

**KRİTİK HASTALIK SİGORTALARINDA SAKLI
MARKOV MODELİ**

**CRITICAL ILLNESS INSURANCE WITH A HIDDEN
MARKOV MODEL**

GÖNÜL AYRANCI

PROF. DR. MERAL SUCU

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Aktüerya Bilimleri Anabilim Dalı için Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

ÖZET

KRİTİK HASTALIK SİĞORTALARINDA SAKLI MARKOV MODELİ

Gönül AYRANCI

Doktora, Aktüerya Bilimleri Bölümü

Tez Danışmanı: Prof. Dr. Meral SUCU

Haziran 2022, 111 sayfa

Yazında, Kritik hastalık sigortalarına ilişkin hesaplamalarda çok durumlu modeller kullanılmaktadır. Kritik hastalıkların görülme olasılıklarının tahmin edildiği bu modellerde, eklenen açıklayıcı değişkenler yöntemin kısıtı gereği ancak farklı grup kombinasyonlarına bölünülerek modellere dahil olabilmektedir. Bu durum verinin çok fazla sayıda alt risk grubuna bölünmesine, açıklayıcı değişken olarak sürekli değişken kullanılamamasına ve sonuçların ancak belirli sayıda grup için geçerli olmasına neden olmaktadır.

Bu tez çalışmasında, kritik hastalık sigortalarının fiyatlandırılmasında kullanılabilecek, risk gruplarının bütünsel bir yaklaşımla değerlendirildiği bir modelin önerilmesi amaçlanmaktadır. Bu amaçla, kritik hastalık sigortalarının modellenmesinde Saklı

Markov Modellerinin (SMM) kullanılması önerilmiştir. SMM kullanılarak, farklı risk gruplarına göre kritik hastalıkların görülme olasılıkları tahmin edilmiştir.

Çalışmada, Michigan Üniversitesi Sosyal Araştırmalar Enstitüsü tarafından yürütülen Sağlık ve Emeklilik Çalışması (The Health and Retirement Study- HRS) kapsamında toplanan boylamsal veri kullanılmıştır. HRS verisi kullanılarak hem geleneksel model hem de SMM parametre tahminleri yapılmış ve performansları karşılaştırılmıştır. SMM, incelenen durumların çoğunda, parametreleri Genelleştirilmiş Doğrusal Model ile tahmin edilen Markov modeline oranla daha başarılı performans göstermiştir. Tezde ayrıca Thiele diferansiyel denklemleri ve benzetim yöntemi ile prim hesaplamaları yapılmıştır.

Anahtar Kelimeler: Kritik hastalık sigortası, Saklı Markov modeli, Markov modeli, Prim hesaplamaları, HRS

ABSTRACT

CRITICAL ILLNESS INSURANCE WITH A HIDDEN MARKOV MODEL

Gönül AYRANCI

Doctor of Philosophy, Department of Actuarial Science

Supervisor: Prof. Dr. Meral SUCU

June 2022, 111 pages

In the literature, multi-state models are used in the calculations on critical illness insurance. In these models, in which the critical illness incidence rates are estimated, covariates can only be included after splitting into different group combinations due to the constraints of the method. This situation causes the data to be divided into too many sub-risk groups, inability to use continuous covariates and restriction in the validity of the results to a particular number of groups.

In this thesis it is aimed to propose a model that can be used in the pricing of critical illness insurance and considers the risk groups with a holistic perspective. For this purpose, using Hidden Markov Models (HMM) in modeling critical illness insurances is proposed. Incidence rates of critical illnesses are estimated for different risk groups with HMM.

This thesis uses the longitudinal panel data collected in The Health and Retirement Study (HRS) by Michigan University Social Research Institute. Using HRS data, parameters of both HMM and the orthodox models in the literature are estimated and their performances are compared. Results indicate that in most of the scenarios, the HMM model proposed in the thesis performs better than the Markov Model whose parameters are computed using Generalized Linear Model. Furthermore, premium calculations are done using Thiele's differential equation and simulation method.

Keywords: Critical illness insurance, Hidden Markov model, Markov model, Premium calculation, HRS

TEŞEKKÜR

Tez çalışmamda desteklerini ve yardımlarını esirgemeyen danışmanım Sayın Prof. Dr. Meral SUCU'ya,

Değerli eleştirileri ve katkıları için tez izleme komitesi üyelerim Sayın Prof. Dr. Cenap Erdemir'e, Sayın Dr. Öğr. Üyesi Yasemin SAYKAN'a ve tez jüri üyelerim Sayın Prof. Dr. Devin SEZER'e, Sayın Doç. Dr. Könül BAYRAMOĞLU KAVLAK'a,

Akademik hayatıma çok büyük katkıları olan ve desteğini hiç esirgemeyen hocam Sayın Prof. Dr. Behzat GÜRKAN'a,

Hayatımın her anında bana destek olan, beni koşulsuz sevgiyle büyüten canım annem Fikret AYRANCI'ya ve canım babam Asım AYRANCI'ya

En içten teşekkürlerimi sunarım.

Gönül AYRANCI

Haziran, 2022, Ankara

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ.....	ix
ÇİZELGELER DİZİNİ	xi
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ	1
2. GENEL BİLGİLER.....	4
2.1. Kritik Hastalık Sigortaları	4
2.2. Temel Problemler	6
2.2.1. Problem 1: Hastalığın teşhisi.....	6
2.2.2. Problem 2: Kısıtlı Veri	7
2.2.3. Problem 3: Risk gruplarının belirlenmesi	10
2.3. Çok Durumlu Modeller	12
2.4. Kritik Hastalıkların Stokastik Modellenmesi.....	14
2.4.1. Markov Süreçleri.....	14
2.4.2. Kolmogorov İleri Denklemleri.....	17
2.4.3. Thiele Diferansiyel Denklemleri	18
2.4.4. Poisson Regresyonu	19
3. VERİ.....	21
3.1. Sağlık ve Emeklilik Çalışması Verileri (HRS)	21
3.2. Açıklayıcı Değişkenler	28
3.2.1. Sigara Kullanımı	30
3.2.2. Alkol Kullanımı.....	32

3.2.3. Beden Kütle İndeksi (BKI)	35
3.2.4. Tansiyon Hastalığı	38
3.2.5. Diyabet Hastalığı	40
3.2.6. Psikolojik Problemler	42
3.2.7. Romatizma	45
3.2.8. Akciğer Hastalığı	47
4. YÖNTEM	49
4.1. Saklı Markov Modellerinin (SMM) kullanıldığı öne çıkan çalışmalar	49
4.2. Saklı Markov Modelleri	50
4.3. Saklı Markov Modeli (SMM) Bileşenleri	51
4.4. SMM Temel Problemleri için Çözüm Algoritmaları	53
4.4.1. Problem 1	53
4.4.2. Problem 2	55
4.4.3. Problem 3	57
5. UYGULAMA	59
5.1. Kritik Hastalıkların Demografik Özellikleri	59
5.2. Genelleştirilmiş Doğrusal Model (Poisson Regresyon)	60
5.2.1. Ölüm Durumu için Modeller	64
5.2.2. Kanseri için Modeller	67
5.2.3. Kalp Hastalığı için Modeller	70
5.2.4. Felç Hastalığı için Modeller	73
5.3. Saklı Markov Modeli	75
5.3.1. Temel Model	76
5.3.2. Farklı Açıklayıcı Değişken Senaryoları	82
5.3.3. Kritik Hastalılar için Saklı Markov Modeli	86
5.4. Modellerin Karşılaştırılması	93
5.4.1. Ölüm için Modellerin Karşılaştırılması	94
5.4.2. Kanseri için Modellerin Karşılaştırılması	96
5.4.3. Kalp Hastalığı için Modellerin Karşılaştırılması	97
5.4.4. Felç için Modellerin Karşılaştırılması	98
5.5. Prim Hesaplamaları	100

6. SONUÇLAR VE ÖNERİLER	103
7. KAYNAKLAR.....	106
ÖZGEÇMİŞ	111



ŞEKİLLER DİZİNİ

Şekil 2.1. İki durumlu Ölüm Modeli.....	12
Şekil 2.2. Çoklu Azalım Modeli	13
Şekil 2.3. Kalıcı Maluliyet Modeli	13
Şekil 2.4. Hastalık Ölüm Modeli	14
Şekil 3.1. 1992 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı.....	29
Şekil 3.2. 2000 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı.....	29
Şekil 3.3. 2010 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı.....	30
Şekil 3.4. 2018 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı.....	30
Şekil 3.5. 1992 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı	31
Şekil 3.6. 2000 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı	31
Şekil 3.7. 2000 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı	32
Şekil 3.8. 2018 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı	32
Şekil 3.9 1992 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı.....	33
Şekil 3.10. 2000 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı.....	33
Şekil 3.11. 2010 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı.....	34
Şekil 3.12. 2018 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı.....	34
Şekil 3.13. 1992 yılında gözlemlenen bireylerin yaşa göre BKİ kategorileri.....	36
Şekil 3.14. 2000 yılında gözlemlenen bireylerin yaşa göre BKİ kategorileri.....	36
Şekil 3.15. 2010 yılında gözlemlenen bireylerin yaşa göre BKİ kategorileri.....	37
Şekil 3.16. 2018 yılında gözlemlenen bireylerin yaşa göre BKİ kategorileri.....	37
Şekil 3.17. 1992 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı..	38
Şekil 3.18. 2000 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı..	39
Şekil 3.19. 2010 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı..	39
Şekil 3.20. 2018 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı..	40
Şekil 3.21. 1992 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı....	41
Şekil 3.22. 2000 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı....	41
Şekil 3.23. 2010 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı....	41
Şekil 3.24. 2018 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı....	42
Şekil 3.25. 1992 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı	43
Şekil 3.26. 2000 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı	43

Şekil 3.27. 2010 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı.....	44
Şekil 3.28. 2018 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı.....	44
Şekil 3.29. 1992 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı.....	45
Şekil 3.30. 2000 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı.....	45
Şekil 3.31. 2010 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı.....	46
Şekil 3.32. 2018 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı.....	46
Şekil 3.33. 1992 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı	47
Şekil 3.34. 2000 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı	47
Şekil 3.35. 2010 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı	48
Şekil 3.36. 2018 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı	48
Şekil 4.1. Kavramsal bir SMM örneği	51
Şekil 5.1. Kritik hastalıklar için önerilen Markov modeli.....	60
Şekil 5.2. Saklı durum sayısına göre model performansları	77
Şekil 5.3. Saklı durumlar arası ortalama geçiş olasılıkları	78
Şekil 5.4. Saklı durumlar arası ortalama geçiş olasılıkları.....	87
Şekil 5.5. Saklı durumda bulunma olasılıklarının marginal dağılımları	93

ÇİZELGELER DİZİNİ

Çizelge 2.1. Kritik Hastalık Sigortaları ile teminat altına alınan hastalıklar	5
Çizelge 3.1. RAND HRS içerisinde yer alan veri grupları ve özellikleri	22
Çizelge 3.2. RAND HRS içerisinde yer alan bireylerin demografik özellikleri.....	23
Çizelge 3.3. RAND HRS içerisinde yer alan eşlerin demografik özellikleri	24
Çizelge 3.4. RAND HRS verisinden 2010 kodlu katılımcının değerleri.....	28
Çizelge 3.5. BKİ Kategorileri	35
Çizelge 5.1. Yaş aralıklarına göre sağlıklı durumdan kritik hastalıklara geçiş sıklığı ...	59
Çizelge 5.2. Kullanılan açıklayıcı değişkenler ve değerleri	62
Çizelge 5.3. Model numaralarına göre açıklayıcı değişkenler.....	64
Çizelge 5.4. Ölüm için temel model çıktıları.....	65
Çizelge 5.5. Üç değişkenli senaryolar için model performansları.....	66
Çizelge 5.6. Dört değişkenli senaryolar için model performansları	66
Çizelge 5.7. Ölüm modeli ve parametre tahminleri.....	67
Çizelge 5.8. Kanser için temel model çıktıları.....	68
Çizelge 5.9. Kanser için temel model çıktıları (quasiPoisson)	68
Çizelge 5.10. Üç değişkenli senaryolar için model performansları (Kanser).....	69
Çizelge 5.11. Dört değişkenli senaryolar için model performansları (Kanser)	69
Çizelge 5.12. Kanser modeli ve parametre tahminleri.....	70
Çizelge 5.13. Kalp hastalığı için temel model çıktıları.....	71
Çizelge 5.14. Kalp hastalığı için temel model çıktıları (quasiPoisson)	71
Çizelge 5.15. Üç değişkenli senaryolar için model performansları (Kalp hastalığı).....	72
Çizelge 5.16. Dört değişkenli senaryolar için model performansları (Kalp hastalığı) ...	72
Çizelge 5.17. Kalp hastalığı modeli ve parametre tahminleri.....	73
Çizelge 5.18. Felç hastalığı için temel model çıktıları	73
Çizelge 5.19. Üç değişkenli senaryolar için model performansları (Felç hastalığı).....	74
Çizelge 5.20. Dört değişkenli senaryolar için model performansları (Felç hastalığı) ...	75
Çizelge 5.21. Felç hastalığı modeli ve parametre tahminleri	75
Çizelge 5.22. Saklı durumlar arası ortalama geçiş olasılıkları	78
Çizelge 5.23. Saklı durum 1'den geçiş olasılıkları için parametre tahminleri.....	79
Çizelge 5.24. Saklı durum 2'den geçiş olasılıkları için parametre tahminleri.....	80

Çizelge 5.25. Saklı durum 3'ten geçiş olasılıkları için parametre tahminleri	80
Çizelge 5.26. Saklı durum 4'ten geçiş olasılıkları için parametre tahminleri	81
Çizelge 5.27. Saklı durumlar için koşullu olasılık matrisi	82
Çizelge 5.28. Saklı Markov modellerinin karşılaştırılması.....	83
Çizelge 5.29. Saklı Markov modellerinin karşılaştırılması.....	84
Çizelge 5.30. Saklı Markov modellerinin karşılaştırılması.....	84
Çizelge 5.31. Saklı Markov modellerinin karşılaştırılması.....	85
Çizelge 5.32. Saklı Markov modellerinin karşılaştırılması.....	86
Çizelge 5.33. Saklı durumlar arası ortalama geçiş olasılıkları	87
Çizelge 5.34. Saklı durum 1'den geçiş olasılıkları için parametre tahminleri	88
Çizelge 5.35. Saklı durum 2'den geçiş olasılıkları için parametre tahminleri	89
Çizelge 5.36. Saklı durum 3'ten geçiş olasılıkları için parametre tahminleri	90
Çizelge 5.37. Saklı durum 4'ten geçiş olasılıkları için parametre tahminleri	91
Çizelge 5.38. Saklı durumlar için koşullu olasılık matrisi	92
Çizelge 5.39. 10 yıllık kritik hastalık sigortası için primler	102
Çizelge 5.40. 15 yıllık kritik hastalık sigortası için primler	102
Çizelge 5.41. 20 yıllık kritik hastalık sigortası için primler	102

SİMGELER VE KISALTMALAR

Simgeler

N	Saklı Markov modelindeki gözlem sayısı
M	Her saklı durum için gözlem adedi
a_{ij}	Saklı durumlar arası geçiş olasılık dağılım fonksiyonu
b_j	Her j saklı durumu için gözlem olasılık dağılım fonksiyonu
π	Saklı Markov modeli başlangıç durumu olasılık dağılımı
λ	Saklı Markov Modeli
V	Gözlem kümesi

Kısaltmalar

HRS	The Health and Retirement Study
NIA	National Institute on Aging
SMM	Saklı Markov Modeli
BKI	Beden Kütle İndeksi
SD	Serbestlik Derecesi
BIC	Bayesçi Bilgi Kriteri
AIC	Akaike Bilgi Kriteri
GLM	Genelleştirilmiş Doğrusal Modeller

1. GİRİŞ

Üründen bağımsız olarak tüm sigorta şirketleri, geniş risk faktörlerinin etkisinde prim hesaplamaları yapmak ve mali yeterliliklerini karşılamak zorundadır. Özellikle sağlık sigortalarında ve bunun daha özel bir biçimi olan kritik hastalık sigortalarında hastalık (morbidite) riski büyük önem kazanmaktadır ve doğru şekilde yönetilmesi gerekmektedir. Hastalıklara ilişkili hasarlar doğru modellendiği takdirde morbidite riskinin de azalacağı düşünülebilir. Ayrıca morbidite riski ele alınırken bunda etkili olabilecek yaş, cinsiyet, beslenme, sigara ve alkol kullanımı, fiziksel aktivite gibi faktörler de incelenmelidir.

Türkiye’de, sigorta şirketleri, kritik hastalıklar sigortası, tehlikeli hastalıklar sigortası, yaşam riskleri sigortası, moral destek sigortası gibi farklı isimler ile yaklaşık 20 farklı kritik (ölümcül ya da maluliyete neden olan) hastalığa karşı teminat sağlayan ürünler sunmaktadır. Genellikle bu sigortalar ileri yaştaki nüfusa yönelik oluşturulmuştur ve 18-70 yaş arasındaki bireyler bu sigortalardan faydalanabilmektedir. Teminatlar sigortalının isteğine göre sigortalının ölümü, maluliyeti ya da önceden belirlenen ve yaşamı tehdit eden bazı hastalıkların teşhisi durumunda ödenmektedir. Şirketlerin sunduğu ürünler arasında farklılıklar olmakla birlikte genel olarak Türkiye’de de en sık görülen kanser, kalp krizi, böbrek yetmezliği, uzuv kaybı, işitme kaybı, inme, ciddi yanıklar vb. hastalıkları kapsamaktadır.

Tez çalışmasının temel amacı, kritik hastalık sigortalarının fiyatlandırılmasında kullanılmak üzere kritik hastalıkların görülme olasılıklarının tahmin edilebileceği bir model önerisi sunmaktadır. Bu amaçla, tez çalışmasında Saklı Markov Modelleri (SMM) yönteminin kullanılması önerilmiştir. Bu yöntem sayesinde önceden belirlenen risk sınıflandırmaları yerine veriden elde edilecek sınıflandırmaların kullanılması mümkündür. Doğru risk sınıflandırması hem sigortacılar için hem de sağlık konusundaki kamu düzenleyicileri için oldukça önemlidir. Özellikle zaman içerisinde değişebilen, kesikli ya da sürekli olan risk değişkenleri ile modelleme yapıldığı düşünüldüğünde bu risk faktörlerinin etkili bir biçimde sınıflandırılması, sağlık ile ilişkili sigortaların fiyatlandırılmasında büyük öneme sahiptir. Bu çeşit bir veriyi sınıflandırmak ve

belirlenen sınıflar için ayrı süreçlerin modellenmesinin daha doğru sonuçlar vereceği düşünülmektedir. SMM sayesinde bu şekilde bir modelleme mümkün olabilmektedir.

Yazında nispeten yeni kullanılmaya başlanan, özellikle ses ve görüntü tanımlama problemlerinde kullanılan Saklı Markov Modellerinin sağlık sigortalarında kullanımı geliştirmeye açıktır. Odaklanılan problem için bu yöntemin kullanılmasının sağlık sigortalarında aktüeryal modelleme yazınına katkı sağlayacağı ve sigorta sektöründe farklı kullanım alanlarına örnek olacağı düşünülmektedir.

Tez çalışması aşağıda kısaca özetlenen altı bölümden oluşmaktadır.

İlk bölümde tez çalışmasının temel amacı ve araştırma problemi açıklanmıştır. İkinci bölümde ise araştırma problemine ilişkin genel bilgilerle birlikte, yazında yapılan çalışmalar tartışılmıştır. İkinci bölümde yazında kullanılan yöntemler özetlenmiştir.

Üçüncü bölümde, tez çalışmasında kullanılan veri ile ilgili bilgiler verilmiş ve yazında daha önce kullanıldığı çalışmalar özetlenmiştir. Bu veri seti içerisinde yer alan ve önerilen modelde kullanılan değişkenlere ilişkin tanımlayıcı istatistikler verilmiştir.

Dördüncü bölümde, tez çalışmasında kullanılan yöntemle bilgilere yer verilmiştir. Öncelikle, tez çalışmasında kullanılan Saklı Markov modellerinin kullanıldığı çalışmalar özetlenmiştir. Sonrasında model bileşenleri ve kullanılan gösterim verilmiş ve modelde kullanılan çözüm algoritmaları açıklanmıştır.

Beşinci bölümde iki farklı yöntem kullanılarak uygulama yapılmış ve sonuçlar verilmiştir. Bu bölüm dört alt bölümden oluşmaktadır. İlk alt bölümde, modelde kullanılan ve Kanseri, Felç, Kalp Hastalığı ve Ölüm durumlarının görülme sıklığına yönelik istatistikler özetlenmiştir. Bir sonraki bölümde, yazında hali hazırda kritik hastalık sigortaları konusunda kullanılan bir Markov Modeli önerilmiş ve üçüncü

bölümde ayrıntıları verilen veri seti ile parametre tahminleri yapılmıştır. Daha sonra, tezde önerilen yöntem olan Saklı Markov Modelleri kullanılarak aynı veri seti üzerinde parametre tahminleri yapılmış ve en iyi model yapısı tartışılmıştır. Bu bölümde ayrıca yazında daha önce kullanılan model ile tezde önerilen modelin karşılaştırılması yapılmış ve sonuçları verilmiştir. En son bölümde ise bu modellerin sigorta prim hesaplamaları konusundaki kullanımı tartışılmış ve prim tahminleri sunulmuştur.

Altıncı ve son bölümde, tez çalışmasından elde edilen önemli sonuçlar, bu çalışmanın yazına katkısı ile çalışmanın geliştirilmesine yönelik öneriler sunulmuştur.



2. GENEL BİLGİLER

2.1. Kritik Hastalık Sigortaları

Kritik hastalık sigortası önceden belirlenmiş olan hastalıkların teşhisi sonrası sigortalıya toplu ödeme yapılmasını sağlar. Genellikle bu sigortalar ileri yaştaki nüfus için oluşturulmuştur ancak 18-70 yaş arasındaki bireyler de bu sigortalardan faydalanabilirler. Tazminatlar sigortalının isteğine göre ölüm ya da hastalık durumunda ödenmektedir. Şirketlerin sunduğu ürünler arasında farklılıklar olmakla birlikte, teminat altına alınan hastalıklar arasında genel olarak kalp krizi, ameliyat gerektiren koroner arterlerin sertleşme ve daralması ile belirgin kalp hastalığı, kanser ve felç bulunmaktadır. Poliçenin ilk uygulamalarından itibaren toplu tazminat ödemeleri iki farklı şekilde yapılabilmektedir.

- a. Tek başına Kritik Hastalık Sigortası (Stand-alone): Sigorta süresi içinde teminat altına alınan kritik hastalıklardan birinin teşhisi durumunda tek seferde tüm teminat tutarı kadar tazminat ödemesi yapılır. Genel olarak sigortalının, hastalığın teşhisi sonrası 14 ile 90 günlük süre aralığında hayatta kalması şartı konulabilmektedir. Tipik olarak hasarın kabulü için teşhis sonrası 28 gün hayatta kalma şartı getirilmektedir [1].
- b. Öncelenmiş Menfaatli Kritik Hastalık Sigortası (Accelerated Benefit): Öncelenmiş menfaatli poliçelerde, teminat altına alınan kritik hastalıklardan birinin teşhisi durumunda ölüm teminat tutarının hepsi ya da bir kısmı ödenebilir. Tüm teminat tutarının hastalık durumunda ödenmesi halinde ölüm gerçekleştiğinde herhangi bir tazminat ödemesi yapılmaz. Kritik hastalığın teşhisinde daha önceden belirlenen bir kısmının ödenmesi durumunda, ölüm gerçekleştiğinde kalan kısmı ödenir. Türkiye’de de genellikle hayat sigortaları ile birlikte bu çeşit ürünler sunulmaktadır ve kritik hastalık teminatı ölüm teminatının yarısı olarak belirlenmektedir [2] .

Kritik hastalık sigortalarının kapsadığı bazı hastalıklar aşağıdaki Çizelge 2.1’de özetlenmiştir.

Çizelge 2.1. Kritik Hastalık Sigortaları ile teminat altına alınan hastalıklar

Dört temel hastalık	Kanser Kalp krizi Felç Koroner bypass ameliyatı
Ek olarak sunulabilen hastalıklar	Böbrek yetmezliği Organ nakli MS (Multiple Skleroz) hastalığı Alzheimer hastalığı Körlük, duyma ya da konuşma kaybı Koma Uzuv kaybı Menenjit Parkinson hastalığı Tam ve kalıcı maluliyet Ciddi yanıklar Karaciğer yetmezliği Solunum yetmezliği Motor Nöron hastalığı
Hariç tutulanlar	Madde bağımlılığı Alkol bağımlılığı Medikal tavsiyeye uymama AIDS Ağır ihmal

Kritik hastalık sigortaları bir ürün olarak ilk kez 1983 yılında Güney Afrika'da Crusader Life şirketi tarafından geliştirilmiştir [3]. Bununla birlikte Amerika, Japonya ve İsrail'de

daha önce kanser hastalığına yönelik ürünlerin geliştirildiği bilinmektedir [4]. Standart hayat sigortalarından farklı olarak kritik hastalık sigortaları, sigortalının gelir kaybını veya tedavi masraflarını karşılamamaktadır. Ürün, hastalığın ilk ortaya çıkışından ölüme kadar olan süreyi de göz ardı ederek sadece kritik hastalık teşhisinin gerçekleşmesine odaklanmaktadır. Hatta bu ürünlerden birinin pazarlanma sloganı “ölmeden hayat garantinizi kazanın” olarak kullanılmıştır [4].

Dash ve Grimshaw [4], kritik hastalık sigortalarının bir ürün olarak sigorta sektöründe doldurduğu boşluğu açıklarken, geride miras bırakacak kimsesi olmayan insanların bir hayat sigortası yerine, ciddi bir hastalık ile karşılaşmaları durumunda ihtiyaç duyacakları finansal desteği karşılayacak hiçbir ürünün olmadığından bahsetmektedir. Bu kişiler için kritik hastalık sigortası hayat sigortasına oranla çok daha rasyonel olabilmektedir.

2.2. Temel Problemler

Kritik hastalık sigortalarında prim hesaplaması, poliçenin fiyatlandırılması ve hasarın gerçekleşmesi ile ilgili yazında bahsedilen üç temel problem bulunmaktadır. Bu bölümde yapılan yazın taraması ile geçmişte kritik hastalık sigortasına yönelik yapılan çalışmaların bu üç problemi ne şekilde ele aldıklarına yönelik bir çerçevede sunulmaktadır.

2.2.1. Problem 1: Hastalığın teşhisi

Kritik hastalık poliçesi ile teminat altına alınan tazminatın ödenebilmesi için öncelikle hasarın gerçekleştiğinin kesin olarak belirlenmesi gerekmektedir. Ancak poliçeye konu olan kritik hastalıklar ile ilgili yapılan tıp çalışmaları sayesinde, teşhise yönelik daha ileri teknikler geliştirilmektedir. Poliçe şartlarında adı geçen kritik hastalık tanımları zaman içerisinde büyük değişkenlik gösterebilmektedir.

Yazında kritik hastalık sigortalarını ele alan ilk çalışma Dash ve Grimshaw [4] tarafından yapılmıştır. Yazarlar kritik hastalık sigortalarının poliçe şartlarını açıkladıktan sonra ele aldıkları kritik hastalıkların teşhisine yönelik detaylara yer vermişlerdir. Örneğin, kalp krizi teşhisinde şu üç şart bulunmalıdır:

- a. Tipik bir göğüs ağrısı geçmişi

- b. Yeni elektrokardiyografik deęişimler
- c. Kardiyak enzimlerde yükseliş

Bu tezin yazıldığı tarihte güncel poliçe şartlarından alınan kalp krizi teşhisinde bulunması gereken şartlar Dash ve Grimshaw'un [4] çalışmasında verilen o dönemde geçerli şartlara göre açıkça deęişkenlik göstermektedir. Legal and General şirketinin sunduęu bir kritik hastalık sigortası ürününün poliçe şartlarındaki kalp krizi tanımı örnek olarak aşağıda verilmiştir.

- a. Kan dolaşımının yetersiz olması nedeniyle kalp kası ölümü sonucunda aşağıdaki akut miyokard infarktüsü kanıtlarının hepsinin bulunması:
 - Troponinler veya enzimler gibi biyokimyasal kardiyak spesifik belirteçlerin karakteristik artışı
 - Tanısal görüntüleme testlerinde yeni karakteristik elektrokardiyografik deęişiklikler veya dięer pozitif bulgular
- b. Akut miyokard enfarktüsüne yönelik kesin kanıt bulunması
- c. Aşağıdaki durumlar kapsam dışı bırakılmıştır:
 - Dięer akut koroner sendromlar
 - Miyokard enfarktüsü olmayan anjin

2.2.2. Problem 2: Kısıtlı Veri

Kritik hastalık poliçelerinin dięer sigorta ürünlerine oranla daha yeni olması nedeniyle yeterli boyutta sigortalı veri havuzu bulunmamaktadır. Dash ve Grimsaw [4] kritik hastalık sigortası ürünleri için baz (temel) prim belirlemenin zorlukları ve elde henüz yeterince veri, bilgi ve deneyim olmadığından bahsetmişlerdir. Çalışmada, İngiltere için sigortalı verisi olmadığından popölasyona ilişkin verilerden faydalanarak kanser, kalp krizi ve felç hastalıkları için görülme oranlarını tahmin etmişlerdir.

Öncelenmiş menfaat poliçelerinin maliyetini hesaplamak için yazarlar, kritik hastalığa sahip bireylerin hastalık dışı nedenlerle ölüm oranlarının, kritik hastalığa sahip olmayan insanların ölüm oranları ile eşit olacağını varsaymışlardır. Buna göre kritik hastalığın ölüm oranı üzerindeki ekstra maliyeti şu şekilde verilebilir:

$$i_x - k_x q_x$$

Formülde i_x kritik hastalığın görülme oranı, k_x popülasyonda kritik hastalık nedenli ölümlerin oranı ve q_x mortalite oranını göstermektedir. Çalışmada kullanılan tüm oranlar popülasyon verisinden doğrudan hesaplanan kaba oranlardır (crude rates). Diğer bir deyişle herhangi bir istatistiksel model önermemişlerdir.

Dash ve Grimshaw [4]'ın kritik hastalık sigortalarına yönelik çalışmaları daha sonra Dinani ve ark.[1]'nin çalışmasına temel oluşturmaktadır. Dinani ve ark.[1]'nin çalışmasında 1993 ve 1994 yıllarını kapsayan İngiltere popülasyonuna ait verileri kullanarak, kritik hastalık sigortalarının fiyatlandırılmasında da kullanılabilecek hastalık görülme oranları için tabloların oluşturulması amaçlanmaktadır. Ayrıca çalışmada kritik hastalıklar için geliştirilen ürünlerin pazar araştırması ve pazardaki gelişimi de raporlanmıştır.

Dinani ve ark.[1]'nin oluşturduğu kritik hastalık tabloları 20 ve 80 yaş aralığındaki kişileri kapsamaktadır. Kaba görülme oranı, vaka sayısının ilgili popülasyona bölünmesi ile bulunmuştur. Yaş değişkenine göre hastalık görülme oranları 5 yıllık hareketli ortalamalar kullanılarak düzeltilmiştir. Öncelenmiş mefaatli kritik hastalık sigortasının söz konusu olduğu durumda, görülme oranlarının tahmininde Dash ve Grimshaw [4] modeli kullanılmıştır. Çalışmada kritik hastalıklara ilişkin hasarların %80'nin kanser, kalp krizi ve felç olduğu raporlanmıştır. Bu hastalıklar kritik hastalık sigortası poliçeleri kapsamında teminat altına alınan temel üç hastalıktır.

Kritik hastalıklara yönelik istatistiksel modeller, Macdonald ve ark. [5,6]'nın yaptığı çalışmalar ile ortaya çıkmıştır. Temel olarak bu çalışmalarda da çoklu azalım modelinin yanında Dash ve Grimshaw'un [4] çalışmasında kullanılan kaba oranlar yerine gözlem verisine uygun fonksiyonel formlar önermişlerdir. Temelinde her biri birer olasılık dağılımı olan bu fonksiyonel formlar sayesinde durumlar arası kesikli oranlar yerine, anlık geçiş hızlarını (intensity rates) elde etmişlerdir. Bu sayede prim hesaplaması

yapılırken o dönemde yeni bir teknik olarak kabul edilen Thiele denklemlerinden faydalanmışlardır.

Kritik hastalık sigorta poliçelerinin fiyatlandırılmasında, anlık geçiş hızlarının kullanılmaya başlanması teknik anlamda önemli bir ilerleme olsa da yapılan çalışmalarda kullanılan verinin farklı kaynaklardan derlendiği görülmektedir. Örneğin, Gutierrez ve Macdonald [7] çalışmalarında, ele aldıkları her bir kritik hastalık ile ilgili daha önce yapılmış tıp makalelerinden faydalanarak anlık geçiş hızlarını hesaplamışlardır.

Macdonald ve ark. [5,8] meme kanseri ve yumurtalık kanseri için anlık geçiş hızını hesaplarken ağırlıklı olarak İngiltere popülasyon verisini kullanmışlardır. Çalışmada temel olarak meme kanseri, yumurtalık kanseri ve kritik hastalık sigorta poliçesi kapsamında hasar talebine neden olabilecek hastalıklar için bir Markov modeli sunulmuştur. Meme kanseri ve yumurtalık kanseri dışındaki kritik hastalıkların görülme oranları Dinani ve ark. [1]'nin 1991-1997 yılları arasında gerçekleştirilen çalışmasından alınmıştır.

Lu ve ark.[9]'nın çalışmasında, çeşitli hastalıkların görülme oranlarının tahminine dayalı çalışmalarda çoğunlukla küçük örneklemle çalışılmak zorunda kalındığı ve bunun da istatistiksel olarak belirsizliğe ve bu belirsizliğin de geniş güven aralıklarına neden olduğundan bahsedilmektedir. Verinin kısıtlı olması ve orijinal veriye ulaşmanın zorluklarından bahseden yazarlar, iki farklı çalışmadan aldıkları riske maruz birey sayısı ve her gözlem zamanındaki olay sayısı ile simülasyon yöntemi kullanarak örneklem oluşturmuşlardır. Simülasyon sonucu elde ettikleri örneklemde yola çıkılarak primler için örneklem dağılımı tahmin edilmiştir [9].

İtalya'da yapılan bir çalışmada, aktüeryal uygulamalarda, sağlık sigortası için doğru bir modelin oluşturulabilmesi için çok çeşitli verilerin gerekliliğinden ve İtalya'daki sağlık sigortası piyasasının dar olması, ulusal sağlık istatistiklerinin de yetersiz ve genel popülasyona ilişkin olması nedenleriyle bu verilere erişimin mümkün olmaması problemlerine değinilmektedir [10].

Kritik hastalık sigortasına konu olan hastalıkların tanısına yönelik yeni tekniklerin geliştirilmesi ve çeşitlenmesi poliçelerin fiyatlandırılmasında kullanılacak homojen bir veri havuzunun oluşmasını engellemektedir. Bu problemin çözümü için popülasyon bazında bir veriye ihtiyaç duyulmaktadır. Günümüzde, Türkiye’de de kullanılan merkezi sağlık veri tabanlarının bulunması ve bu veri tabanlarında hastalığın teşhisine yönelik yapılan testlerin ayrıntılı dökümlerinin bulunması bu problemlerin çözülmesine yardımcı olacaktır.

2.2.3. Problem 3: Risk gruplarının belirlenmesi

Kritik hastalık sigortasının fiyatlandırılmasına ilişkin yapılan ilk çalışmalarda popülasyon verisi kullanılmış ve bu veri yaş ve cinsiyet dışında herhangi bir grupta yapılmadan analiz edilmiştir [1,4]. Ancak Dash ve Grimshaw [4] eş değişkenlerin homojen risk grupları için ayrı hesaplamaların yapılması gerektiğinden bahsetmişlerdir. Kritik hastalık poliçelerinin fiyatlandırılması için kullanılan verideki farklı risk gruplarının belirlenmesi ve popülasyonun alt kümesi olan bu risk grupları için ayrı hesaplamalar ilk olarak genetik bilginin kullanıldığı çalışmalarda görülmektedir [5,7,8,11,12].

Genetik bilginin kullanıldığı kritik hastalık sigortasına yönelik çalışmalar, tek gen hastalıkları yani tek bir gendeki mutasyonun yüksek olasılıkla belirli bir hastalığın başlamasına neden olduğu durumlar üzerine yoğunlaşmıştır. Macdonald ve Yang [11] bu hastalıklar için yapılacak genetik testler sayesinde, popülasyonun genetik mutasyon taşıyıcıları ve taşıyıcı olmayanlar gibi iki veya daha fazla gruba ayrılarak modellenebileceğini ve bu alt grupların modellenmesi için çok durumlu modellerin uygun olacağını belirtmektedir. Yazarlar kritik hastalık sigortası poliçelerinin fiyatlandırılmasında kullanılabilecek dört durumdan oluşan çok durumlu bir model önermektedir. Bu modelde sağlıklı durumdan, genetik bozukluğun yol açtığı kritik bir hastalığın başladığı duruma, başka bir kritik hastalığın başladığı duruma ve ölüm durumuna geçişler tanımlanmıştır. Bu çok durumlu, sürekli zaman modelindeki anlık geçiş hızı yaşın bir fonksiyonu olarak tanımlanmaktadır. Herhangi bir genotip için, herhangi bir yaştaki bireyin, kritik hastalık sigortası kapsamında ödeyeceği net primler, Thiele denklemlerinin Runge-Kutta algoritması ile çözülmesiyle hesaplanmıştır.

