

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

**SAĞDAN SANSÜRLÜ GÖZLEMLERİN YERLEŞİMİNİN KAPLAN MEİER  
VE  
HAZARD ORANI TAHMİNİNE ETKİSİ**

Seval KUL

DOKTORA TEZİ

DANIŞMANI

Doç. Dr. Arzu KANIK

MERSİN – 2009

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

**SAĞDAN SANSÜRLÜ GÖZLEMLERİN YERLEŞİMİNİN KAPLAN MEİER  
VE  
HAZARD ORANI TAHMİNİNE ETKİSİ**

Seval KUL

DOKTORA TEZİ

DANIŞMANI

Doç. Dr. Arzu KANIK

Tez No:

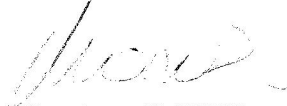
Bu tez DAAD (Deutscher Akademischer Austausch Dienst) tarafından tez çalışmalarının bir kısmının yurt dışında gerçekleştirilmesi amacıyla A/60/09280 nolu proje ile desteklenmiştir.

MERSİN – 2009


**Mersin Üniversitesi Sağlık Bilimleri Enstitüsü**

Biyoistatistik Anabilim Dalı Doktora Programı çerçevesinde yürütülmüş olan “Sağdan Sansürlü Gözlemlerin Yerleşiminin Kaplan Meier ve Hazard Oranı Tahminine Etkisi” adlı çalışma aşağıdaki jüri tarafından Doktora Tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 17.08.2009



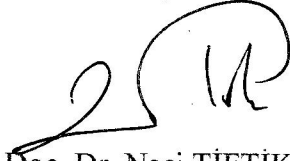
Doç. Dr. Arzu KANIK  
Mersin Üniversitesi Tıp Fakültesi  
Biyoistatistik Anabilim Dalı  
Jüri Başkanı



Prof. Dr. Nazan ALPARSLAN  
Çukurova Üniversitesi Tıp Fakültesi  
Biyoistatistik Anabilim Dalı  
Jüri Üyesi



Prof. Dr. Adnan ERKUŞ  
Mersin Üniversitesi Eğitim Fakültesi  
Eğitim Bilimleri Anabilim Dalı  
Jüri Üyesi



Doç. Dr. Naci TİFTİK  
Mersin Üniversitesi Tıp Fakültesi  
İç Hastalıkları Anabilim Dalı  
Jüri Üyesi



Yrd. Doç. Dr. Bahar TAŞDELEN  
Mersin Üniversitesi Tıp Fakültesi  
Biyoistatistik Anabilim Dalı  
Jüri Üyesi

Yukarıdaki tez, Enstitü Yönetim Kurulunun.....18.08.2009..... tarih ve 2009/290...sayılı kararı ile kabul edilmiştir.

Enstitü Müdürü  
Doç. Dr. Ülkü ÇÖMTELEKÇİ



## **TEŐEKKÜR**

Yanlarında geirdiđim yıllar boyunca bana akademik alıŐmalarımın hepsinde sınırsız destek veren, yeri gelince arkadaş ve abla olabilen, sadece eđitimci olarak deđil insan olarak da örnek olan, bana Biyoistatistiđi öđreten ve sevdiren Anabilim dalımız öđretim üyeleri Do.Dr. Arzu Kanık ve Yrd. Do. Dr. Bahar TaŐdelen'e, tüm sorunlarımızda yanlarına gittiđimiz halde sonsuz bir anlayıŐ ve güler yüzle bizleri karŐılayan, Sađlık Bilimleri Enstitüsü akademik ve idari personeline ve en nihayetinde beni büyüten ve yetiŐtiren aileme ok teŐekkür ediyorum.

# İÇİNDEKİLER

KABUL VE ONAY .....	iii
TEŞEKKÜR.....	iv
İÇİNDEKİLER .....	v
ŞEKİLLER DİZİNİ .....	vii
ÇİZELGELER DİZİNİ .....	x
ÖZET .....	xii
1. GİRİŞ .....	1
2. GENEL BİLGİLER .....	5
2.1. Neden Hayatta Kalma Analizleri? .....	5
2.2. Sansürlü Veriler .....	6
2.2.1. Sağdan Sansürlü Veriler .....	7
2.2.2. Soldan Sansürlü Veriler .....	8
2.2.3. Aralık Sansürlü Veriler .....	8
2.3. Hayatta Kalma Analizlerinde Kullanılan Fonksiyonlar.....	8
2.3.1. Hayatta Kalma Fonksiyonu .....	9
2.3.2. Olasılık Yoğunluk Fonksiyonu .....	11
2.3.3. Hazard Fonksiyonu .....	12
2.4. Hayatta Kalma Fonksiyonları Arasındaki İlişkiler .....	14
2.5. Hayatta Kalma Fonksiyonunun Parametrik Olmayan Yöntemler ile Tahmini: Kaplan Meier Yöntemi .....	16
2.5.1. Kaplan Meier Yönteminin Bazı Kritik Özellikleri .....	18
2.6. Kernel Tahmin Yöntemi .....	19
2.6.1. Kernel ağırlıklandırma fonksiyonları.....	19
2.6.2. Bant genişliğinin belirlenmesi .....	21
2.6.3. Tahmin edicinin başarısının test edilmesi.....	21
2.7. Bazı Parametrik Hayatta Kalma Dağılımları .....	22
2.7.1. Üstel Dağılım .....	22
2.7.2. Düzgün Dağılım.....	23
3. GEREÇ VE YÖNTEM .....	25
4. BULGULAR.....	28
4.1. Sabit Hazard Oranı, N=100 için Kaplan Meier Tahminleri.....	28
4.2. Sabit Hazard Oranı, N=100, için Hazard Oranı Tahminleri .....	32
4.3. Sabit Hazard Oranı, N=250, için Kaplan Meier Tahminleri.....	36
4.4. Sabit Hazard Oranı, N=250, için Hazard Oranı Tahminleri .....	40
4.5. Lineer Artan Hazard Oranı, N=100 için Kaplan Meier Tahminleri .....	45
4.6. Lineer Artan Hazard Oranı, N=100 için Hazard Oranı Tahminleri.....	49
4.7. Lineer Artan Hazard Oranı, N=250 için Kaplan Meier Tahminleri .....	53
4.8. Lineer Artan Hazard Oranı, N=250 için Hazard Tahminleri .....	58
4.9. Lineer Azalan Hazard Oranı, N=100 için Kaplan Meier Tahminleri .....	61
4.10. Lineer Azalan Hazard Oranı, N=100, için Hazard Oranı Tahminleri .....	66
4.11. Lineer Azalan Hazard Oranı, N=250 için Kaplan Meier Tahminleri .....	71
4.12. Lineer Azalan Hazard Oranı, N=250, için Hazard Oranı Tahminleri .....	75
5. TARTIŞMA .....	80
6. SONUÇ VE ÖNERİLER.....	85

KAYNAKLAR .....	86
EKLER.....	89

## ŞEKİLLER DİZİNİ

Şekil 2.1. Hayatta kalma analizlerinde karşılaşılabilecek olası veri tipleri .....	7
Şekil 2.2 Hayatta kalma fonksiyonu .....	9
Şekil 2.3 Olasılık yoğunluk fonksiyonu örnekleri .....	12
Şekil 2.4 Hazard oranı için örnek grafikler. ....	13
Şekil 2.5 Bazı Kernel fonksiyonlar .....	20
Şekil 2.6 Üstel dağılıma ait hayatta kalma fonksiyonu (a), olasılık yoğunluk fonksiyonu (b), hazard fonksiyonu (c).....	23
Şekil 2.7 Düzgün dağılıma ait hayatta kalma fonksiyonu (a), olasılık yoğunluk fonksiyonu (b), hazard fonksiyonu (c).....	24
Şekil 4.1 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri .....	28
Şekil 4.2 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	29
Şekil 4.3 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	30
Şekil 4.4 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	30
Şekil 4.5 Sabit hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için Kaplan Meier tahmin hatalarına ait box plot grafikleri .....	31
Şekil 4.6 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	33
Şekil 4.7 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	33
Şekil 4.8 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	34
Şekil 4.9 Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	34
Şekil 4.10 Sabit hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için Kernel Düzgünleştirme fonksiyonu tahmin hatalarına ait box plot grafikleri .....	35
Şekil 4.11 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminler .....	37
Şekil 4.12 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	37
Şekil 4.13 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	38
Şekil 4.14 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	38
Şekil 4.15 Sabit hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	39
Şekil 4.16 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	41
Şekil 4.17 Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	41

<b>Şekil 4.18</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	42
<b>Şekil 4.19</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	42
<b>Şekil 4.20</b> Sabit hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	43
<b>Şekil 4.21</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri .....	45
<b>Şekil 4.22</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	46
<b>Şekil 4.23</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	46
<b>Şekil 4.24</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	47
<b>Şekil 4.25</b> Doğrusal artan hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	48
<b>Şekil 4.26</b> Lineer artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	50
<b>Şekil 4.27</b> Lineer artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	50
<b>Şekil 4.28</b> Lineer artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	51
<b>Şekil 4.29</b> Lineer artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	51
<b>Şekil 4.30</b> Lineer artan hazard oranı, $n=100$ farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	52
<b>Şekil 4.31</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri .....	54
<b>Şekil 4.32</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	54
<b>Şekil 4.33</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	55
<b>Şekil 4.34.</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	55
<b>Şekil 4.35</b> Doğrusal artan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	56
<b>Şekil 4.36</b> Lineer artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	58
<b>Şekil 4.37</b> Lineer artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	59
<b>Şekil 4.38</b> Lineer artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	59
<b>Şekil 4.39</b> Lineer artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	60
<b>Şekil 4.40</b> Lineer artan hazard oranı, $n=250$ farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	60
<b>Şekil 4.41</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri .....	62



<b>Şekil 4.42</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	62
<b>Şekil 4.43</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	63
<b>Şekil 4.44</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	64
<b>Şekil 4.45</b> Lineer azalan hazard oranı, $n=100$ farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	65
<b>Şekil 4.46</b> Lineer azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	67
<b>Şekil 4.47</b> Lineer azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	67
<b>Şekil 4.48</b> Lineer azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	68
<b>Şekil 4.49</b> Lineer azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	68
<b>Şekil 4.50</b> Lineer azalan hazard oranı, $n=100$ farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	69
<b>Şekil 4.51</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri .....	71
<b>Şekil 4.52</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri .....	72
<b>Şekil 4.53</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri .....	72
<b>Şekil 4.54</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri .....	73
<b>Şekil 4.55</b> Doğrusal azalan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	73
<b>Şekil 4.56</b> Lineer azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri .....	75
<b>Şekil 4.57</b> Lineer azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri .....	76
<b>Şekil 4.58</b> Lineer azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri .....	77
<b>Şekil 4.59</b> Lineer azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri .....	77
<b>Şekil 4.60</b> Lineer azalan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri .....	78
<b>Şekil 5.1</b> $N=100$ $K=1000$ iken Kaplan Meier tahminlerine ait tüm tahmin hataları karşılaştırma .....	80
<b>Şekil 5.2</b> $N=250$ $K=1000$ iken Kaplan Meier tahminlerine ait tüm tahmin hataları karşılaştırma .....	81
<b>Şekil 5.3</b> $N=100$ , $K=1000$ iken Hazard oranı tahminlerine ait tüm tahmin hataları karşılaştırma .....	83
<b>Şekil 5.4</b> $N=250$ , $K=1000$ iken Hazard oranı tahminlerine ait tüm tahmin hataları karşılaştırma .....	83

## ÇİZELGELER DİZİNİ

<b>Çizelge 2.1</b> Bazı dağılımlara ait hayatta kalma, olasılık yoğunluk ve hazard fonksiyonları .....	15
<b>Çizelge 2.2</b> Kernel ağırlıklandırma fonksiyonları .....	20
<b>Çizelge 3.1</b> Simülasyonda kullanılan dağılımlara ilişkin hayatta kalma fonksiyonları ve fonksiyonların tersleri .....	26
<b>Çizelge 3.2</b> Simülasyonda kullanılan dağılımlara ilişkin hazard fonksiyonu ve katsayıları .....	26
<b>Çizelge 4.1</b> Sabit hazard oranı, $n=100$ , Kaplan Meier tahminlerinin hatalarına ait tanıtıcı istatistikler.....	31
<b>Çizelge 4.2</b> Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	32
<b>Çizelge 4.3</b> Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	32
<b>Çizelge 4.4</b> Sabit hazard oranı, $n=100$ , Kernel düzgünleştirme tahminlerinin hatalarına ait tanıtıcı istatistikler.....	36
<b>Çizelge 4.5</b> Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	36
<b>Çizelge 4.6</b> Sabit hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	36
<b>Çizelge 4.7</b> Sabit hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	40
<b>Çizelge 4.8</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	40
<b>Çizelge 4.9</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	40
<b>Çizelge 4.10</b> Sabit hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	44
<b>Çizelge 4.11</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.20 için çoklu karşılaştırma sonuçları .....	44
<b>Çizelge 4.12</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları .....	44
<b>Çizelge 4.13</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	44
<b>Çizelge 4.14</b> Sabit hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	45
<b>Çizelge 4.15</b> Doğrusal artan hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	48
<b>Çizelge 4.16</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları. ....	49
<b>Çizelge 4.17</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları. ....	49
<b>Çizelge 4.18</b> Doğrusal artan hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	52

<b>Çizelge 4.19</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları .....	53
<b>Çizelge 4.20</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	53
<b>Çizelge 4.21</b> Doğrusal artan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	53
<b>Çizelge 4.22</b> Doğrusal artan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	57
<b>Çizelge 4.23</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	57
<b>Çizelge 4.24</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	57
<b>Çizelge 4.25</b> Doğrusal artan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	61
<b>Çizelge 4.26</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	61
<b>Çizelge 4.27</b> Doğrusal artan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	61
<b>Çizelge 4.28</b> Doğrusal azalan hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	65
<b>Çizelge 4.29</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları .....	66
<b>Çizelge 4.30</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	66
<b>Çizelge 4.31</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	66
<b>Çizelge 4.32</b> Lineer azalan hazard oranı, $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	69
<b>Çizelge 4.33</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları .....	70
<b>Çizelge 4.34</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	70
<b>Çizelge 4.35</b> Doğrusal azalan hazard oranı, $n=100$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	70
<b>Çizelge 4.36</b> Doğrusal azalan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	74
<b>Çizelge 4.37</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları .....	74
<b>Çizelge 4.38</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	75
<b>Çizelge 4.39</b> Doğrusal azalan hazard oranı, $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler.....	78
<b>Çizelge 4.40</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları .....	79
<b>Çizelge 4.41</b> Doğrusal azalan hazard oranı, $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları .....	79

## ÖZET

### Sağdan Sansürlü Gözlemlerin Yerleşiminin Kaplan Meier ve Hazard Oranı Tahminine Etkisi

Hayatta kalma analizlerinde sansürlü gözlemlerin rasgele dağıldığı ve veri seti içinde her hangi bir zamanda rasgele olarak oluşabileceği varsayılır. Fakat bazı klinik çalışmalarda sansürlü gözlemlerin büyük bir kısmı belirli bir zaman aralığı içinde oluşabilir. Bunun birçok nedeni olabilir (hastaların tedaviden memnun olmaması, tedavinin yeterli gelmemesi, basında bir ilaç ya da tedavi hakkında çıkan olumsuz haberler, hastane enfeksiyonu varlığı... vb). Bu tezde sansürlü gözlemlerin veri setinde rasgele olarak oluşmadığı, belirli bir aralıkta olduğu durumda hayatta kalma fonksiyonu ve hazard oranı tahminlerine etkisinin araştırılması amaçlanmıştır.

Tezin simülasyon kısmında, sabit hazard, doğrusal artan ve doğrusal azalan hazard olmak üzere üç farklı dağılımdan örnek genişliği 100 ve 250 için hayatta kalma fonksiyonu ve hazard değerleri hesaplanarak dağılımlar oluşturulmuştur. Veri seti içinde 4 farklı oranda (0.20, 0.30, 0.40, 0.50) sansürlü gözlem oluşturulmuştur. Sansürlü gözlemler veri seti içinde dört farklı bölgeye yerleştirilmiştir. Hayatta kalma fonksiyonunun tahmininde Kaplan Meier yöntemi, hazard fonksiyonlarının tahmininde ise Kernel düzgünleştirme yöntemi kullanılmıştır. Simulasyon programları R programlama dilinde yazılmıştır ve her bir simulasyon 1000 defa tekrarlanmıştır.

Sonuç olarak, sansürlü gözlemlerin veri seti içindeki yerleşiminin bazı tahminlerde tamamen rasgele oluşturulmuş sansür yerleşimine göre istatistiksel olarak anlamlı farklılıklar yarattığı gözlenmiştir. Fakat bu farklılıkların binde birlik düzeylerde olduğu dikkate alındığında sonuçların klinik olarak bir farklılık oluşturmayacağına karar verilmiştir. Bunun yanı sıra sansürlü gözlemlerin oranının artırılmasının tahminleri kötüleştirdiği bir kez daha ortaya konulmuştur. Ayrıca küçük örneklerde ve tamamlanmış gözlem sayısının az olduğu bölgelerde tahminlerin kötüleştiği gözlenmiştir.

**Anahtar Sözcükler:** Hayatta kalma analizleri, Kaplan Meier tahmini, hazard tahmini, Kernel düzgünleştirme yöntemi, sansürlü gözlemler.

## **ABSTRACT**

### **Effect of Right Censored Observations Locations on Kaplan Meier and Hazard Rate Estimates**

In survival analysis censored observations are assumed randomly disturbed and censoring can occur any time in data set randomly. But in some medical study, most of the right censoring can occur within very short time interval. This can be result of many reasons (e.g., dissatisfactory or inefficient therapy, negative news about remedy or interventions and high risk of hospital infection). In this study, we studied in this special case and we aimed to show how location of censored data affects the survival and hazard rate estimates.

In the simulation part, we simulated data from 3 different distributions; survival functions derived from constant, linear increasing and linear decreasing hazard functions including censored observation with four different proportions (200., 300., 400., 500.) of censoring and censoring located at three different parts of data with two sample sizes ( $n=100$ ,  $n=250$ ). Simulation programs were implemented in R language and each simulation was repeated 1000 times.

As a result of our simulations, we found location of right censored observations in data set has significant effect on Kaplan Meier and Kernel smoothing estimates. But differences were too small to make clinically significant effect. Additionally we showed that, increasing of the censored observation proportions in life time data makes survival and hazard function estimates worse. Also estimates were worse at the time when there are few observations left under risk.

**Key Words:** Survival analysis, Kaplan Meier estimate, hazard estimation, Kernel smoothing, censored observations.

# 1. GİRİŞ

İlgilenilen sonuç değişkeninin bir olayın gerçekleşmesine kadar geçen süre olduğu durumlarda veri analizinde kullanılan yöntemlerin tamamı “Hayatta kalma analizleri (Survival analysis)” olarak adlandırılır (1). Klinik çalışmalarda her hasta için yaşam süresine ulaşılması mümkün değildir. Bazı hastalar çalışmanın sonlanma zamanında halen hayatta olabilirken, bazıları ise çalışmadan kendi istekleri ile çekilebilirler ya da başka nedenlerden dolayı (başka bir hastalık dolayısıyla ölüm, tedaviden memnuniyetsizlik, alternatif yeni bir tedavinin varlığı, taşınma... vb) çalışmadan çıkabilirler. Bu durumda hastalar hakkında bazı bilgilere sahip olmamıza karşın hayatta kalma sürelerini tam olarak bilmemiz mümkün değildir. Yukarıda belirtilen olası nedenlerden dolayı hayatta kalma süreleri bilinmeyen gözlemler sansürlü (censored) gözlemler olarak adlandırılırlar (2).

Sansürlü gözlemlerin çalışmadan çıkarılması araştırmacı için bilgi kaybına neden olacaktır. Her türlü bilimsel çalışmada elde edilen verinin en doğru şekilde kullanılması ve hiçbir verinin boşa harcanmaması amaçlanır (3). Fakat hayatta kalma verileri uzun süreli takip gerektirdiğinden doğası gereği sansürlü gözlemler oluşacaktır. Sansürlü gözlem içeren 2 grubun yaşam süreleri karşılaştırılmasında Student t testi kullanıldığında sansürlü gözlemler görmezlikten gelinecektir ya da gruplarda olayların görülme olasılığını (riskini) lojistik regresyon analizi ile karşılaştırdığımız durumda zaman değişkeni analizlerde dikkate alınmayacaktır (4). Yaşam sürelerinin analizlerinde oluşabilecek bu tür sorunlardan dolayı, veri analizlerinde hem zamanı hem de sansürlü gözlemleri dikkate alan hayatta kalma analizleri geliştirilmiştir.

Hayatta kalma analizlerinin tarihi yüzyıllarca öncesinde mortalite tablolarının oluşturulmasına dayanmaktadır. II. Dünya Savaşında askeri malzemelerin dayanıklılığının ya da güvenilirliğinin test edilmesine hayatta kalma analizlerinin kullanılmasının ardından özel sektörde de tüketicinin daha dayanıklı ve güvenilir ürün talebinin karşılanması amacıyla hayatta kalma analizleri kullanılmaya başlanmıştır (5). Ardından sansürlü gözlemlerin analizini mümkün kılmasından dolayı tıp alanında da uygulanmıştır (5).

Kaplan ve Meier 1958 yılında ‘‘Kaplan Meier product limit yntemi’’ yardımıyla sansrl gzlemlerden de faydalanarak bireylerin yařam fonksiyonlarının tahmin edilebileceđini gstermiřlerdir (6). Kaplan ve Meier makalesinde sansrl gzlemler ieren bir veri setini ařađıdaki gibi ifade etmiřtir. Y hayatta kalma sresini ve C ise sansrl gzlemin en son bilinen yařam sresini gstermek zere, Y ve C bađımsız ve benzer dađılıma sahip olan rasgele deđiřkenler olsun ve sırasıyla (R,R) de deđerler alsın,

$$(C_1, C_2, \dots, C_n), (C_1, C_1), \dots, (C_n, C_n) \quad (1.1)$$

Bu durumda Z gibi bir gzlenen yařam sresi deđiřkeni tanımlanmıřtır

$$Z = \min(C_1, C_2), \quad \delta = I_{C_1 < C_2} \quad (1.2)$$

$$Z_i = \min(C_i, C_i), \quad \delta_i = I_{C_i < C_i} \quad (i = 1, 2, \dots, n) \quad (1.3)$$

ve veri seti

$$D_n = (C_1, \delta_1), \dots, (Z_n, \delta_n) \quad (1.4)$$

řekline dnřtrlmř, analizlerde eřitlik 1.4 de belirtilen veri seti kullanılmıřtır. Ayrıca Kaplan Meier yntemi sayesinde yařam srelerinin tahminlerinin eđriler yardımıyla grselleřtirilmesi de mmkn olmuřtur (6).

Watson and Leadbetter sansrsz gzlemler iin hazard fonksiyonu tahmin edicilerini bulmuřlardır (7,8). Ramlau-Hansen bu tahmin edicileri hazard

fonksiyonunda sansürlü gözlemler için genişletmişlerdir (9) ve  $H(\cdot)$  nin popüler bir tahmin edicisi bağımsız olarak Nelson ve Aalen tarafından bulunmuştur (10).

$$\hat{H}(t) = \sum_{j=1}^k \frac{\delta_j}{n - j + 1} \quad (1.5)$$

$t \leq t < t_{(k+1)}$   $k=1,2,\dots,r$  ve  $t_{(k)}$   $k$ . sıralı ölüm zamanıdır. Nelson-Aalen tahmin edicisi  $j$ . ölüm zamanındaki ölüm hazard oranı hakkında bilgi sağlamak amacıyla kullanılmaktadır.

Nonparametrik hazard tahmini için düzgünleştirme (smoothing) yöntemleri istatistik literatürlerinde oldukça yoğun ilgi görmektedir (11, 12, 13, 14, 15, 16, 17). Ayrıca sansürlü gözlemler içeren Kernel düzgünleştirme yönteminde bant genişliği belirleme teknikleri de araştırmacıların ilgisini çeken bir konu olmuştur (18, 19, 20, 21). Düzgünleştirme yöntemlerinin matematiksel özellikleri üzerine oldukça dikkatli çalışılmış olmasına karşın, istatistiksel özellikleri ile henüz yeterince ilgilenilmemiştir. Hess ve arkadaşları (1999) popüler hazard regresyon tahmin edicilerinin sansürlü gözlemlerin olduğu durumlarda farklı sansürlü gözlem oranları ile dağılımları için istatistiksel özelliklerini gelişmiş bir simülasyon çalışmasında çalışmışlardır (22). Literatürde sansürlü gözlemlerin miktarı ile ilgili bir çok simülasyon çalışması vardır ve hayatta kalma verilerinde sansürlü gözlemlerin oranlarının artmasının yaşam fonksiyonunun ve hazard oranının tahminini kötüleştirdiği bilinmektedir (22).

Genellikle sansürlü gözlemlerin veri seti içinde herhangi bir zamanda rasgele olarak oluşabileceği varsayılır. Fakat pratikte bazı özel nedenlerle sansürlü gözlemlerin rasgele olarak ortaya çıkmaması mümkün olabilir. Örneğin ağrılı bir operasyondan sonra kullanılan iki ağrı kesicinin karşılaştırılmak istendiği bir klinik denemede ağrı kesicilerden birinin kullanımından kısa bir süre sonra hastaların memnuniyetsizliği nedeni ile ikinci bir ağrı kesici verilmesi veya ağrı kesicinin dozunun artırılması durumunda aynı zaman aralığında birçok hastanın yeni ilaç kullanımı nedeni ile sansürlü gözlem olarak kabul edilmek zorunda kalması mümkündür. Bu tür durumlarla ilaç denemelerinde de sıklıkla karşılaşılabileceği gibi özellikle placebo kullanıldığı



durumlarda arařtırmacının karřılařabileceđi bir durumdur (23, 24). Rasgele olmayan sansürlü gözlemlere verilebilecek bir bařka örnek ise menopođ döneminde östrojen hormonu eksikliđine bađlı olarak ortaya çıkabilecek çeřitli Őikayetleri gidermek amacıyla kullanılan hormon replasman tedavisinin meme kanseri riskini arttırdıđına dair basında çıkan haberler dolayısıyla bir çok kadın hekime danıřmadan tedaviyi aynı dönemde bırakma eğilimine girecek ve dolayısıyla izlemde olan kadınların büyük bir kısmı aynı dönemde çalışmadan çıkacaktır.

Hayatta kalma analizlerinde sansürlü gözlemlerin rasgele olmayan bir dađılıma sahip olarak karřımıza çıkması mümkün olduđu halde literatürde sansürlü gözlemlerin veri seti içindeki yerleřiminin Kaplan Meier ve Hazard oranı tahminleri üzerine etkisinin arařtırıldıđı bir çalışma henüz mevcut deđildir. Bu tezde sansürlü gözlemlerin veri setinde farklı bölgelerde bulunduđu durumda hayatta kalma fonksiyonunun ve Hazard oranının tahminine etkisinin arařtırılması amaçlanmıřtır. Bu çalışmada farklı dađılım tipleri ve sansürlü gözlem miktarları için sansürlü gözlemlerin dađılım içinde yerleřim yerleri deđiřtirilmiř ve tahminler karřılařtırılmıřtır.

## 2. GENEL BİLGİLER

Hayatta kalma analizleri (survival analysis), sonuç değişkeninin belirli bir ilgilenilen olayın gerçekleşmesine kadar geçen süre olduğu durumda kullanılan analiz yöntemlerinin tamamıdır (1). Hayatta kalma analizlerinde süre, bireyin ya da bir nesnenin takibine başlanılmasından olayın gerçekleşmesine kadar geçen yıllar, aylar, haftalar ya da günler olabileceği gibi bir bireyin olayın gerçekleştiği andaki yaşı da olabilir (1). Ayrıca ilgilenilen olay da çalışmanın amacına göre farklılık gösterebilir. Ölüm, hastalık insidansı, hastanın remisyondan çıkarak tekrar kötüleşmesi, iyileşmesi olabileceği gibi bir birey ya da nesne için tanımlanmış herhangi bir olayda ilgilenilen olay olabilir (25, 26). Hayatta kalma analizlerinde zaman genellikle hayatta kalma süresi (survival time) olarak adlandırılırken, olay ise başarısızlık (failure) olarak adlandırılır. Çünkü genellikle klinik verilerde ilgilenilen olay, ölüm, hastalığın görülmesi ya da tekrarlanması gibi negatif olaylardır (27). Fakat bireylerin ciddi bir operasyondan sonra işe başlamalarına kadar geçen zaman ya da kanser hastalarında remisyon yaşam süresi olarak kabul edilebilir ve böyle bir durumda ilgilenilen olay pozitif bir durumun gerçekleşmesidir.

Hayatta kalma analizleri tıp, biyoloji, halk sağlığı, epidemiyoloji, mühendislik, ekonomi gibi birçok disiplinde kullanılmaktadır (28).

### 2.1. Neden Hayatta Kalma Analizleri?

Klinik araştırmalarda birçok amaçla denemeler planlanmakta ve ilgili değişkenlerin ölçüm tipine bağımlı ve bağımsız değişkenlerin sayısına ve araştırmanın amacına bağlı olarak farklı istatistiksel yöntemler kullanılmaktadır (29). Fakat hayatta kalma verilerinin sansürlü olma özelliklerinden dolayı, istatistikte yaygın olarak kullanılan yöntemlerle analiz edilmeleri uygun değildir. Hayatta kalma analizleri için özel yöntemler kullanılmasının başlıca nedenleri şöyle sıralanabilir:

1. Hayatta kalma verileri genellikle simetrik olmayan dağılıma sahiptirler. Bu nedenle birçok istatistik yöntemin temel varsayımlarından olan normal dağılım varsayımı yerine gelmez. Bu sorun verilerin daha simetrik dağılımlara

dönüştürülmesi ile çözülebilir ama orijinal verinin alternatif dağılım modelleri ile analiz edilmesi daha tatmin edici bir yaklaşımdır (1).

2. Sansürlü gözlemlerin varlığından dolayı hayatta kalma analizlerine standart yöntemleri uygulamak mümkün değildir (3).
3. Çalışmaya bireyler aynı anda girmiyor olabilir. Bu durumda da klasik istatistiksel yöntemlerin kullanılması mümkün değildir (3).

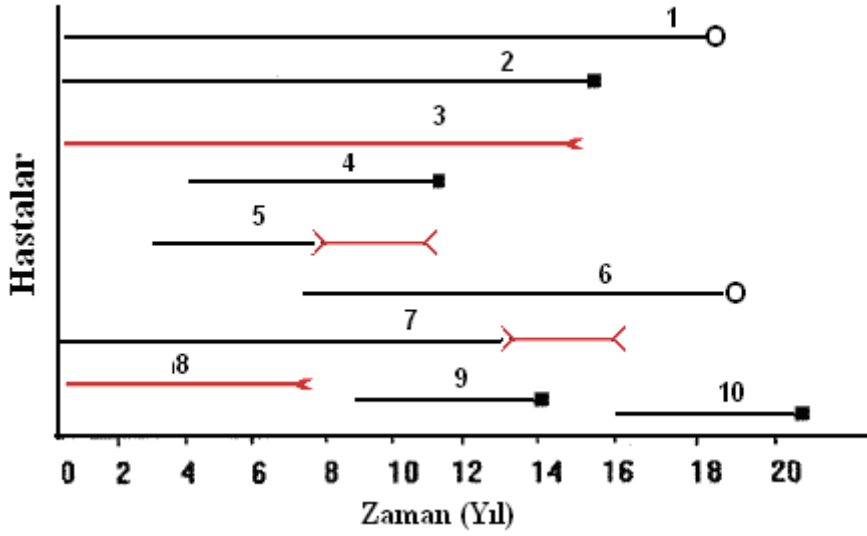
## 2.2. Sansürlü Veriler

Hayatta kalma analizlerinde çalışma bittiğinde, örneğin üzerinde durulan olayın “Ex olma” olduğu bir çalışmada bazı hastalar çalışmanın sonlanma zamanında halen hayatta olabilirler, çalışmadan kendi istekleri ile çekilebilirler ya da hastalar başka nedenlerden dolayı (başka bir hastalık dolayısıyla ölüm, taşınma... vb) çalışmayı terk edebilirler. Böylelikle bu hastaların ex olana kadar geçen sürelerinin gözlenmesi artık mümkün değildir. Yukarıda bahsedilen nedenlerden dolayı hayatta kalma süreleri değerlendirilemeyen gözlemler sansürlü gözlemler olarak adlandırılırlar (30).

Klinik çalışmalarda ortaya çıkabilecek birçok sansürlü gözlem tipi mevcuttur. Şekil 2.1’de hayatta kalma analizlerinde karşılaşılabilecek olası veri tipleri gösterilmiştir (25). Bu çalışmada ilgilenilen olayın ölüm olduğunu varsayalım Şekil 2.1 de “o” ile gösterilmiş hastalara ait veriler tamamlanmamış sansürlü veriler ve “■” ile gösterilen veriler “ex” olan hastalara ait verilerdir. Şekil 2.1 de 2 nolu hasta çalışmaya başlangıç noktasında girmiş ve 16. yılın sonunda ilgili hastalıktan dolayı “ex” olmuştur. Bu gözlem tamamlanmış bir gözlemdir. Ayrıca 4, 9 ve 10 nolu hastalar çalışmaya farklı zamanlarda girmiş fakat çalışılan hastalıktan dolayı gerçekleşen ölüm zamanı bilindiği için tamamlanmış verilerdir. Burada dikkat edilmesi gereken bir diğer nokta “ex” olan vakaların ölüm sebebinin mutlaka üzerinde çalışılan hastalıkla ilgili olmasıdır.

### 2.2.1. Sağdan Sansürlü Veriler

Veri analizi yapıldığı anda bazı gözlemlerde ilgilenilen olayın henüz gerçekleşmemesi durumunda oluşan sansür tipi sağdan sansürlü (right censored) olarak adlandırılır (4). Şekil 2.1de 1 nolu hasta çalışmaya başlangıç noktasından girmiş ve 19. yılın sonunda izlemden çıkmıştır. Hastanın “ex” olduğuna dair bir bilgi olmadığı için bu hasta sağdan sansürlü gözlem olarak tanımlanır. Ayrıca 6 nolu hasta çalışmaya 7. ayda girmiş ve 19. ayın sonuna kadar çalışmada bulunmuştur fakat çalışmada bulunduğu süre boyunca ölüm olayı gözlenmemiştir. Bu tür gözlemlerde sağdan sansürlü olarak adlandırılır. Klinik araştırmalarda en sık karşılaşılan sağdan sansürlü gözlemlerle karşılaşılır. Bu nedenle istatistik paket programlarında yaygın olarak sağdan sansürlü gözlemlerin analizlerine yer verilir (27,29,31,32).



Şekil 2.1. Hayatta kalma analizlerinde karşılaşılabilecek olası veri tipleri

Çalışmada belirlenen belirli bir zamanı aştığı için sansürlü olarak değerlendirilen sağdan sansürlü gözlemler Tip I (Type I) sansürü olarak adlandırılırken, çalışmaya başlamadan önce belirlenmiş sayıda olayın gerçekleşmesine kadar geçen zamanla sınırlandırılan sansürlü gözlem tiplerine Tip II (Type II) sansürü denir (33). Gözlemlerin araştırmacının kontrolü dışındaki nedenlerden dolayı sonlanması ile rasgele sansürlü (random censored) gözlemler oluşur.

### 2.2.2. Soldan Sansürlü Veriler

Soldan sansürlü (left censored) verilerde sadece belirli bir zamandan önce olayın gerçekleştiği bilinir (3). Bu tip sansürlü gözlemlerin bireyin olay deneyimine önceden sahip olduğu durumlarda ortaya çıkması muhtemeldir. 3 nolu hasta çalışmaya başlangıç noktasından girmiş, 15. yılsonuna kadar herhangi bir noktada “ex” olduğu bilinmektedir fakat ex olma zamanı ile ilgili kesin bir bilgi yoktur. Aynı şekilde 8 nolu hasta da çalışmaya başlangıç noktasında girmiş, fakat ölüm zamanı hakkındaki bilgi ölüm olayının başlangıç ile 8. yıl arasında olabileceğidir. Bu nedenle 3 ve 8 nolu hastalar soldan sansürlü gözlem olarak tanımlanır.

### 2.2.3. Aralık Sansürlü Veriler

Aralık sansürlü veri (Interval censored) istenen olayın sadece belirli bir aralık içinde gerçekleştiğinin bilindiği veri tipidir (34). Ölüm olayının 5 nolu hasta için 8–12 yıl aralığında, 7 nolu hasta içinse ise 13–17 yıl aralığında gerçekleştiği bilinmektedir. 5 ve 7 nolu gözlemler aralık sansürlü gözlemler olarak adlandırılır. HIV virüsü için bireylerin yıllık olarak test yaptıklarını düşünelim. 2. yılın sonunda enfeksiyon kapmamış bir bireyin 3. yılın sonundaki testte enfeksiyon kapmış görülürse enfeksiyon kapma zamanı 2 ve 3 yıl arasında aralık sansürlüdür.

## 2.3. Hayatta Kalma Analizlerinde Kullanılan Fonksiyonlar

Hayatta kalma değişkenlerinin dağılımı genel olarak üç farklı fonksiyon ile karakterize edebilir (1).

