



Kesirli Polinomlu Lojistik Regresyon Modeli

Nihal Ata Tutkun

Hacettepe Üniversitesi

Fen Fakültesi

İstatistik Bölümü

06800-Beytepe, Ankara, Türkiye

nihalata@hacettepe.edu.tr

Öz

Lojistik regresyon modelinde sürekli bağımsız değişkenler ile bağımlı değişken arasında doğrusal bir fonksiyonel ilişkinin olduğu varsayılmaktadır. Doğrusallık varsayımının sağlanmaması modelin yanlış belirlenmesine neden olmaktadır. Sürekli bir açıklayıcı değişken kategorik biçime dönüştürülerek kullanıldığında ise düzey sayısı arttıkça tahmin edilmesi gereken parametre sayısı arttığından modeli çözmek ve yorumlamak zorlaşmaktadır. Bu nedenlerden dolayı, sürekli açıklayıcı değişkenlerin fonksiyonel yapısının belirlenmesi için kesirli polinomlar kullanılmaktadır. Bu çalışmada kesirli polinomlu lojistik regresyon modeli incelenmiş ve gerçek bir veri kümesine uygulanmıştır. Doğrusal ilişkili lojistik regresyon modeli ve kategorik değişkenli lojistik regresyon modeli için de sonuçlar elde edilmiş ve tartışılmıştır.

Anahtar sözcükler: Doğrusal olmayan ilişki; Kesirli polinomlar; Sürekli bağımsız değişken.

Abstract

Logistic Regression Model with Fractional Polynomial

In logistic regression analysis, there is a linearity assumption between the independent and dependent variables. The violation of linearity assumption leads to wrong final model. When a continuous variable is categorized to use in a model, the analysis and interpretation of the model may become difficult especially with the high number of levels. For these reasons, fractional polynomials are used to determine the functional structure of the continuous variables. In this study, the fractional polynomials are examined and applied to a real data set. The results for logistic regression model with an assumption of linearity and logistic regression model with a categorical independent variable were also obtained and discussed.

Keywords: *Nonlinear relationship; Fractional polynomials; Continuous independent variable.*

1. Giriş

Regresyon modellerinde bağımsız değişkenlerin bağımlı değişken ile doğrusal ilişkili olduğu varsayılmaktadır. Ancak doğrusallık varsayımı sağlanmayabilir. Bu durumda da değişkenler kategorik biçime dönüştürülerek modele dahil edilmektedir. Örneğin yaş için 5 yıllık sınıflandırmaların kullanılması durumunda doğrusallık varsayımı sağlanmayabilir ve model uyumu kötü olabilir. Bu durumda daha karmaşık fonksiyonel biçimlere gereksinim açıktır. Özellikle büyüme eğrisi doğrusal olmayan bir yapıya sahip olduğu çeşitli çalışmalar (Count [7,8], Wingerd [27], Berkey ve Reed [5]) ile gösterilmiştir.

Doğrusal olmayan etkileri modellemek için ikinci dereceden regresyon ya da spline-tabanlı modeller alternatif modellerdir. Ancak, tahmin için kullanılan denklemler basit değildir ve çözümleme için uygun yazılımlar yaygın değildir. Ayrıca, değişkenlerin etkilerini göstermek için en uygun modelin kesirli polinomlu (fractional polynomial, FP) modeller olduğu Baneshi ve diğ. [4] tarafından gösterilmiştir.

Sürekli değişkenlerin sürekli değişken olarak modele dahil edilmesi ile kategorik biçime dönüştürülerek modele dahil edilmesi farklı sonuçların elde edilmesine neden olmaktadır. Özellikle sağlık alanındaki çalışmalarda sürekli değişkenler, örneğin yaş, çok kategorili değişkenlere dönüştürülerek kullanılmaktadır. Ancak kategorileştirmek bilgi kaybına neden olmaktadır ve kaç tane kesim noktası olmalı ya da bu kesim noktaları nerelerde olmalı gibi önemli sorunlar açığa çıkartmaktadır [2, 12, 16]. Sürekli bir değişkenin kategorik biçime dönüştürülmesinin etkinlik kaybına neden olduğu simülasyon çalışmaları (Zhao ve Kolonel [28], Greenland [9]) ile gösterilmiştir. Bu durumda da kesirli polinomlu modeller uygun bir alternatif olmaktadır.

Regresyon tabanlı modellerde değişkenleri kategorik biçime dönüştürmek yerine doğrusal olmayan bir ilişki söz konusu olduğunda logaritmik ya da karekök gibi basit dönüşümler ya da polinom modeller kullanılabilir. Fakat polinom modellerin uygulamada çoğunlukla yetersiz olduğu gösterilmiştir (Hastie ve Tibshirani [10], Royston ve Altman [13], Greenland [9]).

