J Contemp Med 2017; 7(3): 265-277
DOI: 10.16899/gopctd.349948

ORIGINAL ARTICLE
ORİJİNAL ARAŞTIRMA

Simulation Study on Performance of Balance Metrics in Propensity Score Weighting Method

Propensity Skor Ağırlıklandırma Yönteminde Denge Metriklerinin Performansı Üzerine Benzetim Çalışması

Osman Demir[1], Anıl Dolgun[2], İlker Etikan[3], Yunus Emre Kuyucu[1], Osman Saraçbaşı[4]

[1]Gaziosmanpaşa Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Tokat,
[2]RMIT University, Mathematical Sciences, Avusturalya,
[3] Yakın Doğu Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kıbrıs,
[4]Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

ÖZ

Amaç: Randomizasyonun sağlanamadığı durumlarda bireylerin tedavi kollarına atanmasında yanlılığı minimize etmek için propensity skor ağırlıklandırma yönteminin kullanılmasını ve bu yöntemin genelleştirilmiş boosted ve çok kategorili lojistik regresyondan elde edilen bulgulara ilişkin performanslarının değerlendirmesini amaçlamaktadır.

Yöntem: Çok kategorili lojistik regresyon (MLR) ve genelleştirilmiş boosted modelden (GBM) elde edilen sonuçlar bir benzetim çalışması yardımıyla karşılaştırılacaktır. Benzetim çalışmasında üç kategorili tedavi grubu, sürekli yanıt değişkeni ve sürekli/dikotom ortak değişkenlerin olduğu, yedi farklı senaryo üzerinde 1000 tekrar kullanılarak, n=500, 1000, 2000 örneklem büyüklüğüne sahip veriler türetilecektir. MLR ve GBM'den elde edilen propensity skorları yardımıyla propensity ağırlıklarına ulaşılacak ve bu ağırlıkları kullanarak, tedavi etkilerini değerlendirmede kullanılan, ortalama tedavi etkisi (OTE) kestirimi ile denge metrikleri kullanılarak, dengenin değerlendirmesi yapılacaktır. Çalışmada R programındaki "twang" paketi kullanılacaktır.

Bulgular: Örneklem sayısı arttıkça denge değerlerinin daha azaldığı dolayısı ile yanlılığın düştüğü görülmektedir. Senaryolar daha kompleks hale geldikçe GBM'nin daha iyi denge sonuçları ürettiği görülmektedir. Ana etkilerin olduğu bir modelde MLR için daha iyi sonuçlar görülmektedir. MLR regresyon modelden elde edilen OTE ağırlıkları kararsız ve zayıf bir denge göstermektedir. Aşırı ağırlıkların kırpılması ya da kaldırılması dengenin düzelmesini sağlamaktadır.

Anahtar Kelimeler: Propensity skor ağırlıklandırma, GBM, Çok Kategorili Lojistik Regresyon

ABSTRACT

Objective: In the situation that randomization is not avaliable, to minimize the biasness in treatment arm assignments, the use of propensity score weighting method and the assessment of performances related to results obtained from generalized boosted and multinomial logistic regression (MLR) of propensity score weighting are aimed.

Method: Results obtained from MLR and GBM are to compare with the help of a simulation study. In simulation study, data with n=500, 1000, 2000 sample size will be derived using 1000 repetitions on seven scenarios with three categorized treatment group, continuous outcome variable and continuous/binary covariates. The propensity weights will be found with the help of Propensity scores obtained from MLR and GBM and using these weights, the balance will be assessed using balance metrics with average treatment effect estimation (ATE). In study, "twang" package in R program is used.

Results: As the number of samples increases, the balance values decreases more, so it seems that the biasness has fallen. As the scenarios become more complex, GBM produces better balance results. There are better results for MLR at main effect model. Trimming or removing excess weights ensures improving of balance.

Keywords: Propensity score weighting, GBM, Multinomial Logistic Regression

Corresponding Author: Dr Osman Demir,
Address Gaziosmanpaşa Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Tokat- Türkiye
E-mail: mosmandemir@hotmail.com

Başvuru Tarihi/Received: 13-03-2017
Kabul Tarihi/Accepted: 27-03-2017

Giriş

Tedavinin, yanıt değişken üzerindeki etkisi belirlenirken, tedavi kollarına atanan bireylerin özelliklerinden kaynaklı bir yanlılık ortaya çıkabilmektedir. Bu yanlılığın en aza indirilmesi için kullanılabilecek çalışma düzeni olan deneysel çalışmalarda amaç gruptaki bireylerin benzer özelliklerde olmasını sağlamaktır. Dolayısıyla yanıt değişken üzerindeki tedavi etkinliği yansız bir şekilde değerlendirilebilmektedir (1). Gruplara rasgele atamanın sağlanabildiği ve birbirine benzer grupların oluşturulmaya çalışılarak tedavilerin etkilerinin değerlendirildiği çalışmalar rasgele kontrollü çalışmalar olarak adlandırılır. Burada bireylerin gruplara atamasında araştırıcı ya da hasta faktörünü en aza indirmek gerekmektedir. Özellik olarak birbirine benzer gruplar oluşturulduğundan müdahalenin ya da tedavinin etkinliğinin yansız olarak değerlendirilmesi mümkün olabilmektedir. Bu bağlamda, rasgele kontrollü çalışmalar bilimsel olarak sonuçlarına güvenirliğin en fazla olduğu ve klinik denemeler için altın standardın olduğu denemelerdir (2,3).

Yarı deneysel ya da gözlemsel çalışmalardaki gibi rasgeleleştirmenin sağlanamadığı durumlarda tedavi kollarına atamadaki dengesizliği gidermek ve dolayısıyla tedavilerin yanıt üzerindeki etkisini incelemek için alternatif bir yöntem olan propensity skor teknikleri kullanılmaktadır. Bu tekniklerin genel amacı; tedaviye atanmadaki dengesizlikleri kontrol altına alarak sistematik hatayı azaltacak şekilde propensity skorunun kullanımını sağlamaktadır. Bu özelliği ile gözlemsel çalışma, rasgeleleştirmenin sağlandığı deneysel çalışmaya benzerlik gösterecektir (4). Propensity skoru, tedavi kollarında benzer olacak şekilde eşleştirme, tabakalama, ortak değişken düzeltmesi ve ağırlıklandırma için kullanılmaktadır (5,6).

