

**YAŞAM ÇÖZÜMLEMESİNDE EŞ ZAMANLI YAŞAM
SÜRELERİ**

TIED SURVIVAL TIMES IN SURVIVAL ANALYSIS

SENA KESKİN

PROF. DR. DURDU KARASOY

Tez Danışmanı

Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2015

SENA KESKİN' in hazırladığı “Yaşam Çözümlemesinde Eş Zamanlı Yaşam Süreleri” adlı bu çalışma aşağıdaki jüri tarafından İSTATİSTİK ANABİLİM DALI' nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Prof. Dr. Yasemin YAVUZ

Başkan

Prof. Dr. Durdu KARASOY

Danışman

Doç. Dr. Nihal ATA TUTKUN

Üye

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından YÜKSEK LİSANS TEZİ olarak onaylanmıştır.

Prof. Dr. Fatma SEVİN DÜZ
Fen Bilimleri Enstitüsü Müdürü

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

07/01/2015


SENA KESKİN

ÖZET

YAŞAM ÇÖZÜMLEMESİNDE EŞ ZAMANLI YAŞAM SÜRELERİ

SENA KESKİN

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Durdu KARASOY

Ocak 2015, 41 sayfa

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 (tied) durumunun meydana gelmemiş olmasıdır. Ancak çalışmalarda genellikle eş zamanlı olarak meydana gelen başarısızlıklara rastlanmaktadır ve bu durum özel çözüm gerektirmektedir. Kesin (Exact) yöntem, Breslow yöntemi, Efron yöntemi, Kesikli (discrete) yöntem ve EM algoritmasına dayalı yöntem olarak bilinen yöntemler bu özel çözümlerdir.

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 eş zamanlı gözlemlerin olmadığı Akut miyokard infarktüsü (AMI) verilerine uygulanmış ve sonuçların aynı olduğu gözlenmiştir.

Anahtar Kelimeler: Yaşam çözümlemesi, Cox regresyon, eş zamanlı yaşam süreleri.

ABSTRACT

TIED SURVIVAL TIMES IN SURVIVAL ANALYSIS

SENA KESKİN

Master of Science, Department of Statistics

Supervisor: Prof. Dr. Durdu KARASOY

January 2015, 41 pages

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 Efron method, Exact method, Breslow method, Discrete method and EM algorithm for handling tied event times.

Firstly, the methods are analysed during the study, Breslow approach, Efron Approach and Exact Method, which is applied on a stomach cancer data set (with tied observations) It is observed that Cox regression with Exact Method is the best model.

Then these methods are applied to Acute Myocardial Infarction data set which has no tied data between event times and it is found that the same results are obtained for all methods.

Key Words: Survival analysis, Cox regression, tied survival times.

TEŐEKKÜR

Tezimin oluŐmasında bana en büyük desteęi veren, alıŐmalarımın her aŐamasında yol gÖsteren, sonsuz sabırla ve yakın ilgisiyle beni her zaman teŐvik eden ve güven veren danıŐmanım Sayın Prof. Dr. Durdu KARASOY'a; hayatımın her aŐamasında maddi manevi desteklerini esirgemedikleri, cesaret verdikleri, bana güvendikleri ve hep yanımda oldukları için sevgili babam Kenan KESKİN, annem Yasemin KESKİN, kardeŐim Ömer Halit KESKİN'e ve tüm destekleri için sevgili arkadaşlarıma içtenlikle teŐekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER.....	iv
ÇİZELGELER.....	vi
ŞEKİLLER.....	vii
SİMGELER VE KISALTMALAR.....	viii
1. GİRİŞ.....	1
2. YAŞAM ÇÖZÜMLEMESİ.....	3
2.1. Cox Regresyon Modeli.....	5
2.1.1. Orantılı Tehlikeler Varsayımı.....	6
2.2. Parametrik Dağılımlar.....	8
2.2.1. Üstel Dağılım.....	9
2.2.2. Weibull Dağılımı.....	10
2.2.3. Log-Normal Dağılım.....	10
2.2.4. Log-Lojistik Dağılım.....	11
2.2.5. Gamma ve Genelleştirilmiş Gamma Dağılımı.....	12
3. EŞ ZAMANLI YAŞAM SÜRELERİ.....	14
3.1. Eş zamanlı Yaşam Süreleri Tahmin Edici Süreci.....	15
3.1.1. Sürekli Zaman Ölçümü İle Belirlenen Eş Zamanlı Gözlemler.....	16
3.1.2. Kesikli Zaman Ölçümü İle Belirlenen Eş Zamanlı Gözlemler.....	16
3.2. Eş Zamanlı Yaşam Süreleri Oluştığı Durumda Kullanılan Yöntemler.....	17
3.2.1. Kesin Yöntem.....	18
3.2.2. Breslow Yöntemi.....	19
3.2.3 Efron Yöntemi.....	20
3.2.4. Kesikli Yöntem.....	20
3.2.5. EM Algoritmasına Dayalı Yöntem.....	21
3.2.6. Yöntemlerin Karşılaştırılması.....	22
4. UYGULAMA.....	24
4.1. Eş Zamanlı Gözlemlerin Olduğu Durum.....	24

4.2. Eş Zamanlı Gözlemlerin Olmadığı Durum.....	35
5. SONUÇLAR.....	38
KAYNAKLAR.....	39
ÖZGEÇMİŞ.....	41

ÇİZELGELER

	<u>Sayfa</u>
Çizelge 3.1. Örnek veri.....	17
Çizelge 4.1. Açıklayıcı değişkenler.....	25
Çizelge 4.2. Kaplan-Meier sonuçları.....	26
Çizelge 4.3. Yaşam süresi rankı ile Schoenfeld artıkları ilişkisinin incelenmesi.....	30
Çizelge 4.4. 118 hastaya ilişkin yaşam süreleri.....	32
Çizelge 4.5. Breslow yöntemi için Cox regresyon modeli sonuçları.....	33
Çizelge 4.6. Efron yöntemi için Cox regresyon modeli sonuçları.....	33
Çizelge 4.7. Kesin Marjinal yöntemi için Cox regresyon modeli sonuçları.....	34
Çizelge 4.8. Yöntemlerin karşılaştırılması.....	34
Çizelge 4.9. 100 hastaya ilişkin yaşam süreleri.....	36
Çizelge 4.10. Breslow yöntemi için Cox regresyon modeli sonuçları.....	37
Çizelge 4.11. Efron yöntemi için Cox regresyon modeli sonuçları.....	37
Çizelge 4.12. Kesin marjinal yöntemi için Cox regresyon modeli sonuçları.....	37

ŞEKİLLER

Sayfa

Şekil 4.1. Genel Kaplan-Meier eğrisi.....	26
Şekil 4.2. Cinsiyet değişkeni için Kaplan-Meier eğrisi.....	27
Şekil 4.3. Kilo kaybı değişkeni için Kaplan-Meier eğrisi.....	27
Şekil 4.4. Anemi değişkeni için Kaplan-Meier eğrisi.....	27
Şekil 4.5. Diseksiyon değişkeni için Kaplan-Meier eğrisi.....	27
Şekil 4.6. Lokal değişkeni için Kaplan-Meier eğrisi.....	28
Şekil 4.7. Evre değişkeni için Kaplan-Meier eğrisi.....	28
Şekil 4.8. Kemoterapi değişkeni için Kaplan-Meier eğrisi.....	28
Şekil 4.9. Açıklayıcı değişkenlerin log(-log) yaşam eğrileri.....	31

SİMGELER VE KISALTMALAR

Simgeler

μ	Ortalama
σ^2	Varyans
Γ	Gamma Fonksiyonu

Kısaltmalar

AIC	Akaike Bigi Kriteri
BIC	Bayesci Bilgi Kriteri
AMI	Akut Miyokard İnfarktüsü

1. GİRİŞ

İnsan ve çevresindeki birimlerin başarısızlık (failure) ya da ölüm zamanının istatistiksel çözümlemesine ilişkin çalışmalar yaşam tablosu yardımıyla başlamıştır. Bu çalışmalar daha sonra geliştirilerek başarısızlık modeli ya da tehlike modeli olarak adlandırılmıştır. Bu modeller, iyi tanımlanmış herhangi bir olayın olma ya da gözleme zamanının çözümlemesinde kullanılmıştır [1].

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 birey ya da birime ait yaşam süresi T , tanımı gereği sürekli ve pozitif bir değere sahiptir [2].

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 [2, 3].

Cox'un kısmi olabilirlik fonksiyonu gözlemler arasında eş zaman durumunun olmadığını varsaymaktadır [4, 5]. 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 [6].

Veri kümelerinde eş zamanlı olarak meydana gelen yaşam süreleri sıklıkla gözlenmektedir. Bu durumda Cox'un kısmi olabilirlik fonksiyonu eş zamanlı gözlemler olduğu durum için yeniden düzenlenmelidir [5]. Bu yöntemler, Breslow (1974) tarafından önerilen Breslow yöntemi, Efron (1977) tarafından önerilen Efron yöntemi, Kesin yöntem [9, 10, 25], Kesikli (discrete) yöntem [11] ve Scheike ve Sun (2007) tarafından önerilen EM algoritmasına dayalı yöntemdir.

Çalışmanın ikinci bölümünde yaşam çözümlemesi hakkında genel bilgiler verilmiştir. Yaşam çözümlemesinde kullanılan fonksiyonlara ve birbirleriyle ilişkilerine, yaşam çözümlemesinde yaygın olarak kullanılan yarı parametrik bir yöntem olan Cox regresyon modeline ve bu modelin temel varsayımı olan orantılı tehlikeler varsayımına değinilmiş ayrıca yaşam çözümlemesinde kullanılan parametrik dağılımlar anlatılmıştır.

Çalışmanın üçüncü bölümünde, eş zamanlı yaşam süreleri hakkında ve yaşam çözümlemesinde eş zamanlı yaşam süreleri durumunda kullanılan yöntemler hakkında bilgiler yer almaktadır.

Dördüncü bölümde ise çalışma boyunca anlatılan yöntemlerden veri kümesine 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 AMI verilerine bu yöntemler uygulanmış, sonuçlar değerlendirilmiştir.

Bu çalışmanın amacı, eş zamanlı gözlem olduğunda kısmi olabilirliğin hesaplanması için kullanılan yöntemleri incelemektir. Bu amaçla eş zamanlı gözlemler durumunda kullanılan yöntemler araştırılmış, literatürde yer alan yöntemler incelenmiş ve örnek verilere uygulanmıştır.