Macdonald ve Yang [11]'in genetik mutasyon taşıyıcıları açısından kritik hastalık sigortası fiyatlandırması için geliştirdiği çok durumlu model yapısının örnekleri aynı yıl yapılan Macdonald, Waters ve Wekwete [5,8], Gutierrez ve Macdonald [7]'in çalışmalarda da görülmektedir. Macdonald, Waters ve Wekwete [5,8] çalışmasında bu çerçevede meme kanseri, yumurtalık kanseri ve kritik hastalık sigortası poliçesi kapsamında hasar tazmin talebine neden olabilecek diğer durumlar için çok durumlu bir model sunulmuş ve İngiltere popülasyon verisi kullanılarak durumlar arası anlık geçiş hızları hesaplanmıştır. Gutierrez ve Macdonald [7] ise aynı yöntemi kullanarak, böbrek yetmezliğine yol açan polikistik böbrek hastalığı ve bu hastalığa yol açtığı belirtilen genetik mutasyonların etkilerini incelemiştir. Sağlıklı durumdan böbrek yetmezliği durumuna, diğer kritik hastalıklara ve ölüm durumuna geçişler tanımlanmıştır. Durumlar arası anlık geçiş hızı, aile geçmişinde hastalığın bulunmadığı, aile geçmişinde hastalığın bulunduğu ama genetik mutasyonun olmadığı, aile geçmişinde hastalığın bulunduğu ve genetik mutasyonun olduğu alt gruplar için ayrı ayrı hesaplanmıştır. Böylece, böbrek yetmezliğine yol açtığı bilinen genetik mutasyonun varlığının bilindiği durumda veya aile geçmişinde böbrek hastalığı bulunan bir bireyin kritik hastalık sigortası poliçesi satın alması durumunda ödemesi gereken ekstra prim hesaplanmıştır. Ayrıca genetik bilgi olmamasının sigortacı açısından oluşturabileceği maliyet de değerlendirilmiştir.

Yazında, kritik hastalık sigortalarına yönelik en kapsamlı çalışmalardan biri Ozkok [13]'ün tez çalışmasında görülmektedir. Bu çalışmada kritik hastalık sigortalarında hasara neden olabilecek hastalıkların görülme oranlarının tahmini için stokastik bir model önerilmektedir. İngiltere'de bulunan Continuous Mortality Investigation (CMI) adındaki kuruluşun sağladığı veri kullanılarak, 1999-2005 yılları arasında gerçekleşen hasarlar alınarak oluşturulan modelde, öncelikle hastalığın teşhisi ve hasarın ödenmesi arasında geçen sürede yaşanan gecikmenin dağılımı için ayrı bir model önerilmiş ve kayıp değerlerin tahmininde kullanılmıştır. Tamamlanan veri ile Genelleştirilmiş Doğrusal Model kullanılarak Poisson hataları ile kaba oranlar modellenmiştir.

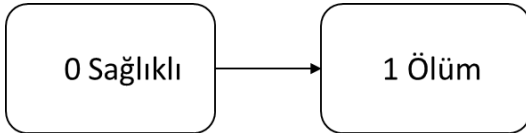
Genel olarak yazında kritik hastalık sigortalarının modellenmesinde çok durumlu stokastik modellerin kullanıldığı görülmektedir. Bir sonraki kesimde aktüerya alanında

sağlık sigortalarına ilişkin problemlerde kullanılan çok durumlu model yapıları verilmiştir.

2.3. Çok Durumlu Modeller

Aktüerya alanında yer alan birçok problem çok durumlu modeller ile kurgulanabilmektedir ve bu sayede özellikle sigortaların modellenmesine ilişkin literatürde geniş yer bulduğu görülmektedir. Günümüzde sunulan sigorta poliçeleri sadece ölüm ve yaşam durumunu içermemekte, emeklilik, hastalık veya maluliyet gibi durumlar için de çeşitli poliçeler sunmaktadır. Ayrıca birey yaşam süresi boyunca farklı sağlık durumlarında bulunabileceğinden ve bu durumlar arası geçişlerin yaşanması olası olduğundan çok durumlu modeller bu tür problemlerin modellenmesinde uygun görülmektedir[14].

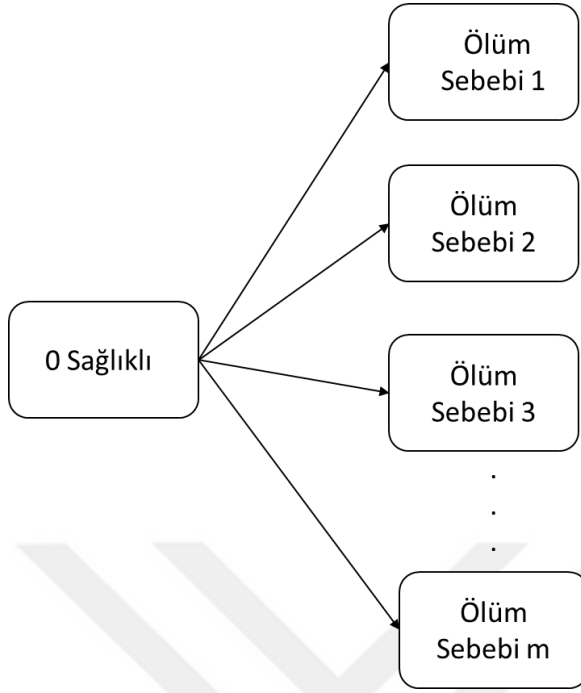
Hayat sigortaları ile ilgili hesaplamalarda kullanılan bir model temel olarak iki durumlu bir model olarak değerlendirilebilir. Şekil 2.1’de verilen modelde birey herhangi bir anda yaşam (0) ve ölüm (1) durumlarından birinde bulunabilir.



Şekil 2.1. İki durumlu Ölüm Modeli

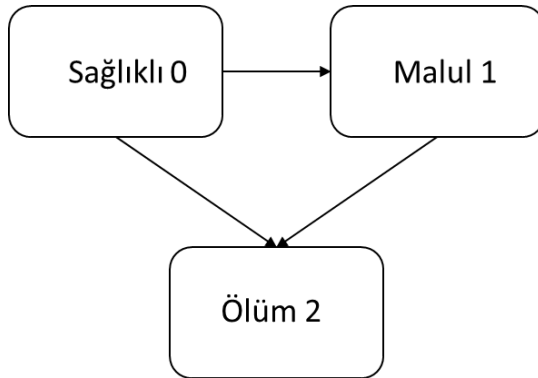
Bu modelde durum (1) yutucu durum olarak adlandırılır ve kişi yutucu duruma vardığında bu durumu asla terk edemez. Bu nedenle herhangi bir birey için düşünüldüğünde 0 durumundan 1 durumuna birden fazla geçiş yapılamaz. Bu durumda buradaki geçiş oranı ölüm oranını verecektir.

Bir diğer çok durumlu model örneği Şekil 2.2’de verilen çoklu azalım modelidir. Süreç şekilde gösterildiği gibi 0 durumu ile başlar ve m sayıdaki yutan durumlardan herhangi birine geçilmesi ile sonlanır. Genellikle kritik hastalık sigortaları hastalığın teşhisi ile sonlandığı için çoklu azalım modelleri kullanılmaktadır.



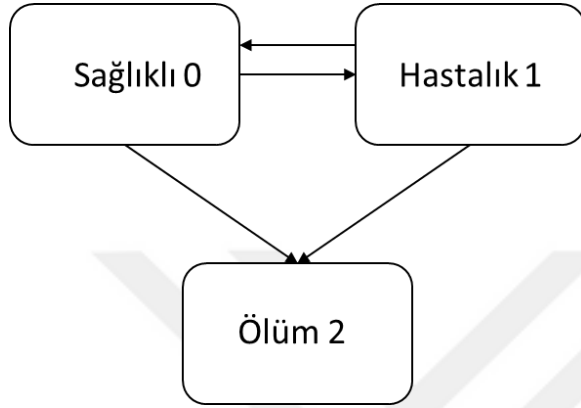
Şekil 2.2. Çoklu Azalım Modeli

Maluliyet için geliştirilen modeller en basit şekliyle 3 durum (sağlıklı, malul, ölüm) ile tanımlanmıştır. Aşağıda verilen Şekil 2.3'te bu modelin temel gösterimi yer almaktadır. Bu modelde maluliyet durumundan sağlıklı duruma geçişin olmadığı varsayılmaktadır. Diğer bir deyişle maluliyet durumunun kalıcı olduğu yapıları temsil etmektedir.



Şekil 2.3. Kalıcı Maluliyet Modeli

Sağlık sigortaları alanında kullanılan bir diğer model ise, Şekil 2.4'te verilen üç durumlu hastalık-ölüm modelidir. Bu modelde özellikle hastalıktan iyileşmenin mümkün olduğu durum göz önünde bulundurulmaktadır. Bu yapı çok durumlu modellerin Şekil 2.3'te verilen modelde görülmeyen önemli bir özelliğini göstermektedir. Bu yapıda, bir ya da daha fazla durumda çok kere bulunulabileceği olasılığı dikkate alınmaktadır.



Şekil 2.4. Hastalık Ölüm Modeli

2.4. Kritik Hastalıkların Stokastik Modellenmesi

Bu kesimde, yazında kritik hastalıkların modellenmesinde yaygın olarak kullanılan Markov çok durumlu modelleri ile ilgili tanımlar, varsayımlar ve tahmin yöntemleri özetlenmiştir. Markov Modelleri özetlenirken, özellikle Macdonald, Richards ve Currie [15] tarafından yazılan *Modelling Mortality with Actuarial Applications* isimli kitabın 14. bölümünde sunulan gösterimler ve Dickson, Hardy ve Waters [14] tarafından yazılan *Actuarial Mathematics for Life Contingent Risks* isimli kitabının 8. bölümünde sunulan gösterimler kullanılmıştır.

2.4.1. Markov Süreçleri

Sonlu $n+1$ durumlu bir Markov sürecinde, $0,1,2,\dots,n$ olarak isimlendirilen durumlar, bireyler için farklı durumları ifade etmektedir. Her $t \geq 0$ için $Y(t) = i$, bireyin $x + t$ zamanında i durumunda olduğunu gösterir.

Varsayım 1: Herhangi bir j durumu, t ve $s \geq 0$ olmak üzere $\Pr[Y(t+s) = j | Y(u), u \leq t] = \Pr[Y(t+s) = j | Y(t)]$ koşulu sağlanır. Diğer bir deyişle, sürecin $t+s$ anındaki pozisyonunun, sürecin t anına kadarki bütün yörüngesine göre koşullu dağılımı, aynı pozisyonun sürecin t anındaki pozisyonuna göre koşullu dağılımına eşittir.

Markov varsayımı olarak da adlandırılan bu varsayım aynı zamanda Markov modelinin sınırlılığını göstermektedir. Özellikle hastalıkların modellenmesinde, hastalığın ne kadar süredir devam ettiği bireyin hastalık durumunda ne kadar süre geçirdiği, hastalığın seviyesi ve iyileşme olasılığı üzerinde etkisi olabileceği düşünüldüğünde Markov varsayımı modele sınırlılık getirmektedir.

Varsayım 2: Her pozitif h zaman aralığı için,

$$\Pr[h \text{ aralığında iki veya daha fazla geçiş olması}] = o(h)$$

özellğine sahiptir. Daha açık bir ifadeyle, h 'nin herhangi bir fonksiyonu olan $g(h)$ fonksiyonu $o(h)$ özelliğine,

$$\lim_{h \rightarrow 0} \frac{g(h)}{h} = 0$$

durumunda sahip olur. Diğer bir deyişle, h sifıra yaklaşırken $g(h)$ sifıra daha hızlı yaklaşır. Bu varsayıma göre, kısa zaman aralıklarında iki veya daha fazla geçiş olma olasılığı ihmal edilebilecek kadar düşüktür.

Aktüerya alanında standart olarak kullanılan gösterimler, ${}_t p_x$, ${}_t q_x$ ve μ_x çok durumlu modellere göre uyarlanıp ve genelleştirilerek geçiş olasılıklarının gösteriminde kullanılmaktadır [14,15].

Gösterim: i ve j durumları ve $x, t \geq 0$ için,

$${}_t p_x^{ij} = \Pr[Y(x+t) = j | Y(x) = i]$$

$${}_t \bar{p}_x^{ii} = \Pr[Y(x+s) = i, \forall s \in [0, t] | Y(x) = i]$$

Bu tanımlarla, ${}_t p_x^{ij}$ gösterimi x yaşında i durumunda olan birinin $x + t$ yaşında j durumunda olma olasılığını gösterirken ${}_t \bar{p}_x^{ii}$ gösterimi x yaşında i durumunda olan birinin $x + t$ yaşına kadar i durumundan hiç ayrılmadığını ifade eder. ${}_t p_x^{ii}$, bir kişinin t yıl sonra tekrar i durumunda olma olasılığını ifade ettiği için, ${}_t p_x^{ii} \geq {}_t \bar{p}_x^{ii}$ ilişkisi elde edilir. Varsayım 2 kullanılarak;

$${}_t p_x^{ii} = {}_t \bar{p}_x^{ii} + o(t) \quad (1)$$

yazılabilir.

$i \neq j$ durumu için ise, tanımlanan geçiş olasılığına anlık geçiş hızı adı verilir ve

$$\mu_x^{ij} = \lim_{h \rightarrow 0^+} \frac{{}_h p_x^{ij}}{h}$$

olarak tanımlanmaktadır.

Varsayım 3: Tüm durumlar için i ve j ve $x \geq 0$ olan tüm yaşlar için, ${}_t p_x^{ij}$ 'nin t noktasında türevlenebilir olduğu varsayılmaktadır. Bu varsayım, yukarıda tanımlanan μ_x^{ij} limitinin olduğunu garanti eder. Bu durumda μ_x^{ij} , x 'in sınırlı ve integrallenebilir bir fonksiyonudur.

Varsayım 2 kullanılarak, anlık geçiş hızı μ_x^{ij} , $h > 0$ için aşağıdaki şekilde yazılabilir.

$${}_h p_x^{ij} = h \mu_x^{ij} + o(h) \quad (2)$$

Eşitlik (2) ve varsayım 2 kullanılarak, çok durumlu modellerde $h > 0$ için ${}_h \bar{p}_x^{ii} = 1 - \sum_{j=0, j \neq i}^n {}_h p_x^{ij}$ den yola çıkarak,

$${}_h \bar{p}_x^{ii} = 1 - h \sum_{j=0, j \neq i}^n \mu_x^{ij} + o(h) \quad (3)$$

eşitliği yazılabilir.

Tüm olasılıklar anlık geçiş hızı cinsinden tanımlanabilir olduğundan, bu ilişki çok durumlu Markov modellerine de uyarlanabilmektedir. $n+1$ durum içeren bir Markov modelinde herhangi bir i durumu için,

$${}_t p_x^{\bar{ii}} = \exp \left(- \int_0^t \sum_{j=0, j \neq i}^n \mu_{x+s}^{ij} ds \right) \quad (4)$$

eşitliği yazılabilir.

2.4.2. Kolmogorov İleri Denklemleri

Herhangi i ve j durumları ve $x, t, h \geq 0$ için,

$${}_{t+h} p_x^{ij} = {}_t p_x^{ij} {}_h p_{x+t}^{jj} + \sum_{k=0, k \neq j}^n {}_t p_x^{ik} {}_h p_{x+t}^{kj}$$

eşitliği x yaşında bireyin i durumundayken $t + h$ zaman sonra j durumuna geçeceği yolları ikiye ayırır;

1. t zaman sonra j durumuna gelip geri kalan h zamanda bu durumdan çıkmadığı (sağdaki ilk ifade)
2. t zaman sonra j dışında bir durumda olup, h zaman sonra herhangi bir rotayı takip ederek j durumuna geldiği (sağdaki ikinci ifade).

Eşitlik (1), (2) ve sonrasında (3) kullanılarak,

$$\begin{aligned} {}_{t+h} p_x^{ij} &= {}_t p_x^{ij} \left({}_h p_{x+t}^{\bar{jj}} + o(h) \right) + \sum_{k=0, k \neq j}^n {}_t p_x^{ik} {}_h p_{x+t}^{kj} \\ &= {}_t p_x^{ij} \left(1 - h \cdot \sum_{k=0, k \neq j}^n \mu_{x+t}^{jk} - o(h) \right) + \sum_{k=0, k \neq j}^n {}_t p_x^{ik} {}_h p_{x+t}^{kj} \\ &= {}_t p_x^{ij} \left(1 - h \cdot \sum_{k=0, k \neq j}^n \mu_{x+t}^{jk} - o(h) \right) + h \cdot \sum_{k=0, k \neq j}^n {}_t p_x^{ik} \mu_{x+t}^{kj} + o(h) \end{aligned}$$

elde edilir.

Terimler düzenlendiğinde,

$${}_{t+h}p_x^{ij} = {}_tp_x^{ij} - h \sum_{k=0, k \neq j}^n ({}_tp_x^{ij} \mu_{x+t}^{jk} - {}_tp_x^{ik} \mu_{x+t}^{kj}) + o(h) \quad (5)$$

biçiminde yazılabilir. Kolmogorov denklemlerini elde etmek amacıyla Eşitlik (5) için türev tanımından yararlanılarak

$$\frac{{}_{t+h}p_x^{ij} - {}_tp_x^{ij}}{h} = \sum_{k=0, k \neq j}^n ({}_tp_x^{ik} \mu_{x+t}^{kj} - {}_tp_x^{ij} \mu_{x+t}^{jk}) + o(h)$$

ifadesi yazılır ve $h \rightarrow 0$ için limit alındığında,

$$\frac{d}{dt} {}_tp_x^{ij} = \sum_{k=0, k \neq j}^n ({}_tp_x^{ik} \mu_{x+t}^{kj} - {}_tp_x^{ij} \mu_{x+t}^{jk}) \quad (6)$$

olur. Çok durumlu Markov modelinde anlık geçiş hızları (anlık geçiş hızı matrisi) biliniyorsa Kolmogorov İleri Denklemleri ile geçiş olasılıkları hesaplanabilir.

2.4.3. Thiele Diferansiyel Denklemleri

Markov süreçlerinin bir çıktısı olan, anlık geçiş hızlarının olasılığa dönüştürülmeden prim hesaplamalarında kullanılması ise Thiele diferansiyel denklemleri sayesinde mümkün olmaktadır. Bu bölümde prim hesaplanması için izlenecek yöntem ve gösterim [14,15] verilmiştir.

x yaşındaki bir birey için düzenlenen sigorta poliçesinin $x + t$ yaşındaki poliçe değeri(rezerv) ${}_tV_x$ olsun. $[t, t + 1)$ yılının başında eğer sigortalı hayatta ise, P_t prim ödemesi yapılır, eğer birey yıl içinde ölürse, yıl sonunda toplam tazminat S_t poliçenin lehdarlarına ödenir. Yıllık faiz oranının da i olduğunu varsayıldığında; yıl sonundaki durum

$$({}_tV_x + P_t)(1 + i) = q_{x+t}S_t + p_{x+t}{}_tV_x$$

eşitliği ile yazılır. Bu denklem tekrar düzenlendiğinde bir fark denklemine dönüşür:

$${}_{t+1}V_x - {}_tV_x = i {}_tV_x + (1 + i)P_t - q_{x+t}(S_t - {}_{t+1}V_x)$$

Yani, poliçe değerindeki artış; birikmiş prim ile poliçe değeri üzerinden kazanılan faizden ölüm dolayısıyla beklenen kayıpların çıkarılması ile elde edilebilir. Yazında $(S_t - {}_{t+1}V_x)$

değeri riske maruz toplam (sum-at-risk) olarak ifade edilmektedir. Bu fark denklemi, Thiele diferansiyel denklemlerine genelleme yapılabilmesi nedeniyle, kesikli ve sürekli zaman modelleri arasındaki bağlantıyı verir. Bir sigorta poliçesi için primlerin sürekli zamanda yıllık P olarak ödendiği ve δ nın anlık faiz hızı olduğu durumda eşitlik,

$${}_{t+dt}V_x - {}_tV_x = \delta_t {}_tV_x dt + (1 + \delta dt)P_t dt - \mu_{x+t}dt(S_t - {}_{t+dt}V_x)$$

biçiminde yazılır. Her iki tarafın dt ile bölünmesi ve t 'ye göre limitinin alınması sonucu Thiele'nin diferansiyel denklemi elde edilir.

$$\frac{d}{dt} {}_tV_x = \delta_t {}_tV_x + P - \mu_{x+t}(S_t - {}_tV_x)$$

Genel olarak $m+1$ durumlu Markov Modelinde, kişi i durumunda iken prim b_t^i oranıyla sürekli zamanda ödenmekte ve b_t^{ij} poliçe tazminatı i durumundan j durumuna geçiş anında ödenmektedir. Markov Modelinde i durumunda olan bir bireyin t zamanındaki muhtemel poliçe değeri, ${}_tV_x^i$ ve t zamanındaki sürekli faiz oranı δ_t olmak üzere Thiele'nin diferansiyel denklemi aşağıdaki gibi yazılabilir.

$$\frac{\partial}{\partial t} {}_tV_x^i = \delta_t {}_tV_x^i + b_t^i - \sum_{j \neq i} \mu_{x+t}^{ij} (b_t^{ij} + {}_tV_x^j - {}_tV_x^i)$$

Thiele denklemleri kritik hastalıklar için teminat verilen bir poliçenin değerini ve priminin hesaplanmasında kullanılabilir.

2.4.4. Poisson Regresyonu

Modellenmek istenen veri kesikli (count) olduğunda, genellikle Poisson regresyonu kullanılmaktadır. Poisson regresyon modeli aşağıdaki gibi ifade edilebilir.

$$y \sim P(\mu) \quad g(\mu) = x' \beta$$

Bağ fonksiyonu $g(\mu)$ için en çok tercih edilen formlar $\mu = x' \beta$ ve $\ln \mu = x' \beta$ 'dir. Logaritmik bağ fonksiyonu tercih edildiğinde $\hat{\mu} = \exp(x' \hat{\beta})$ pozitif olmaktadır. $\mu = x' \beta$ bağ fonksiyonu ile pozitif olma şartı sağlanamamaktır.

Tek açıklayıcı değişken (x) içeren bir model için

$$x = (1, x)' \quad \beta = (\beta_0, \beta_1)' \quad g(\mu) = \beta_0 + \beta_1 x_1$$

olur. Özdeş bağ fonksiyonu ile (identity link), x_1 değişkeninin değerindeki değişimin etkisi, toplamsal (additive) olmaktadır. x_1 değişkeninin değerindeki bir birimlik artış ile y değişkeninin beklenen değerindeki artış,

$$\{\beta_0 + \beta_1(x_1 + 1)\} - \{\beta_0 + \beta_1(x_1)\} = \beta_1$$

eşitliği ile gerçekleşir.

Logaritmik bağ fonksiyonu ile (log link), y değişkeninin beklenen değeri $\mu = \exp(\beta_0 + \beta_1 x_1)$ olur ve x_1 değişkeninin değerindeki bir birimlik artış ile y değişkeninin beklenen değerindeki etkisi,

$$\exp\{\beta_0 + \beta_1(x_1 + 1)\} = \exp\{\beta_0 + \beta_1(x_1)\} \exp(\beta_1)$$

şeklinde gerçekleşir. Bu durumda çarpımsal (multiplicative) etkiden söz edilebilir.

Jong ve Heller [16] kitabında, Avustralya'da 2002 yılında diyabet hastalığına bağlı ölümlerin modellenmesinde Poisson regresyonu örneğini vermişlerdir. Bu modelde y gerçekleşen ölüm sayısını, μ ölüm sayısının beklenen değerini, x değişkeni yaş ve cinsiyet etkisi vektörünü ve n popülasyon sayısını göstermektedir. Buna göre diyabet hastalığına bağlı ölümler için aşağıdaki model önerilmektedir.

$$y \sim P(\mu), \quad \ln \mu = \ln n + x' \beta, \quad x = (1, x_1, x_2, \dots, x_p)'$$

x_1 değişkeni cinsiyeti, $x_2, \dots, x_p$ değişkenleri ise belirlenen yaş aralıklarını ifade etmektedir. Benzer şekilde bu çalışmada sunulan analizlerde de anlık geçiş hızı tahmininde Poisson regresyon kullanılmış ve yukarıda Jong ve Heller [16] tarafından varsayılan modele benzer bir model varsayımı yapılmıştır.

3. VERİ

Bu bölümde tez çalışmasında kullanılan veri tanıtılmış, verinin kullanıldığı yazındaki diğer çalışmalar incelenmiş ve tartışılmıştır. Bu bölümün ikinci kısmında (Kısım 3.2) tez çalışmasında incelenen değişkenler tanıtılmış ve dağılımlarına ilişkin grafikleri verilmiştir.

3.1. Sağlık ve Emeklilik Çalışması Verileri (HRS)

Sağlık ve Emeklilik Çalışması (The Health and Retirement Study- HRS), Michigan Üniversitesi'ndeki Sosyal Araştırmalar Enstitüsü tarafından yürütülen boylamsal hane halkı araştırmasıdır. Ulusal Yaşlanma Enstitüsü (National Institute on Aging- NIA) tarafından da desteklenen HRS, Birleşik Devletlerde yaşayan hane halkının yaşlanmasıyla ilgili birçok farklı yönün araştırılması için araştırmacılara zengin ve karmaşık veriler sunmaktadır.

Çalışma 1992 yılında başlamış ve günümüze kadar her iki yılda bir çalışmaya katılan yetişkin bireyler ile görüşmeler gerçekleştirilmiştir. Katılımcı grubu, kurum özelinde olmayan (non-institutionalized) çoğunlukla 50 yaş ve üzerindeki bireyler ve eşlerinden oluşmaktadır. Çalışmanın ana hedefi, emeklilik, sağlık sigortası, tasarruf ve ekonomik refah konularındaki ulusal politikalarla ilgili araştırmalara panel veri seti sunmaktadır [17]. Bu amaçla sunulan veride, demografik, gelir, varlıklar, sağlık, biliş, aile yapısı ve bağlantılar, sağlık harcamaları, yaşam koşulları, mesleki durum, beklentiler ve sigorta poliçeleri ile ilgili değişkenler bulunmaktadır. Ayrıca veri seti yedi farklı grup (cohort) içermektedir.

Çizelge 3.1. RAND HRS içerisinde yer alan veri grupları ve özellikleri

Grup Adı	Doğum Yılı	İlk Gözlem Yılı	Gözlem Sıklığı
HRS	1931-1941	1992	2 yılda bir
AHEAD	1924 öncesi	1993	1993, 1995, 1998, izleyen 2 yılda bir
CODA	1924-1930	1998	2 yılda bir
WB	1942-1947	1998	2 yılda bir
EBB	1948-1953	2004	-
MBB	1954-1959	2010	-
LBB	1960-1965	2016	-

Veri setinde yer alan gruplar ve özellikleri Çizelge 3.1’de verilmiştir. Tez çalışmasında HRS olarak adlandırılan, 1931-1941 tarihleri arasında doğmuş bireylerin yer aldığı gruptaki veriler kullanılmıştır. İlk gözlemin yapıldığı 1992 yılından 2018 yılına kadar 2 yılda bir toplam 14 adet gözlem yapılmış ve bu 14 gözlemi içeren veri seti tez çalışmasında kullanılmıştır. Her bir gözleme katılan birey sayısı ve ortalama, minimum ve maximum yaşları Çizelge 3.2’de verilmiştir.

Çizelge 3.2. RAND HRS içerisinde yer alan bireylerin demografik özellikleri [17]

Gözlem Numarası	Frekans	Ortalama Yaş	Minimum Yaş	Maximum Yaş
1	12652	55,25	23,0	85,0
2	19642	65,13	23,0	103,0
3	17991	66,35	22,0	105,0
4	21384	65,94	25,0	105,0
5	19578	67,13	23,0	107,0
6	18165	68,37	25,0	109,0
7	20129	66,57	24,0	107,0
8	18469	67,99	25,0	104,0
9	17217	69,20	25,0	106,0
10	22034	65,66	18,0	109,0
11	20554	66,85	20,0	103,0
12	18747	67,90	18,0	104,0
13	20912	65,70	21,0	107,0
14	17146	66,99	22,0	104,0

Her bir gözleme katılan bireylerin eş ya da partnerleri de ayrı bir değişken olarak kaydedildiğinden onlara ilişkin veriler de tez çalışmasına dahil edilmiş ve gözlemlenen eşlerin sayısı ve ortalama, minimum ve maximum yaşları aşağıdaki Çizelge 3.3’de verilmiştir.

Çizelge 3.3. RAND HRS içerisinde yer alan eşlerin demografik özellikleri [17]

Gözlem Numarası	Frekans	Ortalama Yaş	Minimum Yaş	Maximum Yaş
1	12652	55,16	23,0	85,0
2	19579	62,66	23,0	103,0
3	17923	63,88	21,0	98,0
4	21318	63,55	25,0	98,0
5	19544	64,69	23,0	100,0
6	18144	66,00	22,0	102,0
7	20118	64,28	27,0	104,0
8	18468	65,72	22,0	99,0
9	17216	66,96	22,0	98,0
10	22028	63,69	18,0	98,0
11	20550	64,89	20,0	99,0
12	18699	65,83	18,0	99,0
13	20768	63,74	18,0	99,0
14	17021	65,12	22,0	102,0

HRS verisi içerisinde yer alan değişkenlerden sağlık ile ilgili değişkenler incelendiğinde, katılımcıların kendi genel sağlık durumlarına ilişkin 1(kötü)-5 (çok iyi) arasındaki puan değerlendirmeleri, geçen 2 yılda hastane ziyaret sıklıkları, son iki yılda bakım evinde kalıp kalmadıkları, son iki yıldaki doktor ziyaret sayıları, son iki yılda evde bakıma ihtiyaç duyup duymadıkları, düzenli kullandığı ilaçları olup olmadığı, son iki yılda yaptıkları sağlık harcamaları, katılımcıların kendi psikolojik durumlarına ilişkin değerlendirmeleri ve uzman bir doktor tarafından teşhis edilen hastalıklara ilişkin değişkenler olduğu görülmektedir. Tez çalışmasında temel amaç kritik hastalık sigortaları için bir model geliştirmek olduğundan, uzman bir doktor tarafından teşhis edilen hastalıklara ilişkin değişken çalışmada kullanılmıştır. Bu değişken incelendiğinde, yüksek tansiyon, diyabet, cilt kanseri dışındaki kanser türleri ya da herhangi kötü huylu tümör, astım ve bronşit dışındaki kronik akciğer hastalığı, kalp krizi, kalp hastalığı, kalp

yetmezliđi, felç, psikiyatrik problemler, romatizma, uyku bozukluđu, bunama ve Alzheimer hastalığına ilişkin verilerin kaydedildiđi görölmektedir. Bu deđiřkende yer alan hastalıklardan, bir kritik hastalık sigorta poliçesinde teminat altına alınabilecek olan temel hastalıklar dikkate alınarak kanser, kalp krizi/kalp hastalığı/kalp yetmezliđi ve felç hastalığına ilişkin veriler dikkate alınmıştır. Kritik hastalık sigortası kapsamında temel olarak kanser, kalp krizi, felç ve koroner bypass ameliyatı teminat altına alınmaktadır. HRS verisinde kalp krizi ve koroner bypass ameliyatı geçirme durumu tek deđiřken altında verildiğinden, bu hastalıklar “kalp hastalığı” adı altındaki tek bir deđiřkende incelenmiştir. Kritik hastalık durumu yanında ayrıca bireylerin ölüm durumları da dikkate alınmış ve incelenmiştir.

Juster ve Suzman [18], HRS çalışmasının yapılmasının temel amacını, veri toplama yöntemlerini, verinin içerdii deđiřkenleri ve verinin toplanması için yapılan çalışmaların bütçesini inceledikleri bir çalışma yapmışlardır. HRS çalışmasında elde edilen verilerin kullanıldığı çalışmalar incelendiğinde, özellikle emeklilik dönemindeki sađlık sigortalarına erişim, sađlık harcamaları ve özellikle yaşlılık ile gelişen maluliyet durumu ve bunlara sunulan sigorta poliçeleri gibi konuların ele alındığı çalışmalar öne çıkmaktadır.

Robinson ve Clark [19], emeklilik döneminde sađlık sigortasına erişimi olan ve olmayan bireylerin emekli olma eğilimleri arasındaki ilişkiyi incelendikleri çalışmada HRS verilerini kullanmış ve Cox Orantısal Hazard modeli ile tahminler yapılmıştır. Kapur ve Rogowski [20] ise emeklilikte sađlık sigortalarına erişimin etkisini çiftlerin her ikisini de düşünerek ele alırken, HRS verisindeki 1992-2002 yılları arasındaki altı gözlemden faydalanmışlardır.

Kim ve Richardson [21] HRS çalışmasındaki verinin 1994-2006 yılları arasındaki gözlemlerini kullanarak, yaşın artması ile birlikte bireylerde azalan fiziksel fonksiyon durumunda özel sađlık sigortalarına erişimi olan ve olmayan bireyler arasındaki farklılık ortaya konmuştur.

Xu ve ark. [22]'nin çalışmasında, 55 ve 64 yaşları arasındaki kadınlarda sağlık sigortalarına erişim ve bunun sağlık hizmetlerini kullanmalarındaki etkisi araştırılırken yalnızca HRS çalışmasının 2002 yılındaki gözlemleri kullanılmıştır.

İleri yaştaki bireylerin sağlık sigortalarına erişimi ve sağlık harcamalarının incelendiği bu çalışmalara ek olarak ayrıca Türkiye'deki sigorta piyasasında henüz sunulmayan uzun süreli bakım sigortası (long-term care insurance) ürünlerinin fiyatlandırılmasına yönelik olarak yapılan birçok çalışmada da HRS çalışmasında elde edilen verilerin kullanıldığı görülmektedir. Jung [23] HRS çalışmasından elde edilen verileri kullanarak, 40-90 yaşları arasında bireylerin sağlık durumlarındaki değişimi incelemiştir. Sağlık durumları arasındaki geçişler tahmin edilirken Markov modeli kurgulanmış ve yaş, cinsiyet, vücut kitle endeksi ve sigara kullanma durumu gibi eş değişkenlerin de etkisine bakılmıştır. Fong, Shao ve Sherris [24] uzun dönem bakım sigortalarını inceledikleri çalışmalarında, ileri yaştaki bireylerde yaşlılıktan kaynaklanan engellilik durumları arasındaki geçiş olasılıklarını tahmin ederken, çok durumlu bir Markov modeli kurgulamışlar ve 1998-2010 yılları arasındaki HRS çalışmasından elde edilen veriyi kullanmışlardır. Sonrasında ise uzun dönem bakım sigorta ürünlerinin fiyatlandırılması üzerine çalışmalarını geliştirmişlerdir [25].

HRS verisi devasa boyuttadır. Toplamda 113,799 adet gözlemden oluşmaktadır. Son olarak yayınlanan on dördüncü gözlem verilerini de içeren veri Şubat 2021 itibarıyla sisteme üye olarak araştırma amacıyla indirilebilmektedir [26]. Tez çalışmasına konu olan sağlık ve hastalıklarla ilgili değişkenler, bunların toplanma yöntemi ve kodlanması konusunda Wallace ve Herzog [27] ilk detaylı raporlamayı vermiştir. Çok fazla değişken ve katılımcı grubu içeren veri herhangi bir araştırmaya konu olmadan önce ön analiz gerektirmektedir. Araştırmacılara veriye daha rahat erişim sağlaması ve çeşitli istatistiksel analiz yazılımlarına uyumlu hale getirilmesi amacıyla RAND Yaşlanma Çalışmaları Merkezi (RAND Center for the Study of Aging) tarafından ham veri düzenlenmiştir.

Veri ile ilgili pek çok farklı format seçeneği sunulmakla birlikte, tercih SPSS versiyonundan yana kullanılmıştır. Bu tercihin nedeni analiz amacıyla kullanılacak R

yazılımı ile SPSS formatı arasındaki transfer kolaylığıdır. Verinin SPSS formatındaki boyutu 742 MB'tır.

Verinin SPSS formatından dönüştürülmesi Intel Core i7-6700HQ 2,60 Ghz. QuadCore 16 GB ram'li bir bilgisayarda yaklaşık 3 dakika almıştır. Daha sonrasında veri yerel R formatına kaydedilmiş ve her defasında bu dosya üzerinde çalıştırılmıştır. Dosyanın boyutu bu şekilde 174 MB'a kadar düşürülmüştür. Her defasında verinin hafızaya yüklenme süresi yaklaşık 20 saniye almaktadır.

RAND HRS verisi içerisinde HRS, AHEAD, CODA, WB, EBB, MBB ve LBB ismi verilen 7 farklı grupta katılımcılar bulunmaktadır. Başlangıçtaki HRS isimli grup 1931 ile 1941 yılları arasında doğmuş olan ve ilk kez 1992 yılında görüşme yapılmış, her iki yılda bir takip eden görüşmeler yapılan bireylerden ve eşlerinden oluşmaktadır. Tez çalışmasında 13,429 bireyden oluşan, HRS isimli veri grubu kullanılmıştır.

Teze konu olan kritik hastalık sigorta poliçesinde teminat altına alınabilecek durumlar seçilerek incelenmiştir. Temel olarak tüm poliçelerde Kanser, Kalp Krizi ve/veya Kalp Hastalığı, Felç hastalıklarından biri veya Ölüm gerçekleşmesi durumları kritik hastalık sigortası kapsamında değerlendirilmektedir. Bundan yola çıkılarak RAND HRS verisinde Kanser, Kalp Krizi ve/veya Kalp Hastalığı, Felç hastalıklarından biri veya Ölüm durumu bilgisini içeren değişkenler kullanılarak tez çalışması için model oluşturulmuştur.

