



**MARMARA ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**COX REGRESYON MODELİNDE SAĞKALIM  
FONKSİYON PARAMETRELERİNİN DÜZELTİLMİŞ  
DENEYSEL OLABİLİRLİK YÖNTEMİ İLE TAHMİN  
EDİLMESİ**

---

KÜBRA ORAN

**YÜKSEK LİSANS TEZİ**

Matematik Anabilim Dalı

**DANIŞMAN**

Doç. Dr. Nilgün YILDIZ

İSTANBUL, 2020



**MARMARA ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**COX REGRESYON MODELİNDE SAĞKALIM  
FONKSİYON PARAMETRELERİNİN DÜZELTİLMİŞ  
DENEYSEL OLABİLİRLİK YÖNTEMİ İLE TAHMİN  
EDİLMESİ**

---

KÜBRA ORAN

(520917006)

**YÜKSEK LİSANS TEZİ**

Matematik Anabilim Dalı

**DANIŞMAN**

Doç. Dr. Nilgün YILDIZ

İSTANBUL, 2020

## **ÖNSÖZ/TEŞEKKÜR**

Yüksek lisans eğitim süresince hem ders hem de tez aşamasında karşılaştığım tüm soru ve sorunlarımda bilgisi, tecrübesi ve önerileriyle bana yol gösteren, desteklerini hiçbir zaman esirgemeyen ve tüm çalışmalarımı her an yanımda olduğunu hissettiren danışmanım ve değerli hocam Doç. Dr. Nilgün YILDIZ'a tüm yardımları ve teşvikleri için sonsuz teşekkürlerimi sunuyorum. Ayrıca, yüksek lisans öğrenim süresince verdikleri tüm bilgiler ve yardımları için bölümümüz tüm hocalarına teşekkür ediyorum.

Son olarak, öğrenim hayatım boyunca ve aldığım kararlarda her daim maddi ve manevi desteğini esirgemeyen aileme sonsuz teşekkür ediyorum.

**Ağustos 2020**

**Kübra ORAN**

# İÇİNDEKİLER

<b>ÖNSÖZ/TEŞEKKÜR</b>	i
<b>ÖZET</b>	iv
<b>ABSTRACT</b>	v
<b>YENİLİK BEYANI</b>	vi
<b>SEMBOLLER/SYMBOLS</b>	vii
<b>KISALTMALAR/ABBREVIATIONS</b>	viii
<b>ŞEKİL</b>	ix
<b>TABLO LİSTESİ/LISTS OF TABLE</b>	x
<b>1.GİRİŞ/INTRODUCTION</b>	1
<b>2.SAĞKALIM ANALİZİ</b>	3
2.1. Sağkalım Analizinde kullanılan Dağılımlar	4
2.1.1. Kesikli dağılımlar	4
2.1.2. Sürekli dağılımlar	5
2.1.2.1. Önemli dağılımlar	6
a. Üstel dağılım	6
b. Weibull dağılımı	7
c. Log-Normal dağılım	8
d. Log-Lojistik dağılımı	9
2.2. Veri Türleri	9
2.2.1. Sansürsüz veri	10
2.2.2. Sansürlü veri	10
2.2.2.1. Sağdan sansürlü	10
2.2.2.2. Soldan sansürlü	11
2.2.2.3. Aralık sansürlü	11
2.3. Sağkalım Parametre Tahmin Yöntemleri	12
2.3.1. Kaplan-Meier yöntemi	12
2.3.2. Nelson-Aalen yöntemi	15
2.3.3. Yaşam tablosu yöntemi	12
2.4. Parametre Tahmin Yöntemleri	18
2.4.1. Parametrik sağkalım fonksiyonu için tahmin yöntemi	12

2.4.1.1. En Çok Olabilirlik Yöntemi	19
2.4.1.1.1. Sağdan sansürlü veriler için en çok olabilirlik yöntemi	20
2.4.1.1.2. Sansürlü veriler (sağ, sol ve aralık) için en çok olabilirlik yöntemi	21
2.4.2. Yarı parametrik sağkalım fonksiyonu için tahmin yöntemi	21
2.4.2.1. Cox yöntemi	21
2.4.2.1.1. Cox modeli	22
2.4.2.1.2. Cox kısmi olabilirlik fonksiyonu	22
2.4.2.2 $\beta$ katsayısının önemliliğinin test edilmesi	23
2.4.3. Parametrik olmayan sağkalım fonksiyonu için tahmin yöntemi	24
2.4.3.1. Deneysel Olabilirlik Yöntemi	24
2.4.3.1.1. Sansürsüz örneklem	25
2.4.3.1.2. Sansürlü örneklem	27
2.4.3.1.3. Tehlikeye bağlı olarak deneysel olabilirlik yöntemi	24
<b>3. COX-DENEYSEL OLABİLİRLİK</b>	31
3.1. Düzeltmiş Deneysel Olabilirlik Güven Bölgesi	32
<b>4. UYGULAMA</b>	20
4.1. Uygulama 1(Gastrit kanser verisi)	40
4.2. Uygulama 2(Akciğer kanseri verisi)	44
4.3. Uygulama 3(Böbrek kateter verisi)	52
<b>5. SİMÜLASYON</b>	57
<b>6. SONUÇ</b>	61
<b>KAYNAKLAR/REFERENCES</b>	62

## ÖZET

# COX REGRESYON MODELİNDE SAĞKALIM FONKSİYON PARAMETRELERİNİN DÜZELTİLMİŞ DENEYSEL OLABİLİRLİK YÖNTEMİ İLE TAHMİN EDİLMESİ

Özellikle son birkaç yılda yapılan çalışmalar ile birlikte sosyoloji, kriminoloji, pazarlama gibi birçok farklı alanlarda da kullanılmaya başlayan sağkalım analizi, belli bir zaman aralığında ilgilenilen veya belirlenen bir olayın meydana gelme süresi olarak söylenir. Olaylar farklı olduğundan kullandığımız veri yapısı farklı olacaktır. Bu sebepten, sağkalım analizinde sansürlü ve sansürsüz olmak üzere iki tür veri türü bulunmaktadır. Sağkalım parametrelerinin tahmin etmek için Kaplan-Meier, Nelson-Aalen ve Yaşam Tablosu yöntemleri kullanılmaktadır. Parametre tahmini yapılırken, sağkalım dağılım türlerine göre farklı tahmin yöntemleri yapılmaktadır.

İkinci bölümde sağkalım analizi, veri yapısı, parametre tahmin yöntemleri ve artıklar hakkında bilgi verilmiştir. Üçüncü bölümde, düzeltilmiş deneysel olabilirlik yöntemi yardımıyla güven bölgesi oluşturulması yapıldı. Dördüncü bölümde, gerçek üç veri üzerinde Cox model uygulaması verildi. Beşinci bölümde simülasyon çalışması yapılmıştır. Altıncı bölümde bulunan sonuçlar paylaşılmıştır.

Bu çalışmada amacımız, Cox modelinde (Yarı parametrik) sağkalım dağılımına sahip fonksiyonun Düzeltilmiş Deneysel Olabilirlik ve Deneysel Olabilirlik yöntemleri ile elde edilen kapsama olasılıkları simülasyon çalışması ile kıyaslama yapılmıştır.

## **ABSTRACT**

### **ESTIMATION OF SURVIVAL FUNCTION PARAMETERS VIA ADJUSTED EMPIRICAL LIKELIHOOD METHOD IN COX REGRESSION MODEL**

The survival analysis, which has been used in many different fields such as sociology, criminology and marketing, especially with the studies carried out in the last few years, is said to be the period of occurrence of an event that is interested or determined in a certain period of time. Since the events are different, the data structure we use will be different. For this reason, there are two types of data in survival analysis: censored and uncensored. Kaplan-Meier, Nelson-Aalen and Life Table methods are used to estimate survival parameters. While estimating the parameters, different estimation methods are made according to the survival distribution types.

In the second part, information about survival analysis, data structure, parameter estimation methods and residues is given. In the third section, a confidence region was created with the help of adjusted empirical likelihood method. In the fourth chapter, Cox model application is given on three real data. In the fifth part, simulation study has been done. The results in the sixth section are shared.

In this study, our aim was to compare the coverage possibilities of the function with the Cox model (Semi-parametric) survival distribution with the Adjusted Empirical Likelihood and Empirical Likelihood methods with the simulation study.

## YENİLİK BEYANI

Simülasyon çalışmamızda kümülatif temel tehlike fonksiyonu Log-lojistik dağılım ile dağılma özelliği kullanılmıştır.



## SEMBOLLER/SYMBOLS

<b>C</b>	: Sansürleme süresi
<b>f</b>	: Olasılık yoğunluk fonksiyonu
<b>F</b>	: Kümülatif (birikimli) dağılım fonksiyonu
<b>h</b>	: Tehlike fonksiyonu
<b>H</b>	: Kümülatif (birikimli) tehlike fonksiyonu
<b>L</b>	: Olabilirlik fonksiyonu
<b>l</b>	: Cox kısmi olabilirlik fonksiyonu
<b>M</b>	: Martingale artık
<b>S</b>	: Sağkalım fonksiyonu
<b>T</b>	: Sağkalım süresi
<b>t</b>	: Zaman (gün)
<b>U</b>	: Skor fonksiyonu
<b><math>\Gamma</math></b>	: Gamma fonksiyonu
<b><math>\phi</math></b>	: Standart normal dağılım

## KISALTMALAR/ABBREVIATIONS

<b>EL</b>	: Empirical Likelihood (Deneysel olabilirlik)
<b>k.d.f</b>	: Kümülatif dağılım fonksiyonu
<b>KM</b>	: Kaplan-Meier
<b>LR</b>	: Likelihood Ratio (Benzerlik oranı)
<b>NA</b>	: Nelson-Aalen
<b>o.y.f</b>	: Olasılık yoğunluk fonksiyonu



## ŞEKİL

Şekil 2.1. Sağdan sansürlü veri örneği .....	11
Şekil 2.2. Soldan sansürlü veri örneği .....	11
Şekil 2.3. Aralık sansürlü veri örneği.....	12
Şekil 4.1. Tedavi yöntemlerine göre sağkalım olasılıkları .....	42
Şekil 4.2. Güven aralığı .....	43
Şekil 4.3. Cinsiyete göre sağkalım olasılıkları.....	50
Şekil 4.4. ECOG performansına göre sağkalım olasılıkları .....	50
Şekil 4.5. Cinsiyete göre sağkalım olasılıkları.....	54
Şekil 4.6. Hastalık türüne göre sağkalım olasılıkları .....	55
Şekil 4.7. Güven aralığı .....	56

## TABLO LİSTESİ/LIST OF TABLES

Tablo 2.1. Örnek yaşam tablosu.....	17
Tablo 4.1. Tedavi gruplarının hasta karakterleri.....	40
Tablo 4.2 Gastrit kanseri verisi R sonuçları .....	41
Tablo 4.3 $\beta = 0.2$ ve $\lambda = 0$ için EL .....	42
Tablo 4.4. $\beta = \hat{\beta}$ ve $\lambda = 0$ için EL.....	42
Tablo 4.5. $\beta = 0.2$ ve $\lambda = -1.62$ için EL .....	43
Tablo 4.6. Akciğer kanseri hastalarının verisi .....	44
Tablo 4.7. Akciğer kanseri verinin R sonuçları .....	49
Tablo 4.8. Beta tahminleri ve güven aralıkları .....	50
Tablo 4.9. Böbrek kateter verisi.....	52
Tablo 4.10. Böbrek kateter verisinin R sonuçları .....	53
Tablo 4.11. Böbrek kateter verisinin R sonuçları 2 .....	54
Tablo 4.12. $\beta$ katsayıları için güven aralıkları.....	54
Tablo 4.13. $\beta = -0.2$ ve $\lambda = 0$ için EL.....	55
Tablo 4.14. $\beta = \hat{\beta}$ ve $\lambda = 0$ için EL .....	56
Tablo 4.15. $\beta = -0.2$ ve $\lambda = 14.185$ için EL .....	56
Tablo 5.1. Weibull dağılımlı simülasyon için kapsama olasılıkları.....	58
Tablo 5.2. Log-Lojistik dağılımlı simülasyon için kapsama olasılığı.....	60

## 1. GİRİŞ

Başlangıcı 17.yy'la kadar uzanan sağkalım analizi, belli bir zaman aralığında ilgilenilen veya belirlenen bir hastalığın başlaması, nüksetmesi veya ölüm gibi durumları gösteren olayın meydana gelme süresi olarak söylenir. Sağkalım analizi için kullanılan sağkalım verisi ile yapılan çalışmalar ile hastanın ilgili olaya yanıt verme olasılığının ve sağkalım-ortalama yaşam süresinin tahmin edilmesi gibi sonuçların elde edilmesinde kullanılmaktadır. [8] Uzun yıllar boyunca ölümlerin araştırılması olarak düşünülen sağkalım analizi, özellikle son birkaç yılda yapılan çalışmalar ile birlikte sosyoloji, kriminoloji, ekonomi, endüstri gibi birçok farklı alanlarda da kullanılmaya başlanmıştır.

Sağkalım analizi belli bir zamandaki olayların verilerinin incelemesidir. Buradaki olay hastalıktan kurtulma, ölüm gibi durumlar olabilirken deprem, evlilik süresi, elektronik aletlerin süreleri birçok zamanı içeren veriler üzerinde incelemeler yaparız. Olaylar farklı olduğundan kullandığımız veri yapısı da farklı olacaktır. Sağkalım verisinde; sağkalım süresi, hasta karakter özelliği gibi birçok veri sonucu bulunabilir. Sağkalım süresi, çalışma sonunda belirlenen olayın meydana gelme süresidir. Sağkalım analizinde çalışma sonunda olayın gerçekleşmesiyle ilişkili olarak sansürlü ve sansürsüz olmak üzere iki veri türü bulunmaktadır. Sansürlü veriler sağdan, soldan ve aralık sansürlü olmak üzere üç grupta incelenmektedir.

Sağkalım parametrelerinin tahmin edilmesinde literatürde birçok yöntem bulunmaktadır. Fakat en yaygın olarak kullanılan yöntemler: Kaplan-Meier, Nelson-Aalen ve Yaşam Tablosu yöntemi kullanılmaktadır. Sağkalım fonksiyonunun parametrik olmayan tahmin edicisi Kaplan-Meier tahmin edicisi genellikle tıp/halk sağlığı çalışmalarında kullanılan tahmin edicidir ve Çarpım-Limit (Product-Limit) tahmin edicisi olarak da ifade edilmektedir. Kaplan-Meier özellikle sağkalım olasılıkları ve tehlike oranının tahmini üzerine çalışmaları bulunmaktadır. Kümülatif tehlike fonksiyonunun tahmin edilmesinde en yaygın kullanılan tahmin edici Nelson-Aalen yöntemidir. Sağkalım ve sansürleme süreleri aralıklar halinde gruplanan örneklem veya kohort bireylerin sağkalım verisinin resmedilmesi için kullanılan en eski yöntem ise Yaşam Tablosu'dur.[7] Sağkalım analizinin en eski yöntemi olan Yaşam Tablosu yöntemi genellikle örneklem sayısı az verilerde kullanılır.

Parametre tahmini yapılırken sağkalım dağılım türüne göre uygulama yapılmaktadır. Sağkalım analizinde parametrik, yarı parametrik ve parametrik olmayan olmak üzere üç farklı sağkalım dağılım türü bulunmaktadır. Parametre tahmini yapılırken eğer dağılım parametrik sağkalım dağılımına sahip ise En Çok Olabilirlik yöntemini

kullanılır. Burada parametrenin hangi dağılım ailesine ait olduğu bilinmektedir. İkinci olarak, yarı parametrik sağkalım dağılımına ait ise Cox yöntemi ile parametre tahminini yaparız. 1972’de David Cox tarafından ileri sürülen bu yöntem parametrik ve parametrik olmayan dağılıma ait parametrelerden oluştuğu için yarı parametrik olarak ifade edilmektedir. Bu modelin en önemli özelliği,  $t$ 'ye bağlı olmayan iki tehlike fonksiyonunun oranı sabittir. Son olarak parametrik olmayan sağkalım dağılımına ait parametrenin tahmin edilmesinde Deneysel Olabilirlik yöntemi kullanılmaktadır. Deneysel olabilirlik yöntemi, dağılım varsayımı yapılmaksızın çıkarsama ve tahmin etme yöntemidir. Başka bir deyişle, parametrenin dağılım ailesi bilinmeden sadece veri odaklı olup olabilirlik oran fonksiyonuna bağlı olarak parametre tahmini yapılmasıdır. Bu yöntem, parametrik olmayan en çok olabilirlik tahmin işlemi olarak söylenebilir. Maksimum olabilirlik gibi benzer aşamaları vardır, fakat hata dağılımı için daha az varsayıma ihtiyaç duyar. Ayrıca bu yöntemde, verinin bağımsız ve aynı dağılıma sahip olduğu düşünülür. 1988’de Owen’ın öncülük etmesiyle geliştirilen deneysel olabilirlik yöntemi, her geçen gün artan araştırmacı sayısı ile daha kapsamlı uygulamaları bulunmaktadır. Küçük örneklerde, sansürlü ve simetrik olmayan verilerde iyi sonuç vermesi avantajlarındandır.[25]

Tezde gerçek veri üzerinde uygulama ile Cox modelinde sağkalım fonksiyon parametrelerinin tahmin, tehlike oranı ve güven aralığı hesaplanmıştır. Simülasyon çalışmamız ile üretilen veri yardımıyla deneysel olabilirlik yöntemi ve düzeltilmiş deneysel olabilirlik yöntemlerinin etkinliği karşılaştırılmıştır.

## 2. SAĞKALIM ANALİZİ

“Sağkalım analizi” terimi; bir hastalığın başlaması veya nüksetmesi ve ölüm gibi belli bir olayın zamanını kapsayan verilerin incelenmesinde kullanılmıştır. 1662’de John Graunt tarafından yapılan ilk yaşam tablosu ile sağkalım analizi çalışmaları 17.yy’a kadar uzanmaktadır.

Yüzyıllar boyunca, sağkalım analizi yalnızca ölüm oranlarının araştırılmasıyla ilişkili olarak düşünülürdü. Fakat son birkaç yılda sağkalım analizinde istatistiksel yöntemlerin uygulamaları ile birlikte biyomedikal araştırmaların ötesinde, kriminoloji, sosyoloji, ekonomi, endüstri, kurumsal araştırma ve sağlık sigortası uygulamaları gibi diğer alanlara uygulanmaktadır.

Sağkalım süresi, ilgilenilen veya belirlenen bir olayın meydana gelme süresi olarak tanımlanabilir. Sağkalım verisinde; sağkalım süresi, verilen bir tedaviye yanıt süresi ve hasta karakter özelliği gibi birçok veri sonuçları bulunabilir. Sağkalım verisi üzerinde çalışmalar ile hastanın ilgili olaya yanıt verme olasılığının, sağkalım-ortalama yaşam süresinin tahmin edilmesi veya deney hayvan ya da insanlar ile hastaların sağkalım dağılımlarının karşılaştırılması gibi birçok konuda kullanılır. [9] Farklı bir deyişle, sağkalım veri analizi, hayvanlar üzerinde yapılan laboratuvar çalışmalarından elde edilen sağkalım verisi veya insanların klinik ve epidemiyolojik çalışmalardan elde edilen verileri istatistiksel metotlar yardımıyla analiz etmek için kullanılan bir alandır diyebiliriz.

Sağlık alanı için baktığımızda sağkalım analizi, ölüm, kalp durması, uyuşturucu bağımlılığının nüksetmesi gibi belli bir zaman içinde gözlemlenen olaylara ilişkin verilerin analiz edilmesinde önemli rol oynamaktadır. Sağkalım modeli, önemli risk faktörlerini tanımlamanın yanı sıra bu tehlikelerin sağkalım sürelerinin öngörülmesinde ayrıca önem taşımaktadır. Bu bilgiler cerrahlar, psikologlar ve üreticiler de bu risk faktörlerini çözebilmek için önemlidir. [1]

Sağkalım analizinin gelişmesinde Kaplan, Meier, Nelson ve Cox gibi birçok bilim insanları ciddi katkıda bulunmuşlardır. Kaplan ve Meier'in 1958'deki sağkalım olasılıkları ve tehlike oranlarını tahmin etmesi ile sağkalım analizinde önemli derecede gelişmelere yön vermiştir. Nelson-Aalen tahmincisi, kümülatif tehlike oranının parametrik olmayan alternatif bir tahmincisidir. Belirli olayların kümülatif sayısını tahmin etmek için sağkalım teorisi, güvenilirlik mühendisliği ve hayat sigortası gibi farklı alanlarda kullanılmaktadır. Ayrıca Nelson – Aalen tahmincisinin eğriliği, içbükey ise şekil bebek ölümlerinin bir göstergesi ve dışbükey ise şekil insanların ölümlerini göstermektedir.[24] 1972'de David Cox, orantılı tehlike varsayımı ile

tehlike fonksiyonu olmadan etki parametrelerinin tahmin edilmesinin mümkün olduğunu söylemiştir. Sağkalım modellerinin sınıflarından biri olan Cox orantılı tehlike modeli olarak adlandırılır ve bu model ile bir değişkene göre tehlike oranına bakılarak birim zamandaki artışın etkisi incelenebilmesi gibi sağkalım analizine önemli katkısı olmuştur.[24] İki tehlike oranı sabit olması sebebiyle model “orantılı tehlike” olarak da ifade edilmektedir.

Sağkalım analizindeki diğer gelişmeler, 1970’li yılların başlarından itibaren metodoloji, teori ve uygulama alanlarında sağkalım yöntemlerinin kullanılması hızlı bir şekilde yayıldı. 1980’den beri yaşam süresi verilerin analiz edilmesi için yazılım paketleri kullanılmaktadır. [7] Yaşam süresi olayların(ölüm veya başarısızlık) belirlenmesi yerine bir zaman ölçeği ve zaman başlangıcını içerir. Zaman ölçeği her zaman gerçek veya kronolojik zaman değildir.

