

**ARALIKLI DURDURULMUŞ VERİLER İLE
YAŞAM ÇÖZÜMLEMESİ**

**SURVIVAL ANALYSIS
WITH INTERVAL CENSORED DATA**

İLKNUR ERÖZ

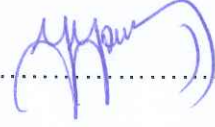
DOÇ. DR. NİHAL ATA TUTKUN
Tez Danışmanı

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

2019

İLKNUR ERÖZ'ün hazırladığı **“Aralıklı Durdurulmuş Veriler ile Yaşam Çözümlemesi”** adlı bu çalışma aşağıdaki jüri tarafından **İSTATİSTİK ANABİLİM DALI'** nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Yasemin YAVUZ
Başkan



Doç. Dr. Nihal ATA TUTKUN
Danışman



Prof. Dr. Duru KARASOY
Üye



Doç. Dr. Derya ERSEL
Üye



Dr. Öğr. Üyesi Uğur KARABEY
Üye



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

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK

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

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

beyan ederim.

11/06/2019



İLKNUR ERÖZ

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) veya elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- ☐ Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ay ertelenmiştir.
- ☐ Tezim ile ilgili gizlilik kararı verilmiştir.

11/06/2019



İLKNUR ERÖZ

ÖZET

ARALIKLI DURDURULMUŞ VERİLER İLE YAŞAM ÇÖZÜMLEMESİ

İlknur ERÖZ

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Doç. Dr. Nihal ATA TUTKUN

Haziran 2019, 76 sayfa

Durdurma, yaşam çözümlemesini diğer istatistiksel yöntemlerden ayıran en önemli özelliktir. Durdurulmuş gözlem tamamlanamamış gözlem demektir ve başarısızlığın gerçekleşme zamanı hakkında kısmen bilgi vermektedir. Bunun anlamı, bir birimin bir süre boyunca gözlenmesine rağmen bu süreçte başarısızlığın meydana gelmemesidir. Ayrıca gözlem altındaki bir birim bazı nedenlerden dolayı gözlemden çıkabilir ya da birimler araştırmaya farklı zamanlarda da dâhil olmuş olabilir. Eğer başarısızlık süresi, bu gibi nedenlerden dolayı tamamlanamamış ise durdurulmuş gözlem söz konusudur.

Yaşam çözümlemesinde farklı durdurma türleri olmakla birlikte uygulamada en çok sağdan durdurma ile karşılaşilmektedir. Yaşam çözümlemesinde kullanılan yöntemlerin çoğu gözlemlerin sağdan durdurulmuş olduğu veri kümeleri için geliştirilmiştir. Aralıklı durdurulmuş veriler ile ilgili çalışmalar da son yıllarda hız kazanmıştır. Birimin başarısızlık süresi belli bir aralıkta gerçekleşiyorsa aralıklı durdurulmuş gözlemler söz konusudur. Bu veriler genellikle, ilgilenilen gözlemlerin sürekli izlenemediği durumda oluşur. Sadece ilgilenilen olayın iki gözlem periyodu arasında meydana geldiği bilinmektedir.

Yaşam çözümlerinde yaygın olarak kullanılan Cox regresyon modeli ve parametrik yaşam modelleri aralıklı durdurulmuş veriler için geliştirilmiştir. Bu çalışmada aralıklı durdurulmuş veriler için yaşam modelleri teorik açıdan incelenmiş ve uygulama bölümünde ise Primer Biliyer Siroz verisi kullanılmıştır. Veri yapısı hem aralıklı durdurulmuş hem de sağdan durdurulmuş olarak ele alınmıştır. Cox regresyon modeli ve parametrik yaşam modelleri uygulanmış ve sonuçlar tartışılmıştır.

Anahtar Kelimeler: Aralıklı durdurma, Cox regresyon modeli, hızlandırılmış başarısızlık süresi modeli, orantılı tehlikeler, sağdan durdurma.

ABSTRACT

SURVIVAL ANALYSIS WITH INTERVAL CENSORED DATA

İlknur ERÖZ

Master of Science, Department of Statistics

Supervisor: Assoc. Prof. Dr. Nihal ATA TUTKUN

June 2019, 76 pages

Censoring is the most important feature that distinguishes survival analysis from other statistical methods. Censored observation means incomplete observation and provides partial information about the time of failure. This means that even though a unit has been observed for a period of time, there is no failure in this process. In addition, a unit under observation may be excluded from the study for some reason or units may have been involved in research at different times. If the failure time is not completed due to such reasons, then censored observation occurs.

In survival analysis, there are different types of censoring, however in practice the most common type is right censoring. Most of the methods used in survival analysis have been developed for right censored data sets. Studies on interval censoring gains speed in recent years. If the failure time of the unit occurs within a certain range, there are interval censoring observations. These data usually occur when observations of interest are not constantly monitored. It is only known that the event of interest takes place between two observation periods.

Cox regression model and parametric survival models used commonly in survival analysis have been developed for interval censored data. In this study, survival models for interval censored data is examined theoretically and primary biliary cirrhosis data is used in the application section. The data structure is taken both as right censored and interval censored. Cox regression model and parametric survival models are applied and the results are discussed.

Keywords: Interval censoring, Cox regression model, accelerated failure time model, proportional hazards, right censoring.

TEŞEKKÜR

Tezimin oluşmasında bana en büyük desteği veren, çalışmalarımda bana yol gösteren ve teşvik eden, bilgisini ve tecrübesini benimle paylaşan danışmanım Sayın Doç. Dr. Nihal Ata Tutkun'a,

Her zaman ve her konuda yanımda olan, benden hiçbir desteğini ve yardımını esirgemeyen aileme,

Değerli önerileri ile tezime katkıda bulunan Sayın Prof. Dr. Duru Karasoy, Sayın Prof. Dr. Yasemin Yavuz, Sayın Doç. Dr. Derya Ersel ve Sayın Dr. Öğr. Üyesi Uğur Karabey'e,

tüm samimiyetimle teşekkür ederim.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR.....	v
İÇİNDEKİLER	vi
ÇİZELGELER.....	viii
ŞEKİLLER.....	ix
SİMGELER VE KISALTMALAR	x
1. GİRİŞ	1
2. YAŞAM ÇÖZÜMLEMESİ	5
2.1. Giriş.....	5
2.2. Yaşam Çözümlemesi İçin Temel Kavramlar	6
2.3. Durdurma ve Durdurma Türleri	9
2.3.1. Sağdan Durdurma.....	9
2.3.2. Aralıklı ve Soldan Durdurma	9
2.4. Parametrik Olmayan Yaşam Çözümlemesi Yöntemleri	12
2.4.1. Kaplan-Meier Yaşam Fonksiyonunun Tahmini.....	12
2.4.2. Yaşam Fonksiyonlarının Parametrik Olmayan Karşılaştırması	13
2.5. Cox Regresyon Modeli.....	13
2.6. Parametrik Yaşam Modelleri	17
2.6.1. Üstel Yaşam Modeli	17
2.6.2. Weibull Yaşam Modeli.....	19
2.6.3. Log-Normal Yaşam Modeli.....	20
2.6.4. Log-Lojistik Yaşam Modeli	21
2.6.5. Gamma Yaşam Modeli.....	22
2.6.6. Parametrik Yaşam Modelleri İçin Model Seçim Kriterleri ve Uyum İyiliğinin Test Edilmesi.....	22
3. ARALIKLI DURDURULMUŞ VERİLER İLE YAŞAM ÇÖZÜMLEMESİ	24
3.1. Giriş.....	24
3.2. Aralıklı Durdurulmuş Veri Yapısı	27
3.3. Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli	29

3.3.1. Durum I Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli..	29
3.3.2. Durum II Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli.	36
3.4. Aralıklı Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri.....	40
3.4.1. Giriş.....	40
3.4.2. Aralıklı Durdurulmuş Veriler İçin Hızlandırılmış Başarısızlık Süresi Modeli	45
4. UYGULAMA.....	47
4.1. Sağdan Durdurulmuş Veri Yapısı İçin Model Sonuçları	49
4.1.1. Sağdan Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli Sonuçları.....	49
4.1.2. Sağdan Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri Sonuçları.....	56
4.2. Aralıklı Durdurulmuş Veri Yapısı İçin Model Sonuçları.....	59
4.2.1. Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli Sonuçları.....	60
4.2.2. Aralıklı Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri Sonuçları.....	60
5. SONUÇLAR.....	64
KAYNAKLAR	66

ÇİZELGELER

Çizelge 1.1.	Olaylara ilişkin başlangıç ve bitiş noktalarını içeren örnekler.....	1
Çizelge 2.1.	Yaşam çözümlemesinde durdurma türlerine göre veri yapısı	12
Çizelge 3.1.	Meme kanseri veri kümesi.....	26
Çizelge 4.1.	Kategorik değişkenler ve düzeyleri	48
Çizelge 4.2.	Sağdan durdurulmuş veri yapısı için CRM sonuçları	49
Çizelge 4.3.	OT varsayımı için Schoenfeld artıkları ile yaşam süresinin rankı arasındaki ilişki sonuçları	50
Çizelge 4.4.	İlaç değişkeni için CRM sonuçları.....	53
Çizelge 4.5.	İlaç ve İlaçxlogt değişkeni için CRM sonuçları	53
Çizelge 4.6.	Cinsiyet değişkeni için CRM sonuçları.....	53
Çizelge 4.7.	Cinsiyet ve Cinsiyetxlogt değişkeni için CRM sonuçları.....	53
Çizelge 4.8.	Hepatomegali değişkeni için CRM sonuçları	53
Çizelge 4.9.	Hepatomegali ve Hepatomegalixlogt değişkeni için CRM sonuçları.....	54
Çizelge 4.10.	Histolojik Evre değişkeni için CRM sonuçları.....	54
Çizelge 4.11.	Histolojik Evre ve Histolojik Evrexlogt değişkeni için CRM sonuçları	54
Çizelge 4.12.	Histolojik Evre ve Histolojik Evrexlogt değişkeni için CRM sonuçları	56
Çizelge 4.13.	Sağdan durdurulmuş veri yapısı için Gompertz OT modeli sonuçları	58
Çizelge 4.14.	Sağdan durdurulmuş veri yapısı için Weibull HBS modeli sonuçları	59
Çizelge 4.15.	Aralıklı durdurulmuş veri yapısı için CRM sonuçları	60
Çizelge 4.16.	Aralıklı durdurulmuş veri yapısı için parametrik yaşam modellerine ait AIC ve BIC değerleri.....	61
Çizelge 4.17.	Aralıklı durdurulmuş veri yapısı için Gompertz OT modeli sonuçları	62
Çizelge 4.18.	Aralıklı durdurulmuş veri yapısı için Gamma HBS modeli sonuçları	63

ŞEKİLLER

Şekil 2.1. Sağdan durdurma	10
Şekil 2.2. Soldan durdurma	10
Şekil 2.3. Aralıklı durdurma	11
Şekil 4.1. Değişkenlerin orantılı tehlikeler varsayımı.....	51-52
Şekil 4.2. Cox-Snell artıklarına ait grafikler.....	57

SİMGELER VE KISALTMALAR

Kısaltmalar

AIC	Akaike Bilgi Kriteri
BIC	Bayesci Bilgi Kriteri
CRM	Cox Regresyon Modeli
EÇO	En Çok Olabilirlik
HBS	Hızlandırılmış Başarısızlık Süresi
LR	Olabilirlik Oranı
OT	Orantılı Tehlikeler
PBS	Primer Biliyer Siroz
TO	Tehlike Oranı
ZB	Zamana Bağlı

1. GİRİŞ

Yaşam çözümlemesi, yaşam süresinin istatistiksel modellemesi ve çözümlemesi ile ilgilenmektedir. Yaşam süresi, belirli bir başlangıç noktasından önceden belirlenmiş bir olay meydana gelene kadar geçen süreyi belirtmek için kullanılır. Endüstriyel uygulamalarda ilgilenilen olay, genellikle bir bileşenin veya tüm makinenin veya tüm üretim sürecinin başarısızlığıdır (Barlow ve Proschan 1975). Bu durumda, yaşam süresi yerine “başarısızlık süresi” terimi de kullanılabilir. Başarısızlık süresi, ilgilenilen olay için olumsuz bir çağrışıma sahip olsa da olumlu bir olay da olabilir.

Olaylara ilişkin başlangıç ve bitiş noktalarını içeren örnekler aşağıdaki gibi verilebilir:

Çizelge 1.1. Olaylara ilişkin başlangıç ve bitiş noktalarını içeren örnekler

Olay	Başlangıç Noktası	Bitiş Noktası
Herhangi bir nedenden ölüm	Doğum	Ölüm tarihi
Hastalıktan ölüm	Hastalık teşhisi	Ölüm tarihi
Hastalığın nüksetmesi	Tedaviye başlama	Nüksettiği tarihi
İyileşme	Ameliyat	İyileşme tarihi
Mezuniyet	Üniversiteye giriş	Mezuniyet tarihi
İş bulma	İşsiz kalma	Yeniden işe başlama tarihi
Boşanma	Evlilik tarihi	Boşanma tarihi
Poliçenin bozulması	Sigorta başlangıç tarihi	Sigorta bitiş tarihi
Kredi kartını iade etmek	Kredi kartının kullanıma açılma tarihi	Kredi kartının kullanıma kapanma tarihi

Yaşam çözümlemesi, çoğunlukla biyoistatistik alanında kullanılmaktadır. Araştırmacılar, belirli bir tedavinin başarısızlık süresi üzerindeki etkisini incelemek için genellikle parametrik olmayan yöntemlerden Kaplan-Meier yaşam olasılığı tahminlerini ve log-rank testini kullanırlar. Modelleme yapmak için en çok kullanılan model ise Cox regresyon modelidir (Cox, 1972). Durdurulmuş gözlemlerin olması ve başarısızlık riskinin ya da tehlike fonksiyonunun modellemede esnek olması Cox regresyon modeli (CRM)’nin tercih edilme nedenidir. Hızlandırılmış

başarısızlık süresi (HBS) modelleri ise yaşam süresinin parametrik bir dağılıma uygun olması durumunda kullanılır.

Yaşam modellerini diğer istatistiksel modellerden ayıran, durdurulmuş gözlemleri içeren veri kümeleri için tasarlanmış istatistiksel yöntemler bütünü olmasıdır. Klasik istatistiksel yöntemlerden farklı olarak yaşam çözümlemesinin en önemli özelliklerden biri durdurmadır. Durdurulmuş gözlem, başarısızlığın meydana gelme süresi hakkında kısmen bilgi vermektedir ve tamamlanamamış gözlem olarak da adlandırılabilir. Yani, gözlem süresi içerisinde başarısızlığın meydana gelmemesi durumunda durdurulmuş gözlemden söz edilebilir. Bu durumda başarısızlık, gözlemlenen durdurma süresinden daha sonraki bir sürede gerçekleşmektedir. Ayrıca başarısızlık süresi, gözlemlenen bir birimin bazı nedenlerden dolayı gözlemden çıkabilmesi ya da birimlerin araştırmaya farklı zamanlarda dahil olabilmesi gibi nedenler ile tamamlanamamış ise durdurulmuş gözlemden söz edilebilir.

Durdurma türleri sağdan, soldan ve aralıklı durdurma olmak üzere üç grupta incelenebilir. Yaşam çözümlemesinde genellikle sağdan durdurulmuş veri yapısı için çalışmalar yapılmıştır. Bununla birlikte sağdan ve soldan durdurma, aralıklı durdurmanın özel biçimleridir. Aralıklı durdurulmuş veri yapısı farklı olduğundan kullanılacak yaşam çözümlemesi yöntemleri sağdan durdurulmuş veri yapısı için kullanılan yöntemlerden farklıdır.

Aralıklı durdurma, ilgilenilen olayın gözlemlenemediği ve sadece iki zaman aralığının içinde ortaya çıktığının bilindiği durumlarda söz konusudur. Aralıklı durdurulmuş veri yapısı, bir aralıktaki birimlerin başarısız oldukları kesin zaman noktaları ile ilgili belirsizliği yansıtmaktadır. Aralıklı durdurulmuş veri türleri, Durum I aralıklı durdurulmuş veriler (mevcut durum verileri) ve Durum II (genel) aralıklı durdurulmuş veriler olmak üzere ikiye ayrılmaktadır. Durum I aralıklı durdurulmuş veri yapısı, sadece yaşam süresi bilgisi ile gözlem süresinin öncesinde ya da sonrasında meydana gelen soldan veya sağdan durdurulmuş veri yapılarıdır. Bu nedenle, Durum I aralıklı durdurulmuş veri yapısı, durdurulmamış ve $(a, b]$ aralığında aralıklı durdurulmuş gözlemlere bakılmaksızın, Durum II aralıklı durdurulmuş veri yapısının özel bir durumu olarak gözlemlenmektedir. Durum I

aralıklı durdurulmuş veri yapısında, birimler sadece bir kez gözlemlenmekte ve sadece olayın gözlemlenmeden önce gerçekleşip gerçekleşmediği bilinmektedir. Bu tür veriler demografik çalışmalarda yaygındır.

Kesin başarısızlık süresinin bilinmediği durumlarda meydana gelen olayın, sadece başarısızlığın rasgele bir zaman aralığında, sol uç noktasından önce ya da sağ uç noktasından sonra olduğu biliniyorsa Durum II aralıklı durdurulmuş veri yapısı söz konusudur. Bu durumda her gözlem keyfi olarak durdurulabilmektedir. Durum II aralıklı durdurulmuş veri yapısına, Durum I aralıklı durdurulmuş veri yapısına göre daha sık rastlanmaktadır.

Yaşam çözümlemesinde, aralıklı durdurulmuş veri yapısı için parametrik olmayan yaşam modelleri ilk kez Turnbull (1976) tarafından incelenmiştir. Böhning ve diğ. (1996), Li ve diğ. (1997), Betensky ve diğ. (1999), Yu ve diğ. (2000), Bechuk ve Betensky (2000), Fang ve diğ. (2002), Zhao ve Sun (2004), Sun ve diğ. (2005), Vandal ve diğ. (2005), Kim ve diğ. (2006), Yuen ve diğ. (2006), Zhao ve diğ. (2009) çalışmaları ile geliştirilmiştir.

Aralıklı durdurulmuş veri yapısı için CRM'ye ilişkin çalışmalar Finkelstein ve Wolfe (1985), Finkelstein (1986) ile başlamış; Huang ve Wellner (1995), Huang (1996), Satten (1996), Younes ve Lachin (1997), Huang ve Rossini (1997), Goggins ve diğ. (1998), Geskus ve Groeneboom (1999), Pan (2000), Xue ve diğ. (2006), Zeng ve diğ. (2006), Sen ve Banerjee (2007), Heller (2011) ile devam etmiştir.

Aralıklı durdurulmuş veri yapısı için parametrik yaşam modelleri ise Rabinowitz ve diğ. (1995), Betensky ve diğ. (2001), Oller ve diğ. (2004), Zhang ve diğ. (2005), Lawless ve Babineau (2006), Tian ve Cai (2006), Zhang ve diğ. (2007), Zhu ve diğ. (2008) tarafından ele alınmıştır.

Radhey ve Totawattage (2013) çalışmasında aralıklı durdurulmuş veri yapısı için parametrik olmayan, yarı parametrik ve parametrik yaşam modellerinin uygulamasını aynı veri üzerinde göstermiştir.

Yaşam çözümlemesinde, çift (double) aralıklı durdurulmuş veri yapısı Sun (2002, 2004) tarafından incelenmiştir. Aralıklı durdurulmuş veri yapısına Bayesci yaklaşım

ise Komàrek ve Lesaffre (2007, 2008), Bogaerts ve diğ. (2018) tarafından çalışılmıştır. Kim ve Jhun (2008) çalışmasında aralıklı durdurulmuş veriler için iyileşme modelini incelemiştir.

Aralıklı durdurulmuş veri yapısının istatistiksel paket programlarındaki çözümlemeleri üzerine çalışmalar literatürde mevcuttur. R paketindeki çalışmalar Gomez ve diğ. (2009), Fay (2014), Fay ve Shaw (2010), Anderson-Bergman (2017), Delord (2017), Bogaerts ve diğ. (2018); SAS paketindeki çalışmalar So ve diğ. (2010), Guo ve diğ. (2014), Bogaerts ve diğ. (2018); STATA paketindeki çalışma ise Yang (2017) tarafından yapılmıştır.

Bu çalışmanın amacı, aralıklı durdurulmuş veri yapısı için CRM'yi ve parametrik yaşam modellerini incelemek ve sağdan durdurulmuş veri yapısı için kullanılan klasik yöntemler ile karşılaştırma yapmaktır.

Çalışmanın ikinci bölümünde yaşam çözümlemesiyle ilgili temel kavramlar, durdurma ve durdurma türleri, yaşam çözümlemesinde kullanılan yöntemler ve modeller hakkında genel bilgiler verilmiştir.

Çalışmanın üçüncü bölümünde aralıklı durdurulmuş veri yapısı detaylı olarak incelenmiş, aralıklı durdurulmuş veri yapısı için CRM ve parametrik yaşam modelleri tanıtılmıştır.

Çalışmanın dördüncü bölümünde ise literatürde bulunan Primer Biliyer Siroz (PBS) verisi kullanılarak sağdan ve aralıklı durdurulmuş veri yapısı için CRM ile parametrik yaşam modelleri incelenmiş ve sonuçlar yorumlanmıştır.

2. YAŞAM ÇÖZÜMLEMESİ

2.1. Giriş

İlgilenilen sonuç değişkeni için olayın ortaya çıkışına kadar geçen süreyi analiz etmek amacıyla kullanılan istatistiksel yöntem yaşam çözümlemesidir. Bu nedenle yaşam çözümlemesi tıp, halk sağlığı, sosyal bilimler ve mühendislik gibi birçok uygulamalı alanda "olay gerçekleşene kadar geçen sürenin çözümlemesi" (time to event analysis) olarak da adlandırılır. Tıp bilminde olay süresi, kanser hastalığının nüksetmesi, enfeksiyon oluşması veya ölüme kadar geçen süre olabilir. Sosyal bilimlerde olay süresi çözümlemede ilgilenilen iş değişikliği, evlilik, çocukların doğumu vb. olabilir. Mühendislik bilimlerinin temel odak noktası makinelerin veya elektronik bileşenlerin ömrünü modellemek olduğu için mühendislik bilimleri, başarısızlık süresi çözümlemesi olarak da adlandırılan yaşam çözümlemesinin geliştirilmesine katkıda bulunmuştur. Farklı alanlardaki araştırmacılar tarafından uyarlandığı alana göre farklı biçimde ifade edilmiştir: Sosyolojide olay geçmiş çözümlemesi, mühendislikte başarısızlık süresi çözümlemesi, ekonomide süre çözümlemesi veya geçiş çözümlemesi. Bu farklı ifadeler teknik açıdan gerçek bir farklılığa işaret etmemektedir, ancak farklı disiplinler farklı yaklaşımları vurgulayabilir. Yaşam çözümlemesi, en yaygın kullanılan ifadedir.