Bu çalışmada bireylerin tedavi kollarına atanmasındaki dengesizliği gidermek için propensity skor kullanımı önerilmektedir. Bireylerin tedavi gruplarına atanma olasılıkları olan propensity skorlarıyla ağırlıklandırarak oluşturulan dengelenmiş yapay gruplar üzerinden tedavi etkinliğini belirlemek bu çalışmanın amacını oluşturmaktadır. İkiden fazla tedavi kolu olduğu durumda çok kategorili lojistik regresyon ve genelleştirilmiş boosted model, propensity skorlarını elde etmek için kullanılacaktır. GBM ve MLR'den elde edilen propensity skorlarıyla ağırlıklandırma sonrası yöntemlerin denge performansları karşılaştırılacaktır. Buna göre GBM'nin genel olarak denge performansları açısından MLR'ye göre daha başarılı sonuçlar göstereceği ve farklı senaryolar altında yöntemlerin tedavi etkinliğini farklı belirleyeceği hipotez edilmektedir. Bu hipotezleri test edebilmek için farklı veri senaryoları altında veri türetilecek, GBM ve MLR ile propensity skorları elde edilerek ağırlıklandırma yapılacak, ağırlıklandırma sonrası denge değerleri ve tedavi etkinlikleri karşılaştırılacaktır.

Nedensel Çıkarım

İki ya da daha fazla değişkenin (olayın) birbirini etkilemesi sonucu değişkenler (olaylar) arası ilişkiden ve dolayısıyla etkileyen ve etkilenen değişkenler olarak adlandırılan kavramlar ortaya çıkmaktadır. Bu kavramlar deneysel bilim çerçevesinde olayların ya da bireylerin gözlenebilir özelliklerinin tanımlanmasını, olaylar arasındaki ilişkilerin tanımlanmasını ve daha önemlisi olaylar arasındaki nedensel ilişkilerin tanımlanmasını gerektirmektedir (7,8). Değişkenler arasındaki gözlenebilir bağlantı olarak tanımlanabilen ilişki için göz önünde bulundurulması gereken nokta, bu ilişkiyle birlikte nedenselliğin var olmasıdır. Bir değişkenin başka bir değişken ya da değişkenler tarafından etkileniyor olması, değişkenler arasındaki nedenselliğin bir göstergesi olmayabilir. Dolayısıyla mutlak bir ilişkiden çok, bu ilişkinin ilgili alan içerisinde mantıksal yoruma sahip bir neden-sonuç ilişkisi içermesi gereklidir (9). Örneğin planlanan bir çalışma çerçevesinde cebinde

çakmak bulunan kişilerin akciğer kanseri olması ile sonuçlanan bir etkiye sahip olduğu gibi ilişki bulunabilir. Burada düşünülmesi gereken bilgi, çakmağın varlığı ile akciğer kanseri olması arasında nedensel olmayan ilişkiden çok çakmak taşıyan kişilerin bir sigara içicisi olduğu ve bu durumun akciğer kanseri riskini artırdığı neden-sonuç ilişkisidir. Bu örnekten de görüldüğü gibi sonuca etki eden bir neden olması ve nedenin sonuçtan önce olması gerekmektedir (10).

Nedenselliğin olduğu yerde, sonuç üzerindeki nedensel etkilerin kestirilmesinde karşıolgusallık (counterfactual) kavramı ortaya çıkmaktadır. Karşıolgusallık çalışma gruplarındaki bireylerin verilen herbir tedavi için potansiyel bir yanıta sahip olacağı ideal durumunu içermektedir. Örneğin tedavi almayan (örneğin sigara içen) bir birey için gözlenebilir yanıt değişken (bebeğin doğum ağırlığı) Y_0 olduğu düşünülsün. Şayet aynı birey tedaviyi almış (sigara içmeyen) olsaydı yanıt değişkenin Y_1 olduğu düşünülürse, Y_1 yanıt değişkeni bu birey için karşıolgusaldır denir. Buradaki problem bireyin asıl tedaviyi aldığı durumda gözlenebilir bir potansiyel yanıta sahip olması ve aksi durumda kayıp veri içermesidir. Buna; Rubin'in nedensel çıkarımının temel sorunu (Rubin's fundamental problem of causal inference) denmektedir (11,12). Bu problemin çözümü için potansiyel yanıtlar üzerinden kestirimde bulunacak olan tedavi etkileri kullanılmaktadır (13).

Tedavi Etkisi

Tedavi etkisi, ilgilenilen bağımlı değişken ya da yanıt değişken (örneğin, sistolik kan basıncı) üzerindeki tedavi kolunun (ilaç alma) etkisi olarak tanımlanabilir. Alan yazında tedavi değişkeni genel olarak ikili olarak çalışılmaktadır (14). Bu çalışmada ise tedavi değişkeni üç kategorili olarak ele alınmaktadır. Gruplardaki her bir birey için bu tedavileri alıp almama durumuna göre potansiyel sonuçlar ortaya çıkacaktır. Gruplardaki her bir birey için potansiyel yanıtlar $Y(m_k), k = 1,2,3$ olarak düşünüldüğünde bir birey için aldığı asıl tedavi $Y(m_1)$ iken aynı bireye ilişkin karşıolgusallar $Y(m_2)$ ve $Y(m_3)$ olacaktır. Dolayısıyla $Y(m_1), Y(m_2), Y(m_3)$ yanıtları aynı birey için potansiyel yanıtlar olacaktır. Birey için nedensel etkiyi ortaya koymak için tedavi kollarına karşılık gelen potansiyel yanıtlar arasındaki farkın bulunması gereklidir. Burada bir birey karşılaştırılacak olan tüm tedavi kollarında gözlemlenemediğinden dolayı bireylere ilişkin nedensel etki kestirilemeyecektir. Bu etkileri kestirmek için ortalama tedavi etkisi (OTE) kestiricisi kullanılmaktadır (13,15).

Ortalama Tedavi Etkisi (OTE)

İkili durumda tedavi ve kontrol grubu için gruplar arasındaki yanıt üzerinden elde edilen ortalama farktır. Yani popülasyondan rasgele seçilen bir birey için beklenen kazanç da denilebilmektedir (6).

$$E(F[1,0]) = E(Y[1] - Y[0]) \quad (2.1)$$

tedavi için düşünüldüğünde $K(K-1)/2$ tane OTE bulunacaktır. Çalışma üç tedavi grubu için düşünüldüğünde $\frac{3(3-1)}{2} = 3$ tane OTE vardır.

Bunlar;

$$E(F[m_1,m_2] = E(Y[m_1] - Y[m_2]) = E(Y[m_1]) - E(Y[m_2]) = \mu_{m_1} - \mu_{m_2} \quad (2.2)$$

$$E(F[m_1,m_3] = E(Y[m_1] - Y[m_3]) = E(Y[m_1]) - E(Y[m_3]) = \mu_{m_1} - \mu_{m_3} \quad (2.3)$$

$$E(F[m_2,m_3] = E(Y[m_2] - Y[m_3]) = E(Y[m_2]) - E(Y[m_3]) = \mu_{m_2} - \mu_{m_1} \quad (2.4)$$

Eşitliklerde;

$m_k,\ k = 1,2,3$: Tedavi gruplarını

F: Tedaviler arası fark

$Y[m_k]$: İlgili tedavi grubundaki yanıtı ve

μ_{m_k}: İlgili tedavi grubundaki ilgili değişken için ortalamayı temsil etmektedir.

