



Orantısız hazardlar için parametrik ve yarı parametrik yaşam modelleri

Nihal Ata

Hacettepe Üniversitesi,
Fen Fakültesi,
İstatistik Bölümü,
Beytepe, 06800, Ankara
nihalata@hacettepe.edu.tr

Durdu Karasoy

Hacettepe Üniversitesi,
Fen Fakültesi,
İstatistik Bölümü,
Beytepe, 06800, Ankara
durdu@hacettepe.edu.tr

M. Tekin Sözer

Hacettepe Üniversitesi,
Fen Fakültesi, Aktüerya
Bilimleri Bölümü,
Beytepe, 06800, Ankara
sozer@hacettepe.edu.tr

Özet

Yaşam verileri için en çok kullanılan regresyon modeli Cox regresyon modelidir. Bu model orantılı hazardlar varsayımına karşı duyarlıdır. Bu varsayımın sağlanmadığı durumlarda farklı yaşam modellerinin kullanılması önerilmektedir. Bu çalışmada, orantılı hazardlar varsayımının sağlanmadığı durumlarda parametrik yaşam modelleri (üstel, weibull, log-lojistik, log-normal, Gompertz, Gamma regresyon modelleri) ve yarı parametrik yaşam modelleri (Cox regresyon modeli ve uzanımları) incelenmiştir. Orantılı hazardlar varsayımını sağlamayan gerçek yaşam verileri kullanılarak, parametrik ve yarı parametrik yaşam modellerinin uygulaması yapılmıştır.

Anahtar Sözcükler: Cox regresyon modeli; Orantısız hazardlar; Parametrik yaşam modelleri.

Abstract

Parametric and nonparametric survival models for nonproportional hazards

The Cox regression model is the most commonly used regression model for survival data and sensitive to proportional hazards. In the violation of proportional hazards, usage of different survival models is suggested. In this study, the parametric survival models (exponential, weibull, log-logistic, log-normal, Gompertz, Gamma regression models) and semi-parametric survival models (Cox regression model and its extensions) were investigated in case of nonproportional hazards. Parametric and nonparametric survival models are applied to real survival data with nonproportional hazards.

Key Words : Cox regression model; Nonproportional hazards; Parametric survival models.

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 (censored) veri için tasarlanmış istatistiksel yöntemler bütünü olmasıdır. Gözlem süresi boyunca çalışmada yer alan birimlerin tamamı başarısızlık ile karşılaşmamış olabilir. Bu durumda bu birimler yaşam çözümlemesinde durdurulmuş olarak nitelendirilmektedir.

Yaşam verileri için en çok kullanılan regresyon modeli Cox regresyon modelidir. Bu model orantılı hazardlar varsayımına karşı duyarlıdır. Bu varsayımın sağlanmadığı durumlarda araştırmacılar basit Cox regresyon modelinden özel matematiksel yapıya sahip yaşam modellerine kadar birçok yöntem geliştirmeye çalışmaktadırlar. Bu çalışmada, orantılı hazardlar varsayımını sağlamayan akciğer kanseri verileri kullanılarak yarı parametrik ve parametrik yaşam modelleri incelenmiştir.

2. Orantılı hazardlar varsayımı

Orantılı hazardlar varsayımı, hazard oranının zamana karşı sabit olması ya da bir bireyin hazard hızının diğ er bireyin hazard hızına orantılı olması anlamına gelmektedir [15]. $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_p^*)$ ve $\mathbf{x} = (x_1, x_2, \dots, x_p)$ iki bireye ait açıklayıcı deđ işkenler vektörü olmak üzere hazard oranı,

$$\exp \left[\sum_{j=1}^p \hat{\beta}_j (x_j^* - x_j) \right] \quad (1)$$

biçimindedir [9]. Eşitlik (1)'de görüldüğü gibi hazard oranı t 'yi içermemektedir, yani zamandan bağımsızdır. Bu durumda orantılı hazardlar varsayımı sağlanmaktadır. Eşitlik (1)'de verilen hazard oranının zamanın bir fonksiyonu olması durumunda ise orantılı hazardlar varsayımı sağlanmamaktadır.

Yaşam çözümlemesinde, orantılı hazardlar varsayımının incelenmesi için kullanılan birçok yöntem vardır. Bunlardan en yaygın olanları, grafiksel yöntemler [log(-log) yaşam eğrileri [8], Arjas grafikleri [1], zamana bağı açıklayıcı deđ işkenlerin kullanılması yöntemi, Schoenfeld artıkları ile yaşam süresinin rankının korelasyon testi [14] yöntemi biçiminde sıralanmaktadır.

3. Orantısız hazardlar için parametrik ve yarı parametrik yaşam modelleri

Bağımlı deđ işken olan yaşam süresi üzerinde açıklayıcı deđ işkenlerin etkilerinin araştırıldığı regresyon modelleri yaşam çözümlemesinde önemli bir yere sahiptir.