Durdurulmuş gözlemlerin varlığının neden olduğu karmaşıklıklar, yeni bir istatistiksel metodoloji alanı geliştirilmesine yol açmıştır. Yaşam çözümlemesindeki metodolojik gelişmeler büyük ölçüde 20. yüzyılın ikinci yarısında sağlanmıştır.

Yaşam verilerini çözümlmek için en eski ve en basit parametrik olmayan yöntemlerden biri, Berkson ve Gage (1950) tarafından kanserli bireylerin yaşamını incelemek için önerilen yaşam tablosunu oluşturmaktır. Parametrik olmayan yöntemlerde en önemli gelişme Kaplan ve Meier (1958) tarafından elde edilmiştir. Parametrik olmayan yöntemler homojen örnekler için yeterli bilgi verse de belirli değişkenlerin yaşam süreleriyle ilişkili olup olmadığını belirlemezler. Bu ihtiyaç, yaşam verilerini çözümlmek için regresyon yöntemlerinin uygulanmasına yol açmaktadır. Klasik çoklu doğrusal regresyon modeli, birkaç nedenden dolayı yaşam verisine uygun değildir. Yaşam süreleri nadiren normal dağılır ve durdurulmuş veriler için bağımlı değişken olan yaşam süresi kayıp gözlem

içerebilir. CRM, artık kovaryantlar veya prognostik faktörlerin varlığında yaşam verilerinin çözümünde kullanılan en yaygın modeldir. Bu, basitliği nedeniyle yaşam çözümlemesi için en popüler modeldir ve yaşam süresi dağılımı ile ilgili herhangi bir varsayıma dayanmamaktadır. Model, tehlike oranının (TO) bağımsız değişkenlerin bir fonksiyonu olduğunu varsayar, ancak tehlike fonksiyonunun niteliği veya şekli hakkında hiçbir varsayım yapmaz. Son birkaç yılda, modelin teorik temeli Fleming ve Harrington (1991) ve Andersen ve diğ. (1993) kitaplarında tartışılan sayma süreçleri ve martingale teorisi çalışmalarına bağlanarak sağlamlaştırılmıştır. Bununla birlikte, CRM birçok durumda uygun olmayabilir ve zamana bağlı değişkenli CRM ya da tabakalı CRM gibi diğer modeller yaşam verilerinin çözümlemesi için kullanılabilir. HBS modeli, yaşam verilerinin çözümü için başka bir alternatif yöntemdir.

2.2. Yaşam Çözümlemesi İçin Temel Kavramlar

Yaşam çözümlemesinde kullanılan bazı temel kavramlar bu alt bölümde açıklanmıştır.

Yaşam Süresi: Belirli bir başlangıç zamanında ilgilenilen olayın meydana geldiği zamana kadar ölçülen süredir. Yaşam süresini kesin olarak belirlemek için üç koşul vardır: Bir zaman kaynağı kesin olarak tanımlanmalı, zamanın geçişini ölçmek için bir ölçek üzerinde anlaşmaya varılmalı ve son olarak olayın tanımı (genellikle başarısızlık olarak adlandırılır) tamamen açık olmalıdır. Yaşam çözümlemesindeki spesifik zorluklar büyük ölçüde, sadece bazı birimlerin olayı yaşamış olmasından ve diğer birimlerin çalışmanın sonunda olayı yaşamamış olmasından, dolayısıyla gerçek yaşam sürelerinin bilinmemesinden kaynaklanmaktadır. Bu durum, durdurma kavramına yol açmaktadır.

Durdurma: Durdurma, yaşam süresi hakkında bazı bilgilere sahip olduğunda ortaya çıkmaktadır, ancak yaşam süresi kesin olarak bilinmemektedir. Üç tür durdurma vardır: Sağdan durdurma, soldan durdurma ve aralıklı durdurma.

Olay gözlemlenen yaşam süresinden sonra meydana gelirse sağdan durdurmanın gerçekleştiği söylenir. C durdurma süresini, yani olayın gözlemlenemediği süreyi gösterebilir. Gözlemlenen yaşam süresi ayrıca takip süresi olarak da ifade edilir. Bu

süre 0 anında başlar ve hangisi önce gelirse, X olayına veya C durdurma süresine kadar devam eder. Gözlemlenen veriler (T, δ) ile belirtilir, burada $T = \min(X, C)$ takip süresidir. $\delta = I_{(X \leq C)}$ takip sonundaki durum göstergesidir ve

$$\delta = I_{(X \leq C)} = \begin{cases} 0, & X > C \text{ (gözlemlenen durdurma)} \\ 1, & X \leq C \text{ (gözlemlenen başarısızlık)} \end{cases}$$

biçiminde ifade edilir. Sağdan durdurmanın ortaya çıkmasının bazı nedenleri vardır. Çalışma sona ermeden hiç olay olmaması, çalışma süresi boyunca gözlemlerde kayıp olması veya gözlemlerin bazı nedenlerden dolayı çalışmadan geri çekilmesi bu nedenlere örnek verilebilir. Son neden, yarışan risklerden kaynaklanabilmektedir. Gerçek yaşam süresi, sağdan durdurulmuş yaşam süresinden daha fazladır.

İlgilenilen olayın meydana geldiğini gözlemleyebiliyor, ancak ne zaman başladığı bilinmiyorsa soldan durdurma söz konusu olabilir ve gerçek yaşam süresi gözlemlenen durdurulmuş süreden daha azdır.

Bir kişinin belli bir zaman dilimi içinde bir olay yaşadığı biliniyor ancak gerçek yaşam süresi bilinmiyorsa, aralıklı durdurmadan bahsedilebilir. Bu durumda, olayın gerçekleşme süresinin bir zaman aralığı içinde olduğu bilinmektedir.

Yaşam Süresi Dağılımı : T yaşam süresini gösteren rasgele bir değişken ise yaşam süresinin dağılımı, olasılık yoğunluk fonksiyonu (o.y.f.), yaşam fonksiyonu ya da tehlike fonksiyonu ile ifade edilmektedir.

Yaşam Fonksiyonu: Yaşam fonksiyonu, yaşam süresinin t'ye eşit ya da t'den daha büyük olma olasılığıdır:

$$S(t) = P(T \geq t), t \geq 0.$$

Tehlike Fonksiyonu: Birimin t zamanına kadar yaşaması koşulu ile t zamanındaki anlık başarısızlık riskini tehlike fonksiyonu vermektedir. Tehlike fonksiyonu;

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}, \quad t \geq 0$$

biçimindedir. $S(t)$ ve $h(t)$ arasında açıkça tanımlanmış bir ilişki vardır ve bu ilişkiler Eşitlik 2.1 ve Eşitlik 2.2 ile verilmektedir:

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt}, \quad (2.1)$$

$$S(t) = \exp \left[-\int_0^t h(u) du \right] = \exp(-H(t)), \quad t \geq 0 \quad (2.2)$$

Burada, $H(t) = -\log S(t)$ 'den beri yaşam fonksiyonundan elde edilebilecek

$H(t) = \int_0^t h(u) du$ birikimli tehlike fonksiyonu olarak adlandırılmaktadır.

Olasılık Yoğunluk Fonksiyonu: T için o.y.f. aşağıdaki gibidir:

$$f(t) = h(t) \exp \left[-\int_0^t h(u) du \right], \quad t \geq 0.$$

Bu üç fonksiyon, T yaşam süresi dağılımlarının matematiksel olarak aynı özelliklerini vermektedir. Bunlardan biri biliniyorsa, diğer ikisi belirlenebilmektedir. Bu fonksiyonlardan biri, belirli durumlara göre istatistiksel çözümlemelerin temeli olarak seçilebilmektedir. Yaşam fonksiyonu, iki veya daha fazla grubun yaşam durumlarını karşılaştırmak için kullanılırken, tehlike fonksiyonu herhangi bir zaman noktasında başarısızlık riskini ifade etmek için kullanılmaktadır.

2.3. Durdurma ve Durdurma Türleri

2.3.1. Sağdan Durdurma

Yaşam verilerinin tipik bir özelliği, olay süresinin her zaman doğru bir şekilde gözlenememesi ve gözlemlerin durdurmaya tabi tutulmasıdır. Genellikle, çalışmaya alınan tüm birimler ilgilenilen olayı göstermeden ya da bir birim bir olay ile karşılaşmadan önce çalışma sona ermektedir. Her iki durumda da, sadece T gerçek olay süresi için alt sınır ℓ olmak üzere, $T > \ell$ olduğu bilinmektedir. Bu durumda, verilerde sağdan durdurma olduğu ve olay süresinin sağdan durdurulduğu söylenebilir.

2.3.2. Aralıklı ve Soldan Durdurma

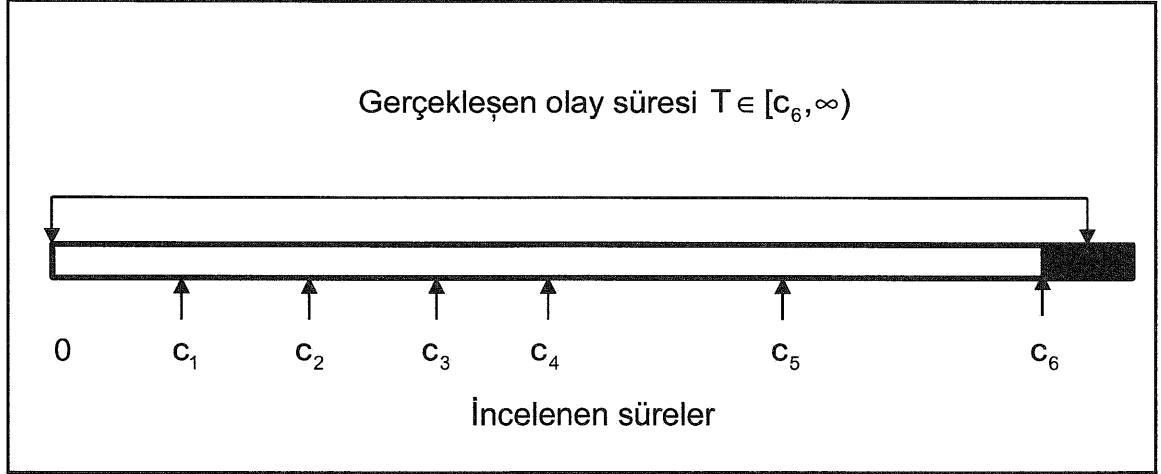
Çalışmaya bir birim dahil edilmeden önce ilgilenilen olay çoktan gerçekleşmiş, ancak olayın gerçekleştiği süre bilinmiyor ise soldan durdurma; olayın sadece iki gözlem zamanı arasında gerçekleştiği biliniyor ise aralıklı durdurma söz konusudur. Diş muayenesine giden bir çocuk hastanın, muayene olmadan önce diş çıkarması soldan durdurmaya; hastanın önceki muayenesinde diş yokken muayene esnasında diş olduğunun görülmesinden itibaren geçen süre aralıklı durdurmaya örnek verilebilir. Durdurma, HIV/AIDS çalışmalarında da sıkça görülmektedir. HIV serokonversiyonuna kadar geçen süre, sadece hekimi ziyaret ettikten sonra başlatılan laboratuvar değerlendirmesiyle belirlenebilmektedir. Öyleyse, sadece iki gözlem arasında HIV serokonversiyonunun gerçekleştiği sonucuna varılabilmektedir. Aynısı, klinik semptomlara dayanan ve tıbbi bir muayene ile onaylanması gereken AIDS tanısı için de geçerlidir.

Soldan durdurma ve aralık durdurma, yaşam süreleri için oldukça benzerdir. Aralıklı durdurmada ℓ alt sınır ve u üst sınır olmak üzere, T olay süresinin $\ell > 0$ ve $u < \infty$ ile ℓ 'den sonra ve u 'dan önce meydana geldiği bilinmektedir. Soldan durdurma için de aynı durum geçerlidir, ama $\ell = 0$ 'dır. Bu durumda, $\ell < T \leq u$, $\ell \leq T < u$, $\ell \leq T \leq u$ veya $\ell < T < u$ olduğu bilinmektedir. Bunu göstermek için $T \in [\ell, u]$ yazılabilir.

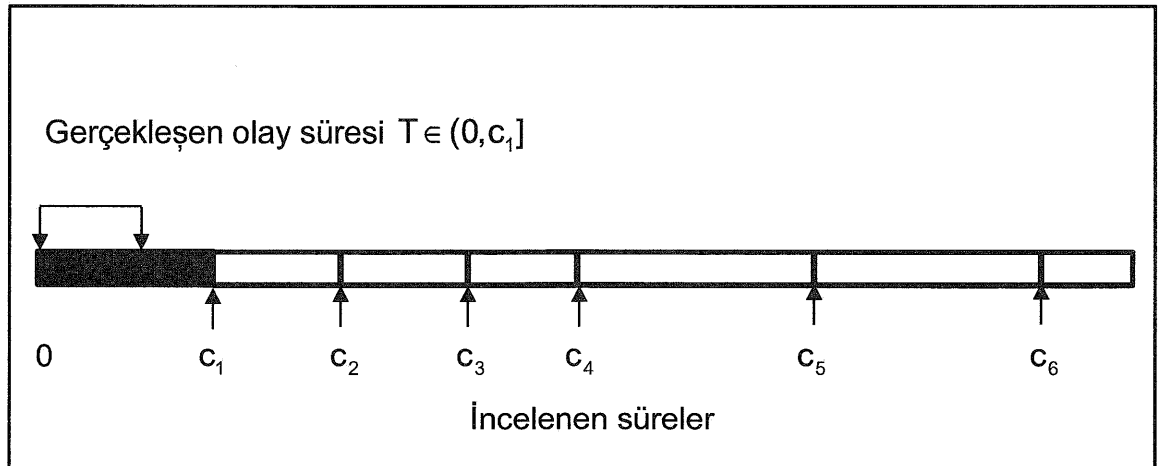
Aralıklı durdurulmuş bir gözlemin tanımında $\ell = 0$ ve $u = \infty$ varsa aralıklı durdurma, sağdan durdurma ve soldan durdurmanın bir genellemesi olmaktadır.

Tam olarak gözlemlenen olay zamanı bile $\ell = u = t$ ile aralıklı durdurulmuş bir gözlem olarak düşünülebilir.

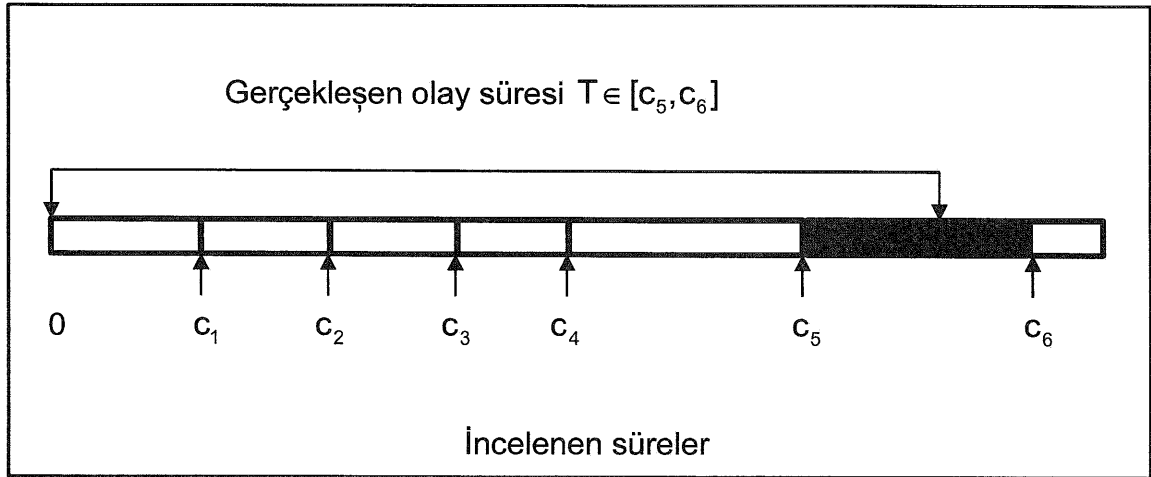
Sağdan durdurma Şekil 2.1, soldan durdurma Şekil 2.2 ve aralıklı durdurma Şekil 2.3 ile verilmiştir:



Şekil 2.1. Sağdan durdurma



Şekil 2.2. Soldan durdurma



Şekil 2.3. Aralıklı durdurma

$c_1, \dots, c_6$ gözlem sürelerindeki olay durumunu kontrol etmek için yapılan incelemelerde elde edilen durdurulmuş birimler ile ilgili, Şekil 2.1.'de c_6 gözlem süresinde birimin sağdan durdurulduğu ($\ell = c_6$), Şekil 2.2.'de c_1 gözlem süresinde birimin soldan durdurulduğu ($u = c_1$), Şekil 2.3.'te $[c_5, c_6]$ gözlem süresi aralığında birimin aralıklı durdurulduğu ($\ell = c_5$, $u = c_6$) görülmektedir.

Her birim kesin olarak iki kez gözlemlendiğinde, c_1 ve c_2 iki gözlem süresi ile bir birim $(0, c_1]$, $[c_1, c_2]$, $[c_2, \infty)$ gözlem biçimlerinden herhangi birini almaktadır (sabit veya rasgele). Bu durumda, Durum II (genel) aralıklı durdurmadan söz edilir. Benzer şekilde, k bir sabit olmak üzere, k inceleme süresinden elde edilen verilere durum k aralıklı durdurma uygulanmıştır denebilir (Gomez ve diğ. 2004).

Birkaç sabit veya rasgele inceleme süresi olduğunda, elde edilen veri kümesi sadece $\ell < u$ için $[\ell, u]$ aralığını ve $0 < \ell < u < \infty$ en az bir aralık içerir. Buna Durum II (genel) aralıklı durdurma denir (Sun, 2006). Her iki aralık da kesin gözlemlerle birlikte $0 \leq \ell < u \leq \infty$ için $[\ell, u]$ aralığı, yani $\ell = u$ için $[\ell, u]$ aralıkları mevcut olduğunda, karışık aralıklı durdurulmuş veri yapısından bahsedilebilir (Yu ve diğ. 2000).

Yaşam çözümlemesinde durdurma türleri için veri yapısı, t_1 başlangıç süresi ve t_2 bitiş süresi olmak üzere, Çizelge 2.1’de verildiği gibi ele alınmaktadır:

Çizelge 2.1. Yaşam çözümlemesinde durdurma türlerine göre veri yapısı

Veri Türü		t_1	t_2
Durdurulmamış Veri Yapısı	$a = [a, a]$	a	a
Aralıklı Durdurulmuş Veri Yapısı	$(a, b]$	a	b
Soldan Durdurulmuş Veri Yapısı	$(0, b]$	.	b
Soldan Durdurulmuş Veri Yapısı	$(0, b]$	0	b
Sağdan Durdurulmuş Veri Yapısı	$[a, +\infty)$	a	.
Kayıp		.	.
Kayıp		0	.

2.4. Parametrik Olmayan Yaşam Çözümlemesi Yöntemleri

Yaşam çözümlemesinde, birimler için yaşam sürelerinin sayısal veya grafiksel özetlerini sunmak önemlidir. Genel olarak yaşam verileri, yaşam fonksiyonu ve tehlike fonksiyonunun tahminleri ile özetlenmektedir. Yaşam dağılımının tahmini, ortanca yaşam süresi gibi tanımlayıcı istatistiklerin tahminlerini sağlamaktadır. Bu yöntemlerin parametrik olmayan yöntemler olduğu söylenmektedir, çünkü yaşam süresinin dağılımı hakkında hiçbir varsayım gerektirmezler.

2.4.1. Kaplan-Meier Yaşam Fonksiyonunun Tahmini

Yaşam tablosu, ölümü analiz etmek için kullanılan en eski istatistiksel yöntemdir, ancak Kaplan-Meier (KM) yöntemi gibi modern yöntemlerin geliştirilmesi ile önemi azalmıştır. Yaşam eğrileri için KM tahmin edicisi genellikle bireysel verileri analiz etmek için kullanılmaktadır, buna karşın yaşam tablosu yöntemi gruplanmış verilere uygulanmaktadır. Yaşam tablosu yöntemi gruplanmış bir veri istatistiği olduğu için, bireysel değerleri kullanan KM tahmini kadar kesin değildir (Singh ve Totawattage, 2013).

Kaplan-Meier tahmin edicisi, yaşam süresi verileri için yaşam fonksiyonunu tahmin etmektedir. Kaplan-Meier’in yaşam fonksiyonu tahminine ait grafik, yeterince

büyük bir örneklem alındığında, bu kitle için gerçek yaşam fonksiyonuna yaklaşan adimsal bir grafikdir.

n büyüklüğündeki bir örneklem için sıralanmış başarısızlık süreleri $t_1 \leq t_2 \leq t_3 \leq \dots \leq t_n$ olsun. Kaplan-Meier tahmin edicisi, $S(t)$ 'nin parametrik olmayan en çok olabilirlik tahminidir (EÇO) ve

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \quad (2.3)$$

biçimindedir. Burada t_i 'ye karşılık gelen n_i , t_i süresinden hemen önce “risk altındaki” sayıyı ve d_i , t_i süresindeki ölüm sayısını vermektedir. Durdurma olmadığında n_i , t_i süresinden hemen önce hayatta kalanların sayısını göstermektedir.

2.4.2. Yaşam Fonksiyonlarının Parametrik Olmayan Karşılaştırması

KM yaşam eğrileri, iki veya daha fazla grupta yaşam fonksiyonlarının farkı hakkında bir fikir verebilir, ancak bu gözlemlenen farkın istatistiksel olarak anlamlılığını belirlemek için istatistiksel test kullanmayı gerektirmektedir. Farklı gruplarda yaşam fonksiyonlarının eşitliğini test etmek için kullanılabilecek birkaç yöntem vardır. İki veya daha fazla yaşam dağılımının karşılaştırılması için yaygın olarak kullanılan yöntem, parametrik olmayan log-rank testidir (Mantel, 1966).

