

T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**COX REGRESYON ANALİZİ İLE BÖBREK YETMEZLİĞİ VERİLERİ
ÜZERİNE BİR UYGULAMA**

YÜKSEK LİSANS TEZİ

Emel ÖZER

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Yrd. Doç. Dr. Meral YAY

MAYIS 2017

T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**COX REGRESYON ANALİZİ İLE BÖBREK YETMEZLİĞİ VERİLERİ
ÜZERİNE BİR UYGULAMA**

YÜKSEK LİSANS TEZİ

Emel ÖZER

İstatistik Anabilim Dalı

İstatistik Programı

Tez Danışmanı: Yrd. Doç. Dr. Meral YAY

MAYIS 2017

KABUL VE ONAY SAYFASI

Emel ÖZER tarafından hazırlanan **COX REGRESYON ANALİZİ İLE BÖBREK YETMEZLİĞİ VERİLERİ ÜZERİNE BİR UYGULAMA** adlı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

YRD.DOC.DR. MERAL YAY

Tez Yöneticisi



Bu çalışma, jürimiz tarafından İSTATİSTİK Anabilim Dalında

YÜKSEK LİSANS tezi olarak kabul edilmiştir.

Başkan: : YRD.DOC.DR. MERAL YAY



Üye : Prof.Dr. Nalan CİNEURE



Üye : Prof.Dr. Ünal H. ÖZDEN



Üye : _____

Üye : _____

Bu tez, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

ÖNSÖZ

20. yüzyılın ikinci yarısından sonra tıp başta olmak üzere fen ve sosyal bilimlerde yapılan araştırmalarda sıkça kullanılan yaşam analizi yöntemleri, sağlık alanında önemli bir yer edinmiştir. İnsan sağlığının ne denli önemli olduğu düşünülürse, yaşam analizi yardımıyla tedavi süresi ve yaşam süresinin tahmin edilmesinin de önemi ortaya çıkmaktadır. Bu alanda yapılan çalışmalar aynı yapıya sahip diğer vakalar için de temel bir kaynak niteliğini taşımaktadır.

Bu tez çalışması, prognostik faktörlerin (bağımsız değişkenlerin) yaşam süresi üzerindeki olumlu ya da olumsuz etkileri incelemeyi amaçlayan Cox regresyon modeli üzerine kurulmuş olup, çalışma kapsamında diğer yöntemlere de yer verilecektir. Böbrek yetmezliği verileri üzerinde Cox orantılı tehlikeler modeli kurulmuş ve parametre tahmini elde edilmiştir.

Tez çalışmamda bilgi ve tecrübesiyle beni en iyi şekilde yönlendiren, eleştirileriyle doğru bilgiye ulaşmam için her zaman yol gösteren tez danışmanım sayın Yrd. Doç. Dr. Meral YAY hocama sabrı ve ilgisi için sonsuz teşekkürlerimi sunuyorum. Tez seçimim konusunda beni yönlendiren ve bu süreçte manevi desteğini asla esirgemeyen sayın hocam Prof. Dr. Nalan CİNEMRE'ye, beni yüksek lisans eğitimi konusunda yönlendiren, manevi desteğini asla esirgemeyen eşime ve aileme sonsuz teşekkürlerimi sunuyorum.

Mayıs 2017

Emel Kater ÖZER

İÇİNDEKİLER

ÖNSÖZ.....	I
ŞEKİL LİSTESİ	IV
ÇİZELGE LİSTESİ.....	VI
ÖZET.....	VII
ABSTRACT.....	VIII
1. GİRİŞ.....	1
2. YAŞAM ANALİZİ.....	4
2.1. YAŞAM ANALİZİNDE KULLANILAN GENEL KAVRAMLAR	7
2.2. YAŞAM ANALİZİNDE SANSÜRLEME	8
2.2.1. Planlanmış Sansürleme	10
2.2.2. Planlanmamış Sansürleme	12
2.3. YAŞAM ANALİZİ FONKSİYONLARI	16
2.3.1. Olasılık Yoğunluk Fonksiyonu	16
2.3.2. Yaşam Fonksiyonu	18
2.3.3. Tehlike Fonksiyonu	20
2.3.4. Birikimli Tehlike Fonksiyonu	23
2.3.5. Yaşam Fonksiyonlarının Birbirleriyle İlişkileri.....	24
2.4. YAŞAM ANALİZİNDE KULLANILAN OLASILIK DAĞILIMLARI	26
2.4.1. Üstel Dağılım	27
2.4.2. Weibull Dağılımı.....	29
2.4.3. Log-Normal Dağılım	32
2.5. YAŞAM ANALİZİNDE KULLANILAN YÖNTEMLER	35
2.5.1. Yaşam Tablosu Yöntemi	36
2.5.2. Kaplan-Meier Yöntemi	37
3. COX REGRESYON MODELİ	38

3.1. COX ORANTILI TEHLİKELER MODELİNİN YAPISI.....	38
3.2. OLABİLİRLİK FONKSİYONU.....	39
3.3. β KATSAYILARININ ANLAMLILIĞININ TEST EDİLMESİ	42
3.3.1. Wald Testi.....	42
3.3.2. Olabilirlik Oranı Testi.....	43
3.3.3. Score Testi.....	44
3.4. COX ORANTILI TEHLİKE MODELİNİN VARSAYIMI	45
3.4.1. Grafiksel Yöntemler	46
3.4.2. Uyum İyiliği Testi.....	52
3.4.3. Zamana Bağlı Bağımsız Değişkenlerin Kullanılması	52
3.5. COX REGRESYON MODELİNİN ZAMANA BAĞLI DEĞİŞKENLER İÇİN GENİŞLETİLMESİ	53
4. UYGULAMA.....	57
4.1. YAŞAM TABLOSU YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ	57
4.2. KAPLAN MEIER YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ	61
4.3. COX REGRESYON YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ.....	62
5. SONUÇ.....	65
KAYNAKLAR.....	66

ŞEKİL LİSTESİ

Şekil 2.1. Genel İzleme Süreci.....	6
Şekil 2.2. Sansürsüz Veri.....	9
Şekil 2.3. Sansürlü Veri.....	10
Şekil 2.4. II. Tip Sansürleme	12
Şekil 2.5. Sağdan Sansürlü Veri	13
Şekil 2.6. Rasgele Sansürlü Veri.....	14
Şekil 2.7. Soldan Sansürlü Veri.....	16
Şekil 2.8. Olasılık Yoğunluk Fonksiyonu	17
Şekil 2.9. Teorik Yaşam Fonksiyonu	18
Şekil 2.10. Gerçek Yaşam Fonksiyonu	19
Şekil 2.11. Sabit Tehlike Fonksiyonu	21
Şekil 2.12. Artan Tehlike Fonksiyonu	22
Şekil 2.13. Azalan Tehlike Fonksiyonu.....	22
Şekil 2.14. Önce Artan Sonra Azalan Tehlike Fonksiyonu	23
Şekil 2.15. Üstel Dağılımın Olasılık Yoğunluk Fonksiyonu	28
Şekil 2.16. Üstel Dağılımın Yaşam Fonksiyonu.....	28
Şekil 2.17. Üstel Dağılımın Tehlike Fonksiyonu	29
Şekil 2.18. $\lambda = 1$ için Weibull Dağılımının Olasılık Yoğunluk Fonksiyonu	30
Şekil 2.19. $\lambda = 1$ için Weibull Dağılımının $\log_e S(t)$ Fonksiyonu	31
Şekil 2.20. $\lambda = 1$ için Weibull Dağılımının Tehlike Fonksiyonu	32
Şekil 2.21. $\mu = 0$ için Log-normal Dağılımının Olasılık Yoğunluk Fonksiyonu.....	33
Şekil 2.22. $\sigma^2 = 0.5$ için Log-normal Dağılımının Olasılık Yoğunluk Fonksiyonu	34
Şekil 2.23. Farklı Parametreler için Log-normal Dağılımının Tehlike Fonksiyonu	35
Şekil 3.1. $-\ln(-\ln \hat{S})$ Eğrisinin Elde Edilmesi.....	48
Şekil 3.2. İki Bireye Ait Log-log Yaşam Eğrileri	49

Şekil 3.3. Gözlenen ve Beklenen Yaşam Eğrileri	51
Şekil 4.1. Farklı Tedavi Türleri için Yaşam Fonksiyonu	60
Şekil 4.2. Kümülatif Yaşam Sürelerine Göre Yaşam Süreleri Tahminleri.....	62
Şekil 4.3. Cox Regresyon Eğrisi.....	64

ÇİZELGE LİSTESİ

Çizelge 4.1. Yaşam Tablosu Sonuçları.....	58
Çizelge 4.2. Wilcoxon Testi	58
Çizelge 4.3. Medyan Yaşam Süreleri.....	59
Çizelge 4.4. Kaplan Meier Tahminleri	61
Çizelge 4.5. Log-Rank Testi	61
Çizelge 4.6. -2 Log Likelihood Testi	63
Çizelge 4.7. Cox Regresyon Sonuçları.....	63

ÖZET

Bu tez çalışmasında, yaşam analizi yöntemlerinden Cox Regresyonun, böbrek yetmezliği sorunu yaşayan hastalardan elde edilen gerçek verilerle analiz edilmesi ve sonuçların değerlendirilmesi amaçlanmıştır.

Dört bölümden oluşan çalışmanın giriş bölümünde, yaşam analizinin ortaya çıkış amacı açıklanarak literatür taraması yardımıyla kullanım alanı açısından değerlendirilmesi yapılmıştır. İkinci bölümde analizin daha iyi kavranabilmesi için gereken kavramlar, fonksiyonlar ve dağılımlar açıklanmış, ayrıca analiz için büyük önem taşıyan sansürleme üzerinde durulmuştur. Bu bölümde ayrıca, Yaşam analizi yöntemlerinden “Yaşam Tablosu” ve “Kaplan Meier” yönteminden de bahsedilerek, yöntemler arasındaki farkın önemi vurgulanmıştır. Üçüncü bölümde Cox Regresyon modelinin yapısı, varsayımları ve tahmin yöntemleri detaylı bir şekilde incelenmiştir. Uygulama bölümünde ise böbrek yetmezliği sorunu yaşayan hastalardan elde edilen gerçek veriler üzerine Cox regresyon yöntemi uygulanarak analiz yapılmıştır. Tedavi yöntemi seçiminde etkili olan kriterler belirlenerek bağımsız değişkenlerin hastalığın tedavisi üzerindeki etkileri araştırılmıştır. Ayrıca, karşılaştırma yapmak amacıyla yaşam tabloları ve Kaplan Meier yöntemlerinin özet şeklinde analizleri yapılmıştır.

Anahtar Kelimeler : Cox Regresyon, yaşam analizi, Cox Regresyon modelinin zamana bağlı değişkenler için genişletilmesi, böbrek yetmezliği

ABSTRACT

The aim of this thesis, to apply Cox regression, which is one of the methods of survival analysis, on real data from patients who have renal failure and to discuss the results.

In the four-part introductory chapter, the reasons underlying the emergence of survival analysis are explained and this analysis is later evaluated with respect to its field of use via a literature survey. In the second chapter, in order to better understand the analysis, the main concepts, functions and distributions are explained. Besides, censoring which is one of the most important concepts of the analysis is also elaborated. In this section, "Life Tables" and "Kaplan-Meier" techniques which are the one of the techniques of Survival Analysis are mentioned and the differences of them are emphasized. In the third chapter, the structure of Cox regression, the assumptions and estimation methods are examined in detail. In the application chapter, Cox regression is applied on the real data from patients who have renal failure. Upon the determination of effective criteria with respect to the selection of treatment method, the effects of independent variables on the treatment have been specified. In addition, life tables and Kaplan-Meier techniques applications are also carried out for the purpose of comparison.

Keywords: Cox regression, survival analysis, extensions of Cox Regresyon, kidney failure

1. GİRİŞ

Yaşam analizi (survival analysis), bir araştırmada asıl ilgilenilen değişken olan bağımlı değişkenin bir olay gerçekleşene kadar geçen zaman olarak tanımlandığı ve bu değişkeni açıklamak amacıyla uygulanacak istatistiksel yöntemlerin bütünü olarak ifade edilebilir.

“Yaşam analizi” teriminin ortaya çıkışı, asıl olarak biyomedikal araştırmacılardan kaynaklanmaktadır. Bunun nedeni, sözkonusu yöntemin temellerinin ölüm oranını (mortality) ya da hastalara belirli bir hastalık için konan tanı ile ölüm zamanları arasındaki yaşam sürelerini araştıran biyomedikal araştırmalara dayanmasıdır. Yaşam analizinin ilk kullanımı bir İngiliz olan John Graunt tarafından, ilk yaşam tablosunun temellerinin atılması ve “Ölüm Faturalarına Göre Gerçekleştirilen Doğal ve Politik gözlemler (Natural and Political Observations Made Upon the Bills of Mortality)” isimli oldukça popüler olan makalesinin 1662 yılında basılmasıyla ortaya çıkmıştır. Günümüzde biyoistatistik çerçevesinde sayılabilecek olan bu çalışmasında Graunt, doğum, evlilik ile ölüm sürecinin kendi doğasında var olan riski araştırmış ve ölüm istatistiklerini (mortality statistics) hesaplayarak bir hastalığı bir başka hastalık ve bir yılı başka bir yıllla karşılaştırmak amacıyla Londra’nın kenar kentlerindeki ölüm nedenlerini ve sayılarını haftalık olarak raporlar yazarak bunlar üzerinde çalışmıştır (Everitt, 2005a). Edmund Halley, 1687-1691 yılları arasında, günümüz de kullanılan demografik ve aktüerya çalışmalarına çok benzeyen ilk yaşama tablosunu tasarlamıştır (Marubini & Valsecchi, 1995).

Greenwood (1926), yaşam olasılıklarının standart hatasının formüllerini vermiştir. II. Dünya savaşı yıllarında askeri malzemelerin dayanıklılığı ve güvenilirliğinin test edilmesinde sağkalım analizi kullanılmış, daha sonraki dönemlerde tüketicilerin daha dayanıklı ve güvenilir ürün talebini karşılamak amacıyla da uygulamada yer bulmuştur.

20. yüzyılın ikinci yarısından sonra, tıp başta olmak üzere fen ve sosyal bilimlerde yapılan araştırmalarda yaşam analizi çok daha fazla kullanılmaya başlanmıştır. Tıp alanında, Frei'nin (1961) belirli bir hastalığı olan ve tedavi edilen kişilerin tedaviden sonra iyileşme sürelerinin incelenmesine ilişkin çalışması; MacDonald (1963), bir hastalığa yakalananları belirli bir süre izleyerek yaşam sürelerinin ölçülmesine ilişkin çalışması örnek olarak gösterilebilir.

Sosyal bilimler alanında ise 1988 yılında Fichman, maden ocağında çalışan kömür işçilerinin çalıştığı günlerin kayıtlarına ilişkin verileri kullanarak, işe devam etmenin motivasyonel sonuçlarını, yine aynı yılda Lehter, evlilikteki anlaşmazlıklara neden olan faktörlerin etkilerini ve 1997 yılında Albonetti ve Hepburn, şartlı olarak serbest bırakılan 617 suçlunun ortalama şartlı serbest bırakılma süresini etkileyen faktörleri yaşam analizi kullanarak incelemişlerdir (Göz Çekçeki, 2007).

1972 yılında David R. Cox tarafından geliştirilen orantılı tehlikeler modeli ve Kaplan-Meier tarafından geliştirilen yaşam fonksiyonlarının tahmini için çarpım limit yöntemi de analizin geliştirilmesine öncülük eden bilimsel çalışmalardır. 1980 yılında Kalbfleisch ve Prentice, Cox tarafından geliştirilen orantılı tehlikeler modeli üzerine çalışmalar yaparak analizi daha da ayrıntılı bir şekilde incelemiştir.

Yaşam analizi, hem sosyal hem de doğal bilimlerde birçok farklı olayı incelemek için kullanışlı bir yöntemdir. Örneğin, makina parçalarının bozulması, hastalığın başlangıcı, depremler, trafik kazaları, borsada hisse senetlerinin düşmesi, doğum, ölüm, boşanma, terfi, emeklilik, işten çıkarma gibi olayları incelemek için yaşam analizi yöntemleri kullanılır (Karasoy & Ata Tutkun, 2016). Yaşam analizi yöntemleri, farklı disiplinlerde farklı isimler ile anılır. Örneğin, sosyoloji'de "olay geçmişi analizi (event history analysis)" (Tuma & Hannan, 1984; Yamaguchi, 1991), ekonomide "süreklilik analizi ya da geçiş analizi (duration analysis or transition analysis)" (Greene, 2003; Heckman & Singer, 1985), mühendislikte "yaşam süresi ya da başarısızlık süresi analizi (lifetime or failure-time analysis)" (Lawless, 1982) ve biyoistatistikte "yaşam analizi" ya da "sağkalım analizi (survival analysis)" olarak anılmaktadır (Guo, 2010). Bu çalışma kapsamında en çok kullanılan isim olan "yaşam analizi" ismi kullanılacaktır.

Yaşam analizinde bağımlı değişkenin normal dağılmaması, bağımsız değişkenlerin ardışık bağımlı olması ve aralarında orantısal ilişkiler sonucu korelasyon olması, en çok kullanılan klasik istatistiksel yöntemlerden biri olan klasik regresyon yönteminin kullanımını engellemektedir. Bunun yanısıra bağımlı ve bağımsız değişkenlerin kategorik yapıda olması, klasik regresyon yerine daha anlamlı ve güvenilir sonuçlar verecek farklı regresyon yöntemleri kullanılmasını gerektirir (Özşen, 2006). Ayrıca yaşam analizinde kullanılan sansürlü ya da durdurulmuş (censored) değişkenlerin normal dağılıma uymaması ve yaşam sürelerinin durdurulmuş olması da klasik yöntemlerden uzaklaşmaya sebep olmaktadır. Dolayısıyla parametrik yöntemler yerine parametrik olmayan yöntemlerin kullanılması gerekir. Örneğin yaşam analizine konu olan hastalığa ilişkin veride, değişkenlerin analize giriş sıralamaları değişebilir. Bunun yanı sıra tüm değişkenler aynı anda analize katılmayabilir ya da analiz başladıktan sonra yeni değişkenler eklenebilir. Kısacası yaşam analizinde kullanılacak yöntemlerin, kullanılan değişkenlerin yapısına ve içeriğine göre belirlenmesi kaçınılmaz olur. Bu bağlamda bu analizde ilgilenilen problemin çeşidine ve veri yapısına göre başvurulabilecek çok sayıda yöntem mevcuttur. En sık kullanılan yöntemler arasında Yaşam Tablosu Yöntemi, yaşam fonksiyonunun tahmininde kullanılan Kaplan-Meier Yöntemi, iki yaşam dağılımı arasında karşılaştırma yapan Long-Rank Yöntemi ve bağımsız değişkenlerin yaşam süresi üzerindeki etkilerini ölçmede kullanılan Cox Regresyon Yöntemi sayılabilir.

Bu tez çalışması, prognostik faktörlerin (bağımsız değişkenlerin) yaşam süresi üzerindeki olumlu ya da olumsuz etkileri incelemeyi amaçlayan Cox regresyon modeli üzerine kurulmuş olup, çalışma kapsamında diğer yöntemlere de kısaca yer verilecektir. İkinci bölüm olan “Yaşam Analizi” başlıklı bölümde, analiz tüm detayları ile irdelenerek analizde kullanılan temel kavramlar, olasılık dağılımları ve sansürleme üzerinde ayrıntısı ile durulacaktır. Üçüncü bölüm olan “Cox Regresyon Modeli” başlıklı bölümde, Cox orantılı tehlikeler modelinin yapısı tanıtılacak ve model tüm detayları ile ele alınacaktır. Dördüncü bölüm olan “Uygulama” başlıklı bölümde ise, böbrek yetmezliği verileri üzerinde Cox orantılı tehlikeler modeli kurulacak ve parametre tahmini elde edilecektir.

2. YAŞAM ANALİZİ

Biyoistatistik, canlıların özelliklerinin ve bu özellikler üzerinde etkili olan faktörlerin belirlenmesi, sayılması, düzenlenmesi ve değerlendirilmesi amacıyla istatistiksel yöntemlere dayalı bir bilim dalıdır. İnsan ve toplum sağlığını iyileştirici çözümler bulmaya yönelik yöntemler bütünü olan biyoistatistiğin en önemli konularından biri de yaşam analizidir. Belirli bir zaman periyodunda incelenen bireyler analize dahil edilerek istatistiksel yöntemler yardımıyla incelenir ve sonuçlar yorumlanarak daha sonraki çalışmalarda kullanılır. Analizde gözlenen bireyler izleme süresi sona ermeden ulaşamaz hale gelebilir, farklı sebeplerden dolayı çalışmadan ayrılabilir veya çalışma sona ermeden ölebilirler, bir başka deyişle izlenemez hale gelebilirler. Bu tür bireylere ilişkin veriye sansürlü ya da durdulmuş veri (censored data) denir. Yaşam analizinin en önemli özelliklerinden biri sansürlü veriyi analize dahil edebilmesidir. Herhangi bir hastalığa maruz kalmış bireye uygulanan tedavi girişiminden sonra daha ne kadar yaşayabileceğini tahmin etmede ve yaşam süresine etki eden faktörlerin belirlenmesinde kullanılan yaşam analizinin amaçlarını üç maddede toplamak mümkündür:

- Tedaviye maruz kalan hastalara ilişkin beklenen yaşam süresinin tahmin edilmesi
- Yaşam süresine etki eden faktörlerin belirlenmesi
- Tedavi edilen ve edilmeyen hastaların yaşam sürelerinin karşılaştırılması

Hastalık verileri incelendiğinde, her hastalığın farklı semptomları olduğu ve bu semptomların kişilerde farklı şiddet ve biçimde ortaya çıkabildiği gözlenmektedir. Bir hastalığa yakalanan kişinin iyileşme olasılığı, ölüm olasılığı ya da ne zaman öleceğine ilişkin olasılık istatistiksel yöntemler kullanılarak tahmin etmek mümkündür.

