



**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**



**YAŞAM ÇÖZÜMLEMESİNE YENİ BİR
YAKLAŞIM: MARS**

Mesut AKYOL

**BIYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ**

**DANIŞMAN
Doç. Dr. Atilla Halil ELHAN**

2011- ANKARA

**TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**YAŞAM ÇÖZÜMLEMESİNE YENİ BİR
YAKLAŞIM: MARS**

Mesut AKYOL

**BIYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ**

**DANIŞMAN
Doç. Dr. Atilla Halil ELHAN**

2011- ANKARA

İÇİNDEKİLER

Kabul ve Onay	ii
İçindekiler	iii
Önsöz	v
Simgeler ve Kısaltmalar	vi
Şekiller	vii
Çizelgeler	viii
1. GİRİŞ	1
1.1. Yaşam Çözümlemesi Terminolojisi	3
1.1.1. Olgu (Event)	3
1.1.2. Takip (izlem) süresi (Follow-up time)	4
1.1.3. Yaşam süresi ya da başarısızlık zamanı (Failure Time)	4
1.1.4. Sansürleme (Censoring)	5
1.2. Yaşam Çözümlemesinde Kullanılan Fonksiyonlar	6
1.2.1. Olasılık yoğunluk fonksiyonu	6
1.2.2. Yaşam fonksiyonu (Survival function)	7
1.2.3. Hazard fonksiyonu (Hazard function)	8
1.3. Cox Orantısal Hazard Modeli	10
1.3.1. Cox orantısal hazard modelinde artıklar	12
Martingale artıklar	13
Deviance artıklar	14
1.3.2. Cox regresyon modelinin olumsuz yanları	15
1.4. Çok Değişkenli Uyarlanabilir Regresyon Kesitleri	16
1.4.1. MARS yöntemini kullanan bazı çalışmalar	17
1.4.2. MARS modeli	19
1.4.3. Temel fonksiyon (Basis function)	23
1.4.4. Adımsal ilerleme (Modeli geliştirmek)	25
1.4.5. Adımsal gerileme (Modeli budamak)	26
1.4.6. KNOT (Düğüm noktası)	26
1.4.7. Düğüm değerinin elde edilmesi	28
1.4.8. Ayna-görüntü temel fonksiyonlar	31
1.4.9. Genelleştirilmiş çapraz geçerlilik	32
1.4.10. Yaşam çözümlemesinde MARS kullanım örnekleri	33
1.5. MARS ve Yaşam Çözümlemesi	36
1.5.1. MARS yönteminin yaşam çözümlemesine uyarlanması	36

2. GEREÇ VE YÖNTEM	40
2.1. İstatistiksel Analizler	40
3. BULGULAR	42
3.1. Karsinom Gerçek Veri Uygulaması	42
3.1.1. Karsinom: Tanımlayıcı istatistikler	42
3.1.2. Karsinom: Cox-PH modeli sonuçları	43
3.1.3. Karsinom: MARS modeli sonuçları (Martingale)	47
3.1.4. Karsinom: MARS modeli sonuçları (Deviance)	50
3.2. UMARU – UIS Veri Seti Uygulaması	53
3.2.1. UIS: Tanımlayıcı istatistikler	54
3.2.2. UIS: Cox-PH modeli sonuçları	55
3.2.3. UIS: MARS modeli sonuçları (Martingale)	59
3.2.4. UIS: MARS modeli sonuçları (Deviance)	64
4. TARTIŞMA	70
5. SONUÇ VE ÖNERİLER	72
ÖZET	75
SUMMARY	76
KAYNAKLAR	77
ÖZGEÇMİŞ	81

ÖNSÖZ

Bu tez çalışması, yaşam çözümlemesinde kullanılan yöntemlere alternatif olarak kullanılabilir yeni bir yöntemin geliştirilmesi ve yaşam çözümlemesinde kullanılabilirliğinin ispatlanması amacı ile yapılmıştır. Önerilen MARS yönteminin kullanılması ile yaşam çözümlemesinde bilinen yöntemlerin eksiklerinin giderilebileceği ya da farklı bir bakış açısı ile yeniden değerlendirilebileceği düşünülmektedir.

Hocalarımla aktardığı bilgiler ve tecrübeleri olmasa bu aşamaya gelemezdim. Arkadaşlarımla destekleri olmasaydı bu çalışma sonuçlandırılmazdı. Bu nedenle Biyoistatistik camiasındaki tüm hocalarımla ve arkadaşlarıma teşekkür etmek istiyorum.

Gülhane Askeri Tıp Akademisinde 2001 yılından bugüne kadar birlikte çalıştığım komutanlarıma ve hocalarıma da saygılarımı sunuyorum. Onların gösterdikleri anlayış sayesinde doktora eğitimimi ve tezimi tamamlayabildim.

1992 yılından beri bana sürekli güvenen, her koşulda ve durumda beni destekleyen, tez çalışmamı biran önce tamamlamam için sürekli olarak beni motive eden Doç.Dr. S. Yavuz Sanisoğlu'na şükranlarımı sunuyorum.

Danışmanım Doç.Dr. Atilla Halil Elhan'a ve tez izleme komitemdeki Prof.Dr. Osman Saraçbaşı ile Yrd.Doç.Dr. S. Kenan Köse'ye yaptıkları katkılar ve destekleri için şükranlarımı sunuyorum.

Doktora eğitimim ve tez çalışmam sırasında desteklerini sürekli hissettiğim, daima bana karşı anlayışla yaklaşan aileme, özellikle de eşim Nilgün ve kızlarım Nadide Sena ile Nur Sima'ya anlayış ve destekleri için sonsuz teşekkürler.

Simgeler ve Kısaltmalar

T, t	:	Sağkalım zamanı
e, ε	:	Artık, hata
C_i	:	Sansür zamanı
δ_i	:	Sansür değişkeni
Cox-PH	:	COX'un orantısal hazard fonksiyonu
$S(t)$:	Yaşam fonksiyonu
$f(t)$:	Olasılık yoğunluk fonksiyonu
$F(t)$:	Dağılım fonksiyonu
$\lambda(t)$:	Hazard fonksiyonu
$\Lambda(t)$:	Birikimli Hazard fonksiyonu
HR	:	Hazard hızı
D, d	:	Ceza parametresi
ÇAG	:	Çeyreklikler arası genişlik (Interquartile Range)
MARS	:	Çok değişkenli uyarlanabilir regresyon kesitleri
$MARS_M$:	Martingale artıklarını kullanan MARS modeli
$MARS_D$:	Deviance artıklarını kullanan MARS modeli
GCV	:	Genelleştirilmiş çapraz geçerlilik
Knot	:	Düğüm noktası
B, BF	:	MARS'ta temel fonksiyon
K	:	MARS modelinde düğüm (knot) sayısı
β	:	Regresyon modellerinde katsayı
C	:	MARS'ta maliyet-karmaşıklık değeri
σ^2	:	Varyans
R^2	:	Açıklayıcılık değeri
SST	:	Kareler toplamı
SSE	:	Açıklanan kareler toplamı
-2LL	:	-2 log olabilirlik (-2 log Likelihood)

Şekiller

Şekil 1.4.3. Temel Fonksiyonlara Karşılık Ayna Değişken	25
Şekil 1.4.6. Kesit ve düğüm kullanılarak MARS veri tahmini	27
Şekil 1.4.7.1. X ve Y dağılım grafiği ve düğümler (K_1, K_2)	29
Şekil 1.4.7.2. X değişkeni 0-100 aralığı için $c=10$ 'dan 80'e kadar temel fonksiyonların değişimi (hokey sopası)	31
Şekil 3.1.2.1. Karsinom veri seti: Birikimli sağkalım fonksiyonu	45
Şekil 3.1.2.2. Karsinom veri seti: Final Cox-PH modeli birikimli sağkalım fonksiyonu	46
Şekil 3.1.2.3. Karsinom veri seti: Final Cox-PH modeli birikimli hazard fonksiyonu	46
Şekil 3.1.3.1. Martingale artıklara göre Platelet kesim noktaları	48
Şekil 3.1.4.1. Deviance artıklara göre Platelet kesim noktaları	51
Şekil 3.2.2.1. UIS veri seti: Birikimli sağkalım fonksiyonu	56
Şekil 3.2.2.2. UIS veri seti: Final Cox-PH modeli birikimli sağkalım fonksiyonu	58
Şekil 3.2.2.3. UIS veri seti: Final Cox-PH modeli birikimli hazard fonksiyonu	58
Şekil 3.2.3.1. Beck depresyon skoru (Bectota) ile Martingale artıkları arasındaki ilişki	60
Şekil 3.2.3.2. Tedavi süresi (Lot) ile Martingale artıkları arasındaki ilişki	61
Şekil 3.2.4.1. Yaş (Age) ile Deviance artıkları arasındaki ilişki	66
Şekil 3.2.4.2. Beck depresyon skoru (Becktota) ile Deviance artıkları arasındaki ilişki	66
Şekil 3.2.4.3. Tedavi süresi (Lot) ile Deviance artıkları arasındaki ilişki	67

Çizelgeler

Çizelge 3.1. Karsinom veri setindeki değişkenler	42
Çizelge 3.1.1. Karsinom veri seti tanımlayıcı istatistikleri	43
Çizelge 3.1.2.1. Karsinom veri seti: Cox-PH modeli başlangıç değişkenleri ve modele katkıları	44
Çizelge 3.1.2.2. Karsinom veri seti: Final Cox-PH modeli değişkenleri ve modele katkıları	45
Çizelge 3.1.3.1. Karsinom veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_M$)	47
Çizelge 3.1.3.2. Karsinom veri seti: Cox-PH modeli değişkenleri ve modele katkıları ($MARS_M$)	49
Çizelge 3.1.4.1. Karsinom veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_D$)	51
Çizelge 3.1.4.2. Karsinom veri seti: Cox-PH modeli değişkenleri ve modele katkıları ($MARS_D$)	52
Çizelge 3.2. UIS veri setindeki değişkenler	53
Çizelge 3.2.1. UIS veri seti tanımlayıcı istatistikleri	54
Çizelge 3.2.2.1. UIS veri seti: Cox-PH modeli başlangıç değişkenleri ve modele katkıları	55
Çizelge 3.2.2.2. UIS veri seti: Final Cox-PH modeli değişkenleri ve modele katkıları	57
Çizelge 3.2.3.1. UIS veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_M$)	59
Çizelge 3.2.3.2. UIS veri seti: Cox-PH modeli değişkenleri ve modele katkıları ($MARS_M$)	62
Çizelge 3.2.4.1. UIS veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_D$)	64
Çizelge 3.2.4.2. UIS veri seti: Cox-PH modeli değişkenleri ve modele katkıları ($MARS_D$)	68

1. GİRİŞ

Yaşam çözümlemesi (survival analysis), belirli bir hastalığa maruz kalan bireyin hastalığın tanısından veya uygulanan bir girişimden sonra ne kadar süre daha yaşayabileceğini ya da hastalığın ne kadar sürede tekrarlanabileceğini (nüks derecesini) tahmin etmek, tedavi tiplerinin ve diğer faktörlerin yaşam süresine etkilerini inceleyebilmek amacıyla geliştirilmiş yöntemler ailesidir (Özdamar, 2003).

Canlılarda başarısızlık ya da ölüm zamanının çözümlenmesine ilişkin çalışmalar, sigorta şirketleri tarafından yaşam tablolarının hazırlanması ile başlamıştır. Bu çalışmalar daha sonra geliştirilerek başarısızlık modeli ya da risk (hazard) modeli olarak adlandırılan genel modellerle, herhangi bir olayın olma ya da gözlemlenme zamanının çözümlenmesinde kullanılmıştır (Köse, 1997).

Yaşam çözümlemesi, T zaman süresinde (araştırma periyodu) izlenen n sayıda hasta birimden elde edilen yaşam sürelerinin (izlem süresi) dağılımını açıklayarak, yaşam süresini etkileyen ve etkilemesi olası değişkenleri içeren modeller kurarak, bu modellere göre parametre tahminleri yapmayı amaçlamaktadır (Özdamar, 2003).

Yaşam çözümlemesinde araştırmacının amacı;

- a. Tedaviden sonra hastaların beklenen yaşam sürelerinin kestirilmesi,
- b. Genel sağkalım veya hastalıksız sağkalım eğrilerinin elde edilmesi,
- c. Tedavinin etkinliğinin belirlenebilmesi amacı ile tedavi edilen ve edilmeyen ya da farklı yöntemlerle tedavi edilen hastaların yaşam sürelerinin karşılaştırılması,

d. Hastalık ve dolayısıyla yaşam süresi ile bu süreye etki eden risk faktörleri arasındaki ilişkiyi ortaya koymak,

e. Hastalıkların doğal seyri için incelenmesi olarak sıralanabilir (Saraçbaşı, 2002; Dişçi, 2008).

Yaşam fonksiyonunu bulabilmek için yaşam sürelerinin dağılımını (üstel, gamma, Weibull, log-normal vb) bilmek gerekir. Yaşam sürelerinin dağılımı bilindiğinde uygun yaşam fonksiyonu kullanılarak yaşam süreleri değerlendirilip yorumlanabilir. Yaşam sürelerinin bilinen teorik bir dağılıma dayanarak incelenmesi parametrik yaşam çözümlenmesi yöntemleri ile yapılır. Ancak yaşam süreleri genellikle, canlılar ile uğraşan (tıp ve sağlık alanı gibi) bilimlerde belirli bir teorik dağılıma uymaz. Bu nedenle sağlık bilimlerinde parametrik olmayan yaşam çözümlenmesi yöntemleri daha sık kullanılmaktadır (Saraçbaşı, 2002).

Sağlık alanında birçok kronik hastalıkta olduğu gibi, bir hastalığın varlığı ya da yokluğu tek başına yaşam ya da ölüm açısından belirleyici değildir. Hastalık dışında birçok faktör yaşam süresini uzatabilir veya ölümleri hızlandırabilir. Bu duruma bireylerin yaşı, hastalığın tipi, hastalık evresi, tedavi yöntemleri gibi birçok faktör örnek olarak verilebilir. Hastalık tipini ve araştırma konusuna göre belirlenen faktörleri dikkate alarak çözüm bulmak gerekir (Terzi ve ark., 2005).

Yaşam çözümlenmesinde; gerek parametrik gerekse parametrik olmayan yöntemlerin kullanımında araştırmacı tek gruba ilgilenmektedir. Oysa araştırmacılar için tek grubun yaşam eğilimi yanında birden fazla grubun yaşam eğrilerinin farklı olup olmadığı ya da yapılan tıbbi bir girişimin (ilaç tedavisi, cerrahi müdahale, terapi vb.) yaşam eğilimini nasıl etkilediğinin bilinmesi gereklidir. Bu ise farklı grupların yaşam fonksiyonlarının birlikte incelenmesi ve karşılaştırılması ile yapılabilir. Bununla birlikte her bir birey için yaşam sürelerine etki edebileceği düşünülen açıklayıcı değişkenlerin de

araştırılması ve yaşam süreleri üzerine tek tek ve birlikte etkilerinin de incelenmesi gereklidir (Köse, 1997; Saraçbaşı, 2002).

Yaşam çözümlemesi için kullanılan klasik yöntemlerde (yaşam tablosu, Kaplan-Meier gibi) incelenen birimin başarısızlığına etken olarak zamanın etkileri incelenmekte ancak bireyden kaynaklanan farklılıklar ve bireye ait özellikleri risk (hazard) modeline katılmamaktadır. Cox'un (1972) makalesi ve arkasından gelen Kalbfleisch, Prentice, Tsiatis, Oakes ve Breslow'un çalışmaları, yine Cox'un (1984) "*Analysis of Survival Data*" adlı çalışması, sağkalım analizi araştırmalarına yeni bir yaklaşım getirmiştir. Bu yaklaşımda; bireylerin özelliklerinin de hazard modeline eklenmesi önerilmiştir (Cox, 1972; Cox 1984; Köse, 1997).

1.1. Yaşam Çözümlemesi Terminolojisi

Yaşam çözümlemesinde kullanılan terminolojiyi, kavram karmaşasına neden olmamak amacı ile yeniden ifade etmenin faydalı olacağı düşünülmüştür. Bu çalışmada kullanılacak olan terimler aşağıda kısaca açıklanmıştır.

1.1.1. Olgu (Event)

Yaşam çözümlemesinde ilgilenilen olay, olgu (event) olarak adlandırılır. Olgu; belirli bir tedavi gördükten sonra takip altında tutulan bir bireyin ölmesi olabileceği gibi tedavi edilen bir hastalığın tekrar görülmesi (nüks) olarak da kabul edilebilir.

1.1.2. Takip (izlem) süresi (Follow-up time)

İlgilenilen olgunun gerçekleşip gerçekleşmeyeceğini, gerçekleşecek ise ne kadar süre sonra gerçekleşeceğini belirleyebilmek amacı ile seçilen n büyüklüğündeki örneklemin çalışma başlangıcından itibaren izlem altında tutulacağı süreyi (T) ifade eder.

Takip süresi çalışmanın boyutuna ve cinsine bağlı olarak birkaç saat ya da gün olabileceği gibi yıllarla ifade edilen bir zaman dilimi de olabilir. Örneğin ratlarda yapılan bir zehir deneyinde ratlara uygulanan %5, %10 ve %20'lik konsantrasyonlar sonrasında beklenen ölüm zamanları dakikalarla ifade edilebileceğinden takip süresi de dakikalarla ifade edilir ve ölçümler daha hassas analizler yapılabilmesi amacı ile saniyeler ile belirlenebilir. Başka bir çalışmada yeni evlenen çiftlerin ne kadar süre sonunda çocuk sahibi oldukları araştırılacaksa bu durumda takip süresi aylar ya da yıllarla ifade edilebilir. Eğer incelenen konu genç yaşlarda (<18 yaş) sigaraya başlayan bireylerde ilk kalp krizi ne zaman görülecek ise bu durumda da takip süresi yıllar, on yıllar ile ifade edilir. Çalışmanın başlangıç tarihi, takip süresi ve sonuçlandığı tarih çalışma planında açık ve net olarak belirtilmelidir (Marubini ve Valsecchi, 1994; Vittinghoff ve ark., 2004).

1.1.3. Yaşam süresi ya da başarısızlık zamanı (Failure Time)

Yaşam çözümlemesinde öncelikle her bir bireyin takip altına alındığı ve izlendiği bir zaman dilimi tanımlanmalıdır. Bu zaman dilimi; uygulanan tedaviye başlangıç zamanları, tedaviye cevap verme süreleri, tedaviden herhangi bir nedenle çıkma zamanları bireyden bireye farklılık göstereceğinden her bir birey (gözlem) için farklı olacaktır.

Başarısızlık belirli bir zaman aralığından sonra oluşur ve buna başarısızlık zamanı (failure time) denir. Bir bireyin, belirli bir başlangıç

zamanından ölümüne ya da yeniden hastalanmasına (başarısızlığına) kadar geçen zamana “yaşam süresi” ya da “başarısızlık süresi” adı verilir. Başarısızlık süresi genellikle T ile, takip altındaki n bireyli bir örneklemede ilgilenilen herhangi bir bireyin yaşam süresi ise T_i ya da t_i ile gösterilir. Başarısızlık zamanı ile ilgili tek gerek koşul negatif olmamasıdır. Tıbbi çalışmalarda başarısızlık zamanı (herhangi bir nedenden) ölüme kadar geçen zaman anlamına gelebildiği gibi tedaviden sonra bir hastalığın ilk nüksü veya bir hastalığın oluşumuna kadar geçen zaman anlamına da gelebilir (Köse, 1997). Başarısızlık süresine örnek olarak; diş eti çekilmesi tanısı sonrasında flap tedavisi uygulanan hastaların yeniden diş eti çekilmesi yaşamalarına kadar geçen süre ya da böbrek transplantasyonu yapılan hastaların operasyondan sonra ölmelerine kadar geçen süre verilebilir. Bir rüzgâr tüneline test edilen uçak parçalarının artan rüzgâr şiddeti ile kırılmalarına kadar geçen süre de başarısızlık zamanı olarak görülebilir (Marubini ve Valsecchi, 1994).

1.1.4. Sansürleme (Censoring)

Yaşam çözümlemesinde, sağkalım süresi çeşitli nedenlerle tam olarak bilinmeyen gözlemler sansürlü gözlem olarak adlandırılır. Takip altındaki bir birey ilgilenilen olgudan farklı herhangi bir nedenle öldüğünde ya da herhangi bir şekilde (tayin, taşınma, hastane ya da doktor değişikliği nedeni ile iletişim kurulamaması vb) takipten çıktığında sansürlenmiş olur. Çalışma için belirlenen takip süresi sonunda hala olgu gözlenmeyen bireyler de sansürlü olarak adlandırılır. Gerek takip edilemeyen veya başka nedenlerle ölen gerekse çalışma sonunda olgu gözlenmeyen tüm bu bireyler sansürleme zamanından daha uzun bir yaşam süresine sahip olacaklardır. Bu nedenle sansürlenmiş bu bireylerin sansürleme süreleri (C_i) kayıt altına alınmalıdır.

Sağkalım zamanı sadece takip altındaki bireyler için, takip süresini içermektedir. Bu nedenle sağkalım verileri genellikle sağdan sansürlüdür. Üç

nedenden ötürü sansürleme yapılabilir; çalışmanın bitiş süresine kadar olayın gözlenmemesi (administrative censoring), çalışma sona ermeden takipteki denekten bilgi alınamaması (lost to follow-up) ya da çalışma bitmeden başka bir olay (trafik kazası, zehirlenme vb) nedeni ile ölümün gerçekleşmesi (withdrawing) (Marubini ve Valsecchi, 1994; Kriner, 2007).

1.2. Yaşam Çözümlemesinde Kullanılan Fonksiyonlar

Yaşam çözümlemesinde, izlem altındaki örnekleme bir birimin belirlenen başlangıç zamanı ile başarısızlık (olgu gözleendiği) zamanına kadar izlendiği süreye yaşam süresi (T_i) denir ve yaşam süresi olan T_i sürekli ve negatif olmayan bir değere ($T_i \geq 0$) sahiptir. Yaşam sürelerinin analiz edilmesinde kullanılan üç temel fonksiyon vardır.

1.2.1. Olasılık yoğunluk fonksiyonu

Olasılık yoğunluk fonksiyonu, incelenen bir birimin $\Delta t \rightarrow 0$ iken ($t, t + \Delta t$) aralığında başarısız olma (ölme) hızıdır ve $f(t)$ ile gösterilir:

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{P(t \leq T < t + \Delta t)}{\Delta t} \right] \quad 0 < t < \infty$$

buradan homojen bir popülasyonda, T rastlantı değişkeninin belirli, sabit bir t 'den küçük ya da eşit olması olasılığı olan $F(t)$;

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad 0 < t < \infty$$

elde edilir. $F(t)$, sürekli deęişken olan T 'nin $0 < t < \infty$ aralığında daęılım fonksiyonudur. Olasılık yoğunluk fonksiyonu ile daęılım fonksiyonu arasındaki ilişki ise;

$$f(t) = \frac{dF(t)}{dt}$$

ile gösterilebilir (Marubini ve Valsecchi,1994; Köse, 1997; Ibrahim ve ark., 2005).

1.2.2. Yaşam fonksiyonu (Survival function)

Yaşam fonksiyonu incelenen birimin yaşam süresinin t 'den büyük olma olasılığıdır. $S(t)$ ile gösterilen yaşam fonksiyonu $F(t)$ 'nin tamamlayıcı fonksiyonu olarak;

$$S(t) = 1 - F(t) = P(T > t)$$

$$S(t) = \int_t^{\infty} f(u) du$$

şeklinde gösterilir ve negatif olmayan ($0 \leq t \leq \infty$) herhangi bir t için t zamanında yaşam olasılığını verir. Buna göre $f(t)$ ile $S(t)$ arasındaki ilişki,

$$f(t) = -\frac{dS(t)}{dt}$$

biçiminde ifade edilebilir. Yaşam fonksiyonun özel durumlarına ilişkin aşağıdaki eşitlikler kolayca bulunabilir:

- Takip süresinin başlangıcında (0. zaman) yaşam fonksiyonu sonucu tüm bireyler için 1'dir. Başka bir deyişle çalışma başlangıcında örneklemdaki tüm bireylerin yaşam olasılıkları %100'dür ($S(0) = 1$).
- Takip süresi sonsuza giderken ($t \rightarrow \infty$) yaşam fonksiyonu sonucu 0'dır. Diğer bir ifade ile izlem süresi yeteri kadar uzatılırsa (örneğin sonsuza kadar) örneklemdaki tüm birimlerin yaşam olasılıkları 0'a yaklaşır $\lim_{t \rightarrow \infty} S(t) = 0$ (Marubini ve Valsecchi,1994; Köse, 1997; Ibrahim ve ark., 2001).

1.2.3. Hazard fonksiyonu (Hazard function)

Hazard fonksiyonu, t zamanına kadar izlenen ve olgu gerçekleşmemiş (yaşayan) bir birimin belirli bir zaman içerisinde ($t + \Delta t$) olguya maruz kalması (ölmesi) riskidir. Hazard fonksiyonu $\lambda(t)$;

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) \mid T \geq t)}{\Delta t}$$

olarak gösterilir. İncelenen t zamanında yaşadığı bilinen bir bireyin ($t, t + \Delta t$) kısıtlı, küçük zaman aralığı içerisinde ölüm olasılığı $\lambda(t)dt$ ile gösterilir. Bu durumda hazard oranı (hazard rate, $\lambda(t)$, $H(t)$ ya da HARE şeklinde gösterilir);

$$\lambda(t) = f(t \mid t) = \frac{f(t)}{S(t)} \quad (1.2.3.1)$$

eşitliği ile ifade edilebilir. Bu eşitlik aynı zamanda başarısızlık oranı (failure rate), yoğunluk oranı (intensity rate) ya da ölümlülük gücü (force of mortality)

olarak da adlandırılmaktadır. Eğer eşitlikteki $f(t)$ yerine $f(t) = -\frac{dS(t)}{dt}$ ifadesi konursa;

$$\lambda(t) = f(t | t) = \frac{f(t)}{S(t)} = \frac{-\frac{dS(t)}{dt}}{S(t)} = \frac{-dS(t)}{dtS(t)}$$

$$\lambda(t) = -\frac{d}{dt} \log S(t)$$

elde edilir. Eşitliğin her iki tarafının integrali alındığında;

$$-\int_0^t \lambda(u) du = \log S(u) \Big|_0^t$$

olur. Bu eşitlikte bölüm 1.2.2.'de anlatılan yaşam fonksiyonunun özel durumu olan $S(0)=1$ (izlem zamanı başlangıcında tüm bireylerin yaşam olasılığı 1'dir) kullanarak;

$$S(t) = \exp\left(-\int_0^t \lambda(u) du\right) = \exp(-H(t)) \quad (1.2.3.2)$$

elde edilebilir (Marubini ve Valsecchi, 1994; Köse, 1997).

1.3. Cox Orantısal Hazard Modeli

Sağkalım üzerinde etkili olabilecek değişkenlerin etkilerini incelemek amacı ile ilk defa 1974 yılında Cox tarafından değinilen açıklayıcı değişkenlerin birlikte etkisi (ortak değişken) kavramı, Cox orantısal hazard (Cox proportional hazard, Cox-PH) modelinin temelini oluşturmaktadır.

Tamamlanmamış izlem verilerinde, bağımlı değişken (yaşam süresi) ile bağımsız değişkenler (tanı faktörleri, ortak değişkenler) arasındaki neden-sonuç bağıntısını ortaya koymak için Cox Regresyon yönteminden yararlanılır (Özdamar, 2003).

Cox regresyon analizinde hedef, yaşam verilerinin genel durumunu yansıtacak bir model oluşturmaktır. Cox regresyon yöntemi, bir bireyin yaşam süresi (bağımlı değişken) ile birden fazla bağımsız değişken arasındaki ilişkiyi ortaya koyan bir yöntemdir. Cox yöntemi ile yaşam süresi üzerinde etkili olduğu düşünülen bağımsız değişkenlerin etkileri eşanlı olarak belirlenip, yorumlanabilmektedir (Yay ve Ark., 2007).