2. YAŞAM ÇÖZÜMLEMESİ

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 [13].

Yaşam verilerinin çözümlenmesinde karşılaşılan temel güçlük, gözlem altına alınan bazı birimlerin başarısızlık zamanlarının gözlenememiş olmasıdır [1]. 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 (tamamlanmamış) veri* ortaya çıkar. Durdurulmuş veriler gözlemlerin tamamının değil bir kısmının bilindiği veriler olarak da açıklanabilir. Durdurma soldan, sağdan ve aralıklı olmak üzere üç çeşittir. En yaygın kullanılanı sağdan durdurmadır (right censoring). Örneğin bir birimde çalışma süresi boyunca ilgilenilen olay meydana gelmemişse bu sağdan durdurmadır [14].

Yaşam çözümlemesinde bağımlı değişken, olayın gerçekleşmesine (durdurulmuş da olabilir) kadar geçen süre olup, “*başarısızlık süresi*” veya “*yaşam süresi*” olarak adlandırılır. Ölüme kadar geçen süre, bilgisayar bozuluncaya kadar geçen süre vb. ile örneklendirilebilir. Yaşam süresi verileri belirli olayların zamanı ile ilgilidir. Bu zamanlar rasgeledir ve bir dağılım oluşturur. Genellikle T ile gösterilir.

Yaşam sürelerinin dağılımları genel olarak üç fonksiyonla gösterilir. Bunlar; yaşam fonksiyonu, olasılık yoğunluk fonksiyonu ve tehlike fonksiyonudur. Bu fonksiyonlar matematiksel olarak birbirinden elde edilebilir [6].

Olasılık Yoğunluk Fonksiyonu

Birimin gerçek yaşam süresi t, T değişkeninin bir değeri olarak gösterilebilir ve bu değer negatif olamaz. T rastlantı değişkeni f(t) olasılık dağılımına sahiptir. f(t) olasılık yoğunluk fonksiyonu

$$f(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t)}{\delta t}$$

şeklinde dir. T' nin dağılım fonksiyonu,

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

biçimindedir ve yaşam süresinin t değerinden küçük olması olasılığını gösterir [4].

Yaşam Fonksiyonu

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

$$S(t) = P(T \geq t) = 1 - F(t)$$

biçimindedir [4].

Yaşam fonksiyonu monoton azalan soldan sürekli bir fonksiyondur ve

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

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

olur [14].

Tehlike Fonksiyonu

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

Tehlike fonksiyonu,

$$\lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta_t / T \geq t)}{\delta_t}$$

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

Bu tanımdan yaşam ve tehlike fonksiyonları arasında bazı ilişkiler görülebilir.

$F(t)$, T'nin dağılım fonksiyonu olmak üzere,

$$h(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{F(t+\delta t) - F(t)}{\delta_t} \right\} \frac{1}{S(t)}$$

olur ve buradan

$$h(t) = \frac{f(t)}{S(t)}$$

elde edilir.

Tehlike fonksiyonu,

$$h(t) = - \frac{d}{dt} \{ \log S(t) \}$$

biçiminde de yazılabilir [4].

Başlangıç zamanı her bir birim için şüpheye yer vermeyecek şekilde tanımlanmalıdır geçen süreyi ölçmek için bir zaman ölçeği kabul edilmelidir, her bir birim için başarısızlığın meydana geldiği an net olmalıdır. Diğer bir ifadeyle, başarısızlığın tam olarak tanımlanması gerekmektedir [1]. Her bir birim için başlangıç noktasının aynı olması gerekmez. Birimler farklı zamanlarda çalışmaya katılabilir. Önemli olan izlenmeye başlanıldığı anın kaydedilmesidir.

Olağan dışı bir durumla karşılaşmadıkça bağımlı değişkenin gözlenememesi söz konusu olmaz. Bununla birlikte yaşam çözümlemesi, bazı gözlemler durdurulmuş olsa da uygulanabilmektedir. Bu da yaşam analizinin güçlü yanlarından biri ve klasik yöntemlerden farkı olarak kabul edilir [13].

Yaşam çözümlemesinde sıkça kullanılan yöntemlerden biri, yaşam süresi üzerinde etkili olan faktörlerin belirlenmesinde kullanılan Cox Regresyon Modelidir [3].

2.1 Cox Regresyon Modeli

İlk olarak 1972 yılında Cox tarafından ele alınan orantılı tehlikeler modeli (proportional hazards model) yaşam çözümlemesinde yaygın olarak kullanılan yarı parametrik bir yöntemdir. Cox regresyon modeli olarak da bilinir. Model ismini kendisini geliştiren David R. Cox 'tan almıştır. Cox regresyon modeli, bir bağımlı değişken (yaşam süresi) ile bir veya birden fazla bağımsız değişken arasındaki ilişkiyi ortaya çıkaran istatistiksel bir yöntemdir [13].

Cox regresyon modeli,

$$h(t, X) = h_0(t) \exp(\beta' X) \quad (2.1)$$

biçimindedir. Burada,

$\beta \Rightarrow$ regresyon katsayıları vektörü,

$h_0(t) \Rightarrow$ açıklayıcı değişkene sahip olmayan ($x=0$ olan) bir birimin temel tehlike fonksiyonudur [1].

Cox regresyon modeli, parametrik modellerin gerektirdiği varsayımların (normallik, bağımsızlık vb.) 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 [15].

2.1.1 Orantılı Tehlikeler Varsayımı

Cox orantılı tehlikeler modelinin temel varsayımı orantılı tehlikelerdir. Orantılı tehlikeler varsayımı, tehlike oranının zamana karşı sabit olması ya da bir birimin tehlikesinin diğer birimin tehlikesine orantılı olması anlamına gelmektedir [16].

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

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

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

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

biçiminde yazılabilir. Bu orantılı tehlikeler varsayımını gösteren matematiksel bir ifadedir. Orantılı tehlike varsayımı,

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

biçiminde de ifade edilebilir. Burada, $\widehat{\theta}$ orantılılık sabiti (proportionality constant) olarak adlandırılır ve zamandan bağımsızdır [1, 4].

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 [1].

En yaygın kullanılan grafiksel yöntem olan log(-log) yaşam eğrileri, açıklayıcı değişkenlerin farklı düzeyleri üzerinden $-\ln(-\ln)$ yaşam eğrilerinin tahmininin karşılaştırılmasını içermektedir. Elde edilen paralel eğriler orantılı tehlikeler varsayımının sağlandığını göstermektedir.

Orantılı tehlikeler varsayımını değerlendirmek için gözlenen yaşam olasılıkları ile beklenen yaşam olasılıklarının çizimlerinin kullanımı uyum iyiliği testi yaklaşımının

grafiksel karşılığıdır. Beklenen yaşam eğrilerini elde etmek için, değerlendirilen tahmin ediciyi içeren Cox orantılı tehlikeler modeli uydurulur. Yaşam eğrisinin tahmini, tahmin edicinin her bir sınıfı için değerleri formülde yerine yerleştirilerek elde edilir. Tahmin edicinin her bir sınıfı için gözlenen ve beklenen yaşam eğrileri birbirine yakınsa, orantılı tehlikeler varsayımının sağlandığına karar verilir. Bir ya da daha çok sınıf için bu eğriler farklılık gösterirse, orantılı tehlikeler varsayımının bozulduğuna karar verilir.

Arjas grafikleri ise orantılı tehlikeler 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. Grafiksel yöntemler içinde en uygun sonucu veren yöntemdir [1].

Zamandan bağımsız değişkenin orantılı tehlike varsayımını sağlayıp sağlamadığını incelemek için zamana bağlı değişkenler de kullanılır. Bu durumda model, zamandan bağımsız değişkenleri ve zamanın fonksiyonlarını kapsayan çarpım terimlerini içeren Cox orantılı tehlikeler modeli olmaktadır. Bu genişletilmiş Cox orantılı tehlikeler modeli x açıklayıcı değişkeni için

$$h(t,x) = h_0(t) \exp[\beta x + \delta(xg(t))] \quad h(t,x) = \alpha\beta^{-1}(\beta^{-1}t)^{\alpha-1}$$

biçimindedir.

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

$$g(t) = t$$

$$g(t) = \log t$$

$$g(x) = \begin{cases} 1, & t \geq t_0 \\ 0, & t < t_0 \end{cases} \quad (\text{Adım fonksiyonu})$$

biçiminde olabilir [14].

Genişletilmiş Cox orantılı tehlikeler modelinde, çarpım terimlerinin önemliliği için test yapılarak orantılı tehlikeler varsayımı değerlendirilir. Yokluk hipotezi, $H_0: \delta = 0$ biçimindedir. Yokluk hipotezi doğru ise, model tek bir değişken içeren orantılı tehlikeler modeline indirgenir [1].

Orantılı tehlikeler varsayımı Schoenfeld artıkları ile de kontrol edilebilir. Schoenfeld artıkları açıklayıcı değişkenin gerçek değeri ile ağırlıklı risk skorlarının ortalaması

arasındaki farktır. Schoenfeld artıkları zamana karşı çizilir ve orantılı tehlikeler varsayımının geçerliliğini kontrol amacıyla kullanılır. Eğer artıklar yatay bir doğru etrafında (sıfır merkezli) tesadüfi olarak yer alıyorsa, orantılı tehlikeler varsayımı sağlanır. Orantılı tehlikeler varsayımı sağlanmazsa grafik bir trend gösterir [18].

Bu test istatistiği, orantılı tehlikeler varsayımının incelenmesi için kullanılan grafiksel yöntemlere göre daha nesnel bir kriter sağlamaktadır. Grafiksel yöntemler ise daha öznel [1].

Bağımlı değişken olan yaşam süresi üzerinde açıklayıcı değişkenlerin etkilerinin araştırıldığı regresyon modelleri yaşam çözümlemesinde önemli bir yere sahiptir.

Cox orantılı tehlikeler modelinde orantılı tehlike varsayımı sağlanmıyorsa yaşam verisinin modellenmesi için birçok yaklaşım önerilmektedir. Bu yaklaşımlar aşağıda verilmiştir:

- Orantısızlığa neden olan değişkenlerle tabakalandırma yapmak (Tabakalandırılmış Cox regresyon modeli)
- Orantısızlığı zamana bağlı açıklayıcı değişkenlerle modellemek (Genişletilmiş Cox regresyon modeli)
- Farklı modeller kullanmak (Ağırlıklandırılmış Cox regresyon modeli, hızlandırılmış başarısızlık zamanı modeli ya da toplamsal tehlike modeli) [16].

