



**KÜMELENMİŞ VERİLERDE SAĞKALIM ANALİZİ**

**Kübra Elif AKBAŞ**

**BIYOİSTATİSTİK VE TIP BİLİŞİMİ ANABİLİM DALI**

**Tez Danışmanı  
Doç. Dr. Harika Gözde GÖZÜKARA BAĞ**

**Doktora Tezi-2022**

**T.C.  
İNÖNÜ ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**KÜMELENMİŞ VERİLERDE SAĞKALIM ANALİZİ**

**Kübra Elif AKBAŞ**

**Biyoistatistik ve Tıp Bilişimi Anabilim Dalı  
Doktora Tezi**

**Tez Danışmanı  
Doç. Dr. Harika Gözde GÖZÜKARA BAĞ**

**MALATYA  
2022**

**T.C. İNÖNÜ ÜNİVERSİTESİ**  
**Sağlık Bilimleri Enstitüsü Müdürlüğüne**

**ETİK BEYANI**

İnönü Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak “Doç. Dr. Harika Gözde GÖZÜKARA BAĞ” danışmanlığında hazırlayıp sunduğum “Kübra Elif AKBAŞ ” başlıklı Yüksek Lisans/Doktora tezim içinde elde ettiğim verileri, bilgileri, belgeleri akademik ve etik kurallar çerçevesinde elde ettiğimi, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tezimde yararlandığım eserlere bilimsel kurallara uygun atıfta bulunarak kaynak gösterdiğimi, tezimin özgün olduğunu, tezimin çalışma ve yazımında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. 08/06/2022

Kübra Elif AKBAŞ

# İÇİNDEKİLER

ÖZET .....	vii
ABSTRACT.....	viii
SİMGELER VE KISALTIMA DİZİNİ .....	ix
ŞEKİLLER DİZİNİ .....	x
TABLolar DİZİNİ.....	xi
1. GİRİŞ .....	1
2. GENEL BİLGİLER .....	3
2.1. Koşullu Modeller .....	6
2.1.1. Tabakalandırılmış Cox Modeli .....	8
2.1.2. Parametrik Kırılğanlık Modelleri.....	9
2.1.3. Sabit-Weibull Modeli .....	13
2.1.4. Sabit Etkili Cox Modeli .....	14
2.2. Marjinal Modeller .....	14
2.2.1. Marjinal Weibull Model .....	15
2.2.2. Marjinal Cox Model.....	17
2.3. Kümelenmiş verilerde Kaplan-Meier Yöntemi .....	17
2.4. Klasik Cox Modeli.....	18
2.5. Araştırma Konusuna İlişkin Literatür .....	18
3. MATERYAL VE METOT .....	21
3.1. Hipotez.....	21
3.2. Benzetim çalışması .....	21
3.3. Model Seçimi .....	22
4. BULGULAR.....	24
4.1. Küme büyüklüğünün eşit olduğu durumlar .....	24
4.1.1. Küme boyutu $k=2$ ve $n=50$ iken elde edilen bulgular.....	24
4.1.2. Küme boyutu $k=2$ ve $n=100$ iken elde edilen bulgular.....	26
4.1.3. Küme boyutu $k=2$ ve $n=200$ iken elde edilen bulgular.....	27
4.1.4. Küme boyutu $k=2$ ve $n=500$ iken elde edilen bulgular.....	28
4.1.5. Küme boyutu $k=2$ ve $n=1000$ iken elde edilen bulgular.....	29
4.1.6. Küme boyutu $k=4$ ve $n=50$ iken elde edilen bulgular.....	31
4.1.7. Küme boyutu $k=4$ ve $n=100$ iken elde edilen bulgular.....	32
4.1.8. Küme boyutu $k=4$ ve $n=200$ iken elde edilen bulgular.....	33

4.1.9. Küme boyutu $k=4$ ve $n=500$ iken elde edilen bulgular.....	34
4.1.10. Küme boyutu $k=4$ ve $n=1000$ iken elde edilen bulgular.....	35
4.2. Küme boyutunun eşit olmadığı durumlar .....	37
4.2.1. Maksimum küme boyutu $k=4$ ve $n=50$ iken elde edilen bulgular .....	37
4.2.2. Maksimum küme boyutu $k=4$ ve $n=100$ iken elde edilen bulgular .....	38
4.2.3. Maksimum küme boyutu $k=4$ ve $n=200$ iken elde edilen bulgular .....	39
4.2.4. Maksimum küme boyutu $k=4$ ve $n=500$ iken elde edilen bulgular .....	40
4.2.5. Maksimum küme boyutu $k=4$ ve $n=1000$ iken elde edilen bulgular .....	41
4.3. Farklı sağkalım oranlarında testlerin performanslarının karşılaştırılması .....	43
4.3.1. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 10 iken elde edilen bulgular ..	43
4.3.2. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 20 iken elde edilen bulgular ..	44
4.3.3. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 30 iken elde edilen bulgular ..	46
4.3.4. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 40 iken elde edilen bulgular ..	47
4.3.5. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 50 iken elde edilen bulgular ..	48
4.3.6. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 60 iken elde edilen bulgular ..	49
4.3.7. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 70 iken elde edilen bulgular ..	51
4.4. Küme boyutu $k=4$ ve $n=500$ , sağkalım oranlarının (%10-%70) değiştiği durumlar	52
4.4.1. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 10 iken elde edilen bulgular ..	52
4.4.2. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 20 iken elde edilen bulgular ..	54
4.4.3. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 30 iken elde edilen bulgular ..	55
4.4.4. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 40 iken elde edilen bulgular ..	56
4.4.5. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 50 iken elde edilen bulgular ..	57
4.4.6. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 60 iken elde edilen bulgular ..	59
4.4.7. Küme boyutu $k=4$ ve $n=500$ ve sağkalım oranı % 70 iken elde edilen bulgular ..	60
5. TARTIŞMA .....	62
6. SONUÇ VE ÖNERİLER.....	67
KAYNAKLAR .....	69
EKLER.....	74
Ek-1. Özgeçmiş.....	74
Ek-2. Etik Kurul Almama Gerekçesi .....	75

## TEŐEKKÜR

Bu alıőmanın planlanması ve yürütölmesi sürecinde benden destek ve ilgilerini esirgemeyen, lisansüstü eğitimim boyunca, bilgi ve hoşgörülerinden yararlandığım danışman hocam sayın Do. Dr. Harika Gözde GÖZÜKARA BAĞ' a sonsuz teşekkürlerimi sunarım. Bu tez alıőmasının izleme raporlarında değerli görüşleriyle alıőmama katkı sağlayan hocalarım sayın Prof. Dr. Saim YOLOĞLU ve sayın Prof. Dr. M. Cengiz OLAK'a, ayrıca lisansüstü eğitimi boyunca her zaman deneyimlerini paylaşan sayın Prof. Dr. Cemil OLAK' a ve Do. Dr. Mehmet Onur KAYA' ya teşekkürlerimi sunarım.

Desteklerini hiçbir zaman esirgemeyen değerli arkadaşım Arő. Gör. Dr. Betül DAĞOĞLU HARK' a ve Biyoistatistik ve Tıp Biliőimi Anabilim dalında görev yapan ok kıymetli asistan arkadaşlarıma teşekkür ederim. Eğitim hayatım boyunca benden desteklerini esirgemeyen ve her zaman yanımda olan babam Prof. Dr. Mikail ET' e, annem Fahriye ET' e, ablam Arő. Gör. Dr. Sümeyra Zeynep ET' e, kardeşim Ömer Faruk ET' e, eşim Mehmet Lokman AKBAŐ' a ve oğlum Ali Tuğrul AKBAŐ' a teşekkür ederim.

Kübra Elif AKBAŐ

# ÖZET

## Kümelenmiş Verilerde Sağkalım Analizi

**Amaç:** Kümelenmiş verilerde sağkalım analizi, sağkalım süreleri kümeler halinde örneklendiğinde ortaya çıkar. Böylece aynı küme içindeki sağkalım süreleri ilişkilendirilmiş olur. Bu çalışmanın amacı, kümelenmiş sağkalım verilerin analizinde kullanılan ileri istatistiksel analiz yöntemler ile klasik sağkalım analizini karşılaştırmaktır.

**Materyal ve Metot:** Yapılan benzetim çalışmaları iki açıklayıcı değişken içerecek şekilde tasarlanmıştır. Farklı örneklem büyüklüklerinde ve farklı küme büyüklüklerinde üretilmiş verilerle parametre tahminleri elde edilmiştir. Çalışmada üç farklı durum için model performansları karşılaştırılmıştır. Kırılganlık, tabakalandırılmış Cox modeli, marjinal Cox, marjinal weibull ve klasik Cox modeli üretilmiş olan verilere uygulanmış ve performansları akaie bilgi kriterine (AIC) göre karşılaştırılmıştır.

**Bulgular:** Çalışmada yapılan benzetim senaryosu sonuçları her üç durumda elde edilen bulgular birbirleri ile benzer olarak bulunmuştur. Gözlem sayısının az olduğu durumlarda en kötü performansa sahip model marjinal weibull olmuştur. Ancak gözlem sayısı arttıkça klasik Cox en kötü performansa sahip modeldir. Ayrıca tüm durumlarda en iyi performans gösteren model kırılganlık modeli olarak elde edilmiştir.

**Sonuç:** Kümelenmiş verilerde sağkalım analizinde, küme içi bağımlılığı dikkate almamak her zaman yanlış istatistiksel sonuç çıkacak kadar yanlış olmayabilir. Ancak, gözlem sayısının artması ile bu yanlışlığın da artması muhtemeldir. Bu nedenle kümelenmiş sağkalım verilerinin çözümlemesinde küme içi korelasyonu dikkate alan analiz yöntemlerinin kullanılması önerilmektedir.

**Anahtar Kelimeler:** Sağkalım Analizi, Kümelenmiş Veri Analizi, Kümelenmiş Sağkalım Veri Analizi, Koşullu Modeller, Marjinal Modeller

# ABSTRACT

## Survival Analysis of Clustered Data

**Aim:** Survival analysis in clustered data occurs when survival times are sampled in clusters. Thus, the survival times within the same cluster are correlated. The aim of this study is to compare advanced statistical analysis methods used in the analysis of clustered survival data with classical survival analysis.

**Material and Method:** The simulation studies were designed to include two explanatory variables. Parameter estimates were obtained with the data produced in different sample sizes and different cluster sizes. In the study, model performances were compared for three different situations. The frailty, stratified, marginal Cox, marginal Weibull and classical Cox model were applied to the data obtained by simulating and their performances were mixed according to the akaie information criterion (AIC).

**Results:** The results of the simulation scenario made in the study were found to be similar to each other in all three cases. In cases where the number of observations is small, the model with the worst performance was the marginal weibull. However, as the number of observations increases, the classical Cox model has the worst performance. In addition, the best performing model in all cases was obtained as the frailty model.

**Conclusion:** In the survival analysis of clustered data, ignoring intra-cluster dependency may not always be biased enough to result in false statistical results. However, this bias is likely to increase as the number of observations increases. Therefore, it is recommended to use analysis methods that consider intra cluster correlation in the analysis of clustered survival data.

**Key Words:** Survival Analysis, Clustered Data Analysis, Clustered Survival Data Analysis, Conditional Models, Marginal models



## SİMGELER VE KISALTMA DİZİNİ

<b>AIC</b>	: Akaike bilgi kriteri
<b>BIC</b>	: Bayesyen bilgi kriteri
<b>EM</b>	: Beklenti maksimizasyonu
<b>RMSE</b>	: Karekök hata kareler ortalaması



## ŞEKİLLER DİZİNİ

Şekil No	Sayfa No
Şekil 2.1. Sansürleme Tipleri .....	4
Şekil 3.1. Benzetim Algoritması.....	23
Grafik 4.1. Maksimum küme büyüklüğü 2 iken yöntemlere göre AIC değerleri.....	31
Grafik 4.2. Maksimum küme büyüklüğü 4 iken yöntemlere göre AIC değerleri.....	36
Grafik 4.3. Maksimum küme büyüklüğü 4 ve kayıp oranı %25 iken yöntemlere göre AIC değerleri .....	42
Grafik 4.4. Maksimum küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranları %10-%70 iken yöntemlere göre AIC değerleri .....	52
Grafik 4.5. Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranları %10-%70 iken yöntemlere göre AIC değerleri .....	61

## TABLULAR DİZİNİ

Tablo No	Sayfa No
<b>Tablo 4.1.</b> Küme büyüklüğü 2 ve birey sayısı 50 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	24
<b>Tablo 4.2.</b> Küme büyüklüğü 2 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri .....	25
<b>Tablo 4.3.</b> Küme büyüklüğü 2 ve birey sayısı 100 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	26
<b>Tablo 4.4.</b> Küme büyüklüğü 2 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri .....	26
<b>Tablo 4.5.</b> Küme büyüklüğü 2 ve birey sayısı 200 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	27
<b>Tablo 4.6.</b> Küme büyüklüğü 2 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri .....	27
<b>Tablo 4.7.</b> Maksimum küme büyüklüğü 2 ve birey sayısı 500 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	28
<b>Tablo 4.8.</b> Küme büyüklüğü 2 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri .....	29
<b>Tablo 4.9.</b> Maksimum küme büyüklüğü 2 ve birey sayısı 1000 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	30
<b>Tablo 4.10.</b> Küme büyüklüğü 2 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri.....	30
<b>Tablo 4.11.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	31
<b>Tablo 4.12.</b> Küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri.....	32
<b>Tablo 4.13.</b> Küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	32
<b>Tablo 4.14.</b> Küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri.....	33

<b>Tablo 4.15.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	33
<b>Tablo 4.16.</b> Küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri.....	34
<b>Tablo 4.17.</b> Küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	34
<b>Tablo 4.18.</b> Küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri.....	35
<b>Tablo 4.19.</b> Küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri.....	35
<b>Tablo 4.20.</b> Küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri.....	36
<b>Tablo 4.21.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	37
<b>Tablo 4.22.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri .....	37
<b>Tablo 4.23.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	38
<b>Tablo 4.24.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri .....	38
<b>Tablo 4.25.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	39
<b>Tablo 4.26.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri .....	40
<b>Tablo 4.27.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	40
<b>Tablo 4.28.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri .....	41
<b>Tablo 4.29.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	41
<b>Tablo 4.30.</b> Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri .....	42

<b>Tablo 4.31.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %10 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	43
<b>Tablo 4.32.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %10 iken yöntemler için elde edilen AIC ve uyum değerleri .....	44
<b>Tablo 4.33.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %20 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	45
<b>Tablo 4.34.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %20 iken yöntemler için elde edilen AIC ve uyum değerleri .....	45
<b>Tablo 4.35.</b> Küme küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %30 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	46
<b>Tablo 4.36.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %30 iken yöntemler için elde edilen AIC ve uyum değerleri .....	46
<b>Tablo 4.37.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %40 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	47
<b>Tablo 4.38.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %40 iken yöntemler için elde edilen AIC ve uyum değerleri .....	47
<b>Tablo 4.39.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %50 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	48
<b>Tablo 4.40.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %50 iken yöntemler için elde edilen AIC ve uyum değerleri .....	49
<b>Tablo 4.41.</b> Maksimum küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %60 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	50
<b>Tablo 4.42.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %60 iken yöntemler için elde edilen AIC ve uyum değerleri .....	50
<b>Tablo 4.43.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %70 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	51
<b>Tablo 4.44.</b> Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %70 iken yöntemler için elde edilen AIC ve uyum değerleri .....	51
<b>Tablo 4.45.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %10 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	53
<b>Tablo 4.46.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %10 iken yöntemler için elde edilen AIC ve uyum değerleri .....	53

<b>Tablo 4.47.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %20 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	54
<b>Tablo 4.48.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %20 iken yöntemler için elde edilen AIC ve uyum değerleri .....	54
<b>Tablo 4.49.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %30 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	55
<b>Tablo 4.50.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %30 iken yöntemler için elde edilen AIC ve uyum değerleri .....	55
<b>Tablo 4.51.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %40 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	56
<b>Tablo 4.52.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %40 iken yöntemler için elde edilen AIC ve uyum değerleri .....	57
<b>Tablo 4.53.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %50 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	58
<b>Tablo 4.54.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %50 iken yöntemler için elde edilen AIC ve uyum değerleri .....	58
<b>Tablo 4.55.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %60 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	59
<b>Tablo 4.56.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %60 iken yöntemler için elde edilen AIC ve uyum değerleri .....	59
<b>Tablo 4.57.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %70 iken yöntemler için elde edilen $\exp(\hat{\beta})$ , $se(\exp(\hat{\beta}))$ ve p değerleri .....	60
<b>Tablo 4.58.</b> Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %70 iken yöntemler için elde edilen AIC ve uyum değerleri .....	61

# 1. GİRİŞ

Sağkalım analizi, iyi tanımlanmış bir olayın başlangıcından, olay gerçekleşene kadar geçen süreyi çözümlmek için kullanılan istatistiksel bir yaklaşımdır (1). Sağkalım analizi tıp, biyoloji, epidemiyoloji, nüfus bilimleri ve mühendislik gibi birçok alanda kullanılan temel araştırma yöntemlerinden biridir (2). Bu analiz ile ilgili ilk araştırmalar ampullerin ve diğer elektrikli cihazların aktif ömrü hakkındaki araştırmalardan ortaya çıkmış olsa da, sağkalım analizinin modern uygulamaları hayatta kalmama olaylarının pek çoğunu içerir (3).

Sağkalım analizi diğer tüm istatistiksel yöntemlerde olduğu üzere ölüm, boşanma ve intihar gibi rasgele olaylarla ilgilenir. Bu nedenle, ilgilenilen olayın zamanı negatif olmayan rasgele bir deęişkendir ve genellikle saęa çarpıktır, çünkü çoęu olay bir süre sonra birbirine yakın bir yerde meydana gelme eğilimindedir (3). Sağkalım analizinde cevap deęişkeni olarak adlandırılan  $T$ ; bir bireyin sağkalım süresi veya bir hastalığın başlangıcındaki yaş veya bir hastalığın tedavisinden nüksüne kadar geçen süre olabilir.

Sağkalım analizinde, bazı olay zamanları sansür nedeniyle eksik olarak gözlemlenebilir (4). Gözlenen son zaman noktasında, ilgilenilen olayın gerçekleşmemesi durumunda, o bireyin sağkalım süresinin sansürlü olduğu söylenir. Bunun nedeni, bir araştırmadan elde edilen verilerin bazı kişilerin hala hayatta olduğu bir zamanda analiz edilmeleri olabilir veya analiz sırasında bir bireyin takip edilememesinden dolayı sağkalım durumu bilinmeyebilir. Gerçek bir sağkalım süresi, ölümün tedavi ile ilgisiz olduğu bilinen bir nedenden geldiğinde de sansürlenmiş olarak da kabul edilebilir. Bu durumların her birinde,  $t_0$  zamanında çalışmaya giren hasta,  $t_0+t$  zamanında ölür. Fakat birey hala hayatta olduğu için ya da takipten kaybolduęu için  $t$  bilinmemektedir. Bireyin en son  $t_0+c$  zamanında hayatta olduğu biliniyorsa,  $c$  süresine sansürlü sağkalım süresi denir. Bu sansür, bireyin bir araştırmaya girmesinden sonra, yani bilinen en son sağkalım süresinin saęında meydana gelirse saędan sansürleme adını alır. Bir başka sansür biçimi olan soldan sansürlemeyle bir bireyin gerçek sağkalım süresi gözlenenden daha az olduğunda karşılaşılr. Diğer bir sansürleme çeşidi aralıklı sansürlemedir. Bu tip sansürlemede bireylerin aralıklı zaman içinde bir olay yaşadığı bilinmektedir (5).

Çok deęişkenli sağkalım süresi verileri, her bir denek potansiyel olarak birkaç olayı deneyimleyebildiğinde (örneğin ameliyat sonrası çoklu enfeksiyonlar) veya aynı kümenin sağkalım süreleri arasında bağımlılık yaratan bazı doğal veya yapay kümelenmeler (örneğin, bir defada doğan fare yavruları) olduğunda ortaya çıkar. İlk

durumdaki veriler tekrarlayan çoklu olay verileri olarak ifade edilir. İkinci durumdaki veriler ise kümelenmiş verilerdir (1).

Kümelenmiş veriler, boylamsal çalışmalar, müdahale çalışmaları ve klinik denemeler dâhil birçok biyomedikal araştırma alanında ortaya çıkmaktadır (6). Bu tür verilere örnek olarak, tek bir kişiden alınan kan basıncının tekrarlı ölçümleri, kemirgenlerin kullanıldığı bir deneyde hayvan yavrularının yanıtları veya kardeşlerin vücut kitle indeksi verilebilir (7). Ayrıca periodontal çalışmalarda bireyin ağızındaki birden fazla alan (diş etleri veya dişler), oftalmolojik araştırmalarda, her iki gözden alınan ölçümler ve teratolojik çalışmalarda, bir hayvandan doğan yavruların ölçümlerinden elde edilen veriler kümelenmiş veri yapısına örnek gösterilebilir (8).

Kümelenmiş verilerin analizindeki temel problem, kümelenme içindeki gözlemlerin bağımsız olmaması ve benzerlik derecesinin tipik olarak küme içi korelasyon katsayısı ile ölçülmesidir (9). Bu tür verilerin analizinde küme yapısını göz ardı eden standart analiz yöntemlerini kullanmak, tahmini bir tedavi farklılığının gerçek standart hatasının azalmasına neden olma eğilimindedir (10). Kümelenmiş veriler yapısı gereği korelasyon nedeniyle istatistiksel analizler yapılırken küme içi korelasyonun göz ardı edilmesi ile yanlış p değerlerinin, küçük güven aralıklarının, yanlış kestirimlerin ve etki genişliklerinin elde edilmesine neden olabilir. Bu durum ise değişkenler arasındaki ilişkilerin yanlış yorumlanmasına neden olabilir (9).

Kümelenmiş sağkalım süresi verileri, çalışma süreleri kümeler halinde örneklendiğinde ortaya çıkar, böylece aynı küme içindeki sağkalım süreleri ilişkilendirilir. Bu tür veriler, aile bireyleri arasında genetik bir hastalığın başlangıcı, kanserojene maruz kalan yavrularda tümörlerin ortaya çıkması, sol ve sağ gözlerde görme kaybının ortaya çıkması ve sınıf arkadaşlarının sigara içmeye başlaması olarak örneklendirilebilir (11).

Bu çalışmanın amacı, kümelenmiş verilerde sağkalım analizleri hakkında bilgi vermek, kümelenmiş sağkalım süresinin analizinde kullanılan ileri istatistiksel yöntemleri tanıtmak ve klasik sağkalım analizi yöntemleri ile bağımlılığı dikkate alan ileri istatistiksel analiz yöntemlerini karşılaştırmaktır. Bu amaçla, sağlık alanında karşımıza çıkabilecek hipotetik bir kümelenmiş sağkalım verisi türetilenektir. Kümelenmiş verideki ilişkili olma durumunu dikkate alan sağkalım çözümlemesi yöntemleri kullanılarak, küme-içi korelasyondan kaynaklı yanlış p değerleri, küçük güven aralıkları ve yüksek tip I hata gibi sorunların giderilmesi amaçlanmıştır.

