Available online at www.alphanumericjournal.com

alphanumeric journal
The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

Volume 5, Issue 1, 2017

Received: September 7, 2016
Accepted: March 9, 2017
Published Online: June 30, 2017

AJ ID: 2017.05.01.STAT.02
DOI: 10.17093/alphanumeric.323833

Tied Survival Times In Survival Analysis

Durdu Karasoy | Department of Statistics, Hacettepe University, Turkey, durdu@hacettepe.edu.tr

Sena Keskin Kaplan | Student Selection and Placement Center, Turkey, sena.keskin@osym.gov.tr

ABSTRACT

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. One of the methods commonly used in the survival analysis is Cox regression model which is used to determine the factors that impact on survival times. Cox regression model has assumptions. One of them is proportional hazards assumption and the another one is there is no tied data between event times. However, in real applications, tied event times are commonly observed and Cox's partial likelihood function needs to be modified to handle ties. It is well known methods that the Exact method, Breslow method, Efron method and Discrete method for handling tied event times. Firstly, the methods are analyzed during the study, Breslow, Efron and Exact methods, which is applied on a stomach cancer data set (there is tied data between event times) It was decided that Cox regression with Exact Method is the best model. Than this methods is applied Acute Myocardial Infarction data set which has no tied data between event times and it is found the same results at all methods.

Keywords: Survival Analysis, Cox Regression, Tied Survival Times

Yaşam Çözümlemesinde Eş Zamanlı Yaşam Süreleri

ÖZET

Yaşam çözümlemesi, tanımlanan herhangi bir olayın belirli bir başlangıç noktasından, ortaya çıkmasına kadar geçen sürenin incelenmesinde kullanılan istatistiksel yöntemler topluluğudur. Yaşam çözümlemesinde sıkça kullanılan yöntemlerden biri yaşam süresi üzerinde etkili olan faktörlerin belirlenmesinde kullanılan Cox regresyon modelidir Cox regresyon modelinin orantılı tehlikeler varsayımına ek olarak bir diğer varsayımı ise eş zaman durumunun meydana gelmemiş olmasıdır. Ancak çalışmalarda genellikle eş zamanlı olarak meydana gelen başarısızlıklara rastlanmaktadır ve bu durum özel çözüm gerektirmektedir. Kesin yöntem, Breslow yöntemi, Efron yöntemi ve Kesikli yöntem olarak bilinen yöntemler bu özel çözümlemelerdir. Çalışma boyunca incelenen yöntemlerden Breslow yöntemi, Efron yöntemi ve Kesin yöntem, eş zamanlı gözlemlerin olduğu duruma örnek olarak mide kanseri verilerine uygulanmış Kesin yöntem ile Cox regresyon modelinin en iyi model olduğuna karar verilmiştir. Daha sonra ise eş zamanlı gözlemlerin olmadığı Akut Miyokard İnfarktüsü verilerine uygulanmış ve sonuçların aynı olduğu gözlenmiştir.

Anahtar Kelimeler: Yaşam Çözümlemesi, Cox Regresyon, Eş Zamanlı Yaşam Süreleri

1. Giriş

Yaşayan bir organizmanın ya da cansız bir nesnenin belirli bir başlangıç zamanı ile ölümü arasında geçen zamana ‘‘yaşam süresi’’ ya da ‘‘başarısızlık süresi’’ adı verilir. Her bir birime ait yaşam süresi T, tanımı gereği sürekli ve pozitif bir değere sahiptir (Johnson and Johnson, 1980).

Bir birimin başarısızlığında zamanın etkisi olduğu kadar bazı özelliklerin de etkisi vardır. Araştırmacıların, bu özellikleri (değişkenleri) modele katma çalışmaları 1970’lere kadar pek yokken, Cox (1972)’un, Cox regresyon modeli ile ilgili makalesi çalışmalara yeni bir yön vermiştir (Johnson and Johnson, 1980).

Cox’un kısmi olabilirlik fonksiyonu gözlemler arasında eş zaman durumunun olmadığını varsaymaktadır (Collett, 1994; Xin, 2011). Eğer gözlemler arasında eş zaman durumu mevcut ise veya başarısızlık zamanları tam olarak belli değilse Cox regresyon modelini kullanmak uygun değildir. Ancak pratikte eş zamanlı yaşam süreleri sıklıkla gözlenmektedir ve bu durum özel çözüm gerektirmektedir. Yaşam süreleri, sürekli zaman ölçeği ile ölçülmüş ise eş zamanlı gözlemler büyük ihtimalle belirsiz ölçümler sonucunda elde edilmiştir. Bu durumda ölçüm daha hassas şekilde kaydedilmeli, kaydedilemiyorsa olabilecek tüm sıralamalar düşünülmelidir. Eğer yaşam süreleri kesikli zaman ölçeği ile ölçülmüş ise eş zamanlı gözlemler, gerçekten eş zamanlıdır yani olaylar aynı zamanda meydana gelmiştir (Lee and Wang, 2003).

Veri setlerinde eş zamanlı olarak meydana gelen yaşam süreleri sıklıkla gözlenmektedir. Bu durumda Cox’un kısmi olabilirlik fonksiyonu eş zamanlı gözlemler olduğu durum için yeniden düzenlenmelidir (Xin, 2011). Bu yöntemler, Breslow (1974) tarafından önerilen Breslow yöntemi, Efron (1977) tarafından önerilen Efron yöntemi, Kesin yöntem (DeLong vd., 1994; Kalbfeisch and Prentice, 2002) ve Kesikli (discrete) yöntemdir (Kalbfleisch and Prentice, 1973).

Çalışmada, eş zamanlı yaşam süreleri hakkında ve yaşam çözümlemesinde eş zamanlı yaşam süreleri durumunda kullanılan yöntemler hakkında bilgilere yer verilmiştir. Daha sonra anlatılan yöntemlerden veri setine uygun olanları öncelikle eş zamanlı gözlemlere sahip olan mide kanseriyle ilgili verilere uygulanmış ve en iyi yönteme ait sonuçlar yorumlanmıştır. Daha sonra ise eş zamanlı gözlemlerin olmadığı Akut Miyokard İnfarktüsü (AMI-Kalp krizi) verilerine bu yöntemler uygulanmış sonuçlar değerlendirilmiştir.

2. Yaşam Çözümlemesi

Yaşam çözümlemesi, belirli bir başlangıç noktasında başlayan (hastalığın başlangıcı, tedavinin başlangıcı) izleme sürecindeki olayların sonuca varmasına kadar geçen süredeki verilerin analizidir. Olay; ölüm, tedaviye yanıt almak gibi farklı durumlarla sonuçlanabilir (Yay vd., 2007).

Yaşam verilerinin çözümlenmesinde karşılaşılan temel güçlük, gözlem altına alınan bazı birimlerin başarısızlık zamanlarının gözlenememiş olmasıdır (Sertkaya vd., 2005). İzleme sürecinde izlenen olay meydana gelmeyebilir, çalışma periyodu boyunca birim izlenememiş olabilir veya ilgilenilen olay dışında başka bir nedenle birim çalışmadan çıkmış olabilir. Bu üç durumdan biri meydana geldiğinde durdurulmuş (censored) veri ortaya çıkar.

Yaşam sürelerinin dağılımları genel olarak üç fonksiyonla gösterilir. Bunlar; yaşam fonksiyonu, olasılık yoğunluk fonksiyonu ve tehlike fonksiyonudur. Bu fonksiyonlar matematiksel olarak birbirinden elde edilebilir (Lee and Wang, 2003).

T rastlantı değişkeni f(t) olasılık dağılımına sahiptir. f(t) olasılık yoğunluk fonksiyonu

$$f(t)=\lim_{\delta t\to 0}\frac{P(\,t\le T\le t+\delta_t)}{\delta_t} \quad (1)$$

şeklindedir. T'nin dağılım fonksiyonu,

$$F(t) = P(T < t\,) = \int_0^t f(u)du \quad (2)$$

biçimindedir ve yaşam süresinin t'den küçük olması olasılığını gösterir.

Yaşam fonksiyonu S(t), yaşam süresinin t'den daha büyük olması olasılığını verir ve

$$S(t) = P(\,T \ge t\,) = 1- F(t) \quad (3)$$

biçimindedir (Collett, 1994).

Yaşam fonksiyonu monoton azalan soldan sürekli bir fonksiyondur ve

$$t=0 \text{ iken; } S(t)=S(0)=1 \quad (4)$$

$$t=\infty \text{ iken; } S(t)=S(\infty)=0 \quad (5)$$

olur (Kleinbaum, 1996).