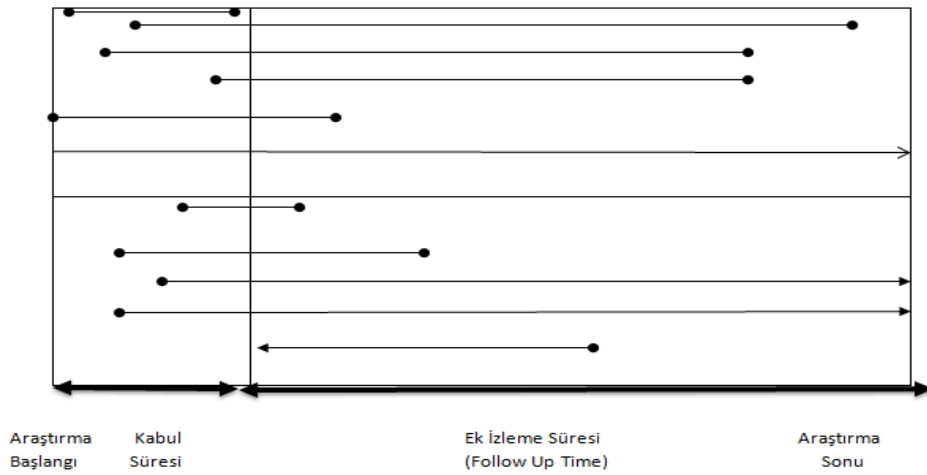
Bu noktada hastalığa yakalanan bireye, uygulanan tedavi türü ve izleme (follow up) zamanına ilişkin değişimlerin yaşam süresi üzerine etkilerini ortaya koymak gerekmektedir.

Yaşam analizi ile teşhisi konulan hastalıklara uygulanacak olan tedavi tiplerinin belirlenmesinden sonra, bu tedavilerin hastalık sürecine sağladığı olumlu yada olumsuz etkileri, hastanın yaşam süresi üzerindeki etkilerini, tedavi ya da uygulanan bir girişimden sonra (ameliyat, nakil) hastanın daha ne kadar yaşayabileceği, fizyolojik değerlerin (kan ölçümleri, tansiyonu, yaşı) yaşam süresi üzerindeki etkilerini incelemek mümkündür.

Yaşam analizinde sıklıkla kullanılan izleme terimi, araştırma için önemli olan özelliklerin çalışmanın başından sonuna kadar gözlemlenmesini ifade eder. Ancak izlenen özelliğin çalışmanın sonuna kadar gözlemlenebileceği garanti edilemez. Bir başka deyişle izlenen özellik araştırma sonlanamadan izlenemez hale gelebilir. Örneğin kanser ameliyatı olan hastaların, ameliyat tarihinden itibaren ne kadar sürede iyileşeceği üzerine yapılan bir çalışmada, araştırma süresi sona ermeden hasta ölebilir, taburcu olabilir ya da herhangi bir nedenle gözlemden ayrılabilir. Sonuç olarak izleme terimi, çözümlenecek çalışmada belirli bir özelliğin, konunun süreç içinde gözlemlendiğini ya da saptandığını vurgulamaktadır. Bu tür araştırmaların genel amacı belirli bir özelliğin ortaya çıkışını, zamanın da fonksiyonu olarak ortaya koyup değerlendirebilmektir. Yaşam analizi ile eşdeğer olarak kullanılan sağkalım terimi aslında geniş anlamda kullanılan ve gerçekte konunun uygulanacağı tüm örneklerdeki durumu yansıtmayan bir terimdir. Dolayısıyla araştırmalarda “sağkalmak (yaşamak) / ölmek” çiftli seçeneği bulunması zorunlu değildir. “İzleme”den anlaşılan, belirli bir başlangıç noktasından sonra (bir ameliyat; bir tedavinin başlangıcı, bir hastalığın başladığı veya tanısının konulduğu zaman v.b.), bir izlem süresi içindeki olgunun, araştırmanın ana konusu olan özel bir konuma erişmesi (veya erişmeden eski özelliklerini sürdürmesi) durumudur. Burada özel konum ile kastedilen ölüm olabileceği gibi, tedaviye yanıt vermek, tümörsüz geçen süre gibi diğer özellikler de olabilir (Şenocak, 2009).

İzleme özelliği taşıyan araştırmalar genellikle uzun süreli araştırmalardır, bu nedenle bu çalışmalar için önemli olan zaman dilimlerinin doğru bir şekilde ayrıştırılmasıdır. İzleme

özelliği olan araştırmalarda ilk aşamada bireylerin çalışmaya dahil edildiği bir kabul dönemi mevcuttur. Örneğin, kanser hastaları için yapılan bir çalışmada, çalışma için belirlenen başlangıç tarihinden itibaren iki ay içerisinde kanser ameliyatı olan bütün hastalar bu çalışma için birer veri özelliğini taşır. Ancak kabul dönemi olarak belirlenen ikinci aydan sonra ameliyat olan hastalar veri olarak kabul edilemez. İzleme süresi belirlenirken, örnekleme yöntemine tabi olacak büyüklükte verinin toplanabilmesine imkan tanıyacak uzunlukta bir dönemi kapsamasına dikkat edilmelidir. Kabul süresinin bitiminde, araştırılan konunun klinik özelliklerine göre ek izlem süresi tanımlanabilir. Ek izlem süresinin çok kısa tutulması, karşılaşılmaması beklenen olay sıklığının doğal olmayan bir şekilde düşük, uzun tutulması ise araştırma maliyetinin gereğinden fazla olması sonuçlarını doğurabilir. Bu durum çalışmayla ilgisi olmayan başka değişkenlerin araştırmaya dahil edilerek araştırma amacının sapmasına ve beklenen sonuçların elde edilememesine de neden olabilir. Şekil 2.1.'de izleme sürecinin genel yapısı hakkında görsel bir yaklaşım verilmiştir.



Şekil 2.1. Genel İzleme Süreci

Araştırma izlenirken dikkat edilmesi gereken bir nokta, izlenen tüm bireylerin gözlenen olay ile karşılaşmayabileceğidir. Bir başka deyişle, yaşam analizinde çalışmaya katılan bireyler belirli bir zaman periyoduna kadar izlenir. Ancak izlem süresi sona ermeden bireylerin kaybolmaları, farklı bir sebepten dolayı çalışmadan çıkmaları ya da çalışma

sona ermeden ölmeleri söz konusu olabilir. Yaşam analizinde kullanılan modellemenin temelini oluşturan sansürlü veriler de çalışmadan çıkan ya da kaybolan verilerdir.

Yaşam analizinde sıklıkla kullanılan bir başka terim ise “başarısızlık zamanı” (failure time) dır. Yaşam analizi, başarısızlık olarak adlandırılan bir olayın meydana gelmesine kadar geçen sürede elde edilen verilerin analizidir. Bir başka deyişle “başarısızlık” terimi, incelenen konunun denekte görülmesi durumudur. Canlılar için genelde ölüm veya hastalık, mekanik aletler için ise bozulma anlamına gelir (Nelson 1982). Analizde temel olan, gözlenen başarısızlık sürelerinin incelenmesi olduğu için bu değişkenin duyarlı olarak ölçülmesi gerekmektedir. İlgilenilen olaya göre farklı şekillerde ortaya çıkabilen bu değişkenin duyarlı olarak ölçülebilmesi için ise; her bir birey için başlangıç zamanı şüpheye yer vermeyecek şekilde tanımlanmalıdır. Geçen süreyi ölçmek için bir zaman ölçeği kabul edilmeli ve her birey için ömrün sona erdiği ya da başarısızlığın meydana geldiği an net olarak tanımlanmalıdır (Cox, 1984).

2.1. YAŞAM ANALİZİNDE KULLANILAN GENEL KAVRAMLAR

Yaşam analizinde sıkça kullanılan bazı kavramların açıklanmasının, konunun daha iyi anlaşılabilmesi açısından faydalı olacağı düşünülmüştür. Bu amaçla söz konusu kavramlardan aşağıda kısaca bahsedilmiştir (Özdamar, 2003).

Olgu (Event), yaşam analizinde ilgilenilen olaydır. Örneğin, belirli bir tedavi gördükten sonra takip altında tutulan bir bireyin ölmesi ile tedavi edilen bir hastalığın tekrar görülmesi (nüks) birer olgudur.

Yaşam süresi ya da izlem süresi (survival time), bir bireyin belirli girişime ya da etkene maruz kaldıktan sonra iyileşmesine, hastalığının tekrarlanmasına ya da ölümüne kadar geçen süredir.

Yaşam fonksiyonu (survival function, survival curve), yaşam sürelerinin olasılık dağılımıdır. Fonksiyon, yaşam verilerinin genel eğilimini matematiksel bir modelle ifade eder. Yaşam fonksiyonu bir olasılıktır.

Yaşam yoğunluk fonksiyonu, sağ olan bir kişinin t zamanı içinde daha ne kadar yaşayabileceğini ortaya koyan olasılıklardır. Bu olasılıklar, belirli t_i zamanı için birlikte hesaplanmaktadır.

Anlık ölüm olasılığı ya da hazard fonksiyonu (hazard function, hazard rate), sağ olan bir kişinin belirli zamanda (anda) ölüm olasılığı, taşıdığı ölüm riskidir.

Ölüm yoğunluk fonksiyonu, sağ olan bir kişinin t zamanındaki ölüm riskini ortaya koyan olasılıklardır. Bu olasılıklar, belirli t_i zamanı için birlikte hesaplanmaktadır.

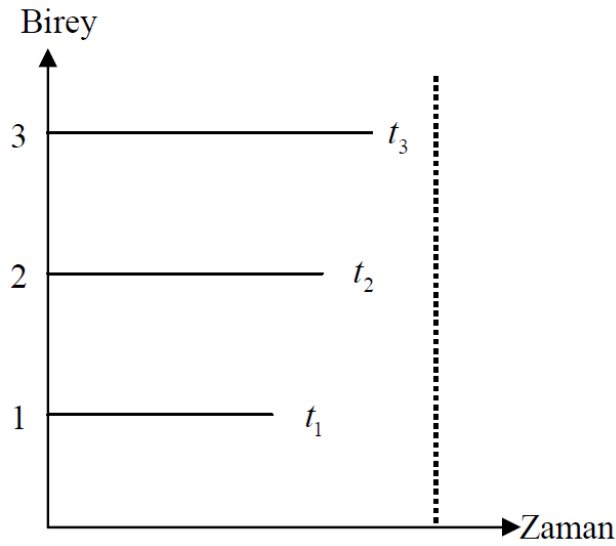
Birikimli ölüm fonksiyonu ya da birikimli hazard fonksiyonu (cumulative hazard function), T zamanı içinde belirli bir t anı için hesaplanmış olan ölüm olasılıklarının birikimli fonksiyonudur.

2.2. YAŞAM ANALİZİNDE SANSÜRLEME

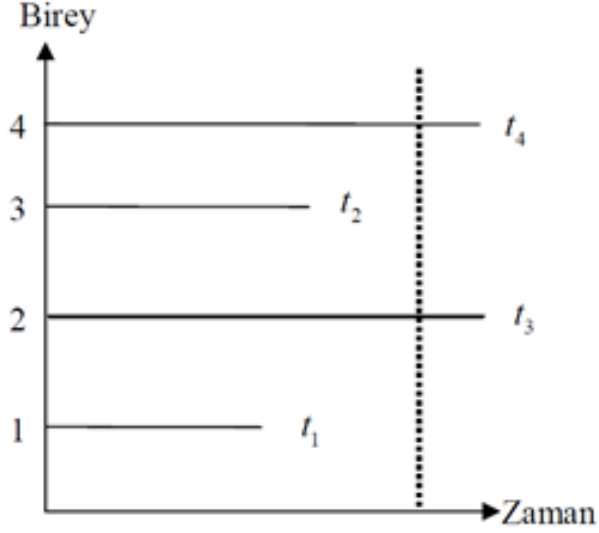
Yaşam analizinde incelenen birimlerin bazılarının başarısızlık zamanları kesin olarak bilinirken bazılarının belirlenemez. Bir başka deyişle, gözlemlene işinin tamamlanamamış olması ile karşılaşılabilir. Herhangi bir sebepten dolayı kendilerinden bir daha haber alınamayan olan birimler sansürlü, durdurulmuş ya da tamamlanmamış gözlemler olarak adlandırılırlar. Veri setlerinde de sıkça karşılaşılabilecek olan bu tür gözlemler için başarısızlık zamanı yerine durdurma zamanları göz önünde bulundurulur. Bunun yanı sıra araştırma sonuna kadar hiç bir sorunla karşılaşmadan izlenen gözlemler de bulunabilir. Bu tür gözlemler sansürsüz, durdurulmamış ya da tamamlanmış (uncensored) gözlemler olarak adlandırılır. Bu durumda veri yapısına bağlı olarak izleme özelliği olan her araştırmada gözlemler, tamamlanmış ve tamamlanmamış olmak üzere ikiye ayrılabilir.

Yaşam modellerinde sansürleme farklı durumlarda ortaya çıkabilir. İlk durum, bireyin izlem süresi içerisinde ölmesidir. Bu durumda bireyin yaşam süresi bilindiği için veri sansürlü değildir. İkinci durumda, birey herhangi bir nedenden dolayı gözlemden çıkabilir, araştırma konusu dışında bir sebepten dolayı ölebilir ya da uygulanan başka bir yöntemden farklı bir sonuç alabilir. Bu durumda bireyin yaşam süresi, çalışmadan

çekildiği andan itibaren sansürlüdür. Üçüncü durumda, birey çalışma sonlandığı halde hala yaşıyor olabilir. Bu durumda da, bireyin yaşam süresi çalışmanın sonlandığı zamana kadar bilinmesine rağmen, sonrası için herhangi bir bilgi edinilemediğinden dolayı sansürlüdür (Kleinbaum,1996). Sansürlü veriler, eksik veriler olmasına rağmen yine de gözlem hakkında kesin olmasa da bazı önemli bilgiler verdiği için gözardı edilmeden analiz içinde kullanılmaktadır (Terzi, 2003). Buradan da anlaşıldığı üzere, yaşam analizinin temel taşı verinin yapısı ya da daha açık bir ifadeyle sansürlü verinin analizidir. Şekil 2.2.'de sansürlü Şekil 2.3.'de sansürlü veri türlerine göre gözlemler incelenebilir (Nelson, 1982).



Şekil 2.2. Sansürlü Veri



Şekil 2.3. Sansürlü Veri

Şekil 2.2.'de sansürlü veri, olması beklenen sonucun araştırma periyodu içerisinde gerçekleşmesini, Şekil 2.3.'de ise sansürlü veri, sonucun araştırma periyodu dışında gerçekleştiğini göstermektedir.

İncelenen olayın yapısına ve elde edilen bilginin türüne bağlı olarak birçok sansürleme tipinden bahsedilebilir. Sansür tiplerinin sınıflandırılmasında farklı görüşler olmakla birlikte, en uygun sınıflandırma sansürlemeyi planlanmış sansürleme ve planlanmamış sansürleme olmak üzere iki ana grupta incelemektir (Başar, 1993).

2.2.1. Planlanmış Sansürleme

Planlanmış sansürleme, araştırmacının belirlediği kriterlere göre sonlandırılan çalışmalardır ve araştırmacının kararı doğrultusunda gerçekleştirilir. Planlanmış sansürlemede önemli olan nokta, çalışma daha başlamadan sansürleme zamanını veya sansürleme sayısının belirlenmiş olmasıdır. Planlanmış sansürleme, I. tip sansürleme (Type I censoring) ve II. tip sansürleme (Type II censoring) olmak üzere iki başlık altında incelenebilir. I. tip sansürleme, çalışma ile ilgili zamanın sansürlenmesini, II. tip sansürleme ise çalışmadaki birim ya da bireylere ilişkin başarısızlık sayısının sansürlenmesini esas alır.

I. Tip Sansürleme

Araştırmanın başlangıç ve bitiş süresinin kesin olarak bilindiği sansürleme türüdür. Çalışma süresini esas aldığı için, zamansal sansürleme (time censored) olarak da isimlendirilir. I. tip sansürlemede araştırmaya başlamadan önce bir gözlem süresi belirlenir ve belirlenen gözlem süresi bittiğinde araştırmaya son verilir. I. Tip sansürlemenin hayvanlarla ilgili deney çalışmalarında sıkça kullanıldığı görülmüştür. Örneğin, kanserojen bir maddenin fareler üzerinde kanser oluşumundaki etkisini inceleyen bir çalışmada farelerin yaşam süreleri değerlendirilerek ölçülmüş ve planlanmış sansürleme verisi kullanılmıştır (Terzi, 2003).

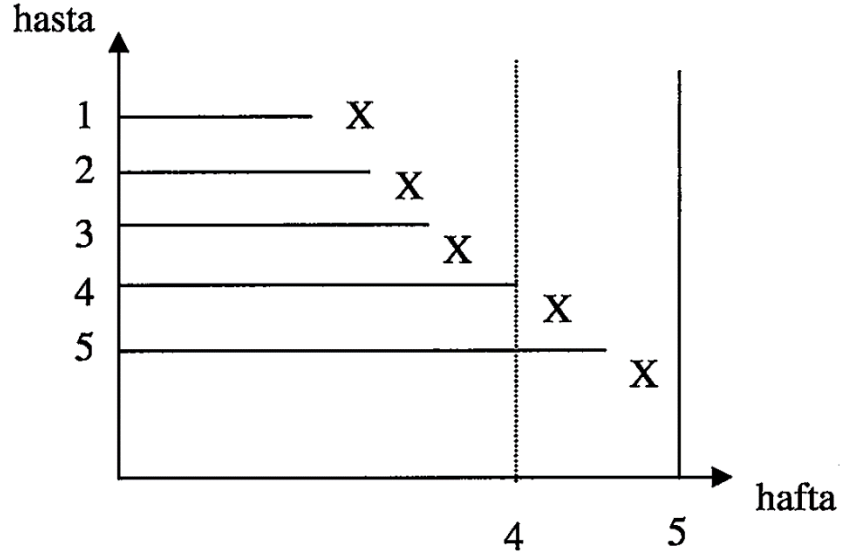
I. Tip sansürlemede $T_1, T_2, \dots, T_n$, birbirinden bağımsız raslantı değişkenleri, $T_i = (i = 1, 2, \dots, n)$ raslantı değişkeni ise i. yaşam süresi olarak tanımlansın. C_i , araştırma öncesinden belirlenen planlı sansürleme zamanını gösterirken, T_i süresinin C_i süresinden büyük veya küçük eşit olması durumunda yeni bir Y_i raslantı değişkeni aşağıdaki gibi tanımlanabilir:

$$Y_i = \begin{cases} 1, & T_i \leq C_i \\ 0, & T_i > C_i \end{cases} \quad (2.1)$$

Y'nin dağılım fonksiyonu pozitif olup $Y_i = 0$ ($T_i > C_i$) ise birey sansürlenmiş, $Y_i = 1$ ($T_i \leq C_i$) ise birey gözlemlenmiştir (Nelson, 1982).

II. Tip Sansürleme

Araştırmada, önceden planlanan sayıda başarısızlık meydana geldiği anda çalışmanın sona erdirildiği bir sansürleme kuralıdır. Başarısızlık sayısını baz aldığı için sayısal durdurma (failure censoring) olarak da isimlendirilir (Başar, 1993). II. Tip sansürlemenin I. Ti p sansürlemeden farkı, araştırma başlamadan önce başarısızlık sayısının belirlenmiş olmasıdır. Başarısızlık sayısına ulaşıldığı anda çalışmaya son verilir.



Şekil 2.4. II. Tip Sansürleme

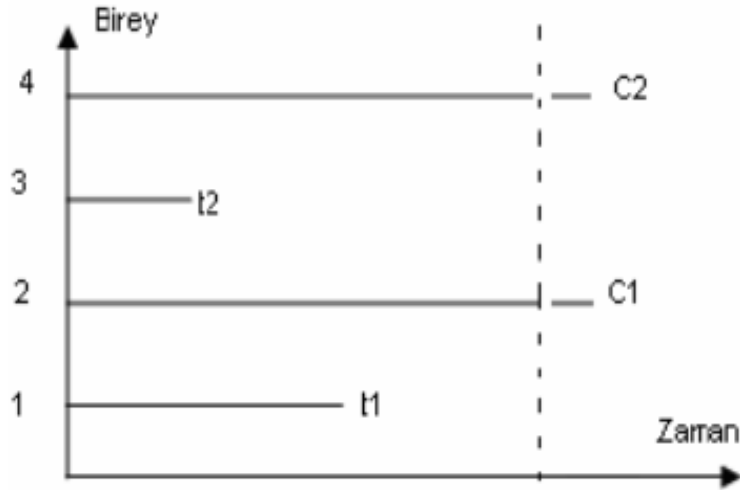
Şekil 2.4. II. Tip sansürleme çalışmasına bir örnek olarak verilebilir. Bu örnekte 4. haftada çalışma durdurulur. Bunun nedeni başlangıçta belirlenen %80 başarısızlık (ölüm) oranının (4/5) ile 4. haftada elde edilmesidir. Bu durumda çalışma sonu beklenmeyip çalışma bu haftada sona erdirilir. Bu durumda beşinci gözlem, II. Tip sansürlü gözlemdir (Terzi, 2003).