1. Hayatta kalma (Survival) fonksiyonu,
2. Olasılık yoğunluk fonksiyonu (Probability density function)
3. Risk oranı (Hazard rate) fonksiyonu,

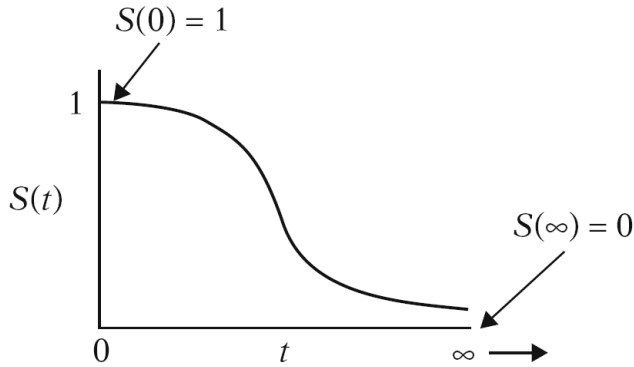
Bu fonksiyonlardan herhangi biri bilinmesi durumunda diğer ikisi tahmin edilebilir. Bu üç fonksiyondan en yaygın olarak kullanılan hayatta kalma fonksiyonudur.

### 2.3.1. Hayatta Kalma Fonksiyonu

T bireylerin hayatta kalma sürelerini gösteren rasgele değişken olmak üzere Hayatta kalma fonksiyonu  $S(t)$  bireylerin t zamanından daha uzun süre hayatta kalma olasılığını belirtir (33).

$$S(t) = P[T \geq t] \quad (2.1)$$

T, zaman belirten bir değişken olduğu için sıfıra büyük ve eşit tüm pozitif değerleri alabilir. Bu rasgele değişkenin herhangi bir zamanda aldığı değeri belirtmek içinse küçük harflerle “t” kullanılır.



Şekil 2.2 Hayatta kalma fonksiyonu

Teorik olarak hayatta kalma fonksiyonu 0 ile  $+\infty$  arasında değerler alan düzgün bir eğri şeklinde çizilebilir. Şekil 2.2de görüldüğü gibi zaman hayatta kalma fonksiyonunun alacağı değerleri belirler (1). Dik bir hayatta kalma eğrisi, düşük hayatta kalma oranını ya da kısa yaşam süresini gösterirken, yavaş iniş gösteren hayatta kalma eğrileri ise daha yüksek hayatta kalma oranını ya da uzun yaşam süresini gösterir (4).

Tüm hayatta kalma fonksiyonları özellikleri benzerdir.

Bunlar,

- Yaşam süresi  $t$  arttıkça fonksiyon azalır.
- $t = 0$  iken, yani çalışmanın başında iken  $S(t) = S(0) = 1$  dir. Çünkü henüz kimsede ilgilenilen olay gözlenmemiştir.
- $t = \infty$  iken  $S(t) = S(\infty) = 0$  değerini alır.

Teorik olarak çalışmanın süresi uzadıkça, bir noktada hayatta kalan birey kalmayacaktır ve hayatta kalma eğrisi sıfır değerini alacaktır (4).

Matematiksel olarak kümülatif dağılım fonksiyonu;

$$F(t) = P(t \text{ zamanından sonra hayatta kalma}) = p(T \leq t) \quad (2.2)$$

şeklinde ifade edilir ve hayatta kalma fonksiyonu kümülatif dağılım fonksiyonunun tümleyenidir.  $S(t) = P(t \text{ zamanından önce hayatta kalma})$  olmak üzere;

$$S(t) = 1 - F(t) \quad (2.3)$$

eşitliğinden yararlanılarak hesaplanabilir (28).

Ayrıca  $T$  sürekli rasgele değişken iken, hayatta kalma fonksiyonu olasılık yoğunluk fonksiyonunun integrali yardımıyla da elde edilebilirken (eşitlik 2.4) kesikli bir değişken iken ise hayatta kalma fonksiyonu toplam ile elde edilir (eşitlik 2.5).

$$S(t) = p(T > t) = \int_0^a f(t) dt \quad (2.4)$$

$$S(t) = p(T > t) = \sum_{x_j > x} p(t_j) \quad (2.5)$$

Sansürlü gözlemlerin olmadığı durumda hayatta kalma fonksiyonu (35);

$$\hat{S}(t) = \frac{t \text{ zamanından daha uzun süre hayatta kalan hasta sayısı}}{\text{toplam hasta sayısı}} \quad (2.6)$$

### 2.3.2. Olasılık Yoğunluk Fonksiyonu

T hayatta kalma zamanı için olasılık yoğunluk fonksiyonu,  $t$  ile  $t + \Delta t$  gibi kısa bir zaman aralığında istenen olayın gözlenmesi olasılığını verir (35).

$$f(t) = \frac{\lim_{\Delta t \rightarrow 0} P[(t, t + \Delta t) \text{ aralığında bir bireyin ölmesi}]}{\Delta t} \quad (2.7)$$

$f(t)$  nin grafiği yoğunluk fonksiyonu eğrisi olarak adlandırılır ve yoğunluk fonksiyonunun iki özelliği vardır (35):

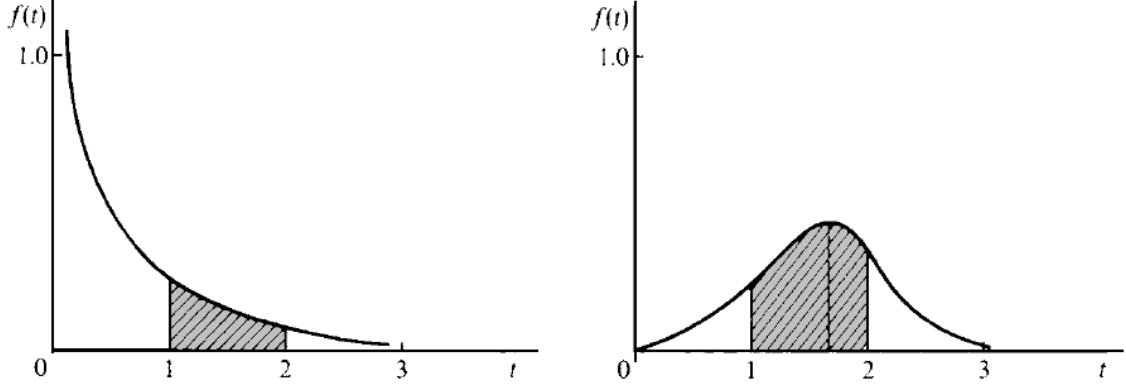
1.  $f(t)$  negatif olmayan bir fonksiyondur.

$$f(t) = \begin{cases} \geq 0 & t > 0 \\ 0 & t = 0 \end{cases} \quad (2.8)$$

2. Olasılık yoğunluk eğrisi ile t ekseninde kalan alan 1 e eşittir.



Şekil 2.3’de olasılık yoğunluk fonksiyonları örnek olarak verilmiştir.



Şekil 2.3 Olasılık yoğunluk fonksiyonu örnekleri

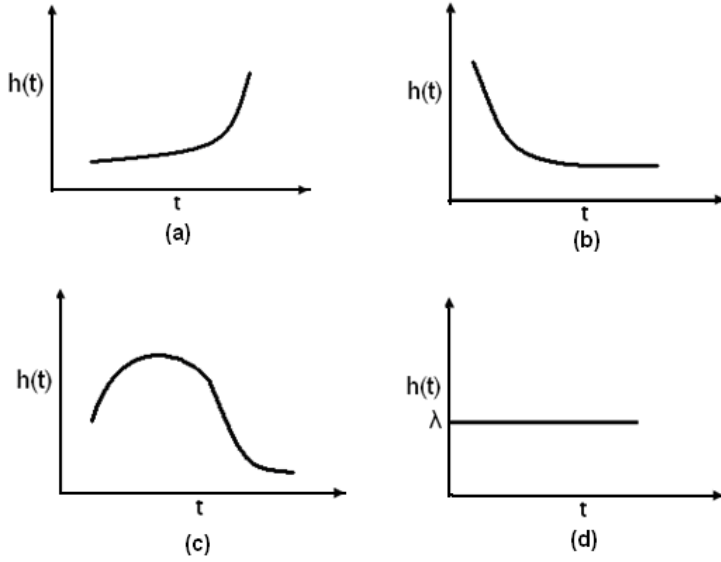
### 2.3.3. Hazard Fonksiyonu

$T$  hayatta kalma zamanının  $h(t)$  hazard fonksiyonu koşullu başarısızlık oranını (conditional failure rate) verir. Aralığın başında hayatta olan bireylerde kısa bir zaman aralığı içinde ilgilenilen olayın gözlenmesi olasılığıdır (33).

$$h(t) = \frac{\lim_{\Delta t \rightarrow 0} P[t \text{ zamanına kadar hayatta kalmış bir bireyde } (t, t + \Delta t) \text{ aralığında olayın gözlenmesi}]}{\Delta t} \quad (2.9)$$

Hazard fonksiyonu, artan, azalan ve sabit gibi birçok farklı şekilde gözlenebilir. Sıklıkla karşılaşılabilecek bazı hazard fonksiyonları Şekil 2.4’de verilmiştir (36). Tedaviye cevap vermeyen lösemi hastaları için hazard fonksiyonu, ölüm oranı zaman geçtikçe artacağı için artan bir yapıda (Şekil 2.4 (a)) iken ameliyattan sonra hastaların iyileşmesi için hazard fonksiyonu ölüm oranı zaman geçtikçe azalacağı için (Şekil 2.4 (b)) azalan yapıda bir hazard fonksiyonu gözlemlenir. Tüberküloz hastaları için ölüm oranı hastalığın başlarında yüksektir ve kritik dönem atlatıldıktan sonra azalır (Şekil 2.4

(c). Ve bir kaza sonucu ölen bireyler için Hazard fonksiyonu. ölüm oranı zaman boyunca sabittir (Şekil 2.4 (d)).



Şekil 2.4 Hazard oranı için örnek grafikler.

Kümülatif hazard fonksiyonu;

$$H(t) = \int_0^t h(x)dx \quad (2.10)$$

şeklinde tanımlanır (28). Kümülatif hazard fonksiyonu 0 ile 1 arasında herhangi bir değeri alabilir. Ve survival fonksiyonundan aşağıdaki eşitlik yardımı ile elde edilebilir.

$$H(t) = -\log S(t) \quad (2.11)$$

Bu nedenle;  $H(t) = \begin{cases} 0, & t = 0 \\ 1, & t = \alpha \end{cases}$  (2.12)

#### 2.4. Hayatta Kalma Fonksiyonları Arasındaki İlişkiler

Hayatta kalma değişkenlerinin dağılımlarının karakterize edilebileceği üç farklı fonksiyon arasında bazı matematiksel ilişkiler vardır ve bu ilişkiler yardımı ile fonksiyonlardan birini bilindiği durumda diğer fonksiyonları elde etmek mümkündür.

Bu ilişkilere matematiksel olarak göz atacak olursak;

- Hazard fonksiyonu olasılık yoğunluk fonksiyonunun hayatta kalma fonksiyonuna oranı ile elde edilebilir (37).

$$h(t) = \frac{f(t)}{S(t)} \quad (2.13)$$

- Olasılık yoğunluk fonksiyonu kümülatif dağılım fonksiyonunun türevidir (37).

$$f(t) = \frac{d}{dt}[1 - S(t)] = -S'(t) \quad (2.14)$$

- 2.13 nolu eşitlikte 2.14 nolu eşitliği yerine koyduğumuzda (35);

$$h(t) = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \log S(t) \quad (2.15)$$

- 2.15 nolu eşitliğin 0 dan t zamanına kadar olan integralini alacak olursak (35);

$$-\int_0^t h(x)dx = \log S(t) \quad (2.16)$$

ya da  $H(t) = -\log S(t)$  (2.17)

ya da  $S(t) = \exp[-H(t)] = \exp[-\int_0^t h(x)dx]$  (2.18)

- Eşitlik 2.17 ve 2.21'den;

$$f(t) = h(t) \exp[-H(t)] \quad (2.19)$$

Eşitliği elde edilir (37).

Çizelge 2.1'de bazı dağılımlara ait hayatta kalma, olasılık yoğunluk ve hazard fonksiyonları verilmiştir (28).

Çizelge 2.1 Bazı dağılımlara ait hayatta kalma, olasılık yoğunluk ve hazard fonksiyonları			
Dağılım	Hazard Oranı $h(x)$	Hayatta Kalma Fonksiyonu $S(x)$	Olasılık Yoğunluk Fonksiyonu $F(x)$
Üstel	$\lambda$	$e^{-\lambda x}$	$\lambda e^{-\lambda x}$
Weibull	$\alpha \lambda x^{\alpha-1}$	$e^{-\lambda x^\alpha}$	$\alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha}$
Log logistic	$\frac{\alpha \lambda x^{\alpha-1} \lambda}{1 + \lambda x^\alpha}$	$\frac{1}{1 + \lambda x^\alpha}$	$\frac{\alpha \lambda x^{\alpha-1} \lambda}{(1 + \lambda x^\alpha)^2}$

## 2.5. Hayatta Kalma Fonksiyonunun Parametrik Olmayan Yöntemler ile Tahmini: Kaplan Meier Yöntemi

Sansürlü gözlemler içeren bir veri setine ait hayatta kalma fonksiyonunu tahmin etmek için en yaygın kullanılan yöntem Kaplan Meier yöntemidir. Kaplan ve Meier 1958 yılında “product limit tahmini” olarak da bilinen bu yöntemi bulmuşlardır (6). Herhangi bir kovaryant olmadığı durumda yaşam sürelerinin  $T$  ve sansürlü gözlem değerleri  $C$  lerin birbirinden bağımsız olduğunu varsayarsak (2) ,

$$\delta_i = \begin{cases} T \leq C_{ii} & \text{ise } 1 \\ T > C & \text{ise } 0 \end{cases} \quad (2.20)$$

$\delta_i$  sansürlü gözlemlerin gösterilmesinde kullanılır ve sansür indikatörü olarak adlandırılır. 1 değerini alması durumunda gözlemler sansürlüdür. Bir birey için ilgilenilen yaşam süresi ve sansürlü gözlem değerlerinden en küçük olanı gerçekte gözlemlediğimiz değerdir. Bu nedenle veri setimiz gözlemlediğimiz yaşam süreleri  $Z_i = \min(T_i, C_i)$  ve sansürlü olma durumunu gösteren  $\delta_i$  değerlerini içeren  $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$  gözlem çiftlerine dönüşür (6). Böyle bir popülasyondan  $n$  bireylik bir örnek çaktığımızı varsayalım. Gözlemlerimizin hayatta kalma süreleri sağdan sansürlü gözlemlere bağlı olacaktır. Bu şu demektir; bazı bireyler için sadece gerçek hayatta kalma zamanları belli bir sansür zamanını aştığı bilinmektedir. Sansürlü gözlemlerin birbirlerinden bağımsız olduğu varsayılır. Bu nedenle  $t$  zamanından önce herhangi zamanda oluşmuş olan sansürlü gözlemler  $t$  zamanındaki olayın gözlenme riskini değiştirmez.

$F(\cdot)$  ve  $G(t)$  sırasıyla yaşam sürelerine ve sansürlü gözlemlere ait birikimli dağılım fonksiyonları olmak üzere gerçekte gözlemlediğimiz yaşam sürelerine ait fonksiyonu bu iki fonksiyonun çarpımı yardımı ile elde edilir (2).

$$K_{\mathbf{C}} = P(Z > t) = F(t).G(t) \quad (2.21)$$

F ve G fonksiyonları ise Kaplan ve Meier product limit tahmini yardımı ile tahmin edilir.  $F_n$  ve  $G_n$  sırasıyla F ve G nin Kaplan-Meier tahminleri olmak üzere bu fonksiyonların tahminlerinde kullanılan eşitlikler aşağıdaki gibi tanımlanır (38).

$(Z_{(i)}, \delta_{(i)})$  ( $i = 1, \dots, n$ ) n tane gözlenen veri çifti Z ye göre sıralıdır,

$$Z_{\mathbf{C}} \leq Z_{\mathbf{C}} \leq \dots \leq Z_{(n)} : T_{K_n}$$

$$F_n_{\mathbf{C}} = \begin{cases} \prod_{i: Z_{\mathbf{C}} \leq t} \left( \frac{n-i}{n-i+1} \right)^{\delta_{\mathbf{C}}} & \text{eger } t < Z_{(n)} \\ 0 & \text{dd} \end{cases} \quad (2.22)$$

$$G_n_{\mathbf{C}} = \begin{cases} \prod_{i: Z_{\mathbf{C}} \leq t} \left( \frac{n-i}{n-i+1} \right)^{1-\delta_{\mathbf{C}}} & \text{eger } t < Z_{(n)} \\ 0 & \text{dd} \end{cases} \quad (2.23)$$

$F_n(t)$  sansürlü olmayan gözlem noktalarında sıçrarken benzer şekilde  $G_n(t)$  sansürlü gözlem noktalarında sıçrar.

Büyük örneklerde verilen bir  $t$  zamanında Kaplan Meier tahmincisi, yaklaşık olarak normal dağılıma sahiptir. Bu nedenle  $S_{\mathbf{C}}$  için  $100(1-\alpha)\%$  standart güven aralığı,

$$\hat{S}(t) \pm z_{1-\alpha/2} \hat{\sigma}(t) \quad (2.24)$$

Eşitlik 2.24'de belirtilen güven aralığı belirleme yöntemi küçük örnek genişlikleri için tatmin edici sonuçlar vermektedir.

Kaplan Meier tahmin edicisinin kullanım alanı sadece verilen  $t$  zamanı için hayatta kalma olasılıklarının tahmin edilmesi değildir; Medyan hayatta kalma (Medyan survival time) süresi gibi bazı istatistiklerin hesaplanmasında da kullanılabilir ya da Kaplan Meier eğrisi altında kalan alandan yararlanılarak ortalama hayatta kalma süresi (mean survival time) hesaplanabilir (37).

Kaplan Meier yöntemi bireyleri alt gruplara ayırmadan yaşam süresi tahmini yapan bir yöntemdir. Kaplan Meier yöntemi yaşam tablolarının analizlerine göre daha az birey içerir çünkü yaşam süresi bir hastanın öldüğü her zaman için hesaplanırken çalışmadan çekilen bireyler göz ardı edilir (3).

### **2.5.1. Kaplan Meier Yönteminin Bazı Kritik Özellikleri**

Kaplan Meier yöntemi gerek hayatta kalma olasılıklarının hesaplanması gerekse grafiksel olarak yaşam sürelerinin dağılımının ortaya konması amacıyla hayatta kalma analizlerinde yaygın olarak kullanılan bir yöntemdir. Breslow ve Crowley 1974 yılında Kaplan Meier tahminlerinin tutarlı olduğunu ve asimtotik olarak normal dağılıma sahip olduğunu göstermişlerdir. Buna karşın Kaplan Meier tahminlerinin bazı kritik özelliklerinden de bahsetmek gereklidir (1, 3, 35, 38, 39).

1. Kaplan Meier tahminleri verilerin aralığı içinde geçerlidir. Eğer en büyük gözlem tamamlanmış veri ise Kaplan Meier tahmini bu noktada sifıra eşittir. Eğer en büyük gözlem sansürlü ise, Kaplan Meier tahmini hiçbir zaman sıfır olamaz.
2. Hayatta kalma analizlerinde en yaygın olarak kullanılan istatistik medyan survivaldir. Medyanın hayatta kalma eğrisinden okunmasının en kolay yolu  $\hat{S}(t) = 0.5$  olduğu noktadaki  $t$  zamanına bakılarak karar vermektir. Bazı durumlarda birden fazla  $t$  zamanı olabilir. Bu durumda da aralığın orta noktasını alarak medyan hayatta kalma zamanı bulunabilir.
3. Gözlemlerin 0.50 sinden fazlasının sansürlü olması ve en büyük gözlemin sansürlü olması durumunda medyan hayatta kalma süresi tahmin edilemez.

4. Kaplan Meier yöntemi sansürlü gözlemler ile yaşam sürelerinin birbirinden bağımsız olduklarını varsayarlar. Bir başka ifadeyle gözlemlerin sansürlü olma nedeni ölüm nedeninden bağımsızdır.

## 2.6. Kernel Tahmin Yöntemi

Kernel tahmin yöntemi parametrik olmayan fonksiyonların tahmininde yaygın olarak kullanılan bir yöntemdir.  $F$  gibi bir fonksiyonun Kernel tahmin edicisi  $x$  gibi bir noktada  $[x-b, x+b]$  gibi bir aralıkta verilerin bölgesel ağırlıklandırılmış ortalamasıdır.  $B$ , Kernel tahmininde bant genişliği (bandwidth) olarak adlandırılır. Kernel tahminindeki kritik faktör bant genişliğinin belirlenmesidir. Bant genişliği yapılacak tahminlerin ne kadar düzgün (smoothness) olacağını etkiler. Büyük bant genişliğinin belirlenmesi durumunda tahminler daha düzgün olacaktır, düzgünlüğün yüksek olması daha düşük varyasyona dolayısıyla yanlılığın artmasına neden olacaktır (13,14,15,16).

Sansürlü gözlemler için hazard tahmininde Kernel tahmin yönteminin kullanılması ilk defa Ramlau-Hansen tarafından gündeme getirilmiş ve diğer araştırmacılar tarafından sabit bir bant genişliği için sağdan sansürlü gözlemlerin hazard fonksiyonunun tahmininde kullanılmak üzere aşağıdaki tahmin edici tanımlanmıştır (17).  $K(\cdot)$  Kernel ya da ağırlıklandırma fonksiyonu,  $b$  bant genişliği olmak üzere;

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t-t_{(i)}}{b}\right) \frac{\delta_i}{n-i+1} \quad 0 < t_{(1)} < t_{(2)} < \dots < t_{(n)} \quad (2.25)$$

### 2.6.1. Kernel ağırlıklandırma fonksiyonları

B beta fonksiyonu olmak üzere



$$B(a,b) = \Gamma(a)\Gamma(b)\Gamma(a+b) \quad (2.26)$$

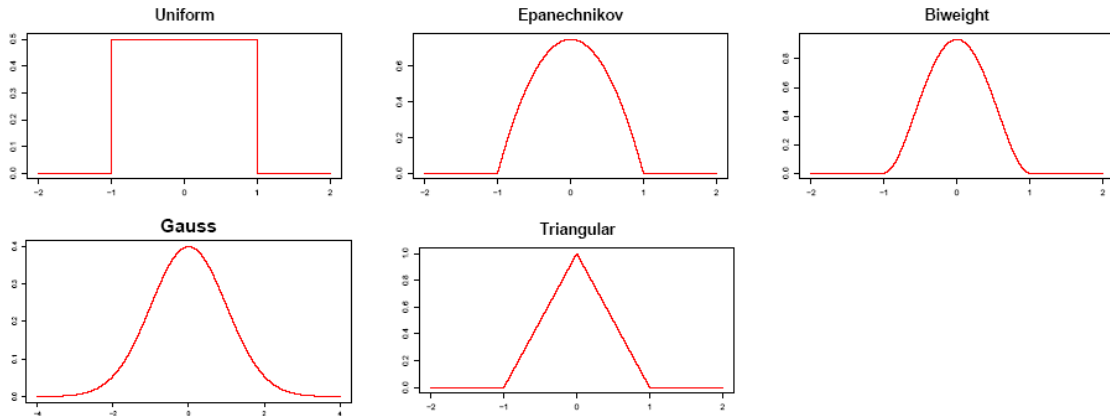
$$K(x,p) = \frac{(1-x^2)^p}{2^{2p+1}B(p+1,p+1)} \mathbb{1}_{|x| \leq 1} \quad (2.27)$$

P=0 iken uniform Kernel, p=1 iken Epanechnikov Kernel, p=2 iken Biweight Kernel, p=3 iken Triweight Kernel elde edilir. Bu Kerneller [-1,1] aralığında simetrik beta yoğunluklarıdır. I indikatör fonksiyonunu göstermektedir.

Çizelge 2.2’de bazı Kernel fonksiyonları ve Şekil 2.5 de Kernel fonksiyonlarının grafikleri verilmiştir (11).

Çizelge 2.2 Kernel ağırlıklandırma fonksiyonları

Kernel	K(x)	Kernel	K(x)
Uniform	$\frac{1}{2} I( u  \leq 1)$	Quartic	$\frac{15}{16} (1-u^2)^2 I( u  \leq 1)$
Triangle	$(1- u ) I( u  \leq 1)$	Triweight	$\frac{35}{32} (1-u^2)^3 I( u  \leq 1)$
Epanechnikov	$\frac{3}{4} (1-u^2) I( u  \leq 1)$	Gaussian	$-\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right)$



Şekil 2.5 Bazı Kernel fonksiyonlar

Kernel fonksiyonlarının etkinlikleri Ortalama bütünleşmiş edilmiş karesel hata (MISE, mean integrated squared error) ve Asimtotik bütünleşmiş edilmiş karesel hata (AMISE, asymptotic MISE) ile ölçülür. Epanechnikov Kernel AMISE i minimize eder dolayısıyla optimal olarak değerlendirilir. Diğer Kernel fonksiyonlarının etkinlikleri epanechnikov Kernel fonksiyonuna göre değerlendirilir. Yapılan çalışmalar sonunda Epanechnikov Kernelin etkinliği 1.000 olarak kabul edildiği durumda, Biweight Kernelin etkinliği 0.994, Triangular Kernelin etkinliği 0.986 ve Normal Kernelin etkinliği 0.951 bulunmuştur (19).

Fakat Kernelin seçiminin Kernel tahminleri üzerine etkisi bant genişliğinin seçimi kadar etkili değildir.

### 2.6.2. Bant genişliğinin belirlenmesi

Literatürde sansürlü gözlemlerin varlığı durumunda kullanılan Kernel düzgünleştirme yöntemi için birçok bant genişliği belirleme yönteminden bahsedilmiştir (13, 18, 20, 23, 40).

### 2.6.3. Tahmin edicinin başarısının test edilmesi

Regresyon analizinde  $Y$  gibi bir fonksiyonun tahmini olan  $f(x)$  fonksiyonunun gerçek fonksiyonun iyi bir tahmini, bir başka ifadeyle  $|f(x) - Y|$  değerinin küçük olması amaçlanır. Fakat bu farkın ne kadar küçük olduğu açık değildir. Bu problem  $L_2$  hatası ya da  $f(x)$  fonksiyonun hata kareler ortalaması yardımıyla çözülebilir.  $L_2$  Hatası kolay ve hızlı bir şekilde hesaplanabilme özelliğinden dolayı yaygın olarak tahminlerin başarısının gösterilmesinde kullanılmaktadır (41, 42).

$$L_2 = \frac{(Y - f(x))^2}{n} \quad (2.28)$$

## 2.7. Bazı Parametrik Hayatta Kalma Dağılımları

### 2.7.1. Üstel Dağılım

Hayatta kalma analizlerindeki en basit ve önemli dağılım üstel (exponential) dağılımdır. Üstel dağılımı, sabit bir hazard oranı ( $\lambda$ ) ile tanımlanır ve bir tek parametresi vardır.  $\lambda$  değerinin yüksek olması yüksek risk ve kısa yaşam süresi anlamına gelirken, düşük bir  $\lambda$  düşük riski ve uzun bir yaşam süresini gösterir (26).

$T$  yaşam sürelerinin  $\lambda$  parametrelili üstel dağılıma sahip olduğu durumda, Olasılık yoğunluk fonksiyonu (26),

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \lambda > 0 \\ 0 & t < 0 \end{cases} \quad (2.29).$$

Birikimli dağılım fonksiyonu,

$$F(t) = 1 - e^{-\lambda t} \quad t \geq 0 \quad (2.30).$$

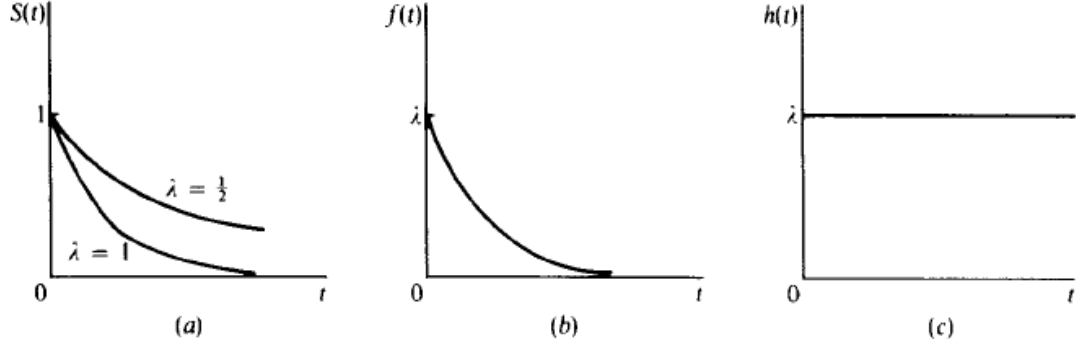
Hayatta kalma fonksiyonu,

$$S(t) = e^{-\lambda t} \quad t \geq 0 \quad (2.31).$$

Hazard fonksiyonu,

$$h(t) = \lambda \quad t \geq 0 \quad (2.32).$$

Şekil 2.6 da üstel dağılıma ait hayatta kalma fonksiyonu, yoğunluk fonksiyonu ve hazard fonksiyonu  $\lambda$  değerine bağlı olarak verilmiştir (33).



Şekil 2.6 Üstel dağılıma ait hayatta kalma fonksiyonu (a), olasılık yoğunluk fonksiyonu (b), hazard fonksiyonu (c)

Üstel fonksiyon sabit hazard oranı ile karakterize edildiği için bireyin yaşından ve yıpranmadan bağımsızdır. Bu nedenle istenen olayın gözlenmesi zamandan bağımsızdır. Üstel dağılımın hayatta kalma fonksiyonunun doğal logaritması alındığında oluşan fonksiyon doğrusaldır. Üstel dağılımın ortalaması  $1/\lambda$  ve varyansı  $1/\lambda^2$  dir (33).

### 2.7.2. Düzgün Dağılım

Düzgün dağılım (uniform distribution), tüm  $x$  değerleri için bir olayın görülme olasılığının eşit olduğu bir dağılımdır.  $A$  ve  $b$  gibi iki parametresi vardır.  $A=0$ ,  $b=1$  olduğu durumda düzgün dağılım, standart düzgün dağılım olarak adlandırılır (34).

$T$  yaşam sürelerinin  $[a - b]$  aralığında düzgün dağılıma sahip olduğu durumda olasılık yoğunluk fonksiyonu,

$$f(t) = \begin{cases} \frac{1}{b-a} & a < t < b \\ 0 & \text{diğer} \end{cases} \quad (2.33)$$

Birikimli dağılım fonksiyonu,

$$F(t) = t \quad 0 \leq t \leq 1 \quad (2.34).$$

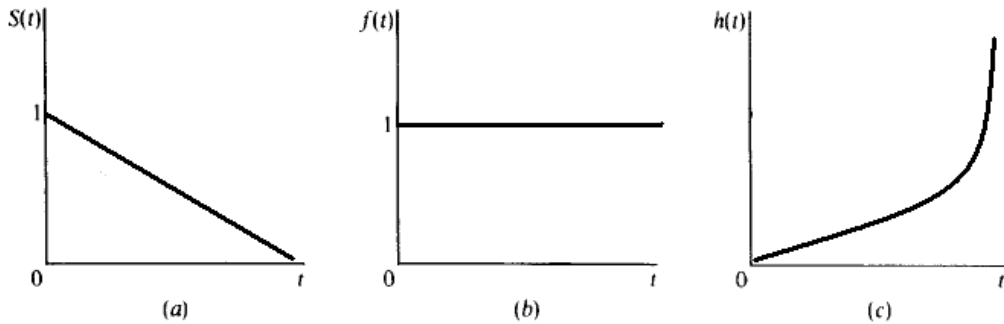
Hayatta kalma fonksiyonu.

$$S(t) = 1 - t \quad t \geq 0 \quad (2.35).$$

Hazard fonksiyonu,

$$h(t) = \frac{1}{1-t} \quad (2.36).$$

Şeklinde (1). Eğer düzgün dağılım eşit aralıklara ayrılırsa her aralığa düşen birey sayısı eşit olacaktır. Şekil 2.7’de düzgün dağılıma ait hayatta kalma fonksiyonu, yoğunluk fonksiyonu ve hazard fonksiyonu  $\lambda$  değerine bağlı olarak verilmiştir.



Şekil 2.7 Düzgün dağılıma ait hayatta kalma fonksiyonu (a), olasılık yoğunluk fonksiyonu (b), hazard fonksiyonu (c)

### 3. GEREÇ VE YÖNTEM

Bu çalışmada farklı dağılımlar ve sansürlü gözlem miktarları için sansürlü gözlemlerin dağılım içinde yerleşim yerleri değiştirilerek, hayatta kalma fonksiyonu ve hazard oranı tahminlerine sansürlü gözlemlerin veri içindeki yerleşiminin etkisi araştırılmıştır. Sonuçların karşılaştırılabilmesi amacıyla Hess K.R, ve arkadaşları tarafından yayınlanmış olan makalede tanımlanmış olan dağılımlardan faydalanılmıştır (22). Sabit, lineer artan ve lineer azalan polynomial hazard şekilleri kullanılmıştır.  $h(t) = \lambda_0 h_0(t)$  hazard fonksiyonunda  $n=100$  için  $S(t=90)$  0.1 ve  $n=250$  için  $S(t=90)$  0.4 değerlerinden faydalanarak  $\lambda_0$  değerleri hesaplanmıştır. Bu değer  $t=90$  zamanında 10 hastanın risk altında olması demektir. Sabit hazard için  $h_0(t)=1$ , lineer artış gösteren hazard için  $h_0(t) = at + b$  hazard fonksiyonunun  $a$ ,  $b$  değerleri  $h_0(0) = 0$  ve  $h_0(100) = 1$  olacak şekilde belirlenmiştir. Lineer azalış gösteren hazard için ise  $h_0(t) = at + b$  hazard fonksiyonunun  $a$ ,  $b$  değerleri  $h_0(0) = 1$  ve  $h_0(100) = 0$  olacak şekilde belirlenmiştir.

Simülasyonun bir sonraki aşamasında standart uniform dağılıma sahip rasgele  $u$  değerleri üretilmiştir. Ve bu  $u$  değerleri hazard fonksiyonunun integrali yardımıyla elde edilmiş, hayatta kalma fonksiyonun tersinde  $t_i = S^{-1}(u_i)$  yerine konulmuştur. Sansürlü gözlemlerin yerinin ve oranının değiştirilebilmesi için uniform dağılımdan faydalanılmıştır.

Bilinen fonksiyonların sansürlü gözlemlerin varlığında tahmin edilebilirliğini ortaya koymak ve aynı zamanda sansürlü gözlemlerin veri seti içinde yerleşiminin tahminleri etkileyip etkilemediğini ortaya koymak amacıyla Kaplan Meier tahmini yardımı ile hayatta kalma fonksiyonları tahmin edilmiştir.  $L_2$  Hatası kolay ve hızlı bir şekilde hesaplanabilme özelliğinden dolayı yaygın olarak tahminlerin başarısının gösterilmesinde kullanılmaktadır (41, 42). Bu nedenle tezde yapılan simülasyon tahminlerinin başarısının ortaya konulmasında  $L_2$  hatası kullanılmıştır.

Kaplan Meier tahmininde kullanılan dağılımlar için hazard fonksiyonları da tahmin edilmiştir. Hazard fonksiyonu tahminlerinde Kernel düzgünleştirme yöntemi ve

Kernel fonksiyonu olarak ise Epanechnikov Kernel kullanılmıştır. Bant genişliği olarak  $L_2$  hatasını minimum yapan bant genişliği kullanılmıştır.

Çizelge 3.1 Simülasyonda kullanılan dağılımlara ilişkin hayatta kalma fonksiyonları ve fonksiyonların tersleri

	$S(t)$	$S^{-1}(t)$
Sabit Hazard	$S(t) = \exp(-\lambda_0 t)$	$S^{-1}(t) = -\frac{\ln(t)}{\lambda_0}$
Lineer Artan	$S(t) = \exp\left(-\frac{\lambda_0 t^2}{200}\right)$	$S^{-1}(t) = \sqrt{-\frac{200 \ln(t)}{\lambda_0}}$
Lineer Azalan	$S(t) = \exp\left(-\lambda_0 \left(t - \frac{t^2}{200}\right)\right)$	$S^{-1}(t) = 100 - \sqrt{\frac{200 \ln(t) + 10000}{\lambda_0}}$

Çizelge 3.2 Simülasyonda kullanılan dağılımlara ilişkin hazard fonksiyonu ve katsayıları

Hazard Oranı	$h(t) = h_0 \lambda_0(t)$	$n = 100$	$n = 250$
Sabit Hazard	$h(t) = \lambda_0$	$\lambda_0 = 0.0256$	$\lambda_0 = 0.0357$
Lineer Artan	$h(t) = \frac{\lambda_0 t}{100}$	$\lambda_0 = 0.0568$	$\lambda_0 = 0.0794$
Lineer Azalan	$h(t) = \begin{cases} \lambda_0 \left(1 - \frac{t}{100}\right) & t \leq 100 \\ \lambda_0 & else \end{cases}$	$\lambda_0 = 0.04651$	$\lambda_0 = 0.0650$
		$S(t = 90) = 0.1$	$S(t = 90) = 0.04$

Tezin simülasyon çalışmasında 0.20, 0.30, 0.40 ve 0.50 sansür yüzdeleri için, aynı dağılıma sahip veri setinin içinde sansürlü gözlemler 4 farklı bölgeye yerleştirilmiştir. Bu yerleşimler grafiklerde aşağıdaki şekilde ifade edilmiştir;

**Rasgele:** Sansürlü gözlemler veri setinin içinde rasgele olarak yerleştirilmiş.

**I. Bölge:** Sansürlü gözlemlerin tamamı çalışmanın başlangıcında oluşturulmuş.

**II. Bölge:** Sansürlü gözlemlerin tamamı çalışmanın ortalarına doğru oluşturulmuş.

**III. Bölge:** Sansürlü gözlemlerin tamamı çalışmanın sonuna doğru oluşturulmuş.

Simülasyonlar R programlama dilinde yazılmıştır (43). Sansür yerleşim grafiğinde tamamlanmış veriler, T harfi ve kırmızı renkli kesikli çizgi ile (—) sansürlü gözlemler ise S harfi ve mavi renkli kesikli çizgi (—) gösterilmiştir.