Tehlike fonksiyonu, t zamanında yaşadığı bilinen bir birimin t ile t+δt zaman aralığında başarısız olma riskinin bir tanımıdır. Anlık ölüm hızı olarak da tanımlanır.

Tehlike fonksiyonu,

$$h(t) = \lim_{\delta t\to 0}\frac{P(\,t\le T\le t+\delta_t/\,T\ge t)}{\delta_t} \quad (6)$$

biçimindedir. h(t), 0 ile ∞ arasında değer alır (Collett, 1994).

Yaşam çözümlemesinde sıkça kullanılan yöntemlerden biri, yaşam süresi üzerinde etkili olan faktörlerin belirlenmesinde kullanılan Cox Regresyon Modelidir (Cox, 1972).

2.1. Cox Regresyon Modeli

İlk olarak 1972 yılında Cox tarafından ele alınan Cox regresyon modeli yaşam çözümlemesinde yaygın olarak kullanılan yarı parametrik bir yöntemdir. Orantılı tehlikeler modeli (proportional hazards model) olarak da bilinir.

Cox regresyon modeli,

$$h(t,\boldsymbol{X})=h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{X}) \quad (7)$$

biçimindedir. Burada, **β** regresyon katsayıları vektörü, $h_0(t)$ açıklayıcı değişkene sahip olmayan (**x=0** olan) bir birimin temel tehlike fonksiyonudur (Sertkaya vd., 2005).

Cox regresyon modeli, parametrik modellerin gerektirdiği varsayımların sağlanmadığı durumlarda, parametrik yöntemlerden daha etkilidir. Cox regresyon modelinin temel varsayımı, tehlikelerin orantılı olması, yani tehlike oranının zaman boyunca değişmemesi, sabit olmasıdır. Tehlike oranı ile ilgili bu varsayım orantılı tehlikeler varsayımı olarak bilinir (Arı ve Önder, 2013).

Orantılı Tehlikeler Varsayımı

Cox orantılı tehlikeler modelinin temel varsayımı orantılı tehlikelerdir. Orantılı tehlike varsayımı, tehlike oranının zamana karşı sabit olması ya da bir birimin tehlikesinin diğer birimin tehlikesine orantılı olması anlamına gelmektedir (Therneau and Grambsch, 2000).

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

$$\widehat{HO} = \frac{\hat{h}(t,x^*)}{\hat{h}(t,x)} = \frac{\exp(\sum_{j=1}^{p}\widehat{\beta_j}x_j^*)\widehat{h_0}(t)}{\exp(\sum_{j=1}^{p}\widehat{\beta_j}x_j)\widehat{h_0}(t)} = \exp\left[\sum_{j=1}^{p}\widehat{\beta_j}(x_j^* - x_j)\right] \tag{8}$$

biçimindedir. Üstteki eşitlikte görüldüğü gibi tehlike oranı t'yi içermez. Bir başka deyişle, model uydurulduğunda x* ve x için değerler belirlendiğinde, tehlike oranı tahmini için üstel ifadenin değeri sabittir, zamana bağlı değildir. Bu sabit θ ile gösterilirse, tehlike oranı

$$\hat{\theta} = \frac{\hat{h}(t,\boldsymbol{x}^*)}{\hat{h}(t,\boldsymbol{x})} \tag{9}$$

biçiminde yazılabilir. Burada, $\hat{\theta}$ orantılılık sabiti (proportionality constant) olarak adlandırılır ve zamandan bağımsızdır (Collet,1994; Ata, 2005).

Orantılı tehlikeler varsayımını incelemek için grafiksel ya da sayısal yöntemler kullanılmaktadır. Orantılı tehlike varsayımının incelenmesinde en çok kullanılan yöntemler, log(-log) yaşam eğrileri, gözlenen ve beklenen yaşam eğrileri, Arjas grafikleri, modele zamana bağlı değişkenlerin eklenmesi, Schoenfeld artıkları ile yaşam süresinin rankı arasındaki korelasyon testi biçiminde sıralanabilmektedir (Ata, 2005).

3. Eş Zamanlı Yaşam Süreleri

Yaşam verileri için kullanılan Cox regresyon modeli, tehlike fonksiyonunun sürekli bir fonksiyon olduğunu varsayar ve bu varsayım altında eş zamanlı yaşam sürelerinin (tied survival times) oluşması mümkün değildir. Ancak, yaşam süreleri en yakın gün, ay veya yıla göre kaydedilmekte olup bu nedenle yuvarlanan süreler sonucunda eş zamanlı yaşam süreleri ortaya çıkabilmektedir (Collett, 1994).

Xin (2011) çalışmasında, iki olay eş zamanlı ise, olayların tam olarak aynı zaman noktasında meydana gelmiş olacağını belirtmiştir. Bu bilgi doğrultusunda "eş zamanlı gözlem", incelenen olayların tam olarak aynı zaman noktasında meydana gelmesiyle oluşmaktadır (Xin, 2011). Ayrıca bir başarısızlık zamanında bir veya daha fazla durdurulmuş gözlem de olabilir. Bir zaman noktasında hem durdurulmuş yaşam süreleri hem de ölüm meydana gelebilir, durdurmanın ölümden sonra geldiği varsayılır (Collett, 1994).

Eş zamanlı gözlem durumu, belli aralıklarla ölçüm yapılması (ölçüm kısıtları) veya yuvarlanarak kaydedilen olay zamanları nedeniyle meydana gelmektedir. Bazen daha hassas ölçümler yapılmasının maliyeti bilgi eklemenin öneminden daha ağır basabilir.

Böyle bir durumda başarısızlık süreleri çok hassas ölçülemeyebilir (Zhang, 1997). Örneğin ölçüm birimi yıl olarak alındığında ölüm anı aynı olmamasına karşın aynı yıl içinde ölen iki birimin olay zamanı aynı olarak kaydedilmiş olur (Zhang,1997; Xin, 2011).

Kitlenin büyük olduğu veya geniş kohort kanser çalışmalarında ölçümlerin ayda bir ya da uzun aralıklarla yapıldığı durumlarda, eş zamanlı gözlemlerin yoğun olduğu diğer durumlardır (Picciotto and Rockhill, 1997).

Cox'un kısmi olabilirlik fonksiyonu veri setinde eş zamanlı gözlem olmadığını varsaymaktadır (Picciotto and Rockhill, 1997). Eğer gözlemler arasında bir eş zamanlılık söz konusu ise veya gözlemlere ait başarısızlık süreleri tam olarak belli değilse Cox regresyon modelini kullanmak uygun değildir. Bu durumda Cox'un kısmi olabilirlik fonksiyonu eş zamanlı gözlemler olduğu durum için yeniden düzenlenmelidir (Collett, 1994; Zhang, 1997; Picciotto and Rockhill, 1997; Xin, 2011). Eğer eş zamanlı başarısızlık süreleri bir belirsizlik sonucunda elde edilmişse bu durumdan kurtulmak için basit bir yaklaşım, yeniden yapılandırma olabilir. Yani, eğer mümkünse gözlemler yeniden incelenerek olay zamanları daha hassas ölçümler sonucunda yeniden kaydedilmelidir (Zhang, 1997; Xin, 2011).

3.1. Eş Zamanlı Yaşam Süreleri Tahmin Edici Süreci

n gözlem için, k farklı başarısızlık (durdurulmamış) süreleri $t_{(1)} < t_{(2)} <\ldots< t_{(k)}$ olsun. $m_{(i)}$, $t_{(i)}$ zamanında başarısız olan birim sayısını göstersin. $m_{(i)}>1$ ise $t_{(i)}$ değerinde birden fazla, $m_{(i)}=1$ ise $t_{(i)}$ zamanında bir tane gözlem başarısız olmuş demektir. $R(t_{(i)})$, $t_{(i)}$ zamanda riskte olan birimlerin kümesi; r_i ise $R(t_{(i)})$'deki birim sayısı olsun (Lee and Wang, 2003).