Ayrıca, Vaupel tarafından 1979’da tanıtılan ve 1984’te Hougaard tarafından geliştirilen paylaşılan ve paylaşılmayan kırılma modellerini içerir. Bu modeller, tahminlerde önyargıları ortadan kaldırırken heterojenliğin ve rastgele etkilerin yerleştirilmesinde daha uygundur.

1988 yılında Owen tarafından öne sürülen deneysel olasılık yöntemi, ilk olarak bir rasgele vektörün ortalamasının güven bölgesi oluşturmak için tanımlanmıştır. Wang, Qin, Klein gibi birçok bilim insanları tarafından çalışılmış ve hala bu yöntem üzerinde çalışmalar devam etmektedir.

## **2.1. Sağkalım Analizinde Kullanılan Dağılımlar**

Sağkalım analizinde, sağkalım süresi (rasgele değişkeni) kesikli ve sürekli olacak şekilde iki farklı dağılım göstermektedir.

### **2.1.1. Kesikli dağılımlar**

Yaşam süreleri bazı türlerin (sınıfın) devirlerinin sayısı olarak gruplandırılır veya ölçüldüğünde, T kesikli rasgele değişken olarak davranabilir. T,  $(0 \leq t_1 < t_2 < \dots)$  olacak şekilde  $(t_1, t_2, \dots)$  değerlerini aldığı varsayalım. Olasılık fonksiyon  $j = 1, 2, \dots$  için  $f(t_j) = \Pr(T = t_j)$  olsun. O halde sağkalım fonksiyonu:

$$S(t) = \Pr(T \geq t) = \sum_{j:t_j \geq t} f(t_j). \quad (2.1.1)$$

Tüm  $t \geq 0$  için bir fonksiyon olarak düşünüldüğünde,  $S(t)$  soldan sürekli, artmayan adım fonksiyonudur,  $S(0) = 1$  ve  $S(\infty) = 0$  dir. Kesikli zaman tehlike fonksiyonu  $j = 1,2, \dots$  için:

$$\begin{aligned} h(t_j) &= \Pr(T = t_j \mid T \geq t_j) \\ &= \frac{f(t_j)}{S(t_j)} \end{aligned} \quad (2.1.2)$$

Sürekli durumda olduğu gibi olasılık, sağkalım ve tehlike fonksiyonlarının  $T$ 'ye bağlı dağılımda denk ifadeler içermektedir.  $f(t_j) = S(t_j) - S(t_{j+1})$  olduğundan

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)}$$

ve böylece

$$S(t) = \prod_{j:t_j < t} [1 - h(t_j)] \quad (2.1.3)$$

şeklinde yazabiliriz.[7]

### 2.1.2. Sürekli dağılımlar

Bazı kitledeki bireylerin yaşam sürelerini temsil eden, negatif olmayan rasgele değişken  $T$  olsun. Tüm fonksiyonlar  $[0, \infty)$  aralığı üzerinde tanımlanır.  $f(t)$ ,  $T$ 'nin olasılık yoğunluk fonksiyonu (o.y.f) ve  $F(t)$  de kümülatif dağılım fonksiyonu (k.d.f) olsun.

$$F(t) = \Pr(T \leq t) = \int_0^t f(x) dx$$

$t$  zaman hayatta kalan bir bireyin olasılığı, sağ kalım fonksiyonu ile verilirse:

$$S(t) = \Pr(T \geq t) = \int_0^{\infty} f(x) dx \quad (2.1.4)$$

$S(0) = 1$  ve  $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$  olduğundan  $S(t)$  fonksiyonu monoton azalan sürekli fonksiyondur.

Yaşam süresi dağılımlarında en önemli kavram  $h(t)$  tehlike fonksiyonudur ve

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.1.5)$$

olarak tanımlanır.

Tehlike fonksiyonu  $t$  zamanında ani ölüm veya başarısızlık oranı olarak belirtilir; çünkü birey  $t$  zamanına kadar hayatta kalır.  $h(t) \cdot \Delta t$  çarpımı,  $[t, t + \Delta t)$ 'de yaklaşık ölme olasılığıdır.

$F(t), f(t), S(t)$  ve  $h(t)$  fonksiyonları matematiksel olarak  $T$  'nin dağılımlarını tanımlamada denktir.  $f(t) = -S'(t)$  olacak şekilde arasında ilişki vardır. Buna göre  $h(t)$ 'yi aşağıdaki gibi yazabiliriz.

$$h(x) = -\frac{d}{dx} \log S(x) \quad (2.1.6)$$

Böylece

$$\int_0^t \frac{d}{dx} \log S(x) = -\int_0^t h(x) dx,$$

$S(0)=1$  olduğu için,

$$S(t) = \exp \left\{ -\int_0^t h(x) dx \right\}. \quad (2.1.7)$$

Kümülatif tehlike fonksiyonu ( $H(t) = \int_0^t h(x) dx$ ) ve sağ kalım fonksiyonu arasındaki ilişki

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

şeklindedir.[6]

### 2.1.2.1. Önemli dağılımlar

#### a. Üstel dağılım

Üstel dağılım, sağkalım çalışmalarında en basit ve en önemli dağılımdır. 1940'lerden sonra araştırmacılar elektronik sistemin yaşam modellerini tanımlamak için üstel dağılımı kullanmaya başlamıştır. 1952'de Davis üstel dağılımın kullanıldığı maaş bordrosu kontrol hatası, otomatik hesaplama makine hataları gibi banka sistemleri ve ana defterdeki hatalardan oluşan verileri içeren birçok örnekler vermiştir. 1953'te Epstein ve Sobel üstel dağılımın normal dağılıma karşı neden seçildiğini rapor etmiş ve veri sadece sansürlü olduğunda parametreyi nasıl tahmin edileceğini göstermiştir. [9] İstatistiğin diğer alanlarında normal dağılımın etkinliği gibi üstel dağılımda sağkalım çalışmalarında önemli rol oynamaya devam etmektedir.

Üstel dağılım sabit tehlike fonksiyonu ile karakterize edilir:  $t \geq 0, \lambda > 0$  olacak şekilde  $h(t) = \lambda$  olarak ifade edilsin. Buna göre o.y.f ve sağkalım fonksiyonu sırasıyla;

$$f(t) = \lambda e^{-\lambda t} \quad (2.1.8)$$

$$S(t) = e^{-\lambda t} \quad (2.1.9)$$

olarak yazılır.

$\theta = \lambda^{-1}$  parametresi kullanarak dağılım tekrar yazılırsa;

$$f(t) = \theta^{-1} e^{-\frac{t}{\theta}} \quad t \geq 0. \quad (2.1.10)$$

T rasgele değişkenin (2.1.10) dağılımına sahip olduğunu belirtmek için  $T \sim \text{Exp}(\theta)$  notasyonu kullanılır. Dağılımın ortalama ve varyansı sırasıyla  $\theta$  ve  $\theta^2$  dir.[7]

$$E(\theta) = \theta$$

$$\text{Var}(\theta) = \theta^2.$$

### **b. Weibull dağılımı**

Weibull dağılımı, en geniş ölçüde kullanılan sağkalım süresinin dağılım modelidir. Sağkalım sürelerine veya üretilen ürünlerin devam sürelerine ilişkin uygulamalarda yaygın kullanılmakta ve bilye, otomobil parçaları ve elektriksel yalıtım gibi ürünlerin çeşitli türlerinde model olarak kullanıldığı gibi biyoloji ve tıp uygulamalarında da kullanılır. Örneğin; laboratuvar hayvanlarında veya insanlarda tümörün meydana gelmesinin zaman içindeki çalışmalarında kullanılır. Weibull dağılımı, üstel dağılımın genelleştirilmesi olmasına rağmen üstel dağılımın aksine sabit bir tehlike oranı bulunmaz ve bu sebeple model daha kapsamlı uygulama yapılmasını sağlar. Weibull, 1939'da dağılımı yayınlanmış ve 1951 yılında da dağılımın bazı başarısız durumlar üzerinde uygulanabilirliğini gösterdi. Güvenilirlik ve hastalığa bağlı ölüm gibi alanlarda çalışmaların çoğunda kullanılmaktadır.

Weibull dağılımı aşağıdaki tehlike fonksiyonuna sahiptir:

$$h(t) = \lambda \beta (\lambda t)^{\beta-1} \quad (2.1.11)$$

$\lambda > 0$  ve  $\beta > 0$ .  $\beta = 1$  olduğu durumda üstel dağılıma dönüşür. Olasılık yoğunluk fonksiyonu (o.y.f) ve sağkalım fonksiyonlarının  $t > 0$  için aşağıdaki şekilde ifade edilir.

$$f(t) = \lambda\beta(\lambda t)^{\beta-1} \exp[-(\lambda t)^\beta] \quad (2.1.12)$$

$$S(t) = \exp[-(\lambda t)^\beta] \quad (2.1.13)$$

$\beta$  ve  $\lambda$  sırasıyla, şekil ve ölçü parametreleridir. Weibull tehlike fonksiyonu, eğer  $\beta > 1$  ise monoton artan,  $\beta < 1$  ise azalan ve  $\beta = 1$  ise sabittir. [3]

$\Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du$  olmak üzere,  $E(X^r) = \lambda^{-r} \Gamma(1 + \frac{r}{\beta})$  dir. O halde dağılımın ortalaması ve varyansı ise;

$$E(X) = \lambda^{-1} \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$Var(X) = \lambda^{-2} \left[ \Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}\right)^2 \right].$$

### c. Log-Normal dağılım

Lognormal dağılım, mühendislik, tıp ve başka alanların çeşitli uygulamalarında model olarak kullanılır. Dağılımın başlangıcı 1879'dan öncesine kadar dayanmaktadır. McAlister 1879 yılında dağılımın teorisini açık bir şekilde tanımlamıştır.  $Y = \log T$  ortalaması  $\mu$  ve varyansı  $\sigma^2$  olduğu söylenen normal dağılımlı ve o.y.f

$$f(y) = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma} \exp\left[-\frac{1}{2} \left(\frac{y - \mu}{\sigma}\right)^2\right] \quad (-\infty < y < \infty)$$

ise yaşam süresi  $T$  log-normal dağılımlıdır denir.

$T = \exp Y$  nin o.y.f nu:

$$f(t) = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma t} \exp\left[-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma}\right)^2\right] \quad t > 0 \quad (2.1.14)$$

şeklinde bulunur.

Log-normal dağılım için sağkalım ve tehlike fonksiyonları standart normal dağılımı içermektedir.

$\Phi(x) = \int_{-\infty}^x \frac{1}{(2\pi)^{1/2}} e^{-u^2/2}$  standart normal dağılım olmak üzere, log-normal sağkalım fonksiyonu

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right). \quad (2.1.15)$$

#### d. Log-Lojistik dağılımı

$\log(T)$ , lojistik dağılıma sahip ise sağkalım süresi  $T$  log-lojistik dağılıma sahiptir.

Log-lojistik dağılımı aşağıdaki form

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta]^2}, t > 0 \quad (2.1.16)$$

şeklindedir.  $\alpha > 0$  ve  $\beta > 0$  parametrelerdir. Sağkalım ve tehlike fonksiyonu aşağıdaki gibidir:

$$S(t) = [1 + (t/\alpha)^\beta]^{-1} \quad (2.1.17)$$

$$h(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta]} \quad (2.1.18)$$

$Y = \log T$  bir lojistik dağılıma sahiptir ve  $u = \log \alpha$  ve  $b = \beta^{-1}$  olduğunda o.y.f

$$f(y) = \frac{b^{-1} \exp[(y-u)/b]}{[1 + \exp[(y-u)/b]]^2} \quad -\infty < y < \infty, \quad (2.1.19)$$

şeklindedir.  $Y \sim \text{Logist}(u, b)$  notasyonu kullanılır ve  $T \sim \text{Logist}(\alpha, \beta)$  şeklinde belirtilir.

$T$ 'nin  $r$ .momenti vardır ve  $\beta > r$  sağlandığında  $E(T^r) = \alpha^r \Gamma(1 + r/\beta) \Gamma(1 - r/\beta)$  şeklindedir.  $\beta > 1$  durumda ortalama,

$$E(T) = \alpha \Gamma(1 + \beta^{-1}) \Gamma(1 - \beta^{-1}).$$

## 2.2. Veri Türleri

Sağkalım analizi, olay zaman verilerinin (time-to-event data) analizidir. Hastalığın tetikleme, ölüm, araba kazaları, depremler, elektronik alet, parça veya sistemin ömrü, evlilik süresi, gazete, dergi gibi aboneliklerin süresi gibi birçok olayları incelemek için sağkalım analiz metotlarını uygulayabiliriz. Tabi burada incelediğimiz veri sansürlü ve sansürlü olmak üzere iki şekildedir.

### 2.2.1. Sansürsüz veri

Çalışmanın sonuna kadar deneklerin başarısız süresi veya sağkalım süresi tam olarak biliniyor ise böyle bir veriye sansürsüz veri denir. Bu türdeki veri türüne sahip deneklerin çalışılan zaman aralığında beklenen olayı gerçekleştirdiği bilinir.

### 2.2.2. Sansürlü veri

Olay zaman verileri, veri analiz etmede özel problemler yaratan farklı yolları kendileri sunar. Olay zamanı verisinin ortak bir özelliği olan sansürleme olarak bilinen olay zamanı verisi sunmasıdır. Sansürleme başarısız zaman verisinin özgün özelliklerinden biridir. Sansürleme ile ilgili sağkalım zamanında gözlemin tamamlanmadığını anlayabiliriz. Yani sağkalım süresi tam olarak bilinen olmasının dışında sadece belli bir aralıkta yer aldığı gözlemlenir. Sansürlü veri, eksik veriden farklıdır, çünkü sansürlü gözlemler kısmi olarak hala bilgi sağlar iken eksik gözlemler ilgili değişken hakkında herhangi bir bilgi içermez. Sansürün birçok türü vardır. Bunlar sağdan sansürlü, soldan sansürlü ve aralık sansürlü olmak üzere 3 şekildedir.[5,8,21]

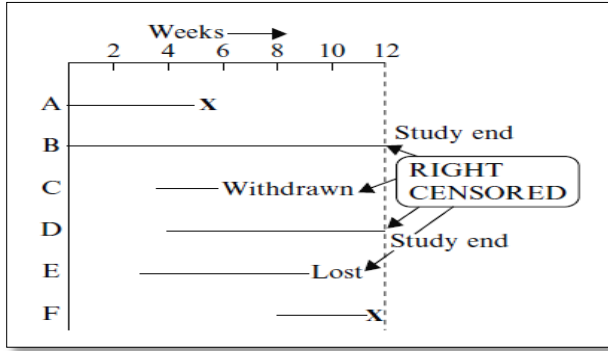
Sansürleme nedenleri:

- Takip etme kaybedilebilir(hasta başka bir yere taşınabilir)
- Vazgeçme(terapi kötü bir etkiye sebep olabilir ve tedavi durdurmak gerekebilir.)
- Çalışmanın bitmesi
- İnceleme altında olmayan bir sebepten ölüm(intihar)

#### 2.2.2.1. Sağdan sansürlü

Sağdan sansürleme veya sağdan sansürlü başarısız zaman verisi ile ilgilenilen başarısız süresinin sansür süresinden ya aynı ya da daha büyük olduğunu anlarız. Sağdan sansürlü gözlemlerin getirdiği genel bir çözümden biri sağkalım çalışmalarında zaman kısıtlamaları veya kaynak sınırlandırması sebebiyle sonlandırılmak zorunda olmasıdır. Bu durumlarda deneklerin çalışma sonunda beklenen olaylar meydana gelmediği için sağkalım süreleri tam olarak gözlemlenemiyor. Fakat sağkalım sürelerinin çalışma bitiş süresinden fazla olduğu bilinir, yani sağdan sansürlü diyebiliriz. Eğer sağkalım değişkenleri sansürleme değişkeninden küçük ise gözlem tamdır. Aksi takdirde gözlem sağdan sansürlüdür. Bu rasgele sansürleme olarak da söylenebilir. Şekil 2.1'de üç farklı sağdan sansürlü gözlem örnekleri vardır. B ve D gözlemleri çalışma sonunda yaşadığı için yani yaşam süresi tam bilinmediğinden sağdan sansürlü, C gözlemi

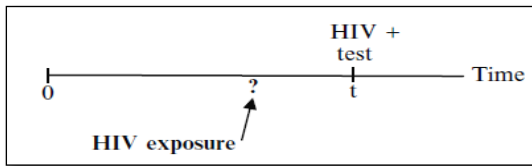
çalışmadan çekilmiş ve D gözleminin takibi kaybedildiğinden dolayı sağdan sansürlü gözlem olarak adlandırılır. [6,21]



Şekil 2.1. Sağdan sansürlü veri örneği

### 2.2.2.2. Soldan sansürlü

Bireyin sağkalım süresi, gözlemlenen sağkalım süresinden az veya eşit olduğunda bu veri soldan sansürlü olarak ifade edilir. Şekil 2.2'ye baktığımızda denekler HIV+ olana kadar ki süreci takip edilmektedir. İlk olarak bir denek virüs için pozitif olduğunda bunu başarısızlık olarak kayıt edebiliriz. Fakat denegin virüse maruz kaldığı ilk anı tam olarak bilemeyiz ve sebeple başarısızlığın meydana geldiği süre tam olarak bilinmemektedir. Böyle bir sağkalım süresi soldan sansürlüdür, çünkü denegin maruz bittiğindeki sağkalım süresi, testin pozitif olduğundaki bitirilen takip süresinden daha kısadır. Diğer bir deyişle, eğer bir kişi  $t$  anında soldan sansürlü ise  $(0 - t)$  zaman aralığında olay yaşandığını biliyoruz fakat olayın gerçek zamanını bilmiyoruz.[6]



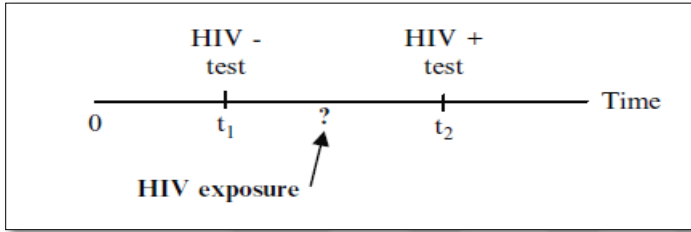
Şekil 2.2. Soldan sansürlü veri örneği

### 2.2.2.3. Aralık sansürlü

Aralık sansürleme ile çalışma denekleri veya ilgili başarısız zaman süreçleri sürekli gözlem altında değildir. Bu sebeple, başarısız veya sağkalım süresi tamamen gözlemlenmiş veya sağdan sansürlü olduğunu söyleyemeyiz. Aralık sansürlü gözlem, sadece belli bir zaman aralığındaki sağkalım olayının meydana geldiği aralık olarak bilinir. Tam veya sağdan sansürlü başarısızlık süreleri aralık sansürlü başarısızlık süresinin özel bir durumuyla ilişkilendirilebilir, çünkü böyle durumlarda aralık tek bir noktaya düşürülür veya sağda sınırsızdır. Daha genel olarak ifade edilirse, aralık

sansürlü gözlemi birden fazla örtüşmeyen aralıkların birleşimi olarak da tanımlayabiliriz.

Aralık sansürlü başarısızlık zaman verisi sosyolojik, medikal, finansal, demografik çalışmaları kapsayan birçok alanda kullanılır. Aralık sansürlü verinin en çok olduğu örnek, periyodik olarak yapılan sağlık çalışmaları veya medikaller söylenebilir. Şekil 2.3'teki veri örneğinde bir deneğin HIV gözetimi yapıldığını düşünelim. Deneğin iki HIV testi yapılabilir. Bu testlerde deneğin 1.testin zamanı  $t_1$ 'de HIV negatif olduğunu ve 2.testin zamanı  $t_2$ 'de HIV pozitif olduğunu belirtir. Böyle bir durumda, deneğin gerçek sağkalım süresinin  $t_1$ 'den sonra ve  $t_2$ 'den önce meydana geldiğini yani, deneğin sağkalım zamanının  $(t_1, t_2)$  aralığında olduğunu söyleyebiliriz.[6,21]



Şekil 2.3. Aralık sansürlü veri örneği

### 2.3. Sağkalım Parametre Tahmin Yöntemleri

Sağkalım parametrelerin tahmin etmek için çok sayıda yöntem bulunmaktadır. Burada en çok kullanılan Kaplan-Meier, Nelson-Aalen ve Yaşam Tablosu yöntemleri gösterilecektir.

#### 2.3.1. Kaplan-Meier yöntemi

Bağımsız ve aynı dağılımlı  $n$  sağkalım sürelerine bağlı  $S(t)$  sağkalım fonksiyonunun parametrik olmayan tahminidir. Çarpım-Limit tahmin edicisi olarak da bilinen Kaplan-Meier tahmin edicisi tıp/halk sağlığı çalışmalarında en yaygın şekilde kullanılan tahmin edicidir. Eğer  $n$  boyutlu bir örnekte sansür yok ise sağkalım fonksiyonu,

$$S(t) = \frac{\text{Gözlem sayısı} \geq t}{n} \quad t \geq 0 \quad (2.3.1)$$

şeklinde tanımlanır. Bu tüm gözlemler kesikli ise her gözlemlenen yaşam süresinden sonra  $1/n$  şeklinde azalan bir basamak fonksiyonudur. Eğer yaşam süresi verisi sansürlü ise, denklemde bazı değişiklik yapılması gerekir. Çünkü  $t$  ye eşit veya daha büyük yaşam süresinin sayısı tam olarak bilinmeyecektir.

