



**T.C.
ESKİŐEHİR OSMANGAZI ÜNİVERSİTESİ
SAĐLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

**KESİŐEN İKİ YAŐAM FONKSİYONU İÇİN
KULLANILAN İSTATİSTİKSEL TESTLERİN
KARŐILAŐTIRILMASI**

DOKTORA TEZİ

Hülya ÖZEN

**DANIŐMAN
Doç. Dr. Cengiz BAL**

2019



**T.C.
ESKİŐEHİR OSMANGAZI ÜNİVERSİTESİ
SAĐLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

**KESİŐEN İKİ YAŐAM FONKSİYONU İÇİN
KULLANILAN İSTATİSTİKSEL TESTLERİN
KARŐILAŐTIRILMASI**

DOKTORA TEZİ

Hülya ÖZEN

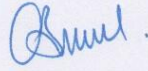
**DANIŐMAN
Doç. Dr. Cengiz BAL**

KABUL VE ONAY SAYFASI

Hülya ÖZEN'İN Doktora Tezi olarak hazırladığı “KESİŞEN İKİ YAŞAM FONKSİYONU İÇİN KULLANILAN İSTATİSTİKSEL TESTLERİN KARŞILAŞTIRILMASI” başlıklı bu çalışma Eskişehir Osmangazi Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddesi uyarınca değerlendirilerek “KABUL” edilmiştir.

28.06.2019

Üye : Doç. Dr. Cengiz BAL (Danışman)



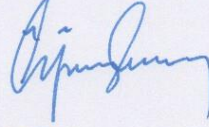
Üye : Prof. Dr. K. Setenay ÖNER



Üye : Doç. Dr. Ertuğrul ÇOLAK



Üye : Doç. Dr. K. Özgür PEKER



Üye : Doç. Dr. Ünal ERKORKMAZ



Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu'nun
.../.../ tarih ve/..... sayılı kararı ile onaylanmıştır.

Prof Dr. Özkan ALATAŞ
Enstitü Müdürü

Özet

KESİŞEN İKİ YAŞAM FONKSİYONU İÇİN KULLANILAN İSTATİSTİKSEL TESTLERİN KARŞILAŞTIRILMASI

İki yaşam fonksiyonunun eşitliğinin araştırılması yaşam analizi çalışmalarında sıklıkla kullanılmaktadır. Literatürde en çok tercih edilen yöntem log-rank testidir, ancak orantılı hazardlar varsayımının ihlal edildiği kesişen yaşam fonksiyonlarını değerlendirmede yanıltıcı sonuçlar sunmaktadır. Kesişen yaşam fonksiyonları birçok farklı yapıda oluşabilmektedir. Tek başına tüm koşulları uygun bir şekilde değerlendirilecek bir test ise bulunmamaktadır. Bu nedenle kesişen yaşam fonksiyonlarını değerlendirmek için farklı karşılaştırma testleri önerilmiştir.

Bu tez çalışmasında literatüre önerilen güncel yöntemlerden parçalanmış log-rank testleri ve ağırlıklandırılmış Lin ve Wang testleri, log-rank ve ağırlıklandırılmış log-rank testleri ile karşılaştırılmıştır. Monte Carlo simülasyon tekniği kullanılarak farklı örneklem büyüklüğü ve sansür oranlarında karşılaştırma testlerinin tip 1 hata oranları ve güçleri kıyaslanmıştır. Orantılı hazardlar varsayımının sağlandığı ve kesişen yaşam fonksiyonları ile bu varsayımın ihlal edildiği senaryolarda karşılaştırma testlerinin zayıf ve güçlü yönleri incelenmiştir. Simülasyon çalışmasına ek olarak sağlık alanında gerçek bir yaşam veri setinde de karşılaştırma testleri değerlendirilmiştir.

Simülasyon çalışmaları sonucunda tüm testler makul tip 1 hata oranları sunmuştur. Orantılı hazardlar varsayımı altında, log-rank en başarılı karşılaştırma testidir. Kesişen yaşam fonksiyonlarına ait senaryolarda ise en başarılı sonuçlar parçalanmış log-rank testlerine aittir. Lin ve Wang testleri ve ağırlıklandırılmış log-rank testleri, kesişen yaşam fonksiyonlarına ait senaryolarda parçalanmış log-rank testlerinden oldukça düşük güç oranları sunmuştur. Örneklem büyüklüğündeki artış testlerin performansını arttırmış, ancak sansür oranındaki artış olumsuz bir etki yaratmıştır. Yaşam fonksiyonlarının kesişim noktasının çalışmanın başlangıcından sonuna doğru

ilerlemesi, testlerin performansını olumsuz yönde etkilemiştir. En büyük etki log-rank, en küçük etki ise parçalanmış log-rank testlerinde görülmüştür. Gerçek yaşam veri setinden elde edilen sonuçlar, simülasyon çalışması sonuçları ile uyumlu bulunmuştur.

Sonuç olarak orantılı hazardlar varsayımı sağlandığında log-rank testinin, kesişen yaşam fonksiyonlarının karşılaştırılmasında ise parçalanmış log-rank testlerinin kullanılması önerilmektedir.

Anahtar Kelimeler: Ağırlıklandırılmış Lin ve Wang testleri, Kesişen yaşam fonksiyonları, Log-rank testi, Parçalanmış log-rank testleri, Yaşam analizi

Summary

COMPARISON OF THE STATISTICAL TESTS FOR TWO CROSSING SURVIVAL FUNCTIONS

The investigation of the equality of two survival functions is frequently used in survival analysis studies. The most preferred method in the literature is the log-rank test, but it provides misleading results in evaluating the crossing survival functions in which the proportional hazards assumption is violated. Crossing survival functions can occur in many different ways. There is not a single test to properly assess all conditions. Therefore, different comparison tests have been proposed to evaluate crossing survival functions.

In this thesis study, partitioned log-rank tests and weighted Lin and Wang tests, which are the current methods proposed in the literature, were compared with log-rank and weighted log-rank tests. Using the Monte Carlo simulation technique, type 1 error rates and powers of comparison tests were compared in different sample sizes and censorship rates. The weaknesses and strengths of the comparison tests were examined in scenarios where the proportional hazards assumption was provided and violated with two crossing survival functions. In addition to the simulation study, comparison tests were evaluated in a real life data set in the field of health.

As a result of the simulation studies, all tests provided reasonable type 1 error rates. Under the assumption of proportional hazards, log-rank is the most successful comparison test. The most successful results in the scenarios of crossing survival functions were belong to partitioned log-rank tests. Lin and Wang tests and weighted log-rank tests showed lower power rates than the partitioned log-rank tests in scenarios of crossing survival functions. The increase in the sample size increased the performance of the tests, while the increase in the censorship rate had a negative effect. From the beginning to the end of the study, the progress of the crossing point of survival functions affected the performance of the tests adversely. The largest effect was seen in log-rank, and the smallest effect was in partitioned log-rank tests. The results obtained

from the real life data set were consistent with the results of the simulation study.

In conclusion, it is recommended to use the log-rank test when the proportional hazards assumption is provided and the use of partitioned log-rank tests in the comparison of crossing survival functions.

Keywords: Weighted Lin and Wang tests, Crossing survival functions, Log-rank test, Partitioned log-rank tests, Survival analysis



İçindekiler

Özet.....	iii
Summary	v
Tablo Dizini	viii
Şekil Dizini	ix
Simge ve Kısaltmalar Dizini	x
1. GİRİŞ VE AMAÇ	1
2. GENEL BİLGİLER	5
2.1. Yaşam Analizinde Kullanılan Kavramlar.....	5
2.1.1. Yaşam fonksiyonu (Survival function).....	8
2.1.2. Hazard fonksiyonu (Hazard function).....	10
2.2. Orantılı Hazardlar Modeli (Proportional Hazards Model)	12
2.2.1. Grafikselsel yaklaşım	15
2.2.2. Uyum iyiliği testi yaklaşımı.....	17
2.2.3. Zamana bağlı açıklayıcı değişken varlığının araştırılması	18
2.3. İki Yaşam Fonksiyonunun Karşılaştırılmasında Kullanılan Testler.....	19
2.3.1. Log-rank ve ağırlıklandırılmış log-rank testleri.....	19
2.3.2. Parçalanmış log-rank testleri (Partitioned log-rank tests)	23
2.3.3. Ağırlıklandırılmış Lin ve Wang testleri (Weighted Lin and Wang tests)....	27
3. GEREÇ VE YÖNTEMLER.....	31
3.1. Veri Türetim Senaryoları ve R Paket Programında Algoritmaların Oluşturulması.....	31
3.2. Kullanılan Testlerin Algoritmalarının Oluşturulması	36
3.2.1. Log-rank ve ağırlıklandırılmış log-rank testlerinin algoritması.....	36
3.2.2. Parçalanmış log-rank testlerinin algoritmaları.....	37
3.3.3. Ağırlıklandırılmış Lin ve Wang testlerinin algoritmaları.....	40
3.3. Simülasyon Algoritmasının Oluşturulması	48
3.4. Uygulama Veri Seti	50
4. BULGULAR	52
4.1. Simülasyon Çalışması Bulguları.....	52
4.2. Uygulama Veri Seti Bulguları.....	60
5. TARTIŞMA	63
6. SONUÇ VE ÖNERİLER	68
KAYNAKLAR DİZİNİ.....	70
ÖZGEÇMİŞ.....	74

Tablo Dizini

Tablo 2.1	Hazard fonksiyonunun diğer bilim dallarındaki isimleri	11
Tablo 2.2	t_i zamanında iki grubun gösterge değişkene göre dağılımı	20
Tablo 2.3	Ağırlık fonksiyonlarına göre ağırlıklandırılmış log-rank testleri	22
Tablo 2.4	Ağırlıklandırılmış Lin ve Wang Testlerine ait Ağırlık Fonksiyonları	29
Tablo 3.1	Uygulama veri setine ait veriler	51
Tablo 4.1	Senaryo 1'e göre karşılaştırma testlerinin tip 1 hata oranları	55
Tablo 4.2	Senaryo 2'ye göre karşılaştırma testlerinin güç oranları	56
Tablo 4.3	Senaryo 3'e göre karşılaştırma testlerinin güç oranları	57
Tablo 4.4	Senaryo 4'e göre karşılaştırma testlerinin güç oranları	58
Tablo 4.5	Senaryo 5'e göre karşılaştırma testlerinin güç oranları	59
Tablo 4.6	Cox Regresyon modellerine ait sonuçlar	61
Tablo 4.7	Uygulama veri setine ait karşılaştırma testi sonuçları	62

Şekil Dizini

Şekil 2.1	Teorik yaşam fonksiyonu grafiği	9
Şekil 2.2	Kaplan-Meier yöntemiyle çizilen yaşam fonksiyonu grafiği	10
Şekil 2.3	Hazard Fonksiyonu Grafiği	11
Şekil 2.4	Orantılı hazardlar varsayımı altında iki gruba ait yaşam ve hazard fonksiyonları	14
Şekil 2.5	Orantılı hazardlar durumu ihlalinde iki gruba ait yaşam ve hazard fonksiyonları	14
Şekil 2.6	Log-log yaşam fonksiyonları grafiği	16
Şekil 3.1	Simülasyon çalışmalarında kullanılan senaryolar	32
Şekil 3.2	Tedavi gruplarına ait yaşam fonksiyonları grafiği	51
Şekil 4.1	Tedavi gruplarına ait log-log yaşam fonksiyonları grafiği	60

Simge ve Kısaltmalar Dizini

Simgeler

t	Gözlenen yaşam süresi
$S(t)$	Yaşam fonksiyonu
$h(t)$	Hazard fonksiyonu
$f(t)$	Olasılık yoğunluk fonksiyonu
δ, ε	Gösterge değişken
C_r, C_l	Sabit sansürlü zaman
$F(t)$	Kümülatif olasılık yoğunluk fonksiyonu
$H(t)$	Kümülatif hazard fonksiyonu
$h(t Y)$	Açıklayıcı değişkenlere bağlı hazard fonksiyonu
Y_i	i. açıklayıcı değişken
$g(t)$	Zamana bağlı fonksiyon
$h_0(t)$	Temel hazard fonksiyonu
$S_1(t)$	Birinci gruba ait yaşam fonksiyonu
$h_1(t)$	Birinci gruba ait hazard fonksiyonu
$S_2(t)$	İkinci gruba ait yaşam fonksiyonu
$h_2(t)$	İkinci gruba ait hazard fonksiyonu
n	Çalışmadaki toplam birim sayısı
n_1	Birinci grupta yer alan birim sayısı
n_2	İkinci grupta yer alan birim sayısı
t_1	Birinci gruba ait gözlenen yaşam süresi
t_2	İkinci gruba ait gözlenen yaşam süresi
t_i	Herhangi bir yaşam süresi noktası
X_{kj}	k grubunda j. birime ait gözlenen yaşam süresi
δ_{kj}	k grubunda j. birime ait gösterge değişken değeri
n_i	t_i zamanda risk altında kalan toplam birim sayısı
d_i	t_i zamanda olay ile karşılaşan toplam birim sayısı
n_{1i}	t_i zamanda birinci grupta risk altında kalan birim sayısı
n_{2i}	t_i zamanda ikinci grupta risk altında kalan birim sayısı
d_{1i}	t_i zamanda birinci grupta olay ile karşılaşan birim sayısı
d_{2i}	t_i zamanda ikinci grupta olay ile karşılaşan birim sayısı
w_i	t_i zamanda ağırlık fonksiyonu değeri
$t^{(i)}$	Küçükten büyüğe sıralanmış yaşam süresi noktalarından i. sırada yer alan yaşam süresi değeri
τ	En büyük gözlenen yaşam süresi
Z_1	Beklenen ve gözlenen olay sayıları arasındaki farklar toplamı
σ_1	Z_1 'in standart hatası
$\chi^2_{\log-rank}$	Log-rak test istatistiği
$\chi^2_{\text{ağırlıklandırılmış log-rank}}$	Ağırlıklandırılmış log-rank test istatistiği
T	Parçalanmış log-rank test istatistiği
$\tilde{T}(t)$	t noktasından büyük ve küçük verilere ait log-rank test istatistiklerinin toplamı

$D_l(t)$	t noktasından küçük verilerde beklenen ve gözlenen olay sayıları arasındaki farklar toplamı
$D_u(t)$	t noktasından büyük verilerde beklenen ve gözlenen olay sayıları arasındaki farklar toplamı
$V_l(t)$	$D_l(t)$ 'nin varyansı
$V_u(t)$	$D_u(t)$ 'nin varyansı
Δ	Beklenen ve gözlenen olay sayıları arasındaki farkların karelerinin toplamı
Δ_w	Ağırlık fonksiyonu ile çarpılmış beklenen ve gözlenen olay sayıları arasındaki farkların karelerinin toplamı
T_w	Lin ve Wang testleri test istatistiği
$d_{1i}^{(m)}$	d_{1i} 'nin m. dereceden ham momenti
$n^{(m)}$	n'nin m. dereceden ham momenti
$B(t)$	Brownian Motion
$U(0,1)$	Standart tekdüze dağılım
λ	Üstel dağılım parametresi
$W(\alpha, \beta)$	α ve β parametrelili Weibull dağılımı

Kısaltmalar

K-M	Kaplan–Meier Yöntemi
HR	Hazard Oranı (Hazard Ratio)
c	Sansür Oranı
LR	Log-rank Testi
GW	Gehan-Wilcoxon testi
TW	Tarone-Ware testi
PP	Peto-Peto testi
MPP	Modifiye edilmiş Peto-Peto testi
FH	Fleming-Harrington(1,1)testi
SUP-LR	Parçalanmış log-rank testi
SUP-GW	Parçalanmış Gehan-Wilcoxon testi
SUP-PP	Parçalanmış Peto-Peto testi
LW	Lin ve Wang testi
LW_{w1}	n_i ' e göre ağırlıklandırılmış Lin ve Wang testi
LW_{w2}	$\sqrt{n_i}$ ' e göre ağırlıklandırılmış Lin ve Wang testi
LW_{w3}	$1/\sqrt{Var(d_{1i})}$ 'e göre ağırlıklandırılmış Lin ve Wang testi

1. GİRİŞ VE AMAÇ

Yaşam analizleri; belirlenen bir zaman noktasından, ilgilenilen olay gözlenene kadar geçen süreleri değerlendirmek amacıyla geliştirilen istatistiksel yöntemler ailesidir. Yaşam analizleri, literatürde en sık sağlık bilimleri alanında farklı hastalıkların tedavisinde yararlanılan yöntemlerin değerlendirilmesi amacıyla kullanılmaktadır. Ayrıca, sosyal ve doğa bilimleri alanlarında, hastalığın başlangıcı, sanayide ekipman arızaları, depremler, trafik kazaları, iş feshleri, doğumlar, evlilikler, boşanmalar, terfiler, sigortacılık ve hukuk dahil olmak üzere bir çok farklı olayı incelemek için de kullanılmıştır. Yaşam analizi farklı alanlardaki araştırmacılar tarafından farklı isimlerle tanımlanmaktadır: olay tarihi analizi (sosyoloji), güvenilirlik analizi (mühendislik), başarısızlık süresi analizi (mühendislik), süre analizi (ekonomi) ve geçiş analizi (ekonomi). Farklı isimler farklı yaklaşımları vurgulasa da, uygulanan teknikler aynıdır (Allison, 2004; Klein ve Moeschberger, 2005).

Yaşam analizinde, çalışmaya dahil edilen grupların yaşam sürelerinin karşılaştırılması önemli bir yer tutmaktadır. Farklı grupların yaşam sürelerinin karşılaştırılması amacıyla birden fazla sayıda yöntem önerilmektedir. Bunlar içerisinde en bilinen ve yaygın kullanılanı log-rank testidir (Mantel, 1966). Log-rank testi, orantılı hazardlar varsayımı altında karşılaştırılan yaşam fonksiyonları arasındaki farkı belirlemede optimum güce sahiptir (Lin ve Xu, 2010; Rosner, 2010; Bouliotis ve Billingham, 2011).

Literatürde, yaşam fonksiyonları arasındaki farklılıkların çeşitli şekillerde olduğu görülebilir. Örneğin, bir çalışmanın izleminde erken ya da geç dönemde farklılıklar olabilir ve yaşam fonksiyonları bazen kesişebilir. Bu durum hem erken hem de geç izleminde olası farklılıkları gösterir, ancak farklılıklar aynı yönde olmaz. Alternatif senaryolar dikkate alındığında, gruplar arası yaşam fonksiyonlarının karşılaştırılmasında seçilecek testler de farklılık göstermektedir (Balakrishnan ve Rao, 2004).

Log-rank testinin uygulanmasında orantılı hazardlar varsayımı bir ön koşul olup, hazard fonksiyonlarının her bir grupta t zamanı boyunca birbiriyle

orantılı şekilde ilerlemesi beklenir. Hazard fonksiyonlarındaki orantının bozulması ve t zamanı boyunca herhangi bir noktada kesişmesi durumunda söz konusu varsayım ihlal edilmiş olur. Bu duruma örnek olarak bir tedavi yönteminin erken dönemde iyi etki gösterip ilerleyen dönemlerde bu etkiyi sürdürememesi gösterilebilir (Stablein ve Koutrouvelis, 1985). Varsayımın ihlal edildiği bu gibi durumlarda, log-rank test istatistiğinin hesaplanmasında, yaşam fonksiyonları arasında başlangıçta belirlenen pozitif farklar, ilerleyen zamanlarda hesaplanan ters işaretli farklar tarafından nötrlenir. Bu durum log-rank testinin farkı yakalama gücünü kaybetmesine sebep olur.

İki yaşam fonksiyonunun karşılaştırıldığı çalışmalara yönelik yapılan literatür taramalarında, genellikle orantılı hazardlar varsayımının göz ardı edildiği ve hatalı analiz yöntemi seçimlerinin yapıldığı gözlemlenmiştir (Krishnan vd., 2011; Weeda vd., 2013). Yaşam fonksiyonu grafiği bulunan ve yaşam analizi yöntemleri uygulanan çalışmalara ait yapılan bir literatür taramasında beş önemli derginin (Annals of Internal Medicine, British Medical Journal, JAMA, New England Journal of Medicine (NEJM) ve The Lancet) yayınları incelenmiştir. İncelenen 175 çalışmanın %47'sinde yaşam fonksiyonlarının kesiştiği ve bu çalışmaların %70'inde log-rank testinin ve %66'sında da Cox regresyon modelinin uygulandığı belirlenmiştir. Sadece %31'lik bir kısımda orantılı hazardlar varsayımının araştırıldığı gözlemlenmiştir (Kristiansen, 2012). Bir başka tarama çalışmasında ise bilimsel kongrelerde sunulan, yaşam analizi yöntemlerinin kullanıldığı 107 bildiri incelenmiş ve bu bildirilerden 95'inde log-rank testinin kullanıldığı belirlenmiştir. Yaşam fonksiyonlarının durumları incelendiğinde ise 42 bildiride log-rank testinin kesişen yaşam fonksiyonlarına uygulandığı gözlemlenmiştir. Araştırmacılar, bu yanlış uygulamaların en büyük nedeni olarak paket programlarda sadece log-rank ve ağırlıklandırılmış log-rank testlerinin yer almasını göstermektedir (Balakrishnan ve Rao, 2004).

Literatürde orantılı hazardlar varsayımı ihlalinde kullanılmak üzere çok sayıda test önerilmektedir. Bu testlerin belirli bir kısmı Nelson-Aalen kümülatif hazard fonksiyonları arasındaki farklara dayalı elde edilmiştir. Gill (1980), ağırlıklandırılmış log-rank testleri ve matematiksel supremum fonksiyonu

yaklaşımı ile Renyi testlerini geliştirmiştir. Modifiye edilmiş Kolmogorov-Smirnov testi de kesişen yaşam fonksiyonları için önerilen bir başka testtir (Fleming, O'Fallon, O'Brien & Harrington, 1980). Özellikle sansürlü veri setleri için genelleştirilmiş Cramer-von Mises testleri, Koziol (1978) ve Schumacher (1984) tarafından önerilmiştir. Kaplan-Meier yöntemi ile tahmin edilen iki yaşam fonksiyonunun farkına dayalı karşılaştırma testleri de önerilmiştir. Pepe ve Fleming (1989) ağırlıklandırılmış Kaplan-Meier testlerini literatüre sunmuşlardır. Bu yöntem farklı bakış açısı sunan alternatif iki farklı çalışma da önerilmiştir (Shen ve Cai, 2001; Lee, 2007). Lin ve Wang (2004) log-rank test istatistiğinin karesini alarak yeni bir test istatistiği önermiştir. Lin ve Xu (2010) ise iki yaşam fonksiyonu altında kalan alanlar üzerine yeni bir test elde etmişlerdir. Kraus (2009), Neyman'ın düzgünleştirici (smooth) testlerini kullanarak farklı testler geliştirmiştir. Qiu ve Sheng (2008) iki aşamalı bir yöntem geliştirerek kesişen yaşam fonksiyonları durumunda kullanılmasını önermişlerdir. Koziol ve Jia (2014), Lin ve Wang tarafından önerilen karşılaştırma testine farklı ağırlık fonksiyonları ekleyerek literatüre öneride bulunmuşlardır. Liu ve Yin (2016), belirli zaman noktalarından veri setini bölerek elde ettiği alt gruplara ait log-rank test istatistiği toplamları ile yeni yöntemler önerisinde bulunmuşlardır. Literatüre önerilen bu testlerin birçoğu farklı senaryolar altında Monte Carlo simülasyon yöntemi kullanılarak log-rank ve bazı ağırlıklandırılmış log-rank testleri ile karşılaştırılmıştır. Hatta literatürde bu testlerden bazılarını da birbiri ile kıyaslayan çalışmalar bulunmaktadır (Jurkiewicz ve Wycinka, 2011; Li, Han, Hou, Chen & Chen, 2015).

Literatürde birçok parametrik olmayan yaşam analizi yöntemi olmasına rağmen, belirli koşullar altında hangi yöntemin daha iyi olduğunu belirleyen kurallar seti bulunmamaktadır. Bunun sebebi orantılı hazardlar varsayımının olmadığı durumlardaki alternatiflerin heterojen bir aile oluşturması ve genelleme yapılmasını engellemesidir (Balakrishnan ve Rao, 2004). Bu tez çalışmasında literatüre yeni önerilen ve birbirleri ile karşılaştırılmamış testler olan parçalanmış log-rank testleri ve ağırlıklandırılmış Lin ve Wang testleri kullanılmıştır. Bu testlere ilave olarak log-rank ile Gehan-Wilcoxon (Gehan,

1965), Tarone Ware (Tarone ve Ware, 1977), Peto-Peto (Peto ve Peto, 1972), Modifiye Edilmiş Peto-Peto (Andersen, Borgan, Gill & Keiding, 1992; Andersen ve Gill, 1982), Fleming-Harrington (Harrington ve Fleming, 1982) gibi ağırlıklandırılmış log-rank testlerine de yer verilmiştir. Bu tez çalışmasının amacı, yaşam analizinde sıklıkla görülen çeşitli senaryolarda bu karşılaştırma testlerinin güçlü ve zayıf yönlerini değerlendirmek ve elde edilen sonuçlar doğrultusunda önerilerde bulunmaktır. Bu amaç doğrultusunda farklı senaryolar, farklı örneklem büyüklüğü ve sansür oranlarında simülasyon çalışmaları gerçekleştirilmiştir. Ardından, uygulama veri seti ile testlerin kullanım alanına örnek verilmiştir. Tezin ikinci bölümünde genel bilgilere yer verilmiş ve yaşam analizine ait temel kavramlar ile kullanılan karşılaştırma testlerinden bahsedilmiştir. Simülasyon çalışmasının detayları ve yazılan kodlamalar, üçüncü bölüm olan Gereç ve Yöntemler başlığı altında anlatılmıştır. Dördüncü bölüm Bulgular' da, simülasyon çalışması ve uygulama setine ait sonuçlara yer verilmiştir. Literatür ile karşılaştırmalar beşinci bölüm olan Tartışma' da belirtilmiş, genel sonuç ve değerlendirmeler altıncı bölüm Sonuç ve Öneriler başlığı altında yapılmıştır.

2. GENEL BİLGİLER

Yaşam analizi başta biyomedikal çalışmalar olmak üzere pek çok alanda kullanılmaktadır. Parametrik ve parametrik olmayan çok sayıda istatistiksel yöntem, diğer veri yapılarından farklılık gösteren yaşam veri setlerine uygulanır. Yaşam analizinde kullanılan yaşam veri setlerinin iki temel özelliği bulunmaktadır. Bunlar; sansürlü gözlem içermeleri ve zamana bağlı değişkenler bulundurmalarıdır. Bu temel özellikleri nedeniyle yaşam veri setleri geleneksel yöntemler kullanılarak analiz edilememektedirler. Bu ihtiyaç, yaşam analizi yöntemlerinin ortaya çıkışının temel sebebini oluşturmaktadır.

Yaşam analizi için en iyi gözlem, ileriye yönelik (prospektif) çalışmalar kurgulanarak yapılabilir. İyi tanımlanmış bir zaman noktasında bir kişiler izlenmeye başlanır ve ilgilenilen olayların gerçekleştiği zamanlar kaydedilir. Her bireyin ilgilenilen olay ile karşılaşma zorunluluğu bulunmamaktadır.

Yaşam analizi, geçmişe dönük (retrospektif) verilerle de sıkça kullanılır. Potansiyel sınırlamalar doğru tanımlandığı sürece yaşam analizinin bu şekilde kullanılmasında herhangi bir sorun bulunmamaktadır. Ancak bu tür uygulamalarda bireyler, olayların zamanlarını hatırlamakta önemli hatalar yapabilir ve bazı olayları tamamen unutabilirler. Ayrıca zamana bağlı değişkenler hakkında doğru bilgi sağlamada zorluk çekebilirler. Daha ayrıntılı bir sorun, görüşme yapılan bireylerin, olay riski altında bulunanların yanlış bir alt örnekleme olabileceğidir. Örneğin, ölen ya da taşınan bireyler görüşmelere dahil edilemeyecektir. İleriye dönük veriler daha fazla tercih edilmesine rağmen, geriye dönük verilerden de çok şey öğrenilebilir (Allison, 2004; Armitage, Berry & Matthews, 2008).