Gözlemler sürekli zaman ölçeği ile elde edildiğinde, eş zamanlı gözlemler belirsiz ölçümler sonucunda elde edildiğinden belirli bir zaman noktasında başarısız olan birimler için yaşam sürelerinin aynı olduğunu söylemek doğru olmaz. Çünkü, eğer ölçümler daha hassas bir şekilde yapılabilirse, bu birimlere ait yaşam süreleri sıralanabilir ve Eşitlik 10'da verilen olabilirlik fonksiyonu kullanılabilir (Lee and Wang, 2003):

$$L(\beta)=\prod_{i=1}^{k}\frac{\exp(\beta' x_{(i)})}{[\sum_{l\in R(t_{(i)})}\exp(\beta' x_l)]} \qquad (10)$$

Eğer gerçek sıra bilgisi mevcut değilse gözlenen eş zamanlı yaşam sürelerinin mümkün olan tüm sıralamaları düşünülmelidir. Her bir başarısızlık zamanında gözlenen eş zamana sahip birimler için ($m_{(i)}$), $m_{(i)}!$ tane farklı sıralama yapılabilir. Bu nedenle elimizde sürekli zaman ölçümüne sahip yaşam süreleri olduğu zaman $m_{(i)}$ büyükse kısmi olabilirlik fonksiyonunun kurulumu ve hesabı oldukça sıkıntılı olabilmektedir (Collet, 1994; Picciotto and Rockhill, 1997; Chalita vd., 2002; Lee and Wang, 2003).

Delong vd. (1994) tarafından önerilen eşitlik eş zamanlı sürekli yaşam süresi olduğunda kısmi olabilirlik fonksiyonunun hesaplamasını daha uygulanabilir yapmaktadır. Aynı zamanda istatistiksel yazılım paketleri, örneğin SAS ve STATA kısmi olabilirlik fonksiyonuna dayalı prosedürler içerir (Lee and Wang, 2003).

Yaşam süreleri kesikli zamanda gözleniyorsa eş zamanlı gözlemler, gerçekten eş zamanlıdır yani olaylar aynı zamanda meydana gelir. Bu durumda Cox'un önerdiği model aşağıda verilmiştir:

$$\frac{h_i(t)d_t}{1-h_i(t)d_t} = \frac{h_0(t)d_t}{1-h_0(t)d_t}\exp(\textstyle\sum_{j=1}^{p} b_j x_{ji}) = \frac{h_0(t)d_t}{1-h_0(t)d_t}\exp(b'x_i) \quad (11)$$

Bu model sürekli zaman ölçeğinde aşağıdaki eşitliğe dönüşmektedir:

$$h(t\backslash x)= h_0(t)\exp(\textstyle\sum_{j=1}^{p} b_j x_j)= h_0(t)\exp(b'x_i) \quad (12)$$

Kesikli zaman ölçeğinde eş zamanlı gözlemlerle kısmi olabilirlik fonksiyonu,

$$L_d(d)=\prod_{i=1}^{k} \frac{\exp(z'_{u^*(i)}b)}{\sum_{u_j\in U_i}\exp(z'_{u_j}b)} \quad (13)$$

biçimindedir. Bu eşitlikteki i. terim gözlenen $m_{(i)}$ başarısızlıklarının koşullu olabilirliğini gösterir. $m_{(i)}$, $t_{(i)}$ zamanında ve $R_{(t(i))}$ risk kümesindeki başarısızlıklardır. Paydadaki i. terimlerin sayısı ${}_{ri}C_{m(i)}=r_i!/[m_{(i)}!(r_i-m_i)!]$ dir. $m_{(i)}$ büyükse sonuç çok büyük çıkabilir. Ancak Gail ve diğerleri (1981) yaptıkları çalışmada kullandığı bir algoritma ile bu değeri hesaplamayı başarmıştır (Lee and Wang, 2003).

3.2. Eş Zamanlı Yaşam Süreleri Oluştuğu Durumda Kullanılan Yöntemler

Yaşam çözümlemesinde iki ya da daha fazla birim aynı yaşam süresini paylaştığında eş zamandan söz edilmektedir. Süre genellikle sürekli değişken olarak incelenmektedir.

Cox regresyon modelinin bir varsayımı, veri setinde eş zamanlı süreler olmamasıdır (Picciotto and Rockhill, 1997; Xin, 2011). Fakat gerçek uygulamalarda eş zamanlı olay süreleri genellikle gözlenmektedir. Bu durumda Cox'un kısmi olabilirlik fonksiyonu yeniden düzenlenmelidir.

Eş zamanlı yaşam verileri olduğu durumda kullanılan yöntemler geliştirilmiştir. Bu yöntemler, Kesin (Exact) yöntem, Breslow yöntemi, Efron yöntemi ve Kesikli yöntemdir.

Kesin Yöntem

Bu yöntem yaşam sürelerinin sürekli bir dağılıma sahip olduğunu ve eş zamana sahip birimlerin yaşam sürelerinin gerçekte farklı olduğunu varsaymaktadır. Bu yöntemde, eş zamanlı gözlemler, sürekli zamanda kaydedilen belirsiz ölçümler sonucunda oluşmaktadır. Gerçekte olayların öncelikli bir sırası vardır. Bu nedenle uygun model için kısmi olabilirlik hesaplanmasında mümkün olan tüm sıralamaların değerlendirilmesi gerekir (Allison, 2010; http://www4.stat.ncsu.edu/~dzhang2/st745/chap7.pdf)

Kesin yöntemde kısmi olabilirlik mümkün olan tüm sıralamaların olasılıklarının toplamını içermektedir.

Kesin yöntem, oldukça açık ve kusursuz bir yöntemdir, fakat permütasyona dayandığından çok fazla eş zamanlı gözlem içeren çok sayıda zaman noktası olduğunda bu yöntemin hesaplanabilirliği olanaksız (infeasible) olabilmektedir (Allison, 2010; Xin, 2011).

Kesin yöntem, tüm eş zamanlı olay sürelerinin aynı veya daha büyük değere sahip durdurma sürelerinden önce meydana geldiği varsayımına dayanan orantılı tehlikeler varsayımı altında kesin koşullu olasılığı (exact conditional probability) hesaplar. Bu eşitlik, gözlenen veri ile tutarlı olan marjinal olabilirlik terimlerinin tümünün toplamına denktir (http://www.medicine.mcgill.ca/ epidemiology/hanley/c681/cox/TiesCox ModelR.txt).

Bazı kaynaklarda Kesin yöntem iki şekilde hesaplanmaktadır. İlki yaşam sürelerinin sürekli zaman ölçümü ile kaydedildiği durumda önerilmektedir. Yaşam süreleri sürekli zaman ölçümü ile kaydedildiğinde eş zamanlı gözlem meydana gelmez. Eğer meydana gelirse, olabilirlik, eş başarısızlık zamanlarının başarısızlık olmayan olay zamanlarından önce meydana geldiği marjinal olasılığı (marginal probability) yansıtmaktadır. Çünkü başarısızlık olmayan zamanlar için sıra önemli değildir. Bu, kesin marjinal olabilirlik (exact marginal likelihood) olarak adlandırılır ve "exactm" olarak ifade edilebilir.

İkinci durum yaşam sürelerinin kesikli olduğu durumdur. Burada eş zamanlı başarısızlıkların oluşması beklenmektedir. Olabilirlik, kesikliliği yansıtacak şekilde değiştirilir ve gözlenen başarısızlıkların koşullu olasılığı hesaplanır. Bu yöntem de kesin kısmi olabilirlik (exact partial likelihood) olarak adlandırılır ve "exactp" olarak ifade edilebilir **(**https://files.nyu.edu/mrg217/public/cox.pdf).

Breslow Yöntemi

Breslow yöntemi, başarısızlık sürelerinin sürekli olduğunu ve olay riskinin (t_i, t_{i+1}) aralığında sabit olduğunu varsaymaktadır (Breslow, 1974). Ayrıca, (t_i, t_{i+1}) aralığına düşen durdurma süreli bir bireyin, aralığın başında durdurulmuş olduğu varsayılır ve bu süre t_i olarak ifade edilir.

$\mathbf{z}_j$, j. bireyin açıklayıcı değişken değerlerinin vektörü; Di, t_i zamanında başarısız olan di sayıda bireyden oluşan küme olsun. R_i, t_i zamanındaki risk kümesi yani t_i zamanında yaşayan (risk altında olan) tüm bireyleri içerir. Breslow yöntemi için kısmi olabilirlik Eşitlik 3.2'deki gibidir (Picciotto and Rockhill, 1997; Xin, 2011):

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp[(\sum_{j\in D_i} z_j)\beta]}{[\sum_{j\in R_i} \exp(z_j\beta)]^{d_i}} \tag{14}$$

Hesaplama bakımından Breslow yöntemi, Kesin yöntemin kısmi olabilirliğini hesaplamadan çok daha basittir. Fakat belirli zaman noktasındaki eş zamanlı gözlemlerin sayısı, risk kümesindeki hastaların sayısına göre daha fazla olduğunda Breslow yönteminin verdiği sonuçlar bozulmaktadır (Zhang, 1997; Xin, 2011).