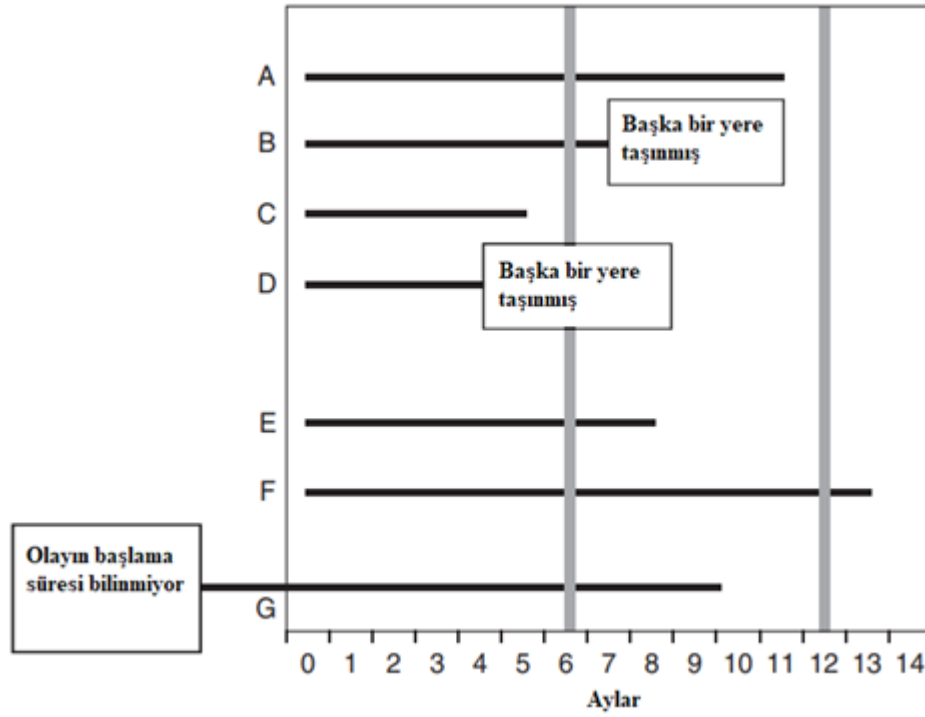


## 2. GENEL BİLGİLER

Sağkalım analizi, iki olay arasındaki süreyi veya daha genel olarak birkaç durum veya koşul arasındaki geçiş zamanlarını içeren verileri analiz etmek için çeşitli alanlarda kullanılır. Sağkalım süresi terimi, risk süresi, yaşam süresi, başarısızlık süresi ve belirli bir olaya kadar geçen süre terimleri ile birbirinin yerine kullanılabilir. Sağkalım süresini belirlemek için iki zaman noktasının tanımlanması gerekir: Başlangıç zamanı, yani doğum gibi orijinal bir olayın meydana geldiği zaman ve sağkalım zamanı, yani ölüm gibi son olayın meydana geldiği zaman (12). Ayrıca sağkalım analizi, sağkalım sürelerinin ve onları etkileyen faktörlerin incelenmesi ile ilgilidir. Sağkalım analizi çalışma türlerinde klinik deneyler, ileriye dönük ve geriye dönük gözlemsel çalışmalar ve hayvan deneyleri bulunur. Sağkalım sürelerine örnekler, doğumdan ölüme kadar geçen süreyi, bir klinik araştırmaya girişten son ana kadar geçen süreyi içerir. Sağkalım çalışmaları, sağkalım dağılımının tahminini, çeşitli tedavilerin veya müdahalelerin sağkalım dağılımlarının karşılaştırmasını veya sağkalım sürelerini etkileyen faktörlerin açıklanmasını içerebilir (13).

Sağkalım analizini diğer istatistik analizlerinden ayıran en önemli özellik, sağkalım verilerinin genellikle bir şekilde sansürlenmiş veya eksik olmasıdır (12). Sansürlenme, verilerin tamamlanmaması anlamına gelir. Sansür, deneklerin yalnızca bir kısmı için kesin olay zamanları bilindiğinde ve geri kalanlarında olay zamanlarının yalnızca belirli bir değeri aştığı bilindiğinde gerçekleşir. Sağdan sansür, soldan sansür ve rasgele sansür olmak üzere üç temel sansürleme biçimi vardır. Sağdan sansürleme, bir olay veya bölümün bitiş noktasının bilinmediği veya veri toplamanın sonunda ilgilenilen olayın henüz gerçekleşmediği durumu ifade eder. Örneğin bir çalışmanın süresinin 12 ay olarak tanımlandığını varsayalım (şekil 1). On iki ay içerisinde olaya kadar geçen süre verilerinin toplanmasının bitiş noktasında, F bireyinin çıkış olayı gerçekleşmediği için verisi sağdan sansürlüdür. Bilinen tek şey bu konudaki asıl olay bitiş noktasının 12 aydan fazla olduğudur. Aynı şekilde, çalışma 6 ay olarak tanımlarsa, A, B, E ve F denekleri sağdan sansürlenir ve bilinen tek bilgi, bu bireylerin olayın meydana gelme sürelerinin 6 aydan fazla olduğudur. Soldan sansür, bir olayın başlangıç noktasının bilinmediği durumu ifade eder. G bireyi soldan sansürlüdür. Rasgele sansürlemede ise, bireyin hem başlangıç hem de bitiş noktalarını gözlemlenir, ancak gözlem ilgilenilen olay dışındaki nedenlerle sonlandırılır. Örneğin, 12 aylık bir çalışmada, B ve D bireyleri için hem

başlangıç hem de bitiş noktalarını gözlemlenir, ancak iki gözlemin sona ermesi, iki vakanın ilgilenilen olayın ortaya çıkması değil, başka yerlere taşınmalarıdır (14).



Şekil 2.1. Sansürleme Tipleri

Sağkalım analizi yöntemleri sağkalım dağılımına bağlıdır ve sağkalım fonksiyonu ve hazard fonksiyonu ile belirlenir. Sağkalım fonksiyonu, bir t noktasına kadar sağkalım olasılığını tanımlar.

$$S(t) = pr(T > t), 0 < t < \infty$$

Fonksiyon 0 anında 1 değerini alır, zamanla azalır ve asla 0'ın altına düşmez. Sağkalım fonksiyonu genellikle başarısızlık oranı olan hazard fonksiyonu ile tanımlanır. Bir birimin t zamanına kadar sağkaldığı göz önüne alındığında, bir sonraki küçük zaman aralığında başarısız olma olasılığı, o aralığın uzunluğuna bölünür ve aşağıdaki gibi ifade edilir.

$$h(t) = \lim_{\delta \rightarrow 0} \frac{pr(t < T < t + \delta | T > t)}{\delta}$$

Bu fonksiyon aynı zamanda yoğunluk fonksiyonu olarak da bilinir. Hazard ve sağkalım fonksiyonları, bir sağkalım dağılımını belirlemenin iki yoludur (13).

Bazı uygulamalarda veriler, genetik özellikler veya ortak çevresel faktörler gibi bazı ortak özellikler nedeniyle kümelenebilir veya ilişkilendirilebilir. Kümelenmiş

sağkalım verileri, doğal oluşturulmuş kümeleri içeren ve her bir küme için paralel olay zamanları gözlemlenen çok değişkenli verileri temsil etmek için kullanılır. Her kümedeki olay süreleri, kümelerin doğası nedeniyle ilişkilendirilebilir (15). Kümelenmiş sağkalım verisi tek çeşit bir olay iki veya daha fazla başka birim için aynı birey üzerinde benzer değerlendirmeler yapıldığında ortaya çıkar. Örneğin, oftalmolojide, görme kaybını hafifletme süresi bir kişinin her iki gözünde ayrı ayrı değerlendirilebilir. Aksine, aynı bireyde farklı olaylar veya tekrarlanan olaylar meydana geldiğinde çok değişkenli sağkalım verileri ortaya çıkar. Örneğin kardiyolojide miyokard enfarktüsüne kadar geçen süre ve inme zamanı aynı birey için değerlendirilebilir (16).

Sağkalım analizindeki birçok uygulamada, veriler sağkalım zamanlarının ilişkilendirilebileceği kümelerde düzenlenir. Örneğin, bir araştırmacı, belli bir popülasyondaki bireyler için diş çürümesinden başlangıcına kadar olan zamanı inceleyebilir. Başlama zamanı, her birey için çoklu dişlerde ölçülür ve kümelenmiş veriler bağlamında, bireyler kümeleri ve dişleri kümeler içindeki potansiyel olarak bağımlı gözlemleri oluşturur (17).

Aynı kümeden elde edilen gözlemler genellikle bazı gözlemlenmemiş özellikleri paylaşır ve sonuç olarak korele olma eğilimindedir. Örneğin ikizlerin, sağkalım şanslarını etkileyen genetik ve çevresel etkiler ortaktır. Çoğu araştırmacı bu durumun sağkalım sürelerinin analizinde dikkate alınması gerektiği konusunda hemfikir olmasına rağmen, genellikle göz ardı edilir. Bu uygulama yetersiz ve yanlı tahminlere yol açabilir (18). Tahmin edileceği gibi daha ciddi sonuçlardan dolayı, aynı kümedeki gözlemler pozitif korelasyon gösterirse, tahminlerin kesinliğinin abartılması muhtemeldir (18). Bu nedenle küme içi ilişkiyi dikkate alan ileri istatistiksel analiz yöntemlerine literatürde yer verilmektedir. Kümelenmiş sağkalım verilerinin modellenmesinde çalışmanın amacına bağlı olmak üzere iki yaklaşım vardır. Bunlardan ilki olan koşullu modellerde, modelin rasgele modellendiği ve modelin koşullu gözlemlenmeyen etkilerinin her birey için geçerli olduğu temeline dayanır. İkinci yaklaşım olan marjinal modellerde ise, ortak değişkenler koşulsuz olarak belirlenir ve regresyon modelinin her birey için marjinal olduğu, ancak gruplar içindeki bireylerin ilişkili olduğu varsayılır (4). Koşullu modeller, rastgele etkileri dâhil ederek bağımlılığı indüklerken, marjinal modellerde doğrudan sabit etkileri tanımlanır (19).

Merkez etkilerin neden olduğu bağımlılığın varlığında koşullu (merkeze özgü) ve marjinal (popülasyon ortalamalı) modeller olmak üzere iki farklı yaklaşım mevcuttur. Her iki yöntemde tahminleme ve yorumlama farklı şekildedir (20).

N kümesinden (ya da merkez) her biri n büyüklüğünde bir grup birey gözlemlendiği varsayalım.  $T_{ij}$  ve  $X_{ij}$  sırasıyla i. kümedeki j. birey için sağkalım süresi ve maruz kalmayı gösterecek.  $T_{ij}|X_{ij}$  için sağkalım fonksiyonu aşağıdaki gibi tanımlanır.

$$S(t_{ij}|X_{ij}) = P(T_{ij} > t_{ij}|X_{ij})$$

ve  $T_{ij}|X_{ij}$  için hazard fonksiyonu,

$$h(t_{ij}|X_{ij}) = \lim_{d \rightarrow 0} \frac{P(t_{ij} \leq T_{ij} < t_{ij} + d | X_{ij}, T_{ij} \geq t_{ij})}{d}$$

şeklinde dir.

Sağdan sansürün olduğu varsayalım. Sansürleme i. kümedeki j. bireyin sansürlendiği zamanı gösteren  $C_{ij}$  sansür değişkeni ile modellenir. Sağkalım göstergesi  $\Delta_{ij} = T_{ij} \leq C_{ij}$  ve  $V_{ij} = \min(T_{ij}, C_{ij})$  olsun.  $(V_{ij}, \Delta_{ij}, X_{ij})$  tesadüfi vektörlerin  $(v_{ij}, \delta_{ij}, x_{ij})$  bağıntılarının ne olduğunu gözlemlenir. Böyle bir model aynı birey üzerinde sağkalım (veya ölüm) ve sansürleme olmayacağı için gerçeği yansıtmamaktadır. Böylece  $(T_{ij}, C_{ij})$  tesadüfi vektörleri gözlemlenemez. Sağkalım süresi olan  $T_{ij}$ , sansür olmadığı durumlarda gerçekleşecek olan varsayım olarak kabul edilebilir. Belirlenen şartlar altında,  $T_{ij}$  sağkalım süreleri dağılımlarını indeksleyen parametreler gözlemlenen  $(v_{ij}, \delta_{ij}, x_{ij})$  vektörlerinden ayırt edilebilir.  $(T_{i1}, \dots, T_{in})$ ,  $(C_{i1}, \dots, C_{in})$  ve  $(X_{i1}, \dots, X_{in})$  vektörlerini göstermek için sırasıyla  $T_i$ ,  $C_i$  ve  $X_i$  kullanılır.

## 2.1. Koşullu Modeller

Kümelenmiş sağkalım verilerinin modellenmesinde kullanılan ilk yaklaşımdan biri olan koşullu modeller, her birey için rasgele olarak modellenen bazı gözlemlenmemiş etkiyi koşullu tutar (4).

### Varsayımlar

$Z_i$  i kümesinin birimleri arasında paylaşılan gözlenen ve gözlemlenmeyen faktörleri özetlemesi beklenen sayısal rasgele bir değişken olduğu varsayalım. i kümesindeki koşullandırma ile  $Z_i'$  deki koşullandırma aynıdır.  $Z_i$ , i kümesinin kırılğanlığı olarak ifade edilir. Aynı kümedeki bireyler aynı zayıflığa sahip olduklarından,

gözlemlenen sonuçları bağımlı olacaktır. Bununla birlikte, kümeler arasında bağımlılık olmadığı farzedilir, yani

$$(T_1, C_1, X_1, Z_1), \dots, (T_N, C_N, X_N, Z_N) \text{ bağımsızdır (1)}$$

Bu varsayım genellikle hem tasarım ile eşleşen veriler (1: k uyumlu kohort tasarımları gibi) hem de doğal olarak eşleşen veriler (ikiz veriler gibi) için geçerlidir, çünkü herhangi iki küme tipik olarak ilgisizdir. Ayrıca, kırılğanlığa bağlı olarak kümeler içinde bağımsızlık olduğu varsayılır. Yani;

$$(T_{i1}, C_{i1}, X_{i1}, Z_i), \dots, (T_{in}, C_{in}, X_{in}, Z_i) \quad i=1, \dots, N \text{ bağımsızdır (2)}$$

Her bir i kümesi için,

$$p(T_i, C_i | X_i, Z_i) = \prod_{j=1}^n p(T_{ij}, C_{ij} | X_{ij}, Z_i) \text{ dir.}$$

İkinci varsayım genellikle tasarımla eşleşen veriler için geçerlidir, çünkü kırılğanlık (yani eşleşen değişkenler kümesi) aynı kümedeki bireyleri "birbirine bağlayan" tek şeydir. Ancak, doğal olarak eşleşen veriler için her zaman doğru değildir. Örneğin, ikizlerden birini etkileyen olay eğer diğer ikizin olayını da etkiliyorsa, bu durumda sağkalım süreleri bağımsız değildir.  $T_{ij} | X_{ij}, Z_i$  için dağılımı indeksleyen parametreleri tanımlayabilmek için, kümeler içinde rasgele sansürleme aşağıdaki gibi farz edilsin

$$T_{ij} \perp C_{ij} | X_{ij}, Z_i \text{ (3).}$$

Bu varsayım gözlemlenen verileri test etmeden her zaman geçerli değildir. Son olarak, koşullu ( $Z_i$ ) hazardlar arasında orantılılık;

$$\frac{h(t_{ij} | X_{ij}, Z_i)}{h(t_{ij}' | X_{ij}', Z_i)} = e^{\beta(X_{ij} - X_{ij}')} \quad (4)$$

dir.

### 2.1.1. Tabakalandırılmış Cox Modeli

Kümelenmiş veriler için tabakalandırılmış Cox modellerinde, her bir kümedeki hazardlar ayrı ayrı modellenir ve aşağıdaki gibidir.

$$h(t_{ij} | X_{ij}, Z_i) = h_{0i}(t_{ij})e^{\beta X_{ij}}$$

(4)' de verilen varsayım ile koşullu log hazard oranı  $\beta$  elde edilir. (1) - (3) varsayımları ile,  $\beta$  'ya göre kısmi olasılık maksimize edilerek  $\beta$  parametresinin tahmini aşağıdaki gibidir.

$$\begin{aligned} PL(\beta) &= \prod_{i=1}^N \prod_{j=1}^n \left\{ \frac{h(v_{ij} | X_{ij}, Z_i; \beta)}{\sum_{(i,j') \in r_{ij}} h(v_{ij'} | X_{ij'}, Z_i; \beta)} \right\}^{\delta^{ij}} \\ &= \prod_{i=1}^N \prod_{j=1}^n \left\{ \frac{h_{0i}(v_{ij})e^{\beta X_{ij}'}}{\sum_{(i,j') \in r_{ij}} h_{0i}(v_{ij'})e^{\beta X_{ij}'}} \right\}^{\delta^{ij}} \\ &= \prod_{i=1}^N \prod_{j=1}^n \left\{ \frac{e^{\beta X_{ij}}}{\sum_{(i,j') \in r_{ij}} e^{\beta X_{ij}'}} \right\}^{\delta^{ij}} \quad (5) \end{aligned}$$

Burada  $r_{ij} = \{(i,j') | v_{ij'} \geq v_{ij}\}$  i. kümede  $t_{ij}$  için riski belirtir (5). Denklem kümeye özgü  $h_{0i}$  hazardları içermemektedir. Bu nedenle,  $\hat{\beta}$  tahmini  $Z_i'$  ye bağlı değildir. Holt ve Prentice (1974) tarafından bağlantısız sansürlü eşleştirilmiş çift veriler için tahmin ve asimptotik varyansı için basit ifadeler türetilmiştir (21).  $T_{i0}$  ve  $T_{i1}$  sırasıyla maruz kalan ve kalmayan bireylerin sağkalım sürelerini gösterebilir. Kısmi olasılık tahmini aşağıdaki gibi yazılabilir;

$$\hat{\beta} = \log \left( \frac{\hat{p}}{1 - \hat{p}} \right)$$

Burada  $\hat{p} = \frac{\sum_i^N I(t_{i1} < t_{i0})}{N}$ ,  $p = P(T_{i1} < T_{i0})$  olasılığının parametrik olmayan maksimum olasılık tahminidir. Ayrıca tabakalı Cox modelinde  $\beta$  'nın aslında  $\log \left( \frac{\hat{p}}{1 - \hat{p}} \right)$  'a eşit olduğu gösterilmiştir. Tahmin için sadece  $i=1, \dots, N$  için  $I(t_{i1} < t_{i0})$  sıra istatistikleri gerekmektedir. Bu tahminin asimptotik varyansı aşağıdaki gibidir.

$$Var(\hat{\beta}) = \frac{(1 + e^{\beta})^2}{Ne^{\beta}}$$

Varyans,  $\beta$ 'nin gerçek değerine ve N küme sayısına bağlıdır. Genel olarak, bu formüller sansürleme olduğunda kullanılamaz. Çünkü bu durumda,  $I(v_{i1} < v_{i0})$  gözlenen rankların dağılımı ile  $I(t_{i1} < t_{i0})$  sağkalım sürelerinin rankları birbirine denk değildir. Bununla birlikte, Holt ve Prentice'in (1974) önerdiği formüller, sansürleme süresinin her iki çift üye için aynı olduğu durumda kullanılabilir (örneğin Tip 1 sağdan sansürleme gibi) (21). Her iki üye çifti de sansürlendiğinde, kısmi olasılığa hiçbir katkısı olmaz.  $U = \{kmax(\delta_{k1}, \delta_{k0}) = 1\}$  olsun.  $\hat{\beta}$  tahmini sansürün olmadığı durumla aynı fakat  $\hat{p} = \frac{1}{|U|} \sum_{k \in U} I(v_{k1} < v_{k0})$  ile aynı değildir ve asimptotik varyansı aşağıdaki gibidir (22).

$$Var(\hat{\beta}) = \frac{(1 + e^{\beta})^2}{|U|e^{\beta}} \quad (6)$$

### 2.1.2. Parametrik Kırılğanlık Modelleri

Kırılğanlık modelleri, aynı kümenin gözlemleri arasındaki korelasyonu modellemek için kırılğanlıkların veya rasgele etkilerinin çoğunlukla temel hazarda girdiği, kümelenmiş sağkalım verilerinin analizleri için daha popüler hale geldiğini göstermektedir (23, 24).

Kümelenmiş verilerde, aynı kümeden elde edilen gözlemler genellikle gözlemlenmemiş belirli özellikleri paylaştığı için korele olma eğilimindedir. Bir küme içinde böyle bir korelasyonu modellemek için genel bir yaklaşım, sağkalım süresi verileri bağlamında zayıflıklar olarak adlandırılan gizli değişkenleri kullanmaktır (25).

Kırılğanlık modelinin en yaygın kullanılanı, orantılı hazard modellerinin bir uzantısı olan, paylaşılmış kırılğanlık modelidir,  $Z_i$ 'lerin, çoğunlukla gama, log-normal veya pozitif kararlı olan bazı parametrik dağılımın bağımsız örnekleri olması koşuluyla, kümeye özgü temel hazardlar aşağıdaki gibidir.

$$h(t_{ij}|X_{ij}, Z_i) = Z_i h_0(t_{ij}) e^{\beta X_{ij}}$$

Tanımlama amaçları için, tek parametreye bağlı olan kırılğanlık dağılımının parametre uzayını tek boyuta sınırlanabilir. Kırılğanlık modellerinde, koşullu hazard varsayım (4) ile orantılıdır. Koşullu sağkalım fonksiyonları,

$$S(t_{ij}|X_{ij}, Z_i) = \exp\{-Z_i H_0(t_{ij}) e^{\beta X_{ij}}\} \quad (7)$$

şeklinde dir. Burada  $H_0(t) = \int_0^t h_0(s) ds$  dir. Kırılgnalık modeli,  $h_{i0}(t) = Z_i h_0(t)$  ile tabakalı Cox modelinin özel bir durumu olarak kabul edilebilir.

