TÜRKİYE CUMHURİYETİ
ANKARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

SAĞKALIM VERİLERİ ile BOYLAMSAL VERİLERİN BİRLEŞİK MODELLENMESİ ve SAĞLIK ALANINA BİR UYGULAMA

Ebru TURGAL

BİYOİSTATİSTİK ANABİLİM DALI
YÜKSEK LİSANS TEZİ

DANIŞMAN
Yrd. Doç Dr. Beyza DOĞANAY ERDOĞAN

ANKARA
2016

Ankara Üniversitesi
Sağlık Bilimleri Enstitüsü Müdürlüğü'ne,

Yüksek Lisans tezi olarak hazırlayıp sunduğum "Sağkalım Verileri ile Boylamsal Verilerin Birleşik Modellenmesi ve Sağlık Alanına bir Uygulama" başlıklı tez; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan çalışma tarafımdan yapılmış olup, tüm cümleler, yorumlar bana aittir. Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Öğrencinin Adı Soyadı: Ebru TURGAL
Tarih: 21.12.2016
İmza:

KABUL ve ONAY SAYFASI

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalında Ebru TURGAL tarafından hazırlanan "Sağkalım Verileri ile Boylamsal Verilerin Birleşik Modellenmesi ve Sağlık Alanına Bir Uygulama" adlı tez çalışması aşağıdaki jüri tarafından YÜKSEK LİSANS TEZİ olarak OY BİRLİĞİ ile kabul edilmiştir.

Tez Savunma Tarihi: 21.12.2016

Prof. Dr. Yasemin YAVUZ
Ankara Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Jüri Başkanı

Yrd. Doç. Dr. Beyza DOĞANAY ERDOĞAN
Ankara Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Tez Danışmanı

Yrd. Doç. Dr. N. Anıl DOLGUN
Hacettepe Üniversitesi
Tıp Fakültesi
Biyoistatistik Anabilim Dalı

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

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

İÇİNDEKİLER

Etik Beyan **ii**
Kabul ve Onay **iii**
İçindekiler **iv**
Önsöz **vi**
Simgeler ve Kısaltmalar **vii**
Şekiller **vii**
Çizelgeler **ix**

1.GİRİŞ **1**
1.1. Araştırmanın Konusu, Amacı ve Önemi 1
1.2.Birleşik Modellerin Avantajları 3
1.3.Birleşik Modellemenin Tarihçesi 3
1.4.Boylamsal Verilerin Analizi 5
1.4.1.Doğrusal Karışık Etkili Model 6
1.4.1.1.Rastgele Kesim Noktası Modeli 8
1.4.1.2. Rastgele Kesim Noktası ve Rastgele Eğim Modeli 10
1.4.1.3. Rastgele Etkilerin Kestirimi 12
1.5.Sağkalım Verilerinin Modellenmesi 13
1.5.1.Cox Oransal Hazard Regresyon Modeli 13
1.6.Zamana Bağlı Açıklayıcı Değişkenlerin Sağkalım Sürecinde Ele Alınması 14
1.7.Birleşik Model 16
1.7.1.Alt Modeller 18
1.7.1.1. Boylamsal Alt Model 18
1.7.1.2. Sağkalım Alt Modeli 20
1.7.2.Temel Risk Fonksiyonu Çeşitleri 21
1.7.2.1.Weibull Dağılımı 21
1.7.2.2.Parçalı Sabit Hazard Fonksiyonu 22
1.8.Kestirim Yöntemleri 23
1.8.1.İki Aşamalı Yaklaşım 23
1.8.2.Birleşik Olabilirlik Çözümü 24
1.8.3.İntegrasyon Yöntemi 28
1.9.Model Seçim Kriterleri ve Model Uyumu 31
1.9.1.Artık Analizi 32
1.9.1.1.Sağkalım Sürecine İlişkin Artıklar 32
1.9.1.2.Boylamsal Sürece İlişkin Artıklar 35

2.GEREÇ VE YÖNTEM **37**
2.1.Uygulama Verileri 37
2.2.Uygulama Yöntemleri 38
2.2.1.Açıklayıcı Değişkenlerin Düzenlenmesi 38
2.2.2.Birleşik Modeller için kullanılan R Fonksiyonları 40

3.BULGULAR **43**
3.1.Birleşik Modellere Ait Sonuçlar ve Model Seçimi 46
3.1.1.Model 1 46
3.1.1.1.Boylamsal Kısım için Artıklar 47
3.1.1.2.Sağkalım Kısmı için Artıklar 49
3.1.2.Model 2 51
3.1.2.1.Boylamsal Kısım için Artıklar 52
3.1.2.2.Sağkalım Kısmı için Artıklar 54
3.1.3.Model 3 56
3.1.3.1.Boylamsal Kısım için Artıklar 57

4.TARTIŞMA **61**
5.SONUÇ VE ÖNERİLER **63**
ÖZET **64**
SUMMARY **65**
KAYNAKLAR **66**
EKLER **72**
Ek-1. EM Algoritması 72
Ek-2. Rasgele Etkilerin Kestirim Değerler 76
Ek-3. Tez çalışmasında kullanılan veri seti ile ilgili izin yazısı 79
ÖZGEÇMİŞ **80**

ÖNSÖZ

Sağlık alanında yapılan sağkalım çalışmaları, genellikle boylamsal verileri de içerir. Boylamsal verilerin en iyi biçimde çözümlenmesi ve yorumlanması ile ilgili çalışmalar geçtiğimiz yıllarda büyük bir ilerleme göstermiş ve bu anlamda güçlü birçok yöntem geliştirilmiştir. Bunun yanı sıra, sağkalım verilerini inceleyen yöntemler üzerine de çok sayıda çalışma bulunmaktadır. Ancak bu aynı çalışmada aynı bireyler üzerinde toplanan bu iki veri tipinin birbiri ile ilişkisini inceleyen yöntemler son yıllarda önem kazanmıştır. Bu çalışmada yoğun bakım ünitesinde akut hiperkapnik solunum yetmezliği nedeniyle takipte olan 88 hastaya ait veriler kullanılarak birleşik modelleme tekniği ile sağkalım süreci ile boylamsal süreç arasındaki ilişki tahmin edilmeye çalışılmıştır.

Tezimin oluşmasında bana en büyük desteği veren, çalışmalarımda bana yol gösteren ve teşvik eden değerli danışman hocam Yrd. Doç. Dr. Beyza Doğanay Erdoğan'a, ilgisinden dolayı A.Ü. Biyoistatistik Bölümü Başkanı değerli hocam sayın Prof. Dr. Yasemin Yavuz Genç'e ve bu olanağı sağlayan saygıdeğer hocam Hitit Üniversitesi Tıp Fakültesi Dekanı Prof. Dr. Mete Dolapçı'ya,

Bilgisini ve tecrübesini en iyi biçimde benimle paylaşan hocam Sayın Doç. Dr. Kenan Köse'ye, hiçbir zaman yardımlarını esirgemeyen değerli hocam Prof. Dr. Atilla Halil Elhan'a, ilgi ve anlayışlarından dolayı tüm Biyoistatistik Anabilim Dalı çalışma arkadaşlarıma,

Tezim için uygulama verilerini sağlayan Ankara Üniversitesi Tıp Fakültesi Göğüs Hastalıkları Kliniği'nden Uzm. Dr. Fatma Çiftçi'ye,

Her zaman beni destekleyen, gösterdikleri sabır ve sevgiden dolayı AİLEM'e, tez çalışmamın her anında bana destek olan Esat'a içtenlikle teşekkür ederim.

SİMGELER ve KISALTMALAR

AIC	Akaike Bilgi Kriteri
AFT	Hızlandırılmış Başarısızlık Süresi
EM	Beklenti Maksimizasyonu
BIC	Bayesci Bilgi Kriteri
θ	İncelenen Özellik

ŞEKİLLER

Şekil 1.1. Rastgele kesim noktası modelinde herhangi iki bireye ve popülasyona ait zaman içindeki yanıt eğilimleri 9
Şekil 1.2. Ölçüm hataları eklendikten sonra zaman içindeki marjinal ve koşullu yanıtlara ait profiller 11
Şekil.1.3. β_1 boylamsal sürece ve β_2 sağkalım sürecine ilişkin model parametreleri gösterimi 17
Şekil 2.1. Bireylerin Zaman İçindeki Boylamsal Trendleri 39
Şekil 3.1. Rastgele kesim noktası modeli 44
Şekil 3.2. Rastgele kesim noktası ve rastgele eğim modeli 45
Şekil 3.3. Birleşik Model için Tanısal Grafikler 47
Şekil **3.4.** Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler 48
Şekil 3.5. Martingale artıklarının lowess yöntemi ile oluşturulan eğri grafiği 49
Şekil 3.6. Veri setinin Cox-Snell artıkları 50
Şekil 3.7. Birleşik Model için Tanısal Grafikler 52
Şekil **3.8.** Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler 53
Şekil 3.9. Martingale artıklarının lowess yöntemi ile oluşturulan eğri grafiği 54
Şekil 3.10.Veri setinin Cox-Snell artıkları 55
Şekil 3.11. Birleşik Model için Tanısal Grafikler 57
Şekil **3.12.** Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler 58

ÇİZELGELER

Çizelge 1.1. Model Seçim Kriterleri 32
Çizelge 2.1. Değerlendirilecek açıklayıcı değişkenler 39
Çizelge 2.2. R programında birleşik fonksiyon girişi -Model 1 41
Çizelge 2.3. R programında birleşik fonksiyon girişi -Model 2 42
Çizelge 2.4. R programında birleşik fonksiyon girişi -Model 3 42
Çizelge 3.1. Tanımlayıcı istatistikler 43
Çizelge 3.2 PaO_2 / FiO_2 değişkeni için oluşturulan birleşik modellerin AIC ve BIC değerleri 44
Çizelge 3.3. Birleşik model sonuçları (1.model) 46
Çizelge 3.4. Birleşik model sonuçları (2.model) 51
Çizelge 3.5. Birleşik model sonuçları (3.model) 56
Çizelge 3.6. Ayrı Analiz, İki-Aşama Modeli ve Birleşik Model Sonuçları 60
Çizelge 4.1 Rastgele Etkilerin Kestirim Değerleri 76

1. GİRİŞ

1.1.Araştırmanın Konusu, Amacı ve Önemi

Araştırmalarda veri toplama işlemi tek bir zaman diliminde kesitsel olarak yapılabileceği gibi, verilerin farklı zaman dilimlerinde tekrarlı olarak toplanması biçiminde de yapılabilmektedir. Değişkenleri oluşturan gözlemlerin ya da ölçümlerin zamana bağlı olarak birçok kez toplanması ile elde edilmiş olan veriler "boylamsal veriler" olarak isimlendirilmektedir. Boylamsal verilerin analizi, bireylerden toplanan verilerin, hem aynı birey içindeki, hem de bireyler arasındaki değişimleri incelenerek gerçekleştirilmektedir. Böylelikle, bireyler arasındaki doğal biyolojik farklılıklar da göz önüne alınarak, toplanan verilerin zaman içindeki değişimleri gösterilmiş olur.

Sağkalım verileri ise belli bir izlem süresi boyunca ilgilenilen olayı (ölüm, hastalığın nüks etmesi, organ nakli olması, vb.) ve görülme zamanını kapsar. Geleneksel olarak aynı çalışma kapsamında toplanan bu iki tip veri farklı istatistiksel yöntemler kullanılarak ayrı ayrı analiz edilir. Ancak kişilere ilişkin sağkalım süreci ile bu süre içinde boylamsal olarak toplanan verilerin ilişki olduğu durumlarda, yansız ve daha etkin sonuçlar elde etmek için birleşik bir model oluşturulmalıdır.

Son zamanlarda, boylamsal süreç ile sağkalım sürecinin ilişkili olduğu durumlarda bu iki sürecin birlikte modellenmesi üzerine yapılan çalışmalar gittikçe artmaktadır. Toplanan bu tip verilerden doğru sonuçlar elde edebilmek için birleşik modelleme yönteminin kullanılması gerekliliği yapılan çalışmalar sonucu ortaya konmuştur.

Birleşik modelleme, boylamsal ve sağkalım verilerini birbirine bağlayan bir modellemedir ve bağlı iki alt modelden oluşur. Bunlardan biri boylamsal süreç için ölçüm modeline, diğeri ise, sağkalım süreci için yoğunluk modeline karşılık gelir.

Genelde, birleşik modelin sağkalım kısmı için Cox oransal hazard modeli, aralıklı sansürlü veri analizi ya da hızlandırılmış hata zamanı (accelerated failure time, AFT) tercih edilir. Boylamsal kısım için, verinin tipine (sürekli ya da kesikli) ve bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkinin doğrusal olup olmamasına bağlı olarak, doğrusal karışık etkili modeller (lineer mixed effects models, LMEM), genelleştirilmiş doğrusal karışık etkili modeller (generalized linear mixed models, GLMM), veya doğrusal olmayan karışık etkili modeller (nonlinear mixed effects models, NLMM) tercih edilir. Bu yöntemlerin birleşiminden birleşik modelleme (joint modeling, JM) ortaya çıkar. Bu tez kapsamında, sağkalım kısmı için Cox oransal hazard modeli ile, boylamsal kısım için sürekli veri tipinde kullanılabilecek doğrusal karışık etkili modeller ele alınacaktır.

Birleşik modelleme son zamanlarda oldukça önem kazanmakta ve özellikle sağlık alanında sıklıkla karşılaşılan sağkalım çalışmalarındaki sağkalım süreci ile bu süreçte toplanan boylamsal verilerin ilişkilendirilmesi amacıyla kullanılmaktadır. Genel olarak birleşik modelleme yönteminin kullanılabileceği durumlar aşağıda verilmiştir.

- Boylamsal verilerde göz ardı edilemeyen kayıp veriler ortaya çıktığında (kayıp veri yapısı ile gözlenen verilerin eş zamanlı modellenmesinde)
- Sağkalım ve boylamsal sürecin arasındaki ilişkinin gücünü ortaya koymak için
- Sağkalım sürecinin zamana bağlı açıklayıcı değişkenleri ölçüm hatası içerdiği durumlarda kullanılır.

Bu tez çalışmasının amacı, birleşik modellemenin teorik özelliklerini tanıtmak, sağlık araştırmalarındaki yerini ve avantajlarını açıklamak, ayrıca uygun bir araştırma verisi üzerinde yöntemin uygulamasını gerçekleştirerek sonuçları yorumlamaktır.

Ayrıca, bu tez çalışmasının hipotezi, boylamsal süreç için kullanılan karışık etkili model ile sağkalım süreci için kullanılan Cox regresyon modelini rasgele etkiler bakımından birbirine bağlayıp bu süreçler arasındaki ilişkiyi ortaya koymaktır.

Tez çalışmasının bu bölümünde ilk olarak, birleşik modellemenin sağladığı avantajlara, tarihsel gelişimine, sonraki aşamada, boylamsal veriler ile sağkalım verilerinin modellenmesinde birleşik modeller kapsamında kullanılabilecek yöntemlere değinilecektir.

1.2. Birleşik Modellerin Avantajları

Sağkalım ve boylamsal veri için sonuçları ayrı ayrı analiz etmek yerine, iki durumu birlikte ele aldığı için daha fazla açıklayıcı bilgi içermektedir. Sağkalım ve boylamsal belirteç üzerinde tedavi etkisinin daha etkili tahminleri elde edilir.

Sağkalım süreçlerinde ve tekrarlı ölçümlerde birleşik modellerin birçok avantajı mevcuttur;
1- tekrarlı ölçümler; belirli olay zamanında gözlenen ölçümlerden tahmin edilebilir;
2- sağkalım, tekrarlı ölçümlerin gerçek ama bilinmeyen değerine bağlıdır, dolayısıyla ölçüm hatası için düzeltme yapılabilir ve Cox model parametre kestirimlerinin yanlılığına bu sayede çözüm bulunabilir.
3- ölüm ya da bireyin süreçten ayrılması ile yaşanan bilgi kaybı düzeltilir (Nugity ve ark, 2005).

1.3. Birleşik Modellemenin Tarihçesi

Sağlık araştırmalarında bireylerden toplanan boylamsal veriler ile zamana bağlı sağkalım sürecinin birbiri ile genellikle ilişkili olduğu bilinmektedir. Bu ilişkiyi ele alabilmek için birleşik modelleme tekniği ilk olarak Schluchter (1992) tarafından önerilmiştir.

Sağlık alanındaki literatürde birleşik modeller ile ilgili görülen ilk uygulamalar ise son yirmi yıldır gözlenmektedir. Birçok klinik ve epidemiyolojik çalışmada

periyodik olarak ölçülen hastalık belirteçleri hastalığın başlangıcından gelişimine dek izlenir. Wang ve Taylor (2001), belirteç ile hastalık/ölüm üzerindeki tedavi ve risk faktörlerinin etkisini araştırmak amacıyla birleşik modellemeyi AIDS hastalarından toplanan verilere uygulamışlardır. AIDS için HIV virüsü taşıyan erkeklerde CD4 hücre sayılarını inceleyen bu çalışmada, Wang ve Taylor boylamsal ve sağkalım verisi için birleşik modelleme tekniğini kullanmışlardır.

Henderson ve ark. (2000) boylamsal veri ve sağkalım sürecinde tek ya da yinelenen olay verisi için iki süreç arasındaki birlikteliğin rastgele etkilere bağlı olduğu birleşik modelleme yöntemini kullanmışlardır.

Martins ve ark. (2016), 2002-2006 yılları arasında yaptıkları Brezilya'da HIV/AIDS hastalarından oluşan bir kohort çalışmasında, birleşik modelleme ile diğer sağkalım modelleri karşılaştırıldığında, birleşik modelleme ile elde edilen sağkalım süresi kestirimlerinin iyileştiğini göstermişlerdir. Bunun sebebini çalışmaya alınan bireyler arası gözlenemeyen biyolojik heterojenliğe dayandırmışlardır. Bu durum da, birleşik modellemede bireye özel rastgele etkilerin modele dahil edilmesi ile göz önüne alınmaktadır.

Akut böbrek yetmezliği, ölüm ve hastalık oranı ile oldukça ilişkilidir. Khoundabi ve ark. (2015) kalp ameliyatı geçirmemiş 400 hastada, akut böbrek yetmezliğinin (ABY) hazard oranını belirleyip, ABY'ye etki eden risk faktörlerini araştırmışlardır. İdrar sonuçlarını ABY'nin bir tahmin edicisi olarak kullanıp birleşik modelleme tekniğini uygulanmıştır. Sonuç olarak boylamsal ve sağkalım modellerine göre ABY ve idrar sonuçlarının birleşik modelleme ile kestiriminde kayda değer gelişmeler olduğunu göstermişlerdir.

AIDS Klinik Denemeler grubundaki 175 katılımcıya antiretroviral terapi denemesi sonuçları ile tedavi etkisini araştırmak amacıyla Fu ve Gilbert (2016) birleşik modellemeyi kullanmışlardır.

Kavrama gücü risk faktörünün girişimsel olmayan bir yöntemidir. Kavrama gücündeki değişim ve ölüm arasındaki ilişki bilinmemektedir. Peterson ve ark. bu çalışmada yaşlı Meksikalı Amerikalılarda kavrama gücü ile ölüm arasındaki ilişkiyi incelemişlerdir. Çalışmaya 3052 kişi alınmış ve koşullu çıkarsama ağaç analizi (Conditional inference tree analyses) uygulanarak zayıflık eşiğinin cinsiyet ve yaşa göre değişimini incelemişler ve Kaplan-Meier tahmin edicisi işe sağkalım tahminlerini elde etmişlerdir. Sağkalım için Cox oransal hazard regresyon ile birleşik modellemeyi karşılaştırmışlardır. Kavrama gücündeki azalma ile erken ölüm arasında güçlü ve boylamsal bir ilişki bulmuşlardır. Kavrama gücündeki %10'luk bir azalmanın ölüm riskini erkeklerde 1,12; kadınlarda 1,15 kat arttırdığı sonucuna varmışlardır (Peterson ve ark, 2016).

Piulachs ve Alemany (2016) 2006-2014 yılları arasında hastaneye yatırılma, acil servis bakımı ve ambulans servisi kullanan 65 yaş ve üstü 39137 kişinin analizini birleşik modelleme ile göstermişlerdir ve ayrı ayrı modellemeye göre daha etkili ve az yanlı sonuçlar elde etmişlerdir.

Görüldüğü üzere birleşik modelleme tekniği, sağlık alanındaki literatürde hem yöntemsel hem de uygulamalı anlamda gelişmesini hızla sürdürmektedir.

Bu bölümde, birleşik modelleme yöntemine geçmeden önce, boylamsal süreç ve sağkalım sürecinin ayrı ayrı modellenmesinde kullanılan modellerden bahsedilecektir.

1.4. Boylamsal Verilerin Analizi

Boylamsal verileri içeren çalışmaların amacı genellikle, ortalama yanıt profillerinin gruplar arasında nasıl farklılık gösterdiğinin ve ayrıca yanıtların zaman içindeki

seyirlerinin incelenmesidir. Bu soruların yanıtları çeşitli istatistiksel modeller yardımı ile bulunabilir.

Boylamsal verilerin analizi, son yıllarda yapılan çalışmalar oldukça artmış ve hız kazanmıştır. "Karışık etkili modeller boylamsal veri analizinde güçlü araçlardır" (Wu ve Zhang, 2006). Karışık etkili modeller ayrıca, rastgele etki modelleri, rastgele katsayılar modelleri, hiyerarşik modeller ve çok düzeyli modeller gibi başka adlarla da anılır (Armitage ve Colton, 1998).

Karışık etkili modeller, hem sabit hem de rastgele etkileri bir arada bulundurma özelliğine sahiptir. Karışık etkili modellerde sabit etkiler popülasyon etkileri, rastgele etkiler ise bireylerin popülasyon değerlerinden sapmalarını gösteren birey etkileri olarak alınır. Rastgele etkinin aynı düzeyini paylaşan ölçümler ilişkili olarak modellenirler (Vittinghoff ve ark., 2005).

Veri tipi sürekli olduğunda genel olarak aşağıda tanımlanan doğrusal karışık etkili modeller kullanılabilir.

1.4.1. Doğrusal Karışık Etkili Model

Genel olarak, bir doğrusal karışık etkili model aşağıdaki koşulları sağlayan herhangi bir model olabilir.

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i,$$
$$b_i \sim N(0, D),$$
$$\varepsilon_i \sim N\left(0, \sigma^2 I_{n_i}\right),$$
$$\mathrm{Cov}(b_i, \varepsilon_i) = 0 \qquad \text{(Eşitlik 1.1)}$$

Modelde y_i i. birey için n_i boyutlu yanıt vektörü, 1 ≤ i ≤ N, N: birey sayısı, X_i ve Z_i sırasıyla ($n_i \times p$) ve ($n_i \times q$) boyutlu bilinen açıklayıcı değişkenler matrisleri, β sabit etkileri içeren p-boyutlu vektör, b_i rastgele etkileri içeren q-boyutlu vektör, ε_i hata bileşenlerini içeren n_i -boyutlu vektördür. Son olarak, D , $d_{ij} = d_{ji}$ olmak üzere, ($q \times q$) boyutlu genel kovaryans matrisi, ve $\sigma^2 I_{n_1}$ ($n_i \times n_i$) boyutlu kovaryans matrisidir (Verbeke ve Molenbergs, 2000). Rastgele etkilerin sıfır ortalama ve varyans kovaryans matrisi D ile normal dağıldığı varsayılır ve ε_i hata terimlerinin bağımsız olduğu varsayılır.

β sabit etkilerin yorumu, basit doğrusal regresyon modelindeki gibidir. İlgili değişkendeki bir birim artışın y_i ortalamadaki değişimi gösterir. b_i rastgele etkiler i. denek için regresyon parametrelerinin bir alt kümesi olarak popülasyonda diğerlerinden nasıl saptığı şeklinde yorumlanır.

Karışık etkili modellerin avantajı, hem popülasyonda değişkenin ortalama yanıtının nasıl değiştiğini açıklayan parametreleri hem de zamanla bireysel trendlerin nasıl değiştiğini tahmin eder.

Karışık etkili modellerin boylamsal ve sağkalım süresi verilerinin birleşik modellerinde kullanılmasının temel nedeni budur. Aynı bireyden alınan tekrarlı ölçümlerin birbiri ile ilişkili olması, bireyler arası heterojenlikten kaynaklanmaktadır. Bireyler arası heterojenliğin ölçütü olan rastgele etkiler ile tekrarlı ölçümler arası ilişki yapısı modelde göz önüne alınmış olur. Buna göre, birey bazında, yani rastgele etkiye koşullu olarak, tekrarlı ölçümlerin birbirinden bağımsız olduğu varsayılır. Koşullu bağımsızlık altında i. bireyden alınan tekrarlı ölçümlerin olabilirliği aşağıdaki gibi yazılabilir:

$$p(y_i|b_i;\theta) = \prod_{j=1}^{n_1} p(y_{ij}|b_i;\theta) \qquad \text{(Eşitlik 1.2)}$$

Koşullu bağımsızlık varsayımı, birey bazında ölçüm hatalarının bağımsızlığını ifade eder $\left(\varepsilon_i \sim N\left(0, \sigma^2 I_{n_i}\right)\right)$. Ancak hataların da ilişkili olduğu durumlara genellenebilecek, bireye özel hata bileşeni için bir kovaryans matrisi tanımlanabilir ($\varepsilon_i \sim \mathrm{N}\left(0, \sum_i\right)$). Literatürde; $\sum_i$ için pek çok farklı kovaryans yapısı vardır; ve farklı tip seri korelasyon fonksiyonuna dayanır. En sık kullanılanları; birinci derece otoregresif, üstel ve Gausçu korelasyon yapılarıdır ancak standart istatistiksel yazılımlar doğrusal karışık etkili modeli uydurmak için daha fazla seçenek sağlar (Pinheiro ve Bates, 2000; Verbeke ve Molenberghs, 2000).