Propensity Skoru

Tedavi etkinliğini incelemek için propensity skorununun kullanımı, ilk olarak Rosenbaum ve Rubin tarafından önerilmektedir (11). Propensity skoruna dayanan bu yöntemler son yıllarda yaygın bir şekilde kullanılmaktadır. Her bir tedavi grubundaki bireyin propensity skorunun ve kontrol grubundaki benzer skorla eşleştirilmesi üzerinden yapılan eşleştirme, tüm örneklemde hesaplanan propensity skorlarının beştebirlikleri hesaplanarak oluşan tabakalar üzerinden yapılan tabakalama, yine elde edilen propensity skorunun sonuç modelde bir ortak değişken olarak kullanılmasıyla yapılan regresyon düzeltmesi ve üzerinde durulacak ağırlıklandırma yaparak tedavi etkinlikleri bulunmaktadır (6,16).

Rasgeleleştirmenin mümkün olmadığı durumda kullanılabilecek olan model-tabanlı bir regresyon çözümlemesine göre propensity skor yöntemlerinin avantajları vardır. Tek bir skor yardımıyla tüm değişkenler özetlenebilmektedir. Klasik bir regresyon tabanlı analizde modele ilave edilecek değişken sayısında kısıtlılık vardır. (17,18). Regresyon çözümlemesinde yanıt değişkene ilişkin ortalama modellenirken propensity skor metotlarında böyle bir durum yoktur. Dolayısıyla model belirlemede yanlılık yaşanmamaktadır. Regresyon tabanlı bir modelin ileriye dönük kestirim yapması mümkün iken propensity skor yöntemleri gözlenen veri üzerinden bunu yapar yani ileriye dönük bir kestirimi yoktur (6,19).

Bir dengeleme skoru olan propensity skoru; gözlenen ortak değişken üzerinden bireyin tedavi grubuna atanma olasılığıdır. m tedavi grupları olmak üzere propensity skoru (3.1) eşitliğiyle hesaplanır:

$$\pi_j = e(x_j) = \Pr(T_j = m | X_j) \quad (3.1)$$

Eşitlik (3.1)'de, X_j, C tane gözlenen ortak değişkene ilişkin vektörü ifade etmektedir. Propensity skoru gözlenen X_j ortak değişkenlerinin tedavi kollarında benzer dağılım göstermesine yardımcı olmaktadır (4).

OTE kestirimlerinin elde edilebilmesi için gözlenen veri için ayrı bir gösterim kullanılacaktır. $j = 1,2,\dots,n$ olmak üzere $j.$ birey için gözlenen tedavi durumu T_j ile gösterilsin. $m_k \epsilon \{1,\dots,K\}$ tedavi gruplarını göstermek üzere j bireyi gerçekte m tedavisini alıyorsa $T_j = m_k$ şeklinde gösterilecektir. Y_j aynı şekilde asıl tedaviyi alan $j.$ bireyi göstermektedir. Burada dikkat edilmesi

gereken her bir bireyin K potansiyel sonucu vardır. Herbir birey için K tedaviden sadece biri gözlenebilir yanıta sahiptir (Y_j) ve diğer tedaviler altında gözlenmiş olsalardı ortaya çıkacak yanıt değişkenler karşıolgusallar adını alacaktır. Daha genel olarak $j.$ birey için $T_j = m_k$ ise $Y_j = Y_j[m_k]$ yanıtını göstermektedir. Popülasyondaki ortak değişkenlerin gösterimi için ise X_j kullanılacaktır (6).

Propensity Skoru ile Ağırlıklandırma

Propensity skorunu kullanarak verideki dengesizliği azaltmanın yöntemleri arasında propensity skoru eşleştirme, tabakalama, ortak değişken düzeltmesi ve propensity skoru ile ağırlıklandırma yöntemleri kullanılmaktadır. Alan yazında özellikle propensity skoru ile ağırlıklandırmanın kullanımı konusunda eksikliklerden dolayı bu yöntem üzerinde durulacaktır. Alanyazında propensity skor ağırlıklandırma, ters olasılıklı ağırlıklandırma ya da ters olasılıklı tedavi ağırlıklandırma olarak da adlandırılan bu yöntemin tabakalama ya da eşleştirmede olduğu gibi benzer propensity skorları üzerinden herhangi bir alt gruptaki bireyleri karşılaştırmaya çalışmaz. Kestirilen propensity skorlarını bireyleri ağırlıklandırmak için kullanmaktadır. Bunu yaparken ise tedavi kollarındaki ortak değişkenlerin dengelendiği yapay ya da yalancı bir popülasyon oluşturmaktadır. Alan çalışması analizlerinde olduğu gibi ağırlıklar kullanarak, kimi gözlemlere daha az, kimisine daha fazla ağırlık vererek tedavi etkisini ölçer. Propensity skor değerleri özellikle çok küçük olduğunda, bu değerlerin çarpmaya göre tersi kullanılacağından büyük ağırlıkların bulunmasına neden olmaktadır. Bunun üstesinden gelmek için ise ağırlıklar için aşırı değerlerden arındırılmış bir veri seti ile çalışılması önerilse de, bu durum tedavi etkilerinin kestiriminde yanlılığa neden olabilmektedir (4,20,21).

OTE kestirimlerinde her bir ortak değişken için propensity skorları kullanılarak elde edilecek ağırlıklar yardımıyla yeni ortalamalar kestirilmektedir.

K tedavi kolunun olduğu bir çalışma düzeninde OTE için kestirim yapıldığında, burada $K = 3$ alındığında Eşitlik 2.2, 2.3 ve 2.4'te verildiği gibi ($\mu_{m_1} - \mu_{m_2}$, $\mu_{m_1} - \mu_{m_3}$, $\mu_{m_2} - \mu_{m_3}$) 3(3-1)/2=3 kestirim yapılabilmektedir. $j.$ bireyin X_j ortak değişkenleri için m tedavisine atanma olasılığını $p_{m_k}(X_j)$ propensity skorunu göstersin.

$$p_{m_k}(X_j) = \Pr(T_j = m_k | X_j) \quad (3.2)$$

ve tedavi grubundaki bireylerin ağırlıkları,

$$w_j[m_k] = \frac{1}{p_{m_k}(X_j)} \quad (3.3)$$

olmak üzere ağırlıklandırılmış ortalama,

$$\hat{\mu}_{m_k} = \frac{\sum_{j=1}^{n} T_j[m_k] Y_j w_j[m_k]}{\sum_{j=1}^{n} T_j[m_k] w_j[m_k]} \quad (3.4)$$

olarak verilmektedir. Burada $T_j[m_k]$: $j.$ bireyin aldığı tedavi grubu, Y_j: $j.$ bireye ilişkin yanıt değişken değeridir. Kontrol grubu olarak alınacak tedavi grubu dışındaki grup için kullanılacak ağırlık

$$w_j[m_k] = \frac{1}{1 - p_{m_k}(X_j)} \quad (3.5)$$

'd ir.