2.5. Cox Regresyon Modeli

Yaşam verilerinin modellenmesinde kullanılan en popüler model Cox tarafından önerilen Cox regresyon modelidir. CRM,

$$h(t | x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = h_0(t) \exp(\beta' x) \quad (2.4)$$

ile verilmektedir. Burada, temel tehlike fonksiyonu $h_0(t)$, açıklayıcı değişkenler vektörü $x = (x_1, x_2, \dots, x_p)'$, regresyon katsayıları vektörü $\beta' = (\beta_1, \beta_2, \dots, \beta_p)$ 'dir.

Bunlara karşılık gelen yaşam fonksiyonu ise,

$$S(t | x) = S_0(t) \exp\left(\sum_{i=1}^p \beta_i x_i\right).$$

biçiminde ifade edilir. CRM, $h_0(t)$ 'nin biçimi (modelin parametrik olmayan kısmı) ile ilgili hiçbir varsayımda bulunmaz, ancak tahmin edicilerin tehlike fonksiyonu üzerindeki etkisi için parametrik bir biçim (modelin parametrik kısmı) kabul etmektedir. Bu nedenle model, yarı parametrik bir model olarak adlandırılmaktadır. CRM'nin avantajı, bu belirsizliğin tahmin için hiçbir sorun yaratmamasıdır. Temel tehlike belirtilmemiş olsa da, regresyon katsayıları β ve TO için iyi bir tahmin yapılabilir.

x ve x^* iki birime ait açıklayıcı değişkenler olmak üzere TO,

$$TO = \frac{h_0(t) \exp(\hat{\beta}'x)}{h_0(t) \exp(\hat{\beta}'x^*)} = \exp\left(\sum \hat{\beta}'(x - x^*)\right) \quad (2.5)$$

biçimindedir. TO zamandan bağımsızdır. Bu nedenle CRM, orantılı tehlikeler (OT) modeli olarak da adlandırılmaktadır.

CRM kullanılarak $h_0(t)$ ve β tahmin edilmektedir. Cox (1972) tarafından β tahmini için $h_0(t)$ 'ye bağlı olmayan bir kısmi olabilirlik fonksiyonu önerilmiştir. Kısmi olabilirlik, nüans parametrelerinin varlığında regresyon parametreleri hakkında çıkarım yapmak için geliştirilen bir tekniktir.

$t_1, t_2, \dots, t_n$, n birim için gözlemlenen yaşam süresi olsun. n birimin sıralı başarısızlık süreleri $t_{(1)} < t_{(2)} < \dots < t_{(n)}$ ve $R(t_{(j)})$, $t_{(j)}$ süresinden hemen önceki risk kümesi olsun. Bu durumda $R(t_{(j)})$, $t_{(j)}$ süresinden hemen önceki bir sürede başarısızlık ile karşılaşmamış ve durdurulmamış birimlerin bir grubudur. i . birimin $t_{(j)}$ süresinde başarısız olmasının koşullu olasılığı, $t_{(j)}$ süresinde başarısız olan $R(t_{(j)})$ 'de gözlemlenen riskten bir birim olarak verilmesidir:

$P(t_{(j)}$ 'de başarısız olan i birim | $t_{(j)}$ süresinde $R(t_{(j)})$ 'de gözlemlenen riskten başarısız olan birim)

$$= \frac{P(t_{(j)} \text{'de başarısız olan i birim})}{P(t_{(j)} \text{'de başarısız olan birim})}$$

$$= \frac{P(t_{(j)} \text{'de başarısız olan i birim})}{\sum_{k \in R(t_{(j)})} P(t_{(j)} \text{'de başarısız olan k birim})}$$

$$\approx \frac{P\{(t_{(j)}, t_{(j)} + \Delta t) \text{'de başarısız olan i birim}\} / \Delta t}{\sum_{k \in R(t_{(j)})} \{P\{(t_{(j)}, t_{(j)} + \Delta t) \text{'de başarısız olan k birim}\} / \Delta t}$$

$$= \frac{\lim_{\Delta t \downarrow 0} P\{(t_{(j)}, t_{(j)} + \Delta t) \text{'de başarısız olan i birim}\} / \Delta t}{\lim_{\Delta t \downarrow 0} \sum_{k \in R(t_{(j)})} \{P\{(t_{(j)}, t_{(j)} + \Delta t) \text{'de başarısız olan k birim}\} / \Delta t}$$

$$= \frac{h_1(t_{(j)})}{\sum_{k \in R(t_{(j)})} h_k(t_{(j)})}$$

$$= \frac{h_0(t_{(j)}) \exp(\beta' x_1(t_{(j)}))}{\sum_{k \in R(t_{(j)})} h_0(t_{(j)}) \exp(\beta' x_k(t_{(j)}))}$$

$$= \frac{\exp(\beta' x_1(t_{(j)}))}{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k(t_{(j)}))}$$

biçimindedir.

CRM için kısmi olabilirlik fonksiyonu,

$$L(\beta) = \prod_{j=1}^r \frac{\exp(\beta' x_i(t_{(j)}))}{\sum_{k \in R(t_{(j)})} \exp(\beta' x_k(t_{(j)}))} \quad (2.6)$$

biçimindedir. Burada $x_i(t_{(j)})$, $t_{(j)}$ 'de başarısız olan i birim için açıklayıcı değişkenler vektörüdür. Bu olabilirlik fonksiyonunun sadece durdurulmamış birimler için olduğuna dikkat edilmelidir. $t_1, t_2, \dots, t_n$, n birim için gözlemlenen yaşam süresi ve δ_i , i . yaşam süresi durdurulursa sıfır (0) ve aksi halde bir (1) değerini alan durum göstergesidir. Eşitlik 2.6'daki olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' x_i(t_{(j)}))}{\sum_{k \in R(t_i)} \exp(\beta' x_k(t_{(j)}))} \right]^{\delta_i} \quad (2.7)$$

ile ifade edilebilmektedir. Veri kümesi içerisinde eş zamanlı birim olmadığında Eşitlik 2.7'de verilen kısmi olabilirlik fonksiyonu kullanılmaktadır. Eş zamanlı birimlerin olması durumunda Breslow, Efron ve Kesin olabilirlik fonksiyonlarının kullanılması gerekmektedir (Karasoy ve Kaplan, 2017).

CRM'de dağılım varsayımı olmamasına karşın, temel bir varsayım olan OT varsayımı bulunmaktadır. OT, bir birimin tehlike fonksiyonunun diğer birimin tehlike fonksiyonu ile orantılı olduğu, yani TO'nun zaman içinde sabit olduğu anlamına gelmektedir.

Grafiksel ya da sayısal yöntemler ile OT varsayımı incelenebilmektedir. En çok bilinen yöntemler modele zamana bağlı açıklayıcı değişkenlerin eklenmesi (Kalbfleisch ve Prentice, 1980), log-log yaşam eğrileri (Kalbfleisch ve Prentice, 1980), Schoenfeld artıklarına dayanan test istatistiği (Harrell, 1986; Schoenfeld, 1982) ve Arjas grafikleridir (Arjas, 1988). Ayrıntılı bilgi için Ata (2010) çalışmasına bakılabilir.

Kurulan modelin yeterliliği artıklar kullanılarak incelenmektedir. Doğrusal regresyon yöntemlerinde, artıklar bağımlı değişkenin gözlemlenen ve tahmin edilen değerleri arasındaki fark olarak tanımlanmaktadır. Bununla birlikte, durdurulmuş birimler mevcut olduğunda ve CRM'de kısmi olabilirlik fonksiyonu kullanıldığında, genel artık kavramı geçerli değildir. CRM için özellikle Cox-Snell Artığı ve Schoenfeld Artığı başta olmak üzere pek çok artık türü önerilmiştir (Hosmer ve Lemeshow, 1999).

Cox Regresyon Modeli İçin Model Seçim Kriterleri

CRM'de model uygunluğu Akaike bilgi kriteri (AIC) ya da Bayesci bilgi kriteri (BIC) ile analiz edilebilmektedir. En uygun modelin, en küçük AIC ya da BIC değerine sahip olması gerekmektedir.

L olabilirlik fonksiyonu, n gözlem sayısı ve p bilinmeyen parametre sayısını göstermek üzere, %95 güven düzeyinde AIC ve BIC formülleri aşağıdaki gibidir:

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

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

(Klein and Moeschberger, 1997).

2.6. Parametrik Yaşam Modelleri

Yaşam çözümlemesinde CRM'den sonra en sık kullanılan yaşam modelleri parametrik yaşam modelleridir. Bunlardan log-normal, log-lojistik, Gamma, Weibull ve üstel yaşam modelleri sıklıkla kullanılmaktadır.

2.6.1. Üstel Yaşam Modeli

Tek parametrelili üstel yaşam modeli, tehlike fonksiyonunun sabit olduğunu varsaymaktadır. Yani, $h(t) = \lambda > 0$ dır.

En basit başarısızlık süresi modelidir ve anlık başarısızlık oranının t süresinden bağımsız olduğunu varsaymaktadır. Bu model altında, T'nin yaşam fonksiyonu ve o.y.f.'si sırasıyla aşağıdaki gibidir:

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

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

Ayrıca, belirtilen uzunlukta bir zaman aralığı içinde koşullu başarısızlık olasılığının, birimin ne kadar süre gözlemlendiğine bakılmaksızın aynı olduğu kolayca görülebilmektedir. Bu özellik, üstel modelin hafızasızlık özelliği olarak adlandırılır.

X'in açıklayıcı değişkenlerin bir vektörü olduğunu ve X'lerden birinin T üzerinde etkisi olduğunu varsayalım. Üstel bir regresyon modelini tanımlamanın bir yolu, X bilindiğinde T'nin koşullu tehlike fonksiyonunun,

$$h(t; X) = \lambda \exp(X'\beta)$$

ile OT modeline benzer bir biçimde verilmesidir. Burada β , regresyon parametrelerinin vektörünü göstermektedir. T'nin koşullu o.y.f.'si ise

$$f(t; X) = \lambda \exp(X'\beta) \exp[-\lambda t \exp(X'\beta)]$$

biçimindedir. $\log$ (yaşam süresi), $Y = \log T$ ile tanımlanmaktadır. Buna göre, yukarıdaki model,

$$Y = \alpha - X'\beta + W \tag{2.8}$$

ile eşdeğerdir. Burada $\alpha = -\log \lambda$ ve W , $\exp(w - e^w)$, $-\infty < w < \infty$ ile verilen o.y.f. ile uç değer dağılımına sahiptir. T ile ilgili olarak, Eşitlik 2.8'deki model log-doğrusal bir modeldir ve Y için uç değer dağılımına sahip W, hata değişkenli doğrusal bir modeldir (Sun, 2006).

2.6.2. Weibull Yaşam Modeli

Üstel yaşam modeli sadece bir parametreye bağlıdır ve bazen çok kısıtlayıcı olabilmektedir. Bu durumda, iki parametrelili Weibull yaşam modeli iyi bir alternatif olabilmektedir. Modelin tehlike fonksiyonu aşağıdaki gibidir:

$$h(t) = \lambda \gamma (\lambda t)^{\gamma-1}, \quad \lambda, \gamma > 0 \text{ için.}$$

Tehlike fonksiyonu $\gamma < 1$ için monoton azalan, $\gamma > 1$ için monoton artan ve $\gamma = 1$ için üstel azalmaktadır.

Weibull yaşam modelinde, T'nin yaşam fonksiyonu ve o.y.f.'si sırasıyla

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

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

biçimindedir. Regresyon çözümlemesi için, üstel yaşam modelinde olduğu gibi, tehlike fonksiyonu aşağıdaki gibi genelleştirilebilir:

$$h(t; X) = \lambda \gamma (\lambda t)^{\gamma-1} \exp(X'\beta).$$

T için koşullu o.y.f.,

$$f(t; X) = \lambda \gamma (\lambda t)^{\gamma-1} \exp(X'\beta) \exp[-(\lambda t)^\gamma \exp(X'\beta)]$$

olur. Üstel yaşam modelinde olduğu gibi, $Y = \log T$ ile bu regresyon modeli,

$$Y = \alpha + X'\beta^* + \sigma W, \quad (2.9)$$

biçiminde yazılabilir. Burada

$\alpha = -\log \lambda$, $\sigma = \gamma^{-1}$, $\beta^* = -\sigma\beta$ ve W uç değer dağılımını ifade etmektedir.

OT modeli olarak, hem üstel hem de Weibull yaşam modellerinin değişkenlerinin tehlike fonksiyonu üzerinde çarpımsal etkilerinin olduğunu belirtmektedir. Öte yandan, HBS modeli gibi, her iki model de log-doğrusal modeldir ve açıklayıcı değişkenler log yaşam süresini etkiler. Weibull yaşam modeli bu koşulları sağlayan tek model ailesidir (Kalbfleisch ve Prentice, 2002).

2.6.3. Log-Normal Yaşam Modeli

Log-normal yaşam modeli, $Y = \log T$ 'nin $Y = \alpha + \sigma W$ biçimine sahip olduğunu ve W 'nin standart bir normal değişken olduğunu varsaymaktadır. Buna göre, T 'nin o.y.f.'si

$$f(t) = (2\pi)^{-1/2} \gamma t^{-1} \exp \left[\frac{-\gamma^2 (\log \lambda t)^2}{2} \right]$$

biçimindedir. Burada,

$\lambda = \exp(-\alpha)$ ve $\gamma = \sigma^{-1}$ olur. T 'nin yaşam ve tehlike fonksiyonları $S(t) = 1 - \Phi(\gamma \log \lambda t)$ ile standart normal dağılım fonksiyonunu $\Phi(w)$ içerir ve her ikisinin de kapalı biçimi yoktur.

Tehlike fonksiyonu $t=0$ 'da sıfırdan maksimuma kadar artar ve sonra t arttıkça sıfıra düşer. X açıklayıcı değişkenlerinin olduğu durumlarda, log-normal regresyon modeli aşağıdaki genel doğrusal regresyon modeli gibi tanımlanabilir:

$$Y = \alpha + X'\beta + \sigma W. \quad (2.10)$$

Bu model, durdurma yoksa özellikle uygulanabilir. Ancak durdurma ile hesaplama ve çıkarım zorlaşır (Sun, 2006).

2.6.4. Log-Lojistik Yaşam Modeli

Log-lojistik yaşam modeli, W lojistik yoğunluğa sahip olması dışında log-normal yaşam modeli ile aynı biçimde tanımlanmaktadır:

$$\frac{e^w}{(1+e^w)^2}.$$

Bu yoğunluk fonksiyonu, normal yoğunluk fonksiyonundan biraz daha ağır kuyruklara sahip, 0 ortalama ve $\pi^2 / 3$ varyans ile simetriktir. Log-lojistik yaşam modeli altında T, aşağıda verilen o.y.f.'ye sahiptir:

$$f(t) = \lambda \gamma (\lambda t)^{\gamma-1} [1 + (\lambda t)^\gamma]^{-2}.$$

Burada $\lambda = \exp(-\alpha)$ ve $\gamma = \sigma^{-1}$ dir. Log-normal yaşam modeli ile karşılaştırıldığında log-lojistik yaşam modeli, başarısızlık süresi çözümlemesinde daha az kullanılmasına rağmen, hem yaşam hem de tehlike fonksiyonlarının kapalı biçimlerine sahip olma avantajına sahiptir. Dolayısıyla durdurma işlemlerinde log-normal yaşam modelinden daha uygundur. Yaşam fonksiyonu,

$$S(t) = \frac{1}{1 + (\lambda t)^\gamma}$$

ve tehlike fonksiyonu

$$h(t) = \frac{\lambda \gamma (\lambda t)^{\gamma-1}}{1 + (\lambda t)^\gamma}.$$

biçimindedir. Log-normal tehlike fonksiyonunun en önemli özelliği, önce artan sonra azalan tehlike fonksiyonu yapısına izin vermesidir.

2.6.5. Gamma Yaşam Modeli

Yaşam çözümlemesinde, standart (2 parametrelili) ve genelleştirilmiş (3 parametrelili) olmak üzere iki farklı Gamma modeli vardır. Genelleştirilmiş Gamma dağılımına ait o.y.f.,

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

biçimindedir. Burada dağılımın biçim parametresi γ ile ifade edilmiştir. Genelleştirilmiş Gamma dağılımı için yaşam ve tehlike fonksiyonlarının kapalı biçimi bulunmamaktadır. Genelleştirilmiş Gamma yaşam modelinin özel durumları üstel, Weibull ve log-normal yaşam modelleridir. Genelleştirilmiş Gamma dağılımı $\gamma \rightarrow \infty$ ile log-normal dağılıma, $\gamma = 1$ ile Weibull dağılımına ve $\gamma = \alpha = 1$ ile üstel dağılıma dönüşmektedir (Qi, 2009; Karasoy ve Sezayi, 2014).

2.6.6. Parametrik Yaşam Modelleri İçin Model Seçim Kriterleri ve Uyum İyiliğinin Test Edilmesi

n gözlem sayısını, L olabilirlik fonksiyonunu ve p parametre sayısını göstermek üzere, parametrik yaşam modelleri için hesaplanan AIC ve BIC değerleri, Eşitlik 2.11 ve Eşitlik 2.12'deki gibi ifade edilmektedir:

$$AIC = -2\log L + 2(p + 2 + k), \quad (2.11)$$

$$BIC = -2\log L + (p + 2 + k)\log(n). \quad (2.12)$$

k=0 üstel yaşam modeli için, k=1 Weibull, log-lojistik, log-normal yaşam modelleri için ve k=2 genelleştirilmiş Gamma yaşam modeli için kullanılmaktadır (Klein and Moeschberger, 1997).

Parametrik yaşam modellerinin uyum iyiliği için kullanılan Cox-Snell artıkları

$$r_i = -\log G(\hat{\epsilon}_i) = -\log G\left(\frac{\log t_i - \mu_i}{\hat{\sigma}}\right), i = 1, 2, \dots, n \quad (2.13)$$

biçimindedir. r_i 'ler için ($i = 1, 2, \dots, n$) yaşam fonksiyonu tahmini $S(\hat{r})$ olmak üzere önerilen parametrik yaşam modeli veri için uygunsa r_i 'nin $-\log \hat{S}(r_i)$ 'ye karşı grafiğinin, kesişim noktası sıfır ve eğimi bir olan düz bir doğru olması beklenir (Lee and Wang, 2003).

3. ARALIKLI DURDURULMUŞ VERİLER İLE YAŞAM ÇÖZÜMLEMESİ

3.1. Giriş

Yaşam süresi verilerinin çözümlenmesi için özel yöntemlere ihtiyaç duyulmasının birkaç nedeni vardır. Yaşam süresi verilerinin en önemli özelliği durdurulmuş birimlerin varlığıdır ve özellikle sağdan durdurulmuş birimleri içeren birçok çalışma vardır. Bu bölümde, sağdan durdurulmuş veri yapısına göre teorisi daha karmaşık olan aralıklı durdurulmuş veri yapısı ele alınacaktır. Sağdan durdurulmuş veri yapısı için kullanılan yöntemler genel olarak bu veriler için geçerli değildir.

Aralıklı durdurma ile ilgilenilen birimlerin veya yaşam süresi süreçlerinin sürekli gözlem altında olmadığı kastedilmektedir. Sonuç olarak, başarısızlık ya da yaşam süresi her zaman tam olarak gözlemlenemez ya da sağdan durdurulamaz. Aralıklı durdurulmuş bir birim için, sadece olayın gerçekleştiği bir aralık bilinmektedir. Kesin veya sağdan durdurulmuş başarısızlık süreleri olduğunda, bu gibi durumlar, aralıklı durdurulmuş başarısızlık süresinin özel bir durumu olarak kabul edilebilir, dolayısıyla aralık tek bir noktaya indirgenir veya sağdan sınırsız olur. Daha genel olarak, aralıklı durdurulmuş bir birim, birbiriyle çakışmayan birkaç aralığın birleşimi olarak tanımlanabilir (Turnbull, 1976).

Aralıklı durdurulmuş başarısızlık süresi verileri demografik, epidemiyolojik, finansal, tıbbi ve sosyolojik çalışmalar dahil olmak üzere birçok alanda ortaya çıkmaktadır. Periyodik olarak takip edilen tıbbi veya sağlık çalışmalarında tipik olarak durdurulmuş verilerin bir örneği vardır ve birçok klinik çalışma bu kategoriye girmektedir. Bu gibi durumlarda, aralıklı durdurulmuş veri yapısı çeşitli biçimlerde ortaya çıkabilmektedir. Örneğin, bir birey hastalık durumundaki olası değişiklikleri klinik olarak gözlemlmek ve daha sonra değişen bir durumla geri dönmek üzere programlanmış bir veya daha fazla gözlem zamanını kaçırabilir. Alternatif olarak, bireyler önceden belirlenmiş gözlem zamanlarından ziyade kendilerine uygun zamanlarda klinik merkezleri ziyaret edebilir. Her iki durumda da, durumdaki değişiklik ile ilgili veriler aralıklı olarak elde edilmektedir. Tüm birimler önceden belirlenmiş gözlem planını tam olarak takip etse bile sürekli bir değişken olduğu varsayılarak, durum değişikliğinin tam olarak gerçekleştiği süre net olarak

bilinemez. Son durumda, başarısızlık süresi verileri bir gruptur, yani gözlemlenen her birim için aralıklı durdurulmuş veri yapısı, çakışmayan aralıkların toplamıdır. Gruplanmış başarısızlık süresi verileri nispeten daha kolay bir biçimde ele alınabilir.

Literatürde aralıklı durdurma özelliği gösteren veri kümelerine fazla rastlanılamamaktadır. Bu durum veri kayıt sistemi ile ilgili eksikliklerden kaynaklanmaktadır. Literatürde yer alan ve gruplanmış yaşam süresi verileri olmayan bazı örnekler Hoel ve Walburg (1972), Carvalho ve diğ. (1989), Kim ve diğ. (1993), Betensky ve Finkelstein (1999), Goggins ve Finkelstein (2000), Meyns ve diğ. (2005), Goethals ve diğ. (2009), Hough (2010), Karvanen ve diğ. (2014) çalışmalarında yer almaktadır.