2.1. Yaşam Analizinde Kullanılan Kavramlar

Olay (Event, Failure): Yaşam analizinde bir bireyde meydana gelmesi beklenen, tasarlanan ya da ilgilenilen durumdur. Olaya örnek olarak ölüm, tümör oluşumu, bir hastalığın meydana gelmesi, bir hastalığın tekrarlanması (recurrence), bir cihazın bozulması, vb. gösterilebilir. Olay her zaman kötü bir

durum olmak zorunda değildir. Belirli bir tedavi bitiminde remisyonunda kalma, gebe kalma, sigarayı bırakma vb. iyi durumlar da olaya birer örnektir. Aynı analiz içerisinde birden fazla farklı olay da görülebilir; ancak bu durumda uygulanacak analiz yöntemleri de farklı olmaktadır.

Süre (Zaman, Time / Survival Time): Bir birimin takibinin başlangıcından ilgilenilen olay ile karşılaşmasına kadar geçen süreçtir. Yıl, ay, hafta, gün, saat vb. birimler cinsinden ölçülerek değerlendirilir. Bir yaşam analizinde süre, homojen bir toplumdaki geldiği varsayılan ve negatif olmayan bir değişkendir. Genellikle sağa çarpık bir dağılım gösterir. Yaşam analizinde yaşam sürelerini karakterize eden fonksiyonlar bulunmaktadır. Bunlar; yaşam fonksiyonu ($S(t)$, survival function), hazard fonksiyonu ($h(t)$, hazard function) ve olasılık yoğunluk fonksiyonu ($f(t)$, probability density function)dur. Aralarında ilişki bulunan bu fonksiyonlardan biri bilindiğinde, diğerleri de elde edilebilir.

Gösterge (indikatör) değişken (δ): Yaşam süreleri elde edilmiş birimlerin çalışma periyodu içerisinde ilgilenilen olay ile karşılaşmış olup olmadığını belirleyen değişkendir. Olay ile karşılaşan birimlerin gösterge değişken değeri 1, karşılaşmayan bir diğer adıyla sansürlü birimlerin ise değeri 0 olarak gösterilir.

Sansür: Bir birimin ilgilenilen olay ile karşılaştığı zamanın tam olarak bilinmemesi durumudur. Kişinin yaşam süresi ile ilgili bir bilgi vardır, ancak bu süre onun olayı gördüğü zamana kadar geçen süre değildir. Belirli bir t zamanında risk altındaki birim sayısına etkileri bulunduğu için sansürlü veriler de yaşam analizine dahil edilir. Yaşam analizinde karşılaşılan üç farklı sansür mekanizması bulunmaktadır. Bunlar; sağdan sansürlü, soldan sansürlü ve aralıklı sansürlü yapılarıdır.

Sağdan Sansürlü Yapı: Tıp ve biyoloji alanında en sık görülen sansür mekanizmasıdır. Sağdan sansürlü bir yapının ortaya çıkması üç ana başlık altında incelenebilir:

- 1) Çalışma sonunda birimde ilgilenilen olayın görülmemesi,
- 2) Çalışma sürecinde bir birimin takibinin kaybolması,
- 3) Birimin çalışmadan geri çekilmesi.

Örneğin; bir çalışmada belirli bir zaman aralığında remisyondan çıkıncaya kadar takip edilen lösemi hastaları bulunsun. Belirli bir hasta için çalışma, hasta halen remisyundayken biter ise o zaman hastanın gerçekte remisyondan çıkma süresi sansürlenmiş olur. Gerçekte bu hastanın remisyondan çıkma süresi en az takip edildiği süre kadardır.

Sağdan sansürlü bir yapı aşağıdaki notasyon ile tanımlanabilir: Bir çalışmada; takip edilen bireylerin yaşam süresi X ve çalışmanın sabit sansürlü zamanı ise C_r olsun. X lerin birbirinden bağımsız olarak dağıldığı kabul edilsin. Bireyin gerçek yaşam süresi; X 'in C_r 'den küçük ya da büyük olduğu bilinmesi durumunda elde edilecektir. Eğer X süresi C_r 'den küçük ise birey ilgilenilen olay ile karşılaşmıştır, X süresi C_r 'den büyük ise de ilgilenen olay ile karşılaşmamıştır. Bu bilgiler ışığında bireye ait bilgi bir çift rasgele değişken; (T, δ) ile gösterilebilir. Burada δ bireyin ilgilenilen olay ile karşılaşmış karşılaşmadığını belirten gösterge değişken ve $T = \min(X, C_r)$ ise bireyin gözlenen yaşam süresidir.

Soldan Sansürlü Yapı: İlgilenilen bir olayın bir birimde daha çalışmaya katılmadan önce görülmüş olduğunun bilinmesi ancak ne kadar süre önce olduğunun tam olarak bilinmemesidir. Örneğin okul öncesi eğitim merkezlerinde, ilgi genellikle çocukların belirli aktiviteleri yerine getirmeyi öğrenip öğrendiğini belirleme üzerinedir. Bir çocuğun ilgili aktiviteyi öğrenme yaşı, olay zamanı sayılır. Çoğu zaman, bazı çocuklar çalışmaya başladıklarında zaten bu aktiviteyi yerine getirebilirler. Bu tür olay zamanları soldan sansürlenmiş olarak kabul edilir.

Soldan sansürlü yapıya ait notasyon tanımlaması şu şekilde yapılabilir. Çalışmada yer alan bir birimin yaşam süresi X ; C_l sansürlü yaşam süresinden küçük ise bu birimin ilgilenilen olay ile C_l 'den daha önce karşılaştığı bilinmektedir; ancak ne kadar zaman önce olduğu bilinmemektedir. Bu nedenle gerçek yaşam süreleri belirlenmemektedir. Gerçek yaşam süresi ancak ve ancak C_l 'ye eşit ya da büyük olduğunda bilinebilecektir. Soldan sansürlü yapıda da birimlere ait bilgi bir çift rasgele değişken (T, ε) ile gösterilir. Gerçek X yaşam süresinin gözlenip ($\varepsilon = 1$), gözlenmediğini ($\varepsilon = 0$) belirleyen gösterge değişken

ve sağdan sansürlü yapının tersi şeklinde hesaplanan $T = \max(X, C_I)$ ise bireyin gözlenen yaşam süresidir.

Aralıklı Sansürlü Yapı: Daha genel bir sansürleme türüdür. Yaşam süresinin sadece belirli bir zaman aralığı içerisinde yer aldığı ve ilgilenilen olayın da bu aralık içerisinde gerçekleştiği bilinmektedir. Bu sansürlü yapı genellikle klinik denemelerde ya da izlemsel çalışmalarda görülebilir. Örneğin, kalp hastalıkları çalışmalarında bireylerin ilk kez koroner kalp hastalığı geliştirdiği yaşları genellikle tam olarak bilinir. Bununla birlikte, alt kategori anjin pektorisin ilk ortaya çıkma yaşının yaklaşık iki yıl ara ile takip edilen iki klinik muayene arasında olduğu bilinmektedir. Bu tür gözlemler aralıklı olarak sansürlenir (Klein ve Moeschberger, 2005; Kleinbaum ve Klein, 2006; Armitage vd., 2008; Rosner, 2010).

2.1.1. Yaşam fonksiyonu (Survival function)

Yaşam fonksiyonu, yaşam sürelerini tanımlamada kullanılan en temel özelliktir. Bir birimin belirlenen bir t zamanından daha uzun süre ilgilenilen olay ile karşılaşmama olasılığını gösterir.

$$S(t) = P(T > t) \quad (2.1)$$

Yaşam süresi t sürekli bir yapıda ise yaşam fonksiyonu, kümülatif olasılık dağılım fonksiyonunun ($F(t)$) tamamlayandır. Yaşam analizinde $F(t)$; başarısızlık dağılımı (failure distribution) olarak da adlandırılır.

$$S(t) = 1 - F(t) \quad (2.2)$$

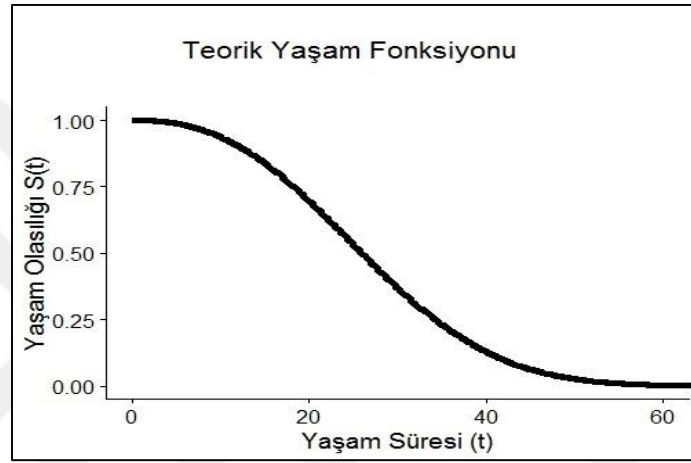
Olasılık yoğunluk fonksiyonu ile yaşam fonksiyonu arasındaki ilişki ise eşitlik 2.3' te belirtilmektedir.

$$f(t) = -\frac{dS(t)}{dt} \quad (2.3)$$

Farklı türlerde yaşam fonksiyonları görülebilmektedir. Ancak hepsinin aynı ortak özellikleri bulunmaktadır. Bu ortak özellikler şu şekilde sıralanabilir:

- Şekil 2.1'de görüldüğü gibi $S(t)$ monoton azalan bir fonksiyondur.

- $t=0$ durumunda $S(t)=1$ 'dir. Çalışmanın başlangıcında hiçbir birim olay ile karşılaşmamış olduğu için birimlerin yaşam olasılığı bire eşit olacaktır.
- $t=\infty$ durumunda $S(t)=0$ 'dır. Teorik bir bakış açısidir. Çalışma periyodu sonsuzluğa doğru uzatılırsa, olay ile karşılaşmayan hiçbir birim kalmayacaktır. Bu nedenle de olasılık sıfır olacaktır.
- Yaşam fonksiyonlarına bakarak ilgilenilen olayın yapısını (failure pattern) belirlemek oldukça zordur.



Şekil 2.1 Teorik Yaşam Fonksiyonu Grafiği

Yaşam analizinde, yaşam fonksiyonuna ait olasılık hesaplamalarına ihtiyaç duyulmaktadır. Yaşam süreleri parametrik olarak teorik bir dağılım yardımıyla modellenabilir. Üstel, Weibull, Gamma, log-normal gibi dağılımlar kullanılarak yaşam fonksiyonları elde edilebilir ve yaşam olasılıkları hesaplanabilir. Ancak yaşam sürelerine ait teorik dağılım fonksiyonu bilinmediği ya da belirlenemediği durumlarda bu olasılıkları hesaplamada kullanılacak parametrik olmayan yaklaşımlara ihtiyaç duyulur. Kaplan-Meier yöntemi bu amaçla geliştirilen ve literatürde sıklıkla kullanılan parametrik olmayan bir yöntemdir (Klein ve Moeschberger, 2005; Kleinbaum ve Klein, 2006).

Kaplan-Meier (K-M) yöntemi, çarpımsal limit (product limit) yaklaşımını kullanarak yaşam fonksiyonu olasılıklarının hesaplanmasını ve yaşam fonksiyonuna ait grafiklerin çizilmesini sağlamaktadır. Genel bir formül ile K-M yöntemini belirtmek gerekirse; $t_{(1)} < t_{(2)} < \dots < t_{(m)}$ ilgilenilen olayın

görüldüğü birbirinden bağımsız zaman aralıklarını gösterebilir. n_i ; $t_{(i)}$ zamanında risk altındaki birim sayısı, d_i ; ise bu zaman noktasında ilgililen olay ile karşılaşan birim sayısını gösterir. K-M yaşam fonksiyonu $\hat{S}(t)$ eşitlik 2.4'te görüldüğü gibi hesaplanır:

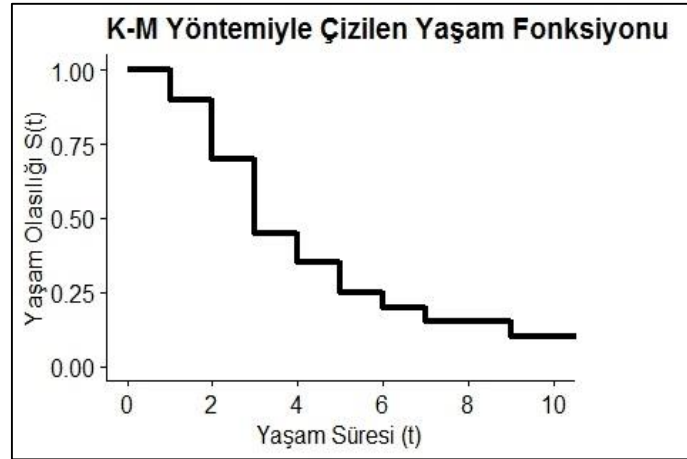
$$\hat{S}(t) = \begin{cases} 1 & \text{eğer } t < t_{(1)} \\ \prod_{t_{(j)} \leq t} \left[1 - \frac{d_j}{n_j}\right] & \text{eğer } t_{(1)} \leq t \end{cases} \quad (2.4)$$

Eşitlik 2.4'ten de elde edilebileceği gibi, ardışık zaman noktalarında hesaplanan yaşam olasılıkları arasında eşitlik 2.5'te görüldüğü gibi bir ilişki bulunmaktadır:

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \times P(T > t_{(j)} | T \geq t_{(j)}) \quad (2.5)$$

Burada $\hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} P(T > t_{(i)} | T \geq t_{(i)})$ 'dir ve koşullu olasılık kullanılarak hesaplanır.

K-M yöntemi kullanılarak hesaplanan yaşam olasılıkları ile çizilen yaşam fonksiyonu; sağa doğru ilerleyen, olayın görüldüğü zaman noktalarında sıçrama yapan basamaklı bir grafik görüntüsü sunar. Şekil 2.2'de bu duruma bir örnek gösterilmektedir (Kaplan ve Meier, 1958).



Şekil 2.2 Kaplan-Meier Yöntemiyle Çizilen Yaşam Fonksiyonu Grafiği

2.1.2. Hazard fonksiyonu (Hazard function)

Hazard fonksiyonu; t zamanına kadar ilgililenen olay ile karşılaşmamış bir birimin t zamanında ilgililenen olayı yaşama potansiyelini gösterir. Koşullu

olasılık yaklaşımı kullanılarak hesaplanan bir oran (rate)dir. Anlık potansiyel (instantaneous potential) limit yaklaşımı ile elde edilir. Birim zamanda bir bireyde görülebilecek beklenen olay sayısı olarak yorumlanır.

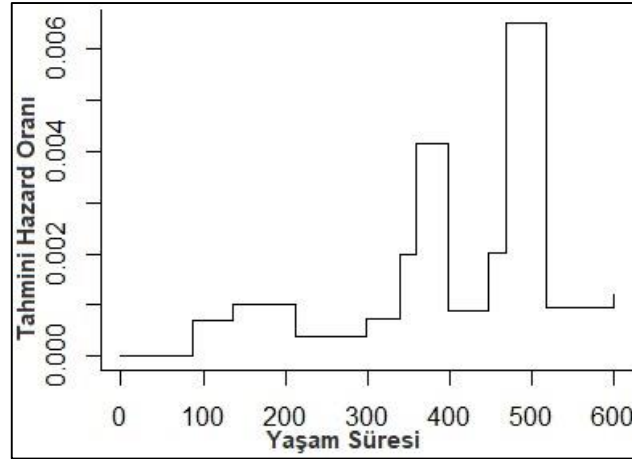
$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (2.6)$$

Hazard fonksiyonları, yaşam fonksiyonlarına göre ilgilenilen olayın altında yatan mekanizma (failure pattern) hakkında daha fazla bilgi vericidir. Bu nedenle veri setlerini özetlemede hazard fonksiyonları daha baskın bir yöntem olarak düşünülebilir. Yaşam analizinin kullanıldığı farklı uygulama alanlarında, hazard fonksiyonunun ismi değişiklik göstermektedir. Bu alanlara örnekler Tablo 2.1’de gösterilmektedir.

Tablo 2.1. Hazard fonksiyonunun diğer bilim dallarındaki isimleri

Nüfus Bilimi	Mortalite gücü (force of mortality)
Stokastik Süreç	Yoğunluk fonksiyonu (intensity function)
Epidemiyoloji	Yaşa bağlı başarısızlık (age-specific failure)
Güvenirlilik	Koşullu başarısızlık oranı (conditional failure rate)
Ekonomi	Mill oranının tersi (inverse of Mill’s ratio)

Hazard fonksiyonlarının, yaşam fonksiyonları gibi ortak seyreden bir grafik yapısı bulunmamaktadır. Aldığı değerler sıfır ile ∞ arasında değişir. Şekil 2.3’te görüldüğü gibi, bazen sabit bazen ise inişli çıkışlı bir grafik sergileyebilir.



Şekil 2.3 Hazard Fonksiyonu Grafiği

Hazard fonksiyonu ile yaşam sürelerinin modellenmesinde kullanılan diğer fonksiyonlar arasındaki matematiksel ilişki eşitlik 2.7, 2.8 ve 2.9’da verilmektedir.

$$h(t) = \frac{f(t)}{S(t)}, \quad (2.7)$$

$$h(t) = \frac{f(t)}{1-F(t)}, \quad (2.8)$$

$$S(t) = \exp\left\{-\int_0^t h(u)du\right\} \quad (2.9)$$

Yaşam fonksiyonunun kümülatif hazard fonksiyonu ile doğrudan bir ilişkisi bulunmaktadır. Bu doğrudan ilişki kullanılarak çarpımsal limit teorisi ile kümülatif hazard fonksiyonu tahmini yapılabilmektedir.

$$H(t) = -\ln[S(t)] \quad (2.10)$$

Çarpımsal limit teorisine alternatif olarak kullanılan ve küçük örneklem büyüklüklerinde daha iyi sonuçlar sunan bir tahminci, Nelson tarafından önerilmiştir (Nelson, 1972). Daha sonra Aalen, Nelson'ın önerisini geliştirerek eşitlik 2.11'de yer alan Nelson-Aalen tahmincisini elde etmiştir (Aalen, 1978).

$$\hat{H}(t) = \begin{cases} 0, & t \leq t_1 \\ \sum_{t_1 < t} \frac{d_i}{n_i}, & t_1 \leq t \end{cases} \quad (2.11)$$

2.2. Orantılı Hazardlar Modeli (Proportional Hazards Model)

Doğrudan hazard fonksiyonu üzerine odaklanan modeller ailesi Cox tarafından önerilmiştir (Cox, 1972). Bu ailenin en basit ve en çok kullanılan üyesi Orantılı Hazardlar Modeli (Proportional Hazards Model) dir. Bu model, genellikle sağlık bilimleri alanında kullanılmaktadır.

Orantılı Hazardlar Modelinin yaygın kullanımının temel sebeplerinden biri, modelin parametrik yöntemlerden farklı olarak, hayatta kalma sürelerini temsil etmek için bazı özel olasılık dağılımlarını kullanmayı gerektirmemesidir. Bu yüzden bu model yarı-parametrik (semi-parametrik) olarak sınıflandırılır. Bir ya da daha fazla sayıda açıklayıcı değişkenin yaşam süresi üzerindeki etkisi araştırılmak istendiğinde kullanılır. Cox Regresyon Modeli olarak da literatürde yer almaktadır (Cox, 1972).

Yaşam analizine ait bir veri setinde t ; yaşam süresini, Y_1, Y_2, \dots, Y_p açıklayıcı değişkenleri gösterebilir. Bu durumda Cox'a ait Orantılı Hazardlar Modeli eşitlik 2.12'deki gibi belirtilir:

$$h_i(t|\mathbf{Y}) = h_0(t) \exp(\sum_{i=1}^p \beta_i Y_i) \quad (2.12)$$

Bu eşitlik; t zamanında i birimine ait hazardın iki temel faktörün çarpımından oluştuğunu göstermektedir. Bunlar;

- $h_0(t)$ temel hazard fonksiyonudur. $h_i(t|\mathbf{Y})$ ' nin pozitif kalabilmesi için $h_0(t)$ negatif olmamalıdır. Tüm açıklayıcı değişkenlerin aynı anda sifıra eşit olduğundaki riski tanımlar. Cox'un önerdiği bu modelin yarı-parametrik olmasının nedeni, temel hazard fonksiyonunun belirtilmek zorunda olmamasından kaynaklanmaktadır.
- $\exp(\sum_{i=1}^p \beta_i Y_i)$ ise üstel yapıda bulunan p adet açıklayıcı değişkenin doğrusal kombinasyonudur. Yaşam süresi t 'yi içermez. Doğrusal kombinasyon değeri negatif çıksa bile, üstel fonksiyon sayesinde $h_i(t|\mathbf{Y})$ negatif çıkmamaktadır.

$Y^* = (Y_1^*, Y_2^*, \dots, Y_p^*)$ ve $Y = (Y_1, Y_2, \dots, Y_p)$ iki farklı birime ait değerler seti olsun. Bu iki birime ait hazard modelleri birbirine oranlansın. Her iki modelde de $h_0(t)$ temel hazard fonksiyonu bulunduğundan, orana ait hesaplamalar açıklayıcı değişkenler kullanılarak gerçekleştirilir. Elde edilen bu orana Hazard Oranı (Hazard Ratio; HR) denir.

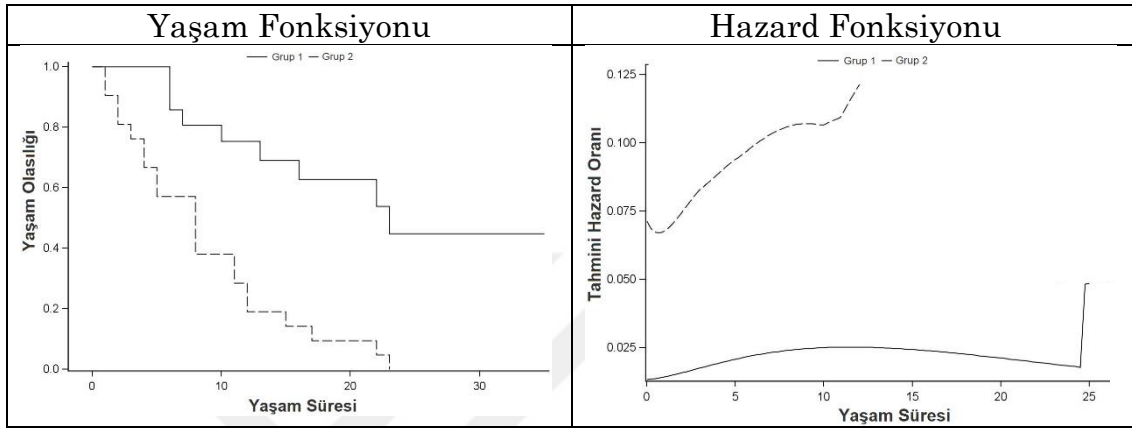
$$HR = \frac{h(t|\mathbf{Y}^*)}{h(t|\mathbf{Y})} \quad (2.13)$$

$$HR = \exp[\sum_{i=1}^p \beta_i (Y_i^* - Y_i)] \quad (2.14)$$

t zamanı boyunca HR sabittir. Eşitlik 2.13 ve 2.14'te de görüldüğü gibi HR değeri açıklayıcı değişkenlere bağlı olarak hesaplanır. Bu özelliği nedeni ile anılan model, Cox Orantılı Hazardlar modeli olarak adlandırılmaktadır. Ayrıca bu özellik modelin kullanımını da oldukça elverişli kılmaktadır.

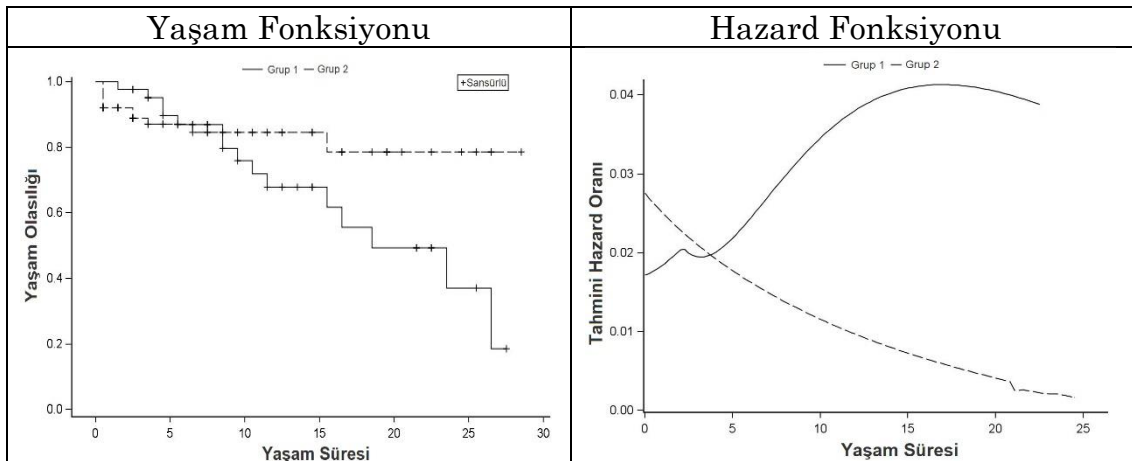
Orantılı hazardlar varsayımı, HR'nin tüm zaman boyunca sabit (ya da elde edilen sonuçların birbirine denk) kalması ile sağlanır. Bir birim için hesaplanan hazard değerinin, bir diğer birim için hesaplanan hazard değeri ile orantılı kalması ve bu oransallığın zamandan bağımsız olması gerekmektedir. Zamanın etkisinde kalan açıklayıcı değişkenler orantılı hazardlar varsayımının bozulmasına neden olur.

Yaşam analizinde, Orantılı Hazardlar varsayımının araştırılması gereken durumların başında, yaşam fonksiyonlarının karşılaştırılması gelmektedir. Şekil 2.4'te orantılı hazardlar varsayımının sağlandığı bir veri setine ait yaşam ve hazard fonksiyonlarının grafikleri görülmektedir. Hem yaşam fonksiyonları, hem de hazard fonksiyonları zaman içerisinde belirli bir uyum içerisinde ilerlemektedir.



Şekil 2.4 Orantılı Hazardlar Durumunda iki gruba ait yaşam ve hazard fonksiyonları

Orantılı Hazardlar varsayımını sağlamayan bir veri setine ait yaşam ve hazard fonksiyonları Şekil 2.5'te gösterilmektedir. Şekilde, yaşam fonksiyonlarının ve bunlara karşılık gelen hazard fonksiyonlarının kesiştiği görülmektedir. HR çalışma süresi boyunca sabit kalmamaktadır.



Şekil 2.5 Orantılı Hazardlar Durumu ihlalinde iki gruba ait yaşam ve hazard fonksiyonları

Üstel dağılıma ait yaşam süreleri olan iki grubun yaşam fonksiyonları kesişmemektedir. Çünkü her bir yaşam fonksiyonu hem ortalama hem de varyansı etkileyen tek bir parametre ile karakterize edilmektedir. Kesişen yaşam fonksiyonları ise iki parametre ile karakterize edilmektedir; bunlar

ortalama ve varyanstır. Bu ikisi içinden en büyük etken grupların farklı varyanslarının olmasıdır (Balakrishnan ve Rao, 2004).

İki yaşam fonksiyonu kesiştiği takdirde, bunlara karşılık gelen hazard fonksiyonları da kesişir, ancak hazard fonksiyonların kesiştiği her durumda bunlara karşılık gelen yaşam fonksiyonları kesişmeyebilir. Hatta hazard fonksiyonlarının da kesişmediği durumlarda orantılı hazardlar varsayımının ihlal edildiği görülebilir (Liu, Qiu & Sheng, 2007; Li vd., 2015). Bu nedenle orantılı hazardlar varsayımının sağlanıp sağlanmadığını önerilen yöntemlerle de incelemek faydalı olacaktır. Bunlar;

- 1) Grafikselle yaklaşım (log-log yaşam fonksiyonu grafikleri),
- 2) Uyum iyiliği testleri,
- 3) Zamana bağlı (time-dependent) açıklayıcı değişken varlığının araştırılmasıdır (Klein ve Moeschberger, 2005; Kleinbaum ve Klein, 2006; Hosmer, Lemeshow & May, 2011).

2.2.1. Grafikselle yaklaşım

Orantılı hazardlar varsayımını değerlendirmek için bilinen en yaygın yöntem log-log yaşam fonksiyonları grafiklerinin incelenmesidir.