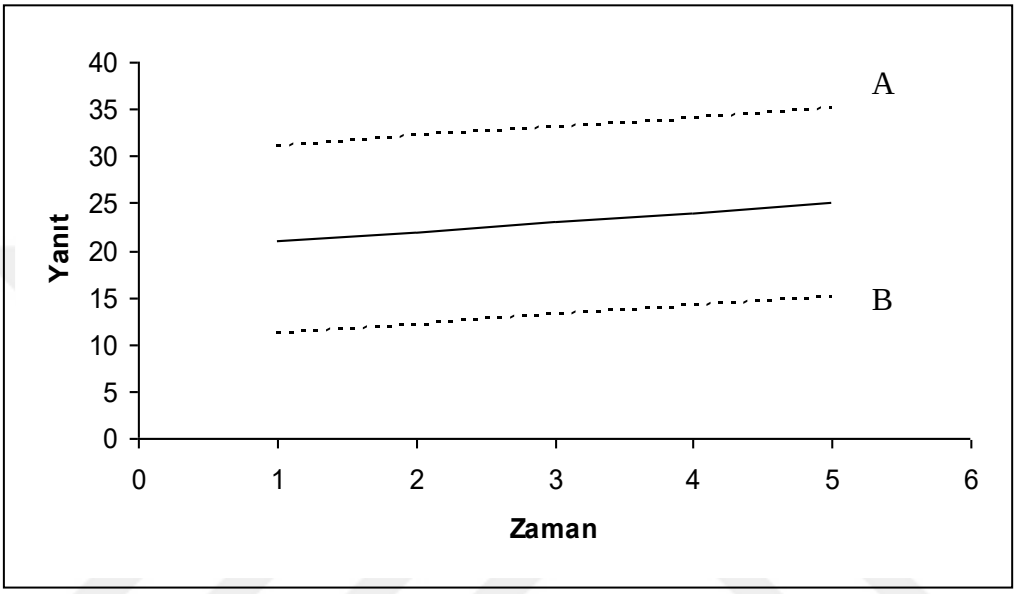
1.4.1.1. Rastgele Kesim Noktası Modeli

Doğrusal karışık etkili modelinin en basit hali, rastgele değişen birey etkisinin bulunduğu bir doğrusal modeldir. Bu modelde, her bireyin kendi yanıt düzeyine sahip olduğu varsayılır. Bireyin etkisi, karışık etkili modele rastgele etki olarak alınır. Böylelikle, bireyler arasındaki heterojenlik kaynağı modele eklenmiş olur (Doğanay, 2007).

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij} \qquad \text{(Eşitlik 1.3)}$$

Burada b_{oi} parametresi i. bireyin kendi tekrarlı gözlemleri üzerindeki etkisini göstermektedir. Yani, b_{oi} Eşitlik 1.3'te rastgele birey etkisi olarak yer almaktadır. Bireylerin kendi ölçümleri üzerinde etkisi olmadığında bütün b_{oi} değerleri sıfıra eşit olacaktır. Ancak, bireylerin kendi tekrarlı ölçümleri üzerinde pozitif ya da negatif etkiye sahip olmaları daha olası bir durumdur ve bu nedenle b_{oi} değerleri sıfırdan farklı olacaktır. Bu modelde i. bireyden j. durum ya da zamanda alınan yanıt popülasyon ortalaması $\beta_0 + \beta_1 t_{ij}$'den, birey etkisi b_{oi} ve birey içi ölçüm hatası ε_{ij} kadar farklılık gösterir. Burada hem birey etkisi hem de birey içi ölçüm etkisi sıfır ortalama ve sırasıyla $Var(b_{0i}) = {\sigma_b}^2$ $Var\left(\varepsilon_{ij}\right) = \sigma^2$ varyanslarına sahip rastgele etkilerdir. Daha önce de belirtildiği gibi, Eşitlik 1.3'te yalnızca tek bir birey etkisi

vardır ve bu birey etkisi, bireyin grubun başlangıç zamanı ortalamasından sapmasını verir. Bireye ait ölçümlerin trendi, bulunduğu gruba ait ortalama trende paralel olacaktır (Şekil 1.1). Bu nedenle bu model rastgele kesim noktaları modeli olarak tanımlanır ve her bir b_{oi}, i. bireyin popülasyon eğiliminden nasıl saptığını gösterir (Doğanay, 2007).



Şekil 1.1. Rastgele kesim noktası modelinde herhangi iki bireye ve popülasyona ait zaman içindeki yanıt eğilimleri

Şekil 1.1'de, düz çizgi ile gösterilen eğri β_0 ve β_1'e bağlı popülasyon eğilimini göstermektedir. A ve B bireylerine ait koşullu yanıtlar, ortalama (popülasyon) eğilimin üzerinde ve altında, ortalama eğilime paraleldir. Seçilen örneklemde N birey için N tane eğri olacaktır. Rastgele birey etkilerine ait varyans terimi, σ_b^2, bu eğrilerin yayılımını göstermektedir. Eğer σ_b^2 sıfıra yakın ise, bireysel eğilimler ortalama eğilimden farklılık göstermeyecektir. Bireylerin popülasyon eğiliminden farklılık gösterdiği durumda birey eğrileri popülasyon eğrisinden uzaklaşacak ve σ_b^2 değeri artacaktır (Doğanay, 2007).

Bu örnekte A'ya ait yanıtların hep popülasyon ortalamasının üzerinde olduğu, yani A bireyinin pozitif etkiye (b_{0A}) sahip olduğu, B'ye ait yanıtların ise hep popülasyon ortalamasının altında olduğu yani B bireyinin negatif bir etkiye (b_{0B}) sahip olduğu görülmektedir. Rastgele ölçüm hataları e_{ij} 'ler de ise, Şekil 1.2'de içi boş noktalar ile gösterilen gözlenen ölçümlerin bireyin kendi regresyon doğrusuna olan uzaklığını göstermektedir (Doğanay, 2007).

1.4.1.2. Rastgele Kesim Noktası ve Rastgele Eğim Modeli

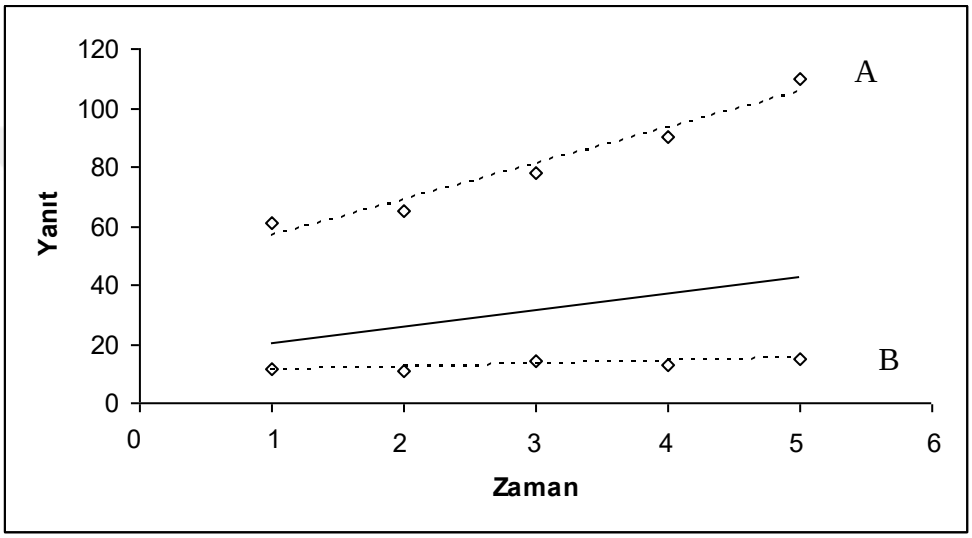
Boylamsal veri için rastgele kesim noktası modeli basit düzeyde kalabilir. Öncelikle, bütün bireyler için zaman içindeki değişimin aynı olduğunu varsaymak yanlış olacaktır. Her bireyin aynı oranda değişim göstermemesi, yani bireylerin zaman içindeki eğilimleri bakımından farklılık göstermeleri daha olası bir durumdur. Bu nedenlerle, hem kesim noktasının hem de zaman içindeki eğilimin bireyler arasında değişim gösterdiğini varsayan bir karışık etkili model daha uygun olabilir (Doğanay, 2007).

Buna göre, j. zaman noktasında i. bireyden alınan yanıt için karışık etkili model aşağıdaki gibi yazılabilir (Fitzmaurice ve ark., 2004).

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{01i} + b_{1i} t_{ij} + e_{ij} \qquad \text{(Eşitlik 1.4)}$$

Bu modelde β_0 popülasyon kesim noktasını, β_1 ise popülasyon eğimini, b_{0i} i. bireyin popülasyon kesim noktasından sapmasını, b_{1i} ise i. bireyin popülasyon eğiminden sapmasını göstermektedir. İlk aşama modelindeki e_{ij}, sıfır ortalama ve σ^2 varyansı ile normal dağılan ve koşullu bağımsız hata terimidir (Doğanay, 2007).

Kesim noktası parametreleri başlangıç yanıt düzeyini, eğim parametreleri zaman içindeki değişimi göstermektedir. Popülasyon kesim noktası ve eğim parametreleri genel eğilimi (popülasyon eğilimini) gösterirken, bireye özel parametreler ise bireylerin popülasyondan nasıl saptıklarını gösterir. Yani, böyle bir modelde bireyler arası değişkenlik yalnızca ilk yanıt düzeylerinden değil, zaman içindeki yanıt profillerinin bireyler arası değişiminden de kaynaklanmaktadır. Şekil 1.2 bu modeli grafiksel olarak tanımlamaktadır (Doğanay, 2007).



Şekil 1.2. Ölçüm hataları eklendikten sonra zaman içindeki marjinal ve koşullu yanıtlara ait profiller

Şekil 1.2'de düz çizgi ile gösterilen eğilim popülasyon eğilimi (marjinal ortalama yanıt), kesikli çizgiler ise A ve B bireylerinin zaman içindeki eğilim eğrilerini (koşullu ortalama yanıt trendleri) göstermektedir. Bu modelde zaman içindeki eğilim de bireyler arasında farklılık gösterdiği için, bazı bireyler zaman içinde çok değişim göstermezken, bazıları çok belirgin değişim gösterebilirler. Popülasyon eğilimi bütün bireyler için ortalama eğilimi gösterirken, varyans terimleri popülasyonda ne kadar heterojenlik olduğunu göstermektedir. Yani, $\sigma^2_{b_0}$ bireylerin popülasyon kesim noktası etrafında ne kadar değişim gösterdiklerini gösterirken, $\sigma^2_{b_1}$ eğimlerdeki değişimi göstermektedir (Doğanay, 2007).

Bu örneğe göre; A ve B bireyleri ayrı ayrı popülasyonla karşılaştırılacak olursa, A bireyine ait başlangıç düzeyinin ($\beta_0 + b_{0A}$), popülasyon ortalamasından (β_0) daha yüksek olduğu, yani pozitif bir bireysel etkiye sahip olduğu görülmektedir. B bireyine ait başlangıç yanıt düzeyinin ($\beta_0 + b_{0B}$), popülasyon ortalamasından (β_0) daha düşük olduğu, yani negatif bir bireysel etkiye sahip olduğu görülmektedir. Ayrıca, A bireyinin popülasyona ait ortalama değişime (β_1) göre daha dik bir değişim gösterdiği ($\beta_1 + b_{1A}$), yani pozitif bir bireysel etkiye sahip olduğu görülmekteyken, B bireyinin daha yatık bir değişim ($\beta_1 + b_{1B}$) göstermekte olduğu, yani negatif bir bireysel etkiye sahip olduğu görülmektedir. Son olarak, ölçüm hataları e_{ij} 'ler de göz önüne alındığında, gözlenen yanıtların bireylere özel eğriler etrafında rastgele olarak değişim gösterdiği görülmektedir. Bu modelde rastgele kesim noktaları ve rastgele eğimler yer almaktadır. Doğrusal karışık etkili modeller rastgele değişen ek regresyon katsayıları için de genellenebilir. Ayrıca rastgele etkilere ait ortalamaların ortak değişkenlere (tedavi grupları, cinsiyet, … gibi) bağlı olarak değişmesini sağlayabilir (Fitzmaurice ve ark., 2004).

1.4.1.3. Rastgele Etkilerin Kestirimi

Rastgele etkiler b_i boylamsal değişimlerin heterojenliğini açıklamak ve boylamsal değişimler ile sağkalım sürecinin arasındaki ilişkiyi oluşturmak için kullanılır. Genellikle, her bir sonuç değişkeni için, bireye özgü tahminler yapılmaktadır. Bu tür tahminler elde etmek, b_i rastgele bir etki vektörü tahminini gerekmektedir. Rastgele etkilerin rastgele değişkenler olduğu varsayımı altında, Bayes paradigması kullanılarak kestirilir (Rizopoulos, 2012a).

Bayes paradigmasına göre, $p(b_i;\theta)$ önsel dağılım, $p(T_i,\delta_i|b_i;\theta)p(y_i|b_i;\theta)$ koşullu olabilirlik olmak üzere, rastgele etkilere ait sonsal dağılım aşağıdaki gibi elde edilir:

$$p(b_i|T_i, \delta_i, y_i; \theta) = \frac{p(T_i,\delta_i|b_i;\theta)p(y_i|b_i;\theta)p(b_i;\theta)}{p(T_i,\delta_i,y_i;\theta)}$$

$$\propto \quad p(T_i, \delta_i|b_i; \theta)p(y_i|b_i; \theta)p(b_i; \theta). \qquad \text{(Eşitlik 1.5)}$$

n_i boylamsal ölçümlerin sayısı arttıkça, bu dağılım normal dağılıma yakınsar. Bireye özel rastgele etkinin kestirimi, bu dağılımın beklenen değeri olarak alınabilir. Varyans kestirimi ise, sonsal dağılım kullanılarak aşağıdaki gibi hesaplanabilir (Rizopoulos, 2012a).

$$var(b_i) = \int \left(b_i - \bar{b}_i\right)^2 \; p(b_i|T_i, \delta_i, y_i; \theta) db_i, \qquad \text{(Eşitlik 1.6)}$$

ampirik Bayes yaklaşımı Eşitlik 1.6'da $\hat{\theta}$'nın θ yerine konulmasıyla bulunabilir.

1.5. Sağkalım Verilerinin Modellenmesi

Sağkalım verilerinin modellenmesinde amaç, hazard fonksiyonunu etkileyen açıklayıcı değişkenleri belirlemek ve hazard fonksiyonunu elde etmektir. Bu amaçla kullanılabilecek modellerden bu bölüm altında bahsedilmiştir.

1.5.1. Cox Oransal Hazard Regresyon Modeli

Sağkalım verisini modellemek için genellikle orantılı hazard modeli kullanılır. Cox (1972) tarafından önerildiğinden Cox regresyon modeli olarak da adlandırılır. Orantılı hazard varsayımı, hazard oranının zamana karşı sabit olması ya da bir grubun hazard fonksiyonunun diğer grubun hazard fonksiyonuna orantılı olması anlamına gelmektedir (Therneau ve Grambsch, 2000:145-152). Model, orantılı hazard varsayımına dayanır ancak sağkalım süreleri için belirli bir olasılık dağılımı varsayımı yoktur ve temel hazard fonksiyonun belirtilmemiş olmamasından yani

başarısızlık zamanı, T raslantı değişkeni için belirli bir olasılık dağılımı olmamasından dolayı Cox regresyon modeli yarı parametrik bir model olarak ele alınır ve modeldeki en önemli sorun β parametrelerinin tahminidir (Kleinbaum ve Klein, 2005).

$$h(t|x) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i x_i\right] \quad \text{(Eşitlik 1.7)}$$

Denklemdeki x, $(x_1, x_2, \ldots, x_p)$ şeklindeki açıklayıcı değişkenler vektörüdür. Bu formülün önemli bir özelliği, orantısal hazardlar varsayımı ile ilgili olarak, sağkalım süresini (t) içeren ancak açıklayıcı değişkenleri içermeyen temel hazard fonksiyonudur ($h_0(t)$). Açıklayıcı değişkenlerin yer aldığı kısım ise sağkalım süresinden bağımsızdır. Bu durumdaki x açıklayıcı değişkenleri zamandan bağımsız olarak adlandırılırlar. $h_0(t)$ zamanla ölüm riskinin nasıl değiştiğini gösterirken üstel kısım açıklayıcı değişkenlerin hazard fonksiyonuna etkilerini gösterir. Zamandan bağımsız değişkenler zaman içinde değeri değişmeyen değişkenlerdir. Bu değişkenler cinsiyet, sigara içme durumu gibidir. Sigara içme durumu, yaş ve ağırlık değişkenleri değişebilir ancak değişiklik fazla olmadığından zamandan bağımsız kabul edilirler.

1.6. Zamana Bağlı Açıklayıcı Değişkenlerin Sağkalım Sürecinde Ele Alınması

Zamana bağlı açıklayıcı değişkenler dışsal ve içsel olmak üzere iki grupta incelenebilir. Çevresel faktörler (örn. hava kirliliğinin seviyesi, müdahale ya da girişimin farklı uzmanlar tarafından uygulanması) dışsal açıklayıcı değişkenlerdir. Hasta üzerinde ölçülen tüm açıklayıcı değişkenler ise içseldir. İçsel açıklayıcı değişkenlere, primer biliyer siroz hastalarının serum biluribin düzeyi, HIV hastalarının CD4 hücre sayıları, karaciğer sirozu hastaların protrombin endeksi ve aort darlığına sahip hastaların ana atardamar gradyan seviyeleri gibi biyobelirteçler ve klinik parametreler örnek gösterilebilir.

Zamana bağlı dışsal açıklayıcı değişkenler, ilgilenilen olayın gerçekleşmesinden bağımsız olarak değişiklik gösterirler. Örneğin astım ataklarının incelendiği bir çalışmada, bir biyobelirteç değerinin ve hastaların yaşadığı yerdeki hava kirliliği seviyesinin izlem süresince ölçüldüğünü düşünelim. Çalışmadaki bir birey çalışmanın başlamasından 5 ay sonra astım atağı geçirsin. Bu durumda atak gerçekleştikten sonraki günlerde biyobelirtecin seviyesi bireyin astım atağı geçirmesinden etkilenecek ancak hava kirliliği seviyesi ise bu ataktan etkilenmeyecektir (Rizopoulos ve ark., 2016). Yani, bir içsel açıklayıcı değişkenin değerlerindeki zaman içindeki değişim bireyler arası biyolojik kaynaklı heterojenlikten etkilenirken; bir dışsal açıklayıcı değişkenin değerlerindeki zamana bağlı değişim bireyler arası heterojenlikten etkilenmemektedir. Bu nedenle zamana bağlı içsel değişkenler modele dahil edilirken bireyler arası heterojenliğin göz önüne alınması önemlidir. Bireyler arası biyolojik heterojenlik ise modelde rastgele etkiler kullanılarak göz önüne alınabilir.

Cox oransal hazard modeli zamana bağlı açıklayıcı değişkenlerle başa çıkmak için aşağıda verildiği gibi genişletilebilir.

$$h_i\big(t|w_i,\ y_i(t)\big) = h_0(t)\exp\{\gamma^{\mathrm{T}} w_i + \alpha m_i(t)\} \qquad \text{(Eşitlik 1.8)}$$

w_i: zamandan bağımsız temel (baseline) açıklayıcı değişkenler,
γ^{T}: zamandan bağımsız temel (baseline) açıklayıcı değişkenlere ait regresyon katsayıları,
$m_i(t)$: t anında zamana bağlı açıklayıcı değişkenler,
α: t anındaki zamana bağlı açıklayıcı değişken ile aynı zaman noktasındaki olayın riskini belirleyen regresyon katsayısı olarak tanımlanır.

Yukarıda Eşitlik 1.8 ile gösterilen model genişletilmiş Cox hazard modeli (Extended Cox Model) olarak bilinir (Kleinbaum ve Klein, 2005). Bu modelde klasik Cox modelinde yer alan X açıklayıcı değişkenleri yerine (Eşitlik 1.7), zamandan bağımsız ve zamana bağlı açıklayıcı değişkenler ile ilgili regresyon katsayıları ayrı ayrı yer

almaktadır. Bu modelde w_i zamandan bağımsız açıklayıcı değişkenleri (örn. cinsiyet, tedavi grupları, vb.); y_i ise zamana bağlı açıklayıcı değişkenleri göstermektedir. Modelde α ve γ sırasıyla zamandan bağımsız ve zamana bağımlı açıklayıcı değişkenlere ait regresyon katsayılarıdır. Zamana bağlı ortak değişkenlere ait regresyon katsayılarının yorumu şu şekilde yapılabilir. t zamanında ölçülen $y_i(t)$deki bir birimlik artış, ilgilenilen olayın t zamanındaki göreli riskini $exp(\alpha)$ kadar arttırmaktadır denilebilir. Örneğin, siroz hastalarında ölüm riski üzerine tedavinin ve hastanede yattıkları süre boyunca her hafta tekrarlı olarak ölçülen protrombin indeksi değerlerinin etkisinin incelendiği bir çalışmada; protrombin indeksine ait $exp(\alpha) = 3$ olarak bulunmuş olsun. Buna göre siroz hastalarının hastaneye yattıkları ikinci hafta protrombin indeksi ölçümlerindeki bir birimlik artış, yine ikinci haftadaki ölüm riskini 3 kat arttırmaktadır denilebilir.

Ancak genişletilmiş Cox modelinde, bireyler arası biyolojik heterojenlik göz önüne alınmadığından, zamana bağlı değişkenler içsel olduğu durumda, bireydeki biyolojik varyasyonun içsel değişkendeki değişime etkisi göz ardı edilmektedir.

Birleşik modelleme tekniğinde ise, zamana bağlı içsel açıklayıcı değişkenlerin yer aldığı boylamsal alt modeli ile sağkalım alt modeli birbirine bireyler arası heterojenliği gösteren rastgele etkiler ile bağlanmaktadır. Yani, zamana bağlı içsel değişkenlere bireysel biyolojik varyasyonların etkisi de göz önüne alınmaktadır (Rizopoulos ve ark., 2016).

1.7. Birleşik Model

Birleşik modelleme, boylamsal ve sağkalım verilerini birbirine bağlayan bir modellemedir ve bağlı iki alt modelden oluşur. Bunlardan biri boylamsal süreç için karışık etki modeline, diğeri ise, sağkalım süreci için orantılı hazard modeline karşılık gelir. Bu iki alt model bireye özel rastgele etkiler ile birbirine bağlanmaktadır. Ortak rastgele etkiler ile boylamsal ve sağkalım süreçlerinin

birbirine bağlandığı birleşik modelleme, bu iki alt model için eş zamanlı kestirim yapma olanağı sağlar.

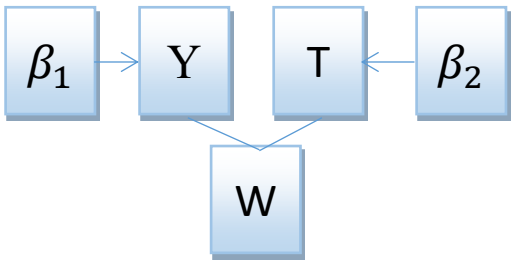
Birleşik modellemede, aynı çalışma kapsamında yer alan sağkalım süreci, boylamsal sürecin birleşik olabilirliği çözümlenir. Bu birleşik olabilirlik aşağıda gösterildiği gibi üç farklı şekilde parçalanarak çözümlenmektedir.

$$[Y,T,W] = [W]\,[Y|W]\,[T|Y]$$
$$[Y,T,W] = [W]\,[T|W]\,[Y|T]$$
$$[Y,T,W] = [W]\,[Y|W]\,[T|W] \qquad \text{(Eşitlik 1.9)}$$

Eşitlik 1.9'da yer alan Y, boylamsal süreçte toplanan tekrarlı ölçümleri, T, sağkalım sürecindeki izlem süresini, W ise rastgele birey etkilerini göstermektedir.

Birleşik olabilirliğin parçalanma biçimine göre Eşitlik 1.9'da tanımlanan modeller sırasıyla, Seçim Modelleri (Selection Models), Karışık Örüntü Modelleri (Patern Mixture Models) ve Rastgele Etki Modelleri (Random Effects Models) ya da Paylaşımlı Parametre Modelleri (Shared Parameters Models) olarak bilinmektedir Little, 1993; Diggle ve ark., 2008).

Ancak bu tez kapsamında birleşik modellemede tanımlanan olabilirlik çözümlenmesi Rastgele Etki Modelleri kullanılarak gerçekleştirilecektir. Bu yaklaşım ile daha önce de belirtildiği gibi, sağkalım ve boylamsal süreç birbirine rastgele etkiler ile bağlanmaktadır (Şekil 1.3).



Şekil.1.3. β_1 boylamsal sürece ve β_2 sağkalım sürecine ilişkin model parametreleri gösterimi

Şekil 1.3.'te β_1 ve β_2 sırasıyla boylamsal ve sağkalım süreçlerine ilişkin model parametrelerini göstermektedir.

1.7.1. Alt Modeller

1.7.1.1. Boylamsal Alt Model

Boylamsal alt model olarak rastgele ve sabit etkileri içeren karışık etkili model kullanılabilir. Karışık etkili model hiyerarşik olarak düzey 1 ve düzey 2 olmak üzere yazılabilir. Aşağıda bireyden zaman içinde alınan ölçümlerin modellendiği ve her birey için ayrı regresyon parametrelerinin yer aldığı düzey1 modeli yer almaktadır (Eşitlik 1.10).

$$y_{ij} = \tilde{\beta}_{i0} + \tilde{\beta}_{i1} t_{ij} + \varepsilon_{ij}, \qquad \text{(Eşitlik 1.10)}$$

Bu model için bireyler popülasyondan rastgele seçildiği durumda regresyon katsayılarının $\tilde{\beta}_{i0}$ ve $\tilde{\beta}_{i1}$ ilgili regresyon katsayılarının da rastgele olduğu söylenebilir.

