

T.C.
ÇUKUROVA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

**R PROGRAMI KULLANILARAK BAĞIMLI VE BAĞIMSIZ
GERÇEK VE YARATILMIŞ VERİLERDE YARIŞAN RİSKLERİN
DEĞERLENDİRİLMESİ**

Gözde ERTÜRK

BİYOİSTATİSTİK YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

DANIŞMANI
Prof. Dr. Hüseyin Refik BURGUT

ADANA-2015

T.C.
ÇUKUROVA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

**R PROGRAMI KULLANILARAK BAĞIMLI VE BAĞIMSIZ
GERÇEK VE YARATILMIŞ VERİLERDE YARIŞAN RİSKLERİN
DEĞERLENDİRİLMESİ**

Gözde ERTÜRK

BİYOİSTATİSTİK YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

DANIŞMANI
Prof. Dr. Hüseyin Refik BURGUT

Bu tez, ÇÜ Rektörlüğü Bilimsel Araştırma Projesi Birimi tarafından TYL-2014-2986
nolu proje olarak desteklenmiştir.

ADANA-2015

KABUL VE ONAY

TEŞEKKÜR

Bu araştırmanın yapılmasında ve yürütülmesinde desteğini esirgemeyen, yapıcı ve yönlendirici fikirleri ile bana daima yol gösteren, bilgi ve deneyimlerinden yararlandığım danışman hocam Sayın Prof. Dr. Refik Burgut'a teşekkür ederim.

Eğitim ve tez çalışma süresince bilgi ve deneyimlerinden yararlandığım her türlü desteği benden esirgemeyen değerli hocam Sayın Prof. Dr. Zeliha Nazan Alparıslan'a teşekkür ederim.

Yüksek Lisans çalışmalarım süresince değerli bilgi ve becerilerinden yararlandığım çalışmamda büyük emeđi geçen desteklerini benden hiç esirgemeyen değerli hocalarım Öğr. Gör. Dr. Yaşar Sertdemir ile Yrd. Doç. Dr. İlker Ünal hocama teşekkür ederim.

Maddi ve manevi desteklerini benden hiçbir zaman esirgemeyen canım babam Seyit Ertürk'e ve canım annem Fitnet Ertürk'e teşekkürlerimi sunarım.

İÇİNDEKİLER

KABUL VE ONAY	iii
TEŞEKKÜR	iv
İÇİNDEKİLER	v
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ	viii
ÖZET	ix
ABSTRACT	x
1.GİRİŞ	1
1.1 HİPOTEZLER VE ÇALIŞMANIN AMACI	3
2.GENEL BİLGİ	4
2.1. AÇIKLAYICI TERİMLER	6
2.2. SAĞ KALIM ANALİZİ	7
2.2.1. Parametrik Yaklaşımlar	8
2.2.2. Parametrik Olmayan Yaklaşımlar	13
2.3. SAĞ KALIM ANALİZİNDE KAPLAN-MEİER YAKLAŞIMI	16
2.3.1. Log-Rank ve Wilcoxon Testleri	17
2.4. SAĞ KALIM ANALİZİNDE COX YAKLAŞIMI	21
2.4.1. β Katsayılarının Önemliliğinin Test Edilmesi	23
2.4.2. Wald Testi	23
2.4.3. Olabilirlik Oranı Testi(LRATIO)	23
2.4.4. Skor Testi (Rao's Score Test)	24
2.4.5. Cox Regresyon Modelinde Kullanılan Oranlar	25
2.4.6. Modelin Önemliliğinin Test Edilmesi	25
2.5. YARIŞAN RİSKLER ANALİZİ (COMPETING RISKS ANALYSIS)	26
2.5.1. Yarışan Risk Durumunda Kümülatif İnsidans Foksiyonu Yaklaşımı (CIF)	29
2.5.2. Kümülatif İnsidans Fonsiyonu (CIF) Non-Parametrik Maximum-Olabilirlik Tahmin Edicileri	30
2.5.3. Kümülatif İnsidans Fonsiyonu (CIF) Tahmin Edicileri için Varyans ve Güven Aralığı	31
2.5.4. "Proportion Hazard"(Oransal Ölüm Olasılığı) Modeli	33
2.5.5. Eş değişken Testi	33
2.6. YARIŞAN RİSKLERİN TANIMLANMASINDA KULLANILAN REGRESYON MODELLERİ	34
2.6.1. Ölüm Olasılığına Dayalı Regresyon Modelleri	34
2.6.1.1. Nedene Özel Ölüm Olasılığı Regresyonu (Cause-Specific Hazard Regression)	34
2.7. SAĞ KALIM VERİSİ ÜRETME	38
2.7.1. Sağ kalım Analizlerinde Veri üretiminde kullanılan Fonsiyonlar	38
2.7.2. Üstel Dağılımla Veri Üretme	39
2.7.3. Weibull Dağılımından Veri Üretme	40
3.GEREÇ VE YÖNTEM	42

3.6.5. Simülasyon Düzeni	45
3.6.6. Alt Dağılım Ölüm Olasılığı Modelinde Simülasyon	46
3.6.8. Analiz ve Simülasyonda Kullanılan Paket Programlar	47
4. BULGULAR	48
4.1. FOLLİKÜLER HÜCRE LENFOMASI VERİ SETİ	48
4.1.1. Gerçek Veri setinde Uygulama	48
4.2. WEİBULL DAĞILIMINDAN ÜRETİLEN SİMÜLASYON SONUÇLARI	56
4.3. ÜSTEL DAĞILIMINDAN ÜRETİLEN SİMÜLASYON SONUÇLARI	62
4.4. GERÇEK VERİ SETİNDE EŞ DEĞİŞKEN OLMASI DURUMUNDA SONUÇLAR VE YORUMLAR	71
5. TARTIŞMA	74
6. SONUÇLAR VE ÖNERİLER	77
7. KAYNAKLAR	79
EKLER	83
Ek 1. GERÇEK VERİ SETİNDE KAPLAN MEİER VE YARIŞAN RİSKLERİN DEĞERLENDİRİLMESİ İÇİN KULLANILAN R KODU	83
Ek 2. KÜMÜLATİF ÖLÜM OLASILIKLARI GRAFİKLERİ OLUŞTURMA DEĞERLENDİRİLMESİ İÇİN KULLANILAN R KODU	85
Ek 3. WEİBULL DAĞILIMINDAN VERİ ÜRETME VE KARŞILAŞTIRMA KODLARI DEĞERLENDİRİLMESİ İÇİN KULLANILAN R KODU	86
Ek 4. OLAYA KARAR VERMEDE BİNOMİAL DENEMELER İÇİN KULLANILAN R KODU	88
Ek 5. ÜSTEL DAĞILIMINDAN VERİ ÜRETME VE KARŞILAŞTIRMA DEĞERLENDİRİLMESİ İÇİN KULLANILAN R KODU	91
Ek 6. GERÇEK VERİ SETİNDE EŞ DEĞİŞKENLERİN DEĞERLENDİRİLMESİ İÇİN KULLANILAN R KODU	93

ŞEKİLLER DİZİNİ

Şekil 2.1.Belirli bir hata için gösterilen K farklı risk için yarışan risk modeli	6
Şekil 3.3.1.1.Yarışan Riskler	43
Şekil 4.4.1.1.1. Gerçek Veri Seti Diagramı	50
Şekil 4.4.1.1.2. Kümülatif Ölüm Olasılığı Fonksiyonunun Hastalığa Bağlı ölümler ile Hastalık Dışı Durum grafiği	544
Şekil 4.4.1.1.3.Hastalığa Bağlı Ölümlerde CR ve KM sağ kalım olasılıkları	55
Şekil 4.4.1.1.4.Hastalık Dışı Ölümlerde CR ve KM sağ kalım olasılıkları	56
Şekil 4.4.2.1.Hastalık Dışı Ölümlerde Kümülatif Ölüm Olasılığı Grafiği	60
Şekil 4.4.2.2.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde Örneklem Sayısına Bağlı CR-KM Farkının Ortalaması Grafiği	61
Şekil 4.4.3.1.Hastalık Dışı Ölümlerde $\lambda_1 =5, \lambda_2 =5$ Kümülatif Ölüm Olasılığı Grafiği	63
Şekil 4.4.3.2.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 =5, \lambda_2 =5$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalaması Grafiği	64
Şekil 4.4.3.3.Hastalık Dışı Ölümlerde $\lambda_1 =5, \lambda_2 =6$ Kümülatif Ölüm Olasılığı Grafiği	65
Şekil 4.4.3.4. Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 =5, \lambda_2 =6$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalaması Grafiği	66
Şekil 4.4.3.5.Hastalık Dışı Ölümlerde $\lambda_1 =5, \lambda_2 =7$ Kümülatif Ölüm Olasılığı Grafiği	67
Şekil 4.4.3.6.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 =5, \lambda_2 =7$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalaması Grafiği	68
Şekil 4.4.3.7.Hastalık Dışı Ölümlerde $\lambda_1 =5, \lambda_2 =8$ Kümülatif Ölüm Olasılığı Grafiği	69
Şekil 4.4.3.8.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 =5, \lambda_2 =8$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalaması Grafiği	70
Şekil 4.4.4.1.Yaşın Hastalığa Bağlı Ölümlerde Kümülatif İnsidans Fonksiyonuna göre Ölüm olasılıkları Grafikleri	72
Şekil 4.4.4.2.Yaş Değişkeninin Hastalık Dışı Ölümlerde Yarışan Risk Durumunda Kümülatif İnsidans Fonksiyonuna Göre Ölüm olasılıkları Grafikleri	73

ÇİZELGELER DİZİNİ

Çizelge 1.2.7.1.Dağılım Tablosu	41
Çizelge 4.4.1.1.1.Foliküler Hücre Lenfoması Hastalarının Veri Seti	51
Çizelge 4.4.1.1.2.Foliküler Hücre Lenfoması Hastalarının Veri Seti	52
Çizelge 4.4.1.1.3.Hastalığa Bağlı ve Hastalık Dışı Ölümlerde Kaplan Meier ve Yarışan Risk Sonuçları	53
Çizelge 4.4.2.1.Weibull Dağılımından Üretilen Veri Seti	57
Çizelge 4.4.2.2..(Devam)	58
Çizelge 4.4.2.3.Weibull Dağılımından Üretilen Veri Sonuçlarının Tanımlayıcıları	59

ÖZET

R Programı Kullanılarak Bağımlı ve Bağımsız Gerçek ve Yaratılmış Verilerde Yarışan Risklerin Değerlendirilmesi

Klasik sağ kalım analizlerinde genellikle tek bir başarısızlık (ölüm, hastalığın nüksetmesi vb.) nedeni araştırılmaktadır. Ancak çoğu kez, başarısızlığa birden fazla faktör etki etmekte ve bu faktörlerden birisi öne çıkarak başarısızlığa neden olursa, sağ kalım analizlerinin yarışan riskler dikkate alınarak yapılması gerekmektedir. Sağ kalım analizlerinde yarışan risk verileri mevcutken Kaplan-Meier (KM) yöntemi uygulandığında yarışan risklerin göz ardı edildiği görülmektedir. Yarışan risk varlığında Kaplan Meier yaklaşımını uygulamanın hatalı sonuçlar ve yorumlar sunacağını ortaya koymak amacıyla bu çalışmada yarışan riskler analizi (CR) ile KM yöntemlerinin sağ kalım olasılıklarının farklarının ortalama değerlerini hem gerçek veri setinde hem de simüle veri setinde incelenerek değerlendirilmiştir. Simülasyon üretim kısmında farklı dağılımlarda ve farklı örneklem büyüklüklerinde senaryo sonuçları sunulmuş ve bulgular değerlendirilmiştir. Veriler R yazılım programı kullanılarak analiz edilmiş ve simülasyon çalışması için gerekli R kodları yazılmış simülasyonlar yapılmıştır.

Farklı senaryolar oluşturulurken weibull dağılımından ve üstel dağılımdan yararlanılmış, örneklem büyüklüğü $n=100, 150, 250, 500, 1000$ alınmıştır. Sonuçlar Kümülatif ölüm olasılığı grafikleri de sunularak değerlendirilmiştir. Yarışan risklerin varlığında eş değişkenleri incelerken Gray's test istatistiği uygulanmıştır. Yapılan hem gerçek veri seti üzerinde ki uygulama hem de simülasyon sonuçları yarışan risklerin varlığında genel sağ kalım analizleri kullanmak yerine yarışan riskler analizi uygulamanın doğru olacağını sunmuştur. Weibull dağılımından üretilen sonuçlarda iki yöntemin sağ kalım olasılıklarının farklarının ortalama sonuçları farklı çıkarken örneklem sayısı etkilememiştir. Üstel dağılımda bu sonuçlar sansürleme oranına bağlı olarak değiştiği ortaya konulmuştur. Sansürleme oranı arttıkça farkların azaldığı, örneklem sayısının büyümesi ile farkların arttığı gözlemlenmiştir.

Anahtar sözcükler: R programı, Yarışan riskler, Üstel Dağılım, Kümülatif Ölüm Olasılığı, Grays modeli

ABSTRACT

Evaluation of Competing Risks Based On Both Dependent , Independent Real And Simulated Data By Using Self Developed R Program

Classical survival analysis methods commonly evaluate single cause of failure (die, relapse, etc.). However, failures may be of several distinct causes and the observation of any one of them prevents the observation in others. In this situation, survival analysis must be done by considering competing risks. In survival analysis when several competing risks exist, Kaplan Meier (KM) method ignores the competing risks (CR) aspect of the data. In order to show KM approach gives insufficient results and inferences when there is competing risks, we studied mean differences in survival probability which are both analyzed by CR and KM approaches separately in real data set and simulated data sets. R statistical software is used to simulate and analyze data sets. During simulation data, we used different sample sizes ($n=100, 150, 250, 500, 1000.$) in two different distributions (weibull and exponentials distributions) so that different scenarios are presented to give corresponding outcomes and inferences. Outcomes and graphs for Cumulative hazard function are presented. We used Grays test statistics to study covariate effects when competing risks exists. Both real and simulated data sets shows that when there is competing risks competing risks approach gives more appropriate results than classical survival analysis. When weibull distributions are used mean differences in survival probabilities of two methods are different but not effected by sample sizes. But when exponentials distributions are used mean differences in survival probabilities depend on censoring rate and sample sizes. When censoring rates gets higher we observed that mean differences in survival probabilities of two methods is decreased. As the sample sizes gets higher we see that mean differences in survival probabilities of two methods is increased.

Keywords: R program, Competing Risks, Exponential Distribution, Cumulative Hazard Function, Gray's models

1.GİRİŞ

Sağlık düzeyini gösteren en önemli ölçütlerden biri yaşam süresi (survival time) ve modern tıbbın en önemli amaçlarından biri de bu süreyi uzatmaktır. Bu amaçla bireylerin sağlıklarının korunması, hastaların sağlıklı bir yaşama kavuşması ve uzun süre yaşamalarını sağlayıcı tedavi yöntemlerinin gelişmesi çalışmaları her geçen gün artmaktadır.^{1,2}

. Bir hastalığa yakalanan kişinin hangi olasılıkla iyileşeceği, hangi olasılıkla öleceği ya da hangi olasılıkla ne zaman öleceği istatistiksel yöntemler kullanılarak tahmin edilebilir.^{1,3}

Hastalık ve ölüm verileri kullanarak insan sağlığını etkileyen etmenlerin yaşam süresi üzerine etkileri sağ kalım analizi ile (Survival Analysis) incelemektedir.²

Sağ kalım analizinin temel değişkeni sağ kalım süresidir. Sağ kalım süresi, belirli bir tedavi verilen hasta grubunda tedavinin başlangıcından araştırmanın sonuçlanmasına kadar geçen süredir. Hastalık olgularında hastaların herhangi bir medikal ya da cerrahi girişimde (ilaç ile tedavi, ilaç+ameliyat vb.) sonraki yaşamlarının izlem zamanlarını; izlem süresi, yaşam süresi, lezyonsuz geçen süre, remisyon zamanı gibi isimlerle belirtilir. Bu süreler, izleme zamanına ilişkin gün, ay, yıl, gibi süreleri içerir.²

Yaşam fonksiyonu ya da yaşam eğrisi, yaşam sürelerinin olasılık dağılımıdır. Fonksiyon, yaşamsal verilerin genel eğilimini, matematiksel bir modelle ifade etmektedir. Sağ kalım analizi yöntemleri aracılığıyla izlem sürelerini ve diğer faktör değişkenleri içeren veri setlerinde yaşam olasılıkları, ölüm olasılıkları, ortalama yaşam süresi, ortanca yaşam süresi vb. ile ilgili tahminler yapılır. Yaşam süresi içerisinde incelenen grupta hastalığın yaşam süresine etkisi, hastalığın tekrar ne zaman olacağını, tedavi tiplerinin ve diğer faktörlerin yaşam süresine etkilerinin analizidir.^{2,4}

Yaşam analizleri;

- Belirli bir hastalığa maruz kalan ve tedavi edilen bir bireyin hastalığının ne kadar süre sonra nüksedebileceğini tahmin etmek,

- Tedaviye alınan bir grup hastanın tedavi sonrasında ne kadar yaşayabileceğini belirlemek
- Hastalıkların, tedavi tiplerinin, prognostik ve diğer faktörlerin yaşam sürelerine etkilerini belirlemek
- Hastalıkların ölüm riskleri ve bu riskler üzerinde etkide bulunan prognostik faktörlerin etkilerini saptamak,
- T zaman süresinde n sayıda izlenen hasta birimden elde edilen yaşam sürelerinin dağılımını açıklayarak, yaşam süresini etkileyen ve etkilemesi olası değişkenleri içeren modeller kurmak ve bu modellere parametre tahminleri yapmak amacıyla geliştirilmiştir.^{2,5,6,7}

Klasik sağ kalım analizlerinde genellikle tek bir başarısızlık (ölüm, hastalığın nüksetmesi vb.) nedeni araştırılmaktadır. Ancak bazen hastaların birçok başarısızlık nedenine maruz kaldığı durumlar söz konusu olabilir. Böyle bir durumda başarısızlık nedenlerinden sadece birini dikkate alıp diğer başarısızlık nedenlerinin göz ardı edilmesi doğru bir karar değildir. Bütün başarısızlık nedenlerinin bir arada düşünülmesi ve incelenmesi gerekir. Başarısızlığa birden fazla faktör etki eder ve bu faktörlerden birisi öne çıkarak başarısızlığa neden olursa ve diğer faktörlerin öne çıkmasını engellerse buna "yarışan risk" adı verilmektedir.

Yarışan risklerin anlaşılabilmesi için birkaç örnek aşağıda verilmiştir.

1. Bir kişiye ölmeden önce hem akciğer kanseri hem ateroskleroz tanısı konmuş olabilir fakat ölüm sebebi bunlardan sadece biridir.
2. İleri-derece kanser hastalarında ölüm sebebi kanser değil, geçirdiği ameliyat sonrasında hastane ortamında vücuduna yerleşen hastane enfeksiyonu olabilir.
3. Bir klinik çalışmada metastazı olmayan ekstremitte sarkom hastalarında, uygulanan ilaç tedavisi ve cerrahi müdahale sonrasında lokal nüks, akciğer metastazı veya diğer organ metastazları gelişebilir.
4. Kemik iliği nakli sonrasında hasta, farklı türde enfeksiyonlar (bakteriyel, viral, fungal) sebebiyle ya da nüks,"graft-versus-host" hastalığı (GvHD) veya diğer nedenlerden dolayı ölebilir.
5. Nod negatif akciğer kanserinde Tamoxifen tedavisi uygularken çalışmanın sonuna gelene kadar farklı durumlarla karşılaşmıştır. Tedavinin ilk yılında

bölgesel yayılma, ikinci yılda bölgesel yayılma, üçüncü yılda koltuk altına yayılma, dışarı yayılma, ikincil kanser olma ve ölümlerle sonlanırken bütün durumları yarışan risk olarak değerlendirmek gerekir.

1.1 Hipotezler ve Çalışmanın Amacı

H₀: Yarışan risklerin varlığında yarışan risk modellerinin veya yarışan risk tahminlerinin kullanımının risklerin bağımsız olduğu durumlar için yapılan değerlendirmelere göre aynı bilgiyi içermektedir

H₁: Yarışan risklerin varlığında yarışan risk modellerinin veya yarışan risk tahminlerinin kullanımının risklerin bağımsız olduğu durumlar için yapılan değerlendirmelere göre daha fazla bilgiyi içermektedir.

Çalışmanın amacı;

- a) Değişik sayıda yarışan risklerin olduğu durumlarda her bir yarışan risk için bir bağımsızlık varsayımı altında yaşam fonksiyonu ölüm olasılığı oranı, bu riskleri etkileyen eş değişkenlerin Kaplan Meier ve Kümülatif İnsidans fonksiyonu altında alt dağılım ölüm olasılığı ile durumlar incelenerek etkilerini (R programı yardımı ile değerlendirmek),
- b) Geliştirilen R programı kullanarak gerçek veya simüle edilmiş verilerde, risklerin bağımlı oldukları varsayım altında non-parametrik ölüm olasılığı oranı, yaşam fonksiyonu, kümülatif insidans fonksiyonunu (değerlendirmek ve)
- c) Eş değişkenlerin Kümülatif insidans fonksiyonu üzerindeki etkilerini değişik modellerle gerçek veya simüle veri kullanarak yazılan (R programı ile değerlendirmek.)

Bu çalışmada hastalarda tedavinin ve prognostik faktörlerin etkilerini yarışan risk modeli kullanarak, yarışan risk modelleri kullanımının bağımsız risk modelleri kullanımına göre ilave bilgiler sağlayıp sağlamadığını göstermektir. Sonuçta, çalışma sayesinde tedavinin daha kapsamlı etkilerinin incelenebileceği modellerin varlığını göstermektedir.

2.GENEL BİLGİ

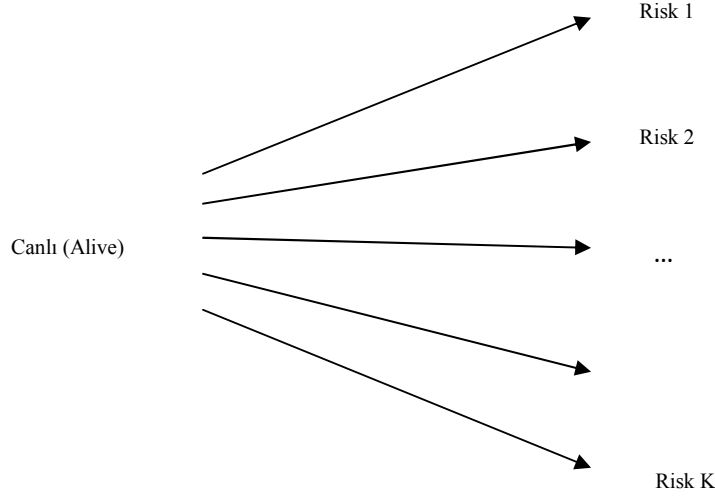
Sağ kalım analizleri son yıllarda oldukça büyük gelişmeler göstermiştir. Bu alandaki en önemli gelişmeler Kaplan-Meier'in 1958'de yaşam dağılımlarının tahmini için geliştirdiği yöntem, Mantel'in 1966'da geliştirdiği iki yaşam dağılımını karşılaştırmaya olanak sağlayan Log-Rank Testi, Cox'un 1972 yılında yaşam süresini etkileyen eş değişkenlerin etkisini değerlendirmek amacı ile geliştirdiği oransal riskler modelidir.⁷

Sağ kalım analizlerinde uygulanan bir tedavinin başarılı olup olmadığı ya da tedavi risklerinin ortaya konması amacıyla klinik denemelerde bir sonuç- ölüm ya da iyileşme – elde edildikten sonra tedavi başlangıcından tedavi sonuna kadar geçen süreler incelenir.

Sağ kalım analizi, tedavi grubuna ilişkin sürelerin analizi ile incelenen hastalığa maruz kalan bir bireyin hastalığının tanısından sonra, uygulanan bir girişimi (tıbbi tedavi, operasyon, kemoterapi vb.) müteakip daha ne kadar yaşayabileceğini ya da hastalığın ne kadar sürede nüksedebileceğini tahmin etmek tedavi tiplerinin ve diğer faktörlerin yaşam süresine etkilerini incelemek amacıyla geliştirilmiştir.^{2,4,8}

Klinik denemelerde çalışma uzun süreler ele alındığında risk faktörlerinin, zamanın ve hastalıktan ölüm nedeninin yanı sıra diğer nedenle de ölüm üzerinde etkili olabilir. Bu durumda ölüm nedenleri arasında bir yarış söz konusudur. Hangi ölüm nedeninin ölüm üzerinde daha etkili olduğu ya da öne çıktığı, eş değişkenlerden hangisinin yaşam süresini kısalttığı ya da uygulanan tedavinin türleri arasında hangisinin tedavinin başarısına etkisi olduğu yarışan riskler analizi ile bulunabilir.⁸

Yarışan riskler, başarısızlık için birden fazla nedenin mümkün olması halinde ortaya çıkar. Bu nedenlerin bir sonucu olarak ölümü düşünürsek, bunlardan sadece birinin ölümün gerçek sebebi olacağını söyleyebiliriz. Diğer nedenler de birer risktir fakat birincil ölüm sebebiyle sadece ilişki içindedir ve riskin büyüklüğü bu öncül riskten küçüktür. Aşağıdaki şekilde bir ilk durum ve farklı sonuçların uç dizisinin gösterildiği bir yarışan risk modeli temsil edilmektedir.



Şekil 2.1. Belirli bir hasta için gösterilen K farklı risk için yarışan risk modeli

Gooley ve arkadaşları yarışan risklerin; ilgilenilen olayın gözlenen bazı sonuçlarının ortaya çıkması olasılığını değiştiren ya da tamamen ortadan kaldıran bir etken olduğunu söylemişlerdir.⁹ Yarışan risklerden birinin ortadan kaldırılması sonrasında başarısızlık olasılıkları tahmini sorunu büyük önem taşımış ve 1970'lerde tartışma konusu olmuştur.