2.2.2. Planlanmamış Sansürleme

Adından da anlaşılacağı üzere önceden tahmin edilemeyen bir sebep yüzünden araştırmaya son verilmesi gerektiğinde “planlanmamış” sansürleme söz konusu olur. Planlanmış sansürlemede araştırmacı sansürleme sayısına ya da sansürleme zamanına müdahale edebilirken planlanmamış sansürlemede herhangi bir müdahale söz konusu değildir. Çünkü bir anda ve beklenmeyen bir şekilde ortaya çıkar. Çalışmaya dahil olan birimler çalışmanın ortasında ya da sonlarına doğru çalışmadan çıkabilir, kaybolabilir ya da çalışmaya devam etmesi mümkün olmayabilir. Bu durumda araştırmacının müdahalesi dışında gelişen bir durum söz konusudur yani planlanmamış bir şekilde ortaya çıkar. Planlanmamış sansürleme, sağdan sansürleme, rasgele sansürleme, aralıklı sansürleme ve soldan sansürleme olmak üzere dört alt başlık altında incelenebilir.

Sağdan Sansürleme

Sağlık alanında yapılan çalışmalarda en sık karşılaşılan sansürleme türlerinden biridir. Başarısızlık zamanına ilişkin analiz yapılırken araştırmanın başlangıç ve bitiş süresi önceden planlanır. Analize alınan gözlemlerden bazıları araştırma çerçevesinde başarısızlığa uğramış olabilir. Bir başka deyişle bu süre içerisinde analize dahil edilmiş olan gözlemlerden bazıları çalışmanın planlanan süresi içinde gözlemlenemez hale gelmiş olabilir. Bu durumda veri *sağdan sansürlüdür* (right censored) denir. Ulaşılmak istenilen sonuç, çalışma periyodu içerisinde gerçekleşmeyip periyodun sağına geçmiştir. Dolayısıyla, sansürlü yaşam süresi, gerçek bilinmeyen yaşam süresinden daha kısadır (İnceoğlu, 2013). Örneğin, böbrek nakli yapılan hastaların bünyelerinin böbreği kabul edip etmemesi üzerine yapılan bir araştırmada, böbrek naklinin yapıldığı tarih baz alınarak iki yıl süre ile hastalar takip edilmiş ve ikinci yıl sonunda istenen sonuç bünyenin böbreği reddetmesi olarak belirlenmiştir. Çalışmanın bitim süresi olan ikinci yılın sonunda böbreği reddeden gözlemler, istenilen sonuca ulaştığı için sansürlü, halen daha tedavi gören olumlu ya da olumsuz böbreği taşıyan gözlemler ise sağdan sansürlü olarak belirlenir. Sağdan sansürleme Şekil 2.5.'de görüldüğü gibi gözlemlenebilir.

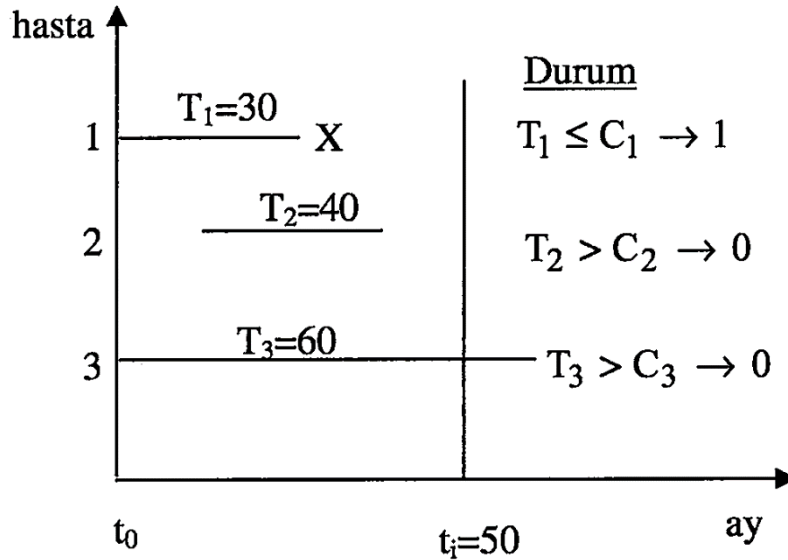


Şekil 2.5. Sağdan Sansürlü Veri

Rasgele Sansürleme

Klinik çalışmalarda, çalışma başlangıcından itibaren istenilen sayıda bireyin çalışmaya katılması genelde mümkün olmaz. Çalışma başladığında istenilen sayıda bireyin bulunması mümkün olmayabilir ve bu çok sık karşılaşılan bir durumdur. Dolayısıyla bazı bireylerin çalışmaya katılma tarihleri farklı olacaktır. Çalışma sürerken bireyler isteyerek ya da istem dışı olarak birçok nedenden dolayı çalışmadan ayrılmak zorunda kalabilir ya da takip süreleri boyunca rasgele kaybolabilirler. Bu nedenle sağlık alanında kullanılan verilerin çoğu *rasgele sansürlü* verilerdir. Yine bireyler herhangi bir nedenden dolayı izlenemediği için ölüm olayı saptanamamışsa bu tip veri de rasgele sansürlü veri olarak adlandırılır (Terzi, 2003; Sümbüloğlu, 2009; Fox, 2006).

Rasgele sansürlemede çalışmanın başlangıç zamanının tüm bireyler için aynı olma zorunluluğu yoktur. Bireyler istedikleri zamanda yani rasgele bir zamanda çalışmaya dahil olabilirler. Ayrıca takip süreleri boyunca da rasgele kaybolabilirler. Rasgele sansürleme, I.Tip ve II.Tip sansürlemenin bir karışımı şeklinde de düşünülebilir (Miller, 1981).



Şekil 2.6. Rasgele Sansürlü Veri

Şekil 2.6.'da farklı zamanlarda tedaviye başlayan hastalara ilişkin bir klinik deney örneği verilmiştir. Bu çalışmada 1. hasta $T_2 = 30$ ay yaşayıp ölmüştür. 1. hasta çalışma

periyodu içerisinde öldüğünden dolayı sansürlü veri olarak değerlendirilir. 2. hasta çalışma periyodu başladıktan sonra araştırmaya dahil olmuştur. Bu hasta $T_2 = 40$. ayda nedeni bilinmeyen bir sebeple kaybolmuştur ve yaşadığına ya da öldüğüne ilişkin bir bilgi bulunmamaktadır. Dolayısıyla bu veri rasgele sansürlü veridir. 3. hasta ise çalışma süresi olan 50 ay tamamlandığı halde yaşamaya devam etmektedir, yani ölüm olayı gerçekleşmemiştir, ve bu nedenle de rasgele sansürlü veri olarak işlem görür (Terzi, 2003).

Aralıklı Sansürleme

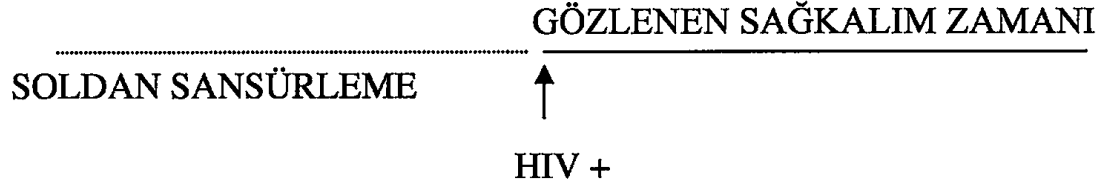
Bazı araştırmalarda gerçek olay zamanı bilinmemektedir. Fakat araştırılan olayın iki zaman noktası arasında, iki zaman aralığında gerçekleştiği bilinmektedir. Yani başarısızlık belli bir zamanda değil de iki zaman aralığında gerçekleşir (Liu, 2012). Aralıklı sansürlemede zaman, çalışma başlangıcından çalışma bitimine kadar, bir değil daha fazla periyoda ayrılarak da incelenebilir. Örneğin dört yıl süren bir araştırma, on iki aylık periyodlara ayrılarak değerlendirilebilir. Periyodlar baz alınarak analiz yapıldığı için de olası beklenen sonucun tam olarak ne zaman gerçekleştiği bilinmez, sadece hangi periyodda gerçekleştiği bilinir. Özellikle ölüm olaylarının araştırıldığı analizlerde aralıklı sansürleme daha çok kullanılır. Çünkü, bu analizler uzun bir zaman periyodu gerektirdiğinden, bu süre içerisinde her hastanın her an gözlemlenmesi mümkün değildir. Ayrıca gözlemlerin çalışmadan kaybolması da çok olası bir durum olduğundan araştırmayı periyodlara ayırarak yapmak daha yararlıdır.

Soldan Sansürleme

Soldan sansürleme, araştırılacak olan durumun, tam olarak ne zaman ortaya çıktığının bilinmediği durumlarda tanımlanan sansürleme çeşididir (Collett, 2003). Bir bireyin yaşam zamanının sol taraftan tamamlanmamış olması durumundaki, veriler soldan sansürlü (left censored) veriler olarak tanımlanır.

Soldan sansürlemeye AIDS araştırması örnek olarak gösterilebilir. Yapılan bir AIDS araştırmasına, birey, HIV virüsü pozitif çıktığı andan sonra dahil olabilir yani izlemeye alınabilir. Ancak burada bilinen zaman, sadece testin yapıldığı ve sonucun belgelendiği tarihtir ve gerçekte virüsün tam olarak ne zaman bulaştığı kesinlikle bilinemez. Virüs muhtemelen çok daha önce bulaşmıştır, yani virüsün bulaşma zamanı, teşhis

zamanından küçüktür ve solunda kalır. Dolayısıyla da veri soldan sansürlü veridir (Kleinbaum&Klein, 2005).



Şekil 2.7. Soldan Sansürlü Veri

Yine aynı şekilde, başağrısı şikayeti ile hastaneye başvuran ve beyin tümörü teşhisinin konulduğu hastalara ilişkin veriler de soldan sansürlü verilere örnek olarak gösterilebilir.

2.3. YAŞAM ANALİZİ FONKSİYONLARI

Her bir birim ya da bireye ait yaşam süresi T , tanımı gereği sürekli ve pozitif bir değere sahiptir. Herhangi bir birim için başarısızlık ancak bir kez oluşabilir. Başarısızlık süresine örnek olarak, makina bileşenlerinin yaşam süreleri, işçilerin grev süreleri ya da ekonomide işsizlik dönemleri, psikolojik bir deneyde deneğin belirlenen görevi tamamlama süresi ve klinik bir deneyde hastaların yaşam süreleri gösterilebilir (Cox&Oakes, 1984; Ata, 2010).

Yaşam verilerinin analiz edilmesi, mevcut veriler ile kitle hakkında önemli sonuçlara ulaşmasını sağlar. Bu sonuçlara ulaşmak için de mevcut olan verilerle yaşam fonksiyonlarının doğru bir şekilde tahmin edilmesi önemlidir. Yaşam verilerinin analiz edilmesinde üç önemli fonksiyon vardır. Bu fonksiyonlar, yaşam fonksiyonu, olasılık yoğunluk fonksiyonu ve hazard (tehlike) fonksiyonudur. Bu fonksiyonlardan bir tanesinin bilinmesi durumunda diğer iki fonksiyon tahmin edilebilir.

2.3.1. Olasılık Yoğunluk Fonksiyonu

Yaşam süresi ya da başarısızlık zamanını gösteren T , rasgele ve sürekli bir fonksiyon ise kısa bir zaman aralığı için (t ile $t+\Delta t$ gibi) birimin başarısız olma olasılığının limitine

T'nin olasılık yoğunluk fonksiyonu denir ve $f(t)$ ile gösterilir. Bu ifadeye dayanarak $f(t)$ 'nin formülasyonu şu şekilde gösterilebilir:

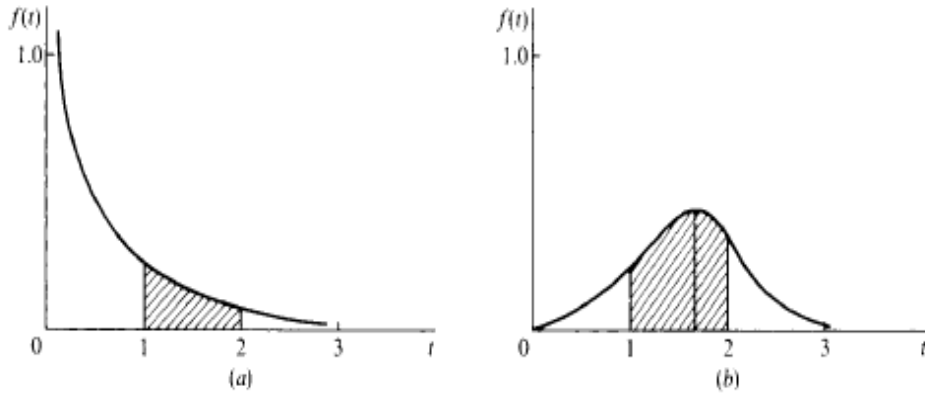
$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} \quad (2.2)$$

Bu fonksiyonda 2 özellik izlenir:

i. $f(t)$ negatif olmayan bir fonksiyondur.

$$\begin{aligned} f(t) &\geq 0 & t &\geq 0 \text{ için} \\ &= 0 & t &< 0 \text{ için} \end{aligned} \quad (2.3)$$

ii. Yoğunluk eğrisi ile t ekseninde kalan alan 1'e eşittir. Eğer sansürlü gözlem yoksa, olasılık yoğunluk fonksiyonu $f(t)$, bir zaman aralığında ölen hastaların oranı olarak tahmin edilir.



Şekil 2.8. Olasılık Yoğunluk Fonksiyonu

Şekil a Çalışmanın başında yüksek bir başarısızlık oranını ve süre arttıkça azalan bir başarısızlık oranının modelini verir. **Şekil b** Yüksek başarısızlık frekansı nın tepe noktası zamanın yaklaşık 1.7 noktasında meydana gelir. 1 ve 2 birimleri arasında

başarısız olan bireylerin oranı, yoğunluk eğrisi ve t eksenini arasında ki gölgeli alana eşittir. Yoğunluk fonksiyonu, koşulsuz başarısızlık olarak da bilinir(Lee, 2003).

Bir bireyin t zamanından önce başarısız olma (ölme) olasılığına ise dağılım fonksiyonu denir ve aşağıdaki eşitlik ile verilebilir (Lee, 2003):

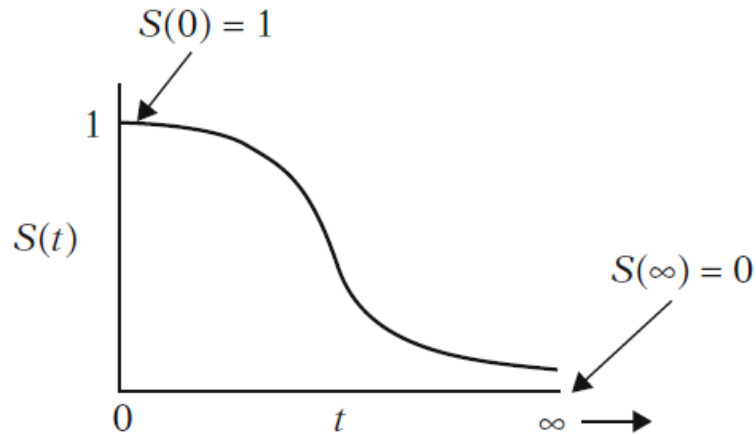
$$F(t) = P(T \leq t) = \int_0^t f(x) dx, \quad 0 < t < \infty$$

2.3.2. Yaşam Fonksiyonu

Yaşam fonksiyonu, yaşam sürelerinin olasılık dağılımına denilmektedir. Fonksiyon, yaşamsal verilerin genel eğilimini matematiksel bir modelle ifade eder. Yaşam fonksiyonu bir olasılıktır ve $S(t)$ ile gösterilir (Tamam, 2008). Bu fonksiyon, bir bireyin belirli bir t zamanından daha uzun süre hayatta kalması, bir başka deyişle T raslantı değişkeninin t'den daha büyük olma olasılığı olarak tanımlanır:

$$S(t) = P(T > t) = \int_t^{\infty} f(x) dx, \quad 0 < t < \infty \quad (2.5)$$

Yaşam fonksiyonunun grafiği (ya da aslında teorik yaşam fonksiyonu), Şekil 2.9.'daki gibi elde edilir:



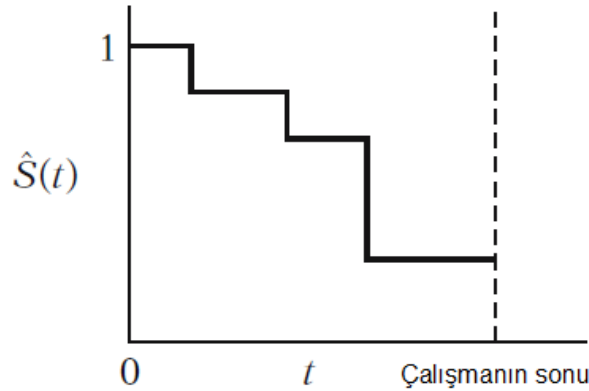
Şekil 2.9. Teorik Yaşam Fonksiyonu

Yaşam fonksiyonu, t'nin farklı değerleri için yaşam olasılıklarının elde edilmesinde veriler için önemli bilgi içermesinden dolayı, yaşam analizi için bir temel oluşturur.

Teorik olarak t , 0 ile ∞ arasında değişir ve yaşam fonksiyonu pürüzsüz bir eğri şeklinde elde edilir. X-ekseninin t olduğu Şekil 1.8.'den de görülebileceği gibi tüm yaşam fonksiyonları aşağıdaki özelliklere sahiptir (Kleinbaum&Klein, 2005):

- Monoton azalan bir fonksiyondurlar, bir başka deyişle t arttıkça azalmaktadırlar ve soldan-süreklidirler.
- $t = 0$ noktasında, $S(t) = S(0) = 1$ 'dir. Bu durum, takip süresinin başlangıcında yaşam fonksiyonu sonucunun tüm bireyler için 1 olması demektir. Bir başka deyişle, çalışmanın başlangıcında örneklemdeki tüm bireylerin yaşama olasılıkları %100'dür.
- $t = \infty$ olması durumunda, $S(t) = S(\infty) = 0$ 'dır. Bu durum, takip süresi sonsuza giderken ($t \rightarrow \infty$) yaşam fonksiyonunun sonucu 0'dır olarak açıklanabilir. Diğer bir deyişle takip süresi sonsuza kadar uzatılırsa, eninde sonunda kimse yaşamaya devam edemeyeceği için örneklemdeki tüm birimler için yaşam eğrisi sıfıra düşer.

Yukarıda verilen bilgilerin, yaşam eğrilerinin teorik özellikleri olduğu unutulmamalıdır. Şekil 2.10.'da Gerçek hayattaki uygulamalarda karşımıza çıkan yaşam fonksiyonunun grafiği verilmiştir:



Şekil 2.10. Gerçek Yaşam Fonksiyonu

Gerçek hayatta karşılaşılan uygulamalarda, pürüzsüz bir yaşam eğrisinden çok yukarıdaki şekildeki gibi bir adım fonksiyonu (step function) ile karşılaşılır. Çalışma süresinin uzunluğu hiç bir zaman sonsuz olamayacağı ve başarısızlık için yarışan

riskler de olabileceği için, çalışmaya dahil olan tüm birimlerin çalışma süresi içerisinde ölmeleri olasıdır. Bu nedenden dolayı tahmin edilen yaşam fonksiyonu, çalışmanın sonunda sıfıra kadar düşmez (Kleinbaum&Klein, 2005).

Yaşam fonksiyonu ile dağılım fonksiyonu arasındaki ilişki aşağıdaki gibidir:

$$S(t) = 1 - F(t) \quad (2.6)$$

2.3.3. Tehlike Fonksiyonu

Tehlike (hazard) fonksiyonu, t zamanına kadar yaşayan bir birimin $[t, t + \Delta t]$ aralığında yaşamının sona ermesi riski olarak tanımlanmaktadır. Diğer bir ifadeyle, herhangi bir zamandaki başarısızlık riski olarak tanımlanır. Tehlike fonksiyonu, başarısızlık hızı (failure rate), ani ölüm hızı (instantaneous death rate) ile ölümlülük gücü (force of mortality) olarak da adlandırılır. Bu tanıma göre tehlike fonksiyonu,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t / T > t)}{\Delta t} \quad (2.7)$$

biçimindedir (Ata, 2010). Sürekli dağılımlar için $h(t)$ aşağıdaki özellikleri sağlar:

- i. $h(t) > 0$
- ii. $\int_0^{\infty} h(t)dt = \infty$

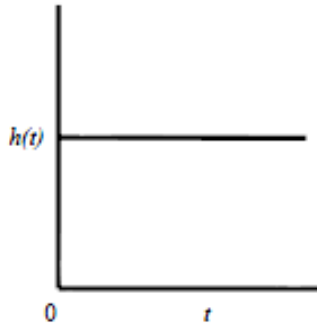
Tehlike fonksiyonunu kullanmanın faydaları aşağıdaki gibi sıralanabilir:

- t zamanına kadar yaşamını sürdürdüğü bilinen bir birimin başarısızlık riskini dikkate alır.
- Tehlike fonksiyonunu temel alan modeller, durdurma ya da birçok türden başarısızlık varken daha uygundur.

- Tehlike fonksiyonu tek tür başarısızlık içeren sistemler için özel bir analiz biçimidir (Karasoy & Ata Tutkun, 2016).

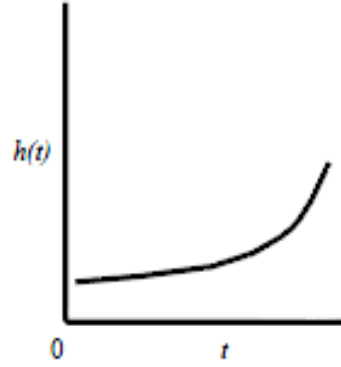
Tehlike fonksiyonu, yaşam sürdürme fonksiyonunun sahip olduğu dağılıma göre farklı yapıdadır.