Kesikli polinomlu lojistik regresyon modelleri ilk olarak Royston ve Altman [13] tarafından önerilmiş ve daha sonra Sauerbrei ve Royston [20] tarafından geliştirilmiştir. Royston ve Altman [13] çalışmasında bir grup üs terimleri kümesinden oluşan kesirli polinomları önermiştir. Geleneksel polinomlar da önerilen bu ailenin bir üyesidir. Royston ve Altman [13] ise birleştirilmiş bir tanım ve derecelendirme yapmıştır. Klasik yöntemlere uyan esnek ve basit bir yöntem geliştirmişlerdir. Değişken seçimi ve model uyumu için iteratif bir algoritma önermişlerdir. Royston vd. [14] ise sürekli risk değişkenleri ile karıştırıcı değişkenleri tek bir model içinde birleştirmişlerdir. Sauerbrei ve Royston [20] çok değişkenli prognostik ve tanısal modellerin oluşturulmasında kesirli polinomların kullanımını incelemiş ve modelin güvenilirliğini değerlendirmek için bootstrap yeniden örnekleme yöntemini kullanmışlardır. Ambler ve Royston [3], aşırı belirlenmeyi dikkate alan, Royston ve Altman [13] tarafından uygun kesirli polinomu seçmek için önerilen seçim prosedürlerinin performanslarını incelemişlerdir. Royston ve Sauerbrei [15] çalışmasında ise çok değişkenli kesirli polinom modellerinde kararlılığı değerlendirmek için 2001'de yapılan çalışmayı genişletmişlerdir. Royston ve Sauerbrei [16] çalışmasında kesirli polinomları kullanarak klinik denemelerde tedavi ve açıklayıcı değişkenler arasındaki etkileşimleri modellemek için yeni bir yaklaşım önerilmiştir. Royston ve Sauerbrei [17] kesirli polinomlar ile sürekli açıklayıcı değişkenleri modellemiş ve epidemioloji ile ilgili uygulama yapmıştır. Sauerbrei ve diğ. [22] kesirli polinomlar için SAS ve R'da bir macro geliştirmişlerdir. Royston ve Ambler [14] tarafından STATA'da geliştirilen program ile farklılıklarını ortaya koymuştur. Royston ve Sauerbrei [18]'de açıklayıcı değişkenin başlangıç dönüşümü ile kesirli polinom modellerinin sağlamlığını geliştiren bir yaklaşıma yer verilmiştir. MFPI olarak adlandırılan çok değişkenli kesirli polinomlar için geliştirilen algoritmada, açıklayıcı değişkenler iki düzeyli, kategorik ya da sürekli olabiliyor ve kesim noktalarından (cut-points) kaçınılmaktadır. Sauerbrei vd. [23]'de MFPI üzerinde incelemeler yapmıştır.

Bu çalışmada, polinom regresyonun bir alternatifi olan kesirli polinomlu lojistik regresyon modeli elde alınacaktır. Çalışmanın ikinci bölümünde kesirli polinomlar verilecektir. Bölüm 3'te ise gerçek bir veri kümesine doğrusal ilişkili lojistik regresyon modeli, kesirli polinomlu lojistik regresyon modeli ve sürekli bağımsız değişkenin kategorik biçime dönüştürülerek kullanıldığı lojistik regresyon modeli uygulanmış ve Bölüm 4'de ise sonuçlar tartışılmıştır.

2. Kesirli Polinomlu Lojistik Regresyon Modeli

Değişkenler arasındaki ilişkiyi doğru bir biçimde modellemek için kesirli polinomlu lojistik regresyon modeli kullanılmaktadır [19]. Uygun fonksiyon seçimi yönteminin (prosedürünün) uygulanması doğrusal bir fonksiyonun ya da doğrusal olmayan bir kesirli polinomun uygun olup olmamasının kontrol edilmesini kolaylaştırmaktadır.

Sağlık bilimlerinde ve bir klinik denemede sürekli açıklayıcı değişken ile tedavi arasındaki etkileşimler incelenirken, en bilinen yaklaşım tek bir kesim noktası kullanarak iki grup oluşturmaktadır ki bu da yanlış pozitif sonuçların olasılığını yükseltebilir [13, 19]. x üzerindeki her bir olası kesim noktası düşünülür ve p -değerini minimum yapan x değeri seçilir. Kesim noktasının seçimi aslında bir ölçüde şansa bağlıdır. Bu prosedür, Altman vd. [2] “minimum p -değeri” yaklaşımı olarak adlandırmıştır. Çoklu testler I. tür hata olasılığını 0.05’den 0.4’e yükseltir. İki grubun bağımlı değişkenleri arasındaki farklılık aşırı tahmin edilir ve güven aralıkları da oldukça dardır.

