

Cox Karma İyileşme Modeli ve Glioma Veri Kümesi Üzerine Bir Uygulama

Cox Mixture Cure Model and its Application to Glioma Data Set

Pınar KARA,^a
Nihal ATA TUTKUN^a

^aİstatistik Bölümü,
Hacettepe Üniversitesi Fen Fakültesi,
Ankara

Geliş Tarihi/Received: 03.07.2017
Kabul Tarihi/Accepted: 19.09.2017

Yazışma Adresi/Correspondence:
Nihal ATA TUTKUN
Hacettepe Üniversitesi Fen Fakültesi
İstatistik Bölümü, Ankara,
TÜRKİYE/TURKEY
nihalata@hacettepe.edu.tr

ÖZET Amaç: Orantılı tehlikeler varsayımına ve çalışmadaki her bireyin ilgilenilen olaya duyarlı olduğu varsayımına dayanan Cox regresyon modeli, yaşam çözümlemesinde en çok tercih edilen modellerden biridir. Ancak, kitlede ilgilenilen olayı yaşamayan çok sayıda birim bulunduğu bu ikinci varsayım uygun olmayabilir. İyileşme modelleri, kitlede ilgilenilen olayı yaşamayan bireylerin varlığını da göz önüne alan, klasik yaşam modellerine alternatif modellerdir. Literatürde farklı modeller önerilmekle birlikte, Cox karma iyileşme modeli en çok kullanılan karma iyileşme modellerinden biridir. Bu çalışmanın amacı Cox karma iyileşme modelini açıklayarak, bir veri kümesi üzerinde uygulamasını yapmak ve yorumlamaktır. **Gereç ve Yöntemler:** Bu çalışmada yaşam verisi için orantılı tehlikeler varsayımı incelenmiş, bu varsayımın sağlandığının görülmesi üzerine glioma veri kümesi için Cox regresyon modeli ve Cox karma iyileşme modeli incelenmiştir. **Bulgular:** Glioma veri kümesi için Cox karma iyileşme modelinde lojit bağlantı fonksiyonu kullanıldığında iyileşmiş kısım için hiçbir değişken anlamlı bulunmaz iken iyileşmemiş kısım için malignite düzeyi ve yaş değişkenlerinin istatistiksel olarak anlamlı değişkenler olduğu görülmüştür. Tamamlayıcı log-log bağlantı fonksiyonu kullanıldığında ise malignite düzeyi ve yaş değişkenlerinin hem iyileşmemiş hem de iyileşmiş kısım için istatistiksel olarak anlamlı değişkenler olduğu sonucuna ulaşılmıştır. **Sonuç:** Veri yapısının uygun olması durumunda Cox karma iyileşme modeli kullanılarak klasik yaşam modellerine kıyasla daha bilgi verici sonuçlar elde edilebileceği için bu modelin kullanımı önerilmektedir.

Anahtar Kelimeler: Cox regresyon modeli; Cox karma iyileşme modeli; orantılı tehlikeler; yaşam verisi

ABSTRACT Objective: Cox regression model which is based on the proportional hazards assumption and the assumption of every individual in the study is sensitive to the event of interest, is one of the most preferred model in survival analysis. However, when there are a lot of patients in the study who have never experienced the event of interest, the latter assumption may not be suitable. Cure models are alternative models to classical survival models which also account for the existence of this cured part. Although there are different types of cure models suggested in the literature, Cox mixture cure model is one of the most used model among them. In this study, Cox mixture cure model was examined and its application to the glioma data set was given. The results obtained from classical Cox regression model and Cox mixture cure model for glioma data set were interpreted. **Material and Methods:** In this study the proportional hazards assumption was examined for the survival data and after finding out this assumption is provided, Cox regression model and Cox mixture cure model were examined for the glioma data set. **Results:** When logit link function was selected for Cox mixture cure model, none of the covariates were found to be statistically significant for the cured part while malignancy grade and age were statistically significant covariates for uncured part. When complementary log-log link function was used, malignancy grade and age covariates were statistically significant for both cured and uncured parts. **Conclusion:** Since more informative results can be obtained by using Cox mixture cure model, they are recommended when the data set is suitable.

Sağlık bilimleri alanında, özellikle kanser konusunda yapılan araştırmalar gün geçtikçe hız kazanmaktadır. Kanser çalışmalarının çoğunda, değerlendirilen asıl çıktı ilgilenilen olayın meydana gelme süresidir. Bu süre genellikle yaşam süresi olarak adlandırılmaktadır. Eğer ilgilenilen olay çalışma grubundaki bütün bireylerde meydana gelirse, birçok istatistiksel çözümleme yöntemi kullanılabilir. Ancak, bazı bireylerin yeterli bir gözlem süresinin sonunda ilgilenilen olayı yaşamamış olması ve bundan dolayı bu bireylerde ilgilenilen olayın gerçekleşme sürelerinin bilinmemesi durumuyla sıklıkla karşılaşmaktadır. Verilerin bu özellikleri, yaşam çözümlemesinin ortaya çıkmasına sebep olmuştur.

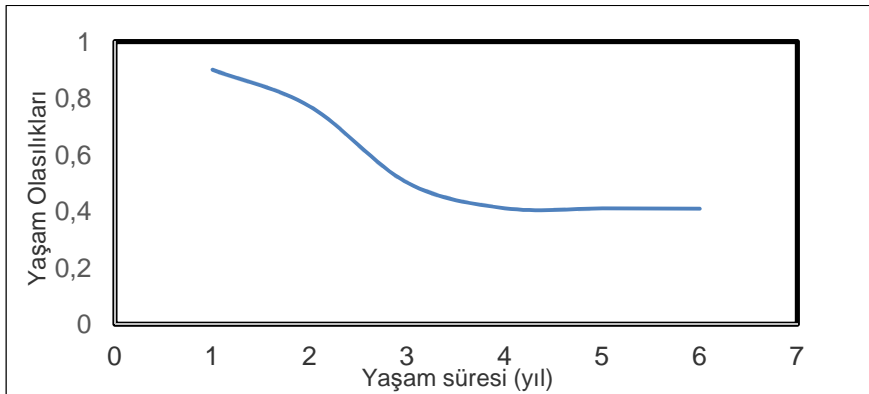
Yaşam çözümlemesi, kitlede ilgilenilen olayın meydana gelme sürelerinin araştırılması amacıyla kullanılan istatistiksel yöntemler topluluğu olarak tanımlanmaktadır. Yaşam çözümlemesinin temel kavramlarından biri yaşam süresidir ve yaşam süresini etkileyen faktörleri belirlemek için çeşitli yaşam modelleri geliştirilmektedir. Bu amaçla en çok tercih edilen yaşam modeli Cox regresyon modeli (CRM)'dir.

CRM, kitledeki her bir bireyin ilgilenilen olayı yaşayacağı, araştırılan olaya göre başarısızlığa (ölüme, metastaza veya nüksetmeye) yatkın olduğu varsayımına dayanmaktadır. Ancak bu varsayım, yeterli bir gözlem süresinin ardından hastaların büyük bir bölümünün hayatta kalması durumunda doğru olmayabilir. Özellikle son yıllarda, meme kanseri, çocukluk lenfomaları ve lösemileri dâhil olmak üzere bazı kanser türlerindeki araştırmalar, bu hastalıklara sahip hastaların önemli bir bölümünün, uygulanan tedavilerden sonra iyileştiğini göstermiştir. Dolayısıyla, bu durumu dikkate alan ve bu amaçla geliştirilen "iyileşme modelleri" kanser hastalarının yaşam verilerini analiz etmek için kullanılan önemli bir istatistiksel model haline gelmiştir.¹ İyileşme modelleri Boag (1949) tarafından önerilmiş ve farklı araştırmacıların çalışmaları ile geliştirilmiştir.¹⁻¹⁴

Tıbbi anlamda iyileşme, hastada hastalıkla ilgili bütün göstergelerinin yok olmasıdır. Bu tanım, iyileşmenin birey düzeyindeki tanımıdır. Ancak bir hastanın tıbbi olarak iyileştiğini varsaymak zor olabilir. İstatistiksel anlamda iyileşme ise, kitlede hastalıkla ilişkili ölüm oranının kitlede beklenen ölüm oranına eşit olması ile gerçekleşir. Bu durumda ek ölüm oranı (excess mortality rate) sifıra yaklaşır. Bu, iyileşmenin kitle düzeyindeki tanımıdır.¹⁴ Gözlenen ölüm oranı aşağıdaki gibi ifade edilebilir;

Gözlenen Ölüm Oranı = Beklenen Ölüm Oranı + Ek Ölüm Oranı

Ek ölüm oranı sifıra yaklaştığında yaşam eğrisi, Şekil 1'deki gibi düzleşerek düz bir kuyruğa sahip olacaktır ve bunun gibi yaşam eğrilerine sahip veri kümelerinde iyileşme modellerinin kullanımı önerilmektedir.¹⁴



ŞEKİL 1: Yaşam eğrisi örneği.

