

Makalenin Geliş Tarihi : 18.02.2006
Makalenin Kabul Tarihi : 21.12.2006

ORANTILI TEHLİKE VARSAYIMININ İNCELENMESİNDE KULLANILAN YÖNTEMLER VE BİR UYGULAMA

Nihal ATA¹, Durdu SERTKAYA KARASOY², M. Tekin SÖZER³

ÖZET : Yaşam çözümlemesinde orantılı tehlike modelinin kullanılabilmesi için orantılılık varsayımının sağlanması gerekmektedir. Bu çalışmada, yaşam çözümlemesinde orantılılık varsayımını incelemek için kullanılan yöntemler ele alınmıştır. Bu yöntemlerden, grafiksel yöntemler, zamana bağlı açıklayıcı değişkenlerin kullanıldığı yöntem ve yaşam süresinin rankı ile Schoenfeld artıkları arasındaki korelasyon testi yöntemi incelenmiştir. Grafiksel yöntemler görselliğe dayanmaktadır. Diğer iki yöntem ise teste dayalı sonuç vermektedir. Bu yöntemler akciğer kanserli hastalardan elde edilen verilere uygulanmış ve analize alınan değişkenlerden genişletilmiş rezeksiyon değişkeninin orantılılığı sağlamadığı görülmüştür.

ANAHTAR KELİMELE: Tehlike oranı, Orantısız tehlike, Schoenfeld artıkları, Yaşam eğrileri, Cox regresyon modeli.

THE METHODS USED FOR ASSESSMENT OF PROPORTIONAL HAZARD ASSUMPTION AND AN APPLICATION

ABSTRACT : In survival analysis, the assumption of proportionality has to be satisfied in order to use proportional hazard model. In this study, the methods used for checking the proportionality assumption in survival analysis are investigated. Graphical methods, the method of using time dependent covariates and correlation test between rank of survival time and Schoenfeld residuals are examined. Graphical methods are objective and rely on visuality. The other two methods are subjective and give the results supported by test statistics. We apply these methods to a data set gathered from a group of patients with lung cancer and it is found that extended resection variable does not satisfy the assumption proportionality.

KEYWORDS: Hazard ratio, nonproportional hazard, Schoenfeld residuals, survival curves, Cox regression model.

^{1,2,3} Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü 06800 Beytepe, ANKARA.

I. GİRİŞ

Yaşam çözümlemesinde yaygın olarak kullanılan yarı parametrik bir yöntem olan orantılı tehlike (hazard) modelinin (Cox regresyon modelinin) kullanılabilmesi için orantılı tehlike varsayımının incelenmesi gerekmektedir. Uygulamalarda bu varsayım incelenmeden varsayımın doğru olduğu kabul edilerek çözümler yapılmaktadır. Ancak tehlikelerin orantılı olmaması durumunda orantılı tehlike modelinin veri kümesine uygulanması uygun olmamaktadır.

Orantılı tehlike varsayımının incelenmesinde en çok kullanılan yöntemler, log(-log) yaşam eğrileri [1], gözlenen ve beklenen yaşam eğrileri [2], Arjas grafikleri [3], modele zamana bağlı değişkenlerin eklenmesi [1], Schoenfeld artıkları ile yaşam süresinin rankı arasındaki korelasyon testi [4-5] biçiminde sıralanabilmektedir.

Orantılı tehlike varsayımının sağlanıp sağlanmadığının ortaya çıkarılması için yukarıda verilen yöntemlerin dışında başka yöntemler de vardır. Ancak, bu yöntemler çok sık kullanılmamaktadır.

II. ORANTILI TEHLİKE MODELİ

Yaşam çözümlemesi, pozitif tanımlı rassal değişkenlerinin çözümlenmesi için kullanılan istatistiksel yöntemler bütünü olarak tanımlanmaktadır. Rassal değişkeninin değeri, bir makina parçasının başarısızlık zamanı, biyolojik bir birimin (hasta, hayvan, hücre) ölüm zamanı olabilmektedir. İyi tanımlanmış herhangi bir olayın gerçekleşme ya da gözlenme süresinin çözümlemesi, yaşam çözümlemesi teknikleri ile yapılabilir. Söz konusu olayın gerçekleşmesi başarısızlık olarak tanımlanmaktadır [6].

Yaşam çözümlemesinde, klasik istatistiksel yöntemlerin kullanılmamasının nedenlerinden biri durdurma (censoring), diğeri ise zamana bağlı açıklayıcı değişkenlerdir (time-dependent covariates) [7].

Yaşam süresine ilişkin etkenlerin tehlike fonksiyonu üzerindeki etkilerin çarpımsal olduğu modeller, yaşam süresi verilerinin çözümlemesinde önemli bir rol oynamaktadır. İlk kez 1972’de Cox tarafından ele alınan orantılı tehlike modeli de bu modellerden biridir.

T , bir birimin yaşam süresini temsil eden sürekli rassal değişkeni ve $\mathbf{x}$, bu bireyle ilgili bilinen açıklayıcı değişkenler vektörü olmak üzere orantılı tehlike varsayımı altında $\mathbf{x}$ verildiğinde T ’nin tehlike fonksiyonu,

$$h(t, \mathbf{x}) = h_0(t)g(\mathbf{x}, \boldsymbol{\beta})$$

biçimindedir. $g(\mathbf{x}, \boldsymbol{\beta})$ değişik biçimlerde ifade edilebilmektedir. Cox’un (1972) incelediği model,

$$h(t, \mathbf{x}) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}) \quad (1)$$

biçimindedir. Burada, $\boldsymbol{\beta}$ regresyon katsayıları vektörü ve $h_0(t)$ ise açıklayıcı değişkene sahip olmayan ($\mathbf{x}=0$ olan) bir birimin temel tehlike fonksiyonu olarak tanımlanmaktadır [8]. Cox’un (1972) önerdiği yaklaşım, $h_0(t)$ için özel bir biçimin varsayılmadığı, dağılımdan bağımsız bir yaklaşımdır. Diğer bir ifadeyle, bu yöntem, ilgili yaşam süresi dağılımına bir başka deyişle temel tehlike fonksiyonu $h_0(t)$ ’ye dayanmaktadır [9]. Eşitlik (1)’de verilen model Cox regresyon modeli olarak da adlandırılmaktadır. Model, orantılı tehlike varsayımına dayanmasına rağmen, yaşam süreleri için olasılık dağılımının belirli bir biçimi yoktur. Bu yüzden, orantılı tehlike modeli yarı parametrik bir model olarak ele alınmaktadır [10]. Temel tehlike fonksiyonunun gerçek biçimini içeren hiçbir varsayım yoktur. Aynı şekilde Cox regresyon modelinde $\boldsymbol{\beta}$ katsayıları, hiçbir varsayım yapılmadan tahmin edilmektedir [7].