Sabit tehlike modeli : Tehlike fonksiyonu sabit olduğunda sağkalım modeline üstel model denir. Bu modelde t hangi değeri alırsa alsın, $h(t)$ 'nin değeri sabittir (Kleinbaum, 2005). Sabit tehlike modeli aşağıdaki gibi gösterilebilir:



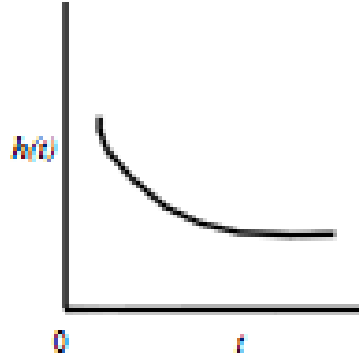
Şekil 2.11. Sabit Tehlike Fonksiyonu

Artan tehlike modeli : Bu modelde, tehlike fonksiyonunda zaman içinde artan bir yapı gözlenir ve artan Weillbul model olarak adlandırılır. Ölüm oranı zaman geçtikçe artacağı için modele ilişkin grafik de artan bir yapıdadır.



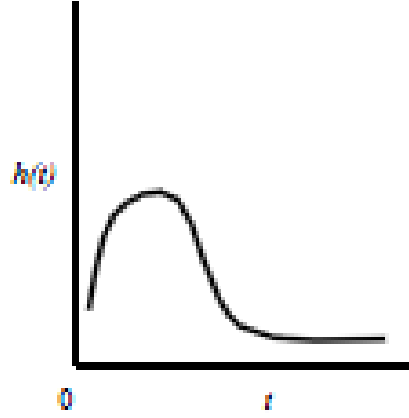
Şekil 2.12. Artan Tehlike Fonksiyonu

Azalan tehlike modeli : Bu modelde tehlike fonksiyonu, zaman içinde azalan bir yapıdadır ve azalan Weibull model olarak adlandırılır. Ölüm oranı zaman geçtikçe azalacağı için azalan yapıda bir tehlike fonksiyonu gözlemlenir.



Şekil 2.13. Azalan Tehlike Fonksiyonu

Önce artan sonra azalan tehlike modeli : Bu modelde ise tehlike fonksiyonu, önce artan daha sonra azalan yapıdadır ve log-normal model olarak adlandırılır. Modelin grafiği aşağıdaki gibi gözlemlenir (Kleinbaum, 2005)



Şekil 2.14

Şekil 2.14. Önce Artan Sonra Azalan Tehlike Fonksiyonu

Yaşam fonksiyonu, yaşama olasılığını incelerken; tehlike fonksiyonu başarısızlığı (ölümü) inceler ve zamana bağlı ölüm riskini belirler. Tehlike fonksiyonunun grafiği, yapılan çalışma için model oluşturmada önemli ipuçları verir. Bu sebeple, yaşam fonksiyonu gibi, hazard fonksiyonu da, yaşam modellerinin önemli bir karakteristiğidir (Tamam, 2008).

2.3.4. Birikimli Tehlike Fonksiyonu

Araştırma için belirlenen T zaman süresi içerisinde, belirlenen bir t dönemi için hesaplanan başarısızlık hızlarının birikimli fonksiyonuna birikimli tehlike fonksiyonu denir ve $H(t)$ ile gösterilir ve aşağıdaki eşitlik ile ifade edilir:

$$H(t) = \int_0^t h(x) dx \quad (2.8)$$

Ayrıca, birikimli tehlike fonksiyonu, yaşam fonksiyonundan (2.9) nolu eşitlikle elde edilir:

$$H(t) = -\log S(t) \quad (2.9)$$

Birikimli tehlike fonksiyonu $H(t)$ aşağıdaki özellikleri sağlar (Karasoy&Ata Tutkun, 2016):

- i. Artan bir fonksiyondur.
- ii. $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ olduğundan $\lim_{t \rightarrow \infty} H(t) = \infty$ 'dur.
- iii. Sağdan sürekli bir fonksiyondur.
- iv. Ayrıca:

$$t = 0 \text{ iken } S(t) = 1, H(t) = 0 \text{ 'dır.}$$

$$t = \infty \text{ iken } S(t) = 0, H(t) = \infty \text{ 'dur.}$$

Buradan hareketle, birikimli tehlike fonksiyonunun 0 ile ∞ arasında değer aldığı görülebilir (Terzi, 2003).

2.3.5. Yaşam Fonksiyonlarının Birbirleriyle İlişkileri

Daha önce alt bölüm 2.3.'de de değinildiği gibi, yaşam fonksiyonu, olasılık yoğunluk fonksiyonu ve tehlike fonksiyonu birbirleriyle ilişkilidir ve bu fonksiyonlardan bir tanesinin bilinmesi durumunda diğer iki fonksiyon tahmin edilebilir. Söz konusu fonksiyonlar arasındaki ilişkiler aşağıda açıklanmıştır (Karasoy & Ata Tutkun, 2016).

Koşullu olasılık tanımından hareketle, yaşam ve tehlike fonksiyonları arasında bazı ilişkiler elde edilebilir. Olasılık teorisine göre bir B olayının bilinmesi koşulu altında, A olayının gerçekleşmesi olasılığı Eşitlik (2.10) ile bulunabilir:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (2.10)$$

Bu özelliğin kullanılması ile, tehlike fonksiyonunun tanımında verilen limit eşitliğindeki koşullu olasılık:

$$\frac{P(t \leq T < t + \Delta t)}{P(T > t)} \quad (2.11)$$

Bu biçimde yazılabilir ve bu da,

$$\frac{F(t + \Delta t) - F(t)}{S(t)} \quad (2.12)$$

İfadesine eşittir. Burada $F(t)$, T 'nin dağılım fonksiyonudur. Buradan,

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \quad (2.13)$$

yazılabilir. Buradaki $\lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\}$ ifadesi $F(t)$ 'nin t 'ye göre türevinin tanımıdır ve buradan aşağıdaki eşitlik yazılabilir:

$$h(t) = \frac{f(t)}{S(t)} \quad (2.14)$$

$f(t)$, $S(t)$ ve $h(t)$ fonksiyonları, T değişkenine ilişkin dağılım fonksiyonunun farklı biçimlerde ifade edilişi olarak düşünülebilir ve fonksiyonlar arasındaki ilişkiler aşağıdaki matematiksel ifadeler ile kolaylıkla oluşturulabilir.

$$f(t) = - \frac{dS(t)}{dt} \quad (2.15)$$

$$h(t) = - \frac{\frac{dS(t)}{dt}}{S(t)} = - \frac{d}{dt} \log S(t) \quad (2.16)$$

Böylece,

$$\log S(x)|_0^t = - \int_0^t h(x)dx \quad (2.17)$$

olur. $S(0)=1$ olduğundan,

$$S(t) = \exp\left(- \int_0^t h(x)dx\right) \quad (2.18)$$

bulunur. $H(t) = \int_0^t h(x)dx$, birikimli tehlike fonksiyonu olduğundan,

$$S(t) = \exp[-H(t)] \quad (2.19)$$

yazılabilir. Birikimli tehlike fonksiyonu, yaşam fonksiyonundan aşağıdaki eşitlik ile elde edilebilir:

$$H(t) = -\log S(t) \quad (2.20)$$

Söz konusu üç fonksiyon arasında $f(t) = h(t).S(t)$ ilişkisi olduğuna göre, olasılık yoğunluk fonksiyonu ile tehlike fonksiyonu arasındaki ilişki:

$$f(t) = h(t). \exp\left(- \int_0^t h(x)dx\right) \quad (2.21)$$

biçiminde de yazılabilir.

2.4.YAŞAM ANALİZİNDE KULLANILAN OLASILIK DAĞILIMLARI

Belirli bir zamanda, belirli bir kişinin ölümü ya da başarısızlığına neden olan bir çok fiziksel faktör vardır. Bu fiziksel faktörleri matematiksel olarak açıklamak imkansız değil

ancak çok zordur. Bu nedenle yaşam verilerinin hangi dağılıma yaklaştığına karar vermek oldukça zor ve önemli bir sanattır. Bu bölümde yaşam sürelerini tanımlamak için kullanılan önemli dağılımlar özetlenmiştir(Lee&Wang, 2003).

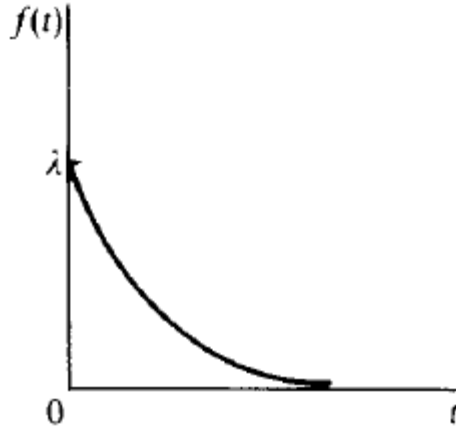
2.4.1. Üstel Dağılım

Üstel dağılım, yaşam analizinde kullanılan en basit ve de en önemli dağılımdır. 1940'lı yılların sonunda araştırmacılar elektronik sistemlerin yaşam örüntülerini (pattern) açıklayabilmek için üstel dağılım kullanmaya başlamışlardır (Lee&Wang, 2003). Bu dağılım, özellikle rasgele başarısızlığın olduğu durumlarda sıklıkla kullanılmaktadır (Terzi, 2003). Bununla birlikte üstel dağılım, istenen sonucun ilk kez gerçekleşmesine kadar geçen sürenin olasılıklarının hesaplanmasında da kullanılabilir. Ancak bu dağılımın sabit tehlike hızına sahip olması, bir çok durum için uygun model olarak seçilmesini engellemektedir (İnceoğlu, 2013).

Üstel dağılım tek bir parametreye sahiptir. Bu parametre “sabit hazard oranı” λ , dır. λ 'nın yüksek değeri, yüksek risk ve kısa yaşam süresinin, düşük değeri ise düşük risk ve daha uzun bir yaşam süresinin göstergesidir (İnceoğlu, 2013).

Yaşam süresi, λ parametresi ile üstel dağılıma sahip ise, bu dağılımın olasılık yoğunluk fonksiyonu (2.22)'deki gibi olup fonksiyonun grafiği Şekil 2.15'de gösterilmiştir.

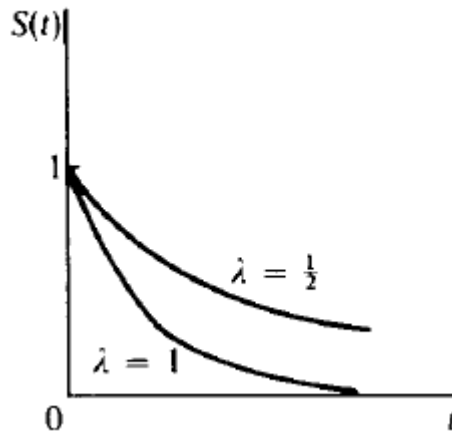
$$f(t) = \begin{cases} \lambda e^{-\lambda t} & , t \geq 0, \lambda > 0 \\ 0 & , t < 0 \end{cases} \quad (2.22)$$



Şekil 2.15. Üstel Dağılım Olasılık Yoğunluk Fonksiyonu

Üstel dağılımın yaşam fonksiyonu aşağıdaki gibi olup bu fonksiyonun grafiği Şekil (2.16)'da verildiği gibidir:

$$S(t, \lambda) = e^{-\lambda t} \quad (2.23)$$



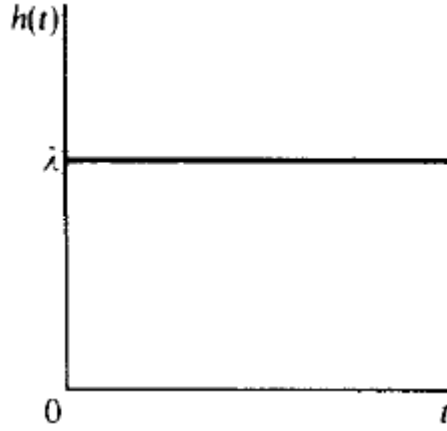
Şekil 2.16. Üstel Dağılımın Yaşam Fonksiyonu

Burada, λ $(0, \infty)$ aralığında değer alır. Yaşam fonksiyonundan hareketle üstel dağılımın tehlike fonksiyonu ise aşağıdaki gibi yazılabilir:

$$h(t; \lambda) = -\frac{1}{S(t; \lambda)} \frac{dS(t; \lambda)}{dt} = \frac{\frac{d}{dt} e^{-\lambda t}}{e^{-\lambda t}} = \frac{\mu e^{-\lambda t}}{e^{-\lambda t}} = \lambda \quad (2.24)$$

$$h(t; T \approx \exp) = \lambda \quad (2.25)$$

Yukarıdaki eşitlikte de görüldüğü gibi üstel dağılımın tehlike fonksiyonu sabittir. Tehlike fonksiyonunun grafiği ise Şekil 2.17'de yer almaktadır:



Şekil 2.17. Üstel Dağılım Tehlike Fonksiyonu

Üstel dağılımın birikimli tehlike fonksiyonu:

$$H(t; \lambda) = -\log S(t; \lambda) = -\log \{ \exp [-(\lambda t)] \} = \lambda t \quad (2.26)$$

şeklinde yazılır. Bu dağılımın birikimli dağılım fonksiyonu ise:

$$F(t; \lambda) = 1 - S(t; \lambda) = 1 - e^{-\lambda t} \quad (2.27)$$

şeklinde ifade edilir (Liu, 2012).

2.4.2. Weibull Dağılımı

Parametrik zaman dağılımları ailesi içerisinde Weibull dağılımı, basitliği ve esnekliği bakımından yaşam analizinde en çok uygulanabilirliği olan parametrik fonksiyondur.

Bununla birlikte, Weibull fonksiyonu iki yaklaşımla tanımlanabilir: Orantılı tehlike modeli (proportional hazard rate model) ve hızlandırılmış başarısızlık zamanı modeli (accelerated time failure process) (Liu, 2012).

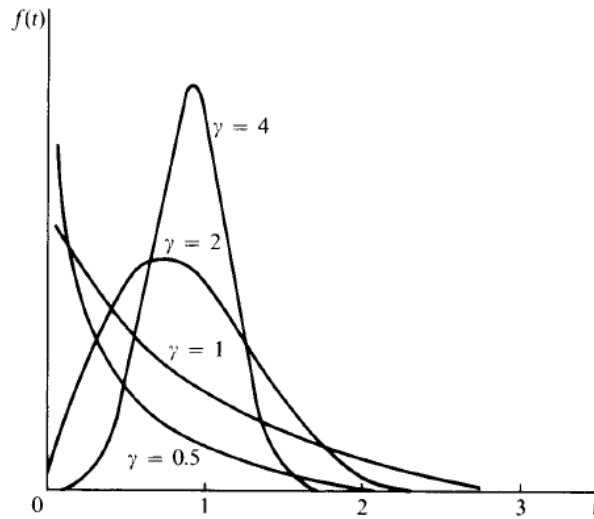
Weibull dağılımı, üstel dağılımın genelleştirilmiş biçimidir. Bu dağılım, Weibull (1939) tarafından önerilmiş ve çeşitli başarısızlık durumlarındaki kullanımı Weibull (1951) çalışmasında incelenmiştir. Weibull dağılımı, sabit tehlike hızına sahip olmadığından üstel dağılımdan daha fazla kullanım alanına sahiptir. Söz konusu dağılım, güvenilirlik analizlerinde ve yaşam analizlerinde kullanılmaktadır (Lee&Wang, 2003).

Weibull dağılımı, iki parametre ile tanımlanmaktadır. Ölçek parametresi λ ve biçim parametresi γ sembolleriyle ifade edilir. Bu parametreler birlikte işler ve üstel dağılıma göre daha esnek parametrik bir model elde etmeyi sağlar.

Weibull dağılımının olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(t; \lambda, \gamma) = h(t)S(t) = \lambda\gamma(\lambda t)^{\gamma-1} \exp[-(\lambda t)^\gamma], \quad t \geq 0, \quad \gamma, \lambda > 0 \quad (2.28)$$

Aşağıdaki grafikte $\lambda = 1$ durumunda Weibull dağılımının olasılık yoğunluk fonksiyon grafiği yer almaktadır.

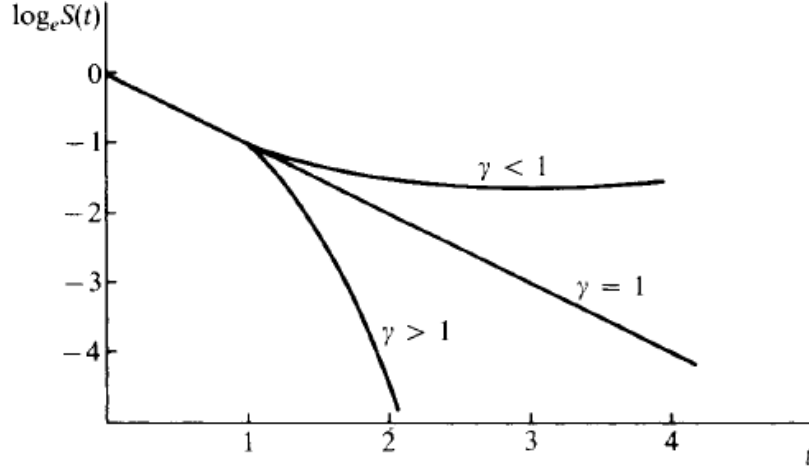


Şekil 2.18. $\lambda = 1$ için Weibull Dağılımın Olasılık Yoğunluk Fonksiyonu

Weibull dağılımının yaşam fonksiyonu aşağıdaki (2.29)'da verilmiştir.

$$S(t; \lambda, \gamma) = \exp[-(\lambda t)^\gamma] \quad (2.29)$$

Weibull dağılımının yaşam fonksiyonu yerine bu fonksiyonun logaritmasının grafiğini elde etmek daha kolay olduğundan, bu grafik Şekil 2.19'da verilmiştir:



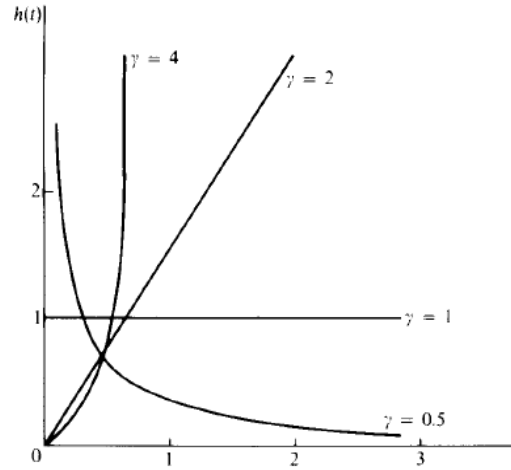
Şekil 2.19. $\lambda = 1$ için Weibull Dağılımının $\log_e S(t)$ Fonksiyonu

Weibull dağılımın tehlike fonksiyonu, yaşam fonksiyonundan hareketle aşağıdaki gibi ifade edilir (Lee & Wang, 2003; Liu, 2012):

$$h(t; \lambda, \gamma) = -\frac{\frac{d}{dt} e^{-(\lambda t)^\gamma}}{e^{-(\lambda t)^\gamma}} = \lambda \gamma (\lambda t)^{\gamma-1} \quad (2.30)$$

$\gamma = 1$ alınırsa $h(t) = \lambda$ olur ve tehlike fonksiyonu zaman arttıkça sabit kalmaktadır. sabit hazar hızına dönüştüğünden üstel dağılım elde edilir. Tehlike fonksiyonu, $\gamma > 1$ iken artar ve bu da riskin artarak devam edeceği ve $\gamma < 1$ iken riskin giderek azalacağı anlamına gelir (Liu, 2012).

Weibull dağılımının tehlike fonksiyonuna ilişkin grafiği Şekil 2.20'de verilmektedir:



Şekil 2.20

Şekil 2.20. $\lambda = 1$ için Weibull Dağılımın Tehlike Fonksiyonu

Weibull dağılımının birikimli tehlike fonksiyonu ile birikimli dağılım fonksiyonu Eşitlik (2.31) ile Eşitlik (2.32)'de verildiği gibidir:

$$H(t; \lambda, \gamma) = -\log S(t; \lambda, \gamma) = -\log \left\{ \exp \left[-(\lambda t)^\gamma \right] \right\} = (\lambda t)^\gamma \quad (2.31)$$

$$F(t; \lambda, \gamma) = 1 - S(t; \lambda, \gamma) = 1 - \exp \left[-(\lambda t)^\gamma \right] \quad (2.32)$$

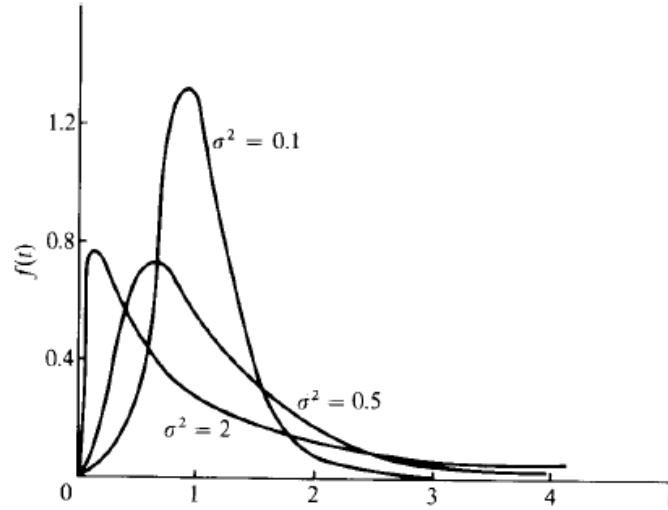
2.4.3. Log-Normal Dağılım

Log-normal dağılım, en basit olarak logaritması normal dağılım gösteren bir değişkenin dağılımı olarak tanımlanabilir. Bu dağılımın teorik çatısını McAlister (1879) oluşturmuştur. Gaddum (1945a, b) biyoloji alanındaki, Boag (1949) kanser araştırmalarındaki, Aitchison & Brown (1957) ise ekonomi alanındaki uygulamalarını incelemiştir. Sonraları araştırmacılar, sağa çarpık olması ve yaşam zamanlarının logaritmalarının yaklaşık olarak normal dağılıma sahip olmaları nedeniyle Alzheimer, Hodgkin ve Lösemi gibi hastalıkların yaşam süresi dağılımının log-normal dağılım olabileceğini bulmuşlardır (Lee&Wang, 2003).