İYİLEŞME MODELİ VE COX REGRESYON MODELİ ARASINDAKİ FARKLAR

Cox (1972) belli bir aralığın başında hayatta olan bireylerin, bu aralık sürecinde ölmesi olasılığını dikkate alarak bir model geliştirmiştir.¹⁵ Herhangi bir zaman noktasında, bir gruptaki tehlike fonksiyonunun, aynı zaman noktası için başka bir gruptaki tehlike fonksiyonu ile orantılı olacağı varsayımında bulunmuştur. Bu, gruplar arasındaki yaşam olasılıklarının karşılaştırılmasına olanak sağlamaktadır.

Genellikle, iyileşmiş hastaların oranının, klasik yaşam çözümlemesi yöntemleri ile hesaplanan “1-tehlike oranı” ile tahmin edilebileceği varsayılmaktadır. Ancak, Peto vd. (1977) ve Clark vd. (2003)’ün çalışmalarına göre, “1-tehlike oranı” tedavi sayesinde ölümü ya da nüksetmesi önlenen hastaların oranı değil, önlenen ya da ertelenen ölümlerin oranıdır. Dolayısıyla, tehlike oranı ölümü önleyen (iyileştirici tedavi) ve ölümü erteleyen (yaşam süresini uzatan veya ölümü erteleyen tedavi) arasında ayırım yapamamaktadır.^{16,17}

İyileşme oranı ile tehlike oranı arasındaki ilişki, Cox’un çalışmasında da görülmektedir.¹⁵ 1959 ve 1960 yılları arasında, çocuklardan oluşan hasta grubu için 6-Merkaptopürin (6-MP) tedavi grubu ve placebo grubu olacak biçimde kontrollü bir deney yapılmıştır. CRM kullanılarak veriler analiz edildiğinde 6-MP grubunun placebo grubuna göre tehlike oranı 0.22 (%95 güven düzeyinde güven aralığı ise 0.10-0.49’dır) bulunmuştur. Eğer tehlike oranı ile iyileşme oranı arasındaki ilişki geçerli olsaydı bulunan sonuç şunu ifade edecekti: placebo grubunda meydana gelebilecek nüksetmelerin %78’i (1-0.22=0.78), 6-MP tedavisi sayesinde önlenmiştir. Ancak, hastalar daha uzun süre takip edildiğinde, 6-MP ile tedavisi yapılan çocukların neredeyse hepsinin hastalığın nüksetmesinden dolayı öldüğü görülmüştür.¹⁸

İyileşme modellerinin klasik CRM’ye tercih edildiği durumlar aşağıdaki biçimde özetlenebilmektedir;

1. Yaşam eğrisi düz bir kuyruğa sahip ise orantılı tehlikeler varsayımı sağlanmıyor olabilir. Bu durumda iyileşme modellerin kullanılması daha iyi sonuç verebilmektedir.
2. Çalışma grubu içerisinde heterojenlik var ise iyileşme modellerin kullanılması önerilebilmektedir.¹⁹

KARMA İYİLEŞME MODELLERİ

İyileşme modelleri karma ve karma olmayan iyileşme modelleri olmak üzere iki alt başlıkta incelenmektedir. Ancak karma olmayan iyileşme modelleri ile ilgili literatürde uygulamalı çalışmalara nadir rastlanılmaktadır.

Karma iyileşme modelleri, kitlede hem iyileşmiş hem de iyileşmemiş kısmı dikkate alan modellerdir. Karma iyileşme modellerinde kitle iki alt gruba ayrılabilir; kitledeki bireyler $0 \leq \pi \leq 1$ olmak üzere ya $1-\pi(\mathbf{z})$ oranıyla iyileşmiştir ya da $\pi(\mathbf{z})$ oranıyla iyileşmemiştir.

$S_u(t|\mathbf{x})=S(t|Y=1, \mathbf{x})$ t anında iyileşmemiş hastaların yaşam fonksiyonu, Y bir bireyin, ilgilenilen olayı yaşayacağını (Y=1) ya da yaşamayacağını (Y=0) gösteren gösterge değişkeni, \mathbf{x} iyileşmemiş kısım için açıklayıcı değişkenler vektörü, \mathbf{z} ise iyileşmiş kısım için açıklayıcı değişkenler vektörü iken karma iyileşme modelleri en genel biçimiyle aşağıdaki gibi ifade edilebilir;

$$S(t|\mathbf{x}, \mathbf{z})=\pi(\mathbf{z})S_u(t|\mathbf{x})+(1-\pi(\mathbf{z}))$$

ya da

$$S(t|\mathbf{x}, \mathbf{z})=\pi(\mathbf{z})S(t|Y=1, \mathbf{x})+(1-\pi(\mathbf{z})). \quad (1)$$

Karma iyileşme modellerinde iyileşmiş kısım ve iyileşmemiş kısım ayrı ayrı modellenmektedir. Karma modellerin bu özelliği açıklayıcı değişkenlerin iyileşmiş ve iyileşmemiş hastalar üzerindeki etkilerinin

farklı olmasına olanak sağlamaktadır. Örneğin, uygulanan bir tedavi, iyileşmiş hastaların oranını arttırmış ancak bunun yanı sıra iyileşmemiş hastaların yaşamında herhangi bir etkisi olmamış olabilir. Karma iyileşme modelleri, bunun gibi durumlarda klasik yaşam modellerine kıyasla daha geniş yorum yapılabilmesine imkan sağlamaktadır.

Karma iyileşme modellerinde, \mathbf{z} açıklayıcı değişkenler vektörünün iyileşmiş kısım üzerindeki etkisini modellemek için bazı bağlantı (link) fonksiyonları kullanılmaktadır. İyileşmiş kısmın oranı $1-\pi(\mathbf{z})$ açıklayıcı değişkenler ile farklılık gösterebilir ve farklı durumlarda farklı bağlantı fonksiyonlarının kullanılması bir avantaj olabilir. Bu bağlantı fonksiyonları aşağıdaki gibi özetlenebilir:

- **Lojit bağlantı fonksiyonu**, iyileşmiş kısmı modellemek için en çok tercih edilen bağlantı fonksiyonudur ve aşağıdaki biçimde ifade edilmektedir;

$$\pi(\mathbf{z}) = \frac{e^{\mathbf{bz}}}{1+e^{\mathbf{bz}}}$$

Burada \mathbf{b} , açıklayıcı değişken vektörü ile ilişkili bilinmeyen parametrelerin vektörüdür.

- **Tamamlayıcı log-log bağlantı fonksiyonu (Cloglog)**, lojit bağlantı fonksiyonundan sonra en çok tercih edilen bağlantı fonksiyonudur ve $\log(-\log(1-\pi(\mathbf{z}))) = \mathbf{bz}$ biçiminde ifade edilmektedir.
- **Probit bağlantı fonksiyonu**, φ standart normal birikimli dağılım fonksiyonu iken, $\varphi^{-1}(\pi(\mathbf{z})) = \mathbf{bz}$ biçiminde ifade edilmektedir.¹²

Karma iyileşme modellerinde iyileşmemiş kısım için yaşam fonksiyonunu belirten $S_u(t|\mathbf{x})$ ise, hangi karma iyileşme modelinin kullanıldığına bağlı olarak farklı biçimlerde ifade edilmektedir. $S_{u0}(t|\mathbf{x})$ iyileşmemiş hastalar için temel yaşam fonksiyonu iken, $S_u(t|\mathbf{x}) = S_{u0}(t)^{\exp(\mathbf{x}'\boldsymbol{\beta})}$ ise bu karma iyileşme modeli “Cox karma iyileşme modeli”, $S_u(t|\mathbf{x}) = S_{u0}(t e^{\mathbf{x}'\boldsymbol{\beta}})$ ise “yarı parametrik hızlandırılmış başarısızlık süresi karma iyileşme modeli” ve $S_u(t|\mathbf{x}) = S_{u0}(t e^{\mathbf{x}'\boldsymbol{\beta}})^{\exp(-\mathbf{x}'\boldsymbol{\beta})}$ ise “hızlandırılmış tehlikeler karma iyileşme modeli” olarak adlandırılmaktadır.¹²

KARMA İYİLEŞME MODELİ İÇİN OLABİLİRLİK FONKSİYONU

n tane hastanın katıldığı bir çalışma ele alınsın. $i=1,2,\dots,n$ iken, i . hasta için gözlemlenen değerler $\mathbf{O}=(t_i, \delta_i, \mathbf{z}_i, \mathbf{x}_i)$ olsun. Burada t_i yaşam süresi, δ_i durdurma gösterge değişkeni (durdurma varsa $\delta_i=0$, durdurma yok ise $\delta_i=1$) olarak ifade edilmiştir. $\Theta=(\mathbf{b}, \boldsymbol{\beta}, S_0(t))$ tahmin edilmek istenilen bilinmeyen parametreler olsun. $\mathbf{y}=(y_1, y_2, \dots, y_n)$ vektörü ve gözlemlenen veriler \mathbf{O} biliniyorken olabirlik fonksiyonu aşağıdaki biçimde ifade edilebilir;

$$L(\mathbf{b}, \boldsymbol{\beta}) = \prod_{i=1}^n [1 - \pi_i(\mathbf{z}_i)]^{1-y_i} \pi_i(\mathbf{z}_i)^{y_i} h(t_i | Y=1, \mathbf{x}_i)^{\delta_i} S(t_i | Y=1, \mathbf{x}_i)^{y_i}. \quad (2)$$

Bu eşitlikte $h(\cdot)$, $S(\cdot)$ 'ye karşılık gelen tehlike fonksiyonudur. Bu olabirlik fonksiyonu iki parça olarak da ifade edilebilir;

$$l_{c1}(\mathbf{b}; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^n y_i \log[\pi(\mathbf{z}_i)] + (1-y_i) \log[1-\pi(\mathbf{z}_i)] \quad (3)$$

ve

$$l_{c2}(\boldsymbol{\beta}; \mathbf{O}, \mathbf{y}) = \sum_{i=1}^n y_i \delta_i \log[h(t_i | Y=1, \mathbf{x}_i)] + y_i \log[S(t_i | Y=1, \mathbf{x}_i)]. \quad (4)$$

Tam olabilirlik ve log-olabilirlik fonksiyonları bu şekilde ifade edildikten sonra, farklı tahmin yöntemleri ele alınabilir. EM algoritması, y'nin modele dâhil olması nedeniyle kullanışlıdır. Burada gözlemlenemeyen değişken EM algoritmasındaki beklenen değer ile yer değiştirecektir.¹²

COX KARMA İYİLEŞME MODELİ

Cox karma iyileşme modeli Kuk ve Chen (1992) tarafından önerilmiştir ve orantılı tehlikeler karma iyileşme modeli olarak da adlandırılmaktadır.⁵

T, ilgilenilen olayın gerçekleşme zamanını göstermektedir ve sadece Y=1 iken tanımlıdır. Olasılık yoğunluk fonksiyonu $f(t|Y=1)$, yaşam fonksiyonu ise $S(t|Y=1)$ ile gösterilmiştir. T'nin marjinal yaşam fonksiyonu $t < \infty$ iken, $S(t) = (1 - \pi) + \pi S(t|Y=1)$ 'dir. t sonsuza yakınsarken $S(t)$, $1 - \pi$ 'ye yakınsayacaktır.

Farewell (1977) iyileşmiş kısmın hesaplanması için lojistik regresyon modelini kullanmıştır ve bu modelde iyileşme oranı $1 - \pi(z) = 1 - [\exp(\mathbf{z}'\mathbf{b}) / (1 + \exp(\mathbf{z}'\mathbf{b}))]$ biçiminde ifade edilmektedir.⁴ Burada açıklayıcı değişken vektörü z, $S(t|Y=1)$ için parametrik bir yaşam modeli ve kesişim terimi içermektedir. Kuk ve Chen (1992) \mathbf{x} kesişim değişkeni dışındaki açıklayıcı değişkenler vektörü ve temel tehlike fonksiyonu olmak üzere, $h(t|Y=1; \mathbf{x}) = h_0(t|Y=1) \exp(\mathbf{x}'\boldsymbol{\beta})$ tehlike fonksiyonuna sahip CRM'yi kullanarak bu modeli genelleştirmiştir.⁵

Bu model \mathbf{b} ve $\boldsymbol{\beta}$ aracılığı ile açıklayıcı değişkenlerin iyileşmiş ve iyileşmemiş kısım üzerindeki etkilerini ayırabilme özelliğine sahiptir. Bu nedenle de ilgilenilen olaya duyarlı olmayan bir grupta önsel bilgi olduğunda, esnek bir model sağlar. Kuk ve Chen modeli için, koşullu birikimli tehlike fonksiyonu $H(t|Y=1, \mathbf{x}) = H_0(t|Y=1) \exp(\mathbf{x}'\boldsymbol{\beta})$ olur. Koşullu temel yaşam fonksiyonu $S_0(t|Y=1)$ iken, koşullu yaşam fonksiyonu $S(t|Y=1; \mathbf{x}) = S_0(t|Y=1) \exp(\mathbf{x}'\boldsymbol{\beta})$ olarak ifade edilir.

Cox karma iyileşme modellerinin en çok olabilirlik (EÇO) tahminleri, log-olabilirlik fonksiyonu maksimize edilerek elde edilebilir. Eğer fonksiyonun kapalı biçimi yok ise Newton-Raphson iterasyon yöntemi kullanılarak sonuçlar elde edilebilmektedir.

COX KARMA İYİLEŞME MODELİ İÇİN EN ÇOK OLABİLİRLİK TAHMINİ

i bireyi için gözlenen veriler $\mathbf{O} = (t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$ olarak ifade edilsin ve $t_1 < t_2 < \dots < t_k$ k tane ayrık olayın meydana geliş zamanları olsun. Burada y_i, Y_i rastgele değişkeninin aldığı değer iken, eğer $\delta_i = 1$ ise $y_i = 1$ olacağı ve $\delta_i = 0$ ise y_i 'nin gözlemlenemeyeceği sonucuna ulaşılabilmektedir. i. bireyin olabilirlik fonksiyonuna katkısı, $\pi_i = P(Y_i = 1; \mathbf{z}_i)$ iken, $\delta_i = 1$ için $\pi_i f(t_i|Y=1; \mathbf{x}_i)$, $\delta_i = 0$ için ise $(1 - \pi_i(z)) + \pi_i(z) S(t_i|Y=1; \mathbf{x}_i)$ 'dir. Böylece, Cox karma iyileşme modeli için gözlenen tam olabilirlik fonksiyonu;

$$L(\mathbf{b}, \boldsymbol{\beta}, h_0) = \prod_{i=1}^n \left\{ \pi_i h_0(t_i|Y=1) e^{\mathbf{x}_i' \boldsymbol{\beta}} e^{-H_0(t_i|Y=1) \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right\}^{\delta_i} \times \left\{ (1 - \pi_i) + \pi_i e^{-H_0(t_i|Y=1) \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right\}^{1 - \delta_i} \quad (5)$$

olarak ifade edilir. Burada $L(\mathbf{b}, \boldsymbol{\beta}, h_0)$ 'yi maksimize eden $\hat{\mathbf{b}}$ ve $\hat{\boldsymbol{\beta}}$ tahminleri elde edilmek istenmektedir. Klasik CRM'de, kısmi olabilirliği kullanmak için klasik çözümlenme, $h_0(t)$ 'ye bağlı değildir. Breslow (1975) yarı parametrik bir tam olabilirlik yapısı ve $H_0(t)$ 'nin tahmin edicisinin Nelson-Aalen tahmin edicisi olduğu profil olabilirlik yöntemini kullanmış ve $\boldsymbol{\beta}$ 'nin kısmen maksimize edilmiş olabilirlik fonksiyonunun, kısmi olabilirlik ile orantılı olduğunu göstermiştir.²¹ Ancak bu yaklaşım, Cox iyileşme modeli için uygun değildir. $S_0(t)$ 'yi eleyerek sadece az miktarda bilginin kaybedildiği klasik CRM'nin aksine, tahmindeki $S_0(t|Y=1)$, b hakkındaki bilgiler kaybedilmeden modelden çıkartılamaz.⁷

EM ALGORİTMASI

Cox karma iyileşme modelinde EM algoritması kullanılmaktadır. EM algoritmasındaki E adımı, gözlemlenen veri \mathbf{O} ve parametrelerin o andaki tahminleri $\Theta^{(m)}=(\mathbf{b}^{(m)}, \boldsymbol{\beta}^{(m)}, S_0^{(m)}(t))$ biliniyorken, y_i 'lere göre tam log-olabilirliklerin koşullu beklentisini hesaplar. Bu adımı tamamlamak için Eşitlik 1 ve Eşitlik 2 y_i 'nin doğrusal fonksiyonları olduğundan dolayı, y_i 'nin koşullu beklentisini elde etmek yeterlidir. $E(y_i|\mathbf{O}, \Theta^{(m)})$ aşağıdaki biçimde ifade edilebilir;

$$w_i^{(m)}=E(y_i|\mathbf{O}, \Theta^{(m)})=\delta_i+(1-\delta_i)\frac{\pi(\mathbf{z}_i)S(t_i|Y=1;\mathbf{x}_i)}{1-\pi(\mathbf{z}_i)+\pi(\mathbf{z}_i)S(t_i|Y=1;\mathbf{x}_i)}\Big|_{(\mathbf{O}, \Theta^{(m)})} \quad (6)$$

Bu eşitlikten, eğer $\delta_i=1$ ise $w_i^{(m)}=1$ ve $\delta_i=0$ ise $w_i^{(m)}$ 'nin iyileşmemiş hastaların oranı olacağı görülmektedir. $E(y_i|\mathbf{O}, \Theta^{(m)})$ eşitliğinin ikinci kısmı, i. hastanın iyileşmemiş olarak kalması koşullu olasılığı olarak yorumlanabilir. $\delta_i \log w_i^{(m)}=0$ ve $\delta_i w_i^{(m)}=\delta_i$ olduğu için Eşitlik 3 ve Eşitlik 4'ün beklenen değerleri sırasıyla aşağıdaki gibi yazılabilir;

$$E(l_{c1})=\sum_{i=1}^n w_i^{(m)} \log[\pi(\mathbf{z}_i)]+(1-w_i^{(m)}) \log[1-\pi(\mathbf{z}_i)] \quad (7)$$

$$E(l_{c2})=\sum_{i=1}^n \delta_i \log[w_i^{(m)} h(t_i|Y=1, \mathbf{x}_i)]+w_i^{(m)} \log[S(t_i|Y=1, \mathbf{x}_i)] \quad (8)$$

EM algoritmasında M adımı, bilinmeyen parametrelere göre Eşitlik 7 ve Eşitlik 8'in maksimize edilmesi ile elde edilir. İyileşmiş kısım için parametrelerin tahmininde, R programında 'glm' fonksiyonunun içindeki 'link' seçeneği kullanılır. Sy ve Taylor (2000) ve Peng ve Dear (2000)'in çalışmalarında $\boldsymbol{\beta}$ 'yi temel tehlike fonksiyonu belirtmeden tahmin etmek için bir kısmi olabilirlik yöntemi ele alınmıştır.^{7, 22} Bunun için tahmin eşitliği;

$$\log \prod_{i=1}^n [h_0(t_i) \exp(\boldsymbol{\beta}' \mathbf{x}_i + \log(w_i^{(m)}))]^{\delta_i} S_0(t_i)^{\exp(\boldsymbol{\beta}' \mathbf{x}_i + \log(w_i^{(m)}))} \quad (9)$$

biçimindedir.

Algoritmanın M adımı ise l_c 'nin maksimizasyonunu içerir. $H_0(t|Y=1)$ ya da $S_0(t|Y=1)$ için, profil olabilirlik yöntemleri kullanılarak M adımına ek bir maksimizasyon adımı uygulanır. CRM'den, $H_0(t|Y=1)$ için Breslow-tipi tahmin edici ve $S_0(t|Y=1)$ için çarpım-limit tahmin edicisi olmak üzere iki yöntem genişletilebilir.

E adımına geri dönmek için, tahmin edilen yaşam fonksiyonları güncellenmelidir. $t_1 < t_2 < \dots < t_k$ ayrık, durdurulmamış başarısızlık zamanları, d_{t_j} ilgilenilen olayların sayısı, $R(t_j)$ t_j zamanındaki risk kümesi olsun. $S_0(t|Y=1)$ için Breslow tipi tahmin edici Eşitlik 10'da verilmiştir;

$$\hat{S}_0(t|Y=1)=\exp\left[-\sum_{j:t_{(j)} \leq t} \frac{d_{t_{(j)}}}{\sum_{i \in R(t_{(j)})} w_i^{(m)} \exp(\boldsymbol{\beta}' \mathbf{x}_i)}\right] \quad (10)$$

Kitlede iyileşmiş kısmın varlığından dolayı t sonsuza giderken $\hat{S}_0(t|Y=1)$ sifıra yakınsamayabilir. Tanımlanma sorunlarını ortadan kaldırmak amacıyla t_k son gözlemlenen başarısızlık zamanını ifade ediyor iken, $t > t_k$ için $\hat{S}_0(t|Y=1) = 0$ olarak ayarlanabilir.¹²

UYGULAMA

Çalışmada, agresif bir beyin kanseri türü olan glioma hastası 411 bireyi içeren veri kümesi ele alınmıştır.²³ Hastalara iki çeşit kemoterapi tedavisi uygulanmıştır. Çalışma sürecinde 411 hastanın 274'ü hayatını kaybetmiş ve bu hastalar başarısız olarak ele alınmıştır. Çalışmadaki diğer hastalar ise durdurulmuş olarak

nitelendirilmiştir. Buna göre 411 hastadan 274'ünde (%66.7) başarısızlık ve 137'sinde (%33.3) durdurma gözlenmiştir. Yaşam süresi hastaların ölümüne ya da durdurulmasına kadar geçen süre (gün) olarak tanımlanmıştır ve bu çalışma için takip süresi 1617 gündür. Tedavi değişkeni dışında yaşam süresini etkilemesi olası olan dört değişken daha çalışmaya dâhil edilmiştir. Çalışmada kullanılan değişkenler ve düzeyleri Tablo 1'de verilmiştir. Yaş değişkeni ortanca yaşa göre sınıflandırılarak kullanılmıştır. Karnofsky indeksi, hastaların işlevsel bozukluklarının derecesine göre sınıflandırılmasını sağlayan bir performans ölçөгüdür ve 0 ile 100 arasında değer almaktadır. Karnofsky indeksinin 0 olması ölümü, 80'den büyük olması hastanın günlük aktivitelerini yardıma gerek duymadan yerine getirebildiğini, 100 olması ise hastalıkla ilgili herhangi bir yakınmanın olmadığı normal durumu ifade etmektedir.²⁴

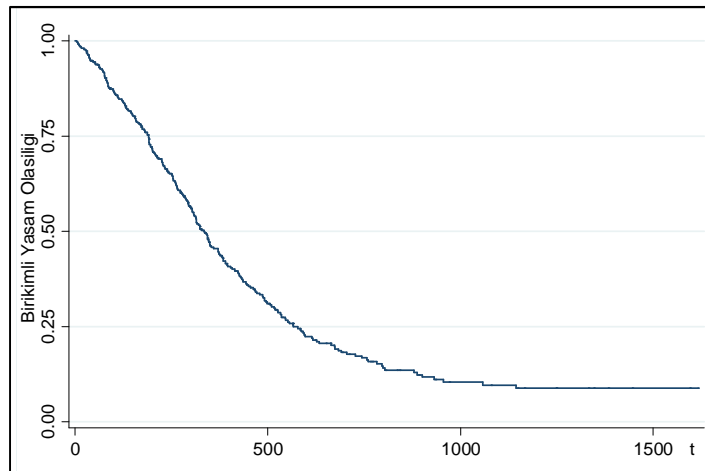
TABLO 1: Kullanılan değişkenler ve düzeyleri.

Değişken	Değişken Düzeyleri	n (%)	Başarısız Gözlemlerin Sayısı	Durdurulmuş Gözlemlerin Sayısı
Malignite Düzeyi	0: Malignite düzeyi 1 ve 2	139 (%33.8)	74	65
	1: Malignite düzeyi 3	272 (%66.2)	200	72
Cinsiyet	0: Erkek	233 (%56.7)	158	75
	1: Kadın	178 (%43.3)	116	62
Yaş	0: Yaş ≤ 54 olanlar	212 (%51.6)	123	89
	1: Yaş > 54 olanlar	199 (%48.4)	151	48
Karnofsky İndeksi	0: Karnofsky indeksi ≤ 80	312 (%75.9)	219	93
	1: Karnofsky indeksi > 80	99 (%24.1)	55	44
Tedavi	0: Standart	204 (%49.6)	137	67
	1: Yeni	207 (%50.4)	137	70

Aşağıdaki kısımda verilerin çözümlenmesine ilişkin bulgulara yer verilmektedir. Çalışmada sonuçlar IBM SPSS Statistics 23 ve R programları kullanılarak elde edilmiştir.

PARAMETRİK OLMAYAN YAŞAM ÇÖZÜMLEMESİ SONUÇLARI

Çalışmada kullanılan glioma veri kümesine ait Kaplan-Meier (KM) yaşam eğrisi Şekil 2'de verilmiştir.



ŞEKİL 2: Glioma veri kümesi için KM yaşam eğrisi

Glioma veri kümesi için elde edilen KM yaşam eğrisine bakıldığında eğrinin düz bir kuyruğa sahip olduğu görülmektedir. Kitlede ilgilenilen olayı yani ölümü yaşamayan bireylerin varlığı düşünülebilir ve iyileşme modellerinin kullanımının değerlendirilmesi daha iyi sonuçlar elde edilmesini sağlayabilir.

Açıklayıcı değişken düzeyleri için yaşam olasılıkları bakımından istatistiksel olarak anlamlı bir farklılık olup olmadığı log-rank testi kullanılarak incelenmiş ve Tablo 2'deki sonuçlar elde edilmiştir.

Değişken	Ki-kare	p değeri
Cinsiyet	0.10	0.755
Malignite Düzeyi	26.51	<0.001
Karnofsky İndeksi	7.71	0.005
Tedavi	0.77	0.380
Yaş	49.60	<0.001

Log-rank testi sonuçlarına göre malignite düzeyi ($p<0.001$), karnofsky indeksi ($p=0.005$) ve yaş ($p<0.001$) değişkenlerinin düzeyleri arasında yaşam olasılıkları bakımından istatistiksel olarak anlamlı bir farklılık olduğu %95 güven düzeyinde söylenebilmektedir.

COX REGRESYON MODELİNDEN ELDE EDİLEN SONUÇLAR

Cox regresyon çözümlemesinde değişken düzeylerinin yorumlanması amacıyla düzeylerden biri referans kategorisi olarak alınıp yorumlar yapılmaktadır. Bu çalışmada tüm değişkenler için birinci düzey referans kategorisi olarak alınmıştır.

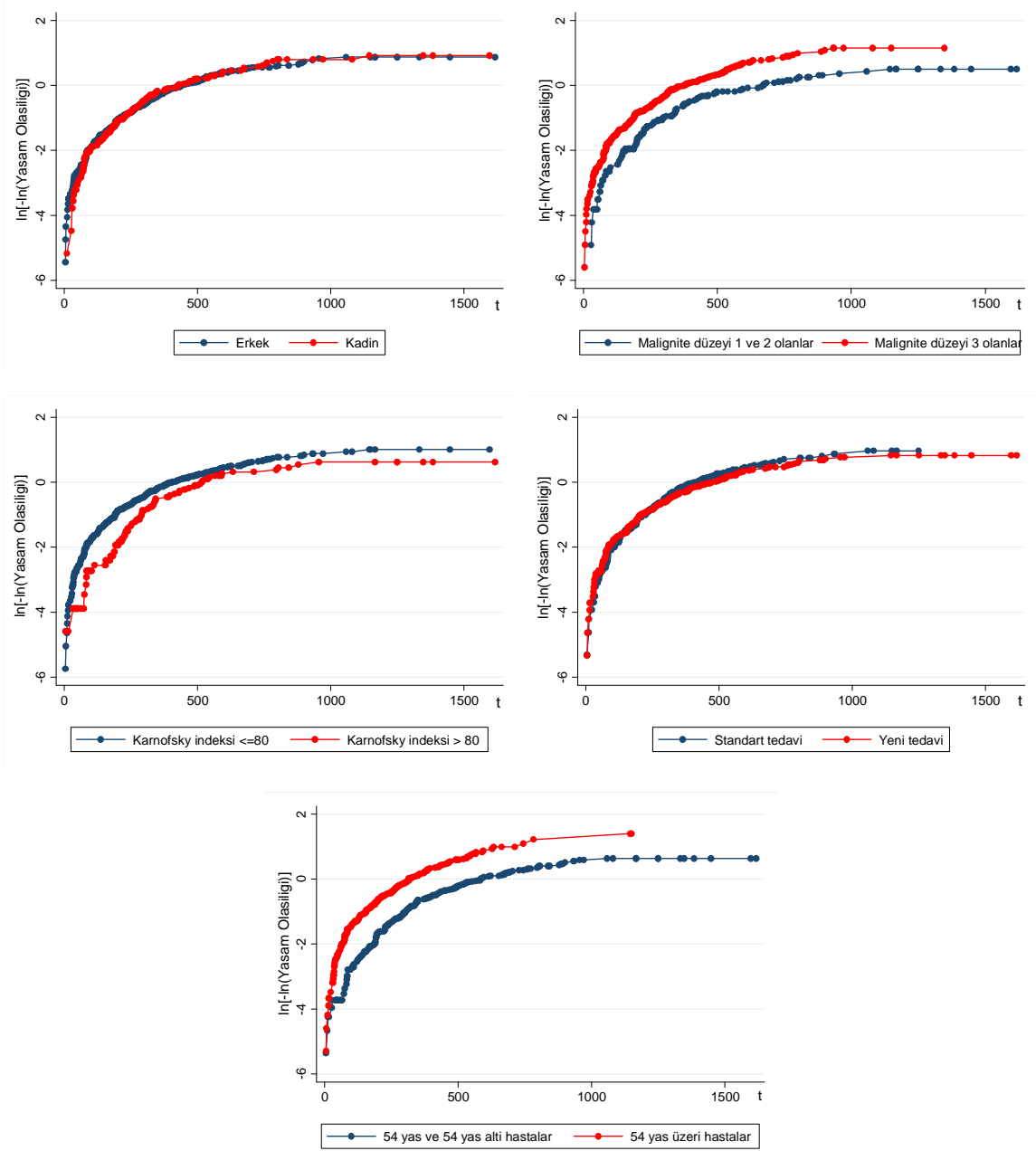
Glioma veri kümesi için CRM kullanıldığında p değeri 0.001'den küçük olarak elde edildiğinden modelin anlamlı olduğu %95 güven düzeyinde söylenilebilmektedir. CRM için AIC değeri 2739.576, BIC değeri 2759.576 ve -2LogL değeri 2729.576 olarak elde edilmiştir. Glioma verisi için CRM sonuçları Tablo 3'te verilmiştir.

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Cinsiyet	-0.128	0.125	0.308	0.880	0.688, 1.125
Malignite Düzeyi	0.569	0.139	<0.001	1.766	1.344, 2.320
Yaş	0.781	0.129	<0.001	2.183	1.696, 2.810
Karnofsky İndeksi	-0.292	0.153	0.056	0.746	0.553, 1.007
Tedavi	-0.086	0.121	0.479	0.918	0.724, 1.164

Tablo 3'te açıklayıcı değişkenler için p değerleri incelendiğinde yaş ($p<0.001$) ve malignite düzeyi ($p<0.001$) değişkenlerinin glioma kanseri veri kümesi için yaşam süresi üzerinde etkili olan önemli risk faktörleri olduğu görülmüştür. CRM'den elde edilen sonuçlara göre, malignite düzeyi 3 olan hastaların, malignite düzeyi 1 ve 2 olan hastalara göre başarısızlık riskinin 1.766 kat ($e^{0.569}=1.766$) daha fazla olduğu söylenebilmektedir. Benzer biçimde 54 yaş üzeri hastaların, yaşı 54 ve 54'den küçük olan hastalara göre başarısızlık riskinin 2.183 kat ($e^{0.781}=2.183$) daha fazla olduğu ifade edilebilmektedir.

ORANTILI TEHLİKELER VARSAYIMININ İNCELENMESİ

$\ln(-\ln)$ yaşam eğrileri, orantılı tehlikeler varsayımının incelenmesi için grafiksel bir yaklaşım sağlamaktadır. Glioma verisi için açıklayıcı değişkenlere ait $\ln(-\ln)$ yaşam eğrileri elde edilmiş ve Şekil 3'te verilmiştir.



ŞEKİL 3: Açıklayıcı değişken düzeyleri için $\ln(-\ln)$ grafikleri.

Veri kümesi için orantılı tehlikeler varsayımı sağlanıyorsa, $\ln(-\ln)$ eğrilerinin çakışmaması beklenmektedir. Şekil 3 incelendiğinde cinsiyet ve tedavi değişkenleri için $\ln(-\ln)$ yaşam eğrilerinde çakışma olduğu görülmektedir. Tedavi ve cinsiyet değişkenlerinin orantılı tehlikeler varsayımını sağlayıp sağlamadığını kesin olarak söyleyebilmek için Schoenfeld artıkları yöntemi ve zamana bağlı açıklayıcı değişkenler yöntemi de kullanılmıştır.

Glioma veri kümesindeki değişkenler için orantılı tehlikeler varsayımının sağlanıp sağlanmadığı, Schoenfeld artıkları ile bireylerin başarısızlık sürelerinin rankı arasındaki korelasyon testi ile incelenebilmektedir.^{25,26} Bu test istatistiği için yokluk hipotezi “orantılı tehlikeler varsayımı sağlanmaktadır” biçimindedir. Çalışmada orantılı tehlikeler varsayımının sağlanıp sağlanmadığını incelemek için kullanılan Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testine ait sonuçlar Tablo 4’de verilmiştir.

Değişken	r'	Ki-kare	Sd	p değeri
Cinsiyet	0.026	0.20	1	0.658
Malignite Düzeyi	-0.001	0.00	1	0.990
Yaş	-0.085	2.04	1	0.153
Karnofsky İndeksi	0.112	3.50	1	0.061
Tedavi	-0.023	0.14	1	0.705

* Yaşam süresinin rankı ile Schoenfeld artıkları arasındaki Pearson korelasyon katsayısı.

Glioma veri kümesindeki tüm açıklayıcı değişkenler için p değerleri 0.05’den büyük olarak elde edildiği için bu modelde orantılı tehlikeler varsayımının sağlandığı söylenebilmektedir.

Orantılı tehlikeler varsayımının incelenmesi için kullanılan bir diğer yöntem zamana bağlı açıklayıcı değişkenler yöntemidir.²⁶ Bu çalışmada orantılı tehlikeler varsayımını sağlamıyor olabileceği düşünülen cinsiyet değişkeni için zamana bağlı değişkenler yöntemi ele alınmıştır. Bu amaçla ilk olarak sadece cinsiyet değişkenini ele alan bir CRM kurulmuş ve bu model için elde edilen sonuçlar Tablo 5’de verilmiştir. Bu modele ait -2LogL değeri 2800.9358 olarak elde edilmiştir. İkinci aşamada cinsiyet değişkeni ve zamana bağlı bir fonksiyon olan cinsiyetxlogt değişkeni için CRM sonuçları elde edilmiş ve Tablo 6’da verilmiştir. Bu model için de -2LogL değeri 2799.975’dir.

Değişken	$\hat{\beta}$	Std. hata	p değeri	exp($\hat{\beta}$)	exp($\hat{\beta}$) için %95 güven aralığı
Cinsiyet	0.038	0.122	0.755	1.039	0.817, 1.321

Değişken	$\hat{\beta}$	Std. hata	p değeri	exp($\hat{\beta}$)	exp($\hat{\beta}$) için %95 güven aralığı
Cinsiyet	-0.626	0.694	0.367	0.534	0.137, 2.083
Cinsiyetxlogt	0.124	0.127	0.330	1.132	0.882, 1.453

Orantılı tehlikeler varsayımının sağlanıp sağlanmadığını incelemek amacıyla kullanılan zamana bağlı açıklayıcı değişkenler yönteminde yokluk hipotezi $\delta = 0$ biçiminde kurulmuştur. CRM ile zamana bağlı CRM arasındaki log-olabilirlik oranı istatistiğinin farkı test istatistiğini vermektedir. Test istatistiği yokluk hipotezi altında 1 serbestlik dereceli ki-kare dağılımına sahiptir. Yokluk hipotezi kabul edilirse,

cinsiyet değişkeni için orantılı tehlikeler varsayımının sağlandığı söylenebilmektedir.²⁶ Bu model için test istatistiği,

$$LR = -2\ln\hat{L}_{CRM}(-2\ln\hat{L}_{Zamana\ bağı\ CRM}) = 2800.9358 - 2799.975 = 0.9608$$

biçiminde elde edilmiştir. Elde edilen test istatistiği, $\chi^2 = 3.86$ 'dan küçük olduğu için yokluk hipotezi %95 güven düzeyinde kabul edilmiş ve cinsiyet değişkeni için orantılı tehlikeler varsayımının sağlandığı sonucuna ulaşılmıştır.

Orantılı tehlikeler varsayımını sağlamayabileceği düşünülen tedavi değişkeni için de yukarıda belirtilen yöntem kullanılmış ve modeller için elde edilen sonuçlar Tablo 7 ($-2\log L = 2800.2638$) ve Tablo 8 ($-2\log L = 2799.1406$)'de verilmiştir.

TABLO 7. Tedavi değişkeni için CRM'den elde edilen sonuçlar.					
Değişken	$\hat{\beta}$	Std. Hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Tedavi	-0.106	0.121	0.380	0.899	-0.344, 0.131

TABLO 8. Zamana bağlı tedavi değişkeni için CRM'den elde edilen sonuçlar.					
Değişken	$\hat{\beta}$	Std. Hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Tedavi	0.601	0.680	0.377	1.824	-0.732, 1.934
Tedavixlogt	-0.132	0.125	0.290	0.876	-0.377, 0.113

Olabilirlik oran test istatistiği,

$$= 2800.2638 - 2799.1406 = 1.1232$$

olarak elde edilmiştir. Buna göre $LR = 1.1232 < \chi_1^2 = 3,86$ olduğu için tedavi değişkeninin orantılı tehlikeler varsayımını sağladığı %95 güven düzeyinde söylenebilmektedir.

Glioma veri kümesi için yapılan incelemeler sonucundan orantılı tehlikeler varsayımının sağlandığı sonucuna ulaşılmıştır. Buna göre CRM sonuçları kullanılarak yorum yapılabilmektedir.

COX İYİLEŞME MODELİNDEN ELDE EDİLEN SONUÇLAR

Karma iyileşme modelleri, kitlede ilgilenilen olayı yaşamayan çok sayıda birey var iken kitledeki her bireyin başarısızlığa yatkın olduğunu varsayan yaşam modellerine göre daha doğru sonuçlar verebilmektedir. KM yaşam eğrisi incelenip, eğri düzleşmiş bir kuyruğa sahip ise iyileşme modellerinin kullanılması önerilmektedir. Glioma verisi için KM eğrisi düzleşen bir kuyruğa sahip olduğundan bu veri kümesi için iyileşme modellerinin kullanımı değerlendirilmiştir. En sık kullanılan karma iyileşme modelleri Cox karma iyileşme modeli ve hızlandırılmış başarısızlık süresi karma iyileşme modelidir. Glioma veri kümesi için orantılı tehlikeler varsayımı sağlandığı için bu çalışmada Cox karma iyileşme modeli kullanılmıştır.

Karma iyileşme modellerinde iyileşme oranının modellenmesi için üç temel bağlantı fonksiyonu kullanılmaktadır. Bunlar lojit bağlantı fonksiyonu, probit bağlantı fonksiyonu ve cloglog bağlantı fonksiyonlarıdır. Glioma verisi için probit bağlantı fonksiyonu kullanıldığında model yakınsama sorunları nedeni ile elde edilemediği için çalışmada sonuçlara yer verilememiştir.

Glioma verisinde iyileşme oranının tahmini için lojit bağlantı fonksiyonu kullanıldığında elde edilen sonuçlar Tablo 9a'da, iyileşmemiş kısım için elde edilen sonuçlar ise Tablo 9b'de verilmiştir.

TABLO 9A. Glioma verisinde iyileşmiş kısım için lojit Cox iyileşme modeli sonuçları.				
Değişken	$\hat{\beta}$	Std hata	z değeri	p değeri
Sabit	1.938	1.221	1.587	0.112
Cinsiyet	0.273	0.816	0.335	0.738
Malignite Düzeyi	2.007	2.720	0.738	0.460
Yaş	1.887	3.945	0.478	0.632
Karnofsky İndeksi	-1.185	1.322	-0.897	0.370
Tedavi	-1.092	1.280	-0.853	0.394

TABLO 9B. Glioma verisinde iyileşmemiş kısım için lojit Cox iyileşme modeli sonuçları.				
Değişken	$\hat{\beta}$	Std hata	z değeri	p değeri
Cinsiyet	-0.215	0.160	-1.342	0.180
Malignite Düzeyi	0.379	0.157	2.413	0.016
Yaş	0.705	0.137	5.164	<0.001
Karnofsky İndeksi	-0.141	0.177	-0.798	0.425
Tedavi	-0.027	0.140	-0.195	0.845

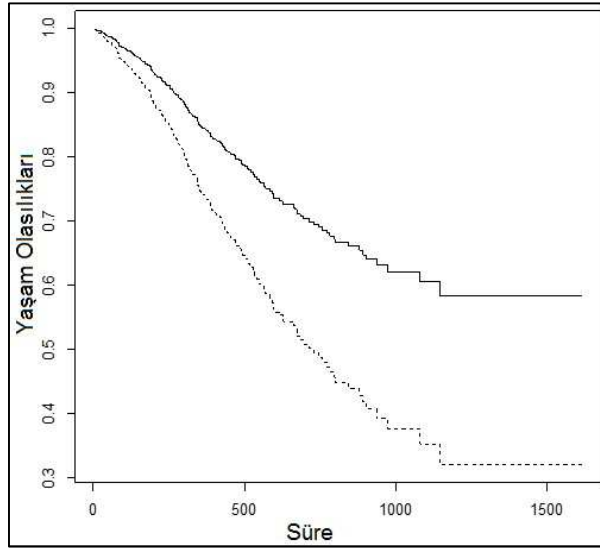
İyileşmiş kısım için değişken katsayıları ve düzeyleri kullanılarak istenilen iyileşme oranları elde edilebilmektedir. İyileşmiş kısmın oranı hesaplanırken parametre tahmininde kullanılan bağlantı fonksiyonu göz önüne alınmakta ve hesaplamalar buna göre yapılmaktadır. Lojit bağlantı fonksiyonu için, iyileşme oranı tahmin edilmek istenilen değişken düzeyleri seçildikten sonra $1 - \pi(z) = e^{\hat{\beta}z} / (1 + e^{\hat{\beta}z})$ biçiminde hesaplanmaktadır.

Glioma veri kümesi için, malignite düzeyi 1 ya da 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi uygulanmış ve yaşı 54 ya da 54'den küçük olan erkek hastaların iyileşme oranı;

$1 - \pi(z) = 1 - \frac{e^{(1.938+(2.207 \times 0)+(-1.185 \times 1)+(-1.092 \times 0)+(1.887 \times 0)+(0.273 \times 0)}}{1 + e^{(1.938+(2.207 \times 0)+(-1.185 \times 1)+(-1.092 \times 0)+(1.887 \times 0)+(0.273 \times 0)}}$ biçiminde hesaplanmış ve %58.4 olarak elde edilmiştir.

Aynı değişken düzeyleri alınıp sadece tedavi düzeyi değiştirilip, yeni tedavi yerine standart tedavi seçildiğinde iyileşme oranının $1 - e^{1.938-1.185} / (1 + e^{1.938-1.185}) = \%32$ 'ye düştüğü görülmüştür. Bu değişken düzeyleri için yaşam eğrileri Şekil 4'de verilmiştir.

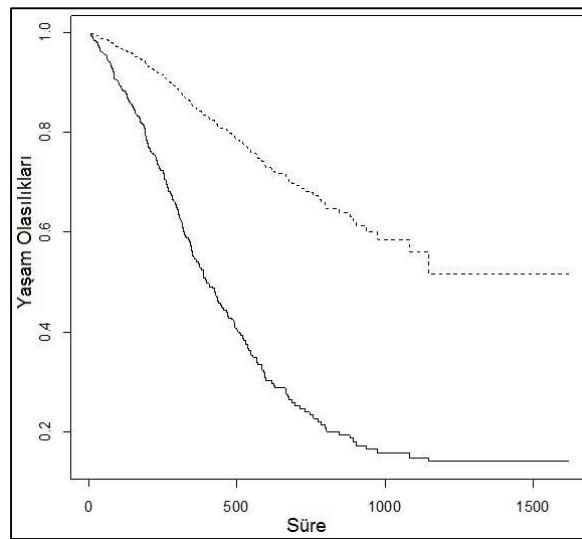
Şekil 4'deki kesiksiz eğri malignite düzeyi 1 veya 2 olan, karnofsky indeksi 80'den büyük, yaşı 54 ya da 54'den küçük, yeni tedavi türü uygulanmış erkek hastaların yaşam eğrisini göstermektedir. Kesikli eğri ise aynı değişken düzeyleri alınıp sadece tedavi türü değiştirildiğinde elde edilen yaşam eğrisini ifade etmektedir. Şekil 4'den bu değişken düzeyleri için sadece uygulanan tedavinin türü değiştiğinde, yaşam olasılıklarında değişme olduğu ve yeni tedavi uygulandığında iyileşme oranının daha yüksek olduğu görülmüştür.



ŞEKİL 4: Seçilen açıklayıcı değişken düzeyleri için yaşam olasılıkları grafiği-1.

Aynı biçimde malignite düzeyi 1 ya da 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi uygulanmış ve yaşı 54 ya da 54'den küçük olan kadın hastalar alındığında bu grup için iyileşme oranı %51.7 olarak bulunmuştur. Bu değişken düzeyleri için tedavi türü değiştirilip standart tedavi tercih edildiğinde ise iyileşmiş kısmın oranı sadece %26.4 olarak tahmin edilmiştir. Bu veri kümesi için, malignite düzeyi 3 olan, 54 yaş üzeri, karnofsky indeksi 80'den küçük olan ve standart tedavi uygulanan kadınlar için iyileşme oranı en düşük olarak tahmin edilmiştir ve sadece %2'dir.

Malignite düzeyi 1 veya 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi uygulanmış 54 yaş üzeri kadın hastaların iyileşme oranı hesaplanmış ve %13.9 olarak elde edilmiştir. Malignite düzeyi 1 veya 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi uygulanmış 54 yaş ve 54 yaş altı kadın hastalar için ise bu oran %51.6 olarak hesaplanmıştır. Bu değişken düzeyleri için yaşam eğrileri Şekil 5'de verilmiştir.



ŞEKİL 5: Seçilen açıklayıcı değişken düzeyleri için yaşam olasılıkları grafiği-2.

Şekil 5'de kesikli eğri malignite düzeyi 1 veya 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi uygulanmış, 54 yaş ve 54 yaş altı kadın hastalar yaşam eğrisi, sürekli eğri ise aynı değişken düzeyleri için yaş 54 üzeri alındığında elde edilen yaşam olasılıklarını göstermektedir.

Tablo 9b kitledeki iyileşmemiş hastalar için değişken düzeylerinin tehlike oranlarının yorumlanmasını sağlamaktadır. Bu model için malignite düzeyi ($p=0.016$) ve yaş ($p<0.001$) değişkenleri istatistiksel olarak anlamlı bulunmuştur. Tablo 9b'den elde edilen sonuçlara dayanarak, malignite düzeyi 3 olan hastaların riskinin, malignite düzeyi 1 ve 2 olan hastalara göre 1.46 kat daha fazla olduğu söylenebilmektedir. Yaşı 54 üzeri olan hastaların, yaşı 54 ve 54'den küçük hastalara göre 2.02 kat ($e^{0.705}=2.02$) daha fazla risk altında olduğu ifade edilebilmektedir.

Glioma verisi için Cox iyileşme modelinin elde edilmesinde cloglog bağlantı fonksiyonu kullanıldığında elde edilen sonuçlar Tablo 10a ve Tablo 10b'de verilmiştir.

TABLO 10A. Glioma verisinde iyileşmiş kısım için cloglog Cox iyileşme modeli sonuçları.

Değişken	$\hat{\beta}$	Std hata	z değeri	p değeri
Sabit	0.159	0.139	1.139	0.255
Cinsiyet	-0.131	0.125	-1.052	0.293
Malignite Düzeyi	0.425	0.134	3.165	0.001
Yaş	0.360	0.124	2.891	0.004
Karnofsky İndeksi	-0.195	0.137	-1.422	0.155
Tedavi	-0.012	0.123	-0.096	0.923

TABLO 10B. Glioma verisinde iyileşmemiş kısım için cloglog Cox iyileşme modeli sonuçları.

Değişken	$\hat{\beta}$	Std hata	z değeri	p değeri
Cinsiyet	-0.062	0.134	-0.463	0.643
Malignite Düzeyi	0.411	0.152	2.705	0.006
Yaş	0.651	0.133	4.913	<0.001
Karnofsky İndeksi	-0.206	0.147	-1.407	0.159
Tedavi	-0.136	0.136	-0.996	0.319

Cloglog bağlantı fonksiyonu kullanıldığında, hem iyileşmiş hastaların iyileşme oranı için, hem de iyileşmemiş hastaların yaşam fonksiyonu için malignite düzeyi ve yaş değişkenlerinin %95 güven düzeyinde anlamlı olduğu ($p<0.05$) görülmektedir.

İstenilen değişken düzeyleri için iyileşme oranı $1-\pi$, tamamlayıcı log-log bağlantı fonksiyonunu kullanılarak ($\log(-\log(1-\pi)) = \hat{\beta}z$) elde edilmiştir. Buna göre, malignite düzeyi 1 ya da 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi türü uygulanmış ve yaşı 54 ya da 54'den küçük olan erkek hastaların iyileşme oranı %38.6 olarak bulunmuştur. Aynı değişken düzeyleri alınıp sadece uygulanan tedavi türü standart tedavi olarak değiştirildiğinde iyileşme oranı %38.1 olmuştur. Bu model için, tedavi türündeki değişikliğin, iyileşmiş kısmın oranını etkileyen önemli bir değişken olmadığı görülmektedir. Yaş değişkeninin iyileşmiş kısma olan etkisini incelemek için, malignite düzeyi 1 ya da 2 olan, karnofsky indeksi 80'den büyük, yeni tedavi türü uygulanmış erkek hastalardan, yaşı 54 üzeri olan hastalar için

iyileşme oranı hesaplanmış ve %25.5 olarak elde edilmiştir. Bu modelde, düşük yaş grubundaki hastaların iyileşme oranının, yüksek yaş grubundaki hastaların iyileşme oranından daha yüksek olduğu ve yaş değişkeninin istatistiksel olarak anlamlı olduğu görülmektedir. Yüksek yaş, bu model için iyileşme oranını azaltan bir faktördür. Malignite düzeyinin iyileşmiş kısımdaki etkisini incelemek için, Karnofsky indeksi 80'den büyük, yeni tedavi türü uygulanmış ve yaşı 54 ya da 54'den küçük olan erkek hastalardan malignite düzeyi 3 olanlar alınmıştır. Bu grup için iyileşme oranı %23.3 olarak elde edilmiştir. Aynı değişken düzeylerinde, sadece malignite 1 veya 2 değerini aldığında bu oran %38.6 olarak elde edilmişti. Buradan da, bu grup için malignite düzeyi 3 olanların iyileşme oranının, malignite düzeyi 1 ve 2 olanlardan daha düşük olduğu görülebilmektedir ve malignite düzeyi bu model için istatistiksel olarak anlamlı ($p=0.001$) bir değişkendir.

Glioma verisinde cloglog Cox karma iyileşme modeli sonuçlarından, iyileşmemiş kısım için malignite düzeyi ($p=0.06$) ve yaş ($p<0.001$) değişkenlerinin istatistiksel olarak anlamlı olduğu %95 güven düzeyinde söylenebilmektedir. Malignite düzeyi 3 olan hastalar, malignite düzeyi 1 ve 2 olan hastalara göre 1.51 kat ($e^{0.411}=1.51$) daha fazla risk altında bulunduğu ifade edilebilir. Yaşı 54'ün üzerinde olan hastaların başarısızlık riskinin, yaşı 54 ve 54'ün altında olan hastaların başarısızlık riskine göre 1.92 kat ($e^{0.651}=1.92$) daha fazla olduğu söylenebilmektedir.

İstenilen tüm gruplar için iyileşme oranı, parametre tahminleri ve değişken düzeyleri kullanılarak bu biçimde yorumlanabilmektedir. Daha önce de belirttiğimiz gibi, karma iyileşme modellerinin, açıklayıcı değişken etkilerinin iyileşmiş kısım ve iyileşmemiş kısımda farklı olmasına olanak sağladığı elde edilen sonuçlarda da görülmektedir.

Glioma verisi için lojit Cox karma iyileşme modeli ve cloglog Cox karma iyileşme modeli sonuçlarına bakıldığında, cloglog bağlantı fonksiyonu kullanıldığında elde edilen modelin p değerlerinin ve standart hatalarının çok daha düşük olduğu görülmüştür ve bu nedenle cloglog Cox karma iyileşme modelinin kullanılması daha uygun olabileceği düşünülebilir.

SONUÇ VE TARTIŞMA

Yaşam çözümlemesinde en yaygın kullanımına sahip Cox regresyon modeli birçok araştırmacı tarafından yarı parametrik bir model olması ve uygulanabilirliğinin kolay olması nedeni ile tercih edilmektedir. İçinde bulunduğumuz yüzyılın gelişimlerine ve gereksinimlerine göre farklı yarı-parametrik yaşam modellerinin geliştirilmesi üzerine çalışmalar yapılmaktadır. Cox karma iyileşme modeli de bu modellerden biridir. Özellikle kanser araştırmalarında geliştirilen yeni tedaviler ile hastaların iyileşme olasılıkları artmaktadır. Bu durum da iyileşmiş ve iyileşmemiş hastaların özelliklerinin ayrı ayrı değerlendirilmesinin uygun olacağını sonucunu açığa çıkarmaktadır.

Bu çalışmada glioma veri kümesi için CRM'ye ve Cox karma iyileşme modeline ait sonuçlar elde edilmiştir. İyileşme modellerinde EM algoritması kullanıldığından istatistiksel paket programlarından bilgi kriterine ilişkin bir sonuç elde edilememektedir. Bu nedenle de CRM ve Cox karma iyileşme modeli bilgi kriterleri kullanılarak karşılaştırılamamaktadır. Ancak iyileşme modelleri iyileşmiş ve iyileşmemiş kısımlar için ayrı ayrı yorum yapabilme imkanı verdiğiinden veri yapısının uygun olması durumunda tercih edilmektedir.

Cox karma iyileşme modeli için sonuç elde ederken bağlantı fonksiyonlarına bağlı olarak bazı sorunlar ortaya çıkmaktadır. Bu da modelin gerçek veri kümeleri için kullanılabilirliğini zorlaştırmaktadır. Bu nedenle daha hızlı ve daha detaylı sonuç almayı sağlayacak yazılımların geliştirilmesi gerekmektedir.

Teşekkür

Yrd. Doç. Dr. Uğur Karabey'e çalışmamıza katkılarından dolayı teşekkür ederiz.

Çıkar Çatışması

Yazarlar herhangi bir çıkar çatışması veya finansal destek bildirmemiştir.

Yazar Katkıları

Fikir Kavram: Pınar Kara, Nihal Ata Tutkun; **Tasarım:** Pınar Kara, Nihal Ata Tutkun; **Denetleme/Danışmanlık:** Pınar Kara, Nihal Ata Tutkun; **Analiz ve Yorum:** Pınar Kara, Nihal Ata Tutkun; **Kaynak Taraması:** Pınar Kara, Nihal Ata Tutkun; **Makalenin Yazımı:** Pınar Kara, Nihal Ata Tutkun; **Eleştirel İnceleme:** Pınar Kara, Nihal Ata Tutkun.

KAYNAKLAR

1. Rama R, Swaminathan R, Venkatesan P. Cure models for estimating hospital-based breast cancer survival. *Asian Pac J Cancer Prev* 2010;11(2):387-91.
2. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society* 1949;11(1):15-44.
3. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *J Am Stat Assoc* 1952;47(259):501-15.
4. Farewell VT. A model for a binary variable with time-censored observations. *Biometrika* 1977;64(1):43-6.
5. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika* 1992;79(3):531-41.
6. Taylor JM. Semi-parametric estimation in failure time mixture models. *Biometrics* 1995;51(3):899-907.
7. Sy JP, Taylor JMG. Estimation in a cox proportional hazards cure model. *Biometrics* 2000;56(1):227-336.
8. Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. *J Am Stat Assoc* 2003;98(464):1063-78.
9. Yin G, Ibrahim JG. Cure rate models: a unified approach. *Can J Stat* 2005;33(4):559-70.
10. Yu B. A frailty mixture cure model with application to hospital admission data [corrected]. *Biom J* 2008;50(3):386-94.
11. Zhang J, Peng Y. Accelerated hazards mixture cure model. *Lifetime Data Anal* 2009;15(4):455-67.
12. Cai C, Zou Y, Peng Y, Zhang J. Smcure: an R-package for estimating semiparametric mixture cure models. *Comput Methods Programs Biomed* 2012;108(3):1255-60.
13. Maetani S, Gamel JW. Parametric cure model versus proportional hazards model in survival analysis of breast cancer and other malignancies. *Adv Breast Cancer Res* 2013;2(4):119-25.
14. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007;8(3):576-94.
15. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society* 1972;34(2):187-220.
16. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trial requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer* 1977;35(1):1-39.
17. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer* 2003;89(2):232-8.
18. Maetani S, Gamel JW. Parametric cancer survival analysis based on the boag model: a clinician's view. *Journal of Biometrics & Biostatistics* 2012;3(3):7-17.
19. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. *Clin Cancer Res* 2012;18(14):3731-6.
20. Silvestri GA, Handy J, Lackland D, Corley E, Reed CE. Specialists achieve better outcomes than generalists for lung cancer surgery. *Chest* 1998;114(3):675-80.
21. Breslow NE. Analysis of survival data under the proportional hazards model. *Int Stat Rev* 1975;43(1):45-57.
22. Peng Y, Dear KB. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000;56(1):237-43.
23. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* 1992;11(16):2093-109.
24. Kamofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of the nitrogen mustards in the palliative treatment of carcinoma. With particular reference to bronchogenic carcinoma. *Cancer* 1948;1(4):634-56.
25. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982;69(1):239-41.
26. Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. 2nd ed. New York: Springer-Verlag; 2005. p.590.