Veri öncelikle çeşitli dönüşümlerden geçirilmiştir. Ham veri, her katılımcının bir satırda olduğu ve yapılan her ankette sorulan tüm soruların ayrı ayrı değişkenler şeklinde sütunlarda olduğu "geniş" veri formatındadır. Kritik Hastalıklara ilişkin modelin kurulabilmesi için kullanılan değişkenler ham veriden ayıklanmıştır. Bu değişkenler veriden filtrelendikten sonra ilk adım; verinin "geniş" formattan "uzun" formata dönüştürülmesidir. Uzun veri formatında, her gözlem ayrı şekilde kodlanan değişkenleri içeren tek bir satırda toplanır. Bu formatta her katılımcıdan farklı zamanlarda toplanan veri tek bir satırda değil, katılımcının her gözleminde verdiği yanıtlar ayrı satırlarda

bulunur. Bu şekilde hazırlanan veriden alınan bir katılımcının değerleri Çizelge 3.4'te verilmiştir.

Çizelge 3.4. RAND HRS verisinden 2010 kodlu katılımcının değerleri

Katılımcı Kodu	Gözlem				Doğum				
	Yılı	Cinsiyet	Yaş	Ölüm	Yılı	Kanser	Akciğer	Kalp	Felç
2010	1992	Kadın	58	0	1934	0	1	0	0
2010	1994	Kadın	60	0	1934	0	1	0	0
2010	1996	Kadın	62	0	1934	0	1	0	0
2010	1998	Kadın	64	0	1934	0	1	0	0
2010	2000	Kadın	66	0	1934	0	1	0	0
2010	2002	Kadın	68	1	1934	0	1	0	0

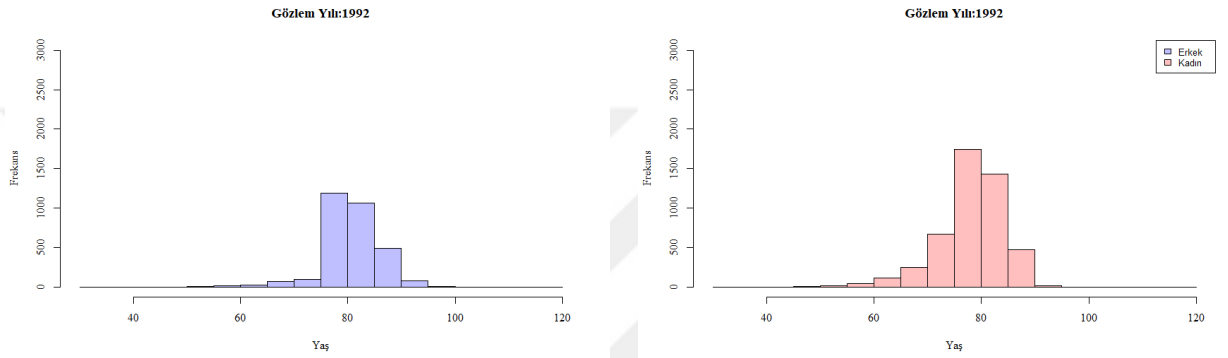
3.2. Açıklayıcı Değişkenler

Bu bölümde tez çalışmasında geliştirilen modellerde kullanılan açıklayıcı değişkenlerin özet istatistikleri verilmiştir. HRS verisi içinde açıklayıcı değişkenler her gözlem yılında tekrar ölçümlenmiştir. Bu durum için istisna sadece cinsiyet değişkenidir.

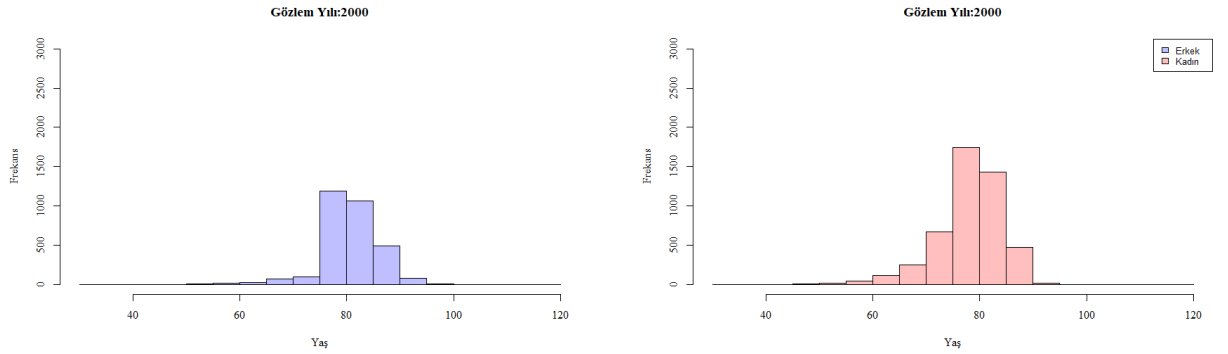
Kritik hastalıkların modellenmesinde ve geliştirilen modeller ile sigorta prim değerleri tahmin edilirken genel olarak cinsiyet ve yaş değişkeninin etkisi düşünülmektedir. Ancak kritik hastalık sigortaları konusunda yapılan bazı önemli çalışmalarda yaş ve cinsiyet değişkeni dışında birçok açıklayıcı değişkenin de yer aldığı görülmektedir. Macdonald, Waters ve Wekwete [28] koroner kalp hastalığı ve felç için yaptıkları kritik hastalık sigortası modelinde, cinsiyet ve yaşın dışında Beden Kitle İndeksi (BKI), sigara kullanma durumu, diyabet hastalığı, kolesterol seviyesi, tansiyon hastalığı ve altı kategoride değerlendirdikleri sosyal sınıf değişkenlerini analizlerine dahil etmişler ve model üzerindeki etkilerini tartışmışlardır. Ozkok ve ark. [29] kritik hastalık sigortalarında hastalığın teşhisi ile hasar bildirimi arasındaki gecikmeyi modelledikleri çalışmalarında, 10 adet açıklayıcı değişken kullanmışlardır. Yaş, cinsiyet, sigorta türü, sigara kullanma

durumu gibi kategorik deęiřkenlerin yanı sıra poliçe süresi, teminat tutarı gibi nicel deęiřkenlerin de deęerlendirildięi görölmektedir.

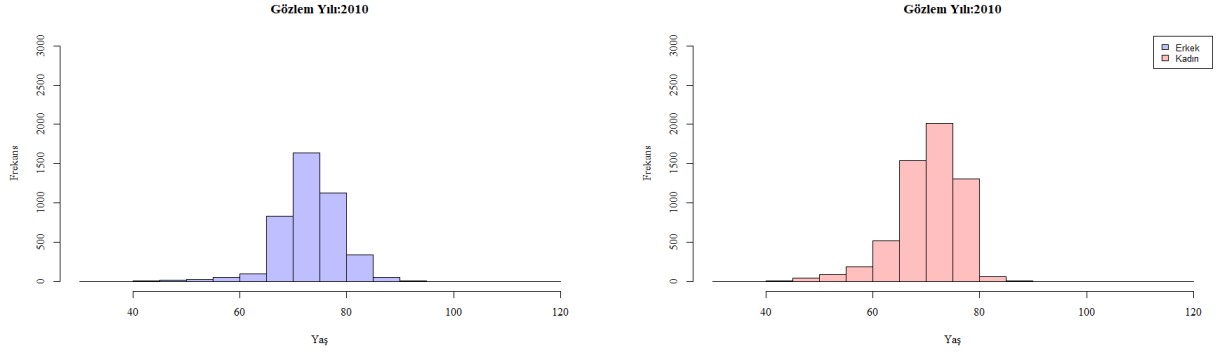
HRS verisinde yer alan bireylerin yařlarına göre daęılımını gösteren histogram grafikleri 1992, 2000, 2010 ve 2018 gözlem yılları için ayrı ayrı verilmiřtir. Bu grafiklerde ayrıca cinsiyet kırılımı da renk vurgusuyla görölmektedir. Buna göre kadın gözlemlerin erkeklere göre daha genç yařta olduęu ve tüm gözlem yılları incelendiğinde, beklenen řekilde eldeki örnekleimde yer alan bireylerin yıllar geçtikçe yařlandığı gözlemlenebilir.



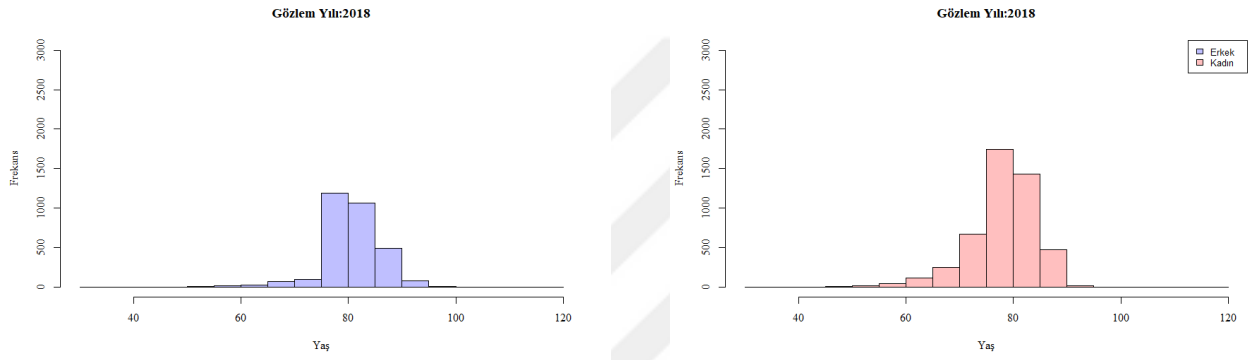
řekil 3.1. 1992 yılında gözlemlenen bireylerin yařa göre sıklık daęılımı



řekil 3.2. 2000 yılında gözlemlenen bireylerin yařa göre sıklık daęılımı



Şekil 3.3. 2010 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı

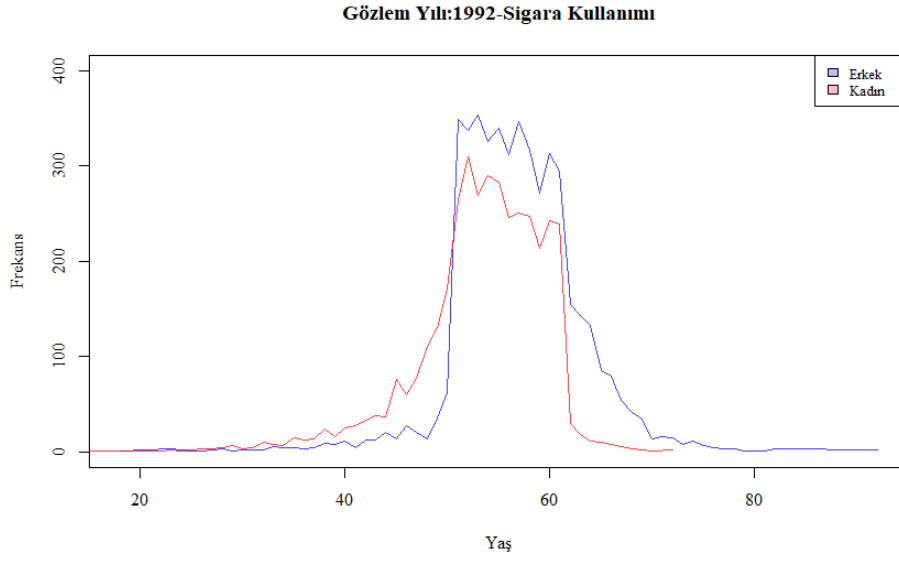


Şekil 3.4. 2018 yılında gözlemlenen bireylerin yaşa göre sıklık dağılımı

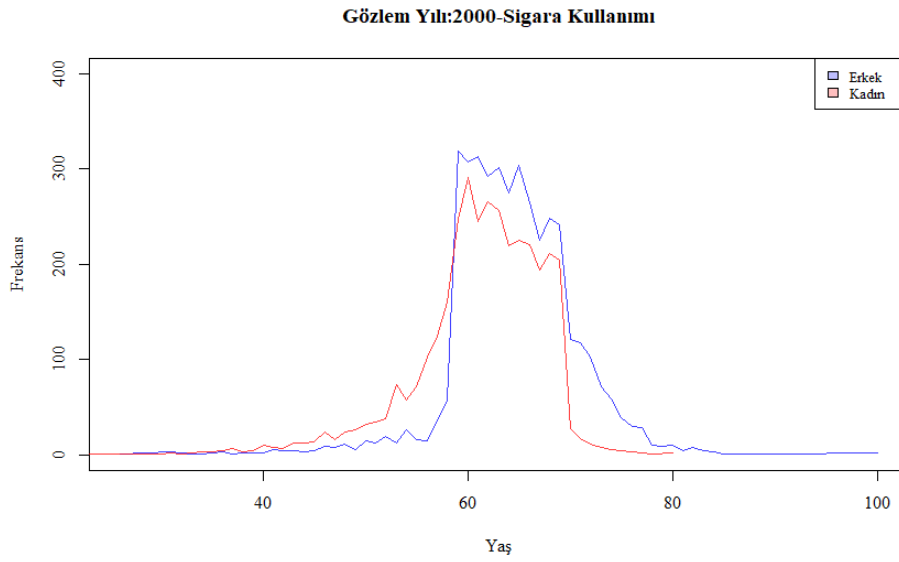
3.2.1. Sigara Kullanımı

HRS çalışmasındaki katılımcıların sigara kullanımını gösteren iki farklı değişken bulunmaktadır. RwSMOKEV kodlu değişken, katılımcının hayatında hiç sigara içip içmediğini, RwSMOKEN kodlu değişken ise şu anda sigara içip içmediği sorusunun cevabını göstermektedir. Her iki değişkende de evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Dördüncü gözlem yılından yedinci gözlem yılına kadar yapılan anketlerde eğer katılımcı daha önce hiç sigara içmedim yanıtını veriyse, tekrar hiç sigara içip içmediği sorulmamıştır. Yedinci gözlem yılından sonra ise eğer bir önceki gözlemden katılımcı hiç sigara içmediğini belirttiyse, iki soru da atlanmıştır. Buna göre katılımcıların belli bir yaştan sonra sigaraya başlamayacağı varsayılmış ve sadece sigara kullanan katılımcıların bırakıp, bırakmadıkları araştırılmıştır. Buna göre tez çalışmasında RwSMOKEN kodlu değişken kullanılmıştır. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayrımında sigara kullanım sayılarına ilişkin grafikler verilmiştir.

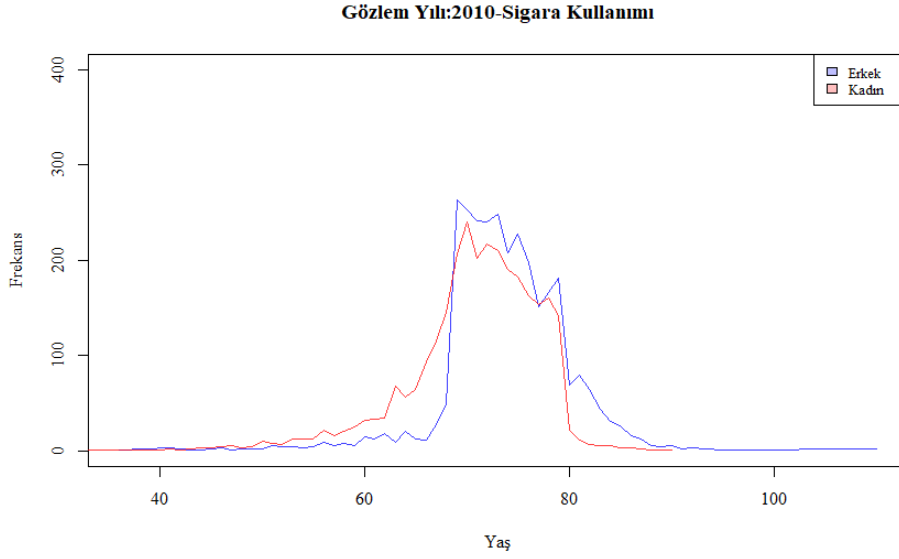
Grafiklerde de görüleceği üzere, katılımcılar yaşlandıkça sigara kullanan bireylerin sayısı azalmaktadır.



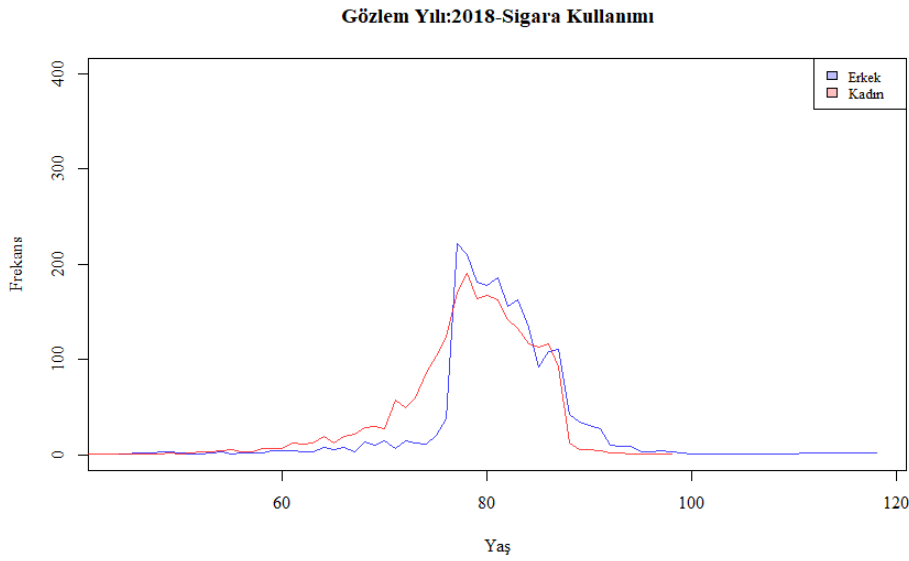
Şekil 3.5. 1992 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı



Şekil 3.6. 2000 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı



Şekil 3.7. 2000 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı

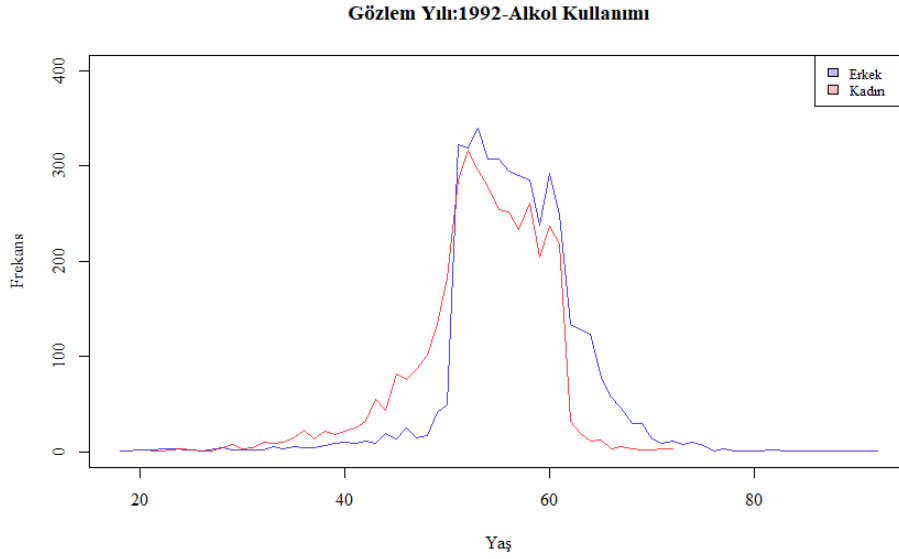


Şekil 3.8. 2018 yılında gözlemlenen bireylerin yaşa göre sigara kullanım sıklığı

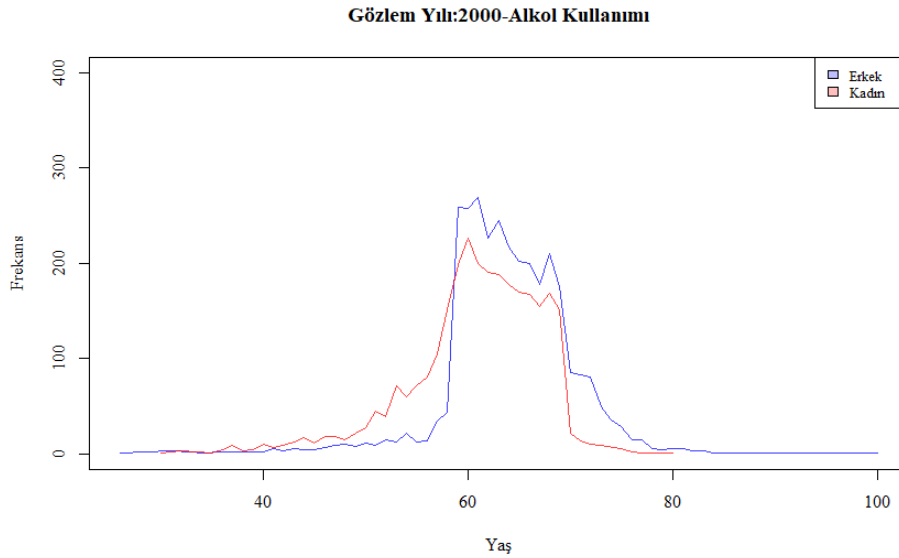
3.2.2. Alkol Kullanımı

HRS çalışmasındaki katılımcıların alkol kullanımını gösteren üç farklı değişken bulunmaktadır. RwDRINK kodlu değişken, katılımcının hayatında hiç alkol kullanıp kullanmadığını, RwDRINKD ve RwDRINKN kodlu değişkenler ise tüketilen alkol miktarını göstermektedir. İlk değişkende de evet/hayır şeklinde kategorik yanıtlar, diğer

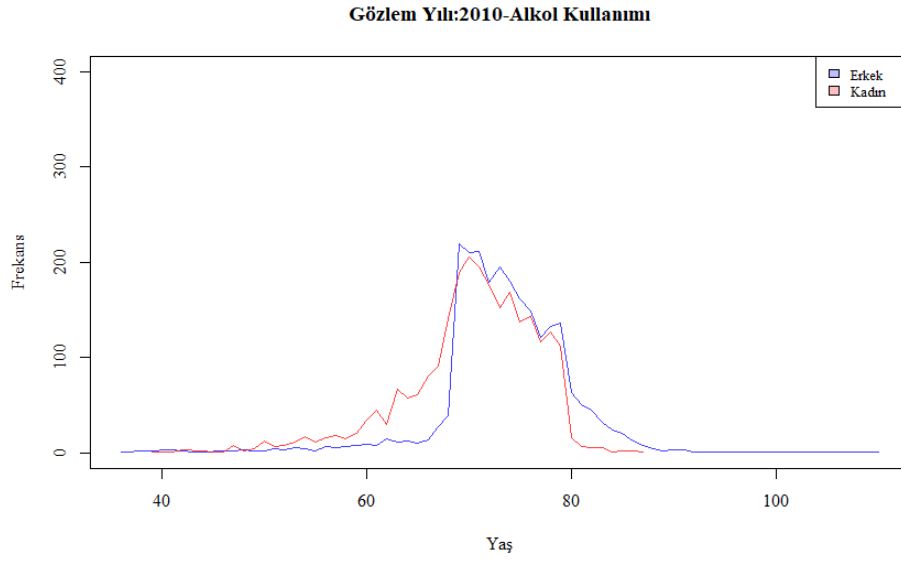
iki deęiřkende ise numerik yanıtlar kaydedilmiřtir. Ancak tüketlenen alkol miktarının tutulduęu deęiřken üçüncü gözlem yılından itibaren kaydedildięi için eksik veri oluřturmaktadır. Bu nedenle alkol kullanım durumu incelenirken RwDRINK kodlu deęiřkende tutulan yanıtlara göre bireylerin alkol kullanımı kodlanmıřtır. Ařaęıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayırımında alkol kullanım sayısına iliřkin grafikler verilmiřtir.



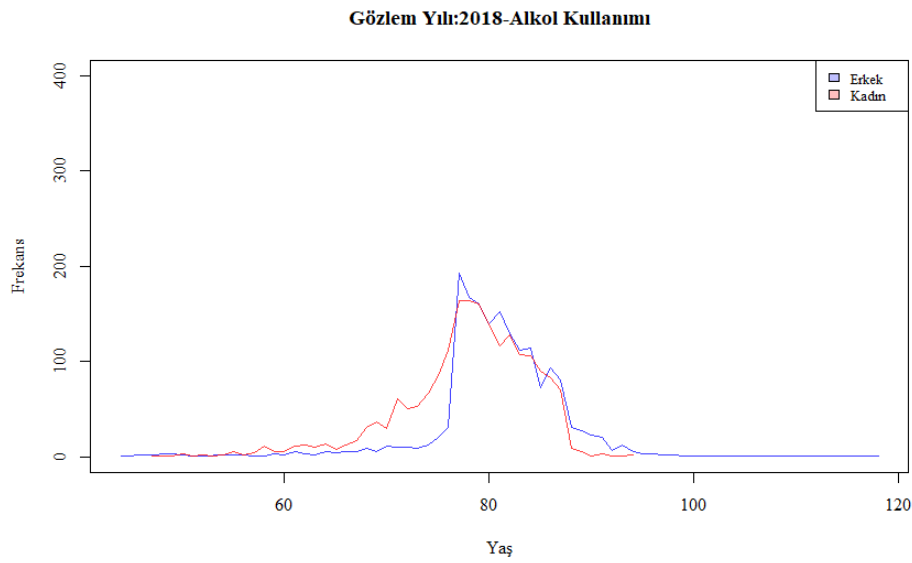
řekil 3.9 1992 yılında gözlemlenen bireylerin yařa göre alkol kullanım sıklıęı



řekil 3.10. 2000 yılında gözlemlenen bireylerin yařa göre alkol kullanım sıklıęı



Şekil 3.11. 2010 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı



Şekil 3.12. 2018 yılında gözlemlenen bireylerin yaşa göre alkol kullanım sıklığı

Her iki cinsiyet içinde, 1992 yılından 2018 yılına gelindiğinde bireylerin yaşlandığı ve yaş ilerledikçe alkol kullanımının azaldığı gözlemlenmektedir.

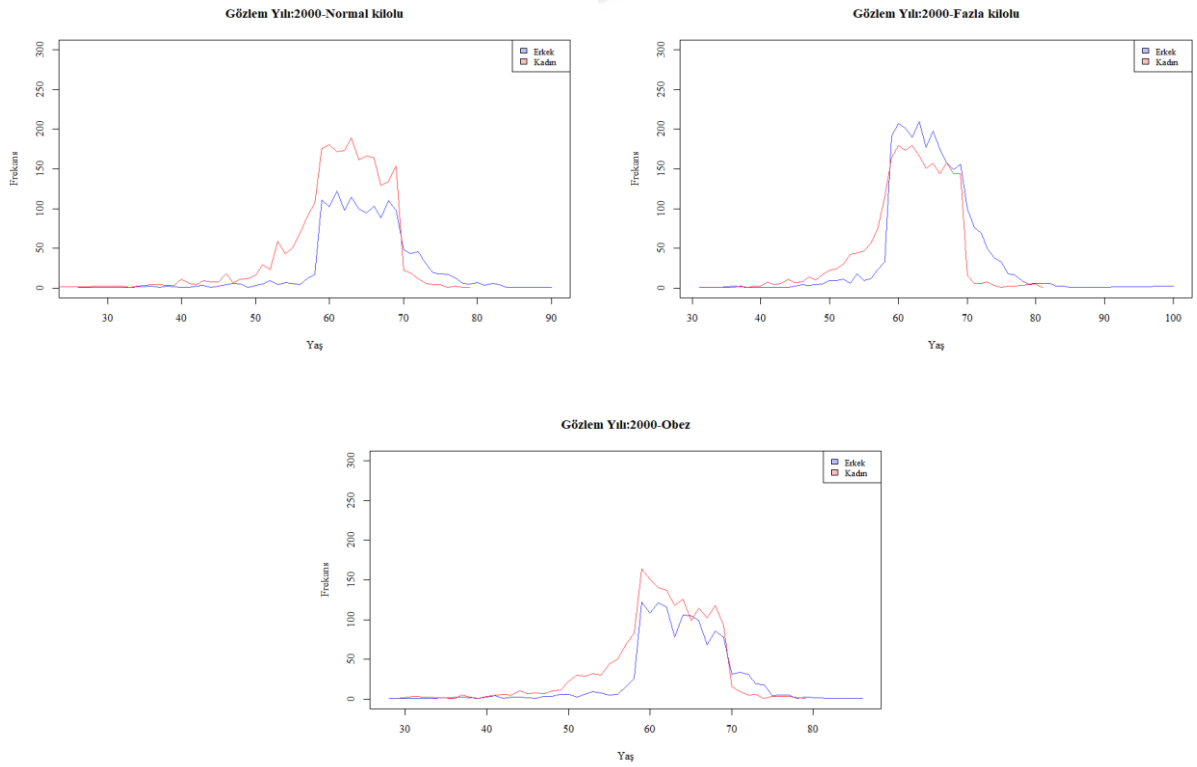
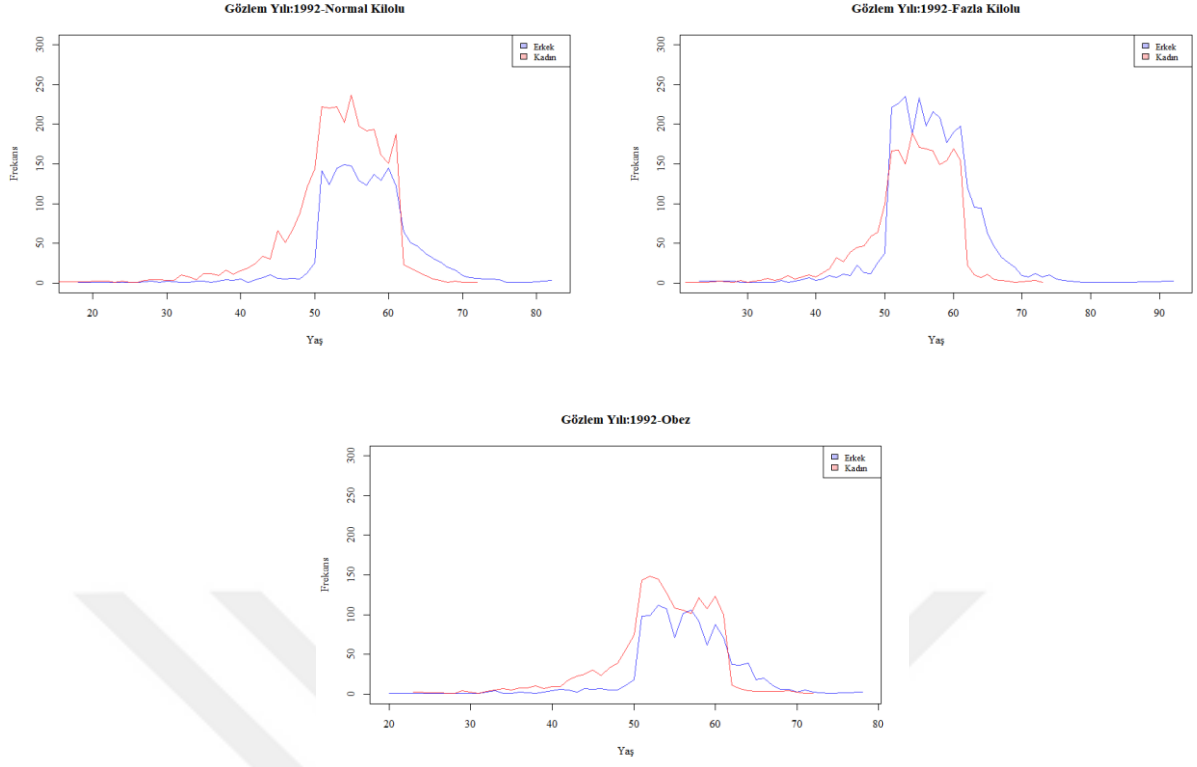
3.2.3. Beden K tle İndeksi (BKI)

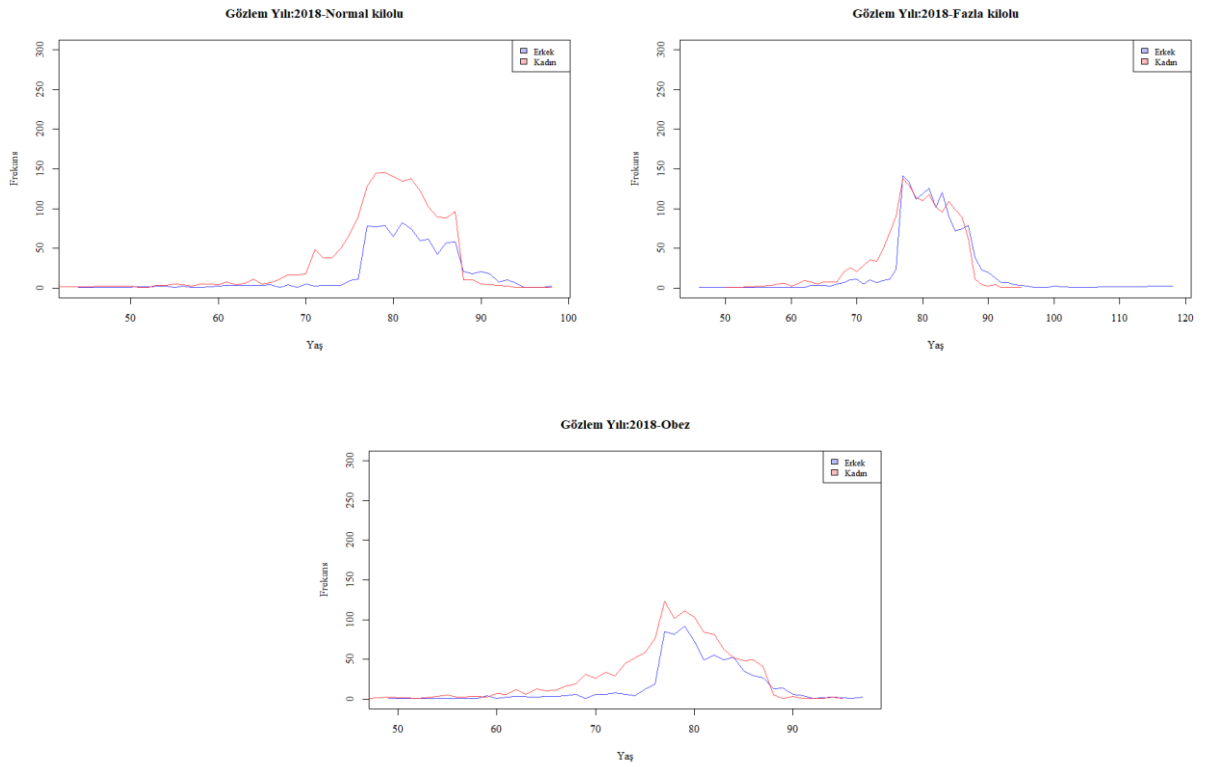
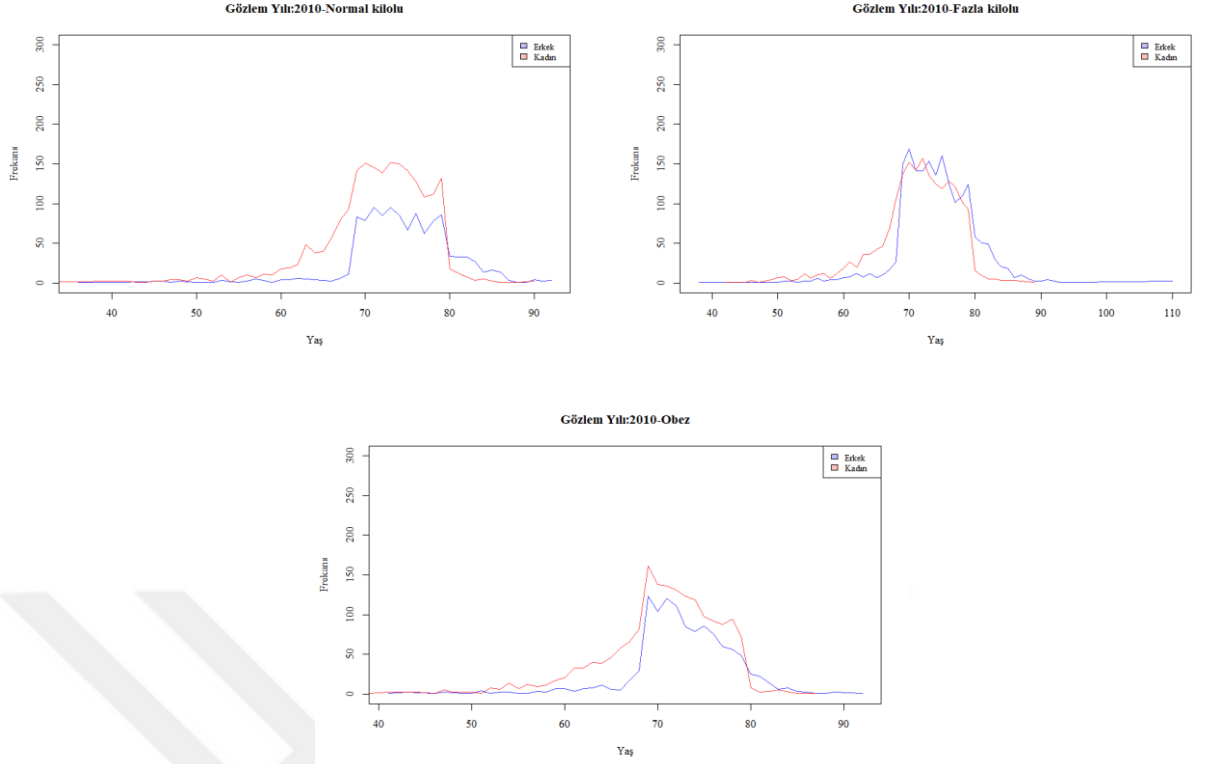
HRS alışmasındaki katılımcıların Beden K tle İndeksini (BKI) g steren RwpMBMI kodlu deėiřken bulunmaktadır. BKI deėiřkeni katılımcıların boy ve kilo  l mlerini kullanarak kilo/boy^2 řeklinde hesaplanmış ve numerik olarak kaydedilmiřtir. alışmamızda bu deėiřkenin aıklayıcı deėiřken olarak kullanılabilmesi iin gruplamaya ihtiya duyulmaktadır. Bu řekilde katılımcılar, normal kilolu ya da fazla kilolu olmaları y n nden deėerlendirilebilecektir. Bu hususta, alışmamıza benzer olarak kritik hastalıklar ile ilgili modellerin sunulduėu Macdonald, Warters ve Wekwete [28] alışmasında yer alan BKI grupları kullanılmıştır. Kaydedilen BKI deėerine g re yapılan gruplama ařaėıdaki izelge 3.5'te verilmiřtir.

izelge 3.5. BKI Kategorileri

BKI Deėer Aralığı	Kategori
$BKI \leq 25$	Normal kilolu
$25 < BKI \leq 30$	Fazla kilolu
$30 < BKI$	Obez

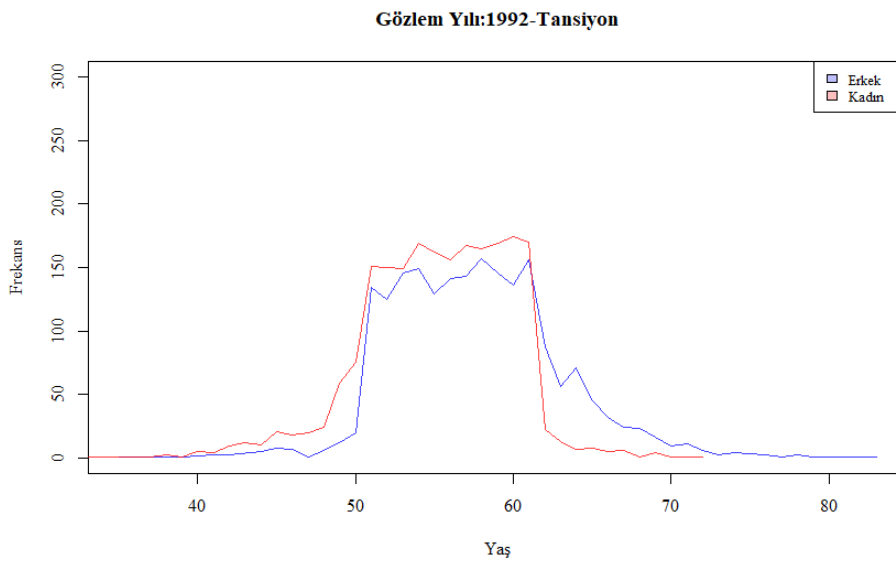
Belirlenen gruplara g re katılımcılar 3 kategoride deėerlendirilmiřtir. Ařaėıda verilen grafiklerde 1992, 2000, 2010 ve 2018 g zlem yılları iin Normal kilolu, Fazla kilolu ve Obez kategorilerinde frekanslar ve cinsiyet farklılıkları g zlemlenebilir.