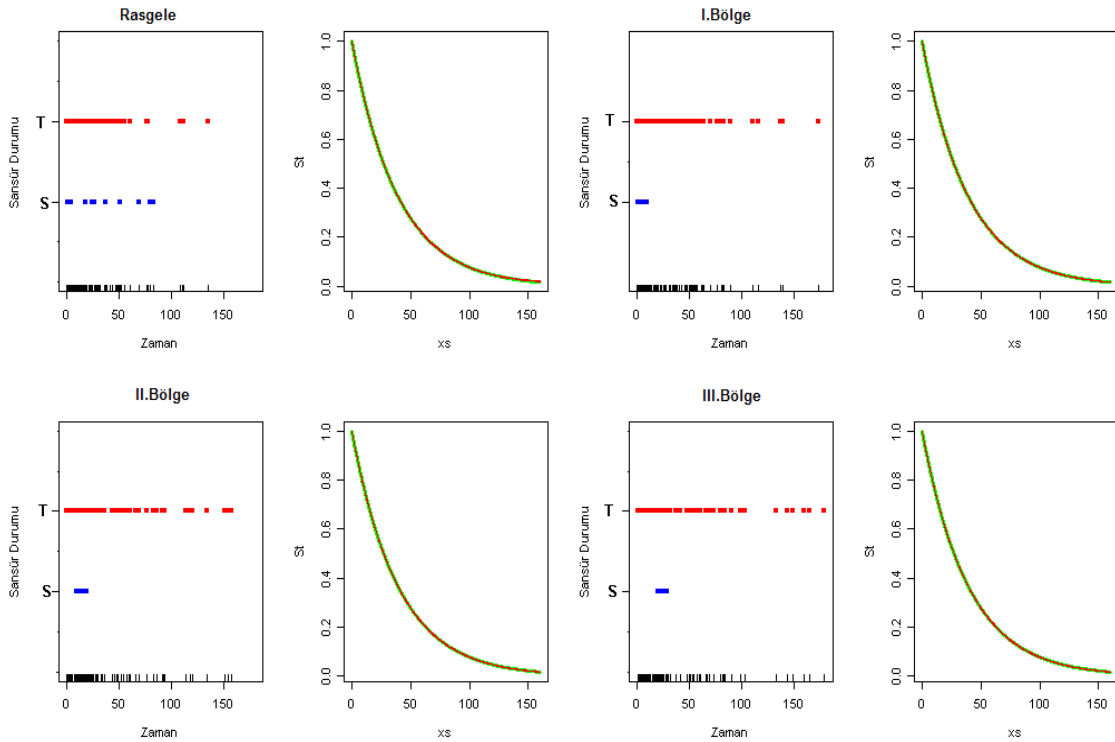
Kaplan Meier ve Kernel tahminlerinin başarılarını göstermek amacıyla 1000 tekrarlı olarak programlar çalıştırılmış ve her tahmin noktası için tahminlerin ortalaması alınmıştır. Bilinen gerçek hayatta kalma ve hazard fonksiyonları ile tahmin edilen hayatta kalma ve hazard fonksiyonlarının karşılaştırılabilmesi amacıyla gerçek ve ilgili tahminler aynı grafikte verilmiştir. Gerçek fonksiyonlar yeşil düz çizgilerle (—) gösterilirken tahmin fonksiyonlarının gösteriminde kırmızı düz çizgi kullanılmıştır(—).



## 4. BULGULAR

### 4.1. Sabit Hazard Oranı, N=100 için Kaplan Meier Tahminleri

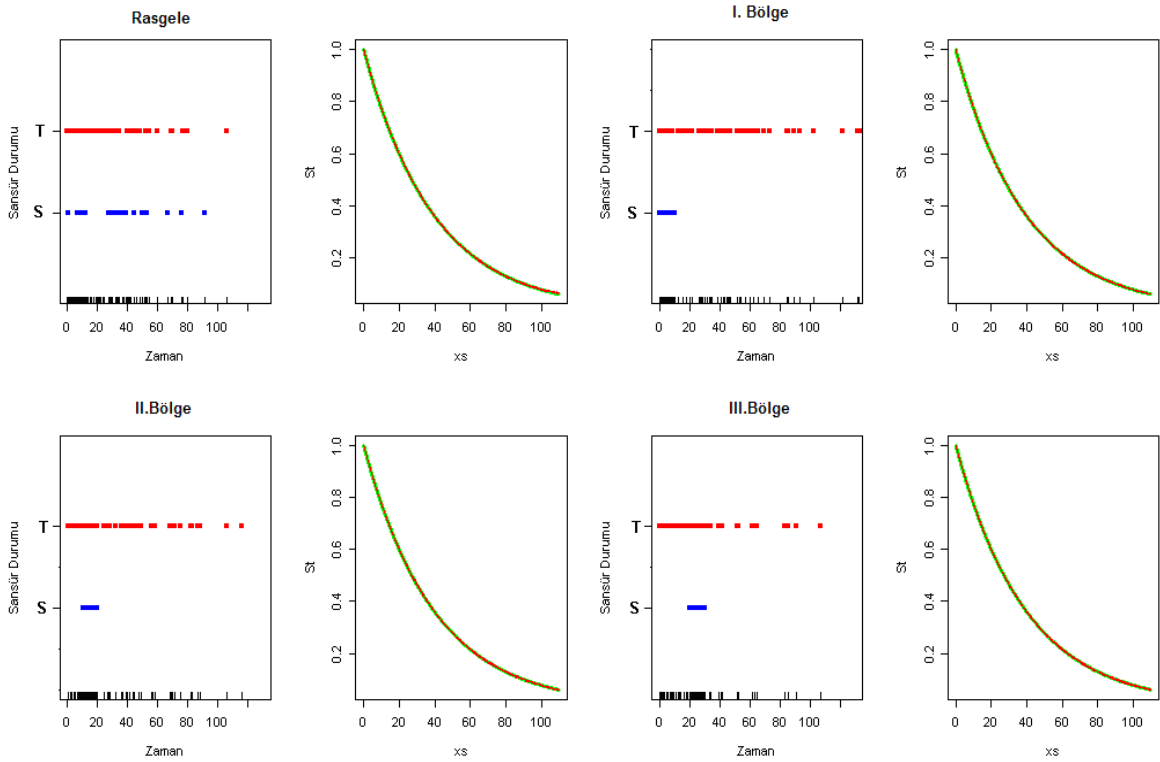
N=100 iken sabit hazard oranına ait simülasyonda sırasıyla 0.20 sansür oranına ait tahminler Şekil 4.1’de, 0.30 sansür oranına ait tahminler Şekil 4.2’de, 0.40 sansür oranına ait tahminler Şekil 4.3’de ve 0.50 sansür oranına ait tahminler Şekil 4.4’de, görüldüğü gibidir. Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş olan hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde oldukça başarılı olduğu görülmektedir.



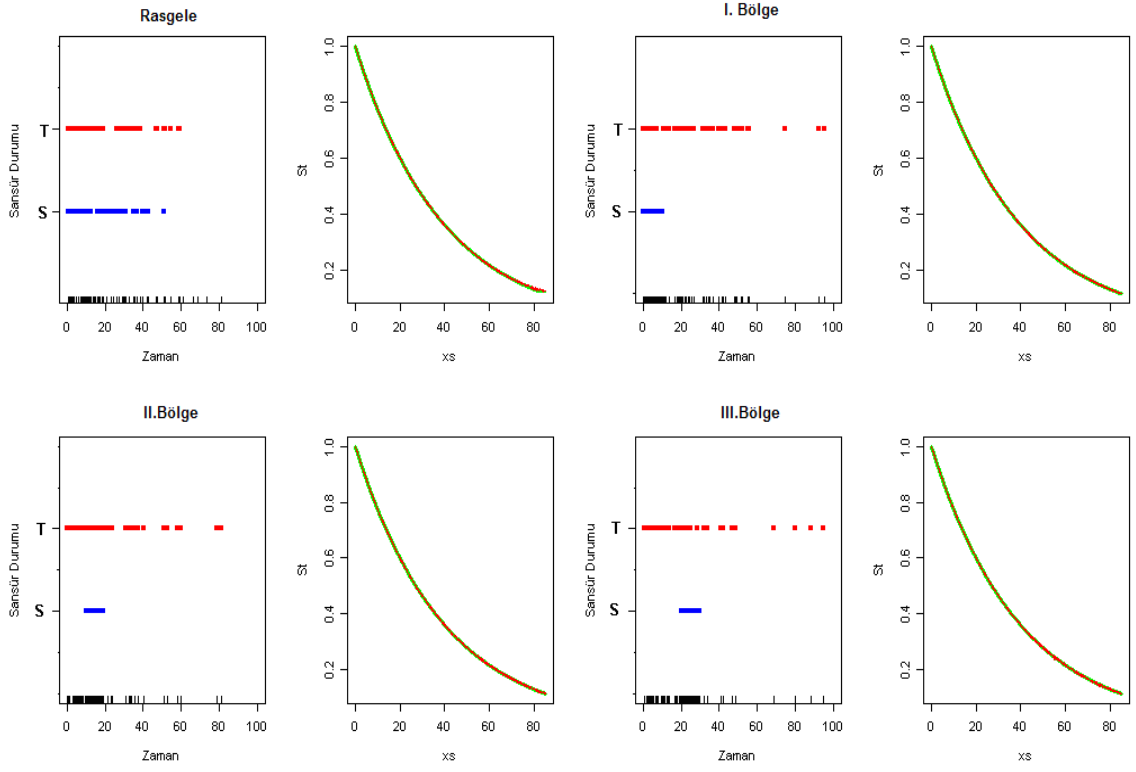
Şekil 4.1 Sabit hazard oranı, n=100, sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri

Şekil 4.5’de ise Sabit hazard oranı için,  $n=100$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir. Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.1’de belirtilmiştir.

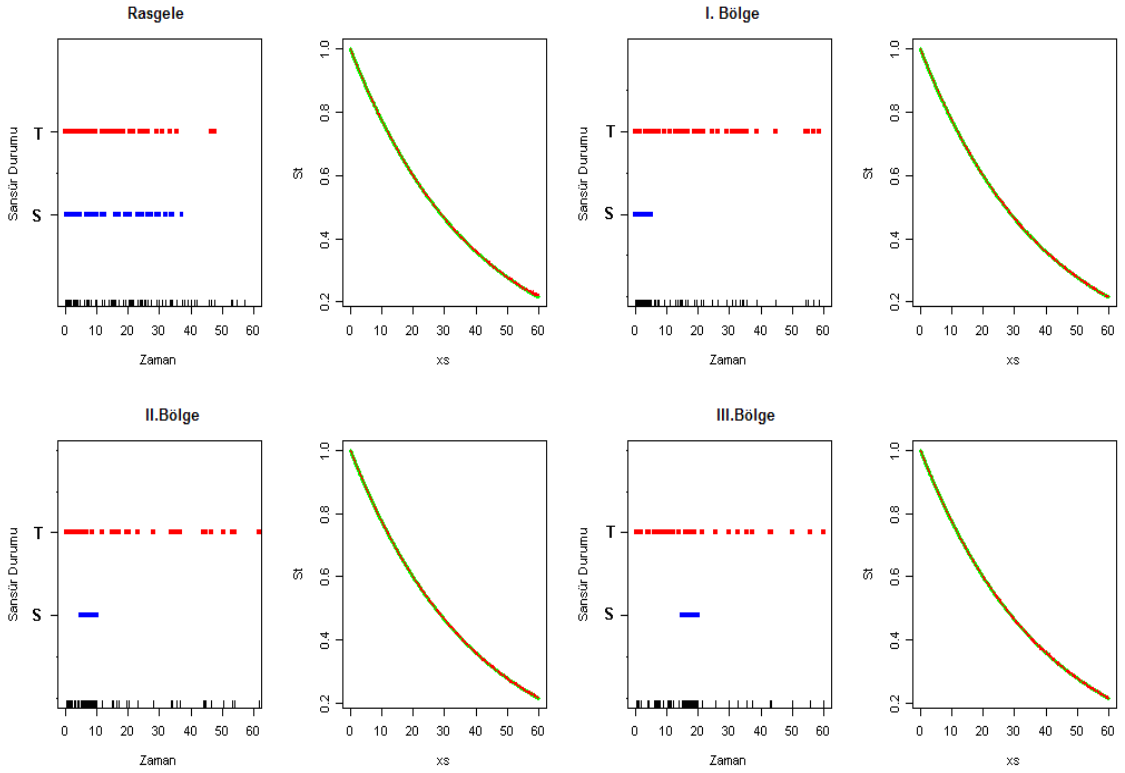
Çizelge 4.1’deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı 0.40 iken, I. ve II. bölgede oluşan sansür oranlarının tahminleri rasgele sansür oranına göre daha az etkilediği fakat verinin sonuna odaklanmış sansürlü gözlemlerin ise hatayı diğer bölgelere göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.2).



Şekil 4.2 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri



Şekil 4.3 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri

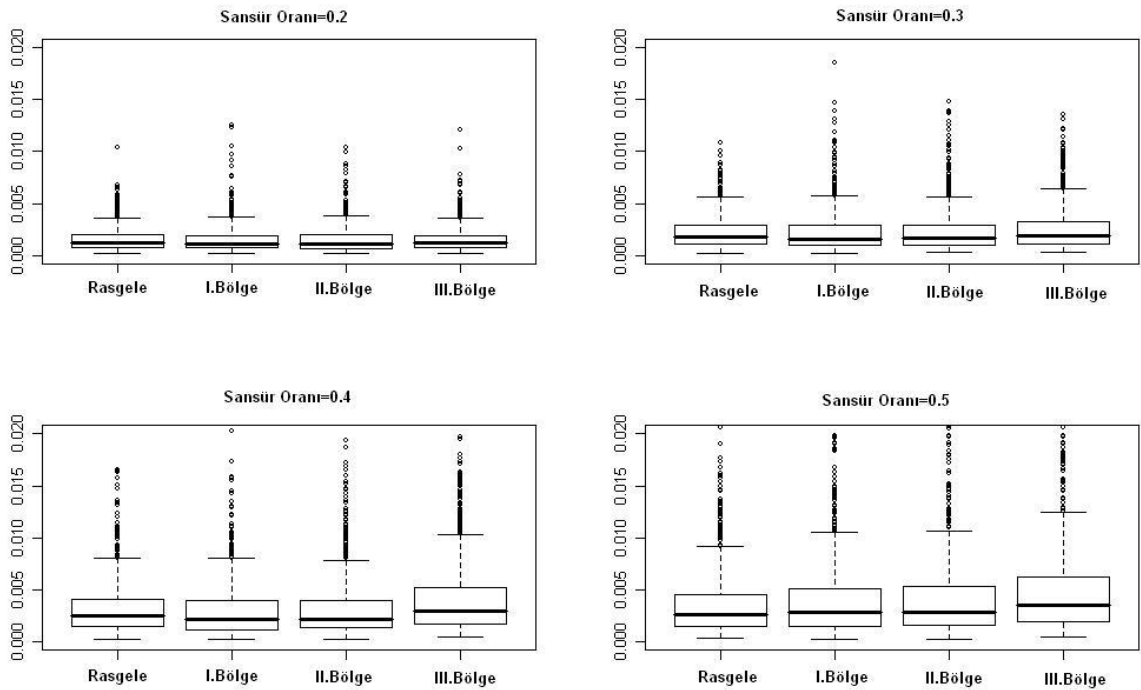


Şekil 4.4 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri

Sansür oranının 0.50 olduğu durumda ise tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.3).

Çizelge 4.1 Sabit hazard oranı, n=100, Kaplan Meier tahminlerinin hatalarına ait tanıttıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00017	0.00076	0.00118	0,00155	0.00187	0.01220	0.097
	I. Bölge	0.00013	0.00069	0.00113	0,00147	0.00187	0.01074	
	II. Bölge	0.00017	0.00068	0.00113	0,00148	0.00184	0.01029	
	III. Bölge	0.00019	0.00076	0.00116	0,00152	0.00194	0.00992	
0.30	Rasgele	0.00020	0.00105	0.00169	0,00221	0.00281	0.01753	0.089
	I. Bölge	0.00030	0.00098	0.00169	0,00228	0.00296	0.01334	
	II. Bölge	0.00026	0.00101	0.00171	0,00234	0.00285	0.01865	
	III. Bölge	0.00020	0.00109	0.00182	0,00248	0.00318	0.01666	
0.40	Rasgele	0.00025	0.00145	0.00246	0,00318	0.00409	0.01652	0.001*
	I. Bölge	0.00021	0.00120	0.00217	0,00303	0.00394	0.02451	
	II. Bölge	0.00028	0.00134	0.00219	0,00328	0.00397	0.02330	
	III. Bölge	0.00044	0.00177	0.00295	0,00423	0.00519	0.03224	
0.50	Rasgele	0.00036	0.00152	0.00262	0,00362	0.00459	0.03117	0.001*
	I. Bölge	0.00025	0.00148	0.00280	0,00406	0.00512	0.03977	
	II. Bölge	0.00027	0.00157	0.00285	0,00421	0.00532	0.02930	
	III. Bölge	0.00046	0.00199	0.00358	0,00497	0.00619	0.04451	



Şekil 4.5 Sabit hazard oranı, n=100, farklı sansürlü gözlem oranları ve yerleşimleri için Kaplan Meier tahmin hatalarına ait box plot grafikleri

En yüksek hata miktarı hem 0. 40 hem de 0. 50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır.

Çizelge 4.2 Sabit hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

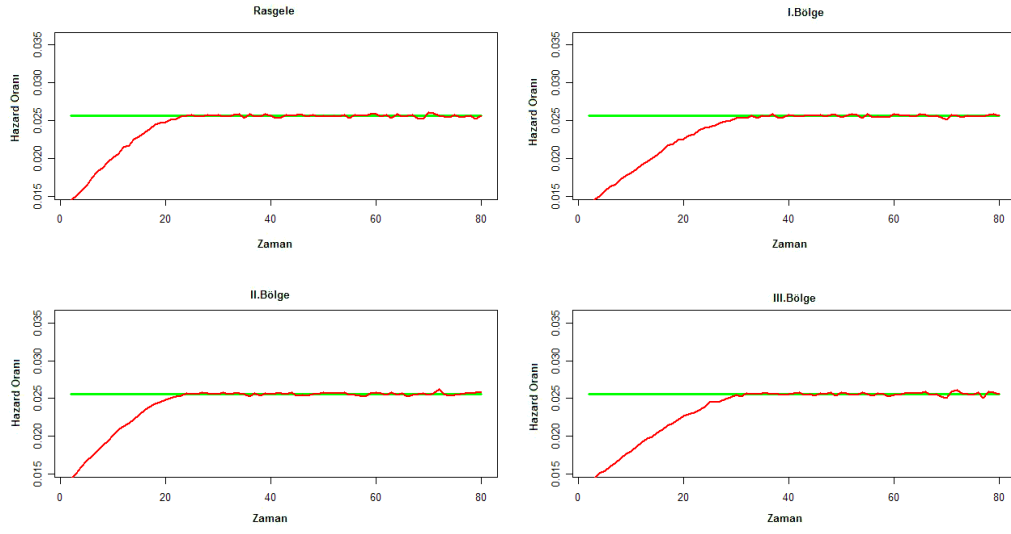
	Bölge I	Bölge II	Bölge III
Rasgele	0.005	0.023	0.000
Bölge I	-----	0.052	0.000
Bölge II	-----	-----	0.000

Çizelge 4.3 Sabit hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

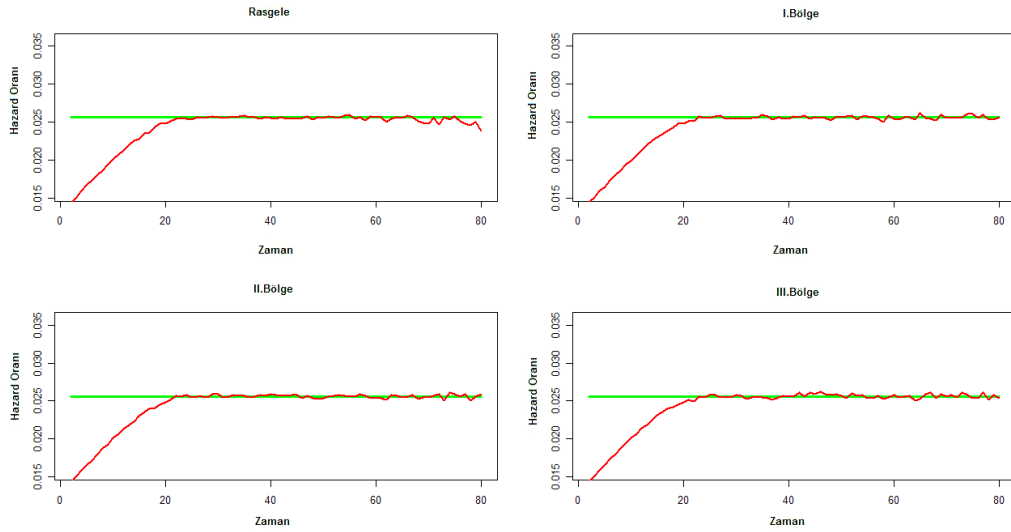
	Bölge I	Bölge II	Bölge III
Rasgele	0.044	0.075	0.000
Bölge I	-----	0.081	0.000
Bölge II	-----	-----	0.000

#### 4.2. Sabit Hazard Oranı, N=100, için Hazard Oranı Tahminleri

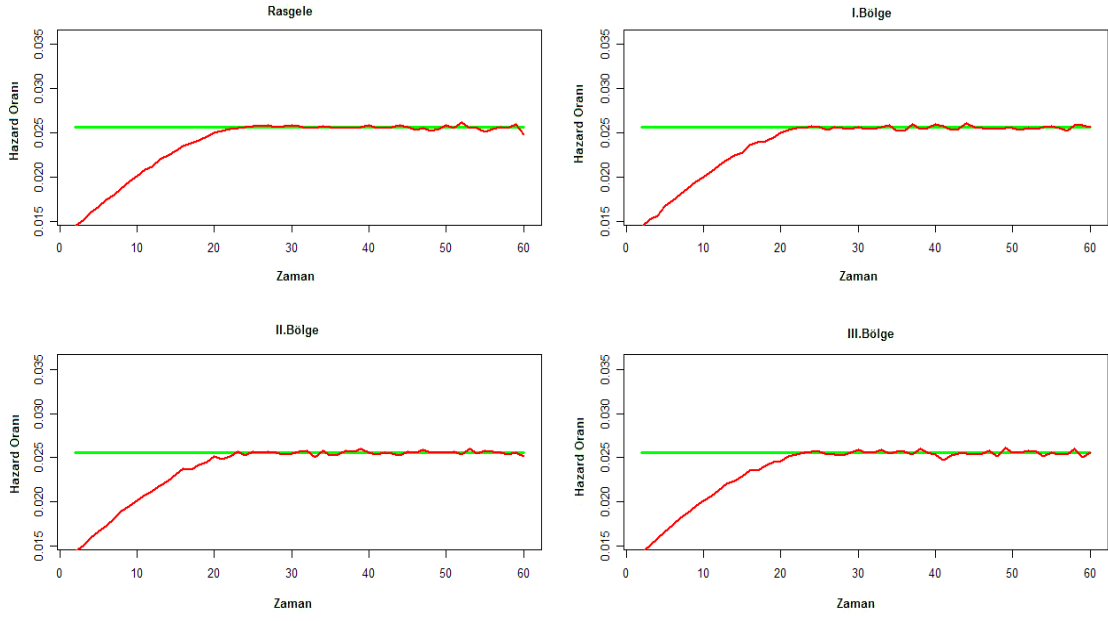
N=100 iken sabit hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.6, Şekil 4.7, Şekil 4.8 ve Şekil 4.9'da, çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinin veri setinin başlangıcında kötü olduğu, fakat zaman ilerledikçe oldukça başarılı olduğu görülmektedir.



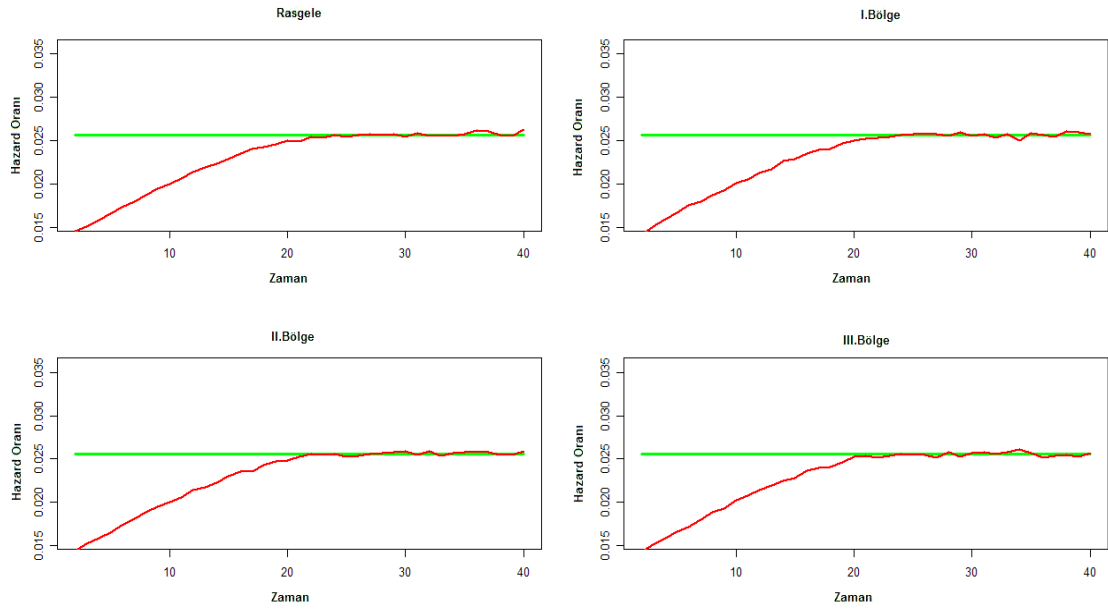
Şekil 4.6 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri



Şekil 4.7 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri

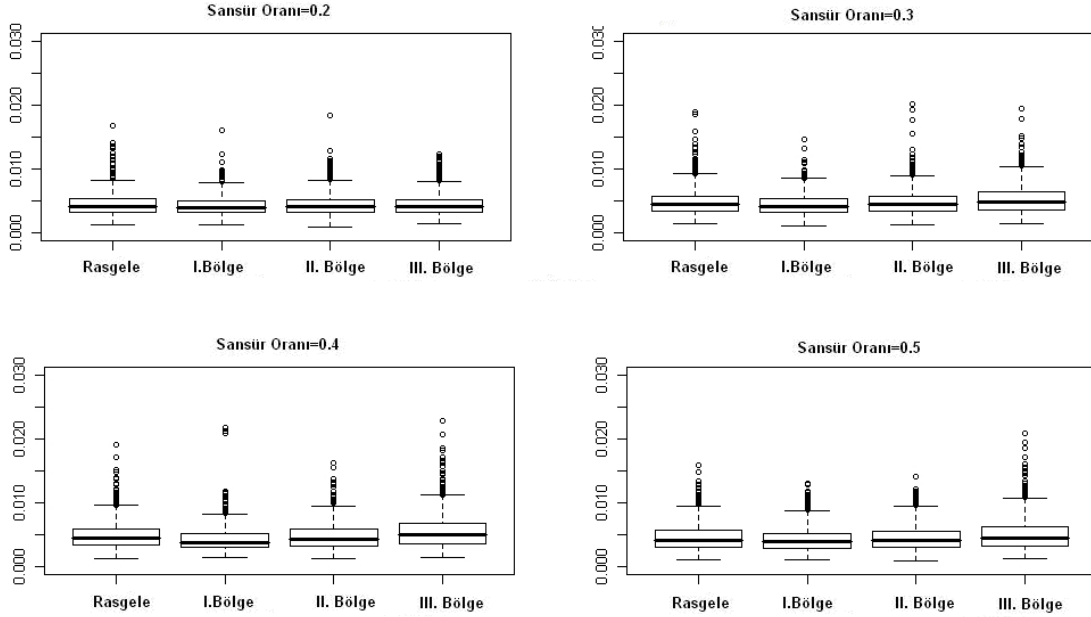


Şekil 4.8 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri



Şekil 4.9 Sabit hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri

Şekil 4.10’da ise Sabit hazard oranı için,  $n=100$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.



Şekil 4.10 Sabit hazard oranı,  $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için Kernel Düzgünleştirme fonksiyonu tahmin hatalarına ait box plot grafikleri

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.4’de belirtilmiştir.

Çizelge 4.4’deki değerlere bakıldığında genel olarak sansürlü gözlem miktarının artırılmasının bölgelerdeki tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken sansür oranı 0.40 iken, II. ve III. Bölge tahmin hatalarının rasgele sansüre göre istatistiksel olarak anlamlı düzeyde artış gösterdiği (Çizelge 4.5.) ve sansür oranı 0.50 iken, III. Bölgede oluşan tahmin hatasının rasgele sansüre göre istatistiksel olarak anlamlı düzeyde arttığı gözlenmiştir (Çizelge 4.6.).

En yüksek hata miktarı 0.40 ve 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.4.).



Çizelge 4.4 Sabit hazard oranı, n=100, Kernel düzgünleştirme tahminlerinin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00130	0.00319	0.00401	0,00437	0.00526	0.01673	0.448
	I. Bölge	0.00129	0.00316	0.00397	0,00426	0.00505	0.01600	
	II. Bölge	0.00083	0.00318	0.00412	0,00442	0.00524	0.01837	
	III. Bölge	0.00137	0.00318	0.00406	0,00435	0.00515	0.01223	
0.30	Rasgele	0.00150	0.00346	0.00443	0,00487	0.00575	0.01893	0.089
	I. Bölge	0.00115	0.00318	0.00407	0,00438	0.00534	0.01468	
	II. Bölge	0.00124	0.00336	0.00439	0,00477	0.00563	0.02011	
	III. Bölge	0.00142	0.00362	0.00474	0,00519	0.00634	0.01942	
0.40	Rasgele	0.00126	0.00347	0.00451	0,00492	0.00592	0.01902	0.001*
	I. Bölge	0.00141	0.00300	0.00380	0,00422	0.00513	0.02172	
	II. Bölge	0.00127	0.00321	0.00430	0,00469	0.00580	0.01621	
	III. Bölge	0.00141	0.00364	0.00494	0,00548	0.00668	0.02273	
0.50	Rasgele	0.00112	0.00306	0.00417	0,00458	0.00567	0.01586	0.001*
	I. Bölge	0.00114	0.00276	0.00389	0,00422	0.00518	0.01300	
	II. Bölge	0.00097	0.00297	0.00402	0,00447	0.00559	0.01413	
	III. Bölge	0.00129	0.00316	0.00444	0,00505	0.00623	0.02086	

Çizelge 4.5 Sabit hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

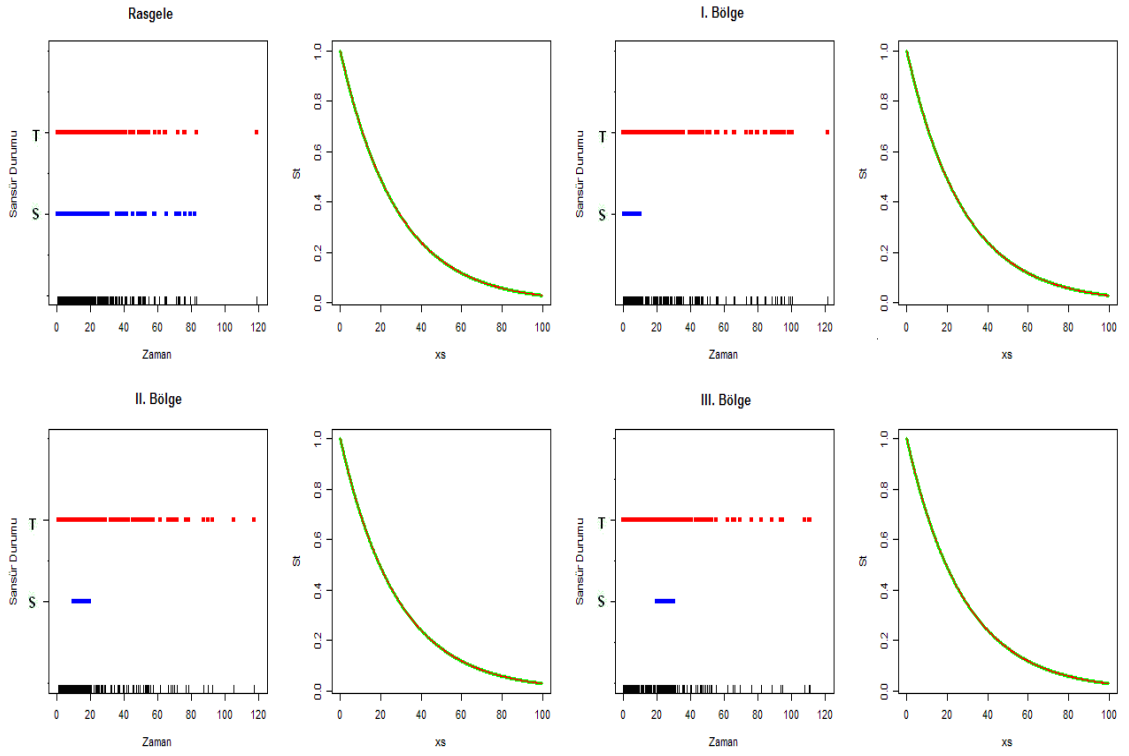
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.005	0.023	0.000
I. Bölge	-----	0.052	0.000
II. Bölge	-----	-----	0.000

Çizelge 4.6 Sabit hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

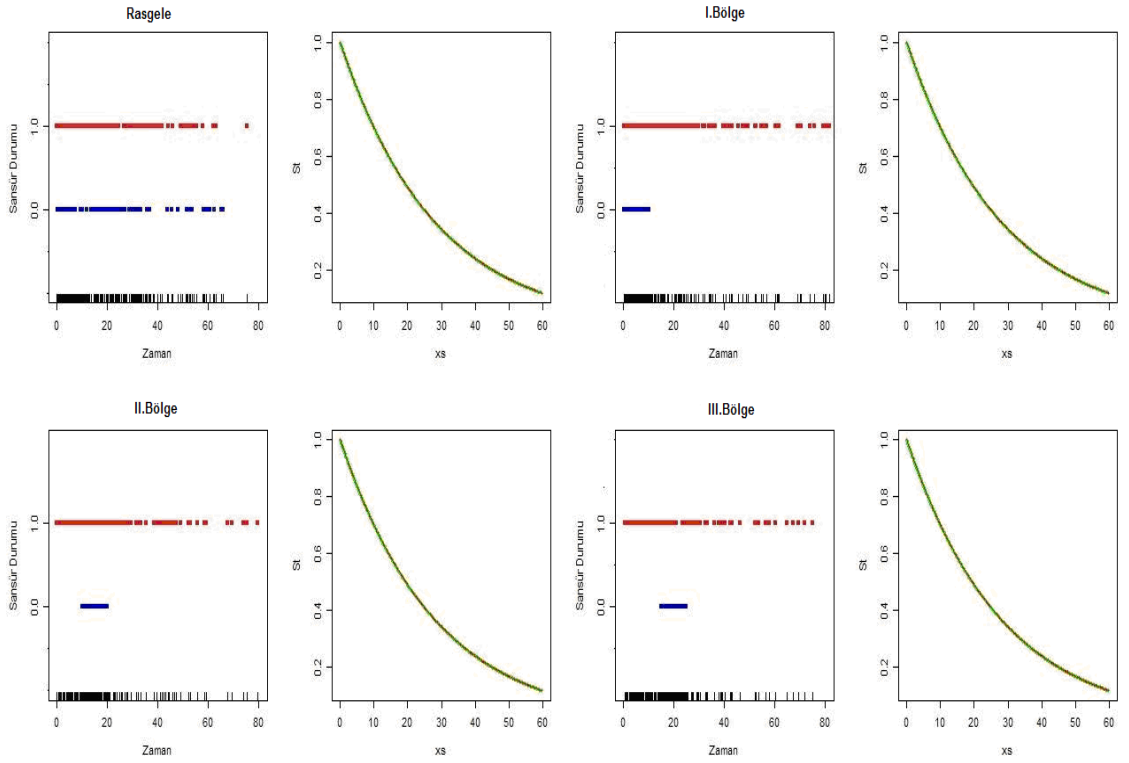
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.044	0.075	0.000
I. Bölge	-----	0.081	0.000
II. Bölge	-----	-----	0.000

### 4.3. Sabit Hazard Oranı, N=250, için Kaplan Meier Tahminleri

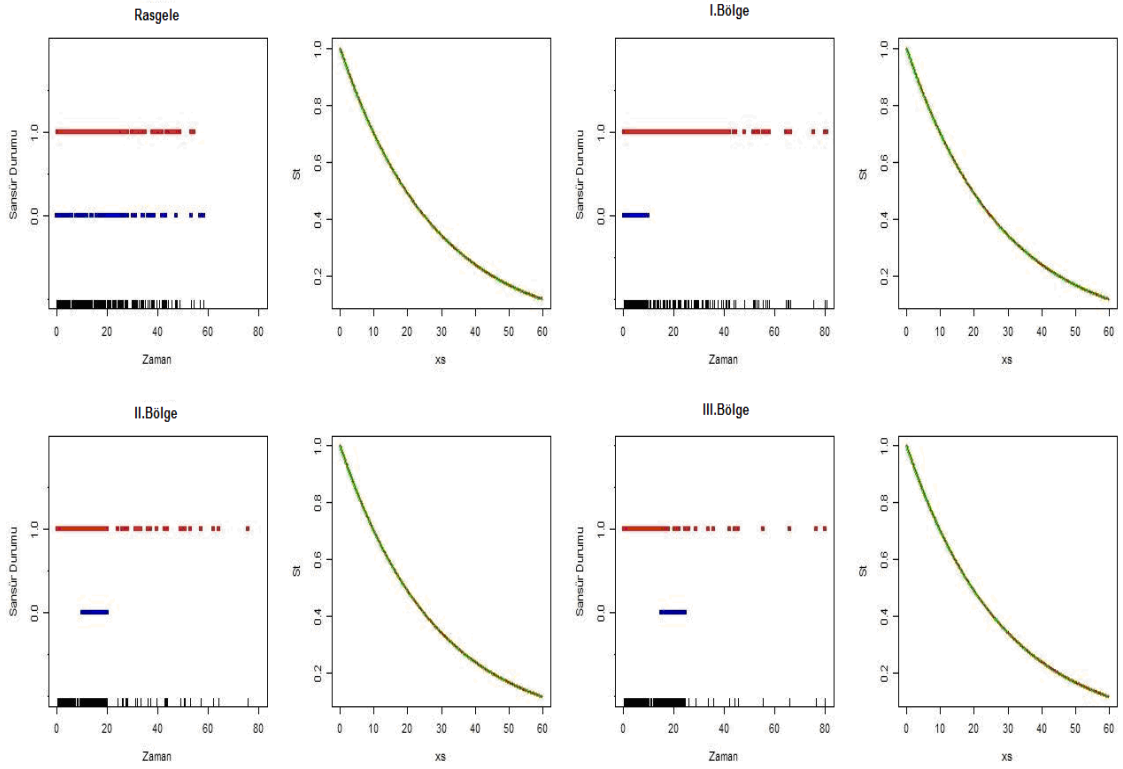
N=250 iken sabit hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.11, Şekil 4.12, Şekil 4.13 ve Şekil 4.14'de çizilmiştir. Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde oldukça başarılı olduğu görülmektedir.



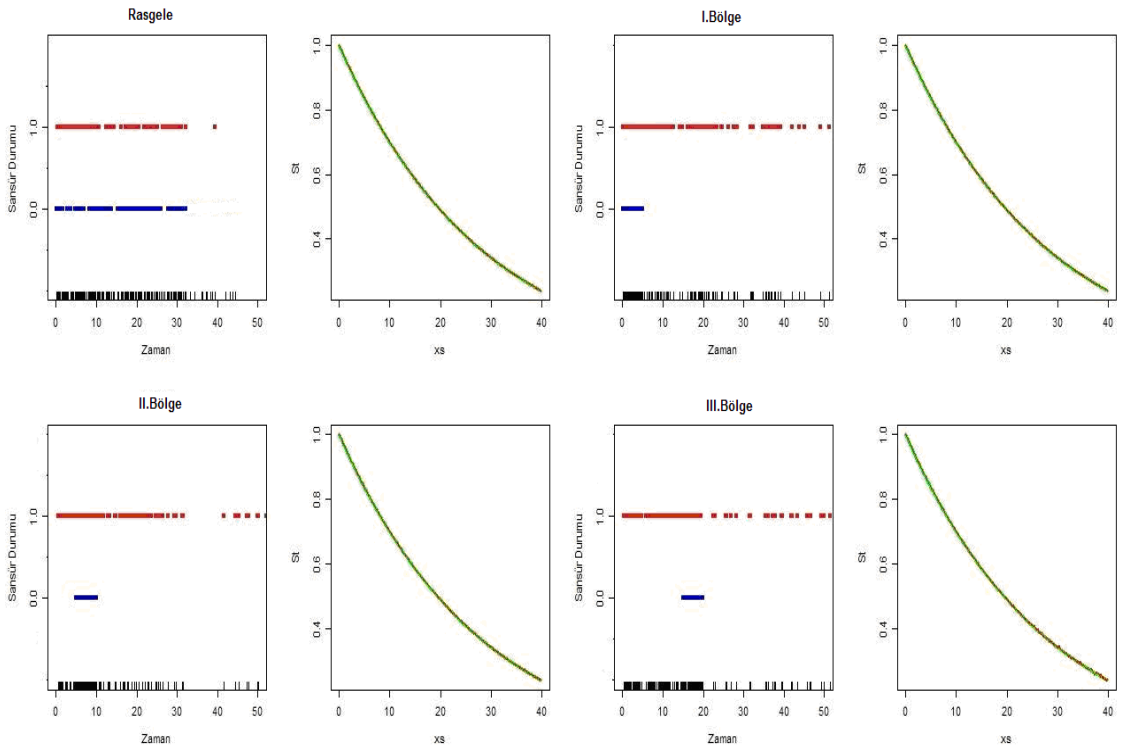
Şekil 4.11 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminler



Şekil 4.12 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri



Şekil 4.13 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri



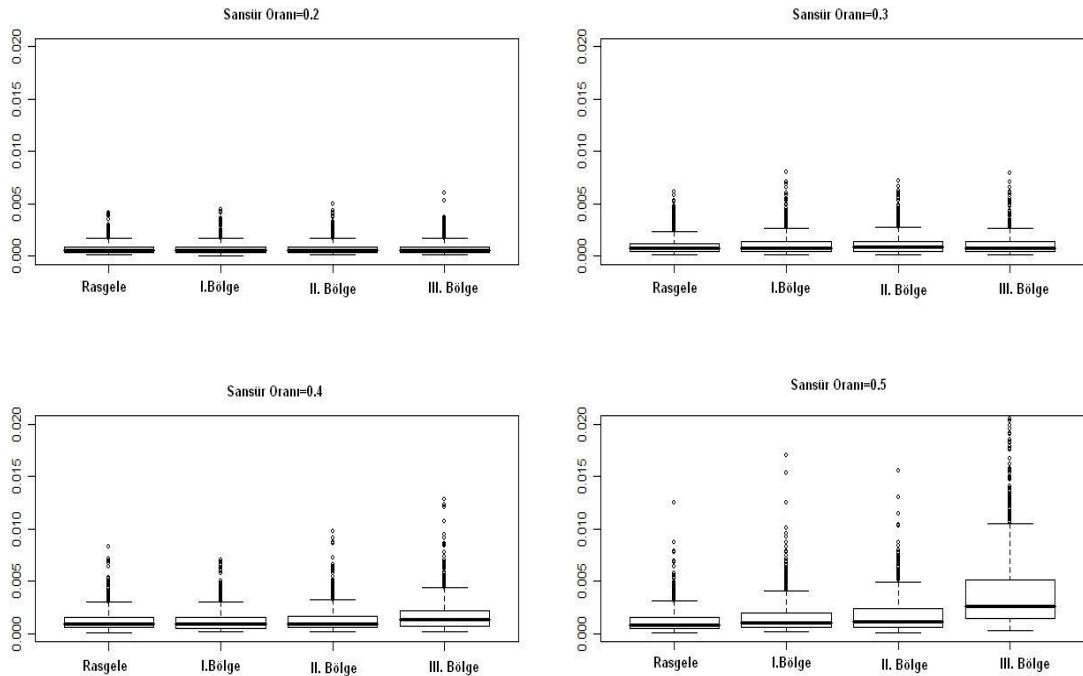
Şekil 4.14 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri

Şekil 4.15’de ise Sabit hazard oranı için,  $n=250$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir. Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıttıcı istatistikler Çizelge 4.7’de belirtilmiştir.

Çizelge 4.7’deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı 0.40 ve 0.50 iken, III. verinin sonuna odaklanmış sansürlü gözlemlerin ise hatayı diğer bölgelere göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.8, Çizelge 4.9).

Sansür oranının 0.50 olduğu durumda ise tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.7).

En yüksek hata miktarı hem 0.40 hem de 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.7).



Şekil 4.15 Sabit hazard oranı,  $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.7 Sabit hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00009	0.00030	0.00052	0,00068	0.00085	0.00410	0.051
	I. Bölge	0.00006	0.00030	0.00051	0,00069	0.00856	0.00450	
	II. Bölge	0.00007	0.00030	0.00053	0,00069	0.00086	0.00496	
	III. Bölge	0.00008	0.00033	0.00056	0,00073	0.00089	0.00598	
0.30	Rasgele	0.00009	0.00041	0.00072	0,00096	0.00121	0.00611	0.053
	I. Bölge	0.00001	0.00042	0.00073	0,00107	0.00135	0.00804	
	II. Bölge	0.00010	0.00046	0.00081	0,00110	0.00136	0.00719	
	III. Bölge	0.00011	0.00046	0.00080	0,00108	0.00135	0.00788	
0.40	Rasgele	0.00012	0.00055	0.00089	0,00122	0.00153	0.00828	0.001*
	I. Bölge	0.00014	0.00051	0.00089	0,00122	0.00153	0.00702	
	II. Bölge	0.00014	0.00057	0.00095	0,00130	0.00163	0.00980	
	III. Bölge	0.00016	0.00074	0.00129	0,00177	0.00221	0.01278	
0.50	Rasgele	0.00009	0.00040	0.00085	0,00124	0.00156	0.01251	0.001*
	I. Bölge	0.00013	0.00050	0.00100	0,00160	0.00198	0.01710	
	II. Bölge	0.00008	0.00060	0.00120	0,00170	0.00023	0.01550	
	III. Bölge	0.00020	0.00140	0.00260	0,00412	0.00510	0.04040	

Çizelge 4.8 Sabit hazard oranı, n=250, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

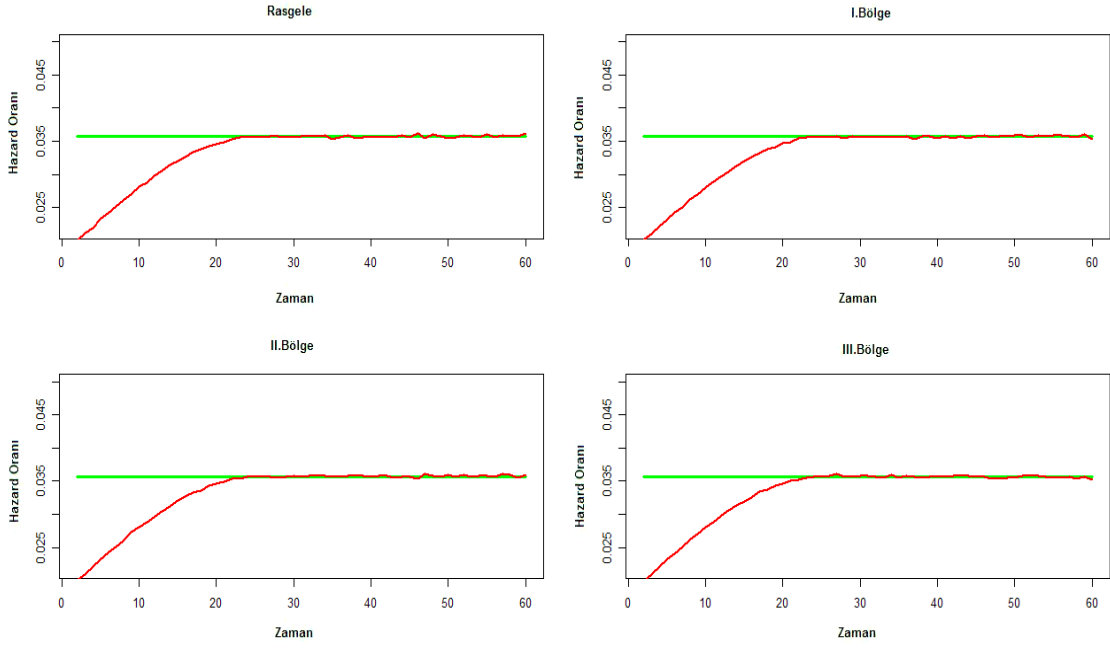
	Bölge I	Bölge II	Bölge III
Rasgele	0.943	0.367	0.000
I. Bölge	-----	0.134	0.000
II. Bölge	-----	-----	0.000

Çizelge 4.9 Sabit hazard oranı, n=250, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

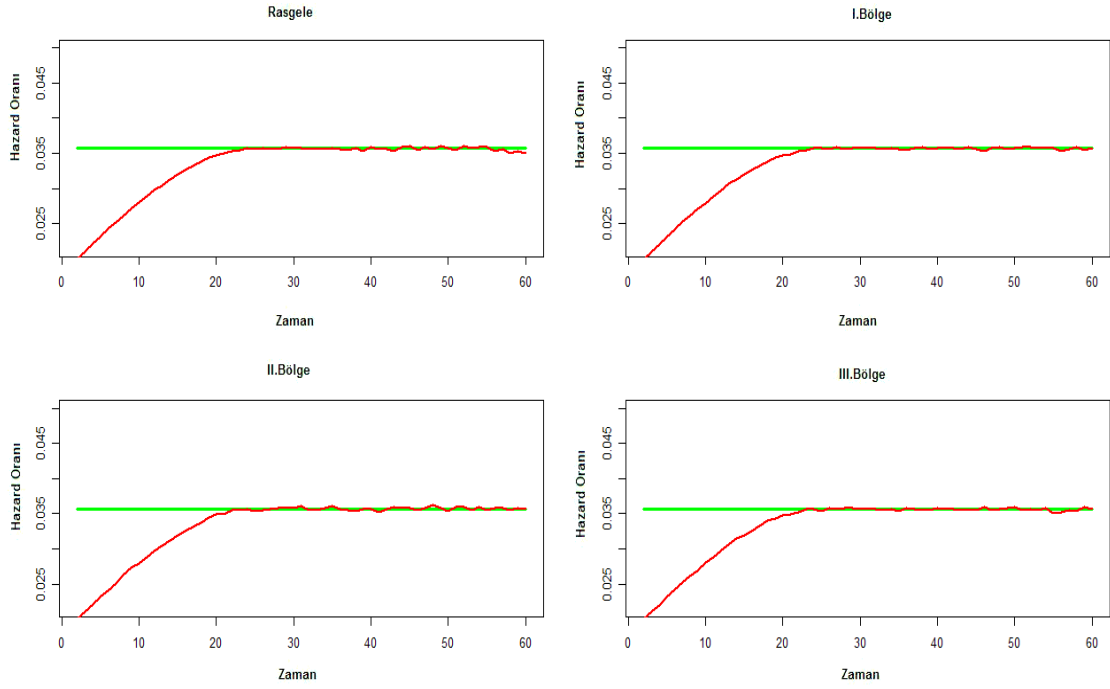
	Bölge I	Bölge II	Bölge III
Rasgele	0.001	0.001	0.000
I. Bölge	-----	0.034	0.000
II. Bölge	-----	-----	0.000

#### 4.4. Sabit Hazard Oranı, N=250, için Hazard Oranı Tahminleri

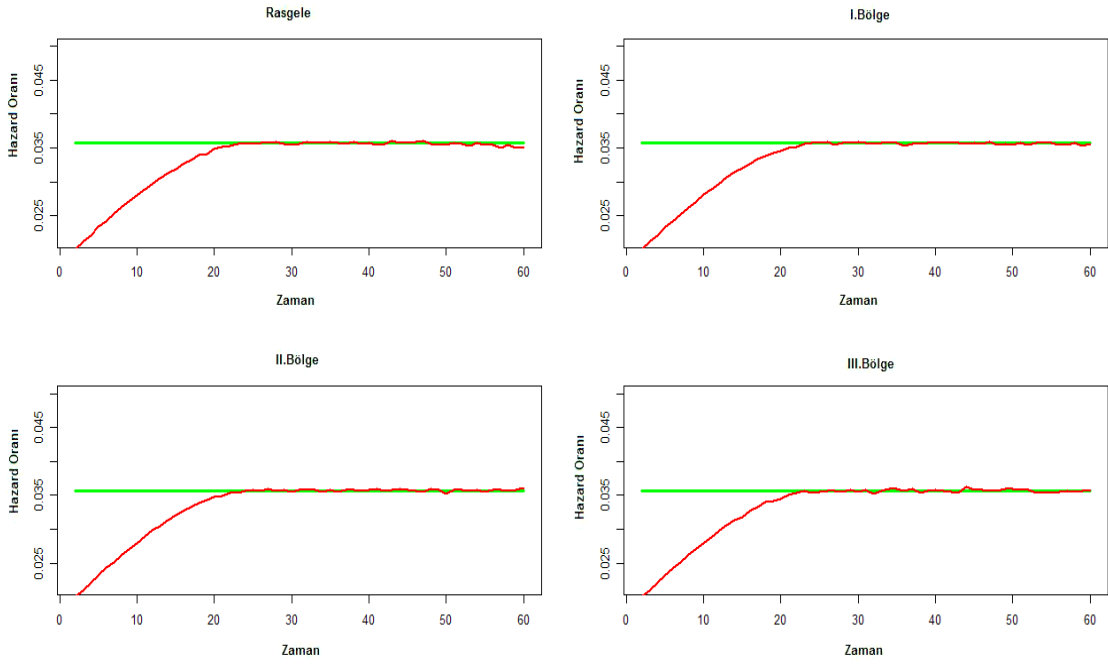
N=250 iken sabit hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.16, Şekil 4.17, Şekil 4.18 ve Şekil 4.19'da çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinin veri setinin başlangıcında kötü olduğu fakat zaman ilerledikçe oldukça başarılı olduğu görülmektedir.



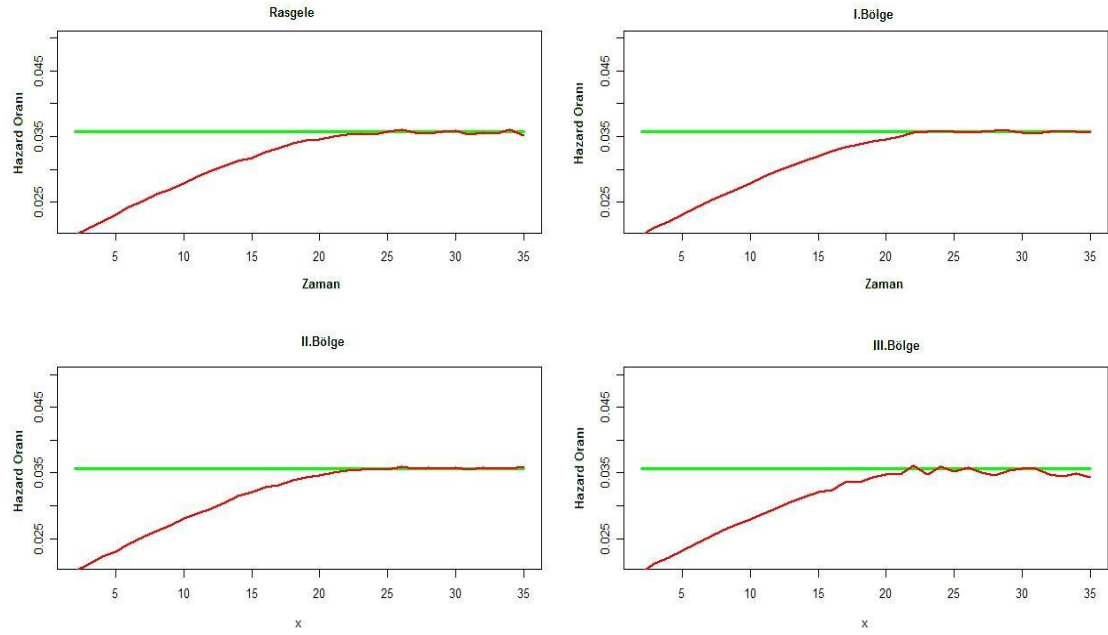
Şekil 4.16 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri



Şekil 4.17 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri



Şekil 4.18 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri



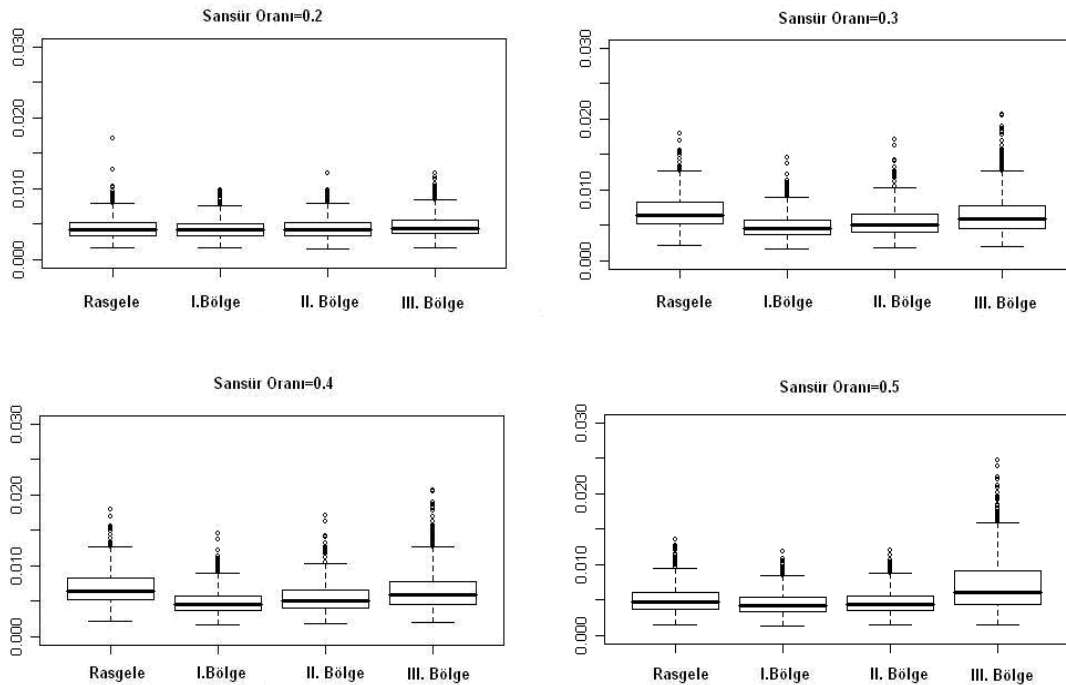
Şekil 4.19 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri

Şekil 4.20’de ise Sabit hazard oranı için,  $n=250$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve oranlarının değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.10’da belirtilmiştir.

Çizelge 4.10’daki değerlere bakıldığında, genel olarak sansürlü gözlem miktarının artırılmasının bölgelerdeki tahmin hatasını yükseldiği gözlenmektedir. Sansürlü gözlem oranı 0.20 iken III. Bölgede oluşan hata miktarı rasgele sansür oluşumuna göre istatistiksel olarak anlamlı düzeyde artış göstermiştir (Çizelge 4.11) ve 0.30, 0.40 ve 0.50 sansür oranlarında ise sansürlerin I. bölgede oluşması rasgele sansür bölgesinde oluşmasına göre tahmin hatasını azaltırken, diğer bölgelerde hata miktarında artış görülmüştür (Çizelge 4.12, Çizelge 4.13, Çizelge 4.14).

En yüksek hata miktarı 0.20, 0.30 ve 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.7).



Şekil 4.20 Sabit hazard oranı,  $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri



Çizelge 4.10 Sabit hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00157	0.00343	0.00417	0,00447	0.00521	0.01705	0.001*
	I. Bölge	0.00163	0.00333	0.00411	0,00429	0.00503	0.00978	
	II. Bölge	0.00150	0.00336	0.00417	0,00442	0.00524	0.01225	
	III. Bölge	0.00172	0.00360	0.00440	0,00470	0.00551	0.01219	
0.30	Rasgele	0.00173	0.00375	0.00462	0,00492	0.00580	0.01642	0.001*
	I. Bölge	0.00120	0.00334	0.00418	0,00447	0.00533	0.01323	
	II. Bölge	0.00189	0.00360	0.00447	0,00477	0.00565	0.01493	
	III. Bölge	0.00173	0.00383	0.00474	0,00512	0.00609	0.01392	
0.40	Rasgele	0.00218	0.00524	0.00632	0,00680	0.00822	0.01789	0.001*
	I. Bölge	0.00163	0.00363	0.00455	0,00485	0.00576	0.01446	
	II. Bölge	0.00184	0.00402	0.00501	0,00545	0.00656	0.01705	
	III. Bölge	0.00202	0.00456	0.00591	0,00658	0.00783	0.02063	
0.50	Rasgele	0.001515	0.00362	0.00464	0,00504	0.00599	0.01351	0.001*
	I. Bölge	0.001367	0.00327	0.00421	0,00444	0.00531	0.01182	
	II. Bölge	0.001432	0.00351	0.00441	0,00469	0.00560	0.01199	
	III. Bölge	0.001554	0.00436	0.00607	0,00717	0.00903	0.02464	

Çizelge 4.11 Sabit hazard oranı, n=250, sansürlü gözlem oranı 0.20 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.100	0.912	0.007
I. Bölge	-----	0.353	0.001
II. Bölge	-----	-----	0.001

Çizelge 4.12 Sabit hazard oranı, n=250, sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.181	0.243
I. Bölge	-----	0.001	0.001
II. Bölge	-----	-----	0.001

Çizelge 4.13 Sabit hazard oranı, n=250, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

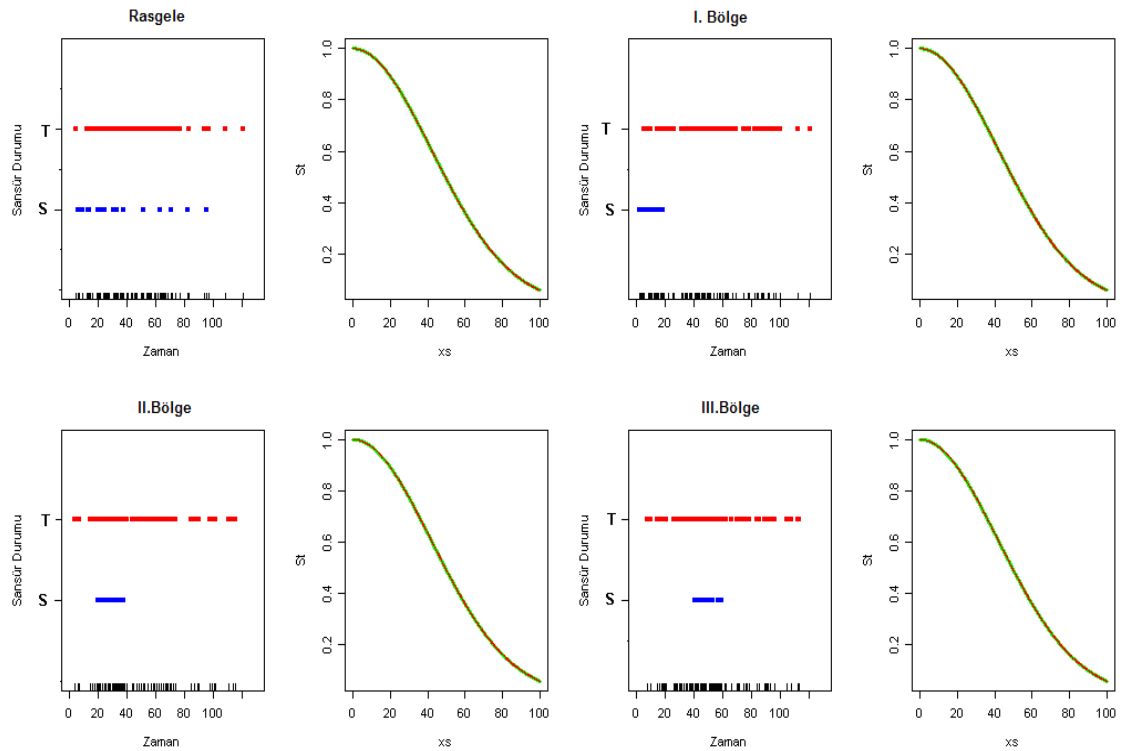
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.001	0.001
I. Bölge	-----	0.001	0.001
II. Bölge	-----	-----	0.001

Çizelge 4.14 Sabit hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

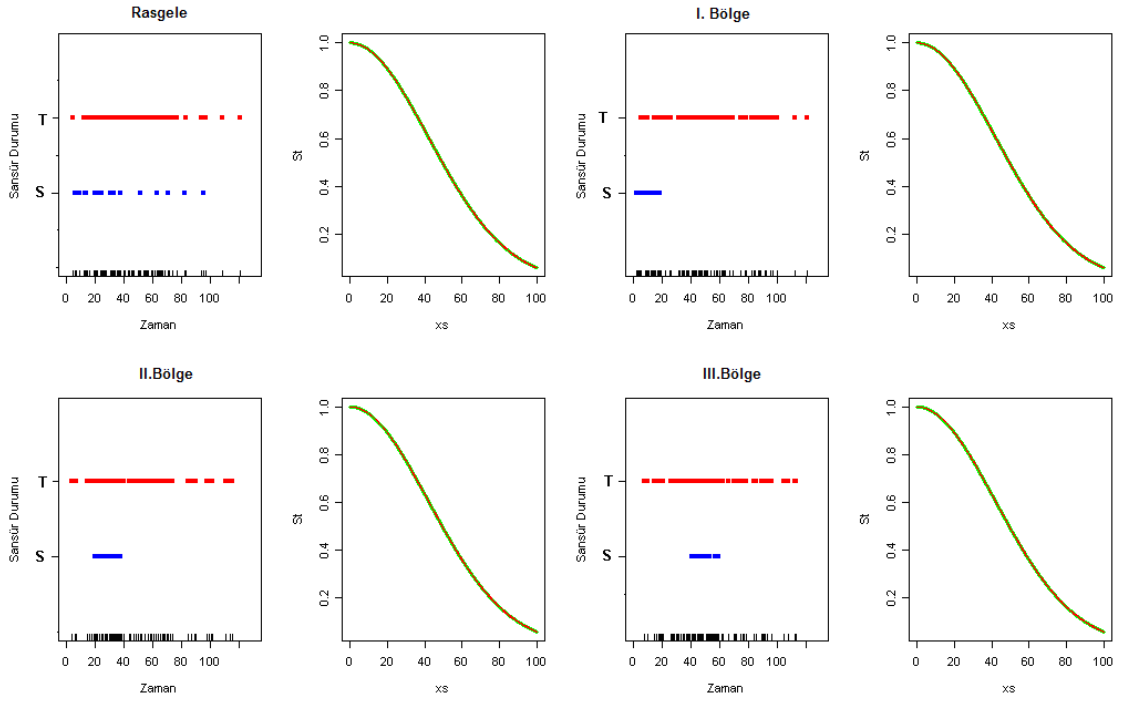
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.004	0.001
I. Bölge	-----	0.002	0.001
II. Bölge	-----	-----	0.001

#### 4.5. Linear Artan Hazard Oranı, $N=100$ için Kaplan Meier Tahminleri

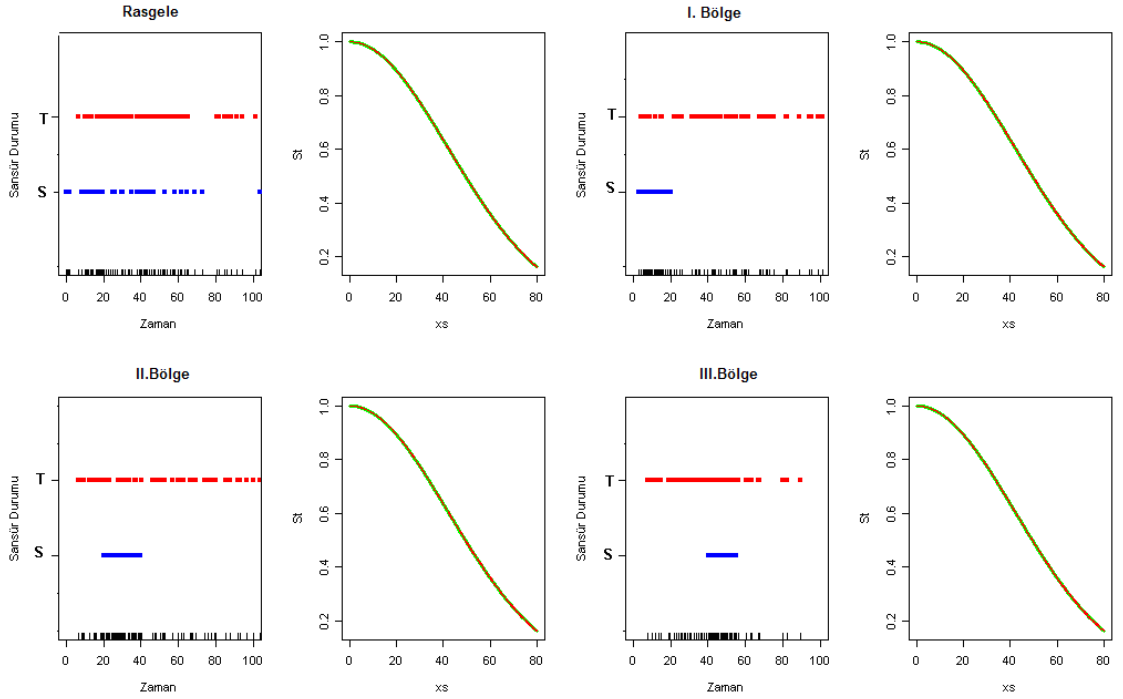
$N=100$  iken Linear artan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.21, Şekil 4.22, Şekil 4.23 ve Şekil 4.24'de çizilmiştir. Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde oldukça başarılı olduğu görülmektedir.



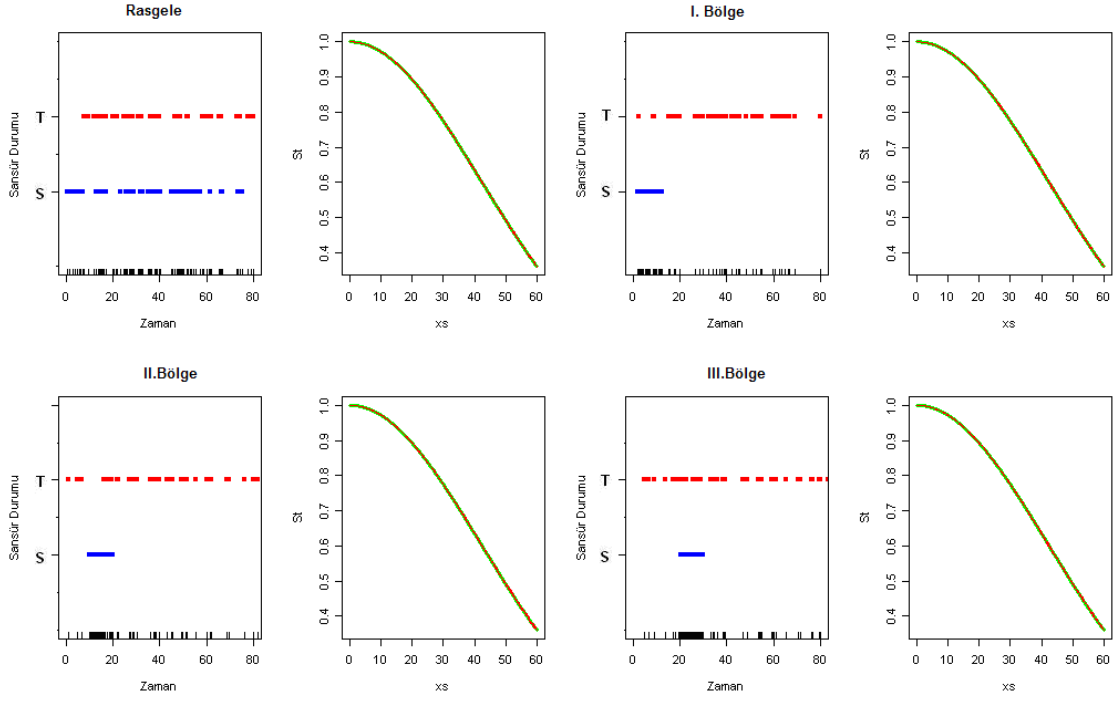
Şekil 4.21 Doğrusal artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri



Şekil 4.22 Doğrusal artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri



Şekil 4.23 Doğrusal artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri



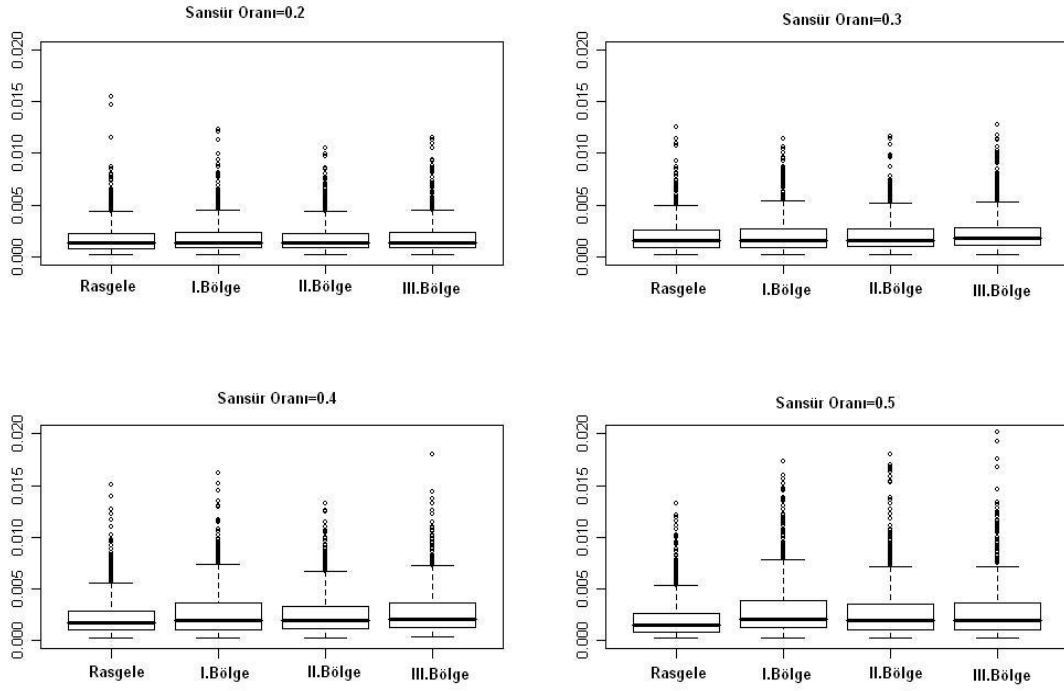
Şekil 4.24 Doğrusal artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri

Şekil 4.25’de Lineer artan hazard oranı için,  $n=100$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.15’de belirtilmiştir.