Parametrik modellerin gerektirdiği varsayımların sağlanamadığı durumlarda Cox regresyon analizinin, parametrik sağkalım analizlerden daha güçlü sonuçlar verdiği ifade edilmektedir (Yay ve Ark., 2007).

İlgilenilen birimin sağkalım süresi üzerinde bir ya da daha fazla bağımsız değişkenin etkisi kestirilmek istendiğinde Cox'un klasik orantısal hazard modeli (Cox-PH) kullanılmaktadır. Sağkalım üzerinde etkisi olabilecek her bir bağımsız değişken X ile gösterilsin. Toplam p sayıda ölçülebilen bağımsız değişkenler ise $X = (X_1, \dots, X_p)$ olsun. Bu durumda bilinmeyen ve negatif olmayan temel hazard fonksiyonu $\lambda_0(t)$ olmak üzere ilgilenilen hazard fonksiyonu;

$$\lambda(t) = \lambda_0(t)e^{f(X)}$$

şeklinde gösterilir. Temel hazard fonksiyonu, verideki **ortalama** bir birimin hazard değerini gösterir. Bu nedenle ilgilenilen her birimin hazardı temel hazardın çarpanı (ya da katı) olarak ele alınır. Cox modelinin en önemli varsayımı hazard hızlarının orantısal;

$$\frac{\lambda(t | X)}{\lambda(t | X = 0)} = \frac{\lambda_0(t)e^{f(X)}}{\lambda_0(t)} = e^{f(X)}$$

olmasıdır. Bu, ilgilenilen birimin hazard hızının tüm izlem süresi boyunca (kovariyetlerin zamandan bağımsız olması şartıyla) sabit kalacağı, değişmeyeceği anlamına gelmektedir.

Temel hazard parametrik modellerle tanımlanamadığından, Cox modeli için olabilirlik (likelihood) sonuç çıkarımı uygulanamaz. Bu durumda λ_0 soldan sansürlü olmadığı sürece bilinmeyen β katsayıları kısmi olabilirlik fonksiyonunun ençoklanması (maksimize edilmesi) ile kestirilebilir. Cox 1972 yılındaki makalesinde β katsayısı için kısmi olabilirlik fonksiyonunu

$$L(\beta) = \prod_{i=1}^n \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\beta' X_i)}{\sum_{j \in R_k} Y_j(t) \exp(\beta' X_j)} \right\}^{dN_i(t)}$$

biçiminde tanımlamıştır. Eşitlikteki;

- $Y_i(t)$: İncelenen i bireyin t . Zamandaki riski (1: Risk var, 0: Risk yok)
- $dN_i(t)$: Belirli bir sınırlı zaman aralığında ($[t, t+\Delta t]$) N_i 'de gözlenen artışı
- R_k : Belirli bir zamanda risk altındaki birimleri ifade etmektedir.

Kısmi olabilirlik fonksiyonunun logaritması ise;

$$\log L(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[Y_i(t) \beta' X_i - \log \left(\sum_{j \in R_k} Y_j(t) \exp(\beta' X_j) \right) \right] dN_i(t)$$

ile gösterilebilir.

$\hat{\beta}$ 'yı tahmin etmek için $\log L(\beta)$ eşitliğinin birinci türevi alındığında elde edilen $U(\beta) = \frac{\partial}{\partial \beta} \log L(\beta)$ eşitliği en çok kısmi olabilirlik

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[X_i - \frac{\sum_{j \in R_k} Y_j X_j \exp(\beta' X_j)}{\sum_{j \in R_k} Y_j \exp(\beta' X_j)} \right] dN_i(t)$$

ile çözülebilir (Köse, 1997; Kriner, 2007)

1.3.1. Cox orantısal hazard modelinde artıklar

Cox regresyon modeli tıbbi araştırmalarda çok sık kullanılmaktadır. Ancak Cox regresyon modeli de bir regresyon modeli olmasına karşın kurulan modelin yeterliliğini ya da aşırı ve aykırı değerlerin varlığını sorgulamak için artıkların incelenmesi uygulamada pek sık görülmez. Oysa kurulan herhangi bir regresyon modelinin yeterliliğinin saptanmasına ilişkin kullanılacak ölçütlerin en önemlilerinden biri de artıkların etkili, uzak ve aykırı gözlemler için incelenmesidir (Alpar, 2003).

Model uyumunu incelemek amacı ile Martingale, Deviance, Schoenfeld ya da score artıkları incelenebilir. Bu çalışma kapsamında

Martingale ve Deviance türü artıklar kullanılacağından sadece bu tür artıklara değinilecektir.

Martingale artıklar

Martingale artıklar Cox modelindeki j . bağımsız değişken ile hazard arasındaki ilişkiyi incelemek amacı ile kullanılır. İki değişken arasındaki korelasyonu incelemek amacı ile çizilen saçılım grafiğine benzer bir mantıkla (X_{ij}, r_i) saçılım grafiğinin çizilmesi ve yerel ağırlıklı saçılım yumuşatması (Locally Weighted Scatterplot Smoothing - LOWESS) ya da benzeri bir yumuşatma (smoothing) yönteminin uygulanması mantığına dayanır. Saçılım grafiğinde kullanılan r_i terimi aynı zamanda Cox-Snell artıkları olarak da bilinir ve $r_i = H(t_i, X_i, \beta)$ ile hesaplanabilir (Taylor ve Cobb, 2004).

Martingale artıkları, incelenen zamandaki sansürlü gözlemlerin sayısı $N_i(t)$ olmak üzere;

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda(s)ds$$

şeklinde gösterilebilir (Kriner, 2007). Sağkalım verisi sağdan sansürlü ve bağımsız değişkenler zamandan bağımsız olduğunda Martingale artıkları;

$$\hat{M}_i = \delta_i - r_i$$

ile bulunabilir. Eşitlikteki r_i yukarıda açıklandığı gibi Cox-Snell artıkları, δ_i ise i . gözlem için sansür göstergesidir (Taylor ve Cobb, 2004).

Martingale artıklar $-\infty$ ile $+1$ arasında değerler alabildiğinden çarpık bir dağılım gösterirler (Karabulut ve ark., 2000).

Cox regresyon modelinde birikimli hazard (ya da Cox-Snell artıkları) HAZ ile gösterilsin. Her bir birey için HAZ hesaplandığında; Martingale artıklar olgu gözlenen bireyler için $1-HAZ$, sansürlenmiş bireyler için ise $0-HAZ$ işlemi ile kolayca hesaplanabilir (SPSS Knowledgebase). Martingale artıkların sıfırdan uzaklaşan (küçülen) negatif değerleri yüksek risk altındaki uzun sağkalım zamanına sahip sansürlü (aslında o zamana kadar olgu gözlenmesi beklenen) bireyleri göstermektedir (NCSS Help System, 2007). Martingale artıklarının $+1$ 'e yaklaşması riskin arttığını, $-\infty$ 'a yaklaşması ise riskin azaldığını göstermektedir.

Martingale artıklar; cox modelinde yer alan bağımsız değişkenlerin (kovariyet) bağımlı değişkenle olan ilişkisinin fonksiyonel şeklini (doğrusal, log-dogrusal ya da doğrusal olmayan gibi) belirlemek için de kullanılabilir. Bunun için, bağımsız değişken kullanılmayan Cox modeline (Cox null) ait Martingale artıkları elde edilir ve bu artıklara karşılık bağımsız değişkenlerin grafikleri çizdirilir (Therneau ve Grambsch, 2000).

Deviance artıklar

Martingale artıkların çarpık bir dağılım göstermesi, incelenen bireyler için elde edilen kestirimin doğruluk ve kesinliğini değerlendirmede bazı zorluklara neden olmaktadır. Martingale artıkları çarpık dağılımdan kurtarmak ve 0 etrafında dağılımlarını sağlayabilmek amacı ile karekök ve logaritmik dönüşüm yöntemleri uygulanarak Deviance artıkları elde edilmiştir. Deviance artıkları hesaplanırken,

$$d_i = \text{sign}(\hat{M}_i) \sqrt{2 \left[-\hat{M}_i - N_i(\infty) \log \left(\frac{N_i(\infty) - \hat{M}_i}{N_i(\infty)} \right) \right]}$$

eşitliği kullanılır. Eşitlikteki *sign*, işaret fonksiyonu olarak isimlendirilir ve Martingale artıklarının pozitif değerleri için +1, negatif değerleri için ise -1 sonucunu üretir. Eşitlikte uygulanan dönüşüm işlemlerinden karekök, Martingale artıklarının çok büyük negatif değerlerini küçültürken, logaritmik dönüşüm ise artıkların dağılımını yakınlştırır, bir merkez etrafında dengeli dağılımlarını sağlar (Sas/Stat User Guide).

1.3.2. Cox regresyon modelinin olumsuz yanları

Cox regresyon modelinde, bilinmeyen başlangıç hazard fonksiyonu regresyonun parametrik olmayan kısmı iken yine bilinmeyen β regresyon katsayısı ise regresyonun parametrik kısmını oluşturmaktadır. Başlangıç hazard fonksiyonu ve β parametresi kullanılan Cox regresyon modeli bu nedenle yarı-parametrik (semi-parametric) bir modeldir. Cox regresyon modelinin kullanılabilirliği veri hakkında gerçekçi olmayan varsayımları nedeni ile etkilenmektedir. Cox-PH modelinin iki temel varsayımı vardır. Birincisi hazard hızının zaman içerisinde sabit olduğu yani değişmediği, ikincisi ise hazard hızı ile bağımsız sürekli değişkenler arasındaki ilişkinin log-doğrusal olduğu. Ancak uygulamada çoğu zaman ikinci varsayım sağlanamaz. Sürekli dağılım gösteren bağımsız değişkenler ile hazard arasındaki ilişki her zaman doğrusal ya da log-doğrusal değildir (Electronic Statistics TextBook). Daha da önemlisi Cox regresyon, zaman bağımlı ortak değişkeni (time-varying covariates) modele almamaktadır ki bu pek çok durumda kabul edilemez. Örneğin; yeni bir ilaç denenen kontrollü bir çalışmada hastaların kontrol grubuna oranı <1 ise (kontrol grubunda hastalardan daha fazla denek varsa), ilacın etkisi ilacın ne kadar zamandır kullanıldığına bağımlı gibi görünecektir ve buna bağlı olarak da risk

fonksiyonu ilacın alınıp-alınmadığına ve ne kadar zamandır alındığına bağımlı olacaktır. Cox regresyon ile ilgili diğer bir sorun ise üstel bağlantı fonksiyonudur. Üstel bağlantı fonksiyonu, hazard fonksiyonunun her yerde pozitif olmasını sağlayan matematiksel uygunlukta seçilir (Terzi ve ark., 2005).

1.4. Çok Değişkenli Uyarlanabilir Regresyon Kesitleri

Çok değişkenli uyarlanabilir regresyon kesitleri, tekniği ilk olarak 1991 yılında Jerome Friedman tarafından geliştirilmiştir. Sürekli değişkenlere model uyarlayabilmek için kullanılan doğrusal regresyon, doğrusal olmayan regresyon, regresyon ağaçları, yapay sinir ağları gibi birçok model vardır. MARS adimsal regresyon yöntemidir. Regresyonun performansını geliştirmek için tekrarlamalı ayırma metodunun ve adimsal doğrusal regresyonun genelleştirilmiş bir hali olarak düşünülebilir (Friedman, 1991; Deichman ve ark., 2002; Kaki ve ark., 2004; Kolyshkina ve Sylvia, 2004)

MARS tekniği özellikle veri madenciliği alanında popüler hale gelmiştir. Çünkü bu yöntem sebep ve sonuç değişkenleri arasında bir varsayım gerektirmez ve herhangi bir matematiksel ilişki aramaz. Bunun yerine sebep ve sonuç değişkenleri arasında dinamik bir ilişki geliştirir. Diğer regresyon modellerinde belirli varsayımlar aranmasına karşın, MARS modelinde bu şekilde varsayımların olmaması önemli bir avantajdır. MARS modeli veriden elde edildiğinden, parametrik yöntemlerle tahmini güç olan doğru öngörülerini, dinamik yapısı sayesinde yorumlaması kolay, kestirim başarısı yüksek modeller kurarak bulmamızı sağlamaktadır (Sas/Stat User Guide, Deichman ve ark., 2002; Kaki ve ark., 2004; Kolyshkina ve Sylvia, 2004)

1.4.1. MARS yöntemini kullanan bazı çalışmalar

Çok değişkenli uyarlanabilir regresyon kesitleri (Multivariate Adaptive Regression Splines – MARS), Jerome Friedman tarafından 1991 yılında yayınladığı makalesi ile tanıtıldıktan sonra oldukça farklı alanlarda farklı sorunlara çözüm bulan uygulamalarda sıklıkla kullanılmaya başlanmıştır. Özelliği gereği doğrusal ve lojistik regresyonun kullanılabildiği her durumda ve bu regresyon yöntemlerinin yetersiz kaldığı ya da varsayımları sağlanamadığı için uygulanamadığı durumlarda MARS alternatifi olmayan bir yöntem olarak öne çıkmıştır.

Chen ve arkadaşları (2003), MARS yöntemini kullanarak havayolu karlılığını arttırmaya yönelik olarak çalışmışlardır. Kar yönetimi şirket kazancını arttırmayı hedefler. Rekabetçi bir havayolu endüstrisi için kar payı iyi bir kar yönetimi politikasına bağlıdır. Havayolu kar yönetimi üzerinde sıklıkla araştırmalar yapılmaktadır. Fakat bunlar halen deneysel ve belirli alanlara yöneliktir. Chen ve arkadaşları ise 20 şehir ve 31 uçuş rotasında çalışan yerel hava yolu taşımacılığı yapan şirketin kar yönetimine bir çözüm önerisi oluşturabilmek için MARS'ı stokastik dinamik programlama metodunun bir türevi olarak kullanılmışlardır.

Abraham ve Steinberg (2001); yeni ve esnek regresyon modeli olan MARS'ı, üç farklı derinlikte toprak sıcaklığı benzetimi çalışmalarında kullanmışlardır. Diğer açıklayıcı hesaplama yöntemleriyle karşılaştırıldığında, MARS hızlı, esnek ve modelin önemli girdilerini belirlemede oldukça etkilidir. Modelin girdileri; yılın günlerini, maksimum ve minimum sıcaklıkları, yağış miktarını ve potansiyel buharlaşmayı içermektedir. Çıktılar ise, 100, 500 ve 1500 mm derinlikteki toprak sıcaklığını içermektedir. Abraham ve Steinberg, MARS'ın performansını yapay sinir ağları (YSA) ile karşılaştırmıştır. Doğrusal regresyonun korelasyon katsayısı MARS ve YSA'da daima 0.95'in üzerinde olmuştur. MARS ayrıca yılın gününün, ardından da minimum hava sıcaklığının çıktı üzerindeki en önemli girdi olduğunu göstermiştir. Sonuçlar

regresyon teknolojisi olarak MARS'ın potansiyelini tarımsal uygulamalarda da göstermiştir.

Xiong ve Meullent (2004), müşterilerin peynir çubukları (Cheese sticks) tercihlerini etkileyen sebepleri araştırmışlar ve tercihleri belirleyen esas unsurları MARS ile bulmaya çalışmışlardır. MARS, tahmin ediciler arasındaki ilişkiyi ortaya koyabilmiş, parçalı regresyon fonksiyonları ile de tahmin edici değişkenlerin bağımlı değişken (peynir çubuklarını tercih) üzerine etkilerini saptayabilmiştir. Xiong ve Meullent çalışmaları ile MARS'ın karmaşık veriden, elde edilebilecek mümkün olan en iyi modeli oluşturulabildiğini göstermişlerdir.

Verzilli ve arkadaşları (2005) İngiltere'de yaptıkları çalışmada, DNA kodlamasında genetik çok şekillilik olan bölgelerde taşıyıcılarda fenotip etkiler görülebileceğini (örneğin hastalığa bağlı hassasiyet etkisi gibi) belirtmişler ve analiz için bağ tutucu ve lizozom proteinlerindeki başlangıç mutasyonu verileri kullanılarak bir model oluşturmuşlardır. Oluşturdukları modelin performansını sınamak için hiyerarşik Bayesci MARS Yöntemini kullanmışlardır. Deneilerinde 12 Aminoasit durumu hakkında bilgi edinebilmek için her bir aminoasidin pozisyonu ve protein işlevselliği etkilerini ölçmüşler ve her bir pozisyondan elde edilen değerleri modelde yerine koyarak kestirimlerde bulunmuşlardır. Öğrenme verisi olarak bağ tutucu verilerini kullanmış ve lizozom değişimleri ile bu değişkenlerin tersleri üzerinde modeli test etmişlerdir.

Leathwick ve arkadaşları (2006), Yeni Zelanda akarsularında yaşayan balıkların dağılıma ilişkin bir veritabanını kullanarak 15 tatlı su balığı türünün bulunma olasılıkları ile bu canlıların akarsulardaki yaşam alanlarını belirleyen çevresel değişkenler arasındaki ilişkileri MARS ile incelemişlerdir. Her tür için yaşam koşullarını en iyi belirleyen çevresel faktörler MARS yöntemi ile saptanmış, her tür için ideal yaşam alanını belirleyen faktörler ortaya

çıkarılmış, elde edilen sonuçlar coğrafi bilgi sistemine işlenerek bu türlerin olası yaşam alanlarının haritaları oluşturulmuştur.

1.4.2. MARS modeli

Regresyon yöntemleri; bağımsız değişkenlerle bağımlı değişkenler arasındaki bağıntıyı incelemek için kullanılır. Doğrusal olmayan regresyon, bağımlı değişkenle bağımsız değişkenler arasındaki bağıntıyı veriden model türeterek kurar. MARS; regresyon yapısını, veri uzayını küçük küçük doğrusal regresyon denklemlerine (temel fonksiyonlar) bölerek oluşturur. Her bölme noktası veriye bağlı olarak bulunur. MARS modeli, orijinal veriden tahmin edilen temel fonksiyonların özel sınıflarının regresyon kesitleridir. MARS'ın üstünlüğü en uygun değişken dönüşümleri ve etkileşimlerine ek olarak çoğunlukla yüksek-boyutta saklanan karmaşık veri yapılarını bulmasından kaynaklanmaktadır (Friedman, 1991; Deichman ve ark., 2002; Kaki ve ark., 2004; Kolyshkina ve Sylvia, 2004)

Kesit (spline), karmaşık eğri çizimlerinde ve fonksiyon tahminlerinde matematiksel bir süreç olarak göz önüne alınabilir. Kesit düzleştirme yöntemi iki ya da daha üst düzeyli polinomieller kullanıldığında elde edilen ve parametrik olmayan hata varyansının kontrol edilmesini sağlayan bir yöntemdir (Kaki ve ark., 2004).

Kesit oluşturmak için X ekseninde uygun sayıda bölge oluşturulur. Bu bölgeler arasındaki sınırlar düğüm olarak bilinir. Yeterli sayıda düğüm ile model tahmin edilebilir. İki boyutlu kesit çizmek oldukça kolayken ikiden çok boyutlu kesitlerin çizimi zorlaştığından her bir kesit matematiksel temel fonksiyonlarla (Basis Functions) ifade edilmektedir (MARS User Guide).

MARS; oldukça sadeleştirilmiş bir karar ağacı oluşturur. Parçalı temel fonksiyonlar ve bunların birleşimlerini kullanarak ve regresyon modellerindeki hem ileri doğru hem de geriye doğru ilerleme algoritmalarından yararlanılarak MARS modeli elde edilir. Model kurulumu iki adımda gerçekleşir.

1. Adım: Mümkün olan tüm temel fonksiyonlar oluşturulur. Çok büyük bir model bulununcaya kadar yani modelin karmaşıklığı maksimum seviyeye ulaşıncaya kadar eklenen temel fonksiyonlarla model geliştirilir.

Temel fonksiyonlar; ya dönüştürülmüş tek değişkenden ya da çok değişkenli karşılıklı etkileşim terimlerinden meydana gelir. Temel fonksiyonlar modeldeki bağımsız değişkenlerin karşılıklı etkileşimlerini ve doğrusal olmayan dönüşümlerini hesaba katan fonksiyonlardır. Temel fonksiyonlar oluşturulurken tüm değişkenler ve değişkenlerin birleşimleri tek tek ele alınır. Bu durumda, her bir bağımsız değişken ve bu değişkenlerin birbirleri ile olan etkileşimlerinin tanımlı bulunduğu aralıklardaki tüm olası değerler birer düğüm olarak düşünülüp, mümkün olan tüm olası temel fonksiyonlar belirlenir. Temel fonksiyonlar eklenerek model daha karmaşık ve daha esnek bir yapıya kavuşturulur. Temel fonksiyon sayısı maksimum seviyeye ulaşıncaya kadar bu işlem sürer.

Temel fonksiyonlar oluşturulurken aynı değişkene ait *Ayna Temel Fonksiyonu* da oluşturulup modele alınabilir. Ayna temel fonksiyonu, bağımlı değişkenle bağımsız değişken arasındaki dağılımın düğüm noktasında eğiminin değiştiğini ve düğüm noktasına kadar olan eğiminin sıfır olduğunu gösterir.

2. Adım: Eğilim ve varyansların optimal bir dengesi bulununcaya kadar temel fonksiyonlar modelden çıkarılır. Bu adıma budama adımı denir ve bu aşamayı gerçekleştirmek için MARS geriye doğru adım algoritması kullanılır.

Mars algoritması, sadece sonuç üzerinde ölçülebilir bir etki yapan temel fonksiyonları seçer, diğerlerini dikkate almaz. Geriye dönük bir prosedür uygulanarak sonuca daha az etki eden temel fonksiyonlar modelden çıkarılır. Sabit temel fonksiyona dayalı en basit model oluşturulur. Temel fonksiyonları her değişken ve her olası düğüm için araştırarak tahmindeki hatayı azaltanları final modeline ekler. En küçük kareler mantığı ile sonuca en az etki eden temel fonksiyonları modelden çıkarır (Temel ve ark., 2005).

MARS, özellikle ayna değişkenli fonksiyonları doğrusal ve doğrusal olmayan dağılımların temel fonksiyonları olarak kullanır ve etki-tepki değişkenleri arasındaki ilişkiyi tahmin eder.

Kategorik Tahmin Ediciler: Pratikte hem sürekli hem de kategorik tahmin ediciler kullanılabilir. İkisi de başarılı sonuçlar verir. Ancak temel MARS algoritması değişkenlerin sürekli olduğunu öngörür, sadece iki değer alabilen (dikotom) değişkenlere göre sıralı (ordinal) değişkenlerde daha başarılı sonuçlar elde eder.

Çoklu Bağlantı (Multicollinearity): MARS algoritması aralarında çoklu bağlantı olan değişkenlere de uygulanabilir. Burada algoritma verilere bağlı olarak temel fonksiyonlar bulur. Fakat sinir ağları yapısında olduğu gibi, her bağımlı değişken için farklı bir sabit oluşturur.

Sınıflandırma Problemleri: Bilindiği gibi sınıflandırma problemlerinde çoklu bağlantı sorunu sınıflandırma performansını düşürmektedir. MARS, çoklu bağlantı sorunlarını çözebildiği için sınıflandırma problemlerine de uygulanabilir. Sınıflandırılmış değişkenleri kodlar, algoritmayı işletir, model uygun ise o kodlarla skorları belirler. Hangi sınıf için en iyi skor elde edilirse kurulan modelde o sınıf yer alır.

Model Seçimi ve Budama: Doğrusal olmayan modeller, esnek bir yapıya sahiptirler. Bu nedenle önlem alınmadığı durumlarda gereğinden fazla yaklaşma çabası içine girilebilirler. Denenmiş verilerde neredeyse sıfır hata yapmalarına rağmen yeni gözlemler ve durumlarla karşılaştıklarında zayıf bir performans gösterebilirler. Diğer pek çok metotta olduğu gibi MARS da veriye aşırı uyum (karmaşıklık artırma, over-fit) yapma eğilimindedir. Bunu engellemek için budama tekniği kullanılır ve modelin karmaşıklığı azaltılır (sınırlanır).

MARS, pratikte yaygın kullanılan doğrusal ve doğrusal olmayan regresyon modellerinin aksine bağımsız değişkenin farklı aralık değerleri için farklı katsayılar türetir ve modele etkileşim terimlerini de katarak gerçek yapıyı daha iyi yansıtan modeller geliştirir. Ayrıca bağımlı ve bağımsız değişkenlerin dağılımları üzerine (diğer regresyon yöntemlerinden farklı olarak) herhangi bir varsayım gerektirmez ve değişken tipi de önemli değildir (Statistica User Guide).

Gerçek veriye bağlı olarak oluşturulan temel fonksiyonlar, model parametreleri ile birlikte (en küçük kareler yöntemi) modeli tahmin eder. Genel MARS modeli eşitlikteki gibi tanımlanabilir.

$$Y = \beta_0 + \sum_{k=1}^K a_k B_k(X_t) + \varepsilon \quad (1.4.2)$$

Eşitlikteki;

- k : *Düğüm sayısını*
- K : *Temel fonksiyon sayısını*
- X : *Bağımsız değişkeni*
- a_k : *k. temel fonksiyonun katsayısını*

β_0 : Modeldeki sabit terimi

$B_k(X_t)$: t . bağımsız değişken için k . temel fonksiyonu

ifade eder.

MARS modeli; modeldeki temel fonksiyon sayısı M , model katsayısı β ve temel fonksiyon $B(X)$ olmak üzere aşağıdaki gibi de tanımlanabilmektedir:

$$f(X) = \sum_{i=1}^M \beta_i B_i(X)$$

Temel fonksiyon $B(X)$ in olası modları ise;

- $B(X) = 1$
- $B(X) = x_i$
- $B(X) = (x_i - t_k)_+$
- $B(X) = x_i x_j$
- $B(X) = (x_i - t_k)_+ x_j$
- $B(X) = (x_i - t_k)_+ (x_j - t_l)_+$

olarak ifade edilebilir (Kriner, 2007).

1.4.3. Temel fonksiyon (Basis function)

Genellikle regresyon denklemleri, veriler arasındaki ilişkiyi tek bir fonksiyon kullanarak belirlemeye çalışırlar. Fakat MARS parçalı polinomik bir fonksiyon kullanır. Böylece bütün değerlere en yakın noktalardan geçebilecek (bu sayede artıkları da en aza indirebilecek) regresyon kesitleri oluşturulabilir. Regresyon kesit fonksiyonları parçalı polinomik temel fonksiyonların düğümlerde birleştirilmesi ile elde edilmiş sürekli bir fonksiyondur. Temel fonksiyonlardaki sabitler en küçük kareler yöntemi ile bulunur. Parçalı

fonksiyonlar gerçek verinin dağılımına daha uygun olduğu için tek polinom kullanmaya göre daha fazla esneklik sağlar. MARS her duruma yüksek derecede uyum sağlayabilir ve düğüm yerlerini ve derecelerini otomatik olarak seçen bir algoritma oluşturur.