$T_1, T_2, \dots, T_n$  bağımsız ve aynı dağılımlı sağkalım süresi,  $S(t)$  sağkalım fonksiyonu olmak üzere,  $C_1, C_2, \dots, C_n$  bağımsız ve aynı dağılımlı ve  $T_i$ 'den bağımsız sansürleme zamanları olsun.  $(t_i, \delta_i)$   $i = 1, 2, \dots, n$  yaşam sürelerinin rasgele örneklemini gösterebilir.  $d_j = \sum I(t_i = t_j, \delta_i = 1)$   $t_j$  anında ölümlerin sayısını ve  $n_j = \sum I(t_i \geq t_j)$  ise  $t_j$  anında risk altında kalanların sayısını ( $t_j$  den önceki sansürlenmiş ve yaşayan bireylerin sayısı) göstermek üzere  $S(t)$ 'nin KM tahmini:

$$S(t) = \prod_{t_j < t} \frac{n_j - d_j}{n_j}$$

olarak tanımlanır.  $\frac{d_j}{n_j} = \lambda_j$  olarak alınırsa,

$$S(t) = \prod_{t_j < t} (1 - \lambda_j) \quad (2.3.2)$$

şeklinde yazabiliriz. Buradan hareketle yoğunluk fonksiyonunu da aşağıdaki şekilde ifade edebiliriz:

$$f(t) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j). \quad (2.3.3)$$

(2.2.3) denklemini yoğunluk fonksiyonu  $f(t)$  yardımıyla sağkalım fonksiyonu  $S(t)$  aşağıdaki gibi elde edilir.

$$F(t) = \sum_{j=1}^i f(t_j)$$

$$F(t) = \lambda_1 + \lambda_2(1 - \lambda_1) + \lambda_3(1 - \lambda_1)(1 - \lambda_2) + \dots + \lambda_i(1 - \lambda_{i-1}) \dots (1 - \lambda_1)$$

$$S(t) = 1 - F(t)$$

$$S(t) = 1 - \lambda_1 - \lambda_2(1 - \lambda_1) - \lambda_3(1 - \lambda_1)(1 - \lambda_2) - \dots - \lambda_i(1 - \lambda_{i-1}) \dots (1 - \lambda_1)$$

$$= (1 - \lambda_1)(1 - \lambda_2)(1 - \lambda_3) \dots (1 - \lambda_i)$$

$$= \prod_{j=1}^n (1 - \lambda_j)$$

dir.

Risk altında bulunan  $n_i$  denekten  $d_i$  öldüğünde ( $t_i^-$ ,  $t_i$  den önceki süre olmak üzere) elde edilen koşullu olasılık:

$$L_i = \frac{[f(t_i)]^{d_i} [S(t_i)]^{n_i - d_i}}{[S(t_i^-)]^{n_i}}.$$

Yukarıda verilen (2.3.2) ve (2.3.3) ifadeleri yerine yazılırsa;

$$\begin{aligned} L_i &= \frac{[\lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j)]^{d_i} [\prod_{j=1}^i (1 - \lambda_j)]^{n_i - d_i}}{[\prod_{j=1}^{i-1} (1 - \lambda_j)]^{n_i}} \\ &= \frac{\lambda_i^{d_i} [\prod_{j=1}^{i-1} (1 - \lambda_j)]^{d_i} [\prod_{j=1}^{i-1} (1 - \lambda_j)]^{n_i - d_i} (1 - \lambda_i)^{n_i - d_i}}{[\prod_{j=1}^{i-1} (1 - \lambda_j)]^{n_i}} \\ &= \lambda_i^{d_i} (1 - \lambda_i)^{n_i - d_i} \end{aligned}$$

elde edilir. Bulunan sonuç  $n_i$  denemeli,  $d_i$  başarı sayısı ve  $\lambda_i$  başarımın olasılığı olan bir iki terimli binom aşamasına benzemektedir. [8]  $L_i$  olasılık değerinin sonlu çarpımı,  $S(t)$  sağkalım fonksiyonu uzayında olabilirlik fonksiyonu olarak ifade edilir.

$$\begin{aligned} L &= \prod L_i \\ &= \prod_{i=1}^k \lambda_i^{d_i} (1 - \lambda_i)^{n_i - d_i} \end{aligned}$$

### Varyans Tahmini

Sağkalım parametresini tahmin etmek için Kaplan-Meier tahmin edicisi kullanıldığında,  $S(t)$  sağkalım fonksiyonunun varyansının tahminini bulmak kolaydır. Greenwood formülü olarak bilinen eşitlik yardımıyla hesaplanır. Bu formül aşağıdaki şekilde ifade edilir.

$$Var[\hat{S}(t)] = \hat{S}(t)^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)} \quad (2.3.4)$$

Eğer elimizde veri sansürlü ise formülü aşağıdaki forma dönüşür;

$$Var[\hat{S}(t)] = \frac{\hat{S}(t)[1 - \hat{S}(t)]}{n}. \quad (2.3.5)$$

$\hat{S}(t)$ ,  $S(t)$ 'nin Kaplan-Meier tahmini veya Çarpım-Limit tahmini olarak söylenir.

Varyans tahmini,  $I(\lambda)$  bilgi matrisinde köşegen değerleri  $I_{tt}(\lambda) = -\frac{\partial^2 \log L}{\partial \lambda(t)^2} =$

$\frac{n_t}{\lambda(t)(1-\lambda(t))}$  olduğu olabilirlik fonksiyonunun ikinci türevinden bulunmaktadır.

Standart maksimum olabilirlik büyük-örneklem teorisi sonucu  $Var(\hat{\lambda}) = I(\hat{\lambda})^{-1}$  olması kullanılarak sağkalım fonksiyonunun tahmininin varyansı bulunur.

$$Var(\ln \hat{S}(t)) = \left( \frac{1}{\hat{S}(t)} \right)^2 Var(\hat{S}(t)) \quad ^1$$

$$Var(\hat{S}(t)) = \hat{S}(t)^2 Var(\ln \hat{S}(t))$$

$$\begin{aligned} &\approx \sum_{s=0}^{t-1} \hat{S}(t)^2 Var[\ln(1 - \hat{\lambda}(s))] \\ &= \hat{S}(t)^2 \sum_{s=0}^{t-1} Var(\hat{\lambda}(s)) \frac{1}{(1 - \hat{\lambda}(s))^2} \\ &= \hat{S}(t)^2 \sum_{s=0}^{t-1} \frac{\hat{\lambda}(s) (1 - \hat{\lambda}(s))^{-1}}{n_s} \end{aligned}$$

$\hat{\lambda}(s) = \frac{d_s}{n_s}$  eşitliği yukarıdaki eşitlikte yazılırsa, önceki kısımda ifade edilen varyans ile aynıdır.

### 2.3.2. Nelson-Aalen yöntemi

Nelson-Aalen yöntemi, belirli bir popülasyonun zamanla nasıl geliştiğini analiz etmek için kullanılır. Bu yöntem yardımıyla bulduğumuz parametre tahminlerine Nelson-Aalen tahmini denir. Nelson – Aalen tahmincisi, sansürlü veya eksik verilerde kümülatif tehlike oranının parametrik olmayan bir tahmincisidir. Bu yöntemi sağkalım teorisi, güvenilirlik mühendisliği ve hayat sigortası gibi farklı alanlarda kullanılmaktadır. Ayrıca, Poisson sürecinin homojenliğini test ederken de kullanılır.[23,26]

Kümülatif tehlike fonksiyonunun tahmini olarak verilen integral Riemann-Stieljes integrali olarak da verilir.  $Y_i(t) = I(t_i \geq t)$  ve  $dN_i(t) = I(t_i = t, \delta_i = 1)$  iken integral,

---

<sup>1</sup> Delta yöntemi kullanılmıştır.

$$\widehat{\Lambda}(t) = \int_0^t \frac{dN.(u)}{Y.(u)}. \quad (2.3.6)$$

Deneysel kümülatif tehlike fonksiyonu olarak da adlandırılabilir. Fakat en yaygın Nelson (1969) ve Aalen (1972 tezi) tarafından yayınlanan Nelson-Aalen(NA) tahmini olarak bilinir. Eğer  $t_1, \dots, t_k$  başarısızlığın gözlemlendiği kesikli süreleri gösterirse, NA tahmini aşağıdaki gibi yazılır.

$$\widehat{\Lambda}(t) = \sum_{j:t_j \leq t} \lambda_j = \sum_{j:t_j \leq t} \frac{d_j}{n_j} \quad (2.3.8)$$

$\widehat{\Lambda}(t)$  'nin grafiği tehlike fonksiyonunun şekli hakkında bilgi verir.  $\lambda(t)$  tehlike fonksiyonu sabit ise  $\Lambda(t)$  lineerdir ve  $\lambda(t)$  monoton ise  $\Lambda(t)$  konveks (iç bükey)dir.  $\widehat{\Lambda}(t)$  'nin varyansı için, Kaplan-Meier yöntemi için en çok olabilirlik tahmininde asimptotik varyans tahmini burada da sağlanır.  $Var(\hat{\lambda}) = I(\hat{\lambda})^{-1}$  özelliği kullanılarak varyans tahmini,

$$\begin{aligned} \widehat{Var}[\widehat{\Lambda}(t)] &= \sum_{j:t_j \leq t} Var[\hat{\lambda}(t)] \\ &= \sum_{j:t_j \leq t} I(\hat{\lambda})^{-1} \\ &= \sum_{j:t_j \leq t} \left( \frac{n_j}{h_j(1-h_j)} \right)^{-1} \\ &= \sum_{j:t_j \leq t} \frac{d_j(n_j - d_j)}{n_j^3}. \end{aligned}$$

Alternatif varyans tahmini olarak aşağıdaki gibi verilir.

$$\widehat{Var}[\widehat{\Lambda}(t)] = \sum_{j:t_j \leq t} \frac{d_j}{n_j^2}$$

Veri analizinde Nelson-Aalen tahmin edicisinin kullanılmasında iki önemli sebep vardır. İlki, olay süreleri için parametrik modeller arasında seçim yapılabilir. Nelson-Aalen tahmin edicisinin grafiği çizildiğinde, eğer verilen parametrik model veriyle uyumlu ise grafik sonuçları yaklaşık olarak lineer olacaktır. İkincisi, Nelson-Aalen

tahmin edicisi  $\lambda$  tehlike oranının ham tahminlerini gösterir.[7] Yani, bu tahminler Nelson-Aalen tahmin edicisinin eğimidir.

### 2.3.3. Yaşam tablosu yöntemi

Sağkalım verisini resmetmek için geniş çapta kullanılan en eski yöntemlerden biridir. Tartışılan kohort yaşam tablosu bile kullanıldı. Çünkü 20.yy'ın başlarında kitle yaşam tablosu nüfus bilimcileri tarafından kullanıldı. Ayrıca aktüerya çevresinde kullanımı oldukça uzun oldu. Yaşam tablosu, sağkalım ve sansürleme süreleri aralıklar halinde gruplanan örneklem veya kohort bireyin sağkalım verisinin resmedilmesi için bir araç olmuştur. Burada genellikle kohort bazı kitleden rasgele bir örneklem olduğu varsayılır.[7] Sağkalım dağılımının tahmininin bulunmasında bu yöntem kullanılmıştır.

Kaplan-Meier yöntemi herhangi bir sağkalım veri seti için uygulanabilir, fakat yöntem büyük bir veri takımı (100 veya daha fazla denek) için aralıklarda grup zamanlamalarında daha uygundur. Aktüerya veya histogram yöntem olarak da bilinen yaşam tablosunda işlem frekans tablosu oluşturulmasına benzerdir. Örnek üzerinde gösterecek olursak:

Tablo 2.1. Örnek yaşam tablosu

1	2	3	4	5	6	7	8
Zaman	Son kayıtlar		Başlangıçta yaşayan	Risk altındaki sayı	Ölüm oranı aralıkta	Sağkalım oranı aralıkta	(%) t de sağkalım oranı
$t - t + 1$	d	c	n	n'	q	p	S(t)
0-1	90	0	374	374.0	0.2406	0.7594	100.0
1-2	76	0	284	284.0	0.2676	0.7324	75.9
2-3	51	0	208	208.0	0.2452	0.7548	55.6
3-4	25	12	157	151.0	0.1656	0.8344	42.0
4-5	20	5	120	117.5	0.1702	0.8298	35.0
5-6	7	9	95	90.5	0.0773	0.9227	29.1
6-7	4	9	79	74.5	0.0537	0.9463	26.8
7-8	1	3	66	64.5	0.0155	0.9845	25.4
8-9	3	5	62	59.5	0.0504	0.9496	25.0
9-10	2	5	54	51.5	0.0388	0.9612	23.7
10+	21	26	47	-	-	-	22.8

1. Zaman aralıklarının seçimi verinin yapısına ve boyutuna bağlıdır. Sıklık tablosunun oluşumundaki ana esaslar burada eşitse uygulanabilir.

2. ve 3. Hastalar en son kayıttaki durumlarının zaman aralıklarına göre sınıflandırılmıştır. Eğer bir ölüm raporlanmışsa (2.) sütunda yazılmıştır ya da hasta sansürlü ise (3.) sütunda sayılmıştır. Her hastanın zamanı çalışma kaydına göre hesaplanır.

4. Başlangıç aralığında yaşayan hasta sayısı (2.), (3.) ve (4.) sütunlardan yaralanarak hesaplanır. Örneğin, 10 yıldır yaşayan hasta sayısı  $21 + 26 = 47$  veya 8 yıldır yaşayan  $54 + 3 + 5 = 62$  şeklinde hesaplanır.

5.  $t$ 'den  $(t + 1)$ 'e kadar ki zamanda risk altındaki hasta sayısı:  $n' = n - c/2$ . Bu formül ölüm oranı hesabı için gereklidir.  $c$  sansürlü denekler, aralığın sadece bir kısmında risk altında olduğu için  $n$  de  $n'$  gibi bir düzenleme yapılmasına ihtiyaç duyulur. Buradaki esas varsayım ortalama olarak aralığın yarısında risk altındadır.

6. Aralıktaki ölüm oranı:  $q = d/n'$ . Örneğin birinci aralık boyunca ölüm oranı,

$$q = \frac{90}{374.0} = 0.2406.$$

7. Aralıktaki sağkalım oranı:  $p = 1 - q$ .

8.  $S(t)$  için,  $t$  yılındaki sağkalım oranı (7.) sütundaki  $t$ 'den  $(t + 1)$ ' e kadar ki değerlerin çarpımı şeklinde bulunur. Örneğin,

$$2.\text{yıldaki sağkalım oranı: } S(2) = (0.7564 \times 0.7324) \times 100 = 55.6$$

5.yıldaki sağkalım oranı:

$$S(5) = (0.7594 \times 0.7324 \times 0.7548 \times 0.8344 \times 0.8298) \times 100 = 29.1$$

Kaplan-Meier metodu ile elde edilmiş oranlar ve sağkalım eğrisi yaşam tablosu yöntemleriyle de bulunabilir. Buradaki tek fark diyagramdaki noktalar genellikle düz çizgi ile ilişkilidir. [8] Ayrıca bu yöntem kullanışlıdır. Çünkü kullanıcı Kaplan-Meier eğimini çizmek için bilgisayar paketi olsa bile elde edilmesi zor olan “tehlike” veya “risk” fonksiyonların tahmin edilmesinde rahatlıkla kullanılır.

#### 2.4. Parametre Tahmin Yöntemleri

Sağkalım analizinde, parametrik, yarı parametrik ve parametrik olmayan olmak üzere üç tür sağkalım dağılımı vardır. Her dağılım türü için tahmin yöntemleri bulunmaktadır. 1972’de David Cox’un yarı parametrik sağkalım fonksiyonları ve 1988’de Owen’in parametrik olmayan sağkalım fonksiyonlarının parametrelerinin tahmin edilmesi üzerine yaptıkları çalışmalardan sonra bu alanlarda çalışmalar

artmıştır. Günümüzde özellikle parametrik olmayan sağkalım dağılımı üzerine çalışmalar devam etmektedir.

#### 2.4.1. Parametrik sağkalım fonksiyonu için tahmin yöntemi

Parametre tahmini yapılırken, sağkalım fonksiyonu parametrik dağılımına sahip ise “En Çok Olabilirlik Yöntemi” kullanılmaktadır.

##### 2.4.1.1. En Çok Olabilirlik Yöntemi

$X_1, X_2, \dots, X_n$  n rasgele değişkenin olabilirlik fonksiyonu,  $\theta$ ’nın fonksiyonu ve n rasgele değişkenin ortak yoğunluğu olarak tanımlanır. Özellikle,  $X_1, X_2, \dots, X_n$   $f(x; \theta)$  yoğunluk fonksiyonundan rasgele bir örneklem ise, o halde olabilirlik fonksiyonu  $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta)$  yoğunluk fonksiyonlarının çarpımıdır. Denklem olarak ifade edecek olursak:

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) . \quad (2.4.1)$$

$L(\theta; x_1, \dots, x_n)$  olabilirlik fonksiyonu,  $x_1, x_2, \dots, x_n$  rasgele değişkenler belirli değerler aldığı varsayan olabilirliği verir. Olabilirlik, yoğunluk fonksiyonunun değeridir; böylece kesikli rasgele değişken için onun bir olasılığıdır.  $\theta$  bilinen bir moment için  $\theta_0$  ile ifade edildiği varsayalım. “Meydana gelmesi en muhtemel” olan rasgele değişkenlerin belirli değerleri  $x_1', x_2', \dots, x_n'$  olsun. Öyle ki  $f_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n; \theta_0)$  yoğunluk fonksiyonunu maksimum yapan değerdir. Şimdi n rasgele değişkenin ortak yoğunluk fonksiyonu  $f_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)$  olmak üzere  $\theta$  bilinmeyen bir parametre olsun. Burada gelmesi en olası değeri bulmak istenir. Başka bir deyişle, olabilirlik fonksiyonunu maksimum yapan  $\theta$  değeri aranır. Bu değer genellikle  $\hat{\theta}$  ile gösterilir ve  $\{x_1, x_2, \dots, x_n\}$ ’nin bir fonksiyonudur. Yani,  $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$  dir. Böyle bir değer bulunduğunda,  $\hat{\Theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$  rasgele değişkenine en çok olabilirlik tahmin edicisi denir. En çok olabilirlik tahmin edicisini formülüne edelim:

$f(x; \theta)$  yoğunluk fonksiyonundan rasgele örneklem  $X_1, X_2, \dots, X_n$  olmak üzere olabilirlik fonksiyonu (2.4.1)’de,

$$L(\theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta)$$

şeklinde olduğu ifade edilmişti.

Çoğu olabilirlik fonksiyonları düzenlilik koşulu gösterir, bu sebeple en çok olabilirlik tahmin edicisi,

$$\frac{\partial L(\theta)}{\partial \theta} = 0$$

eşitliğin çözümünden elde edilir.  $L(\theta)$  ve  $\log L(\theta)$ ,  $\theta$ 'nın aynı değerlerinde maksimum olurlar, bu yüzden ve işlem kolaylığı olması için logaritması alınan olabilirlik fonksiyonu maksimize edilir. Eğer olabilirlik fonksiyonu  $k$  parametre içeriyorsa,

$$L(\theta_1, \theta_2, \dots, \theta_k) = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k)$$

o halde her bir parametreye göre maksimize edilirse,  $\theta_1, \theta_2, \dots, \theta_k$  parametrelerinin en çok olabilirlik tahmin edicisi  $\hat{\Theta}_1 = \hat{\vartheta}(X_1, X_2, \dots, X_n)$ ,  $\hat{\Theta}_2 = \hat{\vartheta}(X_1, X_2, \dots, X_n)$ ,  $\dots$ ,  $\hat{\Theta}_k = \hat{\vartheta}(X_1, X_2, \dots, X_n)$  rasgele değişkenleridir. Düzenlilik koşulu olması halinde olabilirliğin maksimum olduğu nokta aşağıdaki  $k$  tane denklemin ortak çözümüdür.

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_1} = 0$$

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_2} = 0$$

...