III. ORANTILI TEHLİKE VARSAYIMININ İNCELENMESİ

Cox regresyon modelinin temel varsayımı orantılılığın sağlanmasıdır. Orantılı tehlike varsayımı, tehlike oranının zamana karşı sabit olması ya da bir bireyin tehlikesinin diğer bireyin tehlikesine orantılı olması anlamına gelmektedir [11].

$\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_p^*)$ ve $\mathbf{x} = (x_1, x_2, \dots, x_p)$ iki bireye ait açıklayıcı değişkenler vektörü olmak üzere tehlike oranı,

$$HO = \frac{\hat{h}(t, \mathbf{x}^*)}{\hat{h}(t, \mathbf{x})} = \frac{\exp\left[\sum_{j=1}^p \hat{\beta}_j x_j^*\right] \hat{h}_0(t)}{\exp\left[\sum_{j=1}^p \hat{\beta}_j x_j\right] \hat{h}_0(t)} = \exp\left[\sum_{j=1}^p \hat{\beta}_j (x_j^* - x_j)\right] \quad (2)$$

biçimindedir [12]. Eşitlik (2)'de görüldüğü gibi tehlike oranı t 'yi içermemektedir. Diğer bir ifadeyle, model uydurulduğunda $\mathbf{x}^*$ ve $\mathbf{x}$ için değerler belirlendiğinde, tehlike oranı tahmini için üstel ifadenin değeri sabittir, zamana bağlı değildir. Bu sabit $\hat{\theta}$ ile gösterilirse, tehlike oranı

$$\hat{\theta} = \frac{\hat{h}(t, \mathbf{x}^*)}{\hat{h}(t, \mathbf{x})}$$

biçiminde yazılabilir. Burada $\hat{\theta}$, orantılılık sabiti (proportionality constant) olarak adlandırılır ve zamandan bağımsızdır [7].

Orantılı tehlikenin anlamını ortaya koyabilmek için orantılı tehlike varsayımının sağlanmadığı bir durum örnek olarak verilebilir:

Tedavi görmekte olan kanser hastalarının iki gruba ayrıldığı düşünölsün. Gruplardan biri cerrahi tedavi gördükten sonra kemoterapi alan hastalardan, diğeri ise cerrahi tedavi görmeden kemoterapi alan hastalardan oluşsun.

İlgilenilen tek değişken

$$C = \begin{cases} 1 & \text{cerrahi müdahale görüyor ise} \\ 0 & \text{cerrahi müdahale görmüyor ise} \end{cases}$$

biçiminde tanımlanan cerrahi müdahale değişkeni olsun. Bu durumda verinin incelenmesi için Cox regresyon modeli tek bir değişken içerecektir. Model,

$$h(t, x) = h_0(t) \exp(\beta C)$$

biçimindedir. C değişkenini içeren Cox regresyon modelinin bu durum için uygun olup olmadığının incelenmesi gerekmektedir. Hasta bir dizi cerrahi müdahale geçirirken, cerrahi müdahaleden doğan komplikasyonlar için genellikle yüksek risk vardır. İyileşme sürecinde hasta bu kritik dönemi geçirdikten sonra cerrahi müdahalenin faydasını görebilir. Bu durumda tehlike fonksiyonları çakişacaktır. t^* günden önce cerrahi müdahale görmeyen hasta grubunun tehlikesi, cerrahi müdahale gören hasta grubuna göre daha düşüktür. Ancak t^* günden sonra tam tersi bir durum ortaya çıkmaktadır. Bu durum,

$$t < t^* \text{ iken tehlike oranı } \frac{\hat{h}(t < t^*, C = 0)}{\hat{h}(t < t^*, C = 1)} < 1$$

$$t > t^* \text{ iken tehlike oranı } \frac{\hat{h}(t > t^*, C = 0)}{\hat{h}(t > t^*, C = 1)} > 1$$

biçiminde gösterilebilir. Herbir grup için tehlike fonksiyonları doğru tahmin edilmiş ise tehlike oranları zamana karşı sabit değildir [2].

Orantılı tehlike modelinde, orantılı tehlike varsayımının incelenmesi için kullanılan birçok yöntem vardır. Bu yöntemlerden bazıları, grafiksel yöntemler, zamana bağlı açıklayıcı değişkenlerin kullanılması yöntemi, Schoenfeld artıkları ile yaşam süresinin rankının korelasyon testi yöntemi olmaktadır. Sözü edilen yöntemler aşağıda verilmiştir.

III.1 Grafiksel Yöntemler

Az sayıda düzeyleri olan sabit zamanlı değişkenler için yaşam eğrilerine bakılarak orantılı tehlike varsayımının basit grafiksel testi yapılabilir. Orantılı tehlike varsayımını incelemek için kullanılan grafiksel yöntemlerden bazıları aşağıda verilmiştir:

- ◆ Log(-log) yaşam eğrilerinin çizimi,
- ◆ Gözlenen ve beklenen yaşam eğrilerinin çizimi
- ◆ Arjas grafikleri [13].

III.1.1 Log(-log) Yaşam Eğrileri Yöntemi

Log(-log) yaşam eğrileri yöntemi en yaygın kullanılan grafiksel yöntemdir. İncelenen değişkenlerin farklı kategorileri üzerinden $-\ln(-\ln)$ yaşam eğrilerinin tahmininin karşılaştırılmasını içermektedir. Elde edilen paralel eğriler orantılı tehlike varsayımının sağlandığını göstermektedir [1-11].

Yaşam eğrisi tahmini $[0,1]$ aralığında değer alırken, $-\ln(-\ln \hat{S})$ eğrisi $(-\infty, \infty)$ aralığında değer almaktadır.

Log(-log) yaşam eğrisi, yaşam eğrisi tahminine uygulanan basit bir dönüşümdür. Yaşam olasılığı tahmininin iki kez doğal logaritması alınarak elde edilmektedir.