Yaşam süresi dağılımının belirlenemediği durumlarda yarı parametrik regresyon modellerinin (Cox regresyon modeli, tabakalandırılmış Cox regresyon modeli, genişletilmiş Cox regresyon modeli), belirlenebildiği durumlarda ise parametrik regresyon modellerinin (Üstel, Weibull, Log-lojistik, Log-normal, Gompertz, Gamma regresyon modelleri) kullanılması uygun olmaktadır. Bu modeller aşağıda verilmiştir.

3.1. Yarı parametrik yaşam modelleri

3.1.1. Cox regresyon modeli

Yaşam süresine ilişkin faktörlerin hazard fonksiyonu üzerindeki etkilerinin çarpımsal olduğu modeller, yaşam süresi verilerinin çözümlemesinde önemli bir rol oynamaktadır. İlk kez 1972'de Cox tarafından ele alınan Cox regresyon modeli de bu modellerden biridir.

T , bir birimin yaşam süresini temsil eden sürekli raslantı deđ işkeni ve \mathbf{x} bu birimle ilgili açıklayıcı deđ işkenler vektörü olmak üzere orantılı hazardlar varsayımı altında \mathbf{x} bilindiğinde T 'nin hazard fonksiyonu,

$$h(t / \mathbf{x}) = h_0(t) \exp(\beta' \mathbf{x}) \quad (2)$$

biçimindedir. Burada, β regresyon katsayıları vektörü ve $h_0(t)$ ise $x = 0$ olan bir birimin temel hazard fonksiyonu olarak tanımlanmaktadır [4]. Cox regresyon modeli için parametreler olabilirlik fonksiyonu (L) maksimize edilerek tahmin edilmektedir.

Cox (1972)'un önerdiği yaklaşım, $h_0(t)$ için özel bir biçimin varsayılmadığı, dağılımdan bağımsız bir yaklaşımdır. Yani bu yöntem, ilgili yaşam süresi dağılımına bir başka deyişle temel hazard fonksiyonu $h_0(t)$ 'ye dayanmamaktadır. Bu model hazard fonksiyonun fonksiyonel biçimini belirlemeden hazard fonksiyonu üzerindeki açıklayıcı değışkenlerin etkilerini incelemek için tasarlandığından temel hazard fonksiyonunun tahmin edilmesine gerek yoktur.

Cox regresyon modeli yaşam süresinin dağılımına ilişkin bir varsayım içermemesine rağmen, bu modelin kullanılması için de hazardların orantılı olması gerekmektedir. Hazardların orantılı olması durumunda basit Cox regresyon modelinin kullanımı, orantısız olması durumunda ise tabakalandırılmış ve genişletilmiş Cox regresyon modellerinin kullanılması uygun olmaktadır [2].

3.1.2. Tabakalandırılmış Cox regresyon modeli

Orantılı hazard varsayımını sağlamayan açıklayıcı değışkenin tabakalandırılması ile elde edilen bir modeldir. k tane değışkenin orantılı hazard varsayımını sağlamadığı ve p tane değışkenin ise bu varsayımı sağladığı düşünölsün. Orantılı hazard varsayımını sağlamayan değışkenler z_1, z_2, \dots, z_k ile, orantılı hazard varsayımını sağlayan değışkenler ise x_1, x_2, \dots, x_p ile gösterilsin. Tabakalandırılmış Cox regresyon modelini oluşturmak için, tabakalandırma yaparken kullanılmak üzere z 'lerden yeni, tek bir değışken tanımlanır ve z^* ile gösterilir. z^* in düzeyleri orantılı hazard varsayımını sağlamayan değışkenlerin düzey sayılarının çarpımı ile elde edilmektedir ve k^* tane kategorisi vardır.

Tabakalandırılmış Cox regresyon modeli etkileşimli ve etkileşimsiz tabakalandırılmış Cox regresyon modelleri olmak üzere iki alt başlıkta incelenmektedir.

3.1.2.1. Etkileşimsiz tabakalandırılmış Cox regresyon modeli

Etkileşimsiz tabakalandırılmış Cox regresyon modeli;

$$h_g(t, x) = h_{0g}(t) \exp[\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p] \quad , \quad g = 1, 2, \dots, k^* \quad (3)$$

biçimindedir. Burada g alt indisi, g . tabakayı göstermektedir. Tabaka, tabakalandırılan değışken z^* 'in farklı düzeyleri olarak adlandırılır ve tabaka sayısı k^* a eşittir. Orantılı hazard varsayımını sağlamayan açıklayıcı değışkenler (z^*) modelde açıkça yer almazken orantılı hazard varsayımını sağlayan açıklayıcı değışkenler (x) modelde yer almaktadır.