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_k)}{\partial \theta_k} = 0$$

Aynı şekilde işlem kolaylığı sebebiyle log-olabilirlik fonksiyonu ile işlemler yapılır.[9,13]

#### 2.4.1.1.1. Sağdan sansürlü veriler için en çok olabilirlik yöntemi

Yaşam süresi  $T$  ve sansür süresi  $C$  bağımsız rasgele değişkenlere sahip bir veri grubu için sağkalım fonksiyonları sırasıyla  $S(t)$  ve  $G(t)$  dir.  $G(t)$ ,  $S(t)$  'nin parametrelerinden bağımsızdır.  $t_i = \min(T_i, C_i)$  olmak üzere, eğer  $T_i \leq C_i$  ise  $\delta_i = 1$  veya  $T_i > C_i$  ise  $\delta_i = 0$  dır.  $n$  bireyin gözleminden veri,  $i = 1, \dots, n$  iken  $(t_i, \delta_i)$  çiftlerinden oluştuğunu varsayalım.  $f(t)$  ve  $g(t)$ ,  $T_i$  ve  $C_i$  için olasılık yoğunluk fonksiyonu(o.y.f) iken  $(t_i, \delta_i)$ 'nin o.y.f aşağıdaki gibi elde edilebilir:

$$\Pr(t_i = t, \delta_i = 0) = \Pr(C_i = t, T_i > C_i) = g(t)S(t)$$

$$\Pr(t_i = t, \delta_i = 1) = \Pr(T_i = t, T_i \leq C_i) = f(t)G(t).$$

$$\Pr(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i} [g(t)S(t)]^{1-\delta_i} \quad (2.4.2)$$

$i = 1, \dots, n$  için  $(t_i, \delta_i)$ 'nin dağılımı:

$$\prod_{i=1}^n [f(t)G(t)]^{\delta_i} [g(t)S(t)]^{1-\delta_i} \quad (2.4.3)$$

$g(t)$  ve  $G(t)$  fonksiyonları  $f(t)$  deki parametrelerin herhangi birini içermediği için ihmal edilebilir ve olabilirlik fonksiyonu şu şekilde alınabilir:

$$L = \prod_{i=1}^n [f(t)]^{\delta_i} [S(t)]^{1-\delta_i}. \quad (2.4.4)$$

#### 2.4.1.1.2. Sansürlü veriler (sağ, sol ve aralık) için en çok olabilirlik yöntemi

$n$  kişi için gözlemlenen sağkalım süreleri  $T_1, T_2, \dots, T_n$  olsun. Sağkalım süresi sansürsüz, sol-, sağ- ve aralık sansürlü gözlemlerden oluşuyorsa da tahmin yöntemi benzerdir.  $f(t, b)$  yoğunluk fonksiyonlu  $a$  dağılımına ait sağkalım süresinin sağkalım fonksiyonu  $S(t, b)$ , ( $b$  dağılımın bilinmeyen bütün parametrelerini göstermektedir.) olduğu varsayalım. O halde log-olabilirlik fonksiyonu,

$$\begin{aligned} l(b) &= \log L(b) \\ &= \sum \log[f(t_i, b)] + \sum \log S(t_i, b) + \sum \log[1 - S(t_i, b)] \\ &\quad + \sum \log[S(v_i, b) - S(t_i, b)]. \end{aligned}$$

İlk toplam sansürsüz gözlemler, ikinci toplam sağdan sansürlü gözlem, üçüncü toplam soldan sansürlü gözlem ve son toplam aralık sansürlü gözlemleri içerir. ( $v_i$ , sansürlü aralığın alt ucunu belirtir.)  $b$ 'nin en çok olabilirlik tahmin edicisi ( $\hat{b}$ )'nin elde edilmesi için diğer adımlar önceki kısımlara benzerdir.[9] Fakat  $\hat{b}$  en çok olabilirlik tahmin edici ve tahmin edilen kovaryans matrisinin hesaplanması zor olmasından dolayı hesaplamalar için yazılım paketi kullanılır.

#### 2.4.2. Yarı parametrik sağkalım fonksiyonu için tahmin yöntemi

Sağkalım fonksiyonu, yarı parametrik sağkalım dağılımına ait ise “Cox Yöntemi” yöntemi yardımıyla parametre tahmin edilir.

##### 2.4.2.1. Cox yöntemi

Sağkalım analizi çalışmalarında yaygın olarak kullanılan model Cox orantılı tehlike modeli (Cox proportional hazard model) veya başka bir deyişle Cox regresyon modelidir. Modeli ilk olarak 1972 yılında David Cox, “Regression Models and Life

Tables” en çok bilinen makalesinde yayınlamıştır. Cox model parametrik modellerin aksine sağkalım sürelerini belirtmek için belli bir olasılık dağılımı seçmemizi gerektirmez.[9] Model özellikle klinik denemelerde istatistikçiler tarafında kısa sürede uygulanmaya başlandı. Ayrıca Cox orantılı tehlike modeli, ortak değişkenlerin bir sağkalıma veya diğer sansürlü sonuçla ilişkisini modellemek için kullanılan bir prosedür olduğunu söyleyebiliriz.

#### 2.4.2.1.1. Cox modeli

$X_i$ ,  $i = 1, 2, \dots, n$  sağkalım süresini gösteren negatif olmayan rasgele değişken ve  $Z_i$   $i = 1, 2, \dots, n$   $X_1, X_2, \dots, X_n$  sağkalım süresiyle ilişkili ortak değişken (covariate) olsun.  $Z_i$  ortak değişkeni zamanla değişmediği varsayılır ve burada  $i$ . hastanın yaşı, cinsiyeti veya hastalık bilgisini ifade edebilir.  $X_i$   $i$ . yaşam süresinde, Cox regresyon modeli koşuluna göre  $\lambda_i(t)$  tehlike fonksiyonu, ortak değişken  $Z_i$  ile ilişkisi aşağıdaki gibidir:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta Z_i) \quad (2.4.5)$$

$\beta$  bilinmeyen regresyon katsayısı,  $\lambda_0$  temel tehlike fonksiyonu tanımsız negatif olmayan zamanın bir fonksiyonu veya  $Z = 0$  durumunda olan bir birey için tehlike fonksiyonu olarak ifade edilir. Cox regresyon modelinde  $\beta$  ve  $\lambda_0(t)$  şeklinde iki bilinmeyen vardır. Model varsayımına göre  $\lambda_0(t)$  temel tehlike fonksiyonu bilinmeyen ve keyfi bir (parametrik olmayan) fonksiyondur. Fakat çalışmalarda alınan tüm denek( veya birey) için fonksiyon ortaktır.  $\beta$  regresyon katsayısı sonlu boyutlu bir parametredir. Parametrik ve parametrik olmayan modellerin birlikte yer aldığı karışık modellere yarı parametrik model olarak da ifade edilmektedir.

Modelin en önemli özelliği, herhangi iki tehlike fonksiyonlarınız oranı sabit olup  $t$ 'ye bağlı olmamasıdır. Örneğin,  $Z_i$  ve  $Z_j$  ortak değişkeniyle iki bireyin tehlike fonksiyonları oranı:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp(\beta Z_i)}{\lambda_0(t) \exp(\beta Z_j)} = \frac{\exp(\beta Z_i)}{\exp(\beta Z_j)}$$

şeklindedir. Bu sebeple adı “orantılı tehlike” dir.[29] Cox regresyon parametrelerinin tahmin edilmesinde Cox kısmi olabilirlik yöntemi kullanılır.

#### 2.4.2.1.2. Cox kısmi olabilirlik fonksiyonu

En çok olabilirlik fonksiyonunda değişkenin dağılımı bilinmemektedir. Fakat Cox regresyon modelinde değişkenin dağılımı bilinmemektedir. Bu sebeple tam olarak en çok olabilirlik fonksiyonu oluşturulamaz. Cox (1972 makalesinde) sadece  $\beta$ 'ya bağlı

bir fonksiyon oluşturmuştur.  $X_i$  zamanında belli başarısızlık (ölüm, hastalıktan kurtulma veya nüksetme vb.) için, başarısızlığı gözlemlenen bireyler üzerindeki olasılık aşağıdaki gibidir.

$$l_i(\beta) = \frac{\exp(\beta Z_i)}{\sum_{j \in R(t_i)} \exp(\beta Z_j)} \quad (2.4.6)$$

Kısmi olabilirlik fonksiyonunda, sansürlü gözlemler için olasılık değeri dikkate alınır. Sansürlü veriler hesaba katılmaz. Genel olarak kısmi olabilirlik fonksiyonu, ayrı ayrı olasılıkların çarpımı şeklindedir.  $R(t_i)$ ,  $t_i$  anında risk altında bulunan bireylerin kümesi olsun. Yani  $t_i$  anına kadar yaşamış bireylerin kümesidir. O halde Cox regresyon modeli için kısmi olabilirlik fonksiyonu,

$$l(\beta) = \prod_{i=1}^n l_i(\beta) . \quad (2.4.7)$$

#### 2.4.2.2 $\beta$ katsayısının önemliliğinin test edilmesi

Cox regresyon katsayısı  $\beta$  'nın önemliliğinin test edilmesi için Wald, Benzerlik Oranı(LR) ve Skor testi kullanılmaktadır. Bu standart asimptotik olabilirlik çıkarsama testleri,  $\beta$  için kurulan hipotezlerin test edilmesinde Cox kısmi olabilirlik uygundur. Burada sıfır hipotezi  $H_0: \beta = 0$  şeklinde ifade edilmektedir.

##### **Wald testi:**

Cox regresyon analizinde kullanılan Wald testi, genel çoklu regresyon analizindeki kullanıma benzemektedir. Çoklu regresyonda, belirli regresyon katsayısının önemliliğinin test edilmesi için kullanılan t veya z testine Wald testi olarak söylenmektedir. Tek değişken için test istatistiği:

$$W = \left( \beta / se_{\beta} \right)^2 . \quad (2.4.8)$$

(2.4.8) ifadesi  $n$  'nin değeri için, 1 serbestlik dereceli  $W$  dağılımını göstermektedir. Hipotezin kararı için  $\chi^2$  tablosundan yararlanılır. [22] Eğer hesaplanan  $W$  değeri,  $\chi^2$  tablo değerinden büyük ise hipotez reddedilir ve  $\beta$ , Cox regresyon katsayısının anlamlı olduğu söylenir.

**Benzerlik oranı(LR) testi:**

Hipotezi kurulan parametre değerleri içinden maksimum yapan değer in olabirliđinin oranı olarak söylenir.  $l(.)$  Cox kısmi olabirlik fonksiyonunu göstermekte üzere, LR testi:

$$LR = 2 \left( l(\hat{\beta}) - l(\beta) \right). \quad (2.4.9)$$

$\hat{\beta}$ , Cox kısmi olabirlik fonksiyonunu maksimum yapan değer ve  $\beta$  ise hipotezde kurulan değeri göstermektedir. Birden çok örnek için LR testi, serbestlik derecesi tahmin edilen parametre sayısı olan  $\chi^2$  dağılımını ifade etmektedir. LR değeri,  $\chi^2$  tablo değeri ile karşılaştırılır. [22] Wald testinde ifade ettiğimiz gibi eđer LR değeri tablo değerinden büyük ise Cox regresyon katsayısı anlamlı olduđu söylenir.

**Skor testi:**

İki grubu karşılaştırmak için kullanılan Mantel-Haenszel Log-rank testine benzeyen skor testi, birden fazla ve sürekli deđişkenler için log-rank testinin genelleştirilmiş şekli olarak söylenir.[27]  $U(\beta)$ , Cox kısmi olabirlik fonksiyonunun skor fonksiyonu olmak üzere skor testi:

$$S = U'(\beta)I(\beta)^{-1}U(\beta). \quad (2.4.10)$$

Newton-Raphson algoritmasının birinci iterasyonu kullanılarak (2.4.10) skor testi hesaplanabilir.  $S = U'(\beta)I(\beta)^{-1}U(\beta) \sim \chi^2$  şeklinde dağılım göstermektedir. [22] Testin kararı için tahmin edilen parametre sayısına eşit serbestlik dereceli  $\chi^2$  tablo değeri ile karşılaştırılır.

**2.4.3. Parametrik olmayan sağkalım fonksiyonu için tahmin yöntemi**

Parametrik olmayan sağkalım dağılımına ait parametre için “Deneysel Olabirlik Yöntemi” ile tahmin etme yöntemi kullanılmaktadır.

**2.4.3.1. Deneysel Olabirlik Yöntemi**

Thomas ve Grunkemeier (1979), Kaplan-Meier tahmin edicisinden yararlanarak sağkalım olasılıđının güven aralıđını elde etmek için olabirlik oran testinin parametrik olmayan versiyonuna uyarlayarak yayınladı, fakat deneysel olabirlik ismini kullanmadılar. Deneysel olabirlik yaklaşımı ilk olarak bir rasgele vektörün ortalaması için güven bölgesi oluşturmak için 1988 yılında Owen tarafından tanımlanmıştır. Verinin bilinen bir dağılım ailesinden geldiđi varsayımına karşın veri odaklı olabirlik oran fonksiyonuna bađlı olarak etkili ve esnetilebilir parametrik olmayan yöntem olarak ifade edilir. Deneysel olabirlik metodu, hem parametrik

olmayan yöntem hem de olabilirlik oran fonksiyonunun avantajlarını sağlamaktadır. Oluşturulan güven bölgesinin boyutu ve yönelimi veriye bakılarak tam ve otomatik olarak karar verilmesi yöntemin avantajlarından örnek olarak verebiliriz. Asimptotik olarak güçlü olduğundan dolayı küçük örneklerde iyi sonuç verir.

Owen'ın öncülük etmesiyle deneysel olabilirlik araştırmacı sayısı her geçen gün artmaktadır ve böylece daha kapsamlı uygulamaları yapılmaktadır. Ayrıca parametrik olmayan dağılımlar için tanımlanan istatistiksel bir araç olmuştur. Bugün deneysel olabilirlik yöntemi üzerine birçok kaynak bulunmakta ve uygulamaları istatistiğin birçok alanında kullanılmaktadır. Denklem tahmini üzerine Chen ve Hall (1993), Qin ve Lawless (1994) ve iki örneklemlerle Keziou ve Leoni (2008), Ren'in çalışmaları (2008) örnek olarak gösterilir. [29] Fakat deneysel olabilirlik yöntemini istatistik yazılım programlarında uygulamak kolay değildir.

Deneysel olabilirlik (EL), dağılım varsayımı yapılmaksızın çıkarsama ve tahmin etme yöntemidir. Parametrik olmayan en çok olabilirlik tahmin işlemi olarak gösterilir. EL'nin temel özelliği kesikli dağılıma ait olan bilinmeyen dağılım fonksiyonunu tahmin etmektir. Ayrıca bunlara ek olarak, deneysel olabilirlik minimum iraksama tahmin işlemi olarak da kabul edilmektedir. Büyük sapma ilkesi(LDP) ve diğer teknikler de gösteriyor ki, deneysel olabilirlik birçok önemli özelliklere sahiptir.[4]

Deneysel olabilirlik yöntemi, sansürlü ve sansürlü olmak üzere iki tür örnekleme ait veri türü ayrı ayrı incelenmektedir.

#### 2.4.3.1.1. Sansürlü örneklem

$R^n$ 'da bilinmediği varsayılan  $F$  dağılımına ait bağımsız ve aynı dağılıma sahip rasgele vektörün bir örnekleme  $X, X_1, X_2, \dots, X_n$  olduğunu varsayalım. Herhangi kümülatif dağılım fonksiyonu  $F$  için  $X_1, X_2, \dots, X_n$  gözlemlerine bağlı olarak deneysel olabilirlik fonksiyonu,  $p_i = P_F(X = x_i) = F(x_i) - F(x_i -)$  olarak belirtildiğinde aşağıdaki gibi yazılır.[29]

$$EL(F) = EL(F, x_1, \dots, x_n) = \prod_{i=1}^n P_F(X = x_i) = \prod_{i=1}^n p_i \quad (2.4.11)$$

$\{X_1, X_2, \dots, X_n\}$ ,  $F(x)$  ortak dağılım fonksiyonunda bağımsız ve aynı dağılımlı örneklem ve  $\theta$  kitlenin ortalaması olsun. Deneysel olabilirlik dağılımın yapı iskeletini yukarıdaki ifadede farklı bir şekilde oluşturalım;  $\{p_i\}$ ,  $p_i \geq 0$  ve  $\sum_{i=1}^n p_i = 1$  ( $i = 1, \dots, n$ ) ifade edilsin ve  $I(\cdot)$  indikatör fonksiyonu olmak üzere,  $\sum_{i=1}^n p_i I(X_i \leq x)$

formundaki ağırlıklı deneysel dağılımı gerçek dağılım  $F(x)$  ile yer değiştirir. O halde ağırlıklı deneysel dağılımın ortalamasını şu şekilde yazabiliriz.

$$\theta_0 = \sum_{i=1}^n p_i I(X_i \leq x)$$

$\theta = \theta_0$  için deneysel olabilirlik fonksiyonu,

$$\prod_{i=1}^n p_i, \quad \sum_{i=1}^n p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^n p_i X_i = \theta_0 \quad (2.4.12)$$

şeklinde tanımlanır. Tanıma göre deneysel olabilirlik oranını,

$$\sup\left\{ \prod_{i=1}^n p_i, \quad \sum_{i=1}^n p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^n p_i X_i = \theta_0 \right\}, \quad (2.4.13)$$

ve log-formu yani log-deneysel olabilirlik oranı,

$$\sup\left\{ \sum_{i=1}^n \log(p_i), \quad \sum_{i=1}^n p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^n p_i X_i = \theta_0 \right\} \quad (2.4.14)$$

şeklinde yazılır. (2.4.14) uygun hale getirmek için Lagrange çarpanı metodu kullanılır.

$\lambda_0, \lambda_1 \in R$  Lagrange çarpanları olsun.  $G$  fonksiyonu tanımlayalım.

$$G = \sum_{i=1}^n \log(p_i) + \lambda_0 \left( \sum_{i=1}^n p_i - 1 \right) - n\lambda_1 \left( \sum_{i=1}^n p_i (X_i - \theta_0) \right) \quad (2.4.15)$$

$$\frac{\partial G}{\partial p_i} = \sum_{i=1}^n \frac{1}{p_i} + \lambda_0 \left( \sum_{i=1}^n 1 \right) - n\lambda_1 \left( \sum_{i=1}^n (X_i - \theta_0) \right) \quad (2.4.16)$$

$$p_i \frac{\partial G}{\partial p_i} = \sum_{i=1}^n \frac{p_i}{p_i} + \lambda_0 \left( \sum_{i=1}^n p_i \right) - n\lambda_1 \left( \sum_{i=1}^n p_i (X_i - \theta_0) \right)$$

Yukarıdaki denklemde eşitliğin sağ tarafındaki üçüncü toplam sifıra eşittir. O halde sağ taraf sifıra eşitlenirse,

$$\sum_{i=1}^n 1 + \lambda_0 = 0$$

$$\lambda_0 = -n$$

şeklinde bulunur. (2.4.16) denklemde bulunan değer yerine yazılıp  $p_i$  denklemden çekilir ve  $\lambda_1 = \lambda$  olarak yazılırsa,

$$\frac{1}{p_i} - n - n\lambda_1(X_i - \theta_0) = 0$$

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(X_i - \theta_0)} \quad i = 1, \dots, n$$

olarak bulunur. (2.4.12) deneysel olabilirlik fonksiyonundaki ifadede yukarıdaki bulunan değer yazılır.

$$\sum_{i=1}^n \frac{X_i - \theta_0}{1 + \lambda(X_i - \theta_0)} = 0$$

$l(\theta_0)$  log-deneysel olabilirlik oranını göstermek üzere,

$$l(\theta_0) = - \sum_{i=1}^n \log(1 + \lambda(X_i - \theta_0)) \quad (2.4.17)$$

Owen (1988,1990) yazılarından  $-2l(\theta_0)$  merkezi limit teoremine göre,  $\chi_1^2$  dağılıma yakınsar. (syf. 39) Bu sonuçtan  $\theta$  için  $P\{\chi_1^2 \leq \alpha\} = 1 - \alpha$  şeklinde ifade edildiğinde  $(1 - \alpha)$  seviyesinde güven bölgesini oluşturmak mümkündür.[29] O halde güven bölgesi,

$$\{\theta : -2l(\theta_0) \leq \alpha\}. \quad (2.4.18)$$

#### 2.4.3.1.2. Sansürlü örneklem

$(T_i, \delta_i)$ , bir elemanı sürekli ( $T$ ) ve bir elemanı kesikli ( $\delta$ ) dağılımdan bağımsız, aynı dağılıma sahip iki boyutlu rasgele vektör olsun. Olabilirlik fonksiyonu varsayılan dağılım fonksiyonu altında gözlemlenen örneklem olasılığıdır. Yani olabilirlik fonksiyonu,

$$\prod_{i=1}^n P(T = t_i, D = \delta_i) = \prod_{i=1}^n p_i. \quad (2.4.19)$$

$p_i, (T_i, \delta_i)$  gözlemlenen değeri için ortak olasılıktır. Bu kullanımı tam olarak doğru değildir.  $P(T = t_i, D = \delta_i)$  iki tane bir boyutlu alt dağılım veya alt sağkalım fonksiyonuna denk olacak şekilde yazılabilir.  $U_0(t) = P(T > t, \delta = 0)$  ve  $U_1(t) = P(T > t, \delta = 1)$  şeklinde iki tane bir boyutlu alt sağkalım fonksiyonu oluşturulur. O halde olabilirlik fonksiyonu alt sağkalım fonksiyonu kullanılarak aşağıdaki gibi yazılabilir:

$$\prod_{i=1}^n \{-\Delta U_1(T_i)\}^{\delta_i} \{-\Delta U_0(T_i)\}^{1-\delta_i}. \quad (2.4.20)$$

Burada  $\Delta U_1(t) = U_1(t) - U_1(t-) = -P(T = t, \delta = 1)$  şeklinde tanımlanır. Burada rasgele ve bağımsız sansürlü model varsayalım. O halde alt sağkalım fonksiyonlarını  $T_i = \min(X_i, C_i)$  ve  $\delta_i = I[X_i \leq C_i]$  olmak üzere,

$$U_1(t) = P(T > t, \delta = 1) = \int_t^{\infty} (1 - G(s)) dF(s)$$

$$U_0(t) = P(T > t, \delta = 0) = \int_t^{\infty} (1 - F(s)) dG(s)$$

Yukarıdaki denklemde  $G(\cdot)$ ,  $C_i$  sansürleme süresinin kümülatif dağılım fonksiyonu ve  $F(\cdot)$  ise  $X_i$  sağkalım süresinin kümülatif dağılım fonksiyonudur. İki alt sağkalım dağılımları  $U_0$  ve  $U_1$ , bağımsız rasgele sansürlü model altında  $F(s)$  ve  $G(s)$  dağılımlarına denktir. Bağımsız ve aynı dağılıma sahip sansürlü  $n$  gözlem için yukarıdaki ifadelere denk olan sağkalım fonksiyonları için  $U_0$  ve  $U_1$  tekrar ifade edelim.

$$U_0(t) = \Pr(t_i = t, \delta_i = 0) = \Pr(C_i = t, T_i > C_i) = g(t)S(t)$$

$$U_1(t) = \Pr(t_i = t, \delta_i = 1) = \Pr(T_i = t, T_i \leq C_i) = f(t)G(t).$$

$F(\cdot)$  ve  $G(\cdot)$  dağılımları açısından deneysel olabilirlik fonksiyonu,

$$\prod_{i=1}^n [(1 - G(t_i))dF(t_i)]^{\delta_i} [(1 - F(t_i))dG(t_i)]^{1-\delta_i} \quad (2.4.21)$$