Ağırlıklandırma sonrasında yapay olarak oluşturulan örneklem gruplarının büyüklüklerini, ağırlıkları kullanarak hesaplamak mümkündür.

$$ESS_{m_k} = \frac{\left(\sum_{j=1}^{N} T_j[m_k]w_j\right)^2}{\sum_{j=1}^{N} T_j[m_k]{w_j}^2} \quad (3.6)$$

k gruba ilişkin örnek büyüklüğü Eşitlik 3.8'deki gibi hesaplanabilir.

Varsayımlar

Koşullu bağımsızlık varsayımı

Bilinmeyen ya da ölçülemeyen etki karıştırıcının olmadığı anlamını taşıyan bu varsayımda, ortak değişken vektörü için potansiyel tedavi yanıtlarının, atama yönteminden bağımsız olması anlamına gelmektedir. Koşullu bağımsızlığa ilişkin bağıntı Eşitlik 3.9'da verilmektedir. K tedavi kolu için düşünüldüğünde;

$$Y_{m_1}, Y_{m_2}, \ldots, Y_{m_k} \perp T[m_k] \mid X \quad (3.9)$$

Pozitiflik varsayımı

Eşleştirme ya da örtüşme olarak da bilinen bu varsayım da her bir tedaviyi alan bireylerin olasılık değerlerinin pozitif olması gerektiğini ifade eder (Eşitlik 3.10).

$$0 < pr\left(T_j = m_k \middle| X\right) < 1 \quad (3.10)$$

Tedavi gruplarından elde edilen propensity skorlarına ilişkin dağılımın tedavi gruplarına yakın olması ya da örtüşmesi pozitiflik varsayımının sağlandığını göstermektedir.

Propensity Skor Kestirim Yöntemleri

Propensity skoru, kümeleme, lojistik, ayırma analizi gibi klasik istatistiksel yöntemlerle elde edildiği gibi genelleştirilmiş boosted, CART (Classification and regression trees, karar ağaçları) gibi makine öğrenim algoritmalarıyla da elde edilebilmektedir. Burada iki yöntem genelleştirilmiş boosted ve çok kategorili lojistik regresyon yöntemleri kullanılacaktır.

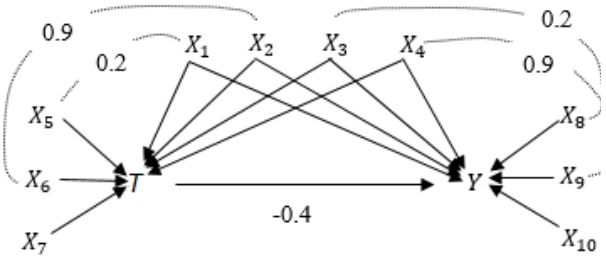
Benzetim Çalışması Tasarımı

Analizler, R 3.2.0 programı (22) ile yapıldı. GBM için "Twang", Çok kategorili Lojistik regresyon için "nnet" paketleri kullanıldı (23,24). Veri setlerini benzetim yardımıyla türetibilmek için R programlama dili kullanıldı. Benzetim çalışmasında bilgisayar işlemci çekirdeklerinden yararlanmak için paralel hesaplamaya yardımcı olması için ayrıca bazı paketleri çalıştırmak için gerekli "doSNOW", "car", "foreach", "digest", "survey", "foreign", "ggplot2", "reshape2" R paketlerinden yararlanıldı (25-31). Yedi senaryo altında türetilen veri setlerinde GBM ve çok kategorili lojistik regresyon yöntemlerinin ters olasılıklı ağırlıklandırma yardımıyla karşılaştırmasında 1000 tekrarlı Monte Carlo simülasyon yöntemi kullanıldı. Herbir senaryo için n=500, 1000, 2000 örneklem büyüklüğüne sahip veri setleri türetildi. Karşılaştırmalarda parametre kestirimleri, mutlak standartlaştırılmış yanlılık ve Kolmogorov Smirnov testi ölçütleri kullanıldı. Senaryolar Ek 2'de verilmektedir.

Benzetim Algoritmaları

Propensity skor modellerini oluştururken, karesel ve etkileşim terimlerinin dikkate alınma durumuna göre, doğrusallığın ve toplamsallığın farklılaştığı yedi senaryoda genelleştirilmiş boosted ve çok kategorili lojistik regresyon modellerinin performansları değerlendirilmiştir. Bu modeller ile kestirim yapılırken ters ağırlıklandırma yöntemi kullanılmaktadır.

Benzetim çalışması için veri yapısı, Setoguchi (2008) vd'nin gerçek bir veri setinden elde ettiği yapı kullanılmaktadır. Bu çalışmada farklılık olarak tedavi değişkeni üç kategori olarak alınmaktadır. Veri yapısı Şekil 4.1'deki gibi olacaktır.



Şekil 4.1. Benzetim verisi için değişken yapısı

Şekil 4.1'de tedavi ve yanıt değişken üzerinde etkisi önemli olan değişkenler ok ile gösterilmektedir. Şekil 4.1'de,

T: üç kategorili tedavi (etken, asıl bağımsız değişken),

Y: yanıt değişken,

$X_1 - X_4$: hem etkeni hem de sonucu etkileyen ortak değişkenler,

$X_5 - X_7$: sadece etkenle ilişkili kestiriciler,

$X_8 - X_{10}$: sadece yanıt değişkenle ilişkili kestiriciler,

$X_1, X_3, X_5, X_6, X_8, X_9$: iki kategorili değişkenler,

Y, X_2, X_4, X_7, X_{10}: sürekli değişkenlerdir.

