



Yaşam Çözümlemesinde Hızlandırılmış Başarısızlık Süresi Modelleri ve Bir Uygulama

Durdu KARASOY*¹, Selin SEZAYİ¹

¹Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Beytepe, Ankara
(Alınış Tarihi: 20.01.2013, Kabul Tarihi: 27.04.2014)

Anahtar Kelimeler

Cox regresyon modeli
Parametrik modeller
Hızlandırılmış başarısızlık süresi modelleri
Orantılı tehlikeler

Özet: Yaşam verileri için en yaygın kullanılan regresyon modeli Cox regresyon modelidir. Cox regresyon modelinin önemli bir varsayımı, tehlike hızlarının zaman boyunca orantılı olmasıdır. Hızlandırılmış başarısızlık süresi modeli, tehlikeler orantılı olmadığına yaşam verilerinin çözülmesi için alternatif bir yöntemdir. Hızlandırılmış başarısızlık süresi modelleri, belirli durumlar altında Cox modelden daha etkili parametre tahminleri sağlar. Bu çalışmada, hızlandırılmış başarısızlık süresi modelleri tanıtıldı ve mide kanseri hastalarına ait veriler kullanılarak bu modeller örneklendi ve sonuçlar tartışıldı.

Accelerated Failure Time Models in Survival Analysis and an Application

Keywords

Cox regression model
Parametric models
Accelerated failure time models
Proportional hazards

Abstract: The Cox regression model is the most commonly used regression model for survival data. The Cox regression model has an important assumption that hazard rates are proportional over time. The accelerated failure time model is an alternative method for the analysis of survival data when hazards are not proportional. The accelerated failure time models should lead to more efficient parameter estimates than Cox model under certain circumstances. In this study, The accelerated failure time models are described. The data of stomach cancer patients is used to illustrate these models and the results are discussed.

1. Giriş

Yaşam çözümlemesi, başarısızlık olarak adlandırılan bir nokta olayı ile ilgilenmektedir. Yaşayan bir organizmanın ya da cansız bir nesnenin belirli bir başlangıç zamanı ile başarısızlığı arasında geçen zamana "yaşam süresi" ya da "başarısızlık süresi" adı verilmektedir. Yaşam süresini etkileyen faktörler yarı parametrik ya da parametrik regresyon modelleri kullanılarak incelenebilmektedir. Yaşam çözümlemesinde kullanılan regresyon modellerinin diğer istatistiksel modellerden temel farkı durdurulmuş veri içermesidir (Cox, 1972; Oakes, 1977; Cox ve Oakes, 1984; Klembaum, 1996; Klein ve Moeschberger, 1997; Collett, 2003; Lee ve Wang, 2003).

Yaşam çözümlemesinde en sık kullanılan model Cox regresyon modelidir. Cox regresyon modelinin kullanılabilmesi için orantılı tehlikeler varsayımının sağlanması gerekmektedir (Klembaum, 1996; Klein ve Moeschberger, 1997; Lawless, 1982).

Ancak yapılan çalışmaların çoğunda bu varsayım incelenmeden doğrudan Cox regresyon modeli

kullanılmaktadır. Orantılı tehlikeler varsayımı sağlanmadığında Aalen (1994) zayıflığı modellemeyi, Hougaard (1995) zayıflık modellerinin, Wei (1992), Keiding, Andersen ve Klein (1997) hızlandırılmış başarısızlık süresi (HBS) modellerinin kullanılmasını, Hougaard, Myglegaard ve Borch-Johnsen (1994) ise açıklayıcı değişkenlerin seçim probleminde HBS modellerinin önemini belirtmişlerdir. HBS modelleri, bir çok uygulamada daha uygun bir model yapısı sağlamaktadır (Kay ve Kinnersley, 2002). HBS modellerinde sonuçları yorumlamak daha kolaydır (Orbe vd., 2002).

Bu çalışmada, HBS modellerini tanıtmak ve mide kanseri verilerini kullanarak bu modellerin uygulanabilirliğini ve yorumlarını göstermek amaçlanmıştır.

2. Materyal ve Metot

Yaşam sürelerinin olasılık dağılımının belirli bir biçimi olmaması nedeniyle Cox regresyon modeli parametrik regresyon modellerine göre daha avantajlıdır.