Temel hazard fonksiyonu, $h_{0g}(t)$, herbir tabaka için farklıdır. Bununla birlikte $\beta_1, \beta_2, \dots, \beta_p$ herbir tabaka için aynıdır. Bu nedenle, hazard oranı tahminleri de herbir tabaka için aynı olacaktır.

3.1.2.2. Etkileşimli tabakalandırılmış Cox regresyon modeli

i) Orantısızlığa neden olan değışken sayısı bir ise etkileşimli model iki farklı biçimde yazılabilir.

a) Orantısızlığa neden olan değışkene göre veri kümesi k farklı tabakaya ayrılabilir. Bu durumda etkileşimli model,

$$h_g(t, x) = h_{0g}(t) \exp [\beta_{1g} x_1 + \beta_{2g} x_2 + \dots + \beta_{pg} x_p] \quad , \quad g = 1, 2, \dots, k \quad (4)$$

biçiminde yazılabilir.

b) Modelde açıklayıcı de ğ işkenler ve orantısızlığa neden olan de ğ işken ile bu de ğ işkenin çarpım terimleri etkileşimli modelde yer alabilir. Bu model ise,

$$h_g(t, x) = h_{0g}(t) \exp \left[\beta_1^* x_1 + \beta_2^* x_2 + \dots + \beta_p^* x_p + \beta_{p+1}^* (x_1 \times z) + \dots + \beta_{2p}^* (x_p \times z) \right] \quad (5)$$

biçiminde yazılabilir. β^* katsayıları g tabaka indisini içermemektedir.

ii) Orantısız hazardlara neden olan de ğ işken sayısı iki ya da daha fazla ise yeni bir de ğ işken $z^* = (0, 1, 2, \dots, k^* - 1)$ tanımlanır. Burada k^* orantısızlığa neden olan de ğ işkenlerin düzey sayılarının çarpımı olarak elde edilmektedir. Orantılı hazard varsayımını sağlayan de ğ işken sayısı ise p olmak üzere etkileşimli modeli

$$(a) \quad h_g(t, x) = h_{0g}(t) \exp \left[\beta_{1g} x_1 + \dots + \beta_{pg} x_p \right] \quad g = 1, 2, \dots, k^* \quad (6)$$

$$(b) \quad h_g(t, x) = h_{0g}(t) \exp \left[\begin{array}{l} \beta_1 x_1 + \dots + \beta_p x_p + \beta_{11} (z_1^* \times x_1) + \dots + \beta_{1(k^*-1)} (z_{k^*-1}^* \times x_1) \\ + \dots + \beta_{p1} (z_1^* \times x_p) + \dots + \beta_{p(k^*-1)} (z_{k^*-1}^* \times x_p) \end{array} \right] \quad (7)$$

biçimindedir.

3.1.3. Genişletilmiş Cox regresyon modeli

x_1, x_2, \dots, x_{p_1} zamandan bağımsız de ğ işkenler ve $x_1(t), x_2(t), \dots, x_{p_2}(t)$ zamana ba ğ lı de ğ işkenler olmak üzere genişletilmiş Cox regresyon modeli,

$$h(t, x(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j x_j(t) \right] \quad (8)$$

biçimindedir. $x_j(t) = x_j \times g_j(t)$ olarak tanımlanmaktadır. Genişletilmiş Cox regresyon modeli kullanılırken verilmesi gereken en önemli karar $g_j(t)$ fonksiyonlarının biçimidir.

i) $g_j(t) = t$

ii) $g_j(t) = \log t$

iii) $g_j(t)$ adım (Heaviside, step) fonksiyonu olabilir. Bu fonksiyon kullanıldığında, hazard oranı, farklı zaman aralıkları için sabit hazard oranları vermektedir [2].

3.2. Parametrik yaşam modelleri

Cox regresyon modeli, orantılı hazardlar varsayımına karşı duyarlıdır. Temel hazard fonksiyonunun biçimi tahmin edicilerin özelliklerini etkileyebilir [3]. Temel hazard fonksiyonu belirli bir dağılım ile hesaplanamadığından Cox regresyon modeli parametrik regresyon modellerine göre daha avantajlıdır. Çünkü yaşam sürelerinin olasılık dağılımının belirli bir biçimi olmadığından esnek ve yaygın kullanıma sahiptir. Ancak veri kümesi için belirli bir olasılık dağılımı varsayımı geçerli ise, bu varsayım dayalı çıkarsamalar daha kesindir. Ayrıca parametre tahminleri ve görelî hazard ya da ortanca yaşam süresi gibi ölçümlerin tahminleri daha küçük standart hataya sahip olur [6]. Bu durumda parametrik regresyon

modellerinin Cox regresyon modeline göre daha etkili parametre tahminlerine sahip oldukları gösterilmiştir [7, 12, 13].