T yaşam süresi olarak ele alınırsa $\log T$, μ ortalama ve σ^2 varyansı ile normal dağılmaktadır. Buradan hareketle T'nin log-normal dağılıma sahip olduğu söylenir ve $T: \Lambda(\mu, \sigma^2)$ şeklinde ifade edilir. Burada, μ ve σ^2 'nin log-normal dağılımın ortalaması ve varyansı olmadığına dikkat edilmelidir.

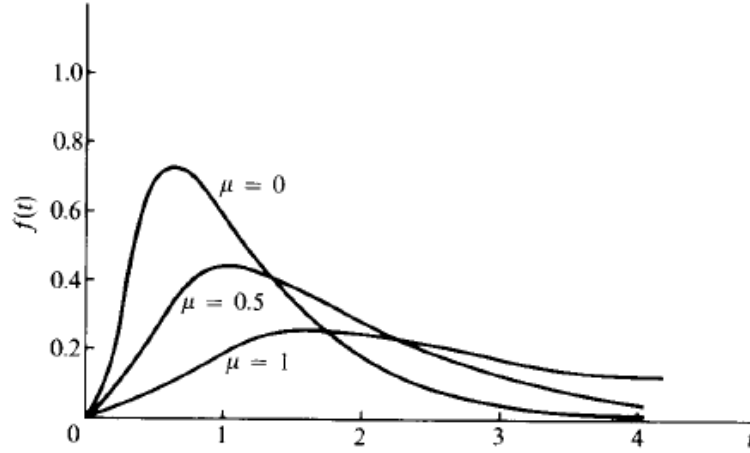
Log-normal dağılımın olasılık yoğunluk fonksiyonu aşağıdaki şekilde ifade edilir:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right], \quad t > 0, \quad \sigma > 2 \quad (2.33)$$



Şekil 2.21. $\mu = 0$ için Log-normal Dağılımın Olasılık Yoğunluk Fonksiyonu

Şekil 2.21.'de görüldüğü gibi, log-normal dağılımın olasılık yoğunluk fonksiyonu $\mu = 0$ için verilmiştir. Bu fonksiyondan da görülebileceği üzere, σ^2 varyansı azaldıkça, dağılımın çarpıklığı azalır, daha simetrik bir dağılıma dönüşmektedir.



Şekil 2.22. $\sigma^2 = 0.5$ için Log-normal Dağılımının Olasılık Yoğunluk Fonksiyonu

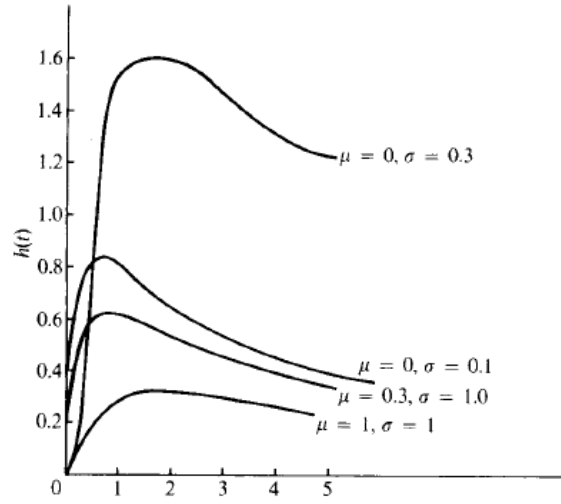
Şekil 2.22.'de ise görüldüğü gibi, $\sigma^2 = 0.5$ için log-normal dağılımın olasılık yoğunluk fonksiyonu verilmiştir. Bu fonksiyondan da görülebileceği gibi, μ sıfıra yaklaştıkça dağılım da normal dağılıma yaklaşmaktadır.

Log-normal dağılımın yaşam fonksiyonu $S(t)$ ile tehlike fonksiyonu $h(t)$ sırasıyla Eşitlik (2.34) ve (2.35)'deki gibidir:

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right] dx \quad (2.34)$$

$$h(t) = \frac{(1/t\sigma\sqrt{2\pi}) \exp\left[-(\log e^{-\mu}t)^2/2\sigma^2\right]}{1 - \Phi\left[\log(e^{-\mu}t)/\sigma\right]} \quad (2.35)$$

Log-normal dağılımın tehlike fonksiyonu monoton değildir, belli bir miktar arttıktan sonra azalma eğilimi gösterir. Bu dağılımın tehlike fonksiyonunun farklı parametreler için olan grafiği Şekil 2.18'de verildiği gibidir (Lee&Wang, 2003).



Şekil 2.23. Farklı Parametreler için Log-normal Dağılımının Tehlike Fonksiyonu

2.5. YAŞAM ANALİZİNDE KULLANILAN YÖNTEMLER

Bilimsel çalışmaların çoğunda olayların gelişimini ve sonucunu etkileyen çok sayıda neden vardır. Yaşam süresi üzerine yapılan araştırmaların en temel amaçlarından biri, sonuç değişkenine etki eden bağımsız değişkenler arasındaki nedenselliği açıklamaktır. Araştırılacak konuya ait veriler toparlanmaya başlandıkça her birim, birey veya gözleme ait başka özellikler ortaya çıkacak ve analizde bağımsız değişken olarak yerini alacaktır. Yaşam süresi üzerinde etkisi olacağı düşünülen bütün değişkenler bağımsız veya prognostik değişken olarak adlandırılır. İstatistiksel analizlerde, sonuç değişkenine etki eden bağımsız değişkenleri ile arasındaki ilişkiyi açıklamak için kullanılan klasik regresyon analizi yöntemleri, yaşam analizi için yapılan araştırmalarda uygun olmamaktadır. Bunun nedeni, klasik regresyon analizinin uygulanabilmesi için gerekli olan varsayımların, yaşam analizinin gerektirdiği varsayımlardan farklı olmasıdır. Burada gözlenen en büyük farklılık, yaşam analizinde verilerin sansürlü olmasıdır. Bununla birlikte, verilerin normal dağılıma sahip olmaması ile verilerin genellikle kategorik yapıda olması da söz konusu nedenlere eklenebilir. Yaşam analizi uygulamalarında kullanılan üç farklı yöntem vardır. Bunlar, Yaşam Tablosu Yöntemi, Kaplan-Meier Yöntemi ve Cox Regresyon Yöntemidir. Bu tez çalışması kapsamında asıl ele alınacak olan Cox Regresyon yönteminin ayrıntılı incelenmesine geçmeden

önce, yöntemler arasındaki farkın iyi anlaşılabilmesi amacıyla, diğer iki yönteme de kısaca değinilecektir.

2.5.1. Yaşam Tablosu Yöntemi

Yaşam tablosu yöntemi, tıbbi istatistikçiler tarafından kullanılan en eski istatistiksel yöntemlerden biridir (Cox, 1972). Ayrıca nüfus bilimciler tarafından ülke, bölge yaşam olasılıklarının hesaplanmasında ve sigortacılar tarafından da yaşam beklentilerini belirlemek amacıyla yaşam tablosu yöntemi kullanılmıştır. Bu yöntem, araştırma sonuçlarını, araştırmacı tarafından belirlenen zaman aralıklarına bölüp kümelere ayırarak değerlendiren bir yöntemdir (Şenocak, 2009). Yaşam tablosu yöntemi ilk olarak, ölüm düzeylerinin ölçülmesi, belli bir hastalığa yakalanan bireylerin kaç yıl daha yaşayabileceğini ya da belli bir tedavi yönteminden sonra hastanın kaç yıl daha yaşayabileceğini tahmin eder. Söz konusu yöntem daha sonraları, nüfus yapısı ve özellikleri, belli bir yaşta beklenen yaşam süresi, hastalıklı yaşam süresi gibi konularda da tahmin aracı olarak kullanılmıştır (Özdamar, 2003). Bireyler, araştırma periyoduna araştırmanın sürdüğü herhangi bir zaman diliminde girebilirler. Araştırmada olgularının ortaya çıkardığı en kısa ve en uzun izleme süreleri ve araştırmanın toplam süresi bellidir. Araştırmacı başlangıç olarak elindeki bu verilerle toplam süreyi aralıklara böler. Yaşam tablosu yönteminde toplam olgu sayısının oldukça yüksek olduğu çalışma örnekleri için $n \geq 300$ olarak kullanılması önerilir. Toplam süreye göre ayrılan periyodların ise $3 < k < 13$ şeklinde olması önerilir (Şenocak, 2009).

Yaşam tablosu yönteminin temel amacının zamanı periyodlara ayırarak incelemek olduğu yukarıda da belirtilmişti. Bu yöntemi bir örnek üzerinde kısaca açıklamak istersek, örneğin; iki yıldır kanser tedavisi gören hastaların ortalama yaşam süresini hesaplamanın hiç bir anlamı olmaz. Sonuçta halen daha yaşamını yitirmeyen hastalar vardır. İlk iki ayda hastaların %20'sinin, 1.yılda da %60'ının (%20'nin dışında) hayatını kaybettiklerini, geriye kalan %20'lik bir hasta diliminin ise halen yaşadıklarını düşünelim. İlk 1.yıldaki %80'lik bir kayıp geriye kalan %20'lik bir hasta diliminin de ortalama yaşam olasılığını çok fazla düşürerek hatalı, yanlış sonuçlara ulaşmamıza neden olur. Yaşam tablosu yöntemi, bu gibi nedenlerden dolayı süreleri araştırmanın amacına göre uygun görülen periyodlara ayırır ve her periyod için yaşam sayısına göre olasılıkları hesaplamaya yarar.

2.5.2. Kaplan-Meier Yöntemi

Kaplan-Meier (KM) yöntemi, yaşam sürelerine ilişkin verileri yaşam tablosundaki gibi zaman aralıklarına bölmeden, yaşam ve tehlike olasılıklarının tam olarak hesaplanmasını sağlayan bir yöntemdir (Terzi, 2003). Bu nedenle bu yöntemde kesin ölüm tarihi kullanıldığı için nokta yaşam olasılığı bulunur.

Bu yöntemde, her ölümün meydana geliş zamanındaki kestirimleri yapılır. Burada her ölüm zamanı değerlendirildiğinden, yaşam olasılıkları adimsal fonksiyon şeklinde belirlenir. Dolayısıyla, bir önceki ölüm anından sonra yeni ölüm zamanına kadar başka ölüm olayının olmaması olasılığı değişmez. Sıfır zamanındaki yaşam olasılığı 1'e eşittir. Araştırmanın bittiği anda izlemesi devam eden olgu varsa, son ölüm zamanındaki birikimli yaşam olasılığı alınır (Cox & Oakes, 1984).

KM yönteminde nokta yaşam olasılığı bulunduğundan kayıp ve eksik veriler dikkate alınmaz. Fakat araştırmacının ve okuyucuların kayıpların oranı ve kaybolma oranları üzerinde fikir sahibi olması, değerlendirmeler açısından önemli fikirler verebileceğinden, bu bilgilerin grafiğin üzerinde gösterilmesinde yarar vardır (Özdamar, 2003; Şenocak, 2009).

Yaşam tablosu ile KM yöntemlerinin her ikisinin de amacı, ölüm-yaşam fonksiyonlarını hesaplayarak geleceğe yönelik çıkarımlar yapmaktır. Fakat bu iki yöntemi birbirinden ayıran en önemli faktör verinin büyüklüğüdür. Verinin büyüklüğü arttıkça her iki yöntem benzer sonuçlar verecektir. Veri büyüklüğü az olan bir analizi yaşam tablosu yöntemi ile incelemek, bu yöntemin güvenilirliğini azaltacaktır. Bu noktada dikkat edilmesi gereken unsur, veri büyüklüğü fazla ise Yaşam Tablosu yöntemini, az ise KM yöntemini kullanmaktır (Karagöz, 2015).

3. COX REGRESYON MODELİ

3.1. COX ORANTILI TEHLİKELER MODELİNİN YAPISI

Yaşam verisinin analizinde modelleme sürecinin amacı, tehlike fonksiyonunu etkileyen bağımsız değişkenleri belirlemek ve bireye ait tehlike fonksiyonunu elde etmektir. Yaşam verisini modellemek için kullanılan temel model orantılı tehlikeler modelidir. Bu model, İngiliz istatistikçi Cox (1972) tarafından önerildiğinden Cox regresyon modeli olarak da bilinmektedir (Ata, 2005). Cox, regresyon modelini geliştirirken parametrik olmayan yöntemler kullanmıştır. Ayrıca, analize zamana bağlı değişkenleri de dahil etmiş ve bununla birlikte orantılı tehlike modelinin kullanımı ile iki veya daha fazla tedavi grubunun karşılaştırılmasını da mümkün kılarak, Cox regresyon modelinin kullanımını oldukça genişletmiştir (Ersoy, 2005). Bu nedenle Cox regresyon modeli yaşam analizlerinde en çok kullanılan model haline gelmiştir (Sertkaya, Ata & Sözer, 2005). Orantılı tehlikeler varsayımına dayanan bu modelde, yaşam süreleri için olasılık dağılımının belirli bir biçimi yoktur. Bu sebepten dolayı, Cox regresyon modeli yarı parametrik bir model olarak nitelendirilir (Ata, 2005).

$X_1, X_2, \dots, X_p$ p tane bağımsız değişken ve $x_1, x_2, \dots, x_p$ bu değişkenlerin aldığı değerler olsun. Cox regresyon modelinde bağımsız değişkenlerin aldığı değerlerin kümesi $x = (x_1, x_2, \dots, x_p)$ vektörü, t yaşam süresi ve $h_0(t)$ temel tehlike fonksiyonu olmak üzere, i. birim için tehlike fonksiyonu aşağıdaki eşilikte verildiği gibidir (Collett, 2003).

$$h_i(t) = h_0(t) \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}) \quad (3.1)$$

(3.1) nolu eşitlikteki model, bağımsız değişkenler grubuna dayanarak, bir birimin t anındaki tehlike olasılığını gösterir. Bir başka deyişle, bir birimin bağımsız değişkenler grubuna dayanarak tehlike olasılığını modeller (Kurt, 2008).

Eşitlik (3.1)'deki modelde yer alan β 'lar regresyon modelinin katsayılarıdır. $h_0(t)$ temel tehlike fonksiyonu bağımsız değişkenlerden bağımsız, zamana bağlı parametrik olmayan tahminleri veren bir fonksiyondur. Bu fonksiyon, X bağımsız değişkenleri sıfır değerini aldığı anda ölüm riski olarak bilinen tehlike fonksiyonuna eşit olur. Cox regresyon modellerinde sabit terim yoktur; ancak sabit terim yerine $h_0(t)$ terimi modelde yerini alır (Terzi, 2003).

Cox regresyon modelinde iki önemli varsayım vardır. Bu varsayımlardan ilki, bağımsız değişkenlerin yaşam fonksiyonu üzerindeki etkilerinin log-doğrusal olduğudur. İkinci varsayım ise bağımsız değişkenlerin log-doğrusal fonksiyonu ile ölüm fonksiyonu arasında çarpımsal bir ilişki olmasıdır. Bu varsayımlara göre, farklı bağımsız değişken gruplarına sahip olan iki birimin ölüm fonksiyonlarının oranı zamana bağlı değildir, ölüm riskleri orantılıdır (Özdamar, 2003).

Cox regresyon modelinin yaşam fonksiyonu aşağıdaki gibidir (Kleinbaum & Klein, 2005):

$$S(t, X) = [S_0(t)]^{e^{\sum_{i=1}^p \beta_i X_i}} \quad (3.2)$$

$S_0(t)$ temel yaşam fonksiyonudur ve değeri 0 ile 1 arasında değer alır.

3.2. OLABİLİRLİK FONKSİYONU

Cox regresyon modelinin yaşam analizi verilerine uygulanabilmesi için bilinmeyen parametrelerin ve temel tehlike fonksiyonunun tahmin edilmesi gerekir. Modeldeki bilinmeyen parametreler ve tehlike temel fonksiyonu ayrı ayrı tahmin edilir. Öncelikle bilinmeyen β_i 'ler tahmin edilir, daha sonra da buradan elde edilen değerler, temel tehlike fonksiyonunun tahmin edilmesinde kullanılır.

Cox regresyon modelinde bilinmeyen parametreler en çok olabilirlik (EÇO) yöntemi kullanılarak tahmin edilir. Bu yöntemi kullanmak için öncelikle olabilirlik fonksiyonunun elde edilmesi gerekir. Bu fonksiyon, gözlenen verinin bileşik olasılığıdır ve varsayılan modeldeki bilinmeyen parametrelerin fonksiyonu olarak ele alınır. Cox regresyon modeli için olabilirlik fonksiyonu, gözlenen yaşam sürelerinin ve modelin doğrusal bileşenindeki bilinmeyen β_i parametrelerinin bir fonksiyonudur (Karasoy & Ata Tutkun, 2016).

r tanesi ayrık başarısızlık süresi ve $(n - r)$ tanesi sağdan sansürlenmiş yaşam sürelerine sahip toplam n tane birim olsun. Her bir başarısızlık süresinde yalnızca bir birimin başarısız olduğu ve veride birbirine bağlı hiçbir gözlem olmadığı varsayalım. $t_{(j)}$, j . sıralı başarısızlık süresi olmak üzere, r tane sıralı başarısızlık süresi $x_{(1)} < x_{(2)} < \dots < x_{(r)}$ şeklinde gösterilsin. $t_{(j)}$ süresinde risk altında olan birimlerin kümesi $R(t_{(j)})$ ile gösterilsin. Burada $R(t_{(j)})$, $t_{(j)}$ süresinden hemen önce yaşayan ve sansürlenmemiş olan birimlerin kümesi olur. $R(t_{(j)})$, *risk kümesi* olarak adlandırılır.

$x_{(j)}$ bağımsız değişken vektörlü bir birimin $t_{(j)}$ süresinde başarısız olması olasılığı tehlike fonksiyonlarının oranı olarak tanımlanır ve (3.3) nolu eşitlikteki formülle verilmektedir:

$$\frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \quad (3.3)$$

Cox (1972), (3.1) nolu eşitlikte verilen orantılı tehlikeler modeli için ilgili olabilirlik fonksiyonunun aşağıdaki gibi olduğunu belirtmiştir (Collett, 2003):

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \quad (3.4)$$

(3.4) nolu eşitlikte $x_{(j)}$, j. sıralı başarısızlık süresi (ölüm zamanı) $t_{(j)}$ 'de başarısız olan (ölen) birimler için bağımsız değişkenler vektörüdür. Olabilirlik fonksiyonunun paydasında yer alan toplam, $t_{(j)}$ süresinde risk altında olan birimler üzerinden $\exp(\beta'x)$ değerlerinin toplamıdır. Burada olabilirlik fonksiyonunun formülünde kullanılan çarpım işleminin, başarısızlık süreleri kaydedilen yani sadece tamamlanmış birimler üzerinden yapıldığına dikkat edilmelidir. Bu nedenle olabilirlik fonksiyonu, kısmi olabilirlik fonksiyonu olarak da anılır. Yaşam süreleri sansürlenmiş olan birimler, log olabilirlik fonksiyonunun pay kısmında yer almaz; ancak sansürleme zamanından önce ortaya çıkan başarısızlık sürelerindeki risk kümeleri aracılığıyla toplama girer. Olabilirlik fonksiyonunun iki temel özelliği vardır: bilinmeyen büyüklük $h_0(t)$ 'nin yok edilmesi ve sansürlenmiş yaşam süreleri tarafından etkilenmemesidir (Ata, 2010).

Eşitlik (3.4)'deki olabilirlik fonksiyonu sansürlenmiş gözlemleri hesaba katmamaktadır. Gözlemlenmiş bireylerin yanı sıra sansürlü bireyler de hesaba katılırsa, olabilirlik fonksiyonu aşağıdaki eşitlikte verilen farklı bir fonksiyona dönüşür. Burada veri kümesinin $t_1, t_2, \dots, t_n$ olmak üzere n tane gözlenen yaşam süresi içerdiğini düşünelim. δ_i ise gösterge değişken olsun. $i = 1, 2, \dots, n$ olmak üzere i. yaşam süresi t_i , sağdan sansürleme varsa $\delta_i = 0$, diğer durumlarda ise $\delta_i = 1$ değerini alır.

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta'x_i)}{\sum_{I \in R(t_i)} \exp(\beta'x_I)} \right]^{\delta_i} \quad (3.5)$$

(3.5) nolu eşitlikteki olabilirlik fonksiyonunun logaritması ise eşitlik (3.6)'da verildiği gibidir.

$$\log L(\beta) = \sum_{i=1}^n \delta_i \left\{ \beta'x_i - \log \sum_{I \in R(t_i)} \exp(\beta'x_I) \right\} \quad (3.6)$$

Log olabilirlik fonksiyonunu maksimize ederek en iyi β tahmin edicilerini bulmak için kullanılan sayısal yöntem ise Newton Raphson algoritmasıdır. β tahmin edicileri elde edildikten sonra, her bir bağımsız değişken için elde edilen pozitif katsayı riskin yüksek ve prognostik etkenlerin kötü olduğunu, negatif katsayı ise, gözlemler için daha iyi prognostik etkenlerin olduğuna işaret eder (Yay, Çoker & Uysal, 2007).

