

Regresyon Modellerine Alternatif Bir Yaklaşım: MARS

An Alternative Approach to Regression Models: MARS

Gülhan OREKİCİ TEMEL,^a
Handan ANKARALI,^b
Ayşe Canan YAZICI,^c

^aBiyostatistik AD,
Mersin Üniversitesi Tıp Fakültesi,
Mersin

^bBiyostatistik AD,
Düzce Üniversitesi Tıp Fakültesi,
Düzce

^cBiyostatistik AD,
Başkent Üniversitesi, Tıp Fakültesi,
Ankara

Geliş Tarihi/Received: 04.12.2009
Kabul Tarihi/Accepted: 19.02.2010

*Bu çalışma VIII. Ulusal Biyoistatistik Kongresi
(20 Eylül 2005 Bursa)'nde sözlü sunum olarak
sunulmuştur.*

Yazışma Adresi/Correspondence:
Gülhan OREKİCİ TEMEL
Mersin Üniversitesi Tıp Fakültesi,
Biyostatistik AD, Mersin,
TÜRKİYE/TURKEY
gulhanorekici@mersin.edu.tr

ÖZET Amaç: Bu çalışmada, regresyon modellerinden birisi olan MARS modelinin özellikleri tanımlanarak diğer regresyon modellerine göre avantajlı olduğu durumlar açıklanmış ve uygulama adımları üzerinde durulmuştur. **Gereç ve Yöntemler:** Çalışmada, MARS modelinin teorik özelliklerini tanımlayan formüller verilmiş, uygulama bölümünde ise kanser ve behçetli toplam 51 hastanın ailelerinden tespit edilmiş bir takım davranış ve psikolojik test sonuçlarına ilişkin ölçümler kullanılarak bu kişilerdeki depresyon durumunu tahmin etmede kullanılabilecek bir MARS modeli elde edilmiştir. **Bulgular:** Kurulan MARS modeline göre çalışmaya alınan hastaların depresyon durumunu, ailelerinin durumluluk ve süreklilik kaygı puanları ve yaşları anlamlı düzeyde etkilemektedir. Ancak aile değerlendirme ölçeğine ait puanların ve öfke kontrol puanlarının depresyon puanlarını anlamlı düzeyde etkilemediği görülmüştür. **Sonuç:** Tahmin amacıyla kurulan MARS modelinin belirleme katsayısı (R^2) %83.3 olarak bulunmuştur. Bu sonuç modelin oldukça başarılı tahmin yaptığını göstermektedir ($p < 0.001$). MARS modeli yardımıyla değişkenler arasındaki doğrusal olmayan ilişkiler incelenebilmekte ve yorumlanabilmektedir. Bu nedenle hekimler biyolojik yapıyı daha iyi tanımlama ve anlama fırsatını yakalayacaktır.

Anahtar Kelimeler: MARS, temel fonksiyon, düğüm, ayırma metodu, regresyon analizi

ABSTRACT Objective: In this study by introducing the basic properties of MARS regression model, its advantages over the other regression models were explained and the application steps were emphasized. **Material and Methods:** In the study, given the formulas indicating the theoretical properties of MARS model, an application was conducted with some certain measurements obtained from the study on psychosocial conditions of the families of total 51 patients who were diagnosed as cancer and Behcet disease. By using the study results, a model of MARS, which is likely to be used to estimate the condition of beck depression on these people, was obtained. **Results:** According to the designed MARS model, the depression state of the patients was affected at a statistically significant level by the family state and permanent anxiety scale scores and age. However, it has been observed that the depression scores were not affected by Family Assessment Device and Anger Control Scale scores. **Conclusion:** It has been found that the accuracy level of forecasts of this equation is (R^2) 83.3% ($p < 0.001$). This result indicated that the model had made a successful estimate. The model of MARS provided opportunities for the clinicians to explain the biological structure in a better way since it indicated variations depending on the coefficient and node variables.

Key Words: MARS, base function, node, discrimination method, regression analysis

Türkiye Klinikleri J Biostat 2010;2(2):58-66

Regresyon modellerinde, gözlenen bir sonuç değişkeni üzerinde hangi faktörlerin etkili olduğu ve etkili olan faktörler içerisinde hangisinin etkisinin daha fazla olduğu araştırılmaktadır. Modelde, etkilenen değişken bağımlı, etkileyen değişkenler ise bağımsız değişken olarak adlandırılır.¹

Regresyon modellerinin gerçek yapıyı daha iyi yansıtması için özellikle son yıllarda çok sayıda bağımsız değişken bir arada değerlendirilmektedir. Günümüzde bilgisayar teknolojisindeki değişimlere paralel olarak çok sayıda değişkeni bir arada değerlendirebilen karmaşık algoritmaların kullanılabilirliği artmıştır. Bu algoritmaları kullanan regresyon yöntemlerinden birisi de MARS (*Multivariate Adaptive Regression Splines*, Çok Değişkenli Uyumlu Regresyon Uzanımları) tekniğidir. Bu teknik, 1991 yılında Stanford'da fizikçi ve istatistikçi Jerome Friedman tarafından ilk olarak incelenmiştir. Pratikte yaygın kullanılan doğrusal ve doğrusal olmayan regresyon modellerinin aksine bağımsız değişkenin farklı aralık değerleri için farklı katsayılar türetir ve modele interaksiyon terimlerini de katarak gerçek yapıyı daha iyi yansıtan modeller geliştirir. Ayrıca bağımlı ve bağımsız değişkenlerin dağılımları üzerine her hangi bir varsayım gerektirmez ve değişken tipi de önemli değildir.^{2,3} MARS regresyon setinin performansını geliştirmek için tekrarlamalı ayırma metodunun ve adimsal lineer regresyonun genelleştirilmiş bir hali olarak görülebilir.⁴ MARS, bağımlı ve bağımsız değişkenler arasındaki doğrusal olmayan ilişkileri doğrusal yapıya dönüştürme amacıyla uygun dönüştürme teknikleri kullanır.⁵ MARS, bağımsız değişkenlerin farklı değer aralıklarına karşılık gelen temel fonksiyonları kullanarak esnek bir regresyon modeli oluşturur.

Bu çalışmanın amacı, regresyon modelleri kapsamında yeni uygulama alanları bulan ve karmaşık bir yöntem olan MARS tekniğinin teorik özelliklerini tanıtmak ve tıbbi bir veri seti üzerinde uygulama adımlarını açıklayarak sonuçlarını yorumlamaktır.

GEREÇ VE YÖNTEMLER

MARS MODELİ

Parçalı temel fonksiyonlar ve bunların kombinasyonlarını kullanarak ve regresyon modellerindeki hem ileri doğru hem de geriye doğru ilerleme algoritmalarından yararlanılarak MARS modeli elde edilebilir. Model oluşumu iki adımda gerçekleştirir.⁴

1. Adım: Mümkün olan bütün temel fonksiyonlar oluşturulur. Çok büyük bir model bulununcaya kadar yani modelin karmaşıklığı maksimum seviyeye ulaşıncaya kadar eklenen temel fonksiyonlarla model büyür. Temel fonksiyonlar ya dönüştürülmüş tek değişkeni ya da çok değişkenli interaksiyon terimlerinden meydana gelir. Bu fonksiyonlar modeldeki bağımsız değişkenlerin interaksiyonlarını ve doğrusal olmayan dönüşümlerini hesaba katan bir fonksiyondur. Temel fonksiyonlar oluşturulurken tüm değişken ve değişkenlerin kombinasyonları tek tek ele alınır. Bu durumda, her bir bağımsız değişken ve bu değişkenlerin birbirleri ile olan kombinasyonlarının tanımlı bulunduğu aralıklardaki tüm olası değerler birer düğüm (node) olarak düşünülüp, mümkün olan tüm olası temel fonksiyonlar belirlenir. Temel fonksiyonlar eklenerek model daha karmaşık ve daha esnek bir model haline gelir. Temel fonksiyon sayısı maksimum seviyeye ulaşıncaya kadar bu adım sürer. Ayrıca temel fonksiyonlar oluşturulurken aynı değişkene ait ayna temel fonksiyonu da oluşturulup modele girebilir. Ayna temel fonksiyon, bağımlı değişkenle bağımsız değişken arasındaki dağılımın düğüm noktasına kadar olan eğiminin sıfır olduğunu gösterir.

MARS modeli (I) nolu eşitlikteki gibi tanımlanabilir.

$$Y = \beta_0 + \sum_{k=1}^K \beta_k B_k(X) \quad (I)$$

Burada;

k: Düğüm sayısı

K: Temel fonksiyon sayısı

X: Bağımsız değişken

B_k : Modeldeki sabit terim

β_k : Regresyon katsayısı

$B_k(X)$: k. Temel fonksiyon

biçimindedir.

Temel fonksiyon ise aşağıdaki gibi tanımlanır.

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

Buradaki;

J_k interaksiyon derecesini göstermektedir.

$$[\cdot]_+ = \max[0, \cdot]_+$$

$$s_{kj} : \in [\pm 1]$$

t_{kj} : düğüm değeri ve

x_{wkj} : bağımsız değişken değerini göstermektedir.²

Temel fonksiyonlar bağımsız değişkenlerin doğrusal olmayan dönüşümleri olabilirler. Fakat bağımlı değişken temel fonksiyonların doğrusal bir dönüşümüdür.

2. Adım: Temel fonksiyonlar tahmin hatası en küçük olacak şekilde modelden çıkarılır. Bu adıma budama (pruning) adımı denir ve bu adımı gerçekleştirmek için MARS bir geriye doğru adım algoritması gerçekleştirir. Geriye adım algoritması gerçekleştirebilmek için aşağıdaki adımları yerine getirilmelidir. Modele yalnızca temel fonksiyon sabitini içeren en küçük modele başlanır.

Tahmin hatasını en küçük yapan en uygun ölçümle her değişken ve mümkün bütün düğümler için temel fonksiyon uzayı araştırılır.

Maksimum karmaşıklığa ulaşan bir modeli oluşturuncaya kadar bütün düğümlere ve değişkenlere bakma işlemi tekrarlanır.

Son olarak, budama prosedürüne başvurulur.

Oluşturulan maksimum model budanarak yani önemli bağımsız değişkenler ve bu değişkenlerin interaksiyonları belirlenerek, hata kareler toplamı en küçük olan en uygun model oluşturulur. Budama algoritması en yaygın olarak genelleştirilmiş çapraz geçerlilik (Cross Validation, GCV) tekniği ile yapılır.⁶ Genelleştirilmiş çapraz geçerlilik hem artıkların hatasını, hem de model karmaşığını hesaba katar ve (III) nolu eşitlik yardımıyla hesaplanır.

$$GCV = \frac{\sum_{i=1}^N \hat{y}_i - y_i}{N} + \frac{\sum_{i=1}^N \hat{y}_i^2}{N} \quad (III)$$

Burada,

N: Veri setindeki denek sayısını,

d: Etkili serbestlik derecesi göstermektedir. Etkili serbestlik derecesi bağımsız temel fonksiyonların sayısına eşittir.

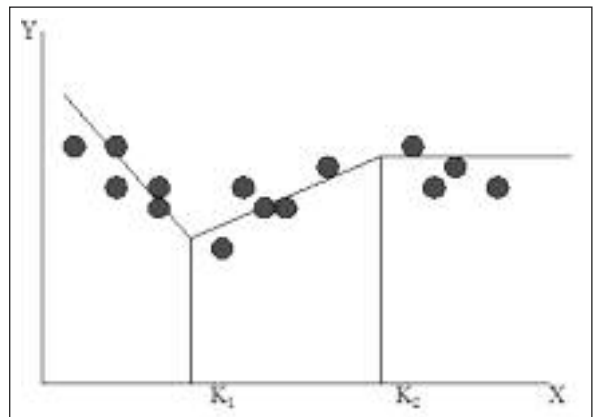
c: eklenen temel fonksiyonların maliyet-karmaşıklık (cost-complexity) ölçüsüdür.

Hesaplamalar şunu göstermiştir ki en iyi C değeri için $2 < d < 3$ bulunmuştur.⁷

DÜĞÜM DEĞERİNİN ELDE EDİLMESİ

Bağımlı ve bağımsız değişken arasındaki ilişkiler doğrusal, eğrisel ve kübik şeklinde olabilir. Aynı bağımsız değişken üzerinde, ilişkinin şeklinin değiştiği bağımsız değişken değerine düğüm değeri denir. Bir başka ifadeyle bağımsız değişken değeri, tanımlı bulunduğu aralıklarda doğrunun eğimini değiştirmeyen en son değer düğüm değeri olarak alınır. Bu düğüm değerlerinde hata kareler toplamı en küçük değerini alır.

Ardışık iki düğüm değerini birleştirerek çizilen doğrunun eğimi, Model 2'de b ile gösterilen regresyon katsayısıdır. Bu haliyle MARS modeli parçalı regresyon modeline benzer. Oluşturulan her temel fonksiyonda bağımsız değişkenden seçilen uygun düğüm değerleri bağımlı değişken ile bağımsız değişken arasında klasik doğrusal dönüştürme yöntemlerini kullanmadan değişkenler arasındaki ilişkiyi parçalı doğrusal ilişkiler haline getirir. İki farklı düğüm değerine sahip bir değişkenin bağımlı değişkenle olan ilişkisi sunulmuştur (Şekil 1). K_1 ve K_2 düğüm değerleridir. Başlangıç noktasından K_1 'e kadar doğrunun eğimi



ŞEKİL 1: X ve Y arasındaki dağılım grafiği.

aynıdır. Benzer şekilde K_1 ve K_2 arasında da doğrunun eğimi aynıdır. Fakat K_1 'den ε (çok küçük bir sayı) kadar uzaklaşsak bile doğrunun eğimi değişir. Doğrunun eğimini değiştirmeyen değer bir düğüm değeridir.

Modelin bağımlı ve bağımsız değişkenlerin dağılımına ilişkin bir varsayımının olmaması, değişkenler arasında doğrusallık varsayımının olmaması nedeni ile önemlidir. Bağımlı değişkeni açıklamada oluşturulan modelde eklemeli bir model kullanır ve bu modelde etkileşime de yer verir.⁸ Diğer alternatif regresyon modellerine göre daha az veri gerektirir ve dögüsel bir model olmasına karşın hızlıdır. Eksik gözlem değerlerinden, aşırı uç değerlerden ve çoklu bağlantıdan çok az etkilenir.⁹ Ayrıca alternatif regresyon modellerinde değişkenlerin yaklaşık %90'ının modele girmesine karşılık MARS modelinde daha az değişken modele girer.⁴

UYGULAMA VERİSİ

Uygulamada kullanılan veriler, Okyayüz ve Ünlüoğlunun 1998'de¹⁰ yapmış oldukları çalışmadan elde edilmiştir. Bu veri setinde kanser ve behçet tanısı konmuş toplam 51 hastanın ailelerinin psiko-sosyal durumlarına ilişkin değerlendirme sonuçları yer almaktadır. Ailelerin psiko-sosyal durumlarını değerlendirmek için durumluluk ve sürekli kaygı ölçeği, aile değerlendirme ölçeği, öfke, kaygı ve mutsuzluk kontrol ölçekleri ve beck depresyon ölçeği uygulanmıştır. Bu ölçümlerde hasta yakınlarının beck depresyon envanteri puanı (BD) bağımlı değişken olarak alınmıştır. Yaş, durumluluk ölçeği olarak (STAT1), sürekli kaygı ölçeği olarak (STAT2), öfke kontrol ölçeği olarak (CECS-1), kaygı kontrol ölçeği olarak (CECS-2), mutsuzluk kontrol ölçeği olarak (CECS-3) ve kontrol ölçeklerinin toplamı olarak (CESCT), aile değerlendirme ölçeklerinden, problem çözme (PÇ), iletişim kurma (iletişim), roller, duygusal tepki verebilme (DUYTEP) gereken ilgiyi gösterme (GERILG), davranış kontrolü (DAVKONT), genel işlevler olarak (GENİŞL) değişkenleri ise bağımsız değişkenler olarak alınmıştır. Bağımlı ve bağımsız değişkenler belirli aralıklarda değişen puan veya toplam puanlardan oluşmaktadır. Ölçülen değiş-

TABLO 1: Değişkenler ve tanımlayıcı istatistikleri.

Değişkenler	Ortalama	SS	Minimum	Maksimum	N
BECKDER	14.76	9.28	3	43	51
YAS	38.33	12.23	17	62	51
STAT-1	45.74	10.97	25	71	51
STAT-2	44.72	9.52	29	69	51
CECS1	16.41	3.61	11	27	51
CECS2	16.08	3.60	9	28	51
CECS3	16.29	4.30	8	27	51
CECST	48.78	9.34	30	74	51
PROBCOZ	1.72	0.53	1	3	51
İLETİŞİM	1.89	0.55	1	3.66	51
ROLLER	1.99	0.37	1.27	2.63	51
DUYTEP	1.88	0.58	1	3.5	51
GERILG	2.38	0.41	1.42	3.57	51
DAVKONT	1.88	0.42	1.22	3.57	51
GENİŞL	1.69	0.44	1	2.58	51

SS: Standart Sapma

kenlere ait tanımlayıcı istatistikler topluca sunulmuştur (Tablo 1).

MARS modelinin hesaplamalarda MARS 2.0 paket programı kullanılmıştır. MARS 2.0 programı Windows ortamında çalışabilen bir programdır. Kullanılan program Hearne Scientific Software firmasının bir yazılımıdır ve 30 günlük deneme programı internetten ücretsiz indirilebilir. Deneme sürümü için erişim adresi:

<http://salford-systems.com/products/mars/overview.html>

BULGULAR

Veri içinde çok az miktarda kayıp gözlem olduğu için bu değerler ihmal edilmiştir. Modele alınan değişkenlerden CECST değişkeni CECS1, CECS2 ve CECS3 değişkenlerindeki puanların cebirsel toplamı olduğu için aralarında yüksek bir korelasyon mevcuttur. Bu yüksek korelasyon çoklu bağlantı olarak adlandırılır ve klasik regresyon modellerinde genellikle problem yaratır. Ancak MARS modelinde bu ilişki çok yüksek olmadıkça bir problem teşkil etmez. Bu durumu göstermek amacıyla CECST değişkeni kendi parçaları ile birlikte modele alınmış ve buna ilişkin sonuçlar açıklanmıştır. Kurulan model ana etkiler modeli olup interaksiyon terimleri modelde incelenmemiştir.

TABLO 2: Maksimum model sonuçları.

Temel Fonk.	GCV	Değişkenler	Düğüm değeri
0	87.989		
2 1	54.355	STAI_1	42
4 3	44.313	GERILG	2.28
5	40.730	STAI_2	29
6	40.658	PROBCOZ	1
7	41.570	CECS3	8
8	41.268	CECST	30
10 9	38.367	YAS	25
12 11	39.924	DUYTEP	2.5
13	37.718	GENISL	1
15 14	39.175	STAI_1	39
17 16	41.495	STAI_1	56
19 18	44.918	PROBCOZ	2.16
21 20	48.467	GENISL	1.66
22	52.186	ROLLER	1.27
24 23	57.573	STAI_1	50
25	64.378	CECS2	9

Model inşasının ilk aşamasında, ileriye doğru adım prensibi ile eklemeli maksimum model elde edilmiş ve maksimum modele ait genelleştirilmiş çapraz geçerlilik testi sonuçları, değişkenlere ait temel fonksiyonlar ve düğüm değerleri verilmiştir. Ayna fonksiyonuna sahip değişkenler sırasıyla STAT-1, GERILG, YAŞ, DUYTEP, PROBCÖZ, GENİŞL değişkenleridir (Tablo 2). Ayna fonksiyonu olmayan değişkenler ise STAT-2, CECS3, CECST, CECS2 ve ROLLER dir. Yine bazı değişkenlerin tabloda birkaç kez tekrarladığı görülmektedir. Bu ise o değişkenlere ait birden fazla düğüm noktası belirlendiğini göstermektedir. Örneğin STAT-1 değişkeninin düğüm değerleri sırasıyla 39, 42, 50 ve 56 olmak üzere 4 tanedir. Düğüm değerleri kısaca şöyle yorumlanır. STAT-1 ile BD arasındaki ilişki, STAT-1'in düğüm değerlerinde farklılık göstermektedir. Bir başka ifadeyle STAT-1'in 42 ile 50 arasında değer aldığı durumda BD ile ilişkisinin şekli, 50 ile 56 arasında değer aldığı durumdan farklıdır. Bu durum diğer düğüm değerleri için de geçerlidir. Maksimum modelde sabit terimle beraber toplam 26 temel fonksiyon yer almaktadır. Modelde CECS1 değişkeni CECST değişkeni ile çok yüksek ilişki içinde olduğu için ve CECST değişke-

ninin modele katkısı daha yüksek olduğu için CECS1 değişkeni modelden atılmış CECST kalmıştır. Buna ilaveten İLETİŞİM ve DAVKONT değişkenleri de başlangıçta belirlenen temel fonksiyon sayısına göre modele katkısı olmadığı için modele hiç alınmamıştır. Ayrıca STAT-2, CECS2, CECS3, ROLLER ve CECST değişkenlerinin düğüm değerleri birer tane olup bu değişkenlerin kendi değerlerinin minimumuna eşittir (Tablo 2). Bu durumda söz konusu 5 değişken ile BD arasında doğrusal bir ilişkinin olduğunu söyleyebiliriz. Maksimum model aşırı doymuş bir model olup yeni veri setlerine uyumu düşüktür.

İkinci aşamada geriye doğru adım yöntemi kullanılarak, maksimum modeli daha basit bir hale getirmek için budanmıştır. Budama yöntemi olarak 10 katlı çapraz geçerlilik testi tercih edilmiştir ve oluşturulacak olan en iyi model elde edilmiştir (Tablo 3). Bu modelin GCV değeri en küçüktür (GCV=33.472) ve modelin belirleme katsayısı %83.3'dür. Bu sonuç modelin oldukça başarılı tahmin yaptığını göstermektedir ($p < 0.001$). Ayrıca regresyon modelinin standart hatası ise 4.3 olarak hesaplanmıştır. Son modele sabit terimle beraber 12 tane temel fonksiyon girmiştir. Bunlar içerisinde BF4 ve BF12 (GERILG ve DUYTEP değişkenlerine ait) ayna temel fonksiyonlarıdır. Ayrıca STAT-1 değişkeni farklı iki düğüm noktasında modele girmiştir. Son modelde bazı değişkenlere ait sadece bir temel fonksiyon kullanılmıştır. GERILG,

TABLO 3: Son (en uygun) modele ait bilgiler.

Temel Fonk.	Regresyon Katsayısı	Değişkenler	Düğüm
0	1.272		
1	2.506	STAI_1	42
3	-7.242	GERILG	2.28
4	-21.919	GERILG	2.28
5	0.228	STAI_2	29
7	1.628	CECS	8
8	-0.605	CECST	30
9	0.142	YAS	25
11	10.433	DUYTEP	2.5
12	5.094	DUYTEP	2.5
14	-1.530	STAI_1	39
20	9.926	GENISL	1.66

STAI_2, CECS3, CECST, YAS ve GENISL değişkenleri bu grupta olup bunların düğüm değerlerinden daha büyük olan değerler modele alınmıştır. Çünkü bu değişkenlerin düğüm değerlerinden daha küçük olan değerleri için eğim sıfıra eşittir. Geriye kalan değişkenlerin, modele giren temel fonksiyon sayısı ise 2'dir. Temel fonksiyon sayısının birden fazla olması, o değişkenin doğrusal bir ilişki içinde olmadığını ve belirlenen düğüm değerine göre parametrik olmayan bir doğrusallaştırma ile ilişkinin doğrusala dönüştürüldüğünü gösterir.

Bu sonuçlara göre temel fonksiyonlar ve temel fonksiyonlardan yararlanarak oluşturulan tahmin denklemi (IV) nolu eşitlikteki gibidir.

$$BF1 = \max(0, STAI_1 - 42.000)$$

$$BF3 = \max(0, GERILG - 2.280)$$

$$BF4 = \max(0, 2.280 - GERILG)$$

$$BF5 = \max(0, STAI_2 - 29.000)$$

$$BF7 = \max(0, CECS3 - 8.000)$$

$$BF8 = \max(0, CECST - 30.000)$$

$$BF9 = \max(0, YAS - 25.000)$$

$$BF11 = \max(0, DUYTEP - 2.500)$$

$$BF12 = \max(0, 2.500 - DUYTEP)$$

$$BF14 = \max(0, STAI_1 - 39.000)$$

$$BF20 = \max(0, GENISL - 1.660)$$

$$Y = 1.272 + 2.506BF1 - 7.242BF3 - 21.919BF4 + 0.228BF5 + 1.628BF7 - 0.605BF8 + 0.142BF9 + 10.433BF11 + 5.094BF12 - 1.530BF14 + 9.926BF20 \quad (IV)$$

Bu modelde STAT-1 değişkeninin 39'dan önceki değerleri için eğim sıfırdır, 39 ile 42 değerleri arasındaki eğimle 42'den büyük değerleri için eğim katsayıları da birbirinden farklıdır. Şekil 1'de ilişkinin şekli daha açık görünmektedir.

Değişkenlerin modele olan katkılarına göre sıralaması Tablo 4'de verilmiştir. Önemlilik değeri düşük olan değişkenin modele katkısı da azdır. Bu değerler modele alınan bütün değişkenler için hesaplanmıştır. Tablo incelendiğinde modele katkısı en fazla olan veya bir başka ifadeyle BD üzerine en etkili olan değişken STAT-1 değişkenidir. CECS1, PROBCOZ, ILETISIM, ROLLER ve DAVKONT değişkenlerinin önemlilik değerlerinin sıfır çıkma-

TABLO 4: Değişkenlerin önemlilik oranları.

Değişkenler	Önemlilik
STAI_1	100.000
CECS3	72.403
GERILG	67.939
CECST	55.140
GENISL	53.168
STAI_2	31.262
DUYTEP	17.635
YAS	16.931
CECS1	0.000
CECS2	0.000
PROBCOZ	0.000
ILETISIM	0.000
ROLLER	0.000
DAVKONT	0.000

sı bu değişkenlerin BD'deki varyasyonu açıklamada anlamlı bir etkiye sahip olmadığını göstermektedir.

Sonuçların yorumlanmasında yardımcı olan bir diğer tablo ise son modele giren temel fonksiyonlar (sabit terim hariç) için oluşturulan varyans analiz tablosudur (Tablo 5). Bu tabloda her bir değişkenin BD üzerine genel etkisi yer almaktadır. ANOVA tablosundaki standart sapmaların büyüklüğü o değişkenin bağımlı değişken üzerine olan genel etkisinin fazla olduğunu gösterir. Örneğin STAT-1 değişkeni toplam 2 temel fonksiyon ile

TABLO 5: 11 temel fonksiyon için ANOVA Decomposition sonuçları.

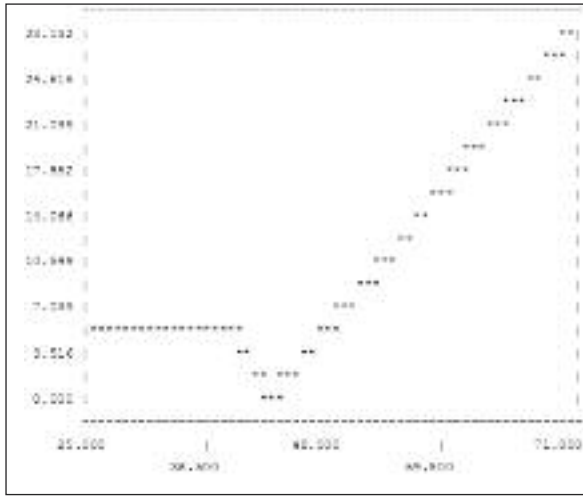
Fonksiyon	Çıkmadaki		Temel	Etkili	Değişken
	SS	kayıp	fonksiyon sayısı	parametre sayısı	
1	6.582	64.180	2	3.066	STAI_1
2	4.105	47.646	2	3.066	GERILG
3	2.154	36.474	1	1.533	STAI_2
4	6.931	49.570	1	1.533	CECS3
5	5.594	42.809	1	1.533	CECST
6	1.521	34.353	1	1.533	YAS
7	2.279	34.427	2	3.066	DUYTEP
8	2.845	42.153	1	1.533	GENISL

SS: Standart sapma

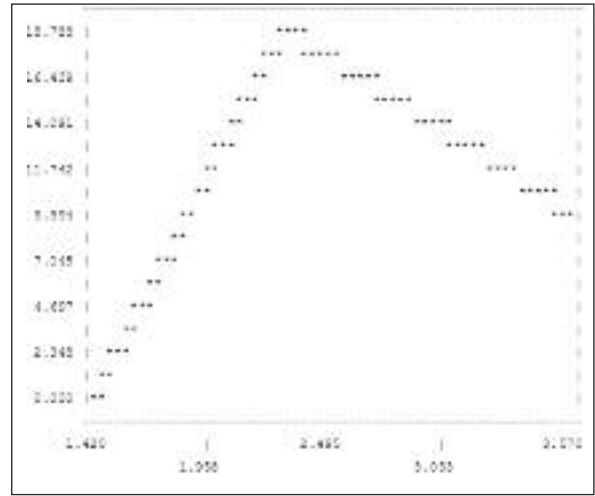
modele girmiş ve standart sapması incelendiğinde BD üzerine en etkili olan değişkenlerden birisi olduğu görülmüştür. Bu sonuç, değişken önemlerine ilişkin Tablo 4'teki açıklamalarla paralel bir sonuçtur. Genel etkisi en düşük olan değişken ise YAŞ olarak belirlenmiştir. Ayrıca Tablo 5'in üçüncü sütununda yer alan değerler, söz konusu değişkene ait temel fonksiyonları modelden çıkardığımız durumda, tahminlerde meydana gelecek kaybı göstermektedir. Standart sapma ile bu değerler zıt ilişki içindedir. Standart sapma söz konusu değişkenin modeldeki varyasyonun ne kadarını açıkladığını göstermektedir. Bu nedenle standart sapma değeri büyük olan değişkenin modelden çıkarılması duru-

munda tahminlerde kötüleşme beklenir. Bu nedenle SS değeri büyük olan değişkenin kayıp değeri de büyüktür.

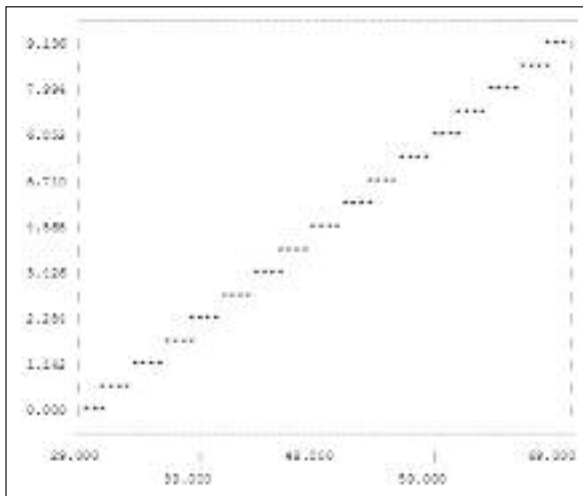
Bunlara ilaveten Şekil 2 ile Şekil 9 arasında son modelde kalan değişkenlerin uygun düğüm değerleri dikkate alınarak dönüşüm yardımıyla doğrusallaştırılmış hallerinin, bağımlı değişken (BD) ile ilişkilerini gösteren iki boyutlu grafikler görünmektedir. Ayrıca bu değişkenlere ait düğüm değerleri de grafiklerde görünmektedir. Şekillerdeki düğüm değerleri yardımı ile bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkinin şeklinin hangi noktalarda değiştiği gözlemlenebilir.



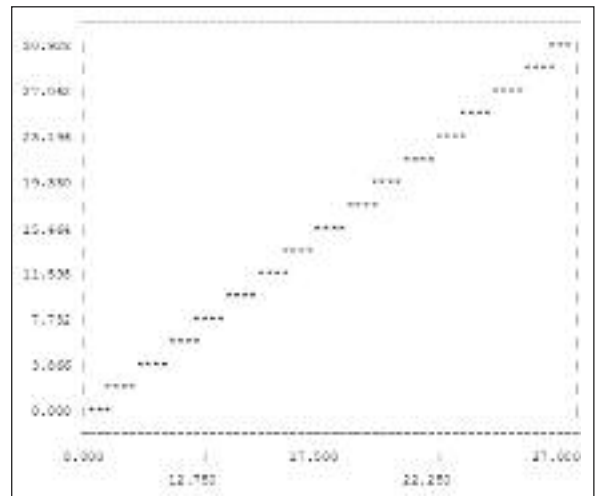
ŞEKİL 2: STAT-1 ile BD arasındaki ilişki ve düğüm değeri.



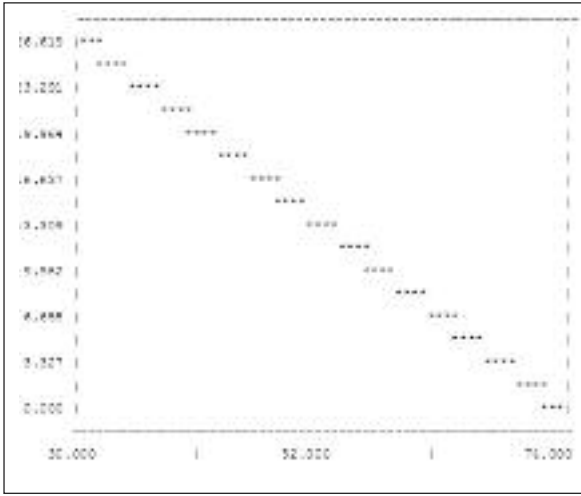
ŞEKİL 3: GERILG ile BD arasındaki ilişki ve düğüm değeri.



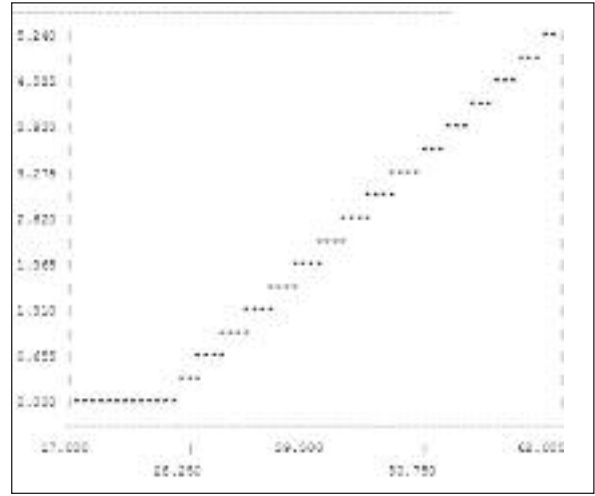
ŞEKİL 4: STAT-2 ile BD arasındaki ilişki ve düğüm değeri.



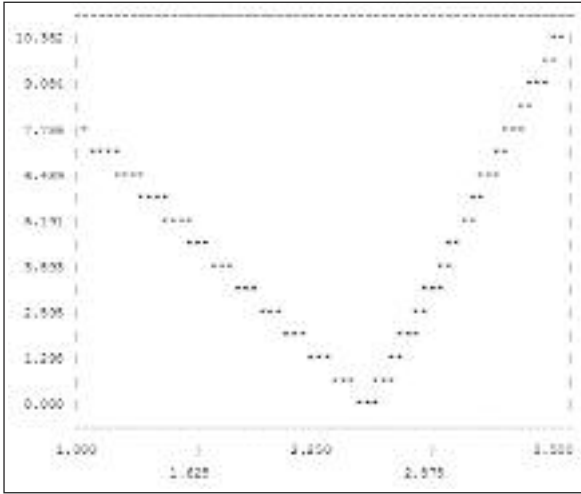
ŞEKİL 5: CECS3 ile BD arasındaki ilişki ve düğüm değeri.



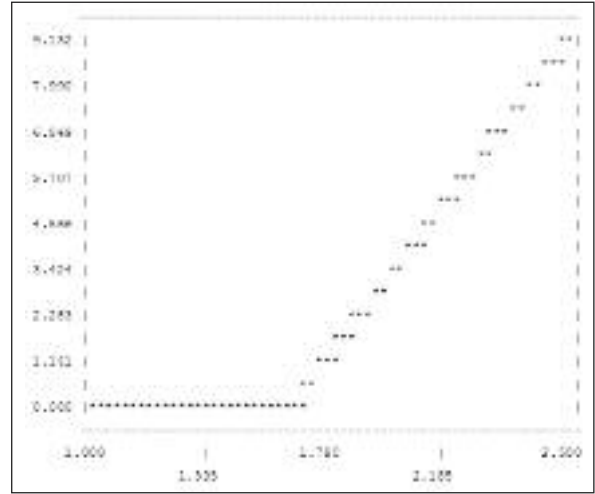
ŞEKİL 6: CECST ile BD arasındaki ilişki ve düğüm değeri.



ŞEKİL 7: YAŞ ile BD arasındaki ilişki ve düğüm değeri.



ŞEKİL 8: DUYTEP ile BD arasındaki ilişki ve düğüm değeri.



ŞEKİL 9: GENISL ile BD arasındaki ilişki ve düğüm değeri.

TARTIŞMA

Bu çalışmada kişilerin depresyon puanlarına etkili olan değişkenler klasik yöntemlerden farklı olan MARS modeli ile ortaya çıkarmaya çalışılmıştır. Ayrıca her bir değişkenin depresyon puanı ile doğrusal ilişkisini tanımlı buldukları aralıklarda kırılma noktalarını vererek grafikler ve temel fonksiyonlar aracılığı ile sunmaya çalışılmıştır. Sağlık verilerinin yorumlanmasında MARS modeli, bağımsız değişkenlerin alt gruplarındaki dağılım hakkında fikir verdiği, çoklu bağlantıdan, aşırı uç değerlerden ve kayıp verilerden çok az etkilendiği ve değişkenlerin dağılımları üzerine bir varsayım taşımadığı için klasik regresyon modellerine göre üstündür.

Doğrusal modellerdeki en temel varsayım katsayı değerinin, bağımsız değişkenin bütün değerlerinde sabit kalmasıdır. Buna karşın MARS modelinde katsayı değerleri düğüm değerlerine bağlı olarak değişim gösterir. Bu durum biyolojik yapıyı daha iyi açıklama fırsatı sağlar.² Doğrusal ilişkinin var olduğu modellerde doğrusal regresyon modeli ile MARS modelinden elde edilen eğim katsayıları aynı olmasına karşın modelin sabit terimi aynı çıkmaz.

MARS modeli doğrusallık varsayımı gerektirmeyen CART modeli ile karşılaştırıldığında CART'ın sadece interaksiyon etkisini ortaya çıkardığı görülmüştür. Buna karşın MARS modeli interaksiyon etkisinin yanında esas etkileri de ortaya

çıkarr. Ayrıca MARS modelinde bir tahmin modeli elde edilirken CART analizinde görsel açıklanabilen bir şema elde edilmektedir.⁵

SONUÇ

MARS modeli sağladığı avantajlar nedeniyle günümüzde sıklıkla kullanılan bir regresyon modeli

olmalıdır. En önemli dezavantajı sonuçların yorumlanmasının zor olmasıdır. Bu nedenle yorumlamaları dikkatli yapmak gerekir. Tıp gibi karmaşık biyolojik ilişkilerin incelendiği bir bilim dalı için gerçek yapıyı tanımlamada model iyi kurulduğu takdirde mükemmel yakın sonuçlar verebilir.

KAYNAKLAR

1. Rawlings J. Applied Regression Analysis: A Research Tool. 2nd ed. America, Brooks/ Cole Publishing Company; 1988. p.1-7.
2. Friedman JH. Multivariate Adaptive Regression Splines. *Annals of Statistics* 1991;19(1): 1-141.
3. Yu B. Approximating the risk score for disease diagnosis using MARS. *J Appl Stat* 2009;36(7):769-78.
4. Leathwick JR, Elith J, Hastie T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* 2006;199(2):188-96.
5. Deichmann J, Eshghi A, Haughton D, Sayek S, Teabagy N. Application of Multiple Adaptive Regression Splines (MARS) in Direct Response Modelling. *Journal of Interactive Marketing* 2002;16 (4):15-27.
6. Balshi M, McGuic D, Duffys P, Flannigan M, Walsh J, Melillo J. Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology* 2009;15(3): 578-600.
7. York TP, Eaves LJ, van den Oord EJ. Multivariate adaptive regression splines: a powerful method for detecting disease-risk relationship differences among subgroups. *Stat Med* 2006;25(8):1355-67.
8. Septhon P. Forecasting recessions: can we do better on MARS? *Federal Reserve Bank of St. Louis Review* 2001;83(2):39-49.
9. Chen I, Lee T. A two-stage credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Syst Appl* 2005;28(4):743-52.
10. Okyayüz ÜH, Ünlüoğlu G. An assesment of various emotions consciously controlled reactions and family functions, concerning the relatives who have malignant and chronic illnesses like cancer and Behçet Disease. *The 12th Conference of the European Health Psychology Society (EHPS) Abstract Book*. Vienna; 1998. p.143.