2.2. Parametrik Dağılımlar

Yaşam çözümlemesinde, elde edilen veriye hangi dağılımın uygulanacağı çok önemlidir. Bu nedenle model seçimi yapılırken ilk olarak modelin veriye uygunluğu için gerekli test yapılmaktadır. Model veriye uygunsa parametrik test varsayımlarının sağlanıp sağlanmadığına bakılarak modelin geçerliliği ile ilgili gerekli analizler yapılmaktadır.

Yaşam çözümlemesi verilerinde yaygın olarak kullanılan parametrik dağılımlar;

- Üstel dağılım
- Weibull dağılımı
- Log-normal dağılım
- Log-lojistik dağılım
- Gamma dağılımı

olarak ifade edilebilir.

2.2.1 Üstel Dağılım

Üstel dağılım için “Tamamen Rastgele Başarısızlık Modeli” ifadesi de kullanılmaktadır. Birçok yaşam çözümlemesi verisi üstel dağılım tarafından yeterli derecede tanımlanmasa da, bazı genel durumlarda anlamayı kolay hale getirmektedir [19].

Sabit tehlike fonksiyonuna sahip üstel dağılım, yaşam çözümlemesi ve güvenilirlik teorisinde oldukça yaygın uygulama alanı bulmuştur. Böylece üstel dağılımın tehlike fonksiyonu,

$$h(t; \beta) = \beta^{-1}, t \geq 0$$

biçimindedir. Buradan yaşam fonksiyonu,

$$S(t; \beta) = \exp \left[- \int_0^t h(u) du \right] = \exp(-t/\beta)$$

biçiminde yazılabilir. Dağılım fonksiyonu ise,

$$F(t; \beta) = 1 - \exp(-t/\beta), t \geq 0$$

biçimindedir. Böylece β ortalamalı bir üstel dağılımın olasılık yoğunluk fonksiyonu,

$$f(t; \beta) = \beta^{-1} \exp(-t/\beta), t \geq 0$$

biçimindedir.

Bu fonksiyondaki β parametresi ölçek parametresidir ve $\beta > 0$ 'dır. T, β ölçek parametrelili bir üstel dağılımlı rasgele değişken ise $T \sim E(\beta)$ ile gösterilir [4].

2.2.2 Weibull Dağılımı

Weibull dağılımı 1960 ve 1970'li yıllarda çok yaygın olarak kullanılmaya başlanmıştır. Bu model daha esnek bir dağılım sınıfı vermekte ve özel bir durum olarak üstel dağılımı içine almaktadır.

Weibull dağılımı üstel dağılımın genelleştirilmiş halidir. Sabit tehlike oranına sahip olmaması Üstel dağılımdan tek farkıdır. Sabit tehlike oranına sahip olmadığından genelleyici uygulamaları Üstel dağılımdan daha fazladır. Bu dağılım daha çok güvenilirlik çalışmalarında ve hastalık ölümlerinde kullanılmaktadır [19].

Weibull dağılımı yaşam çözümlemesi modellerinde en yaygın kullanılan dağılımdır. Risk ve sigortacılık gibi alanlarda sıkça kullanılmaktadır. Başarısızlığın oluşmasına kadar geçen

süreyi ya da başarısızlıktan sonra ikinci bir başarısızlığın oluşmasına kadar geçen süreyi modellemede kullanılan bir dağılımdır [20].

Weibull dağılımı iki parametrelidir ve $W(\alpha, \beta)$ ile gösterilir. Bu dağılım Johnson v.d. (1970) ve Barnett (1982) gibi birçok araştırmacı tarafından da çalışılmıştır [19].

Tehlike fonksiyonu,

$$h(t, \alpha, \beta) = \alpha \beta^{-1} (\beta^{-1} t)^{\alpha-1}$$

ile verilir. Weibull dağılımı için yaşam fonksiyonu,

$$S(t, \alpha, \beta) = \exp[-(\beta^{-1} t)^\alpha]$$

ve dağılım fonksiyonu,

$$F(t; \alpha, \beta) = 1 - \exp[-(t/\beta)^\alpha], \quad t \geq 0$$

biçiminde verilir.

Burada β ölçek ve α ise biçim parametresidir. $\alpha = 1$ olduğu zaman Weibull dağılımı üstel dağılım olur. Buradan Weibull dağılımının olasılık yoğunluk fonksiyonu,

$$f(t, \alpha, \beta) = \alpha \beta^{-1} (\beta^{-1} t)^\alpha \exp[-(t/\beta)^\alpha], \quad t \geq 0$$

ile ifade edilir [21].

2.2.3. Log-Normal Dağılım

Yaşam çözümlerinde, elde edilen veriye hangi dağılım şeklinin uygulanacağı konusu ayrı bir önem taşımaktadır. Bazı durumlarda belli bir dağılımı seçmek için bir neden olmaktadır. Örneğin, geçmiş deneyimler modelin, daha önce incelenen benzer kitleleri oldukça iyi bir şekilde tanımladığını göstermektedir ya da geçerliliğini sağlayan başarısızlık sürecine ait bir bilgi elde edilebilmektedir. Bazen model, var olan veriye iyi uyum sağlasa bile tehlike fonksiyonunun formundan dolayı uygun olmayabilir. Log-normal dağılım böyle bir dağılımdır ve azalan tehlike fonksiyonunun sahip olduğu forma göre uygun olmaktan uzaklaşabilmektedir. Benzer olasılık yoğunluk fonksiyonlarına sahip iki dağılım çok farklı tehlike fonksiyonlarına sahip olabilmektedir. Bu durumda tehlike fonksiyonunun yapısına ilişkin özel bir bilgi ya da varsayım, dağılımlarından biri lehine olmak üzere karar verilmesi sağlanabilmektedir [19].

Eğer $\log T$, μ ortalamalı ve σ^2 varyansı ile normal dağılıma sahip ise, T rasgele değişkeninin μ ve σ parametrelili log-normal dağılıma sahip olduğu söylenir.

Log-normal dağılımın tehlike fonksiyonu,

$$h(t; \mu, \sigma) = f(t) S^{-1}(t)$$

ile verilir. Log-normal dağılım için yaşam fonksiyonu,

$$S(t, \mu, \sigma) = 1 - \Phi\left[\frac{\ln t - \mu}{\sigma}\right]$$

ile verilir. Burada, Φ normal birikimli dağılım fonksiyonudur.

Eğer T rasgele değişkeni log-normal dağılıma sahip ise $T \sim \text{LN}(\mu, \sigma)$ ile gösterilir.

Böylece $T \sim \text{LN}(\mu, \sigma)$ 'in dağılım fonksiyonu,

$$F(t, \mu, \sigma) = \Phi\left[\frac{\ln t - \mu}{\sigma}\right], t > 0$$

ile verilir.

Buradan olasılık yoğunluk fonksiyonu ise,

$$f(t; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} t^{-1} \exp\left[-(\log t - \mu)^2 / 2\sigma^2\right], t > 0$$

ile verilir [21].

2.2.4. Log-Lojistik Dağılım

Log-lojistik dağılım, Weibull dağılımının alternatifidir. Eğer $\log(T)$ lojistik dağılıma sahipse T yaşam süresi log-lojistik dağılmaktadır. Bu dağılım iki parametreye sahiptir. Bunlar α ve λ ' dir. Bu dağılım başta artan daha sonra azalan bir şekle sahip olmaktadır. Log-normal dağılım ile karşılaştırıldığında benzer yaşam fonksiyonuna ve tehlike oranına sahip olmaktadır [19].

Weibull tehlike fonksiyonunun kısıtlı olduğu durumlarda log-lojistik dağılım yararlı olabilir. Bu dağılımın tehlike fonksiyonu,

$$h(t; \alpha, \lambda) = \alpha \lambda t^{\alpha-1} (1 + \lambda t^\alpha)^{-1}, t \geq 0$$

ile verilir. Eğer $\alpha \leq 1$ ise tehlike fonksiyonu monoton olarak azalır. Fakat $\alpha > 1$ ise tehlike fonksiyonu monoton olarak artar. Yaşam fonksiyonu,

$$S(t; \alpha, \lambda) = [1 + \lambda t^\alpha]^{-1}, t \geq 0$$

ile verilir. Log-lojistik dağılımın dağılım fonksiyonu,

$$F(t; \alpha, \lambda) = \lambda t^\alpha [1 + \lambda t^\alpha]^{-1}, t \geq 0$$

biçimindedir. Buradan olasılık yoğunluk fonksiyonu,

$$f(t; \alpha, \lambda) = \alpha \lambda t^{\alpha-1} [1 + \lambda t^\alpha]^{-2}, \quad t \geq 0$$

denklemleri ile verilir. Böyle bir olasılık yoğunluk fonksiyonuna sahip T rasgele değişkenine λ ve α parametrelili Log-lojistik dağılıma sahiptir denir [21].

2.2.5. Gamma ve Genelleştirilmiş Gamma Dağılımı

Gamma dağılımı Ki-kare ve Üstel dağılımı içermektedir. Brown ve Flood (1947) tarafından bir kafeteryadaki bardak kullanım zamanı için tanımlanmıştır. Birnbaum ve Saunders (1958) ise materyallerin yaşam uzunluğu için bir istatistiksel model olarak tanımlamışlardır. İzleyen dönemlerde de yaşam çözümlemesi problemlerinde sık bir şekilde kullanılmaya başlanmıştır [19].

Gamma dağılımı λ ve γ olmak üzere iki parametreye sahiptir.

$0 < \gamma < 1$ iken, grafik azalış yönündedir, yani zaman 0'dan sonsuza artarken tehlike oranı monoton bir şekilde sonsuzdan λ 'ya doğru azalmaktadır. $\gamma > 1$ ise bir artma söz konusu olmaktadır, yani tehlike oranı zaman artarken λ 'ya göre artmaktadır. $\gamma = 1$ ise tehlike oranı λ 'ya eşit olmaktadır.

Gamma dağılımının olasılık yoğunluk fonksiyonu,

$$f(t) = \frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} e^{-\lambda t} \quad t > 0, \lambda > 0, \gamma > 0$$

biçimindedir. Burada γ şekil, λ ölçüm parametresidir. Yani γ 'daki değişim grafiğin şeklini değiştirirken, λ 'daki değişim ölçümünü değiştirmektedir.

