



**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**



RANDOMİZE OLMAYAN KLİNİK ÇALIŞMALARDA EN UYGUN EŞLEŞTİRME ANALİZİ İÇİN MAKİNE ÖĞRENME ALGORİTMALARI İLE YENİ PROPENSITY SKOR TAHMİN MODELLERİNİN GELİŞTİRİLMESİ

Emre DEMİR

**BİYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ**

**DANIŞMAN
Doç. Dr. Serdal Kenan KÖSE**

ANKARA

2019

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**RANDOMİZE OLMAYAN KLİNİK ÇALIŞMALARDA EN
UYGUN EŞLEŞTİRME ANALİZİ İÇİN MAKİNE ÖĞRENME
ALGORİTMALARI İLE YENİ PROPENSITY SKOR TAHMİN
MODELLERİNİN GELİŞTİRİLMESİ**

Emre DEMİR

**BİYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ**

**DANIŞMAN
Doç. Dr. Serdal Kenan KÖSE**

ANKARA

2019

Etik Beyan

Ankara Üniversitesi

Sağlık Bilimleri Enstitüsü Müdürlüğü'ne,

Doktora tezi olarak hazırlayıp sunduğum “Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi” başlıklı tez; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin fikir/hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan deneysel çalışma/araştırma tarafımdan yapılmış olup, tüm cümleler, yorumlar bana aittir.

Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Öğrencinin Adı Soyadı: Emre DEMİR

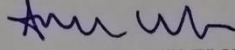
Tarih: 17.01.2019

İmza:

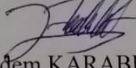
KABUL VE ONAY

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü
Biyoistatistik Anabilim Dalında
Emre DEMİR tarafından hazırlanan
“Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için
Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin
Geliştirilmesi” adlı tez çalışması
aşağıdaki jüri tarafından DOKTORA TEZİ olarak OY BİRLİĞİ ile kabul edilmiştir.

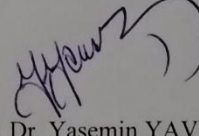
Tez Savunma Tarihi: 17.01.2019



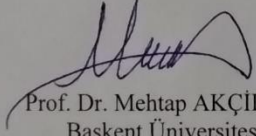
Prof. Dr. Atilla Halil ELHAN
Ankara Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Jüri Başkanı



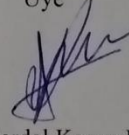
Prof. Dr. Erdem KARABULUT
Hacettepe Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye



Prof. Dr. Yasemin YAVUZ
Ankara Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye



Prof. Dr. Mehtap AKÇİL OK
Başkent Üniversitesi
Beslenme ve Diyetetik
Programı
Üye



Doç. Dr. Serdal Kenan KÖSE
Ankara Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Üye
Tez Danışmanı

Tez hakkında alınan jüri kararı, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü
Yönetim Kurulu tarafından onaylanmıştır.

Prof. Dr. Mehmet AKAN
Sağlık Bilimleri Enstitüsü Müdürü

İÇİNDEKİLER

Etik Beyan	ii
Kabul ve Onay	iii
İçindekiler	iv
Önsöz	vi
Simgeler ve Kısaltmalar	vii
Şekiller	ix
Çizelgeler	xii
1.GİRİŞ	1
1.1. Nedensel Çıkarım	7
1.1.1. Deneysel Çalışmalar	8
1.1.1.1. Yanlılık	9
1.1.1.2. Randomizasyon	10
1.1.2. Gözlemsel Çalışmalar	11
1.1.2.1. Karıştırıcı Faktörler (Confounding)	12
1.1.2.2. Etki Değişimi (Effect modification)	17
1.2. Deneysel Hatayı Azaltma Yöntemleri	18
1.2.1. Tabakalandırma	19
1.2.2. Regresyon Düzeltmesi	20
1.2.3. Propensity Skor	21
1.2.4. Eşleştirme Analizi	23
1.2.4.1. Ortalama Farklar (OF) Standardize Edilmiş Ortalamalar Arasındaki Farklar (SOF) ve Varyans Oranları (VO)	24
1.2.4.2. Kolmogorov Smirnov (KS) Test İstatistiği	25
1.2.4.3. Ortalama Tedavi Etkisi (ATE) ve Tedavi Grubunda Ortalama Tedavi Etkisi (ATT)	26
2.GEREÇ ve YÖNTEM	32
2.1. Benzetim Çalışması	32
2.2. Gerçek Veri Seti İle Uygulama	40
3.BULGULAR	41
3.1. Lojistik Regresyon ile En Uyumlu Algoritmaların Belirlenmesi	41
3.2. n= 500 Örneklem Büyüklüğü için Benzetim Sonuçları	42
3.3. n= 1000 Örneklem Büyüklüğü için Benzetim Sonuçları	55
3.4. Lalonde Gerçek Veri Setine Ait Sonuçlar	67
4. TARTIŞMA	71
5. SONUÇ ve ÖNERİLER	81
ÖZET	83
SUMMARY	85
KAYNAKLAR	87

EKLER	92
Ek-1. Benzetim Verisi Üretimi için Örnek R Kodları (Senaryo A, n=1000)	92
Ek-2. Denge Sonuçları (Senaryo A, n=500)	94
Ek-3. Denge Sonuçları (Senaryo B, n=500)	98
Ek-4. Denge Sonuçları (Senaryo C, n=500)	102
Ek-5. Denge Sonuçları (Senaryo A, n=1000)	106
Ek-6. Denge Sonuçları (Senaryo B, n=1000)	110
Ek-7. Denge Sonuçları (Senaryo C, n=1000)	114
ÖZGEÇMİŞ	118



ÖNSÖZ

Bu çalışmada deney grupları ile ortak değişkenler arasında ana etkiler dışında etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu eşleştirme analizi modellerinde propensity skor tahmini için klasik yöntem olan lojistik regresyona alternatif olarak makine öğrenme yöntemleri (genelleştirilmiş boosted modeller, sınıflandırma ve regresyon ağaçları, yapay sinir ağları vb.) ve bazı yaygın kullanılan yöntemlerin (bayesci genelleştirilmiş doğrusal modeller, kuadratik diskriminant analizi, adımsal akaike bilgi kriteri vb.) başarısını benzetim çalışması ile değerlendirmek suretiyle yeni bir kombine tahminci model geliştirmek amaçlanmıştır.

Doktora eğitimim ve tez çalışması süresince ilgisini, bilgisini ve desteğini esirgemeyen değerli danışman hocam Doç. Dr. Serdal Kenan KÖSE'ye,

Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Öğretim Üyeleri Prof.Dr. Atilla Halil ELHAN, Prof.Dr. Yasemin YAVUZ, Doç.Dr. Derya GÖKMEN ve Doç.Dr. Beyza Doğanay ERDOĞAN'a, Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Öğretim Üyesi Prof.Dr. Erdem KARABULUT ve Başkent Üniversitesi Tıp Fakültesi Öğretim Üyesi Prof.Dr. Mehtap Akçıl OK'a, Sabrı ve yardımları için başta Araş.Gör. Serhat HAYME olmak üzere tüm Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı çalışma arkadaşlarıma,

Her zaman beni destekleyen, çalışmalarım boyunca sabrını ve sevgisini esirgemeyen sevgili eşim Sevil DEMİR'e ve bu uzun ve meşakkatli yolda desteklerini hiçbir zaman esirgemeyen çok değerli annem Firdevs DEMİR ve babam İsmail DEMİR'e sonsuz teşekkürlerimi sunarım.

Doktora eğitimim sürecinde dünyaya gelerek beni motive eden ve en büyük şansım olan biricik kızım Defne DEMİR'e atfediyorum.

SİMGELER VE KISALTMALAR

α	Alfa
AI	Abadie Imbens
ATE	Ortalama tedavi etkisi
ATT	Tedavi grubunda ortalama tedavi etkisi
BAYESGLM	Bayesian generalized linear model - Bayesci genelleştirilmiş doğrusal model
β	Beta
CARET	Classification and regression tree - Sınıflama ve regresyon ağaçları
CARETRPART	Recursive partitioning/decision trees - Özyinelemeli parçalama
CFOREST	Conditional random forests - Koşullu rasgele ormanlar
DBARTS	Discrete bayesian additive regression trees - Kesikli bayesci eklemeli regresyon ağaçları
EARTH	Multivariate adaptive regression splines - Çok değişkenli uyarlanabilir regresyon eğrileri
ε	Epsilon - Hata terimi
GBM	Gradient boosting - Boosted model
GAM	Generalized additive model - Genelleştirilmiş toplamsal model
GLMNET	Generalized linear model with penalized maximum likelihood - Cezalandırılmış, genelleştirilmiş lineer modeller
IPREDBAGG	Improved predictive models bagging classification, regression and survival trees - Geliştirilmiş tahmin modelleri
KERNELKNN	Kernel k-nearest neighbors - Çekirdek k en yakın komşu algoritması

KNN	k nearest neighbors - k en yakın komşuluk
KS	Kolmogorov Smirnov
KSVM	Kernel support vector machines - Çekirdek destek vektör makineleri
LR	Lojistik regresyon
ML	Machine learning - Makine öğrenme
NNET	Neural network - Yapay sinir ağları
PS	Propensity skor
QDA	Quadratic discriminant analysis - Karesel (kuadratik) diskriminant analizi
RANDOMFOREST	Random forest - Rasgele ormanlar
RPART	Recursive partitioning for classification, regression and survival trees - Özyinelemeli parçalama regresyon ağaçları
RPARTPRUNE	Pruning recursive partitioning regression trees – Budamalı özyinelemeli regresyon ağaçları
σ	Sigma - Standart sapma
SCI	Science citation index
SH	Standart hata
SKK	Sınıf içi korelasyon katsayısı
SOF	Standardize edilmiş ortalamalar arasındaki farklar
STEPAIC	Stepwise model selection by Akaike information criterion - Adımsal Akaike bilgi kriteri
SVM	Support vector machine - Destek vektör makineleri
VO	Varyans oranları
XGBOOST	Extreme gradient boosting methods - Ölçeklenebilir gradient boosting

ŞEKİLLER

Şekil 1.1. Yayınların Yıllara Göre Dağılımı	3
Şekil 1.2. Anahtar Kelime Analizi	6
Şekil 1.1.1.1. Kanıt Düzeylerine Göre Araştırma Yöntemleri	9
Şekil 1.1.2.1. Deneyssel ve Gözlemsel Çalışmalara Yönelik Akış Diyagramı	11
Şekil 1.1.2.2. Nedenselliği İspatlamak Üzere Farklı Araştırma Türlerinin Görelî Kabiliyeti	12
Şekil 1.1.2.1.1. Karıştırıcı Değişken Örneği; Yaş ve Sıvı Tüketimi	14
Şekil 1.1.2.1.2. Karıştırıcı Değişken	14
Şekil 1.1.2.1.3. Doğum Sırası (a) ve Anne Yaşına Göre (b) Down Sendromu Dağılımı	15
Şekil 1.1.2.1.4. Doğum Sırası ve Anne Yaşına Göre Down Sendromu Dağılımı	16
Şekil 1.2.1.1. Tabakalandırma için Örnek Tasarım	20
Şekil 1.2.4.3.1. Deneyssel Çalışmalar için Örnek Veri Seti	29
Şekil 1.2.4.3.2. Gözlemsel Çalışmalar için Tek Ortak Değişkenin Bulunduğu Örnek Veri Seti	29
Şekil 1.2.4.3.3. Gözlemsel Çalışmalar için Birden Fazla Ortak Değişkenin Bulunduğu Örnek Veri Seti	30
Şekil 2.1.1. Superlearner Çalışma Algoritmasını Gösteren Akış Diyagramı	33
Şekil 2.1.2. Eşleştirme Yöntemleri	35
Şekil 2.1.3. Benzetim Veri Seti	36
Şekil 3.2.1. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo A, n=500)	43
Şekil 3.2.2. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo A, n=500)	43
Şekil 3.2.3. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo A, n=500)	44
Şekil 3.2.4. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo A, n=500)	45

Şekil 3.2.5. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo A, n=500)	45
Şekil 3.2.6. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo B, n=500)	46
Şekil 3.2.7. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo B, n=500)	47
Şekil 3.2.8. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo B, n=500)	48
Şekil 3.2.9. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo B, n=500)	48
Şekil 3.2.10. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo B, n=500)	48
Şekil 3.2.11. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo C, n=500)	50
Şekil 3.2.12. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo C, n=500)	50
Şekil 3.2.13. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo C, n=500)	51
Şekil 3.2.14. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo C, n=500)	51
Şekil 3.2.15. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo C, n=500)	52
Şekil 3.3.1. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo A, n=1000)	55
Şekil 3.3.2. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo A, n=1000)	56
Şekil 3.3.3. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo A, n=1000)	57
Şekil 3.3.4. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo A, n=1000)	57
Şekil 3.3.5. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo A, n=1000)	57
Şekil 3.3.6. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo B, n=1000)	59

Şekil 3.3.7. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo B, n=1000)	59
Şekil 3.3.8. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo B, n=1000)	60
Şekil 3.3.9. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo B, n=1000)	60
Şekil 3.3.10. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo B, n=1000)	61
Şekil 3.3.11. Tahmincilere Göre ATT ve AI Standart Hata ile Güven Aralıkları (Senaryo C, n=1000)	62
Şekil 3.3.12. Tahmincilere Göre ATT ve Klasik Standart Hata ile Güven Aralıkları (Senaryo C, n=1000)	62
Şekil 3.3.13. Tahmincilere Göre SOF Denge Karşılaştırmaları (Senaryo C, n=1000)	63
Şekil 3.3.14. Tahmincilere Göre VO Denge Karşılaştırmaları (Senaryo C, n=1000)	64
Şekil 3.3.15. Tahmincilere Göre KS Denge Karşılaştırmaları (Senaryo C, n=1000)	64

ÇİZELGELER

Çizelge 1.1. PS Konusunda Literatürde En Çok Atıf Yapılan Makaleler	4
Çizelge 1.1.2.2.1. Etki Karıştırıcı Kavramı için Ham ve Tabakalara Göre Örnek Odds Oranları	18
Çizelge 1.2.1. Deneysel Hatayı Azaltma Yöntemleri	19
Çizelge 1.2.4.3.1. Potansiyel Çıktılar ve Karşı Olgu (Counter-factual)	26
Çizelge 2.1.1. Tahminci Algoritmalar ve Tanımlamaları	34
Çizelge 2.1.2. Benzetim Veri Üretiminde Kullanılan Korelasyon Matrisi	37
Çizelge 2.1.3. Benzetim Veri Üretiminde Kullanılan Katsayılar	38
Çizelge 2.1.4. A Senaryosu n=1000 Benzetim Veri Setine Ait Regresyon Analizi Sonuçları	39
Çizelge 3.1.1. Farklı Algoritmalarla Kestirilen PS'ler Arasındaki Korelasyon ve Uyum Değerleri	41
Çizelge 3.2.1. ATT için Denge Sonuçları (Senaryo A; n=500)	42
Çizelge 3.2.2. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo A; n=500)	44
Çizelge 3.2.3. ATT için Denge Sonuçları (Senaryo B; n=500)	46
Çizelge 3.2.4. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo B; n=500)	47
Çizelge 3.2.5. ATT için Denge Sonuçları (Senaryo C; n=500)	49
Çizelge 3.2.6. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo C; n=500)	51
Çizelge 3.2.7. A, B ve C Senaryolarına Göre Başarılı Olan Algoritmalar	52
Çizelge 3.2.8. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo A; n=500)	53
Çizelge 3.2.9. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo B; n=500)	53
Çizelge 3.2.10. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo C; n=500)	54

Çizelge 3.2.11. LR ve Kombine Tahminci için Özet Denge Sonuçları (Senaryo A, B, C; n=500)	54
Çizelge 3.3.1. ATT için Denge Sonuçları (Senaryo A; n=1000)	55
Çizelge 3.3.2. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo A; n=1000)	56
Çizelge 3.3.3. ATT için Denge Sonuçları (Senaryo B; n=1000)	58
Çizelge 3.3.4. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo B; n=1000)	60
Çizelge 3.3.5. ATT için Denge Sonuçları (Senaryo C; n=1000)	61
Çizelge 3.3.6. Ortak Değişken/Faktörlerin Ortalama Denge Sonuçları (Senaryo C; n=1000)	63
Çizelge 3.3.7. A, B ve C Senaryolarına Göre Başarılı Olan Algoritmalar	65
Çizelge 3.3.8. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo A; n=1000)	65
Çizelge 3.3.9. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo B; n=1000)	66
Çizelge 3.3.10. LR ve Kombine Tahminci için Denge Sonuçları (Senaryo C; n=1000)	66
Çizelge 3.3.11. LR ve Kombine Tahminci için Özet Denge Sonuçları (Senaryo A, B, C; n=1000)	67
Çizelge 3.4.1. Lalonde Veri Seti Eşleştirme Sonrası Denge Ölçümlerinin Karşılaştırılması (n=185)	67
Çizelge 3.4.2. Lalonde Veri Seti Eşleştirme Sonrası Ortalamalar	68
Çizelge 3.4.3. Lalonde Veri Seti Eşleştirme Sonrası OF Denge Ölçümlerinin Karşılaştırılması	68
Çizelge 3.4.4. Lalonde Veri Seti Eşleştirme Sonrası SOF Denge Ölçümlerinin Karşılaştırılması	69
Çizelge 3.4.5. Lalonde Veri Seti Eşleştirme Sonrası VO Denge Ölçümlerinin Karşılaştırılması	69
Çizelge 3.4.6. Lalonde Veri Seti Eşleştirme Sonrası KS Denge Ölçümlerinin Karşılaştırılması	70

1. GİRİŞ

Randomizasyon, hastaların deney gruplarına şansa bağlı olarak atanmasını sağlayarak seçim yanlılığının oluşmasını engellemek ve klinik denemelerde etkinliği araştırılan tedavi dışındaki tüm ortak değişkenler/faktörler için grupların benzer dağılım özelliklerine sahip olmasını sağlamak amacıyla yapılır. Araştırma sonucunda ulaşılan bulguların araştırılan saf tedavi etkisinin dışında başka bir etkene bağlı olmaması için gruplarda diğer etkenler benzer yönde olmalıdır (Kanık ve ark., 2011). İstatistiksel araştırmalarda nedensellik bağıntısı incelenirken, varsayılan olası nedenin doğru olarak belirlenebilmesi için ikincil ögelerin etkilerinin arıtılması gerekmektedir. Randomize kontrollü çalışmalarda sebep sonuç ilişkileri nedensel etkiler (causal effect) izole edilerek ortaya koyulduğu için güvenilirlik yüksek olmaktadır. Klinik çalışmalarda (gözlemsel veya yarı deneysel) ise vaka ve kontrol grupları arasında gözlenen ortak değişkenlere/faktörlere göre dengesizlik (unbalance) bulunabilmektedir. Bu dengesizlik tedavi etkisinin yanı sıra tahminlerine neden olmaktadır. Gruplarda dengesizlik olduğunda sonuç değişkeni üzerine kontrol değişkenlerinin/faktörlerin etkilerinin arıtıldığı ve tedavinin saf etkisinin belirlenebildiği modeller geliştirilmiştir. Propensity skor (PS) tahminine dayalı eşleştirme yöntemleri gözlemsel çalışmaların yanlılığını azaltmaktadır (Rubin, 1974; Rubin, 1997; Rosenbaum ve Rubin, 1983).

Birçok bilim alanında randomizasyonun yapılamadığı gözlemsel ve yarı deneysel araştırmalarda tedavinin veya bir girişimin saf etkisinin belirlenebilmesi, yansız tahminlere ulaşılabilmesi, karıştırıcı (confounder) değişkenlerin etkisinin giderilmesi ve deney sonunda ulaşılan bulguların araştırılan tedavi etkisinin dışında başka bir faktöre bağlı olmadığından kesin emin olunabilmesi için etkene maruz kalan ve kalmayan grupların birçok açıdan birbirine benzer olması gerekmektedir. Gözlemsel araştırmalarda bunu sağlamak için literatürde PS'ye dayalı eşleştirme analizi kullanılmaktadır.

Eşleştirmenin amacı kontrol grubu içerisinde tedavi grubundaki bireylerin sahip olduğu ortak değişken değerlerine benzer özellikteki bireyleri seçmektir. Ortak değişken sayısı fazla olduğunda PS çok sayıda ortak değişken bilgisinin özet ölçüsünü yansıttığı için geleneksel yöntemlere (tabakalara ayırma, kısıtlama ve kovaryans düzeltilmesi vb.) göre daha başarılı sonuçlar vermektedir. Ortak değişken sayısı arttığında geleneksel yöntemler ortak değişken düzeltilmesi için sınırlı sayıda ortak değişken kullanabileceği için bulgulara yanlışlık olma ihtimali artacaktır (Kaspar ve ark., 2010). PS’de ise böyle bir sınırlama yoktur. Bu nedenle randomizasyonun yapılamadığı klinik çalışmalarda PS’ye göre vaka ve kontrol gruplarının eşleştirilmesi yanlışlığı azaltmak ve karıştırıcı faktörlerin etkisini azaltarak tedavinin gerçek etkisini belirlemek için araştırmalarda dikkate alınması gereken ve gün geçtikçe önemi artan bir yöntemdir.

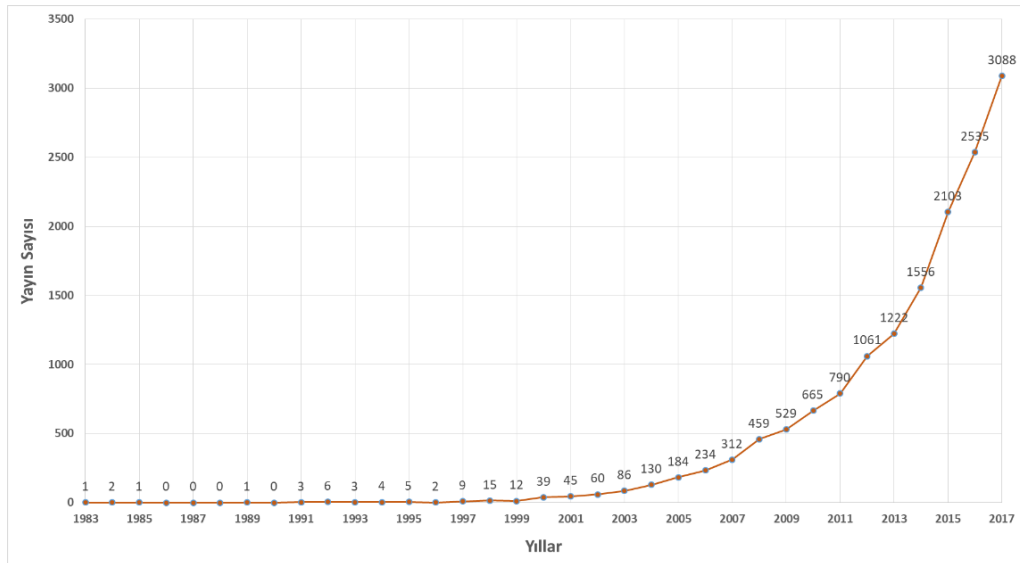
Thomson Reuters Web of Science indeksinde “propensity score” ve “propensity-score” anahtar kelimeleri ile “*Topic*” bölümünde (makalenin başlık, özet ve anahtar kelimelerini kapsamaktadır) 1975 ve 2017 yılları arasında yayınlanan yayınların taranması sonucunda “16828” yayın bulunmuştur (veri erişim tarihi 11.03.2018; *Topic: “propensity score” or Topic: “propensity-score” Timespan: 1975-2017*). Bu yayınların kategorilere göre sınıflanması; 15162 tanesi (%90,1) makale, 983’ü (%5,8) özet bildiri, 617’si (%3,7) tam metin bildiri, 226’sı (%1,3) derleme makale, 97’si (%0,6) editöre mektup, 22’si (%0,1) kitap bölümü ve 4’ü (%0,02) kitap derlemesi olarak raporlanmıştır.

15162 makale incelendiğinde PS’yi en çok kullanan ilk on ülke sırasıyla Amerika (8443), Kanada (1154), Almanya (1125), Japonya (1070), İngiltere (1032), Güney Kore (1000), Çin (953), İtalya (943), Tayvan (750) ve Fransa (715) olarak bulunmuştur. Bu konuda yayınlanan yayınların yarısından fazlasının (%50,2) sadece Amerika’ya ait olduğu belirlenmiştir. Türkiye ise 57 yayımla PS’yi en çok kullanan ülkeler arasında 33. sırada yer almaktaydı. Türkiye adresli bu yayınlardan sadece bir tanesi “İstatistik” (Statistics Probability) araştırma alanında yapılmıştı. Baycan (2016) tarafından “Journal of Applied Statistics” dergisinde yayımlanan bu çalışma ekonomi alanı ile ilgiliydi. Ankaralı (2009) tarafından yapılan PS’ye dayalı uygulama çalışması

ise Bilgisayar bilimleri ve mühendislik araştırma alanındaki “Expert Systems with Applications” dergisinde yayınlanmıştı. Sonuç olarak, araştırmamız sonucunda literatürde PS metodolojisi ile ilgili Web of Science indeksinde istatistik alanında SCI expanded kategorisindeki bir dergide yayınlanan Türkiye adresli herhangi bir benzetim çalışmasına rastlanmamıştır.

PS’nin en çok kullanıldığı ilk üç alanın sırasıyla 3081 (%18,3) yayınlarda Kardiyovasküler Sistem Kardiyolojisi (Cardiovascular System Cardiology), 2156 (%12,8) yayınlarda Cerrahi (Surgery) ve 1919 (%11,4) yayınlarda İşletme ekonomisi (Business Economics) olduğu görülmüştür. PS’nin yaygın olarak kullanıldığı diğer alanların ise sırasıyla Genel İç Hastalıkları (General Internal Medicine, 1547; %9,2), Halk Sağlığı (Public Environmental Occupational Health, 1319; %7,8), Onkoloji (Oncology, 1226; %7,3), Solunum Sistemleri (Respiratory System, 1054; %6,3), Sağlık Bakım Hizmetleri (Health Care Sciences Services, 1045; %6,2), İstatistik ve Matematik (Statistics and Mathematics, 741; %4,4) ve Farmakoloji (Pharmacology Pharmacy, 683; 4,1) olduğu görülmüştür.

Yayınların yıllara göre dağılımı incelendiğinde ise 15162 makalenin 7726 (%50,96)’sının son üç yılda (2015-2017) yayınlandığı görülmektedir (Şekil 1.1).



Şekil 1.1. Yayınların yıllara göre dağılımı

Şekil 1.1’den görüldüğü üzere PS sadece istatistik alanında değil birçok bilim alanındaki araştırmalarda özellikle son yıllarda popüler olarak kullanılan bir yöntemdir. Literatür incelendiğinde PS’ye dayalı eşleştirme analizinin birçok tıp alanında yaygın olarak kullanılan ve her geçen gün kullanımı artan önemli bir yöntem olduğu görülmektedir. Ancak dünyadaki kullanımına göre ülkemizde halen yeterince kullanılmayan bir yöntem olması da tez konusunun önemini göstermektedir.

İkinci literatür taramasında, Web of Science indeksinde “propensity score” ve “propensity-score” anahtar kelimeleri ile “Title” bölümünde (sadece makale başlıklarını kapsamaktadır) 1975 ve 2017 yılları arasında yayınlanan yayınlar bibliyometrik olarak analiz edilmiştir (veri erişim tarihi 11.03.2018; Title: “*propensity score*” or Title: “*propensity-score*”). Bibliyometrik ağ görselleştirmeleri VOSviewer (Version 1.6.6) paket programı kullanılarak oluşturulmuştur.

Arama sonucunda 3290 yayın elde edilmiştir. Bu yayınların 2093 (%63,6) tanesinin makale kategorisinde yer aldığı belirlenmiştir. Yıllık ortalama alınan atıf sayısına göre en çok atıf alan ilk yirmi makale Çizelge 1.1’de gösterilmiştir.

Çizelge 1.1. PS konusunda literatürde en çok atıf yapılan makaleler

Makale	Yazarlar	Dergi	YY	TA	OA
The central role of the propensity score in observational studies for causal effects	Rosenbaum, PR; Rubin, DB	Biometrika	1983	8537	237,1
An introduction to propensity score methods for reducing the effects of confounding in observational studies	Austin, Peter C.	Multivariate Behavioral Research	2011	1510	188,8
Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group	D'Agostino, RB.	Statistics in Medicine	1998	2986	142,2
Some practical guidance for the implementation of propensity score matching	Caliendo, M; Kopeinig, S.	Journal of Economic Surveys	2008	1037	94,3
Propensity score-matching methods for nonexperimental causal studies	Dehejia, RH; Wahba, S.	Review of Economics and Statistics	2002	1303	76,7
Transcatheter aortic valve replacement versus surgical valve replacement in intermediate-risk patients: a propensity score analysis	Thourani, Vinod H; Kodali, S; Makkar, Raj R. et al.	Lancet	2016	215	71,7
Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples	Austin, Peter C.	Statistics in Medicine	2009	714	71,4

Çizelge 1.1. PS konusunda literatürde en çok atıf yapılan makaleler (devamı)					
Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies	Austin, Peter C.	Pharmaceutical Statistics	2011	521	65,1
Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score	Rosenbaum, PR; Rubin, DB.	American Statistician	1985	1836	54,0
Reducing bias in observational studies using subclassification on the propensity score	Rosenbaum, PR; Rubin, DB.	Journal of the American Statistical Association	1984	1822	52,1
Multivariate and propensity score matching software with automated balance optimization: the matching package for R	Sekhon, Jasjeet S.	Journal of Statistical Software	2011	374	46,8
A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003	Austin, Peter C.	Statistics in Medicine	2008	505	45,9
Variable selection for propensity score models	Brookhart, MA; Schneeweiss, S; Rothman, KJ. et al.	American Journal of Epidemiology	2006	583	44,9
A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a monte carlo study	Austin, Peter C; Grootendorst, P; Anderson, Geoffrey M.	Statistics in Medicine	2007	462	38,5
Analysis of observational studies in the presence of treatment selection bias - effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods	Stukel, Therese A; Fisher, Elliott S; Wennberg, David E. et al.	Jama-Journal of The American Medical Association	2007	432	36,0
The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments	Austin, Peter C.	Statistics in Medicine	2014	180	36,0
High-dimensional propensity score adjustment in studies of treatment effects using health care claims data	Schneeweiss, S; Rassen, Jeremy A; Glynn, Robert J. et al.	Epidemiology	2009	351	35,1
Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies	Austin, Peter C; Stuart, Elizabeth A.	Statistics in Medicine	2015	135	33,8
Efficient estimation of average treatment effects using the estimated propensity score	Hirano, K; Imbens, GW; Ridder, G.	Econometrica	2003	526	32,9
Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study	Lunceford, JK; Davidian, M.	Statistics in Medicine	2004	481	32,1

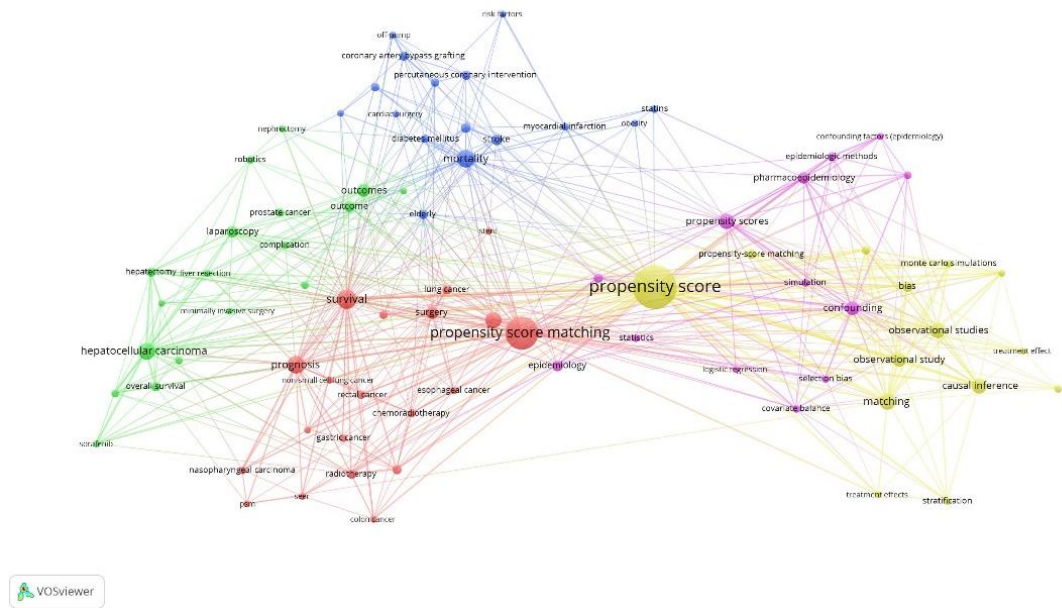
YY: Yayın yılı, TA: Toplam atıf sayısı, OA: Yıllık ortalama atıf sayısı

Çizelge 1.1 incelendiğinde toplamda ve yıllık ortalamada en çok atıf alan çalışmanın Rosenbaum ve Rubin (1983), tarafından Biometrika’da yayımlanan “The central role of the propensity score in observational studies for causal effects” başlıklı makale olduğu görülmüştür. Bu çalışma literatürde PS’nin ilk kez önerildiği

çalışmadır. Bu çalışmadan sonra PS'nin metodolojisi ve uygulanması ile ilgili istatistik literatüründe Rosenbaum ve Rubin (1984; 1985), D'Agostino (1998), Imbens (2000), Dehejia ve Wahba (2002), McCaffrey ve ark. (2004), Austin (2009; 2011), Caliendo ve Kopeinig (2008), Setoguchi ve ark. (2008), Lee ve ark. (2010), Westreich ve ark. (2010), Sekhon (2011), Pirracchio ve ark. (2014), Colson ve ark. (2016) tarafından birçok önemli çalışma yapılmıştır. Yerli literatürde ise Kaspar ve ark. (2010) ve Demir ve ark. (2017) tarafından yayınlanan kısıtlı sayıda makale ve Kaspar ve ark. (2014), Özdemir ve ark. (2014), Demir ve ark. (2014), Demir ve Köse (2016; 2017), Demir ve ark. (2017), Demir ve ark. (2017) tarafından hazırlanan kongre tebliğleri bulunmuştur.

Ekonomi alanında da yaygın kullanılan PS, Dehejia ve Wahba (2002) ve Abadie ve Imbens (2006)'in çalışmalarıyla popüler olmuştur. 3. ve 4. olarak en fazla atıf alan yayınların ekonomi alanında olduğu Çizelge 1.1'de görülmektedir.

Yayınların bibliyometrik olarak analizi sonucunda, 2093 makale içerisinde toplam 4058 anahtar kelime kullanıldığı bulunmuştur. Bu anahtar kelimeler içerisinde en az 10 kez farklı makale içerisinde kullanılan 81 anahtar kelime bulunmaktadır. Bu anahtar kelimeler ile yapılan bibliyometrik görsel ağ haritası Şekil 1.2'de verilmiştir.



Şekil 1.2. Anahtar kelime analizi

Şekil 1.2 incelendiğinde 5 farklı kümelenme olduğu görülmüştür (renkler kümelenmeyi, daire büyüklüğü kullanılma sayısının fazlalığını ve çizgiler birlikteliği göstermektedir). Kümelerden iki tanesi PS'nin doğrudan metodolojisi ile diğerleri ise daha çok tıp araştırmalarındaki kullanım başlıklarında kümelenmiştir. 4. küme (sarı renk) sırasıyla balance, bias, causal inference, generalized propensity score, inverse probability of treatment weighting, matching, monte carlo simulations, observational studies, observational study, propensity score, propensity-score matching, stratification ve treatment effect(s) kelimelerinden oluşmuştur. 5. küme ise (pembe) sırasıyla comparative effectiveness research, confounding, confounding factors (epidemiology), covariate balance, epidemiologic methods, epidemiology, logistic regression, pharmacoepidemiology, propensity scores, selection bias, simulation, statistics ve survival analysis kelimelerinden oluşmuştur.

Her iki kümede oluşan anahtar kelimeler şimdiye kadar PS konusunda yayınlanan makalelerde en sık kullanılan (en az 10 kez) anahtar kelimelerdir ve PS konusunun anlaşılmasında önemli kavramlardır. Bu bölümden sonraki kısımda konu ile ilgili yöntemler ve bibliyometrik analizlerin yardımı ile belirlenen en önemli kavramlar tanımlanacaktır.

1.1. Nedensel Çıkarım

Hernan (2004)'in nedensel çıkarımı ele aldığı makalesindeki örnekte; kalp nakli için bekleyen bir hasta olan Zeus'un 1 Ocak'ta gerçekleştirilen kalp naklinden beş gün sonra öldüğü ifade edilmiştir. Bu durumda Zeus'a 1 Ocak'ta kalp nakli yapılmıyaydı Zeus'un beş gün sonra yaşayabileceği düşünülecekti (hayatındaki diğer her şeyin normal seyrettiği düşünüldüğünde). Bu bilgiye sahip insanlar, nâkilin Zeus'un ölümüne neden olduğunu kabul edecektir. Sonuç olarak müdahalenin Zeus'un beş günlük sağ kalımı üzerinde nedensel bir etkisi olduğu anlaşılmaktadır. Diğer tüm şartların eşit olduğu durumlarda A eyleminin olduğu ve olmadığı iki durum arasında karşılaştırma yapıldığında eğer iki sonuç farklıysa, A eyleminin sonuç üzerinde nedensel veya önleyici bir nedensel etkiye sahip olduğu söylenecektir. Aksi takdirde,

A eyleminin sonuç üzerinde nedensel bir etkisi olmadığı söylenir. Epidemiyolojide, A genellikle maruziyet (exposure) veya tedavi olarak adlandırılmaktadır ve

$$\Pr[Y_{a=1} = 1] \neq \Pr[Y_{a=0} = 1]$$

ise tedavi etkisi popülasyonda nedensel bir etkiye sahip olacaktır. Diğer bir ifadeyle nedensel etki aşağıdaki gibi tanımlanacaktır (Hernan, 2004).

$$\Pr[Y_{a=1} = 1] - \Pr[Y_{a=0} = 1] \neq 0$$

Bilimsel araştırmalar, veri toplama yöntemlerine, nedensellik ilişkilerine, zaman ilişkilerine ve uygulandığı ortama göre aşağıdaki gibi sınıflandırılabilir (Çaparlar ve Dönmez, 2016).

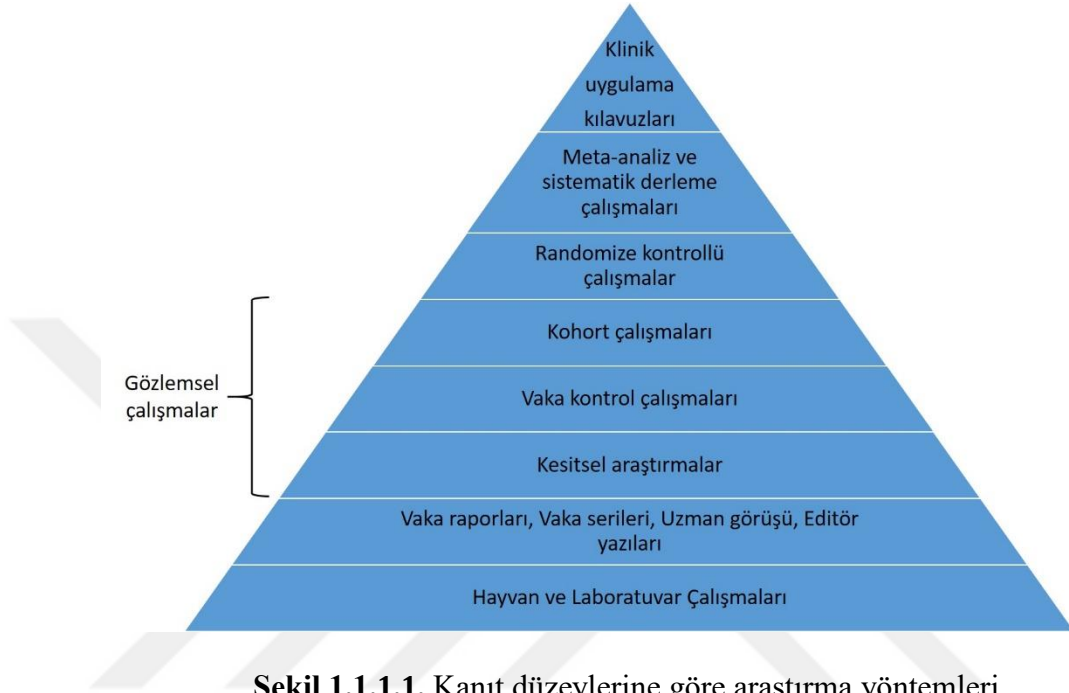
1. Veri toplama yöntemlerine göre; deneysel ve gözlemsel
2. Nedensellik ilişkilerine göre; tanımlayıcı ve çözümleyici
3. Zamana göre; retrospektif, prospektif ve kesitsel
4. Uygulandığı ortama göre; klinik, laboratuvar ve toplum tanımlayıcı araştırmalar

Tez çalışmasının devamında veri toplama tekniklerine göre sınıflandırılan bilimsel araştırmalardan deneysel ve gözlemsel çalışmalar açıklanacaktır.

1.1.1. Deneysel Çalışmalar

Deneysel araştırmalar belli bir girişimin kontrol altına alınmış koşullarda belli bir hastalığın veya problemin çözümündeki etkisini araştırmak için yapılır. Araştırmacı çalışmalarını kontrol edilebilir ortamda gerçekleştirerek deneyi etkileyecek dış faktörleri kontrol altında tutar. Deneysel araştırma yöntemi araştırmacılara bir olaydaki etkenin sonuca etkilerini gözleme ve sebep-sonuç ilişkisini test etme olanağı sağlar. Bu yüzden diğer araştırma yöntemlerine göre kanıt düzeyi yüksektir. Deneysel çalışmalarda incelenen faktör araştırmacının kontrolündedir.

Deney tasarımımda diğer tüm değişkenler sabit tutulduğu için sonuçlarda yanlılık olma olasılığı düşük olacaktır. Deneysel çalışmalar benzer şartlar oluşturularak tekrarlanma özelliğine sahiptir. Şekil 1.1.1.1’de tıp araştırmaları kanıt düzeylerine göre sıralanmıştır (Yılmaz ve Çöl, 2014).



1.1.1.1. Yanlılık

Yanlılık, klinik denemelerde önyargılı girişimde bulunma, yan tutma veya değerlendirmeleri etik kurallara uygun olmayan bir şekilde gerçekleştirme durumudur. Çalışmaların sonucunda ulaşılan tedavi etkinliğinin olduğundan farklı olarak yorumlanmasına yol açabilecek sistematik hatalara yanlılık (bias) denir. Literatürde pek çok yanlılık kaynağı tanımlanmıştır. Seçim Yanlılığı (Selection Bias), Değerlendirme Yanlılığı (Evaluation Bias), Yayın Yanlılığı (Report Bias) ve Bilgi Yanlılığı (Information Bias) bunlardan bazılarıdır (Hoffmeister ve ark., 1998; Hill ve Kleinbaum, 2005).

Yanlılık, gözlemsel çalışmalarda önemli bir konudur. Çalışmanın tasarımı ve yürütülmesi sırasında yanlılığın dikkate alınması önemlidir çünkü daha sonra

düzeltilmemektedir. Yanlılık; örnekleme bulunan, popülasyondaki değişkenlikle ilgili olan ve veri sayısını arttırarak azaltılabilecek rastgele bir hatadan farklı olarak sistematik bir hatadır. Tez konusuyla ilgili olan ve klinik denemelerde en çok karşılaşılan seçim yanlılığı, araştırmacının araştırmaya alınacak hastaları kendi tercihlerine göre seçmesi diğer bir deyişle seçilecek hastaları deney tasarımındaki tedavi ve kontrol grubuna kendi belirlediği şekilde ataması ile oluşur. Dolayısıyla gruplar arasında ortak değişkenler açısından farklılık bulunmaktadır ve araştırma sonucunu bu farklı etkiler karıştırmaktadır. Seçim yanlılığını önlemenin yolu ise hastaları randomizasyon yöntemlerini kullanarak tedavi gruplarına rastgele atamaktır. Kısacası araştırmalarda seçilen örneklemin popülasyonu yansıtması istenmektedir. Gözlemsel çalışmalarda ise çoğunlukla bu mümkün olmadığı için araştırmalarda araştırılan etki yanlı olarak elde edilebilmektedir (Pandis, 2014).