Temel hazard  $h_0(t)$ , Cox modellerinde olduđu gibi belirsiz bırakılabilir ve bu da yarı parametrik bir model verir. Koşullu etki  $\beta$ , beklenti maksimizasyon (EM) algoritması kullanılarak tahmin edilebilir (26). Temel hazard bazı parametrik dağılım için hazard olarak modellenirse, tam parametrik bir model elde edilir. Daha sonra kırılgnlıklar üzerinde marjinalleşmiş yoğunlukları elde etmek mümkündür.  $\beta$  bu yoğunluklara bağlı olasılığı maksimize ederek tahmin edilebilir. Kırılgnlık modelleri için gerekli olan, ancak tabakalı Cox modelleri için gerekli olmayan iki varsayım vardır. İlk varsayım, maruziyet kırılgnlıktan bağımsızdır:

$$X_i \perp Z_i, \quad i = 1, \dots, N \quad (8)$$

Bu varsayım göz ardı edilirse,  $\hat{\beta}$ 'nin etki tahmini  $\beta$  için tutarlı değildir. Örneğin, eğer Z etki karıştırıcılar içeriyorsa ifade (8) ihlal edilir. Bununla birlikte, eşleştirilmiş kohort çalışmalarında, (8) sabittir çünkü  $X_i$  vektörü  $i$  üzerinde sabittir. Örneğin 1:1 eşleşmiş çalışmalarda,  $(X_{i1}, X_{i2}) = (1, 0)$  'dır, çünkü  $X_{i1}$  ve  $X_{i2}$  'nin sıralaması önemsizdir. İkinci varsayım, sansürün maruziyete bağlı zayıflıktan bağımsız olduğudur yani,

$$C_{ij} \perp Z_i | X_{ij} \quad (9)$$

dır. Varsayım (9) gerçek veriler için geçerli değildir. (1)-(3) ile verilen varsayımlarla birlikte, gözlemlenen verilere dayalı tam olasılığın hesaplanması için varsayımlar (8) ve (9) gereklidir. (8) ve (9) varsayımlarını açıklamak amacıyla tam parametrik bir kırılgnlık modelinin olasılığını türetilsin.  $i=1, \dots, n$  ve  $j=1, \dots, n$  için  $(v_{ij}, \delta_{ij}, x_{ij})$  vektörlerinin gözlemlendiği farz edilsin. Şimdi

$$V_{ij} = v_{ij}, \Delta_{ij} = \delta_{ij}, X_{ij} = x_{ij}, Z_i = z_i$$

↔

$$\begin{cases} T_{ij} = v_{ij}, C_{ij} > v_{ij}, X_{ij} = x_{ij}, Z_i = z_i \delta_{ij} = 1 \\ T_{ij} > v_{ij}, C_{ij} = v_{ij}, X_{ij} = x_{ij}, Z_i = z_i \delta_{ij} = 0 \end{cases}$$

olsun, böylece olasılık  $T_{ij}$ ,  $C_{ij}$ ,  $X_{ij}$  ve  $Z_i$  cinsinden ifade edilebilir. Varsayım (1), olasılık kümelenme yoluyla çarpanlarına ayrılır.  $k$  bir kümedeki sansürlenmemiş gözlem sayısını



ifade etsin. Gözlemlerin sıralaması keyfi olduğunda,  $\delta_1 = \dots = \delta_k = k$  ve  $\delta_{k+1} = \dots = \delta_n = 0$ dır. Daha sonra,  $X'$  e koşullu olarak, kümelerin olasılığa katkısı aşağıdaki gibidir.

$$\begin{aligned} & p(T_1 = v_1, \dots, T_k = v_k, T_{k+1} > v_{k+1}, \dots, T_n > v_n, \\ & C_1 > v_1, \dots, C_k > v_k, C_{k+1} = v_{k+1}, \dots, C_n = v_n | \mathbf{X}) \\ & = E_{Z|X} \{ p(T_1 = v_1, \dots, T_k = v_k, T_{k+1} > v_{k+1}, \dots, T_n > v_n \\ & C_1 > v_1, \dots, C_k > v_k, C_{k+1} = v_{k+1}, \dots, C_n = v_n | \mathbf{X}, Z) \} \end{aligned}$$

(2) varsayımı ile

$$E_{Z|X} \left\{ \sum_{j=1}^k p(T_j = v_j, C_j > v_j | X_j, Z) \sum_{j=k+1}^n p(T_j > v_j, C_j = v_j | X_j, Z) \right\} \quad (10)$$

dır. (3) varsayımı ile yaşam süreleri ve sansür süreleri  $Z$  ve  $X$  koşullarından bağımsızdır. Bu nedenle;

$$p(T_j = v_j, C_j > v_j | X_j, Z) = p(T_j = v_j | X_j, Z) p(C_j > v_j | X_j, Z)$$

ve

$$p(T_j > v_j, C_j = v_j | X_j, Z) = p(T_j > v_j | X_j, Z) p(C_j = v_j | X_j, Z)$$

(9) varsayımı ile sansür zayıflıktan bağımsızdır, dolayısıyla,

$$p(C_j > v_j | X_j, Z) = p(C_i > v_j | X_j)$$

ve

$$p(C_j = v_j | X_j, Z) = p(C_i = v_j | X_j)$$

dır. Bu nedenle, (10) eşitliği aşağıdaki gibidir.

$$\begin{aligned} & E_{Z|X} \left\{ \sum_{j=1}^k p(T_j = v_j | X_j, Z) \sum_{j=k+1}^n p(T_j > v_j | X_j, Z) \right\} \\ & \sum_{j=1}^k p(C_j > v_j | X_j) \sum_{j=k+1}^n p(C_j = v_j | X_j) \quad (11) \end{aligned}$$

Böylece, sansürleme süreleri yalnızca  $T_j$  ve kırılma  $Z$  ömürleri için dağılım parametrelerine bağlı olmayan bir faktörle olasılığa katkıda bulunur. (3) ve (9) varsayımı olmadan, 'sansürleme' kısmını (10) 'daki beklentiden çarpanlarına ayıramaz. (3) ve (9)

varsayımları olmaksızın, sansürlemenin (10)'daki beklenen kısmı gerçekleştirilemez.  $X$  ve  $Y$  arasında (8) de verilen bağımsızlık kabulü,  $Z$  entegre edildiğinde  $Z$ 'nin koşulsuz dağılımı kullanılabilir; diğer bir deyişle

$$\begin{aligned} & E_{Z|X} \left\{ \sum_{j=1}^k p(T_j = v_j | X_j, Z) \sum_{j=k+1}^n p(T_j > | X_j, Z) \right\} \\ &= E_Z \left\{ \sum_{j=1}^k p(T_j = v_j | X_j, Z) \sum_{j=k+1}^n p(T_j > | X_j, Z) \right\} \quad (12) \end{aligned}$$

dır.  $Z|X$ 'in dağılımı bilinmiyorsa,  $Z|X$  üzerinde marjinalleşme olasılığı hesaplanamaz. Varsayım (8),  $Z|X$ 'in  $Z$  ile aynı dağılıma sahip olduğu anlamına gelir.

$$p(T_j > v_j | X_j, Z) = S(t_j | X_j, Z)$$

ve

$$p(T_j = v_j | X_j, Z) = -\frac{\partial}{\partial t_j} S(t_j | X_j, Z)$$

olduğundan (12) varsayımı koşullu sağkalım fonksiyonları açısından yeniden yazılabilir;

$$E_Z \left\{ (-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} \prod_{j=1}^n S(v_j | X_j, Z) \right\}$$

Entegrasyonu ve türetmeyi değiştirerek,

$$E_Z \left\{ (-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} \prod_{j=1}^n S(v_j | X_j, Z) \right\} = (-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} E_Z \left\{ \prod_{j=1}^n S(v_j | X_j, Z) \right\}$$

elde edilir. Şimdi  $L_Z$ ,  $Z$ 'nin yoğunluk Laplace dönüşümü olmak üzere;

$$\begin{aligned} & E_Z \left[ \prod_{j=1}^n S(v_j | X_j, Z) \right] \\ &= E_Z \left[ \prod_{j=1}^n \exp\{-ZH_0(v_j) e^{\beta X_j}\} \right] \end{aligned}$$

$$\begin{aligned}
&= E_Z \left[ \exp \left\{ -Z \sum_{j=1}^n H_0(v_j) e^{\beta X_j} \right\} \right] \\
&= L_Z \left\{ \sum_{j=1}^n H_0(v_j) e^{\beta X_j} \right\}
\end{aligned}$$

dır. Dolayısıyla;

$$\begin{aligned}
&(-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} L_Z \left\{ \sum_{j=1}^n H_0(v_j) e^{\beta X_j} \right\} \\
&= (-1)^k \prod_{j=1}^k \{H_0(v_j) e^{\beta X_j}\} L_Z^{(k)} \left\{ \sum_{j=1}^n H_0(v_j) e^{\beta X_j} \right\}
\end{aligned}$$

dır. Kümelerin alt indisleri tekrar tanıtarak ve  $D_i = \sum_{j=1}^n \delta_{ij}$  olmak üzere, olasılık

$$L(\beta) = \prod_{i=1}^N \left[ (-1)^{D_i} \left\{ \prod_{j=1}^n h(v_{ij}|X_{ij})^{\delta_{ij}} \right\} L_Z^{(D_i)} \left\{ \sum_{j=1}^n H(v_{ij}|X_{ij}) \right\} \right]$$

Şeklindedir (22).

### 2.1.3. Sabit-Weibull Modeli

Bu modelde, Weibull temel hazardları ve pozitif kararlı kırılma hızları olan bir parametrik kırılma hızı modeli kullanılır. Bunun nedeni, Sabit-Weibull modelinde hem koşullu hazardların hem de marjinal hazardların orantılı olmasıdır. Bu modelde temel hazard, Weibull hazard parametrik olarak aşağıdaki gibi modellenir:

$$h_0(t) = \lambda_0 c t^{c-1}$$

Buna göre kümülatif hazardlar;

$$H(t|X) = \lambda_0 e^{\beta X} t^c$$

dır.  $0 < \alpha < 1$  olmak üzere, kırılma hızının pozitif kararlı dağılımlar ailesine ait olduğu varsayılır ve yoğunluğu,

$$f_Z(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} (-z^{-\alpha})^k \sin(k\alpha\pi)$$

dır.  $L(s) = e^{-s^\alpha}$  olmak üzere laplace dönüşümünün daha basit bir ifadesi vardır. Bu nedenle, her bir kümeden  $k$  sansürlü gözlemlerin olasılığı aşağıdaki gibidir (22);

$$\begin{aligned}
& (-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} L_Z \left\{ \sum_{j=1}^n H_0(v_j) e^{\beta X_j} \right\} \\
& = (-1)^k \frac{\partial^k}{\partial v_1, \dots, v_k} \exp \left\{ - \left( \sum_{j=1}^n \lambda_0 e^{\beta X_j} v_j^c \right)^\alpha \right\}
\end{aligned}$$

#### 2.1.4. Sabit Etkili Cox Modeli

Sansürlü olay zaman verilerinin,  $N = \sum_{k=1}^K n_k$  toplam örnek sayısını,  $n_k$  ( $k=1, \dots, K$ ) küme başına düşen birey sayısını göstermek üzere  $K$  kümelerinden alındığı (veya merkezlerinden) varsayalım. Bu yöntemde, koşullu modellerle merkez etkilerini modellemeye yönelik bir yaklaşım sunulmaktadır ve formül aşağıdaki gibidir.

$$\lambda_{ki}(t) = \lambda_{0k}(t) \exp\{\beta^T Z_{ki}\} \quad (13)$$

Kümelerin sağkalım riski üzerinde orantılı hareket ettiği varsayılarak modele (13) eklenebilir.

$\xi_k$  merkezi etkileri temsil etmek ve  $\lambda_0(t)$  temel hazard fonksiyonu olmak üzere,  $\lambda_{0k}(t) = \xi_k \lambda_0(t)$  olduğu varsayalım.  $k=1, \dots, K$  ve  $\alpha_1 = 0$  olmak üzere 1. küme keyfi olarak referans küme olarak ayarlanarak;

$$\lambda_{0k}(t) = \lambda_0(t) \exp\{\alpha_k\} \quad (14)$$

elde edilir. Bu model, tabakalandırılmamış Cox modelinin merkez için gösterge değişkenleri eklenerek tahmin edilebilir. Küme sayısı örneklem büyüklüğüne göre az olduğunda, merkezi etkiler ilgi çekici bir yaklaşımdır (20).

## 2.2. Marjinal Modeller

Kümelenmiş sağkalım verilerinin modellenmesine ilişkin ikinci yöntem marjinal modellemedir. Bu yöntem, her olay zamanının marjinal dağılımını modelleyen bir yaklaşımdır (15). Kümelenmiş veriler için marjinal modellerde,  $\beta^*$  tahmini, yaşam sürelerinin marjinal ( $Z$  üstü) yoğunluğuna dayanmaktadır. Verilerdeki kümelenmeden meydana gelen korelasyonlar için standart hatalar düzeltilir. Parametrik durumda, bu olasılık marjinal dağılımlara bağlıdır. Yarı parametrik durumda ise sıradan Cox regresyon modelinin kısmi olasılığı kullanılır. Bu durum hem parametrik durumda hem de yarı parametrik durumda  $\beta^*$  'in tutarlı tahminlerini verir.

## Varsayımlar

Marjinal modeller rasgele bir Z bileşeni içermediğinden, bu modeller için farklı varsayımlar gereklidir. Bununla birlikte kümeler arasında bağımsızlık varsayımı kabulüne ihtiyaç duyulmaktadır. Yani;

$$(T_1, C_1, X_1), \dots, (T_N, C_N, X_N) \quad (15)$$

bağımsızdır. Bu varsayım aslında (1)' den dolayı açıktır. Ayrıca her bir birey için bağımsız sansür olduğu varsayılır, yani

$$T_{ij} \perp C_{ij} | X_{ij} \quad (16)$$

dır. Ayrıca orantılı marjinal hazardlar kabulüne ihtiyaç bulunmaktadır, yani

$$\frac{h(t|X)}{h(t|X')} = e^{\beta(X-X')} \quad (17)$$

dır. (22)

### 2.2.1. Marjinal Weibull Model

Parametrik marjinal modelleme, sadece küme üyeliği bilgileriyle birlikte marjinal sağkalım sürelerinin belirtilmesini gerektirir. Marjinal sağkalım süreleri, çoğunlukla üstel, Weibull, normal veya lognormal olmak üzere bazı parametrik dağılımları ile modellenebilir. Bir marjinal parametrik model, koşullu sağkalım fonksiyonları üzerinden beklentisi alınarak parametrik bir kırılma modelinden türetilebilir, yani

$$\begin{aligned} S(t|X) &= E_{Z|X}[S(t|X, Z)] \\ &= E_Z[S(t|X, Z)] \\ &= E_Z[\exp\{-ZH_0(t)e^{\beta X}\}] \\ &= L_Z\{H_0(t)e^{\beta X}\} \end{aligned}$$

dır. Bu taktirde marjinal hazard fonksiyonları aşağıdaki gibidir.

$$h(t|X) = \frac{\partial}{\partial t} [-\log\{S(t|X)\}]$$

Parametrik kırılma modelinden türetilen marjinal hazard fonksiyonlarının orantılı olması gerekmemektedir. Marjinal hazardlar,

$$h(t|X) = (\lambda_0^\alpha e^{\alpha\beta X} \alpha c t^{\alpha c - 1})$$

ve marjinal sağkalım fonksiyonları

$$\exp(-\lambda_0^\alpha e^{\alpha\beta X} t^{\alpha c})$$

olmak üzere bir istisna sabit-Weibull modelidir.

Bu, ölçek parametreleri  $\lambda_0^\alpha e^{\alpha\beta X}$  ve şekil parametresi  $\alpha c$  olmak üzere Weibull dağılmış sağkalım sürelerine karşılık gelir. Log hazard oranı  $\beta^* = \alpha\beta$ , aynı zamanda marjinal Cox modelinde hedeflenen parametre ile aynıdır.

Weibull modeli altında marjinal yoğunluğa dayalı, i. kümedeki j. bireyin olasılığını türetilsin: Bu durumda son eşitlik rasgele sansür varsayımından (16) kaynaklanmak üzere aşağıdaki eşitlik yazılabilir.

$$\begin{aligned} & p(V_{ij} = v_{ij}, \Delta_{ij} = \delta_{ij} | X_{ij}) \\ &= \{p(T_{ij} = v_{ij}, C_{ij} > v_{ij} | X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij}, C_{ij} = v_{ij} | X_{ij})\}^{1-\delta_{ij}} \\ &= \{p(T_{ij} = v_{ij} | X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij} | X_{ij})\}^{1-\delta_{ij}} \\ &\quad \{p(C_{ij} > v_{ij} | X_{ij})\}^{\delta_{ij}} \{p(C_{ij} = v_{ij} | X_{ij})\}^{1-\delta_{ij}} \end{aligned}$$

Tüm birey ve kümeler üzerinden çarpım alınarak ve  $C_{ij}$  terimini yok sayarak, aşağıdaki eşitlik yazılabilir (22).

$$\begin{aligned} L(\beta) &= \prod_{i=1}^N \prod_{j=1}^n \{p(T_{ij} = v_{ij} | X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij} | X_{ij})\}^{1-\delta_{ij}} \\ &= \prod_{i=1}^N \prod_{j=1}^n f(v_{ij} | X_{ij})^{\delta_{ij}} S(v_{ij} | X_{ij})^{1-\delta_{ij}} \\ &= \prod_{i=1}^N \prod_{j=1}^n h(v_{ij} | X_{ij})^{\delta_{ij}} S(v_{ij} | X_{ij}) \\ &= \prod_{i=1}^N \prod_{j=1}^n (\lambda_0^\alpha e^{\alpha\beta X_{ij}} \alpha c t_{ij}^{\alpha c - 1})^{\delta_{ij}} \exp(-\lambda_0^\alpha e^{\alpha\beta X_{ij}} t_{ij}^{\alpha c}) \quad (18) \end{aligned}$$

### 2.2.2. Marjinal Cox Model

$j=1,2,\dots,J$  her bir kümenin birimleri olmak üzere,  $i=1,2,\dots,n$  kümeleri (merkezler, aileler, hayvan yavruları vb.) gösterebiliriz.  $T_{ij}$  kümelenmiş sağkalım süresi ve  $C_{ij}$  bilgilendirici olmayan sansür süresini olmak üzere, her bir birim için  $X_{ij}$  takip süresi göz önüne alınsın.  $I(\cdot)$  gösterge fonksiyonu olmak üzere, olay göstergesi  $\Delta = I(T_{ij} \leq C_{ij})$  ile gösterilsin. Her bir birim için,  $t$  zamanında sabit etkiler için bir  $Z_{ij}(t)$  ortak değişken vektör vardır.  $i$ . küme ve  $j$ . olay için marjinal cox model aşağıdaki gibidir.

$$\lambda_j(t; Z_{ij}) = \lambda_{0j} e^{\beta' Z_{ij}(t)}, j=1,\dots,J, i=1,\dots,n$$

ve  $j$ . olaya özgü kısmi olasılık ise;

$$L_j(\beta) = \prod_{i=1}^n \left[ \frac{\exp\{\beta' Z_{ij}(X_{ij})\}}{\sum_{l \in R_j(X_{ij})} \exp\{\beta' Z_{il}(X_{ij})\}} \right]^{\Delta_{ij}}$$

dır, burada  $\lambda_{0j}(t)$ ,  $j$ . olayı için keyfi bir temel hazard fonksiyonu ve  $\beta_j$ ,  $j$ . olay için regresyon katsayılarının (olaya özgü) sütun vektörüdür (1).

### 2.3. Kümelenmiş verilerde Kaplan-Meier Yöntemi

Genelde biyomedikal çalışmalar, bir olay meydana gelene kadar geçen süreyle ilgilendirilir. Zaman, bütün bireyler olayı deneyimlediğinde diğer herhangi sürekli bir ölçüm gibi nitelendirilebilir. Bununla birlikte, olayı tüm hastalar yaşamazsa, sansürü açıklayan Kaplan-Meier ve Cox orantılı hazard yöntemleri kullanılmalıdır. Tüm gözlemler bağımsız olmadığında bir karmaşıklık ortaya çıkar. Örneğin, aynı kişi içinde tekrarlanan sağkalımlar veya aynı konudaki farklı olay türlerini içerir. Çoğunlukla bireyler, aynı anda aynı türden birden fazla müdahaleye sahip olabilir. Her gözlem benzersizdir ve kendi sağkalım veya sansür süresine sahip olsa da, belirli özne özellikleri sistemik olabilir ve bireyin tüm gözlemlerini etkileyebilir (27).

$i=1,2,\dots,m$  (hastalar),  $j=1,2,\dots, K_i$  ( $i$  inci hastadaki diş implantı) ve  $\delta_{ij}$  sansür göstergesi olmak üzere  $T_{ij}$  sağkalım süresini gösterebiliriz.  $Y_n(t) = \sum_{i=1}^m \sum_{j=1}^{K_i} I(T_{ij} \geq t)$  olmak üzere Ying ve Wei tarafından verilen sağkalım fonksiyonu için varyans tahmini

$$\sum_{i=1}^m \sum_{j=1}^{K_i} \sum_{l=1}^{K_i} \left[ \frac{I(T_{ij} \leq t) \delta_{ij}}{Y_n(T_{ij})} \sum_{i'=1}^m \sum_{j'=1}^{K_{i'}} \frac{I(T_{i'j'} \leq \min(T_{ij}, t) \delta_{i'j'})}{(Y_n(T_{i'j'}))^2} \right] \left[ \frac{I(T_{il} \leq t) \delta_{il}}{Y_n(T_{il})} \right. \\ \left. - \sum_{i'=1}^m \sum_{l'=1}^{K_{i'}} \frac{I(T_{i'l'} \leq \min(T_{il}, t) \delta_{i'l'})}{(Y_n(T_{i'l'}))^2} \right]$$

dir (28).

#### 2.4. Klasik Cox Modeli

Sağkalım analizinin temel amacı hazard fonksiyonları elde etmek ve sağkalımı açıklayan değişkenleri belirlemektir. Sağkalım modellenmesinde kullanılan klasik yöntemlerden biri Cox tarafından önerilen Cox regresyon modelidir. Sağkalım süresi için belirli bir olasılık dağılımı varsayımı olmadığı için yarı parametrik bir model olarak ele alınır (29).

Her bir bireyden,  $z_1, \dots, z_p$  değişkenlerinden bir veya daha fazla ölçümün mevcut olduğu varsayılın. Her bir j. birey için  $z'$  nin değeri  $z_i = z_{1j}, \dots, z_{pj}$  olsun ( $z'$  ler zaman fonksiyonu olabilir). Burada  $\beta$   $p \times 1$  vektörünün bilinmeyen bir parametresi ve  $\lambda_0(t)$ ,  $z = 0$  şartlarının standart kümesi için verilen hazard fonksiyonunun bilinmeyen bir fonksiyonu olmak üzere;

$$\lambda(t, z) = \lambda_0(t) \exp(z\beta)$$

dır. Aslında  $(z\beta)$  belli olan  $h(z, \beta)$  fonksiyonu ile yer değiştirebilir (30).

#### 2.5. Araştırma Konusuna İlişkin Literatür

Literatür incelendiğinde tezde kümelenmiş sağkalım süresi ile kümelenmiş başarısızlık süresi terimleri aynı anlamı karşıladığı için anlam bütünlüğü açısından sadece kümelenmiş sağkalım süresi olarak kullanılacaktır. Kümelenmiş sağkalım süresi verileri sağlık alanında geliştirme göstererek birçok araştırmada karşımıza çıkmaktadır.