$\gamma > 1$ iken grafik tek bir tepe noktasına sahipken, bu tepe noktası $t = (\gamma - 1) / \lambda$ biçimindedir.

Gamma dağılımı için yaşam fonksiyonu,

$$S(t) = e^{-t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!},$$

Tehlike fonksiyonu ise,

$$h(t) = \frac{\lambda (\lambda t)^{n-1}}{(n-1)! \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}}$$

biçimindedir.

Genelleştirilmiş Gamma dağılımının λ , α ve γ olmak üzere üç parametresi vardır.

Genelleştirilmiş Gamma dağılımının olasılık yoğunluk fonksiyonu,

$$f(t) = \frac{\alpha \lambda^{\alpha \gamma}}{\Gamma(\gamma)} t^{\alpha \gamma - 1} \exp[-(\lambda t)^\alpha] \quad t > 0, \gamma > 0, \lambda > 0, \alpha > 0$$

biçimindedir.

Üstel, Weibull, log-normal ve Gamma dağılımları Genelleştirilmiş Gamma dağılımının özel halleridirler. Buna göre;

$\alpha = \gamma = 1$ ise Üstel dağılım,

$\gamma \rightarrow 1$ ise Weibull dağılımı,

$\gamma \rightarrow \infty$ ise Log-normal dağılım,

$\alpha = 1$ ise Gamma dağılımı elde edilir [6].

3. EŞ ZAMANLI YAŞAM SÜRELERİ

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 [4].

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 [5]. 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 [4].

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 [22]. Örneğin ölçüm birimi yıl olarak alındıysa ölüm anı aynı olmamasına karşın aynı sene içinde ölen iki birimin olay zamanı aynı olarak kaydedilmiş olur [5,22].

Kitlenin büyük olduğu veya geniş kohort kanser çalışmalarında ölçümlerin ayda bir ya da uzun aralıklarla yapıldığı durumlarda eş zamanlı gözlemlerin yoğun olduğu diğer durumlardır [23].

Eşitlik 2.1’ de verilen Cox’un kısmi olabilirlik fonksiyonu veri kümesinde eş zamanlı gözlem olmadığını varsaymaktadır [23]. 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 [4, 5, 22, 23,]. 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 [5, 22].

3.1. Eş Zamanlı Yaşam Süreleri Tahmin Edici Süreci

n gözlem için, k farklı sıralı başarısızlık (durdurulmamış) süreleri $t_{(1)} < t_{(2)} < \dots < t_{(k)}$ olsun. $m_{(i)}$, $t_{(i)}$ zamanında başarısız olan birim sayısını gösterebilir. $m_{(i)} > 1$ 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.

Örneğin, 8 birimin yaşam süreleri kümesi $\{15, 16, 20, 20, 20, 21, 24, 24\}$ olsun. Burada (+) durdurulmuş gözleme ait yaşam süresini göstermektedir. Burada,

$$n=8, k=4,$$

$$t_{(1)} = 15, m_{(1)} = 1, R(t_{(1)}) = \{8 \text{ gözlemi içerir}\}, r_1 = 8$$

$$t_{(2)} = 20, m_{(2)} = 3, R(t_{(2)}) = \{\text{Yaşam süreleri } 20, 21, 24 \text{ olan birimler}\}, r_2 = 6$$

$$t_{(3)} = 21, m_{(3)} = 1, R(t_{(3)}) = \{\text{Yaşam süreleri } 21, 24 \text{ olan birimler}\}, r_3 = 3$$

$$t_{(4)} = 24, m_{(4)} = 2, R(t_{(4)}) = \{\text{Yaşam süreleri } 24 \text{ olan birimler}\}, r_4 = 2$$

olmaktadır.

Eş zamanlı yöntemleri tartışabilmek için bazı ek gösterimler de verilmelidir. Her $R(t_{(i)})$ 'den rasgele $m_{(i)}$ birim seçilebilir. Bu seçilen $m_{(i)}$ 'lerin her biri $u_{(j)}$ ile gösterilir. $r_{(i)} C_{m_{(i)}} = r_i! / [m_{(i)}!(r_i - m_{(i)})!]$ ifadesi mümkün olan tüm $u_{(j)}$ 'lerin sayısını verir. U_i , tüm $u_{(j)}$ 'leri içeren kümedir. Örneğin, $R(t_{(2)})$ 'den rasgele herhangi bir $m_{(2)} = 3$ ($r_{(2)} = 6$) seçilmiş olsun. Burada toplam ${}_6C_3 = 20$ seçim (altküme) vardır ve $u_{(j)}$ 'lerden biri örneğin 20, 20, 24 yaşam sürelerine sahip üç birimden oluşur. $U_2 = \{u_{(1)}, u_{(2)}, \dots, u_{(20)}\}$ tüm 20 altkümeyle içerir.

Bu gösterimlerden sonra eş zamanlı gözlemlere odaklanılsın. $x_k = (x_{1k}, x_{2k}, \dots, x_{pk})'$, k. birimin açıklayıcı değişkenlerini, $Z_{u_{(j)}} = \sum_{k \in u_{(j)}} x_k = (z_{1u_{(j)}}, z_{2u_{(j)}}, \dots, z_{pu_{(j)}})'$ gösterebilir. Burada $z_{lu_{(j)}}$, $u_{(j)}$ 'deki $m_{(i)}$ birimlerinin l. açıklayıcı değişkeninin toplamıdır. $u_{(i)}^*$, $t_{(i)}$ zamanında başarısız olan $m_{(i)}$ birimin kümesini ve $z_{u_{(i)}^*} = \sum_{k \in u_{(i)}^*} x_k = (z_{1u_{(i)}^*}, z_{2u_{(i)}^*}, \dots, z_{pu_{(i)}^*})'$ gösterebilir. Burada $z_{1u_{(i)}^*}$, $u_{(i)}^*$ deki ($t_{(i)}$ zamanında başarısız olan) $m_{(i)}$ birimlerin l. açıklayıcı değişkeninin toplamıdır. Örneğin, yaşam sürelerinin yukarıda verilen kümesi için, $z_{1u_{(2)}^*}$, yaşam süresi olan 20'de başarısız olan üç birimin ilk açıklayıcı değişken değerlerinin toplamına eşittir [6].

3.1.1 Sürekli Zaman Ölçümü İle Belirlenen Eş Zamanlı Gözlemler

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 3.1’de verilen olabilirlik fonksiyonu kullanılabilir:

$$L(\beta) = \prod_{i=1}^k \frac{\exp(\beta'x_{(i)})}{[\sum_{l \in R(t_{(i)})} \exp(\beta'x_l)]} \quad (3.1)$$

[6].

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 eğer $m_{(i)}$ büyükse kısmi olabilirlik fonksiyonunun kurulumu ve hesabı oldukça sıkıntılı olabilmektedir [4, 6, 23, 24].

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 [6].

3.1.2. Kesikli Zaman Ölçümü İle Belirlenen Eş Zamanlı Gözlemler

Eğer 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:

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

Bu model sürekli zaman ölçeğinde aşağıdaki denkleme dönüşmektedir.

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

Kesikli zaman ölçeğinde eş zamanlı gözlemlerle kısmi olabilirlik fonksiyonu:

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

Bu denklemdeki i . terim gözlenen $m_{(i)}$ başarısızlıklarının koşullu olasılığını gösterir. $m_{(i)}$ $t_{(i)}$ zamanında ve $R_{(i)}$ risk kümesindeki başarısızlıklardır. Paydadaki i . Terimlerin sayısı $r_i C_{m(i)=r_i} / [m_{(i)}!(r_i - m_{(i)})!]$ dir. Daha önce belirtildiği gibi eğer $m_{(i)}$ büyükse sonuç çok büyük çıkabilir. Ancak Gail vd. (1981) yaptıkları çalışmada kullandıkları bir algoritma ile bu değeri hesaplamayı başarmıştır. Eşitlik 3.1’de eş zamanlı sürekli yaşam süreleri için kısmi olasılık fonksiyonunun bir yaklaşımı olarak düşünülebilir (eş zamanlı gözlemlerin, kesikli zaman ölçeğinde gözlenmiş gibi, doğru şekilde sıralandığı varsayımı altında) [6].

3.2. Eş Zamanlı Yaşam Süreleri Oluştugu 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 kümesinde eş zamanlı süreler olmamasıdır [5, 23]. Fakat gerçek uygulamalarda eş zamanlı olay süreleri genellikle gözlenmektedir. Bu durumda Cox’un kısmi olasılık fonksiyonu yeniden düzenlenmelidir. Aşağıdaki kalp nakli ile ilgili örnek veri kümesi eş zamanlılığı göstermektedir:

Çizelge 3.1. Örnek veri

No	Başarısızlık süresi	Nakil zamanı	Durum
1	12	10	1
2	12	8	1
3	16	NA	1
4	9	NA	1
5	20	7	0
6	9	5	1
7	11	NA	0

“No” sütunu hastaların numarasını, “Başarısızlık süresi” sütunu her bir bireyin başarısızlık süresini göstermektedir. Bu süre durdurma süresi ya da yaşam süresi olabilir. “Nakil zamanı” sütunu hastalara kalp naklinin yapıldığı zamanı, bu sütundaki “NA” ise izlem süresi içerisinde kalp naklinin yapılmadığını göstermektedir. Son sütun olan “durum” sütunu ise izlem süresi içerisinde bireyin öldüğü (1) ya da ölmediğini (0) yani durdurulmuş gözlem olduğunu gösteren göstergedir.

Bu veri kümesinde hasta₍₁₎ ile hasta₍₂₎ ve hasta₍₄₎ ile hasta₍₆₎ aynı yaşam süresine sahiptir. Bu nedenle hasta₍₁₎ ile hasta₍₂₎ ve hasta₍₄₎ ile hasta₍₆₎ eş zamanlı gözlemlerdir.

Eş zamanlı yaşam verileri olduğu durumda kullanılan beş yöntem geliştirilmiştir. Bu yöntemler, Allison (2010) tarafından önerilen Kesin (Exact) yöntem, Breslow'un (1974) geliştirdiği Breslow yöntemi, Efron (1977) tarafından önerilen Efron yöntemi, Kesikli yöntem (Kalbfleisch vd., 1973) ve Scheike ve Sun (2007) tarafından önerilen EM algoritmasına dayalı yöntemdir [5, 12].