Yaşam fonksiyonu $[0,1]$ aralığında değerler aldığından $\ln S(t, x)$ ve $\ln S_0(t)$ negatif olmaktadır. Negatif bir sayının logaritması alınamadığından, $\ln S(t, x)$ 'in ikinci kez logaritmasını alabilmek için bu fonksiyonun $-\ln S(t, x)$ olarak kullanılması gerekmektedir. İkinci logaritma pozitif ya da negatif olabilir. Bu nedenle ikinci kez negatifinin alınmasına gerek yoktur. Tutarlılık için, $\ln(-\ln)$ ifadesini elde etmek için ikinci logaritmanın önüne eksi işareti konulmaktadır.

Açıklanan bu grafiksel yöntemin bazı zayıf yönleri vardır. Paralelliğin nasıl belirleneceğine karar vermek öznel olabilir. Log(-log) yaşam eğrilerinin paralel olmadığına dair güçlü bir kanıt yoksa, orantılı tehlike varsayımının sağlandığı varsayılmaktadır. Sürekli değişkenlerin nasıl sınıflandırılacağıının belirlenmesi önemlidir. Farklı sınıflandırmalar farklı grafiksel çizimleri ortaya koymaktadır. Sürekli değişkenler sınıflandırılırken, sınıf sayısının az alınması önerilmektedir. Sınıfların seçimi mümkün olduğu kadar anlamlı olmalı ve uygun sayı dengesinin sağlanması gerekmektedir. Log(-log) yaşam eğrilerinin kullanılması ile ilgili bir sorun da birçok değişken için orantılı tehlike varsayımının eş zamanlı olarak nasıl değerlendirileceğidir. Eş zamanlı karşılaştırmalar için, bütün değişkenler ayrı ayrı sınıflandırılmalı, sınıfların farklı birleşimleri oluşturulmalı ve daha sonra aynı grafik üzerinde bütün log(-log) yaşam eğrileri karşılaştırılmalıdır. Herbir birleştirilmiş sınıf için yeterli sayı olsa bile, paralel olmama durumuna neden olan değişkenlere karar vermek zordur [2].

III.1.2 Gözlenen ve Beklenen Yaşam Eğrileri Yöntemi

Orantılı tehlike varsayımını değerlendirmek için gözlenen yaşam olasılıklarına karşı beklenen yaşam olasılıklarının çizimlerinin kullanımı uyum iyiliği testi yaklaşımının grafiksel karşılığıdır. Uyum iyiliği testi de aynı gözlenen ve beklenen yaşam olasılığı tahminlerini kullanarak uygulanmaktadır.

Log(-log) yaşam eğrisi yaklaşımında olduğu gibi, gözlenen ile beklenen yaşam eğrilerine dayanan grafiksel yöntem aşağıda verilen yollardan biri ya da her ikisi de kullanılarak uygulanabilir:

1. Her zaman noktasında tüm değişkenler için orantılı tehlike varsayımının değerlendirilmesi,
2. Diğer değişkenler için düzeltme yapıldıktan sonra orantılı tehlike varsayımının değerlendirilmesi.

Birinci yol, gözlenen çizimleri elde etmek için Kaplan-Meier (KM) eğrilerinin kullanılmasını içermektedir. İkinci yol ise gözlenen çizimleri oluşturmak için

tabakalandırılmış Cox regresyon modelinin kullanılmasını içermektedir. Birinci yol kullanılırken, açıklayıcı değişkenlerin sınıfları kullanılarak veri tabakalandırılır ve daha sonra her sınıf için KM eğrileri ayrı ayrı elde edilir [2].

Beklenen yaşam eğrilerini elde etmek için, değerlendirilen açıklayıcı değişkeni içeren Cox regresyon modeli uydurulur. Yaşam eğrisinin tahmini, açıklayıcı değişkenin her bir sınıfı için değerleri formülde yerine yerleştirilerek elde edilir.

Açıklayıcı değişkenin her bir sınıfı için gözlenen ve beklenen yaşam eğrileri birbirine benzerse, orantılı tehlike varsayımının sağlandığına karar verilmektedir. Bir ya da daha çok sınıf için bu eğriler biraz farklılık gösterirse, orantılı tehlike varsayımının bozulduğuna karar verilir.

Gözlenen ve beklenen yaşam eğrileri yönteminin zayıf bir yönü, incelenen sınıf için gözlenen ve beklenen yaşam eğrilerini karşılaştırarak, eğrilerin yakınlığının ne kadar olduğuna karar verilmesidir. Gözlenen ve beklenen yaşam eğrileri gerçekten farklı bulunursa, orantılı tehlike varsayımının sağlanmadığına karar verilir.

Orantılı tehlike varsayımını incelemek için sürekli değişkenlere ait gözlenen ve beklenen yaşam eğrileri kullanılırken, gözlenen yaşam eğrileri kategorik değişkenlerde olduğu gibi elde edilir. Sürekli değişken sınıflara ayrılarak tabakalar oluşturulur ve daha sonra her bir sınıf için KM eğrileri elde edilir [2].

III.1.3 Arjas Grafikleri

Orantılı tehlike modelinde uyum iyiliğinin test edilmesi için kullanılan grafiksel bir yöntemdir. Alternatif modellerin tahminine gereksinim duymadığından ve sadece parametre tahmininde kullanılan kısmi olabilirlik ifadesindeki niceliklere benzer nicelikleri içerdiğinden kullanılması kolay bir yöntemdir [3]. Grafiksel yöntemler içinde en uygun sonucu veren yöntemdir [13].

Grafiklerin değerlendirilmesi öznelidir. Gözlenen ve beklenen (modelden tahmin edilen) başarısızlık sıklıkları arasında doğrudan bir karşılaştırma yapmaktadır.

Doğru belirlenmiş bir modelde bu sıklıkların arasında yaklaşık bir dengenin olması beklenmektedir.

Arjas grafikleri yönteminde bireyler tabakalara ayrılır ve herbir tabaka için bir grafik çizilir. Cox regresyon modeli veriye uygunsa, grafiğin eğiminin bire yakın ve yaklaşık olarak doğrusal olması beklenmektedir. Model yanlış belirlendiği takdirde grafik üzerinde gözle görülebilir farklı şekiller gözlemlenmektedir [3].