Shih ve Louis kümelenmiş sağkalım verilerini analiz etmek için kırılabilirlik adı verilen rasgele bir etki kullanılarak orantılı hazard zayıflık modelinin yeterliliğinin değerlendirilmesi için grafiksel bir yöntem önermişlerdir (31).

Cai ve arkadaşları korelasyonlu sağkalım verilerini analiz etmek için istatistiksel yöntemler önermişlerdir. Ayrıca tahmin ve model seçiminde yeni yöntemleri göstermek



için alkolizmin genetik etiyolojisine ilişkin bir çalışmadan elde edilen verileri kullanmışlardır (32).

Jung ve Jeong kümelenmiş sağkalım süresi verilerinin analizi için marjinal sağkalım dağılımlarının eşitliğini test eden ve standart hataların hesaplanmasında küme-içi bağımlılığı dikkate alan iki örneklem sıra testlerini önermişlerdir. Yöntem ayrıca bağımlılık altında K-örneklem testleri için genişletilmiştir (33).

Chuang ve arkadaşları kümelenmiş gözlemlerin bağımlılığını varsayarak dış implantlarının sağkalım tahminleri için marjinal bir model kullanmışlardır. Ayrıca kümelenmiş gözlemlerin bağımsızlığını varsayan ikinci bir model geliştirmişleridir (34).

Glidden ve Vittinghoff kümelenmiş sağkalım süresi verilerinin analizi için gamma kırılma modelini kullanmış ve bu yaklaşımla önemli avantajlar elde etmişlerdir (20).

Lorino ve arkadaşları kümelenmiş sağkalım süresi verilerinin analiz eden ve temel varsayımlarından biri bağımsızlık olan Cox, marjinal Cox, sabit küme etkisi içeren Cox ve zayıf Cox modelini olmak üzere dört yöntem önermişlerdir (35).

Zeng ve arkadaşları diyabetik retinopati verileri ile kümelenmiş sağkalım süresi verilerini analiz etmek için rasgele etkilere sahip yarı parametrik dönüşüm modelleri önermişlerdir (11).

Zhang ve arkadaşları Kanadalı diyaliz hasta verilerini kullanarak kümelenmiş sağkalım verilerini analiz etmek için kümeler arasında ortak bir temel hazard ve ortak regresyon katsayısına sahip bir marjinal hazard modeli önermişleridir (36).

Mahmood ve arkadaşları çalışmalarında coğrafik olarak kümelenmiş sağkalım süresi verilerini analiz etmek için Cox'un orantılı hazard modeli ile rasgele etkiler modeli arasında karşılaştırmalar yapmanın yanı sıra, Bangladeşli kadınların doğum aralıklarının uzunluğunu etkileyebilecek önemli demografik ve sosyoekonomik faktörleri de göstermişlerdir (37).

Gauran ve Barrios kümelenmiş sağkalım süresi verilerini parametrik olmayan Cox orantılı hazard modeline rasgele kümeleme etkisini de ekleyerek analiz etmişlerdir (38).

Austin kümelenmiş sağkalım verilerini analiz etmek için karışık etkilere sahip Cox orantılı hazard modelleri olan kırılma modelleri, karışık etkilere sahip parçalı

üstel sağkalım modelleri ve karışık etkilere sahip ayrık zamanlı sağkalım modelleri olmak üzere üç yöntem önermiş ve kalp verisi üzerinde uygulamıştır (39).

Dahlqwist ve arkadaşları çalışmalarında kümelenmiş sağkalım verileri için kırılma modelleri (arasında-içinde) ile regresyon standartlaştırılması için yeni tahmin yöntemlerini önermişler ve bu yöntemleri erken doğum ve dikkat eksikliği hiperaktivite bozukluğu konusunda geniş bir kohortu analiz ederek göstermişlerdir (40).



### 3. MATERYAL VE METOT

#### 3.1. Hipotez

Kümelenmiş sağkalım verilerinde küme içi korelasyonu dikkate alan sağkalım yöntemleri ile her bir bireyden alınan birden fazla ölçümün oluşturacağı korelasyonu göz ardı eden (her bir ölçümü birbirinden bağımsızmış gibi analiz eden) klasik Cox modeli arasında bir farklılık yoktur.

Bu hipotezde, önerilen kümelenmiş sağkalım veri yapısına uygun küme korelasyonunu dikkate alan sağkalım yöntemlerinin sonuçları ışığında küme içi korelasyonu dikkate almayan klasik Cox yöntemi ile kıyaslamalar yapılması amaçlanmıştır.

#### 3.2. Benzetim çalışması

Kümelenmiş veri yapısını dikkate almayan Cox modeli ile bu tür verilerdeki bağımlılık yapısını dikkate alan yöntemlerin sonuçlarının karşılaştırılması için bir benzetim çalışması planlanmıştır.

Benzetim senaryosu; varsayımsal olarak üretilen değişkenler iki sınıflı sigara tüketimi (sigara içme durumu=1, içmeme durumu=0) ve yaş açıklayıcı değişkenleri ile tabakalı modeli uygulamak amacıyla cinsiyet (erkek=1, kadın=0) değişkenini içerecek şekilde tasarlanmıştır. Açıklanan yöntemleri kıyaslamak ve veri setlerini R programlama dilinde üretmek için “frailtysurv” paketinden “genfrail” fonksiyonu kullanılmıştır (41). “Genfrail” kümelenmiş sağkalım verilerini üretmek için kullanılan bir fonksiyondur. Tez kapsamında elde edilen sağkalım süresi gamma dağılımı ile üretilmiştir. Değişkenler üretilirken tezin amacına paralel olarak belli oranlarda üretilmiştir. Cinsiyet değişkeni %50’si kadın %50’ si erkek olmak üzere yarı-yarıya, sigara içme oranı Türkiye popülasyonunda ki yaklaşık içme oranı göz önüne alınarak %25 içme %75 içmeme olarak üretilmiştir (42). Ayrıca benzetim sonuçlarından elde edilen yaş değişkeni 19 ile 87 arasında değişmektedir. Tez kapsamında elde edilen veri setlerinde sağkalım oranları belirtilmedikçe yaklaşık %35 olacak şekilde üretilmiştir. Çalışmanın amacı; dişinde plak olan bireylerin dişlerinin çürümesi için gereken zamanın etkisinin açıklayıcı değişkenler ile modellenmesi ve kümelenmiş sağkalım modelleri ile klasik sağkalım modelini karşılaştırmaktır. Burada dişlerde çürük oluşması için gereken zaman sağkalım süresini, her bir birey ise küme etkisini göstermektedir.

Çalışma üç aşamalı olarak düşünülmüştür. İlk olarak gözlemlerin tam olduğu durumlar ele alınmıştır. Çalışma küme boyutları (k) 2 ve 4 olacak şekilde tasarlanmıştır. Çalışmanın ikinci aşamasında gözlemler %25 kayıplı olarak üretilmiştir. Maksimum küme boyutu 4 olarak alınmıştır. Her iki durum için birey sayısı (n) 50, 100, 200, 500 ve 1000' dir. Küme boyutu 2 iken toplam gözlem sayısı sırası ile 100, 200, 400, 1000 ve 2000 iken küme boyutu 4 olduğunda toplam gözlem sayısı sırasıyla 200, 400, 800, 2000 ve 4000' dir. Çalışmanın üçüncü aşaması ise iki kısımdan oluşmuştur. İlk aşaması k=2, n=100 ve sağkalım oranı %10 ile %70 arası, ikinci aşamasında ise k=4, n=500 ve sağkalım oranı %10 ile %70 arasında değiştiği durumları içermektedir. Buradaki amaç küçük ve büyük örneklerde sağkalım oranlarının değişmesi durumunda modeller arası performansları karşılaştırmaktır. Her bir bireyden alınan plaklı dişler arasındaki bağımlılığı göz önünde bulunduran ve koşullu olan kırılma modeli ve tabakalandırılmış Cox modeli ile marjinal modellerden marjinal Cox ve marjinal weibull modelleri klasik Cox modeli ile karşılaştırılmıştır.

### 3.3. Model Seçimi

İstatistiksel analizlerde en uygun modeli seçmenin bir çok kriter vardır. En yaygın kullanılanlar, bilgi ve olasılığa dayalı kriterleri içerir. Çalışmada kullanılan farklı modellerini ve karşılık gelen temel hazard fonksiyonlarını karşılaştırmak için bilgiye dayalı kriterler uygulanır. En çok kullanılan model seçme kriterleri; Akaike bilgi kriteri (*AIC*) (43) ve Bayesyen bilgi kriterleridir (*BIC*). *AIC* ve *BIC* *L* maksimum olabilirlik, *k* modeldeki parametre sayısı ve *n* toplam örnek büyüklüğü olmak üzere aşağıdaki gibidir (44).

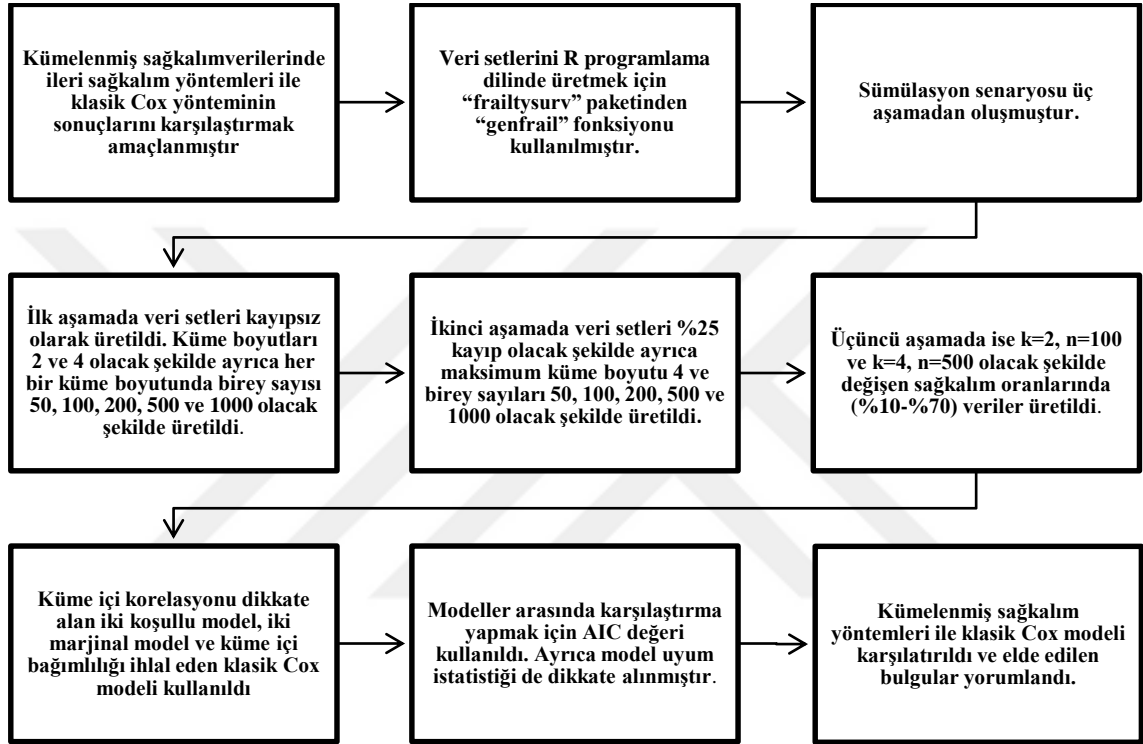
$$AIC = -2 \log(L) + 2k$$

$$BIC = -2 \log(L) + k \ln(n)$$

En küçük *AIC* ve *BIC* değerleri, modelin diğer modellere göre daha uygun olduğunu göstermektedir (45). Literatürde *AIC* ile örneklem büyüklüğünden bağımsız olarak daha karmaşık bir model seçileceği *BIC* ile ise daha basit bir model seçileceği öne sürülmektedir. (46). Ayrıca uyum istatistiği, modelin önce ilk olayı deneyimleyen kişiye daha büyük hazard atandığı tüm gözlem çiftlerinin oranıdır. % 70' in üzerinde bir uyum, herhangi bir sağkalım modeli için iyidir, bu nedenle modelin de bu seviyeye ulaşması güven vericidir (43). Literatürde kullanılan bir başka model seçim yöntemi de karekök

hata kareler ortalamasıdır (RMSE). Modelde en küçük değere sahip olan en uygun model olarak kabul edilir (45).

Bu tez kapsamında üretilen veriler ile uygulanan modeller arasında en iyi performansa sahip olan modeli bulmak için AIC değeri kullanılacaktır. Ayrıca model uyumu için uyum istatistiğine bakılacaktır.



Şekil 3.1. Benzetim Algoritması

## 4. BULGULAR

Tez kapsamında bulgular üç başlık altında verilmiştir. İlk kısımda küme büyüklüklerinin eşit fakat küme boyutlarının ve birey sayılarının değiştiği senaryolardır. İkinci kısımda küme büyüklüğünün eşit olmadığı ve birey sayılarının değiştiği senaryolardır. İlk iki durum için üretilen veriler %35 sağkalım oranları ile üretilmiştir. Üçüncü kısım ise; küme büyüklükleri, birey sayılarının farklı olduğu ve sağkalım oranlarının değiştiği durumlardır. Her üç aşama için üretilen bu veriler modellere uygulanmış ve elde edilen bulgular bağlamında tezin hipotezinin paralelinde yorumlanmıştır.

### 4.1. Küme büyüklüğünün eşit olduğu durumlar

#### 4.1.1. Küme boyutu k=2 ve n=50 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 50 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.1.'de, AIC ve uyum değerleri tablo 4.2' de verilmiştir.

**Tablo 4.1.** Küme büyüklüğü 2 ve birey sayısı 50 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=50		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılganlık Modeli	$\hat{\beta}_1$	9.129	0.109	0.003
	$\hat{\beta}_2$	1.100	0.909	0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	3.419	0.293	0.001
	$\hat{\beta}_2$	1.045	0.957	0.003
Marjinal Cox Modeli	$\hat{\beta}_1$	4.612	1.242	0.002
	$\hat{\beta}_2$	1.059	2.569	0.007
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.339	0.331	0.001
	$\hat{\beta}_2$	0.961	0.014	0.003
Klasik Cox Modeli	$\hat{\beta}_1$	3.453	0.289	0.001
	$\hat{\beta}_2$	1.046	0.956	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.2.** Küme büyüklüğü 2 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=50	AIC	Uyum
<b>Kırılgenlik Modeli</b>	184.280	%88
<b>Tabakalandırılmış Cox Modeli</b>	435.309	%67
<b>Marjinal Cox Modeli</b>	198.090	%74
<b>Marjinal Weibull Modeli</b>	750.784	%67
<b>Klasik Cox Modeli</b>	524.472	%67

Çalışmada ilk olarak küme boyutu  $k=2$ , birey sayısı  $n=50$  toplam gözlem sayısı 100 olarak alınmıştır. Bireylerin ağızlarındaki plaklı dişler arasında bağımlılığı dikkate alan kırılgenlik ve tabakalandırılmış Cox modeli olmak üzere iki koşullu model, marjinal Cox ve marjinal weibull modeli olmak üzere iki marjinal model kullanılmıştır. Ayrıca kümelenmiş veri yapısını göz ardı eden ve plaklı dişler arasındaki korelasyonu dikkate almayan klasik Cox modeli kullanılmıştır. Modellere ait hazard oranlarının beklenen değerleri, hazard oranlarının standart hatalarının beklenen değerlerini ve modele alınan açıklayıcı değişkenlerin anlamlılıklarını değerlendirmek için p değerleri ile modeller arasında en iyi modeli seçmek için AIC ve model uyumu için uyum istatistikleri verilmiştir.

Tablo 4.1' de verilen hazard oranı olan  $\exp(\hat{\beta})$  değeri ile önemli bulunan düzeyin, referans kategorisine göre kaç kat (ya da % ne kadar) daha riskli olduğu yorumu yapılmaktadır. Önemli bulunan değişkenlerin her bir düzeyine karşılık gelen p değerlerine bakılarak önemli değişken düzeyleri belirlenmektedir. İster küme içi bağımlılığı dikkate alan modeller olsun isterse klasik Cox modeli olsun her beş model içinde çalışmadaki açıklayıcı değişkenler anlamlıdır. Örneğin kırılgenlik modeli için sigara içen erkeklerde ve kadınlarda sigara içmeyenlere göre plaklı dişlerde çürük oluşma riski 9.129 kat daha fazladır. Aynı şekilde yaş ilerledikçe plaklarda çürük oluşma riski 1.100 kat daha fazladır. Diğer modellerdeki değişkenler için de benzer yorumlar yapılabilir.

Tablo 4.2' de verilen AIC değerleri tam model durumunda elde edilmiştir, bu da tüm ortak değişkenlerin modele dâhil edildiği anlamına gelmektedir. Kırılgenlik modeli diğer modeller arasında en düşük AIC değerine sahiptir (184.280). Kırılgenlik modeli, dikkate alınan diğer modeller arasında veriye en iyi uyan modeldir. Kırılgenlik modeline en yakın model ise marjinal Cox modelidir (AIC=198.090). AIC değerleri incelendiğinde en yüksek değere sahip model marjinal weibull modeli olmuştur (AIC=750.784). Yani

beş model arasında en kötü olan model marjinal weibull modeli olmuştur. AIC=524.472 ile klasik Cox modeli en kötü ikinci model olarak bulunmuştur. Ayrıca % 70' in üzerinde bir uyum, herhangi bir kümelenmiş sağkalım modeli için iyidir, bu nedenle kırılğanlık modelinin %88' e marjinal Cox modelin ise %74' e ulaşması güven vericidir (43).

#### 4.1.2. Küme boyutu k=2 ve n=100 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.3'de, AIC ve uyum değerleri tablo 4.4' de verilmiştir.

**Tablo 4.3.** Küme büyüklüğü 2 ve birey sayısı 100 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılğanlık Modeli</b>	$\hat{\beta}_1$	6.231	0.161	0.005
	$\hat{\beta}_2$	1.047	0.955	0.005
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.926	0.342	<0.001
	$\hat{\beta}_2$	1.042	0.959	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	4.190	0.238	0.015
	$\hat{\beta}_2$	1.030	0.970	0.031
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.387	1.275	<0.001
	$\hat{\beta}_2$	0.959	1.009	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.618	0.382	<0.001
	$\hat{\beta}_2$	1.043	0.959	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.4.** Küme büyüklüğü 2 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100	AIC	Uyum
<b>Kırılğanlık Modeli</b>	479.749	%85
<b>Tabakalandırılmış Cox Modeli</b>	1042.075	%67
<b>Marjinal Cox Modeli</b>	498.127	%65
<b>Marjinal Weibull Modeli</b>	1508.421	%67
<b>Klasik Cox Modeli</b>	1225.784	%67

Küme büyüklüğü 2 ve birey sayısı 100 olduğunda elde edilen istatistikler tablo 4.3 ve tablo 4.4 ' de verilmiştir. Her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Elde edilen bulgular incelendiğinde en iyi model küme içi



bağımlılığı dikkate alarak analiz yapan koşullu modellerden kırılğanlık modeli olmuştur (AIC=479.749). Kırılğanlık modeline en yakın olan model yine küme içi korelasyonu hesaba katan ve marjinal modellerden olan marjinal Cox modeli olmuştur (AIC=498.127). Uyum istatistiği incelendiğinde kırılğanlık modelinin uyumu %85 çıkmıştır. Tabloda verilen AIC değerleri incelendiğinde marjinal weibull modeli en kötü model olarak bulunmuştur (AIC=1508.421). Ayrıca klasik Cox modeli en kötü ikinci model olarak elde edilmiştir (1225.784).

#### 4.1.3. Küme boyutu k=2 ve n=200 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 200 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.5’de, AIC ve uyum değerleri tablo 4.6’ da verilmiştir.

**Tablo 4.5.** Küme büyüklüğü 2 ve birey sayısı 200 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=200		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılğanlık Modeli</b>	$\hat{\beta}_1$	5.093	0.196	<0.001
	$\hat{\beta}_2$	1.069	0.935	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.884	0.347	<0.001
	$\hat{\beta}_2$	1.032	0.969	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.564	0.390	0.001
	$\hat{\beta}_2$	1.037	0.964	0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.349	1.229	<0.001
	$\hat{\beta}_2$	0.969	1.007	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.906	0.344	<0.001
	$\hat{\beta}_2$	1.032	0.969	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.6.** Küme büyüklüğü 2 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=200	AIC	Uyum
<b>Kırılğanlık Modeli</b>	1097.245	%87
<b>Tabakalandırılmış Cox Modeli</b>	2428.536	%65
<b>Marjinal Cox Modeli</b>	1169.134	%66
<b>Marjinal Weibull Modeli</b>	2964.775	%65
<b>Klasik Cox Modeli</b>	2783.969	%65

Küme büyüklüğü 2 ve birey sayısı 200 toplamda gözlem sayısı 400 olduğunda elde edilen istatistikler tablo 4.5 ve 4.6' da verilmiştir. Tablo 4.5' de elde edilen bulgulara göre her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.6' daki AIC değerleri karşılaştırıldığında literatüre paralel olarak AIC değerleri incelenmiş ve en iyi model kırılmalı model olarak bulunmuştur (AIC=1097.245). Ayrıca uyum istatistiği incelendiğinde kırılmalı modelinin uyumu %87 bulunmuştur. Kırılmalı modeline en yakın model marjinal Cox modeli bulunmuştur (AIC=1169.134). Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü model olarak bulunmuştur (AIC=2964.775). Ede edilen bulgular bağlamında AIC=2783.969 ile klasik Cox modeli en kötü ikinci model olarak bulunmuştur.

#### 4.1.4. Küme boyutu k=2 ve n=500 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 500 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.7' de, AIC ve uyum değerleri tablo 4.8' de verilmiştir.