Kabul görmüş kesim noktalarının kullanılması (örneğin kitle beden indeksi 25’den büyük olan birinin fazla kilolu olarak adlandırılması), daha önceki çalışmalarda kullanılan kesim noktalarının kullanılması ya da bu noktalar bilinmiyorsa ortancanın kullanılması kesim noktasını belirlerken kullanılan yöntemler arasında yer almaktadır. Bununla birlikte, farklı çalışmaların farklı kesim noktaları vardır. Bu nedenle de sonuçların karşılaştırılmaları ya da özetlenmesi mümkün olamamaktadır. Bu yöntemler de bilgi kaybı yaratmaktadır, ancak en azından şişirilmiş I. tür hataya neden olmaz.

$\beta_1 X$ modeli ele alınsın ve kolaylık olması açısından da sabit terimi (β_0) çıkaralım. Üssel dönüşüm modeli (power transformation model), $\beta_1 X^p$ en basit halidir. Üssel dönüşüm modeli, p nin farklı seçimlerini kullanmaktadır. Royston ve Altman [13] birinci dereceden kesirli polinomlu (first-degree fractional polynomial) ya da FP1 fonksiyonlu olarak adlandırarak modellemiştir. Kesirli polinomlar yönteminde, bütün değişkenleri pozitif yapacak biçimde dönüşümü veren fonksiyonlar kullanılmaktadır. Royston ve Sauerbrei daha sonraki çalışmalarında m . dereceden FP fonksiyonu FP_m olarak göstermiş ve $m=1$ ve $m=2$ olarak alıp, modelleri incelemiştir. FP1 ya da FP2 fonksiyonları uygulamada karşılaşılan birçok regresyon fonksiyonunu modellemek için yeterlidir. $x>0$ açıklayıcı değişkeni için FP1,

$$FP1(x) = \beta_1 x^p$$

biçimindedir. Burada $p \in S = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$ ’dir. Polinomlu regresyonda olduğu gibi tek terimli FP1 fonksiyonlarından daha karmaşık ancak daha esnek olan iki terimli FP2 fonksiyonlarına geçilmektedir. $\beta_1 X^{p_1} + \beta_2 X^{p_2}$ yerine, (p_1, p_2) üslerine sahip FP2 fonksiyonları,

$$FP2(x) = \begin{cases} \beta_1 x^{p_1} + \beta_2 x^{p_2}, & p_1 \neq p_2 \\ \beta_1 x^{p_1} + \beta_2 x^{p_2} \ln x, & p_1 = p_2 = p \end{cases}$$

biçiminde yazılmaktadır. Burada. $p, p_1, p_2 \in S = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$ ’dir ve $x^0 = \ln x$ olarak yazılmaktadır. $p_1 = p_2$ ise, model tekrarlı üslü (repeated-powers) FP2 modeli adını almaktadır.

Günümüze kadar yapılan çalışmalarda S kümesi için herhangi bir değişiklik önerilmemiştir. Bu küme hiçbir dönüşüm yapılmamasını ($p=1$) ve çift taraflı (reciprocal, $1/x$), logaritmik, karekök ve karesel dönüşümlerini içermektedir. Model uyumu için p nin sekiz değerinin her biri denenmektedir, yani 8 FP1 dönüşümü vardır. $p=1$ ’ e karşılık $p \neq 1$ ’in anlamlılık testi, doğrusal olmamanın (non-linearity) testidir. Olabilirlik oran testine dayalı olarak ($-2 \log L$ farkları alınarak) test yapılır ve bu test bir serbestlik dereceli ki-kare dağılımına sahiptir. Test istatistiği anlamlı değil ise doğrusal model, anlamlı ise doğrusal olmayan fonksiyonlara sahip kesirli polinomlar tercih edilir [26]. İkinci dereceden kesirli polinomlar için $p_1 \neq p_2$ olan 28 FP2 dönüşümü ve $p_1 = p_2$ olan 8 FP2 dönüşümü vardır. Bu dönüşümler $p \in S = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$ kümesi içinde yer alan 8 p değerinin ikinci dereceden kesirli polinomlar için birbirleri ile olan kombinasyonlarından elde edilmektedir. En yüksek olabilirlik, S kümesinden seçilen en iyi kombinasyonları verir.