1.1.1.2. Randomizasyon

Klinik denemeye alınan bireylerin tedavi gruplarına dağıtımlarının rastgele yapılmasıdır. Deneysel çalışmalarda randomizasyon yöntemi ilk olarak 1923 yılında R.A. Fisher tarafından kullanılmıştır. Randomizasyonun amacı, hastaların tedavi gruplarına atanırken tamamen şansa bağlı olarak atanmasını sağlayarak araştırmacı tarafından yapılabilecek olası seçim yanlılığını önlemektir. Kısaca randomizasyon, klinik denemelerde grupların etkinliği araştırılacak olan tedavi dışındaki öngörülen ve öngörülemeyen tüm ortak değişken/faktörler açısından benzer dağılım özelliklerine sahip olmasını sağlamak için yapılır (Gökmen ve ark., 2018, s:44). Bu şekilde bireyler araştırmaya dahil olma kriterleri açısından birbirine benzer olacak şekilde deney gruplarına dengeli olarak atanabilirler (Kanık ve ark., 2011). Deney sonunda tedavi grupları arasında bulunabilecek olası bir farkın, bireyleri tedavi gruplarına atamadaki yanlılığa bağlı olma olasılığı ortadan kaldırılmış olur (Rosenberger ve Lachin, 2002). Klinik denemelerde sabit olasılıklı (basit, blok, tabakalı vb.) ve değişken olasılıklı (kazanana oyna kuralı, minimizasyon vb.) birçok randomizasyon yöntemi kullanılmaktadır.

Randomizasyon her zaman yanlılığı gidermeyebilir. Özellikle veri sayısı az olan küçük örneklemelerde bazı ortak değişken (covariate) veya karıştırıcı (confounder) etkisini gidermek için kullanılan tabakalı (stratified) randomizasyonda bazı gruplara az sayıda gözlem düşebilir ve bu yüzden bazı demografik faktörlere veya ortak değişkenlere göre yanlılık giderilemeyebilir.

1.1.2. Gözlemsel Çalışmalar

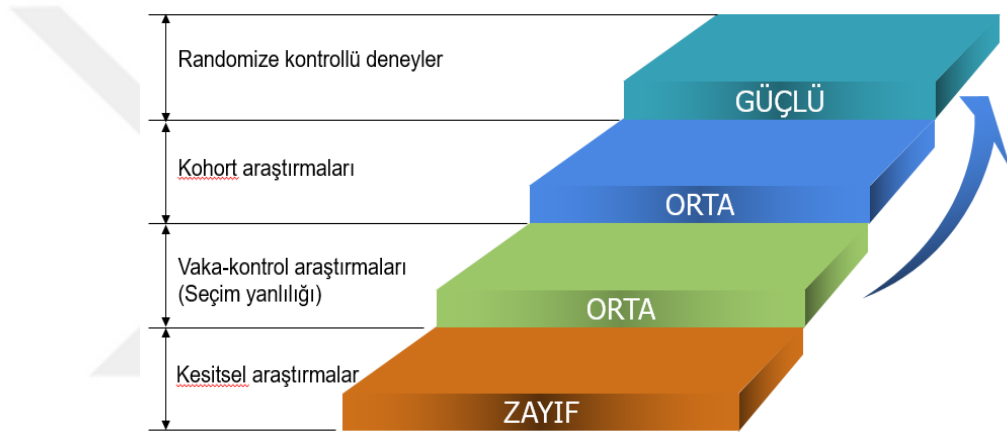
Gözlemsel araştırmalar klinik verilerin sistemde mevcut olmasından dolayı daha ucuz maliyet ile hızlı sonuca varılması nedeniyle diğer çalışmalara göre daha avantajlıdır (Çaparlar ve Dönmez, 2016). Gözlemsel çalışmalarda incelenen faktör araştırmacının kontrolünde değildir (Rosenbaum, 2002). Çünkü diğer değişkenler kontrol altında tutulamaz. Dolayısıyla neden-sonuç ilişkisinde yanlılık bulunma olasılığı yüksek olduğu için sonuçlar kesin değildir ve bu deney tasarımı benzer şartlar oluşturularak tekrarlanamaz. Analitik gözlemsel araştırmalar; kesitsel çalışmalar, vaka-kontrol çalışmaları ve kohort çalışmaları olarak 3 başlık altında sınıflandırılabilir (Gökmen ve ark., 2018, s:3).

Şekil 1.1.2.1’de deneysel ve gözlemsel çalışmaların ayrımını gösteren bir akış diyagramı sunulmuştur.



Şekil 1.1.2.1. Deneysel ve gözlemsel çalışmalara yönelik akış diyagramı

Randomize kontrollü çalışmalar ile elde edilen sebep sonuç ilişkileri karıştırıcı değişkenlerin etkilerinden dolayı gözlemsel klinik çalışmalara göre daha güvenilir olmaktadır. Gözlemsel çalışmalarda çoğunlukla deney grupları arasında ortak değişkenlere/faktörlere göre farklılıklar bulunmaktadır. Gruplar arasındaki bu farklılıkları gidermek amacıyla PS tahminine dayalı eşleştirme analizi yaygın kullanılan yöntemlerden biridir. Literatürde geriye dönük (retrospektif) olarak elde edilen verilere dayalı gözlemsel vaka-kontrol çalışmalarında eşleştirme yapılarak sonuçların verilmesi çalışmanın kanıt düzeyini artırmaktadır. Nedenselliği belirlemek üzere farklı araştırma türlerinin göreceli kabiliyeti üzerine eşleştirmenin etkisi Şekil 1.1.2.2’de gösterilmiştir.



Şekil 1.1.2.2. Nedenselliği ispatlamak üzere farklı araştırma türlerinin göreceli kabiliyeti

1.1.2.1. Karıştırıcı Faktörler (Confounding)

Karıştırıcı etkisi, bir maruziyet ile sonuç arasındaki etki veya ilişkinin başka bir değişkenin varlığı tarafından bozulduğu bir durum olarak tanımlanabilir. Araştırılan etken ve ortaya çıkan sonuç arasındaki ilişki kısmen veya tamamen üçüncü bir karıştırıcı değişkene bağlıdır.

Örneğin; kahve ile koroner kalp hastalığı arasındaki ilişkiyi araştırmak amacıyla bir çalışma planlansın. Sigara kullananlar iki grup arasında eşit olmayan bir şekilde

dağılırsa bu ilişki sigara kullanımına bağlı olarak bozulabilir. Kahve ile koroner kalp hastalığı arasında bir ilişki var gibi görülebilir ancak her grup için sigara kullanan ve sigara kullanmayanları ayrı ayrı ele almanız durumunda aslında hiçbir ilişki olmadığını bulabilirsiniz.

Bir ortak değişkenin “karıştırıcı” olarak adlandırılabilmesi için aşağıda listelenen üç şartın tümünü sağlaması gerekir.

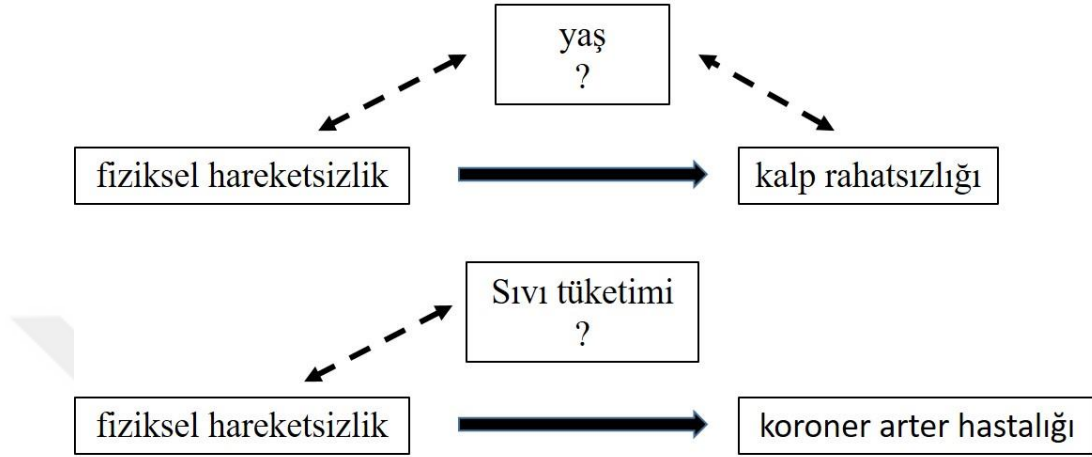
1. Karıştırıcı, varsayılan risk faktöründen bağımsız olarak hastalık için bir risk faktörüdür.
2. Varsayılan risk faktörü ile ilişkilidir.
3. Maruziyet ve hastalık arasındaki nedensel yolda değildir.

Bu koşulların ilk ikisi verilerle test edilebilir ancak üçüncüsü daha biyolojik ve kavramsaldır. Nedensel yolu daha iyi anlamak için makul alkol tüketimi ile koroner kalp hastalığı arasındaki ilişki incelenebilir. Makul alkol tüketiminin kanda HDL’yi yükseltmesi nedeniyle koroner kalp hastalığı riskini azalttığı bilinmektedir. Bu gibi bir durumda HDL seviyesi alkol ve kalp hastalığı arasındaki ilişkiyi karıştırmamaktadır çünkü HDL artışı alkol tüketiminin bir sonucu ve alkol kalp hastalığı riskini düşüren mekanizmanın bir parçasıdır (Lamorte ve Sullivan, 2018).

Karıştırıcı değişken kavramını daha iyi anlamak için aşağıdaki örneği inceleyelim. Bir araştırmacının, fiziksel aktivite ve kalp rahatsızlığı arasındaki ilişkinin gücünü tespit etmek istediğini farz edelim. Bu araştırmada yaş, karıştırıcı bir faktör olarak değerlendirilmelidir. Çünkü yaş;

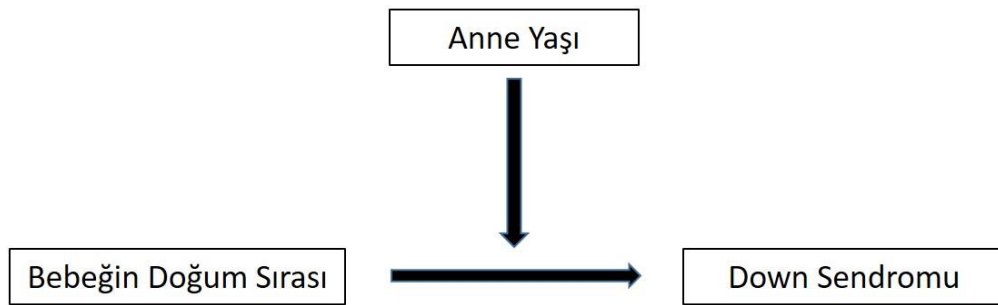
1. Maruziyet (fiziksel aktivite) ile ilişkilidir (yaşlı insanlar daha az aktiftir)
2. Sonuç değişkeni ile ilişkilidir (yaşlı insanların kalp hastalığına yakalanma riski daha fazladır).
3. Nedensel yolda olmadığı bilinmektedir.

Tam tersi bir örnek ise fiziksel aktivite ve kalp rahatsızlığı arasındaki ilişkide sıvı tüketimi fiziksel aktivite ile ilişkili olsa da koroner arter hastalığı için bir risk faktörü olmadığı için karıştırıcı faktör değildir. Karıştırıcı değişken kavramı için verilen örnekler Şekil 1.1.2.1.1’de gösterilmiştir (Lamorte ve Sullivan, 2018).



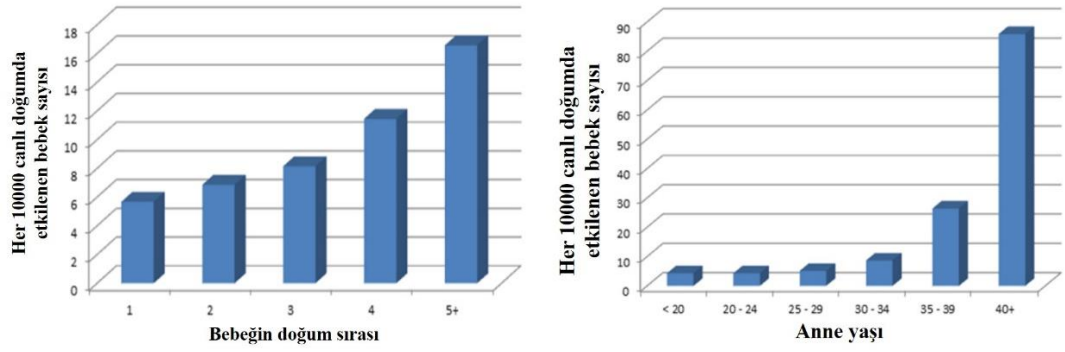
Şekil 1.1.2.1.1. Karıştırıcı değişken örneği; yaş ve sıvı tüketimi

Karıştırıcı değişken kavramının daha iyi anlaşılabilmesi için literatürden Stark ve Mantel (1966)’in çocukların doğum sırası ile Down sendromu arasındaki ilişkiyi araştırdıkları çalışmada örnek olarak verilebilir (Şekil 1.1.2.1.2).



Şekil 1.1.2.1.2. Karıştırıcı değişken

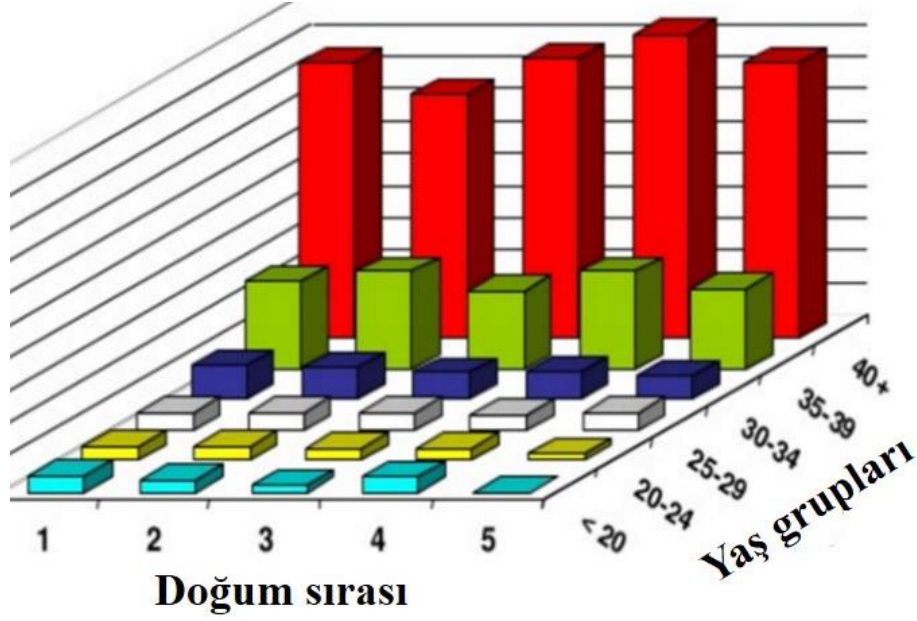
Makale bulgularına göre paylaşılan Şekil 1.1.2.1.3 (a) ve 1.1.2.1.3 (b) ayrı ayrı değerlendirildiğinde doğum sırasının ve anne yaşının Down sendromu ile ilişkili olduğu düşünülecektir.



Şekil 1.1.2.1.3. Doğum sırası (a) ve anne yaşına göre (b) Down sendromu dağılımı
(Lamorte ve Sullivan, 2018)

Çünkü çocukların doğum sırasına göre Down sendromu oranının arttığı, benzer şekilde anne yaşı arttıkça Down sendromu oranının arttığı grafiklerden görülmektedir. Burada karıştırıcı değişkeni belirlemek için sorulması gereken sorular şunlardır.

1. Doğum sırasının Down sendromunun görülmesi olasılığı üzerinde gerçekten bağımsız bir etkisi var mı? (yani doğum sırasının anne olma yaşıyla bağlantılı olmasından bağımsız bir etki)
2. Bir kadının çocuklarının doğuş sırası kadının doğum sırasındaki yaşıyla ilişkili midir? (beşinci çocuğunu doğuran kadınlar ortalama olarak ilk çocuklarını doğuran kadınlardan daha yaşlıdır). Anne yaşı doğum sırasındaki etkiyi karıştırabilir mi?
3. Çocuk doğduğunda çocukların doğum sırası anne yaşıyla karıştırılmış olabilir mi? Doğum sırası ile Down sendromu prevalansı arasındaki ilişki anne yaşının karıştırıcı etkisiyle abartılıyor mu?



Şekil 1.1.2.1.4. Doğum sırası ve anne yaşına göre Down sendromu dağılımı
(Lamorte ve Sullivan, 2018)

Şekil 1.1.2.1.3 yerine şekil 1.1.2.1.4 değerlendirildiğinde anne yaş gruplarına göre artan doğum sırasının Down sendromu prevalansı ile ilişkili olmadığı açıktır. Bir başka deyişle, "anne yaşı" kontrol edilirse, doğum sırasının herhangi bir etkisi olduğu konusunda hiçbir kanıt bulunmamaktadır (Şekilde doğum sırası gruplarına göre bakıldığında Down sendromu prevalansının tüm yaş gruplarında benzer olduğu görülecektir). Öte yandan prevalans değişiklikleri önden arkaya doğru incelenirse doğum sırası gruplarının hepsinde anne yaşı (maternal yaş) arttıkça prevalansda belirgin bir artış olduğu görülecektir. Başka bir deyişle, doğum sırasını kontrol ettikten sonrada yaş ile Down sendromunun ilişkili olduğu görülecektir.

Sonuç olarak, doğum sırası ile Down sendromu arasındaki belirgin ilişki tamamen yaşın karmaşık etkisinden kaynaklanmıştır. Öte yandan anne yaşı ile Down sendromu arasındaki ilişki doğum sırası ile karıştırılmamıştır, çünkü doğum sırasının Down sendromunun yaygınlığı üzerinde herhangi bir etkisi yoktur ve doğum sırasındaki farklılıklar nedeniyle yaş ile Down sendromu arasındaki ilişki bozulmamıştır (Lamorte ve Sullivan, 2018). Özetle;

Anne yaşı;

- Risk faktörüdür
- Doğum sırası ile ilişkilidir.
- Nedensel yolda değildir.

Doğum sırası;

- Risk faktörü değildir
- Anne yaşı ile ilişkilidir
- Nedensel yolda değildir.

1.1.2.2. Etki Değişimi (Effect Modification)

Karıştırıcı değişkenden farklı olarak etki değişimi/modifikasyonu sadece sonuç ile ilişkilidir, tedavi veya maruziyet ile ilişkilendirilemez. Bir risk faktörünün hastalık durumu üzerine gözlemlenen etkisini farklı gruplarda değiştiren bir değişkendir. Etki değişikliği olduğunda farklı grupların birbirinden farklı risk tahminleri olacaktır. Diğer bir ifadeyle iki değişkenden birisinin etkisi diğer değişkenin kategorilerine bağlı olarak değişiyorsa bu iki bağımsız değişken birbirleri ile etkileşim halindedir. Özetle, bir bağımsız değişkenin başka bir bağımsız değişkenin sonuç değişkeni üzerindeki etkisini değiştirmesi ile ortaya çıkan etki değişimi olarak da ifade edilir.

Karıştırıcı faktörler istenmeyen bir durumdur ve araştırma tasarımlarında sonuçların çarpıtılmasını önlemek için karıştırıcı faktörlerin etkilerinin ortadan kaldırılması gerekmektedir. Ancak etki değiştirme istenmeyen bir durum değildir, aslında araştırma amacı için önemli bilgiler sağlayacaktır. Etki modifikasyonunda bir tedavinin bir sonuç üzerindeki etkisinin büyüklüğü üçüncü bir faktörün düzeylerine göre değişmektedir. Özetlemek gerekirse karıştırıcı bulunduğunda gözlenen bir etki doğru değildir, farklı bir değişken hem potansiyel risk faktörü hem de sonuç ile ilişkilidir. Etki değişiminde ise gözlenen bir etki doğrudur ancak etkinin büyüklüğü farklı bir faktörün düzeylerinde farklıdır ve bu faktörün düzeylerinde ayrı olarak raporlanması yeterli olacaktır (örneğin, erkek ve kadınlarda farklı sonuçlar

raporlanacaktır). Etki değişikliğine sebep olan bir faktör bulunduğunda sonuçlar raporlanırken etkileşim olduğu belirtilerek tabakalarda ayrı ayrı tahminler verilmelidir.

Örneğin trafik kazası sonrası yaşlılık ile ölüm arasındaki ilişkinin araştırıldığı bir çalışmada emniyet kemeri takıp takmamanın etkiyi değiştirip değiştirmediği incelenmiştir. Burada emniyet kemeri kullanmanın etki karıştırıcı mı etki değiştirici mi olduğu sorusu kanıtlanmalıdır. Emniyeti kemeri kullanımı ile yaşlılık arasında doğrudan bir ilişki olmadığı düşünüldüğünde emniyet kemeri takıp takmama faktörünün karıştırıcı değişken olmadığı yorumu yapılabilir. Emniyet kemeri kullanımının sadece sonuç değişkeni ile ilişkili olabileceği düşünüldüğünden, yaş ile ölüm arasındaki ilişki emniyet kemeri takan ve takmayanlarda incelenmiş ve Çizelge 1.1.2.2.1'deki sonuçlar bulunmuştur.

Çizelge 1.1.2.2.1. Etki değişimi kavramı için ham ve tabakalara göre örnek odds oranları

	Yaş	Ölenler	Yaşayanlar	Toplam	Odds Oranı
Ham risk (crude risk)	≥70	13	61	74	6,75
	<70	25	935	960	
Emniyet kemeri takmayanlar	≥70	8	16	24	9,54
	<70	13	359	372	
Emniyet kemeri takanlar	≥70	5	45	50	4,90
	<70	12	576	588	

Çizelge incelendiğinde ilk sırada verilen sonuçlara göre 70 yaş ve üzerinde olmanın diğer gruba göre trafik kazasında ölüm riskini 6,75 kat arttırdığı görülmüştür. Ancak burada emniyet kemeri takıp takmamak etkiyi değiştireceği için (karıştırmayacak) tabakalarda ayrı olarak sonuçlar raporlanmalıdır. Güvenilir bir sonuç olarak; emniyet kemeri takanlarda bu riskin 4,90 kat fazla iken takmayanlarda 9,54 kat fazla olacağı bulguları raporlanmalıdır (Lamorte ve Sullivan, 2018).

1.2. Deneyisel Hatayı Azaltma Yöntemleri

Deneyisel hatayı azaltmak için istatistiksel analiz öncesi ve sonrasında uygulanabilecek yöntemler Çizelge 1.2.1'de verilmiştir.

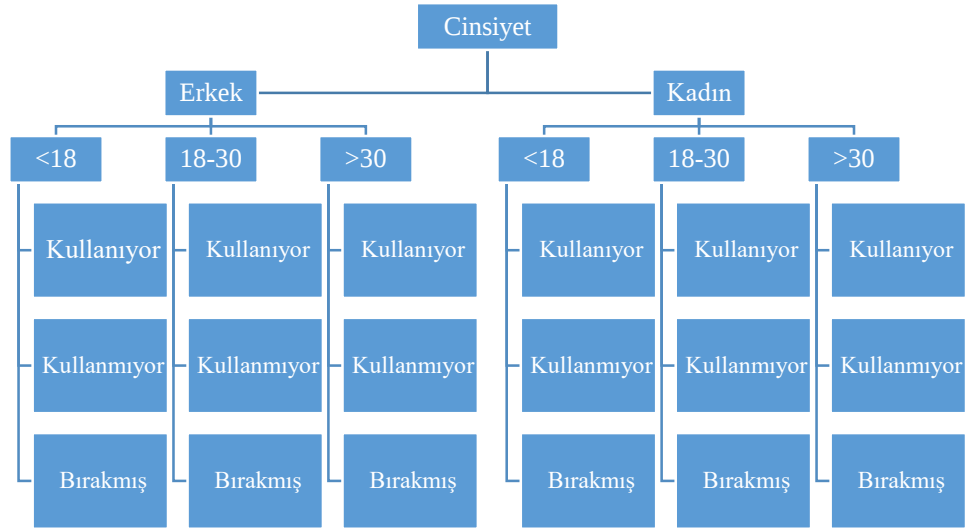
Çizelge 1.2.1. Deneysel hatayı azaltma yöntemleri

Analiz Öncesi		Analiz Evresi	
Randomizasyon	Gruplarda bilinmeyen karıştırıcıların benzer dağılımını sağlamak amacıyla	Tabakalandırma (OR, RR, Mantel-Haenszel yöntemi)	Örneğin kadın ve erkek veya sigara içme durumuna göre ayrı analiz yapılabilir (çok fazla karıştırıcı varsa zor olabilir).
Eşleştirme	Gruplarda potansiyel karıştırıcıların (Örneğin cinsiyet, yaş vb.) benzer dağılımını sağlamak amacıyla	İstatistiksel Modelleme	Örneğin ANCOVA, Regresyon düzeltmesi vb.
Kısıtlama	Örneğin kahvenin koroner kalp hastalığına etkisi araştırılıyorsa daha homojen bir grup için çalışma sigara kullanmayanlar ile sınırlandırılabilir. Ancak güç kaybı olabilir.		

1.2.1. Tabakalandırma

Tabakalandırma gözlemsel çalışmalarda tedavi ve kontrol grupları arasındaki sistematik farklılıkları gidermek için yaygın olarak kullanılan yöntemlerden biridir. Bu yöntemde gözlenen ortak değişkenlerin özelliklerine göre tabakalar oluşturularak aynı tabakalardaki tedavi ve kontrol gruplarında bulunan bireyler doğrudan karşılaştırılmaktadır. Diğer bir deyişle karıştırıcı faktörlerin etkisini gidermenin bir yolu birincil ilişkiyi potansiyel bir karıştırıcı faktörün farklı düzeylerinde incelemektir. Ancak bu yöntemin bir dezavantajı, ortak değişken sayısı arttıkça tabaka sayısının katlanarak artması ve ilişkinin incelenmesinin grup sayısının çokluğu ve buna bağlı olarak birey sayısının azlığı nedeniyle zorlaşmasıdır. Bu durumda PS'nin daha etkin bir yöntem olduğu ifade edilmiştir (D'agostino, 1998).

Örneğin Şekil 1.2.1.1.'deki araştırmada yaş (3 kategorili), cinsiyet (2 kategorili) ve sigara içme durumundan (3 kategorili) oluşan faktörlerin/ortak değişkenlerin analiz içerisinde tabakalı olarak değerlendirilmesi durumunda toplamda 18 farklı tabaka oluşacağı ve analizlerin birçok açıdan etkin olmayacağı söylenebilir.



Şekil 1.2.1.1. Tabakalandırma için örnek tasarım

1.2.2. Regresyon Düzeltmesi

Regresyon düzeltmesinde, t ve c sırasıyla tedavi ve kontrol gruplarını göstermek üzere; tedavi etkisi τ , aşağıdaki gibi tahmin edilir.

$$\hat{\tau} = (\bar{Y}_t - \bar{Y}_c) - \beta(\bar{X}_t - \bar{X}_c)$$

Regresyon modelinde ortak değişkenlerin etkisi denklemin sağ tarafındaki gibi çıkarılarak düzeltilir. Literatürde regresyon düzeltmesine ilişkin birçok farklı yöntem önerilmiştir. PS’de regresyon (kovaryans) düzeltmesinde kullanılabilmektedir (Rubin, 1979).

Genel olarak regresyon düzeltmesi dikkatli yapılmalıdır. Rubin, regresyon düzeltmesinin tedavi edilen ve tedavi edilmeyen gruplardaki kovaryans matrislerinin eşit olmaması halinde beklenen karesel yanlılığı artırabildiğini göstermiştir (Rubin, 1979). D’agostino (1998), tedavi edilen ve tedavi edilmeyen gruplardaki varyansın çok farklı olması durumunda regresyon düzeltmesi kullanmak yerine eşleştirme veya alt sınıflandırma için propensity skor yöntemlerini kullanmayı önermiştir.

Tez çalışmasının bundan sonraki bölümünde gözlemsel çalışmalarda analiz öncesi gruplarda karıştırıcı değişkenlerin benzer dağılması ve yanlılığı azaltmak için yaygın olarak kullanılan PS'ye dayalı eşleştirme analizinin metodolojisi sunulacaktır.

1.2.3. Propensity Skor

Randomize çalışmalarda gruplar ortak değişken veya faktörler açısından benzer olacağı için iki tedavi grubu araştırılan bir etki açısından doğrudan karşılaştırılabilecektir. Ancak randomize olmayan gözlemsel çalışmalarda farklı tedavi alan gruplar ortak değişken veya faktörler açısından sistematik olarak farklı olacağı için grupların bir etki açısından doğrudan karşılaştırılması yanıltıcı olacaktır (Rosenbaum ve Rubin, 1983).

Gözlemsel çalışmalarda araştırmacıların tedavi ataması üzerinde herhangi bir kontrolü yoktur. Tedavi ve kontrol grupları arasında ortak değişkenlere göre bulunabilecek büyük farklılıklar tedavi etkisinin yanlı olarak tahmin edilmesine sebep olabilir. Geleneksel kovaryans analizi ve diğer yöntemler (tabakalandırma, alt sınıflandırma, regresyon düzeltmesi veya bunların kombinasyonları) bu yanlılığı gidermek için kısıtlı sayıda ortak değişken kullanıldığı için yetersiz kalabilir. Ancak PS çok sayıda ortak değişkenin sayısal bir özet bilgisini sunduğu için bu kısıtlılığa sahip değildir (D'agostino, 1998). Rosenbaum ve Rubin (1983), PS'nin bir dengeleme skoru olduğunu ve gözlemsel çalışmalarda yanlılığı azaltmak için kullanılabileceğini göstermiştir (D'agostino, 1998).

D'agostino (1998), gözlemsel çalışmalarda hangi tedavinin verildiğini bildiğimiz halde neden bir bireyin belirli bir tedaviyi alma olasılığını tahmin etmeliyiz sorusuna, tedavi etkisine ilişkin tahminimizi düzeltmek için bir bireyin tedavi edilme olasılığı (propensity skoru) kullanıldığında yarı-randomize bir deney oluşturabileceği cevabını vermiştir. Diğer bir ifadeyle, tedavi ve kontrol grubunda aynı PS puanına sahip iki kişi bulunduğunda bu iki kişinin her bir gruba rastgele olarak eşit olasılıkla

atanmış olduğu düşünülebilir. Randomize kontrollü bir deneyde; bireylerin tedavi ve kontrol gruplarına atanmasını sağlayan randomizasyon, PS'ye dayalı eşleştirmeden elbetteki daha iyidir. Çünkü belirli bir değişkenler kümesine yönelik araştırmacının koşullarına bağlı değildir. Ayrıca randomizasyon gözlemlenen veya gözlemlenemeyen ortak değişkenler için geçerli olmasına rağmen PS kullanımının sonuçları sadece gözlemlenen değişkenlere bağlıdır. Bunun yanında bir kişi tedavi atamasıyla ilişkili olduğuna inanılan birçok ortak değişkeni belirleyip araştırmada dikkate alıyorsa o zaman tedavi etkisine yönelik neredeyse tarafsız tahminlerin elde edilebileceği söylenebilir (D'agostino, 1998).

Bağımlı değişken; tedavi değişkeni ve bağımsız değişken; ortak değişkenler olmak üzere Rosenbaum ve Rubin (1983), PS'yi, gözlenen ortak değişkenlere göre belirli bir tedavi grubuna atanmanın koşullu olasılığı olarak tanımlamıştır. $e(x)$; PS'yi, T_i ; tedavi ($T_i = 1$) ve kontrol ($T_i = 0$) grubunu göstermek üzere, ortak değişkenlere (x) bağlı olarak tedaviye atanma olasılığı eşitlik'deki gibi ifade edilmiştir.

$$e(x) = P(T = 1|x),$$

$$P(T_1, \dots, T_n | x_1, \dots, x_n) = \prod_{i=1}^N e(x_i)^{T_i} (1 - e(x_i))^{1-T_i}$$

Randomize çalışmalarda T_i rastgele bir mekanizmaya göre belirlenen bir dağılıma sahip olduğu için gözlemsel çalışmalardan farklılaşmaktadır. Çünkü randomize çalışmalarda PS, kabul edilen bir olasılığa dayalı bilinen bir fonksiyondur. Gözlemsel çalışmaların neredeyse tamamında ise PS fonksiyonu belirli bir olasılığa dayalı olmadığı için bilinmemektedir. Ancak bir lojit model kullanılarak gözlenen ortak değişkenlere bağlı olarak tahmin edilebileceği Rosenbaum ve Rubin (1983) tarafından gösterilmiştir. Bu çalışmadan sonra sağlık alanında yapılan çalışmalarda PS hesabında yaygın olarak lojit model (Lojistik Regresyon analizi) kullanılmıştır.

Randomize çalışmaların gözlemsel çalışmalardan ikinci farklılığı ise randomize çalışmalarda x 'in sonuç ile ilişkili olan ve tedavi gruplarına atamakta kullanılan tüm

ortak deęişkenleri içerdęi bilinmektedir. Gözlemsel çalıřmalarda ise çoęunlukla bilinmemektedir. Kısaca randomize deneylerde tüm bireyler tedavi grubuna atanmak için eřit olasılıęa sahiptir. Gözlemsel çalıřmalarda eřit olasılık olmadıęı için yanlılık bulunmaktadır. Karıřtırıcı deęişkenlerin etkisini arıtarak bu yanlılıęı gidermek için Rosenbaum ve Rubin (1983), üç farklı yöntem önermiřlerdir. Bunlar eşleřtirme, tabakalandırma (alt sınıflama) ve regresyon düzeltmesi' dir.

1.2.4. Eřleřtirme Analizi

Eřleřtirme, potansiyel olabilecek kontroller içeresinden tedavi grubunun ortak deęişken daęılımına en benzer olan grubu belirlemek için kullanılan bir örnekleme yöntemidir (D'agostino, 1998). PS' ye dayalı eşleřtirme analizinde kontrol grubunda propensity skoru P_j olan j'inci gözlem tedavi grubunda propensity skoru P_i olan bir gözlem ile eşleřir. Burada propensity skorları arasındaki farkı en küçük olan P_i ve P_j seçilir. Uzaklık olarak en yakın olan kontrol ile eşleme yapmaya çalıřır. Çok fazla ortak deęişken olması durumunda propensity skorları yakın olmayan gözlemlerin eşleřtirilmesi problemine karřı "kabul edilebilir fark (ϵ =caliper)" sınırlaması önerilmiřtir. En yakın eşleřtirme $|P_i - P_j| < \epsilon$ olacak řekilde eşleřtirme için önceden belirlenen bir toleransa göre (ϵ) eşleřtirme gerçekteřir. Kabul edilebilir fark sınırı için 0,1 ve 0,2 yaygın olarak kullanılmakla birlikte Cochran ve Rubin (1973), $\epsilon < 0.25 * \sigma_p$ (σ_p : PS'nin standart sapması) olarak seçilmesini önermiřlerdir.

Stuart (2010), eşleřtirme analizinin ařamalarını ařaęıdaki gibi sıralamıřtır;

1. Eřleřtirme sonucu dengenin saęlanması gereken ortak deęişkenlere/faktörlere karar verilmesi
2. Eřleřtirme için kullanılacak ölçütün (örn. PS) tahmin edilmesi
3. Eřleřtirme yönteminin belirlenmesi (örn. Eřleřtirme, aęırlıklandırma veya alt sınıflandırma)

4. Ortak değişkenlere/faktörlere göre eşleştirme dengesinin değerlendirilmesi (eşleştirme dengesi zayıf ise 2-4 arası adımların tekrar edilmesi)
5. Eşleştirilmiş veri setinden tedavi etkisinin tahmin edilmesi
6. Duyarlılık analizleri

Beşinci aşamada tedavi etkisini belirlemek için yaygın olarak eşleştirme öncesi tedavi ve kontrol grubu ortalama karşılaştırması Student't testi ile yapılırken eşleştirme sonrası eşleştirilmiş t-testi (paired t test) ile yapılmaktadır. Birçok araştırmada gruplar arasında ortak değişkenler açısından fark varken anlamlı bulunan tedavi etkisinin eşleştirme analizi sonrasında gruplar ortak değişkenler açısından benzer duruma getirildiğinde anlamsız bulunduğu görülmektedir.

Araştırmamızda denge değerlendirmesi için literatürde yaygın olarak kullanılan standardize edilmiş ortalama farklar, varyans oranları, Kolomogorov Smirnov test istatistiği, tedavi grubunda ortalama tedavi etkisi (ATT) ve standart hatası ve yanlılık değerleri (%) (gerçek tedavi etkisi ile kestirilen etki arasındaki farkın yüzdesi) kullanılmıştır.

1.2.4.1. Ortalama Farklar (OF), Standardize Edilmiş Ortalamalar Arasındaki Farklar (SOF) ve Varyans Oranları (VO)

$\bar{x}_t$ tedavi grubundaki, $\bar{x}_c$ kontrol grubundaki herhangi bir ortak değişken ortalamasını, s_t^2 tedavi grubundaki, s_c^2 kontrol grubundaki herhangi bir ortak değişken varyansını göstermek üzere sürekli değişkenler için OF, SOF ve VO;

$$OF = \bar{x}_t - \bar{x}_c$$

$$SOF = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}$$

$$VO = \frac{s_t^2}{s_c^2}$$

$\bar{p}_t$ tedavi grubunun, $\bar{p}_c$ kontrol grubunun prevalansını göstermek üzere kategorik değişkenler için OF, SOF ve VO;

$$OF = \bar{p}_t - \bar{p}_c$$

$$SOF = \frac{\bar{p}_t - \bar{p}_c}{\sqrt{\frac{\bar{p}_t(1 - \bar{p}_t) + \bar{p}_c(1 - \bar{p}_c)}{2}}}$$

$$VO = \frac{p_t * (1 - p_t)}{p_c * (1 - p_c)}$$

olarak tanımlanmıştır (Austin, 2009).

R yazılımında bulunan eşleştirme kütüphanelerinin birçoğunda eşleştirme sonrası denge değerlendirmesi için varsayılan olarak SOF ve VO değerleri hesaplanmakta ve literatürde eşleştirme sonrası denge değerlendirmesinde sıklıkla kullanılmaktadır. SOF örneklem büyüklüğünden etkilenmemekte ve farklı birimde ölçülen değişkenlerin göreceli olarak karşılaştırılmasına olanak sağlamaktadır. Mutlak SOF'un 0 olması grupların benzer olduğunu ve dengenin sağlandığını göstermektedir (<0,1 dengenin çok iyi olduğunu gösterir). Varyans oranının 1 olması tam dengeyi göstermektedir (1'e yakın olması dengenin çok iyi olduğunu gösterir).

1.2.4.2. Kolmogorov Smirnov (KS) Test İstatistiği

Bir diğer denge istatistiği olan KS istatistiği tedavi gruplarındaki örneklemelerin ağırlıklandırılmış deneysel dağılım fonksiyonlarına dayanmaktadır. Bu test tedavi ve kontrol grupları arasında ortak değişkenlere göre dağılımın uyuşup uyuşmadığını belirlemeye yardımcı olmaktadır (Demir ve ark., 2017). KS sadece ortalamayı değil tüm dağılımı karşılaştırmaktadır ancak SOF'tan farklı olarak istatistik değeri örneklem büyüklüğüne bağlıdır. Bundan dolayı denge için net bir kesim noktası olmasa da, KS

istatistiği farklı yöntemlerin göreceli başarılarının karşılaştırmasında önemli bir yöntemdir. KS istatistiğinin 0'a yakın olması dengein sağlandığını göstermektedir.

1.2.4.3. Ortalama Tedavi Etkisi (ATE) ve Tedavi Grubunda Ortalama Tedavi Etkisi (ATT)

Y sonuç değişkeni, T_i iki düzeyli bir tedavi değişkeni ve bir birey tedavi aldığı durumda $T=1$ ve almadığı durumda $T=0$ olmak üzere, i . birey için ($i = 1, \dots, n$) tedavi etkisi aşağıdaki gibi tanımlanmıştır (Caliendo ve Kopeinig, 2008).

$$\tau_i = Y_i(1) - Y_i(0)$$

Ancak gözlemsel çalışmalarda her birey için sadece bir potansiyel sonuç değeri olacağı için bu durumda temel bir problem oluşmaktadır ve gözlenemeyen sonuç değişkeni karşı olgu (counter-factual) olarak adlandırılmaktadır (Çizelge 1.2.4.3.1). Bu yüzden i . birey için tedavi etkisi hesaplanması mümkün olmadığı için araştırmalarda ATE ve ATT kavramlarına odaklanılmıştır.

Çizelge 1.2.4.3.1. Potansiyel çıktılar ve karşı olgu (counter-factual)

	Çıktılar Y_1 (Tedavi alındı/Maruziyet var)	Çıktılar Y_0 (Tedavi alınmadı/Maruziyet yok)
Tedavi	Gözlenir	Gözlenemez (Karşı olgu-counterfactual)
Kontrol	Gözlenemez (Karşı olgu-counterfactual)	Gözlenir

Epidemiyoloji ve birçok diğer alanda karşı olgu kavramı, nedensel düşünmenin temelini oluşturmakta ve nedensel etkileri tahmin etmeyi hedefleyen birçok istatistiksel araştırma için çerçeve sağlamaktadır (Heckman, 2005).

Tıp alanında yapılan kontrollü deneylerde kontrol grubu "karşı olgu (counterfactual)" durumunu test etmeye yaramaktadır. Tedaviye, deneye veya herhangi bir girişime maruz kalan deney grubu ile kontrol grubu arasında yapılan

karşılaştırma, nedensellik ilişkisini belirlemeyi amaçlamaktadır. Tedavi dışında tüm yönleriyle tedavi grubuna benzeyen (karıştırıcı faktörlerin bulunmadığı) kontrol grubu "tedavi veya girişim olmasaydı ne olurdu?" sorularının cevabı hakkında fikir edinmek için kullanılmaktadır (Höfler, 2005).

Yansız kontrollü deneylerde tedavi ve kontrol gruplarına rastgele bir şekilde birey atamak hem deneyin sonucundan hem de sonucu belirleyen diğer etkenlerden bağımsız olacaktır. Örneğin başta ekonometri olmak üzere sosyal bilimlerde çoğunlukla bireyler hangi gruba dâhil olacaklarını yaşadıkları şartlara göre kendileri belirlediği için gruplara seçilimi etkileyen faktörlerle deneyin (veya tedavinin) sonucu arasında ilişki bulunmaktadır. Bu yüzden nedensel ilişkilerin etkisinin ekonometrik araştırması zorlaşmaktadır (selection bias). Deney ve kontrol grupları belirlenirken tedavi dışında sonuca etki edecek gözlenebilir veya gözlenemeyen diğer faktörleri kontrol etmek tedavi etkisinin yansız tahminlerinde dikkate alınmalıdır (Höfler, 2005).

Eğer bireyler tedavi ve kontrol gruplarına rastgele atanmışsa (Randomize kontrollü deneysel çalışma) ve tedavi sonrası gruplar arasında anlamlı fark bulunmuşsa farklılığa tedavinin yol açtığı söylenir ve $T \in \{0,1\}$ ve Y sonuç değişkeni olmak üzere ATE aşağıdaki eşitlikteki gibi (tedavi grubunun ortalaması ile kontrol grubunun ortalaması arasındaki fark) tanımlanabilir.

$$E(Y | T=1) - E(Y | T=0)$$

Diğer bir ifadeyle aşağıdaki eşitlikte β_1 ortalama tedavi etkisidir.

$$Y = \alpha + \beta_1 T + \varepsilon$$

Gözlemsel çalışmalarda ATT eşitlikteki gibi tanımlanabilir (tedavi grubu ortalaması ile karşı olgu ortalaması arasındaki fark) (Austin, 2011).

$$E(Y(1) | T=1) - E(Y(0) | T=1)$$

Eşitlikte tedavi alanların karşı olgu ortalaması $E(Y(0)|T=1)$ dır. Çizelge 1.2.4.3.1’de gösterildiği gibi karşı olgu araştırmada gözlenememektedir.

Diğer bir ifadeyle aşağıdaki eşitlikte Z karıştırıcı değişken olmak üzere, β_1 ortalama tedavi etkisidir. Randomizasyon yapılmadığı durumlarda β_1 yanlı tahminler verecektir.

$$Y = \alpha + \beta_1 T + \beta_2 Z + \varepsilon$$

Gözlemsel çalışmalarda tedavi almayan bireylerin sonuç değişkenlerinin ortalamasını $E(Y(0)|T=0)$ kullanmak anlamsız olacaktır. Çünkü bir bireyin tedavi almasını etkileyen sebepler sonuç değişkenini de etkilemiştir. Yani tedavi verilmeden öncede tedavi ve kontrol grupları arasında fark olması muhtemeldir (Imbens, 2004).

Randomizasyonun yapıldığı ve sonuç değişkeninden bağımsız olarak tüm bireylerin tedavi grubuna atanmasında eşit olasılığa sahip olduğu bir deneysel çalışmada ATE, ATT' ye eşit olacaktır ($ATE=ATT= E(Y|T=1) - E(Y|T=0)$). Çünkü randomizasyondan dolayı kontrol grubu, tedavi grubunun gözlemlenemeyen karşı olgularının yerine geçecektir ($E(Y_0|T=1) = E(Y_0|T=0)$ ve $E(Y_1|T=1) = E(Y_1|T=0)$). Randomizasyonun yapılamadığı çalışmalarda ise $ATE \neq ATT$ olacaktır. Çünkü bazı bireylerin tedavi grubuna atanma olasılığı sıfırdır. Bu yüzden gözlemsel çalışmalarda tedavi grubundaki bireylerin PS'sine en yakın olan kontrol grubundaki bireyler seçilerek karşılaştırmalar yapılır. Böylece, tedavi atama mekanizmasının bağımsızlığı ile ilgili kritik varsayım teorik olarak tatmin edilmiş olur.

Aşağıda deneysel ve gözlemsel çalışmalar için ATE ve ATT hesaplanması basit bir örnek üzerinde anlatılmıştır.

No	Tedavi (Treat)	Sonuç (Outcome)		Tedavi Etkisi (Causal Effect)
		Tedavi aldıktan sonra	Tedavi almadan önce	
1	+	80	65	15
2	+	85	75	10
3	+	85	70	15
4	+	80	75	5
5	-	75	70	5
6	-	60	70	-10
7	-	90	85	5
8	-	80	80	0

Karşı olgu (counterfactual outcome)

Şekil 1.2.4.3.1. Deneysel çalışmalar için örnek veri seti

Şekil 1.2.4.3.1’de verilen deneysel çalışmalar için örnek veri seti kullanılarak ortak değişkenlerin bulunmadığı durumda bireysel tedavi etkileri kullanılarak ATE’nin hesaplanması aşağıda yapılmıştır.

$$ATE = (15+10+15+5+5-10+5+0)/8 = 5,625$$

No	Cinsiyet	Tedavi (Treat)	Sonuç (Outcome)		Tedavi Etkisi (Causal Effect)
			Tedavi aldıktan sonra	Tedavi almadan önce	
1	E	+	80	-	-
2	E	+	85	-	-
3	E	+	85	-	-
4	K	+	80	-	-
5	E	-	-	70	-
6	K	-	-	70	-
7	K	-	-	85	-
8	K	-	-	80	-

Şekil 1.2.4.3.2. Gözlemsel çalışmalar için tek ortak değişkenin bulunduğu örnek veri seti

Gözlemsel çalışmalarda karşı sonuç değişken değerleri bilinmediği için bireysel tedavi etkisi hesaplanamaz. Şekil 1.2.4.3.2’de verilen gözlemsel çalışmalar için örnek

veri seti kullanılarak ortak değişkenlerin bulunduğu durumda ATE'nin hesaplanması aşağıda yapılmıştır.

$$\text{Erkekler için ATE} = \left(\frac{80+85+85}{3} \right) - 70 = 13,3$$

$$\text{Kadınlar için ATE} = 80 - \left(\frac{70+85+80}{3} \right) = 1,67$$

$$\text{ATE} = \left(\frac{1}{2} \right) * 13,3 + \left(\frac{1}{2} \right) * 1,67 = 7,49$$

Eğer randomizasyon yapılmış olsa idi ATE ortak değişken/faktör olan cinsiyet dikkate alınmadan aşağıdaki gibi hesaplanabilirdi.

$$\text{ATE} = ((80+85+85+80)/4) - ((70+70+85+80)/4) = 6,25$$

Ancak çoğu gözlemsel araştırmada birden çok ortak değişken bulunduğu için yukarıdaki gibi tabakalı hesaplama yapmak mümkün olamamaktadır.

No	Cinsiyet	Yaş	Tedavi (Treat)	Sonuç (Outcome)		PS	IPTW		
				Tedavi aldıktan sonra	Tedavi almadan önce		ATT	Sonuç	ATT2
1	E	25	+	80		0,4972978	1	80	2,011
2	E	45	+	85		0,5097693	1	85	1,962
3	E	18	+	85		0,4929328	1	85	2,029
4	K	25	+	80		0,5003118	1	80	1,999
5	E	26	-		70	0,4979215	0,992	69,4	1,992
6	E	45	-		70	0,5097693	1,040	72,8	2,040
7	E	17	-		85	0,4923093	0,970	82,4	1,970
8	K	24	-		80	0,4996882	0,999	79,9	1,999

$$w_{tedavi}=1, w_{kontrol}=\frac{p_i}{1-p_i} \quad w_{tedavi}=\frac{1}{p_i}, w_{kontrol}=\frac{1}{1-p_i}$$

Şekil 1.2.4.3.3. Gözlemsel çalışmalar için birden fazla ortak değişkenin bulunduğu örnek veri seti

Şekil 1.2.4.3.3'te gözlemsel çalışmalar için birden fazla ortak değişkenin bulunduğu örnek veri seti için ATT hesaplanması aşağıda gösterilmiştir.

$$ATT = \frac{80+85+85+80}{4} - \frac{70 \times 0,992 + 70 \times 1,040 + 85 \times 0,970 + 80 \times 0,999}{0,992 + 1,040 + 0,970 + 0,999} = 6,367$$

ATT hesaplanırken ağırlıklandırma farklı şekillerde yapılmaktadır. Şekil 1.2.4.3.3'te 2 farklı ağırlıklandırma (IPTW: Ters olasılık ağırlıklandırması) gösterilmiştir. PS' ye bağlı olarak hesaplanan ağırlıklar ATT hesaplanmasında kullanılmaktadır. Dolayısıyla ağırlıklandırma yöntemine göre ATT değişmektedir (Stuart, 2010).

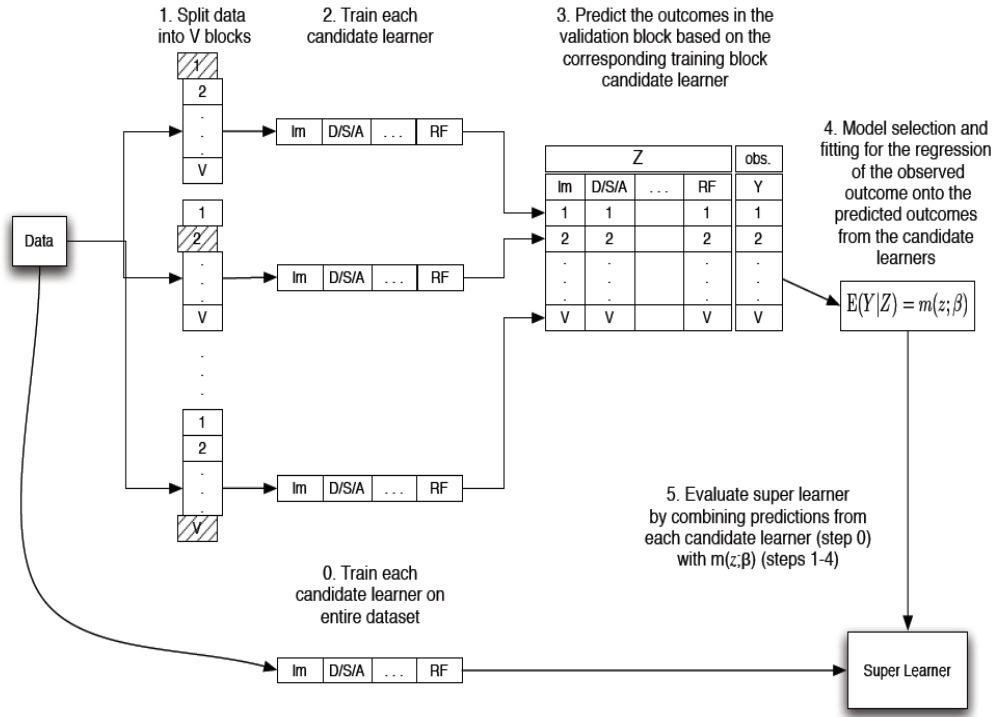
Bu tez çalışmasında modelde sadece ana etkiler bulunmadığı yani etkileşim etkilerinin ve doğrusal olmayan ilişkilerin olduğu durumlarda LR'ye alternatif veya daha başarılı makine öğrenme algoritmalarının belirlenmesi, belirlenen algoritmalar ile bir kombine algoritma oluşturmak suretiyle PS tahmini yapılması, hangi modellerde hangi yöntemlerin daha başarılı olduğu sonucunu ortaya koyarak literatürdeki mevcut eksikliğin giderilmesi amaçlanmıştır. Bu nedenle tez çalışmasında; PS konusunda geniş bir literatür taraması ve bibliyometrik analizler yapılmak suretiyle bu konudaki son gelişmeler belirlenerek, PS tahmini için klasik yöntem olan LR'ye alternatif olarak geliştirilmiş doğrusal modeller, makine öğrenme yöntemleri (geliştirilmiş boosted modeller, sınıflandırma ve regresyon ağaçları, rasgele ormanlar, destek vektör makineleri ve yapay sinir ağları vb.), bazı yaygın kullanılan yöntemler (bayesci geliştirilmiş doğrusal modeller, kuadratik diskriminant analizi ve adımsal akaike bilgi kriteri vb.) ve bu yöntemlerin kombinasyonunun başarısının bir benzetim çalışması ile değerlendirilmesi ve benzetim çalışması sonucunda ulaşılabilecek bulgulara göre tıp araştırmacılarının gözlemsel ve yarı deneysel araştırmalarında eşleştirme analizi için kullanabileceği en uygun eşleştirme sağlayan yeni propensity skor tahmin modellerinin geliştirilmesi hedeflenmiştir.

2. GEREÇ ve YÖNTEM

2.1. Benzetim Çalışması

Uygulamada Monte Carlo Benzetim yöntemi kullanılmıştır. Benzetim ve uygulamalarda ML algoritmaları ve diğer tahminciler için R paket programında açık kaynak kodlu SuperLearner (Polley ve Van Der Laan, 2010) ve gerekli olan diğer alt kütüphaneler "arm", "biglasso", "caret", "dbarts", "e1071", "earth", "gam", "gbm", "ggplot2", "glmnet", "ipred", "kernelknn", "kernlab", "nnet", "party", "polspline", "randomforest", "rpart", "sva", "xgboost" kullanılmıştır. Kestirilen PS'lerin uyum analizleri için "irr" (Gamer ve ark., 2015) kütüphanesi kullanılmıştır.

SuperLearner paketi, Polley ve Van Der Laan (2010), tarafından parametrik regresyon modelleri ve birçok makine öğrenme algoritması içerisinde en başarılı kestirim elde eden yöntemin veya kombine yöntemlerin seçimi için geliştirilmiştir. Paket V Katlamalı Çarpaz Doğrulama (V-fold cross-validation) kullanarak hesapladığı kayıp değer fonksiyonuna (risk tahmini) göre birçok aday makine öğrenme algoritmaların doğrusal bileşimini ağırlıklandırarak yeni bir kombine tahminci (ensemble learning) sağlamaktadır. Paketin en büyük avantajı ağırlıklandırmayı kendisi yaparak kombine tahminciyi oluşturmasıdır. Kombine tahmin gücüne katkı yapmayan algoritmalara 0 ağırlık vererek modelden çıkardığı için birçok algoritmayı denemeye gerek kalmamaktadır. Kombine yöntemler tahmin süresini uzatsa da basit ve doğrusal ilişkilerin bulunmadığı karmaşık ilişki bulunan model tahminlerinde yüksek performans göstermektedir. Polley algoritmanın çalışma mantığını doktora tezinde Şekil 2.1.1 ile özetlemiştir.



Şekil 2.1.1. SuperLearner çalışma algoritmasını gösteren akış diyagramı (Polley ve Van Der Laan, 2010)

Tez çalışmamızda uygulamalar paketin 12 Mart 2018 son sürümü ile gerçekleştirilmiştir. Pakette toplam 42 adet farklı tahminci bulunmaktadır. Bunlar; "SL.bartMachine", "SL.bayesglm", "SL.biglasso", "SL.caret", "SL.caret.rpart", "SL.cforest", "SL.dbarts", "SL.earth", "SL.extraTrees", "SL.gam", "SL.gbm", "SL.glm.interaction", "SL.glmnet", "SL.ipredbag", "SL.kernelKnn", "SL.knn", "SL.ksvm", "SL.lda", "SL.leekasso", "SL.lm", "SL.loess", "SL.logreg", "SL.mean", "SL.nnet", "SL.nnls", "SL.polymars", "SL.qda", "SL.randomForest", "SL.ranger", "SL.ridge", "SL.rpart", "SL.rpartPrune", "SL.speedglm", "SL.speedlm", "SL.step", "SL.step.forward", "SL.step.interaction", "SL.stepAIC", "SL.svm", "SL.template", "SL.xgboost" dır.

Bu algoritmalarından iki düzeyli tedavi değişkenini sınıflamak ve propensity skor değerlerini hesaplamak için iki düzeyli sonuç değişkeni tahmini için uygun görülen 21 algoritma kullanılmış ve Çizelge 2.1.1'de algoritmalar açıklamaları ile birlikte gösterilmiştir.

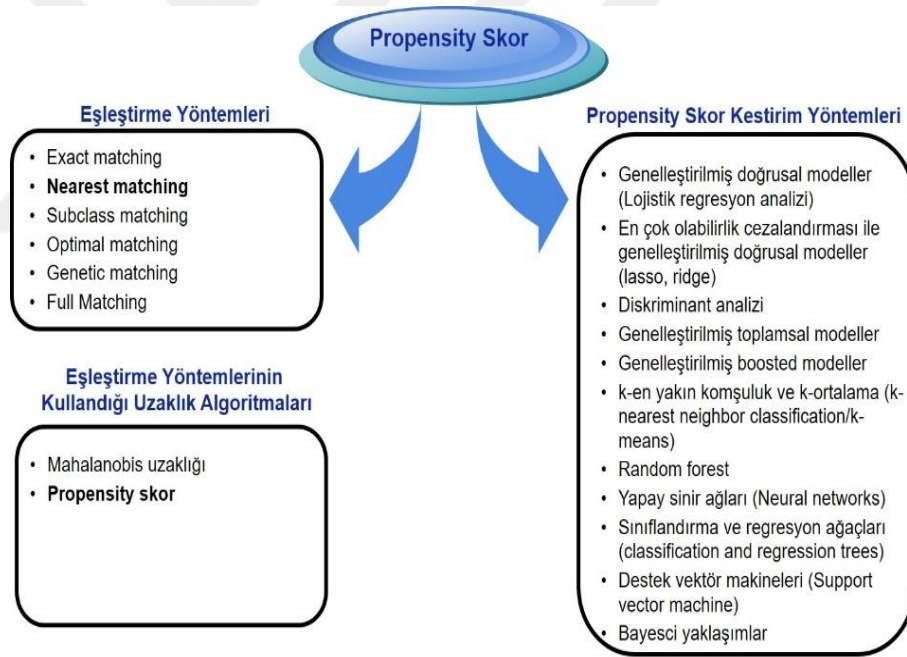
Çizelge 2.1.1. Tahminci algoritmalar ve tanımlamaları

No	Tahminci	Tanım	Orijinal tanım	R paket
1	"SL.bayesglm"	Bayesci genelleştirilmiş doğrusal model	Bayesian generalized linear model (glm)	arm
2	"SL.caret"	Sınıflama ve regresyon ağaçları	Classification and regression tree (CART)	caret
3	"SL.caret.rpart"	Özyinelemeli parçalama karar ağaçları	Recursive partitioning / decision trees	caret
4	"SL.cforest"	Koşullu rasgele ormanlar	Conditional random forests	party
5	"SL.dbarts"	Kesikli Bayesci eklemeli regresyon ağaçları	Discrete bayesian additive regression trees	dbarts
6	"SL.earth"	Çok değişkenli uyarlanabilir regresyon eğrileri	Multivariate adaptive regression splines	earth
7	"SL.gbm"	Boosted model	Gradient boosting	gbm
8	"SL.gam"	Genelleştirilmiş toplamsal model	Generalized additive model	gam
9	"SL.glmnet"	Cezalandırılmış, genelleştirilmiş lineer modeller	Generalized linear model with penalized maximum likelihood (elastic net: lasso-ridge)	glmnet
10	"SL.ipredbagg"	Geliştirilmiş tahmin modelleri	Improved predictive models bagging classification, regression and survival trees	ipred
11	"SL.kernelKnn"	Çekirdek k en yakın komşu algoritması	Kernel k-nearest neighbors	kernelKnn
12	"SL.knn"	K en yakın komşuluk	k-nearest neighbors	knn
13	"SL.ksvm"	Çekirdek destek vektör makineleri	Kernel-based machine learning, kernel support vector machines	kernlab
14	"SL.nnet"	Yapay sinir ağları	Neural network	nnet
15	"SL.qda"	Karesel (kuadratik) diskriminant analizi	Quadratic discriminant analysis.	qda
16	"SL.randomForest"	Rasgele ormanlar	Random forest	randomForest
17	"SL.rpart"	Özyinelemeli parçalama	Recursive partitioning for classification, regression and survival trees	rpart
18	"SL.rpartPrune"	Budamalı özyinelemeli regresyon ağaçları	Pruning recursive partitioning regression trees	rpart
19	"SL.stepAIC"	Adımsal Akaike bilgi kriteri	Stepwise model selection by Akaike Information Criterion (AIC)	stepAIC
20	"SL.svm"	Destek vektör makineleri	Support vector machine	e1071
21	"SL.xgboost"	Ölçeklenebilir gradient boosting	A version of GBM, eXtreme gradient boosting	xgboost

Eşleştirme analizi ve PS tahmini Sekhon (2011) tarafından geliştirilen Matching (Multivariate and propensity score matching with balance optimization) kütüphanesi

kullanılarak yapılmıştır. Denge değerlendirmesi için kullanılan grafik çizimleri ise “ggplot2” kütüphanesi kullanılarak yapılmıştır.

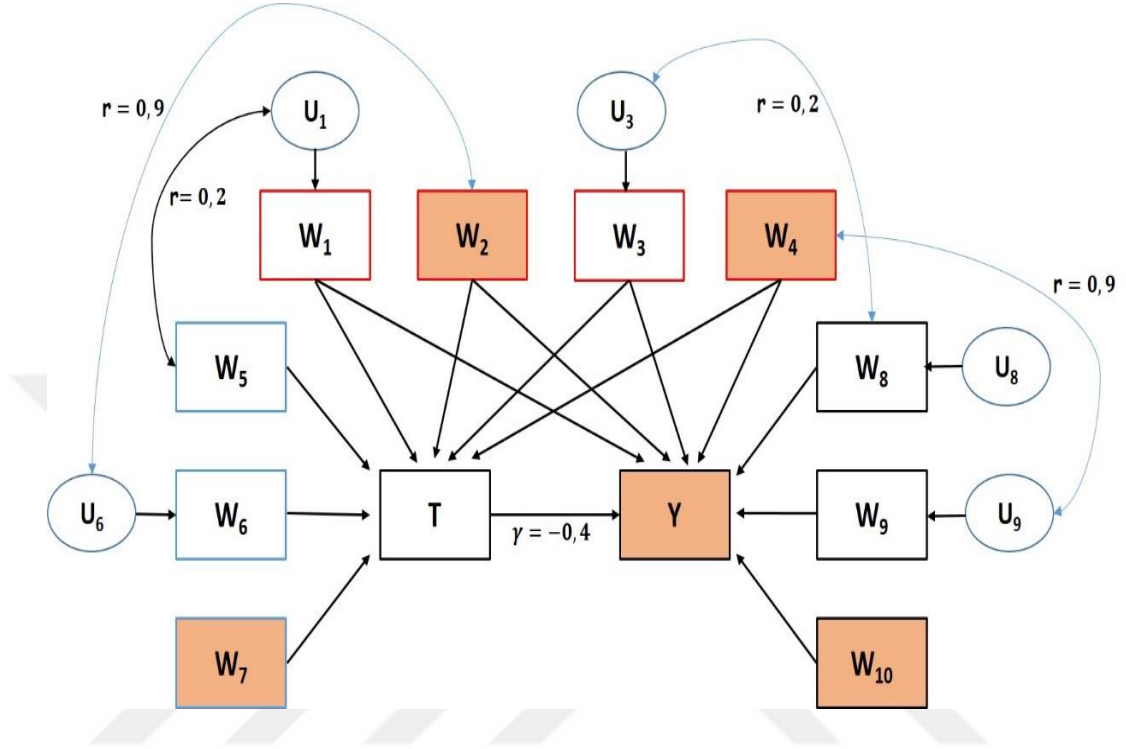
Literatürde birçok eşleştirme yöntemi bulunmaktadır (Şekil 2.1.2). Bu yöntemler içerisinde de farklı seçeneklere göre eşleştirme sonuçları değişmektedir. Araştırmamızda yaygın olarak kullanılan PS’ye dayalı en yakın komşu eşleştirme (Nearest neighbor matching) algoritması kullanılmıştır. Eşleştirme analizi uygulamalarında kontrol grubundaki bir bireyin tedavi grubundaki birden fazla birey ile eşleşmesine izin veren “with replacement” seçeneği kullanılmıştır. Literatürde tedavi grubundaki birden fazla bireye benzerlik gösteren kontrollerin birden çok kez kullanılmasının yanlılığı azaltacağı belirtilmiştir (Caliendo ve Kopeinig, 2008; Stuart, 2010; Austin, 2014).



Şekil 2.1.2. Eşleştirme yöntemleri

Uygulamanın birinci bölümünde farklı senaryolar oluşturularak Monte Carlo benzetim yöntemi kullanılmıştır. Benzetim için Setoguchi ve ark. (2008)’nın önerdiği ve birçok araştırmada kullanılan (Lee ve ark., 2010; Lee ve ark., 2011; Leacy ve ark., 2014; Pirracchio ve ark., 2014) benzetim veri seti kullanılmıştır (Şekil 2.1.3). Setoguchi ve ark. (2008)’nın önerdiği veri setinde tedavi etkisini gösteren sonuç

değişkeni (Y_i), tedavi ve ortak değişkenlerin bir doğrusal kombinasyonu olacak şekilde Lee ve ark. (2010) tarafından değiştirilmiştir. Bizim çalışmamızda da bu düzeltme yapılmıştır. Benzetim verisi için oluşturulan R kodları Ek 1’de verilmiştir.



Şekil 2.1.3. Benzetim veri seti

Veri setinde tedavi ve kontrol gruplarını gösteren bir kategorik değişken (T), tedavi etkisini gösteren ve karşılaştırmalara temel olacak bir sürekli sonuç değişkeni (Y), dört farklı karıştırıcı değişken (W_1 - W_4), sadece tedavi değişkeni ile ilişkili üç farklı ortak değişken (W_5 - W_7) ve sadece sonuç değişkeni ile ilişkili üç farklı ortak değişken (W_8 - W_{10}) bulunmaktadır. Bu değişkenlerden W_1 , W_3 , W_5 , W_6 , W_8 , W_9 iki düzeyli (0,1) kategorik değişken ve W_2 , W_4 , W_7 , W_{10} sürekli değişken olarak üretilmiştir. Değişkenlerden W_1 - W_4 , W_7 ve W_{10} standart normal dağılımdan üretilmiştir ($\mu = 0, \sigma = 1$). Daha sonra W_5 , W_6 , W_8 ve W_9 üretilmiştir. W_1 ile W_5 ve W_3 ile W_8 arasında 0,2 korelasyon, W_2 ile W_6 ve W_4 ile W_9 arasında 0,9 korelasyon bulunacak şekilde üretilmiştir. Çizelge 2.1.2’de benzetim veri üretiminde kullanılan korelasyon matrisi gösterilmiştir.

Çizelge 2.1.2. Benzetim veri üretiminde kullanılan korelasyon matrisi

		Karıştırıcı Değişkenler (Confounders)				Maruziyet belirleyici değişkenler (Exposure predictors)			Sonuç belirleyici değişkenler (Outcome predictors)		
		W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀
Karıştırıcı Değişkenler	W ₁	1									
	W ₂	0	1								
	W ₃	0	0	1							
	W ₄	0	0	0	1						
Maruziyet belirleyici değişkenler	W ₅	0,2	0	0	0	1					
	W ₆	0	0,9	0	0	0	1				
	W ₇	0	0	0	0	0	0	1			
Sonuç belirleyici değişkenler	W ₈	0	0	0,2	0	0	0	0	1		
	W ₉	0	0	0	0,9	0	0	0	0	1	
	W ₁₀	0	0	0	0	0	0	0	0	0	1

Son olarak W₁, W₃, W₅, W₆, W₈, W₉ değişkenleri içerisindeki her bir veri genel ortalama ile karşılaştırılarak ortalamadan büyük olanlar 1, küçük olanlar 0 olacak şekilde kategorik değişkene dönüştürülmüştür. Sonuç değişkeni (Y), tedavi değişkeni ve ortak değişkenlerin bir doğrusal kombinasyonu olarak üretilmiştir ($Y = a_0 + a_i W_i + \gamma T$). γ tedavi etkisi olmak üzere -0,4 olarak üretilmiştir. Tedavi değişkeni iki düzeyli kategorik bir değişken olarak ortak değişkenlerin bir fonksiyonu olacak şekilde Lojistik Regresyon ($P(T = 1 | W_i) = 1 / (1 + \exp(-\beta \cdot f(W_i)))$) kullanılarak üretilmiştir. Tedavi alma veya kontrol grubunda olma olasılığını belirleyecek bu modelde $f(W_i)$ alttaki senaryolara göre ortalama maruziyet olasılığı yaklaşık 0,5 olacak şekilde Bernoulli dağılımından üretilmiştir.

- Senaryo A: Toplamsal ve doğrusal (sadece temel etkiler; additivity and linearity, $P(T = 1 | W_i) = (1 + \exp(-(0 + b_1 \cdot w_1 + b_2 \cdot w_2 + b_3 \cdot w_3 + b_4 \cdot w_4 + b_5 \cdot w_5 + b_6 \cdot w_6 + b_7 \cdot w_7)))^{-1}$)
- Senaryo B: 4 iki yönlü etkileşim terimi ve 1 kuadratik terim (hafif düzeyde karmaşık ilişkiler; mild non-additivity and non-linearity, $P(T = 1 | W_i) = (1 + \exp(-(0 + b_1 \cdot w_1 + b_2 \cdot w_2 + b_3 \cdot w_3 + b_4 \cdot w_4 + b_5 \cdot w_5 + b_6 \cdot w_6 + b_7 \cdot w_7 + b_2 \cdot w_2 \cdot w_2 + b_1 \cdot 0.5 \cdot w_1 \cdot w_3 + b_2 \cdot 0.7 \cdot w_2 \cdot w_4 + b_4 \cdot 0.5 \cdot w_4 \cdot w_5 + b_5 \cdot 0.5 \cdot w_5 \cdot w_6)))^{-1}$)
- Senaryo C: 10 iki yönlü etkileşim terimi ve 3 kuadratik terim (orta düzeyde karmaşık ilişkiler; moderate non-additivity and non-linearity, $P(T = 1 | W_i) = (1 + \exp(-(0 + b_1 \cdot w_1 + b_2 \cdot w_2 + b_3 \cdot w_3 + b_4 \cdot w_4 + b_5 \cdot w_5 + b_6 \cdot w_6 + b_7 \cdot w_7 +$

$$b2*w2*w2 + b4*w4*w4 + b7*w7*w7 + b1*0.5*w1*w3 + b2*0.7*w2*w4 + b3*0.5*w3*w5 + b4*0.7*w4*w6 + b5*0.5*w5*w7 + b1*0.5*w1*w6 + b2*0.7*w2*w3 + b3*0.5*w3*w4 + b4*0.5*w4*w5 + b5*0.5*w5*w6)) ^{-1})$$

Modelde sadece ana etkilerin olmadığı, tedavi grubuna atanma olasılığının ortak değişkenler ile doğrusal olmayan ve etkileşimler içeren ilişkiler halinde bulunduğu senaryolarda daha başarılı olacağı düşünülen makine öğrenme algoritmalarının pratikte LR yerine eşleştirme analizlerinde kullanılabilmesi için modelde sadece temel etkiler olan modellerde de en az LR kadar başarılı olması gerekmektedir. Bu yüzden PS tahmininde başarılı ML ve diğer algoritmaları belirlemek ve ayrıca uygulamanın ikinci bölümünde kullanılacak en iyi kombine algoritma içerisinde bulunacak aday algoritmaları belirleyebilmek için öncelikle sadece A senaryosu kullanılarak orta (n=500) ve büyük (n=1000) olmak üzere 2 farklı örneklem büyüklüğünde veri seti üretilmiştir. A senaryosu ile üretilen 2 farklı örneklem büyüklüğündeki veri seti kullanılarak 21 farklı tahminci yöntemin her biri için 10.000 benzetim ile PS'ler hesaplanmıştır. 21 farklı benzetim içerisinde her bir yöntem ile birlikte LR kullanılarak da PS hesaplanmıştır. Her bir benzetim çalışması içerisinde ML yöntemleri ve LR ile hesaplanan PS'ler arasındaki korelasyon ve uyum değerleri hesaplanmıştır. Birinci benzetim çalışması için toplamda 420.000 kez PS hesaplanmıştır. PS'ler arasındaki korelasyon için Pearson korelasyon katsayısı ve uyumu değerlendirmek için sınıf içi korelasyon katsayısı (SKK) kullanılmıştır. Veri üretiminde doğru PS modeli ve sonuç değişkeni için kullanılan katsayılar Çizelge 2.1.3'de verilmiştir.

Çizelge 2.1.3. Benzetim veri üretiminde kullanılan katsayılar

	Doğru PS Modeli (T)	Sonuç Modeli (Y)
Sabit	0	-3,85
Katsayı 1	0,8	0,3
Katsayı 2	-0,25	-0,36
Katsayı 3	0,6	-0,73
Katsayı 4	-0,4	-0,2
Katsayı 5	-0,8	0,71
Katsayı 6	-0,5	-0,19
Katsayı 7	0,7	0,26

Üretilen veri setinin doğrulanması regresyon analizi ile yapılmış ve A senaryosu ile üretilen 1000 örneklem büyüklüğündeki bir veri setine ait sonuçlar Çizelge 2.1.4'te

verilmiştir. Çizelge incelendiğinde tedavi etkisinin -0,4 olduğu ve diğer katsayıların doğrulandığı görülmüş ve verinin doğru olarak üretildiği belirlenmiştir.

Çizelge 2.1.4. A senaryosu n=1000 benzetim veri setine ait regresyon analizi sonuçları

$R^2=1$	β	p değeri
Tedavi etkisi	-0,4	<0,001***
W1	0,3	<0,001***
W2	-0,36	<0,001***
W3	-0,73	<0,001***
W4	-0,2	<0,001***
W5	3,699e-17	0,802
W6	1,395e-16	0,503
W7	-2,093e17	0,789
W8	0,71	<0,001***
W9	-0,19	<0,001***
W10	0,26	<0,001***

Araştırmanın ikinci bölümünde A senaryosuna ek olarak sadece ana etkilerin olmadığı etkileşim etkilerinin ve kuadratik terimlerinde bulunduğu B ve C senaryoları için 500 ve 1000 örneklem büyüklüğünde veri üretilerek ilk benzetim sonucunda başarılı uyum düzeyine sahip bireysel algoritmalar ve LR ile elde edilen PS'ler kullanılarak eşleştirme analizleri gerçekleştirilmiştir. Üretilen iki farklı örneklem büyüklüğündeki veri setleri için her bir benzetim içerisinde LR ve başarılı bulunan 12 tahmincinin her biri ile 1000 benzetim kullanılarak PS'ler hesaplanmış ve bu PS'ler ile eşleştirme analizleri yapılmıştır. Toplamda 26.000 kez hesaplanan PS'ler ile 26.000 eşleştirme analizi gerçekleştirilmiştir. Eşleştirme analizleri sonucunda elde edilen denge değerlendirmesi sonuçlarına göre tahminciler karşılaştırılmıştır. Denge açısından başarılı olarak değerlendirilen tahminciler ile PS hesabı için yeni bir kombine tahminci model oluşturularak son bir benzetim çalışması daha yapılmıştır. Son benzetim çalışmasında kombine tahminci ve klasik model olan LR ile iki farklı örneklem büyüklüğünde her bir senaryo için 1000 benzetim yapılarak elde edilen PS'ler ile yapılan eşleştirme analizi denge sonuçları karşılaştırılmıştır. Toplamda 12.000 kez hesaplanan PS'ler ile 12.000 eşleştirme analizi gerçekleştirilmiştir. Denge değerlendirmeleri için tüm benzetimlerde ATT ve yanlılık değeri, SOF, OF, VO ve KS değerleri kullanılmıştır. Benzetimlerde PS tahminleri ve eşleştirme analizi için benzetim kodları içerisinde kullanılan kısaltılmış özet R kodları Ek 1'de verilmiştir.

2.2. Gerçek Veri Seti İle Uygulama

Araştırmanın bu bölümünde literatürde eşleştirme sonucunda yöntem karşılaştırmalarında yaygın olarak kullanılan Lalonde'nin gerçek veri setinde yöntemlerin başarısı incelenmiştir. Lalonde (1986) veri seti "American Economic Review" dergisinde yayınlanmış "Evaluating the Econometric Evaluations of Training Programs" makale içerisinde bulunan ve temel iş becerilerine sahip olmayan kişilere iş deneyimi ve danışmanlık sağlayarak gelir düzeylerini artırmayı hedefleyen büyük ölçekli bir istihdam eğitim programı (9-12 ay süren işbaşı eğitimi) sonrası bireylerin gelir düzeylerine eğitimin etkisini araştırmak için toplanmış bir veridir.

Araştırmamızda veri setinin kullanıma açık olan ve Dehejia ve Wahba (1999) tarafından yayınlanan deneysel sürümü kullanılmıştır. Deneysel sürümünde tedavi grubu National Supported Work Demonstration (NSW)'dan ve kontrol grubu Current Population Survey (CPS)'den elde edilen verilerden oluşmaktadır. R paketinde eşleştirme analizi için en sık kullanılan MatchIt ve Matching paketlerinde de uygulamalar ve örnek anlatımlar için lalonde veri seti kullanılmaktadır. Veri seti içerisinde tedavi değişkeni treatment (1=eğitim aldı, 0=eğitim almadı; tedavi grubunda 185, kontrol grubunda 260 gözlem), sürekli ortak değişkenler; age (yaş), years of education (eğitim yılı), kategorik ortak değişkenler; black race (1=siyah ırk: Afrikan-Amerikan, 0=diğer), hispanic ethnicity (1=latin asıllı etnik köken, 0=diğer), marital status (1=evli, 0=bekar), high school completion (1=lise tamamlama derecesi yok, 0=var), real earnings in 1974 (1974 yılındaki kazanç) ve real earnings in 1975 (1975 yılındaki kazanç) ve sonuç değişkeni real earnings in 1978 (1978 yılındaki kazanç) bulunmaktadır.

Eşleştirme analizinde benzetim çalışmalarına benzer olarak kontrol grubundaki bir bireyin tedavi grubundaki birden fazla birey ile eşleşmesine izin veren "with replacement" seçeneği ile birlikte en yakın komşu eşleştirme (Nearest neighbor matching) algoritması kullanılmıştır.

3. BULGULAR

3.1. Lojistik Regresyon ile En Uyumlu Algoritmaların Belirlenmesi

A senaryosu ile üretilen n=500 ve n=1000 veri seti kullanılarak LR ve 21 farklı makine öğrenme algoritması ile 10.000 benzetim sonucunda hesaplanan PS değerleri arasındaki Pearson korelasyon katsayısı ve güven aralıkları ile birlikte SKK değerleri Çizelge 3.1.1’de verilmiştir.

Çizelge 3.1.1. Farklı algoritmalarla kestirilen PS’ler arasındaki korelasyon ve uyum değerleri

Tahminci algoritmalar	n=500		n=1000	
	r	SKK	r	SKK
SL.bayesglm	1.000	1.000 (1.000-1.000)	1.000	1.000 (1.000 -1.000)
SL.caret	0.651	0.619 (0.562-0.670)	0.688	0.657 (0.620-0.690)
SL.caret.rpart	NA	0.549 (0.485-0.606)	0.585	0.546 (0.502-0.588)
SL.cforest	0.896	0.877 (0.855-0.896)	0.904	0.896 (0.883- 0.907)
SL.dbarts	0.959	0.956 (0.948-0.963)	0.967	0.966 (0.962-0.970)
SL.earth	0.895	0.894 (0.875-0.910)	0.944	0.943 (0.936-0.950)
SL.gbm	0.923	0.909 (0.892-0.923)	0.950	0.943 (0.936-0.950)
SL.gam	0.994	0.993 (0.992-0.995)	0.997	0.997 (0.996- 0.997)
SL.glmnet	0.998	0.995 (0.993-0.995)	0.999	0.998 (0.998-0.998)
SL.ipredbag	0.827	0.824 (0.794-0.851)	0.858	0.840 (0.821-0.857)
SL.kernelKnn	0.753	0.751 (0.710-0.787)	0.770	0.769 (0.743- 0.794)
SL.knn	0.754	0.752 (0.712-0.788)	0.769	0.769 (0.743-0.794)
SL.ksvm	0.923	0.910 (0.894-0.924)	0.935	0.929 (0.921- 0.937)
SL.nnet	0.808	0.791 (0.757-0.822)	0.890	0.885 (0.871-0.897)
SL.qda	0.883	0.876 (0.854-0.895)	0.932	0.930 (0.921 0.938)
SL.randomForest	0.606	0.563 (0.501-0.620)	0.611	0.564 (0.521-0.605)
SL.rpart	0.641	0.634 (0.579-0.684)	0.667	0.663 (0.627-0.696)
SL.rpartprune	0.617	0.602 (0.543-0.655)	0.652	0.641 (0.603- 0.676)
SL.stepAIC	0.973	0.973 (0.968-0.977)	0.988	0.988 (0.986-0.989)
SL.svm	0.688	0.614 (0.557-0.665)	0.693	0.624 (0.585-0.660)
SL.xgboost	0.619	0.559 (0.496-0.616)	0.617	0.555 (0.510-0.596)

r: Pearson korelasyon katsayısı, SKK: Sınıf içi korelasyon katsayısı, NA: Hesaplanamadı

Literatürde yaygın olarak uyum katsayısı SKK değerlerine göre; zayıf (< 0,5); orta (0,5-0,75); iyi (0,75-0,9); mükemmel (> 0,9) olarak yorumlanmaktadır (Koo ve Li, 2016). Araştırmamızda ise Çizelge 3.1.1 ile verilen SKK uyum katsayıları güven aralığı 0,80 ve üzerini kapsayan algoritmalar başarılı olarak değerlendirilmiştir. Çizelge 3.1.1 incelendiğinde LR ile en uyumlu başarılı algoritmalar 1000 örneklem

büyükliğinde sırasıyla; BAYESGLM (SKK=1.000), GLMNET (SKK=0.998), GAM (SKK=0.997), STEPAIC (SKK=0.988), DBARTS (SKK=0.966), GBM (SKK=0.943), EARTH (SKK=0.943), QDA (SKK=0.930), KSVM (SKK=0.929), CFOREST (SKK=0.896), NNET (SKK=0.885) ve IPREDBAGG (SKK=0.840) olarak bulunmuştur. 500 örneklem büyüklüğünde ise sırasıyla; BAYESGLM (SKK=1.000), GLMNET (SKK=0.995), GAM (SKK=0.993), STEPAIC (SKK=0.973), DBARTS (SKK=0.956), KSVM (SKK=0.910), GBM (SKK=0.909), EARTH (SKK=0.894), CFOREST (SKK=0.877), QDA (SKK=0.876), IPREDBAGG (SKK=0.824) ve NNET (SKK=0.791) olarak bulunmuştur.

Araştırmanın devamında sadece ilk benzetim çalışmasında LR ile uyumlu bulunan algoritmalar A, B ve C senaryoları için gerçekleştirilen benzetimlere dahil edilmiştir.

3.2. n= 500 Örneklem Büyüklüğü için Benzetim Sonuçları

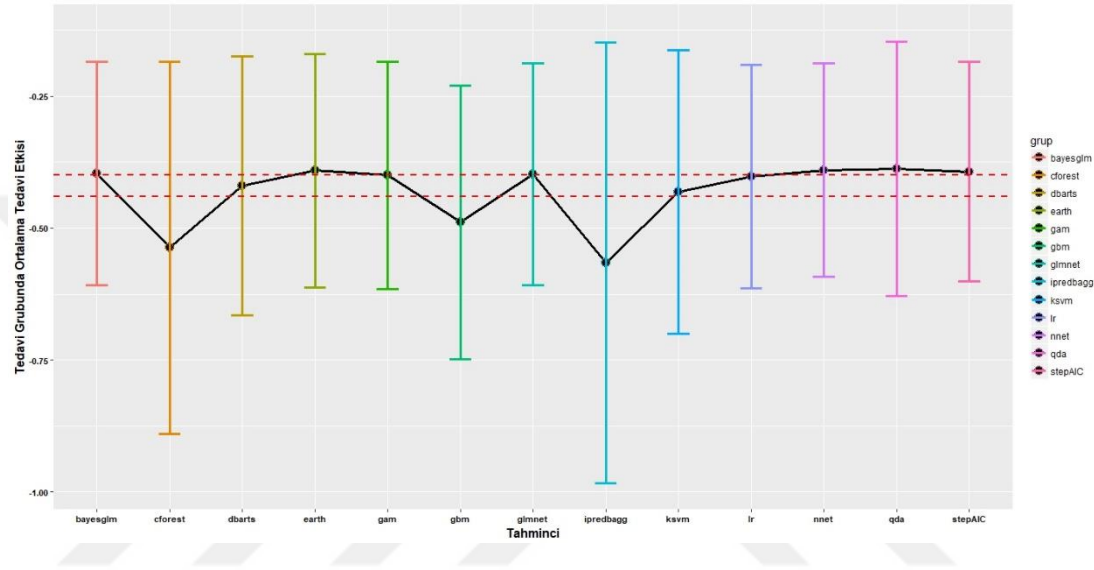
PS'lere dayalı olarak yapılan eşleştirme analizi sonucunda, A senaryosu ve n=500 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak PS'ler hesaplanmış ve PS'ler ile gerçekleştirilen eşleştirme analizi sonucunda elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.2.1'de verilmiştir. Denge sonuçlarına göre LR ile birlikte yanlışlık değeri %10'dan daha az olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH, STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

Çizelge 3.2.1. ATT için denge sonuçları (Senaryo A; n=500)

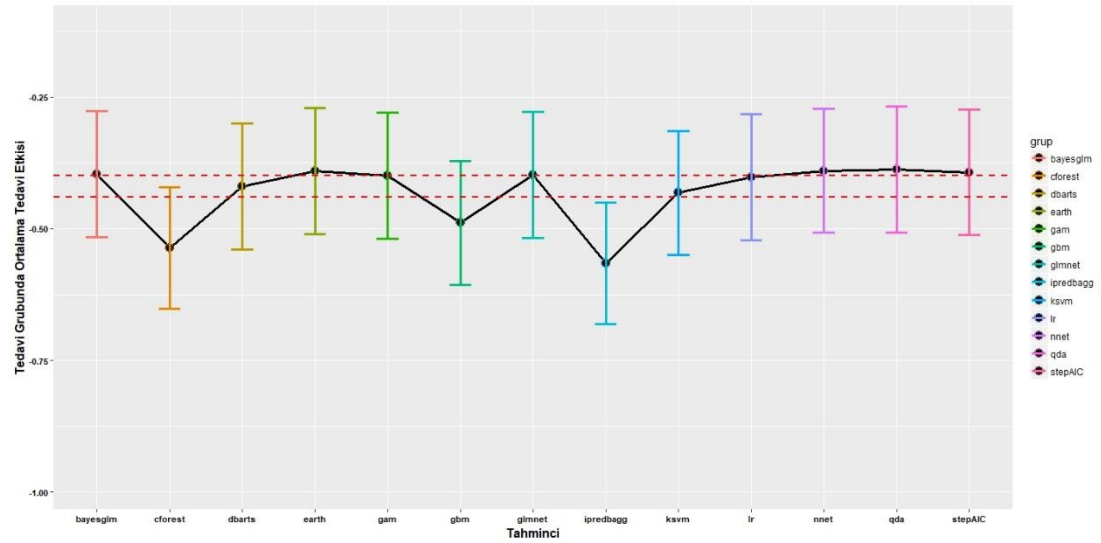
	ATT	AI SH	SH	YANLILIK (%)
LR	-0,402	0,108	0,061	0,494
KSVM	-0,432	0,137	0,060	8,042
BAYESGLM	-0,397	0,108	0,061	-0,689
GLMNET	-0,398	0,107	0,061	-0,575
NNET	-0,390	0,103	0,060	-2,463
DBARTS	-0,420	0,125	0,061	4,983
QDA	-0,388	0,123	0,061	-2,876
GAM	-0,400	0,110	0,061	-0,078

Çizelge 3.2.1. ATT için denge sonuçları (Senaryo A; n=500) (devamı)				
EARTH	-0,391	0,113	0,061	-2,272
STEP AIC	-0,393	0,106	0,061	-1,755
IPREDBAGG	-0,566	0,213	0,059	41,421
GBM	-0,489	0,132	0,060	22,150
CFOREST	-0,537	0,180	0,059	34,336

Ayrıca ATT'ler AI standart hataları ile birlikte Şekil 3.2.1 ve klasik standart hatalar ile birlikte Şekil 3.2.2' de karşılaştırılmıştır.



