullara sahip veri setlerinde performansları araştırılmıştır (1, 8, 11, 13-17, 22, 27).

Bu araştırmada; değişik sayıda birim içeren, değişik ortalama vektörü ve varyans-kovaryans yapılarına sahip türetilmiş veri setleri kullanılarak **X** veri matrisinin birim sayısının, ortalama vektörünün, varyans ve değişkenler arası kovaryans düzeylerinin yöntemlerin/modellerin sınıflama performanslarına etkileri incelenmiştir.

Ortalama vektörleri birbirlerine yakın, varyansları bir olan, değişkenler arası kovaryansları sıfır olan veri setlerinde DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemleri %89 ve daha

yukarı yüksek sınıflandırma performansları göstermiştir. Çokdeğişkenli normal dağılımdan türetilmiş ve ortalama vektörleri önemli farklılık gösteren iki gruplu verilerde yapılan araştırmalar ile araştırma bulguları paralellik göstermektedir (9-11, 13-17).

Veri setlerinde ortalama vektörleri ve birim sayısı değişmemesine karşın kovaryans matrislerinin heterojen yapı gösterdiği durumlarda yöntemlerin hemen tümünün doğru sınıflandırma oranları başka bir anlatımla performansları önemli düzeyde düşme göstermektedir. Kovaryans düzeylerinin 0.75 ve 0.90 olduğu durumlarda DAA yöntemi performansını kaybetmekte ancak KAA yöntemi yüksek performans değerlerine ulaşmaktadır. Bu durum KAA'nın heterojen kovaryans matrislerine sahip çok değişkenli veri setlerinde DAA'ya göre güçlü bir yöntem olduğu görüşünü doğrulamaktadır (13-18, 23, 26).

Grup sayısı iki olduğu durumlarda kovaryans matrislerinin homojen ve düşük kovaryans düzeyleri içerdiği durumlarda EAA yöntemleri DAA ve KAA yöntemleri ile benzer sonuçlar vermektedir. Bu göstermektedir ki, Çokdeğişkenli normal dağılım gösteren homojen gruplarda parametrik olmayan ayırma yöntemlerinin güçleri parametrik yöntemlerin güçlerine erişmektedir. Ancak heterojen veri yapılarında KAA'ya yakın DAA'dan daha yüksek performans değerlerine sahip olarak, güçleri DAA'dan daha yüksektir. Grup sayısının düşük ve heterojen kovaryans matrislerine sahip verilerin analizinde EAA yöntemleri etkin biçimde kullanılabilir olarak belirlenmiştir. Friedman, 1991 yılındaki yapmış olduğu çalışmada Çokdeğişkenli normal dağılım göstermeyen veri yapılarında MARS degree=1 ve MARS degree=2 modellerinin etkin sınıflama sonuçları verdiğini bulmuştur (11). Akdağ ve ark. 2006 yılı çalışmalarında MARS degree=1 yönteminin parametrik yöntemlere göre daha yüksek performans sergilediğini belirtmişlerdir (1).

Langaas 1995 yılında gerçek ve türetilmiş veri setlerinde yaptığı çalışmasında geleneksel ayırma analizleri (DAA ve KAA) ile halen gelişmekte olan Yapay Sinir Ağları ve EAA analizlerinin performanslarını karşılaştırmıştır. Gerçek bir veri seti üzerinde iki farklı koşula göre uygulama yapmıştır. İlk

koşulda EAA DAA'dan anlamlı olarak daha iyi bir performans göstermiştir. Yalnız bu koşulda EAA/BRUTO metodu çalışmamıştır. İkinci koşulda EAA ile DAA yöntemleri anlamlı bir fark bulunmamıştır (20).

Anne ve Solberg 2004 yılında yaptıkları çalışmada sınıflandırma yöntemlerinin performanslarını karşılaştırmışlardır. Karşılaştırma sonucunda EAA'nın DAA ve KAA'dan anlamlı olarak daha yüksek doğru sınıflandırma oranına sahip olduğunu bulmuşlardır. Ayrıca EAA'nın sınıflandırma amaçları için iyi bir alternatif olarak görülebileceğini ve normal dağılmayan veri setlerinde tutarlı, güvenilir ve geçerli sonuçlar verdiğini göstermişlerdir (4).

Türe ve ark. 2005 yılında hipertansiyonla ilgili bir veri seti üzerinde yaptıkları çalışmalarında farklı sınıflandırma tekniklerinin performanslarını karşılaştırmışlardır. Bu amaçla hipertansiyon ve kontrol gruplarını tahmin etmek için lojistik regresyon analizi, EAA, EAA/MARS, CHAID, QUEST, CART, Radyal Tabanlı Fonksiyon ve Çok Tabakalı Algılayıcı yöntemlerini kullanmışlardır ve sınıflandırma sonucunda elde edilen duyarlılık, özgüllük ve doğruluk oranlarına dayanarak yöntemleri performanslarına göre Aşamalı Kümeleme Analizi ile kümelendirmişlerdir. Bunun sonucunda ilgili veri seti için Radyal Tabanlı Fonksiyon ve Çok Tabakalı Algılayıcı yöntemlerinin LR, FDA, FDA/MARS, CHAID, QUEST ve CART'dan daha iyi performans gösterdiğini belirlemişlerdir (33).