Doğrusal fonksiyon varsayılan fonksiyondur. Veri daha karmaşık FP fonksiyonlarını desteklemediği sürece, doğrusal model seçilir. α anlamlılık düzeyinde FP modelleri arasına seçim yapabilmek için iki yöntem önerilmiştir. Bunlardan birincisi Ambler ve Royston [3] ve Sauerbrei ve Royston [21] tarafından önerilen RA2 yöntemidir ve kapalı test yöntemine (Marcus ve diğ. [12]) benzemektedir. Bu yöntem aday modeller arasından az belirlenmiş (yokluk) modelden çok belirlenmiş (FP) modele geçilmesine izin vermektedir [22, 23]. RA2 yönteminde en iyi FP modelini seçmek için önerilen fonksiyon seçme işlemi için aşağıdaki adımlar izlenir:

1. X açıklayıcı değişkeni için en iyi FP2 modelini yokluk modeline göre α düzeyinde, 4 sd ile test ediniz. Test anlamlı değil ise (H_0 kabul), X in etkisinin önemli olmadığı sonucuna ulaşılır ve süreç durdurulur. Aksi takdirde, ikinci aşamaya geçilir.
2. X açıklayıcı değişkeni için en iyi FP2 modelini doğrusal modele göre α düzeyinde, 3 sd ile test ediniz. Test önemli değil ise, uygun modelin doğrusal model olduğuna karar verilir. Aksi takdirde, üçüncü aşamaya geçilir.
3. X açıklayıcı değişkeni için en iyi FP2 modelini en iyi FP1 modeline göre α düzeyinde, 2 sd ile test ediniz. Test önemli değil ise, uygun modelin FP1 modeli olduğuna karar verilir. Aksi takdirde FP2 modeli seçilir.

Birinci adımdaki test toplam ilişkiyi (overall association), ikinci adım ise doğrusallığı inceler. Üçüncü adımdaki test ise daha basit ve daha karmaşık doğrusal olmayan modeller arasında seçim yapar. Bu seçim yöntemi kullanılmadan önce araştırmacı α değerini ve FP modelinin derecesini belirlemelidir. Genellikle $\alpha=0.05$ ve FP2 seçilir.

Model seçimi için ikinci yöntem ise ardışık yöntemdir. Bu yöntemde en iyi fonksiyonu seçmek için aşağıdaki adımlar uygulanmaktadır:

1. En iyi FP2 modelini en iyi FP1 modeline göre α düzeyinde, 2 sd ile test ediniz. Test önemli ise süreç durdurulur ve en iyi model FP2'dir. Aksi takdirde, ikinci aşamaya geçilir.
2. En iyi FP1 modelini doğrusal modele göre α düzeyinde, 1 sd ile test ediniz. Test önemli ise süreç durdurulur ve uygun modelin FP1 olduğuna karar verilir. Aksi takdirde, üçüncü aşamaya geçilir.
3. Doğrusal modeli x in olmadığı modele göre α düzeyinde, 1 sd ile test ediniz. Test önemli ise son modelin doğrusal model olduğuna karar verilir. Aksi takdirde x modelden çıkarılır [22, 23]

Birden çok test uygulandığından, gerçek ilişki doğrusal iken, I. tür hata nominal α değerini aşabilir. Bu nedenle, yöntem daha karmaşık modelleri basit modellere göre daha uygun gösterme eğilimindedir [22]. Amber ve Royston [3] tarafından yapılan benzetim çalışmasında en uygun fonksiyonun seçimi için RA2 yönteminin daha uygun olduğu belirtilmiştir.

Birçok çalışmada, birden çok değişken modele dahil edilerek çözümlenmeler yapılmaktadır. Modeli en iyi açıklayan değişkenlerin modelde olduğu, zayıf ya da hiçbir etkisi olmayan değişkenlerin modelden çıkarıldığı ve sürekli açıklayıcı değişkenler için uygun fonksiyonel biçimin yer aldığı çok değişkenli bir model elde edilmek istenmektedir. Sürekli değişkenler için doğrusallık varsayımı yapıldığında değişken seçim yöntemlerinden geriye doğru seçim en yaygın kullanılan yaklaşımdır. Doğrusallığın olmaması durumunda kullanılan kesirli polinomlar için de geriye doğru seçim prosedürü kullanılmaktadır. Sauerbrei ve Royston [20] buna çok değişkenli kesirli polinomlar prosedürü adını vermiştir. Bu prosedürü kullanmak için bir regresyon tipi modeli nasıl kuracağı konusunda bilgi sahibi olmak yeterlidir [22, 23].

3. Sayısal Örnek

Günümüzde bebeklerin beslenmelerini konu alan araştırmalarda en önemli besinin anne sütü olduğu vurgulanmaktadır. Anne sütünün besin değerinin yüksek olması, bebeğin zihinsel, fiziksel ve ruhsal gelişimi için besleyici olması, sindirimini kolay olması, hastalıklara karşı bebeği koruması, bebeğin özellikle ilk 6 ay süre ile tüm besin ihtiyacını karşılaması ve ekonomik olması gibi özellikleri anne sütünün önemini göstermektedir. Dünya Sağlık Örgütü (WHO), Birleşmiş Milletler Çocuklara Yardım Fonu (UNICEF) ve Amerikan Çocuk Akademisi (AAP) bebeklerin ilk altı ay sadece anne sütü ile beslenmelerinin gerekli ve yeterli olduğunu açıklamıştır. Annelerin bebek emzirme konusundaki görüşlerini içeren çalışmalar incelendiğinde, annelerin büyük çoğunluğu emzirmenin en iyi seçenek olduğunu düşünürken, bebeğin ne kadar süreyle emzirilmesi gerektiği konusunda ülkeler arası farklılıklar olduğu görülmektedir.