III.2 Zamana Bağlı Açıklayıcı Değişkenlerin Kullanılması Yöntemi

Zamandan bağımsız değişkenin orantılı tehlike varsayımını sağlayıp sağlamadığını incelemek için zamana bağlı değişkenler kullanılır. Bu durumda model, zamandan bağımsız değişkenleri ve zamanın bazı fonksiyonlarını kapsayan çarpım terimlerini içeren Cox regresyon modeli olmaktadır. Genişletilmiş Cox regresyon modeli,

$$h(t, x) = h_0(t) \exp[\beta x + \delta(x \times g(t))]$$

biçiminde tanımlanır. $g(t)$ fonksiyonu için farklı seçenekler vardır. Bunlar;

- $g(t) = t$,
- ♦ $g(t) = \log t$,
- ♦ $g(t) = \begin{cases} 1 & t \geq t_0 \\ 0 & t < t_0 \end{cases}$ (Adım (Heaviside, step) fonksiyon)

fonksiyonları olabilir [2].

Genişletilmiş Cox regresyon modeli kullanılarak çarpım terimlerinin önemlilik testi yapıp orantılı tehlike varsayımı değerlendirilir. Yokluk hipotezi, $H_0 : \delta = 0$ 'dır. Yokluk hipotezi doğru ise, model tek bir değişken içeren orantılı tehlike modeline indirgenir. Wald istatistiği ya da olabilirlik oranı istatistiği kullanılarak test uygulanabilir. Her iki durumda da test istatistiği yokluk hipotezi altında 1 serbestlik dereceli ki-kare dağılımına sahiptir.

Eşzamanlı açıklayıcı değişkenlerin birçoğunda orantılı tehlike varsayımını incelemek için aşağıda verilen genişletilmiş Cox regresyon modeli kullanılabilir:

$$h(t, x) = h_0(t) \exp \left[\sum_{j=1}^p \beta_j x_j + \delta_j (x_j \times g_j(t)) \right] \quad (3)$$

Verilen bu model ana etki terimlerini ve bu terimlerin zamanın bazı fonksiyonları ile çarpımını içermektedir. Farklı açıklayıcı değişkenler, zamanın farklı fonksiyonlarına gereksinim duymaktadır. Burada, $g_j(t)$, j. açıklayıcı değişken için zamanın fonksiyonunu tanımlamak için kullanılmaktadır.

Eşitlik (3)'de verilen model ile orantılı tehlike varsayımını eşzamanlı olarak test etmek için,

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_p = 0$$

yokluk hipotezi incelenir. Bu ise, p serbestlik dereceli olabilirlik oranı ki-kare istatistiğini gerektirmektedir. Olabilirlik oranı (likelihood ratio, LR) test istatistiği,

$$LR = -2 \ln \hat{L}_{\text{Cox Regresyon Modeli}} - (-2 \ln \hat{L}_{\text{Genişletilmiş Cox Regresyon Modeli}}) \sim \chi_p^2$$

biçimindedir. Model, yokluk hipotezi altında Cox regresyon modeline indirgenmektedir.

Orantılı tehlike varsayımını incelemek için genişletilmiş Cox regresyon modelinin kullanılmasının sakıncası, modelde zamana bağlı çarpım terimleri için $g_j(t)$ fonksiyonunun seçimidir. Bu seçim açık değildir, farklı seçenekler orantılı tehlike varsayımı ile ilgili farklı sonuçlar ortaya çıkarabilmektedir [2].

III.3 Schoenfeld Artıkları ile Yaşam Sürelerinin Rankının Korelasyon Testi Yöntemi

Cox regresyon modelinde kullanılan Cox-Snell, değiştirilmiş Cox-Snell, Martingale ve Sapma artıklarının iki dezavantajı vardır:

- Ağırlıklı olarak gözlenen yaşam süresine bağlı olmaları,
- Birikimli tehlike fonksiyonunun tahminine ihtiyaç duymalarıdır.

Bu dezavantajlar, Schoenfeld (1982) tarafından önerilen artıklar ile ortadan kaldırılmaktadır. Bu artıklar skor artık olarak da adlandırılmaktadır. Tanımlanan diğer artık türlerinden farklı olan bir diğer özelliği de herbir birey için artığın tek bir değeri yoktur. Cox regresyon modelinde kullanılan herbir birey, her açıklayıcı değişkeni için bir artık değerine sahiptir [7].

p tane açıklayıcı değişken olduğunu ve n tane bağımsız gözlem için zamanın, açıklayıcı değişkenin ve durdurma gösterge değişkeninin (t_i, x_i, δ_i) ile gösterildiği varsayalım. $i = 1, 2, \dots, n$ olmak üzere durdurulmamış gözlemler için $\delta_i = 1$ ve durdurulmuş gözlemler için $\delta_i = 0$ 'dır. Schoenfeld'in (1982), orantılı tehlike modeli ile kullanılması için önerdiği artıklar, log kısmi olabilirliğin türevine bireysel katkıya dayanmaktadır. j . açıklayıcı değişken için türev,

$$\frac{\partial L_p(\beta)}{\partial \beta_j} = \sum_{i=1}^n \delta_i \left[x_{ji} - \frac{\sum_{\ell \in R(t_i)} x_{j\ell} \exp(\beta' x_{\ell})}{\sum_{\ell \in R(t_i)} \exp(\beta' x_{\ell})} \right] = \sum_{i=1}^n \delta_i \{x_{ji} - a_{ji}\} \quad (4)$$

biçimindedir. Burada,

$$a_{ji} = \frac{\sum_{\ell \in R(t_i)} x_{j\ell} \exp(\beta' x_{\ell})}{\sum_{\ell \in R(t_i)} \exp(\beta' x_{\ell})} \quad (5)$$

olarak tanımlanmaktadır. Eşitlik (4)'de x_{ji} çalışmadaki i. birey için j. ($j=1, 2, \dots, p$) açıklayıcı değişkenin değeri ve $R(t_i)$ ise t_i zamanında riskte olan tüm bireylerin kümesidir.

j. açıklayıcı değişken için i. bireyin Schoenfeld artığının tahmin edicisi, Eşitlik (4)'den

$$\hat{r}_{Sji} = \delta_i (x_{ji} - \hat{a}_{ji})$$

biçiminde elde edilir. Burada $\hat{a}_{ji}$ Eşitlik (5)'de verildiği gibidir [4].

Schoenfeld artıklarının sıfırdan farklı değerleri sadece durdurulmamış gözlemler için açığa çıkmaktadır [14].

Orantılı tehlike varsayımını incelemek için Schoenfeld artıklarına dayanan bir test geliştirilmiştir [5]. Belirli bir değişken için Schoenfeld artıkları ile bireylerin yaşam sürelerinin rankı (yaşam süreleri) arasındaki korelasyon kullanılarak orantılı tehlike varsayımı incelenmiştir. Bu teste göre, orantılı tehlike varsayımının sağlanması için korelasyonun sıfıra yakın olması beklenmektedir. Yokluk hipotezi “orantılı tehlike varsayımı sağlanmaktadır” biçimindedir. Yokluk hipotezi hiçbir zaman tam anlamıyla kanıtlanamamaktadır. Ancak yokluk hipotezini reddedecek yeterli kanıtın olmadığı söylenebilir [15].

Bu test istatistiği, orantılı tehlike varsayımının incelenmesi için kullanılan grafiksel yöntemlere göre daha nesnel bir kriter sağlamaktadır. Grafiksel yöntemler ise daha öznelidir. İlgilenilen olayla erken bir zamanda karşılaşan bireyler için artıklar pozitif olma eğilimi gösterirse ve olayla geç bir zamanda karşılaşan bireyler için artıklar negatif olma eğilimi gösterirse, tehlike oranı zaman eksenini boyunca sabit olmaz, bu durumda orantılı tehlike varsayımı sağlanmaz [15].

IV. UYGULAMA

Çalışmada, 236 akciğer kanseri hastasına ait veriler kullanılmıştır. Hastalar, ameliyat olduktan sonra hastalıklarının ilk nüks etmesine kadar geçen süre boyunca izlenmiştir. Uygulama aşamasında, orantılı tehlike varsayımının sağlanıp sağlanmadığı Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi, Arjas grafikleri, $\log(-\log)$ yaşam eğrileri, beklenen ve gözlenen yaşam eğrileri ve zamana bağlı değişkenlerin modele eklenmesi yöntemleri kullanılarak incelenmiştir.

Çalışmada, hastaların ameliyat olduğu tarihten hastalığın ilk nüksetmesine kadar geçen süre (ay olarak) yaşam süresi olarak alınmıştır. Hastalığın nüksetmesi başarısızlık olarak ifade edilmiştir. Hastalığı nüksetmeyen hastalar durdurulmuş olarak tanımlanmıştır. Hastaların izlenme süresi sona erdiğinde 236 hastadan 94'ünde (%39.8) başarısızlık ve 142'sinde (%60.2) durdurma gözlenmiştir. Uygulamada yaş (YŞ), sigara tüketimi (ST), genişletilmiş rezeksiyon (extended resection, ER), tümörün boyutu (BY), invazyon (İV) ve patolojik evre (PE) değişkenleri gruplandırılarak çözümlemeye katılmıştır. Bu değişkenler ve değişkenlerin düzeyleri Çizelge 1'de verilmiştir.

Çizelge 1. Kullanılan değişkenler ve düzeyleri.

Değişken	Değişken 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

IV.I Orantılı Tehlike Varsayımının İncelenmesi

Schoenfeld Artıkları İle Yaşam Süresinin Rankı Arasındaki Korelasyon Testi

Belirli bir değişken için Schoenfeld artıkları ile bireylerin yaşam sürelerinin rankı arasındaki korelasyon kullanılarak orantılı tehlike varsayımı incelenmektedir. Uygulamada Schoenfeld artıkları SAS'ın PROC PHREG ve yaşam sürelerinin rankı ise PROC RANKED modülü kullanılarak elde edilmiştir. Bu korelasyonu test etmek için p değeri ise PROC CORR modülü kullanılarak elde edilmiştir. Bu test istatistiği için yokluk hipotezi “orantılı tehlike varsayımı sağlanmaktadır” biçiminde kurulmuştur.

Çizelge 2'ye bakıldığında ER değişkenine ait p değeri 0.0041 olarak bulunmuştur ve sadece bu değişken için %95 güven düzeyinde orantılı tehlike varsayımının sağlanmadığı sonucuna ulaşılmıştır. Diğer tüm değişkenler için p

değeri 0.05 değerinden büyüktür ve bu test istatistiğine göre orantılı tehlike varsayımı %95 güven düzeyinde sağlanmaktadır.

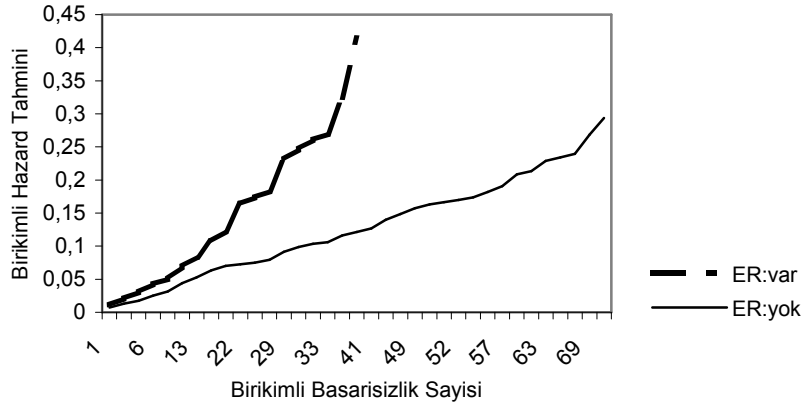
Çizelge 2. Shoenfeld artıkları ile yaşam süresinin rankı arasındaki korelasyon çözümlemesi sonuçları.

	Pearson Korelasyon Katsayısı					
	p değeri					
	YŞ(2)	YŞ(3)	YŞ(4)	YŞ(5)	ST(2)	ST(3)
Yaşam süresinin rankı	0.2084	0.3626	0.8916	0.6435	0.4692	0.8000
	ST(4)	BY(2)	BY(3)	BY(4)	PE(2)	PE(3)
Yaşam süresinin rankı	0.4830	0.2134	0.9015	0.4098	0.0635	0.6910
	PE(4)	İV	ER			
Yaşam süresinin rankı	0.3271	0.4977	0.0041			

Grafiksel Yöntemler

Arjas Grafikleri

Orantılı tehlike modeli veriye uygunsa grafiğin bire yakın eğime sahip ve yaklaşık olarak doğrusal olması beklenmektedir. Korelasyon testi kullanılarak veri kümesi için orantısız tehlikenin varlığına neden olan değişken genişletilmiş rezeksiyon olduğundan bu değişkene ait Arjas grafiği Şekil 1’de verilmiştir.