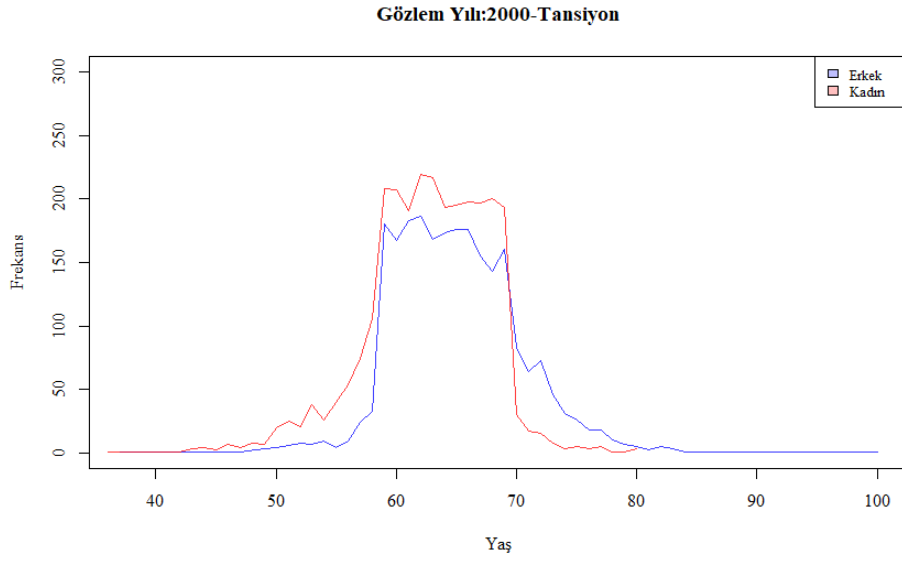


3.2.4. Tansiyon Hastalığı

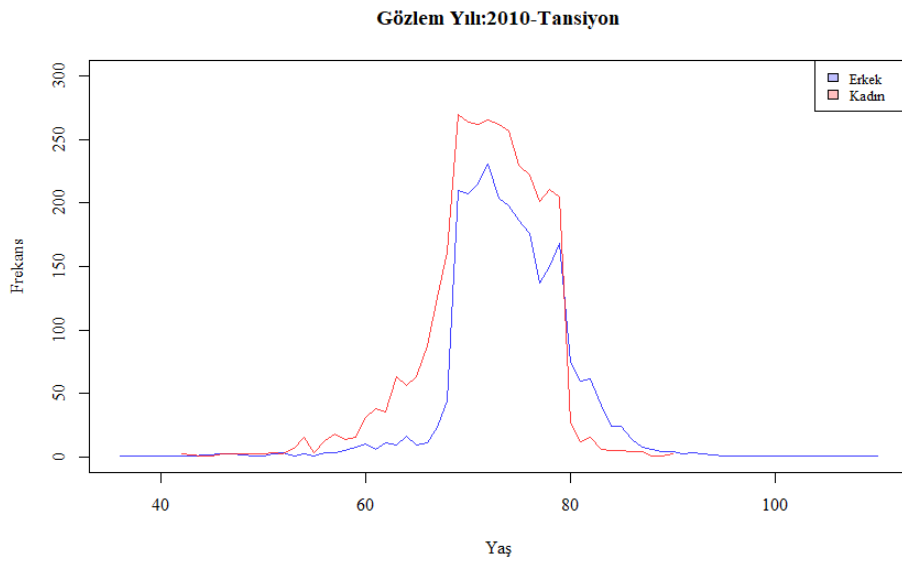
HRS çalışmasındaki katılımcıların sağlık ile ilgili bilgilerinin ölçüldüğü değişkenlerden, kritik hastalık dışında olan ancak bireyin genel sağlık durumu ile ilgili bilgi içeren hastalıklar da dikkate alınmıştır. Katılımcıların tansiyon hastası olup olmadığının ölçüldüğü RwhIBP kodlu değişken kapsamında, katılımcılara doktorun kendilerine yüksek tansiyon hastalığı teşhisi koyup koymadığı sorusu yöneltilmiştir. Bu değişkende evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayrımında tansiyon hastalığı teşhisi konan katılımcı sayılarına ilişkin grafikler verilmiştir.



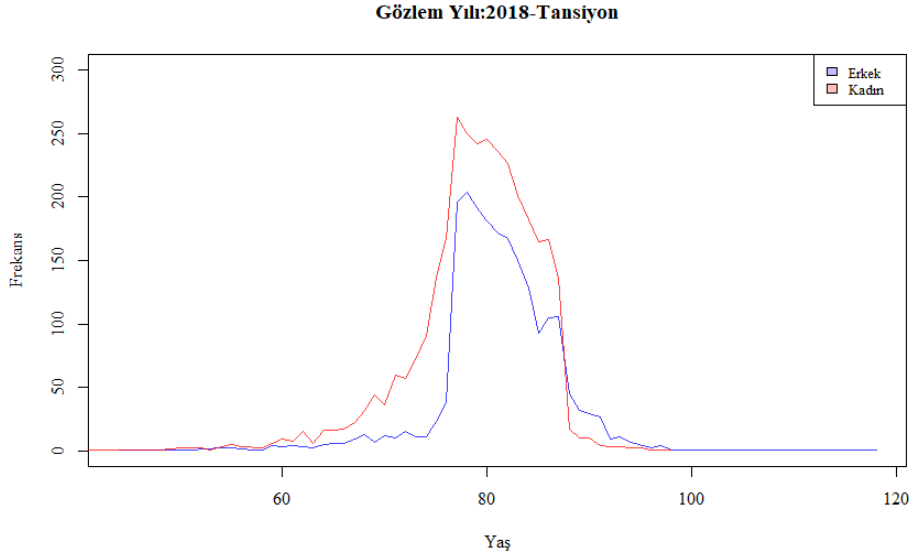
Şekil 3.17. 1992 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı



Şekil 3.18. 2000 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı



Şekil 3.19. 2010 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı

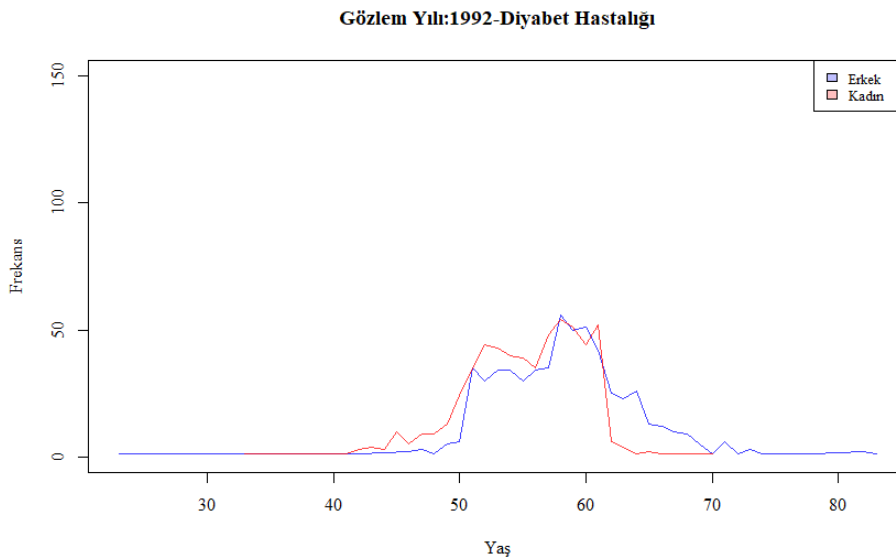


Şekil 3.20. 2018 yılında gözlemlenen bireylerin yaşa göre Tansiyon Hastalığı sıklığı

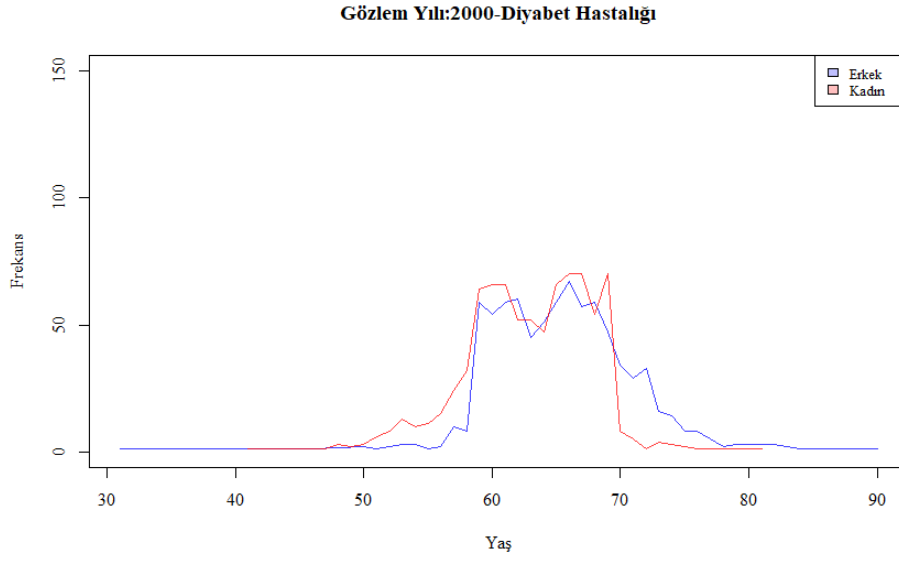
1992 yılından 2018 yılına gelindiğinde, bireyler yaşlandıkça tansiyon hastalığı görülme sıklığının arttığı, kadınlarda erkeklere göre daha fazla görüldüğü izlenmiştir.

3.2.5. Diyabet Hastalığı

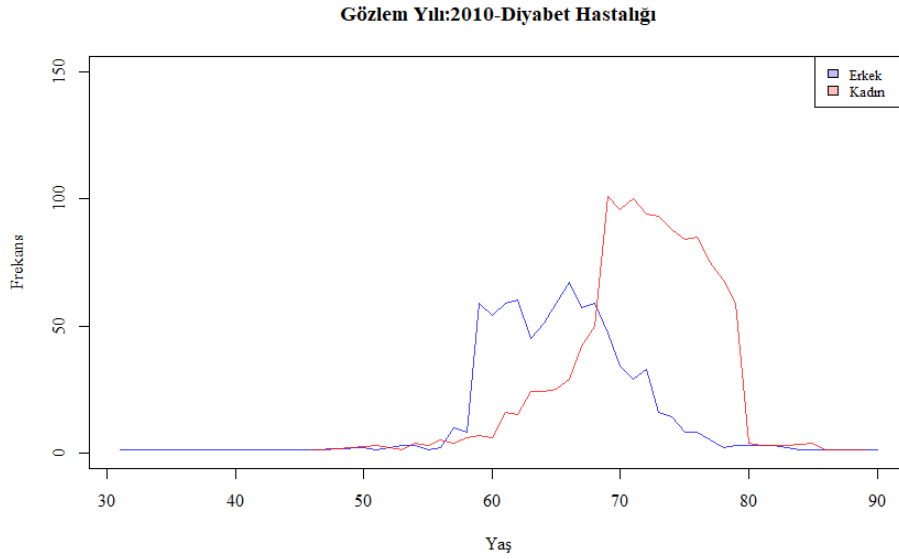
Katılımcıların diyabet hastası olup olmadığının ölçüldüğü RwDIAB kodlu değişken kapsamında, katılımcılara doktorun kendilerine diyabet hastalığı teşhisi koyup koymadığı sorusu yöneltilmiştir. Bu değişkende evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri cinsiyet ayrımında diyabet hastalığı teşhisi konan katılımcı sayılarına ilişkin grafikler verilmiştir.



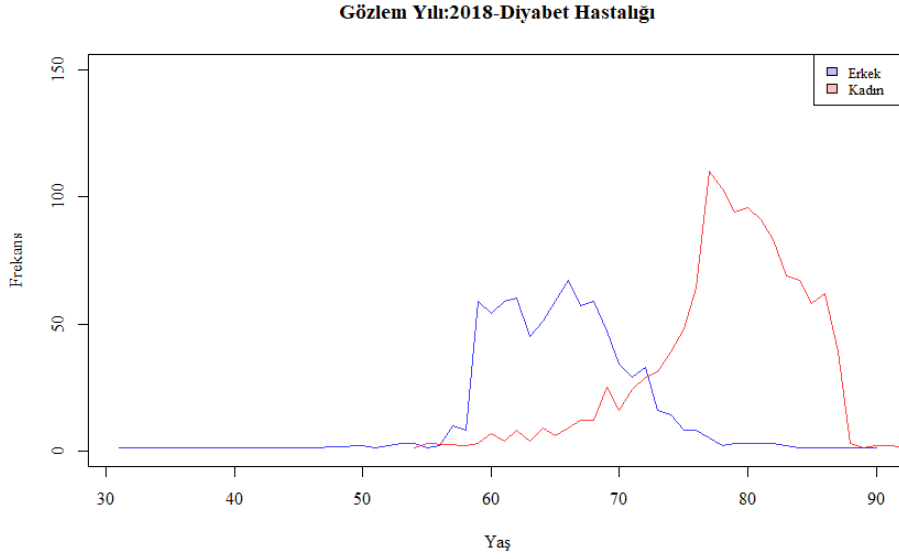
Şekil 3.21. 1992 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı



Şekil 3.22. 2000 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı



Şekil 3.23. 2010 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı

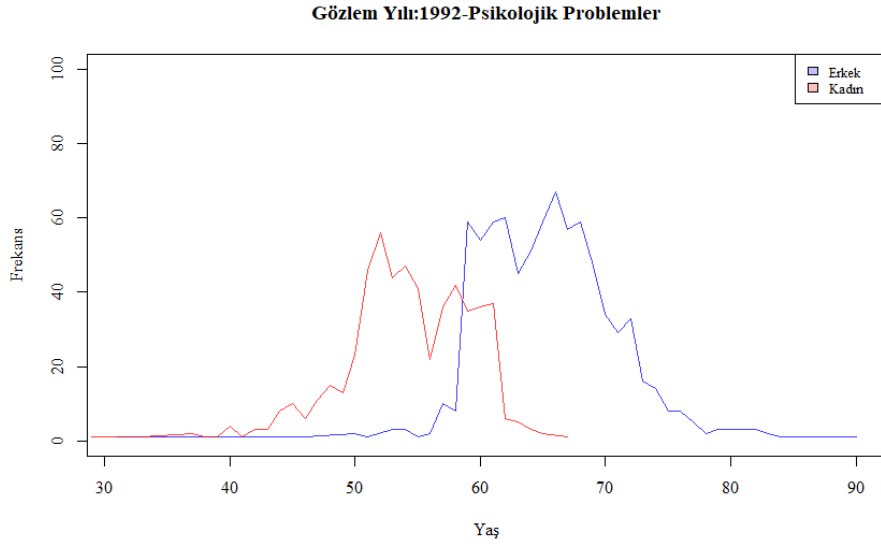


Şekil 3.24. 2018 yılında gözlemlenen bireylerin yaşa göre Diyabet Hastalığı sıklığı

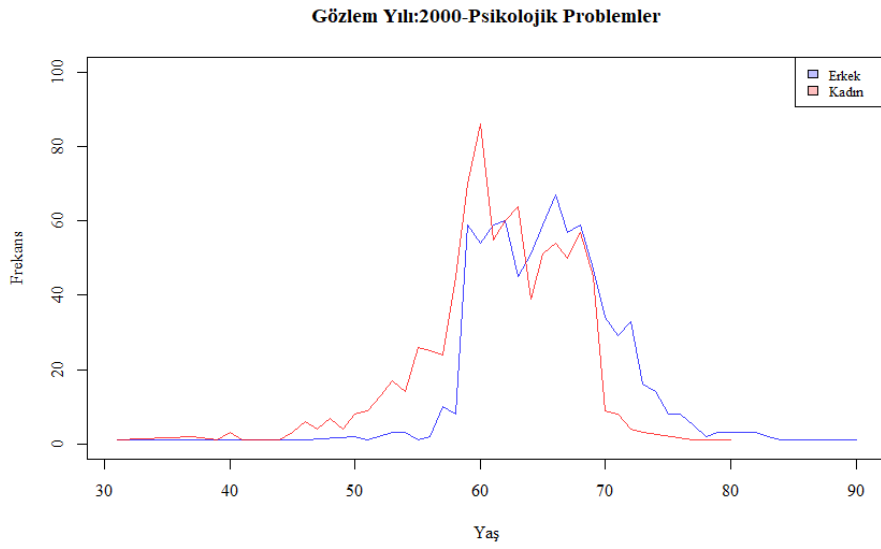
Diyabet hastalığı yapılan gözlemler incelendiğinde, bireyler yaşlandıkça görülme sıklığının arttığı ve kadın ve erkek arasında farklılık olduğu görülmektedir. Kadınlardaki görülme sıklığı erkeklere göre çok daha fazla yükselmektedir.

3.2.6. Psikolojik Problemler

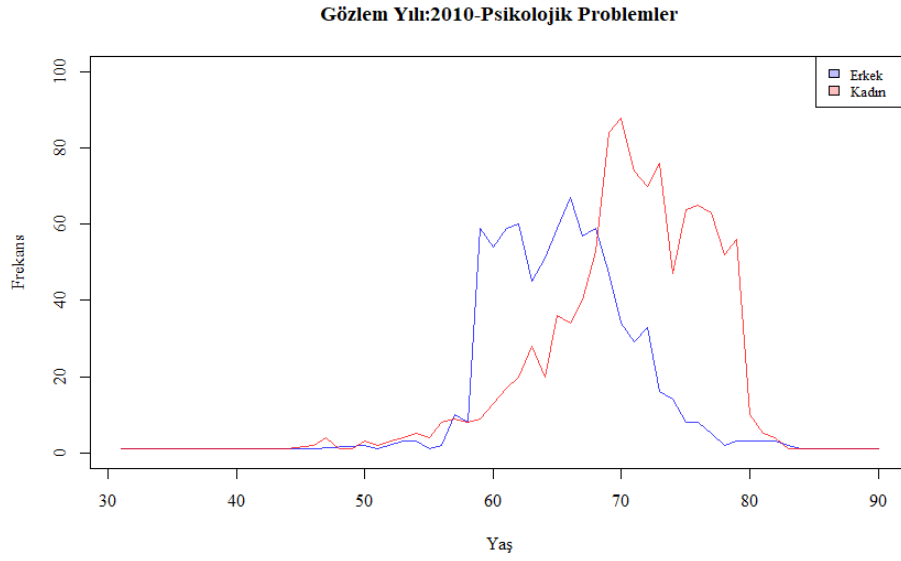
Katılımcıların psikolojik problemleri olup olmadığının ölçüldüğü RwPSYCHR koldu değişken kapsamında, katılımcılara doktorun kendilerine duygusal, sinirsel ya da psikolojik problemleri olduğunu söyleyip söylemediği sorusu yöneltilmiştir. Bu değişken de evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayrımında psikolojik problemleri olduğunu belirten katılımcı sayılarına ilişkin grafikler verilmiştir.



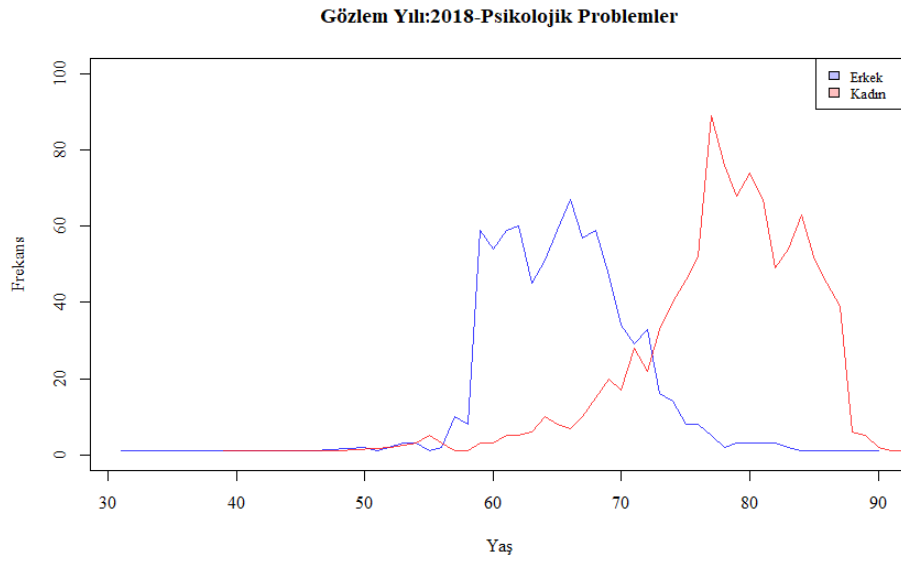
Şekil 3.25. 1992 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı



Şekil 3.26. 2000 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı



Şekil 3.27. 2010 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı

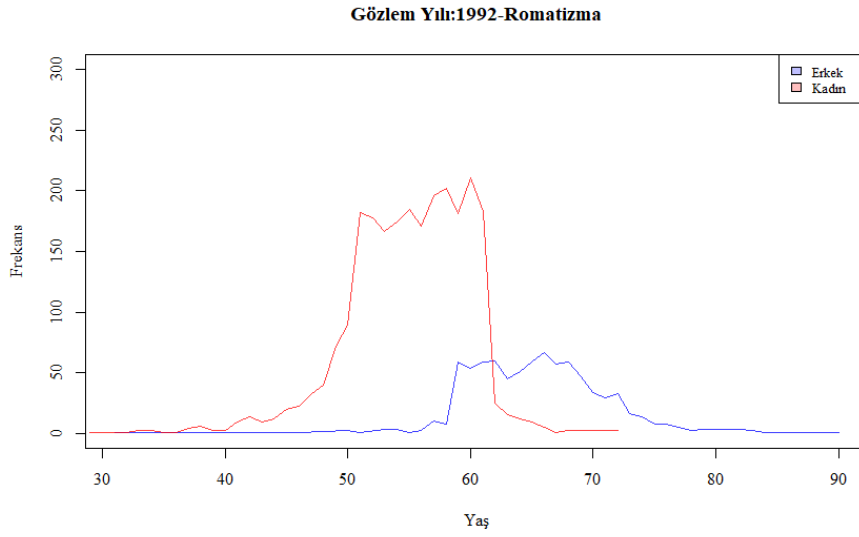


Şekil 3.28. 2018 yılında gözlemlenen Psikolojik Problemleri olan bireylerin sıklığı

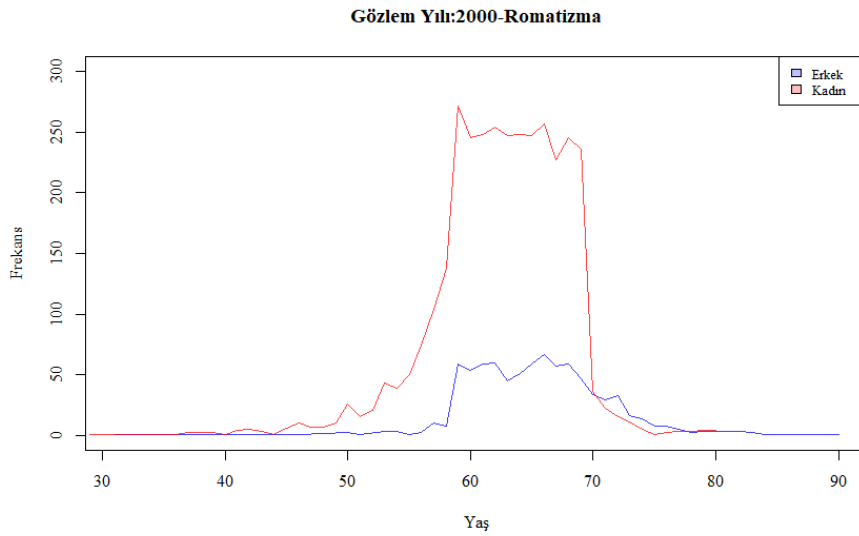
Psikolojik problemleri olan bireylerinde yaş arttıkça arttığı izlenmektedir. Ayrıca ilk gözlem yılında (1992) kadınlardaki görülme sıklığının daha az olmasına rağmen, yaş arttıkça kadınlarda görülme sıklığının daha fazla arttığı görülmektedir.

3.2.7. Romatizma

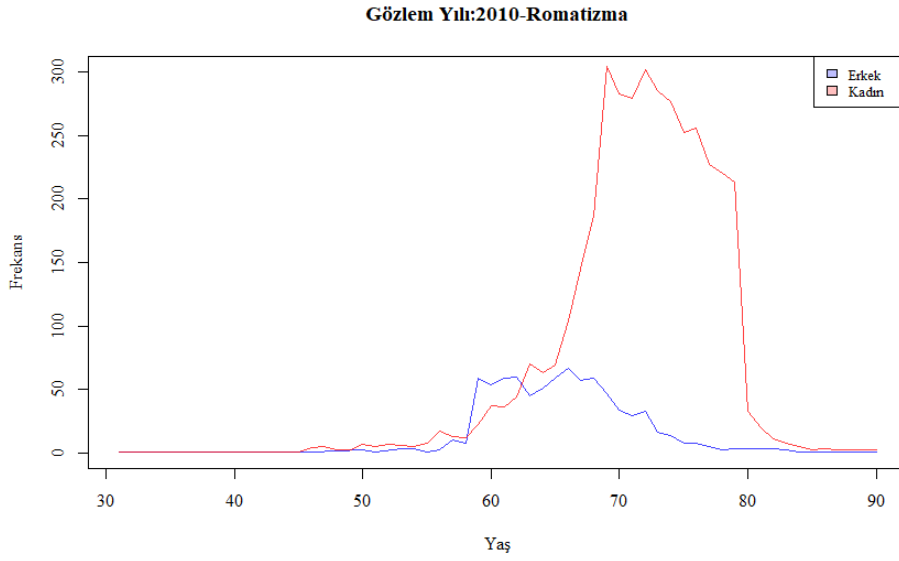
Katılımcıların artrit ya da romatizma hastası olup olmadığının ölçüldüğü RwARTHR koldu değişken kapsamında, katılımcılara doktorun kendilerine artrit ya da romatizma teşhisi koyup koymadığı sorusu yöneltilmiştir. Bu değişkende evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayrımında romatizma hastalığı teşhisi konan katılımcı sayılarına ilişkin grafikler verilmiştir.



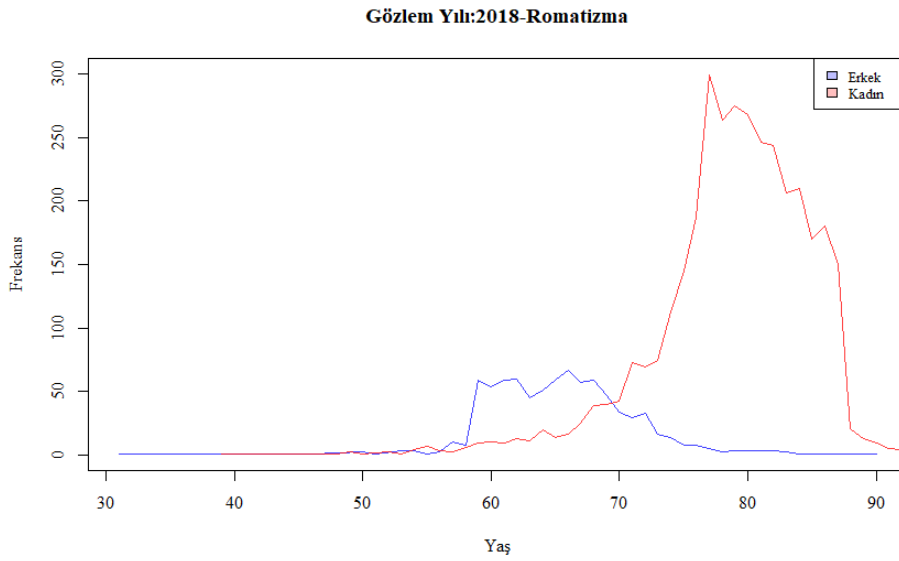
Şekil 3.29. 1992 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı



Şekil 3.30. 2000 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı



Şekil 3.31. 2010 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı

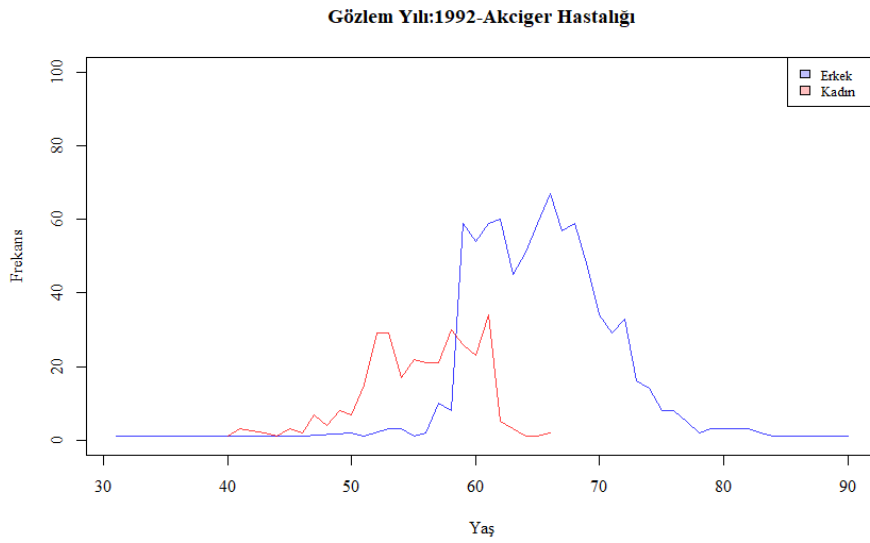


Şekil 3.32. 2018 yılında gözlemlenen bireylerin yaşa göre Romatizma sıklığı

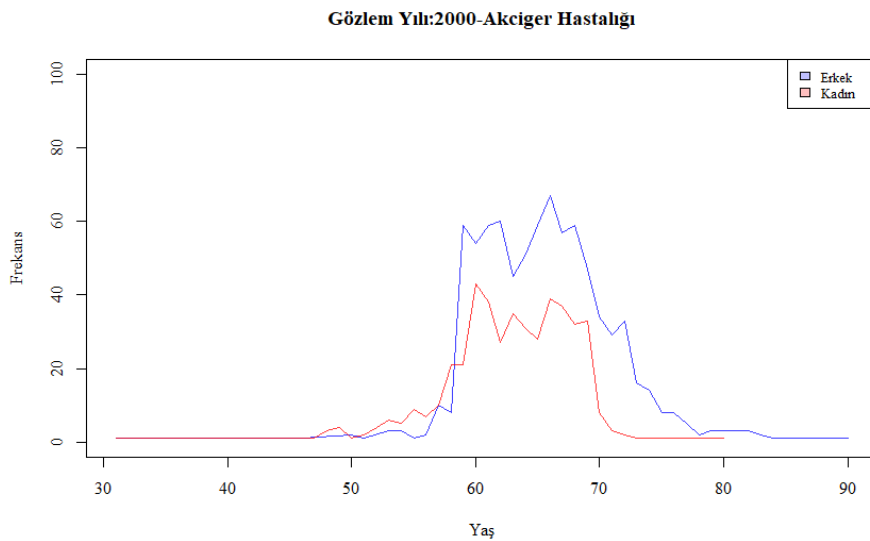
Romatizma hastalığı sıklığı gözlemlendiğinde ise cinsiyet arası çok belirgin farklılık olduğu ve kadınlarda görülme sıklığının çok daha fazla olduğu görülmektedir.

3.2.8. Akciğer Hastalığı

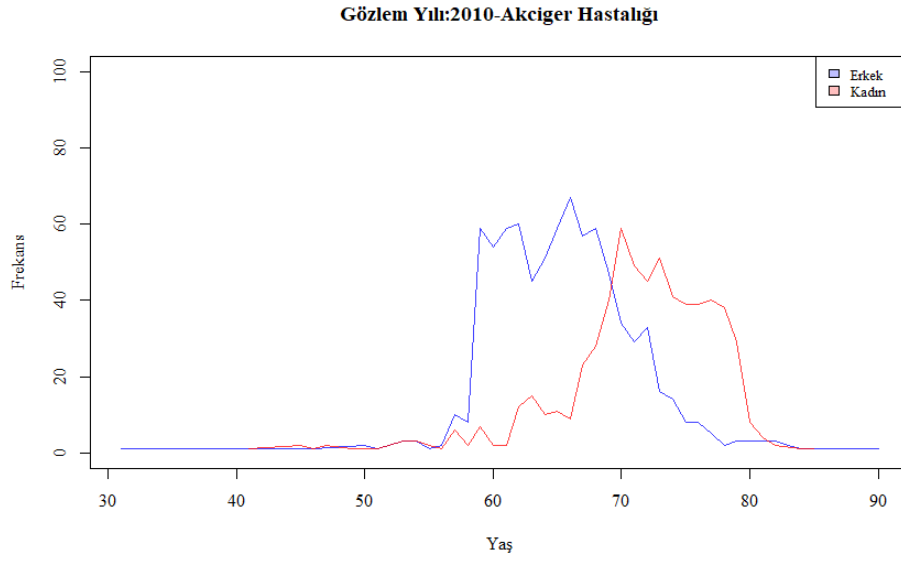
Katılımcıların akciğer ve solunum yolları ile ilgili bir hastalığının olup olmadığının ölçüldüğü RwLUNG koldu değişken kapsamında, katılımcılara doktorun kendilerine astım ve kronik bronşit dışında kronik akciğer hastalığı teşhisi koyup koymadığı sorusu yöneltilmiştir. Bu değişkende evet/hayır şeklinde kategorik yanıtlar kaydedilmiştir. Aşağıda, 1992, 2000, 2010 ve 2018 yılı gözlemleri için cinsiyet ayrımında akciğer hastalığı teşhisi konan katılımcı sayılarını gösteren grafikler verilmiştir.