şeklinde ifade edilir. (2.4.21) fonksiyonunda  $F$  sağkalım hakkında bilgi verirken  $G$  takip süresinin özelliğiyle ilgili bilgi içermektedir. Bu sebeple  $F$  içermeyen terimler sabit düşünülür ve denklemde göz ardı edilebilir. O halde deneysel olabilirlik fonksiyonu en sade haliyle,

$$EL(F) = \prod_{i=1}^n [dF(t_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \quad (2.4.22)$$

olarak yazılır. Genellikle bu ifadeye denk olacak şekilde aşağıdaki ifade de kullanılabilir.[10,29]

$$EL(F) = \prod_{i=1}^n [\Delta F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \quad (2.4.23)$$

### 2.4.3.1.3 Tehlikeye Bağlı Olarak Deneysel Olabilirlik Fonksiyonu

$$1 - F(t) = \prod_{s \leq t} (1 - \Delta\Lambda(s)) \quad (2.4.24)$$

$$\Delta\Lambda(t) = \frac{\Delta F(t)}{1 - F(t-)} \quad (2.4.25)$$

Yukarıdaki tehlike ve dağılım fonksiyonu arasındaki ilişkiyi veren (2.4.24) ve (2.4.25) ifadeleri kullanılarak olabilirlik fonksiyonunu tekrardan yazalım.

$$\begin{aligned} EL(F) &= \prod_{i=1}^n [\Delta F(T_i)]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \\ EL(\Lambda) &= \prod_{i=1}^n [\Delta\Lambda(T_i)(1 - F(T_i-))]^{\delta_i} \left[ \prod_{s \leq t} (1 - \Delta\Lambda(s)) \right]^{1-\delta_i} \\ &= \prod_{i=1}^n \left[ \Delta\Lambda(T_i) \prod_{s < T_i} (1 - \Delta\Lambda(s)) \right]^{\delta_i} \left[ \prod_{s \leq T_i} (1 - \Delta\Lambda(s)) \right]^{1-\delta_i} \\ &= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \left[ \prod_{s < T_i} (1 - \Delta\Lambda(s)) \right]^{\delta_i} \left[ \prod_{s \leq T_i} (1 - \Delta\Lambda(s)) \right]^{1-\delta_i} \\ &= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \left[ \prod_{s \leq T_i} (1 - \Delta\Lambda(s)) \right]^{\delta_i - (1-\delta_i)} \\ &= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \underbrace{\prod_{s \leq T_i} (1 - \Delta\Lambda(s))}_{1 - F(T_i) = S(T_i)} \\ &= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \exp(-\Lambda(T_i)) \end{aligned} \quad (2.4.26)$$

şeklinindedir. Bu çarpım Murphy(1994) tarafından olabilirliğin bir Poisson uzantısı olarak da söylenir ve çarpım Cox model için deneysel olabilirlik fonksiyonu yazımında kullanılır. Dağılım fonksiyonu kesikli iken,

$$EL(\Lambda) = \prod_{i=1}^n \left[ \Delta\Lambda(T_i) \prod_{j: T_j < T_i} (1 - \Delta\Lambda(T_j)) \right]^{\delta_i} \left[ \prod_{j: T_j \leq T_i} (1 - \Delta\Lambda(T_j)) \right]^{1-\delta_i}$$

$$\begin{aligned}
&= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \left[ \prod_{j:T_j < T_i} (1 - \Delta\Lambda(T_j)) \right]^{\delta_i} \left[ \prod_{j:T_j < T_i} (1 - \Delta\Lambda(T_j)) \right]^{1-\delta_i} [1 \\
&\quad - \Delta\Lambda(T_i)]^{1-\delta_i} \\
&= \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} \prod_{i=1}^n \prod_{j:T_j < T_i} (1 - \Delta\Lambda(T_j)) \prod_{i=1}^n [1 - \Delta\Lambda(T_i)]^{1-\delta_i}
\end{aligned}$$

Eşitlikteki ikinci çarpımda aşağıdaki düzenleme yapıp denklemde yazalım.

$$\begin{aligned}
&= \prod_{i=1}^n \prod_{j:T_j < T_i} (1 - \Delta\Lambda(T_j)) \\
i = 2 \text{ için} &= 1 - \Delta\Lambda(T_1) \\
i = 3 \text{ için} &= (1 - \Delta\Lambda(T_1))(1 - \Delta\Lambda(T_2)) \\
&\dots \\
i = n \text{ için} &= (1 - \Delta\Lambda(T_1))(1 - \Delta\Lambda(T_2)) \dots (1 - \Delta\Lambda(T_{n-1})) \\
&= (1 - \Delta\Lambda(T_1))^{n-1} (1 - \Delta\Lambda(T_2))^{n-2} \dots (1 - \Delta\Lambda(T_{n-1}))^1 \\
&= \prod_{i=1}^{n-1} [1 - \Delta\Lambda(T_i)]^{n-i}
\end{aligned}$$

Olabilirlik fonksiyonu:

$$= \prod_{i=1}^{n-1} [\Delta\Lambda(T_i)]^{\delta_i} [1 - \Delta\Lambda(T_i)]^{n-i} [1 - \Delta\Lambda(T_i)]^{1-\delta_i} \Delta\Lambda(T_n)^{\delta_n} (1 - \Delta\Lambda(T_n))^{1-\delta_n}$$

Kesikli dağılım fonksiyonlarında son sıçrayışın 1 olması ve  $\Delta\Lambda(T_n) = \delta_n$  olmasından denklemde son kısımdaki çarpımdan 1 bulunur ve son olarak denklemi aşağıdaki şekilde yazabiliriz.[15-18,29]

$$EL(\Lambda) = \prod_{i=1}^n [\Delta\Lambda(T_i)]^{\delta_i} [1 - \Delta\Lambda(T_i)]^{n-i+1-\delta_i} \quad (2.4.27)$$

### 3.COX-DENEYSEL OLABİLİRLİK

Bu kısımda yayınlanmış Düzeltmiş Deneysel Olabilirlik metodunu büyük örneklem teorisinin temel sonuçlarının tanımlamasını ve Cox regresyon modeli üzerinde tartışacağız.  $t_i$  , sağkalım süresi ve  $c_i$  , sansürleme süresi olmak üzere;

$x_i = \min(t_i, c_i)$  ve  $\delta_i = I(x_i = t_i)$  şeklinde tanımlansın.  $Z$  ,  $p$  boyutlu ortak değişken vektörü ve  $i = 1,2,3, \dots, n$  için  $(t_i, c_i, Z_i)$  bağımsız ve aynı dağılımlıdır. Yani,  $t_i$  sağkalım süresi ve  $c_i$  sansürleme süresi verilen  $Z_i$  değişkeninden bağımsızdır.

1975 yılında Cox,  $\beta_0$  'ın gerçek değeri için çıkarsama yapabilmek için kısmi olabilirliği yayınlamıştır. Cox'un ileri sürdüğü kısmi olabilirlik fonksiyonu  $L(\beta)$  aşağıdaki gibidir.

$$L(\beta) = \prod_{i=1}^n \left[ \frac{\exp(\beta Z_i)}{\sum_j^n \exp(\beta Z_j) I[x_j \geq x_i]} \right]^{\delta_i}$$

$\log L(\beta)$ 'nın birinci türevi alınırsa elde edilen fonksiyon skor fonksiyonudur ve  $U(\beta)$  olarak gösterilecektir. O halde

$$\log L(\beta) = \sum_{i=1}^n \delta_i \left[ \beta Z_i - \log \left( \sum_j \exp(\beta Z_j) I[x_j \geq x_i] \right) \right]$$

$$U(\beta) = \frac{\partial}{\partial \beta} \log L(\beta)$$

şeklinde yazabiliriz.

$$U(\beta) = \sum_{i=1}^n \delta_i \left[ Z_i - \frac{\sum_j Z_j \exp(\beta Z_j) I[x_j \geq x_i]}{\sum_j \exp(\beta Z_j) I[x_j \geq x_i]} \right]$$

Skor fonksiyonu yukarıdaki şekilde bulunmuş olur.

$$\hat{\alpha}_0(t, \beta) = n^{-1} \sum_i \exp(\beta Z_i) I[x_i \geq t], \hat{\alpha}_1(t, \beta) = n^{-1} \sum_i Z_i \exp(\beta Z_i) I[x_i \geq t] \text{ ve}$$

$N_i(t) = I[x_i \leq t, \delta_i = 1]$  olacak şekilde alınırsa skor fonksiyonu aşağıdaki gibi yazılabilir.

$$U(\beta) = \sum_{i=1}^n \int_0^T \left( Z_i - \frac{\hat{\alpha}_1(t, \beta)}{\hat{\alpha}_0(t, \beta)} \right) dN_i(t)$$

Andersen ve Gill (1982) teoreminden,

$$I(\beta_0) = \lim_{n \rightarrow \infty} \hat{I}(\beta_0)$$

ve

$$\hat{I}(\beta) = n^{-1} \sum_{i=1}^n \delta_i \left( \frac{\hat{\alpha}_2(x_i, \beta)}{\hat{\alpha}_0(x_i, \beta)} - \left( \frac{\hat{\alpha}_1(x_i, \beta)}{\hat{\alpha}_0(x_i, \beta)} \right)^{\otimes 2} \right)$$

$\alpha^{\otimes 2} = \alpha\alpha^T$  ve  $\hat{\alpha}_2(x_i, \beta) = n^{-1} \sum_i Z_i^{\otimes 2} \exp(\beta Z_i) I[x_i \geq t]$  olmak üzere bilgi matrisi (information matrix) iken

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, I(\beta_0)^{-1})$$

olduğunu söyleyebiliriz. O halde normal yaklaşımı metoduna göre  $\beta$  için %100(1 -  $\alpha$ ) asimptotik güven bölgesini  $\chi_p^2(\alpha)$  p serbestlik dereceli ki-kare dağılımı olmak üzere aşağıdaki gibi yazılır.

$$R_1 = \left\{ \beta : n(\hat{\beta} - \beta)^T \hat{I}(\beta)(\hat{\beta} - \beta) \leq \chi_p^2(\alpha) \right\}$$

### 3.1. Düzeltilmiş Deneysel Olabilirlik Güven Bölgesi

$$M_i(t) = N_i(t) - \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] \lambda_0(s) ds$$

$i = 1, 2, \dots, n$  için martingale artık ve gerçek  $\beta_0$  için  $E(U(\beta_0)) = 0$  olacak şekilde skor fonksiyonunu tekrardan yazalım.

$$U(\beta_0) = \sum_{i=1}^n \int_0^T \left( Z_i - \frac{\hat{\alpha}_1(t, \beta)}{\hat{\alpha}_0(t, \beta)} \right) dM_i(t)$$

Cox regresyon modelinde temel tehlike fonksiyonu  $\lambda_0$  tanımlı olmadığı için tahmin edilmesi gerekmektedir. Bu sebepten yukarıdaki eşitlikte  $\lambda_0(s)ds$  yerine  $d\hat{\Lambda}_0(s, \beta_0)$  yazılır. Martingale artıklarının özelliklerinden artıkların ortalaması sıfırdır özelliği kullanılarak kümülatif tehlike fonksiyonun tahminini bulabiliriz.

$$\sum_{i=1}^n M_i(t) = 0$$

$$\sum_{i=1}^n N_i(t) - \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] d\Lambda_0(s) = 0$$

$$N_i(t) - \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] d\Lambda_0(s) = 0$$

$$N_i(t) = \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] d\Lambda_0(s)$$

$$dN_i(t) = d \left( \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] d\Lambda_0(s) \right)$$

$$dN_i(t) = \sum_{t_j \leq t} \exp(\beta_0 Z_j) I[x_j \geq t] d\Lambda_0(t)$$

$$d\Lambda_0(t) = \frac{dN_i(t)}{\sum_{t_j \leq t} \exp(\beta_0 Z_j) I[x_j \geq t]}$$

Her iki tarafından  $[0, t]$  aralığında integrali alınırsa,

$$\widehat{\Lambda}_0(t, \beta_0) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{t_j \leq s} \exp(\beta_0 Z_j) I[x_j \geq s]}$$

şeklinde bulunur.

Tahmin edilen kümülatif temel tehlike fonksiyonu martingale artıktaki yazılır ve bulunan yeni tahmin edilmiş martingale artığı skor fonksiyonunda yazılır.

$$\widehat{M}_i(t) = N_i(t) - \int_0^t \exp(\beta_0 Z_i) I[x_i \geq s] d\widehat{\Lambda}_0(s, \beta_0)$$

$$U(\beta_0) = \sum_{i=1}^n \int_0^T \left( Z_i - \frac{\widehat{\alpha}_1(t, \beta)}{\widehat{\alpha}_0(t, \beta)} \right) d\widehat{M}_i(t)$$

$p = (p_1, p_2, \dots, p_n)$  olasılık vektörü olmak üzere, o halde  $p_i \geq 0$  ve  $\sum_{i=1}^n p_i = 1$  olduğu söylenebilir.

$$W_{n,i} = \int_0^T \left( Z_i - \frac{\widehat{\alpha}_1(t, \beta)}{\widehat{\alpha}_0(t, \beta)} \right) d\widehat{M}_i(t)$$

$1 \leq i \leq n$  için yukarıdaki şekilde tanımlansın.

$\prod_{i=1}^n p_i$  çarpımı,  $\sum_{i=1}^n p_i = 1$  koşulu altında  $p_i = n^{-1}$  de maksimumunu  $n^{-n}$  olarak elde edileceği biliniyor. Bu sebepten gerçek  $\beta_0$  değeri için hesaplanan plug-in tahmini ile deneysel olabilirlik oranı aşağıdaki gibi yazılabilir.

$$R(\beta_0) = \sup \left\{ \prod_{i=1}^n np_i : \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i W_{n,i} = 0, p_i \geq 0 \right\}$$

Verilen iki koşul altında Lagrange çarpanı metodu kullanılarak  $R(\beta_0)$ 'ı maksimum yapan  $p_i$  değerini bulabiliriz.

$$G = \log \left( \prod_{i=1}^n np_i \right) = \sum_{i=1}^n \log np_i$$

G fonksiyonunun iki koşul altında Lagrange çarpanı metodu ile istenileni bulabiliriz.

$$G = \sum_{i=1}^n \log np_i - \lambda_0 \left( \sum_{i=1}^n p_i - 1 \right) - \lambda_1 \left( \sum_{i=1}^n p_i W_{n,i} - 0 \right)$$

$$\frac{\partial G}{\partial p_i} = \sum_{i=1}^n \frac{n}{p_i} - \lambda_0 \left( \sum_{i=1}^n 1 \right) - \lambda_1 \left( \sum_{i=1}^n W_{n,i} \right)$$

$$= \sum_{i=1}^n \frac{n}{p_i} - \lambda_0 - \lambda_1 W_{n,i}$$

$$\frac{n}{p_i} - \lambda_0 - \lambda_1 W_{n,i} = 0$$

$$p_i = \frac{1}{n} \frac{1}{\lambda_0 + \lambda_1 W_{n,i}}$$

$\lambda_0 = 1$  ve  $\lambda_1 = \lambda$  olarak alınırsa  $i = 1, 2, \dots, n$  için,

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^\tau W_{n,i}}$$

bulunur.  $\sum_{i=1}^n p_i W_{n,i} = 0$  denkleminde  $p_i$  değerini yazalım.

$$\sum_{i=1}^n p_i W_{n,i} = \frac{1}{n} \sum_{i=1}^n \frac{W_{n,i}}{1 + \lambda^\tau W_{n,i}} = 0 \quad (3.1.1)$$

Buna göre  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)^\tau$  denklemin çözümüdür diyebiliriz. Buradan,

$$R(\beta_0) = \prod_{i=1}^n np_i = \prod_{i=1}^n \{1 + \lambda^\tau W_{n,i}\}^{-1}$$

şeklinde yazabiliriz.

$\hat{I}(\beta) = -2 \log R(\beta)$  deneysel olabilirlik oranıdır.  $W_{n,i}$ ,  $\lambda_0(t)$ 'nin tahmini içerdiğini önceki adımlarda gösterilmişti ve buna bağlı olarak  $W_{n,i}$ 'lerin bağımsız olduğu

söylenir.  $\hat{I}(\beta)$ 'nın limit dağılımı standart p serbest dereceli  $\chi^2$  dağılımından büyük değildir. Güven bölgesi bulmak için kullanılan yöntemlerden biri olan Wilks teoreminin mümkün kılmak için düzeltilmesi gereklidir. Düzeltilmiş deneysel olabilirlik yöntemi konusunda Rao ve Scott (1981), Wang ve Rao (2001) gibi birçok kişi tarafından yayımlar yapılmıştır.

$\hat{I}_1(\beta_0) = n^{-1} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau$  ve  $I_1(\beta_0) = \lim_{n \rightarrow \infty} \hat{I}_1(\beta_0)$  şeklinde ifade edilsin.

$$\hat{\rho}(\beta) = \frac{\text{trace}\{\hat{I}^{-1}(\beta)\hat{I}(\beta)\}}{\text{trace}\{\hat{I}_1^{-1}(\beta)\hat{I}(\beta)\}}$$

$\text{trace}(\cdot)$  matrisin izini ifade etmektedir. Tersini alma işlemi veya başka bir deyişle matristeki köşegen elemanlarının toplamıdır.

$$\hat{S}(\beta) = \left\{ \sum_{i=1}^n W_{n,i}(\beta) / n^{1/2} \right\} \times \left\{ \sum_{i=1}^n W_{n,i}(\beta) / n^{1/2} \right\}^\tau$$

Şeklinde tanımlayalım.  $\hat{\rho}(\beta)$  denkleminde değişiklik yapılarak  $\hat{r}(\beta)$ 'yı tanımlayalım.

$$\hat{r}(\beta) = \frac{\text{trace}\{\hat{I}^{-1}(\beta)\hat{S}(\beta)\}}{\text{trace}\{\hat{I}_1^{-1}(\beta)\hat{S}(\beta)\}}$$

O halde yeni tanımlayacağımız düzeltilmiş deneysel olabilirlik oranını aşağıdaki gibi ifade edebiliriz.

$$\hat{I}_{ad}(\beta) = \hat{r}(\beta)\hat{I}(\beta)$$

Teorem :  $\lambda_0(t)$  sürekli ve ortak değişken vektörü  $Z_i$  sınırlı olduğu varsayımı altında,  $\hat{I}_{ad}(\beta_0)$  istatistiği p serbestlik dereceli ki-kare ( $\chi_p^2$ ) dağılımına yakınsar. Buradan  $\beta$  için  $\%100(1 - \alpha)$  asimptotik düzeltilmiş deneysel olabilirlik güven bölgesi,

$$R_2 = \{\beta : \hat{I}_{ad}(\beta) \leq \chi_p^2(\alpha)\}. \quad [28]$$

İspat: İlk olarak  $\hat{I}(\beta)$ 'nın  $\chi_p^2$  dağılımına yakınsadığını gösterelim. Bunun için öncelikle iki tane Lemma ifade edelim.

Lemma 1: [Fleming ve Harrington] Teoremdeki aynı koşullara bağlı olarak,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n W_{n,i} \xrightarrow{\mathcal{L}} N\left(0, \sum(T, \beta_0)\right).$$

Lemma 2: Teoremdeki kořullara uyularak,

$$(i). \frac{1}{n} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \xrightarrow{P} \sum(T, \beta_0) \quad (ii). \max \|W_{n,i}\| = o_p(n^{1/2})$$

Ayrıca,

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \text{ ve } S = \frac{1}{n} \sum_{i=1}^n W_i W_i^\tau \text{ olmak üzere,}$$

$$\frac{1}{n} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \xrightarrow{P} \sum(T, \beta_0) \text{ iken Büyük Sayılar Yasasına göre,}$$

$$\frac{1}{n} \sum_{i=1}^n W_i W_i^\tau \xrightarrow{P} \sum(T, \beta_0)$$

dir. Yani,

$$\hat{S}_n = S + o_p(1) \quad (3.1.2)$$

dir.

Teoremin ispatına dönersek,

$$\begin{aligned} R(\beta_0) &= \prod_{i=1}^n \{1 + \lambda^\tau W_{n,i}\}^{-1} \\ \hat{I}(\beta) &= -2 \log R(\beta_0) \\ &= -2 \log \prod_{i=1}^n \{1 + \lambda^\tau W_{n,i}\}^{-1} \\ &= 2 \sum_{i=1}^n \log(1 + \lambda^\tau W_{n,i}) \end{aligned}$$

Yukarıdaki ifade de Taylor açılımı uygulayarak,

$$\hat{I}(\beta) = 2 \sum_{i=1}^n \left( \lambda^\tau W_{n,i} - \frac{1}{2} (\lambda^\tau W_{n,i})^2 \right) + r_n \quad (3.1.3)$$

$|r_n| \leq C \sum_{i=1}^n (\lambda^\tau W_{n,i})^3$  olacak şekilde toplamı sınırlandıralım.

$\rho \geq 0$  ve  $\|\theta\| = 1$  iken  $\lambda = \rho\theta$  olsun. Lemma 1, Lemma 2(i) ve Owen(1990)'da benzer ispatlar uygulanarak,

$$\|\lambda\| = O_p(n^{-1/2}). \quad (3.1.4)$$

Lemma 2, (3.1.2) ve (3.1.4)'ye göre,

$$|r_n| \leq C \sum_{i=1}^n (\lambda^\tau W_{n,i})^3 \leq C \underbrace{\|\lambda\|^3}_{O_p(n^{-\frac{3}{2}})} \underbrace{\max_{1 \leq i \leq n} \|W_{n,i}\|}_{O_p(n^{\frac{1}{2}})} \underbrace{\sum_{i=1}^n \|W_{n,i}\|^2}_{O_p(n)} = o_p(1) \quad (3.1.5)$$

elde edilir. (3.1.1) ifadesini düzenleyelim.