Çizelgede 4.15’deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı 0.40 iken, I., II. ve III. bölgede oluşan sansürün tahminleri rasgele sansür oranına göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.16). Sansür oranının 0.50 olduğu durumda ise tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.17).

En yüksek hata miktarı hem 0.40 hem de 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.15).



Şekil 4.25 Doğrusal artan hazard oranı,  $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.15 Doğrusal artan hazard oranı,  $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00019	0.00079	0.00130	0,00176	0.00223	0.01550	0.240
	I. Bölge	0.00025	0.00085	0.00137	0,00185	0.00232	0.01228	
	II. Bölge	0.00019	0.00082	0.00129	0,00178	0.00224	0.01044	
	III. Bölge	0.00018	0.00086	0.00137	0,00183	0.00221	0.01155	
0.30	Rasgele	0.00025	0.00092	0.00153	0,00202	0.00256	0.01252	0.051
	I. Bölge	0.00023	0.00092	0.00159	0,00216	0.00273	0.01134	
	II. Bölge	0.00023	0.00094	0.00156	0,00206	0.00265	0.01165	
	III.Bölge	0.00025	0.00105	0.00174	0,00227	0.00274	0.01277	
0.40	Rasgele	0.00026	0.00104	0.00172	0,00232	0.00288	0.02091	0.001*
	I. Bölge	0.00024	0.00107	0.00198	0,00273	0.00363	0.02208	
	II. Bölge	0.00030	0.00111	0.00190	0,00258	0.00334	0.01322	
	III.Bölge	0.00037	0.00129	0.00210	0,00282	0.00366	0.01803	
0.50	Rasgele	0.00020	0.00082	0.00144	0,00240	0.00263	0.02463	0.001*
	I. Bölge	0.00029	0.00124	0.00203	0,00307	0.00389	0.02502	
	II. Bölge	0.00022	0.00109	0.00194	0,00284	0.00356	0.02106	
	III.Bölge	0.00025	0.00110	0.00195	0,00288	0.00365	0.02327	

Çizelge 4.16 Doğrusal artan hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları.

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.006	0.012	0.001
I. Bölge	-----	0.0956	0.081
II. Bölge	-----	-----	0.013

Çizelge 4.17 Doğrusal artan hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları.

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.0826	0.001	0.001
I. Bölge	-----	0.001	0.001
II. Bölge	-----	-----	0.002

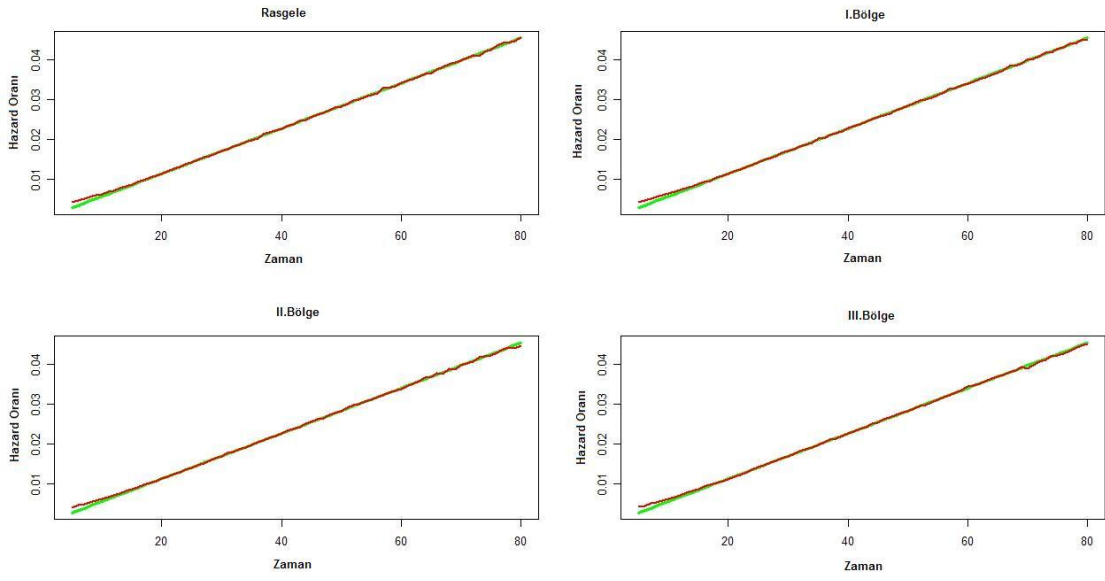
#### 4.6. Linear Artan Hazard Oranı, N=100 için Hazard Oranı Tahminleri

N=100 iken Linear artan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.26, Şekil 4.27, Şekil 4.28 ve Şekil 4.29 'da, çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde oldukça başarılı olduğu görülmektedir.

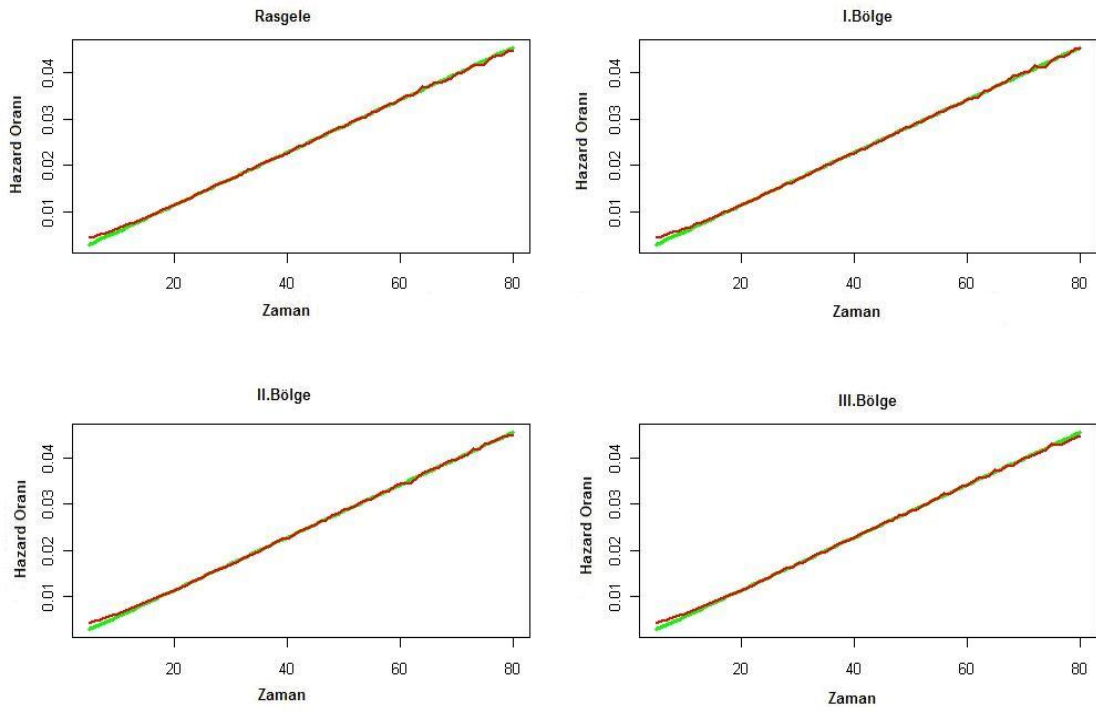
Şekil 4.30'da Linear artan hazard oranı için, n=100 olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.18'de belirtilmiştir.

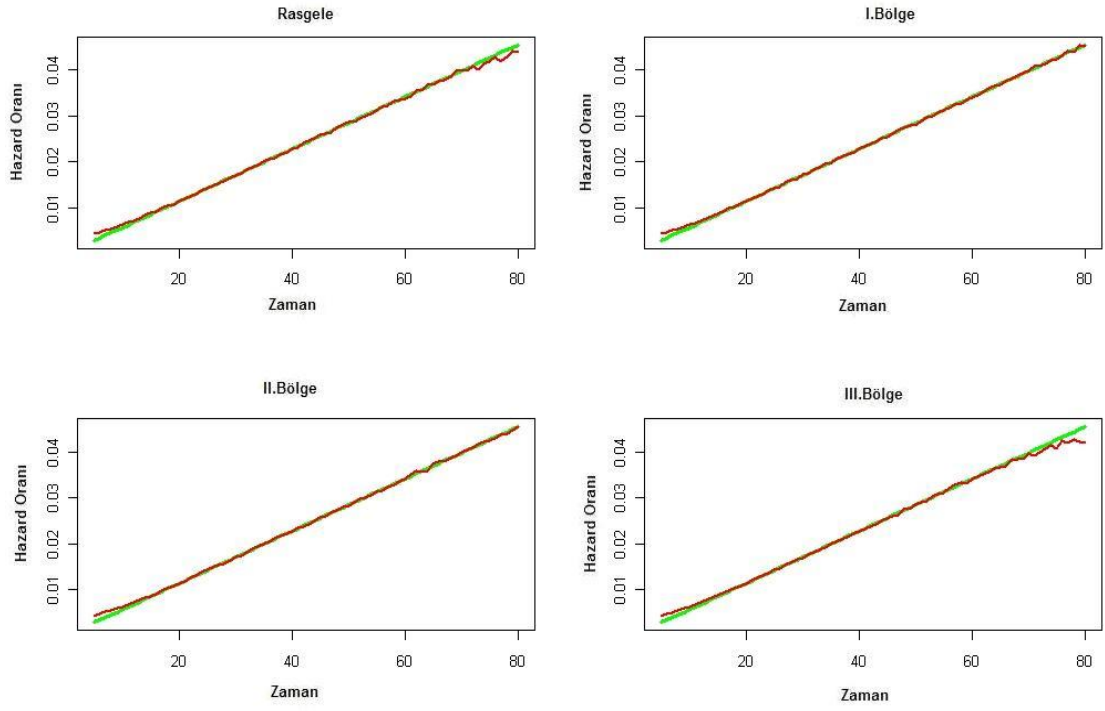
Sansürlü gözlem oranı 0.20 olduğu durumda 3 farklı sansür bölgesindeki tahmin hatası ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken sansür oranı 0.30 ve 0.40 iken III. Bölgede oluşan hatanın rasgele sansür hatasına göre istatistiksel olarak anlamlı düzeyde artış gösterdiği (Çizelge 4.19, Çizelge 4.20) ve sansür oranı 0.50 iken, I, II, ve III Bölgelerde oluşan tahmin hatasının rasgele sansüre göre istatistiksel olarak anlamlı düzeyde azaldığı gözlenmiştir (Çizelge 4.21).



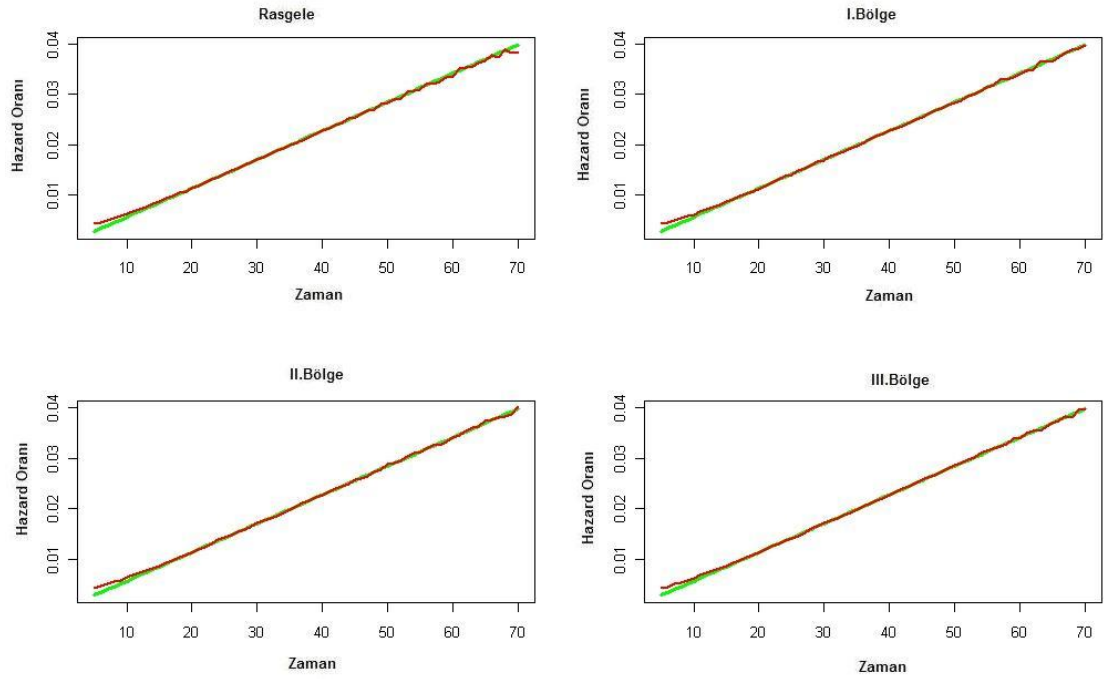
Şekil 4.26 Lineer artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri



Şekil 4.27 Lineer artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri

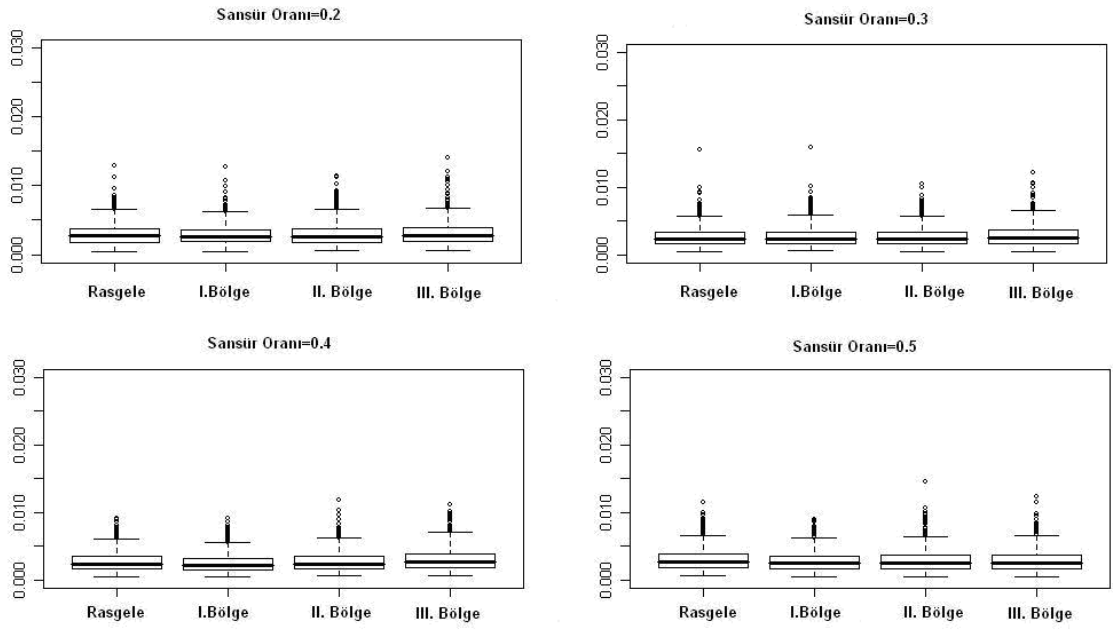


Şekil 4.28 Lineer artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri



Şekil 4.29 Lineer artan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri





Şekil 4.30 Lineer artan hazard oranı,  $n=100$  farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.18 Doğrusal artan hazard oranı,  $n=100$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00049	0.00182	0.00273	0,00295	0.00374	0.01288	0.173
	I. Bölge	0.00050	0.00186	0.00263	0,00293	0.00359	0.01273	
	II. Bölge	0.00056	0.00182	0.00267	0,00298	0.00378	0.01146	
	III. Bölge	0.00060	0.00192	0.00276	0,00309	0.00384	0.01403	
0.30	Rasgele	0.00053	0.00164	0.00236	0,00263	0.00329	0.01561	0.009*
	I. Bölge	0.00056	0.00159	0.00229	0,00264	0.00334	0.01591	
	II. Bölge	0.00046	0.00163	0.00235	0,00261	0.00328	0.01044	
	III. Bölge	0.00052	0.00166	0.00246	0,00283	0.00364	0.01209	
0.40	Rasgele	0.00049	0.00162	0.00237	0,00270	0.00348	0.00919	0.001*
	I. Bölge	0.00045	0.00148	0.00219	0,00248	0.00314	0.00905	
	II. Bölge	0.00059	0.00164	0.00236	0,00269	0.00346	0.01175	
	III. Bölge	0.00059	0.00183	0.00267	0,00308	0.00394	0.01112	
0.50	Rasgele	0.00055	0.00186	0.00269	0,00302	0.00379	0.01153	0.001*
	I. Bölge	0.00042	0.00160	0.00245	0,00271	0.00346	0.00902	
	II. Bölge	0.00044	0.00169	0.00247	0,00283	0.00360	0.01462	
	III. Bölge	0.00054	0.00170	0.00255	0,00288	0.00371	0.01229	

Çizelge 4.19 Doğrusal artan hazard oranı, n=100, sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.934	0.969	0.080
I. Bölge	-----	0.999	0.019
II. Bölge	-----	-----	0.027

Çizelge 4.20 Doğrusal artan hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

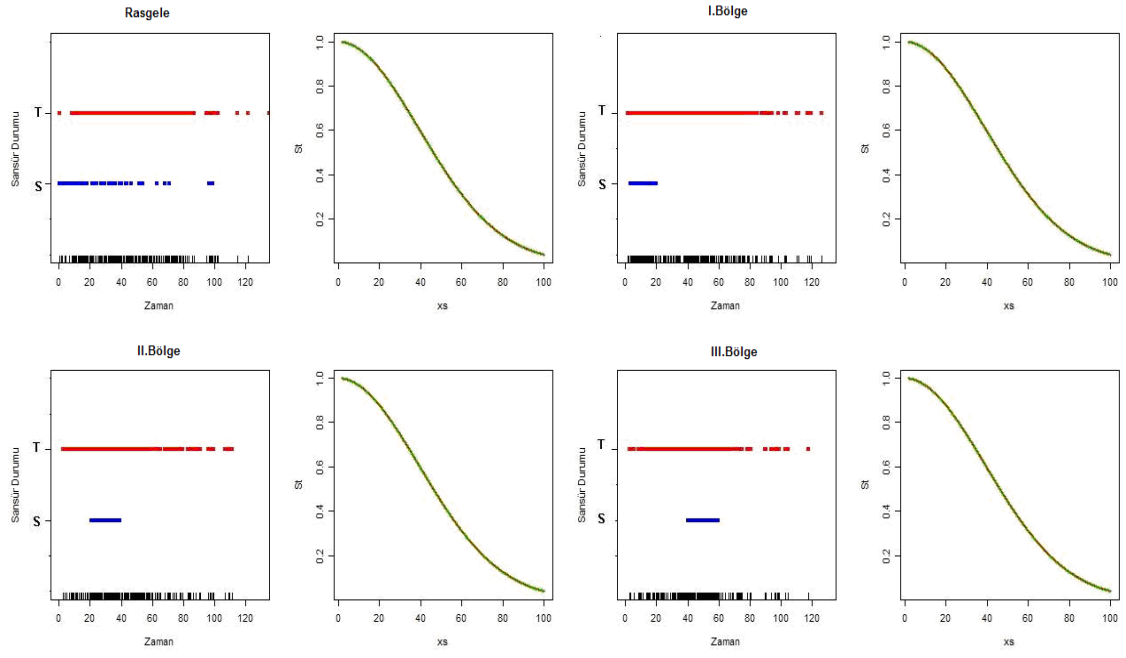
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.909	0.001
I. Bölge	-----	0.004	0.000
II. Bölge	-----	-----	0.001

Çizelge 4.21 Doğrusal artan hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

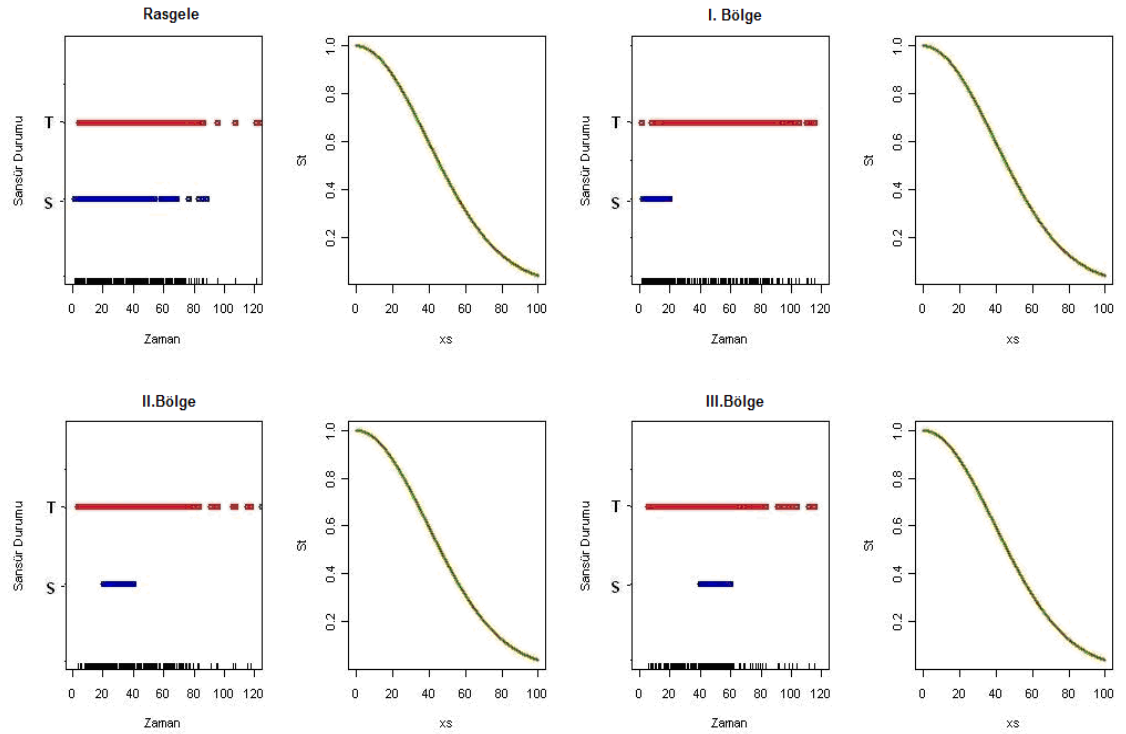
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.001	0.001
I. Bölge	-----	0.453	0.149
II. Bölge	-----	-----	0.902

#### 4.7. Lineer Artan Hazard Oranı, N=250 için Kaplan Meier Tahminleri

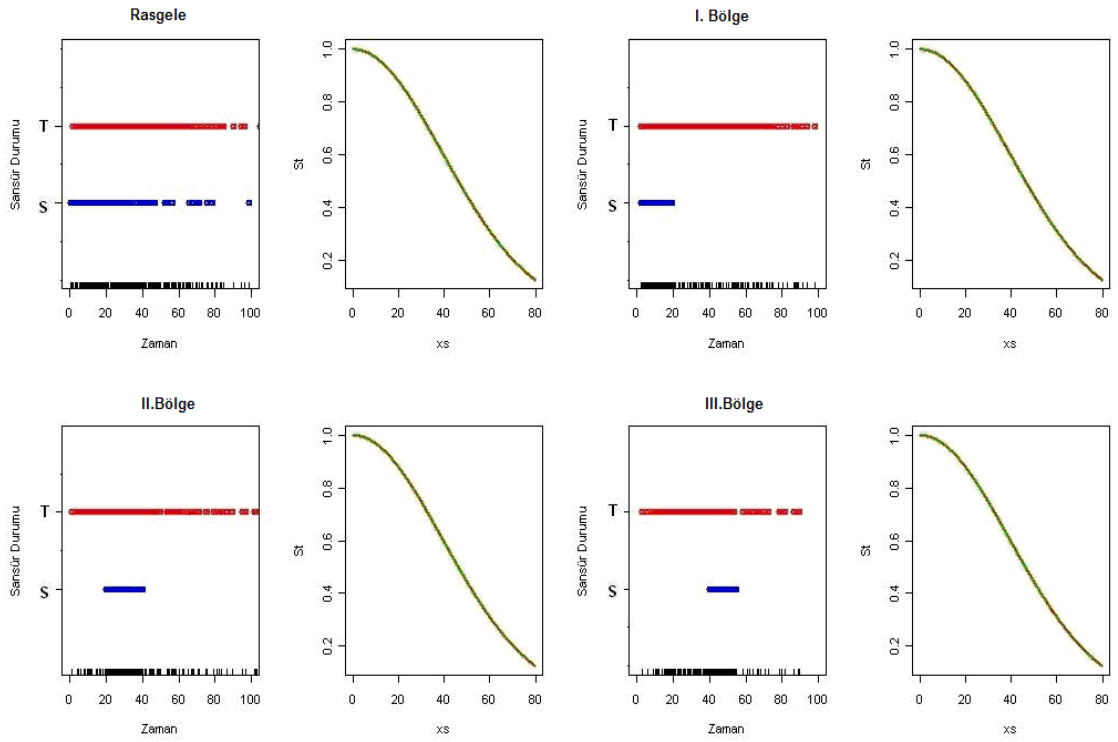
N=250 iken Lineer artan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.31, Şekil 4.32, Şekil 4.33 ve Şekil 4.34'de çizilmiştir. Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde oldukça başarılı olduğu görülmektedir.



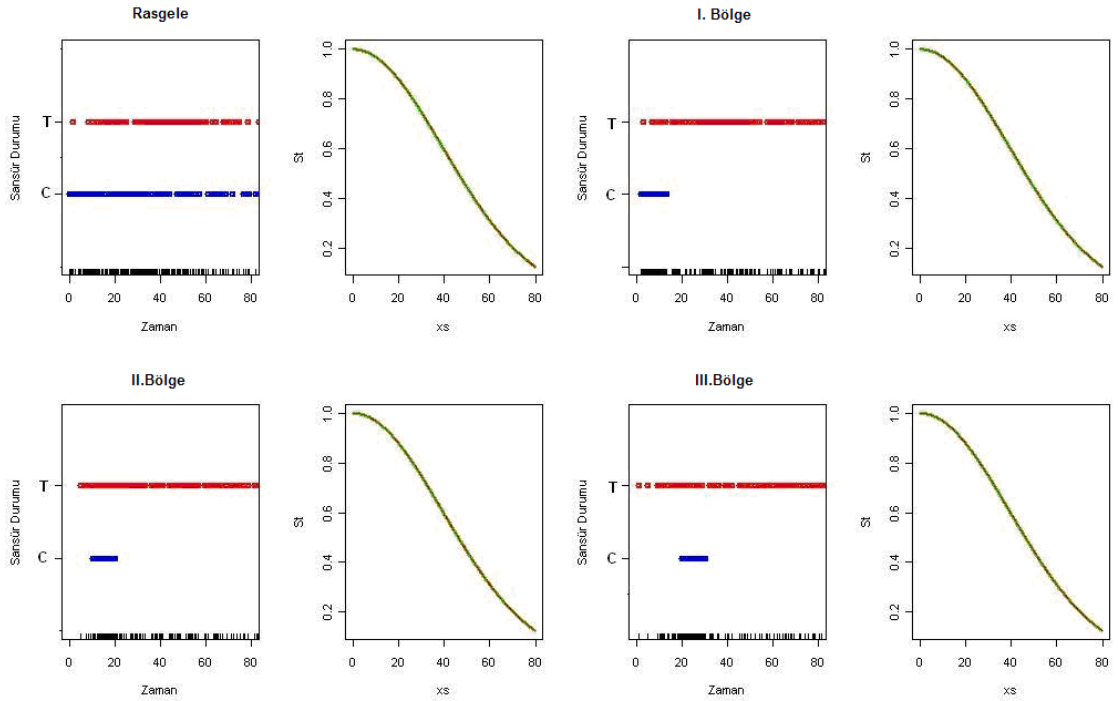
Şekil 4.31 Doğrusal artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri



Şekil 4.32 Doğrusal artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri



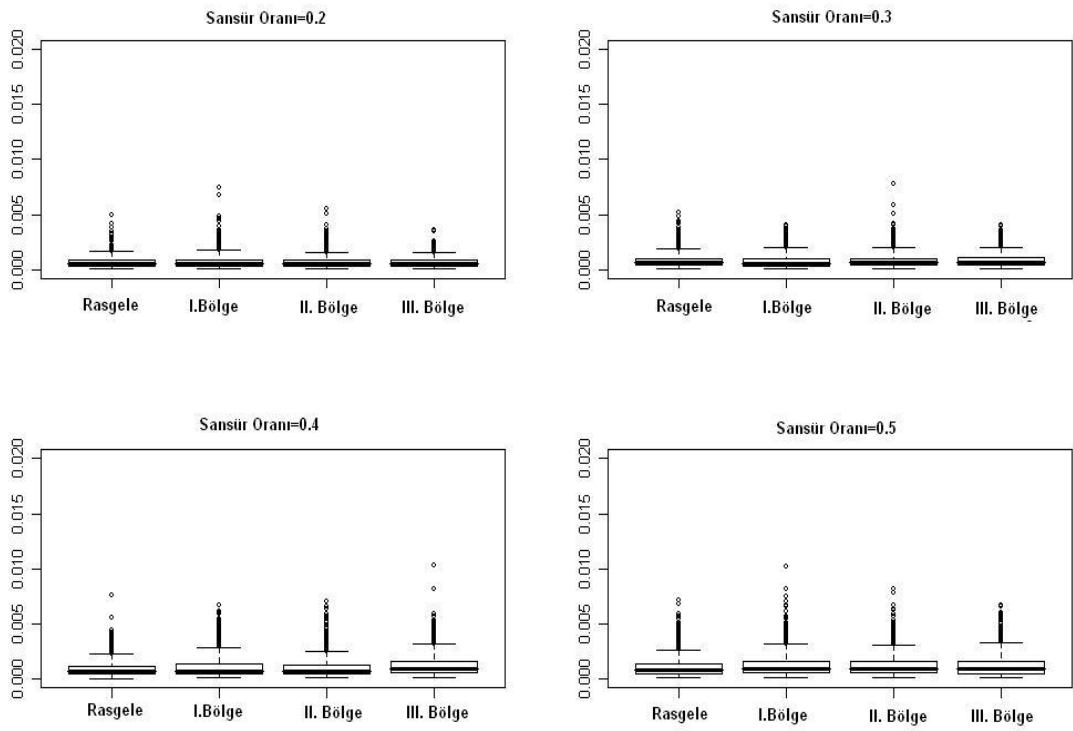
Şekil 4.33 Doğrusal artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri



Şekil 4.34. Doğrusal artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri

Şekil 4.35’de Lineer artan hazard oranı için,  $n=250$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.22’de belirtilmiştir.



Şekil 4.35 Doğrusal artan hazard oranı,  $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelgede 4.22’deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı 0.40 ve 0.50 iken, I. ve II. III. bölgede oluşan sansür oranlarının

tahminleri hatayı rasgele sansüre göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.23, Çizelge 4.24).

En yüksek hata miktarı hem 0. 40 hem de 0. 50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır.

Çizelge 4.22 Doğrusal artan hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00007	0.00030	0.00048	0.00065	0.00085	0.00494	0.356
	I. Bölge	0.00007	0.00067	0.00048	0.00071	0.00087	0.00738	
	II. Bölge	0.00006	0.00033	0.00052	0.00069	0.00083	0.00553	
	III. Bölge	0.00009	0.00031	0.00051	0.00066	0.00081	0.00363	
0.30	Rasgele	0.00008	0.00038	0.00060	0.00079	0.00097	0.00514	0.102
	I. Bölge	0.00009	0.00035	0.00059	0.00079	0.00101	0.00402	
	II. Bölge	0.00008	0.00037	0.00061	0.00082	0.00102	0.00777	
	III. Bölge	0.00012	0.00042	0.00066	0.00083	0.00106	0.00409	
0.40	Rasgele	0.00007	0.00043	0.00067	0.00093	0.00116	0.00755	0.001*
	I. Bölge	0.00009	0.00043	0.00076	0.00108	0.00142	0.00067	
	II. Bölge	0.00009	0.00045	0.00074	0.00011	0.00127	0.00708	
	III. Bölge	0.00013	0.00061	0.00098	0.00129	0.00164	0.01030	
0.50	Rasgele	0.00009	0.00049	0.00083	0.00108	0.00137	0.00719	0.001*
	I. Bölge	0.00011	0.00053	0.00090	0.00127	0.00159	0.01024	
	II. Bölge	0.00011	0.00054	0.00090	0.00125	0.00157	0.00814	
	III. Bölge	0.00013	0.00053	0.00095	0.00129	0.00165	0.00669	

Çizelge 4.23 Doğrusal artan hazard oranı, n=250, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.029	0.067	0.000
I. Bölge	-----	0.091	0.001
II. Bölge	-----	-----	0.001

Çizelge 4.24 Doğrusal artan hazard oranı, n=250, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.0138	0.024	0.003
I. Bölge	-----	0.097	0.774
II. Bölge	-----	-----	0.650

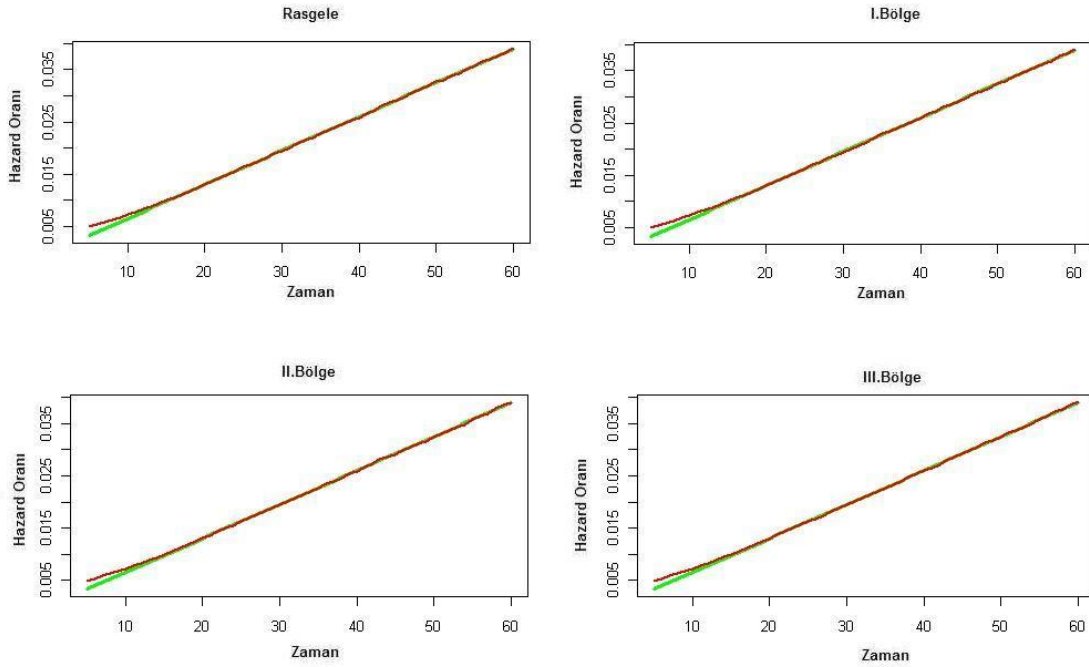
#### 4.8. Lineer Artan Hazard Oranı, N=250 için Hazard Tahminleri

N=250 iken Lineer artan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.36, Şekil 4.37 Şekil 4.38 ve Şekil 4.39’da, çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinin oldukça başarılı olduğu görülmektedir.

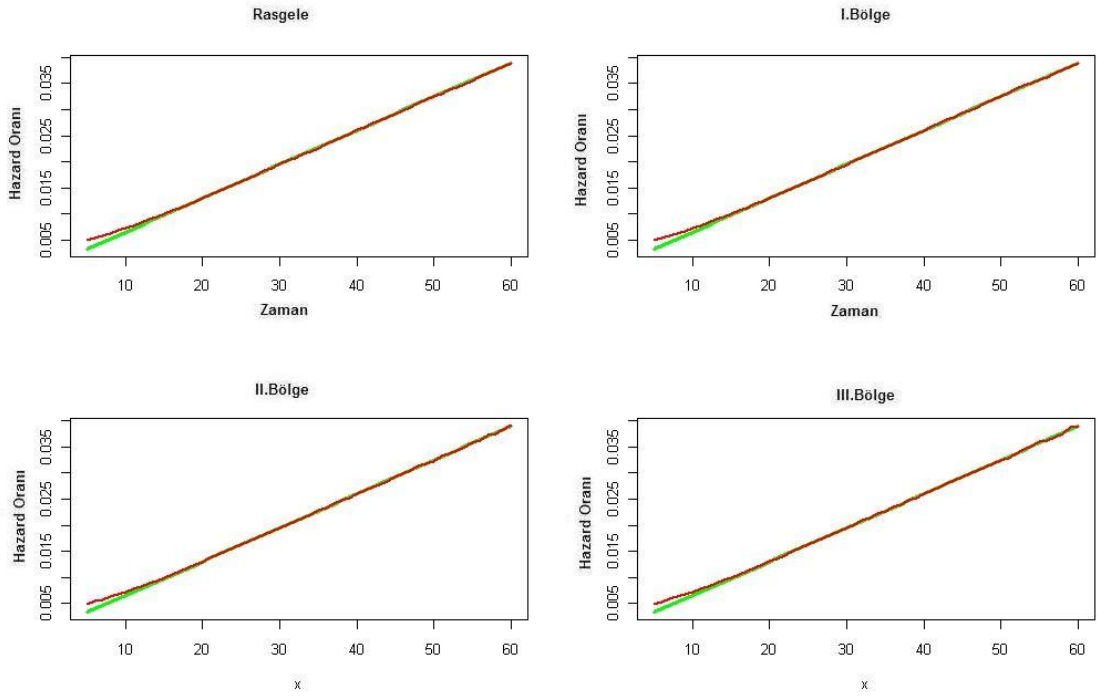
Şekil 4.40’da Lineer artan hazard oranı için, n=250 olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.25’de belirtilmiştir.

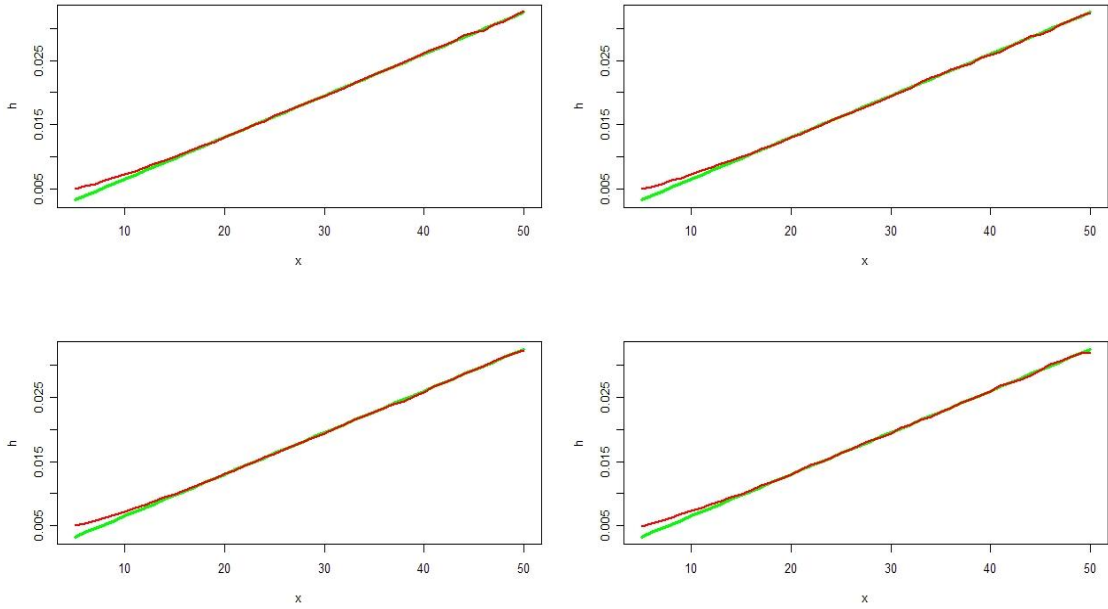
Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken sansür oranı 0.40 iken, I. Bölge tahminlerinin rasgele sansür oluşmasına göre istatistiksel olarak anlamlı düzeyde artış gösterdiği (Çizelge 4.26) ve sansür oranı 0.50 iken, III. Bölgede oluşan tahmin hatasının rasgele sansüre göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.27).



Şekil 4.36 Lineer artan hazard oranı, n=250, sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri

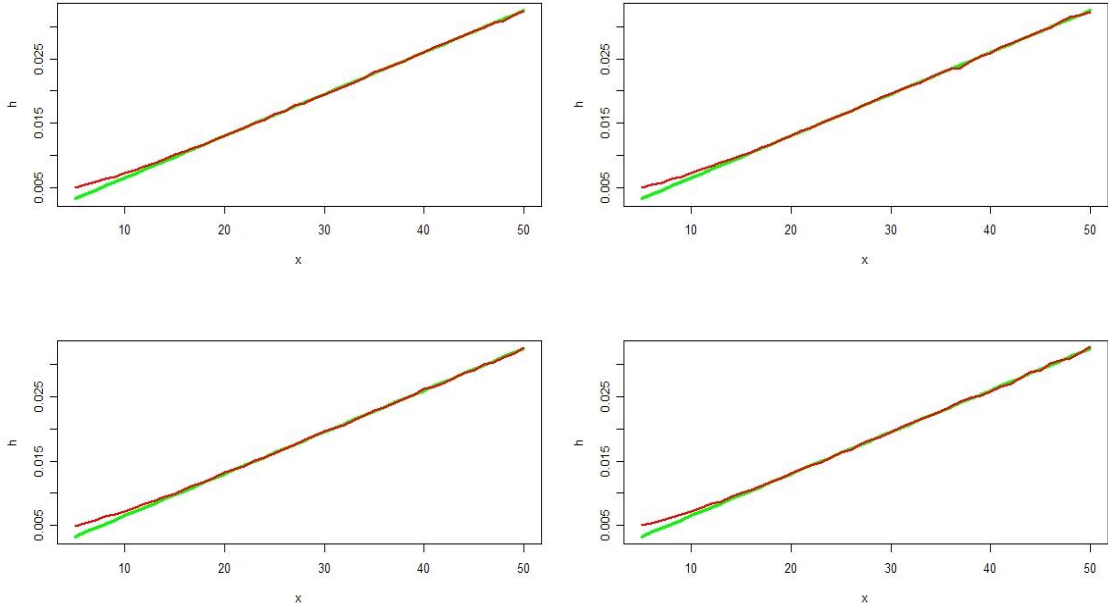


Şekil 4.37 Lineer artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri

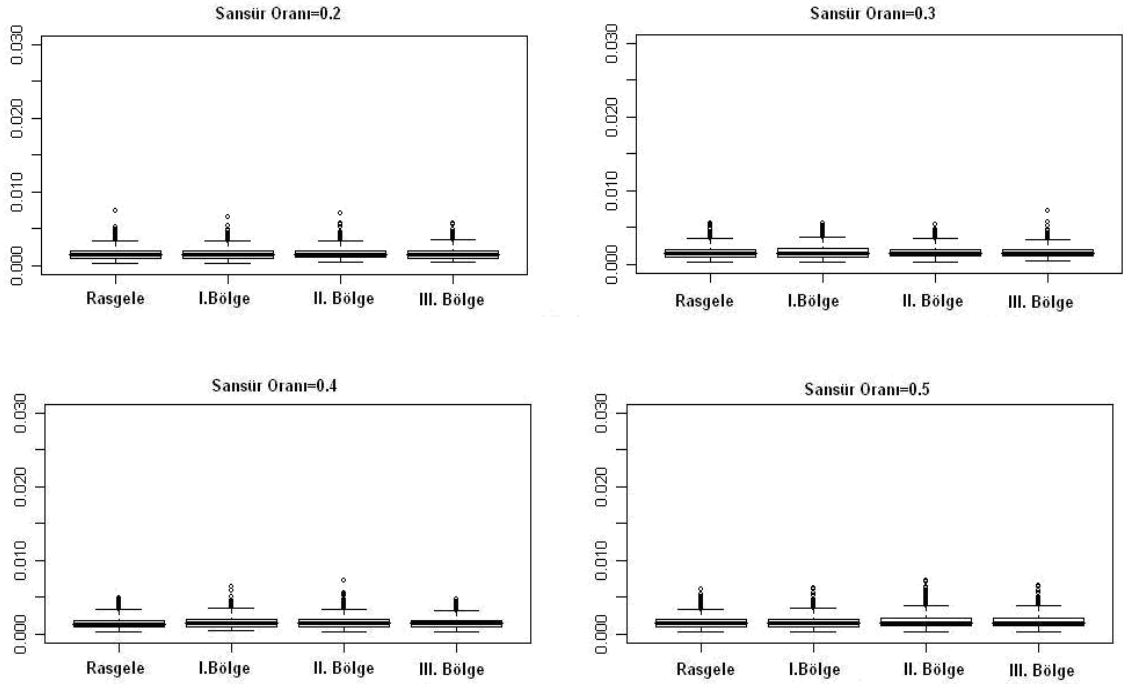


Şekil 4.38 Lineer artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri





Şekil 4.39 Lineer artan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri



Şekil 4.40 Lineer artan hazard oranı,  $n=250$  farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri.

Çizelge 4.25 Doğrusal artan hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00033	0.00105	0.00147	0.00162	0.00200	0.00739	0.436
	I. Bölge	0.00034	0.00104	0.00144	0.00159	0.00198	0.00657	
	II. Bölge	0.00038	0.00108	0.00149	0.00164	0.00201	0.00714	
	III. Bölge	0.00039	0.00105	0.00146	0.00161	0.00203	0.00572	
0.30	Rasgele	0.00031	0.00105	0.00143	0.00160	0.00202	0.00554	0.306
	I. Bölge	0.00034	0.00104	0.00148	0.00164	0.00210	0.00561	
	II. Bölge	0.00032	0.00110	0.00148	0.00164	0.00203	0.00536	
	III. Bölge	0.00038	0.00109	0.00145	0.00165	0.00200	0.00719	
0.40	Rasgele	0.00033	0.00096	0.00132	0.00150	0.00190	0.00480	0.006*
	I. Bölge	0.00038	0.00102	0.00144	0.00162	0.00205	0.00644	
	II. Bölge	0.00031	0.00098	0.00140	0.00158	0.00197	0.00727	
	III. Bölge	0.00028	0.00103	0.00140	0.00152	0.00187	0.00473	
0.50	Rasgele	0.00027	0.00101	0.00142	0.00159	0.00197	0.00598	0.001*
	I. Bölge	0.00033	0.00101	0.00147	0.00162	0.00202	0.00617	
	II. Bölge	0.00029	0.00108	0.00150	0.00174	0.00218	0.00728	
	III. Bölge	0.00036	0.00108	0.00156	0.00176	0.00223	0.00659	

Çizelge 4.26 Doğrusal artan hazard oranı, n=250, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

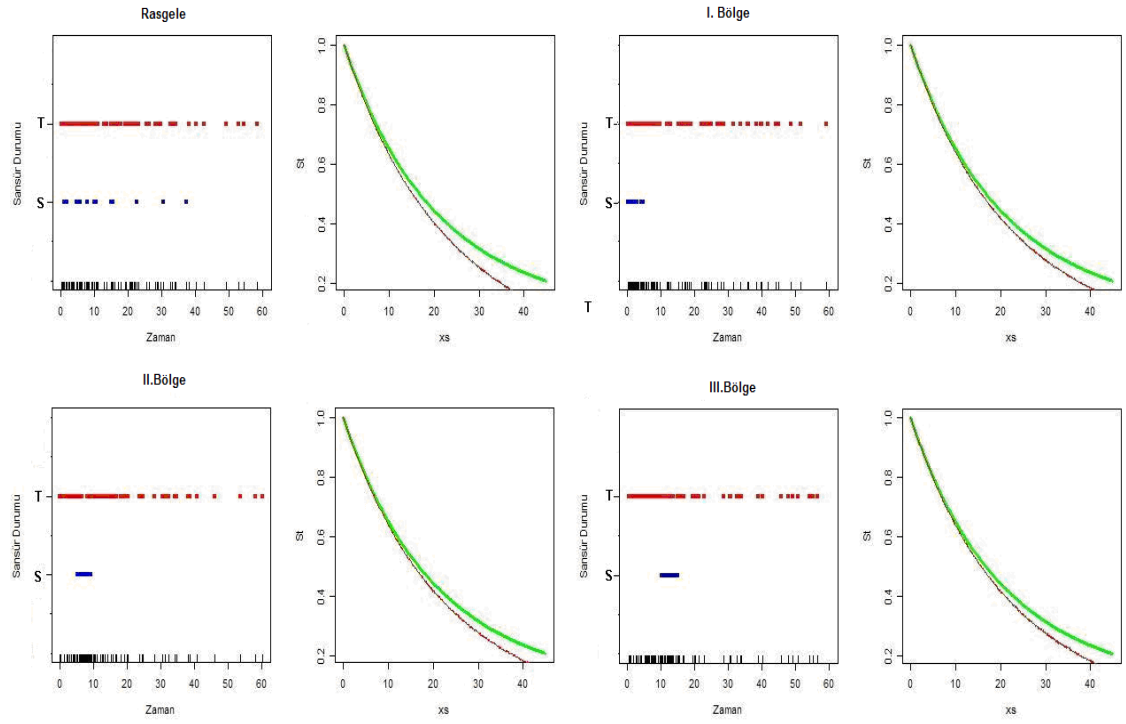
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.006	0.181	0.356
I. Bölge	-----	0.636	0.285
II. Bölge	-----	-----	0.966

Çizelge 4.27 Doğrusal artan hazard oranı, n=250, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

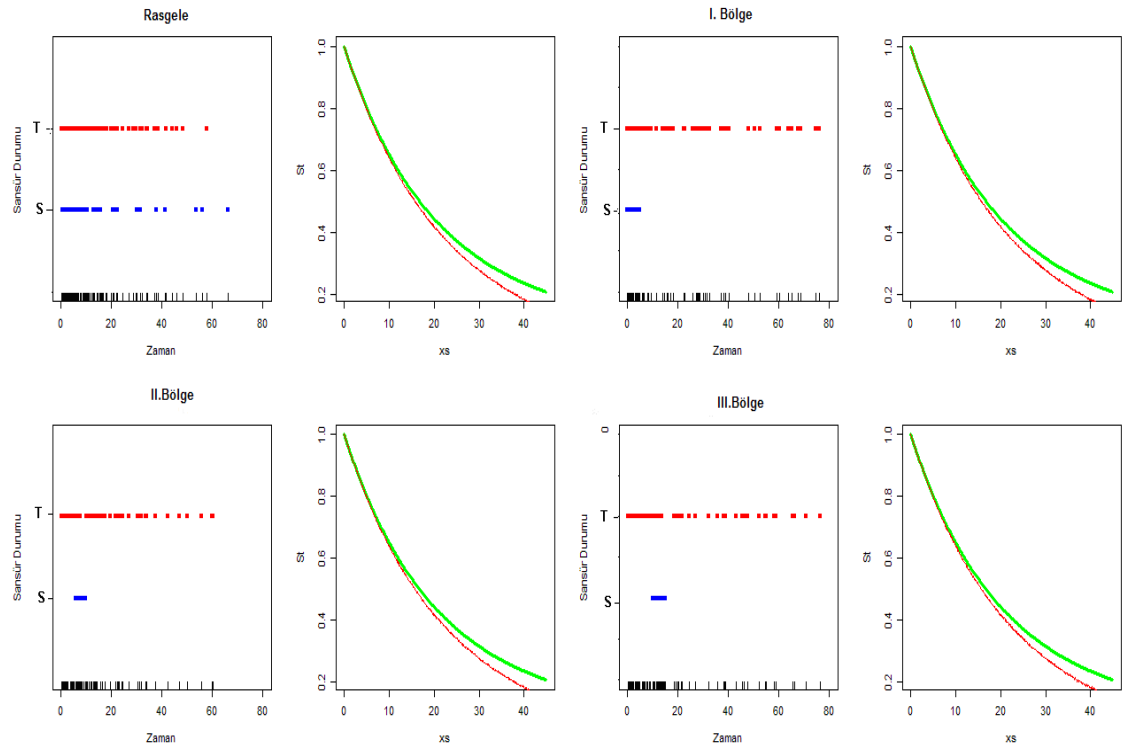
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.979	0.017	0.001
I. Bölge	-----	0.065	0.001
II. Bölge	-----	-----	0.709

#### 4.9. Linear Azalan Hazard Oranı, N=100 için Kaplan Meier Tahminleri

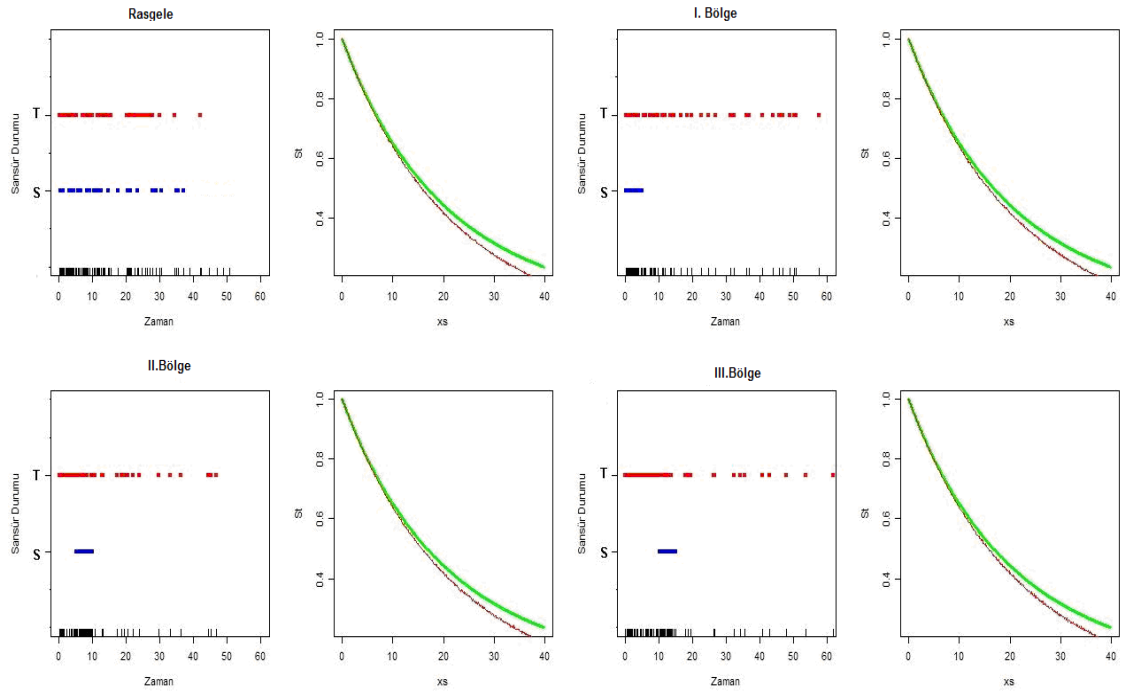
N=100 iken Linear azalan hazard oranınının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.41, Şekil 4.42, Şekil 4.43 ve Şekil 4.44'de çizilmiştir.



Şekil 4.41 Doğrusal azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri



Şekil 4.42 Doğrusal azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri

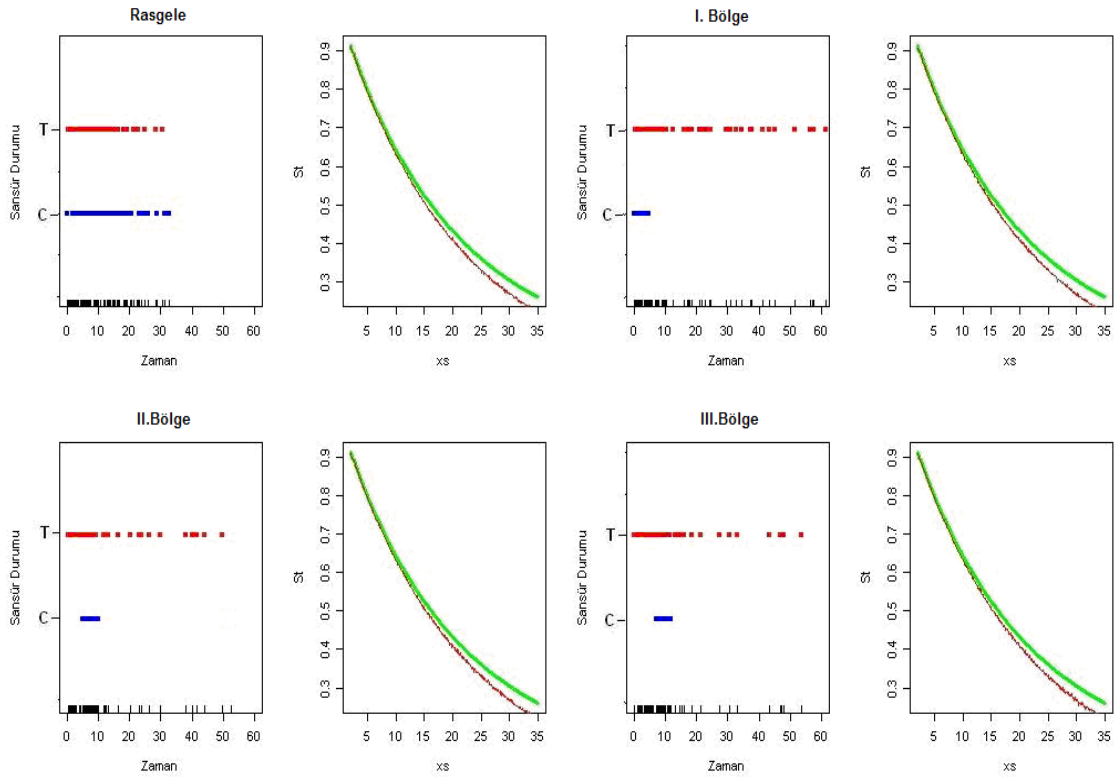


Şekil 4.43 Doğrusal azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri

Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde başlangıçta oldukça başarılı olduğu fakat sonlara yaklaştıkça bir miktar sapmalar olduğu görülmektedir.

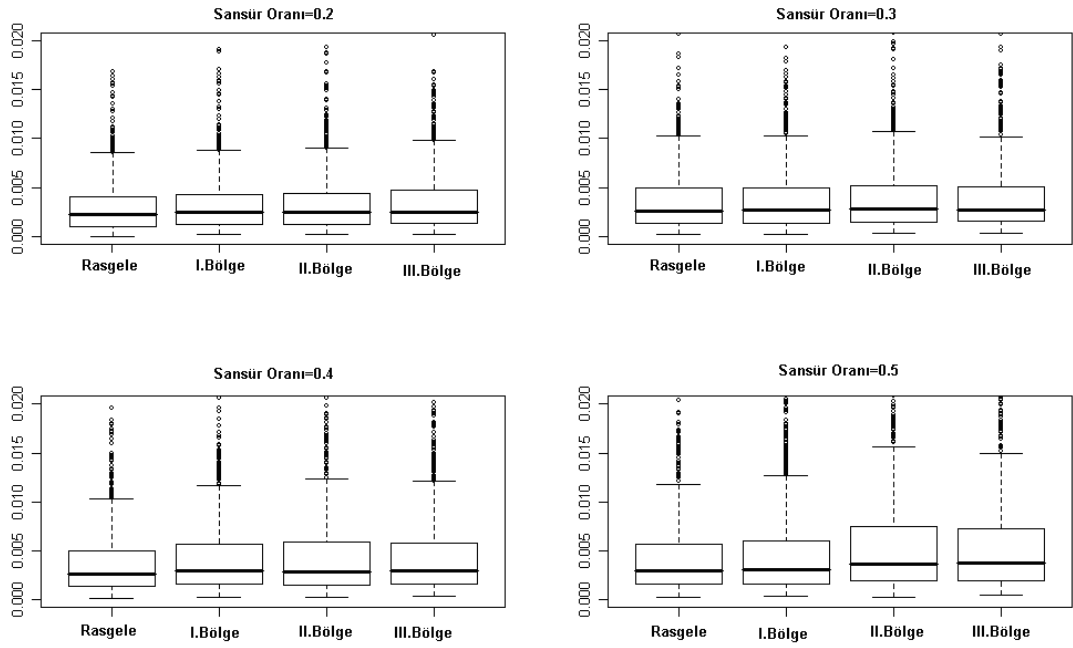
Şekil 4.45’de ise Lineer azalan hazard oranı için,  $n=100$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir.

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.28’de belirtilmiştir. Çizelgede 4.28’deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir.



Şekil 4.44 Doğrusal azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri

Sansürlü gözlem oranı 0.20 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı ve 0.30, 0.40 ve 0.50 iken Sansür oranının 0.50 olduğu durumda ise tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.29, Çizelge 4.30, Çizelge 4.31).



Şekil 4.45 Linear azalan hazard oranı, n=100 farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.28 Doğrusal azalan hazard oranı, n=100, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00018	0.00121	0.00225	0.00329	0.00421	0.02190	0.171
	I. Bölge	0.00021	0.00135	0.00243	0.00342	0.00451	0.02067	
	II. Bölge	0.00018	0.00124	0.00242	0.00337	0.00426	0.02668	
	III. Bölge	0.00020	0.00137	0.00249	0.00353	0.00457	0.02059	
0.30	Rasgele	0.00021	0.00135	0.00259	0.00358	0.00493	0.02061	0.000*
	I. Bölge	0.00020	0.00138	0.00273	0.00384	0.00499	0.02510	
	II. Bölge	0.00025	0.00141	0.00279	0.00385	0.00513	0.02416	
	III. Bölge	0.00032	0.00150	0.00270	0.00390	0.00499	0.02922	
0.40	Rasgele	0.00017	0.00141	0.00268	0.00371	0.00498	0.02212	0.001*
	I. Bölge	0.00023	0.00157	0.00301	0.00438	0.00566	0.03646	
	II. Bölge	0.00025	0.00151	0.00284	0.00443	0.00585	0.02671	
	III. Bölge	0.00035	0.00161	0.00301	0.00455	0.00581	0.03989	
0.50	Rasgele	0.00021	0.00162	0.00298	0.00429	0.00571	0.03418	0.000*
	I. Bölge	0.00032	0.00162	0.00313	0.00490	0.00607	0.04479	
	II. Bölge	0.00030	0.00199	0.00360	0.00559	0.00744	0.04076	
	III. Bölge	0.00043	0.00200	0.00379	0.00568	0.00722	0.03604	

En yüksek hata miktarı hem 0. 40 hem de 0. 50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır.

Çizelge 4.29 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.000	0.000	0.000
I. Bölge	-----	0.999	0.750
II. Bölge	-----	-----	0.691

Çizelge 4.30 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.0115	0.059	0.001
I. Bölge	-----	0.942	0.951
II. Bölge	-----	-----	0.679

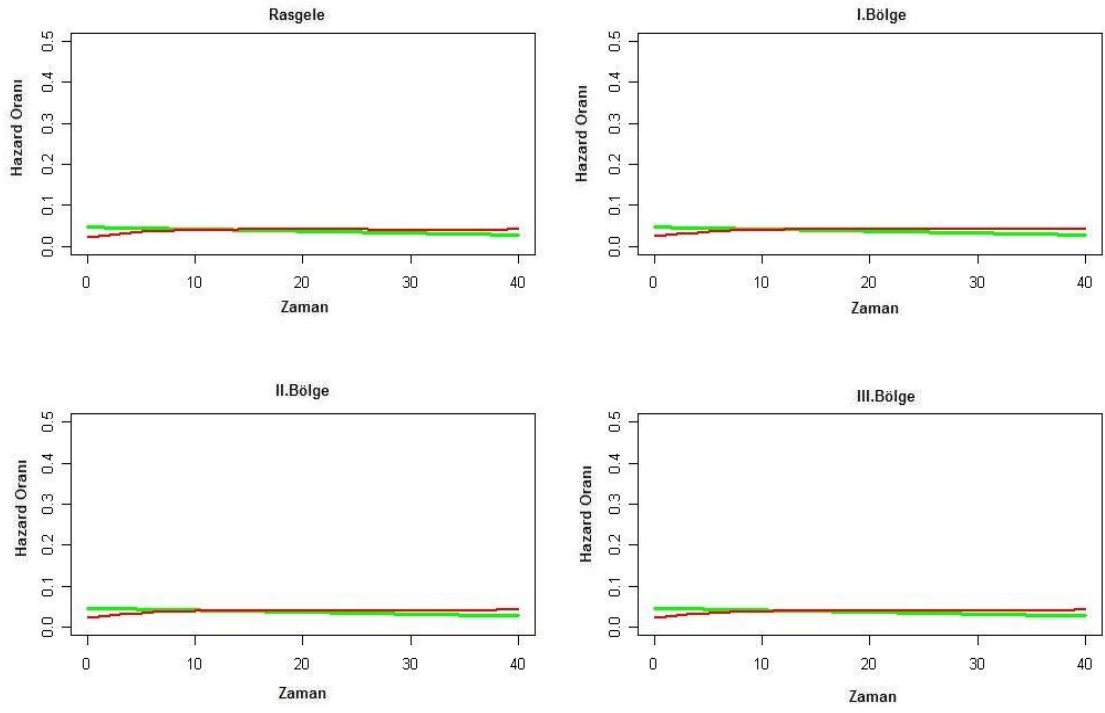
Çizelge 4.31 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.0117	0.059	0.001
I. Bölge	-----	0.942	0.951
II. Bölge	-----	-----	0.679

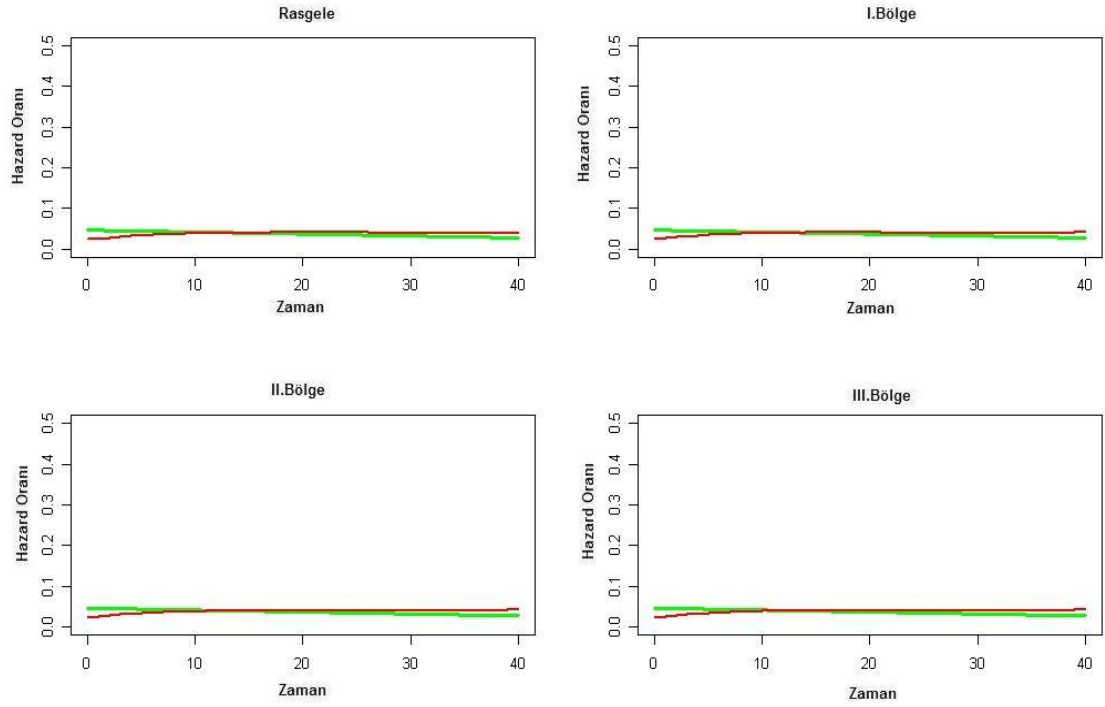
#### 4.10. Lineer Azalan Hazard Oranı, N=100, için Hazard Oranı Tahminleri

N=100 iken Lineer azalan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.46, Şekil 4.47, Şekil 4.48 ve Şekil 4.49'da, çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinde tüm bölgelerde başlangıçta ve sonda sapmalar olduğu fakat orta kısımlarda oldukça başarılı olduğu görülmektedir.

Şekil 4.50'de Lineer azalan hazard oranı için, n=100 olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.

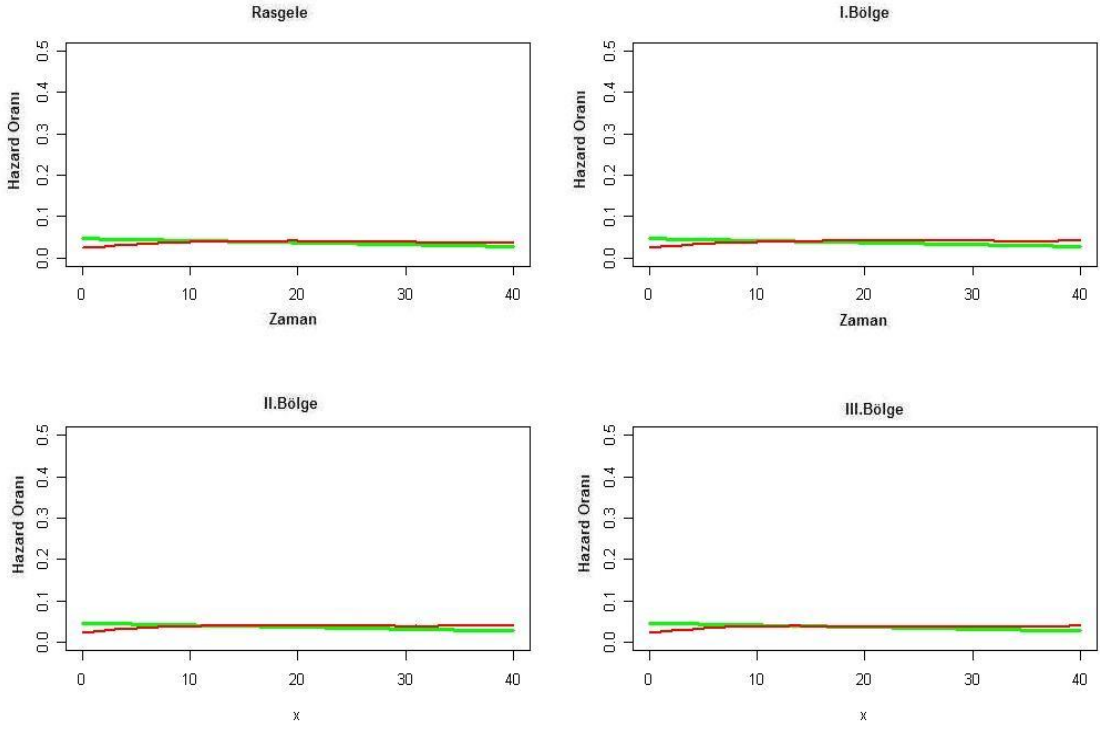


Şekil 4.46 Lineer azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri

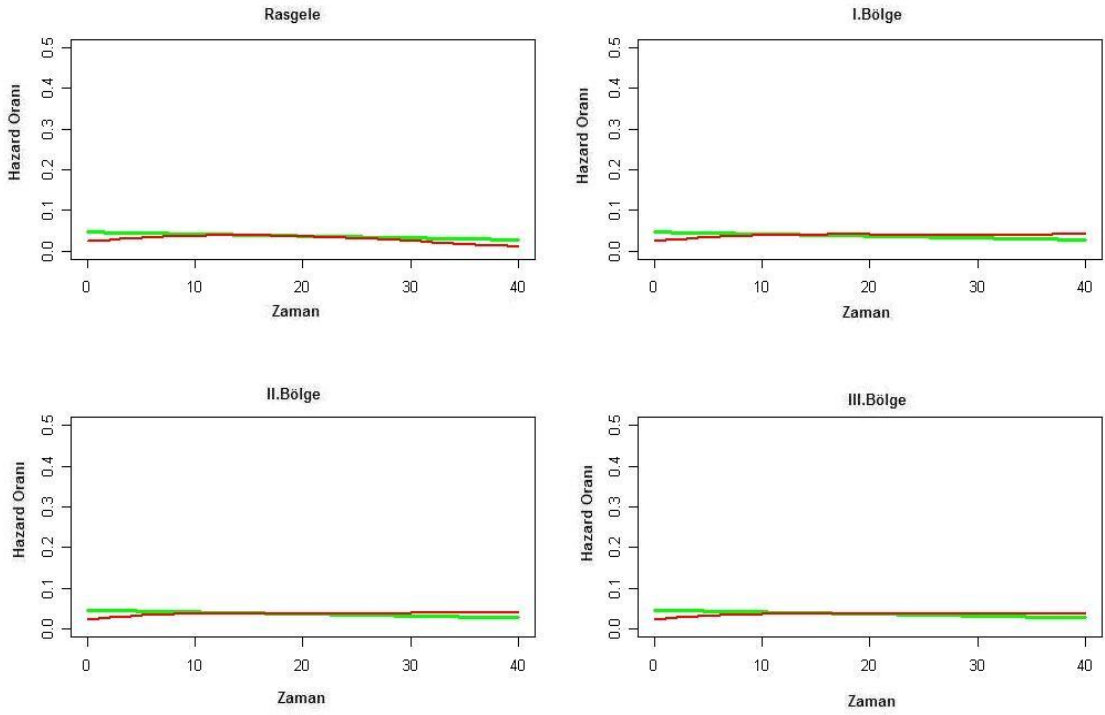


Şekil 4.47 Lineer azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri

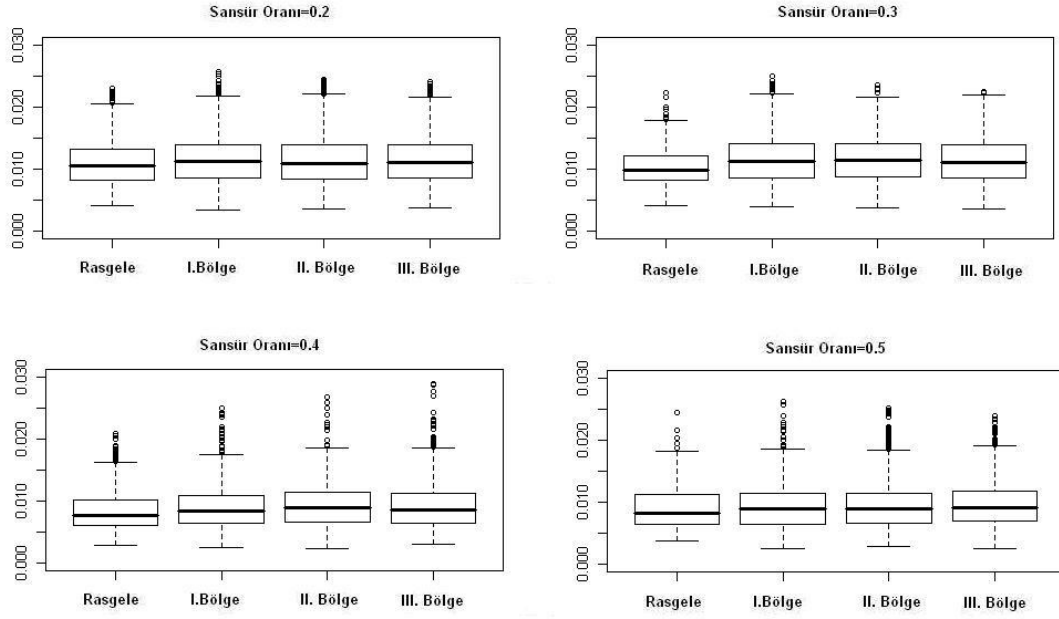




Şekil 4.48 Lineer azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri



Şekil 4.49 Lineer azalan hazard oranı,  $n=100$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri



Şekil 4.50 Linear azalan hazard oranı, n=100 farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.32’de belirtilmiştir.

Çizelge 4.32 Linear azalan hazard oranı, n=100, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıtıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00263	0.00602	0.00821	0.00849	0.01047	0.02424	0.073
	I. Bölge	0.00242	0.00639	0.00849	0.00876	0.01074	0.02133	
	II. Bölge	0.00258	0.00624	0.00861	0.00876	0.01075	0.02353	
	III. Bölge	0.00284	0.00634	0.00856	0.00885	0.01069	0.02476	
0.30	Rasgele	0.00242	0.00610	0.00786	0.00837	0.00991	0.02688	0.001*
	I. Bölge	0.00229	0.00663	0.00871	0.00891	0.01074	0.02314	
	II. Bölge	0.00286	0.00620	0.00832	0.00873	0.01070	0.02652	
	III. Bölge	0.00313	0.00634	0.00839	0.00890	0.01071	0.02520	
0.40	Rasgele	0.00289	0.00599	0.00767	0.00832	0.01011	0.02092	0.001*
	I. Bölge	0.00248	0.00638	0.00842	0.00889	0.01081	0.02496	
	II. Bölge	0.00237	0.00658	0.00882	0.00929	0.01143	0.02680	
	III. Bölge	0.00305	0.00636	0.00861	0.00923	0.01128	0.02881	
0.50	Rasgele	0.00366	0.00637	0.00823	0.00897	0.01116	0.02435	0.001*
	I. Bölge	0.00250	0.00648	0.00887	0.00924	0.01136	0.02618	
	II. Bölge	0.00290	0.00664	0.00882	0.00940	0.01136	0.02507	
	III. Bölge	0.00250	0.00693	0.00903	0.00975	0.01180	0.02378	

Çizelge 4.29'daki değerlere bakıldığında genel olarak sansürlü gözlem miktarının artırılmasının bölgelerdeki tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken sansür oranı ve 0.30, 0.40 ve 0.50 iken, tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.33, Çizelge 4.34, Çizelge 4.35). En yüksek hata miktarı 0.40 ve 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.32).

Çizelge 4.33 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0,000	0.066	0.001
I. Bölge	-----	0.129	0.651
II. Bölge	-----	-----	0.731

Çizelge 4.34 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

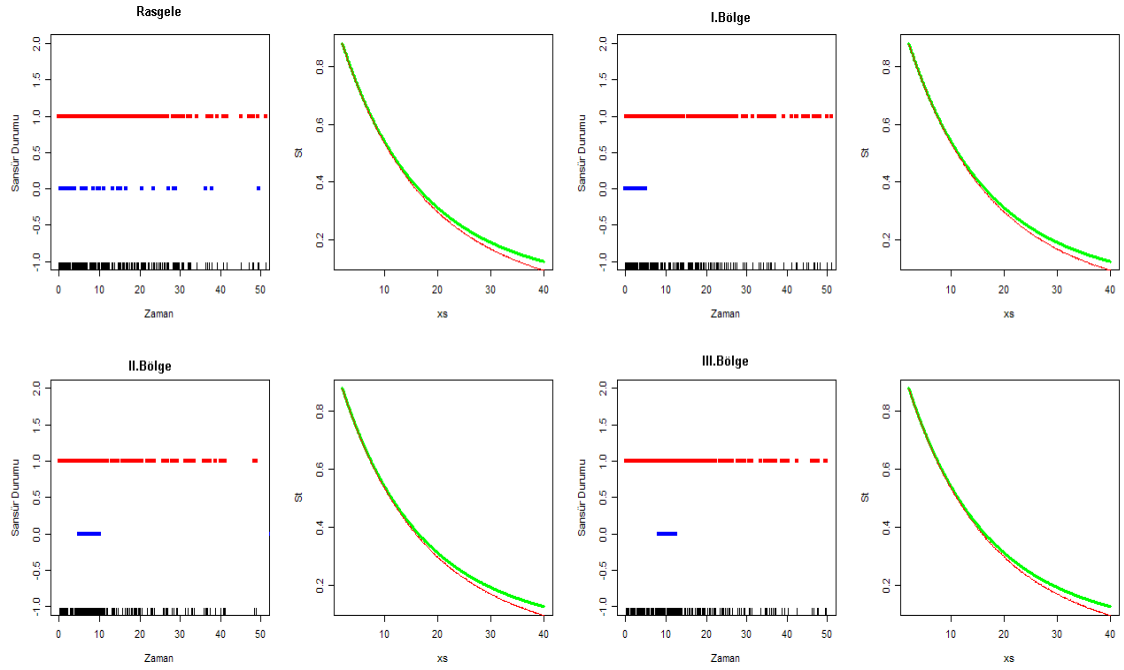
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0,001	0.001	0.001
I. Bölge	-----	0.054	0.489
II. Bölge	-----	-----	0.737

Çizelge 4.35 Doğrusal azalan hazard oranı, n=100, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0,283	0.132	0.001
I. Bölge	-----	0.996	0.162
II. Bölge	-----	-----	0.257

#### 4.11. Lineer Azalan Hazard Oranı, N=250 için Kaplan Meier Tahminleri

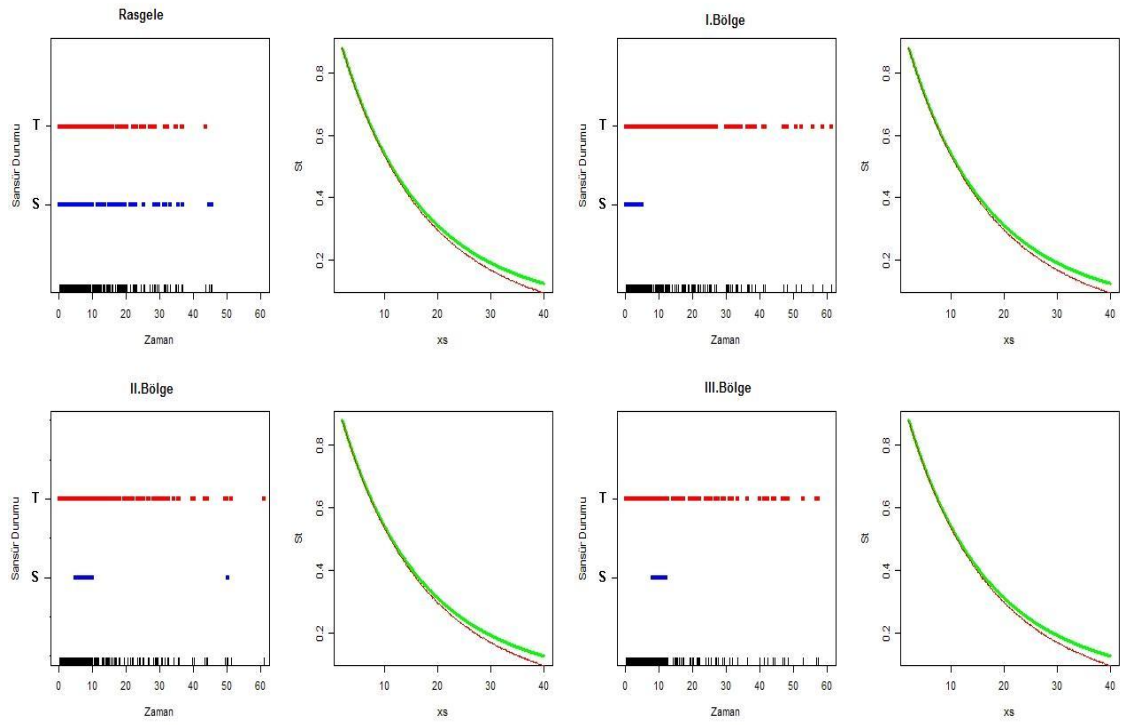
N=250 iken Lineer azalan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.51, Şekil 4.52, Şekil 4.53 ve Şekil 4.54’de, çizilmiştir. Şekillerde sırasıyla sansürlü gözlemlerin veri içindeki yerleşimi ve ilgili veriye ilişkin Kaplan Meier yöntemiyle tahmin edilmiş hayatta kalma fonksiyonları verilmiştir. Tahmin grafiklerinin tüm bölgelerde başlangıçta oldukça başarılı olduğu fakat sonlara yaklaştıkça bir miktar sapsmalar olduğu görülmektedir.



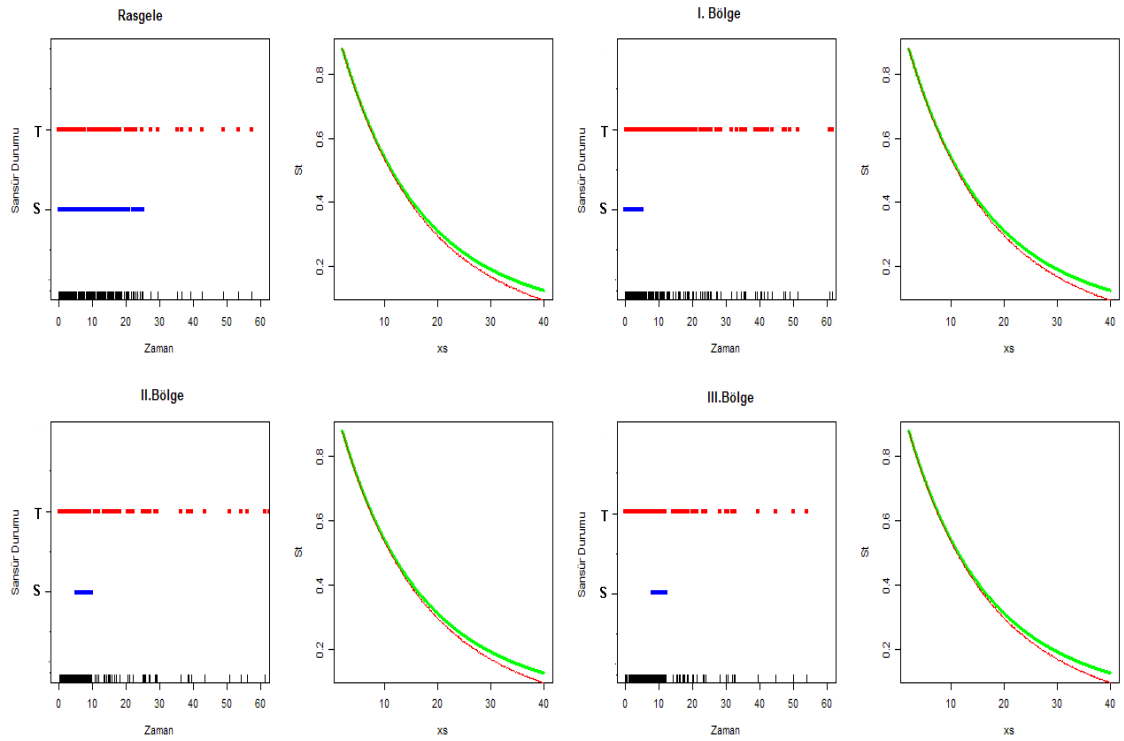
Şekil 4.51 Doğrusal azalan hazard oranı, n=250, sansürlü gözlem oranı 0.20 iken Kaplan Meier tahminleri

Şekil 4.55’de ise Lineer azalan hazard oranı için, n=250 olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait tahmin hatalarının box plot grafikleri verilmiştir.

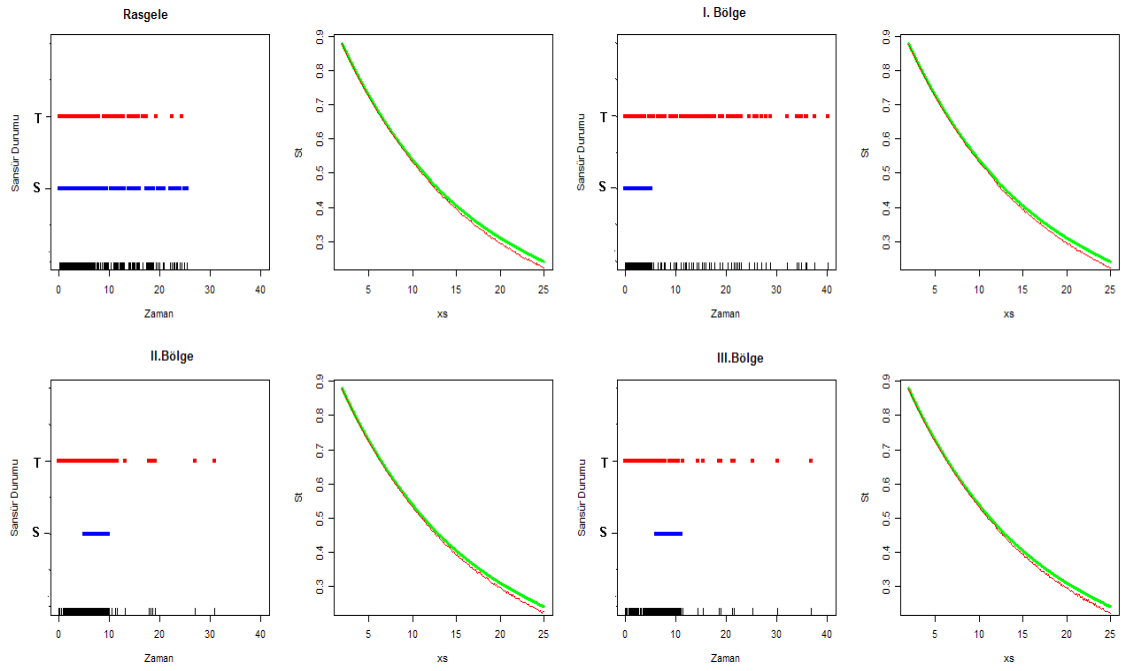
Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıttıcı istatistikler ve Çizelge 4.36’da belirtilmiştir.



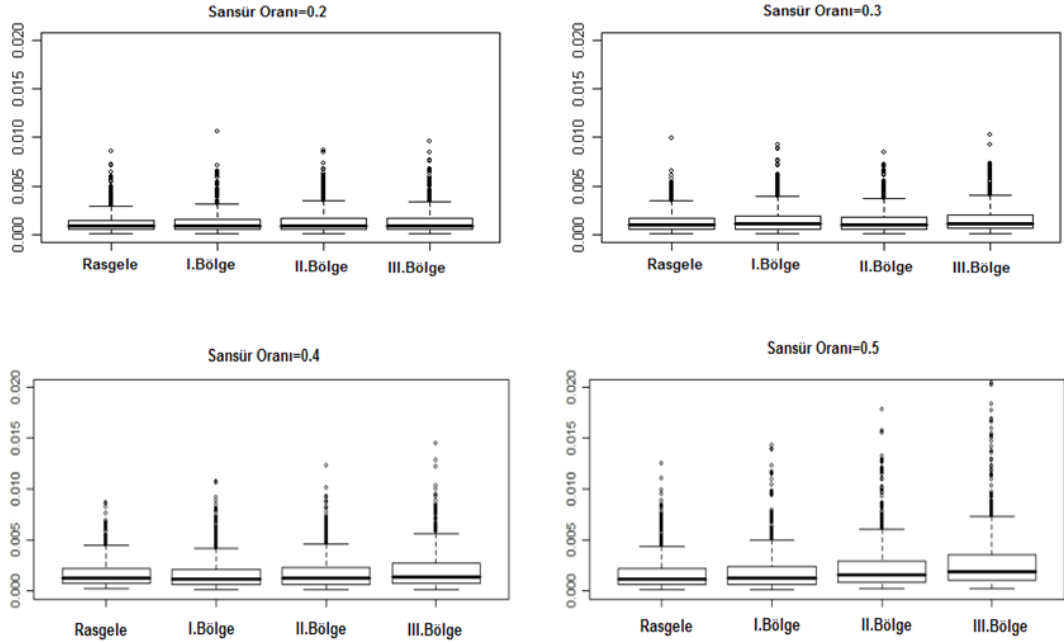
Şekil 4.52 Doğrusal azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Kaplan Meier tahminleri



Şekil 4.53 Doğrusal azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Kaplan Meier tahminleri



Şekil 4.54 Doğrusal azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Kaplan Meier tahminleri



Şekil 4.55 Doğrusal azalan hazard oranı,  $n=250$ , farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.33'deki değerlere bakıldığında sansürlü gözlem miktarının artırılmasının tüm bölgelerde tahmin hatasını yükselttiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken, sansür oranı 0.40 ve 0.50 iken tahmin hatasının rasgele sansürden son bölgeye doğru gidildikçe arttığı görülmüştür (Çizelge 4.37, Çizelge 4.38).

En yüksek hata miktarı hem 0.40 hem de 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır.

Çizelge 4.36 Doğrusal azalan hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıttıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.00009	0.00048	0.00085	0.00116	0.00147	0.00860	0.037*
	I. Bölge	0.00008	0.00051	0.00088	0.00124	0.00157	0.01059	
	II. Bölge	0.00007	0.00051	0.00092	0.00132	0.00171	0.00864	
	III. Bölge	0.00007	0.00050	0.00092	0.00129	0.00167	0.00957	
0.30	Rasgele	0.00010	0.00053	0.00096	0.00130	0.00170	0.00990	0.037*
	I. Bölge	0.00007	0.00056	0.00106	0.00144	0.00190	0.00928	
	II. Bölge	0.00007	0.00057	0.00103	0.00140	0.00018	0.00844	
	III. Bölge	0.00008	0.00061	0.00114	0.00152	0.00199	0.01029	
0.40	Rasgele	0.00019	0.00073	0.00132	0.00171	0.00223	0.00864	0.001*
	I. Bölge	0.00013	0.00065	0.00111	0.00162	0.00207	0.01073	
	II. Bölge	0.00016	0.00069	0.00122	0.00172	0.00228	0.01234	
	III. Bölge	0.00015	0.00078	0.00141	0.00203	0.00277	0.01449	
0.50	Rasgele	0.00007	0.00066	0.00118	0.00173	0.00218	0.01250	0.001*
	I. Bölge	0.00011	0.00068	0.00122	0.00192	0.00242	0.01433	
	II. Bölge	0.00019	0.00085	0.00156	0.00234	0.00294	0.01788	
	III. Bölge	0.00018	0.00102	0.00187	0.00279	0.00353	0.02433	

Çizelge 4.37 Doğrusal azalan hazard oranı, n=250, sansürlü gözlem oranı 0.40 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.008	0.0641	0.017
I. Bölge	-----	0.217	0.004
II. Bölge	-----	-----	0.001

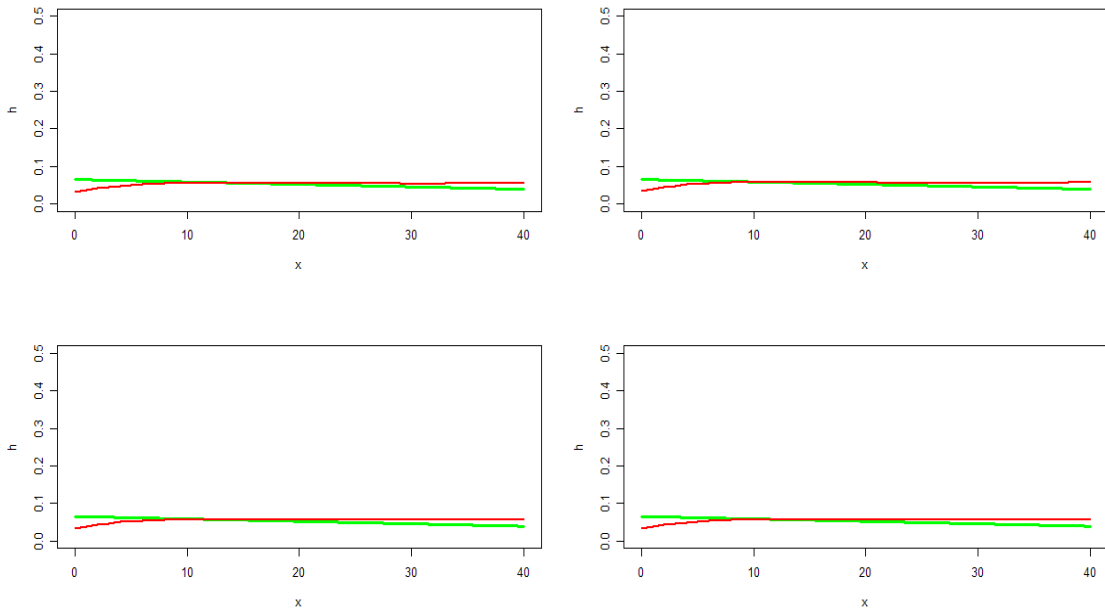
Çizelge 4.38 Doğrusal azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.008	0.006	0.001
I. Bölge	-----	0.002	0.001
II. Bölge	-----	-----	0.001

#### 4.12. Lineer Azalan Hazard Oranı, $N=250$ , için Hazard Oranı Tahminleri

$N=250$  iken Lineer azalan hazard oranının 0.20, 0.30, 0.40 ve 0.50 sansür oranlarında tahmin grafikleri Şekil 4.56, Şekil 4.57, Şekil 4.58 ve Şekil 4.59’de çizilmiştir. Şekillerde ilgili veriye ilişkin Kernel düzgünleştirme yöntemiyle tahmin edilmiş hazard fonksiyonları verilmiştir. Tahmin grafiklerinde tüm bölgelerde başlangıçta ve sonda sapmalar olduğu fakat orta kısımlarda oldukça başarılı olduğu görülmektedir.

Şekil 4.60’de Lineer azalan hazard oranı için,  $n=250$  olduğu durumda farklı sansürlü gözlem oranları ve yerleşimlerine ait Kernel düzgünleştirme yöntemi tahmin hatalarının box plot grafikleri verilmiştir.

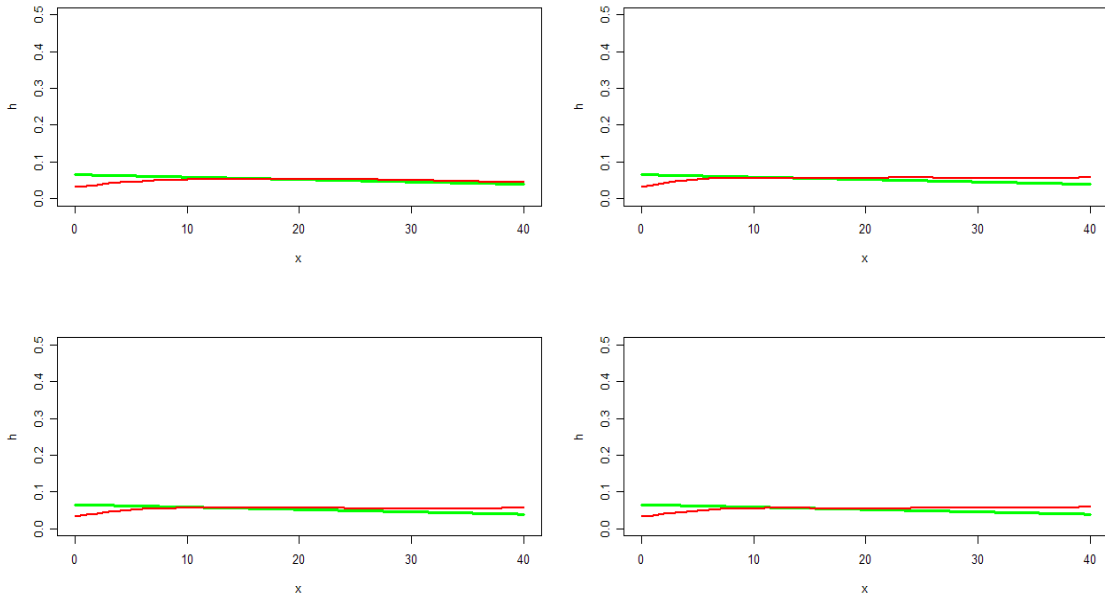


Şekil 4.56 Lineer azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.20 iken Hazard Oranı tahminleri



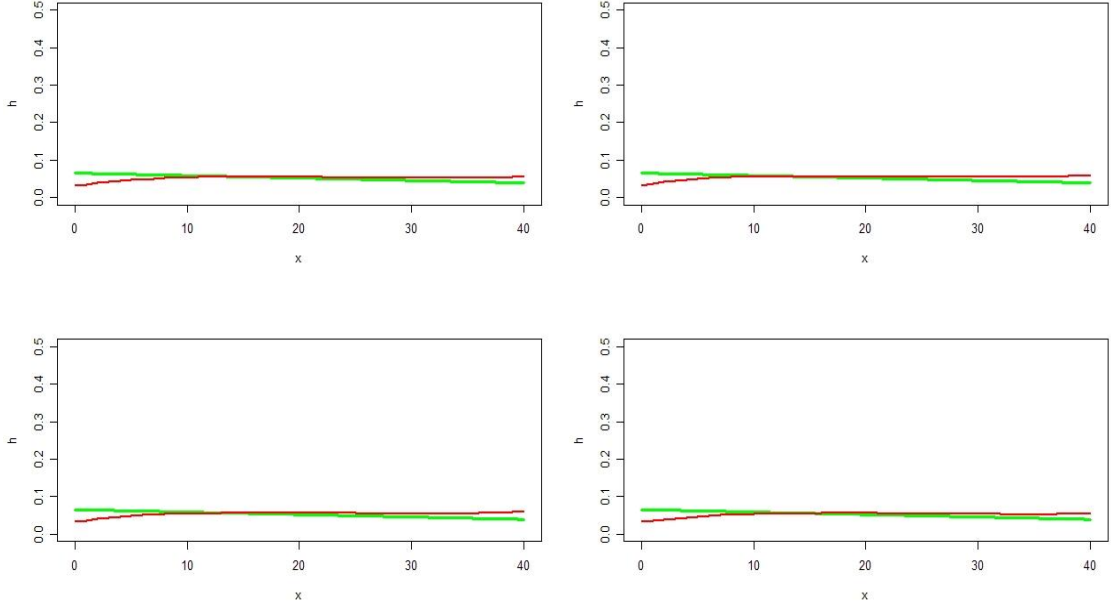
Sansürlü gözlem bölgelerinin ve yüzdelerinin değiştirildiği durumda oluşacak olan tahmin hatalarına ait tanıtıcı istatistikler Çizelge 4.39’da belirtilmiştir.

Çizelge 4.39’daki değerlere bakıldığında genel olarak sansürlü gözlem miktarının artırılmasının bölgelerdeki tahmin hatasını yükseldiği gözlenmektedir. Sansürlü gözlem oranı 0.20 ve 0.30 olduğu durumda 3 farklı sansür bölgesindeki tahminler ile tamamen rasgele sansür oluşması arasında istatistiksel olarak anlamlı farklılık görülmez iken sansür oranı 0.40 iken, II. ve III. Bölge tahminlerinin rasgele sansür oluşmasına göre istatistiksel olarak anlamlı düzeyde artış gösterdiği (Çizelge 4.40) ve sansür oranı 0.50 iken, III. Bölgede oluşan tahmin hatasının rasgele sansüre göre istatistiksel olarak anlamlı düzeyde arttırdığı gözlenmiştir (Çizelge 4.41).

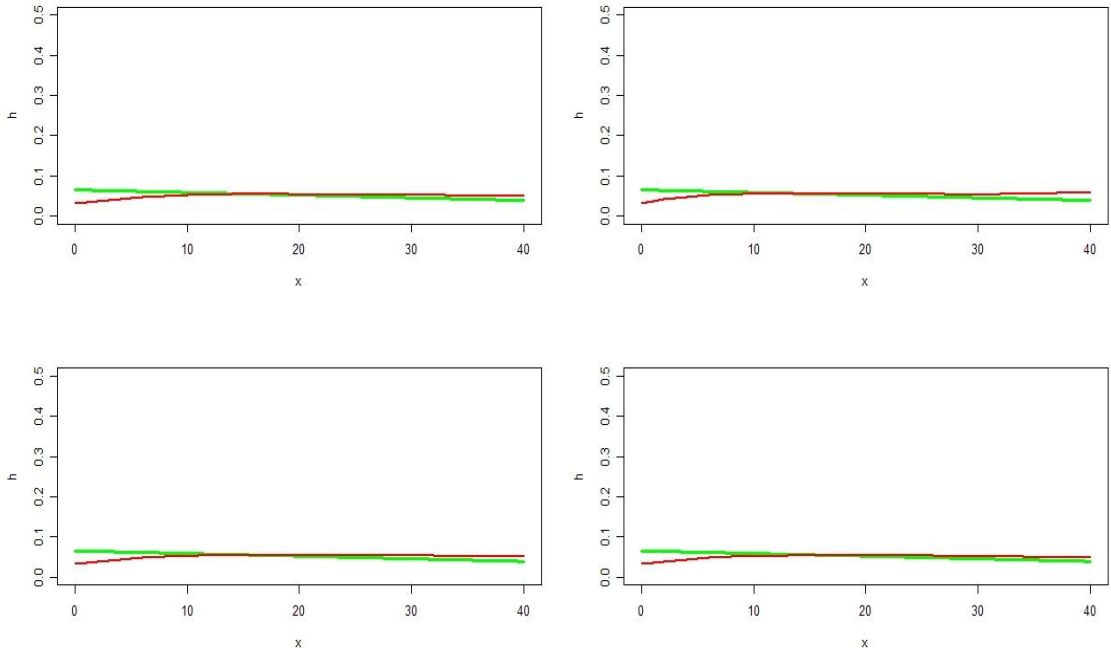


Şekil 4.57 Lineer azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.30 iken Hazard Oranı tahminleri

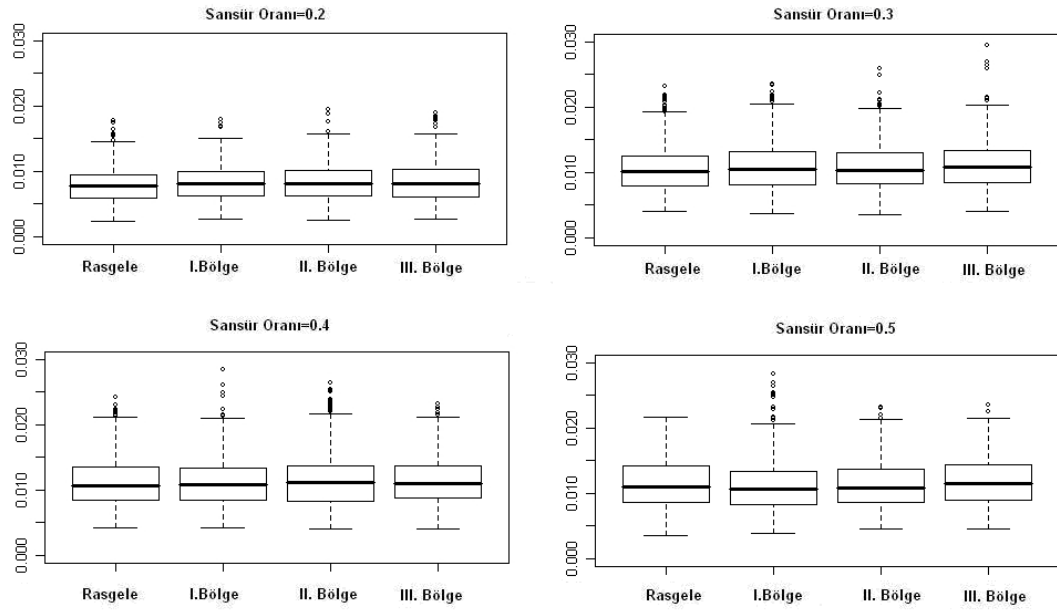
En yüksek hata miktarı 0.40 ve 0.50 sansür oranları için III. Bölgede oluşan sansürlü gözlemlerde ortaya çıkmıştır (Çizelge 4.39).



Şekil 4.58 Lineer azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.40 iken Hazard Oranı tahminleri



Şekil 4.59 Lineer azalan hazard oranı,  $n=250$ , sansürlü gözlem oranı 0.50 iken Hazard Oranı tahminleri



Şekil 4.60 Lineer azalan hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait box plot grafikleri

Çizelge 4.39 Doğrusal azalan hazard oranı, n=250, farklı sansürlü gözlem oranları ve yerleşimleri için tahmin hatalarına ait tanıttıcı istatistikler

Oran	Bölge	Minimum	1st Qu	Medyan	Ortalama	3rd Qu.	Maksimum	P
0.20	Rasgele	0.002317	0.005969	0.007737	0.007842	0.009421	0.017770	0.070
	I. Bölge	0.002673	0.006206	0.008082	0.008210	0.009984	0.017920	
	II. Bölge	0.002451	0.006232	0.008094	0.008267	0.010120	0.019420	
	III. Bölge	0.002708	0.006070	0.008091	0.008356	0.010260	0.018890	
0.30	Rasgele	0.004020	0.008002	0.010170	0.010560	0.012520	0.023200	0.0932
	I. Bölge	0.003686	0.008117	0.010470	0.010800	0.013110	0.023580	
	II. Bölge	0.003512	0.008271	0.010290	0.010730	0.012930	0.025860	
	III. Bölge	0.004057	0.008410	0.010790	0.011060	0.013380	0.029450	
0.40	Rasgele	0.004192	0.008385	0.010710	0.011300	0.013560	0.024140	0.001*
	I. Bölge	0.004124	0.008360	0.010870	0.011080	0.013410	0.028490	
	II. Bölge	0.004107	0.008332	0.011230	0.011520	0.013740	0.026420	
	III. Bölge	0.004025	0.008795	0.011030	0.011480	0.013750	0.023270	
0.50	Rasgele	0.00360	0.00858	0.01094	0.01148	0.01416	0.02169	0.003*
	I. Bölge	0.003908	0.008276	0.010670	0.011090	0.013310	0.028270	
	II. Bölge	0.004534	0.008553	0.010880	0.011480	0.013660	0.023120	
	III. Bölge	0.004516	0.008871	0.011400	0.011700	0.014300	0.023550	

Çizelge 4.40 Doğrusal azalan hazard oranı, n=250, sansürlü gözlem oranı 0.30 için çoklu karşılaştırma sonuçları

	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.001	0.002	0.001
I. Bölge	-----	0.990	0.968
II. Bölge	-----	-----	0.986

Çizelge 4.41 Doğrusal azalan hazard oranı, n=250, sansürlü gözlem oranı 0.50 için çoklu karşılaştırma sonuçları

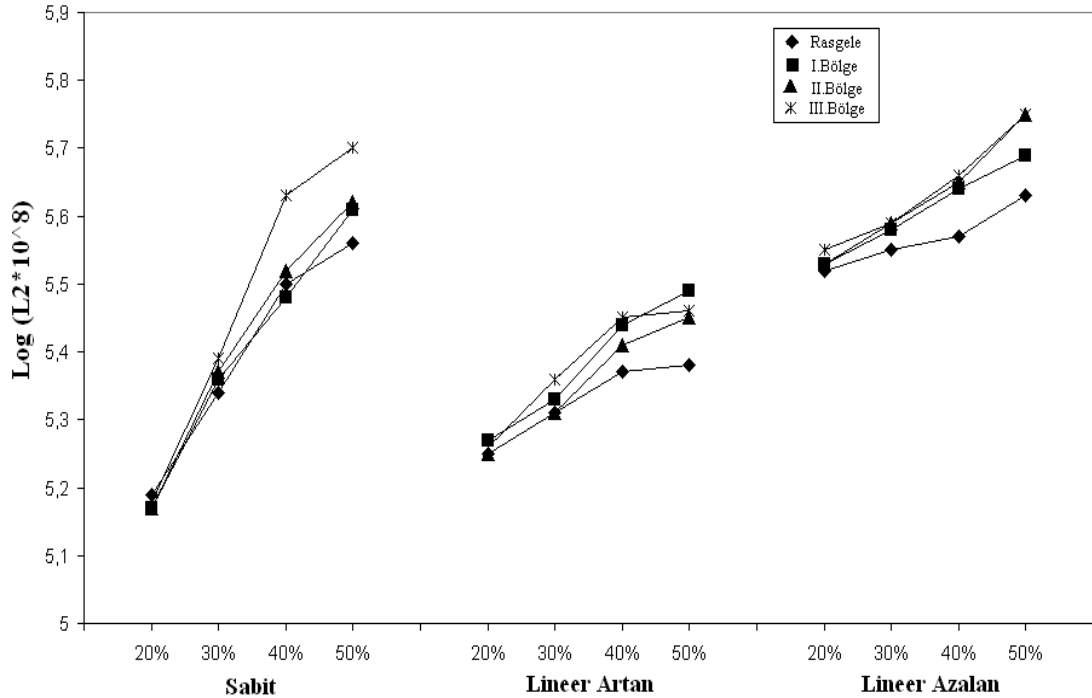
	I. Bölge	II. Bölge	III. Bölge
Rasgele	0.062	0.997	0.315
I. Bölge	-----	0.147	0.001
II. Bölge	-----	-----	0.175

## 5. TARTIŞMA

Bu çalışmada 3 farklı dağılım, 4 farklı sansür oranı, 4 farklı sansürlü gözlem yerleşim yeri ve iki farklı örnek genişliği kullanılarak oldukça detaylı bir simülasyon çalışması ortaya konmuştur.