Eşitlik 1.4.2'de geçen $B_k(x_t)$ ifadesi t . bağımsız değişken için k . temel fonksiyondur ve aşağıdaki gibi tanımlanır.

$$B_m(X) = \prod_{t=1}^{L_m} [S_{l,m}(x_{v(l,m)} - k_{l,m})] + \quad k= 1,2, \dots, K \quad (1.4.3)$$

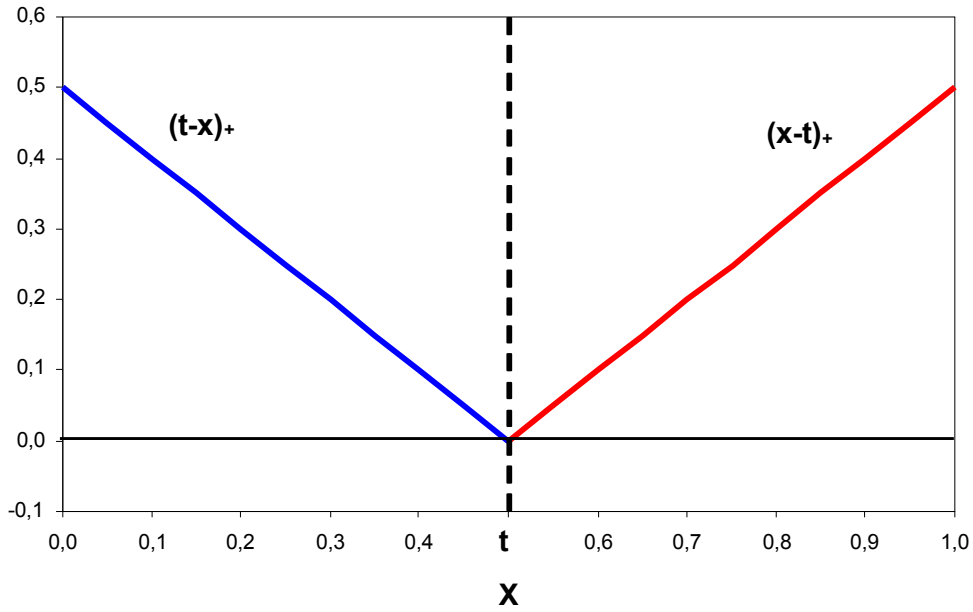
Eşitlikte;

- L_m : Etkileşim derecesini
- $S_{l,m}$: -1 ya da +1'i
- $x_{v(l,m)}$: Bağımsız değişken değerini
- $k_{l,m}$: Düğüm değerini göstermektedir.

Etkileşim derecesi (L_m), bağımsız değişkenler arasındaki etkileşimi ifade etmektedir. Çok değişkenli regresyon analizlerinde olduğu gibi ikili etkileşimden daha yüksek boyutta (3'lü, 4'lü gibi) etkileşim seçmek sonuçların yorumlanması aşamasında sorun yaratabileceğinden, ikiden daha fazla bağımsız değişkenin etkileşimini modele koymak pratikte çok sık kullanılmaz.

Temel fonksiyonlar, x_t 'nin doğrusal olmayan karmaşık dönüşümleri olabilir. Fakat Y_t temel fonksiyonların doğrusal bir fonksiyonudur.

Özellikle doğrusal ve doğrusal olmayan genişleme için temel fonksiyonlar iki yönlü kullanılır. İki temel fonksiyon ($t-x$) ve ($x-t$), t parametresi temel fonksiyonların düğüm (knot) noktasıdır. Düğüm noktası veriden elde edilir. Sadece pozitif sonuçları gösterir. Diğer durumlar için 0 (sıfır) değeri alır (Şekil 1.4.3).



Şekil 1.4.3. Temel Fonksiyonlara Karşılık Ayna Değişken
(MARS User Guide'dan alınmıştır)

1.4.4. Adımsal ilerleme (Modeli geliştirmek)

Bağımlı değişken üzerinde etkisi olan (ya da olabileceği tahmin edilen) tüm öğeler (bağımsız değişkenler ve bunların etkileşimleri) tek tek tespit edilerek bağıntıları belirlenir. En güçlü bağıntıyı gösteren öğe ile işe başlanır. Bu bağıntının açıklayıcılık (belirleyicilik) düzeyi (R^2) hesaplanır. Sonraki aşamada, ikinci derecede güçlü bağıntıya sahip bağımsız değişken de modele katılır. Oluşan yeni denklemde elde edilen belirleyicilik, daha öncekine göre anlamlı olarak yüksek bulunursa ikinci bağımsız değişken de modele eklenir. Eğer ikinci bağımsız değişkenin olduğu modelde R^2 'de anlamlı bir yükseliş olmaz ise ikinci bağımsız değişken modele alınmaz. Bu şekilde devam edilerek öncelikle tüm bağımsız değişkenler ve sonrasında değişkenler arası etkileşimler modele eklenerek en karmaşık model kurulur.

1.4.5. Adımsal gerileme (Modeli budamak)

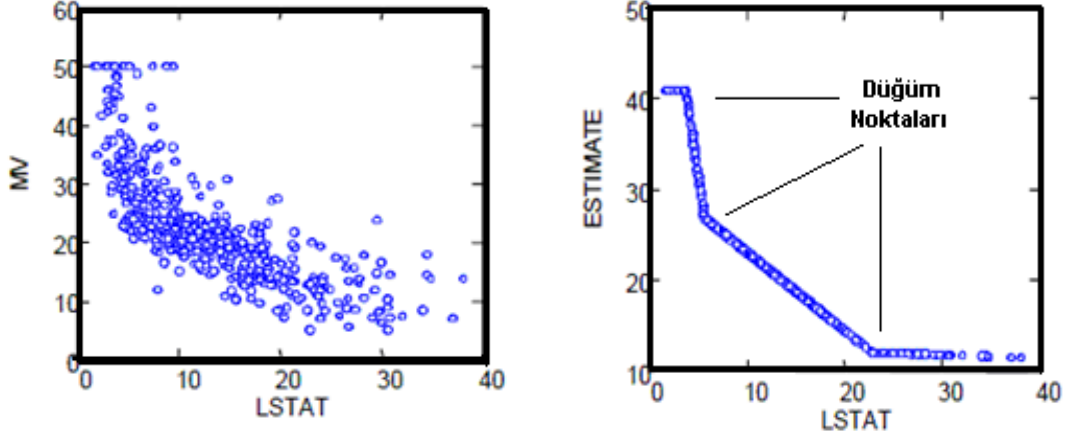
Adımsal ilerleme ile elde edilen en karmaşık ve anlamlı regresyon modeli ile işleme başlanır. Bu modelin belirleyiciliği (R^2 değeri), en zayıf bağıntıya sahip değişken ya da etkileşim modelden çıkarıldığı zaman artıyor ise modelden çıkarmalara devam edilir. Bu tür ekleme ve çıkarmalar ile en az sayıda değişken kullanılarak en yüksek belirleyicilik araştırılır. Bir değişkenin eklenip çıkarılması için karar vermede belirleyiciliğin farklılaşması dışında ölçütler de kullanılabilir. Örneğin, araştırmacı bir değişkenin mutlaka modelde yer almasını isteyebilir. Ayrıca iki değişkenin etkileşimi modelde yer alıyorsa tıpkı diğer regresyon modellerinde olduğu gibi etkileşim içerisinde yer alan bağımsız değişkenlerin de modelde yer alması gerekir.

1.4.6. KNOT (Düğüm noktası)

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. Düğüm değerinde hata kareler toplamı en küçük değerini alır. Ardışık iki düğüm değerini birleştirerek çizilen doğrunun eğimi β değerini gösteren regresyon katsayısıdır.

Kesitlerin altında yatan temel düşünce düğümdür (knot). Düğüm; verinin bir bölgesinin sonu ile diğerinin başlangıcını belirler. Bu nedenle düğüm, fonksiyonun değiştiği yerdir. İki düğüm noktası arasında veri herhangi bir regresyon modeline (örneğin doğrusal regresyon) uygun olabilir. Klasik kesitte, düğümler önceden belirlenmiş ve eşit aralıklı olarak ayrılmıştır, oysa MARS modelinde düğümler prosedür tarafından araştırılarak bulunur. MARS modeli ihtiyaç duyulandan daha çok sayıda düğüm noktasını bulur. Eğer iki düğüm noktası arasında düz bir çizgi yakalanırsa, bu aralıkta yeni düğüm noktasına ihtiyaç yoktur. Tahmin değeri önemsiz olan gözlemlerin oluşturduğu sahte düğüm noktaları MARS tarafından en aza indirgenir. Şekil

1.4.6'da karmaşık gerçek bir veri ve bu veriden oluşturulan üç düğüm noktası olan bir MARS kesiti gösterilmiştir (MARS User Guide).



Şekil 1.4.6. Kesit ve düğüm kullanılarak MARS veri tahmini (solda gerçek veri) (MARS User Guide'dan alınmıştır)

MARS modeli oluşturulurken, bağımlı (sonuç) değişken ve modele girmesi gereken bağımsız değişkenler belirlenmelidir. MARS model verisindeki düğüm noktaları (knot), eşit aralıklı temel birçok kesitten oluşur. Her bir kesit, farklı değişkenler arasında ya da aynı değişkenin belirli aralıklarında oluşur. MARS, modele daha birçok alt gruplar ilave eder ve düğüm noktaları oluşturur.

MARS modeli, temel fonsiyon olarak adlandırılan düzgünleştirilmiş fonsiyonlardan oluşur. İki temel fonsiyonun kesişimi ile düğüm noktası elde edilir. Temel fonsiyonlar; iki düğüm noktası arasında verinin karakteristiğine göre doğrusal ya da doğrusal olmayan regresyonlar olabilir. Temel fonsiyon herhangi iki düğüm noktası arasında tektir ve her bir düğümde başka bir temel fonsiyon oluşur. MARS, verinin oluşturduğu doğrusal olmayan regresyon çizgisini, oluşturulan alt gruplarla kesitlere böler. Uygun regresyon modelindeki, komşu iki düğüm noktası arasındaki en kısa uzaklığı önceden belirleyerek alt gruplardaki küçük sayıdaki verilerden

kaynaklanabilecek (verinin özelliğine göre çok fazla sayıya ulaşabilecek) gereksiz düğüm oluşturma döngüsünü önler.

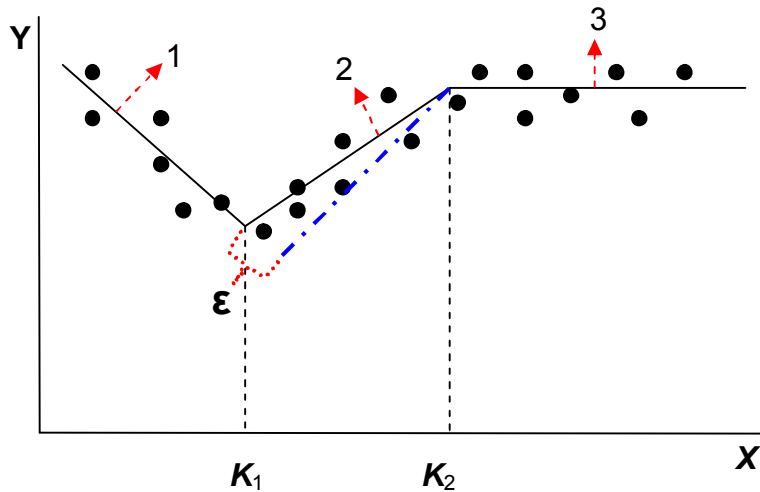
İlk önce birçok düğüm noktasından oluşan bir model (bu aralıktaki veri düz bir çizgi kabul edilir) oluşturulur, sonra modele en az katkısı olan düğümler modelden çıkarılır. Modelin kurulması aşamasında ileriye doğru adımsal yöntem birçok gereksiz (modele az katkısı olan ya da katkı sağlamayan) düğüm oluşturacak fakat gereksiz düğümler geriye doğru adımsal budama yöntemi ile modelden çıkarılacaktır (Abraham ve Steinberg, 2001).

Basit doğrusal regresyonda bir tane en iyi düğüm noktasını bulmak basit bir problemdir. Çok sayıda potansiyel düğümler incelenerek en iyi R^2 ile biri seçilebilir. Ancak düğümlerin en iyi çiftini bulmak birçok uzun hesaplamalar gerektirir. İhtiyacımız olan doğru sayıda düğümlerin en iyi setini bulmak zor bir iştir. MARS ise gereken sayıda ve konumda düğüm noktasını ileriye ve geriye adımsal yöntemle bulur. Önce birçok düğüm noktası ile modeli en karmaşık yapıya kavuşturur. Sonra bu düğüm noktalarından katkısı en az olanları modelden çıkarır. Böylece ileriye doğru düğüm seçimi yanlış yerlerdeki birçok düğüm içerecek, fakat yanlış düğüm noktaları modelden geriye doğru budama yöntemi ile silinecektir (Mukkamala ve ark., 2004).

1.4.7. Düğüm değerinin elde edilmesi

Bağımlı ve bağımsız değişken arasındaki ilişkilerin doğrusal, eğrisel ve kübik biçimde olabileceği daha önce ifade edilmişti. Aynı bağımsız değişken üzerinde, ilişkinin şeklinin değiştiği bağımsız değişken değerine düğüm değeri denir. Bir başka ifadeyle, bağımsız değişken değerinin, tanımlı bulunduğu aralıklarda doğrunun eğimini değiştirmeyen en son değeri düğüm değeri olarak alınır. Bu düğüm değerlerinde hata kareler toplamı en küçüktür.

Ardışık iki düğüm değerini birleştirerek çizilen doğrunun eğimi, Model'de β 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 monoton dönüşümler (logaritmik, üstel, vb.) yapmadan aralarındaki ilişkiyi doğrusal hale getirir. Şekil 1.4.7.1'de iki farklı düğüm değerine sahip bir değişkenin bağımlı değişkenle olan ilişkisi görülmektedir. K_1 ve K_2 düğüm değerleridir. Başlangıç noktasından K_1 'e kadar doğru parçasının eğimi sabittir. Benzer şekilde K_1 ve K_2 arasında da doğru parçasının eğimi sabittir. Fakat K_1 'den ε kadar uzaklaşsak bile doğru parçasının eğimi değişir. İşte doğru parçasının eğimini değiştirmeyen değer bir düğüm değeridir.



Şekil 1.4.7.1. X ve Y dağılım grafiği ve düğümler (K_1 , K_2)
(Briand L.C. ve arkadaşları (2004)'den alınmıştır)

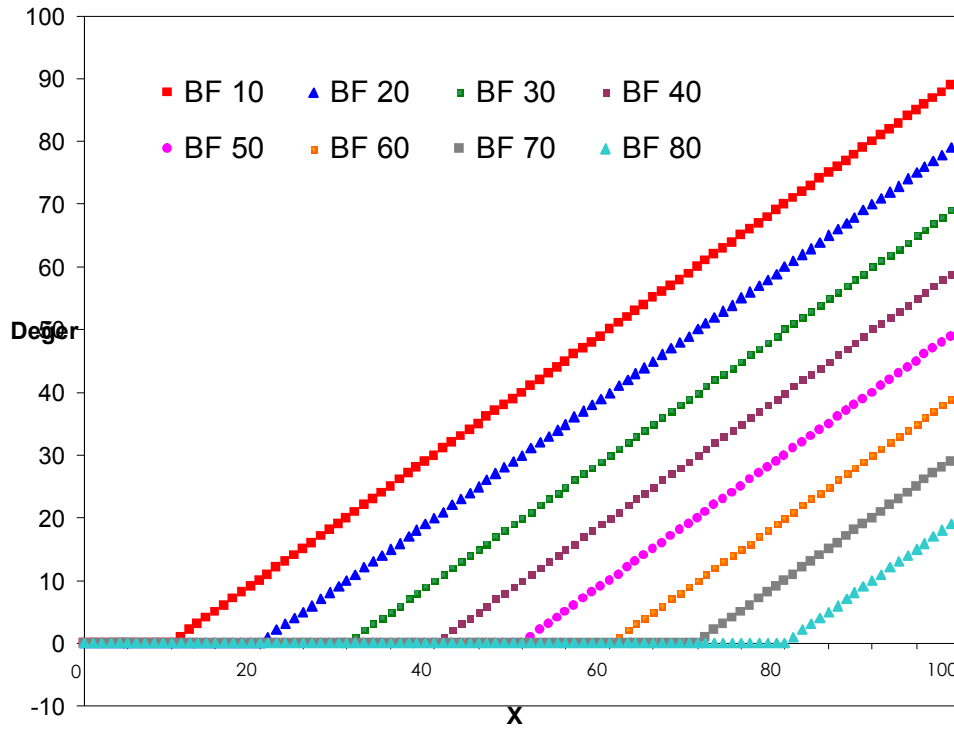
Düğüm seçiminin şartlarını incelersek tek boyutta kesitleri çok iyi bir şekilde anlayabiliriz. Fakat bu şartlar aynı anda çok sayıda değişkenle çalışıldığında kullanışsızdır. Programlama yapabilmek için kısa ve kolay anlatımlar gerekir. Düğüm yerlerini kullanırken etkileşimlerin nasıl yapılacağı ya da ifade edileceği açık değildir. MARS'ta Temel Fonksiyonlar (BF) düğümlerin araştırılmasını genelleştirmek için kullanılan mekanizmalardır. BF bilgi içeren bir ya da daha çok değişkeni ifade etmekte kullanılan

fonksiyonların setidir. Temel bileşenler analizine benzer şekilde, BF'de hedef değişkenle tahmin edici değişkenlerin ilişkisi aslında daha açıktır. Şekil 1.4.7.2'de gösterilen BF hokey sopası yapısı, MARS modelinin yapıtaşısıdır ve tek değişkenli modellerde çoğu zaman uygulanır. Hokey sopası fonksiyonu değişken gösteriminde X 'ten, X^* a dönüşür.

$$\begin{aligned} X^* &= \max(0, X-c) && \text{eğer } X > c \text{ ise} \\ X^* &= \max(0, c-X) && \text{eğer } X < c \text{ ise} \end{aligned}$$

Burada X , eşik değeri c 'nin altında bir değerse $X^*=0$ değerini, c 'den büyük bütün değerler için ise $X^*=(X-c)$ değerini alır ($\max(0, X-c)$). İkinci form birincisinin ayna değişkenini oluşturur ($\max(0, c-X)$). Şekil 1.4.7.2'de, X kestiriminin 0–100 aralığı için c değişkeninin 10'ar birimlik değişimlerine karşılık temel fonksiyonlarda meydana gelen değişimler gösterilmiştir.

Şekil 1.4.7.2. incelendiğinde, BF 10 ile gösterilen temel fonksiyonun $X=10$ değerinden itibaren sürekli yükseldiği, ancak BF 80 ile gösterilen temel fonksiyonun ise nerede ise tüm grafik boyunca 0 değerini aldığı görülmektedir. Buradan BF 10 ile gösterilen temel fonksiyonunun $X=10$ değerine kadar c 'den küçük (dolayısı ile $X^*=0$) kaldığı, $X=10$ değerinden itibaren ise c 'den daha büyük değerler (dolayısı ile $X^*=X-c$) aldığı görülebilir. Benzer şekilde BF 80 ile gösterilen temel fonksiyonun $X=80$ değerine kadar c 'den küçük ($X^*=0$) olduğu ancak $X=80$ değerinden itibaren c 'den büyük (yani $X^*=X-c$) olduğu görülmektedir.



Şekil 1.4.7.2. X değişkeni 0-100 aralığı için $c=10$ 'dan 80'e kadar temel fonksiyonların değişimi (hokey sopası)
(MARS User Guide'dan alınmıştır)

1.4.8. Ayna-görüntü temel fonksiyonlar

MARS ikili parçalarından oluşan temel fonksiyonlar yaratır. Temel fonksiyonların birinci parçası standart temel fonksiyonlarken ikinci parça olarak standart temel fonksiyonun ayna görüntüleri daima temel fonksiyona eklenir. Bu ikililer, olası birçok temel fonksiyonlar gibi ayrı veri değerleri oluştururlar. Ayna görüntüsü sağ ve sol ana kırıklara ayrılmış düğümlerdir.

Bu fonksiyonlar doğrusal olarak bağımsız değildir, fakat modelin esnekliği artırılabilir. Düğümlerin verilmiş seti için, temel fonksiyonların sadece bir ayna görüntüsü standart temel fonksiyonlardan doğrusal olarak bağımsız olacaktır. Modelde kullanılan temel fonksiyonların tümü için, aynı değeri vereceğinden temel fonksiyona hangi ayna değişkenin ekleneceği

önemli değildir. Bununla birlikte herhangi bir düğümde standart temel fonksiyonların ayna görüntüleri yerine model değiştirilebilir. Temel fonksiyonların eklenen belirli ayna görüntülerini değiştirmek sonuç modelini etkilemez.

1.4.9. Genelleştirilmiş çapraz geçerlilik

Oluşturulan en karmaşık model budanarak yani önemli bağımsız değişkenler ve bu değişkenlerin karşılıklı etkileşimleri belirlenerek, hata kareler toplamı minimum olan en uygun model oluşturulur. Budama algoritması en yaygın olarak genelleştirilmiş çapraz geçerlilik (*Generalized Cross Validation, GCV*) tekniği ile yapılır. Genelleştirilmiş çapraz geçerlilik hem hata terimini hem de model karmaşıklığını hesaba katar ve eşitlik 1.4.9.1, 1.4.9.2. yardımıyla hesaplanır.

$$GCV(M) = \frac{1}{n} \frac{\sum_{m=1}^n (y_i - \hat{f}_M(\mathbf{x}_i))^2}{(1 - C(M)/n)^2} \quad (1.4.9.1)$$

$$C = 1 + cd \quad (1.4.9.2)$$

Eşitlikte,

n : Veri setindeki denek sayısını,

d : Etkili serbestlik derecesi olup bağımsız temel fonksiyonların

sayısını,

C : Eklenen temel fonksiyonların maliyet-karmaşıklık (*cost-complexity*) ölçüsünü ve

M : MARS Modelinin kurduğu regresyon modeli sayısını

göstermektedir.

Literatürde maliyet karmaşıklık ölçüsü C 'nin hesaplanmasında kullanılan d için $2 < d < 5$ aralığının, GCV için ise 2 - 5 aralığının en ideal değer aralığı olduğu ifade edilmektedir (Abraham ve Steinberg, 2001; Deichman ve ark., 2002; Leathwick ve ark., 2006; Statistica User Guide).

1.4.10. Yaşam çözümlemesinde MARS kullanım örnekleri

MARS regresyon yönteminin uygulama alanları artıp, diğer regresyon yöntemlerinin uygulanabildiği hemen hemen her durumda uygulanabilirliği gösterildikçe pek çok araştırmacı tarafından özellikle spline (kesit) mantığı ve smooting (yumuşatma, düzleştirme) yaklaşımını içeren bu yöntemin sağkalım analizinde nasıl kullanılabileceği ve getirebileceği faydalar araştırılmaya başlanmıştır.

Kooperberg ve arkadaşları (1995) yayınladıkları makale ile HARE modelinde zamana göre değişen, zaman bağımlı (time-varying) ortak değişkenlerin modele dahil edilmesini tartışmışlardır. Bu sayede Friedman tarafından geliştirilen MARS modelinde kullanılan temel fonksiyonları şartlı log-hazard fonksiyonları olarak tanımlamışlar ve veri tarafından belirlenen non-parametrik HEFT (Hazard Estimation with Flexible Tails, esnek kuyruklar ile hazard kestirimi) fonksiyonunu temel fonksiyonlardan elde etmişlerdir.

Keleş ve Segal (2002), Batı Avusturalya HIV kohort çalışması verilerini kullanarak AIDS olma zamanını belirlemeye çalışmışlardır. Çalışmada, sağdan sansürlü verilerde sansür zamanını bilinen sağkalım analizi yöntemleri ile belirlemenin zorluğu vurgulanmış ve Martingale artıklarının bağımlı, sonuç değişkeni olarak kullanılması ile bu zorluğun üstesinden gelinebileceği tartışılmıştır. Araştırmacılar, MARS tarafından getirilen eklemeli (additive) model ile analitik karşılaştırmaların doğrudan ya da artık tabanlı olarak daha kaliteli şekilde yapılabileceğini belirtmişlerdir. Öncelikle ağaç yapılandırmalı regresyon (tree-structured regression) yaklaşımını

kullanarak logrank ve Martingale artıklarının kareler toplamını bölme fonksiyonunun bileşke ağaçlara (resultant trees) oldukça benzer sonuçlar verdiğini yaptıkları benzetim çalışmaları ile göstermişler, devamında elde ettikleri modeli gerçek veri setine uygulayarak sınımışlardır.

Klinisyenlerin ve sađlık hizmeti arařtırmacıların, olumsuz olayların (ölüm, hastalık nüksü, cerrahi uygulama sonrası gelişen komplikasyonlar, hastaneye yeniden kabul gibi) hastaya özgü olasılıklarının tahmini ile günümüzde daha fazla ilgilendiğini ifade eden Austin (2007), klinik çalışmalarda sonuçları öngörmek için sınıflandırma ve regresyon ağaçları (CART) kullanımına giderek artan bir ilgi olduğunu belirtmiştir. Çalışmada, akut miyokard enfarktüsü (kalp krizi, AMI) ile hastaneye başvuran kişilerin tedavi sonrasındaki ölüm hızlarının belirlenmesinde regresyon ağaçları ve lojistik regresyon kestirimleri ile genelleştirilmiş eklemeli modeller (GAMS) ve çok deđişkenli uyarlanabilir regresyon kesitleri (MARS) karşılaştırılmıştır. Bunun için Ontario'da (Kanada) AMI ile hastaneye başvuran 9484 hastanın verilerinden 1000 kez rastgele örnek seçilmiş ve oluşturulan modellerle 1000 kez veri türetilmiştir. Seçilen örneklem ve türetilen veriler ROC eğrileri de kullanılarak, kestirim deđerleri açısından karşılaştırılmıştır. Lojistik regresyon sonuçları ile GAMS ve MARS sonuçlarını karşılaştırmış çalışmasında kullandığı veriler için lojistik regresyonun GAMS ve MARS'tan daha kötü sonuçlar vermediğini öne sürmüştür.

York ve arkadaşları (2006), ilaç tedavilerinde cinsiyetin etkisi ya da genotip ile çevre etkileşimlerini alt grupları da dikkate alarak en iyi şekilde açıklayabilecek regresyon eğrilerini bulmaya çalışmışlardır. Buradaki asıl sorun regresyon eğrisi şekli hakkında detaylı bilgi olmaması ve şeklin alt gruplarda gösterdiği farklılığın belirlenememesidir. Bu sorunu aşabilmek için açıklayıcı tekniklere ihtiyaç vardır. Arařtırmacılar çalışmalarında açıklayıcı iki farklı tekniđin; MARS ile polinomial kullanan en küçük kareler eğri uydurma yöntemlerinin gücünü karşılaştırmışlardır. Yaptıkları simülasyon çalışmalarında doğrusal, lojistik ve karmaşık doğrusal olmayan eğrileri kullanmışlardır.

Çalışma sonucunda MARS'tan polinomiyal kullanan en küçük kareler eğri uydurma yönteminden (farklı model eğrileri için) ortalama 1.4 kat daha fazla güç elde edildiği belirlenmiştir. Araştırmacılar, eğri doğrusal olduğunda gücün en yüksek olduğu, eğrinin karmaşıklaşması ile kazançların da yükseldiği bu nedenle doğrusal olmayan eğrilerde (non-linear curves), MARS'ın model bağımsız bir yöntem olarak belki de tek alternatif olduğunu öne sürmüşlerdir.

Mallick ve arkadaşları (1999) 38 böbrek enfeksiyonlu hastanın enfeksiyon zamanlarını içeren bir veri seti üzerinde çalışmışlardır. Hastaların; iki enfeksiyon zamanı vardır ve bazı veriler sağdan sansürlüdür. Diğer ortak değişken olarak; yaş (yıl olarak), cinsiyet ve eşlik eden diğer hastalıklar (GN, AN, PKD) olarak çalışmaya alınmıştır. Aynı veri seti üzerinde bilinen regresyon yöntemleri ile yapılan önceki çalışmalara göre kestirimde kullanılabilecek tek anlamlı değişken cinsiyettir. Araştırmacılar oluşturdukları MARS modeli ile önceki çalışmalardan farklı olarak regresyon modelindeki tek anlamlı değişkenin cinsiyet olmadığını ortaya koymuşlardır. Mallick ve arkadaşlarının oluşturduğu Mars modelinde de cinsiyet en önemli kestirim parametresidir ancak enfeksiyon zamanı * cinsiyet etkileşimi (dolayısı ile enfeksiyon zamanı bağımsız değişkeni) de modelde yer alması gereken anlamlı etkiye sahip terim olarak ortaya çıkmıştır. MARS, bilinen regresyon yöntemlerinin ortaya koyamadığı etkileşimleri tespit etmede başarılı sonuçlar vermiştir.