3.3. β KATSAYILARININ ANLAMLIĞININ TEST EDİLMESİ

β 'ların kısmi olabilirlik tahmin edicileri $\hat{\beta}$ ise, elde edilen bu tahmin edicilerin anlamlı olup olmadığına bakılması gerekir. Buna karar vermek için de başvurulacak üç farklı test yöntemi vardır. Parametrelerin anlamlılığını test etmek için kurulan:

$H_0 : \beta = \beta_0$ hipotezi

1. Wald, 2. olabilirlik oranı ve 3. skor istatistikleri kullanılarak test edilir.

3.3.1. Wald Testi

Wald testi, en çok olabilirlik tahminlerinin normal dağıldığı varsayımına dayanır. Bu test istatistiği, t veya Z istatistiğinin genelleştirilmiş biçimi olup, en çok olabilirlik tahmini ile hipotezlenen değer arasındaki farkın bir fonksiyonudur. Bu fark, en çok olabilirlik tahmininin standart hatasının tahmini değeri ile normalleştirilmiştir (Yay, Çoker & Uysal, 2007). Dolayısıyla Wald testi, en çok olabilirlik tahminlerinin normal dağıldığı varsayımına dayanmaktadır. Wald istatistiği, regresyon katsayısının tahmininin, standart hatasına oranı olarak aşağıdaki biçimde ifade edilir:

$$W = \frac{\hat{\beta}}{SE(\hat{\beta})} \quad (3.7)$$

Wald test istatistiği, n'nin büyük değerleri için 1 serbestlik dereceli ki-kare dağılımı gösterir ve anlamlılık kararı, 1 serbestlik dereceli χ^2 dağılımının kritik değeri ile karşılaştırılarak belirlenir. $W \geq \chi^2_1$ ise, H_0 hipotezi reddedilerek söz konusu değişkenin

kurulan model için anlamlı olduğuna karar verilir. $W < \chi_1^2$ ise H_0 hipotezinin reddi için yeterli neden bulunamaz ve böylece ilgili değişkenin kurulan modelde tehlike fonksiyonunu açıklamada anlamlı olmadığı kanısına varılır.

3.3.2. Olabilirlik Oranı Testi

Olabilirlik oranı testi, Wald testinden daha genel bir testtir. Bu test, modelde iki veya daha fazla sınıfa sahip kategorik değişkenin varlığı durumu ile aynı anda birçok değişkenin modele alınması durumunda katsayı anlamlılığının belirlenmesi amacıyla kullanılır (Işık, 2007). Olabilirlik oranı test istatistiği aşağıdaki şekilde ifade edilir:

$$LR = -2 \log \left(\frac{I_0}{I_v} \right) \quad (3.8)$$

Burada I_0 , tam modelin en çok olabilirlik istatistiğini, I_v ise v değişkenli modelin en çok olabilirlik istatistiğini göstermektedir. $L_0 = \log l_0$ ve $L_v = \log l_v$ kabul edildiğinde, (3.8) nolu eşitlik aşağıdaki (3.9) nolu eşitliğe dönüşür:

$$LR = -2 \left(\frac{L_0}{L_v} \right) \quad (3.9)$$

LR istatistiği, tahmin edilecek parametre sayısı olan v serbestlik derecesi ile χ^2 dağılımı gösterir. $LR \geq \chi_v^2$ ise H_0 hipotezi reddedilir ve oluşturulan modelin geçerli olduğu kanısına varılır ve en az bir regresyon katsayısının sıfırdan farklı olduğu söylenebilir. $LR < \chi_v^2$ ise H_0 hipotezinin reddi için yeterli neden bulunamamıştır ve modele alınan değişkenlerin tehlike fonksiyonunu açıklamada yeterli olmadığına karar verilir (Özşen, 2006).

Modelin anlamlılık testi yapıldıktan sonra, elde edilen olabilirlik fonksiyonuna en yüksek katkısı olan değişkenleri modele ekleyerek, yeni indirgenmiş bir model kurarak bu

modeli tam modelle karşılaştırmak gerekir (Terzi, 2003). Burada yokluk hipotezi aşağıdaki gibi kurulur:

$$H_0 : \text{Modelden çıkarılan değişkenlere ait hipotezi } \beta_i = 0, \quad i = 1, 2, \dots, p$$

Kullanılacak LR testinin formülü de aşağıdaki gibi olacaktır:

$$LR = -2 \left(\frac{L_1}{L_2} \right) \quad (3.10)$$

(3.10) nolu eşitlikte L_1 , tam modelin olabilirlik değerini, L_2 indirgenmiş modelin olabilirlik değerini göstermektedir. Bu eşitlikteki LR test istatistiği $p_1 - p_2$ serbestlik derecesi ile χ^2 dağılımına sahiptir. Burada p_1 , tam modelde yer alan değişken sayısı, p_2 ise indirgenmiş modeldeki değişken sayısını ifade etmektedir. LR istatistiği, $p_1 - p_2$ serbestlik derecesi ile χ^2 tablo değeri ile karşılaştırılır. $LR \geq \chi^2_{p_1-p_2}$ ise H_0 hipotezi reddedilir ve indirgenmiş model dışında kalan değişkenlerin olabilirlik fonksiyonuna katkısının önemli olmadığına kanaat getirilir. $LR < \chi^2_{p_1-p_2}$ ise H_0 hipotezinin reddi için yeterli neden bulunamamıştır ve modelden çıkarılan değişkenlerin modele katkısının olmadığı söylenebilir (Özşen, 2006).

3.3.3. Score Testi

Score test istatistiği logaritmik olabilirlik istatistiklerinden yararlanılarak hesaplanmaktadır. İkili değer alabilen değişkenler modelde var olduğunda uygulanan bir yöntemdir. Skor testi, birden fazla değişken ve sürekli değişkenler için log rank testinin genelleştirilmiş şeklidir. İki değerli değişkenin gözlenme oranı p , ve bu orana göre ilgilenilen değişkenin gözlenme olasılığı r ise en çok olabilirlik fonksiyonu aşağıdaki gibi yazılabilir (Yay, Çoker & Uysal, 2007).

$$L = \log(l) = r \log p + (n - r) \log(1 - p) \quad (3.11)$$

Skor istatistiğini elde etmek için L 'nin p 'ye göre birinci ve ikinci türevini alıp sıfıra eşitlenir:

$$S = \frac{(L')^2}{(L'')} \quad (3.12)$$

Skor istatistiği, serbeslik derecesi tahmin edilen parametre sayısı olacak şekilde, bu serbestlik derecesindeki χ^2 dağılımının kritik değeriyle karşılaştırılır. $S \geq \chi_n^2$ ise bu durumda H_0 hipotezi reddedilerek oluşturulan modelin geçerli olduğu kanısına varılır ve en az bir regresyon katsayısının sıfırdan farklı olduğu söylenebilir. $S < \chi_n^2$ ise H_0 hipotezinin reddi için yeterli neden bulunamayarak modele alınan değişkenlerin tehlike fonksiyonunu açıklamada yeterli olmadığı kanısına varılır (Özşen, 2006).

3.4. COX ORANTILI TEHLİKE MODELİNİN VARSAYIMI

İki birime ait bağımsız değişkenler vektörü $x = (x_1, x_2, \dots, x_p)$ ve $x^* = (x_1^*, x_2^*, \dots, x_p^*)$ olsun. Burada x küçük tehlike fonksiyonuna sahip grubu, x^* da büyük tehlike fonksiyonuna sahip grubu gösterirse tehlike oranı aşağıdaki biçimde elde edilir. Bu eşitlikte büyük tehlike fonksiyonuna sahip grubun pay kısmında yer almasının nedeni 1'den büyük tehlike oranını yorumlamak, 1'den küçük tehlike fonksiyonunu yorumlamaktan daha kolay olduğu içindir (Karasoy & Ata Tutkun, 2016).

$$\theta = \frac{\hat{h}(t, x^*)}{\hat{h}(t, x)} = \frac{\hat{h}_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j x_j^*\right)}{\hat{h}_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j x_j\right)} = \frac{\exp\left(\sum_{j=1}^p \hat{\beta}_j x_j^*\right)}{\exp\left(\sum_{j=1}^p \hat{\beta}_j x_j\right)} = \exp\left(\sum_{j=1}^p \hat{\beta}_j (x_j^* - x_j)\right) \quad (3.13)$$

(3.13) nolu eşitlikte θ tehlike oranı (hazard ratio) ya da görel tehlike (relative hazard) olarak tanımlanır.

Cox orantılı tehlike modelinin en temel varsayımı, tehlike oranlarının tüm zaman boyunca sabit olmasıdır. Bir bireyin tehlikesi, diğer bireyin tehlikesi ile orantılıdır ve bu orantı zamandan bağımsızdır (Terzi, 2003). Modelde farklı bağımsız değişken setlerine sahip olan iki birimin başarısızlık fonksiyonları oranı zamana bağlı değil, başarısızlık oranları orantısaldır. Buna 'orantılı tehlike varsayımı' denilmektedir (Özdamar, 2003). Örneğin farklı gruplarla yapılan çalışmalarda her grubun birbirleriyle aralarında olan başarısızlık oranı her yıl aynı oranda olmalıdır. Önemli olan grubun kendi içinde başarısızlığındaki değişim değil de, diğer grup ile arasındaki başarısızlık oranının sabit olmasıdır. Ancak böyle bir durumda orantılı tehlike varsayımı sağlanabilir. Eşitlik (3.13) incelendiğinde tehlike oranının t'yi içermediği, yani zamandan bağımsız olduğu görülebilir.

Bu temel varsayımın sağlanıp sağlanmadığını test etmenin üç yöntemi vardır. Bunlar grafiksel yöntemler, uyum iyiliği testleri ve zamana bağlı değişkenlerin kullanılması yöntemleridir.

3.4.1. Grafiksel Yöntemler

Bağımsız değişkenlerin farklı kategorileri için yaşam eğrilerinin tahminlerini grafik yöntemi ile karşılaştırmanın iki yolu vardır. Bunlar log-log grafikleri ile gözlenen ve beklenen yaşam eğrilerinin karşılaştırılması ile varsayımın test edilmesidir.

Log-log Grafikleri

Grafiksel yöntemler arasında en çok kullanılan yöntem log-log grafikleridir. Bu yöntemde farklı gruplar için çizilen log-log yaşam eğrileri karşılaştırılarak varsayımın sağlanabilirlik koşulu test edilmektedir. Kaplan-Meier tahminlerine dayanan log-log yaşam eğrileri, eğer birbirine paralel ise varsayımın sağlandığı söylenebilir.

Cox regresyon modeli için tehlike ve yaşam fonksiyonları sırasıyla aşağıdaki eşitliklerdeki gibidir:

$$h(t, x) = h_0(t) \exp \left[\sum_{j=1}^p \beta_j x_j \right] \quad (3.14)$$

$$S(t, x) = [S_0(t)]^{\exp\left[\sum_{j=1}^p \beta_j x_j\right]} \quad (3.15)$$

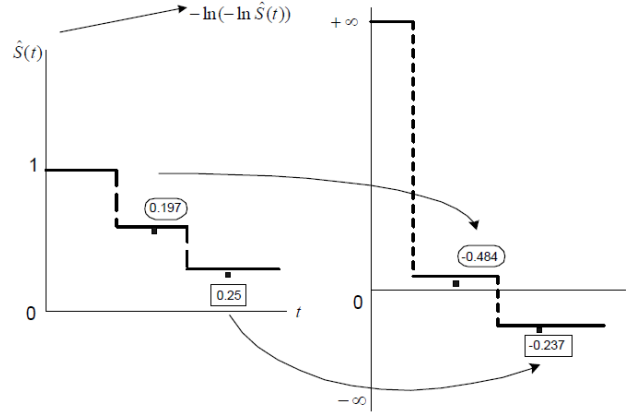
Log-log yaşam eğrisi, yaşam eğrisi tahminine uygulanan basit bir dönüşümdür ve yaşam olasılığı tahmininin iki kez doğal logaritması alınarak elde edilir. Bu eğriler, matematiksel olarak $-\ln(-\ln \hat{S})$ şeklinde ifade edilir. Yaşam fonksiyonunun logaritması (3.16) nolu eşitlikte verildiği gibidir:

$$\ln S(t, x) = \exp\left[\sum_{j=1}^p \beta_j x_j\right] \ln[S_0(t)] \quad (3.16)$$

Yaşam eğrisi tahmini olasılığı $[0, 1]$ aralığında değer aldığından, $\hat{S}$ 'nın bir kez logaritmasının alınması sonucu negatif bir değer elde edilecektir. Yalnızca pozitif sayıların logaritması alınabildiğinden, $\hat{S}$ 'nın ikinci kez logaritması alınarak önüne (-) çarpanı konur ve sonuç olarak $-\ln(-\ln \hat{S})$ ifadesi $(-\infty, \infty)$ aralığında değer alır. Buna göre $-\ln S(t, x)$ 'in logaritması alınırsa (3.17) nolu eşitlikteki ifadeye ulaşılır (Kleinbaum&Klein, 2005). İkinci logaritma pozitif ya da negatif olabilir. Dolayısıyla ikinci kez söz konusu ifadenin negatifinin alınmasına gerek yoktur. Tutarlılık adına, $\ln(-\ln)$ ifadesini elde etmek için ikinci logaritmanın önüne eksi işareti konulur.

$$\begin{aligned} \ln[-\ln S(t, x)] &= \ln\left[-\exp\left(\sum_{j=1}^p \beta_j x_j\right) \ln[S_0(t)]\right] \\ &= \ln\left[\exp\left(\sum_{j=1}^p \beta_j x_j\right)\right] + \ln[-\ln[S_0(t)]] \\ &= \sum_{j=1}^p \beta_j x_j + \ln[-\ln[S_0(t)]] \end{aligned} \quad (3.17)$$

Yaşam fonksiyonları adım fonksiyonu olarak çizildiğinden log-log grafikleri de adım fonksiyonu olarak çizilebilir.



Şekil 3.1. $-\ln(-\ln \hat{S})$ Eğrisinin Elde Edilmesi

Şekil 3.1. $-\ln(-\ln \hat{S})$ eğrisinin elde edilmesinin şematik gösterilmesidir (Göz Çekçeki, 2007). Yukarıdaki şekilde görüldüğü gibi yaşam eğrisi y ekseninde $[0, 1]$ arasında değer alırken $-\ln(-\ln \hat{S})$ eğrisi $(-\infty, \infty)$ arasında değer almaktadır.

İki bireye ait bağımsız değişkenler vektörünün $x_1 = (x_{11}, x_{12}, \dots, x_{1p})$ ve

$x_2 = (x_{21}, x_{22}, \dots, x_{2p})$ olduğu varsayılın. $-\ln[-\ln S(t, x)]$ ifadesinde x yerine x_1 ve

x_2 yazılırsa, bireylere ait log-log eğrileri sırasıyla (3.18) ve (3.19) eşitliklerindeki gibi

elde edilir:

$$-\ln[-\ln S(t, x_1)] = -\sum_{j=1}^p \beta_j x_{1j} - \ln[-\ln[S_0(t)]] \quad (3.18)$$

$$-\ln[-\ln S(t, x_2)] = -\sum_{j=1}^p \beta_j x_{2j} - \ln[-\ln[S_0(t)]] \quad (3.19)$$

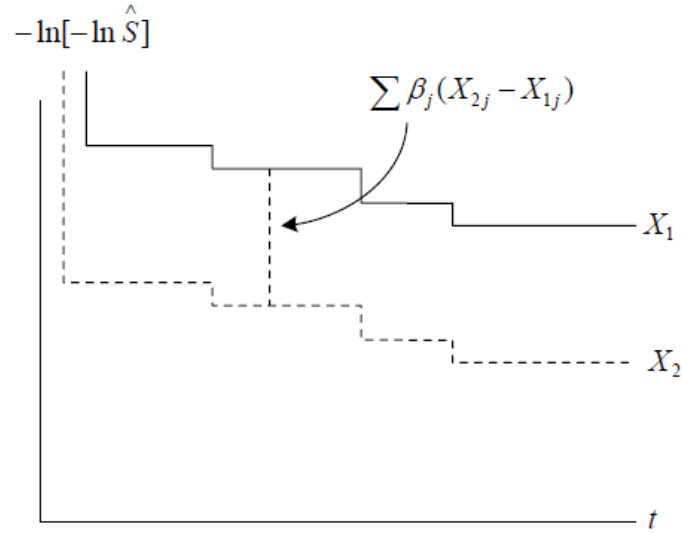
(3.19) nolu eşitlikten (3.18) nolu eşitlik çıkartılırsa aşağıdaki eşitlik elde edilir:

$$-\ln[-\ln S(t, x_1)] - (-\ln[-\ln S(t, x_2)]) = -\sum_{j=1}^p \beta_j (x_{2j} - x_{1j}) \quad (3.20)$$

(3.20) nolu eşitlik, iki bireyin bağımsız değişkenlerinin tahmin değerlerindeki farkların doğrusal toplamıdır. Burada temel yaşam fonksiyonunun ortadan kalktığına ve dolayısıyla log-log yaşam eğrileri arasındaki farkın t 'yi içermediğine dikkat edilmelidir (Kleinbaum & Klein, 2005). Eşitlik (3.20) düzenlenerek (3.21) nolu eşitlikteki gibi ifade edilebilir:

$$-\ln[-\ln S(t, x_1)] = -\ln[-\ln S(t, x_2)] + \sum_{j=1}^p \beta_j (x_{2j} - x_{1j}) \quad (3.21)$$

Cox regresyon modeli kullanılır ve aynı grafik üzerinde bireyler için log-log yaşam eğrilerinin tahmini çizilirse iki eğrinin birbirine paralel olduğu ve böylelikle orantılı tehlike varsayımının sağlandığı görülecektir ve bu durum Şekil 3.2.'de verilmiştir (Kleinbaum&Klein, 2005). İki eğri arasındaki uzaklık, tahmin değerleri arasındaki farkları içeren, ancak t 'yi içermeyen doğrusal bir ifadedir.



Şekil 3.2. İki Bireye ait Log-log Yaşam Eğrileri

Orantılı tehlike varsayımının log-log grafikleri ile test edilmesinde bazı sakıncalar vardır. Burada paralelliğin nasıl belirleneceğine ilişkin verilecek karar öznel bir karar olabilir. Log-log yaşam eğrilerinin paralel olmadığına dair güçlü bir kanıt olmadıkça, orantılı tehlike varsayımının sağlandığı varsayılır. Bununla birlikte, sürekli değişkenlerin nasıl

sınıflandırılacağıının belirlenmesi önemlidir. Sürekli değişkenler sınıflandırılırken, sınıf sayısının olabildiğince küçük alınması önerilir (Karasoy & Ata Tutkun, 2016). Burada önemli olan bir diğer konu da, birçok değişken için orantılı tehlike varsayımının eş zamanlı olarak nasıl değerlendirileceğidir. Bu duruma karşı önerilen çözüm, tüm değişkenlerin ayrı ayrı sınıflandırılması ve bu sınıflandırmalardan çeşitli kombinasyonların oluşturularak, tüm log-log yaşam eğrilerinin aynı grafik üzerinde karşılaştırılmasıdır. Ancak bu durumda paralellik olup olmadığına karar vermek oldukça güçtür (Kleinbaum & Klein, 2005).

Gözlenen ve Beklenen Yaşam Eğrileri

Orantılı tehlike varsayımının test edilmesinde kullanılan grafik yöntemlerinin ikincisi, tahmin edilen eğrilerin, gözlenen eğriler ile karşılaştırılmasıdır. Gözlenen ve beklenen yaşam eğrileri yöntemi, uyum iyiliği testinin grafiksel olarak benzeri olup bu yöntem aynı zamanda log-log grafiklerine alternatif bir çözüm olarak kullanılır. Uyum iyiliği testlerinde olduğu gibi, gözlenen ve beklenen yaşam eğrileri yönteminde de gözlenen ve beklenen yaşam olasılıklarının tahminleri kullanılır.

Log-log grafiklerinde olduğu gibi, gözlenen ve beklenen yaşam eğrileri yönteminde de aşağıda verilen yöntemlerden biri ya da her ikisi için de kullanılabilir:

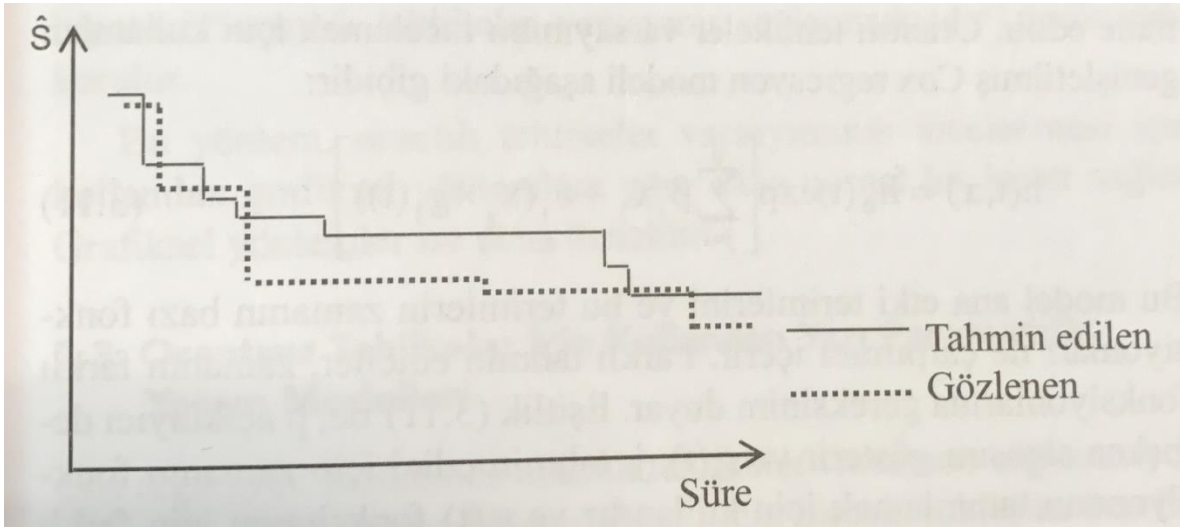
- Her bir zaman noktasındaki tüm değişkenler için orantılı tehlikeler varsayımının değerlendirilmesi
- Diğer değişkenler için gerekli düzeltmeler yapıldıktan sonra orantılı tehlikeler varsayımının değerlendirilmesi

İlk yöntemde gözlenen eğrileri çizmek için KM eğrileri kullanılır. İkinci yöntemde ise gözlenen eğrilerin çiziminde tabakalandırılmış Cox regresyon modeli kullanılır. İlk yöntem kullanılırken, kestricilerin düzeyleri kullanılarak veri tabakalandırılır ve devamında her düzey için KM eğrileri ayrı ayrı elde edilir (Karasoy & Ata Tutkun, 2016).