Efron Yöntemi

Efron yöntemi, genellikle Breslow yönteminden daha iyi bir yöntem olarak incelenmektedir.

Eş zamanlı gözlemlerin yüzdesi arttıkça, tüm yöntemlerin performansı daha kötüye gider, çünkü kısmi olabilirlikleri kesin kısmi olabilirlikten (exact partial likelihood) farklı olur. Efron yöntemi için de aynı durum söz konusu olsa da diğer yöntemlerden daha doğru sonuç verir (Kalbfleisch and Prentice, 2002; Allison, 2010).

Efron yöntemi için kısmi olabilirlik fonksiyonu

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp[(\sum_{j\in D_i} z_j)\beta]}{\prod_{l=1}^{d_i}[\sum_{j\in R_i} \exp(z_j\beta) - \frac{l-1}{d_i}\sum_{j\in D_i} \exp(z_j\beta)]} \tag{15}$$

biçimindedir (Picciotto and Rockhill, 1997).

Breslow yöntemi ile karşılaştırma yapıldığında, paydalar Breslow yöntemindeki kadar basit olmadığından Efron yönteminin kısmi olabilirliği, Kesin yönteme daha yakındır. Efron yönteminin hesaplanması ise Kesin yöntemden daha basit olmaktadır. Efron yöntemi, Kesin ve Breslow yöntemleri arasında yer alır (Picciotto and Rockhill, 1997; Xin, 2011).

Kesikli Yöntem

Kesikli yöntem, yaşam süresinin gerçekten kesikli olduğunu varsayar (Cox, 1972; Allison, 2010).

Kesikli yöntem, olaylar aynı zaman noktasında meydana gelmiş gibi işlem yapar. Cox, kesikli zaman verileri için benzer bir model önermiştir. Bu model orantılı odds model olarak adlandırılmaktadır (Cox, 1972; Allison, 2010). Fakat, bu modelin hesaplamaları zaman aldığından Kalbfleisch ve Prentice (1973) tarafından başka bir yöntem geliştirilmiştir.

d_i, t_i zamanında başarısız olan birimlerin sayısı olsun. q, d_i nin bir alt kümesi; Q_i, tüm mümkün q'ları içeren bir küme olsun. Q_i, t_i zamanındaki risk kümesinden d_i büyüklüğünde tüm alt kümelerin kümesi olsun. $s_q^* = \sum_{j=1}^{d_i} z_{q_j}$,spesifik bir alt küme q'nun açıklayıcı değişken değerlerinin toplamı olsun. Bu durumda, Picciotto and Rockhill'in (1997) çalışmalarındaki gibi, Kalbfleisch ve Prentice (1973) tarafından önerilen kesikli yöntem için kısmi olabilirlik aşağıda verildiği gibidir:

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp[(\sum_{j \in D_i} z_j)\beta]}{\sum_{q \in Q_i} \exp(s_q^* \beta)} \quad (16)$$

Breslow ve Efron yöntemlerinde olduğu gibi kesikli yöntemin hesaplanması da kesin yöntemden daha basittir. Fakat, çalışmalarda olay oranı arttıkça ve ölçüm birimi belirsizse çok daha fazla eş zamanlı olay meydana gelmekte ve kesikli yöntemin parametre tahminleri daha yanlı olabilmektedir (Kalbfleisch and Prentice, 1973; Picciotto and Rockhill., 1997; Xin, 2011).

Yöntemlerin Karşılaştırılması

Kesin yöntem, diğer yöntemlere göre en iyi sonuç veren yöntem olarak belirtilmektedir. Ancak, bu yöntemin hesaplaması zor ve veri seti büyük olduğunda pratik olmayabilmektedir.

Breslow ve Efron yöntemlerinin her ikisinde de hesaplamalar daha basit ve hızlıdır. Ancak özellikle veri setinin büyüklüğüne bağlı olarak eş zamanlı gözlem sayısının fazla olduğu durumlarda, Efron yöntemi Breslow'dan daha doğru sonuçlar vermektedir.

Eş zamanlı olay sürelerinin çözümünde en iyi bilinen Efron ve Breslow yöntemi, eş zamanlı gözlemlerin sayısı arttıkça bozulmaktadır (Xin, 2011).

Picciotto ve Rockhill (1997) çalışmalarında Breslow'un (1974), Efron'un (1977) ve Kalbfleisch ve Prentice'in (1973) yöntemlerini karşılaştırmışlar ve eş zamanlı gözlemlerin yoğunluğu arttıkça bu yöntemlerinin yanlılığının artmakta olduğu sonucuna varmışlardır. Özellikle orta ve yoğun sıklıkta eş zamanlı gözlem olduğunda Efron yönteminin diğer iki yöntemden çok daha iyi sonuç verdiğini belirtmişlerdir. Breslow yöntemi çoğu standart yazılım programlarında kullanıma hazır olarak bulunmasına rağmen Efron yöntemi özellikle örneklem büyüklüğü küçükse ya da veri setinde çok fazla durdurulmuş gözlem mevcutsa eş zamanlı gözlemleri çözümlemede tercih edilmektedir.

Olabilirliklerin tümü aynı paya sahiptir ve başarısızlıklar arasında eş zamanlılık ve durdurma olmadığını varsaymaktadır.

Bazı kaynaklarda eş zamanlı gözlem oluştuğu durumda Cox kısmi olabilirliğin hesaplanması için dört ayrı yöntem ile çözüm sunulmaktadır. BRESLOW, EFRON, EXACTM, VE EXACTP yöntemleri yardımı ile Cox kısmi olabilirlik hesaplanmaktadır. Eğer veri setinde eş zamanlı gözlemler yoksa hangi yöntem kullanılırsa kullanılsın sonuçlar aynı olmaktadır (https://files.nyu.edu/mrg217/public/cox.pdf ; Stefanescu and Mehrotra, 2014).

Scheike ve Sun (2007) çalışmalarında, Efron yönteminin mevcut yöntemlerden eş zamanlı yaşam verilerinde en iyi performansa sahip olduğunu belirtmişlerdir (Scheike and Sun, 2007).

4. Uygulama

Çalışmamızda öncelikle eş zamanlı gözlemlerin olduğu duruma örnek olarak mide kanseri verilerine, daha sonra ise eş zamanlı gözlemlerin olmadığı duruma örnek olarak Akut Miyokard Infarktüs (AMI) verilerine Breslow, Efron ve Kesin yöntem uygulanmış ve elde edilen sonuçlar yorumlanmıştır.

4.1. Eş Zamanlı Gözlemlerin Olduğu Durum

Bu çalışmada kullanılan veriler, Ankara Onkoloji Hastanesi'nde Ocak 1990 ve Kasım 1995 tarihleri arasında mide kanseri tanısı konulan ve cerrahi tedavi geçiren, yaşları 29 ile 84 arasında değişen 118 hastaya ait verilerdir. Hastalar ölene kadar ya da Nisan 1996'ya kadar izlenmişlerdir (Eroğlu vd., 1997; Karasoy ve Tuncer, 2015). Bu çalışmada mide kanserinden ölümü etkileyen faktörleri belirlemek için, orantılılık varsayımı incelendikten sonra Cox regresyon için eş zamanlı gözlemler durumunda kullanılan Breslow, Efron ve Kesin yöntemler kullanılarak çözümleme yapılmış ve sonuçlar karşılaştırılmıştır. Çalışmamızda anlatılan eş zamanlı yaşam sürelerinin varlığında kullanılan bir diğer yöntem olan Kesikli yöntem, kesikli ölçümlü verilere uygulandığından veri setimize uygulanmamıştır. Başarısızlık, ölüm olarak alınmıştır. Yaşam süreleri ay bazında ölçülmüştür.

Açıklayıcı değişkenler olarak cinsiyet, kilo kaybı, anemi, tümörün midedeki lokalizasyonu, lenf nodu diseksiyonunun genişliği, hastalığın evresi ve adjuvan kemoterapi alınmıştır. Bu hastaların yaş ortalaması 56,70'dir ve %58,5'i erkektir. Hastaların %55,1'i hastalığın 3. evresinde ve %19,5'i 4. evresindedir. Hastaların %71,2'si kemoterapi almıştır. Hastaların 52'sinde başarısızlık gözlenmiştir. Kullanılan değişkenlere ilişkin bilgiler Tablo 1'de verilmiştir.