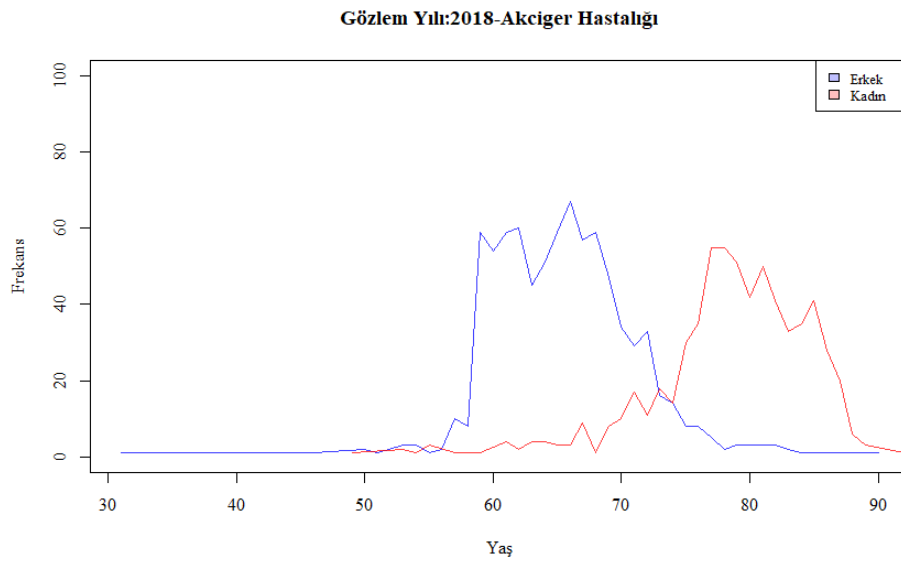
Şekil 3.33. 1992 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı



Şekil 3.34. 2000 yılında gözlemlenen bireylerin yaşa göre Akciğer Hastalığı sıklığı



Şekil 3.35. 2010 yılında gözlemlenen bireylerin yaşı göre Akciğer Hastalığı sıklığı



Şekil 3.36. 2018 yılında gözlemlenen bireylerin yaşı göre Akciğer Hastalığı sıklığı

Akciğer hastalığının görülme sıklığı incelendiğinde, tüm gözlem yıllarında erkeklerde görülme sıklığının daha fazla olduğu görülmektedir.

4. YÖNTEM

Tez çalışmasında, şimdiye kadar kritik hastalık sigortaları için geliştirilmiş modellerden farklı olarak Saklı Markov Modelleri kullanılmıştır. Bu yöntem sayesinde şimdiye kadar yapılan çalışmalarda kullanılan doğal sınıflandırmalar yerine veriden risk sınıflandırmalarının elde edilmesi amaçlanmıştır. Risk açısından doğru sınıflandırma hem sigortacılar hem de sağlık konusundaki kamu düzenleyicileri için oldukça önemlidir. Özellikle zaman içerisinde değişebilen, kesikli ya da sürekli olan risk değişkenleri ile modelleme yapıldığı düşünüldüğünde bu risk faktörlerinin etkili bir biçimde sınıflandırılması sağlık ile ilişkili sigortaların fiyatlandırılmasında büyük öneme sahip olacaktır. Bu tür bir veriyi sınıflandırmak ve belirlenen sınıfların ayrı modellenmesinin daha doğru sonuçlar vereceği düşünülmüştür. Bu şekilde bir modelleme ancak Saklı Markov Modelleri sayesinde mümkün olabilmektedir.

4.1. Saklı Markov Modellerinin (SMM) kullanıldığı öne çıkan çalışmalar

Saklı Markov modelleri ilk olarak Markov Modellerinin ses tanımlama problemlerinde kullanılmaya başlanmasıyla ortaya çıkmıştır [30]. Saklı Markov Modelleri ile ilgili ilk çalışma Baum ve Petrie [31]'nin yapmış olduğu çalışmadır. Bu çalışmada modelde kullanılan gösterimler ve matematiksel alt yapı verilmiştir. Çeşitli zaman serilerinde Saklı Markov Modellerinin uygulamalarının gösterildiği ve R yazılımı ile örneklerin verildiği Zucchini, Macdonald ve Langrock [32] kitabı daha güncel çalışmalarda öne çıkmaktadır.

Saklı Markov Modellerinin aktüerya ve sigorta alanındaki kullanımı incelendiğinde, kısıtlı sayıda çalışma ile karşılaşılmaktadır. 2018 yılında İsveç'teki Maluliyet Sigortası hasar oranlarının Saklı Markov Modelleri ile tahmin edildiği çalışmada maluliyet oranları zaman serisinin tahmininde, genel olarak kullanılan Genelleştirilmiş Doğrusal Modellerin ve zaman serilerindeki tahmin yöntemlerinin birleştirildiği iki aşamalı yöntemlere alternatif olarak Saklı Markov Modellerinin kullanılması tartışılmıştır [33].

Bu tez çalışması çalışma ile aynı yıl tamamlanan Oflaz [34]'ın tez çalışmasında ise, ortak risk faktörlerini paylaşan ve birlikte görülebilen kronik hastalıkların görülme süreçleri

kopula fonksiyonu ve Saklı Markov Modelinin bir kombinasyonu modellenmiştir. Elliott ve Siu [35] sigortacı için optimal yatırım problemini tartıştıkları çalışmalarında finansal fiyat ve sigorta risk süreci belirlenirken sürekli zamanlı, sonlu durumlu Saklı Markov modellerinden faydalanmışlardır.

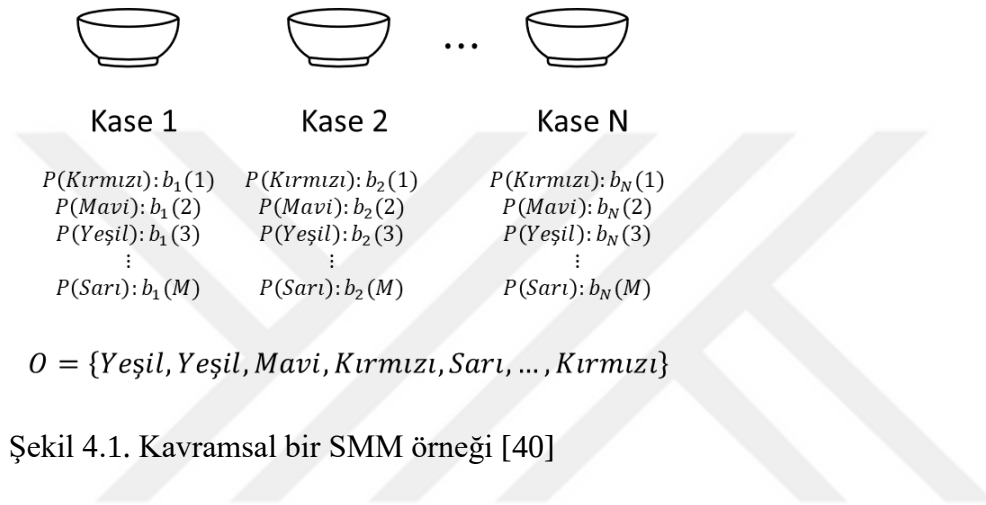
Sigorta uygulamaları dışında, ayrıca bireylerin sağlık durumlarına ilişkin verilerin kullanıldığı (bireyin kronik hastalıkları, kilosu ve yaşı vb.), veri madenciliği alanında yapılan bazı çalışmalarda da SMM kullanıldığı görülmektedir [36–38]. Bu çalışmalarda SMM veri kümeleme yöntemi olarak kullanılmıştır. Ghassempour [37] tez çalışmasında, benzetim çalışmasına ek olarak HRS çalışmasının verileri ile de çalışmış ve modelin sigorta uygulamalarında da kullanılabileceğinden bahsetmiştir. Saklı Markov modellerinin tez çalışmasında da kullanılan HRS verisinde bir uygulaması da Bartolucci, Bacci ve Pennoni [39] çalışmasında görülebilir. Bu çalışmada yazarlar HRS verisindeki katılımcıların kendi sağlık durumları ile ilgili 1-5 arası puanlamalarının kaydedildiği değişkeni kullanarak, modelin gizil (latent) yapısını, açıklayıcı değişkenlerin modelde kullanımını ve tahmin algoritmalarını tartışmışlardır.

4.2. Saklı Markov Modelleri

Saklı Markov Modelleri (SMM) bir değişkenin dağılımının gözlemlenemeyen bir Markov Sürecine bağlı olduğu model sınıfıdır. Tezin bu bölümünde SMM'nin matematiksel tanımı ve parametre kestirim algoritmaları genel hatlarıyla verilmiştir. Daha kapsamlı bir okuma için Rabiner [40], Zucchini, Macdonald ve Langrock [32], Ephraim ve Merhav [41] ve Bartolucci, Farcomeni & Pennoni [42] önerilmektedir.

SMM her ne kadar sinyal işleme ve finansal zaman serisi modellemesinde 30 yılı aşkın süredir kullanıyor olsa da ancak son 10 yıldır üzerinde uzlaşılan bir terminolojiye sahip olabilmıştır. Söz gelimi Bartolucci, Farcomeni & Pennoni [42] SMM için Gizil (Latent) Markov Modeli ismini kullanmayı tercih etmişlerdir. Ancak son dönemde onlar da çoğunluğa uyarak SMM ifadesini kullanmayı tercih etmişlerdir. SMM'in matematiksel tanımı ve arka planda bulunan algoritmaların ifadesi için karmaşık bir gösterim sistemine ihtiyaç duyulmaktadır. Bu bölüm içinde Rabiner [40]'ın çalışmasındaki gösterimler ve anlatım kullanılmıştır.

Rabiner [40], SMM'in kompleks gösteriminde okuyucuya destek olması için bir kabın içinden top çekilme senaryosundan bahsetmektedir. Kurguya göre bir odada N adet kase ve her kasenin içinde dağılımı farklı olmak üzere M adet renkli top bulunmaktadır. Odanın içinde bulunan kişi başlangıçta (tanımlı bir rassal sürece göre) bir kase seçer. Bu kabın içinden seçilen topun rengi dışarıda bekleyen gözlemciye bildirilir. Mevcut kaseye göre belirlenen rassal süreç dahilinde bir sonraki kase seçilir, içinden bir top alınır ve rengi bildirilir. Bu süreç sonunda, Şekil 4.1'de O kümesi ile gösterilen çıktılar gözlemlenir. Bu şekilde sonlu bir şekilde kurgulanan süreç SMM kullanılarak modellenir.



Şekil 4.1. Kavramsal bir SMM örneği [40]

4.3. Saklı Markov Modeli (SMM) Bileşenleri

Bir SMM'nin matematiksel tanımı şu beş bileşenden oluşmaktadır:

- Modeldeki durum sayısı N : Her ne kadar durumlar saklı olsa da modelin anlatımında durumların fiziksel veya kavramsal bir karşılığı bulunmaktadır. Şekil 4.1'de verilen senaryoda tipik olarak kaseler saklı durumlara karşılık gelmektedir. Kritik hastalık modellemesinde hastanın genel sağlık durumu saklı yani bilinmeyen durum olarak ifade edilebilir. Hastanın genel sağlığı için risksiz, riskli ve çok riskli gibi gizli durumlardan bahsedilebilir. SMM'nin en genel formunda tüm durumlar arası geçişe izin verilebilmektedir. Durumlar $S = \{S_1, S_2, \dots, S_N\}$ kümesi ile gösterilir ve t zamanındaki durum q_t şeklinde ifade edilir.
- Her durum için olası gözlem adedi M : Bazı kaynaklarda gözlemlere alfabe denmektedir. Bunun temel nedeni SMM'nin başlangıçta konuşmayı metne

dönüştürme modellerinde kullanılmasıdır. Yukarıdaki senaryoda topların renkleri gözlemlenir. Gözlem kümesi $V = \{v_1, v_2, \dots, v_M\}$ şeklinde tanımlanır.

- Durum geçiş olasılık dağılımı $A = \{a_{ij}\}$ şu şekilde tanımlanmıştır:

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, \quad j \leq N$$

Ergodik modelde tüm geçiş olasılıkları sıfırdan büyüktür ancak diğer modellerde bazı durumlar arası geçiş olasılığı sıfır olabilir. Kritik hastalık için kurulan modelde en kritik duruma denk gelen saklı durumdan daha az riskli sağlık halini ifade eden gizli durumlara geçiş olasılıklarının sıfır veya sıfıra çok yakın çıkması beklenmektedir. Kritik hastalık için kurulacak SMM'de daha az riskli durumlara geçiş olasılığının daha riskli durumlara geçiş olasılığına kıyasla düşük olması beklenmektedir.

- Her j durumunda gözlem olasılık dağılım fonksiyonu, $B = \{b_j(k)\}$.

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M$$

- Başlangıç durumu olasılık dağılımı $\pi = \{\pi_i\}$

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N$$

Bu tanım çerçevesinde SMM'nin model parametrelerinin çözülmesi için $\lambda = (A, B, \pi)$ olasılık dağılımlarının belirlenmesi gerekmektedir. Bunun başarılması için kullanılan algoritma temelde bir çözüm önerisinin sunulması, önerinin değerlendirilmesi ve daha başarılı olacak başka bir parametre kümesinin önerilmesi prensibine dayanır ve modelin uygulanabilmesi için aşağıda verilen üç problemin çözülmesi gerekir.

Problem 1: $\lambda = (A, B, \pi)$ şeklinde verilen bir SMM'nin, $O = \{O_1, O_2, \dots, O_T\}$ gözlem kümesini üretme olasılığı, $P(O|\lambda)$ nasıl hesaplanır? Bu problem aynı zamanda bir değerlendirme ölçeğinin bulunabilmesine de olanak sağlar. Aynı veriyi ürettiği iddia edilen iki modelin hangisinin daha başarılı olduğunun belirlenebilmesine imkan tanır.

Problem 2: $\lambda = (A, B, \pi)$ şeklinde bir SMM ve $O = \{O_1, O_2, \dots, O_T\}$ gözlem kümesi verildiğinde, gözlemleri en iyi açıklayan $Q = \{q_1, q_2, \dots, q_T\}$ saklı durum zinciri nasıl belirlenir? Gözlem zincirini üreten temel mekanizma olan saklı durum modelinin nasıl belirlenebileceği sorgulanmaktadır.

Problem 3: $\lambda = (A, B, \pi)$ model parametreleri, $P(O|\lambda)$ olasılığını maksimize edecek şekilde nasıl belirlenir? Bu problem model parametrelerinin optimizasyonu ile ilgilidir.

4.4. SMM Temel Problemleri için Çözüm Algoritmaları

4.4.1. Problem 1

Bu problemde amaç, verilen bir λ modelinin $O = \{O_1, O_2, \dots, O_T\}$ gözlem kümesini üretme olasılığını $P(O|\lambda)$ hesaplamaktır. Bunun için modelde olabilecek tüm T uzunluğunda saklı durum zincirlerinin belirlenmesi gerekmektedir. İlk adımı başlangıç durumu olarak tanımlanan herhangi bir

$$Q = \{q_1, q_2, \dots, q_T\}$$

saklı durum zincirinin gerçekleşme olasılığı,

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T}$$

şeklinde hesaplanabilir. Bu saklı durum zinciri verildiğinde, $O = \{O_1, O_2, \dots, O_T\}$ kümesinin gözlenme olasılığı,

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda)$$

şeklinde hesaplanır. Bu aşamada gözlemlerin birbirinden bağımsız olduğu

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T)$$

varsayılmaktadır. O ve Q için bileşik olasılık hesabını yapabilmek için yukarıda hesaplanan iki terimin çarpılması gerekmektedir.

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q, \lambda)$$

Dolayısıyla O 'nun marjinal dağılımının hesaplanması (bulunması) için olası tüm Q zincirleri üzerinden hesaplanan olasılıkları toplanır.

$$\begin{aligned}
P(O|\lambda) &= \sum_{\substack{\text{tüm } Q \\ \text{zincirleri}}} P(O|Q, \lambda) P(Q, \lambda) \\
&= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T)
\end{aligned}$$

Bu şekilde verilen matematiksel karakterizasyonda başlangıçta ($t = 1$) SMM π_{q_1} olasılığı ile q_1 durumunda başlar ve $b_{q_1}(O_1)$ olasılığı ile O_1 gözlemini üretir. Bir sonraki adımda ($t = 2$) SMM q_1 durumundan q_2 durumuna $a_{(q_1 q_2)}$ olasılığı ile geçer ve $b_{q_2}(O_2)$ olasılığı ile O_2 gözlemini üretir. Zincir bu şekilde son adıma ($t = T$) kadar devam eder ve bu adımda q_{T-1} durumundan q_T durumuna $a_{(q_{T-1} q_T)}$ olasılığı ile geçer ve $b_{q_T}(O_T)$ olasılığı ile O_T gözlemini üretir.

$P(O|\lambda)$ olasılığının hesaplanabilmesi için her $t = 1, 2, \dots, T$ anı için olası N adet durum olacağından toplamda $2T \cdot N^T$ kadar işlem yapmak gerekmektedir. Dolayısıyla basit modeller için bile yapılması gereken işlem sayısı devasa boyutlara ulaşacaktır. Örneğin $T = 100$ gözlem ve $N = 5$ durumlu bir model için $2 \cdot 100 \cdot 5^{100}$ yaklaşık 10^{47} adet işlem yapmak gereklidir ve bu da mümkün değildir. Baum ve Eagon [43] ile Baum ve Sell [44] tarafından geliştirilmiş olan *İleri-Geri* algoritması kullanılarak bu zorluk aşılabilmektedir.

4.4.1.1. İleri-Geri Algoritması

$t < T$ olmak üzere gözlem zincirinin bir kısmı olan $\{O_1, O_2, \dots, O_t\}$ gözlem kümesinin olasılığını ifade etmek üzere $\alpha_t(i)$ ileri değişkeni,

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i | \lambda)$$

olarak tanımlanır. Baum ve Eagon [43] ile Baum ve Sell [44] bu şekilde tanımlanan $\alpha_t(i)$ değişkeninin indüksiyon ile çözümlenebileceğini göstermişlerdir. Indüksiyon şu adımlardan oluşur:

1. Başlangıç:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

2. İndüksiyon

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad 1 \leq j \leq N$$

3. Termin

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

İleri-Geri algoritmasının $P(O|\lambda)$ hesaplamasını nasıl gerçekleştirdiğine yönelik bir inceleme tezin amacı dışında kaldığı için yapılmamıştır. Ancak, İleri-Geri algoritmasında kullanılan işlem adedine bakıldığında toplamda N^2T adet işlem sonucunda olasılık hesabının yapılabildiği görülmektedir. Bu sayı $2TN^T$ ile karşılaştırıldığında oldukça düşük kalmaktadır. Yukarıdaki örnekteki sayılar $N = 5$ ve $T = 100$ kullanıldığında İleri-Geri algoritması ile 3000 civarında işlem yapıldığını görülmektedir. Bu sayı ilk hesaplanan 10^{72} 'nin yanında oldukça düşük kalmaktadır.

Şu ana kadar $\alpha_t(i)$ ile İleri-Geri algoritmasının sadece İleri kısmı tanımlanmıştır. Geri değişkeni olan $\beta_t(i)$,

$$\beta_t(i) = P(O_{t+1}O_{t+2} \cdots O_T | q_T = S_i, \lambda)$$

olarak tanımlanır. Geri algoritması, Problem 1'in çözümlenmesinde kullanılmamakta ancak Problem 2 ve Problem 3'ün çözümlenmesinde kilit rol oynamaktadır. Geri değişkeninin kullanıldığı algoritma aşağıda verilen iki adımdan oluşur:

1. Başlangıç

$$\beta_T(i) = 1, \quad 1 \leq i \leq N$$

2. İndüksiyon

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N$$

4.4.2. Problem 2

Cevabın matematiksel olarak tekil bir şekilde tanımlanabildiği Problem 1'den farklı olarak Problem 2'de birden fazla çözüm; birden fazla sayıda "optimal" saklı durum zinciri bulmak mümkündür. Problemin çözümü için gözlem kümesi ve O ve model λ verildiğinde herhangi bir durumda olma olasılığını gösteren $\gamma_t(i)$ değişkeni tanımlanır.

$$\gamma_t(i) = P(q_t = S_i | O, \lambda)$$

Gamma değişkeninin alfa ve beta değişkenleri kullanılarak da ifade edilebileceği bilinmektedir.

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)}$$

Dolayısıyla,

$$\sum_{i=1}^N \gamma_t(i) = 1$$

olur. Bu şekilde tanımlanan $\gamma_t(i)$ değişkeni kullanılarak verilen gözlem kümesi için t anında en olası saklı durum,

$$q_t = \underset{1 \leq i \leq N}{\operatorname{argmax}} [\gamma_t(i)], \quad 1 \leq t \leq T$$

olarak ifade edilir. Bu tanımlamayla “optimal” saklı durum kriteri anlık tanımlanmaktadır. Bu tanımlama saklı durum geçiş olasılıkları nedeniyle geçilmesi imkânsız durumların olduğu problemlerde kullanılamaz. Bir örnekle konuyu açmak gerekirse; gözlem kümesi kritik hastalık verisi için herhangi bir an kavramsal anlamı “çok sağlıklı” olan saklı duruma denk geliyor ancak bir önceki saklı durum “çok riskli” saklı durumu olarak hesaplandıysa, çok riskli sağlık durumundan “çok sağlıklı” durumuna bir adımda geçiş olasılığı muhtemelen sıfır veya sıfıra yakın olacağı için çözüm gerçekçi değildir.

4.4.2.1. Viterbi Algoritması

Problem 2’nin çözümü için her tekil durum q_t için olasılık maksimizasyonuna karşılık en olası durum zincirini (zincirin tamamı) diğer bir deyişle $P(Q|O, \lambda)$ olasılığını maksimize etmek tercih edilmektedir. $P(Q|O, \lambda)$ olasılığını maksimize etmek ile $P(Q, O|\lambda)$ olasılığını maksimize etmek birbirine eşit amaçlardır ve Viterbi algoritması [45,46] bu amaçla geliştirilmiştir.

Verilen bir $O = \{O_1, O_2, \dots, O_T\}$ gözlem kümesi için en olası $Q = \{q_1, q_2, \dots, q_T\}$ saklı durum zincirini bulmak amacıyla,

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda]$$

değişkeni tanımlanır. Bu ifadeye göre $\delta_t(i)$ değişkeni gözlemlerin t anına kadar en olası saklı durum zincirini ifade eder. Dolayısıyla,

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1})$$

ifadesinin maksimizasyonu ile zincirin tamamı elde edilebilir. Bu mantık ile çalışan Viterbi algoritmasının adımları:

1. Başlangıç

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

$$\psi_1(i) = 0$$

2. Geri Besleme

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \cdot b_j(O_t), \quad 2 \leq t \leq T \quad 1 \leq j \leq N$$

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \quad 1 \leq j \leq N$$

3. Termin

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)]$$

4. Zincirde Geri-İzsürme

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1$$

biçimindedir.

4.4.3. Problem 3

Son olarak en zor problem SMM parametrelerinin $\lambda = (A, B, \pi)$ tahmin edilmesi problemidir. Problemin zor olmasındaki temel neden verilen gözlem kümesine göre model parametrelerini analitik olarak çözmenin mümkün olmamasıdır. Amaç genel optimizasyon probleminde lokal optimizasyon problemine indirgenir. Bu şekilde Baum-Welch [47] metodu veya gradyan algoritması [48] gibi iterasyona dayalı teknikler yoluyla lokal çözümler aranır.

4.4.3.1. Baum-Welch Algoritması

Baum-Welch metodunda ilk aşamada t anında S_i ve $t+1$ anında S_j durumlarında bulunma olasılığı tanımlanır.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda)$$

Yukarıda tanımlanan ileri $\alpha_t(i)$ ve geri $\beta_{t+1}(j)$ değişkenleri kullanılarak aynı ifade

$$\begin{aligned}\xi_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_i \sum_j \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}\end{aligned}$$

biçimine dönüştürülebilir. Yukarıda ayrıca tanımlanmış olan gözlem kümesi ve O ve model λ verildiğinde herhangi bir durumda olma olasılığını gösteren $\gamma_t(i)$ değişkeni ile $\xi_t(i, j)$ değişkenleri de birbirleriyle şu şekilde ilişkilendirilir:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j)$$

Eğer $\delta_t(i)$ değişkeni t zaman endeksi üzerinden toplanırsa sonuç S_i saklı durumunun beklenen ziyaret edilme sayısını ya da diğer bir deyişle S_i durumundan yapılan geçişlerin beklenen sayısını verir.

$$\begin{aligned}\sum_{t=1}^{T-1} \gamma_t(i) &= S_i \text{ durumundan beklenen geçiş sayısı} \\ \sum_{t=1}^{T-1} \xi_t(i, j) &= S_i \text{ durumundan } S_j \text{ durumuna beklenen geçiş sayısı}\end{aligned}$$

Bu yaklaşımla tahmin edilmesi amaçlanan π, A ve B parametreleri için formüller ortaya çıkmış olur.

$\bar{\pi}_i$ = Başlangıç ($t = 1$) anında S_i durumunda bulunmanın beklenen değeri = $\gamma_1(i)$

$$\begin{aligned}\bar{a}_{ij} &= \frac{S_i \text{ durumundan } S_j \text{ durumuna beklenen geçiş adedi}}{S_i \text{ durumundan beklenen geçiş adedi}} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}\end{aligned}$$

$$\begin{aligned}\bar{b}_j(k) &= \frac{S_j \text{ durumunda olup } v_k \text{ gözleminin beklenen adedi (frekans)}}{S_j \text{ durumunda olmanın beklenen değeri}} \\ &= \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}\end{aligned}$$

5. UYGULAMA

Bu kesimde, HRS verisi üzerinde uygulamalar yapılmıştır. Öncelikle çalışmamızda kullandığımız HRS verisinden kritik hastalık olarak belirlediğimiz değişkenler için yapılan ön analizler sunulmuştur. Sonrasında, belirlenen kritik hastalıkların gerçekleşme olasılıklarının tahmin edilmesinde kullanılan Markov modeli ve Saklı Markov Modeli ile uygulama yapılmıştır. Aynı veri kullanılarak oluşturulan Markov ve Saklı Markov modellerinin karşılaştırılması için yapılan analizler sunulmuş ve Kritik Hastalık Sigortalarının fiyatlandırılmasında kullanımı tartışılmış ve prim hesaplamaları gösterilmiştir.

5.1. Kritik Hastalıkların Demografik Özellikleri

Tez çalışmasında kullanılan HRS verisi ve içerdiği değişkenlerden kullanılabilecek değişkenler hakkındaki ayrıntılı analizler Bölüm 3’de verilmiştir. Bu kesimde ise kritik hastalık sigortaları kapsamında teminat altına alındığı bilinen temel üç hastalık ve ölüm ile ilgili değişkenlere ilişkin istatistikler verilmiştir.

Çizelge 5.1. Yaş aralıklarına göre sağlıklı durumdan kritik hastalıklara geçiş sıklığı

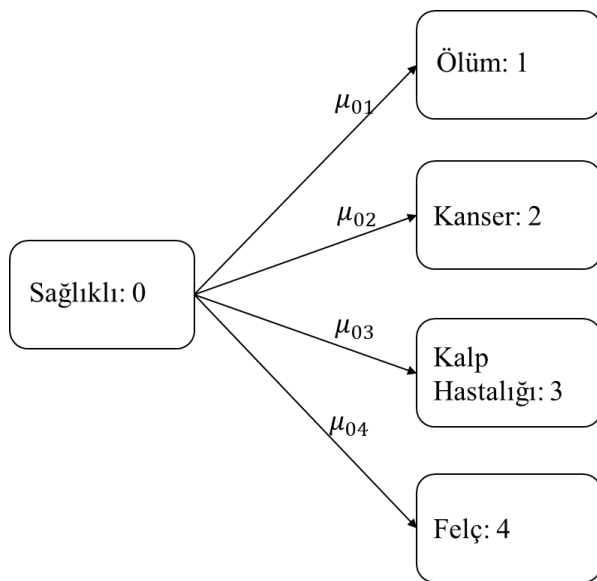
Yaş Aralığı	Toplam Yaşanan Yıl (Exposure years)	Geçiş Frekansı			
		Kanser	Felç	Kalp	Ölüm
Kadın					
40-50	3371	11	9	22	9
50-60	17794	146	63	200	102
60-70	25165	391	182	554	276
70-80	16017	278	194	518	324
80-90	3608	67	55	130	147
>90	26	0	3	1	0
Toplam	65981	893	506	1425	858
Erkek					
40-50	900	2	6	8	1
50-60	11111	76	41	192	86
60-70	19362	408	205	626	354
70-80	12396	420	178	553	427
80-90	3111	85	54	141	165
>90	197	2	0	0	16
Toplam	47077	993	484	1520	1049

Çizelge 5.1’de hem kadın hem de erkek için belirli yaş aralıklarında sağlıklı durumdan kanser, felç, kalp olarak adlandırılan kritik hastalık durumlarından birine veya ölüm durumuna geçenlerin sıklıkları verilmiştir.

Tez çalışmasında kullanılan verideki popülasyon genel olarak ileri yaştaki bireylerden oluşmaktadır. Bu nedenle analize alınan bireylerin yaşları 40 yaş üzeri olacak şekilde sınırlandırılmıştır. Çizelge 5.1 incelendiğinde hem kadınlarda hem de erkeklerde verideki popülasyonun çoğunluğunun 60-70 yaş aralığında olduğu ve özellikle 60 yaşından sonra kritik hastalıkların görülme sıklığının arttığı gözlenmektedir. Ayrıca 80 yaşından sonra hayatta kalan katılımcı sayısı da önemli ölçüde azaldığı için çalışmaya alınan bireylerin yaş sınırı 40-80 olacak şekilde düzenlenmiştir.

5.2. Genelleştirilmiş Doğrusal Model (Poisson Regresyon)

Bu kesimde, tez çalışmasında kullanılan ve Üçüncü Bölümde detayı verilen HRS çalışmasında elde edilen veri kullanılarak, kritik hastalık sigortaları bakımından hasar ile sonuçlanacak Kanser, Kalp Hastalığı ve Felç hastalıklarının teşhis oranları için tahminler yapılmıştır. Tahminler yapılırken öncelikle yazında sıklıkla kullanılan Markov modeli ile uygulama yapılmıştır [6,49]. Bu model sayesinde bir sonraki bölümde tez çalışması için önerilen model ile karşılaştırma yapılmıştır. Çalışmanın ilk aşaması için temel alınan Markov çoklu azalım modelinin görseli aşağıdaki Şekil 5.1’de verilmiştir.



Şekil 5.1. Kritik hastalıklar için önerilen Markov modeli

Şekil 5.1’de görüldüğü üzere bireyler sisteme sağlıklı durumundan girmekte ve çalışmada yer alan kritik hastalıklardan herhangi birine (kanser, kalp hastalığı, felç) yakalandıklarında sistemden çıkacakları ya da sağlıklı iken, ölüm gerçekleştiğinde sistemden çıkacakları varsayılmıştır. Kritik hastalık sigortası açısında da sigortalı sağlıklı iken sigortalanır, hayatta kaldığı her yıl boyunca prim ödemesi yapar, poliçede teminat altına alınan kritik hastalıklardan biri görüldüğünde veya öldüğünde tazminat ödemesi yapılır. Modelde $\mu_{01}, \mu_{02}, \mu_{03}, \mu_{04}$ sırasıyla sağlıklı durumdan ölüm durumuna, sağlıklı durumdan kanser durumuna, sağlıklı durumdan kalp hastalığı durumuna ve sağlıklı durumdan felç durumuna anlık geçiş hızlarını (transition intensity) göstermektedir. Bu anlık geçiş hızları MacDonald ve ark. [28,50]’nin ve Ozkok ve ark. [49,51] uygulamalarına benzer şekilde GLM Poisson Regresyonu kullanılarak tahmin edilmiştir. Markov modelinde x yaşında toplam hasar sayısı $N_x \sim \text{Poisson}(E_x^c \mu_{0i})$ ile Poisson dağılır. E_x^c , x yaşında hayatta olan (central exposure) toplam birey sayısını göstermektedir. Buna göre $\log \mu_{0j}$ için Poisson hata terimleri varsayıldığında aşağıdaki GLM modeli oluşturulur.

$$N_x \sim \text{Poisson}(E_x^c \mu_{0i})$$

$$E[N_x] = E_x^c \mu_{0i}$$

$$\log[E[N_x]] = \log E_x^c + \log \mu_{0i} = \log E_x^c + \alpha + \beta x$$

Bu yapıda rassal bileşen $\mu = E[N] = (E_{x_1}^c \mu_{0i}, E_{x_2}^c \mu_{0i}, \dots, E_{x_n}^c \mu_{0i})'$, sistematik bileşen ise $\eta = \log E[N]$ şeklinde tanımlanır. Bu durumda her yaş için model $\eta_x = \log E_x^c + \alpha + \beta x$ formunu alır. Bu durumda modelin genel formu $\eta_x = \log E^c + X\theta$ şeklinde olacaktır.

Veriden, E_x^c ile gösterilen x yaşında hayatta olan kişi sayılarını veren matrisi (exposure matrisi) elde edebilmek için veriler gruplandırılmış ve her grup için istatistikler hesaplanmıştır. Oluşturulan gruplar açıklayıcı değişkenlere göre belirlenmektedir. Örneğin, bir modelde cinsiyet, yaş ve sigara kullanma açıklayıcı değişkenleri kullanılıyorsa, bu durumda yaş için 50, cinsiyet için 2 ve sigara kullanma için 2 durumun çarpımı olan 200 ($50 \times 2 \times 2$) adet alt grup oluşmaktadır. Her alt grup için toplam kişi sayısının yanı sıra kritik hastalığa sahip kişi sayıları da hesaplanmaktadır. Model parametrelerinin tahmininde kullanılan E_x^c matrisi bu şekilde oluşturulur. Çizelge 5.2’de

tezde kurulan modellerde kullanılan açıklayıcı değişkenler ve bu değişkenlerin alabileceği değerler sunulmuştur.

Çizelge 5.2. Kullanılan açıklayıcı değişkenler ve değerleri

Açıklayıcı Değişkenler	Değerler	Tanım
Cinsiyet	0 Erkek	Katılımcının cinsiyeti
	1 Kadın	
Sigara	0 Hayır	Şu anda sigara kullanıyor musunuz?
	1 Evet	
Alkol	0 Hayır	Alkol kullanıyor musunuz?
	1 Evet	
BKI	0 Normal	kilo/boy ² ≤ 25
	1 Aşırı Kilolu	25 < kilo/boy ² ≤ 30
	2 Obez	30 < kilo/boy ²
Tansiyon	0 Hayır	Tansiyon hastalığı teşhisi konuldu mu?
	1 Evet	
Diyabet	0 Hayır	Diyabet hastalığı teşhisi konuldu mu?
	1 Evet	
Psikolojik	0 Hayır	Psikolojik bir rahatsızlık için teşhis konuldu mu?
	1 Evet	
Romatizma	0 Hayır	Romatizma teşhisi konuldu mu?
	1 Evet	
Akciğer	0 Hayır	Kronik bir akciğer hastalığı teşhisi konuldu mu?
	1 Evet	

Tüm hastalıklar için en iyi modelin belirlenebilmesi amacıyla öncelikle yaş ve cinsiyet değişkenlerinin açıklayıcı değişken olarak yer aldığı temel bir model varsayımı yapılmıştır. Temel model aşağıdaki şekilde ifade edilmektedir.

$$\mu_{0i} = \exp(\alpha_{sabit} + \beta_{cinsiyet}x_1 + \beta_{yaş}x_2) \quad i = 1,2,3,4$$

x_1 : Cinsiyet indikatör değişkeni, x_2 : Yaş değişkeni

α_{sabit} : Regresyon sabiti, $\beta_{cinsiyet}$: Cinsiyet değişkeni katsayısı, $\beta_{yaş}$: Yaş değişkeni katsayısı

Sonrasında her hastalık durumu için hasara neden olabilecek, ya da risk durumunu değiştirebileceğini düşündüğümüz açıklayıcı değişkenler modele sistematik bir şekilde eklenerek değerlendirilmiştir.

Model performansları değerlendirilirken ilk aşamada açıklayıcı değişkenler temel modele teker teker eklenmiş ve AIC, BIC istatistikleri değerlendirilerek model performansına en olumlu etkisi olan açıklayıcı değişkenler seçilmiştir.

Devam eden aşamada, bir önceki adımda seçilen açıklayıcı değişkenler gruplanarak iki açıklayıcı değişkenin de aynı anda eklendiği senaryolar çalışılmıştır. Bu aşamada da AIC, BIC istatistiklerine göre model performansları karşılaştırılmış ve en iyi model seçilmiştir.

Bu prosedür her hastalık ve ölüm için ayrı ayrı yapıldığından ve sonuçların sunumunda kullanılmak üzere her model içerisinde yer alan eş değişken gruplarına göre numaralandırılmıştır. Çizelge 5.3'te hangi modelde hangi eş değişkenlerin yer aldığı görülebilir.

Çizelge 5.3. Model numaralarına göre açıklayıcı değişkenler

Model No	Açıklayıcı Değişkenler
Model 1	Yaş&Cinsiyet&Sigara
Model 2	Yaş&Cinsiyet&Alkol
Model 3	Yaş&Cinsiyet&BKI
Model 4	Yaş&Cinsiyet&Tansiyon
Model 5	Yaş&Cinsiyet&Diyabet
Model 6	Yaş&Cinsiyet&Psikolojik
Model 7	Yaş&Cinsiyet&Romatizma
Model 8	Yaş&Cinsiyet&Akciğer
Model 9	Yaş&Cinsiyet&Sigara&Tansiyon
Model 10	Yaş&Cinsiyet&Psikolojik&Romatizma
Model 11	Yaş&Cinsiyet&Sigara&Akciğer
Model 12	Yaş&Cinsiyet&Psikolojik&Akciğer

Modellerde, hasar olasılığının tahmini için önerilen GLM modelinde Poisson dağılımı varsayımı yapıldığı için öncelikle yayılım (dispersion) parametresine bakılmaktadır. Yayılım parametresi, Sapma değeri/SD (serbestlik derecesi) eşitliği ile hesaplanabilmektedir. Önerilen modelde yayılım parametresi 1'e yakın ise, modelin Poisson varsayımı için uygun olduğu söylenebilir. Poisson varsayımının sağlanıp sağlanmadığı oluşturulan tüm modellerde test edilmiştir.