Açıklayıcı değişkenler vektörü $X=(X_1, \dots, X_p)$ ile gösterilsin. İstatistiksel literatürde, yaşam süresini etkileyen açıklayıcı değişkenleri parametrik regresyon modelleri ile modellemek için iki yaklaşım vardır. Birinci yaklaşım, klasik doğrusal regresyon yaklaşımına benzemektedir. Bu yaklaşımda, yaşam süresinin doğal logaritması $Y=\ln(T)$ modellenir. Y için doğrusal model,

$$Y = \mu + \beta'X + \sigma W \quad (9)$$

biçimindedir. Burada $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ regresyon katsayıları vektörü ve W hata terimidir. Hata dağılımı için en sık kullanılan seçenekler log-normal regresyon modelini veren standart normal dağılım, Weibull regresyon modelini veren uçdeğer dağılımı (extreme value) ya da log-lojistik regresyon modelini veren lojistik dağılımdır. Bu yaklaşım için kullanılan parametrik yaşam modelleri hızlandırılmış başarısızlık zamanı modelleri olarak da adlandırılmaktadır. Ancak bu yaklaşım hata dağılımları için varsayım içerdiğinden kullanım alanları daha sınırlıdır. Bu durumda ikinci yaklaşımın kullanılması daha uygun olabilir. Açıklayıcı değişkenlerin yaşam fonksiyonu üzerindeki etkilerini modellemek için kullanılan ikinci yaklaşım, açıklayıcı değişkenlerin bir fonksiyonu olarak koşullu hazard fonksiyonunu modellemektir. Buna göre koşullu hazard fonksiyonu,

$$h(t/x) = h_0(t)\varphi(\beta'x) \quad (10)$$

biçiminde yazılır [9, 10]. Burada $\varphi(\beta'x)$ açıklayıcı değişkenlerin negatif olmayan bir fonksiyonudur. Birçok uygulamada $\varphi(\beta'x) = \exp(\beta'x)$ olarak seçilmektedir. Parametrik regresyon modeli için koşullu hazard fonksiyonu belirlenirken, yaşam süresi hangi dağılıma uygunsa o dağılıma ait temel hazard fonksiyonu kullanılmaktadır.

En sık kullanılan parametrik regresyon modelleri üstel, Weibull, log-lojistik, log-normal, Gompertz, Gamma dağılımlarını içermektedir. Bu dağılımlara ait temel hazard fonksiyonları ($h_0(t)$), olasılık yoğunluk fonksiyonları ($f(t)$) ve yaşam fonksiyonları ($S(t)$) Çizelge 1'de verilmiştir.

Çizelge 1. Parametrik dağılımlar için temel hazard fonksiyonları, olasılık yoğunluk fonksiyonları ve yaşam fonksiyonları

| Dağılım | Parametre | $h_0(t)$ | $f(t)$ | $S(t)$ |
|--------------|---|--|--|---|
| Üstel | $\lambda > 0$ | λ | $\lambda \exp(-\lambda t)$ | $\exp(-\lambda t)$ |
| Weibull | $\lambda > 0$ $\alpha > 0$ | $\lambda \alpha t^{\alpha-1}$ | $\lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$ | $\exp(-\lambda t^\alpha)$ |
| Log-lojistik | $\alpha, \lambda > 0$ | $\frac{\alpha t^{\alpha-1} \lambda}{1 + \lambda t^\alpha}$ | $\frac{\alpha t^{\alpha-1} \lambda}{[1 + \lambda t^\alpha]^2}$ | $\frac{1}{1 + \lambda t^\alpha}$ |
| Log-normal | $\alpha > 0$ | $\frac{f(t)}{S(t)}$ | $\frac{\exp\left[-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2\right]}{t(2\pi)^{1/2}\sigma}$ | $1 - \Phi\left[\frac{\ln t - \mu}{\sigma}\right]$ |
| Gompertz | $\lambda > 0$ $\alpha \in (-\infty, \infty)$ | $\lambda \exp(\alpha t)$ | $\lambda \exp(\alpha t) \exp\left(\frac{\lambda}{\alpha}(1 - \exp(\alpha t))\right)$ | $\exp\left[\frac{\lambda}{\alpha}(1 - e^{\alpha t})\right]$ |
| Gamma | $\alpha, \beta, \lambda > 0$ | $\frac{f(t)}{S(t)}$ | $\frac{\alpha \lambda^\beta t^{\alpha\beta-1} \exp(-\lambda t^\alpha)}{\Gamma(\beta)}$ | $1 - I[\lambda t^\alpha, \beta]$ |

I : Tamamlanmamış Gamma fonksiyonu

Φ : Standart normal bir değişkenin birikimli dağılım fonksiyonu

μ : Ortalama, σ : Varyans

3.3. Model seçim yöntemleri

Yaşam modellerinde uygun olan modele karar verebilmek için $-2\log L$ istatistiği ya da Akaike bilgi kriteri (AIC) kullanılmaktadır. En küçük $-2\log L$ istatistiğine ya da AIC değerine sahip model en uygun model olarak belirlenmektedir [9].