Kriner (2007), Alman Kalp Merkezi verilerinden kalp krizi (Myocardial infarction - MI) geçirmiş 2376 hastanın bilgilerini kullanarak kalp krizine etki eden faktörleri belirlemek üzere MARS yaklaşımını kullanmıştır. Çalışmada; yavaşlama (DC) ve hızlanma kapasitesi (AC), kalp sol karıncığı boşaltma oranı (LVEF), patolojik kalp atım hızı sorunu (HRT), diyabet varlığı, normal-normal aralığındaki kalp atım hızı değişimlerinin standart sapması (SDNN), yaş ve aritmi varlığı kestirimde kullanılabilecek bağımsız değişkenler olarak belirlenmiştir. Çalışma sonucunda; yaş artışı ile birlikte ölüm riskinin arttığı, yaş>69.8 için ise oldukça yüksek olduğu, patolojik kalp

atım hızı sorununun ölüm riskini arttırıcı bir etken olduğu, LVEF'inde ölüm riski üzerinde etkili olduğu, yüksek LVEF değerlerinin düşük mortalite getirdiği bulunmuştur. Diyabet varlığı mortalite hızını arttırmaktadır. SDNN'nin 54 ile 63 arasındaki değerlerinde ölüm riski artarken 63'ten büyük değerlerinde ölüm riski sabit seviyeye düşmektedir. Yavaşlama kapasitesi (DC) kalp krizi sonrası ölümün kestirimcisi gibi görünmekte, yüksek DC değerleri ölüm riskini düşürmektedir. Diyabet ve HRT etkileşimi de mortalite üzerinde etkilidir.

1.5. MARS ve Yaşam Çözümlemesi

Friedman (1991) tarafından geliştirilen çok değişkenli uyarlanabilir regresyon kesitlerinin (MARS), Kooperberg ve ark. (1997), Mallick ve ark. (1999) tarafından getirilen yeni yaklaşımlarla sağkalım verilerine uygulanabilme olanağı ortaya çıkmıştır.

1.5.1. MARS yönteminin yaşam çözümlemesine uyarlanması

MARS, sürekli ve kategorik değişkenleri aynı anda kullanabilen bir yaklaşımdır. Bu tez çalışmasında MARS'ın bu özelliğinden yararlanılacaktır. Bu sayede; Cox-null modelinin Martingale ya da Deviance artıklarını MARS'ın bağımlı (çıktı ya da cevap) değişkeni olarak kullanan bir yöntem geliştirilerek, sağkalım üzerinde etkili olabilecek bağımsız değişkenleri belirlemede farklı bir yaklaşım getirilmeye çalışılmaktadır.

Kooperberg ve ark. (1995) tarafından geliştirilen hazard regresyon (HARE) modeli, zaman bağımlı ortak değişkenlerin modele katılabilmesine olanak sağlamıştır. Geliştirilen HARE modeli (bazı kaynaklarda olasılıklı log-hazard fonksiyonu olarak da geçmektedir) $\alpha(t | x) = \log(h(t | x))$, parçalanmış doğrusal kesitlerin temel fonksiyonlarını kullanmaktadır. Belirtilen temel fonksiyonların seçilmesinde en büyük etken daha önce Friedman tarafından

iyi bir kestirim gücüne sahip olduklarının ispatlanmış olmasıdır. İyi bir kestirim gücüne sahip olmak, sağkalım analizi yapılırken dikkate alınan en önemli göstergedir.

Kooperberg ve arkadaşları tarafından geliştirilen HARE modeli, veri tarafından belirlenen parametrik olmayan bir temel fonksiyona benzemektedir. HARE fonksiyonu, Friedman tarafından geliştirilen MARS modeline çok benzer bir yapıdadır. Her iki algoritmada model uzayında belirleyici arama yapmaktadır fakat Kooperberg ve arkadaşları temel olarak sayısal faktörleri dikkate alarak çok az değiştirilmiş (slightly modified) model uzayı üzerinden arama yapmayı seçmişlerdir. Amaç temel fonksiyonları ve fonksiyonlara ilişkin katsayıları belirlemek olduğundan, olasılıklı log-hazard fonksiyonu aşağıdaki biçimde yazılabilir:

$$\alpha(t | \mathbf{x}) = \sum_{j=1}^k \beta_j B_j(t | \mathbf{x})$$

Eşitlikte; β_j temel fonksiyonun katsayısını, B_j temel fonksiyonu ve k ise model tarafından karar verilen temel fonksiyon sayısını göstermektedir. Gerektiğinde, temel fonksiyonun zamana göre değişmesine izin verilerek zaman bağımlı katsayılar modele kolayca eklenebilir (Kooperberg ve ark., 1995).

Yoğunluk fonksiyonu $f(t | \mathbf{x})$, yaşam fonksiyonu $S(t | \mathbf{x})$, hazard fonksiyonu $h(t | \mathbf{x})$ ve log-hazard fonksiyonu $\alpha(t | \mathbf{x})$ ile gösterilsin. Ortak değişkenlerin $p \times 1$ boyutlu vektörleri ise $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ olsun. Bu

durumda $S(t | \mathbf{x}) = \int_t^{\infty} f(u | \mathbf{x}) du$, $h(t | \mathbf{x}) = \frac{f(t | \mathbf{x})}{S(t | \mathbf{x})}$ ve $\alpha(t | \mathbf{x}) = \log h(t | \mathbf{x})$

olacaktır. Buradan;

$$S(t | \mathbf{x}) = \exp \left[- \int_0^t \exp(\alpha(t | \mathbf{x})) du \right]$$

elde edilebilir (Kooperberg ve ark., 1995). Cox-PH modelindeki sağkalım fonksiyonu aynı zamanda $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ olmak üzere;

$$S(t | \mathbf{x}) = \exp \left[- \Lambda_0(t) \exp(f(\mathbf{x})) \right]$$

ile de gösterilmektedir (Kruiner, 2007). Buna bağlı olarak Cox modelinin dağılım fonksiyonu;

$$F(t | \mathbf{x}) = 1 - \exp \left[- \Lambda_0(t) \exp(f(\mathbf{x})) \right]$$

ile tanımlanabilir. Dağılım fonksiyonu F olan rastgele bir değişken Y ile gösterildiğinde, $U=F(Y)$ fonksiyonu $[0, 1]$: $U \sim U[0, 1]$ aralığında tekdüze dağılım gösterir. Buradan eğer $U \sim U[0, 1]$ ise $(1-U) \sim U[0, 1]$ olarak kalır.

Cox modelinde sağkalım zamanı T ile gösterildiğinde;

$$U = 1 - F(t | \mathbf{x}) = \exp \left[- \Lambda_0(t) \exp(f(\mathbf{x})) \right] \approx U[0, 1]$$

olacaktır. Tüm t 'ler için $\lambda_0(t) > 0$ olduğundan $\Lambda_0(t)$ terslenebilir ve Cox modelin sağkalım zamanı olan T ;

$$T = \Lambda_0^{-1}(t) \left[-\log(U) \exp(f(\mathbf{x})) \right]$$

olur. Eşitlikteki U , $U \sim U[0, 1]$ tekdüze dağılıma sahip rastgele değişken haline gelir.

Üstel dağılım için sağkalım zamanı $T = -\frac{\log(U)}{\lambda \exp(f(x))}$ ve hazard fonksiyonu ise $\lambda(t|x) = \lambda \exp(f(x))$ ile gösterilir. Weibull dağılımı için ise sağkalım zamanı $T = \left(-\frac{\log(U)}{\lambda \exp(f(x))} \right)^{\frac{1}{u}}$ ve hazard fonksiyonu ise $\lambda(t|x) = \lambda \exp(f(x)) u t^{u-1}$ şeklinde gösterilir (Kriner, 2007).

Bu tez çalışmasının amacı; bağımsız değişken kullanmayan Cox regresyonun (Cox–null model) Martingale ve Deviance artıklarını incelemek ve bu sayede Cox modelinde kullanılacak bağımsız değişkenleri ve bu bağımsız değişkenlerin kesim noktalarını belirlemektir. Bu amacı gerçekleştirebilmek için Friedman'ın (1991) tanıttığı Çok Değişkenli Uyarlanabilir Regresyon Kesitleri (Multivariate Adaptive Regresyon Splines – MARS) yönteminin sağkalım çözümlemesine uyarlaması kullanılacaktır. Cox–null modelinin Martingale ve Deviance artıkları MARS modelinin bağımlı değişkeni olarak alınacak ve bağımlı değişken üzerinde etkili olan bağımsız değişkenler (Cox regresyonun ortak değişkenleri) ile bu değişkenlerin kesim noktaları belirlenmeye çalışılacaktır. Bağımsız değişkenlerin ikili etkileşimleri ve her bir bağımsız değişken için olası kesim noktaları MARS yöntemi ile belirlenecektir. Çalışmanın son aşamasında ise farklı alanlardan elde edilmiş gerçek sağkalım (Karsinom ve UMARU-UIS) verileri üzerinde çalışmada tanıtılan yöntem denenerek, sonuçlar tartışılacaktır. Çalışmada kullanılan veri setleri üzerinde üzerinde daha önce MARS yöntemi uygulanmamıştır.

2. GEREÇ VE YÖNTEM

Yaşam çözümlenmesinde kullanılmak üzere tanımlanan MARS modelinin performansı, üstel ve Weibull dağılımlarında gösterdiği işlevsellik, kararlılık ve tutarlılık Kriner tarafından yapılan benzetim çalışmaları ile ispatlanmıştır (Kriner, 2007). Yaşam çözümlenmesi için geliştirilen MARS modeli farklı parametreler ile yapılan benzetim çalışmalarında bağımlı değişken üzerinde etkili olan bağımlı değişkenleri ve bunların etkileşimlerini belirleyebilmektedir. Bu tez çalışmasında; Kooperberg ve ark. ile Mallick ve ark. tarafından geliştirilen (Kooperberg ve ark, 1995; Kooperberg ve ark., 1997; Mallick ve ark., 1999), performansı ve kararlılığı Kriner tarafından ispatlanan MARS modeli kullanılacaktır.

2.1. İstatistiksel Analizler

Yaşam çözümlenmesine uyarlanmış MARS modeli uygulaması; **R** Ver. 2.12.0, (R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>) programında Kriner (2007) tarafından geliştirilmiş, kullanımı tamamen serbest olan, açık kaynak kodlu R kodlarının bu tez çalışmasına uyarlanması ile elde edilen R kodları ile yapılmıştır. Uygulamada **R base** dışında; **earth** (MARS isim olarak Salford System firmasına marka tescilli olduğundan MARS için R paketinin ismi earth olarak geçmektedir), **polspline**, **mda**, Markov chain Monte Carlo (**MCMC**) modülleri (hazır fonksiyonlardan oluşan, detaylı yardım dosyalarını içeren R eklentileri) ve bu modüllerin kendiliğinden otomatik olarak kullandığı diğer modül ve prosedürler kullanılmıştır.

Uygulama için yukarıda belirtilen R modüllerinin yanı sıra SPSS for Win. Ver. 15.0 (SPSS Inc., Chicago, IL., USA) paket programının Cox

regresyon modülü ile STATISTICA Ver. 7 (StatSoft, Inc. (2004), www.statsoft.com) programının MARsplines modülü kullanılmıştır. Martingale ve Deviance artıkları elde edebilmek için NCSS 2007 (NCSS, PASS, and GESS, NCSS, Kayswille, Utah, USA) programından yararlanılmıştır. Tanımlayıcı istatistiklerin gösteriminde sayı ve yüzde ile birlikte sürekli veriler için ortanca (Çeyreklikler Arası Genişlik – ÇAG – Interquartile Range –IQR–) değerleri kullanılmıştır. Cox regresyon analizlerinde Geriye Doğru Eleme–Olabilirlik Oranı (Backward-LR) yöntemi modele giriş olasılığı için 0.05 ve modelden çıkış olasılığı için 0.10 parametreleri ile çalışılmıştır. Cox regresyon analizi sonuçlarının gösteriminde Beta (B) katsayısı ile birlikte bu katsayıya ilişkin standart hata (S_h), Wald istatistiği ve Hazard oranı (HR) kullanılmıştır. Hazard oranı için %95 güven aralıkları (GA) alt ve üst sınırları belirlenmiştir. Cox modellerinin başarısını değerlendirebilmek amacı ile -2 log olabilirlik (2Log Likelihood, -2LL) değerleri hesaplanmıştır. Cox-PH modelleri arasındaki farklılıkların araştırılması için ki-kare testi kullanılmıştır. Tüm istatistiksel kararlarda $p \leq 0.05$ değeri anlamlı farklılığın göstergesi olarak kabul edilmiştir.

3. BULGULAR

3.1. Karsinom Gerçek Veri Uygulaması

Yaşam çözümlenmesinde kullanılmak üzere tanımlanan MARS fonksiyonunu gerçek klinik verisi üzerinde denemek amacı ile Göğüş ve arkadaşları (2004) tarafından kullanılan lokalize renal hücreli karsinom hastalarına ilişkin toplam 151 hastadan oluşan veri seti (karsinom) kullanılacaktır. Veri setinin kullanımı için arařtırmacılardan gereken izin alınmıřtır. Karsinom veri seti Çizelge 3.1'deki deęiřkenlerden oluřmaktadır.

Çizelge 3.1. Karsinom veri setindeki deęiřkenler

Deęiřken	Açıklama
Takip_ay	Takip zamanı (ay)
Sansur	Durum (0: Sansürlü, 1: Ölüm)
Yas	Hasta yaşı (yıl)
Cinsiyet	Hasta cinsiyeti (0: Erkek, 1: Kadın)
Evre	Patolojik Evre (1–3)
Tumor	Patolojik Tümör derecesi (1–2, 3–4)
Lenfnod	Lenf nodu tutulumu (Yok, Var)
Hucretip	Histolojik hücre tiplmesi (Clear, Dięer)
Plt	Platelet sayısı (*1000 mm ³)

3.1.1. Karsinom: Tanımlayıcı istatistikler

Karsinom hastalarının yaş ortancası 60.0 (Çeyrekler arası genişlik, Interquartile Range, ÇAG=17.0) yıl olarak belirlenmiştir. Takip süre ortancası ise 34.0 (ÇAG=32.0) aydır. Karsinom nedeni ile ölen hastaların yaş ortancası 57.0 (ÇAG=21.0) yıl iken, sansürlenmiş gözlemlerin yaş ortancası 60.0

(ÇAG=17.0) yıl olarak hesaplanmıştır. Çalışmada ölen hastaların takip süre ortancası 9.0 (ÇAG=21.0) ay iken sansürlenmiş hastaların takip süre ortancası 42.0 (ÇAG=26.0) aydır. Karsinom veri setindeki diğer değişkenlere ilişkin tanımlayıcı istatistikler Çizelge 3.1.1’de verilmiştir.

Çizelge 3.1.1. Karsinom veri seti tanımlayıcı istatistikleri

Değişkenler	D U R U M		Toplam (n=151)
	Ölüm (n=31)	Sansürlü (n=120)	
Cinsiyet			
Kadın	7 (13.2)	46 (86.8)	53
Erkek	24 (24.5)	74 (75.5)	98
Patolojik Evre			
1–2	15 (13.9)	93 (86.1)	108
3	16 (37.2)	27 (62.8)	43
Patolojik Tümör derecesi			
1–2	11 (10.3)	96 (89.7)	107
3–4	20 (45.5)	24 (54.5)	44
Lenf nodu tutulumu			
Yok	25 (18.2)	112 (81.8)	137
Var	6 (42.9)	8 (57.1)	14
Hücre tipi			
Clear	26 (22.0)	92 (78.0)	118
Diğer	5 (15.2)	28 (84.8)	33
Platelet			
Ortanca (ÇAG)	290.0 (ÇAG=297.0)	269.5 (ÇAG=112.0)	270.0 (ÇAG=125.0)

3.1.2. Karsinom: Cox-PH modeli sonuçları

Karsinom veri seti üzerinde klasik Cox-PH modeli uygulandı. Cox-null modelin -2 log-olabilirlik (-2LL) değeri 291.273 olarak belirlendi. Daha sonra tüm bağımsız değişkenlerin yer aldığı Cox-PH modeli Backward-LR yöntemi ile çalışıldı. Çözümleme toplam 5 adımda gerçekleştirildi. İlk adımda (başlangıç modeli) tüm değişkenler modele alındı. Başlangıç modeli için -2LL

değeri 248.861 olarak belirlendi. Başlangıç modelindeki değişkenler ve bu değişkenlerin modele katkıları Çizelge 3.1.2.1'de birikimli sağkalım grafiği ise Şekil 3.1.2.1'de gösterilmiştir.

Çizelge 3.1.2.1. Karsinom veri seti: Cox-PH modeli başlangıç değişkenleri ve modele katkıları

Değişkenler	B	S _n	Wald	p	HR	HR %95 GA	
						Alt	Üst
Yaş	-0.029	0.020	2.179	0.140	0.972	0.935	1.010
Cinsiyet *	0.804	0.455	3.131	0.077	2.235	0.917	5.447
Evre *	0.960	0.395	5.901	0.015	2.613	1.204	5.671
Tümör *	1.626	0.419	15.039	<0.001	5.082	2.235	11.558
Lenfnod †	0.076	0.504	0.023	0.880	1.079	0.402	2.901
Hücre tipi †	0.589	0.513	1.317	0.251	1.802	0.659	4.924
Platelet	0.004	0.001	8.252	0.004	1.004	1.001	1.007

S_n: Standart hata

HR: Hazard oranı

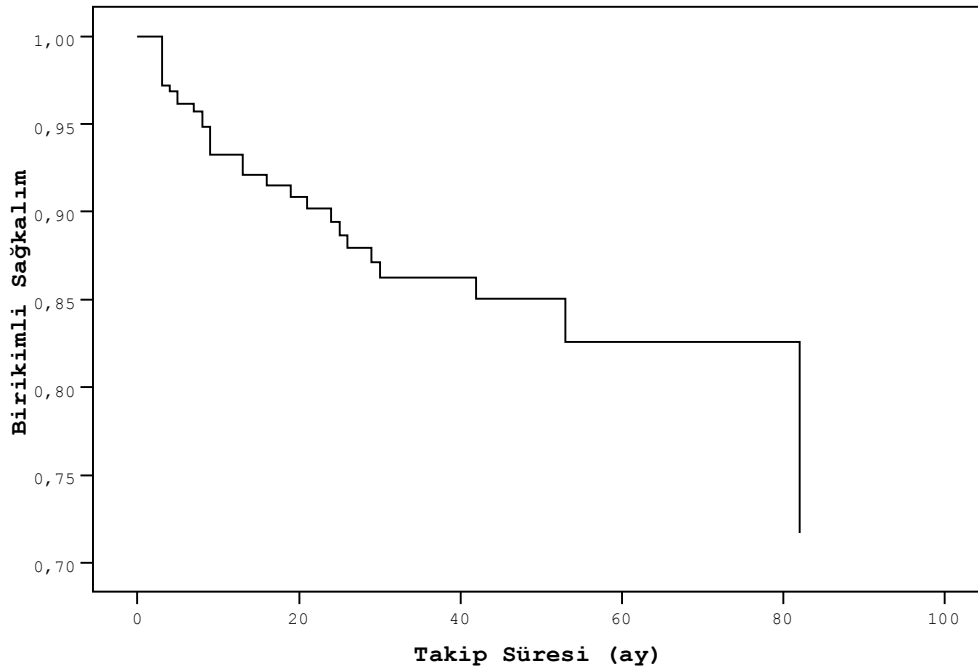
-2LL=248.861

*: İlk kategori referans olarak alınmıştır.

†: İkinci kategori referans olarak alınmıştır.

Kategorik değişkenlerde risk artışına neden olan grubu daha net gösterebilmek amacı ile değişkenler ilk ya da son kategorileri referans kategori olacak biçimde tanımlanmıştır.

Cox-PH regresyonu sonuçlarına göre; bağımsız değişkenlerden Evre, Tümör ve Platelet modele istatistiksel olarak anlamlı katkıda bulunurken, diğer bağımsız değişkenlerin modele katkıları istatistiksel olarak anlamlı değildir. Kurulan model Cox-null modelden istatistiksel olarak farklıdır ($\chi^2=42.412$; $sd=7$; $p<0.001$). Başlangıç modelindeki istatistiksel olarak anlamlı katkıya sahip değişkenlerle kurulan Cox-PH modelinin 5. ve son adımında (final model), -2LL değeri 248.861'den 254.744'e yükselmiştir. Ancak bu artış istatistiksel olarak anlamlı değildir ($\chi^2=5.883$; $sd=4$; $p=0.208$). Final modelde yer alan değişkenler ve modele katkıları Çizelge 3.1.2.2'de, birikimli sağkalım grafiği Şekil 3.1.2.2'de, birikimli hazard grafiği ise Şekil 3.1.2.3'te verilmiştir.



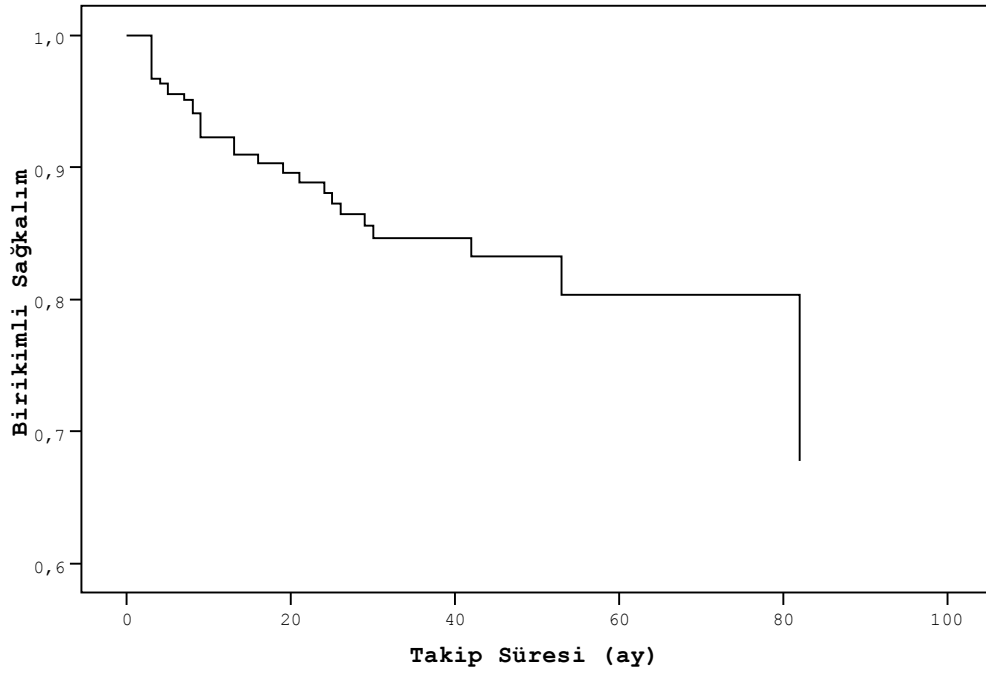
Şekil 3.1.2.1. Karsinom veri seti: Birikimli sağkalım fonksiyonu

Çizelge 3.1.2.2. Karsinom veri seti: Final Cox-PH modeli değişkenleri ve modele katkıları

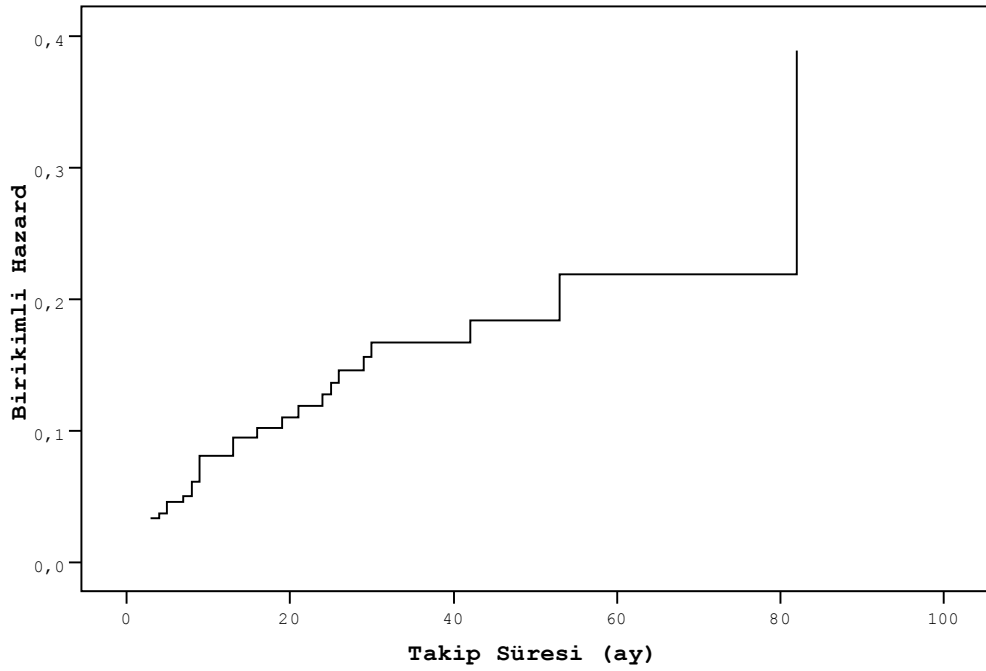
Değişkenler	B	S _h	Wald	p	HR	HR %95 GA	
						Alt	Üst
Evre	0.758	0.379	4.002	0.045	2.133	1.015	4.481
Tümör	1.644	0.404	16.599	<0.001	5.176	2.347	11.416
Platelet	0.003	0.001	5.413	0.020	1.003	1.000	1.006

S_h: Standart hata -2LL=254.744

Cox-PH final modeli detaylı olarak incelendiğinde; hastalık evresinin sağkalım zamanı üzerinde doğrudan etkisinin olduğu görülmektedir. Evre 1–2 hastalara göre Evre 3 hastaların ölüm hızı HR=2.133 kat (%95 güven aralığı 1.015 – 4.481) daha fazladır. Benzer şekilde patolojik tümör derecesi 3–4 olan hastaların ölüm hızı 1–2 olan hastalara göre HR=5.176 kat daha fazla iken Platelet değerindeki her 100 birimlik artış ölüm hızında HR=1.349 kat (%95 güven aralığı: 1.010 – 1.819) artışa neden olmaktadır.



Şekil 3.1.2.2. Karsinom veri seti: Final Cox-PH modeli birikimli sağkalım fonksiyonu



Şekil 3.1.2.3. Karsinom veri seti: Final Cox-PH modeli birikimli hazard fonksiyonu

3.1.3. Karsinom: MARS modeli sonuçları (Martingale)

Martingale artıklarını bağımlı değişken olarak kullanan MARS modeli ($MARS_M$) ile Karsinom verileri analiz edildiğinde, modelin açıklayıcılık değeri $R^2=0.339$ (düzeltilmiş $R^2=0.302$) olarak belirlenmiştir. $MARS_M$ en iyi sonuca ulaşmak için toplam 40 farklı model oluşturmuş, bu modellerde bağımsız değişkenler ile bağımsız değişkenlerin farklı kesim noktalarını da içeren 24 farklı temel fonksiyon kullanmıştır. Oluşturulan modellerin GCV değerleri 0.149 – 0.205 arasında değişmektedir. En uygun model, GCV değeri en düşük olan (ideal) modeldir. İdeal $MARS_M$ modeli; model sabiti, patolojik evre (evre), patolojik tümör derecesi (tumor) ve Platelet değişkenleri ile Platelet'in farklı kesim noktalarını içeren 8 temel fonksiyondan oluşmaktadır. İdeal modelde yer alan temel fonksiyon (BF) öğeleri, bu öğelere ilişkin katsayılar ve kesim noktaları Çizelge 3.1.3.1'de verilmiştir.