5.2.1. Ölüm Durumu için Modeller

Kritik Hastalık Sigortası poliçelerinde genellikle ölüm durumu için de teminat verildiğinden, ölüm durumunun gerçekleşmesi de tıpkı kritik hastalıklar gibi hasara neden olmaktadır. Markov modelinde μ_{01} ile gösterilen ve sağlıklı durumdan ölüm durumuna anlık geçiş hızının tahmin edilmesinde

$$\mu_{01} = \exp(\alpha_{sabit} + \beta_{cinsiyet}x_1 + \beta_{yaş}x_2 + \beta X)$$

modeli kullanılmıştır. Modelde x_1 cinsiyet indikatör değişkenini, x_2 yaş değişkenini ve **X** modele eklediğimiz diğer açıklayıcı değişkenleri temsil etmektedir. Buna göre regresyon sabiti (α_{sabit}), cinsiyet değişkeni katsayısı ($\beta_{cinsiyet}$), yaş değişkeni katsayısı ($\beta_{yaş}$) ve eklediğimiz değişkenlerin katsayıları (β) için tahminler yapılmıştır.

Modelde cinsiyet ve yaş açıklayıcı değişkeni dışında hangi açıklayıcı değişkenlerin yer alacağına karar verilebilmesi için öncelikle sadece yaş ve cinsiyetin değişken olduğu temel model çıktılarına bakılmış, modelin çıktıları ve parametre tahminleri Çizelge 5.4’te verilmiştir.

Çizelge 5.4. Ölüm için temel model çıktıları

	Tahmin	Std Hata	p-değeri
α_{sabit}	-8,742337	0,228983	<2e-16 ***
$\beta_{yaş}$	0,072342	0,003284	<2e-16 ***
$\beta_{cinsiyet}$	-0,485167	0,049726	<2e-16 ***
Sapma (Boş model)	732,65	SD	81
Sapma	85,47	SD	79
Yayılım	1,08	AIC BIC	403,75 410,96

*p < 0,05; **p < 0,01; ***p < 0,001

Önerilen modelde yayılım parametresi 1’e yakın olduğundan önerilen model Poisson varsayımı için uygundur.

Temel modelin çıktıları değerlendirildiğinde, tüm parametreler için p değeri 0,001’den küçüktür. Önerilen modele göre cinsiyetin ölüm oranları üzerindeki etkisi ters yönlüdür. Diğer bir deyişle, değişkenin değeri arttıkça ölüm oranı azalmaktadır. Bu durumda erkek değişkeni 0 olarak kodlandığından, kadınların daha düşük mortalite oranına sahip olduğu anlaşılmaktadır. Ayrıca yaş ile birlikte ölüm oranları da artmaktadır.

Temel model oluşturulduktan sonra farklı açıklayıcı değişkenlerin tahminleri ve model üzerindeki etkisinin analizi kesim 3.2’de incelenen değişkenlerin her biri modele eklenerek yapılmıştır. Model performansı üzerinde en olumlu etkisi olan değişkenlerin seçimi için AIC (Akaike Bilgi Kriteri), BIC (Bayesçi Bilgi Kriteri) değerleri kullanılmıştır. Değerlendirilen her model için bu değerler Çizelge 5.5’de verilmiştir. Bu

tabloda verilen modellerin hepsinde temel model olarak adlandırdığımız modelde yer alan yaş ve cinsiyet açıklayıcı değişkeni yer almaktadır. Çizelge 5.5 yaş ve cinsiyet açıklayıcı değişkenlerini içeren modele eklenen değişkenleri göstermektedir. Ayrıca model numaralarına göre ilgili modele eklenen değişkenler Çizelge 5.3'te görülebilir.

Çizelge 5.5. Üç değişkenli senaryolar için model performansları

	Açıklayıcı Değişkenler	AIC	BIC
Model 1	Yaş&Cinsiyet&Sigara	659,04	671,44
Model 2	Yaş&Cinsiyet&Alkol	690,85	703,24
Model 3	Yaş&Cinsiyet&BKI	917,43	934,95
Model 4	Yaş&Cinsiyet&Tansiyon	679,9	692,29
Model 5	Yaş&Cinsiyet&Diyabet	717,37	729,76
Model 6	Yaş&Cinsiyet&Psikolojik	661,83	674,20
Model 7	Yaş&Cinsiyet&Romatizma	664,03	676,43
Model 8	Yaş&Cinsiyet&Akciğer	684,05	696,45

Çizelge 5.5'de verilen AIC, BIC değerleri verilen modellerden en küçük değerlere sahip olan değişkenler seçilmiş ve yaş ve cinsiyet dışında iki değişkenin eklendiği iki eş değişkenli senaryolar değerlendirilmiştir. Yapılan değerlendirme sonucu Model 1'de eklenen sigara, Model 4'te kullanılan tansiyon, Model 6'da kullanılan psikolojik ve Model 7'de kullanılan romatizma değişkenleri için iki değişkenli senaryolar denenmiştir. Bu modellerin AIC, BIC kriterlerine göre model performansları değerlendirildiğinde, modelde bir iyileştirme olmadığından, tek değişkenli model ile tahminler yapılmıştır.

Çizelge 5.6. Dört değişkenli senaryolar için model performansları

	Açıklayıcı Değişkenler	AIC	BIC
Model 9	Yaş&Cinsiyet&Sigara&Tansiyon	1084,7	1103,618
Model 10	Yaş&Cinsiyet&Psikolojik&Romatizma	1083,7	1102,63

Bu durumda, en iyi modelin temel model olduğu, açıklayıcı bir değişken eklenmek istendiğinde ise sigaranın etkisinin ölçüldüğü Model 1'in görece diğer modellere göre daha iyi olduğu, bu modele yeni bir değişken eklemenin de model performansını olumsuz yönde etkilediği görülmüştür. Açıklayıcı değişken eklenerek elde edilen modellerden en iyisi olan Model 1 ile sağlıklı durumdan ölüm durumuna geçiş modelinin parametreleri tahmin edilmiştir. Bu modelin çıktıları ve parametre tahminleri Çizelge 5.7'de verilmiştir. Modeldeki tüm parametreler anlamlıdır ve sigara değişkeni eklendiğinde de bu durum değişmemiştir. Sigara değişkeni için de p değeri 0,001'den küçüktür. Ayrıca sigara kullanımının ölümlülük üzerindeki etkisine bakıldığında, sigara içen bireylerin ölüm olasılığının yükseldiği görülmektedir.

Çizelge 5.7. Ölüm modeli ve parametre tahminleri

	Tahmin	Std Hata	p-değeri
α_{sabit}	-9,404223	0,236222	<2e-16 ***
$\beta_{yaş}$	0,073856	0,003288	<2e-16 ***
$\beta_{cinsiyet}$	-0,340319	0,050793	2,08e-11***
β_{sigara}	0,714598	0,059019	<2e-16 ***

*p < 0,05; **p < 0,01; ***p < 0,001

5.2.2. Kanser için Modeller

Kritik Hastalık Sigortası poliçelerinde kanser hastalığı için de teminat verildiğinden, kanser hastalığının teşhisi kritik hastalık sigortası açısından hasara neden olmaktadır. Markov modelinde μ_{02} ile gösterdiğimiz ve sağlıklı durumdan kanser durumuna anlık geçiş hızının tahmin edilmesi için varsayılan model,

$$\mu_{02} = \exp(\alpha_{sabit} + \beta_{cinsiyet}x_1 + \beta_{yaş}x_2 + \beta\mathbf{X})$$

biçiminde oluşturulmuştur. Öncelikle sadece yaş ve cinsiyetin değişken olduğu temel modelin çıktılarına bakılmıştır. Model çıktıları ve parametre tahminleri Çizelge 5.8'de verilmiştir. Modelde x_1 cinsiyet indikatör değişkenini, x_2 yaş değişkenini ve $\mathbf{X}$ modele eklediğimiz diğer açıklayıcı değişkenleri temsil etmektedir. Buna göre regresyon sabiti (α_{sabit}), cinsiyet değişkeni katsayısı ($\beta_{cinsiyet}$), yaş değişkeni katsayısı ($\beta_{yaş}$) ve eklediğimiz değişkenlerin katsayıları (β) için tahminler yapılmıştır.

Çizelge 5.8. Kanser için temel model çıktıları

	Tahmin	Std Hata	p-değeri
α_{sabit}	-7,266609	0,208529	<2e-16 ***
$\beta_{yaş}$	0,051267	0,003036	<2e-16 ***
$\beta_{cinsiyet}$	-0,397245	0,047594	<2e-16 ***
Sapma (Boş model)	567,12	SD	81
Sapma	180,40	SD	79
Yayılım	2,28	AIC BIC	506,6 513,81

*p < 0,05; **p < 0,01; ***p < 0,001

Önerilen modelde yayılım parametresi 2,28 olarak hesaplanmıştır. Bu değerin 1'den büyük olması fazla yayılım (overdispersion) olduğunu göstermektedir. Bu durumda standart hatalar yanıltıcı olmaktadır. Dolayısıyla model parametrelerinin anlamlı olup olmadığına karar verilememektedir. Yazında, Poisson varsayımı ile yapılan modellerde fazla yayılım (overdispersion) görüldüğü durumda quasiPoisson dağılımı önerilmektedir [15]. Fazla yayılım görüldüğü durumda GLM modeli ile tahmin yapabilmek için quasiPoisson varsayımı ile model parametreleri yeniden bulunmuş ve sonuçlar Çizelge 5.9'da verilmiştir.

Çizelge 5.9. Kanser için temel model çıktıları (quasiPoisson)

	Tahmin	Std Hata	p-değeri
α_{sabit}	-7,266609	0,299149	<2e-16 ***
$\beta_{yaş}$	0,051267	0,004356	<2e-16 ***
$\beta_{cinsiyet}$	-0,397245	0,068277	1,21e-07***

*p < 0,05; **p < 0,01; ***p < 0,001

Temel modelde quasiPoisson varsayımı yapıldığında, parametre tahminleri için standart hatanın arttığı ancak halen p değerlerinin 0,001'den küçük olduğu görülmektedir. Önerilen modele göre cinsiyetin kanser hastalığı olasılığı üzerindeki etkisi ters yönlüdür. Bu durumda erkek değişkeni 0 olarak kodlandığından, kadınlarda kanser hastalığı teşhisi olasılığı daha düşüktür. Yaş değişkeni ile birlikte kanser hastalığının görülme olasılığı artmaktadır.

Temel model oluşturulduktan sonra eklenen değişkenler için parametre tahminleri incelendiğinde, Model 2’de eklenen Alkol değişkeni ve Model 3’te eklenen BKİ değişkeni anlamlı çıkmamıştır. Model performansı üzerinde en olumlu etkisi olan değişkenlerin seçimi için AIC, BIC değerlendirildiğinde ise, Model 1’de eklenen sigara değişkeni ve Model 8’de eklenen akciğer değişkeni seçilmiştir.

Çizelge 5.10. Üç değişkenli senaryolar için model performansları (Kanser)

	Açıklayıcı Değişkenler	AIC	BIC
Model 1	Yaş&Cinsiyet&Sigara	807,61	820,00
Model 2	Yaş&Cinsiyet&Alkol	809,18	828,58
Model 3	Yaş&Cinsiyet&BKİ	1049,1	1066,67
Model 4	Yaş&Cinsiyet&Tansiyon	804,26	816,65
Model 5	Yaş&Cinsiyet&Diyabet	741,47	753,86
Model 6	Yaş&Cinsiyet&Psikolojik	749,27	761,64
Model 7	Yaş&Cinsiyet&Romatizma	793,93	806,33
Model 8	Yaş&Cinsiyet&Akciğer	732,5	744,90

Çizelge 5.10’de verilen AIC, BIC kriterleri verilen modellerden en küçük değerlere sahip olan değişkenler olan sigara ve akciğer açıklayıcı değişkenleri seçilmiş, yaş ve cinsiyet dışında iki değişkenin eklendiği dört açıklayıcı değişkenli senaryolar değerlendirilmiştir. Bu modellerin AIC, BIC kriterlerine göre model performansları, modeli iyileştirmedikten, tek değişkenli model ile tahmin yapılmasına karar verilmiştir.

Çizelge 5.11. Dört değişkenli senaryolar için model performansları (Kanser)

	Açıklayıcı Değişkenler	AIC	BIC
Model 11	Sigara&Akciğer	1093,2	1111,93

Yapılan değerlendirmeler sonucunda sigaranın etkisinin ölçüldüğü Model 1 ile sağlıklı durumdan, kanser hastalığı durumuna geçiş modelinin parametreleri tahmin edilmiştir. Bu modelin çıktıları ve parametre tahminleri Çizelge 5.12’de verilmiştir. Modeldeki tüm parametreler anlamlıdır. Modele sigara değişkeni eklendiğinde de bu durum değişmemiştir. Sigara değişkeni için de p değeri 0,001’den küçüktür. Ayrıca sigara kullanımının kanser hastalığı üzerindeki etkisine bakıldığında, sigara içen bireylerde kanser hastalığı teşhis olasılığının yükseldiği görülmektedir.

Çizelge 5.12. Kanser modeli ve parametre tahminleri

	Tahmin	Std Hata	p-değeri
α_{sabit}	-7,428893	0,268412	<2e-16 ***
$\beta_{yaş}$	0,051680	0,003818	<2e-16 ***
$\beta_{cinsiyet}$	-0,357197	0,061359	3,1e-08 ***
β_{sigara}	0,182618	0,064261	0,00507 **

*p < 0,05; **p < 0,01; ***p < 0,001

5.2.3. Kalp Hastalığı için Modeller

Kritik Hastalık Sigortası poliçelerinde teminat altına alınan hastalıklardan birisi de Kalp hastalıklarıdır. Kalp hastalığının da kritik hastalık poliçesi bakımından hasara neden olacağı değerlendirilmiştir. Markov modelinde μ_{03} ile gösterilen ve sağlıklı durumdan kalp hastalığı durumuna anlık geçiş hızının tahmini için model,

$$\mu_{03} = \exp(\alpha_{sabit} + \beta_{cinsiyet}x_1 + \beta_{yaş}x_2 + \beta X)$$

biçiminde oluşturulmuştur. Modelde x_1 cinsiyet indikatör değişkenini, x_2 yaş değişkenini ve X modele eklediğimiz diğer açıklayıcı değişkenleri temsil etmektedir. Buna göre regresyon sabiti (α_{sabit}), cinsiyet değişkeni katsayısı ($\beta_{cinsiyet}$), yaş değişkeni katsayısı ($\beta_{yaş}$) ve eklediğimiz değişkenlerin katsayıları (β) için tahminler yapılmıştır.

Modelde cinsiyet ve yaş açıklayıcı değişkeni dışında hangi açıklayıcı değişkenlerin yer alacağına karar verilebilmesi için öncelikle sadece yaş ve cinsiyetin değişken olduğu

temel model çıktılarına bakılmıştır. Model çıktıları ve parametre tahminleri Çizelge 5.13'te verilmiştir.

Çizelge 5.13. Kalp hastalığı için temel model çıktıları

	Tahmin	Std Hata	p-değeri
α_{sabit}	-6,686335	0,166839	<2e-16 ***
$\beta_{yaş}$	0,048918	0,002433	<2e-16 ***
$\beta_{cinsiyet}$	-0,370623	0,038276	<2e-16 ***
Sapma (Boş model)	659,62	SD	81
Sapma	118,27	SD	79
Yayılım	1,87	AIC BIC	491,3 498,52

*p < 0,05; **p < 0,01; ***p < 0,001

Önerilen modelde yayılım parametresi 1,49 olarak hesaplanmıştır. Kanser hastalığı için yapılan quasiPoisson varsayım ile model parametreleri tekrar tahmin edilmiş ve sonuçlar Çizelge 5.14'te verilmiştir.

Çizelge 5.14. Kalp hastalığı için temel model çıktıları (quasiPoisson)

	Tahmin	Std Hata	p-değeri
α_{sabit}	-6,686335	0,193631	<2e-16 ***
$\beta_{yaş}$	0,048918	0,002824	<2e-16 ***
$\beta_{cinsiyet}$	-0,370623	0,044422	1,87e-12 ***

*p < 0,05; **p < 0,01; ***p < 0,001

Temel modelde quasiPoisson varsayımı yapıldığında, parametre tahminleri için standart hatanın arttığı ancak halen p değerlerinin 0,001'den küçük olduğu görülmektedir. Önerilen modele göre cinsiyetin kalp hastalığı olasılığı üzerindeki etkisi ters yönlüdür. Bu durumda erkek değişkeni 0 olarak kodlandığından, kadınlarda kalp hastalığı teşhisi olasılığı daha düşüktür. Yaş değişkeninin artması ile birlikte ise kalp hastalığı görülme olasılığı da artmaktadır.

Temel model oluşturulduktan sonra ölüm ve kanser hastalığında olduğu gibi açıklayıcı değişkenlerin her biri modele teker teker eklenerek değerlendirildiğinde elde edilen sonuçlar Çizelge 5.15’te verilmiştir.

Çizelge 5.15. Üç değişkenli senaryolar için model performansları (Kalp hastalığı)

	Açıklayıcı Değişkenler	AIC	BIC
Model 1	Yaş&Cinsiyet&Sigara	790,29	802,69
Model 2	Yaş&Cinsiyet&Alkol	831,27	843,66
Model 3	Yaş&Cinsiyet&BKI	1121,9	1139,45
Model 4	Yaş&Cinsiyet&Tansiyon	835,52	847,91
Model 5	Yaş&Cinsiyet&Diyabet	837,57	849,97
Model 6	Yaş&Cinsiyet&Psikolojik	821,77	834,14
Model 7	Yaş&Cinsiyet&Romatizma	826,36	838,76
Model 8	Yaş&Cinsiyet&Akciğer	808,16	820,56

Çizelge 5.15’te verilen AIC, BIC ve Sapma değerlerine göre yapılan değerlendirme sonucu, Model 1’e eklenen sigara, Model 8’de kullanılan akciğer değişkenleri için iki değişkenli senaryolar denenmiştir. Bu modellerin AIC ve BIC kriterlerine göre model performansları değerlendirildiğinde, modelde iyileşme görülmediğinden, tek değişkenli model ile tahminler yapılmıştır.

Çizelge 5.16. Dört değişkenli senaryolar için model performansları (Kalp hastalığı)

	Açıklayıcı Değişkenler	AIC	BIC
Model 11	Yaş&Cinsiyet&Sigara&Akciğer	1228,3	1247,01

Bu durumda, sigaranın etkisinin ölçüldüğü Model 1 ile sağlıklı durumdan kalp hastalığı durumuna geçiş parametresi tahmin edilmiştir. Bu modelin çıktıları ve parametre tahminleri aşağıdaki Çizelge 5.17’de verilmiştir. Modeldeki tüm parametreler anlamlıdır ve sigara değişkeni eklendiğinde de bu durum değişmemiştir. Sigara değişkeni için de p

değeri 0,001'den küçüktür. Ayrıca sigara kullanımının kalp hastalığı üzerindeki etkisine bakıldığında, sigara içen bireylerde kalp hastalığına geçiş olasılığının yükseldiği görülmektedir.

Çizelge 5.17. Kalp hastalığı modeli ve parametre tahminleri

	Tahmin	Std Hata	p-değeri
α_{sabit}	-6,855941	0,177240	<2e-16 ***
$\beta_{yaş}$	0,049350	0,002524	<2e-16 ***
$\beta_{cinsiyet}$	-0,328886	0,040706	1,49e-13 ***
β_{sigara}	0,190683	0,042670	1,48e-05***

*p < 0,05; **p < 0,01; ***p < 0,001

5.2.4. Felç Hastalığı için Modeller

Kritik Hastalık Sigortası poliçelerinde teminat altına alınan hastalıklardan bir diğeri de Felç hastalıklarıdır. Felç hastalığının da kritik hastalık poliçesi bakımından hasara neden olacağı değerlendirilmiştir. Markov modelinde μ_{04} ile gösterilen ve sağlıklı durumdan felç durumuna anlık geçiş hızının tahmini için varsayılan temel modelin parametreleri Çizelge 5.18'de verilmiştir.

Çizelge 5.18. Felç hastalığı için temel model çıktıları

	Tahmin	Std Hata	p-değeri
α_{sabit}	-8,580269	0,297339	< 2e-16 ***
$\beta_{yaş}$	0,059671	0,004298	< 2e-16 ***
$\beta_{cinsiyet}$	-0,252290	0,066559	0,00015***
Sapma (Boş model)	327,29	SD	81
Sapma	93,83	SD	79
Yayılım	1,18	AIC BIC	387,07 394,28

*p < 0,05; **p < 0,01; ***p < 0,001

Önerilen modelde yayılım parametresi 1'e yakın olduğundan, model Poisson varsayımı için uygundur. Tüm parametreler için p değeri 0,001'den küçüktür. Önerilen modele göre

cinsiyetin ölüm oranları üzerindeki etkisi ters yönlüdür. Bu durumda erkek değişkeni 0 olarak kodlandığından, kadınların felç geçirme olasılığı daha düşük tahmin edilmiştir. Ayrıca diğer kritik hastalıklarda olduğu gibi yaş ile birlikte hastalık olasılığı da artmaktadır.

Sonraki aşamada modele eklenen açıklayıcı değişkenler için yapılan parametre tahminlerinin tümü anlamlıdır. Model performansı üzerinde en olumlu etkisi olan değişkenlerin seçimi için AIC, BIC değerleri Çizelge 5.19’da görülebilir.

Çizelge 5.19. Üç değişkenli senaryolar için model performansları (Felç hastalığı)

	Açıklayıcı Değişkenler	AIC	BIC
Model 1	Yaş&Cinsiyet&Sigara	615,54	627,93
Model 2	Yaş&Cinsiyet&Alkol	624,35	636,75
Model 3	Yaş&Cinsiyet&BKI	839,38	856,90
Model 4	Yaş&Cinsiyet&Tansiyon	636,23	648,632
Model 5	Yaş&Cinsiyet&Diyabet	645,86	658,26
Model 6	Yaş&Cinsiyet&Psikolojik	613,84	626,21
Model 7	Yaş&Cinsiyet&Romatizma	643,09	655,48
Model 8	Yaş&Cinsiyet&Akciğer	572,58	584,97

Çizelge 5.19’da verilen modellerden en küçük AIC ve BIC değerlerine sahip olan değişkenler seçilmiş ve yaş ve cinsiyet dışında iki değişkenin eklendiği iki eş değişkenli senaryolar değerlendirilmiştir. Yapılan değerlendirme sonucu Model 6’da eklenen psikolojik değişkeni, Model 8’de kullanılan akciğer değişkenleri için iki değişkenli senaryolar denenmiştir. Bu modellerin AIC, BIC ve Sapma değerine göre model performansları değerlendirildiğinde, modeli iyileştirmedikten, tek değişkenli model ile tahminler yapılmıştır.

Çizelge 5.20. Dört değişkenli senaryolar için model performansları (Felç hastalığı)

	Açıklayıcı Değişkenler	AIC	BIC
Model 12	Yaş&Cinsiyet&Psikolojik&Akciğer	874,73	893,57

Bu durumda akciğer hastalığının etkisinin ölçüldüğü Model 8 ile sağlıklı durumdan felç durumuna geçiş modelinin parametreleri tahmin edilmiştir. Bu modelin çıktıları ve parametre tahminleri aşağıdaki Çizelge 5.21’de verilmiştir. Modeldeki tüm parametreler anlamlıdır ve akciğer değişkeni eklendiğinde de bu durum değişmemiştir. Akciğer değişkeni için de p değeri 0,001’den küçüktür. Ayrıca akciğer hastalığının felç üzerindeki etkisine bakıldığında, akciğer hastalığı bulunan bireylerde felç olasılığının yükseldiği görülmektedir.

Çizelge 5.21. Felç hastalığı modeli ve parametre tahminleri

	Tahmin	Std Hata	p-değeri
α_{sabit}	-8,508583	0,297819	<2e-16 ***
$\beta_{yaş}$	0,057832	0,004318	<2e-16 ***
$\beta_{cinsiyet}$	-0,252123	0,06656	0,000152***
$\beta_{akciğer}$	0,575452	0,103428	2,64e-08***

*p < 0,05; **p < 0,01; ***p < 0,001

5.3. Saklı Markov Modeli

Tezde kullanılan en başarılı Saklı Markov modelini bulmak amacıyla, Markov modelinde olduğu gibi sadece yaş ve cinsiyet değişkenini içeren temel bir model ile başlanmıştır. Öncelikle saklı durum sayıları test edilmiştir. Sonrasında ise açıklayıcı değişkenler eklenerek AIC ve BIC kriterlerine göre en başarılı model yapısı araştırılmıştır

Tüm Saklı Markov modellerinde kullanılan gösterim şu şekildedir:

- Gözlem kümesi, $V = \{v_1, v_2, \dots, v_{14}\}$ şeklinde her katılımcının $M = 14$ gözlemi için tanımlanmıştır. Bu tezde kullanılan Saklı Markov modellerinde aynı anda ölüm, kanser, kalp hastalığı, felç durumları modellendiği için;

$$v_i = \{v_i^{\text{ölüm}}, v_i^{\text{kanser}}, v_i^{\text{kalp}}, v_i^{\text{felç}}\}$$

şeklinde gözlem kümesi tanımlanmıştır.

- Başlangıç durumu için olasılık dağılımı, $\pi = \{\pi_i\}$ kümesi ile ifade edilir ve her saklı durum için (toplam N adet) belirlenir.

$$\pi_i = P[q_1 = S_j], \quad 1 \leq j \leq N$$

- Saklı durumlar arası geçiş olasılık dağılımı $A = \{a_{ij}\}$ kümesi,

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, \quad j \leq N$$

eşitliği ile ifade edilir.

Açıklayıcı değişkenlerin kullanıldığı durumlarda $P[q_{t+1} = S_j | q_t = S_i]$ olasılık dağılımları açıklayıcı değişkenlerin $f(\cdot)$ şeklinde bir fonksiyonu olarak ifade edilebilir.

- Her j saklı durumu için gözlem olasılık dağılım fonksiyonu, $B = \{b_j(k)\}$ kümesi ile ifade edilir. İfadede kullanılan v_k^h , h hastalığı için k zamanındaki gözlemdir. Toplam 14 adet gözlem olduğu için her bir gözlem zamanı k ile gösterilmiştir.

$$b_j(k) = P[v_k^h | S_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq 14$$

Her model için gözlem sayısı sabit ve 14'tür. Modelde değişen olan parametreler, modelde kullanılan açıklayıcı değişkenler ve modelin yapısını belirleyen saklı durumların sayısıdır. Her modelde açıklayıcı değişkenlerin kullanıldığı $f(\cdot)$ fonksiyonu multinomial logit fonksiyonudur.

$$f(\cdot) = f(x_1, \dots, x_k) = \exp\left(\alpha + \sum_{u=1}^k \beta x_u\right)$$

Bu kısımda $x_1, x_2, \dots, x_k$ açıklayıcı değişkenleri ve $N = n$ saklı durumlu kurulan Saklı Markov modelleri,

$$\lambda(f(x_1, x_2, \dots, x_k), N = n)$$

şeklinde ifade edilmiştir.

5.3.1. Temel Model

Tezde önerilen SMM ile ilgili analizlere başlarken ilk adımda yaş ve cinsiyet değişkeninin açıklayıcı değişken olarak kullanıldığı temel bir SMM geliştirilmiştir. SMM ile ilgili verilmesi gereken en önemli karar, modelin yapısını belirleyen saklı durumların sayısıdır.

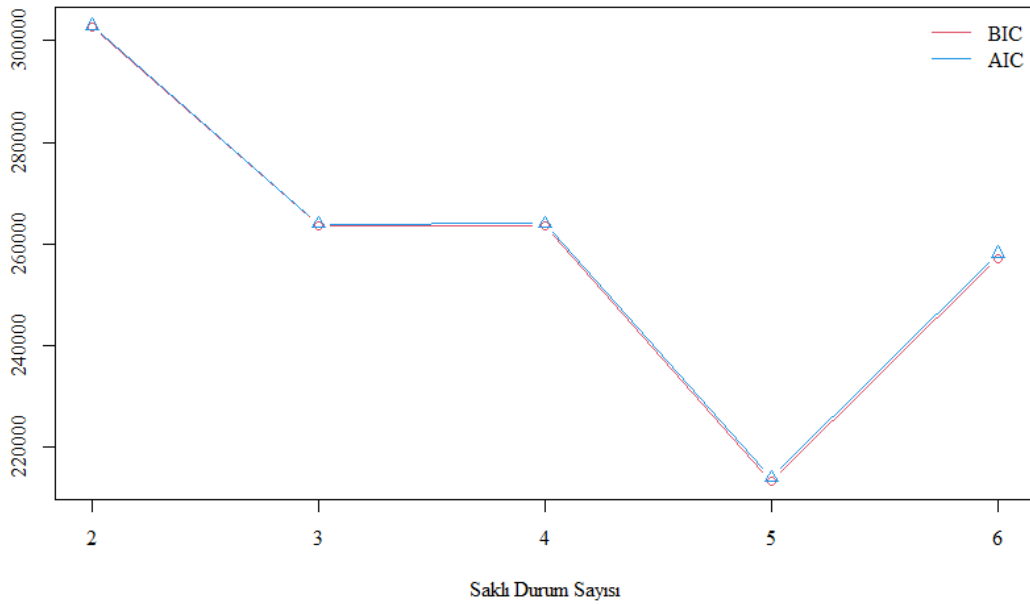
Temel modelde farklı saklı durum sayıları için analizler yapılmış ve oluşturulan modeller bu kesimde verilmiştir.

Yaş ve cinsiyet açıklayıcı değişkenlerini içeren temel modelde 2 ile 6 arasında değişen saklı durum sayıları için

$$\lambda(f(x_{age}, x_{cinsiyet}), N = \{2,3,4,5,6\})$$

modellerinin performansları karşılaştırılmıştır.

Model performansları karşılaştırılırken AIC ve BIC istatistiklerinden faydalanılmıştır. Her saklı durum sayısı için performans istatistikleri Şekil 5.1’de görülmektedir. Buna göre en başarılı performans $N = 5$ saklı durum sayısı için elde edilmiştir.



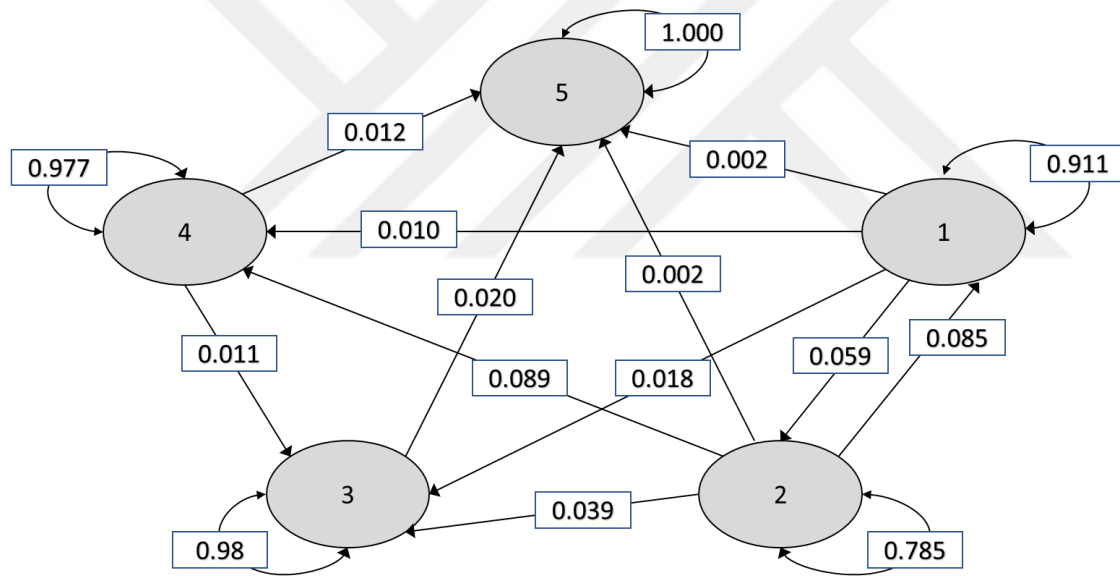
Şekil 5.2. Saklı durum sayısına göre model performansları

Belirlenen saklı durum sayısına göre saklı durumlar arası ortalama geçiş olasılıkları aşağıdaki Çizelge 5.22’de ve modelin görseli Şekil 5.2’de verilmiştir. Buna göre saklı durum 5’ten diğer saklı durumlara geçiş olmadığı, aynı durumda kalma eğilimi olduğu anlaşılmaktadır. Ayrıca diğer saklı durumlarda da genel anlamda bireylerin daha yüksek olasılıkla bulundukları saklı durumda kalma eğiliminde olduğu söylenebilir. En fazla

geçişler saklı durum 1 ve saklı durum 2’de olmaktadır. Saklı durum 1’den yapılan geçişlere bakıldığında, %5,9 olasılıkla geçiş yapacağı diğer saklı durumun saklı durum 2, saklı durum 2’den yapılan geçişlere bakıldığında ise %8,9 olasılıkla saklı durum 4’e geçiş yapacağı görülmektedir.

Çizelge 5.22. Saklı durumlar arası ortalama geçiş olasılıkları

	Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
Saklı 1	0,911	0,059	0,018	0,010	0,002
Saklı 2	0,085	0,785	0,039	0,089	0,002
Saklı 3	0,000	0,000	0,980	0,000	0,020
Saklı 4	0,000	0,000	0,011	0,977	0,012
Saklı 5	0,000	0,000	0,000	0,000	1,000



Şekil 5.3. Saklı durumlar arası ortalama geçiş olasılıkları

Saklı durumlar arası geçişlerde temel model için kullanılan yaş ve cinsiyet açıklayıcı değişkenlerinin etkisini gösteren logit parametre tahminleri ve standart hataları Çizelge 5.23, Çizelge 5.24, Çizelge 5.25 ve Çizelge 5.26’te verilmiştir. Tez çalışması için önerilen 5 durumlu Saklı Markov modelinde durumlar arası olası toplam 25 adet geçiş bulunmaktadır. Bu geçişlere bakıldığında, Çizelge 5.22’de görüldüğü üzere 16 adet geçiş için olasılık verilmiştir. Önerilen temel modelde yer alan saklı durumlar arası 16 geçiş için açıklayıcı değişkenlerin etkisine bakılmıştır.

Her geiř iin yař ve cinsiyet aıklayıcı deęiřkeninin etkisini gsteren logit parametre tahminleri ve standart hatalarının ayrı ayrı incelenmesi, hangi deęiřkeninin hangi saklı durumlar arası geiřlerde etkili olduęunun anlařılması iin nemlidir.

Saklı durum 1’den yapılan geiřler iin parametre tahminleri izelge 5.23’te verilmiřtir. izelgenin st kısmında yer alan deęerler parametre tahminlerini, alt kısmında yer alan deęerler standart hataları gstermektedir.

izelge 5.23. Saklı durum 1’den geiř olasılıkları iin parametre tahminleri

Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
α_{sabit}	3,0281°	-6,3714	-8,9813	-4,5818
$\beta_{cinsiyet}$	-0,3115°	-0,1083°	5,4102	-1,4574
$\beta_{yař}$	-0,0898°	0,0379	-0,0070°	-0,0143°
α_{sabit}	1,9059	0,7728	0,9809	0,9596
$\beta_{cinsiyet}$	0,914	0,2918	0,9499	0,4600
$\beta_{yař}$	0,0530	0,0120	0,0368	0,0148

°: Standart hataya gre anlamlı olmayan parametre tahminleri

ncelikle parametrelerin anlamlı olup olmadıęına bakıldıęında, saklı durum 1’den saklı durum 2, saklı durum 3’e geiřte cinsiyet deęiřkeni, saklı durum 1, saklı durum 4 ve Saklı durum 5’e geiřte de yař deęiřkeni anlamlı deęildir. Yař deęiřkeninin etkisine bakıldıęında, yař ilerledike saklı durum 3’e geiř olasılıęı artmaktadır. Kadınların saklı durum 3’e geme olasılıkları erkeklere gre daha fazladır. Saklı durum 5’e geiř durumuna bakıldıęında ise tam tersi sz konusudur. Kadınların saklı durum 5’e geme olasılıęı erkeklere gre daha azdır.

Saklı durum 2’den yapılan geiřler iin parametre tahminleri ařaęıdaki izelge 5.24’te verilmiřtir. izelgenin st kısmında yer alan deęerler parametre tahminlerini alt kısmında yer alan deęerler standart hataları gstermektedir.

Çizelge 5.24. Saklı durum 2’den geçiş olasılıkları için parametre tahminleri

Saklı 2	Saklı 1	Saklı 3	Saklı 4	Saklı 5
α_{sabit}	-3,9110 ^o	-6,3507	-8,5369	-24,8638
$\beta_{cinsiyet}$	0,2639 ^o	-0,4962	-0,6567	2,5905 ^o
$\beta_{yaş}$	-0,0233 ^o	0,0555	0,1003	0,2372
α_{sabit}	2,2042	0,6897	0,3661	1,4792
$\beta_{cinsiyet}$	0,7428	0,1433	0,1767	2,7535
$\beta_{yaş}$	0,0172	0,0105	0,0050	0,0568

^o: Standart hataya göre anlamlı olamayan parametre tahminleri

Çizelge 5.24’te parametrelerin anlamlı olup olmadığına bakıldığında, saklı durum 2’den saklı durum 1’e geçişte yaş değişkeni anlamlı değildir. Diğer tüm parametreler anlamlıdır. Yaş değişkeninin etkisine bakıldığında, yaş ilerledikçe saklı durum 2’den diğer durumlara geçiş olasılığının arttığı görülmektedir. Kadınların da saklı durum 3 ve saklı durum 4’e geçme olasılıkları erkeklere göre daha küçüktür.