**Tablo 4.7.** Maksimum küme büyüklüğü 2 ve birey sayısı 500 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=500		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılmalı Modeli</b>	$\hat{\beta}_1$	6.371	0.157	<0.001
	$\hat{\beta}_2$	1.056	0.947	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	3.046	0.328	<0.001
	$\hat{\beta}_2$	1.031	0.970	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	3.371	0.297	<0.001
	$\hat{\beta}_2$	1.026	0.975	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.366	1.115	<0.001
	$\hat{\beta}_2$	0.974	1.004	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	3.037	0.329	<0.001
	$\hat{\beta}_2$	1.031	0.970	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.8.** Küme büyüklüğü 2 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=500	AIC	Uyum
<b>Kırılgnlık Modeli</b>	3529.684	%85
<b>Tabakalandırılmış Cox Modeli</b>	7315.127	%67
<b>Marjinal Cox Modeli</b>	3664.261	%67
<b>Marjinal Weibull Modeli</b>	7549.993	%67
<b>Klasik Cox Modeli</b>	8218.805	%67

Küme büyüklüğü 2 ve birey sayısı 500 olduğunda elde edilen istatistikler tablo 4.7 ve tablo 4.8’ de verilmiştir. Tablo 4.7’ de elde edilen bulgulara göre her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.8 incelendiğinde ise en uygun model en küçük AIC değerine sahip olan ve küme içi bağımlılığı dikkate alan koşullu modellerden kırılgnlık modeli olmuştur (AIC=3529.684). Elde edilen bulgular incelendiğinde kırılgnlık modeli %85 uyuma sahiptir. Kırılgnlık modeline en yakın model AIC=3664.261 ile yine küme içi bağımlılığı dikkate alan marjinal Cox modeli olmuştur. Beş model incelendiğinde modeller arasında en kötü model bireylerin dışleri arasındaki küme içi korelasyonu göz ardı eden ve her bir bireyden alınan dışleri bağımsız kabul eden klasik Cox modeli olmuştur (AIC=8218.805). Ayrıca AIC=7549.993 ile marjinal weibull modeli en kötü ikinci model olarak elde edilmiştir.

#### 4.1.5. Küme boyutu k=2 ve n=1000 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 1000 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.9’ da, AIC ve uyum değerleri tablo 4.10’ da verilmiştir.

**Tablo 4.9.** Maksimum küme büyüklüğü 2 ve birey sayısı 1000 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=1000		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılganlık Modeli</b>	$\hat{\beta}_1$	3.749	0.267	<0.001
	$\hat{\beta}_2$	1.055	0.948	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.437	0.410	<0.001
	$\hat{\beta}_2$	1.035	0.966	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.416	0.414	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.439	1.080	<0.001
	$\hat{\beta}_2$	0.969	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.432	0.411	<0.001
	$\hat{\beta}_2$	1.035	0.966	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

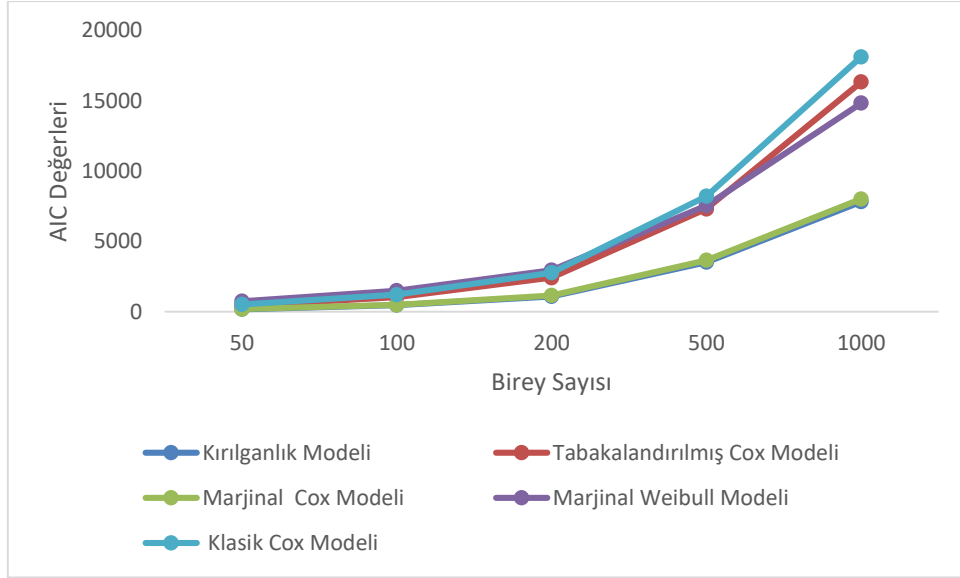
\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.10.** Küme büyüklüğü 2 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=1000	AIC	Uyum
<b>Kırılganlık Modeli</b>	7831.979	%84
<b>Tabakalandırılmış Cox Modeli</b>	16337.310	%66
<b>Marjinal Cox Modeli</b>	8029.007	%66
<b>Marjinal Weibull Modeli</b>	14841.57	%66
<b>Klasik Cox Modeli</b>	18129.300	%66

Küme büyüklüğü 2 ve birey sayısı 1000 olduğunda elde edilen istatistikler tablo 4.9 ve 4.10' da verilmiştir. Tablo 4.9' da elde edilen bulgulara göre her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.10 incelendiğinde ise en iyi model en küçük AIC değerine sahip olan ve küme içi korelasyonu göz önüne alarak analiz eden kırılganlık modeli olmuştur (AIC=7831.979). En kötü model ise küme içi bağımlılığı göz ardı eden klasik Cox modeli olmuştur (AIC=18129.300). Ayrıca uyum istatistiği incelendiğinde kırılganlık modelinin uyumu %84 olarak bulunmuştur. Küme büyüklüğü arttıkça diğer tablolardan farklı olarak klasik Cox modelinden sonra en kötü performans gösteren model tabakalandırılmış Cox modeli olmuştur. (AIC=16337.31)

k=2 ve n=50, 100, 200, 500 ve 1000 için beş modelden elde edilen AIC değerlerini karşılaştıran grafik aşağıda verilmiştir.



**Grafik 4.1.** Maksimum küme büyüklüğü 2 iken yöntemlere göre AIC değerleri

#### 4.1.6. Küme boyutu k=4 ve n=50 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 50 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $\exp(\text{se}(\hat{\beta}))$  ve p değerleri Tablo 4.11’de, AIC ve uyum değerleri tablo 4.12’ de verilmiştir.

**Tablo 4.11.** Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $\text{se}(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=50		$\exp(\hat{\beta})$	$\text{se}(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgnlık Modeli</b>	$\hat{\beta}_1$	4.777	0.209	0.001
	$\hat{\beta}_2$	1.050	0.952	0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.070	0.483	0.001
	$\hat{\beta}_2$	1.037	0.965	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	3.311	0.302	0.001
	$\hat{\beta}_2$	1.051	0.952	0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.611	1.236	0.020
	$\hat{\beta}_2$	0.974	1.008	0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.872	0.534	0.003
	$\hat{\beta}_2$	1.033	0.968	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.12.** Küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=50	AIC	Uyum
Kırılgenlik Modeli	538.146	%80
Tabakalandırılmış Cox Modeli	1055.887	%64
Marjinal Cox Modeli	560.898	%68
Marjinal Weibull Modeli	1497.193	%62
Klasik Cox Modeli	1243.159	%63

Tablo 4.11 ve tablo 4.12’ de küme boyutu 4 birey sayısı 50 ve toplam gözlem sayısı 200 iken elde edilen istatistikler verilmiştir. Tablo 4.11’ de elde edilen bulgular incelendiğinde tüm modellere dâhil edilen açıklayıcı değişkenler anlamlıdır. Literatüre göre en düşük AIC değerine sahip olan modelin en iyi modeli olmaktadır. Tablo 4.1.12 incelendiğinde en düşük AIC değerine sahip kırılgenlik modelidir (AIC=538.146). Ayrıca kırılgenlik modelinin uyumu %80 olarak bulunmuştur. Kırılgenlik modeline en yakın model AIC= 560.898 değeri ile yine küme içi bağımlılığı dikkate alan marjinal Cox modeli olmuştur. Modeller arasında en kötü olan model ise en yüksek AIC=1497.193 değeri ile marjinal weibull olmuştur. Elde edilen bulgular kapsamında en kötü ikinci model AIC=1243.159 ile klasik Cox modeli olmuştur.

#### 4.1.7. Küme boyutu k=4 ve n=100 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 100 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.13’de, AIC ve uyum değerleri tablo 4.14’ de verilmiştir.

**Tablo 4.13.** Küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=100		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılgenlik Modeli	$\hat{\beta}_1$	2.924	0.342	<0.001
	$\hat{\beta}_2$	1.054	0.949	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	3.253	0.307	<0.001
	$\hat{\beta}_2$	1.042	0.959	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	2.471	0.405	0.003
	$\hat{\beta}_2$	1.051	0.952	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.377	1.234	<0.001
	$\hat{\beta}_2$	0.967	1.006	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	3.263	0.306	<0.001
	$\hat{\beta}_2$	1.043	0.959	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.14.** Küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=100	AIC	Uyum
<b>Kırılgnalık Modeli</b>	1170.334	%79
<b>Tabakalandırılmış Cox Modeli</b>	2435.817	%67
<b>Marjinal Cox Modeli</b>	1202.562	%66
<b>Marjinal Weibull Modeli</b>	2973.979	%67
<b>Klasik Cox Modeli</b>	2797.428	%67

Tablo 4.13 ve tablo 4.14’ de küme büyüklüğü 4 birey sayısı 100 iken elde edilen istatistikler verilmiştir. Tablo 4.13’de elde edilen bulgulara göre modele dâhil edilen açıklayıcı değişkenler her beş model için de anlamlı bulunmuştur. Tablo 4.14’ de tam model durumunda elde edilen bulgular verilmiştir. Elde edilen bu bulgular bağlamında en küçük AIC=1170.334 değerine sahip kırılgnalık modeli diğer dört modele göre en iyi model olarak bulunmuştur. Uyum istatistiği incelendiğinde ise kırılgnalık modelinin uyumu %79’ dur. Çalışma kapsamına dâhil edilen modeller arasında en yüksek AIC değerine sahip marjinal weibull modeli en kötü model olarak bulunmuştur (AIC=2973.979). Ayrıca AIC=2797.428 değeri ile klasik Cox modeli en kötü ikinci model olarak elde edilmiştir.

#### 4.1.8. Küme boyutu k=4 ve n=200 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 200 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.15’de, AIC ve uyum değerleri tablo 4.16’ da verilmiştir.

**Tablo 4.15.** Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=200		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgnalık Modeli</b>	$\hat{\beta}_1$	2.016	0.496	<0.001
	$\hat{\beta}_2$	1.046	0.956	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	1.639	0.610	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.570	0.637	0.029
	$\hat{\beta}_2$	1.034	0.967	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.619	1.174	0.003
	$\hat{\beta}_2$	0.967	1.005	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.629	0.614	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.16.** Küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=200	AIC	Uyum
<b>Kırılgenlik Modeli</b>	2735.813	%80
<b>Tabakalandırılmış Cox Modeli</b>	5653.391	%62
<b>Marjinal Cox Modeli</b>	2866.302	%62
<b>Marjinal Weibull Modeli</b>	6072.858	%62
<b>Klasik Cox Modeli</b>	6368.357	%62

Küme büyüklüğü 4 ve birey sayısı 200 olduğunda elde edilen bulgular tablo 4.15 ve tablo 4.16’ da verilmiştir. Tablo 4.15 ‘ de her beş model içinde modele dâhil edilen açıklayıcı değişkenler anlamlı olarak bulunmuştur. Tablo 4.16’ da elde edilen bulgulara göre en iyi model kırılgenlik modeli olmuştur (AIC=2735.813). Ayrıca AIC=2866.302 değeri ile en iyi ikinci model marjinal Cox modeli olmuştur. Her beş modelin AIC incelendiğinde en yüksek değere sahip model klasik Cox modeli olmuştur (AIC=6368.357). En kötü ikinci model ise marjinal weibull olmuştur (AIC=6072.858). Ayrıca kırılgenlik modelinin uyumu %80 olarak elde edilmiştir.

#### 4.1.9. Küme boyutu k=4 ve n=500 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.17’de, AIC ve uyum değerleri tablo 4.18’ de verilmiştir.

**Tablo 4.17.** Küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgenlik Modeli</b>	$\hat{\beta}_1$	2.019	0.495	<0.001
	$\hat{\beta}_2$	1.062	0.942	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	1.541	0.649	<0.001
	$\hat{\beta}_2$	1.033	0.968	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.551	0.645	0.0011
	$\hat{\beta}_2$	1.035	0.967	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.674	1.094	<0.001
	$\hat{\beta}_2$	0.970	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.541	0.649	<0.001
	$\hat{\beta}_2$	1.033	0.968	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı



**Tablo 4.18.** Küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500	AIC	Uyum
Kırılgenlik Modeli	7911.610	%81
Tabakalandırılmış Cox Modeli	16527.180	%62
Marjinal Cox Modeli	8212.043	%63
Marjinal Weibull Modeli	15178.540	%62
Klasik Cox Modeli	18322.730	%62

Küme büyüklüğü 4 ve birey sayısı 500 ve toplam gözlem sayısı 2000 olduğunda elde edilen bulgular tablo 4.17 ve tablo 4.18’ de verilmiştir. Tablo 4.17’ ye bakıldığında her beş model içinde modele dâhil edilen açıklayıcı değişkenler anlamlı olarak bulunmuştur. Tablo 4.18 incelendiğinde ise en düşük AIC değerine sahip model kırılgenlik modeli olarak bulunmuştur (AIC=7911.610). Ayrıca kırılgenlik modelinin uyumu %81 bulunmuştur. AIC=8212.043 değeri ile marjinal Cox modeli en iyi ikinci model olarak bulunmuştur. Küme içi korelasyonu göz ardı eden klasik Cox modeli AIC=18322.730 ile en kötü model olarak bulunmuştur. Yine elde edilen bulguların paralelinde küme boyutunun büyüklüğü ile orantılı olarak tabakalandırılmış Cox modeli en kötü performans gösteren ikinci model olmuştur. (AIC=16527.180)

#### 4.1.10. Küme boyutu k=4 ve n=1000 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.19’ da, AIC ve uyum değerleri tablo 4.20’ de verilmiştir.

**Tablo 4.19.** Küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=1000		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılgenlik Modeli	$\hat{\beta}_1$	2.796	0.358	<0.001
	$\hat{\beta}_2$	1.059	0.944	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	1.823	0.549	<0.001
	$\hat{\beta}_2$	1.037	0.965	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	2.027	0.493	<0.001
	$\hat{\beta}_2$	1.035	0.966	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.573	1.068	<0.001
	$\hat{\beta}_2$	0.967	1.002	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	1.822	0.549	<0.001
	$\hat{\beta}_2$	1.037	0.965	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

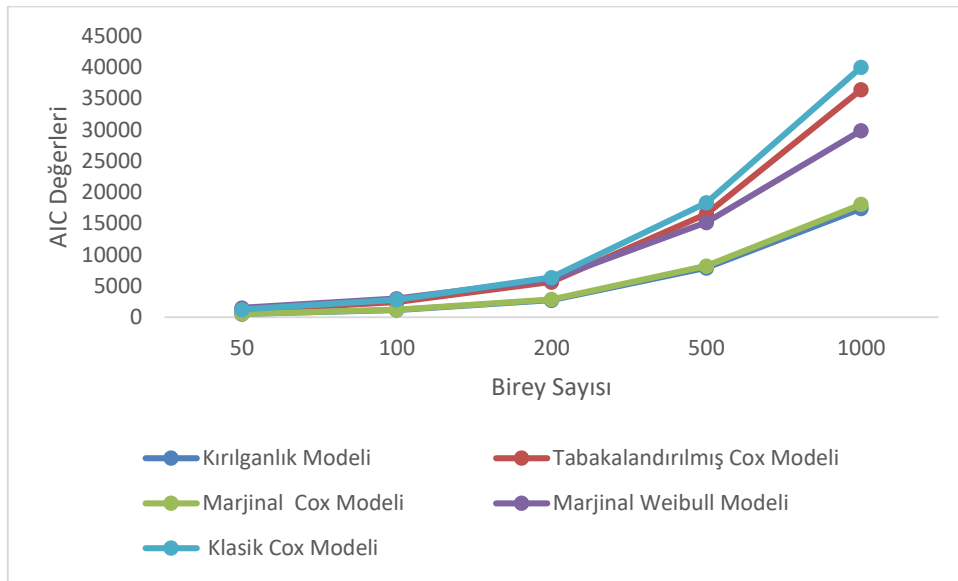
\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.20.** Küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=1000	AIC	Uyum
<b>Kırılğanlık Modeli</b>	17406.310	%82
<b>Tabakalandırılmış Cox Modeli</b>	36424.980	%64
<b>Marjinal Cox Modeli</b>	18074.020	%64
<b>Marjinal Weibull Modeli</b>	29874.980	%64
<b>Klasik Cox Modeli</b>	40003.860	%64

Küme büyüklüğü 4 ve birey sayısı 1000 olduğunda elde edilen bulgular tablo 4.19 ve tablo 4.20’ de verilmiştir. Tablo 4.19’ ye bakıldığında her beş model içinde modele dâhil edilen açıklayıcı değişkenler anlamlı olarak bulunmuştur. Tablo 4.20’ deki bulgular incelendiğinde ise kırılğanlık modeli AIC=17406.310 en düşük değeri ile en iyi model olarak bulunmuştur. Kırılğanlık modelinden sonra en iyi model marjinal Cox modeli olarak bulunmuştur (AIC=18074.020). AIC değerleri incelendiğinde en kötü model en yüksek AIC değerine sahip klasik Cox modelidir (AIC=40003.860). Tabakalandırılmış Cox modeli AIC=36424.980 ile en kötü ikinci model olarak bulunmuştur. Ayrıca tablo 4.20’ deki uyum istatistiği incelendiğinde kırılğanlık modeli %82 ile güven vericidir.

k=4 ve n=50, 100, 200, 500 ve 1000 için beş modelden elde edilen AIC değerlerini karşılaştıran grafik aşağıda verilmiştir.



**Grafik 4.2.** Maksimum küme büyüklüğü 4 iken yöntemlere göre AIC değerleri

## 4.2. Küme boyutunun eşit olmadığı durumlar

Bu aşamada verinin %75'nin küme boyutu 4 olarak alınırken %25'nin küme boyutu 2 ve birey sayıları 50, 100, 200, 500 ve 1000 olarak alınmıştır. Buradaki amaç küme boyutlarının dengesiz olduğu durumlarda modeller arasındaki performansları karşılaştırmaktır. Bu bağlamda elde edilen bulgulara ait tablolar ve yorumları aşağıdadır.

### 4.2.1. Maksimum küme boyutu k=4 ve n=50 iken elde edilen bulgular

Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.21'de, AIC ve uyum değerleri tablo 4.22'de verilmiştir.

**Tablo 4.21.** Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=50		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılganlık Modeli	$\hat{\beta}_1$	4.549	0.219	0.001
	$\hat{\beta}_2$	1.047	0.955	0.005
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	2.271	0.440	0.002
	$\hat{\beta}_2$	1.034	0.967	0.002
Marjinal Cox Modeli	$\hat{\beta}_1$	2.746	0.364	0.018
	$\hat{\beta}_2$	1.030	0.971	0.013
Marjinal Weibull Modeli	$\hat{\beta}_1$	-0.764	1.346	0.010
	$\hat{\beta}_2$	-0.028	1.009	0.002
Klasik Cox Modeli	$\hat{\beta}_1$	2.243	0.446	0.008
	$\hat{\beta}_2$	1.034	0.967	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.22.** Maksimum küme büyüklüğü 4 ve birey sayısı 50 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=50	AIC	Uyum
Kırılganlık Modeli	376.074	%79
Tabakalandırılmış Cox Modeli	721.891	%67
Marjinal Cox Modeli	391.088	%68
Marjinal Weibull Modeli	1129.508	%66
Klasik Cox Modeli	853.814	%67

Tablo 4.21'de maksimum küme büyüklüğü 4, birey sayısı 50 iken elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri değerleri verilmiştir. Bu bulguların sonucu

incelendiğinde her beş model için kullanılan açıklayıcı değişkenlerin anlamlı olduğu sonucuna varılmıştır. Ayrıca tablo 4.22 incelendiğinde en düşük AIC değerine sahip olan model kırılma modelidir (AIC=376.074). Marjinal Cox modeli AIC=391.088 ile en iyi ikinci yöntem olarak bulunmuştur. En yüksek AIC=1129.508 değeri ile en kötü olan model marjinal weibull modeli olarak elde edilmiştir. En kötü ikinci model ise AIC=853.814 ile klasik Cox modeli olmuştur. Ayrıca uyum kriterine bakıldığında kırılma modelinin %79 uyum göstermiştir.

#### 4.2.2. Maksimum küme boyutu k=4 ve n=100 iken elde edilen bulgular

Maksimum küme büyüklüğü 4 ve birey sayısı 100 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.23’ de, AIC ve uyum değerleri tablo 4.24’ de verilmiştir.

**Tablo 4.23.** Maksimum küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=100		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılma Modeli	$\hat{\beta}_1$	2.178	0.459	<0.001
	$\hat{\beta}_2$	1.060	0.943	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	1.618	0.618	0.006
	$\hat{\beta}_2$	1.038	0.963	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	1.941	0.515	0.021
	$\hat{\beta}_2$	1.053	0.950	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.680	1.189	0.026
	$\hat{\beta}_2$	0.970	1.006	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	1.571	0.637	0.009
	$\hat{\beta}_2$	1.038	0.964	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.24.** Maksimum küme büyüklüğü 4 ve birey sayısı 100 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=100	AIC	Uyum
Kırılma Modeli	859.891	%76
Tabakalandırılmış Cox Modeli	1755.180	%62
Marjinal Cox Modeli	876.820	%65
Marjinal Weibull Modeli	2233.600	%62
Klasik Cox Modeli	2028.490	%62

Maksimum küme büyüklüğü 4, birey sayısı 100 iken elde edilen bulgular tablo 4.23 ve tablo 4.24’ de verilmiştir. Tablo 4.23’ de verilen bulgular bağlamında açıklayıcı değişkenler anlamlı bulunmuştur. En iyi modeli bulmak için AIC kriteri değerlendirilmiş ve model uyum istatistiği için tablo 4.24’ e bakılmıştır. En küçük AIC değerine sahip olan kırılmalı model en iyi model olarak bulunmuştur (AIC=859.891). Marjinal Cox modeli en küçük ikinci AIC= 876.820 değeri ile en iyi performansa sahip ikinci model olarak elde edilmiştir. En kötü model ise en yüksek AIC=2233.600 değeri ile marjinal weibull modeli olmuştur. Elde edilen bulgular ışığında en kötü ikinci model klasik Cox modeli olmuştur (AIC=2028.490). Ayrıca kırılmalı modelinin uyumu %76 olarak elde edilmiştir.

#### 4.2.3. Maksimum küme boyutu k=4 ve n=200 iken elde edilen bulgular

Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $\exp(\text{se}(\hat{\beta}))$  ve p değerleri Tablo 4.25’ de, AIC ve uyum değerleri tablo 4.26’ da verilmiştir.