*İlgili yazar: durdu@hacettepe.edu.tr

Cox regresyon modeli,

$$h(t; x) = h_0(t)e^{\beta x} \quad (1)$$

biçimindedir. Burada x açıklayıcı değişkenler vektörü, β regresyon katsayıları vektörü, $h_0(t)$ ise temel tehlike fonksiyonudur (Cox, 1972). Ancak veri kümesinin dağılımı belirli bir olasılık dağılımına uyuyorsa, bu durumda parametrik regresyon modelleri Cox regresyon modeline göre daha etkili parametre tahminleri verir (Cox ve Oakes, 1984; Qi, 2009).

Yaşam çözümlemesinde parametrik orantılı tehlikeler (OT) modelleri de uygulanabilir olmasına rağmen, bu modellerle kullanılacak daha az olasılık dağılımı vardır. Bu durumda, yaşam çözümlemesi verileri için HBS modelleri, OT modellerinin de alternatifini olarak kullanılır. OT modellerinde açıklayıcı değişkenlerin tehlike üzerine etkisi ölçülürken, HBS modellerinde tehlike yerine yaşam süresi üzerine açıklayıcı değişkenlerin direk etkisi değerlendirilebilir (Qi, 2009; Zeng ve Lin, 2007).

HBS modelde (X_1, X_2, \dots, X_p) açıklayıcı değişkenleri için, yaşam olasılıkları

$$S(t/x) = S_0(t/\eta(x)) \quad (2)$$

biçimindedir. Burada, $S_0(t)$ temel yaşam fonksiyonu, η hızlandırma faktörüdür. Hızlandırma faktörü, $\eta(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_p x_p)$ biçiminde ifade edilir.

Yaşam fonksiyonu ve tehlike fonksiyonu ilişkisine göre, (X_1, X_2, \dots, X_p) açıklayıcı değişkenli bir birimin tehlike fonksiyonu

$$h(t/x) = [1/\eta(x)]h_0[t/\eta(x)] \quad (3)$$

biçiminde verilir. HBS modelinin karşılık gelen logaritmik doğrusal formu

$$\log T_i = \mu + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_p X_{pi} + \sigma \varepsilon_i$$

biçimindedir. Burada, μ kesişim, σ ölçek parametresi ve ε_i ise belirli bir dağılıma sahip olduğu varsayılan rastgele bir değişkendir (Cox ve Oakes, 1984; Collett, 2003; Qi, 2009).

ε_i 'nin her dağılımı için, T 'nin karşılık gelen bir dağılımı vardır. Bu dağılımlar Tablo 1'de verilmiştir. HBS modelleri, üstel HBS modeli, Weibull HBS modeli, log-lojistik HBS modeli, log-normal HBS modeli ve gamma HBS modeli olarak sınıflandırılır (Collett, 2003; Lawless, 1982; Qi, 2009). HBS modelleri, ε_i 'nin ya da $\log T$ 'nin dağılımından ziyade T 'nin dağılımıyla isimlendirilmektedir.

Tablo 1: Parametrik HBS modelleri

ε 'nin Dağılımı	T 'nin Dağılımı
Uç değer (1 parametre)	Üstel
Uç değer (2 parametre)	Weibull
Lojistik	Log-lojistik
Normal	Log-normal
Log-Gamma	Gamma

T_i 'nin yaşam fonksiyonu, ε_i 'nin yaşam fonksiyonu ile aşağıdaki gibi ifade edilebilir:

$$S_i(t) = P(T_i \geq t) = S_{\varepsilon_i} \left(\frac{\log t - \mu - \alpha x}{\sigma} \right) \quad (4)$$

HBS modelinin etki büyüklüğü $\exp(\alpha_i)$ 'dir. $\exp(\alpha_i)$; iki grubun beklenen yaşam sürelerinin tahmini oranı olarak yorumlanır. Diğer açıklayıcı değişkenler kontrol altında tutulduğunda x_i açıklayıcı değişkenin iki düzeyi ($x_i=0,1$) bu değerle karşılaştırılabilir. Açıklayıcı değişken için bu değer birden büyük olması, bu değişkenin olay süresini uzattığını gösterir (Collett, 2003; Qi, 2009).