Beklenen yaşam eğrilerini elde etmek için, değerlendirilen bağımsız değişkeni içeren Cox regresyon modeli kurulur. Yaşam eğrisinin tahmini, bağımsız değişkenin her bir kategorisi için değerler formülde yerine konarak elde edilir. Böylelikle bağımsız

değişkenin her bir kategorisi için ayrı bir yaşam eğrisi elde edilmiş olur. Eğer bağımsız değişkenin her bir kategorisi için gözlenen ve beklenen yaşam eğrileri birbirine yakınsa, orantılı tehlikeler varsayımının sağlandığına karar verilir. Bir ya da daha çok kategori için bu eğriler birbirine yakın değilse varsayımın sağlanmadığı söylenebilir (Kleinbaum & Klein, 2005).

Şekil 3.3.'te gözlenen ve beklenen yaşam eğrilerine bir örnek verilmiştir (Karasoy & Ata Tutkun, 2016).



Şekil 3.3. Gözlenen ve Beklenen Yaşam Eğrileri

Bununla birlikte varsayımın sağlanıp sağlanmadığı incelenirken sürekli değişkenlerin kullanılması durumunda da, gözlenen yaşam eğrileri kategorik değişkenlerde olduğu gibi elde edilir. Söz konusu sürekli değişken kategorilere ayrılarak tabakalar oluşturulur ve daha sonra her bir kategori için KM eğrileri elde edilir.

Bu yöntemin dezavantajı da gözlenen ve beklenen yaşam eğrilerini karşılaştırırken “ne kadar yakının yakın kabul edilip edilmeyeceği”dir. Bu durum tıpkı log-log grafiklerinde “ne kadar paralelliğin kabul edilip edilmeyeceği”ne benzemektedir. Burada sonuç olarak, gözlenen ve beklenen yaşam eğrileri gerçekten çok farklı bulunursa, orantılı tehlikeler varsayımının sağlanmadığına karar verilir (Kleinbaum & Klein, 2005).

3.4.2. Uyum İyiliği Testi

Uyum iyiliği testi, orantılı tehlikeler varsayımının kontrol edilmesinde oldukça cazip bir yöntemdir. Çünkü bu yöntem, ilgilenilen değişkenin varsayımı sağlayıp sağlamadığını değerlendirmek için bir test istatistiği ve ona ilişkin bir p-değeri verir. Dolayısıyla, araştırmacı grafiksel yöntemlerden ziyade daha objektif olan bu yöntemi kullanarak daha objektif bir karar vermiş olur.

Orantılı tehlikeler varsayımının kontrolü için literatürde kullanılan farklı uyum iyiliği testleri vardır. Bunlardan birisi de Schoenfeld'in (1982) önerdiği uyum iyiliği testidir. Bu yaklaşım, modeldeki her değişken için hesaplanan gözlenen ve beklenen yaşam olasılıklarına dayanan ki-kare dağılımı aracılığıyla varsayımın sağlanıp sağlanmadığını test eder. Uyum iyiliği testi, bir serbestlik dereceli ki-kare dağılımına sahiptir. Her bir değişkenin orantılı tehlike olasılık değeri 0.05'den küçük ise, o değişken için orantılı tehlikeler varsayımı sağlanmıyor demektir (Terzi, 2003).

Orantılı tehlikeler varsayımının değerlendirmesinde istatistiksel bir test, grafiksel yöntemlere göre daha objektif bir bakış açısı kazandırır da uyum iyiliği testinin de zayıf olduğu birtakım noktalar bulunmaktadır. Örneğin, p-değeri örneklem büyüklüğünden direk olarak etkilenmektedir. Bununla birlikte grafiksel yöntemlerde araştırmacı grafiğe bakarak neler olduğunu az çok anlayabilir. Dolayısıyla, burada araştırmacının kararını verirken grafiksel yöntemler ile uyum iyiliği testini birlikte kullanması önerilmektedir (Kleinbaum&Klein, 2005).

3.4.3. Zamana Bağlı Bağımsız Değişkenlerin Kullanılması

Zamandan bağımsız bir değişkenin, orantılı tehlikeler varsayımını sağlayıp sağlamadığının incelenmesinde zamana bağlı değişkenler kullanılır. Genişletilmiş Cox regresyon modeli olarak da anılan bu model, zamandan bağımsız değişkenler ile zamanın bazı fonksiyonlarını kapsayan çarpım terimlerini içerir. Genişletilmiş Cox regresyon modeli aşağıdaki gibidir (Karasoy & Ata Tutkun, 2016).

$$h(t, x) = h_0(t) \exp \left[\sum_{j=1}^p \beta_j x_j + \delta_j(x_j g_j(t)) \right] \quad (3.22)$$

(3.22)'de p bağımsız değişken sayısını gösterir ve $g_j(t)$, j . tahmin edici için zamanın fonksiyonunu tanımlamak için kullanılan bir fonksiyondur. Bu fonksiyon için kullanılabilecek çeşitli fonksiyonlar ($g_j(t) = t$ ya da $g_j(t) = \log t$ v.b.) vardır.

Genişletilmiş Cox regresyon modeli için çarpım terimlerinin anlamlılık testi yapılarak orantılı tehlikeler varsayımı değerlendirilir. Burada yokluk hipotezi, $H_0 = \delta_1 = \delta_2 = \dots = \delta_p = 0$ şeklinde kurulur. Bu hipotez, olabilirlik oranı istatistiği aracılığıyla test edilir.

$$LR = -2 \ln \hat{L}_{\text{Cox Regresyon Modeli}} - (-2 \ln \hat{L}_{\text{Genişletilmiş Cox Regresyon Modeli}}) \sim \chi_p^2 \quad (3.23)$$

Burada LR istatistiği p serbestlik dereceli ki-kare dağılımına sahiptir. Eğer yokluk hipotezinin doğru olduğuna karar verilirse, model Cox regresyon modeline indirgenir ve böylelikle orantılı tehlikeler varsayımının sağlandığı söylenebilir.

Bu yöntemin kullanılmasının dezavantajı ise $g_j(t)$ fonksiyonunun seçimidir. Farklı fonksiyon kullanılması durumunda, orantılı tehlikeler varsayımının sağlanıp sağlanmadığı ile ilgili farklı kararlara varılabilir.

3.5. COX REGRESYON MODELİNİN ZAMANA BAĞLI DEĞİŞKENLER İÇİN GENİŞLETİLMESİ

Belirli bir gözlem süresi gerektiren tıbbi araştırmalarda, hastalığın ya da tedavinin seyrini değiştirebilecek bağımsız değişkenlerin varlığı ile zamana bağlı değişkenlerin olması durumu yeni bir regresyon modelinin kurulmasını gerekli kılmıştır. Örneğin, prostat kanserinde tümörün boyutu ve diğer önemli değişkenler belirli aralıklarla ölçülür. Böylece değişkenin ya da değişkenlerin değerindeki değişim belirlenen bir zamandaki ölüm riskinin tahmin edilmesi açısından daha tatmin edici bir sonuç verir. Bir başka deyişle, araştırmada belirli bir değişkene ait elde edilen son bulgular, çalışmanın

başlangıcındaki bulgulara göre daha doğru sonuçlar verir. Bu durum da zamana bağlı değişkenlerin kullanılmasını bir anlamda zorunlu kılar (Collett, 2003).

Eşitlik (3.1)'deki Cox-regresyon modelinde, temel tehlike fonksiyonu zamana bağlı iken, üstel kısımda yer alan değişkenler zamandan bağımsızdır. Modelde zamana bağlı değişkenlerin olması durumunda Cox regresyon modeli kullanılmaya devam edilebilir; ancak gözlem süresi uzun olan çalışmalarda orantılı tehlikeler varsayımı sağlanamaz. Bu durumda, zamana bağlı bağımsız değişkenli Cox regresyon modeli olarak da adlandırılan genişletilmiş Cox regresyon modeli kullanılabilir.

İncelenen model “zamandan bağımsız” ve “zamana bağımlı” olmak üzere iki çeşit bağımsız değişken içerebilir. Araştırma süreci boyunca değeri değişmeden kalan bağımsız değişkenler “zamandan bağımsız değişkenler”, değeri zaman içerisinde değişen bağımsız değişkenler “zamana bağlı bağımsız değişkenler” olarak tanımlanır. Örneğin yaş ve kan değerleri ölçümleri zaman içerisinde değişen bir değişkenler olduğundan zamana bağlı değişkenler olarak kabul edilir. Fakat bir gözlemin cinsiyeti zaman içerisinde değişmeyeceğinden zamandan bağımsız değişken olarak kabul edilir (İnceoğlu, 2013).

Üç çeşit zamana bağlı bağımsız değişken vardır. Bunlardan ilki, tanımlanmış zamana bağlı bağımsız değişkendir. Bu, sabit bir bağımsız değişken ile zamanın bir fonksiyonunun çarpımı şeklinde elde edilen değişkendir. İkincisi değişken ise içsel değişkendir. Bireyin kendi davranış ve özelliklerine bağlı olarak değişkenin değerinde bir değişme meydana geldiği durumlarda içsel değişkenden bahsedilebilir. Bir başka deyişle, değişimin nedeni bireyin kendi özelliklerinden kaynaklanmaktadır. Üçüncü değişken ise dışsal değişkendir. Dışsal değişken, değişkenin değerinin çevresel faktörlere göre değiştiği durumu ifade etmektedir. Bireyin yaşadığı coğrafyanın iklimsel özellikleri, hava kirliliği indeksi gibi faktörler, bireyin elinde olmadan hastalığın seyrini değiştirdiği durumlar, dışsal değişkenlere örnek olarak verilebilir (Fisher & Lin, 1999). Bununla birlikte, aynı anda hem içsel hem de dışsal olabilecek değişkenlerin olduğu durumlar da bulunmaktadır. Örneğin kalp rahatsızlığı olup kalp nakline karar verilen bir birey için hem içsel hem de dışsal faktörler etkilidir. Hastanın kendi özelliklerinin (kan grubu, yaş, tansiyon durumu v.b.) durumu içsel faktörleri, bulunan donörün durumu

nakil yapılacak hasta için dışsal özelliklerin kontrol edilmesini gerektirmektedir. (Kleinbaum & Klein, 2005).

$x_1, x_2, \dots, x_{p_1}$ zamandan bağımsız değişkenler ve $x_1(t), x_2(t), \dots, x_{p_2}(t)$ zamana bağlı değişkenler olmak üzere bağımsız değişkenler,

$$x(t) = (x_1, x_2, \dots, x_{p_1}, x_1(t), x_2(t), \dots, x_{p_2}(t)) \quad (3.24)$$

biçiminde gösterilir. Buna göre genişletilmiş Cox regresyon modeli (3.25) nolu eşitlikteki gibi yazılabilir:

$$h(t, x(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j (x_j t) \right] \quad (3.25)$$

(3.25) nolu eşitlikte $x_j(t) = x_j g_j(t)$ 'dir. $g_j(t)$ de alt bölüm (3.4.3)'de belirtildiği gibi zamanın fonksiyonunu tanımlamak için kullanılır. $g_j(t)$ 'nin seçimi kullanılan değişkenlerin durumuna ve araştırmacının bilgisine göre değişir. Farklı $g_j(t)$ fonksiyonları için oluşacak yeni modeller aşağıda verilmiştir (Kleinbaum & Klein, 2005):

1. $g_j(t) = 0$ ise, x_i değişkeni zamana bağlı olur ve aşağıdaki gibi temel Cox regresyon modeli halini alır:

$$h(t, x(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i \right] \quad (3.26)$$

2. $g_j(t) = t$ ise, modeldeki her x_i değişkeni için $x_i * g_j(t) = x_i * t$ halini alır ve model aşağıdaki gibi elde edilir:

$$h(t, x(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j (x_j t) \right] \quad (3.27)$$

3. $g_i(t) = \ln(t)$ ise, t yerine sadece $x_i^* \ln t$ 'nin zaman bağımlı değişken olarak kullanılmasıdır ve model aşağıdaki formüle dönüşür:

$$h(t, x(t)) = h_0(t) \exp \left[\sum_{i=1}^{p_1} \beta_i x_i + \sum_{j=1}^{p_2} \delta_j (x_j \ln t) \right] \quad (3.28)$$

Genişletilmiş Cox regresyon modelinde de Cox regresyon modelinde olduğu gibi modeldeki regresyon katsayıları EÇO yöntemleri kullanılarak tahmin edilir. İstatistiksel çıkarımlar yapmak için de yine Cox regresyon modelinde olduğu gibi Wald ve/veya olabilirlik oran testleri kullanılır (Karasoy & Ata Tutkun, 2016).

$x(t)$ ve $x^*(t)$, iki grup için hem zamandan bağımsız hem de zamana bağlı bağımsız değişkenleri içeren bağımsız değişkenler vektörü olmak üzere, bu vektörler aşağıdaki gibi verilsinler:

$$x(t) = (x_1, x_2, \dots, x_{p_1}, x_1(t), x_2(t), \dots, x_{p_2}(t)) \quad (3.29)$$

$$x^*(t) = (x_1^*, x_2^*, \dots, x_{p_1}^*, x_1^*(t), x_2^*(t), \dots, x_{p_2}^*(t))$$

Buna göre, genişletilmiş Cox regresyon modeli için tehlike oranı (3.30) numaralı eşitlikteki gibi yazılabilir:

$$\frac{\hat{h}(t, x^*(t))}{\hat{h}(t, x(t))} = \exp \left[\sum_{i=1}^{p_1} \hat{\beta}_i (x_i^* - x_i) + \sum_{j=1}^{p_2} \hat{\delta}_j (x_j^*(t) - x_j(t)) \right] \quad (3.30)$$

Genellikle, tehlike oranı t zamanında zamana bağlı değişkenlerin değerlerindeki farklılığı içerdiğinden, tehlike oranı zamanın bir fonksiyonudur. Bir başka deyişle, herhangi bir δ_j sifıra eşit değilse orantılı tehlikeler varsayımı sağlanmaz (Karasoy & Ata Tutkun, 2016).

4. UYGULAMA

Bir üniversite hastanesinin Nefroloji bölümünde yatmakta olan ve böbrek yetmezliği yaşayan 463 hasta üzerinde yapılan bir araştırmada, hastalar 253 günlük gözlem süresi geçirmiştir. Veri setindeki değişkenler; cinsiyet (1: kadın, 2: erkek), yaş, tedavi (1: diyaliz, 2: ilaç, 3: ameliyat ve ilaç), durum (1: tamamlanmış, 0: tamamlanmamış) ve gözlem süresi (gün) olarak belirlenmiştir. 231'i kadın ve 232'si erkek gözlemden oluşan veri setinde tedavi değişkeninin diyalize bağlı 244, ilaç tedavisi gören 109 ve hem ameliyat olup hem de ilaç tedavisi gören 110 gözlemi bulunmaktadır. Ayrıca durum değişkenine ait tamamlanmış (complete) 380, tamamlanmamış (incomplete, censored) 83 gözlem bulunmaktadır.

4.1.YAŞAM TABLOSU YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ

Yaşam Tablosu yöntemi ile yaşam süresi verilerini eşit zaman aralıklarına göre frekans tablosuna dönüştürmek ve her bir zaman aralığı için yaşam fonksiyonlarını hesaplayarak analiz etmek mümkündür. Gözlem sayısının 100'ün üzerinde olduğu ve yaşam süresinin en az 6 eşit aralığa bölündüğü durumlarda kullanılması önerilen bir yöntemdir.

Veri seti üzerinde “Yaşam Tablosu” yöntemi uygulandığında elde edilen sonuçlar Çizelge 4.1.'de özetlenmiştir.

Çizelge 4.1. Yaşam Tablosu Sonuçları

Life Table													
		Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval	Probability Density	Std. Error of Probability Density	Hazard Rate	Std. Error of Hazard Rate
First-order Controls	Interval Start Time												
tedavi diyaliz	0	244	0	244.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	35	244	0	244.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	70	244	1	243.500	3	.01	.99	.99	.01	.000	.000	.00	.00
	105	240	8	236.000	41	.17	.83	.82	.03	.005	.001	.01	.00
	140	191	26	178.000	101	.57	.43	.35	.03	.013	.001	.02	.00
	175	64	2	63.000	56	.89	.11	.04	.01	.009	.001	.05	.00
	210	6	1	5.500	4	.73	.27	.01	.01	.001	.000	.03	.01
	245	1	1	.500	0	.00	1.00	.01	.01	.000	.000	.00	.00
ilaç	0	109	0	109.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	35	109	0	109.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	70	109	2	108.000	3	.03	.97	.97	.02	.001	.000	.00	.00
	105	104	7	100.500	19	.19	.81	.79	.04	.005	.001	.01	.00
	140	78	10	73.000	42	.58	.42	.33	.05	.013	.001	.02	.00
	175	26	2	25.000	23	.92	.08	.03	.02	.009	.001	.05	.01
	210	1	0	1.000	1	1.00	.00	.00	.00	.001	.001	.06	.00
	245	0	0	.000	0	.00	1.00	.00	.00	.000	.000	.00	.00
ameliyat+ilaç	0	110	0	110.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	35	110	0	110.000	0	.00	1.00	1.00	.00	.000	.000	.00	.00
	70	110	0	110.000	2	.02	.98	.98	.01	.001	.000	.00	.00
	105	108	8	104.000	22	.21	.79	.77	.04	.006	.001	.01	.00
	140	78	12	72.000	41	.57	.43	.33	.05	.013	.001	.02	.00
	175	25	1	24.500	22	.90	.10	.03	.02	.009	.001	.05	.01
	210	2	2	1.000	0	.00	1.00	.03	.02	.000	.000	.00	.00
	245	0	0	.000	0	.00	1.00	.00	.00	.000	.000	.00	.00

Yukarıdaki tabloda tedavi yöntemlerine göre yaşam sürelerinin tahmini için kullanılan fonksiyona ilişkin sonuçlar sunulmuştur. Üç farklı tedavi yöntemine ilişkin kurulacak olan hipotez aşağıdaki gibidir:

H_0 : Tedavi yöntemleri arasında istatistiksel olarak anlamlı bir fark yoktur.

H_1 : Tedavi yöntemleri arasındaki fark istatistiksel olarak anlamlıdır.

Hipotezi test etmek için kullanılan Wilcoxon (Gehan) test sonucu Çizelge 4.2.'deki gibi elde edilmiştir:

Çizelge 4.2. Wilcoxon Testi

Wilcoxon (Gehan) Statistic	df	Sig.
1.520	2	.468

Sonucu yorumlamak için elde edilen Wilcoxon istatistiğini (1.520) serbestlik derecesi 2 olan Wilcoxon tablo değeri ile karşılaştırmak veya anlamlılık değeri olan 0.468 'i 0.05 ile karşılaştırmak mümkündür. Her iki durumda da H_0 hipotezinin kabul edildiği ve tedavi yöntemlerinin birbirinden farklılık göstermediği söylenir. Yani tedavi yöntemleri arasında istatistiksel olarak anlamlı bir fark yoktur. Bu durumda tıbbi açıdan izlenebilecek en iyi yolun hastayı fiziksel olarak en az etkileyecek yöntemi tercih etmek olmalıdır.

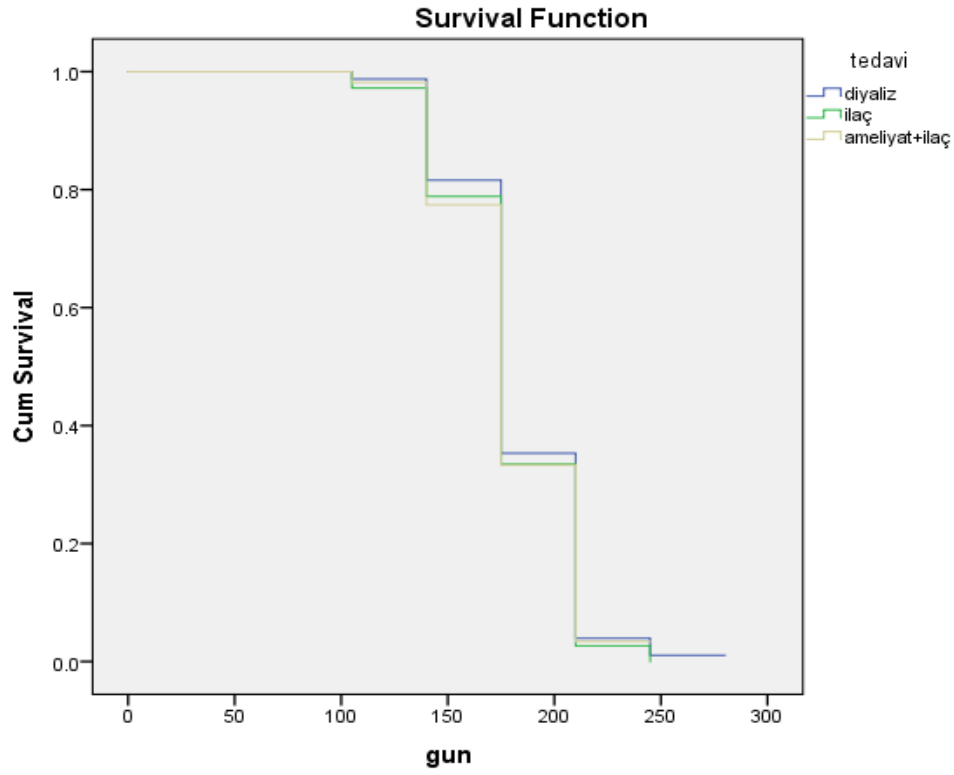
Üç farklı tedavi şekli için elde edilen medyan yaşam süreleri Çizelge 4.3.' de özetlendiği gibidir.

Çizelge 4.3. Medyan Yaşam Süreleri

First-order Controls		Med Time
tedavi	diyaliz	163.8913
	ilaç	162.2540
	ameliyat+ilaç	161.7648

Çizelgeden de görüldüğü gibi uygulanan farklı tedavi tiplerine ilişkin yaşam süresi tahminleri birbirine oldukça yakın değerlerdir. Tedavi yöntemleri arasında farklılık olmadığını ifade eden H_0 hipotezini destekleyen sonuçlarla karşılaşılmış ve tahmini yaşam sürelerinin, hangi tür tedavi uygulanırsa uygulansın çok büyük farklılık göstermediği görülmüştür. Üç farklı tedavi türü olan diyaliz, ilaç ve ameliyat+ilaç tedavilerinden hangisi seçilirse seçilsin, tahmini yaşam süresi üzerinde bir etkisi olmadığı için hastayı en az etki altında bırakacak tedavi şeklinin tercih edilmesi gerekmektedir.

Yaşam fonksiyonlarının farklı tedavi türlerine göre gösterimi Şekil 4.1.'deki gibidir.



Şekil 4.1. Farklı Tedavi Türleri için Yaşam Fonksiyonu

Yaşam fonksiyonuna göre çizilen bu grafik ile farklı tedavi türlerine ilişkin dağılımların birbirinden yine çok büyük bir farklılık göstermediği, hatta çizgilerin neredeyse üst üste binerek farklılığın olmadığı görsel olarak da ifade edilebilir.

4.2. KAPLAN MEIER YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ

Kaplan Meier Yöntemi ve Yaşam Tablosu arasındaki en önemli fark, Kaplan Meier yönteminde araştırmacının belirlediği zaman aralıkları çerçevesinde değil gerçekleşen her ölümden beliren, yani eşit olmayan zamanlarda tahmin yapmasıdır. Bir başka deyişle her ölüm olgusu zamanı kadar tahmin noktası oluşturulur. Yönteme göre elde edilen sonuçlar Çizelge 4.4.'de özetlenmiştir.

Çizelge 4.4. Kaplan Meier Tahminleri

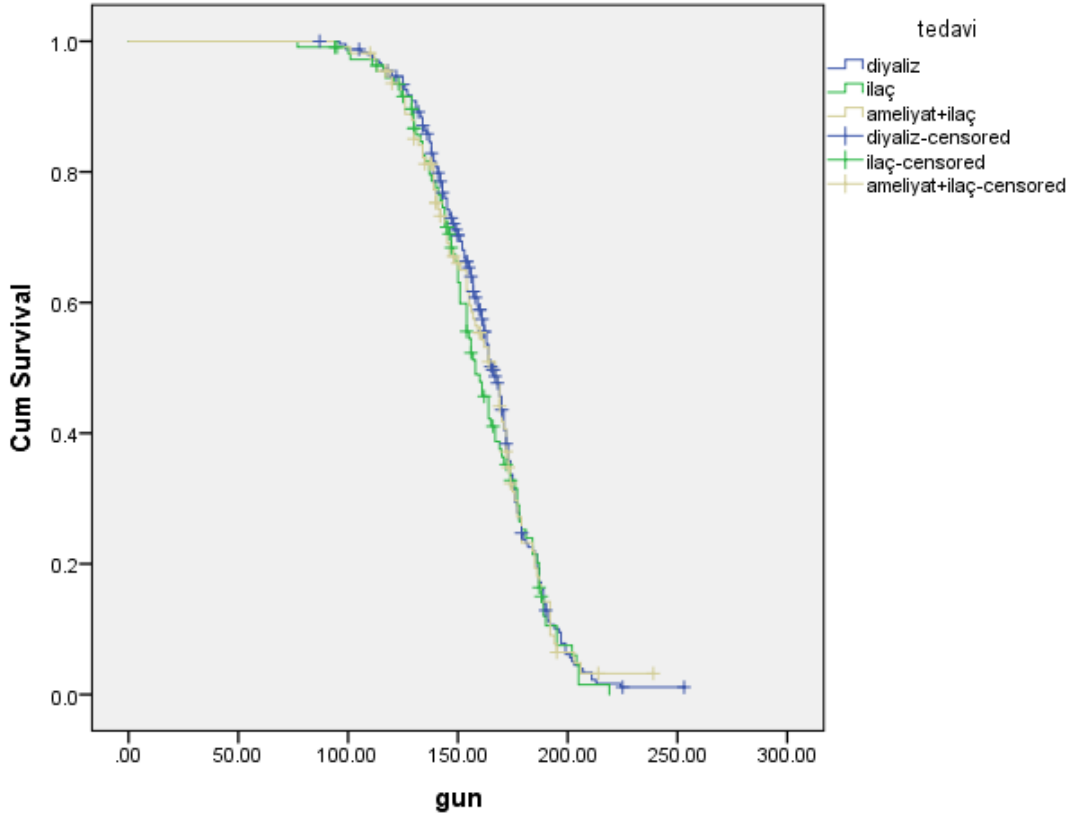
tedavi	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
diyaliz	164.086	1.748	160.660	167.511	166.000	2.248	161.594	170.406
ilaç	160.204	2.704	154.904	165.503	158.000	3.821	150.511	165.489
ameliyat+ilaç	162.469	2.853	156.876	168.062	167.000	4.536	158.110	175.890
Overall	162.853	1.318	160.269	165.437	164.000	1.825	160.423	167.577

Kaplan Meier yöntemine göre tahmin edilen ortalama ve medyan yaşam süreleri, bu sürelerle ait standart hatalar ve %95 güven aralıkları Çizelge 4.4.'deki gibidir. Bu sonuçlar göz önünde bulundurulduğunda, tedavi yöntemleri arasında farklılık olmadığını ifade eden H_0 hipotezini test etmek için Log_Rank testi kullanılmıştır. Serbestlik derecesi 1 olan Ki-Kare tablo değeri 0.185 ile karşılaştırıldığında ya da 0.667 anlamlılık değeri 0.05 ile karşılaştırıldığında, tedavi yöntemleri arasında istatistiksel olarak anlamlı bir fark bulunmamıştır. Çizelge 4.5.'de elde edilen sonuç özetlenmiştir.

Çizelge 4.5. Log-Rank Testi

Overall Comparisons			
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	.185	1	.667

Tahmini kümülatif yaşam sürelerini göz önünde bulundurularak çizilen Şekil 4.2.'de, tedavi türü farklılığının yaşam süresi üzerinde herhangi bir etkisi olmadığı, bir başka deyişle yaşam süresini uzatmadığını söylemek mümkündür.



Şekil 4.2. Kümülatif Yaşam Sürelerine Göre Yaşam Süreleri Tahminleri

4.3. COX REGRESYON YÖNTEMİ İLE BÖBREK YETMEZLİĞİ ANALİZİ

Cox Regresyon yöntemi ile tamamlanmamış izlem verilerinde bağımlı değişken ve bağımsız değişkenler (yaş, cinsiyet, tedavi türü) arasındaki neden-sonuç ilişkisi analiz edilebilir. Yaşam analizi verilerinde bağımsız değişkenler arasında korelasyon olmakla beraber orantısız ilişkiler de içermektedir. Klasik regresyon analizinde sağlanması gereken bir varsayım olan ardışık bağımsızlık burada sağlanmamakta ve araştırmacıyı farklı yöntemlerin arayışına yönlendirmektedir. Ayrıca bağımlı değişkenin de normal

dağılıma sahip olması yine klasik regresyonun gerektirdiği bir varsayımdır. Ancak yaşam verilerinde sağlanamayan bu varsayımlar, nedensellik analizinde Cox Regresyon yönteminin kullanımını gerektirmektedir.

İlk olarak yine sıfır hipotezinin anlamlılığını test etmek üzere elde edilen Cox Regresyon çıktılarına ulaşmak gerekmektedir. -2Log Likelihood için elde edilen anlamlılık değeri 0.652, 0.05 ile karşılaştırıldığında tedavi yöntemleri arasındaki farkın anlamsız olduğu ifade edilir. Cox regresyon sonuçları Çizelge 4.6.'da özetlenmiştir.

Çizelge 4.6. -2 Log Likelihood Testi

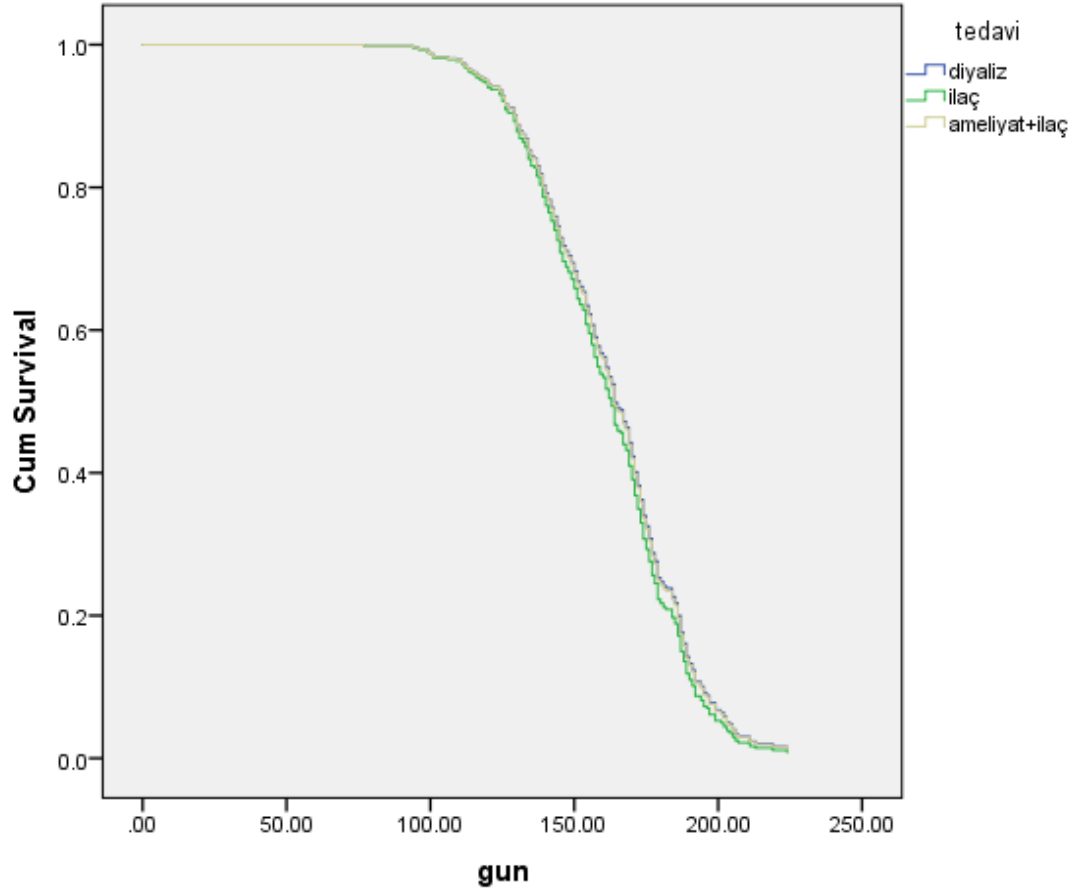
-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
3897.165	1.634	3	.652	1.641	3	.650	1.641	3	.650

Cox regresyon analizi ile elde edilen daha ayrıntılı sonuçlar ile tedavi yöntemlerine ilişkin sonuçları analiz etmek mümkündür. Cox regresyon modelinde kullanılan, tedavi şekli ve yaş prognostik değişkenlerine ilişkin elde edilen model katsayıları ve anlamlılık değerleri Çizelge 4.7.'de gösterildiği gibidir. Yaşın yaşam uzunluğu üzerinde anlamlı bir etkisi olmadığı 0.335 değerinin 0.05 anlamlılık seviyesinden büyük olması ile açıklanabilir. Benzer şekilde tedavi tipinin de yaşam süresi üzerinde bir etkisi yoktur. Elde edilen 0.780 anlamlılık seviyesi bunu açıkça ortaya koymaktadır. Aşağıdaki tabloda yer alan $Exp(B)$ değeri ODDs oranını belirtmektedir. $Exp(B)$ değerine bakarak şu açıklamayı yapmak mümkündür: hiçbir tedavi şekli yaşam süresi üzerinde etkili olacak kadar önemli olmamakla beraber ameliyat + ilaçla tedavi (tedavi 2) diğer tedavi şekillerine göre daha etkindir. Yaş değişkenini de bu karşılaştırmaya katmak mümkündür.

Çizelge 4.7. Cox Regresyon Sonuçları

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
tedavi			.498	2	.780			
tedavi(1)	-.011	.130	.007	1	.935	.989	.766	1.277
tedavi(2)	.079	.151	.269	1	.604	1.082	.804	1.456
yaş	-.004	.004	.930	1	.335	.996	.989	1.004

Bu açıklamalar ışığında, Cox regreyon yöntemi kullanılarak elde edilen grafik aşağıdaki gibidir. Tüm sonuçlar tedavi şeklinin yaşam süresi üzerinde herhangi bir etkisi olmadığını göstermektedir.



Şekil 4.3. Cox Regresyon Eğrisi

5. SONUÇ

Bu çalışmada, böbrek yetmezliği sorunu yaşayan 463 hastanın, uygulanan üç farklı tedavi yönteminden sonra gözlenen yaşam süreleri karşılaştırılarak, prognostik faktör olan tedavi tipi ve yaş değişkeninin yaşam süresi üzerindeki etkileri Cox regresyon yöntemi ile incelenmiştir. 253 günlük izlem süresinden sonra veriler yaşam analizi yöntemleri ile analiz edilerek, yaşam fonksiyonları yardımıyla incelenen veri setinde tedavi yöntemleri arasında istatistiksel olarak anlamlı bir fark olmadığına karar verilmiştir. Bir başka deyişle tedavi tipinin yaşam süresi üzerinde herhangi bir etkisi bulunmamaktadır. Yaşam tablosu, Kaplan Meier ve Cox regresyon yöntemlerinin üçünde de tedavi yöntemleri arasında fark yoktur şeklinde kurulan hipotez kabul edilmiş ve uygulama kısmında ayrıntıları ile açıklanmıştır. ODD_s oralarından da, hiçbir tedavi tipinin yaşam süresi üzerinde etkili olacak kadar önemli olmaydığı ancak en etkin olan tedavi tipinin ameliyat + ilaçla tedavi (tedavi 2) olup diğerlerine göre az da olsa yaşam süresi üzerinde daha olumlu bir etkisinin olduğunu görülmüştür.

KAYNAKLAR

Aitchison, J. and Brown, J.A.C., 1957. The Lognormal Distribution, Cambridge University Press, Cambridge.

Albonetti, C.A., and Hepburn, J.R., 1997. Probation revocation: A proportional hazards model of the conditioning effects of social disadvantage, *Social problems*, **44(1)**, 124-138.

Ata, N., 2010. Orantısız tehlikeler için yaşam modelleri, *Doktora Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Ata, N., 2005. Yaşam çözümlemesinde orantısız hazard modeli, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Başar, E., 1993. Yaşam tabloları analizinde kullanılan bazı istatistiksel tekniklerin böbrek nakli verilerine uygulanması, *Doktora Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Boag, J. W., 1949. Maximum likelihood estimates of proportion of patients cured by cancer therapy, *Journal of the Royal Statistical Society, Series B*, **11**, 15-44.

Collett, D., 2003. Modelling survival data in medical research. CRC press.

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

Cox, D. R., 1972. Regression models and Life Tables (with discussion), *Journal of the Royal Statistical Society*, **34**, 187-220.

Çekçeki Göz, Ö., 2007. Orantılı Hazard Modelinin Zamana Bağlı Değişkenlerle Genişletilmesi ve Çocuk Suçluluğu Üzerine bir Uygulama, *Yüksek Lisans Tezi*, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Ersoy, E., 2005. Sağlık Analizinde Parametre Tahmin Problemlerine Katkıları, *Yüksek Lisans Tezi*, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.

Fox, J. ., 2006. Introduction to Survival Analysis. Sociology 761

- Gaddum, J.H.**, 1945a. Log normal distributions, *Nature*, **156**, 463-466.
- Gaddum, J.H.**, 1945b. Log normal distributions, *Nature*, **156**, 747.
- Graunt, J.**, 1662. Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality. London: John Martyn and James Allestry.
- Fichman, M.**, 1988. The Motivational Consequences of Absence and Attendance: Proportional Hazard Estimation of a Dynamic Motivational Model, *Journal Of Applied Psychology*, **1**, 119-34.
- Frei, E., et al.** 1961. Studies of sequential and combination antimetabolite therapy in acute leukemia: 6-mercaptopurine and methotrexate, *Blood*, **18(4)**, 431-454.
- Greene, W. H.**, 2003, Econometric analysis. Pearson Education, India.
- Greenwood, M.**, 1926. The natural duration of cancer, *Reports of Public Health and Medical Subjects, H.M. Stationery Office, London*, **33**, 1-26.
- Guo, S.**, 2010, Survival analysis. Oxford University Press, New York.
- Hannan, M. T., and Tuma, N. B.**, 1984. Social Dynamics: Models and Methods.
- Heckman, J. J., and Singer, B.**, 1985, Social science duration analysis, *Longitudinal analysis of labor market data*, 39-110.
- Işık, A.**, 2007. Cox Regresyon Modeli ile Finansal Başarısızlığın belirlenmesi üzerine bir uygulama, *Yüksek Lisans Tezi*, Marmara Üniversitesi Sosyal Bilimleri Enstitüsü, İstanbul.
- İnceoğlu, F.**, 2013. Sağlık analizi yöntemleri ve karaciğer nakli verileri ile bir uygulama, *Yüksek Lisans Tezi*, İnönü Üniversitesi Sağlık Bilimleri Enstitüsü, Malatya.
- Kalbfleisch, J.D., and Prentice, R.L.**, 1980, The statistical analysis of failure time data, John Wiley & Sons, New York.
- Kaplan, E.L., and Meier, P.**, 1958. Nonparametric estimation from incomplete observations, *Journal of the American statistical association*, **53(282)**, 457-481.
- Karagöz, Y.**, 2015, SPSS 22 uygulamalı Biyoistatistik, Nobel Akademik Yayıncılık, Ankara.
- Karasoy, D. ve Tutkun Ata, N.**, 2016. Yaşam Çözümlemesi, Nobel Yayınevi, Ankara.
- Kleinbaum, D. G., & Klein, M.**, 2005. Survival analysis: a self-learning text. Springer Science & Business Media.
- Kurt, İ.**, 2008. Bayesgil yaşam analizi ve Cox regresyon yaşam analizinin türetilmiş ve gerçek veri setlerinde uygulanması, *Doktora Tezi*, Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Eskişehir.
- Lawless, J.**, 1982. Statistical Models and Methods for Lifetime Data, Wiley, New York.
- Lee, E.T., and Wang, J.**, 2003. Statistical methods for survival data analysis, John Wiley & Sons.
- Lehrer, E.L.**, 1988, Determinants of marital instability: a Cox-regression model, *Applied Economics*, **20(2)**, 195-210.
- Liu, X.**, 2012. Survival analysis: models and applications. John Wiley & Sons.

- Lloyd D. Fisher and D.Y. Lin.**,1999. Time Dependent Covariates in the Cox Proportional-Hazards regression model, Department of Biostatistics, University of Washington, Seattle, Washington 98195-7232;
- Marubini, E., and Valsecchi, M.G.**, 2004. Analysing survival data from clinical trials and observational studies (Vol. 15). John Wiley & Sons.
- MacDonald, E.J.**, 1963. The epidemiology of melanoma, *Annals of the New York Academy of Sciences*, **100**, 4-15.
- McAlister, D.**, 1879. The law of the geometric mean, *Proceedings of the Royal Society*, **29**, 367.
- Miller, R.G.**, 1981. Survival Analysis, John Wiley & Sons.
- Nelson, W.**, 1982. Applied Life Data Analysis, John Wiley & Sons Inc., Canada.
- Özdamar, K.**, 2003. SPSS ile Biyoistatistik, Kaan Kitabevi, Eskisehir.
- Özşen, D.**, 2006. Evlilik süresini etkileyen faktörlerin Cox Regresyon ile analizi, *Yüksek Lisans Tezi*, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Schoenfeld, D.**, 1982. Partial residuals for the proportional hazards model, *Biometrika*, **69**, 551-555.
- Sertkaya, D., Ata, N., ve Sözer, M.T.**, 2005. Yaşam çözümlemesinde zamana bağlı açıklayıcı değişkenli Cox regresyon modeli. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, **58(1)**, 153-158.
- Sümbüloğlu, K., ve Sümbüloğlu, V.**, 2009, Biyoistatistik. Hatipoğlu Yayınevi, Ankara
- Şenocak, M.Ş.**, 2009. Temel kavramlar ve klinik biyoistatistik, Nobel Tıp Kitabevleri, İstanbul.
- Tamam, D.**, 2008. Tam ve sansürlü örneklem durumlarında Weibull dağılımı için bazı istatistiksel sonuç çıkarımları, *Yüksek Lisans Tezi*, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Terzi, Y.**, 2003. Sansürlü veriler için sağkalım analizi ve gerçek verilere uygulaması, *Doktora Tezi*, Ondokuz Mayıs Üniversitesi Fen Bilimleri Üniversitesi, Samsun.
- Yay, M., Çoker, E. ve Uysal, Ö.**, 2007. Yaşam analizinde cox regresyon modeli ve artıkların incelenmesi, *Cerrahpaşa Tıp Dergisi*, **38**,139-145.
- Weibull, W.**, 1951. A statistical distribution of wide applicability, *Journal of Applied Mathematics*, **18**, 293-297.
- Weibull, W.**, 1939, A statistical theory of the strength of materials, *Ingenioers vetenskaps akademien Handlingar*, **151**, 293-297.
- Yamaguchi, K.**, 1991. Event history analysis (Vol. 28). Sage Publications.