Yarı parametrik yaşam modeli için Akaike bilgi kriteri,

$$AIC = -2\log L + \alpha p \quad (11)$$

biçiminde hesaplanmaktadır. Burada p modeldeki bilinmeyen β parametrelerinin sayısı ve α daha önceden belirlenen bir sabittir. α genellikle 2 ile 6 arasında değerler alır. $\alpha = 3$ olarak seçilmesi, $-2\log \hat{L}$ değerleri arasındaki farklılığı değerlendirmede %5 önem düzeyinin kullanılmasına eşdeğer olduğundan α 'nın genel kullanımı için bu değer önerilmektedir [6].

Parametrik yaşam modelleri için Akaike bilgi kriteri Eşitlik 12'deki gibi hesaplanmaktadır:

$$AIC = -2\log L + 2(p + 2 + k) . \quad (12)$$

AIC değerleri hesaplanırken üstel regresyon modeli için $k=0$, Weibull, log-lojistik ve log-normal regresyon modelleri için $k=1$ ve Gamma regresyon modeli için $k=2$ olarak alınmaktadır [9].

Parametrik yaşam modellerinde modelin uyum iyiliği ise Cox-Snell (1968) artıkları yöntemi kullanılarak incelenmektedir. t_i , gözlenen yaşam süresine sahip i . birey (durdurulmuş ya da durdurulmamış) için Cox-Snell artığı (r_i) aşağıdaki gibi tanımlanmaktadır:

$$r_i = -\log \hat{S}(t_i) , i = 1, 2, \dots, n . \quad (13)$$

Eşitlik 12'de $\hat{S}(t)$ parametrelerin en çok olabilirlik tahmin edicilerine dayanan yaşam fonksiyonu tahminidir. Gözlenen yaşam süresi t_i durdurulmuş ise, buna karşılık gelen r_i de durdurulmuştur. Birikimli hazard fonksiyonu $H(t) = -\log S(t)$ olduğundan r_i , t_i zamanında birikimli hazard fonksiyonunun değeridir. Cox-Snell artıklarının en önemli özelliği, seçilen model veriye uygunsa artıklar (r_i) $f_R(r) = e^{-r}$ olasılık yoğunluk fonksiyonu ile birim üstel dağılıma sahip olmasıdır.

$S_R(r)$, Cox-Snell artıklarının yaşam fonksiyonu olsun. Bu durumda $S_R = \int_r^\infty f_R(x) = \int_r^\infty e^{-x} dx = e^{-r}$ biçimindedir ve

$$-\log S_R(r) = -\log(e^{-r}) = r \quad (14)$$

olur. $\hat{S}_R(r)$, $S_R(r)$ 'nin Kaplan-Meier tahmin edicisidir. r_i 'nin $-\log \hat{S}_R(r)$ 'ye karşı grafiği bir eğim ve sıfır kesişim ile düz bir doğru ise parametrik modelin ilgilenilen veri kümesi için uygun olduğu sonucuna ulaşılır [11].

4. Uygulama

Uygulamada orantılı hazardlar varsayımını sağlamayan 236 akciğer kanseri hastalarına ait veriler kullanılarak yarı-parametrik ve parametrik yaşam modelleri incelenmiştir. Hastalar, ameliyat olduktan sonra hastalıklarının ilk nüks etmesine kadar geçen süre boyunca izlenmiştir. Akciğer kanseri hastalarının yaşam sürelerini etkileyen faktörler yaşam çözümlemesi yöntemleri kullanılarak belirlenmeye çalışılmıştır. Hastaların ameliyat olduğu tarihten hastalığın ilk nüks etmesine kadar geçen süre (min=1 ay,

max=93 ay) yaşam süresi olarak alınmıştır. Hastalığın nüksetmesi başarısızlık olarak ifade edilmiştir. Hastalığı nüksetmeyen hastalar durdurulmuş olarak tanımlanmıştır. Hastaların izlenme süresi sona erdiğ inde 236 hastadan 94'ünde (%39.8) başarısızlık ve 142'sinde (%60.2) durdurma gözlenmiştir. Uygulamada yaş, sigara tüketimi (paket yıl), genişletilmiş rezeksiyon (extended resection), tümörün boyutu (mm), invazyon ve patolojik evre değışkenleri gruplandırılarak çözümlenmeye alınmıştır. Bu değışkenler ve değışkenlerin düzeyleri Çizelge 2'de verilmiştir.

Çizelge 2. Kullanılan değışkenler ve düzeyleri

| Değışken | Değışken Düzeyleri | Toplam Olay Sayısı (n) | % | Başarısız Olay Sayısı | Durdurulmuş Olay Sayısı |
|--------------------------|--------------------|------------------------|------|-----------------------|-------------------------|
| Yaş | 1. <=39 | 13 | 5.5 | 5 | 8 |
| | 2. 40-49 | 38 | 16.1 | 12 | 26 |
| | 3. 50-59 | 76 | 32.2 | 33 | 43 |
| | 4. 60-69 | 81 | 34.3 | 32 | 49 |
| | 5. >=70 | 28 | 11.9 | 12 | 16 |
| Sigara Tüketimi | 1. <=5 | 16 | 6.8 | 4 | 12 |
| | 2. 6-30 | 62 | 26.3 | 23 | 39 |
| | 3. 31-60 | 123 | 52.1 | 48 | 75 |
| | 4. >=61 | 35 | 14.8 | 19 | 16 |
| Genişletilmiş Rezeksiyon | 0. Yok | 190 | 80.5 | 71 | 119 |
| | 1. Var | 46 | 19.5 | 23 | 23 |
| Boyut | 1. <=30 | 73 | 30.9 | 25 | 48 |
| | 2. 31-40 | 46 | 19.5 | 12 | 34 |
| | 3. 41-50 | 41 | 17.4 | 18 | 23 |
| | 4. >=50 | 76 | 32.2 | 39 | 37 |
| İnvazyon | 0. Yok | 136 | 57.6 | 50 | 86 |
| | 1. Var | 100 | 42.4 | 44 | 56 |
| Patolojik Evre | 1. Evre I | 102 | 43.2 | 28 | 74 |
| | 2. Evre II | 61 | 25.8 | 24 | 37 |
| | 3. Evre III | 60 | 25.4 | 30 | 30 |
| | 4. Evre IV | 13 | 5.5 | 12 | 1 |

Ata ve Sözer (2007)'in çalışmasında, akciğ er kanseri verileri kullanılmış ve orantılı hazardlar varsayımının sağlanıp sağlanmadığı Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi kullanılarak incelenmiş ve sadece genişletilmiş rezeksiyon değışkeni için orantılı hazardlar varsayımının sağlanmadığı görülmüştür ($p= 0.0041$). Çalışmada, yarı parametrik yaşam modellerinden Cox regresyon modeli, tabakalandırılmış Cox regresyon modeli ve genişletilmiş Cox regresyon modeli ile çözümlenme yapılmış ve sonuçlar karşılaştırılmıştır. Elde edilen sonuçlara göre tabakalandırılmış Cox regresyon modelinin veri kümesi için en uygun model olduğu ve sigara tüketimi değışkeninin 4. düzeyinin, boyut değışkeninin 4. düzeyinin, patolojik evre değışkeninin 3. ve 4. düzeylerinin başarısızlığı etkileyen önemli faktörler olduğu görülmüştür.

Bu çalışmada ise akciğ er kanseri verileri ile orantısız hazardlar için yarı parametrik ve parametrik yaşam modellerinin kullanımı incelenmiştir. Bu nedenle aynı veri kümesi için parametrik yaşam modelleri kapsamında üstel, Weibull, log-lojistik, log-normal, Gompertz, Gamma regresyon modelleri ile çözümlenme yapılmıştır. Elde edilen sonuçlar, Ata ve Sözer (2007)'in çalışmasında bulunan sonuçlar ile birlikte değıerlendirilmiştir.

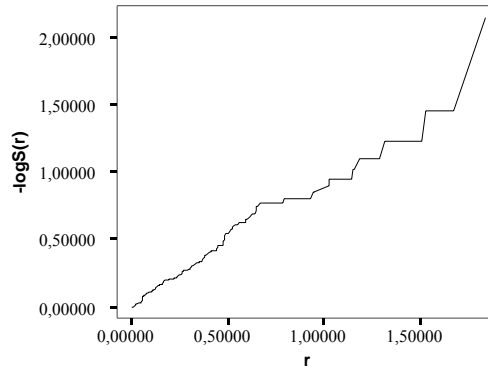
Modellerin anlamlığını test etmek için olabilirlik oranı (LR) test istatistiğı kullanılmış ve tüm modellerin istatistiksel olarak anlamlı olduğu ($p < 0.05$) görülmüştür. Model seçim kriteri olarak ise $-2\log L$ kullanılmıştır. Bu nedenle çalışmada incelenen modellere ait $-2\log L$ ve AIC değıerleri elde edilmiş ve Çizelge 3'de verilmiştir.