Log-log yaşam fonksiyonları grafiği, tahmini yaşam olasılığı değerlerinin, $\hat{S}(t)$, iki defa doğal logaritmasının alınmasının ardından elde edilen değerler kullanılarak çizilen grafikleridir.

$$\log - \log \hat{S} = -\ln(-\ln \hat{S}) \quad (2.15)$$

İki defa logaritmik dönüşüm yapılmasının temel nedeni doğrusal bir yapı elde etmektir. Cox'un önerdiği orantılı hazardlar modeli kullanılarak Yaşam fonksiyonu eşitlik 2.17'deki gibi elde edilsin.

$$h(t|\mathbf{Y}) = h_0(t) \exp(\sum_{i=1}^p \beta_i Y_i) \quad (2.16)$$

$$S(t|\mathbf{Y}) = S_0(t) \exp(\sum_{i=1}^p \beta_i Y_i) \quad (2.17)$$

Log-log dönüşümü yapıldıktan sonra eşitlik 2.18'deki doğrusal bir denklem elde edilir. Bu denklem kullanılarak grafik çizilir.

$$-\ln[-\ln [S(t|\mathbf{Y})]] = -\sum_{i=1}^p \beta_i Y_i - \ln[-\ln [S_0(t)]] \quad (2.18)$$

Bu logaritmik dönüşüm sonucunda y eksenini için sınırlar farklı olacaktır. $\hat{S}(t)$ değerleri $[0,1]$ aralığında değişirken, $-\ln(-\ln \hat{S})$ dönüşüm verileri $(-\infty, \infty)$ aralığında değerler alabilir.

İki farklı bireye ait Y açıklayıcı değişkenler vektörleri $Y^* = (Y_1^*, Y_2^*, \dots, Y_p^*)$ ve $Y = (Y_1, Y_2, \dots, Y_p)$ olsun. Her iki vektöre ait log-log yaşam fonksiyonları eşitlik 2.19 ve eşitlik 2.20'deki gibi elde edilsin:

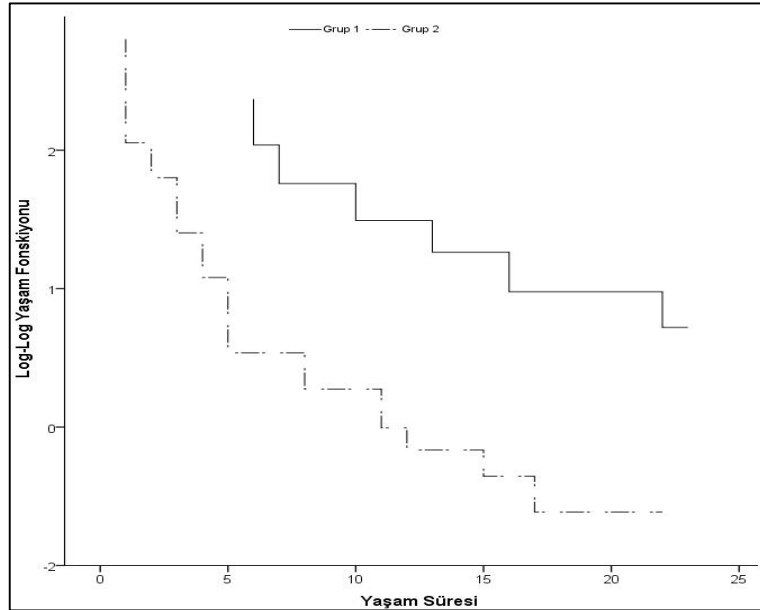
$$-\ln[-\ln [S(t|Y)]] = -\sum_{i=1}^p \beta_i Y_i - \ln[-\ln [S_0(t)]] \quad (2.19)$$

$$-\ln[-\ln [S(t|Y^*)]] = -\sum_{i=1}^p \beta_i Y_i^* - \ln[-\ln [S_0(t)]] \quad (2.20)$$

İki fonksiyonu birbirinden çıkarttığımızda, t içermeyen bir denklem eşitlik 2.21'deki gibi hesaplanmaktadır.

$$-\ln[-\ln [S(t|Y^*)]] - (-\ln[-\ln [S(t|Y)]]) = \sum_{i=1}^p \beta_i (Y_i - Y_i^*) \quad (2.21)$$

Bu durumda, her iki bireye ait log-log yaşam fonksiyonları aynı grafik ortamında çizildiğinde, iki grafiğin yaklaşık olarak birbirine paralel ilerlemesi ve aralarındaki açıklığın $\sum_{i=1}^p \beta_i (Y_i - Y_i^*)$ değerine eşit olması beklenir. Şekil 2.6'da görüldüğü gibi fonksiyonların paralel ilerlemesi, orantılı hazardlar varsayımının sağlandığını göstermektedir. Paralel bir görüntü izlenmez, ya da fonksiyonlar kesişir ise varsayımın sağlanmadığına karar verilir.



Şekil 2.6 Log-log yaşam fonksiyonları grafiği

Paralellik varsayımı, modelde yer alan her bir açıklayıcı değişken için ayrı ayrı araştırılır. K-M yöntemi kullanılarak her bir değişken için yaşam olasılıkları ayrı ayrı hesaplanabilir. Bazı durumlarda ise diğer açıklayıcı değişkenlere göre düzeltilmiş yaşam fonksiyonları kullanılabilir. Nitel değişkenler doğrudan kategorili olduğu için grafik çizimi daha kolaydır. Nicel verilerde ise, örneklem büyüklüğü göz önünde bulundurularak uygun sayıda alt kategorilere ayrılarak oluşturulur.

Log-log yaşam fonksiyonları kullanılarak orantılı hazardlar varsayımını araştırmak, öznel bir yöntemdir. Küçük örneklem büyüklüğü olan çalışmalarda daha dikkatli değerlendirme yapılmalıdır. Diğer yaklaşımlar da göz önünde bulundurularak sonuca karar verilmesi önerilmektedir (Klein ve Moeschberger, 2005; Kleinbaum ve Klein, 2006).

2.2.2. Uyum iyiliği testi yaklaşımı

Uyum iyiliği testi yaklaşımı grafiksel yaklaşımlara göre daha fazla tercih edilen bir yöntemdir. İstatistiksel bir testtir ve sonuca p değeri ile karar verilir. Öznel bir yaklaşım olmadığından sonuçları daha güveniliridir.

Literatürde birden fazla önerilmiş uyum iyiliği testi bulunmaktadır. Bunlar içerisinde en bilineni Harrell ve Lee tarafından önerilen testtir (Harrell ve Lee, 1986). Cox Modelinden elde edilen Schoenfeld artıkları kullanılarak geliştirilmiştir. Modelde bulunan her bir açıklayıcı değişken için ayrı ayrı Schoenfeld artıkları hesaplanır. Bu hesaplamalar ilgilenilen olay ile karşılaşmış birimler üzerinde gerçekleştirilir. Böylece, ilgilenilen olay ile karşılaşmış her bir birim için, modeldeki açıklayıcı değişken sayısı kadar Schoenfeld artığı hesaplanmış olur.

Bir birime ait Schoenfeld artığı; birimde gözlenen ilgili açıklayıcı değişken değerinden, bu değişkenin tüm birimlerinden hesaplanan ağırlıklı ortalama değerinin çıkarılması ile elde edilir. Bu hesaplama her bir birim için ilgilenilen olay ile karşılaştıkları t_j zamanında gerçekleştirilir. Burada kullanılan ağırlık fonksiyonu diğer birimlerin t_j zamanında hesaplanan hazard değerleridir.

Uyum iyiliği testinin geliştirilmesindeki temel mantık, Schoenfeld artığı hesaplanan açıklayıcı değişkenin yaşam süresi ile ilişkisi olup olmadığının

belirlenmesidir. İlişki bulunmadığı durumlarda Orantılı Hazardlar varsayımı sağlanmaktadır. Testin uygulanışı genellikle üç aşamadan oluşur. Bunlar;

- 1) Bir Cox modeli kurulur ve her bir açıklayıcı değişken için Schoenfeld artığı hesaplanır.
- 2) İlgilenilen olay ile karşılaşma zamanlarının sırasını gösteren yeni bir sıralayıcı değişken oluşturulur.
- 3) Birinci ve ikinci aşamada elde edilen değişkenlerin korelasyonuna bakılır. H_0 hipotezi aralarındaki korelasyonun sıfıra eşit olduğu yönündedir.

Kabul edilen H_0 hipotezi orantılı hazardlar varsayımının sağlandığını göstermektedir.

Grambsch ve Therneau tarafından önerilen ve genellikle paket programlarda yer alan bir diğer uyum iyiliği testi de benzer aşamalar kullanılarak hesaplanmaktadır (Grambsch ve Therneau, 1994). Bu uyum iyiliği testinin bir önceki testten farkı ilk aşamada ölçeklendirilmiş Schoenfeld artıklarını kullanmasıdır.

Uyum iyiliği testi, çok küçük ve çok büyük örneklem büyüklüğü olan çalışmalarda yanlış sonuçlar sunabilir. Bu nedenle sonuçlar değerlendirilirken, grafiksel yaklaşımın sonuçlarına da bakmak faydalı olacaktır (Kleinbaum ve Klein, 2006).

2.2.3. Zamana bağlı açıklayıcı değişken varlığının araştırılması

Açıklayıcı değişkenlerin, zamana bağlı bir değişken olup olmadığını değerlendirmek için Genişletilmiş (Extended) Cox Modeli kullanılır. Bu model, Cox Modeli üzerinde düzenleme yapılarak geliştirilmiştir. Açıklayıcı değişkenler tek başına ve zamana bağlı bir fonksiyon ile etkileşimli olacak şekilde modelde yer alır. Genişletilmiş Cox modeli genel formu eşitlik 2.22'de yer almaktadır.

$$h(t|Y) = h_0(t) \exp[\beta Y + \delta Y * g(t)] \quad (2.22)$$

Burada $g(t)$ zamana bağlı fonksiyonu göstermektedir. Basit bir zamana bağlı fonksiyon seçimi için $g(t) = t$ ya da $g(t) = \log(t)$ tercih edilebilir. Farklı $g(t)$ seçimi, farklı sonuçlar elde edilmesine neden olmaktadır.

Orantılı hazardlar varsayımı araştırılırken zamana bağlı fonksiyonun yer aldığı etkileşim teriminin katsayısı kullanılarak hipotez kurulur. δ Parametresinin sıfıra eşitliğini araştıran H_0 hipotezinin kabulü, orantılı hazardlar varsayımının sağlandığını göstermektedir.

$$H_0: \delta = 0 \quad (2.23)$$

Hipotezin değerlendirilmesinde Wald test istatistiği ya da Olabilirlik Oranı (Likelihood Ratio) istatistiği kullanılır. Bu test istatistikleri bir serbestlik dereceli ki kare dağılımına uygunluk gösterir.

$$LR \sim \chi^2_1 \quad (2.24)$$

Birden fazla açıklayıcı değişkeni eş zamanlı değerlendirmek için ise kurulacak model aşağıdaki gibidir:

$$h(t|Y) = h_0(t) \exp\left[\sum_{i=1}^p [\beta_i Y_i + \delta_i Y_i * g_i(t)]\right] \quad (2.25)$$

Bu durumda, kurulacak hipotez eşitlik 2.26 ve test istatistiği de eşitlik 2.27'deki gibi farklı olacaktır.

$$H_0: \delta_1 = \delta_2 = \dots = \delta_p = 0 \quad (2.26)$$

$$LR = -2\ln L_{\text{COX MODELİ}} - (-2\ln L_{\text{GENİŞLETİLMİŞ COX MODELİ}}) \sim \chi^2_p \quad (2.27)$$

En az bir parametrenin bile sıfırdan farklı olması, orantılı hazardlar varsayımının sağlanmamasına neden olacaktır (Klein ve Moeschberger, 2005; Kleinbaum ve Klein, 2006).

2.3. İki Yaşam Fonksiyonunun Karşılaştırılmasında Kullanılan Testler

2.3.1. Log-rank ve ağırlıklandırılmış log-rank testleri

Yaşam analizinde, bir çalışmada bulunan grupların yaşam sürelerinin karşılaştırılması önemli bir yer tutmaktadır. Farklı istatistiksel karşılaştırma testleri önerilmiştir. Bu testler genellikle yaşam ve hazard fonksiyonları

üzerine odaklanarak geliştirilmiştir. Bunlar içerisinde literatürde en yaygın olarak kullanılan log-rank testidir (Mantel, 1966).

Log-rank testi orantılı hazardlar varsayımı altında geliştirilmiştir ve bu koşulda elde edilen karşılaştırma sonuçlarının gücü oldukça yüksektir. Pek çok istatistik paket programında yer alan bir testtir. Bu nedenle log-rank testi oldukça popülerdir (Balakrishnan ve Rao, 2004).

Yaşam analizinde iki grubun yaşam sürelerini karşılaştırırken tüm testler için kullanılan sıfır hipotezi (H_0) ve alternatif hipotez (H_1) eşitlik 2.28 ve 2.29'da belirtilmektedir:

$$\begin{aligned} H_0: S_1(t) &= S_2(t), \forall t \leq \tau \\ H_1: S_1(t) &\neq S_2(t), \exists t \leq \tau \end{aligned} \quad (2.28)$$

Ya da;

$$\begin{aligned} H_0: h_1(t) &= h_2(t), \forall t \leq \tau \\ H_1: h_1(t) &\neq h_2(t), \exists t \leq \tau \end{aligned} \quad (2.29)$$

Burada τ ; en büyük gözlenen yaşam süresini göstermektedir. H_0 ; tüm zaman noktalarında iki yaşam fonksiyonu arasında fark olmadığını gösterirken, alternatif hipotez H_1 bazı zaman noktalarında fark olduğunu ifade eder.

İki grubun birleştirilmesi ile elde edilen veri setindeki yaşam süreleri göz önünde bulundurularak $t_1 < t_2 < \dots < t_r$ olmak üzere ilgilenilen olay ile karşılaşıldığı r tane birbirinden bağımsız zaman noktası elde edilsin.

Notasyonlarda $k=1,2$ ve $i=1, \dots, r$ olmak üzere t_i zamanında k örneğinde yer alan risk altındaki n_{ki} birimden d_{ki} tanesinde ilgilenilen olay gözlemlenmiş olsun. Birleştirilmiş örnekte t_i zamanında yer alan toplam ilgilenilen olay ile karşılaşanların sayısı $d_i = d_{1i} + d_{2i}$ ve risk altındaki kişi sayısı $n_i = n_{1i} + n_{2i}$ olsun.

Tablo 2.2 t_i zamanında iki grubun gösterge değişkene göre dağılımı

Gösterge Değişken	Grup 1	Grup 2	Toplam
1	d_{1i}	d_{2i}	d_i
0	$n_{1i} - d_{1i}$	$n_{2i} - d_{2i}$	$n_i - d_i$
Toplam	n_{1i}	n_{2i}	n_i

Her bir t_i zamanında iki grubun ilgilenilen olay ile karşılaşmış ve karşılaşmadığını gösteren Tablo 2.2'de görüldüğü gibi 2x2 çapraz tablolar oluşturulur. Eğer H_0 hipotezi doğru ise bu hipotez altında k . grubun beklenen hazard oranının tahmincisi; birleştirilmiş örnekten elde edilen hazard oranının tahmincisidir, d_i/n_i . Sadece k . örnekteki veri kullanılarak hazard oranının tahmini ise d_{ki}/n_{ki} dir. n_{1i} ; t_i zamanında birinci grupta bulunan risk altındaki birim sayısı ve d_{1i} ise ilgilenilen olay ile karşılaşan birim sayısı olsun.

$$Z_1 = \sum_{i=1}^r \left\{ d_{1i} - n_{1i} \left(\frac{d_i}{n_i} \right) \right\} \quad (2.30)$$

Eşitlik 2.30'da belirtilen test istatistiği Z_1 , birinci grupta ilgilenilen olay ile karşılaşan birim sayısı d_{1i} , ile H_0 altında beklenen olay sayısı $n_{1i}(d_i/n_i)$, arasındaki temel farkların toplamıdır. Geleneksel ki-kare uyum iyiliği testine benzemektedir. Eğer Z_1 değeri sıfıra yakın ise, H_0 hipotezinin yanlış olacağı yönünde bir sonuç bulma olasılığı düşecektir, ancak Z_1 sıfırdan çok uzak bir değer alırsa, H_0 hipotezi altında birinci grubun beklenenden daha farklı olacağı yönde bir bulgu taşıyacaktır.

H_0 hipotezinin değerlendirmesini tamamlamak için Z_1 istatistiğinin eşitlik 2.31'de belirtilen standart hatasına ihtiyaç duyulur.

$$\sigma_1 = \sum_{i=1}^r \left\{ \frac{n_{1i}}{n_i} \left(1 - \frac{n_{1i}}{n_i} \right) \left(\frac{n_i - d_i}{n_i - 1} \right) d_i \right\} \quad (2.31)$$

Mantel-Haenszel testi yaklaşımından faydalanılarak log-rank test istatistiği eşitlik 2.32'deki gibi elde edilir.

$$\chi^2_{log-rank} = \frac{\left[\sum_{i=1}^r \left\{ d_{1i} - n_{1i} \left(\frac{d_i}{n_i} \right) \right\} \right]^2}{\sum_{i=1}^r \left\{ \frac{n_{1i}}{n_i} \left(1 - \frac{n_{1i}}{n_i} \right) \left(\frac{n_i - d_i}{n_i - 1} \right) d_i \right\}} = \frac{Z_1^2}{\sigma_1} \quad (2.32)$$

H_0 hipotezi altında iki grup karşılaştırmasında kullanılan $\chi^2_{log-rank}$ test istatistiği bir serbestlik dereceli ki-kare dağılımına uygunluk gösterir. Test istatistiği, α anlamlılık düzeyinde ki-kare teorik tablo değerinden küçük ise $\left(\chi^2_{log-rank} \leq \chi^2_{1, 1-\alpha} \right)$ H_0 hipotezi kabul edilir.

Log-rank testi ikiden fazla grup karşılaştırılmasında da kullanılabilir. Eşitlik 2.30, 2.31 ve 2.32 matris formatında genelleştirilerek k ($k > 2$) sayıda grup

karşılaştırılabilir. Bu tez çalışmasında iki grup karşılaştırılmasına yer verildiği için bu konudan bahsedilmeyecektir.

Literatürde, araştırmacıların alternatif hipotezlerini daha iyi yansıtabilmek için log-rank testine farklı ağırlık fonksiyonları ekledikleri görülmektedir. Bu testlerin tümü ağırlıklandırılmış (weighted) log-rank testleri olarak adlandırılmaktadır. Ağırlıklandırılmış log-rank testlerine ait genel test istatistiği eşitlik 2.33'te belirtilmektedir.

$$\chi^2_{\text{ağırlıklandırılmış log-rank}} = \frac{\left[\sum_{i=1}^r w_i \left\{ d_{1i} - n_{1i} \left(\frac{d_i}{n_i} \right) \right\} \right]^2}{\sum_{i=1}^r w_i^2 \left\{ \frac{n_{1i}}{n_i} \left(1 - \frac{n_{1i}}{n_i} \right) \left(\frac{n_i - d_i}{n_i - 1} \right) d_i \right\}} \quad (2.33)$$

Burada w_i ; ağırlık fonksiyonunu göstermektedir. Ağırlık fonksiyonuna göre her bir test farklı isimler taşımaktadır. Bu testler ve ağırlık fonksiyonları Tablo 2.3'te gösterilmektedir.

Tablo 2.3 Ağırlık fonksiyonlarına göre ağırlıklandırılmış log-rank testleri

w_i	Test
1	Log-rank
n_i	Gehan-Wilcoxon
$\sqrt{n_i}$	Tarone-Ware
$\tilde{S}(t_i)$	Peto-Peto
$\tilde{S}(t_i) \frac{n_i}{n_i + 1}$	Modifiye Edilmiş Peto-Peto
$[\hat{S}(t_i)]^p [1 - \hat{S}(t_i)]^q$	Fleming-Harrington (p,q)

Gehan-Wilcoxon testi ağırlık olarak t_i zamanında risk altındaki toplam kişi sayısını (n_i) önerirken (Gehan, 1965), benzer bir yaklaşım ile Tarone-Ware testi bu ağırlık değerinin karekökü alınmış halinin ($\sqrt{n_i}$) kullanımını önermektedir (Tarone ve Ware, 1977). Her iki yöntem de genellikle çalışmanın başlangıcında gözlenen ve beklenen olay sayısının arasındaki farka daha fazla ağırlık vermektedir. Çünkü en fazla kişi sayısı genellikle çalışmanın başlangıcında görülmektedir.

Peto-Peto testi özellikle sansürlü veri için geliştirilen alternatif bir başka yaklaşımdır (Peto ve Peto, 1972). Bu testte ağırlık olarak genel bir yaşam fonksiyonu eşitlik 2.34'deki gibi tanımlanmıştır:

$$\tilde{S}(t_i) = \prod_{t_i \leq t} \left\{ 1 - \frac{d_i}{n_{i+1}} \right\} \quad (2.34)$$

Peto-Peto testinin sansürlü birim oranının fazla olduğu çalışmalarda Gehan-Wilcoxon ve Tarone-Ware testlerinden daha tutarlı sonuçlar verdiği gözlemlenmiştir (Balakrishnan ve Rao, 2004; Kassambara, Kosinski, Biecek & Fabian, 2017). Andersen tarafından bu teste farklı bir düzeltme terimi eklenerek yeni bir ağırlık fonksiyonu elde edilmiştir. Bu ağırlık fonksiyonu kullanılarak elde edilen teste ise Modifiye edilmiş Peto-Peto denilmektedir (Andersen ve Gill, 1982; Andersen vd., 1992). Fleming ve Harrington, özel durumlar olarak ele alınan testlerin çoğunu içeren çok genel bir test sınıfı önermektedir (Harrington ve Fleming, 1982). Genelleştirilmiş Fleming-Harrington testi, $\hat{S}(t)$ 'yi birleşik örneklemin yaşam fonksiyonu olarak tanımlayarak ağırlık fonksiyonunu eşitlik 2.35' teki gibi önermektedir:

$$W_{p,q}(t_i) = \hat{S}(t_i)^p [1 - \hat{S}(t_i)]^q, \quad p \geq 0, q \geq 0 \quad (2.35)$$

Fleming-Harrington testinin ağırlık fonksiyonu $p=q=0$ değerlerini aldığıında log-rank testine denktir. Eğer $q=0$ ve $p>0$ ise bu test başlangıçta oluşan farklılıklara ağırlık verirken tam tersi durum olan $p=0$ ve $q>0$ ise ilerleyen zamanlarda oluşan farklılıklara ağırlık vermektedir. Esnek bir yöntem olmasına karşın p ve q değerlerinin nasıl tespit edileceği konusunda net bir tutum bulunmamaktadır (Balakrishnan ve Rao, 2004).

2.3.2. Parçalanmış log-rank testleri (Partitioned log-rank tests)

Yaşam ya da hazard fonksiyonlarının kesişmesi ile ortaya çıkan orantılı hazardlar varsayımı ihlalinde yaşam ya da hazard fonksiyonlarını karşılaştırmak için farklı yaklaşımlar önerilmiştir. Bunlar içerisinde en güncel olanlardan biri Yukun Liu ve Guosheng Yin tarafından geliştirilen parçalanmış log-rank testleridir (Liu ve Yin, 2016). Yaşam sürelerinin küçükten büyüğe sıralandığı bir ekseni belirli noktalardan ikiye ayırdıktan sonra, log-rank ve ağırlıklandırılmış log-rank testlerini kullanarak yeni bir yöntem önermektedirler.

$\{(T_{kj}, C_{kj}) : j = 1, \dots, n_k\}$ olmak üzere T_{kj} ; k ($k=1,2$) grubundaki ilgilenilen olayın gözleendiği yaşam sürelerini, C_{kj} ; ise eşlik eden sansürlü yaşam sürelerini göstermektedir.

Gözlenen yaşam sürelerine ait veri setini tekrardan yazmak gerekirse; $\{(X_{kj}, \delta_{kj}) : j = 1, \dots, n_k\}$; burada $X_{kj} = \min\{T_{kj}, C_{kj}\}$ ve $\delta_{kj} = 1\{T_{kj} \leq C_{kj}\}$ gösterge fonksiyonudur.

$S_k(t)$ ve $h_k(t)$ k grubuna ait yaşam ve hazard fonksiyonları olsun. Buradaki amaç, iki grubun hazard fonksiyonlarının ya da yaşam fonksiyonlarının eşitliğinin test edilmesidir.

$$\begin{aligned} H_0: S_1(t) = S_2(t) \text{ (ya da } h_1(t) = h_2(t)) ; \forall t \in [0, \infty) \\ H_1: S_1(t) \neq S_2(t) \text{ (ya da } h_1(t) \neq h_2(t)) ; \exists t \in [0, \infty) \end{aligned} \quad (2.36)$$

İki gruba ait birleştirilmiş veri setinde r birbirinden bağımsız olay zamanları olsun ve $t_1 < t_2 < \dots < t_r$ olarak gösterilsin. Her bir olay zamanı t_i için d_{ki} ; k grubunda ilgilenilen olay ile karşılaşanların sayısını ve n_{ki} ; ise k grubunda risk altında kalan kişileri gösterebilir. $d_i = d_{1i} + d_{2i}$ ve $n_i = n_{1i} + n_{2i}$ olmak üzere genellikle kullanılan ağırlıklandırılmış log-rank formülü eşitlik 2.37'de ifade edilmektedir:

$$\chi^2_{\text{ağırlıklandırılmış}} = \left\{ \sum_{i=1}^r w_i \left(d_{1i} - \frac{d_i n_{1i}}{n_i} \right) \right\}^2 / \sum_{i=1}^r \frac{w_i^2 n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} \quad (2.37)$$

Burada w_i ; önceden tanımlanan ağırlıklardır.

Parçalanmış log-rank testleri geliştirilirken Liu ve Yin tarafından kullanılan ağırlıklar ve testler; $w_i = 1$ ile bilinen orijinal log-rank, $w_i = n_i$ ile Gehan-Wilcoxon ve $w_i = \hat{S}(t_i)$ ile Peto – Peto'dur (Liu ve Yin, 2016)

Eğer iki hazard fonksiyonu kesişir ise $\chi^2_{\text{ağırlıklandırılmış}}$ önemli bir güç kaybı yaşar. Çünkü pay kısmında yer alan $\sum_{i=1}^r w_i (d_{1i} - d_i n_{1i}/n_i)$ formülünde bozulan bazı noktalar bulunmaktadır. Bu çalışmada önerilen testler ile yaşam süreleri eksenini belirli parçalara ayırarak $\sum_{i=1}^r w_i (d_{1i} - d_i n_{1i}/n_i)$ formülünden kaynaklanan sorun çözümlenmeye çalışılmıştır.

Liu ve Yin, $H_0: S_1(t) = S_2(t)$ (ya da $h_1(t) = h_2(t)$); $\forall t \in [0, \infty)$ yerine 2 farklı hipotezi eş zamanlı çözmeye çalışmaktadırlar. Bu hipotezler eşitlik 2.38 ve 2.39'da belirtildiği gibidir.

$$H_{01}: S_1(t) = S_2(t) \text{ (ya da } h_1(t) = h_2(t)) ; \forall t \in [0, u) \quad (2.38)$$

$$H_{02}: S_1(t) = S_2(t) \text{ (ya da } h_1(t) = h_2(t)) ; \forall t \in [u, \infty) \quad (2.39)$$

Burada u parçalanma noktasını göstermektedir ($u \in [0, \infty)$). H_{01} ve H_{02} 'e uygulanan testlerin toplamı, H_0 'ı değerlendirmede kullanılacak yeni bir yaklaşım olacaktır. Belirli bir t noktasından bölünen alt ve üst bileşenler ayrı ayrı tanımlanmaktadır.

Pay kısmında yer alan farklar bileşenleri eşitlik 2.40 ve 2.41'de gösterilmektedir.

$$D_l(t) = \sum_{i=1}^r w_i \left(d_{1i} - \frac{d_i n_{1i}}{n_i} \right) 1(t_i < t) \quad (2.40)$$

$$D_u(t) = \sum_{i=1}^r w_i \left(d_{1i} - \frac{d_i n_{1i}}{n_i} \right) 1(t_i \geq t) \quad (2.41)$$

Bu farklara eşlik eden varyans bileşenleri ise eşitlik 2.42 ve 2.43'te tanımlanmaktadır.

$$V_l(t) = \sum_{i=1}^r \frac{w_i^2 n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} 1(t_i < t) \quad (2.42)$$

$$V_u(t) = \sum_{i=1}^r \frac{w_i^2 n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} 1(t_i \geq t) \quad (2.43)$$

Belirli bir t yaşam süresi noktasında H_0 'ı değerlendirecek olan test istatistiği eşitlik 2.44'teki formül ile hesaplanmaktadır:

$$\tilde{T}(t) = \frac{D_l^2(t)}{V_l(t)} + \frac{D_u^2(t)}{V_u(t)} \quad (2.44)$$

$t = t_1$ ya da $t = t_r$ olduğunda $\tilde{T}(t)$ istatistiği bilinen ağırlıklandırılmış log-rank istatistiğine eşit olur. Ağırlıklandırılmış log-rank testleri ile karşılaştırıldığında $\tilde{T}(t)$ 'e bağlı bu test istatistiğinin hazard fonksiyonları kesişmediğinde az bir miktar güç kaybettiği belirlenmiştir. Öte yandan iki hazard fonksiyonu bir kere kesiştiğinde ve bu kesişim belirli bir t zamanında olduysa da bu test, diğer testlere göre çok daha fazla güç kazanmaktadır. Pratikte iki fonksiyonun kesiştiği noktayı tam olarak belirlemek kolay değildir.

Ayrıca hazard fonksiyonlarının iki ya da daha fazla defa kesiştiği durumlar karşılaştırılırken testin performansı kesin değildir. Güç kaybını azaltmak ve t noktasının keyfi seçimini engellemek için H_0 'ı test edecek istatistik eşitlik 2.45'teki gibi elde edilir.

$$T = \sup_{t \in [0, \infty)} \tilde{T}(t) = \max_{1 \leq i \leq r} \tilde{T}(t_i) \quad (2.45)$$

Bu test istatistiğine parçalanmış log-rank testi istatistiği denir. Farklı ağırlık seçimleri ile farklı parçalanmış log-rank testleri elde edilir .

Elde edilen T test istatistiğinin değerlendirilmesi ve H_0 'ın kabul ya da red edilmesine karar verilmesi için teorik bir dağılıma ihtiyaç duyulmaktadır. Liu ve Yin parçalanmış log-rank testlerinin test istatistiği T 'nin limit dağılımının yakınsamasını Brownian motion $\{B(t)\}$ 'ı kullanarak açıklamaya çalışmaktadırlar.

$$T \rightarrow \sup_{0 < t < 1} [\{B(t)\}^2/t + \{B(1) - B(t)\}^2/(1 - t)] \quad (2.46)$$

Brownian motion sürekli yaşam verilerini kullanarak hesaplanan stokastik bir süreçtir (Gill, 1980; Revuz ve Yor, 1999; O'Quigley, 2012). Eşitlik 2.46'dan da görüldüğü gibi herhangi bir örneklem büyüklüğü ya da sansür mekanizmasına bağlı değildir. Brownian motion kullanılarak elde edilen T 'nin bu limit dağılımının oldukça karmaşık olduğu ve bu yapı kullanılarak p değerinin hesaplanmasının çok kolay olamayacağı çalışmalarında belirtilmektedir. Bu nedenle Liu ve Yin herhangi bir dağılımdan veri türeterek ve simülasyonlar kullanılarak p değerine karar verileceğini düşünmektedirler. Ancak gerçekleştirdikleri simülasyon çalışmalarında T 'nin dağılımının limit dağılımına yakınsamasının oldukça zayıf olduğunu göstermektedirler. Özellikle bu yakınsamanın küçük ve orta düzeyde örneklem büyüklüklerinde çok yavaş olduğunu belirtmektedirler. Bunun yerine T istatistiğine ait p değerini bootstrap yöntemi (Efron ve Tibshirani, 1986; Efron, 1992) ile elde etmeye karar vermişlerdir. Bu amaçla aşağıdaki algoritmayı kullanmışlardır.

- 1) T orijinal veri setine ait iki grubu karşılaştırmada elde edilen parçalanmış log-rank test istatistiği olsun $\{(X_{kj}, \delta_{kj}): j = 1, \dots, n_k\} (k = 1, 2)$.

2) $\{(X_j, \delta_j): j = 1, \dots, n\}$ birleştirilmiş veri setini gösterebilir. $\{(X^*_{kj}, \delta^*_{kj}): j = 1, \dots, n_k\}$ ($k = 1, 2$) n_1 ve n_2 örneklem büyüklüğü olacak şekilde birleştirilmiş veri setinden iadeli olarak çekilmiş iki rasgele bootstrap örneği elde edilebilir. Bu iki grubun karşılaştırılmasından elde edilen parçalanmış log-rank istatistiği ise T^* ile gösterilebilir.

3) 2. Aşama büyük bir sayı, B kadar tekrarlanabilir. B tane elde edilen test istatistiği ise T^*_1, \dots, T^*_B ile gösterilebilir. İki farklı yoldan hipotez testini sonuçlandıracak p değeri elde edilebilir. α anlamlılık düzeyinde bu hesaplamalar şu şekilde tanımlanmaktadır:

- a) T^* 'den büyük olan T^* 'lerin oranı p değeridir. $P < \alpha$ ise H_0 reddedilir.
- b) c_α ; B x en büyük T^*_i olsun. $T > c_\alpha$ ise H_0 reddedilir.

Liu ve Yin (2016) çalışmalarında a) maddesindeki uygulamanın b) maddesine göre daha etkin sonuçlar verdiği göstermişlerdir. Yazarlar simülasyon çalışmaları sonucunda a) maddesinin kullanımını önermektedir. Bu nedenle bu tez çalışmasında a) maddesi kullanılarak p değeri hesaplanmaktadır.

2.3.3. Ağırlıklandırılmış Lin ve Wang testleri (Weighted Lin and Wang tests)

Lin ve Wang (2004) iki grup için kullanılan log-rank test istatistiği üzerinde düzenleme yaparak, kesişen yaşam fonksiyonu alternatifinde kullanmak üzere yeni bir yöntem önermişlerdir. Bir simülasyon çalışması sonucunda, düzenlenen bu yeni yöntemin iki grup karşılaştırmasında yaygın olarak kullanılan log-rank ve Gehan-Wilcoxon testlerinden daha büyük bir güce sahip olduğunu göstermişlerdir.

Koziol ve Jia (2014) yaptıkları çalışmada Lin ve Wang tarafından önerilen bu yönteme ağırlıklandırma fonksiyonu ekleyerek ağırlıklandırılmış Lin ve Wang testlerini elde edip, literatüre öneride bulunmuşlardır.

Önerilen testlerin değerlendirilmesinde kullandıkları hipotezler şu şekildedir:

H_0 : İki grubun hayatta kalma dağılımları küresel alternatif hipoteze karşın aynıdır.

H_1 : İki grubun hayatta kalma dağılımları aynı değildir.

Ağırlıklandırılmış Lin ve Wang testleri, Lin ve Wang testi ile aynı varsayımları paylaşmaktadır. Bu varsayımlar maddeler halinde aşağıda sıralanmıştır:

- İlgilenilen olay gerçekleşene kadar geçen süreler birbirinden bağımsızdır.
- Her iki grupta sansürlü gözlem bulunuyor ise bu gözlemlerin dağılımı birbirinden bağımsızdır.
- İlgilenilen olay sayısı ve farklı yaşam süresi değerlerine sahip olay sayısı oldukça fazladır.
- Ağırlık fonksiyonları (w_i), $0 < w_i < A$ ve $A < \infty$ olacak şekilde pozitif ve üstten sınırlanmış değerler alır.

$t_1 < t_2 < \dots < t_r$, r tane farklı zaman noktasını, k ; ($k=1,2$) grup değerini d_{ki} ; t_i zamanında k grubunda ilgilenilen olay ile karşılaşan birey sayısını ve $E(d_{ki})$ ise beklenen olay sayısını gösterebilir. Lin ve Wang, log-rank test istatistiği üzerinde düzenleme yaparak karesel bir test istatistiğini eşitlik 2.47'deki gibi önermişlerdir:

$$\Delta = \sum_{i=1}^r [d_{1i} - E(d_{1i})]^2 \quad (2.47)$$

Burada Δ ; iki yaşam fonksiyonu arasındaki uzaklığın karesini göstermektedir, böylece her iki yönde de oluşabilecek farklılara duyarlı olmaktadır. Koziol ve Jia H_0 Hipotezine ait çıkarılamayı yaparken Δ 'nın standardize edilmiş halini ve standart normal dağılımı kullanmışlardır. Çalışmalarında Δ 'nın ağırlıklandırılmış durumunu eşitlik 2.48'deki gibi tanımlamışlardır:

$$\Delta_w = \sum_{i=1}^r w_i * [d_{1i} - E(d_{1i})]^2 \quad (2.48)$$

Koziol ve Jia (2014)'nin kullandığı ağırlık fonksiyonları ve isimleri Tablo 2.4'te verilmektedir.

Tablo 2.4 Ağırlıklandırılmış Lin ve Wang testlerine ait ağırlık fonksiyonları

w_i	Test
1	Lin ve Wang (LW)
n_i	LW _{w1}
$\sqrt{n_i}$	LW _{w2}
$1/\sqrt{Var(d_{1i})}$	LW _{w3}

Ağırlıklandırılmış Lin ve Wang yöntemine ait test istatistiği eşitlik 2.49'da gösterilmektedir.

$$T_w = \frac{\Delta_w - E(\Delta_w)}{\sqrt{Var(\Delta_w)}} \quad (2.49)$$

$E(\Delta_w)$ ve $Var(\Delta_w)$, d_{1i} 'nin marjinal hipergeometrik dağılımı kullanılarak hesaplanmaktadır. Bu terimlerin daha detaylı olarak hesaplanmaları eşitlik 2.50 ile 2.58 aralığındaki formüllerde gösterilmektedir.

$$E(\Delta_w) = \sum_{i=1}^r w_i * \frac{n_{1i}n_{2i}d_i(n_i-d_i)}{n_i^2(n_i-1)} \quad (2.50)$$

$$Var(\Delta_w) = \sum_{i=1}^r w_i^2 \{E(d_{1i}^4) - 4E(d_{1i}^3)E(d_{1i}) + 6E(d_{1i}^2)[E(d_{1i})]^2 - 3[E(d_{1i})]^4 - [Var(d_{1i})]^2\} \quad (2.51)$$

Burada d_{1i} 'nin ham momentleri, faktöriyel momentler formülünden de hesaplanabilir:

$$E(d_{1i}^{(m)}) = \frac{n_{1i}^{(m)} d_i^{(m)}}{n_i^{(m)}} \quad (2.52)$$

$$n^{(m)} = n * (n - 1) * \dots * (n - m + 1) \quad (2.53)$$

$$E(d_{1i}) = \frac{n_{1i}d_i}{n_i} \quad (2.54)$$

$$Var(d_{1i}) = \frac{n_{1i}n_{2i}d_i(n_i-d_i)}{n_i^2(n_i-1)} \quad (2.55)$$

$$E(d_{1i}^2) = Var(d_{1i}) + [E(d_{1i})]^2 \quad (2.56)$$

$$E(d_{1i}^3) = 3E(d_{1i}^2) - 2E(d_{1i}) + \frac{n_{1i}^{(3)}d_i^{(3)}}{n_i^{(3)}} \quad (2.57)$$

$$E(d_{1i}^4) = 6E(d_{1i}^3) - 11E(d_{1i}^2) + 6E(d_{1i}) + \frac{n_{1i}^{(4)}d_i^{(4)}}{n_i^{(4)}} \quad (2.58)$$

Δ_w İstatistiğinin beklenen değerinden farkı alınıp, $\sqrt{Var(\Delta_w)}$ değerine bölünmesiyle standardize edilmiş bir yapı oluşmaktadır. Yukarıdaki varsayımlar da sağlandığında test istatistiği T_w eşitlik 2.59'da da görüldüğü üzere standart normal dağılıma uygunluk gösterir.

$$T_w \sim N(0,1) \quad (2.59)$$

Bu nedenle H_0 hipotezinin değerlendirilmesinde kullanılacak p değeri standart normal dağılımdan elde edilmektedir.



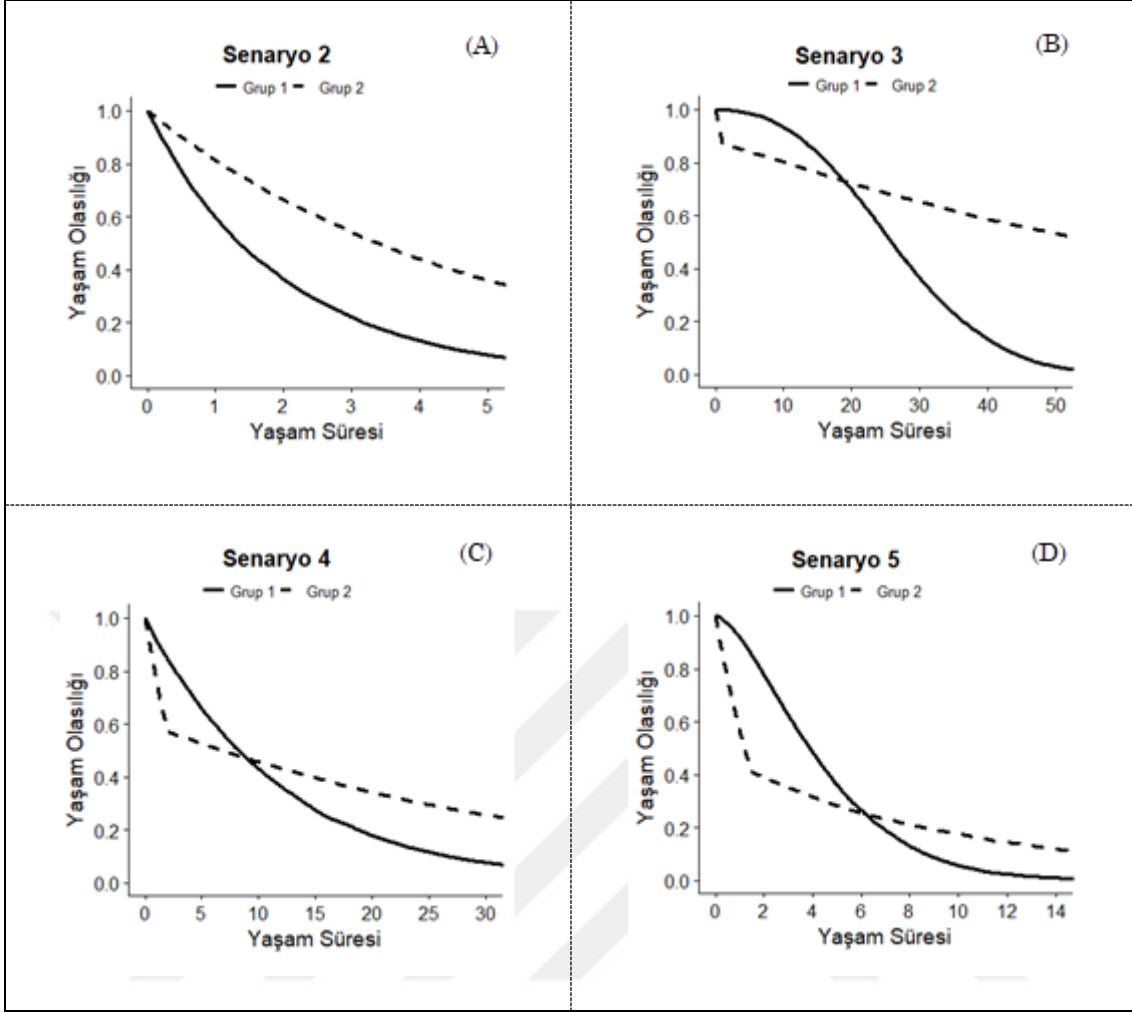
3. GEREÇ VE YÖNTEMLER

Bu başlık altında tez çalışmasında yer alan simülasyon çalışmasının aşamaları açıklanmıştır. Öncelikle veri türetim senaryoları belirlenmiş, ardından kullanılan testlere ait algoritmalar oluşturulmuştur. Bu aşamalar sonucunda tez çalışması boyunca kullanılacak genel simülasyon algoritmasının alt elemanları elde edilmiştir. Son olarak da genel simülasyon algoritması ve bu algoritmayı şekillendiren temel elemanlar detaylı bir şekilde açıklanmıştır.

Bu tez çalışması boyunca kullanılan kodlar R Studio v 1.1.463 ve R v 3.5.3 paket programları ve R Programlama Dili kullanılarak oluşturulmuştur. R Programı içerisinde yer alan dplyr (Wickham ve Wickham, 2019), survival (Therneau ve Lumley, 2014), survminer (Kassambara vd., 2017), survMisc (Dardis ve Dardis, 2018) paketleri kullanılmıştır.

3.1. Veri Türetim Senaryoları ve R Paket Programında Algoritmaların Oluşturulması

Veri türetim işlemleri beş farklı senaryo göz önünde bulundurularak yapılmıştır. Öncelikle testlerin birinci tip hata oranları incelenmiştir (Senaryo 1) (Liu ve Yin, 2016). Daha sonra orantılı hazardlar varsayımının sağlandığı bir senaryo (Senaryo 2) ve varsayımın ihlal edildiği üç senaryo (Senaryo 3, Senaryo 4 ve Senaryo 5) kullanılarak testlerin farkı belirleme gücü karşılaştırılmıştır. Her bir senaryo için farklı parametrelili teorik dağılımlardan faydalanılmıştır (Li vd., 2015). Sansürlü gözlemler rasgele bir yapıda olacak şekilde $U(0,1)$ parametrelili tekdüze dağılım ile oluşturulmuştur (Lin ve Wang, 2004). Testlerin gücünün karşılaştırıldığı senaryolara ait yaşam fonksiyonu grafikleri Şekil 3.1'de gösterilmiştir.



Şekil 3.1 Simülasyon çalışmalarında kullanılan senaryolar

Senaryo 1: Tip 1 Hata Oranlarının İncelenmesi

Testlerin tip 1 hata oranlarının belirlenmesi için birbirinden bağımsız olacak şekilde her iki grup için üstel dağılım kullanılarak yaşam süresi (t) veri üretimi yapılmıştır. Veri üretiminde her iki grup için de $\lambda = 1$ parametresi tercih edilmiştir.

$$\text{Grup 1} : t_1 \sim \text{Üstel}(\lambda = 1)$$

$$\text{Grup 2} : t_2 \sim \text{Üstel}(\lambda = 1) \quad (3.1)$$

Senaryo 1'e ait veri üretimi kodlaması aşağıdaki gibi yapılmıştır:

```

1. # Senaryo 1: Tip 1 Hata Oranlarının İncelenmesi
2. # n1 ve n2 gruplara ait örneklem büyüklüğü
3. # c1 ve c2 gruplara ait sansür oranı
4. # d1 ve d2 gruplara ait gösterge değişken
5.
6. veri_turetim1=function (n1, n2, c1, c2){
7.
8.   stime1=rexp(n1,rate=1)

```

```

9. stime2=rexp(n2,rate=1)
10.
11. ctime1=runif(n1,0,1)
12. ctime2=runif(n2,0,1)
13.
14. d1=0;d2=0
15.
16. for (j in 1:n1) {if(ctime1[j]<c1) d1[j]=0 else d1[j]=1}
17. for (j in 1:n2) {if(ctime2[j]<c2) d2[j]=0 else d2[j]=1}
18.
19. stime=c(stime1,stime2)
20. d=c(d1,d2)
21. g1=rep(1,n1);g2=rep(2,n2)
22. group=c(g1,g2)
23.
24. survdata<-data.frame(stime=stime, d=d,group=group)
25. }

```

Senaryo 2: Orantılı Hazardlar Varsayımının Sağlandığı Durum

Bu senaryo orantılı hazardlar varsayımını sağlayan birbirinden bağımsız iki grubun yaşam süreleri karşılaştırıldığında testlerin gücünü incelemek için oluşturulmuştur. Birinci grup için $\lambda = 0.5$ ve ikinci grup için $\lambda = 0.2$ olacak şekilde üstel dağılım kullanılarak veri türetimi yapılmıştır. Bu işlemler sonucunda Şekil 3.1-A'daki durum gözlemlenmiştir.

$$\text{Grup 1 } t_1 \sim \text{Üstel}(\lambda = 0.5)$$

$$\text{Grup 2 } t_2 \sim \text{Üstel}(\lambda = 0.2) \quad (3.2)$$

Bu senaryoya ait veri türetimi kodlaması aşağıda yer almaktadır:

```

1. # Senaryo 2: Orantılı Hazardlar Varsayımı Sağlandığında
2.
3. veri_turetim2=function (n1, n2, c1, c2){
4.
5. stime1=rexp(n1,rate=0.5)
6. stime2=rexp(n2,rate=0.2)
7.
8. ctime1=runif(n1,0,1)
9. ctime2=runif(n2,0,1)
10.
11. d1=0;d2=0
12.
13. for (j in 1:n1) {if(ctime1[j]<c1) d1[j]=0 else d1[j]=1}
14. for (j in 1:n2) {if(ctime2[j]<c2) d2[j]=0 else d2[j]=1}
15.
16.
17. stime=c(stime1,stime2)
18. d=c(d1,d2)
19. g1=rep(1,n1);g2=rep(2,n2)
20. group=c(g1,g2)
21.
22. survdata<-data.frame(stime=stime, d=d,group=group)
23. }

```

Senaryo 3: Başlangıçta Kesişen İki Yaşam Fonksiyonu Durumu

Orantılı hazardlar varsayımının ihlal edildiği ilk senaryoda, iki yaşam fonksiyonunun çalışma başlangıcında kesiştiği varsayımı üzerinde durulmuştur. Burada hedeflenen durum her iki yaşam fonksiyonunun da olasılık değerinin %60'dan büyük olduğu ($S(t) > 0.6$) durumda kesişmesidir. Birinci grupta veri türetimi yapılırken şekil parametresi $\alpha = 2.5$ ve ölçek parametresi $\beta = 30$ olan Weibull ($W(\alpha, \beta)$) dağılımından faydalanılmıştır. İkinci grupta ise kesişen yapı oluşturmak için veri türetimi iki aşamalı olarak gerçekleştirilmiştir. Öncelikle $\lambda = 0.125$ parametrelili üstel dağılımdan veri türetimi yapılmıştır. Daha sonra yaşam süresi 1 ve üzeri olan gözlemler yerine tekrardan $\lambda = 0.01$ parametrelili üstel dağılımdan veri türetilmiştir. Bu durumun grafiksel hali Şekil 3.1-B'de gösterilmiştir.

$$\text{Grup 1 : } t_1 \sim W(2.5, 30)$$

$$\text{Grup 2 : } t_2 \sim \begin{cases} \text{Üstel}(\lambda = 0.125), t_2 < 1 \\ \text{Üstel}(\lambda = 0.01), t_2 \geq 1 \end{cases} \quad (3.3)$$

Senaryo 3'e ait veri türetimi kodlaması aşağıdaki gibidir:

```
1. # Senaryo 3: Başlangıçta Kesişen İki Yaşam Fonksiyonu
2.
3. veri_turetim3=function (n1, n2, c1,c2){
4.
5.   stime1=rweibull(n1,shape=2.5,scale=30)
6.   stime2=rexp(n2,rate=0.125)
7.
8.   for (m in 1:n2) {if(stime2[m]>=1) stime2[m]=rexp(1,rate=0.01)}
9.
10.  ctime1=runif(n1,0,1)
11.  ctime2=runif(n2,0,1)
12.
13.  d1=0;d2=0
14.
15.  for (j in 1:n1) {if(ctime1[j]<c1) d1[j]=0 else d1[j]=1}
16.  for (j in 1:n2) {if(ctime2[j]<c2) d2[j]=0 else d2[j]=1}
17.
18.  stime=c(stime1,stime2)
19.  d=c(d1,d2)
20.  g1=rep(1,n1);g2=rep(2,n2)
21.  group=c(g1,g2)
22.
23.  survdata<-data.frame(stime=stime, d=d,group=group)
24. }
```

Senaryo 4: Ortalarda Kesişen İki Yaşam Fonksiyonu Durumu

Bu senaryoda hedeflenen durum Şekil 3.1-C'de görüldüğü gibi her iki yaşam fonksiyonunun 0.4 ve 0.6 aralığında ($0.4 < S(t) < 0.6$) bir yaşam

olasılığına sahip olduğunda kesişmesidir. Bu amaçla veri türetiminde birinci grup için $\lambda = 1/12$ parametrelili üstel dağılım kullanılmıştır. İkinci grupta ise iki aşamalı veri türetimi yapılmıştır. Öncelikle $\lambda = 0.25$ parametrelili üstel dağılımdan veriler türetilmiştir. Daha sonra ise yaşam süresi 2 ve üzeri olan gözlemler için yeniden $\lambda = 1/35$ parametrelili üstel dağılımdan veri türetilmiştir.

$$\text{Grup 1 : } t_1 \sim \text{Üstel}(1/12)$$

$$\text{Grup 2 : } t_2 \sim \begin{cases} \text{Üstel}(\lambda = 1/4), t_2 < 2 \\ \text{Üstel}(\lambda = 1/35), t_2 \geq 2 \end{cases} \quad (3.4)$$

Senaryo 4'e ait veri türetimi kodlaması ise aşağıdaki gibidir:

```

1. # Senaryo 4: Ortalarda Kesişen İki Yaşam Fonksiyonu
2.
3. veri_turetim4=function (n1, n2, c1,c2){
4.
5.   stime1=rexp(n1,rate=1/12)
6.   stime2=rexp(n2,rate=1/4)
7.
8.   for (m in 1:n2) {if(stime2[m]>=2) stime2[m]=rexp(1,rate=1/35)}
9.
10.  ctime1=runif(n1,0,1)
11.  ctime2=runif(n2,0,1)
12.
13.  d1=0;d2=0
14.
15.  for (j in 1:n1) {if(ctime1[j]<c1) d1[j]=0 else d1[j]=1}
16.  for (j in 1:n2) {if(ctime2[j]<c2) d2[j]=0 else d2[j]=1}
17.
18.  stime=c(stime1,stime2)
19.  d=c(d1,d2)
20.  g1=rep(1,n1);g2=rep(2,n2)
21.  group=c(g1,g2)
22.
23.  survdata<-data.frame(stime=stime, d=d,group=group)
24. }

```

Senaryo 5: Sonlarda Kesişen İki Yaşam Fonksiyonu Durumu

Orantılı Hazardlar Varsayımının ihlal edildiği son senaryoda, iki yaşam fonksiyonunun yaşam olasılıklarının 0.2 ve 0.4 olduğu aralıkta ($0.2 < S(t) < 0.4$) kesişmiş olması kurgulanmıştır. Bu amaçla birinci gruba ait veri türetimi $\alpha = 1.5$ şekil parametrelili ve $\beta = 5$ ölçek parametrelili Weibull dağılımından yapılmıştır. İkinci grupta ise öncelikle $\lambda = 0.5$ parametrelili üstel dağılım kullanılmıştır. Yaşam süresi 1.5 ve üzeri olan gözlemler için tekrardan $\lambda = 0.1$ parametrelili üstel dağılımdan veri türetimi yapılmıştır. Bu veri türetim işlemi sonucunda gözlenen grafik Şekil 3.1-D'de belirtilmiştir.

$$\text{Grup 1 : } t_1 \sim W(1.5,5)$$

$$\text{Grup 2 : } t_2 \sim \begin{cases} \text{Üstel}(\lambda = 0.5), t_2 < 1.5 \\ \text{Üstel}(\lambda = 0.1), t_2 \geq 1.5 \end{cases} \quad (3.5)$$

Bu senaryoya ait veri türetimi kodlaması ise aşağıda yer almaktadır.

```
1. # Senaryo 5: Sonlarda Kesişen İki Yaşam Fonksiyonu
2.
3. veri_turetim5=function (n1, n2, c1,c2){
4.
5.   stime1=rweibull(n1,shape=1.5,scale=5)
6.   stime2=rexp(n2,rate=0.5)
7.
8.   for (m in 1:n2) {if(stime2[m]>=1.5) stime2[m]=rexp(1,rate=0.1)}
9.
10.  ctime1=runif(n1,0,1)
11.  ctime2=runif(n2,0,1)
12.
13.  d1=0;d2=0
14.
15.  for (j in 1:n1) {if(ctime1[j]<c1) d1[j]=0 else d1[j]=1}
16.  for (j in 1:n2) {if(ctime2[j]<c2) d2[j]=0 else d2[j]=1}
17.
18.  stime=c(stime1,stime2)
19.  d=c(d1,d2)
20.  g1=rep(1,n1);g2=rep(2,n2)
21.  group=c(g1,g2)
22.
23.  survdata<-data.frame(stime=stime, d=d,group=group)
24. }
```

3.2. Kullanılan Testlerin Algoritmalarının Oluşturulması

Tez çalışmasında kullanılan karşılaştırma testlerine ait algoritmalar farklı paketler ve fonksiyonlardan faydalanılarak elde edilmiştir. Simülasyon çalışmalarında testlerden elde edilen p değerleri ve bu p değerleri doğrultusunda verilen kararlar karşılaştırıldığı için, algoritmalar bu sonuçları elde edecek şekilde oluşturulmuştur. Tüm test algoritmalarının oluşturulma sürecinde öncelikle bir yaşam veri seti modeli tanımlanmıştır. Daha sonra bu model üzerinden hesaplamalar yapılmıştır.