Tüm ülkelerde olduğu gibi Türkiye'de de anne sütü ile ilgili birçok bilimsel çalışma yapılmıştır. Alikışifoğlu [1], Tunçel vd. [25], Ünsal vd. [26], Onay vd. [11] ve Bolat vd.'nin [6] çalışmalarında ise

lojistik regresyon çözümlemesi kullanılmıştır. Bu çalışmada ise, KKTC Gazimagusa bölgesinde annelerin bebeklerini emzirme alışkanlıklarının araştırıldığı ve Şafak ve Tutkun'un [24] çalışmasında yer alan veriler ele alınmıştır. 25 yaş ve üstü annelere ait veriler, lojistik regresyon çözümlemesi ile kesirli polinomlu lojistik regresyon çözümlemesinin uygulamasını göstermek amacı ile kullanılmıştır.

Haziran 2013–Ağustos 2013 tarihleri arasında Gazimağusa Tıp Merkezi Hastanesi Çocuk Polikliniklerine başvuran en az 6 aylık bebeği olan annelerin araştırmanın kitlesini oluşturduğu çalışmadan yararlanılmıştır. Bu çalışmanın örneklemini ise araştırmaya katılmayı kabul eden ve 25 yaş üstü olan anneler oluşturmuştur. Çalışmada yer alan analizler için SAS paket programı kullanılmıştır. Çalışmaya katılmayı kabul eden annelerin doğum anındaki ortalama yaşı 29.6 ± 2.38 olarak elde edilmiştir. 42 annenin 22'sinin anne sütü ile bebeklerini besleme sürelerinin 6 aydan daha az olduğu sonucuna ulaşılmıştır. Diğer değişkenlere ait bilgiler ise Çizelge 1'de özetlenmiştir.

Çizelge 1. Özetleyici bilgiler

Özellikler	n (%)	Özellikler	n (%)
<i>Çocuk sayısı</i>		<i>Annenin iş durumu</i>	
1	14 (35.1)	Çalışmıyor	13 (31.7)
2 çocuk ve üstü	27 (65.9)	Çalışıyor	28 (68.3)
<i>Doğum yapılan gebelik haftası</i>		<i>Anne sigara/alkol kullanıyor mu?</i>	
Term (≥ 37 hafta)	28 (68.3)	Evet	10 (24.4)
Preterm (< 37 hafta)	13 (31.7)	Hayır	31 (75.6)
<i>Annenin eğitim düzeyi</i>		<i>Ek katı gıdalara ne zaman başlandı?</i>	
İlk veya Ortaokul	7 (17.1)	6. aydan önce	11 (26.8)
Lise	10 (24.4)	6. ay ve sonrası	30 (73.2)
Önlisans/Lisans/Lisansüstü	24 (58.5)		
<i>Doğum anındaki çoklu gebelik</i>			
Tek çocuk	37 (90.2)		
Çoklu gebelik	4 (9.8)		

Çalışmada lojistik regresyon çözümlemesine geçmeden önce aykırı değerler ve uç değerler incelenmiş ve veri kümesinde olmadığı görülmüştür. Ayrıca çoklu bağlantı sorunu da araştırılmış ve değişkenler arasında çoklu bağlantı sorunu olmadığı sonucuna ulaşılmıştır.

Öncelikli olarak, yaş değişkeni kategorik ve sürekli değişken olarak ele alınmış ve klasik lojistik regresyon çözümlemesi yapılmıştır. Daha sonra ise yaş değişkeni sürekli değişken olarak model dahil edilerek kesirli polinomlu lojistik regresyon çözümlemesi yapılmıştır. Yaş değişkeni için ortanca 29 olduğundan, yaş değişkeni 29 yaş ve altındakiler birinci grup olmak üzere iki kategoriye ayrılmıştır. Buna göre elde edilen lojistik regresyon modeli sonuçları Çizelge 2'de verilmiştir. %5 anlamlılık düzeyinde modelin veri kümesine uygun olduğu görülmüştür ($p=0.027$).