3.2.1. 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. Örneğin Çizelge 3.1' de verilen veri kümesinde, eş zamanlı iki hasta aynı yaşam süresine sahiptir. Bunun nedeni ölçümlerin doğru yapılamamış olması veya orijinal sonucun kolaylık için yuvarlanması, bu şekilde gerçek bilginin kaybedilmiş olmasıdır. Gerçekte olayların öncelikli bir sırası vardır. Bu nedenle uygun model için kısmi olabilirlik hesaplanacağı zaman, mümkün olan tüm sıralamaların değerlendirilmesi gerekir [25, 26, 27].

Örneğin Çizelge 3.1' de verilen veri kümesinde, hasta₍₁₎'in olay zamanı hasta₍₂₎'den önce veya sonra olabilir. Kesin yöntem, bu birimlerin ölçüm kısıtları nedeniyle eş zamanlı olarak kaydedilmiş olduğunu varsaymaktadır. Kesin yöntemde kısmi olabilirlik her iki olasılığı, yani mümkün olan tüm sıralamaların olasılıklarının toplamını içermektedir. Bu veri kümesi için yaşam süresi = 12'nin kısmi olabilirliği aşağıda verildiği gibidir:

$$L_{12} = \frac{1}{2} \left(\frac{e^{\beta z_1}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}} \right) \left(\frac{e^{\beta z_2}}{e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}} \right) + \frac{1}{2} \left(\frac{e^{\beta z_2}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}} \right) \left(\frac{e^{\beta z_1}}{e^{\beta z_1} + e^{\beta z_3} + e^{\beta z_5}} \right)$$

Yukarıda verilen kısmi olabilirlik iki çarpımın toplamının katkısıyla oluşmaktadır. Eğer her bir çarpım Eşitlik 3.1' de verilen Cox'un kısmi olabilirliği ile karşılaştırılırsa, ilk çarpım hasta₍₁₎'in hasta₍₂₎'den önce başarısız (ölüm) olduğu olasılığı için; ikinci çarpımsa hasta₍₂₎'nin hasta₍₁₎'den önce başarısız olduğu olasılığı içindir.

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 olabilmektedir [5, 25].

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 [28].

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 gelmişse, eş zamanlı yaşam süresine sahip gözlemlerden başarısız olanların, durdurulmuşlardan önce meydana geldiği varsayılarak hesaplanan 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 [29].

3.2.2. Breslow Yöntemi

Breslow yöntemi, başarısızlık sürelerinin sürekli olduğunu ve olay riskinin (t_i, t_{i+1}) aralığında sabit olduğunu varsaymaktadır [7]. 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.

z_j , j. bireyin açıklayıcı değişken değerlerinin vektörü; D_i , 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 [5, 23].

$$L(\beta) = \prod_{i=1}^k \frac{\exp [(\sum_{j \in D_i} z_j)\beta]}{[\sum_{j \in R_i} \exp (z_j \beta)]^{d_i}} \quad (3.2)$$

Sonuç olarak, 12 süresinin kısmi olabilirliğe katkısı yaklaşık olarak aşağıdaki gibidir:

$$\frac{e^{\beta z_1} e^{\beta z_2}}{[e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}]^2}$$

Hesaplama bakımından Breslow yöntemi, Kesin yöntemin kısmi olabilirliğinin hesaplamasından çok daha basittir. Fakat belirli zaman noktasındaki eş zamanlı gözlemlerin sayısı, risk kümesindeki hastaların sayısına göre daha fazla olduğunda Breslow yönteminin verdiği sonuçlar bozulmaktadır [5, 22].

3.2.3. 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 [10, 25].

Efron yöntemi için kısmi olabilirlik fonksiyonu

$$L(\beta) = \prod_{i=1}^k \frac{\exp [(\sum_{j \in D_i} z_j)\beta]}{\prod_{l=1}^{d_i} [\sum_{j \in R_l} \exp (z_j \beta) - \frac{l-1}{d_i} \sum_{j \in D_i} \exp (z_j \beta)]}$$
 biçimindedir [23].

Sonuç olarak, 12 süresinin kısmi olabilirliğe katkısı yaklaşık olarak aşağıdaki gibidir:

$$\frac{e^{\beta z_1} e^{\beta z_2}}{[e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}] [(e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} + e^{\beta z_5}) - \frac{1}{2} (e^{\beta z_1} + e^{\beta z_2})]}$$

Breslow yöntemi ile karşılaştırma yapıldığında, paydalar Breslow yöntemindeki kadar basit olmadığından Efron yönteminin kısmi olabilirliği, Kesin yöntemden daha yakındır. Efron yönteminin hesaplanması ise Kesin yöntemden daha basit olmaktadır [5, 23].

Efron yöntemi, bu özellikleri ile Kesin yöntem ile Breslow yöntemi arasında yer almaktadır [27].

3.2.4. Kesikli Yöntem

Kesikli yöntem, yaşam süresinin gerçekten kesikli olduğunu varsayar [3, 25].

Çizelge 3.1.'de verilen veri kümesinde yaşam süresi 12 için hasta₍₁₎ ile hasta₍₂₎ arasında eş zamanlılık mevcuttur. Kesikli yöntem, bu 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 [3, 25]. Fakat, bu modelin hesaplamaları zaman aldığından Kalbfleisch ve Prentice (1973) tarafından başka bir yöntem geliştirilmiştir [11].

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 ve Rockhill'in (1997) çalışmalarındaki gibi, Kalbfleisch ve Prentice (1973) tarafından önerilen kesikli yöntem için kısmi olabilirlik aşağıda verildiği gibidir:

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

Breslow ve Efron yöntemlerinde olduğu gibi kesikli yöntemin hesaplanması da kesin yöntemden daha basittir. Fakat çalışmalarda olay oranı yükseldikçe veya ölçümler belirsizleştikçe daha fazla eş zamanlı gözlem meydana geleceğinden kesikli yöntemin parametre tahminleri daha yanlı olabilmektedir [5, 11, 23].

3.2.5. EM Algoritmasına Dayalı Yöntem

Scheike ve Sun (2007) çalışmalarında Breslow, Efron yöntemlerinin ve eş zamanlı yaşam verileri ile ilgili diğer yöntemlerin yanlı tahminler verdiğini belirtmişlerdir. Çalışmalarında bu yöntemleri inceleyerek, doğrudan çözülebilen skor eşitliğini sağlayan EM-algoritmasını kullanarak kayıp gözlem prensibine dayanan ve yansız tahminler üreten yeni bir yöntem önermişlerdir. Bu skor sıfır ortalamaya sahiptir. Ayrıca, tüm incelenen yöntemlerin aynı asimptotik özelliklere sahip olduklarını ve eş zamanlı gözlem büyüklüğü sınırlı olduğunda ya da verilen bir hızda sonsuzluğa yaklaştığında asimptotik etkinlik kaybı olmadığını göstermişlerdir. Bir simülasyon çalışmasıyla yöntemlerin sonlu örneklem özelliklerini karşılaştırmışlardır. Eş zamanlı yaşam verilerinde Efron yönteminin mevcut yöntemlerden daha iyi performansa sahip olduğunu belirtmişlerdir. Aynı zamanda Efron yöntemi ile EM süreci karşılaştırıldığında, Efron yönteminin ikili (binary) ve tek düze (uniform) açıklayıcı değişkenlerle daha iyi çalıştığı belirtilmiştir. Normal ve üstel açıklayıcı değişkenler için özellikle eş zaman gözlem sayısı fazla ise veya küçük örneklem büyüklüğüne sahip bir veri kümesi ile çalışılıyorsa Efron yönteminin EM'den daha büyük yana sahip olduğu belirtilmiştir [12].

3.2.6. Yöntemlerin Karşılaştırılması

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

Breslow ve Efron yöntemlerinin her ikisinde de hesaplamalar daha basit ve hızlıdır. Ancak özellikle veri kümesinin 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 [5].

Picciotto ve Rockhill (1997) çalışmalarında Breslow (1974), Efron (1977) ve Kalbfleisch-Prentice (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 (default) olarak bulunmaktadır. Buna rağmen Efron yöntemi, özellikle örneklem büyüklüğü küçükse ya da veri kümesinde çok fazla durdurulmuş gözlem varsa Breslow yöntemine tercih edilmektedir.

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

Bu çalışmada, Picciotto ve Rockhill (1997), durdurulmuş gözlemin olmadığı çeşitli senaryolar altında eş zamanlı başarısızlık süreleri için Efron, Breslow ve Kesikli yöntemleri bir simülasyon çalışmasıyla karşılaştırmışlardır. Cox regresyonda eş zamanlı başarısızlık süreleri varlığında kullanılan olabilirlikler içinden Efron yöntemi, Breslow ve kesikli zaman yöntemlerinden daha iyi sonuç vermiştir. Ortalama olarak, Breslow yöntemi regresyon katsayılarını tahminin altında bulurken, Kalbfleisch ve Prentice yöntemi de tahminin üstünde hesaplamaktadır. Yoğun eş zamanlı gözlem ve oldukça az örneklem genişliğine rağmen Efron tahmin edicisinin çok iyi performans gösterdiğini ifade etmişlerdir. Başarısızlık süreleri aralıklarla kaydedildiğinde veya başarısızlıkları tek bir sıra halinde sıralamak mümkün olmadığında Efron yönteminin kullanılmasının daha iyi olduğunu belirtmişlerdir [23, 30].

Bazı kaynaklarda, eş zamanlı gözlem olduğu 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 kümesinde eş zamanlı gözlemler yoksa hangi yöntem kullanılırsa kullanılsın sonuçlar aynı olmaktadır [29, 30].

Scheike ve Sun (2007) çalışmalarında, Efron yönteminin mevcut yöntemlerden eş zamanlı yaşam verilerinde en iyi performansa sahip olduğunu, normal ve üstel açıklayıcı değişkenler için ise özellikle çok fazla eş zaman büyüklüğünde ve küçük örneklem büyüklüğünde EM'den daha büyük yana sahip olduğunu belirtmişlerdir [12].

4. UYGULAMA

Çalışmamızda öncelikle eş zamanlı gözlemlerin olduğu duruma örnek olarak mide kanseri verilerine, daha sonra ise eş zamanlı gözlemlerin olmadığı duruma örnek olarak 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 [31]. 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 çözümlene yapılmış ve sonuçlar karşılaştırılmıştır. Çalışmamızda anlatılan eş zamanlı yaşam sürelerinin varlığında kullanılan bir diğer yöntem olan Kesikli yöntem, kesikli ölçümlü verilere uygulandığından veri kümemeze uygulanmamıştır. Başarısızlık, ölüm olarak alınmıştır. Yaşam süreleri ay bazında ölçülmüştür. Çözümlenmeler için IBM SPSS Statistics 22 ve Stata 11 paket programları kullanılmıştır.