Değişkenler	Düzeyler	Sıklıklar (%)
Cinsiyet	1. Erkek 2. Kadın	69 (58,5) 49 (41,5)
Kilo Kaybı	1. Yok 2. Var	90 (76,3) 28 (23,7)
Anemi	1. Yok 2. Var	31 (26,3) 87 (73,7)
Lenf nodu diseksiyonunun genişliği (Diseksiyon)	1. D0-1 2. D2-3	62 (52,5) 56 (47,5)
Tümörün midedeki lokalizasyonu (Lokal)	1. Üst üçlük 2. Orta üçlük 3. Alt üçlük 4. Tüm mide	21 (17,8) 25 (21,2) 62 (52,5) 10 (8,5)
Hastalığın evresi	1. Evre1+Evre2 2. Evre3 3. Evre4	30 (25,4) 65 (55,1) 23 (19,5)
Adjuvan kemoterapi (Kemoterapi)	1.Yok 2. Var	34 (28,8) 84 (71,2)

Tablo 1. Açıklayıcı değişkenler

Hastalara ilişkin Kaplan-Meier yaşam olasılıkları elde edildiğinde bir yıllık yaşam olasılığı %76, iki yıllık yaşam olasılığı %55, beş yıllık yaşam olasılığı ise %39,6 olarak bulunmuştur. Değişkenlere ilişkin 5 yıllık yaşam olasılıkları Tablo 2'de verilmiştir.

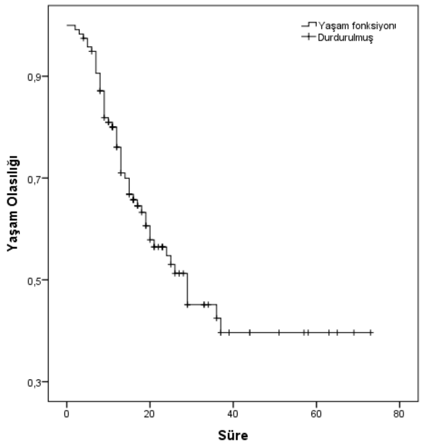
Değişken	Değişken Düzeyleri	Ortalama			5 Yıllık Yaşam Olasılığı (%)	Log-rank p
		Süre (Ay)	Std. Hata	Güven Aralığı (%95)		
Genel		39,18	3,28	32,76 - 45,61	39,6	-
Cinsiyet	1.Erkek	36,20	4,28	27,81 - 44,58	35,6	0,28
	2.Kadın	41,22	4,70	32,01 - 50,43	44,5	
Kilo Kaybı	1.Yok	40,42	3,79	33,00 - 47,83	57	0,60
	2.Var	31,16	5,37	20,64 - 41,68	30	
Anemi	1.Yok	46,81	5,64	35,75 - 57,87	56,1	0,09
	2.Var	35,58	3,78	28,17 - 42,99	52,6	
Diseksiyon	1.D0-1	30,54	3,83	23,03 – 38,05	26,2	0,00*
	2.D2-3	47,54	4,03	39,64 – 55,44	63,8	
Lokal	1.Üst üçlük	38,13	5,50	27,36 – 48,91	44	0,02*
	2.Orta üçlük	31,73	6,03	19,91 – 43,56	27,3	
	3.Alt üçlük	42,91	4,32	34,45 – 51,36	51,3	
	4.Tüm mide	14,80	2,92	9,10 – 20,52	15	
Evre	1.Evre1+Evre2	63,97	4,20	55,75 – 72,20	85,3	0,00*
	2.Evre3	25,47	2,82	19,94 - 31,00	28,4	
	3.Evre4	17,71	1,75	14,29 – 21,13	17	
Kemoterapi	1.Yok	30,37	4,68	21,21 – 39,54	31,4	0,11
	2.Var	41,77	4,10	33,74 – 49,80	60,3	

*p<0,05

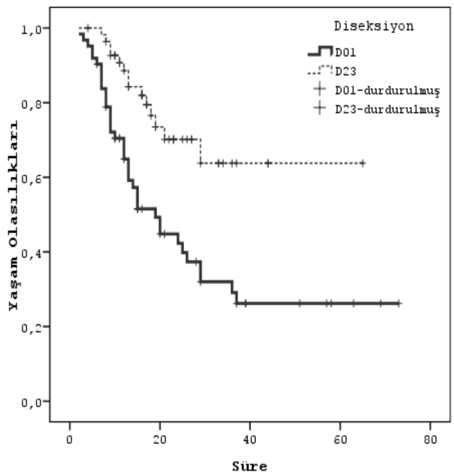
Tablo 2. Kaplan-Meier sonuçları

Elde edilen log-rank sonuçlarına göre lenf nodu diseksiyonunun genişliği, tümörün midedeki lokalizasyonu ve hastalığın evresi değişkenlerin düzeyleri arasında yaşam süreleri açısından fark olduğu %95 güven düzeyinde söylenebilir.

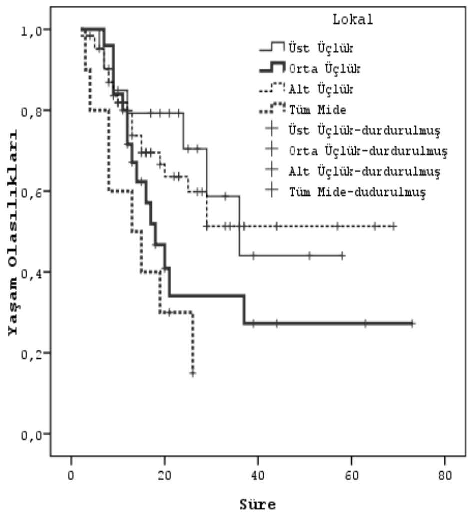
Şekil 1'de genel Kaplan Meier eğrisi, Şekil 2 – Şekil 4'de ise düzeyleri arasında yaşam olasılıkları açısından anlamlı fark bulunan açıklayıcı değişkenlere ait Kaplan Meier eğrileri verilmiştir.



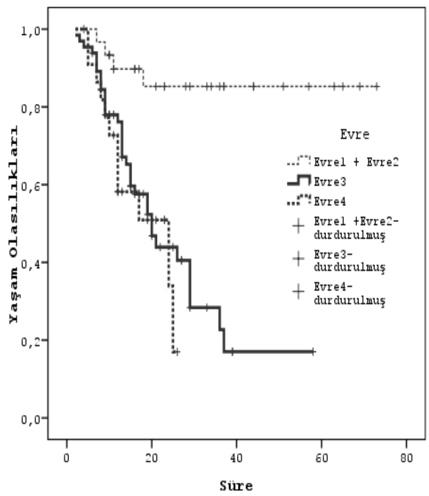
Şekil 1. Genel Kaplan-Meier eğrisi



Şekil 2. Diseksiyon değişkeni için Kaplan-Meier eğrisi



Şekil 3. Lokal değişkeni için Kaplan-Meier eğrisi



Şekil 4. Evre değişkeni için Kaplan-Meier eğrisi

Lenf nodu diseksiyonunun genişliği değişkeninin, Çizelge 4.2'de verilen log-rank testi sonucuna göre diseksiyon değişkeninin D2-3 ile D0-1 düzeyleri arasında yaşam olasılıkları bakımından %95 güven düzeyinde istatistiksel olarak önemli bir fark bulunmaktadır (p=0,00<0,05). Şekil 4.2'den de görüldüğü gibi D2-3 olan hastalar D0-1 olanlara göre daha uzun bir yaşam süresine sahiptir.

Tablo 2'de verilen log-rank testi sonucuna göre tümörün midedeki lokalizasyonu değişkeninin düzeyleri arasında yaşam olasılıkları bakımından %95 güven düzeyinde istatistiksel olarak önemli bir fark bulunmaktadır (p=0,02<0,05). Hangi düzeyler arasında farklılık olduğu incelenmiş olup sonuçlar aşağıdaki gibidir:

- Üst üçlük ile orta üçlük (p=0,119>0,05) arasında ve Üst üçlük ile Alt üçlük (p=0,682) arasında yaşam olasılıkları bakımından istatistiksel olarak önemli bir fark bulunamamıştır. Ancak Üst üçlük ile tüm mide (p=0,006<0,05) arasındaki fark istatistiksel olarak anlamlıdır.
- Orta üçlük ile alt üçlük (p=0,129) ve orta üçlük ile tüm mide(p=0,259) arasında yaşam olasılıkları bakımından istatistiksel olarak anlamlı bir bulunmamaktadır.