İlk olarak standart normal dağılım kullanılarak $X_1, X_3, X_4, X_7, X_{10}$ değişkenleri türetilecektir. X_1 değişkeni kullanılarak arasındaki korelasyon 0,2 olacak şekilde X_5 değişkeni elde edilecektir. Korelasyon 0,2 olacak şekilde X_3 yardımıyla X_8 değişkeni elde edilecektir. Korelasyon 0,9 olacak şekilde X_2 yardımıyla X_6 değişkeni elde edilecektir. Son olarak korelasyon 0,9 olacak şekilde X_4 yardımıyla X_9 değişkeni elde edilecektir. $X_1, X_3, X_5, X_6, X_8, X_9$ değişkenlerini dikotom hale dönüştürmek için herbir değişkenin gözlemleri ilgili değişkenin ortalaması ile karşılaştırılacak ve ilgili değişkenin gözlemi değişken ortalamasından büyük ise o gözleme 1 değilse 0 atanarak değişken ikili yapıya dönüştürülecektir.

Çoklu tedavi T değişkeni X_i nin bir fonksiyonu olarak çok kategorili lojistik model kullanılarak modellenecektir. Bu fonksiyonun eşitliği (gerçek propensity skorları) aşağıda verilen yedi senaryo altında elde edilecektir. Senaryolara ilişkin kullanılan katsayılar ve Eşitlikler Ek 2 ve Ek 3'de verilmektedir.

Senaryolar

1. Sadece ana etkilerin olduğu toplamsallık ve doğrusallığın olduğu model

2. Bir karesel terimin olduğu model (hafif doğrusal olmayan model)

3. Üç karesel terimin olduğu model (ılımlı doğrusal olmayan model)

4. Üç tane iki yönlü etkileşim terimli hafif toplamsal olmayan model

5. Dört tane iki yönlü etkileşim terimli ve bir karesel terimin olduğu hafif toplamsal

ve doğrusal olmayan model

6. 10 tane iki yönlü etkileşim terimli ılımlı toplamsal olmayan model

7. 10 tane iki yönlü etkileşim terimli ve üç tane iki yönlü etkileşim terimli ılımlı toplamsal ve doğrusal olmayan modeldir.

Y yanıt değişkeni ise T ve X_i nin bir fonksiyonu olarak modellenecektir. Yanıt değişkene ilişkin kullanılan modeller ve katsayılar Ek 2 ve Ek 3'de verilmektedir.

Ortak Değişken ve Tedavi Değişkeni Türetimi

İlk olarak standart normal dağılım kullanılarak $X_1, X_3, X_4, X_7, X_{10}$ değişkenleri türetilecektir. X_1 değişkeni kullanılarak arasındaki korelasyon 0.2 olacak şekilde X_5

değişkeni elde edilecektir. Korelasyon 0.2 olacak şekilde X_3 yardımıyla X_8 değişkeni elde edilecektir. Korelasyon 0.9 olacak şekilde X_2 yardımıyla X_6 değişkeni elde edilecektir. Son olarak korelasyon 0.9 olacak şekilde X_4 yardımıyla X_9 değişkeni elde edilecektir. $X_1, X_3, X_5, X_6, X_8, X_9$ değişkenlerini dikotom hale dönüştürmek için herbir değişkenin gözlemleri ilgili değişkenin ortalaması ile karşılaştırılacak ve ilgili değişkenin gözlemi değişken ortalamasından büyük ise o gözleme 1 değilse 0 atanarak değişken ikili yapıya dönüştürülecektir.

Çoklu tedavi T değişkeni X_i nin bir fonksiyonu olarak multinomial lojistik model kullanılarak modellenecektir. Bu fonksiyonun eşitliği (gerçek propensity skorları) Ek 3'de verilen yedi senaryo altında elde edilecektir.

Yanıt Değişken Türetimi

Y yanıt değişkeni ise T ve X_i nin bir fonksiyonu olarak modellenecektir. Eşitlik 4.1'deki kullanılan katsayılar Ek 3'de verilmektedir.

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_8 + \alpha_6 X_9 + \alpha_7 X_{10} + \gamma_1 T \quad (4.1)$$

Veri Setlerinin Türetimi

1000 tekrarlı 7 farklı senaryo oluşturularak n=500, 1000, 2000 örneklem büyüklüğüne sahip veriler türetilecektir. Senaryolar Ek 2'de verilmektedir.

Performans Ölçüleri

GBM ve MLR'den elde edilen propensity skorları kullanılarak bireyler ağırlıklandırıldıktan sonra her bir ortak değişken için ağırlıklandırmadan önceki ve sonraki değerleri üzerinden hesaplanacak iki performans ölçüsü ya da metrik vardır. Bunlar standartlaştırılmış yanlılık ve Kolmogorov Smirnov istatistiğidir.

Standartlaştırılmış yanlılık

Bireylerin ağırlıklandırılmasından önceki tedavi ve kontrol olarak seçilen gruplar arasındaki mutlak fark ile ağırlıklandırdıktan sonraki tedavi ve kontrol olarak seçilen gruplar arasındaki mutlak farkın değişimi incelenir. Mutlak farkın azalması grupların benzer olduğunu ve dengenin sağlandığını göstermektedir. OTE için tedavi ve kontrol grubu için ağırlıklandırılmış ortalamalar arasındaki farkın ortak standart sapmaya bölünmesi ile elde edilir. OTE için standartlaştırılmış yanlılıklığa ilişkin çoklu tedavi için genelleştirme Eşitlik 4.2'deki gibidir.

OTE için mutlak standartlaştırılmış yanlılık (MSB), $c = 1,2,\ldots,C$ ortak değişken sayısını ifade etmek üzere

$$MSB_{cm_k} = \left|\bar{X}_{cm_k} - \bar{X}_{cm_p}\right| / \hat{\sigma}_{cm_p} \quad (4.2)$$

dir.

Eşitlikte;

$\bar{X}_{cm_k}$: c. ortak değişken için k tedavisini alanlar için ortalama,

$\bar{X}_{cm_p}$: c. ortak değişken için popülasyona ilişkin ortalama,

$\hat{\sigma}_{cm_p}$: c. ortak değişken için popülasyona ilişkin standart sapmadır.

$\bar{X}_{cm_k}$ tedavi grubu olarak alındığında Eşitlik 3.5'in gösterim olarak farklı bir yazımı ile ilgili ortak değişken için ağırlıklı ortalama

$$\bar{X}_{cm_k} = \frac{\sum_{j=1}^{n} T_j[m_k]\bar{X}_{cj}w_j[m_k}{\sum_{j=1}^{n} T_j[m_k]w_j[m_k]} \text{ dir.} \quad (4.3)$$

Standartlaştırılmış yanlılık değerleri dengeyi değerlendirmek için kullanılmaktadır. Genel olarak, 0,20'den daha düşük standartlaştırılmış ortalama farklar küçük olarak düşünülür, 0,40 orta seviye, 0,60 büyük olarak düşünülmektedir (32).

Benzetim çalışmasında her bir yöntemde, 1000 farklı veri setinden elde edilen standartlaştırılmış yanlılık değerlerinin ortalamaları alınmaktadır.

Kolmogorov Smirnov İstatistiği

Kolmogorov Smirnov (KS) istatistiği tedavi kollarındaki örneklemlerin ağırlıklandırılmış deneysel dağılım fonksiyonlarına dayanmaktadır. Bu test tedavi kolları arasında ortak değişkenlere ilişkin dağılımın örtüşüp örtüşmediğini belirlemeye yardımcı olmaktadır (33). Ayrıca tüm dağılımı karşılaştırmak mümkündür.

c ortak değişken ve m_1 tedavi ve m_2 kontrol olarak düşünüldüğünde

$$I(X_{jc} \le x) = \begin{cases} 1, & X_{jc} \le x \\ 0, & X_{jc} > x \end{cases} \quad (4.5)$$

için olasılık yoğunluk fonksiyonu

$$YF_{cm_k}(x) = \frac{\sum_{j=1}^{n} w_j[m_k]T_j[m_k]I(X_{jc} \le x)}{\sum_{j=1}^{n} w_j[m_k]T_j[m_k]} \quad (4.6)$$

olarak tanımlanmaktadır. Aynı ortak değişken için Kolmogorov Smirnov istatistiği

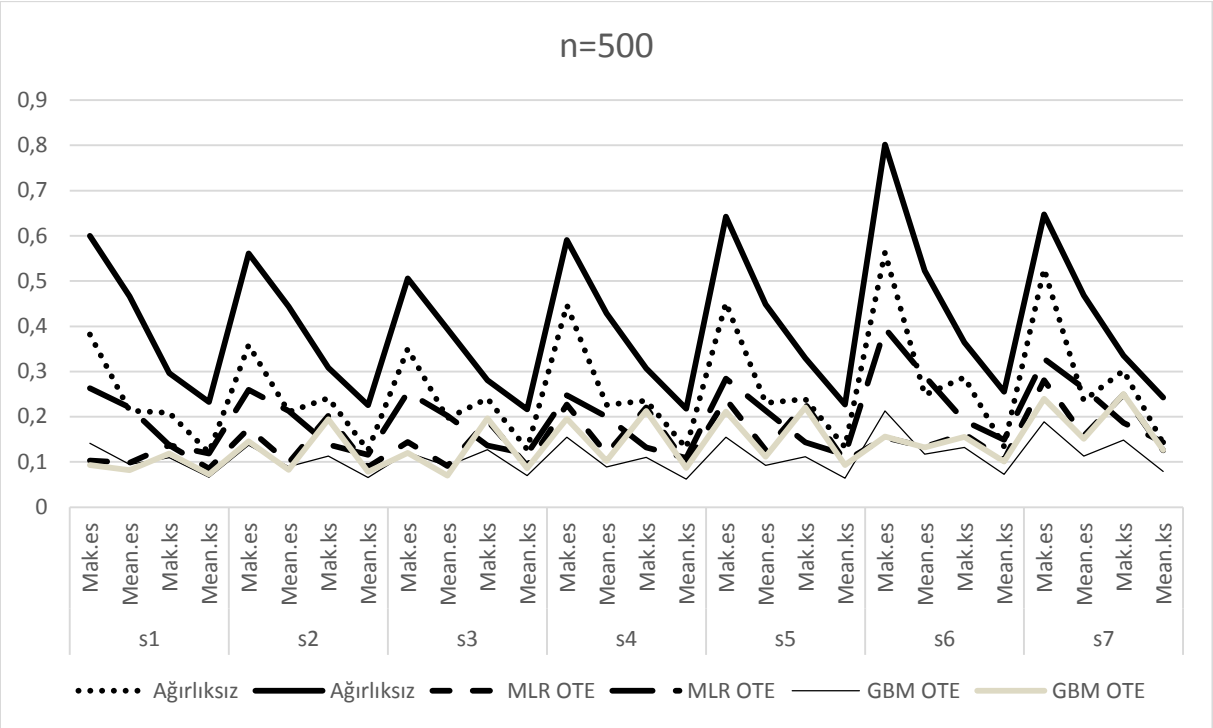
$$KS_c = \sup_x \left| YF_{cm_1}(x) - YF_{cm_2}(x) \right| \text{ 'dir.} \quad (4.7)$$

Eşitlik 4.7'de gösterildiği gibi tedavi kolları arasındaki en büyük fark değeri Kolmogorov Smirnov istatistiğini vermektedir. KS istatistiği ile denge değerlendirmesinde büyük örneklemlerde 0,10'dan büyük KS istatistiğinin dengenin bozulduğunu göstermektedir. Her bir yöntemde, 1000 farklı veri setinden elde edilen KS değerlerinin ortalamaları alınmaktadır.