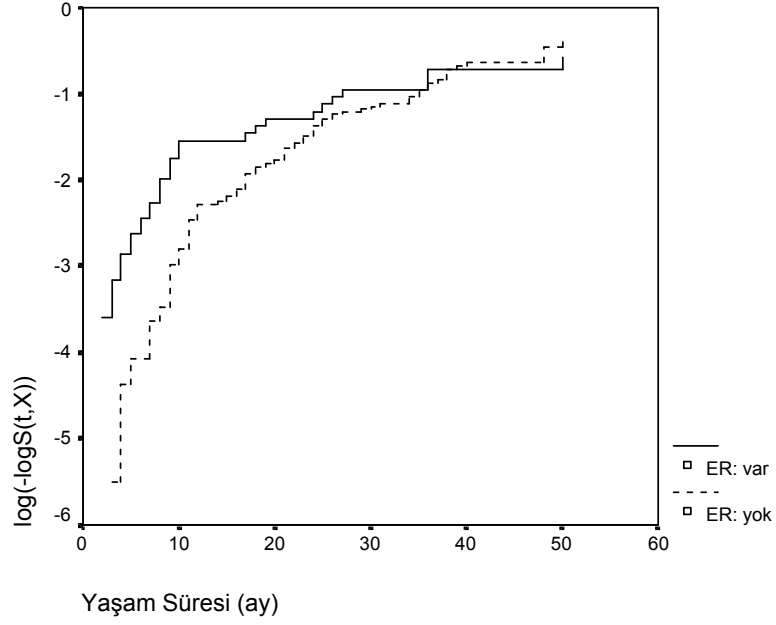


Şekil 1. Genişletilmiş rezeksiyon değişkeni için Arjas grafiği.

Şekil 1 incelendiğinde ER değişkeninin iki düzeyi için de grafiğin eğiminin birden farklı olduğu görülmektedir. Buna göre bu değişken için orantılı tehlike varsayımının sağlanmadığı sonucuna ulaşılmıştır.

Log(-Log) Yaşam Eğrileri

Log(-log) yaşam eğrilerinin paralelliği orantılı tehlike varsayımının incelenmesi için grafiksel bir yaklaşım sağlamıştır. Veri kümesi için orantılı tehlike modeli uygunsa log(-log) eğrilerinin paralel olması beklenmektedir. Orantılı tehlike varsayımını sağlamadığı düşünülen ER değişkeni için log(-log) yaşam eğrisi Şekil 2’de verilmiştir:

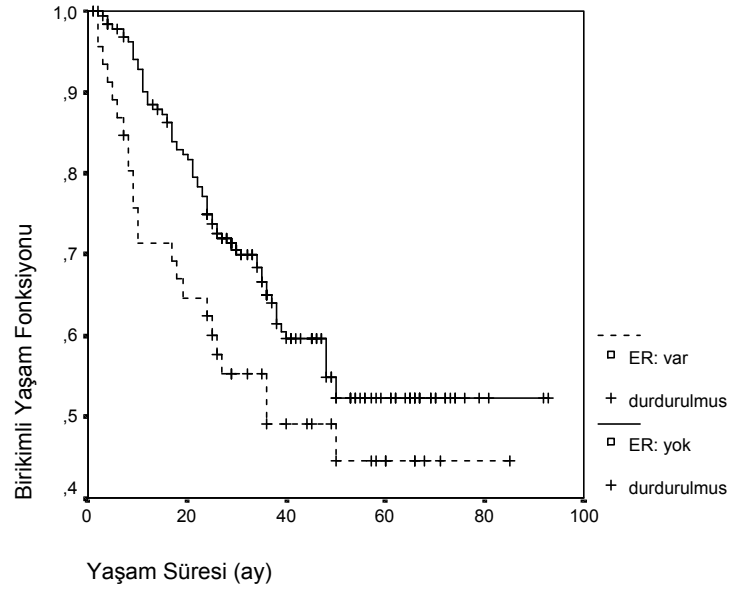


Şekil 2. Genişletilmiş rezeksiyon değişkeni için log(-log) yaşam eğrileri.

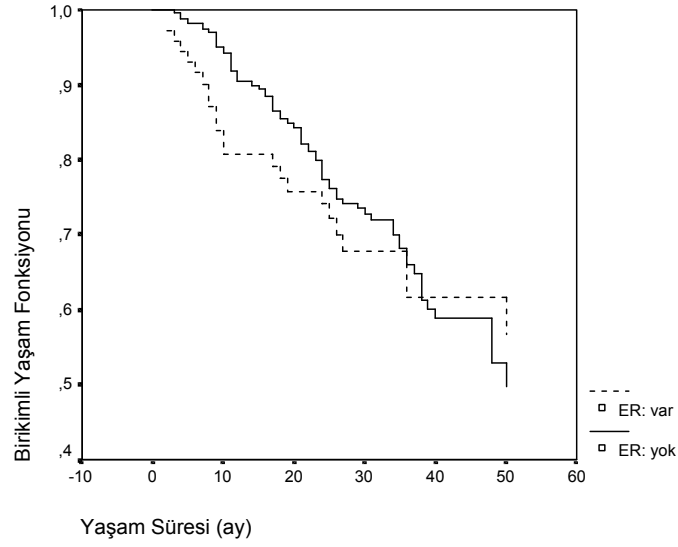
Şekil 2 incelendiğinde ER değişkeni için log(-log) yaşam eğrisinin paralel olmadığı ve kesiştiği görülmektedir. Bu durumda bu değişken için orantılı tehlike varsayımının sağlanmadığı söylenebilmektedir.

Gözlenen ve Beklenen Yaşam Eğrileri

ER değişkeni için gözlenen yaşam eğrileri Kaplan-Meier eğrilerinden elde edilen çizimlerdir. Beklenen yaşam eğrileri ise Cox regresyon modeli kullanılarak elde edilmiştir. Orantısızlığa neden olan ER değişkeni için gözlenen ve beklenen yaşam eğrileri Şekil 3 ve Şekil 4’de verilmiştir. Gözlenen ve beklenen yaşam eğrileri birbirine yakın olmadığından ER değişkeni için orantılı tehlike varsayımının sağlanmadığı sonucuna varılmaktadır.



Şekil 3. Genişletilmiş rezeksiyon değişkeni için beklenen yaşam eğrisi.



Şekil 4. Genişletilmiş rezeksiyon değişkeni için gözlenen yaşam eğrisi.