- Alt üçlük ile tüm mide (p=0,009) arasında yaşam süreleri bakımından istatistiksel olarak anlamlı bir fark bulunmaktadır.

Tablo 2’de verilen log-rank testi sonucuna göre hastalığın evreleri arasında yaşam olasılıkları açısından farkın önemli olduğu sonucuna varılmıştır (p=0,00<0,05). Hangi düzeyler arasında farklılık olduğu incelenmiş olup sonuçlar aşağıdaki gibi gözlenmiştir:

- Evre1+Evre2, Evre 3’ten (p=0,00) ve Evre 4’ten (p=0,00) farklı bulunmuştur.
- Evre3 ile Evre4 arasında fark bulunmamaktadır.(p=0,459).
- Evre1+Evre2’deki hastaların ortalama yaşam süresinin Evre3’deki hastalardan ve özellikle Evre 4’deki hastalardan daha fazla olduğu görülmüştür. Evre1+Evre2’deki hastaların 5 yıllık yaşam olasılığı %85,3 iken, Evre3’ teki hastaların 5 yıllık yaşam olasılığı %28,4 ve Evre4’ teki hastalarınki ise %17’ dir.

Tablo 1’de verilen değişkenlerin yaşam süresine etkisi araştırılmak istenmiştir. Cox regresyon modelinin kullanılabilmesi için değişkenlerin orantılı tehlikeler varsayımını sağlaması gerekmektedir. Bu çalışmada orantılı tehlikeler varsayımının sağlanıp sağlanmadığı iki yöntem kullanılarak araştırılmıştır.

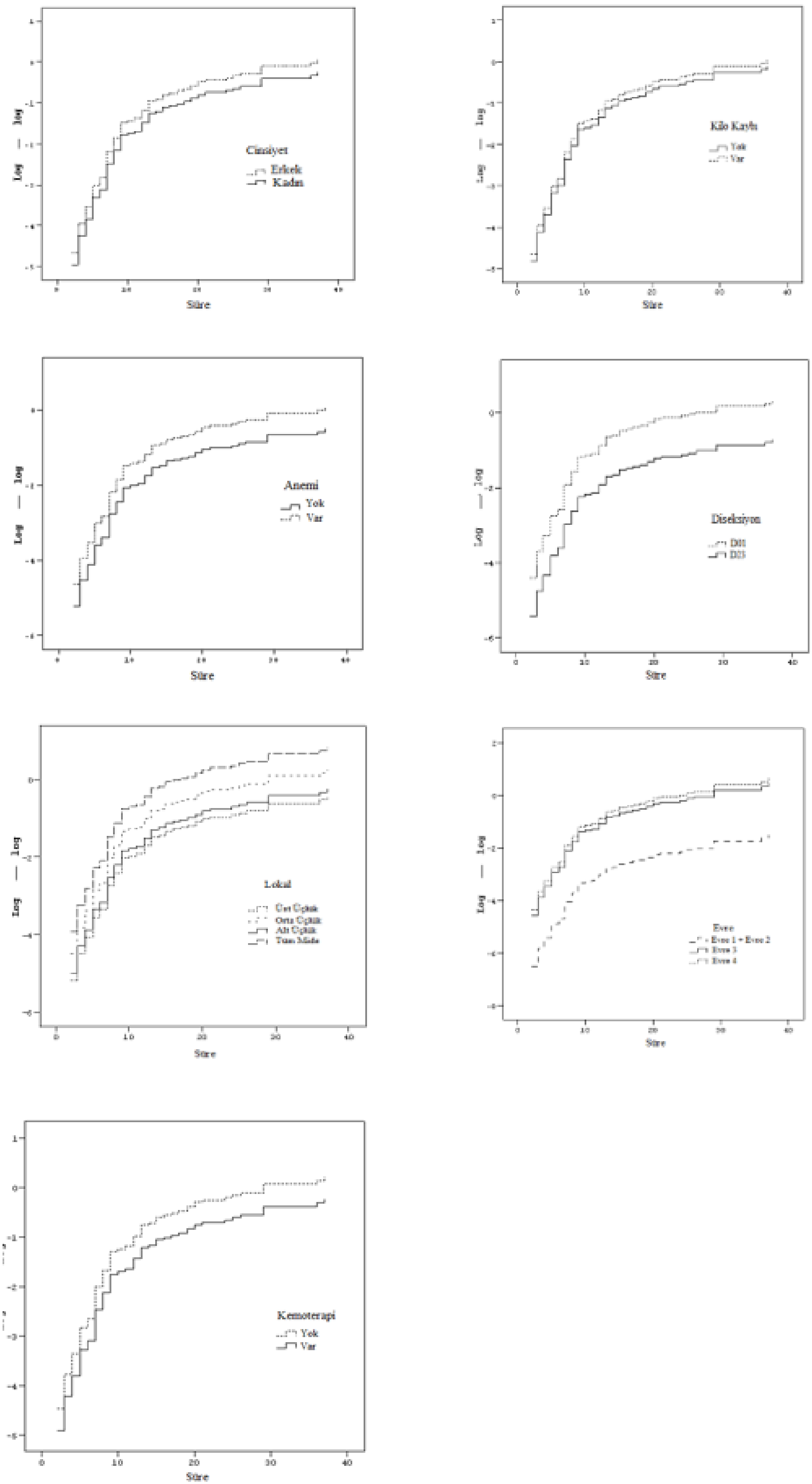
Orantılı tehlikeler varsayımını incelenmek için kullanılan yöntemlerden biri, yaşam süresi rankının, Schoenfeld artıkları ile ilişkisini incelemektir. Bu yöntem sonucunda tüm açıklayıcı değişkenler için edilen sonuçlar Tablo 3’ te verilmiştir.

Değişkenler	**p**
Cinsiyet2	0,63
Kilo Kaybı2	0,23
Anemi 2	0,40
Diseksiyon2	0,09
Lokal2	0,40
Lokal3	0,68
Lokal4	0,99
Evre3	0,09
Evre4	0,28
Kemoterapi2	0,19

Tablo 3. Yaşam süresi rankı ile Schoenfeld artıkları ilişkisinin incelenmesi

Tablo 3 incelendiğinde, tüm değişken düzeylerinde p > 0,05 olarak bulunduğu için orantılı tehlikeler varsayımının sağlandığı görülmüştür (Schoenfeld, 1982).

Orantılı tehlikeler varsayımının sağlanıp sağlanmadığını incelemede kullanılan grafiksel yöntemlerden log(-log) yaşam eğrileri yöntemi kullanılarak incelendiğinde ise elde edilen grafikler Şekil 5’de verilmiştir. Değişkenin düzeylerine ait eğriler paralel ise ve kesişme yoksa o değişken için varsayımın sağlandığı söylenebilir. Şekillerden tüm değişkenler için orantılı tehlikeler varsayımının sağlandığı görülmektedir.



Şekil 5. Açıklayıcı değişkenlerin log(-log) yaşam eğrileri

Orantılı tehlikeler varsayımı sağlandığından Cox regresyon modelinin uygulanabilir olduğu sonucuna varılmıştır. Ancak yaşam süreleri incelendiğinde 118 hastaya ait 41 farklı yaşam süresinin 27'sinde eş zamanlılık olduğu dikkat çekmiştir.

Eş zamanlı gözlemler olduğunda kullanılan Cox regresyon için Breslow, Efron ve Kesin yöntemler uygulanmış ve elde edilen sonuçlar sırasıyla Tablo 4 – Tablo 6'da verilmiştir.

Değişken	$\widehat{\beta}$	**Standart Hata ($\widehat{\beta}$)**	**p**	$e^{\widehat{\beta}}$	**$e^{\widehat{\beta}}$ için Güven Aralığı**
Cinsiyet (Kadın)	-0,30	0,33	0,36	0,74	0,39 ; 1,41
Kilo kaybı (Var)	-0,39	0,34	0,25	0,67	0,34; 1,32
Anemi (Var)	-0,11	0,38	0,77	0,90	0,43 ; 1,88
Diseksiyon (D2-3)	-1,02	0,34	0,00*	0,36	0,18 ;0,71
Lokal (Orta Üçlük)	1,28	0,51	0,01*	3,59	1,32 ; 9,73
Lokal (Alt Üçlük)	0,78	0,46	0,09	2,19	0,88 ; 5,43
Lokal (Tüm Mide)	1,39	0,58	0,02*	4,02	1,28 ; 12,60
Evre (Evre 3)	1,80	0,55	0,00*	6,05	2,04 ; 17,92
Evre (Evre 4)	2,80	0,63	0,00*	16,42	4,74 ; 56,86
Kemoterapi (Var)	-1,06	0,34	0,00*	0,35	0,18 ; 0,68
-2log(L)	388,4				

*p < 0.05 olduğundan anlamlıdır.