Türe ve ark. 2005 yılındaki bir başka çalışmalarında primer hipertansiyonun tahminlenmesinde lojistik regresyon analizi, EAA ve sinir ağıları yöntemlerinin performanslarını karşılaştırmışlardır. Kontrol ve hipertansiyonlu hasta gruplarını ayırmada sinir ağlarının performansının lojistik regresyon ile EAA'dan istatistiksel olarak daha iyi olduğunu belirlemişlerdir (34).

Sonuç olarak, veri setlerindeki gruplar kendi içinde homojen yapıda ve aralarında heterojen yapıda ise DAA, KAA, EAA/MARS degree=1, EAA/MARS degree=2 ve EAA/BRUTO yöntemlerinin daha iyi performans gösterdiği belirlenmiştir. Grup merkezlerinin (ortalama vektörleri) birbirine yakın olduğu durumlarda eğer varyanslar küçük ise yöntemler sınıfları yüksek DSO

değerleriyle birbirlerinden ayırmaktadırlar. Ancak birimlerin özellikleri birbirlerine karıştığında (üst üste binme, overlapping) yine KAA, DAA'ya göre daha etkin DSO değerleri verdiği ve böylece sınıfları ayırma performansının daha iyi olduğu sonucuna varılmıştır. EAA yöntemlerinin ise genellikle farklı koşullarda DAA'ya göre yüksek ancak KAA'dan daha düşük performans değerleri verdiği gözlenmiştir.

Normal dağılım göstermeyen iki ve üç gruplu ($p=4$) verilerin ayırımında ise her beş yöntem birbirlerine yakın ve %70'denb daha büyük DSO değerlerine sahip olarak bulunmuştur. KAA ve DAA'yı bu tür verilerde varsayımları uymadığı için kullanmak doğru olmadığından EAA'nın üç yönteminin de her iki veri setinde ($g=2$, $g=3$) benzer performans vermiş olması yöntemlerin düşük ve yüksek ilişkili verilerde performansları değişmeden etkinliklerini korudukları görülmüştür. Bu yöntemlerinin birbirlerine alternatif yöntemler oldukları sonucuna varılmıştır. Ancak Friedman, Hastie ve ark. yaptıkları çalışmalarda EAA/MARS degree=2 modelinin EAA/MARS degree=1 ve EAA/BRUTO modellerine göre daha etkin sonuçlar verdiğini belirtmişlerdir (11, 13, 14). Friedman 1991 yılındaki çalışmasında EAA/MARS degree=1 modeli için en uygun uygulama koşullarının birim sayısının 50 ile 1000 arasında ve değişken sayısının 3 ile 20 arasında olmasını önermiştir (11).

DAA yöntemi çokdeğişkenli normal dağılım gösteren homojen veri setlerinde kullanıldığından, veri setlerindeki küçük değişimlerden kolay etkilenen bir yöntemdir. Sağlık alanında çokdeğişkenli normal dağılım gösteren homojen veri setlerinin elde edilmesi oldukça zor bir durumdur. Bu şartlar altında, araştırmacılara, çokdeğişkenli normal dağılım gösteren heterojen veri setlerinde KAA analizini kullanmaları, normal olmayan heterojen veri setlerinde ise EAA/MARS degree=1 yöntemini kullanmaları önerilmektedir.

KAYNAKLAR DİZİNİ

1. Akdağ B, Özkurt S, Ankaralı H, Kavas M, Evyapan F, Kiter G: Tanı koymada MARS ve lojistik regresyon modellerinin kullanımı: Karşılaştırmalı bir uygulama, IX. Ulusal Biyoistatistik Kongresi, 25, 2006.
2. Alpar R: Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, Nobel Yayın Dağıtım, Ankara, 2003.
3. Amato U: Independent component discriminant analysis, from http://www.na.iac.cnr.it/rapporti/2002/RT254_02.pdf
4. Anne H and Solberg S: Flexible nonlinear contextual classification, Pattern Recognition Letters 25(13): 1501-1508, 2004.
5. Çakmak Z: Çoklu Ayırma ve Sınıflandırma Analizinin Eğitimde Öğrencilerin Meslek Seçimine Uygulanması, Anadolu Üniversitesi Yayınları, Eskişehir, 1992.
6. Callut J: Flexible discriminant analysis, from http://www.stat.ucl.ac.be/slidelecture/chap12_suite.pdf.
7. Dalgaard P: Introductory Statistics with R, Springer, USA, 2003.
8. Dudoit S, Fridlyand J, Gentleman R: Classification in DNA microarray experiments, from www.bioconductor.org/workshops/2002/WyethCourse101702/Classif/classif.pdf
9. Edler L and Suhai S: Role and result of statistical methods in protein fold class prediction, Mathematical and Computer Modelling 33(12): 1401-1417, 2001.

10. Eubank RL: Spline Smoothing and Nonparametric Regression, New York: Marcel Dekker Inc., 1988.
11. Friedman JH: Multivariate Adaptive Regression Splines, *Annals of Statistics* 19(1):1-141, 1991.
12. Hardle W and Simar L: Applied Multivariate Statistical Analysis, First Ed., Springer-Verlag, USA, 2003.
13. Hastie T, Buja A, Tibshirani R: Penalized discriminant analysis, *Annals of Statistics* 23(1):73-102, 1995.
14. Hastie T, Tibshirani R, Buja A: Flexible discriminant and mixture models, *Proceedings of "Neural Networks and Statistics" conference*, Edinburgh: Oxford University Press, p1-23, 1995.
15. Hastie T, Tibshirani R, Buja A: Flexible discriminant analysis by optimal scoring, *Journal of the American Statistical Association* 89:1255-1270, 1994.
16. Hastie T, Tibshirani R, Friedman J: *Elements of statistical learning: Data mining, prediction, and inference*, Springer, USA, 2001.
17. Hastie T, Tibshirani R: Discriminant analysis by Gaussian mixtures, *J Roy Stat Soc B* 58: 155-76, 1996.
18. Johnson RA and Wichern DW: *Applied Multivariate Statistical Analysis*, Third Ed., Prentice-Hall Inc., USA, 1992.
19. Krzanowski WJ and Marriott FHC: *Multivariate Analysis Part II Classification, Covariance Structures and Repeated Measurements*, Arnold, Volume 2, Kendall's Library of Statistics, 1995.

20. Langaas M: Discrimination and Classification, Technical Report Statistics 1/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim 1995.
21. Maindonald JH and Braun WJ: Data Analysis and Graphics Using R , Cambridge University Press, UK, 2003.
22. McCullagh P and Nelder JA: Generalized Linear Models, Second Ed., Chapman and Hall, London, 1989.
23. Özdamar K: Paket Programlarla İstatistiksel Veri Analizi 2, Beşinci Baskı, Kaan Kitabevi, Eskişehir, 2002.
24. PRESS SJ: Applied Multivariate Analysis, Holt Rinehart and Winston Inc., New York, 1972.
25. R program and packages for windows version 2.3.1, from <http://www.r-project.org>
26. Rencher AC: Methods of Multivariate Analysis, Second edition, John Wiley & Sons Inc., USA, 2002.
27. Reynes C, Sabatier R, Molinari N: Choice of B-splines with free parameters in the flexible discriminant analysis context, Computational Statistics & Data Analysis 24: 2005.
28. Schmitz PIM, Habbema JDF, Hermans J, Raatgever JW: Comparative performance of four discriminant analysis methods for mixtures of continuous and discrete variables, Commun Statist-Simula. Computa. 12(6): 727-751, 1983.
29. Stevens J: Applied Multivariate Statistics For The Social Sciences. Hillsdale, New Jersey USA, 1986.

30. Sugnet G and Niel L: Extensions of biplot methodology to discriminant analysis, *Journal of Classification* 22: 59-86, 2005.
31. Tebaldi C, Nychka D, Brown B, Sharman R: Non-parametric, flexible models for forecasting clear-air turbulence, from <http://www.image.ucar.edu/~tebaldi/talks/gtpgsp.pdf>
32. Temel GO, Çamdeviren H, Yazıcı AC: Regresyon modellerine alternatif bir yaklaşım: MARS, VIII. Ulusal Biyoistatistik Kongresi, 105-123, 2005.
33. Türe M, Kurt İ, Kürüm AT, Özdamar K: Comparing classification techniques for predicting essential hypertension, *Expert Systems with Applications* 29(3): 583-588, 2005.
34. Türe M, Kurt İ, Yavuz E, Kürüm AT: Hipertansiyonun tahmini için çoklu tahmin modellerinin karşılaştırılması (Sinir ağları, lojistik regresyon ve esnek ayırma analizleri), *Anadolu Kardiyoloji Dergisi* 5: 24-28, 2005.
35. Yeomans KA: Multivariate Classification: Data reduction using component and cluster analysis,-Birmingham Wards, The University of Aston Management Centre, Working Paper Series 145: 1, 1979.

ÖZGEÇMİŞ

Ahmet Öztürk 1967 yılında Kayseri'nin Sarız ilçesinde doğdu. İlk ve Orta Okul ve Lise öğrenimini Kayseri-Bünyan ilçesinde bitirdi. Lisans eğitimini 1986 yılında girdiği Anadolu Üniversitesi İşletme Fakültesini 1991 yılında bitirerek tamamladı.

1996 yılında Kayseri Erciyes Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Bilim dalına Yüksek Lisans Öğrencisi olarak girdi. 2000 yılında Biyoistatistik Yüksek Lisansını tamamladı ve Biyoistatistik Bilim Uzmanı ünvanını aldı. 2002 Yılında Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim dalında doktora eğitime başladı. Ders ve Yeterlilik Sınavlarını 2004 yılında tamamlayan Ahmet Öztürk doktora aşamalarını tamamlamaya devam etmektedir.

Ahmet Öztürk 2005 yılında Kayseri Erciyes Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim dalına Biyoistatistik Uzmanı olarak atandı.

Ahmet Öztürk evli ve 3 çocuk babasıdır.