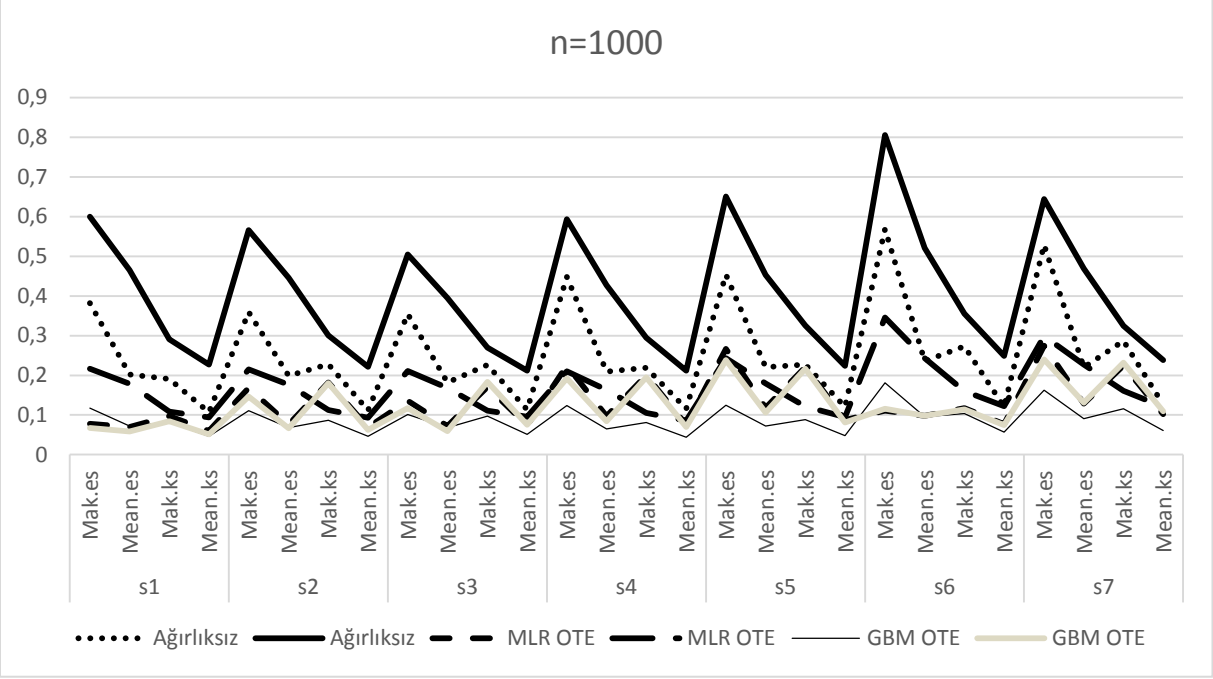
Bulgular

Çalışmada her bir senaryo altında n=500, 1000 ve 2000 örneklem büyüklüğüne sahip her biri için 1000 farklı veri seti türetilmektedir. Senaryolardan elde edilen gerçek propensity skor değerleri tedavi değişkenin türetimi için kullanılmaktadır. Özet denge değerleri için tedavi kollarındaki örneklem büyüklükleri, 1000 farklı veri setinden elde edilen herbir tedavi grubundaki örneklemlerin kendi gruplarındaki örneklem büyüklüklerinin ortalamaları alınarak bulunmaktadır. Aynı şekilde diğer özet istatistikler için de 1000 farklı veri setinden elde edilen değerlerin ortalamaları kullanılmaktadır.

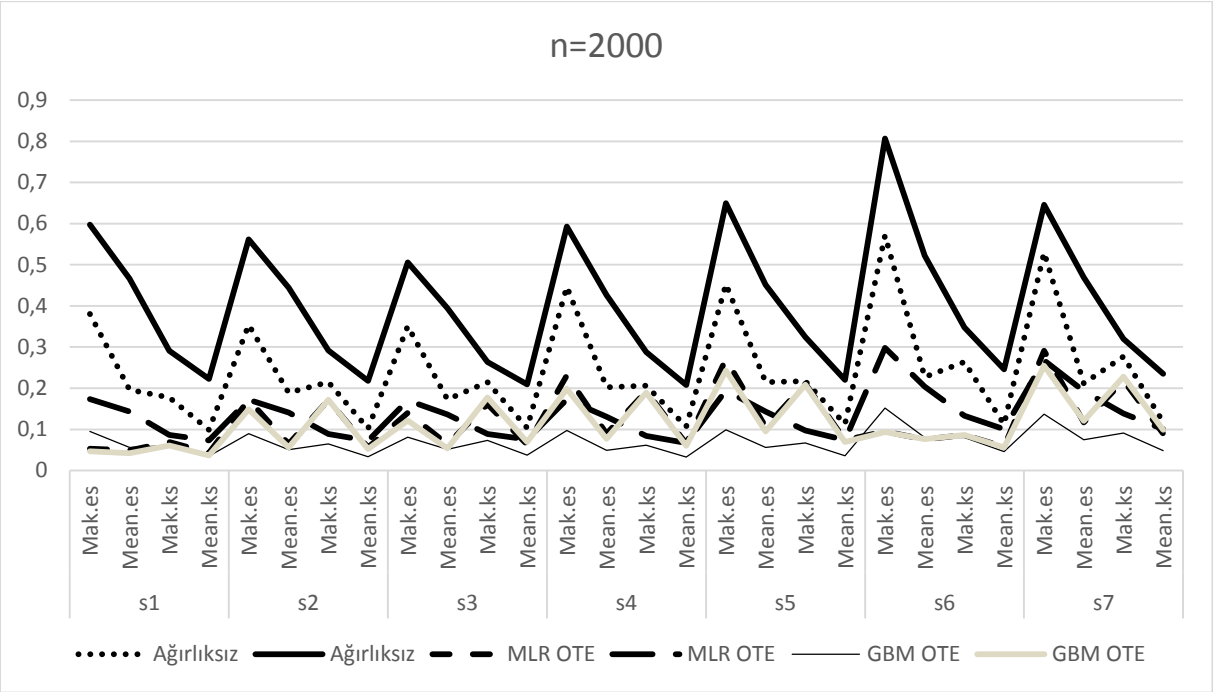
Benzetim çalışmasında, belirlenen senaryolar altında oluşturulan modellerden propensity skor kestirim yöntemlerinden GBM ve MLR ile propensity skorlarını kestirip, propensity skor ağırlıklandırma yöntemi yardımıyla ağırlıkların elde edilmektedir. Elde edilen ağırlıklar yardımıyla ortak değişkenlere ilişkin denge metrikleri değerlendirilmektedir.



Şekil 1. n=500 örneklem için 7 senaryo altında MLR ve GBM denge metrikleri



Şekil 2. n=1000 örneklem için 7 senaryo altında MLR ve GBM denge metrikleri



Şekil 3. n=2000 örneklem için 7 senaryo altında MLR ve GBM denge metrikleri

Senaryolar kompleks hale geldikçe genel olarak GBM'nin MLR'den daha iyi sonuçlar ortaya koymaktadır. Ana etkilerin olduğu modelde ise MLR'nin daha iyi sonuçlar ortaya çıkardığı görülmektedir. Kompleks modeller için esnek bir kestirim yapan GBM yöntemi MLR'den daha başarılıdır. Örneklem sayısı arttıkça GBM ve MLR yöntemlerinden elde edilen metriklerdeki ölçüm değerleri daha da düşmektedir. Bu da örneklem sayısındaki artışa bağlı olarak daha iyi performans ortaya koymaktadır.

Sonuç

Araştırıcılar tedavi etkilerini belirlerken, maliyet ve etik nedenlerden ya da geleneksel istatistiksel yöntemlerin kısıtlılıklarından dolayı alternatif yöntemlere gereksinim duyabilirler. Bu çalışmada, gözlemsel ya da yarı deneysel çalışmalarda gruplar arasındaki farklılığı en aza indirmek ya da gidermek için propensity skor ağırlıklandırma yardımıyla genelleştirilmiş boosted model ve çok kategorili lojistik regresyon yönteminin tedavi etkinliğini belirlemedeki performanslarının belirlenmesi üzerine durulmaktadır. Tedavi değişkeninin iki kategorili olduğu durumda propensity skor ağırlıklandırmasında yapay zeka algoritmalarının, lojistik regresyondan yanlılığın azaltılması açısından daha iyi sonuçlar verdiği çalışmalar bulunmaktadır (5). Tıpta çeşitli nedenlerle çalışma düzenlerinin sağlıklı olarak belirlenememesi ihtimalini de göz önünde bulundurarak çalışmalardaki yanlılıkları gidermek büyük önem arz etmektedir. Bu bağlamda propensity skor ağırlıklandırma yöntemiyle bireysel nedenlerle grup farklılıklarını en aza indirip, tedavi etkinliğini belirlemek uygulamada sıklıkla kullanılabilecek bir yöntem olabilir. Dolayısıyla klasik istatistiksel yöntemlerdeki kısıtlılıklar nedeniyle yapılamayan analizler, propensity skor ağırlıklandırma ile yapılabilir (17,18). Her ne kadar gözlemsel çalışmalarda bireysel farklılıkları dikkate alarak tedavi kollarındaki farklılığı en aza indirerek tedavi etkinliği değerlendirmesinde önerilse de Olmos ve ark. bu yöntemin kontrollü çalışmalarda da benzer sonuçlar ürettiği de gösterilmektedir. Ayrıca dağılımda aşırı değerler olduğunda özellikle MLR'den elde edilen ağırlıklar tedavi etki kestiriminde sorun ortaya