**Tablo 4.25.** Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $\text{se}(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=200		$\exp(\hat{\beta})$	$\text{se}(\exp(\hat{\beta}))$	p-değeri
<b>Kırılmalı Modeli</b>	$\hat{\beta}_1$	2.551	0.392	<0.001
	$\hat{\beta}_2$	1.039	0.962	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.121	0.472	<0.001
	$\hat{\beta}_2$	1.036	0.965	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.920	0.521	0.017
	$\hat{\beta}_2$	1.030	0.971	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.498	1.197	0.0001
	$\hat{\beta}_2$	0.970	1.005	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.149	0.465	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.26.** Maksimum küme büyüklüğü 4 ve birey sayısı 200 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=200	AIC	Uyum
<b>Kırılgenlik Modeli</b>	1820.332	%80
<b>Tabakalandırılmış Cox Modeli</b>	3985.272	%64
<b>Marjinal Cox Modeli</b>	1895.038	%64
<b>Marjinal Weibull Modeli</b>	4512.319	%64
<b>Klasik Cox Modeli</b>	4527.019	%64

Tablo 4.25 ve Tablo 4.26’ da maksimum küme büyüklüğü 4, birey sayısı 200 ve toplam gözlem sayısı 800 iken elde edilen bulgular verilmiştir. Tablo 4.25’ de  $\exp(\hat{\beta})$ ,  $\exp(\text{se}(\hat{\beta}))$  ve p değerleri ve tablo 4.26’ da AIC ve uyum istatistiği verilmişti. Tablo 4.25’ de elde edilen bulgular incelendiğinde modellere dâhil edilen tüm açıklayıcı değişkenler anlamlı çıkmıştır. Ayrıca Tablo 4.26’ daki AIC değerleri incelendiğinde en düşük AIC=1820.332 ile en iyi model kırılgenlik modelidir. En iyi ikinci model ise bireyler arasındaki bağımlılığı dikkate alan marjinal Cox modeli olmuştur (AIC=1895.038). Küme içi korelasyonu göz ardı eden Klasik Cox modeli en yüksek AIC=4527.019 ile en kötü model olarak bulunmuştur. Ayrıca AIC=4512.319 ile marjinal weibull modeli en kötü ikinci model olarak elde edilmiştir. İlaveten kırılgenlik modelinin uyumu %80’ dir.

#### 4.2.4. Maksimum küme boyutu k=4 ve n=500 iken elde edilen bulgular

Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $\text{se}(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.27’ de, AIC ve uyum değerleri tablo 4.28’ de verilmiştir.

**Tablo 4.27.** Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $\text{se}(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500		$\exp(\hat{\beta})$	$\text{se}(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgenlik Modeli</b>	$\hat{\beta}_1$	1.849	0.541	<0.001
	$\hat{\beta}_2$	1.052	0.950	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	1.659	0.603	<0.001
	$\hat{\beta}_2$	1.036	0.9648	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.541	0.649	0.004
	$\hat{\beta}_2$	1.034	0.967	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.638	1.104	<0.001
	$\hat{\beta}_2$	0.968	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.637	0.611	<0.001
	$\hat{\beta}_2$	1.036	0.965	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.28.** Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500	AIC	Uyum
<b>Kırılgnlık Modeli</b>	5536.901	%78
<b>Tabakalandırılmış Cox Modeli</b>	11774.620	%63
<b>Marjinal Cox Modeli</b>	5678.961	%63
<b>Marjinal Weibull Modeli</b>	11366.490	%63
<b>Klasik Cox Modeli</b>	13123.000	%63

Maksimum küme büyüklüğü 4 ve birey sayısı 500 iken yöntemlere göre elde edilen bulgular Tablo 4.27 ve 4.28’ de verilmiştir. Tablo 4.27’ de elde edilen sonuçlar incelendiğinde modellere alınan tüm değişkenler anlamlı bulunmuştur. Tablo 4.28 ‘e göre en düşük AIC=5536.901 değerine sahip kırılgnlık modeli en iyi model bulunmuştur. Marjinal Cox modeli AIC=5678.961 ile ikinci en iyi model olarak bulunmuştur. En yüksek AIC=13123.000 değeri ile en kötü model klasik Cox modeli olarak bulunmuştur. Ayrıca AIC=11774.620 değeri ile tabakalandırılmış Cox modeli en kötü ikinci model olarak elde edilmiştir. Kırılgnlık modelinin uyumu %78 olarak elde edilmiştir.

#### 4.2.5. Maksimum küme boyutu k=4 ve n=1000 iken elde edilen bulgular

Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.29’ da, AIC ve uyum değerleri Tablo 4.30’ da verilmiştir.

**Tablo 4.29.** Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=1000		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgnlık Modeli</b>	$\hat{\beta}_1$	2.897	0.345	<0.001
	$\hat{\beta}_2$	1.048	0.954	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.163	0.462	<0.001
	$\hat{\beta}_2$	1.035	0.966	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.338	0.428	<0.001
	$\hat{\beta}_2$	1.036	0.966	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.493	1.078	<0.001
	$\hat{\beta}_2$	0.969	1.002	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.167	0.461	<0.001
	$\hat{\beta}_2$	1.035	0.966	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

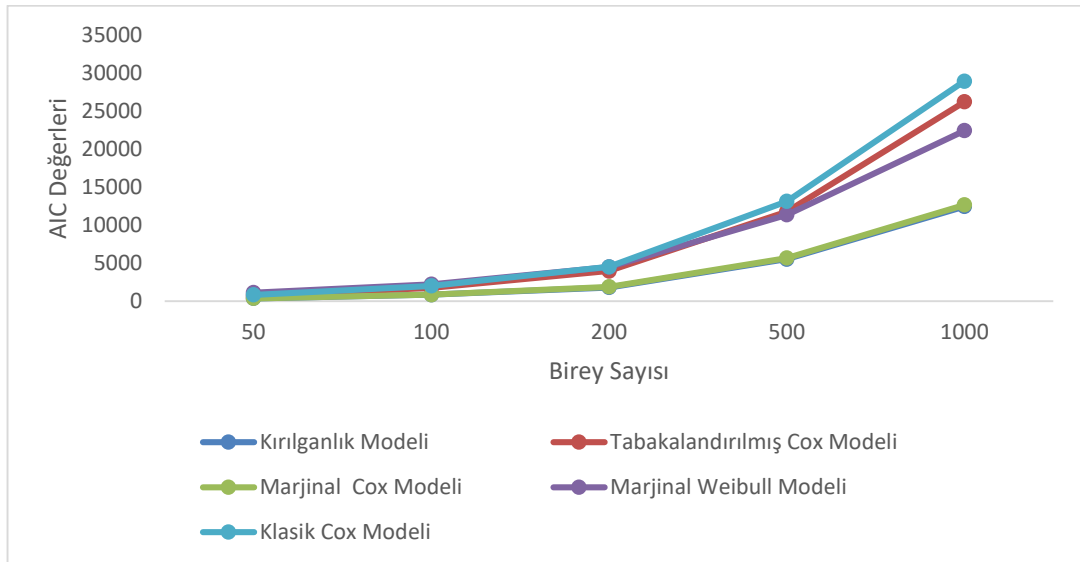
\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.30.** Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=1000	AIC	Uyum
<b>Kırılgenlik Modeli</b>	12461.010	%77
<b>Tabakalandırılmış Cox Modeli</b>	26225.680	%64
<b>Marjinal Cox Modeli</b>	12659.690	%65
<b>Marjinal Weibull Modeli</b>	22438.370	%64
<b>Klasik Cox Modeli</b>	28921.220	%64

Maksimum küme büyüklüğü 4 ve birey sayısı 1000 iken elde edilen bulgular tablo 4.29 ve 4.30' da verilmiştir. Tablo 4.29 incelendiğinde modele alınan tüm açıklayıcı değişkenler anlamlı olarak bulunmuştur. Tablo 4.30 incelendiğinde ise en uygun modeli seçmek için AIC kriterleri incelenmiş ve en iyi model en küçük AIC=12461.010 değerine sahip olan kırılgenlik modeli olmuştur. Ayrıca en iyi ikinci model ise AIC=12659.690 ile marjinal Cox modeli olmuştur. En kötü model ise en yüksek AIC=28921.220 değerine sahip olan klasik Cox modeli olmuştur. İkinci en kötü model ise tabakalandırılmış Cox modeli olarak elde edilmiştir. (AIC=26225.68). İlâveten uyum istatistiği incelendiğinde kırılgenlik modeli %77 ile güven vericidir.

k=4 ve n=50, 100, 200, 500 ve 1000 ayrıca verilerde beş modelden elde edilen AIC değerlerini karşılaştıran grafik aşağıda verilmiştir.



**Grafik 4.3** Maksimum küme büyüklüğü 4 ve kayıp oranı %25 iken yöntemlere göre AIC değerleri



### 4.3. Farklı sağkalım oranlarında testlerin performanslarının karşılaştırılması

Yapılan benzetim çalışmasının üçüncü aşaması iki kısımdan oluşmaktadır. İlk kısımda  $k=2$ ,  $n=100$  olduğu durumlarda değişen sağkalım oranlarında (%10-%70), ikinci kısımda ise örneklem büyüklüğü artırılıp  $k=4$ ,  $n=500$  ve yine sabit olmayan sağkalım oranlarında (%10-%70), veriler üretilmiş ve bu veriler yöntemlere uygulanarak elde edilen bulgular yorumlanmıştır.

#### 4.3.1. Küme boyutu $k=2$ ve $n=100$ ve sağkalım oranı % 10 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %10 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.31' de, AIC ve uyum değerleri tablo 4.32' de verilmiştir.

**Tablo 4.31.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı % 10 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

$k=2$ $n=100$ sağkalım oranı %10		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılganlık Modeli	$\hat{\beta}_1$	4.706	0.213	<0.001
	$\hat{\beta}_2$	1.057	0.946	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	2.324	0.430	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	2.380	0.420	0.005
	$\hat{\beta}_2$	1.037	0.965	0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.241	0.403	0.001
	$\hat{\beta}_2$	0.009	0.963	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	2.258	0.443	<0.001
	$\hat{\beta}_2$	1.034	0.967	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.32.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %10 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %10	AIC	Uyum
<b>Kırılganlık Modeli</b>	628.587	%84
<b>Tabakalandırılmış Cox Modeli</b>	1356.528	%68
<b>Marjinal Cox Modeli</b>	667.430	%68
<b>Marjinal Weibull Modeli</b>	2165.475	%68
<b>Klasik Cox Modeli</b>	1606.385	%68

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %10 olduğunda elde edilen istatistikler tablo 4.31 ve 4.32’de verilmiştir. Tablo 4.31’ de elde edilen bulgulara göre her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.32’ deki AIC değerleri karşılaştırıldığında kırılganlık modeli AIC=628.587 değeri ile en iyi model olarak bulunmuştur. Ayrıca uyum istatistiği incelendiğinde kırılganlık modelinin uyumu %84 bulunmuştur. Kırılganlık modeline en yakın model marjinal Cox modeli bulunmuştur (AIC=667.430). Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü model olarak bulunmuştur (AIC=2165.475). Ede edilen bulgular bağlamında AIC=1606.385 ile klasik Cox modeli en kötü ikinci model olarak bulunmuştur.

#### **4.3.2. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 20 iken elde edilen bulgular**

Maksimum küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %20 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.33’ de, AIC ve uyum değerleri tablo 4.34’ de verilmiştir.

**Tablo 4.33.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %20 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %20		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılğanlık Modeli</b>	$\hat{\beta}_1$	6.531	0.153	<0.001
	$\hat{\beta}_2$	1.056	0.947	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.352	0.425	<0.001
	$\hat{\beta}_2$	1.036	0.966	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	3.382	0.296	0.006
	$\hat{\beta}_2$	1.027	0.974	0.005
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.308	0.365	<0.001
	$\hat{\beta}_2$	0.009	0.959	0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.355	0.425	<0.001
	$\hat{\beta}_2$	1.037	0.965	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.34.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %20 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %20	AIC	Uyum
<b>Kırılğanlık Modeli</b>	586.148	%84
<b>Tabakalandırılmış Cox Modeli</b>	1246.459	%66
<b>Marjinal Cox Modeli</b>	625.208	%65
<b>Marjinal Weibull Modeli</b>	1875.655	%66
<b>Klasik Cox Modeli</b>	1464.416	%66

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %20 olduğunda elde edilen istatistikler tablo 4.33 ve 4.34' de verilmiştir. Tablo 4.33' de elde edilen bulgulara göre her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.34' deki AIC değerleri karşılaştırıldığında kırılğanlık modeli AIC=586.148 değeri ile en iyi model olarak bulunmuştur. Ayrıca uyum istatistiği incelendiğinde kırılğanlık modelinin uyumu %84 bulunmuştur. Marjinal Cox model en düşük ikinci AIC değeri ile model performansı olarak ikinci en iyi model olmuştur. (AIC=625.208). Model performansı açısından en kötü model en yüksek AIC değerine sahip olan marjinal weibull modelidir (AIC=1875.655). Ede edilen bulgular sonucunda AIC=1464.416 ile klasik Cox modeli en kötü ikinci model olarak bulunmuştur.

### 4.3.3. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 30 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %30 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.35' de, AIC ve uyum değerleri tablo 4.36' da verilmiştir.

**Tablo 4.35.** Küme küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %30 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %30		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılganlık Modeli	$\hat{\beta}_1$	4.545	0.220	<0.001
	$\hat{\beta}_2$	1.067	0.937	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	2.082	0.480	0.001
	$\hat{\beta}_2$	1.031	0.970	0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	3.030	0.330	0.001
	$\hat{\beta}_2$	1.044	0.958	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.561	1.243	0.008
	$\hat{\beta}_2$	0.975	1.007	0.001
Klasik Cox Modeli	$\hat{\beta}_1$	1.971	0.507	0.001
	$\hat{\beta}_2$	1.030	0.971	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.36.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %30 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %30	AIC	Uyum
Kırılganlık Modeli	529.781	%83
Tabakalandırılmış Cox Modeli	1123.878	%62
Marjinal Cox Modeli	549.284	%70
Marjinal Weibull Modeli	1602.074	%63
Klasik Cox Modeli	1318.262	%63

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %30 olduğunda elde edilen istatistikler tablo 4.35 ve 4.36.'da verilmiştir. Tablo 4.35' de elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.36.'daki AIC değerlerine göre AIC=529.781 değeri ile kırılganlık modeli en iyi model olarak bulunmuştur. Marjinal Cox modeli AIC=549.284 ile en iyi performansa sahip ikinci

model olarak elde edilmiştir. Ayrıca uyum istatistiği incelendiğinde kırılmalılık modelinin uyumu %83 ve marjinal Cox modelinin uyumu %70 olarak bulunmuştur. Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=1602.074). Klasik Cox modeli ise AIC=1318.262 ile en kötü ikinci model olarak bulunmuştur.

#### 4.3.4. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 40 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %40 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.37.' de, AIC ve uyum değerleri tablo 4.38.' de verilmiştir.

**Tablo 4.37.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %40 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %40		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılmalılık Modeli	$\hat{\beta}_1$	3.054	0.327	0.001
	$\hat{\beta}_2$	1.039	0.963	0.011
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	2.658	0.376	<0.001
	$\hat{\beta}_2$	1.028	0.972	0.004
Marjinal Cox Modeli	$\hat{\beta}_1$	3.051	0.328	<0.001
	$\hat{\beta}_2$	1.039	0.963	0.015
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.450	1.226	<0.001
	$\hat{\beta}_2$	0.978	1.009	0.011
Klasik Cox Modeli	$\hat{\beta}_1$	2.641	0.379	0.001
	$\hat{\beta}_2$	1.029	0.972	0.003

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.38.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %40 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %40	AIC	Uyum
Kırılmalılık Modeli	518.552	%66
Tabakalandırılmış Cox Modeli	983.225	%64
Marjinal Cox Modeli	518.576	%65
Marjinal Weibull Modeli	1356.684	%64
Klasik Cox Modeli	1149.343	%64

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %40 olduğunda elde edilen istatistikler tablo 4.37 ve 4.38.'de verilmiştir. Tablo 4.37' den elde edilen bulgular sonucunda her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.38' e göre model performans karşılaştırma kriterleri değerlendirildiğinde AIC=518.552 değeri ile kırılmalılık modeli en iyi model olarak bulunmuştur. Kırılmalılık modelini marjinal Cox modeli AIC=518.576 ile takip etmektedir. Ayrıca Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=1356.684). Klasik Cox modeli ise AIC=1149.343 ile en kötü performansa sahip ikinci model olarak bulunmuştur.

#### 4.3.5. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 50 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %50 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.39' da, AIC ve uyum değerleri tablo 4.40' da verilmiştir.

**Tablo 4.39.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %50 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %50		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılmalılık Modeli	$\hat{\beta}_1$	3.453	0.290	0.020
	$\hat{\beta}_2$	1.052	0.950	0.006
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	2.555	0.391	0.001
	$\hat{\beta}_2$	1.052	0.951	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	2.392	0.418	0.044
	$\hat{\beta}_2$	1.035	0.967	0.019
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.517	1.224	0.001
	$\hat{\beta}_2$	0.966	1.008	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	2.498	0.400	0.001
	$\hat{\beta}_2$	1.048	0.954	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.40.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %50 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %50	AIC	Uyum
<b>Kırılgenlik Modeli</b>	365.001	%84
<b>Tabakalandırılmış Cox Modeli</b>	829.581	%66
<b>Marjinal Cox Modeli</b>	371.176	%64
<b>Marjinal Weibull Modeli</b>	1175.570	%66
<b>Klasik Cox Modeli</b>	972.416	%66

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %50 olduğunda elde edilen istatistikler tablo 4.39 ve 4.40.'da verilmiştir. Tablo 4.39' de elde edilen bulgular bağlamında modellere dâhil edilen açıklayıcı değişkenler anlamlı olarak bulunmuştur. Tablo 4.40' da verilen AIC değerleri incelendiğinde kırılgenlik modeli en iyi model olarak bulunmuştur (AIC=365.001). Marjinal Cox modeli AIC=371.176 ile en iyi performansa sahip ikinci model olarak elde edilmiştir. Ayrıca uyum istatistiği incelendiğinde kırılgenlik modelinin uyumu %84 olarak elde edilmiştir. Marjinal weibull modeli en yüksek AIC=1175.570 değeri ile en kötü performansa sahip model olarak bulunurken, klasik Cox modeli AIC=972.416 değeri ile en kötü ikinci performansa sahip model olarak bulunmuştur.

#### **4.3.6. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 60 iken elde edilen bulgular**

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %60 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.41' de, AIC ve uyum değerleri tablo 4.42' de verilmiştir.

**Tablo 4.41.** Maksimum küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %60 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %60		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgenlik Modeli</b>	$\hat{\beta}_1$	2.762	0.362	<0.001
	$\hat{\beta}_2$	1.083	0.923	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	4.394	0.228	<0.001
	$\hat{\beta}_2$	1.043	0.959	0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.567	0.390	0.021
	$\hat{\beta}_2$	1.078	0.927	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.360	1.325	0.001
	$\hat{\beta}_2$	0.973	1.009	0.002
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	4.303	0.232	<0.001
	$\hat{\beta}_2$	1.039	0.962	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.42.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %60 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %60	AIC	Uyum
<b>Kırılgenlik Modeli</b>	307.578	%79
<b>Tabakalandırılmış Cox Modeli</b>	661.015	%69
<b>Marjinal Cox Modeli</b>	309.273	%73
<b>Marjinal Weibull Modeli</b>	890.347	%68
<b>Klasik Cox Modeli</b>	777.776	%68

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %60 olduğunda elde edilen istatistikler tablo 4.41 ve 4.42’de verilmiştir. Tablo 4.41’ de elde edilen bulgular bağlamında her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.42’ deki AIC değerleri incelendiğinde AIC=307.578 değeri ile kırılgenlik modeli en iyi performans gösteren model olarak bulunmuştur. Marjinal Cox modeli AIC=309.273 ile en iyi ikinci model olarak bulunmuştur. Ayrıca uyum istatistiği incelendiğinde kırılgenlik modelinin uyumu %79 ve marjinal Cox modelinin uyumu %73 olarak bulunmuştur. Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=890.347). Klasik Cox modeli ise AIC=777.776 ile en kötü ikinci model olarak elde edilmiştir.



### 4.3.7. Küme boyutu k=2 ve n=100 ve sağkalım oranı % 70 iken elde edilen bulgular

Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %70 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.43' de, AIC ve uyum değerleri tablo 4.44' de verilmiştir.

**Tablo 4.43.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %70 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=2 n=100 sağkalım oranı %70		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılgenlik Modeli	$\hat{\beta}_1$	11.597	0.086	0.004
	$\hat{\beta}_2$	1.074	0.931	0.006
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	4.025	0.248	0.002
	$\hat{\beta}_2$	1.048	0.954	0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	9.119	0.110	0.008
	$\hat{\beta}_2$	1.058	0.945	0.008
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.398	1.380	0.004
	$\hat{\beta}_2$	0.969	1.009	0.001
Klasik Cox Modeli	$\hat{\beta}_1$	4.149	0.241	0.001
	$\hat{\beta}_2$	1.049	0.953	0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

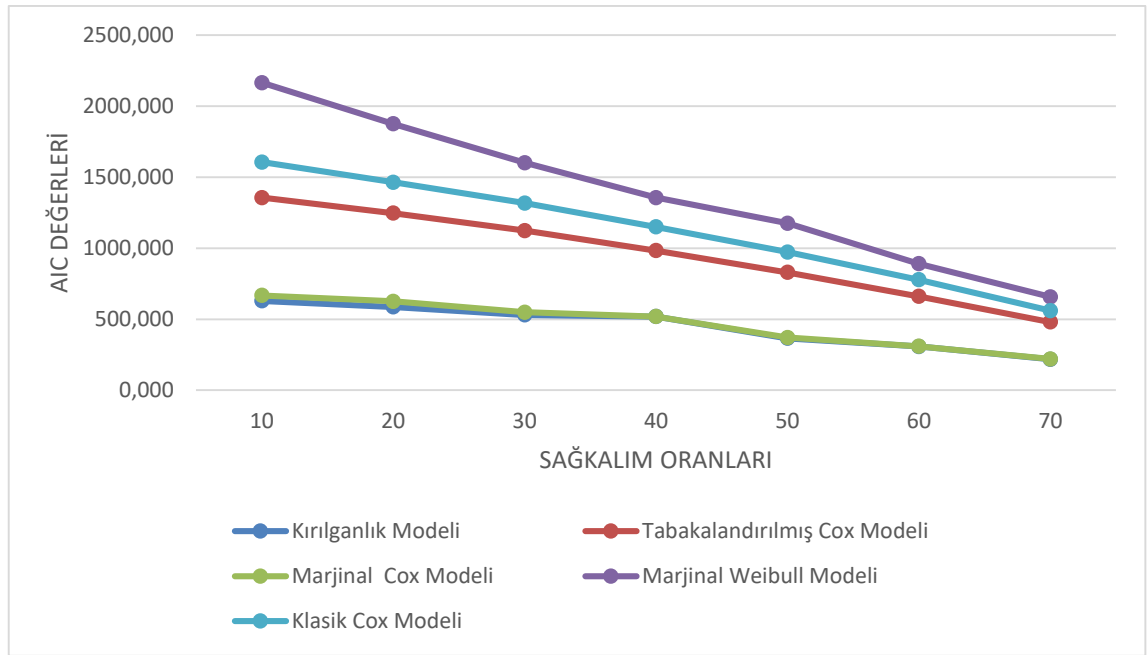
**Tablo 4.44.** Küme büyüklüğü 2 ve birey sayısı 100 ve sağkalım oranı %70 iken yöntemler için elde edilen AIC ve uyum değerleri

k=2 n=100 sağkalım oranı %70	AIC	Uyum
Kırılgenlik Modeli	217.751	%86
Tabakalandırılmış Cox Modeli	480.694	%69
Marjinal Cox Modeli	219.594	%74
Marjinal Weibull Modeli	657.753	%69
Klasik Cox Modeli	560.499	%69

Küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranı %70 olduğunda elde edilen istatistikler tablo 4.43 ve 4.44.'de verilmiştir. Tablo 4.43' de elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.44' deki AIC değerlerine göre AIC=217.751 değeri ile kırılgenlik modeli en iyi model olarak

bulunmuştur. Marjinal cox modeli AIC=219.594 ile takip etmektedir. Ayrıca uyum istatistiği incelendiğinde kırılmalı modelinin uyumu %86 ve marjinal Cox modelinin uyumu %74 olarak bulunmuştur. Marjinal weibull modeli en yüksek AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=657.753). Klasik Cox modeli ise AIC=560.499 ile en kötü ikinci model olarak bulunmuştur.

k=2, n=100 ve sağkalım oranı %10 ile %70 arasında değiştiği durumlar için her beş modelden elde edilen AIC değerlerini karşılaştıran grafik aşağıda verilmiştir.