Bu alanda yapılan çalışmalarda ölümün kısmi kaba (partial crude failure) olasılığı tartışılmıştır.¹⁰ Hastalıklara ilişkin insidans fonksiyonlarının tahmini için Gray, Pepe ve Mori, Fine ve Gray yeni teknikler geliştirmiştir.^{11,12}

Yarışan riskler analiz edilirken, parametrik ve parametrik olmayan yaklaşımlarda yararlanılmaktadır. Bu yaklaşımların seçiminde veri tipi önem taşımaktadır. Parametrik yaklaşımda, başarısızlık nedenlerinin yanı sıra yaşam süresi üzerine etkide bulunan eş değişkenler (covariates) modele katılırken, parametrik olmayan yaklaşımlardan sadece ölüm nedenleri ve ölüm süreleri dikkate alınarak risk fonksiyonu tahminleri yapılmaktadır.⁶ Parametrik yaklaşım ile ölüm nedenleri yanında yaşam süresi üzerine etkide bulunan eş değişkenler modelde analize katılır. Parametrik olmayan

yaklaşımlarda ise eş değişken tahminlerde yer almaz, ölüm nedenleri ve ölüm süreleri dikkate alınarak risk fonksiyonu tahminleri yapılır.^{5,6,13}

2.1. Açıklayıcı Terimler

Çalışmayı daha detaylı anlayabilmek için bazı kavramlar ve gösterimler aşağıda gösterilmiştir.^{2,10,13,14,15,16}

- I. Yaşam Süresi (İzlem Süresi, Survival Time): Bir bireyin belirli girişime ya da etkene maruz kaldıktan sonra iyileşmesine, hastalığı tekrarlamasına ya da ölümüne kadar geçen süreye yaşam süresi (izlem süresi) adı verilmektedir ve t_i ile gösterilmektedir.
- II. Yaşam Fonksiyonu (Survival Function, Survival Curve): Yaşam sürelerinin olasılık dağılımı fonksiyonu ya da yaşam eğrisi adı verilmektedir. Fonksiyon, yaşam verilerin genel eğilimini matematiksel bir modelle ifade etmektedir. Yaşam fonksiyonu bir olasılıktır ve $Y(x)$ ile ifade edilir.
- III. Yaşam İnsidans Fonksiyonu (Survival Density Function): Sağ olan bir kişinin t zamanı i.indisinde daha ne kadar yaşayabileceğini ortaya koyan olasılıklardır. Bu olasılıklar, belirli t_i zamanları için birlikte hesaplanmaktadır.
- IV. Ölüm olasılığı (Hazard Function, Hazard Rate, Instantaneous Death Rate): Sağ olan bir kişinin, belirli bir zamanda (anda) ölüm olasılığı, taşıdığı ölüm riskidir ve I ya da $h(t)$ ile gösterilir. Ölüm olasılığı, birim zamanda bir olayın anlık olmasını ve bir birimin t zamanına kadar hayatta kalma olasılığını verir.
- V. Ölüm İnsidans Fonksiyonu (Failure Density Function) : Sağ olan bir kişinin t zamanındaki ölüm riskini ortaya koyan olasılıklardır. Bu olasılıklar, belirli t_i zamanları için birlikte hesaplanmaktadır ve f_i ile gösterilir.
- VI. Kümülatif Yaşam Fonksiyonu (Kümülatif Yaşam Olasılığı, Cumulative Survivor Density Function): T zamanı içinde belirli bir t zamanı (anı)

için hesaplanmış olan yaşam olasılıklarının kümülatif fonksiyonudur ve Y_l ile gösterilir.

- VII. Kümülatif Ölüm Fonksiyonu (Kümülatif Ölüm Olasılıkları, Cumulative Hazard Function): T zamanı içinde belirli bir t zamanı (anı) için hesaplanmış olan ölüm olasılıklarının kümülatif fonksiyonudur ve $\Lambda(t)$ ile gösterilir.
- VIII. Nedene Özel Ölüm olasılığı (Cause- Specific Hazard Function): Sağ olan bir kişinin, belirli bir zamanda (anda) belirli bir nedenden ölüm olasılığı, taşıdığı ölüm riskidir ve $h_l(t)$ ile $l = 1, \dots, C$ ölüm nedenlerini gösterir.
- IX. Kaba Kümülatif İnsidans (Crude Cumulative Incidence): Birden fazla ölüm nedeni olduğunda belirli bir nedenden ölüm için bulunan alt insidans fonksiyonudur. $I(t)$ olarak gösterilir. Belirli bir hastalığa yakalanan kişinin, zaman ∞ olarak alındığında bu hastalıktan ölüm olasılığını (kesin ölme olasılığı) verir. $p = I(\infty)$ olarak da yazılabilir.
- X. Nedene Özel Kümülatif İnsidans (Cause Specific Cumulative Incidence): Birden fazla ölüm nedeni olduğunda diğer nedenlerin elemine edilmesiyle bir neden için bulunan Kümülatif insidansdır. $I_l(t)$ olarak gösterilir.

2.2. Sağ Kalım Analizi

Sağ kalım analizi veri yapılarına ve test edilecek hipotezlere bağlı olarak kendi içinde gruplara ayrılmaktadır. Genel Sağ Kalım analizleri; Parametrik olan ve parametrik olmayan yaklaşım olarak iki grupta incelenirken tahmin yöntemlerine göre üç başlık altında inceleyebiliriz.

- Yaşam Tablosu Yöntemi (Actuarial Life Table Method, Cutler –Ederer Method)
- Kaplan –Meier Yöntemi (Product Limit Method)
- Cox Regresyon Yöntemi (Cox Orantısal Riskler Regresyon Yöntemi, Cox Proportional Hazards Regression, Cox Regression)

Epidemiyolojik verilerin ya da yaşamsal verilerin analizinde ise Kalbfleish ve Prentice, monotonik artan, azalan ve monotonik olmayan ölüm olasılıklarının fonksiyonları için parametrik dağılımlar önermişlerdir.^{6,16}

Bu dağılımlar üstel, weibull, log-logistik, gamma ve log-normal gibi dağılımlardır.^{17,18} Sağ Kalım analizinde dağılım modeli, ölüm olasılığı $I(t)$ bir fonksiyon formu olarak tanımlanmasıyla belirlenebilir. Daha sonra sağ kalım olasılığı $S(t)$ ve t zamanında koşullu ölüm olasılığı olan $f(t)$ bulunabilir.¹⁵ Dağılım modelinin belirlenmesinden sonra model parametreleri tahmin edilir. Eğer varsayılan parametrik form verilere uygun ise bu tahminler çalışılan hasta grubunun sağ kalım ve risk tahminleri için kullanılır.

2.2.1. Parametrik Yaklaşımlar

Belirli bir gelişime tabi tutulan hastaların T izlem süresi içinde bir kısmı iyileşirken, bir kısmı ölebilir, bazı hastalar çeşitli nedenlerde kaybolabilir ya da izlem dışı bırakılabilir. Yine bir kısım hasta kısa sürede ölürken, diğerleri uzun süre hayatta kalabilir. Araştırma süresi sona erdiği için izlendikleri halde hala yaşayan hastalar olabilir. Bu nedenle izlem süreleri tamamlanmış (complete) ya da tamamlanmamış (eksik, incomplete, censored) süre olarak nitelenirler. Araştırmanın tamamlanmasından önce çalışmadan herhangi nedenle ayrılan hastalar tamamlanmamış birim olarak işlem görürler. İzlem süresini tamamlayamamış birimler incelenen sürenin dışında kalırlar ve sağdan sınırlı olarak isimlendirilirler.^{5,16}

Yaşamsal verilerde temelde iki değişken incelenir:

1. Çalışmaya alınan özel olaylar (tedavi, cerrahi müdahale vb.) arasında geçen zamanlar
2. Sonuç olayının zamanı (başarısızlık zamanı, hastalığın kötüleşme zamanı, nüks zamanı, ölüm zamanı)

Verilerin parametrik yaklaşımla analizinde hesaplamalar, fonksiyonlar ve parametre tahminleri aşağıdaki varsayımlara ve eşitliklere göre yapılır.

T rastgele deęişkeni homojen toplumda, incelenen bir olayda ölüm zamanını gösterir, $F(t)$ ise ölüm olayının t zamanına eşit veya küçük olma olasılığını verir, bu durumda ilişki;

$$F(t) = \Pr\{T \leq t\} \quad (1)$$

Şeklinde gösterilir. T rastgele deęişkenin dağılım fonksiyonu olarak isimlendirilir. Bu sürekli ve negatif olmayan deęişkendir. Tanım aralığı $0 \leq t \leq \infty$ olarak yazılabilir. Buradan,

$$F(0) = 0 \text{ ve } \lim_{t \rightarrow \infty} F(t) = 1 \quad (2)$$

olarak tanımlanabildiği için $F(\cdot)$ düzenli (tanımlı) dağılım fonksiyon olarak gösterilir.

$F(t)$ monoton artan fonksiyon olması nedeniyle t deęişkeni $t^* < t^{**}$ için

$$F(t^{**}) - F(t^*) = \Pr\{T \leq t^{**}\} - \Pr\{T \leq t^*\} = \Pr\{t^* \leq T \leq t^{**}\} \quad (3)$$

olarak yazılabilir.

Tüm t zamanları için türevi alınırsa

$$\frac{d}{dt} F(t) = f(t) \quad (4)$$

t zamanı için olasılık insidans fonksiyonunu verir. Türevi alındığında

$$dF(t) = f(t)dt \approx [F(t+dt) - F(t)] \quad (5)$$

$(t, t+dt)$ aralığında ölümün koşullu olmayan olasılığını verir.

Yaşam fonksiyonu $S(t)$, $F(t)$ 'nin tamamlayıcı fonksiyonu olarak tamamlanır ve

$$S(t) = 1 - F(t) = \Pr\{T > t\} \quad (6)$$

şeklinde gösterilir. Burada, $S(t)$, negatif olmayan, tanım aralığı $0 \leq t \leq \infty$ olarak yazılabilen t zamanında yaşam olasılığını verir. Böylece

$$F(t) dt = dF(t) = -dS(t) \approx [S(t) - S(t+dt)] \quad (7)$$

$$S(0) = 1 \text{ ve } \lim_{t \rightarrow \infty} S(t) = 0 \quad (8)$$

olarak yazılabilir.

Verilen bir t^* zamanında yaşayan bir birimin t zamanında koşullu ölüm olasılığı koşullu insidans fonksiyonu olarak tanımlanır ve

$$f(t | t^*) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{t \leq T < t + \Delta t | T > t^*\}}{\Delta t} \quad (9)$$

Yazılabilir. Bu tanım aracılığı ile $(t + t + dt)$ aralığında ölüm olasılığı;

$$f(t | t^*)dt = \frac{f(t)dt}{S(t^*)} \quad (10)$$

olarak tanımlanabilir.

Risk fonksiyonu (ölüm olasılığı) (hazard function)

$$I(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{t \leq T < t + \Delta t | T > t^*\}}{\Delta t} \quad (11)$$

olarak tanımlanır. $I(t)dt$, t zamanında yaşayan bir birimin $(t, t + dt)$ aralığında ölüm olasılığını verir. Dolayısıyla;

$$I(t) = f(t | t) = \frac{f(t)}{S(t)} \quad (12)$$

ölüm olasılığı olarak tanımlanır. Kümülatif ölüm olasılığı;

$$\Lambda(t) = \int_0^t I(u)du \quad (13)$$

olarak tanımlanır.

Bir nedenden ölüm fonksiyonunu birden fazla nedene genellediğinde bir bireyin maruz kaldığı birden fazla ölüm nedenleri $L \in \{1, 2, \dots, C\}$ ^{2,6,19} ile ölüm gerçekleştiği zaman T ile gösterilsin. Ölüm nedeni olarak birden fazla ölüm göz önüne alındığında homojen bir toplumda nedene özel ölüm olasılığı,

$$h_l(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\Pr\{t \leq T < t + \Delta t, L = l | T \geq t\}}{\Delta t} \quad (14)$$

şeklinde hesaplanmaktadır ve burada $l = 1, 2, \dots, C$ ölüm nedenlerini göstermektedir.²⁰

Kümülatif nedene özel ölüm olasılığı

$$H_l(t) = \int_0^t h_l(u) du \quad (15)$$

olarak tanımlanmakta ve bu fonksiyon aracılığı ile kaba yaşam fonksiyonu;

$$S_{cs;l}(t) = \exp[-H_l(t)] \quad (16)$$

Biçimin de tanımlanır.

Ölüm nedenlerinin bağımsız olup olmadığına bakılmaksızın yaşam fonksiyonu;

$$S(t) = \exp\left[-\sum_{l=1}^c H_l(t)\right] = \prod_{l=1}^c S_{cs;l}(t) \quad (17)$$

olarak da yazılabilir.

Diğer ölüm nedenlerinin olduğu bilindiği durumda l tipine bağlı koşullu ölüm olasılığı;

$$f_l(t)dt = S(t)h_l(t)dt \quad (18)$$

Biçiminde yazılır ve l tipine bağlı alt insidans (subdensity) fonksiyonudur.²¹

l ölüm nedeni için kaba ölüm Kümülatif insidans;

$$I_l(t) = \int_0^t f_{l0}(u)du = \int_0^t S(u)h_l(u)du \quad (19)$$

olarak tanımlanabilir.^{22, 23}

$T = t$ zamanda l ölüm nedeninin olup olmamasına bağlı olarak $Y = T$ ya da $Y = \infty$ için Y rastgele değişkenininin kaba yaşam fonksiyonu $S_l(t) = 1 - I_l(t)$ olarak yazılır.

C farklı ölüm nedeni olduğunda, $p_l = I_l(\infty)$ olarak alınır.

2.2.2. Parametrik Olmayan Yaklaşımlar

Verilerin parametrik olmayan yaklaşımla analizinde hesaplamalar, fonksiyonlar ve parametre tahminleri aşağıdaki varsayımlara ve eşitliklere göre yapılmıştır.

Parametrik olmayan yaklaşım kullanıldığında kesikli nedene özel ölüm olasılığı modeli;

$$h_i(t) = \begin{cases} h_{i,t} & t = t_i \\ 0 & t \end{cases}$$

Biçiminde yazılır. h_{li}, t_{i-1} 'in dışında herhangi bir tip ölüm nedeninden bağımsız olarak verilen bir birimin $t_{(i)}$ zamanında l tipi için ölüm olasılığını verir.¹⁶

d_{li}, t_i de l tipi için bireysel ölümlerin sayısını gösterebilir, $d_i = \sum_{j=1}^i d_{lj}$ olarak yazılır.

Burada n_i, t_i 'de risk altındaki birey sayısını gösterir. En büyük benzerlik fonksiyonu

$$L = \prod_{i=1}^I [(\prod_{l=1}^c h_{li}^{d_{li}})(1 - \sum_{l=1}^c h_{li})^{n_i - d_i}] \quad (20)$$

biçiminde yazılır ve nedene özel anlık ölümün ($h_l(t)$) MLE (Maximum Likelihood Estimations) tahmin edicisi;

$$\hat{h}_{li} = \frac{d_{li}}{n_i} \quad (21)$$

olarak bulunur. Nedene özel ölüm olasılıklarının Kümülatif fonksiyonu $H_l(t)$

$$\hat{H}_l(t) = \sum_{i: t_j \leq t} \frac{d_{li}}{n_i} \quad (22)$$

olarak hesaplanır.

L tipi için kaba Kümülatif insidans

$$\hat{I}_l(t) = \sum_{i|t_j \leq t} \hat{S}(t_{i-1}) \frac{d_{li}}{n_i} \quad (23)$$

olarak tahmin edilir.⁶

Farklı grupları ve hastalık tiplerini karşılaştırmak amacıyla kaba Kümülatif insidansın varyansı, standart hatası ile güven limitleri aşağıdaki gibi hesaplanır.

Genel olarak $\hat{I}(t)$ nokta tahmini varyansı;

$$\text{var}[\hat{I}_l(t_i)] = \text{var}\left[\sum_{a=1}^j h_{1a} \hat{S}(t_{a-1})\right] \quad (24)$$

$$= \sum_{a=1}^j \text{var}[h_{1a} \hat{S}(t_{a-1})] + 2 \sum_{a=1}^{i-1} \sum_{b=a+1}^i \text{cov}[h_{1a} \hat{S}(t_{a-1}), h_{1b} \hat{S}(t_{b-1})] \quad (25)$$

şeklinde bulunur. Varyans Delta metodu kullanılarak²⁵;

$$\begin{aligned} \text{var}[\hat{I}(t_i)] &= \sum_{a=1}^i \left\{ [\hat{I}_1(t_i) - \hat{I}_1(t_a)]^2 \frac{d_a}{n_a - (n_a - d_a)} \right\} + \sum_{a=1}^i [\hat{S}(t_{a-1})]^2 \left(\frac{n_a - d_a}{n_a} \right) \left(\frac{d_{1a}}{n_a^2} \right) \\ &\quad - 2 \sum_{a=1}^i [\hat{I}_1(t_i) - \hat{I}_1(t_a)] [\hat{S}(t_{a-1})] \left(\frac{d_{1a}}{n_a^2} \right) \end{aligned} \quad (26)$$

Bulunabilir. Burada $d_i = \sum_{l=1}^c d_{li}$ 'dir. Güven limitleri;

$$\exp\{\log \hat{I}(t) \pm Z_p [SE(\hat{I}(t))] / \hat{I}(t)\} \quad (27)$$

Şeklin de hesaplanabilir. Burada $SE(\hat{I}_1(t))$, kaba Kümülatif insidansın standart hatasıdır.

$$SE(I_1(t_i)) = \sqrt{\text{var}[\hat{I}(t_i)]} \quad (28)$$

şeklinde hesaplanır.²⁵

2.3. Sağ kalım analizinde Kaplan-Meier Yaklaşımı

Sağ kalım analizlerinde parametrik ve parametrik olmayan yaklaşımlardan yararlanılmaktadır. Kaplan-Meier yaklaşımı sağ kalım analizlerinde en çok kullanılan parametrik olmayan yaklaşımlardan biridir. Kaplan-Meier yöntemini diğer parametrik yöntemlerden ayıran en önemli özelliği, sağ kalım olasılıklarının hesaplanmasında sansürlü denekleri işleme katmamasıdır. Bu yöntemi uygularken dikkat edilmesi gerek hususlar şunlardır:

1. Deneklerin çalışmaya katılma tarihlerinin kesinlikle bilinmesi
2. Olgunun ortaya çıktığı zamanın kesinlikle bilinmesi
3. Kayıpların (sansürlerin) ortaya çıktığı zamanın kesinlikle bilinmesi
4. Olgunun ve kayıpların aynı anda ortaya çıkması gerekir.

Sansürlü bir veri setinde yaşam eğrileri parametrik olmayan tahmini K-M (Kaplan – Meier) metodu kullanılarak elde edilmektedir. Bu yaklaşımla bir bireyin, hayatta kalma süresinin sansürleme mekanizmasından bağımsız olacağı varsayılmaktadır. Birden fazla bağımsızlık riski olduğunda K-M yönteminde ilgilenilen başarısızlık riski dikkate alınmakta ve diğer başarısızlık riski sansürlü veri olarak değerlendirilmektedir. Bu özelliğinden dolayı gerçek değer üzerinde tahminler vermektedir.

$$\begin{aligned}
KM_1(t_k) &= \prod_{j=1}^k \frac{n_j - d_{1j}}{n_j} \\
&= \frac{n_k - d_{1k}}{n_k} KM_1(t_{k-1}) \\
&= KM_1(t_{k-1}) - \frac{d_{1k}}{n_k} KM_1(t_{k-1})
\end{aligned} \tag{29}$$

n_j = Risk altındaki kişi sayısı

2.3.1. Log-Rank ve Wilcoxon Testleri

Log- Rank ve Wilcoxon testleri, iki veya daha fazla yaşam dağılımının eşitliliği ile ilgilenen testlerdir. Yaygın şekilde bu iki test tercih edilir Mantel-Haenszel testleri olarak da bilinen Wilcoxon testleri insidansla tercih edilmektedir. İki grup arasında karşılaştırmada zaman olarak $t_1 < t_2 < t_3 < \dots < t_r$ ifade edilir ve d_{1j} başarısızlıkların sayısını verirken t_i zamanında ki, n_j grup 1’de risk sayılarıdır. d_{2j} ve n_{2j} grup 2 için ifade etmektedir. Daha sonrasında $d_j = d_{1j} + d_{2j}$ ise t_j zamanında her iki grupta başarısızlıkların sayısını temsil eder ve $n_j = n_{1j} + n_{2j}$ ise t_j ‘de her iki grupta risk sayılarını temsil eder. Eğer yaşam her iki grupta benzer ise daha sonra grup 1 için t_j ‘de başarısızlıkların sayıları belirlenir $e_{1j} = n_{1j} \frac{d_j}{n_j}$ ’dir. Test gözlemi d_{1j} ve beklenen e_{1j} arasında farklılığa dayanır.

$$U = \sum_{i=1}^r (d_{1j} - n_{1j} \frac{d_j}{n_j}) \tag{30}$$

varyans tahmin edilebilir.

$$\hat{V}ar(U) = \sum_{i=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)} \quad (31)$$

log-rank testi daha sonra verilir istatistik olarak;

$$S = \frac{U^2}{\hat{V}ar(U)} \quad (32)$$

1 serbestlik dereceli ki-kare (c^2) dağılımı olarak bilinir. Wilcoxon test istatistiği başarısızlık olarak kullanılan n_j ağırlıklı olarak gözlenen değerler arasındaki farkların ağırlıklı bir toplamıdır. Benzer formüllerle gösterirsek;

$$U_w = \sum_{j=1}^r \left\{ n_j \left(d_{1j} - n_{1j} \frac{d_j}{n_j} \right) \right\} \quad (33)$$

$$\hat{V}ar(U_w) = \sum_{i=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{(n_j - 1)} \quad (34)$$

Wilcoxon testleri daha sonra

$$S_w = \frac{U_w^2}{\hat{V}ar(U_w)} \text{ yine 1 serbestlik dereceli } c^2 \text{ dağılımı olarak bilinir.}$$

Ağırlıklı olarak risklerin hesaplanmasında birey sayıları kullanılır. Amaç Wilcoxon testi ve Log-Rank testinde yaşam fonksiyonları arasında ayrılcılığı önceki farklılıklarla göstermektedir.

Log-Rank testi ölüm olasılıkları oransal olduğunda daha güvenilir sonuçlar verir. Ölüm olasılıkları zaman ilerledikçe bütün gruplarda değişim gösterebilir. Bu değişim her bir grup için farklılık gösterebilir. Fakat olasılıkların oranları grup 1 ve grup 2 arasında sabittir. Denklem 30'da verilen U fonksiyonunu grup 1 ve grup 2 için yazıldığında;

$$U(\text{grup 1})+U(\text{grup 2})=0 \quad (35)$$

$$\text{Var}(U(\text{grup 1}))=\text{Var}(U(\text{grup 2})) \quad (36)$$

sonuçları elde edilir. Beklenen değer bağımlı değildir. Genel durumda m grup için, log-rank test istatistiği karesel formda şu şekilde verilir.

$$U \sum^{-1} U^t \quad (37)$$

U bir vektördür $m-1$ uzunluğunda, U vektörünün U^t ise transpozudur. \sum kovaryans matrisidir. $m-1$ boyutunda, özellikle U'nun elemanları;

$$u_k = \sum_{j=1}^r (d_{kj} - n_{kj} \frac{d_j}{n_j}), k = 1,2,3,4, \dots, m-1 \quad (38)$$

ve \sum elemanları ise;

$$s_{kk'} = \sum_{i=1}^r \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} (d_{kk'} - \frac{n_{k'j}}{n_j}), kk' = 1, 2, 3, 4, \dots, m-1 \quad (39)$$

$$d_{kk'} = \begin{cases} 1 \Rightarrow k = k' \\ 0 \Rightarrow k \neq k' \end{cases} \quad (40)$$

Yokluk hipotezi altında $(m-1)$ serbestlik dereceli c^2 dağılımına sahiptir. Genel Wilcoxon'ın testi için m grupta benzer formda $U_w \sum_w^{-1} U_w^t$ gösterilir.

U_w 'nun elemanları;

$$u_k = \sum_{j=1}^r \{n_j (d_{kj} - n_{kj} \frac{d_j}{n_j})\} \quad \text{ve } k = 1, 2, 3, 4, \dots, m-1. \quad (41)$$

değişkenler alır.

Kovaryans matrisi ise;

$$s_{kk'} = \sum_{i=1}^r \frac{n_{kj} n_{k'j} d_j (n_k - d_j)}{(n_j - 1)} (d_{kk'} - \frac{n_{k'j}}{n_j}) \quad (42)$$

$$kk' = 1, 2, 3, 4, \dots, m-1$$

$$d_{kk'} = \begin{cases} 1 \Rightarrow k = k' \\ 0 \Rightarrow k \neq k' \end{cases} \quad (43)$$

2.4. Sağ kalım analizinde Cox Yaklaşımı

Gözlenen N tane yaşam süresi arasından k tane sıralanmış ($t_1 < t_2 < \dots < t_k$), ölüm süresi olarak gösterilsin. Bir R_i setinde, t_i zamanında değerleri saptanan z eş değişken vektörü belirlenmiş olsun. Bir prognostik değişkenin, yaşam süresi üzerine etkide bulunan tüm değişkenler dikkate alınarak genel risk içindeki oranı;

$$\exp(\mathbf{b}'z_i) / \sum_{j \in R_i} \exp(\mathbf{b}'z_j) \quad (44)$$

biçiminde belirlenir.

k farklı ölüm zamanlarının bu oranlarla çarpımı, kısmi olabirlik fonksiyonunu verir. Regresyon katsayıları kısmi olabirlik fonksiyonu yardımı ile tahmin edilirler.

Kısmi olabirlik Fonksiyonu $L(\mathbf{b})$;

$$L(\mathbf{b}) = \prod_{i=1}^k \exp(\mathbf{b}'z_i) / \sum_{j \in R} \exp(\mathbf{b}'z_j) \quad (45)$$

biçimin de belirlenir.

b katsayılarının maksimum olabilirlik tahminleri, logaritmik olabilirlik fonksiyonu;

$$\ln L(\mathbf{b}) = \sum_{i=1}^k \left\{ \sum_{j \in D_i} \ln r(\mathbf{b}_j z_j(t_i)) - m_i \ln \left[\sum_{l \in R_i} r(\mathbf{b}_{z_l}(t_i)) \right] \right\} \quad (46)$$

şeklinde hesaplanır.

Burada;

D_i : t_i zamanında ölen kişilerin seti

m_i : t_i zamanındaki cevapların sayısı

R_i : t_i zamanında canlı olan kişilerin setini belirtmektedir.

$Z_i(t_i)$: t_i zamanında i .birey için p-eş değişken vektörü

$r(\mathbf{b}, z_i(t_i))$: $\exp(\mathbf{b}, z_i(t_i))$ log-linear riskini belirtmektedir.

b katsayılarının hesaplanmasında Newton-Raphson algoritması kullanılır ve ardışık tekrarlanan çözümler ile (iterative parameter estimation) b 'nin tahminleri yapılır.²⁴

Veri setinde benzer süre gözlemleri olduğunda $L(\mathbf{b})$ 'nin maximizasyonu Breslow geliştirdiği formül ile şu şekilde;

$$L(\mathbf{b}) = \left\{ \prod_{i=1}^k \exp(\mathbf{b}' z_i) / \left[\sum_{j \in R_i} \exp(\mathbf{b}' z_j) \right]^{m_i} \right\} \quad (47)$$

hesaplanır.⁵

2.4.1. β Katsayılarının Önemliliğinin Test Edilmesi

Cox regresyon yöntemi yardımıyla tahmin edilen b katsayılarının önemliliği, olabilirlik fonksiyonunun uygulamalarına dayanan test istatistiklerinin geliştirilmesi ile hesaplanmıştır. Test edilecek hipotez ve alternatif hipotez;

$$H_0 : b_1 = b_{10}, b_2 = b_{20}, \dots, b_g = b_{g0}$$

$$H_1 : b_1 \neq b_{10}, b_2 \neq b_{20}, \dots, b_g \neq b_{g0}$$

şeklinde formüle edilir.

Bu amaçla üç test yöntemi ileri sürülmüştür. Bunlar;

1. Wald testi
2. Olabilirlik Oranı (Likelihood,LR,LRATIO) testi
3. Skor (Rao's score Test) testidir.⁵³

2.4.2. Wald Testi

Wald testi maksimum olabilirlik tahminlerinin (MLE) normal dağılım varsayımına dayanır. Regresyon katsayısının, standart hatasına oranı $R = (\hat{b} / SH_{\hat{b}})$ Wald istatistiği olarak adlandırılır. Bu durumda Wald istatistiği Standart Normal Dağılım (SND) gösterir ve SND'nin kritik değerleri ile karşılaştırılarak önemliliği belirlenir. Aynı zamanda Wald istatistiği, $R = z^2 = (\hat{b} / SH_{\hat{b}})^2$ olarak da kullanılabilir. Bu durumda ise Wald test istatistiği 1 serbestlik dereceli ki-kare dağılımı gösterir ve 1 serbestlik dereceli ki-kare (c^2) dağılımının kritik değerleri ile karşılaştırılarak önemliliği belirlenir.

2.4.3. Olabilirlik Oranı Testi(LRATIO)

LRATIO testi Wald testinden daha genel bir testtir. Kategorik değişkenlerin iki veya daha fazla sınıfı bulunduğu ve Cox modeline aynı anda birçok değişkenin alındığı durumlarda tercih edilen bir testtir.

Cox regresyon katsayılarının önemliliği $H_0 : b_i = 0 \quad i=1,2,\dots,p$ varsayımı test edilerek belirlenir. Modelin maksimum olabilirlik istatistiği I_0 değeri ile modelde v tane birbirinden farklı ve yaşam süresi üzerine önemli etkide bulunan başka bir modelin maksimum olabilirlik istatistiği I_v belirlenerek LRATIO test istatistiği hesaplanır.

Modelde, v farklı regresyon katsayısı olduğunda bu modelin maximum olabilirliği I_v ile gösterilir. Her bir modelin $b = 0$ durumuna göre olabilirlik oranı (LR) test edilir.

$$LR = -2 \log(I_0 / I_v) \quad (48)$$

$$LR = -2 \log(L_0 / L_v) \quad (49)$$

biçiminde hesaplanır.

Burada, $L_v = \log I_v$ ve $L_0 = \log I_0$ 'dır.

LR istatistiği v serbestlik dereceli ($sd = v$) c^2 dağılımı gösterir. Önemliliğinin belirlenmesinde ($sd = v$) serbestlik dereceli c^2 dağılımının kritik değerinden yararlanır.

2.4.4. Skor Testi (Rao's Score Test)

Skor testi Q_R ile gösterilir. Logaritmik olabilirlik istatistiklerinden yararlanılarak hesaplanır. İkili değer alabilen (binary) değişkenin P gözlenme oranına göre r kez gözlenme olasılığından yararlanılarak logaritmik olabilirlik istatistiği;

$$L = \log l = r \log p + (n - r) \log(1 - p) \quad (50)$$

şeklinde hesaplanır.

Skor istatistiği, pay'a L 'nin birinci türevinin karesi, payda'ya ise L 'nin ikinci türevi alınarak belirlenir. Q_R istatistiği payın paydaya oranı olarak bulunan bir değerdir. Modeldeki değişkenler sürekli olduğunda ya da birden fazla değişken bulunduğu skor testi Mantel- Haenszel Log-Rank testinin bir genellemesidir.

Skor testi istatistiği, sd tahmin edilen parametre sayısı olmak üzere sd serbestlik dereceli ki-kare (c^2) dağılımı gösterir. Önemliliği sd serbestlik dereceli dağılımının kritik değerleri ile karşılaştırılarak yapılır.¹⁶

2.4.5. Cox Regresyon Modelinde Kullanılan Oranlar

Cox regresyon modelinde verilere uygun olarak kurulan modelin etken ile hastalık arasındaki bağımlılığı, birlikteliği değerlendirmek için yararlanılan ve olasılık kurallarından yararlanılarak geliştirilen oranlar bulunmaktadır. Bu oranlardan en uygun kullanılan Göreli Orantı (Olasılık Oranı, Odds Ratio, OR) ve Göreli Risk Oranı (Relatif Risk, RR)'dir.

T zaman periyodunda toplumda gözlenen H hastalığının gözlenme oranı $P(H)$ ve gözlenmeme oranı $Q(H)$ olarak ele alınır. H 'nin gözlenme oranının gözlenmeme oranına göre kaç kat daha fazla/az olduğunu belirtmesidir.²⁷

Verilere uygun olarak kurulan model önemli olarak bulunduğu hesaplanılan b katsayılarına dayanılarak bir etken olarak yaşam süresi üzerine etkisini belirleme bakımından Odds Oranı bulunabilir.

Odds Oranı; Odds Oranı = $Exp(b)$ hesaplanır.

2.4.6. Modelin Önemliliğinin Test Edilmesi

Cox regresyon modelinde verilere uygun olarak kurulan modelin önemliliğini test için Logaritmik Olabilirlik (LL) istatistiğinin -2 değeri ile çarpılmış bir istatistik olan -2LL değeri hesaplanır. Modelin önemliliğini test etmek için parametrelerin tümünün sıfır olduğu hipotez kullanılır.

İlk olarak oluşturulan model için -2LL istatistiği hesaplanır. Yinelemeli olarak oluşturulan sonraki modeller için -2LL istatistikleri belirlendikten sonra arasındaki fark bulunur. Bulunan bu fark değeri parametre sayısı kadar serbestlik derecesine sahip ki-kare dağılımı gösterir. Bu -2LL 'deki fark istatistiğinin önemliliği belirlenir. Bir önemlilik bulunduğu durumda kurulan modelde önemli olarak yorum yapılır. Model için kullanılan bir başka önemlilik testinde genel ki-kare ya da artık (residual) ki-kare istatistiğidir. Bu test şeklinde skor istatistikleri esas alınır.⁵

2.5. Yarışan Riskler Analizi (Competing Risks Analysis)

T başarısızlık zamanını, D başarısızlığın nedenini, Z 'de T ve D 'nin simdilik ihmal edeceğimiz ortak bir vektörü olsun. Çıkarımımız verilecek bu üç değer için ortak dağılım fonksiyonuna bağlı olacaktır. Yarışan risk modellerinde temel yaklaşım; belirli bir olaydan kaynaklanan riskin verildiği 'nedene özel ölüm olasılığı' fonksiyonudur.

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t, D = k \mid T \geq t)}{\Delta t} \quad (51)$$

Nedene özel ölüm olasılığı tahmin edilebilir ve gözlenen verideki ilgili tüm bilgiyi kullanır. Ayrıca, nedene özel ölüm olasılığından da elde edilen bir değer de tahmin edilebilir.

Daha önceki yaklaşımlarda, her başarısızlık tipi için kişinin potansiyel bir başarısızlık zamanına sahip olduğu varsayımında, yarışan risk modellerinin çok değişkenli başarısızlık zamanı modeli olarak kabul ediliyor. Genel olarak ilk başarısızlık gözleme alınır ve daha sonra gelecek olanlar potansiyel başarısızlık olarak kabul edilir.

\tilde{T}_k , k nedeniyle gerçekleşen başarısızlık zamanı olsun. Biz sadece $T = \min\{\tilde{T}_k\}$ ve D değerlerini gözleyebiliriz. Burada D , hangi olayın ilk önce gerçekleştiğini gösteren indeks değişkenidir. Bazı bireyler çalışma sonunda sansürlü veri ya da kayıp olarak değerlendirilirse, diğer tüm olayların bağımsız olarak kabul edildiği $C \sim G$ olarak tanımlanan bir ekstra sansür dağılımına sahip $D = 0$ değerini alırlar.

Latent (gizli) başarısızlık zamanı yaklaşımında ortak yaşam fonksiyonu olarak düşünülen K farklı olayın ortak dağılımı üzerinde durulur.

$$\bar{S}(t_1, \dots, t_K) = P(\tilde{T}_1 > t_1, \dots, \tilde{T}_K > t_K) \quad (52)$$

$$S^k(t) = P(\tilde{T}_k > t) = \bar{S}(\mathbf{0}, \dots, \mathbf{0}, t, \mathbf{0}, \dots, \mathbf{0}) \quad (53)$$

Marjinal dağılımı üzerinden

$$\lambda(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (54)$$

gibi bir marjinal risk fonksiyonu tanımlanabilir. Bu yaklaşımdaki esas sorun ek varsayımlar olmadan, verilerden ortak yaşam fonksiyonunun tanımlanamamasıdır. Farklı zaman dağılımlarından gelen bağımlı herhangi bir ortak yaşam fonksiyonu için bağımsız zaman dağılımından alınmış aynı nedene özel bir başka ortak yaşam fonksiyonunun ve marjinal dağılımların tanımlanamadığı sonucuna götürür. Bu marjinal bağımsızlık zaman dağılımları üzerinden bağımsızlık testi de yapılamamaktadır. Fakat bazen rakip olaydan hemen önce gözlenen bazı belirleyici değerler ölçülebilir. Bu belirleyiciler rakip olayın bağımlılığı ile ilgili ek bilgi sağlayabilir. Bundan kaynaklanan yeni sorun ise belirleyici değer üzerinde koşullu bağımsızlık testinin yapılamamasıdır.

Kümülatif nedene özel ölüm olasılığı

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds \quad (55)$$

şeklinde tanımlanmaktadır.

$$S_k(t) = \exp(-\Lambda_k(t)) \quad (56)$$

Tahmin edilebilir bir fonksiyondur. Rakip olayın zaman dağılımı ve sansürlü dağılımı bağımsız olmadıkça bir marjinal yaşam fonksiyonu olarak düşünülmemelidir. Bu

durumda, marjinal dağılımın, rakip olaylar oluşmadığı durumlarda olay-zaman dağılımını tanımladığını söyleyebiliriz. Bununla birlikte,

$$S(t) = \exp\left(-\sum_{k=1}^K \Lambda_k(t)\right) \quad (57)$$

Şeklinde tanımlanan yaşam fonksiyonu t anındaki herhangi bir nedenden dolayı başarısız olmama olasılığıdır.

$$I_k(t) = \int_0^t \lambda_k(s) S(s) ds \quad (58)$$

Fonksiyonu k nedeniyle oluşan kümülatif insidans fonksiyonudur ve t zamanından önce gerçekleşen, k nedeniyle gerçekleşen $P(T \leq t, D = k)$ olasılığı ile tanımlanır. Bu fonksiyon ‘kaba birikimli insidans fonksiyonu’ ya da ‘alt-dağılım fonksiyonu’ olarak adlandırılır.

k 'den başka nedenlerle gerçekleşen olaylar sansürlü olarak değerlendirildiğinde,

$$1 - S_k(t) = \int_0^t \lambda_k(s) S_k(s) ds \quad (59)$$

fonksiyonu, t anındaki veya daha önceki bir zamana ait, k nedeninden kaynaklanan başarısızlık olasılığının Kaplan-Meier değeriyle tahmin edilir.

2.5.1. Yarışan Risk Durumunda Kümülatif İnsidans Foksiyonu Yaklaşımı (CIF)

Kümülatif insidans fonksiyonu (CIF) non-parametrik verilerde kullanılan bir tekniktir. Kaplan–Meier yaklaşımı yarışan riskleri sansürlü veri olarak incelerken insidans fonksiyonları her bir zaman aralığı için olasılıkları hesaplamakta ve kümülatif olasılıkları dikkate almaktadır. Sansürlü veri durumu olmadığında CIF ‘in toplam n gözlemde tahmin etme olasılığı;

$$\hat{F}_i(t) = \frac{(T \leq t | C = 1)}{n} \quad (60)$$

şeklindedir. Şayet veri setinde n gözlem p tane olayın bir kaçı daha gerçekleşmemiş ise veri setinde sansür oluşur. Sansür olmadığında $t_1 < t_2 < t_3 < \dots < t_r$ olarak belirtilir. t_j zamanında gerçekleşen her i olayının sayısını tanımlarken d_{ij} ‘den yararlanılır. Eğer sansürlü bir veri ise

$$\hat{F}_i(t) = \sum_{i, t_j < t} \hat{h}_{ij} \hat{S}(t_{i-1}) \quad (61)$$

\hat{h}_{ij} =Nedene özel ölüm olasılığı t_j zamanında i tane olay olarak gösterilir.

Bir olayın gerçekleşme durumu t_{j-1} kadar gözlemlendiğinde olayın gerçekleşmesini değerlendirirsek $t_j - d$ mesafesi t_j ye kadar d_{ij} / n_j olarak tahmin edilir. Böylece CIF tahminleri i olayında sadece birey sayısına bağlı değildir.

$$\hat{F}_i(t) = \sum_{j, t_j < t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}) \quad (62)$$

2.5.2. Kümülatif İnsidans Fonsiyonu (CIF) Non-Parametrik Maximum-Olabilirlik Tahmin Edicileri

Alt dağılım fonksiyonları için \hat{h}_{ij} 'nin önemli bir tahmin edicisi vardır. m_j sansürlü gözlemler $[t_{j-1}, t_j)$ zaman aralığında $t_{j1}, t_{j2} < \dots < t_{jm}$ şeklinde simgelenir. Olabilirlik fonksiyonunda şu şekilde ifade edilir.

$$L = \left[\prod_{j=1}^r \prod_{i=1}^p \{F_i(t_j) - F_i(t_{j-1})\} \right] \left[\prod_{j=1}^{r+1} \prod_{l=1}^{m_j} S(t_{ij}) \right] \quad (63)$$

Nedene özel ölüm olasılığı fonksiyonu ise ;

$$L = \left[\prod_{j=1}^r \prod_{i=1}^p \{h_{ij}^{d_{ij}} (1 - h_j)^{n_j - d_j}\} \right] \quad (64)$$

h_j ise t_j zamanında birkaç olayın ölüm olasılığıdır. Genel olarak, log olabilirlik h_{ij} 'de tanımlanması ile farklılaşır ve türetilir.

$$h_{ij} = \frac{d_{ij}(1 - h_j)}{n_j - d_j} \quad (65)$$

Çünkü ;

$$h_j = \sum_{i=1}^p h_{ij} \quad \text{ve} \quad d_j = \sum_{i=1}^p d_{ij} \quad (66)$$

$$h_{ij} = \frac{d_j(1-h_j)}{n_j - d_j} \quad (67)$$

h_j için tahmin edici $\hat{h}_j = \frac{d_j}{n_j}$ ile ifade edilir. $\hat{h}_{ij} = \frac{d_{ij}}{n_j}$ alt statüsü CIF' in verilir.

$$\hat{F}_i(t) = \sum_{t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}) \quad (68)$$

$\hat{S}(t)$ Kaplan- Meier'de benzer bütün olaylar düşünüldüğünde yaklaşımın tahmin edicisidir. Beklenen değer olarak Maximum olabilirlik tahmini kümülatif insidans fonksiyonunun tahmin edicilerine benzer olarak benzer argümanlar oluşturulabilir.

2.5.3. Kümülatif İnsidans Fonksiyonu (CIF) Tahmin Edicileri için Varyans ve Güven Aralığı

CIF'in güven sınırları gerçek durumda $F_i(t)$ varyansı için tahminlemede $\hat{F}_i(t)$ kullanılır. CIF tahmin edicisinin kesin varyansını edinmek zordur.

$$V(\hat{F}_i(t)) = \sum_{t_j \leq t} Var(\{d_{ij} / n_{ij}\} \hat{S}(t_j)) + 2 \sum_{t_j} \sum_{\substack{t_v > t_j \\ t_v \leq t}} Cov(\{d_{ij} / n_j\} \hat{S}(t_j), \{d_{iv} / n_v\} \hat{S}(t_v)) \quad (69)$$

Delta metodu kullanarak $V(\hat{F}_i(t))$ tahmin edicisi hesaplanır.

$$V_{DM}(\hat{F}_i(t)) = \sum_{t_j \leq t} \{[\hat{F}_i(t) - \hat{F}_i(t_j)]^2 \frac{d_j}{n_j(n_j - d_j)}\} + \sum_{t_j \leq t} \hat{S}(t_{j-1})^2 \frac{d_{ij}(n_j - d_{ij})}{n_j^3} - 2 \sum_{t_j \leq t} \{[\hat{F}_i(t) - \hat{F}_i(t_j)] \hat{S}(t_{j-1}) d_{ij} / n_j^2\} \quad (70)$$

Varyans tahmin edicisi Aalen(1978) tarafından sunulmuştur.⁶

$$V_A(\hat{F}_i(t)) = \sum_{t_j \leq t} \{[\hat{F}_i(t) - \hat{F}_i(t_j)]^2 \frac{d_j}{(n_j - 1)(n_j - d_j)}\} + \sum_{t_j \leq t} \hat{S}(t_{j-1})^2 \frac{d_{ij}(n_j - d_{ij})}{n_j^2(n_j - 1)} - 2 \sum_{t_j \leq t} \{[\hat{F}_i(t) - \hat{F}_i(t_j)] \hat{S}(t_{j-1}) \frac{d_{ij}(n_j - d_{ij})}{n_j(n_j - d_j)(n_j - 1)}\} \quad (71)$$

CIF için güven aralığı tahmini ise;

$$\hat{F}(t) \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{F}(t))} \quad (72)$$

$$A = \frac{z_{1-\alpha/2} \sqrt{\hat{V}(\hat{F}(t))}}{\hat{F}(t) \log(\hat{F}(t))} \quad (73)$$

CIF'in varyans tahmininde burada ki z standart normal dağılıma sahiptir.

2.5.4. “Proportion Hazard”(Oransal Ölüm Olasılığı) Modeli

Proportion (oransal) Hazard modeli kurmak için teorik alt yapı göz önüne alınması gerekmektedir. Ölüm olasılığı i birey için iki parçada eş değişkenlerin $x_{1i}, x_{2i}, \dots, x_{mi}$ oluşumunu verir. Zaman faktörünü içeren fakat eş değişken olmayan bir diğerinde eş değişken olup zaman verilmediği durumlar için incelenir.

$$h_i(t) = \exp\{b_1 X_{1i} + b_2 X_{2i} + \dots + b_m X_{mi}\} h_0(t) \quad (74)$$

$h_0(t)$ ana ölüm olasılığı fonksiyonudur ve b_1, b_2, \dots, b_m parametreleridir veya beklenen bir katsayıdır. Maximum olan kısmi olabilirlik ile elde edilir.

$$L(b_1, b_2, \dots, b_m) = \prod_{j=1}^r \left(\frac{\exp\{b_1 X_{1j} + \dots + b_m X_{mj}\}}{\sum_{i \in R_j} \exp\{b_1 X_{1i} + \dots + b_m X_{mi}\}} \right) \quad (75)$$

Bütün başarısızlık (t_i) zamanlarında ve R_i , t_i 'de risklerin var oluşunu temsil eder. Kısmi olabilirliğin maximizasyonu bilgisayar üzerinde tekrarlı bir süreç boyunca verilerek elde edilir. Güven aralığı her bir katsayı için ise;

$$\hat{b}_k \pm z_{1-a/2} \hat{S}_b \quad \text{Ayrıca } z_a \text{ ve } a \text{ standart normal dağılıma sahiptir.}$$

Wald testi $H_0 = b_k = 0$ karşı alternatif hipotez $H_0 = b_k \neq 0$ test istatistiği \hat{b}_k / \hat{S}_{bk} standart normal dağılım gösterdiğinde.

2.5.5. Eş değişken Testi

Yarışan risklerin varlığında Eş değişkenlerin ölüm üzerinde etkilerini araştıran test yöntemidir. Yarışan risk varlığında bu soruya cevap aranır. Nedene özgü ölüm

olasılıkları marjinal dağılımların ölüm olasılıklarıdır. Gray tarafından alt dağılımların ölüm olasılıkları şu şekilde tanımlanmıştır.

$$g(t) = \frac{f(t)}{1-F(t)} \quad (76)$$

f değeri ilgilenilen olay için alt insidans fonksiyonunu ifade eder. F değeri ise ilgilenilen olay için alt insidans fonksiyonunun ölüm olasılıklarıdır. Alt insidansların ölüm olasılığı ilgilenilen olayın gözlemleri olarak yorumlanabilir. İstenilen olay kliniğe veya tıp sorularına bağlıdır. Nedene özgü ölüm olasılıkları CIF istenilirse karşılaştırılabilir.

Sağ kalım fonksiyonunda yarışan risklerin yokluğunda ölüm olasılıklarının monotonik bir durumu vardır. Gray bu durumun varlığına dikkat çekmiştir. Bir gruptaki nedene özgü ölüm olasılıklarının diğer gruptan daha büyük olduğu gösterilmiştir. Bu zaman sürecinde CIF 'in eşitsizliği belli noktalarda değişebilir. İki grup için CIF iki eğri olarak gösterilmiştir. İlgilenilen olayın ölüm olasılıkları üstel dağılıma sahiptir.

2.6. Yarışan Risklerin Tanımlanmasında Kullanılan Regresyon Modelleri

2.6.1. Ölüm Olasılığına Dayalı Regresyon Modelleri

2.6.1.1. Nedene Özel Ölüm Olasılığı Regresyonu (Cause-Specific Hazard Regression)

Nedene özel ölüm olasılığı regresyonu sağdan sansürlü verilerin değerlendirilmesinde yaygın olarak kullanılmaktadır. Prentice ve arkadaşları nedene özel ölüm olasılık oranlarının eş değişken etkisini tahmin etmeyi amaçlamışlardır.²⁷

Yarı parametrik Cox regresyon yaklaşımı sebebe bağlı ölüm olasılık oranını $I_{k,0}(t)$ olarak tanımlanmıştır.²⁵

$$I_k(t|X) = I_{k,0}(t) \exp(\mathbf{b}_k^T X) \quad (77)$$

Yaygın olarak kullanılan Cox Regresyon Modellerinde olduğu gibi nedene özel ölüm olasılığı oranları zamanının sabit olduğu durumlarda eş değişken etkilerin oransal dönüşümünü gerçekleştirdiği varsayılır. Bu varsayım Schoenfeld artıkları kullanılarak grafiksel yöntemler ile kontrolü yapılır.²⁵ Geçerli bir Cox regresyon modeli oluşturmak için zamana bağlı etkilerden ve zamana bağlı eş değişkenlerden oluşan bir model dönüşümü esas alınır.^{20,26,27}

Cox tarafından sunulan klasik modelin genişletilmesi ve modelin doğruluğunu kanıtlamak için Cox regresyon modelin içerdiği varsayımların ayrıntılı bir tanımlama için örnekler incelenebilir.²⁸

Yarı parametrik Cox regresyon modeli kullanılması yerine üstel, weibull ya da gamma dağılımına çevirme gibi yöntemler uygulanarak parametrik modeller ile analiz yapılabilir.¹⁷ Sansürlü gözlemler olarak diğer modellerden ölümlerin ve ölüm istenilen olay olduğunda uygulanan Cox regresyon modeli, SAS, SPSS, STATA gibi standart istatistik paket programları veya açık kodlu R programı ile incelenebilir.

Nedene özel ölüm olasılığı oranlarında eş değişkenlerin etkisi kümülatif insidans fonksiyonunda direk bir etki olarak düşünülemez. Yani B grubu ile karşılaştırıldığında A grubunun nedene özel ölüm olasılıkları daha yüksek ise A grubu B grubundan ilgilenilen olay açısından yüksek bir insidansa sahip demek değildir.³⁰ K sebepli bir olayın k . kümülatif insidans fonksiyonu aşağıdaki gibi gösterilir.

$$F_k(t|X) = \int_0^t I_k(s|X) \exp\left(-\sum_{l=1}^K \Lambda_l(s|X)\right) ds \quad (78)$$

olarak gösterilir.

$I_k(t|X)$ tane eş değişken matrisi için t zamanında k olayın kümülatif nedene özel ölüm olasılığı oranı verilir.

$$\Lambda_k(t | X) = \int_0^t I_l(s | X) ds \quad (79)$$

olarak gösterilir.

2.6.1.2. Fine ve Gray Modeli (Alt Dağılım Ölüm olasılığı Regresyon modeli) (Subdistribution Hazard Regression)

Fine ve Gray Alt dağılım ölüm olasılığı regresyon modeli olarak gösterilen sağ kalım modelini nedene özel ölüm olasılıkları regresyonunu açıklarken ortaya çıkan olası problemleri gidermek için çalışmalarda bulunmuş ve geliştirmiştir.¹² Modelin geliştirilmesi ile Kümülatif insidans fonksiyonu ile regresyon katsayıları ile doğrudan bağlantısını bu modelle ortaya koyulmuştur. Fine ve Gray modeli aslında Gray tarafından 1988 yılında önce alt dağılım ölüm olasılıkları diye tanımlanmış ve ilgilenilen olayın eş değişken değerleri olan ilişkisini değerlendirmiştir.¹⁴

Bunu da şu şekilde göstermiştir.

$$I_k^*(t) = \lim_{\Delta t \rightarrow \infty} \frac{P(t \leq T < T + \Delta t, C = k | T \geq t \cup \{T < t, C \neq k\})}{\Delta t} \quad (80)$$

t zamanından önce ölüm olayı gerçekleşmiş fakat ilgilenilen olaydan dolayı bir ölüm gerçekleşmemiş ise olay zamana göre daha detaylı incelenerek gerçekte olayın sansürlü olmadığı anlaşılacaktır.

Alt dağılım ölüm olasılıklarının tahmin edicileri;

$$\hat{I}_k^*(t_i) = \frac{d_{ki}}{n_i^*} \quad (81)$$

$d_{ki} = t_i$ zamanında k sebepten ölenlerin sayısı olarak gösterilir.

$n_i^* = t_i$ zamanına kadar gerçekleşmemiş bütün durumların sayısıdır.

Alt dağılım ölüm olasılıkları ile nedene özel ölüm olasılıkları arasındaki farkı açıklamıştır. İlk yarışan olay gerçekleşene kadar alt dağılım ölüm olasılığı tahmini ile nedene özel ölüm olasılığı tahmini birbirine eşittir

İlk yarışan olay gerçekleştiğinde nedene özel ölüm olasılıklarının \hat{n}_i tahmin edicisinden daha küçük bir değere sahip olduğundan alt dağılım tahmin edicisinden daha küçük bir değer almaktadır. Alt dağılım ölüm olasılığını Cox regresyon modeli ile kurduğumuzda;

$$\hat{I}_k^*(t | X) = I_{k,0}^*(t) \exp(\mathbf{b}_k^{*T} X) \quad (82)$$

Burada $I_{k,0}^*(t)$ alt dağılım ölüm olasılığı fonksiyonu olarak tanımlanırken bütün eş değişken değerlerinin 0 olduğu kabul ettiğimiz durumlardır fakat alt dağılım ölüm olasılığı eş değişken faktörlerinin etkisi göz ardı edilemez. Bilinen klasik sağ kalım analizdeki fonksiyonla tanımlamada bulunursak;

$$F_k(t | X) = 1 - \exp(-\Lambda_k^*(t | X)) = 1 - \exp\left(-\int_0^t I_k^*(s | X) ds\right) \quad (83)$$

Böylece Fine ve Gray tarafından regresyon eş değişken faktörlerinden kümülatif insidans fonksiyonu tahmin edilmiş olacaktır. Yarışan olaylardaki eş değişken etkileri böylece göz ardı edilmemiş olur.

2.7. Sağ Kalım Verisi Üretme

2.7.1. Sağ kalım Analizlerinde Veri üretiminde kullanılan Fonksiyonlar

Cox oransal ölüm olasılığı modellerinin Sağ kalım fonksiyonu aşağıdaki gibi gösterilmektedir.

$$S(t | x) = \exp[-H_0(t) \exp(\mathbf{b}'x)] \quad (84)$$

olarak gösterilir.

Kümülatif ölüm olasılığı fonksiyonu;

$$H_0(t) = \int_0^t h_0(u) du \quad (85)$$

olarak gösterilir.

Buna bağlı olarak Cox modelinin dağılım fonksiyonları

$$F(t | x) = 1 - \exp[-H_0(t) \exp(\mathbf{b}'x)] \quad (86)$$

ile gösterilir.

F fonksiyonunda Y rastgele bir değişken olsun. Uniform dağılım $[0,1]$ aralığında tanımlanıyor olursa ;

$$U \sim Uni[0,1] \text{ ve } U = F(Y) \quad (87)$$

Eğer $U \sim Uni[0,1]$ tanımlanırsa $(1-U) \sim Uni[0,1]$ ile ifade edilir. Böylece Cox modelinin sağ kalım zamanı T olarak gösterilir.

$$U = \exp[-H_0(T) \exp(\mathbf{b}'x)] \sim Uni[0,1] \quad (88)$$

olarak gösterilir.

$h_0(t) > 0$ bütün t değerleri için T zamanı;

$$T = H_0^{-1}[-\log(U) \exp(-\mathbf{b}'x)] \text{ ile ifade edilir.} \quad (89)$$

2.7.2. Üstel Dağılımla Veri Üretme

$I > 0$ parametrelili üstel bir dağılım $t \geq 0$ için sabit bir ölüm olasılığı fonksiyonuna sahiptir.

$$H_0^{-1}(t) = I^{-1}t \quad (90)$$

eşitliği ile gösterilir.

Bir Cox modelinin sağ kalım zamanı aşağıdaki formül ile sunulmuştur.

$$T = I^{-1}[-\log(U) \exp(-b'x)] = -\frac{\log(U)}{I \exp(b'x)} \quad (91)$$

Cox regresyon modelinin uyumlu ölüm olasılığı fonksiyonu;

$$h(t | x) = I \exp(b'x) \quad (92)$$

şeklinde gösterilir.

Bu yüzden sabit ölüm olasılığı sağ kalım zamanlarını üstel dağılımda göstermek için;

$$I(x) = I \exp(b'x) \quad (93)$$

ölüm olasılığı parametresi kullanılır.

2.7.3. Weibull Dağılımından Veri Üretme

Weibull dağılımları ölüm olasılığı fonksiyonlarının daha genel bir formu olarak verilmektedir. İki pozitif parametre olarak ifade edilir. I bir Weibull Dağılımının ölçüm parametresi iken ν şekil parametresi olarak ifade edilmektedir. $\nu=1$ ölüm olasılığı fonksiyonu durumunda üstel dağılımı temsil ettiği gözlemlenmektedir. $\nu > 1$ durumu için ölüm olasılıkları artan bir eğilimde iken $0 < \nu < 1$ durumunda monoton azalan bir eğilim içerisinde seyretmektedir.²⁹

Kümülatif ölüm olasılığı fonksiyonlarının aralığı;

$$H_0(t) = (I^{-1}t)^{1/\nu} \quad (94)$$

şeklinde gösterilmektedir.

Cox modelinin sağ kalım zamanları Weibull dağılımında ölüm olasılığı ile gösterimi şu şekildedir.

$$T = I^{-1}[-\log(U) \exp(-b'x)]^{1/v} = -\left(\frac{\log(U)}{I \exp(b'x)}\right)^{\frac{1}{v}} \quad (95)$$

olarak gösterilmektedir.

Karşılıklı ölüm olasılığı fonksiyonları şu şekilde gösterilmektedir.

$$h(t | x) = I vt^{v-1} \exp(b'x) = I \exp(b'x) vt^{v-1} \quad (96)$$

Çizelge 1.2.7.1. Dağılım Tablosu

Karakter	Üstel Dağılım	Weibull Dağılım
Parametre	Ölçüm Parametre $l > 0$	Ölçüm Parametre $l > 0$ Şekil Parametresi $v > 0$
Aralık	$[0, \infty)$	$[0, \infty)$
Ölüm Olasılığı Fonksiyonu	$h_0(t) = l$	$h_0(t) = I vt^{v-1}$
Kümülatif Ölüm Olasılığı Fonksiyonu	$H_0(t) = lt$	$H_0(t) = I t^v$
Ters Ölüm Olasılığı Fonksiyonu	$H_0^{-1}(t) = l^{-1}t$	$H_0^{-1}(t) = (l^{-1}t)^{1/v}$
İnsidans Fonksiyonu	$f_0(t) = l \exp(-lt)$	$f_0(t) = I vt^{v-1} \exp(-I t)^{1/v}$
Sağ Kalım Fonksiyonu	$S_0(t) = \exp(-lt)$	$S_0(t) = \exp(-I t^v)$
Ortalama	$E(T) = \frac{1}{l}$	$E(T) = \frac{1}{\sqrt[v]{I}} \Gamma^2\left(\frac{1}{v} + 1\right)$
Varyans	$\text{Var}(T) = \frac{1}{l^2}$	$\frac{1}{\sqrt[v]{I^2}} [\Gamma\left(\frac{2}{v} + 1\right) - \Gamma^2\left(\frac{1}{v} + 1\right)]$

3.GEREÇ VE YÖNTEM

Genel sağ kalım analizlerinin eksikliklerini ortaya koymak için yapılan bu tez çalışması, yarışan riskler analizi ile sağ kalım analizlerinin sağ kalım süreleri karşılaştırmalarını hem gerçek verilerde hem de üretilen verilerde değerlendirmiştir. Çalışmadaki gerçek veri seti Pintilie tarafından 2006 yılında toplanan folliküler hücre lenfoması tanısı ile gelen hastaların sonuçlarından oluşmaktadır. Çalışma 1967 ile 1996 yılları arasında Toronto'daki Princess Margaret hastanesine gelen 541 hastanın verilerini içermektedir.³⁰

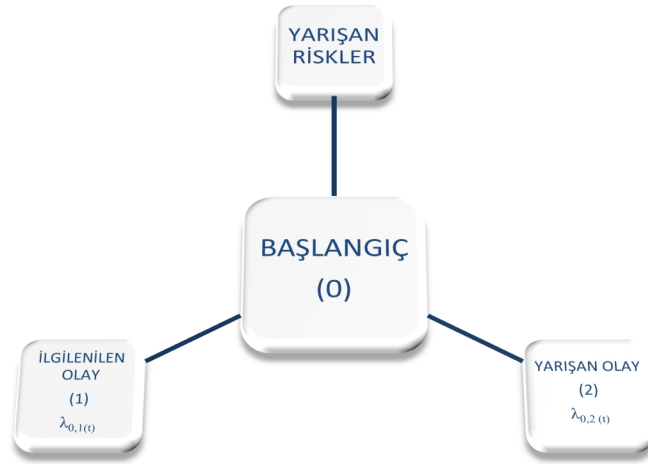
Çalışmaya Folliküler hücre lenfomasından erken teşhisle gelen sadece (RT) radyasyon tedavisi görenler ile kemoterapi veya radyasyon alan, hastalığın I. ve II. döneminde olan hastalar çalışmaya alınmıştır.

Bu çalışmanın amacı hasta takip süresince ölüm gerçekleştiğinde ölüm sebebinin hastalıkla ilişkisini incelemek ve tedaviye cevap vermeyen ve nüks durumlarını ortaya koyarak sağ kalım çalışmasını yarışan risklere göre değerlendirmektir. Hastalığa bağlı olmayan ölümleri yarışan olay olarak değerlendirildikten sonra çalışma ona göre incelenmiştir. Çalışmaya http://www.uhnresearch.ca/labs/hill/People_Pintilie.htm adresinden ulaşılabilmektedir.

Gerçek veri seti üzerindeki uygulamada hastalığa bağlı durumda ve hastalık dışı durumda Kaplan Meier tahmin yönteminin sağ kalım olasılıkları ile Yarışan risk yöntemi uygulandığında yöntemin sağ kalım olasılıkları arasındaki farkların ortalama değerleri test edilmiştir. Simülasyon çalışmasında farklı senaryolar oluşturularak yarışan risklerin sağ kalım olasılık sonuçları ile Kaplan Meier yönteminin sağ kalım olasılık sonuçları üretilen farklı senaryolarda ortaya koyulmuştur. Weibull dağılımından ve üstel dağılımdan farklı örneklem sayıları ile sağ kalım olasılık sonuçları karşılaştırılmıştır. n=100,150,250,500,1000 alındığında dağılım sansürleme oranları da %10,20,30,40,50 alındığında Kaplan Meier yönteminde ve yarışan riskler analizinde sağ kalım olasılıklarının farklarının ortalamasının farklı olduğu tespit edilmiştir. Yarışan riskleri

gerçek veri seti üzerinde incelerken eş değişken durumunda nasıl bir sonuç gösterdiği de Gray's test yöntemi ile yorumlanmıştır.

3.1. Yarışan Risk Verilerinin Üretilmesi



Şekil 3.3.1.1. Yarışan Riskler

$i = 1, 2$ nedene özel ölüm olasılığını gösterirken $X_t \in \{0, 1, 2\}$ zamana göre durumları tanımlar. $X_t = 0$ kişilerin yaşıyor olduğunu ve $X_t = 1, 2$ yarışan olayın olmadığını düşündüğümüzde kişileri yaşıyor kabul ederek bu kişileri sansürlü bir gözlem olarak gözlemlendiğinin göstergesidir. $X_t = 1$ ise ilgilenilen olayın gerçekleştiğini gösterirken $X_t = 2$ diğer yarışan olayın gerçekleştiğini tanımlamış olacaktır. Eğer $(X_t)_{t \geq 0}$ yarışan risklerin durumunun gerçekleştiğini gözlemleniyorsa C tane sansürlü gözlem vardır yorumu yapılır. Yarışan riskler $X_t = 0$ başlangıç zamanının T sağ kalım süresinin dışında gerçekleştiği durum olarak ifade edilir.

$T = \inf\{t > 0 \mid x_t \neq 0\}$ ve gözlemlenen veri $(\min(T, C), 1(T \leq C)X_t)$ bir fonksiyon olarak ifade edilir.

$$\begin{aligned}
 P(X_T = i, X_T = i) &= P(X_T = i) \neq P(X_T = i).P(X_T = i) \\
 P(X_T = i, X_T = i) &= P(X_T = i) \\
 P(X_T = i) &\neq P(X_T = i).P(X_T = i)
 \end{aligned} \tag{97}$$

ile gösterilmesi uygundur.

Nedene özel ölüm olasılıklarını açıklarken $a_{0i}(t)$

$$a_{01}(t) = dt := P(T \in dt, X_T = i \mid T \geq t), i = 1, 2 \tag{98}$$

Bölümleyecek kadar küçük aralığı alırsak $[t, t + dt]$ ³¹ Hem de $a_{0i}(t)$, $i = 1, 2$ ifadesini kullanacağız. Sağ kalım fonksiyonunu tanımlanırsa şayet;

Nedene özel ölüm olasılıklarını kümülatif insidans fonksiyonu ile tanımlarsak;

$$P(T \leq t, X_T = i) = \int_0^t P(T > u -) a_{0i}(u) du, i = 1, 2 \tag{99}$$

gösterilir. $a_{01}(t)$ hem de $a_{02}(t)$ bağlı bir durum gerçekleşir. Simülasyon çalışmalarında gerekli olan sağ kalım zamanlarındaki olayların durumudur. Bunun için gözlemler çok önemlidir. Sebebi 1 olan durumdan ölen bir birey $T = t$ zamanı ile gösterilir.

$$P(X_T = 1 | T \in dt, T \geq t) = \frac{P(T \in dt, X_T = 1 | T \geq t)}{P(T \in dt | T \geq t)} \quad (100)$$

$$= \frac{a_{01}(t)}{a_{01}(t) + a_{02}(t)}$$

Formül ile ifade edilir.

Yarışan risk verilerini sürekli bir zaman verisi için formulüze ettiğimizde;

$$a_{0i}(t) = P(X_t = i | T \geq t), i = 1, 2 \quad (101)$$

olarak gösterilir.^{1, 4, 13}

3.6.5.Simülasyon Düzeni

Son zamanlarda yapılan simülasyon çalışmaları git gide artmaktadır. Burton ve arkadaşları sağ kalım analizlerine simülasyon çalışmalarında insidans ile durmuş ve bu alanda çalışma örnekleri sunmuşlardır.^{32,33} Çalışmamızda simülasyon çalışmaları üzerinde şu şekilde bir seyir izlenecektir.

Weibull dağılımından üretilen veri seti için;

1. Hastalığa bağlı ve hastalık dışı durumları yansıtacak iki sağ kalım süresi üretilir.
2. Hastalığa bağlı ve hastalık dışı durumlar için Kümülatif ölüm olasılıkları fonksiyonlarına özel bir formül oluşturulur.
3. Kümülatif ölüm olasılıkları hastalığa bağlı durum = $0.09 * \log(t + 1) + 0.024 * t * t / 2 - y$ ile ifade edilir.
4. Kümülatif ölüm olasılıkları hastalık dışı bağlı durum = $0.825 * 0.09 * \log(t + 1) + 0.2 * 0.024 * t * t / 2 - y$ ile ifade edilir.
5. Eş değişken faktörü olmadığında $Z=0$ nedene özel ölüm olasılıkları $a_{01}(t)$ ve $a_{02}(t)$ zamana bağlı bir fonksiyon olarak üretilir.

6. $a_{01}(t)=0.09 / (t + 1)$, $a_{02}(t)=0.024 * t$ 'dir.
7. Eş deęişken faktörü olmadığında $Z=1$ nedene özel ölüm olasılıkları $a_{01}(t)$ ve $a_{02}(t)$ zamana baęlı bir fonksiyon olarak üretilir.
8. $0.825 * a_{01}(t)$, $0.2 * a_{02}(t)$ 'dir.
9. Bütün nedenler için ölüm olasılıkları ile saę kalım T süresi üretilir.
10. Saę kalım zamanını T bir binomial durum oluşturularak üretilir ve $I_{01}(T)/(I_{01}(T)+I_{02}(T))$ birinci sebep üzerine durumlar incelenir.
11. C tane sansürlü zaman üretilir.
Üstel dağılımından üretilen veri seti için;
 1. Hastalığa baęlı ve hastalık dışı durumda üstel dağılım parametreleri λ_1, λ_2 ile sansür oranları belirlenerek hastalığa baęlı ve hastalık dışı durumlarda saę kalım süreleri üretilir.
 2. Bütün nedenler için ölüm olasılıkları ile saę kalım T süresi üretilir. Üstel dağılımda simülasyonda saę kalım T süreleri $\min(T_1, T_2)$ belirlenir.

3.6.6. Alt Dağılım Ölüm Olasılığı Modelinde Simülasyon

Yarışan risk verilerini alt dağılım ölüm olasılığı modeli ile üretmek istendiğinde

$$I_{(t;Z)} = I_0(t) \exp(b^T Z)$$

formülü bize ölüm olasılıklarının eş deęişken faktörüne dayanılarak nasıl üretileceğini gösterir.

b : Regresyon katsayısının k çeşit vektörüdür.

Z : k çeşit eş deęişken değeridir.

$I_0(t)$: Ölüm olasılıklarının alt dağılım değerleridir.

Çalışmanın adımları şu şekildedir;

1. İlk adımda nedene özel ölüm olasılığı yoksa alt dağılım ölüm olasılığından hangisini seçebileceğine karar verilir.
2. Her iki yöntem içinde seçilen modele uygun yöntem karar verilir.
3. $Z=0$ durumu için yani eş deęişken faktörü olmadığında $a_{01:0}$ ve $a_{02:0}$ nedene özel ölüm olasılıklarını ve alt dağılım ölüm olasılıklarını belirleyip $A_{01:0}$ ve $A_{02:0}$

kümülatif insidans fonksiyonları tespit edilmelidir. Daha sonra $\Lambda_0(t)$ alt dağılım ölüm modeline göre (Fine ve Gray) kümülatif değer belirlenir.

$$4. I_{01}(t) = \left(1 + \frac{P(T \leq t, X_T = 2)}{P(T > t)}\right) I(t) \text{ ile } I_{01} \text{ ve } I_{02} \text{ hesaplanır.}$$

$$5. I_0(t) \exp(-\Lambda_0(t)) = \exp(-A_{01;0}(t) - A_{02;0}(t)) I_{01;0}(t) \text{ ile hesaplanır.}$$

$$6. I_{02;0(t)} = I(t) - I_{01;0}(t) - \frac{d}{dt} \ln(I_0(t) / I_{01;0}(t)) \text{ ile hesaplanır.}$$

$$7. I_{02;0}(t) \text{ ve } I_0(t) \text{ seçilerek } I_{01;0}(t) \text{ hesaplanır.}$$

$$\exp(-A_{01;0}(t)) I_{01;0}(t) = I_0(t) \exp(-\Lambda_0(t) + A_{02;0}(t))$$

olarak hesaplanır.

8. Nedene özel ölüm olasılıklarının modelleri alt dağılım ölüm olasılığına göre verilir. $I_{01}(t; z)$ veya $I_{02}(t; z)$ bir model olarak seçilir.

Eğer $I_{02}(t; z)$ modeli seçilirse;

$$I_{02}(t; z) = I(t; z) - I_{01}(t; z) - \frac{d}{dt} \ln(I(t; z) / I_{01}(t; z))$$

$$I_{01}(t; z) = \frac{I(t; z) \exp(-\Lambda(t; z) + A_{02}(t; z))}{1 - \int_0^t I(u; z) \exp(-\Lambda(u; z) + A_{02}(u; z)) du}$$

olarak hesaplanır.

3.6.8. Analiz ve Simülasyonda Kullanılan Paket Programlar

Simülasyon ve analizlerin tamamı için R paket programı kullanılmıştır. Çalışmanın analizlerini yapabilmek için R paket programında hazırlanan komutlar ve fonksiyonlar ekte verilmiştir.

4.BULGULAR

Çalışmanın bu bölümünde yarışan riskleri değerlendirirken gerçek veri seti üzerinde yapılan bir uygulama ile simüle veriler üretilerek oluşturulan veri setlerinde Kaplan Meier (KM) ve Yarışan Risk (CR) yöntemleri ile hesaplanan sağ kalım olasılıkları arasındaki farkın sonuçları yer almaktadır. İlk bölümde gerçek veri setinde eş değişken faktörleri göz önüne alınmadığında Kaplan Meier ve Yarışan Risk yöntemleri veri setine özel hastalığa bağlı ve hastalık dışı durumlarda sağ kalım olasılıkları arasındaki farklarının ortalamalarının karşılaştırılma sonuçları nedene özel ölüm olasılıkları sonuçları ile değerlendirilmiştir. İkinci bölüm de ise simüle veri setleri farklı senaryolar oluşturarak değerlendirilmiştir. Bu senaryolar farklı dağılım (Weibull dağılımı ve Üstel dağılım) ve farklı örneklem büyüklüğü (n=100, 150, 250, 500, 1000) , olduğunda ayrıca tekrar sayısı 1000 alındığında Kaplan Meier ve Yarışan Risk yöntemlerinin sağ kalım olasılıkları arasındaki farkın ortalama sonuçları sunulmuştur. Kümülatif ölüm olasılığı grafikleri ile bu sonuçların nasıl yorumlandığı bulgular kısmında verilmiş. Çalışmanın üçüncü aşamasında yine aynı veri seti incelenerek eş değişken faktörleri göz önüne alındığında yarışan risk varlığında sonuçlarının nasıl yorumlandığı verilmiştir.

4.1.Foliküler Hücre Lenfoması Veri Seti

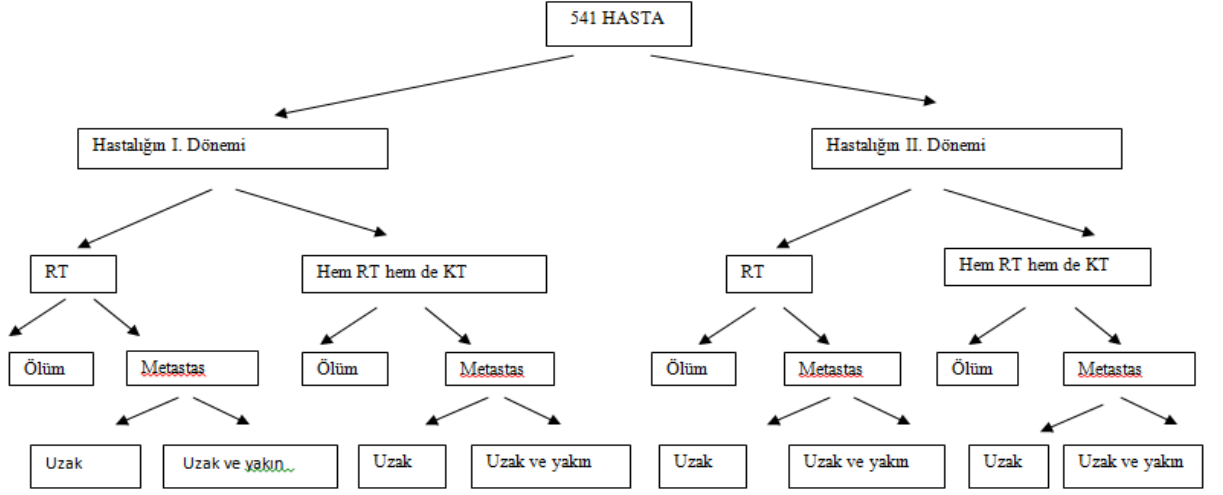
4.1.1. Gerçek Veri setinde Uygulama

Çalışmada kullanılacak veri seti gerçek bir veri setidir. 1967 yılında Toronto'da Princess Margaret hastanesine yatan folliküler hücre lenfoması tanısı konmuş 541 hastanın verilerinden oluşmaktadır. Hastaneye gelen 541 hasta, hastalığın I. ve II. döneminde olarak ayrılmış ve hastalara uygulanan tedavide sadece radyasyon tedavisi gören hastalar ve radyoterapi + kemoterapi tedavisi almış olanlar olarak ayrılmıştır. Çalışmadaki temel amaç çalışılan hasta grubunda uzun vadedeki sonuçlarını gözlemlemektir. İlgilenilen sonuçlar (yarışan riskler) tedaviden sonra metastas durumu

veya ölüm durumudur. Metastas uzak metastas olabilirken diğer durumda da hem uzak hem de yakın metastas durumu gerçekleşmiş olabilir. Tedavinin sonuçlarını incelerken hastanın tedaviye cevap vermesi durumunda tam bir iyileşme gerçekleşmiş olabilirken tedaviye cevap vermeme durumunda ise ölümlerle sonlanabilmektedir. İzlem süresi sürecinde tam iyileşme durumundaki hastanın hastalığı tekrar nüksedebilir ve bu durumda yakın metastas gerçekleşebilir veya hem yakın hem uzak metastas da gerçekleşebilir. Çizelge 4.4.1.1.1’de eş değişken faktörlerini incelemeye göre follüküler hücre lenfomasının veri seti verilmiştir. Çizelge 4.4.1.1.2’de eş değişken faktörleri ile follüküler hücre lenfomasının veri seti verilmiştir.

Veri setindeki değişkenlerin tanımlamaları;

Yaş	yıl
Hgb	Hemoglobin(g/l)
Klinik D	Klinik durum 1=Hastalığın I. dönemi 2=Hastalığın II. dönemi
KT	Kemoterapi Y=Evet N=Hayır
RT	Radyoterapi Y=Evet Boş=Hayır
resp	Tedaviden sonra iyileşme durumu CR=Tam iyileşme, NR=İyileşmeme
relsite	Metastasin yeri; L=Yakında metastas, D=Uzakta metastas, B=Hem yakında hem uzakta metastas, Boş=Metastas yok
survtime	Ölüm veya son takip süresine kadar geçen süre
stat	1=ölu, 0=sağ
dftime	İlk başarısızlık (İyileşme, nüks veya ölüm) zamanına kadar geçen süre
dfcens	Sansürlü veri 1=Başarısızlık (İyileşme, nüks veya ölüm)



Şekil 4.4.1.1.1 Gerçek Veri Seti Diagramı

Çizelge 4.4.1.1.1.Foliküler Hücre Lenfoması Hastalarının Veri Seti

Yaş	hgb	Kdurum	K T	R T	Zaman	stat
56	140	2	N	Y	0,238193018	1
36	130	2	N	Y	12,41889117	1
39	140	2	Y	Y	0,002737851	1
37	140	1	N	Y	15,68514716	1
61	110	2	N	Y	0,002737851	1
69	120	1	N	Y	8,41889117	1
57	110	2	N	Y	25,14989733	1
32	120	2	N	Y	31,10198494	0
24	110	2	N	Y	14,57357974	0
49	110	2	N	Y	0,807665982	1
44	130	2	N	Y	5,615331964	1
82	120	1	N	Y	0,002737851	1
58	130	1	N	Y	13,04859685	1
32	140	2	N	Y	0,150581793	1
51	140	2	N	Y	29,66735113	1
73	130	2	N	Y	2,193018481	1
64	130	1	N	Y	0,405201916	1
56	129	1	N	Y	11,18412047	0
38	160	2	N	Y	11,27446954	1
68	160	2	N	Y	4,005475702	1
55	160	1	N	Y	6,973305955	1
58,1	116	2	N	Y	1,437371663	1
77	130	2	N	Y	0,331279945	0
54	140	2	N	Y	29,25667351	1
75	140	2	N	Y	0,002737851	1
54	160	1	N	Y	4,147843942	1
71	130	1	N	Y	0,002737851	1
77	130	1	N	Y	5,629021218	1
70	140	1	N	Y	0,711841205	1
49	140	2	N	Y	13,33333333	0
48	130	1	N	Y	28,75017112	0
31	110	2	N	Y	28,61054073	0
56	130	1	N	Y	16,07665982	1

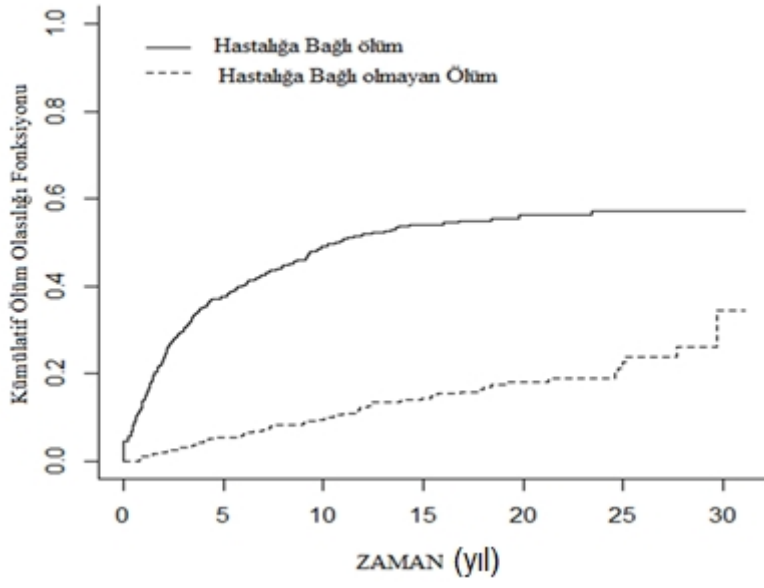
Çizelge 4.4.1.1.2. Folliküler Hücre Lenfoması Hastalarının Veri Seti

yaş	path1	hgb	ldh	KlinikD	blktxeat	relsite	KT	RT	survtime	stat	dftime	dfcens	resp	stnum
34	NH03	124	10	1	2	D		Y	23,5866	1	2,60917	1	CR	94
27	NH04	129	75	1	0			Y	9,02943	0	9,02943	0	CR	343
69	NH03	133	90	1	0	D		Y	11,5236	0	4,21629	1	CR	369
60	NH04	134	96	2	3		Y	Y	1,59617	1	0,00274	1	NR	378
56	NH02	148	107	1	1	D		Y	11,2197	0	0,4627	1	CR	345
57	NH03	138	107	2	3		Y	Y	6,48323	0	6,48323	0	CR	434
54	NH04	144	109	1	2		Y	Y	4,88159	0	4,88159	0	CR	536
53	NH03	135	111	1	0			Y	6,44764	0	6,44764	0	CR	495
72	NH04	126	116	1	0		Y	Y	7,24162	0	7,24162	0	CR	485
72	NH03	152	117	1	1			Y	4,06845	1	4,06845	1	CR	373
47	NH02	130	118	1	0	D		Y	8,51745	0	3,00616	1	CR	401
64	NH04	155	118	1	0	D		Y	8,42437	1	8,26283	1	CR	443
54	NH03	131	118	1	1			Y	4,64613	0	4,64613	0	CR	520
66	NH04	125	118	2	2		Y	Y	5,46475	0	5,46475	0	CR	502
37	NH02	152	119	1	0			Y	13,2457	0	13,24572	0	CR	351
48	NH02	158	119	2	2		Y	Y	6,70226	0	6,70226	0	CR	478
83	NH04	161	120	1	0	D		Y	4,78303	1	3,87132	1	CR	437
59	NH02	137	120	1	1	D		Y	5,61807	0	3,48255	1	CR	529
40	NH02	144	122	1	0	D		Y	6,08624	0	2,48049	1	CR	454
67	NH03	137	122	2	2			Y	10,0534	0	10,05339	0	CR	377
32	NH03	130	122	1	0		Y	Y	7,05544	0	7,05544	0	CR	475
64	NH03	134	123	2	1			Y	2,69678	1	2,69678	1	CR	389
57	NH03	151	124	1	0	D		Y	7,93703	0	3,26352	1	CR	449
63	NH03	139	125	1	0			Y	11,3073	0	11,30732	0	CR	360
71	NH02	154	125	2	0			Y	4,60233	0	4,60233	0	CR	537
57	NH02	125	126	1	0	D		Y	3,87406	0	0,34497	1	CR	533
43	NH04	120	126	1	2			Y	5,3963	0	5,3963	0	CR	400
64	NH04	144	127	1	0			Y	5,63723	0	5,63723	0	CR	503
55	NH03	155	128	1	1	D		Y	6,47775	0	3,65777	1	CR	424
47	NH03	147	128	1	1			Y	7,11841	0	7,11841	0	CR	463
56	NH04	163	129	1	0			Y	7,33196	0	7,33196	0	CR	436
34	NH03	132	129	1	3			Y	6,99795	0	6,99795	0	CR	470
66	NH02	139	129	1	0			Y	4,97467	0	4,97467	0	CR	530
56	NH04	126	129	1	0	D	Y	Y	7,14853	1	1,69747	1	CR	398

Çizelge 4.4.1.1.3.Hastalığa Bağlı ve Hastalık Dışı Ölümlerde Kaplan Meier ve Yarışan Risk Sonuçları

Hastalığa Bağlı Sağ Kalım Olasılığı				Hastalık Dışı Sağ Kalım Olasılığı			
zaman	CR	KM	CR-KM	zaman	CR	KM	CR-KM
0,5585	0,9057	0,9057	0,0000	0,5585	0,9982	0,9979	0,0003
0,9884	0,8595	0,8593	0,0002	0,9884	0,9908	0,9895	0,0013
1,4182	0,8132	0,8125	0,0007	1,4182	0,9871	0,9850	0,0021
1,9630	0,7669	0,7652	0,0017	1,9630	0,9796	0,9756	0,0040
2,6721	0,7132	0,7100	0,0032	2,6721	0,9722	0,9654	0,0068
3,4524	0,6704	0,6654	0,0050	3,4524	0,9667	0,9574	0,0093
4,8569	0,6246	0,6167	0,0079	4,8569	0,9476	0,9279	0,0197
5,7878	0,6001	0,5902	0,0099	5,7878	0,9414	0,9175	0,0239
7,0746	0,5694	0,5563	0,0131	7,0746	0,9260	0,8906	0,0354
9,3169	0,5202	0,4997	0,0205	9,3169	0,9054	0,8518	0,0536
11,5044	0,4861	0,4590	0,0271	11,5044	0,8882	0,8156	0,0726
19,8166	0,4382	0,3935	0,0447	19,8166	0,8114	0,6343	0,1771
31,1020	0,4282	0,3778	0,0504	31,1020	0,6548	0,2197	0,4351

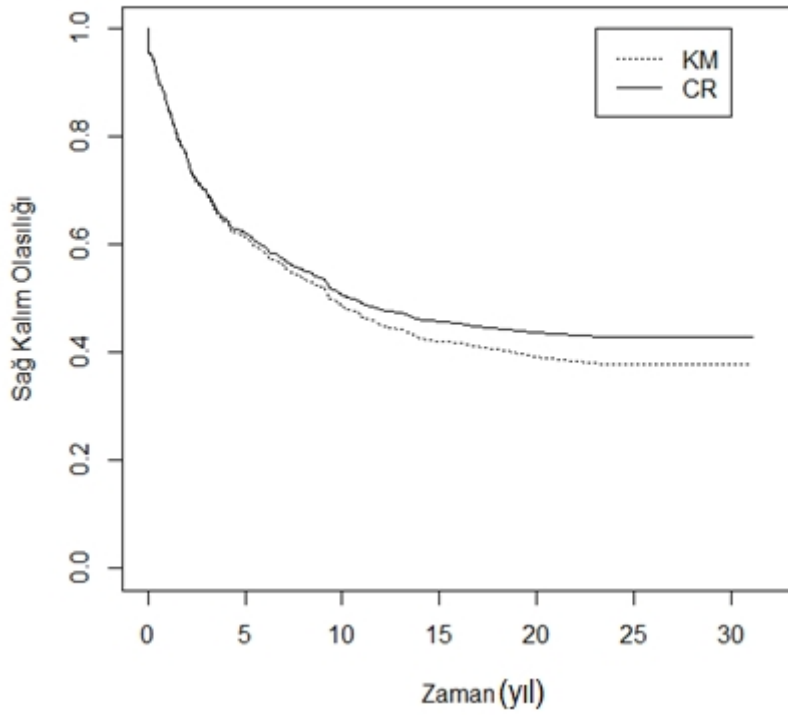
Çizelge 4.4.1.1.3’de Kaplan Meier (KM) ve Yarışan risk (CR) yaklaşımlarının sağ kalım olasılıklarının farklarının ortalamalarının hastalığa bağlı ve hastalık dışı ölümlerdeki değişimleri verilmiştir. İki yönteminde farklı olduğu gözlemlenmektedir. Hastalığa bağlı ölümlerde CR değerlerinin 0.4282 iken KM değerinin 0.3787 olduğu görülmektedir. CR’ nin KM’ den yüksek olduğu görülmektedir. Hastalık Dışı ölümlerde bu değer CR’ de 0.6548 olduğu görülürken KM ‘nin 0.2197’dir. Hastalık Dışı ölümlerde KM’ nin daha düşük sonuç verirken CR daha yüksek sonuç verdiği görülmektedir. İki değer farkı hastalığa bağlı ve hastalık dışı durumda farklarının ortalaması alınarak incelendiğinde hastalığa bağlı ölümlerde fark 0.0504 iken hastalık dışı ölümlerde fark değeri 0.4351 olarak bulunmuştur.



Şekil 4.4.1.1.2.Kümülatif Ölüm Olasılığı Fonksiyonunun Hastalığa Bağlı ölümler ile Hastalık Dışı Durum grafiği

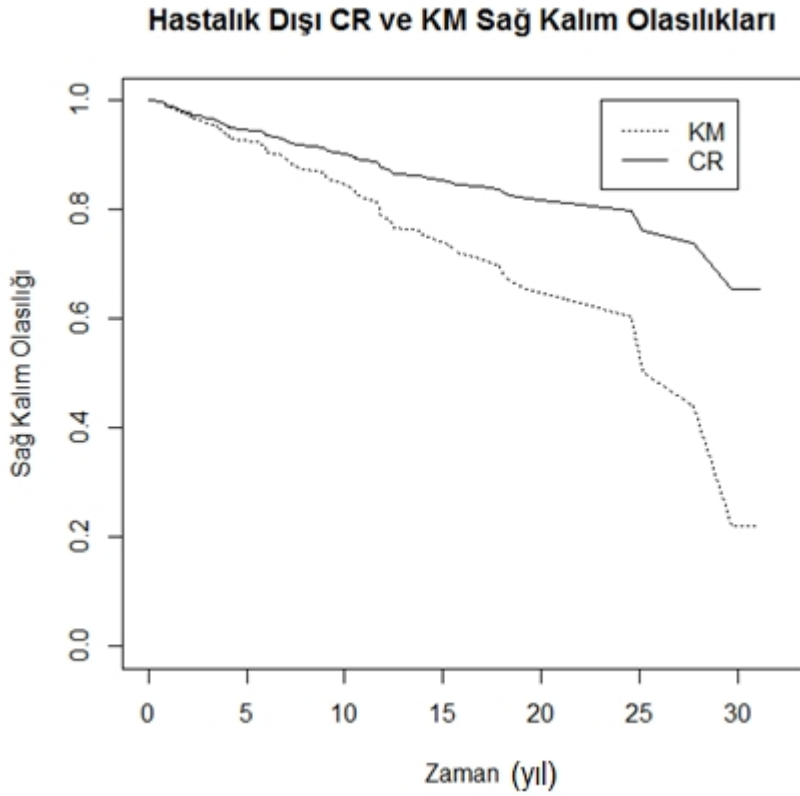
Şekil 4.4.1.1.2’de folliküler hücre lenfoma çalışmasının kümülatif ölüm olasılığı fonksiyonunun hastalığa bağlı ve hastalık dışı durumlarda nasıl bir seyir izlediği gözlemlenmiştir. Gerçek veri setinin Hastalığa bağlı ve hastalık dışı durumlardaki durumları farklı olduğu gözlemlenmiştir. Artan bir ölüm olasılığı olduğu görülmektedir.

Hastalığa Bağlı CR ve KM Sağ Kalım Olasılıkları



Şekil 4.4.1.1.3.Hastalığa Bağlı Ölümlerde CR ve KM sağ kalım olasılıkları

Şekil 4.4.1.1.3’de Hastalığa bağlı ölümler ile hastalık dışı ölümlerin CR ve KM değerlerinin sağ kalım olasılıklarının zamana göre (yıl) grafiklerle verilmiştir. Grafikteki sonuçlara göre iki yöntem arasında farklılık olduğu gözlemlenmektedir. CR değerlerinin daha yüksek sağ kalım tahminleri yaptığı görülürken KM değerlerinin hem hastalığa bağlı durumlarda hem de hastalık dışı durumlarda daha düşük sonuçlar verdiği görülmektedir. Bu farkın hastalık dışı ölümlerde daha fazla olduğu gözlemlenmektedir



Şekil 4.4.1.1.4.Hastalık Dışı Ölümlerde CR ve KM sağ kalım olasılıkları

Şekil 4.4.1.1.4 'de Hastalığa bağlı ölümler ile hastalık dışı ölümlerin CR ve KM değerlerinin sağ kalım olasılıklarının zamana göre (yıl) grafiklerle verilmiştir. Grafikteki sonuçlara göre iki yöntem arasında farklılık olduğu gözlemlenmektedir. CR değerlerinin daha yüksek sağ kalım tahminleri yaptığı görülürken KM değerlerinin hem hastalığa bağlı durumlarda hem de hastalık dışı durumlarda daha düşük sonuçlar verdiği görülmektedir. Bu farkın hastalık dışı ölümlerde daha fazla olduğu gözlemlenmektedir

4.2. Weibull Dağılımından Üretilen Simülasyon Sonuçları

Weibull dağılımından üretilerek oluşturulan veri seti üzerinde yapılan karşılaştırmalarda, veri üretim kısmında örneklem sayılarını $n=100, 150, 250, 500, 1000$ olduğunda ki durumlar değerlendirilirken simülasyonun tekrar sayısı 1000

alınmıştır. Çalışmada Kaplan Meier (KM) ve Yarışan Riskler (CR) yöntemleri örneklem sayılarına ve dağılıma göre CR-KM değerlerinin sağ kalım olasılıklarının farklarının ortalama değerlerinin nasıl seyrettiği incelenmiş ve grafiklerle sonuçlar verilmiştir. Ayrıca yorum yapılırken Hastalığa bağlı durumlar ve Hastalık dışı durumlar göz önüne alındığında bu iki yöntemin farklarının ortalaması alınarak sonuçlar sunulmuştur.

Çizelge 4.4.2.1.Weibull Dağılımından Üretilen Veri Seti

n	p(oran)	Hastalığa Bağlı(CR-KM) Sağ Kalım olasılığı	Hastalığa Bağlı(CR-KM) Sağ kalım olasılığı
100	0,1	0,00930	0,08676
100	0,2	0,01580	0,07453
100	0,3	0,01562	0,06792
100	0,4	0,00686	0,07094
100	0,5	0,00960	0,07172
100	0,6	0,09313	0,06014
100	0,7	0,00497	0,05333
100	0,8	0,00236	0,07393
100	0,9	0,01177	0,07004
150	0,1	0,00803	0,08445
150	0,2	0,00424	0,07266
150	0,3	0,01027	0,07803
150	0,4	0,00998	0,06901
150	0,5	0,00449	0,04713
150	0,6	0,00375	0,07407
150	0,7	0,00105	0,07366
150	0,8	0,00417	0,08893
150	0,9	0,00425	0,10137
250	0,1	0,02511	0,05155

Çizelge4.4.2.2.(Devam)

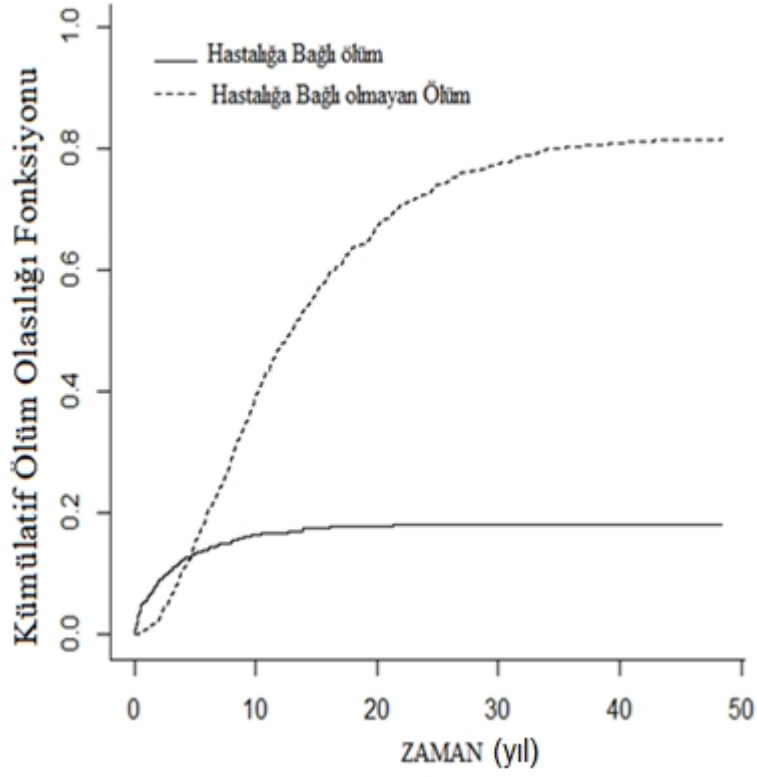
n	p(oran)	Hastalığa Bağlı(CR-KM) Sağ Kalım olasılığı	Hastalığa Bağlı(CR-KM) Sağ kalım olasılığı
250	0,2	0,00472	0,08488
250	0,3	0,01655	0,07684
250	0,4	0,01135	0,05296
250	0,5	0,00690	0,07578
250	0,6	0,00389	0,06706
250	0,7	0,00733	0,07396
250	0,8	0,00507	0,07245
250	0,9	0,00562	0,07669
500	0,1	0,00270	0,07118
500	0,2	0,00429	0,06418
500	0,3	0,00528	0,05468
500	0,4	0,00336	0,06326
500	0,5	0,00538	0,07014
500	0,6	0,00324	0,06963
500	0,7	0,00432	0,07640
500	0,8	0,00566	0,07221
500	0,9	0,00534	0,08269
1000	0,1	0,00481	0,06363
1000	0,2	0,00573	0,06376
1000	0,3	0,00615	0,07015
1000	0,4	0,00692	0,06781
1000	0,5	0,01041	0,06728
1000	0,6	0,00456	0,07420
1000	0,7	0,00369	0,07284
1000	0,8	0,00598	0,07025
1000	0,9	0,00343	0,08034

Çizelge 4.4.2.1 ve Çizelge 4.4.2.1’de simülasyonun üretim sonuçları özet olarak verilmiştir. Örneklem sayısı n=100, 150, 250, 500, 1000 olup p(oran)= 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 alındığında Hastalığa bağlı ölümlerdeki CR-KM farkının ortalama değerleri ile Hastalık dışı durumlarda ölümlerin sağ kalım olasılıklarının CR-KM farklarının ortalama değerleri verilmiştir. Veri setini oluştururken 1000 tekrarlı toplam 100000 veri üretilerek sonuçlar elde edilmiştir.

Çizelge 4.4.2.3.Weibull Dağılımından Üretilen Veri Sonuçlarının Tanımlayıcıları

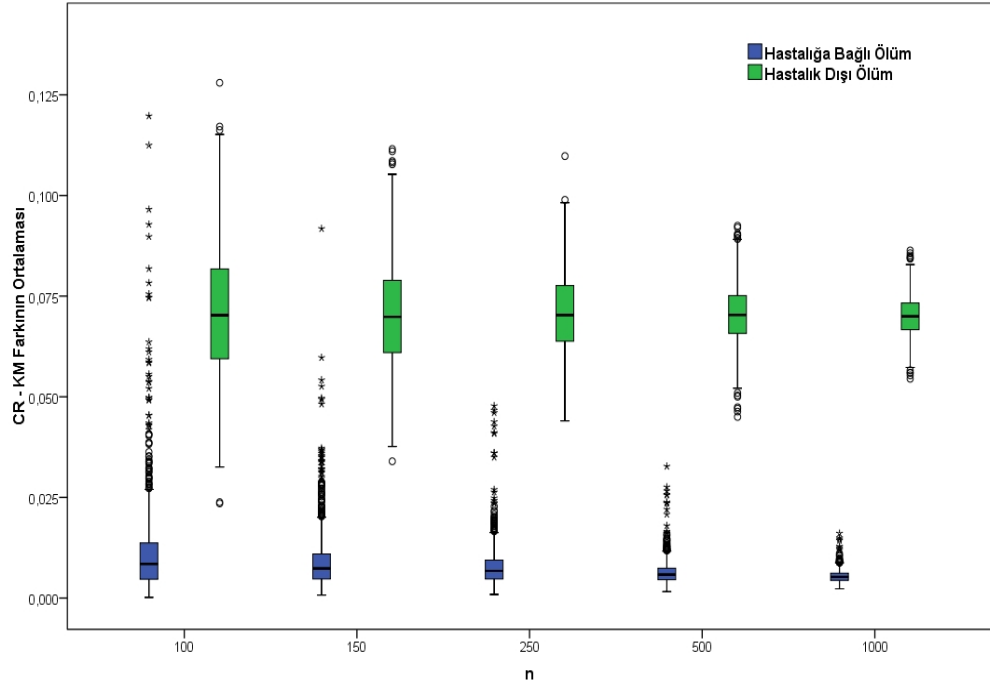
n	Hastalığa Bağlı Ölüm CR-KM (Ort±SD)	Hastalığa Dışı Ölüm CR-KM Ort±SD
100	0.011±0.012	0.721±0.162
150	0.093±0.085	0.717±0.132
250	0.075±0.053	0.713±0.103
500	0.061±0.029	0.709±0.074
1000	0.052±0.017	0.707±0.052

Çizelge 4.4.2.3’de Hastalığa bağlı ve hastalık dışı durumlarda üretilen KM ve CR sonuçlarının farklarının örneklem sayılarına göre ortalama ve standart sapma değerleri verilmiştir. Hastalığa bağlı ölümlerde CR-KM değerinin n=100 olduğunda ortalaması 0.011 iken standart sapma değeri 0.012’dir. Hastalık dışı durumlarda ise sağ kalım olasılıklarının farkların ortalama sonuçları n=100 iken ortalaması 0.721iken standart sapması 0.162’dur. Aynı şekilde n=150 iken ortalama 0.093 standart sapma ise 0.085’dir. Hastalık dışı durumda ise ortalama 0.712 iken standart sapma 0.132’dir.Örneklem n=250 iken hastalığa bağlı durumda ortalama 0.075 iken standart sapma 0.053’tür. Hastalık dışı durumda ise ortalama 0.713iken standart sapma değeri 0.103’dir. Hastalığa bağlı durumlarda n=500 iken ortalama 0.061 standart sapma değeri 0.029’dir.Hastalık dışı durumda ortalama 0.709 standart sapma değeri 0.074’dir. Hastalığa bağlı durumlarda n=1000 iken ortalama değer 0.052 standart sapma değeri 0.017.Hastalık dışı durumda ise ortalama 0.707 standart sapma değeri 0.052’dir. Bu sonuçların hepsi daha iyi açıklanabilmek için 100 değeri ile çarpılarak verilmiştir.



Şekil 4.4.2.1.Hastalık Dışı Ölümlerde Kümülatif Ölüm Olasılığı Grafiği

Şekil 4.4.2.1’de weibull dağılımından üretilen simülasyon sonuçlarına göre hastalığa bağlı ve hastalık dışı durumlarda zamana göre(yıl) kümülatif ölüm olasılığı fonksiyonunun artan bir seyir izlediği gözlemlenmektedir. Hastalığa bağlı ve hastalık dışı durumlarda kümülatif ölüm olasılığı fonksiyon sonuçlarına göre Hastalığa bağlı ölümlerin 0.2 civarında seyrederek sabitlendiği gözlemlenirken hastalık dışı durumda kümülatif ölüm olasılığı 0.8 civarında seyretmektedir.

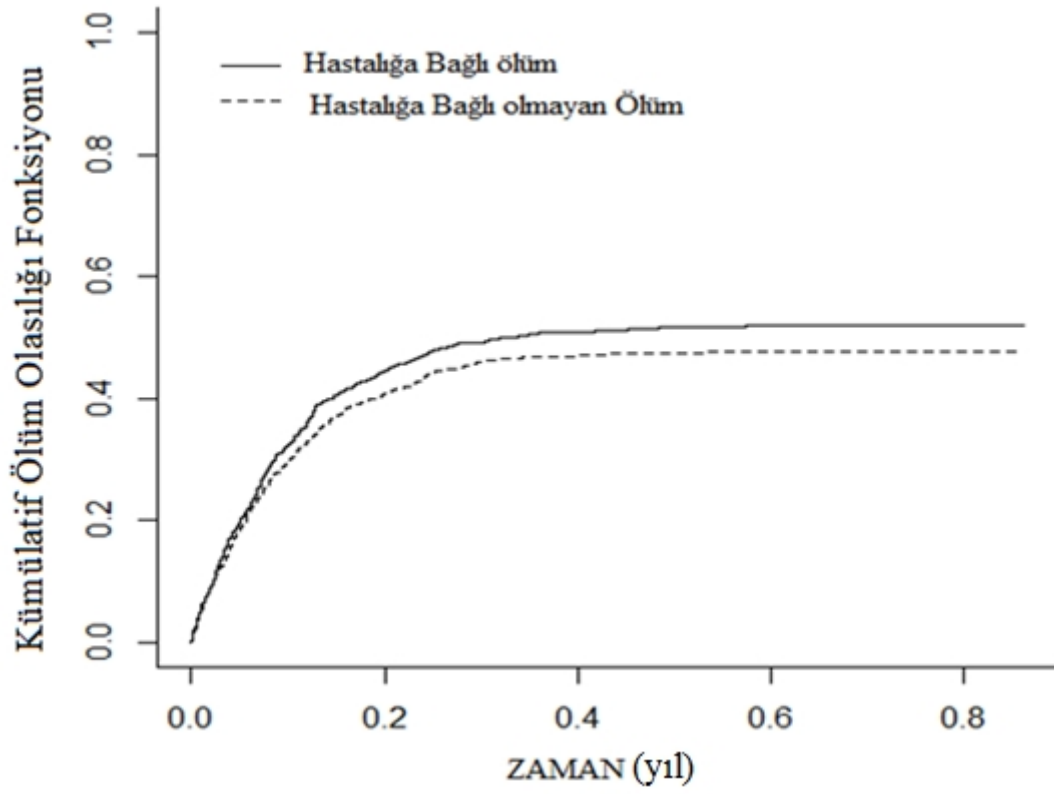


Şekil 4.4.2.2.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde Örneklem Sayısına Bağlı CR-KM Farkının Ortalamasının Box Plot Grafiği

Şekil 4.4.2.2’de hastalığa bağlı ölümler ile hastalık dışı ölümlerde CR-KM farkının ortalamasının sonuçları örneklem büyüklüğü ile olan ilişkisi verilmiştir. Kaplan Meier ve Yarışan risk sağ kalım olasılıklarının weibull dağılımına göre üretilen veri setinde farklarının ortalamaları arasında hastalığa bağlı ve hastalık dışı durumlarda farklılık olduğu anlaşılırken bu farklılığın çok yüksek olmadığı gözlemlenmektedir. Hastalık dışı ölümlerde CR-KM farklarının ortalama değeri yaklaşık 0.075 seyrederken örneklem sayısına göre bu değer çok az değişirken hastalığa bağlı durumda CR-KM farklarının ortalama sonuçları hastalık dışı durumlara göre daha düşük bir değer içermektedir. Hastalığa bağlı durumlarda CR-KM farkların ortalaması değeri yaklaşık 0.010 civarında seyretmektedir.

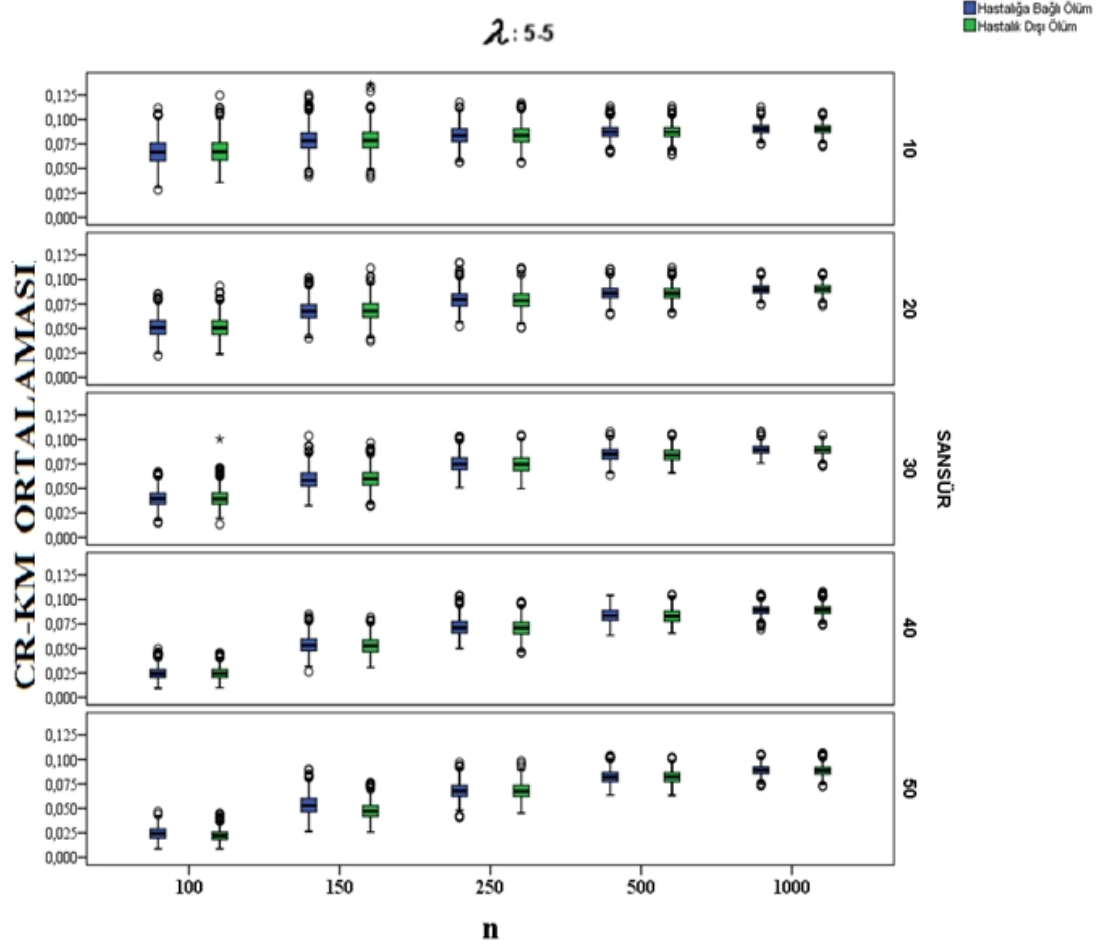
4.3.Üstel Dağılımından Üretilen Simülasyon sonuçları

Üstel dağılımından veri üretilerek oluşturulan veri seti üzerinde yapılan karşılaştırılmalarda, veri üretim kısmında örneklem sayılarını $n=100, 150, 250, 500, 1000$ alındığında simülasyonun tekrar sayısı 1000 yapılarak sonuçlar değerlendirilmiştir. Çalışmada Kaplan Meier (KM) ve Yarışan Riskler (CR) yöntemleri örneklem sayılarına ve dağılıma göre CR-KM değerlerinin sağ kalım olasılıklarının farklarının ortalama değerleri nasıl bir değişim olduğu incelenmiş ve grafiklerle sonuçlar verilmiştir. Ayrıca yorum yapılırken Hastalığa bağlı durumlar ve Hastalık dışı durumlar göz önüne alındığında bu iki yöntemin farklarının ortalaması alınarak sonuçlar sunulmuştur. Üstel dağılıma göre üretilen veri setinde hastalığa bağlı ve hastalık dışı durumda lamda katsayılarının değişimi ve sansür oranlarının değişimine göre sonuçlar sunulmuş ve grafiklerle yorumlama kolaylığı sağlanmıştır. Sansür oranı %10, 20, 30, 40, 50 olduğunda üstel dağılım parametresindeki değişim hastalığa bağlı ve hastalık dışı durumda $(I_1 = I_2 = 5), (I_1 = 5, I_2 = 6), (I_1 = 5, I_2 = 7), (I_1 = 5, I_2 = 8)$ durumlar incelenmiştir.



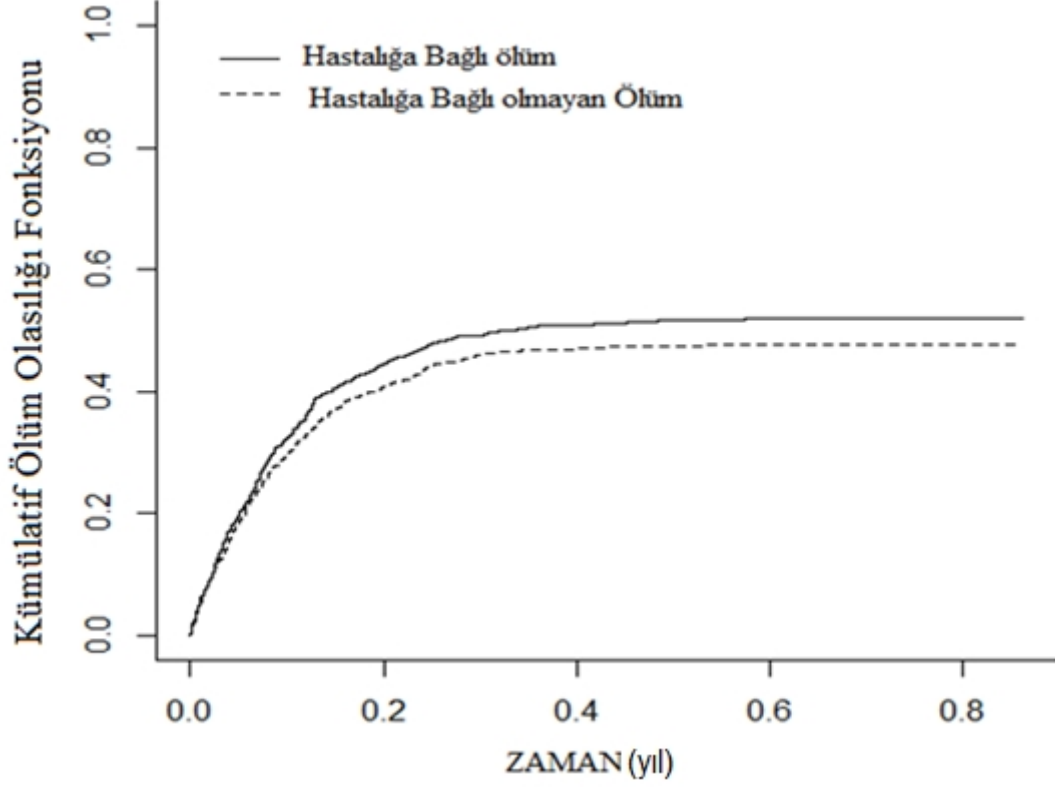
Şekil 4.4.3.1.Hastalık Dışı Ölümlerde $\lambda_1 = 5, \lambda_2 = 5$ Kümülatif Ölüm Olasılığı Grafiği

Şekil 4.4.3.1’de üstel dağılımda üstel dağılım parametresi $\lambda_1 = \lambda_2 = 5$ sabit aldığımızda kümülatif ölüm olasılığı fonksiyonunun hastalığa bağlı ve hastalık dışı durumda nasıl bir durum sergilediğini grafikte gösterilmektedir.



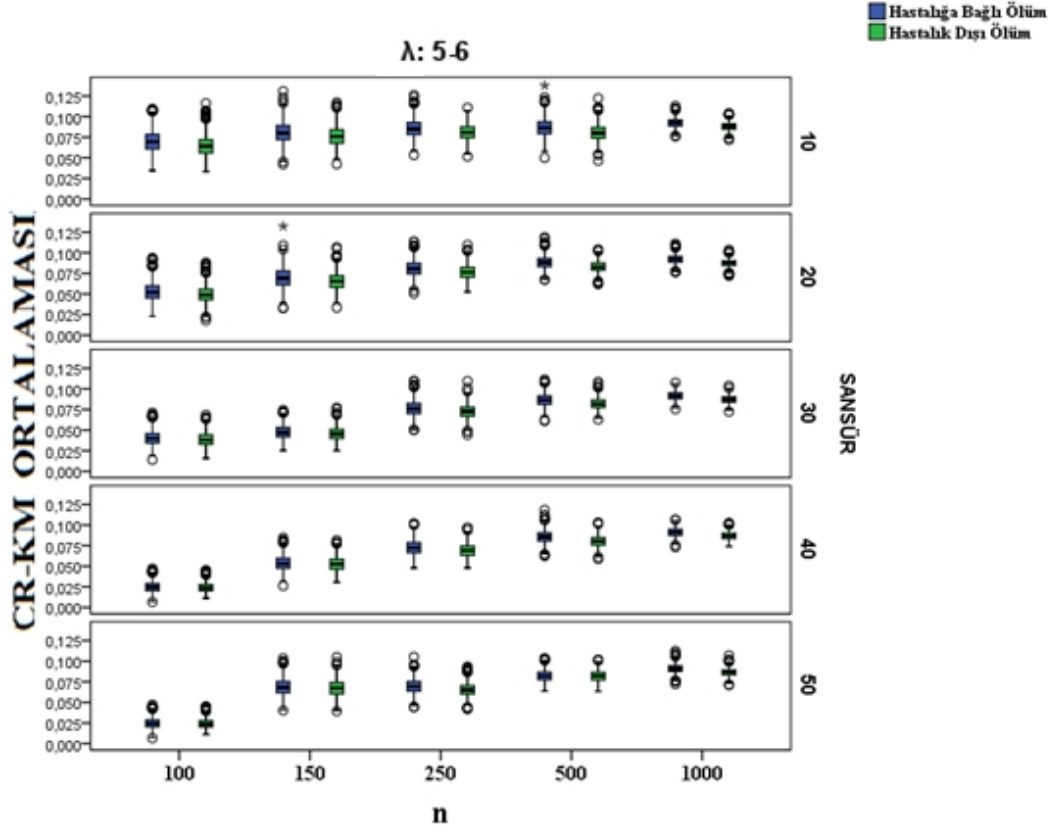
Şekil 4.4.3.2.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 = 5, \lambda_2 = 5$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalamasının Box Plot Grafiği

Şekil 4.4.3.2’de üstel dağılımda dağılım parametresi $\lambda_1 = 5, \lambda_2 = 5$ aldığımızda CR-KM farkların ortalama sonuçlarının hastalığa bağlı ve hastalık dışı durumda sansür oranı ve örneklem sayısı değişiminde nasıl bir durum gösterdiği grafikte gösterilmektedir. Sansür oranı ile örneklem sayısının değişimine göre yorum yapılmaktadır. Hastalığa Bağlı sebeplerde lamdalar eşit ve sabit alındığında örneklem sayısı azaldıkça sansürleme oranındaki artış CR-KM arasındaki farkı azaltmaktadır. %10 ve %20 sansür oranında örneklem büyüklüğü arttıkça farkın değerinin yükseldiği fakat bu sonucun %30, 40, 50 değerlerine göre çok yüksek olmadığı gözlemlenmiştir. %40 ve %50 CR-KM değerinin çok daha düştüğü görülmektedir. Örnekleme büyüklüğü küçüldükçe sansürleme etkisi ortaya çıkmaktadır.



Şekil 4.4.3.3.Hastalık Dışı Ölümlerde $\lambda_1 =5$, $\lambda_2 =6$ Kümülatif Ölüm Olasılığı Grafiği

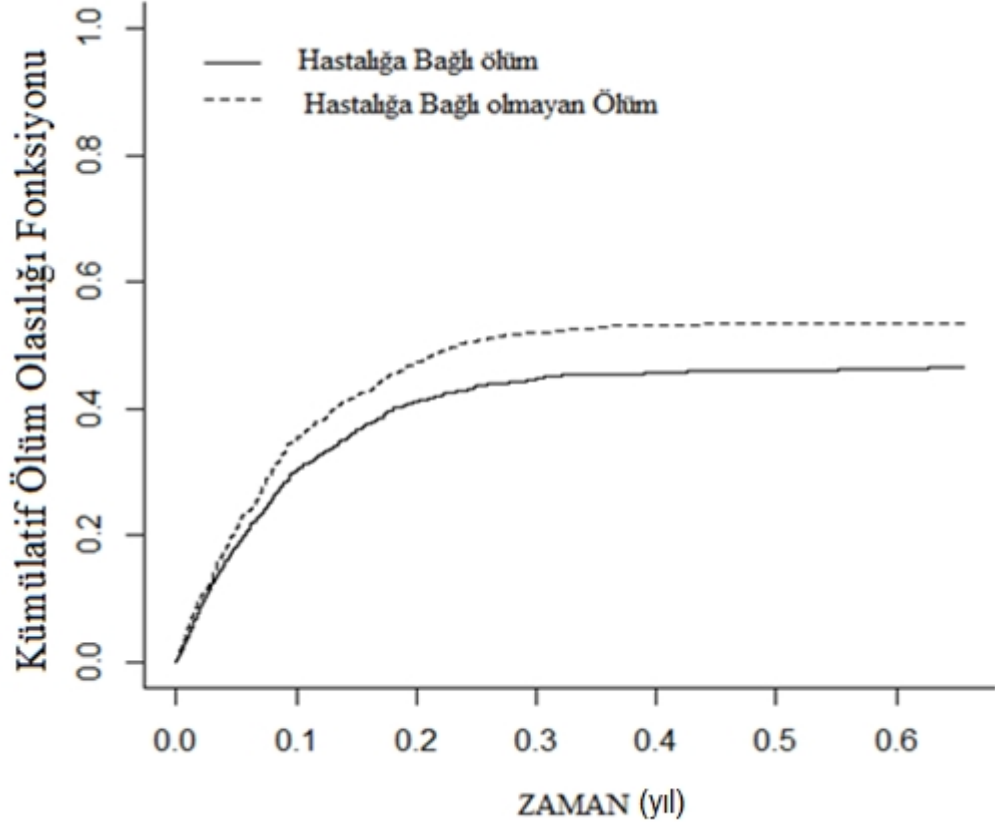
Şekil 4.4.3.3’de üstel dağılımda üstel dağılım parametresi $\lambda_1 =5$, $\lambda_2 =6$ sabit aldığımızda kümülatif ölüm olasılığı fonksiyonunun hastalığa bağlı ve hastalık dışı durumda nasıl bir durum gösterdiği grafikte verilmektedir.



Şekil 4.4.3.4. Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 = 5$, $\lambda_2 = 6$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalamasının Box Plot Grafiği

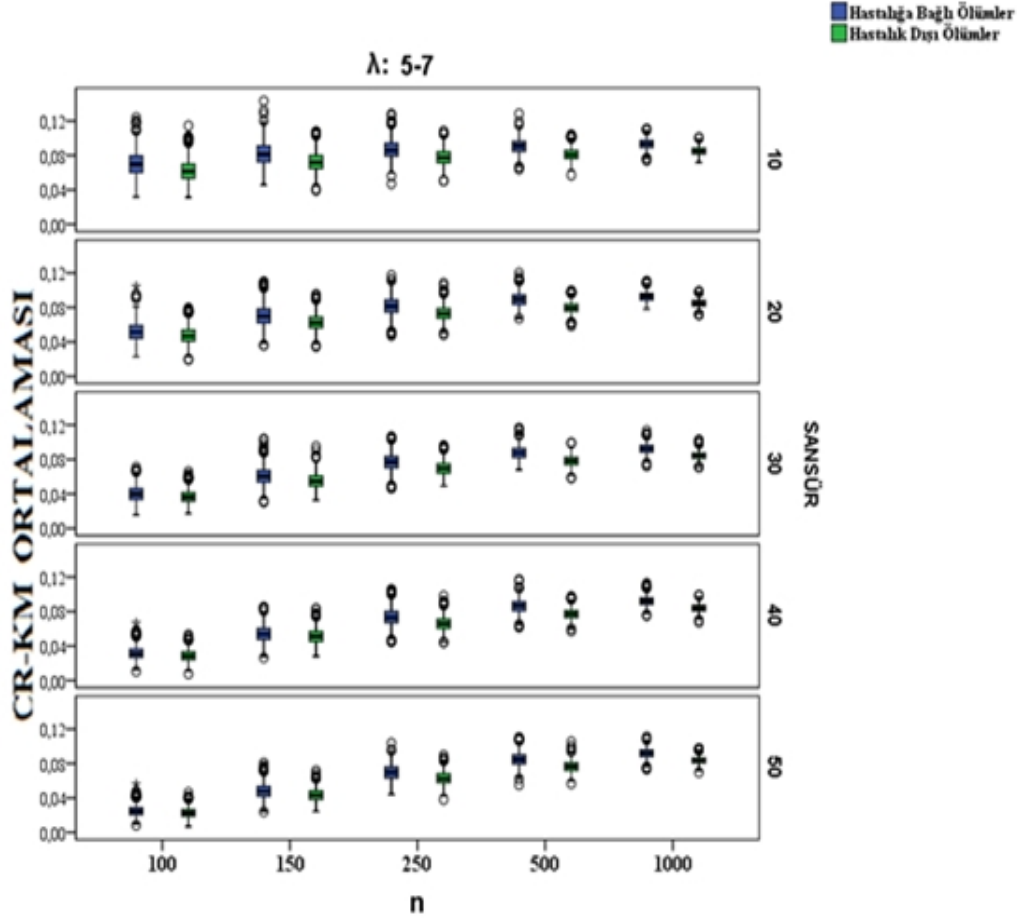
Şekil 4.4.3.4’de üstel dağılımda dağılım parametresi $\lambda_1 = 5$, $\lambda_2 = 6$ aldığımızda CR-KM farkların ortalama sonuçlarının hastalığa bağlı ve hastalık dışı durumda sansür oranı ve örneklem sayısı değişiminde nasıl bir durum izlediği grafikte gösterilmiştir. Sansür oranı ile örneklem sayısının değişimine göre yorum yapılmaktadır. Hastalığa Bağlı sebeplerde lamda $\lambda_1 = 5$, $\lambda_2 = 6$ alındığında örneklem sayısı azaldıkça sansürleme oranındaki artış CR-KM arasındaki farkı azaltmaktadır. %10 ve %20 sansür oranında örneklem büyüklüğü arttıkça farkın değerinin yükseldiği görülmektedir. Sansür oranı %10 olduğunda CR-KM farkların ortalaması değeri örneklem sayısı arttıkça artarken yaklaşık 0.075 ile 0.1 olduğu görülmektedir. Fakat %30, 40, 50 sansür oranlarında farkın örneklem sayısı küçüldükçe çok küçük değerler aldığı görülmektedir. %40 ve

%50 sansürleme oranında CR-KM değerinin çok daha düştüğü görülmektedir. Bu grafik için örnekleme büyüklüğü küçüldükçe sansürleme etkisi ortaya çıkmaktadır.



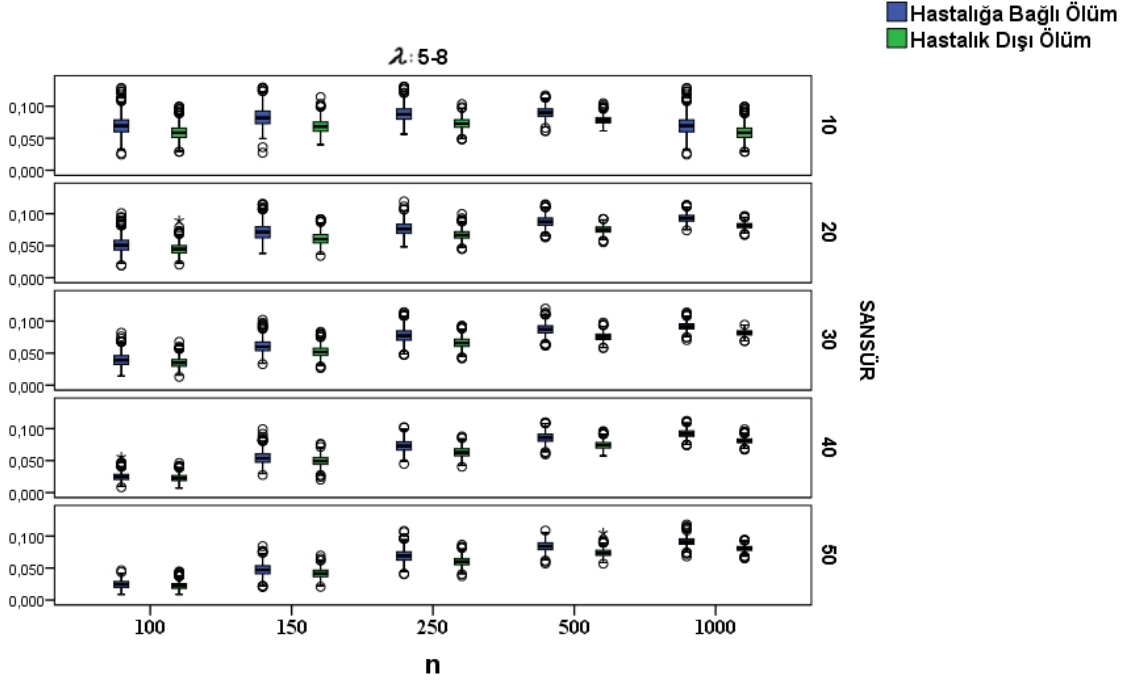
Şekil 4.4.3.5.Hastalık Dışı Ölümelerde $\lambda_1 = 5$, $\lambda_2 = 7$ Kümülatif Ölüm Olasılığı Grafiği

Şekil 4.4.3.5’de üstel dağılımda üstel dağılım parametresi $\lambda_1 = 5$, $\lambda_2 = 7$ sabit aldığımızda kümülatif ölüm olasılığı fonksiyonunun hastalığa bağlı ve hastalık dışı durumda nasıl bir durum sergilediği grafikte gösterilmiştir.



Şekil 4.4.3.6.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 = 5, \lambda_2 = 7$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalamasınının Box Plot Grafiği

Şekil 4.4.3.6’da üstel dağılımda dağılım parametresi $\lambda_1 = 5, \lambda_2 = 7$ aldığımızda CR-KM farkların ortalama sonuçlarının hastalığa bağlı ve hastalık dışı durumda sansür oranı ve örneklem sayısı değişiminde nasıl bir durum gösterdiği grafiklerle verilmiştir. Sansür oranının değişimi ile örneklem sayısının değişimine göre gerekli yorumlar yapılmaktadır. Hastalığa Bağlı sebeplerde lamdalar $\lambda_1 = 5, \lambda_2 = 7$ alındığında örneklem sayısı azaldıkça sansürleme oranındaki artış CR-KM arasındaki farkı azaltmaktadır. %10 ve %20 sansür oranında örneklem büyüklüğü arttıkça farkın değerinin yükseldiği görülmektedir. Sansür oranı %10 olduğunda CR-KM farkların ortalaması değeri örneklem sayısı arttıkça artarken sonuçlar yaklaşık 0.04 ile 0.08 etrafında seyretmektedir. Fakat %30, 40, 50 sansür oranlarında farkın örneklem sayısı küçüldükçe



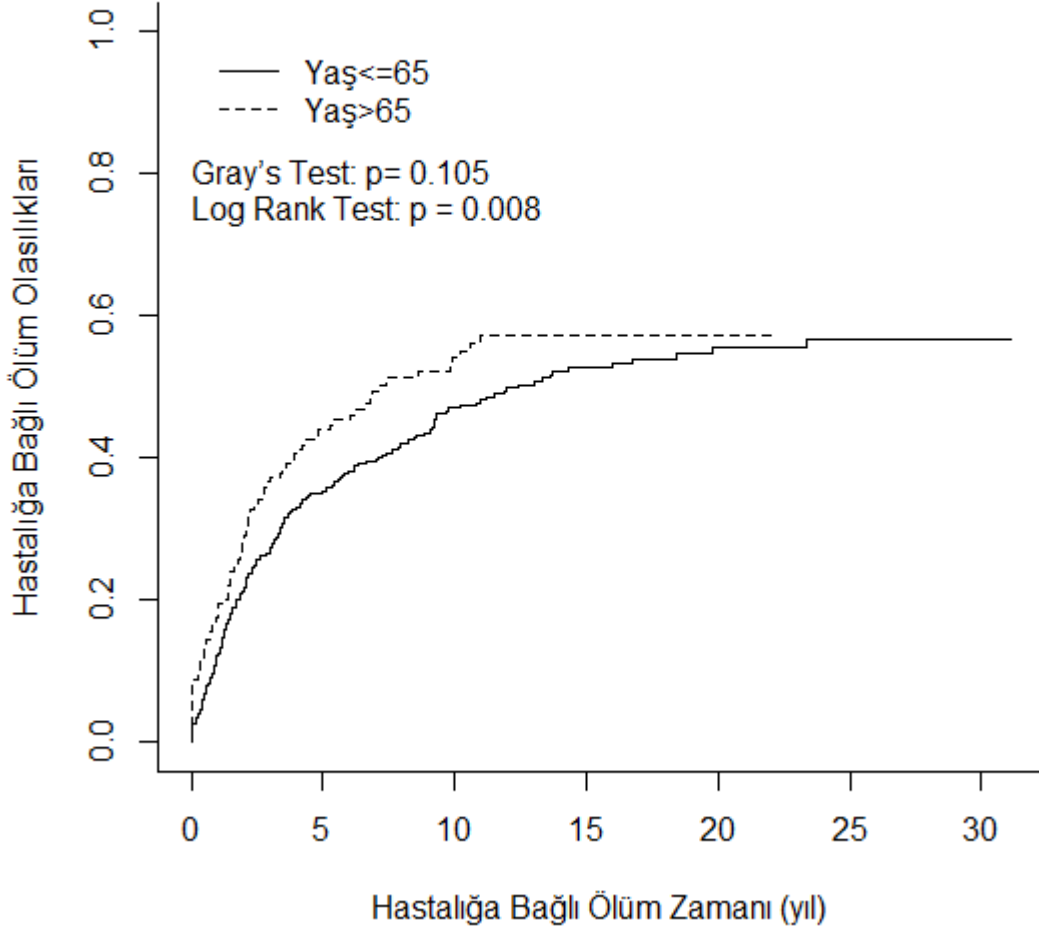
Şekil 4.4.3.8.Hastalık Dışı ve Hastalığa Bağlı Ölümlerde $\lambda_1 = 5$, $\lambda_2 = 8$ Üstel Dağılımla Üretilen Veri setinin Örneklem Sayısına Bağlı CR-KM Farkının Ortalamasının Box Plot Grafiği

Şekil 4.4.3.8’de üstel dağılımda dağılım parametresi $\lambda_1 = 5$, $\lambda_2 = 7$ aldığımızda CR-KM farkların ortalama sonuçlarının hastalığa bağlı ölümlerde ve hastalık dışı ölümlerde sansür oranı ve örneklem sayısı değişiminde nasıl bir durum gösterdiği grafikte gösterilmektedir. Sansür oranı ile örneklem sayısının değişimine göre yorum yapılmaktadır. Hastalığa Bağlı sebeplerde lamdalar $\lambda_1 = 5$, $\lambda_2 = 8$ alındığında örneklem sayısı azaldıkça sansürleme oranındaki artış CR-KM arasındaki farkı azaltmaktadır. %10 ve %20 sansür oranında örneklem büyüklüğü arttıkça farkın değerinin yükseldiği görülmektedir. Sansür oranı %10 olduğunda CR-KM farkların ortalaması değeri örneklem sayısı arttıkça artarken yaklaşık sonuçlar 0.05 ile 0.1 bulunmuştur. Fakat %30, 40, 50 sansür oranlarında farkın örneklem sayısı küçüldükçe çok küçük değerler aldığı görülmektedir. %40 ve %50 sansürleme oranında CR-KM değerinin çok daha düştüğü görülmektedir bu değer 0 ile 0.05 arasında bulunmuştur. Örnekleme büyüklüğü küçüldükçe sansürleme etkisi ortaya çıkmaktadır.

4.4. Gerçek Veri Setinde Eş değişken Olması durumunda Sonuçlar ve Yorumlar

İlgilenilen olay hastalıktan dolayı gerçekleşen bir başarısızlık ise bu durumu daha detaylı açıklarsak hastalıktan dolayı iyileşmeme ya da daha sonra nüks durumları olarak ele alırsak diğer durumları da ölüm olarak ele alırız. Follicüler hücre lenfoması gerçek veri seti için incelendiğinde 541 hastanın 272'si ilgilenilen olaydan kayıt altındadır. Bunlardan 24 tanesi hastalık teşhisi koyulduktan sonra uygulanan tedaviye cevap vermeyip iyileşemeyen grup iken 248 tanesi tedaviden sonra tekrar hastalığı nüks etmiş olan hastaların kayıtlarıdır. Bu hastaların 76 tanesi hastalık teşhisi koyulduktan sonra nüks olmadan ölümü gerçekleşen hastalardır. Eş değişken durumu göz önüne alınarak veri seti için çalışma incelendiğinde yaşa göre sağ kalım olasılık sonuçlarının değiştiği gözlemlenmektedir. Yaş değişkenini iki grup olarak kodladığımızda 1. Grup = $65 >$ ve 2. Grup = ≤ 65 'dir. Test sonuçları Gray's test istatistiğinin Ki-kare test sonucu ile verilmiş anlamlılık düzeyi $p < 0.05$ olarak alınmıştır. Fine ve Gray's test istatistiği sağ kalım olasılıkları dikkate alındığında Log Rank test istatistiği ile benzer formülüzasyonlara sahiptir. Gray's test istatistiğinin sonuçlarını yorumlarken karşılaştırmayı Log rank test istatistiği sonuçları değerlendirilmiştir. Hastalığa bağlı ölümlerde ve hastalık dışı ölümlerde sağ kalım olasılık değerlerinin nasıl bir seyir izlediği aşağıdaki grafiklerle verilmiştir.

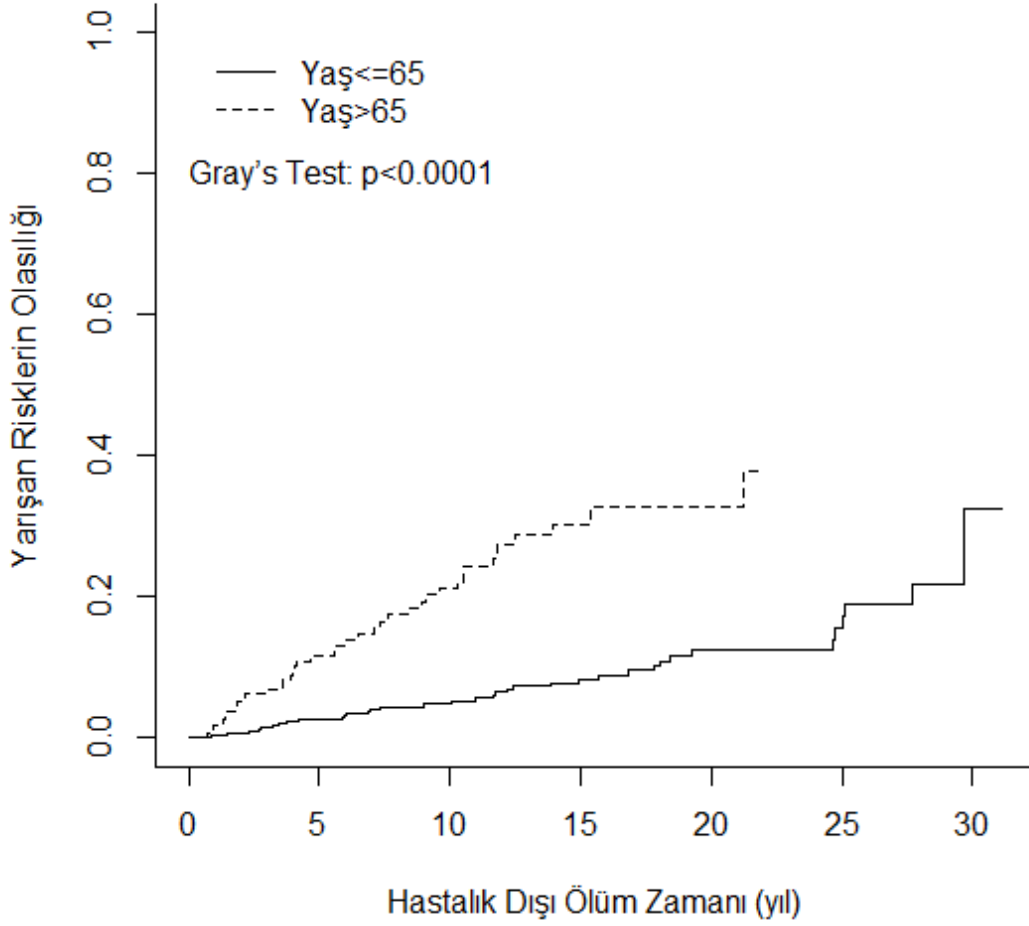
Hastalığa Bağlı Ölümlerde CIF



Şekil 4.4.4.1.Yaşın Hastalığa Bağlı Ölümlerde Kümülatif İnsidans Fonksiyonuna göre Ölüm olasılıkları Grafikleri

Şekil 4.4.4.1'de yaşın iki grupta incelendiğinde $65 >$ ve ≤ 65 hastalığa bağlı ölümler kümülatif insidans fonksiyonunun zamana göre nasıl bir durum izlediği grafikte verilmiştir. Kümülatif insidans fonksiyonu (CIF) hastalığa bağlı durumda Gray's test istatistiği sonucu $p = 0.105$ iken Log Rank test istatistiğinin sonucu $p = 0.008$ 'dir.

Yarışan Riskler İçin CIF



Şekil 4.4.4.2. Yaş Değişkeninin Hastalık Dışı Ölümlerde Yarışan Risk Durumunda Kümülatif İnsidans Fonksiyonuna Göre Ölüm olasılıkları Grafikleri

Şekil 4.4.4.2'de yaşın iki grupta incelendiğinde $65 >$ ve ≤ 65 hastalık dışı ölümlerde kümülatif insidans fonksiyonunun zamana göre nasıl bir durum izlediği grafikte verilmiştir. Kümülatif insidans fonksiyonu (CIF) hastalık dışı durumda Gray's test istatistiği sonucu $p < 0.0001$ ' dir. Gray's test istatistiğinin yarışan risk durumunda eş değişken durumu olduğunda doğru tahminler yaptığı ve hastalık dışı ölümlerde 65 yaş üstü hastaların olasılık değerlerinin daha yüksek çıktığı görülmektedir.

5.TARTIŞMA

Son yıllarda yapılan araştırmalar, sağ kalım verilerinin değerlendirilmesinde yapılan istatistiksel analiz problemlerini ortaya koymakta ve bu problemleri gidermek için gerekli yöntemler geliştirmektedir. Koller ve arkadaşları 2012 yılında yayınladıkları çalışmalarında zamana bağlı olayların analizinde ciddi hatalar yapıldığını ortaya koymuşlardır.³⁴Bu hataları açıklarken yaşam süresini göz önünde bulundurmışlardır.

Yaşam süresi, sağlık düzeyini gösteren en önemli ölçütlerden biridir. Yaşamsal verilerin özelliklerinden dolayı yaygın olarak bilinen istatistiksel yöntemlerle tahminlerin yapılması ya da uygulanan tedavilerin etkinliklerinin belirlenmesi mümkün olmamaktadır. Bu nedenlerden dolayı özgün istatistiksel yöntemler geliştirilmiştir. Sağlık verilerinin analizinde yaşam süreleri hakkında bilgi sahibi olabilmek amacıyla sağ kalım analizi uygulanmaktadır.³⁵

Sağ kalım analizinde veri yapılarına ve test edilecek hipotezlere bağlı olarak farklı yöntemler kullanılmaktadır. Bu yöntemlerin gelişmesi ile birlikte yarışan riskler analizi sağlık alanında ki araştırmalarda sağ kalım analizlerinin eksikliğini tamamlayan bir analiz olarak ortaya koyulmuştur.

Yarışan riskler analizinin matematiksel bir model olarak tanımlanması Crowder'i kitabında detaylı bir şekilde açıklanmıştır.³⁶ Daha sonrasında sağ kalım analizlerinin daha detaylı bir şekilde anlatılması Fine ve Gray tarafından yapılmıştır.¹¹

İzlem çalışmalarında hastaların bir kısmı incelenen hastalık türünden bir kısmı da farklı bir nedenden ölebilirler. Yarışan riskler analizi; ölüm nedeni olarak birden fazla nedenin bulunması durumunda tek bir ölüm sebebi seçilip, diğer nedenlerin göz önüne alınmamasıyla bir nedene özel ölüm risklerinin tahminini içerir. İncelenen hastalık tipinde prognostik değişkenlerin bulunduğu durumlarda parametre tahminleri ve bu faktörlerin ortadan kaldırılması ile yaşam sürelerindeki değişimleri bulmak amacıyla yarışan riskler analizi uygulanmaktadır.³⁷

Yarışan risklerin analizde son zamanlarda kümülatif insidanslar kullanılmaya başlanmıştır.^{19,38} Özellikle kanser vakalarında kanserin tekrarlanmasından korunmak amacıyla parametrik yaklaşımlarda prognostik faktörlerin rolünü değerlendirmek amacıyla kümülatif insidanslar büyük önem taşımaktadır.³⁹

Açıklayıcı eş değişkenlerin de bulunduğu yarışan risk verileri daha detaylı açıklayabilmek için standart analiz, oransal riskler yaklaşımı ile nedene özel risk fonksiyonlarının modellenmesi gerekir.^{12,24,40} Eş değişkenleri incelemede en detaylı bilgiyi modelleme oluşturarak ortaya koyan Fine ve Gray'sdir.¹² Alt dağılım ölüm olasılıkları olarak daha detaylı açıklamalarla araştırmacılara yol göstermiştir. Çalışmalarda hesaplanan riskler incelenen ölüm nedenine özel riskleri vermemektedir. Belirli bir nedenden ölümlerde ölüm verileri kullanılarak bulunan kaba ölüm olasılıkları, net yaşam fonksiyonlarının açıklanması için iyi bir araçtır.⁴¹ İncelenilen ölüm nedenine özel risklerin bulunmasında kaba ölüm olarak isimlendirilen kümülatif insidans grafiklerinden yararlanılmaktadır.³⁸ Bu grafik değerleri incelenen bir olayın marjinal olasılık değerleridir. Ölüm sayıları ve risk altındaki nüfus aracılığı ile bulunan kümülatif insidans değerleri açıklayıcı bilgiler verse de grafiksel gösterimi daha belirgin farkları ortaya koymaktadır.^{40, 42}

Yapılan literatür araştırması sonucunda çalışmaya uygun bulgu karşılaştırmaları bulunamamıştır fakat yöntemlerdeki eksikliği ortaya koyan birçok çalışma mevcuttur. Gerek gerçek veri seti üzerinde ki uygulama örnekleri olsun gerekse yapılan simülasyon çalışmaları yarışan risklerin sağ kalım analizlerinden farklılığını göstermek amaçlı hem de geliştirilen methodları daha detaylı açıklamak için sunulmuştur. Nuria Porto 2008 yılında yaptığı çalışmada follicüler hücre lenfoma hastası 541 kişinin verilerinin oluşturduğu veri setinde genel sağ kalım analizleri ile hem yarışan riskler yöntemini hem de kendi geliştirdiği benzer matematiksel bir modeli tanıtmıştır. Porto'nun sonuçları çalışma sonuçlarını karşılaştırma fırsatı vermiştir. Nuria Porto çalışmasında sağ kalım analizi ile sunulan sağ kalım sonuçlarının yarışan risklerle yapılan sağ kalım olasılık sonuçlarından düşük olduğunu göstermiş modelledikleri nedene özel sağ kalım olasılıklarına benzer bir fonksiyon oluşturup sonuçları karşılaştırmada bulunmuştur.⁴³

Pepe ve Mori çalışmalarında 1981 ve 1988 yılları arasında Fred Hutchinson kanser araştırma merkezinde yapılan çalışmada kemik iliği nakli yapılan akut GVHD(Graft Versus Host Disease) hastalarının kronik GVHD hastalarına göre iyileşme gösterdiklerini belirlemiştir. Ayrıca yarışan risklerinde Kaplan Meier grafiklerinin bir sonuç vermediği gözlemlenmiştir. Nedene özel kümülatif insidans grafiğinde GVHD nedeniyle ölen hastaların akut ve kronik GVHD hastalarında insidanslarının benzer

olduđu, diđer nedenlerden ölen akut GVHD hastalarında kronik GVHD hastalarına göre insidanslarının daha yüksek olarak bulunduđu gözlemlenmiştir.^{44,14}

Yarıřan risk verilerinin regresyon modellerini açıklarken Bernard Haller'in çalışmasında, yarıřan risk modellerinde regresyon modelini simülasyon çalışması ile açıklamıştır. Alt dağılım regresyon modelinin eş deđişken durumu göz önüne alındığında daha uygun sonuçlar verdiđi fakat nedene özel ölüm olasılığı regresyon modellerinin de veri setine uygun olarak tercih edildiđini çalışmasında yer vermiştir. Diđer birçok yönteminde Pseudo, vertical, Karma Modelide alt dağılım ölüm olasılığı modeli ile karşılaştırılmaları yapılmış eş deđişken durumda tercih edilmesi daha uygun bulunduđu söylenmiştir.⁴⁵

Yarıřan risklerin genel sağ kalım analizleri ile karşılaştırılmasında Christine Eulenburg yumurtalık kanserlerinde ilerleme durumu ve ölüm durumunu inceleyerek yarıřan risk analizi ile deđerlendirme yapıldığında sonuçların sağ kalım analiz sonuçlarının eksik bilgi verdiđi gözlemlenmiştir. Çalışmada genel sağ kalım analizi olarak uygulanan Kaplan Meier yönteminin hastalık dışı durumları sansürlü veri olarak aldığından sağ kalım olasılıkları daha düşük sonuçlar verdiđi ve bu yöntemdeki problemleri gidermek için yarıřan risk analizini kullanmanın dođru olduđunu öne sürmektedir.⁴⁶

Sađ kalım analizlerinin yorumlanmasında Thomas H. Scheike'nin simülasyon çalışması, yarıřan risk analizlerinin uygulama alanına katkıda bulunacak açık kodlu R programı kolaylıklarını sunmaktadır. Fine ve Gray modelini kullanarak üretilen sağ kalım olasılık sonuçlarını kendi yazdıkları açık kodlar (timereg) yardımıyla gerekli karşılařtırmaları yapmaktadır. Sađ kalım olasılıklarının eksikliklerini geliřtirmiş oldukları yarıřan risk modellerinin R paket programının kodları yardımıyla kullanıcılara farklı bir açı sunmaktadırlar. Yeni geliřtirdikleri paket programları ile güven aralıkları tahmininde daha güvenli bir sonuç sunduklarını göstermişlerdir.⁴⁷

Sađlık sistemlerinde geliřmelerin daha hızlı olabilmesi için birincil kořul iyi bir kayıt sisteminin bulunmasıdır. Dođru, yeterli ve güncel kayıtlara ulařılabilirse yapılacak çalışmaların daha hızlı, daha dođru ve daha güvenilir olacađı açıktır. Sađ kalım verilerinin tutulmasında detaylı bilgi dođru bir analiz sonucu vermek için şarttır. Aksi takdirde hastanın sağ kalım sonuçları dođru yorumlanamayacağından dođru bir bulgu sunmak yanlıř olacaktır.

6.SONUÇLAR VE ÖNERİLER

Çalışmada sağ kalım analiz sonuçlarını yarışan riskler analiz sonuçları ile karşılaştırırken dağılım ve örneklem sayılarındaki değişimleri incelemek amacıyla bir simülasyon çalışması sunulmuştur. Gerçek veri seti 1967 yılında Toronto'da Princess Margaret hastanesine yatan folliküler hücre lenfoması tanısı konmuş 541 hastanın hastane kayıtlarıdır. Hastaneye gelen 541 hasta tanı konulduktan sonra, dönem I ve dönem II hastası olarak sınıflandırılmış, bunlara tedavi olarak sadece radyasyon tedavisi veya radyoterapi+kemoterapi verilmiştir.

Çalışmadaki temel amaç çalışılan hasta grubunda hastanın uzun vadedeki sonuçlarını incelemek ve gözlemlemektir. Bu çalışmadaki hasta bulgularını değerlendirmede dikkat edilen noktalar şunlardır. Tedaviden sonra ki takipte hastada metastas mı gerçekleşti yoksa ölüm durumu mu gerçekleşti bunu sorgulayarak yarışan riskleri değerlendirilmektir. Metastas uzak metastas olabilirken diğer durumda da hem uzak hem de yakın metastas durumu gerçekleşmiş olabilir. Bu veri seti incelenirken çalışmanın amacına uygun olarak paket programlarda uygulama yapılmıştır. R programında yazılan kodlar ile sağ kalım analizlerinde kullanılan Kaplan Meier yaklaşımı ve yarışan riskler yaklaşımının sağ kalım olasılıklarının farklarının ortalama sonuçları hastalığa bağlı ve hastalık dışı durumlara göre ayrılarak gerekli karşılaştırılmalar yapılmıştır. Yapılan analiz sonuçları şu şekildedir.

Gerçek veri setindeki sonuçlar değerlendirilirken hastalığa bağlı ölüm gerçekleştiğinde Kaplan Meier (KM) ve yarışan risklerinin (CR) farklarının ortalama sonuçlarının hastalık dışı ölüm durumundaki KM ve CR farklarının ortalama sonuçlarından farklı olduğu tespit edilmiştir. Diğer taraftan Simülasyon yardımı ile üretilen sonuçlarda verilmiştir. Weibull dağılımı ile üretilen simülasyon sonuçları değerlendirilirken örneklem sayısının değişimine göre analiz sonuçları sunulmuştur. Hastalığa bağlı ölümlerde ve hastalık dışı ölümlerde Kaplan Meier sağ kalım olasılıkları ile yarışan risklerin sağ kalım olasılıkları sonuçları değerlendirildiğinde iki yöntem arasında sağ kalım olasılıklarına göre farklılık olduğu anlaşılmıştır. Bu farklılığı tespit etmek için iki yöntemin sağ kalım olasılıklarının farklarının ortalaması alınarak değerlendirildiğinde yarışan risk değerlerinin daha yüksek sağ kalım değerleri verdiği

gözlemlenmiştir. Örneklem sayısının değişimine göre de değerlendirildiğinde sonuçların örneklem sayısının artış ve azalışına göre bir değişim göstermediği, hastalık dışı durumlarda CR-KM farklarının ortalama değer sonucu yaklaşık 0.075 iken hastalığa bağlı durumlarda bu sonucun 0.010'a yakın bir değer olduğu bulunmuştur.

Üstel dağılımdan elde edilen simülasyon sonuçlarında CR-KM farklarının ortalama sonuçları hastalığa bağlı ölümlerde ve hastalık dışı ölümlerde lamdalar sabit alındığında ($\lambda_1 = 5, \lambda_2 = 5$) örnekleme sayısı azaldıkça sansürleme oranındaki artış CR-KM arasındaki farkı azaltmaktadır. Üstel dağılım parametresi olan lamdaların değişmesi bu sonuçlarda farklılığa sebep olmamıştır. Sansürleme oranı %10 ve %20 alındığında örneklem büyüklüğü arttıkça CR-KM farkının ortalama değerinin yükseldiği görülmektedir. Sansürleme oranının %40 ve %50 olduğu durumda ise örneklem sayısı azaldıkça farkın azaldığı gözlemlenmektedir. Bu değerlendirmelere göre örneklem büyüklüğü küçüldükçe sansürleme etkisi ortaya çıkmaktadır. Lamdalar farklı alındığında da ortaya bu sonuç çıkmaktadır. Küçük örnekleme sansürleme arttıkça büyük örneklem sayısına göre fark azalmakta değerlendirilmesi yapılmaktadır. Bu sonuca göre şu katkıda bulunabiliriz "CR' nin minimum örneklem büyüklüğü ne olmalıdır?" sorusu ortaya çıkmaktadır. Çalışmamızın sonuçlarına göre örneklem büyüklüğünü 500'den büyük almamız sonuçları doğru değerlendirmemizi sağlayacaktır.

Analiz öncesi çalışmadaki ölümler ayırarak yapılarak incelenirse CR aslında düzeltilmiş bir test sonucu olduğundan KM yöntemine göre daha doğru bir sonuç verecektir. Yarışan risk durumunda KM yerine CR tercih etmek daha doğru olacaktır.

Tek bir eş değişken olması durumunda yarışan risklerin varlığında log rank yöntemi ile Gray's test yönteminin karşılaştırılması uygun olabilir.

Birden fazla eş değişken olması durumunda Gray's model ve klasik Cox regresyon model karşılaştırmaları uygun olur. Bu konuda çalışmalarımız devam etmektedir

.Simulasyonlar daha farklı olabilir senaryolar ilave edilerek tekrarlanmalıdır..

7.KAYNAKLAR

- 1 **Ö Gülesen.** Çağdaş Epidemiyoloji. *Uludağ Üniversitesi Yayını*, **1995**.
- 2 **Kazım Özdamar.** Spss Ile Biyoistatistik. Eskişehir: 4.Baskı, *Kaan Kitabevi* ; **2001**: 343-60.
- 3 **Kadir Sümbüloğlu.** Sağlık Alanına Özel İstatistiksel Yöntemler. Ankara: 4.Baskı, *Ankara Tabipler Odası*, **1985**.
- 4 **Regina C.,** Elandt-Johnson. Survival Models and Data Analysis. *John Wiley & Sons*, **1980**; 110
- 5 **David Roxbee Cox** and David Oakes. Analysis of Survival Data, *CRC Press*, **1984**; 21
- 6 **O Aalen .** Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models ,*The Annals of Statistics*, **1978**; 6,
- 7 **John D Kalbfleisch,** and Ross L Prentice. The Statistical Analysis of Failure Time Data, *John Wiley & Sons*, **2011**; 360.
- 8 **Mahesh KB Parmar,** Valter Torri, and Lesley Stewart. Extracting Summary Statistics to Perform Meta-Analyses of the Published Literature for Survival Endpoints, *Statistics in medicine*, **1998**, 17:2815-34.
- 9 **Thomas R Fleming,** and DY Lin. Survival Analysis in Clinical Trials: Past Developments and Future Directions, *Biometrics*, **2000**, 56:971-83.
- 10 **Jason P.Fine.** Dealing with Competing Risks: Testing Covariates and Calculating Sample Size, *Stat Med*, **2002**, 21:3317-24
- 11 **Ted A Gooley.** Wendy Leisenring, John Crowley, and Barry E Storer, 'Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators, *Statistics in medicine*, **1999**, 18:695-706
- 12 **AW Kimball.** Models for the Estimation of Competing Risks from Grouped Data, *Biometrics* **1969**, 10:329-37.
- 13 **Jason P. Fine,** and Robert J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk, *Journal of the American Statistical Association*, **1999**, 94:496-509.

- 14 **Robert J Gray**, A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk, *The Annals of statistics*, **1988**, 1141-54.
- 15 **Mary Lunn**, Don McNeil. Applying Cox Regression to Competing Risks, *Biometrics* **1995**, 524-32.
- 16 **KC Carriere**, Subhash C Kochar. Comparing Sub-Survival Functions in a Competing Risks Model, *Lifetime data analysis*, **2000**, 6:85-97.
- 17 **Ettore Marubini**, Maria Grazia Valsecchi. Analysing Survival Data from Clinical Trials and Observational Studies, *John Wiley & Sons*, **2004**, 15.
- 18 **ML Moeschberger**, John P Klein. Statistical Methods for Dependent Competing Risks, *Lifetime Data Analysis*, **1995**, 1:195-204.
- 19 **Mitchell Gail**. A Review and Critique of Some Models Used in Competing Risk Analysis, *Biometrics*, **1975**, 209-22.
- 20 **J Benichou**, Mitchell H Gail. Estimates of Absolute Cause-Specific Risk in Cohort Studies, *Biometrics*, **1990**, 813-26.
- 21 **Debasis Kundu**, Sankarshan Basu, 'Analysis of Incomplete Data in Presence of Competing Risks', *Journal of Statistical Planning and Inference*, **87 (2000)**, 221-39.
- 22 **B. Lau**, S. R. Cole, and S. J. Gange. Competing Risk Regression Models for Epidemiologic Data, *Am J Epidemiol*, **2009**, 170:244-56.
- 23 **Chin Long Chiang**. The Life Table and Its Applications, **1984**.
- 24 **Jeffrey Gaynor**, Douglass Chapman, Claudia Little, Susan McKenzie, Wendy Miller, Michael Andreeff, Zalmen Arlin, Ellin Berman, Sanford Kempin, and Timothy Gee. A Cause-Specific Hazard Rate Analysis of Prognostic Factors among 199 Adults with Acute Lymphoblastic Leukemia: The Memorial Hospital Experience since 1969, *Journal of Clinical Oncology*, **1988**, 6:1014-30.
- 25 **David G Clayton**. A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, **1978**, 65:141-51.
- 26 **Alan H Schoenfeld**, 'Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics', *Handbook of research on mathematics teaching and learning (1992)*, 334-70.
- 27 **Bryan Lau**, Stephen R Cole, Richard D Moore, and Stephen J Gange. Evaluating Competing Adverse and Beneficial Outcomes Using a Mixture Model, *Statistics in medicine*, **2008**, 27:4313-27.

- 28 **Kenneth R Hess.** Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression, *Statistics in medicine*, **1995**, 14:1707-23.
- 29 **H. Putter,** M. Fiocco, and R. B. Geskus,.Tutorial in Biostatistics: Competing Risks and Multi-State Models, *Stat Med*, **2007**, 26:2389-430.
- 30 **Ralf Bender,** Thomas Augustin and Maria Blettner .Generating survival times to simulate Cox proportional hazards models, *Stat Med.*, **2005**, 24:1713-1723.
- 31 **Rinku Sutradhar,** Shawn Forbes, David R Urbach, Lawrence Paszat, Linda Rabeneck, and Nancy N Baxter. Multistate Models for Comparing Trends in Hospitalizations among Young Adult Survivors of Colorectal Cancer and Matched Controls, *BMC health services research*, **2011**, 12:353.
- 32 **Andrea Burton,** Douglas G Altman, Patrick Royston, and Roger L Holder. The Design of Simulation Studies in Medical Statistics. *Statistics in medicine*, **2006**, 25:4279-92
- 33 **M. Pintilie.** Analysing and Interpreting Competing Risk Data, *Stat Med*, **2007**, 26:1360-7.
- 34 **Michael T Koller,** Heike Raatz, Ewout W Steyerberg, and Marcel Wolbers. Competing Risks and the Clinical Community: Irrelevance or Ignorance, *Statistics in medicine*, **2012**, 31:1089-97.
- 35 **El-Sayed Nour,** and CM Suchindran,. Multi-State Mortality by Cause of Death: A Life Table Analysis, *Journal of the Royal Statistical Society. Series A (General)*, **1984**, 582-97.
- 36 **Martin J Crowder,** and David J Hand. Analysis of Repeated Measures. *CRC Press* , **1990**, 41.
- 37 **Edward C Chao.** Gibbs Sampling for Long-Term Survival Data with Competing Risks, *Biometrics* **1998**, 350-66.
- 38 **SC Cheng,** Jason P Fine, and LJ Wei. Prediction of Cumulative Incidence Function under the Proportional Hazards Model, *Biometrics*, **1998**, 219-28.
- 39 **Kathleen A Cronin,** and Eric J Feuer. Cumulative Cause-Specific Mortality for Cancer Patients in the Presence of Other Causes: A Crude Analogue of Relative Survival, *Statistics in medicine*, **2000**, 19:1729-40.
- 40 **Martin G Larson.** Covariate Analysis of Competing-Risks Data with Log-Linear Models, *Biometrics*, **1984**, 459-69.
- 41 **Margaret Sullivan Pepe,** Motomi Mori, Kaplan—Meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data?, *Statistics in medicine*, **1999**, 12:737-51.

- 42 **Fred Ederer**. The Effect of Adjusting for Competing Mortality Risks, *American Journal of Public Health and the Nations Health*, **1964**, 54:1129-33.
- 43 **Núria Porta**, Guadalupe Gómez, and M Luz Calle. The Role of Survival Functions in Competing Risks, *DR.Universitat Politècnica de Catalunya, Universitat de Vic*, **2008**.
- 44 **Margaret Sullivan Pepe**. Inference for Events with Dependent Risks in Multiple Endpoint Studies, *Journal of the American Statistical Association*, **1991**, 86:770-78.
- 45 **B. Haller**, G. Schmidt, and K. Ulm. Applying Competing Risks Regression Models: An Overview, *Lifetime Data Anal*, **2013**, 19:33-58.
- 46 **Christine Eulenburg**, Sven Mahner, Linn Woelber and Karl Wegscheider. Progression and Death as Competing Risks in Ovarian Cancer, *International Journal of Statistics in Medical Research*, **2013**, 2:249-54.
- 47 **Thomas H Scheike**, Mei-Jie Zhang. Analyzing Competing Risk Data Using the R Timereg Package, *Journal of statistical software*, **2011**, 38.

EKLER

Ek 1. Gerçek veri setinde Kaplan Meier ve Yarışan risklerin değerlendirilmesi için kullanılan R kodu

```
data(follic)
risk<-function(t=0,vT){
  val<-sum((vT>=t),na.rm=T)
  return(val)
}
fail<-function(t=0,vT,vC,c=1){
  val<-sum((vT==t)*(vC==c),na.rm=T)
  return(val)
}
time<-follic[,6]
cens<-follic[,7]
ts<-c(0,unique(sort(time)))
ni<-numeric(length(ts))
d1<-numeric(length(ts))
d2<-numeric(length(ts))
for(i in 1:length(ts)){
  ni[i]<-risk(ts[i],time)
  d1[i]<-fail(ts[i],time,cens,1)
  d2[i]<-fail(ts[i],time,cens,2)}
data.frame(ni,d1,d2,ts)
lam1<-d1/ni
lam2<-d2/ni
plot(ts,lam1)
plot(ts,lam2)
delta<-as.integer(cens!=0)
Surv(time,delta)
cuminc(time,cens,cencode=0)
```

```

cif<-cuminc(time,cens,cencode=0)
plot(cuminc(time,cens,cencode=0))
s1.est<-1-cif[[1]][2]$est
s2.est<-1-cif[[2]][2]$est
cif[[1]]$time
delta1<-as.integer(cens==1)
delta2<-as.integer(cens==2)
s1<-survfit(Surv(time,delta1)~1)$surv
s2<-survfit(Surv(time,delta2)~1)$surv
hesaplama<-function(ssurv,cifzaman,ciftahmin)
{
KM<-cbind(as.numeric(names(table(time))),ssurv)
CR<-cbind(cifzaman,ciftahmin)
CR<-rbind(CR[1,],CR)
CR<-rbind(CR,CR[length(CR[,1]),])
len_CR<-length(CR[,1])
verisayisi<-len_CR/2
aqwe<-matrix(c(0),nrow=verisayisi,ncol=2)
for (i in 1:len_CR)
{
      if (i %% 2 == 0)
      aqwe[i/2,]<-CR[i,]
}

      CR<-aqwe
      KM<-rbind(c(0,1),KM)
      aqw<-cbind(CR,0)
      for (i in 1:length(CR[,1]))
      {
            aqw[i,3]<-KM[which(KM[,1]==CR[i,1]),2]
      }
}

aqw<-data.frame(round(aqw,4))

```

```

aqw<-cbind(aqw,aqw[,2]-aqw[,3],abs(aqw[,2]-aqw[,3]),(aqw[,2]-aqw[,3])^2)
colnames(aqw)<-c("time","CR","KM","CR-KM","abs(CR-KM)","(CR-KM)^2")

return(aqw)
}

```

```

s1<-survfit(Surv(time,delta1)~1)$surv
s2<-survfit(Surv(time,delta2)~1)$surv
cif<-cuminc(time,cens,cencode=0)
sonuc<-hesaplama(s1,cif[[1]]$time,1-cif[[1]][2]$est)

```

```

mean(sonuc[,4])
mean(sonuc[,5])
mean(sonuc[,6])
sonuc1<-hesaplama(s2,cif[[2]]$time,1-cif[[2]][2]$est)
mean(sonuc1[,4])
mean(sonuc1[,5])
mean(sonuc1[,6])
r(mfrow=c(1,2))

```

Ek 2.Kümülatif ölüm olasılıkları grafikleri oluşturma değerlendirilmesi için kullanılan R kodu

```

plot(sonuc[,1],sonuc[,2],ylim=range(c(0,1)),xlim=range(c(0,32)),
type="l",col="black",xlab = "Zaman", ylab = "Sağ Kalım Olasılığı")
lines(sonuc[,1],sonuc[,3],lty=3)
title(main =" Hastalığa Bağlı CR ve KM Sağ Kalım Olasılıkları ")
legend(23, 1, c("KM", "CR"), col =(1,1),lty=c(3,1))
plot(sonuc1[,1],sonuc1[,2],ylim=range(c(0,1)),xlim=range(c(0,32)),
type="l",col="black",xlab = "Zaman", ylab = "Sağ Kalım Olasılığı")

```

```

lines(sonuc1[,1],sonuc1[,3],lty=3)
title(main =" Hastalık Dışı CR ve KM Sağ Kalım Olasılıkları ")
legend(23, 1, c("KM", "CR"), col = c(1, 1),lty=c(3,1)
plot(sonuc1[,1],sonuc1[,4],ylim=range(c(0,1)),xlim=range(c(0,32)),
type="l",col="1",xlab = "Zaman", ylab = "CR-KM farkı")
lines(sonuc[,1],sonuc[,4],col="2")
legend(18, 1, c("Hastalık Dışı", "Hastalığa Bağlı"), col = c(1, 2),lty=1)

```

Ek 3.Weibull Dağılımdan Veri üretme ve Karşılaştırma Kodları değerlendirilmesi için kullanılan R kodu

```

hesaplama<-function(time,ssurv,cifzaman,ciftahmin)
{
  KM<-cbind(as.numeric(names(table(time))),ssurv)
  CR<-cbind(cifzaman,ciftahmin)
  CR<-rbind(CR[1,],CR)
  CR<-rbind(CR,CR[length(CR[,1]),])
  len_CR<-length(CR[,1])
  verisayisi<-len_CR/2
  aqwe<-matrix(c(0),nrow=verisayisi,ncol=2)
  for (i in 1:len_CR)
  {
    if (i %% 2 == 0)
      aqwe[i/2,]<-CR[i,]
  }
  CR<-aqwe
  if (CR[1,1] == 0 && CR[2,1] == 0) CR<-CR[-1,]
  if (KM[1,1] != 0) KM<-rbind(c(0,1),KM)
  aqw<-cbind(CR,0)
  for (i in 1:length(CR[,1]))
  {

```



```

    aqw[i,3]<-KM[which(KM[,1]==CR[i,1]),2]
  }
  aqw<-data.frame(round(aqw,4))
  aqw<-cbind(aqw,aqw[,2]-aqw[,3],abs(aqw[,2]-aqw[,3]),(aqw[,2]-
aqw[,3])^2)
  colnames(aqw)<-c("time","CR","KM","CR-KM","abs(CR-KM)","(CR-
KM)^2")
  return(aqw)
}
weibuluret<-function(nn,p1)
{
  n<-0
  n<-nn
  cum.haz.01 <- function(t, y)
  {
    0.09 * log(t + 1)+ 0.024 * t * t / 2 - y
  }

  cum.haz.02 <- function(t, y) {
    0.825 * 0.09 * log(t + 1) + 0.2 * 0.024 * t * t / 2 - y
  }
  alpha.0.01 <- function(t)
  {
    0.09 / (t + 1)
  }
  alpha.0.02 <- function(t)
  {
    0.024 * t
  }
  alpha.1.01 <- function(t)
  {
    0.825 * alpha.0.01(t)
  }

```

```

}
  alpha.1.02 <- function(t)
{
  0.2*alpha.0.02(t)
}
create.times <- function(n, ch, sup.int = 200)
{
  times <- numeric(n)
  i <- 1
  while (i <= n)
  {
    u <- runif(1)
    if (ch(0, -log(u)) * ch(sup.int, -log(u)) < 0)
    {
      times[i] <- uniroot(ch, c(0, sup.int),
                        tol = 0.0001, y= -log(u))$root
      i <- i + 1
    }
    else
    {
      cat("pos")
    }
  }
  times
}

```

Ek 4.Olaya Karar Vermede Binomial Denemeler için kullanılan R kodu

```

binom.status <- function(ftime, n, a01, a02, size = 1)
{
  prob <- a01(ftime) / (a01(ftime) + a02(ftime))
  out <- rbinom(n, size, prob)
}

```

```

        out
    }
gen.data <- function(n, PROB)
{
    n1 <- n * PROB
    n0 <- n - n1
    t0 <- create.times(n0, cum.haz.01)
    t1 <- create.times(n1, cum.haz.02)
    fs0 <- binom.status(t0, n0, alpha.0.01, alpha.0.02)
    fs1 <- binom.status(t1, n1, alpha.1.01, alpha.1.02)
    cov <- c(rep(0, n0), rep(1, n1))
    ft <- c(t0, t1)
    fs <- c(fs0, fs1)
    fs <- (-1) * fs + 2
    matrix(c(ft, fs, cov), ncol=3)
}

y <- gen.data(nn, PROB = p1)
x <- data.frame(id = 1:length(y[, 1]), T = y[, 1], X.T = y[, 2], Z = y[, 3])

cens <- runif(length(x$T), max = 100)
x$C <- cens
x$time <- pmin(x$T, x$C)
x$status <- as.numeric(x$T <= x$C) * x$X.T
return(x)
}
library(splines)
library(survival)
library(cmprsk)
library(mstate)

nn<-SAYI
sonucMean<-matrix(c(0),nrow=nn,ncol=6)

```

```

sonucMedian<-matrix(c(0),nrow=nn,ncol=6)
sonucMax<-matrix(c(0),nrow=nn,ncol=6)
for(i in 1:nn){
veri<-0
veri<-weibuluret(SAYI,SAYI)
cens<-0
time<-0
time<-veri[,6]
time=round(time,4)
cens<-veri[,7]
delta1<-0
delta2<-0
cif<-0
s1.est<-0
s2.est<-0
sonuc<-0
sonuc1<-0

cif<-cuminc(time,cens,cencode=0)
s1.est<-1-cif[[1]][2]$est
s2.est<-1-cif[[2]][2]$est
delta1<-as.integer(cens==1)
delta2<-as.integer(cens==2)
s1<-survfit(Surv(time,delta1)~1)$surv
s2<-survfit(Surv(time,delta2)~1)$surv
sonuc<-hesaplama(time,s1,cif[[1]]$time,1-cif[[1]][2]$est)
sonuc1<-hesaplama(time,s2,cif[[2]]$time,1-cif[[2]][2]$est)
sonucMean[i,]<-
c(mean(sonuc[,4]),mean(sonuc[,5]),mean(sonuc[,6]),mean(sonuc1[,4]),mean(sonuc1[,5]
),mean(sonuc1[,6]))

```

```

sonucMedian[i,]<-
c(median(sonuc[,4]),median(sonuc[,5]),median(sonuc[,6]),median(sonuc1[,4]),median(sonuc1[,5]),median(sonuc1[,6]))
sonucMax[i,]<-
c(max(sonuc[,4]),max(sonuc[,5]),max(sonuc[,6]),max(sonuc1[,4]),max(sonuc1[,5]),max(sonuc1[,6]))
}
cuminc(time,cens,cencode=0)

```

Ek 5.Üstel Dağılımdan Veri Üretme ve Karşılaştırma değerlendirilmesi için kullanılan R kodu

```

survuretExp<-function(n,ncens,lam1,lam2){
status1<-0
status2<-0
t1<-0
t2<-0
time<-0
censtime<-0
status1<-rep(1,n)
status2<-rep(2,n)
status<-0
t1<- rexp(n, lam1)
t2<- rexp(n, lam2)
for(i in 1:n){
time[i]<-min(t1[i],t2[i])
if(t1[i]<=t2[i]) status[i]<-1
if(t1[i]>t2[i]) status[i]<-2
}
censtime<-runif(n,min(t1,t2),max(t1,t2))
sec<-0

```

```

sec2<-0
sec<-1:n
for(i in 1:ncens){
sec2[i]<-sample(sec,1)
time[sec2[i]]<-censtime[sec2[i]]
status[sec2[i]]<-0}
veri<-data.frame(t1,status1,t2,status2,censtime,time,status)
return(veri)}
nn<-SAYI
sonucMean<-matrix(c(0),nrow=nn,ncol=6)
sonucMedian<-matrix(c(0),nrow=nn,ncol=6)
sonucMax<-matrix(c(0),nrow=nn,ncol=6)
for(i in 1:nn){
veri<-0
veri<-survuretExp(100,10,5,6)
cens<-0
time<-0
time<-veri[,6]
time=round(time,4)
cens<-veri[,7]
delta1<-0
delta2<-0
cif<-0
s1.est<-0
s2.est<-0
sonuc<-0
sonuc1<-0
cif<-cuminc(time,cens,cencode=0)
s1.est<-1-cif[[1]][2]$est
s2.est<-1-cif[[2]][2]$est
delta1<-as.integer(cens==1)
delta2<-as.integer(cens==2)

```

```

s1<-survfit(Surv(time,delta1)~1)$surv
s2<-survfit(Surv(time,delta2)~1)$surv
sonuc<-hesaplama(time,s1,cif[[1]]$time,1-cif[[1]][2]$est)
sonuc1<-hesaplama(time,s2,cif[[2]]$time,1-cif[[2]][2]$est)
sonucMean[i,]<-
c(mean(sonuc[,4]),mean(sonuc[,5]),mean(sonuc[,6]),mean(sonuc1[,4]),mean(sonuc1[,5]
),mean(sonuc1[,6]))
sonucMedian[i,]<-
c(median(sonuc[,4]),median(sonuc[,5]),median(sonuc[,6]),median(sonuc1[,4]),median(s
onuc1[,5]),median(sonuc1[,6]))
sonucMax[i,]<-
c(max(sonuc[,4]),max(sonuc[,5]),max(sonuc[,6]),max(sonuc1[,4]),max(sonuc1[,5]),ma
x(sonuc1[,6]))
}
cuminc(time,cens,cencode=0)
plot(cuminc(time,cens,cencode=0))

```

Ek 6.Gerçek Veri Setinde Eş değişkenlerin Değerlendirilmesi için kullanılan R kodu

```

follic<-read.table("d:/rveriler/follicorj.txt",header=T)
for (i in 1:541){
if(follic$relsite[i]=="NA") follic$relsite[i]=="
}

evcens=(follic$resp=="NR" | follic$relsite!="XA")+0
crcens=(follic$resp=="CR" & follic$relsite=="XA" &follic$stat==1)+0
a=(follic$age>65)+0
fit=survdiff(Surv(follic$dftime,evcens)~a)
fit$chisq

```

```

lrpvalue=1-pchisq(fit$chisq,1)
lrpvalue
cens=evcens+2*rcens
fit=cuminc(follic$df$time,cens,a,cencode=0)
forplev=list(list(fit$'0 1'$time,fit$'0 1'$est),list(fit$'1 1'$time,fit$'1 1'$est))
plot.cuminc(forplev,curvlab=c('Yaş<=65','Yaş>65'),lty=c(1,2),
xlab='Hastalığa Bağlı Ölüm Zamanı (yıl)',ylab='Hastalığa Bağlı Ölüm Olasılıkları')
title(main='Hastalığa Bağlı Ölümlerde CIF')
text(0,0.75,adj=0,paste("Log Rank Test p=",round(lrpvalue,3)))
text(0,0.8,adj=0,paste("Gray's Test p=", round(fit$Tests[1,2],3)))
forplcr=list(list(fit$'0 2'$time,fit$'0 2'$est),list(fit$'1 2'$est))
round(lrpvalue,3)))
plot.cuminc(forplev,curvlab=c('Yaş<=65','Yaş>65'),lty=c(1,2),
xlab='Hastalığa Dış Ölüm Zamanı (yıl)',ylab='Yarışan Risklerin Olasılıkları')
title(main='Yarışan Riskler için CIF')
text(0,0.8,adj=0,paste("Gray's Test p=", round(fit$Tests[1,2],3)))

```


ÖZGEÇMİŞ

1986 yılında Yozgat-Boğazlıyan'da doğdu. İlköğretimine Kaymakam Kemal Bey ilkokulunda başladı. Ortaokul ve lise eğitimini Boğazlıyan Anadolu Lisesinde tamamladı. Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünden mezun oldu. Yüksek lisans eğitimini Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim dalında “R programı kullanılarak bağımlı ve bağımsız gerçek ve yaratılmış verilerde yarışan risklerin değerlendirilmesi” yaptığı çalışmayla sürdürmektedir.