Eşitlik 1.10'daki rastgele regresyon katsayılarının ortalama $\beta = (\beta_0, \beta_1)^{\mathrm{T}}$ vektörü ve varyans-kovaryans matrisi D ile normal dağıldığı varsayımı altında, düzey 2 modeli aşağıdaki gibi yazılabilir.

$$\tilde{\beta}_{i0} = \beta_0 + b_{i0},$$
$$\tilde{\beta}_{i1} = \beta_1 + b_{i1}, \qquad \text{(Eşitlik 1.11)}$$

İlgili parametreler düzey 1'de yazıldığında karışık etkili model aşağıdaki biçimde yazılabilir:

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \varepsilon_{ij}$$ (Eşitlik 1.12)

Bu modelde, β_0 ve β_1 parametreleri popülasyondaki ortalama kesim noktası ve zaman içindeki değişimi gösterir ve sabit etkiler olarak adlandırılır. $b_i = (b_{i0}, b_{i1})^{\mathrm{T}}$ rastgele etkiler olarak adlandırılır ve her bireyin ortalama regresyon kesim noktası ve zaman içindeki eğimden ne kadar saptığını gösterir.

Birleşik modelde kullanılmak üzere Eşitlik 1.12'de gösterilen karışık etkili model Eşitlik 1.13'teki gibi yeniden yazılabilir.

$M_i(t) = \{m_i(\mathrm{s}), 0 \leq \mathrm{s} < \mathrm{t}\};$
$y_i(t) = m_i(t) + \varepsilon_i(t),$
$m_i(t) = x_i^{\mathrm{T}}(t)\beta + z_i^{\mathrm{T}}(\mathrm{t})\, b_i,$
$b_i \sim N(O, D), \qquad \varepsilon_i(t) \sim N(O, \sigma^2),$ (Eşitlik 1.13)

$M_i(t)$; boylamsal sürecin t zamanına kadar beklenen değerlerini ifade etmektedir. Yani, boylamsal süreçte yer alan ölçümlerin, ölçüm hatasından arındırılmış gerçek değerleridir. Sağkalım modellerini ifade etmek ve t zamanındaki boylamsal açıklayıcı değişkenin beklenen değerini göstermek için $m_i(t)$ kullanılmıştır. $m_i(t)$ $y_i(t)$'den farklıdır çünkü; t zamanındaki boylamsal sonuç değişkeni için ölçüm hatasını içermemektedir. Modelde, $x_i(t)$ ve β sabit etkileri, $z_i(t)$ ve b_i rastgele etkileri göstermektedir. D rastgele etkilere ait $(q \times q)$ boyutlu genel kovaryans matrisidir. $x_i(t)$ t zamanındaki i. birey için p boyutlu açıklayıcı değişkenler vektörüdür ve hata terimleri $\varepsilon_i(t)$ zamanla değişir.

Böylelikle, bir sonraki bölüm altında Eşitlik 1.15'te tanımlanacak olan sağkalım fonksiyonu tekrarlı ölçümlerin ölçüm hatalarından arındırılmış gerçek değerlerine dayanır ve bu nedenle, $S_i(t)$,'nin doğru tahmini için $M_i(t)$'nin iyi bir tahminini elde etmek gerekir. Yani, $x_i(t)$, ve $z_i(t)$'deki zaman yapısının ve varsayılan zaman yapısı ile temel açıklayıcı değişkenler arasındaki etkileşim terimlerinin dikkatli

seçilmesi gerekir. Örneğin, bireyler doğrusal olmayan boylamsal trendler içeriyorsa, $x_i(t)$, ve $z_i(t)$'deki t zamanının fonksiyonları için yüksek dereceden polinomlar veya spline şeklinde çok boyutlu vektör kullanılması tavsiye edilir. Polinomlarla kıyaslandığında, yerel doğal spline'lar ve daha iyi sayısal özellikleri nedeniyle tercih edilirler (Ruppert ve ark., 2003).

1.7.1.2. Sağkalım Alt Modeli

Birleşik modelde kullanılan sağkalım alt modelini belirlemeden önce, kümülatif hazard fonksiyonu Eşitlik 1.14'te verildiği gibi tanımlanabilir.

$$h_i(t \mid M_i(t), w_i) = \lim_{dt \to 0} \Pr\{t \leq T_i^* < t + dt \mid T_i^* \geq t, M_i(t), w_i\}/\, dt$$
$$= h_0(t)\exp\{\gamma^T w_i + \alpha\, m_i(t)\}, \quad t>0, \qquad \text{(Eşitlik 1.14)}$$

Bu modelde $M_i(t)$ boylamsal süreçte t zamanına kadar ölçülmüş ve ölçüm hatasından arındırılmış gerçek değerleri, ve w_i; temel açıklayıcı değişkenler vektörünü (tedavi indikatörü, hastalık öyküsü) göstermektedir. γ, temel açıklayıcı değişkenlere ilişkin regresyon katsayılarını içeren parametre vektörü,

α parametresi ise, boylamsal sürecin sağkalım sürecine olan etkisini gösteren regresyon katsayısıdır.

$\exp(\gamma_j)$ t'nin herhangi bir zamanında w_{ij}'deki bir birimlik değişim için hazard oranını belirtir. exp (α) ise; bir olayın t zamanındaki riskinin göreli artışıdır ve bu aynı zaman noktasında $m_i(t)$'deki bir birimlik artışın sonucudur.

Birleşik model olabilirliğinde kullanılacak olan sağkalım fonksiyonu, kümülatif hazard fonksiyonu ile arasındaki arasındaki bilinen ilişkiden aşağıdaki gibi yazılabilir.

$S_i(\text{t} \mid M_i(t), w_i) = \Pr\{t > T_i^* \mid M_i(t), w_i)$

$= \exp\{- \int_0^t h_0(s) \exp\{\gamma^T w_i + \alpha m_i(\text{s})\}\text{ds}\},$ (Eşitlik 1.15)

Sağkalım fonksiyonu da, $M_i(t)$ boylamsal açıklayıcı değişken değerleri ile temel açıklayıcı değişkenlerin değerlerine bağlıdır.

Burada temel risk fonksiyonu h_0 (·)'ın yapısını belirlemek için çeşitli seçenekler mevcuttur. Standart seçenek; bilinen parametrik dağılıma karşılık gelen risk fonksiyonunu kullanmaktır.

Sağkalım analizi kapsamında temel risk fonksiyonu için genellikle kullanılan dağılımlar Weibull, log-normal ve Gamma'dır. İkinci seçenek, yine parametrik ancak daha esnek bir yapıda olan bir risk fonksiyonu kullanmaktır. Literatürde, esnek temel risk fonksiyonu modellemek için de birkaç yaklaşım öne sürülmüştür.

Bundan sonraki bölümde temel risk fonksiyonunu tanımlanmasında kullanılabilecek yaklaşımlardan bahsedilmiştir.

1.7.2. Temel Risk Fonksiyonu Çeşitleri

1.7.2.1. Weibull Dağılımı

Weibull dağılımı sabit hazard fonksiyonuna sahip değildir. Weibull dağılımı biçim (p) ve ölçek (λ) parametreleri ile tanımlanmaktadır. p=1 iken hazard fonksiyonu zaman arttıkça sabit kalmaktadır. Hazard fonksiyonu, p>1 iken artar ve p<1 iken

azalır. Böylelikle Weibull dağılımı, kitlenin sağkalım dağılımını artan, azalan ya da sabit risk ile modellemek için kullanılır (Klein ve Moeschberger, 1997). Lee ve Wang: "Akciğer kanseri hastaları artan hazard fonksiyonuna, başarılı ve büyük bir ameliyat geçiren hastalar ise azalan hazard fonksiyonuna sahiptir." der. Dağılımın olasılık yoğunluk, sağkalım ve hazard fonksiyonları sırasıyla aşağıda verilmiştir:

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

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

$h(t) = \lambda\, p(\lambda t)^{p-1}$ (Eşitlik 1.16)

(Klein ve Moeschberger, 1997; Lee ve Wang, 2003).

1.7.2.2. Parçalı Sabit Hazard Fonksiyonu

Hazard oranının şekli üzerinde başlangıçta bir biçim dayatılmamıştır ve zamanla nasıl değiştiğinin gözlemlenmesinde etkilidir. En önemli özelliği kullanılacak uygun zaman aralıklarının belirlenmesidir.

Parçalı sabit modeli altında, temel risk fonksiyonu

$h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \leq v_q),$ (Eşitlik 1.17)

şeklini alır.

$0 = v_0 < v_1 < \cdots < v_Q$ gösterimi zaman ölçeğinin bir bölümünü gösteririr. v_Q gözlenen en büyük zamandan daha büyüktür, ve ξ_q; $(v_{q-1}, v_q]$ aralığında hazard değerini gösterir. Düğüm sayısı arttıkça hazardı belirlemek kolaylaşır. $(v_{q-1}, v_q]$ aralığı sadece tek gerçek sağkalım zamanı ile kısıtlandırıldığında, model h_0 (·)'ı tamamen belirsiz bırakır ve parametrik olmayan en çok olabilirlik tahminine eşdeğerdir.

Ayrıca uyarlanmış Gauss-Hermite integrasyon kuralı ile çözümü için "*JM"* paketinde aşağıdaki seçenekler de mevcuttur (Rizopoulos, 2012a):

-weibull-PH-aGH
-parçalı sabit-PH-aGH
-cox-PH-aGH
-spline-PH-aGH
-weibull-AFT-aGH

1.8. Kestirim Yöntemleri

Birleşik modellerin kestiriminde, iki yaklaşım söz konusudur. Bunların birincisi ve hesaplaması daha kolay olan iki aşamalı yaklaşım, ikincisi ise birleşik olabilirliğin çözümüdür. İki aşamalı yaklaşım hesaplamada kolaylık sağlasa da, birleşik olabilirlik çözümüne göre yanlı sonuçlar verdiği bilinmektedir. Aşağıda bu iki yaklaşımdan bahsedilmiştir.

1.8.1. İki Aşamalı Yaklaşım

İki aşamalı yaklaşımda boylamsal süreç ile sağkalım sürecindeki parametreler ayrı ayrı kestirilir. İlk adımda karışık etkili modelden t = "sağkalım süresi" için elde edilen gerçek değer ($m_i(t)$), ikinci aşamada Cox mdoelinde açıklayıcı değişken olarak alınır (Self ve Pawitan, 1992).

Tsiatis ve ark. (1995), hazard fonksiyonunu doğrusallaştırmadan Eşitlik 1.13 ile Eşitlik 1.14'ü birleştiren birleşik modeli tahmin etmek için iki aşamalı bir yaklaşım geliştirmişlerdir. Her zaman noktası için $m_i(t)$ teriminin kısmi olabilirlik katkısını $\mathrm{E}\{m_i(t) \mid Y_i(t), T_i \geq t\}$ tahmininin yerine kullanmayı önermişlerdir. Bu yaklaşımlar, standart yazılımlar ile uygulanması kolay olsa da, birçok durumda yanlı

sonuçlar üretir. Bu durum; Dafni ve Tsiatis (1998); Tsiatis ve Davidian (2001); Ye ve ark. (2008); Sweeting ve Thompson (2011) tarafından simülasyonlar ile gösterilmiştir. Bu nedenle bu yaklaşımlara güvenmek yerine, birleşik modeller bu yanlılığı ortadan kaldırmak için boylamsal ve sağkalım süreçlerinin birleşik olabilirliklerinin hesaplandığı tam olabilirlik yaklaşımlarına odaklanmışlardır.

1.8.2. Birleşik Olabilirlik Çözümü

Birleşik modeller için ana tahmin yöntemi en çok olabilirliktir (Henderson ve ark, 2000; Hsieh ve ark. 2006; Wulfsohn ve Tsiatis, 1997).

Maksimum olabilirlik kestirimleri gözlenen sonuçların $\{T_i, \delta_i, y_i\}$ birleşik dağılımına karşılık gelen log-olabilirlik fonksiyonunun türevlerinin sıfıra eşitlenmesi ile bulunabilir. Bu birleşik dağılım tanımlanırken zamandan bağımsız rastsal etkiler vektörü b_i, boylamsal süreç ile sağkalım sürecini birbirine bağlamada kullanılacaktır.

$p(T_i, \delta_i, y_i)|b_i ;\theta) = p(T_i, \delta_i)|b_i;\theta)\ p(y_i|b_i;\theta)$ (Eşitlik 1.18)

ve,

$p(y_i|b_i ;\theta) = \prod_j p\ \{ y_i\ (t_{ij}) \mid b_i\ ;\ \theta\}$ (Eşitlik 1.19)

$\theta = (\theta_t^{\mathbf{T}}, \theta_y^{\mathbf{T}}, \theta_b^{\mathbf{T}})^{\mathbf{T}}$ tam parametre vektörü

θ_t : sağkalım sonuç parametresi

θ_y : boylamsal sonuç parametresi

θ_b : rastgele etki kovaryans matrisi parametreleri

y_i : i. birimin boylamsal yanıt için n_ix1 vektörü

$p(T_i, \delta_i|b_i; \theta_t, \beta)$: sağkalım kısmının koşullu olasılık yoğunluk fonksiyonu

i. birey için log-olabilirliğe katkısı aşağıdaki gibi formülize edilebilir:

$$log\ p(T_i, \delta_i, y_i; \theta) = log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \quad \text{(Eşitlik 1.20)}$$
$$= log \int p(T_i, \delta_i \mid b_i; \theta_t, \beta) \left[\prod_j p\{y_i(t_{ij}) \mid b_i;\ \theta_y\}\right] p(b_i; \theta_b)\ db_i,$$

$$p(T_i, \delta_i | b_i; \theta_t, \beta) = h_i(T_i | M_i(T_i); \theta_t, \beta)^{\delta_i} S_i(T_i | M_i(T_i); \theta_t, \beta) \quad \text{(Eşitlik 1.21)}$$
$$\left[h_0(T_i) exp\{\gamma^{\mathrm{T}} w_i + \alpha m_i(T_i)\}\right]^{\delta_i} \times \exp\left(-\int_0^{T_i} h_o(s) exp\{\gamma^{\mathrm{T}} w_i + \alpha m_i(s)\}\, ds\right),$$

h_0 (·); Eşitlik 1.17'deki parçalı-sabit modeli gibi ya da herhangi bir bilinen bir dağılımın hazard fonksiyonu ve Eşitlik 1.15'te verilen sağkalım fonksiyonu gibi, zamanın pozitif herhangi bir fonksiyonudur (Rizopoulos, 2012a).

Boylamsal yanıt değişkeninin rastgele etki içeren birleşik yoğunluk fonksiyonu:

$$p(y_i | b_i; \theta) p(b_i; \theta) = \prod_j p\{y_i(t_{ij}) | b_i; \theta_y\}\ p(b_i; \theta_b) \quad \text{(Eşitlik 1.22)}$$
$$= (2\pi\sigma^2)^{-n_i/2}\{-\exp\|y_i - X_i\beta - Z_i b_i\|^2 / 2\sigma^2\}$$
$$\times (2\pi)^{-q_b/2} \det(D)^{1/2} \exp\left(-\mathrm{b}_i{}^T D^{-1} b_i / 2\right),$$

q_b : rastgele etkiler vektörünün boyutluluğu
$\|x\| = \{\sum_i {x_i}^2\}^{1/2}$: Öklit vektör normu olarak ifade edilir (Rizopoulos, 2012a).

Log-olabilirlik fonksiyonu $\ell(\theta) = \sum_i logp\,(T_i, \delta_i, y_i; \theta)$ θ'nın esas alınarak en büyüklemesi, standart logaritmalar kullanılarak; örneğin EM (Expectation Maximization) ya da Newton Raphson ya da türevleri kullanılarak elde edilir (Rizopoulos, 2012a).

Birleşik modelleme literatüründe EM algoritmasının tercih edilmesinin nedeni (rastgele etkiye 'kayıp veri' gibi davranılması) M-adım algoritmasını uygularken bazı parametrelerin kapalı-formlarının (closed-form) güncellenebilir olmasıdır.

Ancak; EM algoritmasının önemli bir dezavantajı özellikle maksimumun yakınında yavaş yakınsama ile sonuçlanan doğrusal yakınsama oranıdır (Rizopoulos, 2012a).

Bununla birlikte, Rizopoulos ve ark. (2009); $\ell(\theta)$ karşılık gelen skor vektörü, aslında EM ve Newton tipi algoritmasının da anahtar fonksiyonu olduğunu belirtmiştir.

Bu skor vektörü aşağıdaki biçimde yazılabilir:

$$\begin{aligned} \mathrm{S}(\theta) &= \sum_i \frac{\partial}{\partial \theta^{\mathrm{T}}} \log \int p(T_i, \delta_i \,|b_i; \theta) p(y_i \,|b_i; \theta) p(b_i; \theta)\, db_i \\ &= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \frac{\partial}{\partial \theta^{\mathrm{T}}} \int p(T_i, \delta_i \,|b_i; \theta) p(y_i \,|b_i; \theta) p(b_i; \theta)\, db_i \\ &= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \left[\int \frac{\partial}{\partial \theta^{\mathrm{T}}} \{ p(T_i, \delta_i \,|b_i; \theta) p(y_i \,|b_i; \theta) p(b_i; \theta)\, db_i \} \right] \\ &\qquad \times \frac{p(T_i, \delta_i \,|b_i; \theta) p(y_i \,|b_i; \theta) p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)}\, db_i \\ &= \sum_i \int A(\theta, b_i)\, p(b_i | T_i, \delta_i, y_i; \theta) db_i, \end{aligned} \qquad \text{(Eşitlik 1.23)}$$

$A(\cdot)$: tam veri skor vektörü

$$A(\theta, b_i) = \partial \{ \log p(T_i, \delta_i \,|b_i; \theta) + \log p(y_i \,|b_i; \theta) + \log p(b_i; \theta) \} / \partial \theta^{\mathrm{T}} \qquad \text{(Eşitlik 1.24)}$$

olarak ifade edilir.

Gözlenmiş veri vektörü rastgele etkilerin sonsal dağılıma göre tam veri vektörünün beklenen değeri olarak ifade edilir (Rizopoulos, 2012a).

Eşitlik 1.23'e karşılık gelen skor vektörleri θ'ya göre çözülürse, bir önceki iterasyonun θ değeri sabit ve $p(b_i | T_i, \delta_i, y_i; \theta)$ ile bu EM algoritmasına karşılık gelir, eğer skor vektörü θ'ya göre ve θ'nın bir fonksiyonu olarak çözülürse, bu $\ell(\theta)$ gözlenen verinin log-olabilirliğinin doğrusal maksimizasyonuna karşılık gelir. Bu şekilde, parametre tahminleri için standart hatalar hesaplanır.

EM algoritmasını kullanarak birleşik model tahmin edilse de, (Eşitlik 1.23) veri skoru vektöründen faydalanarak Hessian matrisini ve gözlenen bilgi matrisinin standart hataları hesaplanabilir (diğer bir deyişle Hessian matrisinin tersinin negatifi).

Yukarıdaki skor vektörü türetme gibi benzer hesaplamaları kullanarak, Hessian matrisi aşağıdaki şekilde yazılabilir:

$$\frac{\partial S_i(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \int A(\theta, b_i)\, p(b_i|T_i, \delta_i, y_i; \theta)\, db_i$$
$$= \int \frac{\partial A(\theta, b_i)}{\partial \theta} p(b_i|T_i, \delta_i, y_i; \theta)\, db_i$$
$$+ \underbrace{\int A(\theta, b_i)\, \frac{\partial p(b_i|T_i, \delta_i, y_i; \theta)}{\partial \theta}\, db_i,}_{I_1} \qquad \text{(Eşitlik 1.25)}$$

$$I_1 = \int A(\theta, b_i)\, \left\{\frac{\partial log p(b_i|T_i, \delta_i, y_i; \theta)}{\partial \theta}\right\}^{\mathrm{T}} p(b_i|T_i, \delta_i, y_i; \theta)\, db_i$$
$$= \int A(\theta, b_i) \qquad \text{(Eşitlik 1.26)}$$

$$\left\{\frac{\partial\{\log p(T_i, \delta_i \mid b_i; \theta) + \log p(y_i|b_i; \theta) + \log p(b_i;, \theta)\}}{\partial \theta} - \frac{\partial \log p(T_i, \delta_i, y_i; \theta)}{\partial \theta}\right\}^{\mathrm{T}} p(b_i|T_i, \delta_i, y_i; \theta)\, db_i$$

$I(\theta)$'nın tahmininden sonra, parametre tahminleri için standart hatalar tahmin edilen gözlenen bilgi matrisini temel alır, yani;

$$\mathrm{v\hat{a}r}(\theta) = \left\{I\left(\hat{\theta}\right)\right\}^{-1}, \mathrm{I}\left(\hat{\theta}\right) = -\sum_{i=1}^{n} \frac{\partial S_i(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}} \qquad \text{(Eşitlik 1.27)}$$

Olayların ortaya çıkması nedeniyle oluşan ayrılmalar sebebiyle gözlemlenen bilgi matrisi, standart hatalar için hesaplanan bilgi matrisi yerine tercih edilebilir. (Rizopoulos, 2012a).

1.8.3. İntegrasyon Yöntemi

Birleşik modelin tanımındaki iki integral; bizi rastgele etkilerin varlığını hesaplama zorluğuna sokar. Sağkalım fonksiyonundaki integral tanımı tek boyutludur ve 7-noktalı veya 15-noktalı Gauss-Kronrod kuralı uygulanarak yaklaşık sonuç elde edilebilir (Press ve ark., 2007). Rastgele etkilere göre integral almak boyut arttığı için gittikçe zorlaşır.

Gaussçu çeyreklik (quadrature) kuralları ve Monte Carlo örneklemesi bu çok boyutlu integralleri hesaplamada ve birleşik modelleri kestirmede literatürde sıkça kullanılır (Henderson ve ark., 2000; Song ve ark., 2002; Wulfsohn ve Tsiatis, 1997).

Rizopoulos ve ark. (2009) ve Ye ve ark. (2008) özellikle, çok boyutlu rastgele etkiler belirlendiği durumlarda; Gaussçu çeyreklik ve Monte Carlo'ya göre Laplace yaklaşımlarının daha etkili olduğunu tartışmıştır. Bununla birlikte, bu yaklaşımların hepsi hesaplamada zordur ve modern uygulamalı istatistikte birleşik modellerin hak ettikleri yeri bulamamasının temel nedenidir.

Hesaplama yükünü azaltan alternatif bir yaklaşım, Rizopoulos tarafından önerilmiştir (Rizopoulos, 2012a). Bu yaklaşımın ardındaki fikir ilk olarak; boylamsal sonuç için karışık etkili modeli kestirmek ve her birey için verilen boylamsal sonuç değişkeninin rastgele etkilerin sonsal dağılımının konumu ve ölçeği hakkında bilgi elde etmektir.

Bu bilgiler daha sonra birleşik modelin log-olabilirlik ve skor vektörü için konuya yönelik integrandların yeniden ölçeklendirilmesi için kullanılır. Bu yaklaşım için, öncelikle birleşik bir modelin skor vektörünün hesaplanması için standart ve uyarlanabilir Gauss-Hermite kuralları bilinmelidir.

Standart Gauss-Hermite kuralınca ve rastgele etkilerin fonksiyonunun herhangi bir formu A (·) için, skor vektörün tanımındaki integrali x-eksenindeki belirli integrand ölçümlerinin ağırlıklı toplamı ile yaklaşılır (Rizopoulos, 2012a).

$$E\{A(\theta; b_i)|T_i, \delta_i, y_i; \theta\} = \int A(\theta; b_i)p\,(b_i|T_i, \delta_i, y_i; \theta)db_i$$
$$\approx 2^{q_b/2} \sum_{t_1 \ldots t_q} \pi_t A\big(\theta, b_t\sqrt{2}\big)p\big(b_t\sqrt{2}|T_i, \delta_i, y_i; \theta\big)\exp(\|b_t\|)^2, \quad \text{(Eşitlik 1.28)}$$

$\sum_{t_1 \ldots t_q}$; $\sum_{t_1=1}^{K} \ldots \sum_{t_q=1}^{K}$ ifadelerinin kısa gösterim şeklidir. K çeyreklik noktalarını belirtir. $b_t{}^{\mathrm{T}} = \left(b_{t_1}, \ldots, b_{t_q}\right)$: π_t ağırlığına karşılık gelen x-ekseni değerlerini belirtir (Rizopoulos, 2012a).

Eğer integrand $exp(-b^{\mathrm{T}}b)l(b)$ ve $l(b)$: $2K - 1$ dereceli polinom veya daha az apsis enterpolasyonu şeklinde ifade edilebiliyorsa; apsisler ve ağırlıkların konumu integrand için tam bir çözüm verebilir. K çeyreklik noktaları arttığında yaklaşım kalitesi de artar (Rizopoulos, 2012a).

Gauss-Hermite kuralı; her rastgele etki için apsislerin kartezyen çarpımı üzerindeki integrantın değerlendirmesini gerektirdiğinden dolayı, q_b arttıkça hesaplama yükü katlanarak artar. Örneğin; bir rastgele sabit 5 noktalı çeyreklik ile 5 belirli noktada olabilirliği belirlerken, eğer rastgele sabit ve eğim modeline genişletilirse, her bir rastgele etki için 5 noktalı çeyreklik olacaktır ve olabilirlik 25 nokta ile çözümlenecektir (Rizopoulos, 2012a).

Eğer $g(b) = A(\theta, b)p(b|T_i, \delta_i, y_i; \theta)$ sıfırdan uzak bir nokta etrafında konumlanmışsa ya da $g(b)$'deki yayılma ağırlık fonksiyonu $exp(-b_2)$'den oldukça farklıysa, Gauss-Hermite kuralını uygulamak $g(b)$'nin yaklaşımını zayıflatır, hatta çok büyük K için de, çünkü çeyreklik kuralında apsisler $g(b)$'nin büyük çoğunluğunun konumlandığı yerde konumlanamayacaktır (Pinheiro ve Bates, 1995). Bu problemi çözmek için uyarlanmış Gauss-Hermite kuralı intergrantı her iterasyonda ortalar ve ölçer (Rizopoulos, 2012a).

$$\hat{r}_t = \hat{b}_i + \sqrt{2}\hat{B}_i^{-1} b_t,$$
$$\hat{b}_i = \text{argmax}_b\{\log p(T_i, \delta_i, y_i; \theta)\}$$
$$\widehat{H}_i = -\partial^2 \log p(T_i, \delta_i, y_i; \theta)/\partial b \partial b^{\mathrm{T}}|_{b=\hat{b}_i} \quad \text{(Eşitlik 1.29)}$$

$\hat{B}_i$: $\widehat{H}_i$'nin Choleski faktörünü ifade eder.
Bu dönüşümü kullanmak, integrant $N(0,\ 2^{-1}I)$ dağılımının yoğunluğu gibi davranır ve Gauss-Hermite ağırlık fonksiyonu bu yoğunlukla orantılı olduğu için optimal yaklaşım elde edilir.