3.2.1. Log-rank ve ağırlıklandırılmış log-rank testlerinin algoritması

Log-rank ve ağırlıklandırılmış log-rank (Gehan-Wilcoxon, Tarone-Ware, Peto-Peto, Modifiye Edilmiş Peto-Peto, Fleming-Harrington) testlerine ait anlamlılık değeri hesaplamaları *survminer* ve *survival* paketleri kullanılarak elde edilmiştir. Bu tez çalışmasında $p=1$ ve $q=1$ olan Fleming-Harrington testi kullanılmıştır. R programında doğrudan hesaplanan p değerleri aşağıdaki kodlar yardımıyla elde edilmiştir.

```

1. library(survminer)
2. library(survival)
3.
4. # Yaşam veri seti modelinin tanımlanması
5.
6. fit <- surv_fit(Surv(stime, d) ~ group, data = survdata)
7.
8. #Log-rank testi için p değeri
9.
10.    p_LR<-surv_pvalue(fit, survdata, method="1")$pval
11.
12. #Gehan-Wilcoxon testi
13.
14.    p_GB<-surv_pvalue(fit,survdata,method="gehan-breslow")$pval
15.
16. #Tarone-Ware testi
17.
18.    p_TW<-surv_pvalue(fit, survdata, method="sqrtN")$pval
19.
20. #Peto-Peto testi
21.
22.    p_PP<-surv_pvalue(fit, survdata, method="S1")$pval
23.
24. #Modifiye edilmiş Peto-Peto
25.
26.    p_MPP<-surv_pvalue(fit, survdata, method="S2")$pval
27.
28. #Fleming-Harrington testi (p=1 ve q=1)
29.
30.    p_FH<-surv_pvalue(fit, survdata, method="FH_p=1_q=1")$pval

```

3.2.2. Parçalanmış log-rank testlerinin algoritmaları

Hesaplamalarda, *survMisc* ve *dplyr* paketlerinde yer alan fonksiyonlar kullanılmıştır. Testlerin her biri için algoritmalar iki aşamalı oluşturulmuştur. Önce T test istatistiğinin hesaplandığı algoritmalar yazılmıştır. Daha sonra ise test istatistiklerine ait p değerlerinin hesaplandığı algoritmalar oluşturulmuştur.

Algoritmalar arasındaki temel farklılık ağırlık fonksiyonlarından kaynaklanmaktadır. parçalanmış log-rank testlerine ait algoritma kodları aşağıda belirttiği gibidir:

```

1. #Parçalanmış Log-rank Testleri
2.
3. #SUP-LR: Logrank a göre olan parçalanmış log-rank Testi
4. #SUP-GW: Gehan-Wilcoxon a göre olan parçalanmış log-rank Testi
5. #SUP-PP: Peto-Peto ya göre olan parçalanmış log-rank Testi
6.
7. library(survival)
8. library(survMisc)
9. library(dplyr)
10.
11. SUP_LR<-function (survdata){
12.   tmax=0; t_test=0;v_a_new=0
13.
14.   fit<-ten(Surv(stime,d)~group, data=survdata)
15.   c<-comp(fit)

```

```

16.
17. e<-attributes(fit)$p[,3]
18. e_a<-e[e$eMP_1!=0]
19. v<-attributes(fit)$COV[]
20. v_a<-v[v!=0]
21.
22. if (nrow(e_a)<length(v_a)) {
23.     ev<-data.frame(e_a,v_a[1:nrow(e_a)])}
24. else {ev<-data.frame(e_a,v_a)}
25. c_ev<-cumsum(ev)
26.
27. for (i in 2:nrow(c_ev)) {
28.     tmax[i]=c_ev[i-1,1]^2/c_ev[i-1,2]+(c_ev[nrow(c_ev),1]-c_ev[i-
29.     1,1])^2/(c_ev[nrow(c_ev),2]-c_ev[i-1,2])
30. }
31. fit <- surv_fit(Surv(stime, d) ~ group, data = survdata)
32. LR<-surv_pvalue(fit, survdata, method="1")$pval
33. chi_LR<-qchisq(LR,df=1,ncp=0,lower.tail = FALSE)
34. print(t_test<-max(tmax,chi_LR,na.rm=TRUE))
35.
36. }
37.
38. test=SUP_LR(survdata)
39.
40. # p degerini hesaplamak icin
41.
42. SUP_LR_p<-function (survdata,test,n1,n2,b){
43.     t_star=vector("double",b); p_result=0;
44.
45.     survdata2<-survdata[,-3]
46.
47.     for (i in 1:b){
48.
49.         surv1_star<-data.frame(sample_n(survdata2,          size=n1,replace
50.         TRUE),group=rep(1,n1))
51.         surv2_star<-data.frame(sample_n(survdata2,          size=n2,replace
52.         TRUE),group=rep(2,n2))
53.         survdata<-data.frame(merge(surv1_star,surv2_star,all=TRUE),row.names
54.         1:(n1+n2))
55.         t_star[i]<-SUP_LR(survdata)
56.     }
57.
58.     p_result=mean(t_star>test)
59. }
60. SUP_GW<-function (survdata){
61.     tmax=0; t_test=0;
62.
63.     fit<-ten(Surv(stime,d)~group, data=survdata)
64.     c<-comp(fit)
65.
66.     e<-attributes(fit)$p[,3]
67.     e_a<-e[e$eMP_1!=0]
68.     v<-attributes(fit)$COV[]
69.     v_a<-v[v!=0]
70.
71.     gw_w<-attributes(fit)$lrw[,2]
72.
73.     if (nrow(e_a)<nrow(gw_w)) {
74.         e_w<-gw_w[1:nrow(e_a)]*e_a
75.         v_w<-gw_w[1:nrow(e_a)]^2*v_a[1:nrow(e_a)]
76.         ev<-data.frame(e_w,v_w)}
77.     else {

```

```

78.   e_w<-gw_w*e_a
79.   v_w<-gw_w^2*v_a
80.   ev<-data.frame(e_w,v_w)}
81.   c_ev<-cumsum(ev)
82.
83.   for (i in 2:nrow(c_ev)) {
84.     tmax[i]=c_ev[i-1,1]^2/c_ev[i-1,2]+(c_ev[nrow(c_ev),1]-c_ev[i-
1,1])^2/(c_ev[nrow(c_ev),2]-c_ev[i-1,2])
85.   }
86.
87.   fit <- surv_fit(Surv(stime, d) ~ group, data = survdata)
88.   GW<-surv_pvalue(fit, survdata, method="gehan-breslow")$pval
89.   chi_GW<-qchisq(GW,df=1,ncp=0,lower.tail = FALSE)
90.   print(t_test<-max(tmax,chi_GW,na.rm=TRUE))
91.
92. }
93.
94. test=SUP_GW(survdata)
95.
96. # p değerini hesaplamak için
97.
98. SUP_GW_p<-function (survdata,test,n1,n2,b){
99.   t_star=vector("double",b); p_result=0;
100.
101.     survdata2<-survdata[, -3]
102.
103.     for (i in 1:b){
104.
105.         surv1_star<-data.frame(sample_n(survdata2, size=n1,replace =
TRUE),group=rep(1,n1))
106.         surv2_star<-data.frame(sample_n(survdata2, size=n2,replace =
TRUE),group=rep(2,n2))
107.         survdata<-
data.frame(merge(surv1_star,surv2_star,all=TRUE),row.names = 1:(n1+n2))
108.
109.         t_star[i]<-SUP_GW(survdata)
110.     }
111.
112.     p_result=mean(t_star>test)
113. }
114.
115. SUP_PP<-function (survdata){
116.   tmax=0; t_test=0;
117.
118.   fit<-ten(Surv(stime,d)~group, data=survdata)
119.   c<-comp(fit)
120.
121.   e<-attributes(fit)$p[,3]
122.   e_a<-e[e$eMP_1!=0]
123.   v<-attributes(fit)$COV[]
124.   v_a<-v[v!=0]
125.
126.   pp_w<-attributes(fit)$lrw[,4]
127.
128.   if (nrow(e_a)<nrow(pp_w)) {
129.
130.     e_w<-pp_w[1:nrow(e_a)]*e_a
131.     v_w<-pp_w[1:nrow(e_a)]^2*v_a[1:nrow(e_a)]
132.     ev<-data.frame(e_w,v_w)}
133.
134.   else {
135.     e_w<-pp_w*e_a
136.     v_w<-pp_w^2*v_a
137.     ev<-data.frame(e_w,v_w)}
138.     c_ev<-cumsum(ev)
139.

```

```

140.     for (i in 2:nrow(c_ev)) {
141.         tmax[i]=c_ev[i-1,1]^2/c_ev[i-1,2]+(c_ev[nrow(c_ev),1]-c_ev[i-
142.         1,1])^2/(c_ev[nrow(c_ev),2]-c_ev[i-1,2])
143.     }
144.
145.     fit <- surv_fit(Surv(stime, d) ~ group, data = survdata)
146.     PP<-surv_pvalue(fit, survdata, method="S1")$pval
147.     chi_PP<-qchisq(PP,df=1,ncp=0,lower.tail = FALSE)
148.     print(t_test<-max(tmax,chi_PP,na.rm=TRUE))
149.
150. }
151.
152. test=SUP_PP(survdata)
153.
154. # p deęerini hesaplamak iin
155.
156. SUP_PP_p<-function (survdata,test,n1,n2,b){
157.     t_star=vector("double",b); p_result=0;
158.
159.     survdata2<-survdata[, -3]
160.
161.     for (i in 1:b){
162.
163.         surv1_star<-data.frame(sample_n(survdata2, size=n1,replace =
164.         TRUE),group=rep(1,n1))
165.         surv2_star<-data.frame(sample_n(survdata2, size=n2,replace =
166.         TRUE),group=rep(2,n2))
167.         survdata<-
168.         data.frame(merge(surv1_star,surv2_star,all=TRUE),row.names = 1:(n1+n2))
169.
170.         t_star[i]<-SUP_PP(survdata)
171.     }
172.
173.     p_result=mean(t_star>test)
174. }

```

3.3.3. Aęrliklandırılmıř Lin ve Wang testlerinin algoritmaları

Bu testlere ait algoritma hesaplamaları iin *survMisc* paketinde yer alan fonksiyonlardan faydalanılmıřtır. Testler ierisinde ayrı ayrı birden fazla moment hesaplamaları yapıldığı iin, genel bir moment fonksiyonu en bařta oluřturulmuřtur. Dört farklı test iin aęrlik fonksiyonları farklı olacak řekilde benzer algoritmalar elde edilmiřtir. Algoritmalara ait kodlar ařaęıda belirtilmiřtir.

```

1. library(survival)
2. library(survMisc)
3.
4. #LW: Lin ve Wang testi (wi=1)
5. #LW_W1: wi=n olan Aęrliklandırılmıř Lin ve Wang testi
6. #LW_W2:wi=sqrt(n) olan Aęrliklandırılmıř Lin ve Wang testi
7. #LW_W3:wi=wi=1/sqrt(v) olan Aęrliklandırılmıř Lin ve Wang testi
8.
9. LW<-function (survdata){
10.
11.     moment= function(n, x) {
12.
13.         return(factorial(n) / factorial(n-x))
14.     }

```

```

15. }
16.
17. fit<-ten(Surv(stime,d)~group, data=survdata)
18. c<-comp(fit)
19.
20. e1<-attributes(fit)$p[,3]
21. e1<-e1$eMP_1
22. e2<-attributes(fit)$p[,4]
23. e2<-e2$eMP_2
24.
25. v<-attributes(fit)$COV[]
26. v<-unname(v)
27.
28. #delta w için
29.
30.   delta_w<- sum(e1^2)
31.
32. #e (delta w) için
33.
34.   E_delta_w<- sum(v)
35.
36. #e (d1j) ve e (d2j)
37.
38. e_d1<-attributes(fit)$p[,1]
39. e_d1<-e_d1$P_1
40. e_d2<-attributes(fit)$p[,2]
41. e_d2<-e_d2$P_2
42.
43. #d1j, d2j, dj , n1j,n2j,nj
44.
45. d1<-(e_d1-e1)
46. d2<-(e_d2-e2)
47. d<-(d1+d2)
48.
49. butun<-cbind.data.frame(e1,e2,e_d1,e_d2,d1,d2,d,v)
50. bb<-subset(butun,d>0)
51.
52. e1<-bb$e1
53. e2<-bb$e2
54. e_d1<-bb$e_d1
55. e_d2<-bb$e_d2
56. d1<-bb$d1
57. d2<-bb$d2
58. d<-bb$d
59. v<-bb$v
60.
61. n<-(attributes(fit)$lrw[,2])$n
62. n1<-(e_d1*n)/d
63. n2<-(e_d2*n)/d
64.
65. n1[which(!is.finite(n1))] <- 0
66. n2[which(!is.finite(n2))] <- 0
67.
68. #var_d1
69.
70. var_d1<-v
71.
72. #e_2_d1
73.
74. e_2_d1<-var_d1+e_d1^2
75.
76. #e_3_d1
77.
78. m31<-moment(n1,3)
79. m31[which(!is.finite(m31))] <- 0
80.

```

```

81. m32<-moment(d,3)
82. m32[which(!is.finite(m32))] <- 0
83.
84. m33<-moment(n,3)
85. m33[which(!is.finite(m33))] <- 0
86.
87. M3<-(m31*m32)/m33
88. M3[which(!is.finite(M3))] <- 0
89.
90. e_3_d1<-3*e_2_d1-2*e_d1+M3
91.
92. #e_4_d1
93.
94. m41<-moment(n1,4)
95. m41[which(!is.finite(m41))] <- 0
96.
97. m42<-moment(d,4)
98. m42[which(!is.finite(m42))] <- 0
99.
100.     m43<-moment(n,4)
101.     m43[which(!is.finite(m43))] <- 0
102.
103.     M4<-(m41*m42)/m43
104.     M4[which(!is.finite(M4))] <- 0
105.
106.     e_4_d1<-6*e_3_d1-11*e_2_d1+6*e_d1+M4
107.
108.     #var(delta w)
109.
110.     var_delta_w <- sum(e_4_d1-4*e_3_d1*e_d1+6*e_2_d1*e_d1^2-3*e_d1^4-
var_d1^2)
111.
112.     #T w test istatistiği
113.
114.     T_w<-(delta_w-E_delta_w)/sqrt(var_delta_w )
115.
116.     #p değeri
117.
118.     p_w<-2*(1-pnorm(abs(T_w)))
119.
120.     print(p_w)
121.
122.     }
123.
124. LW_W1<-function (survdata){
125.
126.     moment= function(n, x) {
127.
128.         return(factorial(n) / factorial(n-x))
129.
130.     }
131.
132.     fit<-ten(Surv(stime,d)~group, data=survdata)
133.     c<-comp(fit)
134.
135.     e1<-attributes(fit)$p[,3]
136.     e1<-e1$eMP_1
137.     e2<-attributes(fit)$p[,4]
138.     e2<-e2$eMP_2
139.
140.     v<-attributes(fit)$COV[]
141.     v<-unname(v)
142.
143.     #ağırlıklar hesabı için
144.     LW_W1<-attributes(fit)$lrw[,2]$n
145.

```

```

146.      #e (d1j) ve e (d2j)
147.
148.      e_d1<-attributes(fit)$p[,1]
149.      e_d1<-e_d1$P_1
150.      e_d2<-attributes(fit)$p[,2]
151.      e_d2<-e_d2$P_2
152.
153.      #d1j, d2j, dj , n1j,n2j,nj
154.
155.      d1<-(e_d1-e1)
156.      d2<-(e_d2-e2)
157.      d<-(d1+d2)
158.
159.      butun<-cbind.data.frame(e1,e2,e_d1,e_d2,d1,d2,d,v)
160.      bb<-subset(butun,d>0)
161.
162.      e1<-bb$e1
163.      e2<-bb$e2
164.      e_d1<-bb$e_d1
165.      e_d2<-bb$e_d2
166.      d1<-bb$d1
167.      d2<-bb$d2
168.      d<-bb$d
169.      v<-bb$v
170.
171.      n<-(attributes(fit)$lrw[,2])$n
172.      n1<-(e_d1*n)/d
173.      n2<-(e_d2*n)/d
174.
175.      n1[which(!is.finite(n1))] <- 0
176.      n2[which(!is.finite(n2))] <- 0
177.
178.      #delta w için
179.      delta_w1<- sum(LW_W1*e1^2)
180.
181.      #e (delta w) için
182.
183.      E_delta_w1<- sum(LW_W1*v)
184.
185.      #var_d1
186.
187.      var_d1<-v
188.
189.      #e_2_d1
190.
191.      e_2_d1<-var_d1+e_d1^2
192.
193.      #e_3_d1
194.
195.      m31<-moment(n1,3)
196.      m31[which(!is.finite(m31))] <- 0
197.
198.      m32<-moment(d,3)
199.      m32[which(!is.finite(m32))] <- 0
200.
201.      m33<-moment(n,3)
202.      m33[which(!is.finite(m33))] <- 0
203.
204.      M3<-(m31*m32)/m33
205.      M3[which(!is.finite(M3))] <- 0
206.
207.      e_3_d1<-3*e_2_d1-2*e_d1+M3
208.
209.      #e_4_d1
210.
211.      m41<-moment(n1,4)

```



```

212.     m41[which(!is.finite(m41))] <- 0
213.
214.     m42<-moment(d,4)
215.     m42[which(!is.finite(m42))] <- 0
216.
217.     m43<-moment(n,4)
218.     m43[which(!is.finite(m43))] <- 0
219.
220.     M4<-(m41*m42)/m43
221.     M4[which(!is.finite(M4))] <- 0
222.
223.     e_4_d1<-6*e_3_d1-11*e_2_d1+6*e_d1+M4
224.
225.     #var(delta w)
226.
227.     var_delta_w1<-sum(LW_W1^2*(e_4_d1-4*e_3_d1*e_d1+      6*e_2_d1*e_d1^2-
3*e_d1^4-var_d1^2))
228.
229.     #T w test istatistiği
230.
231.     T_w1<-(delta_w1-E_delta_w1)/sqrt(var_delta_w1 )
232.
233.     #p değeri
234.
235.     p_w1<-2*(1-pnorm(abs(T_w1)))
236.
237.     print(p_w1)
238. }
239.
240. LW_W2<-function (survdata){
241.
242.
243.     moment= function(n, x) {
244.
245.         return(factorial(n) / factorial(n-x))
246.
247.     }
248.
249.     fit<-ten(Surv(stime,d)~group, data=survdata)
250.     c<-comp(fit)
251.
252.     e1<-attributes(fit)$p[,3]
253.     e1<-e1$eMP_1
254.     e2<-attributes(fit)$p[,4]
255.     e2<-e2$eMP_2
256.
257.     v<-attributes(fit)$COV[]
258.     v<-unname(v)
259.
260.     #ağırlıklar hesabı için
261.     LW_W2<-attributes(fit)$lrw[,3]$sqrtN
262.
263.     #e (d1j) ve e (d2j)
264.
265.     e_d1<-attributes(fit)$p[,1]
266.     e_d1<-e_d1$P_1
267.     e_d2<-attributes(fit)$p[,2]
268.     e_d2<-e_d2$P_2
269.
270.     #d1j, d2j, dj , n1j,n2j,nj
271.
272.     d1<-(e_d1-e1)
273.     d2<-(e_d2-e2)
274.     d<-(d1+d2)
275.
276.     butun<-cbind.data.frame(e1,e2,e_d1,e_d2,d1,d2,d,v)

```

```

277.      bb<-subset(butun,d>0)
278.
279.      e1<-bb$e1
280.      e2<-bb$e2
281.      e_d1<-bb$e_d1
282.      e_d2<-bb$e_d2
283.      d1<-bb$d1
284.      d2<-bb$d2
285.      d<-bb$d
286.      v<-bb$v
287.
288.      n<-(attributes(fit)$lrw[,2])$n
289.      n1<-(e_d1*n)/d
290.      n2<-(e_d2*n)/d
291.
292.      n1[which(!is.finite(n1))] <- 0
293.      n2[which(!is.finite(n2))] <- 0
294.
295.      #delta w için
296.
297.      delta_w2<- sum(LW_W2*e1^2)
298.
299.      #e (delta w) için
300.
301.      E_delta_w2<- sum(LW_W2*v)
302.
303.      #var_d1
304.
305.      var_d1<-v
306.
307.      #e_2_d1
308.
309.      e_2_d1<-var_d1+e_d1^2
310.
311.      #e_3_d1
312.
313.      m31<-moment(n1,3)
314.      m31[which(!is.finite(m31))] <- 0
315.
316.      m32<-moment(d,3)
317.      m32[which(!is.finite(m32))] <- 0
318.
319.      m33<-moment(n,3)
320.      m33[which(!is.finite(m33))] <- 0
321.
322.      M3<-(m31*m32)/m33
323.      M3[which(!is.finite(M3))] <- 0
324.
325.      e_3_d1<-3*e_2_d1-2*e_d1+M3
326.
327.      #e_4_d1
328.
329.      m41<-moment(n1,4)
330.      m41[which(!is.finite(m41))] <- 0
331.
332.      m42<-moment(d,4)
333.      m42[which(!is.finite(m42))] <- 0
334.
335.      m43<-moment(n,4)
336.      m43[which(!is.finite(m43))] <- 0
337.
338.      M4<-(m41*m42)/m43
339.      M4[which(!is.finite(M4))] <- 0
340.
341.      e_4_d1<-6*e_3_d1-11*e_2_d1+6*e_d1+M4
342.

```

```

343.         #var(delta w)
344.
345.         var_delta_w2<-sum(LW_W2^2*(e_4_d1-4*e_3_d1*e_d1      +6*e_2_d1*e_d1^2-
346.         3*e_d1^4-var_d1^2))
347.         #T w test istatistiği
348.
349.         T_w2<-(delta_w2-E_delta_w2)/sqrt(var_delta_w2 )
350.
351.         #p değeri
352.
353.         p_w2<-2*(1-pnorm(abs(T_w2)))
354.
355.         print(p_w2)
356.     }
357.
358.     LW_W3<-function (survdata){
359.
360.         moment= function(n, x) {
361.
362.             return(factorial(n) / factorial(n-x))
363.
364.         }
365.
366.         fit<-ten(Surv(stime,d)~group, data=survdata)
367.         c<-comp(fit)
368.
369.
370.         e1<-attributes(fit)$p[,3]
371.         e1<-e1$eMP_1
372.         e2<-attributes(fit)$p[,4]
373.         e2<-e2$eMP_2
374.
375.         v<-attributes(fit)$COV[]
376.         v<-unname(v)
377.
378.         #e (d1j) ve e (d2j)
379.
380.         e_d1<-attributes(fit)$p[,1]
381.         e_d1<-e_d1$P_1
382.         e_d2<-attributes(fit)$p[,2]
383.         e_d2<-e_d2$P_2
384.
385.         #d1j, d2j, dj , n1j,n2j,nj
386.
387.         d1<-(e_d1-e1)
388.         d2<-(e_d2-e2)
389.         d<-(d1+d2)
390.
391.         butun<-cbind.data.frame(e1,e2,e_d1,e_d2,d1,d2,d,v)
392.         bb<-subset(butun,d>0)
393.
394.
395.         e1<-bb$e1
396.         e2<-bb$e2
397.         e_d1<-bb$e_d1
398.         e_d2<-bb$e_d2
399.         d1<-bb$d1
400.         d2<-bb$d2
401.         d<-bb$d
402.         v<-bb$v
403.
404.         n<-(attributes(fit)$lrw[,2])$n
405.         n1<-(e_d1*n)/d
406.         n2<-(e_d2*n)/d
407.

```

```

408.      n1[which(!is.finite(n1))] <- 0
409.      n2[which(!is.finite(n2))] <- 0
410.
411.      #AĞIRLIKLAR hesabı için
412.      LW_W3<-(1/sqrt(v))
413.      LW_W3[which(!is.finite(LW_W3))] <- 0
414.
415.      #delta w için
416.
417.      delta_w3<- sum(LW_W3*e1^2)
418.
419.      #e (delta w) için
420.
421.      E_delta_w3<- sum(LW_W3*v)
422.
423.      #var_d1
424.
425.      var_d1<-v
426.
427.      #e_2_d1
428.
429.      e_2_d1<-var_d1+e_d1^2
430.
431.      #e_3_d1
432.
433.      m31<-moment(n1,3)
434.      m31[which(!is.finite(m31))] <- 0
435.
436.      m32<-moment(d,3)
437.      m32[which(!is.finite(m32))] <- 0
438.
439.      m33<-moment(n,3)
440.      m33[which(!is.finite(m33))] <- 0
441.
442.      M3<-(m31*m32)/m33
443.      M3[which(!is.finite(M3))] <- 0
444.
445.      e_3_d1<-3*e_2_d1-2*e_d1+M3
446.
447.      #e_4_d1
448.
449.      m41<-moment(n1,4)
450.      m41[which(!is.finite(m41))] <- 0
451.
452.      m42<-moment(d,4)
453.      m42[which(!is.finite(m42))] <- 0
454.
455.      m43<-moment(n,4)
456.      m43[which(!is.finite(m43))] <- 0
457.
458.      M4<-(m41*m42)/m43
459.      M4[which(!is.finite(M4))] <- 0
460.
461.      e_4_d1<-6*e_3_d1-11*e_2_d1+6*e_d1+M4
462.
463.      #var(delta w)
464.
465.      var_delta_w3<-sum(LW_W3^2*(e_4_d1-4*e_3_d1*e_d1+6*e_2_d1*e_d1^2-
3*e_d1^4-var_d1^2))
466.
467.      #T w test istatistiği
468.
469.      T_w3<-(delta_w3-E_delta_w3)/sqrt(var_delta_w3)
470.
471.      #p değeri
472.

```

```
473.     p_w3<-2*(1-pnorm(abs(T_w3)))
474.
475.     print(p_w3)
476.
477.     }
```

3.3. Simülasyon Algoritmasının Oluşturulması

Bu tez çalışmasında karşılaştırma testlerinin performansının değerlendirilmesi için Monte Carlo Simülasyon tekniği kullanılmıştır. Simülasyon algoritmasının genel yapısı aşağıdaki aşamalar ile oluşturulmuştur:

- 1) Belirlenen senaryoya göre veri türetimi yapılmıştır.
- 2) Türetilen veri setinde yer alan gruplar, tez çalışmasında kullanılan testler ile karşılaştırılmıştır. Her bir test için ayrı ayrı p değeri elde edilmiştir.
- 3) Karşılaştırmalar sonucunda elde edilen p değerlerinin, anlamlılık değeri 0.05'den küçük olup olmadığına dikkat edilmiştir.

Genel olarak üç aşamadan oluşan bu algoritmanın parametrelerini veri türetim senaryosu, tekrar sayısı, örneklem büyüklükleri, sansür oranları ve parçalanmış log-rank testlerinde kullanılan bootstrap sayısı oluşturmaktadır.

Birbirinden bağımsız olarak oluşturulan her bir koşul için simülasyon tekrar sayıları 2000 olarak seçilmiştir (Liu ve Yin, 2016). Beş farklı senaryonun her biri için örneklem büyüklüğü gruplarda eşit ve farklı olacak şekilde belirlenmiştir. Örneklem büyüklükleri eşit olan simülasyonlar için ($n_1 = n_2 = 50,100$) değerleri kullanılmıştır. Farklı olanlarda ise örneklem büyüklükleri ($n_1 = 40, n_2 = 80$) ve ($n_1 = 50, n_2 = 100$) olarak belirlenmiştir. Gruplarda yer alan sansürlü birimler rasgele oluşturulmuştur. Tüm simülasyon çalışmalarında her iki grupta da eşit sansürlü birim oranı olması tercih edilmiştir. Bu oranlar %0, %20, %40 ve %60 olarak belirlenmiştir (Li vd., 2015). Parçalanmış log-rank testlerinde p değerinin elde edilmesi için her bir veri setinde kullanılacak bootstrap tekrar sayısı 1000 olarak düzenlenmiştir (Liu ve Yin, 2016).

Simülasyonlar tamamlandığında ayrı ayrı her bir karşılaştırma testinin p değerleri incelenmiş, 0.05 anlamlılık değerinden küçük olan p değerlerinin sayısı belirlenmiştir. Elde edilen sayı, simülasyon sayısına oranlanarak testlerin güç oranları hesaplanmıştır. Bu güç oranları birbirleri ile karşılaştırılarak farklı koşullarda testlerin durumları değerlendirilmiştir. Genel simülasyon algoritmasını gösteren kodlar aşağıda belirtilmiştir.

```
1. library(survival)
2. library(survminer)
3. library(survMisc)
4. library(dplyr)
5.
6. sim<-function(n1,n2,c1,c2,b,s){
7.
8. #n1 ve n2: örneklem büyüklüğü
9. #c1 ve c2: sansür oranları
10. #s: simülasyon sayısı
11. #b:bootstrap sayısı
12.
13. p_LR=0;p_GB=0;p_TW=0;p_PP=0;p_MPP=0;p_FH=0
14. p_SUP_LR=0;p_SUP_GW=0;p_SUP_PP=0
15. p_LW=0; p_LW_W1=0; p_LW_W2=0;p_LW_W3=0
16.
17. for (i in 1:s){
18.
19.   survdata=veri_turetim1(n1,n2,c1,c2)
20.   #survdata=veri_turetim2(n1,n2,c1,c2)
21.   #survdata=veri_turetim3(n1,n2,c1,c2)
22.   #survdata=veri_turetim4(n1,n2,c1,c2)
23.   #survdata=veri_turetim5(n1,n2,c1,c2)
24.
25.   fit <- surv_fit(Surv(stime, d) ~ group, data = survdata)
26.
27. #log-rank ve ağırlıklandırılmış log-rank testleri
28.   p_LR[i]<-surv_pvalue(fit, survdata, method="1")$pval
29.
30.   p_GB[i]<-surv_pvalue(fit, survdata, method="gehan-breslow")$pval
31.
32.   p_TW[i]<-surv_pvalue(fit, survdata, method="sqrtN")$pval
33.
34.   p_PP[i]<-surv_pvalue(fit, survdata, method="S1")$pval
35.
36.   p_MPP[i]<-surv_pvalue(fit, survdata, method="S2")$pval
37.
38.   p_FH[i]<-surv_pvalue(fit,survdata,method="FH_p=1_q=1")$pval
39.
40. #parçalanmış log-rank testleri
41.
42.   test_LR=SUP_LR(survdata)
43.   p_SUP_LR[i]=SUP_LR_p(survdata,test_LR,n1,n2,b)
44.
45.   test_GW=SUP_GW(survdata)
46.   p_SUP_GW[i]=SUP_GW_p(survdata,test_GW,n1,n2,b)
47.
48.   test_PP=SUP_PP(survdata)
49.   p_SUP_PP[i]=SUP_PP_p(survdata,test_PP,n1,n2,b)
50.
51. #ağırlıklandırılmış lin ve wang testleri
52.
53.   p_LW[i]=LW(survdata)
```

```

54.
55.     p_LW_W1[i]=LW_W1(survdata)
56.
57.     p_LW_W2[i]=LW_W2(survdata)
58.
59.     p_LW_W3[i]=LW_W3(survdata)
60.
61. }
62.
63.     sonuc=c(LR=sum(p_LR<0.05)/s,GB=sum(p_GB<0.05)/s,TW=sum(p_TW<0.05)/s,PP=sum(p_PP
<0.05)/s,MPP=sum(p_MPP<0.05)/s,FH=sum(p_FH<0.05)/s,SUP_LR=sum(p_SUP_LR<0.05)/s,
SUP_GW=sum(p_SUP_GW<0.05)/s,
SUP_PP=sum(p_SUP_PP<0.05)/s,LW=sum(p_LW<0.05)/s,LW_W1=sum(p_LW_W1<0.05)/s,LW_W2
=sum(p_LW_W2<0.05)/s,LW_W3=sum(p_LW_W3<0.05)/s)
64.     print(sonuc)
64. }

```

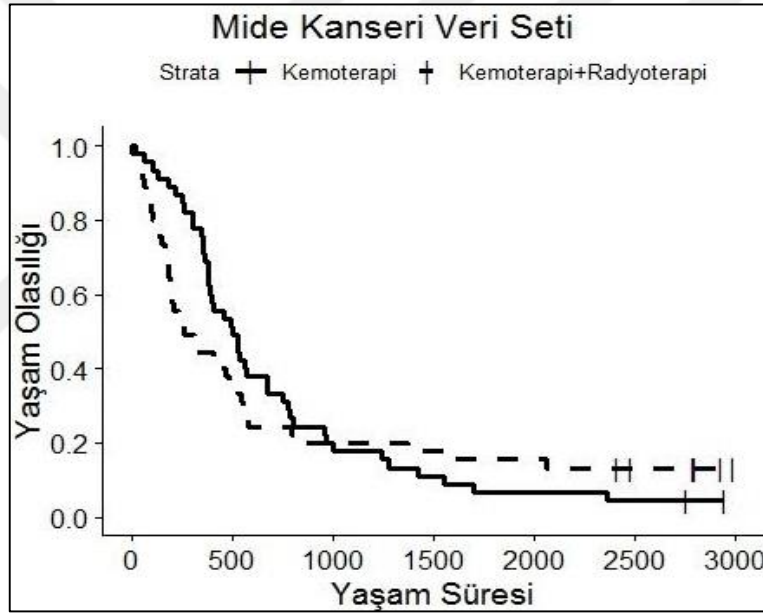
3.4. Uygulama Veri Seti

Lokal olarak cerrahi operasyonla tümörü çıkartılamayan mide kanseri hastaları üzerinde Schein (1982) tarafından bir çalışma planlanmıştır. Uygulama veri setini oluşturan bu çalışmada sekiz yıl boyunca hastalar takip edilmiştir. Her iki grupta 45 kişi olacak şekilde bir planlama yapılmıştır. Tek başına kemoterapi (5 - FU ve metil - CCNU) veya aynı kemoterapi ile birlikte 5000 rad radyoterapi alan lokal ileri mide kanserli hastaların yaşam süreleri karşılaştırılmıştır. Hastaların takip süreleri gün cinsinden kaydedilmiştir. İlgilenilen olay ise ölüm olarak belirlenmiştir. Yapılan takip sonucunda, kemoterapi ile tedavi alan grupta %4.44 ve kemoterapi ile birlikte radyoterapi alan grupta ise %13.33 sansür oranı görülmüştür. Çalışmaya ait veriler Stablein ve Koutrouvelis'in yayınından ya da Klein ve Moeschberger'ın kitabından elde edilebilmektedir (Stablein ve Koutrouvelis, 1985; Klein ve Moeschberger, 2005). Mide kanseri çalışmasına ait veri seti Tablo 3.1'de gösterilmektedir.

Tedavi gruplarına ait yaşam fonksiyonları Şekil 3.2'de gösterilmektedir. İki tedavi grubuna ait yaşam fonksiyonlarının, yaşam olasılığının yaklaşık 0.2'ye eşit olduğu noktada kesiştiği görülmektedir. Bu grafik orantılı hazardlar varsayımının ihlaline yönelik bir ön bilgi vermektedir.

Tablo 3.1 Uygulama veri setine ait veriler

Grup 1: Kemoterapi Tedavi Grubu	
Ölen bireylerin takip süreleri (gün)	1, 63, 105, 129, 182, 216, 250, 262, 301, 301, 342, 354, 356, 358, 380, 383, 383, 388, 394, 408, 460, 489, 499, 523, 524, 535, 562, 569, 675, 676, 748, 778, 786, 797, 955, 968, 1000, 1245, 1271, 1420, 1551, 1694, 2363
Sansürlü bireylerin takip süreleri (gün)	2754, 2950
Grup 2: Kemoterapi + Radyoterapi Tedavi Grubu	
Ölen bireylerin takip süreleri (gün)	17, 42, 44, 48, 60, 72, 74, 95, 103, 108, 122, 144, 167, 170, 183, 185, 193, 195, 197, 208, 234, 235, 254, 307, 315, 401, 445, 464, 484, 528, 542, 547, 577, 580, 795, 855, 1366, 1577, 2060
Sansürlü bireylerin takip süreleri (gün)	2412, 2486, 2796, 2802, 2934, 2988



Şekil 3.2 Tedavi gruplarına ait yaşam fonksiyonları grafiği

Çalışmada orantılı hazardlar varsayımı; log-log yaşam fonksiyonları grafiği, Grambsch ve Therneau tarafından önerilen uyum iyiliği testi ve genişletilmiş Cox Regresyon modeli kullanılarak araştırılacaktır. Daha sonra ise, tez çalışmasında kullanılan karşılaştırma testleri ile mide kanserine yönelik iki farklı tedavi için kıyaslama yapılacaktır. Karşılaştırma testlerinden anlamlı farkı saptayabilenler belirlenecektir.

4. BULGULAR

4.1. Simülasyon Çalışması Bulguları

Birbirinden farklı beş senaryo kullanılarak gerçekleştirilen simülasyon sonuçları, her bir senaryo için ayrı tablolarda sunulmuştur. Tablolarda yer alan sonuçlar, karşılaştırma testlerinin farksızlık hipotezi H_0 'ı reddetme oranlarını göstermektedir. Senaryo 1 kullanılarak yapılan veri türetimi sonucunda elde edilen tip 1 hata oranlarına ait sonuçlar Tablo 4.1'de gösterilmiştir. Eşit örneklem büyüklüğü olan durumlarda tüm testler birbirlerine benzer ve hedeflenen tip 1 hata oranı %5'e yakın sonuçlar vermiştir. Farklı örneklem büyüklüğü ile gerçekleştirilen simülasyon çalışmalarında parçalanmış log-rank testleri, diğer karşılaştırma testlerine göre çok az bir fark ile yüksek tip 1 hata oranları sergilemiştir.

Orantılı hazardlar varsayımını sağlayan iki grubun karşılaştırıldığı Senaryo 2'e ait karşılaştırma testlerinin farksızlık hipotezini reddetme güç oranları Tablo 4.2'de belirtilmiştir. Eşit ve farklı örneklem büyüklüğüne ait sonuçlarda, örneklem büyüklüğündeki artışın testlerin farkı belirleme gücünü arttırdığı görülmüştür. Bunun aksine sansür oranındaki artış ise olumsuz bir etki yaratmıştır. Her bir grupta örneklem büyüklüğü 50 ve üzeri olduğu, sansür oranının ise %40'ı aşmadığı koşullarda tüm testlerde birbirine yakın ve %80 üzerinde güç oranları gözlemlenmiştir. Örneklem büyüklüğünün 100 ve sansür oranının sıfır olduğu koşulda %100 güç oranına sahip karşılaştırma testleri de görülmüştür. Simülasyon sonuçlarına göre bu senaryoda en yüksek güç oranları log-rank testinde bulunmuştur. Orantılı hazardlar varsayımı ihlalinde kullanılması önerilen parçalanmış log-rank testleri ve Lin ve Wang testleri, orantılı hazardlar varsayımında yüksek güç oranları sunmuştur. Ancak, log-rank ve ağırlıklandırılmış log-rank testleri kadar başarılı sonuçlar sergileyememiştir.

Orantılı hazardlar varsayımının ihlal edildiği durumlardan ilki olan başlangıçta kesişen yapının simüle edildiği Senaryo 3'e göre elde edilen karşılaştırma testlerinin güç oranları Tablo 4.3'te gösterilmiştir. Bu senaryoda da artan sansür oranı güç oranlarını olumsuz etkilemiş, örneklem büyüklüğü

arttıkça bu etki daha azalmıştır. Örneklem büyüklüğünün 50 ve üzerine çıktığı durumlarda tüm testlerin gücünde artış ve sansür oranındaki artıştan etkilenme düzeylerinde de azalış görülmüştür. Parçalanmış log-rank testleri en yüksek güç oranlarını sergilemiştir. Parçalanmış log-rank testlerinin ardından en iyi sonuçları sırasıyla Fleming-Harrington ve Lin ve Wang testleri göstermiştir. Parçalanmış log-rank, Fleming-Harrington ve Lin ve Wang testlerinin sansür oranının %40'tan küçük olduğu bazı koşullarda %100'lük güç oranları sunduğu görülmüştür. Gehan-Wilcoxon, Tarone-Ware, Peto-Peto ve Modifiye Edilmiş Peto-Peto karşılaştırma testleri, orantılı hazardlar varsayımına göre daha düşük ama genellikle %80 ve üzerinde güç oranları sergilemiştir. Gehan-Wilcoxon testi sansür oranındaki artıştan en çok etkilenen ve en düşük güç oranları veren karşılaştırma testidir. Çalışmanın başlangıcında yer alan olaylara daha fazla ağırlık vermesi, bu testin daha sonraki izlem sürelerindeki farklılıklara duyarsız olmasına neden olmuştur.

Senaryo 4'te yer alan ve çalışmanın izlem periyodunda ortada kesişen yaşam fonksiyonlarına göre türetilen veri setleri kullanılarak karşılaştırma testlerinden elde edilen güç oranları Tablo 4.4'te gösterilmiştir. Senaryo 3 ile kıyaslandığında bütün karşılaştırma testlerinin farkı belirleme güçlerinde bir düşüş olduğu görülmüştür. Log-rank testleri grubunda en iyi sonuçları sırasıyla Fleming-Harrington ve log-rank vermiştir. Log-rank, Gehan-Wilcoxon, Tarone-Ware, Peto-Peto ve Modifiye Edilmiş Peto-Peto testleri en düşük güç oranlarına sahip testler olarak gözlemlenmiştir. Örneklem büyüklüğündeki artış, bu testlerin gücünde büyük bir değişim yaratmamıştır. Örneklem büyüklüğünün 100 ve sansür oranının sıfır olduğu koşulda bile yaklaşık %10 oranında farkı yakalama gücüne sahip oldukları belirlenmiştir. Parçalanmış log-rank testleri grup başına düşen örneklem büyüklüğü 50 olduğu durumlarda %76.15 ile %100 arasında değişen oldukça yüksek güç oranları vermiştir. Lin ve Wang testleri, log-rank testleri grubundan daha başarılı olsa da, parçalanmış log-rank testleri kadar yüksek güç oranı sonuçları verememiştir. Lin ve Wang testleri, grup başına düşen örneklem büyüklüğü 100 ve sansür oranı %20 ve altında olduğu durumlarda %80 ve üzerinde güç oranı sergileyebilmiştir. Bu senaryo ile yapılan karşılaştırma sonuçlarına bakıldığında Lin ve Wang testleri, sansür

oranındaki artıştan en çok etkilenen test grubudur. Bu test grubu içerisinde en iyi sonucu Lin ve Wang testi verirken, onu sırasıyla $1/\sqrt{Var(d_{1i})}$ 'e göre ağırlıklandırılmış Lin ve Wang, $\sqrt{n_i}$ 'e göre ağırlıklandırılmış Lin ve Wang ve n_i 'e göre ağırlıklandırılmış Lin ve Wang testi takip etmiştir.

Kesişen yaşam fonksiyonları simülasyon karşılaştırmalarındaki son durum olan Senaryo 5'e ait sonuçlar ise Tablo 4.5'te yer almaktadır. Sonuçlar incelendiğinde log-rank, en düşük güç oranlarına sahip karşılaştırma testi olarak bulunmuştur. Örneklem büyüklüğündeki artıştan neredeyse hiç etkilenmemiştir. Log-rank testine ait güç oranları, grup başına düşen örneklem büyüklüğü 100 ve sansür oranı sıfır olduğu koşullarda bile %8.45'in üzerine çıkamamıştır. Lin ve Wang testleri grubu, parçalanmış log-rank testlerinden daha düşük güç oranları sergilemiştir. Sansür oranındaki artıştan etkilenmişlerdir. Sadece örneklem büyüklüğünün her iki grupta da 100 olduğu ve sansür oranının %20'yi aşmadığı durumlarda güç oranları %90.45 ile %99.95 arasında değerler almıştır. Bu test grubu içerisinde en yüksek güç oranı n_i 'e göre ağırlıklandırılmış Lin ve Wang testine ait iken bu testi sırasıyla Lin ve Wang testi, $1/\sqrt{Var(d_{1i})}$ 'e göre ağırlıklandırılmış Lin ve Wang testi ve $\sqrt{n_i}$ 'e göre ağırlıklandırılmış Lin ve Wang testi takip etmiştir. Ağırlıklandırılmış log-rank testleri içerisinde %6.1 ile %53.75 arasında değerler alan Fleming-Harrington testi en düşük güç oranına sahiptir. Tarone-Ware testi Fleming-Harrington'dan daha başarılı sonuçlar sunmuştur. Gehan-Wilcoxon, Peto-Peto ve Modifiye Edilmiş Peto-Peto birbirlerine yakın sonuçlar vermiştir. Bu üç test, ağırlıklandırılmış log-rank testleri grubunda en yüksek güç oranlarına sahiptirler. Aynı zamanda, Gehan-Wilcoxon, Peto-Peto ve Modifiye Edilmiş Peto-Peto testleri, Lin ve Wang test grubundan daha iyi sonuçlar vermiştir. Gehan-Wilcoxon testi, Peto-Peto ve Modifiye Edilmiş Peto-Peto testlerine göre sansür oranındaki artıştan en az etkilenen karşılaştırma testi olmuştur.

Tablo 4.1 Senaryo 1'e göre karşılaştırma testlerinin tip 1 hata oranları

(n_1, n_2)	c	LR	GW	TW	PP	MPP	FH	SUP-LR	SUP-GW	SUP-PP	LW	LW_{w1}	LW_{w2}	LW_{w3}
(50,50)	%0	0.0590	0.0600	0.0545	0.0600	0.0600	0.0580	0.0615	0.0560	0.0550	0.0435	0.0560	0.0555	0.0470
	%20	0.0550	0.0535	0.0485	0.0525	0.0530	0.0630	0.0580	0.0530	0.0530	0.0510	0.0515	0.0525	0.0485
	%40	0.0550	0.0535	0.0510	0.0505	0.0510	0.0595	0.0635	0.0580	0.0560	0.0465	0.0480	0.0495	0.0455
	%60	0.0480	0.0485	0.0510	0.0500	0.0505	0.0475	0.0590	0.0535	0.0560	0.0445	0.0490	0.0470	0.0440
(100,100)	%0	0.0505	0.0470	0.0525	0.0470	0.0470	0.0510	0.0645	0.0640	0.0620	0.0450	0.0465	0.0465	0.0425
	%20	0.0480	0.0480	0.0465	0.0455	0.0460	0.0505	0.0590	0.0535	0.0530	0.0570	0.0560	0.0560	0.0545
	%40	0.0525	0.0595	0.0540	0.0535	0.0535	0.0555	0.0585	0.0605	0.0575	0.0430	0.0480	0.0505	0.0390
	%60	0.0555	0.0585	0.0565	0.0570	0.0575	0.0475	0.0655	0.0525	0.0565	0.0415	0.0480	0.0490	0.0400
(40,80)	%0	0.0550	0.0485	0.0510	0.0485	0.0485	0.0585	0.0655	0.0655	0.0655	0.0555	0.0470	0.0525	0.0540
	%20	0.0555	0.0480	0.0525	0.0485	0.0490	0.0460	0.0755	0.0685	0.0690	0.0515	0.0525	0.0545	0.0475
	%40	0.0550	0.0530	0.0515	0.0530	0.0530	0.0490	0.0815	0.0755	0.0775	0.0565	0.0540	0.0550	0.0535
	%60	0.0450	0.0545	0.0505	0.0470	0.0490	0.0555	0.0690	0.0610	0.0655	0.0455	0.0460	0.0455	0.0435
(50,100)	%0	0.0555	0.0515	0.0530	0.0515	0.0505	0.0555	0.0745	0.0710	0.0685	0.0585	0.0470	0.0515	0.0560
	%20	0.0605	0.0550	0.0565	0.0545	0.0560	0.0525	0.0690	0.0640	0.0645	0.0595	0.0500	0.0560	0.0560
	%40	0.0505	0.0520	0.0515	0.0505	0.0510	0.0545	0.0720	0.0630	0.0635	0.0490	0.0510	0.0510	0.0480
	%60	0.0530	0.0530	0.0505	0.0510	0.0510	0.0585	0.0710	0.0660	0.0660	0.0515	0.0485	0.0510	0.0515

c: sansür oranı; LR: log-rank; GW: Gehan-Wilcoxon; TW: Tarone-Ware; PP: Peto-Peto; MPP: Modifiye edilmiş Peto-Peto; FH: Fleming-Harrington(1,1); SUP-LR: parçalanmış log-rank; SUP-GW: parçalanmış Gehan-Wilcoxon; SUP-PP: parçalanmış Peto-Peto; LW: Lin ve Wang; LW_{w1} : n_i ' e göre ağırlıklandırılmış LW; LW_{w2} : $\sqrt{n_i}$ ' e göre ağırlıklandırılmış LW; LW_{w3} : $1/\sqrt{Var(d_{1i})}$ ' e göre ağırlıklandırılmış LW

Tablo 4.2 Senaryo 2'ye göre karşılaştırma testlerinin güç oranları

(n_1, n_2)	c	LR	GW	TW	PP	MPP	FH	SUP-LR	SUP-GW	SUP-PP	LW	LW_{w1}	LW_{w2}	LW_{w3}
(50,50)	%0	0.9935	0.9755	0.9900	0.9755	0.9740	0.9870	0.9735	0.9755	0.9745	0.9305	0.9600	0.9575	0.8955
	%20	0.9745	0.9345	0.9655	0.9515	0.9485	0.9610	0.9465	0.9445	0.9465	0.8735	0.9225	0.9180	0.8200
	%40	0.9235	0.8410	0.8945	0.8870	0.8845	0.8675	0.8675	0.8765	0.8735	0.7625	0.8125	0.8095	0.6755
	%60	0.7775	0.6665	0.7355	0.7460	0.7460	0.6950	0.7540	0.7695	0.7560	0.5975	0.6575	0.6410	0.5310
(100,100)	%0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9995	0.9990	0.9995	0.9995	1.0000	0.9990
	%20	1.0000	0.9995	1.0000	0.9995	0.9995	0.9990	0.9995	1.0000	1.0000	0.9925	0.9975	0.9980	0.9860
	%40	0.9980	0.9895	0.9965	0.9955	0.9955	0.9935	0.9905	0.9930	0.9930	0.9650	0.9900	0.9880	0.9340
	%60	0.9750	0.9330	0.9645	0.9685	0.9685	0.9295	0.9390	0.9530	0.9435	0.8655	0.9355	0.9210	0.7985
(40,80)	%0	0.9985	0.9865	0.9935	0.9865	0.9865	0.9860	0.9895	0.9915	0.9910	0.9945	0.9890	0.9925	0.9880
	%20	0.9840	0.9525	0.9705	0.9610	0.9605	0.9590	0.9630	0.9675	0.9670	0.9690	0.9640	0.9730	0.9465
	%40	0.9355	0.8825	0.9115	0.9070	0.9040	0.8865	0.9175	0.9345	0.9285	0.9145	0.8980	0.9230	0.8675
	%60	0.7985	0.7330	0.7695	0.7745	0.7745	0.7080	0.7715	0.8025	0.7820	0.7670	0.7545	0.7765	0.7015
(50,100)	%0	1.0000	0.9975	0.9995	0.9975	0.9970	0.9990	0.9980	0.9975	0.9975	0.9995	0.9990	0.9995	0.9990
	%20	0.9955	0.9860	0.9910	0.9870	0.9870	0.9900	0.9860	0.9900	0.9900	0.9925	0.9890	0.9935	0.9825
	%40	0.9710	0.9320	0.9570	0.9535	0.9525	0.9380	0.9495	0.9620	0.9580	0.9535	0.9465	0.9600	0.9285
	%60	0.8700	0.8045	0.8505	0.8585	0.8585	0.7785	0.8580	0.8700	0.8610	0.8395	0.8455	0.8590	0.7760

c: sansür oranı; LR: log-rank; GW: Gehan-Wilcoxon; TW: Tarone-Ware; PP: Peto-Peto; MPP: Modifiye edilmiş Peto-Peto; FH: Fleming-Harrington(1,1); SUP-LR: parçalanmış log-rank; SUP-GW: parçalanmış Gehan-Wilcoxon; SUP-PP: parçalanmış Peto-Peto; LW: Lin ve Wang; LW_{w1} : n_i ' e göre ağırlıklandırılmış LW; LW_{w2} : $\sqrt{n_i}$ ' e göre ağırlıklandırılmış LW; LW_{w3} : $1/\sqrt{Var(d_{1i})}$ ' e göre ağırlıklandırılmış LW

Tablo 4.3 Senaryo 3'e göre karşılaştırma testlerinin güç oranları

(n_1, n_2)	c	LR	GW	TW	PP	MPP	FH	SUP-LR	SUP-GW	SUP-PP	LW	LW_{w1}	LW_{w2}	LW_{w3}
(50,50)	%0	0.9995	0.8820	0.9850	0.8820	0.8710	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%20	0.9965	0.8190	0.9545	0.8775	0.8725	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000
	%40	0.9735	0.6860	0.8850	0.8590	0.8550	1.0000	1.0000	1.0000	1.0000	0.9990	0.9985	0.9995	0.9970
	%60	0.8925	0.5195	0.7260	0.7595	0.7555	0.9960	0.9960	0.9945	0.9950	0.9805	0.9735	0.9795	0.9695
(100,100)	%0	1.0000	0.9910	0.9995	0.9910	0.9905	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%20	1.0000	0.9740	1.0000	0.9925	0.9920	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%40	1.0000	0.9215	0.9965	0.9910	0.9905	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%60	0.9935	0.8025	0.9555	0.9705	0.9695	1.0000	1.0000	1.0000	1.0000	0.9995	0.9995	0.9995	0.9975
(40,80)	%0	1.0000	0.9360	0.9930	0.9360	0.9355	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	1.0000	1.0000
	%20	0.9975	0.8365	0.9670	0.9075	0.9035	1.0000	1.0000	1.0000	1.0000	1.0000	0.9960	0.9995	1.0000
	%40	0.9830	0.7395	0.9025	0.8755	0.8700	1.0000	0.9995	0.9995	0.9995	0.9995	0.9825	0.9980	0.9990
	%60	0.8780	0.5780	0.7515	0.7825	0.7765	0.9945	0.9950	0.9950	0.9950	0.9775	0.9060	0.9590	0.9775
(50,100)	%0	1.0000	0.9765	0.9995	0.9765	0.9750	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%20	1.0000	0.9275	0.9935	0.9695	0.9685	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	%40	0.9920	0.8265	0.9550	0.9365	0.9355	1.0000	1.0000	1.0000	1.0000	0.9995	0.9945	0.9995	1.0000
	%60	0.9445	0.6700	0.8460	0.8725	0.8700	0.9985	0.9995	0.9995	0.9995	0.9945	0.9545	0.9850	0.9940

c: sansür oranı; LR: log-rank; GW: Gehan-Wilcoxon; TW: Tarone-Ware; PP: Peto-Peto; MPP: Modifiye edilmiş Peto-Peto; FH: Fleming-Harrington(1,1); SUP-LR: parçalanmış log-rank; SUP-GW: parçalanmış Gehan-Wilcoxon; SUP-PP: parçalanmış Peto-Peto; LW: Lin ve Wang; LW_{w1} : n_i ' e göre ağırlıklandırılmış LW; LW_{w2} : $\sqrt{n_i}$ ' e göre ağırlıklandırılmış LW; LW_{w3} : $1/\sqrt{Var(d_{1i})}$ ' e göre ağırlıklandırılmış LW

Tablo 4.4 Senaryo 4'e göre karşılaştırma testlerinin güç oranları

(n_1, n_2)	c	LR	GW	TW	PP	MPP	FH	SUP-LR	SUP-GW	SUP-PP	LW	LW_{w1}	LW_{w2}	LW_{w3}
(50,50)	%0	0.3075	0.0760	0.0795	0.0760	0.0770	0.6625	0.9930	0.9855	0.9850	0.7100	0.4765	0.6235	0.6790
	%20	0.2715	0.0885	0.0815	0.0790	0.0790	0.6825	0.9745	0.9620	0.9650	0.6175	0.4325	0.5555	0.5820
	%40	0.2095	0.0730	0.0630	0.0530	0.0545	0.6575	0.9070	0.8885	0.8980	0.5000	0.3240	0.4335	0.4705
	%60	0.1670	0.0665	0.0685	0.0715	0.0710	0.5520	0.7870	0.7615	0.7775	0.3655	0.2540	0.3215	0.3480
(100,100)	%0	0.5855	0.1050	0.1250	0.1050	0.1090	0.9005	1.0000	1.0000	1.0000	0.9590	0.9295	0.9585	0.9370
	%20	0.4880	0.0875	0.0925	0.0620	0.0610	0.9380	1.0000	0.9995	1.0000	0.9140	0.8285	0.8975	0.8735
	%40	0.3755	0.0715	0.0940	0.0665	0.0655	0.9185	1.0000	0.9980	0.9995	0.8110	0.6850	0.7895	0.7510
	%60	0.2875	0.0650	0.0810	0.1005	0.0970	0.8435	0.9700	0.9605	0.9665	0.6535	0.5010	0.6005	0.5935
(40,80)	%0	0.2385	0.0435	0.0510	0.0435	0.0455	0.6540	0.9960	0.9940	0.9940	0.6545	0.1040	0.2755	0.6740
	%20	0.2145	0.0420	0.0500	0.0335	0.0350	0.6715	0.9775	0.9720	0.9735	0.5450	0.0870	0.2415	0.5840
	%40	0.1730	0.0435	0.0565	0.0465	0.0460	0.6465	0.9305	0.9270	0.9290	0.4390	0.0865	0.2020	0.4705
	%60	0.1495	0.0340	0.0500	0.0630	0.0620	0.5595	0.8170	0.8305	0.8250	0.3415	0.0735	0.1720	0.3655
(50,100)	%0	0.3245	0.0430	0.0535	0.0430	0.0450	0.7600	1.0000	1.0000	0.9995	0.7515	0.1000	0.3470	0.7745
	%20	0.2695	0.0495	0.0640	0.0355	0.0345	0.7740	0.9945	0.9935	0.9945	0.6485	0.1095	0.3100	0.6795
	%40	0.1985	0.0395	0.0495	0.0390	0.0380	0.7580	0.9750	0.9705	0.9740	0.5335	0.0755	0.2290	0.5575
	%60	0.1755	0.0440	0.0630	0.0745	0.0725	0.6430	0.8765	0.8870	0.8850	0.4090	0.0755	0.1905	0.4295

c: sansür oranı; LR: log-rank; GW: Gehan-Wilcoxon; TW: Tarone-Ware; PP: Peto-Peto; MPP: Modifiye edilmiş Peto-Peto; FH: Fleming-Harrington(1,1); SUP-LR: parçalanmış log-rank; SUP-GW: parçalanmış Gehan-Wilcoxon; SUP-PP: parçalanmış Peto-Peto; LW: Lin ve Wang; LW_{w1} : n_i ' e göre ağırlıklandırılmış LW; LW_{w2} : $\sqrt{n_i}$ ' e göre ağırlıklandırılmış LW; LW_{w3} : $1/\sqrt{Var(d_{1i})}$ ' e göre ağırlıklandırılmış LW

Tablo 4.5 Senaryo 5'e göre karşılaştırma testlerinin güç oranları

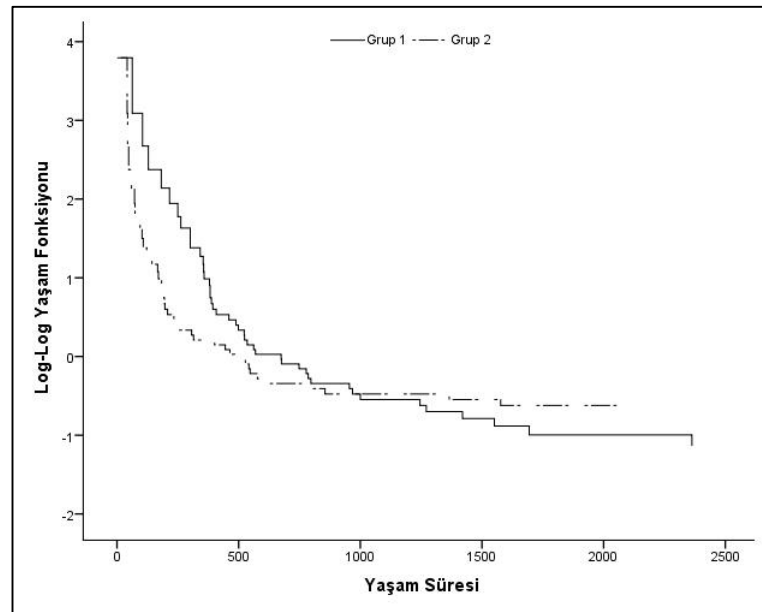
(n_1, n_2)	c	LR	GW	TW	PP	MPP	FH	SUP-LR	SUP-GW	SUP-PP	LW	LW_{w1}	LW_{w2}	LW_{w3}
(50,50)	%0	0.0650	0.7925	0.4395	0.7925	0.8030	0.1340	1.0000	1.0000	1.0000	0.6065	0.6710	0.6090	0.6050
	%20	0.0665	0.6920	0.3575	0.5765	0.5850	0.2135	0.9980	0.9975	0.9970	0.5100	0.4940	0.4560	0.5055
	%40	0.0800	0.5720	0.2900	0.3540	0.3665	0.2860	0.9920	0.9840	0.9845	0.3895	0.3755	0.3500	0.4030
	%60	0.0705	0.4320	0.2280	0.1870	0.1975	0.3240	0.9275	0.9060	0.9135	0.2870	0.2650	0.2530	0.2975
(100,100)	%0	0.0830	0.9665	0.6975	0.9665	0.9685	0.1795	1.0000	1.0000	1.0000	0.9910	0.9995	0.9985	0.9850
	%20	0.0695	0.9315	0.6070	0.8515	0.8610	0.3210	1.0000	1.0000	1.0000	0.9200	0.9470	0.9255	0.9045
	%40	0.0660	0.8545	0.4995	0.6070	0.6175	0.4950	1.0000	1.0000	1.0000	0.7670	0.7735	0.7460	0.7620
	%60	0.0795	0.7015	0.3745	0.3130	0.3195	0.5375	0.9990	0.9985	0.9990	0.5900	0.5500	0.5450	0.5880
(40,80)	%0	0.0400	0.8225	0.4445	0.8225	0.8310	0.0610	1.0000	1.0000	1.0000	0.1640	0.6850	0.2355	0.2840
	%20	0.0355	0.7195	0.3650	0.5990	0.6095	0.1265	1.0000	0.9995	1.0000	0.1670	0.5910	0.1745	0.2660
	%40	0.0520	0.5800	0.2570	0.3375	0.3495	0.2105	0.9945	0.9920	0.9940	0.1415	0.4605	0.1500	0.2335
	%60	0.0405	0.3875	0.1855	0.1485	0.1570	0.2745	0.9385	0.9390	0.9375	0.1415	0.2930	0.1015	0.2105
(50,100)	%0	0.0330	0.9110	0.5390	0.9110	0.9135	0.0615	1.0000	1.0000	1.0000	0.2190	0.7700	0.2380	0.3945
	%20	0.0455	0.8230	0.4455	0.7085	0.7175	0.1345	0.9995	0.9995	0.9995	0.1985	0.6865	0.1985	0.3345
	%40	0.0535	0.6750	0.3335	0.4145	0.4235	0.2475	0.9995	0.9975	0.9975	0.1705	0.5430	0.1715	0.2785
	%60	0.0595	0.5050	0.2260	0.1895	0.1940	0.3180	0.9680	0.9675	0.9680	0.1550	0.3835	0.1235	0.2285

c: sansür oranı; LR: log-rank; GW: Gehan-Wilcoxon; TW: Tarone-Ware; PP: Peto-Peto; MPP: Modifiye edilmiş Peto-Peto; FH: Fleming-Harrington(1,1); SUP-LR: parçalanmış log-rank; SUP-GW: parçalanmış Gehan-Wilcoxon; SUP-PP: parçalanmış Peto-Peto; LW: Lin ve Wang; LW_{w1} : n_i ' e göre ağırlıklandırılmış LW; LW_{w2} : $\sqrt{n_i}$ ' e göre ağırlıklandırılmış LW; LW_{w3} : $1/\sqrt{Var(d_{1i})}$ ' e göre ağırlıklandırılmış LW

Gehan-Wilcoxon, Peto-Peto ve Modifiye Edilmiş Peto-Peto testleri her iki grupta da örneklem büyüklüğünün 100 olduğu ve sansür oranının %20'yi aşmadığı durumlarda güç oranları %85.15 ile %96.85 arasında değişen değerler almıştır. Senaryo 5'e ait sonuçlarda en yüksek güç oranına sahip test grubu parçalanmış log-rank testleridir. Tüm örneklem büyüklükleri ve sansür oranları göz önünde bulundurulduğunda en düşük %90.60 ve en yüksek %100 olan güç oranlarına sahiptir. Bu senaryoda sansür oranındaki artıştan en az etkilenen test grubudur. Ayrıca, tüm senaryolardaki simülasyon çalışmaları göz önünde bulundurulduğunda, parçalanmış log-rank testlerinin diğer testlere kıyasla en fazla öne çıktığı durum Senaryo 5'te görülmüştür.

4.2. Uygulama Veri Seti Bulguları

Uygulama veri seti üzerinde orantılı hazardlar varsayımının araştırılmasında kullanılan yöntemler ile değerlendirmeler yapılmıştır. İlk olarak tedavi yöntemlerine ait log-log yaşam fonksiyonları incelenmiştir. Şekil 4.1'de tedavi yöntemlerine ait log-log yaşam fonksiyonlarının birbirine paralel ilerlemediği, hatta kesiştiği görülmüştür. Grafik, tedavi gruplarının karşılaştırılması sırasında orantılı hazardlar varsayımının sağlanmadığını göstermiştir.



Şekil 4.1 Tedavi gruplarına ait log-log yaşam fonksiyonları

İkinci yaklaşım olan uyum iyiliği testi sonuçlarına göre de orantılı hazardlar varsayımının ihlal edildiği belirlenmiştir ($\chi^2 = 13.1$, $p < 0.001$).

Son yaklaşım olan zamana bağlı açıklayıcı değişken varlığının araştırılması için iki adet Cox regresyon modeli kurulmuştur. Model 1 sadece grup değişkenini içeren Cox Regresyon modelidir. Model 2 ise *grup* ve zamana bağlı fonksiyon ile grup değişkeni çarpımı olan *grup*t* değişkenlerini içeren genişletilmiş Cox Regresyon modelidir. Modellere ait katsayı sonuçları ve anlamlılık değerleri Tablo 4.6'da verilmiştir.

Tablo 4.6 Cox Regresyon modellerine ait sonuçlar

Model 1'e ait sonuçlar					
	β	Standart Hata	Wald Test İstatistiği	Serbestlik Derecesi	P
<i>grup</i>	0.107	0.223	0.232	1	0.630
Model 2'ye ait sonuçlar					
	β	Standart Hata	Wald Test İstatistiği	Serbestlik Derecesi	P
<i>grup</i>	0.928	0.350	7.010	1	0.008
<i>grup*t</i>	-0.002	0.001	8.536	1	0.003

Model 1'e bakıldığında tedavi grupları arasında anlamlı bir fark görülmemiştir ($p=0.630$). Model 2'e ait sonuçlar incelendiğinde, etkileşim terimi *grup*t* değişkenine ait katsayı anlamlı bulunmuştur ($p=0.003$). Grup değişkeni zamana bağımlı bir değişkendir. Bu nedenle tedavi gruplarının karşılaştırıldığı bu çalışmada orantılı hazardlar varsayımı ihlal edilmiştir. Etkileşim teriminin Model 2'ye eklenmesi ile birlikte *grup* değişkeninin katsayısı da anlamlı hale gelmiştir. Tedavi grupları arasında anlamlı bir fark vardır ($p=0.008$).

Cerrahi operasyonla tümörü çıkartılamayan mide kanseri hastaları için kullanılan iki farklı tedavi yönteminin yaşam sürelerinin kıyaslanmasına yönelik karşılaştırma testi sonuçları Tablo 4.7'de verilmiştir. İncelenen karşılaştırma testleri içerisinde parçalanmış log-rank testleri, Gehan-Wilcoxon, Peto-Peto ve Modifiye edilmiş Peto-Peto farkı yakalamada başarılı olurken, log-rank, Tarone-Ware, Fleming-Harrington ve Lin ve Wang testleri farkı yakalamada başarılı olamamıştır. En düşük p değerleri ise, 1000 adet bootstrap örnekleme kullanılarak hesaplama yapan parçalanmış log-rank testleri ile belirlenmiştir.

Tablo 4.7 Uygulama veri setine ait karşılaştırma testi sonuçları

Karşılaştırma Testleri	Test İstatistiği	P Değeri
LR	0.2319	0.6301
GW	3.9963	0.0456
TW	1.9269	0.1651
PP	4.0299	0.0447
Mod-PP	4.1192	0.0424
F-H	0.0111	0.9160
SUP-LR	17.2668	0.0030
SUP-GW	15.2706	0.0040
SUP-PP	15.2932	0.0020
LW	1.0855	0.2778
LW_W1	1.3310	0.1832
LW_W2	1.1853	0.2359
LW_W3	1.1340	0.2568

5. TARTIŞMA

İki yaşam fonksiyonunun karşılaştırılması, yaşam analizinde önemli bir yer tutmaktadır. Yaşam fonksiyonlarının eşitliğini test ederken alternatif hipotez, fonksiyonların izlem periyodunun başlangıcında ya da ileri bölümünde farklılık gösterebileceği, hatta farkların artacağı ya da fonksiyonların kesişeceği gibi pek çok seçenek içermektedir. Bu durum parametrik olmayan yaşam analizi karşılaştırmaları için yalnızca tek bir test kullanma fikrinin açık bir şekilde yanlış sonuçlara yol açabileceğini göstermektedir. Tüm karşılaştırmalar için kullanılacak standart bir test bulunmamaktadır. Bu nedenle, iki yaşam fonksiyonunu karşılaştırmak üzere literatürde araştırmacılar tarafından çok sayıda test önerilmiştir. Ancak bu yöntemler içerisinde en çok tercih edilenler log-rank ve ağırlıklandırılmış log-rank testleridir. Bu durumun ortaya çıkmasındaki en büyük etken, log-rank ve ağırlıklandırılmış log-rank testlerinin hazır bir şekilde paket programlarda yer almasıdır. Log-rank testi, orantılı hazardlar varsayımı altında oldukça iyi sonuçlar sunarken, varsayım bozulduğunda bu başarısını koruyamamaktadır. Araştırmacıların çoğunlukla bu varsayımı dikkate almadığı görülmektedir (Balakrishnan ve Rao, 2004; Bouliotis ve Billingham, 2011; Kristiansen, 2012).

Yaşam fonksiyonlarının kesişmesi, orantılı hazardlar varsayımının ihlal edildiği durumlar içerisinde yer almaktadır. Li vd. (2015) tarafından yapılan bir çalışmada, kesişen yaşam fonksiyonları göz önünde bulundurularak önerilen bazı karşılaştırma testlerinin performansları değerlendirilmiştir. Bu çalışma sonucunda Li ve arkadaşları, Adaptif Neyman (Kraus, 2009) ve iki aşamalı Qui-Sheng (Qiu ve Sheng, 2008) testlerinin oldukça başarılı testler olduğu yönünde öneride bulunmuşlardır. Liu ve Yin (2016) ise log-rank testine dayalı parçalanmış log-rank testlerini önerip, bu yöntemlerin iki aşamalı Qui-Sheng testinden daha başarılı olduğu sonucunu elde etmiştir. Koziol ve Jia (2014), log-rank testi istatistiği kullanılarak karesel yaklaşım ile elde edilen Lin ve Wang testine ağırlıklandırma fonksiyonları ekleyerek yeni yöntemler elde edip öneride bulunmuş, ancak bu yöntemleri literatürdeki herhangi bir test ile karşılaştırmamıştır.

Bu tez çalışmasında, literatürde görülen kesişen yaşam fonksiyonu şemaları göz önünde bulundurularak Monte Carlo simülasyon tekniği ile testlerin performansları karşılaştırılmıştır. Literatürde son zamanlarda önerilen parçalanmış log-rank ve ağırlıklandırılmış Lin ve Wang testleri ile log-rank ve ağırlıklandırılmış log-rank testleri farklı senaryolarda, farklı örneklem büyüklüklerinde ve sansür oranlarında kıyaslanmıştır.

Yapılan simülasyon çalışmalarındaki ilk senaryoda, karşılaştırma testlerinin tip 1 hata oranları incelenmiştir. Testlerin beklenen nominal değer olan %5'e yakın sonuçlar verdiği görülmüştür. Örneklem büyüklüğü ya da sansür oranındaki artışın, tip 1 hata oranlarında önemli bir değişim yaratmadığı gözlemlenmiştir. Çalışmamızda kullanılan karşılaştırma testlerinin literatürdeki diğer çalışmalardaki tip 1 hata oranlarına bakıldığında, benzer sonuçlara sahip olduğu gözlemlenmiştir (Lin ve Wang, 2004; Lin ve Xu, 2010; Koziol ve Jia, 2014; Liu ve Yin, 2016; Hsieh ve Chen, 2017) .

Senaryo 2'de ise orantılı hazardlar varsayımı altında karşılaştırma testlerinin farkı yakalama gücü belirlenmiştir. Kullanılan testler oldukça iyi performans sergilerken, log-rank testinin tüm örneklem büyüklüğü ve sansür oranlarında en iyi sonucu verdiği gözlemlenmiştir. Bu bulgu literatür tarafından da desteklenmektedir. Orantılı hazardlar varsayımının sağlandığı benzer simülasyon karşılaştırmalarında log-rank testinin, kendisi ile karşılaştırılan tüm karşılaştırma testlerinden daha yüksek güç oranlarına sahip olduğu belirlenmiştir (Mantel ve Stablein, 1988; Balakrishnan ve Rao, 2004; Armitage vd., 2008; Li vd., 2015; Liu ve Yin, 2016; Hsieh ve Chen, 2017).

Orantılı hazardlar varsayımının ihlal edildiği ilk durum olan başlangıçta kesişen yaşam fonksiyonları Senaryo 3'te incelenmiştir. Orantılı hazardlar varsayımı ihlali için geliştirilen parçalanmış log-rank ve Lin ve Wang testleri beklenen yüksek farkı yakalama gücüne sahip sonuçlar sunduğu gözlemlenmiştir. Ancak beklenenin aksine log-rank testinin de yüksek güç oranları sergilediği belirlenmiştir. Ağırlıklandırılmış log-rank testlerinin de farkı belirlemede başarılı oldukları gözlemlenmiştir. Bu bulgulara benzer

sonular Li ve arkadaşları, Hsieh ve Chen ile Tubert-Bitter ve arkadaşlarının alıřmalarında da grlmřtr (Tubert-Bitter, Kramar, Chal & Moreau, 1994; Li vd., 2015; Hsieh ve Chen, 2017) .

Senaryo 4'te grlen ortada kesişen yařam fonksiyonlarına ait elde edilen simlasyon sonularında, paralanmıř log-rank testleri, diđer testlerle karřılařtırıldıđında en iyi sonuları verdiđi gzlemlenmiřtir. Orantılı hazardlar varsayımı ihlalinde kullanılması iin geliřtirilen Lin ve Wang testlerinin performansında olduka dřř grlmřtr. rneklem byklđnn 100 ve sansr oranının dřk olduđu kořulda g oranı %90'ın zerinde sonular elde etmiřtir. Lin ve Wang testleri ierisinde en iyi sonuları ađırlık fonksiyonu olmayan Lin ve Wang testi vermiřtir. Benzer bulgular Koziol ve Jia (2014) tarafından da desteklenmektedir. Lin ve Wang testi bu senaryoda sansr oranındaki artıřtan en fazla etkilenen test olarak bulunmuřtur. Li ve arkadaşlarının (2015) yaptıđı alıřmada da benzer bulgulara rastlanmıřtır. Fleming-Harrington dıřındaki ađırlıklandırılmıř log-rank testleri ve log-rank testi olduka kt performans sergilemiřtir. Bunun sebebinin, test istatistiđinde pay kısmında yer alan pozitif farkların, negatif farklar tarafından ntrlenmesi olduđu dřnlmektedir. Literatrde de senaryo 4'e benzer ortada kesişen yapı ile gerekleřtirilen simlasyonlara ait sonularda, log-rank ve Gehan-Wilcoxon testlerinin olduka dřk g oranlarına sahip olduđu grlmřtr (Tubert-Bitter vd., 1994; Lin ve Wang, 2004; Lin ve Xu, 2010; Jurkiewicz ve Wycinka, 2011; Li vd., 2015; Hsieh ve Chen, 2017). Log-rank ve diđer ađırlıklandırılmıř log-rank testlerinin aksine Fleming-Harrington kullandıđı dengeli ađırlık fonksiyonu ile olduka bařarılı sonular sunmuřtur ve bu bulgu Li ve arkadaşları (2015) tarafından da desteklenmiřtir.

Son simlasyon alıřma kurgusu olan Senaryo 5'te, alıřmanın sonlarında kesişen yařam fonksiyonları durumunda testlerin performansları incelenmiřtir. Bu senaryoda da paralanmıř log-rank testleri n plana ıkararak en yksek g oranlarını sunmuřtur. Lin ve Wang testleri, sadece byk rneklem byklđ ve dřk sansr oranında yksek g oranlarına sahip performans gstermiřtir. Benzer sonular Li ve arkadaşlarının alıřmasında da yer almıřtır (Li vd., 2015). Log-rank ve Fleming-Harrington testleri tm

senaryolar içerisinde en düşük güç oranlarını bu senaryoda vermiştir. Çalışmanın başlangıç dönemine odaklanarak hesaplama yapan Gehan-Wilcoxon, Tarone Ware, Peto-Peto ve Modifiye edilmiş Peto-Peto testleri ortada kesişen yapıya göre daha yüksek güç oranları sunmuştur. Sonlarda kesişen yapıya benzer simülasyon çalışması yapan literatürdeki diğer araştırmalarda da log-rank ve Fleming-Harrington testlerinin, Gehan-Wilcoxon ve Tarone Ware testlerinden daha düşük güç oranlarına sahip olduğu gözlemlenmiştir (Li vd., 2015; Hsieh ve Chen, 2017).

Güç oranlarının karşılaştırıldığı simülasyon çalışması sonuçları incelendiğinde örneklem büyüklüğündeki artışın testleri olumlu, sansür oranındaki artışın ise olumsuz yönde etkilediği görülmüştür. Literatürdeki simülasyon çalışmalarında da benzer sonuçlar gözlemlenmiştir (Tubert-Bitter vd., 1994; Lin ve Wang, 2004; Koziol ve Jia, 2014; Li vd., 2015; Liu ve Yin, 2016; Hsieh ve Chen, 2017)

Orantılı hazardlar varsayımının ihlal edildiği senaryolarda farklı kesişim noktaları göz önünde bulundurularak değerlendirilmeler yapılmıştır. Kesişim noktasının yerinin testlerin gücü üzerinde önemli bir rol oynadığı belirlenmiştir (Tubert-Bitter vd., 1994; Jurkiewicz ve Wycinka, 2011; Li vd., 2015; Hsieh ve Chen, 2017). Kesişim noktasının soldan sağa doğru ilerlemesinden en çok etkilenen karşılaştırma testi log-ranktır. Sırasıyla ağırlıklandırılmış log-rank testleri ve Lin ve Wang testleri kesişim noktasından etkilenen diğer testlerdir. Benzer sonuç Li ve arkadaşlarının (2015) çalışmasında da görülmüştür. Parçalanmış log-rank testleri bu değişimden neredeyse hiç etkilenmemiştir. Bu durumun test istatistiği hesaplamalarında kullanılan algoritmadan kaynaklandığı düşünülmektedir. Parçalanmış log-rank testleri, ilgilenilen olayın gözlemlendiği zaman noktalarının her birinin öncesi ve sonrasını ayrı ayrı değerlendirerek hesaplamaya dahil etmektedir. Liu ve Yin (2016) çalışmalarında olayların görüldüğü zaman noktalarından birinde iki yaşam fonksiyonu kesişirse, testin farkı yakalama gücünün artacağını belirtmişlerdir.

Parçalanmış log-rank testleri tüm senaryolarda diğer karşılaştırma testlerine kıyasla oldukça başarılı sonuçlar sunmuştur. Ancak 50'den küçük

örneklem büyüklüğü ve %40 ve üzeri yüksek sansür oranı koşulunda yapılan denemelerde simülasyon çalışması tamamlanamadığı için sonuç elde edilememiştir. Çünkü, yüksek sansür oranına sahip veri setlerinden iadeli çekilen bootstrap örnekleri sadece sansürlü gözlemler içerdiğinde, log-rank testleri hesaplama yapamamıştır. Bootstrap tekniğinde seçimlerin rasgele yapılması gerektiği için p değeri hesaplama algoritmasına herhangi bir müdahalede bulunulamamıştır. Bu nedenle de simülasyon çalışması durmuş ve döngüler tamamlanamamıştır. Bu durum parçalanmış log-rank testlerinin önemli bir kısıtıdır. Simülasyon çalışmasında yaşanan bu sorundan dolayı bu tez çalışmasında eşit örneklem büyüklüğü koşulundaki veri türetimlerinde en küçük örneklem büyüklüğü değeri olarak 50 tercih edilmiştir. Parçalanmış log-rank testlerini literatüre kazandıran Liu ve Yin (2016) de çalışmalarında, simülasyon kurgusunda kullanılan en küçük örneklem büyüklüğünü 50 olarak belirlemişlerdir.

Parçalanmış log-rank testlerini, diğer testlerden ayıran bir diğer özellik ise p değerinin hesaplanmasında bootstrap tekniğinin kullanılmasıdır. Bu nedenle aynı veri seti kullanılarak yapılan iki grup karşılaştırmalarının her birinde nadiren aynı ama çoğu zaman farklı p değerleri sunabilmektedir. Bu p değerleri farklı olsa da birbirine oldukça yakın sonuçlardır.

Bu tez çalışmasında simülasyon çalışmalarına ek olarak, sağlık alanında bir uygulamaya da yer verilmiştir. Lokal olarak tümörü çıkartılamayan mide kanseri hastalarından bir gruba kemoterapi, diğerine ise kemoterapi ile birlikte radyoterapi veren bir çalışmanın verileri kullanılmıştır. Tedavi yöntemlerine ait yaşam fonksiyonlarının grafiğinin sonlarda kesiştiği ve bu yönüyle simülasyon çalışmalarında kullanılan Senaryo 5'e benzediği belirlenmiştir. Tez çalışmasında yer alan karşılaştırma testleri kullanılarak iki tedavi yöntemi kıyaslandığında, Gehan-Wilcoxon, Peto-Peto, Modifiye Edilmiş Peto-Peto ve parçalanmış log-rank testleri anlamlı farkı yakalayan testler olarak belirlenmiştir. Log-rank, Tarone-Ware, Fleming-Harrington ve Lin ve Wang testleri ise farkı yakalamada başarılı olamamıştır. Uygulama veri setinden elde edilen bu sonuçların, Senaryo 5'e ait simülasyon sonuçları ile uyum sağladığı gözlemlenmiştir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında iki yaşam fonksiyonunun karşılaştırılmasında kullanılan testler incelenmiş ve belirli koşullar altında kıyaslanmıştır. Orantılı hazardlar varsayımının sağlandığı ve kesişen yaşam fonksiyonları ile bu varsayımın ihlal edildiği senaryolar kullanılmıştır. Monte Carlo simülasyon tekniği ile farklı örneklem büyüklüklerinde ve farklı sansür oranlarında karşılaştırma testlerinin farkı yakalama güçleri elde edilmiştir. Ayrıca, Monte Carlo simülasyon tekniği ile karşılaştırma testlerinin tip 1 hata oranları da incelenmiş ve tüm testlerin 0.05 değerine yakın sonuçlar verdiği gözlemlenmiştir. Orantılı hazardlar varsayımı altında en iyi sonuçları log-rank testinin, kesişen yaşam fonksiyonları durumunda ise parçalanmış log-rank testlerinin sergilediği belirlenmiştir. Örneklem büyüklüğündeki artışın testlerin performansını olumlu, ancak sansür oranındaki artışın olumsuz yönde etkilediği sonucu elde edilmiştir. Ayrıca kesişen yaşam fonksiyonlarına ait senaryolarda, kesişim noktasının testlerin performansında önemli bir etki yarattığı gözlemlenmiştir.

Mide kanseri hastalarına yönelik yapılan bir çalışmada iki farklı tedavi yöntemi uygulandıktan sonra elde edilmiş takip süreleri, tez çalışmasında yer alan karşılaştırma testleri kullanılarak kıyaslanmıştır. Bu veri setinde orantılı hazardlar varsayımının ihlal edildiği ve iki tedavi yönteminin çalışmanın sonlarına doğru kesiştiği belirlenmiştir. Gruplar arası farkı yakalamada en başarılı yöntemin parçalanmış log-rank testleri olduğu bulunmuştur. Bu sonuç, simülasyon çalışması sonuçlarını desteklemiştir.

Parçalanmış log-rank testleri küçük örneklem büyüklüğü ve yüksek sansür oranları ile oluşturulan koşullarda sonuç verememiştir. Bu nedenle tez çalışmasında küçük örneklem büyüklüğünde simülasyon çalışması gerçekleştirilmemiştir. Bu durum çalışmamızın temel kısıtını oluşturmaktadır. İleri dönemlerde belirli küçük örneklem büyüklükleri için parçalanmış log-rank testlerinden sonuç elde edilebilecek en büyük sansür oranı sınırlarının belirlenmesi yeni bir çalışma konusu olabilir.

İki yaşam fonksiyonu karşılaştıran araştırmacıların, ilgili karşılaştırma testinin uygulanmasının bazı temel varsayımlar taşıdığı farkında olması gerekmektedir. En uygun testin seçimi, örneklem büyüklüğüne, sansür yapısına ve yaşam ya da hazard fonksiyonlarının durumuna bağlıdır. Sadece bir test, tüm durumlara eşit derecede uymamaktadır. Karşılaştırma testi, hazır bir şekilde bir paket programda mevcut olduğundan basitçe tercih edilmemelidir.

Bu tez çalışmasında; orantılı hazardlar varsayımı sağlandığında log-rank testinin, orantılı hazardlar varsayımının ihlal edildiği koşullardan biri olan kesişen yaşam fonksiyonlarının karşılaştırılmasında ise parçalanmış log-rank testlerinin kullanılması önerilmektedir. Gelecek çalışmalarda parçalanmış log-rank testlerinin paket program haline getirilerek kullanımının arttırılması ile literatüre önemli bir katkıda bulunacağı düşünülmektedir.

KAYNAKLAR DİZİNİ

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 701-726.
- Allison, P. D. (2004). *Survival analysis using SAS: a practical guide*: Sas Institute.
- Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1992). *Statistical models based on counting processes*: Springer Science & Business Media.
- Andersen, P. K., & Gill, R. D. J. T. a. o. s. (1982). Cox's regression model for counting processes: a large sample study. 1100-1120.
- Armitage, P., Berry, G., & Matthews, J. N. S. (2008). *Statistical methods in medical research*: John Wiley & Sons.
- Balakrishnan, N., & Rao, C. R. (2004). *Advances in survival analysis* (Vol. 23): Elsevier.
- Bouliotis, G., & Billingham, L. J. T. (2011). Crossing survival curves: alternatives to the log-rank test. *12(S1)*, A137.
- Cox, D. (1972). Regression models and life-tables JR Statist Soc B 34: 187–220.
- Dardis, C., & Dardis, M. C. (2018). Package ‘survMisc’.
- Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569-593): Springer.
- Efron, B., & Tibshirani, R. J. S. s. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. 54-75.
- Fleming, T. R., O'Fallon, J. R., O'Brien, P. C., & Harrington, D. P. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data. *Biometrics*, 607-625.
- Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2), 203-224.
- Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica*, 34(2), 124-124.
- Grambsch, P. M., & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3), 515-526.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Harrell, F. E., & Lee, K. L. (1986, February). Verifying assumptions of the Cox proportional hazards model. In *Proceedings of the eleventh annual SAS Users group international conference* (pp. 823-828). SAS Institute Inc, Cary, NC.
- Harrington, D. P., & Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, 69(3), 553-566.
- Hosmer Jr, D. W., Lemeshow, S., & May, S. (2011). *Applied survival analysis: regression modeling of time-to-event data* (Vol. 618): John Wiley & Sons.
- Hsieh, J.-J., & Chen, H.-Y. (2017). A testing strategy for two crossing survival curves. *J Communications in Statistics-Simulation Computation*, 46(8), 6685-6696.
- Jurkiewicz, T., & Wycinka, E. (2011). Significance Tests of Differences Between Two Crossing Survival Curves for Small Samples. 255, 107-114.
- Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282), 457-481.
- Kassambara, A., Kosinski, M., Biecek, P., & Fabian, S. (2017). Package 'survminer'.
- Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*: Springer Science & Business Media.
- Kleinbaum, D. G., & Klein, M. (2006). *Survival analysis: a self-learning text*: Springer Science & Business Media.
- Koziol, J. A., & Jia, Z. (2014). Weighted Lin-Wang tests for crossing hazards. *Computational and mathematical methods in medicine*, 2014.
- Koziol, J. A. (1978). A Two Sample CRAMÉR-VON MISES Test for Randomly Censored Data. 20(6), 603-608.
- Kraus, D. (2009). Adaptive Neyman's smooth tests of homogeneity of two samples of survival data. *J Stat Plan Infer*, 139(10), 3559-3569.
- Krishnan, A., Pasquini, M. C., Logan, B., Stadtmauer, E. A., Vesole, D. H., Alyea 3rd, E., . . . & Laport, G. (2011). Autologous haemopoietic stem-cell transplantation followed by allogeneic or autologous haemopoietic stem-cell transplantation in patients with multiple myeloma (BMT CTN 0102): a phase 3 biological assignment trial. *The lancet oncology*, 12(13), 1195-1203.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Kristiansen, I. (2012). PRM39 Survival curve convergences and crossing: a threat to validity of meta-analysis? *Value in health*, 15(7), A652.
- Lee, S. H. (2007). On the versatility of the combination of the weighted log-rank statistics. *Computational Statistics & Data Analysis*, 51(12), 6557-6564.
- Li, H., Han, D., Hou, Y., Chen, H., & Chen, Z. (2015). Statistical inference methods for two crossing survival curves: a comparison of methods. *PloS one*, 10(1), e0116774.
- Lin, X., & Wang, H. (2004). A new testing approach for comparing the overall homogeneity of survival curves. *Biometrical Journal*, 46(5), 489-496.
- Lin, X., & Xu, Q. (2010). A new method for the comparison of survival distributions. *Pharmaceutical statistics*, 9(1), 67-76.
- Liu, K., Qiu, P., & Sheng, J. (2007). Comparing two crossing hazard rates by Cox proportional hazards modelling. *Statistics in medicine*, 26(2), 375-391.
- Liu, Y., & Yin, G. (2016). Partitioned log-rank tests for the overall homogeneity of hazard rate functions. *Lifetime data analysis*, 1-26.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50, 163-170.
- Mantel, N., & Stablein, D. M. (1988). The crossing hazard function problem. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37(1), 59-64.
- Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4), 945-966.
- O'Quigley, J. (2012). Survival Model Inference Using Functions of Brownian Motion. *Applied Mathematics*, 3(06), 641.
- Pepe, M. S., & Fleming, T. R. (1989). Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. *Biometrics*, 497-507.
- Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 185-207.
- Qiu, P., & Sheng, J. (2008). A two-stage procedure for comparing hazard rate functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1), 191-208.

KAYNAKLAR DİZİNİ (Devam Ediyor)

- Revuz, D., & Yor, M. (1999). *Continuous martingales and Brownian motion* (Vol. 293): Springer-Verlag Berlin Heidelberg.
- Rosner, B. (2010). *Fundamentals of biostatistics*. Boston MA USA: Cengage Learning.
- Schein, P. S. (1982). A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma. *J Cancer*, 49(9), 1771-1777.
- Schumacher, M. (1984). Two-Sample Tests of Cramér--von Mises-and Kolmogorov--Smirnov-Type for Randomly Censored Data. *International Statistical Review/Revue Internationale de Statistique*, 263-281.
- Shen, Y., & Cai, J. (2001). Maximum of the Weighted Kaplan-Meier Tests with Application to Cancer Prevention and Screening Trials. *Biometrics*, 57(3), 837-843.
- Stablein, D. M., & Koutrouvelis, I. A. (1985). A two-sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics*, 643-652.
- Tarone, R. E., & Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, 64(1), 156-160.
- Therneau, T. M., & Lumley, T. (2014). Package 'survival'. *Survival analysis Published on CRAN*.
- Tubert-Bitter, P., Kramar, A., Chalé, J.-J., & Moreau, T. (1994). Linear rank tests for comparing survival in two groups with crossing hazards. *J Computational statistics data analysis*, 18(5), 547-559.
- Weeda, V., Murawski, M., McCabe, A., Maibach, R., Brugières, L., Roebuck, D., . . . & Perilongo, G. (2013). Fibrolamellar variant of hepatocellular carcinoma does not have a better survival than conventional hepatocellular carcinoma—results and treatment recommendations from the Childhood Liver Tumour Strategy Group (SIOPEL) experience. *European journal of cancer*, 49(12), 2698-2704.
- Wickham, H., & Wickham, M. H. (2013). Package 'plyr'.

ÖZGEÇMİŞ



HÜLYA ÖZEN (YILMAZ)

KİŞİSEL BİLGİLER

Adı Soyadı: Hülya ÖZEN

Bekarlık Soyadı: YILMAZ

Doğum Tarihi ve Yeri: 21/06/1987, İZMİR

Uyruğu: T.C.

Medeni Hali: Evli

Yabancı Dil: İngilizce

Yabancı Dil Belgesi: YÖKDİL, Mart 2017 (91,25/100)

Adres: Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, 26480, Odunpazarı / ESKİŞEHİR

e-posta: hulyayilmaz@ogu.edu.tr; hulya_ozen@yahoo.com

EĞİTİM VE AKADEMİK KARİYER BİLGİSİ

1.Lisans: Ege Üniversitesi Fen Fakültesi İstatistik Bölümü / 2005-2009

2.Yüksek Lisans: Ege Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı/ 2009-2011

Yüksek Lisans Tez Başlığı:

Çarpık Dağılımlarda Parametre Tahmin Yöntemlerinin Karşılaştırılması, Ağustos 2011, Danışman: Yrd. Doç. Dr. Hakan Savaş SAZAK

3.Yüksek Lisans: Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı / 2011-2014

Yüksek Lisans Tez Başlığı:

Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama, Ocak 2014, Danışman Yrd. Doç. Dr. Cengiz BAL

4. Doktora: Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı / 2014- 2019

Doktora Tez Başlığı:

Kesişen İki Yaşam Fonksiyonu için Kullanılan İstatistiksel Testlerin Karşılaştırılması

ESERLER LİSTESİ

TAM METİNLİ MAKALELER (SCI,SCI- EXPANDED, SSCI, AHCI)

A1. Yılmaz Hülya, Sazak Hakan Savaş (2014). Double looped Maximum Likelihood Estimation For The Parameters Of The Generalized Gamma Distribution. *Mathematics and Computers in Simulation*(98), 18-30. Doi: 10.1016/j.matcom.2013.12.001

A2. Kalemci Orhan, Aydın Hasan Emre, Kızmazoğlu Ceren, Kaya İsmail, **Yılmaz Hülya**, Arda Mehmet Nuri (2017). Effects of Quercetin and Mannitol on Erythropoietin Levels in Rats Following Acute Severe Traumatic Brain Injury. *Journal of Korean Neurosurgical Society*, 60(3), 355-361., Doi: 10.3340/jkns.2016.0505.015

A3. Özbek Zühtü, Özkara Emre, Öner Hasan, Baş Gültekin, Erman İpek Canan, **Özen Hülya**, Entok Emre, Arslantaş Ali (2017). Treatment of Unstable Thoracolumbar Fractures: Does Fracture-Level Fixation Accelerate the Bone Healing?. *World Neurosurgery*, 107, 362-370., Doi: 10.1016/j.wneu.2017.08.007

A4. Çolak Ertuğrul, **Özen Hülya**, Emir Büşra, Öner Kevser Setenay (2017). Pairwise Multiple Comparison Adjustment Procedure for Survival Functions with Right-Censored Data. *Computational and Mathematical Methods in Medicine*, 2017, 1-8., Doi: 10.1155/2017/9270450

A5. Özkaraman Ayşe, Dügüm Özlem, **Özen Hülya**, Usta Yeşilbalkan Öznur (2018). Aromatherapy: The Effect of Lavender on Anxiety and Sleep Quality in Patients Treated With Chemotherapy. *Clinical Journal of Oncology Nursing*, 22(2), 203-210., Doi: 10.1188/18.CJON.203-210

A6. Erzurumluoğlu Ebru, Çilingir Oğuz, Özbabalık Adapınar Belgin Demet, Bilgiç Başar, Kocagil Sinem, **Özen Hülya**, Durak Aras Beyhan, Yenilmez Çınar, Artan Sevilhan (2019). The association between repeat number in C9orf72 and phenotypic variability in Turkish patients with frontotemporal lobar degeneration. *Neurobiology of Aging*, 76, 2161-2167, Doi: 10.1016/j.neurobiolaging.2018.12.007

TAM METİNLİ MAKALELER (DİĞER ALAN İNDEKSLERİ & ULUSAL HAKEMLİ)

B1. Sazak Hakan Savaş, **Yılmaz Hülya**. (2015). Comparison of the Estimators of the Location and Scale Parameters Under the Mixture and Outlier Models via Simulation, *Ege University Journal of the Faculty of Science*, 39(1), 20-38.

B2. Cemalettin Elmira, Ünlüoğlu İlhami, Bilge Uğur, Akalın Aysen, **Yılmaz Hülya** (2017). Tıp 2 Diabetes Mellituslu Hastalar Tarafından Başvurulan Hekim Sayısının Hedef Değerlere Ulaşma ve Tedaviye Uyum Oranları Üzerine Etkileri. *Konuralp Tıp Dergisi*, 9(1), 1-6., Doi: 10.18521/ktd.296809

B3. Özkaraman Ayşe, Orulu, Nihal, Atalay Hatice Özden, Babadağ Burcu, **Yılmaz Hülya**, Dinçer Murat, Alparslan Balcı Güler (2017). Oral Kemoterapi Alan Hastalarda Yaşanılan Güçlükler ve İlaç Uyumunun Değerlendirilmesi, *Bakırköy Tıp Dergisi*, 13(2), 91-97., Doi: 10.5350/BTDMJB201713206

B4. Özkaraman Ayşe, Özel Filiz, Alparslan Balcı Güler, **Yılmaz Hülya**, Kul Esra Songül, Kaşifoğlu Timuçin, Korkmaz Cengiz (2017). Evaluation of the Body Image and Self-Esteem in Behcet's Syndrome. ACU Sağlık Bil Derg (4), 226-231

B5. Sırmatel Pınar, **Özen Hülya**, Gündüz Eren (2017). Non-malignant late effects in lymphoma patients treated with autologous hematopoietic stem cell transplantation. European Journal of Therapeutics, 23(3), 106-110., Doi: 10.5152/EurJTher.2017.66

B6. Aydın Hasan Emre, Kaya Ismail, **Özen Hülya**, Kızmaçoğlu Ceren, Uzunoğlu İnan (2018). Anti-inflammatory efficiency of erythropoietin in sciatic nerve damage. Journal of Clinical and Analytical Medicine, 4(9), 287-291., Doi: 10.4328/JCAM.5673

B7. Özdemir Zeynep Canan, Deniz Asli, Düzenli Kar Yeter, **Özen Hülya**, Bör Özcan (2018). Can Platelet Indices and Serum Lactate Dehydrogenase Levels be Used for the Differential Diagnosis of Malignancy in Children with Lymphadenopathies?. Istanbul medical journal, 19(2), 134-137., Doi: 10.5152/imj.2017.1357

B8. **Özen Hülya**, Bal Cengiz (2019). A Study on Missing Data Problem in Random Forest. Osmangazi Journal of Medicine, Doi: 10.20515/otd.496524

**ULUSAL VE
ULUSLARARASI
SÖZLÜ
BİLDİRİLER**

C1. **Yılmaz Hülya**, Bal Cengiz, Özdamar Kazım, Kayıp Veri İçeren Bir Sınıflandırma Probleminde Random Forests ve K En Yakın Komşu Yöntemlerinin Kayıp Değer Atama Bakımından Karşılaştırılması. XVI. Ulusal Biyoistatistik Kongresi, 30, 10-12 Eylül 2014, Antalya.

C2. **Yılmaz Hülya**, Colak Ertugrul, Bal Cengiz, Mutlu Fezan, Ozdamar Kazim, Is nonparametric bootstrap an appropriate technique for estimating variance of sample median?, 36th Annual Conference of the International Society for Clinical Biostatistics (ISCB 2015), 75, August 23-27 2015, Utrecht/The Netherlands.

C3. Eyiş, Esra, Kaygısız Bilgin, **Yılmaz Hülya**, Ayhancı Adnan, Kılıç Fatma Sultan, Verbascum exuberans hub.-mor. türünün antinosiseptif etkisi ve bu etkide mekanizmaların incelenmesi, 23. Ulusal Farmakoloji Kongresi, 99, 7-10 Eylül 2015, Ankara.

C4. Sazak Hakan Savas, **Yılmaz Hülya**, Comparison of the Robust Estimators of Location and Scale under Various Situations via Simulation, İSTKON9, 9. Uluslararası İstatistik Kongresi, 58-59, 28 Ekim-1 Kasım 2015, Antalya/Türkiye.

C5. Özkaraman, Ayşe, Özel Filiz, Balcı Alparslan Güler, **Yılmaz Özen Hülya**, Kul Esra Songül, Kaşifoğlu Timur, Korkmaz Cengiz, Behçet Hastalıklı Bireylerde Beden İmağı ve Özgüven, 1. Uluslararası Sağlık Bilimleri Kongresi, 267, 29 Haziran-1 Temmuz 2017, Aydın.

C6. Özkaraman Ayşe, Düşüm Özlem, **Yılmaz Özen Hülya**, Usta Yeşilbalkan Öznur, Kemoterapi Alan Hastalarda Lavanta Kokusunun Kaygı ve Uyku Kalitesine Etkisi, 1. Uluslararası Sağlık Bilimleri Kongresi, 736, 29 Haziran-1 Temmuz 2017, Aydın.

C7. **Özen Hülya**, Bal Cengiz, Özen Doğukan, Öner Kevser Setenay, A General Overview Of The Working Life Of Women In Turkey And

Europe With Statistics. 1st International Woman Symposium, 110-111. October 13-14 2017, Eskisehir.

C8. Özen Hülya, Bal Cengiz, Orantılı Hazard Durumunun İhlalinde Log-Rank Testinin Gücünün İncelenmesi. XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi, 12, 25-28 Ekim 2017, Antalya.

C9. Özdemir Zeynep Canan, Düzenli Kar Yeter, Canik Ağgöl, Küskü Kiraz Zeynep, **Özen Hülya**, Bör Özcan, Pediatrik Febril Nötropenide Kan dolaşımı Enfeksiyonlarını Öngörmede Prokalsitonin, C-reaktif protein, Presepsin ve sTREM-1 (Solubl-myeloid hücrelerde eksprese edilen tetikleyici reseptör) düzeylerinin önemi, 11.Ulusal Çocuk Enfeksiyon Hastalıkları Kongresi, 5, 4-8 Nisan 2018, Antalya.

**ULUSAL VE
ULUSLARARASI
POSTER
BİLDİRİLER**

D1. Yılmaz Hülya, Sazak Hakan Savaş, Bozulmuş normal dağılım altında konum ve dağılım parametre tahmin edicilerinin etkinliklerinin benzetim yoluyla karşılaştırılması, Uluslararası Katılımlı 15. Ulusal Biyoistatistik Kongresi, 94, 20-23 Ağustos 2013, Aydın.

D2. Erdinç Oğuz O., Ertan Bengü, Uzuner G., **Yılmaz Hülya**, Oner Setenay, Autonomic Functions in Adult Patients with Moderate And Severe Obstructive Sleep Apnea Syndrome, Sleep 2014 28th Annual Meeting of the Associated Professional Sleep Societies, 134, May 31 2014 –June 4 2014, Minnesota/United States of America

D3. Eyiis Esra, Kaygisiz Bilgin, **Yılmaz Hülya**, Ayhancı Adnan, Kiliç Fatma Sultan, The antinociceptive effects of verbascum exuberans hub.-mor, 17th World Congress of Basic & Clinical Pharmacology Congress, 156, July 13-18 2014, Cape Town/South Africa.

D4. Durak Aras Beyhan, Işık Sevgi, Töre Tolga, Canaz Funda, **Yılmaz Hülya**, Özdemir Muhsin, Çilingir Oğuz, Aslan Hüseyin, Artan Sevilhan, Study of fluorescence in situ hybridization in malig-nant melanoma, 10th European Cytogenetics Conference, 103, July 4-7 2015, Strasbourg/France.

D5. Colak Ertugrul, **Yılmaz Hülya**, Oner Setenay, Pairwise multiple comparisons for survival functions with right censored data, 36th Annual Conference of the International Society for Clinical Biostatistics (ISCB 2015), 138, August 23-27 2015, Utrecht/The Netherlands.

D6. Toprak Çiğdem, Sırmagül Başar, Yiğitarıslan Semra, Özakın Engin, **Yılmaz Hülya**, Bir üniversite hastanesine başvuran ilaç zehirlenmelerinin 1 yıllık analizi, 23. Ulusal Farmakoloji Kongresi, 209, 7-10 Eylül 2015, Ankara.

D7. Eyiş Esra, Kaygısız Bilgin, **Yılmaz Hülya**, Kılıç Fatma Sultan, Ayhancı, Adnan, Verbascum exuberans hub.-mor. türünün antiinflamatuvar etkisi, 23. Ulusal Farmakoloji Kongresi, 188, 7-10 Eylül 2015, Ankara.

D8. Cemalettin Elmira, Unluoglu Ilhami, Bilge Ugur, Unalacak Murat, Akalin Aysen, **Yılmaz Hülya**, Number of physicians visited, self-rated level of therapy adherence and metabolic control in people with T2DM: a descriptive study, 14th International Primary Care Diabetes Europe Conference, 44, April 29-30 2016, Barcelona/Spain.

D9. Kilic Fatma Sultan, Kaygisiz Bilgin, Aydin Şule, Yildirim Engin, **Yilmaz Hulya**, Erol Kevser, The effects and mechanism of action of galangin on spatial memory in the Morris water maze test in rats, 30th Seoul World Congress of the International-College-of-Neuropsychopharmacology (CINP), 45, July 3-7 2016, Seoul/South Korea.

D10. Dal Pınar, Dincer Murat, Yildiz B., Kahraman O., Canhoroz M., Arik D., Başeskioglu B., **Yilmaz Hulya**, Oxaliplatin and Erectile Dysfunction, 41st ESMO Congress (ESMO 2016), 89, October 7–11 2016, Copenhagen, Denmark.

D11. Kolaç Umut, Erdoğan Orhan İlkay, **Özen Hülya**, Çağlar Selim Süalp, Insecticidal Activities of Different Salvia Species, 3rd International Congress on Applied Biological Sciences, 34, July 09-12, 2017, Afyon.

D12. Sener Ceylin, Yelken Tuna, **Ozen Hulya**, Ozbabalik Demet, The Effect of Listening to Rhythmic Music on Brain Training Game Scores in Patients with Alzheimer's Dementia, Alzheimer's Association International Conference 2017, 1413-1414, 14-20 July 2017, London/United Kingdom.

D13. Kaltuş Zühal, Şöhret Nurullah Cihan, Bozkurt Ezgi, Çarman Kürşat Bora, Yarar Coşkun, **Özen Hülya**, Harmancı Nuşin, Yiğitaslan Semra, Çocuklarda Yaş ve Eşzamanlı Antiepileptik Kullanımının Levetirasetam Farmakokinetiğine Etkisinin Değerlendirilmesi, 24. Ulusal Farmakoloji Kongresi, 177, 17-20 Ekim 2017, Trabzon.

D14. Özen Doğukan, **Özen Hülya**, Ambarcıoğlu K. Pınar, Kaya Ufuk, Gülendağ Erman, Gürcan İ. Safa, Yaşam Analizlerinde Esnek Parametrik Bir Yaklaşım: Royston-Parmar Modelleri, 87, XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi, 12, 25-28 Ekim 2017, Antalya.

D15. Sener Ceylin, Sener Cansu, **Ozen Hulya**, Ozbabalik Demet, Blood Pressure and Pulse Reactions of Alzheimer's Dementia Patients to Different Emotional Stimuli, Alzheimer's Association International Conference 2018, 1466, July 22-26 2018, Chicago, Illionis/United States of America.

D16. Sener Ceylin, Sener Cansu, **Ozen Hulya**, Ozbabalik Demet, Emotional Reactions of Alzheimer's Dementia Patients to Different Genres of Film Clips, Alzheimer's Association International Conference 2018, 1182-1183, July 22-26 2018, Chicago, Illionis/United States of America.

ARAŞTIRMA PROJELERİ

D1. Pregabalinin Deneysel İnflamasyon Modelindeki Antiinflamatuvar Etkisi, Eskişehir Osmangazi Üniversitesi Bilimsel Araştırma Projeleri, Araştırmacı, Proje Bitiş Tarihi: 21.11.2016, Proje No: 201611D01, Proje Yürütücüsü: Prof. Dr. Fatma Sultan Kılıç

D2. Ratlarda Karbontetraklorür (CCl₄) ile İndüklenmiş Karaciğer Hasarında Naringenin Endoplazmik Retikulum Stresi Üzerine Etkileri (Effects of Naringenin on Carbon Tetrachloride (CCl₄) Induced Liver Injury and Endoplasmic Reticulum Stress in Rats), Eskişehir Osmangazi Üniversitesi Bilimsel Araştırma Projeleri,

Arařtırmacı, Proje Bitiř Tarihi: 17.08.2018, Proje No: 201711D26,
Proje Yürütücüsü: Doç. Dr. M. Cengiz ÜSTÜNER

**KURS
SERTİFİKA
BİLGİSİ**

E1. Ege Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı tarafından düzenlenen “Epidemiologic Methods” Konulu Yaz Okulu, 4-6 Haziran 2012, Çeřme/İZMİR

E2. Uluslararası Katılımlı XIV. Ulusal Biyoistatistik Kongresi kapsamında düzenlenen “Veri Madencilięi ve Uygulamaları” Konulu Kurs, 3 Eylül 2012, KAYSERİ

E3. Ege Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı tarafından düzenlenen “Mixed Models with Medical Applications” Konulu Yaz Okulu, 27-28 Mayıs 2013, Çeřme/İZMİR

E4. Ege Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı tarafından düzenlenen “Generalized Linear Mixed Models with Applications in Medicine” Konulu Yaz Okulu, 26-27 Mayıs 2014, Çeřme/İZMİR

E5. Eskiřehir Osmangazi Üniversitesi Saęlık Bilimleri Enstitüsü tarafından düzenlenen “Bilgisayar Uygulamalı Yapısal Eřitlik Modelleri” konulu kurs, 26-28 Ağustos 2014, ESKİŐEHİR

E6. Eskiřehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı tarafından düzenlenen “Bilgisayar Uygulamalı Yapısal Eřitlik Modellemesi” konulu kurs, 18-20 Şubat 2015, ESKİŐEHİR

E7. 19. Ulusal ve 2. Uluslararası Biyoistatistik Kongresi kapsamında düzenlenen “ R ile Kestirime Yönelik İstatistiksel Modellerin Oluřturulması ” Konulu Kurs, 27 Ekim 2017, ANTALYA