Zamana Bağlı Değişkenler Kullanılarak Orantılı Tehlike Varsayımının İncelenmesi

(a) Zamana bağlı fonksiyon olarak en çok tercih edilen $g(t)=\log t$ fonksiyonu kullanılarak orantılı tehlike varsayımı test edilebilir. Orantılı tehlike varsayımını sağlamadığı düşünülen ER değişkeninin zamandan bağımsız olup olmadığını anlamak için Cox regresyon çözümlemesi kullanılmıştır. Bunun için öncelikle sadece ER değişkenini içeren Cox regresyon çözümlemesi yapılmış ve sonuçlar Çizelge 3’de verilmiştir.

Çizelge 3. ER değişkeni için Cox regresyon çözümlemesinin sonuçları.

Değişken	β	Std. Hata	p-değeri	$\exp(\beta)$	Güven Aralığı (%95)
ER	0,41885	0,24004	0,0810	1,5202	0.9497 - 2.4335

Daha sonra ise ER değişkenini ve zamana bağlı bir fonksiyon olarak tanımlanan ERxlogt değişkenini içeren Cox regresyon çözümlemesi yapılmış ve sonuçlar Çizelge 4’de verilmiştir.

Çizelge 4. ER ve ERxlogt değişkeni için Cox regresyon çözümlemesinin sonuçları.

Değişken	β	Std. Hata	p-değeri	$\exp(\beta)$	Güven Aralığı (%95)
ER	2.85612	0.89030	0.0013	17.394	3.038 - 99.91
ERxlogt	-0.90607	0.32846	0.0058	0.404	0.212 - 0.769

Çizelge 4’deki sonuçlardan elde edilen Cox regresyon modeli

$$h(t, X) = h_0(t) \exp[\beta ER + \delta(ER \times \log t)]$$

$$= h_0(t) \exp[2.85612(ER) - 0.90607(ER \times \log t)]$$

biçiminde yazılmaktadır. Test istatistiği kullanılarak orantılı tehlike varsayımı incelenmek istendiğinde yokluk hipotezi $H_0 : \delta = 0$ biçimindedir. Test istatistiği, Cox regresyon modeli ile genişletilmiş Cox regresyon modeli arasındaki log-olabilirlik oranı istatistiğinin farkına dayanmaktadır. Test

istatistiği yokluk hipotezi altında 1 serbestlik dereceli ki-kare dağılımına sahiptir. Yokluk hipotezi kabul edilemez ise ER değişkeninin orantılı tehlike varsayımını sağlamadığı düşünülmektedir. Buna göre test istatistiği,

$$\begin{aligned} LR &= -2 \ln \hat{L}_{\text{Cox Regresyon Modeli}} - (-2 \ln \hat{L}_{\text{Genisletilmiş Cox Regresyon Modeli}}) \\ &= 944,560 - 936.209 = 8.351 \end{aligned}$$

biçiminde elde edilmiştir. Olabilirlik oranı test istatistiği 8.351 olarak bulunmuştur ve $\chi^2_1 = 3,86$ olduğundan yokluk hipotezi %95 güven düzeyinde kabul edilemez. Bu durumda ER değişkeni için orantılı tehlike varsayımının sağlanmadığı sonucuna ulaşılmıştır.

(b) Modelde tüm değişkenler varken ER değişkeni için orantılı tehlike varsayımı incelenebilir. Zamana bağlı fonksiyon olarak en çok tercih edilen $g(t)=\log t$ fonksiyonu kullanılarak orantılı tehlike varsayımı test edilebilir. Öncelikle tüm değişkenleri içeren Cox regresyon çözümlemesi sonuçlarının elde edilmesi gerekmektedir. Bu sonuçlar Çizelge 5’de verilmiştir. Çalışmada, tüm değişkenleri ve zamana bağlı bir fonksiyon olarak tanımlanan ERxlogt değişkenini içeren Cox regresyon çözümlemesinin sonuçları ise Çizelge 6’da verilmiştir.

Çizelge 5. Tüm değişkenler için Cox regresyon çözümlemesinin sonuçları.

Değişken	β	Std. Hata	p-değeri	$\exp(\beta)$	Güven Aralığı (%95)
Yaş			0.6327		
YŞ(2)	-0.3893	0.5489	0.4781	0.6775	0.2311 – 1.9867
YŞ(3)	0.1572	0.4970	0.7517	1.1702	0.4418 – 3.0994
YŞ(4)	-0.0975	0.5081	0.8478	0.9071	0.3351 – 2.4556
YŞ(5)	0.0635	0.5676	0.9109	1.0656	0.3503 – 3.2411
Sigara Tüketimi			0.0197		
ST(2)	0.5773	0.5702	0.3113	1.7813	0.5826 – 5.4458
ST(3)	0.4200	0.5901	0.4452	1.5219	0.5178 – 4.4729
ST(4)	1.2452	0.5762	0.0307	3.4737	1.1228 – 10.7467
Boyut			0.0045		
BY(2)	-0.4185	0.3660	0.2528	0.6580	0.3212 – 1.3482
BY(3)	0.4876	0.3460	0.1588	1.6285	0.8265 – 3.2087
BY(4)	0.7678	0.2962	0.0095	2.1551	1.2059 – 3.8512
Patalojik Evre			0.0000		
PE(2)	0.2117	0.3372	0.5301	1.2358	0.6382 – 2.3933
PE(3)	0.7496	0.3017	0.0130	2.1162	1.1715 – 3.8228
PE(4)	1.8474	0.3632	0.0000	6.3431	3.1128 – 12.9258
ER	0.1039	0.2925	0.7223	1.1095	0.6254 – 1.9683
İnvazyon	0.2622	0.2699	0.3314	1.2998	0.7658 – 2.2060

Çizelge 6. Tüm değişkenler ve ERxlogt değişkeni için Cox regresyon çözümlemesinin sonuçları.

