T.C.
ÇUKUROVA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

TANI TESTLERİNİN DEĞERLENDİRİLMESİNDE KULLANILAN STANDARTLAR VE ANALİTİK YÖNTEMLER

İLKER ÜNAL

DOKTORA TEZİ

DANIŞMANI
Prof. Dr. Refik BURGUT

ADANA – 2010

T.C.
ÇUKUROVA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

TANI TESTLERİNİN DEĞERLENDİRİLMESİNDE KULLANILAN STANDARTLAR VE ANALİTİK YÖNTEMLER

İLKER ÜNAL

DOKTORA TEZİ

DANIŞMANI
Prof. Dr. Refik BURGUT

Bu tez Çukurova Üniversitesi Bilimsel Araştırma Proje Birimi tarafından TF2007D10 nolu proje olarak desteklenmiştir.

ADANA – 2010

KABUL VE ONAY

Biyoistatistik Anabilim Dalı Doktora Programı Çerçevesinde yürütülmüş olan

"Tanı Testlerinin Değerlendirilmesinde Kullanılan Standartlar Ve Analitik Yöntemler"

adlı çalışma, aşağıdaki jüri tarafından Doktora Tezi olarak kabul edilmiştir.

Tarihi: 04/07/2011

TEZ SINAV JÜRİSİ

Prof Dr. H.Refik Burgut
Çukurova Üniversitesi
Başkan

Prof Dr. Z.Nazan Alparslan
Çukurova Üniversitesi
Üye

Prof Dr. Hamza Erol
Çukurova Üniversitesi
Üye

Prof Dr. Ergün Karaağaoğlu
Hacettepe Üniversitesi
Üye

Prof Dr. Nafiz Bozdemir
Çukurova Üniversitesi
Üye

Yukarıdaki Tez, Yönetim Kurulunun ……/……./…….. tarih ve ……. sayılı kararı ile kabul edilmiştir.

Prof. Dr. Halil Kasap
Sağlık Bilimleri Enstitü Müdürü

TEŞEKKÜR

Tezimin planlanması ve yürütülmesinde yardımlarını esirgemeyen tez danışmanım Prof. Dr. Refik BURGUT'a, desteğini esirgemeyen Prof. Dr. Nazan ALPARSLAN'a ve Prof. Dr. Hazma EROL'a; tezde kullanan verileri sağlayan Dr. Aysun SUKAN ve ekip arkadaşları, Dr Fatin KOÇAK ve ekip arkadaşları, Dr Suhan GÜNAŞTI ve ekip arkadaşları, Dr Mevlüt KOÇ ve ekip arkadaşlarına; tezin okunması ve hataların düzeltilmesinde emeği geçen sevgili eşim Öğr. Gör. Esin ÜNAL'a teşekkürü bir borç bilirim.

İlker ÜNAL

İÇİNDEKİLER

KABUL VE ONAY **II**

TEŞEKKÜR **III**

İÇİNDEKİLER **IV**

ŞEKİLLER DİZİNİ **VI**

ÇİZELGELER DİZİNİ **VIII**

SİMGELER VE KISALTMALAR DİZİNİ **X**

ÖZET **XI**

ABSTRACT **XII**

1 GİRİŞ **1**

2 GENEL BİLGİLER **6**

2.1 TANI TESTLERİNİN DEĞERLENDİRİLME SÜRECİ 6

2.2 TANI TESTLERİNİN DEĞERLENDİRİLMESİNDE KULLANILAN ÖLÇÜTLER 9

2.2.1 Niteliksel Ölçümlü Testlerde Doğruluk Ölçütleri 9

2.2.2 Niceliksel Ölçümlü Testlerde Doğruluk Ölçütleri 12

2.3 İKİ TESTİN DOĞRULUKLARININ KARŞILAŞTIRILMASI 15

2.3.1 İki Değerli Testlerin Doğruluklarının Karşılaştırılması 15

2.3.2 Sürekli Değerler Alan Testlerin Doğruluklarının Karşılaştırılması 17

2.4 DOĞRULAMA YANLILIĞI 21

2.4.1 Doğrulama Yanlılığı Nedir? 21

2.4.2 Begg ve Greenes Yaklaşımı ile Yanlılığın Düzeltilmesi 23

2.4.3 Zhou Alt Sınır – Üst Sınır Yaklaşımı ile Yanlılığın Düzeltilmesi 25

2.4.4 Lojistik Regresyon Yaklaşımı ile Yanlılığın Düzeltilmesi 26

2.4.5 Çoklu İmputasyon Yaklaşımı ile Yanlılığın Düzeltilmesi 28

2.4.6 Yapay Sinir Ağları Yaklaşımı ile Yanlılığın Düzeltilmesi 30

2.5 KESİN OLMAYAN ALTIN STANDART (KOAS) YANLILIĞI 36

2.5.1 KOAS yanlılığı nedir? 36

2.5.2 Tanı Testinin İkili (Binary) Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Önerilen Yaklaşımlar 37

2.5.3 Tanı Testinin Sıralı Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Parametrik Olmayan ROC Eğrisi Yaklaşımı 39

2.5.4 Tanı Testinin Sürekli Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Bayes Yaklaşımı 42

2.5.5 Tanı Testinin İkili, Sıralı veya Sürekli Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Önerilen Alternatif Yaklaşım 45

3 GEREÇ VE YÖNTEM 48

3.1 DOĞRULAMA YANLILIĞININ DÜZELTİLMESİ 48

3.2 KESİN OLMAYAN ALTIN STANDART YANLILIĞIN DÜZELTİLMESİ 51

4 BULGULAR 56

4.1 DOĞRULAMA YANLILIĞINI DÜZELTMEDE KULLANILAN VE ÖNERİLEN YAKLAŞIMLARIN KARŞILAŞTIRILMASI 56

4.1.1 Eksik ölçüm mekanizması rastgele kabul edildiğine (MAR) 56

4.1.2 Eksik ölçüm mekanizması rastgele kabul edilmediğine (NMAR) 74

4.1.3 Gerçek Veri Seti Kullanılarak Yapılmış Çalışma Sonuçları 91

4.2 KESİN OLMAYAN ALTIN STANDART YANLILIĞININ DÜZELTİLMESİNDE KULLANILAN VE ÖNERİLEN YAKLAŞIMLARIN KARŞILAŞTIRILMASI 93

4.2.1 İlgilenilen Testin İkili (Binary) Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları 93

4.2.2 İlgilenilen Testin Sıralı (Ordinal) Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları 107

4.2.3 İlgilenilen Testin Sürekli Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları 115

5 TARTIŞMA 132

6 SONUÇ VE ÖNERİLER 143

7 KAYNAKLAR 144

ÖZGEÇMİŞ 148

ŞEKİLLER DİZİNİ

Şekil 1.1 Tanı koymada etkili faktörler 1
Şekil 2.1 ROC Eğrisi 13
Şekil 2.2 Binormal dağılım eğrileri ve kesim noktasına göre DPO ve YPO değerleri 19
Şekil 2.3 Yapay Sinir Ağlarının işleyiş şeması 32
Şekil 2.4 İleri beslemeli gizli katmanlı yapay sinir ağı 34
Şekil 4.1 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti doğruluk ölçütleri 58
Şekil 4.2 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti doğruluk ölçütleri 60
Şekil 4.3 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti doğruluk ölçütleri 61
Şekil 4.4 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti doğruluk ölçütleri 64
Şekil 4.5 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti doğruluk ölçütleri 66
Şekil 4.6 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti doğruluk ölçütleri 68
Şekil 4.7 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti doğruluk ölçütleri 71
Şekil 4.8 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti doğruluk ölçütleri 73
Şekil 4.9 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti doğruluk ölçütleri 76
Şekil 4.10 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti doğruluk ölçütleri 78
Şekil 4.11 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti doğruluk ölçütleri 79
Şekil 4.12 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti doğruluk ölçütleri 82
Şekil 4.13 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti doğruluk ölçütleri 85
Şekil 4.14 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti doğruluk ölçütleri 87
Şekil 4.15 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti doğruluk ölçütleri 89
Şekil 4.16 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti doğruluk ölçütleri 91
Şekil 4.17 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Prevelans = 0.5 95
Şekil 4.18 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Prevelans = 0.8 96
Şekil 4.19 Prevelans değişiminin yaklaşımların Duyarlılık nokta tahminine olan etkisi – n=100 99
Şekil 4.20 Prevelans değişiminin yaklaşımların Seçicilik nokta tahminine olan etkisi – n=100 100
Şekil 4.21 Prevelans değişiminin yaklaşımların Duyarlılık nokta tahminine olan etkisi – n=1000 101
Şekil 4.22 Prevelans değişiminin yaklaşımların Seçicilik nokta tahminine olan etkisi – n=1000 102
Şekil 4.23 Doğruluk ölçütlerindeki değişimin yaklaşımların Duyarlılık nokta tahminine olan etkisi 104
Şekil 4.24 Doğruluk ölçütlerindeki değişimin yaklaşımların Seçicilik nokta tahminine olan etkisi 105
Şekil 4.25 Büyük veri setinde prevelans değişiminin yaklaşımlara olan etkisi 108
Şekil 4.26 Küçük veri setinde prevelans değişiminin yaklaşımlara olan etkisi 109
Şekil 4.27 Büyük veri setinde Test ve Kategori sayısındaki değişimin yaklaşımlara olan etkisi 111
Şekil 4.28 Küçük veri setinde Test ve Kategori sayısındaki değişimin yaklaşımlara olan etkisi 112
Şekil 4.29 Büyük veri setinde Eğri altında kalan alan değerinin değişiminin yaklaşımlara olan etkisi 113

Şekil 4.30 Küçük veri setinde Eğri altında kalan alan değerinin değişiminin yaklaşımlara olan etkisi 114
Şekil 4.31 İterasyonlara göre Bayes yaklaşımı parametrelerinin tahmin edicileri 116
Şekil 4.32 Büyük veri seti ve prevelans 0.5 olduğu durumda Bayes yaklaşımının ROC eğrisi altında kalan alan tahmini (üretilmiş 1 veri seti için) 117
Şekil 4.33 Büyük veri setinde Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 120
Şekil 4.34 Küçük veri setinde Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 121
Şekil 4.35 EAKA değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 123
Şekil 4.36 Düşük açıklama oranında Bayes yaklaşımının iterasyonlar sonucunda elde ettiği EAKA değeri (üretilmiş 1 veri seti için) 124
Şekil 4.37 Açıklama oranındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 126
Şekil 4.38 Karıştırıcı etkili veri setinde Bayes yaklaşımının iterasyonlar sonucunda elde ettiği EAKA değeri (üretilmiş 1 veri seti için) 127
Şekil 4.39 Karıştırıcı etkili veri setinin yaklaşımların EAKA nokta tahminine olan etkisi 129

ÇİZELGELER DİZİNİ

Çizelge 2.1 Referans test ile Tanı testi karşılaştırılması 10
Çizelge 2.2 İki tanı testinin karşılaştırılması 16
Çizelge 2.3 İki değerli bir test için gözlenen veri 24
Çizelge 2.4 Kosinski ve Barnhart Yaklaşımına göre düzenlenmiş veri seti 35
Çizelge 2.5 Tutarsızlık Çözücü yaklaşımı 38
Çizelge 3.1 Üretilen veride iki değişkenli normal dağılım gösteren değişkenlerin dağılım bilgileri 49
Çizelge 3.2 KOAS yanlılığını düzeltmede kullanılan yaklaşımları karşılaştırmak için üretilen verilerdeki çapraz ilişkiler 55
Çizelge 4.1 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 57
Çizelge 4.2 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 59
Çizelge 4.3 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 62
Çizelge 4.4 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 63
Çizelge 4.5 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 67
Çizelge 4.6 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 69
Çizelge 4.7 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 70
Çizelge 4.8 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 72
Çizelge 4.9 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 75
Çizelge 4.10 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 77
Çizelge 4.11 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 80
Çizelge 4.12 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 81
Çizelge 4.13 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 84

Çizelge 4.14 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 86
Çizelge 4.15 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 88
Çizelge 4.16 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$) 90
Çizelge 4.17 Gerçek veri seti kullanılarak yapılan karşılaştırma sonuçları 92
Çizelge 4.18 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 94
Çizelge 4.19 Küçük veri setinde Prevelans değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 97
Çizelge 4.20 Büyük veri setinde Prevelans değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 98
Çizelge 4.21 Doğruluk ölçütü değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 103
Çizelge 4.22 Gerçek veri seti kullanılarak yapılan karşılaştırma sonuçları 106
Çizelge 4.23 Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 107
Çizelge 4.24 Test ve kategori sayısındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 110
Çizelge 4.25 Eğri altında kalan alan değerinin değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı($x10^{-3}$) 113
Çizelge 4.26 Gerçek veri seti kullanılarak elde edilen yaklaşımların ROC eğri alanları ve standart sapma değerleri 115
Çizelge 4.27 Büyük veri seti ve prevelans 0.5 olduğu durumda Gizli sınıf modelleri uyum istatistikleri (üretilmiş 1 veri seti için) 118
Çizelge 4.28 Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 119
Çizelge 4.29 EAKA değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 122
Çizelge 4.30 Açıklama oranındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi 125
Çizelge 4.31 Karıştırıcı etkili veri setinin yaklaşımların EAKA nokta tahminine olan etkisi 128
Çizelge 4.32 Gerçek veri seti kullanılarak elde edilen yaklaşımların ROC eğri alanları ve standart sapma değerleri 130
Çizelge 4.33 Gerçek veri seti kullanılarak elde edilen yaklaşımların sınıflama hataları 131

SİMGELER VE KISALTMALAR DİZİNİ

KOAS	:	Kesin Olmayan Altın Standart
ROC	:	Reciever Operating Characteristic – İşlem karakteristiği
Duy	:	Duyarlılık
Seç	:	Seçicilik
YPO	:	Yanlış Pozitif Oranı
YNO	:	Yanlış Negatif Oranı
LR	:	Likelihood Ratio – Olabilirlilik Oranı
H_0	:	Sıfır Hipotezi
H_A	:	Alternatif Hipotez
Var	:	Varyans
Cov	:	Kovaryans
MLE	:	Maximum Likelihood Estimator – Maksimum olabilirlilik tahmin edicisi
MAR	:	Missing At Random – Rastgele eksik ölçümler
NMAR	:	Not Missing At Random – Rastgele olmayan eksik ölçümler
ML	:	Maximum Likelihood – Maksimum olabilirlilik
$L_{göz}$	:	Gözlenen Likelihood – Gözlenen olabilirlilik
EM	:	Expectation Maximization – Beklenti Maksimizasyon
YSA	:	Yapay Sinir Ağları
LR	:	Lojistik Regresyon
Çİ	:	Çoklu İmputasyon
LCR	:	Ligase Chain Reaction – Ligaz zincir reaksiyonu
PCR	:	Polymerase Chain Reaction – Polimeraz zincir reaksiyonu
DPO	:	Doğru Pozitif Oranı
MCMC	:	Monte Carlo Markov Chain
GSA	:	Gizli Sınıf Analizi
MICE	:	Multivariate Imputation by Chained Equations
EAKA	:	Eğri Altında Kalan Alan
HP	:	Helicobacter Pylori
HpSA	:	Helicobacter Pylori Stool Antigen Test
METS	:	Metabolic Equivalents
LL	:	LogLikelihood – Logaritmik olabilirlilik
BIC	:	Bayesian Information Criteria – Bayes bilgi kriteri
AUROC	:	Area Under ROC Curve – ROC eğrisi altında kalan alan

ÖZET

Tanı Testlerinin Değerlendirilmesinde Kullanılan Standartlar ve Analitik Yöntemler

Bir tanı testinin doğruluk ölçütlerinin değerlendirilmesinde yansız tahmin ediciler tercih edilir. Ancak, testin doğruluk ölçütlerinin belirlenmesinde yanlılıklar ile karşılaşılabilir. Bu yanlılıklardan sıklıkla görülenleri; Doğrulama yanlılığı ve Kesin olmayan altın standart yanlılığıdır. Bu tezde bu yanlılıkların düzeltilmesinde kullanılan yaklaşımlar incelenmiş ve bu yaklaşımların sorun yaşadığı durumlarda alternatif olabilecek yaklaşımlar önerilmiştir.

Doğrulama yanlılığı çalışmaya alınan tüm bireylere altın standart testin uygulanmaması sonucunda ortaya çıkar. Tanı testinin ikili ölçüm olması durumunda yanlılığın düzeltilmesinde; altın standart uygulanmayan bireylerin sonuçları eksik ölçüm olarak kabul edilirse, eksik ölçüm mekanizmasına göre, Begg&Greenes, Zhou sınırları, Lojistik Regresyon, Çoklu İmputasyon gibi yaklaşımlar kullanılır. Bu tezde bu yaklaşımlara Yapay Sinir Ağları yaklaşımı da alternatif olarak eklenmiş ve yaklaşımlar çeşitli karşılaştırma başlıkları altında üretilmiş ve gerçek veri setleri kullanılarak karşılaştırılmıştır.

Kesin olmayan altın standart yanlılığı ise, altın standart olarak kabul edilen bir testin olmaması durumunda bu test yerine daha düşük sınıflama başarısı olan bir testin, kesin olmayan altın standart testinin, kullanılması ile ortaya çıkar. Yanlılığın düzeltilmesinde tanı testinin ölçüm ölçeğine göre Tutarsızlık Çözücü, Bileşke Referans Standardı, Zhou'nun Parametrik olmayan ROC eğrisi tahmini, Bayes yaklaşımı gibi yaklaşımlar kullanılır. Bu tezde bu yaklaşımlar ile bunlara alternatif olarak önerilen Gizli Sınıf Analizi yaklaşımı, üretilmiş ve gerçek veri setleri kullanılarak karşılaştırılmıştır.

Üretilmiş veri seti ile yapılan tüm karşılaştırmalarda yaklaşımların sonuçlarını etkileyebilecek: değişen veri seti büyüklüğü, değişen hastalık prevelansı, değişen doğruluk ölçütü gibi durumlar göz önüne alınmıştır.

Doğrulama yanlılığını düzeltmede; kesin tanı doğrulaması yapılmamış bireylerin rastgele seçilmiş olması durumunda Begg&Greenes yaklaşımı, rastgele olmaması durumunda ise Lojistik Regresyon ve Yapay Sinir Ağları yaklaşımlarının uygun yaklaşımlar olabileceği görülmüştür.

Kesin olmayan altın standart yanlılığını düzeltmede; tanı testinin ikili ölçüm olması durumunda uygun prevelans aralığı ve yüksek doğruluk ölçütünde Gizli Sınıf Analizi; tanı testinin sıralı ölçüm olması durumunda yine uygun prevelans aralığı, yeterli sayıda kategori ve test varlığında Gizli Sınıf Analizi; tanı testinin sürekli ölçüm olması halinde ise küçük örneklem, düşük eğri altında kalan alan ve karıştırıcı etkili değişken varlığı dışındaki durumlarda Bayes yaklaşımı, bu yaklaşımın sorun yaşadığı durumlarda ise Gizli Sınıf Analizi önerilebilir yaklaşımlardır.

Anahtar Kelimeler: Tanı Testleri, Doğruluk Ölçütleri, ROC Eğrisi, Eğri Altında Kalan Alan, Gizli Sınıf Modelleri

ABSTRACT

Standards and Analytic Techniques Used in Evaluation of Diagnostic Tests

In the evaluation of accuracy for a diagnostic test, unbiased estimators for accuracy measures are preferred. However, some bias in measuring diagnostic accuracy of the test may occur. The most common biases among encountered are verification bias and imperfect gold standard bias. In this thesis, proposed approaches for correcting these biases were examined and alternative approaches were proposed for the situations involving problems for the current approaches.

Verification bias arises when not all patients in the study are taken the gold standard test. When the diagnostic test is binary, depending on the missing mechanism, Begg&Greenes, Upper and Lower Bound by Zhou, Logistic Regression or Multiple Imputation approaches are used for correcting this bias. In this thesis, a new approach, called Artificial Neural Networks, was alternatively proposed and compared with other approaches in different comparison schemes using simulated and real data sets.

Imperfect gold standard bias arises when using an imperfect gold standard test instead of gold standard test. Depending on the measurement scale of new diagnostic test, Discrepant Resolution, Composite Reference Standard, Non Parametric AUROC Estimation by Zhou and Bayesian Approaches are used for correcting this bias. In this thesis, a new approach, called Latent Class Analysis, was alternatively proposed and compared with other approaches in different comparison schemes using simulated and real data sets.

In the simulations, in order to measure the correct performance of the approaches and to determine deficiencies involving, data were generated with a variety of experimental conditions; changing sample size, disease prevalence, and diagnostic accuracy of new test.

When the missing mechanism is MAR, Begg&Greenes approach and when the missing mechanism is NMAR, Logistic Regression or Artificial Neural Networks approaches are preferable for correcting verification bias.

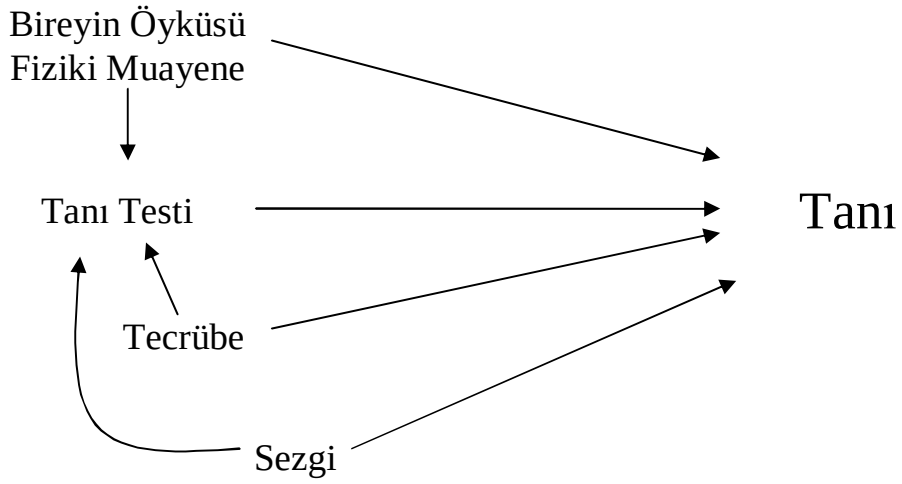
When the measurement scale of the new test is binary and disease prevalence and diagnostic accuracy values are well-defined, Latent Class Analysis approach; when the measurement scale is ordinal and disease prevalence and the number of test and test categories are appropriate, Latent Class Analysis approach; and finally when the measurement scale is interval or ratio and with large sample, high AUROC and without confounding variable, Bayesian Approach and for other situations Latent Class Analysis approach are the best approaches for correcting imperfect gold standard bias.

Key Words: Diagnostic Test, Diagnostic Accuracy, ROC Curve, Area Under the Curve (AUC), Latent Class Models

1 Giriş

Bir bireye hasta ya da sağlıklı tanısını koymada birçok etken rol oynar. Bireyin öyküsü, fiziki muayene bulguları, risk faktörlerinin varlığı ve tabi ki tanı testinin sonucu gibi bilgilere sahip bir doktor, kendi bilgi ve tecrübesini de ekleyerek birey hakkında karara varır. Yukarıda bahsedilen bilgilerin bazıları birey hakkındaki kararı direk etkiler. Doğruluğu kesin olarak kanıtlanmış bir referans testi (altın standart test) sonucu veya doğruluğu çok yüksek olan bir tanı testi sonucu diğer bilgilerin de önüne geçerek kararı tek başına belirleyebilir. Ancak tanı testinin doğruluğu yüksek değilse ve kullanılmasıyla ilgili şüpheler varsa testten elde edilen bilgi sadece fikir sahibi olmak için kullanılabilir ve kararı belirlemede etkisi daha az olacaktır. Tanı koymada etkili olan faktörleri Şekil 1.1'de görebilirsiniz.

Tanı testleri sağlık alanında önemli bir rol oynamaktadır. Şikayeti nedeniyle sağlık kuruluşuna başvuran kişinin gerçek durumunun güvenilir bir test sonucu saptanması ve buna göre gerekiyorsa uygun tedaviye başlanması veya taburcu edilmesi son derece önemlidir. Tanı testinin sonucu, kişi ile ilgili kararı etkilediğinden testin doğruluğunun yüksek ve güvenilir olması gereklidir. Bu sebeple testin doğruluğunun ve güvenirliğinin ölçütleri belirlenmeli ve kullanılır veya kullanılamaz olduğu gösterilmelidir.



Şekil 1.1 Tanı koymada etkili faktörler

Bazı durumlarda tanı koymada etkili olan faktörler birey hakkındaki kararı belirlese de gerçek durum ile ilgili şüpheler olabilecektir. Bu gibi durumlarda şüphenin artmasına veya azalmasına neden olan tanı testidir. Doğruluğu yüksek bir tanı testi sonucu şüpheleri azaltırken, verdiği sonuçtan emin olunmayan bir test kullanıldığında şüpheler artacaktır.

Sox ve ark. göre tanı testlerinin iki amacı vardır. Birincisi hastanın durumu hakkında güvenilir bir bilgiye sahip olmak ve ikincisi de hastaya yapılacak müdahalelerin planını belirlemek1. McNeil ve Adelstein2 bu iki amaca üçüncü bir amacı şöyle eklemişlerdir: hastalığın mekanizmasını ve gelişim durumunu anlamak (örneğin, kronik hastalarda testleri tekrar tekrar uygulamak). Eğer bir test konunun uzmanı tarafından anlaşılıyor ve açıklanabiliyorsa, bu üç amaca hizmet etmiş olur.

Tanı testinin doğruluğu iki farklı sağlık durumunu sınıflayıp sınıflayamaması ile belirlenir. Tüm hasta ve sağlıklılara tanı testi uygulanır. Eğer ayrım yapılamaz veya hata oranı çok yüksek olursa testin doğruluğu düşük demektir. Eğer tamamen hasta ve sağlıklı diye ayırabiliyorsa testin doğruluğu mükemmeldir. Ancak bir tanı testi bu iki uç durumun arasında kalmaktadır. Bu durumda tanı testinin kişi için vermiş olduğu sonuç şüpheli kalmaktadır. Yani birçok tanı testi uzmana fikir vermekte ancak kişinin gerçek durumu hakkında kesin karar vermesini sağlayamamaktadır. Eğer test sonucu negatif çıkarsa uzman kişiyi sağlıklı olarak görüp eve mi göndermelidir? Ya da test sonucu pozitif çıkarsa, kişi gerçekten hasta kabul edilip, tedavisine başlanmalı mıdır? Ayrıca test sonucunun belirgin olmadığı ve uzman tarafından okunması gerektiği durumlarda (radyologlar gibi) uzmandan uzmana test sonucu değişirse ne yapılmalıdır?

Bu gibi kritik soruları cevaplayabilmek için uzmanın tanı testi hakkında bazı bilgilere sahip olması gerekmektedir. Testin duyarlılığı, gerçek hastaların yüzde kaçını doğru olarak saptayabildiği, testin seçiciliği, gerçek sağlıklıların yüzde kaçını doğru olarak saptayabildiği, gibi testin geçerliliği için temel olan bilgilerin bilinmesi gerekir.

Testin performansını belirlemede kullanılan doğruluk ölçütleri (diagnostic accuracy) bireye ait gerçek durum bilgisi varlığında tespit edilir. Gerçek durum bilgisi altın standart test olarak kabul edilen referans test sonucunda elde edilir. Referans testi bazı durumlarda çalışmadaki her bireye uygulanamayabilir. Bu durumda eldeki veride hastalık bilgisi doğrulananlar, yani referans test sonucu olanlar ve hastalık bilgisi doğrulanmayanlar, yani referans test sonucu olmayan hastalar hakkında sonuçlar

olacaktır. Böyle bir veri setinde, önerilen yeni bir tanı testinin performansı ölçüldüğünde elde edilen sonuçlar veri setindeki kimi hastaların gerçek durum bilgisi olmadığından yanlı olacaktır. Bu yanlılığa doğrulama yanlılığı veya seçme yanlılığı (verification bias, selection bias, work-up bias) denir3.

Doğrulama yanlılığının düzeltilmesi için literatürde bazı yaklaşımlar önerilmiştir. Bu yaklaşımlar referans test uygulanmayan bireylerin seçiminin rastgele olup olmamasına göre değişim göstermektedir. Bu bireylerin seçiminin rastgele olduğu durumda Begg ve Greenes3 tarafından önerilen düzeltme kullanılmaktadır. Rastgeleliliğin olmadığı durumda ise birden fazla yaklaşım önerilmektedir. Zhou 1993 yılında doğruluk ölçütleri için alt ve üst sınırlar saptanabildiğini göstermiştir4. Kosinski ve Barnhart5 bu yanlılığın düzeltilmesi için doğruluk ölçütlerinin tahmininde bir olabilirlilik (likelihood) yaklaşımı önermiştir. Harel ve Zhou6 ise 2006 yılında doğrulama yanlılığını düzeltmede Çoklu imputasyon (Multiple Imputation-Çİ) yaklaşımını kullanmıştır. Son olarak 2006 yılında Martinez ve ark. tarafından ve 2008 yılında Buzoianu ve Kadane tarafından yanlılığın düzeltilmesinde Bayes yaklaşımı önerilmiştir7,8.

Literatürde önerilen yaklaşımların avantaj ve dezavantajlarının belirlenmesi amacıyla karşılaştırıldığı görülmüştür. Harel ve Zhou'nun6 Çoklu imputasyonu önerdiği çalışmada yazarlar 5 farklı imputasyon metodunu Begg ve Greenes düzeltmesi ile karşılaştırmıştır. Yazarlar çalışma sonucu olarak Çoklu imputasyon yaklaşımının Begg ve Greenes düzeltmesinden daha iyi sonuç verdiğini belirtmiştir6. Ancak bu yorum Hanley ve ark.9 tarafından Harel ve Zhou'nun karşılaştırdıkları yaklaşımların sonuçlarının elde edilişinde yapılan bir hatadan dolayı eleştirilmiştir. Daha sonra 2008 yılında de Groot ve ark. Harel ve Zhou'nun yapmış oldukları çalışmayı aynen tekrarlamış ve Harel ve Zhou'nun hatalarını göstermiştir[10]. Bu sonuçlar bu konu hakkında daha çok çalışma yapılmasının gerekliliğini ortaya koymaktadır.

Tanı testlerinin değerlendirilmesinde karşılanan bir diğer sorun da gerçek durum bilgisinin belirlenmesinde altın standart testinin kullanıl(a)mamasıdır. Birçok hastalık için, her durumda kesin tanıyı belirlemek çok zor (veya imkansız) olabilmektedir. Bazı durumlarda kesin bir altın standart testi olamadığı gibi, altın standart testinin olduğu ama uygulamasının çok pahalı ve/veya çok zor olduğu durumlarla da karşılaşılabilinmektedir. Örneğin, kalp krizi kesin tanısını koymak için bireyin

hastaneye yatırılması ve uygulaması hasta için rahatsızlık verici olabilen anjiyo-grafi çekilmesi gerekmektedir. Kesin tanı konmasının hasta hayatta iken mümkün olmadığı bir örnek ise Alzheimer hastalığıdır. Bu hastalığın kesin tanısı hasta öldükten sonra beyninden alınacak parçanın nöropatolojik olarak incelenmesi sonucunda sağlanacaktır. Bazı durumlarda ise kesin tanı için uygulanacak testin sonuç vermesi uzun sürebilmektedir (örneğin kültür üreme testleri). Sonuç olarak bunlar gibi birçok durumda kesin tanıyı belirlemek yerine, belirli bir hata oranına sahip Kesin Olmayan Altın Standart (imperfect gold standard) testler tercih edilmektedir. İşte bu noktada, eğer yeni bir tanı testi önerilir ve bunun doğruluktaki başarısı ölçülmek istenirse, Kesin Olmayan Altın Standart (KOAS) testinden kaynaklanan hatadan dolayı yeni tanı testinin başarısı olduğundan daha düşük veya yüksek çıkabilecektir. Bu yanlılığa kesin olmayan altın standart yanlılığı denir[11].

KOAS yanlılığının düzeltilmesinde tanı testinin yapısına göre (kategorik veya sürekli) farklı düzeltmeler önerilmiştir. Henkelman ve ark. tarafından 1990 yılında kesin olmayan altın standart varlığında sürekli ölçümlü yeni tanı testi için bir düzeltme yayınlamıştır. Bu düzeltmede çok değişkenli normal dağılımlı karma model kullanılarak ROC eğrisinin maksimum olabilirlilik tahmini elde edilmektedir[12]. 1996 yılında Hadgu KOAS ve yeni tanı testi arasında tutarsızlık olduğu durumlarda aradaki tutarsızlığı çözmek için Çözücü test kullanılabileceğini belirterek Tutarsızlık Çözücü yaklaşımı ile KOAS yanlılığını düzeltmiştir[13]. Hui ve Zhou 1998 yılında yayınladıkları makalede daha önceden yayınlanmış ve kesin olmayan altın standart testi kullanıldığı durumlarda yeni tanı testi olarak önerilen bir veya birden fazla testin doğruluk ölçütleri için düzeltmeleri incelemiştir. Burada incelenen birçok düzeltmenin, önerilen yeni tanı testinin (veya testlerinin) iki değer alan (ikili-binary) testler olması durumunda geçerli olduğu gösterilmiştir[14]. Zhou ve ark. tarafından 2002 yılında yayınlanan kitapta kesin olmayan altın standart varlığında kategorik tanı testlerinin doğruluk ölçütleri için düzeltmeler verilmiştir. Bu düzeltmeler bir veya birden fazla tanı testi için olduğu gibi bir veya birden fazla popülasyon yapılmış testler için de verilmiştir[11]. Hall ve Zhou tarafından 2003 yılında ve yine Zhou ve ark. tarafından 2005 yılında yayınlanan makalelerde ise koşullu bağımsızlık varsayımı altında ikiden fazla sıralı (ordinal) yapıdaki test için parametrik olmayan ROC eğrisi elde ederek kesin olmayan altın standart yanlılığı düzeltilmiştir[15,16]. Konuya yeni bir yaklaşım olan Bayes yaklaşımı ise

2006 yılında Wang ve ark. tarafından önerilmiştir. Bu makalede ek değişkenler ve iterasyon için başlangıç bilgisi kullanılarak ROC eğrisi tahmin edilmeye çalışılmıştır[17]. Bu çalışmalar bu konudaki arayışın hala sürdüğünü göstermektedir.

Tanı testlerinin sınıflama başarıları bu başarıyı ölçmede kullanılan gerçek durum bilgisinin belirlenmesinden kaynaklanan sorunlar nedeniyle yanlı olabilmektedir. Bu tezde çalışmalarda sıklıkla görülen doğrulama yanlılığı ve KOAS yanlılığının düzeltilmesi konuları incelenmiş, bu yanlılıkların düzeltilmesinde kullanılabilecek iki yeni yaklaşım önerilmiş ve bu yaklaşımlar önerilen önceki yaklaşımlarla karşılaştırılmıştır. Doğrulama yanlılığının düzeltilmesinde Yapay Sinir Ağları (Artificial Neural Networks) yaklaşımı önerilmiştir. Bu yaklaşım, referans testi uygulanmayan bireylerin seçiminin rastgele olmaması durumunda ikili (binary) yeni bir tanı testinin performansının belirlenmesinde kullanılmak amacıyla önerilmiştir. Bu yaklaşım literatürde önerilen Begg ve Greenes düzeltmesi, Zhou alt ve üst sınırları, Lojistik Regresyon yaklaşımı ve Çoklu imputasyon yaklaşımlarıyla, yaklaşımların sonuçlarını etkileyebileceği düşünülen farklı veri setleri üretilerek ve Nükleer Tıp alanında yapılmış bir çalışmadan elde edilen bir veri seti kullanılarak karşılaştırılmıştır.

KOAS yanlılığının düzeltilmesinde ise alternatif yaklaşım olarak Gizli Sınıf Analizi önerilmiştir. Bu yaklaşım tanı testinin isimsel ölçekte olduğu durumda önerilen Tutarsızlık Çözücü ve Bileşke Referans Standartlarına, sıralı yapıdaki testler için elde edilen ve Zhou tarafından önerilen ROC eğrisi alan tahmini yaklaşımına ve de Wang ve ark. tarafından önerilen Bayes yaklaşımına alternatif olarak önerilmiştir. Bu tezde önerilen yaklaşım diğer önerilen yaklaşımlarla üretilmiş veri setleri ve Gastroenteroloji, Dermatoloji ve Kardiyoloji alanlarında yapılmış bir çalışmadan elde edilmiş gerçek veri setleri kullanarak karşılaştırılmıştır.

2 Genel Bilgiler

2.1 Tanı testlerinin değerlendirilme süreci

Uygulamada önerilen tanı testlerinin sınıflama başarısını değerlendirmede gereken önemin verilmediği ve testin derhal rutin uygulamaya alındığı birçok yazar tarafından belirtilmiştir[18,19,20]. Ayrıca bazı çalışmalarda testlerin sınıflama başarısını belirlemek için tasarlanan çalışmaların hatalı olduğu ve bu da testin doğruluk ölçütlerinin yanlı tahmin edilmesine sebep olduğu belirtilmiştir[18,21]. Bu sebeple birçok yazar tanı testlerinin değerlendirilmesinde standartların kullanılması gerektiğini vurgulamıştır[18,21,22].

Tanı testlerinin değerlendirilmesi ile ilgili yapılan birçok genel hata vardır. Bunlardan biri tanı testinin açıklanması ile ilgilidir. Birçok araştırıcı yeni bir tanı testi geliştirirken, sadece sağlıklı gönüllüleri göz önüne almaktadır. Örneğin bir hastalığın tanısını koymada kullanılacak olan bir testte bir kan değeri sağlıklı gönüllüler üzerinde ölçülüyor. Elde edilen değerlere göre bir kesim noktası –örneğin, ortalamanın 3 standart sapma altında olanlar– belirleniyor. Buna göre gönüllüler içinde ortalama değerden 3 standart sapma altında değere sahip olanlar hasta ve üzerinde olanlar ise sağlıklı olarak sınıflanıyor. Bu sınıflamaya göre testin doğruluğu aşağıdakilerle ilişkilidir.

1. Ölçümler normal dağılım gösteriyor mu?
2. Kesim noktası doğru mu?
3. Klinik olarak test sonucu destekleniyor mu?
4. Sağlıklı gönüllülerin sonuçları hastalar için ne kadar doğru?

Tanı testlerinin değerlendirilmesinde yapılan diğer bir hata çalışma tasarımı ile ilgilidir. Örneğin kullanılmakta olan, pozitifliği yüksek ve uygulaması zor olan bir teste alternatif olarak geliştirilen, kullanımı çok daha kolay bir test önerilsin. Yeni testin doğruluğunu tespit etmek amacıyla yeni bir çalışma tasarlandığında sağlıklı gönüllülere uygulaması zor olan testi uygulamak etik olmayacağından doğru sonuçları elde etmek mümkün olmayacaktır.

Diğer bir hata ise uyumla ilgili çalışmalarda görülmektedir. Burada önerilen yeni test sonucu ile kullanımda olan test sonucu arasındaki uyuma bakarak karar vermek gerektiğinde bazı sorunlar olacaktır. Örneğin eğer yeni test ile kullanımda olan test karşılaştırıldığında genellikle uyumlu sonuçlar elde ediliyorsa, yeni testin kullanılması yönünde karar verilebilir. Eğer uyum sağlanmazsa, bu durumda hangi testin tercih edileceği ile ilgili karar vermek zor olacaktır. Ayrıca testler aynı oranda doğruluğa sahip ama birinin pozitif dediğine diğeri negatif diyorsa bu da sorun olacaktır.

Yukarıda verilen hataları yapmamak ve dolayısıyla tanı testinin değerlendirilmesini daha doğru yapmak amacıyla aşağıda verilenlere dikkat edilmelidir.

1. Seçilmiş denekler üzerinde çalışma yapmaktan kaçınma

Seçilmiş deneklerin kullanıldığı çalışmalarda tanı testini değerlendirmek yanlı sonuçlar doğuracaktır. Seçilmiş denekler ile yapılan çalışmalar çoğunlukla referans testinin uygulaması zor olduğu durumlarda gerçekleşmektedir. Tanı testi pozitif elde edilen bir bireye referans test uygulanmasına karar verilirken, tanı testi negatif olan bir bireye ise referans test uygulanması tercih edilmemektedir. Bu durumda da tanı testinin doğruluk ölçütleri yansız olarak tahmin edilememektedir.

2. Referans testin var olup olmadığı

Önerilen yeni bir tanı testinin doğruluğu, referans test sonucunun verdiği yanıtlarla karşılaştırılarak elde edilir. Ancak birçok hastalıkta referans test bulunmamaktadır. Bu gibi durumlarda yeni test sonuçları, göreli olarak diğer testlerden daha doğru sonuç veren ve sonuçları kesin doğru olamayan bir testin sonuçları ile karşılaştırılır. Buradaki sonuçları kesin doğru olmayan teste “kesin olmayan altın standart” (imperfect gold standard) denir. Yeni testin doğruluk ölçütlerini belirlemede Tutarsızlık Çözücü (Discrepant Resolution), Bileşik Referans Standardı (Composite Reference Standard), Parametrik olmayan ROC tahmini, Bayes yaklaşımı ve Gizli Sınıf Analizi (Latent Class Analysis) gibi yöntemler kullanılmaktadır. Bunlarla ilgili detaylı bilgi ileride verilmiştir.

3. Testi yorumlayanın yan tutması

Bazı durumlarda referans test sonucu bilinen bireylere yeni tanı testi uygulanır. Bu da bireyin test sonucuna göre ait olacağı sınıfı etkilemesine neden olur. Örneğin, radyolojik bulguların yeni test olarak önerildiği durumlarda bireyin referans test sonucu biliniyorsa, tanı testi sonucu yanlı olacaktır. Benzer durum bireye ait klinik bulguların test sonucundan önce bilinmesinde de geçerli olacaktır. Saha araştırmalarında önce bu

bilgiler elde edilir ve sonra tanı testi sonucuna bakılır. Bu da testin ayırıcılık gücünün olması gerekenden daha yüksek bulunmasına neden olacaktır.

4. Hasta popülasyonundaki farklılıktan test sonucunun etkilenmesi

Bazı hastalıklar bireyin demografik bilgilerinden ve hastalığa neden olan etkene maruz kalma süresinden etkilenmektedir. Bu gibi durumlarda uygulanan tanı test sonuçları da bu klinik parametrelerden etkilenmektedir. Örneğin yaşı dikkate almadan egzersiz eko testi sonucunda bireyler atipik ve tipik koroner hastası olarak sınıflanmak istendiğinde hatalı sonuçlar elde edilebilecektir. Tanı testinin performansı tüm grupta düşük bulunurken, bu grubun bir alt grubunda yüksek bulunulabilir. Bu durumda doğruluk ölçütleri için ortalama bir değer verilebilir. Sonuç olarak tanı testi performansını, sonuca etki eden değişkenlere göre düzeltme yaptıktan sonra belirlemede fayda vardır.

5. Tanı koymada yaşanan kararsızlık veya yeterli bulgu bulunmama sorunu

Bazı testler tanıda kararsızlığa neden olabilmektedir. Örneğin radyolojik bulguların değerlendirilmesinde şüpheli tanısının konması tanıda kararsızlıktır. Bazı testlerde ise bireyle ilgili yeterli bulgu bulunamadığından yorumlanamamaktadır. Örneğin yine radyolojik görüntülemede idrara sıkışık gelmesi gereken bireyin bu durumda olmaması durumunda test sonucu yorumlanamamaktadır. Bu gibi durumlarda tanı testi sonucunu hasta veya sağlıklı diye belirlemek duyarlılığı veya seçiciliği olması gerekenden yüksek bulmaya neden olabilmektedir.

6. Tanı testinin tekrarlanabilir olması

Bazı testler sübjektif kriterlere göre değerlendirilmektedir. Bu gibi durumlarda testin performansı testi değerlendiren kişiye, kullanılan ekipmana ve uygulanan laboratuar prosedürüne göre değişkenlik gösterebilmektedir. Örneğin hipertansiyon tanısını koymada ardışık yapılmış tansiyon ölçümleri kullanılacaksa, uzman olmayan ve standart bir ekipmana sahip olan bir bireyin kararı uzman olan ve özel ekipmanı olan bir bireyin kararı ile uyum göstermeyebilir. Tanı testinin tekrarlanabilir olması ve tekrarlarında tutarlı olması çok önemlidir. Testin doğruluk ölçütlerini verirken testin tekrarlanabilirlik düzeyine göre düzeltme yapılmalıdır.

7. Doğruluk ölçütlerinin güven aralıklarını vermek

Tanı testleri için doğruluk ölçütleri örneklemlerden elde edilen bilgilerden hesaplanmaktadır. Ancak bu değeri popülasyona genelleyebilmek için güven aralıkları

ile birlikte vermek gerekir. Güven aralığı ile verilmiş tahmini ölçüt değeri tanı testinin popülasyondaki doğruluğunu verecektir.

2.2 Tanı testlerinin değerlendirilmesinde kullanılan ölçütler

Tanı testlerinin değerlendirilmesinde birçok yöntem kullanılmaktadır. Test ile ölçülenin niceliksel veya niteliksel olmasına göre yöntemler değişmektedir. Niteliksel ölçümlü testler genellikle sonucu iki grup olan testlerdir. Bu testlerde doğruluk ölçütleri duyarlılık, seçicilik, pozitif prediktif değer, negatif prediktif değerdir. Niceliksel ölçümlü testlerde ise doğruluk ölçütü ROC eğrisi altında kalan alan ve olasılıklar oranıdır.

2.2.1 Niteliksel Ölçümlü Testlerde Doğruluk Ölçütleri

Bireyin gerçek durumu D ile gösterilmek üzere, D = 1 bireyin hasta olduğunu ve D = 0 bireyin sağlıklı olduğunu göstersin. Bireye uygulanan test sonucu ise T ile gösterilsin ve T = 1 bireyin test sonucu hasta, T = 0 ise bireyin test sonucunun sağlıklı olduğunu ifade etsin.

Duyarlılık, bir koşullu olasılıktır ve P(T = 1 | D = 1) ile hesaplanır. Bu ifade gerçek durumu 1, yani hasta olan bir birey verildiğinde bireyin test sonucunun 1, yani hasta çıkma olasılığıdır. Diğer bir deyişle, gerçekte hasta olan bireylerin, tanı testi tarafından belirlenen yüzdesidir. İdeal olanı 1 (%100) değerini elde etmektir ancak pratikte bu mümkün olmayabilir. Yüksek duyarlıklı testler hastaları yüksek doğrulukta yakalarlar.

Seçicilik de duyarlılık gibi bir koşullu olasılıktır ve P(T = 0 | D = 0) ile hesaplanır. Bu da gerçekte sağlıklı olan bir birey verildiğinde, bireyin test sonucunun sağlıklı çıkma olasılığı demektir. Yani gerçekte sağlıklı olan bireylerin, tanı testi tarafından belirlenen yüzdesidir. Duyarlılığa benzer şekilde 1 (%100) seçicilik değerini yakalamak mümkün olmayabilir. Yüksek seçiciliğe sahip testler de sağlıklıları yakalar.

Duyarlılık ve seçicilik değerleri aşağıdaki bilgilerle hesaplanabilir.

Çizelge 2.1 Referans test ile Tanı testi karşılaştırılması

		Gerçek Durum		Toplam
		Pozitif	Negatif	
Tanı Testi Sonucu	Pozitif	a	b	a+b
	Negatif	c	d	c+d
Toplam		a+c	b+d	N

$$\text{Duyarlılık} = \frac{a}{a+c} \qquad \text{Seçicilik} = \frac{d}{b+d}$$

$$\text{Pozitif Prediktif Değer} = \frac{a}{a+b} \qquad \text{Negatif Prediktif Değer} = \frac{d}{c+d}$$

$$\text{Yanlış Pozitif Oranı} = \frac{b}{a+b} \qquad \text{Yanlış Negatif Oranı} = \frac{c}{c+d}$$

Testin doğruluğunu saptamada kullanılan diğer ölçütler ise Pozitif Prediktif Değer ve Negatif Prediktif Değerdir. Pozitif prediktif değer, tanı testi sonucu hasta olan bireylerin, gerçekte yüzde kaçının hasta olduğunu gösterir. Diğer ölçütler gibi bu da koşullu olasılıktır. Buna göre test sonucu pozitif verildiğinde gerçekte pozitif olma olasılığı pozitif prediktif değerdir. Diğer bir deyişle, P(D = 1 | T = 1)'dir. Burada gerçekte pozitif olmayan ama teste göre sonucu pozitif olan bireyler yanlış pozitiflerdir. Pozitif prediktif değer yükseldikçe yanlış pozitifler düştüğü için seçicilik de yükselir.

Negatif prediktif değer, tanı testi sonucu sağlıklı olan bireylerin, gerçekte yüzde kaçının sağlıklı olduğunu gösterir. Bu da koşullu olasılıktır ve P(D = 0 | T = 0) ile hesaplanır. Bu ifadeye göre, test sonucu negatif verildiğinde gerçekte negatif olma olasılığı negatif prediktif değerdir. Benzer şekilde tanı testinin negatif dediği ancak gerçekte pozitif olan bireyler yanlış negatiflerdir. Negatif prediktif değerle duyarlılık arasında da pozitif bir ilişki vardır.

Duyarlılık ve seçicilik değerleri tanısı konacak hastalığın prevelansından etkilenmezler. Bu değerler çalışmadaki gerçek hasta sayısı veya gerçek sağlıklı sayılarına göre hesaplandığından, hastaların popülasyondaki oranı sonuca etki etmez. Yani çalışmaya alınan hasta sayısı ile sağlıklı sayısı aynı olsa veya hastalık prevelansına göre oranlı miktarda hasta ve sağlıklı alınsa bu hesaplamalar değişmez.

Duyarlılık ve seçicilik değerleri çalışmadaki bireylerin özelliklerinden etkilenir. Çalışmaya alınan bireyler hasta ya da sağlıklı grubun özelliklerini taşıyorsa, bu değerlerin olması gerekenden farklı çıkmasına neden olur. Örneğin, koroner hastalık tanısı koymada kullanılacak bir testin duyarlılık ve seçiciliğini belirlemek için alınan bireyler daha çok yaşlılar ise bu durumda testin duyarlılığı yüksek çıkabilecektir.

Kimi durumlarda testin performansı için duyarlılık ve seçicilik değerleri yerine tek bir değeri kullanmak daha anlaşılır olabilir. Bunu yapabilmek için birkaç yol izlenebilir. En basit olanı doğruluk değerini hesaplamaktır. Doğruluk değeri testin doğru sonuç verme olasılığıdır ve doğru olarak sınıfladığı (hasta veya sağlıklı) bireylerin çalışmadaki toplam birey sayısına bölünmesi ile hesaplanır. Doğruluk değerinin elde edilmesi kolay olduğundan sıkça tercih edilir. Ancak kullanılması ile ilgili bazı problemler vardır. Örneğin, aynı duyarlılık ve seçicilik değerleri farklı büyüklükteki çalışmalardan elde edilebilir. Ancak bunların doğruluk değerleri örneklem büyüklükleri farklı olduğundan farklı çıkacaktır. Bu da büyük örneklemli çalışmalarda daha düşük, küçük örneklemli çalışmalarda daha büyük doğruluk değeri elde edilmesine neden olacaktır.

Duyarlılık ve seçiciliği tek bir değer altında toplamada diğer bir yöntem Odds Ratio'yu elde etmektir. Odds Ratio olasılıkların oranıdır ve duyarlılığın yanlış negatif oranına oranlanması ile seçiciliğin yanlış pozitif oranına oranlanmasının çarpılması sonucu elde edilir. Diğer bir deyişle duyarlılık ile seçiciliğin çarpımının, yanlış pozitif oran ile yanlış negatif oranın çarpılmasına oranıdır.

$$Odds \quad Ratio = \frac{Duy/(1-Duy)}{(1-Seç)/Seç} = \frac{Duy \times Seç}{YPO \times YNO}$$

Duy: Duyarlılık, Seç: Seçicilik, YPO: Yanlış Pozitif Oranı, YNO: Yanlış Negatif Oranı

Odds Ratio'yu yorumlarken dikkat etmek gerekir. Odds Ratio'nun 1.0 olması, test sonucunun pozitif olma olasılığı, hasta ve sağlıklar için aynıdır demektir. 1.0 değerinden büyük olması test sonucunun hastalar için pozitif çıkma olasılığı sağlıklılardan yüksek demektir. Benzer şekilde 1.0'den küçük çıkması da sağlıklıların test sonuçlarının pozitif çıkma olasılığının hastalardan yüksek olması anlamındadır.

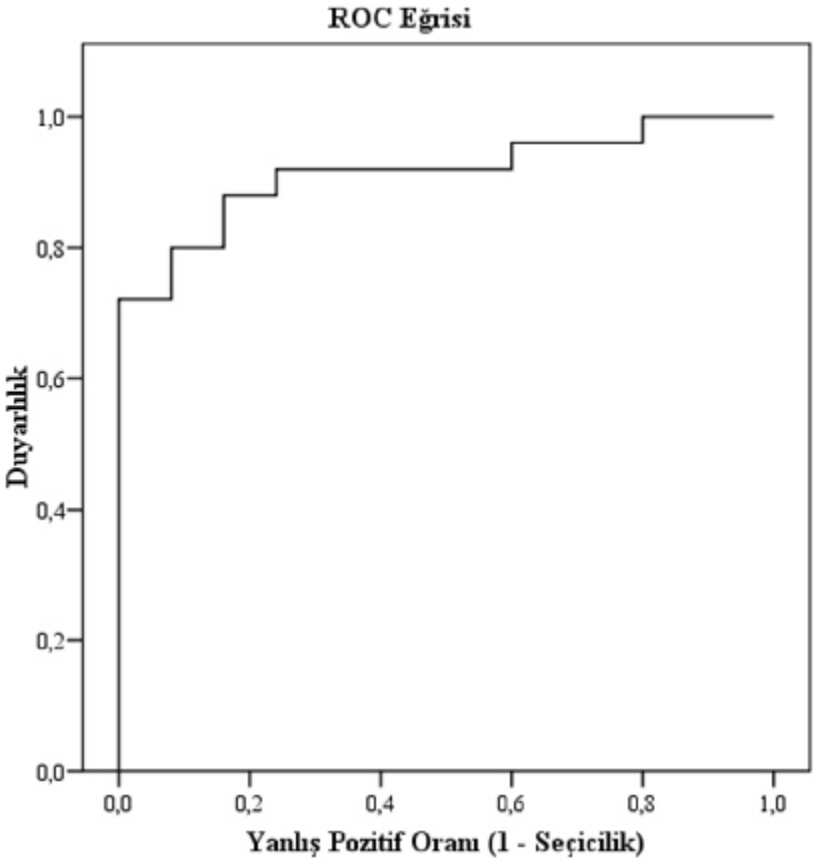
Duyarlılık ile seçiciliği birleştirmede kullanılan diğer bir yöntem ise Youden indeksidir (Youden J istatistiği)[23]. Bu indeks duyarlılık ile seçicilik değerinin toplamlarından 1 çıkarılarak elde edilir. Bu indeks -1 ile +1 arasında değerler alır. +1 değeri en iyi durumdur.

Doğruluk değeri hastalık prevelansından etkilenirken, Odds Ratio ve Youden indeksi etkilenmez. Ancak bu üç değer de duyarlılık ve seçicilik değerini birleştirdiğinden farklı durumlar için aynı sonuçlar alınabilir. Örneğin duyarlılığın 0.9 ve seçiciliğin 0.3 olması ile duyarlılığın 0.3 ve seçiciliğin 0.9 olması aynı sonuçları verecektir.

2.2.2 Niceliksel Ölçümlü Testlerde Doğruluk Ölçütleri

Niceliksel ölçümlü testler uygun kesim noktaları belirlenerek niteliksel ölçümlü testlerde kullanılan doğruluk ölçütleri ile değerlendirilebilir. Kesim noktaları niceliksel veriyi niteliksel veriye dönüştürür. Burada önemli olan uygun kesim noktasının belirlenmesidir. İşlem karakteristiği eğrisi (Reciever Operating Characteristic Curve – ROC) uygun kesim noktasını belirlemede kullanılan bir grafiktir.

ROC eğrisi testin duyarlılığı ile yanlış pozitif oranını eksenlerde kullanan bir grafiktir. Niceliksel ölçümlü testte elde edilen her bir değer kesim noktası olarak varsayılarak testin o değerdeki duyarlılık ve yanlış pozitif oranı değerleri hesaplanır. Bunlar koordinat düzlemine yerleştirilir. Yerleştirilen noktaların birleştirilmesiyle de ROC eğrisi elde edilir. (Şekil 2.1)



Şekil 2.1 ROC Eğrisi

Bu eğri altında kalan en fazla 1 olabilir. Eğri altında kalan alan 0.5 ise bu tanı testinin yazı tura atmaktan farklı olmadığını gösterir. Her zaman için yüksek değerli alan değerleri tercih edilir.

ROC eğrisini kullanmanın birçok avantajı vardır. ROC eğrisi test performansının görsel halidir. ROC eğrisinin x ve y eksenlerinde duyarlılık ve yanlış pozitif oranı bulunmakta ve farklı kesim noktalarında testin performansı görülebilmektedir. Eğri duyarlılık ve seçicilikten elde edildiğinden, hastalık prevelansından etkilenmez. Ancak duyarlılık ve seçicilikte olduğu gibi hastaların durumları eğriyi etkiler. ROC eğrisinin diğer bir avantajı da test sonucunun değerleri ile değil sıraları ile ilgilenmesidir. Bu da eğrinin çiziminde ölçümün gerçek değeri değil de sıralamadaki yerinin kullanıldığı anlamına gelir.

Tanı testlerini değerlendirmede kullanılan diğer bir yöntem de olabilirlilik oranıdır (likelihood ratio). Olabilirlilik oranı niceliksel ölçümlü bir test için belirlenen bir noktaya göre hesaplanabildiği gibi niteliksel ölçümlü testin sonuçları için de bu değer ayrı ayrı hesaplanabilmektedir. Olabilirlilik oranı şu şekilde elde edilir.

$$LR(t) = \frac{P(T = t \mid D = 1)}{P(T = t \mid D = 0)}$$

veya

$$LR(+) = \frac{P(T = 1 \mid D = 1)}{P(T = 1 \mid D = 0)} \qquad LR(-) = \frac{P(T = 0 \mid D = 1)}{P(T = 0 \mid D = 0)}$$

İki sonuçlu testler için elde edilen olasılıklar oranı, test sonucu pozitif iken duyarlılığın yanlış pozitif oranına bölünmesi ile ve test sonucu negatif iken yanlış negatif oranının seçiciliğe bölünmesi ile elde edilebilir.

Olabilirlilik oranı test sonucunun hasta ya da sağlıklı olma olasılığının göreceli ölçüsünü verir. Bu değerin 1.0 olması test sonucunun hasta veya sağlıklı çıkma olasılıklarının eşit olduğunu gösterir. 1.0'dan büyük sonuçlar elde edildiğinde test sonucunun hasta olma olasılığı sağlıklı olma olasılığından yüksek demektir. 1.0'dan küçük sonuçlar ise sonucun sağlıklı çıkma olasılığının daha yüksek olduğunu gösterir. Örneğin LR(+) = 1.80 bulunduysa, gerçek hastalarda gerçek sağlıklara göre %80 daha fazla olasılıkla test sonucu pozitif çıkacaktır. Benzer şekilde LR(-) = 0.25 bulunduysa, gerçek sağlıklılarda gerçek hastalara göre test sonucu 1/0.25 = 4 kat daha fazla olasılıkla negatif çıkacaktır.

Olabilirlilik oranı ile ROC eğrisi arasında bir ilişki vardır. Buna göre $t_1 - t_2$ için hesaplanan olasılıklar oranı, t_1 ve t_2 noktaları arasındaki doğrunun eğimini verecektir[24]. Zweig ve Campell[25] ROC eğrisi olmadan sadece olabilirlilik oranına bakmanın sonucu yanlış yöne götürebileceğini belirtmiş ve aynı olabilirlilik oranına sahip iki farklı ROC eğrisi olabileceğini göstermiştir. Bu eğrilerden bir tanesinin eğri altında kalan alanı 1'e yakın iken diğeri 0.5 civarında bir alan değerine sahip bulunmuştur[25].

2.3 İki testin doğruluklarının karşılaştırılması

İki testin doğruluğunu (accuracy) karşılaştırırken genellikle iki testin doğruluk ölçütleri karşılaştırılır. Bu ölçütler duyarlılık, seçicicilik, eğri altında kalan alan vb. gibi olabilir. Kurulan sıfır hipotezi "iki testin doğruluk ölçütleri aynıdır" ve alternatif hipotezi "iki test birbirinden farklıdır" şeklinde olur. Yani,

$$H_0 : u_1 = u_2 \qquad\qquad H_A : u_1 \neq u_2$$

olur. Burada u_i i. testin doğruluk ölçütüdür.

Bu hipotezler genel test mantığı ile test edilecek olursa, $Var(\hat{u}_1 - \hat{u}_2)$ farkların varyansı ve $Var(\hat{u}_1 - \hat{u}_2) = Var(\hat{u}_1) + Var(\hat{u}_2) - 2Cov(\hat{u}_1, \hat{u}_2)$ olmak üzere,

$$Z = \frac{\hat{u}_1 - \hat{u}_2}{\sqrt{Var(\hat{u}_1 - \hat{u}_2)}}$$

test istatistiği kullanılır ve bu istatistik genellikle asimptotik olarak normal dağılım gösterir. Burada farkların varyansındaki kovaryans terimi testlerin aynı bireylere uygulanması veya farklı bireylere uygulanması durumuna göre değişim gösterir. Aynı bireylere uygulanıyorsa testlerin doğruluk ölçütleri arasında bir korelasyon var demektir ve bu da kovaryansın sıfırdan farklı olmasına neden olur. Eğer testler farklı bireylere uygulandıysa kovaryans sıfır olacaktır.

İki testin doğruluklarını karşılaştırırken farklı iki test veya bir test ama iki farklı değerlendirici ya da bir test ama iki farklı hasta grubu incelenebilir.

2.3.1 İki Değerli Testlerin Doğruluklarının Karşılaştırılması

2.3.1.1 Duyarlılık – Seçicilik:

İki değerli bağımsız iki testin duyarlılık ve seçiciliklerini karşılaştırmak, iki bağımsız oranı karşılaştırmak ile aynıdır. Karşılaştırmada iki testin duyarlılık (veya seçicilik) değerleri aynıdır veya farklıdır hipotezleri kullanılır. Yani,

$$H_0 : Duy_1 = Duy_2 \qquad H_A : Duy_1 \neq Duy_2$$

hipotezleri kullanılır. Önceden de belirtildiği gibi test istatistiği,

$$Z = \frac{D\hat{u}y_1 - D\hat{u}y_2}{\sqrt{Var\left(D\hat{u}y_1 - D\hat{u}y_2\right)}} = \frac{D\hat{u}y_1 - D\hat{u}y_2}{\sqrt{\frac{D\hat{u}y_1\left(1 - D\hat{u}y_1\right)}{n_{11}} + \frac{D\hat{u}y_2\left(1 - D\hat{u}y_2\right)}{n_{21}}}}$$

olur. Buradaki n_{i1} değeri i. örneklemdeki pozitif sonuçlu kişi sayısıdır. Bu istatistik seçicilik için de benzer şekilde elde edilebilir. Örneklemelerdeki birey sayısı çok olduğunda bu istatistik asimptotik olarak normal dağılım gösterir. Ancak örneklemlerdeki kişi sayısı az ise Fisher's exact testi tercih edilmelidir[11].

Bağımlı testlerin karşılaştırılmasında ise bağımlı oranlar için McNemar testi kullanılmalıdır. İki test aynı bireylere uygulandığında Çizelge 2.2'de verilen tablo elde edilir. Bu tabloda köşegen dışında olan sayıların farkı McNemar testinin sonucunu etkilemektedir.

Çizelge 2.2 İki tanı testinin karşılaştırılması

		Tanı Testi II Sonucu		Toplam
		Pozitif	Negatif	
Tanı Testi I Sonucu	Pozitif	m_{11}	m_{10}	s_{21}
	Negatif	m_{01}	m_{00}	s_{20}
Toplam		s_{11}	s_{10}	n_1

Testler arasında uyumsuzluğun olduğu durum m_{10} ve m_{01} değerlerinin sıfırdan farklı olmasında gerçekleşir. McNemar testinde bu iki değerin farkının sıfır veya sıfıra yakın olması uyumun olduğu anlamına gelir. McNemar testinde kullanılan ki kare değeri şöyledir:

$$c^2 = \frac{\left(m_{10} - m_{01}\right)^2}{m_{10} + m_{01}}$$

Bağımsız testlerde olduğu gibi bağımlı testlerde de uyumsuzluğa neden olan gözlerdeki birey sayısı küçük olduğunda (yani $m_{10} + m_{01} < 20$) kesin (exact) testleri kullanmak daha doğru olacaktır. Burada önerilen testler, işaret testi veya binom testi olabilir. Burada p değeri şu şekilde hesaplanabilir. $m = m_{01} + m_{10}$ ve $k = \min(m_{01}, m_{10})$ olmak üzere,

$$p = 2 \times \sum_{j=1}^{k} \binom{m}{j} \left(\frac{1}{2}\right)^m$$

iki yönlü test için p değeridir[11].

Kümelenmiş iki değerli (clustered binary) testlerin karşılaştırılmasında bağımsız testlerin karşılaştırılmasında kullanılan test istatistiği kullanılacaktır. Buradaki fark, duyarlılık (ve seçicilik) için varyansın değişmesidir.

$$\hat{Var}(D\hat{u}y_i) = \frac{1}{I_i(I_i - 1)} \sum_{j=1}^{I} \left(\frac{N_{ij}}{\overline{N}_i}\right)^2 \left(D\hat{u}y_{ij} - D\hat{u}y_i\right)^2$$

Burada $D\hat{u}y_{ij}$ j kümesindeki i. testin duyarlılık değeri, N_{ij} j kümesinde i testi yapılan birey sayısı, I_i küme sayısı ve $\overline{N}_i$ i. testteki ortalama küme sayısıdır[11].

2.3.2 Sürekli Değerler Alan Testlerin Doğruluklarının Karşılaştırılması

Sürekli değerler alan testlerin doğruluklarının karşılaştırılmasında temel alınan ROC eğrileridir. İki ROC eğrisini karşılaştırmada bakılan 3 temel neden vardır[11].

1. ROC eğrilerinin tamamen aynı olup olmadıklarını inceleme. Bu inceleme yapılırken bakılan, iki ROC eğrisinin gerçek pozitif oranları her yanlış pozitif oran için aynı mıdır sorusunun yanıtıdır. Diğer bir deyişle binormal dağılımdan elde edilen a_i ve b_i parametreleri iki ROC eğrisinde de aynıdır hipotezini test etmektir.

$$H_0 : a_1 = a_2 \quad ve \quad b_1 = b_2$$

2. Herhangi bir hatalı pozitif oran için iki ROC eğrisinin verdiği değerleri inceleme. Yani, yanlış pozitif oran e alındığında, aşağıdaki hipotez test edilir.

$$H_0 : Se_1(e) = Se_2(e)$$

3. İki eğri altında kalan alanların aynı olup olmadığını inceleme. Yani,

$$H_0 : A^1_{(e_1 \le HPO \le e_2)} = A^2_{(e_1 \le HPO \le e_2)}$$

2.3.2.1 İki ROC eğrisinin eşitliğinin belirlenmesi:

Hasta ve sağlıklı popülasyonu bir testle incelendiğinde elde edilen değerler binormal dağılım gösterir. (Şekil 2.2) Bu şekilde sağlıklı popülasyon ile hasta popülasyonun kesiştiği nokta kesim noktası olarak alınırsa tüm doğruluk ölçütleri eğri altında kalan alanlarda görünür. Binormal dağılımın ikişer parametresi vardır. (iki ortalama ve iki standart sapma) Bu değerlerdeki bilgiyi tek bir değişkende toplamak için a ve b parametreleri şu şekilde tanımlanabilir:

$$b_i = \frac{s_{i0}}{s_{i1}} \qquad \text{ve} \qquad a_i = \frac{m_{i1} - m_{i0}}{s_{i1}}$$

Burada i test numarasını m_{i0} ve s_{i0} i testindeki sağlıklı popülasyonun ortalama ve standart sapmasını m_{i1} ve s_{i1} i testindeki hasta popülasyonun ortalama ve standart sapmasını göstermektedir.

İki ROC eğrisinin eşitliğinin testinde, yukarıda verilen hipotezlerin testinde kullanılan istatistik, Metz ve Kronman tarafından önerilen ki kare istatistiğidir[26,27]. Eğer $a_{12} = a_1 - a_2$ ve $b_{12} = b_1 - b_2$ alınırsa istatistik,

$$c^2 = \frac{\hat{a}_{12}^2 Var\left(\hat{b}_{12}\right) + \hat{b}_{12}^2 Var\left(\hat{a}_{12}\right) - 2\hat{a}_{12}\hat{b}_{12} Cov\left(\hat{a}_{12}, \hat{b}_{12}\right)}{Var\left(\hat{a}_{12}\right) Var\left(\hat{b}_{12}\right) - Cov\left(\hat{a}_{12}, \hat{b}_{12}\right)^2}$$

olur.

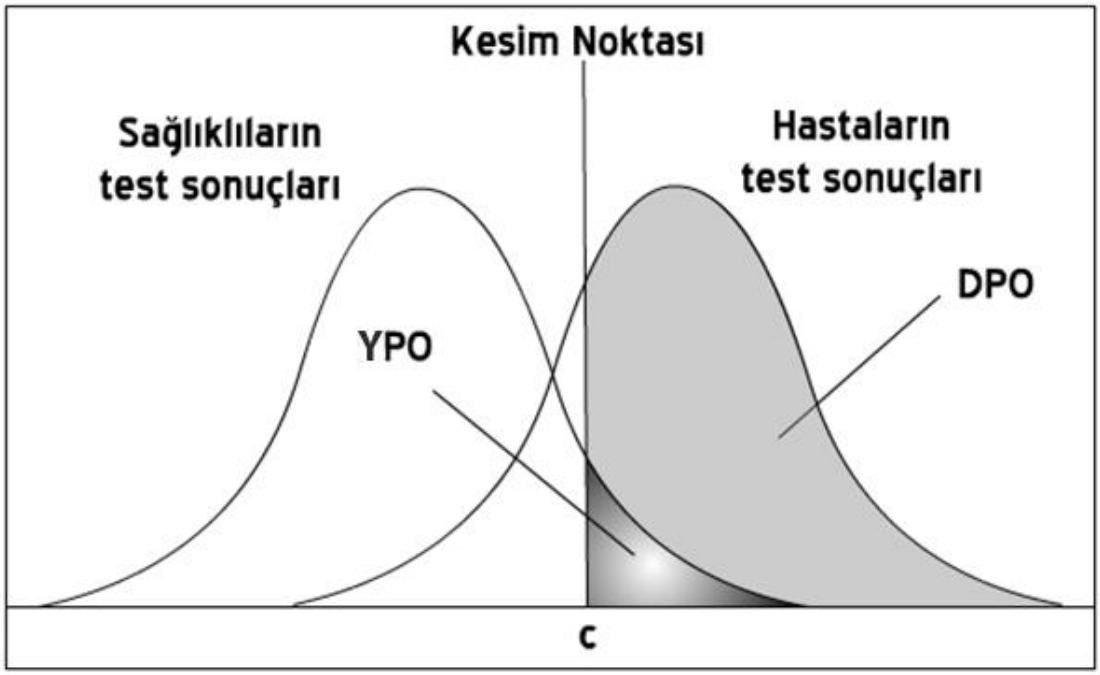
Burada varyanslar ve kovaryans şöyle hesaplanır.

$$Var(\hat{a}_{12}) = Var(\hat{a}_1 - \hat{a}_2) = \hat{s}^2_{a_1} + \hat{s}^2_{a_2}$$

$$Var(\hat{b}_{12}) = Var(\hat{b}_1 - \hat{b}_2) = \hat{s}^2_{b_1} + \hat{s}^2_{b_2}$$

$$Cov(\hat{a}_{12}, \hat{b}_{12}) = Cov(\hat{a}_1 - \hat{a}_2, \hat{b}_1 - \hat{b}_2) = \hat{s}_{a_1 b_1} + \hat{s}_{a_2 b_2}$$

Eğer bağımlı iki test karşılaştırılacak ise Var(a_{12}) ve Var(b_{12}) hesabında kovaryans terimi eklenmelidir. Bu durumda Cov(a_{12},b_{12}) hesabı da değişecektir.



Şekil 2.2 Binormal dağılım eğrileri ve kesim noktasına göre DPO ve YPO değerleri

Metz ve Kronman'ın test istatistiğinin alternatifi Venkatraman ve Begg tarafından 1996 yılında önerilmiştir[28,29]. Bu yaklaşımda binormal dağılım varsayımı bulunmamakta ve permütasyon testi kullanılmaktadır. Burada test edilen, iki eğri arasında kalan alanın sıfır olup olmadığıdır.

İki ROC eğrisini karşılaştırmada kullanılan diğer bir yöntem seçilen bir noktada ROC eğri değerlerinin aynı olup olmadığına bakmaktır. Eğer, yanlış pozitif oran değeri

e olarak seçilir ve iki ROC eğrisinde bu değere karşılık gelen duyarlılığa (doğru pozitif oranı) bakılırsa

$$Z_{Duy} = b_1 Z_e - a_1 \qquad \text{1. ROC eğrisi için}$$

$$Z_{Duy} = b_2 Z_e - a_2 \qquad \text{2. ROC eğrisi için}$$

değerleri elde edilir. Bunların farkı $D(Z_e)$ ile gösterilirse,

$$D(Z_e) = (b_1 Z_e - a_1) - (b_2 Z_e - a_2) = (b_1 - b_2) Z_e - (a_1 - a_2) = b_{12} Z_e - a_{12}$$

olduğu görülür. Bu farkın varyansı aşağıdaki gibi olacaktır.

$$Var\left[\hat{D}(Z_e)\right] = (Z_e)^2 Var\left(\hat{b}_{12}\right) + Var\left(\hat{a}_{12}\right) - 2 Z_e Cov\left(\hat{a}_{12}, \hat{b}_{12}\right)$$

$\hat{a}_i$ ve $\hat{b}_i$ tahmin edicileri maksimum olabilirlilik tahmini (MLE) ise $\hat{D}(Z_e)$ asimptotik olarak normal dağılım göstermelidir. Bu durumda YPO = e iken iki eğrinin duyarlılık değerleri farkı sıfırdır hipotezini test etmek için kullanılacak test istatistiği şöyledir:

$$Z = \frac{\hat{D}(Z_e)}{\sqrt{Var\left[\hat{D}(Z_e)\right]}}$$

Eğer bu oran büyükse, aynı yanlış pozitif oran değerinde iki eğrinin duyarlılık değerleri birbirlerinden farklıdır sonucunu bulmak kuvvetle muhtemeldir.

Seçilen bir yanlış pozitif oran değerindeki duyarlılık değerlerini karşılaştırmak ile belirlenen bir kesim noktasındaki duyarlılık değerlerini karşılaştırmak biraz farklı olacaktır. Eğer kesim noktası T ile gösterilirse, iki duyarlılık değerinin karşılaştırılmasında kullanılacak istatistik şöyle olacaktır:

$$Z = \frac{D\hat{u}y_1(T) - D\hat{u}y_2(T)}{\sqrt{Var\left[D\hat{u}y_1(T) - D\hat{u}y_2(T)\right]}}$$

Buradaki $D\hat{u}y_i(T)$ i. teste ait T kesim noktasındaki duyarlılık değerini göstermektedir. Farkın varyansının hesabı Zhou, Obuchowski ve McClish'in kitabında verilmiştir[11].

ROC eğrilerini karşılaştırmada kullanılan diğer bir yaklaşım da eğri altında kalan alanların aynı olup olmadığını incelemektir. Burada i. eğrinin altında kalan alan $A^i_{(e_1 \le HPO \le e_2)}$ ile gösterilirse, yukarıda verilen istatistik şu şekilde düzenlenebilir.

$$Z = \frac{A^1_{(e_1 \le HPO \le e_2)} - A^2_{(e_1 \le HPO \le e_2)}}{\sqrt{Var\left[A^1_{(e_1 \le HPO \le e_2)} - A^2_{(e_1 \le HPO \le e_2)}\right]}}$$

Burada testler bağımsız ise kullanılacak varyans terimi içinde kovaryans bulunmayacaktır. Yani,

$$Var\left[A^1_{(e_1 \le HPO \le e_2)} - A^2_{(e_1 \le HPO \le e_2)}\right] = Var\left[A^1_{(e_1 \le HPO \le e_2)}\right] + Var\left[A^2_{(e_1 \le HPO \le e_2)}\right]$$

olur. Burada $Var\left[A^1_{(e_1 \le HPO \le e_2)}\right]$ ve $Var\left[A^2_{(e_1 \le HPO \le e_2)}\right]$ terimlerinin nasıl elde edildiği ve testlerin bağımlı olması durumunda kullanılacak olan kovaryans terimi Zhou, Obuchowski ve McClish'in kitabında verilmiştir[11].

2.4 Doğrulama yanlılığı

2.4.1 Doğrulama Yanlılığı Nedir?

Yeni geliştirilen bir tanı testinin veya kullanımda olan ancak sınıflama başarısı kesin olarak belirlenmemiş bir tanı testinin sınıflama başarısının belirlenmesi, testin kullanımında oluşabilecek şüphelerin giderilmesi için gereklidir. Tanı testlerinin sınıflama başarısının belirlenmesi amacıyla yapılan çalışmalarda tanı testi sonucu bireylerin gerçek durumları ile karşılaştırılır. Bu çalışmalarda bireylerin gerçek

durumunun tayininde altın standart denilen referans testler kullanılır. Çalışmada tasarım gereği uygulanması gereken, bireylerin hem tanı testi sonucu hem de referans test sonucunun kayıt edilmesidir. Ancak referans testleri, gerek uygulanmasının zor olması, gerekse pahalı olması nedeniyle çalışmadaki her bireye uygulanamayabilir. Bu durumda eldeki veride hastalık bilgisi doğrulananlar, yani referans test sonucu olanlar ve hastalık bilgisi doğrulanmayanlar, yani referans test sonucu olmayan hastalar hakkında sonuçlar olacaktır. Böyle bir veri setinde, önerilen yeni bir tanı testinin performansı ölçüldüğünde elde edilen sonuçlar veri setindeki kimi hastaların gerçek durum bilgisi olmadığından yanlı olacaktır. Bu yanlılığa doğrulama yanlılığı veya seçme yanlılığı (verification bias, selection bias, work-up bias) denir3.

Literatürde birçok yazar doğrulama yanlılığına dikkat etmemiş ve tanı testlerinin doğruluklarını yanlış hesaplamışlardır. Greenes ve Begg 1976 ve 1980 yılları arasında yayınlanan 145 makaleyi incelemiş ve bunların %26'sında doğrulama yanlılığı olduğunu ve dolayısıyla yanlı doğruluk ölçütleri saptandığını göstermiştir[30]. Benzer şekilde Bates, Margolis ve Evans 54 adet makaleyi incelemiş ve 3'te birinden fazlasında yanlı sonuçlar elde edildiğini vurgulamışlardır[31].

Doğrulama yanlılığının düzeltilmesinde referans test sonucu olmayan bireyler eksik ölçümlü bireyler olarak kabul edilir. Bu eksik ölçümlerin oluşma mekanizması doğrulama yanlılığının düzeltilmesinde kullanılacak yaklaşımları belirler. Eksik ölçüm mekanizmaları ise üç başlık altında incelenir[32].

Tamamen rastgele oluşan eksik ölçümler (Missing Completely At Random-MCAR): Bireylere referans testi uygulanmaması durumu, bireylere ait ölçülen veya ölçülemeyen herhangi bir özelliğe dayanmadığı durumunda oluşan eksik ölçüm mekanizması MCAR olarak adlandırılır. Bu şekilde oluşmuş eksik ölçümlü bir veri seti, tam ölçümlü halinin özelliklerini aynen taşır. Dolayısıyla bu veri setinin kullanılması ile elde edilecek tahmin ediciler yansız olacaktır.

Rastgele oluşan eksik ölçümler (Missing At Random-MAR): Bireylere referans testi uygulanmaması durumu, bireylere ait ölçülmüş özelliklere dayandığı durumda oluşan eksik ölçüm mekanizmasıdır. Bu şekilde oluşmuş eksik ölçümlü veri setinden elde edilecek tahmin ediciler yanlı olacaktır.

Rastgele oluşmayan eksik ölçümler (Not Missing At Random-NMAR); Bireylere referans testi uygulanmaması durumu, bireylere ait ölçülmemiş özelliklere dayandığı

durumda oluşan eksik ölçüm mekanizması NMAR olarak adlandırılır. Bu durumda da elde edilecek tahmin ediciler yanlı olacaktır[32].

Eksik ölçüm mekanizmasının MAR veya NMAR olması halinde tanı testinin sınıflama başarısını belirlemede kullanılacak ölçütler için elde edilecek tahmin ediciler yanlı olacağından, bu iki mekanizma için ayrı ayrı düzeltmeler önerilmiştir. Doğrulama yanlılığı düzeltme yaklaşımları tanı testinin ölçüm ölçeğine göre de değişmektedir. Tanı testinin iki değer almasına göre, aynı anda birden fazla iki değer alan tanı testinin olmasına göre, tanı testinin sıralı yapıda olmasına göre ve aynı anda birden fazla sıralı yapıda tanı testinin olmasına göre önerilmiş yaklaşımlar vardır[11]. Bu tezde iki değer alan tek bir tanı testi için önerilen yaklaşımlar incelenmiştir.

2.4.2 Begg ve Greenes Yaklaşımı ile Yanlılığın Düzeltilmesi

Begg ve Greenes 1983 yılında MAR varsayımı altında iki değer alan (binary) tek bir test için bir düzeltme sunmuştur3. Bu düzeltmede T test sonucunu, D gerçek durumu ve t = (0,1) göstermek üzere,

$$f_{1t} = P(T = t) \qquad \text{ve} \qquad f_{2t} = P(D = 1 \mid T = t)$$

olsun. $f_1 = f_{11}$ ve $f_2 = (f_{20}, f_{21})$ alınırsa eksik ölçümlerin rastgele olduğu varsayımı ile logaritmik olabilirlilik fonksiyonu şu şekilde yazılabilir.

$$l(f_1, f_2) = \sum_{t=0}^{l} m_t \log(f_{1t}) + \sum_{t=0}^{l} [s_t \log(f_{2t}) + r_t \log(1 - f_{2t})]$$

Burada $f_{10} = 1 - f_1$'dır. Bu logaritmik olabilirlilik fonksiyonu maksimize edilir f_1 ve f_2 için maksimum olabilirlilik tahmin edicileri hesaplanırsa aşağıda verilen tahmin ediciler elde edilir.

$$\hat{f}_1 = \frac{m_1}{N} \qquad \text{ve} \qquad \hat{f}_{2t} = \frac{s_t}{s_t + r_t}$$

Burada verilen m, s ve r değerleri aşağıdaki tablodan elde edilir. N değeri m_1 ve m_0 değerlerinin toplamıdır.

Çizelge 2.3 İki değerli bir test için gözlenen veri

		Tanı testi sonucu	
		T = 1	T = 0
Referans test sonucu	D = 1	s_1	s_0
	D = 0	r_1	r_0
Eksik ölçümler		u_1	u_0
Toplam		m_1	m_0

f_1 ve f_2'nın maksimum olabilirlilik tahmin edicilerini kullanarak duyarlılık ve seçiciliği hesaplamak mümkündür. Buna göre duyarlılık ve seçiciliğin maksimum olabilirlilik (ML) tahmin edicileri şöyledir:

$$D\hat{u}y = \frac{m_1 s_1 / [N(s_1 + r_1)]}{m_0 s_0 / [N(s_0 + r_0)] + m_1 s_1 / [N(s_1 + r_1)]}$$

$$S\hat{e}\ç = \frac{m_0 r_0 / [N(s_0 + r_0)]}{m_0 r_0 / [N(s_0 + r_0)] + m_1 r_1 / [N(s_1 + r_1)]}$$

Bunların varyansları ise şöyledir:

$$V\hat{a}r(D\hat{u}y) = [D\hat{u}y(1 - D\hat{u}y)]^2 \left(\frac{N}{m_0 m_1} + \frac{r_1}{s_1(s_1 + r_1)} + \frac{r_0}{s_0(s_0 + r_0)} \right)$$

$$V\hat{a}r(S\hat{e}\ç) = [S\hat{e}\ç(1 - S\hat{e}\ç)]^2 \left(\frac{N}{m_0 m_1} + \frac{s_1}{s_1(s_1 + r_1)} + \frac{s_0}{s_0(s_0 + r_0)} \right)$$

2.4.3 Zhou Alt Sınır – Üst Sınır Yaklaşımı ile Yanlılığın Düzeltilmesi

1993 yılında Zhou MAR varsayımının olmadığı durumda iki değer alan tek bir test için bir düzeltme önermiştir4. Bu düzeltmede tanı testinin sınıflama başarısını ölçmede kullanılan ölçütler için (duyarlılık ve seçicilik) alt ve üst sınırlar maksimum olabilirlilik yaklaşımı ile elde edilmektedir.

Referans test sonucu olmayan bireylerin rastgele olmama durumu birkaç şekilde gerçekleşir. Bunlardan bir tanesi referans testi uygulanan bireylerin uygulanma zamanının tanı testi uygulama zamanından çok geç olmasıdır. Ayrıca farklı kurumlarda birden fazla araştırıcı bu testleri yapıyor veya hasta popülasyonu homojen değil ya da hastalık tam olarak anlaşılmamışsa eksik ölçümlerin rastgeleliğinden bahsetmek mümkün değildir.

Eksik ölçümlerin rastgele olmaması olabilirlilik fonksiyonunu değiştirecektir. Bu sebeple, fonksiyonda olacak bazı olasılıklar şu şekilde tanımlanabilir.

λ_{11} : Test sonucu ve referans test sonucu pozitif olan birey seçme koşullu olasılığı

λ_{01} : Test sonucu pozitif ve referans test sonucu negatif olan birey seçme koşullu olasılığı

λ_{10} : Test sonucu negatif ve referans test sonucu pozitif olan birey seçme koşullu olasılığı

λ_{00} : Test sonucu ve referans test sonucu negatif olan birey seçme koşullu olasılığı

Bu durumda logaritmik olabilirlilik fonksiyonunu aşağıdaki gibi yazabiliriz:

$$l = \sum_{t=0}^{1} m_t \log f_{1t} + \sum_{t=0}^{1} s_t \log(l_{1t} f_{2t}) + r_t \log[l_{0t}(1-f_{2t})] + u_t \log[(1-l_{1t})f_{2t} + (1-l_{0t})(1-f_{2t})]$$

Burada $\frac{l_{1t}}{l_{0t}}$ ifadesi e_t olarak atanırsa, fonksiyon şu şekilde yazılabilir.

$$l = \sum_{t=0}^{1} m_t \log f_{1t} + \sum_{t=0}^{1} s_t \log(e_t l_{0t} f_{2t}) + r_t \log[l_{0t}(1-f_{2t})] + u_t \log[(1-e_t l_{0t})f_{2t} + (1-l_{0t})(1-f_{2t})]$$

Bu fonksiyonun maksimize edilmesi sonucu parametreler tahmin edilebilir ancak e_0 ve e_1 değerlerinin biliniyor olması varsayımı gereklidir. Bu varsayım altında Zhou4 duyarlılık ve seçicilik için ML tahmin edicilerini şu şekilde hesaplamıştır.

$$D\hat{u}y(e_0, e_1) = \frac{\frac{s_1 m_1}{s_1 + e_1 r_1}}{\frac{s_1 m_1}{s_1 + e_1 r_1} + \frac{s_0 m_0}{s_0 + e_0 r_0}}$$

$$S\hat{e}\varsigma(e_0, e_1) = \frac{\frac{e_0 r_0 m_0}{s_0 + e_0 r_0}}{\frac{e_1 r_1 m_1}{s_1 + e_1 r_1} + \frac{e_0 r_0 m_0}{s_0 + e_0 r_0}}$$

Eğer $e_0 = e_1 = 1$ ise eksik ölçümler rastgele demektir ve bu durumda tahmin ediciler bir önceki bölümde verilen ile aynı olacaktır. Genellikle e_0 ve e_1 değerlerini veriden elde etmek mümkün değildir. Ancak bunlar için alt ve üst sınır bulmak mümkündür. Bilindiği gibi e_0 ve e_1 koşullu olasılıkların oranıdır. 1993 yılında Zhou bu değerler için sınırların şu şekilde olabileceğini göstermiştir4.

$$\frac{s_1}{s_1 + u_1} \le e_1 \le \frac{r_1 + u_1}{r_1} \qquad \text{ve} \qquad \frac{s_0}{s_0 + u_0} \le e_0 \le \frac{r_0 + u_0}{r_0}$$

2.4.4 Lojistik Regresyon Yaklaşımı ile Yanlılığın Düzeltilmesi

Lojistik Regresyon analizi sağlık alanında yapılan çalışmalarda birçok farklı amaçla kullanılmaktadır. Başlıca kullanım amaçları şu şekilde sıralanabilir: Kategorik ve/veya sürekli yapıdaki açıklayıcı değişkenler yardımı ile bağımlı bir değişkeni tahmin etmek, açıklayıcı değişkenlerin bağımlı değişkendeki varyasyonun ne kadarını açıkladığını belirlemek, açıklayıcı değişkenlerin göreceli önemlilik sıralarını belirlemek, açıklayıcı değişkenler arasındaki etkileşimi açıklamak, açıklayıcı değişkenlere kontrol değişkenlerinin (düzeltici değişkenler) etkilerini anlamak[33].

Lojistik Regresyon yaklaşımı açıklayıcı değişkenler yardımı ile bağımlı değişkenin tahmin edilmesi amacıyla kullanılabildiğine göre, referans test sonucu olmayan bireyler için test sonucunun tayininde bu yaklaşım kullanılabilir. Bu fikir ile

yola çıkan Kosinski ve Barharnt 2003 yılında eksik ölçümlerin belirlenmesi amacıyla genel olabilirlilik (likelihood) tabanlı ve logit linki kullanan bir yaklaşım önermiştir5.

Bu yaklaşımda tanı testi uygulanmış olan tüm hastaların bu bilgilerine ilave olarak ölçülmüş olan p ek değişkenin varlığı varsayılmaktadır. Eldeki bu gözlenen veriler ile olabilirlilik fonksiyonu şu şekilde yazılabilir.

$$L_{göz} = \prod_{i=1}^{N} P(V_i, T_i, D_i|\boldsymbol{x}_i)^{V_i} P(V_i, T_i|\boldsymbol{x}_i)^{1-V_i}$$

$$= \prod_{i=1}^{N} P(V_i, T_i, D_i|\boldsymbol{x}_i)^{V_i} \left\{ \sum_{d=0}^{1} P(V_i, T_i, D_i = d|\boldsymbol{x}_i) \right\}^{1-V_i}$$

Burada V : hastanın referans test sonucunun olup olmadığını, T : tanı testi sonucunu, D : referans test sonucunu, x : hastanın p ek değişkeninin değerlerini göstermektedir. Buradaki P(V,T,D) olasılığı koşullu olasılıklar çarpımı olarak yazılabilir.

$$P(V, T, D|\boldsymbol{x}) = P(D|\boldsymbol{x})P(T|D, \boldsymbol{x})P(V|T, D, \boldsymbol{x})$$

Kosinski ve Barnhart bu çarpımdaki koşullu olasılıkları şu şekilde adlandırıp, modellemiştir.

Hastalık bileşeni : $\text{logit P}(\text{D}_\text{i} = 1|\mathbf{x_i}) = \mathbf{z'_{0i}\alpha}$

Tanı testi bileşeni: $\text{logit P}(\text{T}_\text{i} = 1|\text{D}_\text{i}, \mathbf{x_i}) = \mathbf{z'_{1i}\beta}$

Eksik veri mekanizma bileşeni: $\text{logit P}(\text{V}_\text{i} = 1|\text{D}_\text{i}, \text{T}_\text{i}, \mathbf{x_i}) = \mathbf{z'_{2i}\gamma}$

Burada $\text{logit}(\text{p}) = \log\frac{\text{p}}{1-\text{p}}$ ile elde edilir. Bu bileşenler içinde eksik veri mekanizması daha açık olarak şu şekilde yazılabilir.

$$\text{logit P}(\text{V}_\text{i} = 1|\text{D}_\text{i}, \text{T}_\text{i}, \mathbf{x_i}) = \gamma_0 + \gamma_1\text{T}_\text{i} + \gamma_2\text{D}_\text{i}$$

Bu modelde D değişkeninin bulunup bulunmaması eksik ölçüm mekanizmasının MAR veya NMAR olmasını sağlamaktadır. Dolayısıyla bu yaklaşımla iki durum için de düzeltme yapılabilmektedir.

Bu yaklaşımda yapılan $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}')'$ vektörü için maksimum olabilirlilik tahmin edicisini bulmaktır. Bunun elde edilmesi için Beklenti-Maksimizasyon (EM : Expectation - Maximization) algoritması kullanılmıştır. EM algoritması başlangıç tahmini $\boldsymbol{\theta}^{(0)}$ ile başlar ve ardışık olarak beklenti ve maksimizasyon hesapları sonucunda $\boldsymbol{\theta}^{(t)}$ (t=1,2,3,…) olarak devam eder. Beklenti adımında verilen $\boldsymbol{\theta}^{(t)}$ tahmin edicisi ile tüm olabilirlilik fonksiyonunun logaritması olan $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$ ifadesi hesaplanır.

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \sum_{i=1}^{N} \left\{ V_i log P(V_i, T_i, D_i|\boldsymbol{x}_i; \boldsymbol{\theta}) + (1 - V_i) \times \sum_{d=0}^{1} P_{id}\left(\boldsymbol{\theta}^{(t)}\right) * log P(V_i, T_i, D_i = d|\boldsymbol{x}_i; \boldsymbol{\theta}) \right\}$$

Burada $P_{id}\left(\boldsymbol{\theta}^{(t)}\right) = P(D_i = d|V_i, T_i, \boldsymbol{x}_i; \boldsymbol{\theta})$'dır. Maksimizasyon adımında $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$ ise ifadesi $\boldsymbol{\theta}$'a göre maksimize edilerek $\boldsymbol{\theta}^{(t+1)}$ elde edilir. Uygun sayıda yapılan tekrar sonucu veya önceden belirlenen bitirme kriterine göre (ör: ardışık iki tahmin edici arasındaki farkın belirli bir değerin altında olması gibi) iterasyon sonlandırılır. Elde edilen tahmin edicilerle tanı testi için duyarlılık ve seçicilik değeri koşullu olasılıklar kullanılarak elde edilir.

$$Duyarlılık = P(T = 1|D = 1, \boldsymbol{x})$$
$$Seçicilik = P(T = 0|D = 0, \boldsymbol{x})$$

2.4.5 Çoklu İmputasyon Yaklaşımı ile Yanlılığın Düzeltilmesi

İmputasyon eksik verinin yerine uygun veriler atayarak eksik veri problemini çözen bir yaklaşımdır. Modeli doğru belirlenen imputasyon yaklaşımı kısa bir süre içinde sonuca ulaşacaktır. Ancak doğru kurgulanmayan bir imputasyon süreci tahmin edicilerin standart hatalarında büyük yanlılıklara ve hipotez test sonuçlarında büyük hatalara yol açabilecektir[34].

Çoklu imputasyon eksik ölçümler için aynı anda birden fazla değer (m>1) üreten bir Monte Carlo veri üretme tekniğidir. Genellikle aynı anda üretilen değer sayısı 3 ile

10 arasındadır. Çoklu imputasyon uygulamasında, birden fazla üretilen her bir tam veri standart metotlarca analiz edilir ve eksik ölçümler nedeniyle yanlı olarak tahmin edilen ölçümlere ait tahmin ediciler ve güven aralıkları tespiti için birleştirilir[32]. Çoklu imputasyon genellikle popülasyon temelli büyük çalışmalardan (örneğin, nüfus sayımı) elde edilen verilerin analizinde kullanılır. Çoklu imputasyon birçok karışık hesaplama içerdiğinden her durumda uygulanması kolay değildir. Ancak gelişen teknoloji ile hızlanan bilgisayar sistemleri ve bunlara uygun bilgisayar yazılımları sayesinde günümüzde sağlık bilimleri, sosyal bilimler gibi birçok alanda giderek artan bir ilgiyle tercih edilmektedir.

Çoklu imputasyonda izlenen matematiksel süreç şöyledir. Popülasyonda ölçülen bir değişken; Q ve bu değişkenin varyansı; U olarak verilsin. İlgilenen değişkenin tahmin edicisi de $\hat{Q}$ olarak gösterilsin. Ayrıca veri seti tam ölçümlü değişkenler ve eksik ölçümlü değişkenler olmak üzere iki parçaya ayrılsın: X; gözlenen tüm ek değişkenler, $Y = (Y_{göz}, Y_{eks})$; gözlenen ve eksik ölçümleri içeren değişken. Bu iki parça kullanılarak elde edilen sonuçlarda imputasyon sayısı (m) arttıkça üretilmiş veriden elde edilen tahmin değeri, gerçek tahmin değerine yaklaşır ve şöyle gösterilir.

$$\lim_{m\to\infty} E(\bar{Q}_m|X,Y) = \hat{Q} \qquad \lim_{m\to\infty} E(\bar{U}_m|X,Y) = U$$

Çoklu imputasyonda en önemli adım eksik olan sonuçları içeren değişkenin ardıl dağılımını (posterior distribution) belirlemektir. Çizelge 2.3'te görüldüğü gibi doğrulamanın kısmen yapıldığı bir veri setinin multinomial dağılımdan geldiği varsayılabilir. Bu varsayımdan hareketle ve multinomial dağılımın bir özelliği olan "koşullu multinomial dağılım da bir multinomial dağılımdır" özelliğini kullanarak, gözlenen veri bilgisi verildiğinde eksik ölçümlerin tahmini dağılımını elde etmek mümkündür. Dolayısıyla, x_{1j}^B ve x_{0j}^B eksik ölçümleri, θ_{ij} bireyin (i,j) gözüne düşme olasılığını $\theta_{+j} = \sum_i \theta_{ij}$ ve M(. , .) multinomial dağılımı göstermek üzere, x_{1j}^B ve x_{0j}^B için koşullu olasılık dağılımı aşağıdaki gibi olur.

$$(x_{1j}^B, x_{0j}^B)|Y_{göz}, \theta \sim M\left(x_{+j}^B, \left(\frac{\theta_{1j}}{\theta_{+j}}, \frac{\theta_{0j}}{\theta_{+j}}\right)\right), \quad j = 0{,}1$$

Bu ifadedeki multinomial parametreleri için öncel dağılım (Prior information) olarak Dirichlet dağılımı varsayılırsa, Bayes istatistiğinde de iyi bilinen, aşağıdaki sonuçlar elde edilir.

$$x|\theta \sim M(n,\theta) \qquad \theta \sim D(\alpha) \qquad \theta|Y \sim D(\alpha')$$

Burada $\alpha' = \alpha + x$ ve $D(\alpha)$ α parametreli Dirichlet dağılımını göstermektedir. Görüldüğü gibi x değişkeni θ değişkenine bağlı, θ değişkeni de x değişkenine bağlıdır. Dolayısıyla iteratif bir süreçte bu değişkenler belirlenebilir. Bu işlem için önce multinomial dağılımından bir x değişkeni θ değişkeni bilinmesi varsayımı altında üretilir. Daha sonra üretilen bu x değerleri kullanılarak Dirichlet, Beta ardıl dağılımından (posterior distribution) θ değerleri hesaplanır. Bu işlemler loglineer modelleme yapabilen herhangi birçoklu imputasyon yazılımı ile kolayca yapılabilir.

Eksik veriler tamamlama işlemi bittikten sonra, elde m tane tam veri seti bulunur. Bu veri setleri kullanılarak $\left(\hat{Q}^{(1)}, \hat{Q}^{(2)}, \dots, \hat{Q}^{(m)}\right)$ ve $\left(U^{(1)}, U^{(2)}, \dots, U^{(m)}\right)$ tahmin edicileri elde edilir. Bu tahmin edicilerden genel tahmin edicilerse şöyle elde edilir.

$\bar{Q} = \frac{1}{m}\sum Q^{(i)}$ ve bunun varyansı $T = \bar{U} + \frac{1}{m+1}B$

Bu varyanstaki $\bar{U} = \frac{1}{m}\sum U^{i}$ ifadesi tüm veri varyans tahmini ve $\frac{1}{m+1}B$ ifadesi ise imputasyonlar sonucu oluşan ek varyansı göstermektedir. Ek varyansın içindeki B terimi imputasyonlar arası varyansı göstermekte ve $B = \frac{1}{m-1}\sum_{i=1}^{m}\left(Q^{i} - \bar{Q}\right)^2$ ile tahmin edilmektedir[34].

2.4.6 Yapay Sinir Ağları Yaklaşımı ile Yanlılığın Düzeltilmesi

Yapay Sinir Ağları (YSA) yaklaşımı, çok yönlü analiz yapabilen, paralel dağıtılmış işlemci olarak da bilinen işlemsel bir metodolojidir. Bu işlemsel metot bazı önemli özelliklere sahiptir. Bunlardan bazıları: öğrenme, genelleştirme, kümeleme ve veriyi düzenlemedir. YSA işlemleri, eş zamanlı yani paralel olarak yapar[35]. YSA birçok

bilim alanında çalışan araştırmacılar tarafından gruplama, sonucu tahmin etme, ileriye yönelik tahminde bulunma, performansı en üst seviyeye çıkarma ve kümeleme gibi birçok amaç için kullanılmaktadır[36].

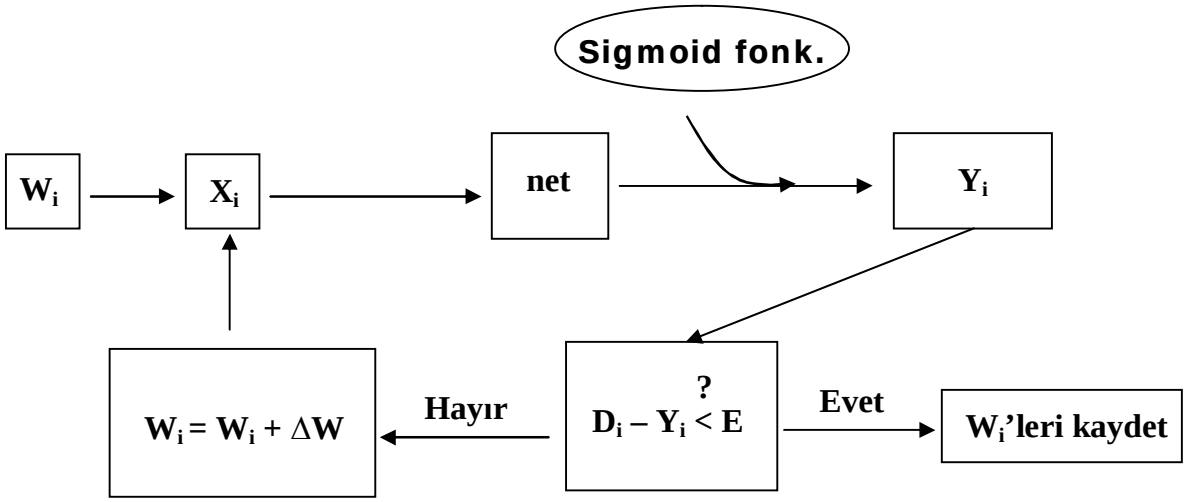
YSA doğrusal olmayan bir modellemedir. Bu özelliği ona değişkenler arasındaki karmaşık ilişkileri çözümleyebilme becerisi sağlamıştır. Ayrıca LR'de olduğu gibi YSA'da da değişkenlerin dağılımları hakkında herhangi bir varsayım yoktur.

YSA, iki aşamalı bir modellemedir: Eğitim aşaması ve test aşaması. Yoğun işlemlerin gerçekleştiği aşama öğrenme kısmındadır. Burada eş zamanlı birçok işlem yapılmakta ve sonuca gidilmektedir. Paralel dağıtılmış işlemci ismini de bu kısımda gerçekleşen işlemlerden dolayı almaktadır. Test aşaması daha basit ve kısa süreli olacaktır.

YSA modelinde yapılan öğrenme işlemi, ağırlıkların bulunması üzerinedir. Bunun için, tüm YSA metotları için benzer süreç işlemektedir. Girdi ve çıktı katmanından oluşan bir ağda bu süreç şöyle işleyecektir: Girdi katmanından alınan değerler başlangıçta rastgele atanmış olan ağırlıklarla çarpılarak toplanmaktadır. Elde edilen bu değere net denmektedir. YSA yapısına göre belirlenen aktivasyon fonksiyonu (lojistik fonksiyonu ve tanh fonksiyonu sıkça kullanılanlardandır) net değerini çıktı değerine yakın bir değere dönüştürecektir. Bu noktadan sonra ağırlıkların düzeltilmesi devreye girmektedir. Düzeltme için, asıl çıktı değerinden hesaplanan çıktı değerinin çıkarılması ile elde edilen hata terimi, öğrenme katsayısı ile çarpılır ve bu değer her bir ağırlıktan çıkarılır. Öğrenme işlemi önceden belirlenen devir sayısının sonuna kadar veya durdurma kriterinin gerçekleşmesi durumuna kadar devam edecektir. Bu anlatılanın şematik gösterimi Şekil 2.3'de verildiği gibidir.

Günümüzde kullanımı en yaygın olan çok katmanlı ileri beslemeli sinir ağı modelini, 1969 yılında Minsky ve Papert adlı araştırmacılar geliştirmiş ve adını çok katmanlı perseptron koymuşlardır. Bu model perseptronun, girdi ve çıktı katmanı arasında yer alan bir veya birden fazla gizli katman içeren geliştirilmiş bir halidir. Bu gelişmiş hali modele her türlü mantıksal işlemi çözebilme yeteneği sağlamıştır[37].

Genellikle standart bir L katmanlı network, bir girdi katmanı, (L-2) gizli katman ve bir çıktı katmanından oluşur. Ayrıca, aynı katman içindeki düğümler arasında olmayıp, katmanlardaki düğümler arasında olan (tüm veya kısmi) bağlantılar vardır.



Şekil 2.3 Yapay Sinir Ağlarının işleyiş şeması

Çok katmanlı perseptrondaki öğrenme kuralında, ağırlıklar elde edilen hata terimini minimum yapacak şekilde düzeltilir. Burada katman sayısına paralel olarak birden fazla hata terimi oluşacaktır. Bu yüzden burada sinir ağlarının hatasını belirleyebilmek için bir hata fonksiyonu kullanılır. Genellikle bu fonksiyon aynı zamanda doğrusal regresyonda da kullanılan hata kareleri toplamı fonksiyonudur. Çok katmanlı perseptronda bu fonksiyonu minimize etmek için çeşitli öğrenme kuralları kullanılır. Bunlardan en meşhur olanı geriye yayınım algoritmasıdır (Back propagation algorithm)[37].

2.4.6.1 Geriye Yayınım Algoritması

1986 yılında E. Rumelhart, G.E. Hinton ve R.J. Williams, network modelleri içinde en güçlü ve en popüler olan geriye yayınım algoritmalı sinir ağını geliştirmişlerdir[38]. Bu algoritma hata kareleri toplamı fonksiyonunu minimize etmek için eğim iniş metodunu (gradient descent method) kullanmaktadır ve bu metot şu şekilde işlemektedir:

1. Ağırlıklar küçük değerler olacak şekilde atanır.
2. Rastgele bir **x** vektörü girdi için belirlenir.
3. Bu vektör ile ağırlıklar çarpılarak toplanır.
4. Çıktı katmanında d_i^L değeri şu şekilde hesaplanır:

$$d_i^L = g'(h_i^L)[d_i^L - y_i^L]$$

buradaki h_i^L değeri L katmanındaki i düğümünün net girdisini ve g' ise aktivasyon fonksiyonu g'nin türevini göstermektedir.

5. Geriye doğru gelerek deltalar teker teker hesaplanır.

$$d_i^l = g'(h_i^l)\sum_j w_{ij}^{l+1} d_j^{l+1}$$

burada *l* değerleri (L – 1)'den başlayarak 1'e doğru olan değerlerdir.

6. Ağırlıklar şu şekilde düzeltilir. $\Delta w_{ji}^l = h d_i^l y_j^{l-1}$

7. 2.basamağa dönülerek hatanın belirli bir değer altına inmesine veya belirlenen devir sayısının sonuna kadar işlem tekrar edilir.

2.4.6.2 Önerilen yaklaşım

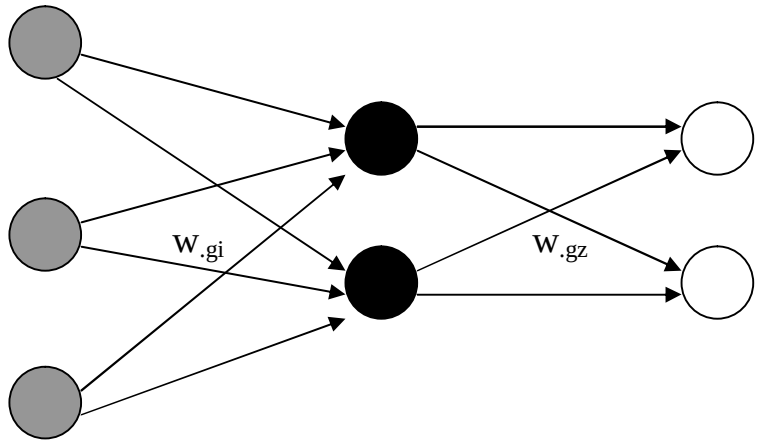
Doğrulama yanlılığının düzeltilmesinde Regresyon, İmputasyon gibi eksik ölçümün modellenmesi ve tahmin edilmesi yöntemlerine başvurulduğu bu tezde belirtilmiştir. Bu fikirden hareketle, literatürde birçok alanda Lojistik Regresyona alternatif olarak önerilen Yapay Sinir Ağları yaklaşımı da doğrulama yanlılığının düzeltilmesinde kullanılabilir.

Kosinski ve Barnhart'ın doğrulama yanlılığını düzeltmede önerdikleri Lojistik Regresyon yaklaşımı5 temel alınarak, bu tezde de Yapay Sinir Ağları yaklaşımı doğrulama yanlılığını düzeltmede kullanılmıştır. Kosinski ve Barnhart yaklaşımlarında olabilirlilik fonksiyonunun bileşenlerini modellemişlerdir. Benzer şekilde Yapay Sinir Ağları yaklaşımında da bu bileşenler modellenip, ağırlıklar elde edilmesi yöntemiyle de düzeltilmiş doğruluk ölçütleri tahmin edilmiştir. Buna göre hastalık, tanı testi ve eksik veri mekanizma bileşenleri için ayrı ayrı sinir ağları oluşturulmuştur. Oluşturulan modellerden ağırlıklar elde edilerek doğruluk ölçütleri hesaplanmıştır.

2.4.6.2.1 Yapay Sinir Ağı Yaklaşımı için Bileşenler

Kosinski ve Barnhart doğrulama yanlılığını düzeltme amacıyla önerdikleri yaklaşımlarında[5], veri setinde doğrulama yapılmamış bireylerin ölçümlerini kopyalayıp veri setinin sonuna ekleyerek veri setinin büyüklüğünü büyütmüşlerdir. Dolayısıyla veri

setindeki doğrulama yapılmamış birey sayısı iki katına çıkmıştır. Bu bireylerin yarısının gerçek durumları pozitif, diğer yarısının da (eklenen bireyler) gerçek durumları negatif olarak alınmıştır. Sonuç olarak elde edilen yeni veri setinde tüm bireylerin sonuçları tam olarak bulunmaktadır. Bu veri seti kullanılarak, önce gerçek durum için bir lojistik regresyon modeli, sonra önerilen yeni test için bir lojistik regresyon modeli ve son olarak doğrulama yapılıp yapılmadığı gösteren değişken için bir lojistik regresyon modeli kurulmuş ve her birinden ayrı ayrı parametre tahmin edicileri elde edilmiştir. Bu tahmin edicilerin bir kombinasyonu oluşturularak veri setindeki her bir birey için bir ağırlık elde edilmiş ve bu ağırlıklarla 3 lojistik regresyon modeli tekrar oluşturulmuştur.



Şekil 2.4 İleri beslemeli gizli katmanlı yapay sinir ağı

Kosinski ve Barnhart'ın yaklaşımına benzer şekilde Yapay Sinir Ağlarında da 3 farklı bileşen için 3 farklı ağ oluşturulmuştur:

- 1. Sinir ağında gerçek durum ek değişkenler kullanılarak modellendi.
  - Gerçek Durum Bileşeni (D): $y_{k=}\varphi(\sum w_{Dh}\varphi(\boldsymbol{z'}_{0i}\boldsymbol{w_{Di}}))$
- 2. Sinir ağında yeni tanı testi sonucu ek değişkenler ve gerçek durum bilgisi ile modellendi.
  - Tanı Testi Bileşeni (T): $y_{k=}\varphi(\sum w_{Th}\varphi(z'_{0i}w_{Ti}))$
- 3. Sinir ağında ise doğrulama bilgisi ek değişkenler, gerçek durum ve tanı testi sonucu kullanılarak modellendi.
  - Eksik Ölçüm Mekanizması Bileşeni (M): $y_{k=}\varphi(\sum w_{Mh}\varphi(z'_{0i}w_{Mi}))$

Bu üç bileşende de φ fonksiyonu sigmoid (logistic) fonksiyonu olarak alınmıştır. Burada verilen $w_{.i}$ ağırlıkları girdi katmanı ile gizli katman arasındaki, $w_{.h}$ ağırlıkları ise gizli katman ile çıktı katmanı arasındaki ağ ağırlıklarını göstermektedir. Doğrulama

yanlılığını düzeltmek için önerilen Yapay Sinir Ağları yaklaşımında, Kosinski ve Barnhart'ın çalışmalarında elde edilen regresyon katsayıları yerine $w_{.i}$ ağ ağırlıkları kullanılmıştır.

Kosinski ve Barnhart'ın yaklaşımlarındaki uygulamaya bağlı kalınarak veri seti yeniden oluşturulmuştur. Buna göre veri setinde N satır olduğu ve bunun içinde ilk N-U satırın doğrulama yapılmış bireyler ve son U satırın ise doğrulama yapılmamış bireyler olduğu kabul edilerek, son U satırın tüm verileri kopyalanarak veriye ilave edilmiştir. Sondaki 2U satırdaki bireylerin doğrulamaları yapılmadığı için, ilk U satırın gerçek durum bilgisi 1 yani hasta, son U satırın gerçek durum bilgisi 0 yani sağlıklı olarak atanmıştır. Son olarak veri setine yeni bir değişken –doğrulama değişkeni- eklenmiş ve bu değişkenin ilk N-U satırı 1 yani doğrulaması yapılmış, son 2U satırı 0 yani doğrulaması yapılmamış olacak şekilde düzenlenmiştir. (Çizelge 2.4)

Çizelge 2.4 Kosinski ve Barnhart Yaklaşımına göre düzenlenmiş veri seti

k	Y_0	Y_1	Y_2	Ek Değişkenler
1	d_1	t_1	1	$\mathbf{x}'_1$
2	d_2	t_2	1	$\mathbf{x}'_2$
.	.	.	.	.
.	.	.	.	.
N-U	d_{N-U}	t_{N-U}	1	$\mathbf{x}'_{N-U}$
N-U+1	0	t_{N-U+1}	0	$\mathbf{x}'_{N-U+1}$
.	.	.	.	.
.	.	.	.	.
N	0	t_N	0	$\mathbf{x}'_N$
N+1	1	t_{N-U+1}	0	$\mathbf{x}'_{N-U+1}$
.	.	.	.	.
.	.	.	.	.
N+U	1	t_N	0	$\mathbf{x}'_N$

Veri seti düzenlendikten sonra her bir satırdaki gerçek durum, test sonucu ve doğrulama bilgisi verilerini y_{ij} ile göstererek ve ilk sütunu 1 olarak atanan ve her bir bireyin ek değişken vektörünü satır olarak kabul eden dizayn matrisini oluşturarak aşağıda verilen hesaplamalar yapılmıştır.

$$p_k = \prod_{m=0}^{2} \{p_{mk}\}^{y_{mk}} \{1 - p_{mk}\}^{1-y_{mk}}$$

burada $p_{0k} = \{1 + \exp(-v'_{Dk} w_{Di})\}^{-1}$, $p_{1k} = \{1 + \exp(-v'_{Tk} w_{Ti})\}^{-1}$, $p_{2k} = \{1 + \exp(-v'_{Mk} w_{Mi})\}^{-1}$. Elde edilen bu p_k değeri kullanılarak her bir birey için bir birey ağırlığı, w_k şu şekilde hesaplanmıştır.

$$w_k = \begin{cases} 1 & \text{for } k = 1, \dots, N - U \\ p_k / \{p_k + p_{k+U}\} & \text{for } k = N - U + 1, \dots, N \\ 1 - w_{k-U} & \text{for } k = N + 1, \dots, N + U \end{cases}$$

Hesaplanan bireysel ağırlık değerleri bir sonraki iterasyonda oluşturulan modellerde kullanılarak yeni ağ ağırlıkları elde edilmiştir. Bu işlem elde edilen ağ ağırlıklarındaki değişim belirlenen bir hata değerinin altına düşene kadar veya önceden belirlenen iterasyon sayısına ulaşılana kadar tekrarlanmıştır.

2.5 Kesin olmayan altın standart (KOAS) yanlılığı

2.5.1 KOAS yanlılığı nedir?

Kesin tanıyı koymada kullanılan birçok altın standart test uygulaması kolay olmayan ve/veya pahalı testlerdir. Örneğin kalp hastalıkları ile ilgili birçok konuda kesin tanıyı belirlemede kullanılan Anjiyografi uygulaması, girişimsel bir uygulama olduğundan hasta için riskli bir uygulamadır. Diğer bir örnekte Radyolojik görüntüleme için verilebilir. Birçok hastalığın kesin tanısını koymada kullanılan Radyolojik görüntüleme yöntemleri oldukça pahalı yöntemlerdir. Altın standart olarak kullanılan testin uygulanması ile ilgili karşılaşılabilecek bu sıkıntılar nedeniyle birçok hastalık için altın standart yerine kullanılan ve altın standart testine göre daha düşük sınıflama başarısına sahip olan Kesin Olmayan Altın Standart (KOAS) testi önerilmiştir.

Bir hastalığın tanısını koymada yeni bir tanı testi önerilirse, bu testin mutlaka sınıflama başarısı elde edilmelidir. Daha önceden de belirtildiği gibi bu başarının ölçülebilmesi için tanı testinin uygulandığı grubun kesin tanı sonuçları (altın standart sonucu) biliniyor olması gerekir. Hastalığın kesin tanısını koymada KOAS testine

başvurulursa, yeni geliştirilen tanı testinin sınıflama başarısı yanlı olarak tahmin edilir. Bu yanlılığa kesin olmayan altın standart yanlılığı denir[11].

KOAS yanlılığının düzeltilmesinde tanı testinin yapısına göre (kategorik veya sürekli) farklı düzeltmeler önerilmiştir[12,13,14,15,16,17]. Bu tezde, yeni tanı testinin sıralı veya oran ölçeği olması durumunda önerilen yaklaşımlar incelenmiştir.

Sıralı veya sürekli tanı testlerinin doğruluk ölçütlerinin belirlenmesinde ROC eğrileri kullanılmaktadır. ROC eğrileri, bireylerin gerçek durum bilgisi varlığında tanı testinin her bir değerini kesim noktası alarak elde edilen duyarlılık ve 1 – seçicilik değerlerini koordinat noktası kabul eden eğrilerdir. Eğri altında kalan alan tanı testinin sınıflama başarısı hakkında bilgi verir. Gerçek durum bilinmediğinde elde edilecek olan doğruluk ölçütleri yanlı olacaktır. Bu sebeple kesin olmayan altın standardın kullanıldığı çalışmalarda tanı testinin doğruluğu düzeltilerek elde edilir. Literatürde önerilen bazı düzeltmeler şöyledir:

2.5.2 Tanı Testinin İkili (Binary) Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Önerilen Yaklaşımlar

Tanı testinin ikili ölçüm olması durumunda literatürde KOAS yanlılığını düzeltmede birden fazla yaklaşım önerilmiştir. Bu yaklaşımlardan biri Tutarsızlık Çözücü (Discrepant Resolution) yaklaşımıdır. Bu yaklaşımda KOAS ile tanı testi arasında bir uyumsuzluk oluştuğunda, örneğin tanı testi pozitif, KOAS negatif veya tam tersi, üçüncü bir çözücü teste (resolver test) başvurulur. Bu çözücü test sonucu olması gereken sonuç olarak kabul edilerek tanı testinin doğruluk ölçütleri hesaplanır. Bu hesaplama örnekle şu şekilde izah edilebilir.

Çizelge 2.5'te bir tanı testi ve bir KOAS sonucu a, b, c ve d olarak verilsin. Burada da görülebileceği gibi iki test arasında tutarsızlığa yol açan b+c sonuç bulunmaktadır. Bu yaklaşım uygulandığında, yani çözücü olarak kabul edilen üçüncü bir test sonucunda b* ve c* adet sonuç tekrar sınıflanarak çizelgede görülen Çözücü sonrası durumu oluşturulabilir.

Çizelge 2.5 Tutarsızlık Çözücü yaklaşımı

	Tutarsızlık durumu		Çözücü sonrası	
	KOAS = 0	KOAS = 1	KOAS = 0	KOAS = 1
Tanı testi = 0	a	b	a+b*	b-b*
Tanı testi = 1	c	d	c-c*	d+c*

Tutarsızlık Çözücü literatürde sıklıkla başvurulan bir yaklaşımdır[13]. Ancak genellikle tercih edildiği durum KOAS testinin %100 seçici (az duyarlı) veya %100 duyarlı (az seçici) olduğu durumdur. Tutarsızlık Çözücü yaklaşımı, KOAS'ın düşük sınıfladığı durumda tanı testi ile tutarsızlık varsa uygulanmaktadır. Örneğin, bel soğukluğuna neden olan bakterinin (Chlamydia trachomatis) saptanmasında kültür testi yüksek seçiciliğe (nerdeyse %100) fakat daha düşük duyarlılığa sahip bir testtir. Bu bakterinin saptanmasında DNA büyütme yöntemlerinden (DNA amplification tests) Ligaz enzimi zincir reaksiyonu (Ligase Chain Reaction-LCR) kullanılacak olursa

Çizelge 2.5'te verilen ilk durum oluşabilir. Bu durumda kullanılacak çözücü test ise polimeraz zincir reaksiyonu (PCR) testi olacaktır.

Bu yaklaşım uygulamadaki kolaylığı nedeniyle popüler olsa da neden olduğu bazı yanlılıklardan dolayı verdiği sonuçlar ile ilgili bazı şüpheler vardır. Yaklaşıma başvurulan durumların sadece yanlış pozitif veya yanlış negatif durumlarında olması tartışılan bir konudur. Çözücü olarak kullanılan test kesin bir altın standart testi olsa dahi belirtilen durum hala yanlılık doğurabilmektedir. Üstelik ikinci bir sorun da çözücü olarak kullanılan testin çoğu durumda kesin bir altın standart test olmamasıdır. Bu da çözücü test sonucuna olan güveni azaltmaktadır. Meier ilk soruna bir çözüm getirmek için bu yaklaşımın uygulamasını biraz değiştirerek, sadece pozitif veya yanlış negatif sonuçlarına değil bunlara ilave olarak diğer durumlardan rastgele olarak seçilmiş bir alt gruba da uygulamıştır[39]. Ancak bu yaklaşım da yeterli istatistik dayanağı bulunmadığından uygulamada sık tercih edilen bir yaklaşım olamamıştır.

Her ne kadar Tutarsızlık Çözücü yaklaşımı yanlı bir yaklaşım olsa da bu yaklaşım "tek bir referans test yerine birden fazla referans test sonucu kombinasyonu" fikrinin ortaya çıkmasını sağlamıştır. Bu fikirden hareketle, Alonzo ve Pepe 1999 yılında tüm testlerin (değerlendirme altında olan test dışındaki) sonuçlarının kombinasyonuna Bileşik Referans Standardı (Composite Reference Standard) testi adını

vermiştir[40]. Bu fikrin dayandığı en önemli nokta, gerçek durumun belirlenmesinde tek bir KOAS testi kullanmak yerine, oluşturulan bu Bileşik Referansın daha etkili olacağıdır. Bel soğukluğu örneğinde duyarlı bir bileşik referans testi şöyle tanımlanabilir: Eğer kültür testi veya PCR testi pozitif ise Bileşik Referans test pozitif, eğer kültür ve PCR aynı anda negatif ise Bileşik Referans testi negatiftir.

Bileşik Referans test uygulamasının sağladığı en büyük fayda, hastalık tanımın oluşturulmasıdır. Bel soğukluğu örneğinde olduğu gibi, hastalığın varlığı veya yokluğu kültür ve PCR testleri sonucunda kesin olarak belirlenebilmektedir. Tutarsızlık Çözücüde bu durum değerlendirme altında olan test sonucuyla belirlendiği için elde edilen sonuçlar yanlı olacaktır[41].

2.5.3 Tanı Testinin Sıralı Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Parametrik Olmayan ROC Eğrisi Yaklaşımı

2005 yılında Zhou ve ark. ikiden fazla sıralı ölçekteki test için KOAS yanlılığını düzeltmede kullanılacak bir yaklaşım önermiştir[16]. Bu yaklaşımda ROC eğrisi altındaki alanın elde edilişi parametrik olmayan bir yöntemle belirlenmiştir.

Zhou ve ark. önerdiği yaklaşım şöyledir: Gerçek durumu bilinmeyen N hastanın 1'den J'ye kadar değer alan sıralı K adet test sonucu olduğu varsayılsın. Bu durumda belirli bir bireyin K adet test sonucu T_1, T_2, …, T_K ve bilinmeyen gerçek durumu ise D ile gösterilsin. Her bir test 1'den J'ye kadar sıralı değerler aldığından parametrik olmayan kesikli ROC eğrisi tahmini yapılabilir. Burada tüm değerler sırayla kesim noktası olarak alınarak J + 1 tane Doğru pozitiflik oranı (DPO) ve Yanlış pozitiflik oranı (YPO) ikililerinden oluşan kesikli bir ROC eğrisi elde edilir. Yani,

$$DPO_k(j) = P(T_k \geq j \mid D = 1) \qquad YPO_k(j) = P(T_k \geq j \mid D = 0)$$

Burada $DPO_k(1) = 1 = YPO_k(1)$ ve $DPO_k(J+1) = 0 = YPO_k(J+1)$'dir.

Bu ikililerin bir doğru ile birleştirilmesi sonucu ROC eğrisinin parametrik olmayan tahmini yapılmış olur. Burada oluşan her ardışık ikili arasında kalan

yamukların alanları toplamı ROC eğrisi altında kalan alanın parametrik olmayan tahmini olur. Yani,

$$A_k = \sum_{j=1}^{J-1}\left[P(T_k = j \mid D=0) \sum_{i=j+1}^{J} P(T_k = i \mid D=1) \right] + \frac{1}{2}\sum_{j=1}^{J} P(T_k = j \mid D=0) P(T_k = j \mid D=1)$$

Bu formülde $f_{0kj} = P(T_k = j \mid D=0)$ ve $f_{1kj} = P(T_k = j \mid D=1)$ olarak tanımlanırsa, ifade daha kısa hale gelir.

$$A_k = \sum_{j=1}^{J-1}\left[f_{0kj} \sum_{i=j+1}^{J} f_{1ki} \right] + \frac{1}{2}\sum_{j=1}^{J} f_{0kj} f_{1kj}$$

Burada f_{dkj} ifadesini içerecek bir olabilirlilik fonksiyonu yazılabilir ve bunun için ML tahminleri bulunabilirse ROC eğrisinin alanı tahmin edilmiş olur. Bu amaçla 0 ve 1 değerlerini alan yeni bir değişken tanımlansın.

$$y_{ikj} = \begin{cases} 1 & \text{i. bireyin k. testinin j değerine eşit olması durumunda} \\ 0 & \text{diğer durumlarda} \end{cases}$$

Bu durumda i. birey için K x J'lik test skor vektörü şöyle yazılabilir.

$$\mathbf{y}_i = (y_{i11},..., y_{i1J},...., y_{iK1},..., y_{iKJ})'$$

D_i i. bireyin gerçek durumunu göstermek üzere, D_i = d verilmişken i. bireyin y_i test skorunun koşullu olasılığı K test için koşullu bağımsızlık varsayımı altında şöyle tanımlanabilir.

$$g_d(\mathbf{y}_i) = P(\mathbf{y}_i \mid D_i = d) = \prod_{k=1}^{K}\prod_{j=1}^{J}\left[f_{dkj}\right]^{y_{ikj}}$$

Bilinmeyen gerçek durumu da $p_d = P(D = d)$ parametreli Bernoulli dağılımından gelen bir değişken olarak tanımlanırsa, N birey için bileşke olabilirlilik fonksiyonu şu şekilde yazılır.

$$l(p_1, f_0, f_1) = \sum_{i=}^{N} \log\left[p_0 g_0(\mathbf{y}_i) + p_1 g_1(\mathbf{y}_i)\right]$$

Burada $p_0 = 1 - p_1$ ve $f_d = (f_{d11},...,f_{d1J},.....,f_{dK1},...,f_{dKJ})'$ dir. Bu noktadan sonra gerçek durum bilgisi eksik ölçüm olarak kabul edilerek EM algoritması ile tahmin yapılabilir. Buna göre, $q = (p_1, f_0, f_1)$ olmak üzere, t. iterasyon sonucu

$$E\left(l(q) \mid \mathbf{y}, q = q^t\right) = \sum_{i=}^{N}\sum_{d=0}^{1} q_{id}^t \log g_d(\mathbf{y}_i)$$

$$g_d^t(\mathbf{y}_i) = \prod_{k=1}^{K}\prod_{j=1}^{J}\left[f_{dkj}^t\right]^{y_{ikj}}, \quad q_{id}^t = \frac{p_d^t g_d^t(\mathbf{y}_i)}{p_0^t g_0^t(\mathbf{y}_i) + p_1^t g_1^t(\mathbf{y}_i)}$$

elde edilir. t +1. iterasyonda ise,

$$p_1^{t+1} = \frac{1}{N}\sum_{i=1}^{N} q_{i1}^t \qquad f_{dkj}^{t+1} = \frac{\sum_{i=1}^{N} q_{id}^t y_{ikj}}{\sum_{i=1}^{N} q_{id}^t}$$

değerleri elde edilir. Belirli bir iterasyon sonucu hatanın belirli bir değerin altında olması durumunda ilgili katsayılar elde edilir ve K test için ROC eğrisi altında kalan alan hesaplanır[16].

2.5.4 Tanı Testinin Sürekli Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Bayes Yaklaşımı

Ek değişkenler varlığında sürekli bir testin kesin olmayan altın standart teste göre doğruluk ölçütleri Bayes yaklaşımı ile hesaplanabilir. Bayes yaklaşımında hasta ve sağlıklı bireylerin ek değişkenlerce modellenmesi sonucu elde edilen regresyon katsayılarının ve diğer bazı parametrelerin başlangıç dağılımları varsayılır ve iterasyonlar sonucu elde edilen yakınsamış katsayı değerleri ROC eğrisi için oluşturulan modelde kullanılır[17].

Her bir bireye ait sürekli test sonucu T_i ve kesin olmayan altın standart test sonucu da R_i olarak kabul edilsin. Ayrıca bilinmeyen gerçek durum bilgisi de D_i ve K tane ek değişken bilgisi de Z_{i1}, …, Z_{iK} ile gösterilsin. Bu durumda ek değişkenlere göre gerçek durum bilgisi için lojistik regresyon uygulanırsa;

$$\log \frac{P(D_i = 1)}{1 - P(D_i = 1)} = d_0 + d_1 Z_{i1} + ... + d_K Z_{iK}$$

$\boldsymbol{\delta} = (d_0, d_1, ..., d_K)'$ regresyon katsayıları elde edilir. Burada gerçek durum bilinmediğinden katsayılar da bilinemeyecektir. Bu sebeple bu katsayılar belirli bir başlangıç dağılımı bilgisi ile belirli bir iterasyon sayısı sonucunda elde edilir.

Hasta ve sağlıklı bireylerin sürekli test sonuçları üzerine ek değişkenlerin etkisi incelenmek istendiğinde aşağıdaki regresyon denklemleri elde edilir.

$$E(T_i \mid D_i = 1) = b_{H0} + b_{H1} Z_{i1} + ... + b_{HK} Z_{iK}$$
$$E(T_i \mid D_i = 0) = b_{S0} + b_{S1} Z_{i1} + ... + b_{SK} Z_{iK}$$

Buradaki $\boldsymbol{\beta}_H = (b_{H0}, b_{H1}, ..., b_{HK})'$ ve $\boldsymbol{\beta}_S = (b_{S0}, b_{S1}, ..., b_{SK})'$ vektörleri hasta ve sağlıklı bireyler için bilinmeyen regresyon katsayılarıdır. δ'ya benzer şekilde β değerleri de başlangıç dağılımları varsayıldıktan sonra iterasyon sonucunda elde edilir.

Sürekli test sonuçları hasta ve sağlıklı bireylerde farklı varyansta dağılım gösterebilir. Burada varyans hasta grupta $Var(T_i \mid D_i = 1) = s_H^2$ ve sağlıklı grupta $Var(T_i \mid D_i = 0) = s_S^2$ olarak varsayılsın.

Son olarak referans test olarak kabul edilen kesin olmayan altın standart testi hasta ve sağlıklı bireylerde 1 – α_H ve α_S parametreli Bernoulli dağılımı göstersin. Burada α_H ve α_S parametreleri referans testi için sırasıyla yanlış negatif oranı ile yanlış pozitif oranlarıdır. Bu değerler de başlangıçta bilinmeyip, geldikleri dağılımlar varsayılacaktır.

Sonuç olarak, sürekli test, referans testi ve gerçek durum için bileşke olabilirlilik fonksiyonu şu şekilde yazılabilir.

$$
\begin{aligned}
&L\left(T_i, R_i, D_i \mid \mathbf{Z}_i, \boldsymbol{\beta}_H, \boldsymbol{\beta}_S, s_H^2, s_S^2, a_H, a_S, \boldsymbol{\delta}\right) = \\
&\left\{\frac{1}{\sqrt{2p}\, s_H} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_H' \cdot \mathbf{Z}_i\right)^2}{2 s_H^2}\right]\right\}^{D_i} \\
&\left\{\frac{1}{\sqrt{2p}\, s_S} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_S' \cdot \mathbf{Z}_i\right)^2}{2 s_S^2}\right]\right\}^{1-D_i} \\
&\left[a_H^{1-R_i}\left(1-a_H\right)^{R_i}\right]^{D_i}\left[a_S^{R_i}\left(1-a_S\right)^{1-R_i}\right]^{1-D_i} \\
&\left[\frac{\exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)}{1+\exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)}\right]^{D_i}\left[\frac{1}{1+\exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)}\right]^{1-D_i}
\end{aligned}
$$

Olabilirlilik fonksiyonu EM algoritması ile çözümlenebilir. Ancak veri seti büyük olduğunda bu çözümleme çok fazla işlem yapılmasını gerektirecektir. Bunun yerine Bayes yaklaşımı kullanılabilir. Burada β regresyon katsayıları çok değişkenli normal dağılımdan (her bir beta değeri N(0, 1000) dağılımlı) gelen ve

$$
p\left(\boldsymbol{\beta}_H \mid \bullet\right) \propto \prod_{i=1}^{N}\left\{\frac{1}{\sqrt{2p}\, s_H} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_H' \cdot \mathbf{Z}_i\right)^2}{2 s_H^2}\right]\right\}^{D_i} \exp\left(\frac{-\boldsymbol{\beta}_H' \cdot \boldsymbol{\beta}_H}{2x1000}\right)
$$

$$
p\left(\boldsymbol{\beta}_S \mid \bullet\right) \propto \prod_{i=1}^{N}\left\{\frac{1}{\sqrt{2p}\, s_S} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_S' \cdot \mathbf{Z}_i\right)^2}{2 s_S^2}\right]\right\}^{1-D_i} \exp\left(\frac{-\boldsymbol{\beta}_S' \cdot \boldsymbol{\beta}_S}{2x1000}\right)
$$

ifadeleri ile düzeltilen parametreler olur. Her biri Γ(0.001, 0.001) Gamma dağılımından gelen $1/s_H^2$ ve $1/s_S^2$ varyansları ise aşağıdaki ifade ile düzenlenir.

$$p\left(1/s_H^2 \mid \bullet\right) \propto \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2p}\, s_H} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_H' \cdot \mathbf{Z}_i\right)^2}{2 s_H^2} \right] \right\}^{D_i} \left(\frac{1}{s_H^2} \right)^{0.001-1} \exp\left(\frac{-0.001}{s_H^2} \right)$$

$$p\left(1/s_S^2 \mid \bullet\right) \propto \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2p}\, s_S} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_S' \cdot \mathbf{Z}_i\right)^2}{2 s_S^2} \right] \right\}^{D_i} \left(\frac{1}{s_S^2} \right)^{0.001-1} \exp\left(\frac{-0.001}{s_S^2} \right)$$

Yanlış sınıflama olasılıkları α_H ve α_S'nın her biri β(0.5, 0.5) beta dağılımından gelir ve şu şekilde düzenlenir.

$$p\left(a_H \mid \bullet\right) = \left[\prod_{i=1}^{N} a_H^{(1-R_i)D_i} \left(1 - a_H\right)^{R_i D_i} \right] a_H^{0.5-1} \left(1 - a_H\right)^{0.5-1}$$

$$p\left(a_S \mid \bullet\right) = \left[\prod_{i=1}^{N} a_S^{R_i(1-D_i)} \left(1 - a_S\right)^{(1-R_i)D_i} \right] a_S^{0.5-1} \left(1 - a_S\right)^{0.5-1}$$

Hastalık prevelansı ile ilgili regresyon katsayıları δ'nın her biri N(0, 1000) dağılımlıdır ve aşağıdaki ifade ile düzenlenir.

$$p\left(\boldsymbol{\delta} \mid \bullet\right) = \prod_{i=1}^{N} \left[\frac{\exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)}{1 + \exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)} \right]^{D_i} \left[\frac{1}{1 + \exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)} \right]^{1-D_i} \exp\left(\frac{-\boldsymbol{\delta}' \cdot \boldsymbol{\delta}}{2x1000} \right)$$

Bu düzenlemeler yapıldıktan sonra gerçek durum bilgisi D aşağıdaki ifade ile düzenlenir.

$$p\left(D_i \mid \bullet\right) = \left\{ \frac{1}{\sqrt{2p}\, s_H} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_H' \cdot \mathbf{Z}_i\right)^2}{2 s_H^2} \right] a_H^{1-R_i} \left(1 - a_H\right)^{R_i} \frac{\exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)}{1 + \exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)} \right\}^{D_i}$$
$$\times \left\{ \frac{1}{\sqrt{2p}\, s_S} \exp\left[\frac{-\left(T_i - \boldsymbol{\beta}_S' \cdot \mathbf{Z}_i\right)^2}{2 s_S^2} \right] a_S^{R_i} \left(1 - a_S\right)^{1-R_i} \frac{1}{1 + \exp\left(\boldsymbol{\delta}' \cdot \mathbf{Z}_i\right)} \right\}^{1-D_i}$$

Tüm düzenleme işlemleri Gibbs Sampler kullanılarak ve MCMC prosedürü ile yapılır. Yukarıda verilen başlangıç dağılımları 1961 yılında Jeffreys tarafından önerilen ve bilgi içermeyen (noninformative) başlangıç dağılımlarıdır[42]. Düzenleme süreci iterasyon sayısı belirlenen rakama ulaştığında veya durdurma kriterleri (hatanın belirli bir değere düşmesi gibi) gerçekleştiğinde biter ve en son iterasyonda elde edilen gerçek durum bilgisi kullanılarak sürekli tanı testi için ROC eğrisi altında kalan alan hesaplanır[17].

2.5.5 Tanı Testinin İkili, Sıralı veya Sürekli Ölçüm Olması Durumunda KOAS Yanlılığının Düzeltilmesinde Önerilen Alternatif Yaklaşım

KOAS testini içeren çalışmalarda test sonuçları için yanlılığa neden olan durum bireylerin altın standart test sonucunun olmamasıdır. Bu yanlılığı ortadan kaldırmanın yolu bunun için bir düzeltme geliştirmek olabildiği gibi, KOAS sonucu ve diğer bazı ek değişken bilgileri ile bilinmeyen altın standart sonucunu tahmin etmek de olabilir. Bayes yaklaşımında yapılan işlem altın standart sonucunun belirlenmesi üzerinedir. Burada alternatif olarak bu işlemin Gizli Sınıf Analizi kullanılarak da tespit edilebileceği gösterilmiştir.

Bir veri setinde benzer objeleri gruplar altında toplama işlemi kümeleme analizi olarak bilinmektedir. Burada söz edilen grupların sayısı ve yapısı bilinmektedir[43]. Benzer durum Gizli Sınıf Analizi (Latent Class Analysis – GSA) için de geçerlidir. Gizli sınıflar gözlenemeyen alt grup veya bölümlerdir. Analiz sonucunda elde edilen sınıflara göre, aynı sınıftaki objeler belirli kriterlere göre homojenlik gösterirken, farklı sınıftaki objeler önemli bazı kriterlere göre farklılık göstermektedir. Kümeleme, Gizli Sınıf Analizinin kullanım amaçlarından olduğu gibi derecelendirme (scaling), yoğunluk tahmini (density estimation) ve rastgele etkiler modellemesi gibi birçok başka amaçla da GSA kullanılmaktadır. Günümüzde GSA aynı zamanda sınırlı karma modeller (finite mixture models) olarak da bilinmektedir[44].

Gizli Sınıf Analizi ilk olarak 1950 yılında Lazarsfeld tarafından, bir saha araştırmasında elde edilen iki değerli yanıtlardaki heterojenliği açıklamak amacıyla önerilmiştir[45]. Kullanımın artmasıyla, 1970'li yıllarda gizli sınıf metodolojisi daha

mantıksal dayanakları bulunan bir yapı haline gelmeye başlamıştır. 1974 yılında Goodman ikiden fazla değer alan isimsel değişkenler için de Gizli Sınıf Analizinin uygulanabileceğini göstermiştir[46]. Yine aynı dönemde Haberman gizli sınıf modelleri ile eksik ölçümlü frekans tabloları için log-lineer modelleri arasındaki bağlantıyı göstermiştir[47]. Bu tarihten sonra gizli sınıf modellerin kullanımı; sürekli ek değişkenler, yerel bağımlılık, çok sayıda gizli sınıflar ve tekrarlı ölçümler gibi birçok durum için genişletilerek artmıştır.

2.5.5.1 Gizli sınıf modelleri ve önerilen yaklaşım

Gizli sınıf modelleri gizli değişkenler modellerinin (Latent Variable Models) bir alt koludur ve bu modellerde varsayılan gizli değişken kategorik bir değişkendir. Gizli sınıf modellerinde, verideki bireylerin tam olarak hangisi olduğu bilinmemektedir, ama kesin olarak 1,…,M sınıftan birinde olduğu varsayılmaktadır. Bu sınıflar kategorik bir gizli değişkenin kategorileri olarak görülür. i. bireyin j sınıfında olma olasılığı p_j ile gösterilir ve bu model parametresidir[48].

Gizli sınıf modellerinde; yanıt değişkeni olarak varsayılan ölçümün, gözlenen değişkenlere ve gözlenemeyen ama yanıt değişkeninde heterojenliğe neden olan gizli değişkenlere bağlı olduğu varsayılmaktadır. Bu heterojenlik modellenebilir ve dolayısıyla verideki bireyler, ilgili gruplara sınıflanabilir. M sınıflı bir gizli sınıf modeli şöyle tanımlanabilir. i. bireyin K duruma (ör. teste) verdiği yanıtlar y_i vektörü ile gösterilsin. η_i i. bireyin ait olduğu gözlenemeyen gizli sınıfı göstersin. Bu durumda i. bireyin j. sınıfta olma olasılığı $p_j = P(h_i = j)$ ile gösterilir. Bu durum sağlık alanında bir veriye uygulanırsa, örneğin bir hastalığın tanısını koymada K durum altın standart olmayan test sonuçlarını, η_i i. bireyin bilinmeyen gerçek durumunu ve p_j ise verideki tanı değeri j olan bireylerin yüzdesini göstermektedir[49].

Gizli sınıf modellerinde koşullu bağımsızlık varsayımı altında η_i verildiğinde olabilirlilik fonksiyonu şu şekilde yazılır.

$$L(Y) = \prod_{i=1}^{N} \sum_{j=1}^{M} p_j \prod_{k=1}^{K} p_{jk}^{y_{ik}} \left(1 - p_{jk}\right)^{1-y_{ik}}$$

Burada Y, N bireyin gözlenen verilerinin matrisini, $p_{jk} = P(y_{ik} = t \mid h_i = j)$ ise j sınıfındaki k. duruma t yanıtını veren bireylerin olasılığını göstermektedir. Bu olasılık veri setindeki olabilecek ek değişkenler yardımı ile aşağıdaki gibi modellenebilir.

$$P(y_{ik} = t \mid h_i = j) = \frac{\exp(\mathbf{x}'_{ik,t}\boldsymbol{\beta}_j)}{\sum_{t=1}^{T_i} \exp(\mathbf{x}'_{ik,t}\boldsymbol{\beta}_j)}$$

3 Gereç ve Yöntem

Bu tezde, tanı testlerinin doğruluklarının saptanmasında ortaya çıkabilen yanlılıkların düzeltilmesinde kullanılan bazı yöntemler incelenmiştir. Bu yanlılıklar, seçilmiş örneklerle yapılmış çalışmalar sonucunda elde edilebildiği gibi gerçek durumun tayininde kullanılan altın standart testinin olmaması durumuyla da oluşabilmektedir. İki yanlılık durumu da birbirlerinden bağımsız olarak gerçekleştiği için iki durum için de farklı gereç ve yöntemler kullanılmıştır.

3.1 Doğrulama yanlılığının düzeltilmesi

İki değer alan bir test için doğrulama yanlılığının düzeltilmesinde önerilen yaklaşımların farklı koşullarda performanslarını karşılaştırabilmek için üretilmiş (simule) veriler ve gerçek veri (Nükleer Tıp alanında yapılmış bir çalışmanın verilerinin bir kısmı) kullanılmıştır. Veri üretilirken birçok durum göz önüne alınmıştır. İki değerli tek bir tanı testi için doğrulama yanlılığının düzeltilmesinde kullanılan Begg-Greenes düzeltmesine ve Zhou sınırlarına alternatif olarak önerilen Lojistik Regresyon, Yapay Sinir Ağları (Neural Networks) ve Çoklu İmputasyon (Multiple Imputation) yaklaşımlarını karşılaştırmak amacıyla farklı özelliklerde veriler üretilmiştir. Üretilen veride

- Altın standart test sonucu
- Yeni tanı testi sonucu
- İki değer alan ve altın standart test sonucu ile ilişkili iki adet ikili (binary) değişken
- Birbirleri ile korale ve altın standart test sonuçlarına göre farklı dağılımlar gösteren iki adet sürekli değişken
- Karıştırıcı etki olması maksadıyla altın standart test sonucundan bağımsız bir sürekli değişken

vardır. Üretilen tüm verilerde prevelans sabit ve 0.5 olarak alınmıştır. Altın standart test sonucu ile ilişkili olan iki değerli değişkenler, testin pozitif çıktığı grupta 0.2 oranında ve testin negatif çıktığı grupta 0.7 oranında 1 değerini almıştır. Birbirleri ile korale olan sürekli değişkenler ise iki değişkenli (bivariate) normal dağılım gösterecek şekilde üretildi. Buna göre gruplardaki dağılımlar şöyledir:

Çizelge 3.1 Üretilen veride iki değişkenli normal dağılım gösteren değişkenlerin dağılım bilgileri

Altın Standart Test Sonucu Pozitif		Altın Standart Test Sonucu Negatif	
Ortalama vektörü	Varyans-Kovaryans matrisi	Ortalama vektörü	Varyans-Kovaryans matrisi
$\begin{pmatrix} 4 \\ 3 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$

Üretilen veriler büyük (n=1000) ve küçük (n=100) olacak şekilde farklı örneklem büyüklüklerinde üretilmiştir. Doğrulama yanlılığı problemini oluşturmak için üretilen verilerdeki altın test sonuçları, veri setine göre değişen oranlarda, eksik ölçüm yapılmıştır. Eksik ölçüm oranının yaklaşımların performansını etkileyeceği düşünüldüğünden, %30, %50, %80 olacak şekilde ayarlanmıştır. Yaklaşımların eksik ölçümlerin rastgele olup olmamasına göre önerilmesi nedeniyle, eksik ölçümler yaklaşım karşılaştırmalarının ilk bölümünde rastgele olacak şekilde ve karşılaştırmaların ikinci bölümünde ise rastgele olmayacak şekilde oluşturulmuştur. Rastgele eksik ölçümlü veriler elde etmek için verinin altın standart test sonucu belirlenen oranda rastgele olarak eksik ölçüm yapılmıştır. Rastgele olmayan eksik ölçümlerin oluşturulmasında ise yanlılık $P(V=1|T=i,D=j)$ $i,j=0,1$ koşullu olasılıkların belirlenmesi ile sağlanmıştır. Bu koşullu olasılıklar yeni tanı testi sonucu ve gerçek durum bilgisi verilmişken doğrulanmış birey seçme olasılıklarını göstermektedir. Bu değerlerin uygun şekilde atanması sonucu eksik ölçüm mekanizmaları oluşturulabilir. Örneğin, bu değerlerin $P(V=1|T=1,D=1)=0.15$, $P(V=1|T=0,D=1)=0.80$, $P(V=1|T=1,D=0)=0.1$ ve $P(V=1|T=0,D=0)=0.15$ alınması halinde eksik ölçüm mekanizmasının NMAR olduğu görülmüştür. (Kosinski ve Barnhart yaklaşımındaki Eksik ölçüm mekanizması bileşeni için yapılan Lojistik Regresyon analizi sonucunda

Gerçek durum bilgisi modele girmiştir. Dolayısıyla bireylerin eksik ölçümlü olup olmaması gözlenemeyen gerçek durum bilgisine bağlıdır.)

Veriler üretilip, belirli oranda eksik ölçüm yapılarak çıkartılan kısım dışında kalan veri kullanılarak altın standart test sonucu Lojistik Regresyon ile modellenmiştir. Oluşturulan modelin altın standardı açıklama oranlarına göre tanımlama katsayısının

- 0.60 – 0.70 arasında olduğu
- 0.40 – 0.50 arasında olduğu
- 0.20 – 0.30 arasında olduğu
- 0.07 – 0.15 arasında olduğu

durumlar için farklı farklı veriler üretilmiştir.

Farklı durum ve özellikteki veriler üretildikten sonra, veri eksik ölçümlü hale getirilmeden önce, Gerçek Durumu göstermek amacıyla üretilen veride doğruluk ölçütleri hesaplanmıştır. Daha sonra eksik ölçümlü hale getirilen veriler üzerinde önerilen tüm yaklaşımlar uygulanmıştır. Yaklaşımların doğruluk ölçüt tahmin edicileri; eksik ölçümlerin çıkarıldığı veriler kullanılarak Genel Bilgiler bölümünde açıklanan yöntemlerle hesaplanmıştır. Yapay Sinir Ağları yaklaşımında kullanılan gizli katmanın sonuca olan etkisini görebilmek için katmandaki düğüm sayısı 1, 2 ve 3 olarak alınmış ve karşılaştırma sonuçlarında çizelge ve şekillerde YSA1, YSA2 ve YSA3 olarak belirtilmiştir. Tüm ölçütler Gerçek Duruma göre ortaya çıkan hata payları (bias) dikkate alınarak karşılaştırılmıştır. Bu işlem 500 kere tekrar edilerek benzer özellikteki durumlar için yaklaşımların performanslarının dağılımları elde edilmiştir.

Gerçek veri seti olarak Sukan ve ark.[50] tarafından yapılan hiperparatiroidi tanısını koymada kullanılan yeni bir tanı testinin değerlendirildiği çalışmanın verilerinin bir kısmı kullanılmıştır. Bu çalışmada 119 birey yeni tanı testine tabi tutulurken, bu bireylerin sadece 59'nun altın standart test sonucu mevcuttur. Bu tezde bu çalışmadan elde edilen verilerden seçilmiş 98 bireyin sonucu üzerinden analizler yapılmıştır.

Yaklaşımları karşılaştırmada R[51] programı kullanılmıştır. Begg & Greenes ve Zhou alt ve üst sınır yaklaşımları için kod yazılmış, Lojistik Regresyon ve Yapay Sinir Ağları için R'da bulunan hazır kodlar (nnet) kullanılmış, Çoklu imputasyon yaklaşımı

içinse van Buuren ve Oudshoorn tarafından R için geliştirilen MICE V2.3 koduna başvurulmuştur[52].

3.2 Kesin olmayan altın standart yanlılığın düzeltilmesi

Kesin olmayan altın standart yanlılığı, doğruluğu hesaplanmak istenen tanı testinin veri ölçek tipine göre önerilen yaklaşımlarca düzeltilebilmektedir. Buna göre bu yanlılığın varlığında

- isimsel (binary) ölçek tipinde olan bir tanı testinin doğruluk ölçütlerini hesaplamada Çözücü test, Bileşke Referans Standardı yaklaşımları,
- sıralı (ordinal) ölçek tipinde olan bir tanı testinin doğruluk ölçütünü hesaplamada Zhou tarafından önerilen Parametrik olmayan ROC yaklaşımı ve
- aralık veya oran (interval or ratio) ölçek tipindeki bir tanı testinin doğruluk ölçütünü hesaplamada ise Wang ve ark. tarafından önerilen Bayes yaklaşımı

kullanılmaktadır. Bu tezde tüm veri yapıları için önerilen yaklaşımlara alternatif olarak Gizli Sınıf Analizi yaklaşımı önerilmiştir.

Literatürde önerilen yaklaşımların varsayımları dikkate alınarak, yeni önerilen yaklaşımı bu yaklaşımlarla karşılaştırmak amacıyla farklı özelliklerde veriler üretilmiştir. Hem Bileşke Referans Standardı yaklaşımında hem de Parametrik olmayan ROC yaklaşımında ikiden fazla testin varlığı varsayıldığından, isimsel ölçek tipindeki tanı testi için üretilen veride 3 test, sıralı ölçek tipindeki tanı testi için üretilen veride değişen sayıda (3 ve 10 tane) tanı testi bulunmaktadır. Yaklaşımların performanslarını etkileyebileceği düşünüldüğünden altın standart test sonucu dikkate alınarak hastalık prevelansı 0.1'den 0.9'a kadar değişen değerlerde; doğruluk ölçütü değeri değişen değerlerde (ikili test için Duyarlılık ve Seçicilik değerleri 0.5-0.8 arası, sıralı test için eğri altında kalan alan 0.7, 0.8 veya 0.9); örneklem büyüklüğü değişen değerlerde (ikili test için 50 – 1000 arası; sıralı test için 100 veya 1000) alınmıştır. Sıralı ölçekti tanı testlerinin kategori sayıları da sonuca etkisi olabileceği düşüncesi ile değişen değerlerde (3, 5, 7 veya 10 kategori) alınmıştır. Hem isimsel hem de sıralı ölçek tipindeki tanı testi için karşılaştırmalarda kullanılan veri setleri üretilen testler arasındaki korelasyonu

sağlayabilmek için çok değişkenli (multivariate) standart normal dağılım kullanılarak üretilmiştir. Bu dağılımdan üretilen veriler ölçek tipine göre gruplanmıştır.

İkili ölçek tipindeki tanı testinin KOAS yanlılığını düzeltmede kullanılan yaklaşımları karşılaştırmak için kullanılan gerçek veri seti Gastroenteroloji bölümünde M. Sandıkçı danışmanlığında F.Koçak tarafından yapılmış olan bir uzmanlık tezinden elde edilen verilerdir[53]. Koçak'ın çalışmasında, hem yetişkinlerde hem de çocuklarda gaitada Helicobacter Pylori (HP) antijenini araştıran, yeni geliştirilmiş HpSA testinin değerlendirilmesi yapılmıştır. Bu amaçla çocuklarda ve yetişkinlerde farklı tipteki test ve test kombinasyonları kullanılarak karşılaştırmalar yapılmıştır. Bu tezde ise Koçak'ın çalışmasının yetişkinler bölümünde kullanılan verileri kullanılmıştır. HP tanısını koymada yetişkinlerde Kültür, CLO, Histoloji ve PCR gibi testler kullanılmaktadır. Koçak'ın tezinde belirtildiğine göre Kültür testi Kesin Olmayan Altın Standart olarak ve diğer 3 testten herhangi biri ise Çözücü test olarak kabul edilmektedir. Ayrıca Kültür testi dışındaki diğer 3 test kullanılarak da Bileşke Referans Standardı oluşturulabilmektedir. Bu tezde, bu bilgiler ışığında çeşitli Gizli Sınıf Modelleri oluşturularak karşılaştırmalar yapılmıştır.

Sıralı ölçek tipindeki tanı testinin KOAS yanlılığını düzeltmede kullanılan yaklaşımları karşılaştırmak için Dermatoloji alanında yapılmış bir çalışmadan elde edilen gerçek bir veri seti de kullanılmıştır. Buna göre Günaştı ve ark. yaptıkları çalışmada Dermatoloji bölümündeki 5 öğretim üyesi 0-6 arasında bir skor vererek hastaların hastalıklarını derecelendirmiştir. Günaştı ve ark. çalışmalarında öğretim üyeleri arasındaki uyumu saptamaya çalışmışlardır[54]. Bu tezde ise bu veri öğretim üyelerinin sınıflama başarılarını ölçmek amacıyla kullanılmıştır.

Tanı testinin aralık veya oran ölçek tipinde olduğu durumda önerilen Bayes yaklaşımı ile bu tezde önerilen Gizli Sınıf Analizi yaklaşımını karşılaştırmada kullanılan veri setinde aşağıdaki değişkenler üretilmiştir;

- Altın standart test sonucu
- Kesin olmayan altın standart test sonucu
- Yeni tanı testi sonucu
- İki değer alan ve altın standart test sonucu ile ilişkili iki adet ikili (binary) değişken

- Birbirleri ile korale ve altın standart test sonuçlarına göre farklı dağılımlar gösteren iki adet sürekli değişken
- Karıştırıcı etki olması maksadıyla altın standart test sonucundan bağımsız bir sürekli değişken.

Birbirleri ile korale olan tüm değişkenler çok değişkenli (multivariate) standart normal dağılım kullanılarak elde edilmiştir. Buna göre ikili değişkenler üretilen normal dağılım gösteren değişkenin ortanca değerine göre gruplanmasıyla elde edilmiştir. Üretilen veriler büyük (n=1000) ve küçük (n=100) olacak şekilde farklı örneklem büyüklüklerinde üretilmiştir.

Sonuçları etkileyebileceği düşünülen hastalık prevelansı, Eğri Altında Kalan Alan (EAKA) değeri, sınıflamada kullanılan ek değişkenlerin gerçek durum açıklama oranı değerleri, karıştırıcı etkisi olan değişkenin varlığı gibi durumlar için farklı veri setleri üretilmiştir. Bu durumlar aynı veri setinde aynı anda etki edebileceği için bir durum ele alındığında diğer durumların sabit olması varsayılmıştır. Yani hastalık prevelansı değiştirilirken eğri altında kalan alan değeri yüksek ve sabit, ek değişkenlerin gerçek durumu açıklama oranı yüksek ve sabit, karıştırıcı ek değişken olmayacak şekilde alınmıştır. Benzer şekilde diğer durumlar da buna göre uyarlanmıştır. Detaylı bilgi Çizelge 3.2'de verilmiştir.

Aralık veya oran tipindeki tanı testi için KOAS yanlılığı düzeltmede kullanılan Bayes yaklaşımı ile Gizli Sınıf Analizi yaklaşımını karşılaştırmak için gerçek veri seti olarak Kardiyolojide A.Bozkurt danışmanlığında M. Koç tarafından yapılmış olan bir uzmanlık tezinden elde edilen verilerin bir kısmı kullanılmıştır. Koç'un çalışmasında fonksiyonel egzersiz kapasitesinin belirlenmesinde serum NT-proBNP değerinin önemi araştırılmıştır. Çalışmada kalp yetmezliği olan ve olmayan toplam 120 hasta üzerinde kardiyolojik incelemeler yapılıp, tüm hastalar efor testine alınmış ve fonksiyonel kapasiteleri hesaplanmıştır. Fonksiyonel kapasiteyi göstermede METS efor kapasitesi değeri kullanılmıştır. Bu değerin 5'in altında olması düşük, üzerinde olması da yüksek efor kapasitesi olarak sınıflanmış ve bu sınıflamayı belirlemede kullanılabilecek parametreler incelenmiştir[55]. Bu tezde ise bu çalışmadaki 95 hastanın verileri kullanılmış ve METS değerine göre oluşturdukları gruplar gerçek durum bilgisi olarak, efor testindeki maksimum kalp atım hızının gruplanmış hali (143 altı ve üzeri) KOAS

olarak, NT-proBNP değeri yeni tanı testi olarak, çeşitli eko parametreleri de ek değişkenler olarak alınarak Bayes ve Gizli Sınıf Analizi yaklaşımları karşılaştırılmıştır.

İkili ölçek tipindeki tanı testinin KOAS yanlılığını düzeltmede kullanılan yaklaşımları karşılaştırmak için üretilen veri setleri üzerinde yaklaşımlar farklı durumlarda uygulanmıştır. Yaklaşımların önerildiği durum dikkate alınarak çeşitli düzenlemeler yapılmış ve farklı modeller kurulmuştur. Örneğin; Çözücü olarak kabul edilen CLO test, KOAS olarak kabul edilen Kültür testi ile yeni tanı testi olan HpSA arasındaki uyumsuzluğu çözeceği için, her iki testin uyumsuz oldukları durumda CLO test, bu testleri uyumlu hale getirildi. Önerilen yaklaşımlar Gizli Sınıf Analizinde oluşturulan 5 farklı modelle karşılaştırılmıştır. Bu modeller sırasıyla

GS1: Kültür, Çözücü test – düzeltilmiş CLO Test, CLO test, Histoloji ve PCR,

GS2: Kültür, Çözücü test – düzeltilmiş CLO Test, Histoloji,

GS3: Kültür, Çözücü test – düzeltilmiş CLO Test, Bileşke Referans Standardı,

GS4: Kültür, Bileşke Referans Standardı,

GS5: Kültür, CLO test, Histoloji ve PCR

olarak alındı.

Sıralı ölçek tipindeki tanı testinin KOAS yanlılığını düzeltmede kullanılan yaklaşım olan Parametrik olmayan ROC eğrisi tahmininde yöntemin anlatımında verildiği gibi uygun başlangıç değerleri ile g fonksiyonu, q değeri ve p değeri tekrar tekrar düzenlenmiştir. Her bir tekrar sonunda ϕ değeri elde edilmiş ve uygun sayıda tekrar sonucunda bu değer kullanılarak ROC eğrisi altında kalan alan hesaplanmıştır. Burada bu yaklaşıma alternatif olarak önerilen Gizli Sınıf Analizinde ise üretilen tüm test sonuçları bir araya getirilerek gizli bir değişken oluşturulmuş ve bu değişken Gerçek Durum olarak kabul edilerek tanı testlerinin doğruluk ölçütleri hesaplanmıştır. Gizli değişken oluşturma aşamasında Gizli Sınıf Kümeleme Analizi kullanılmıştır.

Aralık veya oran tipindeki tanı testi için KOAS yanlılığı düzeltmede kullanılan Bayes yaklaşımında ise yukarıda verildiği gibi başlangıç değerleri verildikten sonra önce D gerçek durumu, sonra β lineer regresyon katsayıları, sonra σ varyans değerleri, sonra α olasılıkları ve son olarak δ lojistik regresyon katsayıları her bir iterasyonda hesaplanmıştır. En son iterasyonda elde edilen D gerçek durum değişkeni ile ROC eğrisi altında kalan alan değeri hesaplanmıştır. Burada Bayes yaklaşımına alternatif olarak önerilen Gizli Sınıf Analizinde ise birden fazla model kurulmuştur. Modeller,

modeldeki değişken sayısına ve bu değişkenler arasında etkileşimin olup olmamasına göre değişkenlik göstermektedir. 4 farklı model kurulmuştur. Sadece ikili değişkenlerden oluşan model (Model I), Model I'e ilave olarak sürekli değişkenli modeller (Model II ve Model III) ve son olarak tüm değişkenleri içeren model (Model IV). Bu modellere ilave olarak sadece karıştırıcı ek değişkenli verinin analizinde kullanılmak üzere karıştırıcı ek değişken içeren bir model daha kurulmuştur (Model V).

Yaklaşımların karşılaştırılmasında R[51], Matlab[56] ve LatentGold[57] paket programları kullanılmıştır. Parametrik olmayan ROC eğrisi ve Bayes yaklaşımı için kod yazılmıştır. Yazılan kodlar yaklaşımları öneren Zhou ve ark.[16] ve Wang ve ark.[17] makalelerinde verildiği üzere yazılmıştır. Önerilen diğer yaklaşım olan Gizli Sınıf Analizi için ise yeni tanı testinin isimsel ölçek tipinde olduğu durumda R[51] içinde bulunan kütüphaneler (e1071) kullanılmıştır. Diğer durumlarda ise LatentGold[57] paket programı kullanılmıştır. Yaklaşımlar üretilen veriye uygulanmış ve yaklaşımlara göre ROC eğrisi altında kalan alan değerleri ve bu değerlerin standart hata değerleri kaydedilmiştir. Bu işlem 500 kere tekrarlanarak nokta tahmin edicileri için aralık tahminleri belirlenmeye çalışılmıştır.

Çizelge 3.2 KOAS yanlılığını düzeltmede kullanılan yaklaşımları karşılaştırmak için üretilen verilerdeki çapraz ilişkiler

	Hastalık Prevelansı	EAKA+	Ek değişkenlerin Gerçek durumu açıklama oranı	Karıştırıcı ek değişken varlığı
Hastalık Prevelansı 0.1– 0.9	-----	0.8	Yüksek	YOK
EAKA+ 0.7 – 0.9	0.5	------	Yüksek	YOK
Ek değişkenlerin Gerçek durumu açıklama oranı Düşük - Orta	0.5	0.8	-------	YOK
Karıştırıcı ek değişken varlığı VAR	0.5	0.8	Yüksek	-----

[+] EAKA: Eğri Altında Kalan Alan.

4 Bulgular

4.1 Doğrulama yanlılığını düzeltmede kullanılan ve önerilen yaklaşımların karşılaştırılması

Farklı özellikte üretilmiş veriler kullanılarak yapılan analizlerin sonuçları alt başlıklarda verilmiştir. Sonuçlar, Altın Standart sonucunu eksik ölçüm yapma mekanizması (rastgele veya değil) genel başlığı altında veri boyutuna (büyük veya küçük), eksik ölçüm oranına (%20, %50 ve %80) ve tanımlama katsayına göre alt başlıklar altında sunulmuştur. Nükleer tıp alanında yapılmış bir çalışmanın gerçek verilerinin bir kısmı kullanılarak yapılan karşılaştırmalar üretilmiş veri sonuçlarından sonra verilmiştir.

4.1.1 Eksik ölçüm mekanizması rastgele kabul edildiğine (MAR)

4.1.1.1 Üretilmiş Büyük Veri Seti Kullanılarak Yapılan Karşılaştırmalar

1000'lik olarak üretilmiş büyük veri setlerindeki üretilen değişkenlerin, altın standardı açıklama oranına göre karşılaştırmalı sonuçları aşağıdaki çizelge ve şekillerde verilmiştir.

4.1.1.1.1 Tanımlama Katsayısının (R^2) 0.60 ve 0.70 arasında olduğu durumlar

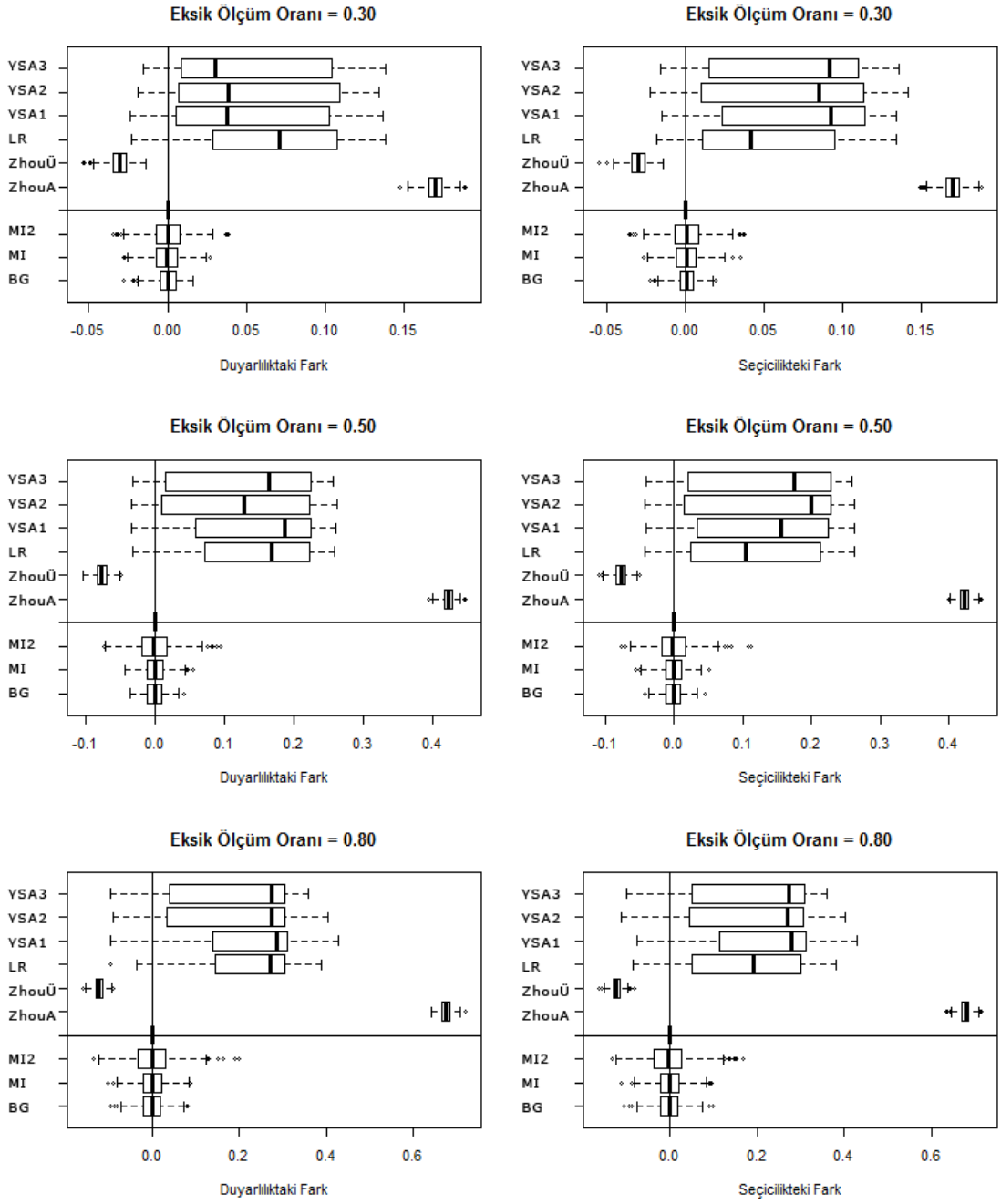
500 defa üretilen verilerde uygulanan regresyon modelinin tanımlama katsayısı ortalaması 0.661 ve standart sapması 0.079 olarak elde edildi. Bu oran modelin sonuç değişkenini ne kadar açıkladığının göstergesi olduğundan çalışmalarda değerinin yüksek olması istenir. Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.1'de ve yaklaşımların gerçek durum ile farklarının

sonuçları Şekil 4.1'de verilmiştir. Yaklaşımlar arasında Begg&Greenes ve Zhou yaklaşımları sadece Tanı testi sonucu ve Altın Standart test sonucunu kullandığından Tanımlama katsayısındaki değişim bu yaklaşımların tahmin değerlerini etkilememiştir. Çizelgede de görüldüğü gibi, eksik ölçüm oranının artması yaklaşımın doğruluk ölçütü tahminlerinin standart sapmasının artmasına ve dolayısıyla yaklaşımlara ait mutlak hataların artmasına neden olmuştur.

Çizelge 4.1 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark
	Gerçek Durum	847±12	---	846±11	---	846±11	---
MAR	Begg&Greenes	847±14	6±4	846±19	11±8	847±29	22±17
	Çoklu İmputasyon	848±15	7±6	846±21	14±10	847±32	24±18
	Çoklu İmputasyon – Ek Değişkenli	847±17	9±7	845±29	21±17	847±49	39±31
NMAR	Zhou Alt Sınır	678±11	169±6	423±9	423±8	169±9	677±13
	Zhou Üst Sınır	878±11	31±6	923±10	77±10	969±7	123±12
	Lojistik Regresyon	780±44	68±42	701±86	146±83	629±110	220±105
	Yapay Sinir Ağları – 1	796±49	53±46	698±90	150±87	621±123	230±115
	Yapay Sinir Ağları – 2	791±52	57±49	730±102	120±98	654±137	199±127
	Yapay Sinir Ağları – 3	793±49	55±47	717±102	132±97	650±134	202±124
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark
	Gerçek Durum	847±12	---	846±11	---	846±11	---
MAR	Begg&Greenes	847±14	6±4	846±19	11±8	848±32	23±18
	Çoklu İmputasyon	847±16	7±6	847±21	14±10	847±34	26±20
	Çoklu İmputasyon – Ek Değişkenli	847±17	9±7	845±28	20±17	847±51	40±31
NMAR	Zhou Alt Sınır	678±12	169±7	423±9	423±8	169±9	677±13
	Zhou Üst Sınır	877±12	30±6	923±10	78±10	969±7	123±11
	Lojistik Regresyon	796±45	52±42	729±93	120±88	669±127	183±120
	Yapay Sinir Ağları – 1	775±48	73±45	711±94	137±90	631±127	220±117
	Yapay Sinir Ağları – 2	783±52	65±49	708±104	142±98	655±134	198±126
	Yapay Sinir Ağları – 3	781±50	67±47	713±99	135±96	649±134	204±123

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.1 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.1.1.2 Tanımlama Katsayısının (R^2) 0.40 ve 0.50 arasında olduğu durumlar

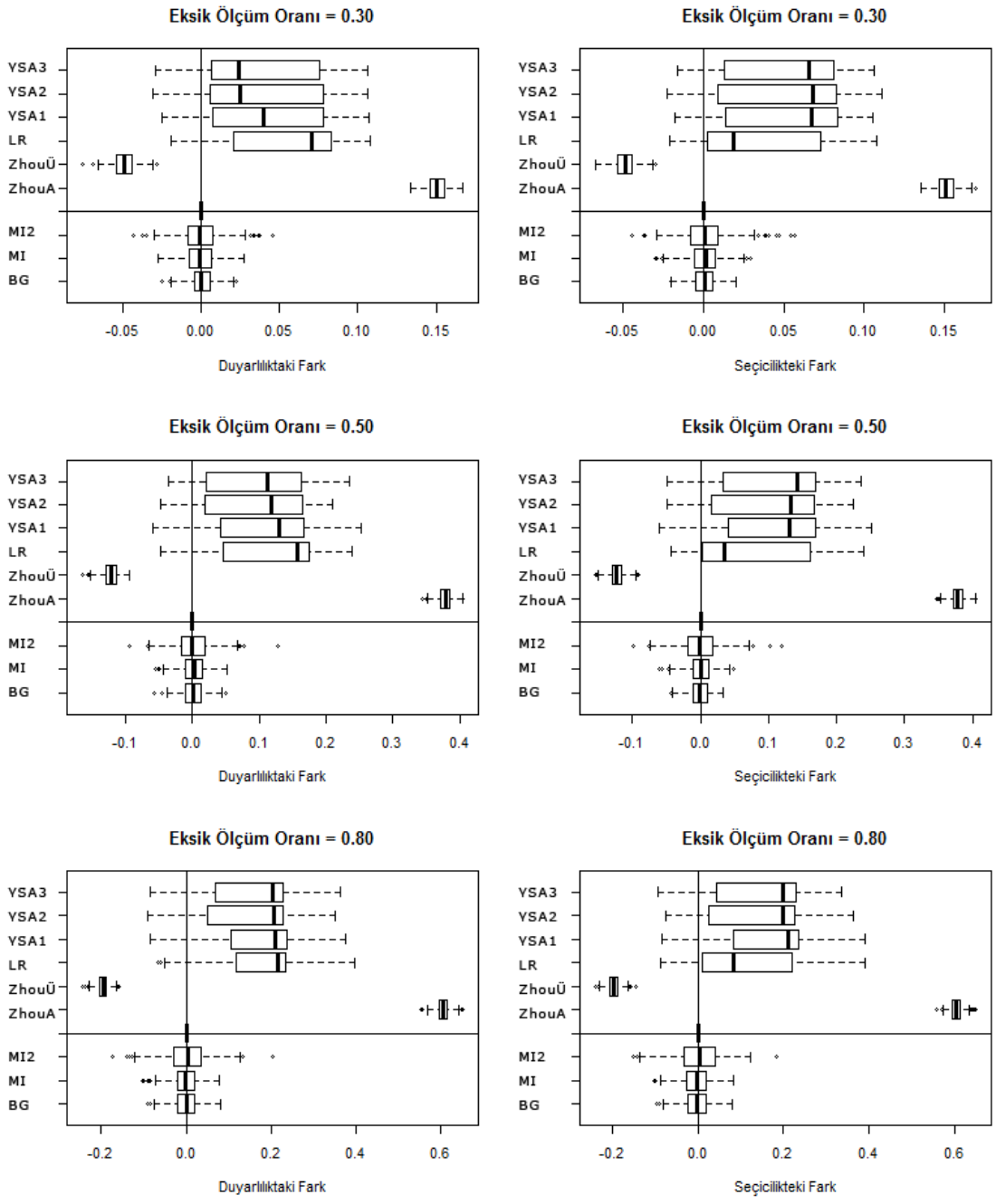
500 defa üretilen verilerde uygulanan regresyon modelinin tanımlama katsayısı ortalaması 0.447 ve standart sapması 0.013 olarak elde edildi. Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.2'de ve

yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.2'de verilmiştir. Elde edilen sonuçlara göre tanımlama katsayısının 0.7 civarında olduğu durum ile benzer olarak tüm yaklaşımlar gerçek durumu ufak hata payları ile tahmin edebilmiştir.

Çizelge 4.2 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark
	Gerçek Durum	754±14	---	755±14	---	755±14	---
MAR	Begg&Greenes	754±16	6±5	753±21	13±10	756±35	24±18
	Çoklu İmputasyon	754±17	8±6	753±23	15±11	756±36	25±19
	Çoklu İmputasyon – Ek Değişkenli	754±19	10±8	754±32	22±18	753±54	40±31
NMAR	Zhou Alt Sınır	603±13	151±6	377±10	378±10	151±10	605±15
	Zhou Üst Sınır	803±13	49±7	876±12	122±12	951±8	196±13
	Lojistik Regresyon	699±38	56±33	636±74	121±68	575±94	183±88
	Yapay Sinir Ağları – 1	713±39	44±34	646±71	112±65	581±100	178±93
	Yapay Sinir Ağları – 2	715±40	41±35	660±75	99±69	604±101	158±90
	Yapay Sinir Ağları – 3	715±38	41±34	662±72	96±68	599±102	164±88
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark
	Gerçek Durum	754±14	---	755±14	---	755±14	---
MAR	Begg&Greenes	754±16	6±4	755±20	12±9	757±35	25±18
	Çoklu İmputasyon	753±17	8±6	755±22	14±11	757±37	27±20
	Çoklu İmputasyon – Ek Değişkenli	753±19	10±8	754±32	23±17	753±55	43±31
NMAR	Zhou Alt Sınır	603±13	151±6	377±10	378±10	151±10	604±13
	Zhou Üst Sınır	803±13	49±7	878±11	123±11	951±8	196±13
	Lojistik Regresyon	719±40	38±34	682±82	80±73	644±117	123±103
	Yapay Sinir Ağları – 1	701±38	54±34	646±73	111±68	589±104	171±94
	Yapay Sinir Ağları – 2	705±40	51±36	654±77	104±70	617±109	148±95
	Yapay Sinir Ağları – 3	704±38	51±34	647±73	111±68	606±104	157±91

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.

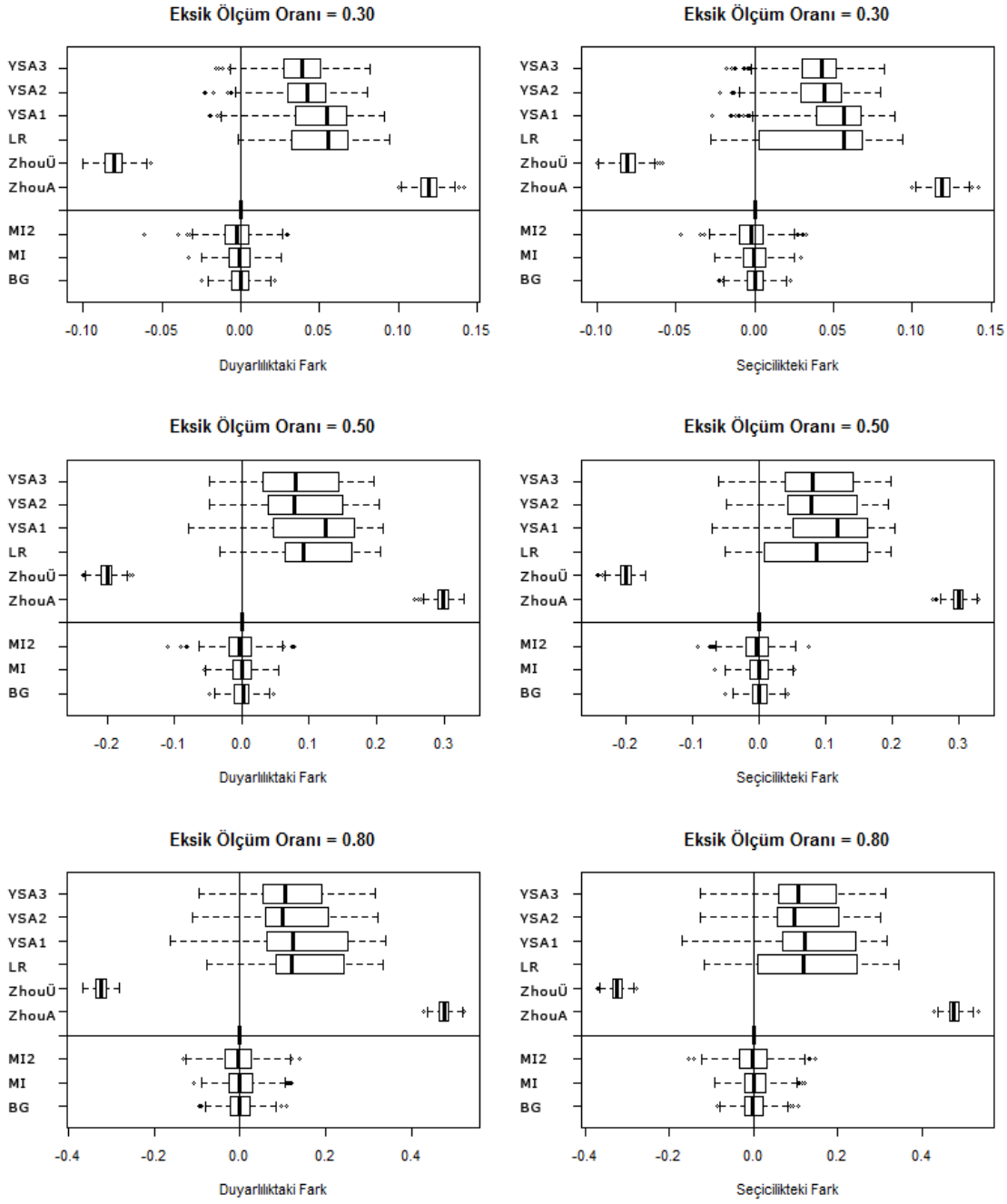


Şekil 4.2 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.1.1.3 Tanımlama Katsayısının (R^2) 0.20 ve 0.30 arasında olduğu durumlar

500 defa üretilen verilerde uygulanan regresyon modelinin tanımlama katsayısı ortalaması 0.258 ve standart sapması 0.012 olarak elde edildi. Bu durumdaki doğruluk

ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.3'te ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.3'te verilmiştir. Daha önceki sonuçlara benzer olarak Begg & Greenes ve Çoklu İmputasyon yaklaşımları gerçek durumu neredeyse sıfır hata ile tahmin edebilirken, bu değer diğer yaklaşımlarda daha yüksek hata payları ile elde edilebilmiştir.



Şekil 4.3 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti doğruluk ölçütleri

Çizelge 4.3 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları (x10^{-3})

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	597±15	---	597±15	---	596±16	---
MAR	Begg&Greenes	597±16	6±5	597±22	13±10	595±36	26±20
MAR	Çoklu İmputasyon	597±18	8±6	598±24	16±11	590±45	33±27
MAR	Çoklu İmputasyon – Ek Değişkenli	599±18	9±7	601±30	20±16	598±51	38±30
NMAR	Zhou Alt Sınır	477±13	119±7	298±12	299±12	119±11	477±16
NMAR	Zhou Üst Sınır	678±13	81±8	799±12	201±12	919±10	324±16
NMAR	Lojistik Regresyon	546±22	51±20	488±51	110±51	437±85	159±86
NMAR	Yapay Sinir Ağları – 1	547±27	50±21	496±73	107±63	454±115	151±99
NMAR	Yapay Sinir Ağları – 2	556±25	41±17	511±66	90±58	476±96	127±84
NMAR	Yapay Sinir Ağları – 3	558±23	39±16	513±64	88±56	479±94	125±79
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	597±15	---	597±15	---	596±16	---
MAR	Begg&Greenes	597±16	6±4	597±22	13±10	594±35	25±20
MAR	Çoklu İmputasyon	597±18	8±6	598±24	16±12	590±44	32±27
MAR	Çoklu İmputasyon – Ek Değişkenli	599±18	9±7	601±30	20±16	598±52	38±30
NMAR	Zhou Alt Sınır	477±13	119±7	298±12	299±11	119±10	477±17
NMAR	Zhou Üst Sınır	677±13	81±7	798±12	201±12	919±9	323±16
NMAR	Lojistik Regresyon	558±36	42±29	514±78	90±71	469±120	139±107
NMAR	Yapay Sinir Ağları – 1	545±26	52±20	496±71	105±62	456±105	148±91
NMAR	Yapay Sinir Ağları – 2	556±26	41±18	510±64	90±57	480±96	125±82
NMAR	Yapay Sinir Ağları – 3	557±23	40±17	512±62	89±55	481±96	126±78

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.

4.1.1.1.4 Tanımlama Katsayısının (R^2) 0.07 ve 0.15 arasında olduğu durumlar

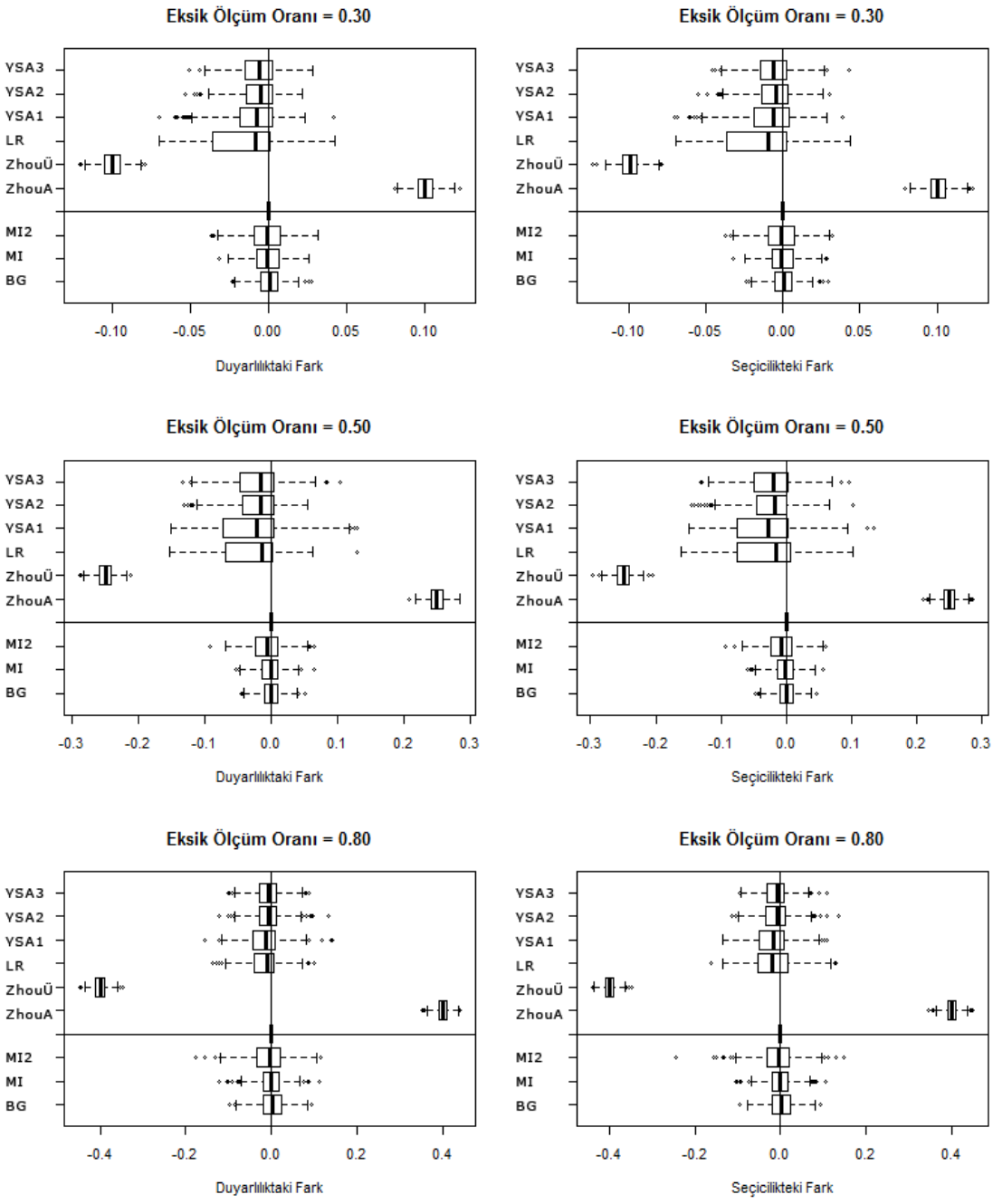
500 defa üretilen verilerde uygulanan regresyon modelinin tanımlama katsayısı ortalaması 0.112 ve standart sapması 0.011 olarak elde edildi. Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.4'te ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.4'te verilmiştir. Bu

bölümde elde edilen sonuçlar daha önceki karşılaştırma sonuçlarına göre farklılık göstermektedir. Tanımlama katsayısının oldukça küçük olması modelde kullanılan değişkenlerin önemini azaltması nedeniyle yaklaşımların tahminlerini daha az hata ile yapmasını sağlamıştır. Mutlak farktaki bu azalma eksik ölçüm oranındaki artış ile daha da azalmaktadır.

Çizelge 4.4 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	501±17	---	499±16	---	500±15	---
MAR	Begg&Greenes	501±19	6±5	499±22	12±9	498±35	25±18
	Çoklu İmputasyon	501±19	8±6	501±21	14±11	502±33	23±20
	Çoklu İmputasyon – Ek Değişkenli	502±21	10±7	506±28	19±15	507±44	33±27
NMAR	Zhou Alt Sınır	400±16	101±7	250±12	250±13	99±10	401±15
	Zhou Üst Sınır	601±16	100±7	750±12	251±13	900±10	399±15
	Lojistik Regresyon	517±29	20±18	532±53	41±44	518±38	30±30
	Yapay Sinir Ağları – 1	511±24	15±14	531±53	46±39	516±40	33±28
	Yapay Sinir Ağları – 2	508±22	11±10	521±38	31±27	509±33	26±23
	Yapay Sinir Ağları – 3	507±21	12±9	520±39	33±27	508±33	26±20
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	501±17	---	499±16	---	500±15	---
MAR	Begg&Greenes	501±19	6±5	499±22	12±9	498±34	24±18
	Çoklu İmputasyon	502±19	8±6	501±21	14±11	501±32	23±19
	Çoklu İmputasyon – Ek Değişkenli	502±21	10±7	506±28	20±15	508±46	34±29
NMAR	Zhou Alt Sınır	400±16	101±7	249±12	250±13	100±10	401±16
	Zhou Üst Sınır	600±16	99±7	750±12	250±13	900±10	399±15
	Lojistik Regresyon	517±31	21±18	532±57	46±43	517±50	42±30
	Yapay Sinir Ağları – 1	511±24	15±14	534±54	48±40	518±41	35±28
	Yapay Sinir Ağları – 2	507±21	11±10	523±37	31±30	511±33	27±23
	Yapay Sinir Ağları – 3	508±21	12±9	522±39	33±28	512±32	26±22

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



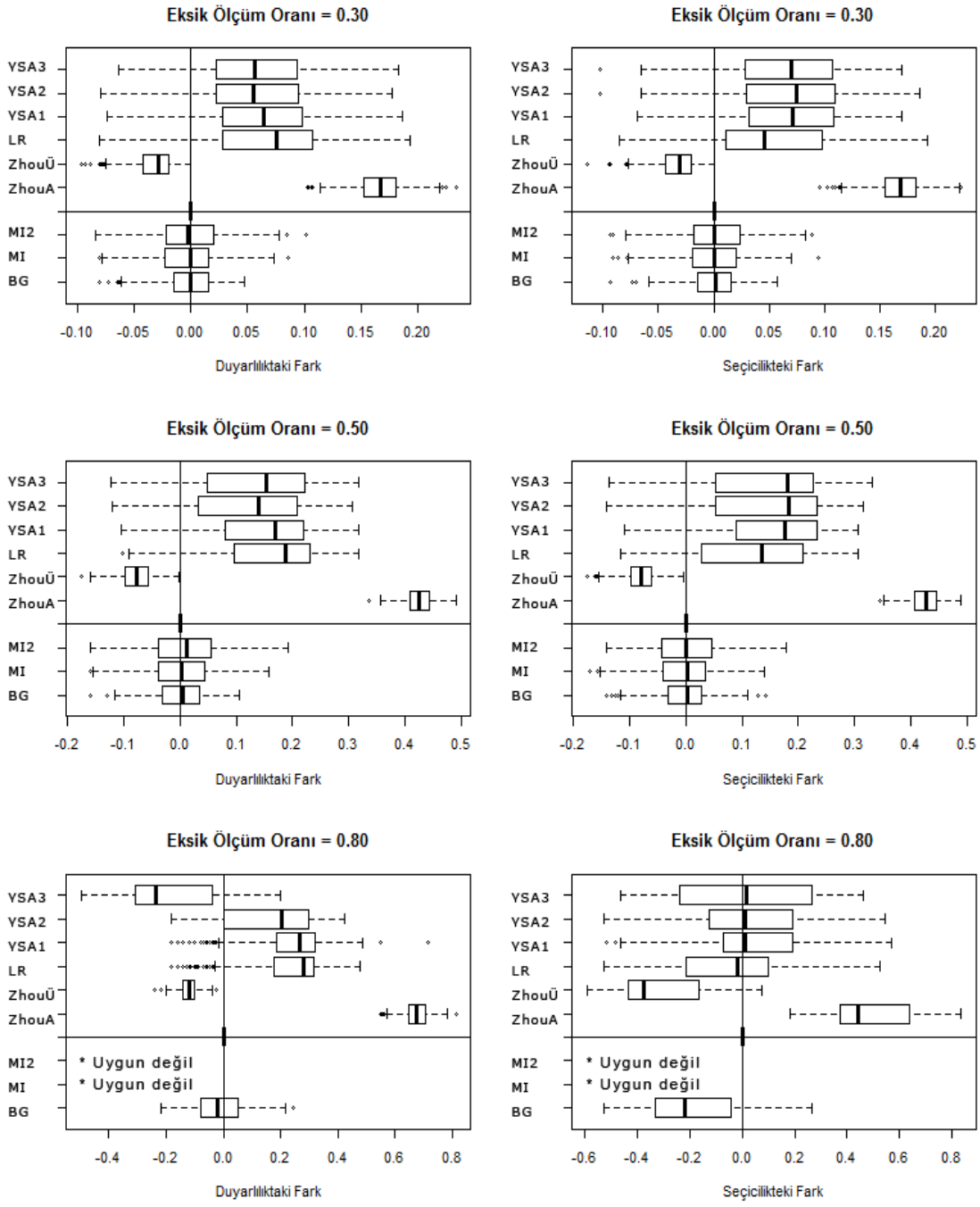
Şekil 4.4 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.1.2 Üretilmiş Küçük Veri Seti Kullanılarak Yapılan Karşılaştırmalar

100'lük olarak üretilmiş küçük veri setlerinde eksik ölçümlü vaka oranı büyük veri setinde alındığı gibi %30, 50 ve 80 olarak alınmıştır.

4.1.1.2.1 Tanımlama Katsayısının (R^2) 0.60 ve 0.70 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.5'te verilmiştir. Büyük veri setinde olduğu gibi burada da yaklaşımlar küçük hata payları ile gerçek durumu tahmin edebilmiştir. Büyük veri setinden farklı olarak burada yaklaşımların tahmin değerlerindeki standart sapma değerleri artmıştır. Eksik ölçüm oranındaki artışın ise bazı yaklaşımların hata paylarının artmasına neden olduğu görülmüştür. Özellikle eksik ölçüm oranın %80 olması durumunda, çoklu imputasyon yaklaşımının karşılaştığı sıfıra bölme hatası nedeniyle sonuç vermemesi, bu yaklaşımın bu gibi durumlarda (küçük örneklem, yüksek eksik ölçüm ve yüksek tanımlama katsayısı) uygulanabilirliği hakkında şüphelerin oluşmasına neden olmuştur.



Şekil 4.5 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti doğruluk ölçütleri

Çizelge 4.5 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

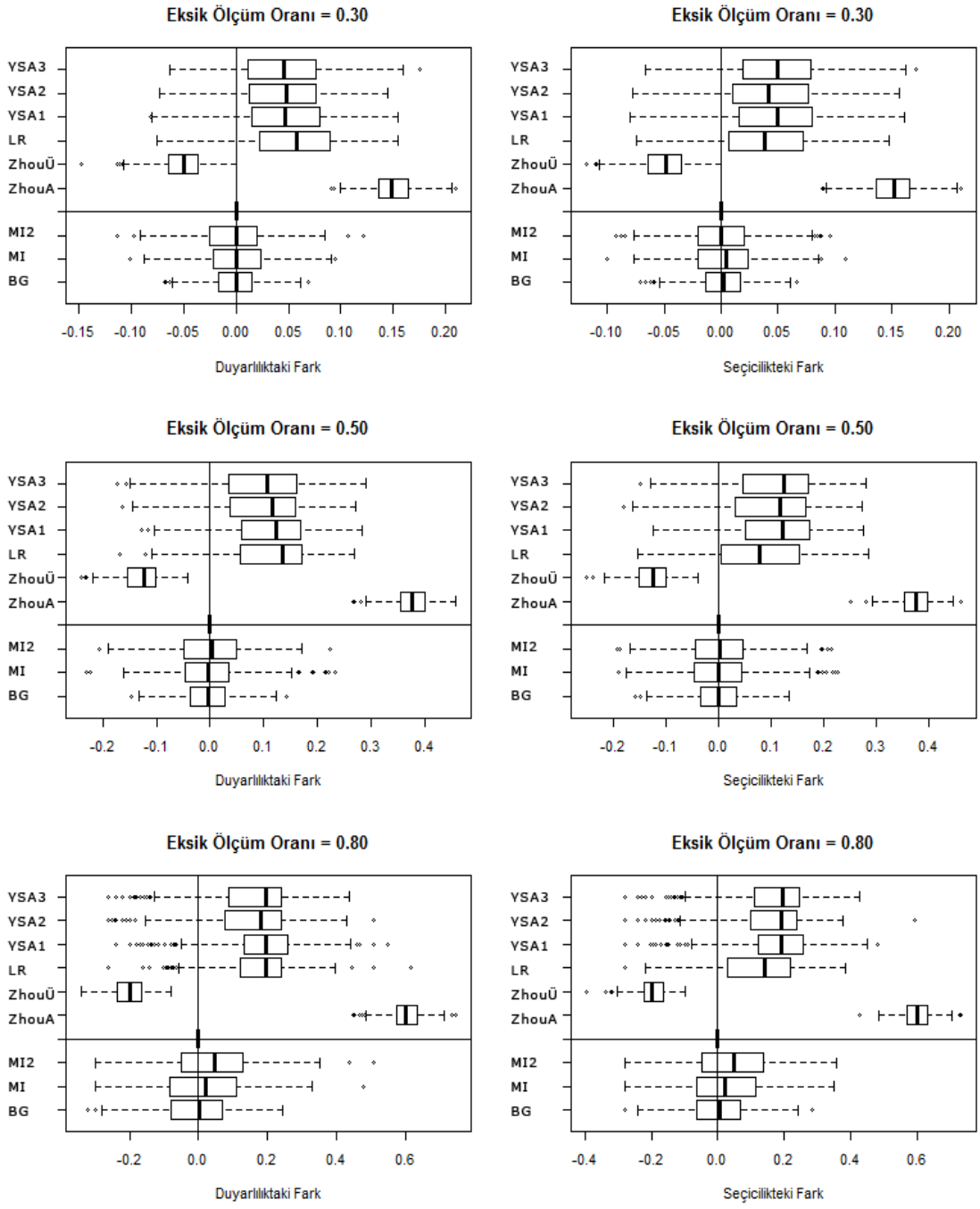
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark	**Duyarlılık**	Mutlak Fark
	Gerçek Durum	845±35	---	846±36	---	845±36	---
MAR	Begg&Greenes	847±43	19±14	845±56	38±28	862±96	76±52
	Çoklu İmputasyon	848±46	23±19	843±64	46±33	*	*
	Çoklu İmputasyon – Ek Değişkenli	847±48	25±20	835±70	53±37	*	*
NMAR	Zhou Alt Sınır	679±35	167±21	421±27	425±27	162±30	677±42
	Zhou Üst Sınır	878±35	32±21	922±30	76±33	971±22	126±36
	Lojistik Regresyon	779±62	72±44	685±90	166±81	610±118	245±102
	Yapay Sinir Ağları – 1	783±57	66±42	698±95	154±82	607±126	249±104
	Yapay Sinir Ağları – 2	788±58	63±43	724±108	134±88	693±165	191±120
	Yapay Sinir Ağları – 3	788±58	63±41	712±100	143±86	670±158	206±117
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark	**Seçicilik**	Mutlak Fark
	Gerçek Durum	845±35	---	846±36	---	845±36	---
MAR	Begg&Greenes	846±42	18±14	848±61	36±29	859±92	213±140
	Çoklu İmputasyon	846±46	24±18	848±69	45±36	*	*
	Çoklu İmputasyon – Ek Değişkenli	843±47	25±20	843±73	51±38	*	*
NMAR	Zhou Alt Sınır	677±35	168±21	421±26	424±27	162±28	497±145
	Zhou Üst Sınır	877±34	32±20	924±32	78±30	971±20	304±152
	Lojistik Regresyon	793±64	61±45	727±107	130±87	711±165	178±138
	Yapay Sinir Ağları – 1	776±58	73±44	691±96	161±83	629±138	160±133
	Yapay Sinir Ağları – 2	778±59	72±45	701±104	155±90	652±157	177±142
	Yapay Sinir Ağları – 3	779±59	71±44	703±106	154±87	654±159	238±133

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalama ve standart sapma değerleri verilmiştir.

* Eksik ölçüm oranı %80 alınarak bu tanımlama katsayısı için yapılan analizler Çoklu İmputasyon yönteminin karşılaştığı sorunlar nedeniyle sonlandırılmıştır. (sıfıra bölme hatası)

4.1.1.2.2 Tanımlama Katsayısının (R^2) 0.40 ve 0.50 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.6'da verilmiştir. Bir önceki tanımlama katsayındaki sonuçlara benzer sonuçlar elde edilmiştir. Buradaki farklılık çoklu imputasyon yaklaşımının sonuç verebilmesi ve bu yaklaşımın Begg ve Greenes yaklaşımı kadar başarılı olmasıdır.



Şekil 4.6 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti doğruluk ölçütleri

Çizelge 4.6 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	751±43	---	752±42	---	749±44	---
MAR	Begg&Greenes	752±50	19±15	756±62	38±29	753±112	84±63
MAR	Çoklu İmputasyon	750±53	26±20	754±78	51±41	728±142	110±81
MAR	Çoklu İmputasyon – Ek Değişkenli	752±56	28±21	751±82	57±41	706±138	111±82
NMAR	Zhou Alt Sınır	601±39	150±20	376±32	376±33	144±32	603±48
NMAR	Zhou Üst Sınır	802±40	51±23	878±34	127±36	950±27	202±44
NMAR	Lojistik Regresyon	696±62	61±37	641±90	124±63	572±104	185±88
NMAR	Yapay Sinir Ağları – 1	704±59	54±35	639±81	118±65	563±112	196±94
NMAR	Yapay Sinir Ağları – 2	705±59	52±34	654±89	113±64	603±129	174±88
NMAR	Yapay Sinir Ağları – 3	706±58	51±35	658±88	109±64	595±128	179±88
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	751±43	---	752±42	---	749±44	---
MAR	Begg&Greenes	750±48	18±14	756±66	41±32	746±107	80±58
MAR	Çoklu İmputasyon	749±53	26±19	752±80	53±42	721±137	105±79
MAR	Çoklu İmputasyon – Ek Değişkenli	751±53	26±19	748±82	55±43	705±139	115±77
NMAR	Zhou Alt Sınır	601±39	150±21	376±31	375±31	145±31	602±47
NMAR	Zhou Üst Sınır	800±39	49±22	878±36	126±36	948±26	200±42
NMAR	Lojistik Regresyon	712±60	49±35	677±98	98±67	625±128	149±96
NMAR	Yapay Sinir Ağları – 1	703±59	54±36	643±87	119±67	569±114	193±93
NMAR	Yapay Sinir Ağları – 2	708±57	52±35	654±89	112±67	591±122	179±86
NMAR	Yapay Sinir Ağları – 3	703±55	54±34	642±87	120±67	586±123	185±87

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.

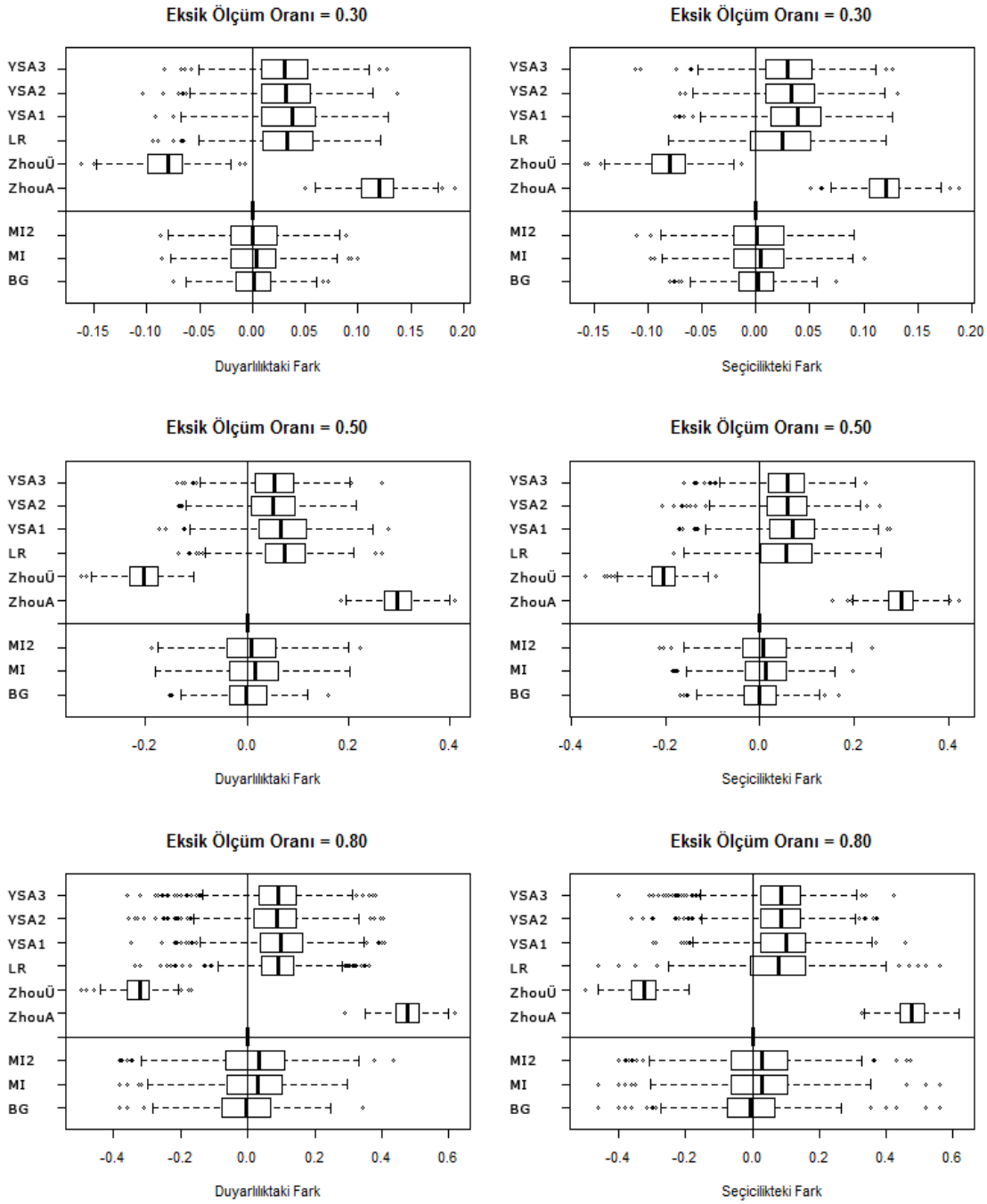
4.1.1.2.3 Tanımlama Katsayısının (R^2) 0.20 ve 0.30 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.7'de verilmiştir. Bir önceki tanımlama katsayındaki sonuçlara benzer sonuçlar elde edilmiştir. Eksik ölçüm oranı arttıkça yaklaşımların hata payları artmıştır.

Çizelge 4.7 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	594±48	---	594±48	---	596±48	---
MAR	Begg&Greenes	594±53	20±15	595±67	40±31	605±122	89±70
	Çoklu İmputasyon	591±57	26±19	583±76	53±36	580±126	96±70
	Çoklu İmputasyon – Ek Değişkenli	592±57	26±20	588±84	57±42	577±144	111±82
NMAR	Zhou Alt Sınır	475±43	119±22	297±36	298±39	118±32	478±49
	Zhou Üst Sınır	676±43	82±24	798±37	204±40	920±32	324±49
	Lojistik Regresyon	561±54	40±27	520±69	83±52	507±94	109±77
	Yapay Sinir Ağları – 1	559±59	43±28	528±74	80±54	499±118	127±84
	Yapay Sinir Ağları – 2	563±56	39±27	545±73	69±45	520±120	118±78
	Yapay Sinir Ağları – 3	565±54	37±25	543±68	67±45	515±112	115±74
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	594±48	---	594±48	---	596±48	---
MAR	Begg&Greenes	594±54	20±15	595±69	41±33	603±132	92±82
	Çoklu İmputasyon	591±58	27±20	583±78	53±39	580±138	102±82
	Çoklu İmputasyon – Ek Değişkenli	592±59	27±20	586±82	56±42	575±147	110±90
NMAR	Zhou Alt Sınır	475±43	119±22	295±38	299±40	118±33	478±50
	Zhou Üst Sınır	675±44	81±24	798±39	204±41	922±30	326±49
	Lojistik Regresyon	571±61	38±26	540±88	79±55	522±137	124±94
	Yapay Sinir Ağları – 1	558±58	43±27	528±77	81±54	510±115	120±76
	Yapay Sinir Ağları – 2	562±55	40±27	541±72	72±48	522±115	112±74
	Yapay Sinir Ağları – 3	565±53	37±25	540±66	70±44	526±115	115±70

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.7 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti doğruluk ölçütleri

4.1.1.2.4 Tanımlama Katsayısının (R^2) 0.07 ve 0.15 arasında olduğu durumlar

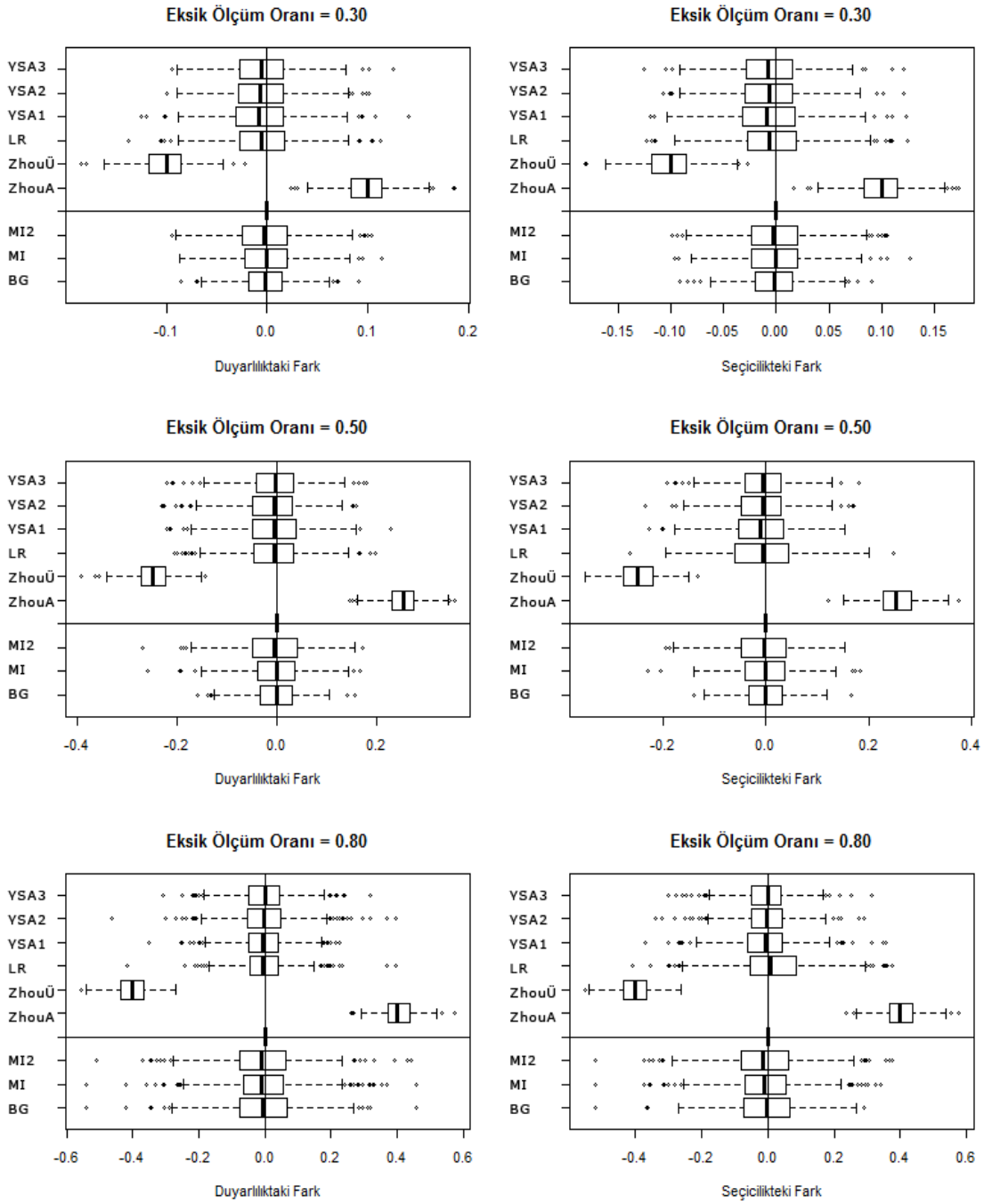
Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.8'de verilmiştir. Yaklaşımların doğruluk ölçütü tahminlerindeki hata değerleri önceki durumlarda elde edilen değerler gibi, eksik ölçüm oranı arttıkça artmaktadır. Burada diğer durumlardan farklı olarak Lojistik Regresyon ve Yapay Sinir

Ağları yaklaşımı daha az hata yapmıştır. Bu durumdan, veri setinin ve tanımlayıcılık katsayısının küçük olması halinde bu yaklaşımların daha başarılı olduğu sonucuna varılmaktadır.

Çizelge 4.8 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	502±53	---	505±51	---	503±50	---
MAR	Begg&Greenes	504±58	20±16	507±70	37±29	507±123	86±73
MAR	Çoklu İmputasyon	505±60	27±20	507±70	46±36	509±119	83±77
MAR	Çoklu İmputasyon – Ek Değişkenli	505±60	27±21	511±78	52±40	510±127	92±79
NMAR	Zhou Alt Sınır	403±46	100±24	253±38	253±38	99±31	404±50
NMAR	Zhou Üst Sınır	603±47	101±24	754±40	249±39	903±30	399±49
NMAR	Lojistik Regresyon	507±56	28±22	514±69	48±41	505±66	54±53
NMAR	Yapay Sinir Ağları – 1	509±58	30±23	512±73	52±41	508±74	60±52
NMAR	Yapay Sinir Ağları – 2	508±57	28±22	513±70	49±40	503±86	66±61
NMAR	Yapay Sinir Ağları – 3	507±56	27±20	510±69	46±38	503±76	60±54
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	502±53	---	505±51	---	503±50	---
MAR	Begg&Greenes	504±58	20±16	506±70	38±28	508±122	86±71
MAR	Çoklu İmputasyon	504±60	27±21	507±70	45±35	509±118	83±75
MAR	Çoklu İmputasyon – Ek Değişkenli	505±60	27±22	511±79	52±39	514±128	92±80
NMAR	Zhou Alt Sınır	403±47	100±24	251±39	254±38	100±31	403±50
NMAR	Zhou Üst Sınır	603±47	101±24	754±38	249±39	903±32	399±51
NMAR	Lojistik Regresyon	507±65	30±25	511±88	58±44	487±126	92±79
NMAR	Yapay Sinir Ağları – 1	510±60	31±23	516±73	52±40	510±85	68±61
NMAR	Yapay Sinir Ağları – 2	508±57	28±22	513±66	46±38	509±74	61±55
NMAR	Yapay Sinir Ağları – 3	508±58	27±20	511±63	44±36	508±70	58±51

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.8 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti doğruluk ölçütleri

4.1.2 Eksik ölçüm mekanizması rastgele kabul edilmediğine (NMAR)

4.1.2.1 Üretilmiş Büyük Veri Seti Kullanılarak Yapılan Karşılaştırmalar

1000'lik olarak üretilmiş büyük veri setlerindeki üretilen değişkenlerin, altın standardı açıklama oranına göre karşılaştırmalı sonuçları aşağıdaki çizelge ve şekillerde verilmiştir.

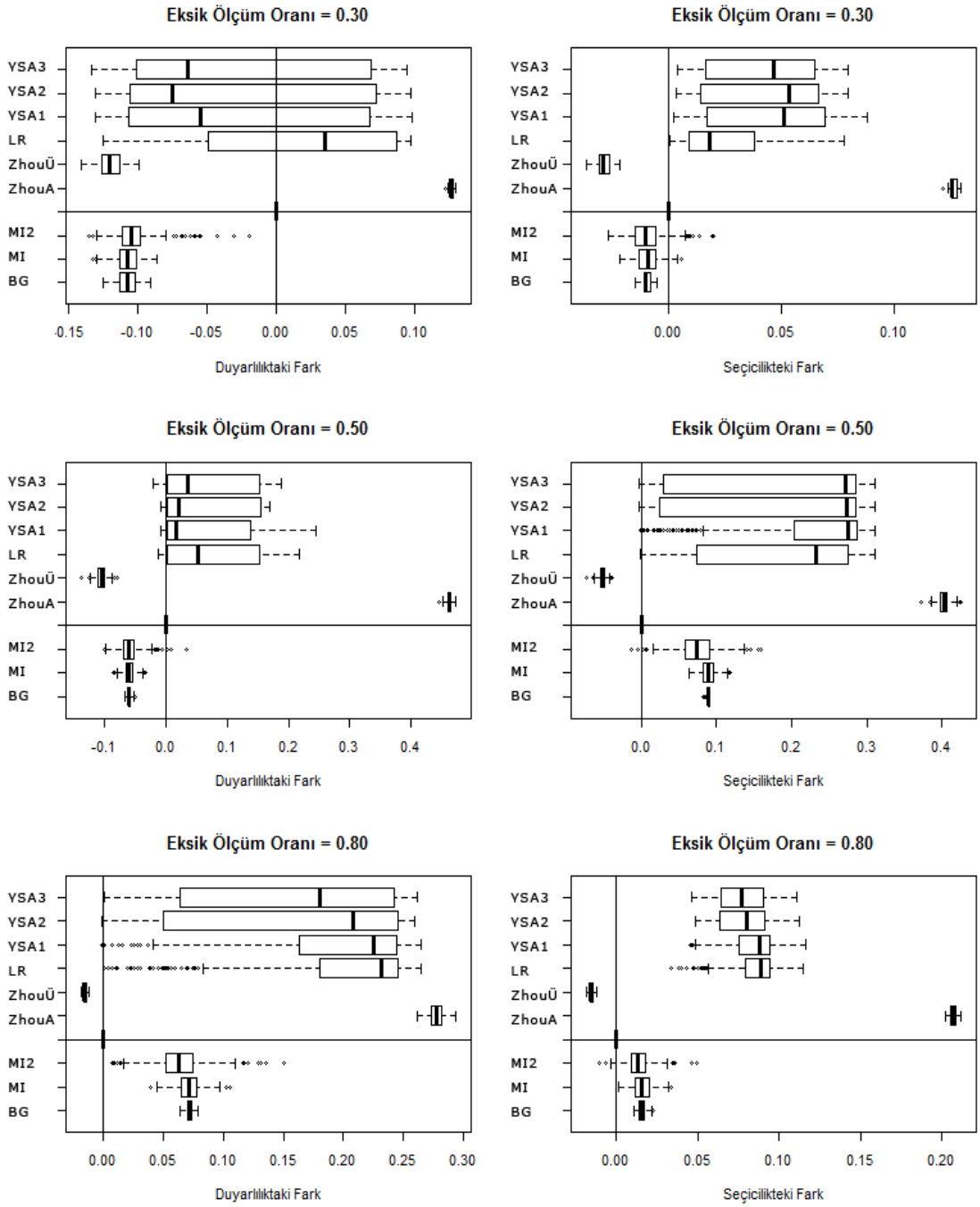
4.1.2.1.1 Tanımlama Katsayısının (R^2) 0.60 ve 0.70 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.9'da ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.9'da verilmiştir. Elde edilen sonuçlara göre eksik ölçüm oranı arttıkça yaklaşımların hata değerlerinde değişimler gözlenmiştir. Eksik ölçüm mekanizmasının rastgele olduğu durumlar için önerilen Begg&Greenes ve Çoklu İmputasyon yaklaşımları eksik ölçüm oranın artması ile ufak değişimler gösterirken, LR ve YSA yaklaşımlarında dikkate değer değişimler gözlenmiştir. Beklendiği üzere eksik ölçüm mekanizmasının değişmesi Begg&Greenes ve Çoklu İmputasyon yaklaşımlarının hata paylarının artmasına neden olmuştur. LR ve YSA yaklaşımları için eksik ölçüm oranı %30 ve %80 iken seçicilik tahminindeki hata değeri, duyarlılık tahminindeki hata değerinden daha düşük; %50 eksik ölçümde ise duyarlılık tahminindeki hata değeri, seçicilik tahminindeki hata değerinden daha düşük bulunmuştur.

Çizelge 4.9 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	846±11	---	845±11	---	847±12	---
MAR	Begg&Greenes	953±4	107±7	906±9	60±2	775±14	72±3
	Çoklu İmputasyon	953±6	107±8	905±12	60±9	775±18	71±10
	Çoklu İmputasyon – Ek Değişkenli	950±12	104±13	905±19	60±16	783±23	63±20
NMAR	Zhou Alt Sınır	719±9	127±2	383±8	462±4	569±6	277±6
	Zhou Üst Sınır	965±3	119±8	950±4	105±7	862±11	15±1
	Lojistik Regresyon	830±71	67±27	772±67	74±66	642±60	205±60
	Yapay Sinir Ağları – 1	873±83	81±32	786±70	60±69	654±72	192±71
	Yapay Sinir Ağları – 2	874±86	86±27	779±72	67±71	693±98	153±97
	Yapay Sinir Ağları – 3	868±83	82±26	775±73	71±70	693±91	154±91
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	846±11	---	845±11	---	847±12	---
MAR	Begg&Greenes	856±9	10±2	756±12	89±1	831±14	16±2
	Çoklu İmputasyon	855±10	10±5	756±15	89±10	831±15	16±6
	Çoklu İmputasyon – Ek Değişkenli	856±11	11±6	771±29	74±26	833±15	14±7
NMAR	Zhou Alt Sınır	719±9	127±2	442±4	403±7	640±13	206±2
	Zhou Üst Sınır	875±8	29±3	898±7	53±4	862±11	15±1
	Lojistik Regresyon	820±23	26±20	665±111	180±110	761±20	86±13
	Yapay Sinir Ağları – 1	801±28	45±26	614±87	231±87	762±20	85±13
	Yapay Sinir Ağları – 2	803±27	43±26	659±124	187±123	769±19	78±15
	Yapay Sinir Ağları – 3	804±25	42±25	659±120	186±121	769±20	77±14

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.9 Tanımlama katsayısı 0.60 – 0.70 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.2.1.2 Tanımlama Katsayısının (R^2) 0.40 ve 0.50 arasında olduğu durumlar

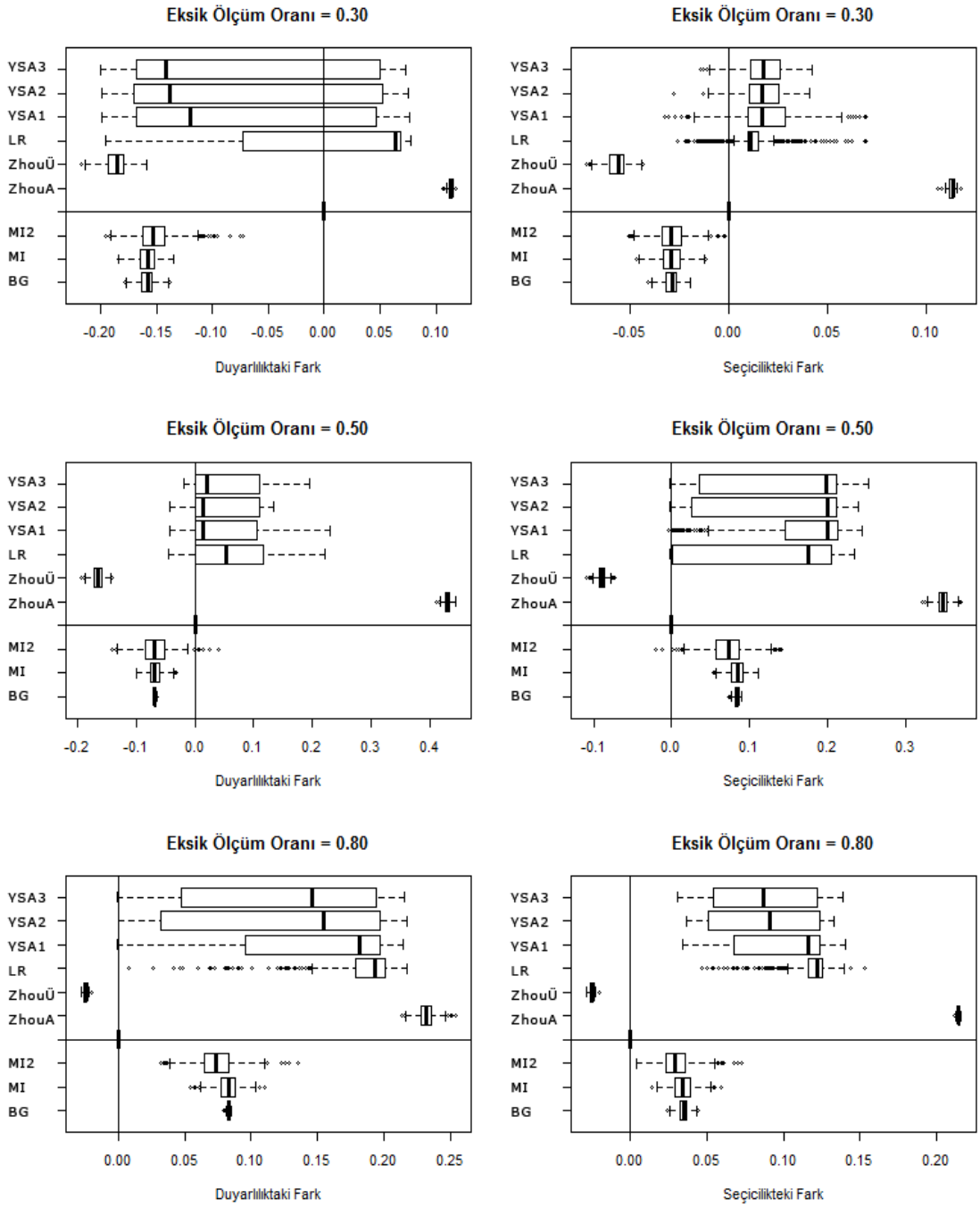
Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.10'da ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil

4.10'da verilmiştir. Elde edilen sonuçlar tanımlama katsayısının 0.6 civarında olduğu durum ile benzer olarak bulunmuştur.

Çizelge 4.10 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	754±14	---	754±14	---	754±13	---
MAR	Begg&Greenes	912±7	158±7	822±14	68±1	671±14	83±1
MAR	Çoklu İmputasyon	913±10	159±10	821±18	68±12	671±16	83±9
MAR	Çoklu İmputasyon – Ek Değişkenli	906±18	151±17	822±28	69±23	680±20	74±15
NMAR	Zhou Alt Sınır	641±12	113±2	324±9	430±5	521±7	233±6
NMAR	Zhou Üst Sınır	941±4	186±10	918±5	164±9	778±12	25±1
NMAR	Lojistik Regresyon	744±85	76±34	694±61	63±55	570±30	184±30
NMAR	Yapay Sinir Ağları – 1	828±101	111±58	708±58	48±55	609±70	145±68
NMAR	Yapay Sinir Ağları – 2	827±106	113±60	710±53	45±51	635±82	118±81
NMAR	Yapay Sinir Ağları – 3	827±103	112±59	705±55	50±51	631±74	123±74
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	754±14	---	754±14	---	754±13	---
MAR	Begg&Greenes	784±10	29±3	670±12	84±3	719±17	35±3
MAR	Çoklu İmputasyon	784±11	30±6	669±14	84±9	719±18	35±7
MAR	Çoklu İmputasyon – Ek Değişkenli	783±13	29±8	682±27	72±24	723±19	31±10
NMAR	Zhou Alt Sınır	641±12	113±2	405±6	348±8	540±14	214±1
NMAR	Zhou Üst Sınır	811±9	57±5	843±8	89±6	778±12	25±1
NMAR	Lojistik Regresyon	743±20	14±10	629±90	125±90	636±22	118±14
NMAR	Yapay Sinir Ağları – 1	737±19	20±13	587±71	167±71	655±35	99±31
NMAR	Yapay Sinir Ağları – 2	737±15	18±9	608±89	145±89	666±38	88±35
NMAR	Yapay Sinir Ağları – 3	736±14	19±9	609±86	145±87	666±36	88±33

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.

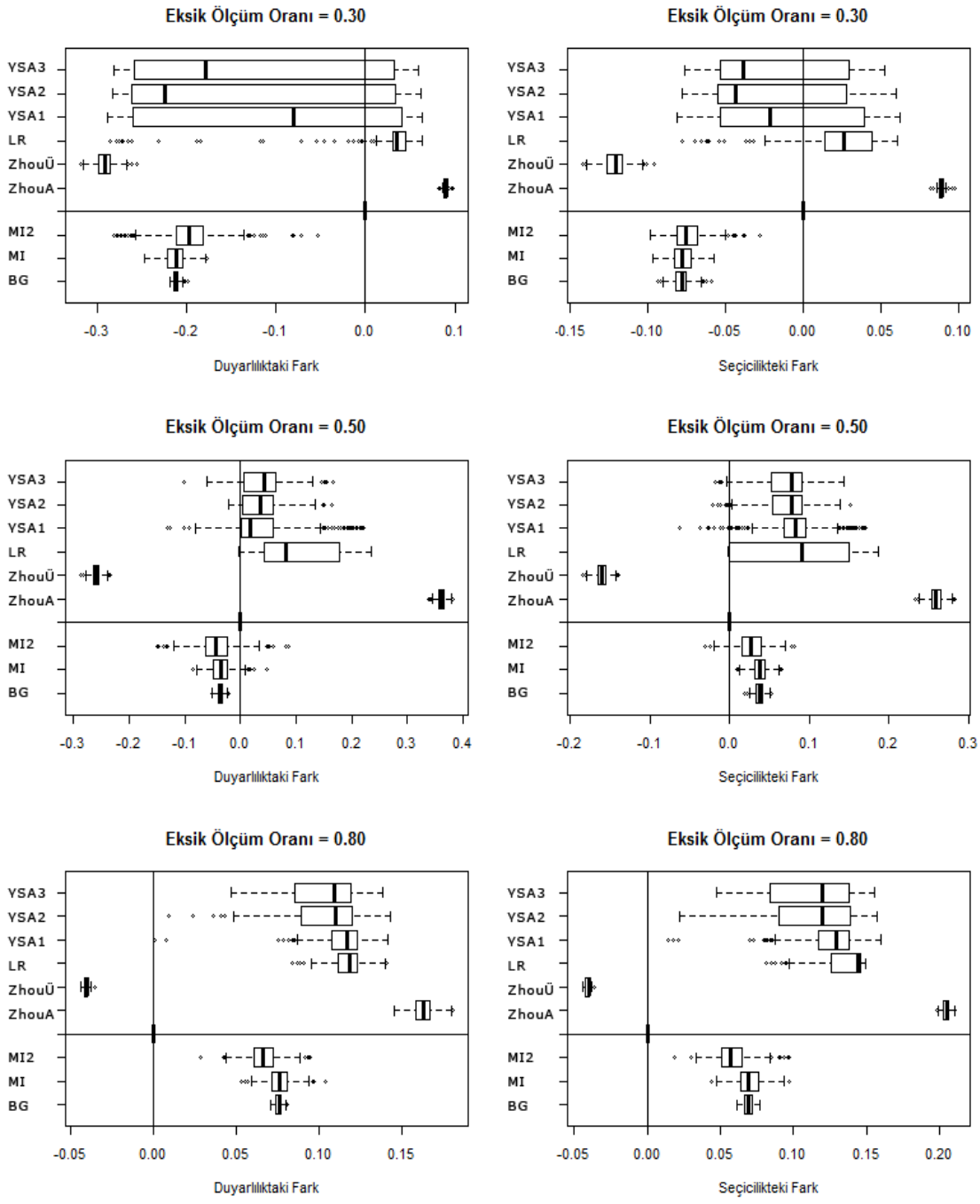


Şekil 4.10 Tanımlama katsayısı 0.40 – 0.50 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.2.1.3 Tanımlama Katsayısının (R^2) 0.20 ve 0.30 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.11'de ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.11'de verilmiştir. Şekilden de görülebileceği gibi eksik ölçümünün %30 olması durumunda YSA yaklaşımı yüksek yaygınlık değeri gösteren tahminlerde bulunurken,

eksik ölçüm oranın artmasıyla bu yaygınlık azalma göstermiştir. Eksik ölçüm oranın yüksek olması (%80) LR ve YSA yaklaşımlarının tahminlerindeki hatanın artmasına neden olmaktadır.



Şekil 4.11 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti doğruluk ölçütleri

Çizelge 4.11 Tanımlama katsayısı 0.20 – 0.30 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	596±16	---	596±15	---	596±15	---
MAR	Begg&Greenes	808±13	212±3	633±20	36±5	527±18	76±2
MAR	Çoklu İmputasyon	808±18	212±13	633±27	37±17	527±19	76±7
MAR	Çoklu İmputasyon – Ek Değişkenli	793±32	197±29	640±39	46±28	538±21	66±9
NMAR	Zhou Alt Sınır	506±14	89±2	235±8	361±7	392±13	163±6
NMAR	Zhou Üst Sınır	887±6	291±10	856±7	259±9	637±13	40±2
NMAR	Lojistik Regresyon	567±51	44±37	489±72	107±70	461±21	118±9
NMAR	Yapay Sinir Ağları – 1	702±141	140±107	557±62	45±56	471±24	114±15
NMAR	Yapay Sinir Ağları – 2	732±138	162±107	560±37	38±33	483±32	104±21
NMAR	Yapay Sinir Ağları – 3	719±139	150±107	556±40	43±34	485±33	103±20
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	596±16	---	596±15	---	597±15	---
MAR	Begg&Greenes	674±11	78±5	559±10	38±6	528±18	69±3
MAR	Çoklu İmputasyon	673±12	77±7	559±13	38±10	527±20	70±9
MAR	Çoklu İmputasyon – Ek Değişkenli	670±14	74±10	569±20	28±15	540±22	58±11
NMAR	Zhou Alt Sınır	506±14	89±2	337±7	259±8	392±13	205±2
NMAR	Zhou Üst Sınır	717±9	121±7	757±8	160±7	637±14	40±2
NMAR	Lojistik Regresyon	569±25	30±16	516±71	81±69	461±21	135±15
NMAR	Yapay Sinir Ağları – 1	605±46	43±15	517±38	81±36	472±22	125±19
NMAR	Yapay Sinir Ağları – 2	616±43	44±15	527±31	69±31	483±33	114±29
NMAR	Yapay Sinir Ağları – 3	612±44	42±16	527±31	70±31	490±34	112±29

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.

4.1.2.1.4 Tanımlama Katsayısının (R^2) 0.07 ve 0.15 arasında olduğu durumlar

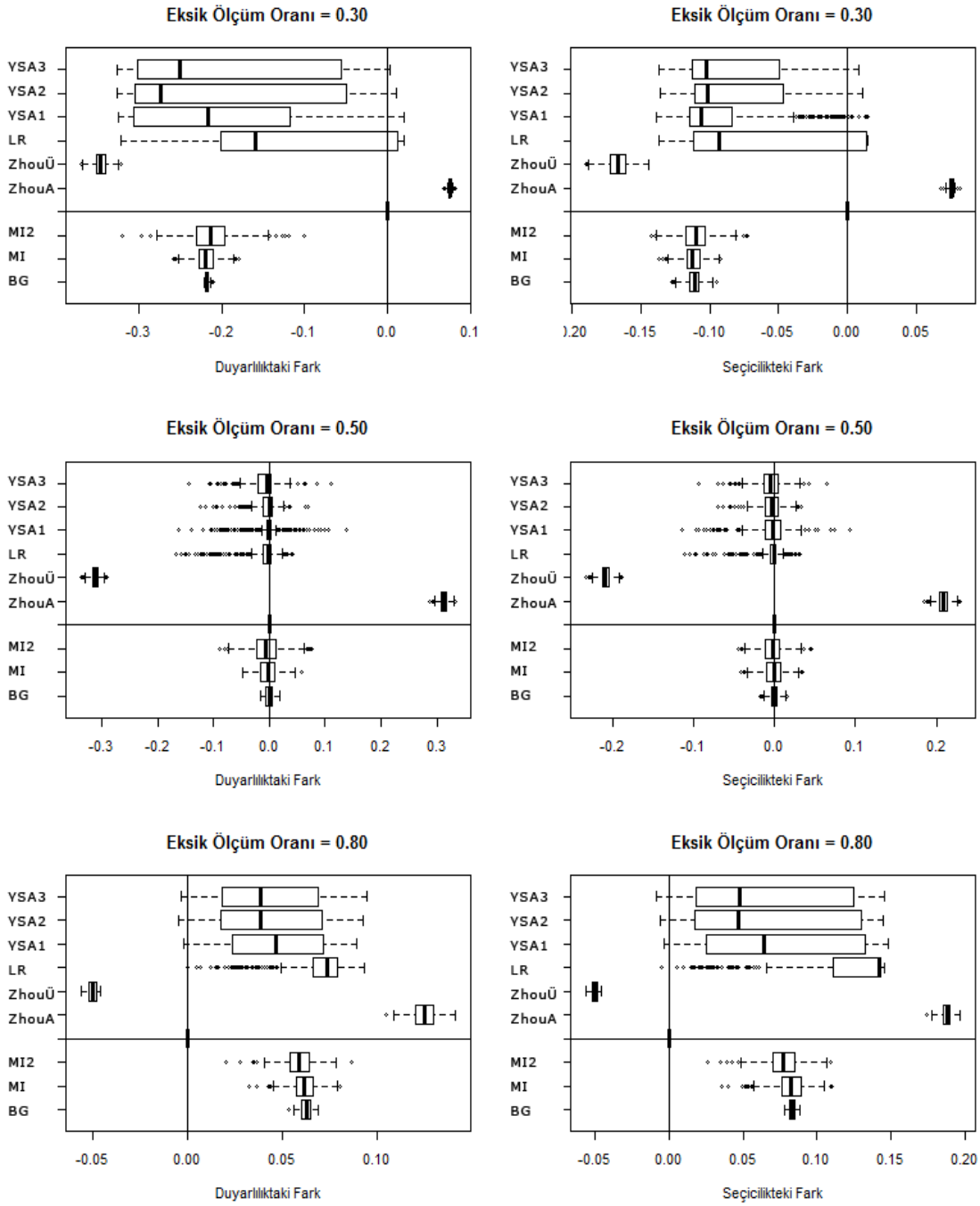
Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.12'de ve yaklaşımların gerçek durum ile farklarının sonuçları Şekil 4.12'de verilmiştir. Bu bölümde elde edilen sonuçlara göre LR ve YSA yaklaşımlarının özellikle eksik ölçüm oranın %50 olması durumunda en düşük hata değerlerine sahip

olduğu ve tüm eksik ölçüm oranlarında bu yaklaşımların MAR yaklaşımlarından daha düşük hata değerlerine sahip olduğu görülmüştür.

Çizelge 4.12 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	499±16	---	500±15	---	500±15	---
MAR	Begg&Greenes	717±17	218±2	500±22	5±4	437±13	62±2
	Çoklu İmputasyon	719±22	219±13	502±24	15±11	438±15	61±7
	Çoklu İmputasyon – Ek Değişkenli	713±33	214±28	505±33	22±17	441±15	59±8
NMAR	Zhou Alt Sınır	424±13	75±2	187±7	312±8	375±10	125±6
	Zhou Üst Sınır	846±8	346±8	812±7	313±8	550±14	50±2
	Lojistik Regresyon	603±116	115±105	511±37	18±29	431±19	69±17
	Yapay Sinir Ağları – 1	701±108	202±107	504±33	14±25	453±29	47±26
	Yapay Sinir Ağları – 2	691±127	192±124	505±25	12±16	456±30	44±28
	Yapay Sinir Ağları – 3	693±119	194±118	508±29	17±20	457±28	43±26
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	499±16	---	500±15	---	500±15	---
MAR	Begg&Greenes	611±10	112±6	500±9	5±4	416±17	83±2
	Çoklu İmputasyon	612±11	112±7	501±10	9±7	418±21	82±11
	Çoklu İmputasyon – Ek Değişkenli	610±14	110±11	502±15	11±9	422±21	77±12
NMAR	Zhou Alt Sınır	424±13	75±2	291±8	208±8	312±12	187±3
	Zhou Üst Sınır	666±8	167±8	708±8	208±8	550±14	50±2
	Lojistik Regresyon	550±65	64±49	505±23	9±17	377±39	123±36
	Yapay Sinir Ağları – 1	591±38	92±35	504±20	14±17	425±55	75±52
	Yapay Sinir Ağları – 2	581±42	82±39	504±13	11±10	430±57	69±54
	Yapay Sinir Ağları – 3	584±38	84±36	505±15	13±12	434±53	66±51

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.12 Tanımlama katsayısı 0.07 – 0.15 arasında olan 1000'lik veri seti doğruluk ölçütleri

4.1.2.2 Üretilmiş Küçük Veri Seti Kullanılarak Yapılan Karşılaştırmalar

Küçük veri seti karşılaştırmalarında büyük veri setlerinde yapıldığı gibi yaklaşımlar, değişen eksik ölçüm oranlarında (%30, %50 ve %80) ve değişen tanımlayıcılık katsayılarında karşılaştırılmıştır.

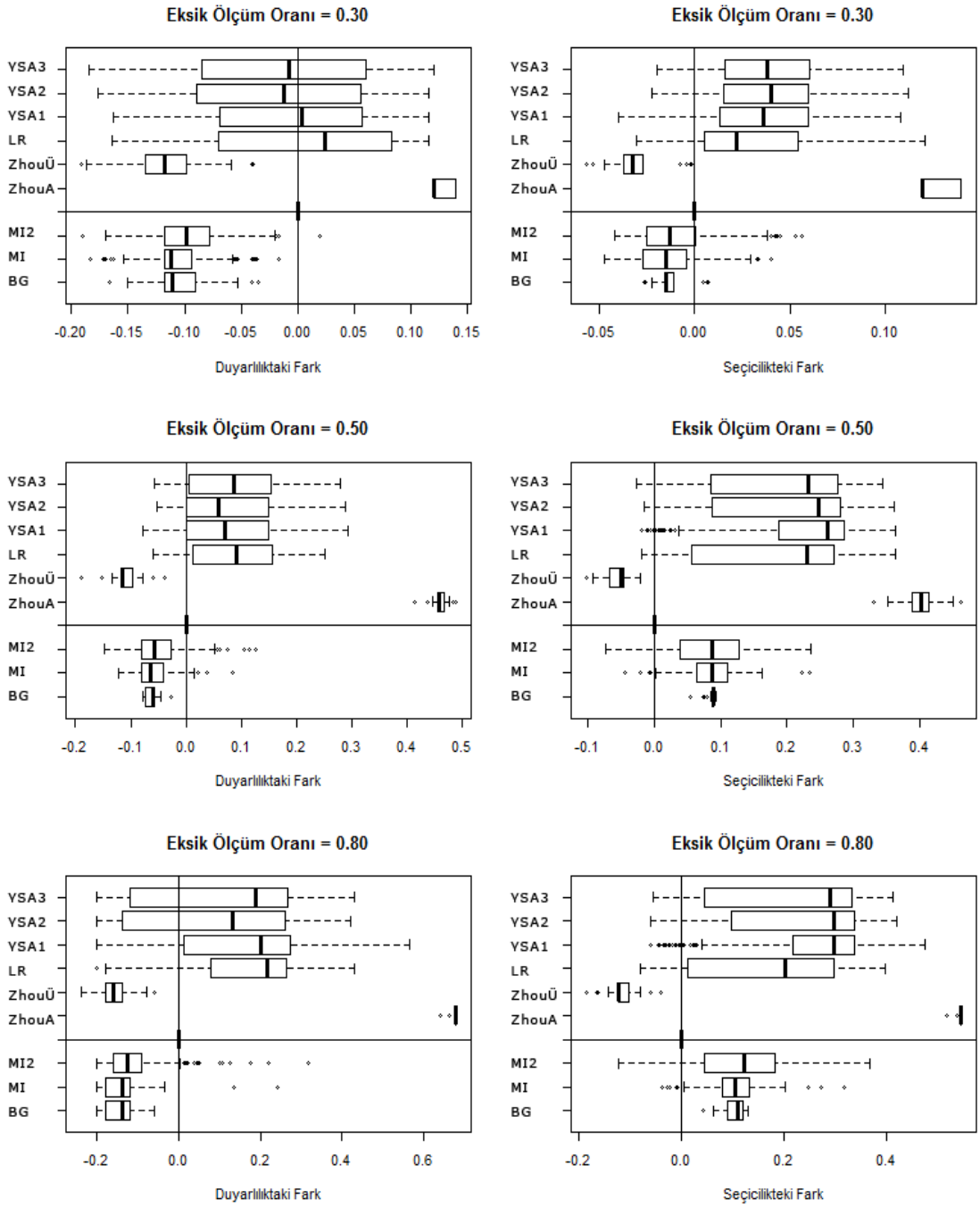
4.1.2.2.1 Tanımlama Katsayısının (R^2) 0.60 ve 0.70 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.13'de verilmiştir. Büyük veri setinde olduğu gibi burada da yaklaşımlar eksik ölçüm oranına göre değişen hata payları ile gerçek durumu tahmin edebilmiştir. Büyük veri setinden farklı olarak burada yaklaşımların tahmin değerlerindeki standart sapma değerleri artmıştır. Eksik ölçüm oranındaki artışın ise bazı yaklaşımların hata paylarının artmasına neden olduğu görülmüştür.

Çizelge 4.13 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	848±36	---	847±34	---	844±34	---
MAR	Begg&Greenes	953±18	106±22	908±30	62±10	776±41	149±31
	Çoklu İmputasyon	954±23	106±26	907±42	61±27	778±50	149±32
	Çoklu İmputasyon – Ek Değişkenli	949±29	101±28	898±53	59±32	779±56	123±49
NMAR	Zhou Alt Sınır	722±29	126±9	385±24	462±11	99±1	675±11
	Zhou Üst Sınır	966±12	118±26	952±14	106±22	999±4	152±32
	Lojistik Regresyon	841±84	73±33	762±86	93±68	692±151	198±88
	Yapay Sinir Ağları – 1	854±75	63±36	770±87	86±71	707±179	202±108
	Yapay Sinir Ağları – 2	865±79	70±36	771±82	85±68	771±192	190±84
	Yapay Sinir Ağları – 3	859±74	69±35	766±83	90±68	743±192	200±86
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	848±36	---	847±34	---	847±33	---
MAR	Begg&Greenes	860±30	14±4	761±41	86±10	745±45	102±23
	Çoklu İmputasyon	861±34	18±11	760±54	87±34	744±60	104±44
	Çoklu İmputasyon – Ek Değişkenli	858±36	18±11	763±73	87±55	735±110	128±77
NMAR	Zhou Alt Sınır	722±29	126±9	442±13	404±21	230±6	544±7
	Zhou Üst Sınır	879±27	31±10	899±23	52±13	958±11	110±23
	Lojistik Regresyon	817±42	32±28	666±111	181±111	675±140	184±125
	Yapay Sinir Ağları – 1	810±40	39±27	622±88	224±89	591±120	260±114
	Yapay Sinir Ağları – 2	809±39	41±27	650±112	197±107	617±141	238±130
	Yapay Sinir Ağları – 3	808±40	40±27	656±107	191±104	635±145	221±135

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.13 Tanımlama katsayısı 0.60 – 0.70 arasında olan 100'lük veri seti doğruluk ölçütleri

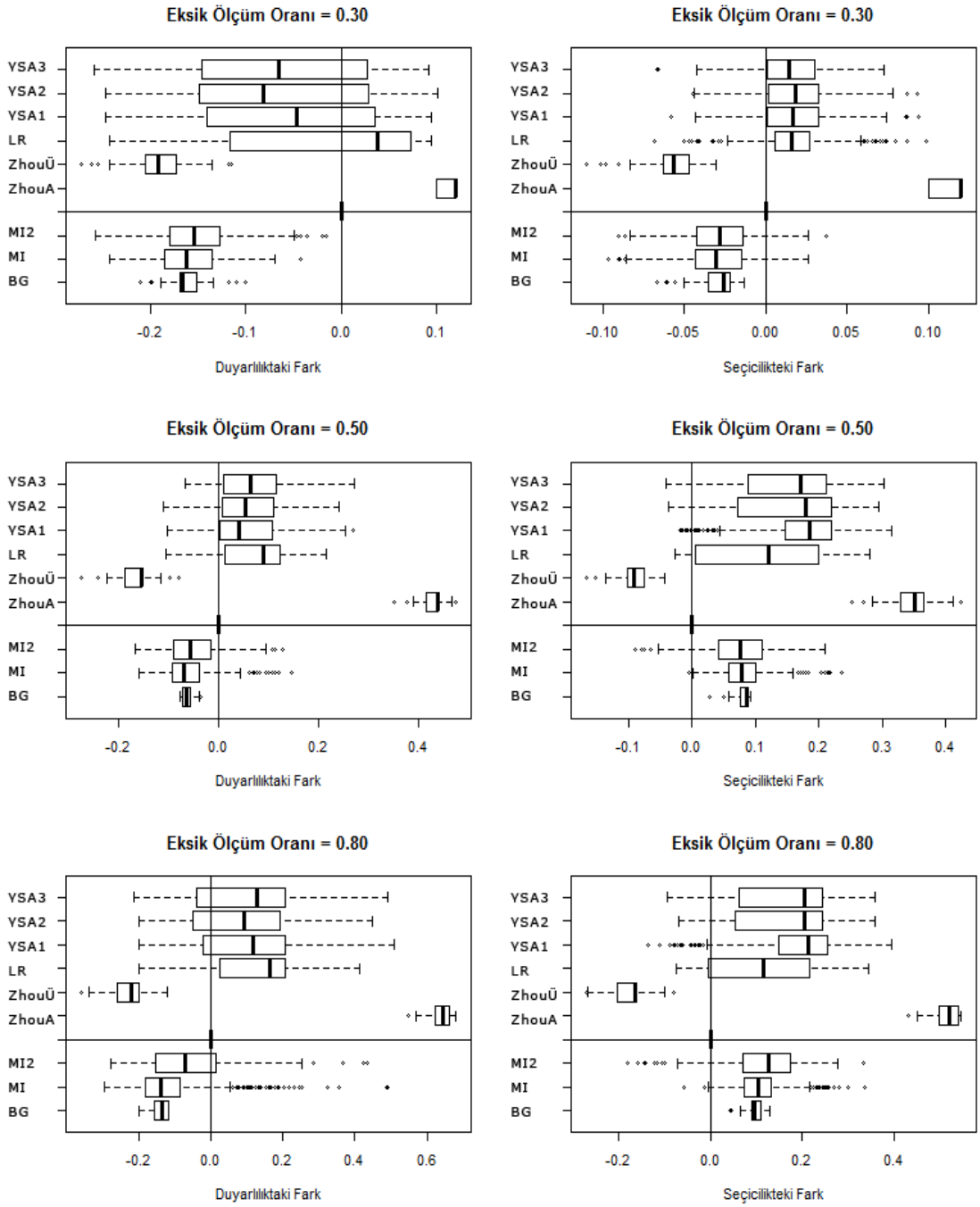
4.1.2.2.2 Tanımlama Katsayısının (R^2) 0.40 ve 0.50 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.14'te verilmiştir. Bir önceki tanımlama katsayındaki sonuçlara benzer sonuçlar elde edilmiştir.

Çizelge 4.14 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları (x10^{-3})

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	748±46	---	755±43	---	750±42	---
MAR	Begg&Greenes	908±26	159±23	819±47	65±9	892±53	142±27
	Çoklu İmputasyon	908±35	160±33	817±67	68±36	860±125	141±59
	Çoklu İmputasyon – Ek Değişkenli	899±44	150±40	805±72	63±38	807±131	110±72
NMAR	Zhou Alt Sınır	635±39	113±10	326±25	429±19	93±10	644±26
	Zhou Üst Sınır	938±15	190±33	917±18	162±27	982±8	232±37
	Lojistik Regresyon	765±107	87±50	681±75	82±53	631±128	151±82
	Yapay Sinir Ağları – 1	801±96	88±61	698±80	67±57	645±145	144±99
	Yapay Sinir Ağları – 2	813±100	96±61	697±73	68±51	677±151	137±82
	Yapay Sinir Ağları – 3	811±95	91±65	689±73	73±54	659±152	148±87
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	748±46	---	755±43	---	750±42	---
MAR	Begg&Greenes	779±34	31±13	672±36	82±8	656±35	94±16
	Çoklu İmputasyon	779±37	32±19	673±52	82±36	641±66	109±52
	Çoklu İmputasyon – Ek Değişkenli	777±38	31±18	680±65	80±45	633±87	126±66
NMAR	Zhou Alt Sınır	635±39	113±10	406±17	348±26	219±10	518±25
	Zhou Üst Sınır	807±29	59±17	844±26	89±18	928±13	178±31
	Lojistik Regresyon	733±47	21±17	644±95	112±92	642±121	127±100
	Yapay Sinir Ağları – 1	731±42	24±18	582±66	173±66	564±94	192±86
	Yapay Sinir Ağları – 2	731±38	24±17	605±84	151±83	589±107	172±97
	Yapay Sinir Ağları – 3	732±38	22±17	603±77	153±76	587±109	173±97

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.14 Tanımlama katsayısı 0.40 – 0.50 arasında olan 100'lük veri seti doğruluk ölçütleri

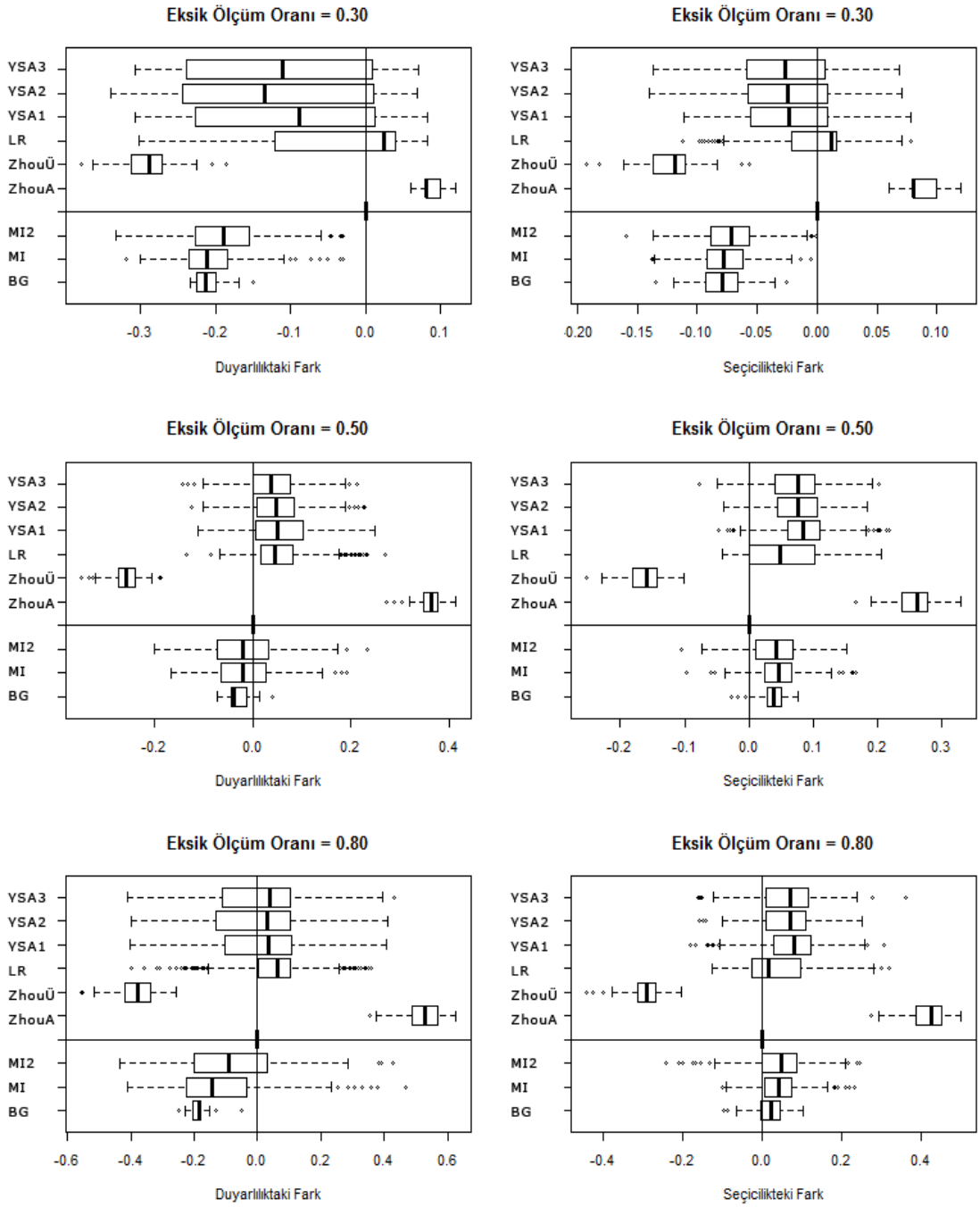
4.1.2.2.3 Tanımlama Katsayısının (R^2) 0.20 ve 0.30 arasında olduğu durumlar

Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.15'te verilmiştir. Özellikle LR ve YSA yaklaşımlarının seçicilik tahminindeki hata değerleri düşüş göstermiştir.

Çizelge 4.15 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	596±46	---	598±49	---	599±48	---
MAR	Begg&Greenes	808±40	212±14	633±66	36±17	578±14	190±22
	Çoklu İmputasyon	805±59	209±41	617±89	54±39	557±43	162±92
	Çoklu İmputasyon – Ek Değişkenli	785±69	189±54	618±97	61±45	559±60	144±100
NMAR	Zhou Alt Sınır	508±39	88±10	236±27	362±22	181±14	529±45
	Zhou Üst Sınır	887±20	291±29	856±21	258±29	891±12	378±48
	Lojistik Regresyon	636±124	83±88	541±71	64±53	563±83	104±74
	Yapay Sinir Ağları – 1	701±122	123±102	539±85	69±57	527±70	121±83
	Yapay Sinir Ağları – 2	717±131	138±106	548±75	59±47	536±64	122±74
	Yapay Sinir Ağları – 3	711±126	131±105	555±76	56±44	536±70	117±73
		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	596±46	---	598±49	---	596±50	---
MAR	Begg&Greenes	675±31	79±15	560±31	38±17	529±63	33±25
	Çoklu İmputasyon	673±37	77±22	553±43	48±30	526±64	54±40
	Çoklu İmputasyon – Ek Değişkenli	668±39	72±25	559±52	48±31	536±67	67±47
NMAR	Zhou Alt Sınır	508±39	88±10	337±23	261±26	393±45	418±35
	Zhou Üst Sınır	717±27	121±19	758±24	160±25	637±49	292±37
	Lojistik Regresyon	597±57	28±23	542±62	59±56	470±64	70±58
	Yapay Sinir Ağları – 1	618±50	39±27	513±43	86±43	491±64	91±56
	Yapay Sinir Ağları – 2	621±51	39±27	524±44	76±41	503±68	81±52
	Yapay Sinir Ağları – 3	622±49	38±27	527±46	74±40	504±67	82±54

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.15 Tanımlama katsayısı 0.20 – 0.30 arasında olan 100'lük veri seti doğruluk ölçütleri

4.1.2.2.4 Tanımlama Katsayısının (R^2) 0.07 ve 0.15 arasında olduğu durumlar

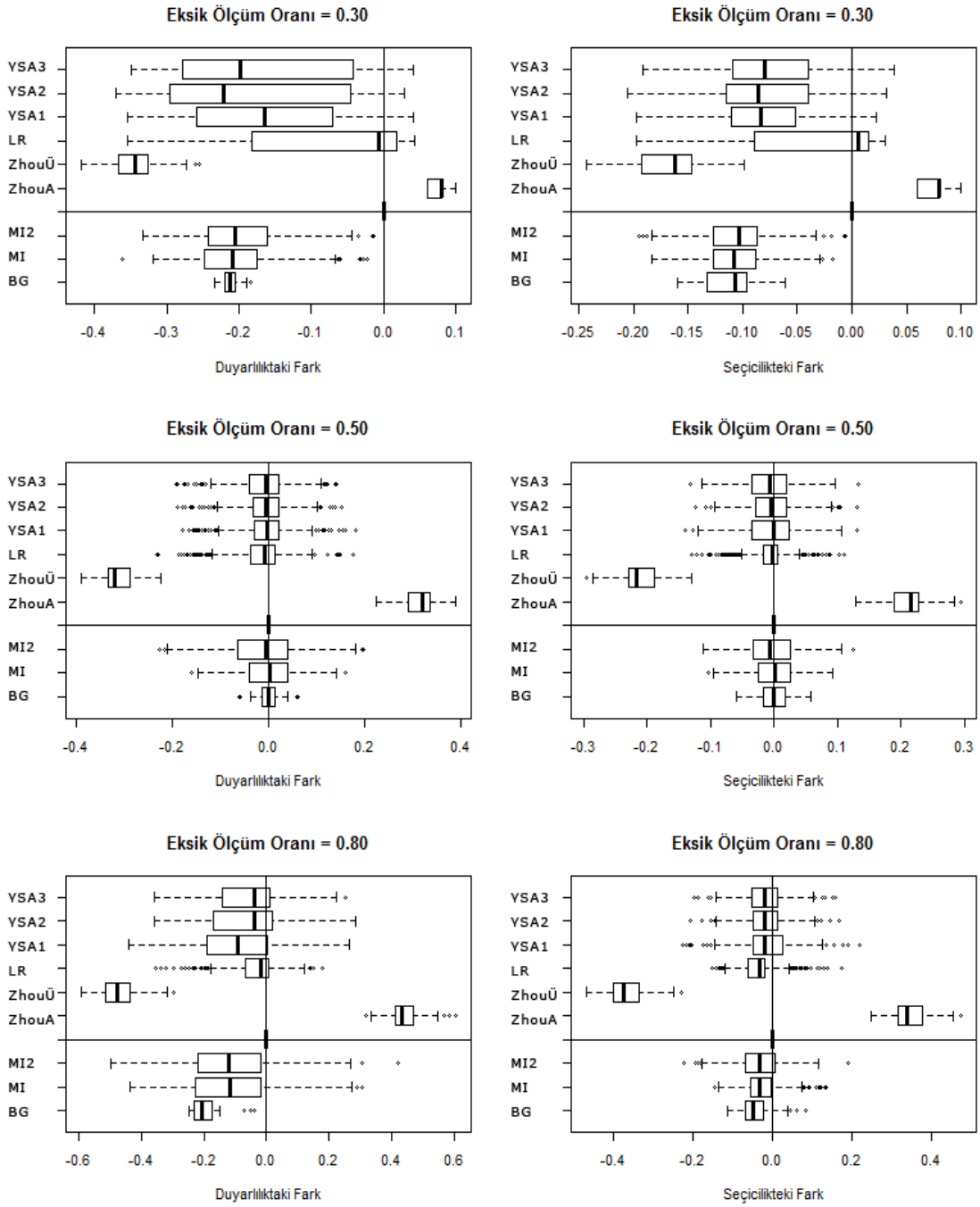
Bu durumdaki doğruluk ölçütlerinin tahmininde kullanılan yaklaşımların sonuçları Çizelge 4.16'da verilmiştir.

Çizelge 4.16 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti için yaklaşımların doğruluk ölçütü tahminleri ve bu değerlerin standart sapmaları ($x10^{-3}$)

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark	Duyarlılık	Mutlak Fark
	Gerçek Durum	498±51	---	500±50	---	501±51	---
MAR	Begg&Greenes	713±54	215±12	500±69	15±12	700±79	199±41
	Çoklu İmputasyon	705±78	207±54	499±74	45±32	617±154	150±101
	Çoklu İmputasyon – Ek Değişkenli	699±75	201±59	511±93	59±45	618±168	161±112
NMAR	Zhou Alt Sınır	423±43	74±10	186±24	314±27	59±12	443±42
	Zhou Üst Sınır	844±25	346±29	814±24	314±27	976±5	475±49
	Lojistik Regresyon	576±127	91±108	516±70	37±42	538±81	63±63
	Yapay Sinir Ağları – 1	662±116	165±106	507±72	40±38	591±119	124±86
	Yapay Sinir Ağları – 2	677±127	180±121	507±70	38±35	566±112	101±82
	Yapay Sinir Ağları – 3	666±127	170±116	512±68	41±37	562±105	94±78

		Eksik Ölçüm Oranı = %30		Eksik Ölçüm Oranı = %50		Eksik Ölçüm Oranı = %80	
	Yaklaşımlar	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark	Seçicilik	Mutlak Fark
	Gerçek Durum	498±51	---	500±50	---	501±51	---
MAR	Begg&Greenes	610±32	112±19	500±30	15±13	547±23	49±26
	Çoklu İmputasyon	605±40	108±27	499±33	29±22	527±38	44±29
	Çoklu İmputasyon – Ek Değişkenli	604±38	106±31	505±44	34±25	534±50	52±39
NMAR	Zhou Alt Sınır	423±43	74±10	290±25	210±26	155±15	347±37
	Zhou Üst Sınır	666±24	168±27	710±25	210±26	869±13	368±39
	Lojistik Regresyon	531±73	49±47	508±54	22±26	530±52	44±29
	Yapay Sinir Ağları – 1	578±53	82±42	505±35	33±27	515±40	47±39
	Yapay Sinir Ağları – 2	578±55	82±47	505±34	31±25	518±37	42±33
	Yapay Sinir Ağları – 3	574±56	78±45	508±37	32±25	519±42	44±34

Mutlak Fark: Her bir tekrar için gerçek durum ile yaklaşım tahmini arasındaki mutlak fark değerinin ortalaması ve standart sapma değerleri verilmiştir.



Şekil 4.16 Tanımlama katsayısı 0.07 – 0.15 arasında olan 100'lük veri seti doğruluk ölçütleri

4.1.3 Gerçek Veri Seti Kullanılarak Yapılmış Çalışma Sonuçları

Nükleer tıp alanında Sukan ve ark.[50] tarafından yapılmış bir çalışmada hiperparatiroidi tanısını koymada kullanılan yeni bir tanı yöntemi, altın standart olarak

kabul edilmiş patoloji sonucu ile karşılaştırılmıştır. Bu çalışmada bulunan tüm bireylerin patoloji sonuçları olmadığı için çalışma seçilmiş örnekler üzerinden değerlendirilmiştir. Patoloji sonucu olmayan bireyleri de dikkate almak için Begg & Greenes ve Zhou tarafından önerilen yöntemlerle, bu tezde sunulan alternatif yaklaşımlar doğruluk ölçütlerini belirlemede kullanılmıştır. Elde edilen sonuçlar Çizelge 4.17'de verilmiştir.

Çizelge 4.17 Gerçek veri seti kullanılarak yapılan karşılaştırma sonuçları

Yaklaşımlar	Duyarlılık	Seçicilik
Düzeltmesiz Sonuç	0.794±0.053	0.514±0.142
Begg & Greenes	0.637±0.059	0.695±0.110
Zhou Alt Sınır	0.489±0.042	0.263±0.090
Zhou Üst Sınır	0.836±0.044	0.858±0.047
Lojistik Regresyon	0.605±0.066	0.595±0.166
Yapay Sinir Ağları	0.634±0.053	0.699±0.108
Çoklu İmputasyon	0.627±0.062	0.682±0.129

Çizelge 4.17'den de görüldüğü gibi düzeltmesiz doğruluk ölçütleri; Duyarlılık: 0.794, Seçicilik: 0.514 olarak elde edilmiştir. Kosinski ve Barnhart'ın önerdiği Lojistik Regresyon yaklaşımında eksik veri mekanizma bileşeni için yapılan modelleme sonucunda mekanizmanın MAR olduğu görülmüştür. Bu durumda Begg&Greenes yaklaşımı sonuçları dikkate alınacak olursa, önerilen yeni tanı yönteminin düzeltilmiş doğruluk ölçütleri; duyarlılık: 0.63 civarında ve seçicilik: 0.69 civarında olabileceği görülür. Bu değerler ile düzeltmesiz değerler arasındaki fark düzeltmenin ne kadar gerekli olduğunu göstermektedir.

4.2 Kesin olmayan altın standart yanlılığının düzeltilmesinde kullanılan ve önerilen yaklaşımların karşılaştırılması

4.2.1 İlgilenilen Testin İkili (Binary) Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları

İlgilenilen testin ikili ölçüm olması durumunda önerilen yaklaşımlar (Tutarsızlık Çözücü veya Bileşke Referans standardı) kullanımda olan alternatif test sonuçlarını dikkate alarak düzeltme yapmaktadır. Tutarsızlık Çözücü yaklaşımında bu alternatif test Çözücü test olarak kabul edilirken, Bileşke Referans standardında ise alternatif testlerin kombinasyonu kullanılmaktadır. Bu tezde bu yaklaşımlar ile bu yaklaşımlara alternatif olabilecek Gizli Sınıf Analizi, yaklaşımların yanlılığı düzeltmedeki başarılarını etkileyebilecek durumlar göz önüne alınarak karşılaştırılmıştır. Üretilen verilerde bileşke referans standardını oluşturmak için 3 adet referans test üretilmiştir. Üretilen test verilerinden 5 farklı Gizli Sınıf modeli oluşturulmuştur.

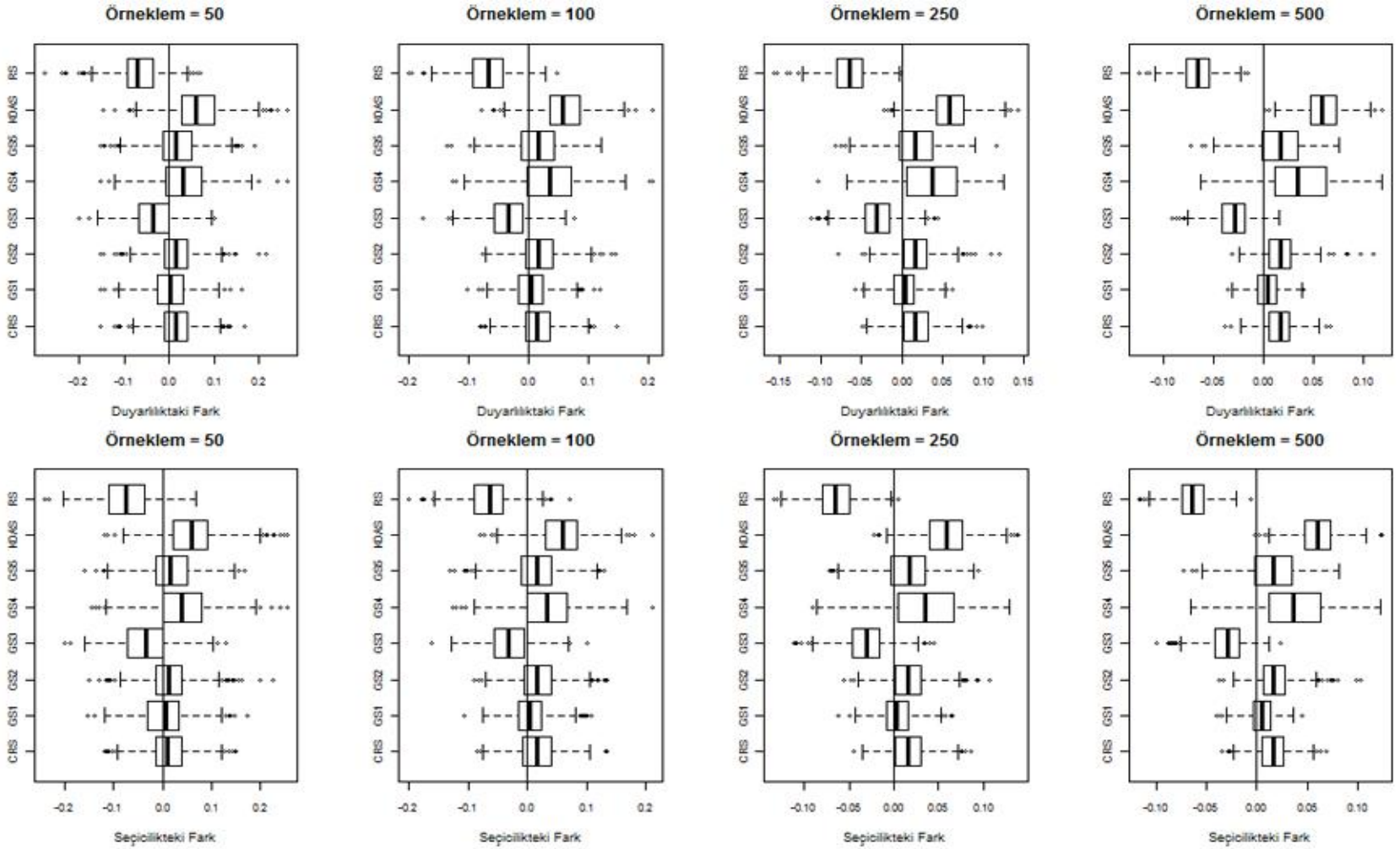
4.2.1.1 Örneklem büyüklüğünün değişmesinin yaklaşımlara olan etkisi

Örneklem büyüklüğünün 50 ile 500 değerleri arasında değişmesinin yaklaşımların Duyarlılık ve Seçicilik tahminlerine etkisi olabileceği düşüncesiyle Prevelansın hem 0.5 hem de 0.8 olduğu durumlarda veriler üretilerek yaklaşımlar karşılaştırılmıştır. Değişen örneklem büyüklüğünün yaklaşımlarca elde edilen doğruluk ölçütlerine etkisi Çizelge 4.18, Şekil 4.17 ve Şekil 4.18'de gösterilmiştir. Buradan da görüldüğü gibi örneklem büyüklüğünün değişmesi yaklaşım tahminlerindeki hata değerini değiştirmemiştir. Burada asıl dikkate değer sonuç yaklaşım tahminlerindeki hata değerlerinin Prevelanstan etkilenmesidir. Prevelans 0.5 olduğunda tüm örneklem büyüklüklerinde yaklaşımların hem duyarlılık hem de seçicilik tahminleri benzer iken, Prevelansın 0.8 olması durumunda yaklaşımların seçicilik tahminindeki hata değerlerinin duyarlılık tahminlerine göre oldukça yüksek olduğu görülmüştür.

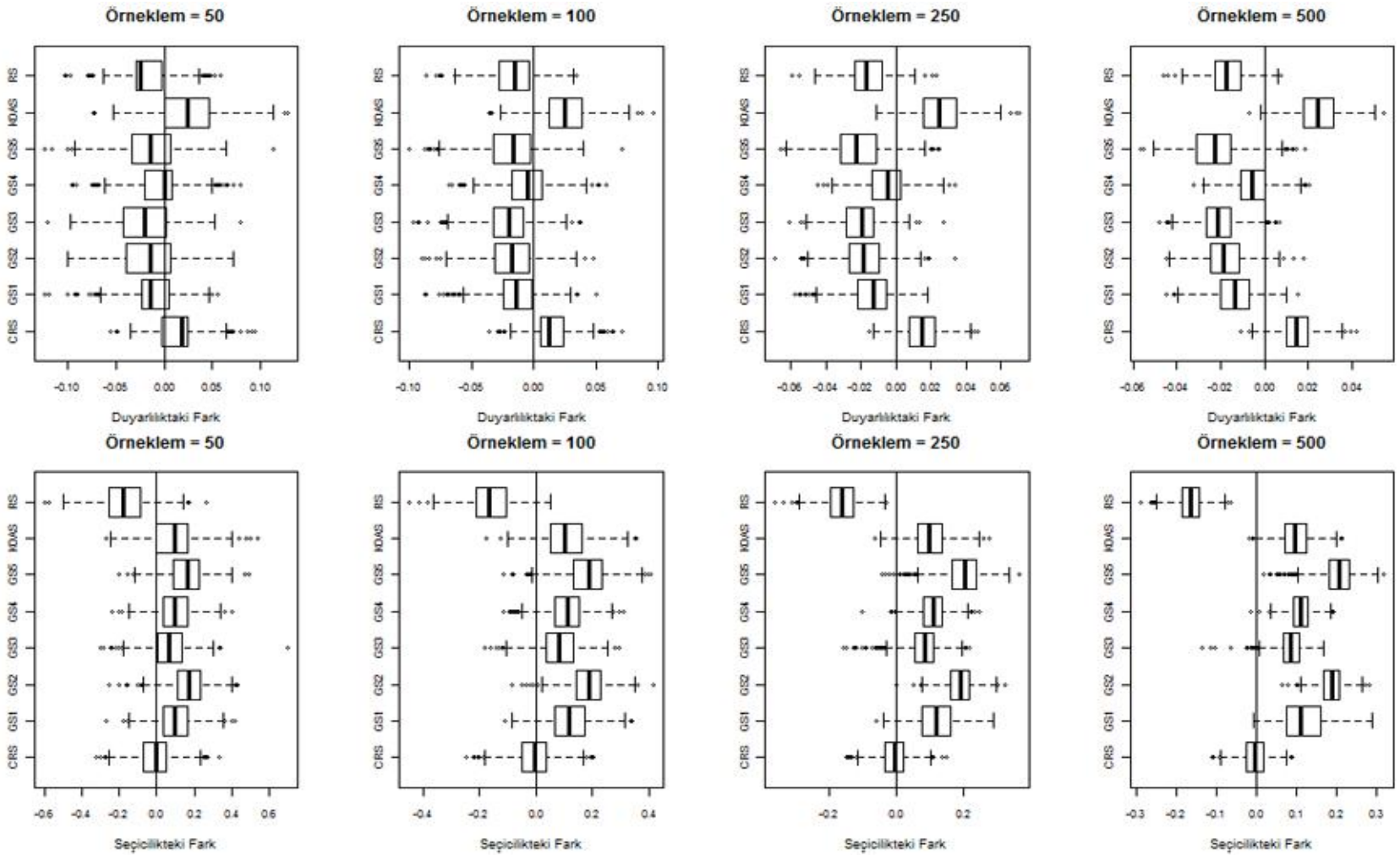
Çizelge 4.18 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$)

	Duyarlılık	N=50	N=100	N=250	N=500
Prevelans = 0.5	KOAS	-61±56	-59±40	-59±26	-60±18
	Çözücü Test	65±50	67±35	65±24	65±16
	Bileşke Ref.	-15±45	-16±33	-16±22	-17±15
	GS1	-5±42	-4±30	-3±18	-4±13
	GS2	-15±46	-17±33	-16±22	-17±16
	GS3	31±48	33±35	31±23	30±17
	GS4	-31±60	-34±51	-35±41	-36±36
	GS5	-16±49	-16±39	-16±30	-16±25
	Seçicilik	N=50	N=100	N=250	N=500
	KOAS	-60±56	-59±41	-58±27	-60±19
	Çözücü Test	68±50	65±36	65±22	64±16
	Bileşke Ref.	-14±46	-16±34	-16±21	-16±15
	GS1	-5±43	-4±30	-4±19	-4±13
	GS2	-16±47	-17±34	-17±22	-17±17
	GS3	33±48	31±35	31±23	30±18
	GS4	-36±59	-34±48	-34±41	-36±35
	GS5	-17±50	-15±40	-16±30	-15±27
	Duyarlılık	N=50	N=100	N=250	N=500
Prevelans = 0.8	KOAS	-23±29	-26±21	-25±13	-25±9
	Çözücü Test	18±25	18±18	17±11	17±8
	Bileşke Ref.	-15±23	-15±16	-15±10	-15±8
	GS1	13±25	14±18	14±12	14±10
	GS2	16±29	18±21	18±13	18±9
	GS3	21±25	21±18	20±12	21±8
	GS4	5±26	5±18	5±12	5±8
	GS5	14±29	18±22	22±15	23±12
	Seçicilik	N=50	N=100	N=250	N=500
	KOAS	-91±117	-103±83	-100±53	-98±38
	Çözücü Test	170±114	164±77	164±50	164±35
	Bileşke Ref.	7±99	7±72	6±43	4±32
	GS1	-106±97	-121±76	-121±59	-119±55
	GS2	-174±96	-185±68	-190±42	-190±30
	GS3	-69±100	-80±69	-83±48	-86±34
	GS4	-102±92	-109±66	-109±42	-110±29
	GS5	-158±105	-183±78	-200±58	-203±46

GS1:KOAS+ÇT+BRS1+BRS2+BRS3,
GS2:KOAS+ÇT+BRS1, (ÇT=Çözücü Test, BRS1=Bileşke Referans için üretilen 1. Test)
GS3:KOAS+ÇT+BRS, (BRS=Bileşke Referans Standardı)
GS4:KOAS+BRS, (KOAS=Kesin Olmayan Altın Standart)
GS5:KOAS+BRS1+BRS2+BRS3 (Bileşke Referans için üretilen 1, 2 ve 3. Testler)



Şekil 4.17 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Prevelans = 0.5



Şekil 4.18 Örneklem büyüklüğünün değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Prevelans = 0.8

4.2.1.2 Prevelans değerinin değişmesinin yaklaşımlara olan etkisi

Prevelanstaki değişimin, yaklaşımların doğruluk ölçütleri tahminleri üzerindeki etkisi incelenirken örneklem büyüklüğündeki değişim de göz önüne alınmıştır. Bu sebeple, hem küçük örneklem büyüklüğünde (n=100), hem de büyük örneklem büyüklüğünde (n=1000) prevelans değeri 0.1 – 0.9 aralığında değiştirilerek karşılaştırılmıştır. (Çizelge 4.19, Çizelge 4.20)

Çizelge 4.19 Küçük veri setinde Prevelans değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$)

Örneklem büyüklüğü = 100					
	Duyarlılık	P=0.2	P=0.4	P=0.6	P=0.8
	KOAS	53±84	28±53	18±33	12±21
	Çözücü Test	234±83	137±51	84±33	42±21
	Bileşke Ref.	109±104	49±60	26±38	9±24
	GS1	-15±89	40±59	55±39	47±24
	GS2	-79±77	47±56	37±35	51±23
	GS3	40±89	99±55	65±36	50±22
	GS4	10±83	-13±55	73±38	37±23
	GS5	-83±93	18±57	61±39	53±25
	Seçicilik	P=0.2	P=0.4	P=0.6	P=0.8
	KOAS	14±21	19±36	28±50	50±88
	Çözücü Test	44±21	86±34	138±51	231±86
	Bileşke Ref.	9±24	22±39	52±58	106±97
	GS1	48±25	54±41	40±61	-18±88
	GS2	51±24	36±37	47±55	-81±76
	GS3	51±23	65±38	100±52	38±87
	GS4	38±23	70±39	-11±51	9±83
	GS5	53±26	58±39	19±56	-81±93

P:Prevelans
GS1:KOAS+ÇT+BRS1+BRS2+BRS3,
GS2:KOAS+ÇT+BRS1, (ÇT=Çözücü Test, BRS1=Bileşke Referans için üretilen 1. Test)
GS3:KOAS+ÇT+BRS, (BRS=Bileşke Referans Standardı)
GS4:KOAS+BRS, (KOAS=Kesin Olmayan Altın Standart)
GS5:KOAS+BRS1+BRS2+BRS3 (Bileşke Referans için üretilen 1, 2 ve 3. Testler)

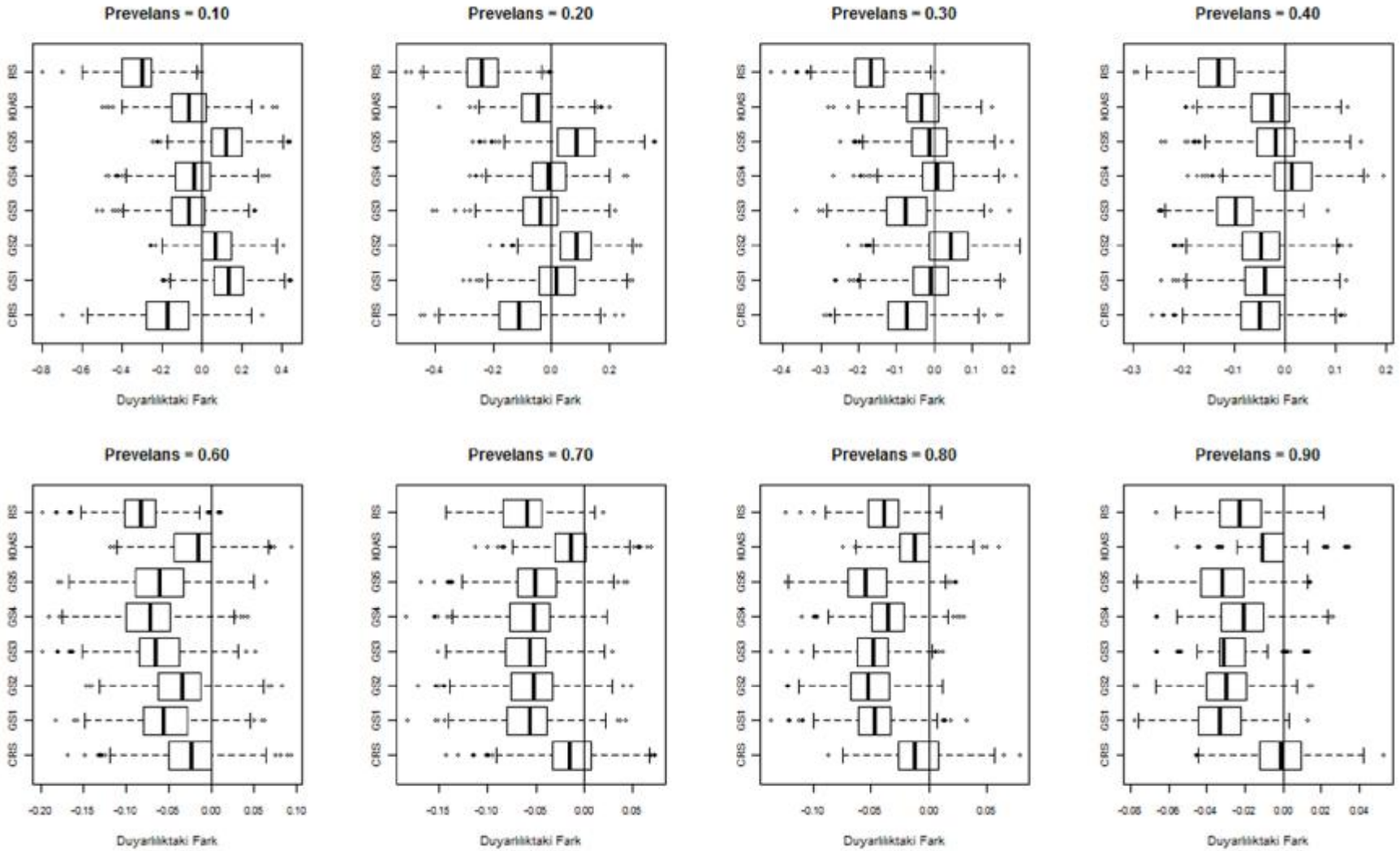
Çizelge 4.20 Büyük veri setinde Prevelans değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı($x10^{-3}$)

Örneklem büyüklüğü = 1000

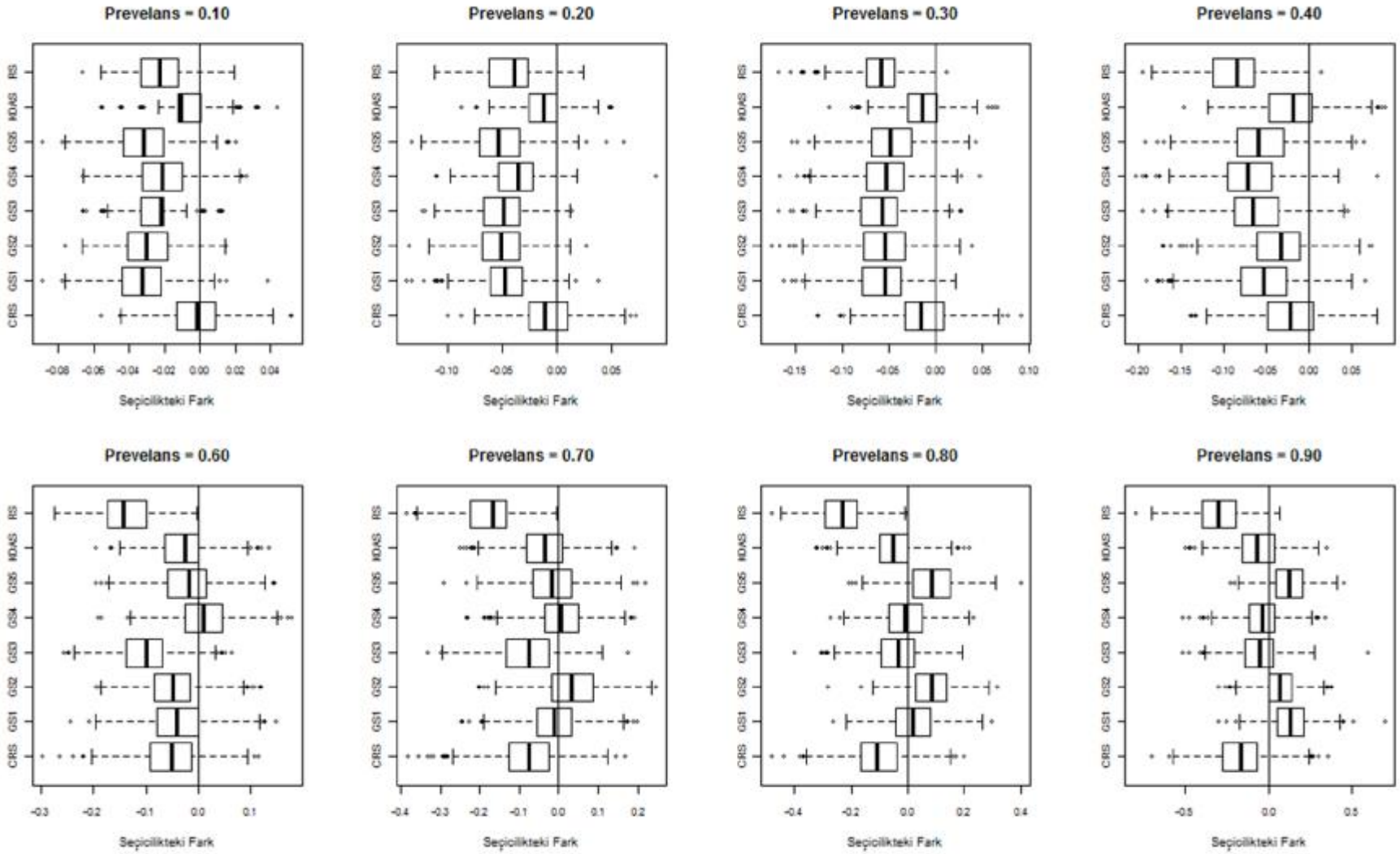
Duyarlılık	P=0.2	P=0.4	P=0.6	P=0.8
KOAS	46±27	30±16	20±11	11±7
Çözücü Test	226±27	139±16	85±10	42±7
Bileşke Ref.	102±31	52±19	25±12	7±8
GS1	-28±46	50±26	53±16	48±9
GS2	-87±24	56±23	35±12	52±7
GS3	29±33	105±18	70±16	51±8
GS4	5±26	-10±21	74±17	37±7
GS5	-115±50	14±18	66±13	58±11
Seçicilik	**P=0.2**	**P=0.4**	**P=0.6**	**P=0.8**
KOAS	12±7	20±11	31±16	45±27
Çözücü Test	43±7	86±11	139±16	223±28
Bileşke Ref.	8±7	25±12	52±18	99±33
GS1	48±9	52±15	49±26	-30±48
GS2	52±8	36±12	57±21	-88±24
GS3	51±8	70±16	105±17	29±36
GS4	37±7	73±17	-11±20	3±28
GS5	59±11	66±13	14±18	-117±50

P:Prevelans
GS1:KOAS+ÇT+BRS1+BRS2+BRS3,
GS2:KOAS+ÇT+BRS1, (ÇT=Çözücü Test, BRS1=Bileşke Referans için üretilen 1. Test)
GS3:KOAS+ÇT+BRS, (BRS=Bileşke Referans Standardı)
GS4:KOAS+BRS, (KOAS=Kesin Olmayan Altın Standart)
GS5:KOAS+BRS1+BRS2+BRS3 (Bileşke Referans için üretilen 1, 2 ve 3. Testler)

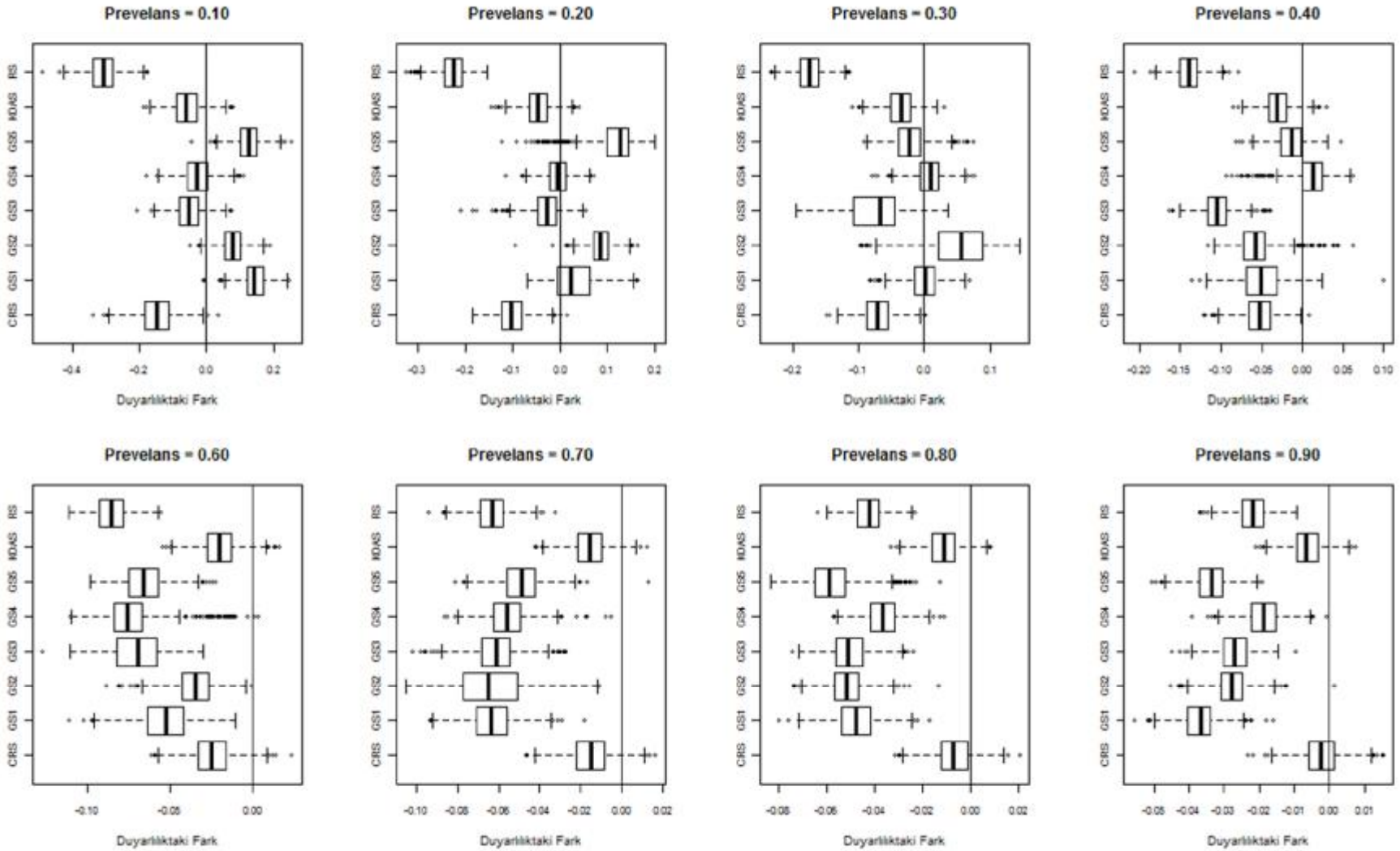
Yaklaşımların prevelans değişiminden etkilendiği gözlenmiştir. Buna göre hem büyük veri setinde hem de küçük veri setinde prevelans değeri arttıkça Gizli Sınıf Analizi yaklaşımında kullanılan modellerin Duyarlılık tahminlerindeki hata değeri artarken, Seçicilik tahminlerindeki hata değeri azalmaktadır. Öte yandan Çözücü test ve Bileşke Referans Standardında ise Gizli Sınıf Analizi yaklaşımında bulunan sonuçların tersi bir ilişki bulunmuştur. Burada ise prevelansın artması Duyarlılık tahminlerindeki hata değerinin azalmasına, Seçicilik tahminlerindeki hata değerinin artmasına neden olmuştur. (Şekil 4.19, Şekil 4.20, Şekil 4.21, Şekil 4.22)



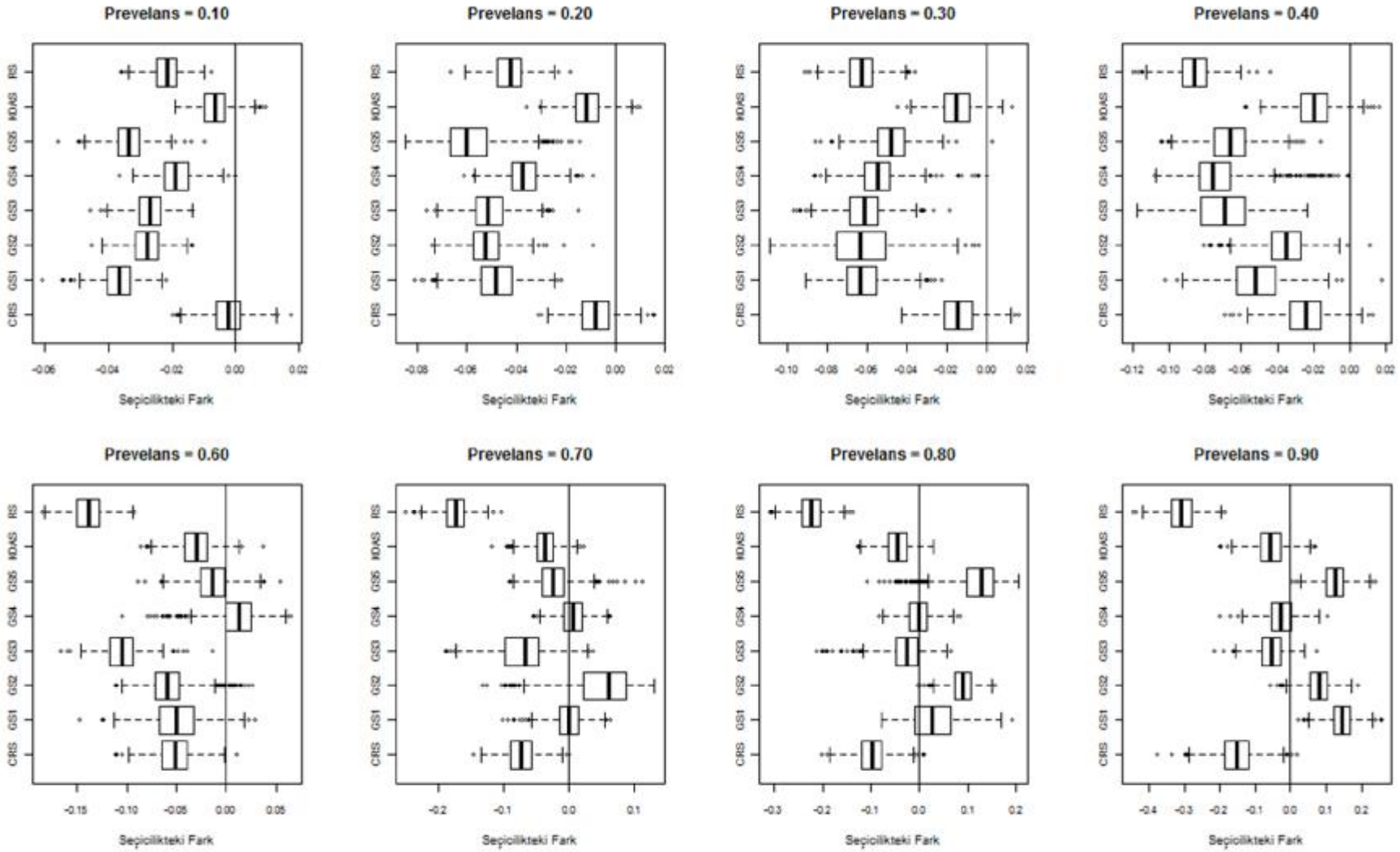
Şekil 4.19 Prevelans değişiminin yaklaşımların Duyarlılık nokta tahminine olan etkisi – n=100



Şekil 4.20 Prevelans değişiminin yaklaşımların Seçicilik nokta tahminine olan etkisi – n=100



Şekil 4.21 Prevelans değişiminin yaklaşımların Duyarlılık nokta tahminine olan etkisi – n=1000



Şekil 4.22 Prevelans değişiminin yaklaşımların Seçicilik nokta tahminine olan etkisi – n=1000

4.2.1.3 Doğruluk ölçütlerindeki değişimin yaklaşımlara olan etkisi

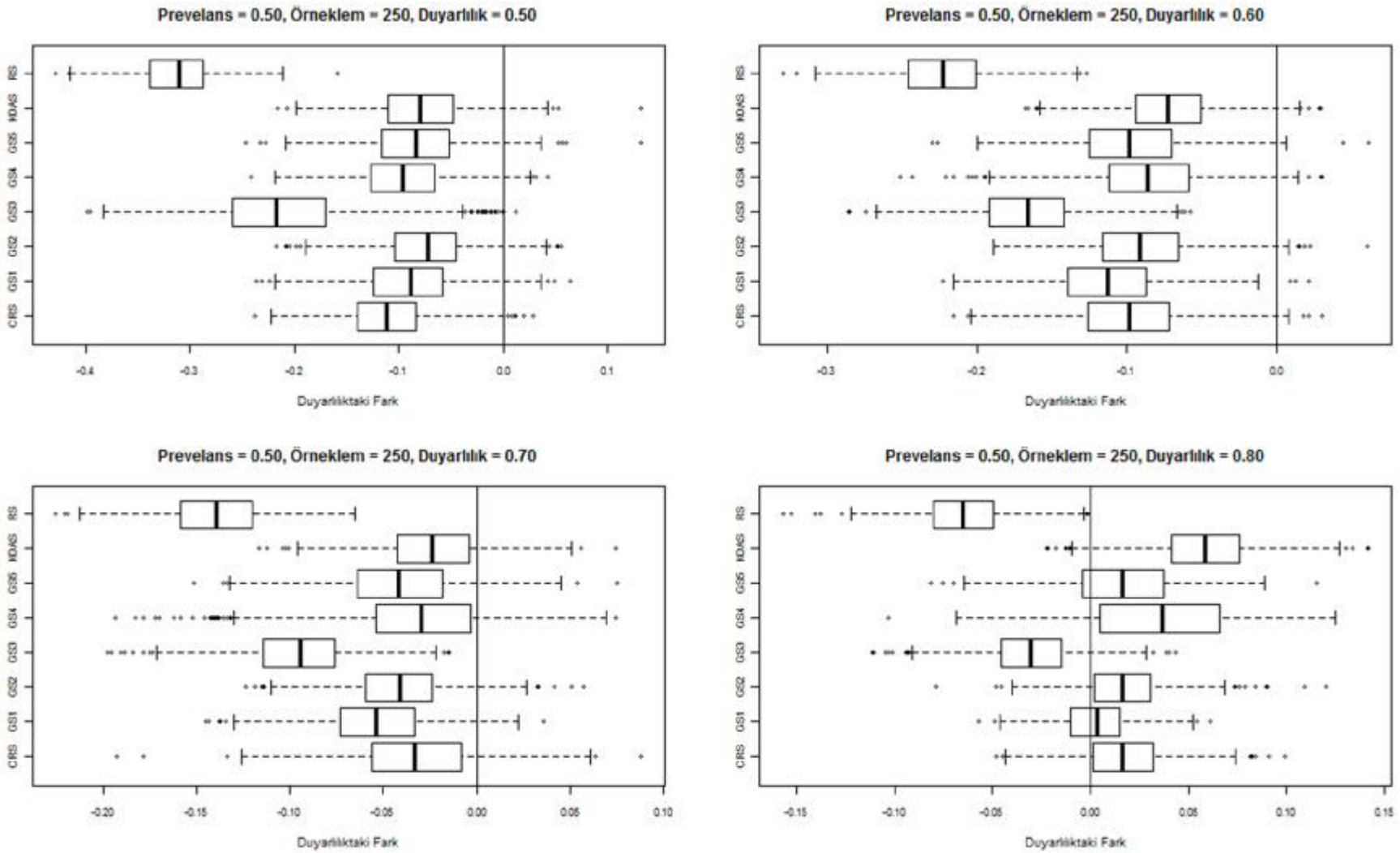
Doğruluk ölçütlerindeki değişimin yaklaşımlara etkisini incelemek amacıyla prevelans ve örneklem büyüklüğü değeri sabit tutulup (prevelans=0.5, örneklem büyüklüğü=250) testin duyarlılık değeri 0.5, 0.6, 0.7 ve 0.8 alınarak yaklaşımlar karşılaştırılmıştır.

Çizelge 4.21 Doğruluk ölçütü değişiminin yaklaşımların doğruluk ölçütleri nokta tahminine olan etkisi: Tahmin değeri – Gerçek değer farkı(x10^{-3})

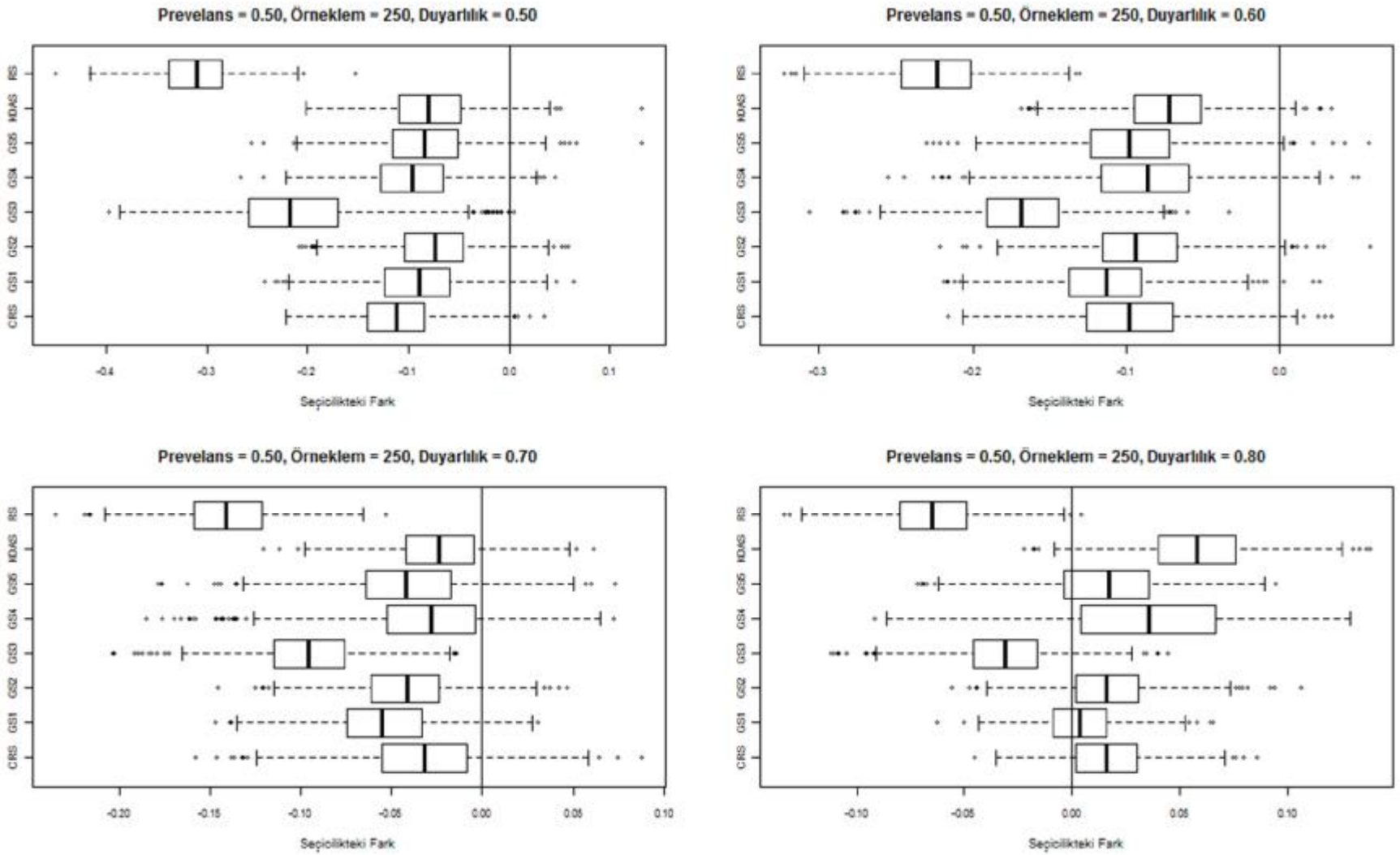
Örneklem büyüklüğü = 250	Duyarlılık	Duy=0.5	Duy=0.6	Duy=0.7	Duy=0.8
	KOAS	79±45	72±34	24±28	-59±26
	Çözücü Test	312±39	223±33	140±27	65±24
	Bileşke Ref.	112±42	99±40	33±35	-16±22
	GS1	91±48	113±39	54±29	-3±18
	GS2	75±44	90±38	42±28	-16±22
	GS3	211±74	167±37	96±29	31±23
	GS4	96±46	87±41	32±41	-35±41
	GS5	84±49	96±40	41±33	-16±30
	Seçicilik	Duy=0.5	Duy=0.6	Duy=0.7	Duy=0.8
	KOAS	79±45	72±34	24±28	-58±27
	Çözücü Test	312±38	224±33	141±27	65±22
	Bileşke Ref.	111±42	98±40	33±35	-16±21
	GS1	90±47	113±37	55±31	-4±19
	GS2	76±44	91±38	42±28	-17±22
	GS3	211±75	168±37	96±30	31±23
	GS4	96±46	88±43	31±40	-34±41
	GS5	84±49	97±39	42±35	-16±30

Duy:Duyarlılık
GS1:KOAS+ÇT+BRS1+BRS2+BRS3,
GS2:KOAS+ÇT+BRS1, (ÇT=Çözücü Test, BRS1=Bileşke Referans için üretilen 1. Test)
GS3:KOAS+ÇT+BRS, (BRS=Bileşke Referans Standardı)
GS4:KOAS+BRS, (KOAS=Kesin Olmayan Altın Standart)
GS5:KOAS+BRS1+BRS2+BRS3 (Bileşke Referans için üretilen 1, 2 ve 3. Testler)

Elde edilen sonuçlara göre Gizli Sınıf Analizi yaklaşımında, duyarlılık değerindeki artış ile doğruluk ölçütleri tahminindeki hata değerinin azaldığı görülmüştür. Benzer sonuç Çözücü test ve Bileşke Referans Standardı için de geçerlidir. (Çizelge 4.21, Şekil 4.23 ve Şekil 4.24)



Şekil 4.23 Doğruluk ölçütlerindeki değişimin yaklaşımların Duyarlılık nokta tahminine olan etkisi



Şekil 4.24 Doğruluk ölçütlerindeki değişimin yaklaşımların Seçicilik nokta tahminine olan etkisi

4.2.1.4 Gerçek Veri Seti Kullanılarak Yapılmış Çalışma Sonuçları

Gastroenteroloji alanında M.Sandıkçı danışmanlığında F.Koçak tarafından yapılan bir çalışmada[53] mide ülserinin suşları arasında sayılan Helicobacter Pylori (HP) bakterisinin varlığını göstermede kullanılan çeşitli testler karşılaştırılmıştır. Çalışmada 80 hasta üzerinde gaitada HP varlığını gösteren HpSA testinin duyarlılık ve seçicilik değerleri belirlenmeye çalışılmıştır. HP varlığı için altın standart olarak kabul edilen test alınan mide biyopsisine uygulanan patolojik incelemedir. Ancak gerek biyopsi alma işleminin girişimsel olması, gerekse patolojik değerlendirme süresinin uzun olabilmesi nedeniyle Kültür (KOAS) test sonuçlarına da başvurulmaktadır. Çalışmada histolojik inceleme, CLO test, PCR gibi diğer tanı koyma testlerine de başvurulmuştur. Bu tezde ise Kültür KOAS olarak, CLO test Çözücü test olarak ve CLO test, histoloji ve PCR sonuçlarından elde edilen kombinasyon da Bileşke Referans Standardı olarak alınmıştır. Ayrıca oluşturulan 5 farklı Gizli sınıf modeli de HpSA'nın doğruluk ölçütlerini belirlemede kullanılmıştır.

Çizelge 4.22 Gerçek veri seti kullanılarak yapılan karşılaştırma sonuçları

Yaklaşımlar	Duyarlılık	Seçicilik
KOAS	0.91	0.67
Çözücü Test	1.00	0.78
Bileşke Ref.	0.92	0.59
GS1	0.83	0.93
GS2	0.83	0.93
GS3	0.83	0.93
GS4	0.62	0.91
GS5	0.70	0.76

GS1:KOAS+ÇT+BRS1+BRS2+BRS3,
GS2:KOAS+ÇT+BRS1, (ÇT=Çözücü Test, BRS1=Bileşke Referans için üretilen 1. Test)
GS3:KOAS+ÇT+BRS, (BRS=Bileşke Referans Standardı)
GS4:KOAS+BRS, (KOAS=Kesin Olmayan Altın Standart)
GS5:KOAS+BRS1+BRS2+BRS3 (Bileşke Referans için üretilen 1, 2 ve 3. Testler)

Yukarıda verilen çizelgeye dikkat edilirse, Gizli Sınıf Analizi yaklaşımı dışındaki yaklaşımlar Duyarlılık değerini yüksek bulurken, Gizli Sınıf Analizi Seçicilik değerini daha yüksek bulmuştur. 80 kişilik veri setinde prevelansın yaklaşık 0.5 olduğu (KOAS pozitif oranı:0.56) durumda Gizli Sınıf Analizi yaklaşımlarının daha iyi sonuç

verdiği üretilmiş veri sonuçlarından hareketle HpSA için Duyarlılık değeri 0.83 ve Seçicilik değeri 0.93 alınabilir. (Çizelge 4.22)

4.2.2 İlgilenilen Testin Sıralı (Ordinal) Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları

Sıralı yapıdaki testler ile yapılmış çalışmalarda ortaya çıkabilecek KOAS yanlılığını düzeltmede Zhou ve ark. tarafından 2003 yılında önerilen ve ROC eğrisi alanını parametrik olmayan bir yolla tahmin eden bir yaklaşım[16] kullanılmaktadır. Bu tezde bu yaklaşım ile bu yaklaşıma alternatif olabilecek Gizli Sınıf Analizi, yaklaşımların yanlılığı düzeltmedeki başarılarını etkileyebilecek durumlar göz önüne alınarak karşılaştırılmıştır.

4.2.2.1 Prevelans değerinin değişmesinin yaklaşımlara olan etkisi

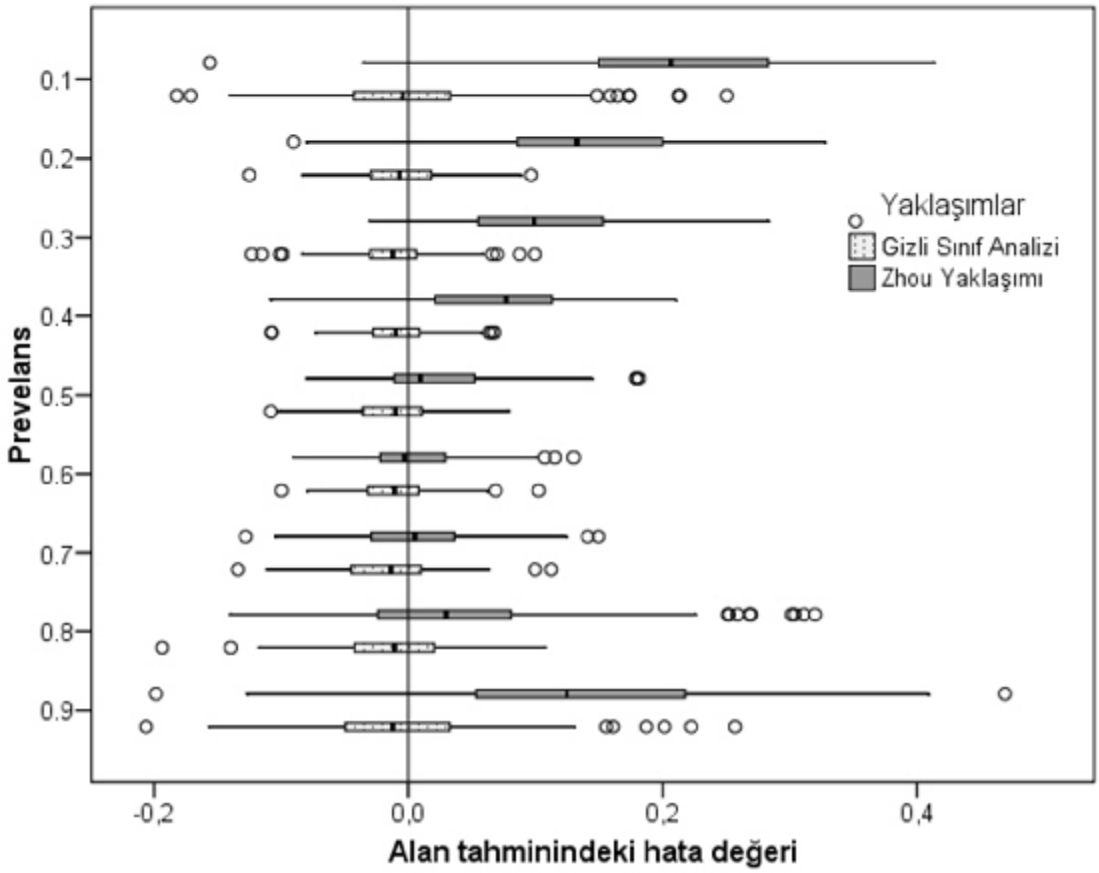
Prevelans değişiminin yaklaşımlara olan etkisi büyük ve küçük örneklemlerde değişkenlik gösterebileceği düşüncesiyle hem büyük hem de küçük örneklemde 0.1'den 0.9'a kadar prevelans değerleri değiştirilerek yaklaşımlar karşılaştırılmıştır. Büyük ve küçük veri setinde yaklaşımların eğri altındaki alan tahminleri ile gerçek alan değeriyle olan farkları Çizelge 4.23, Şekil 4.25 ve Şekil 4.26'da verilmiştir.

Çizelge 4.23 Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı(x10^{-3})

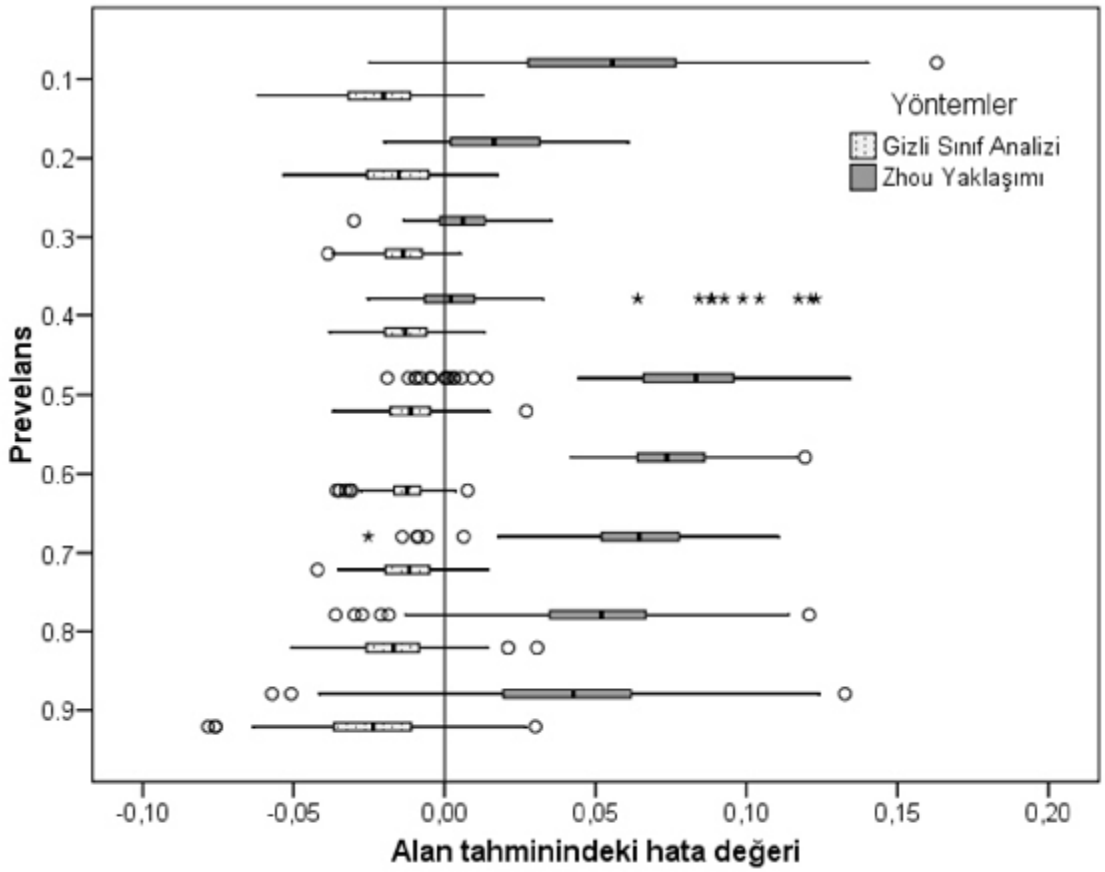
Prevelans	Büyük veri setindeki hata değeri		Küçük veri setindeki hata değeri	
	Zhou Yaklaşımı*	Gizli Sınıf Analizi	Zhou Yaklaşımı*	Gizli Sınıf Analizi
0.1	54±37	-21±15	210±94	1±75
0.2	17±17	-16±15	138±78	-6±39
0.3	7±11	-14±9	105±74	-12±36
0.4	8±27	-12±10	70±68	-11±30
0.5	76±31	-11±11	21±53	-13±34
0.6	75±17	-12±8	4±42	-11±32
0.7	62±23	-12±11	7±50	-15±39
0.8	48±29	-17±14	41±96	-13±48
0.9	42±33	-24±21	134±118	-3±74

* Zhou Yaklaşımı, parametrik olmayan ROC yaklaşımıdır.

Bu çizelgeden de görüldüğü gibi, hem büyük veri setindeki hem de küçük veri setindeki tüm prevelans değerlerinde Gizli Sınıf Analizi yaklaşımı daha küçük hata ile alan değerini tahmin etmiştir. Özellikle prevelans değerinin 0.3 ile 0.7 arasında olması durumunda Gizli Sınıf Analizi yaklaşımının hata değeri daha da azalmıştır.



Şekil 4.25 Büyük veri setinde prevelans değişiminin yaklaşımlara olan etkisi



Şekil 4.26 Küçük veri setinde prevelans değişiminin yaklaşımlara olan etkisi

4.2.2.2 Test ve Kategori sayısının değişmesinin yaklaşımlara olan etkisi

Prevelans değişimine benzer şekilde test ve kategori sayısındaki değişimin yaklaşımlara olan etkisini incelemede hem büyük hem de küçük veri seti kullanılmış ve elde edilen sonuçlar Çizelge 4.24, Şekil 4.27 ve Şekil 4.28'de verilmiştir. Prevelans değişimine benzer olarak örneklemin büyümesi hata oranlarını düşürmüştür. Ayrıca test sayılarındaki ve kategori sayılarındaki değişimin de yaklaşımların sonuçlarını etkilediği görülmüştür. Buna göre testin 3 ve testlerin kategorilerinin 3 olduğu durumda Gizli Sınıf Analizi yaklaşımının hata değerleri Parametrik olmayan ROC yaklaşımından daha düşük bulunmuştur.

Test sayısı sabit tutulup kategori sayısı arttırıldığında ise hem büyük hem de küçük veri setinde Parametrik olmayan ROC yaklaşımının hata oranlarının azaldığı ve Gizli Sınıf Analizi yaklaşımı ile benzer hata dağılımlarına sahip olduğu gözlenmiştir.

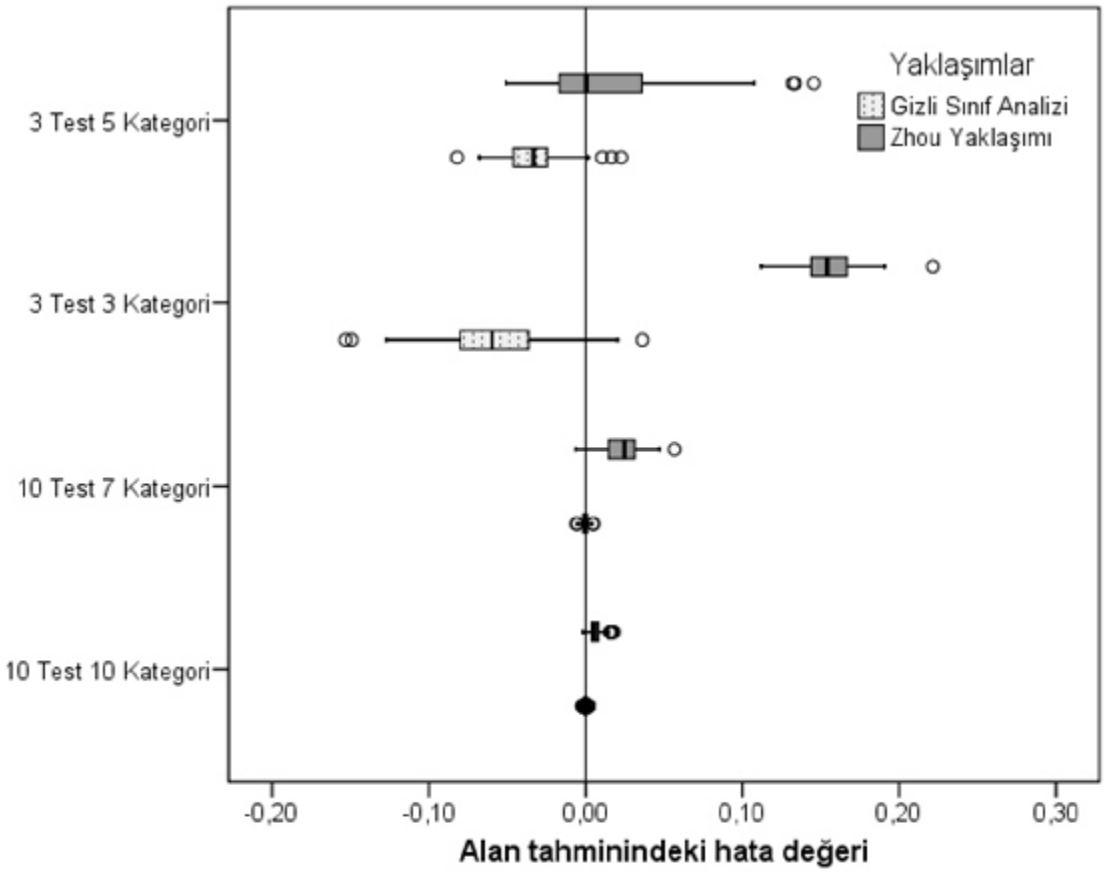
Test ve kategori sayısının artması sonuçlara etki etmiştir. Buna göre 10 test 7 kategori olması durumunda her iki büyüklükteki veri setinde yaklaşımlar daha az hata ve bu hata değerinde de daha düşük varyasyona sahip olmaya başlamıştır. Özellikle büyük veri setinde Gizli Sınıf Analizi yaklaşımı nerdeyse hatasız bir şekilde eğri altındaki alanı tahmin edebilmiştir.

10 test 7 kategori durumuna benzer şekilde 10 test 10 kategori durumunda da yaklaşımlar her iki büyüklükteki veri setlerinde hatasız sonuç vermiştir.

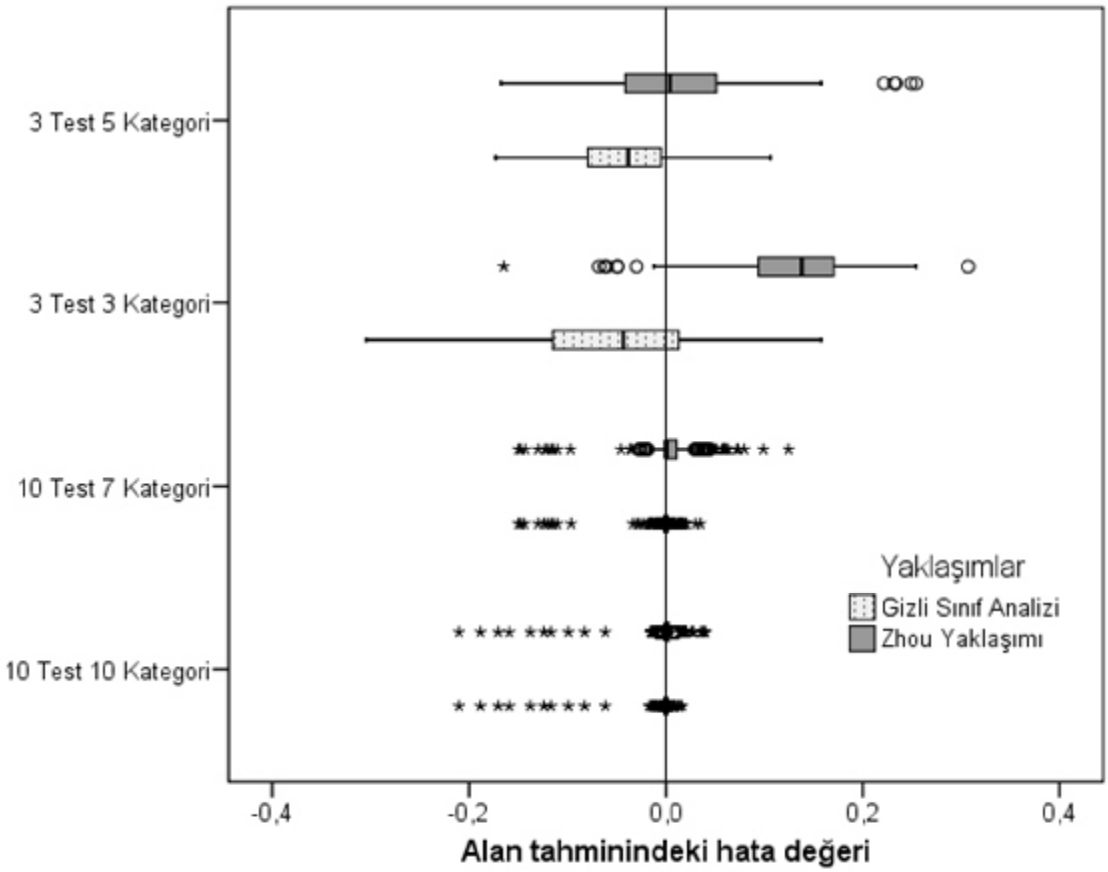
Çizelge 4.24 Test ve kategori sayısındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı(x10^{-3})

Durum	Büyük veri setindeki hata değeri		Küçük veri setindeki hata değeri	
	Zhou Yaklaşımı*	Gizli Sınıf Analizi	Zhou Yaklaşımı*	Gizli Sınıf Analizi
3 Test 3 Kategori	155±18	-58±37	124±79	-54±93
3 Test 5 Kategori	14±46	-33±18	14±86	-43±56
10 Test 7 Kategori	22±13	0±2	3±31	-4±24
10 Test 10 Kategori	6±4	0±1	-2±28	-5±26

* Zhou Yaklaşımı, parametrik olmayan ROC yaklaşımıdır.



Şekil 4.27 Büyük veri setinde Test ve Kategori sayısındaki değişimin yaklaşımlara olan etkisi



Şekil 4.28 Küçük veri setinde Test ve Kategori sayısındaki değişimin yaklaşımlara olan etkisi

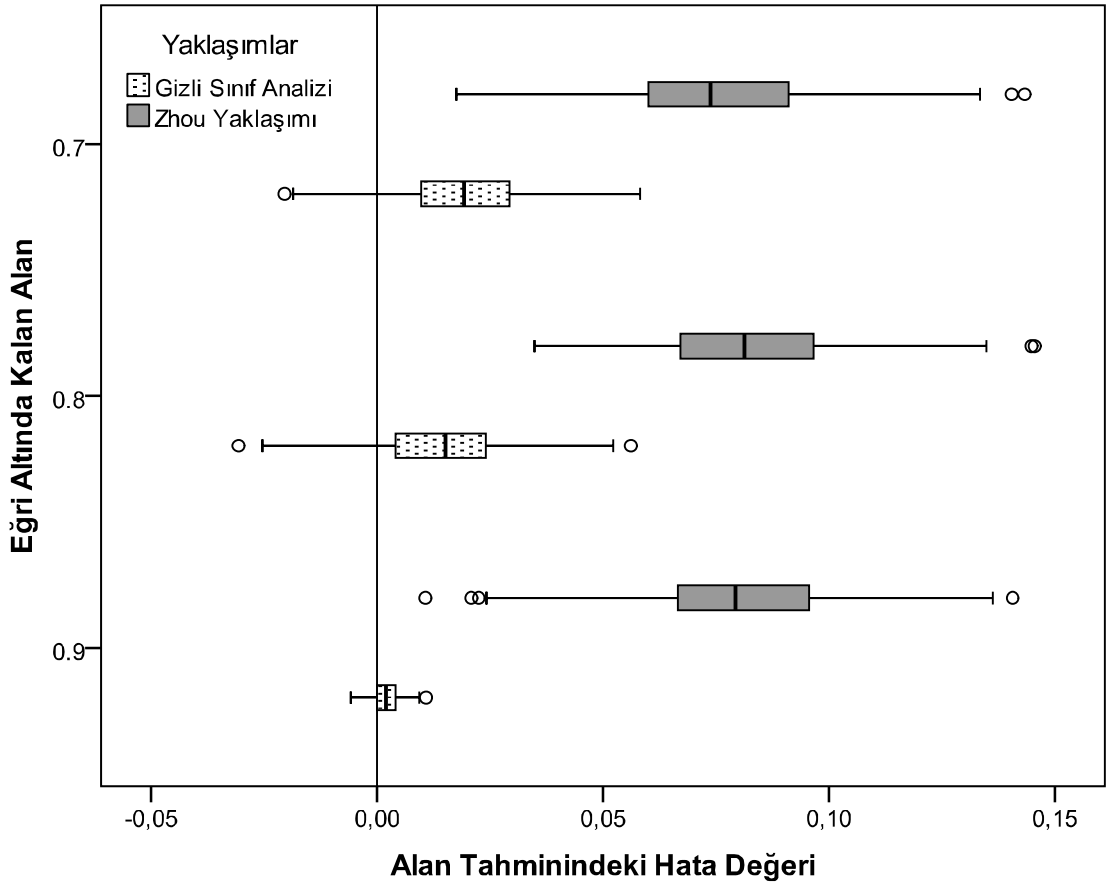
4.2.2.3 Eğri altında kalan alan değerinin değişmesinin yaklaşımlara olan etkisi

Eğri altında kalan alan değerinin değişiminin yaklaşımların sonuçlarını etkileyebileceği düşüncesiyle yaklaşımlar, büyük ve küçük veri setlerinde eğri altındaki alan değerinin 0.7, 0.8 ve 0.9 olduğu durumları tahmin etmede karşılaştırılmıştır. Her iki büyüklükteki veri setinde eğri altında kalan alan değerinin 0.7 olduğu durumda Gizli Sınıf Analizi yaklaşımının hatalı sonuç verdiği, ancak alan değerinin 0.9 değerine yükselmesi durumunda ise bu hatanın ortadan kalktığı görülmüştür. Parametrik olmayan ROC yaklaşımının ise her iki alan değerinde de belirli bir hatayla da olsa tahminde bulunduğu görülmüştür.(Çizelge 4.25, Şekil 4.29 ve Şekil 4.30)

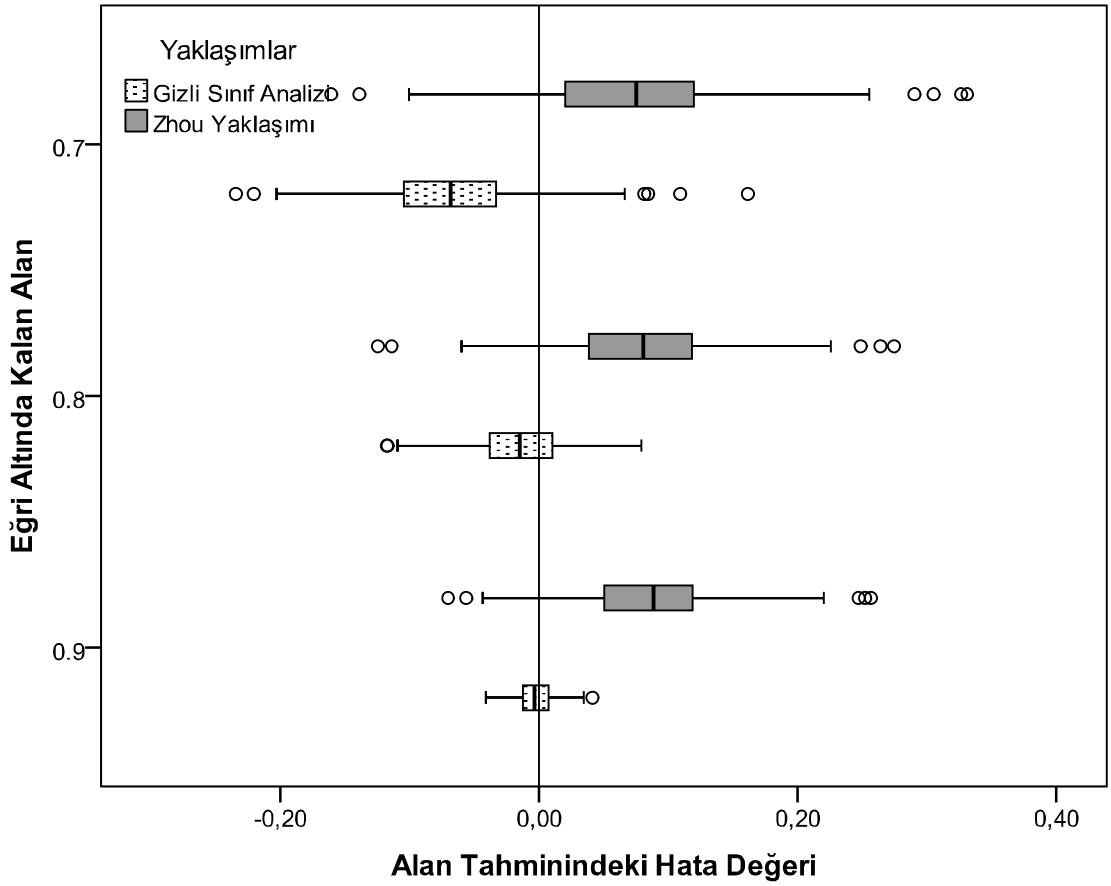
Çizelge 4.25 Eğri altında kalan alan değerinin değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı(x10^{-3})

Eğri altında kalan alan değeri	Büyük veri setindeki hata değeri		Küçük veri setindeki hata değeri	
	Zhou Yaklaşımı*	Gizli Sınıf Analizi	Zhou Yaklaşımı*	Gizli Sınıf Analizi
0.7	76±22	19±15	72±73	-69±54
0.8	82±20	15±14	80±53	-15±35
0.9	80±22	2±3	86±54	-3±15

* Zhou Yaklaşımı, parametrik olmayan ROC yaklaşımıdır.



Şekil 4.29 Büyük veri setinde Eğri altında kalan alan değerinin değişiminin yaklaşımlara olan etkisi



Şekil 4.30 Küçük veri setinde Eğri altında kalan alan değerinin değişiminin yaklaşımlara olan etkisi

4.2.2.4 Gerçek Veri Seti Kullanılarak Yapılmış Çalışma Sonuçları

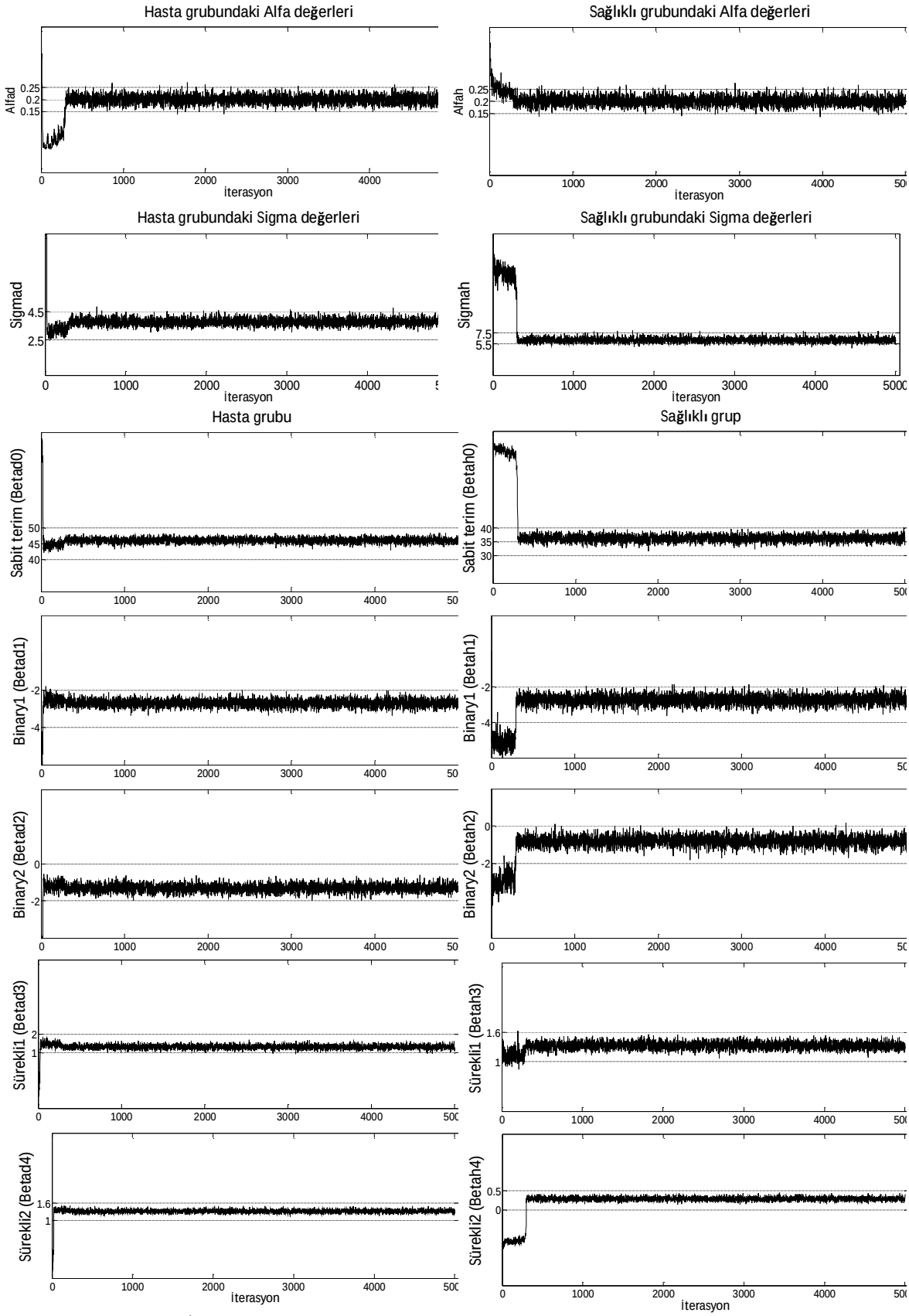
Günaştı ve ark.[54] tarafından yapılan çalışmada 5 öğretim üyesi eğitim öncesi ve sonrası dönemlerinde 0-6 arasında sıralı yapıda değer alan bir değerlendirme ölçeği ile 54 hastanın fotoğraflarına bakarak sınıflama yapmıştır. Yapılan sınıflama sonucunda hastalar az girintiliden çok girintiliye olacak şekilde sınıflanmıştır. Çalışmanın verileri kullanılarak yapılan analizler sonucunda yaklaşımların belirlemiş oldukları eğri altında kalan alan değerleri Çizelge 4.26'da verildiği gibi bulunmuştur. Buna göre yaklaşımlar birbirlerine oldukça yakın sonuçlar elde etmiştir. Üretilmiş veri çalışmalarından elde edilen sonuçlara göre, 5 testin (değerlendiricinin) ve 7 kategorinin olduğu durum için bu verinin değerlendirilmesinde Gizli Sınıf Analizi yaklaşımının sonuçlarının dikkate alınması daha uygun olacaktır.

Çizelge 4.26 Gerçek veri seti kullanılarak elde edilen yaklaşımların ROC eğri alanları ve standart sapma değerleri

Tüm testler üzerinden	Yaklaşımların alan tahminleri	
	Parametrik olmayan ROC	Gizli Sınıf Analizi
Eğitim öncesi öğretim üyeleri		
Öğr. 1	0.969 ± 0.025	0.986 ± 0.016
Öğr. 2	0.878 ± 0.045	0.838 ± 0.055
Öğr. 3	0.941 ± 0.035	0.970 ± 0.024
Öğr. 4	0.677 ± 0.081	0.708 ± 0.070
Öğr. 5	0.928 ± 0.036	0.913 ± 0.040
Eğitim sonrası öğretim üyeleri		
Öğr. 1	0.934 ± 0.040	0.945 ± 0.032
Öğr. 2	0.543 ± 0.125	0.926 ± 0.037
Öğr. 3	0.931 ± 0.026	0.937 ± 0.034
Öğr. 4	0.694 ± 0.080	0.971 ± 0.023
Öğr. 5	0.902 ± 0.049	0.966 ± 0.025

4.2.3 İlgilenilen Testin Sürekli Ölçüm Olduğu Durumda Önerilen Yaklaşımların Karşılaştırmaları

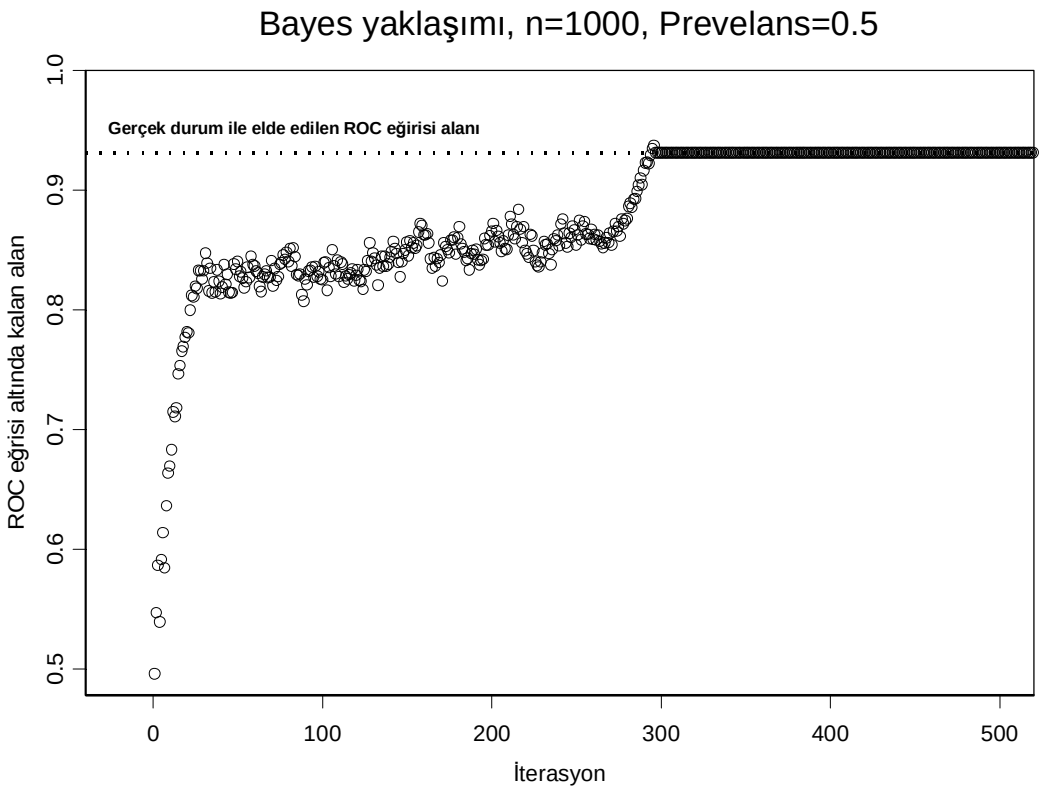
İlgilenilen testin sürekli ölçüm olması durumunda KOAS yanlılığını düzeltmede, ek değişkenlerin varlığında, Bayes yaklaşımı kullanılmaktadır. Bayes yaklaşımı iteratif bir yaklaşım olduğu için tekrar sayısı arttıkça daha çok yakınsayacaktır. Ancak yakınsama problemi olan durumlarda tekrar sayısı ne kadar fazla da olsa iterasyon sonucu sabit bir değer olmayacaktır. Büyük veri seti kullanılarak yapılan analizler sonucunda KOAS yanlılığını düzeltmede kullanılan Bayes yaklaşımının çeşitli parametrelerinde iterasyonlar sonucunda elde edilen sonuçlar Şekil 4.31'de verilmiştir.



Şekil 4.31 İterasyonlara göre Bayes yaklaşımı parametrelerinin tahmin edicileri

Görüldüğü üzere iterasyon yaklaşık 300 gibi bir değere geldikten sonra, ilgili değişkenin belirli aralıkta kalacak şekilde tahmini yapılabilmektedir. Bu durum, bu büyüklükte ve özellikteki bir veri seti için 5000 iterasyonun gereksiz olduğunu göstermektedir. Bayes yaklaşımında tahmini elde edilen diğer bir parametre ise gerçek durum için ek değişkenlerle yapılan lojistik regresyondaki beta katsayılarıdır. Beta(lar)nın tahmin edicileri de belirli bir iterasyon sayısından sonra bir aralıkta elde edilmiştir.

Bayes yaklaşımında prevelansın 0.5 alındığı büyük veri seti kullanılarak yapılan iterasyon sonucunda elde edilen gerçek durum tespiti kullanılarak hesaplanan ROC eğrisi altında kalan alan iterasyonlara göre Şekil 4.32'de verildiği gibi bulunmuştur.



Şekil 4.32 Büyük veri seti ve prevelans 0.5 olduğu durumda Bayes yaklaşımının ROC eğrisi altında kalan alan tahmini (üretilmiş 1 veri seti için)

Bu durum için tezde önerilen diğer yaklaşım olan Gizli Sınıf Analizi farklı modelleri ile incelenmiştir. Sadece bir veri seti için LatentGold[57] paket programından elde edilen genel sonuç tablosu Çizelge 4.27'de verilmiştir. Bu çizelgede modellerin uyumunu incelemede kullanılan logaritmik olabilirlilik (LogLikelihood-LL) değeri, yine bu değerden elde edilmiş Bayes bilgi kriteri (Bayesian Information Criteria, BIC)

değeri, tahmin edilebilen parametre sayısı (Npar) ve sınıflama hata oranı (Class. Err.) değeri verilmiştir. Bilindiği üzere modellerin uyumunda düşük LL (veya BIC) değeri uyumun yüksek olduğunu göstermektedir. Bu tabloda en küçük LL (veya BIC) değerine sahip olan model, Model I'dir. Ancak bu modelin sınıflama hata oranı oldukça yüksek olduğundan bu modeli seçmek doğru olmayacaktır. Burada uygun model, Model III olacaktır. Bu modelin sınıflama hata oranı son modele benzer değer almıştır, fakat bu modelin LL değeri son modele göre oldukça düşüktür.

Çizelge 4.27 Büyük veri seti ve prevelans 0.5 olduğu durumda Gizli sınıf modelleri uyum istatistikleri (üretilmiş 1 veri seti için)

	LL	BIC(LL)	Npar	Class. Err.
Model I	-1316.9658	2668.4703	5	0.1969
Model II	-3579.8469	7221.8636	9	0.1273
Model III	-4403.7894	8869.7486	9	0.0026
Model IV	-6131.9212	12353.6433	13	0.0015

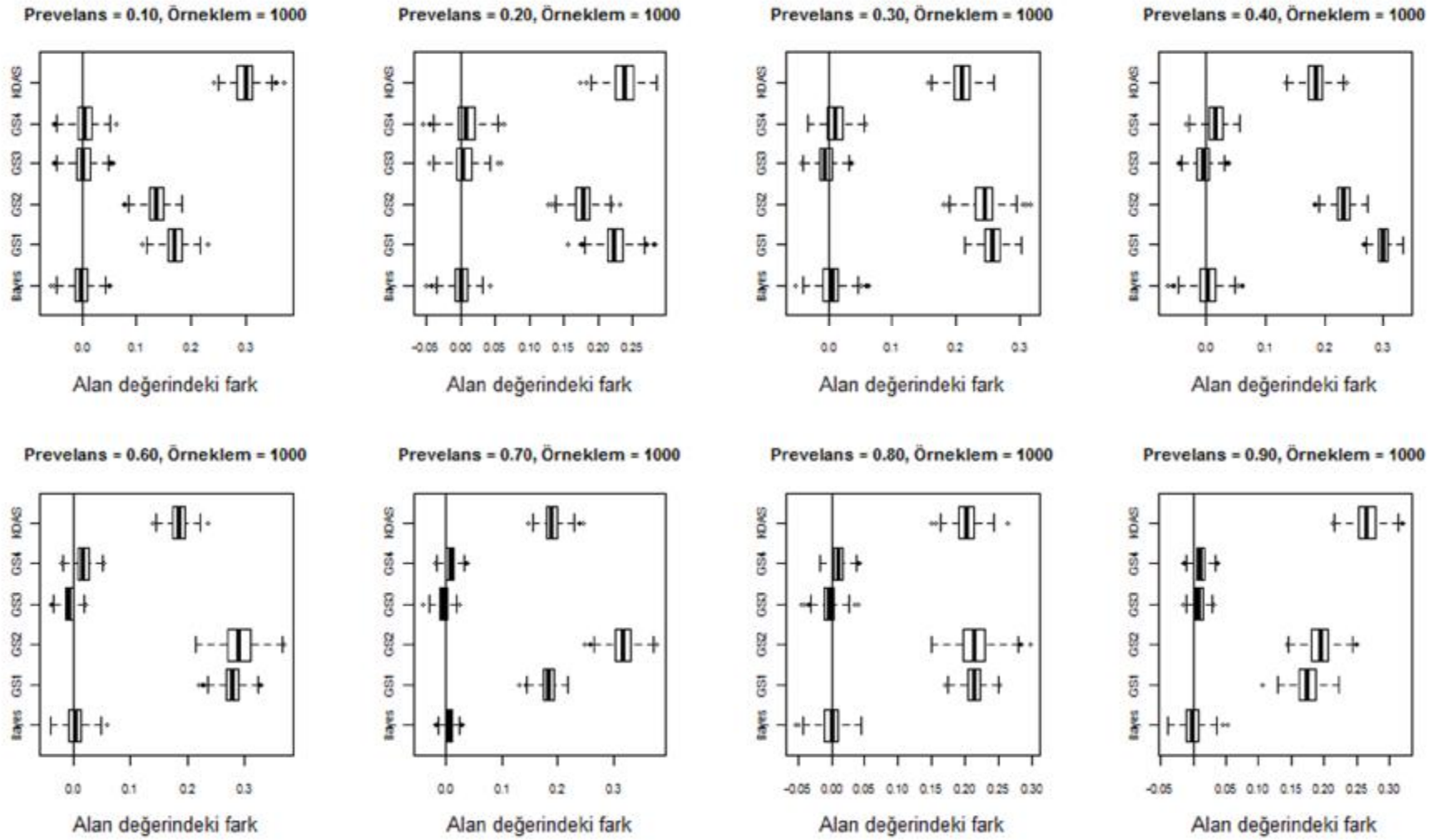
4.2.3.1 Prevelans değerinin değişmesinin yaklaşımlara olan etkisi

Yaklaşımların Prevelans değişiminden etkilenebileceği düşüncesiyle, hem büyük hem de küçük örneklemde prevelans değerleri 0.1 ile 0.9 arasında alınmıştır. KOAS bilindiği üzere belirli bir hata oranı ile doğruluk ölçütlerini belirler. Burada bu hata oranı, özellikle düşük prevelans değerinde, oldukça yüksektir. Bayes yaklaşımı prevelans değişiminden etkilenmemektedir. Öte yandan Gizli Sınıf Analizi yaklaşımında iyi modeller olan GS3 ve GS4'ün prevelans değişiminden az da olsa etkilendiği gözlenmiştir. Prevelans değişimi küçük veri setinde incelendiğinde ise benzer sonuçlar bulunmuş, ilave olarak yaklaşımların hatalarındaki varyasyon değeri artmıştır. (Çizelge 4.28, Şekil 4.33, Şekil 4.34)

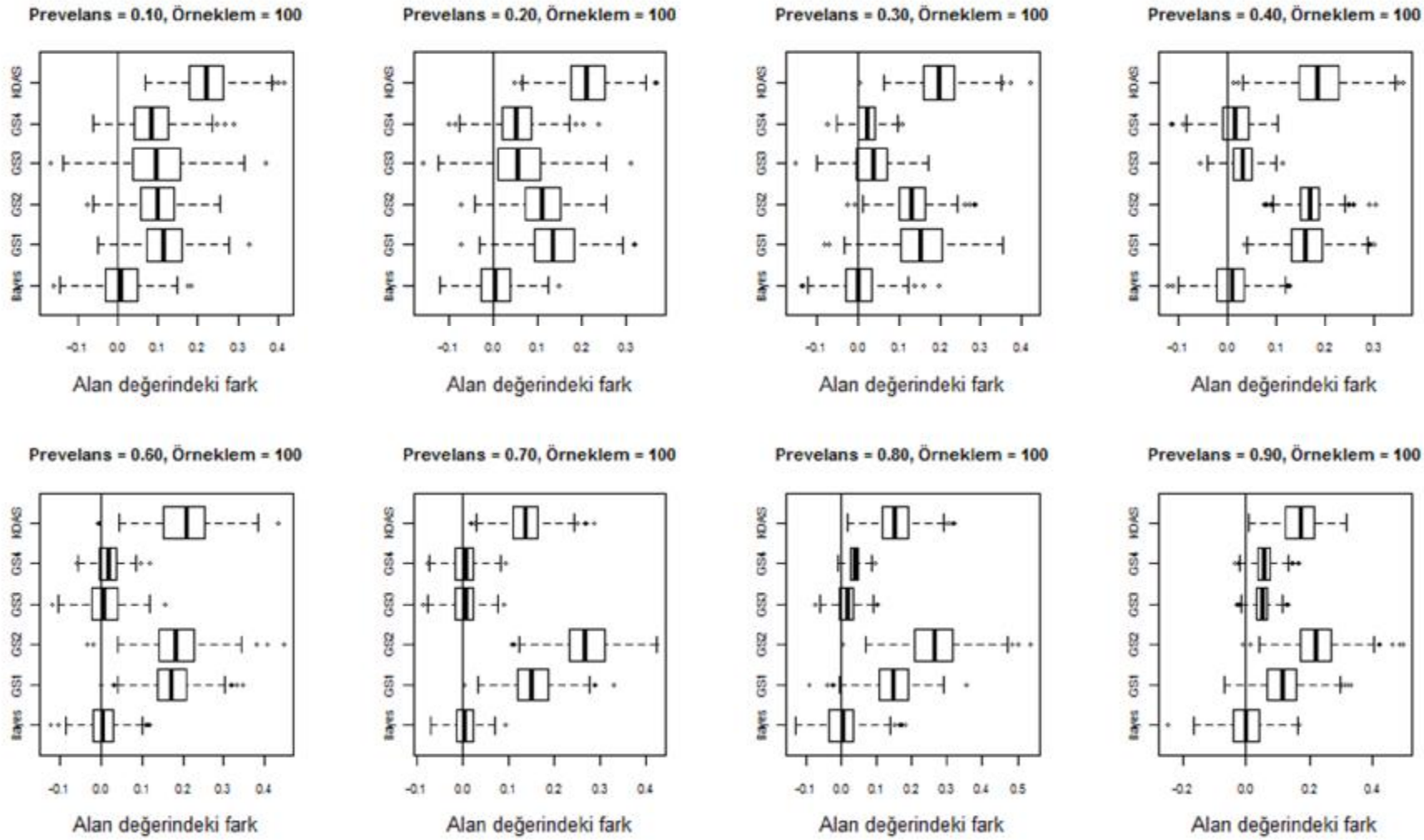
Çizelge 4.28 Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi
Tahmin değeri – Gerçek değer farkı(x10^{-3})

	Prevelans	KOAS	Bayes Yaklaşımı	GS1	GS2	GS3	GS4
Örneklem büyüklüğü = 1000	0.1	299±20	-2±18	168±19	136±19	2±19	4±19
	0.2	238±19	-1±14	223±19	178±15	3±17	6±19
	0.3	208±19	4±17	256±17	243±21	-5±13	11±16
	0.4	184±18	2±19	299±12	231±15	-6±14	15±17
	0.5	176±15	1±15	327±24	252±22	-8±8	15±14
	0.6	185±14	3±17	278±18	290±28	-7±11	17±12
	0.7	190±15	4±8	183±14	317±21	-5±10	6±10
	0.8	202±17	0±16	214±14	215±24	-3±12	10±11
	0.9	266±20	-2±14	174±17	194±18	8±8	10±9
	Prevelans	KOAS	Bayes Yaklaşımı	GS1	GS2	GS3	GS4
Örneklem büyüklüğü = 100	0.1	233±62	5±59	115±65	99±58	96±86	82±59
	0.2	209±56	4±44	136±62	108±57	55±72	51±51
	0.3	198±57	2±48	154±73	131±48	35±52	22±29
	0.4	186±59	7±42	161±48	168±33	32±28	17±37
	0.5	162±48	5±26	175±49	173±52	14±29	8±18
	0.6	204±74	6±38	174±55	184±65	9±46	17±28
	0.7	137±44	5±28	155±48	269±57	5±28	6±29
	0.8	153±52	4±53	148±62	263±81	17±31	37±18
	0.9	171±63	-1±61	115±69	221±76	52±26	58±31

GS1: ikili değişkenler (d1 + d2),
GS2: ikili değişkenler + bir sürekli değişken (d1 + d2 + c1),
GS3: ikili değişkenler + bir sürekli değişken (d1 + d2 + c2),
GS4: tüm değişkenler (d1 + d2 + c1 + c2)



Şekil 4.33 Büyük veri setinde Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi



Şekil 4.34 Küçük veri setinde Prevelans değişiminin yaklaşımların EAKA nokta tahminine olan etkisi

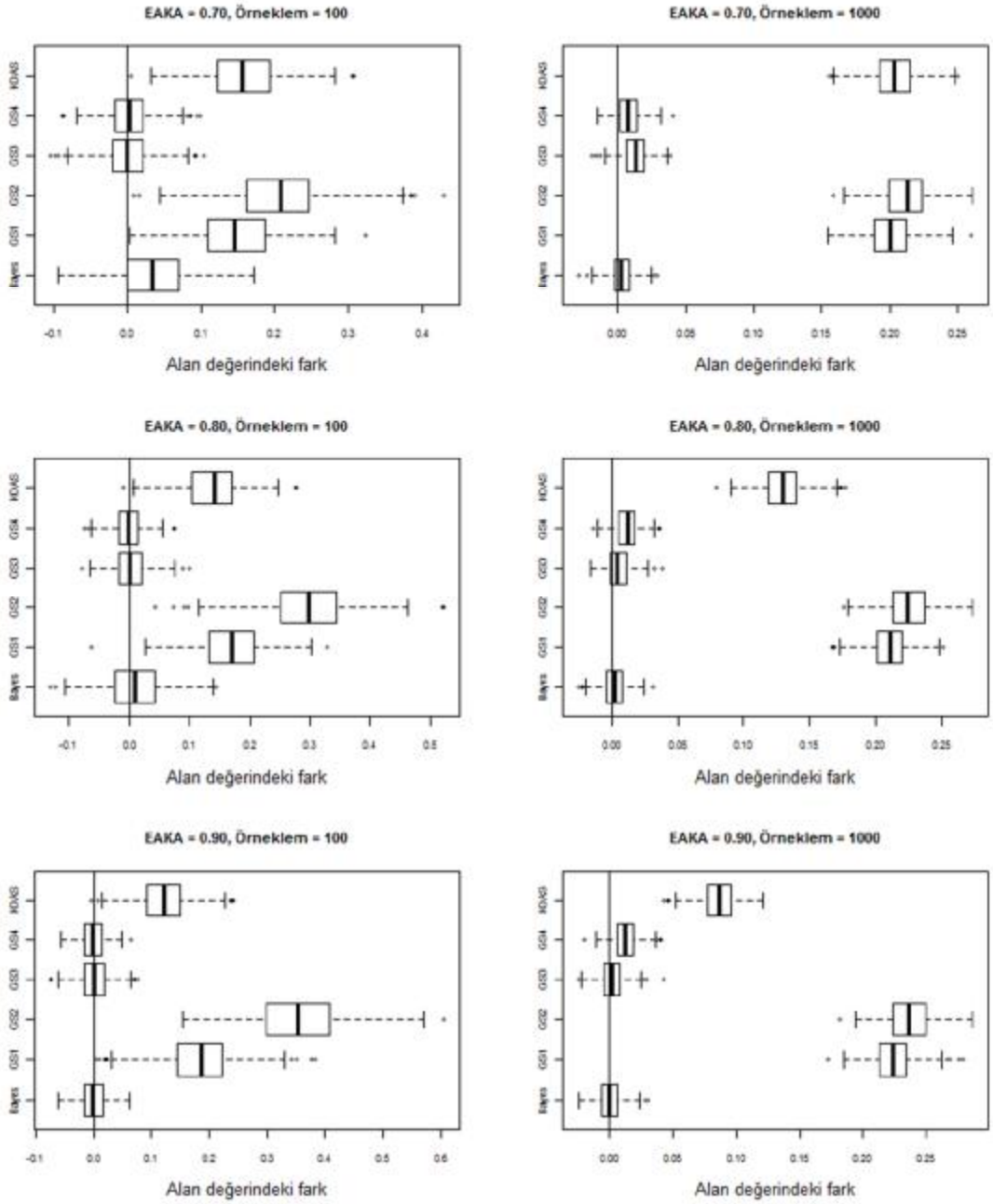
4.2.3.2 Eğri altında kalan alan değerinin değişiminin yaklaşımlara olan etkisi

Yaklaşımların eğri altında kalan alan değerindeki değişimden etkilenebileceği düşüncesiyle, hem büyük hem de küçük örneklemde EAKA değerleri 0.7, 0.8 ve 0.9 olarak alınmıştır. Prevelans değişiminde elde edilen sonuçlara benzer olarak, büyük veri setinde Bayes yaklaşımı ve GS3 ile GS4 modelleri alan değerindeki değişimden etkilenmemiştir. Öte yandan küçük veri setinde özellikle düşük EAKA değerinde Bayes yaklaşımı diğer durumlara göre oldukça hatalı sonuçlar vermiştir ki bu da Bayes yaklaşımının küçük veri setinde EAKA değerindeki değişimden etkilendiğini göstermektedir. Ancak Gizli Sınıf Analizi yaklaşımı bu değişimden etkilenmemektedir. (Çizelge 4.29, Şekil 4.35)

Çizelge 4.29 EAKA değişiminin yaklaşımların EAKA nokta tahminine olan etkisi
Tahmin değeri – Gerçek değer farkı($x10^{-3}$)

		EAKA=0.7	EAKA=0.8	EAKA=0.9
Örneklem büyüklüğü = 1000	KOAS	203±17	130±15	86±14
	Bayes Yaklaşımı	3±9	2±9	0±9
	GS1	200±18	210±15	224±16
	GS2	212±17	224±18	237±18
	GS3	13±8	5±8	2±9
	GS4	8±9	12±9	12±9
		EAKA=0.7	EAKA=0.8	EAKA=0.9
Örneklem büyüklüğü = 100	KOAS	157±51	137±48	122±43
	Bayes Yaklaşımı	35±52	10±48	-1±23
	GS1	148±54	170±57	185±61
	GS2	205±64	295±73	354±82
	GS3	0±33	1±29	1±25
	GS4	2±31	-2±24	0±20

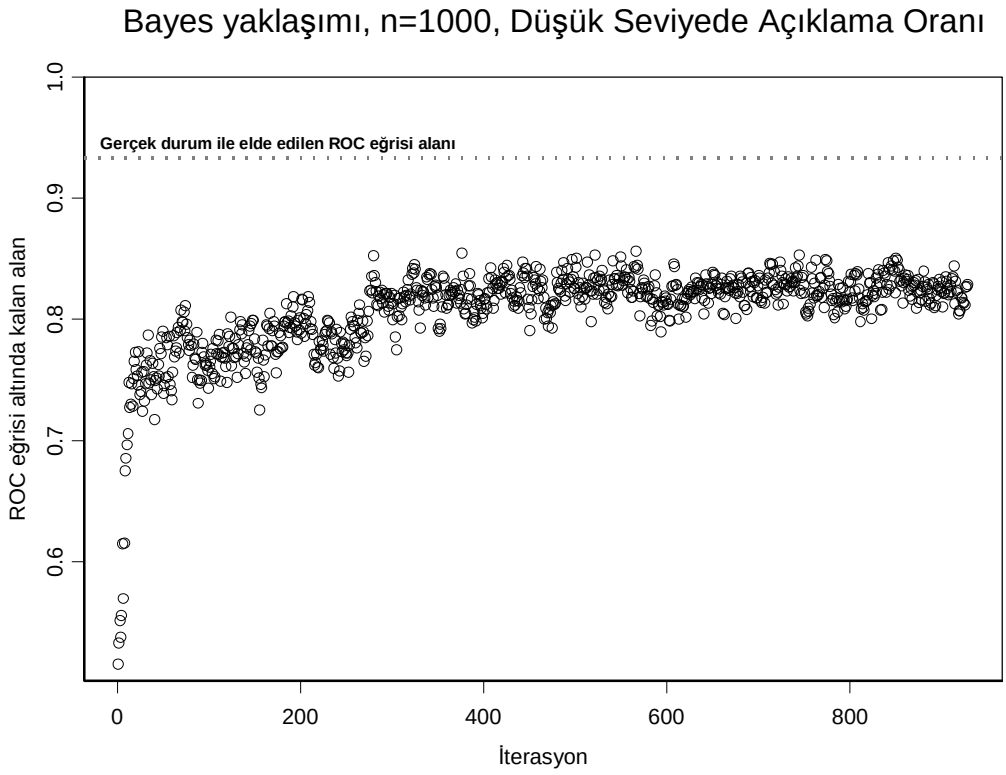
GS1: ikili değişkenler (d1 + d2),
GS2: ikili değişkenler + bir sürekli değişken (d1 + d2 + c1),
GS3: ikili değişkenler + bir sürekli değişken (d1 + d2 + c2),
GS4: tüm değişkenler (d1 + d2 + c1 + c2)

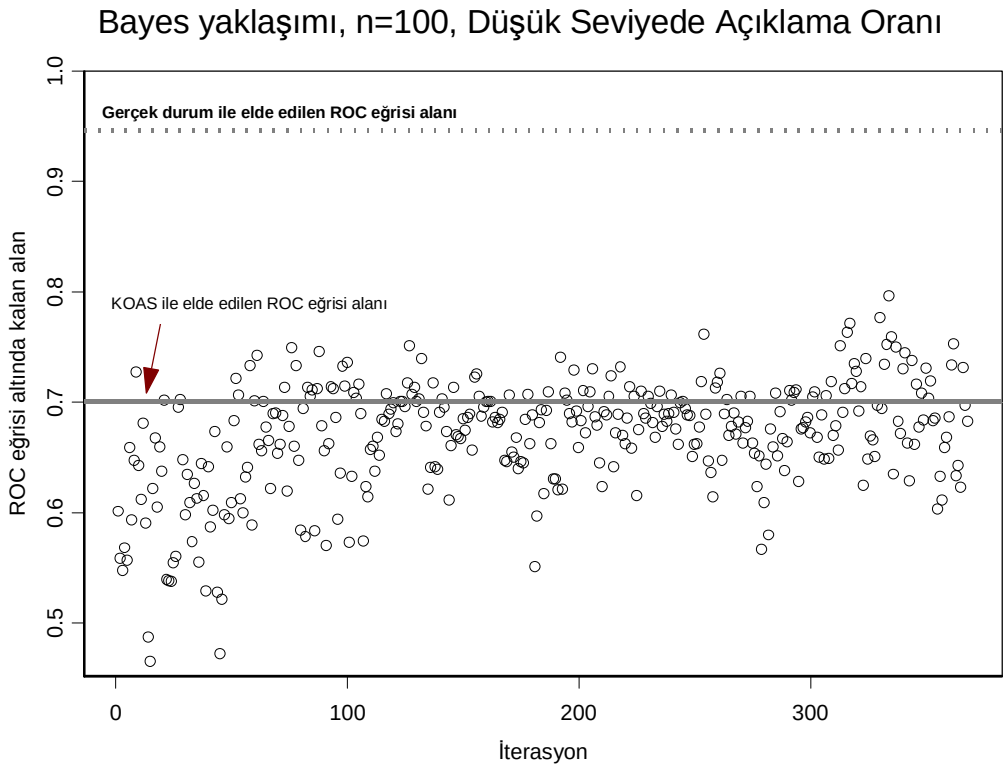


Şekil 4.35 EAKA değişiminin yaklaşımların EAKA nokta tahminine olan etkisi

4.2.3.3 Ek değişkenlerin gerçek durumu açıklama oranındaki değişimin yaklaşımlara olan etkisi

Ek değişkenlerin gerçek durumu açıklama oranındaki değişim hem büyük hem de küçük veri setinde Bayes yaklaşımının tahminlerini etkilemiştir. Her iki büyüklükteki veri setinde de Bayes yaklaşımı iterasyon sürecinde gerçek alan değerinin oldukça uzağında kalan bir değere yakınsamıştır. (Şekil 4.36)





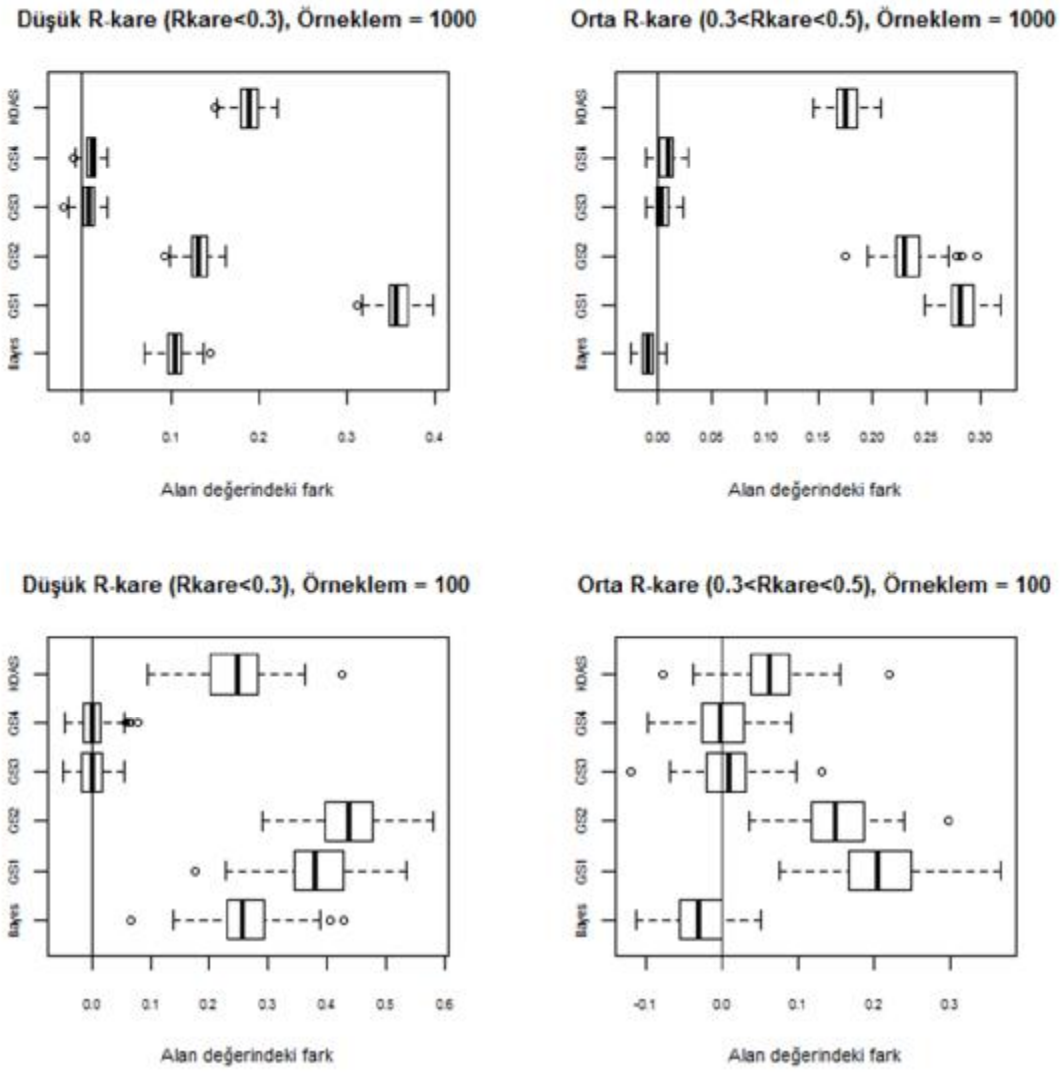
Şekil 4.36 Düşük açıklama oranında Bayes yaklaşımının iterasyonlar sonucunda elde ettiği EAKA değeri (üretilmiş 1 veri seti için)

Bayes yaklaşımının aksine Gizli Sınıf Analizi yaklaşımı açıklama oranından etkilenmemiştir. (Çizelge 4.30, Şekil 4.37)

Çizelge 4.30 Açıklama oranındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı($x10^{-3}$)

		Düşük Açıklama Oranı (R^2<0.3)	Orta Açıklama oranı (0.3<R^2<0.5)
Örneklem büyüklüğü = 1000	KOAS	188±14	176±12
	Bayes Yaklaşımı	105±14	-9±7
	GS1	358±17	284±15
	GS2	132±14	233±21
	GS3	6±8	4±8
	GS4	10±8	9±9
		Düşük Açıklama Oranı (R^2<0.3)	Orta Açıklama oranı (0.3<R^2<0.5)
Örneklem büyüklüğü = 100	KOAS	242±56	61±46
	Bayes Yaklaşımı	259±57	-28±38
	GS1	386±65	205±61
	GS2	433±59	151±49
	GS3	1±24	7±40
	GS4	3±24	-2±39

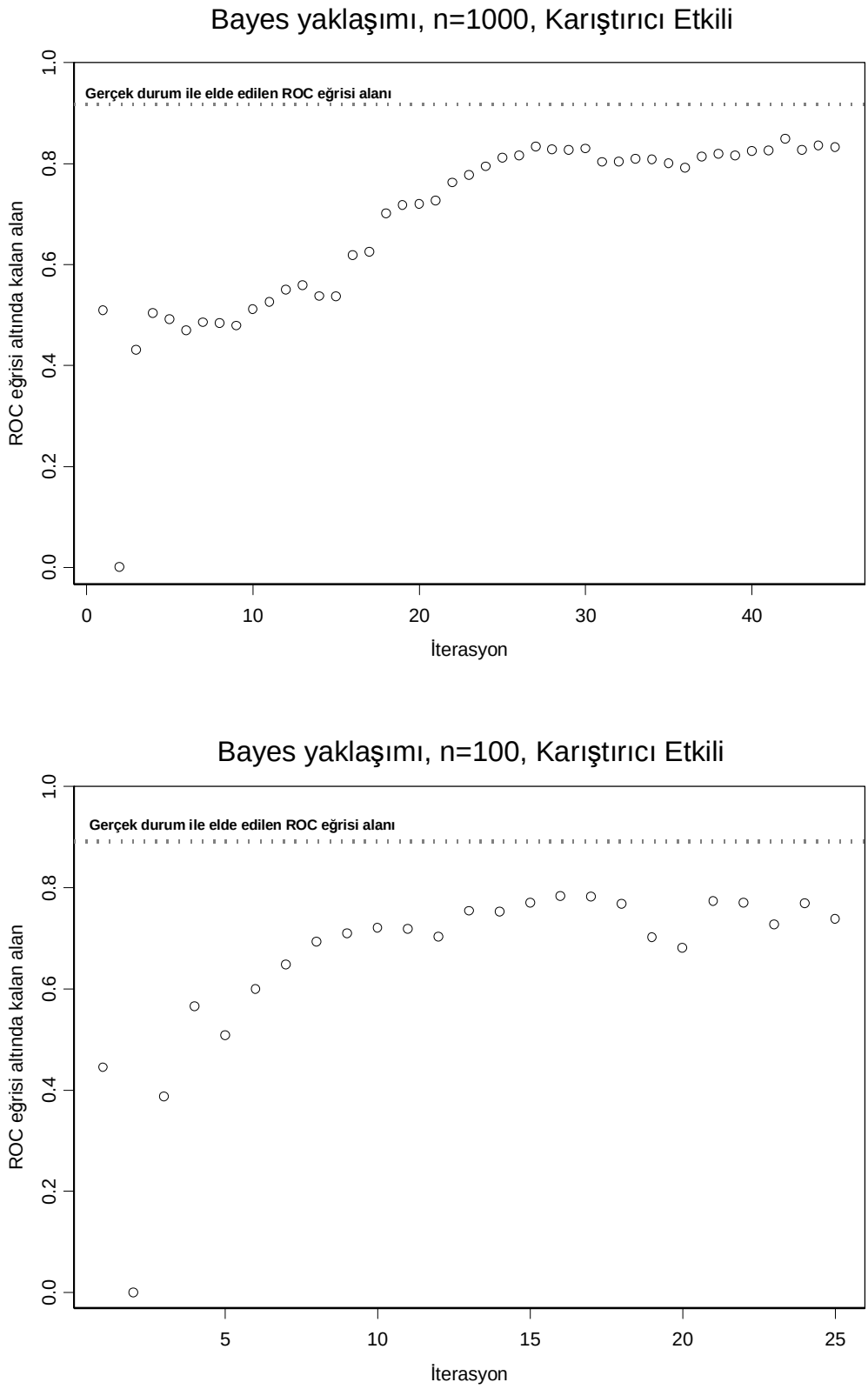
GS1: ikili değişkenler (d1 + d2),
GS2: ikili değişkenler + bir sürekli değişken (d1 + d2 + c1),
GS3: ikili değişkenler + bir sürekli değişken (d1 + d2 + c2),
GS4: tüm değişkenler (d1 + d2 + c1 + c2)



Şekil 4.37 Açıklama oranındaki değişiminin yaklaşımların EAKA nokta tahminine olan etkisi

4.2.3.4 Karıştırıcı etkili veri seti

Karıştırıcı etkili bir değişken ile kurulan modellerde yaklaşımların sonuçlarının etkilendiği görülmüştür. Karıştırıcı etkili değişken model olması gereken iki sürekli değişkenin birbirine bölünmesi ile elde edildiğinden bu değişken ile kurulan modelde diğer iki sürekli değişkenin önemleri azalmakta ve buna göre de regresyon katsayıları değişmektedir. Bayes yaklaşımı ile ilgili yapılan denemelerde bu tip bir veri setinde 45 iterasyondan daha fazla iterasyon görülememiştir. Ancak yine de 45 iterasyon sonucunda elde edilen değer KOAS ile elde edilen değerden daha yüksek olarak saptanmıştır. (Şekil 4.38)



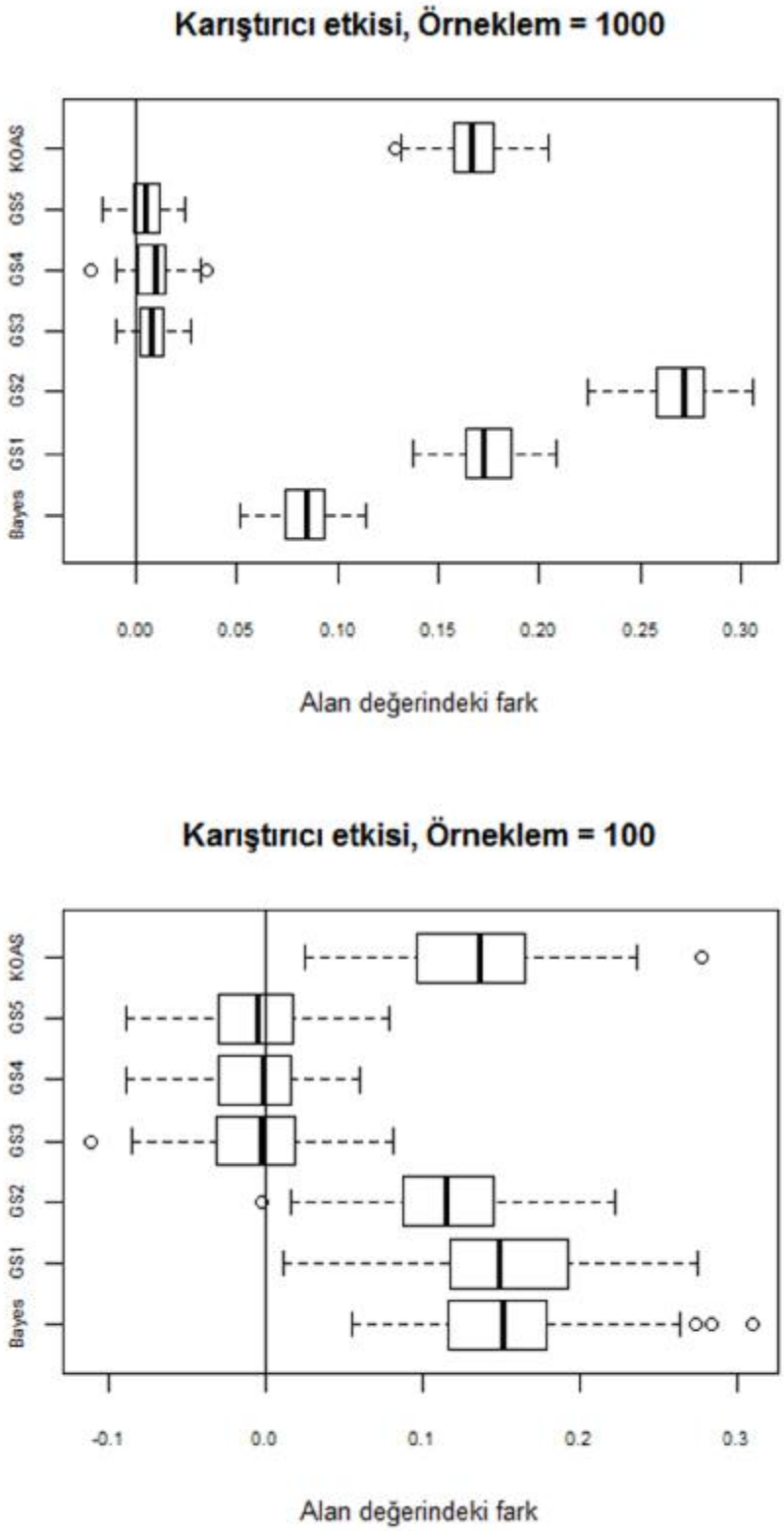
Şekil 4.38 Karıştırıcı etkili veri setinde Bayes yaklaşımının iterasyonlar sonucunda elde ettiği EAKA değeri (üretilmiş 1 veri seti için)

Her ne kadar Bayes KOAS sonuçlarından daha iyi sonuç verse de Gizli Sınıf Analizinin elde ettiği alan değerlerine ulaşamamıştır. Özellikle karıştırıcı değişkeni de içeren GS5 modeli hem büyük veri setinde hem de küçük veri setinde oldukça iyi sonuçlar vermiştir. (Çizelge 4.31, Şekil 4.39)

Çizelge 4.31 Karıştırıcı etkili veri setinin yaklaşımların EAKA nokta tahminine olan etkisi Tahmin değeri – Gerçek değer farkı(x10^{-3})

		Karıştırıcı etkili
Örneklem büyüklüğü = 1000	KOAS	167±14
	Bayes Yaklaşımı	84±13
	GS1	175±15
	GS2	270±16
	GS3	8±8
	GS4	8±10
	GS5	5±9
		Karıştırıcı etkili
Örneklem büyüklüğü = 100	KOAS	133±49
	Bayes Yaklaşımı	154±52
	GS1	151±53
	GS2	114±45
	GS3	-5±35
	GS4	-6±32
	GS5	-5±34

GS1: ikili değişkenler (d1 + d2),
GS2: ikili değişkenler + bir sürekli değişken (d1 + d2 + c1),
GS3: ikili değişkenler + bir sürekli değişken (d1 + d2 + c2),
GS4: tüm değişkenler (d1 + d2 + c1 + c2)
GS5: tüm değişkenler + karıştırıcı değişken (d1 + d2 + c1 + c2 + c3)



Şekil 4.39 Karıştırıcı etkili veri setinin yaklaşımların EAKA nokta tahminine olan etkisi

4.2.3.5 Gerçek Veri Seti Kullanılarak Yapılmış Çalışma Sonuçları

Kardiyolojide A.Bozkurt danışmanlığında M. Koç tarafından yapılmış olan bir uzmanlık tezinden elde edilen verilerin bir kısmı, sürekli bir tanı testi için KOAS yanlılığını düzeltmede önerilen yaklaşımları karşılaştırmada, gerçek veri seti olarak kullanılmıştır. Koç'un çalışmasında fonksiyonel egzersiz kapasitesinin belirlenmesinde kullanılan METS değeri ile serum NT-proBNP değeri arasındaki ilişki araştırılmıştır. METS değerinin 5'in altında olması düşük, üzerinde olması yüksek efor kapasitesi olarak sınıflanmış ve bu sınıflamayı belirlemede kullanılabilecek parametreler incelenmiştir[55]. Bu tezde ise bu çalışmadaki hastaların bir kısmı (95 hasta) kullanılmış ve METS değerine göre oluşturdukları gruplar gerçek durum bilgisi olarak, efor testindeki maksimum kalp atım hızının gruplanmış hali (143 altı ve üzeri) KOAS olarak, NT-proBNP değeri yeni tanı testi olarak, çeşitli eko parametreleri de ek değişkenler olarak alınmış ve yaklaşımlar karşılaştırılmıştır. Karşılaştırmalar sonucunda elde edilen sonuçlar Çizelge 4.32'de sunulmuştur.

Çizelge 4.32 Gerçek veri seti kullanılarak elde edilen yaklaşımların ROC eğri alanları ve standart sapma değerleri

		EAKA
Gerçek Veri seti	Gerçek durum	0.870±0.040
	KOAS	0.727±0.051
	Bayes Yaklaşımı	0.876±0.038
	GS1	0.853±0.042
	GS2	0.949±0.020
	GS3	0.819±0.043
	GS4	0.872±0.038
	GS5	0.841±0.044
	GS6	0.871±0.041

EAKA : Eğri altında kalan alan
GS1: KOAS + iki sürekli değişken (c1 + c2),
GS2: KOAS + üç sürekli değişken (c1 + c2 + c3),
GS3: KOAS + dört sürekli değişken (c1 + c2 + c3 + c4),
GS4: KOAS + iki sürekli değişken (c1 + c4),
GS5: KOAS + iki sürekli değişken (c2 + c4),
GS6: KOAS + üç sürekli değişken (c1 + c2 + c4).

Gerçek EAKA değerini belirlemede KOAS testinin oldukça hatalı olduğu elde ettiği sonuçtan görülmektedir. Öte yandan Bayes yaklaşımının ve bazı GS modellerinin gerçek alan değerini neredeyse tahmin edebildikleri görülmüştür. Ancak alan değerinin tahmini, gerçek durumun hatasız bir şekilde belirlenmesi ile tam olarak hesaplanabilir.

Bunun içinse yaklaşımların hastaları sınıflamadaki hata oranlarının sıfır olması gerekir. Çizelge 4.33'ten de görüldüğü gibi yaklaşımların alan değerlerini doğruya yakın bir şekilde tahmin etmesine rağmen sınıflamada hatalı oldukları görülmektedir. Bu sebeple en uygun yöntemin belirlenmesinde sadece alan değeri tahmini değil, aynı zamanda sınıflamadaki hata oranlarının da göz önüne alınması gerekir. Bu hata oranlarından hareketle Bayes yaklaşımının bu veri seti için en iyi seçenek olmadığı görülmektedir.

Gizli Sınıf modellerini oluştururken kullanılan değişkenler test ile aralarındaki korelasyon değeri dikkate alınarak seçilmiştir. Buna göre test ile aralarında en yüksek korelasyonu olan C1 ve C2 değişkenleridir. C1 ve C2 değişkenlerinin birini veya ikisini içeren modellerin sonuçlarındaki değişimi görmek amacıyla 6 farklı model kurulmuştur. Özellikle korelasyonu en düşük olan C3 modelini içeren durumlarda hatanın artmış olması ve modele alınan değişkenler arasındaki korelasyon incelendiğinde C2 ile C3 ve C2 ile C4 arasında yüksek korelasyonun bulunması nedeniyle C1 ile C2'yi veya C1 ile C4'ü içeren modeller olan GS1 veya GS4 modellerinin KOAS yanlılığının düzeltilmesinde tercih edilmesi önerilebilir. Bu iki modelin sınıflama hataları da diğer modellerden oldukça azdır. (Çizelge 4.33)

Çizelge 4.33 Gerçek veri seti kullanılarak elde edilen yaklaşımların sınıflama hataları

		Hata Oranı*
Gerçek Veri seti	Gerçek durum	1
	KOAS	0.25
	Bayes Yaklaşımı	0.29
	GS1	0.16
	GS2	0.41
	GS3	0.28
	GS4	0.17
	GS5	0.22
	GS6	0.18

* Hata oranı sınıflamada hatalı sınıflanan hastaların toplam hasta sayısına bölünmesi ile elde edilmiştir.
GS1: KOAS + iki sürekli değişken (c1 + c2),
GS2: KOAS + üç sürekli değişken (c1 + c2 + c3),
GS3: KOAS + dört sürekli değişken (c1 + c2 + c3 + c4),
GS4: KOAS + iki sürekli değişken (c1 + c4),
GS5: KOAS + iki sürekli değişken (c2 + c4),
GS6: KOAS + üç sürekli değişken (c1 + c2 + c4).

5 Tartışma

Yeni bir tanı testinin sınıflama başarısı, test uygulanan bireylerin gerçek durum bilgisi olduğunda yansız bir şekilde tahmin edilebilir. Ancak birçok çalışmada, çalışmaya katılan tüm bireylerin yerine sadece belirli bir alt grubun gerçek durum bilgisine ulaşılabilindiğinden yeni tanı testinin sınıflama başarısı yanlı bir şekilde elde edilmektedir. Doğrulama yanlılığı olarak adlandırılan bu yanlılığın düzeltilmesinde literatürde değişik yaklaşımlar önerilmiştir. Tek bir popülasyonda uygulanan ve iki değer alan yeni bir tanı testi için doğrulama yanlılığının düzeltilmesinde eksik ölçüm mekanizmasına göre, Begg ve Greenes yaklaşımı, Zhou tarafından önerilen doğruluk ölçütleri için alt ve üst sınırları, Lojistik Regresyon yaklaşımı ve Çoklu İmputasyon yaklaşımı önerilmiştir. Bu tezde bu yaklaşımlara bir de Yapay Sinir Ağları yaklaşımı ilave edilmiştir.

Literatürde, önerilen yaklaşımların bazıları üstünlükleri ve eksiklikleri gösterilmek amacı ile karşılaştırılmıştır. Harel ve Zhou yapmış oldukları çalışmada Çoklu İmputasyonun doğrulama yanlılığını düzeltmede Begg ve Greenes yaklaşımından daha iyi olduğunu göstermiştir. Çalışmalarında aynı anda birden fazla çoklu imputasyon metodu karşılaştırılmıştır. Karşılaştırılan bu metotların benzer sonuçlar verdikleri ama Begg ve Greenes yaklaşımının imputasyon metotları kadar iyi sonuç vermedikleri sonucuna varılmıştır6. Ancak Hanley ve ark. Harel ve Zhou'nun sonuçlarında aritmetiksel bir hata olduğunu ileri sürmüştür9. Daha sonra de Groot ve ark. bu hatayı Harel ve Zhou'nun çalışmasını tekrarlayarak belirlemiş ve gerçek sonucun onların bulduğu gibi olmadığını göstermiştir. De Groot ve ark. çalışmalarının sonucunda, Begg ve Greenes yaklaşımının sonucunun Çoklu İmputasyon yaklaşımlarının sonuçlarına benzer olduğu bulunmuştur[10].

Doğrulama yanlılığını düzeltmede, önerilen yaklaşımların başarıları ölçülebilen veya ölçülemeyen birçok durumdan etkilenmektedir. Bu durumlardan biri doğrulama yanlılığının ortaya çıkış şeklidir. Çalışmalarda doğrulama yanlılığına neden olan olay; çalışmadaki bazı bireylerin gerçek durum bilgisinin olmamasıdır. Gerçek durum bilgisi olmayan bireylerin rastgele seçilmiş bireyler mi yoksa özellikle seçilmiş kişiler mi

oldukları sorusunun yanıtı uygulanacak olan düzeltmeyi belirlemektedir. Dolaysıyla uygun durum için uygun yöntemi seçmek yanlılığı düzeltmede son derece önemlidir.

Bu tezde, üretilmiş veri ile yapılan analizler sonucunda, gerçek durum sonucu olmayan bireylerin rastgele seçilmiş bireyler olması durumunda Begg ve Greenes yaklaşımının yanlılığı düzeltmede oldukça başarılı olduğu bulunmuştur. Bilindiği üzere Begg ve Greenes yaklaşımı, gerçek durum sonucu olmayan bireylerin rastgele seçilmiş bireyler, yani eksik ölçüm mekanizmasının rastgele (MAR) olduğu durum, için önerilmiştir3.

MAR varsayımı altında kullanılabilecek diğer bir yaklaşım Çoklu İmputasyon yaklaşımıdır. Harel ve Zhou tarafından önerilen bu yaklaşımda, eksik olan gerçek durum bilgisi, test sonucu ve tam olan gerçek durum bilgileri kullanılarak elde edilmiştir. Bu tezde bu yaklaşıma ilave olarak imputasyon sürecinde ek değişkenler (covariates) kullanılmıştır. Buradaki amaç, imputasyon sonucundaki başarının ek değişkenler aracılığı ile artıp artmadığını gözlemlemektir. Analizler sonucunda önemli bir artışın olmadığı hatta bazı durumlarda çok küçük de olsa yaklaşımın, sonucu belirlemedeki hatasının arttığı görülmüştür.

Kosinski ve Barnhart'ın önerdiği Lojistik Regresyon yaklaşımı da uygun modelleme yapıldığında MAR varsayımı altında doğrulama yanlılığını düzeltmede kullanılabilir. Bu yaklaşımı referans alarak, Yapay Sinir Ağları yaklaşımı da bu tezde benzer şekilde MAR varsayımı altında da kullanılmak üzere önerilmiştir. Her iki yaklaşım da gerçek durum tayininde ek değişkenleri kullandıkları için ek değişkenlerin gerçek durumu açıklama oranındaki değişimin yaklaşımların doğruluk ölçütü tahminlerindeki başarılarını etkileyebileceği düşünülmüştür. Bu düşünceden hareketle farklı değerlerde belirlenen tanımlayıcılık katsayılarında yaklaşımların başarıları incelenmiş ve tanımlama katsayısı artışı ile yaklaşımların doğruluk ölçütü tahminlerindeki hata değerlerinde az da olsa artış olduğu gözlenmiştir. Bu durum iki şekilde açıklanabilir. Bunlardan birincisi, verideki yüksek uyumdaki, yanlılık olabilir. Gerçek durum bilgisi eksik ölçüm yapılan bireylerin, yaklaşımlarca modellenmesi sürecinde gerçek durum bilgisi eksik ölçüm olmayan bireylerin bilgilerinin kullanımı yanlılık doğurmuş olabilir. Gerçek durum bilgisi tam olan bireylerin ölçümleri arasındaki uyumun yüksek olması modelleme sonucunda yapılan tahminin hata değerini

arttırabilir. İkinci açıklama ise verideki gerçek durumun eksik ölçüm yapılması, tanımlayıcılık katsayısını değiştirmiş olabileceği yönündedir. Tanımlayıcılık katsayısı değerleri gerçek durum bilgisi tam olduğu durumda kabul edilen değerlerdir. Bu değerler gerçek durum bilgisi belirlenen oranda eksik ölçüm haline getirildikten sonra değişim gösterebilir. Bu değişim de yaklaşımların sonuçlarını etkileyebilir.

Önerilen yaklaşımların yanlılığı düzeltmedeki başarısını etkileyebileceği düşünülen bir diğer durum ise örneklem büyüklüğüdür. Bu tezde büyük (n=1000) ve küçük (n=100) olarak kabul edilen iki farklı büyüklükte veri üretilerek yöntemler birbirleri ile karşılaştırılmıştır. Karşılaştırma sonucunda, örneklem büyüklüğündeki değişimin yaklaşımların sonuçlarını büyük ölçüde etkilemediği gözlenmiştir. Büyük ve küçük örneklemler arasındaki dikkate değer en önemli farklılık, yaklaşımların hata paylarının (gerçek sonuç – yaklaşımın elde ettiği sonuç) standart sapmalarındaki artıştır. Ayrıca küçük örneklemde yaklaşımların hata paylarının küçük de olsa arttığı gözlenmiştir.

Yaklaşımların sonuçlarını etkileyebileceği düşünülen bir diğer nokta da gerçek durum bilgisi bulunmayan bireylerin yüzdesidir. Bu yüzde değerini değiştirmek yanlılığı düzeltmede önerilen tüm yaklaşımların sonuçlarını etkileyebilecektir. Bu değer sırasıyla %30, %50 ve %80 olarak alınmıştır. Analizler sonucunda, önceden de beklendiği üzere eksik ölçüm oranının artmış olmasının tahminlerdeki hata değerini arttırdığı görülmüştür.

MAR varsayımı altında tüm karşılaştırmalardan elde edilen genel sonuç, hem Begg ve Greenes uygulamasının hem de Çoklu İmputasyon yaklaşımının bu durum için en iyi sonuçları verdiği ancak, Begg ve Greenes uygulamasının hem diğer yaklaşımlara göre daha kolay olması hem de herhangi bir ek değişken varlığı gibi bir varsayımı olmamasından dolayı alternatifsiz en uygun yaklaşım olduğudur.

Gerçek durum bilgisi olmayan bireylerin rastgele seçilmediği bir durumda yani eksik ölçüm mekanizmasının rastgele olmadığı (NMAR) durumda yeni tanı testinin performansının tayininde, Zhou tarafından önerilen istatistiğin alt ve üst sınırları kullanılmaktadır4. Ancak burada elde edilen aralığın simetrik olmaması ve geniş olması doğruluk ölçütü hakkında güvenilir bir sonuç çıkarmayı güçleştirmektedir. Bu probleme alternatif olarak literatürde Kosinski ve Barnhart bir çözüm önermiştir. Bu yaklaşımda,

verinin uygunluğuna göre, altın standart test sonucu olmayan bireylerin test sonuçları Lojistik Regresyon yöntemi ile belirlenmektedir5. Bu tezde bu yaklaşıma ilave olarak Yapay Sinir Ağları yaklaşımı da NMAR varsayımı altında doğrulama yanlılığını düzeltmede önerilmiş ve önerilen tüm yaklaşımlarla farklı özellikteki üretilmiş veriler kullanılarak karşılaştırılmıştır.

Karşılaştırmalarda elde edilen sonuçlara göre; üretilmiş büyük veri setinin kullanıldığı analizlerde, eksik ölçüm oranın %30 olduğu durumda, tanımlama katsayısı değerinin düşmesiyle birlikte tahmin yöntemlerinin doğruluk ölçütleri tahminlerindeki hata oranlarının arttığı gözlenmiştir. Bu sonuç yaklaşımların yüksek tanımlayıcılık katsayısında daha başarılı olabileceği yorumunu doğurmaktadır.

Eksik ölçüm oranı %50 olduğu durumda, duyarlılık tahminlerinde önce değişmeyen bir hata değeri, en son tanımlayıcılık katsayısında ise en düşük hata değeri olarak ortaya çıkmıştır. Seçicilik tahmininde ise tanımlayıcılık katsayısındaki düşüşle birlikte hata değerinde bir düşüş gözlenmiştir. Bunun muhtemel sebebi, eksik ölçüm oranındaki artışın yaklaşımların sınıflama başarılarını bozduğu ve dolayısıyla hata değerlerinin artmasına neden olduğudur.

Eksik ölçüm oranın %80 olduğu durumda ise tanımlayıcılık katsayısındaki düşüşe birlikte yaklaşımların duyarlılık tahminlerindeki hata değerleri azalmakta, seçicilik tahminlerindeki hata değerleri ise değişmemektedir. Bu sonuç da %50 eksik ölçüm oranına benzer olarak eksik ölçüm oranın artmasıyla açıklanabilir. Yüksek eksik ölçüm oranında başlangıçta varsayılan tanımlayıcılık katsayıları değerleri değişim gösterebilecektir.

Küçük veri setinin kullanıldığı NMAR analizleri sonucunda ise, eksik ölçüm oranın %30 olması durumunda tanımlayıcılık katsayısının değişmesi yaklaşımların doğruluk ölçütleri tahminlerindeki hata değerlerini etkilememiştir. Eksik ölçüm oranın %50 ve %80 olması durumunda ise tanımlayıcılık katsayısı değeri düştükçe, yaklaşımların doğruluk ölçütleri tahminlerindeki hata değerlerinin düştüğü görülmüştür. Bu durum, büyük veri setindeki sonuçlarda da olduğu gibi, artan eksik ölçüm oranın üretilmiş verinin başlangıç varsayımlarını bozması ile açıklanabilir.

Bu tezde üretilen veriler, gerçek verileri temsil etmesi maksadıyla farklı özellikte üretilmiştir. Tasarlanmış bir çalışmadan elde edilen veride altın standart test sonucunu

tahmin etmede kullanılabilecek birkaç değişken bulunabilir ve bu değişkenlerle model kurulduğunda bu değişkenlerin altın standart test sonucunu açıklama oranı, ölçülemeyen birçok durumun olması nedeniyle, çok yüksek beklenmez. Buradan hareketle bu oranın farklı olduğu birçok veri üretilmiş ve önerilen alternatif yaklaşımların performansları incelenmiştir. Bilindiği üzere tanımlama katsayısı değeri modelin iyiliği konusunda fikir verecektir ve iyi olmayan bir model, sonucu iyi tahmin edemeyecektir. Bu değerin çok küçük olduğu durumlarda bile önerilen alternatif yaklaşımların gerçek duruma yakın değerler vermesi bu yöntemlerin de bu gibi durumlarda kullanılabileceğini göstermektedir.

Gerçek veri setinin kullanıldığı analizlerde ise düzeltme yapılmadan sonuçlar incelendiğinde yeni tanı testinin yüzde 80'lere varan bir duyarlılık değeri elde ettiği görülür. Kosinski ve Barnhart'ın önerdiği Lojistik Regresyon yaklaşımında, eksik veri mekanizma bileşeni için yapılan modelleme sonucunda mekanizmanın MAR olduğu görülmüştür. Bu durumda Begg&Greenes yaklaşımı sonuçları dikkate alınacak olursa, duyarlılık: 0.63 civarında ve seçicilik: 0.69 civarında olabilir. Bu değerler ile düzeltmesiz değerler arasındaki fark, düzeltmenin ne kadar gerekli olduğunu göstermektedir.

Altın standart testinin uygulanmadığı bireylerin yanlı olarak seçilmesi durumunda kullanılması önerilen Zhou alt ve üst sınırları, nokta tahmini içermediğinden doğruluk ölçütü hakkında detaylı bir bilgi vermemektedir. Bu tezde bu sorun, tahmin yöntemlerinin kullanılması ile çözülmeye çalışılmıştır. Çalışmada altın standart test sonucu eksik olan bireyler yanlı veya yansız olarak seçilmiş olsa bile tahmin yöntemlerini kullanmak yeni tanı testinin gerçek performansını belirlemede fikir verebilir.

Eksik ölçümlü bireylerin değerleri, tam ölçümlü bireylerden modelleme yapılarak hesaplanabilir. Daha sonra veri tam hale getirilerek doğruluk ölçütleri elde edilebilir. Burada gözden kaçırılmaması gereken husus, eksik ölçümlü bireylerin tahmin edilen değerlerinin tam ölçümlülere bağımlı olmasıdır. Bu bağımlılık doğruluk ölçütünün yansız tahmininde sorun yaratabilir. Bu konuya Little ve Rubin tarafından önerilen çözüm, eksik ölçüm mekanizmasının doğru belirlenmesidir. Eksik ölçümlü bireylerin yok sayılamayacak olması varsayımı (Nonignorable Missingness) tahmindeki

bağımlılığı ortadan kaldırabilir[32]. Dolayısıyla eksik ölçüm mekanizması bağımlılığı etkileyecektir.

Altın standardın olmadığı veya uygulanmadığı çalışmalarda altın standart kadar iyi sınıflama yapamayan ancak mevcut testler içinde en yüksek sınıflama oranına sahip kesin olmayan altın standart testi kullanılır. Bu testin varlığında, geliştirilen yeni bir tanı testinin sınıflama başarısını incelemek yanlı olacaktır. Bu tezde, KOAS yanlılığını düzeltmede yeni tanı testinin ölçek tipine göre önerilen yaklaşımlar ile bu tezde önerilen yeni bir yaklaşım, üretilmiş ve gerçek veri setleri kullanılarak karşılaştırılmıştır.

Yeni tanı testinin ölçek tipi isimsel ve özellikle ikili (binary) olduğu durumda KOAS yanlılığını düzeltmek için literatürde çeşitli yaklaşımlar önerilmiştir. KOAS belirli bir hata değerine sahip olduğundan yeni tanı testi ile uyumsuz sonuç verdiğinde hangi testin (KOAS mı, yeni tanı testi mi?) doğru sonucu gösterdiği konusunda şüpheler oluşabilecektir. Bu uyumsuzluğu çözmek için Tutarsızlık Çözücü yaklaşımı sıklıkla başvurulan bir yaklaşımdır[13]. Bu yaklaşımda iki test arasında uyumsuzluğa neden olan durumu çözmek için üçüncü bir teste – çözücü test – başvurulur. KOAS yanlılığını düzeltmede kullanılan diğer bir yaklaşım ise Bileşke Referans Standardıdır. Bu yaklaşımda tanı için önerilen birden fazla test sonucundan bir kombinasyon oluşturularak gerçek durum belirlenmeye çalışılır.

Bu tezde, Tutarsızlık Çözücü ve Bileşke Referans Standardı yaklaşımları ve bu yaklaşımlara alternatif olabilecek Gizli Sınıf Analizi yaklaşımı, üretilmiş ve gerçek veri setleri kullanılarak karşılaştırılmıştır. Karşılaştırmalarda yaklaşımların sonuçlarını etkileyebilecek örneklem büyüklüğü değişikliği, hastalık prevelansı değişimi, doğruluk ölçütü değişimi gibi durumlar dikkate alınarak veriler üretilmiştir.

Prevelans değeri 0.5 olarak alındığında, örneklem büyüklüğünün değişmesi ile yaklaşım tahminlerindeki hata değerinin değişmediği gözlenmiştir. Ancak prevelansın 0.8 olması durumunda ise yaklaşımların seçicilik tahminindeki hata değerlerinin duyarlılık tahminlerine göre oldukça yüksek olduğu ve örneklem büyüklüğünün artması ile doğruluk ölçütü tahminlerindeki hata değerinin az da olsa arttığı görülmüştür. Bu sonuç da yaklaşımların örneklem büyüklüğündeki değişimden çok prevelanstaki değişimden daha çok etkilendiği şeklinde yorumlanabilir.

Hem büyük hem de küçük veri setinde prevelans değerinin 0.1 – 0.9 arasında değişmesi ile yaklaşımların yanlılığı düzeltmedeki başarısını etkilendiği gözlenmiştir.

Buna göre hem büyük veri setinde hem de küçük veri setinde prevelans değeri arttıkça Gizli Sınıf Analizi yaklaşımında kullanılan modellerin, duyarlılık tahminlerindeki hata değeri artarken, seçicilik tahminlerindeki hata değeri azalmaktadır. Diğer taraftan Çözücü test ve Bileşke Referans Standardında ise prevelansın artması duyarlılık tahminlerindeki hata değerinin azalmasına, seçicilik tahminlerindeki hata değerinin ise artmasına neden olmuştur. Beklenildiği üzere prevelans değeri doğruluk ölçütlerinin tahminlerini etkilemiştir. Elde edilen sonuçlardan hareketle prevelans değerine göre Gizli Sınıf Analizi modelleri arasında tercih yapılabilir. Buna göre tüm yaklaşımlar arasında GS4 modeli prevelansın 0.4'e kadar olan değerlerinde özellikle duyarlılık tahmini için daha iyiyken, yine GS4 modeli prevelansın 0.6 ile 0.9 arası aldığı değerlerde ise özellikle seçicilik tahmininde tüm yaklaşımlar arasında daha iyi sonuç vermiştir.

Doğruluk ölçütlerindeki değişim de yaklaşımların tahminlerini etkilemiştir. Özellikle artan doğruluk ölçütü değeri ile GS modelleri daha iyi sonuçlar vermeye başlamış ve önerilen diğer yaklaşımların önüne geçmiştir.

Gerçek verinin kullanıldığı analizlerde ise Gizli Sınıf Analizi yaklaşımı dışındaki yaklaşımlar duyarlılık değerini yüksek bulurken, Gizli Sınıf Analizi seçicilik değerini daha yüksek bulmuştur. Üretilmiş veri çalışmalarından elde edilen sonuçlarda 80 kişilik veri setinde prevelansın yaklaşık 0.5 olduğu (KOAS pozitif oranı:0.56) durumda Gizli Sınıf Analizi yaklaşımlarının daha iyi sonuç verdiği görüldüğü için bu yeni tanı testinin Duyarlılık ve Seçicilik değerlerinin Gizli Sınıf Analizi sonucunda elde edilen değerler olarak kabul edilmesi daha uygun olur.

Hadgu Tutarsızlık Çözücü yaklaşımının tercih edildiği durumun genellikle KOAS testinin %100 seçici (az duyarlı) veya %100 duyarlı (az seçici) olduğu durum olduğunu belirtmiştir[13]. Ancak yaklaşıma başvurulan durumların sadece hatalı pozitif veya hatalı negatif durumlarında olması tartışılan bir konudur. Meier bu soruna bir çözüm getirmek için bu yaklaşımın uygulamasını biraz değiştirerek, sadece pozitif veya hatalı negatif sonuçlarına değil bunlara ilave olarak diğer durumlardan rastgele olarak seçilmiş bir alt gruba da uygulamıştır[39]. Ancak bu yaklaşım, yeterli istatistik dayanağı bulunmadığından uygulamada sık tercih edilen bir yaklaşım olamamıştır.

İkili tanı testi durumunda KOAS yanlılığını düzeltmede önerilen diğer yaklaşım olan Bileşik Referans Standardı yaklaşımı ise gerçek durumun belirlenmesinde tek bir

KOAS testi kullanmak yerine, oluşturulan bu Bileşik Referansın daha etkili olacağı düşüncesiyle oluşturulmuştur. Bileşik Referans test uygulamasının sağladığı en büyük fayda, hastalık tanımının oluşturulmasıdır. Ancak her tanı için yeterli test sayısının bulunmaması ve bu testler arasında her zaman uygun kombinasyonun oluşturulamaması bu yaklaşımın önündeki en büyük engeldir.

Yeni tanı testinin sıralı ölçekte olması durumunda KOAS yanlılığını düzeltmede Zhou ve ark.[16] tarafından parametrik olmayan ROC eğrisi yaklaşımı önerilmiştir. Bu yaklaşım, sıralı yapıdaki en az 3 testin varlığında ROC eğrisi altında kalan alan değerini parametrik olmayan iteratif bir yöntemle hesaplayan bir yaklaşımdır. Bu tezde, bu yaklaşım kendisine alternatif olarak önerilen Gizli Sınıf Analizi yaklaşımı ile üretilmiş ve gerçek veri setleri kullanılarak karşılaştırılmıştır. Yapılan karşılaştırmalarda yaklaşımların veri setlerindeki özelliklerden ne kadar etkilendikleri belirlenmeye çalışılmıştır. Buna göre veri setindeki hastalık prevelansı 0.5 değerinden uzaklaştıkça Parametrik olmayan ROC yaklaşımının hata oranının arttığı gözlenmiştir. Benzer durumda ise Gizli Sınıf Analizinin hata oranının azaldığı dikkat çekmiştir.

Zhou ve ark. önerdiği Parametrik olmayan ROC yaklaşımında temel varsayım sıralı yapıdaki en az 3 testin varlığıdır. Bu varsayımın, test sayısı ve testlerin kategori sayılarındaki değişimden nasıl etkilendiği veri üretilerek incelenmiştir. Buna göre kategori sayısı düşük ve test sayısı az olan bir durumda Parametrik olmayan ROC yaklaşımının hata oranı artmaktadır. Öte yandan kategori sayısındaki artış veya test sayısındaki artış ile yaklaşımın hata oranının azaldığı görülmüştür. Gizli Sınıf Analizi yaklaşımının ise test ve kategori sayısındaki değişimden diğer yaklaşım kadar etkilenmediği gözlenmiştir.

Eğri altındaki alan değerinin değişiminin, Gizli Sınıf Analizi yaklaşımının hata oranlarında değişime neden olduğu görülmüştür. Buna göre bu yaklaşımın hata oranı, düşük alan değerinde yüksek alan değerine göre daha yüksektir.

Zhou ve ark. önerdiği yaklaşımın en büyük dezavantajı aynı kategori sayısına sahip sıralı en az 3 test varsayımının olmasıdır. Ancak birçok hastalığın tanısını koymada kullanılabilecek bu özellikte testler bulmak mümkün değildir. Dolayısıyla bu yaklaşımın uygulama alanı oldukça kısıtlıdır.

Gerçek veri setinin kullanıldığı durumda ise elde edilen sonuçların genellikle benzer olması yaklaşımların bu tip verilerde kullanımının uygun olabileceğini

düşündürmektedir. Diğer taraftan gerçek veri ile elde edilen sonuçlarda bazı noktaların, örneğin eğitim sonrası ölçümlerinde 2 ve 4 numaralı öğretim üyelerinin Parametrik olmayan ROC yaklaşımı ile elde edilen alan değerlerinin, Gizli Sınıf Analizine göre düşük olması bu yaklaşım ile ilgili bazı sorunlarla karşılaşılabilineceği düşüncesine neden olmaktadır.

Yeni tanı testinin aralık veya oran ölçek tipinde olması durumunda ise literatürde Wang ve ark.[17] tarafından Bayes yaklaşımı önerilmiştir. Bu yaklaşımda bilinmeyen gerçek durum bilgisi, belirli bir başlangıç değeri alınarak iterasyonlar sonucunda ek değişkenlerce modellenerek, tahmin edilmeye çalışılmaktadır. Bu çalışmada, bu yaklaşımın bağımlı büyük veri setlerinde sorunsuz çalıştığı görülmüştür. Ayrıca veri setinin prevelansındaki veya EAKA değerindeki değişimin Bayes yaklaşımındaki iterasyonları etkilemediği ve belirli bir sayıda iterasyon sonucunda gerçek durumu doğru olarak belirleyebildiği görülmüştür. Ancak ek değişkenlerin gerçek durumu açıklama oranının düşük veya orta seviyede olması ya da karıştırıcı bir ek değişkenin varlığı Bayes yaklaşımının sonucunu etkilemektedir.

Bu tezde önerilen Gizli Sınıf Analizi yaklaşımı ise Bayes yaklaşımının çok iyi olduğu durumlarda Bayes yaklaşımına yakın değerler almıştır. Diğer taraftan Bayes yaklaşımının sorun yaşadığı durumlarda ise Gizli Sınıf Analizi sorun yaşamamıştır. Ek değişkenlerin gerçek durum açıklama oranının herhangi bir seviyede olması (düşük, orta veya yüksek) Gizli Sınıf Analizinin sınıflama yapabilmesini engellememiştir. Hatta modele, karıştırıcı bir ek değişken girdiğinde sınıflama hata oranı düşmüştür. Bu durum, Gizli Sınıf Analizi yaklaşımının, veri setinin belirtilen özelliklere sahip olması durumunda Bayes yaklaşımına alternatif olabileceğini düşündürmektedir.

Bayes yaklaşımının başarılı olduğu durumlarda, maksimum iterasyon sayısına ulaşmadan sınıflamayı yaptığı gözlenmiştir. Bu durum maksimum iterasyon sayısının veri setinin özelliğine göre belirlenmesi gerekliliğini doğurmuştur. Benzer şekilde Bayes yaklaşımının başarısının düşük olduğu durumlarda ise iterasyonların çok fazla ilerlemediği görülmüştür.

Gizli Sınıf Analizinde kurulan modeller arasındaki farklılığı saptamada LogLikelihood değeri ile birlikte hatalı sınıflama oranı değeri de göz önünde bulundurulmuştur. Model belirlemede sadece LogLikelihood değerini kullanmanın sakıncaları şöyle görülebilir. Örneğin Çizelge 4.27'te verilen değerlere göre en düşük

LL değerli model, Model I'dir. Ancak bu model sınıflama için seçilirse elde ettiği gerçek durum tahmini %50 oranda doğrudur. Yani hasta dediklerinin %50'si ve yine sağlıklı dediklerinin %50'si gerçekte doğrudur. Bu durum LatentGold paket programı tarafından hesaplanan Class. Err. değeri ile gösterilmeye çalışılmıştır. Bu sebeple model belirlemede hatalı sınıflama oranı göz önünde bulundurulmalıdır.

Bu tezde yeni tanı testinin aralık veya oran ölçek tipinde olması durumunda KOAS yanlılığını düzeltmede, veri setinin küçük veya büyük olmasının yaklaşımların sonuçlarını etkilediği görülmüştür. Büyük veri setlerinde Bayes yaklaşımı küçük veri setlerine göre daha başarılıdır. Küçük veri setlerindeki birçok durumda Bayes yaklaşımı maksimum iterasyon sayısına ulaşamamıştır. İterasyon, tekrarlı düzenlemelerde karşılaşılan bazı hatalar (matrisin tersinin olmaması, sıfıra bölme hatası, yakınsama problemi gibi) nedeniyle durdurulmuştur. Buna rağmen Gizli Sınıf Analizi ise küçük veri setlerinde oldukça başarılıdır. Prevelansın düşük olması durumu dışındaki tüm durumlarda gerçek durumu doğru olarak belirleyebilmektedir. Küçük veri setinde, Gizli Sınıf Analizinin düşük prevelans (0.1) değerlerinde sorun yaşadığı gözlenmiştir.

Bayes yaklaşımı ile ilgili karşılaşılan bir sorun, verinin analizinin yapılması için ortalama 20 deneme yapılması gerekliliğidir. Bayes yaklaşımında veri analiz edilirken iterasyonlar sırasında bir sorun ile karşılaşıldığında iterasyon durdurulmaktadır. Burada sorun olarak kastedilen genellikle nxn boyutlu matrisinin, tersinin alınamaması veya tersi alındığında elde edilen matrisin simetrik olmaması durumudur. Bu sorun ancak iterasyona en baştan başlanarak aşılabilmektedir. Çünkü sorun ile karşılaşılan iterasyonlar genellikle yakınsama eğilimli değildir. Bu çalışmada 5000 iterasyon yapacak şekilde kod yazılmıştır. Ancak büyük veri setinin kullanıldığı birçok durumda bu sayının çok yüksek olduğu ve küçük sayılarda iterasyonda yakınsamanın sağlandığı görülmüştür. Ayrıca Bayes yaklaşımının sınıflama başarısının düşük olduğu durumlarda çok fazla deneme (bazı durumlarda 50'den fazla) yapılarak ulaşılabilecek maksimum iterasyona ulaşılmaya çalışılmıştır.

Gerçek veri seti kullanılarak yapılan analizler sonucunda ise Bayes yaklaşımının alan değerini çok küçük bir hata ile tahmin edebildiği ve bu tahmini yaparken hastaları sınıflamadaki başarısının oldukça yüksek olduğu görülmüştür. Gizli Sınıf Analizinin ise test ile yüksek korelasyonu olan değişkenlerce oluşturulacak model sonucunda sınıflamada en düşük hata oranına ve alan tahmininde gerçek değere en yakın değere

sahip olduğu görülmüştür. Bu veri seti için Gizli Sınıf Analizinin Bayes yaklaşımından daha iyi sonuç verdiği gözlenmiştir.

Gerçek veri setinde Bayes yaklaşımının sınıflamada yüksek hata yapması, yaklaşımın iterasyon süresince karşılaştığı sıfıra bölme hatası gibi problemlerden kaynaklanmaktadır. Bu tip problemler Bayes yaklaşımının analiz sonuçlarını etkilediğinden bunların düzeltilmesi için alternatif yazılımlara ve bu yazılımlar için geliştirilmiş hazır kodlara ihtiyaç duyulmaktadır.

6 Sonuç ve Öneriler

Bu tezde yeni bir tanı testi geliştirilirken bu testin doğruluk ölçütlerini belirlemede karışılabilecek sorunlar olan Doğrulama Yanlılığı ve Kesin Olmayan Altın Standart Yanlılığı için önerilmiş çözümler incelenmiş ve bu çözümlerin sorun yaşadığı durumlarda alternatif yaklaşımlar önerilmiştir.

Yeni tanı testinin isimsel ölçekte, özellikle ikili olması durumunda, doğrulama yanlılığının düzeltilmesinde kullanılan yaklaşımlar gerçek durumları doğrulanmamış bireylerin seçilme mekanizmasına bağlıdır. Bu mekanizma bilinen ve gözlenen özelliklere bağlı şekilde gerçekleşmişse, (yani rastgele eksik ölçümler MAR) literatürde önerilen yaklaşımlar içinde karşılaştırma yapılan tüm durumlar için en başarılı yaklaşım Begg ve Greenes yaklaşımı olmuştur. Bu mekanizma bilinmeyen veya gözlenemeyen özelliklere bağlı şekilde gerçekleşmişse, (yani rastgele olmayan eksik ölçümler NMAR) önerilen yaklaşımların varsayımları dikkate alınarak eksik ölçümlerin üretilmesini sağlayan yaklaşımlar (Lojistik Regresyon veya Yapay Sinir Ağları) kullanılabilir.

Kesin olmayan altın standart yanlılığın düzeltilmesinde ise önerilen yaklaşımlar, ilgilenilen tanı testinin ölçek tipine göre değişmektedir. Tanı testinin isimsel ölçekte olması durumunda KOAS yanlılığını düzeltmede uygun prevelans aralığında ve yüksek doğruluk ölçütü değerlerinde Gizli Sınıf Analizi yaklaşımı tercih edilebilir. Tanı testinin sıralı ölçekte olması durumunda ise yanlılık uygun prevelans aralığında ve düşük olmayan kategori-test sayısı durumunda Gizli Sınıf Analizi yaklaşımı ile düzeltilebilir. Tanı testinin aralık veya oran ölçek tipinde olması durumunda ise KOAS yanlılığını düzeltmede küçük örneklem, düşük eğri altında kalan alan ve karıştırıcı etkili değişken varlığı dışındaki durumlarda Bayes yaklaşımı, bu yaklaşımın sorun yaşadığı durumlarda ise Gizli Sınıf Analizi kullanılabilir.

7 Kaynaklar

1. **Sox H.C.Jr., Blott M.A., Higgins M.C., Marton K.I.** "Medical Decision Making", Butterworths-Heinemann, **1989**

2. **McNeil B.J., Adelstein S.J.** "Determining the value of diagnostic and screening tests" J. Nucl. Med. **1976**; 17, 439-448

3. **Begg C., Greenes R.** "Assessment of diagnostic tests when disese is subject to selection bias", Biometrics, **1983**; 39, 207-216.

4. **Zhou X.H.** "Maximum likelihood estimators of sensitivity and specificity corrected for verification bias" Commun. Stat. Theory Meth. **1993**; 22, 3177-3198

5. **Kosinski A.S., Barnhart H.X.** "Accounting for Nonignorable Verification Bias in Assessment of Diagnostic Tests" Biometrics **2003**; 1, 163-171.

6. **Harel O., Zhou X.H.** "Multiple Imputation for correcting verification bias" Stat. Med. **2006**; 25, 3769-3786

7. **Martinez E.Z., Achcar J.A., Neto F.L.** "Estimators of sensitivity and specificity in the presence of verification bias: A Bayesian approach." Computational Statistics & Data Analysis **2006**; 51, 601 – 611.

8. **Buzoianu M., Kadane J.B.** "Adjusting for verification bias in diagnostic test evaluation: A Bayesian approach." Statistics in Medicine **2008**; 27, 2453-2473.

9. **Hanley J.A., Dendukuri N., Begg C.B.** "Letter to Editor: Multiple imputation for correcting verification bias by Ofer Harel and Xiao-Hua Zhou. Statistics in Medicine 2006; 25: 3769–3786" Statistics in Medicine **2006**; 26, 3046–3047.

10. **de Groot J.A.H., Janssen K.J.M., Zwinderman A.H., Moons K.G.M. and Reitsma J.B.** "Multiple imputation to correct for partial verification bias revisited." Statistics in Medicine **2008**; 27, 5880–5889.

11. **Zhou X-H., Obuchowski N.A., McClish D.K.,** "Statistical Methods in Diagnostic Medicine", Wiley, **2002**.

12. **Henkelman R.M., Kay I., Bronksill M.J.** "Receiver operator characteristic (ROC) analysis without truth" Medical Decision Making **1990**; 10, 24-29.

13. **Hadgu A. "**The discrepancy in discrepant analysis." Lancet **1996**; 348, 592–593**.**

14. **Hui S.L. and Zhou X.H.** "Evaluation of diagnostic tests without gold standards" Statistical Methods in Medical Research **1998**; 7, 354-370.

15. **Hall P. and Zhou X.H.** "Nonparametric Estimation of component distributions in a multivariate mixture" Annals of Statistics **2003**; 31, 201-224

16. **Zhou X.H., Castelluccio P., Zhou C.** "Nonparametric Estimation of ROC Curves in the Absence of a Gold Standard" Biometrics **2005**; 61, 600-609

17. **Wang C., Turnbull B.W., Gröhn Y.T., Nielsen S.S.** "Estimating Reciever Operating Characteristic Curves with Covariates When There is No Perfect Reference Test for Diagnosis of Johne's Disease" J. Dairy Science **2006**; 89, 3038-3046.

18. **Reid M.C., Lachs M.S., Feinstein A.R.** "Use of methodological standards in diagnostic test research: Getting better but still not good" JAMA, **1995**; 274, 645-651

19. **Sheps S.B., Schester M.T.** "The assessment of diagnostic tests: A survey of current medical research", JAMA, **1984**; 252, 2418-2422

20. **Jaeschke R., Guyatt G.H., Sackett D.L.** "User's guides to the medical literature III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients?" JAMA, **1994**; 271, 703-707

21. **Greenhalgh T.** "How to read a paper. Papers that report diagnostic or screening tests" BMJ, **1997**; 315, 540-543

22. **Genç Y.** "Tanı testi çalışmalarında metodolojik standartların kullanılması" A.Ü. Tıp Fakül. Mecmuası, **2003**; 56, 259-264

23. **Youden W.J. "**Index for rating diagnostic tests". Cancer. **1950**; 3: 32-35

24. **Choi B.C.K.** "Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test" Am. J. Epidemiol. **1998**; 148, 1127-1132

25. **Zweig M.H. and Campell G.** "Receiver-operating characteristic plots: A fundamental evaluation tool in clinical medicine" Clin.Chem. **1993**; 39, 561-577.

26. **Metz C.E. and Kronman H.** "Statistical significance tests for binormal ROC curves" J. Math. Psychol. **1980**; 22, 218-243

27. **Metz C., Wang P. and Kronman H.** "A new approach for testing the significant difference between ROC curves measured from correlated data" in F.Deconinck (ed.) Information processing in medical imaging, **1984**, Nijihoff, The Hague.

28. **Venkatraman E.S. and Begg C.** "A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment" Biometrika **1996**; 83, 835-848

29. **Venkatraman E.S.** "A permutation test to compare receiver operating characteristic curves" Biometrics **2000**; 56, 1134-1138

30. **Greenes R. and Begg C.** "Assessment of diagnostic technologies: Methodology for unbiased estimation from samples of selective verified patients" Invest. Radiol. **1985**; 20, 751-756

31. **Bates A.S., Margolis P.A., Evans A.T.** "Verification bias in pediatric studies evaluating diagnostic tests" J. Pediatr. **1993**; 122, 585-590

32. **Little R.J.A., Rubin D.B.** "Statistical Analysis with Missing Data" Wiley **2002**

33. **Agresti A.** "Categorical Data Analysis" Wiley **2002**

34. **Rubin D.B.** "Multiple Imputation for Nonresponse in Surveys." Wiley **1987**.

35. **Ripley B.D.** "Pattern Recognition and Neural Network" Cambridge University Press **1996**.

36. **Raudys S.** "Statistical and Neural Classifiers: An Integrated Approach to Design" Springer **2001**

37. **Haykin S.** "Neural Networks: A Comprehensive Foundation, Second Edition," Englewood Cliffs, NJ: Prentice-Hall, **1999**.

38. **Rumelhart D.E., Hinton G.E., Williams R.J.** "Learning representations by back-propagating errors" Nature, **1986**; 323, 533-536.

39. **Meier K.** FDA document, Microbiology Devices Panel Medical Advisory Committee Meeting, **1998**.

40. **Alonzo** T.**A. and Pepe M.S.**, "Using a combination of reference tests to assess the accuracy of a new diagnostic test". Stat. Med. **1999**; 18, 2987–3003

41. **Pepe M.S.** "The statistical evaluation of medical tests for classification and prediction", Oxford University Press, USA, **2005**

42. **Jeffreys H.** "Theory of Probability" 3rd edition, Oxford University Press, Oxford, UK, **1961**

43. **Kaufman L. and Rousseeuw P.J.** "Finding groups in data: An introduction to cluster analysis". New York: John Wiley and Sons, Inc **1990**

44. **Vermunt J.K. and Magidson J. "**Latent GOLD User's Manual". Boston: Statistical Innovations Inc. **2000**

45. **Lazarsfeld P.F.,** "The logical and mathematical foundation of latent structure analysis & The interpretation and mathematical foundation of latent structure analysis". S.A. Stouffer et al. (eds.), Measurement and Prediction, 362-472. Princeton, NJ: Princeton University Press. **1950**

46. **Goodman L.A.,** "The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach", American Journal of Sociology, **1974**; 79, 1179-1259

47. **Haberman S.J., "**Analysis of Qualitative Data, Vol 2, New Developments." New York: Academic Press. **1979**

48. **Skrondal A., Hesketh S.R.** "Generalized Latent Variable Modelling" ChapmanHall Press **2004**

49. **Garrett E.S., Eaton W.W., Zeger S.** "Methods for evaluating the performance of diagnostic tests in the absence of a gold standard: a latent class model approach", Statistics in Medicine **2002**; 21, 1289-1307

50. **Sukan A., Reyhan M., Aydin M., Yapar A.F., Sertdemir Y., Canpolat T. and Aktas A.** "Preoperative evaluation of hyperparathyroidism: the role of dual-phase parathyroid scintigraphy and ultrasound imaging." Annals of Nuclear Medicine **2008**; 22, 123-131.

51. **R Development Core Team**. R: A language and environment for statistical computing, reference index version 2.11.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org, **2005**.

52. **van Buuren S., Oudshoorn K.** "Flexible multivariate imputation by mice. Technical Report, TNO prevention 13 and Health, Leiden, The Netherlands." **1999** Available from: http://www.stefvanbuuren.nl/publications/Flexible%20multivariate%20-%20TNO99054%201999.pdf (son erişim tarihi 20 Ekim 2010).

53. **Koçak F.** "Helicobacter pylori infeksiyonunda yeni bir tanı yöntemi olan stool antijen testinin değerlendirilmesi" Ç.Ü. Tıpta Uzmanlık Tezi, **2002**.

54. **Günaştı S., Mülayim M.K, Fettahlıoğlu B., Yucel A., Burgut R, Sertdemir Y, Aksungur VL** "Interrater agreement in rating of pigmented skin lesions for border irregularity" Melanoma Research **2008**; 18, 284–288.

55. **Koç M.** "Kronik Kalp Yetersizliğinde Fonksiyonel Kapasite Ve Prognozu Değerlendirmede Serum NT-proBNP Düzeyi" Ç.Ü. Tıpta Uzmanlık Tezi **2006**.

56. **The MathWorks Inc**, MATLAB version 2008b, Natick, Massachusetts, **2003**.

57. **Statistical Innoviations**, Latent Gold version 4.5, Belmont MA, **2008.**

ÖZGEÇMİŞ

1978 yılında Konya'nın Ereğli ilçesinde doğdum. Doğumundan kısa bir süre babamın işi nedeni ile İzmir'e yerleşip burada ilk ve ortaokulu okudum. Liseyi ortaokul sonrası girmiş olduğum sınavla kazandığım Konya Meram Fen Lisesinde tamamladım. 1996 yılında Öğrenci Yerleştirme Sınavı sonucu Orta Doğu Teknik Üniversitesi Matematik Öğretmenliği bölümünü kazandım. Bu bölümün üçüncü yılında yine aynı üniversitenin Matematik bölümüne yatay geçiş yaptım ve buradan 2001 yılında mezun oldum. Aynı yıl başvurup kabul edildiğim Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim dalında yüksek lisans programını 2004 yılında bitirip, yine aynı anabilim dalında Doktora programına başladım. Yüksek Lisans ve Doktora eğitimim süresince birçok ulusal ve uluslararası kongrelere bildiri sunarak katıldım. Yine bu süre içinde birçok çalışmanın içinde bulunup, bu çalışmaların bilimsel dergi ve kongrelerde yayınlanmasına katkı sağladım. Anabilim dalı bünyesinde verilen lisans-yüksek lisans dersleri ve çeşitli kursların verilmesinde yardımcı öğretici olarak görev aldım. Aynı anabilim dalında 2001 yılında atandığım Araştırma Görevlisi görevini halen devam ettirmekteyim.