2.1. Hızlandırılmış başarısızlık süresi model tahmini

HBS modellerinde en çok olabilirlik yöntemi kullanılır. t_1, t_2, \dots, t_n gözlenen yaşam süreleri için olabilirlik fonksiyonu;

$$L(\alpha, \mu, \sigma) = \prod_{i=1}^n \{f_i(t_i)\}^{\delta_i} \{S_i(t_i)\}^{1-\delta_i} \quad (5)$$

biçimindedir. Burada, $f_i(t_i)$ ve $S_i(t_i)$, sırasıyla i . birimin t_i zamanındaki yoğunluk ve yaşam fonksiyonudur. δ_i ise i . gözlemin durdurma göstergesidir. Log-olabilirlik fonksiyonu ise

$$\log L(\alpha, \mu, \sigma) = \sum_{i=1}^n \{ -\delta_i \log(\sigma t_i + \delta_i \log_{\varepsilon_i}(z_i) + (1 - \delta_i) \log S_{\varepsilon_i}(z_i)) \} \quad (6)$$

ile verilir. Burada, $z_i = (\log t_i - \mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi})/\sigma$ biçimindedir. $(p+2)$ tane bilinmeyen parametrenin $(\mu, \sigma, \alpha_1, \alpha_2, \dots, \alpha_p)$ en çok olabilirlik tahminleri log olabilirlik fonksiyonu en büyüklenerek bulunabilir (Wei, 1992; Qi, 2009). Huang, Ma ve Xie (2006), Zeng ve Lin (2007) ve Cai, Huang ve Tian (2009), HBS modelindeki regresyon katsayıları için tahmin ediciler önermişlerdir.

2.2. Weibull HBS model

Yaşam süresi T , ölçek parametresi λ , biçim parametresi γ ile $W(\lambda, \gamma)$ dağılımına sahip olsun. HBS modeli altında Eşitlik (3)'den, i . birim için tehlike fonksiyonu

$$h_i(t) = [1/\eta_i(x)]h_0[t/\eta_i(x)] = 1/[\eta_i(x)]^\gamma \lambda \gamma (t)^{\gamma-1} \quad (7)$$

biçimindedir. Burada, p açıklayıcı değişkenli i. birim için $\eta_i = \exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi})$ olur. i. birimin yaşam süresi $W(1/[\eta_i(x)]^\gamma, \lambda, \gamma)$ dağılımlıdır. Weibull dağılımı OT ve HBS özelliğine sahiptir.

Weibull modelinde $\gamma=1$ olduğunda üstel model elde edilir.

Eşitlik (4)'den, Weibull modelinin yaşam fonksiyonunun HBS ile gösterimi

$$S_i(t) = \exp \left[-\exp \left(\frac{-\mu - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \dots - \alpha_p x_{pi}}{\sigma} \right) t^{1/\sigma} \right] \quad (8)$$

biçimindedir.

Weibull modelinin yaşam fonksiyonunun OT ile gösterimi ise

$$S_i(t) = \exp \{ -\exp(\beta_1 x_{1i} + \dots + \beta_p x_{pi}) \lambda t^\gamma \} \quad (9)$$

biçimindedir.

Eşitlik (8) ve Eşitlik (9) karşılaştırıldığında OT modelindeki λ, γ, β_j , HBS modelindeki μ, σ, α_j parametreleri ile

$$\lambda = \exp(-\mu/\sigma), \gamma = 1/\sigma, \beta_j = -\alpha_j/\sigma \quad (10)$$

biçiminde ifade edilebilir.

Weibull modelinin tehlike fonksiyonunun HBS gösterimi

$$h_i(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1} \exp \left(\frac{-\mu - \alpha_1 x_{1i} - \dots - \alpha_p x_{pi}}{\sigma} \right) \quad (11)$$

şeklindedir (Collett, 2003; Qi, 2009).

2.3. Log-lojistic HBS model

Log-lojistik yaşam fonksiyonu ve temel tehlike fonksiyonu sırasıyla,

$$S(t) = \frac{1}{1+e^{\theta t^k}}, \quad h_0(t) = \frac{e^{\theta k t^{k-1}}}{1+e^{\theta t^k}} \quad (12)$$

biçiminde ifade edilir. Burada θ ve k ($k>0$) bilinmeyen parametrelerdir. $k \leq 1$ olduğunda, tehlike hızı monoton olarak azalır, $k > 1$ olduğunda ise sıfırdan maksimuma artar ve sonra sifira azalır.

Yaşam süreleri θ ve k parametreleri ile log-lojistik dağılıma sahip olsun. Bu durumda, Eşitlik (3)'den, HBS modeli altında, i. birim için tehlike fonksiyonu

$$h_i(t) = (1/\eta_i) h_0(t/\eta_i) = \frac{e^{\theta - k \log \eta_i} t^{k-1}}{1+e^{\theta - k \log \eta_i} t^k} \quad (13)$$

biçimindedir. Burada, p açıklayıcı değişkenli i. birim için $\eta_i = \exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi})$ olur. Bu nedenle, i. birimin yaşam süresi, $\theta - k \log \eta_i$ ve k

parametreleriyle log-lojistik dağılımlıdır. Log-lojistik dağılım HBS özelliğine sahiptir.