Tablo 4. Breslow yöntemi için Cox regresyon modeli sonuçları

Değişken	$\widehat{\beta}$	**Standart Hata ($\widehat{\beta}$)**	**p**	$e^{\widehat{\beta}}$	**$e^{\widehat{\beta}}$ için Güven Aralığı**
Cinsiyet (Kadın)	-0,32	0,33	0,33	0,73	0,38 ; 1,39
Kilo kaybı (Var)	-0,41	0,35	0,24	0,67	0,34 ; 1,31
Anemi (Var)	-0,11	0,38	0,80	0,89	0,43 ; 1,88
Diseksiyon (D2-3)	-1,03	0,34	0,00*	0,36	0,18 ; 0,70
Lokal (Orta Üçlük)	1,31	0,51	0,01*	3,72	1,37 ; 10,09
Lokal (Alt Üçlük)	0,82	0,46	0,08	2,28	0,92 ; 5,66
Lokal (Tüm Mide)	1,43	0,58	0,01*	4,19	1,33 ; 13,13
Evre (Evre 3)	1,83	0,56	0,00*	6,28	2,12 ; 18,66
Evre (Evre 4)	2,85	0,64	0,00*	17,35	4,99 ; 60,32
Kemoterapi (Var)	-1,10	0,35	0,00*	0,33	0,17 ; 0,66
-2log(L)	385,06388				

*p < 0.05 olduğundan anlamlıdır.

Tablo 5. Efron yöntemi için Cox regresyon modeli sonuçları

Değişken	$\hat{\beta}$	Standart Hata ($\hat{\beta}$)	p	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ için Güven Aralığı
Cinsiyet (Kadın)	-0,32	0,33	0,33	0,72	0,38 ; 1,39
Kilo kaybı (Var)	-0,41	0,35	0,24	0,66	0,34 ; 1,31
Anemi (Var)	-0,11	0,38	0,77	0,90	0,43 ; 1,89
Diseksiyon (D23)	-1,04	0,34	0,00*	0,35	0,18 ; 0,70
Lokal (Orta Üçlük)	1,32	0,51	0,01*	3,74	1,38 ; 10,18
Lokal (Alt Üçlük)	0,83	0,47	0,07	2,30	0,92 ; 5,72
Lokal (Tüm Mide)	1,44	0,58	0,01*	4,22	1,34 ; 13,27
Evre (Evre 3)	1,84	0,56	0,00*	6,30	2,12 ; 18,74
Evre (Evre 4)	2,86	0,64	0,00*	17,45	5,01 ; 60,79
Kemoterapi (Var)	-1,10	0,35	0,00*	0,33	0,17 ; 0,66
-2log(L)	328,63886				

*p < 0.05 olduğundan anlamlıdır.

Tablo 6. Kesin Marjinal yöntemi için Cox regresyon modeli sonuçları

Yöntemleri karşılaştırmak için Akaike Bilgi Kriteri (AIC) ve Bayesci Bilgi Kriteri (BIC) kullanılmış ve elde edilen sonuçlar Tablo 7'de verilmiştir:

Model	-2log(L)	AIC	BIC
Breslow Yöntemi	388,40	418,4	409,12
Efron Yöntemi	385,06	415,06	405,78
Kesin Marjinal Yöntem	328,63	358,63	349,35

Tablo 7. Yöntemlerin karşılaştırılması

Tablo 7 incelendiğinde AIC ve BIC değerlerine göre Kesin marjinal yöntem ile Cox regresyon modelinin en uygun model olduğu görülmektedir. Bu yöntem uygulandığında elde edilen sonuçlar Tablo 6'da verilmiştirTablo 6 incelendiğinde diseksiyon, lokal, evre ve kemoterapi değişkenlerinin mide kanserinden ölümleri etkileyen önemli faktörler olduğu %95 güvenle söylenebilmektedir ve aşağıdaki yorumlar yapılabilmektedir:

Lenf nodu diseksiyonunun genişliği D0-1 olan hastaların ölüm riski, D2-3 olanlara göre 1/(0,35)=2,86 kat daha fazladır.

Tümörün midedeki lokalizasyonunu orta üçlükte olan hastaların üst üçlükte olan hastalara göre ölüm riski 3.74 kat daha fazladır.

Tümörün midedeki lokalizasyonunu tüm midede olan hastaların üst üçlükte olan hastalara göre ölüm riski 4.22 kat daha fazladır.

Hastalığı 3. evrede olan hastaların evre 1 + evre 2'deki hastalara göre ölüm riski yaklaşık olarak 6,30 kat daha fazladır.

Hastalığı 4. evrede olan hastaların ise evre 1 + evre 2'deki hastalara göre ölüm riski yaklaşık olarak 17,45 kat daha fazladır.

Kemoterapi almayan hastalarda ölüm riski kemoterapi alan hastalara göre yaklaşık olarak 1/(0,33)=3,03 kat daha fazladır.

Veri setine uygulanan Kesin Marjinal, Efron ve Breslow yöntemleri incelendiğinde regresyon katsayıları ve tehlike oranları açısından Breslow yönteminin diğer iki yönteme göre genel olarak daha küçük değerlere sahip olduğu, Kesin Marjinal yöntemin ise en yüksek değerlere sahip olduğu gözlenmiştir.

4.2. Eş Zamanlı Gözlemlerin Olmadığı Durum

Akut miyokard infarktüsü nedeniyle hastaneye yatan hastaların yaşam sürelerini etkileyen faktörleri belirlemek için 1975 ile 2001 yılları arasında Worcester civarındaki hastanelere yatan 100 hastanın gün bazında süreleri kaydedilmiştir. Hastalık neticesinde gerçekleşen ölümler başarısızlık olarak ifade edilmiştir. Ölümle sonuçlanmayanlar ise durdurulmuş olarak tanımlanmıştır.

Açıklayıcı değişkenler olarak cinsiyet (Erkek, Kadın), yaş (min:32 ; max:92), hastanede kalış süresi (min:1 ; max:56), vücut kitle indeksi (min: 14,92; max: 39,94) alınmıştır. Bu çalışmada 100 hastanın 65'i erkek, 35i kadındır. Yaş ortalaması 68,25, hastanede kalış süresi ortalaması 6,84 ve vücut kitle indeksi ortalaması 27,04'dür. Gözlemlerin 49'u durdurulmuş 51'i ise başarısız olmuştur.

Dört zaman noktası dışında kalan tüm yaşam süreleri birbirinden farklıdır. Yaşam süreleri gün bazında daha hassas olarak kaydedildiğinden veri setinde eş zamanlı gözlem oluşumu yok denecek kadar azdır.

Bu çalışmada açıklayıcı değişkenlerin yaşam süresine etkileri araştırılmak istenmiştir. Cox regresyon modelinin kullanılabilmesi için değişkenlerin orantılı tehlikeler varsayımını sağlaması gerekmektedir. Tüm değişkenler için orantılı tehlikeler varsayımı sağlanmaktadır. Dolayısıyla yaşam süreleri için Breslow, Efron ve Kesin Marjinal yöntemler için Cox regresyon modeli uygulanabilmektedir. Elde edilen sonuçlar sırasıyla Tablo 8- Tablo 10'da verilmiştir.