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'tir 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 Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Açıklayıcı değişkenler

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

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 ortalama yaşam olasılıkları ise Çizelge 4.2’de verilmiştir.

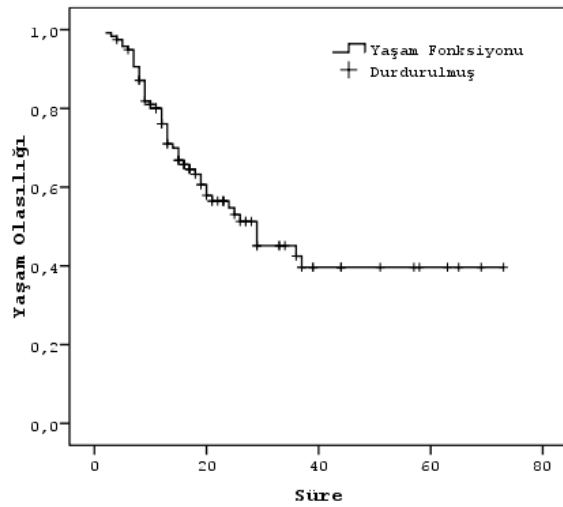
Çizelge 4.2. Kaplan-Meier sonuçları

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

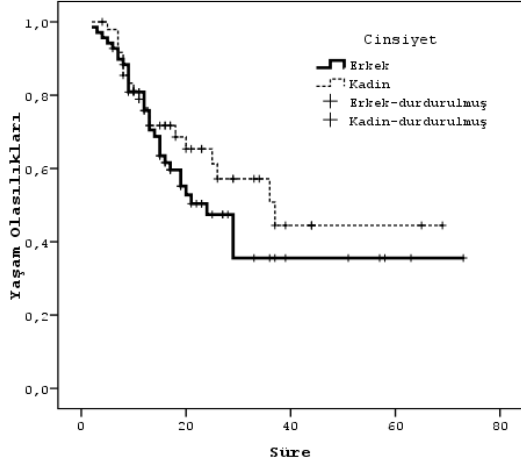
Elde edilen log-rank sonuçlarına göre lenf nodu diseksiyonunun genişliği, tümörün midedeki lokalizasyonu ve hastalığın evresi değişkenlerin düzeyleri arasında yaşam süreleri açısından fark olduğu %95 güven düzeyinde söylenebilir.

Kaplan-Meier eğrileri Şekil 4.1 ile Şekil 4.8 arasında verilmiştir.

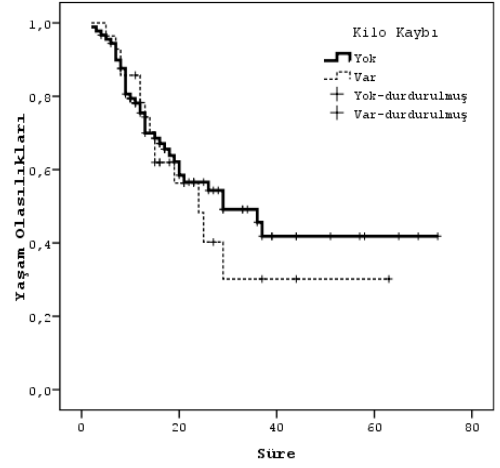


Şekil 4.1. Genel Kaplan-Meier eğrisi

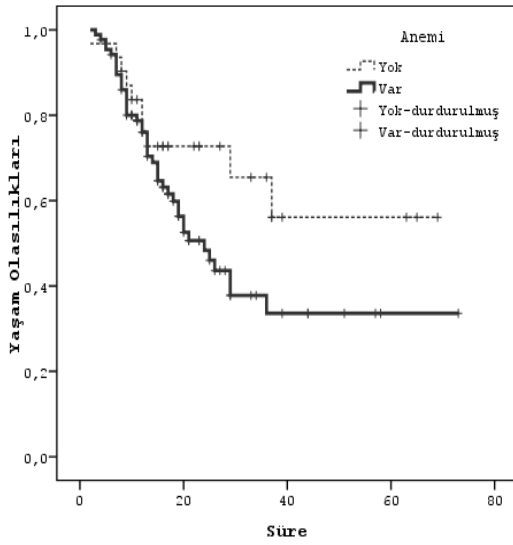
Şekil 4.1’de genel Kaplan Meier Eğrisi verilmiştir. Şekil 4.2 – Şekil 4.8’ de ise diğer açıklayıcı değişkenlere ait Kaplan Meier eğrileri gösterilmektedir.



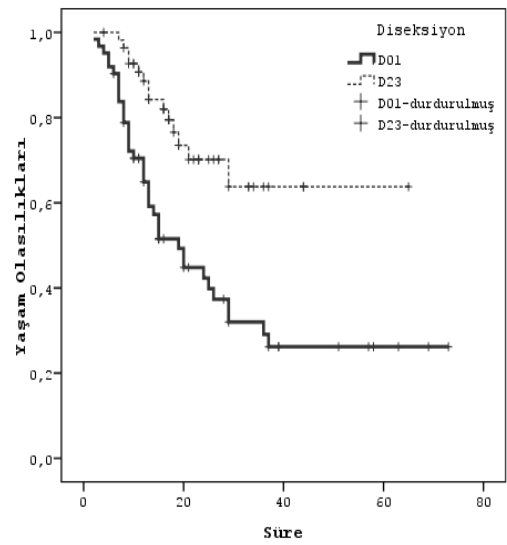
Şekil 4.2. Cinsiyet değişkeni için Kaplan-Meier eğrisi



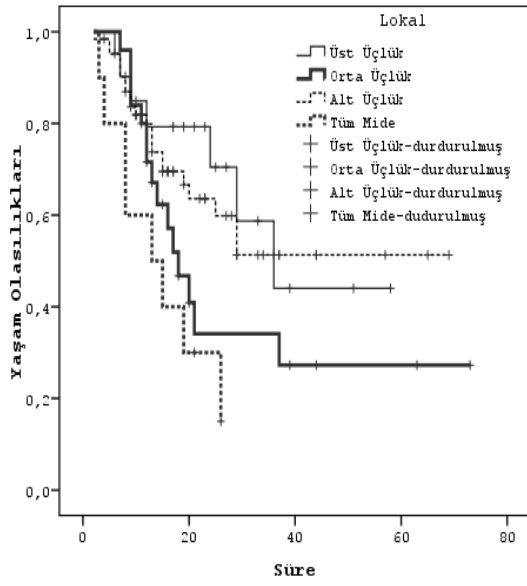
Şekil 4.3. Kilo Kaybı değişkeni için Kaplan-Meier eğrisi



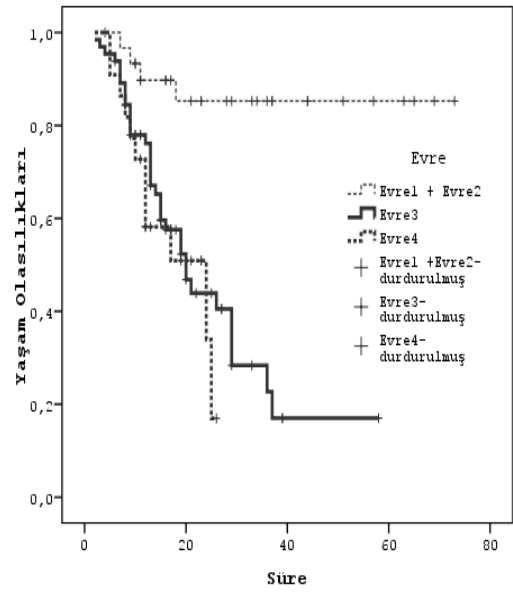
Şekil 4.4. Anemi değişkeni için Kaplan-Meier eğrisi



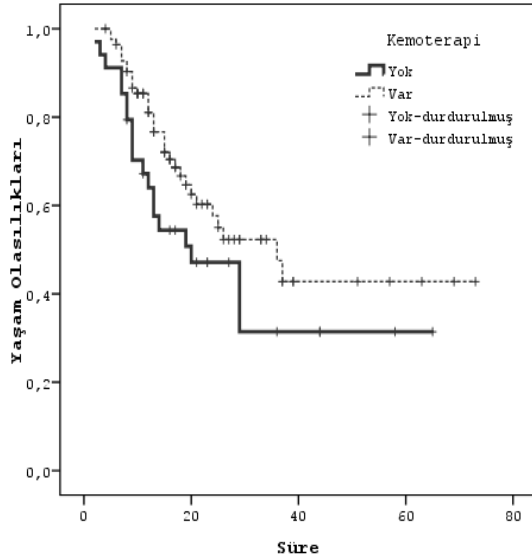
Şekil 4.5. Diseksiyon değişkeni için Kaplan-Meier eğrisi



Şekil 4.6. Lokal değişkeni için Kaplan-Meier eğrisi



Şekil 4.7. Evre değişkeni için Kaplan-Meier eğrisi



Şekil 4.8. Kemoterapi değişkeni için Kaplan-Meier eğrisi

Şekil 4.2 incelendiğinde kadınlarda yaşam süresinin erkeklere göre daha uzun olduğu görülmesine rağmen Çizelge 4.2’de verilen log-rank testi sonucunda kadınlar ile erkekler arasında yaşam olasılıkları açısından istatistiksel olarak önemli bir fark bulunmamaktadır.

Çizelge 4.2 de verilen log-rank testi sonucuna göre, bir semptom olan kilo kaybının gözlemediği ve gözlenmediği hastalar arasında yaşam olasılıkları açısından istatistiksel olarak önemli bir fark bulunmamaktadır.

Şekil 4.4 incelendiğinde anemi görülmeyen hastalar anemi olanlara göre ortalama olarak daha uzun bir yaşam süresine sahip olduğu gözlenmektedir ancak Çizelge 4.2’de log-rank testi sonucunda anemi görülen ve görülmeyen hastalar arasında yaşam olasılıkları açısından istatistiksel olarak önemli bir fark bulunmamaktadır.

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 4.5’ ten de görüldüğü gibi D2-3 olan hastalar D0-1 olanlara göre daha uzun bir yaşam süresine sahiptir.