Çizelge 3. Yarı parametrik ve parametrik yaşam modelleri için $-2\log L$ ve AIC değerleri

| Kriter | Yarı parametrik yaşam modelleri | | | | | |
|----------------------------|---------------------------------|-----------------------|-------------------|------------|----------|---------|
| | Cox | Tabakalandırılmış Cox | Genişletilmiş Cox | | | |
| $-2\log L$ | 896.710 | 791.579 | 890.340 | | | |
| AIC | 926.710 | 819.579 | 922.340 | | | |
| Parametrik yaşam modelleri | | | | | | |
| Kriter | Üstel | Weibull | Log-lojistik | Log-normal | Gompertz | Gamma |
| $-2\log L$ | 446.280 | 441.778 | 437.586 | 432.593 | 446.214 | 433.979 |
| AIC | 496.250 | 495.778 | 491.586 | 487.980 | 482.214 | 490.593 |

Çizelge 3’de yer alan sonuçlar incelendiğinde $-2\log L$ ve AIC değerlerinin küçük olması modelin veri kümesine uygunluğunu gösterdiğinden, bu çalışmaya en uygun modelin parametrik regresyon modelleri olduğu sonucuna ulaşılmıştır.

Parametrik modellerden veri kümesi için en uygun olana karar vermek için Cox-Snell artıkları yöntemi kullanılmış ve veri kümesi için en uygun parametrik model log-lojistik yaşam modeli olarak bulunmuştur. Bu model için Cox-Snell artık grafiği Şekil 1’de verilmiştir.

**Şekil 1.** Log-lojistik yaşam modeli’nden elde edilen Cox-Snell artık çizimi

Çalışmadaki açıklayıcı değişkenlerle log-lojistik yaşam modeli kullanıldığında elde edilen sonuçlar Çizelge 4’de verilmiştir.

Çizelge 4. Log-lojistik Regresyon Çözümlemesinin Sonuçları

| Değişken ve düzeyleri | | β | S.H. | Alt sınır – Üst sınır | p değeri |
|--------------------------|--------------|---------|--------|-----------------------|----------|
| Kesişim terimi | | 5.3139 | 0.5773 | 4.1825 – 6.4454 | < 0.0001 |
| Yaş | 1. ≤ 39 | | | | |
| | 2. 40-49 | 0.1174 | 0.4483 | -0.7613 – 0.9962 | 0.7934 |
| | 3. 50-59 | -0.2057 | 0.4139 | -1.0168 – 0.6054 | 0.6191 |
| | 4. 60-69 | -0.0162 | 0.4198 | -0.8390 – 0.8066 | 0.9692 |
| | 5. ≥ 70 | -0.3570 | 0.4870 | -1.3115 – 0.5975 | 0.4636 |
| Sigara Tüketimi | 1. ≤ 5 | | | | |
| | 2. 6-30 | -0.5232 | 0.4564 | -1.4177 – 0.3714 | 0.2517 |
| | 3. 31-60 | -0.415 | 0.4330 | -1.2501 – 0.4471 | 0.3538 |
| | 4. ≥ 61 | -1.0662 | 0.4642 | -1.9760 – -0.1563 | 0.0216 |
| Genişletilmiş Rezeksiyon | 1. Yok | | | | |
| | 2. Var | -0.2812 | 0.2552 | -0.7815 – 0.2190 | 0.2705 |
| Boyut | 1. ≤ 30 | | | | |
| | 2. 31-40 | 0.3124 | 0.3041 | -0.2837 – 0.9084 | 0.3043 |
| | 3. 41-50 | -0.4470 | 0.3060 | -1.0469 – 0.1528 | 0.1441 |
| | 4. ≥ 50 | -0.7138 | 0.2627 | -1.2286 – 0.1989 | 0.0066 |

| | | | | | |
|--------------------|--|-------------------------------|----------------------------|---|------------------------------|
| İnvazyon | 1. Yok 2. Var | -0.1851 | 0.2549 | -0.6846 – 0.3144 | 0.4676 |
| Patolojik Evre | 1. Evre I 2. Evre II 3. Evre III 4. Evre IV | -0.1737 -0.6372 -1.5326 | 0.2893 0.2634 0.3570 | -0.7408 – 0.3933 -11534 – 0.1211 -2.2322 – 0.8330 | 0.5482 0.0155 < 0.0001 |
| Ölçek (σ) | | 0.6669 | 0.0586 | 0.5614 – 0.7922 | |

Yaşam çözümlenmesinde kullanılan regresyon modellerinde de ğ işken düzeylerinden biri referans kategorisi olarak alınmakta ve de ğ işken düzeylerinin yorumlanması buna göre yapılmaktadır. Başarısızlık süresini etkileyen faktörleri belirlemek için Çizelge 4 incelendiğ inde, sigara tüketimi de ğ işkeninin 4. düzeyinin, boyut de ğ işkeninin 4. düzeyinin ve patolojik evre de ğ işkeninin 3. ile 4. düzeylerinin ($p < 0.05$) %0,95 güven düzeyinde önemli oldu ğ u görülmüştür.

Sigara tüketimi 5'ten az olan hastaların ortalama yaşam süresi, sigara tüketimi 61'den fazla olan hastalara göre yaklaşık 2.95 kat, tümör boyutu 30 mm'den küçük olan hastaların ortalama yaşam süresi tümör boyutu 50 mm'den büyük olan hastalara göre yaklaşık 2 kat, , patolojik evresi I olan hastaların ortalama yaşam süresi patolojik evresi IV olan hastalara göre yaklaşık 1.8 kat ve patolojik evresi III olan hastalara göre yaklaşık 1.2 kat daha fazladır.

5. Sonuç

Bu çalışmada yarı parametrik ve parametrik yaşam modelleri incelenmiş ve akciğ er kanseri verileri kullanılarak bu modellerin orantısız hazardlar için kullanımı ele alınmıştır. Uygulamada Cox regresyon modeli, genişletilmiş Cox regresyon modeli, tabakalandırılmış Cox regresyon modeli ve parametrik regresyon modellerine ait sonuçlar elde edilmiştir. Elde edilen modeller arasında genel bir karşılaştırma yapmak için model seçimi kriteri olarak Akaike bilgi kriteri kullanılmış ve log-lojistik regresyon modelinin uygun oldu ğ u sonucuna ulaşılmıştır.

Akciğ er kanseri verisi için yaşam süresi parametrik dağılımlara uygunluk gösterdiğ inden, parametrik yaşam modelleri Cox regresyon modeli, tabakalandırılmış Cox regresyon modeli ve genişletilmiş Cox regresyon modeline göre daha uygun sonuçlar vermiştir.

Buna göre yaşam verilerinin analizinde yaşam sürelerinin dağılımları incelendikten sonra yarı-parametrik ya da parametrik yaşam modellerinin kullanılması gerektiğ ine karar verilmelidir. Yaşam süresi parametrik dağılımlara uygun dağılış gösterdiğ inde parametrik yaşam modellerinin daha uygun sonuçlar vereceğ i, ancak yaşam süresi parametrik dağılımlara uygun dağılış göstermediğ inde ise orantılı hazardlar varsayımı sağlanıyorsa Cox regresyon modeli, sağlanmıyorsa tabakalandırılmış Cox regresyon modeli ile genişletilmiş Cox regresyon modelinin kullanılmasının daha uygun sonuçlar vereceğ i yorumu yapılabilmektedir.

Kaynaklar

- [1] Arjas, E., (1988), A graphical method for assessing goodness of fit in Cox's proportional hazards model, *Journal of the American Statistical Association*, 83, 204-212.
- [2] Ata, N., Sözer, M.T., (2007), Cox Regression Models with Nonproportional Hazards Applied to Lung Cancer Survival Data, *Hacettepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, Matematik ve İstatistik Dergisi, B Serisi*, 36(2), 157-167.
- [3] Bender R., Augustin, T., Blettner, M., (2005), Generating survival times to simulate Cox proportional hazards models, *Statistics in Medicine*, vol.24, pp.1713-1723.
- [4] Cox, D.R., (1972), Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, 34, 187-220.
- [5] Cox, D.R., Snell, E.J., (1968), A General Definition of Residuals, *Journal of the Royal Statistical Society, Series B*, 30, 248-275.

- [6] Collett, D., (1994), *Modelling Survival Data in Medical Research*, Chapman&Hall, UK.
- [7] Efron, B., (1977), The Efficiency of Cox's Likelihood Function for Censored Data, *Journal of the American Statistical Association*, 72, 557-565.
- [8] Kalbfleisch, J.D. , Prentice,R.L., (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- [9] Klein, J.P., Moeschberger, M.L., (1997), *Survival Analysis Techniques for Censored and Truncated Data*, Springer, New York.
- [10] London, D., (1997), *Survival Models and Their Estimation*, Actex Publications, USA.
- [11] Lee, E.T., Wang, J.W., (2003) *Statistical Methods for Survival Data Analysis*, Wiley&Sons, New York.
- [12] Nardi, A., Schemper, M., (2003), Comparing Cox and Parametric Models in Clinical Studies, *Statistics in Medicine*, vol.22, pp.3597-3610.
- [13] Oakes, D., (1977), The Asymtotic Information In Censored Survival Data, *Biometrika*, 64, 441-448.
- [14] Schoenfeld, D., (1982), Partial residuals for the proportional hazards model, *Biometrika*, Vol.69, pp.551-55.
- [15] Therneau, T.M., Grambsch, P.M., (2000), *Modelling Survival Data: Extending the Cox Model*, Springer, New York.