Çizelge 2. Kategorik değişkenli klasik lojistik regresyon çözümlemesi

Değişken	Parametre Tahmini	Std. Hata	Wald	P > z
Yaş	-1.034	1.056	0.958	0.328
Çocuk sayısı	-0.771	1.044	0.545	0.460
Gebelik haftası	0.097	1.006	0.009	0.923
Eğitim düzeyi			5.168	0.075
Lise	-0.79	1.502	0.003	0.960
Önlisans/Lisans/Lisansüstü	-3.734	1.951	3.663	0.056
Sigara/alkol kullanımı	-1.189	1.065	1.247	0.264
İş durumu	1.517	1.708	0.788	0.375

Ek katı gıda	-3.141	1.218	6.657	0.010
Doğum anındaki çocuk sayısı	0.066	1.575	0.002	0.967
Sabit	1.800	1.042	2.983	0.084

Yaş değişkeni sürekli değişken olarak alınıp modele dahil edildiğinde elde edilen sonuçlar Çizelge 3’de verilmiştir. Bu modelin de anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz ($p=0.023$).

Çizelge 3. Sürekli değişkenli klasik lojistik regresyon çözümlemesi

Değişken	Parametre Tahmini	Std. Hata	Wald	P > z
Yaş	-0.326	0.283	1.330	0.249
Çocuk sayısı	-0.996	1.133	0.772	0.380
Gebelik haftası	0.403	1.118	0.130	0.719
Eğitim düzeyi			5.273	0.072
Lise	-0.221	1.442	0.023	0.878
Önlisans/Lisans/Lisansüstü	-3.829	1.973	3.766	0.052
Sigara/alkol kullanımı	-1.271	1.072	1.405	0.236
İş durumu	1.436	1.669	0.741	0.389
Ek katı gıda	-3.590	1.412	6.463	0.011
Doğum anındaki çocuk sayısı	-0.127	1.565	0.007	0.935
Sabit	11.605	8.901	1.700	0.192

Çizelge 2’de verilen model için Cox Snell $R^2 = 0.367$, Nagelkerke $R^2 = 0.490$ ve Pseudo $R^2 = 0.330$ iken Çizelge 3’de verilen model için Cox Snell $R^2 = 0.374$, Nagelkerke $R^2 = 0.339$ ve Pseudo $R^2 = 0.339$ ’dır. Her iki model için de modele ait doğru sınıflandırma oranı ise %78 bulunmuştur.

Yaş değişkeninin doğrusal olduğu varsayımı kullanılarak Çizelge 2 ve Çizelge 3’teki modeller elde edilmiştir. Lojistik regresyon modelinde doğrusallık varsayımı yer aldığından çalışmada sürekli yapıda bulunan yaş değişkeni hem sürekli hem de kategorik olarak ele alınıp çalışmaya dahil edilmiştir. Böylece yaş değişkeninin yapısı değiştiğinde bağımlı değişken üzerindeki etkisinin değişip değişmediği gözlemlenmek istenmiştir. Ayrıca, yaş değişkeni bağımlı değişken ile doğrusal ilişkili olmayabilir. Bunu görebilmek için de fonksiyonel biçimlerin hepsinin teker teker denemek yerine doğrudan veriye en uygun sonuçları veren kesirli polinomlu lojistik regresyon modeli kullanılmıştır. Kesirli polinomlu lojistik regresyon modeli için kapalı test seçim yöntemi kullanıldığında elde edilen sonuçlar ise Çizelge 4’de verilmiştir.

Çizelge 4. Kesirli polinomlu lojistik regresyon modeli için kapalı test seçim yöntemi sonuçları

Model	Deviance	P	Adım	Karşılaştırma	Dev. farkı	p
FP2	28.636	-2 ve -2	1	FP2 vs yokluk	27.178	<0.05
FP1	36.108	-2	2	FP2 vs doğrusal	8.942	<0.05
Doğrusal	37.518	1	3	FP2 vs FP1	7.472	<0.05
Yokluk	56.814					

FP2'nin yokluk modeli ile karşılaştırılması, yaş değişkeninin önemli olduğunu ve FP2'nin doğrusal model ile karşılaştırılması da yaş değişkeninin bağımlı değişken ile doğrusal olmayan bir ilişkisi olduğunu göstermektedir. FP2 ile FP1'in karşılaştırılması ise önemli bulunmuştur ($p=0.024$), yani FP2 modeli FP1 modeline tercih edildiği sonucuna ulaşılmıştır.

FP2 modelleri için de en uygun model $p_1=p_2=-2$ olarak elde edilmiştir. Yani, $FP2(x) = \beta_1 x^{-2} + \beta_2 x^{-2} \ln x$ 'dir. Yaş değişkeni y ile gösterilmek üzere bu değişkene ait dönüşüm;

$$Yaş_1 = y^2 - 0.001139,$$

$$Yaş_2 = y^2 \ln y - 0.00386$$

biçimindedir. FP2 modeli de %95 güven düzeyinde anlamlı bulunmuştur ($p=0.0017$) ve modele ait sonuçlar Çizelge 5'de verilmiştir. Ayrıca bu modele ait Pseudo R^2 0.496'dır.