çıkarabilmektedir. Bu da ağırlıklandırmanın sınırlılığını ortaya koymaktadır (34).

Klasik istatistiksel bir metot olan MLR ile birlikte, CART, pruned CART, bagged CART, random forest, boosted CART gibi yapay zeka algoritmalarının kullanıldığı yöntemler kullanılarak propensity skor kestirimleri üzerinden performans ölçümleri yapılabilir. Daha sonrasında tedavi etkinliği değerlendirilebilir. Kurulan modeller karmaşık hale geldikçe, GBM'nin MLR'den daha başarılı olduğundan, GBM'yi tercih etmek daha doğru olacaktır (15). Tedavi değişkeni için ikiden fazla kategori için bu analiz yapılabildiği gibi ortak değişkenlerin de ikiden fazla kategorili olan ortak değişken alınarak bu yöntemin nasıl performans gösterdiği belirlenebilir. Eksikliği görülen ve tedavi grupları arasındaki farklılığı gidermeye çalışan propensity skoru ile ağırlıklandırma yöntemi, geniş bir uygulama alanı bulacaktır. R, SAS, Stata 13 ve sonrası programlar yardımıyla bu analizi gerçekleştirmek mümkündür.

Kaynaklar

1.Ellenberg, S.S., Fleming, T.R., DeMets, D.L. Data monitoring committees in clinical trials: a practical perspective: John Wiley & Sons; 2003.

2.Doll, R. Controlled trials: the 1948 watershed. Br Med J 1998; 317 (7167), 1217.

3.Hannan, E.L. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. JACC Cardiovasc Interv 2008; 1 (3), 211-217.

4.Austin, P.C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011; 46 (3), 399-424.

5.Lee, B.K., Lessler, J., Stuart, E.A. Improving propensity score weighting using machine learning. Stat Med 2010; 29 (3), 337-346.

6.McCaffrey, D.F., Griffin, B.A., Almirall, D., Slaughter, M.E., Ramchand, R., Burgette, L.F. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med 2013; 32 (19), 3388-3414.

7.Rosenberger, W.F., Lachin, J.M. Randomization in clinical trials: theory and practice: John Wiley & Sons; 2002.

8.Alpar, R. Spor, Sağlık ve Eğitim Bilimlerinde Uygulamalı İstatistik ve Geçerlik-Güvenirlik: Birinci Baskı, Ankara: Detay Yayıncılık; 2010.

9.Sümbüloğlu, V., Alpar, R., Özdemir, P. (1998) Değişkenler Arası İlişkilerin İncelenmesi. İç Hastalıkları Dergisi, 5 (6), 416.

10.Aalen, O.O., Frigessi, A. What can statistics contribute to a causal understanding? Scand Stat Theory Appl 2007; 34 (1), 155-168.

11.Rosenbaum, P.R., Rubin, D.B. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70 (1), 41-55.

12.Rubin, D.B. Bayesian inference for causal effects: The role of randomization. Ann Stat 1978; 34-58.

13.StataCorp, L. Stata Treatment-Effects Reference Manual; 2013.

14.Söderbom, M. Applied Econometrics Lecture 11: Treatment Effects Part I; 2009.

15.McCaffrey, D.F., Ridgeway, G., Morral, A.R. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 2004; 9 (4), 403.

16.Li, F., Zaslavsky, A.M., Landrum, M.B. Propensity score weighting with multilevel data. *Statistics in medicine* 2013; 32 (19), 3373-3387.

17.Alpar, R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler (3. Baskı). Ankara: Detay; 2011.

18.Rosenbaum, P.R., Rubin, D.B. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc 1984; 79 (387), 516-524.

19.Vansteelandt, S., Daniel, R. On regression adjustment for the propensity score. Stat Med 2014; 33 (23), 4053-4072.

20.Stone, C.A., Tang, Y. Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. PARE 2013; 18 (13), 1-12.

21.Williamson, E.J., Forbes, A. Introduction to propensity scores. Respirology 2014; 19 (5), 625-635.

22.Team, R.D.C. R: A language and environment for statistical computing. Vienna, Austria; 2008.

23.Burgette, L., Griffin, B.A., McCaffrey, D. Propensity scores for multiple treatments: A tutorial for the mnps function in the twang package; 2014.

24.Ripley, W.N.V.a.B.D. Modern Applied Statistics with S (c. Fourth). New York: Springer; 2002.

25.Weisberg, J.F.a.S. An {R} Companion to Applied Regression (Second bs.). Thousand Oaks {CA}: Sage; 2011.

26.Eddelbuettel, D., Lucas, A., Tuszynski, J., Bengtsson, H., Urbanek, S., Frasca, M. ve diğerleri. digest: Create Cryptographic Hash Digests of R Objects; 2014.

27.Lumley, T. Analysis of complex survey samples. J Stat Softw 2004; 9 (1), 1-19.

28.Analytics, R., Weston, S. doSNOW: Foreach parallel adaptor for the snow package. R package; 2011.

29.Analytics, R., Weston, S. foreach: Foreach looping construct for R. R package version 2013; 1 (1).

30.Team, R.C., DebRoy, S., Bivand, R. foreign: Read Data Stored by Minitab. S, SAS, SPSS, Stata, Systat, Weka, dBase 2013; 0.8-57.

31.Wickham, H. ggplot2: elegant graphics for data analysis: Springer Science & Business Media; 2009.

32.Cohen, J. Statistical power analysis for the behavioral sciences (rev: Lawrence Erlbaum Associates, Inc; 1977.

33.Li, M. Using the Propensity Score Method to Estimate Causal Effects A Review and Practical Guide. Organ Res Methods 2013; 16 (2), 188-226.

34.Olmos, A., Govindasamy, P. Propensity Scores: A Practical Introduction Using R. J Multidiscip Eval 2015; 11 (25), 68-88.