Aralıklı durdurulmuş veri yapısını açıklayabilmek için Çizelge 3.1'de verilen Finkelstein ve Wolfe'nin (1985) kullandığı meme kanseri verisi örnek olarak ele alınabilir. Bu veri kümesi, Boston'daki Ortak Radyasyon Tedavisi Merkezinde 1976-1980 yılları arasında tedavi edilen 96 hastaya aittir. 46 hasta sadece radyasyon tedavisi, 48 hasta ise radyasyon ile birlikte adjuvan kemoterapi tedavisi almıştır. Tedavi gören hastalar belli aralıklarda kontrole tabi tutulmuştur. Aralıklar, meme retraksiyonunun gerçekleştiği süreyi temsil etmektedir. Her tedavi grubundaki meme kanseri hastalarının meme retraksiyonu süresi için aylar içerisinde gözlemlenen aralıklar Çizelge 3.1'de yer almaktadır:

Çizelge 3.1. Meme kanseri veri kümesi

Sadece radyasyon tedavisi alan hastalar (n=46)
(0, 5], (0, 7], (0, 8], (4, 11], (5, 11], (5, 12], (6, 10], (7, 14], (7, 16], (11, 15], (11,18], (17, 25], (17, 25], (19, 26], (19, 35], (25, 37], (26, 40], (27, 34], (36, 44], (36, 48], (37, 44], (15,∞), (17,∞), (18,∞), (22,∞), (24,∞), (24,∞), (32,∞), (33,∞), (34,∞), (36,∞), (36,∞), (37,∞), (37, ∞), (37,∞), (38,∞), (40,∞), (45,∞), (46,∞), (46,∞), (46,∞), (46,∞), (46,∞), (46,∞), (46,∞)
Radyasyon ile birlikte adjuvan kemoterapi tedavisi alan hastalar (n=48)
(0, 5], (0, 22], (4, 8], (4, 9], (5, 8], (8, 12], (8, 21], (10, 17], (10, 35], (11, 13], (11,17], (11, 20], (12, 20], (13, 39], (14, 17], (14, 19], (15, 22], (16, 20], (16, 24], (16, 24], (16, 60], (17, 23], (17, 26], (17, 27], (18, 24], (18, 25], (19, 32], (22, 32], (24, 30], (24, 31], (30, 34], (30, 36], (33, 40], (35, 39], (44, 48], (48,∞), (11,∞), (11,∞), (13,∞), (13,∞), (13,∞), (21,∞), (23,∞), (31,∞), (32,∞), (34,∞), (34,∞), (35,∞)

Erken meme kanseri teşhisi konulan hastalarla ilgili yapılan bu çalışmada, tedavilerin kozmetik etkilerini görebilmek için radyasyon ile birlikte adjuvan kemoterapi tedavisi alan hastalar ve sadece radyoterapi tedavisi alan hastalar karşılaştırılmıştır (Beadle ve diğ., 1984). Çalışmanın amacı, iki tedavi yönteminde kozmetik bozulmaya kadar geçen sürenin karşılaştırılmasıdır. Çizelge 3.1’de yaşam süresi (a, b] biçiminde olan hastalar başarısız, (a, ∞) biçiminde olan hastalar ise durdurulmuş gözlemleri ifade etmektedir. Örneğin, Çizelge 3.1’de verilen (8, 12] değeri, 8. aya kadar hastada hiçbir kozmetik bozulma olmadığını, ancak 12. ayda meme retraksiyonunun gözlemlendiğini göstermektedir. Dolayısı ile kozmetik bozulmanın kesin olarak hangi zamanda gerçekleştiği bilinmemekte, ancak 8. ay ile 12. ay arasında olduğu bilinmektedir. Bu durum, aralıklı durdurulmuş veri yapısının ortaya çıkmasına neden olmaktadır. Hastaya ait yaşam süresi sağdan durdurulmuş veri yapısına uygun olarak ele alınsaydı, yaşam süresinin (0, 12] biçiminde ifade edilmesi ve 12 ay olarak değerlendirilmesi gerekirdi.

3.2. Aralıklı Durdurulmuş Veri Yapısı

Bir yaşam çözümlemesi çalışmasında bir birimin başarısızlık süresini gösteren T negatif olmayan bir rasgele değişken olsun. Eğer T tam olarak gözlemlenmez, L yaşam süresinin sol uç noktası ve R yaşam süresinin sağ uç noktası olmak üzere sadece $(L, R]$ aralığında gözlemlenirse, T yaşam süresine sahip bir birim aralıklı durdurulmuştur denir ve $L \leq R$ olacak biçimde

$$T \in (L, R] \quad (3.1)$$

ile ifade edilir. Burada, $L = R$ kesin bir birim ve $R = \infty$ sağdan durdurulmuş bir birim olduğunu belirtir.

Durum I Aralıklı Durdurulmuş Veri Yapısı

“Durum I aralıklı durdurulmuş yaşam süresi verileri” terimi, tüm gözlemlenen aralıklarda ya sıfır zamanını ya da sonsuz zamanı içeren aralıklı durdurulmuş yaşam süresi verilerinin belirtilmesi için kullanılmaktadır (Groeneboom ve Wellner, 1992; Huang, 1996). Diğer bir deyişle, her bir birim ya soldan ya da sağdan durdurulmuştur, yani ya $L = 0$ ya da $R = \infty$ 'dur. Durum I aralıklı durdurulmuş veri yapısı, her birim sadece bir kez gözlemlendiğinde oluşur ve ilgilenilen olay için gözlemlenen tek bilgi, olayın gözlem süresinden daha önce gerçekleşip gerçekleşmediğidir. Eşitlik 3.1'deki aralıklar yerine, Durum I aralıklı durdurulmuş veri yapısı için daha uygun bir gösterim, C gözlem süresi ve I gösterge fonksiyonu olmak üzere, $\{C, \delta = I(T = C)\}$ biçimindedir.

Durum I aralıklı durdurulmuş veri yapısı, genellikle demografik çalışmalardan ortaya çıkmış bir terim olan “mevcut durum verileri (current status data)” olarak da adlandırılır. Örneğin, öldürücü nitelikte olmayan tümörler üzerinde yapılan kesitsel çalışmalar ve tümör oluşturma deneyleri sıklıkla Durum I aralıklı durdurulmuş veri yapısı üreten çalışmalardır. Birincisi, demografik çalışmalarda yaygın olarak kullanılmaktadır. Aynı biçimde çözümlenmelerine rağmen, bu iki çalışma türünden ortaya çıkan mevcut durum verileri arasında temel bir fark vardır. Birinci örnekteki mevcut durum verileri esas olarak çalışma tasarımlarından kaynaklanmakta, buna

karşılık ikinci örnekteki mevcut durum verileri genellikle doğrudan ve/veya tam olarak değişkenin ölçülememesinden kaynaklanmaktadır.

Durum II Aralıklı Durdurulmuş Veri Yapısı

Hem L hem de $(0, \infty)$ olan en az bir aralık içeren aralıklı durdurulmuş veri yapısı, genel veya Durum II aralıklı durdurulmuş veri yapısı olarak adlandırılmaktadır (Groeneboom ve Wellner, 1992; Huang ve Wellner, 1997; Sun, 1998, 2005). Başka bir deyişle, Durum II aralıklı durdurulmuş veri yapısı, sıfırdan farklı bazı sonlu aralıklar içeren aralıklı durdurulmuş veri yapısıdır. Durum II aralıklı durdurulmuş bir birimi göstermenin başka bir yolu, her bir birimin iki kez gözlemlendiğini varsayan, U ve V raslantı değişkenleri olmak üzere, $U \leq V$ 'yi 1 olasılığı ile sağlayan aşağıdaki gösterimi kullanmaktır:

$$\{U, V, \delta_1 = I(T \leq U), \delta_2 = I(U < T \leq V), \delta_3 = 1 - \delta_1 - \delta_2\} \quad (3.2)$$

Bu gösterim teorik çıkarımlarda sıklıkla kullanılır. $U = V = C$ alarak, Durum I aralıklı durdurulmuş veri yapısı Eşitlik 3.2 ile de tanımlanabilmektedir. Yu ve diğ. (2000) bu formülü kesin (yaşam süreleri belli olan) birimleri içerecek biçimde genelleştirmiştir.

Eşitlik 3.2'nin diğer bir genelleştirilmiş biçimi, her bir birim için K rasgele bir tamsayı olmak üzere $U_1 \leq U_2 \leq \dots \leq U_K$ olan gözlem süreleri noktalarına sahip bir kümenin olduğunu varsaymaktır. Gözlemlenen bilgiler $U_0 = 0$ olmak üzere aşağıdaki biçimdedir:

$$\{(K, U_j, \delta_j = I(U_{j-1} < T \leq U_j)), j = 1, \dots, K\}. \quad (3.3)$$

Bu formül ile tanımlanan yaşam süresi verisi genellikle K durumu ya da karma durum aralıklı durdurulmuş veri (mixed case interval censored data) olarak adlandırılmaktadır (Schick ve Yu, 2000; Wellner, 1995). Eşitlik 3.3, Eşitlik 3.2'nin özel bir durumudur ve periyodik gözlemlenmeye sahip boylamsal çalışmalardan

ortaya çıkan aralıklı durdurulmuş başarısızlık süresi verilerinin doğal bir gösterimidir.

Olabilirlik fonksiyonu, Eşitlik 3.1, Eşitlik 3.2, Eşitlik 3.3 ile verilen her üç gösterim için de aynıdır. Eşitlik 3.2 ve Eşitlik 3.3'teki gösterimleri doğal görünse de, uygulamada bu biçimde toplanan veya verilen aralıklı durdurulmuş veri yapılarının bulunması çok yaygın değildir. Bununla birlikte, teorik çıkarsamaların asimptotik özelliklerinin türetilmesi için sıklıkla ihtiyaç duyulan T'nin bağımsızlığı gibi bazı varsayımları koymak Eşitlik 3.1'deki gösterimine göre çok daha kolay ve doğaldır.

Veriler Eşitlik 3.2 veya Eşitlik 3.3'teki gösterimlerindeki gibi verilir ise, Eşitlik 3.1'deki gibi bir gösterimin elde edilmesi kolaydır. Bununla birlikte, gözlemlene süreci hakkında ek bilgi olmaksızın Eşitlik 3.1'i Eşitlik 3.3'e dönüştürmek imkansızdır ve Eşitlik 3.1'de verilen gözlemleri Eşitlik 3.2'ye dönüştürmek de kolay değildir.

3.3. Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli

3.3.1. Durum I Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli

Durum I aralıklı durdurulmuş veriler hayvan kanserojenliği deneyleri, demografik çalışmalar, ekonometri, epidemiyolojik çalışmalar ve güvenilirlik çalışmaları gibi birçok alanda ortaya çıkmaktadır. Gizli tümörlerde kanserojenlik denemeleri gibi bazı durumlarda Durum I aralıklı durdurulmuş veriler, tümörün başlangıç süresi gibi ilgilenilen değişkenler hakkında mevcut olan tek bilgidir (Dinse ve Lagakos, 1983). Yani, değişkenler doğrudan ölçülemez. Bir dönüm noktası olayı üzerindeki kesitsel çalışmalardan kaynaklananlar gibi bazı durumlarda Durum I aralıklı durdurulmuş veriler, olaya ilişkin zamanla ilgili kesin bilgileri veren verilere göre daha kolay ve daha güvenilir bilgiler sağlar. Bu gibi durumlara bir örnek, ilgilenilen olayın kronik hastalığın başlangıcı olduğu epidemiyolojik çalışmalardır (Keiding, 1991; Keiding ve diğ., 1996; Shiboski ve Jewell, 1992). Bir başka örnek ise söz konusu olayın örneğin ilk hamilelik veya evlilik olabileceği demografik çalışmalar olabilmektedir (Diamond ve diğ., 1986; Diamond ve McDonald, 1991).

Yarı parametrik bir yaşam modeli olan CRM incelenirken en sık kullanılan yaklaşım yarı parametrik EÇÖ tahminidir. Bu olabilirlik yaklaşımı oldukça basittir,

ancak kullanımı kolay değildir, çünkü olabilirlik fonksiyonu, sonlu boyutlu regresyon parametrelerinin ve sonsuz boyutlu bir nüans parametresinin birikimli temel tehlike fonksiyonu ya da temel yaşam fonksiyonunun bir fonksiyonudur. Sonuç olarak, regresyon parametrelerini ve nüans parametresini aynı anda tahmin etmek gerekir. Bu da kısmi olabilirlik fonksiyonunun uygulandığı CRM'yi kullanan sağdan durdurulmuş veri yapısının regresyon çözümlemesinden farklıdır. İkinci yaklaşım olarak regresyon parametrelerine ilişkin çıkarsama için, nüans parametresini içermeyen ve özellikleri martingale teorisi ile kolayca elde edilebilen bir kısmi olabilirlik fonksiyonu elde edilebilir. Fakat, Durum I aralıklı durdurulmuş veriler için kısmi olabilirlik yaklaşımı mevcut değildir, bu sebeple tam olabilirlik fonksiyonunun kullanılması gerekmektedir.

Sonsuz boyutlu bir nüans parametresini içeren olabilirlik fonksiyonu ile uğraşmamak için, uygulamada genellikle sieve EÇO yöntemi kullanılır. Bu yöntemde, sonsuz boyutlu nüans parametresi sonlu boyutlu parametreler dizisine yaklaştırılmaktadır; yani, orijinal parametre uzayı, artan sonlu boyutlu alt uzayların (sieve'lerin) bir dizisi ile yaklaştırılır. Burada ele alınan bir problem için, yarı parametrik yaşam modelinin regresyon parametresinin β ve birikimli temel tehlike fonksiyonunun $H_0(t)$ ile tanımlandığını varsayalım. Herhangi sonlu boyutlu bir örneklem için β ve $H_0(t)$ tahminleri, β ve sieve'ler için parametre uzaylarının çarpma üzerindeki olabilirlik fonksiyonunu maksimum yapmak suretiyle bulunabilir. Başka bir deyişle, sieve yöntemiyle sadece sonlu boyutlu bir parametre uzayı ile çalışmak gerekir. Sieve yönteminin başka bir avantajı, birikimli temel tehlike fonksiyonu $H_0(t)$ 'nin tahmin edicisinin, orijinal parametre uzayı üzerindeki tüm olabilirliği maksimize ederek verilen tahmin ediciden daha hızlı bir yakınsama oranına sahip olmasıdır (Huang ve Rossini, 1997).

Sınırsız boyutlu bir nüans parametresiyle uğraşmaktan kaçınabilecek başka bir yaklaşım, β 'nin temel tahminini bazı tahmin denklemlerine dayandırmaktır. Burada β ile ilgili bazı tahmin denklemlerinin yansızlık gibi iyi özelliklere sahip olduğu ve nüans parametresini içermediği varsayılmaktadır. Bu yaklaşımın en büyük avantajı, tam ya da sieve EÇO yaklaşımına göre çok daha kolay bir şekilde

uygulanabilmesidir. Ayrıca, bu yaklaşımdan elde edilen regresyon parametreleri tahminlerinin özellikleri nispeten daha kolay bir biçimde tespit edilebilmektedir.

n bağımsız birimden oluşan bir yaşam çözümlemesi çalışması ve bu çalışmada, i. birim için iki rasgele değişkenin olduğunu varsayalım. $i = 1, \dots, n$ olmak üzere, T_i yaşam süresini, C_i birimler ile ilgili gözlem süresini ve X_i açıklayıcı değişkenler vektörünü ifade etmektedir. T_i 'lerin dağılımı, regresyon parametresi β ve temel birikimli tehlike fonksiyonu $H_0(t)$ veya temel yaşam fonksiyonu $S_0(t) = \exp(-H_0(t))$ ile belirlenir. Ayrıca, β ve $H_0(t)$ veya $S_0(t)$ hakkında çıkarım yapmak için, sadece Durum I aralıklı durdurulmuş verilerin olduğu ve

$$\{(C_i, \delta_i = I(T_i \leq C_i), X_i); i = 1, \dots, n\}$$

biçiminde ifade edildiği varsayımıyla, her birimin sadece bir kez C_i 'de gözlemlendiği ve C_i 'de sadece ilgilenilen yaşam olayının C_i 'den önce mi yoksa C_i 'de mi gerçekleştiği bilinir.

Durum I aralıklı durdurulmuş veriler, yaşam süresi verilerinin çözümlenmesinde en sık kullanılan regresyon modeli olan CRM ile analiz edilebilmektedir. X_i , T_i ve C_i 'nin bağımsız olduğu varsayımına dayanarak birikimli tehlike fonksiyonu

$$H(t; X_i) = H_0(t) \exp(X_i' \beta)$$

ve olabilirlik fonksiyonu

$$L(\beta, H_0) = \prod_{i=1}^n \exp \left[-(1 - \delta_i) e^{X_i' \beta} H_0(C_i) \right] \left[1 - \exp \left(-e^{X_i' \beta} H_0(C_i) \right) \right]^{\delta_i} \quad (3.4)$$

olarak ifade edilir. Temel yaşam fonksiyonu S_0 ve regresyon parametresi β ile olabilirlik fonksiyonu Eşitlik 3.5'teki gibidir:

$$L(\beta, S_0) = \prod_{i=1}^n [S_0(C_i)]^{(1-\delta_i) \exp(X_i' \beta)} \left\{ 1 - [S_0(C_i)]^{\exp(X_i' \beta)} \right\}^{\delta_i} \quad (3.5)$$

β ve S_0 için EÇO Tahmini

β ve $S_0(t)$ 'nin tahmini için EÇO yaklaşımı, Eşitlik 3.5'te verilen olabilirlik fonksiyonunu, yani $L(\beta, S_0)$ 'ı maksimize etmektedir. Durum I aralıklı durdurulmuş veri kümesi için, C_i 'nin gözlem sürelerindeki $S_0(t)$ değerleri olabilirlik fonksiyonunu etkilemektedir. Bu nedenle, genel kayıp olmadan, sadece $S_0(t)$ için C_i 'lerin atlamaları ile tüm artan adım fonksiyonlarında sadece $L(\beta, S_0)$ değerinin maksimize edilmesine odaklanılabilir.

$0 < s_1 < \dots < s_m$, $\{C_i\}_{i=1}^n$ 'in farklı zaman noktalarının bir sıralaması ve Ω_S

$$S_0(t) = \prod_{j: s_j \leq t} e^{-\exp(\alpha_j)} \quad (3.6)$$

ile verilen tüm temel yaşam fonksiyonlarının bir kümesidir. Burada $\alpha = (\alpha_1, \dots, \alpha_m)'$ bilinmeyen parametrelerdir. β ve S_0 'ın EÇO tahmin edicilerini belirlemek için Ω_S 'de β ve S_0 'a göre $L(\beta, S_0)$ 'ın maksimize edilmesi gerekmektedir. Bu durumda, log olabilirlik fonksiyonu

$$l(\beta, \alpha) = \sum_{i=1}^n \left\{ \delta_i \log \left[1 - \prod_{j_i} e^{-\exp(\alpha_j + X_i' \beta)} \right] - \sum_{j_i} (1 - \delta_i) e^{\alpha_j + X_i' \beta} \right\}$$

olarak yazılabilir. β ve α 'nın $\prod_{j_i}$ ve $\sum_{j_i}$ 'nin sırasıyla sonucu $\{j; s_j \leq C_i\}$ üzerinde gösterdiği anlamına gelmektedir.

$C_i = s_j$ ve $\delta_i = 1$ olan birimler için indis kümesi D_j ve $C_i = s_j$, $j = 1, 2, \dots, m$ olan birimler için indis kümesi R_j ile gösterilsin. $a_j = \sum_{k=1}^j \exp(\alpha_k)$, $j = 1, 2, \dots, m$ varsayımıyla log olabilirlik fonksiyonu

$$l(\beta, \alpha) = \sum_{j=1}^m \left\{ \sum_{i \in D_j} \log \left[\frac{1 - e^{a_j \exp(X_i' \beta)}}{e^{a_j \exp(X_i' \beta)}} \right] - a_j \sum_{i \in R_j} e^{X_i' \beta} \right\} \quad (3.7)$$

biçiminde yeniden yazılabilir. $l(\beta, \alpha)$ 'yı maksimize etmek için genel bir yaklaşım Newton-Raphson algoritmasını kullanmaktır ve bunun için aşağıda verilen $l(\beta, \alpha)$ 'nın birinci ve ikinci türevlerine ihtiyaç vardır:

$$\frac{\partial l(\beta, \alpha)}{\partial \beta} = \sum_{j=1}^m a_j \left\{ \sum_{i \in D_j} X_i e^{Z_i' \beta} [q(a_j, X_i) + 1] - \sum_{i \in R_j} X_i e^{Z_i' \beta} \right\},$$

$$\frac{\partial l(\beta, \alpha)}{\partial \alpha_j} = e^{\alpha_j} \sum_{k=j}^m \left\{ \sum_{i \in D_k} e^{X_i' \beta} [q(a_k, X_i) + 1] - \sum_{i \in R_k} e^{X_i' \beta} \right\},$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \beta \partial \beta'} = \sum_{j=1}^m a_j \left\{ \sum_{i \in D_j} X_i X_i' e^{X_i' \beta} [q(a_j, X_i) + 1 - a_j e^{X_i' \beta} q(a_j, X_i) [q(a_j, X_i) + 1]] - \sum_{i \in R_j} X_i X_i' e^{X_i' \beta} \right\},$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_j \partial \beta} = e^{\alpha_j} \sum_{k=j}^m \left\{ \sum_{i \in D_k} X_i e^{X_i' \beta} [q(a_k, X_i) + 1 - a_j e^{X_i' \beta} q(a_k, X_i) [q(a_k, X_i) + 1]] - \sum_{i \in R_k} X_i e^{X_i' \beta} \right\},$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_j^2} = \frac{\partial l(\beta, \alpha)}{\partial \alpha_j} - e^{2\alpha_j} \sum_{k=j}^m \left\{ \sum_{i \in D_k} e^{2X_i' \beta} [q(a_k, X_i) + 1] q(a_k, X_i) \right\}$$

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_j \partial \alpha_k} = -e^{\alpha_j + \alpha_k} \sum_{l=k}^m \sum_{i \in D_l} e^{2X_i' \beta} q(a_l, X_i) [q(a_l, X_i) + 1].$$

Burada, $j = 1, \dots, m$; $i = 1, \dots, n$ olmak üzere $q(a_j, X_i) = \frac{e^{-a_j \exp(X_i' \beta)}}{1 - e^{-a_j \exp(X_i' \beta)}}$ 'dır.

Newton-Raphson algoritmasını uygulamak için, β ve α 'nın bazı başlangıç tahminlerinin ve bir yakınsama kriterinin seçilmesi ve bir $(p+m) \times (p+m)$ matrisinin tersinin hesaplanması gerekir (Sun, 2006).

$\hat{\beta}_n$ ve $\hat{\alpha}_n$ 'nin yukarıda tanımlanan β ve α tahminlerini gösterdiği düşünölsün. Ayrıca, $\hat{S}_n(t)$ 'nin Eşitlik 3.6'da verilen temel yaşam fonksiyonunun tahmini α ile değıştirilmiş $\hat{\alpha}_n$ ile gösterilsin ve birikimli temel tehlike fonksiyonunun bir tahmini $\hat{H}_n(t) = -\log[\hat{S}_n(t)]$ olarak ele alınsın. Bu durumda, β hakkında çıkarımlar yapmak için, asimptotik dağılımın ve $\hat{\beta}_n$ 'nin varyans-kovaryans matrisinin bir tahminini bilmek gerekmektedir.

$\hat{\beta}_n$ 'nin asimptotik normalliğı Huang (1996) çalışmasında gösterilmiştir. Kovaryans tahmini için genel bir yaklaşım, $l(\beta, \alpha)$ 'nın β ve α 'ya göre parametrik bir olabilirlik fonksiyonu olarak ele alınmasıdır. Daha sonra, gözlemlenen Fisher bilgi matrisini veya $\hat{\beta}$ 'nin varyans-kovaryans matrisini tahmin etmek için β 'ya karşılık gelen $l(\beta, \alpha)$ negatif tersinin ikinci dereceden türevi alınmış matrisinin alt matrisi kullanılabilir. X_i 'nin 0 veya 1 değeri alması durumunda, aşağıda verilen Huang (1996)'ın çalışması alternatif olarak kullanılabilir.

C_i 'nin o.y.f.'si $X_i = 0$ veya $X_i = 1$ için sırasıyla $g_0(c)$ ve $g_1(c)$, düzleştirme (smooth) çekirdek tahmin edicileri ise sırasıyla $\hat{g}_0(c)$ ve $\hat{g}_1(c)$ ile gösterilsin. Verilen X değışkeni için

$$\hat{R}(c; X) = \frac{\exp[-e^{X\hat{\beta}_n} \hat{H}_n(c)]}{1 - \exp[-e^{X\hat{\beta}_n} \hat{H}_n(c)]} \hat{H}_n^2(c) e^{2X\hat{\beta}_n}$$

ve

$$\hat{\mu}(c) = \frac{\hat{R}(c; X=1)\hat{g}_1(c)n_1}{\hat{R}(c; X=1)\hat{g}_1(c)n_1 + \hat{R}(c; X=0)\hat{g}_0(c)(n-n_1)}$$

biçimindedir. Burada $n_1 = \sum_{i=1}^n X_i$, $X_i = 1$ olan birimlerin sayısını gösterir. $(n\hat{\sigma}_n^2)^{-1}$ ile

$\hat{\beta}_n$ 'in varyansının tutarlı bir tahmini Eşitlik 3.8'deki gibi verilmektedir:

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \hat{R}(C_i; X_i) [X_i - \hat{\mu}(C_i)]^2 \right\} \quad (3.8)$$

Eşitlik 3.8'de verilen varyans tahmini, aslında β 'ya bağlı bilginin tutarlı bir tahminidir. Yani, EÇO tahmin edicisi $\hat{\beta}_n$ asimptotik olarak etkilidir. Eşitlik 3.8'deki tahminin dezavantajı, uygun bir bant genişliği ve çekirdek tahmininin seçilmesini gerektiren $g_0(c)$ ve $g_1(c)$ 'nin çekirdek tahminlerini elde etmesi gerektiğidir. $g_0(c) = g_1(c)$ ise C_i 'nin X_i 'den bağımsız olduğu söylenebilir. Bu durumda, $\hat{\mu}(c)$ ve dolayısıyla $\hat{\sigma}_n$ 'nin $g_0(c)$ ve $g_1(c)$ 'nin ortak tahminlerini içermediği görülebilir. Eğer C_i , X_i 'den bağımsız ise $\hat{\beta}_n$ 'nin varyans-kovaryans matrisi $(n\hat{\Sigma}_n)^{-1}$ ile aşağıdaki gibi kolay bir biçimde tahmin edilebilir (Huang, 1996):

$$\hat{\Sigma}_n \frac{1}{n} \sum_{i=1}^n \left\{ \hat{R}(C_i; X_i) \left[X_i - \frac{\sum_{j=1}^n X_j \hat{R}(C_j; X_j)}{\sum_{j=1}^n \hat{R}(C_j; X_j)} \right]^{\otimes 2} \right\}. \quad (3.9)$$

Newton-Raphson algoritması genel olarak iyi çalışmasına rağmen, hesaplamanın yoğun olduğu ve çok sayıda farklı gözlem süresi noktasına sahip olan bazı veri kümeleri için kararsız tahmin problemleri ile karşılaşılabilir. Alternatif olarak, belirli

bir veri kümesi için, Eşitlik 3.4'te verilen $L(\beta, H_0)$ olasılıkları, $L(\beta, S_0)$ yerine β ve H_0 üzerinden verilebilir. Bu durumda, ortaya çıkan log olabilirlik fonksiyonu,

$$L(\beta, H_0) = \sum_{i=1}^n \left\{ \delta_i \log \left[1 - \exp \left(-e^{X_i' \beta} H_0(C_i) \right) \right] - (1 - \delta_i) \exp(X_i' \beta) H_0(C_i) \right\}$$

biçimindedir.

$l(\beta, H_0)$ 'ın verilen β için H_0 'a göre içbükey olduğu ve bazı maksimizasyon prosedürlerinin geliştirilmesinde kullanılabileceği gösterilmiştir. Huang (1996) ve Huang ve Wellner (1997) bu durumu incelemiştir ve iki aşamalı bir dışbükey minorant algoritması önermiştir. Bununla birlikte, bu algoritmada benzer sayısal problemler bulunabilir. EÇO tahmin edicilerinin asimptotik özellikleri Huang (1996)'ın çalışmasında yer almaktadır.

3.3.2. Durum II Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli

Durum I aralıklı durdurulmuş veri yapısı ile karşılaştırıldığında, Durum II aralıklı durdurulmuş veri yapısının ilgilenilen temel yaşam süresi hakkında daha fazla bilgi sağladığı açıktır. Dolayısıyla sezgisel olarak, Durum II aralıklı durdurulmuş veri yapısının regresyon analizi, Durum I aralıklı durdurulmuş veri yapısına göre daha basit görünebilir. Diğer yandan, Durum II aralıklı durdurulmuş veri yapısı için, Durum I aralıklı durdurulmuş veri yapısında olduğu gibi, sadece bir değişken yerine gözlem sürelerini temsil eden iki veya daha fazla değişkenle ilgilenmek gerekmektedir. Durum II aralıklı durdurulmuş veri yapısının regresyon analizi, hem hesaplama hem de teoride Durum I aralıklı durdurulmuş veri yapısından daha karmaşıktır.

n tane bağımsız birimden oluşan ve aralıklı durdurulmuş veri yapılarının ortaya çıkmasına neden olan bir çalışma ele alınsın. İlgili yaşam süreleri için

$$\{(L_i, R_i], X_i; i = 1, \dots, n\} \quad (3.10)$$

tanımlansın. Bu durumda $(L_i, R_i]$, gözlemlenen i . birim için yaşam olayının gerçekleştiği aralığı belirtir ve X_i , i ($i=1, \dots, n$) biriminden gelen değişkenlerin p -boyutlu vektörünü temsil eder. $S(t; X)$, X değişkenleri için yaşam fonksiyonunu gösterebilir. Buna göre olabilirlik fonksiyonu

$$L = \prod_{i=1}^n [S(L_i, X_i) - S(R_i, X_i)]$$

ile orantılıdır. Burada $L_i < R_i$, $i = 1, \dots, n$ varsayımı yapılmaktadır.

$S(t; X)$ 'in CRM ile ifade edildiği varsayılmaktadır. Yukarıda verilen olabilirlik fonksiyonunun logaritması, regresyon parametresi β ve temel yaşam fonksiyonu S_0 olmak üzere

$$l(\beta, S_0) = \sum_{i=1}^n \log \left\{ [S_0(L_i)]^{\exp(X_i' \beta)} - [S_0(R_i)]^{\exp(X_i' \beta)} \right\}$$

biçimindedir.

β ve S_0 için Finkelstein (1986) tarafından incelenen EÇÖ tahmin edicileri ve bunların bazı asimptotik özellikleri aşağıda verilmiştir.

β ve S_0 için EÇÖ Tahmini

Olabilirlik fonksiyonu sadece farklı zaman noktalarındaki değerlere bağlıdır. Dolayısıyla, sadece bu zaman noktalarındaki S_0 'ın değerlerini tahmin etmeye odaklanılmalıdır. $\{L_i, R_i; i = 1, \dots, n\}$ ve $\alpha_{ij} = I(s_j \in (L_i, R_i])$, $j = 1, \dots, m$ $i = 1, \dots, n$ ile ifade edilen $s_0 = 0 < s_1 < \dots < s_{m+1} = \infty$, gözlemlenen tüm aralığın uç noktalarının sıralı farklı zaman noktalarını ifade eder. $S_0(s_j)$ 'nin

$$\prod_{k=1}^j e^{-\exp(\alpha_k)} = e^{-\sum_{k=1}^j \exp(\alpha_k)}, j = 1, \dots, m$$

olarak yazıldığını varsayalım. $\alpha_j = \sum_{k=0}^j \exp(\alpha_k)$, $\alpha_0 = -\infty$ ve $\alpha_{m+1} = \infty$ olmak üzere, β ve $\alpha = (\alpha_1, \dots, \alpha_m)'$ parametreleri ile log olabilirlik fonksiyonu $l(\beta, S_0)$ yeniden

$$l(\beta, \alpha) = \sum_{i=1}^n \log \left\{ \sum_{j=1}^{m+1} \alpha_{ij} \left[e^{-\alpha_{j-1} \exp(X_i' \beta)} - e^{-\alpha_j \exp(X_i' \beta)} \right] \right\}$$

biçiminde yazılabilir. Log-olabilirlik fonksiyonunu maksimum yapmak için, parametrik bir modelden oluşan log-olabilirlik fonksiyonu ele alınır ve Newton-Raphson algoritması kullanılır. Bunun için β ve α skor fonksiyonlarına ve gözlemlenen Fisher bilgi matrisine ihtiyaç vardır. Skor fonksiyonları $j = 1, \dots, m$, $i = 1, \dots, n$ olmak üzere,

$$U_{\beta}(\beta, \alpha) = \frac{\partial l(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^n X_i g_i^{-1} \sum_{j=1}^{m+1} \alpha_{ij} (f_{ij-1} - f_{ij})$$

ve

$$U_{\alpha_j}(\beta, \alpha) = \frac{\partial l(\beta, \alpha)}{\partial \alpha_j} = \sum_{i=1}^n g_i^{-1} b_{ij} c_{ij},$$

biçimindedir. Burada,

$$f_{ij} = S(s_j; X_i) \log S(s_j; X_i), f_{i0} = f_{im+1} = 0, b_{ij} = \exp(\alpha_j + X_i' \beta),$$

$$c_{ij} = \sum_{l=j}^{m+1} (\alpha_{il} - \alpha_{il+1}) S_{il}(s_l; X_i), \alpha_{im+2} = 0$$

ve

$$g_i = \sum_{j=1}^{m+1} \alpha_{ij} \left[S(s_{j-1}; X_i) - S(s_j; X_i) \right]$$

ile ifade edilmektedir. β ve α 'nın EÇÖ tahmin edicileri skor denklemleri çözülerek aşağıdaki gibi belirlenebilir:

$$U_{\beta}(\beta, \alpha) = 0,$$

$$U_{\alpha_j}(\beta, \alpha) = 0, \quad j = 1, \dots, m.$$

$$I_{11} = -\frac{\partial^2 I(\beta, \alpha)}{\partial \beta \partial \beta'}, \quad I_{12} = I_{21}' = -\left(\frac{\partial^2 I(\beta, \alpha)}{\partial \alpha_j \partial \beta} \right), \quad I_{22} = -\left(\frac{\partial^2 I(\beta, \alpha)}{\partial \alpha_j \partial \alpha_k} \right) \text{ olmak üzere,}$$

$$I(\beta, \alpha) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \text{ gözlemlenen Fisher bilgi matrisini gösterebilir. Buna göre,}$$

$$h_{ij} = f_{ij} \log S(s_j; X_i) + f_{ij} \text{ ve } h_{i0} = h_{im+1} = 0 \text{ olduğunda,}$$

$$I_{11} = \sum_{i=1}^n X_i X_i' \left\{ \left[\frac{\sum_{j=1}^{m+1} \alpha_{ij} (f_{ij-1} - f_{ij})}{g_i} \right]^2 - \frac{\sum_{j=1}^{m+1} \alpha_{ij} (h_{ij-1} - h_{ij})}{g_i} \right\},$$

$$\frac{\partial^2 I(\beta, \alpha)}{\partial \alpha_j \partial \beta} = \sum_{i=1}^n X_i b_{ij} \left[\frac{c_{ij} + \sum_{l=j}^{m+1} (\alpha_{il} - \alpha_{il+1}) S(s_l; X_i) \log S(s_l; X_i)}{g_i} - \frac{c_{ij}}{g_i^2} \sum_{l=1}^{m+1} \alpha_{il} (f_{il-1} - f_{il}) \right],$$

$$\frac{\partial^2 I(\beta, \alpha)}{\partial \alpha_j^2} = \sum_{i=1}^n b_{ij} c_{ij} \left[\frac{1 - b_{ij}}{g_i} - \frac{b_{ij} c_{ij}}{g_i^2} \right],$$

ve

$$\frac{\partial^2 l(\beta, \alpha)}{\partial \alpha_j \partial \alpha_k} = -\sum_{i=1}^n \left(\frac{b_{ij} b_{ik} c_{ij} c_{ik}}{g_i^2} + \frac{b_{ij} b_{ik} c_{ik}}{g_i} \right), j < k \text{ için}$$

elde edilir. Newton-Raphson algoritmasının uygulanması için, daha önce verilen açıklamalar geçerlidir.

$$I_{11|2} = I_{11} - I_{12} I_{22}^{-1} I_{21}, I_{12|2} = I_{21|1}' = -I_{11|2}^{-1} I_{12} I_{22}^{-1} \text{ ve } I_{22|1} = I_{22} + I_{22}^{-1} I_{21} I_{11|2}^{-1} I_{12} I_{22}^{-1} \text{ olmak üzere,}$$

$$I^{-1}(\beta, \alpha) = \begin{pmatrix} I_{11|2}^{-1} & I_{12|2} \\ I_{21|1} & I_{22|1} \end{pmatrix}$$

olarak elde edilir. EÇÖ tahmin edicilerinin asimptotik özellikleri Huang ve Wellner'in (1997) çalışmasında incelenmiştir.

3.4. Aralıklı Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri

3.4.1. Giriş

Bu bölümde, parametrik yaşam modelleri için olabilirlik temelli çıkarımlar incelenmiştir.

n tane bağımsız birimden oluşan bir çalışma düşünölsün. $i = 1, 2, \dots, n$ olmak üzere, i birim için yaşam süresi T_i ile gösterilsin ve T_i 'nin yaşam fonksiyonu $S(t, \theta)$ olan bir parametrik modele sahip olduđu varsayölsün. Burada $\theta = (\theta_1, \dots, \theta_p)$ bilinmeyen parametreleri belirtir. Ayrıca, sadece aralıklı durdurulmuş veri yapılarının olduđu ve

$$\{(L_i, R_i], X_i; i = 1, \dots, n\}$$

biçimine sahip olduđu varsayölsün. Buradaki $(L_i, R_i]$, T_i 'nin bulunduđu aralığı gösterir ve X_i , $i = 1, \dots, n$ olmak üzere i birimle ilişkili ortak değışken vektörüdür. Tüm $i = 1, \dots, n$ için $L_i < R_i$ olduđu varsayölırsa, olabilirlik fonksiyonu

$$L(\theta) = \prod_{i=1}^n L_i(\theta) = \prod_{i=1}^n [S(L_i, \theta) - S(R_i, \theta)]$$

ile orantılıdır.

Genel Parametrik Modeller ile Çıkarım

θ ile ilgili çıkarımlara standart bir yaklaşım, $L(\theta)$ 'yı maksimum yapan θ değerini EÇO yöntemi ile tahmin etmek ve $L(\theta)$ 'dan elde edilen skor ve olabilirlik oranı (LR) test istatistiklerini kullanmaktır. Burada bağımsız durdurma varsayımıyla genel olarak, standart büyük örneklem olabilirlik teorisi uygulanır. Özellikle, EÇO tahmin edicisinin dağılımları için asimptotik yaklaşımlar, skor test istatistiği ve LR test istatistiği kullanılır ve bunlar basit çıkarım yaklaşımları sağlar. Skor test istatistiği ve Fisher bilgi matrisi

$$U(\theta) = \sum_{i=1}^n U_i(\theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} L_i(\theta)$$

ve

$$I(\theta) = \sum_{i=1}^n I_i(\theta) = \sum_{i=1}^n \frac{\partial^2}{\partial \theta \partial \theta} L_i(\theta)$$

olarak ifade edilsin. Belirli koşullar altında, θ 'nın EÇO tahmin edicisi olan $\hat{\theta}$ tutarlıdır. Ayrıca, n büyük olduğunda, $U(\theta) = 0$ için tek bir çözüm vardır ve dağılımına ortalama θ ve kovaryans matrisi $I^{-1}(\theta)$ olan çok değişkenli normal dağılım ile yaklaşılabılır. Başka bir deyişle,

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta))$$

θ için hipotezleri test etmek ve aralık tahminlerini türetmek için kullanılabilir. $\hat{\theta}$ 'yı belirlemek için, herhangi bir kök bulma prosedürü veya Newton-Raphson algoritması kullanılabilir.

$I(\theta_0)$ 'ın tekil olmadığını varsayalım. Newton-Raphson algoritmasında, θ 'nın başlangıç değeri θ_0 , $\theta_1 = \theta_0 - I^{-1}(\theta_0) \cup (\theta_0)$ ile yakınsayana kadar tekrar ederek güncellenir. θ_0 biliniyorsa, $H_1: \theta = \theta_0$ hipotezini test etmek için, p serbestlik derecesi ile asimptotik χ^2 dağılımı olan

$$U'(\theta_0)I^{-1}(\theta_0) \cup (\theta_0)$$

skor test istatistiğinin kullanılması uygundur.

$\theta = (\theta'_1, \theta'_2)$ eşitliğinde yer alan θ 'nın θ_1 ve θ_2 bileşenlerinin, sırasıyla k ve $p-k$ boyutlarına sahip olduğunu varsayalım. Buna göre uygulamada, θ_{10} bilindiğinde $H_2: \theta_1 = \theta_{10}$ daha bilinen bir hipotez haline gelir. Bu durumda, θ aynı biçimde $U(\theta)$, $U'(\theta) = [U'_1(\theta_1, \theta_2), U'_2(\theta_1, \theta_2)]$ olarak parçalanır. Burada U_1 ve U_2 sırasıyla θ_1 ve θ_2 'ye karşılık gelen k ve $p-k$ boyutlarına sahiptir. $\hat{\theta}_{20}$, $U_2(\theta_{10}, \theta_2) = 0$ çözümü ile verilen θ_2 tahminini; $I^{11}(\theta)$, θ_1 ile $I^{-1}(\theta)$ 'ya karşılık gelen alt matrisi gösterir. O halde, H_2 testi $\hat{\theta} = (\hat{\theta}'_1, \hat{\theta}'_2)'$ olduğunda aşağıda verilen istatistiklere dayanabilir:

$$(\hat{\theta}_1 - \theta_{10})' [I^{11}(\theta_{10}, \hat{\theta}_{20})]^{-1} (\hat{\theta}_1 - \theta_{10}) \quad (3.11)$$

veya

$$U'_1(\theta_{10}, \hat{\theta}_{20}) I^{11}(\theta_{10}, \hat{\theta}_{20}) U_1(\theta_{10}, \hat{\theta}_{20}). \quad (3.12)$$

Her iki istatistik de k serbestlik derecesi ile bir asimptotik χ^2 dağılımına sahiptir. Yukarıda verilen yöntemlerde, bazen gözlemlenmesi zor olabilen Fisher bilgi matrisi $I(\theta)$ belirlenmelidir. Dezavantajı olmayan bu yaklaşım, LR test istatistiğine dayanmaktadır:

$$LR(\theta) = \frac{L(\theta)}{L(\hat{\theta})}.$$

H_1 ve H_2 hipotezlerini test etmek için $-2\log[LR(\theta_0)]$ ya da $-2\log[LR(\theta_{10}, \hat{\theta}_{20})]$ istatistiği kullanılabilir. Bu istatistikler, sırasıyla p ve k serbestlik dereceleri ile asimptotik χ^2 dağılımına sahiptir (Sun, 2006).

Üstel Yaşam Modeli ile Çıkarım

Tehlike fonksiyonu $h(t; X_i) = \exp(X_i' \beta)$ olarak ifade edildiğinde, $X_{i1} = 1$ yazılabilir.

Bu gösterimde tüm değişkenlerin değeri 0 olduğunda, $\theta = \beta = (\beta_1, \dots, \beta_{p+1})'$ olmak üzere TO, $h = \exp(\beta_1)$ olarak ifade edilir.

i. birim ($i = 1, \dots, n$) için $L_i = R_i$ ise $\delta_i = 1$, diğer durumlarda $\delta_i = 0$ olarak tanımlanır. Yani δ_i , i. birimin kesin bir birim olup olmadığını göstermektedir. Buna göre X_i 'ye bağlı olabilirlik fonksiyonu

$$L(\beta) = \prod_{i: \delta_i=1} e^{X_i' \beta} \exp(-e^{X_i' \beta} L_i) \prod_{i: \delta_i=0} \left[\exp(-e^{X_i' \beta} L_i) - \exp(-e^{X_i' \beta} R_i) \right]$$

skor vektörü

$$U(\beta) = \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n X_i \left[\delta_i (1 - e^{X_i' \beta} L_i) + (1 - \delta_i) \exp(X_i' \beta) d_i(\beta, X_i) \right]$$

ve gözlemlenen Fisher bilgi matrisleri

$$\hat{I}(\beta) = -\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta'} = \sum_{i=1}^n X_i X_i' e^{X_i' \beta} \left\{ \delta_i L_i - (1 - \delta_i) \left[d_i(\beta, X_i) + e^{X_i' \beta} d_i^*(\beta, X_i) \right] \right\}$$

biçimindedir. Burada

$$d_i(\beta, X_i) = \frac{R_i \exp(-e^{X_i' \beta} R_i) - L_i \exp(-e^{X_i' \beta} L_i)}{\exp(-e^{X_i' \beta} L_i) - \exp(-e^{X_i' \beta} R_i)}$$

ve

$$d_i^*(\beta, X_i) = \frac{L_i^2 \exp(-e^{X_i' \beta} L_i) - R_i^2 \exp(-e^{X_i' \beta} R_i)}{\exp(-e^{X_i' \beta} L_i) - \exp(-e^{X_i' \beta} R_i)} - d_i^2(\beta, X_i) \text{ 'dir.}$$

İki örneklem karşılaştırma problemi için, daha önce olduğu gibi $X_{i1} = 1$ ve $X_{i2} = 0$ ya da 1 olduğunda $X_i = (X_{i1}, X_{i2})$ olur. Karşılaştırma $\beta_2 = 0$ testi ile aynı olup, $\theta_1 = \beta_2$ ve $\theta_2 = \beta_1$ ile Eşitlik 3.11 ve Eşitlik 3.12'deki istatistikler kullanılarak yapılabilir (Sun, 2006).

Log-Doğrusal Yaşam Modeli ile Çıkarım

Log-doğrusal yaşam modeli, üstel yaşam modeline göre daha genel bir durumu ele alır. $Y_i = \log(T_i)$ tanımına göre, Y_i 'nin o.y.f'sinin $\sigma^{-1} f(w)$ olduğu varsayılınsın. Burada $w = (y - X' \beta) / \sigma$ 'dır. Tüm i değerleri için $X_{i1} = 1$ olduğu varsayımıyla, β ve σ için olabilirlik fonksiyonu

$$L(\beta, \sigma) = \prod_{i=1}^n [\sigma^{-1} f(w_{L_i})]^{\delta_i} [S(w_{L_i}) - S(w_{R_i})]^{1-\delta_i}.$$

olarak yazılabilir. Burada $w_{L_i} = (\log L_i - X_i' \beta) / \sigma$, $w_{R_i} = (\log R_i - X_i' \beta) / \sigma$,

$S(w) = \int_w^\infty f(s) ds$ 'dir ve δ_i önceden tanımlandığı gibidir. Skor vektörü ise,

$$U_\beta(\beta, \sigma) = \frac{\partial \log L(\beta, \sigma)}{\partial \beta} = -\sigma^{-1} \sum_{i=1}^n X_i \left[\delta_i \frac{f'(w_{L_i})}{f(w_{L_i})} - (1-\delta_i) \frac{w_{L_i} f(w_{L_i}) - w_{R_i} f(w_{R_i})}{S(w_{L_i}) - S(w_{R_i})} \right]$$

biçimindedir ve $f'(w) = df(w)/dw$ 'dir. Gözlemlenen Fisher bilgi matrisinin bileşenleri:

$$-\frac{\partial^2 \log L(\beta, \sigma)}{\partial \beta \partial \beta'} = \sigma^{-2} \sum_{i=1}^n X_i X_i' \left\{ \delta_i \left[\left(\frac{f'(w_{L_i})}{f(w_{L_i})} \right)^2 - \frac{f''(w_{L_i})}{f(w_{L_i})} \right] + (1 - \delta_i) \left[\left(\frac{f(w_{L_i}) - f(w_{R_i})}{s(w_{L_i}) - s(w_{R_i})} \right)^2 + \frac{f'(w_{L_i}) - f'(w_{R_i})}{s(w_{L_i}) - s(w_{R_i})} \right] \right\}$$

$$-\frac{\partial^2 \log L(\beta, \sigma)}{\partial \beta \partial \beta'} = \sigma^{-2} \sum_{i=1}^n X_i \left\{ \delta_i w_{L_i} \left[\left(\frac{f'(w_{L_i})}{f(w_{L_i})} \right)^2 - \frac{f''(w_{L_i})}{f(w_{L_i})} \right] + (1 - \delta_i) \left[\frac{(f(w_{L_i}) - f(w_{R_i}))(w_{L_i} f(w_{L_i}) - w_{R_i} f(w_{R_i}))}{(s(w_{L_i}) - s(w_{R_i}))^2} + \frac{w_{L_i} f'(w_{L_i}) - w_{R_i} f'(w_{R_i})}{s(w_{L_i}) - s(w_{R_i})} \right] \right\} + \sigma^{-1} U_\beta(\beta, \sigma)$$

ve

$$-\frac{\partial^2 \log L(\beta, \sigma)}{\partial \sigma \partial \sigma'} = \sigma^{-2} \sum_{i=1}^n \left\{ \delta_i \left[\left(\frac{w_{L_i} f'(w_{L_i})}{f(w_{L_i})} \right)^2 - \frac{w_{L_i} f'(w_{L_i}) + w_{L_i}^2 f''(w_{L_i})}{f(w_{L_i})} \right] + (1 - \delta_i) \left[\left(\frac{w_{L_i} f(w_{L_i}) - w_{R_i} f(w_{R_i})}{s(w_{L_i}) - s(w_{R_i})} \right)^2 + \frac{w_{L_i} f(w_{L_i}) + w_{L_i}^2 f'(w_{L_i}) - w_{R_i} f(w_{R_i}) + w_{R_i}^2 f'(w_{R_i})}{s(w_{L_i}) - s(w_{R_i})} \right] \right\} + \sigma^{-1} U_\sigma(\beta, \sigma)$$

olur. Burada $f''(w) = d^2 f(w)/dw^2$ 'dir (Sun, 2006).

3.4.2. Aralıklı Durdurulmuş Veriler İçin Hızlandırılmış Başarısızlık Süresi Modeli

Gözlemlenen aralıklı başarısızlık süresi verisi

$$\left\{ K_i, \left(U_{ij}, \delta_{ij} = I(U_{j-1} < T_i \leq U_{ij}) \right)_{j=1}^{K_i}, X_i; i = 1, \dots, n \right\}$$

ile ifade edilsin. Burada, i . birim için açıklayıcı değişkenler vektörü X_i , birim sayısı K_i , yaşam süresi T_i ve $U_0 = 0$ olan gözlemlenen zaman noktaları $U_{i1} < \dots < U_{iK_i}$ ile gösterilmiştir. n bağımsız birimden oluşan çalışmadaki her birim, yaşam süresinden bağımsız olduğu varsayılan bir dizi zaman noktasında gözlemlenmektedir. T_i 'lerin HBS modeline sahip olduğu, X_i bilindiğinde U_{ij} 'lerin ve T_i 'nin bağımsız olduğu ve sadece birinin regresyon parametresi ile ilgilendiği varsayılmaktadır.

$\log T = X'\beta + W$ eşitliğindeki W dağılımı F ile gösterilsin ve p -boyutlu b vektörü için $U_{ij}(b) = \log(U_{ij}) - X'_i b$, $j = 1, \dots, K$; $i = 1, \dots, n$ tanımlansın. Ayrıca, $Y_{ij} = I(T_i \leq U_{ij})$ biçiminde ifade edilsin ve her i için U_{iL} ve U_{iR} sırasıyla, $Y_{ij} = 0$ için son gözlem süresi ve $Y_{ij} = 1$ için ilk gözlem süresi olan iki U_{ij} olsun. Yani, $(U_{ij}, U_{iR}]$ T_i 'nin gözlemlendiği aralıktır. Buna göre olabilirlik fonksiyonu,

$$L(\beta, F) = \prod_{i=1}^n [F(U_{iR}(\beta)) - F(U_{iL}(\beta))] \quad (3.13)$$

biçiminde yazılır. Burada $U_{iL}(b) = \log(U_{iL}) - X'_i b$ ve $U_{iR}(b) = \log(U_{iR}) - X'_i b$ 'dir. β ile ilgili çıkarım doğrusal sıralı istatistiklere dayanan bir tahmin denklemi ile elde edilmektedir (Sun, 2006).

4. UYGULAMA

Çalışmamızda aralıklı durdurulmuş ve sağdan durdurulmuş veri yapısı için hem CRM'yi hem de parametrik yaşam modellerini incelemek amacıyla, literatürde yer alan Primer Biliyer Siroz, PBS, (Primary Biliary Cirrhosis, PBC) verisi kullanılmıştır. 312 hastaya ait PBS veri kümesine ilişkin veriler Ocak 1924-Mayıs 1984 yılları arasında Mayo Klinik'te toplanmıştır (Murtaugh ve diğ., 1994; Mayo Clinic). 312 hastanın 158'ine D-penisilin ilacı verilmiş ve 154 hasta ise placebo grubu olarak seçilmiştir. İlk 6 ay ve 6. aydan sonra her yıl olmak üzere, serum bilirubin düzeyleri tekrarlı biçimde ölçülmüştür (Dil, 2019).

Literatürde PBS verisi kullanılarak farklı yaşam çözümlemesi yöntemleri incelenmiş olup, bu çalışmalar Dil (2019)'da detaylı olarak verilmiştir. PBS verisi aralıklı olarak ölçülen değerleri içerdiği için veri yapısı, aralıklı durdurulmuş olarak ele alınabilmektedir. Bu çalışma, yaşam çözümlemesi için aralıklı durdurma literatüründe PBS verisi kullanılarak yapılan ilk çalışma olma özelliğine sahiptir.

Çalışmada sonuçlar, Stata 15'in stcox, streg ve stintreg komutlarından ve R programının icenReg ve icensBKL paketlerinden yararlanılarak elde edilmiştir.

Çalışmada başarısızlık ölüm olarak tanımlanmıştır. Yaşam süresi, hastaların gözlemlenmeye başlamasından itibaren ölüm gerçekleşene kadar geçen süre (yıl) olarak belirlenmiştir. 140 (%44.9) hastanın öldüğü ve 172 (%55.1) hastanın ise durdurulduğu görülmüştür. Veri kümesinde hastaların yaşı, cinsiyeti, kullandıkları ilaç, asit, hepatomegali, deride bozukluk, ödem, histolojik evre, log serum bilirubin (logb) (mg/dl), log albumin (logalbumin), log protrombin süresi (logprotime) değişkenleri yer almaktadır. Kullanılan kategorik değişkenler ve düzeyleri Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Kategorik değişkenler ve düzeyleri

Değişken	Düzeyler	n	%
İlaç	0-Placebo	158	50.6
	1-D-penisilin	154	49.4
Cinsiyet	0-Kadın	276	88.5
	1-Erkek	36	11.5
Asit	0-Yok	257	82.4
	1-Var	55	17.6
Hepatomegali	0-Yok	165	52.9
	1-Var	147	47.1
Deride Bozukluk	0-Yok	208	66.7
	1-Var	104	33.3
Ödem	0-Var	165	52.9
	1-İlaca rağmen ödem (Odem1)	80	25.6
	2-Yok (Odem2)	67	21.5
Histolojik Evre	1-Evre 1	11	3.5
	2-Evre 2	45	14.4
	3-Evre 3	102	32.7
	4-Evre 4	154	49.4

Çalışmada yaş, logb, logalbumin ve logprotime değişkenleri sürekli değişken olarak ele alınmış olup, ortalamalar, yaş için 49.999 ± 0.600 , logb için 1.196 ± 0.075 , logalbumin için 1.105 ± 0.011 ve logprotime için 2.465 ± 0.009 olarak elde edilmiştir.

PBS, ilk 6 ay ve 6. aydan sonra her yıl tekrarlı olarak ölçülmüş bir veri kümesidir. Bu çalışmada, sağdan durdurulmuş ve aralıklı durdurulmuş veri yapılarını aynı veri kümesi üzerinde inceleyebilmek için veride bazı düzenlemeler yapılmıştır. Sağdan durdurulmuş olarak oluşturulan veri yapısı için yaşam süresi, gözlem süresinin başlangıcından bitiş noktasına kadar olan toplam süre olarak ele alınmıştır. Aralıklı

durdurulmuş olarak oluşturulan veri yapısı için ise sadece son aralık ele alınmış ve çözümleme bu aralık dikkate alınarak yapılmıştır.

4.1. Sağdan Durdurulmuş Veri Yapısı İçin Model Sonuçları

4.1.1. Sağdan Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli Sonuçları

PBS veri kümesi için kullanılan CRM, p değeri 0.0000 olduğundan %95 güven düzeyinde anlamlıdır. CRM için 1290.177 olan AIC değeri ve 1342.579 olan BIC değeri elde edilmiştir. Çizelge 4.2'de PBS veri kümesi için CRM sonuçları verilmiştir.

Çizelge 4.2. Sağdan durdurulmuş veri yapısı için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	0.0556	0.0093	0.000	1.0572	1.0381, 1.0767
İlaç	0.0125	0.1772	0.944	1.0126	0.7155, 1.4331
Cinsiyet	-0.2792	0.2493	0.263	0.7564	0.4640, 1.2331
Asit	0.6143	0.2453	0.012	1.8483	1.1428, 2.9893
Hepatomegali	-0.0997	0.2302	0.665	0.9051	0.5764, 1.4212
Deride Bozukluk	0.2830	0.2314	0.221	1.3271	0.8433, 2.0886
Ödem1	-0.0520	0.2470	0.833	0.9493	0.5850, 1.5405
Ödem2	-0.0717	0.2556	0.779	0.9308	0.5640, 1.5362
Logb	0.8445	0.1035	0.000	2.3268	1.8997, 2.8500
Logalbumin	-0.2479	0.4498	0.581	0.7804	0.3232, 1.8845
Logprotime	-0.9506	0.6810	0.163	0.3865	0.1017, 1.4683
Histolojik Evre2	-1.1032	0.9222	0.232	0.3318	0.0544, 2.0223
Histolojik Evre3	-1.0829	0.7692	0.159	0.3386	0.0750, 1.5290
Histolojik Evre4	-0.7767	0.7608	0.307	0.4599	0.1035, 2.0429

CRM'de değişken düzeylerini yorumlayabilmek için düzeylerden birinin referans kategorisi olarak alınması gerekmektedir. Bu sebeple çalışmamızda, referans kategorisi tüm değişkenler için birinci düzey olarak alınmıştır.

Çizelge 4.2’de verilen p değerleri incelendiğinde yaş ($p=0.000$), asit ($p=0.012$) ve logb ($p=0.000$) değişkenlerinin %95 güven düzeyinde anlamlı ($p<0.05$) olduğu görülmüştür. CRM sonuçlarına göre, vücudunda asit olan hastaların vücudunda asit olmayan hastalara göre yaklaşık 1.8483 kat ($e^{0.6143}=1.8483$) daha fazla riskli olduğu söylenebilmektedir. Yaş değişkenindeki 1 birimlik artış, başarısızlık riskini 1.0572 birim artırır. Logb değişkenindeki 1 birimlik artış ise başarısızlık riskini 2.3268 birim artırmaktadır.

CRM’nin veri kümesine uygulanabilmesi için OT varsayımının sağlanması gerekmektedir. Bu çalışmada, Schoenfeld artıkları, grafiksel ve zamana bağlı (ZB) açıklayıcı değişkenler yöntemleri kullanılarak bu varsayım incelenmiş ve sonuçlar sırasıyla Çizelge 4.3, Şekil 4.1, Çizelge 4.4, Çizelge 4.5, Çizelge 4.6, Çizelge 4.7, Çizelge 4.8, Çizelge 4.9, Çizelge 4.10 ve Çizelge 4.11’de verilmiştir.

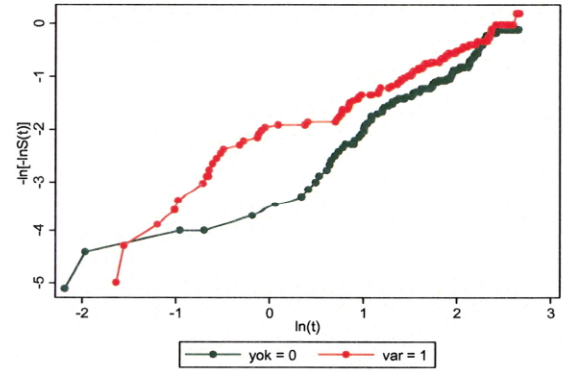
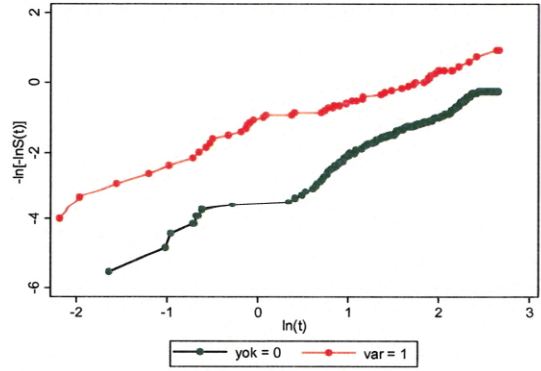
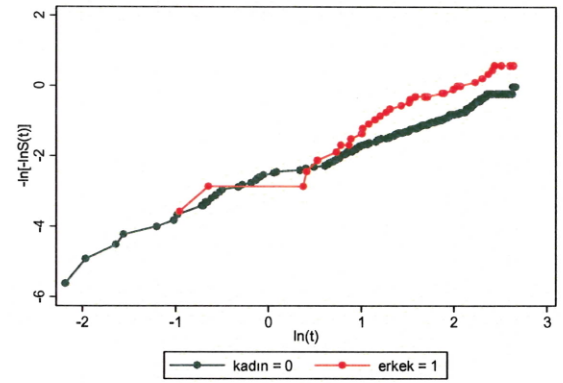
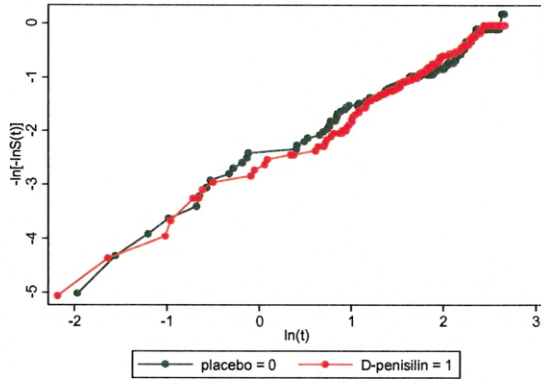
Çizelge 4.3. OT varsayımı için Schoenfeld artıkları ile yaşam süresinin rankı arasındaki ilişki sonuçları

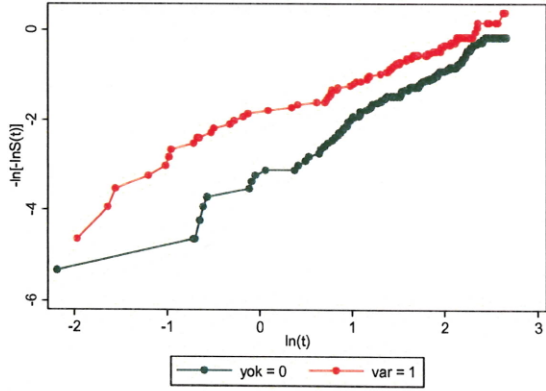
Değişken	r^*	Ki-kare	Std. hata	p değeri
Yaş	0.0052	0.00	1	0.950
İlaç	0.0082	0.01	1	0.922
Cinsiyet	0.0383	0.25	1	0.618
Asit	-0.0547	0.46	1	0.500
Hepatomegali	-0.0292	0.14	1	0.711
Deride Bozukluk	-0.0615	0.65	1	0.422
Ödem1	0.0132	0.03	1	0.859
Ödem2	-0.0836	1.13	1	0.289
Logb	-0.0016	0.00	1	0.983
Logalbumin	-0.0674	0.64	1	0.425
Logprotime	0.0571	0.88	1	0.348
Histolojik Evre2	-0.1014	1.38	1	0.239
Histolojik Evre3	-0.0622	0.56	1	0.454
Histolojik Evre4	-0.0498	0.36	1	0.549

* r: Schoenfeld artıkları ile logt arasındaki Pearson çarpım-moment korelasyon katsayısı

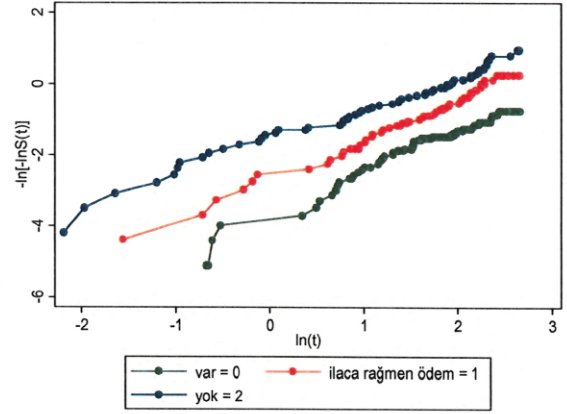
Yokluk hipotezi “OT varsayımı sağlanmaktadır” biçiminde kurulmaktadır ve buna göre elde edilen sonuçlar incelendiğinde, tüm değişkenlerin ($p > 0.05$) OT varsayımını sağladığı %95 güven düzeyinde görülmüştür.

OT varsayımı grafiksel yöntemle de incelenmiş olup, Şekil 4.1’de $\ln(t)$ ’ye karşı çizilen $-\ln(-\ln S(t))$ grafikleri verilmiştir.

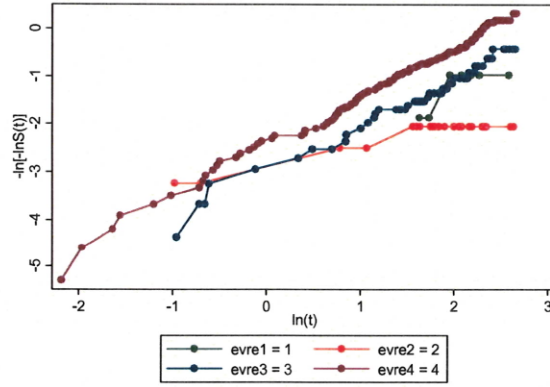




Deride Bozukluk



Ödem



Histolojik Evre

Şekil 4.1. Değişkenlerin orantılı tehlikeler varsayımı

Şekil 4.1'de görüldüğü üzere ilaç, cinsiyet, hepatomegali ve histolojik evre değişken düzeylerinin çakışması sebebiyle OT varsayımı sağlanmamaktadır.

OT varsayımını incelemek için kullanılan iki yöntem de farklı sonuç verdiği için, ZB açıklayıcı değişkenler yöntemi kullanılarak bu varsayım üçüncü bir yöntem ile test edilmiştir.

Logt fonksiyonu kullanılarak OT varsayımını sağlamadığı düşünülen ilaç, cinsiyet, hepatomegali ve histolojik evre değişkenleri için ZB açıklayıcı değişkenler yöntemi kullanılmıştır. Öncelikle ilaç, cinsiyet, hepatomegali ve histolojik evre değişkenlerini

içeren CRM çözümlemesi yapılmıştır ve sonuçlar sırasıyla Çizelge 4.4, Çizelge 4.6, Çizelge 4.8 ve Çizelge 4.10'da verilmiştir. Daha sonra ilaç, cinsiyet, hepatomegali ve histolojik evre değişkenlerini ve zamana bağlı bir fonksiyon olarak tanımlanan ilaçxlogt, cinsiyetxlogt, hepatomegalixlogt ve histolojik evrexlogt değişkenlerini içeren CRM çözümlemesi yapılarak sonuçlar sırasıyla Çizelge 4.5, Çizelge 4.7, Çizelge 4.9 ve Çizelge 4.11'de verilmiştir.

Çizelge 4.4. İlaç değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
İlaç	0.021	0.170	0.901	1.021	0.7318, 1.4250

Çizelge 4.5. İlaç ve İlaçxlogt değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
İlaç	-0.077	0.250	0.757	0.925	0.5672, 1.5113
İlaçxlogt	0.089	0.166	0.592	1.093	0.7895, 1.5135

Çizelge 4.6. Cinsiyet değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Cinsiyet	0.652	0.218	0.003	1.920	1.2519, 2.9426

Çizelge 4.7. Cinsiyet ve Cinsiyetxlogt değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Cinsiyet	0.397	0.353	0.261	1.487	0.7446, 2.9710
Cinsiyetxlogt	0.219	0.223	0.326	1.245	0.8040, 1.9273

Çizelge 4.8. Hepatomegali değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Hepatomegali	0.280	0.170	0.100	1.323	0.9482, 1.8463

Çizelge 4.9. Hepatomegali ve Hepatomegalixlogt değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Hepatomegali	0.685	0.269	0.011	1.983	1.1709, 3.3609
Hepatomegalixlogt	-0.363	0.180	0.044	0.696	0.4888, 0.9899

Çizelge 4.10. Histolojik Evre değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Histolojik Evre 2	-0.761	0.913	0.404	0.467	0.0780, 2.7969
Histolojik Evre 3	0.250	0.736	0.735	1.283	0.3034, 5.4336
Histolojik Evre 4	0.936	0.714	0.189	2.551	0.6291, 10.3336

Çizelge 4.11. Histolojik Evre ve Histolojik Evrexlogt değişkeni için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Histolojik Evre 2	-0.773	0.932	0.407	0.462	0.0743, 2.8682
Histolojik Evre 3	0.228	0.820	0.781	1.256	0.2518, 6.2664
Histolojik Evre 4	0.905	0.889	0.309	2.471	0.4328, 14.1173
Histolojik Evrexlogt	0.009	0.151	0.953	1.009	0.7505, 1.3566

Test istatistiği kullanılarak OT varsayımını incelemek için yokluk hipotezi “OT varsayımı sağlanmaktadır” olarak kurulmuştur. Yokluk hipotezi altında 1 serbestlik dereceli ki-kare dağılımına sahip olan test istatistiği, CRM ile ZB CRM arasındaki log-olabilirlik oranı istatistiğinin farkından oluşmaktadır. Buna göre test istatistiği;

İlaç değişkeni için,

$$\begin{aligned} LR &= -2\ln L_{CRM} - (-2\ln L_{ZB CRM}) \\ &= 1441.553 - 1441.265 \\ &= 0.288 \end{aligned}$$

olarak elde edilmiştir. Olabilirlik oranı test istatistiği 0.228 bulunmuştur ve $\chi_1^2 = 3.86$ olduğu için %95 güven düzeyinde yokluk hipotezi kabul edilir. Yani, ilaç değişkeni için OT varsayımı sağlanmaktadır.

Cinsiyet değişkeni için,

$$\begin{aligned} LR &= -2\ln\hat{L}_{CRM} - (-2\ln\hat{L}_{ZBCRM}) \\ &= 1433.807 - 1432.783 \\ &= 1.024 \end{aligned}$$

olarak elde edilmiştir. Olabilirlik oranı test istatistiği 1.024 bulunmuştur ve $\chi_1^2 = 3.86$ olduğu için %95 güven düzeyinde yokluk hipotezi kabul edilir. Yani, cinsiyet değişkeni için OT varsayımı sağlanmaktadır.

Hepatomegali değişkeni için,

$$\begin{aligned} LR &= -2\ln\hat{L}_{CRM} - (-2\ln\hat{L}_{ZBCRM}) \\ &= 1441.569 - 1441.569 \\ &= 0 \end{aligned}$$

olarak elde edilmiştir. Olabilirlik oranı test istatistiği 0 bulunmuştur ve $\chi_1^2 = 3.86$ olduğu için %95 güven düzeyinde yokluk hipotezi kabul edilir. Yani, hepatomegali değişkeni için OT varsayımı sağlanmaktadır.

Histolojik evre değişkeni için,

$$\begin{aligned} LR &= -2\ln\hat{L}_{CRM} - (-2\ln\hat{L}_{ZBCRM}) \\ &= 1417.908 - 1417.904 \\ &= 0.004 \end{aligned}$$

olarak elde edilmiştir. Olabilirlik oranı test istatistiği 0.004 bulunmuştur ve $\chi^2_1 = 3.86$ olduğu için %95 güven düzeyinde yokluk hipotezi kabul edilir. Yani, histolojik evre değişkeni için OT varsayımı sağlanmaktadır.

Sonuç olarak, Schoenfeld artıkları ve ZB açıklayıcı değişkenler yöntemleri benzer sonuçlar verdiğinden, veri kümesi için OT varsayımı sağlanmıştır. Bu nedenle CRM sonuçlarının kullanılması ve yorumlanması uygundur.

4.1.2. Sağdan Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri Sonuçları

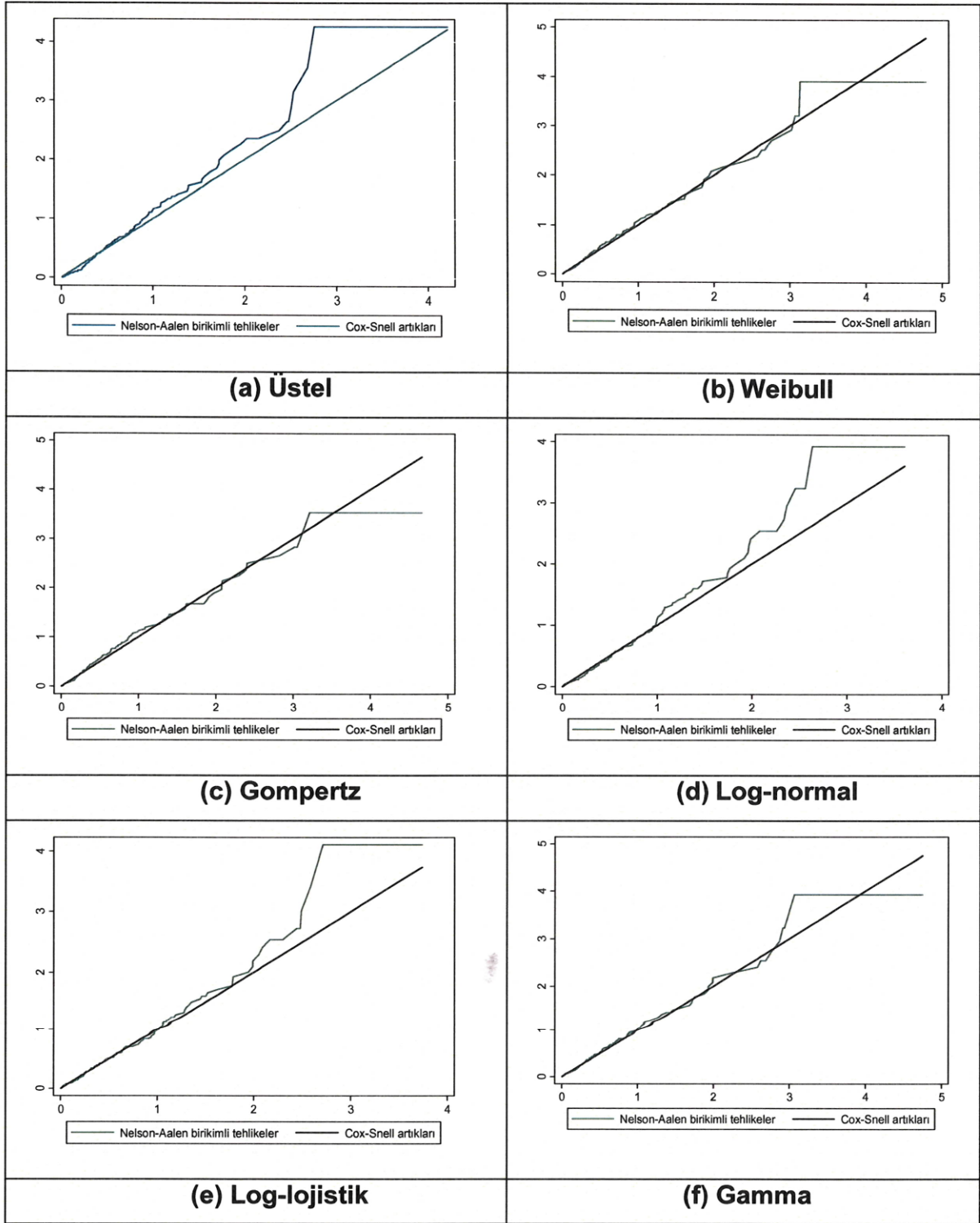
PBS veri kümesini çözümlemek için OT biçimindeki parametrik yaşam modelleri olarak üstel, Weibull ve Gompertz; HBS biçimindeki parametrik yaşam modelleri olarak üstel, Weibull, Gamma, log-normal ve log-lojistik yaşam modelleri kullanılmıştır. LR ile modellerin anlamlılığı test edilmiş ve istatistiksel olarak tüm modeller anlamlı ($p < 0.05$) bulunmuştur. Çizelge 4.12'de model seçiminde kullanılan AIC ve BIC değerleri verilmiştir:

Çizelge 4.12. Sağdan durdurulmuş veri yapısı için parametrik yaşam modellerine ait AIC ve BIC değerleri

Kriter	OT Biçimindeki Parametrik Yaşam Modelleri		
	Üstel	Weibull	Gompertz
AIC	571.089	560.978	559.398
BIC	627.234	620.866	619.286

Kriter	HBS Biçimindeki Parametrik Yaşam Modelleri				
	Üstel	Weibull	Log-lojistik	Log-normal	Gamma
AIC	571.089	560.978	562.507	572.313	562.233
BIC	627.234	620.866	622.395	632.201	625.864

Çizelge 4.12'deki sonuçlara göre, OT biçimindeki parametrik yaşam modellerinden Gompertz OT modeli, HBS biçimindeki parametrik yaşam modellerinden ise Weibull HBS modelinin en uygun modeldir. Şekil 4.2'de verilen Cox-Snell artıklarına ait grafikler ile uygun olarak belirlenen bu modellerin doğruluğu ispatlanmıştır.



Şekil 4.2. Cox-Snell artıkları grafikleri

Çizelge 4.13'te PBS veri kümesi için kullanılan Gompertz OT modelinden elde edilen sonuçlar verilmiştir:

Çizelge 4.13. Sağdan durdurulmuş veri yapısı için Gompertz OT modeli sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	0.0561	0.0093	0.000	1.0577	1.0386, 1.0771
İlaç	0.0045	0.1767	0.980	1.0045	0.7105, 1.4202
Cinsiyet	-0.2810	0.2493	0.260	0.7550	0.4631, 1.2308
Asit	0.6074	0.2451	0.013	1.8356	1.1353, 2.9679
Hepatomegali	-0.1058	0.2296	0.645	0.8996	0.5736, 1.4109
Deride Bozukluk	0.2730	0.2312	0.238	1.3139	0.8352, 2.0671
Ödem1	-0.0700	0.2464	0.776	0.9324	0.5753, 1.5113
Ödem2	-0.0897	0.2552	0.725	0.9142	0.5544, 1.5077
Logb	0.8495	0.1034	0.000	2.3385	1.9096, 2.8636
Logalbumin	-0.2900	0.4502	0.520	0.7483	0.3097, 1.8084
Logprotime	-0.9034	0.6740	0.180	0.4052	0.1081, 1.5185
Histolojik Evre2	-1.0975	0.9224	0.234	0.3337	0.0547, 2.0345
Histolojik Evre3	-1.0535	0.7686	0.170	0.3487	0.0773, 1.5727
Histolojik Evre4	-0.7595	0.7619	0.319	0.4679	0.1051, 2.0827

Çizelge 4.13'te verilen p değerleri incelendiğinde yaş ($p=0.000$), asit ($p=0.013$) ve logb ($p=0.000$) değişkenleri %95 güven düzeyinde anlamlı ($p<0.05$) bulunmuştur. Gompertz OT modeli sonuçlarına göre, vücudunda asit olan hastaların vücudunda asit olmayan hastalara göre yaklaşık 1.8356 kat ($e^{0.6074}=1.8356$) daha fazla riskli olduğu söylenebilmektedir. Yaş değişkenindeki 1 birimlik artış, başarısızlık riskini 1.0577 birim artırır. Logb değişkenindeki 1 birimlik artış ise, başarısızlık riskini 2.3385 birim artırmaktadır.

Çizelge 4.14'te Weibull HBS modeli için elde edilen sonuçlar verilmiştir:

Çizelge 4.14. Sağdan durdurulmuş veri yapısı için Weibull HBS modeli sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	-0.0420	0.0072	0.000	0.9589	-0.0562, -0.0279
İlaç	-0.0089	0.1361	0.948	0.9911	-0.2756, 0.2579
Cinsiyet	0.2157	0.1914	0.260	1.2407	-0.1595, 0.5909
Asit	-0.4717	0.1889	0.012	0.6239	-0.8418, -0.1016
Hepatomegali	0.0741	0.1766	0.675	1.0769	-0.2720, 0.4203
Deride Bozukluk	-0.2173	0.1782	0.223	0.8047	-0.5667, 0.1321
Ödem1	0.0572	0.1895	0.763	1.0589	-0.3143, 0.4287
Ödem2	0.0378	0.1948	0.846	1.0385	-0.3440, 0.4196
Logb	-0.6408	0.0814	0.000	0.5269	-0.8004, -0.4812
Logalbumin	0.2397	0.3476	0.490	1.2709	-0.4416, 0.9210
Logprotime	0.7064	0.5194	0.174	2.0267	-0.3115, 1.7243
Histolojik Evre2	0.8430	0.7141	0.238	2.3233	-0.5567, 2.2427
Histolojik Evre3	0.8042	0.5936	0.175	2.2349	-0.3592, 1.9677
Histolojik Evre4	0.5814	0.5883	0.323	1.7885	-0.5718, 1.7345

Başarısızlık süresini etkileyen faktörleri belirlemek için kullanılan Weibull HBS modeli için elde edilen sonuçlar incelendiğinde yaş ($p=0.000$), asit ($p=0.012$) ve logb ($p=0.000$) değişkenlerinin %95 güven düzeyinde anlamlı ($p<0.05$) olduğu görülmüştür.

Vücudunda asit olmayan hastaların ortalanca yaşam süresi, vücudunda asit olan hastaların ortalanca yaşam süresine göre yaklaşık 1.6 ($1/\exp(-0.4717)$) kat daha uzundur. Yaş değişkenindeki 1 birimlik artışın ortalanca yaşam süresini 0.9589 birim ve logb değişkenindeki 1 birimlik artışın ise ortalanca yaşam süresini 0.5269 birim azalttığı söylenebilmektedir.

4.2. Aralıklı Durdurulmuş Veri Yapısı İçin Model Sonuçları

Bu bölümde veriler aralıklı durdurulmuş birimler olarak düzenlenmiş ve buna göre analiz edilmiştir.

4.2.1. Aralıklı Durdurulmuş Veri Yapısı İçin Cox Regresyon Modeli Sonuçları

PBS veri kümesindeki veriler aralıklı durdurulmuş olarak ele alındığında CRM, p değeri 0.0000 olduğu için %95 güven düzeyinde anlamlıdır. CRM için AIC değeri 2828.114, BIC değeri 2880.516 olarak elde edilmiştir. Çizelge 4.15'te PBS veri kümesindeki aralıklı durdurulmuş veri yapısı için CRM sonuçları verilmiştir.

Çizelge 4.15. Aralıklı durdurulmuş veri yapısı için CRM sonuçları

Değişken	$\hat{\beta}$	Std. hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	0.0289	0.0070	0.000	1.0290	1.0153, 1.0435
İlaç	-0.0319	0.1476	0.829	0.9686	0.7253, 1.2936
Cinsiyet	-0.3225	0.2518	0.200	0.7244	0.4422, 1.1865
Asit	0.2858	0.2923	0.328	1.3310	0.7504, 2.3601
Hepatomegali	0.1184	0.1410	0.401	1.1260	0.8538, 1.4841
Deride Bozukluk	0.1947	0.1751	0.266	1.2150	0.8620, 1.7124
Ödem1	-0.0747	0.1769	0.673	0.9281	0.6561, 1.3126
Ödem2	-0.4606	0.2691	0.869	0.6309	0.3723, 1.0691
Logb	0.4502	0.0806	0.000	1.5690	1.3394, 1.8371
Logalbumin	0.3782	0.4218	0.370	1.4600	0.6386, 3.3364
Logprotime	-0.8705	0.5685	0.126	0.4187	0.1374, 1.2761
Histolojik Evre2	0.1600	0.4462	0.720	1.1730	0.4894, 2.8140
Histolojik Evre3	-0.1294	0.4314	0.764	0.8786	0.3772, 2.0464
Histolojik Evre4	-0.1036	0.4253	0.808	0.9016	0.3917, 2.0751

Çizelge 4.15'te verilen p değerlerine göre yaş ($p=0.000$) ve logb ($p=0.000$) değişkenleri %95 güven düzeyinde anlamlı ($p<0.05$) bulunmuştur. Bu sonuçlara göre, yaş değişkenindeki 1 birimlik artış, başarısızlık riskini 1.0293 birim artırır. Logb değişkenindeki 1 birimlik artış ise, başarısızlık riskini 1.5686 birim artırmaktadır.

4.2.2. Aralıklı Durdurulmuş Veri Yapısı İçin Parametrik Yaşam Modelleri Sonuçları

PBS veri kümesini çözümllemek için OT biçimindeki parametrik yaşam modelleri olarak üstel, Weibull ve Gompertz; HBS biçimindeki parametrik yaşam modelleri

olarak üstel, Weibull, Gamma, log-lojistik ve log-normal yaşam modelleri kullanılmıştır. LR ile modellerin anlamlılığı test edilmiş ve istatistiksel olarak tüm modeller anlamlı ($p < 0.05$) bulunmuştur. Çizelge 4.16'da model seçiminde kullanılan AIC ve BIC değerleri verilmiştir:

Çizelge 4.16. Aralıklı durdurulmuş veri yapısı için parametrik yaşam modellerine ait AIC ve BIC değerleri

Kriter	OT Biçimindeki Parametrik Yaşam Modelleri		
	Üstel	Weibull	Gompertz
AIC	1492.407	1393.990	1337.837
BIC	1548.552	1453.878	1397.725

Kriter	HBS Biçimindeki Parametrik Yaşam Modelleri				
	Üstel	Weibull	Log-lojistik	Log-normal	Gamma
AIC	1492.407	1393.990	2135.830	1492.066	1372.076
BIC	1548.552	1453.878	2195.718	1551.954	1435.707

Çizelge 4.16'daki sonuçlara göre OT biçimindeki parametrik yaşam modelleri içerisinde Gompertz OT modelinin, HBS biçimindeki parametrik yaşam modelleri içerisinde Gamma HBS modelinin en uygun model olduğu görülmüştür.

Çizelge 4.17. Aralıklı durdurulmuş veri yapısı için Gompertz OT modeli sonuçları

Değişken	$\hat{\beta}$	Std. Hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	0.0298	0.0066	0.000	1.0302	1.0170, 1.0436
İlaç	-0.0184	0.1204	0.879	0.9818	0.7753, 1.2431
Cinsiyet	-0.3672	0.1940	0.058	0.6927	0.4736, 1.0132
Asit	0.3661	0.1787	0.040	1.4421	1.0160, 2.0468
Hepatomegali	0.1205	0.1345	0.370	1.1281	0.8667, 1.4685
Deride Bozukluk	0.1768	0.1410	0.210	1.1934	0.9053, 1.5732
Ödem1	-0.0920	0.1602	0.566	0.9121	0.6664, 1.2484
Ödem2	-0.4832	0.1970	0.014	0.6168	0.4192, 0.9075
Logb	0.4473	0.0633	0.000	1.5641	1.3815, 1.7708
Logalbumin	0.3525	0.3565	0.323	1.4226	0.7074, 2.8609
Logprotime	-0.8039	0.5249	0.126	0.4476	0.1600, 1.2522
Histolojik Evre2	0.2247	0.4354	0.606	1.2519	0.5333, 2.9389
Histolojik Evre3	-0.0964	0.4098	0.814	0.9081	0.4068, 2.0273
Histolojik Evre4	-0.0445	0.4081	0.913	0.9565	0.4299, 2.1284

Çizelge 4.17’de verilen p değerleri incelendiğinde yaş ($p=0.000$), asit ($p=0.040$), ödem2 ($p=0.014$) ve logb ($p=0.000$) değişkenleri %95 güven düzeyinde anlamlı ($p<0.05$) bulunmuştur. Gompertz OT modeli sonuçlarına göre, vücudunda asit olan hastaların vücudunda asit olmayan hastalara göre yaklaşık 1.4421 kat ($e^{0.3661}=1.4421$) daha fazla riskli olduğu söylenebilmektedir. Vücudunda ödem olan hastaların vücudunda ödem olmayan hastalara göre yaklaşık 1.62 ($1/\exp(-0.4832)$) kat daha fazla riskli olduğu görülmüştür. Yaş değişkenindeki 1 birimlik artış, başarısızlık riskini 1.0302 birim artırır. Logb değişkenindeki 1 birimlik artış ise başarısızlık riskini 1.5641 birim artırmaktadır.

Çizelge 4.18. Aralıklı durdurulmuş veri yapısı için Gamma HBS modeli sonuçları

Değişkenler	$\hat{\beta}$	Std. Hata	p değeri	$\exp(\hat{\beta})$	$\exp(\hat{\beta})$ için %95 güven aralığı
Yaş	-0.0147	0.0033	0.000	0.9854	-0.0212, -0.0083
İlaç	0.0425	0.0568	0.454	1.0434	-0.0687, 0.1538
Cinsiyet	0.1677	0.0880	0.057	1.1826	-0.0047, 0.3402
Asit	-0.0947	0.0825	0.251	0.9096	-0.2564, 0.0669
Hepatomegali	-0.0484	0.0643	0.452	0.9528	-0.1744, 0.0777
Deride Bozukluk	-0.0403	0.0669	0.547	0.9605	-0.1713, 0.0907
Ödem1	0.0610	0.0772	0.429	1.0629	-0.0903, 0.2123
Ödem2	0.2792	0.0930	0.003	1.3221	0.0970, 0.4614
Logb	-0.1769	0.0316	0.000	0.8379	-0.2388, -0.1149
Logalbumin	-0.1120	0.1680	0.505	0.8940	-0.4412, 0.2172
Logprotine	0.3120	0.2808	0.267	1.3662	-0.2384, 0.8624
Histolojik Evre2	-0.0639	0.1944	0.742	0.9381	-0.4450, 0.3171
Histolojik Evre3	0.0520	0.1820	0.775	1.0534	-0.3048, 0.4087
Histolojik Evre4	0.0309	0.1813	0.865	1.0314	-0.3245, 0.3862

Başarısızlık süresini etkileyen faktörleri belirlemek için kullanılan Gamma HBS modeli için elde edilen sonuçlar incelendiğinde yaş ($p=0.000$), ödem2 ($p=0.003$) ve logb ($p=0.000$) değişkenlerinin %95 güven düzeyinde önemli ($p<0.05$) olduğu görülmüştür.

Vücudunda ödem olmayan hastaların ortalama yaşam süresi, vücudunda ödem olan hastaların ortalama yaşam süresine göre yaklaşık 1.3221 ($\exp(0.2792)$) kat daha uzundur. Yaş değişkenindeki 1 birimlik artışın ortalama yaşam süresini 0.9854 birim ve logb değişkenindeki 1 birimlik artışın ise ortalama yaşam süresini 0.8379 birim azalttığı söylenebilmektedir.

5. SONUÇLAR

Yaşam çözümlemesinde durdurma türlerine göre kullanılan yaşam modellerinin teorik yapısı değişiklik göstermektedir. Sağdan durdurulmuş veri yapısı ile sıkça karşılaşılmakla birlikte, literatürde soldan ve aralıklı durdurulmuş veri yapıları ile de karşılaşmak mümkündür. Aralıklı durdurulmuş veri yapısı, bir aralıktaki birimlerin başarısız oldukları kesin zaman noktaları ile ilgili belirsizliği yansıtmaktadır. Aralıklı durdurma ile ilgilenilen birimlerin veya yaşam süresi süreçlerinin sürekli gözlem altında olmadığı, sonuç olarak başarısızlığın ya da yaşam süresinin her zaman tam olarak gözlemlenemediği ya da sağdan durdurulamadığı belirtilmektedir. Aralıklı durdurulmuş bir birim için sadece yaşam olayının gerçekleştiği bir aralık bilinmektedir. Kesin veya sağdan durdurulmuş başarısızlık süreleri, aralıklı durdurulmuş başarısızlık süresinin özel bir durumu olarak kabul edilmektedir, dolayısıyla aralık tek bir noktaya indirgenir. Yani, aralıklı durdurulmuş bir birim, birbiriyle çakışmayan birkaç aralığın birleşimi olarak tanımlanabilir.

Aralıklı durdurulmuş veri yapısı çoğunlukla demografik, epidemiyolojik, finansal, tıbbi ve sosyolojik çalışmalarda yer almaktadır. Literatürde aralıklı durdurulmuş veri yapısı meme kanseri, kalıcı dişin ortaya çıkış süresi, AIDS klinik deneyler, yoğurt için raf ömrü, cep telefonu satın alma, vb. veri kümeleri ile çalışılmıştır (Bogaerts ve diğ., 2018). Aralıklı durdurma, sağdan durdurmadan farklı bir model altyapısı gerektirir. Bu çalışmada, aralıklı durdurulmuş veri yapısı için CRM ve parametrik yaşam modelleri incelenmiş ve sağdan durdurulmuş veri yapısı için kullanılan klasik yöntemler ile karşılaştırma yapılmıştır.

Uygulamada literatürde yer alan Primer Biliyer Siroz verisi için ilk kez aralıklı durdurulmuş veri yapısı dikkate alınarak modelleme yapılmıştır. PBS veri kümesi sağdan durdurulmuş veri yapısı ve aralıklı durdurulmuş veri yapısı ayrı ayrı dikkate alınarak incelenmiştir. Sağdan durdurulmuş veri yapısı için OT varsayımı Schoenfeld artıkları, grafiksel ve ZB açıklayıcı değişkenler yöntemleri ile test edilmiştir. Bunun sonucunda, OT varsayımının sağlandığı, dolayısıyla CRM'nin veri kümesine uygulanabilir olduğu sonucuna varılmıştır. PBS veri kümesi için, CRM'de sağdan durdurulmuş veri yapısına göre yaş, asit, logb değişkenleri

anlamalı çıkmıştır. Sağdan durdurulmuş veri yapısı için parametrik yaşam modellerinin anlamlılığı LR ile test edilmiş ve tüm modeller istatistiksel olarak anlamlı bulunmuştur. Sağdan durdurulmuş veriler için parametrik yaşam modellerine ait AIC ve BIC değerleri incelendiğinde, OT modellerinden Gompertz ve HBS modellerinden ise Weibull en iyi sonucu vermiştir. Sağdan durdurulmuş veri yapısı için parametrik yaşam modellerinden hem Gompertz OT modeli hem de Weibull HBS modeli sonuçlarına göre yaş, asit, logb değişkenleri anlamlı bulunmuştur. PBS veri kümesi için, CRM'de aralıklı durdurulmuş veri yapısına göre yaş ve logb değişkenleri anlamlı çıkmıştır. Aralıklı durdurulmuş veri yapısı için parametrik yaşam modellerinin anlamlılığı LR ile test edilmiş ve tüm modeller istatistiksel olarak anlamlı bulunmuştur. Aralıklı durdurulmuş veriler için parametrik yaşam modellerine ait AIC ve BIC değerleri incelendiğinde, OT modellerinden Gompertz ve HBS modellerinden ise Gamma en iyi sonucu vermiştir. Aralıklı durdurulmuş veri yapısı için parametrik yaşam modellerinden Gompertz OT modeli için yaş, asit, ödem2 ve logb değişkenleri; Gamma HBS modeli için yaş, ödem2 ve logb değişkenleri anlamlı bulunmuştur. Ayrıca sağdan durdurulmuş veri yapısı ile aralıklı durdurulmuş veri yapısına bağlı olarak tehlike oranları da farklılık göstermektedir. Bu sonuçlara göre incelenecek veri kümesi için durdurma türüne doğru bir biçimde karar verilmesi ve modellemelerin buna göre yapılıp, yorumlanmasının önemli olduğu ortaya çıkmıştır.

Bu çalışma sonrasında, aralıklı durdurulmuş veriler için parametrik olmayan yöntemler, Bayesci yaklaşım ve iyileşme modelleri incelenebilir.

KAYNAKLAR

- Andersen, P.K., Borgan, Ø., Gill, R. D. and Keiding, N. Statistical models based on counting processes. New York: Springer-Verlag, **1993**.
- Arjas, E. A grafical method for assessing goodnes of fit in Cox's proportional hazards model. Journal of the American Statistical Association, 83, 204-212, **1988**.
- Ata, N. Orantısız Tehlikeler için Yaşam Modelleri, Doktora Tezi, Hacettepe Üniversitesi, **2010**.
- Barlow, R. E. and Proschan, F. Importance of System Components and Fault Tree Events. Stochastic Processes and their Applications 3, **1975**.
- Beadle, G. F., Silver, B., Botnick, L., Hellman, S. and Harris, J. R. Cosmetic results following primary radiation therapy for early breast cancer. Cancer, 54(12), 2911-2918, **1984**.
- Bebchuk, J. D., Betensky, R. A. Multiple imputation for simple estimation of the hazard function based on interval censored data. Statistics in Medicine, **2000**.
- Bergman, A. C. icenReg: regression models for interval censored data in R. Journal of Statistical Software, Volume 81, Issue 12, November **2017**.
- Berkson, J. and Gage, R. P. Calculation of survival rates for cancer. Proceedings of the staff meetings of the Mayo clinic 25, 270.286, **1950**.
- Betensky, R. A. and Finkelstein, D. M. A non-parametric maximum likelihood estimator for bivariate interval censored data. Statistics in Medicine, 18(22), 3089-3100, **1999**.

- Betensky, R. A., Lindsey, J. C., Ryan, L. M., Wand, M. P. Local EM estimation of the hazard function for interval-censored data. *Biometrics*, **1999**.
- Betensky, R. A., Rabinowitz, D., Tsiatis, A. A. Computationally simple accelerated failure time regression for interval censored data. *Biometrika*, **2001**.
- Bogaerts, K., Komàrek, A., Lesaffre, E. Survival analysis with interval-censored data. A practical approach with examples in R, SAS, and BUGS. CRC Press by Taylor and Francis Group, LLC, 1-219, **2018**.
- Böhning, D., Schlattmann, P., Dietz, E. Interval censored data: A note on the nonparametric maximum likelihood estimator of the distribution function. *Biometrika*, **1996**.
- Carvalho, J. C., Ekstrand, K. R. and Thylstrup, A. Dental plaque and caries on occlusal surfaces of first permanent molars in relation to stage of eruption. *Journal of Dental Research*, 68(5), 773-779, **1989**.
- Cox, D. R. Regression Models and Life-Tables. *Journal of the Royal Statistical Society, Series B*, 34(2), 187-220, **1972**.
- Delord, M. MIICD: Data augmentation and multiple imputation for interval censored data [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=MIICD> (R package version 2.4), **2017**.
- Diamond, I. D., McDonald, J. W. and Shah, I. H. Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan. *Demography*, 23, 607-620, **1986**.
- Diamond, I. D. and McDonald, J. W. The analysis of current status data. *Demographic Applications of Event History Analysis*, eds. Trussel, J., Hankinson, R. and Tilton, J. Oxford University Press: Oxford, U.K., **1991**.

- Dil, E. Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, **2019**.
- Dinse, G. E. and Lagakos, S. W. Regression analysis of tumor prevalence data. *Applied Statistics*, 32, 236-248, **1983**.
- Fang, H.-B., Sun, J. and Lee, M.- L. T. Nonparametric survival comparisons for interval-censored continuous data. *Statistica Sinica*, 12(4), 1073-1083, **2002**.
- Fay, M. P. and Shaw, P. A. Exact and asymptotic weighted logrank tests for interval censored data: The interval R package. *Journal of Statistical Software*, 36(2), 1-34. Retrieved from <http://www.jstatsoft.org/v36/i02>, **2010**.
- Fay, M. P. interval: Weighted logrank tests and NPMLE for interval-censored data [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=interval> (R package version 1.1-0.1), **2014**.
- Finkelstein, D. M., Wolfe, R. A. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, **1985**.
- Finkelstein, D. M. A proportional hazards model for interval-censored failure time data. *Biometrics*, **1986**.
- Fleming, T. R. and Harrington, D. P. Counting processes and survival analysis. New York: John Wiley and Sons, **1991**.
- Geskus, R. And Groeneboom, P. Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *The Annals of Statistics*, **1999**.
- Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and Duchateau, L. Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(1), 1-14, **2009**.

- Goggins, W. B., Finkelstein, D. M., Schoenfeld, D. A., Zaslavsky, A. M. A Markov chain Monte Carlo EM algorithm for analyzing interval censored data under the Cox proportional hazards model. *Biometrics*, **1998**.
- Goggins, W. B. and Finkelstein, D. M. A proportional hazards model for multivariate interval-censored failure time data. *Biometrics*, 56, 940-943, **2000**.
- Gomez, G., Calle, M. L. and Oller, R. Frequentist and Bayesian approaches for interval-censored data. *Statistical Papers*, 45(2), 139-173, **2004**.
- Gomez, G., Calle, M. L., Oller, R. and Langohr, K. Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, 9(4), 259-297, **2009**.
- Groeneboom, P. and Wellner, J. A. Information bounds and nonparametric maximum likelihood estimation. Basel: Birkhäuser-Verlag, **1992**.
- Guo, C., So, Y., Johnston, G. Analyzing Interval-Censored Data with the ICLIFETEST Procedure. Paper SAS279-2014, **2014**.
- Harrell, F. E. PHGLM procedure, SAS supplemental Library User's Guide (5nd ed.). Cary: SAS Institute, **1986**.
- Heller, G., Proportional Hazards Regression with Interval Censored Data Using an Inverse Probability Weight, *Lifetime Data Analysis*, **2011**.
- Hoel, D. G. and Walburg, H. E. Statistical analysis of survival experiments. *Journal of National Cancer Institute*, 49, 361-372, **1972**.
- Hosmer, D. W., Lemeshow, S. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, Wiley and Sons, New York, **1999**.

- Hough, G. Sensory shelf life estimation of food products. Boca Raton: CRC Press, **2010**.
- Huang, J. and Wellner, J. A. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. *Statistica Neerlandica*, **1995**.
- Huang, J. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, **1996**.
- Huang, J. and Rossini, A. J. Sieve estimation for the proportional odds failure-time regression model with interval censoring. *Journal of the American Statistical Association*, **1997**.
- Huang, J. and Wellner, J. A. Interval censored survival data: a review of recent progress. *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, eds. Lin, D. and Fleming, T. Springer-Verlag, New York, 123-169, **1997**.
- Kalbfleisch, J. D. and Prentice, R. L. The statistical analysis of failure time data. New York: John Wiley and Sons, **1980**.
- Kalbfleisch, J. D. and Prentice, R.L. The Statistical Analysis of Failure Time Data, Second Edition, Wiley and Sons, New York, **2002**.
- Kaplan, E. and Meier, P. Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* 53, 457-481, **1958**.
- Karasoy, D. ve Sezayi, S. Yaşam Çözümlemesinde Hızlandırılmış Başarısızlık Süresi Modelleri ve Bir Uygulama. *Süleyman Demirel Üniversitesi*, 18(1), 1-7, **2014**.

- Karasoy, D. ve Kaplan, S. K. Yaşam Çözümlemesinde Eş Zamanlı Yaşam Süreleri. Alphanumeric Journal. The Journal of Operations Research, Statistics, Econometrics and Management Information Systems, Volume 5, Issue 1, **2017**.
- Karvanen, J., Rantanen, A. and Luoma, L. Survey data and bayesian analysis: a cost-efficient way to estimate customer equity. Quantitative Marketing and Economics, 12(3), 305-329, **2014**.
- Keiding, N. Age-specific incidence and prevalence: A statistical perspective (with discussion). Journal of the Royal Statistical Society, Series A, 154, 371-412, **1991**.
- Keiding, N., Begtrup, K., Scheike, T. H. and Hasibeder, G. Estimation from current status data in continuous time. Lifetime Data Analysis, 2, 119-129, **1996**.
- Kim, J., Kang, D. R., Nam, C. M. Logrank-type tests for comparing survival curves with interval-censored data. Computational Statistics and Data Analysis, **2006**.
- Kim, M. Y., De Gruttola, V. G. and Lagakos, S. W. Analyzing doubly censored data with covariates, with application to AIDS. Biometrics, 49(1), 13-22, **1993**.
- Kim, Y. and Jhun, M. Cure rate model with interval censored data. Statistics in Medicine, **2008**.
- Klein, J. P. and Moeschberger, M. L. Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York, 502 pp, **1997**.
- Komàrek, A. and Lesaffre, E. Bayesian accelerated failure time model for correlated interval-censored data with a normal mixture as error distribution. Statistica Sinica, **2007**.

- Komàrek, A. and Lesaffre, E. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, **2008**.
- Lawless, J. F. and Babineau, D. Models for interval censoring and simulation-based inference for lifetime distributions. *Biometrika*, **2006**.
- Lee, E. T. and Wang, J. W. *Statistical Methods for Survival Data Analysis*, Wiley and Sons, New York, 513 pp, **2003**.
- Li, L., Watkins, T., Yu, Q. An EM algorithm for estimating survival functions with interval-censored data. *Scandinavian Journal of Statistics*, **1997**.
- Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50, 163-170, **1966**.
- Mayo Clinic, Primary Biliary Cirrhosis, sequential data,
<https://www.mayo.edu/research/documents/pbcseqhtml/doc-10027141>
(Eriřim tarihi: **Ocak 2019**).
- Meyns, B., Jashari, R., Gewillig, M., Mertens, L., Komàrek, A., Lesaffre, E., Budts, W. and Daenen, W. Factors influencing the survival of cryopreserved homografts. The second homograft performs as well as the first. *European Journal of Cardio-thoracic Surgery*, 28(2), 211-216, **2005**.
- Murtaugh, P., Dickson, E. R., Dam, G. V., Malinchoc, M., Grambsch, P. M., Langworthayn A. and Gips, C. H., Primary Biliary Cirrhosis: Prediction of Short-Term Survival Based On Repeated Patient Visits, *Hepatology*, 20 126-134, **1994**.
- Oller, R., Gómez, G., Calle, M. L. Interval censoring: model characterizations for the validity of the simplified likelihood. *The Canadian Journal of Statistics*, **2004**.

- Pan, W. A multiple imputation approach to Cox regression with interval-censored data. *Biometrics*, **2000**.
- Qi, J. Comparison of Proportional Hazards and Accelerated Failure Time Models. MSc Thesis, Department of Mathematics and Statistics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, **2009**.
- Rabinowitz, D., Tsiatis, A. A., Aragon, J. Regression with interval-censored data. *Biometrika*, **1995**.
- Radhey, S. S. and Totawattage, D. P. The Statistical Analysis of Interval-Censored Failure Time Data with Applications, *Journal of Statistics*, 2013, 3, 155-166, **2013**.
- Satten, G. A. Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, **1996**.
- Schick, A. and Yu, Q. Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics*, 27(1), 45-55, **2000**.
- Schoenfeld, D. Partial residuals for the proportional hazards model. *Biometrika*, 69, 551-555, **1982**.
- Sen, B. and Banerjee, M. A pseudolikelihood method for analyzing interval censored data. *Biometrika*, **2007**.
- Shiboski, S. C. and Jewell, N. P. Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of the American Statistical Association*, 87, 360-372, **1992**.
- Singh, R. S. and Totawattage, D. The statistical analysis of interval-censored failure time data with applications. *Open Journal of Statistics*, 3, 155-166, **2013**.

- So, Y., Johnston, G. and Kim, S. H. Analyzing Interval Censored Survival Data, SAS Global Forum Paper 257, **2010**.
- Sun, J. Interval censoring. Encyclopedia of Biostatistics, John Wiley, First Edition, 2090-2095, **1998**.
- Sun, J. Statistical analysis of doubly interval-censored failure time data. In: Balakrishnan N, Rao CR, editors. Handbook of statistics: survival analysis, **2002**.
- Sun, J. Statistical analysis of doubly interval-censored failure time data. Advances in survival analysis. Handbook of biostatistics, **2004**.
- Sun, J., Zhao, Q., Zhao, X. Generalized log rank tests for interval-censored failure time data. Scandinavian Journal of Statistics, **2005**.
- Sun, J. The statistical analysis of interval-censored failure time data. Statistics for Biology and Health, Springer Science+Business Media, Inc. 1-152, **2006**.
- Tian, L. and Cai, T. On the accelerated failure time model for current status and interval censored data. Biometrika, **2006**.
- Turnbull, B. W. The empirical distribution with arbitrarily grouped censored and truncated data. Journal of the Royal Statistical Society: Series B, **1976**.
- Vandal, A. C., Gentleman, R., Liu, X. Constrained estimation and likelihood intervals for censored data. The Canadian Journal of Statistics, **2005**.
- Wellner, J. A. Interval censoring case 2: alternative hypotheses. Analysis of Censored Data (Pune, 1994/1995), eds. H. L. Koul and J. V. Deshpande, IMS Lecture Notes, Monograph Series 27, 271-219, **1995**.

- Xue, H., Lam, K. F., Cowling, B. J., de Wolf, F. Semi-parametric accelerated failure time regression analysis with application to interval-censored HIV/AIDS data. *Statistics in Medicine*, **2006**.
- Yang, X. Analyzing interval-censored survival-time data in Stata. Senior Statistician and Software Developer StataCorp LLC, Stata Conference, **2017**.
- Younes, N. and Lachin, J. Linked-based models for survival data with interval and continuous time censoring. *Biometrics*, **1997**.
- Yu, Q., Li, L., Wong, G. On consistency of self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, **2000**.
- Yuen, K., Shi, J. and Zhu, L. A k-sample test with interval-censored data. *Biometrika*, **2006**.
- Zeng, D., Cai, J., Shen, Y. Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, **2006**.
- Zhang, Z., Sun, L., Zhao, X., Sun, J. Regression analysis of interval censored failure time data with linear transformation models. *The Canadian Journal of Statistics*, **2005**.
- Zhang, Z., Sun, L., Sun, J., Finkelstein, D. M. Regression analysis of failure time data with informative interval censoring. *Statistics in Medicine*, **2007**.
- Zhao, Q., Sun, J. Generalized log-rank test for mixed interval-censored failure time data. *Statistics in Medicine*, **2004**.
- Zhao, X., Zhao, Q., Sun, J., Kim, J. S. Generalized log-rank tests for partly interval-censored failure time data. *Biometrical Journal*, **2009**.

Zhu, L., Tong, X., Sun, J. A transformation approach for the analysis of interval-censored failure time data. *Lifetime Data Analysis*, **2008**.



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 20/06/2019

Tez Başlığı / Konusu: Aralıklı Durdurulmuş Veriler ile Yaşam Çözümlemesi

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 66 sayfalık kısmına ilişkin, 20/06/2019 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %9'dur.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dâhil
- 3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

20.06.2019

Adı Soyadı: İlknur ERÖZ

Öğrenci No: N16126731

Anabilim Dalı: İstatistik

Programı: Yüksek Lisans

Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Doç. Dr. Nihal ATA TUTKUN

ÖZGEÇMİŞ

Adı Soyadı : İlknur ERÖZ
Doğum Yeri : Ankara
Medeni Hali : Bekar
E-posta : ilknureroz@hacettepe.edu.tr
Adresi : Ankara

Eğitim Durumu

Lisans : 2006-2011 Hacettepe Üniversitesi İstatistik Bölümü
Yüksek Lisans : 2016-2019 Hacettepe Üniversitesi İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce, iyi seviye
Almanca, orta seviye
Korece, temel seviye
Kazakça, temel seviye

İş Deneyimi

2013-2014 : Halkbank
2014- : Makina ve Kimya Endüstrisi Kurumu

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar