



TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ



TAKİP ARAŞTIRMALARINDA TEKRARLANAN
BAŞARISIZLIKLARIN MODELLENMESİ:
GÖĞÜS HASTALIKLARI VERİ SETİ
ÜZERİNDE BİR UYGULAMA

Şirin ÇETİN

BIYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ

DANIŞMAN
Doç. Dr. S. Kenan KÖSE

ANKARA
2017

TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

TAKİP ARAŞTIRMALARINDA TEKRARLANAN
BAŞARISIZLIKLARIN MODELLENMESİ:
GÖĞÜS HASTALIKLARI VERİ SETİ
ÜZERİNDE BİR UYGULAMA

Şirin ÇETİN

BIYOİSTATİSTİK ANABİLİM DALI
DOKTORA TEZİ

DANIŞMAN
Doç. Dr. S. Kenan KÖSE

ANKARA
2017

ETİK BEYAN

Ankara Üniversitesi

Sağlık Bilimleri Enstitüsü Müdürlüğü'ne,

Doktora tezi olarak hazırlayıp sunduğum "Takip Araştırmalarında Tekrarlanan Başarısızlıkların Modellenmesi: Göğüs Hastalıkları Veri Seti Üzerinde Bir Uygulama" başlıklı tez; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin fikir/hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan araştırma tarafımdan yapılmış olup, tüm cümleler, yorumlar bana aittir.

Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Öğrencinin Adı Soyadı: Şirin ÇETİN


Tarih:

İmza:

KABUL ve ONAY

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü
Biyostatistik Anabilim Dalında
Şirin ÇETİN tarafından "Takip Araştırmalarında Tekrarlanan Başarısızlıkların Modellenmesi:
Göğüs Hastalıkları Veri Seti Üzerinde Bir Uygulama" adlı tez çalışması aşağıdaki jüri
tarafından DOKTORA TEZİ olarak OY BİRLİĞİ/OY ÇOKLUĞU ile kabul edilmiştir.

Tez Savunma Tarihi:16.01.2017



Prof.Dr.Yavuz SANİSOĞLU
Yıldırım Beyazıt Üniversitesi
Biyostatistik Anabilim Dalı
Jüri Başkanı



Doç.Dr.Derya ÖZTUNA
Ankara Üniversitesi Tıp Fakültesi
Biyostatistik Anabilim Dalı
Raportör



Doç.Dr.S.Kenan KÖSE
Ankara Üniversitesi Tıp Fakültesi
Biyostatistik Anabilim Dalı
Danışman



Doç.Dr. Erdem KARABULUT
Hacettepe Üniversitesi Tıp Fakültesi
Biyostatistik Anabilim Dalı
Üye



Yrd.Doç.Dr. Beyza DOĞANAY ERDOĞAN
Ankara Üniversitesi Tıp Fakültesi
Biyostatistik Anabilim Dalı
Üye

Tez hakkında alınan jüri kararı, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu tarafından onaylanmıştır.

Prof. Dr. K. Zafer KARAER
Sağlık Bilimleri Enstitüsü Müdürü

İÇİNDEKİLER

Etik Beyan	ii
Kabul ve Onay	iii
İçindekiler	iv
Önsöz	vii
Simgeler ve Kısaltmalar	viii
Şekiller	x
Çizelgeler	xii
1. GİRİŞ	1
1.1. Yaşam Süresi	1
1.1.1. Yaşam Süresi Verilerinin Analizinde Kullanılan Fonksiyonlar	2
1.1.1.1. Olasılık Yoğunluk Fonksiyonu	3
1.1.1.2. Yaşam Fonksiyonu	3
1.1.1.3. Hazard Fonksiyonu	4
1.1.1.4. Birikimli Hazard Fonksiyonu	4
1.2. Sansürlü Veri	5
1.2.1. Sağdan Sansürleme	5
1.2.2. Tip 1 Sansürleme	6
1.2.3. Tip 2 Sansürleme	6
1.2.4. Soldan Sansürleme	6
1.2.5. Çifte Sansürleme	6
1.2.6. Aralık Sansürlemesi	6
1.3. Yaşam Analizinde Etki Büyüklüğü ve Güç	7
1.4. Cox Regresyon Modeli	7
1.4.1. Hazardların Orantısallığı	9
1.4.1.1. Cox Regresyon Modelinin Orantılı Hazard Varsayımının İncelenmesi	9
1.4.2. Model Seçim Kriterleri	11
1.4.3. Tabakalandırılmış Cox Regresyon Modeli	11
1.4.3.1. Etkileşimsiz Tabakalandırılmış Cox Regresyon Modeli	12

1.4.3.2. Etkileşimli Tabakalandırılmış Cox Regresyon Modeli	12
1.4.3.3. Tabakalandırılmış Cox Regresyon Modelinde Etkileşimsizlik Varsayımı	13
1.4.4. Genişletilmiş Cox Regresyon Modeli	14
1.5. Yaşam Analizinde Kullanılan Bazı Dağılımlar	15
1.5.1. Üstel Dağılım	15
1.5.2. Weibull Dağılımı	17
1.5.3. Log-normal Dağılım	19
1.5.4. Log-Logistik Dağılım	21
1.5.5. Gamma Dağılımı	23
1.5.6. Gompertz Dağılımı	25
1.6. Parametre Tahmin Yöntemleri	26
1.6.1. Newton-Raphson Yöntemi	27
1.6.2. Beklenti En Büyükleme (EM) Algoritması	27
1.6.3. Monte Carlo Beklenti En Büyükleme (MCEM)	28
1.7. Cox Regresyon Modelinde Parametre Tahmini	28
1.7.1. Breslow'un Kısmi Olabilirliği	30
1.7.2. Efron'un Kısmi Olabilirliği	30
1.8. Tekrarlanan Başarısızlıklar	32
1.8.1. Tekrarlanan Başarısızlıkların Modellenmesi	33
1.8.2. Sayma Süreci	33
1.8.3. Prentice, Williams ve Peterson Yaklaşımı	36
1.8.3.1. Koşul 1 Yaklaşımı	37
1.8.3.2. Koşul 2 Yaklaşımı	37
1.8.4. Wei, Lin ve Weissfeld Yaklaşımı	38
1.8.5. Andersen ve Gill Yaklaşımı	39
1.8.6. Zayıflık Modeli	39
1.8.6.1. Paylaşılmamış Zayıflık Modeli	42
1.8.6.2. Paylaşılmış Zayıflık Modeli	44
1.8.6.3. İlişkili Zayıflık Modeli	45
1.8.6.4. Gamma Zayıflık Modeli	45
1.8.6.5. Ters - Gauss Zayıflık Modeli	49

2. GEREÇ VE YÖNTEM	52
3. BULGULAR	57
3.1. Yaşam Sürelerinin Kaplan-Meier Analizi ile Kestirimi	58
3.2. Oransal Hazard Varsayımı Denetimi	61
3.3. Tekrarlanan Başarısızlıkta Kullanılacak Analiz Yönteminin Seçimi	62
3.4. Zayıflık Modelinin Seçimi	62
3.5. Zayıflık Teriminde Kullanılacak Dağılımın Belirlenmesi	62
3.6. Gamma Zayıflık Terimi İçeren Weibull Dağılımı ile Analiz	63
3.7. Zayıflık Modeli ile Elde Edilen Bulgular	65
3.8. Zayıflık Terimlerinin Bulunması	66
3.9. Tabakalı Cox Yaklaşımı ve Zayıflık Modelinin Karşılaştırılması	67
4. TARTIŞMA	69
5. SONUÇ VE ÖNERİLER	75
ÖZET	77
SUMMARY	78
KAYNAKLAR	79
ÖZGEÇMİŞ	86

ÖNSÖZ

Bu çalışmada tekrarlanan başarısızlıklardan elde edilen yaşam sürelerinin yaşam analizi yöntemleriyle modellenmesi yapıldı. Bu modellemenin araştırmacılara farklı bir bakış açısı getireceği düşünülmektedir.

Yaşam analizinde tekrarlanan başarısızlıkları modelleyebilecek için birçok yöntem geliştirilmiştir. Bu yöntemlerden birisi zayıflık modelidir. Zayıflık modelleri yaşam süreleri ile ilgili modelleme çalışmalarında, tekrarlayan başarısızlıkların modellenmesinde ve bireyler arasındaki heterojenliğin açıklanmasında kullanılabilir. Bu çalışmada zayıflık modeli tüm ayrıntılarıyla açıklamaya çalışıldı.

Doktora eğitimim ve tezimin hazırlanma sürecinde sürekli desteğini gördüğüm değerli danışman hocam Doç. Dr. S. Kenan Köse'ye içtenlikle teşekkür ederim.

Ankara Üniversitesi Tıp Fakültesi Göğüs Hastalıkları Anabilim Dalı öğretim üyelerinden Prof. Dr. Oya Karacan'a Arş. Görv. Burak Ekin Dalbayrak'a ve Uzman Dr. Fatma Çiftçi'ye teşekkür ederim.

SİMGELER VE KISALTMALAR

A-G	: Andersen-Gill
AIC	: Akaike bilgi kriteri
AKG	: Arteryel kan gazı
BIC	: Bayesci bilgi kriteri
Exp	: Üstel
F(t)	: Dağılım fonksiyonu
f(t)	: Olasılık yoğunluk fonksiyonu
FEV ₁	: Birinci saniyedeki zorlu ekspiratuar volüm
FVC	: Zorlu vital kapasite
H(t)	: Birikimli hazard fonksiyonu
h(t)	: Hazard fonksiyonu
H(t;x)	: Cox regresyon modeli
h ₀ (t)	: Temel hazard fonksiyonu
HR	: Hazard oranı
KOAH	: Kronik obstrüktif akciğer hastalığı
LL	: Log likelihood
log	: Logaritma
MCEM	: Monte Carlo Expected Maximization
MMF	: Maksimal ekspirasyon ortası akım hızı
mMRC	: Modified Medical Research Council scale
PaCO ₂	: Parsiyel karbondioksit basıncı
PaO ₂	: Parsiyel oksijen basıncı
P _h	: Hidrojen potansiyeli
PWP-GT	: Prentice-Williams-Peterson gap time
S(t)	: Yaşam fonksiyonu
SaO ₂	: Oksijen satürasyonu
SFT	: Solunum fonksiyon testi
sPAB	: Sistolik pulmoner arter basıncı
T, t	: Yaşam süresi

VKI	: Vücut kitle indeksi
WLW-TT	: Wei-Lin-Weissfeld total time
Z	: Zayıflık terimi
α	: Alpha
β	: Beta
δ, Δ	: Delta
ϵ	: Epsilon
η	: Etha
λ	: Lambda
μ	: Mu
ρ	: Rho
σ	: Sigma
Φ	: Phi

ŞEKİLLER

Şekil 1.1.	Üstel Dağılımın Olasılık Yoğunluk Eğrileri	16
Şekil 1.2.	Üstel Dağılımın Olasılık Yaşam Eğrileri	16
Şekil 1.3.	Üstel Dağılımın Hazard Fonksiyon Eğrileri	17
Şekil 1.4.	$\lambda=1$ ve $\gamma=3, 1.5$ ve 0.5 İçin Olasılık Yoğunluk Eğrileri	18
Şekil 1.5.	$\lambda=1$ ve $\gamma=3, 1.5$ ve 0.5 için Hazard Fonksiyon Eğrileri	19
Şekil 1.6.	Log-normal Dağılımın $\mu = 0, \sigma = 0.5, 1$ ve 2 Değerleri için Olasılık Yoğunluk Eğrileri	20
Şekil 1.7.	Log-normal Dağılımın $\sigma = 1, \mu = 0, 0.5$ ve 1.5 Değerleri için Olasılık Yoğunluk Eğrileri	20
Şekil 1.8.	Log-normal Dağılımın Hazard Fonksiyon Eğrileri	21
Şekil 1.9.	Log-logistik Dağılımın Olasılık Yoğunluk Eğrileri	22
Şekil 1.10.	Log-logistik Dağılımın Olasılık Yaşam Eğrileri	22
Şekil 1.11.	Log-logistik Dağılımın Farklı Gamma Değerleri için Hazard Fonksiyonu	23
Şekil 1.12.	$\lambda=1$ ve $\gamma=1,2,3$ olduğunda Gamma Dağılımının Olasılık Yoğunluk Eğrileri	24
Şekil 1.13.	$\gamma=3$ ve $\lambda=1,3,5$ olduğunda Gamma Dağılımının Olasılık Yoğunluk Eğrileri	24
Şekil 1.14.	$\lambda=1$ ve $\gamma=0.5, 1, 2$ ve 4 Olduğu Durumda Hazard Fonksiyon Eğrileri	25
Şekil 2.1.	KOAH Hastalarında Tekrarlanan Başarısızlık Zamanı Verisi	54
Şekil 3.1.	KOAH Hastalarının Birinci Atak İçin Yaşam Eğrisi	58
Şekil 3.2.	KOAH Hastalarının İkinci Atak İçin Yaşam Eğrisi	59
Şekil 3.3.	KOAH Hastalarının Üçüncü Atak İçin Yaşam Eğrisi	59
Şekil 3.4.	KOAH Hastalarının Dördüncü Atak İçin Yaşam Eğrisi	60
Şekil 3.5.	KOAH Hastalarının Beşinci Atak İçin Yaşam Eğrisi	60
Şekil 3.6.	KOAH Hastalarının Altıncı Atak İçin Yaşam Eğrisi	61
Şekil 3.7.	mMRC Değişkeninin Schoenfeld Grafiği	61
Şekil 3.8.	AIC ve BIC Değerlerinin Grafik Üzerinde Karşılaştırılması	63

Şekil 3.9. Gamma Zayıflık Terimi İçeren Weibull Dağılımının Zayıflık Terimlerinin Tahmini	66
Şekil 3.10. Gamma Zayıflık Terimi İçeren Weibull Dağılımının Zayıflık Terimlerinin Tahmininin Doğrusal Grafikle Gösterimi	67



ÇİZELGELER

Çizelge 1.1.	Sayma Süreci Yaklaşımı Veri Düzenlenmesi	35
Çizelge 1.2.	Koşul 2 Yaklaşımı Veri Girişi	37
Çizelge 1.3.	Marjinal Yaklaşımında Veri Girişi	38
Çizelge 2.1.	GOLD Evrelemesi	53
Çizelge 2.2.	Modifiye MRC Skalası (mMRC)	54
Çizelge 2.3.	KOAH Hastalarından Veri Elde Edilmesi	55
Çizelge 2.4.	Tekrarlanan Başarısızlık Veri Girişi	55
Çizelge 3.1	Hastaların Demografik ve Klinik Özellikleri	57
Çizelge 3.2.	AIC ve BIC Değerlerinin Tabloda Gösterimi	63
Çizelge 3.3.	Gamma Zayıflık Terimi İçeren Weibull Dağılımı ile Analiz Sonuçları	65
Çizelge 3.4.	Tabakalı Cox Modeli İçin Veri Girişi	68
Çizelge 3.5.	Tabakalı Cox Yaklaşımı ile Analiz Sonuçları	68
Çizelge 3.6.	Zayıflık Modeli ile Analiz Sonuçları	68

1. GİRİŞ

Yaşam analizi, sosyal ve doğal bilimlerde olayları incelemek için kullanılan bir yöntemdir. Mesela; endüstride bir ürünün parçalarının bozulması, hastalığın başlangıcı, ölüm gibi olayları analiz etmek için yaşam analizi yöntemleri kullanılmaktadır. Bu yöntemler farklı alanlarda farklı isimler ile ifade edilmektedir (Allison, 2010).

Yaşam analizi, sağlık alanında farklı tedavi yöntemleri uygulanan bireylerin hastalıklarının tekrarlaması ya da ölüm gibi olayların süresi ile ilgilidir. Yaşam analizinde araştırılan konu; bir tür tedavi alan hastalar diğer tür tedavi alan hastalara göre daha uzun yaşam süresine mi sahipler? Hastanın yaşam süresi yaş, cinsiyet, sigara veya ek hastalıklarına mı bağlı? Akciğer kanseri teşhisi konulmuş bir hastanın sağ kalım süresi ne kadardır? Kemoterapi uygulandığında sağ kalım süresi daha mı uzundur?, şeklindedir.

Literatür incelendiğinde zaman akışı ile ilgili iki tür analizle karşılaşılır. Bunlardan birincisi yaşam analizi diğeri ise zaman-serileri analizidir. Yaşam analizi bir olayın gerçekleşmesi için ne kadar zamanın geçtiğini araştırırken; zaman-serisi analizi ise bağımlı değişken üzerinde zaman içindeki değişiklikleri araştırmaktadır (Tabachnick ve Fidell, 2013).

1.1. Yaşam Süresi

Bir olayın başlangıç noktası ile başarısızlığı arasında geçen zamana yaşam süresi adı verilmektedir (Johnson ve Johnson, 1980).

Yaşam analizinde araştırma sonlandırıldığında bazı hastaların yaşam süresi (bağımlı değişken) hâlen bilinmemektedir. Çünkü araştırmanın sonunda bazı hastalar henüz başarısız olmadıkları için hala yaşamaktadırlar. Bazı hastalar ise araştırmayı bıraktıkları ya da ulaşamıyor olduğundan dolayı yine sağ kalım süresi bilinmemektedir. Herhangi bir sebepten dolayı bağımlı değişken değerleri

bilinmeyen hastalar yaşam analizinde sansürlü veri olarak adlandırılır ve bu verilerin analizi diğer verilerin analizlerden farklılık gösterirler.

Yaşam analizinde olay yaşam süresi analiz edilmeden önce ortaya çıkmalıdır. Fakat tıbbi bir olay araştırmasında takip süresinde tedavi ne kadar başarılı ise araştırmada veri toplama süresi o kadar uzar ve gecikir.

Yaşam analizinin temel amacı, bir ya da daha fazla grup için ayrı ayrı farklı zamanlarda sağ kalım oranını belirlemektir.

Yaşam analizinde ilgilenilen, gözlenen başarısızlık süreleri olduğundan, başarısızlık süresi değişkeninin açık bir şekilde tanımlanması gerekmektedir.

- Başlangıç zamanı her bir denek için şüpheye yer verilmeyecek biçimde kesin tanımlanmalıdır.
- Sürenin belirlenmesi için bir zaman ölçeği belirlenmelidir.
- Her bir denek için başarısızlığın meydana geldiği zaman net olarak belirlenmelidir (Karasoy ve Tutkun, 2016).

Yeni doğmuş bir bebeğin gelecekteki yaşam süresi bir T değişkeni ise; T'ye yaşam süresi denir. Olasılık teorisinden yaşam süresinin dağılım fonksiyonu

$$P(T \leq t) = F(t) \quad (1.1)$$

şeklinde yazılabilir.

1.1.1. Yaşam Süresi Verilerinin Analizinde Kullanılan Fonksiyonlar

Yaşam analizinde dört temel fonksiyon kullanılmaktadır. Bunlar,

- Olasılık yoğunluk fonksiyonu
- Yaşam fonksiyonu
- Hazard fonksiyonu
- Birikimli hazard fonksiyonudur.

1.1.1.1. Olasılık Yoğunluk Fonksiyonu

Başarısızlık zamanı T 'nin olasılık yoğunluk fonksiyonu,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}, \quad 0 < t < \infty \quad (1.2)$$

şeklindedir. T 'nin dağılım fonksiyonu ise;

$$F(t) = P(T \leq t) = \int_0^t f(x) dx, \quad 0 < t < \infty \quad (1.3)$$

şeklindedir .

1.1.1.2. Yaşam Fonksiyonu

Yaşam fonksiyonu $S(t)$ ile gösterilir ve

$$S(t) = 1 - F(t) \quad (1.4)$$

olarak tanımlanır.

$$\begin{aligned} S(t) &= 1 - P(T \leq t) \\ &= P(T > t) \end{aligned} \quad (1.5)$$

dir. O halde $S(t)$ bir bireyin t 'den daha çok süre yaşama olasılığıdır. (1.3)'ten $S(t)$, T 'nin dağılımını tek bir biçimde tanımlar. T sürekli bir rastgele değişken olduğundan (1.4) eşitliği

$$S(t) = P(T \geq t) \quad (1.6)$$

şeklinde de ifade edilebilir (Akdeniz ve Erdugan, 2014).

Yaşam fonksiyonu eşitlik 1.7 de ifade edildiği gibidir.

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x) dx \quad 0 < t < \infty \quad (1.7)$$

şeklinde gösterilir.

1.1.1.3. Hazard Fonksiyonu

Hazard fonksiyonu, t zamanına kadar hayatta olan bir bireyin $[t, t+\Delta t]$ zamanda hayatının sonlanmasının tanımıdır. Kısacası herhangi bir zamandaki başarısızlık riski olarak tanımlanmaktadır. Hazard fonksiyonu, başarısızlık hızı, ani ölüm hızı ve ölümlülük gücü olarak da isimlendirilmektedir. Hazard fonksiyonu;

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \quad (1.8)$$

şeklindedir. Bir bireyin t anında yaşamını sürdürdüğü biliniyorsa, aynı bireyin $[t, t+\Delta t]$ zaman aralığında yaşamının sona erme olasılığı, zaman aralığı ile orantılı olmak üzere $h(t)\Delta t$ şeklindedir. Sürekli dağılımlar için

- $h(t) > 0$ (1.9)

- $\int_0^{\infty} h(t) dt = \infty$ (1.10)

şeklindedir.

1.1.1.4. Birikimli Hazard Fonksiyonu

Birikimli hazard fonksiyonu, $H(t)$ ile gösterilir ve

$$H(t) = \int_0^t h(x) dx \quad (1.11)$$

şeklinde ifade edilir. Birikimli hazard fonksiyonu

$$H(t) = -\log S(t) \quad (1.12)$$

ile ifade edilir. $H(t)$ aşağıdaki özellikleri sağlar;

- Artan bir fonksiyondur
- $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$ olduğundan $\lim_{t \rightarrow \infty} H(t) = \infty$ şeklindedir.
- Sağdan sürekli bir fonksiyondur.

1.2. Sansürlü Veri

Yaşam verilerinin analizinde karşılaşılan temel güçlük tüm bireylerin başarısızlık sürelerinin gözlenememiş olmasından kaynaklanmaktadır. Tıbbi bir çalışmada takip edilen hastaların bazıları takip süresi sonunda hâlen yaşamlarını sürdürüyor olabilir. Takip edilen bir birey bazı sebeplerden dolayı çalışmayı terk edebilir. Eğer başarısızlık süresi bu gibi nedenlerden dolayı tamamlanmamışsa sansürlü veri oluşur. Diğer bir ifadeyle, yaşam süresi tam olarak bilinmeyen tamamlanmamış ya da durdurulmuş gözlemlerden oluşan veriye “sansürlü veri” adı verilir. Yaşam analizini diğer yöntemlerden ayıran özelliklerden biri de, sansürlü verinin kullanımına imkân sağlamasıdır. Sansürlü verinin oluştuğu durumlar, takip süresi boyunca başarısız olayın gerçekleşmemesi ya da bireyin deneyden çekilmesi şeklinde ifade edilebilir. Bireylerin araştırmaya farklı zamanda katılmaları sebebiyle yaşam süreleri birbirinden farklı olabilir. Bireyler farklı zamanlarda başarısız olabilir ya da birey sansürlü gözlem olarak araştırmaya girebilir. Yaşam analizinde kullanılan temel sansürleme çeşitleri:

- Sağdan sansürleme
- Tip 1 Sansürleme
- Tip 2 sansürleme
- Soldan sansürleme
- Çifte sansürleme
- Aralık sansürleme

1.2.1. Sağdan Sansürleme

Sağdan sansürleme tipi en çok karşılaşılan sansürleme çeşitidir. Sağdan sansürlemede başarısızlık olayı görülmemiştir. Bu tip sansürlemede yaşam süresinin takip süresinde elde edilen yaşam süresinden daha fazla olduğu bilinir. Başarısızlık olayı çalışmanın belirli durma zamanına kadar gerçekleşmemiştir. Bu durumda yaşam süresi kesin bilinemez.

1.2.2. Tip 1 Sansürleme

Sağdan sansürlemenin özel bir tipidir. Tıpta ve endüstride başarısızlık zamanı deneylerinde sıklıkla karşılaşılmaktadır. Tip 1 sansürleme, deneyin sabit bir tamamlanma zamanında sonlandırılmasıyla oluşur. Bu sansürleme tipinde gözlenen yaşam süreleri rassal değişkenlerdir.

1.2.3. Tip 2 Sansürleme

Bu tip sansürlemede deney, önceden belirlenmiş sayıda olay meydana geldiğinde sona erdirilir. Her birey için yaşam süreleri rassal değişkendir.

1.2.4. Soldan Sansürleme

Bu sansürleme çeşidi ile tanımlanan olayın başlangıç zamanının kesin olarak bilinmediği zamanlarda karşılaşılr. Soldan sansürlü veride yaşam süresi zamanı gözlenen takip süresinden küçüktür ($T < t$).

1.2.5. Çifte Sansürleme

Sağ ve sol sansürün birleştirilmiş şekli olarak düşünülebilir. Bu sansürlemede hem sağ hem de sol sansür söz konusudur. Bu tür sansürlemeye çifte sansürleme denir.

1.2.6. Aralık Sansürlemesi

Birçok durumda ilgilenilen olay gözlemlenmemiş olabilir ve sadece iki zaman aralığı içinde ortaya çıktığı biliniyor olabilir. Böyle bir durumda, aralıklı sansürleme söz konusudur. Bu tip sansürlemede beklenen olay iki zaman noktası arasında bir yerlerde gerçekleşmiştir. Ancak gerçekleşme anı hakkında kesin bir bilgi yoktur.

Aralık sansürlemesi soldan sansürlemeye benzer; ancak incelenen olayın gerçekleşmediği bir zaman aralığı içermesi yönüyle soldan sansürlemeden ayrılır.

Aralık sansürlemesinde yaşam süresi takip süresinden küçüktür ama takibe başlandığında başarısızlık olayı henüz gerçekleşmemiştir.

1.3. Yaşam Analizinde Etki Büyüklüğü ve Güç

Eliason (1993), yaşam analizinde kullanılan bağımsız faktörlerin sayısı 5 veya daha az ise örneklem büyüklüğünü 60 olarak önermektedir. Lachin ve Foulkes (1986) yaşam analizi için örneklem büyüklüğü çalışmaları yapmıştır. Literatür incelendiğinde yaşam analizinde örneklem büyüklüğü ve gücü hesaplayan en iyi programlardan birisinin nQuery Advisor 4.0 olduğu görülmektedir (Elashoff, 2000).

Magee (1990) tarafından sunulan etki büyüklüğü ölçümünde uygulanan formül Allison (2010) tarafından yaşam analizine uygulanmıştır. Etki büyüklüğü G^2 ile hesaplanan bir olabilirlik oranı Ki-Kare istatistiğidir.

$$R^2 = 1 - \exp\left\{-\frac{G^2}{n}\right\} \quad (1.13)$$

Burada n örneklem genişliği, $G^2=[(\text{küçük olan model için}-2\log \text{olabilirlik})-(\text{büyük olan model için}-2\log \text{olabilirlik})]$ (Tabachnick ve Fidell, 2013).

R^2 bağımsız faktörler tarafından açıklanan yaşam analizindeki varyans oranı değil yaşam analizi ile ortak değişkenler arasındaki ilişkiyi belirtmektedir (Allison, 2010).

1.4. Cox Regresyon Modeli

Bu model literatürde genellikle yarı parametrik bir model olarak adlandırılmaktadır. Bunun sebebi ise dağılımın belirli bir şekli olmamasından ileri gelmektedir. Yaşam analizinde temel amaç hazard fonksiyonunu etkileyen prognostik faktörleri belirlemek ve hazard fonksiyonunu bulmaktır. Yani prognostik faktörlerin bağımlı değişken üzerindeki sağ kalım süresine etkilerini belirlemektir.

Cox Regresyon Modeli Varsayımları:

- Bağımsız değişkenlerin hazard fonksiyonu üzerine etkileri log-doğrusaldır.
- Bağımsız değişkenlerin log-doğrusal fonksiyonu ile hazard fonksiyonu arasında çarpımsal ilişki vardır.
- Her deneğe ait bağımsız değişken değerleri, araştırmanın başlangıcında belirlenmiş olup, çalışma süresince sabittir.

Yaşam analizi verisini modellemek için kullanılan en yaygın model, veriye kolaylıkla uygulanması nedeniyle Cox regresyon modelidir. Cox regresyon modeli için; X_1, X_2, \dots, X_p p tane bağımsız değişken,

x_1, x_2, \dots, x_p bu değişkenlerin değerleri iken Cox regresyonda bağımsız değişkenlerin kümesi $x=(x_1, x_2, \dots, x_p)$ vektörü ve $h_0(t)$ temel hazard fonksiyonu ise, i. birim için hazard fonksiyonu,

$$h_i(t) = h_0(t)\Psi(x_i) \quad (i = 1, 2, \dots, p) \quad (1.14)$$

şeklindedir.

Burada $\Psi(x_i)$ i. birim için açıklayıcı değişkenler vektörünün bir fonksiyonudur.

$$\Psi(x_i) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (1.15)$$

β_i 'ler regresyon katsayıları olarak tanımlanır. Bu durumda Cox regresyon modeli;

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (1.16)$$

şeklindedir. Orantılı hazard varsayımını içeren bu eşitliğin en önemli özelliği temel hazard fonksiyonunun t'nin bir fonksiyonu olmasıdır. Bu model;

$$\log \left\{ \frac{h_i(t)}{h_0(t)} \right\} = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad (1.17)$$

şeklinde tekrar düzenlendiğinde, Cox regresyon modeli, hazard oranının logaritması için doğrusal bileşen olarak da ifade edilebilir.

Cox regresyon modelinde açıklayıcı değişken olmadığında model;

$$h(t) = h_0(t) \exp\left(\sum_{j=1}^p \beta_j x_j\right) = h_0(t) \exp(0) = h_0(t) \quad (1.18)$$

şeklinde temel hazard fonksiyonuna indirgenir.

1.4.1. Hazardların Orantısallığı

Hazard oranının zamana karşı sabit olması orantılı hazard varsayımdır. $x^* = (x_1^*, x_2^*, \dots, x_p^*)$ ve $x = (x_1, x_2, \dots, x_p)$ iki bireye ait bağımsız değişkenler vektörünü ifade ederken hazard oranı;

$$\theta = \frac{\hat{h}(t, x^*)}{\hat{h}(t, x)} = \frac{\hat{h}_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j x_j^*\right)}{\hat{h}_0(t) \exp\left(\sum_{j=1}^p \hat{\beta}_j x_j\right)} = \frac{\exp\left(\sum_{j=1}^p \hat{\beta}_j x_j^*\right)}{\exp\left(\sum_{j=1}^p \hat{\beta}_j x_j\right)} = \exp\left(\sum_{j=1}^p \hat{\beta}_j (x_j^* - x_j)\right) \quad (1.19)$$

şeklinde yazılır. Bu ifade orantılı hazard varsayımını göstermektedir. Orantılı hazard varsayımı ise;

$$\hat{h}(t, x^*) = \hat{\theta} \hat{h}(t, x) \quad (1.20)$$

şeklinde de ifade edilebilir. Bu ifade de $\hat{\theta}$, orantılılık sabiti olarak ifade edilir ve zamandan bağımsızdır.

1.4.1.1. Cox Regresyon Modelinin Orantılı Hazard Varsayımının İncelenmesi

Bu modelin en önemli varsayımı orantılı hazard varsayımdır. Bu varsayımlar hazard oranlarının tüm zaman boyunca sabit olmasıdır. Bu varsayımın geçerliliği için orantılılığın incelenmesi gerekmektedir. Orantılı hazard varsayımını test etmek için

kullanılan yöntemler, grafiksel ve sayısal yöntemlerdir (Therneau ve Grambsch, 2000).

Orantılı hazard varsayımını test etmek için kullanılan grafiksel yöntemler;

- Ölçeklendirilmiş Schoenfeld artıklarının zamana karşı düzleştirilmiş çizimleri,
- Log-(log) yaşam eğrilerinin çizimi,
- Arsaj grafikleri
- Beklenen ve gözlenen yaşam eğrileri yöntemleri
- Anderson çizimi şeklindedir.

Bu yöntemlerden Schoenfeld artıklarının grafiğinde, Schoenfeld artıkları zamana karşı çizilir. Çizilen Schoenfeld artıkları grafiğinde artıklar yatay bir doğru etrafında sıfır merkezli tesadüfi olarak seyrediyorsa orantılı hazard varsayımının sağlandığı söylenebilir. Aksi takdirde varsayımın sağlanmadığı sonucuna varılır.

Orantılı hazard varsayımını test etmek için kullanılan sayısal yöntemler;

- Zamana bağlı açıklayıcı değişkenlerin kullanılması
- Schoenfeld artıkları ile yaşam sürelerinin rankının korelasyon testidir.

Bu sayısal yöntemlerden zamana bağlı açıklayıcı değişkenlerin kullanılması yönteminde model zamandan bağımsız değişkenleri ve zamanın bazı fonksiyonlarını kapsayan çarpım terimlerini içeren genişletilmiş Cox regresyon modelidir.

Schoenfeld artıkları ile yaşam sürelerinin rankının korelasyon testinde orantılı hazard varsayımının sağlanması için korelasyonun sıfıra yakın olması beklenir. Schoenfeld artıkları açıklayıcı değişkenin gerçek değeri ile ağırlıklı risk skorlarının ortalaması arasındaki farktır.

Cox regresyon modelinin orantılı hazard varsayımının incelenmesinde bazı araştırmacılar sayısal yöntemlerin kullanımını önerirken bazı araştırmacılar ise grafiksel yöntemlerin kullanımını önermektedir (Hosmer ve Lemeshow, 1999).

Sayısal yöntemler grafiksel yöntemlere göre daha nesnel grafiksel yöntemler ise daha öznedir.

1.4.2. Model Seçim Kriterleri

Akaike bilgi kriteri (AIC) ya da Bayesci bilgi kriteri (BIC) yaşam analizi modellemelerinde yapılacak analiz için en uygun olan yaşam analizi modelini belirlemek için kullanılmaktadır. AIC veya BIC değerleri hesaplandıktan sonra en küçük değere sahip model, analizde kullanılacak model olarak belirlenir.

Cox regresyon modeli için AIC değeri

$$AIC = -2\log L + \alpha p \quad (1.21)$$

Cox regresyon modeli için BIC değeri

$$BIC = -2\log L + p \log(n) \quad (1.22)$$

şeklindedir. Burada p modeldeki bilinmeyen parametrelerin sayısını ve n gözlem sayısını belirtmektedir.

1.4.3. Tabakalandırılmış Cox Regresyon Modeli

Bu modelde orantılı hazard varsayımını sağlamayan değişkenler tabakalandırılır. Tabakalandırılmış Cox regresyon modeli birçok tabaka üzerinden çok sayıda değişkenin tabakalandırılmasına imkân sağlar (Therneau ve Grambsch, 2000).

Tabakalı Cox modelinde öncelikle etkisi önemli olduğu düşünülen değişken uygun şekilde tabakalara ayrılır. Ancak her bir tabaka için oluşturulan modelde regresyon katsayıları aynı kalır. Temel hazard fonksiyonları her bir tabaka için farklı olduğundan dolayı tabakalı Cox regresyon modelinde her bir tabaka için farklı tahmin edilmiş yaşam eğrileri oluşacaktır. Tabakalı Cox regresyon modelinde değişkenlerin katsayıları her bir tabaka için aynı olduğundan, hazard oranlarının tahmini de her bir tabaka için aynıdır (Kleinbaum ve Klein, 1996).

Tabakalı Cox modeli için olabilirlik fonksiyonu, Cox regresyon modelinden farklıdır. Tabakalı Cox model için olabilirlik fonksiyonu, her bir tabaka için olabilirlik fonksiyonlarının çarpılmasıyla elde edilir. L_1 ve L_2 orantılı hazard varsayımını sağlamayan değişkenin iki tabakasına ait olabilirlik fonksiyonlarını gösterdiğinde, tabakalı Cox modeli için olabilirlik fonksiyonu $L = L_1 \times L_2$ şeklinde yazılabilir (Karasoy ve Tutkun 2016).

1.4.3.1. Etkileşimsiz Tabakalandırılmış Cox Regresyon Modeli

Bu model aşağıdaki şekilde yazılabilir.

$$h_g(t, x) = h_{0g}(t) \exp [\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p] \quad g = 1, 2, \dots, k^* \quad (1.23)$$

Burada g alt indisi g . tabakayı ifade ederken, tabaka, tabakalandırılan değişken z^* 'in farklı düzeylerini gösterir. Tabaka sayısı ise k^* , a eşittir. Tabakalandırılmış z^* değişkeni modelde yer almazken oransal hazard varsayımını sağladığı kabul edilen değişkenler (x 'ler) yer alır. Temel hazard fonksiyonu her bir tabaka için farklı ve $\beta_1, \beta_2, \dots, \beta_p$ katsayıları her bir tabaka için aynıdır. Burada, x 'lerin katsayıları her bir tabaka için aynı olduğundan hazard oranı tahminleri de her tabaka için aynıdır. Kısmi olabilirlik fonksiyonu ençoklanarak $\beta_1, \beta_2, \dots, \beta_p$ regresyon katsayıları tahmin edilir ve olabilirlik fonksiyonu $L = L_1 \times L_2 \times \dots \times L_{k^*}$ şeklinde yazılabilir.

1.4.3.2. Etkileşimli Tabakalandırılmış Cox Regresyon Modeli

Bu modelde orantısızlığa sebep olan değişken sayısının bir ya da iki ve daha fazla sayıda olmasına bağlı olarak etkileşimli model iki farklı şekilde ifade edilmektedir.

- i) Burada öncelikle orantısızlığa sebep olan değişken sayısı bir taneyse etkileşimli tabakalandırılmış Cox regresyon modeli iki farklı şekilde ifade edilmektedir.
 - a. Veri kümesi k^* farklı tabakaya ayrılır ve model;

$$h_g(t, x) = h_{0g}(t) \exp [\beta_{1g}x_1 + \beta_{2g}x_2 + \dots + \beta_{pg}x_p] \quad g = 1, 2, \dots, k^* \quad (1.24)$$

şeklinde yazılır.

- b. Tabakalandırılmış modelde bağımsız değişkenler ile orantısızlığa sebep olan değişken ile bu değişkenlerin çarpımları etkileşimli modele eklenir ve model eşitlik (1.25)'de ifade edildiği gibi olur.

$$h_g(t, x) = h_{0g}(t) \exp [\beta_1^*x_1 + \beta_2^*x_2 + \dots + \beta_p^*x_p + \beta_{p+1}^*(x_1 * z) + \dots + \beta_{2p}^*(x_p * z)] \quad (1.25)$$

Orantılı hazarda sebep olan değişken sayısı iki ya da daha fazla sayıda ise yeni bir değişken $z^* = 0, 1, 2, \dots, k^* - 1$ tanımlanır. Burada k^* orantısızlığa sebep olan değişkenlerin düzey sayılarının çarpımıdır. Orantılı hazard varsayımını sağlayan değişken sayısı p ise etkileşimli modeli aşağıdaki yazabiliriz.

$$h_g(t, x) = h_{0g}(t) \exp [\beta_{1g}x_1 + \beta_{2g}x_2 + \dots + \beta_{pg}x_p] \quad g = 1, 2, \dots, k^* \quad (1.26)$$

1.4.3.3. Tabakalandırılmış Cox Regresyon Modelinde Etkileşimsizlik Varsayımı

Tabakalandırılmış Cox regresyon modelinde regresyon katsayıları her bir tabaka için değişmez. Modelde bu özellik “etkileşimsizlik” varsayımı olarak tanımlanmaktadır. Bu durumda etkileşimler modele eklenildiğinde; tabakalar için farklı katsayılar oluşur. Bu sebeple tabakalı değişkenin her bir tabakası ayrı ayrı modele eklenir. Olabilirlik oran istatistiği bu modelde etkileşimsizlik varsayımını test etmek için kullanılmaktadır. Etkileşimsizlik varsayımı için olabilirlik oranı istatistiği;

$$LR = -2[\log \hat{L}_{indirgenmiş\ model} - \log \hat{L}_{Tam\ model}] \quad (1.27)$$

şeklinde verilmektedir. Burada $\hat{L}_{indirgenmiş\ model}$; etkileşimsiz modeli ifade ederken $\hat{L}_{Tam\ model}$ ise etkileşimli tabakalandırılmış Cox modeli belirtmektedir.

O halde modeller birbirlerinden çarpım terimlerini içermesinden dolayı farklılaşmaktadır. Araştırma hipotezi aşağıdaki şekilde kurulmaktadır.

bir β katsayısı vardır. Bu durumda t anında, hazard üzerinde etkisi olan ve değeri t zamanında ölçülen $x_j(t)$ değişkeninin tek bir değeri vardır.

Genişletilmiş Cox regresyon modelinde regresyon katsayıları Cox regresyon modelinde olduğu gibi en çok olabilirlik yöntemi kullanılarak tahmin edilir. En çok olabilirlik tahminleri kısmi olabilirlik fonksiyonunu ençoklayarak bulunur. Genişletilmiş Cox regresyon modelinde yapılan hesaplamalar Cox regresyon modeline kıyasla daha karmaşıktır.

1.5. Yaşam Analizinde Kullanılan Bazı Dağılımlar

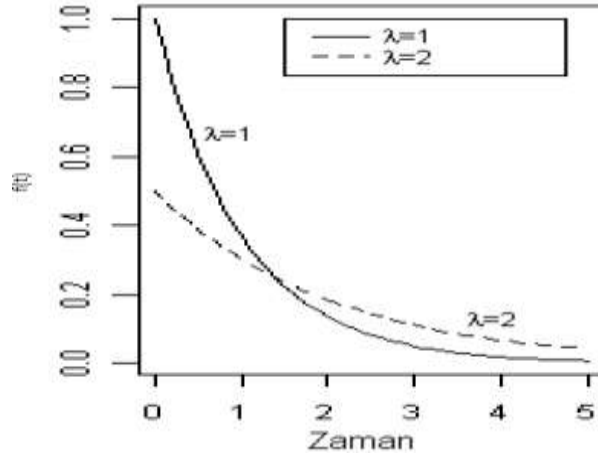
1.5.1. Üstel Dağılım

Tamamen rastgele başarısızlık modeli olarak da adlandırılan üstel dağılım yaşam analizinde yaygın olarak kullanılan önemli dağılımlardan birisidir. Araştırmacılar üstel dağılımı 1940'lı yıllarda elektronik sektöründe başarısızlık analiz için kullanmaya başlamışlardır. Epstein ve Sobel (1953) araştırmalarında, sansürlü veriler olduğunda, normal dağılım yerine neden Üstel dağılımın seçilmesi gerektiğini göstermişlerdir.

Üstel dağılım genellikle tamamen tesadüfi başarısızlık denemelerinin modellenmesinde kullanılmaktadır. Üstel dağılım hazard oranının yüzdesinin belli bir grup hasta için zaman içerisinde aynı kaldığını varsayar. Yani hazard oranı zamana bağlı değildir.

Üstel dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (1.32)$$



Şekil 1.1. Üstel Dağılımın Olasılık Yoğunluk Eğrileri

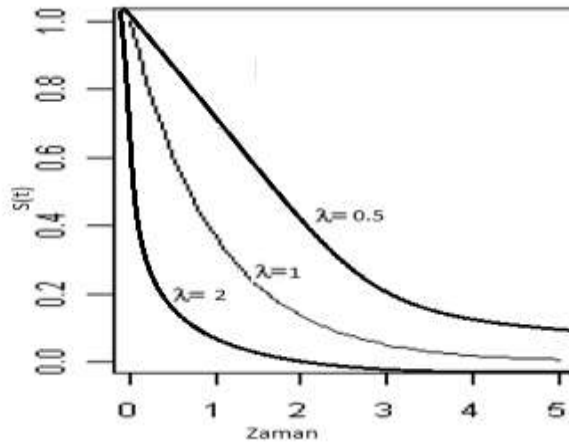
Üstel dağılımın olasılık yoğunluk eğrileri şekil 1.1'de olduğu gibidir.

Üstel dağılımın birikimli dağılım fonksiyonu;

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0, \quad \lambda > 0 \quad (1.33)$$

Üstel dağılımın yaşam fonksiyonu;

$$S(t) = e^{-\lambda t}, \quad t \geq 0 \quad (1.34)$$



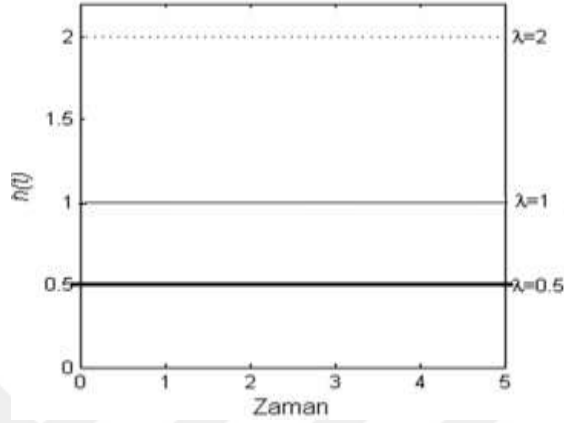
Şekil 1.2. Üstel Dağılımın Olasılık Yaşam Eğrileri

Şekil 1.2'deki Üstel dağılıma ait yaşam eğrileri incelendiğinde; λ parametre değeri arttıkça yaşam olasılığının hızla azaldığı görülmektedir.

Üstel dağılımın hazard fonksiyonu ise aşağıdaki gibi bulunabilir.

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda, \quad t \geq 0, \quad \lambda > 0 \quad (1.35)$$

Yukarıdaki ifadeden de görüldüğü gibi üstel dağılımın hazard fonksiyonu sabittir. Üstel dağılımın hazard fonksiyon eğrileri şekil 1.3’de olduğu gibidir.



Şekil 1.3. Üstel Dağılımın Hazard Fonksiyon Eğrileri

1.5.2. Weibull Dağılımı

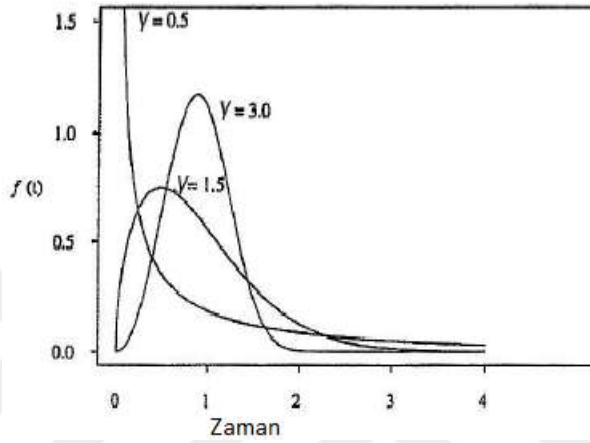
Waloddi Weibull; Weibull dağılımını 1937 yılında bulmuş ve 1951 yılında bu dağılımı güvenilirlik analizinde kullanmıştır. Waloddi Weibull parametre tahmininde ortalama sıra numaralarını kullanırken, Leonard Johnson (1960) daha güvenilir olan medyan sıra numaralarını önermiştir. 1966 yılında Pike, 1973 yılında ise Peto ve Lee Weibull dağılımının parametrelerinin olabilirlik tahminlerini incelemişlerdir. Scott ve Hahn (1980)’de radyasyon çalışmalarında Weibull dağılımını kullanmışlardır. Gumbel iki üç başarısızlıkta bile Weibull dağılımının kullanılabileceğini ispatlamıştır (Abernety, 2000).

Weibull dağılımı mühendislikte belirli bir ömrü olan materyal veya biyomateryallerin üretim ve kullanım süreci içerisinde ortaya çıkabilecek arıza veya bozulma zamanının belirlenmesinde sıklıkla kullanılmaktadır. Weibull dağılımı artan, azalan veya sabit şekildedir. Weibull dağılımı üstel dağılımın genelleştirilmiş halidir. Tek farkı sabit hazard oranına sahip olmamasıdır. Bu fark üstel dağılıma göre daha geniş uygulama alanı sağlamaktadır.

λ ve γ olmak üzere iki parametreye sahip olan Weibull dağılımının parametreleri λ (ölçek) ve γ (biçim) olarak tanımlanır. Weibull dağılımının biçim parametresi $\gamma=1$ alındığında üstel dağılım, $\gamma=2$ alındığında ise Rayleigh dağılımı elde edilir.

Weibull dağılımının olasılık yoğunluk fonksiyonu;

$$f(t) = \lambda \gamma (\lambda t)^{\gamma-1} e^{-(\lambda t)^\gamma}, \quad t \geq 0, \quad \gamma, \lambda > 0, \quad (1.36)$$



Şekil 1.4. $\lambda=1$ ve $\gamma=3, 1.5$ ve 0.5 İçin Olasılık Yoğunluk Eğrileri

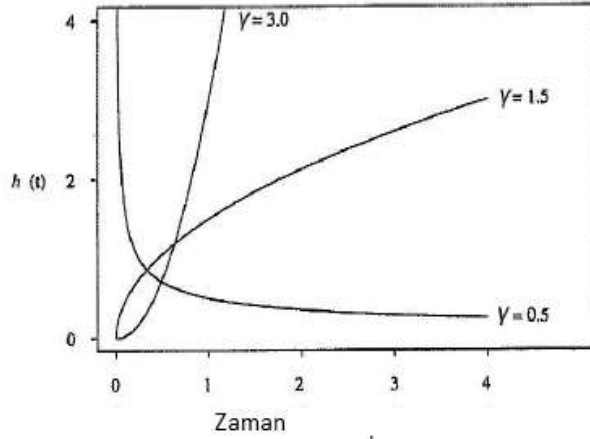
Weibull dağılımının olasılık yoğunluk fonksiyonu eğrileri $\lambda=1$ ve $\gamma=3, 1.5$ ve 0.5 için şekil 1.4'de olduğu gibidir.

Yaşam fonksiyonu;

$$S(t) = \exp[-(\lambda t)^\gamma] \quad (1.37)$$

Hazard fonksiyonu;

$$h(t) = \lambda \gamma (\lambda t)^{\gamma-1} \quad (1.38)$$



Şekil 1.5. $\lambda=1$ ve $\gamma=3, 1.5$ ve 0.5 için Hazard Fonksiyon Eğrileri

$\gamma = 1$ iken hazard fonksiyonu zaman arttıkça sabit kalır. Weibull dağılımının hazard fonksiyonu, $\gamma > 1$ iken artar, $\gamma < 1$ iken azalır (Abernety, 2000). $\lambda=1$ ve $\gamma=3, 1,5$ ve $0,5$ için hazard fonksiyon eğrileri şekil 1.5'de olduğu gibidir.

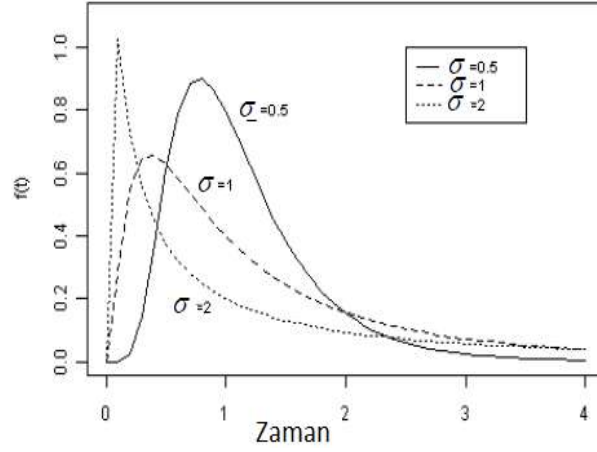
1.5.3. Log-normal Dağılım

Log-normal dağılımı ilk olarak McAlister (1879) tarafından tanımlanmıştır. Daha sonra Gaddum (1945) tarafından fen alanında, sağlık alanında ise ilk olarak kanser çalışmalarında Boag (1949)'tarafından kullanılmıştır. Aitchison ve Brown (1957) yılında ise bu dağılımı ekonomi çalışmalarında kullanmışlardır. Feinleib (1960) yılında lösemi hastalarının yaşam sürelerinin log-normal dağılıma benzediğini, Horner (1987) yılında ise Alzheimer hastalığının yaşam süresinin dağılımının log-normal dağılıma yakın olduğunu ispatlamışlardır (Lee ve Wang, 2003).

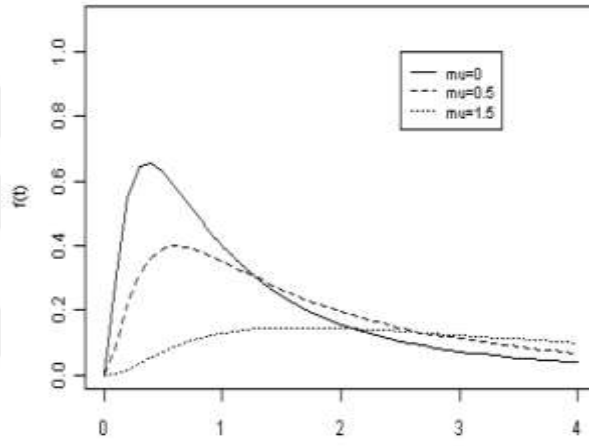
Dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibidir;

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

$$= \frac{1}{t\sigma} \phi_{\text{nor}} \left(\frac{\log t - \mu}{\sigma} \right); \quad t \geq 0, \quad \sigma > 0 \quad (1.39)$$



Şekil 1.6. Log-normal Dağılımın $\mu = 0, \sigma = 0.5, 1$ ve 2 Değerleri için Olasılık Yoğunluk Eğrileri



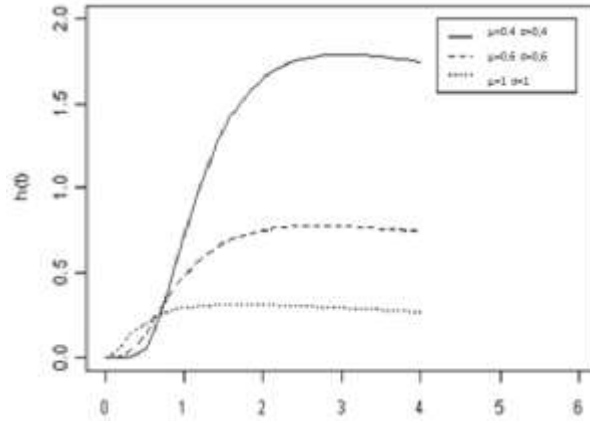
Şekil 1.7. Log-normal Dağılımın $\sigma = 1, \mu = 0, 0.5$ ve 1.5 Değerleri için Olasılık Yoğunluk Eğrileri

Log-normal dağılımın farklı σ ve μ değerleri için olasılık yoğunluk eğrileri şekil 1.6 ve şekil 1.7’de olduğu gibidir. Log-normal dağılımın sağkalım fonksiyonu eşitlik 1.40’da olduğu gibidir.

$$S(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_t^{\infty} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right] dx = 1 - \Phi_{\text{nor}}\left(\frac{\log t - \mu}{\sigma}\right) \quad (1.40)$$

Log-normal dağılımın hazard fonksiyonunda eşitlik (1.41)’de verildiği gibidir.

$$h(t) = \frac{\frac{1}{t\sigma} \Phi_{\text{nor}}\left(\frac{\log t - \mu}{\sigma}\right)}{1 - \Phi_{\text{nor}}\left(\frac{\log t - \mu}{\sigma}\right)} \quad (1.41)$$



Şekil 1.8. Log-normal Dağılımın Hazard Fonksiyon Eğrileri

Log-normal dağılımın sağkalım ve hazard fonksiyonları standart normal dağılım içerir. Log-normal dağılımın hazard fonksiyonu belli bir miktar artar ve daha sonra azalan bir eğri gösterir. Log-normal dağılımın hazard fonksiyonu eğrileri şekil 1.8’de olduğu gibidir.

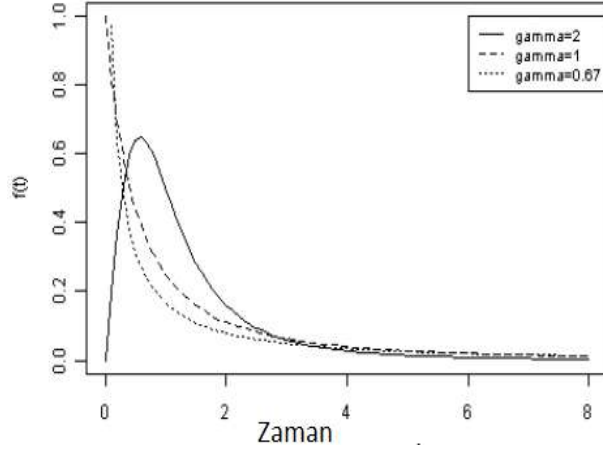
Log-normal dağılımının hazard fonksiyonu önce en yüksek değere kadar artar, daha sonra zaman ilerledikçe azalır. Hazard fonksiyonu azalan bir eğri göstermektedir. Ters U şeklinde bir hazard fonksiyonuna sahip olan log-normal dağılım genellikle literatürde tekrar eden olaylarda kullanılmaktadır.

1.5.4. Log-Logistik Dağılım

Log-logistik dağılım Weibull dağılımının alternatifidir. Log-logistik dağılım ters U şeklindedir ve bu dağılımda ölçüm parametresi 1’den küçüktür. Ölçüm parametresi 1’den büyük ve eşit olduğunda bu dağılım Weibull dağılımına benzer. Zaman içindeki değişim sabittir. Log-logistik dağılım ölçüm parametresine bağlı olarak Weibull ya da log-normal dağılıma benzer.

Log-logistik dağılımın olasılık yoğunluk fonksiyonu:

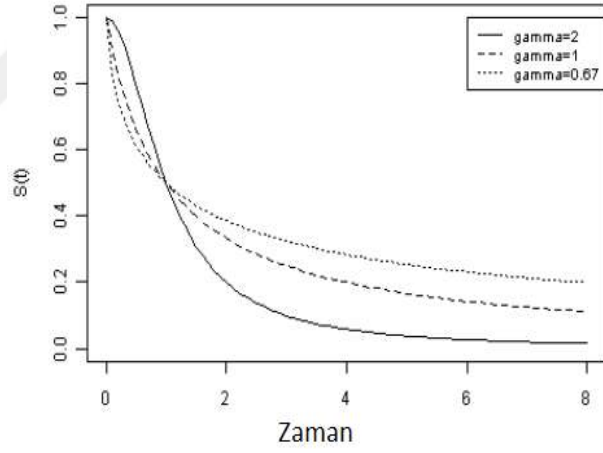
$$f(t) = \frac{\alpha \gamma t^{\gamma-1}}{[1 + \alpha t^\gamma]^2} \quad (1.42)$$



Şekil 1.9. Log-logistik Dağılımın Olasılık Yoğunluk Eğrileri

Log-logistik dağılımın olasılık yoğunluk eğrileri şekil 1.9’da olduğu gibidir. Log-logistik dağılımın yaşam fonksiyonu;

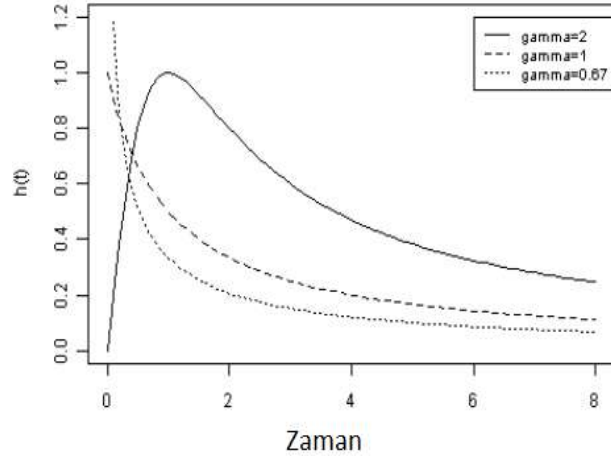
$$s(t) = \frac{1}{1 + \alpha t^\gamma} \quad (1.43)$$



Şekil 1.10. Log-logistik Dağılımın Olasılık Yaşam Eğrileri

Log-logistik dağılımın olasılık yaşam eğrileri şekil 1.10’da olduğu gibidir. Log-logistik dağılımın hazard fonksiyonu

$$h(t) = \frac{\alpha \gamma t^{\gamma-1}}{1 + \alpha t^\gamma} \quad (1.44)$$



Şekil 1.11. Log-logistik Dağılımın Farklı Gamma Değerleri için Hazard Fonksiyonu

Log-logistik dağılımın farklı γ değerleri için hazard fonksiyonu şekil 1.11’de olduğu gibidir. $\gamma > 1$ iken log-logistik dağılımın hazard fonksiyonu başlangıçta sıfır değerini alır.

$t = (\gamma - 1)^{1/\gamma} / \alpha^{1/\gamma}$ noktasında tepe değerini alacak biçimde artar sonra azalır.

$\gamma = 1$ iken log-logistik dağılımın hazard fonksiyonu $\alpha^{1/\gamma}$ noktasında başlar daha sonra monoton olarak azalır.

$\gamma < 1$ iken log-logistik dağılımın hazard fonksiyonu sonsuzda bir zaman noktasında başlar ve daha sonra Weibull dağılımına benzer biçimde azalır (Hanagal, 2011)

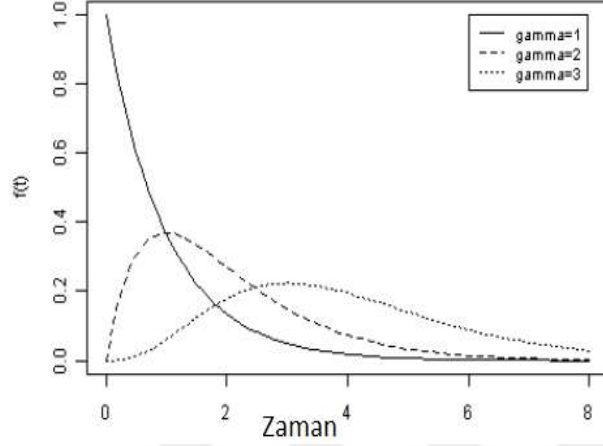
1.5.5. Gamma Dağılımı

Gamma dağılımı güvenilirlik ve yaşam analizlerinde de kullanılan bir dağılımdır. Brown ve Flood (1947)’de ve daha sonra Birnbaum ve Saunders (1958)’de endüstrideki üretim maddelerinin yaşam süresini belirlemede gamma dağılımını kullanmışlardır (Lee, 1992).

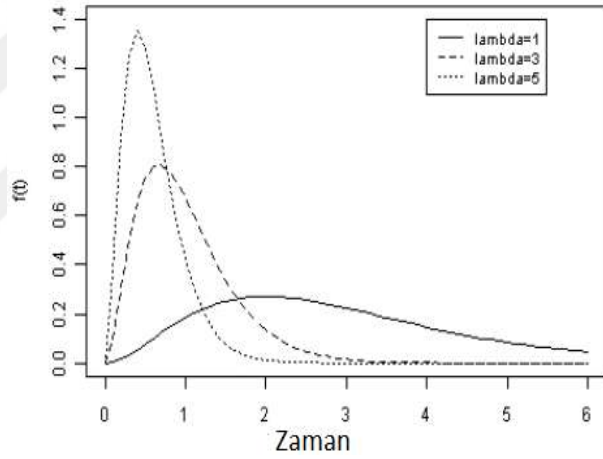
Gamma dağılımı biçim (γ) ve ölçek (λ) olmak üzere iki parametreye sahiptir.

Gamma dağılımın olasılık yoğunluk fonksiyonu:

$$f(t) = \frac{\lambda}{\Gamma(\gamma)} (\lambda t)^{\gamma-1} e^{-\lambda t} \quad t > 0, \gamma > 0, \lambda > 0 \quad (1.45)$$



Şekil 1.12. $\lambda=1$ ve $\gamma=1,2,3$ olduğunda Gamma Dağılımının Olasılık Yoğunluk Eğrileri



Şekil 1.13. $\gamma=3$ ve $\lambda=1,3,5$ olduğunda Gamma Dağılımının Olasılık Yoğunluk Eğrileri

Gamma dağılımının farklı γ ve λ değerleri için olasılık yoğunluk eğrileri şekil 1.12 ve şekil 1.13'de olduğu gibidir. Gamma dağılımının yaşam fonksiyonu:

$$S(t) = \int_t^{\infty} \frac{\lambda}{\Gamma(\gamma)} (\lambda x)^{\gamma-1} e^{-\lambda x} dx \quad (1.46)$$

Üstel, Weibull ve log-normal gibi dağılımların hepsi gamma dağılımının özel durumlarıdır. Bu ilişkiden dolayı gamma ve diğer dağılımlar arasındaki farklılıklar Ki-Kare olabilirlik oran testi ile değerlendirilebilir (Allison, 2010).

Genelleştirilmiş gamma dağılımının olasılık yoğunluk fonksiyonu

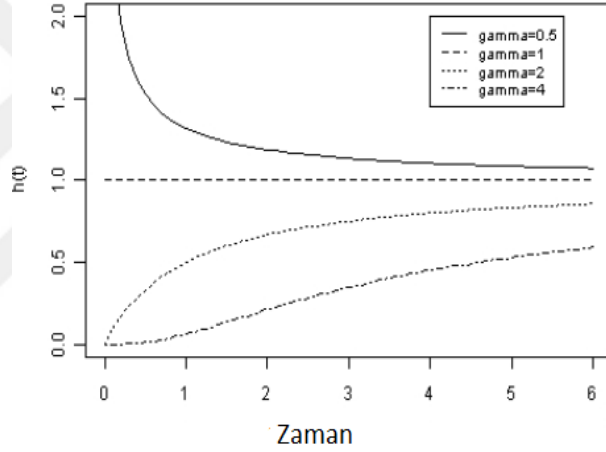
$$f(t) = \frac{\alpha \lambda^{\alpha \gamma}}{\Gamma(\gamma)} (t)^{\alpha \gamma - 1} \exp[-(\lambda t)^{\alpha}], \quad t > 0, \quad \gamma > 0, \quad \lambda > 0, \quad \alpha > 0 \quad (1.47)$$

Genelleştirilmiş gamma dağılımında, $\alpha=\gamma=1$ ise üstel dağılıma

Genelleştirilmiş gamma dağılımında, $\gamma=1$ ise Weibull dağılıma

Genelleştirilmiş gamma dağılımında, $\alpha=1$ ise gamma dağılıma

Genelleştirilmiş gamma dağılımında, $\gamma \rightarrow \infty$ ise log-normal dağılıma dönüşmektedir.



Şekil 1.14. $\lambda=1$ ve $\gamma=0.5, 1, 2$ ve 4 olduğu Durumda Hazard Fonksiyon Eğrileri

Şekil 1.14'de $\lambda=1$ iken $\gamma=0,5;1;2$ ve 4 olduğu durumlarda hazard fonksiyon eğrileri görülmektedir.

1.5.6. Gompertz Dağılımı

Pek çok alanda kullanılan Gompertz dağılımı İngiliz aktüer Benjamin Gompertz (1825) tarafından tanımlanmıştır.

Gompertz dağılımının olasılık yoğunluk fonksiyonu

$$f(t) = \lambda e^{\gamma t} e^{-\frac{\lambda}{\gamma}(e^{\gamma t}-1)}, \quad \lambda > 0 \quad (1.48)$$

Gompertz dağılımının yaşam fonksiyonu

$$S(t) = e^{-\frac{\lambda}{\gamma}(e^{\gamma t} - 1)} \quad (1.49)$$

Gompertz dağılımının hazard fonksiyonu

$$h(t) = \lambda e^{\gamma t} \quad (1.50)$$

Gompertz dağılımının birikimli hazard fonksiyonu

$$H(t) = \frac{\lambda}{\gamma} (e^{\lambda t} - 1) \quad (1.51)$$

$\lambda = 0$ durumunda ise Gompertz dağılımı Üstel dağılıma dönüşür. Gompertz dağılımına bir c sabiti eklendiğinde Gompertz-Makeham dağılımı elde edilir

Gompertz-Makeham dağılımının hazard fonksiyonu

$$h(t) = \lambda e^{\gamma t} + c \quad (1.52)$$

şeklindedir. Gompertz-Makeham dağılımı, paylaşılmamış zayıflık modellerinin temelini başlangıç noktalarından biridir (Wienke, 2010).

1.6. Parametre Tahmin Yöntemleri

Parametrelerin tahmin edilmesinde en çok olabilirlik yöntemi kullanılır. Olabilirlik fonksiyonunun parametrelere göre birinci türevi alınıp sifıra eşitlenmesi ile bulunan parametre tahminleri, en çok olabilirlik tahmini olarak adlandırılmaktadır. Tam olmayan veri gruplarında kayıp veri içeren durumlarda, sansürlü ya da gruplanmış verinin tahmin yöntemleri için olabilirlik fonksiyonu kullanılmaktadır. Olabilirlik fonksiyonu durdurulmuş yaşam sürelerinden etkilenmez (Gharibvand ve Liu, 2009).

En çok olabilirlik yönteminde en sık kullanılan algoritmalar Newton-Raphson, beklenti en büyükleme (EM), Monte Carlo beklenti en büyükleme ve cezalandırılmış olabilirlik yöntemleridir (Klein ve ark 1992; Nielsen ve ark 1992).

Hastie ve Tishirani 1993’de yapmış oldukları çalışmada cezalandırılmış olabilirlik fonksiyonunu önermişlerdir. Therneau ve ark. 2003 yılında gamma zayıflık model ile yapmış oldukları çalışmada cezalandırılmış olabilirlik fonksiyonunu önermişlerdir.

(McGilchrist ve Aisbett 1991)’yılında ilk olarak zayıflık modelinde parametre tahminleri için cezalandırılmış olabilirlik fonksiyonunu önermiş ve kullanmışlardır.

Beklenti en büyükleme (EM) yöntemi ile cezalandırılmış olabilirlik yöntemi karşılaştırıldığında cezalandırılmış olabilirlik yöntemi zayıflık modeli için daha küçük standart hata tahmini yapar. Beklenti en büyükleme (EM) yöntemi ise temel hazzardı bilinmeyen dağılımlar için daha doğru tahminler yapar (Gharibvand ve Liu, 2009).

1.6.1. Newton-Raphson Yöntemi

Newton-Raphson yöntemi ile parametrelerin en çok olabilirlik tahminleri ve katsayıları elde edilmektedir. Cox regresyon modelinde, β parametrelerinin en çok olabilirlik tahminleri bu teknik kullanılarak olabilirlik fonksiyonunun en çoklanması ile hesaplanmaktadır (Collett, 2003).

Eşitlik (1.53)’de Newton-Raphson yönteminde (s+1)’inci turda β parametreleri vektörünün tahmini ifade edilmiştir.

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}(\hat{\beta}_s)u\hat{\beta}_s \quad (1.53)$$

Bu eşitlikte $I^{-1}(\hat{\beta}_s)$ bilgi matrisinin tersi olarak ele alınmaktadır (Karasoy ve Tutkun, 2016).

1.6.2. EM (Beklenti En Büyükleme) Algoritması

Beklenti en büyükleme (EM) algoritması tam olmayan veri analizlerinde en çok olabilirlik tahminini en iyi şekilde elde eder. Beklenti en büyükleme algoritması, tam olmayan veri gruplarında kayıp veri içeren durumlarda, sansürlü ya da gruplanmış gözlemlerde kullanılmaktadır.

Beklenti en büyükleme algoritması, iki adımdan oluşmaktadır. Bunlar, E ve M adımlarıdır. E adımında gözlenemeyen ya da kayıp verinin tahmini yapılır. M adımında ise gözlenemeyen veriye ilişkin elde edilen tahmin değerleri olasılık fonksiyonunda yerine konularak bu fonksiyonu maksimum yapan parametre değerleri kestirilir. M-adımı sonucunda ortaya çıkan tahminler, beklenti En büyükleme algoritmasının çıktısını oluşturmaktadır.

E adımında; parametrelere başlangıç değeri verilerek başlanır. Kayıp veri üzerinden log-olasılığın beklenen fonksiyonu alınarak tam-veri örnekleme hesaplanır. M adımında; E-adımında elde edilen olasılık fonksiyonu maksimum yapılarak tam-veri üzerinden örneklemin en çok olasılık tahmini hesaplanır. Bulunan parametrenin beklenen değeri ile bir önceki parametrenin beklenen değeri karşılaştırılır, aralarındaki fark azalırsa, algoritma durdurulur; eğer fark mevcutsa, hesaplanan parametre, bir önceki parametre ile yer değiştirip yeniden hesaplanır.

1.6.3. Monte Carlo Beklenti En Büyükleme (MCEM)

MCEM beklenti en büyükleme (EM) algoritmasının stokastik bir versiyonudur. Monte Carlo beklenti en büyükleme algoritması, Beklenti en büyükleme algoritmasının kontrol edilmesi zor olan E-adımına seçenek oluşturmak amacıyla geliştirilmiştir (Jank, 2005).

1.7. Cox Regresyon Modelinde Parametre Tahmini

Cox regresyon modelinde x tane başarısızlık $t_{(1)} < t_{(2)} < \dots < t_{(x)}$ olarak sıralanmış olsun. $t_{(i)}$ zamanında risk altındaki birey sayısını $Y_{t_{(i)}}$ ile gösterirsek burada $Y_{t_{(i)}}$ tüm bireylerin sağkalım zamanlarını göstermektedir. A_i , $t_{(i)}$ zamanındaki başarısızlığa uğramış birey sayısını ve B_i $t_{(i)}$ zamanındaki başarısız ve sansürlü birey sayısını ifade ederken olasılık fonksiyonu;

$$L = \prod_{i=1}^x P(A_i/B_i) \prod_{i=1}^{x+1} P(A_i/B_{i-1}, A_{i-1}) \quad (1.54)$$

L fonksiyonundaki $\prod_{i=1}^x P(A_i/B_i)$ ifadesi kısmi olabilirlik fonksiyonu olarak adlandırılır.

Hazard fonksiyonu ise;

$$P(A_i/B_i) = \frac{h(t_i, x_i)}{\sum_{j \in Y(t_i)} h(t_i, x_j)} \quad (1.55)$$

$$L(\beta) = \prod_{i=1}^x P(A_i/B_i) = \prod_{i=1}^x \frac{\exp(\beta' x_i)}{\sum_{j \in Y(t_i)} \exp(\beta' x_j)} \quad (1.56)$$

Bu eşitlikte en çok olabilirlik tahminleri Newton-Raphson metodu ile bulunulabilir. Bunun için eşitliğin logaritması alınarak

$$\ln L(\beta) = \sum_{i=1}^x \beta' x_i - \sum_{i=1}^x \ln \left[\sum_{j \in Y(t_i)} \exp(\beta' x_j) \right] \quad (1.57)$$

β 'ya göre türev alınarak olabilirlik fonksiyonunun birinci türevi ile skor fonksiyonu elde edilir.

$$U(\beta) = \frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^x x_i - \sum_{i=1}^x \frac{\sum_{j \in Y(t_i)} \exp(\beta' x_j) x_j}{\sum_{j \in Y(t_i)} \exp(\beta' x_j)} \quad (1.58)$$

$U(\beta) = 0$ alınarak, en çok olabilirlik tahminleri bulunur.

Olabilirlik fonksiyonunun ikinci türevi bilgi matrisini verir.

$$\begin{aligned} U(\beta) &= \frac{\partial^2 \ln L(\beta)}{\partial \beta^2} \\ &= \sum_{i=1}^x \frac{\sum_{j \in Y(t_i)} \exp(\beta' x_j) x_j x_j'}{\sum_{j \in Y(t_i)} \exp(\beta' x_j)} - \sum_{i=1}^x \frac{[\sum_{j \in Y(t_i)} \exp(\beta' x_j) x_j][\sum_{j \in Y(t_i)} \exp(\beta' x_j) x_j']}{[\sum_{j \in Y(t_i)} \exp(\beta' x_j)]^2} \end{aligned} \quad (1.59)$$

Aynı başarısızlık zamanında birden fazla başarısızlık meydana geldiğinde yani eş zamanlı gözlemler olduğunda tercih edilen olabilirlik fonksiyonları Breslow

(1970), Efron (1977), Kalbfleisch ve Prentice (1981) tarafından önerilmiştir (Collett, 2003).

1.7.1. Breslow'un Kısmi Olabilirlik Fonksiyonu

Breslow (1970) tarafından önerilen olabilirlik fonksiyonu,

$$L_B(\beta) = \prod_{i=1}^k \frac{\exp(\beta' x_i)}{[\sum_{j \in Y(t_i)} \exp(\beta' x_j)]^{d_i}} \quad (1.60)$$

Eğer t_i zamanında d_i tane başarısızlık varsa, x_i 'nin h . elemanı $x_{hi} = \sum_{j=1}^{d_i} x_{hik}$ dir. x_{hik} , i . ($i=1, 2, \dots, r$) ölüm zamanında, ölen d_i bireyden k tanesi için h . ($h=1, 2, \dots, p$) bağımsız değişkenin değeridir. Eşitlik (1.60)'da x_i : i .zamanda başarısızlığa uğrayan bireyler için her bir bağımsız değişkenin toplamlarından oluşan vektörü ifade etmektedir

1.7.2. Efron'un Kısmi Olabilirlik Fonksiyonu

Efron (1977)'de klasik Cox regresyon modeli için uygun olabilirliği, $D(t_i)$, t_i zamanında ölen bireylerin kümesi olmak üzere aşağıdaki gibi tanımlanır.

$$L_E(\beta) = \prod_{i=1}^k \frac{\exp(\beta' x_i)}{\prod_{j=1}^{d_i} [\sum_{r \in Y(t_i)} \exp(\beta' x_r) - \frac{j-1}{d_i} \sum_{r \in D(t_i)} \exp(\beta' x_r)]} \quad (1.61)$$

d_i : t_i zamanındaki ölen birey sayısını ifade etmektedir.

$Y(t_i)$: t_i zamanından önce risk altındaki bireylerin sayısı göstermektedir.

Breslow (1970)'de ve Efron (1977)'de olabilirlik fonksiyonu için önerdiği yaklaşımlar kullanıldığında uygulamada benzer sonuçlar vermektedir (Terzi, 2003).

β 'ların en çok olabilirlik tahminleri $\hat{\beta}$ ile gösterilirse, β 'ların testinde kullanılan üç yöntem vardır:

- Wald Testi

Wald testi, modeldeki her β katsayısının istatistiksel anlamlılığını test etmede kullanılır. Wald testi, en çok olabilirlik tahmin edicilerinin normal dağıldığı varsayımına dayanır.

$$X_w^2 = \hat{\beta}^T [\hat{V}(\hat{\beta})] \hat{\beta}^{-1} \quad (1.62)$$

- Olabilirlik Oran Testi

Olabilirlik Oran Testi Wald istatistiğinden daha genel bir yapıya sahiptir. Kategorik değişkenlerin iki veya daha fazla düzeyinin olduğu ve Cox modeline aynı anda çok sayıda değişkenin dâhil edildiği durumlarda tercih edilmektedir. L_0 basit model için en çok olabilirlik fonksiyonunu, L_1 bütün model için en çok olabilirlik fonksiyonunu gösterdiğinde, olabilirlik oran testi aşağıdaki gibi ifade edilir.

$$X_{LR}^2 = -2[\ln L_0 - \ln L_1] \quad (1.63)$$

β parametreleri için,

$$X_{LR}^2 = -2[\ln L(0) - \ln L(\beta)] \quad (1.64)$$

Yukarıdaki iki test istatistiği de H_0 ın doğruluğu altında p serbestlik dereceli Ki-Kare X_p^2 dağılımına sahiptir. Bir β parametresi için $1-\alpha$ anlamlılık düzeyinde güven aralığı yine aynı varsayım altında,

$$\hat{\beta} \pm z_{\alpha/2} sh(\hat{\beta}) \quad (1.65)$$

olarak gösterilebilir (Hosmer ve Lemeshow, 1999).

Olabilirlik oran testi ve Wald test istatistikleri genellikle aynı sonucu vermezler. Olabilirlik oran testi ve Wald test istatistikleri sonuçları karşılaştırıldığında olabilirlik oran test istatistiğinin daha iyi istatistiksel sonuçlara sahip olduğu görülür. Bu sebeple şüpheye düştüğünde olabilirlik oran test istatistiği tercih edilmelidir (Kleinbaum ve Klein, 1996).

- Skor Testi

Skor testi, birden fazla deęişken ve sürekli deęişkenler için logrank testinin genelleştirilmiş halidir.

$$X_{SC}^2 = U(\beta_0)'I^{-1}(\beta_0)U(\beta_0) \sim X_p^2 \quad (1.66)$$

Oran tahmin edilen parametre sayısı serbestlik derecesine sahip Ki-Kare dağılımına uyar. Hesaplanan deęer Ki-Kare tablo deęeri karşılaştırılıp anlamlılığı test edilir.

1.8. Tekrarlanan Başarısızlıklar

Yaşam analizinde bireyler ya da birimler, aynı olayı birden fazla kez yaşayabilirler. Bu durumda elde edilen veriler, tekrarlanan başarısızlık zamanı verileri olarak adlandırılır (Ghosh, 2000).

Tekrarlanan başarısızlık verilerinde bireyler bağımsız olarak örneklenmiştir, ancak aynı bireylerden elde edilen tekrarlar zamanları ile ilişkilidir. Bu ilişki, parametre tahminlerine ilişkin varyansın, deęerinden daha düşük tahmin edilmesine sebep olmaktadır (Eimermacher, 2008; Ünlü, 2013).

Tekrarlanan başarısızlıkların analizi son yıllarda üzerinde çalışılan bir konudur. Tekrarlanan başarısızlıklarda bireyler aynı olayı birden fazla yaşayabilirler. Bu sebeple aynı bireyden elde edilen tekrarlar zamanları ilişkilidir. Tekrarlanan başarısızlıklara astım atakları, KOAH hastalarının atakları, sıtma hastalarının nöbetleri, epilepsi hastalarındaki nöbetler, kanser hastalarındaki tekrarlayan tümörler, kalp krizlerindeki ataklar örnek olarak verilebilir.

Tekrarlanan başarısızlık verileri analiz edilirken önemli bir ayırım yapmak gerekir. Eđer tüm tekrar eden başarısızlıklar özdeş tedavi gören aynı başarısızlık türünde ise farklı bir analiz, eđer tekrar eden başarısızlıklar farklı kategoriden hastalıklar veya daha önemli derecedeki olaylar içeriyorsa farklı bir analiz uygulanması gerekmektedir (Eimermacher, 2008). Özdeş tekrarlanan olaylar için kullanılan analiz sayma süreci yaklaşımdır. Ancak tekrar eden başarısızlıklar, özdeş

olmayıp farklı kategoriler ve önemliliklere sahip ise tabakalı Cox modeli yaklaşımı tercih edilir (Kleinbaum ve Klein, 2012).

1.8.1. Tekrarlanan Başarısızlıkların Modellenmesi

Tekrarlanan başarısızlıkların modellenmesi için literatürde pek çok yöntem sunulmuştur. Bu yöntemler; zayıflık modeli, sayma süreci, Cox temelli tabakalı Cox yaklaşımlarıdır. (Chang ve Wang 1999; Gail ve ark. 1980; Huang ve Louis 1998; Huang, 2000; Lin ve ark., 1999; Prentice ve ark. 1981; Therneau ve Hamilton, 1997; Wang ve Chang 1999; Visser, 1996).

Cox temelli tabakalı yaklaşımlar; koşul 1 ve koşul 2 olarak adlandırılan Prentice, Williams ve Peterson'ın 1981'de sunduğu yaklaşımlar, Wei, Lin ve Weissfeld tarafından 1989'da sunulan marjinal yaklaşım, Andersen ve Gill'in 1982'de sunduğu yaklaşımdır.

Tekrar eden olaylar benzer ise Sayma süreci yaklaşımı ve zayıflık modeli kullanılabilir. Eğer tekrar eden olaylar farklı kategorilere ayrılabilirse veya olayların oluş sırası önemli ise tabakalı Cox modeli yaklaşımları kullanılabilir. (Kleinbaum ve Klein 2005).

Prentice, Williams ve Peterson'ın ileri sürdüğü koşul 1'de birinci olaya kadar olan süre, sonraki olaylar için risk kümesini etkilerken koşul 2'de daha sonraki tekrarlayan olaylar için risk kümesini etkilememektedir. Wei, Lin ve Weissfeld tarafından önerilen yaklaşımda ise çalışma başlangıcından itibaren zaman risk kümesini etkilemektedir.

1.8.2. Sayma Süreci

Sayma süreci yaklaşımı ilk olarak Aalen (1975) tarafından tanıtılmıştır. Sayma süreci yaklaşımı ile veriler sanki farklı deneklerden elde edilmiş gibi bağımsız olduğu düşünülür. Sayma süreci yaklaşımı korelasyonların yapısal olarak bağımsızlığını varsayar. Sayma süreci araştırılan zaman içerisinde meydana gelen olay sayısının modellenmesinde kullanılan stokastik bir süreçtir. $N(t)$ sayma süreci,

$N(0)=0$ ve $N(t) < \infty$ değerini almaktadır. $N(t)$, iki atlama noktası arasında sabit değer alan, atlama noktalarında +1 büyüklüğünde sıçrama yapan sağdan sürekli bir fonksiyondur.

$\{N_i(t), t \geq 0\}$ sayma süreci,

$N_i(t) = I(T_i^* \leq t, \varphi_i = 1)$ eşitliği ile gösterilir.

$N_i(t)$, başarısızlık anında bir kez sıçrama yaparak 1 değerini alır ve daha sonra sabit kalır.

$$N(t) = \sum_{i=1}^n N_i(t) \quad (1.67)$$

eşitliğinde ise $N(t)$, t zamanına kadar olan toplam başarısızlık sayısını ifade etmektedir.

Sayma süreci, tekrarlanan olayların benzerlik gösterdiği durumlarda kullanılır. Örneğin meme kanseri üzerine yapılan araştırmada, hastada tedavi edilen tümör tekrar ortaya çıktığında analiz çalışmasında sayma süreci yaklaşımının kullanılması uygundur. Sayma sürecinde kullanılan model, standart Cox orantılı hazard regresyon modelidir. Öncelikli olarak yapılması gereken, zamandan bağımsız değişken için orantılı hazard varsayımını incelemektir. Bir ya da daha fazla değişken orantılı hazard varsayımını sağlamazsa tabakalı Cox modeli ya da genişletilmiş Cox modeli kullanmak gerekmektedir.

Sayma yöntemi yaklaşımının veri girişini açıklamak için, 30 aylık bir gözlem süresi içerisinde A, B, C, D kişisinden oluşan bir örnek düşünelim. Bu örneğe ait sayma süreci yöntemi yaklaşımındaki veri girişi aşağıdaki çizelge 1.1 de olduğu gibi verilebilir.

Çizelge 1.1. Sayma Süreci Yaklaşımı Veri Düzenlenmesi

Kişi	Zaman Noktası	Başlangıç Zamanı	Bitiş Zamanı	Durum
A	1	0	3	1
A	2	3	9	1
A	3	9	21	1
A	4	21	23	0
B	1	0	3	1
B	2	3	15	1
B	3	15	30	1
C	1	0	17	1
D	1	0	12	0

Sayma sürecinde veri girişi yapılırken her birey için, her tekrar eden olaya ve buna eklenmiş takip süresi içindeki olay yaşanmamış zamana bağlı olan farklı satırlarda yazılır. Her satırda takip süresinin başlangıç zamanı ve bitiş zamanı yazılır. Tekrar eden olaya sahip birey için her defasında farklı giriş yapılarak risk aralığı değiştirilmektedir. Böylece birey, her başarısız olayı yaşadığında farklı risk kümesine dâhil edilir. Böylece oluşturulan risk kümeleri, Cox regresyon modelindeki olabilirlik fonksiyonuna etki eder. Veri girişindeki bu farklılık, tekrarlı olaylar ile bir kez gerçekleşen tekrarsız olaylar arasındaki önemli bir ayrımdır. Orantılı hazard varsayımını sağlayan zamandan bağımsız değişkenleri içeren tekrar eden başarısızlıkları içeren yaşam verisi analizinde birden fazla zaman aralığına sahip birey, son aralığı tamamlanana kadar risk kümesinde kalır. Buna karşılık, tekrarsız olay verisi için her birey, olaydaki başarısızlık zamanında risk kümesinden ayrılır. Burada iki ya da daha fazla zaman aralığına sahip birey için veride oluşturulan her farklı satıra, aynı kişi üzerinden elde edilen çeşitli sonuçlara rağmen birbirinden bağımsız gibi işlem yapılır. Bir kez gerçekleşen tekrarsız olay verisinde ise her farklı satır, farklı kişilerden elde edildiği için, birbirinden bağımsızdır. Sayma süreci yöntemi yaklaşımında veri girişi yapılırken, aynı birey için oluşturulan her farklı sıra, Cox model analizinde, farklı bireylerden elde edilmiş bağımsız veriler gibi düşünülürse, aynı bireylerden elde edilen gözlemlerin birbiriyle bağımlı olduğu ihmal edilmiş olur (Özkartal, 2009).

1.8.3. Prentice, Williams ve Peterson Yaklaşımı

Bu yaklaşım Prentice ve arkadaşları tarafından 1981’de sunulmuştur. Örneğin bu yaklaşımda kullanılan bir veri setinde maksimum tekrar eden başarısızlık sayısı 4 olsun. Bu durumda bu veri setinin analizinde tabaka 1, tabaka 2, tabaka 3 ve tabaka 4 olmak üzere başarısızlık sayısı kadar tabaka oluşturulur. Prentice, Williams ve Peterson yaklaşımı koşul 1 ve koşul 2 olmak üzere adlandırılan iki tekrarlanan olay arasındaki yaşam süresi ile ilgilenmektedir.

$(T_{i1}, T_{i2}, \dots, T_{ik})$, i . Bireye ait başarısızlık zamanlarını ve C_i , i . Bireye ait durdurma zamanını, $Z(t)=(Z_1(t), Z_2(t), \dots, Z_p(t))$, $t \geq 0$, t anındaki zamana bağlı açıklayıcı değişken vektörünü, $N(t)$ $[0, t]$ zaman aralığındaki sayma sürecini gösterdiğinde çalışmanın başlangıcından t anına kadar geçen süreyi modelleyen hazard fonksiyonu;

$$\lambda_{ik}(t \setminus N_i(t), Z_i(t)) = \lambda_{0k}(t) \exp(\beta_k^T Z_{ik}(t)) \quad (1.68)$$

Ardışık iki tekrarlanan başarısızlık arasında geçen süreyi modelleyen hazard fonksiyonu;

$$\lambda_{ik}(t \setminus N_i(t), Z_i(t)) = \lambda_{0k}(t - T_{i,k-1}) \exp(\beta_k^T Z_{ik}(t)) \quad (1.69)$$

şeklinde tanımlanmaktadır (Prentice ve ark 1981).

Prentice ve arkadaşları tarafından önerilen modelin birinci tekrarında elde edilen sonuçlar, Cox orantılı hazard modeli kullanılarak elde edilen sonuçlarla aynıdır. Prentice ve arkadaşları tarafından önerilen bu modelde, (k^*) . tekrarda risk kümesini oluşturan bireyler, k . başarısızlığı yaşayan bireylerden oluşur şartı vardır. Bu yüzden çalışmanın başlangıcında risk grubundaki sayı değişmektedir (Prentice ve ark, 1981).

Prentice, Williams ve Peterson yaklaşımı koşul 1 ve koşul 2 olmak üzere adlandırılan iki şekilde incelenebilir.

1.8.3.1. Koşul 1 Yaklaşımı

Bu yaklaşımda veri girişi sayma süreci veri girişi ile aynıdır. Koşul 1 yaklaşımı veri girişi Çizelge 1.1’de olduğu gibidir.

Koşul 1 yaklaşımı veri analizinde aynı birey için zamanın ilk başlangıç değeri 0 alınır, bitiş değeri ise ilk başarısız olayın gerçekleştiği zaman olarak tanımlanmaktadır. Aynı birey için ikinci aralık, ilk aralığın bitiş değeri ile başlayıp başarısızlık olayının ikinci kez tekrar ettiği zamanda sona erer. Tekrar eden başarısızlık sayısı kadar zaman noktası satırı veri girişine eklenir. Takip süresinde birey başarısızlık olayını yaşamadan gözlemden ayrılırsa son bitiş değeri ayrılma zamanı olarak veri girişine yazılmaktadır. Koşul 1 yaklaşımında ilk olaya kadar geçen süre, sonraki olaylar için risk kümesinin oluşumunu etkilemektedir.

1.8.3.2. Koşul 2 Yaklaşımı

Koşul 2 yaklaşımı ilk meydana gelen olayı 0 olarak başlatıp son olayla zamanı bitirmektedir. Koşul 2 yaklaşımında, tekrarlanan başarısızlıkların veri girişi yapılırken, yaşam süresi ilk olay için 0 ile başlar, sonraki olayla biter. Örneğin aynı birey için veri girişi yapılırken, ilk satırda başlangıç zamanı 0, bitiş ise ilk olay ile ikinci olay arasında geçen zaman olarak yazılır. Koşul 2 yaklaşımında veri girişi çizelge 1.3 de belirtildiği gibidir.

Çizelge 1.2. Koşul 2 Yaklaşımı Veri Girişi

Kişi	Zaman Noktası	Başlangıç Zamanı	Bitiş Zamanı	Durum
A	1	0	3	1
A	2	0	7	1
A	3	0	8	1
A	4	0	2	0
B	1	0	12	1
B	2	0	4	1
B	3	0	2	0

Koşul 2 yaklaşımında ilk başarısızlık olayına kadar geçen süre, ikinci ya da daha sonraki başarısızlık olayları için oluşturulan risk kümesine etki etmez. Tekrarlanan her başarısızlık olayı gerçekleştikten sonra riskli bireyi değerlendirirken

süre tekrar 0 olarak yazılır. Koşul 2 yaklaşımı özellikle bireyin yaşamış olduğu tekrarlanan başarısızlıkları arasındaki sürenin önemli olduğu durumlarda tercih edilmektedir (Kelly ve Lim, 2000).

1.8.4. Wei, Lin ve Weissfeld Yaklaşımı

Wei ve ark tarafından 1989 yılında önerilen bu yaklaşım marjinal yaklaşım olarak da adlandırılmaktadır. Bu yaklaşım k. olayın tekrarlanması gerçekleşene kadar geçen zamanı modellemektedir. Yani Wei, Lin ve Weissfeld yaklaşımı, çalışmanın başlangıcından tekrar eden olayın yaşanmasına kadar geçen toplam yaşam süresi ile ilgilenir (Mahe ve Chevret, 2001).

Wei, Lin ve Weissfeld yaklaşımında her tekrar eden olay için risk kümesi, başlangıç zamanından olay zamanına kadar geçen süre için belirlenir. Wei, Lin ve Weissfeld yaklaşımı farklı tipte tekrar eden olaylar görülmesi durumunda kullanılması önerilmektedir. Örneğin bir meme kanseri hastasının, hastalığın farklı evresine geçtiğinde değişik değerler için kontrol edilmesi için Wei, Lin ve Weissfeld yaklaşımının kullanılması önerilmektedir (Kleinbaum ve Klein, 2005).

Marjinal yaklaşımda veri girişi, tekrarsız başarısız olay verisinin standart girişi ile aynıdır. Başlangıç zamanı ve bitiş zamanı olarak adlandırılan sütunlar yerine yaşam süresi sütunu olarak tek bir zaman sütunu oluşturulur.

Marjinal yaklaşımında veri girişi aşağıdaki gibidir:

Çizelge 1.3. Marjinal Yaklaşımda Veri Girişi

Kişi	Zaman Noktası	Başlangıç Zamanı	Bitiş Zamanı	Yaşam Süresi	Durum
A	1			4	1
A	2			13	1
A	3			21	1
A	4			23	0
B	1			12	1
B	2			16	1
B	3			18	1
B	4			24	0

$(T_{i1}, T_{i2}, \dots, T_{ik})$, i . bireye ait başarısızlık zamanlarını ve C_i , i . bireye ait durdurma zamanını, $Z(t)=(Z_1(t), Z_2(t), \dots, Z_p(t))$, $t \geq 0$, t anındaki zamana bağlı açıklayıcı değişken vektörünü, k tabaka değişkenini, β_k , k . tabakaya ait parametre vektörünü, $N(t)$ $[0, t]$ zaman aralığındaki sayma sürecini gösterdiğinde i . birey için k . tekrarlama olayı olana kadar geçen süreyi modelleyen hazard fonksiyonu;

$$\begin{aligned} \lambda_{ik}(t|Z) \\ = \lambda_{0k}(t) \exp\left(\beta_k^T Z_{ik}(t)\right) \end{aligned} \quad (1.70)$$

Marjinal yaklaşımda tekrarlamalar arası bağımlılık modellenmemektedir. Bu yaklaşım sadece marjinal tehlikeleri modellediğinden, çalışmanın başlangıcındaki risk kümesindeki sayı değişmez.

1.8.5. Andersen ve Gill Yaklaşımı

(Andersen ve Gill 1982)'de sayma süreci yaklaşımını kullanarak Cox orantılı hazard modelindeki kısmi olabilirlik fonksiyonun birinci türevini yeniden formüle ederek tekrarlanan başarısızlık olaylarının analizini yapmışlardır. Guo ve ark., (2008); Kelly ve Lim (2000); Mahe ve Chevret (2001); Therneau ve Hamilton (1997), Andersen ve Gill yaklaşımı ve diğer yaşam analizleri ile gerçek ve simülasyon verileri üzerinde karşılaştırmalı sonuçları bu çalışmalarda yapılmış ve bu modellerin farklı sonuçlar verdiği görülmüştür.

$Z(t)=(Z_1(t), Z_2(t), \dots, Z_p(t))$, $t \geq 0$, t anındaki zamana bağlı açıklayıcı değişken vektörünü, β , parametre vektörünü göstermek üzere hazard fonksiyonu aşağıdaki gibi verilmektedir.

$$\lambda(t|Z) = \lambda_0(t) \exp\left(\beta^T Z(t)\right) \quad (1.71)$$

1.8.6. Zayıflık Modeli

Yaşam analizinde amaç tanımlanan değişkene ait başarısızlığı analiz etmektir. Yaşam verilerinin analizinde tüm bireylerin aynı başarısızlık riskine sahip olduğu aynı dağılımdan geldiği yani kitlenin homojen olduğu varsayılır. Fakat bir bireyin

başarısızlığı diğerinden farklı olabileceği için bireylerin oluşturduğu kitlenin homojen olmadığı ve heterojen bir karakteristiği olduğu ifade edilir. Heterojenliği değerlendirmek zor, ancak önemlidir. Ancak bu heterojenliği dikkate almamak ciddi sorunlara neden olmakta ve yanıltıcı sonuçlara yol açmaktadır. Heterojenlik gözlenebilen risk faktörlerinden ve açıklayıcı değişkenlerden kaynaklanabilir. Bu gözlenemeyen heterojenliği tanımlamak için zayıflık modeli kullanılmaktadır. Zayıflık modelinde kitleyi oluşturan bireylerin farklı zayıflıklara sahip olduğu ve zayıflık derecesi yüksek olan bireyin düşük olan bireyden daha önce başarısız olacağı varsayılmaktadır. Zaman ve yaş ilerledikçe zayıflık derecesi yüksek olan bireyin diğerlerine göre kitleden erken ayrılması kitlede daha uzun süre yaşayan bireylerin kalması ortalama ölüm hızını azaltır (Duchateau ve Janssen 2008).

Zayıflık modelinde ana düşünce, bireyler veya gruplar arasındaki heterojenliği ifade etmek için ölçülemeyen rastgele etkiyi hazard fonksiyonuna eklemektir. Zayıflık terimi farklı dağılım türlerine uyabilir. Genellikle Gamma ve ters-Gauss dağılımları zayıflık terimi için tercih edilmektedir. Bir çok paket programda Gamma ve ters-Gauss dağılımları zayıflık terimi için tercih seçenekleri arasında sunulur. Aslında gözlenemeyen riskler zayıflık olarak adlandırılır. Zayıflık terimi hazard fonksiyonuna çarpımsal olarak dahil edilir. Bunun sebebi bireyler ile kitle arasındaki ilişkinin yapısı, zayıflığın birimler arasındaki dağılımına bağlı olmasından kaynaklanmaktadır.

Zayıflık; bireylerin başarısızlık zamanları arasındaki farklılığa neden olan faktörler, zayıflık modelleri ise bireyler arasındaki heterojenliği açıklamak için ölçülemeyen rastgele etkiyi hazard fonksiyonuna dâhil edilmesiyle oluşan modeller olarak tanımlanabilir. Zayıflık terimini matematiksel olarak ifade etmenin bir yolu, her bireye ilişkin hazard hızı üzerinde orantılı biçimde hareket eden bir zayıflık terimi olduğunu varsaymaktır. Bu varsayıma dayalı model orantılı zayıflık modelidir. Orantılı zayıflık modeli bireylere ilişkin heterojenliği zaman boyutuna bağlı olmadan açıklamakta ve heterojenlikle ilgili sorunun çözümünde kullanılmaktadır (Başar, 2016; Wienke, 2010)

Zayıflık modelleri, yaşam analizinde en sık kullanılan model olan Cox (1972) modeli olarak da bilinen orantılı hazard modelinin uzantısıdır. (Zeng ve Lin, 2008)

Genellikle sađlık alanındaki arařtırmalar, yařam analizine konu olan kitlenin homojen olduđu varsayımına dayanmaktadır. Oysaki bu verilerin çođunda gözlenemeyen heterojenlik vardır. Zayıflık terimi Z ilk olarak Vaupel ve ark.(1979) tarafından önerilmiřtir. Bu Z terimi literatürde yařam verileri için açıklanamayan etki ya da rastgele etki olarak literatürde adlandırılmıřtır. Aslında zayıflık kavramı esas olarak Greenwood ve Yule'nin (1920) "kaza yapma oranı" çalıřmasından ileri gelmektedir. Fakat bu çalıřmada heterojenlik ve zayıflık terimleri ifade edilmemiřtir.

Clayton (1978) yılında ailelerdeki kronik hastalıkların insidansı çalıřmasında ise zayıflık terimini kullanmadan çoklu yařam verilerinin analizini yapmıřtır. Ayrıca Clayton bu çalıřmasında zayıflık dađılımında gamma dađımlı zayıflık modeli kullanmıřtır. Literatür incelendiđinde o yıllarda yapılan çalıřmalarda zayıflık modelinin analizinde ana düşünce zayıflık dađılımının analizi için hangi dađılımın kullanılacađına karar vermektir (Chen ve ark 2014). Oakes (1982)'de ilk olarak yaptıđı çalıřmada zayıflık modeli için gamma dađımlı zayıflık modelini kullanmıřtır. Clayton ve Cuzick (1985)'de paylařılmıř zayıflık modelini tanıtmıřlardır. Paylařılmıř zayıflık modeli genellikle çok deđişkenli yařam analizi verilerini analiz etmek için kullanılmıřtır (Cook ve Lawless, 2007; Duchateau ve ark., 2003; Oakes, 1989; Petersen, 1998; Zeng ve ark., 2009).

Hougaard (1987) paylařılmıř zayıflık modelde zayıflık terimi için pozitif kararlı dađılımı önermiřtir. Klein ve Moeschberger (2003) ve Duchateau ve Janssen (2008)'de yarı parametrik paylařılmıř zayıflık modeline EM (Beklenti en büyükleme) algoritmasının uygulamasını yapmıřlardır. Sahu (1997)'de Bayesci paylařılmıř zayıflık modelini açıklamaya çalıřmıřtır. Xue ve Brookmeyer (1996) paylařılmıř zayıflık modelinin eksikliklerini arařtırmıřlardır. Pickles ve ark. (1994) ve Yashin ve ark (1995)'de paylařılmıř zayıflık modelini aynı gruptaki gözlemlerde iliřkili zayıflık modeli ile analizi öne sürerek açıklamıřlardır. Hougaard (1986) Weibull hazard fonksiyonu ile paylařılmıř zayıflık modelini açıklamaya çalıřmıřlardır. İbrahim ve ark (2001)'de yarı parametrik paylařılmıř zayıflık modellerine Bayesci yaklařımı açıklamıřlardır.

Vaupel (1991a, 1991b), Vaupel ve ark., (1991, 1992), ve Nielsen ve ark. (1992)'de genetik çalıřmalarında iki deđişkenli yařam analizi verilerinde zayıflık

modelini kullanmışlardır. Vaupel ve ark., (1979), Oakes (1982), Clayton ve Cuzick (1985), ve Andersen ve ark., (1993)'de yaptıkları çalışmalarda zayıflık terimi için gamma dağılımını kullanmışlardır. Whitmore ve Lee (1991) temel hazard fonksiyonunu üstel dağılıma sahip olduğu ters-Gauss paylaşılmış zayıflık modelini önermişlerdir. Cangdon (1995)'de farklı zayıflık dağılımlarının gamma, ters Gauss, pozitif kararlı gibi dağılımların etkilerini araştırmıştır. Balakrishnan ve Peng (2006)'da yaptıkları çalışmada zayıflık terimi için genelleştirilmiş gamma dağılımını önermişlerdir. Yashin ve ark. (1995)'de güç varyans dağılımlı zayıflık modelini geliştirmiştir. Hougaard (2000)'de ise zayıflık terimi için pozitif kararlı dağılımı ve lognormal dağılımın kullanımını önermiştir. Bu çalışmada Hougaard gamma zayıflık modeli ile güç varyans fonksiyonu dağılımlı zayıflık modelinin karşılaştırmasını yapmıştır. Aalen ve Tretli (1999) yaşam verilerinin analizinde birleşik Poisson zayıflık modelini önermişlerdir. Hanagal (2010)'da zayıflık terimini karma dağılımlardan yararlanarak açıklamaya çalışmıştır. Hanagal ve Dabade (2013)'de temel hazard fonksiyonunun farklı dağılımları ile paylaşılmış gamma zayıflık modelinin çözümünü yapmışlardır.

Commenges ve Andersen (1995) orantılı hazard modeline zayıflık teriminin eklenilmesinin heterojenliğin açıklanmasında modelin anlamlı olup olmadığını skor testi ile yapılabileceğini savunmuşlardır.

McGilchrist ve Aisbett (1991) zayıflık modelinin parametrelerinin en çok olabirliklerinin tahminlerini bulmak için ceza fonksiyonunu önermiştir.

1.8.6.1. Paylaşılmamış Zayıflık Modeli

Paylaşılmamış zayıflık modeli, yaşam süresi ilişkili olmayan bireyler arasındaki modellemeyi yapmaktadır.

Zayıflık modeli aşağıdaki gibi verilmektedir.

$$h(t|Z) = Zh(t) \quad (1.72)$$

Z: bireyler arasındaki heterojenliği ifade eden ölçülemeyen ve negatif olmayan rastgele değişkeni ifade etmektedir. Z değişkeni zayıflık terimi olarak da ifade

edilmektedir. $h(t)$: t zamanındaki hazard fonksiyonunu ifade eder. Zayıflık modeli hazard fonksiyonu üzerinde gözlenemeyen etkiyi (Z) açıklar.

$Z > 1$ ise; araştırılan hastaların (bireylerin) değişkenler tarafından ifade edilemeyen sebeplerden ötürü daha kırılğan olduğu ve artan başarısızlık riskine sahip olduğu kararına varılır.

$Z < 1$ ise; araştırılan hastaların (bireylerin) değişkenler tarafından ifade edilemeyen sebeplerden ötürü daha kırılğan olduğu ve azalan başarısızlık riskine sahip olduğu kararına varılır.

Modelde yaşam fonksiyonu;

$$S(t|Z) = [S(t)]^Z \quad (1.73)$$

şeklindedir. $S(t)$, standart yaşam modelinin yaşam fonksiyonudur. Koşulsuz yaşam fonksiyonu ise aşağıdaki gibi hesaplanır.

$$S_\theta(t) = \int_0^\infty [S(t)]^z g(z) dz \quad (1.74)$$

Eşitlikteki θ indisi zayıflık teriminin varyansına bağımlılığı ifade eder.

Zayıflık terimi, gamma dağılımına sahip ise, olasılık yoğunluk fonksiyonu;

$$g(z) = \frac{z^{1/\theta-1} \exp(-z/\theta)}{\Gamma(1/\theta) \theta^{1/\theta}} \quad (1.75)$$

ve yaşam fonksiyonunda;

$$S_\theta(t) = [1 - \theta \ln S(t)]^{-1/\theta} \quad (1.76)$$

şeklinde olur zayıflık teriminin, ters Gauss dağılımına sahip olması durumunda olasılık yoğunluk fonksiyonu ve yaşam fonksiyonları sırasıyla;

$$g(z) = \left(\frac{1}{2\pi\theta z^3}\right)^{1/2} \exp\left[-\frac{1}{2\theta}\left(z - 2 + \frac{1}{z}\right)\right] \quad (1.77)$$

$$S_{\theta}(t) = \exp \left[\frac{1}{\theta} \left(1 - (1 - 2\theta \ln S(t))^{1/2} \right) \right] \quad (1.78)$$

şeklindedir.

1.8.6.2. Paylaşılmış Zayıflık Modeli

Paylaşılmış zayıflık modeli ilk defa Clayton (1978) tarafından tanıtılmış Hougaard (2000) tarafından tüm ayrıntılarıyla açıklanılmıştır.

Paylaşılmış zayıflık modeli benzer organlardan elde edilen başarısızlıkların ve tekrarlanan başarısızlıkların modellenmesinde kullanılırken, bu model bireylerin aynı zayıflık terimini paylaşmalarına imkân sağlamaktadır. Bireyler arasındaki bağımlılık aynı zayıflık teriminin paylaşılmasından ileri gelmektedir. Paylaşılmış zayıflık modelinin ismi gruptaki bireylerin aynı zayıflığı paylaşmasından kaynaklanmaktadır.

Paylaşılmış zayıflık modelinde bireyler aynı zayıflık terimini paylaşmaktadırlar. n gruptan oluşan bir veri analizi için zayıflık modeli

$$h_{ij}(t/z_i) = z_i h_{ij}(t), \quad j = 1, 2, \dots, n_i \quad (1.79)$$

şeklindedir. Burada n_i , i . gruptaki birim sayısını ifade eder ve $h_{ij}(t) = h(t/x_{ij})$ şeklinde yazılır. Koşulsuz yaşam fonksiyonu ise;

$$S_{ij}(t/z_i) = [S_{ij}(t)]^{z_i} \quad (1.80)$$

şeklindedir. n grup sayısını gösterdiğinde, olabilirlik fonksiyonu;

$$L_{ij}(z_i) = \frac{S_{ij}(t_{ij}/z_i)}{S_{ij}(t_{0ij}/z_i)} [h_{ij}(t_{ij}/z_i)]^{d_{ij}} = \left[\frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})} \right]^{z_i} [z_i h_{ij}(t_{ij})]^{d_{ij}} \quad (1.81)$$

şeklindedir. $D_i = \sum_{j=1}^{n_i} d_{ij}$ olursa, i . grup için olabilirlik fonksiyonu,

$$L_i(z_i) = z_i^{D_i} \prod_{j=1}^{n_i} \left[\frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})} \right]^{z_i} [h_{ij}(t_{ij})]^{d_{ij}} \quad (1.82)$$

şeklindedir. Z_i 'ye göre integral alınarak i. grup olabirlik fonksiyonu

$$L_i = \int_0^{\infty} L_i(z_i)g(z_i)dz_i \quad (1.83)$$

şeklindedir. Eşitlik (1.84)'de zayıflık terimi Gamma dağılımı olduğunda L_i 'nin hesaplanması ifade edilmektedir.

$$L_i = \left[\prod_{j=1}^{n_i} [h_{ij}(t_{ij})]^{d_{ij}} \right] \frac{\Gamma(1/\theta + D_i)}{\Gamma(1/\theta)} \theta^{D_i} \left[1 - \theta \sum_{j=1}^{n_i} \ln \frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})} \right] \quad (1.84)$$

Paylaşılmış modelin bazı eksik yönleri ilişkili zayıflık modelinin gerekliliğini ortaya çıkarmıştır.

1.8.6.3. İlişkili Zayıflık Modeli

İlişkili zayıflık modeli genellikle iki değişkenli yaşam analizi verilerinde kullanılmaktadır. Pickles ve ark. (1994), İngiltere'deki ergenlik çağındaki ikizlerin antisosyal davranışları üzerindeki araştırmalarında ilişkili gamma zayıflık modelini uygulamışlardır. Yashin ve Iachine (1995)'de iki değişkenli yaşam analizi verilerinde ilişkili gamma zayıflık modelini kullanmışlardır. Santos ve ark. (1995)'de gamma ve lognormal dağılımlı ilişkili zayıflık modelini tekrarlayan başarısızlıkları olan göğüs kanserli hastalarda uygulamışlardır. Manatunga ve Oakes (1999)'da diyabet hastalarından alınan verilerin analizini pozitif durağan dağılımlı ilişkili zayıflık modeli ile yapmışlardır. Pozitif durağan dağılım hem hazard hem de marginal şartlı model için seçenekler sunar.

1.8.6.4. Gamma Zayıflık Modeli

Gamma zayıflık modeli zayıflık modeli için ilk akla gelen ve tercih edilen modeldir. Lakin Hougaard (2000) yılındaki çalışmasında bahsettiği gibi bazı istenmeyen özellikleri de vardır bu sebeple zayıflık teriminin dağılımı için Gamma dağılımı, lognormal dağılım, Ters Gauss dağılım, pozitif durağan dağılım, güç varyans fonksiyonları da kullanılabilir (Duchateau ve Janssen, 2008).

Gamma dağılımının Laplace dönüşümü ile birikimli yoğunluk fonksiyonu, hazard fonksiyonu, yaşam fonksiyonu kolaylıkla elde edilebilir.

Zayıflık modellerinin çoğunda zaman sonsuza giderken zayıflık modeli gamma dağılımına yakınsamaktadır (Abbring ve Berg, 2007; Başar, 2016) Gamma dağılımında k (konum) parametresi 1 olarak alındığında gamma dağılımı üstel dağılıma dönüşmektedir. k değeri büyüdükçe gamma dağılımın simetrikliği artmakta ve normal dağılıma benzemektedir.

$$f(z) = \frac{\lambda^k z^{k-1} e^{-\lambda z}}{\Gamma(k)}; \quad k, z > 0 \quad (1.85)$$

şeklinde ifade edilmektedir. Burada k , konum ve λ da ölçek parametresidir.

O halde Laplace dönüşümü;

$$\begin{aligned} L(u) &= \frac{1}{\Gamma(k)} \lambda^k \int_0^{\infty} e^{-uz} z^{k-1} e^{-\lambda z} dz \\ &= \frac{\lambda^k}{(\lambda + u)^k} \frac{1}{\Gamma(k)} (\lambda + u)^k \int_0^{\infty} z^{k-1} e^{-(\lambda+u)z} dz \\ &= \left(1 + \frac{u}{\lambda}\right)^{-k} \end{aligned} \quad (1.86)$$

k ve $(\lambda + u)$ parametreleri olmak üzere

birinci türev,

$$L'(u) = -\frac{k}{\lambda} \left(1 + \frac{u}{\lambda}\right)^{-k-1} \quad (1.87)$$

k ve $(\lambda + u)$ parametreleri olmak üzere

ikinci türev,

$$L''(u) = \frac{k(k+1)}{\lambda^2} \left(1 + \frac{u}{\lambda}\right)^{-k-2} \quad (1.88)$$

Bu eşitliklerde $u=0$ iken beklenen değer ve varyans;

$$E(Z) = \frac{k}{\lambda} \quad (1.89)$$

$$V(Z) = \frac{k}{\lambda^2} \quad (1.90)$$

$k = \lambda$ olduğunda dağılım standardize edildiğinde;

$E(Z) = 1$ ve zayıflık teriminin varyansı da $\sigma^2 = \frac{1}{\lambda}$ şeklindedir.

Z rastgele zayıflık değişkeni $Z \sim \Gamma\left(\frac{1}{\sigma^2}, \frac{1}{\sigma^2}\right)$ ile gamma dağılımına sahiptir.

Gamma dağılımının olasılık yoğunluk fonksiyonu, koşulsuz yaşam fonksiyonu, koşulsuz olasılık yoğunluk fonksiyonu, hazard fonksiyonu Laplace dönüşümü kullanılarak sırasıyla;

$$f(z) = \frac{1}{\Gamma\left(\frac{1}{\sigma^2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{1}{\sigma^2} z \sigma^2 - 1} \exp\left(-\frac{z}{\sigma^2}\right) \quad (1.91)$$

$$S(t) = L(H_0(t)) = \frac{1}{(1 + \sigma^2 H_0(t))^{\frac{1}{\sigma^2}}} \quad (1.92)$$

$$f(t) = \frac{h_0(t)}{(1 + \sigma^2 H_0(t))^{\frac{1}{\sigma^2} + 1}} \quad (1.93)$$

$$h(t) = \frac{h_0(t)}{1 + \sigma^2 H_0(t)} \quad (1.94)$$

şeklindedir (Wienke, 2010).

Yaşayan bireyler için,

$$\begin{aligned} f(z|X, T > t) &= \frac{S(t|X, z)f(z)}{S(t|X)} \\ &= \exp(-zH_0(t)e^{\beta^T X}) Z^{\frac{1}{\sigma^2} - 1} \exp\left(-z\left(\frac{1}{\sigma^2} + H_0(t)e^{\beta^T X}\right)\right) \end{aligned}$$

$$= \frac{\frac{1}{\sigma^2} + H_0(t)e^{\beta^T X}}{\Gamma\left(\frac{1}{\sigma^2}\right)} Z^{\frac{1}{\sigma^2}-1} \exp\left(-z\left(\frac{1}{\sigma^2} + H_0(t)e^{\beta^T X}\right)\right) \quad (1.95)$$

t zamanında başarısızlığa uğrayan bireyler için;

$$\begin{aligned} f(z|X, T = t) &= \frac{f(t|X, z)f(z)}{f(t|X)} \\ &= \frac{zh_0(t) \exp(-zH_0(t)e^{\beta^T X}) Z^{\frac{1}{\sigma^2}-1} \exp\left(-\frac{z}{\sigma^2}\right)}{\Gamma\left(\frac{1}{\sigma^2}\right) \sigma^{\frac{2}{\sigma^2}} h_0(t) (1 + \sigma^2 H_0(t)e^{\beta^T X})^{-\frac{1}{\sigma^2}-1}} \\ &= \frac{\frac{1}{\sigma^2} + H_0(t)e^{\beta^T X}}{\Gamma\left(\frac{1}{\sigma^2} + 1\right)} Z^{\frac{1}{\sigma^2}+1-1} \exp\left(-z\left(\frac{1}{\sigma^2} + H_0(t)e^{\beta^T X}\right)\right) \quad (1.96) \end{aligned}$$

Sırasıyla t anındaki ölümlerin ortalaması, t anından sonraki ölümlerin ortalaması; t anında ölen bireylerin varyansı; t anında yaşayan bireylerin varyansı,

$$E(Z|X, T = t) = \frac{1 + \sigma^2}{1 + \sigma^2 H_0(t)e^{\beta^T X}} \quad (1.97)$$

$$E(Z|X, T > t) = \frac{1}{1 + \sigma^2 H_0(t)e^{\beta^T X}} \quad (1.98)$$

$$V(Z|X, T = t) = \frac{\sigma^2(1 + \sigma^2)}{(1 + \sigma^2 H_0(t)e^{\beta^T X})^2} \quad (1.99)$$

$$\begin{aligned} V(Z|X, T > t) \\ &= \frac{\sigma^2}{(1 + \sigma^2 H_0(t)e^{\beta^T X})^2} \quad (1.100) \end{aligned}$$

şeklinde ifade edilebilir.

Eğer varyans 0 ise, heterojenlik yoktur. Bireye ilişkin hazard hızı ile kitleye ait hazard hızı aynıdır. Varyans ve zaman artarken payda büyüyecek ve bireye ilişkin

hazard hızı artmayı sürdürse dahi belli bir süre sonunda kitlede hazard hızı azalmaya başlayacaktır.

1.8.6.5. Ters - Gauss Zayıflık Modeli

Ters-Gauss zayıflık modeli, Hougaard (2000) tarafından önerilmiştir. Ters-Gauss dağılımının olasılık yoğunluk fonksiyonu;

$$f(z) = \frac{\sqrt{\lambda}}{\sqrt{2\pi z^3}} \exp\left(-\frac{\lambda}{2\mu^2 z} (z - \mu)^2\right) \quad \mu > 0, \lambda > 0 \quad (1.101)$$

Ters Gauss dağılımının Laplace dönüşümü yapıldığında;

$$\begin{aligned} L(u) &= E(e^{-uz}) \\ &= \exp\left(-\frac{\lambda \sqrt{1 + \frac{2\mu^2 u}{\lambda}}}{\mu} + \frac{\lambda}{\mu}\right) \\ &= \exp\left(\frac{\lambda}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 u}{\lambda}}\right)\right) \end{aligned} \quad (1.102)$$

Laplace dönüşümlerinin birinci ve ikinci türevleri alındığında;

$$L'(u) = -\frac{\mu}{\sqrt{1 + \frac{2\mu^2 u}{\lambda}}} \exp\left(\frac{\lambda}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 u}{\lambda}}\right)\right) \quad (1.103)$$

$$L''(u) = \frac{\mu^3}{\lambda \left(1 + \frac{2\mu^2 u}{\lambda}\right)^{\frac{3}{2}}} \exp\left(\frac{\lambda}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 u}{\lambda}}\right)\right) + \frac{\mu^2}{1 + \frac{2\mu^2 u}{\lambda}} \exp\left(\frac{\lambda}{\mu} \left(1 - \sqrt{1 + \frac{2\mu^2 u}{\lambda}}\right)\right) \quad (1.104)$$

Beklenen değer ve varyans;

$$E(Z) = -L'(0) = \mu \quad (1.105)$$

$$V(Z) = L''(0) - (L'(0))^2 = \frac{\mu^3}{\lambda} \quad (1.106)$$

$\mu = 1$ olarak alınırsa;

$$E(Z) = \mu = 1 \quad (1.107)$$

$$V(Z) = \frac{\mu^3}{\lambda} = \frac{1}{\lambda} \quad (1.108)$$

$$L(u) = e^{\frac{1}{\sigma^2}(1-\sqrt{1+2\sigma^2u})} \quad (1.109)$$

şeklindedir.

Dönüşüm sayesinde yaşam fonksiyonu;

$$S(t) = e^{\frac{1}{\sigma^2}(1-\sqrt{1+2\sigma^2H_0(t)})} \quad (1.110)$$

$$h(t) = \frac{h_0(t)}{(1 + 2\sigma^2H_0(t))^{\frac{1}{2}}} \quad (1.111)$$

şeklindedir.

t zamanında yaşayan bireyler için;

$$\begin{aligned} f(z|X, T > t) &= \frac{S(t|X, z)f(z)}{S(t|X)} \\ &= \frac{1}{\sqrt{2\pi^2z^3}} \exp\left(-\frac{(z - (1 + 2\sigma^2H_0(t)e^{\beta^T X})^{-\frac{1}{2}})^2}{\frac{2\sigma^2z}{1 + 2\sigma^2H_0(t)e^{\beta^T X}}}\right) \end{aligned} \quad (1.112)$$

$$E(Z|X, T > t) = \frac{1}{\sqrt{1 + \sigma^2H_0(t)e^{\beta^T X}}} \quad (1.113)$$

$$V(Z|X, T > t) = \frac{\sigma^2}{\sqrt{1 + \sigma^2 H_0(t) e^{\beta^T X}}} \quad (1.114)$$

Yapılan bu çalışmada amaç, yaşam analizlerinde karşılaşılan tekrarlanan başarısızlık verilerini en iyi açıklayan modelleri belirlemek ve heterojen bir verinin analizinde ölçülemeyen rastgele etkiyi hazard modele ekleyerek modelleme yapmaktır.

2. GEREÇ VE YÖNTEM

Ankara Üniversitesi Göğüs Hastalıkları Hastanesinden 2000 yılından günümüze kadar olan 500 KOAH şüpheli hasta dosyası araştırılıp, Kronik Obstrüktif Akciğer Hastalığı (KOAH) tanısı için ICD 10 kodlarına göre J44 tanısı konulan hasta dosyaları retrospektif olarak incelenerek 125 adet hasta verisi elde edilmiştir. Bu araştırmadaki istatistiksel değerlendirme SPSS 21.0 ve R paket programı ve Stata 14 demo sürümü kullanılarak yapılmıştır. Hastaların semptom başlangıç tarihinden ilk KOAH atağı geçirdikleri süre (ay olarak) 1. Yaşam süresi olarak alınmış, daha sonra ikinci atak ile birinci atak arasında geçen süre 2. Yaşam süresi olarak tanımlanmış, 2. ve 3. Atak arasında geçen süre ise 3. Yaşam süresi olarak ifade edilmiştir. Atak sayısı min=1, max=12 olarak gözlemlenmiştir. KOAH hastalarının ağır 2 ya da daha fazla sayıda atak geçirmeleri başarısızlık olarak ifade edilmiştir. KOAH tanısı almış fakat ayakta hafif ve bir atak geçiren hastalar durdurulmuş olarak tanımlanmıştır. Araştırmaya alınan 125 hastada 50'sinde (%40) başarısızlık, 75'inde (%60) durdurma gözlemlenmiştir. KOAH tanısı Global Initiative For Chronic Obstructive Lung Disease (GOLD)'na göre (FEV_1/FVC oranının %70'in altında olanlar ve yine (GOLD)'a göre hafif ($FEV_1 \geq 80\%$), orta ($50\% \leq FEV_1 \leq 80\%$), ileri ($30\% \leq FEV_1 \leq 50\%$), ve çok ileri ($FEV_1 < 30\%$) incelenerek 125 adet olgu KOAH tanısı kesinleştirilerek çalışmaya alındı. Bronşektazi, astım, interstisyel akciğer hastalığı, sigara bırakma polikliniğine başvuran hastalar KOAH tanı kriterlerine uymadığı için çalışma dışı bırakıldı. Olguların yaş, cinsiyet, tanı tarihi, VKİ, mMRC skoru, sigara miktarı, SFT (FEV_1 , FEV_1/FVC , MMF), AKG (PaO_2 , $PaCO_2$, SaO_2 , P_h) ekoda ölçülen sPAB, ataklardaki beyaz küre sayısı, balgam kültürü sonuçları ve varsa ek hastalıklar kaydedildi.

KOAH tanısı spirometri yapılarak kolayca konulabilir ancak hastalığın ağırlığı ekstrapulmoner bulgular ve komorbiditelerle de ilişkilidir. KOAH hastalığının kesin tanısı için spirometrik olarak havayolu obstrüksiyonunun gösterilmesi gerekir. KOAH'ın evrelendirilmesinde zorlu ekpirasyon eğrisi üzerinde hesaplanan birinci saniyedeki zorlu ekspiratuar volüm FEV_1 ve FEV_1/FVC oranındaki azalma kriter

olarak kullanılmaktadır. KOAH tanısında hastalık şiddetinin, seyirinin ve prognozunun belirlenmesinde solunum fonksiyon testleri (SFT) kullanılır. Spirometri KOAH tanısının konulması, hastalık şiddetinin belirlenmesi ve hastalık seyirinin izlenmesinde kullanılır. Bu amaçla uygulanan zorlu vital kapasite (FVC) manevrasında toplam akciğer kapasitesi düzeyine kadar derin bir inspirasyondan sonra rezidüel volüm düzeyine kadar hızlı ve zorlu ekspirasyon yaptırılarak zorlu ekspiratuar eğri elde edilir. Bu manevra bir plato oluşuncaya kadar devam eder. FEV₁ KOAH'da en sık önerilen izlemdir. FEV₁ ölçümünün kolaylığı ve iyi tanımlanmış olması sebebiyle havayolları obstrüksiyonunun değerlendirilmesinde tercih edilen bir parametredir. FEV₁/FVC oranı özellikle KOAH'ın erken evrelerinde obstrüksiyonunu daha iyi yansıtmaktadır. Ayrıca FEV₁/ FVC oranı sağlıklı kişilerde yaşla birlikte azalma eğilimindedir ve KOAH tanısı konulurken FEV₁/ FVC oranının %70'in altında olması ve diğer semptomların varlığı KOAH'ın tanısını yaklaştırır. Ancak şiddetin belirlenmesinde FEV₁ değeri baz alınmaktadır (Ulubay ve Yıldız, 2013).

Aşağıdaki çizelge 2.1'de çalışmamızda kullandığımız FEV₁ sınıflandırılması verilmiştir.

Çizelge 2.1. GOLD Evrelemesi

POST BRONKODİLATÖR FEV ₁ / FVC <0,70		
GOLD 1	Hafif	FEV ₁ > %80
GOLD 2	Orta	FEV ₁ %80-50
GOLD 3	İleri	FEV ₁ %50-30
GOLD 4	Çok İleri	FEV ₁ < %30

FEV₁ ölçümü KOAH'lı hastaların evrelendirilmesinde önemli ve yararlı olmasına rağmen KOAH'lı hastalarda farklı değişkenler de alınır. KOAH semptomlarından birisi dispne'dir. Dispnenin KOAH'lı hastalarda değerlendirilmesi farmakolojik ve rehabilitasyon programlarına olan yanıtın değerlendirilmesine imkân sağlamaktadır. Spirometre gibi fizyolojik ölçümler dispne derecesini göstermezler. Bu sebeple dispne ölçümü pek çok şekilde yapılır. Ankara Üniversitesi Hastanesi'nde bu ölçüm hastanın anemnezine göre yapılmıştır ve skala çizelge 2.2 de verilmiştir.

Çizelge 2.2. Modifiye MRC Skalası (mMRC)

MODİFİYE MRC SKALASI	
EVRE 0	Ağır eforla nefesim daralıyor
EVRE 1	Düz yolda hızlı yürürken veya yokuş çıkarken nefesim daralıyor
EVRE 2	Nefes darlığı nedeniyle düz yolda yaşlılarımdan yavaş yürüyorum veya düz yolda yürürken durup dinlenmem gerekiyor
EVRE 3	Düz yolda 100 m veya birkaç dakika yürüdüğümde durup dinlenmem gerekiyor
EVRE 4	Nefes darlığı nedeniyle evden çıkamıyorum veya giyinirken bile nefesim daralıyor

Bu çalışmada atak sıklıklarının öngörüsünü belirlemek ve diğer faktörlerle (FEV₁, Dispne ve sigara faktörleriyle) ilişkisini belirlemek için öncelikle oldukça heterojen bir yapıya sahip olan KOAH atağı klinik olarak incelendi.

KOAH atağı; KOAH hastalarında dispnede artış, balgam volüm ve renginde değişiklik, öksürükte şiddetlenme ve hastanın solunumsal semptomlarında tedavide değişiklik gerektirecek düzeyde kötüleşmesi olarak tanımlanan akut bir olaydır. Ataklar KOAH hastalarının yaşam kalitesinde önemli bir role sahiptir. Atakların sıklığı ve şiddeti KOAH'ın şiddetinin ilerlemesinin yaşam kalitesine etkisinin ve mortalitenin belirteçleri olarak kullanılmaktadır. KOAH hastalığı ilerlediğinde alevlenmeler daha sık ve şiddetli olmaktadır. Bazı KOAH'lı hastalar daha sık atak geçirirken bazıları nisbeten az geçirmektedirler. Genellikle bir yıl içerisinde iki ya da daha fazla sayıda KOAH atağı geçiren hastalar sık atak geçiren grup olarak adlandırılabilir (Ulubay ve Yıldız, 2013).

KOAH hastalarından elde edilen veri tekrarlayan başarısızlıklar içerdiği için bu yaşam analizi verisinde sürekli alışılagelmiş olarak yaşam analizlerinde kullanılan normal bir Cox regresyon analizi kullanılmaz. Tekrarlayan yaşam analizi verilerinde kullanılabilen çözümler tabakalandırılmış Cox yaklaşımları ve zayıflık modeli ile çözümlene yapıldı.



Şekil 2.1. KOAH Hastalarında Tekrarlanan Başarısızlık Zamanı Verisi

Şekil 2.1 deki hastasının tekrarlanan 3 başarısızlığından 3 ayrı yaşam süresi elde edilmektedir. Bu elde edilen başarısızlık zamanları yaşam süreleri olarak çizelge 2.3 de gösterilmiştir.

Çizelge 2.3. KOAH Hastalarından Veri Elde Edilmesi

	1. Yaşam Süresi	2. Yaşam Süresi	3. Yaşam Süresi
A Hastası	3	4	2

Tekrarlayan başarısızlıkların analiz edilebilmesi için veri girişinin doğru bir şekilde yapılması gerekmektedir. Örneğin bizim araştırmamızda şekil 2.1’de görüldüğü gibi A hastası için 3 farklı yaşam süresi elde edildi. Bu farklı yaşam sürelerini analiz etmek için veri girişi normal Cox regresyon yönteminden farklı biçimde yapıldı. Araştırmamız da iki tane hastaya ait veri girişi çizelge 2.4’de olduğu gibidir.

Çizelge 2.4. Tekrarlanan Başarısızlık Veri Girişi

Hasta	Atak Sayısı	ToplamAtak sayısı	Sigara (paket/yıl)
1	1	4	45
1	2	4	60
1	3	4	50
1	4	4	50
2	1	6	56
2	2	6	56
2	3	6	56
2	4	6	56
2	5	6	57
2	6	6	50

Araştırmada R version 2.15.1 programının Parfm Paketi, SPSS 21.0 paket programı, STATA 14. Demo paket programları kullanıldı.

Araştırmada Orantılı hazard varsayımının sağlanıp sağlanmadığı grafik yöntemi ve Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi kullanılarak araştırıldı.

KOAH hastalarının Ataksız yaşam süreleri Kaplan-Meier yöntemi ile incelendi.

KOAH hastalarının veri analizi için parametrik regresyon üstel, lognormal, Weibull, Gompertz regresyon modelleri ile analiz yapıldı. Model seçiminde AIC ve BIC bilgi kriterleri ve grafiksel yöntem birlikte kullanıldı. Zayıflık terimi Z için ise Gamma, Ters-Gauss, lognormal ve Positive Stable dağılımlarından hangisinin kullanılacağına karar vermek için yine AIC ve BIC bilgi kriterleri ve grafiksel yöntem birlikte kullanıldı. Sonuçta tekrarlanan başarısızlıklara sahip olan KOAH hastalarının veri analizi için Weibull regresyon modeli dikkate alınarak ve zayıflık terimi için ise Gamma dağılımının kullanılmasına karar verilmiş ve model kurulmuştur.

Zayıflık modelinin anlamlı olup olmadığı θ (zayıflık dağılımının varyansının tahmin edicisi) parametresinin 0'a karşı testi olabilirlik oran testi ile yapıldı. Testin sonucu anlamlı çıktığından heterojenliğin olduğu sonucuna varıldı. Eğer testin sonucu anlamlı çıkmasaydı heterojenliğin olmadığı sonucuna varılacak ve başka modelleme yöntemleri ile analize gidilecekti. Ayrıca zayıflık dağılımının varyansı kitlenin heterojenlik derecesini belirtmektedir.

θ zayıflık dağılımının varyansının tahmin edicisidir. Dolayısıyla θ terimi için uygulanan olabilirlik oran testi önemli olduğu için Diğer modeller yerine zayıflık modeli tercih edilmelidir. θ için yapılan olabilirlik oran testi yokluk hipotezi altında genellikle 1 serbestlik dereceli χ^2 dağılımı göstermez. Fakat "Chibar2(01)" şeklinde ifade edilen 0 ve 1 serbestlik dereceli Ki-Kare dağılımlarının 50:50 karışımı olan (0 ve 1 serbestlik dereceli Ki-Kare arasında) bir dağılım göstermektedir.

Kurulan zayıflık modelinde zayıflık terimi (Z) her bir birey için bulundu ve $Z > 1$ ise; araştırılan KOAH hastalarının bağımsız değişkenlerce açıklanamayan sebeplerden ötürü daha kırılğan olduğu ve artan başarısızlık riskine sahip olduğu kararı verildi. $Z < 1$ ise; araştırılan KOAH hastalarının bağımsız değişkenlerce açıklanamayan sebeplerden ötürü azalan başarısızlık riskine sahip olduğu ve daha uzun yaşama eğilimine sahip olduğu ifade edildi.

Zayıflık teriminin (Z) karekökünün üstel ifadesinin ($\exp(\sqrt{\theta})$) zayıflığın derecesini gösterdiği düşünülerek istatistiki yorumlar yapılmıştır (Pankratz ve ark., 2005).

3. BULGULAR

Çizelge 3.1 Hastaların Demografik ve Klinik Özellikleri

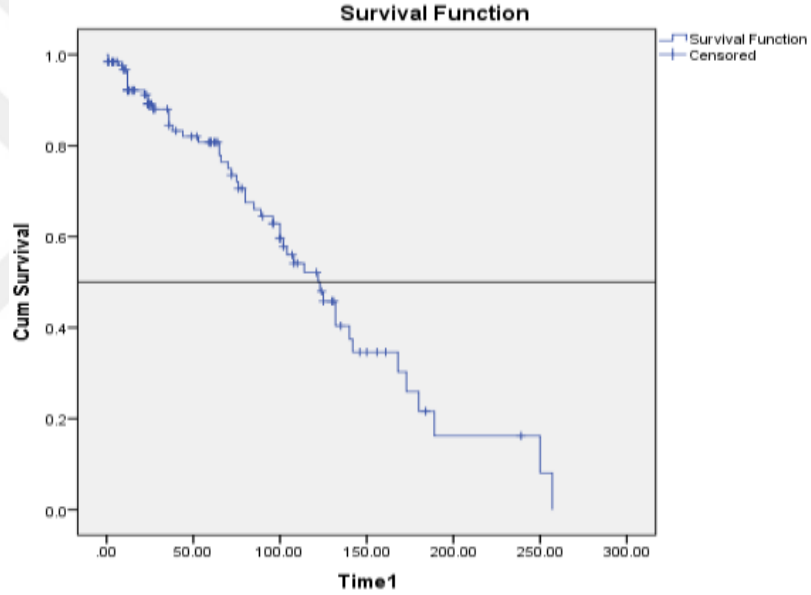
Yaş	71±9,14
Erkek/Kadın	100/25
BMI	26,40±6,87
Sigara İçiyor	% 80 (n=100)
Sigara Bilinmiyor	% 8,8 (n=11)
Biomass	% 5,6 (n=7)
Sigara Maruziyet	% 5,6 (n=7)
Sigara-Paket/yıl	51±36
Atak Sayısı	3±2,62
PaO2 mm Hg	53,21±11,11
PaCO2 mm Hg	45,55±37,57
SaO2 mm Hg	86,80±8,45
FEV1 %	54,87±20,41
FEV1 /FVC %	60,70±14,06
MMF %	43,23±32,65
Ek hastalık +/-	52/73
EkoPAB	42,69±17,67

Çizelge 3.1 de hastaların demografik ve klinik özelliklerinin ortalaması, standart sapması ve % değerleri verilmiştir. Analize alınan hastaların 25'i kadın, 100'ü ise erkek idi. Araştırılan 125 hastanın 52 tanesinde başka hastalıklar (Diabet, Kardiyovasküler hastalıklar,...) tespit edilmiştir. Yine araştırılan hastaların %80'i sigara içmekte iken; %8,8'i bilinmiyor, %5,6'sı biomass ve %5,6'si sigaraya maruz kalmıştır. Hastaların dispne düzeylerine bakıldığında ise; hastaların %18'si düz yolda hızlı yürürken veya yokuş çıkarken nefesinin daraldığını ifade etmişler, %35'si nefes darlığı nedeniyle düz yolda yaşlılarından yavaş yürüdüğünü veya düz yolda yürürken durup dinlenmeleri gerektiğini, %27'i düz yolda 100 m veya birkaç dakika yürüdüklerinde durup dinlenmeleri gerektiğini, %20'si nefes darlığı nedeniyle evden çıkamadıklarını veya giyinirken bile nefeslerinin daraldıklarını ifade etmişlerdir. Hastaların taraması sonucunda balgam kültürü pozitif olan 21 hastada üreyen bakterilerin %52 si Boğaz florası, %14'ü Kandida, %14'ü Haemophilus influenza, %5'i Streophanemous malphinio, %5'i Moraxella catarrhalis, %5'i

Streptococcus pneumoniae, %5'i ise Negatif basildi. Hastaların atak sayılarına bakıldığında en fazla atak sayısının 12 olduğu gözlenmiştir. Yaş ortalamasının ise $71 \pm 9,14$ olduğu, pulmoner arter basıncının ortalama düzeylerinin $42,69 \pm 17,67$ olduğu sonuçlarına varılmıştır.

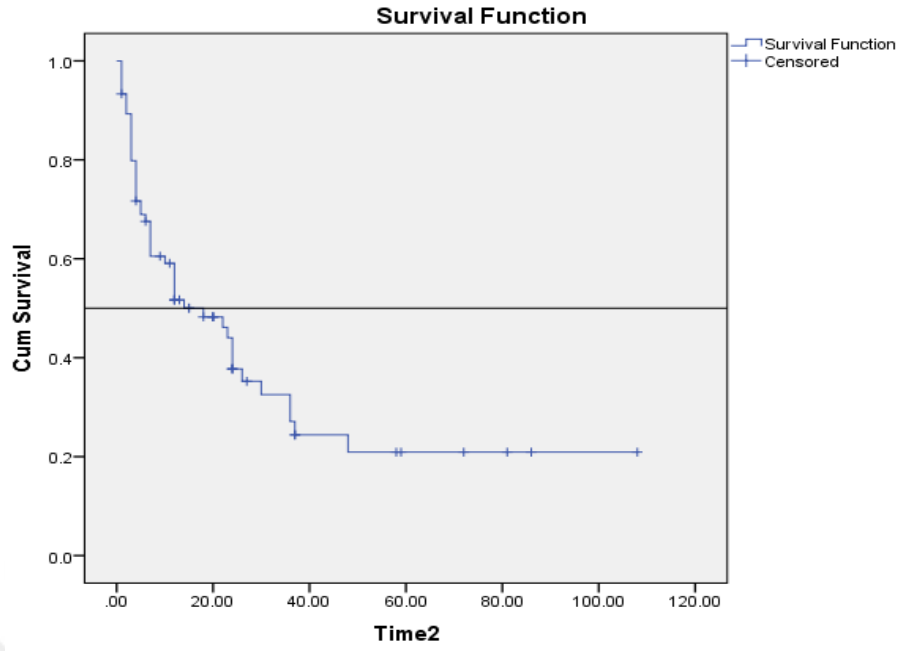
3.1. Yaşam Sürelerinin Kaplan-Meier Analizi ile Kestirimi

KOAH hastalarının ataksız yaşam süreleri Kaplan-Meier yöntemi ile incelendi. KOAH hastalarının semptom başlangıç tarihinden itibaren ilk atağı geçirebilecekleri ortalanca süresi yaklaşık 123 ay (%95 güven aralığı;100,88-145,11) olarak bulundu.



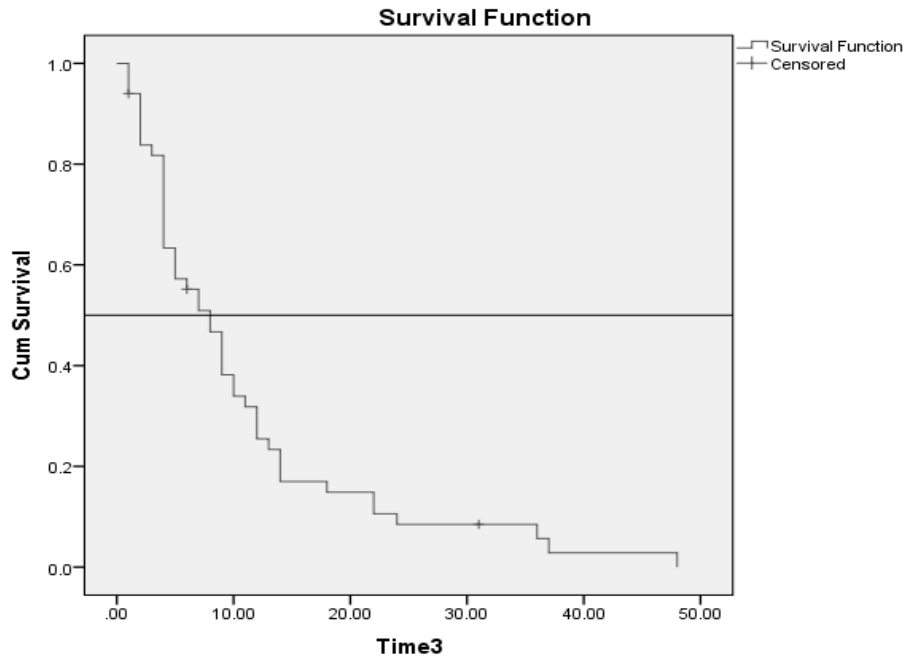
Şekil 3.1. KOAH Hastalarının Birinci Atak İçin Yaşam Eğrisi

KOAH hastalarının ilk atağı ile ikinci atağı arasında geçirebilecekleri ortalanca süresi yaklaşık 18 ay (%95 güven aralığı;7,82-28,17) olarak bulundu.



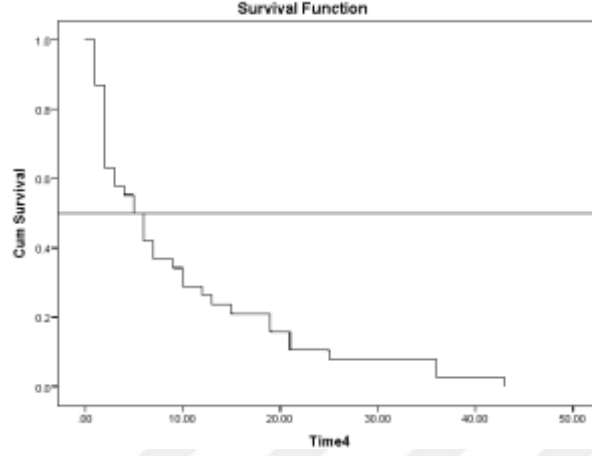
Şekil 3.2. KOAH Hastalarının İkinci Atak İçin Yaşam Eğrisi

KAOH hastalarının ikinci atağı ile üçüncü atağı arasında geçirebilecekleri ortalanca süresi yaklaşık 8 ay (%95 güven aralığı; 5,51-10,48) olarak bulundu.



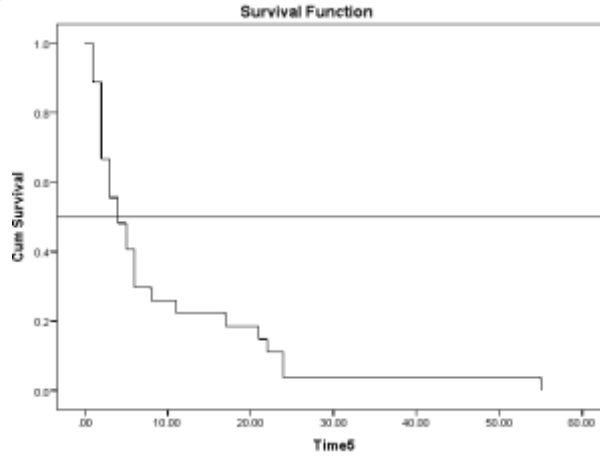
Şekil 3.3. KOAH Hastalarının Üçüncü Atak İçin Yaşam Eğrisi

KAOH hastalarının üçüncü atağı ile dördüncü atağı arasında geçirebilecekleri ortalama süresi yaklaşık 5 ay (%95 güven aralığı; 2,58-7,41) olarak bulundu.



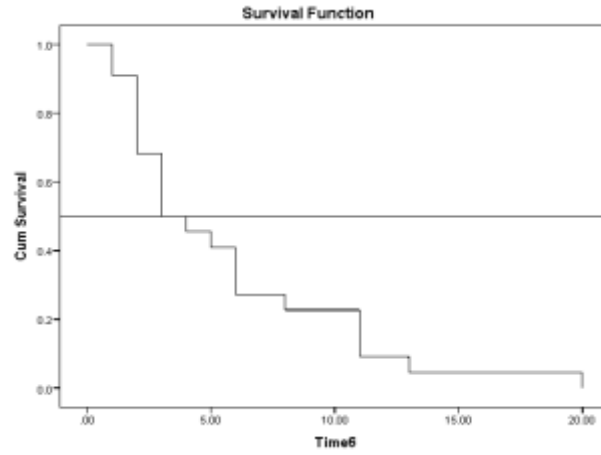
Şekil 3.4. KOAH Hastalarının Dördüncü Atak İçin Yaşam Eğrisi

KAOH hastalarının dördüncü atağı ile beşinci atakları arasında geçirebilecekleri ortalama süresi yaklaşık 4 ay (%95 güven aralığı; 1,45-6,54) olarak bulundu.



Şekil 3.5. KOAH Hastalarının Beşinci Atak İçin Yaşam Eğrisi

KAOH hastalarının beşinci atağı ile altıncı atağı arasında geçirebilecekleri ortalama süresi yaklaşık 3 ay (%95 güven aralığı; 0,7-5,2) olarak bulundu.

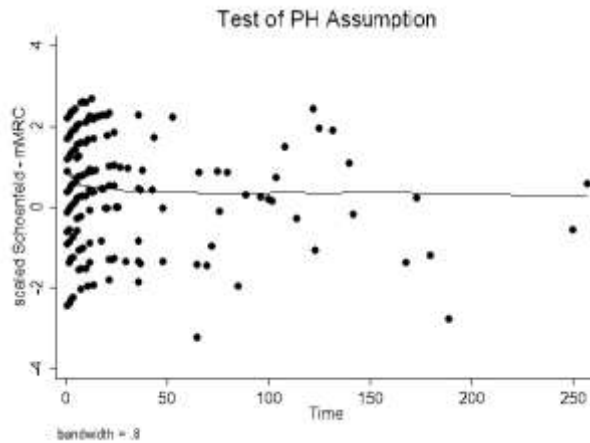


Şekil 3.6. KOAH Hastalarının Altıncı Atak İçin Yaşam Eğrisi

Altıncı ataktan sonra ise mesafeler artık iyice azalarak ortalama 1-2 ay olarak sıklaştığı gözlemlendi.

3.2. Oransal Hazard Varsayımı Denetimi

Orantılı hazard varsayımının sağlanıp sağlanmadığı grafik yöntemi ve Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki korelasyon testi ile araştırıldı. Orantılı hazard varsayımının dispne Modified Medical Research Council scale (MMRC) değişkeni için sağlanmadığı sonucuna ulaştık ($p=0,004$). Şekil 3.7’de mMRC değişkeninin orantılı hazard varsayımı için grafiği gösterildi.



Şekil 3.7. mMRC Değişkeninin Schoenfeld Grafiği

3.3. Tekrarlanan Başarısızlıkta Kullanılacak Analiz Yönteminin Seçimi

KOAH hastalarının verilerinde orantılı hazard varsayımı sağlanamamıştır. Bu durumda bu verilerin analizinde hangi veri analizi kullanılacağına karar verirken bu durumu dikkate almak gerekliydi. Zaten KOAH hastalarından elde edilen veri tekrarlayan başarısızlıklar içerdiği için bu yaşam verilerinin modellenmesinde Cox regresyon analizi kullanılmayacağı gereç ve yöntemler bölümünde ifade edilmişti. Tekrarlayan yaşam analizi verilerinde kullanabilen çözümler sayma süreci, tabakalandırılmış Cox yaklaşımları ve zayıflık modelleri idi. Bu modellerden sayma süreci orantısız hazardlar varlığında kullanılmadığından dolayı zayıflık modeli ve tabakalandırılmış Cox yaklaşımı yöntemi ile de analiz yapıldı.

3.4. Zayıflık Modelinin Seçimi

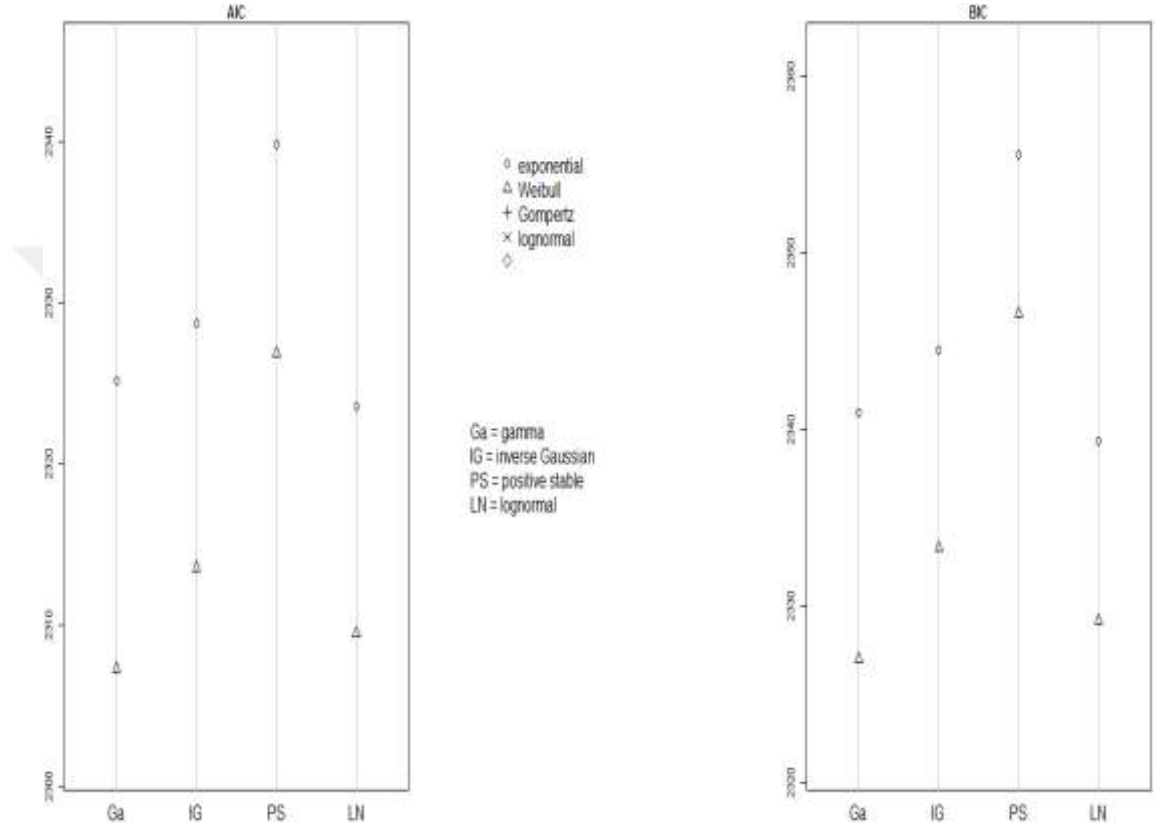
Burada karar vermemiz gereken bir başka nokta ise hangi zayıflık modelinin kullanılacağı idi. Zayıflık modelleri yine çalışmamızın giriş bölümünde ifade edildiği gibi paylaşılmış, paylaşılmamış ve ilişki olarak üç bölümde incelenmektedir. Bu modellerden veri yapımıza en uygun olan zayıflık modeli tekrarlı başarısızlıklarda da kullanılabilen paylaşılmış zayıflık modeli olarak belirlendi.

3.5. Zayıflık Teriminde Kullanılacak Dağılımın Belirlenmesi

Öncelikle zayıflık modelinde hangi parametrik regresyon modelinin kullanılacağına karar vermek için KOAH hastalarının veri analizi için Weibull, Gompertz, üstel, log-lojistik, log-normal regresyon modelleri kullanıldı. Ancak Log-logistik, Log-normal ve Gompertz dağılımlarına yakınsama sağlanamadığından dolayı Weibull ve üstel dağılımı arasında seçim yapıldı. Model seçiminde ise AIC ve BIC kriter değerlerine ve grafik yardımıyla karara varıldı. Zayıflık terimi (Z) için hangi dağılımın kullanılacağına karar vermek için yine AIC ve BIC kriter değerlerine bakarak ve grafik yardımıyla karar verildi.

Çizelge 3.2. AIC ve BIC Değerlerinin Tabloda Gösterimi

Bilgi Kriteri	Temel Hazard Dağılımı	Zayıflık Dağılımı			
		Gamma Dağılımı	Inverse Gaussian	Pozitif Durağan	Lognormal
AIC	Exponential	2325,15	2328,74	2339,81	2323,57
	Weibull	2307,27	2313,56	2326,86	2309,46
BIC	Exponential	2340,92	2344,51	2355,58	2339,35
	Weibull	2326,98	2333,27	2346,58	2329,18



Şekil 3.8. AIC ve BIC Değerlerinin Grafik Üzerinde Karşılaştırılması

3.6. Gamma Zayıflık Terimi İçeren Weibull Dağılımı ile Analiz

Gerek tablodan gerek de grafik üzerinden anlaşılacağı üzere temel hazard dağılımı olarak Weibull zayıflık terimi için ise Gamma dağılımının kullanılmasına karar verilmiştir. Gamma zayıflık terimi içeren Weibull dağılımı ile kurulan modelin anlamlılığını test etmek için başka bir ifade ile zayıflık teriminin modele eklenip eklenmemesine karar vermek için olabilirlik oran testi kullanıldı ve kıkare değeri 78,53 ve $p < 0,001$ olarak bulunmuştur. Bunun sonucunda zayıflık teriminin modelde anlamlı olduğuna ve modele eklenmesi gerektiği kararına varılmıştır. Zayıflık

dağılımının varyansının tahmin edicisi $\theta=1,58$ olarak bulunduğundan verilerin heterojen olduğunu ve KOAH hastalarında gözlenemeyen risk faktörlerinin olduğu kararına varıldı.



3.7. Zayıflık Modeli ile Elde Edilen Bulgular

Çizelge 3.3. Gamma Zayıflık Terimi İçeren Weibull Dağılımı ile Analiz Sonuçları

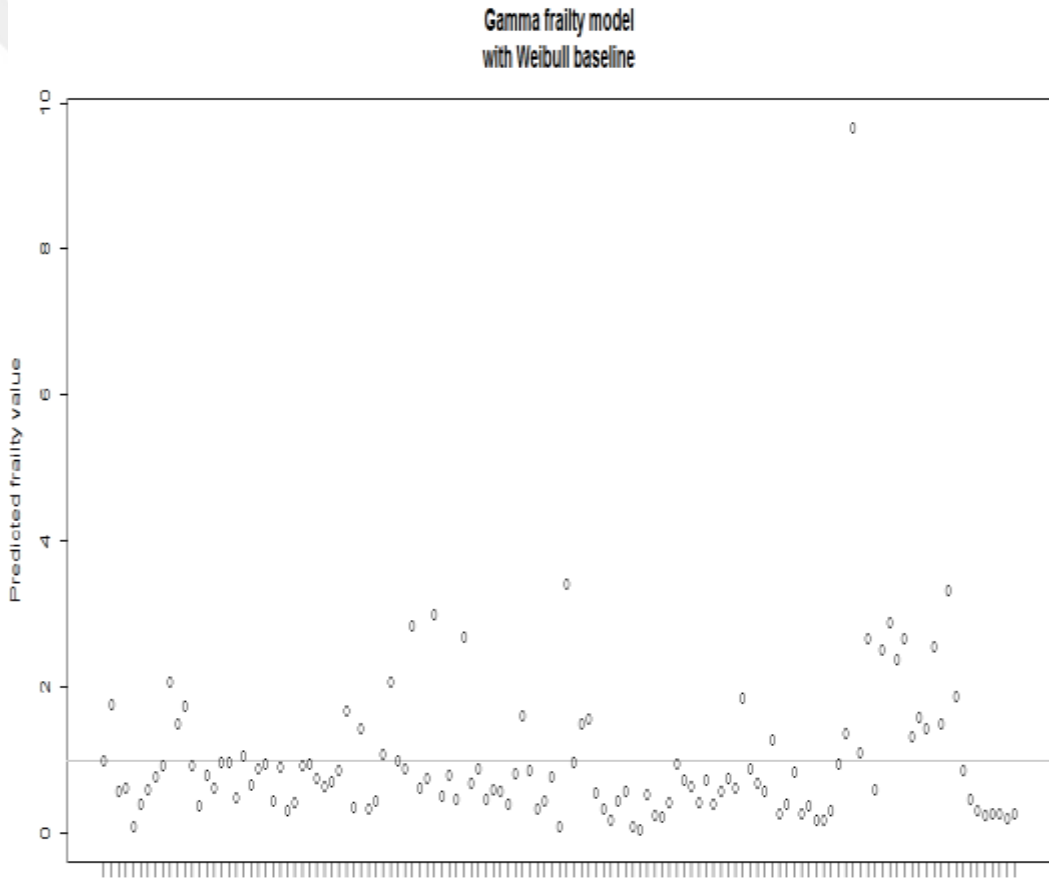
Değişken	Kategori	β	Exp (β)	Standart Hata	z değeri	p değeri	%95 Güven aralığı	
VKI	≥ 21	ref	ref	ref	ref	ref	ref	ref
	< 21	0,64	1,89	0,35	1,82	0,069	- 0,04	1,34
Ek Hastalık	yok	ref	ref	ref	ref	ref	ref	ref
	var	- 0,01	0,99	0,29	-0,06	0,950	-0,60	0,56
Fev	1	ref	ref	ref	ref	ref	ref	ref
	2	0,06	1,06	0,42	-0,15	0,870	-0,90	0,77
	3	1,15	3,15	0,42	2,73	0,006	0,32	1,98
	4	1,73	5,64	0,47	3,62	$< 0,0001$	0,79	2,67
mMRC	1	ref	ref	ref	ref	ref	ref	ref
	2	1,82	6,17	0,78	2,33	0,020	0,28	3,35
	3	2,46	11,70	0,77	3,19	0,001	0,95	3,98
	4	3,59	36,23	0,77	4,61	$< 0,0001$	2,06	5,11
θ		1,58				$< 0,0001$		

Çizelge 3.3’de verilen analiz sonuçlarında FEV₁, mMRC düzeyleri ve VKI değerlerinin kesim noktaları GOLD klavuzuna göre düzenlendi. FEV₁ düzeyi 4 olan KOAH hastaların FEV₁ düzeyi 1 olan hastalara göre yaklaşık 5 kat (exp (1,73)) daha riskli olduğunu, FEV₁ düzeyi 3 olan KOAH hastaların FEV₁ düzeyi 1 olan hastalara göre yaklaşık 3 kat (exp (1.15)), mMRC düzeyi 4 olan KOAH hastaların mMRC düzeyi 1 olan hastalara göre yaklaşık 36 kat (exp (3,59)), ve mMRC düzeyi 3 olan KOAH hastaların mMRC düzeyi 1 olan hastalara göre yaklaşık 11 kat (exp (2,46)), mMRC düzeyi 2 olan KOAH hastaların mMRC düzeyi 1 olan hastalara göre yaklaşık 6 kat (exp (1,82)) daha riskli olduğu saptanmıştır.

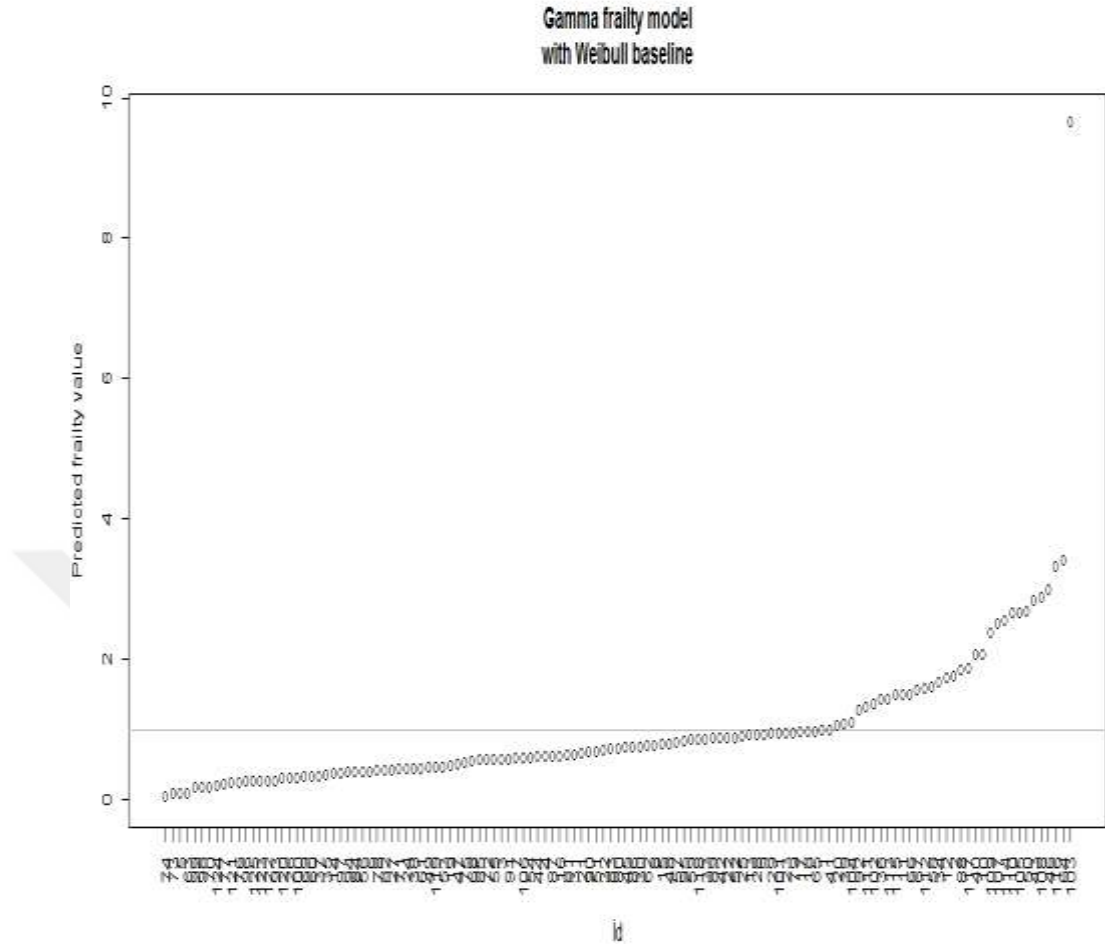
KOAH hastalarının başarısızlık sürelerini etkileyen mMRC ve FEV₁ faktörlerinin yanı sıra çevresel ya da genetik gizli faktörlerin olabileceği söylenebilir. Bu gizli faktör ya da faktörlerin örneklemdaki ortalamada $\pm \sqrt{\theta}$ standart sapma gösteren herhangi bir KOAH hastasında (exp($\sqrt{\theta}$) = exp ($\sqrt{1,58}$)) 3,5 kat daha fazla ya da daha az risk taşıdığı söylenebilir.

3.8. Zayıflık Terimlerinin Bulunması

Kurulan Zayıflık modelinde zayıflık terimi (Z) her bir hasta için bulundu. Aşağıdaki şekilde görüldüğü gibi zayıflık terimlerinin daha çok 0 ile 1 arasında yoğunlaştığı gözlemlendi ve $Z > 1$ durumunda; KOAH hastalarının bağımsız değişkenlerce açıklanamayan sebeplerden ötürü daha kırılğan olduğu ve artan başarısızlık riskine sahip olduğu kararı verilirdi. $Z < 1$ durumunda; KOAH hastalarının bağımsız değişkenlerce açıklanamayan sebeplerden ötürü daha kırılğan olduğu ve azalan başarısızlık riskine sahip olduğu ve daha uzun yaşama eğilimine sahip olduğu ifade edildi.



Şekil 3.9. Gamma Zayıflık Terimi İçeren Weibull Dağılımının Zayıflık Terimlerinin Tahmini



Şekil 3.10. Gamma Zayıflık Terimi İçeren Weibull Dağılımının Zayıflık Terimlerinin Tahmininin Doğrusal Grafikle Gösterimi

3.9. Tabakalı Cox Yaklaşımı ve Zayıflık Modelinin Karşılaştırılması

Tabakalı Cox yaklaşımı ile analiz yapılırken veri girişi aşağıdaki çizelgede olduğu gibidir. Time point şeklinde verdiğimiz KOAH hastalarının geçirdikleri atak sayısıdır. Bu sütun, paket programlarda tabaka olarak adlandırılmaktadır.

Çizelge 3.4. Tabakalı Cox Modeli İçin Veri Girişi

ID	Durum	Zaman	Atak Sayısı
A Hastası	1	3	1
A Hastası	1	4	2
B Hastası	1	3	1
B Hastası	1	5	2
B Hastası	1	7	3
B Hastası	1	9	4
C Hastası	1	3	1
D Hastası	1	3	1
D Hastası	1	5	2

Çizelge 3.5. Tabakalı Cox Yaklaşımı ile Analiz Sonuçları

Değişken	HR	Standart Hata	Z	P	Güven Aralığı	
					Alt	Üst
mMRC	1,32	0,12	2,93	0,003	1,09	1,59
FEV	1,11	0,09	1,24	0,210	0,93	1,32

Çizelge 3.6. Zayıflık Modeli ile Analiz Sonuçları

Değişken	HR	Standart Hata	Z	P	Güven Aralığı	
					Alt	Üst
mMRC	2,31	0,30	6,41	<0,001	1,78	2,98
FEV	1,74	0,23	4,09	<0,001	1,33	2,28

Sonuçlardan görüldüğü gibi zayıflık modelinde mMRC değişkeninin p değeri ($p<0,001$) ve FEV değişkenin p değeri ($p<0,001$) önemli bulunurken tabakalı Cox yaklaşımında sadece mMRC değişkenin p değeri ($p=0,003$) önemli olarak bulunmuştur.

4. TARTIŞMA

Bu güne kadar dünyada yapılan çalışmaların çok azında tekrarlanan başarısızlıkların modellenmesi zayıflık modeli ile analiz edilmiştir. Zayıflık modeli ile ilgili yapılan çalışmaların birçoğu zayıflık modelinin kendi içinde dağılımlarına karar verme çalışmalarıdır.

Ata ve Karasoy (2011), yaptıkları çalışmada orantılı hazard varsayımı sağlanmadığında kullanabilecek olan yöntemleri incelemişlerdir. Yapılan çalışmada mide kanseri olan hastalara ait verilere parametrik regresyon modelleri ve zayıflık modeli uygulanmıştır. Parametrik regresyon modelleri için karşılaştırma kriterlerinden olan AIC, BIC ve $-2\log L$ değerlerine bakılarak en uygun modelin log-normal modeli olduğuna karar verilmiştir. Daha sonra zayıflık modeli ile analiz yapılarak zayıflık terimi 28,79 olarak bulunmuştur. Ata ve Karasoy bireyler arasındaki gözlenemeyen farklılıkları da modele dahil eden zayıflık modelini önermişlerdir.

Faradmal ve ark. (2012), 2000 ile 2005 yılları arasında İran Shiraz Üniversitesi Kanser araştırma merkezinde meme kanseri tanısı ile tedavi edilen 769 olgunun yaşam süresini etkileyen prognostik faktörleri araştırdıkları çalışmada, eğitim durumu, yaş, menopoz durumu, sigara öyküsü, tümör lokalizasyonu, tümör grade' i, tümör boyutu, lenf nodu sayısı, gibi olgulardan elde edilen parametrelerin sağkalım süreleri üzerine etkileri Cox regresyon ve zayıflık modeli karşılaştırılarak analiz etmişlerdir. Analiz R programının R version 2.15.2 'de yapılmıştır. Çalışma sonunda Cox regresyonda concordance index=0,79 zayıflık modelde ise concordance index=0,93 olarak bulunmuştur. Bu sonuç bu meme kanseri verilerinde zayıflık model kullanılmasını işaret etmektedir. Zayıflık modelde zayıflık terimi $\theta=1,92$ ve $p=0,005$ olarak bulunmuştur. Bu sonucun ise meme kanserli hastalarda yaşam süresini etkileyen prognostik faktörlerden başka gizli faktörlerin olduğunu ifade etmişlerdir.

Tutkun ve Yeğen (2016), da yaptıkları çalışmada İbni Sina Hastanesi Göğüs Cerrahisi bölümünde tedavi görmüş 236 akciğer kanseri hastalarının yaşam sürelerini

belirlemek için paylaşılmış ve paylaşılmamış zayıflık modellerini kullanmışlardır. Paylaşılmamış zayıflık modeli ve log-logistik hızlandırılmış başarısızlık modeli karşılaştırmalarında paylaşılmamış zayıflık modelinin bu akciğer kanseri verilerini yorumlamada daha iyi olduğunu öne sürmüşlerdir.

Ullah ve ark. (2014), Avustralya’da 2008’de oynanan ulusal Rugby Liginden elde edilen tekrarlayan yaralanma verilerinde Andersen-Gill (A-G), Zayıflık, Wei-Lin-Weissfeld total time (WLW-TT), Prentice-Williams-Peterson gap time (PWP-GT) modellerini karşılaştırmışlardır. Modelleri karşılaştırmak için klasik Cox regresyon modelinde kullanılan model seçim kriterlerinden Akaike bilgi kriteri (AIC), Bayesci bilgi kriteri (BIC) ve log likelihood (LL) kullanmışlardır. Bu analizin sonucunda bu tekrarlayan veri için Andersen-Gill ya da zayıflık modelin kullanılabilceğini önermişlerdir.

Grover ve Seth (2014), yaptıkları çalışmada Hindistan Yeni Delhi’de PushpawatiSinghania araştırma enstitüsü Gastroenteroloji bölümünden 2007 ile 2010 yılları arasında karaciğer sirozu tedavisi gören 811 hastanın prognostik faktörlerini zayıflık modelle belirlemeye çalışmışlardır. 811 hastanın veri analizinde paylaşılmış ve paylaşılmamış zayıflık modelinin çözümünü yapmışlardır. Paylaşılmamış zayıflık modelin analizinde zayıflık modeli içeren Weibull modeli ile çözümde zayıflık terimi $\theta=5,19$ ve Log-logistik modeli ile çözümde zayıflık terimi $\theta=2,07$ bulunmuştur. Paylaşılmış zayıflık modelin analizinde zayıflık modeli içeren Weibull modeli ile çözümde zayıflık terimi $\theta=0,68$ Log-logistik modeli ile çözümde zayıflık terimi $\theta=0,76$ ve paylaşılmış zayıflık Cox regresyon ile çözümünde $\theta=0,66$ olarak bulunmuştur.

Adham ve AlAhmadi (2016), yaptıkları çalışmada gamma zayıflık ve ters Gaussian zayıflık modelini karşılaştırmışlardır. Gamma zayıflık modelinin zayıflık terimi $\theta=0,307$, ters Gaussian zayıflık modelinin zayıflık terimi $\theta=0,347$ olarak bulunmuştur. Gamma zayıflık ve ters Gaussian modellerinin p değerleri sırasıyla 0,005 ve 0,004 anlamlı olarak bulunmuştur. Bu değerlerin anlamı ise zayıflık teriminin her iki modelde etkisi olduğu ve modele dahil edilmesi gerekmektedir. Bu modeller arasında seçim yapmak için ise Akaike bilgi kriteri (AIC), Bayesci bilgi kriteri (BIC) kullanılarak gamma zayıflık modelinin kullanılmasına karar verilmiştir.

Sari ve Purnami (2015), yaptıkları çalışmada servikal kanser hastalarından aldıkları nüks eden yaşam verilerinin analizinde Andersen-Gill yaklaşımını kullanmışlardır. Tekrarlayan yaşam analizi modellerinde klasik bilindik Cox regresyon modeli yerine Andersen-Gill modelinin kullanılması gerektiğini vurgulamışlardır.

Stanley ve ark. (2016), Blantyre Malawi'deki Queen Elizabeth merkez hastanesinden aldıkları tedavi ve kontrol grubuna ayrılmış 16 yaşından küçük çocuk hastalardan elde edilen verilerin analizinde öncelikle klinik denemelerde sıklıkla kullanılan Cox regresyon analizi yapmışlardır.. Daha sonra veriler Weibull, üstel, ve paylaşılmış Gamma zayıflık modelleri ile analiz yapılmıştır. Bu modellerin uygunluğu Akaike bilgi kriteri (AIC) ile araştırılmış, Cox, üstel, weibull ve paylaşılmış Gamma zayıflık modellerinin sırasıyla AIC kriterleri 252,4; 174,7; 176,7 ve 175,1 olarak bulunmuştur. Dolayısıyla bu veriler üzerinde üstel modelin kullanılmasının uygun olduğuna karar verilmiştir.

Govindarajulu ve ark. (2011), çalışmalarında yaşam verisi analizinde kullanılabilir olan modellerden birisi olan zayıflık modelin klinik ve genetik çalışmalarda kullanılabilirliğini açıklamışlardır. Zayıflık modelin birimler arasındaki heterojenliğin ortaya çıkarılması ve gruplar arası heterojenliği modellemek için özellik göstermesi durumunda kullanıldığını açıklamışlardır. Zayıflık modelinin veri üzerindeki çözümünü R paket programında anlatabilmek için iki veri seti kullanmışlardır. Bu veri setlerinden birincisi Guatemala'daki çocuk ölümlerinin yaşam süresini modellemek için yapılan bir anketten elde edilen 3120 çocuğun verisi diğeri ise ABD'de yapılan birinci derece akraba evliliği, ikinci derece akraba evliliği ve üçüncü derece akraba evliliklerinde doğan 5205 çocuk ölümlerini içermektedir. Elde edilen veriler Cox regresyon ve zayıflık modelle çözümlenmiştir. Sonuçta kovaryetler tarafından ortaya çıkartılamayan, açıklanamayan heterojenliğin zayıflık modeli ortaya çıkarılabileceğini savunmuşlardır.

Huang ve Wang (2004), çalışmalarında $N(t)$ sayma sürecini homojen olmayan Poisson süreciyle modellemiş, tekrarlama süreci ile durdurma süreci arasındaki bağımlılık yapısını paylaşılmış zayıflık terimi kullanarak açıklamışlardır. Huang ve Wang modellerinde birikimli yoğunluk fonksiyonunun tahminini, Kaplan-Meier

tahmin yöntemini kullanarak bulmuşlardır. Daha sonra birikimli yoğunluk fonksiyonundan yararlanarak diğer parametre tahminlerini elde etmişlerdir.

Huang ve ark. (2010), tekrarlanan başarısızlıkların olduğu verinin analizinde, başarısızlık zamanı ile durdurma zamanı arasındaki ilişkiyi zayıflık terimi Z ile modellemişlerdir. Modellerinde zayıflık teriminin dağılımına ilişkin bir varsayımda bulunmamışlardır. Bu çalışmada Huang ve arkadaşları, model parametrelerini tekrarlama zamanları ve zayıflık terimi için herhangi bir dağılım varsayımında bulunmadan yarı parametrik yöntemler kullanarak bulmuşlardır.

Liu ve arkadaşları (2004), tekrarlanan başarısızlık verisini bilgi içeren durdurma varsayımı altında Poisson süreçlerini kullanarak modellemişlerdir. Poisson süreçleri, olayların zaman boyunca rastgele ortaya çıktığı ve birbirleriyle ara kesiti olmayan zaman aralıklarında ortaya çıkan olay sayılarının birbirinden bağımsız olduğu süreçlerdir. Bu çalışmada Liu ve arkadaşları ölüm olayını, bilgi içeren durdurma olarak ele almışlar; durdurma olayı ile tekrarlanan olaylara ait olabilirlik fonksiyonunu paylaşılmış zayıflık modellerinden yararlanarak elde etmişlerdir. Liu ve arkadaşları, bilgi içeren durdurma olayını, ölüm olarak ele almışlardır. Tekrarlanan olayda, ölüm ve durdurma süreçleri sürekli dağılıma sahip olduğunu söylemişlerdir.

Taheri ve ark. (2016), İran Tahran Razi psikiyatri hastanesinden 1993 ve 2011 yılları arasında bipolar efektif bozukluk nedeni ile en az bir kez depresyon atağı ile hastaneye yatan 526 hastadan retrospektif olarak elde edilen veriler üzerinde yapılan analiz R paket programının R 3.2.0 versiyonu üzerinde zayıflık modeli ile yapılmıştır. Elde edilen tekrarlayan veriler üzerinde yapılan zayıflık modeli sonucunda medeni durum ve madde bağımlılık gibi kovaryet değişkenlerinin, depresyon atağı süresini etkilediğini bulmuşlardır. Taheri ve arkadaşları zayıflık varyansı olarak ifade ettikleri zayıflık terimini ($\theta=0,28$) ve $p<0,0001$ önemli olarak bulmuşlardır. Bu değerlerin anlamını ise bireyler arasındaki heterojenliğin varlığı ve depresyon ataklarını etkileyen kovaryetlerin yanı sıra başka gizli faktörlerin olabileceği şeklinde açıklamışlardır.

Keiding ve ark. (1997), Danimarka Odensa'dan aldıkları mide tümörü olan 205 hastanın ameliyat edildikten sonraki yaşam sürelerini incelemişler ve yaşam süresini etkileyen prognostik faktörlerin etkilerini araştırmışlardır. Araştırmada 205 hastadan 57'sinde başarısızlık gözlenirken 14 hasta başka sebeplerden başarısız olmuşlardır. Bu hastalardan 134'ü izleme süresinin sonunda hâlen hayatta olduğu için sansürlü olarak alınmıştır. Araştırma Cox regresyon analizi ile yapıldığında hazardların oransallık varsayımını sağlamadığı gözlenmiştir. Bu sebeple Zayıflık modeli kurularak analiz yapılmış Gamma zayıflık ve Inverse Gaussian zayıflık modelleri karşılaştırılmış ve heterojenlik anlamlı bulunmuştur.

Gutierrez (2002), yaptığı çalışmada; paylaşılmış zayıflık modelini incelemiştir. Bu çalışmada diyaliz makinesi ile evde tedavi gören 38 böbrek hastasında yapılan araştırmada hastalarda kata ter yerleştirilen yerde enfeksiyon oluşumun da etkili olan risk faktörleri incelenmiştir. Bu araştırmada hastalarda tekrarlayan enfeksiyonlardan en fazla ikisi analize alınmıştır. Başarısızlık enfeksiyon oluşumu olarak alınırken, başarısızlık süresi olarak iki enfeksiyon oluşumu arasında geçen süre kaydedilmiştir. Eğer enfeksiyon nedeni dışında kat eter çıkarılmışsa veri sansürlü olarak alınmıştır. Takip süresinde ikinci enfeksiyon gelişmemişse veri yine sansürlü olarak kabul edilmiştir. Böbrek hastalarından elde edilen veri tekrarlayan başarısızlıkların olduğu bir veri kümesi halini almıştır. Bu veriyi analiz etmek için ise paylaşılmış zayıflık modeli analizi kullanılmış ve analizler Stata paket programında yapılmıştır.

Liu ve ark. (2004), yaptıkları araştırmada böbrek nakli yapılan hastalardan elde edilen veriler SRTR veri tabanından alınmıştır. 1999 ile 2003 yılları arasındaki kayıtlardan alınan veriler üzerinde paylaşılmış Zayıflık modeli kullanılmıştır.

Ünlü (2013), yaptığı tez çalışmasında, tekrarlanan olay süreci ile ölüm süreci arasında güçlü bir ilişkinin olduğu düşünülerek bu ilişkinin yapısını da modelde dikkate alan iki yeni model önermiştir. Önerilen her iki model de Cox orantılı hazard modelinin, tekrarlanan olay süreci için genişletilmiş biçimidir. Bilgi içeren durdurma varlığında, tekrarlanan olay süreci, homojen Poisson süreci ile modellenmiştir. Tez çalışmasında ölüm olayı bilindiğinde tekrarlanan olay süreci ile durdurma varlığında tekrarlanan olay süreci iki farklı süreç olduğu için ölüm olayına ve bu iki farklı sürece ilişkin üç farklı yoğunluk fonksiyonu önerilmiştir. Önerilen bu yoğunluk

fonksiyonları ortak bir zayıflık terimi ile birbirlerine bağlanmıştır. Zayıflık terimleri gözlenemediği için parametrelerin en çok olabilirlik tahmininin elde edilmesinde EM (Beklenti en büyükleme) algoritmasından yararlanılmıştır. Tez çalışmasında önerilen her iki model için gerçek bir veri kümesi kullanılarak bilgi içeren durdurma varlığında tekrarlı olay sürecine ilişkin parametre tahminleri elde edilmiş, daha sonra da farklı senaryolar için benzetim çalışması yapılmıştır.

Literatür incelendiğinde yapılan çalışmaların birçoğunda zayıflık modelinin kendi içinde dağılımlarına karar verme yani zayıflık teriminin dağılımının seçimi yapılmıştır. Gamma ve ters-Gauss dağılımları çalışmalarda zayıflık terimi için en sık tercih edilen dağılımlardır. Bu tez çalışmasında ise; hem zayıflık teriminin dağılımına karar verme hem de zayıflık modelinin tekrarlanan başarısızlık içeren heterojen veride kullanımı modellenmiştir. KOAH hastalarından elde edilen tekrarlanan başarısızlıkların modellenmesi Gamma zayıflık terimi içeren Weibull modeli ile yapılmıştır. Öncelikle zayıflık modelinde hangi parametrik regresyon modelinin kullanılacağına karar vermek için KOAH hastalarının veri analizi için Weibull, Gompertz, üstel, log-lojistik, log-normal regresyon modelleri kullanılmıştır. Fakat Log-logistik, Log-normal ve Gompertz dağılımlarına yakınsama sağlanamadığından dolayı Weibull ve üstel dağılımı arasında seçim yapılmıştır. Zayıflık terimi dağılımına karar verildikten sonra model kurulup, verinin heterojenliği ve modelin anlamlılığı kontrol edilerek zayıflık modelinin kullanılmasına karar verilmiştir.

5. SONUÇ VE ÖNERİLER

Yaşam analizinde modellemeler tekrarlanan başarısızlık zamanlarında kullanılabilir şekilde de geliştirilmiştir. Zayıflık modeli bu modellerden birisidir. Yaşam analizinde ölçülen açıklayıcı değişkenler dışında başarısızlık süresini etkileyen ancak gözlenemeyen gizli risk faktörleri vardır. İşte bu heterojenliği açıklamak için ölçülemeyen rastgele etkide yaşam modellerine eklenerek modelleme yapılmaktadır.

Bu araştırmada yaşam analizinin temel kavramları açıklanmıştır. Yaşam analizini diğer yöntemlerden ayıran sansürlü veriler ve çeşitleri hakkında bilgi verilmiştir. Yaşam analizinde geleneksel olarak kullanılan Cox orantılı hazard modelinin genel yapısı ve orantılılık varsayımının incelenmesinde kullanılan yaklaşımlar açıklanmıştır. Yaşam analizinde kullanılan parametrik dağılımlar incelenerek, parametre tahmin yöntemleri hakkında bilgi verilmiştir. Yaşam analizinde tekrarlanan başarısızlıkların analizinde kullanılan yöntemler ve farklar anlatılmıştır. Bireyler tekrarlanan başarısızlıkları birden fazla kez yaşadığında elde edilen tekrarlanan başarısızlıklar arasındaki ilişki göz önüne alınmadan analiz edildiğinde probleme neden olduğundan bu konunun önemi vurgulanmıştır. Tekrarlanan başarısızlıkların modellenmesinde kullanılan zayıflık modeli ayrıntılarıyla incelenmiştir.

Uygulama kısmında ise KOAH hastalarından elde edilen verilerde önce Cox orantısız hazard varsayımları incelenmiş ardından tekrarlanan başarısızlıkları analiz için Cox temelli tabakalı Cox yaklaşımları ve zayıflık modelleri kullanılmış ve farklı sonuçlar verdiği görülmüştür. Öncelikle KOAH veri kümesi için parametrik regresyon modelleri içinde en uygun modelin Weibull regresyon modeli olduğu AIC ve BIC kriterleri ve grafik yardımıyla bulunmuştur. En uygun parametrik regresyon modeline karar verdikten sonra zayıflık terimi için en uygun dağılımın Gamma dağılımı olduğu yine AIC, BIC ve grafik yardımıyla karara bağlanmıştır. Zayıflık teriminin modele alınıp alınmamasına karar vermek için olabilirlik oran testi kullanılmış ve kıkare değeri 78,53 ve p-değeri ($p < 0,001$) olarak bulunmuştur. Bu

değerler bize zayıflık teriminin model üzerinde anlamlı olduğuna ve modele alınması gerektiğine ifade etmektedir. Gamma zayıflık terimi içeren Weibull modeli ile yapılan analiz sonucunda KOAH hastalarının yaşam süresini etkileyen mMRC ve FEV₁ gibi faktörlerin yanı sıra çevresel ya da genetik gizli risk faktörlerin olabileceği sonucuna varılmıştır. Genetik ya da gözlenemeyen faktörler olabilen bu heterojenlik monoton artan ya da monoton azalan olabilir. Bu gizli faktör ya da faktörlerin örnekleme ortalamada $\pm \sqrt{\theta}$ standart sapma gösteren herhangi bir KOAH hastasında ($\exp(\sqrt{\theta}) = \exp(\sqrt{1,58})$) 3,5 kat daha fazla ya da daha az risk taşıdığı sonucuna varılmıştır. Böylece hastalar arasındaki heterojenliği açıklamak için ölçülemeyen rastgele etki modele dahil edilmiştir.

Tekrarlanan başarısızlıkların olduğu KOAH hastalarından elde ettiğimiz veriyi modellerken zayıflık modelinin kullanılması hem heterojenliği açıklamış hem de yorumlarımıza farklı bakış açısı getirmiştir.

Zayıflık modelleri yaşam süreleri ile ilgili modelleme çalışmalarında, tekrarlayan başarısızlıkların modellenmesinde ve bireyler arasındaki heterojenliğin açıklanmasında kullanılabilir. Bu heterojenliğin açıklanması yaşam analizinin modellenmesi sonucunda elde edilen yorumları daha etkin kılmaktadır.

O halde orantısız hazardlar olduğunda ya da tekrarlanan başarısızlık durumlarında veya bireyler arasındaki heterojenliği açıklamak için zayıflık modeli kullanılabilir.

ÖZET

Takip Arařtırmalarında Tekrarlanan Başarısızlıkların Modellenmesi: Göğüs Hastalıkları Veri Seti Üzerinde Bir Uygulama

Bu çalışmanın amacı yaşam analizlerinde karşılaşılan tekrarlanan başarısızlık verilerini en iyi şekilde açıklayan modelleri belirlemektir. Yaşam analizinde modellemeler tekrarlanan başarısızlıkları da modelleyebilecek biçimde de geliştirilmiştir. Bu modellerden birisi zayıflık modelidir. Zayıflık modeli Cox orantısal hazard modelin bir uzantısıdır. Bu çalışmada hem tekrarlayan başarısızlıkların modellenmesi hem de gizli risk faktörlerinin sebep olduğu heterojenliği açıklamak amacıyla yaşam analizinde kullanılan zayıflık modeli tüm ayrıntılarıyla açıklanmıştır. Genetik ya da gözlenemeyen faktörlerden kaynaklanan heterojenlik ölçülemeyen rastgele etkinin hazard modeline eklenmesiyle modellenmiştir.

Bu amaçla yapılan çalışmada KOAH hastalarının tekrarlayan ataklarından elde edilen veriler, tekrarlanan başarısızlıklarda kullanılan tabakalı Cox yaklaşımları ve zayıflık modeli ile analiz edilmiş ve sonuç olarak zayıflık modelinin veri üzerinde anlamlı sonuçlar verdiği gözlemlenmiştir.

Sonuç olarak zayıflık modelleri yaşam analizinde tekrarlayan başarısızlıkların modellenmesinde ve bireyler arasındaki heterojenliğin açıklanmasında kullanılabilir. Bu heterojenliğin açıklanması yaşam analizinin modellenmesi sonucunda elde edilen yorumları daha etkin kılmaktadır.

Anahtar Sözcükler: Göğüs Hastalıkları, Orantısal Hazard, Takip Çalışması, Zayıflık Modeli.

SUMMARY

Modeling of Unsuccessful Repeated in Follow-up Research: Application on a Data Set About Chest Diseases

The purpose of this study is to determine the models expressing ideally the recurrent events data encountered in the survival analyses. Modeling in the survival analysis has been developed in the manner that it can model the recurrent events. One of these models is the frailty model. The frailty model is an extension of Cox proportional hazard model. In this study, the frailty model, which is used in the survival analysis to explain the heterogeneity that both modeling of recurrent events and latent risk factors cause, has been explained in detail. The heterogeneity stemming from genetics or unmeasured covariates has been modeled with adduct of unmeasured random effect to the hazard model.

In the study for this purpose, data acquired from COPD patients' recurrent attacks have been analyzed with stratified Cox approach used in the recurrent events and frailty model, and consequently it has been observed that the frailty model has brought to a successful conclusion on data.

In conclusion, the frailty models can be used in the modeling of recurrent events in the survival analysis, and in the explanation of heterogeneity among subjects. The explanation of this heterogeneity has dynamized the interpretation made in consequence of modeling of the survival analysis.

Keywords: Chest Diseases, Follow-up Studies, Frailty Model, Proportional Hazard Model.

KAYNAKLAR

- AALEN OO (1975). *Statistical Inference for a Family of Counting Processes*, PhD Thesis, University of California, Berkeley, CA.
- AALEN OO, TRETTLI S (1999). Analyzing incidence of testis cancer by means of a frailty model. *Cancer Causes Control*, **10**: 285-292.
- ABERNETY RB (2002). *The New Weibull Handbook: Reliability & Statistical Analysis for Predicting Life, Safety, Survivability, Risk, Cost and Warranty Claims*. (4th Ed.), (Chapter 1) p.: 1-12.
- ABBRING JH, VAN DEN BERG GJ (2007). The Unobserved heterogeneity distribution in duration analysis. *Biometrika*, **94**: 87-99.
- AKDENİZ F, ERDUGAN F (2014). *Aktüerya Matematiği ve Yaşam Analizi*. Özyurt Matbaacılık, (1. Baskı), Akademisyen Yayınevi.
- ALLISON PD (2010). *Survival Analysis Using SAS: a Practical Guide*, SAS Institute INC, (2nd Ed.), North Carolina.
- ANDERSEN PK, BORGAN O, GILL RD, KEIDING N (1993). *Statistical Models Based on Counting Processes*. Springer- Verlag, NewYork.
- ATA N, KARASOY D (2011). Sağlıkım çözümlemesi için zayıflık modeli ve mide kanseri hastalarına ilişkin verilerle bir uygulama. *Çankaya University Journal of Science and Engineering*, **8(2)**: 225-235.
- BALAKRISHNAN N, PENG Y (2006). Generalized gamma frailty model. *Stat. Med*, **25**: 2797-2816.
- BAŞAR E (2016). Yaşam Sürdürme Analizinde Gamma Kırılabilirlik Modelleri. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, DOI: 10.19113/sdufbed.67151.
- BOAG JW (1949). Maximum likelihood estimates of proportion of patients cured by cancer therapy. *Journal of the American Statistical Association*, **53**: 151-160.
- BRESLOW N (1970). A Generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, **57**: 579-594.
- CHANG SH, WANG MC (1999). Conditional Regression Analysis for Recurrence Time Data. *Journal of the American Statistical Association*, **94**: 1221-1230.
- CHEN LM, IBRAHIM JG, CHU H (2014). Sample size determination in shared frailty models for multivariate time-to-event data. *J Biopharm Stat*, **24**: 908-923.

- CLAYTON D (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**: 141-151.
- CLAYTON DG, CUZICK J (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical. Society(A)*, **148**: 82-117.
- COLLETT D (2003). *Modelling Survival Data in Medical Research*, (2nd Ed.), Chapman and Hall, New York.
- COMMENGES D, ANDERSEN PK (1995). Score test of homogeneity for survival data. *Lifetime Data Analysis*, **1**: 145-156.
- COOK RJ, LAWLESS J (2007). *The Statistical Analysis of Recurrent Events*. Springer-Verlag New York.
- COX DR (1972). Regression models and life-tables, *Journal of the Royal Statistical Society, Series B*, **34**:187-220.
- DOS SANTOS DM, DAVIES RB, FRANCIS BJ (1995). Non-parametric hazard versus non-parametric frailty distribution in modelling recurrence of breast cancer. *Journal of Statistical Planning and Inference*, **47**: 111-127.
- DUCHATEAU L, JANSSEN P, KEZIC I, FORTPIED C (2003). Evolution of recurrent asthma event rate over time in frailty models. *Appl Statist*, **52**: 355-363.
- DUCHATEAU L, JANSSEN P (2008), *The Frailty Model*, Statistics for Biology and Health Series. Springer-Verlag New York.
- EPSTEIN B, SOBEL M (1953). Life Testing. *J Amer Stat Assoc*, **48**: 486-502.
- EFRON B (1977). The efficiency of Cox likelihood function for censored data. *Journal of the American Statistical Association*, **72**: 359, 557-565.
- JOHNSON RC, JOHNSON NL (1980). *Survival Models and Data Analysis*, John Wiley & Sons, New York.
- FARADMAL J, TALEBI A, REZAIANZADEH A, MAHJUB H (2012). Survival analysis of breast cancer patients using Cox and frailty models. *J Res Health Sci*, **12**: 127-130.
- FEINLEIB M (1960). A method of analysing log-normally distributed survival data with incomplete follow-up. *Journal of the American Statistical Association*, **55**: 534-545.
- GAIL MH, SANTNER TJ, BROWN CC (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, **36**: 255-266.
- GHARIBVAND L, LIU L (2009). Analysis of survival data with clustered events. In *SAS Global Forum*, 237, p.: 1-11.

- GHOSH D (2000). *Nonparametric and Semiparametric Analysis of Recurrent Events in the Presence of Terminal Events and Dependent Censoring*. Phd Dissertation, Washington University, United States.
- GOVINDARAJULU US, LIN H, LUNETTA KL, D'AGOSTINO RB Sr (2011). Frailty models: Applications to biomedical and genetic studies. *Stat Med*, **30**: 2754-2764.
- GREENWOOD M, YULE GU (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, **83**: 255-279.
- GROVER G, SETH D. (2014). Application of frailty models on advance liver disease using gamma as frailty distribution. *SRL*, **3**: 42-50.
- GUO Z, GILL TM, ALLORE HG (2008). Modeling repeated time-to-event health conditions with discontinuous risk intervals: an example of a longitudinal study of functional disability among older persons. *Methods Inf Med*, **47**: 107-116.
- GUTIERREZ RG (2002). Parametric frailty and shared frailty survival models. *The Stata Journal*, **2**: 22-44.
- HANAGAL DD (2010). Modelling heterogeneity for bivariate survival data by the compound poisson distribution. *Model Assisted Statistics and Applications*, **5**: 1-9.
- HANAGAL DD (2011). *Modelling Survival Data Using Frailty Models*. (1 st Edition). London; Chapman and Hall/CRC.
- HANAGAL DD, DABADE AD (2013). Bayesian estimation of parameters and comparison of shared gamma frailty models. *Comm Statist Simulation Comput*, **42**: 910-931.
- HASTIE TJ, TIBSHIRANI RJ (1993). Varying-coefficient models. *J R Statist Soc B*, **55**: 757-796.
- HORNER RD (1987). Age at onset of Alzheimer's disease: Clue to the relative importance of etiologic factors? *Am J Epidemiol*, **126**: 409-414.
- HOSMER DW, LEMESHOW S (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data*, (2nd Ed.) John Wiley and Sons, New York , 87-253.
- HOUGAARD P (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**: 387-396.
- HOUGAARD P (1987). Modelling multivariate survival. *Scand J Statist*, **14**:291-304.
- HOUGAARD P (2000). *Analysis of Multivariate Survival Data*, Springer, New York.
- HUANG Y, LOUIS TA (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, **85**: 785-798.

- HUANG Y (2000). Two-sample multistate accelerated sojourn times model. *Journal of the American Statistical Association*, **95**: 619-627.
- HUANG CY, WANG MC (2004). Joint modelling and estimation for recurrent event processes and failure time data. *J Am Stat Assoc*, **99**: 1153-1165.
- HUANG CY, QIN J, WANG MC (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics*, **66**: 39-49.
- JAHN-EIMERMACHER A (2008). Comparison of the Andersen-Gill model with poisson and negative binomial regression on recurrent event data. *Computational Statistics and Data Analysis*, **52**: 4989-4997.
- JANK W (2005). Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM, *Computational Statistics & Data Analysis*, **48**: 685-701.
- KALBFLEISCH JD, PRENTICE RL (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.
- KARASOY D, TUTKUN NA (2016). *Yaşam Çözümlemesi (Birinci Basım)*. Nobel Akademik Yayıncılık.
- KEIDING N, ANDERSEN PK, KLEIN JP (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat Med*, **16**: 215-224.
- KELLY PJ, LIM LL (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, **19**: 13-33.
- KLEIN JP, MOESCHBERGER ML, LI YH, WANG ST (1992). *Estimating Random Effects in the Framingham Heart Study, Survival Analysis: State of the Art*, Kluwer Academic, pp. 99-120, Boston, Massachusetts.
- KLEIN JP, MOESCHBERGER, ML (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd Edition, Springer, New York.
- KLEINBAUM DG, KLEIN M (1996). *Survival Analysis: A Self-Learning Text*. Springer, New York, 1-367.
- KLEINBAUM DG, KLEIN M (2005). *Survival Analysis*. New York: Springer Science-Business Media, Inc.
- KLEINBAUM DG, KLEIN M (2012). *Survival Analysis: A Self-Learning Text* (3rd Ed.). Springer, New York.
- LEE ET (1992). *Statistical Methods for Survival Data Analysis*. John Wiley, USA.
- LEE ET, WANG J (2003). *Statistical Methods for Survival Data Analysis*, John Wiley.

- LIN JS, WEI LJ (2002). Linear regression analysis for multivariate failure time observations. *Journal of the American Statistical Association*, **87**: 1091-1097.
- LIN DY, SUN W, YING Z (1999). Nonparametric estimation of the gap time distribution for serial events with censored data, *Biometrika*, 86, 59-70.
- LIU L, WOLFE RA, HUANG X (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60: 747-756.
- MAHE C, CHEVRET S (2001). Analysis of recurrent failure times data: Should the baseline hazard be stratified? *Statistics in Medicine*, **20**: 3807-3815.
- MANATUNGA AK, OAKES, D (1999). Parametric Analysis of Matched Pair Survival Data. *Lifetime Data Analysis*, 5: 371-387.
- McGILCHRIST CA, AISBETT CW (1991). Regression with frailty in survival analysis. *Biometrics*, **47**: 461-466.
- NIELSEN GG, GILL RD, ANDERSEN PK, SORENSEN TIA (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand J Stat*, **19**: 25-43.
- OAKES D (1982). A model for association in bivariate survival data. *J R Stat Soc Ser B*, **44**: 414-422.
- OAKES D (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**: 487-493.
- ÖZKARTAL B (2009). Yinelemeli Olay Analizi ve Çocuk Suçluluğu Üzerine Bir Uygulama. Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü.
- PETERSEN JH (1998). An additive frailty model for correlated life times. *Biometrics*, **54**: 646-661.
- PICKLES A, CROUCHLEY R, SIMONOFF E, EAVES L, MEYER J, RUTTER M, HEWITT J, SILBERG J (1994). Survival models for developmental genetic data: Age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology*, **11**: 155-170.
- PIKE MC (1966). A method of analysis of a certain class of experiments in carcinogenesis, *Biometrics*, **22**: 142-161.
- PRENTICE RL, WILLIAMS BJ, PETERSON AV (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**: 373-379.
- SAHU SK, DEY DK, ASLANIDOU H, SINHA D (1997). A weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis*, **3**: 123-137.
- SARI NWW, PURNAMI SW (2015). Survival analysis for recurrent event data with Andersen-Gill approach. *Proceeding of the IConSSE FSM SWCU*, **BC**: 51-54.
- SCOTT BR, HAHN FF (1980). A model that leads to the Weibull distribution function to characterize early radiation response probabilities. *Health Phys*, **39**: 521-530.

- STANLEY C, MOLYNEUX E, MUKAKA M (2016). Comparison of performance of exponential, Cox proportional hazards, weibull and frailty survival models for analysis of small sample size data. *Journal of Medical Statistics and Informatics*, **4**: 2. DOI:10.7243/2053-7662-4-2.
- TABACHNICK BG, FIDELL LS (2013). *Using Multivariate Statistics*. (6th ed), Pearson.
- TAHERI SS, ARDAKANI MRK, KARIMLOU M, RAHGOZAR M (2016). Identifying some risk Factors of time to recurrent relapses in bipolar I disorder patients using frailty model of survival analysis. *J Psychiatry*, **19**: 376, doi:10.4172/jop.1000376.
- TERZİ Y (2003). *Sansürlü Veriler İçin Sağkalım Analizi ve Gerçek Verilere Uygulaması*. Doktora Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü.
- THERNEAU TM, HAMILTON SA (1997). rhDNase as an example of recurrent event analysis. *Stat Med*, **16**: 2029-2047.
- THERNEAU TM, GRAMBSCH PM (2000). *Modelling Survival Data: Extending the Cox Model*. (1st Ed), Springer, New York, 1-229.
- THERNEAU TM, GRAMBSCH PM, PANKRATZ VS (2003). Penalized survival models and frailty. *J Comput Graph Stat*, **12**: 156-175.
- TUTKUN AT, YEĞEN D (2016). Unshared and shared frailty models - Paylaşılmamış ve paylaşılmış zayıflık modelleri. *Alphanumeric Journal*, **4**: 045-056.
- ULLAH S, GABBETT TJ, FINCH CF (2014). Statistical modelling for recurrent events: an application to sports injuries. *Br J Sports Med*, **48**: 1287-1293.
- ULUBAY G, YILDIZ Ö (2013). *Kronik Obstrüktif Akciğer Hastalığı (KOA) H*, Rota Tıp Yayıncılık
- ÜNLÜ H (2013). *Bilgi İçeren Durdurma Varlığında Yinelemeli Olay Süreci*. Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü.
- VAUPEL JW, MANTON KG, STALLARD E (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**: 439-454.
- VAUPEL JW (1991a). Relatives' risks: frailty models of life history data. *Theor Popul Biol*, **37**: 220-234.
- VAUPEL JW (1991b). Kindred lifetimes: Frailty models in population genetics. In: Adams, Julian, et al., eds. *Convergent Questions in Genetics and Demography* (pp. 155–170). Oxford: Oxford University Press.
- VAUPEL JW, YASHIN AI, HAUGE M, HARVALD B, HOLM N, LIANG X (1991). Survival analysis in genetics: Danish twins data applied to gerontological question. In: Klein, J.P., Goel, P.K., eds. *Survival Analysis: State of the Art* (pp. 121–138). Dordrecht: Kluwer Academic Publisher.

- VAUPEL JW, YASHIN AI, HAUGE M, HARVALD B, HOLM N, LIANG X (1992). Strategies of modelling genetics in survival analysis. What can we learn from twins data? *Paper Presented on PAA Meeting in Denver, CO*, April 30–May 2, 1992.
- VISSER M (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika*, **71**: 507-518.
- WANG MC, CHANG SH (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, **94**: 146-153.
- WEI LJ, LIN DY, WEISSFELD L (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of American Statistical Associations*, **84**: 1065-1073.
- WIENKE A (2010). *Frailty Models in Survival Analysis*. Chapman and Hall Biostatistics Series (Book 37), (1st Ed), 324 pp. New York.
- WHITMORE GA, LEE MLT (1991). A multivariate survival distribution generated by an inverse gaussian mixture of exponentials. *Technometrics*, **33**: 39-50.
- XUE X, BROOKMEYER R (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, **2**: 277-289.
- YASHIN AI, VAUPEL JW, IACHINE IA (1995). Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, **5**: 145-159.
- YASHIN AI, IACHINE IA (1995). Genetic analysis of durations: correlated frailty model applied to survival of Danish twins. *Genet Epidemiol*, **12**: 529-538.
- ZENG D, LIN DY (2008). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*, **65**: 746-752.
- ZENG D, CHEN Q, IBRAHIM JG (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika*, **96**: 277-291.

ÖZGEÇMİŞ

I-Bireysel Bilgiler

Adı : Şirin

Soyadı : Çetin

Doğum Yeri ve Tarihi : Amasya

Uyruğu : T.C

Medeni Durumu : Bekâr

İletişim Adres ve Telf : Ondokuz Mayıs Üniversitesi

II- Eğitimi

Doktora : Ankara Üniversitesi Tıp Fakültesi Biyoistatistik AD.

(2012-

Yüksek Lisans : Ondokuz Mayıs Üniv. Tıp Fakültesi Biyoistatistik AD.

(2007-2009)

Bilimsel Hazırlık : Ondokuz Mayıs Üniv. Tıp Fakültesi (Dönem 2)

Yabancı Dil : İngilizce

III-Unvanları

Biyoistatistik Bilim Uzmanı : 2009-2016

IV-Mesleki Deneyimi

Akademik Uzman : Ondokuz Mayıs Üniversitesi 2005-

V-Üye Olduğu Bilimsel Kuruluşlar

VI-Bilimsel İlgi Alanları

Çetin, Ş., Köse S.K., Ateş, C., Kovaryans Analizi Alternatifleri, 16. Ulusal Biyoistatistik Kongresi, Antalya, 2014.

Çetin, Ş., Köse S.K., Frailty Models: Sağlık Alanında Bir Uygulama , 17. Ulusal Biyoistatistik Kongresi, Girne-KKTC, 2015.

Çakmak M, Göllü G, Boybeyi Ö, Küçük G, Sertçelik M, Günal YD, Aslan MK, Soyer T, Koçak OM, Çakmak A, Çetin Ş, Bingöl-Koloğlu M, Yağmurlu A, Dindar H, Kılıç BG. Cognitive and behavioral characteristics of children with caustic ingestion. J Pediatr Surg. 2015;50:540-2.

VII-Bilimsel Etkinlikler

Verdiği Seminerler

2013- Ankara, Parametrik Olmayan Kovaryans Analizi

2014-Ankara, Cox Oransal Hazard Modeli ve Nonparametrik Randomization Based Modelin Karşılaştırılması