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{W_{n,i}}{1 + \lambda^\tau W_{n,i}} &= 0 \\ \frac{1}{n} \sum_{i=1}^n W_{n,i} \left[ 1 - \lambda^\tau W_{n,i} + \frac{(\lambda^\tau W_{n,i})^2}{1 + \lambda^\tau W_{n,i}} \right] &= 0 \\ \frac{1}{n} \sum_{i=1}^n W_{n,i} - \left( \frac{1}{n} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right) \lambda + \frac{1}{n} \sum_{i=1}^n \frac{W_{n,i} (\lambda^\tau W_{n,i})^2}{1 + \lambda^\tau W_{n,i}} &= 0 \end{aligned} \quad (3.1.6)$$

(3.1.6)'teki eşitlikten  $\lambda$ 'yı çekelim.

$$\lambda = \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n W_{n,i} + \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n \frac{W_{n,i} (\lambda^\tau W_{n,i})^2}{1 + \lambda^\tau W_{n,i}} \quad (3.1.7)$$

(3.1.7) eşitliğinin sağ tarafındaki ikinci işlemde düzenleme yapalım.

$$\begin{aligned} \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n \frac{W_{n,i} (\lambda^\tau W_{n,i})^2}{1 + \lambda^\tau W_{n,i}} &\leq \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n \frac{W_{n,i} (\lambda^\tau W_{n,i})^2}{\lambda^\tau W_{n,i}} \\ &\leq \underbrace{\left( \sum_{i=1}^n \|W_{n,i}\|^2 \right)^{-1}}_{O_p(n^{-1})} \underbrace{\|\lambda\|^1}_{O_p(n^{-1/2})} \underbrace{\sum_{i=1}^n \|W_{n,i}\|^2}_{O_p(n)} = o_p(n^{-1/2}) \end{aligned}$$

O halde  $\lambda$ 'yı aşağıdaki gibi yazabiliriz.

$$\lambda = \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n W_{n,i} + o_p(n^{-\frac{1}{2}}) \quad (3.1.8)$$

Denklem (3.1.1)'i  $\lambda$  ile çarpıp tekrar ele alalım ve yukarıdaki adımlara benzer şekilde ilerleyelim. Burada  $\lambda$  sabit bir sayı olduğu için eşitlik bozulmaz.

$$\begin{aligned}
\sum_{i=1}^n \frac{\lambda^\tau W_{n,i}}{1 + \lambda^\tau W_{n,i}} &= 0 \\
\sum_{i=1}^n \lambda^\tau W_{n,i} \left[ 1 - \lambda^\tau W_{n,i} + \frac{(\lambda^\tau W_{n,i})^2}{1 + \lambda^\tau W_{n,i}} \right] &= 0 \\
\sum_{i=1}^n \lambda^\tau W_{n,i} - \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + \sum_{i=1}^n \frac{(\lambda^\tau W_{n,i})^3}{1 + \lambda^\tau W_{n,i}} &= 0 \tag{3.1.9}
\end{aligned}$$

(3.1.9) eşitliğinin sol tarafındaki üçüncü toplamı düzenleyelim.

$$\sum_{i=1}^n \frac{(\lambda^\tau W_{n,i})^3}{1 + \lambda^\tau W_{n,i}} \leq \sum_{i=1}^n \frac{(\lambda^\tau W_{n,i})^3}{\lambda^\tau W_{n,i}} \leq \|\lambda\|^2 \sum_{i=1}^n \|W_{n,i}\|^2 = o_p(1)$$

Buradan,

$$\sum_{i=1}^n \lambda^\tau W_{n,i} = \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + o_p(1). \tag{3.1.10}$$

(3.1.10)'de elde edilen ifade (3.1.3) denkleminde yerine yazalım.

$$\begin{aligned}
\hat{I}(\beta) &= 2 \sum_{i=1}^n \left( \lambda^\tau W_{n,i} - \frac{1}{2} (\lambda^\tau W_{n,i})^2 \right) + r_n \\
&= 2 \sum_{i=1}^n \lambda^\tau W_{n,i} - \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + r_n \\
&= 2 \left( \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + o_p(1) \right) - \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + o_p(1) \\
&= \sum_{i=1}^n (\lambda^\tau W_{n,i})^2 + o_p(1) \\
&= \sum_{i=1}^n \lambda^\tau W_{n,i} W_{n,i}^\tau \lambda + o_p(1) \\
&= \lambda^\tau \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right) \lambda + o_p(1)
\end{aligned}$$

(3.1.8)'de bulunan  $\lambda$  değeri yukarıda yazalım.

$$\begin{aligned}
\hat{I}(\beta) &= \left( \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n W_{n,i} \right)^\tau \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right) \left( \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n W_{n,i} \right) \\
&\quad + o_p(1) \\
&= \left( \sum_{i=1}^n W_{n,i} \right)^\tau \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right) \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \sum_{i=1}^n W_{n,i} + o_p(1) \\
&= \left( \sum_{i=1}^n W_{n,i} \right)^\tau \left( \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \left( \sum_{i=1}^n W_{n,i} \right) + o_p(1)
\end{aligned}$$

Lemma 1 ve Lemma 2'yi kullanabilmek ve  $\chi^2$ 'ye yakınsadığını göstermek için denklem üzerinde değişiklik yapalım.

$$\hat{I}(\beta) = \left( n^{-1/2} \sum_{i=1}^n W_{n,i} \right)^\tau \left( n^{-1} \sum_{i=1}^n W_{n,i} W_{n,i}^\tau \right)^{-1} \left( n^{-1/2} \sum_{i=1}^n W_{n,i} \right) + o_p(1) \quad (3.1.11)$$

Böylece Lemma 1 ve Lemma 2'den (3.1.11)'de  $\hat{I}(\beta)$ 'nin  $p$  serbestlik dereceli  $\chi^2$  dağılımına yakınsadığı söylenebilir.

Düzeltilmiş deneysel olabilirlik oranının tanımından,

$$\hat{I}_{ad}(\beta) = \left( n^{-1/2} \sum_{i=1}^n W_{n,i} \right)^\tau \hat{I}^{-1} \left( n^{-1/2} \sum_{i=1}^n W_{n,i} \right) + o_p(1)$$

şeklinde yazılabilir ve  $\hat{I} \xrightarrow{P} I$  yakınsadığı bilindiğine göre teorem ispatlanmış olur.

[18-20,28]

■

## 4. UYGULAMA

Cox regresyon modelinde sağkalım parametrelerin tahminini deneysel olabilirlik yöntemiyle bulunması ile ilgili üç sağdan sansürlü gerçek veri üzerinde uygulaması yapılmıştır. Her uygulamada verinin hikayesi verilmiş ve R programındaki sonuçlar aşağıda verilmiştir.

### 4.1 Uygulama 1(Gastrit Kanseri Verisi)

1975 (Şubat) ve 1977 (Ocak) arasındaki dönemde bölgesel olarak ilerlemiş gastrit kanserli 96 hasta bu çalışmaya kabul edildi. Tedaviye başlamadan önce dört hasta iptal edildi, bunlardan biri karma tedavi ile ilgili yere atandı. Tedaviye başlayan iki hastanın daha sonra uygun olmadığı veya değerlendirilemez olduğu belirlendiği dâhil edilmedi. Daha sonra birine lenfoma tanısı konuldu ve protokol tarafından belirlenen süre içinde tedaviye başlanmadı. Kalan 90 hastanın hepsinde midede histolojik olarak kanıtlanmış adenokarsinom vardı ve hiçbiri daha önce radyoterapi veya kemoterapi ile tedavi edilmedi. Kalan tümör kütlelerinin boyutu şu şekilde tanımlanmıştır: (1) cerrah tarafından görülebilen ve kolayca tanımlanabilen brüt artık tümör; veya (2) rezeksiyon kenarlarında tümörün varlığından veya histolojik olarak dâhil olduğu belgelenen birincil tümör veya lenf nodu bölgelerinin hoş olmayan rezeksiyonundan kaynaklanan mikroskopik kalıntıya odaklar. Bu çalışmaya giren tüm hastalar, cerrahi müdahale yapılan bir hasta popülasyonunu içeren destek kemoterapinin ayrı bir denemesi için uygun görülmedi. Aşağıdaki tabloda iki farklı tedavi grubundaki hastaların karakterleri hakkında yüzdeleri verilmiştir.[2]

Tablo 4.1. Tedavi gruplarının hasta karakterleri

Klinik Özellik	Hastaların yüzdeliği (%)	
	Kemoterapi (45 hasta)	Karma kemoterapi/işin tedavi (45 hasta)
<b>Yaş</b>		
60	42	51
60	58	49
<b>Cinsiyet</b>		
Erkek	67	73
Kadın	33	27
<b>Başlangıç Performans Durumu*</b>		
0-1	76	64
2-3	24	36
<b>Patoloji Konumu</b>		
Mide Ağzı	20	24
Başka bir yer	80	76

<b>Büyük Patern</b>		
Linitis Plastica	27	20
Ulcerating	49	53
Diğer	24	27
<b>Diferansiyasyon</b>		
İyi	6	11
Orta	41	38
Zayıf	53	51
<b>Cerrahi(Ameliyat)</b>		
İlk tümörün Çıkarılması	76	71
Çıkarma Olmaması	24	29
<b>Kalan Tümör</b>		
Büyük	49	56
Mikroskopik	51	44

\*Performans Durumu(0-4):

0: Tamamen aktif, restriksiyon olmadan hastalık öncesi tüm performansı sürdürebilir.

1: Fiziksel olarak hareketli aktivitelerde kısıtlı ama ayakta tedavi edilebilir ve kolay veya hareketsiz işlerini yapabilir.

2: Ayakta tedavi ve kendi bakımını yapabilir, fakat herhangi bir iş aktivitesi yapamaz.

3: Sadece sınırlı kendi bakımını yapabilir, yürüme zamanının %50'sinden daha fazlasını yatarak veya oturarak geçirir.

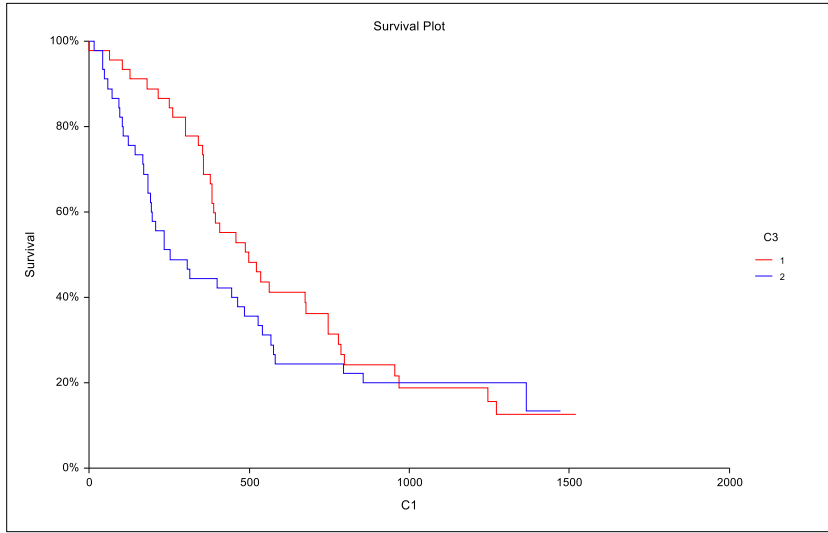
4: Tamamen etkisizdir, herhangi kendi bakımını yapamaz. Tamamen yatağa veya koltuğa mahkumdur.

Gastrit kanseri hastaların tedavi bağımsız değişkenine göre hesaplama yapıldığında aşağıdaki sonuçlar elde edilmiştir.

Tablo 4.2 Gastrit kanseri verisi R sonuçları

	coef	exp(coef)	se(coef)	z	p
Tedavi	0.265	1.303	0.233	1.14	0.26

Tabloya baktığımızda tedavi değişkeni için regresyon katsayısı 0.265 olarak hesaplanmıştır. Burada katsayının işaretine baktığımızda pozitif olması ölüm riskinin daha yüksek olduğunu göstermektedir. Bu veri setine göre karma tedavi alan hastaların, kemoterapi alan hastalara göre ölüm riski daha yüksektir. Tehlike oranı için  $1/\exp(\text{coef})$  değerine bakıldığında 0.77 olarak bulunur ve buna göre kemoterapi alanların %77 oranında ölüm riski daha azdır.



Şekil 4.1. Tedavi yöntemlerine göre sağkalım olasılığı

Şekil 4.1’de verilen grafiğe göre, 1.yöntem yani sadece kemoterapi alan hastaların 2.yöntemde karma tedavi olan hastalara göre sağkalım süresi 500 altındaki sağkalım olasılığı daha büyüktür. Sağkalım süresi 1000 ve üzerinde olan hastalarda karma tedavi uygulanan hastaların sağkalım olasılığı daha büyüktür.

$T \leq 365$  için maksimum kısmi olabilirlik tahmin edicisi  $\hat{\beta}$  ve sıfır hipotezinin testi için iki deneysel olabilirlik(EL) değeri ve olabilirlik oranı R programında bulunan sonuçlar aşağıda verilmiştir.

Maksimum kısmi olabilirlik tahmin edicisi  $\hat{\beta}$  değeri yukarıdaki Tablo 4.2’de belirtildiği gibi 0.265 olarak hesaplanmıştır. Hipotezimiz,

$$H_0: \beta = 0.2, \mu = \Lambda_0(365) = 0.43.$$

Tablo 4.3.  $\beta = 0.2$  ve  $\lambda = 0$  için EL

mu	0.4222382
logPlik	-282.7143
<b>logEmpLik1</b>	<b>-356.7143</b>

Tablo 4.4.  $\beta = \hat{\beta}$  ve  $\lambda = 0$  için EL

mu	0.3837017
logPlik	-282.6758
<b>logEmpLik2</b>	<b>-356.6758</b>

Olabilirlik oranı:

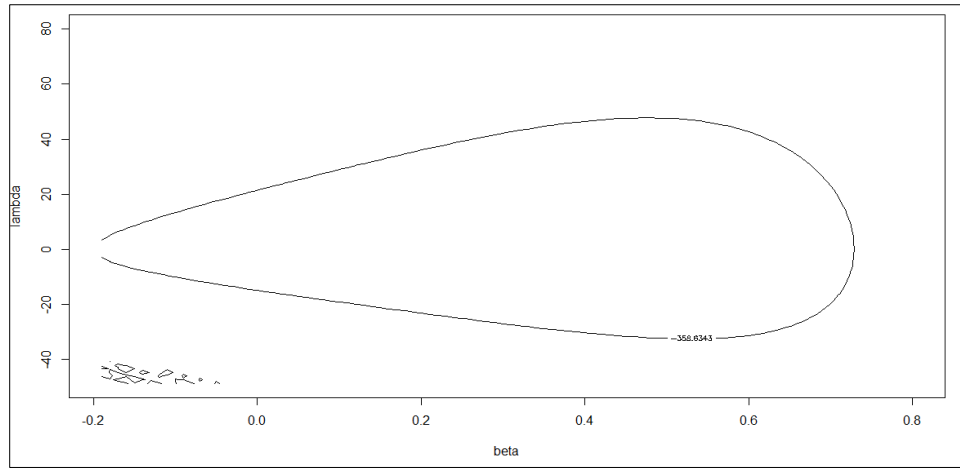
$$-2(\log Emplik1 - \log Emplik2) = -2(-356.7143 - (-356.6758)) = 0.077.$$

Anlamlılık seviye değeri  $p = 0.05$  'den büyük ve 2 serbestlik dereceli  $\chi^2$  değeri 3.84'ten küçüktür. İki duruma göre bakıldığında hipotezimiz reddedilemez.

Tablo 4.5.  $\beta = 0.2$  ve  $\lambda = -1.62$  için EL

mu	<b>0.430008</b>
logPlik	-282.7143
logEmpLik	-356.7206

$\log Lik = \log Lik(\beta = 0.2, \lambda = -1.62) = -356.7206$  .  $T \leq 365$  iken sağkalım olasılığı 0.43 e eşit olarak bulunmuştur. Güven aralığı aşağıdaki gibidir.



Şekil 4.2 Güven aralığı

## 4.2. Uygulama 2(Akciğer Kanseri Verisi)

Bu çalışmada toplam 1.115 hasta incelenmiştir. Hastaların hepsi ileri akciğer kanseri veya ileri kolorektal kanser hastaları için altı klinik çalışmadan birine eş zamanlı olarak katılıyordu. Hastalar, 15 Haziran 1990'dan 17 Temmuz 1993'e kadar tedavi denemelerine alındı. Bu hastaların 34'ü tedavi protokolleri için uygun bulunmadığı için analizlerden çıkarıldı. Kalan 1.077 hastadan 1.008'i öldü. Yaşayan hastaların takip süresi 18 aydan fazladır, fakat dört hasta en az 2 yıl izlenmiştir. 515 ileri kolorektal kanser hastası, 357 küçük hücreli akciğer kanseri hastası ve 205 küçük hücreli olmayan akciğer kanseri hastası üzerinde birçok analizler yapılmıştır. Buradaki çalışmamızda akciğer kanseri hastalarından rasgele seçilen 228 hastadan oluşan veri seti üzerinde çalışma yapılmıştır.[14] Tablo 4.6'da akciğer kanseri veri seti verilmiştir. Sütunlar:

- 1.sütun: Kayıt yeri
- 2.sütun: Sağkalım süresi
- 3.sütun: Sansürleme 0=sansürlü, 1=sansürsüz
- 4.sütun: Yaş
- 5.sütun: Cinsiyet 0=erkek 1=kadın
- 6.sütun: ECOG performans skoru, doktor tarafından değerlendirme: 0,1,2,3
- 7.sütun: Karnofsky performans skoru, doktor tarafından değerlendirme: 100, 90, ..., 30
- 8.sütun: Karnofsky performans skoru, hasta tarafından değerlendirme (self)
- 9.sütun: Günlük yemekte tüketilen kalori
- 10.sütun: Son 30 gündeki kilo kaybı (negatif sayı = kilo alma)

Tablo 4.6. Akciğer kanseri hastalarının verisi

Kayıt yapan kurum	survival	status	age	sex	Karnof			Günlük kalori	Kilo kaybı
					ecog perf	sky perf	karnofsky hasta perf.		
3	306	1	74	0	1	90	100	1175	.
3	455	1	68	0	0	90	90	1225	15
3	1010	0	56	0	0	90	90		15
5	210	1	57	0	1	90	60	1150	11
1	883	1	60	0	0	100	90		0
12	1022	0	74	0	1	50	80	513	0
7	310	1	68	1	2	70	60	384	10
11	361	1	71	1	2	60	80	538	1
1	218	1	53	0	1	70	80	825	16
7	166	1	61	0	2	70	70	271	34

6	170	1	57	0	1	80	80	1025	27
16	654	1	68	1	2	70	70		23
11	728	1	68	1	1	90	90		5
21	71	1	60	0	.	60	70	1225	32
12	567	1	57	0	1	80	70	2600	60
1	144	1	67	0	1	80	90		15
22	613	1	70	0	1	90	100	1150	-5
16	707	1	63	0	2	50	70	1025	22
1	61	1	56	1	2	60	60	238	10
21	88	1	57	0	1	90	80	1175	.
11	301	1	67	0	1	80	80	1025	17
6	81	1	49	1	0	100	70	1175	-8
11	624	1	50	0	1	70	80		16
15	371	1	58	0	0	90	100	975	13
12	394	1	72	0	0	90	80		0
12	520	1	70	1	1	90	80	825	6
4	574	1	60	0	0	100	100	1025	-13
13	118	1	70	0	3	60	70	1075	20
13	390	1	53	0	1	80	70	875	-7
1	12	1	74	0	2	70	50	305	20
12	473	1	69	1	1	90	90	1025	-1
1	26	1	73	0	2	60	70	388	20
7	533	1	48	0	2	60	80		-11
16	107	1	60	1	2	50	60	925	-15
12	53	1	61	0	2	70	100	1075	10
1	122	1	62	1	2	50	50	1025	.
22	814	1	65	0	2	70	60	513	28
15	965	0	66	1	1	70	90	875	4
1	93	1	74	0	2	50	40	1225	24
1	731	1	64	1	1	80	100	1175	15
5	460	1	70	0	1	80	60	975	10
11	153	1	73	1	2	60	70	1075	11
10	433	1	59	1	0	90	90	363	27
12	145	1	60	1	2	70	60		
7	583	1	68	0	1	60	70	1025	7
7	95	1	76	1	2	60	60	625	-24
1	303	1	74	0	0	90	70	463	30
3	519	1	63	0	1	80	70	1025	10
13	643	1	74	0	0	90	90	1425	2
22	765	1	50	1	1	90	100	1175	4
3	735	1	72	1	1	90	90		9
12	189	1	63	0	0	80	70		0
21	53	1	68	0	0	90	100	1025	0
1	246	1	58	0	0	100	90	1175	7
6	689	1	59	0	1	90	80	1300	15
1	65	1	62	0	0	90	80	1300	15
5	5	1	65	1	0	100	80	338	5
22	132	1	57	0	2	70	60		18

3	687	1	58	1	1	80	80	1225	10
1	345	1	64	1	1	90	80	1075	-3
22	444	1	75	1	2	70	70	438	8
12	223	1	48	0	1	90	80	1300	68
21	175	1	73	0	1	80	100	1025	
11	60	1	65	1	1	90	80	1025	0
3	163	1	69	0	1	80	60	1125	0
3	65	1	68	0	2	70	50	825	8
16	208	1	67	1	2	70		538	2
5	821	0	64	1	0	90	70	1025	3
22	428	1	68	0	0	100	80	1039	0
6	230	1	67	0	1	80	100	488	23
13	840	0	63	0	0	90	90	1175	-1
3	305	1	48	1	1	80	90	538	29
5	11	1	74	0	2	70	100	1175	0
2	132	1	40	0	1	80	80		3
21	226	1	53	1	1	90	80	825	3
12	426	1	71	1	1	90	90	1075	19
1	705	1	51	1	0	100	80	1300	0
6	363	1	56	1	1	80	70	1225	-2
3	11	1	81	0	0	90		731	15
1	176	1	73	0	0	90	70	169	30
4	791	1	59	0	0	100	80	768	5
13	95	1	55	0	1	70	90	1500	15
11	196	0	42	0	1	80	80	1425	8
21	167	1	44	1	1	80	90	588	-1
16	806	0	44	0	1	80	80	1025	1
6	284	1	71	0	1	80	90	1100	14
22	641	1	62	1	1	80	80	1150	1
21	147	1	61	0	0	100	90	1175	4
13	740	0	44	1	1	90	80	588	39
1	163	1	72	0	2	70	70	910	2
11	655	1	63	0	0	100	90	975	-1
22	239	1	70	0	1	80	100		23
5	88	1	66	0	1	90	80	875	8
10	245	1	57	1	1	80	60	280	14
1	588	0	69	1	0	100	90		13
12	30	1	72	0	2	80	60	288	7
3	179	1	69	0	1	80	80		25
12	310	1	71	0	1	90	100		0
11	477	1	64	0	1	90	100	910	0
3	166	1	70	1	0	90	70		10
1	559	0	58	1	0	100	100	710	15
6	450	1	69	1	1	80	90	1175	3
13	364	1	56	0	1	70	80		4
6	107	1	63	0	1	90	70		0
13	177	1	59	0	2	50	.		32
12	156	1	66	0	1	80	90	875	14

26	529	0	54	1	1	80	100	975	-3
1	11	1	67	0	1	90	90	925	
21	429	1	55	0	1	100	80	975	5
3	351	1	75	1	2	60	50	925	11
13	15	1	69	0	0	90	70	575	10
1	181	1	44	0	1	80	90	1175	5
10	283	1	80	0	1	80	100	1030	6
3	201	1	75	1	0	90	100		1
6	524	1	54	1	1	80	100		15
1	13	1	76	0	2	70	70	413	20
3	212	1	49	0	2	70	60	675	20
1	524	1	68	0	2	60	70	1300	30
16	288	1	66	0	2	70	60	613	24
15	363	1	80	0	1	80	90	346	11
22	442	1	75	0	0	90	90		0
26	199	1	60	1	2	70	80	675	10
3	550	1	69	1	1	70	80	910	0
11	54	1	72	0	2	60	60	768	-3
1	558	1	70	0	0	90	90	1025	17
22	207	1	66	0	1	80	80	925	20
7	92	1	50	0	1	80	60	1075	13
12	60	1	64	0	1	80	90	993	0
16	551	0	77	1	2	80	60	750	28
12	543	0	48	1	0	90	60		4
4	293	1	59	1	1	80	80	925	52
16	202	1	53	0	1	80	80		20
6	353	1	47	0	0	100	90	1225	5
13	511	0	55	1	1	80	70		49
1	267	1	67	0	0	90	70	313	6
22	511	0	74	1	2	60	40	96	37
12	371	1	58	1	1	80	70		0
13	387	1	56	0	2	80	60	1075	.
1	457	1	54	0	1	90	90	975	-5
5	337	1	56	0	0	100	100	1500	15
21	201	1	73	1	2	70	60	1225	-16
3	404	0	74	0	1	80	70	413	38
26	222	1	76	0	2	70	70	1500	8
1	62	1	65	1	1	80	90	1075	0
11	458	0	57	0	1	80	100	513	30
26	356	0	53	1	1	90	90		2
16	353	1	71	0	0	100	80	775	2
16	163	1	54	0	1	90	80	1225	13
12	31	1	82	0	0	100	90	413	27
13	340	1	59	1	0	100	90		0
13	229	1	70	0	1	70	60	1175	-2
22	444	0	60	0	0	90	100		7
5	315	0	62	1	0	90	90		0
16	182	1	53	1	1	80	60		4

32	156	1	55	0	2	70	30	1025	10
.	329	1	69	0	2	70	80	713	20
26	364	0	68	1	1	90	90		7
4	291	1	62	0	2	70	60	475	27
12	179	1	63	0	1	80	70	538	-2
1	376	0	56	1	1	80	90	825	17
32	384	0	62	1	0	90	90	588	8
10	268	1	44	1	1	90	100	2450	2
11	292	0	69	0	2	60	70	2450	36
6	142	1	63	0	1	90	80	875	2
7	413	0	64	0	1	80	70	413	16
16	266	0	57	1	0	90	90	1075	3
11	194	1	60	1	1	80	60		33
21	320	1	46	0	0	100	100	860	4
6	181	1	61	0	1	90	90	730	0
12	285	1	65	0	0	100	90	1025	0
13	301	0	61	0	1	90	100	825	2
2	348	1	58	1	0	90	80	1225	10
2	197	1	56	0	1	90	60	768	37
16	382	0	43	1	0	100	90	338	6
1	303	0	53	0	1	90	80	1225	12
13	296	0	59	1	1	80	100	1025	0
1	180	1	56	0	2	60	80	1225	-2
13	186	1	55	1	1	80	70		
1	145	1	53	1	1	80	90	588	13
7	269	0	74	1	0	100	100	588	0
13	300	0	60	0	0	100	100	975	5
1	284	0	39	0	0	100	90	1225	-5
16	350	1	66	1	0	90	100	1025	
32	272	0	65	1	1	80	90		-1
12	292	0	51	1	0	90	80	1225	0
12	332	0	45	1	0	90	100	975	5
2	285	1	72	1	2	70	90	463	20
3	259	0	58	0	0	90	80	1300	8
15	110	1	64	0	1	80	60	1025	12
22	286	1	53	0	0	90	90	1225	8
16	270	1	72	0	1	80	90	488	14
16	81	1	52	0	2	60	70	1075	
12	131	1	50	0	1	90	80	513	
1	225	0	64	0	1	90	80	825	33
22	269	1	71	0	1	90	90	1300	-2
12	225	0	70	0	0	100	100	1175	6
32	243	0	63	1	1	80	90	825	0
21	279	0	64	0	1	90	90		4
1	276	0	52	1	0	100	80	975	0
32	135	1	60	0	1	90	70	1275	0
15	79	1	64	1	1	90	90	488	37
22	59	1	73	0	1	60	60	2200	5

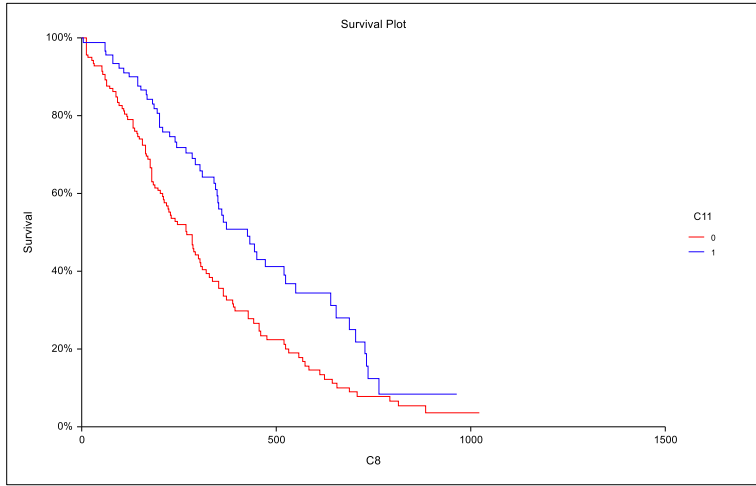
32	240	0	63	1	0	90	100	1025	0
3	202	0	50	1	0	100	100	635	1
26	235	0	63	1	0	100	90	413	0
33	105	1	62	0	2		70		
5	224	0	55	1	0	80	90		23
13	239	1	50	1	2	60	60	1025	-3
21	237	0	69	0	1	80	70		
33	173	0	59	1	1	90	80		10
1	252	0	60	1	0	100	90	488	-2
6	221	0	67	0	1	80	70	413	23
15	185	0	69	0	1	90	70	1075	0
11	92	0	64	1	2	70	100		31
11	13	1	65	0	1	80	90		10
11	222	0	65	0	1	90	70	1025	18
13	192	0	41	1	1	90	80		-10
21	183	1	76	0	2	80	60	825	7
11	211	0	70	1	2	70	30	131	3
2	175	0	57	1	0	80	80	725	11
22	197	0	67	0	1	80	90	1500	2
11	203	0	71	1	1	80	90	1025	0
1	116	1	76	0	1	80	80		0
1	188	0	77	0	1	80	60		3
13	191	0	39	0	0	90	90	2350	-5
32	105	0	75	1	2	60	70	1025	5
6	174	0	66	0	1	90	100	1075	1
22	177	0	58	1	1	80	90	1060	0

Akciğer kanseri hastaların yaş, cinsiyet ve Karnofsky performans değeri bağımsız değişkenlerine göre hesaplama yapıldığında aşağıdaki sonuçlar elde edilmiştir.

Tablo 4.7. Akciğer kanseri verinin R sonuçları

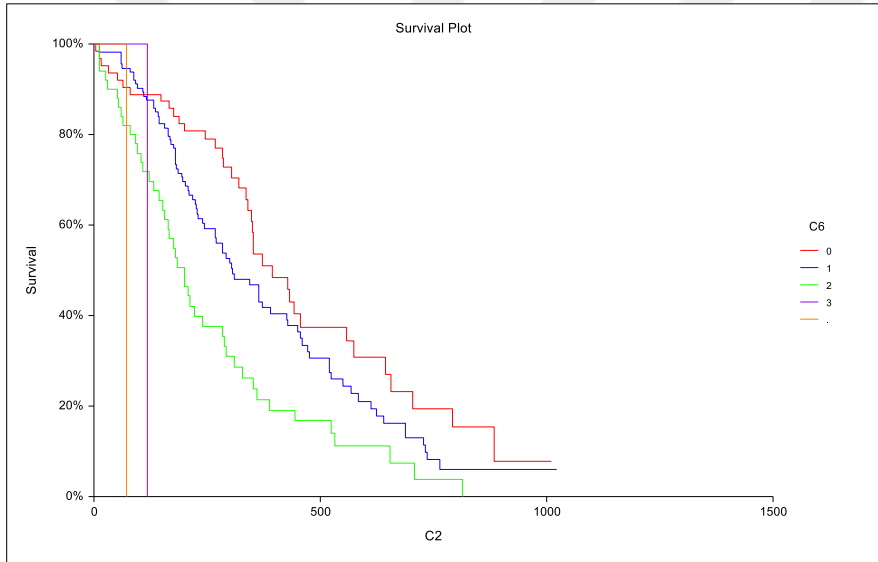
	coef	exp(coef)	se(coef)	z	p
Yaş	0.013	1.013	0.009	1.36	0.1723
Cinsiyet	-0.492	0.611	0.168	-2.93	0.0034
Log(karnofskyperf)	-0.882	0.413	0.435	-2.03	0.0424

Tabloya baktığımızda cinsiyet değişkeni için regresyon katsayısı -0.492 olarak hesaplanmıştır. Burada katsayının işaretine baktığımızda negatif olması 2.grubun ölüm riskinin, 1.gruba göre daha düşük olduğunu göstermektedir. Bu veri setine göre kadın hastaların, erkek hastalara göre ölüm riski daha düşüktür. Tehlike oranı değerine bakıldığında 0.61 veya %61 kadınların erkeklere göre ölüm riski daha azdır.



Şekil 4.3 Cinsiyete göre sağkalım olasılıkları

Şekil 4.3'te grafikte akciğer kanseri hastalarda erkeklerin kadınlara oranla yaşama olasılığı daha düşük olduğunu söyleyebiliriz.



Şekil 4.4. ECOG performansına göre sağkalım olasılıkları

Şekil 4.4 kanser hastaları için doktorların verdiği ECOG performans değerlerine göre sağkalım olasılıklarına baktığımızda performans değeri arttıkça yaşama olasılıkları daha düşük olduğu görülmektedir. Sağkalım süresi 500 ve üzeri hastalarda 0 ve 1 performans değerindeki hastalarda yaşama olasılığı varken 2 performans değerindeki hastalarda belli bir süreden sonra yaşama olasılığı 0(sıfır)'dır.

Tablo 4.8. Beta tahminleri ve güven aralıkları

	MLE	Alt Sınır	Üst Sınır
Yaş	0.012	-0.005	0.032
Cinsiyet	-0.492	-0.828	-0.169
Log(karnofskyperf)	-0.882	-1.717	-0.012

Yukarıdaki tabloya göre, yaş bağımsız değişkeni için maksimum kısmi olabilirlik tahmin edicisi(MLE) 0.012 ve güven aralığı (-0.005,0.032), cinsiyet değişkeni için MLE, -0.492 ve güven aralığı (-0.828,-0.169) ve hastalık türü değişkeni için MLE değeri -0.882 ve güven aralığı (-1.717,-0.169) olarak bulunmuştur.



### 4.3 Uygulama 3(Böbrek Kateter Verisi)

Tablo 4.9'da verilen veriler, portatif diyaliz ekipmanı kullanan böbrek hastaları için kateterin yerleştirildiği noktada enfeksiyonun tekrarlanma zamanlarını göstermektedir. Her hasta için iki nüksetme süresi verilir, tablonun her sırası bir hastaya karşılık gelir. Her hastanın tam olarak 2 gözlemi vardır.

Veri seti, sağkalım verileri için rasgele etki ("kırılganlık") modellerini göstermek için bazı yazarlar tarafından kullanılmıştır. Bununla birlikte, rastgele etkinin sıfır olmayan herhangi bir tahmini neredeyse tamamen denek 21'in aykırı değerinden kaynaklanmaktadır.[12]

Sütunlar:

- 1.sütun: Hasta numarası
- 2.sütun: Hastalığın nüksetme zamanı
- 3.sütun: Sansürleme. 0 = sansürlü, 1 = sansürsüz
- 4.sütun: Yaş
- 5.sütun: Cinsiyet (1 = erkek, 2 = kadın)
- 6.sütun: Hastalık türü
  - 0 = Glomerulo Nefrit
  - 1 = Akut Nefrit
  - 2 = Polikistik Böbrek Hastalığı
  - 3 = Diğer
- 7.sütun: Kırılganlık tahmini(estimate of the frailty)

Tablo 4.9. Böbrek kateter verisi

Hasta numarası	Nüksetme zamanları	Sansürleme	Yaş	Cinsiyet	Hastalık türü	Kırılganlık Tahmini
1	8,16	1,1	28	1	3	2.3
2	23,13	1,0	48	2	0	1.9
3	22,28	1,1	32	1	3	1.2
4	447,318	1,1	31-32	2	3	0.5
5	30,12	1,1	10	1	3	1.5
6	24,245	1,1	16-17	2	3	1.1
7	7,9	1,1	51	1	0	3.0
8	511,30	1,1	55-56	2	0	0.5
9	53,196	1,1	69	2	1	0.7
10	15,154	1,1	51-52	1	0	0.4
11	7,333	1,1	44	2	1	0.6
12	141,8	1,0	34	2	3	1.2
13	96,38	1,1	35	2	1	1.4

14	149,70	0,0	42	2	1	0.4
15	536,25	1,0	17	2	3	0.4
16	17,4	1,0	60	1	1	1.1
17	185,177	1,1	60	2	3	0.8
18	292,114	1,1	43-44	2	3	0.8
19	22,159	0,0	53	2	0	0.5
20	15,108	1,0	44	2	3	1.3
21	152,562	1,1	46-47	1	2	0.2
22	402,24	1,0	30	2	3	0.6
23	13,66	1,1	62-63	2	1	1.7
24	39,46	1,0	42-43	2	1	1.0
25	12,40	1,1	43	1	1	0.7
26	113,201	0,1	57-58	2	1	0.5
27	132,156	1,1	10	2	0	1.1
28	34,30	1,1	52	2	1	1.8
29	2,25	1,1	53	1	0	1.5
30	130,26	1,1	54	2	0	1.5
31	27,58	1,1	56	2	1	1.7
32	5,43	0,1	50-51	2	1	1.3
33	152,30	1,1	57	2	2	2.9
34	190,5	1,0	44-45	2	0	0.7
35	119,8	1,1	22	2	3	2.2
36	54,16	0,0	42	2	3	0.7
37	6,78	0,1	52	2	2	2.1
38	63,8	1,0	60	1	2	1.2

Böbrek hastalarının yaş, cinsiyet ve hastalık türü olmak üzere 3 bağımsız değişkene göre hesaplama yapıldığında aşağıdaki sonuçlar elde edilmiştir.

Tablo 4.10. Böbrek kateter verisinin R sonuçları

	coef	exp(coef)	se	z	p
Yaş	-0.003	0.997	0.010	-0.27	0.785
Cinsiyet	-0.842	0.431	0.300	-2.81	0.005
Hastalık türü	-0.155	0.857	0.124	-1.25	0.212

Tabloya baktığımızda cinsiyet değişkeni için regresyon katsayısı -0.842 olarak hesaplanmıştır. Burada katsayının işaretine baktığımızda negatif olması 2.grubun ölüm riskinin, 1.gruba göre daha düşük olduğunu göstermektedir. Bu veri setine göre kadın hastaların, erkek hastalara göre ölüm riski daha düşüktür. Tehlike oranı değerine bakıldığında 0.57 veya %57 erkekler kadınlara göre ölüm riski daha fazladır.

Tablo 4.11. Böbrek Kateter verisinin R sonuçları 2

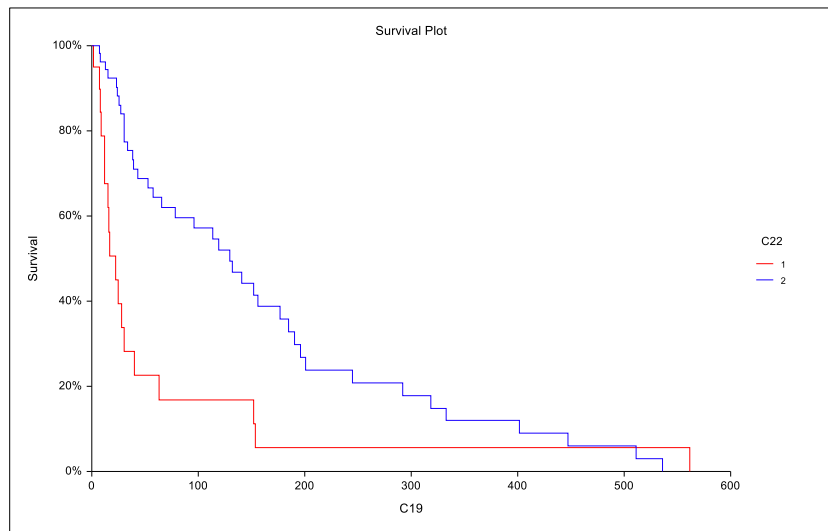
Olabilirlik oran testi (LR)	8.64
Serbestlik derecesi (df)	3
n	76
Olay sayısı	58
p	0.03

Tablo 4.11'e baktığımızda olabilirlik oran testi değeri 3 serbestlik dereceli ki-kare tablo değeri 7.81'den büyük ve anlamlılık seviye değeri 0.05'ten küçük olduğundan gruplar arasında ilişki olmadığını söyleyebiliriz. Her bir bağımsız değişkenin maksimum kısmi olabilirlik tahmin edicisi için güven aralığı aşağıdaki tablodaki gibidir.

Tablo 4.12  $\beta$  katsayıları için güven aralığı

	MLE	Alt Sınır	Üst Sınır
Yaş	-0.003	-0.021	0.017
Cinsiyet	-0.842	-1.412	-0.230
Hastalık türü	-0.155	-0.396	0.092

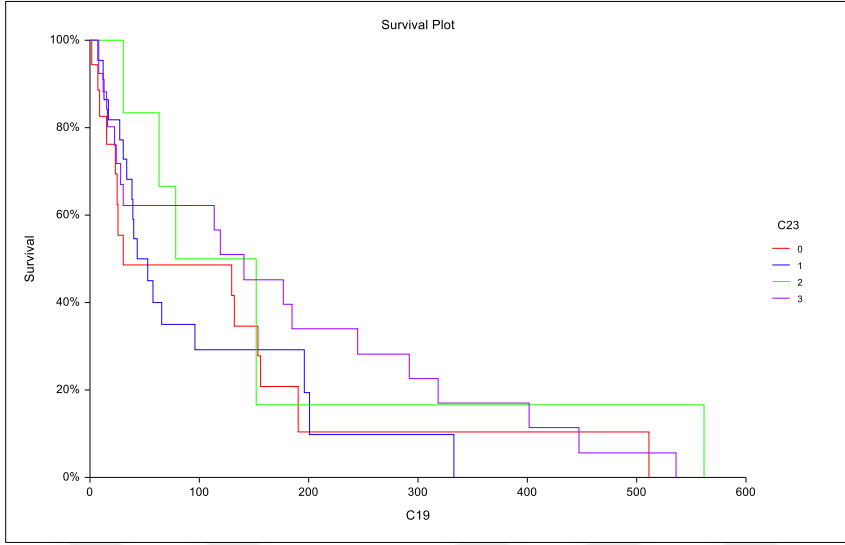
Yukarıdaki tabloya göre, yaş bağımsız değişkeni için maksimum kısmi olabilirlik tahmin edicisi(MLE) -0.003 ve güven aralığı (-0.021,0.017), cinsiyet değişkeni için MLE, -0.842 ve güven aralığı (-1.412,-0.230) ve hastalık türü değişkeni için MLE değeri -0.155 ve güven aralığı (-0.396,0.092) olarak bulunmuştur.



Şekil 4.5 Cinsiyete göre sağkalım olasılıkları

Şekil 4.5'teki grafikte böbrek hastalarında kateterin yerleştirildiği yerde enfeksiyonun tekrarlanma olasılığı erkeklerde kadınlara oranla daha düşük olduğunu söyleyebiliriz.

Sağkalım süresi veya bu veri için tekrarlanma zamanı 400'den sonrasında kadın ve erkeklerde görülme olasılıkları birbirine yakındır.



Şekil 4.6 Hastalık türüne göre sağkalım olasılığı

Şekil 4.6'da grafiğe baktığımızda böbrek hastalarında kateterde enfeksiyonun nüksetme süresi 400 ve üzeri olan hastalarda akut nefrit hastalarının olma olasılığı sıfırdır. Polikistik böbrek hastalarında enfeksiyonun nüksetme olasılığı diğerlerine göre biraz daha fazladır. Glomüre nefrit hastalığı olan hastalarda sağkalım süresi 0-100 arasında nüksetme olasılığı daha yüksektir.

Böbrek hastaları için hastalık türü bağımsız değişkenine göre sonuçlar aşağıdaki gibidir.

$T \leq 100$  için maksimum kısmi olabilirlik tahmin edicisi  $\hat{\beta}$  ve sıfır hipotezinin testi için iki deneysel olabilirlik(EL) değeri ve olabilirlik oranı:

Maksimum kısmi olabilirlik tahmin edicisi  $\hat{\beta}$  değeri -0.143 olarak hesaplanmıştır. Hipotezimiz,

$$H_0: \beta = -0.2, \mu = \Lambda_0(100) = 0.64 .$$

Tablo 4.13.  $\beta = -0.2$  ve  $\lambda = 0$  için EL

mu	0.883
logPlik	-187.1687
<b>logEmpLik1</b>	<b>-245.1687</b>

Tablo 4.14.  $\beta = \hat{\beta}$  ve  $\lambda = 0$  için EL

mu	0.940
logPlik	-187.0944
<b>logEmpLik2</b>	<b>-245.0944</b>

Olabilirlik oranı:

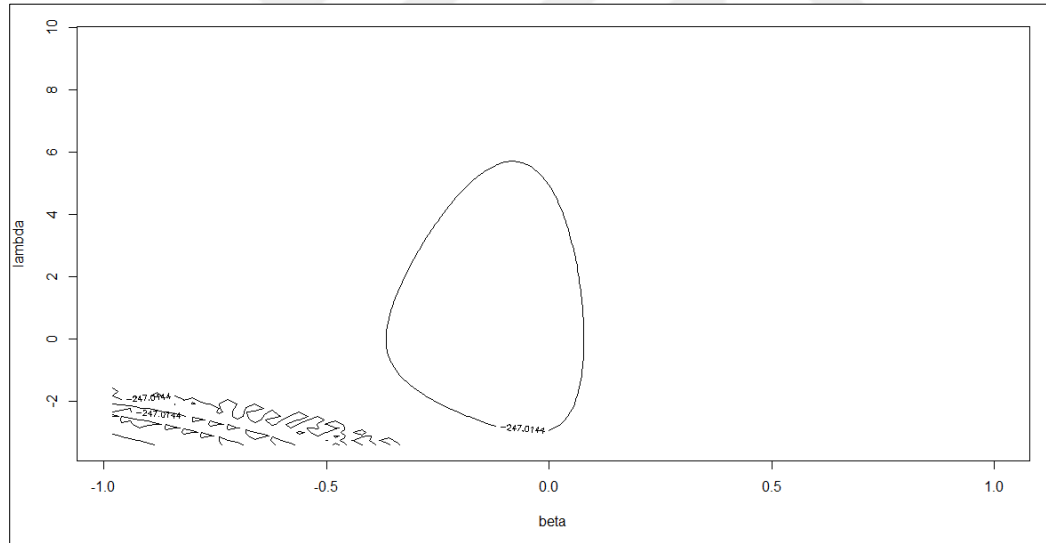
$$-2(\log\text{Emplik1} - \log\text{Emplik2}) = -2(-245.1687 - (-245.0944)) = 0.2508.$$

Anlamlılık seviye değeri  $p = 0.05$  'den büyük ve 2 serbestlik dereceli  $\chi^2$  değeri 3.84'ten küçüktür. İki duruma göre bakıldığında hipotezimiz reddedilemez.

Tablo 4.15.  $\beta = -0.2$  ve  $\lambda = 14.185$  için EL

mu	<b>0.64002</b>
logPlik	-187.1687
logEmpLik	-246.6953

$\log\text{Lik} = \log\text{Lik}(\beta = -0.2, \lambda = 14.185) = -246.695$ .  $T \leq 100$  iken sağkalım olasılığı 0.64'e eşit olarak bulunmuştur. Güven aralığı aşağıdaki gibidir.



Şekil 4.7. Güven aralığı

## 5. SİMÜLASYON

Simülasyon çalışmamız iki dağılım üzerinde bir boyutlu ortak değişken ile yapılmıştır. İlk olarak modelimiz, genel Cox modeli

$$\lambda(t) = \lambda_0(t). \exp(\beta_0 Z)$$

için  $\beta_0 = 1$ ,  $\lambda_0(t)$  ölçü parametresi 1, şekil parametresi 2 olan Weibull dağılımı olacak şekilde oluşturuldu. Ortak değişken  $Z_i$ ,  $U(0,1)$  düzgün dağılım ve sansürleme zamanı  $c_i$ , sansürleme oranına(CR) bağlı olarak seçilen  $k$  için  $U(0,k)$  dağılımı gösterebilir. Çalışmada CR %10 ve %40 olarak alınmıştır. Örneklem sayısı ( $n$ ) 50 ve 100 seçildi. Simülasyon verisi 1000 kere tekrarlandı. Simülasyon verisine göre düzeltilmiş deneysel olasılıklar(AEL) ve deneysel olasılıklar(EL) yöntemleri için tahmin edilen kapsama olasılıkları karşılaştırıldı.  $(1 - \alpha)$  ağırlıklı güven seviyesi, %95 ve %99'dur. Güven bölgesi  $T = 5$  'e göre oluşturuldu. Gözlemlenen başarısızlık süresi veya sansürlü zaman  $T$  'ye eşit ya da daha az olduğu zamanlardır. Aşağıdaki tabloda Weibull dağılımı için farklı kombinasyonlarla kapsama olasılıkları verilmiştir. Simülasyon için adımlar:

1. Ortak değişken düzgün dağılım gösterebilir. Yani  $Z_i(t) \sim U(0,1)$ 'dir.
2. Weibull dağılımına sahip herhangi  $t$  değişkeni için,  $t \sim \text{Weibull}(\alpha, \lambda)$  olarak tanımlayalım. Temel tehlike fonksiyonu aşağıdaki tanımlayabiliriz.

$$\lambda_0(t) = \alpha \lambda^\alpha t^{\alpha-1}$$

$\alpha = 2$  ve  $\lambda = 1$  için denklemde yazılırsa  $\lambda_0(t) = 2t$  şeklinde bulunur. Cox modelimiz,

$$\lambda(t|Z) = 2te^{Z(t)}$$

olur. Weibull dağılımı için sağkalım fonksiyonu,

$$S(t) = \exp[-(\lambda t)^\alpha] = \exp(-4t^4 e^{2Z})$$

şeklinindedir. Sağkalım süresi  $T$  için Cox model için tehlike fonksiyonunu bildiğimiz için sağkalım fonksiyonu  $S(t)$  ve dağılım fonksiyonu  $F(t)$  aşağıdaki gibi ifade edebiliriz.

$$F(t) = 1 - S(t) = 1 - \exp(-4t^4 e^{2Z})$$

$U, U(0,1)$  düzgün dağılım olarak alınsın. Önceki adımlarda elde edilen ortak değişken  $Z$  ve  $U$  kullanılarak sağkalım süresi  $T$  ters fonksiyon yöntemi kullanıldığında,

$$T = [(-e^{-2Z}) \ln(1 - U)/4]^{1/4}$$

olarak tanımlayalım.

3.  $k$  sansürleme oranı(CR) olmak üzere, sansürleme süresi  $C_i, U(0, k)$  düzgün dağılımı gösterebilir. CR bu çalışmada %10 ve %40 olarak alınacaktır.

Tablo 5.1. Weibull dağılımlı simülasyon için kapsama olasılıkları

CR	n	%95		%99	
		AEL	EL	AEL	EL
%10	50	<b>0.97</b>	<b>0.95</b>	<b>0.98</b>	<b>0.96</b>
	100	<b>0.95</b>	<b>0.63</b>	<b>0.97</b>	<b>0.65</b>
%40	50	<b>0.96</b>	<b>0.95</b>	<b>0.98</b>	<b>0.96</b>
	100	<b>0.95</b>	<b>0.66</b>	<b>0.98</b>	<b>0.69</b>

Yukarıdaki tabloya göre,  $\beta_0 = 1$  ve Weibull dağılım gösteren temel tehlike fonksiyonu için Cox regresyon modelimiz için deneysel olabilirlik yöntemi ve düzeltilmiş deneysel olabilirlik yöntemleri kıyaslandığında düzeltilmiş deneysel olabilirlik yöntemi için kapsama olasılığı daha yüksek olarak hesaplanmıştır. Tabloya bakıldığında deneysel olabilirlik yöntemi için denek sayısı arttıkça kapsama olasılığı tüm durumlarda düşmüştür.  $(1 - \alpha)$  ağırlıklı güven seviyesi küçüldükçe her iki yöntemde de kapsama olasılıklarında artış olmuştur.

İkinci olarak modelimizi, genel Cox modeli

$$\lambda(t) = \lambda_0(t). \exp(\beta_0 Z)$$

için  $\beta_0 = 1$ ,  $\lambda_0(t)$   $\alpha = 1$  ve  $\beta = 2$  olan Log-Lojistik dağılımı olarak dağılım gösterebilir. İlk modelde olduğu gibi burada ortak değişken  $Z_i, U(0,1)$  düzgün dağılım ve sansürleme zamanı  $c_i$ , sansürleme oranına(CR) bağlı olarak seçilen  $k$  için  $U(0, k)$  dağılımı gösterebilir ve CR %10 ve %40 olarak alınmıştır. Örneklem sayısı (n) 50 ve 100 seçildi. Simülasyon verisi 1000 kere tekrarlandı. Simülasyon verisine göre düzeltilmiş deneysel olabilirlik(AEL) ve deneysel olabilirlik(EL) yöntemleri için tahmin edilen kapsama olasılıkları karşılaştırıldı.  $(1 - \alpha)$  ağırlıklı güven seviyesi, %95 ve %99'dur.

Güven bölgesi  $T = 5$ 'e göre oluşturuldu. Aşağıdaki tabloda Log-Lojistik dağılımı için farklı kombinasyonlarla kapsama olasılıkları verilmiştir. Simülasyon için adımlar:

4. Ortak değişken düzgün dağılım gösterebilir. Yani  $Z_i(t) \sim U(0,1)$ 'dir.
5. Weibull dağılımına sahip herhangi  $t$  değişkeni için,  $t \sim \text{Logist}(\alpha, \lambda)$  olarak tanımlayalım. Temel tehlike fonksiyonu aşağıdaki tanımlayabiliriz.

$$\lambda_0(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1 + (t/\alpha)^\beta}$$

$\alpha = 1$  ve  $\beta = 2$  için denklemde yazılırsa  $\lambda_0(t) = \frac{2t}{1 + t^2}$  şeklinde bulunur. Cox modelimiz,

$$\lambda(t|Z) = \frac{2t}{1 + t^2} e^{Z(t)}$$

olur. Log-Lojistik dağılımı için sağkalım fonksiyonu,

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

eşitliği ve

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta]^2}$$

kullanılarak,

$$S(t) = e^{-Z}/(1 + t^2)$$

şeklinde bulunur. Sağkalım süresi  $T$  için Cox model için tehlike fonksiyonunu bildiğimiz için sağkalım fonksiyonu  $S(t)$  ve dağılım fonksiyonu  $F(t)$  aşağıdaki gibi ifade edebiliriz.

$$F(t) = 1 - S(t) = 1 - e^{-Z}/(1 + t^2)$$

$U, U(0,1)$  düzgün dağılım olarak alınsın. Önceki adımlarda elde edilen ortak değişken  $Z$  ve  $U$  kullanılarak sağkalım süresi  $T$  ters fonksiyon yöntemi kullanıldığında,

$$T = [e^{-Z}(1 - U)^{-1} - 1]^{1/2}$$

olarak tanımlayalım.

6.  $k$  sansürleme oranı(CR) olmak üzere, sansürleme süresi  $C_i$ ,  $U(0, k)$  düzgün dağılımı gösterebilir. CR bu çalışmada %10 ve %40 olarak alınacaktır.

Tablo 5.2. Log-Lojistik dağılımlı simülasyon için kapsama olasılıkları

CR	n	%95		%99	
		AEL	EL	AEL	EL
%10	50	<b>0.98</b>	<b>0.95</b>	<b>0.99</b>	<b>0.97</b>
	100	<b>0.87</b>	<b>0.69</b>	<b>0.97</b>	<b>0.92</b>
%40	50	<b>0.99</b>	<b>0.97</b>	<b>0.99</b>	<b>0.98</b>
	100	<b>0.90</b>	<b>0.70</b>	<b>0.98</b>	<b>0.92</b>

Yukarıdaki tabloya göre, deneysel olabilirlik yöntemi ve düzeltilmiş deneysel olabilirlik yöntemleri  $\beta_0 = 1$  ve Log-Lojistik dağılım gösteren temel tehlike fonksiyonu için Cox regresyon modelimize göre düzeltilmiş deneysel olabilirlik yönteminin kapsama olasılığı değerleri daha yüksektir. Tabloya göre, denek sayısı arttıkça diğer modelimizde olduğu gibi kapsama olasılıkları tüm durumlarda düşmüş ve  $(1 - \alpha)$  ağırlıklı güven seviyesi azaldıkça her iki yöntemde de kapsama olasılıklarında artış olmuştur.

## 6. SONUÇ

Dördüncü bölümde yaptığımız üç sağdan sansürlü gerçek veri üzerinde Cox regresyon modeline ait sağkalım fonksiyonlarının parametrelerinin deneysel olabilirlik yöntemini ile elde edilmesini işledik. Üç uygulama görüldüğü gibi deneysel olabilirlik yöntemine göre Cox yöntemi (yarı-parametrik olabilirlik yöntemi) yetersiz kaldığını söyleyebiliriz. Başka bir deyişle Cox model yöntemi, deneysel olabilirlik yöntemi için işlem basamaklarından biri olduğunu söyleyebiliriz. Cox yöntemi bize beta katsayısının maksimum kısmi olabilirlik değeri, güven aralığını ve tehlike oranını vermektedir. Fakat Cox modelindeki temel tehlike fonksiyonunun değerini uygulamada da gördüğümüz gibi deneysel olabilirlik yöntemi ile elde edilmiştir. Ayrıca deneysel olabilirlik yöntemi ile parametrelerin güven aralıkları ve model için güven bandı oluşturulmuştur.

Beşinci bölümde iki farklı Weibull ve Log-Lojistik dağılım gösteren temel tehlike fonksiyonlu Cox regresyon modelimiz için deneysel olabilirlik yöntemi ve düzeltilmiş deneysel olabilirlik yöntemlerinin %95 ve %99 güven seviyelerindeki kapsama olasılıkları hesaplanmış ve bulduğumuz değerlere göre düzeltilmiş deneysel olabilirlik yönteminin daha iyi sonuç verdiği görülmüştür. Diğer bir taraftan, güven seviyesindeki artışta her iki yöntemde de kapsama olasılıkları daha yüksek olduğu ve küçük örneklerde daha iyi sonuç verdiği görülmüştür. Örneklem sayısı arttıkça deneysel olabilirlik yöntemi ile düzeltilmiş deneysel olabilirlik yöntemi için bulduğumuz kapsama olasılıkları değerleri arasındaki fark arttığı gözlemlenmiştir.

## KAYNAKLAR/REFERENCES

- [1] Camilleri L. (2019), History of Survival Analysis, Times Malta, "<https://timesofmalta.com/articles/view/history-of-survival-analysis.705424>"
- [2] Gastrointestinal Tumor Study Group (1982), A Comparison of Combination Chemotherapy and Combined Modality Therapy for Locally Advanced Gastric Carcinoma, *Cancer* 49:1777-1777
- [3] Kalbfleisch J.D., Prentice R.L. (2002), *The Statistical Analysis of Failure Time Data*, Second Edition, Wiley-Interscience, A John Wiley & Sons, Inc., Publication, U.S.A.
- [4] Kitamura Y. (2008), *The New Palgrave Dictionary of Economics (Empirical Likelihood)*, Second Edition, Steven N. Durlauf and Lawrence E. Blume.
- [5] Klein J.P., Moeschberger M.L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, Second Edition, Springer-Verlag New York, Inc., Series Editors: Dietz K., Gail M., Krickeberg K., Samet J., Tsiatis A., New York, U.S.A.
- [6] Kleinbaum D.G., Klein M. (2012), *Survival Analysis: A Self-Learning Text*, Third Edition, *Statistics for Biology and Health*, Springer Science+Business Media, New York.
- [7] Lawless J.F. (2003), *Statistical Models and Methods for Lifetime Data*, Second Edition, *Wiley Series in Probability and Statistics*, U.S.A.
- [8] Le C.T. (1997), *Applied Survival Analysis*, First Edition, *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, Inc., New York, U.S.A.
- [9] Lee E.T., Wang J.W. (2003), *Statistical Methods for Survival Data Analysis*, Third Edition, *Wiley Series in Probability and Statistics*, U.S.A.
- [10] Li G., Li R., Zhou M. (2000), *Empirical Likelihood in Survival Data*, *Mathematics Subject Classification*: 62N02, 62N03, 62G20, 3-8.
- [11] Luan J. (2004), *Empirical Likelihood and Right-censoring and Left-truncation Data*, PhD Thesis, University of Kentucky, Lexington, Kentucky, 5-15.
- [12] McGilchrist C.A., Aisbett C.W. (1991), *Regression with Frailty in Survival Analysis*, University of South Wales, P.O. Box 1, Kensington N.S.W. 2033, Australia, 461-466.

- [13] Mood A.M., Graybill F.A., Boes D.C. (1974), Introduction to The Theory of Statistics, Third Edition, McGraw-Hill International Book Company, Singapore.
- [14] North Central Cancer Treatment Group (1994), Prospective Evaluation of Prognostic Variable From Patient-Completed Questionnaires, Journal of Clinical Oncology, Vol 12, No 3(March), pp 601-607
- [15] Owen A. (1988), Empirical Likelihood Ratio Confidence Intervals for a Single Functional, Biometrika, 75, 237-249.
- [16] Owen A. (1990), Empirical Likelihood Confidence regions, The Annals of Statistics, 18, 90-120.
- [17] Owen A.B. (2001), Empirical likelihood, First Edition, Chapman & Hall/CRC, Boca Raton, London, New York, Washington, D.C.
- [18] Pan X. (1997), Empirical likelihood Ratio Method for Censored Data, PhD Thesis, University of Kentucky, Lexington, Kentucky, 56-63.
- [19] Pan X., Zhou M. (2002), Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Censored Data, Journal of Multivariate Analysis, 166-188.
- [20] Qin G., Jing B. (2006), Empirical Likelihood For Cox Regression Model Under Random Censorship, Communications in Statistics-Simulation and Computation, 30:1, 79-90.
- [21] Sun J. (2006), The Statistical Analysis of Interval-censored Failure Time Data, Springe Science-Business Media, Inc., Series Editors: Gail M., Krickeberg K., Samet J., Tsiatis A., New York, U.S.A.
- [22] Therneau T.M., Grambsch P.M. (2000), Modeling Survival Data: Extending the Cox Model, First edition, Springer Science-Business Media, Series Editors: Dietz K., Gail M., Krickeberg K., Samet J., Tsiatis A., New York, U.S.A.
- [23] Wikipedia, [https://en.wikipedia.org/wiki/Proportional\\_hazards\\_model](https://en.wikipedia.org/wiki/Proportional_hazards_model), 06 Nisan 2020.
- [24] Wikipedia, [https://en.wikipedia.org/wiki/Nelson%E2%80%93Aalen\\_estimator](https://en.wikipedia.org/wiki/Nelson%E2%80%93Aalen_estimator), 06 Nisan 2020.
- [25] Wikipedia, [https://en.wikipedia.org/wiki/Empirical\\_likelihood](https://en.wikipedia.org/wiki/Empirical_likelihood), 06 Nisan 2020.

[26] Xlstat by Addinsoft, <https://www.xlstat.com/en/solutions/features/nelson-aalen-analysis>, 10 Nisan 2020.

[27] Yay M., Çoker E., Uysal Ö. (2007), Yaşam analizinde cox regresyon modeli ve artıkların incelenmesi, Cerrahpaşa Medical Journal, Cerrahpaşa Tıp Dergisi, 38, 139-145.

[28] Zhao Y., Jinnah A. (2012), Inference For Cox's Regression Models Via Adjusted Empirical Likelihood, Comput Stat, 27, 1-12

[29] Zhou M. (2016), Empirical Likelihood Method in Survival Analysis, First Edition, Chapman & Hall/CRC Biostatistics Series, Taylor & Francis Group, LLC, Lexington, U.S.A.



# ÖZGEÇMİŞ

## KİŞİSEL BİLGİLER

**Ad Soyad:** Kübra ORAN

**Doğum Tarihi:** 17.02.1994

**Yabancı Dili:** İngilizce

**E-mail:** kubra.oran05@gmail.com

## EĞİTİM BİLGİSİ

**Lise:** Nesrin Uçmaklıođlu Lisesi (2008-2012)

**Üniversite:** Balıkesir Üniversitesi (2012-2013)

Marmara Üniversitesi (2013-2016)

**Bölüm:** Matematik Öğretmenliđi