Şekil 3.2.1. Tahmincilere göre ATT ve AI standart hata ile güven aralıkları (Senaryo A, n=500)

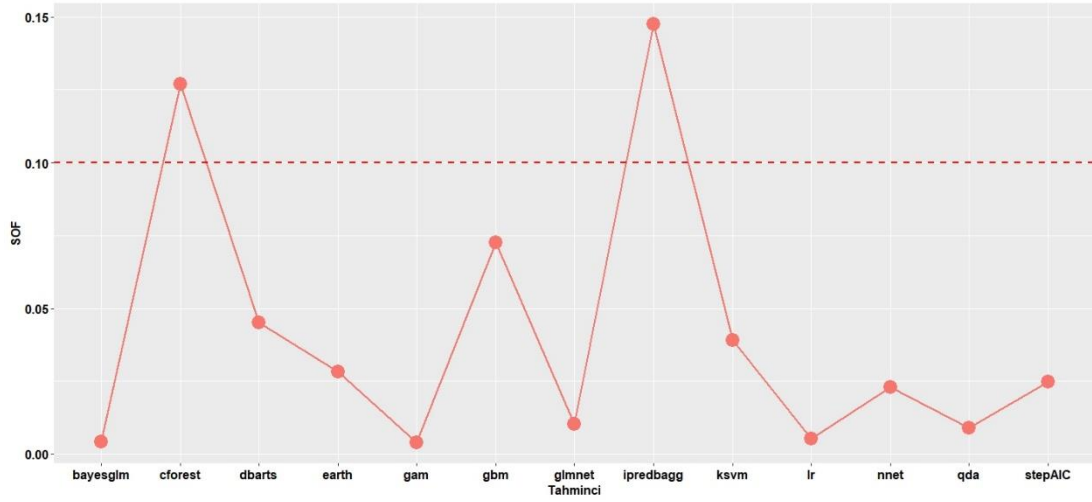


Şekil 3.2.2. Tahmincilere göre ATT ve klasik standart hata ile güven aralıkları (Senaryo A, n=500)

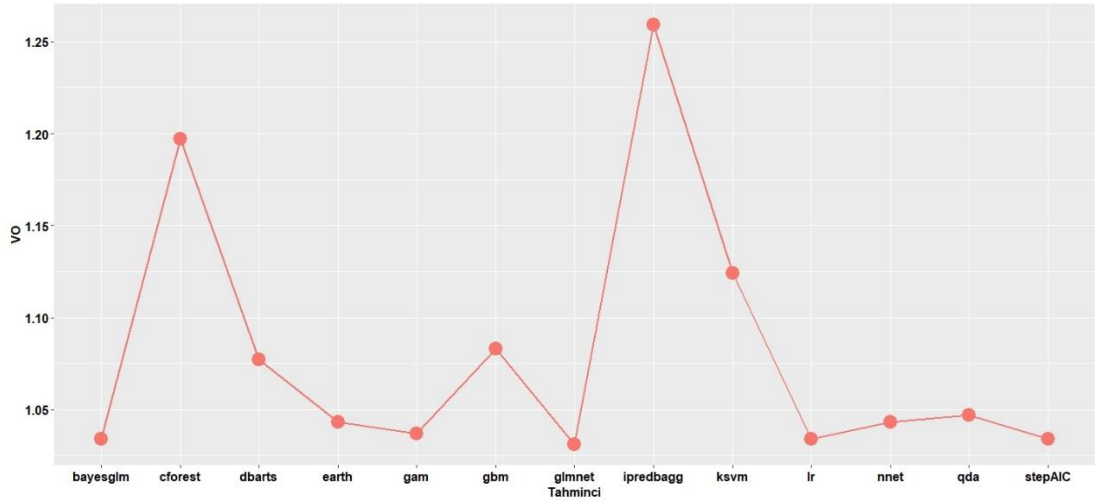
A senaryosu ve n=500 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği değerleri EK 2’de verilmiş ve mutlak ortalama değerleri Çizelge 3.2.2’de özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.2.3, Şekil 3.2.4 ve Şekil 3.2.5’te gösterilmiştir.

Çizelge 3.2.2. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo A; n=500)

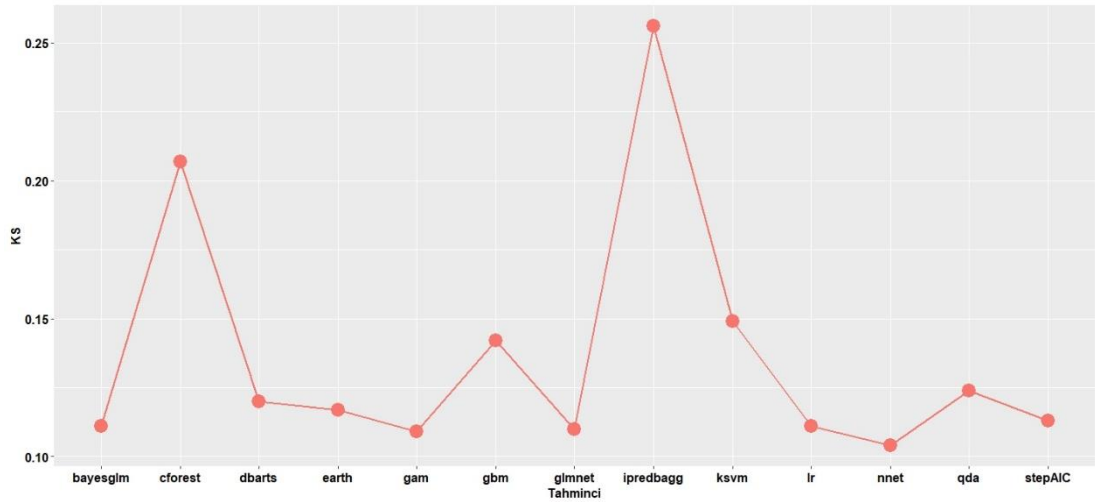
	SOF	OF	VO	KS
LR	0,005	0,005	1,034	0,111
KSVM	0,039	0,025	1,124	0,149
BAYESGLM	0,004	0,003	1,034	0,111
GLMNET	0,010	0,007	1,031	0,110
NNET	0,023	0,016	1,043	0,104
DBARTS	0,045	0,026	1,077	0,120
QDA	0,009	0,008	1,047	0,124
GAM	0,004	0,003	1,037	0,109
EARTH	0,028	0,020	1,043	0,117
STPAIC	0,025	0,017	1,034	0,113
IPREDBAGG	0,147	0,115	1,259	0,256
GBM	0,073	0,055	1,083	0,142
CFOREST	0,127	0,092	1,197	0,207



Şekil 3.2.3. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo A, n=500)



Şekil 3.2.4. Tahmincilere göre VO denge karşılaştırmaları (Senaryo A, n=500)



Şekil 3.2.5. Tahmincilere göre KS denge karşılaştırmaları (Senaryo A, n=500)

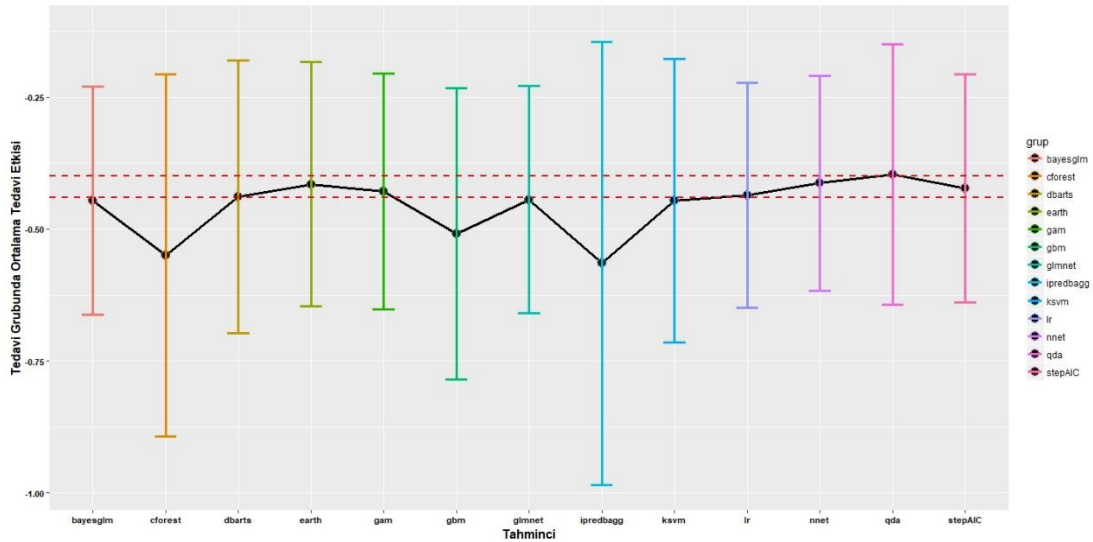
Çizelge 3.2.1, Çizelge 3.2.2, Şekil 3.2.1, Şekil 3.2.2, Şekil 3.2.3, Şekil 3.2.4 ve Şekil 3.2.5 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. IPREDBAGG, GBM ve CFOREST algoritmaları ise yanlışlık ve SOF denge değerlerinden dolayı aday algoritmalar arasına dâhil edilmemiştir.

B senaryosu ve n=500 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.2.3’ te verilmiştir. Denge sonuçlarına göre yanlışlık değeri LR’nin yanlışlık değeri olan %9,087’den daha az olan NNET, QDA, GAM, EARTH ve STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

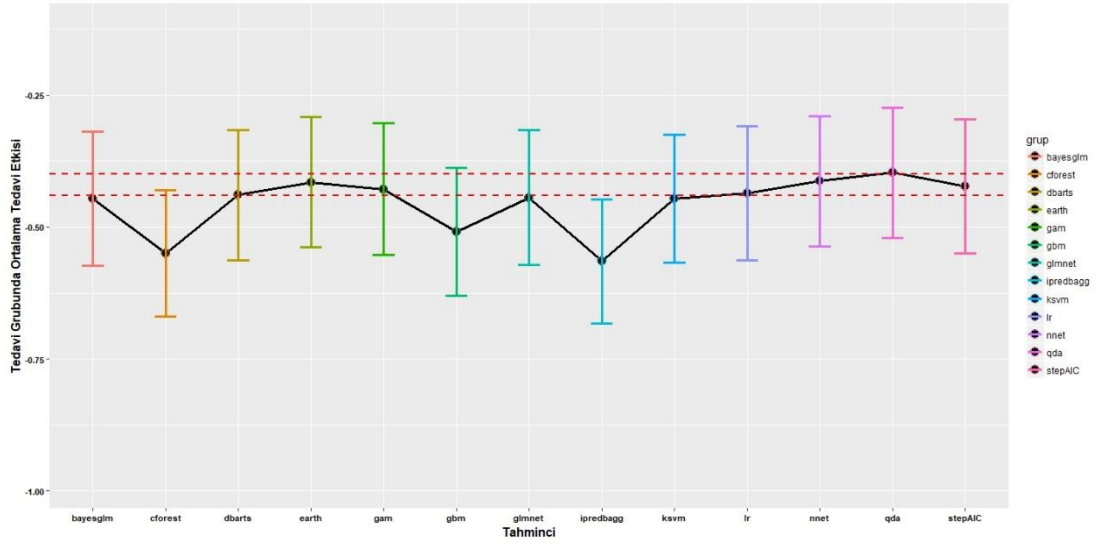
Çizelge 3.2.3. ATT için Denge sonuçları (Senaryo B; n=500)

	ATT	AI SH	SH	YANLILIK (%)
LR	-0,436	0,109	0,065	9,087
KSVM	-0,446	0,137	0,062	11,496
BAYESGLM	-0,446	0,110	0,065	11,536
GLMNET	-0,444	0,110	0,065	10,957
NNET	-0,413	0,104	0,063	3,258
DBARTS	-0,439	0,132	0,063	9,740
QDA	-0,397	0,126	0,063	-0,859
GAM	-0,428	0,114	0,064	6,876
EARTH	-0,415	0,118	0,063	3,783
STEPAIC	-0,423	0,110	0,065	5,692
IPREDBAGG	-0,565	0,214	0,060	41,275
GBM	-0,509	0,141	0,062	27,174
CFOREST	-0,550	0,175	0,061	37,607

Ayrıca ATT’ler AI standart hataları ile birlikte Şekil 3.2.6 ve klasik standart hatalar ile birlikte Şekil 3.2.7’de karşılaştırılmıştır.



Şekil 3.2.6. Tahmincilerle göre ATT ve AI standart hata ile güven aralıkları (Senaryo B, n=500)

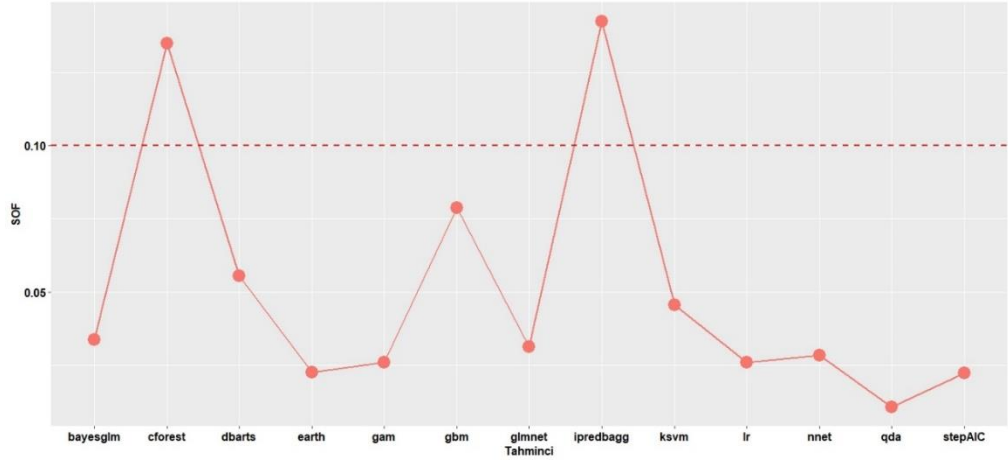


Şekil 3.2.7. Tahmincilerde göre ATT ve klasik standart hata ile güven aralıkları
(Senaryo B, n=500)

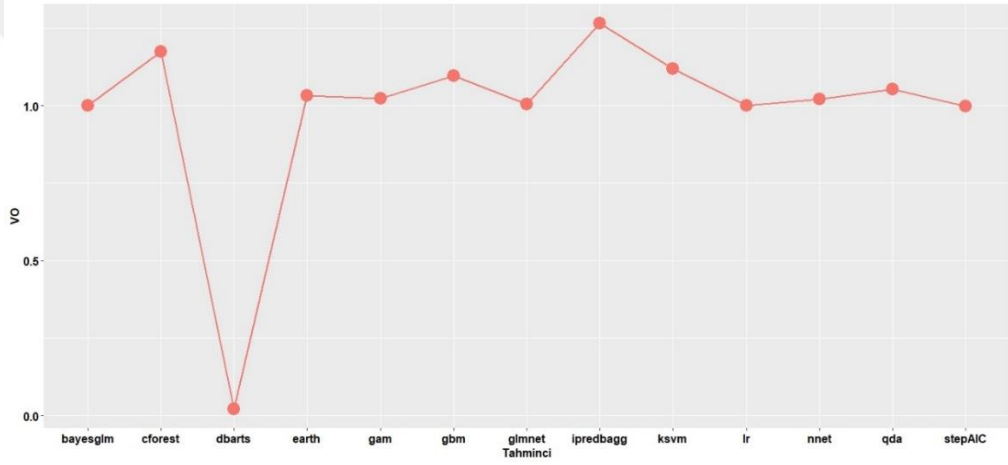
B senaryosu ve n=500 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği değerleri EK 3'te verilmiş ve mutlak ortalama değerleri Çizelge 3.2.4'te özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.2.8, Şekil 3.2.9 ve Şekil 3.2.10'da gösterilmiştir.

Çizelge 3.2.4. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo B; n=500)

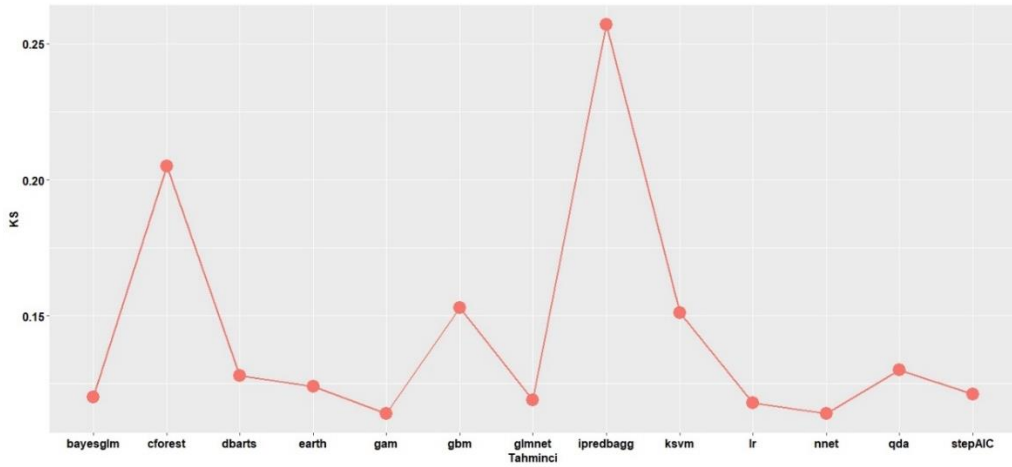
	SOF	OF	VO	KS
LR	0,026	0,018	1,001	0,118
K SVM	0,045	0,029	1,120	0,151
BAYESGLM	0,034	0,025	1,001	0,120
GLMNET	0,031	0,021	1,004	0,119
NNET	0,028	0,020	1,020	0,114
DBARTS	0,055	0,033	0,021	0,128
QDA	0,011	0,008	1,054	0,130
GAM	0,026	0,017	1,024	0,114
EARTH	0,023	0,016	1,033	0,124
STAPAIC	0,022	0,015	0,999	0,121
IPREDBAGG	0,142	0,110	1,265	0,257
GBM	0,079	0,061	1,097	0,153
CFOREST	0,135	0,094	1,175	0,205



Şekil 3.2.8. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo B, n=500)



Şekil 3.2.9. Tahmincilere göre VO denge karşılaştırmaları (Senaryo B, n=500)



Şekil 3.2.10. Tahmincilere göre KS denge karşılaştırmaları (Senaryo B, n=500)

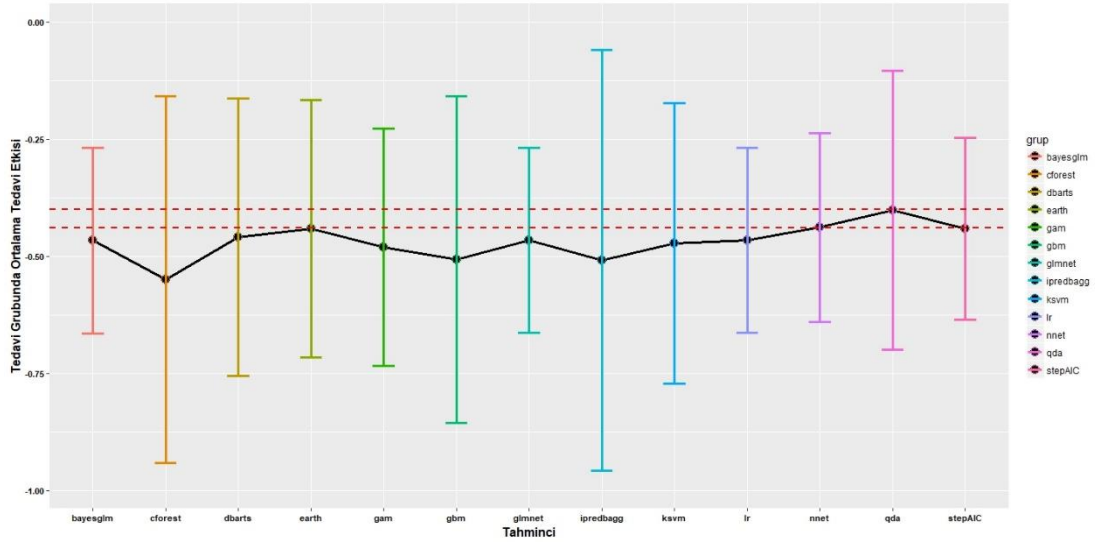
Çizelge 3.2.3, Çizelge 3.2.4, Şekil 3.2.6, Şekil 3.2.7, Şekil 3.2.8, Şekil 3.2.9 ve Şekil 3.2.10 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre NNET, QDA, GAM, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. Diğer algoritmalar ise yanlışlık ve SOF denge değerlerinden dolayı aday algoritmalar arasına dahil edilmemiştir.

C senaryosu ve n=500 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.2.5'te verilmiştir. Denge sonuçlarına göre yanlışlık değeri LR'nin yanlışlık değeri olan %16,52'den daha az olan NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

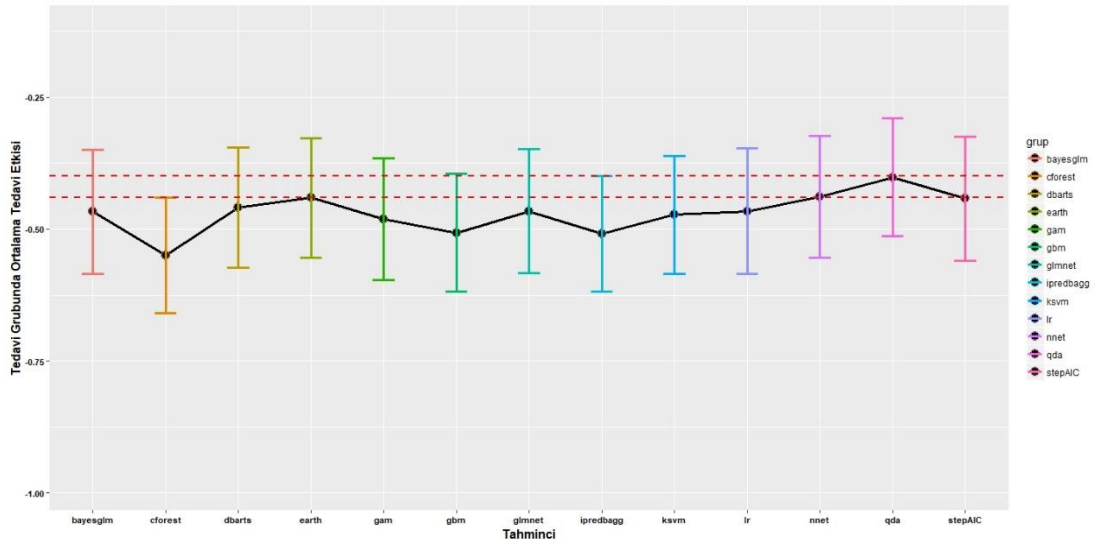
Çizelge 3.2.5. ATT için Denge sonuçları (Senaryo C; n=500)

	ATT	AI SH	SH	YANLIKLİK (%)
LR	-0,466	0,101	0,061	16,523
KSVM	-0,473	0,153	0,057	18,185
BAYESGLM	-0,467	0,101	0,060	16,749
GLMNET	-0,466	0,101	0,060	16,539
NNET	-0,439	0,103	0,059	9,847
DBARTS	-0,459	0,151	0,058	14,868
QDA	-0,402	0,152	0,057	0,459
GAM	-0,481	0,129	0,059	20,126
EARTH	-0,441	0,140	0,058	10,199
STEPAIC	-0,442	0,099	0,060	10,552
IPREDBAGG	-0,509	0,229	0,056	27,198
GBM	-0,507	0,178	0,057	26,759
CFOREST	-0,550	0,200	0,056	37,485

Ayrıca ATT'ler AI standart hataları ile birlikte Şekil 3.2.11 ve klasik standart hatalar ile birlikte Şekil 3.2.12'de karşılaştırılmıştır.



Şekil 3.2.11. Tahmincilere göre ATT ve AI standart hata ile güven aralıkları
(Senaryo C, n=500)

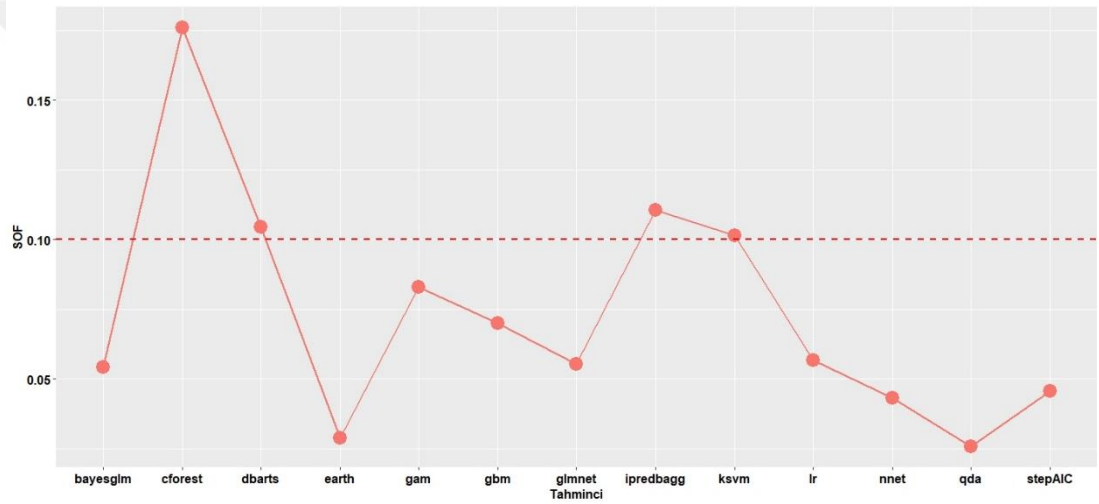


Şekil 3.2.12. Tahmincilere göre ATT ve klasik standart hata ile güven aralıkları
(Senaryo C, n=500)

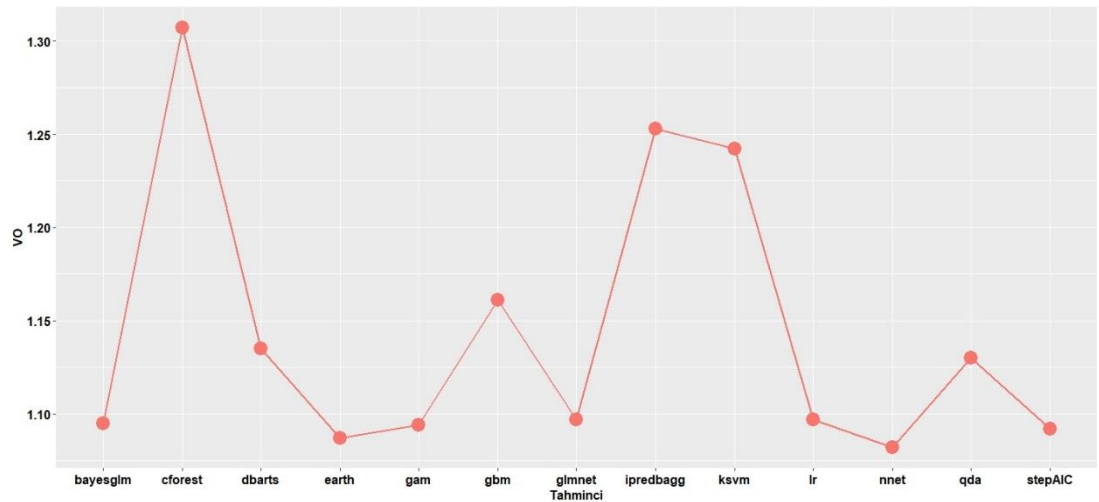
C senaryosu ve n=500 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği değerleri EK 4'te verilmiş ve mutlak ortalama değerleri Çizelge 3.2.6'da özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.2.13, Şekil 3.2.14 ve Şekil 3.2.15'te gösterilmiştir.

Çizelge 3.2.6. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo C; n=500)

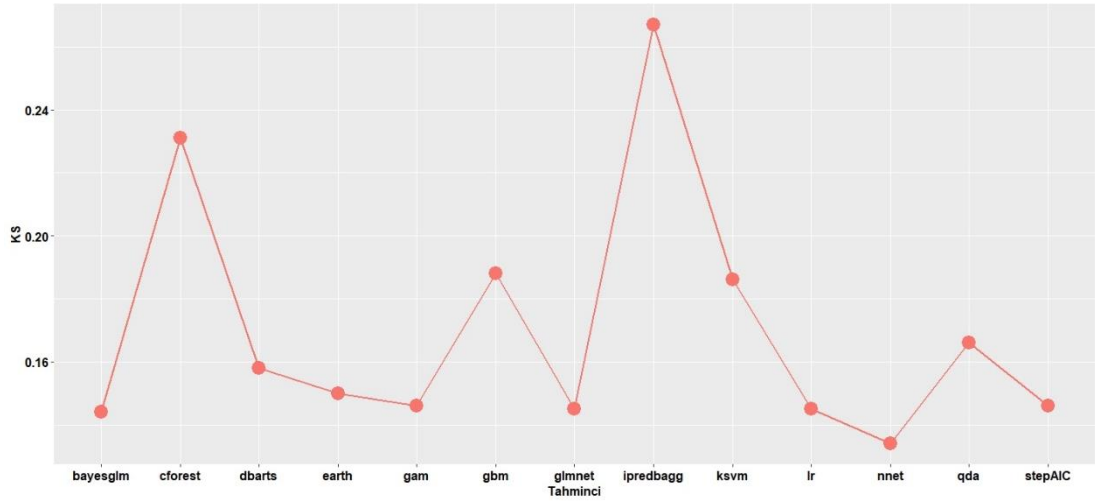
	SOF	OF	VO	KS
LR	0,057	0,045	1,097	0,145
KSVM	0,101	0,069	1,242	0,186
BAYESGLM	0,054	0,043	1,095	0,144
GLMNET	0,055	0,044	1,097	0,145
NNET	0,043	0,029	1,082	0,134
DBARTS	0,105	0,067	1,135	0,158
QDA	0,026	0,019	1,130	0,166
GAM	0,083	0,058	1,094	0,146
EARTH	0,029	0,023	1,087	0,150
STPAIC	0,046	0,037	1,092	0,146
IPREDBAGG	0,111	0,086	1,253	0,267
GBM	0,070	0,050	1,161	0,188
CFOREST	0,176	0,113	1,307	0,231



Şekil 3.2.13. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo C, n=500)



Şekil 3.2.14. Tahmincilere göre VO denge karşılaştırmaları (Senaryo C, n=500)



Şekil 3.2.15. Tahmincilere göre KS denge karşılaştırmaları (Senaryo C, n=500)

Çizelge 3.2.5, Çizelge 3.2.6, Şekil 3.2.11, Şekil 3.2.12, Şekil 3.2.13, Şekil 3.2.14 ve Şekil 3.2.15 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre NNET, QDA, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. Diğer algoritmalar ise yanlışlık ve SOF değerlerinden dolayı aday algoritmalar arasına dahil edilmemiştir.

Üç senaryoya göre de başarılı olarak belirlenen ve kombine tahminci içerisinde yer alacak olan algoritmalar Çizelge 3.2.7’de belirlenmiştir.

Çizelge 3.2.7. A, B ve C senaryolarına göre başarılı olan algoritmalar

	A	B	C	KOMBİNE
K SVM	+	-	-	-
BAYESGLM	+	-	-	-
GLMNET	+	-	-	-
NNET	+	+	+	+
DBARTS	+	-	-	-
QDA	+	+	+	+
GAM	+	+	-	-
EARTH	+	+	+	+
STEPAIC	+	+	+	+
IPREDBAGG	-	-	-	-
GBM	-	-	-	-
CFOREST	-	-	-	-

Çizelge 3.2.7’ye göre tüm senaryolarda etkin olan NNET, QDA, EARTH ve STEPAIC algoritmaları yeni tahminci model için aday algoritma olarak belirlenmiştir.

PS tahmininde oluşturulacak yeni tahminci kombine algoritma içerisine LR ve bu 4 algoritma dahil edilerek tekrar bir benzetim çalışması yapılmıştır. A, B ve C senaryoları kullanılarak LR ve kombine tahminci ile tekrar 1000 benzetim ile elde edilen PS'ler kullanılarak eşleştirme analizi yapılmış ve elde edilen denge sonuçları sırasıyla Çizelge 3.2.8, Çizelge 3.2.9 ve Çizelge 3.2.10'da verilmiştir.

Çizelge 3.2.8. LR ve kombine tahminci için denge sonuçları (Senaryo A; n=500)

	LR				Kombine tahminci			
ATT	-0,401				-0,411			
AI SH	0,108				0,112			
SH	0,061				0,061			
YANLILIK	% 0,276				% 2,814			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	0,004	0,002	1,012	-	-0,015	-0,007	1,02	-
W ₂	-0,008	-0,008	1,062	0,110	<0,001	0	1,072	0,116
W ₃	0,005	0,002	1,012	-	0,009	0,005	1,017	-
W ₄	-0,001	-0,001	1,061	0,111	0,010	0,009	1,083	0,115
W ₅	-0,001	<0,001	1,013	-	0,013	0,007	1,022	-
W ₆	-0,005	-0,003	1,011	-	0,003	0,002	1,017	-
W ₇	0,017	0,017	1,090	0,108	-0,013	-0,012	1,112	0,111
W ₈	-0,003	-0,001	1,014	-	-0,004	-0,002	1,017	-
W ₉	<0,001	<0,001	1,013	-	0,014	0,007	1,018	-
W ₁₀	-0,008	-0,008	1,046	0,113	-0,005	-0,005	1,062	0,118
Mutlak Ortalama	0,005	0,004	1,033	0,111	0,009	0,006	1,044	0,115

Çizelge 3.2.9. LR ve kombine tahminci için denge sonuçları (Senaryo B; n=500)

	LR				Kombine tahminci			
ATT	-0,440				-0,434			
AI SH	0,110				0,119			
SH	0,065				0,064			
YANLILIK	% 9,914				% 8,552			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	0,041	0,020	1,002	-	0,006	0,003	1,016	-
W ₂	0,084	0,074	0,681	0,141	0,071	0,063	0,822	0,135
W ₃	0,038	0,019	1,003	-	0,012	0,006	1,017	-
W ₄	0,024	0,024	1,086	0,110	0,041	0,040	1,106	0,120
W ₅	0,017	0,009	1,020	-	0,046	0,023	1,038	-
W ₆	0,024	0,012	1,025	-	0,050	0,024	1,046	-
W ₇	0,035	0,034	1,094	0,111	-0,018	-0,017	1,122	0,119
W ₈	0,004	0,002	1,015	-	0,008	0,004	1,021	-
W ₉	0,019	0,010	1,018	-	0,038	0,019	1,032	-
W ₁₀	-0,002	-0,002	1,051	0,113	0,001	0,001	1,067	0,124
Mutlak Ortalama	0,029	0,021	1,000	0,119	0,029	0,020	1,029	0,125

Çizelge 3.2.10. LR ve kombine tahminci için denge sonuçları (Senaryo C; n=500)

	LR				Kombine tahminci			
ATT	-0,467				-0,446			
AI SH	0,101				0,150			
SH	0,060				0,058			
YANLILIK	% 16,856				% 11,478			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	0,037	0,018	0,998	-	-0,107	-0,053	1,092	-
W ₂	0,073	0,066	0,735	0,124	0,094	0,086	1,118	0,168
W ₃	0,037	0,018	1,001	-	-0,084	-0,041	1,077	-
W ₄	0,160	0,140	0,593	0,154	0,089	0,078	1,091	0,166
W ₅	0,001	0,001	1,011	-	0,041	0,020	1,056	-
W ₆	0,045	0,022	1,026	-	0,092	0,045	1,087	-
W ₇	0,120	0,137	2,529	0,196	0,027	0,030	1,720	0,159
W ₈	-0,003	-0,001	1,011	-	-0,005	-0,002	1,048	-
W ₉	0,073	0,037	1,032	-	0,078	0,039	1,072	-
W ₁₀	<0,001	<0,001	1,061	0,104	-0,018	-0,018	1,170	0,161
Mutlak Ortalama	0,055	0,044	1,100	0,145	0,063	0,041	1,153	0,164

A, B ve C senaryoları kullanılarak LR ve kombine tahminci ile tekrar hesaplanan denge sonuçlarının özet karşılaştırmaları tek bir çizelge olarak Çizelge 3.2.11’de verilmiştir.

Çizelge 3.2.11. LR ve kombine tahminci için özet denge sonuçları (Senaryo A, B, C; n=500)

	A		B		C	
	LR	KOMBİNE	LR	KOMBİNE	LR	KOMBİNE
ATT	-0,401	-0,411	-0,440	-0,434	-0,467	-0,446
AI SH	0,108	0,112	0,110	0,119	0,101	0,150
SH	0,061	0,061	0,065	0,064	0,060	0,058
YANLILIK (%)	0,276	2,814	9,914	8,552	16,856	11,478
SOF Mutlak Ortalama	0,005	0,009	0,029	0,029	0,055	0,063
OF Mutlak Ortalama	0,004	0,006	0,021	0,020	0,044	0,041
VO Mutlak Ortalama	1,033	1,044	1,000	1,029	1,100	1,153
KS Mutlak Ortalama	0,111	0,115	0,119	0,125	0,145	0,164

Çizelge 3.2.11 incelendiğinde B ve özellikle C senaryosu için kombine tahminci ile elde edilen eşleştirme analizinde yanlılık değerlerinin azaldığı görülmektedir.

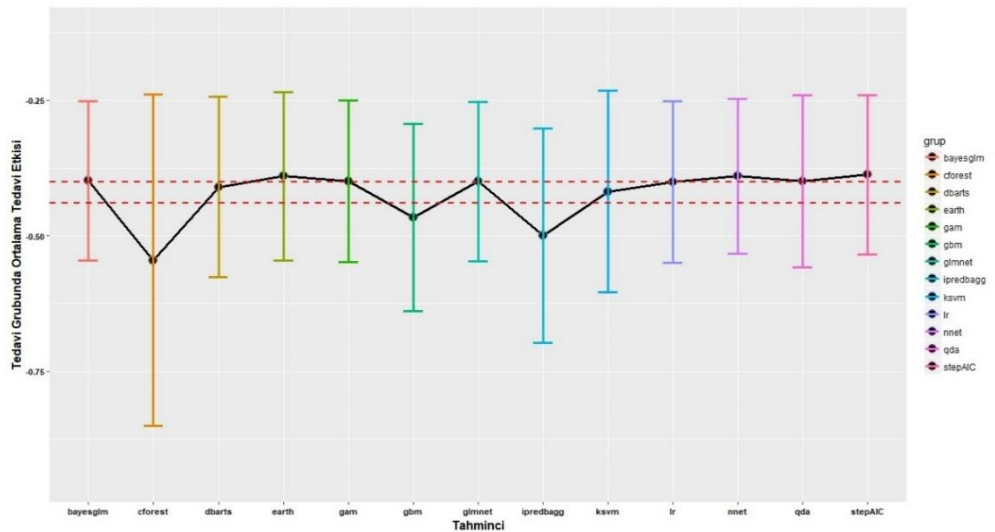
3.3. n= 1000 Örneklem Büyüklüğü için Benzetim Sonuçları

A senaryosu ve n=1000 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.3.1’de verilmiştir. Denge sonuçlarına göre LR ile birlikte yanlışlık değeri %10’dan daha az olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH, STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

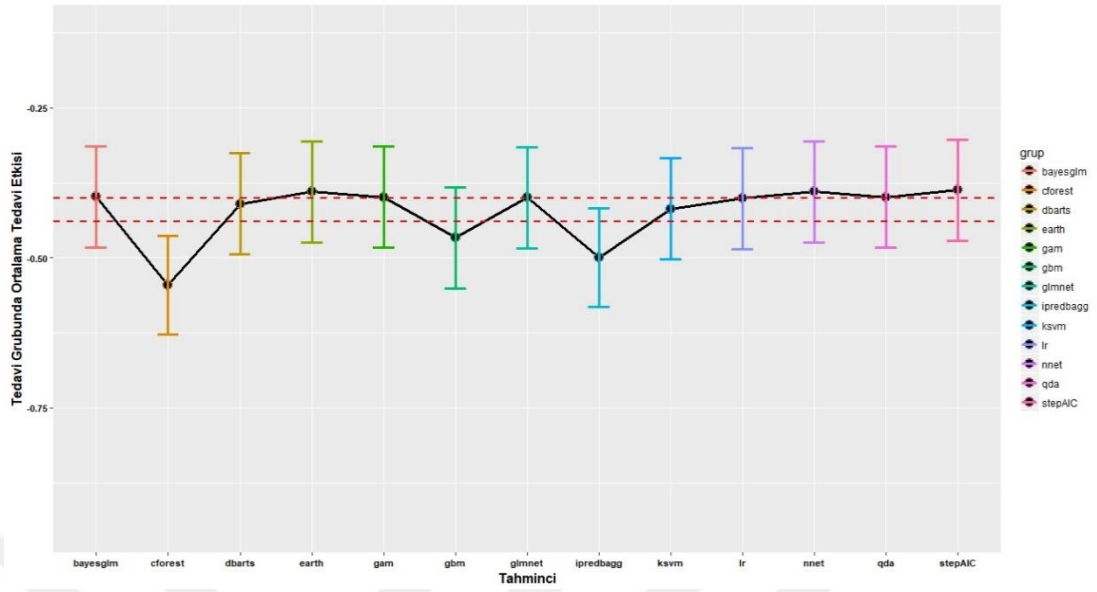
Çizelge 3.3.1. ATT için denge sonuçları (Senaryo A; n=1000)

	ATT	AI SH	SH	YANLILIK (%)
LR	-0,401	0,076	0,043	0,226
KSVM	-0,418	0,095	0,043	4,500
BAYESGLM	-0,398	0,075	0,043	-0,469
GLMNET	-0,3996	0,075	0,043	-0,101
NNET	-0,390	0,073	0,043	-2,612
DBARTS	-0,410	0,085	0,043	2,609
QDA	-0,399	0,081	0,043	-0,291
GAM	-0,399	0,076	0,043	-0,336
EARTH	-0,390	0,079	0,043	-2,483
STEPSAIC	-0,387	0,075	0,043	-3,143
IPREDBAGG	-0,499	0,101	0,042	24,72
GBM	-0,466	0,088	0,043	16,46
CFOREST	-0,545	0,156	0,042	36,17

Ayrıca ATT’ler AI standart hataları ile birlikte Şekil 3.3.1 ve klasik standart hatalar ile birlikte Şekil 3.3.2’de karşılaştırılmıştır.



Şekil 3.3.1. Tahmincilerle ilgili ATT ve AI standart hata ile güven aralıkları (Senaryo A, n=1000)

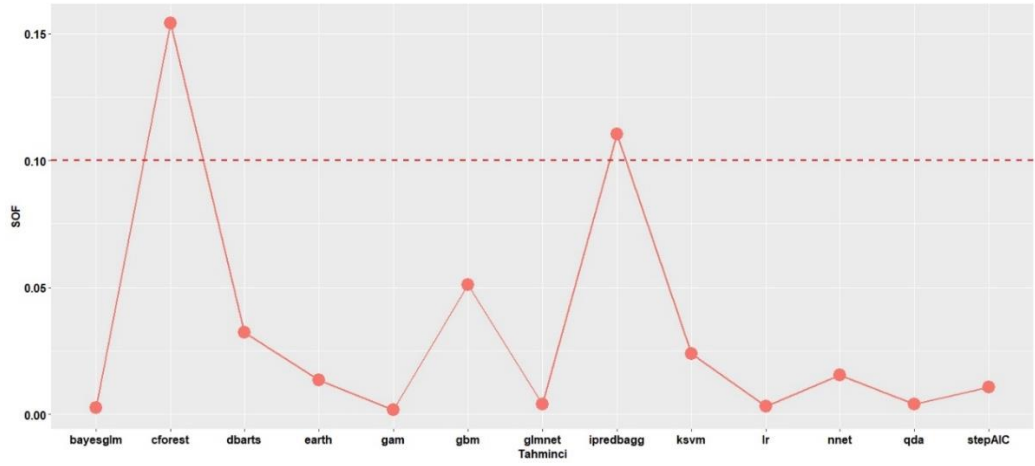


Şekil 3.3.2. Tahmincilerle ilgili ATT ve klasik standart hata ile güven aralıkları (Senaryo A, n=1000)

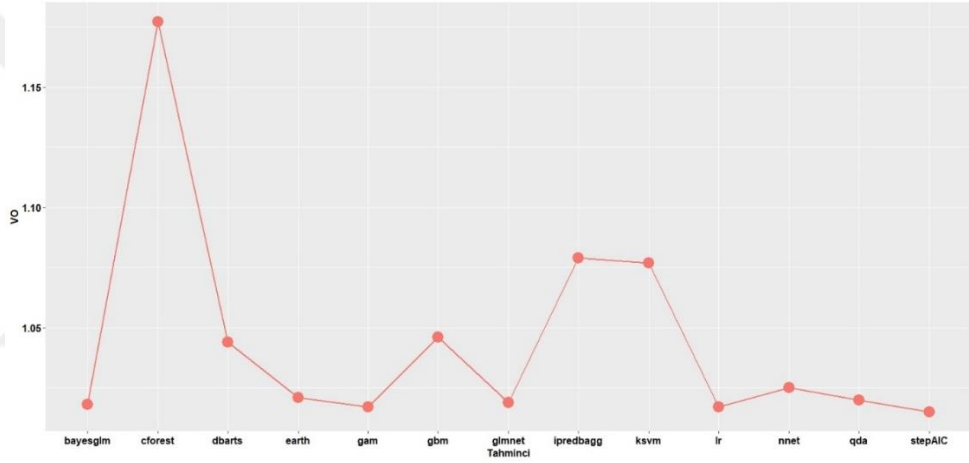
A senaryosu ve n=1000 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği değerleri EK 5’te verilmiş ve mutlak ortalama değerleri Çizelge 3.3.2’de özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.3.3, Şekil 3.3.4 ve Şekil 3.3.5’de gösterilmiştir.

Çizelge 3.3.2. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo A; n=1000)

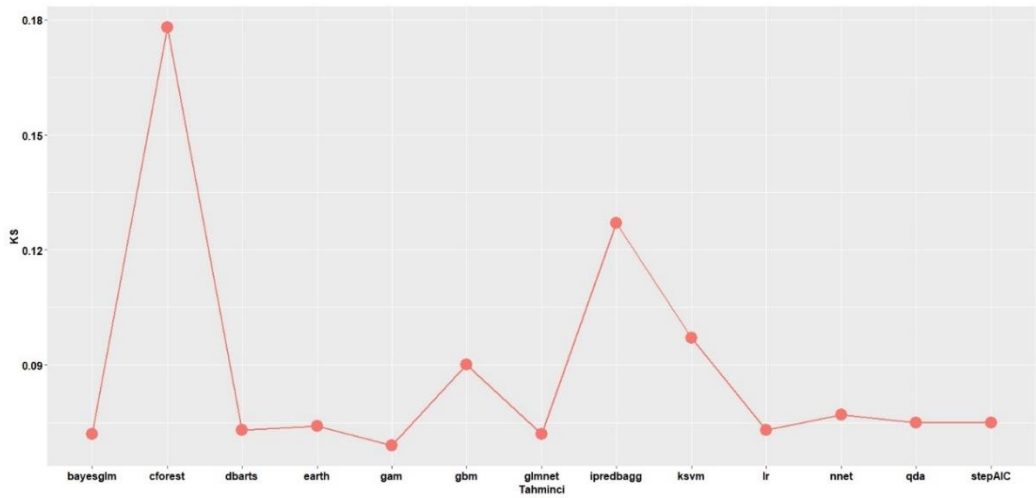
	SOF	OF	VO	KS
LR	0,003	0,003	1,017	0,073
KSVM	0,024	0,015	1,077	0,097
BAYESGLM	0,003	0,002	1,018	0,072
GLMNET	0,004	0,003	1,019	0,072
NNET	0,015	0,011	1,025	0,077
DBARTS	0,032	0,019	1,044	0,073
QDA	0,004	0,003	1,020	0,075
GAM	0,002	0,001	1,017	0,069
EARTH	0,014	0,010	1,021	0,074
STAPAIC	0,011	0,008	1,015	0,075
IPREDBAGG	0,110	0,078	1,079	0,127
GBM	0,051	0,040	1,046	0,090
CFOREST	0,154	0,111	1,177	0,178



Şekil 3.3.3. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo A, n=1000)



Şekil 3.3.4. Tahmincilere göre VO denge karşılaştırmaları (Senaryo A, n=1000)



Şekil 3.3.5. Tahmincilere göre KS denge karşılaştırmaları (Senaryo A, n=1000)

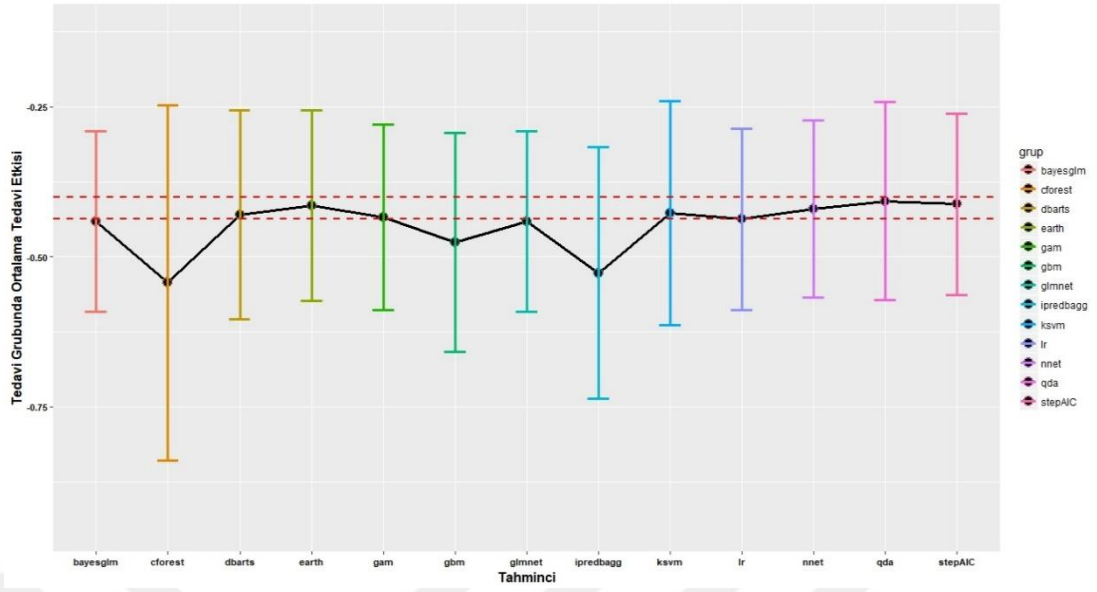
Çizelge 3.3.1, Çizelge 3.3.2, Şekil 3.3.1, Şekil 3.3.2, Şekil 3.3.3, Şekil 3.3.4 ve Şekil 3.3.5 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. IPREDBAGG, GBM, CFOREST algoritmaları ise yanlışlık ve SOF değerlerinden dolayı aday algoritmalar arasına dahil edilmemiştir.

B senaryosu ve n=1000 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.3.3'te verilmiştir. Denge sonuçlarına göre yanlışlık değeri LR'nin yanlışlık değeri olan %9,268'den daha az olan KSVM, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

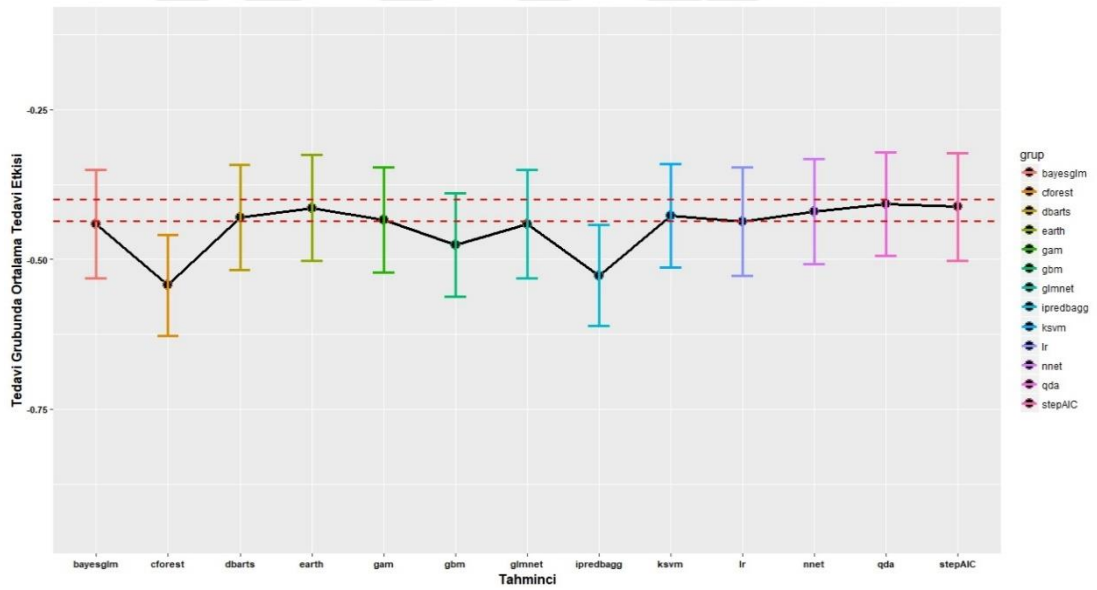
Çizelge 3.3.3. ATT için denge sonuçları (Senaryo B; n=1000)

	ATT	AI SH	SH	YANLILIK (%)
LR	-0.437	0.077	0.046	9.268
KSVM	-0.427	0.095	0.044	6.701
BAYESGLM	-0.441	0.077	0.046	10.248
GLMNET	-0.441	0.077	0.046	10.358
NNET	-0.420	0.075	0.045	4.924
DBARTS	-0.430	0.089	0.045	7.573
QDA	-0.407	0.084	0.044	1.640
GAM	-0.434	0.079	0.045	8.578
EARTH	-0.414	0.081	0.045	3.613
STEPSAIC	-0.412	0.077	0.046	2.964
IPREDBAGG	-0.527	0.107	0.043	31.630
GBM	-0.476	0.093	0.044	19.110
CFOREST	-0.543	0.151	0.043	35.792

Ayrıca ATT'ler AI standart hataları ile birlikte Şekil 3.3.6 ve klasik standart hatalar ile birlikte Şekil 3.3.7' de karşılaştırılmıştır.



Şekil 3.3.6. Tahmincilere göre ATT ve AI standart hata ile güven aralıkları (Senaryo B, n=1000)

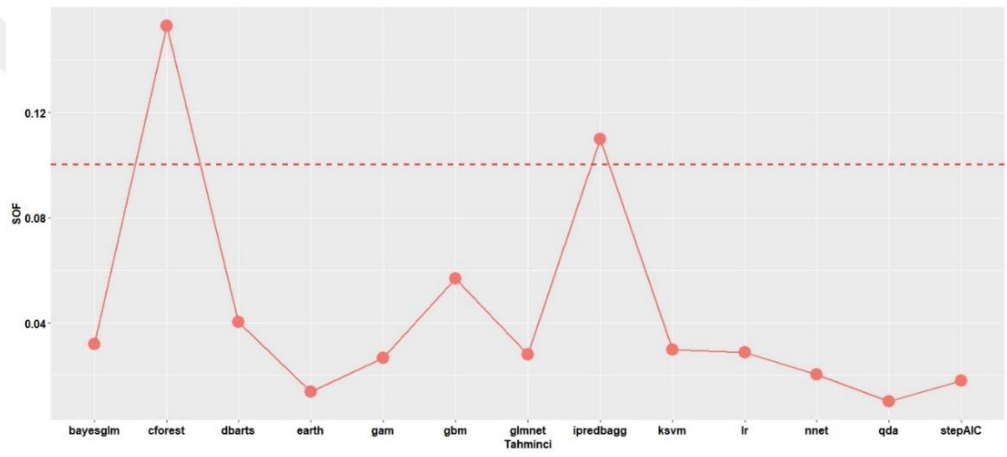


Şekil 3.3.7. Tahmincilere göre ATT ve klasik standart hata ile güven aralıkları (Senaryo B, n=1000)

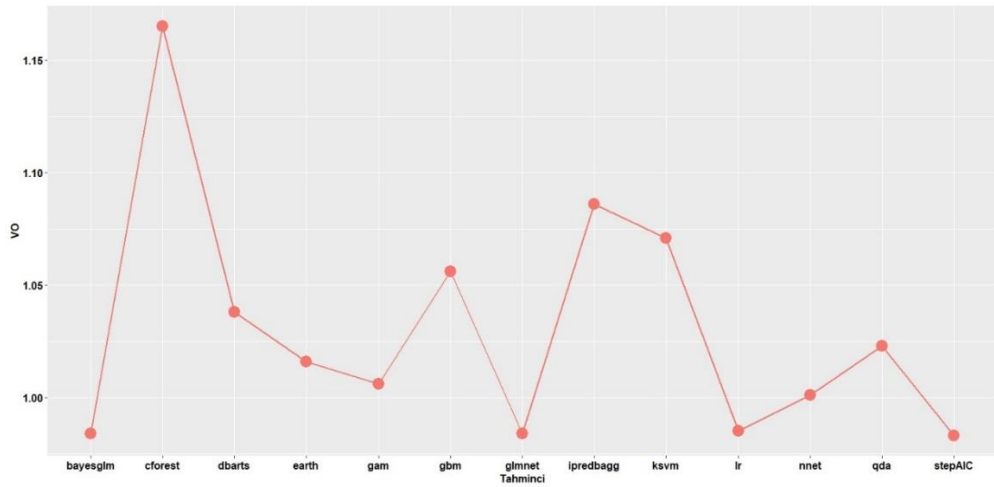
B senaryosu ve n=1000 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği değerleri EK 6'da verilmiş ve mutlak ortalama değerleri Çizelge 3.3.4'te özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.3.8, Şekil 3.3.9 ve Şekil 3.3.10'da gösterilmiştir.

Çizelge 3.3.4. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo B; n=1000)

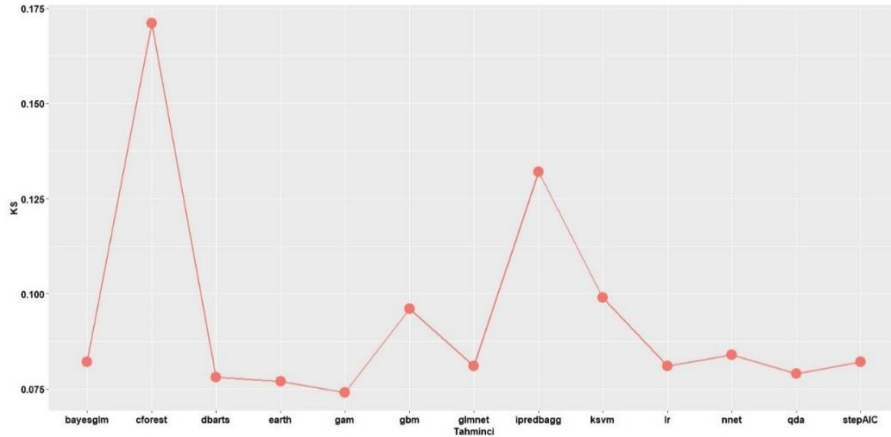
	SOF	OF	VO	KS
LR	0,029	0,021	0,985	0,081
KSVM	0,030	0,019	1,071	0,099
BAYESGLM	0,032	0,023	0,984	0,082
GLMNET	0,028	0,020	0,984	0,081
NNET	0,021	0,014	1,001	0,084
DBARTS	0,040	0,024	1,038	0,078
QDA	0,010	0,007	1,023	0,079
GAM	0,027	0,018	1,006	0,074
EARTH	0,014	0,010	1,016	0,077
STPAIC	0,018	0,012	0,983	0,082
IPREDBAGG	0,110	0,079	1,086	0,132
GBM	0,057	0,043	1,056	0,096
CFOREST	0,153	0,105	1,165	0,171



Şekil 3.3.8. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo B, n=1000)



Şekil 3.3.9. Tahmincilere göre VO denge karşılaştırmaları (Senaryo B, n=1000)



Şekil 3.3.10. Tahmincilere göre KS denge karşılaştırmaları (Senaryo B, n=1000)

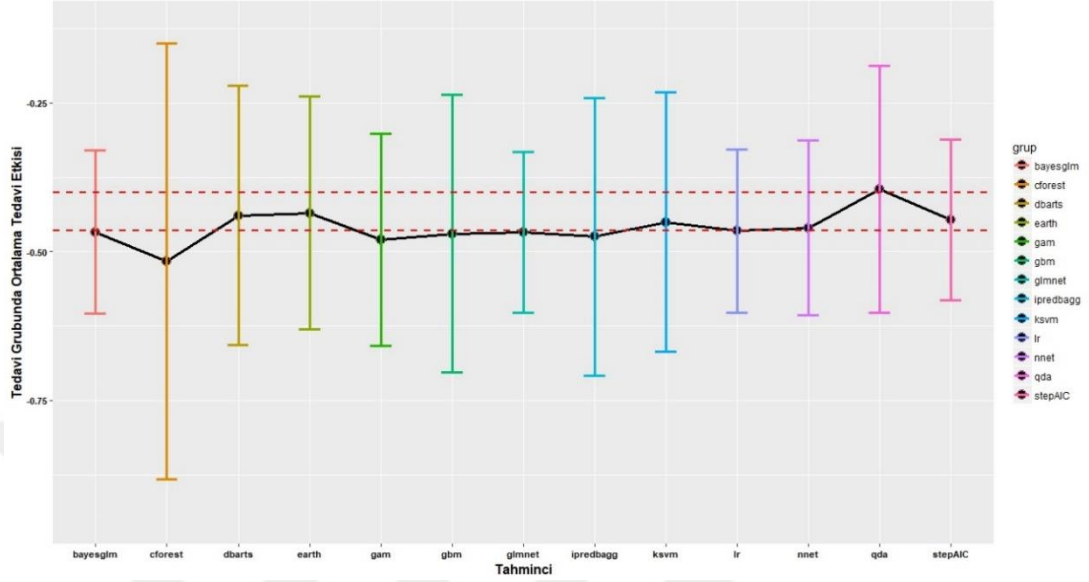
Çizelge 3.3.3, Çizelge 3.3.4, Şekil 3.3.6, Şekil 3.3.7, Şekil 3.3.8, Şekil 3.3.9 ve Şekil 3.3.10 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre KSVM, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. Diğer algoritmalar ise yanlışlık ve SOF değerlerinden dolayı aday algoritmalar arasına dahil edilmemiştir.

C senaryosu ve n=1000 örneklem büyüklüğü için 1000 benzetim kullanılarak LR ve LR ile uyumlu olarak belirlenen algoritmalar kullanılarak elde edilen ATT, standart hataları ve yanlışlık değerleri Çizelge 3.3.5'te verilmiştir. Denge sonuçlarına göre yanlışlık değeri LR'nin yanlışlık değeri olan %16,37'den daha az olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları kombine tahminci için başarılı algoritma olarak belirlenmiştir.

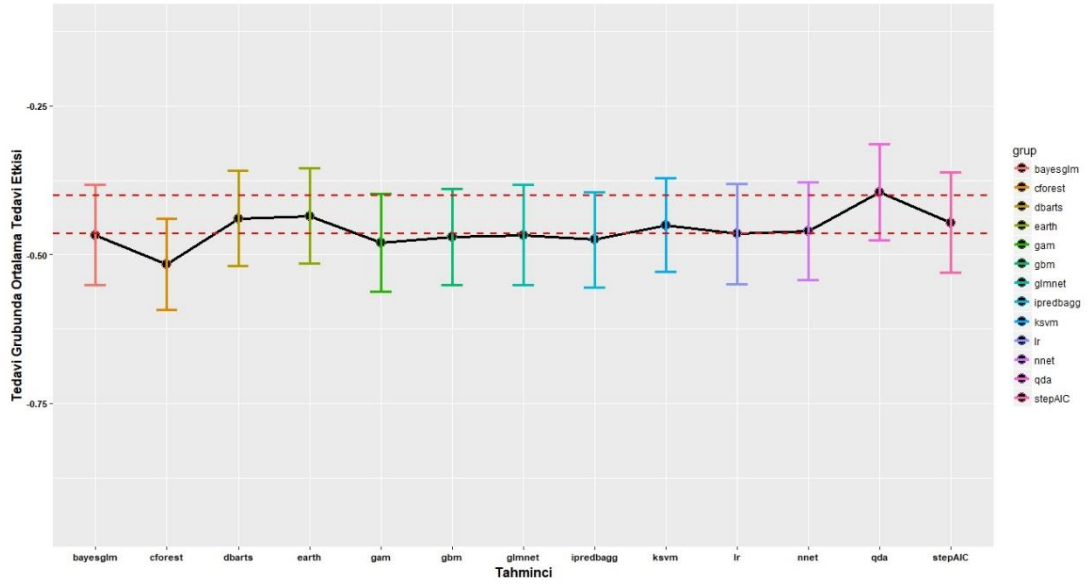
Çizelge 3.3.5. ATT için denge sonuçları (Senaryo C; n=1000)

	ATT	AI SH	SH	YANLILIK (%)
LR	-0,465	0,070	0,043	16,37
KSVM	-0,450	0,111	0,040	12,41
BAYESGLM	-0,467	0,070	0,043	16,72
GLMNET	-0,467	0,069	0,043	16,74
NNET	-0,460	0,075	0,042	14,89
DBARTS	-0,439	0,111	0,041	9,715
QDA	-0,395	0,106	0,041	-1,344
GAM	-0,480	0,091	0,042	19,98
EARTH	-0,435	0,100	0,041	8,758
STEPSAIC	-0,446	0,069	0,043	11,622
IPREDBAGG	-0,475	0,119	0,041	18,77
GBM	-0,470	0,119	0,041	17,40
CFOREST	-0,516	0,187	0,039	28,99

Ayrıca ATT'ler AI standart hataları ile birlikte Şekil 3.3.11 ve klasik standart hatalar ile birlikte Şekil 3.3.12' de karşılaştırılmıştır.



Şekil 3.3.11. Tahmincilerle ilgili ATT ve AI standart hata ile güven aralıkları (Senaryo C, n=1000)



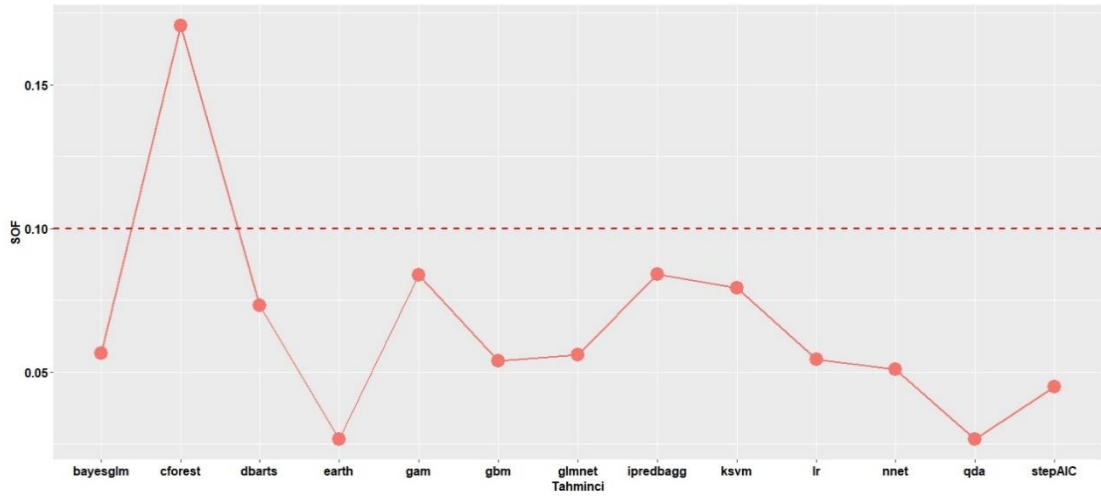
Şekil 3.3.12. Tahmincilerle ilgili ATT ve klasik standart hata ile güven aralıkları (Senaryo C, n=1000)

C senaryosu ve n=1000 örneklem büyüklüğü için LR ve diğer algoritmalar ile her bir ortak değişken/faktör için elde edilen SOF, OF, VO ve KS test istatistiği

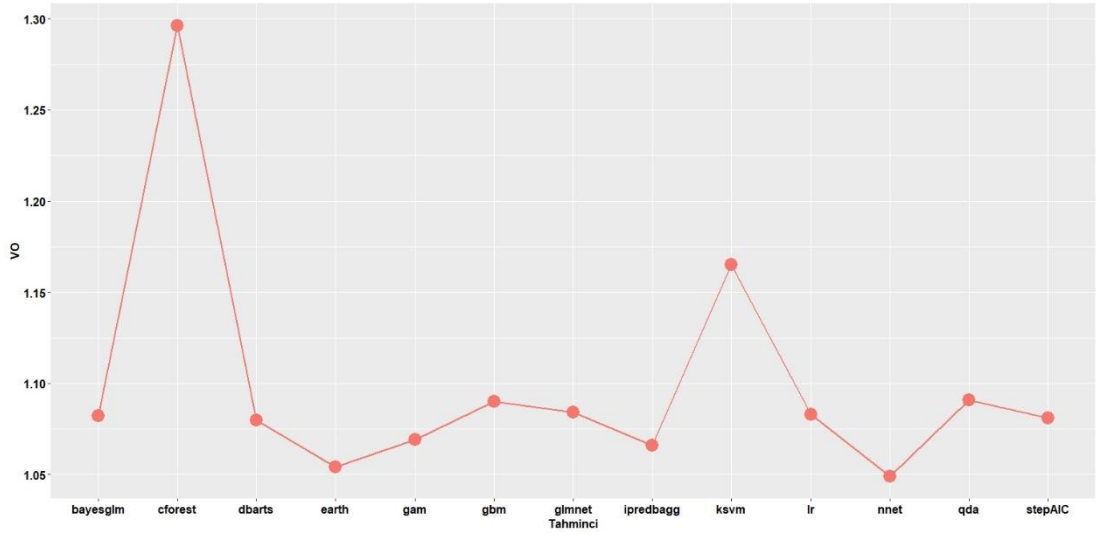
değerleri EK 7’de verilmiş ve mutlak ortalama değerleri Çizelge 3.3.6’da özetlenmiştir. Görsel olarak bireysel karşılaştırmaları ise Şekil 3.3.13, Şekil 3.3.14 ve Şekil 3.3.15’te gösterilmiştir.

Çizelge 3.3.6. Ortak değişken/faktörlerin ortalama denge sonuçları (Senaryo C; n=1000)

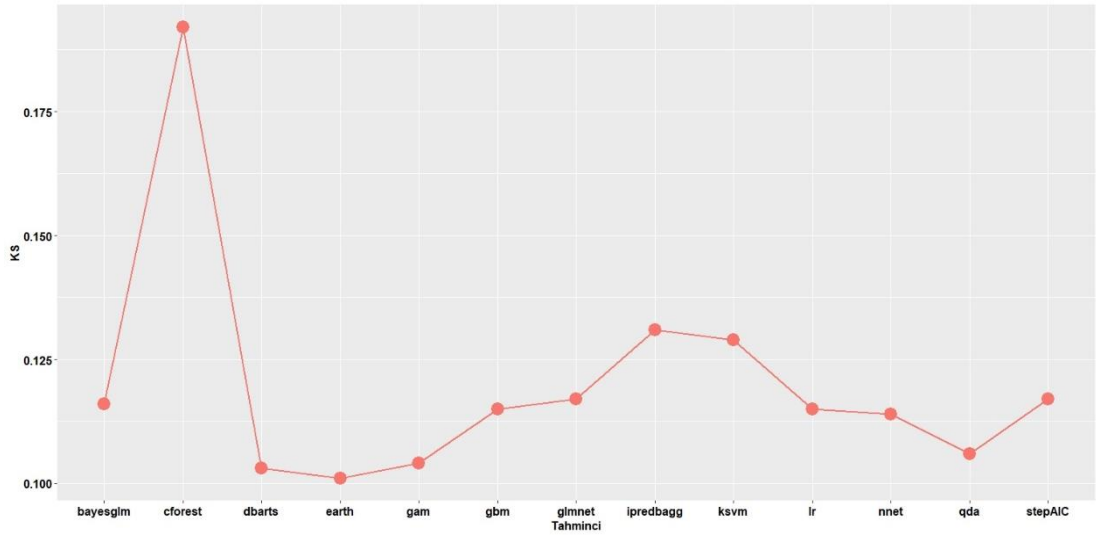
	SOF	OF	VO	KS
LR	0,055	0,044	1,083	0,115
KSVM	0,079	0,053	1,165	0,129
BAYESGLM	0,057	0,045	1,082	0,116
GLMNET	0,056	0,045	1,084	0,117
NNET	0,051	0,037	1,049	0,114
DBARTS	0,073	0,046	1,080	0,103
QDA	0,027	0,019	1,091	0,106
GAM	0,084	0,058	1,069	0,104
EARTH	0,027	0,022	1,054	0,101
STEPAIC	0,045	0,037	1,081	0,117
IPREDBAGG	0,084	0,062	1,066	0,131
GBM	0,054	0,037	1,090	0,115
CFOREST	0,171	0,105	1,296	0,192



Şekil 3.3.13. Tahmincilere göre SOF denge karşılaştırmaları (Senaryo C, n=1000)



Şekil 3.3.14. Tahmincilere göre VO denge karşılaştırmaları (Senaryo C, n=1000)



Şekil 3.3.15. Tahmincilere göre KS denge karşılaştırmaları (Senaryo C, n=1000)

Çizelge 3.3.5, Çizelge 3.3.6, Şekil 3.3.11, Şekil 3.3.12, Şekil 3.3.13, Şekil 3.3.14 ve Şekil 3.3.15 birlikte değerlendirilerek benzetim sonucundaki denge istatistiklerine göre KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları aday algoritma olarak belirlenmiştir. Diğer algoritmalar ise yanlışlık ve SOF değerlerinden dolayı aday algoritmalar arasına dahil edilmemiştir.

Üç senaryoda da başarılı olarak belirlenen ve kombine tahminci içerisinde yer alacak olan algoritmalar Çizelge 3.3.7’de gösterilmiştir.

Çizelge 3.3.7. A, B ve C senaryolarına göre başarılı olan algoritmalar

	A	B	C	Kombine
K SVM	+	+	+	+
BAYESGLM	+	-	-	-
GLMNET	+	-	-	-
NNET	+	+	+	+
DBARTS	+	+	+	+
QDA	+	+	+	+
GAM	+	+	-	-
EARTH	+	+	+	+
STEPSAIC	+	+	+	+
IPREDBAGG	-	-	-	-
GBM	-	-	-	-
CFOREST	-	-	-	-

Çizelge 3.3.7'ye göre tüm senaryolarda etkin olan K SVM, NNET, DBARTS, QDA, EARTH ve STEPSAIC algoritmaları yeni tahminci model için aday algoritma olarak belirlenmiştir. PS tahmininde oluşturulacak yeni tahminci kombine algoritma içerisine LR ve bu 6 algoritma dahil edilerek tekrar bir benzetim çalışması yapılmıştır. A, B ve C senaryoları kullanılarak LR ve kombine tahminci ile tekrar 1000 benzetim ile elde edilen PS'ler kullanılarak eşleştirme analizi yapılmış ve elde edilen denge sonuçları sırasıyla Çizelge 3.3.8, Çizelge 3.3.9 ve Çizelge 3.3.10'da verilmiştir.

Çizelge 3.3.8. LR ve kombine tahminci için denge sonuçları (Senaryo A; n=1000)

	LR				Kombine tahminci			
ATT	-0,398				-0,402			
AI SH	0,075				0,078			
SH	0,043				0,043			
YANLILIK	% 0,496				% 0,527			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	-0,002	-0,001	1,006	-	-0,016	-0,008	1,012	-
W ₂	-0,002	-0,003	1,026	0,072	<0,001	-0,001	1,044	0,074
W ₃	0,001	0,001	1,006	-	-0,012	-0,006	1,011	-
W ₄	-0,002	-0,002	1,039	0,072	0,008	0,008	1,055	0,072
W ₅	-0,001	-0,001	1,006	-	0,016	0,008	1,012	-
W ₆	<0,001	<0,001	1,006	-	0,005	0,002	1,009	-
W ₇	0,013	0,013	1,046	0,069	-0,012	-0,011	1,067	0,071
W ₈	-0,001	-0,001	1,007	-	-0,005	-0,002	1,008	-
W ₉	-0,004	-0,002	1,005	-	0,012	0,006	1,010	-
W ₁₀	0,006	0,006	1,016	0,073	0,001	0,001	1,032	0,076
Mutlak Ortalama	0,003	0,003	1,016	0,072	0,009	0,005	1,026	0,073

Çizelge 3.3.9. LR ve kombine tahminci için denge sonuçları (Senaryo B; n=1000)

	LR				Kombine tahminci			
ATT	-0,441				-0,430			
AI SH	0,077				0,084			
SH	0,046				0,045			
YANLILIK	% 10,160				% 7,405			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	0,037	0,018	0,994	-	-0,006	-0,003	1,011	-
W ₂	0,093	0,083	0,657	0,104	0,063	0,056	0,867	0,088
W ₃	0,033	0,016	0,997	-	-0,008	-0,004	1,012	-
W ₄	0,022	0,021	1,057	0,071	0,032	0,032	1,068	0,078
W ₅	0,017	0,008	1,012	-	0,052	0,026	1,031	-
W ₆	0,034	0,017	1,021	-	0,064	0,031	1,042	-
W ₇	0,035	0,034	1,066	0,072	-0,018	-0,017	1,089	0,076
W ₈	0,005	0,002	1,007	-	-0,004	-0,002	1,009	-
W ₉	0,021	0,011	1,013	-	0,037	0,018	1,021	-
W ₁₀	<0,001	<0,001	1,029	0,074	-0,001	-0,001	1,058	0,081
Mutlak Ortalama	0,030	0,021	0,985	0,080	0,029	0,019	1,021	0,081

Çizelge 3.3.10. LR ve kombine tahminci için denge sonuçları (Senaryo C; n=1000)

	LR				Kombine tahminci			
ATT	-0,471				-0,431			
AI SH	0,070				0,111			
SH	0,043				0,041			
YANLILIK	% 17,762				% 7,672			
	SOF	OF	VO	KS	SOF	OF	VO	KS
W ₁	0,034	0,017	0,993	-	-0,103	-0,051	1,067	-
W ₂	0,084	0,076	0,704	0,095	0,075	0,068	1,098	0,109
W ₃	0,041	0,020	0,993	-	-0,089	-0,044	1,055	-
W ₄	0,162	0,142	0,581	0,127	0,074	0,065	1,079	0,109
W ₅	-0,003	-0,001	1,004	-	0,049	0,025	1,033	-
W ₆	0,050	0,024	1,022	-	0,095	0,047	1,063	-
W ₇	0,122	0,139	2,473	0,175	0,031	0,035	1,426	0,102
W ₈	0,004	0,002	1,005	-	-0,014	-0,007	1,024	-
W ₉	0,075	0,037	1,026	-	0,074	0,037	1,046	-
W ₁₀	0,001	0,001	1,015	0,068	0,007	0,007	1,093	0,104
Mutlak Ortalama	0,058	0,046	1,082	0,116	0,061	0,039	1,098	0,106

A, B ve C senaryoları kullanılarak LR ve kombine tahminci ile tekrar hesaplanan denge sonuçlarının özet karşılaştırmaları tek bir çizelge olarak Çizelge 3.3.11’de verilmiştir.

Çizelge 3.3.11. LR ve kombine tahminci için özet denge sonuçları (Senaryo A, B, C; n=1000)

	A		B		C	
	LR	Kombine	LR	Kombine	LR	Kombine
ATT	-0,398	-0,402	-0,441	-0,430	-0,471	-0,431
AI SH	0,075	0,078	0,077	0,084	0,070	0,111
SH	0,043	0,043	0,046	0,045	0,043	0,041
YANLILIK (%)	-0,496	0,527	10,160	7,405	17,762	7,672
SOF Mutlak Ortalama	0,003	0,009	0,030	0,029	0,058	0,061
OF Mutlak Ortalama	0,003	0,005	0,021	0,019	0,046	0,039
VO Mutlak Ortalama	1,016	1,026	0,985	1,021	1,082	1,098
KS Mutlak Ortalama	0,072	0,073	0,080	0,081	0,116	0,106

Çizelge 3.3.11 incelendiğinde B ve özellikle C senaryosu için kombine tahminci ile elde edilen eşleştirme analizinde yanlılık değerlerinin azaldığı görülmektedir.

3.4. Lalonde Gerçek Veri Setine Ait Sonuçlar

Birinci benzetim çalışması (n=500) sonucunda NNET, QDA, EARTH ve STEPAIC algoritmaları, ikinci benzetim çalışması (n=1000) sonucunda KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir. Her iki örneklem büyüklüğünde başarılı olan NNET, QDA, EARTH ve STEPAIC algoritmaları, LR ve bu beş algoritmanın kombinasyonu ile oluşturulan kombine tahmin modeli ile Lalonde gerçek veri seti kullanılarak tekrar PS hesaplanmış ve bu PS'lere dayalı olarak eşleştirme analizleri yapılmıştır. Eşleştirme analizi sonrasında elde edilen denge değerleri Çizelge 3.4.1'de sunulmuştur.

Çizelge 3.4.1. Lalonde veri seti eşleştirme sonrası denge ölçümlerinin karşılaştırılması (n=185)

	Mutlak Toplam SOF	Ortalama VO	Ortalama KS	ATT	AI SH	P
LR	0,592	1,020	0,086	2624,3	802,1	0,001
NNET*	0,555	1,056	0,076	1843,4	726,1	0,011
QDA	0,710	0,844	0,060	1749,7	831,4	0,035
EARTH	0,549	0,953	0,065	1573,1	818,9	0,055
STEPAIC	0,464	1,013	0,074	1748,9	724,2	0,016
KOMBİNE*	0,316	1,017	0,033	1846,8	846,3	0,029

* tekrarlandığında algoritma seçeneklerine göre sonuçlar farklılık gösterebilir

En az yanlılıkla eşleştirme analizi için PS tahmini yapan algoritma 0,003 SOF değeri ile Kombine tahminci olmuştur. Mutlak SOF toplamına göre QDA hariç NNET, EARTH, STEPAIC ve Kombine tahminci LR'den daha başarılı PS tahmininde bulunmuştur. VO ortalamasına göre ise STEPAIC en başarılı algoritma olarak bulunmuştur. VO ortalamaları 1'e yakınlığına göre sırasıyla STEPAIC, Kombine, LR, EARTH, NNET ve QDA olmuştur. KS ortalamasına göre 0'a en yakın Kombine tahminci olmuştur. KS ortalamaları 0'a yakınlığına göre sırasıyla Kombine, QDA, EARTH, NNET, STEPAIC, NNET ve LR olmuştur. Başarılı algoritmaların tamamında KS istatistiği LR'den küçük olarak bulunmuştur. Algoritmalarından EARTH hariç tüm algoritmalar ile ATT istatistiksel olarak anlamlı farklı bulunmuştur (LR; $p=0.001$, NNET; $p=0.011$, QDA; $p=0.035$, STEPAIC; $p=0.016$, Kombine; $p=0.029$).

Eşleştirme analizi sonucunda tedavi grubu ve algoritmalara göre eşleşen kontrol grubu ortalamaları Çizelge 3.4.2'de ve eşleştirme sonrası tedavi ve kontrol grubu ortalamaları arasındaki farklar Çizelge 3.4.3'te verilmiştir.

Çizelge 3.4.2. Lalonde veri seti eşleştirme sonrası ortalamalar

	LR		NNET	QDA	EARTH	STEPAIC	KOMBİNE
Ortak değişken/faktör (Tipi)	T	K	K	K	K	K	K
Age (S)	25,82	25,69	25,82	26,46	25,64	25,33	25,72
Educ (S)	10,35	10,15	10,22	9,88	10,62	10,41	10,47
Black (K)	0,84	0,87	0,86	0,80	0,82	0,82	0,86
Hispan (K)	0,06	0,05	0,08	0,08	0,09	0,10	0,08
Married (K)	0,19	0,18	0,15	0,18	0,18	0,17	0,21
Nodegree (K)	0,71	0,77	0,79	0,76	0,70	0,71	0,71
Re74 (S)	2095,57	2193,33	2409,83	2144,75	2755,81	2088,76	2070,14
Re75 (S)	1532,06	2179,93	1448,98	1757,32	1682,30	1244,42	1631,70

T: tedavi, K: kontrol

Çizelge 3.4.3. Lalonde veri seti eşleştirme sonrası OF denge ölçümlerinin karşılaştırılması

Ortak değişken/ faktör (Tipi)	LR	NNET	QDA	EARTH	STEPAIC	KOMBİNE
Age (S)	0,13	<0,01	0,64	0,18	0,49	0,10
Educ (S)	0,20	0,13	0,47	0,27	0,06	0,12
Black (K)	0,03	0,02	0,04	0,02	0,02	0,02
Hispan (K)	0,01	0,02	0,02	0,03	0,04	0,02
Married (K)	0,01	0,04	0,01	0,01	0,02	0,02
Nodegree (K)	0,06	0,08	0,05	0,01	<0,01	<0,01
Re74 (S)	97,8	314,3	49,2	660,2	6,8	25,4
Re75 (S)	647,8	83,1	225,3	150,2	287,6	99,6

Eşleştirme analizi sonucunda algoritmalar için ortak değişkenlerin tamamına göre SOF denge ölçüm değerleri Çizelge 3.4.4'te karşılaştırılmıştır.

Çizelge 3.4.4. Lalonde veri seti eşleştirme sonrası SOF denge ölçümlerinin karşılaştırılması

Ortak değişken/ faktör (Tipi)	LR	NNET	QDA	EARTH	STEPSAIC	KOMBİNE
Age (S)	0,017	<0,001	0,089	0,025	0,068	0,014
Educ (S)	0,100	0,064	0,232	0,134	0,032	0,060
Black (K)	0,069	0,052	0,111	0,053	0,050	0,046
Hispan (K)	0,042	0,083	0,068	0,118	0,164	0,099
Married (K)	0,013	0,089	0,018	0,028	0,060	0,061
Nodegree (K)	0,130	0,177	0,111	0,010	<0,001	<0,001
Re74 (S)	0,020	0,064	0,010	0,135	0,001	0,005
Re75 (S)	0,201	0,026	0,070	0,047	0,089	0,031

SOF denge ölçümleri incelendiğinde age değişkeni için NNET, educ değişkeni için STEPSAIC, black değişkeni için Kombine, hispan ve married değişkeni için LR, nodegree değişkeni için Kombine ve STEPSAIC, re74 değişkeni için STEPSAIC, re75 değişkeni için NNET algoritması en başarılı eşleştirmeyi gerçekleştirmiştir. Ancak age, educ, re74 ve re75 sayısal değişkenler birlikte değerlendirildiğinde en dengeli eşleştirmeyi Kombine tahmincinin yaptığı görülmektedir.

Eşleştirme analizi sonucunda algoritmalar için ortak değişkenlerin tamamına göre VO ve KS denge ölçüm değerleri Çizelge 3.4.5 ve Çizelge 3.4.6'da karşılaştırılmıştır.

Çizelge 3.4.5. Lalonde veri seti eşleştirme sonrası VO denge ölçümlerinin karşılaştırılması

Ortak değişken/ faktör (Tipi)	LR	NNET	QDA	EARTH	STEPSAIC	KOMBİNE
Age (S)	1,083	0,988	0,762	1,157	1,092	1,252
Educ (S)	1,234	1,544	0,821	1,353	1,425	1,234
Black (K)	1,157	1,112	0,835	0,910	0,915	1,097
Hispan (K)	1,187	0,767	0,800	0,701	0,630	0,736
Married (K)	1,020	1,176	1,030	1,046	1,109	0,915
Nodegree (K)	1,158	1,241	1,129	0,991	1,000	1,000
Re74 (S)	0,750	0,657	0,826	0,584	0,764	0,930
Re75 (S)	0,565	0,963	0,547	0,878	1,167	0,970

Çizelge 3.4.6. Lalonde veri seti eşleştirme sonrası KS denge ölçümlerinin karşılaştırılması

Ortak değişken/ faktör (Tipi)	LR	NNET	QDA	EARTH	STEPSAIC	KOMBİNE
Age (S)	0,061	0,066	0,0784	0,070	0,065	0,070
Educ (S)	0,061	0,081	0,0954	0,123	0,071	0,019
Re74 (S)	0,122	0,043	0,0307	0,034	0,030	0,020
Re75 (S)	0,098	0,113	0,0341	0,031	0,127	0,023

Superlearner algoritması içerisinde tahmincilerin ağırlık katsayıları LR için 0, NNET için 0,118, QDA için 0,014, EARTH için 0,124 ve STEPSAIC için 0,744 olarak bulunmuştur.

4. TARTIŞMA

Randomize klinik çalışmaların yüksek maliyeti ve etik kurul izninde yaşanan zorluklar birçok araştırmacının maliyeti daha düşük olan gözlemsel çalışmalara yönelmesine neden olmaktadır. Bu nedenle PS özellikle tıp alanındaki istatistiksel analizlerde yaygın olarak kullanılmaktadır. PS gözlemsel çalışmalarda gerçek tedavi etkilerinin daha kesin tahmin edilmesini sağlamakla beraber, zamandan ve maliyetten de tasarruf etmeyi sağlar.

Deneyisel araştırmalarda araştırılan faktör dışındaki deney sonucunu etkileyebilecek etkenler araştırmacı tarafından sabit tutularak randomizasyon ile kontrol edilmeye çalışılır. Bu yüzden neden-sonuç ilişkisi doğru olarak saptanabilmektedir. Ancak randomize olmayan gözlemsel çalışmalarda bireylerin vaka ve kontrol gruplarına atanma olasılığı eşit olmadığı için karıştırıcı ortak değişkenler sıklıkla bulunmaktadır. Bu değişkenler göz ardı edilerek ulaşılan sonuçlarda yanlılık bulunmaktadır. Bu nedenle randomizasyonun yapılamadığı gözlemsel çalışmalarda PS'ye göre deney gruplarının eşleştirilmesi yanlılığı azaltmak ve tedavinin gerçek etkisini belirlemek için araştırmalarda göz ardı edilmemesi gereken bir yöntemdir. Ancak bibliyometrik analizler ile elde edilen sonuçlara göre özellikle yerli literatürde PS kullanımı pek fazla yaygınlaşmamış olup, Gürtekin (2008), Kaspar ve ark. (2010) ile Demir ve ark. (2017)'nin çalışmaları dışında PS konusunda güncel bir çalışma mevcut değildir.

Setoguchi ve ark. (2008) yaptıkları benzetim çalışmasında B (hafif doğrusal olmayan model) ve G (orta düzeyde doğrusal ve toplamsal olmayan model) senaryolarında LR'nin NNET'e göre daha yanlı tahminler ürettiği sonuçlarını raporlamışlardır. NNET'in maruziyet ile ortak değişkenler arasındaki ilişkileri model karmaşıklıkla daha doğru şekilde belirlediği sonucuna ulaşmışlardır. LR'nin sadece temel etkiler olan senaryoda yansız tahminler ile güçlü bir performansa sahip olduğunu ifade etmişlerdir.

Lee ve ark. (2010) yaptıkları benzetim çalışmasında LR'nin sadece temel etkiler olan modelde yeterli dengeyi sağladığını belirtmişlerdir. Farklı örneklem büyüklüğü ile belirlenen senaryoların tamamında bagged CART, random forests ve boosted CART'ın yüksek düzeyde performans sağladığını ve random forests ile boosted CART makine öğrenme yöntemlerinin PS tahmininde kullanımlarının yaygınlaştırılarak daha ileriye götürülmesi gerektiğini ifade etmişlerdir. Ayrıca makine öğrenme algoritmalarının büyük verilerin yanında $n=500$ örneklem büyüklüğünde de PS tahmininde kullanılabileceği sonucuna ulaşmışlardır. Tedavi grubuna atanma olasılığı ile ortak değişkenler arasındaki ilişkilerin bilinmediği karmaşık modellerin gerektiği çok boyutlu verilerde makine öğrenme yöntemlerinin PS tahmininde uygun bir görev üstleneceğini ifade etmişlerdir.

Pirracchio ve ark. (2014) yaptıkları çalışmada temel etkilerin bulunduğu senaryo dışında etkileşim ve doğrusal olmayan ilişkilerin de bulunduğu 3 farklı benzetim senaryosu kullanmışlardır. Çoğunluğu makine öğrenme algoritması olan 15 farklı algoritma ile Super Learner tahmincisi kullanarak elde ettikleri denge sonuçları ve LR ile elde edilen denge sonuçlarını karşılaştırmışlardır. Pirracchio ve ark. diğer araştırmacılara benzer olarak modelin doğru olarak belirlendiği tedavi grubuna atanma olasılığı ile ortak değişkenler arasında sadece temel etkilerin bulunduğu toplamsal modelde LR'nin Super Learner tahmincisine göre biraz daha yansız tahminler elde ederek daha başarılı olduğunu göstermişlerdir. Ancak etkileşim ve doğrusal olmayan karmaşık ilişkilerin olduğu modellerde Super Learner tahmincisinin LR'ye göre daha yansız tahminler ürettiği sonucuna ulaşmışlardır.

Literatürde PS yaygın olarak LR ile tahmin edilmektedir (Austin, 2014). Modelde temel etkiler olduğunda LR ile daha başarılı tahminlere ulaşıldığı ancak doğrusal olmayan ilişkilerde ve etkileşimlerin bulunduğu modellerde ML yaklaşımları ile daha başarılı tahminler elde edilebileceği literatürde birçok çalışmada gösterilmiştir. McCaffrey ve ark. (2004), PS tahmininde genelleştirilmiş boosted modelleri (GBM) kullanarak başarılı sonuçlar elde etmişlerdir. Daha sonra Setoguchi ve ark. (2008), Lee ve ark. (2010), Colson ve ark. (2016) özellikle modelde sadece ana

etkiler olmadığında ve doğrusal olmayan ilişkiler mevcut olduğunda makine öğrenme algoritmaları ile PS tahmininde başarılı sonuçlar elde etmişlerdir. Ancak LR dışında makine öğrenme veya bayesci genelleştirilmiş doğrusal modeller, kuadratik diskriminant analizi ve adımsal akaike bilgi kriteri gibi farklı tahmincilerin kullanıldığı çalışmalar yeterli düzeyde değildir. Ayrıca hangi yöntemin hangi durumlarda kullanılacağına ilişkin kesin sonuçlar bulunmamaktadır. Özellikle LR ile Makine öğrenme yöntemlerinin birlikte kullanılarak daha başarılı eşleştirme sonuçlarına ulaşılacağı Pirracchio ve ark. (2014) tarafından gösterilmiştir. Ancak algoritmaların birlikte kullanılması ile ilgili literatürde Pirracchio ve ark. (2014)'nın çalışması dışında başka çalışmaya rastlanmamıştır. Bu çalışmada da hangi ML algoritmalarının PS hesabında güçlü olduğuna dair ön bilgi olmadığı için kombine modeller içerisindeki çok sayıdaki algoritmalar rastgele seçilmiş olup bulgular tek bir kombine tahminci ile elde edilmiştir.

Tez çalışmamızda etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu modellerde LR'den daha başarılı algoritmaların belirlenmesi için öncelikle LR için avantajlı olan modelde (sadece temel etki ve toplamsal) LR ile rekabet edecek algoritmaların belirlenmesi düşüncesi hakim olmuştur. Bu amaçla öncelikle LR ile en uyumlu algoritmayı belirlemek için birinci benzetim çalışması yapılmış olup benzetim çalışması sonucunda her iki örneklem büyüklüğünde de BAYESGLM, GLMNET, GAM, STEPAIC, DBARTS, GBM, EARTH, QDA, KSVM, CFOREST, NNET ve IPREDBAGG algoritmaları LR ile en uyumlu sonuç veren algoritmalar olarak belirlenmiştir.

Birinci benzetim çalışması sonucunda PS tahmininde LR ile uyumlu ve başarılı olarak belirlenen algoritmalar ile n=500 örneklem büyüklüğü için ikinci benzetim çalışması yapılmıştır. İkinci benzetim çalışması kapsamında 3 farklı senaryonun her birinde LR ve 12 alternatif algoritma ile eşleştirme analizinin yapıldığı toplam 13 benzetim yapılmıştır. A senaryosunda LR %0,494, KSVM %8,042, BAYESGLM %0,689, GLMNET %0,575, NNET %2,463, DBARTS %4,983, QDA %2,876, GAM %0,078, EARTH %2,272, STEPAIC %1,755, IPREDBAGG %41,421, GBM %22,150, CFOREST %34,336 yanlılıkla ATT kestiriminde bulunmuştur. ATT için

%10 ve altında yanlışlık ile kestirimde bulunan ve denge değerlendirmesi olan SOF, VO ve KS değerlerinde başarılı olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

B senaryosunda LR %9,087, KSVM %11,496, BAYESGLM %11,536, GLMNET %10,957, NNET %3,258, DBARTS %9,740, QDA %0,859, GAM %6,876, EARTH %3,783, STEPAIC %5,692, IPREDBAGG %41,275, GBM %27,174, CFOREST %37,607 yanlışlıkla ATT kestiriminde bulunmuştur. ATT için LR'den daha az yanlışlık ile kestirimde bulunan ve SOF, VO ve KS değerlerinde başarılı olan NNET, QDA, GAM, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

C senaryosunda LR %16,523, KSVM %18,185, BAYESGLM %16,749, GLMNET %16,539, NNET %9,847, DBARTS %14,868, QDA %0,459, GAM %20,126, EARTH %10,199, STEPAIC %10,552, IPREDBAGG %27,198, GBM %26,759, CFOREST %37,485 yanlışlıkla ATT kestiriminde bulunmuştur. ATT için LR'den daha az yanlışlık ile kestirimde bulunan ve SOF, VO ve KS değerlerinde başarılı olan NNET, QDA, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

500 örneklem büyüklüğünde A senaryosunda LR ile rekabet eden ve B ve C senaryolarında LR'den daha başarılı olan NNET, QDA, EARTH ve STEPAIC algoritmaları PS tahmininde üstün algoritmalar olarak belirlenmiştir. Ayrıca LR ile birlikte bu algoritmalar kullanılarak yeni bir kombine tahminci oluşturularak tekrar bir benzetim çalışması yapılmıştır. Benzetim çalışması sonucunda A senaryosunda LR %0,276 ve kombine tahminci %2,814 ile kestirimde bulunmuştur. B senaryosunda LR %9,914 ve kombine tahminci %8,552 ile kestirimde bulunmuştur. C senaryosunda ise LR %16,856 ve kombine tahminci %11,478 ile kestirimde bulunmuştur. SOF, OF, VO ve KS dikkate alındığında ise LR ve kombine tahminci birbirine yakın eşleştirme dengesi göstermiştir. 500 örneklem büyüklüğü ile gerçekleştirilen benzetim sonuçlarına göre LR, NNET, QDA, EARTH ve STEPAIC algoritmalarının kombinesi

ile oluşturulan yeni tahmin modelinin senaryo zorlaştıkça daha az yanlışlıkla tedavi etkisini tahmin ettiği belirlenmiştir.

İkinci benzetim çalışması kapsamında n=1000 örneklem büyüklüğü için 3 farklı senaryonun her birinde LR ve 12 alternatif algoritma ile eşleştirme analizinin yapıldığı toplam 13 benzetim yapılmıştır. A senaryosunda LR %0,226, KSVM %4,5, BAYESGLM %0,469, GLMNET %0,101, NNET %2,612, DBARTS %2,609, QDA %0,291, GAM %0,336, EARTH %2,483, STEPAIC %3,143, IPREDBAGG %24,72, GBM %16,46, CFOREST %36,17 yanlışlıkla ATT kestiriminde bulunmuştur. ATT için %10 ve altında yanlışlık ile kestirimde bulunan ve öncelikli denge değerlendirmesi olan SOF, VO ve KS değerlerinde başarılı olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

B senaryosunda LR %9,268, KSVM %6,701, BAYESGLM %10,248, GLMNET %10,358, NNET %4,924, DBARTS %7,573, QDA %1,640, GAM %8,578, EARTH %3,613, STEPAIC %2,964, IPREDBAGG %31,630, GBM %19,110, CFOREST %35,792 yanlışlıkla ATT kestiriminde bulunmuştur. ATT için LR'den daha az yanlışlık ile kestirimde bulunan ve SOF, VO ve KS değerlerinde başarılı olan KSVM, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

C senaryosunda LR %16,37, KSVM %12,41, BAYESGLM %16,72, GLMNET %16,74, NNET %14,89, DBARTS %9,715, QDA %1,344, GAM %19,98, EARTH %8,758, STEPAIC %11,622, IPREDBAGG %18,77, GBM %17,40 ve CFOREST %28,99 yanlışlıkla ATT kestiriminde bulunmuştur. ATT için LR'den daha az yanlışlık ile kestirimde bulunan ve SOF, VO ve KS değerlerinde başarılı olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları başarılı olarak belirlenmiştir.

1000 örneklem büyüklüğünde A senaryosunda LR ile rekabet eden ve B ve C senaryolarında LR'den daha başarılı olan KSVM, NNET, DBARTS, QDA, EARTH

ve STEPAIC algoritmaları PS tahmininde üstün algoritmalar olarak belirlenmiştir. Ayrıca LR ile birlikte bu algoritmalar kullanılarak yeni bir kombine tahminci oluşturularak tekrar bir benzetim çalışması yapılmıştır. Benzetim çalışması sonucunda A senaryosunda LR %0,496 ve kombine tahminci %0,527 ile kestirimde bulunmuştur. B senaryosunda LR %10,160 ve kombine tahminci %7,405 ile kestirimde bulunmuştur. C senaryosunda ise LR %17,762 ve kombine tahminci %7,672 ile kestirimde bulunmuştur. SOF, OF, VO ve KS dikkate alındığında ise LR ve kombine tahminci birbirine yakın eşleştirme dengesi göstermiştir. 1000 örneklem büyüklüğü ile gerçekleştirilen benzetim sonuçlarına göre LR, KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmalarının kombinesi ile oluşturulan yeni tahmin modelinin senaryo zorlaştıkça daha az yanlılıkla tedavi etkisini tahmin ettiği belirlenmiştir.

Benzetim çalışmaları sonucu elde edilen eşleştirme sonuçları incelendiğinde her iki örneklem büyüklüğünde de model zorluluk derecesi arttıkça LR'nin yanlılık değerinin arttığı görülmektedir. LR, 500 ve 1000 örneklem büyüklüğünde A, B ve C senaryolarında neredeyse benzer yanlılık değeri ile kestirimlerde bulunmuştur. Ancak kombine tahminci literatürdeki araştırmalara benzer olarak n=1000 büyüklüğünde A, B ve C senaryolarının tamamında n=500 örneklem büyüklüğüne göre daha yansız kestirimlerde bulunmuştur. LR ile kombine algoritma sonuçları karşılaştırıldığında A senaryosunda benzer kestirimlerde bulunduğu görülmektedir. Bu da kombine tahmincinin LR ile rekabet edebildiğini göstermiştir. B ve C senaryolarında ise kombine tahminci daha başarılı kestirimlerde bulunmuştur. Özellikle n=1000 örneklem büyüklüğünde en zor senaryo olan C senaryosunda kombine tahmincinin LR'den %10,09 daha az yanlı kestirimde bulunması ve LR'nin %17,76 gibi yüksek bir yanlılık kestiriminde bulunması karmaşık modellerde Makine öğrenme veya diğer algoritmaların LR'ye alternatif veya daha başarılı algoritma olarak PS tahmininde kullanılmasının yanlılığı azaltacağını göstermektedir. Ayrıca belirlenen algoritmaların özellikle büyük örneklemelerde daha başarılı kestirimlerde bulunduğu sonucuna ulaşılmıştır.

Denge göstergeleri olan SOF, OF, VO ve KS incelendiğinde de büyük örnekleme LR ve kombine tahmincinin daha dengeli eşleştirme yaptığı görülmektedir. Her iki örneklem büyüklüğünde SOF, OF, VO ve KS değerleri LR ve kombine algoritma ile neredeyse benzer olarak bulunmuştur. Ancak özellikle C senaryosunda Yapay Sinir Ağları, Kuadratik Diskriminant Analizi, Çok Değişkenli Uyarlanabilir Regresyon Eğrileri ve Adımsal Akaike Bilgi Kriteri yöntemleri LR'ye göre benzer VO ve KS değerleri elde etmiş olsa da, SOF değerlerinde daha başarılı eşleştirme analizi gerçekleştirmiştir.

Lalonde verileri eşleştirme analizinin önemini göstermesi açısından uygun bir veri setidir. Çünkü Lalonde'nin yapmış olduğu araştırma sonucunda eğitim programına katılanların daha sonrasında katılmayanlardan yüksek gelir elde etmesi gerekmektedir. Ancak eğitimin etkisini araştırmak için yapılan araştırmada kontrol grubunun gelir ortalaması 6984.170\$, deney grubunun ise 6349.144\$ olduğu görülmüştür. Yani eğitime katılmayanların ortalaması katılanlardan 635\$ daha fazla olarak bulunmuştur. Ancak mevcut ekonometrik çalışmalar eğitime yatırım yapmanın veya iyi okullara gitmenin kazançlar için olumlu bir etkisi olduğunu göstermekte idi. Bundan dolayı eğitim programının başarısız olma ihtimalinin olmadığı düşünüldüğünde bu durumun sebebi deney tasarımı ile ilgili karıştırıcı değişkenlerin kontrol altına alınmamasıdır. Yani araştırılan iş programı etkisi dışında araştırmada diğer karıştırıcı etkenler gruplarda benzer değildir. Bu yüzden çözüm olarak daha sonraki çalışmalarda yaş, eğitim, etnik köken, medeni durum, gelir ve 1974 ve 1975 kazançları bakımından tedavi grubuna benzeyen bir yapay kontrol grubu oluşturmak için PS'ye dayalı eşleştirme analizi uygulanmıştır. Eşleştirme analizi ile oluşturulan yarı-randomize çalışmanın sonucunda, iş eğitiminin 1978'deki gerçek kazançlar üzerinde gerçekten olumlu bir etki yarattığı sonucuna varılmıştır. Eşleştirme sonucunda 185 deney ve 185 kontrol arasında basit bir karşılaştırmayla ortalama eğitim etkisi 908\$, regresyon modeli ile ise 960\$ olarak bulunmaktadır. Ancak ortalama eğitim etkisi veya deney grubunda ortalama eğitim etkisi literatürde farklı PS tahmin yöntemine göre değişmektedir.

Dehejia ve Wahba (1999) yöntem karşılaştırmalarında duyarlılık analizlerinin de yapılabilmesi için Lalonde veri setinde kullanılan 297 eğitim verilen tedavi ve 425 eğitim verilmeyen kontrol grubundan, 1974 takvimindeki kazançların elde edilemediği gözlemleri çıkararak, veri setini 185 tedavi ve 260 kontrol gözlemine indirgemıştır (Lalonde veri setinde analizler eğitim öncesi kazançların sadece 1 yıl öncesine dayanmaktadır. Ancak bir yıldan daha fazla öncesindeki kazançların kullanılması tedavi etkisinin doğru bir şekilde tahmin edilmesinde önemlidir. Çünkü eğitim programları için gönüllü olan birçok kişi, eğitim programına girmeden hemen önce kazançlarında bir düşüş yaşamaktadır). Veri çıkarma esnasında gözlenen ve gözlenmeyen ortak değişkenler arasındaki denge bozulmamıştır (Dehejia ve Wahba, 1999).

Lalonde deneysel veri setinde eğitim programı sonrası bireylerin gelir düzeylerine eğitimin etkisinin 1794 \$ (eşleştirme yapılmadan veri setinde tedavi grubundakilerin kazancı kontrol grubundakilerden ortalama 1794 \$ daha fazla) olduğu bulunmuştur (Colson ve ark., 2016; Dehejia ve Wahba, 2002). Farklı seçenekler ile yapılan eşleştirme sonrasında ise literatürde birçok farklı ATE veya ATT önerilmiştir. Çalışmamızda ise ATT değerleri 1573 ile 2624 arasında bulunmuştur. Deneysel sonuçlara en yakın ATT sırasıyla QDA, STEPAIC, NNET, Kombine, EARTH ve LR tahmincileri ile bulunmuştur. Ancak algoritma performansları için önemli olan denge değeri Mutlak SOF toplamaları dikkate alındığında benzetim çalışmalarında LR'den daha başarılı olan algoritmalarından QDA hariç Kombine, NNET, STEPAIC ve EARTH gerçek veri setinde de LR'den daha başarılı olmuştur. Gerçek veri setinde kategorik yapıda olan değişkenlerde algoritmalar benzer başarıyı göstermiş olsa da özellikle sürekli yapıdaki sayısal ortak değişkenlere göre eşleştirme başarısında Kombine tahminci en yakın dengeyi sağlamıştır. Ayrıca denge sonucunda VO'lara baktığımızda LR'nin re74 ve re75 gibi büyük değerli sayısal ölçümlerde eşleştirme dengesini yakalayamadığı görülmüştür. Kombine tahminci ise neredeyse 1'e yakın VO dengesi göstermiştir. KS istatistiğinde de benzer durum gözlenmiştir. SuperLearner içindeki tahmincilerin ağırlık katsayılarına bakıldığında da kombine tahminciden sonra en az SOF dengesi gösteren STEPAIC'nin en büyük katsayıya sahip olduğu ve en yüksek

SOF dengesi gösteren LR'nin en düşük katsayıya sahip olduğu ve modelde kullanılmadığı görülmüştür.

Bu sonuçlar göstermektedir ki gözlemsel çalışmalarda elde edilen ATT ve ATE değerleri doğrudan eşleştirme analizinde kullanılan PS'lere ve dolayısıyla eşleştirme başarısına bağlıdır. Eşleştirme ise PS'ye göre yapılmaktadır. PS'nin vaka ve kontrol grubundaki bireylere verilen ortak değişkenlerin benzerliğine bağlı bir özet istatistik olduğu düşünüldüğünde her çalışmada LR ile PS hesaplanması durumunda tedavi veya kontrol grubuna atanma olasılığında doğrusal olmayan ilişkiler veya etkileşimlerin bulunması halinde yanlış sonuçlar elde edilmiş olacaktır. Elde ettiğimiz sonuçlara göre gözlemsel çalışmalarda LR'ye ek olarak ML algoritmaları veya diğer algoritmaların da PS hesaplanmasında kullanılarak denge değerlendirmesinin yapılması ve denge başarısına göre yansız ATT veya ATE nin hesaplanması önerilebilir.

Araştırmamızın birinci kısıtlılığı benzetim sonuçlarının bizim ürettiğimiz ilişkilere dayalı olan benzetim verisine dayanmasıdır. Ancak Setoguchi (2008) üretilen bu veri yapısının hem klinik farmakoloji hem de epidemiyolojiyi kapsayan ve çok sayıda insanda ilaçların kullanımının ve etkilerinin incelendiği Farmakoepidemiolojik çalışmalardaki yaygın olan veri yapısını modellediğini ifade etmiştir. Üretilen bu veride birçok sürekli ve kategorik karıştırıcı değişken bulunması ve gerçek tedavi etkisi araştırmalarından yola çıkarak üretilmesi bu kısıtlılık için bir cevap oluşturmaktadır. Ancak farklı korelasyonlar ve farklı örneklem büyüklükleri ile üretilen benzetim çalışmaları ile daha sonra benzer çalışmalar tekrarlanabilir.

İkinci bir kısıtlılık ise; Setoguchi ve ark. (2008), Lee ve ark. (2010), Pirracchio ve ark. (2014) ve Colson ve ark. (2016)'nın çalışmalarında olduğu gibi bu çalışmada da her bir algoritmanın yalnızca varsayılan özellikleri kullanılmıştır. Çünkü PS tahmininde makine öğrenme algoritmaları için birçok araştırmacının kullanacağı seçeneğin varsayılan ayarlar olacağı düşünülmektedir. Herhangi bir algoritmanın makine öğrenme uzmanı bir kullanıcı tarafından uygulandığında daha iyi performans göstermesi olasıdır. Örneğin rastgele ormanlar bazen büyük eğilimli skor ağırlıkları

üretmektedir ancak tahmin algoritmasına aşırı ağırlıkların olasılığını azaltmak için müdahale edilebilmektedir. Ayrıca sonuç değişkeni olarak kabul edilen tedavi ve kontrol grubunu gösteren tedavi değişkeni ile ortak değişkenler arasındaki etkileşimler analiz edilerek etkileşimlere sahip modellerin olduğu LR kullanılarak elde edilen eşleştirme dengesinin sonuçları yaygın olarak kullanılan basit ana etkiler modelinden daha iyi performans gösterebilecektir. Bazı araştırmacılar etkileşimleri kontrol etmeyi önermiş olsalar da (Lee ve ark., 2010) maalesef uygulamada yaygın olarak dikkate alınmamaktadır.

Çalışmanın üçüncü bir kısıtlılığı ise SOF, OF, VO ve KS denge istatistikleri dışında güncel denge değerlendirme yöntemleri ile de araştırmanın tekrarlanabileceğidir. Ancak pratikte halen denge değerlendirmesinde yaygın olarak SOF, OF, VO ve KS istatistikleri kullanılmaktadır.

Bu çalışmanın diğer çalışmalardan bazı güçlü yanları mevcuttur. Bunlardan ilki literatürde LR ile karşılaştırılan ML çalışmalarında ilk kez LR ile en uyumlu kestirimlerde bulunan algoritmalar bu tez çalışması ile belirlenmiştir. İkincisi ise PS hesabında bazı algoritmalar bireysel olarak ilk kez bu çalışmada kullanılmıştır. Özellikle BAYESGLM, CFOREST, DBARTS, EARTH, GAM, IPREDBAGG, GLMNET, KERNELKNN, KSVM, QDA, STEPAIC, SVM ve XGBOOST algoritmaları ile PS tahmini yapılan bir çalışmaya literatürde rastlanmamıştır. PS tahmininde literatürde kullanılmayan birçok yöntem ilk kez bu tez çalışması ile önerilmiştir. Literatürde özellikle random forest, GBM gibi makine öğrenme algoritmalarına odaklanılmıştır. Ancak bu algoritmalar LR'ye göre çok daha uzun sürede PS hesaplamaktadır. Literatürde benzetim çalışmalarında 1000 benzetim kullanılması da hesaplamaların uzun zaman almasından kaynaklanmaktadır. QDA, EARTH ve STEPAIC gibi algoritmalar ise zaman açısından LR ile mücadele edebilen algoritmalarlardır. Bu çalışmanın en önemli üstünlüğünden diğeri ise yeni önerilen kombine tahminci model ile varsayılan seçenekler dışında daha özel ayarlar ile daha yansız sonuçların elde edilebilmesidir.

5. SONUÇ VE ÖNERİLER

Gözlemsel çalışmalarda eşleştirme analizi sonrasında vaka ve kontrol grubu karşılaştırılırken eşleştirilmiş t testi sonucunda anlamlı çıkan p değeri sonrasında verilecek ortalamalar arasındaki fark veya hesaplanan tedavi etkileri vakaya karşı eşleşen kontrol grubu seçimine göre değişmektedir. Çoğu klinik araştırmada gruplar arasında cinsiyet ve yaş gibi birden çok ortak değişken/faktör açısından büyük farklılıklar bulunan bir veri setinin analizi sonucunda etkisi araştırılan bir parametre açısından gruplar arasında anlamlı farklılık görülürken eşleştirme analizi yapılması sonucunda gruplar dengeli hale geldikten sonra elde edilen analiz sonucunda aynı parametre açısından gruplar arasında istatistiksel olarak anlamlı farklılık görülmediği ortaya çıkmaktadır. Eşleştirme analizi ise literatürde çoğunlukla PS kullanılarak yapılmaktadır. Dolayısıyla PS tahmini karıştırıcı faktörlerin etkisini arındırarak saf tedavi etkisini belirlemek için araştırmalarda dikkate alınması gereken ve her geçen gün önemi artan bir yöntemdir.

Literatürde ML algoritmaları PS hesabında isteğe bağlı seçilmiş olup birbirine göreceli etkinliklerinin kıyaslandığı çalışmalar eksiktir. Bu tez çalışmasıyla modelde sadece ana etkiler bulunmadığı yani etkileşim etkilerinin ve doğrusal olmayan ilişkilerin olduğu durumlarda, LR'ye alternatif veya daha başarılı yapay sinir ağları gibi bazı makine öğrenme algoritmaları, bayesci genelleştirilmiş doğrusal modeller, kuadratik diskriminant analizi ve adımsal akaike bilgi kriteri gibi bazı diğer algoritmalar belirlenmiş ve belirlenen algoritmalar ile yeni bir kombine tahminci oluşturularak PS tahmini yapılmıştır. Bu algoritmalarının eşleştirme analizindeki başarıları benzetim ve gerçek veri çalışması ile araştırılmıştır.

Benzetim çalışmaları sonucunda model güçlük derecesi arttıkça LR'nin yanlışlık değerinin arttığı ve NNET, QDA, EARTH, STEPAIC, KSVM ve DBARTS algoritmalarının daha az yanlış kestirimlerde bulunduğu belirlenmiştir. Bu algoritmalarından NNET, QDA, EARTH, STEPAIC ve bu tez çalışması sonucunda önerilen Kombine tahminci model ile gerçek veri setinden elde edilen PS'ler ile

yapılan eşleştirme analizi sonucunda QDA hariç tüm modeller LR'den daha dengeli eşleştirme analizi sonucu vermiştir. Benzetim uygulamaları ile LR'ye alternatif olarak belirlenen algoritmaların eşleştirme analizi öncesinde PS tahmininde kullanılmasının gözlemsel araştırmaların güvenilirliğini artıracak sonucuna ulaşılmıştır. Özellikle kombine tahminci ile elde edilen yeni PS tahmin modeli sonuçlarının gerçek veri uygulamasında da daha başarılı olması bu tez hipotezini desteklemiştir.

Parametrik yöntemler (LR) tedavi gruplarına atanma olasılığı ile ortak değişkenler arasındaki ilişkinin fonksiyonel yapısı ile ilgili varsayımlar gerektirir. PS hesabı için modelin yanlış belirlenmesi tedavi etkisinin yanlış tahminlerine sebep olabilmektedir. Modelde tedavi veya kontrol grubuna atanma olasılığı ile ortak değişkenler/faktörler arasında etkileşim etkileri veya kuadratik ilişkiler bulunmadığında, yani sadece temel etkiler olduğunda LR ile dengeli eşleştirme analizi gerçekleştirilmekte ve buna dayalı olarak yansız tahminlere ulaşılmaktadır. Ancak doğrusal olmayan ilişkiler ve/veya etkileşimlerin bulunduğu modeller için ML yöntemleri ve kombine tahminci model ile daha dengeli ve yansız tahminler elde edilmiştir. Bu çalışma sonucunda araştırmacılara bu modellerin denenmesi önerilmektedir. ML yöntemlerinde doğrusallık veya etkileşimler incelenmeden doğrudan ortak değişkenler seçilerek skorlar hesaplanabilmektedir. PS hesaplamak ve gruplar arası dengeyi değerlendirebilmek için son yıllarda kullanımı artmaya başlayan ML yaklaşımlarının başarısının gözlemsel çalışmaların güvenilirliğini artıracakı düşünülmektedir.

ÖZET

Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi

Randomizasyonun yapılamadığı gözlemsel araştırmalarda, tedavi ve kontrol grupları arasında araştırma öncesinde birçok ortak değişken veya faktöre göre farklılıklar bulunabilir. Gruplar araştırılan etken dışında diğer özellikler açısından homojen olmadığında saf tedavi etkisi kestirilemeyecek ve yanlış tahminlere ulaşılabacaktır. Gözlemsel araştırmalarda tedavi etkisinin yansız tahminlerini elde etmek ve karıştırıcı (confounder) değişkenlerin etkisini gidermek için literatürde yaygın olarak Lojistik Regresyon (LR) ile hesaplanan Propensity skor (PS) tahminine dayalı eşleştirme analizi kullanılmaktadır.

Bu çalışmada tedavi grupları ile ortak değişkenler arasında ana etkiler dışında etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu modellerde PS tahmini için klasik yöntem olan LR'ye alternatif olarak makine öğrenme yöntemleri (genelleştirilmiş boosted modeller, sınıflandırma ve regresyon ağaçları, destek vektör makineleri ve yapay sinir ağları vb.) ve bazı yaygın kullanılan yöntemlerin (bayesci genelleştirilmiş doğrusal modeller, kuadratik diskriminant analizi ve adimsal akaiki bilgi kriteri vb.) başarısını benzetim çalışması ile değerlendirmek suretiyle yeni bir kombine tahminci model geliştirmek amaçlanmıştır. Bu amaçla gerçekleştirilen birinci benzetim çalışmasında sadece ana etkilerin bulunduğu modelde LR ile hesaplanan PS değerleri ile en uyumlu sonuç veren ilk üç algoritma 500 ve 1000 örneklem büyüklüğü için sırasıyla BAYESGLM, GLMNET ve GAM olarak bulunmuştur.

Sadece temel etkilerin olduğu birinci benzetim senaryosunda LR ile rekabet edebilen ve etkileşim ve kuadratik ilişkilerin olduğu daha karmaşık ikinci ve üçüncü benzetim senaryosunda da LR'den daha başarılı olan 1000 örneklem büyüklüğünde KSVM, NNET, DBARTS, QDA, EARTH, STEPAIC ve 500 örneklem büyüklüğünde NNET, QDA, EARTH ve STEPAIC algoritmaları PS tahmininde üstün algoritmalar olarak belirlenmiştir. LR ile birlikte bu algoritmalar kullanılarak oluşturulan yeni kombine PS tahmin modelleri ile yapılan üç farklı benzetim senaryosu sonucunda model zorluluk derecesi arttıkça LR ile elde edilen tahminlerde yanlışlık değerinin arttığı görülmüştür. Özellikle karmaşık modellerde kombine tahminci ile LR sonuçları arasında büyük yanlışlık farkı bulunması kombine tahmincinin veya KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmalarının LR'ye alternatif veya daha başarılı algoritma olarak PS tahmininde kullanılabileceğini göstermiştir. Ayrıca gerçek veri ile yapılan uygulama sonuçları da tez hipotezini desteklemiştir.

PS vaka ve kontrol grubundaki bireylerin sahip olduğu karıştırıcı değişken özelliklerine göre elde edilen bir özet istatistiktir. Bu yüzden vaka veya kontrol grubuna atanma olasılığı ile ortak değişkenler arasında doğrusal olmayan ilişkiler veya etkileşimlerin bulunduğu modellerde LR yöntemiyle PS hesaplanması durumunda yanlış tahminler elde edilmiş olacaktır. Elde ettiğimiz sonuçlara göre gözlemsel çalışmalarda LR'ye ek olarak başarılı bulunan ML algoritmaları veya diğer algoritmaların da PS hesaplanmasında kullanılarak denge değerlendirmesinin yapılması önerilebilir.

Parametrik yöntemler (LR) deney gruplarına atanma olasılığı ile ortak değişkenler arasındaki ilişkinin fonksiyonel yapısı ile ilgili varsayımlar gerektirir. PS hesabı için modelin yanlış belirlenmesi tedavi etkisinin yanlış tahminlerine sebep olabilmektedir. Modelde sadece temel etkiler olduğunda LR ile dengeli eşleştirme analizi gerçekleştirilmekte ve buna dayalı olarak yansız tahminlere ulaşılmaktadır. Ancak tedavi veya kontrol grubuna atanma olasılığı ile ortak değişkenler/faktörler arasında etkileşim etkileri ve/veya kuadratik ilişkiler bulunan

modeller için ML yöntemleri ve kombine tahminciler ile daha dengeli ve yansız tahminler elde edilmiştir. Bu çalışma sonucunda araştırmacılara bu modellerin denenmesi önerilmektedir. ML yöntemlerinde doğrusallık veya etkileşimler incelenmeden doğrudan ortak değişkenler seçilerek skorlar hesaplanabilmektedir. PS hesaplamak ve gruplar arası dengeyi değerlendirebilmek için son yıllarda kullanımı artmaya başlayan ML yaklaşımlarının gözlemsel çalışmaların güvenilirliğini arttıracakı düşünülmektedir.

Anahtar Sözcükler: Eşleştirme, Lojistik Regresyon, Makine Öğrenme, Propensity Skor, Randomize olmayan çalışmalar.



SUMMARY

Development of New Propensity Score Estimation Models with Machine Learning Algorithms for Optimal Matching Analysis in Non-Randomized Clinical Trials

In observational studies in which randomisation cannot be performed, there may be differences between treatment and control groups according to many covariates or factors prior the research. When the groups are not homogeneous in terms of other characteristics beside the investigated factors, pure treatment effect cannot be predicted, which will cause biased estimates. In observational studies, matching analysis based on Propensity Score (PS) estimation calculated by Logistic Regression (LR) is widely used in the literature in order to obtain unbiased estimates of the treatment effect and to eliminate the influence of confounding variables.

This study aims to develop a new combined predictor model by evaluating success of machine learning methods (generalized boosted models, classification and regression trees, support vector machines, artificial neural networks etc.) and some commonly used methods (Bayesian generalized linear models, quadratic discriminant analysis, stepwise akaike information criterion, etc.) with simulation studies as an alternative to logistic regression which is the classical method for PS estimation in the models with interaction effects and nonlinear relations. In the first simulation studies having 500 and 1000 sample size, the first three algorithms showing best agreement with the PS values calculated with LR were found as BAYESGLM, GLMNET and GAM respectively in the model having only the main effects.

KSVM, NNET, DBARTS, QDA, EARTH, STEPAIC algorithms and NNET, QDA, EARTH and STEPAIC algorithms, which are more successful than LR in the more complex second and third simulation scenarios formed, which can compete with LR in the first simulation scenario with only basic effects, were determined as superior algorithms in PS estimation for sample size of 1000 and 500 respectively. As a result of three different simulation scenarios formed with new combined PS estimation models using LR together with these algorithms, the value of bias was found to increase in the estimates obtained through LR. Especially in complex models, the difference between the combined estimator model and the LR results showed that the combined estimator or the KSVM, NNET, DBARTS, QDA, EARTH and STEPAIC algorithms could be used to estimate PS as an alternative or more successful algorithm to LR. In addition, the application results with real data supported the hypothesis of the thesis.

PS is a summary statistics obtained according to the covariate variable characteristics of subjects in the case and control group. Therefore, if there are nonlinear relationships or interactions between the covariates and the probability of being assigned to the case or control group, biased predictions will be obtained in the cases of calculating PS by LR method. According to the results obtained in this study, in addition to LR, it can be suggested to perform a balance assessment using ML algorithms or other algorithms in calculating PS in observational studies.

Parametric methods (LR) require assumptions about the functional structure of the relationship between the probability of being assigned to experimental groups and the covariates. Incorrect determination of the model for a propensity score calculation may lead to biased estimation of treatment effect. When there are only basic effects in the model, balanced matching analysis with LR is performed and unbiased estimations are achieved. However, more balanced and unbiased estimates were obtained with the ML methods and

combined estimators in the models with interaction effects and/or quadratic relations between the probability of being assigned to the treatment or control groups and the covariates/factors. As a result of this study, researchers are recommended to try these models. In ML methods, the scores can be calculated by directly selecting covariates without examining linearity and interactions. It is thought that the reliability of observational studies will increase with the use of ML approaches that have started to increase in recent years in order to calculate PS and to evaluate the balance between groups.

Keywords: Logistic Regression, Machine Learning, Matching, Nonrandomized Trials, Propensity Score.



KAYNAKLAR

- ABADIE A, IMBENS G (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**: 235-267.
- ANKARALI HC, SÜMBÜLOĞLU V, YAZICI AC, YALUĞ I, SELEKLER M (2009). Comparison of different matching methods in observational studies and sensitivity analysis: the relation between depression and stai-2 scores. *Expert Systems with Applications*, **36**: 1876-84.
- AUSTIN PC (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, **28**: 3083-3107.
- AUSTIN PC (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.*, **46**: 399-424.
- AUSTIN PC (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, **33**: 1057-1069.
- BAYCAN IO (2016). The effects of exchange rate regimes on economic growth: evidence from propensity score matching estimates. *J Appl. Stat.*, **43**: 914-924.
- CALIENDO M, KOPEINIG S (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, **22**: 31-72.
- COCHRAN WG, RUBIN DB (1973). Controlling bias in observational studies: a review. *Sankhya, Ser. A*, **35**: 417-46.
- COLSON KE, RUDOLPH KE, ZIMMERMAN SC, GOIN DE, STUART EA, LAAN M VAN DER, AHERN J. (2016). Optimizing matching and analysis combinations for estimating causal effects. *Scientific Reports*, **6**: 23222.
- ÇAPARLAR CÖ, DÖNMEZ A (2016). Bilimsel araştırma nedir, nasıl yapılır? *Turk J Anaesthesiol Reanim*, **44**: 212-8.
- D'AGOSTINO RB JR (1998). Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, **17**: 2265-2281.