Böylece, uyarlanmış Gauss-Hermite kuralı aynı büyüklükte bir yaklaşım hatasını elde etmek için standart Gauss-Hermite'e kıyasla genellikle daha az çeyreklik gerektirir.

Bununla birlikte, her birey için her iterasyonda $\hat{b}_i$ modunun konumu ve $\widehat{H}_i$ matrisinin ikinci dereceden türevinin hesaplanması yükü oldukça artırır (Rizopoulos, 2012a).

Uyarlanmış Gauss-Hermite kuralının ardındaki zorluk, $\hat{b}_i$ modu ve matrisinin ikinci dereceden türev matrisi $\widehat{H}_i$ olan rastgele etkilerin sonsal dağılımının $p(b_i|T_i, \delta_i, y_i; \theta)$ özellikleri kullanılarak azaltılabilir (Rizopoulos, 2012a).

Yoğunluk fonksiyonu log-ölçeğinde yazılırsa,

$$\log p(b_i|T_i, \delta_i, y_i; \theta)$$

$$\propto \sum_{j=1}^{n_i} \log\ p\left\{y_i(t_{ij})|b_i;\ \theta_y\right\} + \log p(b_i;\ \theta_b) + \log p(T_i, \delta_i|\ b_i; \theta_t, \beta) \text{(Eşitlik 1.30)}$$

şeklini alır (Rizopoulos, 2012a).

Bayescil merkezi limit teoreminin bir versiyonu kullanılarak (Cox ve Hinkley, 1974, s.: 399–400) ve genel uygunluk koşulları altında $n_i \to \infty$ elde edilir.

Uyarlanmış Gauss-Hermite kuralındaki standart dönüşüm yerine (Eşitlik 1.28), ilk olarak doğrusal karışık etkili model kestirilir, ampirik Bayes tahminleri hesaplanır; $\tilde{b}_i = \text{argmax}_b\{\log p(y_i, b; \tilde{\theta}_y)\}$ ve $\tilde{H}_i^{-1}$ kovaryans matrisi $\tilde{H}_i = -\partial^2 \log p(y_i, b; \tilde{\theta}_y)/\partial b \partial b^T\big|_{b=\tilde{b}_i}$ ile gösterilir ve

$$E\{A(\theta; b_i)|T_i, \delta_i, y_i; \theta\}$$
$$\approx 2^{q_b/2}\left|\hat{B}_i\right|^{-1} \sum_{t_1 \ldots t_q} \pi_t A(\theta, \tilde{r}_t) p(\tilde{r}_t|T_i, \delta_i, y_i; \theta) \exp(\|b_t\|)^2 \text{ (Eşitlik 1.31)}$$

dönüşümü kullanılır, $\tilde{r}_t = \tilde{b}_i + \sqrt{2}\tilde{B}^{-1} b_t$ $\tilde{H}_i$'nin Choleski faktörü ve $\tilde{\theta}_y$ doğrusal model kestirimden en çok olabilirlik tahminidir (Rizopoulos, 2012a).

Bu prosedür, uyarlanmış Gauss-Hermite kuralına çok benzer, ancak optimizasyonun başında sadece bir kez uygulanır ve sonra çeyreklik noktalarını güncellemeye gerek duymaz. Bu hesaplamanın iki avantajı vardır; birincisi standart Gauss-Hermite kuralından daha az çeyreklik noktası kullanılır, ikincisi; uyarlanmış Gauss-Hermite kuralının her iterasyonunda çeyreklik noktalarının yer değiştirmesi engellenir (Rizopoulos, 2012a).

1.9. Model Seçim Kriterleri ve Model Uyumu

Model değerlendirme adımından sonra uygun olduğu düşünülen birden çok model kurulabilir. Bu modeller belli bilgi kriterleri yardımıyla ile karşılaştırılabilir. Bu kriterlerin önde gelenleri Akaike Bilgi Kriteri (AIC) ve Bayesci Bilgi Kriteri (BIC)'dir.

Çizelge 1.1'de Akaike (AIC) ve Bayesci Bilgi Kriteri (BIC) verilmiştir. Bu eşitliklerde p, bilinmeyen parametrelerin sayısı, n toplam gözlem sayısı ve L modelin olabilirlik fonksiyonu olarak ifade edilmektedir.

Çizelge 1.1. Model Seçim Kriterleri

$\mathrm{AIC} = -2\mathrm{logL} + 2\mathrm{p}$
$\mathrm{BIC} = -2\mathrm{logL} + \mathrm{pIn(n)}$

AIC değerleri her bir dağılım için hesaplanır değeri en küçük olan dağılım en çok uyumu sağlar. Geniş örneklemde BIC yöntemi AIC'e göre daha iyi sonuçlar vermektedir(Klein ve Moeschberger, 2003).

1.9.1. Artık Analizi

Seçilen modelin veriye uyumu artıkların incelenmesi yoluyla değerlendirilebilir.

1.9.1.1. Sağkalım Sürecine İlişkin Artıklar

1.9.1.1.1. Cox-Snell Artıkları

$t_1 < t_2 < \cdots < t_D$ başarısızlık sürelerini ve d_i, t_i zamanındaki başarısızlık sayılarını göstersin. Birikimli temel hazard oran tahmin edicisi;

$$H_0(t) = \sum_{t_i \le t} \frac{d_i}{\sum_{j \in R(t_i)} exp\left(\hat{\beta} Z_j(s)\right)} \qquad \text{(Eşitlik 1.32)}$$

şeklindedir. Burada $\hat{\beta}$, tahmini regresyon katsayılarını, Z_j'ler de açıklayıcı değişkenleri göstermektedir.

$$H_0(t) = \int_0^t h_0(u)\, d_u \qquad \text{(Eşitlik 1.33)}$$

Olarak verilir ve bu gözlemlenen başarısızlık sürelerinde sıçralamalar ile bir adım fonksiyonudur. Burada $h_0(u)$ temel hazard fonksiyonunu göstermektedir. (Tuncer, 2014).

Cox-Snell artıkları, sağkalım verilerinin analizinde en sık kullanılan artık türüdür ve Cox ve Snell tarafından önerilmiştir (Cox ve Snell, 1968). Bu artıklar, modelin uygunluğunu denetlemek için kullanılır.

i. gözlem için Cox-Snell artığı,

$$r_{c_i} = H_0(t_i) exp(\hat{\beta}' Z_i) \qquad \text{(Eşitlik 1.34)}$$

biçimindedir. Burada $H_0(t)$, Eşitlik 1.32'de verildiği gibidir. $\hat{\beta}$, tahmini regresyon katsayılarını, Z_i'ler de açıklayıcı değişkenleri göstermektedir.

Eğer seçilen model veriye uygunsa ve $\hat{\beta}$ değerleri β değerlerine yakınsa o zaman r_{c_i}'ler birim üstel dağılımlı durdurulmuş bir örneklem gibi dağılır.

Bu artıklar negatif değildir ve grafiklerinin yorumlanması kolay değildir. Cox-Snell artıkları model uyumunun tespitinde çok kullanışlıdır. Ancak, tahmin edilen birikimli hazard grafiği doğrusal olmadığında bulunan modelden sapma türlerini belirleyememektedir.

Cox-Snell artıkları aşağıdaki özelliklere sahiptir, sıfır etrafında simetrik dağılım göstermez, sıfır ile sonsuz arasında değer alır yani negatif olamaz, uygun model uydurulduğunda Cox-Snell artıklarının üstel dağılımlı olduğu varsayıldığından oldukça çarpık bir dağılımı vardır ve i. artığın ortalaması ve varyansı 1 olur (Tuncer, 2014).

Diğer bir nokta ise eğer en büyük sağkalım süresi durdurulmamış ise r_{c_i} bu gözlem için tanımsızdır.

Cox-Snell grafiği genellikle sadece kaba bir tanı olarak kullanılır.

1.9.1.1.2. Martingale Artıkları

Martingale artıkları, Barlow ve Prentice tarafından önerilmiştir (Stepanova ve Thomas,2002). Zamana bağlı olmayan açıklayıcı değişkenler içeren Cox model için, i. gözlemin t_i zamanındaki değeri ve δ_i olay durumu ($\delta_i = 0$ iken δ durdurma süresi ve δ_i =1 iken t_i olay süresi) ise Martingale artığı,

$$r_{m_i} = \delta_i - H_0(t)exp(\hat{\beta}' Z_i) = \delta_i - r_{m_i} \quad \text{(Eşitlik 1.35)}$$

biçimindedir.

Eşitlik (1.35)'de görüldüğü gibi Martingale artıkları Cox-Snell artıklarının doğrusal dönüşümüdür (Tuncer, 2014, s:22).

Martingale artıkları ve dönüşümleri model belirlemede kullanılır. Açıklayıcı değişkenlere karşı çizilen Martingale artıkları grafiği modeldeki açıklayıcı değişkenlerin fonksiyonel şeklini belirlemede kullanılır. Modelin uygunluğu için değişkenlerin herhangi bir dönüşüme ihtiyacı olup olmadığını gösterir (Lin ve ark., 1993). En uygun biçim belirlenmek istendiğinde model kurulduktan sonra Martingale artıklarına karşı değişkenlerin istenilen dönüşümlerinin grafiği çizdirilir. Eğer dönüşüm uygun ise grafikteki eğri yaklaşık olarak doğrusal olur (Anonim, 2014).

Martingale artıkları, her birim için gözlenen ve beklenen başarısızlık sayıları arasındaki farkla ilgilenir. Çünkü hiç bir olayda bir başarısızlıktan fazla başarısızlık

olamaz varsayılır. Eşitlik 1.35'nın ikinci faktörü (0, t_i) aralığının üstündeki birim için başarısızlıkların birikimli hazardlarının tahminidir (Tableman ve Kim, 2004).

Martingale artıkları $\sum_{i=1}^{n} r_{m_i}$ özelliğine sahiptir (Gharibvand ve ark., 2008).

Martingale artıkları -∞ ve 1 arasında değer alır ve durdurulmuş gözlemler ($\delta_i = 0$) için artıklar negatif değer alır (Collett, 1994).

Martingale artıkları bir değerinin yakınlarında değer alıyorsa beklenenden daha kısa sağkalım süresi vardır, eğer büyük negatif bir değer alıyorsa beklenenden daha uzun sağkalım süresine sahiptir. Doğrusal regresyon analizindeki artıklarla benzer özellikleri vardır. Fakat Martingale artıkları sıfır etrafında simetrik dağılmaz, çarpıktır (Collett, 1994).

Martingale artıkları genellikle bireye özeldir, çalışmanın sonunda hesaplanır ve hazardın logaritmasında açıklayıcı değişkenlerin doğrusal etkileri varsayımının test edilmesinde yararlıdır (Borucka, 2010).

Doğrusal olmama Cox orantılı hazard modellerinde karşılaşılabilecek olası bir problemdir. Martingale artıkları eğriselliği saptamak için açıklayıcı değişkenlere karşı çizilebilir. Ayrıca doğrusal modellerde bileşen-artı-artık (ya da kısmi artık) grafikleri oluşturmak için de kullanılabilir (Fox, 2002).

1.9.1.2. Boylamsal Sürece İlişkin Artıklar

Boylamsal modelin değerlendirilmesi için marjinal artıklar ve koşullu artıklar olmak üzere iki farklı artık tipi kullanılmaktadır. Bunlardan ilki olan marjinal artıklar, y_i marjinal modeli temel alır ve aşağıdaki eşitlik ile gösterilir.

$$y_i = X_i\beta + \varepsilon_i^*$$
$$\varepsilon_i^* \sim N\left(0, Z_i D Z_i^T + \sigma^2 I_{n_i}\right)$$
$$r_i^{ym} = y_i - X_i\hat{\beta} \qquad \text{(Eşitlik 1.36)}$$

ve standartlaştırılmış gösterimi aşağıdaki gibidir:

$$r_i^{ysm} = \hat{V}_i^{-1/2}\left(y_i - X_i\hat{\beta}\right) \qquad \text{(Eşitlik 1.37)}$$

$\hat{V}_i = Z_i\hat{D}Z_i^T + \hat{\sigma}^2 I_{n_i}$ gözlenen değerlerin tahmin edilen marjinal kovaryans matrisini gösterir.

r_i^{ym} marjinal artıklar; $y_i - X_i\hat{\beta} = Z_i b_i + \varepsilon_i$ marjinal hataları tahmin eder ve $X_i\beta$ ortalama yapısının yanlış belirlenmemesi için V_i bireyler arası kovaryans yapısının varsayımlarının sağlanmasında kullanılır.

İkinci tip olan koşullu artıklar ise aşağıdaki gibi hesaplanabilir:

$$\hat{\varepsilon}_i = \left(y_i - X_i\hat{\beta}\right) - Z_i\hat{b}_i \qquad \text{(Eşitlik 1.38)}$$

Marjinal ve koşullu artıkların, bağımlı değişkenin kestirim değerlerine ($\hat{y}_i$) karşı çizilmiş grafiği sırasıyla model ortalama yapısının iyi tanımlanıp tanımlanmadığı ile koşullu artıklar arasındaki varyans homojenliği varsayımı bozulumlarının incelenmesinde kullanılır. Y-eksenindeki artıklara karşı x-ekseninde $\hat{y}_i$'lerin çizildiği grafiklerde, lowess eğrisinin 0 yakınında seyretmesi beklenir (Rizopoulos, 2012b).

2. GEREÇ VE YÖNTEM

2.1. Uygulama Verileri

Akut solunum yetmezliği, solunum fonksiyonunun veya akciğerde oksijen/ karbondioksit gaz değişiminin yetersiz olması olarak tanımlanır. Diğer bir deyişle, arteriyel kanda parsiyel oksijen (PaO_2) ve karbondioksit ($PaCO_2$) basıncı değerlerinin fizyolojik sınırlarda sürdürülememesidir. Akut solunum yetmezliği, geliştiğinde hastaya non-invaziv pozitif basınçlı mekanik ventilasyon (NPPV) uygulanması, optimum medikal tedavinin ardından ilk seçenek olarak her geçen gün artan yüzdelerle tedavide yerini almaktadır. Non-invaziv pozitif basınçlı mekanik ventilasyon (NPPV) herhangi bir yapay solunum yolu olmaksızın yapılan mekanik ventilasyondur. Çalışmada kullanılacak veriler, Ankara Üniversitesi Tıp Fakültesi Göğüs Hastalıkları Kliniği tarafından sağlanmıştır. 2013 yılında klinikte yoğun bakım ünitesinde akut hiperkapnik solunum yetmezliği tanısı ile takipte olan 40 yaş üstü 88 hasta alınmıştır. Yoğun bakım ünitesinde akut hiperkapnik solunum yetmezliği nedeniyle takipte olan hastalar çalışmaya dahil edilmiştir. Hastaların tedavi öncesi (0. saat), tedavi sonrası 1.saat, 2.saat, 4.saat, 12.saat, 24.saat, 48.saat, 72.saat, 96.saat, 120.saat PaO_2 / FiO_2 ölçüm değerleri alınmıştır. Ayrıca, bütün bireylerin yaş ve beden kitle indeksi, MRC dispne skoru, Charlson Komorbidite skoru, Glasgow koma skoru, APACHE 2 skoru ölçümleri de kaydedilmiştir.

Parsiyel oksijen oksijen fraksiyonu oranı (PaO_2 / FiO_2) ölçümleri çalışmada bağımlı değişken olarak alınacaktır. Tekrarlı ölçümlerin alındığı zaman noktaları saat bazında alınarak analize alınmıştır. Tedavi öncesi ölçüm zamanı 0. saat olarak alınmıştır. Koroner arter hastalığı dikotom; başlangıç düzeyi beden kitle indeksi, lökosit, Glasgow skoru ise sürekli değişkenler olarak alınmıştır.

2.2. Uygulama Yöntemleri

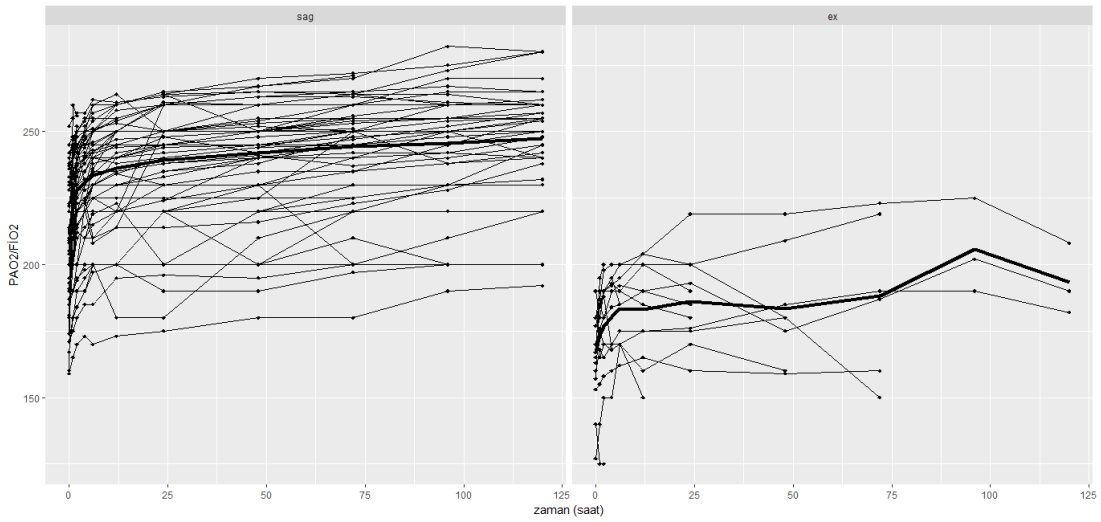
Bu çalışmada veri analizleri R yazılımında, sağkalım verileri için cox regresyon analizi "survival" kütüphanesi, boylamsal veriler için doğrusal karışık etkili model "nlme" kütüphanesi, birleşik model uygulaması, "JM" kütüphanesi kullanılarak yapılmıştır (R Core Team, 2015).

2.2.1. Açıklayıcı Değişkenlerin Düzenlenmesi

Çalışmada yaş, miktar ve test değeri gibi sayısal değişkenlerin haricinde, daha önce tedavi olup olmama durumu, cinsiyet gibi dikotom değişkenler de bulunmaktadır. Söz konusu nitel özellik bulunuyor ise "1" bulunmuyor ise "0" değerini almaktadır. Çizelge 2.1.'de değerlendirilecek açıklayıcı değişkenlerin ifadeleri yer almaktadır.

Çizelge 2.1. Değerlendirilecek açıklayıcı değişkenler

değişken sembolü	**değişken adı**
yas	Yaş
cins	Cinsiyet
apache	Apache 2 Skoru
kah	Koroner Arter Hastalığı
hb	Hemoglobin (g/dL)
entubasy	Entübasyon
npvvoyku	Daha Önce Nppv Kullanımı
kby	Böbrek Yetmezliği
svo	İnme
htk	Hematokrit (%)
bki	Beden Kitle İndeksi
aritmi	Kalbin Düzensiz Atışı
charlson	Charlson Komorbidite İndeksi
sigara	Sigara Kullanımı
glasgow	Glasgow Koma Skoru
diyabet	Şeker Hastalığı
PaO_2 / FiO_2	Parsiyel Oksijen Oksijen Fraksiyonu Oranı
lökosit	Lökosit Sayısı



Şekil 2.1. Bireylerin Zaman İçindeki Boylamsal Trendleri

İki Sürecin Ayrı Modellenmesi

Boylamsal veriler için doğrusal karışık etkili model kurulur. Bu model kendi içinde yorumlanır. Sağkalım süreci için genişletilmiş Cox regresyon modeli kurulurken, boylamsal süreçte ilgilenilen biyobelirteç değeri zaman bağlı ortak değişken olarak modele eklenir.

İki Aşamalı Modelleme

Bu modelin ardındaki fikir, ilk aşama olan karışık etkili model kurulduktan sonra, elde edilen bağımlı değişken tahminlerinin ikinci aşama olan genişletilmiş Cox modelde zamana bağlı açıklayıcı değişken olarak yerine konmasıdır. Çalışmada, doğrusal karışık etkili modelden elde edilen PaO_2 / FiO_2 tahminleri, ikinci adım olan genişletilmiş Cox regresyon analizinde açıklayıcı değişken olarak yerine konulmuştur. Parametre kestirimleri, standart hatalar ile p-değerleri Çizelge 3.5'te gösterilmiştir.

2.2.2. Birleşik Modeller için kullanılan R Fonksiyonları

Doğrusal karışık etkili modelin kurulma aşamasında tüm değişkenler modele dahil edilmiş, sonrasında anlamlı olan ($p<0,05$) değişkenler modelde bırakılmıştır.

Model 1.1.

$$\mathrm{Pa}O_2/\mathrm{Fi}O_{2_{ij}} =$$
$$\beta_0 + \beta_1 zaman +$$
$$\beta_2 cinsiyet + \beta_3 lökosit + \beta_4 svo + \beta_5 apache + \beta_6 kah + \beta_7 HB + \beta_8 charlson$$
$$+ \beta_9 crp + \beta_{10} glasgow + \beta_{11} sigara + \beta_{12} aritmi + \beta_{13} htk + \beta_{14} kby$$
$$+ \beta_{15} nppvoyku + \beta_{16} entubasy + b_{0i} + b_{1i} * zaman + \varepsilon_{ij}$$

Model 1.2.

$$\mathrm{Pa}O_2/\mathrm{Fi}O_{2_{ij}} = \beta_0 + \beta_1 zaman + \beta_2 bki + b_{0i} + b_{1i} * zaman + \varepsilon_{ij}$$

Cox regresyon modeli için; ilk aşamada ilgili değişkenlerin birer birer anlamlılığına bakılmış ve anlamlılık düzeyi 0,30'un üzerindeki değişkenler analize dahil edilmemiştir (Model 1.3). Daha sonra geriye doğru eleme yöntemi ile klinik anlamlılık da göz önünde bulundurularak Model 1.4. elde edilmiştir.

Model 1.3.

$$h(t|x) = h_0(t)\, exp(\beta_1 \mathrm{kah} + \beta_2 lökosit + \beta_3 apache + \beta_4 yas + \beta_5 glasgow + \beta_6 sigara + \beta_7 aritmi + \beta_8 htk + \beta_9 svo + \beta_{10} kby + \beta_{11} kby + \beta_{12} npvvoyku + \beta_{13} entubasy + \beta_{14} hb)$$

Model 1.4.

$$h(t|x) = h_0(t) exp(\beta_1 lökosit + \beta_2 glasgow + \beta_3 kah)$$

Birleşik modeli kurarken doğrusal karışık etkili model için oluşturulan model 1.2. ve Cox regresyon modeli için oluşturulan model 1.4. birlikte alınmıştır. Çizelge 2.2'de R fonksiyon girişi gösterilmiştir.

Çizelge 2.2. R programında birleşik fonksiyon girişi - Model 1

```
avaps <-read.csv("data.csv", header=TRUE,na.strings=c("","NA"))
avaps.id <- avaps [!duplicated(avaps$id), ]
cox.fit<-coxph(Surv(Izlemsure, ex)~kah+lokosit+glasgow, data=avaps.id, x=TRUE)
lme.fit <-lme(pao2/fio2~ Zaman+bki, ~ Zaman | id, data = avaps)
joint.fit<- jointModel(lme.fit, cox.fit, timeVar = "Zaman", method ="piecewise-PH-aGH")
```

Çizelge 2.2'de verilen fonksiyon girişi, temel hazard fonksiyonunun parçalı-sabit oransal hazard ile kestirimini göstermektedir.

Çizelge 2.3. R programında birleşik fonksiyon girişi - Model 2

```
avaps <-read.csv("data.csv", header=TRUE,na.strings=c("","NA"))
avaps.id <- avaps [!duplicated(avaps$id), ]
cox.fit<-coxph(Surv(Izlemsure, ex)~kah+lokosit+glasgow, data=avaps.id, x=TRUE)
lme.fit <-lme(pao2/fio2~ Zaman+bki,  ~ Zaman | id, data = avaps)
joint.fit<- jointModel(lme.fit, cox.fit, timeVar = "Zaman", method ="weibull-PH-aGH")
```

Çizelge 2.3'te verilen fonksiyon girişi, temel hazard fonksiyonunun Weibull oransal hazard ile kestirimini göstermektedir.

Çizelge 2.4. R programında birleşik fonksiyon girişi - Model 3

```
avaps <-read.csv("data.csv", header=TRUE,na.strings=c("","NA"))
avaps.id <- avaps [!duplicated(avaps$id), ]
cox.fit<-coxph(Surv(Izlemsure, ex)~kah+lokosit+glasgow, data=avaps.id, x=TRUE)
lme.fit <-lme(pao2/fio2~ Zaman+bki,  ~ Zaman | id, data = avaps)
joint.fit<- jointModel(lme.fit, cox.fit, timeVar = "Zaman", method ="Cox-PH-GH")
```

Çizelge 2.4'te verilen fonksiyon girişi, temel hazard fonksiyonunun Cox oransal hazard ile kestirimini göstermektedir.