**Grafik 4.4.** Maksimum küme büyüklüğü 2 birey sayısı 100 ve sağkalım oranları %10-%70 iken yöntemlere göre AIC değerleri

#### 4.4. Küme boyutu k=4 ve n=500, sağkalım oranlarının (%10-%70) değiştiği durumlar

##### 4.4.1. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 10 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %10 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.45' de, AIC ve uyum değerleri tablo 4.46' da verilmiştir.

**Tablo 4.45.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %10 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %10		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgenlik Modeli</b>	$\hat{\beta}_1$	2.623	0.381	<0.001
	$\hat{\beta}_2$	1.051	0.951	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	1.519	0.658	<0.001
	$\hat{\beta}_2$	1.027	0.974	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.593	0.628	0.001
	$\hat{\beta}_2$	1.027	0.973	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.123	0.594	<0.001
	$\hat{\beta}_2$	0.003	0.967	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.526	0.655	<0.001
	$\hat{\beta}_2$	1.027	0.9734	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.46.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %10 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %10	AIC	Uyum
<b>Kırılgenlik Modeli</b>	10261.520	%81
<b>Tabakalandırılmış Cox Modeli</b>	22032.410	%62
<b>Marjinal Cox Modeli</b>	10903.880	%63
<b>Marjinal Weibull Modeli</b>	22204.490	%62
<b>Klasik Cox Modeli</b>	24534.550	%62

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %10 olmak üzere elde edilen bulgular tablo 4.45 ve tablo 4.46' da verilmiştir. Bu bağlamda her beş modele alınan açıklayıcı değişkenler anlamlı bulunmuştur. AIC değerleri karşılaştırıldığında beş model içerisinde en iyi model kırılgenlik modeli olarak bulunmuştur (AIC=10261.520). Ayrıca AIC=10903.880 değeri ile marjinal Cox en iyi ikinci model olarak elde edilmiştir. Modeller arasında en kötü model AIC=24534.550 ile küme içi bağımlılığı dikkate almayan klasik cox modeli olmuştur. Marjinal weibull modeli en kötü ikinci model olarak elde edilmiştir (AIC=22204.490). Ayrıca uyum istatistiklerine bakıldığında ise kırılgenlik modelinin uyumu %81 bulunmuştur.

#### 4.4.2. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 20 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %20 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.47 de, AIC ve uyum değerleri tablo 4.48' de verilmiştir.

**Tablo 4.47.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %20 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %20		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılgenlik Modeli	$\hat{\beta}_1$	2.036	0.491	<0.001
	$\hat{\beta}_2$	1.059	0.944	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	1.703	0.587	<0.001
	$\hat{\beta}_2$	1.035	0.967	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	1.413	0.708	0.001
	$\hat{\beta}_2$	1.042	0.960	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.564	1.106	<0.001
	$\hat{\beta}_2$	0.965	1.003	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	1.718	0.582	<0.001
	$\hat{\beta}_2$	1.035	0.967	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.48.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %20 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %20	AIC	Uyum
Kırılgenlik Modeli	9544.301	%79
Tabakalandırılmış Cox Modeli	19926.990	%64
Marjinal Cox Modeli	9899.750	%63
Marjinal Weibull Modeli	19091.280	%64
Klasik Cox Modeli	22142.390	%64

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %20 olduğunda elde edilen istatistikler tablo 4.47 ve 4.48'de verilmiştir. Tablo 4.47' de elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.48' deki AIC değerlerine göre AIC=9544.301 değeri ile kırılgenlik modeli en iyi model olarak

bulunmuştur. Marjinal Cox modeli AIC=9899.750 ile ikinci en iyi performansa sahip model olarak elde edilmiştir. Ayrıca uyum istatistiği incelendiğinde kırılmalı modelinin uyumu %79 olarak bulunmuştur. Klasik Cox modeli en yüksek AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=22142.390). Tabakalandırılmış Cox modeli ise AIC=19926.990 ile en kötü ikinci model olarak bulunmuştur.

#### 4.4.3. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 30 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %30 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.3.19.'da, AIC ve uyum değerleri tablo 4.50' de verilmiştir.

**Tablo 4.49.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %30 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %30		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
Kırılmalı Modeli	$\hat{\beta}_1$	2.635	0.380	<0.001
	$\hat{\beta}_2$	1.054	0.949	<0.001
Tabakalandırılmış Cox Modeli	$\hat{\beta}_1$	1.843	0.543	<0.001
	$\hat{\beta}_2$	1.037	0.965	<0.001
Marjinal Cox Modeli	$\hat{\beta}_1$	1.880	0.532	<0.001
	$\hat{\beta}_2$	1.032	0.969	<0.001
Marjinal Weibull Modeli	$\hat{\beta}_1$	0.556	1.100	<0.001
	$\hat{\beta}_2$	0.966	1.003	<0.001
Klasik Cox Modeli	$\hat{\beta}_1$	1.846	0.5412	<0.001
	$\hat{\beta}_2$	1.037	0.964	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.50.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %30 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %30	AIC	Uyum
Kırılmalı Modeli	8540.484	%80
Tabakalandırılmış Cox Modeli	17764.160	%64
Marjinal Cox Modeli	8893.074	%63
Marjinal Weibull Modeli	16431.130	%64
Klasik Cox Modeli	19714.660	%64

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %30 iken elde edilen istatistikler tablo 4.49 ve 4.50.'de verilmiştir. Tablo 4.49' da elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.50' deki AIC değerlerine göre AIC=8540.484 değeri ile kırılgenlik modeli en iyi model olarak bulunmuştur. Marjinal cox modeli AIC=8893.074 ile kırılgenlik modelinden sonra en iyi performansa sahip ikinci model olarak elde edilmiştir. Ayrıca uyum istatistiği incelendiğinde kırılgenlik modelinin uyumu %80 olarak elde edilmiştir. Klasik Cox modeli AIC=19714.660 ile en kötü performansa sahip model olarak bulunmuştur. Tabakalandırılmış Cox modeli en yüksek ikinci AIC değerine sahip olarak en kötü performansa sahip ikinci model olarak bulunmuştur (AIC=17764.160).

#### 4.4.4. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 40 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %40 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.51' de, AIC ve uyum değerleri tablo 4.52' de verilmiştir.

**Tablo 4.51.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %40 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %40		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılgenlik Modeli</b>	$\hat{\beta}_1$	3.549	0.282	<0.001
	$\hat{\beta}_2$	1.063	0.941	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.077	0.481	<0.001
	$\hat{\beta}_2$	1.039	0.963	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.765	0.362	<0.001
	$\hat{\beta}_2$	1.045	0.957	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.513	1.095	<0.001
	$\hat{\beta}_2$	0.967	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.117	0.472	<0.001
	$\hat{\beta}_2$	1.039	0.963	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.52.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %40 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %40	AIC	Uyum
<b>Kırılgnlık Modeli</b>	6862.034	%82
<b>Tabakalandırılmış Cox Modeli</b>	15295.610	%65
<b>Marjinal Cox Modeli</b>	7069.0570	%68
<b>Marjinal Weibull Modeli</b>	13843.520	%65
<b>Klasik Cox Modeli</b>	16963.790	%65

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %40 olduğunda elde edilen istatistikler tablo 4.51 ve 4.52’de verilmiştir. Tablo 4.51’ de elde edilen bulgular bağlamında her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. modeller arası performansları karşılaştırmak amacı ile elde edilen AIC değerleri Tablo 4.52’ de verilmiştir. Bu değerler incelendiğinde ise en düşük AIC değerine sahip model kırılgnlık modeli (AIC=6862.034) olarak elde edilmiş ve en iyi performans gösteren model olarak elde edilmiştir. Kırılgnlık modelinden sonra en iyi performansa sahip model marjinal Cox modelidir (AIC=7069.057). Ayrıca uyum istatistiği incelendiğinde kırılgnlık modelinin uyumu %82 olarak bulunmuştur. Klasik Cox modeli ise AIC=16963.790 değeri ile en kötü model olarak bulunmuştur. Tabakalandırılmış Cox modeli ise en kötü performansa sahip ikinci model olarak elde edilmiştir (AIC=15295.610).

#### **4.4.5. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 50 iken elde edilen bulgular**

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %50 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.53’ de, AIC ve uyum değerleri tablo 4.54’ de verilmiştir.

**Tablo 4.53.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %50 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %50		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılğanlık Modeli</b>	$\hat{\beta}_1$	2.605	0.384	<0.001
	$\hat{\beta}_2$	1.054	0.949	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.040	0.490	<0.001
	$\hat{\beta}_2$	1.044	0.958	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.005	0.499	<0.001
	$\hat{\beta}_2$	1.041	0.961	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.576	1.093	<0.001
	$\hat{\beta}_2$	0.967	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.001	0.500	<0.001
	$\hat{\beta}_2$	1.043	0.958	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.54.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %50 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %50	AIC	Uyum
<b>Kırılğanlık Modeli</b>	6023.832	%83
<b>Tabakalandırılmış Cox Modeli</b>	12854.750	%66
<b>Marjinal Cox Modeli</b>	6249.411	%65
<b>Marjinal Weibull Modeli</b>	11539.430	%65
<b>Klasik Cox Modeli</b>	14262.480	%66

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %50 olduğunda elde edilen istatistikler tablo 4.53 ve 4.54' de verilmiştir. Tablo 4.53' de elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.54.' deki AIC değerlerine göre, AIC=6023.832 değeri ile kırılğanlık modeli en iyi model olarak bulunmuştur. Marjinal cox modeli AIC=6249.411 ile en iyi performansa sahip ikinci model olarak elde edilmiştir. Klasik Cox modeli AIC=14262.480 değeri ile en kötü model olarak bulunmuştur. Tabakalandırılmış Cox modeli ise en yüksek ikinci AIC değeri ile en kötü ikinci performansa sahip model olarak bulunmuştur (AIC=12854.75). Ayrıca uyum istatistiği incelendiğinde kırılğanlık modelinin uyum istatistiği %83 olarak bulunmuştur.



#### 4.4.6. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 60 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %60 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.55’ de, AIC ve uyum değerleri tablo 4.56’ da verilmiştir.

**Tablo 4.55.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %60 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %60		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılganlık Modeli</b>	$\hat{\beta}_1$	1.985	0.504	<0.001
	$\hat{\beta}_2$	1.054	0.949	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	1.716	0.583	<0.001
	$\hat{\beta}_2$	1.045	0.957	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	1.785	0.560	0.001
	$\hat{\beta}_2$	1.044	0.957	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.649	1.100	<0.001
	$\hat{\beta}_2$	0.966	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	1.724	0.580	<0.001
	$\hat{\beta}_2$	1.045	0.957	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.56.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %60 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %60	AIC	Uyum
<b>Kırılganlık Modeli</b>	4517.881	%83
<b>Tabakalandırılmış Cox Modeli</b>	9918.555	%65
<b>Marjinal Cox Modeli</b>	4606.858	%65
<b>Marjinal Weibull Modeli</b>	8838.711	%65
<b>Klasik Cox Modeli</b>	10990.050	%65

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %60 olduğunda elde edilen istatistikler tablo 4.55 ve 4.56.’da verilmiştir. Tablo 4.55’ de elde edilen bulgular bağlamında her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.56’ daki AIC değerleri incelendiğinde en düşük AIC değeri ile kırılganlık modeli (AIC=4517.881) en iyi performansa sahip model olarak elde edilmiştir. Marjinal Cox

modeli AIC=4606.858 ile ikinci en iyi model olarak bulunmuştur. Ayrıca uyum istatistiği incelendiğinde kırılmalı modelinin uyumu %83' dir. Klasik Cox modeli model performans karşılaştırma kriterine göre en yüksek AIC değeri ile en kötü model olarak elde edilmiştir (AIC=10990.050) ile en kötü ikinci model olarak bulunmuştur. Tabakalandırılmış Cox modeli ise en kötü performansa sahip ikinci model olmuştur (AIC=9918.555).

#### 4.4.7. Küme boyutu k=4 ve n=500 ve sağkalım oranı % 70 iken elde edilen bulgular

Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %70 iken yöntemlere göre elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri Tablo 4.57' de, AIC ve uyum değerleri tablo 4.58' de verilmiştir.

**Tablo 4.57.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %70 iken yöntemler için elde edilen  $\exp(\hat{\beta})$ ,  $se(\exp(\hat{\beta}))$  ve p değerleri

k=4 n=500 sağkalım oranı %70		$\exp(\hat{\beta})$	$se(\exp(\hat{\beta}))$	p-değeri
<b>Kırılmalı Modeli</b>	$\hat{\beta}_1$	3.610	0.277	<0.001
	$\hat{\beta}_2$	1.056	0.947	<0.001
<b>Tabakalandırılmış Cox Modeli</b>	$\hat{\beta}_1$	2.317	0.432	<0.001
	$\hat{\beta}_2$	1.050	0.953	<0.001
<b>Marjinal Cox Modeli</b>	$\hat{\beta}_1$	2.569	0.389	<0.001
	$\hat{\beta}_2$	1.047	0.955	<0.001
<b>Marjinal Weibull Modeli</b>	$\hat{\beta}_1$	0.547	1.107	<0.001
	$\hat{\beta}_2$	0.965	1.003	<0.001
<b>Klasik Cox Modeli</b>	$\hat{\beta}_1$	2.295	0.436	<0.001
	$\hat{\beta}_2$	1.050	0.953	<0.001

\* $\hat{\beta}_1$ : sigara değişkenine ait hazard oranı

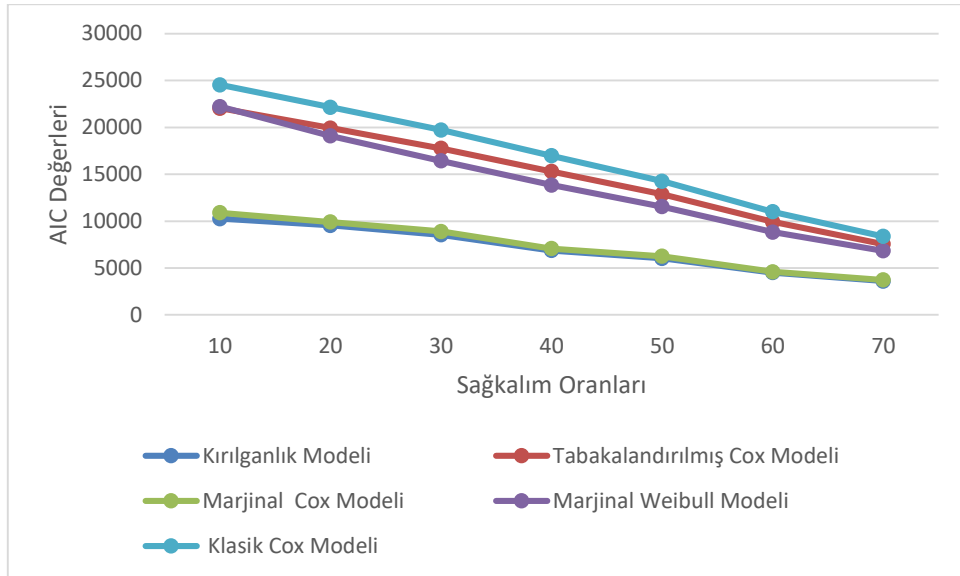
\* $\hat{\beta}_2$ : yaş değişkenine ait hazard oranı

**Tablo 4.58.** Küme büyüklüğü 4 ve birey sayısı 500 ve sağkalım oranı %70 iken yöntemler için elde edilen AIC ve uyum değerleri

k=4 n=500 sağkalım oranı %70	AIC	Uyum
<b>Kırılğanlık Modeli</b>	3610.244	%87
<b>Tabakalandırılmış Cox Modeli</b>	7546.261	%67
<b>Marjinal Cox Modeli</b>	3720.193	%68
<b>Marjinal Weibull Modeli</b>	6842.897	%67
<b>Klasik Cox Modeli</b>	8375.275	%67

Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranı %70 olduğunda elde edilen istatistikler tablo 4.57 ve 4.58.'de verilmiştir. Tablo 4.57' de elde edilen bulgulara göre her beş modele dâhil edilen açıklayıcı değişkenler anlamlı bulunmuştur. Tablo 4.58' deki AIC değerleri incelendiğinde AIC=3610.244 değeri ile kırılğanlık modeli en iyi model olarak bulunmuştur. Kırılğanlık modelinden sonra en iyi performansa sahip ikinci model marjinal Cox modelidir (AIC=3720.193). Ayrıca uyum istatistiği incelendiğinde kırılğanlık modelinin uyumu %87'dir. Klasik Cox modeli AIC=8375.275 ile en kötü ikinci model olarak bulunmuştur. Tabakalandırılmış Cox modeli ise en yüksek ikinci AIC değerine sahip olarak en kötü performansa sahip model olarak bulunmuştur (AIC=7546.261).

k=4, n=500 ve sağkalım oranı %10 ile %70 arasında değiştiği durumlar için her beş modelden elde edilen AIC değerlerini karşılaştıran grafik aşağıda verilmiştir.



**Grafik 4.5.** Küme büyüklüğü 4 birey sayısı 500 ve sağkalım oranları %10-%70 iken yöntemlere göre AIC değerleri

## 5. TARTIŞMA

Kümelenmiş veriler bireylerin farklı gruplara ayrıldığı ve en az birkaç grupta çoklu bulunduğu durumlarda ortaya çıkar (47). Son yıllarda, kümelenmiş verilerin analiz yöntemleri üzerine birçok sayıda araştırma yapılmıştır. Kümelenmiş veriler, gözlemlerin doğası, çalışma tasarımı veya verilerin örnekleme şekli nedeniyle birçok farklı analiz türünde bulunur (48). Bu analizlerden biri de sağkalım analizidir. Sağkalım analizi, bir kanser hastasının ameliyattan ölümüne, düğünden boşanmaya ve birinci ile ikinci intihar girişimleri arasında geçen süre gibi olaylarla ilgilenen bir istatistiksel analiz yöntemidir (3). Epidemiyolojik çalışmaların çoğunda genetik bilgi veya ölçülemeyen ortak çevresel maruziyetler gibi aynı kümeden elde bireylerin sahip olduğu bazı gözlemlenmemiş özelliklerin, incelenen olaya kadar geçen süreyi etkileyebileceği, aileler veya okullar gibi gruplar halinde kümelenen sağkalım zamanlarını içerir (49). Sağkalım verileri genellikle çiftler, kardeşler, aileler, topluluklar ve coğrafi bölgeler gibi gruplar halinde kümelenir. Çoğu zaman aynı küme içinden alınan gözlemler belirli özellikleri paylaştığı için korelasyona eğilimlidir. Bu korelasyonu dikkate almama yanlı tahminlere yol açabilir. Yani, aynı kümeden gelen gözlemler pozitif korelasyona sahip olma eğilimindeyse, tahminlerin kesinliğinin abartılması muhtemeldir (18). Kümelenmiş verilerde küme etkilerinin neden olduğu bağımlılığın dikkate alındığı yöntemler vardır. Bunlardan biri koşullu diğeri marjinal modeldir. Koşullu modellerde, hazard fonksiyonunda küme etkileri dikkate alınır ve aynı küme içindeki bireylerin sağkalım riskini karşılaştırmak için kullanılır. Marjinal yaklaşımlarda ise, sağkalım riskinde küme etkileri hesaba katılmaz. Popülasyondaki ortalama hazardı temsil eder ve popülasyondaki bireylerin sağkalım riskini karşılaştırmak için kullanılır (50). Yashin ve Iachine çalışmalarında, genetik analiz süreleri için korelasyonlu bireysel kırılabilirlik kavramına dayanan iki değişkenli bir sağkalım modelinin nasıl kullanılabileceğini göstermişlerdir. Çalışmada altı genetik kırılabilirlik modeli kullanılmış ve Danimarka'da ikizlerin sağkalım verilerine uygulanmıştır (51). Xue ve Brookmeyer çalışmalarında çok değişkenli sağkalım verilerinin analizinde tek değişkenli kırılabilirlik modelinin kısıtlamaları nedeniyle, iki değişkenli sağkalım verilerinin analizi için iki değişkenli kırılabilirlik modeli önermişlerdir. Bu model, özellikle aynı küme içindeki birimler arasında negatif ilişkiler için esneklik sağlamıştır. Önerilen yöntem, modele, her küme için birbiriyle korele iki kırılabilirliğin dâhil edilmesini içerir. Kırılabilirlik dağılımı olarak iki değişkenli lognormal dağılım kullanılmıştır. Ayrıca model çok değişkenli sağkalım verilerine ve alternatif

süreç verilerine genelleştirilmiştir. Modeli yatan hasta psikiyatrik bakımı için sağlık politikası etkilerinin değerlendirilmesi üzerine bir ruh sağlığı çalışması verileri üzerine uygulamışlardır (52). Sastry iki hiyerarşik düzeyde kümelenmiş sağkalım verileri için çok değişkenli bir hazard modeli önermiştir. Modelde, düzeltilmiş parametre tahminleri ve standart hatalarla birlikte her iki düzey içinde grup içi korelasyon tahminleri de sağlanmıştır. EM algoritması kullanılarak modeli tahmin edilmiştir. Ayrıca Brezilya'dan hiyerarşik olarak kümelenmiş bir sağkalım verisinin ortak değişkenlerini analiz etmek için uygulanmıştır (53). Parner çalışmasında korelasyonlu gamma kırılabilirlik modelini genişletmiş ve ortak değişkenlerin kullanıldığı bir model önermiştir (54). Balakrishnan ve Peng yaptıkları çalışmada, kırılabilirlik dağılımı olarak genelleştirilmiş gamma dağılımını kullanan ve lognormal ve Weibull kırılabilirlik modelleri gibi diğer kırılabilirlik modellerini de içeren kırılabilirlik modeli önermişlerdir ve sonuçlarını benzetim çalışması ile sunmuşlardır (55). Gharibvand ve Liu kümelenmiş sağkalım süresi verilerini analizi için ilki bağımsızlık varsayımı altında marjinal bir yaklaşım olan orantılı hazard modeli ve ikincisi ise olasılığa dayalı rasgele etkiler (kırılabilirlik) modeli olmak üzere iki yöntem önermişlerdir. Ayrıca diyabetik retinopati hasta verilerini kullanarak önerdikleri iki yöntemin avantaj ve dezavantajlarını tartışmışlardır (1). Tessema ve arkadaşları Etiyopya'da kadınların bölgesel durumlarını küme etkisi olarak kullandıkları kırılabilirlik modelleri ile ilk evliliğe kadar geçen sürenin belirleyicilerini modellemeyi amaçlamışlardır (56). Hanagal ve arkadaşları iki değişkenli kümelenmiş sağkalım verilerini analiz etmek için gamma bağımlı kırılabilirlik modelleri önermişlerdir. Ayrıca analizde gerçek böbrek enfeksiyon verilerini kullanmışlardır (57). Fagbamigbe ve arkadaşları Nijerya'da hiyerarşik bir veri kullanarak beş yaş altı ölümün belirleyicilerini değerlendirmek için paylaşılan kırılabilirlik modelini kullanmışlardır (58).