Temel yaşam fonksiyonu, θ ve k bilinmeyen parametreleri ile $S_0(t) = \{1 + e^{\theta t^k}\}^{-1}$ ise, bu durumda t zaman sonrasında yaşamın temel odds'u

$$\frac{S_0(t)}{1-S_0(t)} = e^{-\theta t^k} \quad (14)$$

ile verilir.

i. birimin yaşam fonksiyonu

$$S_i(t) = \frac{1}{1+e^{\theta - k \log \eta_i} t^k} \quad (15)$$

biçimindedir. Bu nedenle, t zaman sonrasında yaşayan i. birimin odds'u

$$\frac{S_i(t)}{1-S_i(t)} = e^{\log \eta_i - \theta t^k} \quad (16)$$

olur.

İki grup çalışmasında, Eşitlik (16) kullanılarak t zaman sonrasında yaşayan i. birimin oddsunun logaritması

$$\log \left[\frac{S_i(t)}{1-S_i(t)} \right] = \beta x_i - \theta - k \log t \quad (17)$$

biçiminde elde edilir. Burada, x_i kategorik değişkenin değeridir ve bir grupta 1, diğer grupta 0 değerini alır. Eşitlik (4) kullanılarak log-lojistik modelin yaşam fonksiyonunun HBS gösterimi

$$S_i(t) = \left[1 + t^{1/\sigma} \exp \left(\frac{-\mu - \alpha_1 x_{1i} - \dots - \alpha_p x_{pi}}{\sigma} \right) \right]^{-1} \quad (18)$$

ile verilir.

Eşitlik (15) ve Eşitlik (18) karşılaştırıldığında $\theta = -\mu/\sigma$, $k = \sigma^{-1}$ olmaktadır.

Yaşam fonksiyonu ve tehlike fonksiyonu ilişkisine göre i. birimin tehlike fonksiyonu

$$h_i(t) = \frac{1}{\sigma t} \left[1 + t^{-1/\sigma} \exp \left(\frac{\mu + \alpha_1 x_{1i} + \dots + \alpha_p x_{pi}}{\sigma} \right) \right]^{-1} \quad (19)$$

biçimindedir (Collett, 2003; Qi, 2009).

2.4. Log-normal HBS model

Yaşam süreleri log-normal dağılıma sahip olduğunda, temel yaşam fonksiyonu ve temel tehlike fonksiyonu sırasıyla aşağıdaki gibidir:

$$S_0(t) = 1 - \Phi \left(\frac{\log t - \mu}{\sigma} \right), \quad h_0(t) = \frac{\phi \left(\frac{\log t}{\sigma} \right)}{\left[1 - \Phi \left(\frac{\log t}{\sigma} \right) \right] \sigma t} \quad (20)$$

Burada μ ve σ parametreler, $\phi(x)$ olasılık yoğunluk fonksiyonu, ve $\Phi(x)$ standart normal dağılımın birikimli yoğunluk fonksiyonudur. i . birimin yaşam fonksiyonu;

$$S_i(t) = S_0(t/\eta_i) = 1 - \Phi\left(\frac{\log t - \alpha'x_i - \mu}{\sigma}\right) \quad (21)$$

biçimindedir. Burada, $\eta_i = \exp(\alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi})$ 'dir. i . birimin log yaşam süresi $(\mu + \alpha'x_i, \sigma)$ parametreleriyle normal dağılımlıdır. Log-normal dağılım HBS özelliğine sahiptir ve tehlike fonksiyonu,

$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}} \left[\frac{-1}{2\sigma^2} (\log(t) - \mu)^2 \right]}{1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)} \quad (22)$$

biçimindedir (Qi, 2009).

2.5. Gamma HBS model

Yaşam çözümlemesinde iki farklı gamma modeli vardır. Bunlar standart (2 parametrelili) ve genelleştirilmiş (3 parametrelili) gamma modelleridir. Burada gamma modeli olarak genelleştirilmiş gamma modeli kullanılmıştır. λ, α, γ parametrelili genelleştirilmiş gamma dağılımının olasılık yoğunluk fonksiyonu;

$$f(t) = \frac{\alpha\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha\gamma-1} \exp[-(\lambda t)^\alpha] \quad (23)$$

biçimindedir. Burada, γ dağılımın biçim parametresidir ve $t > 0, \gamma > 0, \lambda > 0, \alpha > 0$ 'dır. Genelleştirilmiş Gamma dağılımının yaşam fonksiyonunun ve tehlike fonksiyonunun kapalı formu yoktur. Üstel, Weibull ve log-normal modeller, genelleştirilmiş gamma modelinin özel durumlarıdır. Genelleştirilmiş gamma dağılımında $\gamma = \alpha = 1$ olduğunda üstel dağılım, $\gamma = 1$ olduğunda Weibull dağılım, $\gamma \rightarrow \infty$ olduğunda log-normal dağılım elde edilir (Qi, 2009).