3. BULGULAR

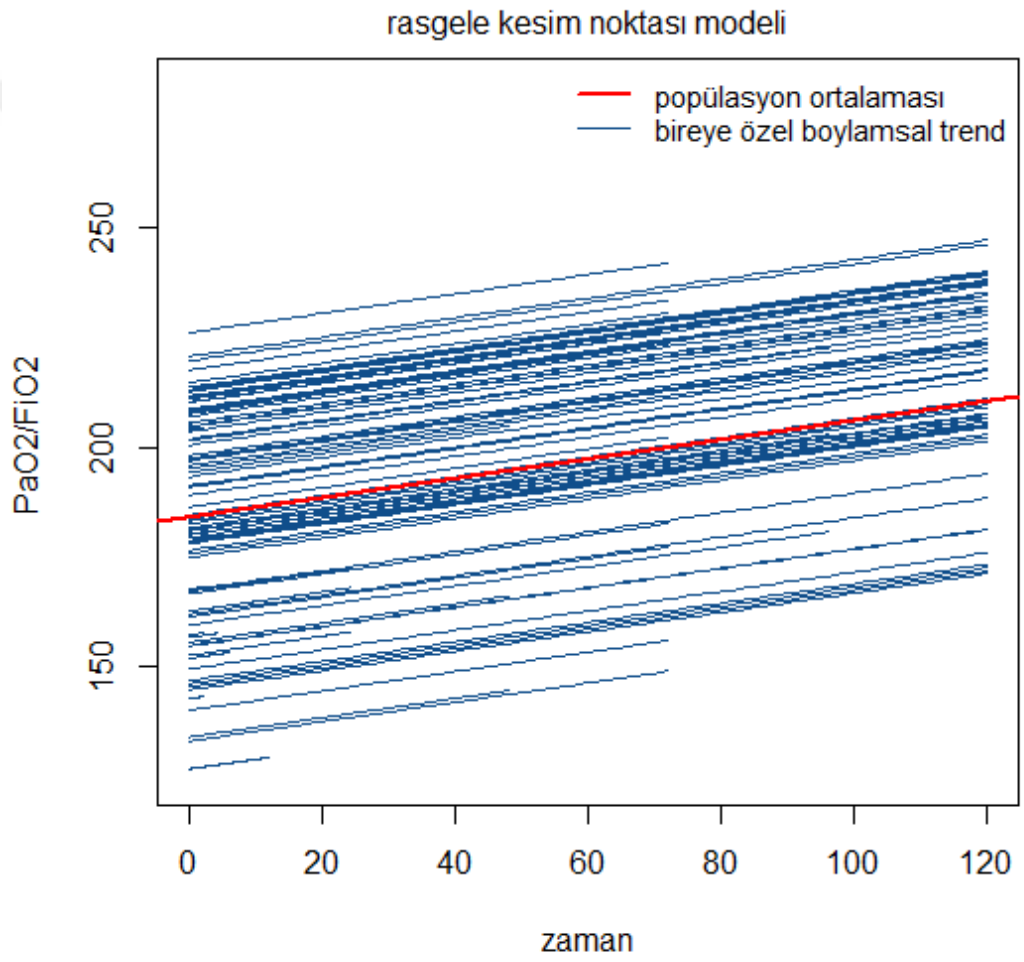
Bu bölümde, bireylere ait PaO_2 / FiO_2 ölçümlerinin zaman içinde gösterdikleri değişimin birleşik modelleme yardımı ile modellenmesi ve bulunan model sonuçları üzerinde durulacaktır. Çizelge 3.1'de çalışma kapsamındaki bireylerin tanımlayıcı istatistikler yer almaktadır.

Çizelge 3.1. Tanımlayıcı istatistikler

	hastalar(n=88)	
	$\bar{X}$**±SS**	**Min-Maks**
bki	27,900±7,906	15,63-55,10
lökosit sayısı	9929,886±4299,743	3600-24900
Glasgow skoru	14,670 ±0,638	12-15
0.Saat PaO_2/FiO_2	202,116±28,085	127-252
1.Saat PaO_2/FiO_2	211,455±27,963	125-260
2.Saat PaO_2/FiO_2	217,273±28,529	125-257
4.Saat PaO_2/FiO_2	222,305±26,647	150-257
6.Saat PaO_2/FiO_2	225,444±26,180	162-262
12.Saat PaO_2/FiO_2	228,180±27,620	150-264
24.Saat PaO_2/FiO_2	231,937±27,406	160-265
48.Saat PaO_2/FiO_2	235,684±26,843	159-270
72.Saat PaO_2/FiO_2	239,671±25,40	150-272
96.Saat PaO_2/FiO_2	243,623±21,392	190-282
120.Saat PaO_2/FiO_2	244,42±22,249	182-280
	Hasta sayısı	
	kah=0	**kah=1**
kah	58	30

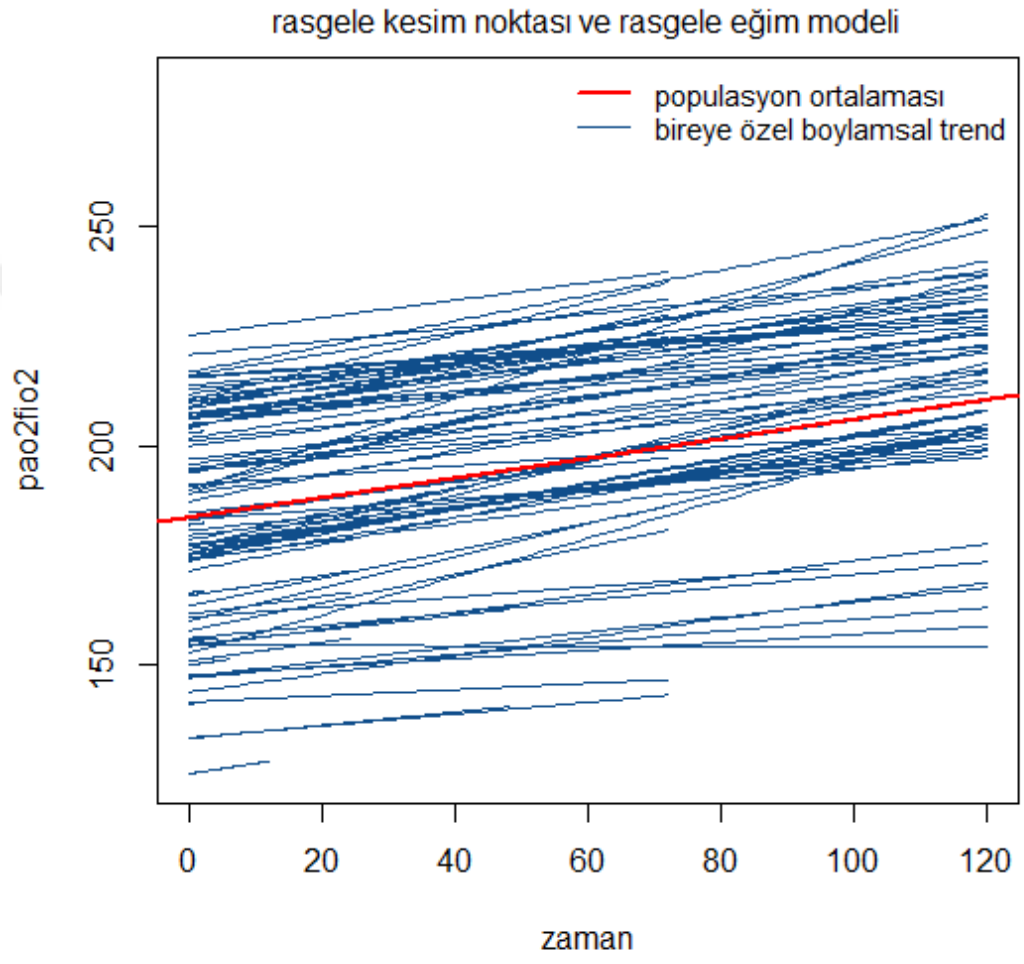
Çizelge 3.2 PaO_2 / FiO_2 değişkeni için oluşturulan birleşik modellerin AIC ve BIC değerleri

	AIC	**BIC**
Model 1	7024,392	7068,984
Model 2	7014,877	7047,082
Model 3	6972,319	6999,569



Şekil 3.1. Rastgele kesim noktası modeli

Şekil 3.1 birey etkisinin rastgele değiştiği ve modele dahil edildiği doğrusal bir modeldir. Kırmızı çizgi popülasyon ortalamasını, mavi çizgiler ise her bir bireyin zaman içindeki yanıt eğilimlerini göstermektedir.



Şekil 3.2. Rastgele kesim noktası ve rastgele eğim modeli

Tekrarlanan ölçüm içeren veri setlerinde genellikle birimlerin başlangıçları birbirine benzeyen zaman noktaları değiştikçe birimler arası değişim artmaktadır. Bu sebeple hem kesim noktası (intercept) hem de eğim (slope) noktalarının şansa bağlı olmasına izin veren bir model oluşturulması gereklidir. Şekil 3.2. kırmızı çizgi rastgele ölçüm

hataları eklendikten sonra popülasyon ortalamasını, mavi çizgiler ise her bir bireyin zaman içindeki yanıt eğilimlerini göstermektedir.

3.1. Birleşik Modellere Ait Sonuçlar ve Model Seçimi

Bölüm 1.9'da da belirtildiği gibi, eldeki veriye en uygun modeli seçebilmek için üç farklı birleşik model kurulmuştur. Bu modellere ilişkin sonuçlar sırasıyla Çizelge 3.2., 3.3. ve 3.4'de yer almaktadır.

3.1.1. Model 1

Çizelge 3.3. Birleşik model sonuçları (1.model)

model 1	açıklayıcı değişken	katsayı	standart hata	%95 GA	p-değeri
boylamsal alt model	sabit terim	182,800	0,003	[182,794; 182,806]	<0,0001
	zaman	0,216	0,017	[0,183; 0,249]	<0,0001
	bki	1,207	0,097	[1,019; 1,396]	<0,0001
sağkalım alt modeli (uyarlanmış parçalı sabit temel risk fonksiyonu)	kah	1,414	0,0008	[1,413; 1,416]	<0,0001
	glasgow	-1,472	0,006	[-1,413; -1,460]	<0,0001
	lökosit	0,0002	1,5E-5	[0,00017; 0,0003]	<0,0001
	Assoct	-0,037	0,0028	[-0,042; -0,031]	<0,0001
	log(xi,1)	18,019	0,0004		
	log(xi,2)	19,086	0,0012		
	log(xi,3)	20,981	0,0002		
	log(xi,4)	20,756	0,0002		
	log(xi,5)	4,052	0,0000		
	log(xi,6)	3,350	0,0000		
	log(xi,7)	21,590	0,0001		

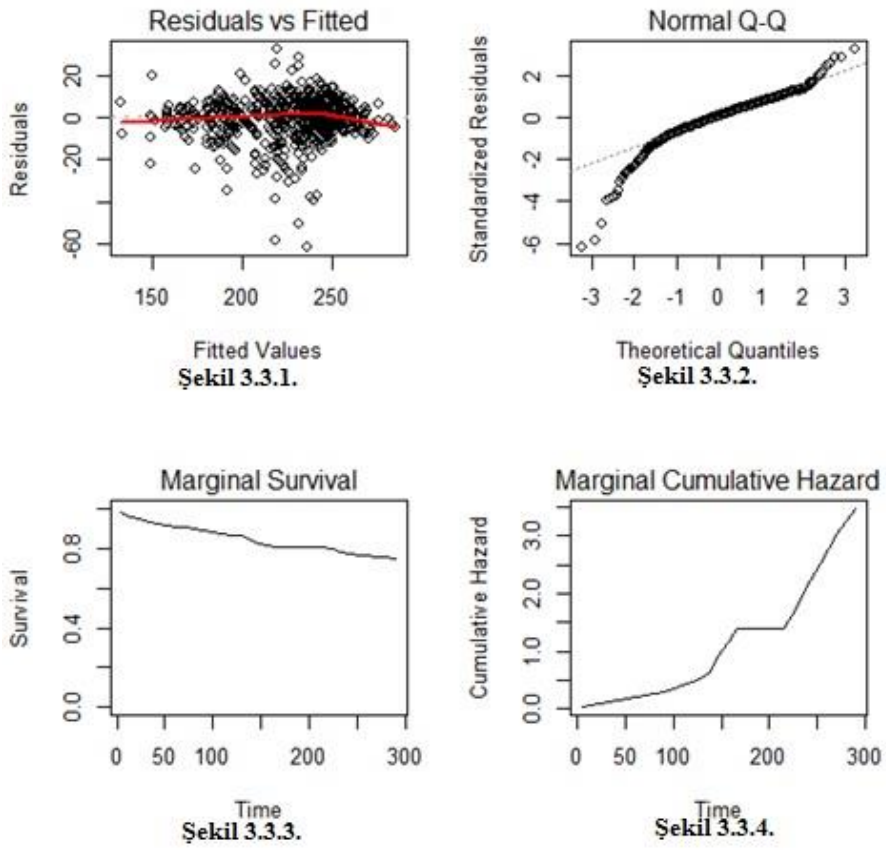
Çizelge 3.3'de boylamsal alt modele ait sabit terim, zaman ve beden kitle indeksi istatistiksel olarak anlamlı bulunmuştur($p<0,0001$). Uyarlanmış parçalı sabit temel risk fonksiyonu altında sağkalım alt modelindeki kah, glasgow ve lökosit

değişkenleri de istatistiksel olarak anlamlı bulunmuştur. (p<0,0001). Ayrıca; Assoct katsayısının istatistiksel olarak anlamlı bulunması PaO_2 / FiO_2 oranı ile ölüm riski arasında güçlü bir ilişki olduğunu gösterir. Değişkendeki bir birimlik azalmanın, ölüm riskinde [exp(-0,037)] =1,038 kat artışla sonuçlanacağını belirtir.

xi.1, . . . , xi.7 parametreleri parçalı-sabit temel risk fonksiyonun parametreleridir ve her biri p<0,0001 ile istatistiksel olarak anlamlı bulunmuştur.

Rastgele etkilerin kestirim değerleri Ek 2'de verilmiştir.

3.1.1.1.Boylamsal Kısım için Artıklar



Şekil 3.3.1.

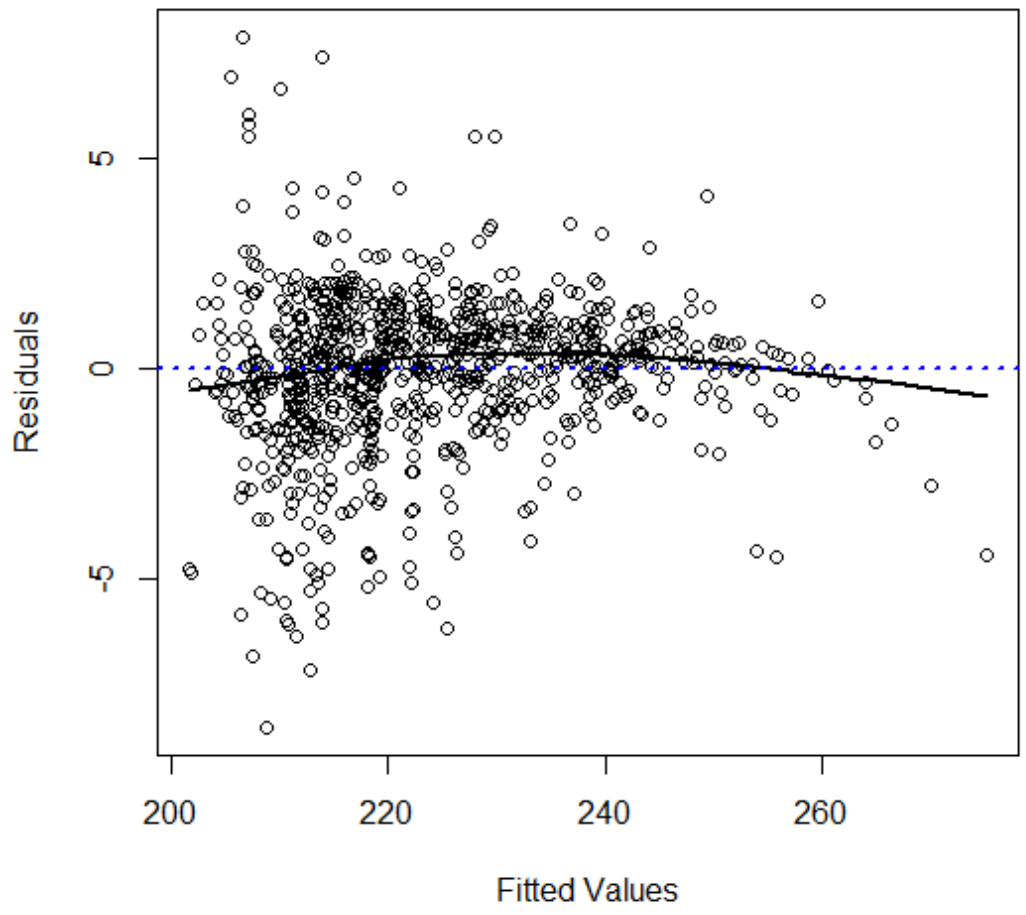
Şekil 3.3.2.

Şekil 3.3.3.

Şekil 3.3.4.

Şekil 3.3. Birleşik Model için Tanısal Grafikler

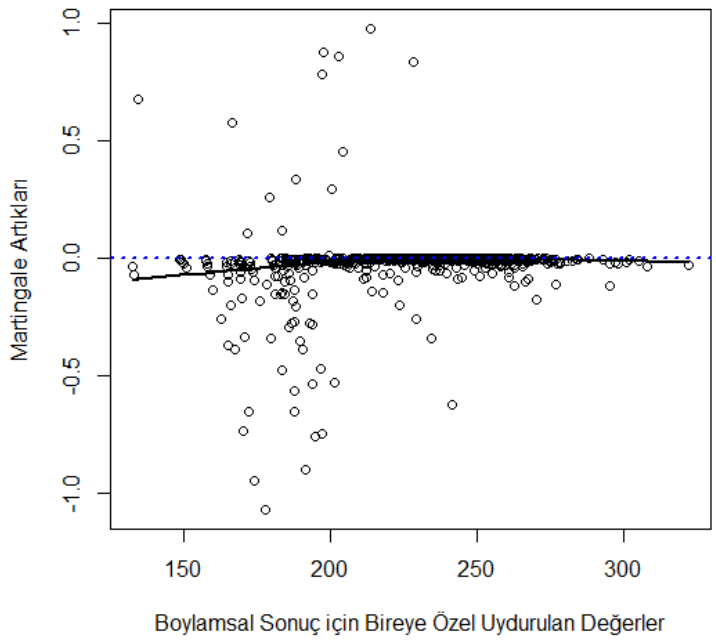
Şekil 3.3.1.'de artıklara karşı modelden elde edilen değerlere bakıldığında koşullu hataların ($\sum = \sigma^2 I_n$) sabit varyanslı olması ve sıfır etrafında dağılması beklenir. Şekil 3.3.2.'de Q-Q grafiğinin doğrusal olması beklenir. Şekil 3.3.3. ve 3.3.4. zamana karşı marjinal sağkalım ve marjinal kümülatif hazard grafikleridir. Şekil 3.3.3.'te marjinal sağkalım olasılığı, her bir zaman noktasındaki sağkalım olasılıkları bir önceki olaya bağlı değildir, yani ilgili zaman noktasındaki olasılık değeri çalışmanın başlangıcı baz alınarak hesaplanır. Şekil 3.3.4.'te kümülatif (birikimli) hazard 0'dan *t* anına kadar olan hazard oranlarının integralidir. Yani, *t* zamanına kadar hesaplanan birikimli olasılıklarının, *t* zamanındaki toplamıdır.



Şekil 3.4. Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler

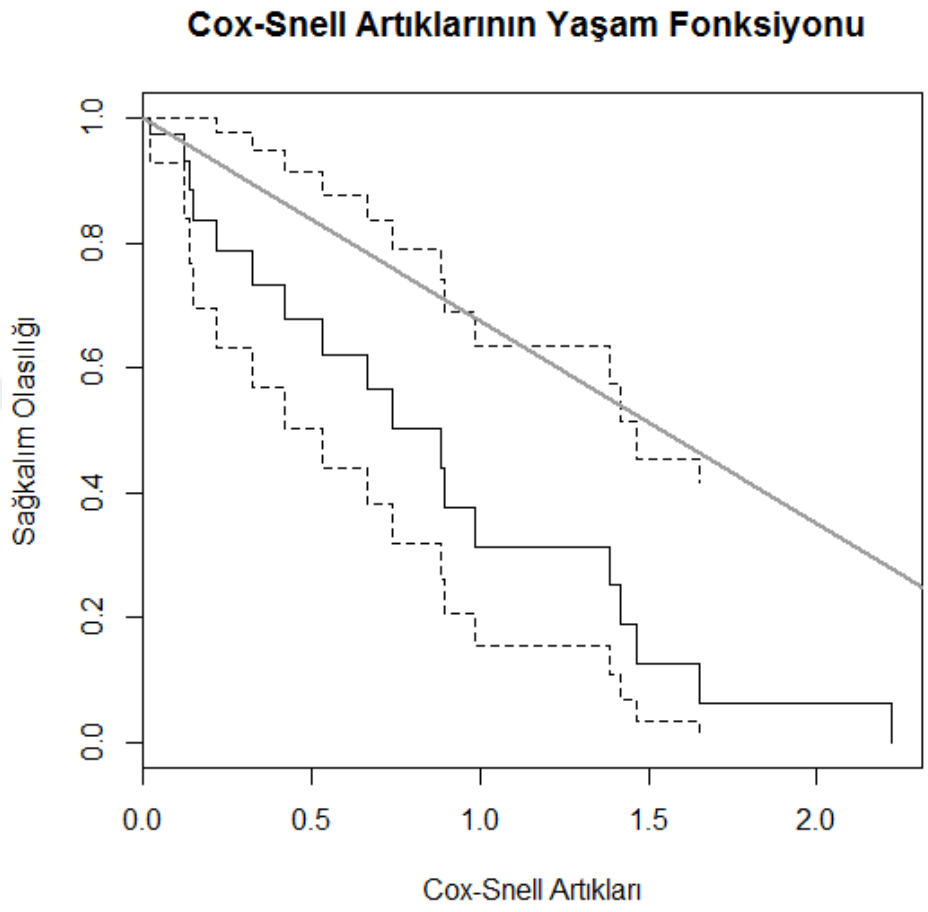
Şekil 3.4'te lowess eğrisinin sıfıra yakın seyrettiği gözlenmektedir. Bu da, kurulan boylamsal modelin ortalama yapısının veriye iyi uyum gösterdiğini işaret etmektedir. Şekilde düşük PaO_2 / FiO_2 kestirim değerlerine karşılık gelen standartlaştırılmış marjinal artıkların daha geniş bir saçılım gösterdiği görülmektedir. Düşük PaO_2 / FiO_2 değerlerine sahip kişilerin sağ kalımının sonlanması nedeniyle çalışmadan çıkmalarına (dropout gözlenmesine) bağlı olup boylamsal modelin ortalama yapısının yetersizliğine atfedilemez (Rizopoulos, 2012b, s:157). Bu nedenle, bağımlı değişken değerlerine bağlı olarak çalışmadan çıkan bireylerin bulunduğu durumlarda, yani göz ardı edilemez kayıp veriler olduğunda; sadece gözlenen veriye dayanan artık grafikleri, birleşik modelin uyumunu incelemede yetersiz kalabilmektedir (Rizopoulos, 2012b, s:157). Ancak tüm kestirim aralığı boyunca bakıldığında, daha önce de belirtildiği gibi, lowess eğrisinin sıfıra yakın seyrettiği ve model ortalama yapısının genel olarak veriye iyi uyum gösterdiği söylenebilir.

3.1.1.2. Sağkalım Kısmı için Artıklar



Şekil 3.5. Martingale artıklarının lowess yöntemi ile oluşturulan eğri grafiği

Şekil 3.5'te uydurulan bazı değerlerin sıfır doğrusu etrafında dağılmadığı görülmekle beraber, lowess eğrisinin sıfırdan sapmadığını söylenebilir.



Şekil 3.6. Veri setinin Cox-Snell artıkları

Şekil 3.6'da siyah doğru parçaları Cox-Snell artıklarının sağkalım fonksiyonunun Kaplan-Meier tahminlerini ifade eder. Kesikli çizgiler, bu tahminlerin güven aralıklarıdır. Gri eğri ise, birim üstel dağılım fonksiyonunu gösterir. Adımsal şekilde inen siyah eğrinin birim üstel dağılımın etrafında seyretmesi beklenir. Sağkalım kısmının uygun olduğu söylenebilir.