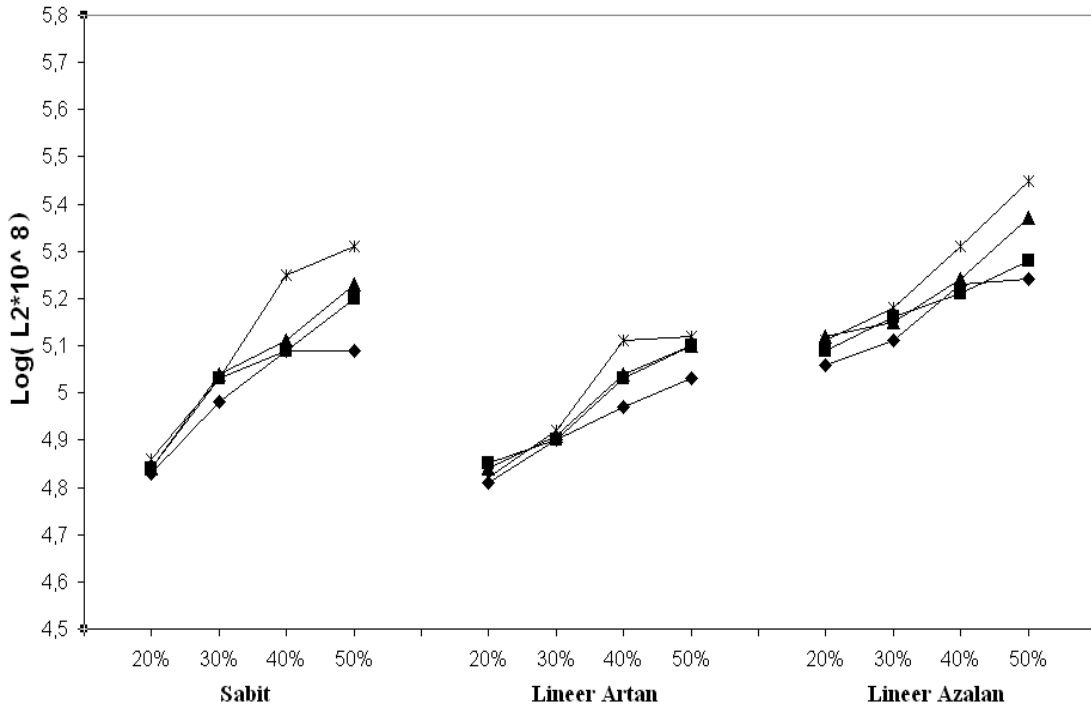
Tahminlerin performansının değerlendirilmesinde gerçek fonksiyonla tahminleri arasındaki farkların ortalama uzaklığı olan  $L_2$  hatası kullanılmıştır. Sonuçların literatürde var olan en detaylı simülasyon çalışması (22) ile karşılaştırılabilmesi için bulunan hata değerleri  $10^8$  le çarpılarak logaritması alınmıştır.

Sonuçlarımız sansürlü gözlemlerin yerleşiminin Kaplan Meier tahminleri üzerine bazı noktalarda istatistiksel olarak anlamlı düzeyde etki yaptığını göstermiştir. Kaplan Meier tahminlerine ait tüm sonuçlar örnek genişliğinin 100 ve 250 olduğu durumlarda Şekil 5.1 ve 5.2’de verilmiştir.



Şekil 5.1 N=100 K=1000 için Kaplan Meier tahminlerine ait tüm tahmin hataları karşılaştırma

Ssonuçlar klinik olarak anlamlılığı bakımından değerlendirilmiş hata miktarlarının çok küçük olmasından dolayı bazı noktalarda istatistiksel olarak anlamlı farklılıklar oluşmasına karşın rasgele olmayan sansürlü gözlemlerin yerleşimlerinin yaşam eğrisi tahminlerini klinik anlamlı farklılık yaratacak düzeyde etkilemediğine karar verilmiştir.



Şekil 5.2 N=250 K=1000 iken Kaplan Meier tahminlerine ait tüm tahmin hataları karşılaştırma

Hayatta kalma analizlerinde sansürlü gözlemlerin çalışmanın başında ortaya çıkması araştırmacılar tarafından büyük bir sorun olarak görülmekte ve tahminler için bilgi kaybının daha fazla olduğunu düşündürmektedir. Fakat çalışmanın başında izlemde olan birey sayısı fazla olduğundan sansürlü gözlemlerin hepsinin dahi çalışmanın başında oluşması tahminleri sansürlerin sonda oluşması kadar etkilememiştir.

Sansürlü gözlemlerin yaşam sürelerinin büyük bir kısmının en son gözlenen olaydan daha büyük olması literatürde uç (extreme) sağdan sansür olarak adlandırılmaktadır (44). Moeschberger ve arkadaşları tarafından 1985 yılında yapılan simülasyon çalışmasında (44) Kaplan Meier tahmin edicisinin başarısı var olan diğer

yöntemlerle karşılaştırmış ve uç sağdan sansürlü gözlemler oluşması durumunda tahmin edicide bir düzeltme yapılması önermiştir. Simülasyon çalışmamızda bu sorunla karşılaşmamak için son gözlemlerin sansürlü olması engellenmiştir.

Şekil 5.1 ve 5.2’de verilen hata miktarları karşılaştırıldığında örnek genişliğinin 250 olduğu durumda hata oranlarında ciddi düzeyde azalış olduğu gözlenmektedir. Örnek genişliği arttırıldığında son bölgede oluşan tahmin hatalarının da azalmış olmasından dolayı, III. Bölgede oluşan tahmin hatalarındaki kötüleşmenin nedeni olarak çalışmanın sonuna doğru izlemdeki birey sayısının azalması olarak değerlendirilmiştir. Çalışmanın sonuna doğru yüksek oranda sansürlü gözlem oluşmasının beklendiği bir klinik çalışmada örnek genişliğinin arttırılması bir çözüm olarak önerilebilir.

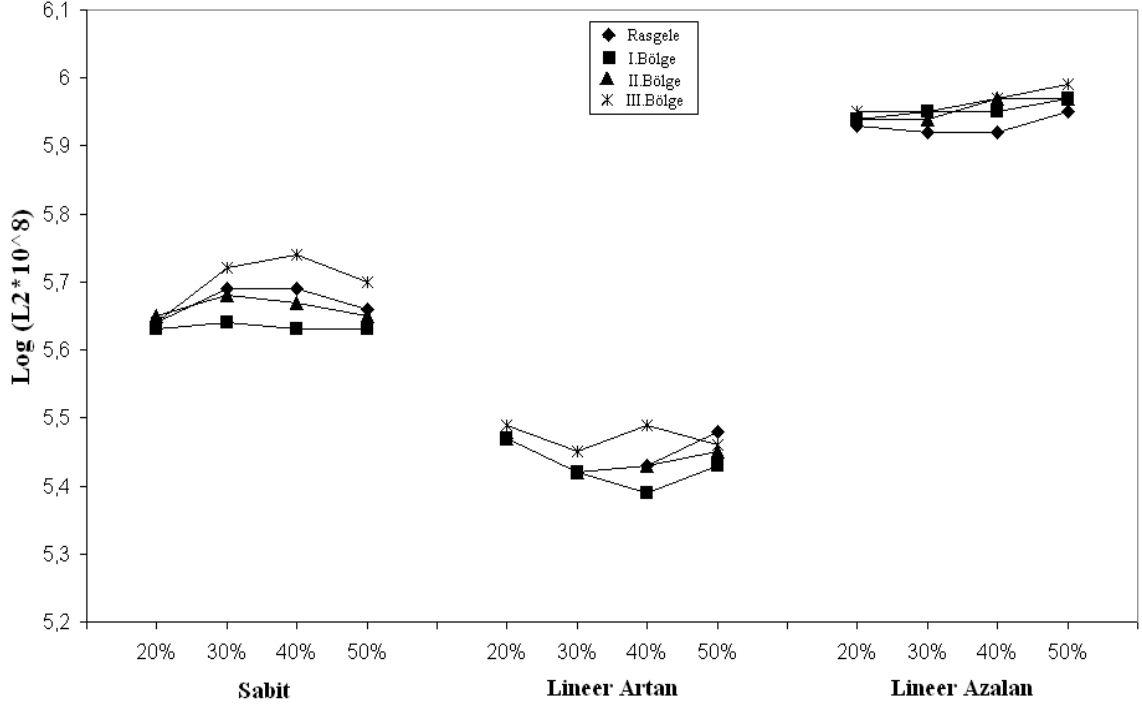
Simülasyonlarda kullanılan dağılımlar karşılaştırıldığında Kaplan Meier yöntemiyle hayatta kalma eğrileri tahmin edildiğinde tahminlerin başarısının üzerinde veri setinin dağılımının önemli bir etkisinin olmadığı da gözlenmiştir. Her iki örnek genişliği ve tüm sansür oranlarında en başarılı tahminler doğrusal artan hazard fonksiyonuna ait hayatta kalma fonksiyonlarının tahmininde ortaya çıkmıştır.

Tüm dağılımlarda sansürlü gözlem miktarındaki artışın tahminlerin başarısını olumsuz yönde etkilediği ortaya konmuştur. Bu durum özellikle örnek genişliği arttırıldığında daha net gözlenmektedir.

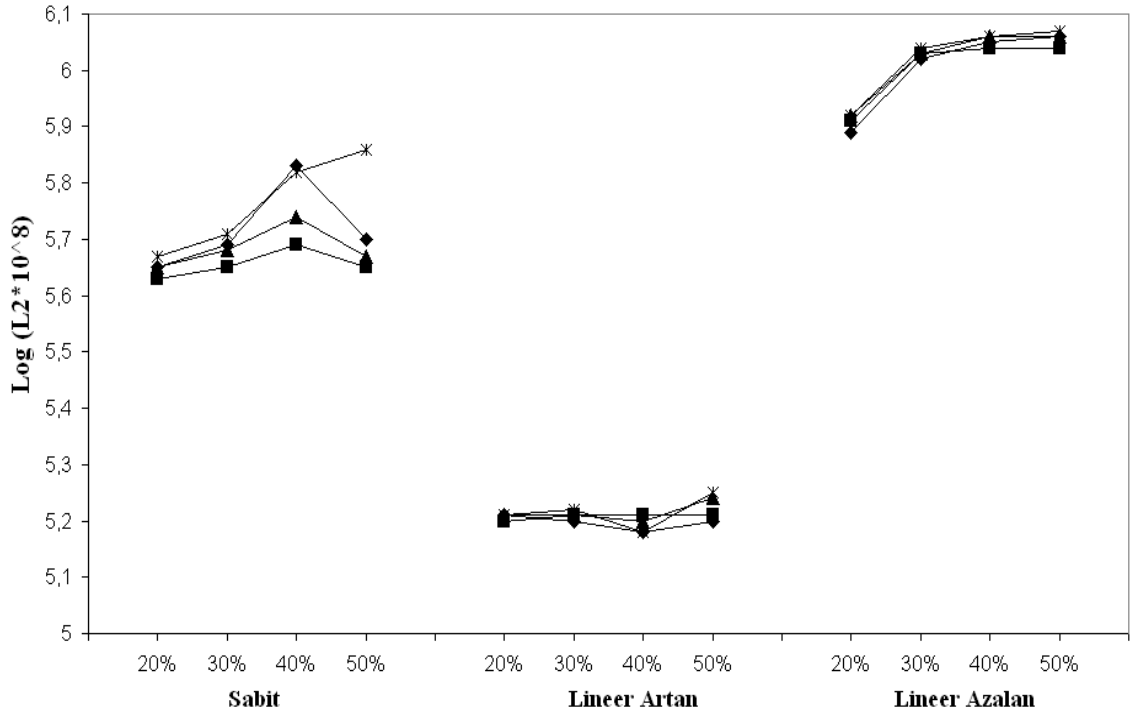
Hazard oranı tahminlerine ait tüm sonuçlar örnek genişliğinin 100 ve 250 olduğu durumlarda Şekil 5.3 ve 5.4’de verilmiştir.

Kernel düzgünleştirme yöntemine ait simülasyon sonuçlarımız da Kaplan Meier tahminleri sonuçlarına benzerdir ve sansürlü gözlemlerin yerleşiminin Hazard oranı tahminleri üzerine bazı noktalarda istatistiksel olarak anlamlı düzeyde etki yaptığını göstermiştir fakat bu rasgele olmayan sansürlü gözlemlerin yerleşimlerinin tahminler üzerine etkisinin klinik olarak anlamlık yaratmayacağına karar verilmiştir.

Hazard oranlarının fonksiyonlarına ait tahminlerde Kernel düzgünleştirme yöntemi kullanılmıştır. Literatürde birçok Kernel fonksiyonu yer almaktadır. Etkinliklerine göre sıralandığında etkinliği en iyi Kernel fonksiyonunun Epanechnikov Kernel olması ve hesaplanma kolaylığı nedeni ile bu çalışmada Epanechnikov Kernel fonksiyonu kullanılmıştır (45).



Şekil 5.3 N=100, K=1000 iken Hazard oranı tahminlerine ait tüm tahmin hataları karşılaştırma



Şekil 5.4 N=250, K=1000 iken Hazard oranı tahminlerine ait tüm tahmin hataları karşılaştırma



Kernel fonksiyonunun belirlenmesinin tahmin performansı üzerine büyük bir etkisi olmadığı asıl olarak bant genişliği belirleme yönteminin Kernel tahminlerinin performansını önemli düzeyde etkilediği bilinmektedir (13, 15, 17). Bu nedenle çalışmamızda R paket programının içinde yer alan birçok bant genişliği belirleme yöntemiyle elde edilen sonuçlar karşılaştırmıştır (45). Kullanılan bant genişliği belirleme yöntemleri şunlardır; Silverman tarafından önerilen pratik kural (46), Yanlı çapraz geçerlilik (biased cross validation) (47), yansız çapraz geçerlilik (unbiased cross validation) (49, 50), Plug in yaklaşımı (45). Ayrıca çalışma simülasyon çalışması olduğu ve verilerin türetildikleri fonksiyonlar bilindiği için gerçek fonksiyonla tahmin edilen fonksiyonlar arasındaki farkı minimum yapan bant genişliğini kullanacak şekilde bir algoritma yazılmıştır ve bant genişliklerinin performansları karşılaştırıldığında bu bant genişliği belirleme yönteminin kullanılmasına karar verilmiştir. Fakat simülasyon çalışmaları dışındaki gerçek veriler ile çalışılan durumda gerçek fonksiyonu bilmek mümkün olmadığı için bu yöntemin bant genişliği belirlenmesinde kullanılması mümkün değildir. Verinin gerçek dağılımı bilinmediğinde literatürden yararlanılarak en uygun bant genişliği belirleme yöntemine karar vermek gerekecektir.

Simülasyon çalışmamızın sonuçları literatürdeki en detaylı simülasyon çalışması ile karşılaştırıldığında (22) çalışmamızdaki Hazard oranı tahminlerinin performansının daha yüksek olduğu görülmüştür. Bunun nedeni her veri için ideal bir bant genişliği belirlemememizdir. Hata oranları oldukça düşük olduğundan sağdan ya da soldan düzeltme yapmaya ihtiyaç duyulmamıştır. Çalışmanın sonlarına doğru oluşan hataların yüksek olması hazard tahminlerinde izlemdeki gözlem sayılarının azalmasından etkilendiğini göstermektedir. Sansür miktarındaki artış hazard tahminlerini de kötüleştirmektedir. Ayrıca örnek genişliğini arttırmak hazard oranı tahminlerinin performansını etkilememiştir, bu bulguda Hess ve ark. bulgusu ile benzerdir. Her iki simülasyon çalışmasında da en kötü tahminler doğrusal azalan hazard fonksiyonun tahmininde ortaya çıkmıştır.

## 6. SONUÇ VE ÖNERİLER

Sonuç olarak, rasgele olmayan sansürlü gözlemlerin varlığının hayatta kalma fonksiyonu tahmininde kullanılan Kaplan Meier yöntemi ve Hazard oranı tahmininde kullanılan Kernel düzgünleştirme yöntemi üzerine yorumları klinik olarak etkileyecek düzeyde bir etkisinin olmadığı ortaya konmuştur. Bu sonuçta dikkat edilmesi gereken konu rasgele olmayan sansürlü gözlemlerin klinik ve hayvan deneylerinde birçok şekilde ortaya çıkabileceğidir. Yarışan bir riskin varlığı nedeni ile ya da bir prognostik faktörden dolayı ortaya çıkan sansürlerde rasgele olmayan sansürlerin özel bir halidir (51, 52). Yarışan risk olduğu durumda oluşacak rasgele olmayan sansürün kendisinin de bir bilgi içermesi söz konusudur, çünkü eşlik eden hastalığın yaşam süresi üzerine genel bir etkisi olacaktır (52, 53). Bu durumda yarışan risklere özel istatistiksel yöntemlerin kullanılması önerilmektedir (53). Bu tezde oluşturulan rasgele olmayan sansürlü gözlemlerin yaşam süresine etkisi olmadığı varsayılmıştır.

Sansürlü gözlem miktarları ele alındığında, yapılan simülasyonların hepsinde 0.30 oranında sansürün varlığında tahminlerdeki hata oranlarının daha makul düzeyde olduğu fakat bu oran arttığında hataların belirgin bir şekilde yükseldiği görülmüştür. Bu nedenle hayatta kalma analizi çalışmalarında 0.30 sansür oranının makul olacağı ve daha fazla sansürlü veri içeren verilerden yapılan tahminlerde tahmin başarılarının düşeceği sonucuna varılmıştır.

Yaşam sürelerinin ve hazard oranının tahmininde rasgele olmayan sansürlerin etkisinin olmaması, bu eğrilerin karşılaştırılması durumunda bir etki oluşmayacağını düşündürmemelidir. Yapılan bir ön simülasyon çalışmasında Logrank testi için rasgele sansür varsayımının ihlalinin I. tip hatayı arttırdığı gözlenmiştir (55). Çalışmamızdaki bulgulardan yola çıkarak öneri olarak rasgele olmayan sansürlü gözlemlerin logrank testi ve cox regresyon üzerine etkisinin araştırıldığı bir çalışma araştırmacılara önerilebilir

## KAYNAKLAR

- 1) **Kleinbaum D. G., Klein M.** Survival Analysis: A Self-Learning Text, Springer, New York, **2005**.
- 2) **Györfi L., Kohler M., Krzyzak A., Walk H.** “A Distribution-Free Theory of Nonparametric Regression”, Springer, New York, **2002**.
- 3) **Dawson B., Trap G.R.** Basic & Clinical Biostatistics, McGraw-Hill Professional, **2004**
- 4) **Vittinghoff E., Glidden D., Shiboski S.C., McCulloch C.E.** Regression Methods in Biostatistics: Linear, Logistic, Survival and repeated measures models, Springer, New York, **2004**.
- 5) **Smith T., and Smith B.**, Survival Analysis and the Application of Cox's Proportional Hazards Modeling Using SAS, Paper 244-26, Erişim adresi: [www2.sas.com/proceedings/sugi26/p244-26.pdf](http://www2.sas.com/proceedings/sugi26/p244-26.pdf), Erişim tarihi: 01/06/2009.
- 6) **Kaplan E.I, Meier P.**, Nonparametric Estimation From Incomplete Observations, *J. Am. Stat. Assoc.*, **1958**, 53, 457-481.
- 7) **Watson, G. S. and Leadbetter, M. R.** Hazard analysis I, *Biometrika*, **1964**, 51, 175-184.
- 8) **Watson, G.S. and Leadbetter, M.R.** Hazard analysis II. *Sankhya Ser. A*, **1964**, 26, 101-116.
- 9) **Nelson, W.** , Hazard plotting for incomplete failure data', *Journal of Quality Technology*, **1969**, 27-52 .
- 10) **Hansen H. R.** Smoothing counting process intensities by means of Kernel functions, *The Annals of statistics*, **1983**, vol:11 , no:2, 453-466.
- 11) **Wand, M. P. and Jones, M. C.** Kernel Smoothing. Chapman and Hall, London., **1995**.
- 12) **Kooperberg, C., Stone, C.J. and Truong, Y.K.** Hazard regression. *J. Am. Statist. Assoc.* **1995**; 90, 78-94
- 13) **Mugdadi A.R., Ahmad A.** A Bandwidth Selection for Kernel Density Estimation of Functions of Random Variables, *Computational Statistics & Data Analysis*, **2004**; Vol:47, issue:1, pp:49-62.
- 14) **Weissbachy R., Gefeller O.** A Rule-of-Thumb for the Variable Bandwidth Selection in Kernel Hazard Rate Estimation, 2-04.pdf. Erişim tarihi: 21.09.2008.
- 15) **Wang, J. L.** “Smoothing hazard rates”, in *Encyclopedia of Biostatistics*, Wiley, London, pp. 4140-4150, **1997**.
- 16) **Sheather S.J, Jones M. C.** A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation, *Journal of the Royal Statistical Society. Series B (Methodological)*, **1991**; Vol. 53, No. 3, pp. 683-690
- 17) **Turlach A.B.**, Bandwidth Selection in Kernel Density Estimation: Review, <http://www.stat.ucl.ac.be/ISpub/dp/1993/dp9317.ps>, Erişim Tarihi: 21.09.2008.
- 18) **Uzunoğulları Ü., Wang L. J.** A Comparison of Hazard Rate Estimators For Left Truncated And Right Censored Data, *Biometrika*, **1992**: 79(2):297-31

- 19) **Taner M.**, A Note on the Variable Kernel Estimator of the Hazard Function from Randomly Censored Data, *The Annals of Statistics*, **1983**, Vol. 11, No. 3, pp. 994-998
- 20) **Muller G.G. and Jane-Ling Wang J.L.** Hazard Rate Estimation under Random Censoring with Varying Kernels and Bandwidths, *Biometrics*, **1994**; Vol. 50, No. 1, pp. 61-76
- 21) **Patil, P.N.** Bandwidth choice for nonparametric hazard rate estimation. *J. Statist. Plann. Inf.*, **1993**. 35, 15-30.
- 22) **Hess K.R, Serachitopol D.M, Brown B.W.**, Hazard function estimators: a simulation study. *Statistics in Medicine*, **1999**; 18, pp. 3075–3088.
- 23) **Sheiner L. B., Beal S. L. ; Dunne A. ; Lin X., Little R. J; Gelman A.; Bois F. Y., Diggle P. J.; Smith D. M.**; “Analysis of Nonrandomly Censored Ordered Categorical Longitudinal Data From Analgesic Trials”, *Journal of the American Statistical Association*, **1997**, Vol: 92, No:440, pp:1235-1255.
- 24) **Ganju J., Lakatos E., Rothe E.**, A Simple Way To Estimate The Median Time And Compare Survival Distributions In Analgesic Trials Under Informative Censoring, *Journal of Biopharmaceutical Statistics*, **1999**, 9(4), 683–693.
- 25) **Kanık A., Kul S.**, Hayatta Kalma Analizleri, Mersin Üniversitesi, Tıp Fakültesi Dergisi, 4 395-401, **2003**.
- 26) **Peace K.E.**, Design and Analysis of Clinical Trials with Time-to-Event Endpoints, Chapman & Hall/CRC, **2009**.
- 27) **Tableman M., Kim J.S.**, Survival analysis using S: analysis of time-to-event data, Chapman & Hall/CRC, **2004**.
- 28) **Klein, J.P., Moeschberger, M.L.** Survival Analysis: Techniques for censored and truncated data, Springer, New York, **1997**.
- 29) **Dalgaard P.**, Introductory Statistics with R, Springer Verlag, NewYork, **2002**.
- 30) **Lee C.T.**, Introductory Biostatistics, John Wiley & Sons, Inc., **2003**.
- 31) **Sabine, L., Brian S. E.**, A handbook of statistical analyses using SPSS, Chapman & Hall/CRC Press LLC , **2004**.
- 32) <http://www.statsoft.com/textbook/stathome.html>, Erişim tarihi: 15/06/**2009**.
- 33) **Marubini E., Valsecchi M.G.** Analyzing Survival Data from Clinical Trials and Observational Studies, John Wiley & Sons, Inc., USA, **2004**.
- 34) **Hosmer W.D, Lemeshow S.** Applied Survival Analysis Regression Modeling of Time to Event Data, John Wiley & Sons, Inc., USA, **1998**.
- 35) **Lee E., Wenyuwang J.** Statistical Methods for Survival Data Analysis, John Wiley & Sons, Inc., USA, **2003**.
- 36) **Kerr K.**, University of Washington, Lecture 19 Regression Methods for Survival Data, Ders notları, Erişim tarihi: 13/09/**2005**.
- 37) **Lawless F.J.** Statistical Methods for Life Time Data, John Wiley & Sons, Inc., USA, **2003**.
- 38) **Selvin S.**, Survival Analysis for Epidemiologic and Medical Research A Practical Guide, Cambridge University Pres, **2008**.

- 39) **Cook R. J., Lawless J.F.**, The Statistical Analysis of Recurrent Events, Springer Science + Business Media, LLC, **2007**.
- 40) **Scott D. W. and Terrell G. R.**, Biased and unbiased cross-validation in density estimation. Journal of the American Statistical Association, 82:1131-1146, **1987**.
- 41) **Bowman A. W. and Azzalini A.** Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations, Oxford University Press, Oxford, **1997**.
- 42) **Wand M. P., Jones M. C.** Kernel Smoothing Monographs on Statistics and Applied Probability, Chapman & Hall, **1995**.
- 43) **R Development Core Team**, R. A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, **2005**. Erişim adresi: <http://www.R-project.org>.
- 44) **Moeschberger M. L. and Klein J. P.A**, Comparison of Several Methods of Estimating the Survival Function When There is Extreme Right Censoring Biometrics, Vol. 41, No. 1 pp. 253-259, **1985**
- 45) **Venables, W. N. and Ripley, B. D.** Modern Applied Statistics with S. Springer, 2002.
- 46) **Silverman, B. W.** Density Estimation. London: Chapman and Hall, 1986.
- 47) **Scott, D. W.** Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, **1992**.
- 48) **Terrell G.R. and Scott D.W.**, Variable Kernel density estimation, *Ann. Statist.* **20**, pp. 1236–1265, **1992**.
- 49) **Bowman A.W.**, An alternative method of cross-validation for the smoothing density estimates. Biometrika, 71:353-360, 1984.
- 50) **Rudemo M.**, Empirical choice of histograms and Kernel density estimators. Scandinavian Journal of Statistics, 9:65-78, 1982.
- 51) **Lagakos S.W**, General Right Censoring and Its Impact on the Analysis of Survival Data, Biometrics, Vol. 35, No. 1, 1979.
- 52) **Wu Y.X, Furnary A.P, GrunkeMeier G.L**, Using the National Death Index to Validate the Noninformative Censoring Assumption of Survival Estimation, *Ann Thorac Surg* 85:1256-1260 2008.
- 53) **Scharfstein D., Robins J.M, Eddings W., Rotnitzky A.**, Inference in Randomized Studies with Informative Censoring and Discrete Time-to-Event Endpoints, Biometrics 57, 404-413, **2001**.
- 54) **Armitage P., Colton T.**, Encyclopedia of Biostatistics, Wiley, **2005**.
- 55) **Kul S., Kanik A.** "Log-Rank Test In The Absence of Random Right Censoring: A Simulation Study“, 5<sup>th</sup> Conference of The Eastern Mediterranean Region of The International Biometric Society, 10-14 Mayıs, İstanbul, **2009**.

## EKLER

### EK 1. Sabit Hazard Kaplan Meier Yöntemi Tahmin Programı

```
KaplanMeier<-function(x=0,Z,delta)
{
n<-length(Z)

perm<-order(Z)
delta<-delta[perm]
Z<-Z[perm]

erg<-1
for (i in (1:n))
  {
  if (Z[i]<=x)
    {
    erg<-erg*((n-i)/(n-i+1))^delta[i]
    }
  }
erg
}

c1<-function(n=100)
{#random censoring
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
  c<-runif(n,min=0,max=196)
  for (i in 1:n)
  {
  y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)
  delta[y<=c]<-1

list(y=y,c=c,z=z,delta=delta)
}

c2<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
  if (runif(1)< 0.23)
    c[i]<-runif(1,min=0,max=10)
  else
    c[i]<-runif(1,min=250,max=600)
  }
}
```

```

    }

    for (i in 1:n)
    {
      y[i]<-(-log(u[i])/0.0256)
    }
    z<-pmin(y,c)
    delta[y<=c]<-1
list(y=y,c=c,z=z,delta=delta)
}

c3<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
    if (runif(1)< 0.29)
      c[i]<-runif(1,min=10,max=20)
    else
      c[i]<-runif(1,min=250,max=600)
  }

  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)

  delta[y<=c]<-1
list(y=y,c=c,z=z,delta=delta)
}

c4<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
    if (runif(1)< 0.38)
      c[i]<-runif(1,min=20,max=30)
    else
      c[i]<-runif(1,min=250,max=600)
  }

  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)
  delta[y<=c]<-1

```

```

list(y=y,c=c,z=z,delta=delta)
}

Testen1<-function(n=100,xs=xs)
{daten<-c1(n=n)
  {
    KM<-KaplanMeier(x=xs,Z=daten$z,delta=daten$delta)
  }
return(KM)
}

Testen2<-function(n=100,xs=xs)
{daten<-c2(n=n)
  {
    KM<-KaplanMeier(x=xs,Z=daten$z,delta=daten$delta)
  }
return(KM)
}

Testen3<-function(n=100,xs=xs)
{daten<-c3(n=n)
  {
    KM<-KaplanMeier(x=xs,Z=daten$z,delta=daten$delta)
  }
return(KM)
}

Testen4<-function(n=100,xs=xs)
{daten<-c4(n=n)
  {
    KM<-KaplanMeier(x=xs,Z=daten$z,delta=daten$delta)
  }
return(KM)
}

pl1<-function(k=100, n=100,xs=seq(from 0.01,to=160,by 0.1))
{
m<-rep(0,k)
l<-length(xs)
St<-rep(0,l)
d<-rep(0,l)
for(j in 1:l)
{print(paste("j=",j))

for (i in 1:k)
{
print(paste("i=",i))
m[i]<-Testen1(n=n,xs=xs[j])
#print(paste("xs=",xs))
}
d[j]<-mean(m)
}
}

```



```

for (i in 1:l)
{
St[i]<-exp(-0.0256*xs[i])
}
list(St=St, d=d, m=m)
plot(xs,St, col="green",type="l",lwd=3)
points(xs,d,col="red",type="l")
}

pl2<-function(k=100, n=100,xs=seq(from 0.01,to=160,by 0.1))
{
m<-rep(0,k)
l<-length(xs)
St<-rep(0,l)
d<-rep(0,l)
for(j in 1:l)
{print(paste("j=",j))

for (i in 1:k)
{
print(paste("i=",i))
m[i]<-Testen2(n=n,xs=x[j])
#print(paste("xs=",xs))
}
d[j]<-mean(m)
}
for (i in 1:l)
{
St[i]<-exp(-0.0256*xs[i])
}
list(St=St, d=d, m=m)
plot(xs,St, col="green",type="l",lwd=3)
points(xs,d,col="red",type="l")
}

pl3<-function(k=100, n=100,xs=seq(from 0.01,to=160,by 0.1))
{
m<-rep(0,k)
l<-length(xs)
St<-rep(0,l)
d<-rep(0,l)
for(j in 1:l)
{print(paste("j=",j))

for (i in 1:k)
{
print(paste("i=",i))
m[i]<-Testen3(n=n,xs=x[j])
#print(paste("xs=",xs))
}
d[j]<-mean(m)
}
for (i in 1:l)
{
St[i]<-exp(-0.0256*xs[i])
}
list(St=St, d=d, m=m)
}

```

```

plot(xs,St, col="green",type="l",lwd=3)
points(xs,d,col="red",type="l")
}

```

```

pl4<-function(k=100, n=100,xs=seq(from 0.01,to=160,by 0.1))
{
m<-rep(0,k)
l<-length(xs)
St<-rep(0,l)
d<-rep(0,l)
for(j in 1:l)
{print(paste("j=",j))

for (i in 1:k)
{
print(paste("i=",i))
m[i]<-Testen4(n=n,xs=xs[j])
#print(paste("xs=",xs))
}
d[j]<-mean(m)
}
for (i in 1:l)
{
St[i]<-exp(-0.0256*xs[i])
}
list(St=St, d=d, m=m)
plot(xs,St, col="green",type="l",lwd=3)
points(xs,d,col="red",type="l")
}

```

## EK 2. Sabit Hazard Kernel Düzgünleştirme Yöntemi Tahmin Programı

```
Epkern<-function(u)
{
  if ((u>=-1) & (u <=1))
    erg<-0.75*(1-u^2)
  else
    erg<-0
  erg
}

c1<-function(n=100)
{#random censoring
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
  c<-runif(n,min=0,max=196)
  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)
  delta[y<=c]<-1
n<-length(z)
perm<-order(z)
delta<-delta[perm]
z<-z[perm]
y<-y[perm]
list(y=y,c=c,z=z,delta=delta)
}

c2<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
    if (runif(1)< 0.23)
      c[i]<-runif(1,min=0,max=10)
    else
      c[i]<-runif(1,min=250,max=600)
  }

  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)
  delta[y<=c]<-1
n<-length(z)
perm<-order(z)
```

```

delta<-delta[perm]
z<-z[perm]
y<-y[perm]
list(y=y,c=c,z=z,delta=delta)
}

c3<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
    if (runif(1)< 0.29)
      c[i]<-runif(1,min=10,max=20)
    else
      c[i]<-runif(1,min=250,max=600)
  }

  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)

  delta[y<=c]<-1
n<-length(z)
perm<-order(z)
delta<-delta[perm]
z<-z[perm]
y<-y[perm]
list(y=y,c=c,z=z,delta=delta)
}

c4<-function(n=100)
{
  y<-rep(0,n)
  c<-rep(0,n)
  u<-rep(0,n)
  delta<-rep(0,n)
  u<-runif(n,min=0,max=1)
for (i in 1:n)
  {
    if (runif(1)< 0.38)
      c[i]<-runif(1,min=20,max=30)
    else
      c[i]<-runif(1,min=250,max=600)
  }

  for (i in 1:n)
  {
    y[i]<-(-log(u[i])/0.0256)
  }
  z<-pmin(y,c)
  delta[y<=c]<-1
n<-length(z)

```

```

perm<-order(z)
delta<-delta[perm]
z<-z[perm]
y<-y[perm]
list(y=y,c=c,z=z,delta=delta)
}

estimate<-function(b=b, n=n,x=x, a=data)
{
  z<-a$z
  delta<-a$delta
  l<-length(z)
  N<-length(x)
  erg<-rep(0,N)
  for (j in (1:N))
  {
    h<-rep(0,l)
    for (i in 1:l)
    {
      h[i]<-Epkern((x-z[i])/b)*delta[i]/(l-i+1)
    }
    sum<-sum(h)
    erg[j]<-sum/b
  }
  erg
}

Testen1<-function(n=100,x=x, b=b)
{daten<-c1(n=n)
  {
    E<-estimate(b=b, n=n, x=x, a=daten)
  }
  return(E)
}

Testen2<-function(n=100,x=x, b=b)
{daten<-c2(n=n)
  {
    E<-estimate(b=b, n=n, x=x, a=daten)
  }
  return(E)
}

Testen3<-function(n=100,x=x, b=b)
{daten<-c3(n=n)
  {
    E<-estimate(b=b, n=n, x=x, a=daten)
  }
  return(E)
}

Testen4<-function(n=100,x=x, b=b)
{daten<-c4(n=n)
  {
    E<-estimate(b=b, n=n, x=x, a=daten)
  }
  return(E)
}

```

```

p11<-function(k=100, n=100,x=seq(from=2,to=80,by=1), b=25)
{
l<-length(x)
f<-rep(0,l)
h<-rep(0.0256,l)
E<-matrix(rep(0,(l*k)),k,l)
for (i in 1:k)
for(j in 1:l)
{daten<-c1(n=n)
print(paste("i=",i))
print(paste("j=",j))
E[i,j]<-Testen1(n=n,x=x[j], b=25)
{
f[j]<-mean(E[,j])
}
}
}
plot(x,h, col="green",type="l",lwd=3)
points(x,f,col="red",type="l",lwd=2)
}

```

```

p12<-function(k=100, n=100,x=seq(from=2,to=80,by=1), b=25)
{
l<-length(x)
f<-rep(0,l)
h<-rep(0.0256,l)
E<-matrix(rep(0,(l*k)),k,l)
for (i in 1:k)
for(j in 1:l)
{daten<-c2(n=n)
print(paste("i=",i))
print(paste("j=",j))
E[i,j]<-Testen2(n=n,x=x[j], b=b)
{
f[j]<-mean(E[,j])
}
}
}
plot(x,h, col="green",type="l",lwd=3)
points(x,f,col="red",type="l",lwd=2)
}

```

```

p13<-function(k=100, n=100,x=seq(from=2,to=80,by=1), b=25)
{
l<-length(x)
f<-rep(0,l)
h<-rep(0.0256,l)
E<-matrix(rep(0,(l*k)),k,l)
for (i in 1:k)
for(j in 1:l)
{daten<-c1(n=n)
print(paste("i=",i))
print(paste("j=",j))
E[i,j]<-Testen3(n=n,x=x[j], b=25)
{
f[j]<-mean(E[,j])
}
}
}

```

```

}
}
plot(x,h, col="green",type="l",lwd=3)
points(x,f,col="red",type="l",lwd=2)
}

pl4<-function(k=100, n=100,x=seq(from=2,to=80,by=1), b=25)
{
l<-length(x)
f<-rep(0,l)
h<-rep(0.0256,l)
E<-matrix(rep(0,(l*k)),k,l)
for (i in 1:k)
for(j in 1:l)
{daten<-c4(n=n)
print(paste("i=",i))
print(paste("j=",j))
E[i,j]<-Testen4(n=n,x=x[j], b=b)
{
f[j]<-mean(E[,j])

}
}
plot(x,h, col="green",type="l",lwd=3)
points(x,f,col="red",type="l",lwd=2)
}

```