Çizelge 3.1.3.1. Karsinom veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_M$)

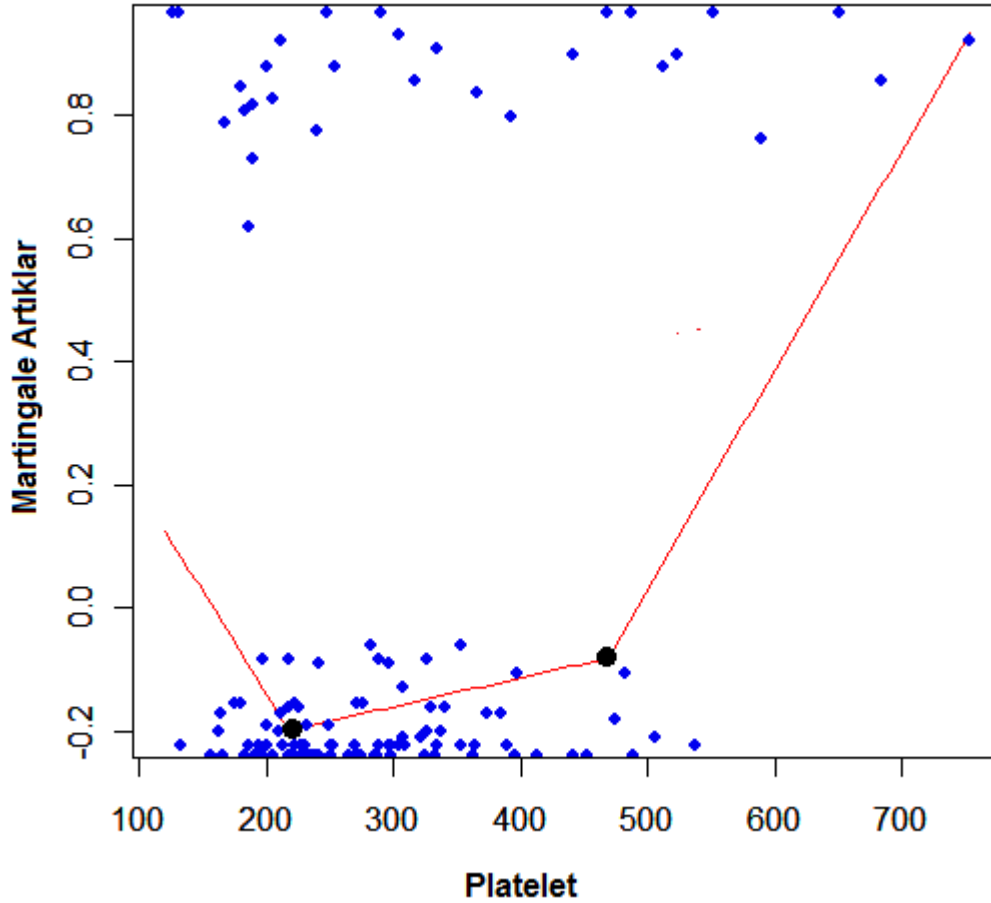
BF No	Değişken	Kesim Noktası	B	S_h
1	Model Sabiti	-	-0.0940	0.3888
2	Cinsiyet	-	0.0800	0.0667
3	Evre	-	0.1675	0.0743
4	Tumor	-	0.2920	0.0751
5	Platelet	-	-0.0033	0.0018
6 *	Platelet	218	0.0038	0.0020
7 *	Platelet	468	0.0031	0.0013

BF: Temel fonksiyon **S_h : Standart hata**

$MARS_M$ modeline göre; Martingale artıklar (risk) üzerinde etkili olan bağımsız değişkenler Cinsiyet, Evre, Tumor ve Platelettir. Platelet sadece bağımsız değişken olarak değil, aynı zamanda 218 ve 468 değerlerinin oluşturduğu kesim noktaları ile de risk üzerinde anlamlı etkiye sahiptir.

* Bu temel fonksiyonların hesaplanmasında; eğer Platelet > Kesim noktası ise, (Platelet – Kesim noktası), değil ise (Kesim noktası – Platelet) (ayna temel fonksiyon mantığı) kullanılmalıdır. Bu temel fonksiyonların gösterimi (Platelet – Kesim noktası)₊ şeklindedir.

Martingale artıklara göre Platelet kesim değerleri Şekil 3.1.3.1’de gösterilmiştir.



Şekil 3.1.3.1. Martingale artıklara göre Platelet kesim noktaları

Şekil 3.1.3.1.’e göre; Platelet ile Hazard arasındaki ilişki doğrusal değildir. Platelet değeri 218’e kadar ölüm riskini azaltıcı bir etkiye sahiptir. Platelet 218 – 467 aralığında ölüm riski çok az artış göstermekte, Plateletin 468’den büyük değerlerinde ise ölüm riski hızla yükselmektedir. Platelet değişkeni Cox modeline alınacağı zaman belirlenen kesim noktalarını dikkate alarak yeni bir değişken adı altında kodlamak gerekir. Platelet değeri <218, 218 – 467 ve 468+ olarak üç sınıflı bir kategorik değişken olarak yeniden kodlandı.

MARS_M tarafından belirlenen değişkenler ile Cox modeli yeniden kuruldu. Cox modelinde yer alan değişkenler ve bu değişkenlere ilişkin katsayılar Çizelge 3.1.3.2'de verilmiştir.

Çizelge 3.1.3.2. Karsinom veri seti: Cox-PH modeli değişkenleri ve modele katkıları (MARS_M)

Değişkenler	B	S _h	Wald	p	HR	HR %95 GA	
						Alt	Üst
Tumor	1.788	0.387	21.297	<0.001	5.979	2.798	12.777
Platelet			13.313	0.001			
Platelet (218-467)	-0.749	0.433	2.995	0.084	0.473	0.202	1.104
Platelet (468+)	0.936	0.474	3.899	0.048	2.551	1.007	6.461
<i>S_h: Standart hata</i>						<i>-2LL=252.370</i>	

Çizelge 3.1.3.2'de yer alan Cox modeli incelendiğinde; final Cox modelinden farklı olarak Patolojik Evre değişkeninin MARS_M tarafından belirlenen değişkenlerle kurulan Cox modelinde yer almadığı dikkat çekmektedir. Cox final modeline katkısı sınırda anlamlı olan (p=0.045) Evre değişkeni yeni Cox modeline anlamlı bir katkıda bulunamadığından model dışında kalmıştır. MARS_M tarafından modelde yer alması gerektiği belirlenen Cinsiyet değişkeni de yeni Cox modeline anlamlı katkı sağlayamadığından model dışında kalmıştır. Yeni oluşturulan modelin -2LL değeri 252.370 ile final Cox modelinin -2LL değerinden (254.744) daha düşüktür. Ancak modeller arasındaki farklılık istatistiksel olarak anlamlı değildir ($\chi^2=2.374$; Sd=1; p=0.123).

MARS_M tarafından belirlenen değişkenlerle kurulan Cox modeli final modelinden daha düşük -2LL değerine sahiptir. Platelet değişkeninde MARS ile belirlenen kesim noktaları araştırmacıya kurulan modeli yorumlamada kolaylık sağlamaktadır.

Patolojik Tümör derecesi 1-2 olan hastalara göre 3-4 olan hastaların hazard hızı HR=5.979 kat (%95 GA: 2.798 – 12.777) daha fazladır. Platelet değeri 218 – 467 arasında yer alan hastaların, 218'den daha düşük Platelet değerine sahip hastalara göre hazard hızları istatistiksel olarak farksızdır. Ancak 468 ve daha yüksek Platelet değerine sahip hastalar, 218'den daha düşük Platelet değerine sahip hastalara göre HR=2.551 kat (%95 GA: 1.007 – 6.461) daha yüksek hazard hızına sahiptirler.

3.1.4. Karsinom: MARS modeli sonuçları (Deviance)

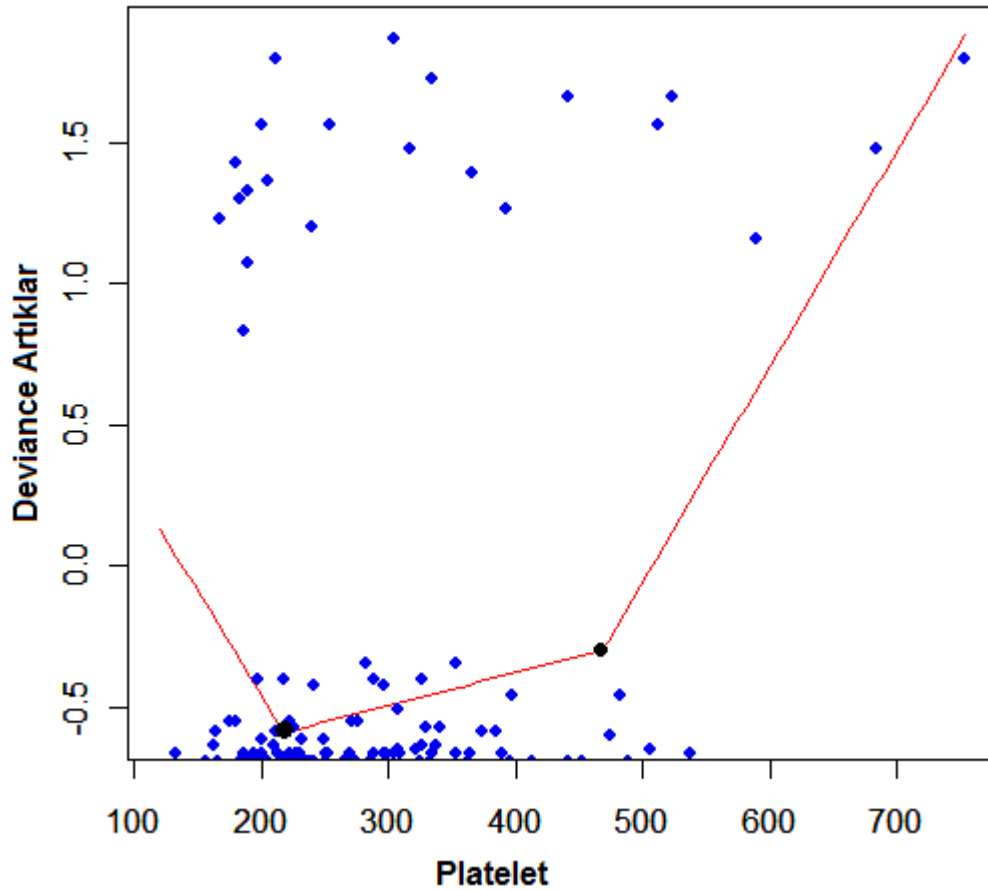
Deviance artıklarını bağımlı değişken olarak kullanan MARS modeli (MARS_D) ile Karsinom verileri analiz edildiğinde, modelin açıklayıcılık değeri $R^2=0.335$ (düzeltilmiş $R^2=0.301$) olarak belirlenmiştir. MARS_D, en iyi sonuca ulaşmak için 56 farklı model oluşturmuş, bu modellerde bağımsız değişkenler ile bağımsız değişkenlerin farklı kesim noktalarını da içeren 32 farklı temel fonksiyon kullanmıştır. Oluşturulan modellerin GCV değerleri 0.668 – 0.939 arasında değişmektedir. En uygun model, GCV değeri en düşük olan (ideal) modeldir. İdeal MARS_D modeli; model sabiti, Patolojik Evre (evre), Patolojik Tümör Derecesi (Tumor), Hücre Tipi (Hucretip) ve Platelet değişkenleri ile Platelet'in farklı kesim noktalarını (218 ve 468) içeren 8 temel fonksiyondan oluşmaktadır. İdeal modelde yer alan temel fonksiyon (BF) öğeleri, bu öğelere ilişkin katsayılar ve katsayıların standart hatası Çizelge 3.1.4.1'de verilmiştir.

Deviance artıklara göre Platelet kesim değerleri Şekil 3.1.4.1'de gösterilmiştir. Şekil 3.1.4.1'e göre; Platelet ile Hazard arasındaki ilişki doğrusal değildir. Platelet değeri 217'ye kadar ölüm riskini azaltıcı (koruyucu) bir etkiye sahiptir. Platelet 217 – 467 aralığında ölüm riski yavaşça artmakta, Plateletin 468'den büyük değerlerinde ise risk artışı hızlanmaktadır.

Çizelge 3.1.4.1. Karsinom veri seti: İdeal modele ilişkin temel fonksiyonlar (MARS_D)

BF No	Değişken	Kesim Noktası	B	S _h
1	Model Sabiti	-	0.0347	0.8032
2	Cinsiyet	-	0.1725	0.1422
3	Evre	-	0.3660	0.1579
4	Tumor	-	0.6216	0.1592
5	Hucretip	-	-0.1731	0.1613
6	Platelet	-	-0.0074	0.0037
7 †	Platelet	218	0.0086	0.0043
8 *	Platelet	468	0.0065	0.0028

BF: Temel fonksiyon

S_h: Standart hata

Şekil 3.1.4.1. Deviance artıklara göre Platelet kesim noktaları

* Bu temel fonksiyonların hesaplanmasında; eğer Platelet > Kesim noktası ise, (Platelet – Kesim noktası), değil ise (Kesim noktası – Platelet) (ayna temel fonksiyon mantığı) kullanılmalıdır. Bu temel fonksiyonların gösterimi (Platelet – Kesim noktası)₊ şeklindedir.

MARS_D tarafından belirlenen değişkenler ile Cox modeli yeniden kuruldu. Cox modelinde yer alan değişkenler ve bu değişkenlere ilişkin katsayılar Çizelge 3.1.4.2'de verilmiştir.

Çizelge 3.1.4.2. Karsinom veri seti: Cox-PH modeli değişkenleri ve modele katkıları (MARS_D)

Değişkenler	B	S _h	Wald	p	HR	HR %95 GA	
						Alt	Üst
Tumor	1.788	0.387	21.297	<0.001	5.979	2.798	12.777
Platelet			13.313	0.001			
Platelet (218-467)	-0.749	0.433	2.995	0.084	0.473	0.202	1.104
Platelet (468+)	0.936	0.474	3.899	0.048	2.551	1.007	6.461
S _h : Standart hata						-2LL=252.370	

Çizelge 3.1.4.2'de yer alan Cox modeli incelendiğinde; tüm sonuçların Çizelge 3.1.3.2'de verilen MARS_M destekli Cox regresyon sonuçları ile bire bir aynı oldukları görülmektedir. MARS_M destekli Cox modelinde yapılan tüm yorumlar aynı şekilde MARS_D destekli Cox modeli için de geçerlidir.

Karsinom veri seti üzerinde uygulanan MARS_M ve MARS_D model sonuçları aynıdır. Martingale artıkları kullanan MARS modeliyle Deviance artıkları kullanan MARS modeli Karsinom veri seti için hazard üzerinde etkili olan değişkenleri ve bu değişkenlerin kesim noktalarını doğru olarak tespit etmişlerdir.

3.2. UMARU – UIS Veri Seti Uygulaması

Yaşam çözümlenmesinde kullanılmak üzere tanımlanan MARS fonksiyonunu gerçek klinik verisi üzerinde denedikten sonra, sağkalım çözümlenmesi literatürüne girmiş bir veri seti üzerinde denemenin de MARS modelinin etkinliğini göstermek açısından yararlı olacağı düşünülmüştür. Bu amaçla; Hosmer ve Lemeshow'un (1999) kitaplarında da yer verdikleri 1989 – 1994 yılları arasında yapılmış Massachusetts Üniversitesi AIDS Araştırma Birimi (UMARU) Etki Çalışması (UIS) verileri kullanılacaktır. UIS veri seti; verilerine ulaşılabilen 628 hastadan Çizelge 3.2'de gösterilen bilgileri tam olan, 111'ü sansürlü, 464'ü incelenen olay (uyuşturucu kullanımına geri dönen) gözlenen olmak üzere toplam 575 hastaya ait bilgilerden oluşmaktadır. Hosmer ve Lemeshow veri setinin gerçek veri setinden bir miktar farklı olduğunu belirtmektedirler.

Veriler, ftp://ftp.wiley.com/public/sci_tech_med/survival adresinden 13.10.2010 tarihinde alınmıştır, genel erişime açıktır. UIS veri setinin Çizelge 3.2'de açıklanan değişkenleri çalışmada kullanılacaktır.

Çizelge 3.2. UIS veri setindeki değişkenler

Değişken	Açıklama
Time	Uyuşturucu kullanımına dönüş zamanı (Gün)
Censor	Uyuşturucu kullanımına dönüş (0- Diğer, 1- Evet)
Age	Yaş (Yıl)
Becktota	Beck depresyon skoru (0.0 – 54.0)
Hercoc	Son 3 ayda uyuşturucu kullanımı (1- Eroin + Kokain, 2- Eroin, 3- Kokain, 4- Kullanmadı)
IVHX	Damar içi uyuşturucu kullanım hikayesi (1- Hiçbir zaman, 2- Çok önce, 3- Yakın zamanda)
Ndrugtx	Daha önce aldığı uyuşturucu tedavisi sayısı (0 – 40)
Race	İrk (0- Beyaz ırk, 1- Diğer ırklar)
Treat	Tedavi grubu (0- Kısa tedavi, 1- Uzun tedavi)
Site	Tedavi türü (A, B)
Lot	Tedavi süresi (gün)

3.2.1. UIS: Tanımlayıcı istatistikler

UIS veri setindeki hastalarının yaşları 20 ile 56 arasında değişmektedir. Yaş ortancası 32.0 (ÇAG=10.0) yıl olarak belirlenmiştir. Takip süresi en az 4, en çok 1172, ortancası ise 170.0 (ÇAG=293.0) gün olarak hesaplanmıştır. UIS veri setindeki diğer değişkenlere ilişkin tanımlayıcı istatistikler Çizelge 3.2.1'de verilmiştir.

Çizelge 3.2.1. UIS veri seti tanımlayıcı istatistikleri

Değişkenler	Uyuşturucu Kullanımına Dönüş		Toplam (n=575) Ortanca (ÇAG)
	Evet (n=464) Ortanca (ÇAG)	Diğer (n=111) Ortanca (ÇAG)	
Time	128.0 (155.6)	553.0 (34.0)	170.0 (293.0)
Age	32.0 (10.0)	33.0 (9.0)	32.0 (10.0)
Becktota	17.0 (12.8)	15.0 (14.0)	17.0 (13.0)
Ndrugtx	3.0 (5.0)	2.0 (3.0)	3.0 (5.0)
Lot	83.0 (82.3)	112.0 (112.0)	87.0 (106.0)
	n (%)	n (%)	n (%)
Hercoc			
Eroin + Kokain	87 (83.7)	17 (16.3)	104
Eroin	94 (87.9)	13 (12.1)	107
Kokain	131 (76.2)	41 (23.8)	172
Kullanmadı *	152 (79.2)	40 (20.8)	192
IVHX			
Hiçbir zaman *	165 (74.0)	58 (26.0)	223
Çok önce	88 (80.7)	21 (19.3)	109
Yakın zamanda	211 (86.8)	32 (13.2)	243
Race			
Beyaz ırk	357 (83.0)	73 (17.0)	430
Diğer ırklar *	107 (73.8)	38 (26.2)	145
Treat			
Kısa *	239 (82.7)	50 (17.3)	289
Uzun	225 (78.7)	61 (21.3)	286
Site			
A *	326 (81.5)	74 (18.5)	400
B	138 (78.9)	37 (21.1)	175

* : Referans kategoriler.

Kategorik değişkenlerde risk artışına neden olan grubu daha net gösterebilmek amacı ile değişkenler ilk ya da son kategorileri referans kategori olacak biçimde tanımlanmıştır.

3.2.2. UIS: Cox-PH modeli sonuçları

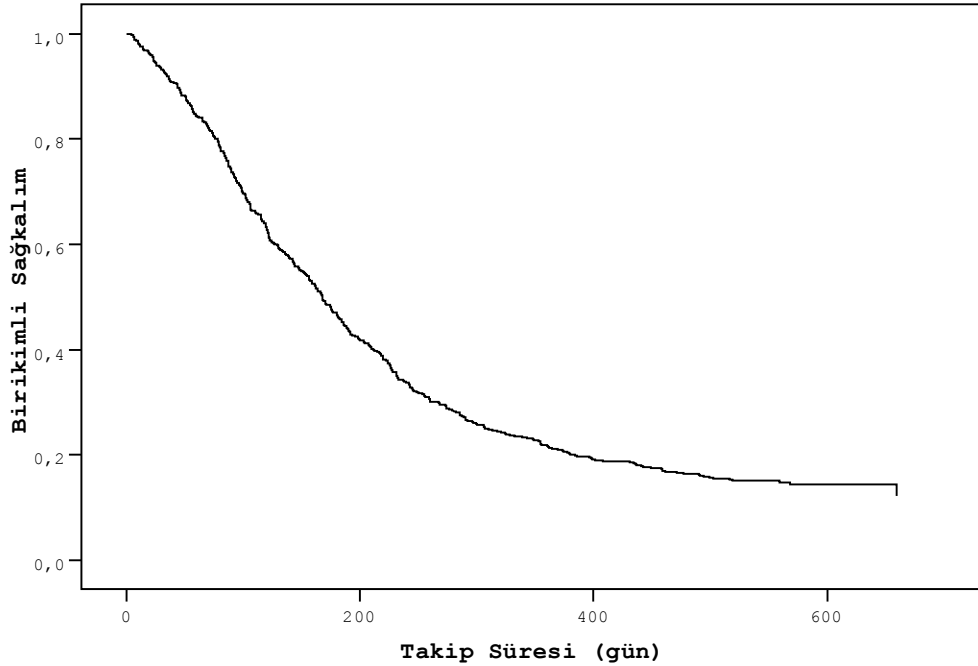
UIS veri seti üzerinde klasik Cox-PH modeli uygulandı. Cox-null modelinin $-2 \log$ -olabilirlik değeri 5327.970 olarak belirlendi. Cox-PH modeli Geriye Dönük Eleme-Olabilirlik Oranı metodu ile 4 adımda en uygun modeli bularak sonlandı. Birinci adım (başlangıç) Cox-PH modeli $-2LL$ değeri 5135.462 olarak hesaplandı. Cox-null ile başlangıç Cox-PH modeli arasındaki farklılık istatistiksel olarak önemlidir ($\chi^2=192.508$; $sd=12$; $p<0.001$). Başlangıç Cox-PH modelindeki değişkenler ve bu değişkenlerin modele katkıları Çizelge 3.2.2.1'de, birikimli sağkalım grafiği ise Şekil 3.2.2.1'de gösterilmiştir.

Çizelge 3.2.2.1. UIS veri seti: Cox-PH modeli başlangıç değişkenleri ve modele katkıları

Değişkenler	B	S _n	Wald	p	HR	HR %95 GA Alt	Üst
Age	-0.021	0.008	6.404	0.011	0.979	0.964	0.995
Becktota	0.005	0.005	1.152	0.283	1.005	0.996	1.015
Ndrugtx	0.026	0.009	9.319	0.002	1.027	1.009	1.044
Lot	-0.009	0.001	130.922	<0.001	0.991	0.989	0.992
Hercoc			1.448	0.694			
Eroin+Kokain	0.001	0.166	0.000	0.994	1.001	0.724	1.385
Eroin	0.152	0.160	0.906	0.341	1.165	0.851	1.594
Kokain	-0.013	0.123	0.001	0.916	0.987	0.776	1.256
Kullanmadı			Referans Kategori				
IVhx			7.290	0.026			
Hiçbir zaman			Referans Kategori				
Çok önce	0.188	0.139	1.832	0.176	1.207	0.919	1.584
Yakın zamanda	0.398	0.148	7.200	0.007	1.488	1.113	1.990
Race							
Beyaz ırk	0.306	0.116	6.942	0.008	1.358	1.081	1.705
Diğer ırklar			Referans Kategori				
Treat							
Kısa			Referans Kategori				
Uzun	0.138	0.097	2.016	0.156	1.148	0.949	1.388
Site							
A			Referans Kategori				
B	0.443	0.011	15.851	<0.001	1.557	1.252	1.936

S_n:Standart hata

$-2LL=5135.462$



Şekil 3.2.2.1. UIS veri seti: Birikimli sağkalım fonksiyonu

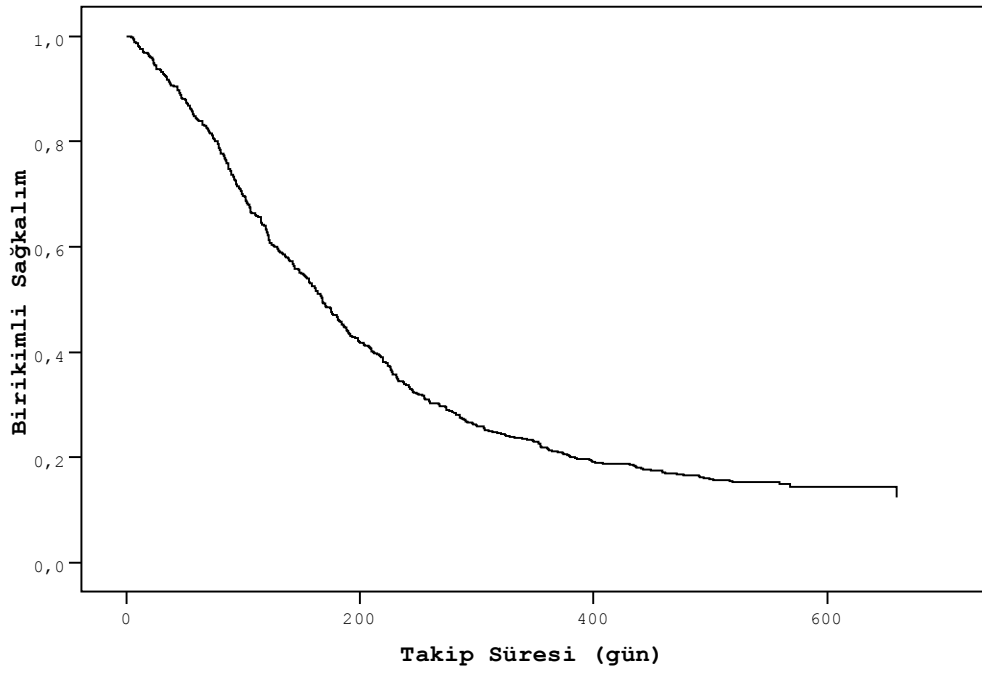
Başlangıç Cox-PH regresyonu sonuçlarına göre; bağımsız değişkenlerden Beck depresyon skoru (Becktota), son 3 ay içerisinde uyuşturucu kullanımı (Hercoc) ve tedavi grubunun (Treat) modele istatistiksel olarak anlamlı katkısı yoktur. Diğer bağımsız değişkenlerin modele katkıları istatistiksel olarak anlamlıdır. Geriye Dönük Eleme-Ençok Olabilirlik sonucunda (4. adımda) Cox-PH regresyon analizi -2LL değeri 5140.293 olarak tespit edilmiştir. Cox-PH regresyonun 1. adımı ile 4. adımı arasında -2LL değerleri açısından istatistiksel olarak anlamlı farklılık yoktur ($\chi^2=4.831$; $sd=5$; $p=0.437$).

Final modelde yer alan değişkenler ve modele katkıları Çizelge 3.2.2.2'de, birikimli sağkalım grafiği Şekil 3.2.2.2'de, birikimli hazard grafiği ise Şekil 3.2.2.3'te verilmiştir.

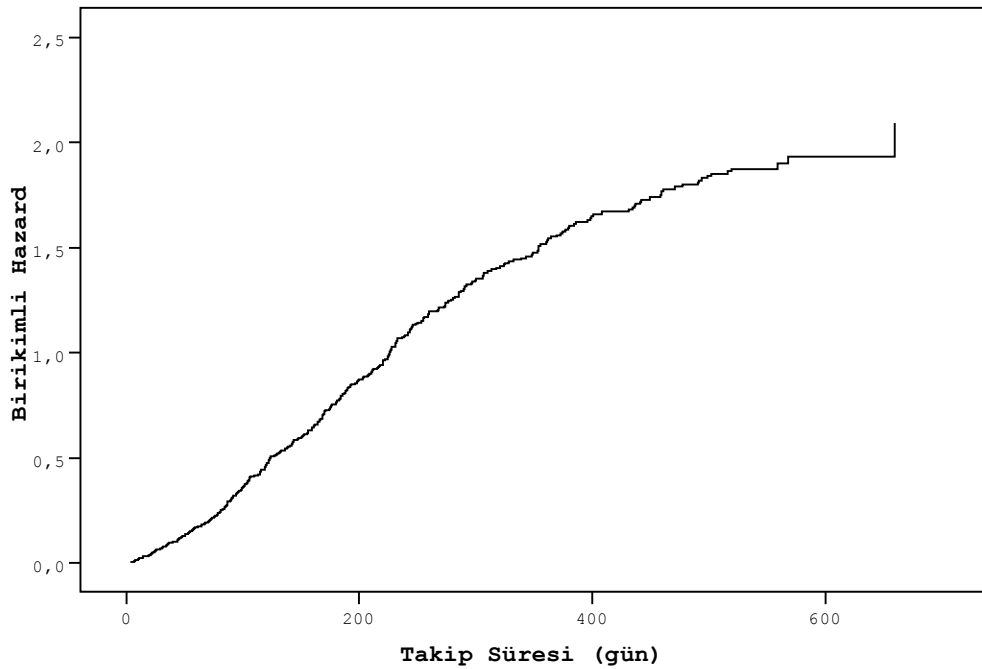
Çizelge 3.2.2.2. UIS veri seti: Final Cox-PH modeli değişkenleri ve modele katkıları

Değişkenler	B	S _h	Wald	p	HR	HR %95 GA	
						Alt	Üst
Age	-0.021	0.008	6.874	0.009	0.979	0.963	0.995
Ndrugtx	0.027	0.008	9.937	0.002	1.027	1.010	1.044
Lot	-0.009	0.001	127.652	<0.001	0.991	0.989	0.992
IVhx			14.098	0.001			
Çok önce	0.194	0.138	1.986	0.159	1.214	0.927	1.590
Yakın zamanda	0.449	0.120	13.929	<0.001	1.566	1.237	1.982
Race							
Beyaz ırk	0.292	0.114	6.528	0.011	1.339	1.070	1.676
Site							
B	0.411	0.110	13.921	<0.001	1.508	1.215	1.872
S_h: Standart hata						-2LL=5140.293	

Cox-PH final modeli detaylı olarak incelendiğinde; modelde yer alan tüm bağımsız değişkenlerin uyuşturucu kullanımına dönüş zamanı üzerinde etkili olduğu görülmektedir. Bağımsız değişkenlerden yaş ve tedavi süresinin azalmasının artmasına göre uyuşturucu kullanımına dönme riskini azaltıcı (süreyi arttırıcı) yönde etkili olduğu, daha önce aldığı uyuşturucu tedavi sayısının yüksek olmasının ise düşük olmasına göre uyuşturucuya dönme riskini arttırıcı (süreyi azaltıcı) etkisi olduğu belirlenmiştir. Beyaz ırktan olanların diğer ırklara göre uyuşturucu kullanımına dönme riski HR=1.339 kat (%95 GA: 1.070 – 1.676) daha fazladır. Yakın zamanda damar yolundan uyuşturucu alanların, hiç almayanlara göre uyuşturucu kullanımına dönme riski HR=1.566 kat (%95 GA: 1.237 – 1.982) daha fazladır. B tipi tedavi alan hastaların uyuşturucuya dönme riski A tipi tedavi alan hastalara göre HR=1.508 kat (%95 GA: 1.215 – 1.872) kat daha yüksektir.



Şekil 3.2.2.2. UIS veri seti: Final Cox-PH modeli birikimli sağkalım fonksiyonu



Şekil 3.2.2.3. UIS veri seti: Final Cox-PH modeli birikimli hazard fonksiyonu

3.2.3. UIS: MARS modeli sonuçları (Martingale)

Martingale artıklarını bağımlı değişken olarak kullanan MARS modeli ($MARS_M$) ile UIS verileri analiz edildiğinde, modelin açıklayıcılık değeri $R^2=0.300$ (düzeltilmiş $R^2=0.293$) olarak belirlenmiştir. $MARS_M$ toplam 80 farklı model oluşturmuş, bu modellerde 45 farklı temel fonksiyon kullanmıştır. Oluşturulan modellerin GCV değerleri 0.5853 ile 0.8066 arasında değişmektedir. En uygun model, GCV değeri en düşük olan (ideal) modeldir. İdeal MARS modeli model sabitinin yanında; yaş (Age), Beck depresyon skoru (Becktota), damar içi uyuşturucu kullanım hikayesi (IVhx), daha önce aldığı uyuşturucu tedavisi sayısı (Ndrugtx), ırk (Race), tedavi türü (Site) ve tedavi süresi (Lot) bağımsız değişkenlerinden Beck depresyon skoru (Bectota) ile tedavi süresi (Lot) bağımsız değişkenlerin farklı kesim noktalarından oluşan 11 adet temel fonksiyon içermektedir (Çizelge 3.2.3.1).

Çizelge 3.2.3.1. UIS veri seti: İdeal modele ilişkin temel fonksiyonlar ($MARS_M$)

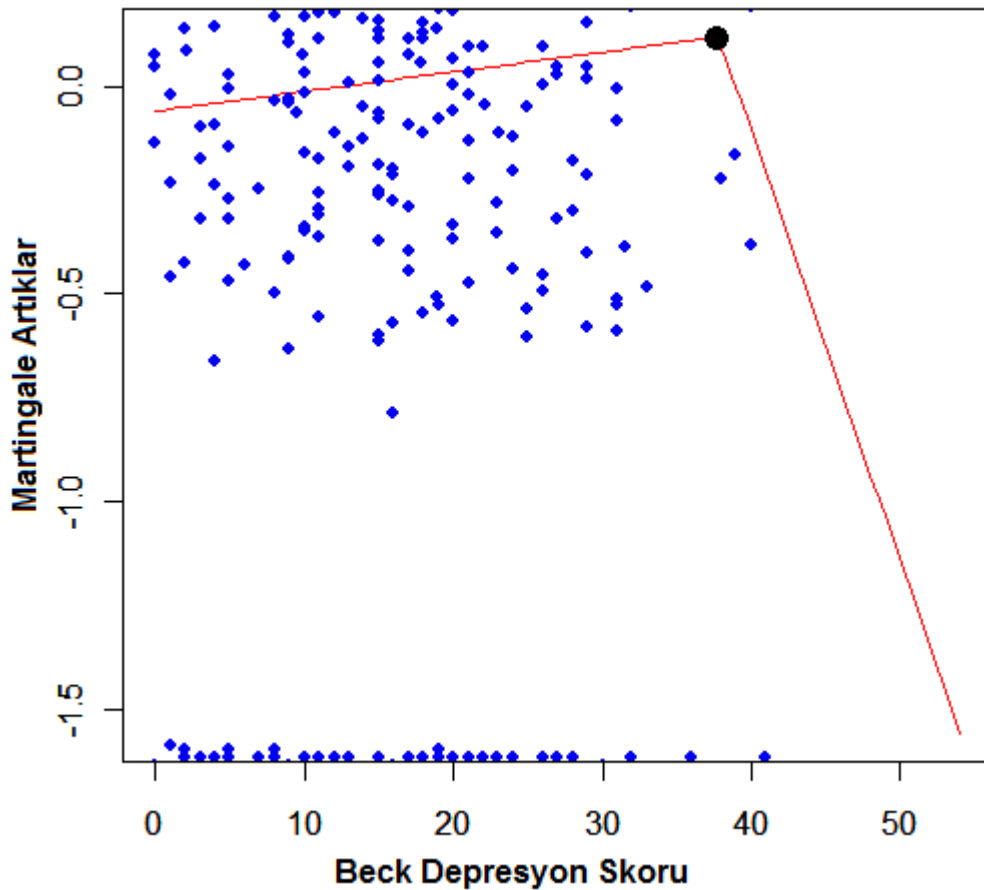
BF No	Değişken Adı	Kesim noktası	B	S_h
1	Model Sabiti	-	0.7060	0.1923
2	Age	-	-0.0129	0.0056
3	Becktota	-	0.0047	0.0036
4 *	Becktota	37.8	-0.1081	0.0438
5	IVhx	-	0.1179	0.0408
6	Ndrugtx	-	0.0168	0.0061
7	Race	-	-0.1826	0.0758
8	Site	-	0.1913	0.0760
9	Lot	-	-0.0073	0.0006
10 ‡	Lot	186.0	0.0338	0.0165
11 *	Lot	199.0	-0.0311	0.0175

BF: Temel fonksiyon **S_h : Standart hata**

* Bu temel fonksiyonların hesaplanmasında; eğer Değişken > Kesim noktası ise, (Değişken – Kesim noktası) değilse (Kesim noktası – Değişken) (ayna temel fonksiyon mantığı) kullanılmalıdır. Bu temel fonksiyonların gösterimi (Değişken – Kesim noktası), şeklindedir.

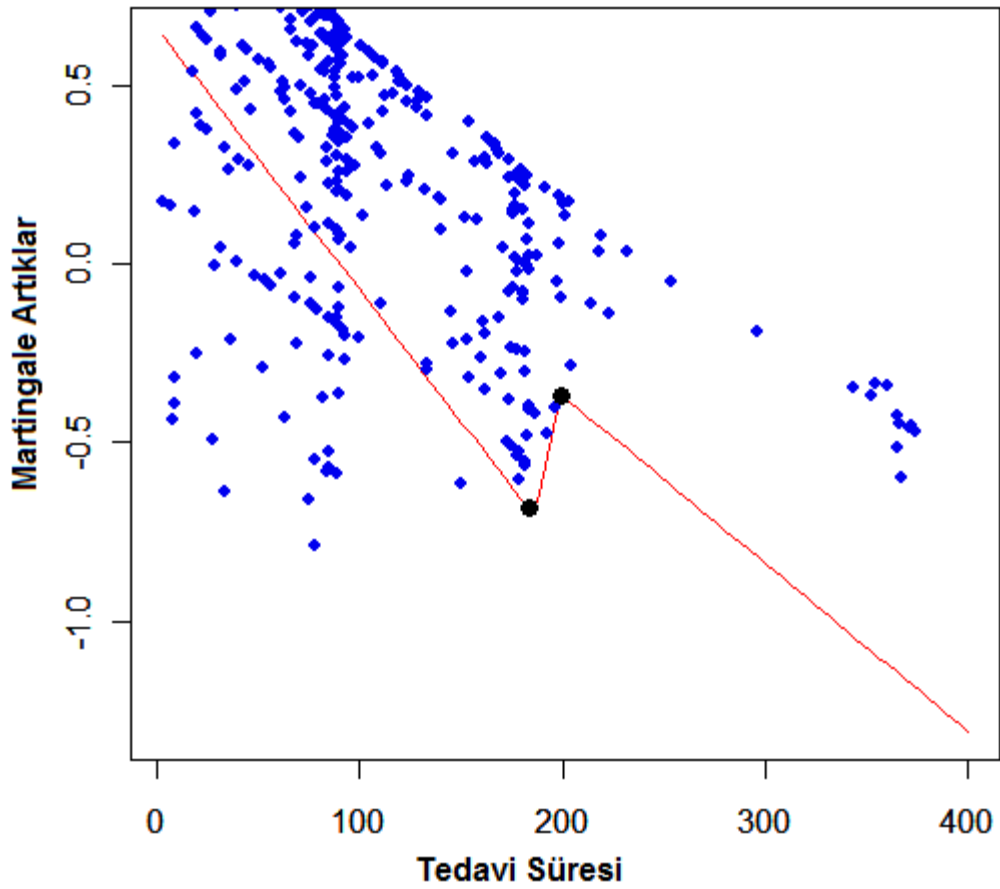
MARS_M modeli, klasik Cox-PH final modelinden farklı olarak uyuşturucu kullanımına dönme riski (hazard) üzerinde; Beck depresyon skoru ve tedavi süresi değişkenlerinin farklı kesim noktalarının etkili olduğunu göstermektedir.

Uyuşturucu kullanımına dönme riski üzerinde etkili olan Beck depresyon skoru (Bectota) ile Martingale artıkları (Risk) arasındaki ilişki doğrusal değildir (Şekil 3.2.3.1). Beck depresyon skoru 38.5'e kadar uyuşturucuya dönme riskini artırırken, 38.5 değerinden sonra uyuşturucuya dönme riskini hızla azaltmaktadır.



Şekil 3.2.3.1. Beck depresyon skoru (Bectota) ile Martingale artıkları arasındaki ilişki

Uyuřturucu kullanımına dönme riski üzerinde etkili olan tedavi süresi (Lot) ile Martingale artıkları (Risk) arasındaki ilişki de doğrusal değildir (Şekil 3.2.3.2). Tedavi süresi 186 güne kadar uyuřturucuya dönme riskini azaltırken, 186 – 199 gün arasında uyuřturucuya dönme riskini arttırmakta ve 199 günden sonra tekrar uyuřturucuya dönme riskini azaltıcı etki göstermektedir.



Şekil 3.2.3.2. Tedavi süresi (Lot) ile Martingale artıkları arasındaki ilişki

Bectota ve Lot değişkenleri $MARS_M$ tarafından belirlenen kesim noktalarına göre kodlanarak kategorik değişkenler haline dönüştürüldü. Martingale artıklarını bağımlı değişken olarak kullanan MARS modeli ile elde edilen bilgiler ile Cox-PH modeli yeniden çalıştırıldı. $MARS_M$ destekli Cox-PH model sonuçları Çizelge 3.2.3.2’de gösterilmiştir.

Çizelge 3.2.3.2. *UIS veri seti: Cox-PH modeli değişkenleri ve modele katkıları ($MARS_M$)*

Değişkenler	B	S_h	Wald	p	HR	HR %95 GA	
						Alt	Üst
Age	-0.025	0.008	9.256	0.002	0.976	0.960	0.991
IVhx			9.001	0.011			
IVhx(Çok önce)	0.264	0.137	3.692	0.055	1.302	0.995	1.703
IVhx(Y. zamanda)	0.344	0.117	8.707	0.003	1.411	1.123	1.774
Ndrugtx	0.024	0.008	8.240	0.004	1.024	1.008	1.041
Race	0.288	0.114	6.342	0.012	1.334	1.066	1.669
Lot			20.668	<0.001			
Lot (<186)	0.829	0.217	14.598	<0.001	2.291	1.497	3.506
Lot (186 – 199)	-0.126	0.411	0.094	0.759	0.882	0.394	1.974

S_h : Standart hata -2LL=5264.834

$MARS_M$ destekli Cox-PH modeli detaylı olarak incelendiğinde; $MARS_M$ tarafından belirlenen değişkenlerden Beckettota dışındaki tüm değişkenlerin Cox-PH modeline anlamlı katkıda buldukları ve modelde yer aldıkları görülmektedir. Cox-PH final sonucuna benzer şekilde, düşük yaşların yüksek yaşlara göre uyuşturucu kullanımına dönme riskini azaltıcı (süreyi arttırıcı) yönde etkili olduğu, daha önce aldığı uyuşturucu tedavi sayısının yüksek olmasının ise düşük olmasına göre uyuşturucuya dönme riskini arttırıcı (süreyi azaltıcı) etkisi olduğu belirlenmiştir.

Yakın zamanda damar yolundan uyuşturucu alanların, hiç almayanlara göre uyuşturucu kullanımına dönme riski HR=1.411 kat (%95 GA: 1.123 – 1.774) daha fazladır. Beyaz ırktan olanların diğer ırklara göre uyuşturucu

kullanımına dönme riski HR=1.334 kat (%95 GA: 1.066 – 1.669) daha fazladır.

Cox-PH final modelinden farklı olarak MARS_M destekli Cox-PH modeli teavi tipininin (Site) modele katkısını anlamlı bulmamıştır.

Cox-PH modelinden farklı olarak MARS_M tarafından kesim noktaları belirlenerek 3 sınıf halinde kodlanan Tedavi Süresi (Lot) değişkeninin uyuşturucu kullanımına dönme üzerindeki etkisi klasik Cox-PH modelinden çok daha çarpıcı bir biçimde ortaya çıkmıştır. Cox-PH final modeline göre; tedavi süresinin azalmasının artmasına göre uyuşturucu kullanımına dönme riskini azaltıcı yönde etkili olduğu bulunmuştur. Oysa MARS_M destekli Cox-PH modeline göre; 186 gün ve daha az süre tedavi gören bağımlıların uyuşturucuya dönmesi riski 199 gün ve daha fazla tedavi gören bağımlılara göre HR=2.291 kat (%95 GA: 1.497 – 3.506) daha fazla olarak belirlenmiştir. 186 – 199 gün arasında tedavi gören bağımlılar ile 199 günden fazla tedavi gören bağımlılar arasında uyuşturucuya dönme riski açısından istatistiksel olarak anlamlı bir risk farkı yoktur.

MARS_M destekli Cox-PH modelinin -2LL değeri 5264.834 ile Cox-PH final modeli değerinden (-2LL=5140.293) istatistiksel olarak da anlamlı miktarda daha yüksektir ($\chi^2=124.541$; Sd=1; p<0.001). MARS_M, Cox-PH modelinin -2LL değerini arttırmış ancak klasik Cox-PH modeli tarafından doğrusal olduğu kabul edilen hazardın aslında (veride gizli olan) bazı değerlerde doğrusal olmadığını göstermiştir.

3.2.4. UIS: MARS modeli sonuçları (Deviance)

Deviance artıklarını bağımlı değişken olarak kullanan MARS modeli (MARS_D) ile UIS verileri analiz edildiğinde, modelin açıklayıcılık değeri $R^2=0.409$ (düzeltilmiş $R^2=0.333$) olarak belirlenmiştir. MARS_D, en iyi sonuca ulaşmak için toplam 80 farklı model oluşturmuştur. Oluşturulan modellerin GCV değerleri 0.9116 ile 1.4578 arasında değişmektedir. En sade modelde sadece 1 temel fonksiyon (bu temel fonksiyon sadece model sabit katsayısını içermektedir) bulunurken, en karmaşık model toplam 45 temel fonksiyondan (bağımsız değişkenler ve bağımsız değişkenlerin farklı kesim noktalarını da içeren) oluşmaktadır. En uygun model, GCV değeri en düşük olan (ideal) modeldir. İdeal model 64 numaralı model olup, modele ilişkin GCV=0.9116 olarak belirlenmiştir. İdeal model toplam 17 adet temel fonksiyondan oluşmaktadır. İdeal modelde yer alan temel fonksiyon (BF) öğeleri, bu öğelere ilişkin katsayılar ve katsayıların standart hatası Çizelge 3.2.4.1'de verilmiştir.

Çizelge 3.2.4.1. UIS veri seti: İdeal modele ilişkin temel fonksiyonlar (MARS_D)

BF No	Değişken Adı	Kesim noktası	B	S _n
1	Model Sabiti	-	2.3065	0.4167
2	Age	-	-0.0223	0.0118
3	Age	35.0	0.0718	0.0388
4	Age	40.0	-0.1357	0.0552
5	Bectota	-	0.0065	0.0045
6	Bectota	37.8	-0.1229	0.0548
7	IVhx	-	0.1380	0.0509
8	Ndrugtx	-	0.0230	0.0076
9	Race	-	-0.2436	0.0944
10	Site	-	0.2052	0.0970
11	Lot	-	-0.0496	0.0103
12	Lot	30.0	0.0611	0.0197
13	Lot	48.0	-0.0306	0.0141

BF: Temel fonksiyon **S_n: Standart hata**

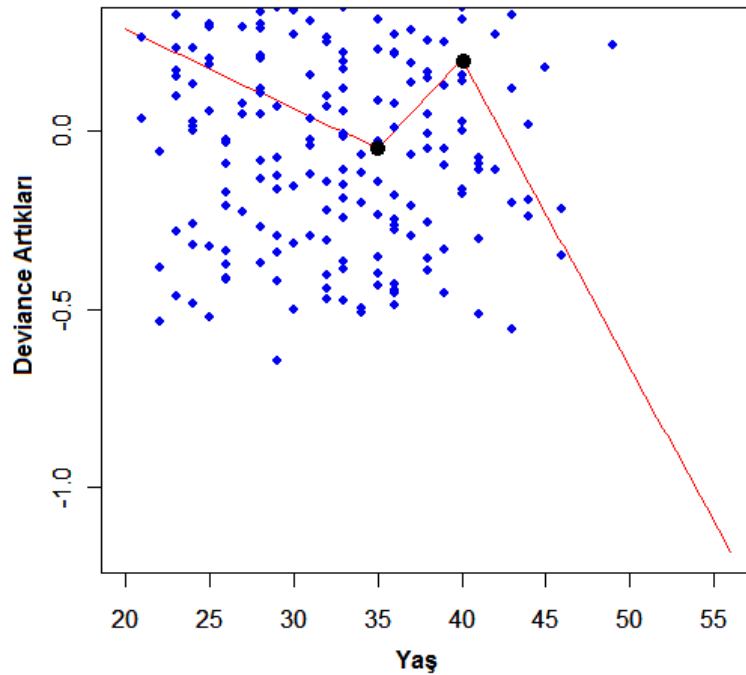
Çizelge 3.2.4.1. (Devam) UIS veri seti: İdeal modele ilişkin temel fonksiyonlar (MARS_D)

BF No	Değişken Adı	Kesim noktası	B	S _n
14	Lot	107.0	0.0482	0.0197
15	Lot	121.0	-0.0434	0.0211
16	Lot	186.0	0.0453	0.0235
17	Lot	199.0	-0.0356	0.0231

BF: Temel fonksiyon **S_n: Standart hata**

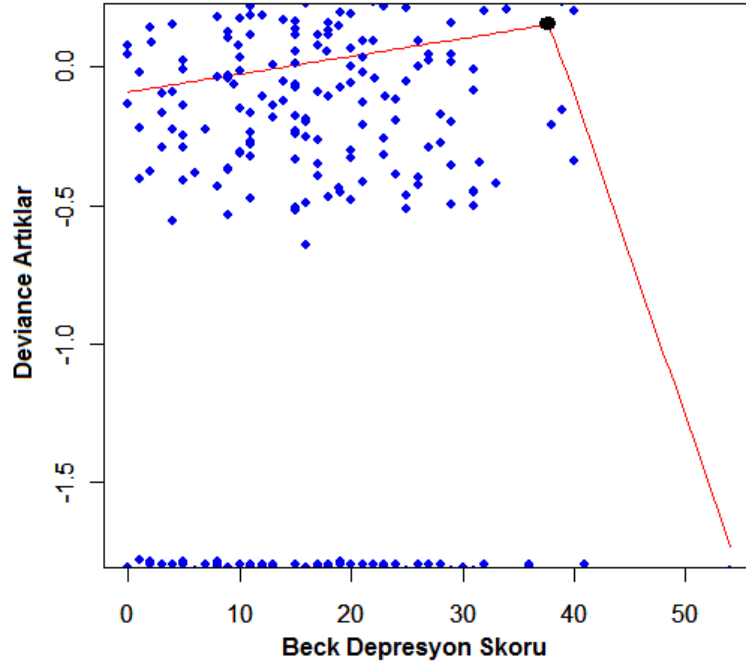
MARS_D modeli klasik Cox-PH final modelinden farklı olarak uyuşturucu kullanımına dönme riski üzerinde; Yaş (Age; <35, 35 – 39, 40+), Beck depresyon skoru (Becktotu; <37.8, 37.8+) ve tedavi süresi (Lot; <30, 30 – 47, 48 – 106, 107 – 120, 121 – 185, 186 – 198, 199+) değişken ve temel fonksiyonlarının da yer alması gerektiğini öngörmektedir.

Uyuşturucu kullanımına dönüş üzerinde etkili olan yaş (Age) değişkeni ile Deviance artıkları arasındaki ilişki doğrusal değildir (Şekil 3.2.4.1). 35 yaşa kadar uyuşturucuya dönme riski azalırken, 35 – 40 yaş aralığında risk artmakta, 40 yaşından sonra tekrar hızlı bir düşüş göstermektedir.



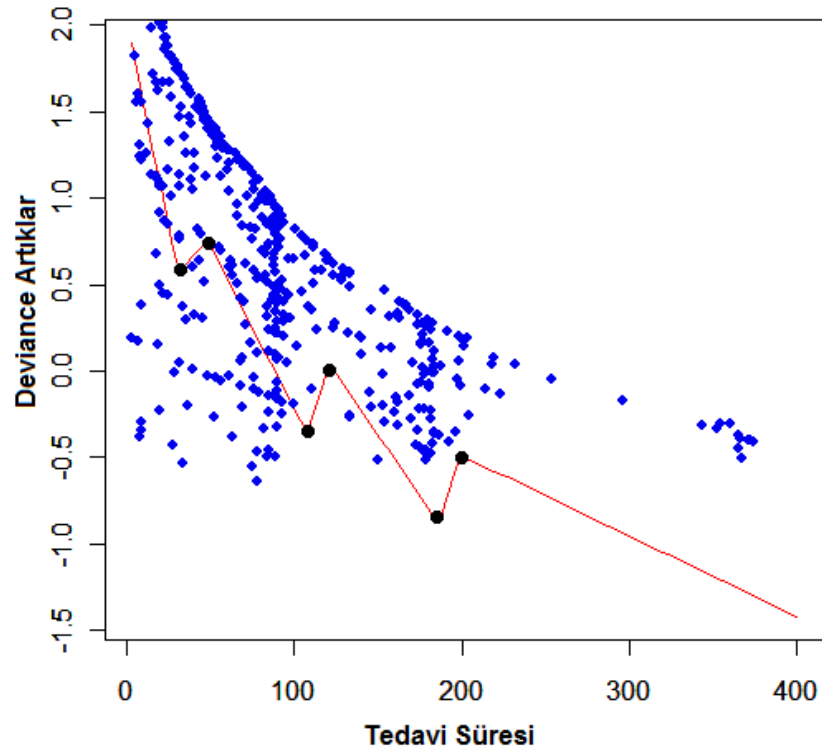
Şekil 3.2.4.1. Yaş (Age) ile Deviance artıkları arasındaki ilişki

Uyuşturucu kullanımına dönüş üzerinde etkili olan Beck depresyon skoru (Becktota) değişkeni ile Deviance artıkları arasındaki ilişki doğrusal değildir (Şekil 3.2.4.2). Beck depresyon skoru 37.8'e kadar risk artmakta, 37.8'den sonra hızlı bir düşüş göstermektedir.



Şekil 3.2.4.2. Beck depresyon skoru (Becktota) ile Deviance artıkları arasındaki ilişki

Uyuşturucu kullanımına dönüş üzerinde etkili olan tedavi süresi değişkeni ile Deviance artıkları arasındaki ilişki doğrusal değildir (Şekil 3.2.4.3). $MARS_D$ modeli uyuşturucu kullanımına dönme üzerinde etkili olan tedavi süresinin 6 farklı kesim noktası olduğunu tespit etmiştir. Tedavi süresi 30. güne kadar olduğunda riski azaltıcı etkisi vardır. 30. gün ile 48. gün arasında risk bir miktar yükselmekte, 48. günden sonra 107. güne kadar riskte azalma görülmektedir. 107. gün ile 121. gün arasında risk tekrar bir miktar yükselmekte, 121. günden 186. güne kadar yine düşüş eğilimine girmektedir. 186. gün ile 199. gün arasında risk artışı gözlenirken 199. günden sonraki tedavi süreleri uyuşturucu kullanımına dönüş riskini azaltmaktadır.



Şekil 3.2.4.3. Tedavi süresi (Lot) ile Deviance artıkları arasındaki ilişki

MARS_D tarafından risk üzerinde etkisi olduğu belirlenen değişkenlerden Yaş (Age; <35, 35 – 39, 40+), Beck depresyon skoru (Becktota; <37.8, 37.8+) ve tedavi süresi (Lot; <30, 30 – 47, 48 – 106, 107 – 120, 121 – 185, 186 – 198, 199+) değişkenleri kategori değişkenler olarak kodlandı. MARS_D destekli Cox-PH modeli yeniden çalıştırıldı. Elde edilen sonuçlar Çizelge 3.2.4.2’de verilmiştir.

Çizelge 3.2.4.2. UIS veri seti: Cox-PH modeli değişkenleri ve modele katkıları (MARS_D)

Değişkenler	B	S _h	Wald	p	HR	HR %95 GA	
						Alt	Üst
IVhx			13.977	0.001			
IVhx (Çok önce)	0.095	0.141	0.453	0.501	1.099	0.834	1.449
IVhx (Y. zamanda)	0.427	0.119	12.830	<0.001	1.532	1.213	1.935
Ndrugtx	0.027	0.009	10.205	0.001	1.028	1.011	1.045
Race	0.293	0.115	6.484	0.011	1.340	1.070	1.679
Site	0.350	0.118	8.842	0.003	1.419	1.127	1.787
Age			7.564	0.023			
Age (35 – 39)	-0.053	0.117	0.207	0.649	0.948	0.753	1.193
Age (40+)	-0.425	0.156	7.478	0.006	0.654	0.482	0.886
Lot			156.372	<0.001			
Lot (<30)	1.998	0.250	63.839	<0.001	7.374	4.517	12.039
Lot (31 – 47)	1.715	0.268	41.103	<0.001	5.559	3.290	9.393
Lot (48 – 106)	1.137	0.243	21.929	<0.001	3.118	1.937	5.018
Lot (107 – 120)	0.641	0.347	3.415	0.065	1.898	0.962	3.747
Lot (121 – 185)	0.376	0.245	2.345	0.126	1.456	0.900	2.354
Lot (186 – 198)	0.029	0.414	0.005	0.944	1.030	0.457	2.318

S_h: Standart hata -2LL=5140.394

MARS_D destekli Cox-PH modeli ile Cox-PH final modeli -2LL değerleri arasındaki farklılık istatistiksel olarak önemli değildir ($\chi^2=0.101$; Sd=6; p=1.000), modeller benzerdir.

Deviance artıklarını bağımlı değişken olarak kullanan MARS modeli ile Risk üzerinde anlamlı etkiye sahip olduğu belirlenen değişkenlerin tamamı Cox-PH modeli tarafından da modeli katkısı anlamlı olarak değerlendirilmiştir.

Yakın zamanda damar yolundan uyuşturucu alanların, hiç almayanlara göre uyuşturucu kullanımına dönme riski HR=1.532 kat (%95 GA: 1.213 – 1.935) daha fazladır. Beyaz ırktan olanların diğer ırklara göre uyuşturucu kullanımına dönme riski HR=1.340 kat (%95 GA: 1.070 – 1.679),

B tipi tedavi görenlerin uyuşturucuya dönme riski A tipi tedavi görenlere göre HR=1.419 kat (%95 GA: 1.127 – 1.787) daha fazladır.

40 Yaş ve üzerinde olan bağımlıların uyuşturucuya dönme riski 35 yaşından küçük bağımlılara göre %65.4 daha azdır. 35 yaşından küçük bağımlıların 35 – 39 yaş arasındaki bağımlıların uyuşturucuya dönme riskleri benzerdir.

Cox-PH modelinden farklı olarak MARS_D tarafından kesim noktaları belirlenerek 7 sınıf halinde kodlanan Tedavi Süresi (Lot) değişkeninin uyuşturucu kullanımına dönme üzerindeki etkisi klasik Cox-PH modelinden çok daha çarpıcı bir biçimde ortaya çıkmıştır. Cox-PH final modeline göre; tedavi süresinin azalmasının artmasına göre uyuşturucu kullanımına dönme riskini azaltıcı yönde etkili olduğu bulunmuştur. Oysa MARS_D destekli Cox-PH modeline göre; 199 günden fazla tedavi görenlere göre uyuşturucuya dönme riski; 30 günden az tedavi görenlerde HR=7.374 kat, 31 – 47 gün tedavi görenlerde HR=5.559 kat ve 48 – 106 gün tedavi görenlerde ise HR=3.118 kat daha fazladır. 107 gün ve daha fazla tedavi gören bağımlıların 199 gün ve daha fazla tedavi gören bağımlıların uyuşturucuya dönme riskleri benzerdir.

MARS_D Cox-PH modelinin -2LL değerini arttırmamış ve MARS_M modeli gibi klasik Cox-PH modeli tarafından doğrusal olduğu kabul edilen hazardın aslında (veride gizli olan) bazı değerlerde doğrusal olmadığını göstermiştir.

4. TARTIŞMA

Gerçek sağkalım verileri (Karsinom) üzerinde yapılan çalışmada; yaşam çözümlmesine uyarlanmış MARS modelinin, Cox-PH modelinin Martingale ve Deviance artıklarını kullanarak Cox-PH modeli ile belirlenemeyen kesim noktalarını belirleyebildiği, hazard ile bağımsız değişkenler arasındaki ilişkiyi daha açık ve kolay yorumlanabilecek biçimde ortaya çıkarabildiği gösterilmiştir. $MARS_M$ ve $MARS_D$ ile modelde yer alması gereken değişkenler ve kesim noktaları belirlenebilmektedir. Belirlenen kesim noktalarına göre oluşturulan yeni değişkenler ile kurulan sağkalım modellerinin yorumlanması kolaylaşmakta ve anlaşılabilirliği artmaktadır. Yaşam çözümlmesine uyarlanmış MARS yöntemi ile kurulan modellerin çoğunda Cox-PH modelinin -2LL değerininin klasik Cox-PH modeline göre istatistiksel olarak anlamlı miktarda artmadığı da gösterilmiştir.

Bu tez çalışmasında kullanılan Karsinom verileri ile yapılan orijinal araştırmalarında Göğüş ve arkadaşları (2004) Platelet değeri 400000 mm^3 ve üzerinde olan hastalarını Trombositoz olarak tanımlamış ve sağkalım analizlerini ona göre yapmışlardır. Geliştirilen MARS modeline göre, Platelet değeri için kesim noktası 468000 mm^3 olarak alınmalıdır. Araştırmacılar, Trombositoz olan hastaların ölüm hızının Trombositoz olmayan hastalara göre HR=4.72 kat daha fazla olduğu bulmuşlardır. Ancak MARS modeline göre Platelet değeri 218000 mm^3 'e kadar ölüm riskini azaltıcı, 218000 'den itibaren 467000 'e kadar ölüm riskini çok az arttırıcı etki göstermektedir. Platelet değeri 468000 'den itibaren ise risk artışı kuvvetlenmektedir. Ayrıca Tümör 3-4 grubunda olan hastalar Tümör 1-2 grubuna göre HR=5.979 kat daha fazla ölüm riskine sahiptir.

Sağkalım literatüründe yer alan bir veri seti (UMARU – UIS) üzerinde de yaşam çözümlmesine uyarlanmış MARS modeli sınanmıştır. Cox-null

modelinin Martingale ve Deviance artıklarını bağımlı değişken olarak kullanan MARS modelinin, sağkalım üzerinde etkili olan değişkenleri ve sürekli değişkenler için kesim noktalarını belirleyebildiği gösterilmiştir.

UIS veri setini çalışmalarında kullanan Hosmer ve Lemeshow (1999), veri setinde yer alan Beck depresyon skoru (Becktota) değişkeninin modelde yer alması gerektiğini buna karşın tedavi süresi (Lot) değişkeninin modele anlamlı bir katkısının olmadığını bulmuşlardır. Geliştirilen MARS modeli ile yapılan çözümler sonucunda, yaş (Age), Beck depresyon skoru (Becktota), damar içi uyuşturucu kullanım hikayesi (IVhx), daha önce aldığı uyuşturucu tedavisi sayısı (Ndrugtx), ırk (Race) ve tedavi türü (Site) değişkenleri ile tedavi süresi (Lot) değişkeninin uyuşturucuya yeniden başlama modelinde anlamlı bir etkisinin olduğu görülmüştür. Klasik çözümlerden farklı olarak MARS ile yaş, tedavi süresi ve Beck depresyon skoru için farklı kesim noktaları belirlenerek bu noktalar için riskte gözlenen değişimler incelenmiştir.

Çalışmada kullanılan Karsinom ve UIS verileri ile yapılan analizler sonucunda MARS yönteminin yaşam çözümlemesi için bilinen klasik yöntemlere karşı kuvvetli bir alternatif, Cox-PH modelinin tamamlayıcısı olarak yararlı bir yöntem olabileceği görülmüştür. Ancak uygulamalar sırasında karşılaşılan bazı kısıtlılıkları da vurgulamak gerekir. MARS'ın uygulanabilmesi (şu an için) veride kayıp gözlem olmamasına bağlıdır. MARS yapısı itibari ile kategorik değişkenlere göre sıralı ya da sürekli değişkenlerde daha başarılı sonuçlar verebilmektedir.

Elde edilen sonuçların sadece uygulamada kullanılan[§] veri setleri üzerinde denendiği ve onlar için geçerli olduğu unutulmamalıdır.

§ UIS veri seti için Hosmer ve Lemeshow (1999) kitaplarında (sayfa 22 – 24) orijinal veri setinin sadece belirli bir kısmını kullanıma açtıklarını, bu veri setinde bazı değişkenlerin ve değişken değerlerinin orijinal veri setinden bir miktar farklı olduğunu ifade etmektedirler.

5. SONUÇ VE ÖNERİLER

Bağımsız değişken kullanmayan Cox regresyonun (Cox–null model) Martingale ve Deviance artıklarını inceleyerek, sağkalım üzerinde etkili olan bağımsız değişkenleri ve bu değişkenlerin kesim noktalarını belirlemek amacıyla MARS modeli yaşam çözümlemesine uyarlanmıştır.

Yaşam çözümlemesine uyarlanmış MARS modeli ile sağkalım üzerinde etkili olan bağımsız değişkenler ile sürekli bağımsız değişkenler için olası kesim noktaları belirlenmiş ve bu kesim noktalarının kurulacak yeni Cox-PH modellerinde kullanılarak modelin açıklayıcılığını arttırıcı ve yorumlanmasını kolaylaştırıcı bir yöntem olarak kullanılması önerilmiştir.

Gerçek sağkalım verileri (Karsinom ve UIS) üzerinde yapılan çalışmalar ile yaşam çözümlemesine uyarlanmış MARS modelinin sağkalım üzerinde etkili olan bağımsız değişkenleri ve bunların kesim noktalarını klasik Cox-PH yaklaşımından daha iyi tespit edebildiği görülmüştür. Bu sayede klasik Cox-PH yönteminin başarısı arttırılabilmektedir.

MARS modelinin en önemli avantajlarından biri sürekli dağılım gösteren bağımsız değişkenlerde kesim noktalarını belirleyebilmesi olmuştur. Klasik Cox-PH yönteminde; kategorik değişkenler tüm kategorileri ile ve sürekli dağılım gösteren değişkenler ise oldukları gibi modele alınmakta ve sonuçta oluşan sağkalım modelini yorumlamakta bazen güçlüklerle karşılaşmaktadır. Örneğin yaştaki 10 birimlik artış ya da Hemoglobin düzeyindeki 3 birimlik azalış riski arttırmaktadır yorumu yapılmaktadır. Ancak bu artışların ya da azalışların etkilerinin doğrusal olduğu kabul edilmektedir. Oysa gerçek hayatta 30 yaşın üzerine eklenecek 10 yaş ile 60 yaşın üzerine eklenecek 10 yaş oldukça farklı etkilere sahiptir. Bu nedenle Cox-PH yönteminin bu eksikliğini kapatmak amacıyla yaşam çözümlemesine uyarlanmış MARS modeli kullanılabilir.

MARS yönteminin sağkalım uyarlamasını kullanmak için veride kayıp gözlem olmadığından emin olunmalı, aykırı değerler temizlenmeli ya da uygun yöntemlerle düzeltilmelidir.

MARS, algoritması sayesinde özellikle büyük veri setlerinde değişkenler arasındaki ilişkileri belirlemede, veride gizli olan değişken etkileşimlerini ortaya çıkarmada ve sonuç değişkenini en iyi tahmin edecek modeli kurmada bilinen yaşam çözümlmelerine göre daha başarılı bir yöntemdir. Ancak yine algoritması gereğince kullanılacak veri setinin 100'den fazla gözleme sahip olması daha başarılı modeller kurulabilmesi ve tutarlı sonuçlara ulaşılabilmesi açısından faydalıdır.

MARS'ın yaşam çözümlemesi alanında kullanılması tüm dünyada henüz yeni bir uygulamadır. Bu nedenle yapılacak araştırma ve çalışmalarda farklı dağılımlara sahip veri türleri için, verilerdeki kayıp gözlem sorunlarını aşabilecek ve aşırı/aykırı değer alan gözlemlerden kaynaklanan sorunları giderebilecek MARS yaklaşımlarının geliştirilip denenmesi yararlı olacaktır.

Bu tez çalışması kapsamında sadece sağdan sansürlü sağkalım verileri üzerinde MARS yöntemi sınanmıştır. Diğer sansür türlerinde de (soldan, ortadan ya da aralıklı) MARS yönteminin sınanması gereklidir.

Yaşam çözümlemesine uyarlanan MARS modeli henüz tam olarak olgunlaşmış bir model değildir. Hala eksiklikleri ve doğal olarak hataları vardır. Örnek uygulamalarda Martingale ve Deviance artıkları sonuç değişkeni olarak kullanılarak hazard üzerinde etkiye sahip bağımsız değişkenler ve bunların kesim noktaları belirlenmeye çalışılmıştır. Bu konuda yapılacak bundan sonraki çalışmalarda Deviance ve Martingale dışındaki artıkların da bu amaçla kullanılabilirliğinin araştırılması faydalı olacaktır.

MARS yönteminin yaşam çözümlemesinde kullanılmasına yönelik (henüz) hazır istatistik paket programı yoktur. Ticari olarak piyasada bulunan çok az sayıda paket program (Salford-System MARS ve Statistica Ver. 7.0'dan itibaren MARspline modülü) ise MARS yönteminin genelleştirilmiş algoritmasını kullanmaktadır. Matlab için geliştirilmiş MARS uygulamaları da mevcuttur. Açık kaynak kodlu R'de ise sadece MARS ve MARS benzeri uyarlanabilir regresyon yöntemlerini kullanmak için yazılmış bulunan pek çok fonksiyon ve prosedür (Ör: earth, polyspline, bruto, polymars, polySpline, mda gibi) mevcuttur. Bu nedenle araştırmacıların hazır paket programları kullanmanın yanı sıra R gibi açık kaynak kodlu uygulamaları da kullanabilecek eğitimleri almaları ya da kendilerini bu yönde geliştirmeleri yararlı olacaktır.

ÖZET

Yaşam Çözümlemesine Yeni bir Yaklaşım: MARS

Bu tez çalışmasında; Çok değişkenli uyarlanabilir regresyon uzanımlarının (MARS) yaşam çözümlemesinde kullanılıp kullanılmayacağı araştırılmıştır. Bu kapsamda; MARS yöntemi kısaca tanıtılmış, yaşam çözümlemesine uyarlaması matematiksel altyapısı ile birlikte gösterilmiştir.

MARS yöntemini yaşam çözümlemesine uyarlayabilmek amacı ile bağımsız değişken kullanmayan Cox modelinin Martingale ve Deviance artıkları incelenmiştir. Martingale ve Deviance artıklar MARS modelinin bağımlı değişkeni olarak alınmış, bu sayede takip süresi ve sansür değişkenlerini birleştiren ve hazardı temsil eden tek bir sürekli değişken üzerinden MARS modeli kurulmuştur.

MARS modelinin uygulanabilirliğini göstermek amacı ile Karsinom veri seti üzerinde çalışılmıştır. Karsinom veri setinde; risk üzerinde etkiye sahip değişkenler belirlenmiş, Platelet için farklı kesim noktaları bulunarak elde edilen sonuçlar klasik Cox-PH modelinin başarısını arttırmak için kullanılmıştır.

Yaşam çözümlemesi ile ilgili kaynaklarda oldukça sık karşılaşılan UMARU – UIS veri seti üzerinde tez çalışmasında tanıtılan yöntem uygulanmış ve MARS yönteminin klasik Cox-PH modeline büyük katkılar sağlayabildiği gösterilmiştir.

Anahtar Kelimeler: Cox-PH modeli, Çok değişkenli uyarlanabilir regresyon uzanımları, Deviance artıklar, MARS, Martingale artıklar

SUMMARY

A New Approach to Survival Analysis: MARS

In this dissertation study it was investigated that whether Multivariate adaptive regression splines (MARS) can be used in survival analysis or not. In this context, MARS method was introduced briefly and its application to survival analysis was shown with its mathematical infrastructure.

In order to adapt the MARS method to survival analysis, Martingale and Deviance residuals of Cox model does not use independent variable was investigated. Martingale and Deviance residuals were assumed as dependent variables of MARS method, so that it was tried to construct a model by using an independent variable which represents hazard and combines the censored variables.

To show the applicability of MARS model, cancer data were used. In this carcinoma data set the variables having risk effect on variables was determined and the results obtained for different cut points for Platelet were used to improve the success of classical Cox-PH model.

The method introduced in this thesis study was applied on UMARU-UIS data set which is frequently come across with the references related to survival analysis and it was shown that MARS model can provide more contributions to classical Cox-PH model.

Key Words: Cox-PH model, Deviance residuals, MARS, Martingale residuals, Multivariate adaptive regression splines

KAYNAKLAR

- ABRAHAM, A., STEINBERG, D. (2001). MARS: Still an Alien Planet in Soft Computing?, *Lecture Notes in Computer Science*; **2074**: 235-244
- ALPAR, R. (2003). Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1, Nobel yayın dağıtım, Ankara
- AUSTIN, P.C. (2007). A comparison of regression trees, logistic regression, generalized additive models and multivariate adaptive regression splines for predicting AMI mortality, *Statistics in Medicine*, **26**(15):2937-57
- BERDER, R., AUGUSTIN, T. and BLETTNER, M. (2005). Generating survival times to Cox proportional hazards models, *Statistics in Medicine*, **24**:1713-1723
- BREIMAN L., (1991). Discussion: The π Method for Estimating Multivariate Functions From Noisy Data, *Technometrics*, **33**:145-148
- BRIAND L.C., FREIMUT, B., VOLLEI, F., (2004), Using multiple adaptive regression splines to support decision making in code inspections, *Journal of Systems and Software*, **73** (2):205-217
- CHEN, V.C.P., GÜNTHER, D., JOHNSON, E.L. (2003). Solving for an optimal airline yield management policy via statistical learning, *Appl.Statist.*, Vol. **52** (1): 19-30
- COX, D.R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, **34**:187-220
- COX, D.R., OAKES, D (1984). Analysis of Survival Data, Chapman and Hall
- DEICHMAN, J., ESHGI, A., HAUGHTON, D., SAYEK, S., TEEBAGY, N. (2002). Application of Multiple Adaptive Regression Splines (MARS) in Direct Response Modelling, *Journal of Interactive Marketing*, **16**:15-27
- DIŞCI R. (2008). Temel ve Klinik Biyoistatistik, İstanbul Tıp Kitapevi, İstanbul
- ELECTRONIC STATISTICS TEXTBOOK. (2010). StatSoft Inc. Tulsa, USA
Erişim: [<http://www.statsoft.com/textbook/>]
Erişim Tarihi: 21.07.2010
- FLEMING, T.R. and HARRINGTON, D.P. (1991). Counting Processes and Survival Analysis, Wiles-Interscience
- FRIEDMAN, J.H. (1991). Multivariate Adaptive Regression Splines, *Annals of Statistics*, **19** (1): 1-67
- GÖĞÜŞ Ç., BALTACI S., FILİZ E., ELHAN A., BEDÜK Y., (2004), Significance of Thtombocytosis for Determining Prognosis in Patients with Localized Renal Cell Carcinoma, *Urology*, **63** (3): 447- 450

- HOSMER, D.W., LEMESHOW S.,(1999), Applied Survival Analysis: Regresyon Modelling of Time to Event Data, John Wiley & Sons Inc.
- IBRAHİM J.G., CHEN M.H., SINHA D. (2005). Bayesian Survival Analysis, Springer, 2nd Ed.
- KAKI, B.,YEŞİLOVA, A., ŞEN, C. (2004), Yarı Parametrik Regresyon Yönteminin Hayvancılıkta Kullanılması, 4. Ulusal Zootekni Bilim Kongresi Sözlü Bildiriler Programı,
Erişim: [http://4uzbk.sdu.edu.tr/4UZBK/BGB/4UZBK_097.pdf]
Erişim Tarihi: 16.03.2009
- KARABULUT, E., ALPAR, R., BAYKAL, C. (2000). COX Regresyonda artıkların analizi, Biyoistatistik 2000 5. Ulusal Biyoistatistik Kongresi Bildiri Kitabı, 325-334
- KELEŞ, S., SEGAL, M.R. (2002). Residual-based tree structured survival analysis, *Statistics in Medicine*, **21**:313-326
- KOLYSHKINA, I., SYLVIA, W. (2004). Enhancing Generalised Linear models with Data Mining, Casualty Actuarial Society Discussion Paper Program, *Casualty Actuarial Society* – Arlington, Virginia, 279-290
Erişim: [www.casact.org/pubs/dpp/dpp04/04dpp279.pdf],
Erişim Tarihi: 20.02.2009
- KOOPERBERG, C., STONE, C.J. & TRUONG, Y.K. (1995) Hazard regression, *Journal of American Statistical Association*, **92** (429): 78-94
- KOOPERBERG, C., BOSE, S., STONE, C.J. (1997) Polychotomous regression, *Journal of American Statistical Association*, **92** (437):117-127
- KÖSE, S.K. (1997), Cox Regresyon Modeli ile Larenks Kanserli Hastalarda Yaşam Sürelerine Etki Eden Faktörlerin İncelenmesi Üzerine Bir Araştırma, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik AD. Yayınlanmamış Doktora Tezi, Ankara
- KRINER, M. (2007), Survival Analysis with Multivariate Adaptive Regression Splines, Yayınlanmamış doktora tezi, Almanya
Erişim: [edoc.ub.uni-muenchen.de/7369/1/Kriner_Monika.pdf]
Erişim Tarihi: 16.07.2008
- LEATHWICK, J.R., ELITH, J. AND HASTIE, T. (2006). Comparative Performance Of Generalized Additive Models And Multivariate Adaptive Regression Splines For Statistical Modelling Of Species Distributions, *Ecological Modelling*, **199**:188–196
- MALLICK, B.K., DENISON, D.G.T. AND SMITH, A.F.M. (1999). Bayesian Survival Analysis using a MARS Model, *Biometrics*, **55**:1071-1077
- MARS USER GUIDE © Salford Systems,
Erişim: [<http://www.salfordsystems.com/424.php>],
Erişim Tarihi: 13.02.2007

- MARUBINI, E., VALSECCHI, M.G. (1984). *Analysing Survival Data from Clinical Trials and Observational Studies*, Jhon Wiley & Sons Ltd.
- MUKKAMALA, S., SUNG, A.H., ABRAHAM, A., (2004), *Intrusion detection systems using adaptive regression splines*,
Eriřim: [<http://www.salford-systems.com/doc/ICEIS-final.pdf>],
Eriřim Tarihi: 23.11.2009
- NCSS HELP SYSTEM, (2007). *NCSS 2007 Help System*, Hintze JL., Kayswille, UTAH
- ÖZDAMAR, K. (2005). *SPSS ile Biyoistatistik*, 5. Baskı, Kaan Kitabevi, Eskiřehir
- SARAÇBAŐI, O. (2002), *Hacettepe Üniversitesi Saęlık Bilimleri Enstitüsü Biyoistatistik AD. BİS 618 Yařam Çözümlemesi (Yüksek Lisans) dersi notları*, Ankara, 2002-2003 Güz Dönemi
- SAS/STAT USER GUIDE, (2010). *The PHREG Procedure*,
Eriřim: [<http://www.technion.ac.il/~docs/sas/stat/chap49/sect23.htm>],
Eriřim Tarihi: 19.03.2010
- SPSS KNOWLEDGEbase, (2010). *SPSS Inc., an IBM Company*,
Eriřim: [<http://support.spss.com/student/troubleshooting/sturesearchdetail.asp?ID=55067>],
Eriřim Tarihi: 13.08.2010
- STATISTICA USER GUIDE, (2009). *Multivariate Adaptive Regression Spline*,
Eriřim:[<http://www.statsoft.com/textbook/multivariate-adaptive-regression-splines>],
Eriřim Tarihi: 06.04.2009)
- TAYLOR, J. AND COBB, K. (2004), *Statistics 262: Intermediate Biostatistics, Modelling Hazard: Parametric Models*
Eriřim:[<http://www.stat.stanford.edu/~jtaylo/courses/stats262/spring.2004/notes/week6.pdf>],
Eriřim Tarihi: 19.03.2010
- TEMEL, G.O., ÇAMDEVİREN, H., YAZICI, A.C. (2005). *Regresyon modellerine alternatif bir yaklařım: MARS*, VIII. Ulusal Biyoistatistik kongresi Bildiri Kitabı, 105-123
- TERZİ, Y., CENGİZ, M.A., BEK, Y. (2005). *Cox Regresyon Modelinde Orantısal Hazard Varsayımının Artıklarla incelenmesi ve Akcięer Kanseri Hastaları Üzerinde Uygulanması*, *Türkiye Klinikleri J Med Sci*, **25**:770-775
- THERNEAU TM., GRAMBSCH PM., (2000). *Modeling Survival Data: Extending the Cox Model*, Springer
- VERZILLI, C.J., WHITTAKER, J.C., STALLARD, N., CHASMAN, D. (2005). *A Hierarchical Bayesian Model For Predicting The Functional Consequences Of Amino-Acid Polymorphisms*, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54** (1):191–206,
- VITTINGHOFF, E., GLIDDEN, D.V., SHIBOSKI, S.C., MCCULLOCH, C.E. (2004). *Regression Model in Biostatistics*, Springer

- WEI, L.J., LIN, D.Y., WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association*, **84**(408):1065-1073
- XIONG, R. AND MEULLENET, J.F., (2004). Application of Multivariate Adaptive Regression Splines (MARS) to the Preference Mapping of Cheese Sticks. *Journal of Food Science*, 69: SNQ131–SNQ139.
- YAY, M., ÇOKER, E., UYSAL, Ö. (2007). Yaşam Analizinde Cox Regresyon Modeli ve Artıkların incelenmesi, *Cerrahpaşa Tıp Dergisi*, **38**:139-145
- YORK, T.P., EAVES, L.J., OORD, E.J.C.G. (2006). Multivariate adaptive regression splines: a powerful method for detecting disease–risk relationship differences among subgroups, *Statist. Med.*, **25**:1355–1367

ÖZGEÇMİŞ

I. Kişisel Bilgiler :

Adı : Mesut
Soyadı : AKYOL
Doğum Yeri : Afyonkarahisar
Doğum Tarihi : 25.01.1972
Medeni Hali : Evli
Askerlik Durumu : Muaf
İletişim Bilgileri : GATA Biyoistatistik Bilim Dalı, 06110
Etlik – Ankara
Tel: 0.530. 886 26 71
e-posta: mesut.akyol@gmail.com

II. Eğitim Bilgileri :

Doktora : Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik AD.
(2006 – 2011)
Yüksek Lisans : Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik AD.
(2003 – 2005)
Lisans : Gazi Üniversitesi Mesleki Eğitim Fakültesi, Bilgisayar
Öğretmenliği AD.
(1988 – 1992)

III. Ünvanları ve Mesleki Deneyimi :

Bilgisayar Öğretmeni/Okutmanı : 1992 – 2005
Biyoistatistik Bilim Uzmanı : 2005 – 2011
Biyoistatistik Bilim Doktoru : 2011 –

IV. Üye Olduğu Bilimsel Kuruluşlar :

- Biyoistatistik Derneği
- International Biometrics Society (Doğu Akdeniz Bölgesi)

V. Bilimsel İlgili Alanları _____ :

ULUSLAR ARASI YAYINLAR

1. Aslan O, Sanisođlu Y, **Akyol M**, Yetkin S., "Quality of sleep in Turkish family caregivers of cancer patients", *Cancer Nurs.* 2009 Sep-Oct;32(5):370-7. PubMed PMID: 19661798.
2. Ozgurtas T, Aydin I, Turan O, Koc E, Hirfanoglu IM, Acikel CH, **Akyol M**, Erbil MK., "Vascular endothelial growth factor, basic fibroblast growth factor, insulin-like growth factor-I and platelet-derived growth factor levels in human milk of mothers with term and preterm neonates", *Cytokine.* 2010 May;50(2):192-4. Epub 2010 Mar 3. PubMed PMID: 20202860.
3. Baysan O, **Akyol M**, Bugan B, Yokuşođlu M, Gökođlan Y, Genç C., "Speckle tracking based myocardial velocities: our experience with novel software", *Anadolu Kardiyol Derg.* 2010 Jun;10(3):233-8. doi: 10.5152/akd.2010.063. PubMed PMID: 20538558.
4. Karacay O, Sepici-Dincel A, Karcaaltincaba D, Sahin D, Yalvaç S, **AKYOL M**, Kandemir O, Altan N., "A quantitative evaluation of total antioxidant status and oxidative stress markers in preeclampsia and gestational diabetic patients in 24-36 weeks of gestation", *Diabetes Res Clin Pract.* 2010 Sep;89(3):231-8. PubMed PMID: 20537747.
5. Erdemir U, Mumcu E, Topcu FT, Yildiz E, Yamanel K, **Akyol M**, "Micro push-out bond strengths of 2 fiber post types luted using different adhesive strategies", *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2010 Oct;110(4):534-44. PubMed PMID: 20868998.
6. Derbent A, Tatli MM, Duran M, Tonbul A, Kafali H, **Akyol M**, Turhan NO., "Transient tachypnea of the newborn: effects of labor and delivery type in term and preterm pregnancies", *Arch Gynecol Obstet.* 2010 Apr 30. [Epub ahead of print] PubMed PMID: 20431892.
7. Boran C, Kandirali E, Yilmaz F, Serin E, **Akyol M**, "Reliability of the 34βE12, keratin 5/6, p63, bcl-2, and AMACR in the diagnosis of prostate carcinoma", *Urol Oncol.* 2010 Feb 27. [Epub ahead of print] PubMed PMID: 20189848.
8. Ozdegirmenci O, Kucukozkan T, Akdag E, Topal T, Haberal A, Kayir H, Oter S., **Akyol M**, Uzbay T. "Effects of sildenafil and tadalafil on ischemia/reperfusion injury in fetal rat brain", *J Matern Fetal Neonatal Med.* 2010 Jun 14. [Epub ahead of print] PubMed PMID: 20540678.
9. Ozdegirmenci O, Dilbaz S, Cinar O, Aydin S, Beydilli G, Cakir L, Guven ES, **Akyol M**, Haberal A., "Can serum oestradiol be a predictor of quality of oocytes and embryos, maturation of oocytes and pregnancy rate in ICSI cycles?", *Gynecol Endocrinol.* 2010 Jun 14. [Epub ahead of print] PubMed PMID: 20540672.