- DEHEJIA RH, WAHBA S (1999). Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, **94**: 1053-1062.
- DEHEJIA RH, WAHBA S (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, **84**: 151-161.
- DEMİR O, DOLGUN A, ETİKAN İ, KUYUCU YE, SARAÇBAŞI O (2017). Simulation study on performance of balance metrics in propensity score weighting method. *J Contemp. Med.*, **7**: 265-277.
- DEMİR E, KÖSE SK (2016). Randomize olmayan klinik çalışmalarda propensity skor tahmini için makine öğrenme yaklaşımı. *I. Uluslararası Biyoistatistik Kongresi*, 26-29 Ekim, Belek-Antalya.
- DEMİR E, KOSE SK (2017). The most compatible machine learning algorithm with propensity score estimates obtained by logistic regression analysis. *6th IBS Channel Network Conference*, April 24-26, Diepenbeek-Belgium.
- DEMİR E, HAYME S, KOSE SK (2017). A systematic review and bibliometric analysis on propensity score obtained by using machine learning algorithms. *6th IBS Channel Network Conference*, April 24-26, Diepenbeek-Belgium.
- DEMİR E, HAYME S, ÖZTUNA D, ELHAN AH, KÖSE SK (2017). Gözlemsel çalışmalarda ölçülemeyen karıştırıcı değişkenlerin etkisi; propensity skor ve duyarlılık analizleri, *II. Uluslararası Biyoistatistik Kongresi*, 25-28 Ekim, Belek-Antalya.
- DEMİR O, DOLGUN A, ETİKAN İ, KUYUCU YE, SARAÇBAŞI O (2014). Çoklu tedavilerde propensity skor ağırlıklandırma yöntemi: multinomial lojistik regresyon ve geliştirilmiş boosted model karşılaştırması, *XVI. Ulusal Biyoistatistik Kongresi*, 10-12 Eylül, Side-Antalya.
- FISHER RA. (1926) The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture*, **33**: 503-515.
- GAMER M, LEMON J, SINGH P (2015). Various coefficients of interrater reliability and agreement. Package 'irr', version 0.84.
- GÖKMEN D, ALKAN A, BAKIRARAR B, SİDDİKOĞLU D, ÖZGÜR EG, ATEŞ FSÖ, KAR İ, DEMİR P, HAYME S, ARSLAN YK, YAVUZ Z (2018). Bilimsel Araştırma Yöntemleri. Ankara Üniversitesi Basımevi, Ankara.
- GÜRTEKİN B (2008). Propensity skor yönteminin meme kanserinde yaş ile dokudaki ER, PR, CERB-B2 değişiminin analizinde kullanılması, Doktora tezi, İstanbul Üniversitesi, İstanbul.

- HECKMAN J (2005). The scientific model of causality. *Sociological Methodology*, **35**: 1-97.
- HERNAN MA (2004). A definition of causal effect for epidemiological research. *J Epidemiol Community Health*, **58**: 265-271.
- HILL HA, KLEINBAUM DG (2005). Bias in observational studies. *Encyclopedia of Biostatistics*, John Wiley and Sons, Chichester.
- HOFFMEISTER H, SZKLO M, THAMM M (1998). Bias In Observational Studies. Epidemiological Practices in Research on Small Effects. Springer, Berlin, Heidelberg
- HOFER M (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology*, **5**: 28.
- IMBENS GW (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, **87**: 706-710.
- IMBENS GW (2004). Nonparametric estimation of average treatment effects under exogeneity: a review, *The Review of Economics and Statistics*, **86**: 4-29.
- KANIK A, TAŞDELEN B, ERDOĞAN S (2011). Klinik denemelerde randomizasyon. *Marmara Medical Journal*, **24**: 149-155.
- KASPAR EÇ, BEKİROĞLU N, GENÇELİ M (2010). Gözleme dayalı çalışmalarda eğilim skoru (propensity score) ve tıp bilimleri'nde bir uygulama. *Türkiye Klinikleri Biyoistatistik Dergisi*, **2**: 1-10.
- KOO TK, LI MY (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, **15**: 155-163.
- LALONDE RJ (1986). Evaluating the econometric evaluations of training programs with experimental data, *The American Economic Review*, **76**: 604-620.
- LAMORTE WW, SULLIVAN L. Boston University School of Public Health, Confounding and Effect Measure Modification, Erişim Adresi: [http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704-EP713_Confounding-EM/BS704-EP713_Confounding-EM_print.html]. Erişim Tarihi: 20/02/2018.
- LEACY FP, STUART EA (2014). On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine*, **33**: 3488-3508.
- LEE BK, LESSLER J, STUART EA (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, **29**: 337-346.

- LEE BK, LESSLER J, STUART EA (2011). Weight trimming and propensity score weighting. *PLoS ONE*, **6**: e18174.
- MCCAFFREY DF, RIDGEWAY G, MORRAL AR (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, **9**: 403-425.
- ÖZDEMİR AA, OREKİCİ TEMEL G, KANIK EA (2014). Dengesiz grupları dengeli hale getirmek için bir çözüm önerisi: propensity skor, XVI. Ulusal Biyoistatistik Kongresi, 10-12 Eylül, Side-Antalya.
- PANDIS N (2014). Bias in observational studies. *American Journal of Orthodontics and Dentofacial Orthopedics*, **145**: 542-543.
- PIRRACCHIO R, PETERSEN M, VAN DER LAAN M (2014). Improving propensity score estimators' robustness to model misspecification using super learner. *American Journal of Epidemiology*, **181**: 108-19.
- POLLEY EC, VAN DER LAAN MJ (2010). Super learner in prediction. U.C. Berkeley Division of Biostatistics Workihong Paper Series. Paper 226.
- ROSENBAUM PR (2002). *Observational Studies*. Springer Series in Statistics. Springer, New York, NY (2nd Ed.)
- ROSENBAUM PR, RUBIN DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**: 41-55.
- ROSENBAUM PR, RUBIN DB (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, **79**: 516-524.
- ROSENBAUM PR, RUBIN DB (1985). Constructing a control-group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, **39**: 33-38.
- ROSENBERGER WF, LACHIN JM (2002). *Randomization in Clinical Trials: Theory and Practice*. John Wiley, New York.
- RUBIN DB (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**: 688-701.
- RUBIN DB (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, **127**: 757-763.

- RUBIN DB (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, **74**: 318-328.
- SEKHON, JS (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software*, **42**: 1-52.
- SETOGUCHI S, SCHNEEWEISS S, BROOKHART MA, GLYNN RJ, COOK EF (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety*, **17**: 546-55.
- STARK CR, MANTEL N (1966). Effects of maternal age and birth order on the risk of mongolism and leukemia. *J Natl Cancer Inst.*, **37**: 687-98.
- STUART EA (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, **25**: 1-21.
- YILMAZ E, ÇÖL M (2014). Evidence Based Medicine. *Journal of Clinical and Analytical Medicine*, **5**: 537-42.
- WESTREICH D, LESSLER J, FUNK MJ (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin. Epidemiol*, **63**: 826-33.
- WOS: Thomson Reuters Web of Science. Erişim Adresi: [www.webofknowledge.com/]. Erişim Tarihi: 11/03/2018.

EK-1. Benzetim Verisi Üretimi için Örnek R Kodları (Senaryo A, n=1000)

```
function(n){

F.sample.cor <- function(x, rho) {
y <- (rho * (x - mean(x)))/sqrt(var(x)) + sqrt(1 - rho^2) * rnorm(length(x))
#cat("Sample corr = ", cor(x, y), "\n")
return(y)
}

F.generate <- function(size, scenario) {
w1 <- rnorm(size, mean=0, sd=1)
w2 <- rnorm(size, mean=0, sd=1)
w3 <- rnorm(size, mean=0, sd=1)
w4 <- rnorm(size, mean=0, sd=1)
w5 <- F.sample.cor(w1, 0.2)
w6 <- F.sample.cor(w2, 0.9)
w7 <- rnorm(size, mean=0, sd=1)
w8 <- F.sample.cor(w3, 0.2)
w9 <- F.sample.cor(w4, 0.9)
w10 <- rnorm(size, mean=0, sd=1)

w1 <- ifelse(w1 > mean(w1), 1, 0)
w3 <- ifelse(w3 > mean(w3), 1, 0)
w5 <- ifelse(w5 > mean(w5), 1, 0)
w6 <- ifelse(w6 > mean(w6), 1, 0)
w8 <- ifelse(w8 > mean(w8), 1, 0)
w9 <- ifelse(w9 > mean(w9), 1, 0)

if (scenario == "A"){
z.a_trueps <- (1 + exp( -(0 + b1*w1 + b2*w2 + b3*w3 + b4*w4 + b5*w5 + b6*w6 + b7*w7) ))^-1
} else
if (scenario == "B") {
z.a_trueps <- (1 + exp( -(0 + b1*w1 + b2*w2 + b3*w3 + b4*w4 + b5*w5 + b6*w6 + b7*w7
+ b2*w2*w2 + b1*0.5*w1*w3 + b2*0.7*w2*w4 + b4*0.5*w4*w5 + b5*0.5*w5*w6) ) ))^-1
} else
{
# scenario C
z.a_trueps <- (1 + exp( -(0 + b1*w1 + b2*w2 + b3*w3 + b4*w4 + b5*w5 + b6*w6 + b7*w7
+ b2*w2*w2 + b4*w4*w4 + b7*w7*w7 + b1*0.5*w1*w3 + b2*0.7*w2*w4 + b3*0.5*w3*w5
+ b4*0.7*w4*w6 + b5*0.5*w5*w7 + b1*0.5*w1*w6 + b2*0.7*w2*w3 + b3*0.5*w3*w4
+ b4*0.5*w4*w5 + b5*0.5*w5*w6) ) ))^-1
}

prob.exposure <- runif(size)
z.a <- ifelse(z.a_trueps > prob.exposure, 1, 0)

y.a <- a0 + a1*w1 + a2*w2 + a3*w3 + a4*w4 + a5*w8 + a6*w9 + a7*w10 + g1*z.a

sim <- as.data.frame(cbind(w1, w2, w3 ,w4, w5, w6, w7, w8, w9, w10, z.a, y.a))
return(sim)
}

b0 <- 0
b1 <- 0.8
b2 <- -0.25
b3 <- 0.6
b4 <- -0.4
b5 <- -0.8
b6 <- -0.5
b7 <- 0.7
a0 <- -3.85
a1 <- 0.3
```

```

a2 <- -0.36
a3 <- -0.73
a4 <- -0.2
a5 <- 0.71
a6 <- -0.19
a7 <- 0.26
g1 <- -0.4

simdata <- F.generate(1000, "A")

simdata

```

LR İle PS Hesaplanması ve Eşleştirme Analizi için Örnek Kod:

```

glm <- glm(z.a ~ w1 + w2 + w3 + w4 + w5 + w6 + w7 + w8 + w9 + w10, data = simdata, family='binomial')

eşleş=Match(Y = simdata$y.a, Tr = simdata$z.a, X = glm$fitted)

```

ML (Bayesglm) ile PS Hesaplanması ve Eşleştirme Analizi için Örnek Kod:

```

SL.library <- c("SL.bayesglm")

SL.out <- SuperLearner(Y = simdata$z.a, X =
simdata[,c("w1","w2","w3","w4","w5","w6","w7","w8","w9","w10")],
  SL.library = SL.library,
  family = 'binomial',
  cvControl = list(V=10))

eşleş=Match(Y = simdata$y.a, Tr = simdata$z.a, X = SL.out$SL.predict)

```

Kombine Tahminci ile PS Hesaplanması ve Eşleştirme Analizi için Örnek Kod:

```

SL.library <- c("SL.glm", "SL.ksvm", "SL.nnet", "SL.qda", "SL.dbarts", "SL.earth", "SL.stepAIC")
SL.out <- SuperLearner(Y = simdata$z.a, X =
simdata[,c("w1","w2","w3","w4","w5","w6","w7","w8","w9","w10")],
  SL.library = SL.library,
  family = 'binomial',
  cvControl = list(V=10))

eşleş=Match(Y = simdata$y.a, Tr = simdata$z.a, X = SL.out$SL.predict)

```

EK-2. Denge Sonuçları (Senaryo A, n=500)

LR için denge sonuçları

	SOE	OF	VO	KS
W ₁	0,006	0,003	1,011	-
W ₂	-0,002	-0,002	1,056	0,112
W ₃	0,004	0,002	1,013	-
W ₄	-0,010	-0,011	1,072	0,110
W ₅	-0,002	-0,001	1,013	-
W ₆	0,001	0,001	1,012	-
W ₇	0,014	0,014	1,089	0,108
W ₈	-0,009	-0,005	1,015	-
W ₉	-0,001	-0,001	1,014	-
W ₁₀	-0,005	-0,005	1,043	0,115
Mutlak Ortalama	0,005	0,005	1,034	0,111

KSVM için denge sonuçları

	SOE	OF	VO	KS
W ₁	-0,030	-0,015	1,048	-
W ₂	0,068	0,067	1,21	0,152
W ₃	-0,035	-0,017	1,049	-
W ₄	0,029	0,028	1,213	0,149
W ₅	0,030	0,015	1,051	-
W ₆	0,091	0,045	1,083	-
W ₇	-0,026	-0,024	1,255	0,145
W ₈	-0,013	-0,007	1,039	-
W ₉	0,067	0,033	1,063	-
W ₁₀	0,001	0,001	1,229	0,149
Mutlak Ortalama	0,039	0,025	1,124	0,149

BAYESGLM için denge sonuçları

	SOE	OF	VO	KS
W ₁	0,003	0,002	1,012	-
W ₂	-0,001	-0,002	1,056	0,111
W ₃	0,003	0,002	1,014	-
W ₄	-0,005	-0,005	1,061	0,110
W ₅	-0,006	-0,003	1,013	-
W ₆	-0,004	-0,002	1,011	-
W ₇	0,013	0,013	1,087	0,108
W ₈	0,006	0,003	1,015	-
W ₉	-0,001	-0,001	1,014	-
W ₁₀	-0,001	-0,001	1,058	0,115
Mutlak Ortalama	0,004	0,003	1,034	0,111

GLMNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,016	0,008	1,010	-
W₂	-0,012	-0,013	1,065	0,111
W₃	0,017	0,008	1,011	-
W₄	-0,018	-0,018	1,060	0,111
W₅	-0,013	-0,007	1,012	-
W₆	-0,005	-0,003	1,012	-
W₇	-0,011	-0,010	1,068	0,107
W₈	-0,005	-0,002	1,015	-
W₉	-0,003	-0,001	1,014	-
W₁₀	-0,003	-0,003	1,046	0,112
Mutlak Ortalama	0,010	0,007	1,031	0,110

NNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,045	0,022	1,010	-
W₂	-0,016	-0,016	1,071	0,104
W₃	0,032	0,016	1,014	-
W₄	-0,026	-0,026	1,079	0,103
W₅	-0,028	-0,014	1,014	-
W₆	-0,020	-0,010	1,011	-
W₇	0,045	0,044	1,119	0,105
W₈	0,009	0,004	1,019	-
W₉	-0,007	-0,004	1,017	-
W₁₀	0,001	<0,001	1,076	0,104
Mutlak Ortalama	0,023	0,016	1,043	0,104

DBARTS için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,072	-0,036	1,047	-
W₂	0,055	0,054	1,113	0,122
W₃	-0,058	-0,029	1,041	-
W₄	0,012	0,012	1,119	0,119
W₅	0,074	0,037	1,050	-
W₆	0,093	0,046	1,066	-
W₇	-0,010	-0,009	1,160	0,116
W₈	-0,007	-0,004	1,023	-
W₉	0,069	0,034	1,044	-
W₁₀	0,002	0,002	1,104	0,121
Mutlak Ortalama	0,045	0,026	1,077	0,120

QDA için denge sonuçları

	SOF	OF	VO	KS
W₁	0,002	0,001	1,019	-
W₂	-0,010	-0,010	1,074	0,123
W₃	0,004	0,002	1,019	-
W₄	-0,020	-0,020	1,093	0,123
W₅	0,001	0,001	1,019	-
W₆	-0,001	<0,001	1,017	-

W₇	0,030	0,029	1,108	0,123
W₈	0,004	0,002	1,021	-
W₉	-0,008	-0,004	1,018	-
W₁₀	0,011	0,011	1,081	0,126
Mutlak Ortalama	0,009	0,008	1,047	0,124

GAM için denge sonuçları

	SOE	OF	VO	KS
W₁	-0,008	-0,004	1,017	-
W₂	-0,007	-0,007	1,068	0,110
W₃	-0,002	-0,001	1,015	-
W₄	0,003	0,003	1,062	0,111
W₅	0,004	0,002	1,016	-
W₆	0,003	0,001	1,014	-
W₇	0,006	0,005	1,090	0,105
W₈	-0,002	-0,001	1,016	-
W₉	0,004	0,002	1,015	-
W₁₀	0,001	0,001	1,055	0,111
Mutlak Ortalama	0,004	0,003	1,037	0,109

EARTH için denge sonuçları

	SOE	OF	VO	KS
W₁	0,025	0,012	1,012	-
W₂	-0,041	-0,041	1,081	0,118
W₃	0,046	0,023	1,015	-
W₄	-0,037	-0,037	1,089	0,117
W₅	-0,029	-0,014	1,014	-
W₆	-0,052	-0,026	1,004	-
W₇	0,026	0,025	1,103	0,110
W₈	0,002	0,001	1,021	-
W₉	-0,017	-0,009	1,015	-
W₁₀	-0,008	-0,008	1,080	0,122
Mutlak Ortalama	0,028	0,020	1,043	0,117

STPAIC için denge sonuçları

	SOE	OF	VO	KS
W₁	0,024	0,012	1,013	-
W₂	-0,047	-0,047	1,059	0,114
W₃	0,056	0,028	1,015	-
W₄	-0,025	-0,025	1,067	0,112
W₅	-0,024	-0,012	1,012	-
W₆	-0,042	-0,021	1,003	-
W₇	0,009	0,010	1,082	0,106
W₈	0,006	0,003	1,016	-
W₉	-0,011	-0,005	1,015	-
W₁₀	-0,005	-0,005	1,054	0,121
Mutlak Ortalama	0,025	0,017	1,034	0,113

IPREDBAGG için denge sonuçları

	SOE	OF	VO	KS
W₁	0,067	0,033	1,127	-
W₂	0,254	0,250	1,384	0,257
W₃	0,090	0,045	1,110	-
W₄	0,170	0,167	1,378	0,238
W₅	-0,086	-0,043	1,112	-
W₆	0,186	0,092	1,248	-
W₇	-0,430	-0,413	1,591	0,304
W₈	0,004	0,002	1,115	-
W₉	0,169	0,084	1,207	-
W₁₀	-0,018	-0,018	1,315	0,224
Mutlak Ortalama	0,147	0,115	1,259	0,256

GBM için denge sonuçları

	SOE	OF	VO	KS
W₁	0,056	0,028	1,018	-
W₂	0,130	0,128	1,135	0,144
W₃	0,064	0,032	1,023	-
W₄	0,091	0,090	1,154	0,137
W₅	-0,054	-0,027	1,018	-
W₆	0,066	0,033	1,057	-
W₇	-0,176	-0,169	1,220	0,157
W₈	-0,005	-0,003	1,029	-
W₉	0,085	0,042	1,061	-
W₁₀	<0,001	-0,001	1,117	0,128
Mutlak Ortalama	0,073	0,055	1,083	0,142

CFOREST için denge sonuçları

	SOE	OF	VO	KS
W₁	-0,057	-0,028	1,094	-
W₂	0,243	0,239	1,291	0,222
W₃	-0,049	-0,024	1,088	-
W₄	0,160	0,157	1,245	0,203
W₅	0,048	0,024	1,095	-
W₆	0,267	0,131	1,270	-
W₇	-0,200	-0,192	1,421	0,214
W₈	-0,023	-0,012	1,073	-
W₉	0,222	0,110	1,187	-
W₁₀	0,001	0,001	1,202	0,188
Mutlak Ortalama	0,127	0,092	1,197	0,207

EK-3. Denge Sonuçları (Senaryo B, n=500)

LR için denge sonuçları

	SOE	OF	VO	KS
W₁	0,033	0,016	1,003	-
W₂	0,080	0,071	0,676	0,139
W₃	0,033	0,016	1,003	-
W₄	0,014	0,013	1,095	0,108
W₅	0,023	0,011	1,021	-
W₆	0,024	0,012	1,023	-
W₇	0,036	0,035	1,099	0,111
W₈	0,002	0,001	1,014	-
W₉	0,013	0,007	1,017	-
W₁₀	-0,001	-0,001	1,063	0,113
Mutlak Ortalama	0,026	0,018	1,001	0,118

KSVM için denge sonuçları

	SOE	OF	VO	KS
W₁	-0,026	-0,013	1,050	-
W₂	0,113	0,100	1,022	0,159
W₃	-0,030	-0,015	1,050	-
W₄	0,028	0,028	1,228	0,149
W₅	0,050	0,024	1,066	-
W₆	0,114	0,055	1,115	-
W₇	-0,008	-0,006	1,305	0,147
W₈	-0,018	-0,009	1,037	-
W₉	0,064	0,032	1,063	-
W₁₀	0,004	0,004	1,262	0,149
Mutlak Ortalama	0,045	0,029	1,120	0,151

BAYESGLM için denge sonuçları

	SOE	OF	VO	KS
W₁	0,043	0,021	1,000	-
W₂	0,102	0,090	0,674	0,140
W₃	0,034	0,017	1,003	-
W₄	0,024	0,023	1,094	0,112
W₅	0,018	0,009	1,021	-
W₆	0,038	0,019	1,030	-
W₇	0,049	0,048	1,096	0,112
W₈	0,001	0,001	1,015	-
W₉	0,023	0,012	1,022	-
W₁₀	-0,005	-0,005	1,057	0,114
Mutlak Ortalama	0,034	0,025	1,001	0,120

GLMNET için denge sonuçları

	SOE	OF	VO	KS
W₁	0,049	0,024	0,998	-
W₂	0,095	0,085	0,688	0,140
W₃	0,045	0,022	1,001	-
W₄	0,011	0,011	1,102	0,111
W₅	0,019	0,009	1,021	-

W₆	0,038	0,018	1,031	-
W₇	0,029	0,028	1,103	0,111
W₈	<0,001	<0,001	1,014	-
W₉	0,026	0,013	1,021	-
W₁₀	0,001	0,001	1,062	0,114
Mutlak Ortalama	0,031	0,021	1,004	0,119

NNET için denge sonuçları

	SOE	OF	VO	KS
W₁	0,073	0,036	1,001	-
W₂	0,037	0,032	0,798	0,126
W₃	0,054	0,027	1,006	-
W₄	-0,018	-0,018	1,113	0,110
W₅	-0,018	-0,009	1,016	-
W₆	0,002	0,001	1,024	-
W₇	0,067	0,065	1,109	0,113
W₈	0,008	0,004	1,021	-
W₉	-0,004	-0,002	1,021	-
W₁₀	0,002	0,002	1,094	0,108
Mutlak Ortalama	0,028	0,020	1,020	0,114

DBARTS için denge sonuçları

	SOE	OF	VO	KS
W₁	-0,059	-0,029	-0,029	-
W₂	0,104	0,093	0,093	0,137
W₃	-0,052	-0,026	-0,026	-
W₄	0,018	0,017	0,017	0,126
W₅	0,092	0,045	0,045	-
W₆	0,124	0,061	0,061	-
W₇	0,016	0,016	0,016	0,122
W₈	-0,014	-0,007	-0,007	-
W₉	0,076	0,037	0,037	-
W₁₀	-0,001	-0,001	-0,001	0,127
Mutlak Ortalama	0,055	0,033	0,021	0,128

QDA için denge sonuçları

	SOE	OF	VO	KS
W₁	0,014	0,007	1,018	-
W₂	0,006	0,005	1,085	0,131
W₃	0,006	0,003	1,019	-
W₄	-0,012	-0,012	1,086	0,128
W₅	0,001	0,001	1,026	-
W₆	0,018	0,009	1,032	-
W₇	0,038	0,037	1,135	0,130
W₈	0,004	0,002	1,024	-
W₉	-0,002	-0,001	1,023	-
W₁₀	0,006	0,006	1,096	0,132
Mutlak Ortalama	0,011	0,008	1,054	0,130

GAM için denge sonuçları

	SO _F	OF	VO	KS
W ₁	0,034	0,017	1,005	-
W ₂	0,052	0,046	0,878	0,119
W ₃	0,022	0,011	1,009	-
W ₄	0,034	0,033	1,086	0,113
W ₅	0,038	0,019	1,030	-
W ₆	0,035	0,017	1,032	-
W ₇	0,011	0,010	1,096	0,109
W ₈	0,003	0,002	1,016	-
W ₉	0,029	0,014	1,026	-
W ₁₀	0,002	0,001	1,060	0,113
Mutlak Ortalama	0,026	0,017	1,024	0,114

EARTH için denge sonuçları

	SO _F	OF	VO	KS
W ₁	0,045	0,022	1,005	-
W ₂	0,037	0,033	0,962	0,126
W ₃	0,041	0,020	1,010	-
W ₄	-0,022	-0,021	1,073	0,123
W ₅	-0,007	-0,004	1,019	-
W ₆	-0,018	-0,009	1,016	-
W ₇	0,042	0,041	1,115	0,118
W ₈	-0,003	-0,002	1,022	-
W ₉	-0,007	-0,003	1,020	-
W ₁₀	0,005	0,005	1,089	0,128
Mutlak Ortalama	0,023	0,016	1,033	0,124

STPAIC için denge sonuçları

	SO _F	OF	VO	KS
W ₁	0,046	0,023	1,000	-
W ₂	0,037	0,032	0,671	0,139
W ₃	0,047	0,023	1,006	-
W ₄	0,009	0,008	1,091	0,111
W ₅	0,016	0,008	1,020	-
W ₆	0,007	0,004	1,016	-
W ₇	0,034	0,034	1,084	0,110
W ₈	0,009	0,005	1,018	-
W ₉	0,013	0,006	1,019	-
W ₁₀	-0,005	-0,005	1,060	0,122
Mutlak Ortalama	0,022	0,015	0,999	0,121

IPREDBAGG için denge sonuçları

	SO _F	OF	VO	KS
W ₁	0,059	0,029	1,136	-
W ₂	0,270	0,241	1,328	0,258
W ₃	0,075	0,037	1,132	-
W ₄	0,224	0,221	1,347	0,250
W ₅	-0,056	-0,028	1,134	-
W ₆	0,162	0,079	1,271	-
W ₇	-0,378	-0,364	1,608	0,291
W ₈	-0,002	-0,001	1,117	-
W ₉	0,188	0,093	1,248	-

W₁₀	0,010	0,010	1,324	0,228
Mutlak Ortalama	0,142	0,110	1,265	0,257

GBM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,044	0,022	1,024	-
W₂	0,172	0,153	1,125	0,158
W₃	0,052	0,026	1,025	-
W₄	0,141	0,140	1,162	0,154
W₅	0,003	0,001	1,040	-
W₆	0,070	0,034	1,073	-
W₇	-0,158	-0,152	1,263	0,160
W₈	-0,007	-0,004	1,035	-
W₉	0,132	0,066	1,100	-
W₁₀	0,007	0,007	1,125	0,140
Mutlak Ortalama	0,079	0,061	1,097	0,153

CFOREST için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,069	-0,034	1,100	-
W₂	0,320	0,286	1,037	0,235
W₃	-0,072	-0,035	1,097	-
W₄	0,153	0,150	1,262	0,202
W₅	0,042	0,020	1,097	-
W₆	0,281	0,138	1,342	-
W₇	-0,154	-0,148	1,350	0,199
W₈	-0,039	-0,020	1,073	-
W₉	0,215	0,106	1,211	-
W₁₀	0,004	0,004	1,180	0,185
Mutlak Ortalama	0,135	0,094	1,175	0,205

EK-4. Denge Sonuçları (Senaryo C, n=500)

LR için denge sonuçları

	SOE	OF	VO	KS
W ₁	0,045	0,022	0,995	-
W ₂	0,076	0,068	0,729	0,122
W ₃	0,031	0,015	1,001	-
W ₄	0,162	0,142	0,590	0,156
W ₅	-0,003	-0,001	1,009	-
W ₆	0,046	0,023	1,026	-
W ₇	0,121	0,137	2,526	0,195
W ₈	<0,001	<0,001	1,010	-
W ₉	0,078	0,039	1,033	-
W ₁₀	-0,006	-0,006	1,049	0,105
Mutlak Ortalama	0,057	0,045	1,097	0,145

KSVM için denge sonuçları

	SOE	OF	VO	KS
W ₁	-0,142	-0,070	1,141	-
W ₂	0,145	0,132	1,129	0,188
W ₃	-0,112	-0,055	1,115	-
W ₄	0,135	0,118	1,020	0,186
W ₅	0,051	0,025	1,070	-
W ₆	0,168	0,083	1,164	-
W ₇	0,107	0,122	2,328	0,203
W ₈	-0,020	-0,010	1,057	-
W ₉	0,127	0,063	1,113	-
W ₁₀	-0,009	-0,009	1,279	0,167
Mutlak Ortalama	0,101	0,069	1,242	0,186

BAYESGLM için denge sonuçları

	SOE	OF	VO	KS
W ₁	0,031	0,015	0,999	-
W ₂	0,073	0,066	0,728	0,124
W ₃	0,038	0,019	1,001	-
W ₄	0,151	0,133	0,598	0,154
W ₅	0,005	0,003	1,011	-
W ₆	0,048	0,024	1,026	-
W ₇	0,115	0,131	2,507	0,195
W ₈	0,002	0,001	1,011	-
W ₉	0,076	0,038	1,032	-
W ₁₀	-0,003	-0,003	1,040	0,104
Mutlak Ortalama	0,054	0,043	1,095	0,144

GLMNET için denge sonuçları

	SOE	OF	VO	KS
W ₁	0,032	0,016	1,000	-
W ₂	0,080	0,072	0,733	0,125
W ₃	0,034	0,017	1,001	-
W ₄	0,147	0,128	0,591	0,153

W₅	-0,012	-0,006	1,008	-
W₆	0,048	0,024	1,028	-
W₇	0,124	0,141	2,516	0,198
W₈	0,001	0,001	1,010	-
W₉	0,073	0,036	1,032	-
W₁₀	-0,003	-0,003	1,047	0,105
Mutlak Ortalama	0,055	0,044	1,097	0,145

NNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,099	0,049	0,998	-
W₂	0,018	0,015	0,800	0,119
W₃	0,093	0,046	1,002	-
W₄	0,091	0,079	0,706	0,139
W₅	-0,019	-0,009	1,017	-
W₆	-0,001	<0,001	1,022	-
W₇	0,061	0,069	2,135	0,175
W₈	0,013	0,007	1,021	-
W₉	0,033	0,017	1,035	-
W₁₀	-0,003	-0,003	1,084	0,103
Mutlak Ortalama	0,043	0,029	1,082	0,134

DBARTS için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,150	-0,074	1,119	-
W₂	0,129	0,117	1,032	0,164
W₃	-0,135	-0,066	1,097	-
W₄	0,142	0,124	0,965	0,167
W₅	0,072	0,036	1,062	-
W₆	0,178	0,088	1,138	-
W₇	0,055	0,062	1,641	0,150
W₈	-0,030	-0,015	1,044	-
W₉	0,138	0,069	1,099	-
W₁₀	0,018	0,018	1,152	0,149
Mutlak Ortalama	0,105	0,067	1,135	0,158

QDA için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,035	-0,017	1,056	-
W₂	0,023	0,022	1,139	0,161
W₃	-0,033	-0,016	1,052	-
W₄	-0,004	-0,003	1,106	0,165
W₅	-0,013	-0,007	1,045	-
W₆	0,042	0,021	1,057	-
W₇	0,084	0,096	1,642	0,170
W₈	0,002	0,001	1,049	-
W₉	0,019	0,009	1,049	-
W₁₀	-0,002	-0,002	1,107	0,166
Mutlak Ortalama	0,026	0,019	1,130	0,166

GAM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,077	-0,038	1,054	-
W₂	0,127	0,116	0,952	0,146
W₃	-0,053	-0,026	1,043	-
W₄	0,184	0,161	0,864	0,156
W₅	0,065	0,032	1,045	-
W₆	0,126	0,062	1,083	-
W₇	0,075	0,085	1,728	0,151
W₈	-0,009	-0,005	1,026	-
W₉	0,115	0,057	1,067	-
W₁₀	-0,001	-0,001	1,077	0,132
Mutlak Ortalama	0,083	0,058	1,094	0,146

EARTH için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,002	-0,001	1,030	-
W₂	0,040	0,036	1,020	0,148
W₃	0,022	0,011	1,026	-
W₄	0,066	0,058	1,013	0,152
W₅	-0,019	-0,009	1,035	-
W₆	-0,003	-0,002	1,034	-
W₇	0,066	0,076	1,474	0,147
W₈	-0,011	-0,005	1,039	-
W₉	0,055	0,027	1,053	-
W₁₀	0,003	0,004	1,141	0,152
Mutlak Ortalama	0,029	0,023	1,087	0,150

STPAIC için denge sonuçları

	SOF	OF	VO	KS
W₁	0,035	0,017	0,998	-
W₂	0,030	0,027	0,725	0,121
W₃	0,043	0,021	1,003	-
W₄	0,106	0,092	0,599	0,150
W₅	-0,034	-0,017	1,010	-
W₆	0,006	0,003	1,013	-
W₇	0,137	0,157	2,484	0,201
W₈	0,008	0,004	1,012	-
W₉	0,050	0,025	1,029	-
W₁₀	-0,007	-0,007	1,044	0,113
Mutlak Ortalama	0,046	0,037	1,092	0,146

IPREDBAGG için denge sonuçları

	SOF	OF	VO	KS
W₁	0,014	0,007	1,190	-
W₂	0,162	0,148	1,294	0,260
W₃	0,075	0,037	1,171	-
W₄	0,144	0,126	1,268	0,257
W₅	-0,222	-0,111	1,169	-
W₆	0,083	0,041	1,217	-
W₇	-0,280	-0,319	1,485	0,301
W₈	<0,001	<0,001	1,153	-

W₉	0,107	0,053	1,217	-
W₁₀	0,019	0,019	1,368	0,248
Mutlak Ortalama	0,111	0,086	1,253	0,267

GBM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,044	-0,022	1,089	-
W₂	0,165	0,151	1,249	0,196
W₃	0,002	0,002	1,075	-
W₄	0,158	0,139	1,250	0,193
W₅	-0,042	-0,021	1,067	-
W₆	0,108	0,054	1,128	-
W₇	-0,027	-0,032	1,328	0,182
W₈	-0,011	-0,006	1,076	-
W₉	0,137	0,068	1,136	-
W₁₀	-0,004	-0,003	1,212	0,182
Mutlak Ortalama	0,070	0,050	1,161	0,188

CFOREST için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,271	-0,133	1,298	-
W₂	0,311	0,284	1,182	0,254
W₃	-0,244	-0,120	1,247	-
W₄	0,267	0,235	0,991	0,243
W₅	0,004	0,002	1,107	-
W₆	0,319	0,158	1,370	-
W₇	0,038	0,044	2,258	0,209
W₈	-0,077	-0,039	1,106	-
W₉	0,227	0,113	1,242	-
W₁₀	0,001	0,001	1,271	0,216
Mutlak Ortalama	0,176	0,113	1,307	0,231

EK-5. Denge Sonuçları (Senaryo A, n=1000)

LR için denge sonuçları

	SOF	OF	VO	KS
W₁	0,003	0,002	1,005	-
W₂	-0,001	-0,001	1,030	0,072
W₃	0,006	0,003	1,005	-
W₄	-0,007	-0,007	1,036	0,074
W₅	-0,001	-<0,001	1,006	-
W₆	-0,003	-0,002	1,005	-
W₇	0,006	0,006	1,048	0,070
W₈	<0,001	-<0,001	1,007	-
W₉	-0,001	-<0,001	1,007	-
W₁₀	-0,004	-0,004	1,025	0,074
Mutlak Ortalama	0,003	0,003	1,017	0,073

KSVM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,027	-0,013	1,025	-
W₂	0,043	0,042	1,141	0,099
W₃	-0,022	-0,011	1,022	-
W₄	0,013	0,012	1,145	0,097
W₅	0,019	0,010	1,024	-
W₆	0,063	0,031	1,043	-
W₇	-0,015	-0,014	1,174	0,095
W₈	-0,003	-0,001	1,017	-
W₉	0,033	0,016	1,025	-
W₁₀	<0,001	<0,001	1,154	0,097
Mutlak Ortalama	0,024	0,015	1,077	0,097

BAYESGLM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,003	0,001	1,005	-
W₂	-0,005	-0,005	1,032	0,073
W₃	0,002	0,001	1,007	-
W₄	-0,001	-0,001	1,035	0,072
W₅	-0,005	-0,003	1,004	-
W₆	-0,002	-0,001	1,005	-
W₇	0,006	0,007	1,050	0,069
W₈	0,001	0,001	1,008	-
W₉	0,001	<0,001	1,006	-
W₁₀	-0,001	-0,001	1,032	0,074
Mutlak Ortalama	0,003	0,002	1,018	0,072

GLMNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,008	0,004	1,004	-
W₂	-0,001	-0,002	1,039	0,072
W₃	0,006	0,003	1,005	-
W₄	-0,006	-0,006	1,036	0,072

W₅	-0,007	-0,004	1,005	-
W₆	0,002	0,001	1,006	-
W₇	0,001	0,002	1,062	0,070
W₈	0,002	0,001	1,006	-
W₉	0,006	0,003	1,008	-
W₁₀	-0,001	-0,001	1,023	0,074
Mutlak Ortalama	0,004	0,003	1,019	0,072

NNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,018	0,009	1,004	-
W₂	-0,015	-0,015	1,043	0,076
W₃	0,015	0,007	1,004	-
W₄	-0,024	-0,024	1,052	0,077
W₅	-0,014	-0,007	1,005	-
W₆	-0,012	-0,006	1,003	-
W₇	0,035	0,033	1,076	0,076
W₈	0,005	0,002	1,008	-
W₉	-0,016	-0,008	1,005	-
W₁₀	<0,001	-<0,001	1,049	0,077
Mutlak Ortalama	0,015	0,011	1,025	0,077

DBARTS için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,049	-0,024	1,025	-
W₂	0,031	0,031	1,070	0,074
W₃	-0,044	-0,022	1,022	-
W₄	0,006	0,006	1,070	0,071
W₅	0,051	0,025	1,026	-
W₆	0,065	0,032	1,035	-
W₇	-0,023	-0,022	1,100	0,073
W₈	-0,008	-0,004	1,010	-
W₉	0,043	0,022	1,021	-
W₁₀	0,001	0,001	1,058	0,073
Mutlak Ortalama	0,032	0,019	1,044	0,073

QDA için denge sonuçları

	SOF	OF	VO	KS
W₁	0,001	0,001	1,007	-
W₂	-0,002	-0,002	1,045	0,074
W₃	0,002	0,001	1,007	-
W₄	-0,008	-0,008	1,033	0,074
W₅	0,004	0,002	1,008	-
W₆	0,002	0,001	1,008	-
W₇	0,016	0,015	1,047	0,074
W₈	-0,001	-0,001	1,008	-
W₉	-0,002	-0,001	1,006	-
W₁₀	-0,002	-0,002	1,035	0,077
Mutlak Ortalama	0,004	0,003	1,020	0,075

GAM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,002	-0,001	1,006	-
W₂	-0,002	-0,002	1,027	0,069
W₃	-0,002	-0,001	1,007	-
W₄	0,003	0,003	1,031	0,071
W₅	<0,001	<0,001	1,007	-
W₆	0,002	0,001	1,007	-
W₇	0,002	0,003	1,048	0,066
W₈	0,001	0,001	1,007	-
W₉	0,001	<0,001	1,007	-
W₁₀	0,001	0,001	1,025	0,071
Mutlak Ortalama	0,002	0,001	1,017	0,069

EARTH için denge sonuçları

	SOF	OF	VO	KS
W₁	0,009	0,005	1,005	-
W₂	-0,021	-0,021	1,035	0,073
W₃	0,013	0,006	1,005	-
W₄	-0,015	-0,015	1,045	0,073
W₅	-0,012	-0,006	1,004	-
W₆	-0,030	-0,015	0,999	-
W₇	0,015	0,014	1,063	0,069
W₈	0,011	0,005	1,009	-
W₉	-0,005	-0,002	1,007	-
W₁₀	-0,006	-0,006	1,037	0,080
Mutlak Ortalama	0,014	0,010	1,021	0,074

STEPAIC için denge sonuçları

	SOF	OF	VO	KS
W₁	0,005	0,002	1,005	-
W₂	-0,039	-0,039	1,027	0,077
W₃	0,010	0,005	1,006	-
W₄	-0,003	-0,003	1,026	0,072
W₅	-0,001	-0,001	1,006	-
W₆	-0,035	-0,017	0,999	-
W₇	0,009	0,009	1,046	0,069
W₈	0,002	0,001	1,008	-
W₉	-0,001	<0,001	1,007	-
W₁₀	0,002	0,002	1,021	0,080
Mutlak Ortalama	0,011	0,008	1,015	0,075

IPREDBAGG için denge sonuçları

	SOF	OF	VO	KS
W₁	0,132	0,066	1,008	-
W₂	0,137	0,135	1,121	0,123
W₃	0,141	0,070	1,008	-
W₄	0,076	0,075	1,136	0,108
W₅	-0,140	-0,070	1,009	-
W₆	0,088	0,043	1,067	-
W₇	-0,279	-0,269	1,268	0,180
W₈	0,015	0,008	1,018	-
W₉	0,091	0,045	1,054	-

W₁₀	-0,003	0,003	1,103	0,096
Mutlak Ortalama	0,110	0,078	1,079	0,127

GBM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,022	0,011	1,007	-
W₂	0,097	0,096	1,084	0,091
W₃	0,032	0,016	1,005	-
W₄	0,077	0,076	1,087	0,085
W₅	-0,030	-0,015	1,004	-
W₆	0,044	0,022	1,027	-
W₇	-0,135	-0,130	1,139	0,106
W₈	-0,003	-0,002	1,012	-
W₉	0,070	0,035	1,032	-
W₁₀	<0,001	<0,001	1,066	0,077
Mutlak Ortalama	0,051	0,040	1,046	0,090

CFOREST için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,100	-0,049	1,089	-
W₂	0,275	0,271	1,249	0,197
W₃	-0,097	-0,048	1,082	-
W₄	0,185	0,181	1,223	0,170
W₅	0,092	0,046	1,089	-
W₆	0,288	0,142	1,264	-
W₇	-0,254	-0,244	1,386	0,201
W₈	-0,028	-0,014	1,054	-
W₉	0,220	0,109	1,162	-
W₁₀	-0,001	-0,001	1,174	0,142
Mutlak Ortalama	0,154	0,111	1,177	0,178

EK-6. Denge Sonuçları (Senaryo B, n=1000)

LR için denge sonuçları

	SOF	OF	VO	KS
W₁	0,040	0,020	0,993	-
W₂	0,093	0,083	0,658	0,106
W₃	0,029	0,014	0,997	-
W₄	0,018	0,017	1,064	0,072
W₅	0,020	0,010	1,013	-
W₆	0,034	0,017	1,021	-
W₇	0,033	0,032	1,056	0,072
W₈	0,003	0,002	1,007	-
W₉	0,015	0,007	1,011	-
W₁₀	0,004	0,004	1,025	0,073
Mutlak Ortalama	0,029	0,021	0,985	0,081

KSVM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,014	-0,007	1,021	-
W₂	0,067	0,060	1,008	0,103
W₃	-0,021	-0,010	1,024	-
W₄	0,021	0,021	1,147	0,099
W₅	0,041	0,020	1,033	-
W₆	0,078	0,038	1,059	-
W₇	-0,004	-0,004	1,203	0,096
W₈	-0,004	-0,002	1,016	-
W₉	0,045	0,022	1,032	-
W₁₀	-0,003	-0,004	1,165	0,099
Mutlak Ortalama	0,030	0,019	1,071	0,099