Çizelge 4.2’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 şu şekilde gözlenmiştir:

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

Çizelge 4.2’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 şu şekilde 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.

Şekil 4.8'deki Kaplan Meier grafiğine göre kemoterapi alan hastaların yaşam süreleri kemoterapi almayan hastalardan daha uzundur. Ancak Çizelge 4.2'de verilen log-rank testi sonucuna göre kemoterapi gören ve görmeyen hastalar arasında yaşam olasılıkları bakımından % 95 güven düzeyinde istatistiksel olarak önemli bir fark bulunmamaktadır.

Çizelge 4.1'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.

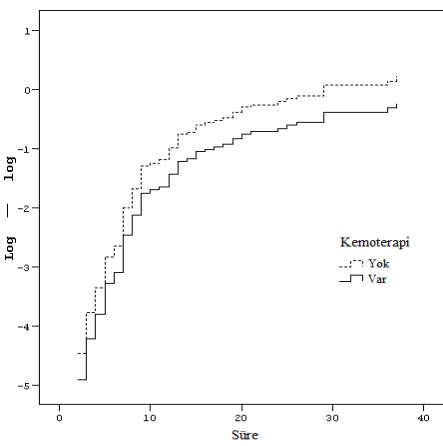
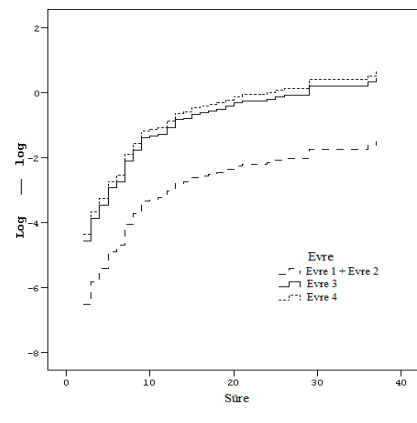
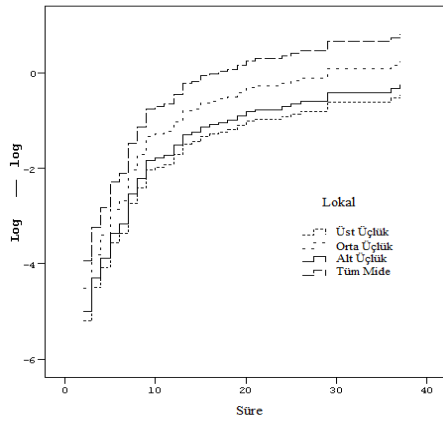
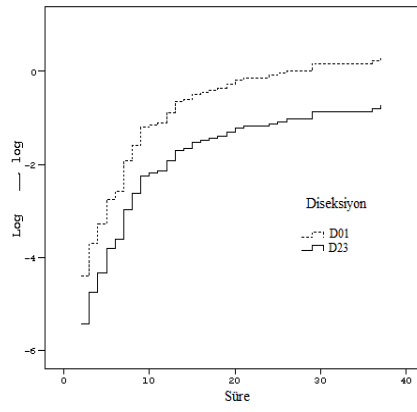
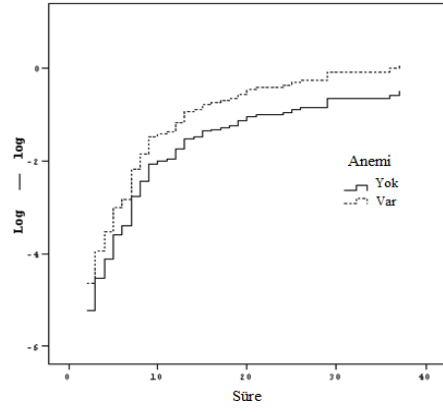
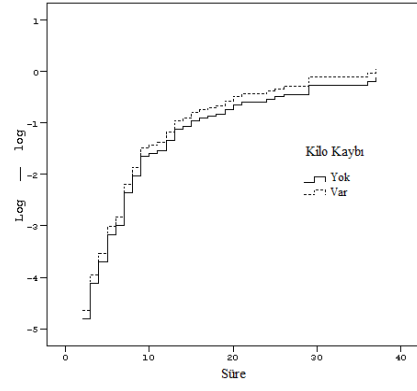
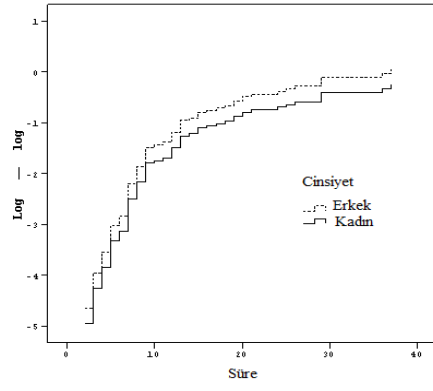
Orantılı tehlikeler varsayımını incelenmek için kullanılan yöntemlerden biri, yaşam süresi rankının, Schoenfeld artıkları ile ilişkisini incelemektir. Bu yöntem sonucunda tüm açıklayıcı değişkenler için edilen sonuçlar Çizelge 4.3' te verilmiştir.

Çizelge 4.3. Yaşam süresi rankı ile Schoenfeld artıkları ilişkisinin incelenmesi

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

Çizelge 4.3 incelendiğinde, tüm değişken düzeylerinde $p > 0,05$ olarak bulunduğu için orantılı tehlikeler varsayımının sağlandığı görülmüştür [32].

Orantılı tehlikeler varsayımının sağlanıp sağlanmadığını incelemeye kullanılan grafiksel yöntemlerden log(-log) yaşam eğrileri yöntemi kullanılarak incelendiğinde ise elde edilen grafikler Şekil 4.9'da 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 4.9. Açıklayıcı değişkenlerin log(-log) yaşam eğrileri

Orantılı tehlikeler varsayımı sağlandığından Cox regresyon modelinin uygulanabilir olduğu sonucuna varılmıştır. Ancak Çizelge 4.4’ te verilen yaşam süreleri incelendiğinde 118 hastaya ait 41 farklı yaşam süresinin 27’sinde eş zamanlılık olduğu dikkat çekmektedir.

Çizelge 4.4. 118 hastaya ilişkin yaşam süreleri

Süre	Frekans	Yüzde	Birikimli Yüzde	Süre	Frekans	Yüzde	Birikimli Yüzde
2	1	0,8	0,8	23	4	3,4	72
3	1	0,8	1,7	24	1	0,8	72,9
4	2	1,7	3,4	25	2	1,7	74,6
5	2	1,7	5,1	26	2	1,7	76,3
6	2	1,7	6,8	27	2	1,7	78
7	5	4,2	11	28	1	0,8	78,8
8	6	5,1	16,1	29	5	4,2	83,1
9	7	5,9	22	33	2	1,7	84,7
10	5	4,2	26,3	34	1	0,8	85,6
11	6	5,1	31,4	36	2	1,7	87,3
12	6	5,1	36,4	37	3	2,5	89,8
13	7	5,9	42,4	39	2	1,7	91,5
14	1	0,8	43,2	44	3	2,5	94,1
15	6	5,1	48,3	51	1	0,8	94,9
16	5	4,2	52,5	57	1	0,8	95,8
17	6	5,1	57,6	58	1	0,8	96,6
18	2	1,7	59,3	63	1	0,8	97,5
19	4	3,4	62,7	65	1	0,8	98,3
20	3	2,5	65,3	69	1	0,8	99,2
21	3	2,5	67,8	73	1	0,8	100
22	1	0,8	68,6				

Verilerimizde eş zamanlı gözlemler olduğunda kullanılan Cox regresyon için Breslow, Efron ve Kesin yöntemler uygulanmış ve elde edilen sonuçlar Çizelge 4.5 – Çizelge 4.7’de verilmiştir.

Çizelge 4.5. 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 (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 3	1,80	0,55	0,00*	6,05	2,04 ; 17,92
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.

Çizelge 4.6. 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	0,73	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 3	1,83	0,56	0,00*	6,28	2,12 ; 18,66
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.

Çizelge 4.7. Kesin Marjinal 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 3	1,84	0,56	0,00*	6,30	2,12 ; 18,74
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.

Yöntemleri karşılaştırmak için aşağıda eşitlikleri verilen Akaike Bilgi Kriteri (AIC) ve Bayesci Bilgi Kriteri (BIC) kullanılmış ve elde edilen sonuçlar Çizelge 4.8’de verilmiştir:

$$AIC = -2\log L + \alpha p$$

$$BIC = -2\log L + p \log(n)$$

Bu eşitliklerde p; modeldeki parametre sayısını göstermektedir. α ise %95 güven düzeyine karşılık geldiği için 3 olarak alınmıştır.

Çizelge 4.8. Yöntemlerin karşılaştırılması

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

Çizelge 4.8 incelendiğinde AIC ve BIC değerlerine göre Kesin marjinal yöntem ile Cox regresyon modelinin en uygun model olduğu görülmektedir. Bu yöntem uygulandığında elde edilen sonuçlar Çizelge 4.7’de verilmiştir. Çizelge 4.7 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 kümesine uygulanan Kesin marjinal, Efron ve Breslow yöntemleri incelendiğinde β katsayıları ve tehlike oranları açısından Breslow yönteminin diğer iki yonteme 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.

Miyokard infarktüsü, kalbi besleyen koroner atardamarlardan birinin ya da birden fazlasının pıhtıyla tıkanması sonucunda kalp kasının (miyokard) bir bölümünün beslenemeyerek ölmesidir. Kalp krizi olarak da adlandırılan bu durum, kan dolaşımının kalbin bir bölümünde hücrelerin gereksinimini karşılayamayacak düzeye düşmesiyle ortaya çıkar.

Açıklayıcı değişkenler olarak cinsiyet (Erkek, Kadın), yaş (min:32 ; max:92), hastanede kalış süresi (min:1; max:56), vücut kitle indeksi (min:14,91878; max:39,93835) alınmıştır. Bu çalışmada 100 hastanın 65'i erkek, 35i kadındır. Yaş ortalaması $68,25\pm 1,44$, hastanede kalış süresi ortalaması $6,84\pm 0,59$ ve vücut kitle indeksi ortalaması $27,04\pm 0,49$ 'dur. Gözlemlerin 49'u durdurulmuş, 51'i ise başarısız olmuştur. Yaşam süreleri Çizelge 4.9'da verilmiştir.