10. Korkmaz C, Cinar O, **Akyol M.**, “The relationship between meiotic spindle imaging and outcome of intracytoplasmic sperm injection: a retrospective study”, *Gynecol Endocrinol.* 2010 Sep 9. [Epub ahead of print] PubMed PMID: 20828242.
11. Uysal T, Amasyali M, Ozcan S, Koyuturk AE, **Akyol M.** Sagdic D., “In vivo effects of amorphous calcium phosphate-containing orthodontic composite on enamel demineralization around orthodontic brackets”, *Aust Dent J.* 2010 Sep;55(3):285-91. PubMed PMID: 20887516.
12. Balcioglu HA, Kilic C, **Akyol M.** Ozan H, Kokten G., “Length of the styloid process and anatomical implications for Eagle's syndrome”, *Folia Morphol (Warsz).* 2009 Nov; 68(4):265-70. PubMed PMID: 19950078.
13. Ozen T, Orhan K, Avsever H, Tunca YM, Ulker AE, **Akyol M.**, “Dentin hypersensitivity: a randomized clinical comparison of three different agents in a short-term treatment period”, *Oper Dent.* 2009 Jul-Aug; 34(4):392-8. PubMed PMID: 19678443.
14. Bayer A, Erdem U, Mumcuoglu T, **Akyol M.**, “Two-site phacotrabecectomy versus bimanual microincision cataract surgery combined with trabeculectomy”, *Eur J Ophthalmol.* 2009 Jan-Feb; 19(1):46-54. PubMed PMID: 19123148.
15. Aydın İ, Özgürtaş T, Turan Ö, Koç E, Hırfanoğlu İM, Açikel CH, Garipoğlu G, **Akyol M.** Tunç T, Erbil MK., “Preterm Ve Term Yenidoğanların Anne Sütünün Biyokimyasal Karşılaştırması”, *Turkish Journal of Biochemistry -Türk Biyokimya Dergisi* 2009 34(4):242-249.
16. Okçu MK, Doğan N, Şençimen M, Korkmaz C, Altuğ HA, İde T, **Akyol M.**, “The Effect of a Hemostatic Agent (FastAct) to Wound and Tissue Repair in a Rat Model”, *Trakya Üniversitesi Tıp Fakültesi Dergisi* 2010; 27(3):221-226
17. Cakir E, Ersu R, Uyan ZS, Oktem S, Varol N, Karakoc F, Karadag B, **Akyol M.** Dagli E., “The prevalence and risk factors of asthma and allergic diseases among working adolescents”, *Asian Pacific Journal of Allergy & Immunology* 2010 Jun-Sep;28(2-3): 122-9.

ULUSLAR ARASI BİLDİRİLER

1. **Akyol M.**, Sanisoglu SY., Etikan İ., Alpar R., Akkus Z., “Evaluation of Three Year Efficiencies of Selected University Hospitals in Turkey by Dea-Malmquist Index Method”, 3rd EMR-IBS Conference. May 09-12, 2005, Corfu, Greece,
2. Sanisoglu SY., Etikan i., **Akyol M.**, Karakaya J., Geyik P., Ugurlu M., “Estimation of Biochemical Parameter Values At a Certain Time in Turkish Population by Using the Combination of Regression Trees and Bernstein Polynomials”, 3rd EMR-IBS Conference. May 09-12, 2005, Corfu, Greece,
3. **Akyol M.**, Alpar R, and Sanisoğlu SY, “Efficiency Analysis Using Data Envelopment Analysis (DEA) Technique Made for 28 Clinics in Gülhane Military Medical Academy 1000+200 Bed Training Hospital”, 13th International Biometric Conference, July 16-21 2006, Montreal, Canada
4. Sanisoglu SY, **Akyol M** and Saracbası O, “Comparing The Techniques for Sample Size Calculations in Wilcoxon Signed Ranks Test: A Simulation Study”, 13th International Biometric Conference, July 16-21 2006, Montreal, Canada
5. **Akyol M.**, Aslan FS, Sanisoğlu SY and Elhan AH, “Performance Estimation of Young Sportives With Mars”, 56th Session of the International Statistical Institute, August 23-28 2007, Lisboa, Portugal
6. Gulcan B, Nuran T, **Akyol M** and Akbayrak N, “The Usage of The Nursing Process by The Students in The Integrated Education System”, 56th Session of the International Statistical Institute, August 23-28 2007, Lisboa, Portugal
7. Ozfalcı Y and **Akyol M.**, “Estimation of Non-Function Attitude Score of Nursing Students With MARS”, 56th Session of the International Statistical Institute, August 23-28 2007, Lisboa, Portugal
8. Kırıkçı K and **Akyol M.**, “Environmental Factors Affecting the Academic Success of Gülhane Military Medical Faculty Students”, 56th Session of the International Statistical Institute, August 23-28 2007, Lisboa, Portugal
9. Bayer A, Erdem U, Mumcuoğlu T and **Akyol M.**, “Two-site phacotrabeculectomy vs. microincision cataract surgery combined with trabeculectomy”, Annual Meetings of American Academy of Ophthalmology, Nov. 10 - 13 2007, New Orleans, USA
10. Köymen CE, Karaçaylı Ü, Köymen R, Ergüder B, Aydınтуğ YS, **Akyol M.**, “Evaluation of the Effects of Articain forms for Impacted Third Molars on Stress Hormones”, 2nd International Oral & Maxillofacial Surgery Congress, May 16-20 2008, Antalya, TURKEY

11. Köymen CE, Aydınтуğ YS, Güsel A, Şençimen M, **Akyol M**, Varol A, A “Preliminary Study of Evulating Stress Factors among Turkish Oral and Maxillofacial Surgeons”, 2rd International Oral & Maxillofacial Surgery Congress, May 16-20 2008, Antalya, TURKEY
12. Güran Ş, Özet A, Dede M, Yenen MC, **Akyol M**, “The hereditary type breast cancer syndrome families had frequent colon/rectum and stomach cancers in their pedigree anlysis”, 3rd ESF Functional Genomics Conference, October 1-4 2008, Innsbruck, AUSTRIA
13. **Akyol M**, Sanisoğlu SY, Elhan AH, Kesici C, Cakir B, “Milk Consumption Habits in Turkish Society, Osteoporosis Risk Estimation Using MARS”, 29th Annual Conference of the International Society for Clinical Biostatistics, August 17-21 2008, Copenhagen, DENMARK
14. Karadağ M, Teke A, Demir C, Celen O, **Akyol M**, Akman N, “The Probable Problems on Homecare Services: Turkish Armed Forces Academic Physicians’ Opinions”, 29th Annual Conference of the International Society for Clinical Biostatistics, August 17-21 2008, Copenhagen, DENMARK
15. Sanisoğlu SY, Dolgun O, **Akyol M**, Aydın RM, Bağci H, Ugurlu M, “Data Mining Implementation in the Ministry of Health”, 29th Annual Conference of the International Society for Clinical Biostatistics, August 17-21 2008, Copenhagen, DENMARK
16. Yıldırım Y, **Akyol M**, Sanisoğlu SY, “Measurment Efficiency of Used Incative Medical Equipment Efficiency of Tobit Modelling in Hospital”, 29th Annual Conference of the International Society for Clinical Biostatistics, August 17-21 2008, Copenhagen, DENMARK
17. Çelen Ö, Teke A, Demir C, İçcan ÖF, Karadağ M, **Akyol M**, “Analysis of Factors Affecting Productivity of Physicians Working at Turkish Armed Formces Military Hospitals”, 29th Annual Conference of the International Society for Clinical Biostatistics, August 17-21 2008, Copenhagen, DENMARK
18. Aslan Ö, Sanisoğlu SY, **Akyol M**, “Problem Solving Among Oncology Nurses”, 2009 International MASCC/ISOO Symposium “Supportive Care in Cancer”, 25 – 27 June 2009, Rome, Italy
19. **Akyol M**, Sanisoğlu SY, Elhan AH, “Comparisons of Alternative Methods in Survival Analysis”, 57th Session of the International Statistical Institute, August 16-22 2009, Durban, South Africa
20. Çelen Ö., Demir C., Teke A.K., **Akyol M**, Karadağ M, Akman N,: “The Effect of job satisfaction on organizational commitment and intention to leave: an aplication in GMMF Training Hospital”, 57th Session of the International Statistical Institute, August 16-22 2009, Durban, South Africa
21. Karadağ M.,Teke A.K., Demir C., **Akyol M**, Özay ,,: “Academic mangers’ organizational commitment and turnover intention in a military training hospital”, 57th Session of the International Statistical Institute, August 16-22 2009, Durban, South Africa

22. Özgürtaş T, Aydın İ, Erbil MK, Turan Ö, Hirfanoğlu İM, Koç E, Açikel CH, **Akyol M**, Tunç T, “Vascular Endothelial Growth Factor, Basic-Fibroblast Growth Factor, Insulin-Like Growth Factor-I and Platelet-Derived Growth Factor Levels in Human Milk of Mothers with Term and Preterm Neonates”, 123rd AOAC Annual Meeting & Exposition, September 13 – 16 2009, Pennsylvania, USA
23. Balcioglu HA, Kilic C, **Akyol M**, Ulusoy A, “Horizontal migration of pre-and postnatal mental foramen”, 7th Forum of European Neuroscience, July 3 – 7 2010, Amsterdam, The Netherlands
24. **Akyol M**, and Sanisoglu SY, “A New Approaches to Biostatistics Education in Medical School”, 8th International Conference on Teaching Statistics, July 10-18 2010, Ljubljana, Slovenia
25. Demir C., and **Akyol M**, “Requirements Of Statistical Methodology For Health Managers”, 8th International Conference on Teaching Statistics, July 10-18 2010, Ljubljana, Slovenia

ULUSAL YAYINLAR

1. Akkuş Z, Sanisoğlu SY, **Akyol M**, Çelik MY, “Değişken Yapılarına Göre İstatistiksel Yaklaşım”, Dicle Tıp Dergisi, 33(2): 101-104, 2006
2. Şahin S, Saygun I, Kurt B, Çanakçı FC, **Akyol M**, Altuğ AH, Kurtiş B, Şençimen M, “Lokal antimikrobiyal ajanların palatinal bölgeden alınan greft alanındaki doku defektinin iyileşmesi üzerine etkilerinin histomorfometrik yöntemle incelenmesi”, Gülhane Tıp Dergisi, 51:27-33, 2009
3. Şahin S, Saygun I, Enhoş Ş, **Akyol M**, Altuğ A, Tekbaş ÖF, “The evaluation of the effect of educational status on oral health in young adult males”, Gazi Üniversitesi Diş Hekimliği Fakültesi Dergisi, 26(3): 133-139, 2009
4. Aydın İ, Özgürtaş T, Turan Ö, Koç E, Hırfanoğlu İM, Açikel CH, Garipoğlu G, **Akyol M**, Tunç T, Erbil MK, “Preterm Ve Term Yenidoğanların Anne Sütünün Biyokimyasal Karşılaştırması”, Türk Biyokimya Dergisi 34(4):242-249. 2009
5. Pişkin b, Uyar S, Ataç MS, Gündüz K, Avsever H, Ataç GK, Dayangeç E, Oğuz D, **Akyol M**, “Şiddetli Anksiyete Tanısı Konulmuş Miyofasiyal Ağrı Sendromlu Bireylerde Stabilizasyon Splinti ile Hidroksizin Etkinliğinin Karşılaştırılması”, Hacettepe Diş Hekimliği Fakültesi Dergisi 33(4): 11-19, 2009
6. Baysan O, **Akyol M**, Bugan B, Yokuşoğlu M, Gökoğlan Y, Genç C, “Benek İzleme Temelli Miyokardiyal Hızlar: Yeni Bir Yazılım İle Olan Deneyimlerimiz”, Anadolu Kardiyoloji Dergisi 10(3):233-238. 2010
7. Okçu KM, Doğan N, Şençimen M, Korkmaz C, Altuğ HA, İde T, **Akyol M**, “Bir Hemostatik Ajanın (Fastact) Deneysel Rat Modelinde Yara Ve Doku İyileşmesi Üzerine Etkisi”, Trakya Üniversitesi Tıp Fakültesi Dergisi 27(3):221-226. 2010

ULUSAL BİLDİRİLER

1. **Akyol M**, Sanisoğlu SY, Alpar R, Etikan İ, “Veri Zarflama Analizi (VZA) ile Hastane Verimliliklerinin Ölçülmesi ve Örnek Bir Uygulama”, 7.Ulusal Biyoistatistik Kongresi, 29 Eylül-2 Ekim 2004, Mersin
2. **Akyol M**, Sanisoğlu SY, Etikan İ, Alpar R, “Malmquist İndeksi ve Veri Zarflama Analizi ile Yıllara Göre Toplam Faktör Verimliliğindeki Değişimin Değerlendirilmesi ve Örnek Bir Uygulama”, 7.Ulusal Biyoistatistik Kongresi, 29 Eylül-2 Ekim 2004, Mersin
3. Sanisoğlu SY, **Akyol M**, Etikan İ, Akkuş Z, “Uyarlanabilir Küme Örneklem Yöntemi”, 7.Ulusal Biyoistatistik Kongresi, 29 Eylül-2 Ekim 2004, Mersin
4. Etikan İ., Uysal M., Sanisoğlu SY., **Akyol M.**, “Simülasyon ve Kullanımı”, 7. Ulusal Biyoistatistik Kongresi, 29 Eylül-2 Ekim 2004, Mersin.
5. Sanisoğlu SY, **Akyol M**, Saraçbaşı O, Etikan İ, Akkuş Z, “Wilcoxon İşaret Testinde Örneklem Büyüklüğü ve Güç Hesaplama Yaklaşımları”, 8. Ulusal Biyoistatistik Kongresi, 20 – 22 Eylül 2005, Bursa
6. **Akyol M**, Sanisoğlu SY, Etikan İ, Akkuş Z, Alpar R, “Polpülasyon Genetiği Prensipleri ve Hardy-Weinberg Kuralı”, 8. Ulusal Biyoistatistik Kongresi, 20 – 22 Eylül 2005, Bursa
7. **Akyol M** ve Sanisoğlu SY, “Biyoistatistik Eğitime Farklı Bir Yaklaşım: Gülhane Askeri Tıp Fakültesi (GATF) Modeli”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
8. Aydın A, Sanisoğlu SY ve **Akyol M**, “Yüksek Kalp Atım Hızı Nedeniyle Kalp Krizi Riski Bulunan Hastalarda Kısıtlanan Epinefrinin Etkisi”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
9. Odabaş A, Sanisoğlu SY ve **Akyol M**, “Solunum Yolu Semptomları İle Başvuran Hastalarda İdrar Yolu İnfeksiyonu Sıklığı”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
10. Filiz B, Sanisoğlu SY ve **Akyol M**, “Diabetes Mellitus Hastalarında Glukoz Değerleri Yaşa Göre Değişkenlik Gösterir mi?”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
11. Karakaya D, Sanisoğlu SY ve **Akyol M**, “Eş Mesleği Yapılan Doğumun Şeklini Belirlemede Etkili midir?”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
12. Karataş HA, Sanisoğlu SY ve **Akyol M**, “Kronik Böbrek Yetmezliği İle Proteinüri Arasındaki Ayırıcı Tanıda Kullanılabilecek Biyokimyasal Değerler”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak

13. Özakman İ, Sanisoğlu SY ve **Akyol M**, “Obez Hastalarda TSH Hormon Tedavisinin Gücü ve ACTH Hormonu Üzerindeki Etkileri”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
14. Erşen M, Sanisoğlu SY ve **Akyol M**, “FMF ve Behçet Hastalarında Kronik Hastalık Anemisi İncelenmesi”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
15. Erdem M, Sanisoğlu SY ve **Akyol M**, “Diabet Hastalarında Açlık Kan Glukozu ve OGTT Değerlerinin Kiloya Bağlı Olarak İncelenmesi”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
16. Aykutluğ Ö, Sanisoğlu SY ve **Akyol M**, “İskemik Kalp Hastalığında Sigara Riskinin İncelenmesi”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
17. Kamacı O, Sanisoğlu SY ve **Akyol M**, “Hepatit C Virüs Enfeksiyonlu Hastalarda Hepatit B Virüs Sıklığı”, 9. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2006, Zonguldak
18. Şimşek İ, Çınar M, Güleç M, Erdem H, Dinç A, Pay S, **Akyol M**, Meriç C, “Ankilozan Spondilit ve Romatoid Artrit Hastalarında Anti-TNF Tedavi Öncesi Anerji Durumunun Tüberkülin ve Candida Deri Testleri ile Değerlendirilmesi”, Romatoloji 2007 Kongresi, 1 – 5 Eylül 2007, Antalya
19. Güngör H, Karakaya D ve **Akyol M**, “ABD’deki Hastalardan Daha Çok Yatıyoruz”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
20. Yıldız H ve **Akyol M**, “EWING’s Sarkomada Yaş ve Cinsiyetin Etkisi”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
21. Sicim H ve **Akyol M**, “Sigara ve Hipertansiyonun Kronik Böbrek Hastalarına Etkisinin İncelenmesi”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
22. Çetiner SR ve **Akyol M**, “Hepatolitiküler Dejenerasyonun (WILSON Hastalığı) Kan Değerlerine Göre Olabilirliğinin İncelenmesi”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
23. Bulutlar E ve **Akyol M**, “Pubertede Erkek Seks Hormonlarının Hemoglobinin Üzerine Etkisi”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
24. Bayram Ç ve **Akyol M**, “Sigaranın Anemi Hastaları Üzerine Etkileri”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
25. Sevinçli Ş ve **Akyol M**, “Hipotiroidili Hastalarda Artan Risk: Ateroskleroza Bağlı Koroner Arter Hastalığı”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
26. Bayraktar V ve **Akyol M**, “Genç Anneler Dikkat”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas

27. Altay O ve **Akyol M**, “Vendeki Kan da Kapillerdeki Kan Değil mi?”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
28. **Akyol M**, Aslan FS, Sanisoğlu SY ve Elhan AH, “Genç Sporcuların Performansının MARS ile Kestirilmesi”, 10. Ulusal Biyoistatistik Kongresi, 05 - 08 Eylül 2007, Sivas
29. Şahin S, Saygun I, Kurt B, Çanakçı FC, **Akyol M**, Kurtiş B, “Lokal antimikrobiyal ajanların palatinal bölgeden alınan greft alanındaki doku defektinin iyileşmesi üzerine etkilerinin histomorfometrik yöntemle incelenmesi”, Türk Periodontoloji Derneği 38. Bilimsel Kongresi, 22 – 24 Mayıs 2008, İstanbul
30. Şahin S, Saygun I, Çanakçı FC, Yaşar F, **Akyol M**, Sevim A, Altuğ A, “Kahramanmaraş – Minnetpınarı’nda Ortaçağ Döneminde Yaşamış 27 İskelette Periodontal Durumun Gözlenmesi”, Türk Periodontoloji Derneği 38. Bilimsel Kongresi, 22 – 24 Mayıs 2008, İstanbul
31. **Akyol M**, Sanisoğlu SY, Elhan AH, Kesici C, Çakır B, “Türk Toplumunda Süt Tüketim Alışkanlıkları ve Osteoporoz Riskinin MARS Yöntemi ile Belirlenmesi”, 11. Ulusal Biyoistatistik Kongresi, 27 - 30 Mayıs 2008, Malatya
32. Özgürtaş T, Aydın İ, Turan Ö, Koç E, Hirfanoğlu İM, Garipoğlu G, Açikel CH, **Akyol M**, “Preterm ve Term Yenidoğanların Anne Sütlerinde, Vasküler Endotelial Growth Faktör, Basic-Fibrolast Growth Faktör, İnsulin-Like Growth Faktör-I ve Platelet-Derived Growth Faktör Seviyeleri”, IX. Ulusal Klinik Biyokimya Kongresi, 23 – 26 Nisan 2009, Antalya
33. Aydın İ, Özgürtaş T, Turan Ö, Aydın FN, Hirfanoğlu İM, Koç E, Kılıç S, **Akyol M**, Erbil MK, “Preterm ve Term Anne Sütü VEGF Konsantrasyonlarının Karşılaştırılması”, 21. Ulusal Biyokimya Kongresi, 28 - 31 Ekim 2009, İstanbul
34. **Akyol M**, “Tıp Fakültesi Biyoistatistik Eğitimine Farklı Bir Yaklaşım GATA – II Modeli”, 12. Ulusal Biyoistatistik Kongresi, 28 Haziran - 01 Temmuz 2010, Van
35. Özfalcı YÜ, Sanisoğlu SY, **Akyol M**, Yıldırım Y, “Hemşirelik Öğrencilerinin Fonksiyonel Olmayan Tutum Skorunun MARS Yöntemi İle Tahmini”, 12. Ulusal Biyoistatistik Kongresi, 28 Haziran - 01 Temmuz 2010, Van
36. Satıcı Ö, Akkuş Z, Sanisoğlu SY, **Akyol M**, “Tıp ve Diş Hekimliği Öğrencilerinin Banka Tercihleri ve Market Kartı Kullanma Alışkanlıklarının Lojistik Regresyon Yöntemiyle Belirlenmesi”, 12. Ulusal Biyoistatistik Kongresi, 28 Haziran - 01 Temmuz 2010, Van
37. Akkuş Z, Satıcı Ö, Sanisoğlu SY, **Akyol M**, Keskin, S, “Tıp ve Diş Hekimliği Öğrencilerinin Market Tercihleri ve Markalı Ürünleri Almadaki Tutumlarının İkili Lojistik Regresyon Yöntemiyle Belirlenmesi”, 12. Ulusal Biyoistatistik Kongresi, 28 Haziran - 01 Temmuz 2010, Van

VI. Bilimsel Etkinlikleri :

Aldığı Burslar : Training School in the Experimental Design & Statistical Analysis of Biomedical Experiments, (29 Mart – 02 Nisan 2009), Manchester – UK, Tam burslu (Yol, Eğitim, Konaklama)

Aldığı Ödüller : TÜBİTAK Yayın Teşvik Ödülü (3 kez)

GATA Bilimsel Çalışmaları Değerlendirme Kurulu,
Temel Tıp Bilimleri, İkincilik Ödülü
("Vascular endothelial growth factor, basic fibroblast growth factor, insulin-like growth factor-I and platelet-derived growth factor levels in human milk of mothers with term and preterm neonates" isimli çalışma ile)

Verdiği Kurslar :

- I. Deney Hayvanları Kullanım Sertifika Kursu içinde Biyoistatistik Kursu (17 – 25 Mart 2008), GATA-Ank.
- II. Deney Hayvanları Kullanım Sertifika Kursu içinde Biyoistatistik Kursu (03 – 11 Kasım 2008), GATA-Ank.
- III. Deney Hayvanları Kullanım Sertifika Kursu içinde Biyoistatistik Kursu (09 – 18 Mart 2009), GATA-Ank.
- IV. Deney Hayvanları Kullanım Sertifika Kursu içinde Biyoistatistik Kursu (13 – 22 Ekim 2009), GATA-Ank.

Bilimsel Araştırmalarda Uygulamalı Biyoistatistik Kursu (24 – 26 Haziran 2009), Kırıkkale Üniv., Kırıkkale

VII. Diğer Bilgiler :**Katılımcı olarak yer aldığı kurslar:**

1. Training School in the Experimental Design & Statistical Analysis of Biomedical Experiments (29 Mart – 02 Nisan 2009), Manchester – UK

Organizasyonuna katkıda bulunduğu bilimsel toplantılar:

1. Deney Hayvanları Kullanım Sertifika Kursu (17 – 25 Mart 2008), GATA-Ankara
2. Deney Hayvanları Kullanım Sertifika Kursu (03 – 11 Kasım 2008), GATA-Ankara
3. Deney Hayvanları Kullanım Sertifika Kursu (09 – 18 Mart 2009), GATA-Ankara
4. Deney Hayvanları Kullanım Sertifika Kursu (13 – 22 Ekim 2009), GATA-Ankara
5. Bilimsel Araştırmalarda Uygulamalı Biyoistatistik Kursu (24 – 26 Haziran 2009), Kırıkkale Üniversitesi, Kırıkkale

Görev yaptığı bilimsel kurullar:

1. Gülhane Askeri Tıp Fakültesi Biyoistatistik ve Bilgisayar Dersi, ders kurul temsilciliği ve I. sınıf ders kurul başkan yardımcılığı (2006 – halen devam ediyor)
2. Gülhane Askeri Tıp Fakültesi müfredat geliştirme kurulu üyeliği (2007 – halen devam ediyor)
3. GATA Hayvan Deneyleri Etik Kurulu üyeliği (Biyoistatistik Danışmanı) (17 Temmuz 2008 – halen devam ediyor)