BAYESGLM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,041	0,020	0,993	-
W₂	0,096	0,086	0,658	0,106
W₃	0,033	0,016	0,997	-
W₄	0,025	0,024	1,045	0,073
W₅	0,022	0,011	1,014	-
W₆	0,035	0,017	1,022	-
W₇	0,039	0,038	1,069	0,073
W₈	0,006	0,003	1,006	-
W₉	0,020	0,010	1,012	-
W₁₀	0,001	0,001	1,024	0,074
Mutlak Ortalama	0,032	0,023	0,984	0,082

GLMNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,041	0,020	0,994	-
W₂	0,087	0,077	0,662	0,105
W₃	0,038	0,019	0,997	-
W₄	0,018	0,018	1,058	0,072

W₅	0,018	0,009	1,012	-
W₆	0,031	0,015	1,021	-
W₇	0,025	0,024	1,056	0,071
W₈	<0,001	<0,001	1,007	-
W₉	0,020	0,010	1,013	-
W₁₀	-0,003	-0,003	1,024	0,075
Mutlak Ortalama	0,028	0,020	0,984	0,081

NNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,047	0,023	0,995	-
W₂	0,039	0,035	0,780	0,097
W₃	0,046	0,022	0,996	-
W₄	-0,002	-0,002	1,084	0,080
W₅	-0,003	-0,001	1,009	-
W₆	0,003	0,002	1,011	-
W₇	0,051	0,049	1,068	0,081
W₈	0,008	0,004	1,008	-
W₉	0,006	0,003	1,011	-
W₁₀	-0,001	-0,002	1,048	0,079
Mutlak Ortalama	0,021	0,014	1,001	0,084

DBARTS için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,040	-0,020	1,026	-
W₂	0,071	0,063	0,954	0,084
W₃	-0,037	-0,018	1,023	-
W₄	0,029	0,029	1,071	0,076
W₅	0,070	0,034	1,040	-
W₆	0,084	0,041	1,056	-
W₇	0,003	0,003	1,108	0,074
W₈	-0,010	-0,005	1,012	-
W₉	0,059	0,029	1,032	-
W₁₀	<0,001	-0,001	1,054	0,076
Mutlak Ortalama	0,040	0,024	1,038	0,078

QDA için denge sonuçları

	SOF	OF	VO	KS
W₁	0,009	0,005	1,005	-
W₂	0,011	0,010	1,037	0,080
W₃	0,012	0,006	1,005	-
W₄	-0,005	-0,005	1,036	0,078
W₅	0,013	0,006	1,014	-
W₆	0,020	0,010	1,018	-
W₇	0,026	0,025	1,066	0,077
W₈	-0,003	-0,002	1,010	-
W₉	-0,001	<0,001	1,008	-
W₁₀	-0,001	-0,001	1,031	0,08
Mutlak Ortalama	0,010	0,007	1,023	0,079

GAM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,025	0,012	0,999	-
W₂	0,060	0,053	0,869	0,08
W₃	0,028	0,014	0,999	-
W₄	0,039	0,038	1,032	0,072
W₅	0,032	0,016	1,019	-
W₆	0,035	0,017	1,023	-
W₇	0,016	0,016	1,056	0,07
W₈	0,005	0,002	1,008	-
W₉	0,026	0,013	1,015	-
W₁₀	-0,001	-0,001	1,036	0,073
Mutlak Ortalama	0,027	0,018	1,006	0,074

EARTH için denge sonuçları

	SOF	OF	VO	KS
W₁	0,029	0,014	0,999	-
W₂	0,032	0,029	0,970	0,077
W₃	0,030	0,015	0,999	-
W₄	-0,004	-0,004	1,035	0,075
W₅	0,002	0,001	1,009	-
W₆	-0,006	-0,003	1,008	-
W₇	0,024	0,024	1,064	0,072
W₈	0,006	0,003	1,010	-
W₉	0,001	0,001	1,010	-
W₁₀	-0,002	-0,002	1,059	0,082
Mutlak Ortalama	0,014	0,010	1,016	0,077

STEPAIC için denge sonuçları

	SOF	OF	VO	KS
W₁	0,037	0,018	0,994	-
W₂	0,013	0,012	0,658	0,104
W₃	0,027	0,013	0,998	-
W₄	0,021	0,020	1,052	0,071
W₅	0,023	0,011	1,015	-
W₆	0,014	0,007	1,013	-
W₇	0,027	0,026	1,056	0,071
W₈	0,004	0,002	1,008	-
W₉	0,014	0,007	1,011	-
W₁₀	0,001	0,001	1,020	0,083
Mutlak Ortalama	0,018	0,012	0,983	0,082

IPREDBAGG için denge sonuçları

	SOF	OF	VO	KS
W₁	0,111	0,055	1,013	-
W₂	0,196	0,175	1,084	0,138
W₃	0,142	0,070	1,008	-
W₄	0,091	0,090	1,136	0,116
W₅	-0,096	-0,047	1,010	-
W₆	0,104	0,051	1,093	-
W₇	-0,249	-0,240	1,290	0,175
W₈	0,002	0,001	1,022	-
W₉	0,100	0,050	1,068	-

W₁₀	-0,008	-0,008	1,131	0,099
Mutlak Ortalama	0,110	0,079	1,086	0,132

GBM için denge sonuçları

	SOF	OF	VO	KS
W₁	0,022	0,011	1,005	-
W₂	0,130	0,116	1,093	0,103
W₃	0,022	0,011	1,006	-
W₄	0,101	0,100	1,091	0,094
W₅	0,025	0,012	1,022	-
W₆	0,063	0,031	1,045	-
W₇	-0,107	-0,103	1,159	0,104
W₈	-0,003	-0,001	1,013	-
W₉	0,094	0,047	1,049	-
W₁₀	-0,001	-0,001	1,077	0,081
Mutlak Ortalama	0,057	0,043	1,056	0,096

CFOREST için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,123	-0,060	1,105	-
W₂	0,291	0,260	1,098	0,195
W₃	-0,104	-0,051	1,096	-
W₄	0,170	0,167	1,218	0,168
W₅	0,099	0,049	1,098	-
W₆	0,276	0,135	1,302	-
W₇	-0,210	-0,202	1,358	0,184
W₈	-0,044	-0,022	1,049	-
W₉	0,212	0,105	1,172	-
W₁₀	<0,001	<0,001	1,157	0,138
Mutlak Ortalama	0,153	0,105	1,165	0,171

EK-7. Denge Sonuçları (Senaryo C, n=1000)

LR için denge sonuçları

	SOF	OF	VO	KS
W ₁	0,036	0,018	0,992	-
W ₂	0,078	0,071	0,713	0,094
W ₃	0,032	0,016	0,995	-
W ₄	0,154	0,134	0,582	0,124
W ₅	<0,001	-<0,001	1,004	-
W ₆	0,046	0,023	1,021	-
W ₇	0,121	0,138	2,466	0,174
W ₈	0,003	0,001	1,005	-
W ₉	0,071	0,035	1,024	-
W ₁₀	-0,004	-0,004	1,023	0,069
Mutlak Ortalama	0,055	0,044	1,083	0,115

KSVM için denge sonuçları

	SOF	OF	VO	KS
W ₁	-0,134	-0,066	1,103	-
W ₂	0,104	0,095	1,101	0,130
W ₃	-0,092	-0,046	1,066	-
W ₄	0,090	0,079	1,045	0,123
W ₅	0,046	0,023	1,040	-
W ₆	0,135	0,067	1,097	-
W ₇	0,089	0,101	1,916	0,148
W ₈	-0,018	-0,009	1,028	-
W ₉	0,082	0,041	1,054	-
W ₁₀	0,003	0,003	1,199	0,114
Mutlak Ortalama	0,079	0,053	1,165	0,129

BAYESGLM için denge sonuçları

	SOF	OF	VO	KS
W ₁	0,041	0,020	0,991	-
W ₂	0,083	0,075	0,706	0,095
W ₃	0,037	0,018	0,994	-
W ₄	0,156	0,136	0,584	0,125
W ₅	-0,001	-0,001	1,004	-
W ₆	0,051	0,025	1,023	-
W ₇	0,119	0,136	2,469	0,175
W ₈	0,004	0,002	1,005	-
W ₉	0,074	0,037	1,026	-
W ₁₀	0,001	0,001	1,022	0,068
Mutlak Ortalama	0,057	0,045	1,082	0,116

GLMNET için denge sonuçları

	SOF	OF	VO	KS
W ₁	0,030	0,015	0,995	-
W ₂	0,081	0,074	0,710	0,096
W ₃	0,037	0,018	0,995	-

W₄	0,155	0,135	0,581	0,125
W₅	-0,007	-0,004	1,004	-
W₆	0,046	0,023	1,021	-
W₇	0,123	0,140	2,480	0,176
W₈	0,006	0,003	1,004	-
W₉	0,075	0,037	1,026	-
W₁₀	0,001	0,001	1,023	0,070
Mutlak Ortalama	0,056	0,045	1,084	0,117

NNET için denge sonuçları

	SOF	OF	VO	KS
W₁	0,087	0,043	0,989	-
W₂	0,055	0,050	0,763	0,101
W₃	0,075	0,037	0,995	-
W₄	0,133	0,116	0,678	0,129
W₅	-0,013	-0,006	1,008	-
W₆	0,021	0,011	1,018	-
W₇	0,057	0,065	1,959	0,146
W₈	0,007	0,004	1,010	-
W₉	0,059	0,029	1,028	-
W₁₀	-0,004	-0,004	1,039	0,080
Mutlak Ortalama	0,051	0,037	1,049	0,114

DBARTS için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,116	-0,057	1,075	-
W₂	0,097	0,088	1,012	0,109
W₃	-0,108	-0,053	1,064	-
W₄	0,100	0,088	0,977	0,107
W₅	0,052	0,026	1,034	-
W₆	0,137	0,067	1,085	-
W₇	0,022	0,025	1,399	0,095
W₈	-0,016	-0,008	1,021	-
W₉	0,079	0,039	1,046	-
W₁₀	0,007	0,007	1,086	0,099
Mutlak Ortalama	0,073	0,046	1,080	0,103

QDA için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,042	-0,021	1,035	-
W₂	0,016	0,015	1,107	0,102
W₃	-0,052	-0,025	1,038	-
W₄	0,001	0,001	1,094	0,104
W₅	0,014	0,007	1,024	-
W₆	0,036	0,018	1,034	-
W₇	0,071	0,082	1,480	0,114
W₈	-0,006	-0,003	1,022	-
W₉	0,019	0,010	1,025	-
W₁₀	0,010	0,010	1,051	0,105
Mutlak Ortalama	0,027	0,019	1,091	0,106

GAM için denge sonuçları

	SOF	OF	VO	KS
W ₁	-0,081	-0,040	1,046	-
W ₂	0,127	0,116	0,920	0,102
W ₃	-0,057	-0,028	1,032	-
W ₄	0,179	0,157	0,841	0,115
W ₅	0,071	0,035	1,033	-
W ₆	0,128	0,063	1,070	-
W ₇	0,062	0,071	1,629	0,110
W ₈	-0,007	-0,004	1,013	-
W ₉	0,119	0,059	1,054	-
W ₁₀	-0,007	-0,007	1,047	0,087
Mutlak Ortalama	0,084	0,058	1,069	0,104

EARTH için denge sonuçları

	SOF	OF	VO	KS
W ₁	-0,007	-0,003	1,016	-
W ₂	0,053	0,048	1,000	0,098
W ₃	0,005	0,003	1,014	-
W ₄	0,058	0,051	1,003	0,102
W ₅	0,007	0,003	1,018	-
W ₆	0,014	0,007	1,024	-
W ₇	0,063	0,072	1,361	0,100
W ₈	0,003	0,002	1,018	-
W ₉	0,055	0,027	1,035	-
W ₁₀	-0,003	-0,003	1,055	0,102
Mutlak Ortalama	0,027	0,022	1,054	0,101

STEPAIC için denge sonuçları

	SOF	OF	VO	KS
W ₁	0,036	0,018	0,993	-
W ₂	0,038	0,035	0,711	0,092
W ₃	0,030	0,015	0,996	-
W ₄	0,143	0,125	0,582	0,123
W ₅	-0,004	-0,002	1,005	-
W ₆	0,012	0,006	1,009	-
W ₇	0,121	0,138	2,468	0,174
W ₈	0,005	0,002	1,006	-
W ₉	0,058	0,029	1,021	-
W ₁₀	0,002	0,002	1,020	0,077
Mutlak Ortalama	0,045	0,037	1,081	0,117

IPREDBAGG için denge sonuçları

	SOF	OF	VO	KS
W ₁	0,034	0,017	1,037	-
W ₂	0,077	0,070	1,025	0,128
W ₃	0,086	0,043	1,022	-
W ₄	0,107	0,094	1,011	0,128
W ₅	-0,251	-0,125	1,045	-
W ₆	0,003	0,002	1,037	-
W ₇	-0,193	-0,221	1,271	0,149
W ₈	0,009	0,005	1,028	-
W ₉	0,076	0,038	1,057	-

W₁₀	-0,005	-0,005	1,127	0,118
Mutlak Ortalama	0,084	0,062	1,066	0,131

GBM için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,036	-0,017	1,040	-
W₂	0,122	0,111	1,162	0,122
W₃	-0,026	-0,013	1,035	-
W₄	0,121	0,106	1,175	0,122
W₅	-0,025	-0,012	1,023	-
W₆	0,076	0,038	1,058	-
W₇	-0,014	-0,015	1,200	0,106
W₈	-0,008	-0,004	1,027	-
W₉	0,109	0,054	1,065	-
W₁₀	0,002	0,002	1,117	0,108
Mutlak Ortalama	0,054	0,037	1,090	0,115

CFOREST için denge sonuçları

	SOF	OF	VO	KS
W₁	-0,320	-0,157	1,356	-
W₂	0,291	0,265	1,243	0,222
W₃	-0,290	-0,143	1,280	-
W₄	0,196	0,172	1,114	0,197
W₅	0,034	0,017	1,097	-
W₆	0,306	0,151	1,324	-
W₇	0,009	0,010	2,007	0,171
W₈	-0,067	-0,033	1,087	-
W₉	0,183	0,091	1,188	-
W₁₀	-0,010	-0,010	1,259	0,177
Mutlak Ortalama	0,171	0,105	1,296	0,192

ÖZGEÇMİŞ

I-Bireysel Bilgiler

Adı : Emre
Soyadı : DEMİR
Doğum yeri ve tarihi : Alaca/1982
Uyruğu : T.C.
Medeni durumu : Evli
Askerlik durumu : Yaptı
İletişim adresi ve telefonu : Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD
ÇORUM, Tel: 0364 222 11 00 / 5025

II- Eğitimi

Doktora : Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD
(2015-2019)
Y. Lisans : Muğla Üniversitesi, Fen Fakültesi, İstatistik Bölümü
(2011-2014)
Lisans ve Y.Lisans Derecesi : Karadeniz Teknik Üniversitesi, Eğitim Fakültesi,
Matematik (2000 – 2005)

III-Ünvanları

Öğretim Görevlisi : 2012- Halen

IV- Mesleki Deneyimi

Öğretim Görevlisi : Hitit Üniversitesi (2012- ...)
Müdür Yardımcısı : Hitit Üniversitesi Geleneksel ve Tamamlayıcı Tıp
Uygulama ve Araştırma Merkezi

V- Üye Olduğu Bilimsel Kuruluşlar

International Society for Clinical Biostatistics (2016- ...)

VI- Bilimsel Etkinlikleri

Uluslararası hakemli dergilerde yayımlanan yayınları

1. **Demir E** (2018). The Evolution of Spirituality, Religion and Health Publications: Yesterday, Today and Tomorrow. *J Relig Health*, <https://doi.org/10.1007/s10943-018-00739-w>
2. Comba A, **Demir E**, Eren NB (2018). Nutritional status and related factors of schoolchildren in Çorum, Turkey. *Public Health Nutrition*, 1-10. doi:10.1017/S1368980018002938
3. Ozsoy Z, **Demir E**, (2018). The Evolution of Bariatric Surgery Publications and Global Productivity: A Bibliometric Analysis. *OBES SURG*. 28(4): 1117-1129.
4. Ozsoy Z, **Demir E**, (2018). Which Bariatric Procedure Is the Most Popular in the World? A Bibliometric Comparison. *OBES SURG*, 28(8): 2339-2352.
5. Comba A, Eren NB, **Demir E**, (2018). Prevalence of celiac disease among school-age children in Çorum, Turkey. *Turk J Gastroenterol*, 29: 595-600.
6. Senel E, **Demir E**, (2018). Bibliometric analysis of apitherapy in complementary medicine literature between 1980 and 2016, *Complementary Therapies in Clinical Practice*, 31: 47-52.
7. Diken ÖE, Diken Aİ, Yalçınkaya A, Gülbay B, Acıcan T, **Demir E**, Özyalçın S, Erol ME (2018). Predictive Value of STOP-BANG on OSAS-Related Complications Following Coronary Artery Bypass Grafting. *Respiratory Care*, 63(10): 1264-1270.
8. Senel E, **Demir E** (2018). Bibliometric and Scientometric Analysis of the Articles Published in the Journal of Religion and Health Between 1975 and 2016, *J Relig Health*, 57(4): 1473-82.
9. Diken ÖE, Arslan S, Akdoğan Ö, Yapar D, Ünal Ö, **Demir E**, Baykam N. (2018). Clinical, radiological and prognostic features of influenza cases in the influenza epidemic during years 2016-2017. *Tuberk Toraks*, 66(2):144-149.
10. Senturk AB, Yaytokgil M, Yilmaz AH, Ekici M, Aydın C, **Demir E**, Cakiroglu B (2018). Can platelets be the early biomarkers of erectile dysfunction? *Journal of Pakistan Medical Association*, 68(4): 515–518.

11. Yazla E, Karadere ME, Küçükler FK, Karşıdag Ç, Inanç L, Kankoç E, Dönertaş M, **Demir E** (2017). The Effect of Religious Belief and Forgiveness on Coping with Diabetes. *J Relig Health*, 57:1010–9.
12. Senturk AB, Yaytokgil M, Ekici M, Sarı S, **Demir E**, Çakiroğlu B (2017). Comparing use of lidocaine periprostatic nerve block and diclofenac suppository alone for patients undergoing transrectal ultrasound guided prostate biopsy. *Central European journal of urology*, 71(1): 38-42.
13. Senel E, **Demir E**, Alkan RM (2017). Bibliometric analysis on global Behcet disease publications during 1980-2014: is there a Silk Road in the literature? *J Eur Acad Dermatol Venereol*, 31(3); 518-522.
14. Ekinçi A, Karataş D, Yetiş A, **Demir E**, Özcan M (2017). Serum Oxidative Stress Levels in Patients with Nasal Septal Deviation, *Türk Arch Otorhinolaryngol*, 55: 125-8.
15. Diken AI, Yalçınkaya A, Hanedan O, Günertem E, **Demir E**, Erol E (2017). Effect of tandem lesions on hemodynamic parameters: an experimental study. *Kardiologia Polska*, 75(12):1315-1323.
16. Özsoy Z, Özsoy Ş, Gevrek F, **Demir E**, Benli I, Daldal E, Yenidogan E (2017). Effect of bevacizumab on acetic acid-induced ulcerative colitis in rats, *Journal of Surgical Research*, 216; 191-200.
17. Afsarlar CE, Elizondo R, Yilmaz E, Cakmakci E, Ballow DJ, **Demir E**, Guney G, Koh CJ (2017). Ultrasonographic findings in the epididymis of pediatric patients with testicular torsion, *Journal of Pediatric Urology*, 13(4): 393. e1-393. e6
18. Ekinçi A, Karataş D, Yetiş A, **Demir E**, Özcan M (2017). The effects of septoplasty surgery on serum oxidative stress levels. *Eur Arch Otorhinolaryngol*, 274 (7): 2799-2802.
19. Karadere ME, Şafak Y, Özçelik HŞ, **Demir E**, Türkçapar MH (2017). Inductive Reasoning in Patients with Paranoid Type Schizophrenia. *JCBPR*, 6(2):67–74.
20. Caglar C, **Demir E**, Kucukler FK, Durmus M (2016). A bibliometric analysis of academic publication on diabetic retinopathy disease trends during 1980-2014: a global and medical view. *International Journal of Ophthalmology*, 9(11): 1663-1668.

21. Şenel E, **Demir E** (2016). An Update and Evaluation of Telemedicine and Teledermatology Publications for 2014, *Skinmed*, 14(2): 157-158.
22. Akkus Ö, **Demir E** (2016). Comparison of Some Classical and Meta-Heuristic Optimization Techniques in the Estimation of the Logit Model Parameters, *International journal of advanced research*, 4(10): 1026-1042.
23. Çağlar Ç, Kocamış Sİ, **Demir E** (2016). Comparison of the measurements of a novel optical biometry: Nidek AL-Scan with Sirius and a ultrasound biometry. *International Ophthalmology*, 37(3): 491-498.
24. Cakir OO, Yilmaz A, **Demir E**, Incekara K, Kose MO, Nagehan E (2016). Association of the BsmI, ApaI, TaqI, Tru9I and FokI Polymorphisms of the Vitamin D Receptor Gene with Nephrolithiasis in the Turkish Population, *The Journal of Urology*, 13(1): 2509-2518.
25. Yentur S, Cakir O, Oncu M, Altın M. **Demir E**, Kandirali I. Semercioz A (2016). Efficacy of Pelvic Diffusion Weighted MRI Prior to Prostate Biopsy in Patients with tPSA Level of 2,5-20 ng/ml for Determination of Malign Lesions, *The Internet Journal of Urology*, 14 (1):1-6.
26. Şenel E, **Demir E** (2015). A global productivity and bibliometric analysis of telemedicine and teledermatology publication trends during 1980–2013. *Dermatologica Sinica*, 33(1): 16-20.
27. **Demir E**, Akkuş Ö (2015). An Introductory Study on “How the Genetic Algorithm Works in the Parameter Estimation of Binary Logit Model?” *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, 19(2):162-180.
28. Şenel E, Sabancılar E, Mansuroğlu C, **Demir E** (2014). A preliminary study of the contribution of telemicroscopy to the diagnosis and management of skin tumours in teledermatology, *Journal of Telemedicine and Telecare*, 20(4): 178–183.
29. **Demir E** (2014). A comparison of classical and fuzzy data envelopment analyses in measuring and evaluating school activities, *TJFS*, 5(1): 37-58.

Ulusal hakemli dergilerde yayımlanan yayınları

1. Akça S, Selen F, **Demir E**, Demir T (2018). Cinsiyet ve Yaş Farklılıklarının Ergenlerin Depresyon, Anksiyete Bozukluğu, Kendine Zarar Verme, Psikoz, Travma Sonrası Stres Bozukluğu, Alkol-Uyuşturucu Bağımlılığı ve Dikkat

- Eksikliği Hiperaktivite Bozukluğu İle İlişkili Sorunlara Etkisi. *Dicle Tıp Dergisi*, 45(3): 255-264.
2. Arslan S, **Demir E**, Tümtürk A, Diken ÖE (2017). Göğüs Hastalıkları Hastanesinde Sağlık Bakımı İlişkili Pnömoni Hastalarının Değerlendirilmesi, *Bozok Tıp Derg*, 7(2):51-8.
 3. **Demir E**, Yavuz Y (2016). Prostat Kanseri Tanısında PSA İçin İşlem Karakteristik Eğrisi Analizinde Ortak Değişken Düzeltmesi, *Türkiye Klinikleri J Biostat*, 8(2):162-71.
 4. Akkus Ö, **Demir E** (2016). İki düzeyli olasılık modellerinde klasik ve meta sezgisel optimizasyon tekniklerinin performansı üzerine bir çalışma, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 15(30): 107-131.
 5. Akmeşe ÖF, **Demir E**, Dünder E (2016). Student Perceptions for Distance Education and Efficiency Analysis of the System, *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(2): 981-997.
 6. Kör H, Erbay H, **Demir E**, Akmeşe Ö (2016). Uzaktan ve Örgün Eğitim Öğrencilerinin Akademik Başarı Performanslarını Etkileyen Faktörlerin Araştırılması. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(2): 1029-1050.

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler

1. **Demir E**, Akça BA, Hayme S, Köse SK. A New Mobile Application for Sample Size Calculations in Medical Research, XX. Ulusal ve III. Uluslararası Biyoistatistik Kongresi, 26-29 Ekim 2018, Gaziantep/Türkiye.
2. Arslan S, Güven G, Akyüz ÜA, **Demir E**, Demirci B. Effects of local decongestive agents at trekea and lung. XX. International Conference on Allergy, Astma and Immunology, 14 Mayıs 2018, Londra, UK.
3. **Demir E**, Yavuz Y, Köse SK. Bağımlı Değişken İle Ortak Değişken Arasındaki Korelasyona Göre Kovaryans Analizinin Gücü, 2. Uluslararası Biyoistatistik Kongresi (19. Ulusal Biyoistatistik Kongresi), 25-28 Ekim 2017, Belek, Antalya
4. **Demir E**, Hayme S, Gökmen D, Elhan AH, Köse SK. Gözlemsel Çalışmalarda Ölçülemeyen Karıştırıcı Değişkenlerin Etkisi, 2. Uluslararası Biyoistatistik Kongresi (19. Ulusal Biyoistatistik Kongresi), 25-28 Ekim 2017, Belek, Antalya

5. Sıddıkoğlu D, Özalp Ateş FH, **Demir E**, Yavuz Y. Bağımlı Değişkenin Sürekli ve Tekrarlı Olduğu Denemelerde Özet İstatistikler Kullanılarak Örneklem Büyüklüğünün Hesaplanması, 2. Uluslararası Biyoistatistik Kongresi (19. Ulusal Biyoistatistik Kongresi), 25-28 Ekim 2017, Belek, Antalya
6. **Demir E**, Hayme S, Kose SK. A Systematic Review and Bibliometric Analysis on Propensity Score Obtained by using Machine Learning Algorithms. 6th IBS Channel Network Conference, April 24-26, 2017, Hasselt University, Diepenbeek, Belgium
7. **Demir E**, Kose SK. The Most Compatible Machine Learning Algorithm with Propensity Score Estimates Obtained by Logistic Regression Analysis. 6th IBS Channel Network Conference, April 24-26, 2017, Hasselt University, Diepenbeek, Belgium
8. **Demir E**, Karasartova D, Güreser AS, Özkan AT, Mumcuoğlu K. A bibliometric analysis of maggot/larval therapy research during 1975-2016. X. International Conference on Biotherapy, Istanbul, 4-9 Ekim 2017.
9. **Demir E**, Güreser AS, Karasartova D, Özkan AT, Mumcuoğlu K. Scientometric analysis of literature on leeches and hirudotherapy between the years 1975 and 2016. X. International Conference on Biotherapy, Istanbul, 4-9 Ekim 2017.
10. Özakar Akça S, Selen F, **Demir E**, Demir T. Ergenlerin Ruhsal Sorun Risklerine Göre Cinsiyet ve Yaş Farklılıkları. I. Uluslararası Pediatri Hemşireliği Kongresi, 29 Kasım - 2 Aralık 2017, Belek – Antalya.
11. **Demir E**, Köse SK. Randomize Olmayan Klinik Çalışmalarda Propensity Skor Tahmini İçin Makine Öğrenme Yaklaşımı, 1. Uluslararası Biyoistatistik Kongresi (18. Ulusal Biyoistatistik Kongresi), 26-29 Ekim 2016, Belek, Antalya
12. **Demir E**, Köse SK. Biyoistatistikçiler için R paketine alternatif yeni bir paket; Julia, 1. Uluslararası Biyoistatistik Kongresi (18. Ulusal Biyoistatistik Kongresi), 26-29 Ekim 2016, Belek, Antalya.
13. Hayme S, Öztuna D, Elhan AH, **Demir E**. Klinik Denemelerde Kesirli Polinomların İki Kategorili Lojistik Regresyon Modelinde Kullanımı, 1. Uluslararası Biyoistatistik Kongresi (18. Ulusal Biyoistatistik Kongresi), 26-29 Ekim 2016, Belek, Antalya.

14. **Demir E**, Köse SK. Comparison Of Survival Curves; A Simulation Study, 37th Annual Conference of the International Society for Clinical Biostatistics, 21 - 25 August 2016, Birmingham, UK.
15. **Demir E**, Yılmaz A, Köse SK. An information to bioinformatic prediction of protein multiple sequence alignment in R package, İnternational conference on information complexity and statistical modeling in high dimensions with applications (IC-SMHD) May 18-21, 2016, Cappadocia, Nevşehir Turkey
16. **Demir E**, Köse SK. Çorum İli Sağlık İstatistikleri ve Etkinlik Değerlendirmesi, Bütün Yönleri İle Çorum "Dün, Bugün, Yarın ve Değişim" Uluslararası Sempozyum, 28-30 Nisan 2016, Çorum.
17. Afsarlar CE, Elizondo R, Yılmaz E, Cakmakci E, Ballow DJ, **Demir E**, Koh CJ Prognostic ultrasonographic findings in the epididymis of pediatric patients with testicular torsion, Pediatric Urology Fall Congress, 9 Eylül 2016, Dallas, Amerika.
18. Kör H, Erbay H, **Demir E**, Ergüzen A. Research of Distance Education Students Cloud Perceptions and Use Cases, II. International Dynamic, Explorative and Active Learning (IDEAL) Conference, Amasya Üniversitesi, 5-7 November 2015, Amasya.
19. Caglar C, Zirig HI, Doguizi S, **Demir E**, Durmuş M. Evaluation of a new optical biometry device for measurement of ocular biometric parameters and its comparison with conventional ultrasound biometer. XXXIII. Congress of the ESCRS. 5-9 Eylül, 2015, Barcelona, İspanya.
20. **Demir E**, Dünder E, Cengiz MA. Detection The Factors Effecting Diabet Patients HbA1c Measurements Using Ordinal Logistic Regression And Genetic Algorithm, 8th conference of the Eastern Mediterranean Region of the International Biometric Society (EMR-IBS), 11-15 May 2015, Cappadocia, Nevsehir.
21. Kör H, Erbay H, **Demir E**. The Research of Factors That May Affect The Success of Distance Education Students on Their Vocational Courses With The Method of Logistic Regression Analysis, International Conference on Advances in Statistics, 24-28 March, 2015, Sharjah/DUBAİ/UAE
22. Akmeşe ÖF, **Demir E**, Erbay H. Statistical Analysis of Students' Opinions About The Efficiency of Web-Based Distance Education System, International Conference on Advances in Statistics, 24-28 March, 2015, Sharjah/DUBAİ/UAE

23. **Demir E**, Dunder E, Cengiz MA. A Comprasion Of Some Meta-Heuristics Optimization Methods For Parameter Estimation In Logit Model With An Application To Vitiligo Disease Risk Factor, International Conference on Advances in Statistics, 24-28 March, 2015, Sharjah/DUBAI/UAE
24. Şenel E, Acar B, **Demir E**. Mean platelet volume in patients with psoriasis. 23rd EADV Congress, P1709, 8-12 Oct, 2014, Amsterdam, The Netherlands.
25. Şenel E, Sabancılar E, Mansuroğlu C, **Demir E**. A novel tool in teledermatology: Telemicroscopy. 23rd EADV Congress, P1319, 8-12 Oct, 2014, Amsterdam, The Netherlands
26. Akmeşe ÖF, **Demir E**. "Uzaktan Eğitim Sistemi Verimlilik Analizi: Hitit Üniversitesi Örneği" 8. ICITS, 18-20 Eylül 2014, Edirne, Türkiye.
27. Şenel E, Acar B, **Demir E**. Mean platelet volume in patients with aphthous stomatitis and Behçet disease. 23rd EADV Congress, P1132, 8-12 Oct, 2014, Amsterdam, The Netherlands
28. **Demir E**, Akkuş Ö. A Comparison of the Parameter Estimations of Newton-Raphson with the Methods of Nelder-Mead and Genetic Algorithm for the Categorical Dependent Variable Models, 15th. International Symposium on Econometrics, Operations Research and Statistics, 22-25 May 2014, Suleyman Demirel University, Isparta
29. **Demir E**, Akkuş Ö. Genetic Algorithm Approach to the Parameter Estimation in the Analysis of Categorical Dependent Variable Models, 9th International Statistics Day Symposium, 10-14 May, 2014, Side-Antalya.
30. **Demir E**, Demir S. Effectiveness Measurement of High Schools in Çorum by Fuzzy Data Envelopment Analysis, 8th International Congress of Statistics, 27-30 October, 2013, Kemer-Antalya.

Ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler

1. Kendirci HNP, Karayurt Ü, **Demir E**. Tip 1 Diyabetli Çocuk/Adolesan Hasta ve Ebeveynlerinin Diyabet Bilgi Düzeyi ile Metabolik Kontrol Arasındaki İlişkinin Değerlendirilmesi, XXII. Ulusal Pediatrik Endokrinoloji ve Diyabet Kongresi, 18-22 Nisan 2018, Belek – Antalya.

2. Kendirci HNP, Comba A, **Demir E.** Çölyak Hastalığı Tanısı Alan Çocuklarda Kemik Mineral Yoğunluğu ve Kemik Metabolizması Belirteçlerinin Değerlendirilmesi, XXII. Ulusal Pediatrik Endokrinoloji ve Diyabet Kongresi, 18-22 Nisan 2018, Belek – Antalya.
3. Kendirci HNP, **Demir E.** Pubertal Jinekomastili Olguların Değerlendirilmesi, XXII. Ulusal Pediatrik Endokrinoloji ve Diyabet Kongresi, 18-22 Nisan 2018, Belek – Antalya.
4. Comba A, **Demir E.**, Eren NB. Okul Çağı Çocuklarında Malnütrisyon Sıklığının Belirlenmesi, 61. Türkiye Milli Pediatri Kongresi, 15-19 Kasım 2017, Antalya.
5. Comba A, **Demir E.**, Eren NB. Çorum İlinde Okul Çağı Çocuklarında Çölyak Hastalığı Sıklığının Araştırılması, 61. Türkiye Milli Pediatri Kongresi, 15-19 Kasım 2017, Antalya.
6. Muslu Ü, Şenel E, **Demir E.** Çorum Yöresindeki Deri Biyopsilerinin Değerlendirilmesi: İki Yıllık Retrospektif Bir Analiz, 23. Prof.Dr. Lütfü Tat Sempozyumu, 22-26 Kasım 2017, Ankara.
7. Diken OE, Diken AI, Yalçınkaya A, Gülbay B, Acıcan T, **Demir E.** Preoperatif STOP-BANG Anketinin Obstruktif Uyku Apne Sendromu ile İlişkili Olabilecek Postoperatif Pulmoner Değişiklikleri Öngörmedeki Değeri, 6. Uyku Bozuklukları Kongresi 26-30 Nisan 2017 Bodrum/Muğla.
8. Diken OE, Aydemir Y, **Demir E.** Astımda mikroRNA ve astım fenotipleri ile ilişkisi, Ulusal Akciğer Sağlığı Kongresi 15-19 Mart 2017 Belek Antalya.
9. Arslan S, Karavelioğlu Y, **Demir E.** Silikoziste kardiyopulmoner egzersiz testi: Silikozis hastalarının performansını değerlendirmede bir yöntem Türk Toraks Derneği 20. Yıllık Kongresi, 5 – 9 Nisan 2017, Belek, Antalya.
10. Diken OE, **Demir E.** Göğüs hastalıkları servisinde yatan hastalarda solunum fizyoterapisinin hastaların dispnelerini nasıl etkilediğinin geçici dispne indeksi ile değerlendirilmesi, Türkiye Solunum Araştırmaları Derneği 38. Ulusal Kongresi 15-19 Ekim 2016 Çeşme, İzmir.
11. Diken OE, **Demir E.** Ücretsiz sigara bırakma ilacı dağıtımı politikası sigara bıraktırma polikliniğine başvuran hastalarda bırakma oranlarını etkiler mi? Türkiye Solunum Araştırmaları Derneği 38. Ulusal Kongresi 15-19 Ekim 2016 Çeşme, İzmir.

12. Diken ÖE, Erdemli HK, Aydemir Y, Şen F, **Demir E.** Akciğer Kanseri Hastalarında, Serum Galektin 1 Ve 3 Düzeylerinin, Tanısal Ve Prognostik Değeri? Türkiye Solunum Araştırmaları Derneği 38. Ulusal Kongresi 15-19 Ekim 2016 Çeşme, İzmir.
13. **Demir E**, Turgal E, Köse SK, Dünder E, Cengiz MA, Şenel E. Sağlık Hizmetlerinin Etkinlik Ölçümünde Klasik, Bulanık Veri Zarflama Analizi ve Stokastik Sınır Analizi Yöntemlerinin Karşılaştırılması, 17. Ulusal Biyoistatistik Kongresi, 5-9 Kasım 2015, Girne/KIBRIS.
14. **Demir E**, Köse SK, Şenel E, Dünder E. Cilt Kanserinin Erken Teşhisi için Dijital Dermatoloji Resimleri ile Görüntü İşleme, 17. Ulusal Biyoistatistik Kongresi, 5-9 Kasım 2015, Girne/KIBRIS
15. Turgal E, Yavuz Z, **Demir E**, Köse SK, Erdoğan BD. Mikrodizi Verilerinin Karar ve Regresyon Ağacı Yöntemleriyle Sınıflandırılması, 17. Ulusal Biyoistatistik Kongresi, 5-9 Kasım 2015, Girne/KIBRIS
16. Şenel E, **Demir E.** Dünya ülkelerinin Behçet hastalığı yayınları üretkenlikleri: Literatürde Bir İpek Yolu Mevcut mu? , XXII. Prof.Dr.A.Lütfü Tat Sempozyumu, 16-20 Kasım 2015, Sheraton Kongre Merkezi Ankara
17. Şenel E, Karadağ AS, Turan H, Koca R, Kavak A, **Demir E.** Deri Tümörlerinin Teledermatolojik Değerlendirmesinde Teledermatoloji ve Telemikroskopinin Etkinliklerinin Karşılaştırılması, XXII. Prof.Dr.A.Lütfü Tat Sempozyumu, 18-22 Kasım 2015, Sheraton Kongre Merkezi Ankara.
18. Şenel E, **Demir E.** Melanositik Lezyon Teşhis ve Takibinde Görüntü İşleme Teknolojisinin Kullanımı, XXII. Prof.Dr.A.Lütfü Tat Sempozyumu, 18-22 Kasım 2015, Sheraton Kongre Merkezi Ankara.
19. Şenel E, **Demir E**, Dünder E. 2014 Yılı Dünya ve Türkiye Teletıp ve Teledermatoloji Yayınlarının Değerlendirilmesi, 8. Dermatoloji Bahar Sempozyumu, 9-12 Nisan 2015, Antalya.
20. Şenel E, Sabancılar E, Mansuroğlu C, **Demir E.** Telemikroskopi: Deri Tümörlerinin Teledermatolojik Değerlendirilmesinde Yeni Bir Teknik, 25. Ulusal Dermatoloji Kongresi, 21-25 Ekim 2014, Antalya.
21. Şenel E, Acar B. **Demir E.** Psoriatik Artrit Ortalama Platelet Hacminin Değerlendirilmesi, 25. Ulusal Dermatoloji Kongresi, 21-25 Ekim 2014, Antalya.

22. Şenel E, **Demir E.** Dünya Ülkelerinin Teletıp ve Teledermatoloji Yayınları Üretkenlikleri: 1980-2013 Yıllarını Kapsayan Bibliyometrik Bir Çalışma, 25. Ulusal Dermatoloji Kongresi 21-25 Ekim 2014, Antalya.

VII- Bilimsel Etkinlikleri

Danışmanlık Yaptığı Bilimsel Dergiler

- Orphanet Journal of Rare Diseases (SCI-E)
- Annals of Saudi Medicine (SCI-E)
- Journal of Integrative Medicine (PiscoMed)
- Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (HÜSBED)

Eğitici Olarak Yer Aldığı Kurs ve Çalıştaylar

- “Akademik Yaşam Klavuzu; Her Araştırmacının Bilmesi Gereken Temel Biyoistatistik Yöntemler” 23 Ekim 2018. Çorum Tabip Odası, Çorum.

Katılımcı Olarak Yer Aldığı Kurs ve Çalıştaylar

- 3. Uluslararası Biyoistatistik Kongresi (20. Ulusal Biyoistatistik Kongresi), 26-29 Ekim 2018, Gaziantep, Antalya
- 2. Uluslararası Biyoistatistik Kongresi (19. Ulusal Biyoistatistik Kongresi), 25-28 Ekim 2017 | Belek, Antalya
- 6th IBS Channel Network Conference, April 24-26, 2017, Hasselt University, Diepenbeek, Belgium
- 1. Uluslararası Biyoistatistik Kongresi (18. Ulusal Biyoistatistik Kongresi), 26-29 Ekim 2016 | Belek, Antalya
- 37th Annual Conference of the International Society for Clinical Biostatistics, 21 - 25 August 2016, Birmingham, UK.
- International conference on information complexity and statistical modeling in high dimensions with applications (IC-SMHD) May 18-21, 2016 - Cappadocia, Nevşehir Turkey
- Bütün Yönleri İle Çorum "Dün, Bugün, Yarın ve Değişim" Uluslararası Sempozyum, 28-30 Nisan 2016, ÇORUM

- 17. Ulusal Biyoistatistik Kongresi, 5-9 Kasım 2015, Girne/KIBRIS
- 12. Uluslararası İstatistik Öğrenci Kolokyum, 18-19 Mayıs 2015, Samsun.
- 8th conference of the Eastern Mediterranean Region of the International Biometric Society (EMR-IBS), 11-15 May 2015, Cappadocia, Nevşehir
- IV. Uluslararası Türk Dünyası Ekonomi Forumu, 7-9 Mayıs 2015, Çorum.
- International Conference on Advances in Statistics, 24-28 March 2015, Sharjah/DUBAİ/UAİ.
- 15th International Symposium on Econometrics, Operations Research and Statistics, 22-25 Mayıs 2014, Isparta.
- 9th International Statistics Day Symposium, 10-14 Mayıs 2014, Antalya.
- 8. Uluslararası İstatistik Kongresi, 27-30 Ekim 2013, Antalya.

Yer aldığı Projeler

- Nazal septum deviasyonlu hastalarda serum oksidatif stres ve serum antioksidan düzeylerinin belirlenmesi (Araştırmacı; 2016-2018 Hitit Üniversitesi BAP)
- Multidisipliner teletıp uygulaması: cerrahi öncesi telecerrahi tanımlarının histopatoloji ile karşılaştırılması (Araştırmacı; 2017-2018 Hitit Üniversitesi BAP)
- Deri tümörlerinin telemikroskopik değerlendirilmesinde farklı iki dijital mikroskop cihazının etkinliğinin karşılaştırılması (Araştırmacı; 2016-2017 Hitit Üniversitesi BAP)
- Çorum ilinde okul çağı çocuklarında çölyak hastalığı sıklığının araştırılması (Araştırmacı; 2016-2017 Hitit Üniversitesi BAP)
- Nazal Polipli hastalarda steroid tedavisi sonrasında serum ve polip dokusunda prolidaz enzim düzeyi ile fibrozis ve vaskülarite ilişkisinin belirlenmesi (Araştırmacı; 2016-2017 Hitit Üniversitesi BAP)
- Nazal Polip Hastalarında serum Interlokin 4, interlokin 5, interlokin 10, interlokin 17 ve interferon gamma düzeyleri ile Lund- maccaı skorlaması arasındaki ilişkinin belirlenmesi (Araştırmacı; 2016-2017 Hitit Üniversitesi BAP)
- Deri tümörlerinin teledermatoskopik değerlendirilmesinde farklı iki dermatoskopi cihazının etkinliğinin karşılaştırılması (Araştırmacı; 2015-2016 Hitit Üniversitesi BAP)

VIII- Diğer Bilgiler

1. En iyi poster araştırma ödülü: XXV. Ulusal Dermatoloji Kongresi, 21-25 Ekim 2014, Antalya.
2. Sözlü sunum en iyi araştırma ödülü, 6. Uyku Bozuklukları Kongresi, 26-30 Nisan 2017, Bodrum.
3. Sözlü sunum en iyi üçüncü araştırma ödülü: III. Uluslararası ve XX. Ulusal Biyoistatistik Kongresi, 26-29 Ekim 2018, Gaziantep.