Çizelge 5. Sürekli değişkenli kesirli polinomlu lojistik regresyon çözümlemesi

Değişken	Parametre Tahmini	Std. Hata	Wald	P > z
Yaş_1	886113.3	505298.6	1.75	0.079
Yaş_2	-305313.2	175687.3	-1.74	0.082
Çocuk sayısı	-0.6013	1.1724	-0.51	0.608
Gebelik haftası	0.7089	1.2831	0.55	0.581
Eğitim düzeyi				
Lise	1.3493	2.1046	0.64	0.521
Önlisans/Lisans/Lisansüstü	-3.7413	2.3962	-1.56	0.118
Sigara/alkol kullanımı	-0.1841	1.2359	-0.15	0.882
İş durumu	1.4379	1.8630	0.77	0.440
Ek katı gıda	-3.5383	1.8107	-1.95	0.051
Doğum anındaki çocuk sayısı	-0.6097	1.7143	-0.36	0.722
Sabit	2.0925	2.9417	0.71	0.477

Çizelge 5'deki kesirli polinomlu lojistik regresyon modeli incelendiğinde, yaş değişkeninin %90 güven düzeyinde anlamlı olduğu görülmüştür. Klasik lojistik regresyon modellerinde ise yaş değişkeni anlamsız bulunmuştur. En iyi FP fonksiyonu seçim prosedürü kullanılarak yaş değişkeninin doğrusal olarak modele dahil edilmemesi gerektiği görülmüştür. Sonuçlar incelendiğinde de önemsiz bulunan yaş değişkeninin aslında önemli olduğu ve bunun da bir analizin sonuçlarını nasıl değiştirebileceği görülmektedir.

4. Sonuç ve öneriler

İstatistiksel modeller veri içindeki ilişkileri gösteren basit modeller elde edilmesini sağlayacak biçimde tasarlanmaktadır. Klasik lojistik regresyon modelinin temel varsayımı bağımlı değişkenlerin bağımsız değişken ile doğrusal ilişkili olmamasıdır. Doğrusallığın incelenmesi için değişkenin doğal logaritmasının alınarak modele dahil edilmesi ve bu yeni değişkenin anlamlı olması doğrusallığın olmadığını göstermek için en temel yöntemlerden biridir. Ancak değişkenin fonksiyonel yapısı farklı olabilir ve bu durumda da doğru olanı bu gibi bir yöntem ile bulmak zor olabilir. Kesirli polinomların ise tek değişkenli ve çok değişkenli doğrusal olmayan regresyon çözümlemelerinde kullanışlı olduğu yapılan çalışmalar ile gösterilmiştir. Bu çalışmada ele alınan kesirli polinomlu lojistik regresyon modeli ise FP1

ile 8 ve FP2 ile 36 fonksiyonel biçimi incelemeye ve doğrusal model ile karşılaştırmaya izin vermektedir. Böylece bağımlı değişken ile doğrusal ilişkiye sahip olmayan sürekli değişkenlerin fonksiyonel yapısını belirlemek için kullanılabilir. Bu çalışmada anne sütü ile beslenme veri kümesi incelenmiş ve kesirli polinomlu lojistik regresyon modelinin klasik lojistik regresyon modellerine göre daha uygun olduğunu sonucuna ulaşılmıştır. Ayrıca veri kümesinde yer alan doğrusal olmayan yaş değişkenini doğrusal olarak modele dahil etmenin değişkenin anlamlılığı açısından farklı sonuçlara neden olduğu görülmüştür. Bu nedenle de veriye uygunluğu sağlayacak en basit modelin seçilmesini sağlayacak yöntemlerden biridir.

Teşekkür

- Veri toplama aşamasındaki değerli yardımları ile çalışmaya katkıda bulunan Dr. Cemal Mert'e, Dr. Emine Özlüses'e ve Çağla Şafak'a teşekkür ederim.