2.6. Model seçim yöntemleri

Gözlenen verilerin parametrik bir dağılıma uygun olup olmadığını kontrol etmek için grafiksel yöntemler kullanılabilir. Grafikler, örneklemin homojen bir kitleden çekildiği varsayımına dayanır ancak açıklayıcı değişkenleri dikkate almaz. Bu nedenle, uygulamalarda grafik yöntem çok güvenilir değildir. Modellerin uygunluğunu kontrol etmek için başka yöntemler de vardır.

HBS modellerini karşılaştırmada istatistiksel testler ya da istatistiksel kriterler kullanılabilir. İç içe modeller olabilirlik oran (LR) testi kullanılarak karşılaştırılabilir. Üstel model, Weibull model ve log-normal model, gamma modelinin içindedir. İç içe olmayan modelleri karşılaştırmak için Akaike bilgi kriteri (AIC) kullanılabilir. Bu bilgi kriteri

$$AIC = -2\log L + 2(k + c) \quad (24)$$

biçiminde tanımlanır. Burada $\log L$, log olabilirlik, k modeldeki açıklayıcı değişken sayısı, c ise modeldeki yardımcı parametre sayısıdır. En küçük AIC değerine sahip olan model en iyi modeldir.

Parametrik yaşam modellerinde modelin uyum iyiliği Cox-Snell artıkları yöntemi kullanılarak da incelenebilmektedir. t_i , gözlenen yaşam süresine sahip i . birimin Cox-Snell artığı (r_{c_i}) aşağıdaki gibi tanımlanabilir:

$$r_{c_i} = \hat{H}(t_i/x_i) = -\log[\hat{S}(t_i/x_i)] \quad (25)$$

Burada, t_i i . birimin gözlenen yaşam süresi, x_i i . birimin açıklayıcı değişken değerleri vektörü, $\hat{S}(t_i)$ uygun model üzerinden yaşam fonksiyonunun tahminidir.

Cox-Snell artıkları her hangi bir parametrik modele uygulanabilir. Belirli bir HBS modele dayalı artığın karşılık gelen formu elde edilebilir. Örneğin, Weibull HBS modeli altında, $S_{\epsilon_i}(\epsilon) = \exp(-e^\epsilon)$ olduğu için, Cox-Snell artığı

$$r_{c_i} = -\log\{\hat{S}(t_i)\} = -\log S_{\epsilon_i}(r_{s_i}) = \exp(r_{s_i}) \quad (26)$$

biçiminde olur.

Eğer model uygunsuzsa, $\log(-\log S(r_{c_i}))$ 'nin $\log r_{c_i}$ 'ye karşı grafiği, orijin boyunca bir eğimli bir doğrudur ((Collett, 2003; Lee ve Wang, 2003; Qi, 2009; Cox ve Snell, 1968).

Bu çalışmada kullanılan veriler, Ocak 1990 ile Kasım 1995 tarihleri arasında Ankara Onkoloji Hastanesi'nde mide kanseri tanısı konan 118 hastaya aittir. Burada amaç, bu verilerin tıbbi yorumları değil, hızlandırılmış başarısızlık süresi modellerinin uygulanabilir olduğunu göstermektir. Hastalar ölene kadar ya da Nisan 1996'ya kadar izlenmişlerdir. Açıklayıcı değişkenler olarak yaş, cinsiyet, kilo kaybı (semptom), anemi (hb), lenf nodu diseksiyonunun genişliği, tümörün midedeki lokalizasyonu, hastalığın evresi, adjuvan kemoterapi alınmıştır.

3. Bulgular

118 hastanın 52'si (%44.1) ölmüş, 66'sı (%55.9) ise çalışmanın sonunda yaşamını sürdürmeye devam etmiştir. Mide kanserinden ölümler başarısızlık olarak alınmış, yaşayanlar ise durdurulmuş olarak kabul edilmiştir. Ortalama yaş 56.7 ± 11.34 (29-84 yıl) olarak bulunmuştur. Kaplan ve Meier (1958)'in tahmin edicisi kullanılarak 3 yıllık yaşam olasılığı %42.5 olarak elde edilmiştir. Ortanca yaşam süresi ise 29 ± 6 'dır. Kullanılan değişkenler ve düzeyleri Tablo 2'de verilmiştir.

Tablo 2. Açıklayıcı Değişkenler ve Düzeyleri

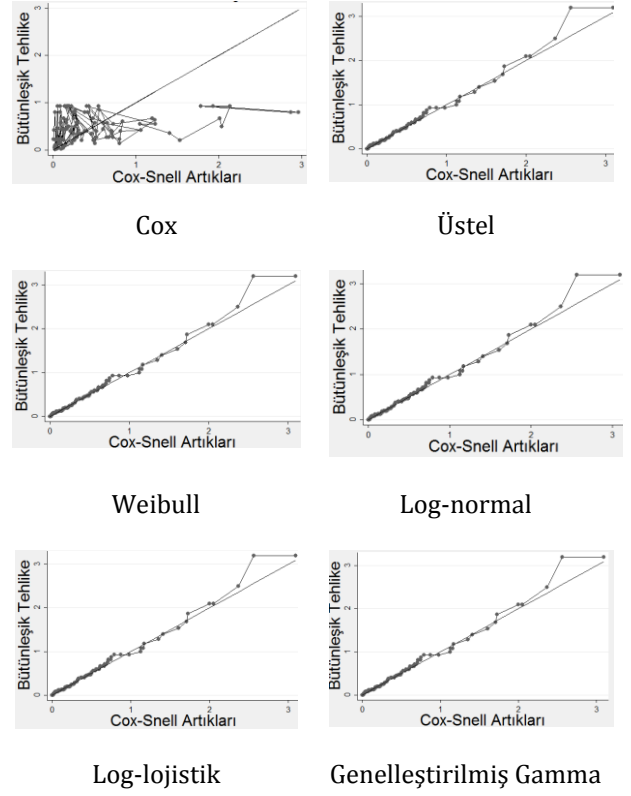
Değişken	Değişken Düzeyleri	n	%
Cinsiyet (Cins)	1. Erkek	69	58.5
	2. Kadın	49	41.5
Kilo Kaybı (Semptom)	1. Yok	90	76.3
	2. Var	28	23.7
Anemi (Hb)	1. Yok	31	26.3
	2. Var	87	73.7
Lenf nodu diseksiyonunun genişliği (Dissek)	1. D ₀₋₁	62	52.5
	2. D ₂₋₃	56	47.5
Tümörün midedeki lokalizasyonu (Lokal)	1. Üst üçlük	21	17.8
	2. Orta üçlük	25	21.2
	3. Alt üçlük	62	52.5
	4. Tüm mide	10	8.5
Hastalığın evresi (Evre)	1. Evre1+Evre2	30	25.4
	2. Evre3	65	55.1
	3. Evre4	23	19.5
Adjuvan kemoterapi (Kt)	1. Yok	34	28.8
	2. Var	84	71.2

Çalışmada STATA 12 paket programı kullanılmış ve mide kanseri verisi için Cox regresyon modeli ve hızlandırılmış başarısızlık süresi modelleri ile çözümleme yapılmıştır.

Orantılı tehlikeler varsayımının sağlanıp sağlanmadığı Schoenfeld (1982) tarafından önerilen Schoenfeld artıkları kullanılarak test edilmiş ve tüm değişkenler için $p > 0.05$ bulunduğundan tüm değişkenlerin orantılı tehlikeler varsayımını sağladığı görülmüştür. Ancak orantılı tehlikeler varsayımı grafik yöntemle de incelendiğinde (t'ye karşı $\ln(-\ln S(t))$ grafiği) lokal ve evre değişkenlerinde bazı düzeylerin çakıştığı, tüm düzeylerin birbirine paralel olmadığı görülmüştür.

Regresyon modellerinde modelin uygunluğu Cox-Snell artıkları kullanılarak incelenebilmektedir. Mide kanseri verisi için Cox regresyon modeli için ve her bir hızlandırılmış başarısızlık süresi dağılımına ilişkin Cox-Snell artık grafikleri Şekil 1'de verilmiştir.

Cox-Snell artık grafiklerine göre Cox regresyon modelinin uygun olmadığı, HBS modellerinin uygun olduğu açık olarak görülmektedir. HBS modellerinden hangisinin en uygun olduğuna karar verilebilmek için model karşılaştırma kriterleri kullanılmıştır. HBS modellerinden Üstel, Weibull, Log-normal, Log-lojistik ve Genelleştirilmiş Gamma parametrik regresyon modellerine ait $-2\log L$, AIC ve iç içe modeller durumunda kullanılan LR'nin Gamma dağılımına karşı test sonuçları elde edilmiştir. Sonuçlar Tablo 3'de verilmiştir.

**Şekil 1.** Cox-Snell artık çizimleri**Tablo 3.** HBS modelleri için model karşılaştırma kriterleri

Model	-2LogL	AIC	LR
Üstel	190.56	214.56	19.73
Weibull	174.20	200.20	3.37
Log-normal	170.98	196.98	0.16
Log-lojistik	172.26	198.26	İç içe model değil
Genelleştirilmiş Gamma	170.84	198.84	-

AIC ya da LR'ye göre HBS modellerinden en uygunu log-normal regresyon modeli olduğundan bu model için elde edilen sonuçlar Tablo 4'de verilmiştir.

Tablo 4. Log-normal regresyon çözümlemesinin sonuçları

Değişken	$\hat{\beta}$	Std. Hata	%95 Güven Aralığı	p-değeri
Yaş	0.01	0.01	(-0.002, 0.03)	0.096
Cins1	-0.11	0.21	(-0.52, 0.31)	0.616
Semptom2	0.21	0.23	(-0.23, 0.65)	0.347
Hb1	-0.03	0.24	(-0.49, 0.44)	0.907
Dissek1	-0.72	0.21	(-1.13, -0.31)	0.001
Lokal2	-0.85	0.32	(-1.47, -0.23)	0.007
Lokal3	-0.48	0.29	(-1.04, 0.08)	0.092
Lokal4	-0.70	0.38	(-1.45, 0.04)	0.063
Evre1	1.83	0.35	(1.16, 2.51)	0.000
Evre2	0.57	0.25	(0.08, 1.06)	0.023
Kt2	0.84	0.21	(0.42, 1.25)	0.000
Ölçek	0.79	0.08	(0.65, 0.97)	

Tehlike fonksiyonunda gözlemlenemeyen faktörlerden açığa çıkan bireysel farklılıkların

orantısız tehlikelere neden olabileceği düşünülerek zayıflık modeli mide kanseri verisi için kullanılabilir. Uygun bulunan log-normal regresyon modeli dikkate alınarak ve zayıflık terimi için en çok kullanılan dağılım olan Gamma dağılımı kullanılarak log-normal zayıflık modeli elde edilmiş ve zayıflık terimi θ 'nın modele dahil edilip edilmemesine karar vermek için olabilirlik oran testi kullanılmış ve p-değeri 1.00 bulunmuştur. Buradan zayıflık teriminin model üzerinde etkisi olmadığı ve modele dahil edilmesi gerekmediği sonucuna varılmıştır. Buna göre log-normal zayıflık modeli yerine log-normal regresyon modelinin kullanılmasının daha uygun olduğu söylenebilmektedir. Log-normal regresyon modeli için anlamlı değişkenlerin yer aldığı modeli elde edebilmek için adimsal regresyon yöntemi uygulanmış ve elde edilen sonuçlar Tablo 5'de verilmiştir.

Tablo 5. Log-normal regresyon çözümlemesinin adimsal seçim sonuçları

Değişken	$\hat{\beta}$	Std. Hata	%95 Güven Aralığı	p-değeri
Dissek1	-0.66	0.21	(-1.06, -0.25)	0.002
Evre1	1.36	0.27	(0.82, 1.89)	0.000
Kt2	0.73	0.21	(0.31, 1.16)	0.001
Ölçek	0.86	0.09	(0.70, 1.05)	

Tablo 5 incelendiğinde lenf nodu diseksiyonunun genişliği, hastalığın evresi ve adjuvan kemoterapi değişkenlerinin yaşam süresini etkileyen önemli faktörler olduğu %95 güven düzeyinde söylenebilmektedir. Lenf nodu diseksiyonunun genişliği $D_{0.1}$ olanların ortalama yaşam süresinin $D_{2.3}$ olanlara göre yaklaşık 2 ($\exp(\hat{\beta})$) kat daha kısa olduğu, hastalığın evresi 1 ve 2 olanların ortalama yaşam süresinin 4. evrede olanlardan yaklaşık 4 kat daha uzun olduğu, kemoterapi alanların ortalama yaşam süresinin almayanlardan 2 kat daha uzun olduğu ifade edilebilir.

4. Tartışma ve Sonuç

Cox regresyon modeli yaşam verilerinin çözümlenmesinde en yaygın kullanılan modeldir. Ancak Cox regresyon modeli kullanılmadan önce orantılı tehlikeler varsayımı incelenmelidir. Orantılı tehlikeler varsayımı sağlanmadığında ve yaşam süresinin ya da parametrik regresyon modelinde hata teriminin dağılımına ilişkin bilgi olması durumunda dağılıma uygun HBS modeli yaşam verilerinin çözümlemesi için alternatif bir modeldir. HBS modelleri, belirli durumlar altında Cox regresyon modelinden daha etkili parametre tahminleri sağlamaktadır.

HBS modellerinde orantılı tehlikeler varsayımına ihtiyaç yoktur ve bu modellerde yaşam süresi üzerine açıklayıcı değişkenlerin doğrudan etkileri modellenmektedir. Bu nedenle sonuçları yorumlamak daha kolaydır.

Bu çalışmada HBS modelleri incelenmiş ve mide kanseri verilerine uygulanmıştır. Bu veriler için HBS modellerinden log-normal regresyon modelinin en uygun model olduğu ve bu modelden lenf nodu diseksiyonunun genişliği, hastalığın evresi ve adjuvan kemoterapi değişkenlerinin yaşam süresini etkileyen önemli faktörler olduğu sonucuna varılmıştır.

Kaynaklar

Aalen, O.O. 1994. Effects of Frailty in Survival Analysis. *Statistical Methods in Medical Statistics*, 3, 227-243.

Cai, T., Huang, J., Tian, L. 2009. Regularized Estimation for the Accelerated Failure Time Model. *Biometrics*, 65, 394-404.

Collett, D. 2003. *Modelling Survival Data in Medical Research*, Chapman and Hall, New York, 347 pp.

Cox, D.R., Snell, E.J. 1968. A General Definition of Residuals with Discussion. *Journal of the Royal Statistical Society, Series B*, 30, 248-275.

Cox, D.R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*, 34(2), 187-220.

Cox, D.R., Oakes, D. 1984. *Analysis of Survival Data*. Chapman and Hall, London, 193 pp.

Hougaard, P., Myglegaard, P., Borch-Johnsen, K. 1994. Heterogeneity Models of Disease Susceptibility, with Application to Diabetic Nephropathy. *Biometrics*, 50, 1178-1188.

Hougaard, P. 1995. Frailty Models for Survival Data. *Lifetime Data Analysis*, 1, 255-273.

Huang, J., Ma, S., Xie, H. 2006. Regularized Estimation in the Accelerated Failure Time Model with High-Dimensional Covariates. *Biometrics*, 62, 813-820.

Kaplan, E., Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of American Statistical Association*, 53, 457-481.

Kay, R., Kinnersley, N. 2002. On the Use of the Accelerated Failure Time Model as an Alternative to the Proportional Hazards Model in the Treatment of Time to Event Data: A Case Study in Influenza. *Drug Information Journal*, 36, 571-579.

Keiding, N., Andersen, P.K., Klein, J.P. 1997. The Role of Frailty Models and Accelerated Failure Time Models in Describing Heterogeneity Due to Omitted Covariates. *Statistics in Medicine*, 16, 215-224.

Klein, J.P., Moeschberger, M.L. 1997. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York, 502 pp.

Klembaum, D.G. 1996. *Survival Analysis: A Self Learning Text*. Springer, New York, 324 pp.

Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data Analysis*, Wiley and Sons, New York, 580 pp.

Lee, E.T., Wang, J.W. 2003. *Statistical Methods for Survival Data Analysis*, Wiley and Sons, New York, 513 pp.

Oakes, D. 1977. The Asymptotic Information in Censored Survival Data. *Biometrika*, 64, 441-448.

Orbe, J., Ferreira, E., Nunez-Anton, V. 2002. Comparing Proportional Hazards and Accelerated Failure Time Models for Survival Analysis. *Statistics in Medicine*, 21, 3493-3510.

Qi, J. 2009. Comparison of Proportional Hazards and Accelerated Failure Time Models. MSc Thesis, Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, 79p.

Schoenfeld, D. 1982. Partial Residuals for the Proportional Hazards Regression Model. *Biometrika*, 69, 239-241.

Wei, L.J. 1992. The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis. *Statistics in Medicine*, 11, 1871-1879.

Zeng, D., Lin, D. 2007. Efficient Estimation for the Accelerated Failure Time Model. *Journal of the American Statistical Association*, 102, 1387-1396.