Saklı durum 3’ten yapılan geçişler için parametre tahminleri Çizelge 5.25’te verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini, alt kısmında yer alan değerler standart hataları göstermektedir. Saklı durum 3’ten saklı durum 1, saklı durum 2 ve saklı durum 4’e geçiş mümkün olmadığı için sadece saklı durum 5’e geçiş olasılığı için parametre tahminleri verilmiştir.

Çizelge 5.25. Saklı durum 3’ten geçiş olasılıkları için parametre tahminleri

Saklı 3	Saklı 1	Saklı 2	Saklı 4	Saklı 5
α_{sabit}	-	-	-	-1,0894 ^o
$\beta_{cinsiyet}$	-	-	-	0,1087 ^o
$\beta_{yaş}$	-	-	-	-0,0431
α_{sabit}	-	-	-	0,9735
$\beta_{cinsiyet}$	-	-	-	0,1902
$\beta_{yaş}$	-	-	-	0,0141

^o: Standart hataya göre anlamlı olamayan parametre tahminleri

Parametrelerin anlamlı olup olmadığına bakıldığında, saklı durum 3’den saklı durum 5’e geçişte cinsiyet değişkeni anlamlı değildir. Yaş değişkeninin etkisine bakıldığında, yaş ilerledikçe saklı durum 3’ten diğer 2 saklı duruma geçiş olasılığının azaldığı görülmektedir. Erkeklerin saklı durum 4’e geçme olasılıkları kadınlara göre daha büyüktür.

Saklı durum 4’ten yapılan geçişler için parametre tahminleri aşağıdaki Çizelge 5.26’te verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini, alt kısmında yer alan değerler standart hataları göstermektedir. Saklı durum 4’ten yalnızca saklı durum 5’e geçiş yapılabildiği için sadece saklı durum 5 için parametre tahminleri verilmiştir.

Çizelge 5.26. Saklı durum 4’ten geçiş olasılıkları için parametre tahminleri

Saklı 4	Saklı 1	Saklı 2	Saklı 3	Saklı 5
α_{sabit}	-	-	-4,9901	-3,3062
$\beta_{cinsiyet}$	-	-	-0,1133 °	0,1197 °
$\beta_{yaş}$	-	-	0,0094 °	-0,0186 °
α_{sabit}	-	-	1,1708	1,1814
$\beta_{cinsiyet}$	-	-	0,1996	0,2179
$\beta_{yaş}$	-	-	0,0163	0,0166

°: Standart hataya göre anlamlı olamayan parametre tahminleri

Çizelge 5.26’da verilen değerlere göre, saklı durum 4’ten saklı durum 3 ve 5’e geçişte ise yaş ve cinsiyet açıklayıcı değişkeninin anlamlı olmadığı görülmüştür.

Ayrıca Saklı Markov modelinin bir çıktısı olan her saklı durum altında gözlemlenen kritik hastalıklar ve ölüm için koşullu olasılıklar matrisi (emisyon olasılıkları matrisi) Çizelge 5.27’de verilmiştir.

Çizelge 5.27. Saklı durumlar için koşullu olasılık matrisi

	Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
Ölüm	0,000	0,000	0,2103	0,5047	0,3141
Kanser	0,000	0,000	0,1466	0,4583	0,1946
Kalp H.	0,000	0,000	1,0000	0,000	1,0000
Felç	0,000	0,000	0,000	0,1818	1,0000

Çizelge 5.27’de verilen olasılıklar değerlendirildiğinde, saklı durum 1 ve saklı durum 2’de yer alan bireylerin sağlıklı olduğu, saklı durum 3’te yer alan bireylerin Kalp hastası olduğu, saklı durum 4’te yer alan bireylerin daha yüksek olasılıkla kanser hastası, saklı durum 5’te yer alan bireylere de felç teşhisi konulduğu görülmektedir.

5.3.2. Farklı Açıklayıcı Değişken Senaryoları

Bu kısımda, Saklı Markov modeller için oluşturulan temel modele, HRS verisinde yer alan değişkenlerden kritik hastalıkların olasılık tahminlerinde anlamlı olabilecek farklı açıklayıcı değişkenler eklenerek model performansları değerlendirilmiştir. Sistematiik bir değerlendirme olması için her açıklayıcı değişken teker teker eklenerek modeller AIC ve BIC kriterlerine göre değerlendirilmiştir. En iyi model belirlendikten sonra, ikinci bir açıklayıcı değişken eklenerek model performans istatistiklerinde iyileşme olup olmadığına bakılmıştır.

Çizelge 5.28. Saklı Markov modellerinin karşılaştırılması

	AIC	BIC
$\lambda(f(x_{age}, x_{cinsiyet}), N = 5)$	84324,7	84926,01
$\lambda(f(x_{age}, x_{cinsiyet}, x_{sigara}), N = 5)$	84274,2	85032,39
$\lambda(f(x_{age}, x_{cinsiyet}, x_{alkol}), N = 5)$	84162,44	84920,62
$\lambda(f(x_{age}, x_{cinsiyet}, x_{BKI}), N = 5)$	84217,7	84975,88
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}), N = 5)$	83720,43	84478,61
$\lambda(f(x_{age}, x_{cinsiyet}, x_{diyabet}), N = 5)$	84014,26	84772,44
$\lambda(f(x_{age}, x_{cinsiyet}, x_{psikolojik}), N = 5)$	84104,3	84862,48
$\lambda(f(x_{age}, x_{cinsiyet}, x_{akciger}), N = 5)$	84057,59	84815,77
$\lambda(f(x_{age}, x_{cinsiyet}, x_{romatizma}), N = 5)$	84099,95	84858,13

Farklı açıklayıcı değişkenler eklenerek oluşturulan Saklı Markov modelleri için AIC ve BIC değerleri Çizelge 5.28’de görülmektedir. Bu kriterlere göre en iyi modelin yaş, cinsiyet ve tansiyon açıklayıcı değişkenlerini içeren $\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}), N = 5)$ modeli olduğuna karar verilmiştir. Bir sonraki adımda bu değişkenlerin yanına yeni bir açıklayıcı değişken eklenerek AIC ve BIC kriterlerine göre tekrar değerlendirme yapılmıştır. Sonuçlar Çizelge 5.29’da verilmiştir.

Çizelge 5.29. Saklı Markov modellerinin karşılaştırılması

	AIC	BIC
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}), N = 5)$	83720,43	84478,61
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{sigara}), N = 5)$	83683,76	84598,8
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{alkol}), N = 5)$	83679,4	84594,44
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{BKI}), N = 5)$	83711,9	84626,95
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{diyabet}), N = 5)$	83539,71	84454,75
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{psiko}), N = 5)$	83590,65	84505,7
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}), N = 5)$	83497,11	84412,16
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{romatizma}), N = 5)$	83583,91	84498,96

Çizelge 5.29’da verilen değerlere göre en iyi model yaş ve cinsiyet açıklayıcı değişkeninin yanında tansiyon ve akciğer açıklayıcı değişkeninin de yer aldığı $\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}), N = 5)$ modeldir. Modelin daha da iyileşmesi adına bu kısımda belirlenen değişkenlere bir açıklayıcı değişken daha eklenerek AIC ve BIC istatistikleri hesaplanmış ve Çizelge 5.30’da verilmiştir.

Çizelge 5.30. Saklı Markov modellerinin karşılaştırılması

	AIC	BIC
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}), N = 5)$	83497,11	84412,16
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{sigara}), N = 5)$	83489,53	84561,44
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{alkol}), N = 5)$	83466,85	84538,76
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{BKI}), N = 5)$	83501,12	84573,02
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}), N = 5)$	83326,92	84398,82
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{psiko}), N = 5)$	83414,1	84486
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{romatizma}), N = 5)$	83404,33	84476,24

Çizelge 5.30’da verilen modeller değerlendirildiğinde ise, en iyi modelin yaş ve cinsiyet açıklayıcı değişkeninin yanında tansiyon, diyabet ve psikolojik açıklayıcı değişkeninin de yer aldığı $\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}), N = 5)$ modeli olduğu görülmektedir. Bir sonraki adımda tüm açıklayıcı değişkenler teker teker bu modele eklenerek model performansları değerlendirilmiş ve en iyi model tespit edilmeye çalışılmıştır.

Çizelge 5.31. Saklı Markov modellerinin karşılaştırılması

	AIC	BIC
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}), N = 5)$	83326,92	84398,82
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{sigara}), N = 5)$	83329,16	84557,93
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{alkol}), N = 5)$	83330,43	84559,2
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{BKI}), N = 5)$	83323,11	84551,88
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{psiko}), N = 5)$	83249,55	84478,32
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{romat}), N = 5)$	83259,92	84488,69

Çizelge 5.31’de görüleceği üzere bir önceki aşamada belirlenen $\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}), N = 5)$ modeline psikoloji açıklayıcı değişkeni eklendiğinde model performansı belirlenen kriterlere göre iyileşmiştir. Modelin daha da iyileşmesi adına aynı sistemle bir açıklayıcı değişken daha eklenerek AIC ve BIC istatistikleri hesaplanmış ve Çizelge 5.32’de verilmiştir.

Çizelge 5.32. Saklı Markov modellerinin karşılaştırılması

	AIC	BIC
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{psiko}), N = 5)$	83249,55	84478,32
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{sigara}), N = 5)$	*	*
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{alkol}), N = 5)$	83273,82	84659,46
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{BKI}), N = 5)$	83230,22	84615,85
$\lambda(f(x_{age}, x_{cinsiyet}, x_{tansiyon}, x_{akciger}, x_{diyabet}, x_{romat}), N = 5)$	83251,16	84596,8

*: Algoritma çözüm üretemedi

Bu aşamadan sonra eklenen değişkenler model performansını iyileştirmedeğinden, Saklı Markov modeli için en iyi modelin yaş, cinsiyet, tansiyon, akciğer, diyabet ve psikoloji açıklayıcı değişkenleri kullanıldığı model olduğuna karar verilmiştir. Çizelge 5.32’de bu modele eklenecek yeni bir açıklayıcı değişkenin model performansının iyileştirmedeğği görülebilir.

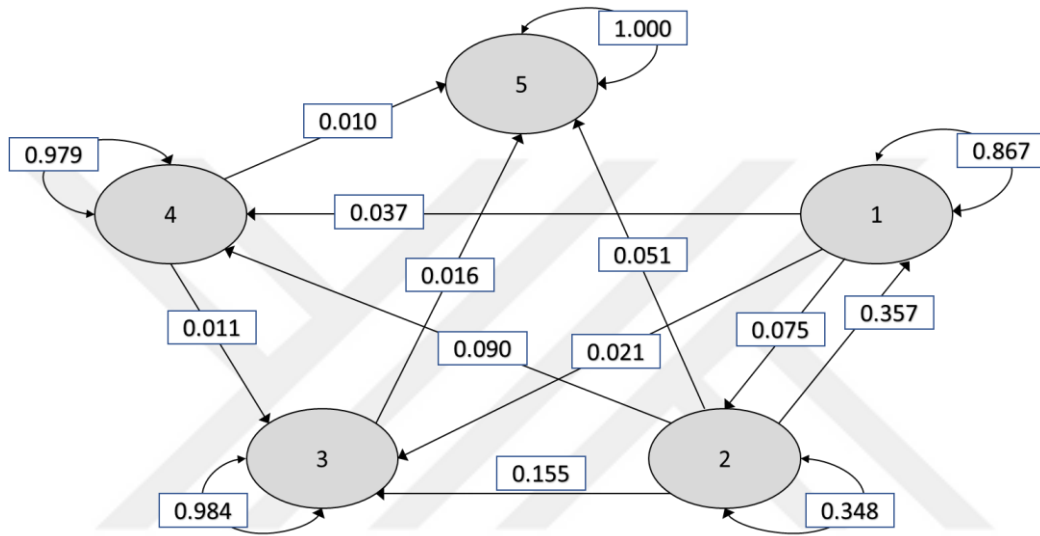
5.3.3. Kritik Hastalılar için Saklı Markov Modeli

Bu kesimde, HRS verisi kullanılarak, kritik hastalık sigortalarında hasara neden olan ölüm, kanser, kalp hastalığı ve felç durumlarının modellenmesi için önerdiğimiz Saklı Markov modeli sunulmuştur. Önerilen model, yaş ve cinsiyetin yanında, tansiyon, diyabet, akciğer ve psikoloji açıklayıcı değişkenlerini ve 5 adet saklı durum içermektedir.

Öncelikle belirlenen saklı durum sayısına göre saklı durumlar arası ortalama geçiş olasılıkları aşağıdaki Çizelge 5.33’de ve modelin görseli Şekil 5.3’te verilmiştir.

Çizelge 5.33. Saklı durumlar arası ortalama geçiş olasılıkları

	Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
Saklı 1	0,867	0,075	0,021	0,037	0,000
Saklı 2	0,357	0,348	0,155	0,090	0,051
Saklı 3	0,000	0,000	0,984	0,000	0,016
Saklı 4	0,000	0,000	0,011	0,979	0,010
Saklı 5	0,000	0,000	0,000	0,000	1,000



Şekil 5.4. Saklı durumlar arası ortalama geçiş olasılıkları

Çizelge 5.33'e göre saklı durum 5'ten diğer saklı durumlara geçiş olmadığı aynı durumda kalma eğilimi olduğu, ayrıca saklı durum 1'den de saklı durum 5'e geçiş olmadığı anlaşılmaktadır. Saklı durum 2 dışındaki saklı durumlarda da genel anlamda bireylerin daha yüksek olasılıkla bulundukları saklı durumda kalma eğilimde olduğu söylenebilir. En fazla geçişler saklı durum 1 ve saklı durum 2'de olmaktadır. Saklı durum 1'den yapılan geçişlere bakıldığında, %7 olasılıkla geçiş yapılan diğer saklı durumun saklı durum 2, saklı durum 2'den yapılan geçişlere bakıldığında ise %35,7 olasılıkla saklı durum 1'e, %15,5 olasılıkla saklı durum 3'e geçiş yapıldığı, saklı durum 3'ten, saklı durum 4'e geçiş olmadığı ve saklı durum 3'ten saklı durum 5'e geçiş olasılığının %1,6 olduğu görülmektedir. Saklı durum 4'ten ise saklı durum 3 ve saklı durum 5'e geçiş olasılığı vardır. Her iki Saklı duruma geçme olasılığı oldukça yakındır.

Önerilen modelde ayrıca, her geçiş için yaş, cinsiyet, tansiyon, akciğer, diyabet ve psikoloji açıklayıcı değişkenlerinin etkisini gösteren logit parametre tahminleri ve hangi değişkenin hangi saklı durumlar arası geçişte etkisi olduğu incelenmiştir.

Saklı durum 1’den yapılan geçişler için parametre tahminleri aşağıdaki Çizelge 5.34’te verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini, alt kısmında yer alan değerler standart hataları göstermektedir.

Çizelge 5.34. Saklı durum 1’den geçiş olasılıkları için parametre tahminleri

Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
α_{sabit}	-0,3126 ^o	-4,7901	-4,5022	-
$\beta_{cinsiyet}$	-1,041	-0,331	-0,5357	-
$\beta_{yaş}$	-0,0405 ^o	0,0072 ^o	0,0200	-
$\beta_{tansiyon}$	-0,3955 ^o	0,6505	0,0195 ^o	-
$\beta_{akciğer}$	-1,4785	1,1426	0,8403	-
$\beta_{diyabet}$	0,0562 ^o	0,3236	0,3603	-
$\beta_{psikoloji}$	-1,7519	0,5196	0,5407	-
α_{sabit}	1,5341	0,6263	0,4178	-
$\beta_{cinsiyet}$	0,3615	0,1501	0,1215	-
$\beta_{yaş}$	0,0232	0,009	0,0067	-
$\beta_{tansiyon}$	0,4586	0,2361	0,1157	-
$\beta_{akciğer}$	0,7174	0,1635	0,1146	-
$\beta_{diyabet}$	0,3836	0,1616	0,1139	-
$\beta_{psikoloji}$	0,5236	0,151	0,112	-

^o: Standart hataya göre anlamlı olamayan parametre tahminleri

Öncelikle parametrelerin anlamlı olup olmadığına bakıldığında, saklı durum 1’den saklı durum 2’ye geçişte cinsiyet, akciğer ve psikoloji değişkeninin anlamlı olduğu görülmektedir. Ayrıca kadınların saklı durum 1’den diğer saklı durumlara geçme olasılığı erkeklerden daha azdır. Tansiyon hastalığının varlığı saklı durum 1’den saklı durum 3’e geçiş olasılığını arttırmaktadır. Saklı durum 3’e ve 4’e akciğer hastalığının, diyabet hastalığının veya psikolojik rahatsızlıkların varlığında geçiş olasılığı artmaktadır.

Saklı durum 2’den yapılan geçişler için parametre tahminleri Çizelge 5.35’de verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini alt kısmında yer alan değerler standart hataları göstermektedir.

Çizelge 5.35. Saklı durum 2’den geçiş olasılıkları için parametre tahminleri

Saklı 2	Saklı 1	Saklı 3	Saklı 4	Saklı 5
α_{sabit}	-1,0697 ^o	-8,3595	-8,7701	-15,3858
$\beta_{cinsiyet}$	1,2593	-0,4687 ^o	0,5499 ^o	-0,833 ^o
$\beta_{yaş}$	-0,0137	0,0709	0,0678	0,1295
$\beta_{tansiyon}$	5,1487	6,4411	6,395	7,3959
$\beta_{akciger}$	2,4205	3,7881	3,4123	5,1809
$\beta_{diyabet}$	0,5711 ^o	2,2336	1,807 ^o	2,6304
$\beta_{psikoloji}$	-0,6995 ^o	1,1323	0,1054 ^o	2,3441
α_{sabit}	1,3821	1,3492	1,2997	1,8324
$\beta_{cinsiyet}$	0,4601	0,4152	0,4843	0,4652
$\beta_{yaş}$	0,0050	0,0195	0,0225	0,0271
$\beta_{tansiyon}$	1,8381	1,2756	1,2086	1,5329
$\beta_{akciger}$	1,028	1,002	1,0996	1,0615
$\beta_{diyabet}$	0,8069	0,8464	0,9259	0,8223
$\beta_{psikoloji}$	0,5189	0,5107	0,6858	0,5927

^o: Standart hataya göre anlamlı olamayan parametre tahminleri

Parametrelerin anlamlı olup olmadığına bakıldığında, saklı durum 1 ve saklı durum 4’e yapılan geçişlerdeki diyabet ve psikoloji açıklayıcı değişkeni, saklı durum 3, saklı durum 4 ve saklı durum 5’e geçişlerdeki cinsiyet açıklayıcı değişkeni haricinde tüm değişkenler anlamlıdır. Diyabet hastalığının varlığının saklı durum 2’den saklı durum 3 ve saklı durum 5’e geçişte anlamlı bir etkisi vardır ve hastalığın varlığı geçiş olasılığını arttırmaktadır. Saklı durum 2’den diğer saklı durumlara geçişte cinsiyetin etkisine bakıldığında, kadınlarda saklı durum 1’e geçiş olasılığının daha yüksek olduğu görülmektedir. Yaş arttıkça saklı durum 1’e geçiş olasılığının azaldığı, saklı durum 3, saklı durum 4 ve saklı durum 5’e geçişlerde olasılığın arttığı görülmektedir.

Saklı durum 3'ten yapılan geçişler için parametre tahminleri aşağıdaki Çizelge 5.36'da verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini, alt kısmında yer alan değerler standart hataları göstermektedir. Saklı durum 3'ten saklı durum 1 ve saklı durum 2'ye geçiş mümkün olmadığı için sadece saklı durum 4 ve saklı durum 5 için parametre tahminleri verilmiştir.

Çizelge 5.36. Saklı durum 3'ten geçiş olasılıkları için parametre tahminleri

Saklı 3	Saklı 1	Saklı 2	Saklı 4	Saklı 5
α_{sabit}	-	-	-	-1,3685
$\beta_{cinsiyet}$	-	-	-	-0,1988 ^o
$\beta_{yaş}$	-	-	-	-0,0464
$\beta_{tansiyon}$	-	-	-	0,2492 ^o
$\beta_{akciger}$	-	-	-	0,14 ^o
$\beta_{diyabet}$	-	-	-	0,4683
$\beta_{psikoloji}$	-	-	-	0,4194
α_{sabit}	-	-	-	0,985
$\beta_{cinsiyet}$	-	-	-	0,1939
$\beta_{yaş}$	-	-	-	0,0142
$\beta_{tansiyon}$	-	-	-	0,258
$\beta_{akciger}$	-	-	-	0,2277
$\beta_{diyabet}$	-	-	-	0,1963
$\beta_{psikoloji}$	-	-	-	0,2094

^o: Standart hataya göre anlamlı olamayan parametre tahminleri

Saklı durum 3'ten saklı durum 5'e yapılan geçişlerde yaş, diyabet hastalığının varlığı, psikolojik rahatsızlıkların varlığı anlamlı bulunmuştur. Diyabet hastalığının ve psikolojik rahatsızlıkların varlığının ise geçiş yapma olasılığını arttırdığı gözlemlenmektedir.

Saklı durum 4'ten yapılan geçişler için parametre tahminleri aşağıdaki Çizelge 5.37'de verilmiştir. Çizelgenin üst kısmında yer alan değerler parametre tahminlerini, alt kısmında yer alan değerler standart hataları göstermektedir. Saklı durum 4'ten yalnızca

Saklı durum 5'e geçiş yapılabildiği için sadece saklı durum 5 için parametre tahminleri verilmiştir.

Çizelge 5.37. Saklı durum 4'ten geçiş olasılıkları için parametre tahminleri

Saklı 4	Saklı 1	Saklı 2	Saklı 3	Saklı 5
α_{sabit}	-	-	-5,0582	-3,4394
$\beta_{cinsiyet}$	-	-	-0,1167 °	0,0542 °
$\beta_{yaş}$	-	-	0,0053 °	-0,0308 °
$\beta_{tansiyon}$	-	-	0,6953	0,9081
$\beta_{akciger}$	-	-	0,099 °	0,1368 °
$\beta_{diyabet}$	-	-	-0,1921 °	0,6999
$\beta_{psikoloji}$	-	-	-0,383 °	0,3935 °
α_{sabit}	-	-	1,1743	1,1802
$\beta_{cinsiyet}$	-	-	0,2006	0,2177
$\beta_{yaş}$	-	-	0,0164	0,0167
$\beta_{tansiyon}$	-	-	0,2291	0,2803
$\beta_{akciger}$	-	-	0,2655	0,2706
$\beta_{diyabet}$	-	-	0,2337	0,2202
$\beta_{psikoloji}$	-	-	0,2712	0,2336

°: Standart hataya göre anlamlı olamayan parametre tahminleri

Çizelge 5.37'de verilen değerlere göre, saklı durum 4'ten 3'e ve saklı durum 5'e geçişte ise tansiyon hastalığının varlığı değişkeni dışındaki değişkenlerin anlamlı bir etkisi bulunamamıştır. Tansiyon hastalığının varlığı ise geçiş olasılığını arttırmaktadır. Ayrıca saklı durum 5'e geçişte diyabet hastalığının anlamlı bir etkisi bulunmuştur. Bu hastalığın varlığı da geçiş olasılığını arttırmaktadır.

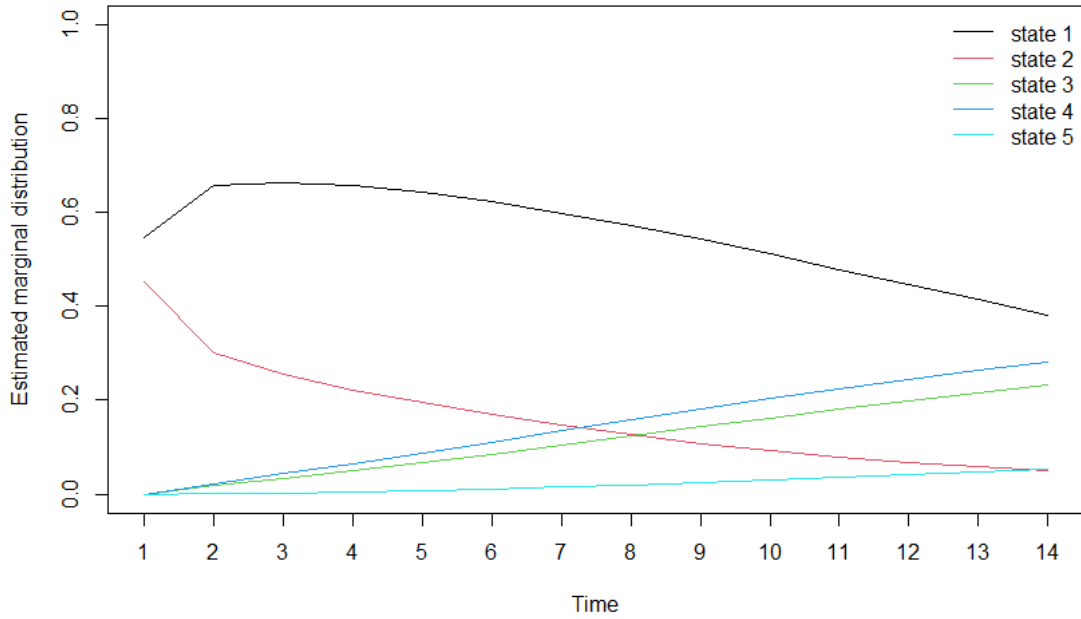
Son olarak Saklı Markov modelinin bir çıktısı olan her saklı durum altında gözlemlenen kritik hastalıklar ve ölüm için koşullu olasılıklar matrisi (emisyon olasılıkları matrisi) Çizelge 5.38'de verilmiştir.

Çizelge 5.38. Saklı durumlar için koşullu olasılık matrisi

	Saklı 1	Saklı 2	Saklı 3	Saklı 4	Saklı 5
Ölüm	0,000	0,000	0,2103	0,5044	0,3141
Kanser	0,000	0,000	0,1466	0,4580	0,1946
Kalp H.	0,000	0,000	1,0000	0,0000	1,0000
Felç	0,000	0,000	0,0000	0,1817	1,0000

Çizelge 5.38’de verilen olasılıklar değerlendirildiğinde, saklı durum 1 ve saklı durum 2’de yer alan bireylerin sağlıklı olduğu, saklı durum 3’te yer alan bireylerin kalp hastası, saklı durum 5’te yer alan bireylerin de kalp hastası ve felç hastası olduğu, saklı durum 4’te yer alan bireylerin %45,8 olasılıkla kanser hastası olabileceği görülmektedir. Bu verilen olasılıklar modelin diğer çıktıları ile birlikte yorumlandığında, saklı durum 5’e gelen bireylerde en az 2 kritik durumun (felç ve kalp hastalığı) olduğu ve bu saklı durumdan diğer durumlara geçiş olmadığı anlaşılmaktadır. Ayrıca saklı durum 3’te en az 1 kritik hastalığın (kalp hastalığı) olduğu ve yalnızca saklı durum 5’e geçiş yapılabildiği, saklı durum 4’e geçildiğinde kanser ve ölüm durumlarının gözlemlendiği ve sağlıklı bireylerin yer aldığı saklı durum 1 ve saklı durum 2’ye geçiş olasılığı olmadığı anlaşılmaktadır. Bu durumda Saklı Markov modelinden amaçlandığı şekilde bir risk sınıflandırması çıkarılabilir. Saklı durum 1’de yer alan bireyler kritik hastalık açısından az riskli, saklı durum 2’de yer alan bireyler riskli, saklı durum 4 çok riskli, saklı durum 3 ve saklı durum 5’e geçildiğinde ise kritik hastalıklar kesinleşmektedir.

Şekil 5.4’te de model tahminine göre, 14 adımda saklı durumlarda bulunma olasılığının marginal dağılımları verilmiştir. Buna göre bireyler saklı durum 1 ya da saklı durum 2’den sisteme sağlıklı olarak giriş yapmakta ve zaman ilerledikçe diğer durumlara geçiş olasılığı artmaktadır. En yüksek artış saklı durum 4’te görülmektedir.



Şekil 5.5. Saklı durumda bulunma olasılıklarının marginal dağılımları

5.4. Modellerin Karşılaştırılması

Bu kısımda GLM ile SMM modellerinin kullanılan verinin dağılımına uyumu karşılaştırılmıştır. Bu karşılaştırmada GLM modelinde anlamlı bulunan açıklayıcı değişkenler üzerinden karşılaştırma yapılmıştır. GLM modeli verideki her yaş ve cinsiyet için kritik hastalığa sahip bireylerin toplam sayısı çeşitli açıklayıcı değişkenlerden faydalanarak modellerken, SMM açıklayıcı değişkenleri kullanarak bireylerin 14 gözlem boyunca kritik hastalığa yakalanıp yakalanmadıklarını modellemeyi amaçlar. Bu anlamda iki tekniğin sonuçları da birbirinden farklıdır. Ek olarak GLM her kritik hastalık için ayrı ayrı oluşturulurken, Saklı Markov modelinde tüm kritik hastalıklar birlikte modellenir. İki model karşılaştırılırken Saklı Markov model çıktısı, GLM model çıktısı ile uyumlu hale gelecek şekilde gruplanmıştır. Buna göre bireylerin tahmin edilen 14 farklı gözleminden faydalanılarak, her yaş ve cinsiyet için kritik hastalık gözlem frekansı hesaplanmıştır.

Tezde sunulan modellerde anlamlı etkisi olduğu tespit edilen veya model performansını iyileştiren farklı açıklayıcı değişkenler kullanıldığından, iki modelin karşılaştırması yaş

ve cinsiyet açıklayıcı değişkeni içeren temel modeller üzerinden yapılmıştır. İki modelin yapısal farkı nedeniyle doğrudan AIC kriterlerini karşılaştırmak mümkün olmamaktadır. Bu karşılaştırmanın yapılabilmesi için Yang [52]'de önerilen hataların kareleri üzerinden hesaplanan AIC kriterine göre hesaplama yapılmıştır.

n sayıda veri noktası için hataların normal dağıldığı ve bağımsız olduğu varsayımı altında,

$$AIC = 2k + n \ln \left(\frac{RSS}{n} \right)$$

RSS hataların kareleri toplamını göstermek üzere,

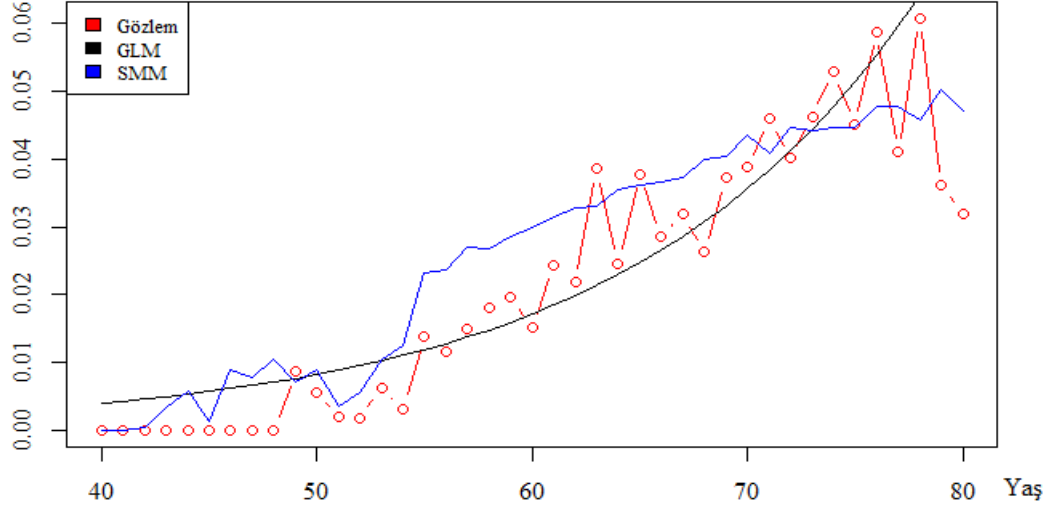
$$RSS = \sum_{i=1}^n [y_i - \hat{y}_i]$$

y_i gerçek değerleri, $\hat{y}_i$ model tarafından tahmin edilen değerleri gösterir. Genel prensip burada da geçerli olacaktır ve AIC küçüldükçe modelin daha başarılı olduğu anlamı çıkarılabilir.

Buna göre, tez kapsamında kritik hastalık sigortaları açısından hasara neden olabilecek hastalıklar ve ölüm durumu için iki modelin görsel karşılaştırılması ve AIC değerleri verilmiştir.

5.4.1. Ölüm için Modellerin Karşılaştırılması

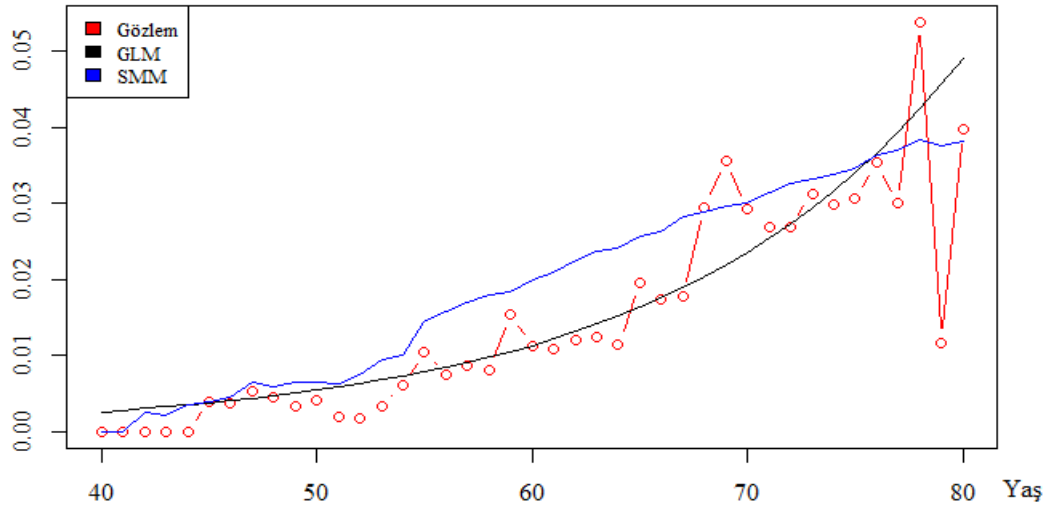
Bu kısımda kritik hastalık sigortası kapsamında değerlendirilen ölüm durumu için sonuçlar karşılaştırılmıştır. Şekil 5.5 'de erkek için yaşa göre veriden, GLM modelinden ve SMM modelinden tahmin edilen ölüm olasılıkları verilmiştir.



Şekil 5.5. Erkekler için yaşa göre ölüm olasılıkları

Model performanslarına bakıldığında, şekil üzerinde 55-63 yaşları arasında GLM modelinin veri ile daha uyumlu olduğu görülmekte, ancak 40-55 arası ve 63-80 yaş grupları için SMM modelinin veri ile daha uyumlu olduğu görülmektedir. AIC kriteri, GLM modeli için -374,37, SMM modeli için -393,87 olarak hesaplanmıştır. Bu kriterlere göre SMM modeli daha başarılıdır.

Şekil 5.6 'da kadın için veriden, GLM modelinden ve SMM modelinden tahmin edilen ölüm olasılıkları verilmiştir.

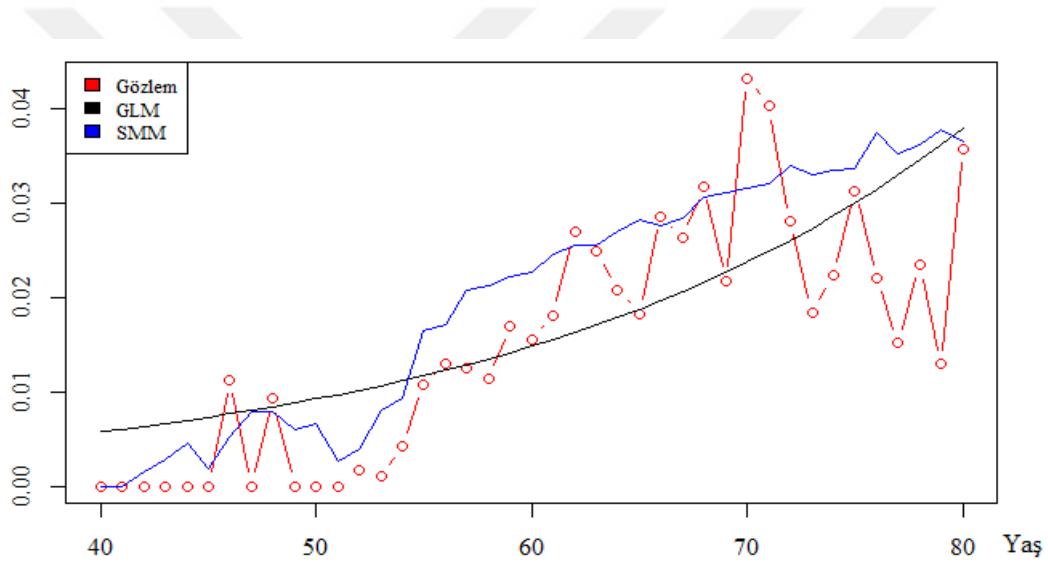


Şekil 5.6. Kadınlar için yaşa göre ölüm olasılıkları

Model performanslarına bakıldığında, şekil üzerinde 56-68 yaşları arasında GLM modelinin veri ile daha uyumlu olduğu görülmekte, ancak 40-56 arası ve 68-80 yaş grupları için SMM modelinin veri ile daha uyumlu olduğu görülmektedir. AIC kriteri, GLM modeli için -407,84, SMM modeli için -401,44 olarak hesaplanmıştır. Bu kritere göre GLM modeli daha başarılıdır.