Kümelenmiş veriler diğer sağlık alanlarında olduğu gibi dış araştırmalarında da yaygın olarak rastlanır. Klinik dış araştırmaları genellikle, her hastadan birden fazla dışın alındığı örnekleri içerir. Bu çalışmadaki plak örneği için, hastalar birden fazla plaklı dişlere sahiptir ve bu da kümelenmiş veriler ile dolayısıyla bağımlı gözlemlerle sonuçlanır. Kümelenmiş verilerin analizi, dış araştırmalarında yaygın olarak karşılaşılan ancak genellikle göz ardı edilen bir durumdur.

Bu çalışmada küme içi bağımlılığı dikkate alan koşullu ve marjinal modeller ile klasik Cox modeli olmak üzere beş yöntem karşılaştırılmıştır. Koşullu modeller altında kırılabilirlik ve tabakalandırılmış Cox modeli, marjinal modeller altında marjinal Cox ve marjinal weibull modelleri yer almaktadır. Çalışma üç aşamadan oluşmuştur. İlk aşama

küme büyüklüklerinin eşit olduğu durumları kapsamaktadır. Küme büyüklüğü 2 ile 4 ve birey sayıları 50, 100, 200, 500 ve 1000 olarak üretilmiştir. İkinci aşamada ise veriler dengesiz küme boyutunda (maksimum küme boyutu 4) ve birey sayıları 50, 100, 200, 500 ve 1000 olacak şekilde ele alınmıştır. Çalışmanın üçüncü aşamasında ise farklı gözlem sayıları için %10 ile %70 arasında değişen sağkalım oranlarında veriler üretilmiştir. Her üç durum için de elde edilen senaryo sonuçları birbiri ile karşılaştırılmış elde edilen bulgular bağlamında yorumlanmıştır. Bu bulgularda en iyi modeli seçmek için literatürde önerilen birçok yöntem bulunmaktadır. Tez kapsamında modeller arası performansı karşılaştırmak için AIC değeri kullanılmıştır. Ayrıca modellerin uyumlarını değerlendirmek için uyum istatistiği kullanılmıştır.

Literatür incelendiğinde en küçük AIC değerine sahip model en iyi model olarak kabul edilmektedir. (46). Tessema ve arkadaşları Etiyopya' da ilk evliliğe kadar geçen sürenin belirleyicilerini modellemeyi amaçlarken kullandıkları modelleri karşılaştırmak için AIC kullanmışlardır (56). Fagbamigbe ve arkadaşları İsveç kohortunda implant olan dişlerdeki yeniden onarım komplikasyonlarının zamanlaması ile ilişkili faktörlerin değerlendirilmesinde 13 sağkalım regresyon modelini karşılaştırmışlar ve hangi modelin daha iyi olduğunu belirlemede AIC ve BIC kullanmışlardır (46). Wanigasekara ve arkadaşları Sri Lankalı çocukların diş çıkarma modelini tahmin etmek için uygun bir model oluşturmuşlar ve kullandıkları kırılma modelleri ve her diş tipi için ayrı ayrı geliştirilen Cox orantılı hazard model performanslarını karşılaştırmak için AIC, BIC ve RMSE kriterlerini kullanmışlardır (45).

Bulgular incelendiğinde yapılan benzetim çalışmalarında küme büyüklüğünün eşit olduğu durumlar için küçük örneklerde en iyi model kırılma modeli, en kötü model marjinal weibull olarak bulunmuştur. Örneklem büyüklüğünün artması ile küme içi korelasyonu göz ardı eden klasik cox modeli en kötü model olarak bulunmuştur. Elde edilen bulgular bağlamında marjinal weibull modeli en kötü performansa sahip ikinci model olmuştur. Örneklem büyüklüğünün daha da arttığı durumlarda ise, klasik Cox modeli yine en kötü performansa sahip iken en kötü ikinci model tabakalandırılmış Cox modeli olmuştur. Çalışmanın ikinci aşamasında ise küme boyutlarının dengesiz (kayıp verinin) olduğu durumlarda model performansları arasında değişiklik olup olmadığı incelenmiştir. Bu bağlamda elde edilen bulgular değerlendirilmiştir. Küçük örneklerde, en düşük AIC değerine sahip dolayısıyla en iyi performansa sahip olan model kırılma modeli olmuştur. Marjinal weibull en yüksek AIC değeri ile en kötü performansa sahip model olarak elde edilmiştir. Örneklem büyüklüğü arttıkça klasik Cox

modeli en kötü performansa sahip model olarak elde edilmiştir. Örneklem büyüklüğünün daha çok arttığı durumlarda ise tabakalandırılmış Cox modeli en kötü ikinci model olarak elde edilirken küme içi bağımlılığı göz ardı eden klasik Cox modeli en kötü model olarak elde edilmiştir. Çalışmanın üçüncü kısmındaki amaç değişen sağkalım oranlarında model performanslarını karşılaştırmaktır. Küçük örneklem büyüklüğünde en iyi model kırılabilirlik modeli olarak elde edilmiştir. En kötü model en yüksek AIC değeri ile marjinal weibull modeli olmuştur. En kötü ikinci model ise tabakalandırılmış Cox modeli olmuştur. Örneklem büyüklüğünün artmasına paralel olarak en kötü model klasik Cox olurken en kötü ikinci model marjinal weibull olmuştur. Gözlem sayısı arttıkça en kötü model küme içi korelasyonu dikkate alamayan klasik Cox modeli olur iken tabakalandırılmış Cox modeli en kötü ikinci model olarak elde edilmiştir. Ayrıca yapılan tüm benzetim sonuçlarında en iyi ikinci performansa sahip olan model marjinal Cox modeli olmuştur. Bu sonuçlar kapsamında çalışmada incelenen her üç durum içinde sonuçlar benzer bulunmuştur. Ek olarak literatürde model uyumu için kullanılan uyum istatistiği kırılabilirlik modellerinin hepsinde %70'in üzerinde bulunmuştur.

Bu çalışmaya paralel olarak, Lorino ve arkadaşları da yaptıkları çalışmada birçok veri tipinde olduğu gibi, sağkalım verilerinin de küçük gruplar halinde kümelendiği durumda klasik Cox modelinin yanlış sonuçlar verdiğini öne sürmüşlerdir. Çalışmalarında çok değişkenli sağkalım verileri için kırılabilirlik ve marjinal model olmak üzere iki farklı model önermişlerdir ve bizim çalışmamızın paralelinde çalışmalarında küçük örneklemde ziyade büyük örneklemde korelasyonu hesaba katan modellerin kullanılmasının daha doğru olacağını ifade etmişlerdir (35). Mahmood ve arkadaşları, aynı kümeden alınan gözlemlerin korelasyonlu olduğunu ve dolayısı ile bireyler arasında gözlemlenmemiş özellikleri paylaştığı ve gözlemler arası korelasyonun göz ardı edilmesi ile ilgilenilen değişkenlerin yanlış standart hatalarının tahminlerine yol açtığını ve bu nedenlerden dolayı klasik Cox modeli yerine kırılabilirlik modellerinin tercih edilmesini önermişlerdir (37). Thapa ve arkadaşları yaptıkları çalışmalarındaki uygulama verisine Cox modeli ve kırılabilirlik modellerini uygulayarak model performansları karşılaştırmışlardır. Elde ettikleri sonuçlara göre kırılabilirlik modelinin daha küçük bir AIC değerine ve daha yüksek tahmin yeteneğine sahip olduğunu saptamışlardır. Çalışmaya paralel olarak bu çalışmada da kırılabilirlik modeli en küçük AIC değerine sahiptir. (59) .

Bu çalışmada kümelenecek verilerde kullanılan sağkalım yöntemleri ile küme içi bağımlılığı dikkate almayan klasik Cox modelinin arasında fark olup olmadığı

karşılaştırılmıştır. Literatüre ek olarak bu çalışmada kümelenmiş sağkalım verilerini doğru analiz edilmesi ve istatistiksel olarak sonucun anlamlı çıkması için klasik Cox ve kırılma modeline ek olarak tabakalandırılmış Cox, marjinal Cox ve marjinal weibull modellerinin performansları değerlendirilmiştir. Çalışma bu bağlamda bilime yenilik katmıştır. Kırılma modeli ve marjinal Cox modeli en iyi model olarak belirlenmiştir. Bundan sonraki çalışmalarda bu tür verilerin analizinde sağkalım süresi ile beraber küme etkisini de dikkate alan yaklaşımların kullanılması sonuçların doğruluğu açısından önemli katkı sağlayacaktır.





## 6. SONUÇ VE ÖNERİLER

Kümelenmiş verilerde uygulanan tüm klasik istatistiksel analizlerde küme içi korelasyon dikkate alınmadığı için sonuçlar yanlış veya yanlış elde edilebilir. Bu nedenle kümelenmiş veriler analiz edilirken bu korelasyonun göz önüne alınması gerekir. Son zamanlarda çoğu istatistiksel analiz yöntemleri küme içi korelasyonu göz önüne alacak şekilde modifiye edilmişlerdir.

Bu tez çalışmasında klasik sağkalım yöntemi olan Cox metodu ile küme içi bağımlılığı dikkate alan yöntemleri karşılaştırmak amacı ile koşullu modellerden olan kırılabilirlik ve tabakalandırılmış Cox modeli ile marjinal modellerden olan marjinal Cox ile marjinal weibull modelleri olmak üzere toplamda beş farklı model için değişik senaryolarda veriler üretilmiştir. Üretilen bu veriler beş modele uygulanmış ve AIC değerine göre elde edilen sonuçlar karşılaştırılmıştır. Ayrıca uyum kriteri ile modellerin uyumu değerlendirilmiştir.

Yapılan benzetim çalışmalarında bireylerden alınan plaklı dişlerde çürük oluşması için geçen sürenin sigara içme ve yaş değişkenleri ile ilişkisinin modellenmesinde aynı bireylerden alınan dişlerin oluşturacağı bağımlılık da dikkate alınmıştır. Tüm küme sayısı ve küme büyüklüğü için oluşturulan farklı senaryolarda bireyler arasındaki dişlerin ortak paylaştıkları özellikler sebebi ile ortaya çıkan küme içi korelasyonu hesaba katan kırılabilirlik modeli en düşük AIC değerine sahip olmuştur. Böylece kırılabilirlik modeli tüm senaryolar için iyi model olarak kabul edilmiştir. Ayrıca kırılabilirlik modeli uyum kriteri açısından tüm modellerde %70'in üzerinde elde edilmiştir.

Elde edilen bulgulara göre aynı bireylerden alınan dişler arasındaki bağımlılığı ihlal eden ve tüm bireylerden alınan dişleri birbirinden bağımsız olarak kabul eden klasik Cox yöntemini özellikle gözlem sayısı arttıkça bireylerin arasındaki korelasyonunu göz ardı etmekte ve buna göre analiz etmektedir. Özellikle de senaryolardaki gözlem sayısının arttığı durumlardan elde edilen AIC değerleri diğer dört modelden daha yüksek bulunmuş ve böylece modeller arasında en kötü olan model olarak elde edilmiştir.

Sonuç olarak çalışmadan elde edilen bulguların sonucunda kümelenmiş sağkalım verilerinin çözümlenmesinde küme içi korelasyonu dikkate alan yöntemlerin kullanılması önerilmektedir. Küme içi bağımlılığı dikkate alan yöntemlerin sonucu her zaman yanlış istatistiksel sonuç çıkacak kadar yanlış olmayabilir. Nitekim gözlem sayısının az olduğu durumlarda küme içi bağımlılığı dikkate alan marjinal weibull modeli en kötü model olarak bulunmuştur. Ancak, küme başına alınan ölçümün artması ile elde edilen en kötü

performansa sahip model klasik Cox modelidir. Çalışmada kırılma noktası modeli ve marjinal Cox modeli en iyi model olarak belirlenmiştir. Bu tür verilerin analizinde sağkalım süresi ile beraber küme etkisini de dikkate alan yaklaşımların kullanılması uygundur. Bu sebeplerden dolayı yansız ve istatistiksel olarak geçerli sonuçlar elde etmek için kullanılacak olan yöntemlerin veri tipine uygun olması önerilmektedir.



## KAYNAKLAR

1. Gharibvand L, Liu L, editors. Analysis of survival data with clustered events. *SAS Global Forum* 2009, 237: 1-11.
2. Wienke A. *Frailty models in survival analysis*, 1<sup>th</sup> ed. New York, Chapman and Hall/CRC, 2010:1
3. Knox KL, Bajorska A, Feng C, Tang W, Wu P, Tu XM. Survival analysis for observational and clustered data: an application for assessing individual and environmental risk factors for suicide. *Shanghai archives of psychiatry* 2013, 25(3): 183-94.
4. Eriksson F. Correlated random effects models for clustered survival data. Department of Mathematical Sciences. Doctoral thesis, Goteborg: Chalmers University of Technology and University of Gothenburg, 2012.
5. Collett D. *Modelling survival data in medical research*, 3<sup>th</sup> ed. London, Chapman and Hall/CRC, 2015: 2-3.
6. Lin X, Carroll RJ. Semiparametric regression for clustered data using generalized estimating equations. *JASA* 2001, 96(455): 1045-56.
7. Datta S, Satten GA. Rank-sum tests for clustered data. *JASA* 2005, 100(471): 908-15.
8. Gönen M, Panageas KS, Larson SM. Statistical issues in analysis of diagnostic imaging experiments with multiple observations per patient. *Radiology* 2001, 221(3): 763-7.
9. Zyzanski SJ, Flocke SA, Dickinson LM. On the nature and analysis of clustered data. *Ann. Fam. Med* 2004, 2(3): 199-200.
10. Rao J, Scott A. A simple method for the analysis of clustered binary data. *Biometrics* 1992, 48(2): 577-85.
11. Zeng D, Lin D, Lin X. Semiparametric transformation models with random effects for clustered failure time data. *Statistica Sinica* 2008, 18(1): 355-77.
12. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. *Annu. Rev. Public Health* 1997, 18(1): 83-104.
13. Moore DF. *Applied survival analysis using R*, 5<sup>th</sup> ed. Switzerland, Springer, 2016:1.
14. Guo, S. *Survival analysis*, 1<sup>th</sup> ed. New York, Oxford University Press, 2010: 6-7.

15. Hou B. Sequential Analysis of Clustered Survival Data by Marginal Methods. Graduate Program in Statistics. Doctoral thesis, New Jersey: The State University of New Jersey, 2008.
16. Gangnon RE, Kosorok MR. Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika* 2004, 91(2): 263-75.
17. Gregg ME, Datta S, Lorenz D. A log rank test for clustered data with informative within-cluster group size. *Stat Med* 2018, 37(27): 4071-82.
18. Guo G, Rodriguez G. Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. *JASA* 1992, 87(420): 969-76.
19. Segal MR, Neuhaus JM, James IR. Dependence estimation for marginal models of multivariate survival data. *Lifetime Data Anal* 1997, 3(3): 251-68.
20. Glidden DV, Vittinghoff E. Modelling clustered survival data from multicentre clinical trials. *Stat Med* 2004, 23(3): 369-88.
21. Holt J, Prentice R. Survival analyses in twin studies and matched pair experiments. *Biometrika* 1974, 61(1): 17-30.
22. Zetterqvist J. Proportional Hazards Model for Matched Failure Time Data. Department of Mathematics. Master's thesis, Stockholm: Stockholm University, 2013.
23. Li Y, Lin X. Covariate measurement errors in frailty models for clustered survival data. *Biometrika* 2000, 87(4): 849-66.
24. Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. *J. R. Stat. Soc* 1985, 148(2): 82-108.
25. Gu MG, Sun L, Huang C. A universal procedure for parametric frailty models. *J Stat Comput Simul* 2004, 74(1): 1-13.
26. Klein JP. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 1992, 48(3): 795-806.
27. Bena J, McIntyre S, editors. Survival methods for correlated time-to-event data. MWSUG conference; 2008.
28. Chuang S, Tian L, Wei L, Dodson T. Kaplan-Meier analysis of dental implant survival: a strategy for estimating survival with clustered observations. *J. Dent. Res* 2001, 80(11): 2016-20.
29. Karasoy D, NA T. *Yaşam Çözümlemesi*, 1. Baskı. Ankara, Nobel Akademik Yayıncılık. 2016:140.

30. Cox DR. Regression models and life-tables. *J. R. Stat. Soc* 1972, 34(2): 187-202.
31. Shih JH, Louis TA. Assessing gamma frailty models for clustered failure time data. *Lifetime Data Anal* 1995, 1(2): 205-20.
32. Cai T, Wei L, Wilcox M. Semiparametric regression analysis for clustered failure time data. *Biometrika* 2000, 87(4): 867-78.
33. Jung S-H, Jeong J-H. Rank tests for clustered survival data. *Lifetime Data Anal* 2003, 9(1): 21-33.
34. Chuang S, Tian L, Wei L, Dodson T. Predicting dental implant survival by use of the marginal approach of the semi-parametric survival methods for clustered observations. *J. Dent. Res* 2002, 81(12): 851-5.
35. Lorino T, Sanaa M, Robin S, Daudin JJ. Comparison of semiparametric regression models for correlated survival data using simulations. *Commun. Stat* 2004, 33(8): 1975-91.
36. Zhang H, Schaubel DE, Kalbfleisch JD. Proportional hazards regression for the analysis of clustered survival data from case-cohort studies. *Biometrics* 2011, 67(1): 18-28.
37. Mahmood S, Zainab B, Latif AM. Frailty modeling for clustered survival data: an application to birth interval in Bangladesh. *J. Appl. Stat* 2013, 40(12): 2670-80.
38. Gauran IIM, Barrios EB. Nonparametric modeling of clustered customer survival data. *Commun. Stat* 2017, 46(1): 603-18.
39. Austin PC. A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review* 2017, 85(2): 185-203.
40. Dahlqwist E, Pawitan Y, Sjölander A. Regression standardization and attributable fraction estimation with between-within frailty models for clustered survival data. *Stat Methods Med Res* 2019, 28(2): 462-85.
41. Team R. RStudio: Integrated development for R (version 1.1. 463)[Computer software]. Boston: RStudio, Inc. <http://www.rstudio.com>; 2018.
42. Saraçoğlu S, Öztürk F. Türkiye’de tütün kontrol politikaları ve tütün tüketimi üzerine bir değerlendirme. *PEK* 2020, 4(1): 20-44.
43. Brettschneider J, Burgess M. Using a frailty model to measure the effect of covariates on the disposition effect. *CRISM Working Paper Series*, 2017, 17 (5)1-46.

44. Kiche J, Ngesa O, Orwa G. Misspecification of frailty random effects in a clustered survival data. *Journal of Statistical and Econometric Methods* 2019, 8(2): 1-25.
45. Wanigasekara C, Nawarathna LS, Vithanaarachchi V. Frailty Models for Predicting Eruption Time and Sequence of Permanent Dentition in Sri Lankan Children. *Asian J. Probab. Stat* 2021, 11(3): 1-10.
46. Fagbamigbe AF, Karlsson K, Derks J, Petzold M. Performance evaluation of survival regression models in analysing Swedish dental implant complication data with frailty. *Plos one* 2021; 16(1): 1-16.
47. Holodinsky JK, Austin PC, Williamson TS. An introduction to clustered data and multilevel analyses. *Fam. Pract* 2020, 37(5): 719-22.
48. Graubard BI, Korn EL. Regression analysis with clustered data. *Stat Med* 1994, 13(5-7): 509-22.
49. Gorfine M, Zucker DM, Hsu L. Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika* 2006, 93(3): 735-41.
50. Gorfine M, De-Picciotto R, Hsu L. Conditional and marginal estimates in case-control family data—extensions and sensitivity analyses. *J Stat Comput Simul* 2012, 82(10): 1449-70.
51. Yashin AI, Iachine IA. Genetic analysis of durations: correlated frailty model applied to survival of Danish twins. *Genet. Epidemiol* 1995, 12(5): 529-38.
52. Xue X, Brookmeyer R. Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Anal* 1996, 2(3): 277-89.
53. Sastry N. A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *JASA* 1997, 92(438): 426-35.
54. Parner E. Asymptotic theory for the correlated gamma-frailty model. *Ann. Stat* 1998, 26(1): 183-214.
55. Balakrishnan N, Peng Y. Generalized gamma frailty model. *Stat Med.* 2006, 25(16): 2797-816.
56. Tessema B, Ayalew S, Mohammed K. Modeling the determinants of time-to-age at first marriage in ethiopian women: A comparison of various parametric shared frailty models. *SJPH* 2015, 3(5): 707-18.
57. Hanagal DD, Pandey A, Ganguly A. Correlated gamma frailty models for bivariate survival data. *Commun. Stat* 2017, 46(5): 3627-44.

58. Fagbamigbe AF, Afolabi RF, ALADE KY, ADEBOWALE AS, Yusuf BO. Unobserved heterogeneity in determinants of under-five mortality in Nigeria: Frailty modeling in survival analysis. *Afr. J. Appl* 2019, 6(1): 565-87.
59. Thapa R, Burkhart HE, Li J, Hong Y. Modeling clustered survival times of loblolly pine with time-dependent covariates and shared frailties. *J Agric Biol Environ Stat* 2016, 21(1): 92-110.



## **Ek-2. Etik Kurul Almama Gerekçesi**

Çalışmanın klinik olmaması, kullanılacak verinin hipotetik olmasından dolayı Resmi Gazetenin 13 Nisan 2013 tarih ve 28617 sayılı “Klinik Araştırmalar Hakkındaki Yönetmelikte” belirtilen Etik Kurul Onay Belgesine gerek bulunmamaktadır.