Değişken	β	Std. Hata	p-değeri	$\exp(\beta)$	Güven Aralığı (%95)
Yaş					
YŞ(2)	-0.38758	0.55083	0.4817	0.679	0.231 - 1.998
YŞ(3)	0.16755	0.49864	0.7369	1.182	0.445 - 3.142
YŞ(4)	-0.13102	0.50925	0.7970	0.877	0.323 - 2.380
YŞ(5)	0.06056	0.56903	0.9152	1.062	0.348 - 3.241
Sigara Tüketimi					
ST(2)	0.59570	0.57355	0.2990	1.814	0.590 - 5.584
ST(3)	0.41818	0.55275	0.4493	1.519	0.514 - 4.489
ST(4)	1.30903	0.58018	0.0241	3.703	1.188 - 11.544
Boyut					
BY(2)	-0.39598	0.36390	0.2765	0.673	0.330 - 1.373
BY(3)	0.47318	0.34806	0.1740	1.605	0.811 - 3.175
BY(4)	0.80128	0.29691	0.0070	2.228	1.245 - 3.988
Patalojik Evre					
PE(2)	0.23528	0.33965	0.4885	1.265	0.650 - 2.462
PE(3)	0.82261	0.30363	0.0067	2.276	1.255 - 4.128
PE(4)	1.92017	0.36612	0.0001	6.822	3.329 - 13.982
İnvazyon	0.24625	0.27299	0.3670	1.279	0.749 - 2.184
ER	2.81917	0.90804	0.0019	16.763	2.828 - 99.374
ERxlogt	-1.01051	0.33079	0.0023	0.364	0.190 - 0.696

Test istatistiği yokluk hipotezi “ $H_0 : \delta = 0$ ” altında 1 serbestlik dereceli ki-kare dağılımına sahiptir. Buna göre test istatistiği,

$$\begin{aligned} LR &= -2 \ln \hat{L}_{\text{Cox Regresyon Modeli}} - (-2 \ln \hat{L}_{\text{Genisletilmiş Cox Regresyon Modeli}}) \\ &= 896.710 - 886.438 = 10.272 \end{aligned}$$

biçimindedir. Olabilirlik oranı test istatistiği 10.272 olarak bulunmuştur ve $\chi^2_1 = 3,86$ olduğundan yokluk hipotezi %95 güven düzeyinde kabul edilemez. Bu durumda ER değişkeninin orantılı tehlike varsayımını sağlamadığı söylenebilir.

V. SONUÇLAR

Yaşam çözümlemesinde yaygın olarak kullanılan yarı parametrik bir yöntem olan Cox regresyon çözümlemesinin kullanılabilmesi için orantılı tehlike varsayımının incelenmesi gerekmektedir. Ancak uygulamalarda bu varsayım incelenmeden varsayımın doğru olduğu kabul edilerek çözümlemeler yapılmaktadır. Klasik istatistiksel yöntemlerde model varsayımları incelenmekte ve varsayımlar sağlanmaz ise farklı yöntemler kullanılarak çözümlemeler yapılmaktadır. Yaşam çözümlemesinde ise tehlikelerin orantılı olmaması durumunda tabakalandırılmış Cox regresyon modeli ve genişletilmiş Cox regresyon modeli kullanılmaktadır.

Yaşam çözümlemesinde, yaşam süresi üzerinde önemli olan faktörleri belirlemek için orantılı tehlike modelini doğrudan kullanmadan önce orantılılık varsayımının sağlanıp sağlanmadığı incelenmelidir. Bu çalışmada, orantılı tehlike varsayımının incelenmesinde kullanılan grafiksel yöntemlerden log(-log) yaşam eğrileri, gözlenen ve beklenen yaşam eğrileri ve Arjas grafikleri, zamana bağlı açıklayıcı değişkenlerin kullanılması yöntemi ve Schoenfeld artıkları ile yaşam süresinin rankının korelasyon testi yöntemi hakkında bilgi verilmiştir.

Uygulama bölümünde ise 236 akciğer kanseri hastalarına ait veriler kullanılarak incelenen yöntemlerin uygulaması yapılmıştır. Bu yöntemlerin benzer sonuçlar verdiği gözlenmiştir. Analiz sonucunda, yaş, sigara tüketimi, genişletilmiş rezeksiyon, tümör boyutu, invazyon ve patolojik evre değişkenlerinden genişletilmiş rezeksiyon değişkeninin orantılı tehlike varsayımını sağlamadığı görülmüştür. Bu durumda, genişletilmiş rezeksiyon değişkeni için tehlike oranlarının zamana karşı sabit olmadığı ifade edilebilmektedir.

Orantılı tehlike varsayımının sağlanmaması durumunda kullanılması uygun olan tabakalandırılmış Cox regresyon modelinin ve genişletilmiş Cox regresyon modelinin incelenmesi bundan sonraki çalışmalar için önerilebilmektedir.

KAYNAKLAR

- [1] J.D. Kalbfleisch ve R.L. Prentice, “*The Statistical Analysis of Failure Time Data*”, Wiley, New York, 1980.
- [2] D.G. Kleinbaum, D.G., “*Survival Analysis: A Self-Learning Text*”, Springer, New York, 1996.
- [3] E. Arjas, “A graphical method for assessing goodness of fit in Cox’s proportional hazards model”, *Journal of the American Statistical Association*, Vol.83, pp.204-212, 1988.
- [4] D. Schoenfeld, “Partial residuals for the proportional hazards model”, *Biometrika*, Vol.69, pp.551-55, 1982.
- [5] P.M. Grambsch ve T.M. Therneau, “Proportional hazards tests and diagnostics based on weighted residuals”, *Biometrika*, Vol.81, pp. 515-526, 1994.
- [6] D. Sertkaya, D., Değişim Noktalı Sabit Hazard Modeli, Bilim Uzmanlığı Tezi, H.Ü. Fen Bil. Enstitüsü, 1996.

- [7] D.Collett, "*Modelling Survival Data in Medical Research*", Chapman&Hall, London, 1994.
- [8] D.R. Cox, "Regression models and life-tables", *Journal of the Royal Statistical Society, Series B*, Vol.34, pp.187-220, 1972.
- [9] J.R. Lawless, "*Statistical Methods and Model for Lifetime Data*", Wiley&Sons, New York, 1982.
- [10] D.R. Cox ve D. Oakes, "*Analysis of Survival Data*", Chapman and Hall, New York, 1984.
- [11] T.M. Therneau ve P.M. Grambsch, "*Modelling Survival Data: Extending the Cox Model*", Springer, New York, 2000.
- [12] J.P. Klein ve M.L. Moeschberger, "*Survival Analysis Techniques for Censored and Truncated Data*", Springer, New York, 1997.
- [13] I. Persson, "*Essays on the Assumption of Proportional Hazards in Cox Regression*", <http://www.diva-portal.org>, 2002.
- [14] D.W. Hosmer ve S. Lemeshow, "*Applied Survival Analysis: Regression Modelling of Time to Event Data*", Wiley&Sons, New York, 1999.
- [15] D.G. Kleinbaum, Yazılı Görüşme, Emory University, Department of Biostatistics, Atlanta, Georgia, dkleinb@sph.emory.edu, 2004.