5.4.2. Kanser için Modellerin Karşılaştırılması

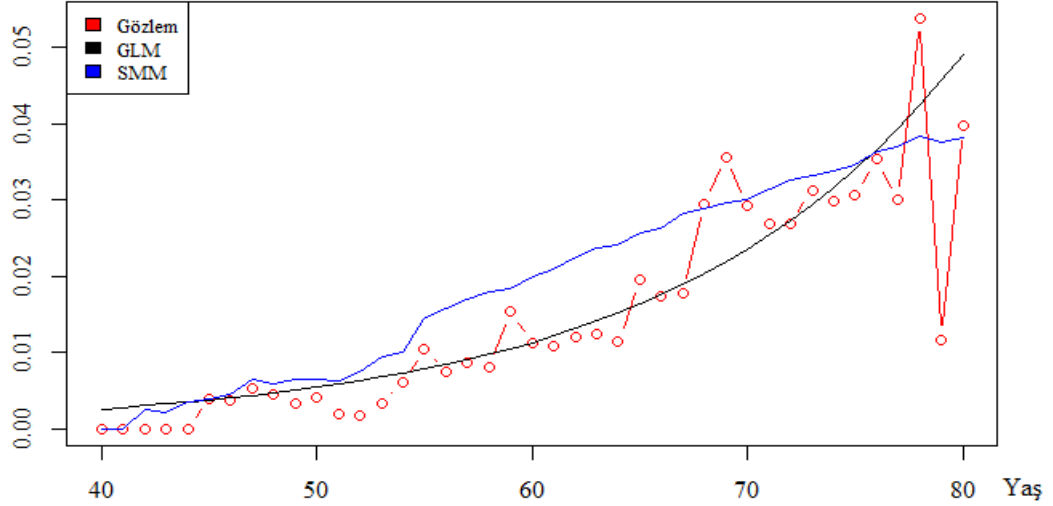
Bu kısımda kritik hastalık sigortası kapsamında değerlendirilen kanser durumu için sonuçlar karşılaştırılmıştır. Şekil 5.7 'de erkek için yaşa göre veriden, GLM modelinden ve SMM modelinden tahmin edilen kanser olasılıkları verilmiştir.



Şekil 5.7. Erkekler için yaşa göre kanser olasılıkları

Model performanslarına bakıldığında, şekil üzerinde SMM modelinin veri ile daha uyumlu olduğu görülmekte, ancak 70-80 arası yaş grupları için veride yüksek değişkenlik görüldüğünden her iki modelin de sağ kuyrukta veriye göre yüksek tahmin yaptığı gözlemlenmektedir. AIC kriteri, GLM modeli için -389,94, SMM modeli için -392,33 olarak hesaplanmıştır. Bu kritere göre SMM modeli daha başarılıdır.

Şekil 5.8'de kadınlar için veriden, GLM modelinden ve SMM modelinden tahmin edilen kanser olasılıkları verilmiştir.

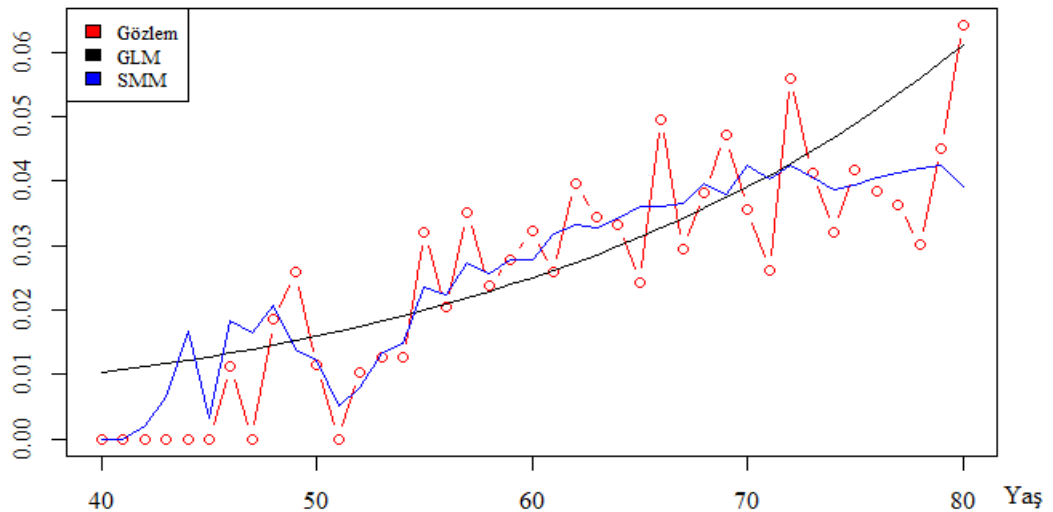


Şekil 5.8. Kadınlar için yaşa göre kanser olasılıkları

Model performanslarına bakıldığında, şekil üzerinde erkeklerde olduğu gibi sağ kuyruk değerleri dışında modeller veri ile uyumludur. SMM modelinin 40-68 arası yaş grupları için oldukça başarılı olduğu görülmektedir. AIC kriteri, GLM modeli için -420,72, SMM modeli için -413,56 olarak hesaplanmıştır. Bu kritere göre SMM modeli daha başarılıdır.

5.4.3. Kalp Hastalığı için Modellerin Karşılaştırılması

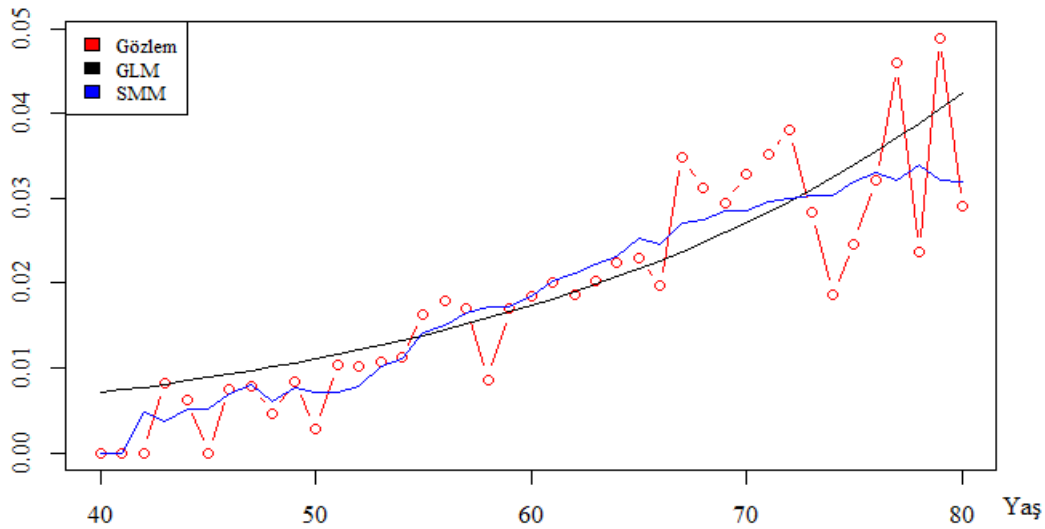
Bu kısımda kritik hastalık sigortası kapsamında değerlendirilen kalp hastalığı durumu için sonuçlar karşılaştırılmıştır. Şekil 5.9 'de erkek için yaşa göre veriden, GLM modelinden ve SMM modelinden tahmin edilen kalp hastalığı olasılıkları verilmiştir.



Şekil 5.9. Erkekler için yaşa göre kalp hastalığı olasılıkları

Model performanslarına bakıldığında, şekil üzerinde SMM modelinin veri ile daha uyumlu olduğu görülmektedir. Özellikle 40-51 ve 72-80 arası yaş grubunda GLM modeli gerçek değerlere göre fazla tahmin etmektedir. AIC kriteri, GLM modeli için -373,08, SMM modeli için -392,97 olarak hesaplanmıştır. Bu kritere göre SMM modeli daha başarılıdır.

Şekil 5.10 'da kadın için veriden, GLM modelinden ve SMM modelinden tahmin edilen kalp hastalığı olasılıkları verilmiştir.

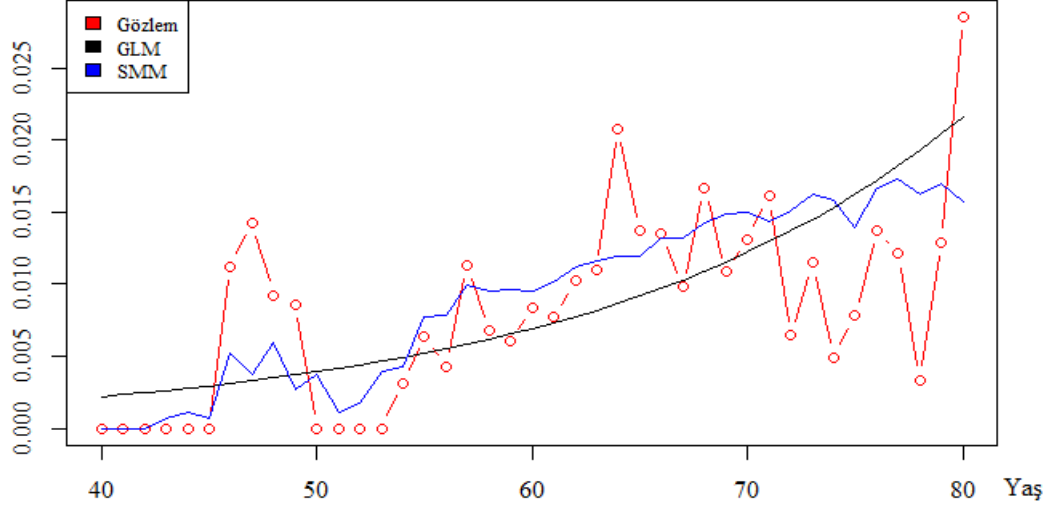


Şekil 5.10. Kadınlar için yaşa göre kalp hastalığı olasılıkları

Model performanslarına bakıldığında, erkelerdeki verilen duruma bezer bir yapı görülmektedir. Şekil üzerinde SMM modelinin veri ile daha uyumlu olduğu görülmektedir. AIC kriteri, GLM modeli için -415,22, SMM modeli için -427,93 olarak hesaplanmıştır. Bu kritere göre SMM modeli daha başarılıdır

5.4.4. Felç için Modellerin Karşılaştırılması

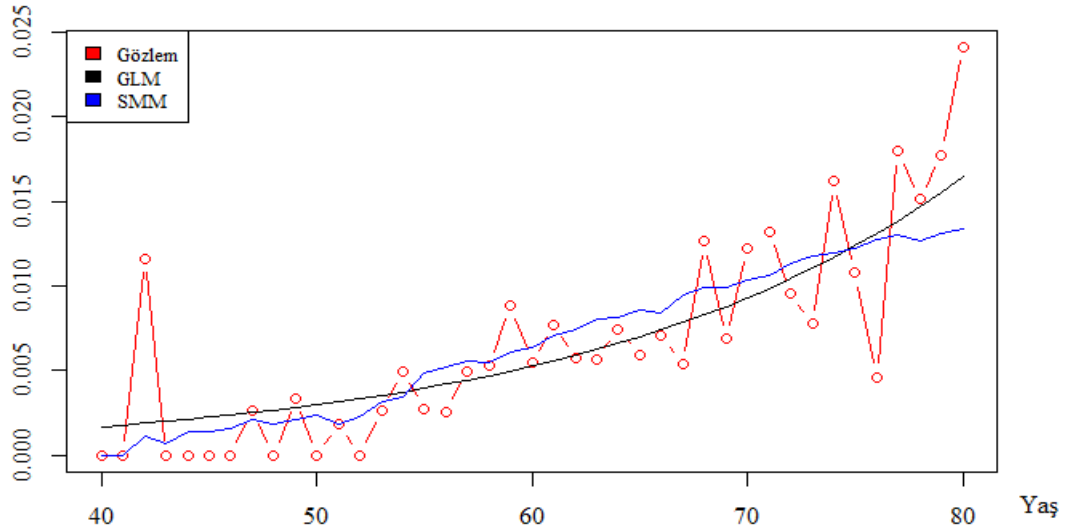
Bu kısımda kritik hastalık sigortası kapsamında değerlendirilen felç hastalığı durumu için sonuçlar karşılaştırılmıştır. Şekil 5.11 'de erkek için yaşa göre veriden, GLM modelinden ve SMM modelinden tahmin edilen kalp hastalığı olasılıkları verilmiştir.



Şekil 5.11. Erkekler için yaşa göre felç olasılıkları

Model performanslarına bakıldığında, şekil üzerinde SMM modelinin veri ile daha uyumlu olduğu görülmekte, ancak 70-80 yaş grupları için verideki düşüşe iki modelinde uyum sağlayamadığı görülmektedir. AIC kriteri, GLM modeli için -425,19, SMM modeli için -434,27 olarak hesaplanmıştır. Bu kritere göre SMM modeli daha başarılıdır.

Şekil 5.12 'da kadın için veriden, GLM modelinden ve SMM modelinden tahmin edilen felç olasılıkları verilmiştir.



Şekil 5.12. Kadınlar için yaşa göre felç olasılıkları

Kadınlarda ise iki modelinde veri ile uyumlu olduğu ve oldukça yakın tahminler yaptığı görülmektedir. AIC kriteri, GLM modeli için -471,56, SMM modeli için -464,47 olarak hesaplanmıştır. Bu kritere göre GLM modeli daha başarılıdır.

5.5. Prim Hesaplamaları

Yazınında her ne kadar Saklı Markov modellerinin risk hesaplamasında kullanılabileceğine dair çalışmalar bulunsun da gerçek anlamda aktüerya uygulamasına yönelik hiçbir çalışma bu tezin yazıldığı tarih itibarıyla bulunmamaktadır. Dolayısıyla SMM kullanılarak bir sigorta ürününün fiyatlandırılması, prim veya rezerv hesabını tartışan hiçbir çalışma da mevcut değildir. Yazındaki çalışmalarda, GLM modeli kullanılarak elde edilen geçiş şiddetlerini Thiele denklemlerinde kullanmak suretiyle prim hesabı yapılmaktadır. GLM modelleri prensipte veriyi yaş bazında gruplar ve parametreleri gruplanan bu veri üzerinden tahmin eder. SMM modellerinin çalışma prensibi ise bundan çok farklıdır. SMM modelleri birey bazında eş değişkenlere göre bireylerin hayat sürecini (life trajectory) kullanır ve bireylerin gözlemlenen Markov zincirinde hangi saklı durumda olduklarını tahmin eder. Dolayısıyla SMM modelinde her birey için eş değişkene bağlı olarak saklı durum tahmini yapılmaktadır. Dolayısıyla modelin Thiele denklemlerinde kullanılabılır bir çıktısı, diğer bir deyişle geçiş şiddetini parametrik olarak tanımlayan bir formülü bulunmamaktadır. SMM modeli bir defa kurulduktan sonra eş değişkenler verilerek modelin hayat sürecini benzetim ile tekrar etmesi istenebilir. Bu çıktıdan yola çıkılarak prim hesabının nasıl yapılacağına yönelik temel Shao ve ark. [25] 'nın çalışmasında mevcuttur. Bu çalışmada karmaşık poliçeler için Thiele denklemlerinin karmaşık durumlar alabileceğini belirtilmiş ve bu durumlarda kurulan GLM modellerinden faydalanılarak benzetim yapılabileceği ve bu benzetimden yola çıkılarak prim hesabı yapılabileceğini önermiştir. Bu öneri SMM çıktılarından prim hesabı yapmanın temelini teşkil etmektedir. Önerilen benzetim temelde şu prensiple kurulmaktadır. İlk olarak veri gruplanır, GLM modelleri kurulur. Ardından söz konusu GLM modellerinden elde edilen geçiş şiddetleri Kolmogorov İleri denklemleri kullanılarak geçiş olasılıklarına dönüştürülür. Bu aşamada başlangıçta kurulan Markov modelinin geçiş olasılıkları elde edilmektedir. Artık herhangi bir yazılım kullanılarak kişilerin eş değişken değerlerine bağlı olarak Markov zincirinde bir sonraki yaş'ta hangi durumda olacakları tahmin edilebilir ve hayat süreçleri oluşturulabilir. Benzetimde

yeterince tekrar yapıldığında sonucun hesaplanan prim değerinin Thiele denklemleri ile hesaplanan prim değerine eşit çıkacağını belirtmektedir.

Simülasyon ile prim hesabı şu şekilde yapılmaktadır. Oluşturulan sentetik bireyler Markov zincirinde eğer hayatta ise o dönem prim öderler. Ancak eğer kritik hastalıklardan biri veya ölüm durumuna geçtilerse tazminat ödemesi yapılır. Buna göre her x yaşında olan N bireyden oluşan portföyün rezervi P prim değişkenine bağlı olarak hesaplanır ve belirlenen dönem sonunda rezerv 0 olacak şekilde formüle edilir. Bu formülde tanımlanan fonksiyonun kökü eşdeğerlik prensibine göre primi verir. Benzetimde ne kadar çok sentetik birey kullanılırsa simülasyon sonucu elde edilen prim hesabının standart hatası birey adedinin karekökü ile doğru orantılı olacak şekilde azalır.

SMM modeli ile benzetim yapılırken R yazılımında LMest paketi içerisinde bulunan drawLMlatent fonksiyonu kullanılmıştır. Bu fonksiyon kullanılırken, her bir bireyin her bir gözlemdaki açıklayıcı değişken değerleri (life trajectory) ve model parametreleri girilir. Bu şekilde SMM modelinden, verilen açıklayıcı değişkenlere ve model parametrelerine bağlı olarak birey seviyesinde gözlemler üretir. Açıklayıcı değişkenler bireyin t zamanında hangi saklı durumda olduğunu belirlerken, saklı durum ise bireyin hangi gözlemlenebilir durumda olduğunu belirler. Bu şekilde oluşturulan 100 farklı benzetim sonucunda farklı durumlara sahip bireyler üretip, bu sentetik bireylerden yukarı açıklanan prensibe göre P primine bağlı bir fonksiyon elde etmektir. Bu fonksiyonun kökü eşdeğerlik prensibine göre primi verecektir.

Tanımlanan modeller için bir sigorta poliçesi varsayılmıştır. Buna göre bu poliçeyi x yaşında bir sigortalı satın alır, hayatta kaldığı süre boyunca her yıl P primi öder ve kritik hastalıklardan birinin görülmesi veya ölmesi durumunda ise $S = 1000$ birim tazminat ödemesi alır. Yıllık sürekli faiz oranı $\delta = 0,04$ olarak varsayılmıştır.

Ayrıca karşılaştırılma yapılabilmesi adına, veride yer alan gerçek bireylerin de x yaşında kritik hastalık sigortası satın alacağı ve hayatta kaldıkları süre boyunca P primini ödeyecekleri, kritik hastalıklardan biri ya da ölüm durumunda $S = 1000$ tazminat

ödemesi alacakları varsayımı yapılmıştır. Bu şekilde yine P'ye bağlı bir fonksiyon elde edilmiş ve bu fonksiyonun kökü hesaplanarak baz prim hesaplaması yapılmıştır.

GLM modelinde yer alan açıklayıcı değişken sayısı kısıtlı olduğundan, karşılaştırmalar sadece cinsiyet ve yaş üzerinde yapılmıştır. Aşağıdaki Çizelge 5.39'de 40, 45 ve 50 yaşındaki erkek ve kadın için hesaplanan 10 yıllık kritik hastalık sigortası için primler verilmiştir.

Çizelge 5.39. 10 yıllık kritik hastalık sigortası için primler

Yaş	Erkek			Kadın		
	Veri	GLM	SMM	Veri	GLM	SMM
40	8,77	28,71	13,84	9,57	19,74	8,86
45	20,27	38,08	24,78	17,01	25,89	17,42
50	38,95	51,11	44,11	31,26	34,22	30,52

Çizelge 5.40'da 40,45 yaşındaki erkek ve kadın için hesaplanan 15 yıllık kritik hastalık sigortası için primler verilmiştir.

Çizelge 5.40. 15 yıllık kritik hastalık sigortası için primler

Yaş	Erkek			Kadın		
	Veri	GLM	SMM	Veri	GLM	SMM
40	12,18	34,65	17,47	13,56	23,27	12,39
45	33,02	47,11	37,63	24,30	31,02	24,60

Çizelge 5.41'de 40 yaşındaki erkek ve kadın için hesaplanan 20 yıllık kritik hastalık sigortası için primler verilmiştir.

Çizelge 5.41. 20 yıllık kritik hastalık sigortası için primler

Yaş	Erkek			Kadın		
	Veri	GLM	SMM	Veri	GLM	SMM
40	21,51	32,24	26,97	31,26	34,23	30,52

6. SONUÇLAR VE ÖNERİLER

Bu bölümde, tez çalışmasında elde edilen önemli sonuçlar, çalışmanın kısıtları, yazına olan katkıları ve ileride yapılabilecek yeni çalışmalara ilişkin öneriler sunulmuştur.

Bu tezin temel amacı, kritik hastalık sigortalarına ilişkin hesaplamalarda kullanılabilecek ve bireylerin sağlık durumlarına ilişkin verilerdeki heterojen yapıyı ortaya koyan yeni bir model geliştirmektedir. Geleneksel yaklaşımda, bireylerin sağlık durumlarına ilişkin verilerde yer alan heterojen gruplar öncesinde belirlenmektedir. Örneğin, veri cinsiyet ve/veya yaş gruplarına bölünerek, ayrı ayrı modellenmektedir. Bu yaklaşım da verinin parçalanmasına ve etkisi değerlendirmek istenen yeni bir değişken eklendiğinde, verinin yeni kombinasyonlar oluşturacak şekilde yeniden parçalanmasına neden olmaktadır. Ayrıca, bireylerin sağlık durumları ve hastalıkları ile ilişkili olabilecek kilo, sigara kullanımı gibi alışkanlıklar, yapılan egzersizler gibi açıklayıcı değişkenlerin geleneksel modellerde sabit oldukları varsayılmaktadır.

Bu tez çalışmasında, kritik hastalık sigortalarında hasara neden olabilecek hastalıkların görülme olasılıklarının tahmini ve bu tür sigortaların fiyatlandırılmasında kullanılabilecek Saklı Markov Modelleri yöntemi önerilmiştir. Bu yöntem bütünsel bir yaklaşım sunmakla birlikte, öncelikle veride yer alan heterojen grupların belirlenmesini sağlamakta, sonrasında ise her farklı grup için farklı olasılık tahminleri üretmektedir. Ayrıca, herhangi bir açıklayıcı değişkenin zaman içerisinde değişen değerleri ile birlikte modele eklenmesine imkân sağlar.

Tez çalışmasında, öncelikle yazında kritik hastalık sigortalarının modellenmesi ve fiyatlandırılmasına yönelik yapılan çalışmalarda sıklıkla kullanılan, Markov modeli uygulanmış ve tezde kullanılan HRS verisi üzerinde parametre tahminleri yapılmıştır. Bu sayede, tez çalışmasında kullandığımız yöntemin tahmin performansını karşılaştırabileceğimiz bir temel oluşturulmuştur. Diğer bir deyişle, aynı veri üzerinde iki farklı yöntemin sonuçlarını karşılaştırarak, tez çalışmasında önerdiğimiz modelin, yazında kullanılan mevcut modellere göre performansının da ortaya konması amaçlanmıştır. Her

iki uygulamada da kritik hastalık sigortasında hasara neden olan, ölüm, kanser hastalığı teşhisi, kalp hastalığı teşhisi ve felç hastalığı teşhisi modellenmiştir.

Önerilen modelin sonuçları değerlendirildiğinde, veride 5 farklı saklı durum olduğu saptanmıştır. Buna göre, saklı durum 1 ve saklı durum 2’de yer alan bireylerin sağlıklı olduğu görülmüştür. Geri kalan üç saklı durum ise ölüm ve kritik hastalıkların yüksek olasılıkla yaşandığı durumlardır. Bu üç kritik durumdan sağlıklı olan durumlara beklendiği gibi bir geçiş olasılığı bulunmamaktadır. Öte yandan sağlıklı durumlar ve kritik durumların kendi içlerinde geçiş olasılıkları hesaplanmıştır. Sigorta hesaplamaları açısından değerlendirildiğinde, asıl önemli olan sağlıklı durumlardan kritik durumlara geçişin modellenmesidir. Bu bağlamda, saklı durum 1’de yer alan bireylerin sağlıklı devam etme olasılığının daha yüksek olduğu, saklı durum 2’de yer alan bireylerin ise saklı durum 3 ve saklı durum 4’e geçiş olasılıklarının daha yüksek olduğu söylenebilir. Ayrıca saklı durum 5’e geçiş yapan bireylerin diğer durumlara geçiş yapamadığı anlaşılmaktadır. Saklı Durum 3’te yer alan bireylerin kalp hastası olduğu, saklı durum 4’te yer alan bireylerin yüksek olasılıkla kanser hastası olduğu bilgisi de değerlendirildiğinde, modelden elde edilen saklı durumların, sağlık durumlarına göre risk grupları oluşturduğu söylenebilir. Bu şekilde sistem her bireyin öncelikle hangi saklı durumda olduğunu belirlemekle birlikte hangi risk grubunda olduğuna ve hangi hastalıkların daha yüksek olasılıkla görülebileceğine ilişkin bir öngörü sunmaktadır. Bu bilgi de kuşkusuz sigorta ürünlerinin fiyatlandırılmasında oldukça önemlidir.

Önerilen modelin performansını değerlendirmek amacıyla, ölüm, kanser, kalp ve felç hastalıkları için oluşturulan Markov modelinin parametreleri GLM kullanılarak tahmin edilmiştir. Bu tahminler, Saklı Markov Modelinden elde edilen tahminler ile karşılaştırılmıştır. Bu karşılaştırma hataların kareleri üzerinden hesaplanan AIC kriterine göre yapılmıştır. Sekiz farklı durum için yapılan tahminler sonucunda, sadece iki durum dışındaki tüm durumlarda tezde önerilen Saklı Markov Modelinin performansının geleneksel Markov yaklaşımına oranla daha başarılı olduğu görülmüştür.

Tez çalışmasında kullanılan verideki katılımcıların yaşlarının yüksek olması, modelde yer alan hastalıkların görülme sıklığının artmasında neden olmaktadır. Modelin

uygulamada geleneksel yöntemlere oranla risk değerlendirilmesinde daha başarılı olacağı öngörülmekle birlikte, ülkenin yaş dağılımına uyumlu bir veri ile performansının değerlendirilmesi gerektiği düşünülmektedir. Her ne kadar Saklı Markov Modelleri yapısı gereği risk grupları oluştursa da açıklayıcı değişkenler kullanılarak yapılacak bir kümeleme analizi sonrası her küme için farklı bir model geliştirilmesi daha başarılı sonuçlar elde edilmesini sağlayabilir.

Kritik hastalık sigortası gibi çok durumlu modeller ile temsil edilebilen sigorta poliçesinde sigortalının bir durumda bulunması veya bir durumdan diğerine geçtiğinde oluşan nakit akışları prim ve rezerv hesaplamaları için oldukça önemlidir. Kesikli ve sürekli Markov Modellerinde bu nakit akışlarının aktüeryal bugünkü değerinin hesaplanması için formüller geliştirilmiştir. Benzer şekilde Saklı Markov Modeli kullanılarak prim ve rezerv hesaplamaları yapılabilmesi için algoritmalar geliştirilebilir. Bu algoritmalar sayesinde karmaşık poliçeler için bile prim ve rezerv hesaplamaları yapılabilir.

Tez çalışmasında önerilen modelde yer alan kritik hastalık sayısı artırılarak yeni modeller geliştirilebilir. Ayrıca kanser hastalığı gibi cinsiyet spesifik farklı türleri olan hastalıklar ayrıca değerlendirilebilir. Önerdiğimiz yöntem, yalnızca kritik hastalık sigortaları için değil, diğer sağlık sigorta poliçelerine ilişkin pek çok uygulamada kullanılabilir.

7. KAYNAKLAR

- [1] A. Dinani, D. Grimshaw, N. Robjohns, S. Somerville, A. Spry, J. Staffurth, A Critical Review Report of the Critical Illness Healthcare Study Group, the Staple Inn Actuarial Society, London, **2000**.
- [2] Türkiye İş Bankası A.Ş., Kritik Hastalık Sigortası | Türkiye İş Bankası A.Ş., <https://www.isbank.com.tr/TR/bireysel/sigorta-urunleri/kritik-hastalik-sigortasi/Sayfalar/kritik-hastalik-sigortasi.aspx> (Erişim tarihi: **17 Temmuz 2021**)
- [3] N. Gatzert, A. Maegebier, Critical Illness Insurances: Challenges and Opportunities for Insurers, Risk Management and Insurance Review, 18 (**2015**) 255–272.
- [4] A.C. Dash, D.L. Grimshaw, Dread Disease Cover—An Actuarial Perspective, Journal of Staple Inn Actuarial Society, 33 (**1993**) 149–193.
- [5] A.S. Macdonald, H.R. Waters, C.T. Wekwete, The Genetics of Breast and Ovarian Cancer I: A Model of Family History, Scandinavian Actuarial Journal, 1 (**2003**) 1–27.
- [6] A.S. Macdonald, H.R. Waters, C.T. Wekwete, A Model for Coronary Heart Disease and Stroke with Applications to Critical Illness Insurance Underwriting II: The Model, North American Actuarial Journal, 9 (**2005**) 41–56.
- [7] C. Gutiérrez, A.S. Macdonald, Adult Polycystic Kidney Disease and Critical Illness Insurance, North American Actuarial Journal, 7 (**2003**) 93–115.
- [8] A.S. Macdonald, H.R. Waters, C.T. Wekwete, The Genetics of Breast and Ovarian Cancer II: A Model of Family History, Scandinavian Actuarial Journal, 1 (**2003**) 28–50.
- [9] L. Lu, A. Macdonald, H. Waters, Sampling Distributions of Critical Illness Insurance Premium Rates: Breast and Ovarian Cancer, ASTIN Bulletin, 38 (**2008**) 527–542.
- [10] F. Baione, S. Levantesi, A health insurance pricing model based on prevalence rates: Application to critical illness insurance, Insurance: Mathematics and Economics, 58 (**2014**) 174–184.
- [11] A.S. Macdonald, S.Y. Yang, Highly Penetrant Genetic Disorders: An Atlas for Critical Illness Insurance, British Actuarial Journal, 9 (**2003**) 875–901.

- [12] C. Wekwete, Genetics and critical illness insurance underwriting: models for breast cancer and ovarian cancer and for coronary heart disease and stroke, Doktora Tezi, Heriot-Watt University Department of Actuarial Mathematics & Statistics, Edinburgh, **2002**.
- [13] E. Ozkok, A Stochastic Model For Critical Illness Insurance, Doktora Tezi, Heriot-Watt University Department of Actuarial Mathematics & Statistics, School of Mathematical and Computer Sciences, Edinburgh, **2011**.
- [14] D.C.M. Dickson, M. Hardy, M.R. Hardy, H.R. Waters, Actuarial Mathematics for Life Contingent Risks, Cambridge University Press, **2013**.
- [15] A.S. Macdonald, S.J. Richards, I.D. Currie, Modelling Mortality with Actuarial Applications, Cambridge University Press, **2018**.
- [16] P. De Jong, G.Z. Heller, Generalized Linear Models for Insurance Data, Cambridge University Press, **2008**.
- [17] D. Bugliari, J. Carroll, O. Hayden, J. Hayes, M. Hurd, A. Karabatakis, R. Main, J. Marks, C. McCullough, E. Meijer, M. Moldoff, P. Pantoja, S. Rohwedder, P. St. Clair, RAND HRS Longitudinal File 2018 (V1) Documentation, RAND Social Economic Well-Being, **2021**.
- [18] F.T. Juster, R. Suzman, An Overview of the Health and Retirement Study, Journal of Human Resources, 30 (**1995**) 7.
- [19] C. Robinson, R. Clark, Retiree Health Insurance and Disengagement from a Career Job, Journal of Labor Research, 31 (**2010**) 247–262.
- [20] K. Kapur, J. Rogowski, The role of health insurance in joint retirement among married couples, Industrial and Labor Relations Review, 60 (**2007**) 397–407.
- [21] J. Kim, V. Richardson, The impact of socioeconomic inequalities and lack of health insurance on physical functioning among middle-aged and older adults in the United States, Health and Social Care in the Community, 20 (**2012**) 42–51.
- [22] X. Xu, D.A. Patel, A. Vahratian, S.B. Ransom, Insurance coverage and health care use among near-elderly women, Women's Health Issues, 16 (**2006**) 139–148.
- [23] J. Jung, Estimating Markov transition probabilities between health states in the HRS dataset, Indiana University, 1 (**2006**) 1–42.
- [24] J.H. Fong, A.W. Shao, M. Sherris, Multistate Actuarial Models of Functional Disability, North American Actuarial Journal, 19 (**2015**) 41–59.
- [25] A.W. Shao, M. Sherris, J.H. Fong, Product pricing and solvency capital requirements for long-term care insurance, Scandinavian Actuarial Journal, 2 (**2017**) 175–208.
- [26] RAND HRS Longitudinal File 2018 | Health and Retirement Study, <https://hrsdata.isr.umich.edu/data-products/rand-hrs-longitudinal-file-2018> (Erişim tarihi: **2 Şubat 2021**)

- [27] R.B. Wallace, A.R. Herzog, Overview of the Health Measures in the Health and Retirement Study, *Journal of Human Resources*, 30 **(1995)** 84.
- [28] A.S. Macdonald, H.R. Waters, C.T. Wekwete, A model for coronary heart disease and stroke with applications to critical illness insurance underwriting {I}: the model, *North American Actuarial Journal*, 9 **(2005)** 13–40.
- [29] E. Ozkok, G. Streftaris, H.R. Waters, A.D. Wilkie, Bayesian modelling of the time delay between diagnosis and settlement for Critical Illness Insurance using a Burr generalised-linear-type model, *Insurance: Mathematics and Economics*, 50 **(2012)** 266–279.
- [30] L.R. Rabiner, B.H. Juang, An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, 3 **(1986)** 4–16.
- [31] L.E. Baum, T. Petrie, Statistical Inference for Probabilistic Functions of Finite State Markov Chains, *The Annals of Mathematical Statistics*, 37 **(1966)** 1554–1563.
- [32] W. Zucchini, I.L. Macdonald, R. Langrock, Hidden Markov models for time series: An introduction using R, second edition, Chapman & Hall/CRC, Boca Raton, **2017**.
- [33] B. Djehiche, B. Löfdahl, A Hidden Markov Approach to Disability Insurance, *North American Actuarial Journal*, 22 **(2018)** 119–136.
- [34] Z. Oflaz, Coupled Hidden Markov Model with Bivariate Discrete Copula to Study Comorbidity of Chronic Diseases, *Doktora Tezi, Orta Doğu Teknik Üniversitesi, Ankara*, **2022**.
- [35] R.J. Elliott, T.K. Siu, An HMM approach for optimal investment of an insurer, *International Journal of Robust and Nonlinear Control*. 22 **(2012)** 778–807.
- [36] S. Ghassempour, F. Girosi, A. Maeder, Clustering multivariate time series using Hidden Markov Models, *International Journal of Environmental Research and Public Health*, 11 **(2014)** 2741–2763.
- [37] S.G. Pour, Clustering longitudinal health data using hidden Markov models, *Doktora Tezi, University of Western Sydney, Sydney*, **2014**.
- [38] A.C. Tsoi, S. Zhang, M. Hagenbuchner, Hierarchical Hidden Markov Models: An Application to Health Insurance Data. *Data Mining*, G.J. Williams, S.J. Simoff (Eds), Springer, Berlin, 244–259, **2006**.
- [39] F. Bartolucci, S. Bacci, F. Pennoni, Longitudinal analysis of self-reported health status by mixture latent auto-regressive models, *Journal of Royal Statistical Society Series C Applied Statistics* 63 **(2014)** 267–288.
- [40] L.R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*. 77 **(1989)** 257–286.

- [41] Y. Ephraim, N. Merhav, Hidden Markov processes, *IEEE Transactions on Information Theory*, 48 **(2002)** 1518–1569.
- [42] F. Bartolucci, A. Farcomeni, F. Pennoni, Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates, *TEST*, 23 **(2014)** 433–465.
- [43] L.E. Baum, J.A. Eagon, An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology, *Bulletin of the American Mathematical Society* 73 **(1967)** 360–363.
- [44] L.E. Baum, G.R. Sell, Growth transformations for functions on manifolds, *Pacific Journal of Mathematics* 27 **(1968)** 211–227.
- [45] A.J. Viterbi, Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm, *IEEE Transactions on Information Theory*, 13 **(1967)** 260–269.
- [46] G.D. Forney, The Viterbi Algorithm, *Proceedings of the IEEE*. 61 **(1973)** 268–278.
- [47] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from Incomplete Data via the *EM* Algorithm, *Journal of Royal Statistical Society Series B*, 39 **(1977)** 1–22.
- [48] S.E. Levinson, L.R. Rabiner, M.M. Sondhi, An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition, *Bell System Technical Journal* 62 **(1983)** 1035–1074.
- [49] E. Ozkok, G. Streftaris, H.R. Waters, A.D. Wilkie, Modelling critical illness claim diagnosis rates I: methodology, *Scandinavian Actuarial Journal*, 5 **(2014)** 439–457.
- [50] A.S. Macdonald, H.R. Waters, C.T. Wekwete, A model for coronary heart disease and stroke with applications to critical illness insurance underwriting {II}: applications, *North American Actuarial Journal*, 9 **(2005)** 41–56.
- [51] E. Ozkok, G. Streftaris, H.R. Waters, A.D. Wilkie, Modelling critical illness claim diagnosis rates II: results, *Scandinavian Actuarial Journal*, 5 **(2014)** 458–482.
- [52] X.S. Yang, *Introduction to Algorithms for Data Mining and Machine Learning*, Academic Press, **2019**.