Kaynaklar

- [1] M. Alikashioglu, E. Erginoz, E.T. Gur, Z. Baltas, B. Beker, A. Arvas, 2001, Factors influencing the duration of exclusive breastfeeding in a group of Turkish women, *J Hum Lact*, 2001, 17, 220-26.
- [2] D.G. Altman, B. Lausen, W. Sauerbrei, M. Schumacher, 1994, The dangers of using "optimal" cutpoints in the evaluation of prognostic factors, *J Natl Cancer Inst*, 86, 829-835.
- [3] G. Ambler, P. Royston, 2001, Fractional polynomial model selection procedures: investigation of Type I error rate, *J. Statist. Simul. Comput.* 69, 89-108.
- [4] Baneshi, M.R., Nakhaee, F., Law, M., 2013, On the Use of Fractional Polynomial Models to Assess Preventive Aspect of Variables: An Example in Prevention of Mortality Following HIV Infection, *Int J Prev. Med.*, 4(4), 414-419.
- [5] C. S. Berkey, R. B. Reed, 1987, A model for describing normal and abnormal growth in early childhood. *Hum. Biol.*, 59, 973-987.
- [6] F. Bolat, S. Uslu, G. Bolat, A. Bülbül, S. Arslan, M. Çelik, S. Cömert., A. Nuhuğlu, 2011, İlk altı ayda anne sütü ile beslenmeye etki eden faktörler. *Çocuk Dergisi*, 11, 5-13.
- [7] E. W. Count, 1942, A quantitative analysis of growth in certain human skull dimensions. *Hum. Biol.*, 14, 143-165.
- [8] E. W. Count, 1943, Growth patterns of human physique: an approach to kinetic anthropometry. *Hum. Biol.*, 15, 1-32.
- [9] S. Greenland, 1995, Avoiding power loss associated with categorization and ordinaş scores in dose-response and trend analysis, *Epidemiology*, 6, 450-454.
- [10] T.J. Hastie, R.J. Tibshirani, 1990, Generalized additive models, New York: Chapman and Hall.
- [11] D. Onay, M. Akman, S. Akdeniz, A. Kacaroglu Vicdan, 2009, Breast-feeding behaviours of the mothers, *Pakistan Journal of Nutrition*, 8, 1126-31.
- [12] R. Marcus, E. Peritz, E., K.R. Gabriel, 1976, On closed test procedures with special reference to ordered analysis of variance, *Biometrika*, 76, 655-660.
- [13] P. Royston, P., D.G. Altman, 1994, Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion), *Applied Statistics*, 43, 3, 429-467.
- [14] P. Royston, G. Ambler, W. Sauerbrei, 1999, The use of fractional polynomials to model continuous risk variables in epidemiology, *Int J Epidemiol*, 28(5):964-974.
- [15] P. Royston, W. Sauerbrei, 2003, Stability of multivariable fractional polynomial models with selection of variables and transformations: a bootstrap investigation, *Statistics in Medicine*, 22, 639-659.
- [16] P. Royston, W. Sauerbrei, 2004, A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials, *Statistics in Medicine*, 23, 2509-2525.
- [17] P. Royston, W. Sauerbrei, 2005, Building multivariable regression models with continuous covariates, with a practical emphasis on fractional polynomials and applications in clinical epidemiology, *Methods of Information in Medicine* 44, 561-571.
- [18] P. Royston, W. Sauerbrei, 2007, Improving the robustness of fractional polynomial models by preliminary covariate transformation: a pragmatic approach. *Computational Statistics & Data Analysis*, 51, 4240 - 4253.

- [19] P.Royston, W. Sauerbrei, 2008, *Multivariable Model - Building: A Pragmatic Approach to Regression Analysis based on Fractional Polynomials for Modelling Continuous Variables*, John Wiley and Sons Ltd., England.
- [20] W. Sauerbrei, P. Royston, 1999, Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials, *Journal of the Royal Statistical Society (Series A)*, 162, 71–94.
- [21] W. Sauerbrei, P. Royston, 2002, Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials, *Journal of the Royal Statistical Society (Series A)*, 165, 399–400.
- [22] W. Sauerbrei, P. Royston, 2006, Multivariate regression model building by using fractional polynomials: Description of SAS, STATA and R programs, *Computational Statistics and Data Analysis*, 50, 3464–3485.
- [23] W. Sauerbrei, P. Royston, K. Zapien, 2007, Detecting an interaction between treatment and a continuous covariate: A comparison of two approaches, *Computational Statistics & Data Analysis*, 51, 4054 – 4063.
- [24] Ç. Şafak, N. Ata, *Bebeklerin anne sütü ile beslenme sürelerinin yaşam çözümlenmesi ile incelenmesi*”, 1. Genç İstatistikçiler Sempozyumu, 10-11 Ekim 2013, Ankara, Türkiye
- [25] E.K. Tunçel, C. DüNDAR, Y. Pekşen, 2005, Ebelerin anne sütü ile ilgili bilgi ve uygulamalarının değerlendirilmesi, *The Medical Journal of Kocatepe*, 2005, 6, 43-8.
- [26] H. Ünsal, F. Atlıhan, H. Özkan, Ş. Targan, H. Hassoy, 2005, Toplumda anne sütü verme eğilimi ve buna etki eden faktörler, *Çocuk Sağlığı ve Hastalıkları Dergisi*, 8, 226-33.
- [27] J. Wingerd, 1970, The relation of growth from birth to two years to sex, parental size and other factors, using Rao's method of the transformed time scale. *Hum. Biol.*, 42, 105-131.
- [28] L. P. Zhao, L.N. Kolonel, 1992, Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies, *American Journal of epidemiology*, 136, 464-474.