3.1.2. Model 2

Çizelge 3.4. Birleşik model sonuçları (2.model)

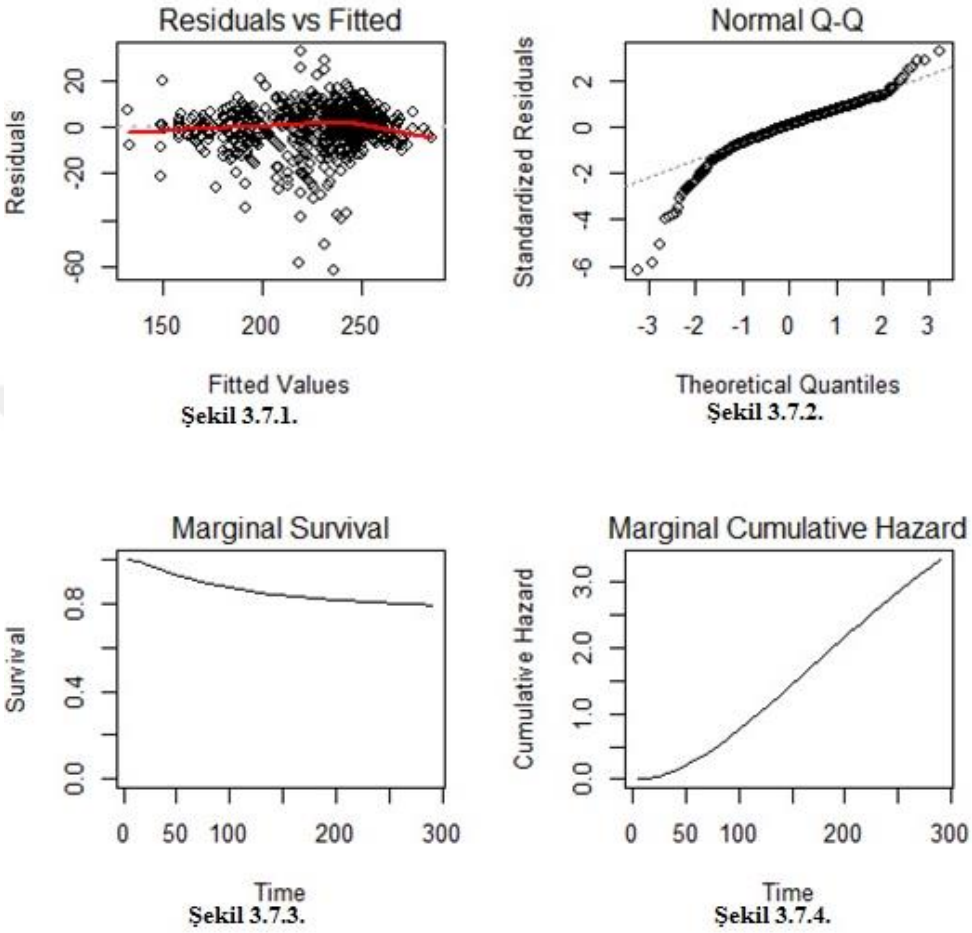
model 2	açıklayıcı değişken	katsayı	standart hata	%95 GA	p-değeri
boylamsal alt model	sabit terim	182,960	10,336	[162,728; 203,244]	<0,0001
	zaman	0,217	0,017	[0,184; 0,250]	<0,0001
	bki	1,201	0,357	[0,502; 1,900]	0,0008
sağkalım alt modeli (uyarlanmış Weibull temel risk fonksiyonu)	sabit terim	17,462	5,504	[6,673; 28,250]	0,0015
	kah	1,617	0,598	[0,443; 2,788]	<0,0001
	glasgow	-1,851	0,468	[-2,767; -0,939]	0,0001
	lökosit	0,0003	0,7E-4	[0,0001;0,0004]	<0,0001
	Assoct	-0,037	0,012	[-0,060; -0,013]	0,0024
	log(shape)	0,776	0,188		<0,0001

Çizelge 3.4'de boylamsal alt modele ait sabit terim, zaman ve beden kitle indeksi istatistiksel olarak anlamlı bulunmuştur. ($p<0,001$). Uyarlanmış Weibull temel risk fonksiyonu altında sağkalım alt modelindeki kah, glasgow ve lökosit değişkenleri de istatistiksel olarak anlamlı bulunmuştur. ($p<0,001$). Ayrıca, Assoct katsayısının istatistiksel olarak anlamlı bulunması PaO_2 / FiO_2 oranı ile ölüm riski arasında güçlü bir ilişki olduğunu gösterir. Değişkendeki bir birimlik azalmanın, ölüm riskinde [exp(-0,037)] =1,38 kat artışla sonuçlanacağını belirtir.

shape katsayısı Weibull dağılımındaki şekil parametresidir. log (shape) katsayısının anlamlılığı da istatistiksel açıdan önemli olarak saptanmıştır. ($p<0,0001$)

Rastgele etkilerin kestirim değerleri Ek 2'de verilmiştir.

3.1.2.1. Boylamsal Kısım için Artıklar



Şekil 3.7.1.

Şekil 3.7.2.

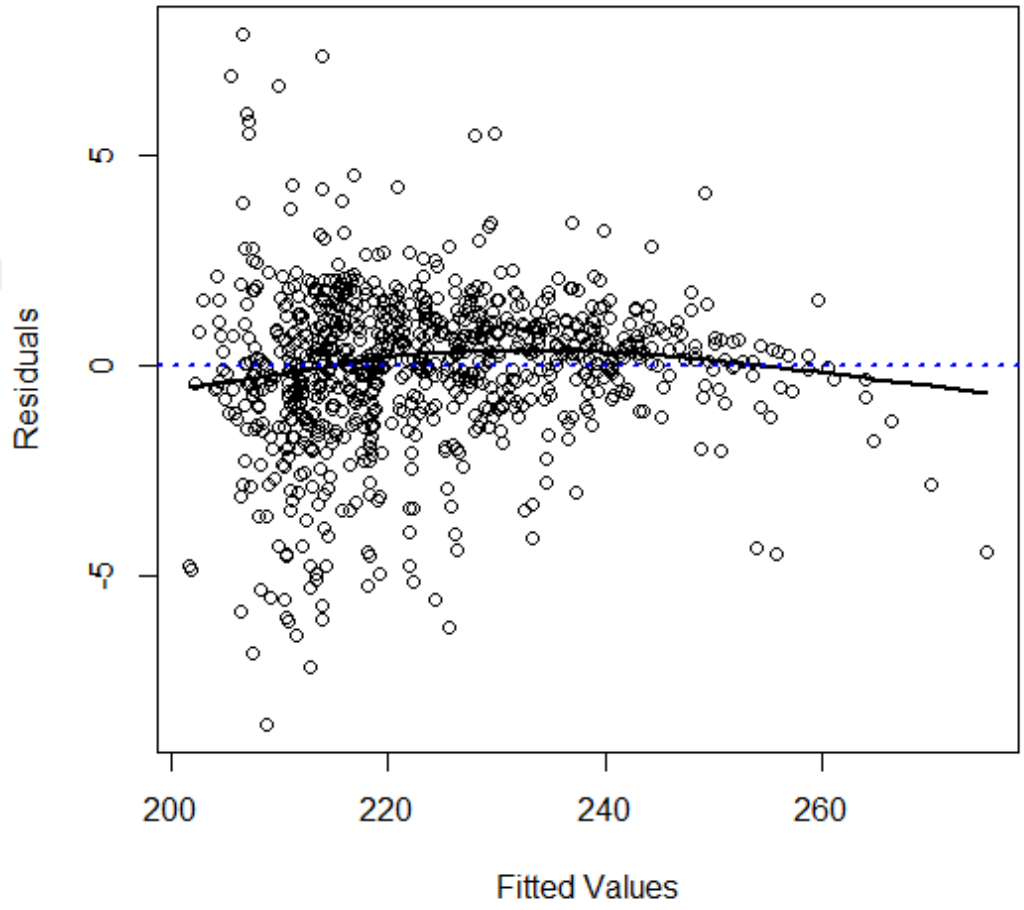
Şekil 3.7.3.

Şekil 3.7.4.

Şekil 3.7. Birleşik Model için Tanısal Grafikler

Şekil 3.7.1'de artıklara karşı modelden elde edilen değerlere bakıldığında koşullu hataların ($\sum = \sigma^2 I_n$) sabit varyanslı olması ve sıfır etrafında dağılması beklenir. Şekil 3.7.2'de Q-Q grafiğinin doğrusal olması beklenir. Şekil 3.7.3. ve 3.7.4. zamana karşı marjinal sağkalım ve marjinal kümülatif hazard grafikleridir. Şekil 3.7.3.'te marjinal sağkalım olasılığı, her bir zaman noktasındaki sağkalım olasılıklarının bir önceki olaya bağlı değildir, yani ilgili zaman noktasındaki olasılık değeri çalışmanın başlangıcı baz alınarak hesaplanır. Şekil 3.7.4.'te kümülatif (birikimli) hazard 0'dan *t*

anına kadar olan hazard oranlarının integralidir. Yani, *t* zamanına kadar hesaplanan birikimli olasılıklarının, *t* zamanındaki toplamıdır.

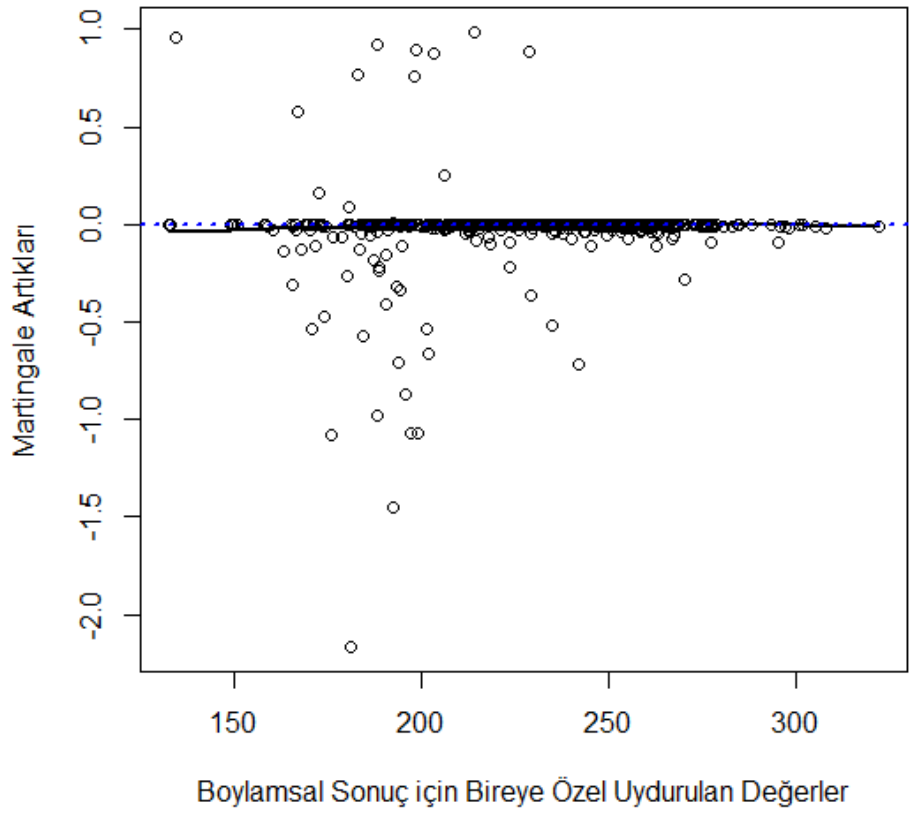


Şekil 3.8. Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler

Şekil 3.8'de lowess eğrisinin sıfıra yakın seyrettiği gözlenmektedir. Bu da, kurulan boylamsal modelin ortalama yapısının veriye iyi uyum gösterdiğini işaret etmektedir. Şekilde düşük PaO_2 / FiO_2 kestirim değerlerine karşılık gelen standartlaştırılmış marjinal artıkların daha geniş bir saçılım gösterdiği görülmektedir. Düşük PaO_2 / FiO_2 değerlerine sahip kişilerin sağ kalımının sonlanması nedeniyle çalışmadan

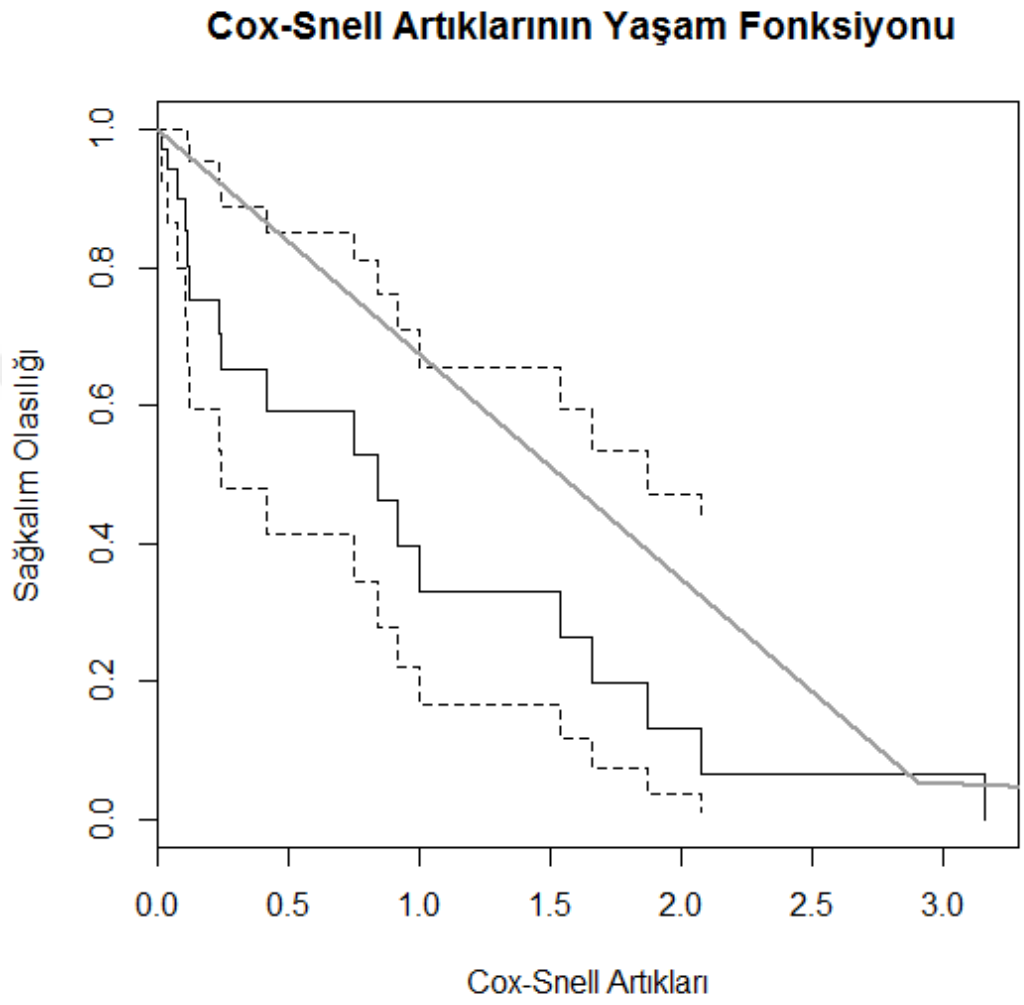
çıkmalarına (dropout gözlenmesine) bağlı olup boylamsal modelin ortalama yapısının yetersizliğine atfedilemez (Rizopoulos, 2012b, s:157). Bu nedenle, bağımlı değişken değerlerine bağlı olarak çalışmadan çıkan bireylerin bulunduğu durumlarda, yani göz ardı edilemez kayıp veriler olduğunda; sadece gözlenen veriye dayanan artık grafikleri, birleşik modelin uyumunu incelemede yetersiz kalabilmektedir (Rizopoulos, 2012b, s:157). Ancak tüm kestirim aralığı boyunca bakıldığında, daha önce de belirtildiği gibi, lowess eğrisinin sıfıra yakın seyrettiği ve model ortalama yapısının genel olarak veriye iyi uyum gösterdiği söylenebilir.

3.1.2.2. Sağkalım Kısmı için Artıklar



Şekil 3.9. Martingale artıklarının lowess yöntemi ile oluşturulan eğri grafiği

Şekil 3.9'da uydurulan bazı değerlerin sıfır doğrusu etrafında dağılmadığı görülmekle beraber, lowess eğrisinin sıfırdan sapmadığını söylenebilir.



Şekil 3.10.Veri setinin Cox-Snell artıkları

Şekil 3.10'da siyah doğru parçaları Cox-Snell artıklarının sağkalım fonksiyonunun Kaplan-Meier tahminlerini ifade eder. Kesikli çizgiler, bu tahminlerin güven aralıklarıdır. Gri eğri ise, birim üstel dağılım fonksiyonunu gösterir. Adımsal şekilde inen siyah eğrinin birim üstel dağılımın etrafında seyretmesi beklenir. Sağkalım kısmının uygun olduğu söylenebilir.

3.1.3. Model 3

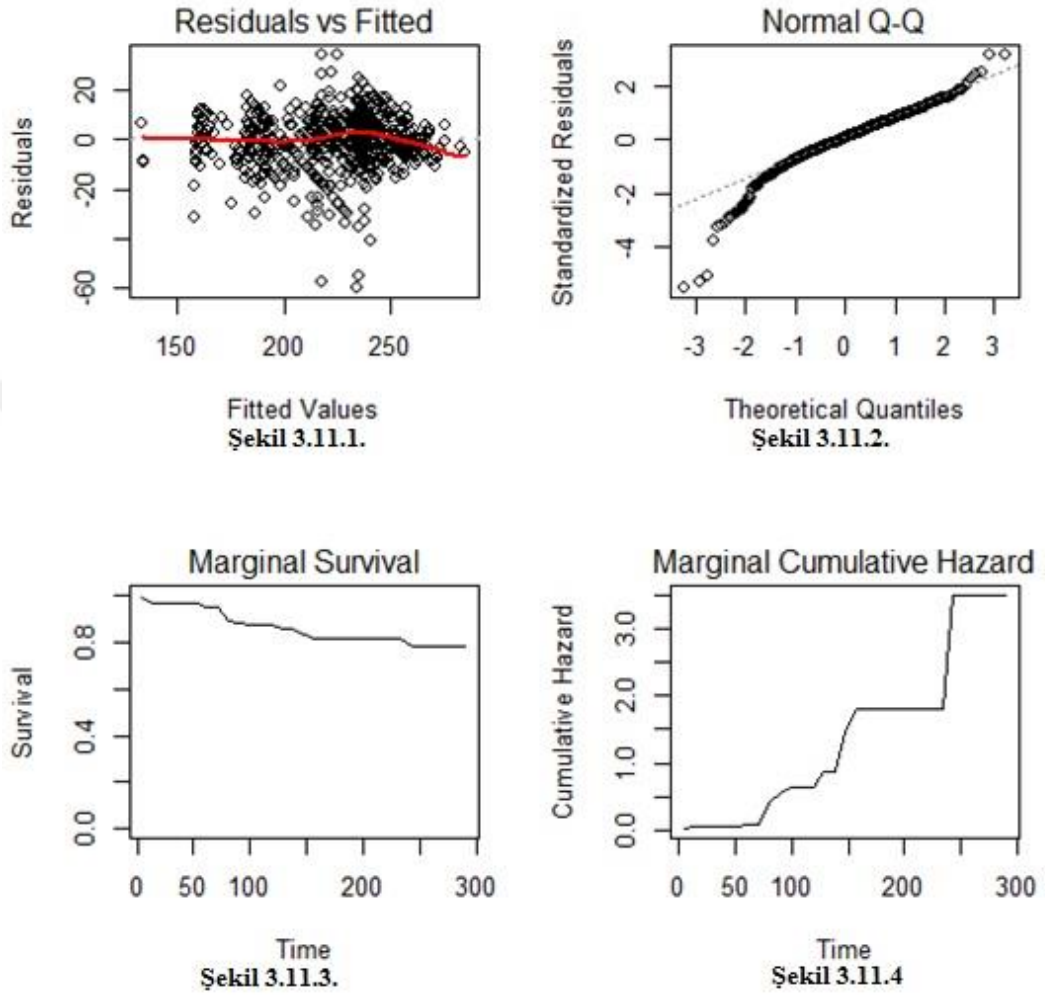
Çizelge 3.5. Birleşik model sonuçları (3.model)

model 3	açıklayıcı değişken	katsayı	standart hata	%95 GA	p-değeri
boylamsal alt model	sabit terim	190,595	2,593	[185,512; 195,677]	<0,0001
	zaman	0,221	0,019	[0,183; 0,259]	<0,0001
sağkalım alt modeli (Cox temel risk fonksiyonu)	bki	0,888	0,082	[0,729; 1,049]	<0,0001
	kah	1,346	0,537	[0,294; 2,398]	0,0121
	glasgow	-1,401	0,177	[-1,749; -1,053]	<0,0001
	lökosit	0,0003	0,0001	[0,00019; 0,00038]	<0,0001
	Assoct	-0,044	0,012	[-0,064; -0,022]	<0,0001

Çizelge 3.5'de boylamsal alt modele ait sabit terim, zaman ve beden kitle indeksi istatistiksel olarak anlamlı bulunmuştur.($p<0,0001$). Cox temel risk fonksiyonu altında sağkalım alt modelindeki kah, glasgow ve lökosit değişkenleri de istatistiksel olarak anlamlı bulunmuştur. ($p<0,05$). Assoct katsayısının istatistiksel olarak anlamlı bulunması PaO_2 / FiO_2 oranı ile ölüm riski arasında güçlü bir ilişki olduğunu gösterir. Değişkendeki bir birimlik azalmanın, ölüm riskinde [exp(-0,044)] =1,045 kat artışla sonuçlanacağını belirtir.

Rastgele etkilerin kestirim değerleri Ek 2'de verilmiştir.

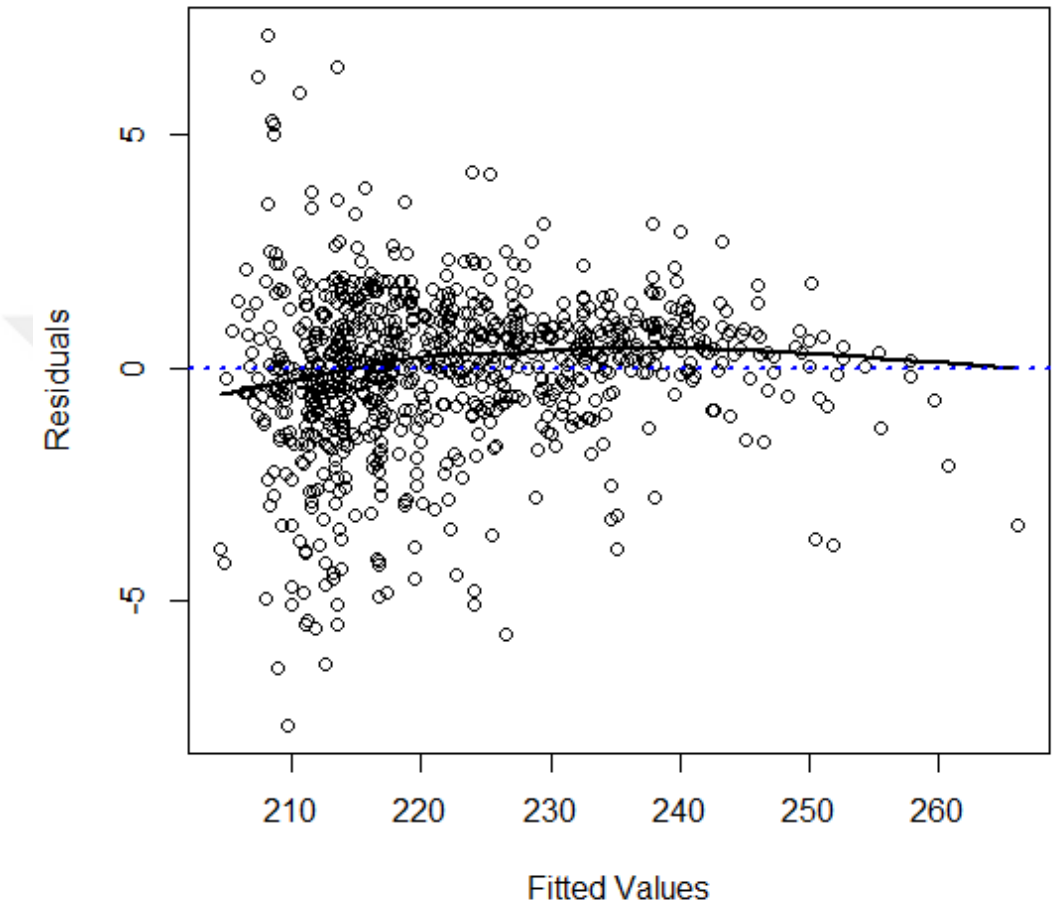
3.1.3.1. Boylamsal Kısım için Artıklar



Şekil 3.11. Birleşik Model için Tanısal Grafikler

Şekil 3.11.1'de artıklara karşı modelden elde edilen değerlere bakıldığında koşullu hataların ($\sum = \sigma^2 I_n$) sabit varyanslı olması ve sıfır etrafında dağılması beklenir. Şekil 3.11.2.'de Q-Q grafiğinin doğrusal olması beklenir. Şekil 3.11.3. ve 3.11.4. zamana karşı marjinal sağkalım ve marjinal kümülatif hazard grafikleridir. Şekil 3.11.3.'te marjinal sağkalım olasılığı, her bir zaman noktasındaki sağkalım olasılıklarının bir önceki olaya bağlı değildir, yani ilgili zaman noktasındaki olasılık değeri çalışmanın başlangıcı baz alınarak hesaplanır. Şekil 3.11.4.'te kümülatif

(birikimli) hazard 0'dan *t* anına kadar olan hazard oranlarının integralidir. Yani, *t* zamanına kadar hesaplanan birikimli olasılıklarının, *t* zamanındaki toplamıdır.



Şekil 3.12. Marjinal standartlaştırılmış artıklara karşı boylamsal sonuç için modelden elde edilmiş değerler

Şekil 3.12'de lowess eğrisinin sıfıra yakın seyrettiği gözlenmektedir. Bu da, kurulan boylamsal modelin ortalama yapısının veriye iyi uyum gösterdiğini işaret etmektedir. Şekilde düşük PaO_2 / FiO_2 kestirim değerlerine karşılık gelen standartlaştırılmış marjinal artıkların daha geniş bir saçılım gösterdiği görülmektedir. Düşük PaO_2 / FiO_2 değerlerine sahip kişilerin sağ kalımının sonlanması nedeniyle çalışmadan çıkmalarına (dropout gözlenmesine) bağlı olup boylamsal modelin ortalama

yapısının yetersizliğine atfedilemez (Rizopoulos, 2012b, s:157). Bu nedenle, bağımlı değişken değerlerine bağlı olarak çalışmadan çıkan bireylerin bulunduğu durumlarda, yani göz ardı edilemez kayıp veriler olduğunda; sadece gözlenen veriye dayanan artık grafikleri, birleşik modelin uyumunu incelemede yetersiz kalabilmektedir (Rizopoulos, 2012b, s:157). Ancak tüm kestirim aralığı boyunca bakıldığında, daha önce de belirtildiği gibi, lowess eğrisinin sıfıra yakın seyrettiği ve model ortalama yapısının genel olarak veriye iyi uyum gösterdiği söylenebilir.

3.2. Ayrı Analiz, İki-Aşama Modeli ve Birleşik Model Sonuçları

Çizelge 3.6'da Ayrı Analiz, İki-Aşama Modeli ve Birleşik Model Sonuçları gösterilmiştir.

Çizelge 3.6. Ayrı Analiz, İki-Aşama Modeli ve Birleşik Model Sonuçları

	Ayrı Analiz				**İki-aşama Modeli**				**Birleşik Modelleme**			
Boylamsal alt model	**Model a.a.** $PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 zaman + \beta_2 bki + b_{0i} + b_{1i} * zaman + \varepsilon_{ij}$				**Model a.c.** $PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 zaman + \beta_2 bki + b_{0i} + b_{1i} * zaman + \varepsilon_{ij}$				**Model a.e.** $PaO_2/FiO_{2_{ij}} = \beta_0 + \beta_1 zaman + \beta_2 bki + b_{0i} + b_{1i} * zaman + \varepsilon_{ij})$			
	Parametre Kestirimi				**Parametre Kestirimi**				**Parametre Kestirimi**			
		katsayı	**Standart hata**	**p-değeri**		**katsayı**	**Standart hata**	**p-değeri**		**katsayı**	**Standart hata**	**p-değeri**
	β_0	182,763	10,408	<0,0001	β_0	182,763	10,408	<0,0001	β_0	190,595	2,593	<0,0001
	β_1	0,222	0,017	<0,0001	β_1	0,222	0,017	<0,0001	β_1	0,221	0,019	<0,0001
	β_2	1,203	0,359	0,0012	β_2	1,203	0,359	0,0012	β_2	0,888	0,082	<0,0001
	σ_{b_0}	26,194			σ_{b_0}	26,194			σ_{b_0}	25,838		
	σ_{b_1}	0,115			σ_{b_1}	0,115			σ_{b_1}	0,114		
	σ_ε	9,950			σ_ε	9,950			σ_ε	9,947		
Sağkalım alt modeli	**Model a.b.** $h(t) = h_0(t)exp(\beta_1 lökosit + \beta_2 glasgow + \beta_3 kah + \beta_4 PaO_2/FiO_2(t))$				**Model a.d.** $h(t) = h_0(t)exp(\beta_1 lökosit + \beta_2 glasgow + \beta_3 kah + \beta_4 \widehat{PaO_2/FiO_2}(t))$				**Model a.f.** $h(t) = h_0(t)exp(\gamma_1 lökosit + \gamma_2 glasgow + \gamma_3 kah + \alpha \widehat{PaO_2/FiO_2}(t))$			
	Parametre Kestirimi				**Parametre Kestirimi**				**Parametre Kestirimi**			
		katsayı	**Standart hata**	**p-değeri**		**katsayı**	**Standart hata**	**p-değeri**		**katsayı**	**Standart hata**	**p-değeri**
	β_1	0,0003	0,00007	<0,001	β_1	0,0003	0,00006	<0,0001	γ_1	0,0003	0,0001	<0,0001
	β_2	-1,408	0,417	<0,001	β_2	-1,957	0,394	<0,0001	γ_2	-1,401	0,177	<0,0001
	β_3	1,291	0,539	0,017	β_3	1,586	0,607	<0,01	γ_3	1,346	0,537	0,0121
	β_4	-0,037	0,009	<0,0001	β_4	-0,107	0,031	<0,001	α	-0,044	0,012	<0,0001

Çizelge 3.6'da Model a.a. boylamsal süreç için kestirilen rasgele ve sabit etkileri içeren karışık etkili modeli ifade etmektedir. Model a.b. ise genişletilmiş Cox regresyon çözümlemesini göstermektedir. Model a.c. Model a.a. ile aynıdır, Model a.d. ise; Model a.c.'de kestirilen PaO_2 / FiO_2 değerinin genişletilmiş Cox regresyon analizinde zamana bağlı açıklayıcı değişken olarak katılmasıyla elde edilmiştir. Model a.e. ve Model a.f. birleşik modellemenin eş zamanlı kestirilen modelleridir. İki-aşama modelinden elde edilen PaO_2 / FiO_2'nin katsayısı yani ilişki katsayısı, birleşik modellemeden elde edilen katsayıdan yüksektir. Bu, aslında beklenen bir durumdur; çünkü iki-aşama modelinde parametre tahminlerinin yanlı olarak elde edildiği bilinmektedir (McCrink, 2013; s:252).

4. TARTIŞMA

Uygulamalı bilim dallarında yapılan araştırmaların çoğunda, tekrarlı ölçüm içeren çalışma tasarımları kullanılır. Tekrarlı ölçümler, aynı denekten farklı zaman noktalarında ya da farklı koşullarda gözlemlerin alınması ile oluşur ve "Boylamsal veriler" olarak da adlandırılır. Bu durumda, "Boylamsal veriler", "tekrarlı ölçümler" in özel bir hali olarak değerlendirilebilir (Ateş, 2010). Sağ kalım analizi, başlangıcı belli olan bir noktadan sonra, bir izlem süresi içindeki gözlemin araştırılmak istenen konuma erişmesi ya da erişmeden önceki özelliklerini sürdürmesi halini inceler (Bilgi, 2009) ve bu izlem süresi sağkalım verilerini oluşturur. Bundan böyle, birleşik modeller, boylamsal verilerle sağkalım verilerini birlikte ele almasından ötürü son zamanlarda bu alanda bilgisayar yazılımlarının da gelişmesi ile oldukça popüler bir yöntem haline gelmiştir.

Bu çalışmada, PaO_2 / FiO_2 ölçümlerinin zaman içindeki değişimlerinin analizi ve sağ kalımı etkileyen faktörler belirlenip için üç çeşit birleşik model kurulmuştur. Bunlar arasından model seçimi AIC ve BIC bilgi kriterlerine göre değerlendirilmiştir.

Birinci model karışık etkili alt modeli rastgele kesim noktası ve eğim modelidir. Bu modelde zaman yer almaktadır. Ayrıca bki açıklayıcı değişkeni de modele dahil edilmiştir. Zaman etkisinin anlamlı olduğu ($p<0,0001$), ayrıca bki'nin de PaO_2 / FiO_2 ölçümleri üzerinde anlamlı bir değişim yaptığı bulunmuştur ($p<0,0001$). Birinci modelin Cox Regresyon ile oluşturulan sağkalım alt modeli; glasgow, kah ve lökosit değişkenlerini içermekte ve istatistiksel olarak her değişken anlamlı bulunmuştur. ($p<0,0001$). Birleşik modelleme aşamasında sağkalım temel alt modeli olarak parçalı-sabit oransal hazard modeli kullanılarak sonuç değişkeni ile ilgilenilen olay arasında anlamlı bir fark bulunmuştur.

İkinci model karışık etkili alt modeli rastgele kesim noktası ve eğim modelidir. Bu modelde zaman yer almaktadır. Ayrıca bki açıklayıcı değişkeni de modele dahil

edilmiştir. Zaman etkisinin anlamlı olduğu ($p<0,0001$), ayrıca bki'nin de PaO_2 / FiO_2 ölçümleri üzerinde anlamlı bir değişim yaptığı bulunmuştur ($p<0,0001$). İkinci modelin Cox Regresyon ile oluşturulan sağkalım alt modeli; glasgow, kah ve lökosit değişkenlerini içermekte ve istatistiksel olarak her değişken anlamlı bulunmuştur. ($p<0,001$). Birleşik modelleme aşamasında sağkalım temel alt modeli olarak weibull oransal hazard modeli kullanılarak sonuç değişkeni ile ilgilenilen olay arasında anlamlı bir fark bulunmuştur.

Üçüncü model karışık etkili alt modeli rastgele kesim noktası ve eğim modelidir. Bu modelde zaman yer almaktadır. Ayrıca bki açıklayıcı değişkeni de modele dahil edilmiştir. Zaman etkisinin anlamlı olduğu ($p<0,0001$), ayrıca bki'nin de PaO_2 / FiO_2 ölçümleri üzerinde anlamlı bir değişim yaptığı bulunmuştur ($p<0,0001$). Üçüncü modelin Cox Regresyon ile oluşturulan sağkalım alt modeli; glasgow, kah ve lökosit değişkenlerini içermekte ve istatistiksel olarak her değişken anlamlı bulunmuştur ($p<0,51$). Birleşik modelleme aşamasında sağkalım temel alt modeli olarak Cox oransal hazard modeli kullanılarak sonuç değişkeni ile ilgilenilen olay arasında anlamlı bir fark bulunmuştur.

Verinin birleşik modellenmesinde sağkalım temel alt modeli olarak parçalı-sabit oransal hazard modeli, weibull oransal hazard modeli ve Cox oransal hazard modeli kullanılmıştır. Veriye en uygun birleşik modelin, AIC ve BIC bilgi kriterlerine göre üçüncü model olduğu söylenebilir.

Ek olarak, boylamsal kısım için marjinal artıklar, sağkalım kısmı için Cox-Snell ve Martingale artıklar analizi kullanılarak modelin veriye uygunluğu incelenmiştir.

Tez kapsamında, çalışma verilerindeki boylamsal süreç ve sağkalım süreci birleşik modelleme yöntemi dışında ayrıca ayrı ayrı analiz edilerek ve iki aşamalı çözüm yaklaşımı ile de değerlendirilmiştir. Ancak, literatürde bu iki yaklaşımın birleşik modellemeye göre yanlı sonuçlar verdiğine ilişkin simülasyon çalışmaları da bulunmaktadır (Ibrahim ve ark., 2010; Sweeting ve Sweeting ve Thompson, 2011).

5. SONUÇ VE ÖNERİLER

Çalışmada birleşik modellemeye bir giriş yapılarak birleşik model ile ilgili temel kavramlar açıklanmıştır. Bu modelin uygulanması için gereken analizlerden biri boylamsal veriler için karışık etkili model, diğeri sağkalım verileri için Cox regresyon analizidir. Bu iki analiz birlikte birleşik modellemeye alınır ve her iki kısım için kestirim değerleri rastgele etkilerin dahil edilmesiyle elde edilir.

Sağkalım verileri ve boylamsal veriler aynı çalışma kapsamında toplansa da bu iki tip veri farklı istatistiksel yöntemler kullanılarak ayrı ayrı analiz edilir. Bu yöntemler tek tek ayrı olarak uygulandığında birleşik modellemeye göre daha az güvenilir sonuçlar elde edilebilir. Bireylere ilişkin sağkalım süreci ile bu süre içinde boylamsal olarak toplanan verilerin ilişki olduğu durumlarda birleşik bir model kullanılması önerilir.

Birleşik modeller arasından en uygun modeller araştırılmış ve AIC ve BIC için en küçük değerleri baz alınmıştır. Üç farklı yöntem ile seçilen modeller, daha sonra artık analizi yapılarak grafiksel uyumları araştırılmış ve üçüncü modelin uygun olduğu görülmüştür.

Genişletilmiş Cox modelinin birleşik modele alternatif olamamasının sebebi sadece dışsal zamana bağlı açıklayıcı değişkenler için uygun olması ve boylamsal biyobelirteçlerde kullanılamamasıdır. Bu ifadenin geçerliğini araştırmak adına simülasyon çalışması ile genişletilmiş Cox modeli ve birleşik model sonuçları karşılaştırabilir.

ÖZET

Sağkalım Verileri ile Boylamsal Verilerin Birleşik Modellenmesi ve Sağlık Alanına bir Uygulama

Boylamsal veriler, aynı bireyden farklı zaman ya da koşullar altında tekrarlı olarak yapılan gözlem ya da ölçümler sonucu elde edilen verileri temsil eder. Sağkalım verileri ise belli bir izlem süresi boyunca ilgilenilen olayı (ölüm, hastalığın nüks etmesi, organ nakli olması, vb.) ve görülme zamanını kapsar. Geleneksel olarak aynı çalışma kapsamında toplanan bu iki tip veri farklı istatistiksel yöntemler kullanılarak incelenir. Ancak son zamanda bu iki farklı veri tipinin birlikte modellenmesi üzerinde çalışmalar yapılmaktadır. Toplanan bu tip verilerinden doğru sonuçlar elde edebilmek için birleşik modelleme yönteminin kullanılması gerekliliği yapılan çalışmalar sonucu ortaya konmuştur.

Bu tez çalışmasında, boylamsal veriler ile sağkalım verilerinin birleşik modellenmesi üzerine geliştirilen yöntemler incelenecek, sağlık alanından bir veri seti üzerinde birleşik modelleme yöntemi uygulanarak, modelin veriye uyumu test edilerek veriye en uygun birleşik model seçimi yapılmıştır. Ayrıca aynı veri setindeki sağkalım verisi sağkalım analizi ile, boylamsal veri kısmı da boylamsal veri analizi ile ayrı ayrı modellenmiştir. Hem birleşik modelden, hem de ayrı ayrı modellemeden elde edilen sonuçlar yorumlanmıştır. Veri analizi için açık kaynak kodlu R yazılımı kullanılmıştır.

Anahtar Kelimeler: Birleşik Modelleme, Boylamsal Veri Analizi, Karışık Etki, Rastgele Etki, Sağkalım Analizi

SUMMARY

Joint Modeling of Survival and Longitudinal Non-Survival Data with an Application to Medical Data

A longitudinal study refers to an investigation where observations or measurements are collected at multiple follow-up times. Survival data includes events (death, recurrence of a disease or transplantation etc.) for a certain follow-up period and the survival time. Traditionally, these two data types collected for the same study purpose are analyzed with different statistical methods. However, recently these two different types of data are being studied modeling with jointly. To obtain accurate results from these kinds of data, further studies showed the necessity of joint modeling.

In this thesis, the methods developed for modeling survival and longitudinal data jointly were examined and, joint modeling techniques were applied to a medical data. After testing whether the model fits the data well, most appropriate joint model was selected. Also, for the same data set, survival data will be modeled with survival analysis and longitudinal data was modeled with longitudinal data analysis separately. Both the results obtained from the joint model and the results obtained from separately modeled before were reviewed. For data analysis, open source R software will be used.

Keywords: Joint Modeling, Longitudinal Data Analysis, Mixed Effects, Random Effects, Survival Analysis

KAYNAKLAR

1- ANDERSEN P, BORGAN O, GILL R, KEIDING N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

2- Anonim, Postestimation tools for stcox, Erişim: [http://www.statapress.com/manuals/errata/stata11/i/stcox_postestimation.pdf]. Erişim Tarihi: 25.10.2016.

3- ATA N, KARASOY D, SÖZER MT (2007). Orantılı Tehlike Varsayımının İncelenmesinde Kullanılan Yöntemler ve Bir Uygulama. *Eskişehir Osmangazi Üniversitesi Müh. Mim. Fak.Dergisi,* **20** (1), 57-80.

4- ATA N, KARASOY D. (2011). Sağkalım Çözümlemesi için Zayıflık Modeli ve Mide Kanseri Hastalarına İlişkin Verilerle Bir Uygulama, *Çankaya University Journal of Science and Engineering*, **8** (2) 225-235.

5- ATA N, SÖZER, MT (2007). Cox Regression Models With Nonproportional Hazards Applied To Lung Cancer Survival Data, *Hacettepe Journal of Mathematical and Statistics*, **36** (2), 157-167.

6- ATA N. (2010). Orantısız Tehlikeler için Yaşam Modelleri. *Doktora Tezi*, Hacettepe Üniversitesi, Fen Bilimleri Ensititüsü, Ankara.

7- ATA TUTKUN N. (2015). Eğitim Bilimlerinde Olay Geçmişi Çözümlemesi. *Journal of International Social Research*, **8** (39), 570-582.

8- ATEŞ C. (2010). Faktöriyel tasarımlarda uzunlamasına veriler için parametrik olmayan analiz. *Yüksek Lisans Tezi,* Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

9- BİLGİ, S. (2009). Sağ Kalım Analizinde Kullanılan İstatistiksel Yöntemler Ve Aktüerya Alanında Bir Uygulama. *Yüksek Lisans Tezi,* Ege Üniversitesi, Fen Bilimleri Enstitüsü, İzmir.

10- BORUCKA J.(2010). Assesment of Cox Proportional Hazard Model Adequancy Using PROC PHREG and PROC GPLOT, *PhUSE 2010 Berlin Conference*, 17-20 October, Berlin, Germany, p. SP05.

11- BRADLEY J, ROGERS W, FISHER L D, GERSH B, COGGIN CG, MYERS WO (1992). Effects of smoking on survival and morbidity in patients randomized to medical and surgical therapy in the Coronary Artery Surgery Study (CASS): 10 year follow-up. *J Am Coll Cardiol*, **20,** 287-94.

12- BROWN E, IBRAHIM J, DEGRUTTOLA V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics,* **61,** 64 – 73.

13- CHEN H, HUANG Y, ZHANG N. (2015). Joint modeling of a linear mixed effects model for selfesteem from mean ages 13 to 22 and a generalized linear model for anxiety disorder at mean age 33. *Journal of Medical Statistics and Informatics*, **3**(1), 1.

14- CHI YY, IBRAHIM JG (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, **62**: 432–445.

15- COLLETT D (1994). *Modelling Survival Data in Medical Research*, Chapman&Hall/CRC.

16- COX D R, SNELL E J (1968). A general definition of residuals, *Journal of the Royal Statistical Society*, **30**, 2, 248-275.

17- COX D, HINKLEY D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

18- COX D. (1972). Regression models and life-tables (with discussion). *J Roy Stat Soc B.;* **34**: 187–220.

19- CROWTHER MJ, ABRAMS KR, LAMBERT PC. (2013). Joint modeling of longitudinal and survival data. *The Stata Journal*, **13**(1): 165-184.

20- DAHER ABDİ Z, ESSIG M, RIZOPOULOS D, LE MEUR Y, PREMAUD A, WOILLARD JB, ET AL.(2013). Impact of longitudinal exposure to mycophe- nolic acid on acute rejection in renal-transplant recipients using a joint modeling approach. *Pharmacol Res.* **72**: 52–60.

21- DEGRUTTOLA V, TU XM (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003-14.

22- DEMPSTER A, LAIRD N, RUBIN D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series,* **39,** 1 – 38.

23- DIGGLE PJ, SOUSA I, CHETWYND, AG. (2008). Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture. *Stat. Med.*, **27**(16), 2981–2998.

24- DING J, WANG JL (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics*, **64,** 546 – 556.

25- DOĞANAY, B. (2007). Uzunlamasına Çalışmaların Analizinde Karma Etki Modelleri. *Yüksek Lisans Tezi*, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

26- ELASHOFF, R M, LI G, LI N. (2007). An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in medicine*, **26**(14), 2813-2835.

27- ER, E. (2010). *Banka başarısızlıklarının* sağkalım *analiziyle* açıklanması*:* Türkiye örneği *(*1989–2008*). Yüksek Lisans Tezi,* Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

28- ERDEM, MT (2015). Joint model for recurrent event times and longitudinal covariates. *Doktora Tezi* (Yayımlanmamış), O.D.T.Ü. Fen Bilimleri Enstitüsü, Ankara.

29- FAUCETT, C J, THOMAS, D C (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statist. Medicine*, **15**, 1663-1685.

30- FERRER L, RONDEAU V, DIGNAM J, PICKLES T, JACQMIN-GADDA H, PROUST-LIMA C (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine*, **35**(22), 3933–3948

31- FITZMAURICE G, LAIRD N, WARE J. (2004). Applied Longitudinal Analysis. Wiley, Hoboken.

32- FLEMING T, HARRINGTON D. (1991). Counting Processes and Survival Analysis. Wiley, New York.

33- FOURNIER M C, FOUCHER Y, BLANCHE P, BURON F, GIRAL M, DANTAN E (2016). A joint model for longitudinal and time-to-event data to better assess the specific role of donor and recipient factors on long-term kidney transplantation outcomes. *European journal of epidemiology*, 1-11.

34- FOX J (2002). Cox Proportional-Hazards Regression for Survival Data, Appendix to an R and S-PLUS Companion to Applied Regression.

35- FU R, GILBERT PB (2016). Joint modeling of longitudinal and survival data with the Cox model and two-phase sampling. Lifetime Data Anal. [Epub ahead of print] *PubMed*.

36- GHARIBVAND L, JESKE DR, LIAO S. (2008). Evaluation of a Hospice Care Referral Program Using Cox Proportional Hazards Model, *Western Users of SAS Software Conference*, Universal City, CA.

37- GHOSH, D. (2000). Nonparametric and Semiparametric Analysis of Recurrent Events in the Presence of Terminal Events and Dependent Censoring, Phd Dissertation, Washington University, United States.

38- HARRELL, F. (2001). Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer-Verlag, New York.

39- HENDERSON R, DIGGLE P, DOBSON A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465-480.

40- HERNDON J, HARRELL F. (1996). The restricted cubic spline hazard model. *Communications in Statistics – Theory and Methods* **19,** 639 – 663.

41- HOGAN JW, LAIRD NM (1997). Mixture models for the joint distributions of repeated measures and event times. *Statistics in Medicine* **16,** 239-257.

42- HSIEH F, TSENG Y, WANG J. (2006). Joint modelling of survival and longitudinal data likelihood approach revisited. *Biometrics,* **62**, 1037–1043.

43- IBRAHİM JG, CHU H, CHEN LM .(2010). Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol.* **28**:2796–801.

44- KARASOY D, TUTKUN, AN (2016). Yaşam Çözümlemesi. Ankara: Nobel Yayın Dağıtım.

45- KHOUNDABI B, KAZEMNEJAD A, MANSOURIAN M, KAZEMPOOR DIZAJI M, HASHEMIAN SM. (2015). Acute Kidney Injury in ICU Patients Following Non-Cardiac Surgery at Masih Daneshvari Hospital: Joint Modeling Application. *Tanaffos,* **14**(1), 49-54.

46- KLEIN JP, MOESCHBERGER M L. (2003). Survival Analysis Techniques for Censored and Truncated Data. New York: Springer-Verlag Inc.

47- KLEIN JP, MOESCHBERGER, ML (1997). Multivariate survival analysis. Survival Analysis Techniques for Censored and Truncated Data, Springer, New York, 405-422.

48- KLEINBAUM D G, KLEIN M. (2006). *Survival analysis: a self-learning text*. Springer Science & Business Media.

49- LEE E, WANG J.(2003). Statistical Methods for Survival Data Analysis, Wiley & Sons, Haboken, New Jersey.

50- LIN D Y, WEI, L J, YING Z (1993). Checking the Cox model with cumulative sums of martingale-based residuals, *Biometrika*, 80, 3, 557-572.

51- LITTLE, R J A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.*, **88**(421), 125-134.

52- MARTINS R, SILVA GL, ANDREOZZI V. (2016). Bayesian joint modeling of longitudinal and spatial survival AIDS data. Stat Med. [Epub ahead of print].

53- MCCRINK L M, MARSHALL A H, CAIRNS K J (2013). Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *International Statistical Review*, **81**(2), 249-269.

54- NGUTI R, BURZYKOWSKİ T, ROWLANDS J, RENARD D, JANSSEN P. (2005). Joint modelling of repeated measurements and event time: Application to performance traits and survival of lambs bred in sub-humid tropics. *Genetics Selection Evolution*, **37**(2), 175-197.

55- ÖZKARTAL, B. (2009). Yinelemeli Olay Analizi ve Çocuk Suçluluğu Üzerine Bir Uygulama, *Yüksek Lisans Tezi*, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

56- PETERSON MD, ZHANG P, DUCHOWNY KA, MARKIDES KS, OTTENBACHER KJ, SNIH SA (2016). Declines in Strength and Mortality Risk among Older Mexican Americans: Joint Modeling of Survival and Longitudinal Data. *J Gerontol A Biol Sci Med Sci*. [Epub ahead of print].

57- PINHEIRO J, BATES D. (2000). Mixed-Effects Models in S and S-PLUS. Springer-Verlag, New York.

58- PIULACHS X, ALEMANY R, GUILLEN M. (2016). Joint Modelling of Survival and Emergency Medical Care Usage in Spanish Insureds Aged 65+. *PloS one*, **11**(4).

59- PRESS W, TEUKOLSKY S, VETTERLING W, FLANNERY B. (2007). Numerical Recipes: The Art of Scientific Computing, 3rd edition. Cambridge University Press, New York.

60- R CORE TEAM (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Erişim: https://www.R-project.org/. Erişim Tarihi: 25.03. 2016.

61- RIZOPOULOS D, VERBEKE G, LESAFFRE E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B,* **71,** 637 – 654.

62- RIZOPOULOS D, VERBEKE G, MOLENBERGHS G. [3.08.2010]. Joint Modeling Approaches in Longitudinal Studies Using Random Effects. Erişim: [https://perswww.kuleuven.be/~u0018341/documents/ldasc10jsm-CursusVersie.pdf]. Erişim Tarihi: 24.09.2016.

63- RIZOPOULOS D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time to-event data. *Biometrics*, **67**, 819–829.

64- RIZOPOULOS D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian çeyreklik rule. *Computational Statistics & Data Analysis*, **56**(3), 491-501.

65- RIZOPOULOS D. (2012b). Joint Models for Longitudinal and Time-to-Event Data: with Applications in R. Boca Raton: Chapman and Hall/CRC.

66- RIZOPOULOS D. (2016). The R package JMbayes for fitting joint models for longitudinal and timeto-event data using MCMC. *Journal of Statistical Software*, **72**(7), 1–45

67- RIZOPOULOS D. GHOSH, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine,* **30,** 1366 – 1380.

68- RIZOPOULOS D. JM (2010). An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, **35**, 1–33.

69- ROSENBERG, P. (1995). Hazard function estimation using B-splines. *Biometrics* **51,** 874 – 887.

70- RUPPERT D, WAND M, CARROLL R. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.

71- SCHLUCHTER MD. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, **11**, 1861-1870.

72- SELF S, PAWİTAN Y. (1992). Modeling a marker of disease progression and onset of disease. *AIDS epidemiology*, 231-255, Birkhäuser Boston.

73- SONG X, DAVIDIAN M, TSIATIS A. (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, **58,** 742-753.

74- SOUSA I. (2011). A review on joint modelling of longitudinal measurements and time-to-event. *REVSTAT–Statistical Journal*, **9**(1), 57-81.

75- STEPANOVA M, THOMAS L. (2002). Survival analysis methods for personal loan data, *Operations Research*, **50**(2) 277-289.

76- SWEETING M J, THOMPSON S G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, **53**(5), 750-763.

77- TABLEMAN M, KIM JS. (2004). *Survival Analysis Using S: Analysis of Time-to-Event Data*, Chapman & Hall/CRC.

78- THERNEAU T M, GRAMBSCH P M. (2000). Modelling Survival Data: Extending the Cox model, Springer, New York.

79- TINTO V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research, *Review of Educational Research*, **45(**1), 89-125.

80- TOPÇU Ç, ARSLAN F. (2009). Sağdan sansürlemenin sağkalım analizindeki etkisi. 6. Ulusal İstatistik Kongresi, Side-Antalya. Bildiri Özetleri Kitabı: 214-221.

81- TSENG Y K, HSIEH F, WANG, J L (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, **92**(3), 587-603.

82- TSIATIS A A, DEGRUTTOLA V, WULFSOHN M S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, **90**, 27-37.

83- TSIATIS A, DAVIDIAN M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal and time-to-event data: An overview: *Statistica Sinica*, **14**, 143-163.

84- TSIATIS A, DAVIDIAN M. (2004). An overview of joint modeling of longitudinal and time-to-event data, *Statistica Sinica*, **14**, 809–834.

85- TUNCER N. (2014). Yaşam Çözümlemesinde Aykırı Değerler. *Yüksek Lisans Tezi,* Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

86- ÜNLÜ, H. (2013). Bilgi İçeren Durdurma Varlığında Yinelemeli Olay Süreci. *Doktora tezi*. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

87- VERBEKE G, MOLENBERGHS G. (2000). Linear Mixed Models for Longitudinal Data. New York John Wiley Sons.

88- VITTINGHOFF E, SHIBOSKI SC, GLIDDEN DV, MCCULLOCH CE (2005), Regression Methods in Biostatistics: Linear, Logistics, Survival and Repeated Measrures Models, Springer Science&Business Media Inc.

89- WANG Y, TAYLOR J M G (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. Amer. Statist. Assoc.*, **96**, 895-905.

90- WHITTEMORE A, KILLER J. (1986). Survival estimation using splines. *Biometrics,* **42,** 495–506.

91- WULFSOHN M S, TSIATIS A A (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330-339.

92- XU J, ZEGER SL. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *J R Stat Soc SerC (Appl Stat),* **50**, 375–387.

93- YAY M, ÇOKER, E, UYSAL Ö. (2007). Yaşam analizinde Cox regresyon modeli ve artıkların incelenmesi, *Cerrahpaşa Tıp Dergisi*, **38**, 139–145.

94- YE W, LIN X, TAYLOR J. (2008). A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface,* **1,** 33 – 45.

95- ZENG D, CAI J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, **33**, 2132-2163.

96- ZHU H, DESANTIS S M, LUO S. (2016). Joint modeling of longitudinal zero-inflated count and time-to-event data: A Bayesian perspective. *Statistical Methods in Medical Research*.

EKLER

Ek-1. EM Algoritması

EM algoritması, tanımlanmamış veri problemleri için en çok olabilirlik tahmini için sık kullanılan iteratif bir algoritmadır.

EM algoritmasının ardındaki fikir, tamamlanmış veride log-olabilirliğin maksimize edilmesidir ve genellikle kapalı form şekilde daha kolay uygulanabilir.

Algoritma E ve M adımı olarak iki adımdan oluşur. E-adımda kayıp veri doldurulur ve gözlenen verinin log olabilirliği vekil bir fonksiyonla yerine konur. Vekil fonksiyon M-adımda maksimize edilir.

EM algoritması şu şekildedir; Y tamamlanmış veri vektörü, Y_0 gözlenen kısım ve Y_m kayıp kısımdan oluşsun. Amaç sadece gözlenen bilgiyi kullanarak, tamamlanmış veri modelinin θ parametresini tahmin etmektir,

$$Q\left(\theta \mid \theta^{(it)}\right) = E\left\{\log p(y;\theta) \mid y^0;\ \theta^{(it)}\right\}$$
$$= \int logp(y^m, y^0;\ \theta)p\left(y^m y^0;\mid \theta^{(it)}\right) dy^m,$$

ve M-adımda güncelleşmiş parametreleri

$$\theta^{(it+1)} = \underset{\theta}{\operatorname{argmax}}\, Q\left(\theta \mid \theta^{(it)}\right).$$

ile elde edilir.

EM algoritmasının en büyük avantajı sayısal kararlılıktır. Dempster et al. (1977) görüldüğü gibi, EM algoritmasındaki her iterasyon log olabilirlikteki gözlenen verinin artışıyla sonuçlanır, örneğin; $\log p\left(y_0;\ \theta_{(it+1)}\right) \geq \log p\left(y_0;\ \theta_{(it)}\right)$ ve maksimum olabilirlikteki ani yükselme veya düşüşe engel olur. EM algoritmasının bir eksikliği; maksimum noktasındaki komşuluğa yakınsama oranının yavaş olmasıdır. Bu oran problemdeki kayıp verinin miktarını yansıtır.

Birleşik Modeller için E-Adım

EM algoritmasının birleşik model için en çok olabilirlik tahminlerini aşağıdaki gibidir.

$h_i(t) = h_0 exp[\gamma^T w_i + \alpha\{x_i{}^T(t)\beta + z_i{}^T(t)b_i\}]$,
$y_i(t) = x_i{}^T(t)\beta + z_i{}^T(t)$
$b_i \sim \mathrm{N}(0, D)$, $\varepsilon_i(t) \sim \mathrm{N}(0, \sigma^2)$,
$\theta = \left(\theta_t{}^T, \theta_y{}^T, \theta_b{}^T\right)^T$,
$\theta_y = (\beta^T, \sigma^2)^T$,
$\theta_t = \left(\gamma^T, \alpha, \theta_{h_0}{}^T\right)^T$,

θ_{h_0}: $h_0(.)$ Temel risk fonksiyonundaki parametreler
$\theta_b = vech(D)$ olsun.

Birleşik modellerde EM'yi uygulamak için rastgele etkilere "kayıp veri" olarak davranmak gerekir.

Amaç; gözlenen veri log olabilirliğin $l(\theta) = \sum_i logp(T_i, \delta_i, y_i;\ \theta)$i maksimize eden $\hat{\theta}$ yı bulabilmektir. Burada tam veri log olabilirliğin yerine gözlenen veri log olabilirliği maksimize edilir.

$$Q\left(\theta \mid \theta^{(it)}\right) = \sum_i \int logp(T_i, \delta_i, y_i, b_i;\ \theta) p\left(b_i \mid T_i, \delta_i, y_i;\ \theta^{(it)}\right) db_i$$
$$= \sum_i \int \left\{logp(T_i, \delta_i \mid b_i;\ \theta_t, \beta) + logp\left(y_i \mid b_i; \theta_y\right) + logp(b_i; \theta_b)\right\} p\left(b_i \mid T_i, \delta_i, y_i;\ \theta^{(it)}\right) db_i.$$

rastgele etkilere göre integralin, $p(T_i, \delta_i \mid b_i;\ \theta_t, \beta)$ terimindeki sağkalım fonksiyonunun tanımındaki zaman gibi kapalı form çözümü yoktur.

$Q\left(\theta \mid \theta^{(it)}\right)$ çözümünde Gauss çeyreklik kuralları ya da Monte Carlo gibi sayısal integrasyon prosedürleri kullanılır.

Eksik olmayan verinin log olabilirliği 3 kısımdan oluşur;

$p(T_i, \delta_i, y_i, b_i; \theta) = logp(T_i, \delta_i | b_i; \theta_t, \beta) + logp(\, y_i | b_i; \theta_y) + logp(b_i; \theta_b).$

$Q\big(\, \theta \mid \theta^{(it)}\big)$ maksimizasyonu θ'ya göre sadece ilgili parameter içeren kısımları içerir.M-adımın belirlenmesinde aşağıdaki gösterimler gereklidir, zamana ve rastgele etkilere göre integraller içerir. Asıl hesaplamalar için, bölüm 1.7.3'te değinilen yöntemlerle yakınsaklaştırması gerekir.

$$\hat{\sigma}^2 = N^{-1} \sum_i \int (y_i - X_i\beta - Z_i b_i)^T \, (y_i - X_i\beta - Z_i b_i) p(b_i \mid T_i, \delta_i, y_i, \theta) \, db_i$$

$$= N^{-1} \sum_i \int (y_i - X_i\beta)^T \left(y_i - X_i\beta - 2Z_i\tilde{b}_i\right) + tr\left(Z_i{}^T Z_i \widetilde{vb}_i\right) + \tilde{b}_i{}^T Z_i{}^T Z_i \tilde{b}_i,$$

$$\widehat{D} = \; n^{-1} \sum_i \widetilde{vb}_i + \tilde{b}\tilde{b}_i{}^T,$$

$$N = \; \sum_i n_i, \; \tilde{b}_i = E\big(b_i \mid T_i, \delta_i, y_i; \theta^{(it)}\big) = \int b_i \, p\big(b_i \mid T_i, \delta_i, y_i; \; \theta^{(it)}\big) db_i,$$

$$\widetilde{vb}_i = var\big(b_i \mid T_i, \delta_i, y_i; \; \theta^{(it)}\big) = \; \int \big(b_i - \tilde{b}_i\big)^2 p\big(b_i \mid T_i, \delta_i, y_i; \; \theta^{(it)}\big) \, db_i.$$

Yukarıdaki birleşik model formülasyonu, β sabit etkiler ve θ_t sağkalım alt modelinin parametreleri için skor denklemlerinin kapalı form çözümlerine ulaşamayız.

Bu yüzden, bu parametreler için M-adım bir-adım Newton Raphson güncellemesi aracılığıyla gösterilir;

$$\hat{\beta}^{(it+1)} = \hat{\beta}^{(it)} - \left\{ \partial S\big(\hat{\beta}^{(it)}\big) \Big/ \partial\beta \right\}^{-1} \mathrm{S}\big(\hat{\beta}^{(it)}\big),$$

$$\hat{\theta}_t{}^{(it+1)} = \hat{\theta}_t{}^{(it)} - \left\{ \partial S\left(\hat{\theta}_t{}^{(it)}\right) \Big/ \partial\theta_t \right\}^{-1} \mathrm{S}\left(\hat{\theta}_t{}^{(it)}\right),$$

$\hat{\beta}^{(it)}$ ve $\hat{\theta}_t{}^{(it)}$ şimdiki iterasyondaki β ve θ_t değerlerini ifade eder, $\partial S\big(\hat{\beta}^{(it)}\big)$ ve $\partial S\left(\hat{\theta}_t{}^{(it)}\right)$; $\hat{\beta}^{(it)}$ ve $\hat{\theta}_t{}^{(it)}$zamanlarında ölçülmüş Hessian matrisinin ilgili bloklarını ifade eder. β ve θ_t'ye karşılık gelen skor vektörlerinin bileşimi

$$S(\beta) = \sum_i X_i^T \{y_i - X_i\beta - Z_i\tilde{b}_i\}/\sigma^2 + \alpha\delta_i x_i(T_i) - exp(\gamma^T w_i)$$
$$\int \int_0^{T_1} h_0(s)\alpha x_i(s) exp[\alpha\{x_i{}^T(s)\beta + z_i{}^T(s)b_i\}] \times p(b_i \mid T_i, \delta_i, y_i;\ \theta)\ ds\ db_i$$

$$S(\gamma) = \sum_i w_i \left[\delta_i - exp(\gamma^T w_i) \int \int_0^{T_1} h_0(s) exp[\alpha\{x_i{}^T(s)\beta + z_i{}^T(s)b_i\}] \times p(b_i \mid T_i, \delta_i, y_i;\ \theta)\ ds\ db_i\right],$$

$$S(\alpha) = \sum_i \delta_i \{x_i{}^T(T_i)\beta + z_i{}^T(T_i)\tilde{b}_i\} - exp(\gamma^T w_i) \int \int_0^{T_1} h_0(s) exp[\alpha\{x_i{}^T(s)\beta + z_i{}^T(s)b_i\}] \times p(b_i \mid T_i, \delta_i, y_i;\ \theta)\ ds\ db_i,$$

$$S(\theta_{h_0}) = \sum_i \delta_i \frac{\partial h_0(T_i;\theta_{h_0})}{\partial \theta_{h_0}{}^T} - exp(\gamma^T w_i) \int \int_0^{T_1} \frac{\partial h_0(s;\theta_{h_0})}{\partial \theta_{h_0}{}^T} exp[\alpha\{x_i{}^T(s)\beta + z_i{}^T(s)b_i\}] \times p(b_i \mid T_i, \delta_i, y_i;\ \theta) ds\ db_i.$$

şeklini alır.

$\partial S(\hat{\beta}^{(it)})$ ve $\partial S\left(\hat{\theta}_t{}^{(it)}\right)$; Hessian matrisinin ilgili blokları merkezi farklar yaklaşımı ile hesaplanabilir. (Press ve ark., 2007)

Ek-2. Rastgele Etkilerin Kestirim Değerleri

	uyarlanmış parçalı sabit oransal hazard temel risk fonksiyonu			uyarlanmış weibull oransal hazard temel risk fonksiyonu			Cox oransal hazard temel risk fonksiyonu			karışık etkili model	
gözlem	**sabit terim**	**Zaman**		**sabit terim**	**Zaman**		**sabit terim**	**Zaman**		**sabit terim**	**Zaman**
1	-38,535	-0,012		-38,639	-0,011		-49,025	0,079		-38,676	-0,013
2	-34,835	-0,055		-34,904	-0,052		-31,05	-0,112		-35,155	-0,053
3	19,046	0,016		19,025	0,015		24,352	-0,034		19,197	0,01
4	33,36	0,029		33,31	0,027		24,273	0,106		33,541	0,02
5	-26,263	-0,074		-26,29	-0,077		-24,264	-0,106		-26,464	-0,077
6	28,013	-0,027		28,025	-0,028		24,315	0,026		28,249	-0,034
7	7,573	0,304		7,592	0,301		-0,233	0,378		7,51	0,301
8	-49,097	-0,095		-49,236	-0,088		-49,075	-0,104		-49,263	-0,071
9	21,235	0,163		21,267	0,161		24,238	0,145		21,329	0,158
10	-28,058	-0,081		-28,117	-0,077		-24,254	-0,121		-27,94	-0,054
11	23,431	-0,091		23,449	-0,092		24,378	-0,074		23,699	-0,099
12	-8,147	0,059		-8,127	0,058		-0,003	0,004		-8,099	0,057
13	-29,386	0,212		-29,395	0,21		-24,44	0,171		-29,55	0,213
14	-7,031	0,146		-6,924	0,144		-0,089	0,141		-6,993	0,145
15	-21,733	-0,104		-21,795	-0,104		-24,281	-0,078		-21,675	-0,107
16	-28,352	-0,224		-28,272	-0,222		-24,217	-0,178		-28,148	-0,227
17	-22,063	0,142		-22,002	0,14		-24,461	0,21		-22,099	0,141
18	24,596	-0,015		24,574	-0,016		24,338	-0,012		24,796	-0,023
19	29,043	0,02		29,023	0,019		24,29	0,065		29,23	0,013
20	-6,209	0,015		-6,218	0,013		0,024	-0,041		-6,166	0,005
21	-5,681	-0,001		-5,631	-0,003		0,018	-0,029		-5,498	-0,01
22	27,223	-0,07		27,183	-0,071		24,361	-0,048		27,456	-0,078
23	6,81	0,195		6,805	0,193		-0,162	0,259		6,805	0,192
24	23,717	0,033		23,642	0,031		24,333	-0,004		23,835	0,026
25	27,061	0,054		27,034	0,052		24,284	0,074		27,213	0,047
26	-3,969	0,078		-3,924	0,077		-0,048	0,076		-3,902	0,075
27	8,105	0,138		8,158	0,135		4,728	0,211		8,183	0,137
28	-7,638	0,058		-7,542	0,057		-0,03	0,047		-7,538	0,055
29	-5,365	0,112		-5,361	0,111		-0,047	0,073		-5,354	0,11
30	8,201	-0,107		8,219	-0,107		14,041	-0,129		8,421	-0,113
31	34,155	0,076		34,154	0,074		24,811	0,168		34,336	0,068

Çizelge 4.1. Rasgele Etkilerin Kestirim Değerleri

32	26,308	-0,003		26,305	-0,004		24,316	0,024		26,512	-0,01
33	-48,949	-0,109		-49,111	-0,105		-49,062	-0,125		-49,125	-0,087
34	-27,98	-0,066		-28,047	-0,063		-24,301	-0,093		-27,204	-0,005
35	-0,667	0,012		-0,675	0,012		0,002	0,014		-0,544	0,012
36	20,167	0,018		20,115	0,016		24,145	0,014		20,287	0,01
37	27,265	0,009		27,25	0,006		24,317	0,021		27,491	-0,001
38	-5,059	-0,015		-5,073	-0,017		0,037	-0,058		-4,963	-0,02
39	17,256	-0,024		17,232	-0,025		24,387	-0,093		17,427	-0,03
40	-8,51	0,155		-8,505	0,153		-0,076	0,084		-8,539	0,153
41	-56,819	0,002		-57,025	-0,035		-49,276	0,003		-57,186	0,015
42	18,24	-0,048		18,26	-0,048		24,383	-0,083		18,461	-0,055
43	29,98	-0,066		29,942	-0,067		24,341	-0,014		30,218	-0,074
44	4,839	0,072		4,824	0,071		-0,075	0,119		4,904	0,067
45	24,38	-0,085		24,365	-0,086		24,381	-0,079		24,633	-0,094
46	-24,417	0,092		-24,431	0,09		-24,4	0,108		-24,488	0,095
47	-28,248	-0,024		-28,293	-0,025		-24,288	-0,068		-28,243	-0,027
48	-35,523	-0,087		-35,523	-0,087		-24,234	-0,162		-35,485	-0,089
49	-8,527	-0,012		-8,568	-0,013		-0,107	-0,103		-8,461	-0,016
50	32,513	-0,104		32,457	-0,105		24,367	-0,036		32,777	-0,113
51	-3,206	-0,053		-3,122	-0,053		0,01	-0,016		-3,022	-0,057
52	14,132	-0,046		14,136	-0,047		24,388	-0,125		14,324	-0,052
53	0,332	-0,098		0,301	-0,098		0,055	-0,086		0,485	-0,103
54	-26,866	-0,012		-27,687	-0,012		-24,371	-0,007		-26,683	0
55	-34,979	-0,126		-34,98	-0,125		-24,209	-0,194		-34,904	-0,129
56	-40,625	-0,148		-40,772	-0,123		-49,105	-0,026		-41,051	-0,151
57	11,866	0,052		11,83	0,05		4,812	0,112		11,958	0,047
58	26,467	0,072		26,441	0,069		24,276	0,086		26,606	0,065
59	11,243	0,012		11,238	0,011		17,008	-0,035		11,379	0,007
60	6,429	0,012		6,403	0,011		-0,043	0,079		6,534	0,007
61	-74,364	-0,031		-74,754	-0,028		-75,038	-0,015		-73,592	-0,003
62	21,868	0,025		21,889	0,023		24,319	0,019		22,053	0,018
63	-7,006	0,024		-6,918	0,023		-0,011	0,017		-6,885	0,021
64	-11,024	0,083		-10,964	0,082		-0,015	0,018		-10,978	0,081
65	-16,865	-0,029		-16,953	-0,026		-24,324	-0,007		-16,578	0,004
66	11,94	-0,049		12,004	-0,05		24,401	-0,118		12,165	-0,055
67	37,877	-0,061		37,834	-0,062		44,347	-0,136		38,137	-0,069
68	23,191	-0,054		23,176	-0,054		24,367	-0,057		23,41	-0,061
69	-2,024	-0,055		-2,056	-0,055		0,044	-0,07		-1,909	-0,059
70	-32,81	-0,045		-32,666	-0,032		-24,318	-0,049		-32,465	-0,001
71	13,034	0,035		12,999	0,034		16,611	-0,009		13,142	0,03
72	22,077	-0,041		22,036	-0,042		24,376	-0,071		22,266	-0,048
73	30,92	-0,082		30,874	-0,082		24,346	-0,019		31,174	-0,091

Çizelge 4.1. Devam. Rasgele Etkilerin Kestirim Değerleri

74	-31,436	-0,037		-31,253	-0,014		-24,298	-0,053		-31,539	-0,008
75	1,98	-0,044		2,074	-0,044		0,373	0,053		2,181	-0,048
76	32,854	-0,099		32,79	-0,1		24,364	-0,031		33,11	-0,108
77	11,451	-0,067		11,427	-0,068		19,027	-0,135		11,628	-0,073
78	-18,891	0,071		-18,94	0,07		-24,408	0,122		-18,875	0,091
79	-8,563	0,014		-8,651	0,013		-1,836	-0,081		-8,756	0,025
80	-41,618	-0,037		-41,461	-0,025		-48,701	-0,013		-41,306	0
81	0,566	0,009		0,503	0,008		0,001	-0,002		0,627	0,004
82	42,379	-0,011		42,355	-0,013		49,197	-0,094		42,644	-0,023
83	27,914	0,156		27,929	0,152		24,202	0,204		28,025	0,15
84	23,58	0,014		23,55	0,012		24,334	-0,004		23,751	0,006
85	-20,564	-0,035		-20,475	-0,013		-24,326	-0,008		-20,871	-0,018
86	23,899	0,039		23,881	0,037		24,311	0,031		24,062	0,032
87	-16,478	-0,003		-17,241	-0,003		-24,103	0,002		-16,035	0,004
88	25,871	0,026		25,838	0,025		24,309	0,034		26,036	0,019

Çizelge 4.1. Devam. Rastgele Etkilerin Kestirim Değerleri

Ek-3. Tez çalışmasında kullanılan veri seti ile ilgili izin yazısı

4.7.2016

İLGİLİ MAKAMA,

AÜTF Göğüs Hastalıkları Yoğun Bakım Ünitesi' nde 2012-2015 yılları arasında gerçekleşen ve yürütücülüğünü yaptığım 'Akut hiperkapnik solunum yetersizliği olan kronik obstrüktif akciğer hastalarında AVAPS etkinliğinin değerlendirilmesi' başlıklı araştırma kapsamındaki verilerin Anabilim Dalınız yüksek lisans öğrencilerinden Ebru Turgal tarafından gerçekleştirilecek olan 'Sağkalım verileri ile boylamsal verilerin birleşik modellenmesi ve sağlık alanına bir uygulama' başlıklı tez çalışması için kullanılmasına izin veriyorum.

Saygılarımla

Uz. Dr. Fatma Çiftci

Ankara Üniversitesi Tıp fakültesi

Göğüs Hastalıkları Anabilim Dalı

Tel: 5958421

Email: fciftci@ankara.edu.tr

ÖZGEÇMİŞ

I-Bireysel Bilgiler

Adı ve Soyadı: Ebru TURGAL

Doğum Yeri ve Tarihi: Ankara - 16.01.1990

Uyruğu: T.C.

Medeni Durumu: Bekar

İletişim Adresi ve Telefonu: Hitit Üni. Tıp Fak. Dekanlığı Biyoistatistik A.B.D., Merkez, ÇORUM

Tel: (0364) 223 08 00

II-Eğitimi

Yüksek Lisans	Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı
Lisans	Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü
Lise	Mehmet Emin Resulzade Anadolu Lisesi

Yabancı Dili: İngilizce

III- Mesleki Deneyimi

2014 - … Hitit Üniversitesi (Araştırma Görevlisi)

IV- Bilimsel İlgi Alanları

Sağkalım analizi

Tekrarlı ölçümler

Sözlü Sunumları:

Turgal, Ebru, Doğanay Erdoğan, B., "Sağkalım Verileri ile Boylamsal Verilerin Birleşik Modellenmesi ve Sağlık Alanında bir Uygulama", 18. Ulusal Biyoistatistik Kongresi ve 1. Uluslararası Biyoistatistik Kongresi, Antalya, 26-29 Ekim 2016.

Poster Sunumları:

Turgal, Ebru, Doğanay Erdoğan, B., "Joint Modeling of Survival and Longitudinal Non-Survival Data with Applications in R", International Conference on Information Complexity and Statistical Modeling in High Dimensions with Applications (IC-SMHD-2016), Nevşehir, 18-21 Mayıs 2016.

Turgal, Ebru, Doğanay Erdoğan, B., "Gamma Frailty Model versus Cox Regression Model with an Application to Medical Data", 2.Uluslararası Araştırmacılar-İstatistikçiler ve Genç İstatistikçiler Kongresi. Hacettepe Üniversitesi, Ankara, 4-8 Mayıs 2016.

Demir, Emre, **Turgal, E.**, Köse, S.K., Dünder, E., Cengiz, M.A., Şenel, E., "Sağlık Hizmetlerinin Etkinlik Ölçümünde Klasik, Bulanık Veri Zarflama Analizi ve Stokastik Sınır Analizi Yöntemlerinin Karşılaştırılması", 17.Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı P28, sayfa 34, 5-9 Kasım 2015.

Turgal, Ebru, Yavuz, Z., Demir, E., Köse, S.K., Doğanay Erdoğan, B., "Mikrodizi Verilerinin Karar ve Regresyon Ağacı Yöntemleriyle Sınıflandırılması", 17.Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı P32, sayfa 81-82, 5-9 Kasım 2015.

Turgal, Ebru, Dağlı, İ., Yavuz, Z., Köse, S.K., Doğanay Erdoğan,B., "Non-Linear Regression Analysis", 8. Doğu Akdeniz Bölgesi Uluslararası Biyometrik Topluluğu Konferansı Bildiri Özetleri Kitabı, Hacı Bektaşi Veli Üniversitesi, Nevşehir, sayfa 152, 11-16 Mayıs 2015.

Yavuz, Zeynep, **Turgal, E.**, Dağlı İ., Doğanay Erdoğan,B., Köse, S.K., Yavuz, Y., "Yarışan Riskler Varlığında Sağkalım Analizi", 16. Biyoistatistik Kongresi Bildiri Özetleri Kitabı, Ankara Üniversitesi Örsem Tesisleri, Antalya, sayfa 88-89, 10-12 Eylül 2014.

VI- Diğer Bilgiler

KATILIMCI OLARAK YER ALINAN KURS ve ÇALIŞTAYLAR

Short Course on Survival Analysis, Faculty of Mathematics and Informatics, Sofia University “St. Kliment Ohridski”, Sofia, Bulgaria, 18.07.2016-20.07.2016.

Lecturer: Laura Antolini, Davide Bernasconi (University of Milano-Bicocca)