Değişken	$\hat{\beta}$	**Standart Hata ($\hat{\beta}$)**	**p**	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ **için Güven Aralığı**
Cinsiyet2	0,16	0,31	0,61	1,17	0,64 – 2,13
Yaş	0,04	0,01	0,00*	1,04	1,01 – 1,07
Hastanede Kalış Süresi	-0,02	0,03	0,40	0,98	0,93 – 1,03
Vücut Kitle Endeksi	-0,07	0,04	0,05*	0,93	0,87 – 1,00
-2log(L)	395,78				

Tablo 8. Breslow yöntemi için Cox regresyon modeli sonuçları

Değişken	$\hat{\beta}$	**Standart Hata ($\hat{\beta}$)**	**p**	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ **için Güven Aralığı**
Cinsiyet	0,16	0,31	0,61	1,17	0,64 – 2,13
Yaş	0,04	0,01	0,00*	1,04	1,01 – 1,07
Hastanede Kalış Süresi	-0,02	0,03	0,40	0,98	0,93 – 1,03
Vücut Kitle Endeksi	-0,07	0,04	0,05*	0,93	0,87 – 1,00
-2log(L)	395,73				

Tablo 9. Efron yöntemi için Cox regresyon modeli sonuçları

Değişken	$\hat{\beta}$	**Standart Hata ($\hat{\beta}$)**	**p**	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ **için Güven Aralığı**
Cinsiyet	0,16	0,31	0,61	1,17	0,64 – 2,13
Yaş	0,04	0,01	0,00*	1,04	1,01 – 1,07
Hastanede Kalış Süresi	-0,02	0,03	0,40	0,98	0,93 – 1,03
Vücut Kitle Endeksi	-0,07	0,04	0,05*	0,93	0,87 – 1,00
-2log(L)	392,95				

Tablo 10. Kesin Marjinal yöntemi için Cox regresyon modeli sonuçları

Tablo 8 – Tablo 10 incelendiğinde veri setinde eş zamanlı gözlemlerin olmaması durumunda, uygulanan bu yöntemlerin sonuçları ($\hat{\beta}$ katsayıları, katsayılara ait standart hatalar, p değerleri, tehlike oranları ve tehlike oranlarının güven aralıkları) birebir aynı bulunmuştur. -2Log(L) değerleri arasında küçük farklılıklar bulunmasının nedeninin ise dört zaman noktasında eş zamanlı gözlemlerin olmasından kaynaklandığı düşünülmektedir.

5. Sonuçlar

Bu çalışmada yaşam çözümlemesi ve yaşam çözümlemesinde eş zamanlı yaşam süreleri hakkında genel bilgiler, fonksiyonlar ve modeller verilmiş, yaşam çözümlemesinde eş zamanlı yaşam süreleri olduğu durumda kullanılacak yöntemlere değinilmiştir.

Yaşam çözümlemesinde sıkça kullanılan bir yöntem olan Cox regresyon modeli, yaşam sürelerinin sürekli ölçümlü olduğu ve bu nedenle eş zamanlı yaşam sürelerinin ortaya çıkmayacağı varsayımına sahiptir. Ancak çalışmalarda ölçüm kısıtları, belirsiz ölçümler ve belli aralıklarla kaydedilebilen ölçümler sonucunda eş zamanlı yaşam sürelerine sıklıkla rastlanmaktadır. Bu durumda ölçümler daha hassas şekilde olmak üzere yeniden kaydedilmeli, kaydedilemiyorsa Cox regresyon modeli, eş zamanlı yaşam sürelerinin gözlendiği duruma göre yeniden düzenlenmelidir. Breslow yöntemi, Efron yöntemi, Kesin yöntem, Kesikli yöntem ve EM algoritmasına dayalı yöntem, yaşam verilerinde eş zamanlı gözlemler olduğu durumda kullanılabilecek Cox regresyon yöntemleridir.

Uygulamada öncelikle eş zamanlı yaşam sürelerinin yer aldığı yedi açıklayıcı değişkene ve 118 gözleme sahip mide kanseri verisine uygun yöntemler uygulanmıştır. Orantılı tehlikeler varsayımının sağlandığı gözlenmiştir. Daha sonra sürekli ölçümlü yaşam sürelerine sahip olan veri setine uygun olan Breslow, Efron ve Kesin yöntem ile Cox regresyon sonuçları elde edilmiştir. Bilgi kriterleri yardımı ile en uygun yöntemin Kesin marjinal yöntem olduğuna karar verilmiş ve diseksiyon, lokal, evre ve kemoterapi değişkenlerinin mide kanserini etkileyen faktörler olduğu sonucuna varılmıştır. Daha sonra eş zamanlı yaşam sürelerinin olmadığı veri setine örnek olarak AMI verilerine, orantılı tehlikeler varsayımının sağlanmasının test edilmesinin ardından aynı yöntemler uygulanmış ve sonuçlarının aynı olduğu gözlenmiştir.

Sonuç olarak yaşam çözümlemesi çalışmalarında, çeşitli bilgisayar programlarında hazır olarak karşımıza çıkan Breslow yöntemini kullanmak doğru değildir. Veri setinde eş zamanlı yaşam süreleri meydana gelmiş ise diğer yöntemler ile sonuçlar elde edilmeli ve en uygun bulunan yönteme göre yorumlanmalıdır. Eğer veri setinde eş zamanlı yaşam süreleri yoksa hangi yöntem kullanılırsa kullanılsın sonuçlar değişmemektedir.

References

Allison, P.D. (2010). Survival Analysis, The Reviewer's Guide to Quantitative Methods in the Social Sciences, New York, 413-425.

Arı, A., Önder, H. (2013). Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri, Anadolu Tarım Bilim. Derg. , 28(3):168-174.

Ata, N. (2005). Yaşam Çözümlemesinde Orantısız Hazard Modeli, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Breslow, N. (1974). Covariance Analysis of Censored Survival Data, Biometrics, 30, 89-99.

Chalita, L.V. A.S., Colosimo, E.A.,Demetrio C.G.B. (2002). Reliablility and Survival Analysis, Likelihood Approximation and Discrete Models for Tied Survival Data, Commun.Statist.-Theory Meth., 31(7),1215-1229.

Collett, D. (1994). Modelling Survival Data in Medical Research, Chapman &Hall.

Cox, D.R. (1972). Regression models and life tables, Journal of Royal Statistical Society, Series B, 34, 187-202.

DeLong, D. M., Guirguis, G. H., and So, Y. C. (1994). Efficient Computation of Subset Selection Probabilities with Application to Cox Regression, Biometrika, 81, 607–611.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Survival Data, Journal of the American Statistical Association, 72,557-565.

Eroğlu A., Altınok M., Özgen K., Sertkaya D. (1997). A Multivariate Analysis of Clinical and Pathological Variables in Survival After Resection of Gastric Cancer, Türkiye Klinikleri Medical Research, 15, 1, 15-20.

Johnson, R.E., Johnson, N. (1980). Survival models and data analysis, Wiley&Sons, New York.

Kalbfleisch, J.D., Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model, Biometrica, 60, 267-278.

Kalbfleisch, J.D., Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data, Second Edition, Wiley&Sons, New York.

Kleinbaum, D.G. (1996). Survival Analysis: A Self-Learning Text, First Edition, Springer.

Karasoy, D., Tuncer, N., (2015), Outliers in Survival Analysis, Alphanumeric Journal, 3, 2, 139-152.

Lee, E.T., Wang, J.W. (2003). Statistical Methods for Survival Data Analysis, Third Edition, Wiley&Sons.

Picciotto, I.H., Rockhill, B. (1997). Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression.Department of Epidemiology, Biometrics, 53, 1151-1156.

Sertkaya, D., Ata, N., Sözer, M.T. (2005). Yaşam çözümlemesinde zamana bağlı açıklayıcı değişkenli Cox regresyon modeli, Ankara Üniversitesi Tıp Fakültesi Mecmuası, 58:153-158.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model, Biometrika, 69, 239-241.

Scheike, T.H., Sun, Y. (2007). Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm, Lifetime Data Anal., 13:399–420.

Therneau, T. M., Grambsch, P. M. (2000). Modeling Survival Data: Extending Cox Model, Springer, New York.

Xin, X. (2011). A Study of Ties and Time-Varying Covariates in Cox Proportional Hazard Model, Master of Science, The Faculty of Graduate Studies of The University of Guelph.

Yay,M., Çoker, E., Uysal, Ö. (2007). Yaşam Analizinde Cox Regresyon Modeli ve Artıkların incelenmesi, Cerrahpaşa Tıp Dergisi, 38: 139 – 145.

Zhang, M.J. (1997). Grouped Failure Times Tied Failure Times-Two Contributions To The Encyclopedia of Biostatistics, Technical Report 24.

http://www.medicine.mcgill.ca/epidemiology/hanley/c681/cox/TiesCoxModelR.txt (Aralık, 2014).

Cox Proportional Hazards Regression Models, http://www4.stat.ncsu.edu/~dzhang2/st745/chap7.pdf (Aralık, 2014).

Stefanescu, C., Mehrotra,D., Cox model versus generalized logrank test for time to event data with ties, http://faculty.london.edu/cstefanescu/glr.pdf (Ekim, 2014).

Semi-Parametric Duration Models: The Cox Model https://files.nyu.edu/mrg217/public/cox.pdf (Aralık, 2014).