Çizelge 4.9. 100 hastaya ilişkin yaşam süreleri

Süre	Frekans	Süre	Frekans	Süre	Frekans	Süre	Frekans
6	2	841	1	1883	1	2114	1
14	1	936	1	1889	1	2124	1
44	1	1002	1	1907	1	2137	1
62	1	1011	1	1912	1	2137	1
89	1	1048	1	1916	1	2145	1
98	1	1054	1	1922	1	2157	1
104	1	1172	1	1923	1	2173	1
107	1	1205	1	1929	1	2174	1
114	1	1278	1	1934	1	2183	1
123	1	1401	1	1939	2	2190	1
128	1	1497	1	1969	1	2201	1
148	1	1557	1	1984	1	2421	1
182	1	1577	1	1993	1	2573	1
187	1	1624	1	2003	1	2574	1
189	1	1669	1	2012	1	2578	1
274	2	1806	1	2013	1	2595	1
302	1	1836	2	2031	1	2610	1
363	1	1846	1	2052	1	2613	1
374	1	1859	1	2054	1	2624	1
451	1	1860	1	2061	1	2631	1
461	1	1870	1	2065	1	2638	1
492	1	1874	1	2072	1	2641	1
538	1	1876	1	2074	1	2710	1
774	1	1879	1	2084	1	2719	1

Çizelge 4.9’da görüldüğü gibi 6., 174., 1836. ve 1939. günlerde iki vaka ölümle sonuçlanmıştır. Bu 4 zaman noktası dışında kalan tüm yaşam süreleri birbirinden farklıdır. Yaşam süreleri gün bazında daha hassas olarak kaydedildiğinden veri kümesinde 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 ($p < 0,05$). Dolayısıyla Çizelge 4.9’da verilen yaşam süreleri için Breslow, Efron ve Kesin Marjinal yöntemler için Cox regresyon modeli uygulanabilmektedir.

Çizelge 4.10. Breslow yöntemi için Cox regresyon modeli sonuçları

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				

Çizelge 4.11. 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)	395,73				

Çizelge 4.12. Kesin marjinal 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				

Çizelge 4.10 – Çizelge 4.12 incelendiğinde veri kümesinde eş zamanlı gözlemlerin olmaması durumunda, uygulanan bu yöntemlerin sonuçları ($\hat{\beta}$ katsayıları, katsayıların standart hataları, 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 çözümlenmesi ve yaşam çözümlenmesinde eş zamanlı yaşam süreleri hakkında genel bilgiler, fonksiyonlar ve modeller verilmiş, yaşam çözümlenmesinde eş zamanlı yaşam süreleri olduğu durumda kullanılacak yöntemler anlatılmıştır.

Yaşam çözümlenmesinde 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 gözlemlendiği duruma göre yeniden düzenlenmelidir. Breslow yöntemi, Efron yöntemi, Kesin yöntem, Kesikli yöntem ve EM algoritmasına dayalı yöntem, yaşam verilerinde eş zamanlı gözlemler olduğu durumda kullanılacak Cox regresyon yöntemleridir.

Uygulamada öncelikle eş zamanlı yaşam sürelerinin yer aldığı 7 açıklayıcı değişkene ve 118 gözleme sahip mide kanseri verisine uygun yöntemler uygulanmıştır. Cox regresyon modelinin orantılı tehlikeler varsayımını sağlaması gerekmektedir. Bu varsayım iki farklı yöntemle araştırılmış ve her iki yönteme göre de varsayımın sağlandığı gözlenmiştir. Daha sonra sürekli ölçümlü yaşam sürelerine sahip olan veri kümesine 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, yorumlar bu model doğrultusunda yapılmıştır. Ayrıca Breslow yönteminde risk oranları ve model katsayıları daha küçük bulunmuş olup kesin yöntemde bu sonuçların diğer yöntemlere göre daha büyük olduğu gözlenmiştir. Daha sonra eş zamanlı yaşam sürelerinin olmadığı veri kümesine örnek olarak AMI verilerine, orantılı tehlikeler varsayımının sağlanmasının ardından aynı yöntemler uygulanmış ve sonuçlarının aynı olduğu gözlenmiştir.

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

KAYNAKLAR

- [1] Sertkaya, D., Ata, N., Sözer, M.T., Yaşam çözümlemesinde zamana bağlı açıklayıcı değişkenli Cox regresyon modeli, *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 58:153-158, **2005**.
- [2] Johnson, R.E., Johnson, N. , *Survival models and data analysis*, Wiley&Sons, New York ,**1980**.
- [3] Cox, D.R., Regression models and life tables, *Journal of Royal Statistical Society*, Series B, 34, 187-202, **1972**.
- [4] Collett,D., *Modelling Survival Data in Medical Research*, Chapman &Hall, **1994**.
- [5] Xin, X., *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, **2011**.
- [6] Lee, E.T., Wang, J.W., *Statistical Methods for Survival Data Analysis*, Third Edition, Wiley&Sons, **2003**.
- [7] Breslow, N., Covariance Analysis of Censored Survival Data, *Biometrics*, 30, 89-99, **1974**.
- [8] Efron, B., The Efficiency of Cox's Likelihood Function for Censored Survival Data, *Journal of the American Statistical Association*, 72,557-565,**1977**.
- [9] DeLong, D. M., Guirguis, G. H., and So, Y. C., Efficient Computation of Subset Selection Probabilities with Application to Cox Regression, *Biometrika*, 81, 607–611,**1994**.
- [10] Kalbfleisch, J.D., Prentice, R.L., *The Statistical Analysis of Failure Time Data*, Second Edition, Wiley&Sons, New York, **2002**.
- [11] Kalbfleisch, J.D., Prentice, R.L., Marginal likelihoods based on Cox's regression and life model, *Biometrika*, 60, 267-278, **1973**.
- [12] Scheike, T.H., Sun, Y., Maximum likelihood estimation for tied survival data under Cox regression model via EM-algorithm, *Lifetime Data Anal.*, 13:399–420, **2007**.
- [13] Yay,M., Çoker, E., Uysal, Ö., Yaşam Analizinde Cox Regresyon Modeli ve Artıkların incelenmesi, *Cerrahpaşa Tıp Dergisi*, 38: 139 – 145, **2007**.
- [14] Kleinbaum, D.G., *Survival Analysis: A Self-Learning Text*, Third Edition, Springer, **2012**.
- [15] Arı, A., Önder, H., Farklı veri yapılarında kullanılabilir regresyon yöntemleri, *Anadolu Tarım Bilim. Derg.* , 28(3):168-174, **2013**.
- [16] Therneau, T. M., Grambsch, P. M., *Modeling Survival Data: Extending Cox Model*, Springer, New York, **2000**.
- [17] Ata, N., *Yaşam Çözümlemesinde Orantısız Hazard Modeli*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2005**.
- [18] Terzi, Y., Cengiz, M.A., Bek, Y., Cox Regresyon Modelinde Oransal Hazard Varsayımının Artıklarla İncelenmesi ve Akciğer Kanseri Hastaları Üzerinde Uygulanması, *Türkiye Klinikleri Tıp Bilimleri Dergisi*, 25, 770-775, **2005**.

- [19] Inceoğlu, F., *Sağkalım Analiz Yöntemleri ve Karaciğer Nakli Verileri İle Bir Uygulama*, Yüksek Lisans Tezi, İnönü Üniversitesi Sağlık Bilimleri Enstitüsü, Malatya, **2013**.
- [20] Tamam, D., *Tam ve Sansürlü Örneklem Durumlarında Weibull Dağılımı için Bazı İstatistiki Sonuç Çıkarımları*, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2008**.
- [21] Ersoy, E., *Sağ Kalım Analizinde Parametre Tahmin Problemlerine Katkılar*, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun, **2005**.
- [22] Zhang, M.J., D. P., *Grouped Failure Times Tied Failure Times, Two Contributions To The Encyclopedia of Biostatistics*, Technical Report 24, February **1997**.
- [23] Picciotto, I.H., Rockhill, B., Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression. Department of Epidemiology, *Biometrics*, 53, 1151-1156, **1997**.
- [24] Chalita, L. V. A.S., Colosimo, E.A., Demetrio C.G.B., Reliability and survival analysis, Likelihood approximation and discrete models for tied survival data, *Commun. Statist. - Theory Meth.*, 31(7), 1215-1229, **2002**.
- [25] Allison, P.D., Survival analysis, The Reviewer's Guide to Quantitative Methods in the Social Sciences, edited by Gregory R. Hancock and Ralph O. Mueller, New York, 413-425, **2010**.
- [26] Anonim, Cox Proportional Hazards Regression Models (cont'd), <http://www4.stat.ncsu.edu/~dzhang2/st745/chap7.pdf> (Aralık, **2014**).
- [27] Anonim, Tied Event Times, <http://sites.harvard.edu/fs/docs/icb.topic158134.files/2008labs/bio233labTies.pdf>
- [28] Anonim, <http://www.medicine.mcgill.ca/epidemiology/hanley/c681/cox/TiesCoxModelR.txt> (Aralık, **2014**).
- [29] Semi-Parametric Duration Models: The Cox Model, <https://files.nyu.edu/mrg217/public/cox.pdf> (Aralık, **2014**).
- [30] Stefanescu, C., Mehrotra, D., Cox model versus generalized logrank test for time to event data with ties, <http://faculty.london.edu/cstefanescu/blr.pdf> (Ekim, **2014**).
- [31] Eroğlu A., Altınok M., Özgen K., Sertkaya D., A Multivariate Analysis of Clinical and Pathological Variables in Survival After Resection of Gastric Cancer, *Türkiye Klinikleri Medical Research*, 15, 1, 15-20, **1997**.
- [32] Schoenfeld, D., Partial residuals for the proportional hazards regression model, *Biometrika*, 69, 239-241, **1982**.

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Sena KESKİN
Doğum Yeri : Çankaya
Medeni Hali : Bekar
E-posta : sena.keskin@osym.gov.tr
Adresi : ÖSYM 06800 Bilkent / ANKARA

Eğitim

Lise : 2004 - 2008 Hacı Ömer Tarman Anadolu Lisesi
Lisans : 2008 – 2012 Hacettepe Üniversitesi İstatistik Bölümü
Lisans : 2009 – 2013 Anadolu Üniversitesi İşletme Bölümü
Yüksek Lisans : 2012 – 2015 Hacettepe Üniversitesi İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce, İyi

İş Deneyimi

2013 - : ÖSYM Uzman Yardımcısı

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar