

KESİRLİ POLİNOMLAR İLE YAŞAM MODELLERİ

SURVIVAL MODELS WITH FRACTIONAL POLYNOMIALS

HAZAL DİNÇ

DOÇ. DR. NİHAL ATA TUTKUN
Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

2019

HAZAL DİNÇ'in hazırladığı “**Kesirli Polinomlar ile Yaşam Modelleri**” adlı bu çalışma aşağıdaki jüri tarafından **İSTATİSTİK ANABİLİM DALI**' nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Mehtap AKÇİL OK
Başkan



Doç. Dr. Nihal ATA TUTKUN
Danışman



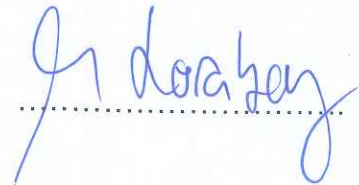
Prof. Dr. Duru KARASOY
Üye



Doç. Dr. Nursel KOYUNCU
Üye



Dr. Öğr. Üyesi Uğur KARABEY
Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak/...../..... tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

15.1.2019



HAZAL DİNÇ

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) veya elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin olarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ay ertelenmiştir.
- Tezim ile ilgili gizlilik kararı verilmiştir.

15.106.1.2019


Hazal DİNÇ

ÖZET

KESİRLİ POLİNOMLAR İLE YAŞAM MODELLERİ

Hazal DİNÇ

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Doç. Dr. Nihal ATA TUTKUN

Haziran 2019, 53 sayfa

Kesirli polinomlu regresyon modelleri, değişkenlerin doğrusal olmayan etkilerini modellemek için önerilmiştir. Sürekli değişkenlerin sürekli değişken olarak modele dahil edilmesi ve kategorik biçime dönüştürülerek modele dahil edilmesi sonuçları değiştirebilir. Özellikle sağlık alanındaki çalışmalarda sürekli değişkenler, örneğin yaş, çok kategorili değişkenlere dönüştürülerek kullanılmaktadır. Ancak değişkenleri kategorik biçime dönüştürmek bilgi kaybına neden olmaktadır. Değişken düzeylerinin nasıl belirleneceği konusunda da kararsızlıklar ortaya çıkmaktadır. Bu durumda da kesirli polinomlu modeller uygun bir alternatif olmaktadır.

Literatürde, kesirli polinomların klasik regresyon modelinde ve lojistik regresyon modelinde kullanımına ilişkin çalışmalar vardır. Cox regresyon modelinde de açıklayıcı değişkenler ile yaşam süresi arasındaki ilişkiyi doğru bir biçimde modellemek için kesirli polinomlu Cox regresyon modeli kullanılabilir. Uygun fonksiyon seçimi yönteminin uygulanması, doğrusal bir fonksiyonun ya da doğrusal olmayan bir kesirli polinomun uygun olup olmamasının kontrol edilmesini kolaylaştırmaktadır.

Tez çalışmasında, kesirli polinomların Cox regresyon modelindeki kullanımı incelenmiş, meme kanseri ve prostat kanseri veri kümeleri üzerinde uygulaması

yapılmıştır. Sürekli ve kategorik açıklayıcı değişkenli klasik Cox ve kesirli polinomlu Cox regresyon modelleri veri kümelerine uygulanmıştır. Meme kanseri veri kümesi için kesirli polinomlu Cox regresyon modelleri uygulanmış, ancak veri kümesi için yaş değişkeninin model üzerinde doğrusal bir etkisi olduğu ve klasik Cox regresyon modellerinin kullanılmasının daha uygun olduğu sonucuna ulaşılmıştır. Prostat kanseri veri kümesinde klasik Cox regresyon modeli için orantılı tehlikeler varsayımı sağlanmadığından bu sonuçların kullanılması ve yorumlanması uygun değildir. Bu veri kümesi için kesirli polinomlu Cox regresyon modeli incelendiğinde ise yaş değişkeninin bağımlı değişken ile doğrusal olmayan bir ilişkiye sahip olduğu belirlenmiş ve orantılı tehlikeler varsayımının sağlandığı görülmüştür. Buna göre, kesirli polinomlu Cox regresyon modelinin prostat kanseri veri kümesi için daha uygun olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Cox regresyon modeli, doğrusal olmama, kesirli polinomlar, orantılı tehlikeler.

ABSTRACT

SURVIVAL MODELS WITH FRACTIONAL POLYNOMIALS

Hazal DİNÇ

Master of Science, Department of Statistics

Supervisor: Assoc. Prof. Dr. Nihal ATA TUTKUN

June 2019, 53 pages

The use of regression models with fractional polynomial was proposed to model nonlinear effects of covariates. The inclusion of continuous covariates as continuous in the model and its inclusion in the model as a categorical covariate can change the results. Especially in health studies, continuous variables, for example age, are transformed into multi categorical variables. However, converting variables into categorical format causes a loss of information. There is also uncertainty about how to determine the levels of categorical variable. In this case, fractional polynomial models are a suitable alternative.

There are studies on the use of fractional polynomials in classical regression model and logistic regression model. In Cox regression model, fractional polynomials can be used to accurately model the relationship between the covariates and survival time. Applying the appropriate function selection method facilitates the control of whether a linear function or a non-linear fractional polynomial is appropriate.

In this thesis, the use of fractional polynomials in Cox regression model was investigated and applied on breast cancer and prostate cancer data sets. Cox regression model with continuous and categorical covariates and Cox regression

model with fractional polynomial were applied to the data set. Cox regression models with fractional polynomials were applied for the breast cancer data set, however it was concluded that the age variable had a linear effect on the model and it was more appropriate to use classical Cox regression models. Proportional hazards assumption was not provided for classical Cox regression model in the prostate cancer dataset, so these results are not appropriate to use and interpret. When the Cox regression models with fractional polynomials were examined for this data set, it was found that the age variable had a nonlinear relationship with the dependent variable and it was observed that the proportional hazards assumption was provided. Accordingly, it was concluded that the Cox regression model with fractional polynomials is more suitable for the prostate cancer dataset.

Keywords: Cox regression model, nonlinearity, fractional polynomials, proportional hazards.

TEŐEKKÜR

Yüksek lisans eğitimim boyunca bana yol gösteren, beni teşvik eden ve tezimin oluşmasında değerli bilgi ve yardımlarıyla en büyük desteęi veren tez danışmanım Sayın Doç. Dr. Nihal ATA TUTKUN'a , hayatımın her aşamasında maddi manevi her konuda yanımda olan, hiçbir desteęi ve yardımı esirgemeyen bana her zaman güvendikleri için annem Gülser Dinç ve babam Bülent Dinç'e, değerli önerileri ile tezime katkıda bulunan Sayın Prof. Dr. Durdu Karasoy, Sayın Prof. Dr. Mehtap Akçil Ok, Sayın Doç. Dr. Nursel Koyuncu ve Sayın Yrd. Doç. Dr. Uğur Karabey'e, tüm samimiyetimle teşekkürlerimi sunarım.

Hazal DİNÇ

Haziran 2019, Ankara

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	i
ABSTRACT	iii
TEŞEKKÜR.....	v
İÇİNDEKİLER.....	vi
ÇİZELGELER	viii
ŞEKİLLER.....	x
KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. COX REGRESYON MODELİ	4
2.1. En Çok Olabilirlik Tahmini.....	8
2.2. Yaşam ve Tehlike Fonksiyonlarının Tahmini	11
2.3. İstatiksel Testler ve Güven Aralıkları	13
2.4. Orantılı Tehlikeler Varsayımı	14
2.5. Model Seçim Yöntemleri.....	14
3. KESİRLİ POLİNOMLU YAŞAM MODELLERİ.....	16
3.1. Giriş	16
3.2. Kesirli Polinomlu Cox Regresyon Modeli	17
3.3. Kesirli Polinomlar	17
3.3.1. Model Seçiminde Kullanılacak Yöntemler.....	19
3.3.1.1.Kapalı Test Yöntemi	19
3.3.1.2. Ardışık Yöntem	19
3.3.2. Ölçekleme ve Odaklama.....	20
3.3.3. Kesirli Polinomlar ile Çok Değişkenli Model Oluşturma	21

4. UYGULAMA	23
4.1. Meme Kanseri Veri Kümesi.....	23
4.1.1. Meme Kanseri Veri Kümesi için Parametrik Olmayan Yaşam Çözümlemesi Sonuçları.....	25
4.1.2. Meme Kanseri Veri Kümesi için Klasik CRM Sonuçları.....	28
4.1.3. Meme Kanseri Veri Kümesi için Kesirli Polinomlu Cox Regresyon Modeli Sonuçları.....	33
4.2. Prostat Kanseri Veri Kümesi.....	37
4.2.1. Prostat Kanseri Veri Kümesi için Parametrik Olmayan Yaşam Çözümlemesi Sonuçları.....	38
4.2.2. Prostat Kanseri Veri Kümesi için Klasik Cox Regresyon Modeli Sonuçları.....	40
4.2.3. Prostat Kanseri Veri Kümesi için Kesirli Polinomlu Cox Regresyon Modeli Sonuçları.....	44
5. SONUÇ VE ÖNERİLER.....	48
KAYNAKLAR.....	51

ÇİZELGELER

	<u>Sayfa</u>
Çizelge 4.1. Meme kanseri veri kümesi için kullanılan değişkenler ve düzeyleri	24
Çizelge 4.2. Meme kanseri verisi için log-rank testi sonuçları.....	28
Çizelge 4.3. Meme kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan çok değişkenli klasik CRM sonuçları	29
Çizelge 4.4. Meme kanseri verisi için Schoenfeld artıkları yöntemi sonuçları.....	30
Çizelge 4.5. Meme kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları.....	31
Çizelge 4.6. Meme kanseri veri kümesi için yaş değişkeni kategorik kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları.....	31
Çizelge 4.7. Meme kanseri veri kümesi için modele tüm değişkenler dahil edildiğinde yaş değişkeni kategorik kabul edilerek oluşturulan klasik CRM sonuçları.....	32
Çizelge 4.8. Meme kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları	33
Çizelge 4.9. Meme kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM sonuçları.....	34
Çizelge 4.10. Meme kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları	35
Çizelge 4.11. Meme kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları.....	35
Çizelge 4.12. Meme kanseri veri kümesi için kesirli polinomlu çok değişkenli CRM'ye ait Schoenfeld artıkları yöntemi sonuçları.....	36
Çizelge 4.13. Prostat kanseri veri kümesi için kullanılan değişkenler ve düzeyleri ..	37
Çizelge 4.14. Prostat kanseri verisi için log-rank testi sonuçları	39
Çizelge 4.15. Prostat kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan çok değişkenli klasik CRM sonuçları	40

Çizelge 4.16.	Prostat kanseri veri kümesi için Schoenfeld artıkları yöntemi sonuçları	41
Çizelge 4.17.	Prostat kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları	42
Çizelge 4.18.	Prostat kanseri veri kümesi için yaş değişkeni kategorik kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları	43
Çizelge 4.19.	Modele tüm değişkenler dahil edildiğinde yaş değişkeni kategorik kabul edilerek oluşturulan klasik CRM sonuçları.....	43
Çizelge 4.20.	Prostat kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları	44
Çizelge 4.21.	Prostat kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM sonuçları.....	45
Çizelge 4.22.	Prostat kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları	45
Çizelge 4.23.	Prostat kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları.....	46
Çizelge 4.24.	Prostat kanseri veri kümesi için kesirli polinomlu çok değişkenli CRM'ye ait Schoenfeld artıkları yöntemi sonuçları.....	46

ŞEKİLLER

	<u>Sayfa</u>
Şekil 4.1. Meme kanseri verisi için KM yaşam eğrisi	25
Şekil 4.2. Meme kanseri verisinde kullanılan açıklayıcı değişkenlerin KM eğrileri.....	26-27
Şekil 4.3. Prostat kanseri verisi için KM yaşam eğrisi	38
Şekil 4.4. Prostat kanseri verisi için açıklayıcı değişkenlere ait KM eğrileri	38-39
Şekil 4.5. Değişkenlerin log-log grafikleri	41

KISALTMALAR

AIC	Akaike Bilgi Kriteri
BIC	Bayesci Bilgi Kriteri
CRM	Cox Regresyon Modeli
EÇO	En Çok Olabilirlik
TO	Tehlike Oranı
OT	Orantılı Tehlikeler

1. GİRİŞ

'Yaşam çözümlemesi' terimi, ilk olarak ilgilenilen olayın ölüm olduğu çalışmalarda ortaya çıkmıştır. Günümüzde ise yaşam çözümlemesinin kullanım alanı çok genişlemiştir. Yaşam çözümlemesi, bir olay gerçekleşene kadar geçen ilgilenilen sonuç değişkeninin zamana bağlı olduğu verilerin analizi için kullanılan istatistiksel yöntemlerdir. Burada zaman, yıl, aylar, hafta olabilir. İlgilenilen sonuç değişkeni bir bireyin takibinin başlangıcından olay gerçekleşene kadar geçen süredir. Alternatif olarak olay meydana geldiğinde bireyin yaşı da süre olarak alınabilir.

Yaşam çözümlemesinde, zaman değişkeninden yaşam süresi olarak bahsedilir. Çünkü zaman değişkeni, bir bireyin takip süresi boyunca hayatta kaldığı zamanı gösterir. Ayrıca olay kavramından başarısızlık olarak bahsedilir. Çünkü ilgilenilen olay çoğunlukla ölüm, hastalık gibi negatif durumlardır. Ancak yaşam süresi, cerrahi bir müdahaleden sonra hastanın iyileşme süresi olarak da ele alınabilir. Bu durumda başarısızlık pozitif bir olaydır. Yaşam süresine örnek olarak hastalık başlangıcına kadar geçen süre, borsa çöküşüne kadar geçen süre, ekipman arızasına kadar geçen süre, depreme kadar geçen süre vb. verilebilir (Singh ve Mukhopadyay,2011).

Yaşam çözümlemesi yöntemleri, önemli model parametrelerinin tahmininde durdurulmuş ve durdurulmamış gözlemlerden gelen bilgileri doğru bir şekilde birleştirir. Klinik çalışmaların çoğunda durdurma probleminin ele alınması gerekmektedir. Durdurma, bireyin yaşam süresi hakkında bilginin olduğu ancak yaşam süresi tam olarak bilinmediğinde meydana gelmektedir. Belirli bir birim için olay gerçekleşmeden çalışma sona eriyor ise bu durumda birimin yaşam süresi durdurulmuştur ve birim durdurulmuş gözlem olarak adlandırılır.

Yaşam çözümlemesinde, önemli iki değişken bulunmaktadır: Birincisi ilgilenilen olayın zamanını gösteren süre değişkeni ve diğeri ise ilgilenilen olayın gerçekleşip gerçekleşmediğini gösteren durdurma değişkenidir. Yaşam ve tehlike fonksiyonları, yaşam sürelerinin dağılımını tanımlamak için yaşam çözümlemesinde anahtar kavramlardır.

Yaşam fonksiyonu belirtilen zamana kadar olayın gerçekleşmeme olasılığını verir. Tehlike fonksiyonu, belirtilen süreye kadar birimin hayatta kalması durumunda olayın gerçekleşmesi riskini verir. Bir açıklayıcı değişkenin olay gerçekleşene kadar geçen süre ile ilişkisini tanımlamak için genellikle yaşam çözümlemesi kullanılmaktadır.

Açıklayıcı değişkenlerin yaşam süresi ile ilişkisini çözümlmek için birçok yaşam modeli mevcuttur. Yöntemler yarı-parametrik, parametrik olmayan ve parametrik yaklaşımları içermektedir. Yaşam fonksiyonunun parametrik olmayan bir tahmin edicisi olarak Kaplan-Meier (KM) yöntemi, yaşam olasılıklarını zamanın bir fonksiyonu olarak tahmin etmek ve grafiklemek için kullanılır. Yarı parametrik bir yöntem olarak Cox regresyon modeli (CRM) kullanılmaktadır ve bu model yaşam çözümlemesinde en yaygın kullanılan model olma özelliğine sahiptir. Parametrik yöntemler yaşam süresinin dağılımının bilindiği durumlarda kullanılır. Üstel, Weibull, log-normal dağılım parametrik dağılımlara örnek olarak verilebilir. Model parametreleri maksimum olabilirlik fonksiyonu ile tahmin edilir. Modele dahil edilen açıklayıcı değişkenlerin etkilerinin doğrusal olmaması durumunda kesirli polinomların kullanılması önerilmektedir (Despa, 2019).

Cox regresyon modeli, çoğunlukla sağlık araştırmalarında açıklayıcı değişkenlerin tehlike fonksiyonu üzerindeki etkisini araştırmak için kullanılır. Bu modelin ilk avantajı, tehlike fonksiyonunun fonksiyonel şekli hakkında veya olayın gerçekleşme zamanının dağılımı hakkında özel varsayımlarda bulunma koşulunun olmamasıdır. İkinci avantajı ise, zamana göre değişen değişken etkilerini ve zamana bağlı (veya zamana göre değişen) değişkenleri modele dahil etme özelliğidir.

İki ya da daha çok düzeyli kategorik değişkenler sağlık araştırmalarında sıklıkla yer almaktadır. Kategorik değişkenlerin CRM'ye dahil edilmesi daha basitken, sürekli değişkenlerin modele dahil edilmesi daha karmaşıktır. Sürekli değişkenler ile tehlike fonksiyonu arasındaki doğrusal olmayan bir ilişkinin nasıl modelleneceğine karar verilmelidir (Austin ve ark.,2014).

Kesirli polinomlar, klasik regresyon, lojistik regresyon ve Cox regresyon modelinde kullanılabilir. Klasik regresyon modelinde kesirli polinoma ilişkin çalışmalar ilk kez Royston ve Altman (1994) tarafından incelenmiştir. Ambler ve diğ. (1999), Ambler ve Royston (2000), Royston ve Sauerbrei (2002), Royston ve Sauerbrei (2004), Zapien ve diğ. (2007), Reitz ve diğ. (2006), Royston and Sauerbrei (2008),

Dupont (2010), Jansen (2011), Marinda (2011), Binder ve diğ. (2012) alıřmaları ile geliřtirilmiřtir. Lojistik regresyon modelinde kesirli polinomlar Royston ve Altman (1994), Karunanayake ve diğ. (2012), Austin ve diğ. (2014) , Ata Tutkun (2015), Elhan ve diğ. (2016), Ganji ve diğ. (2017) alıřmalarında ele alınmıřtır.

CRM'de kesirli polinomlar ise ilk kez Royston ve Altman (1994) tarafından incelenmiřtir. Sauerbrei ve diğ. (1999), Berger ve diğ. (2000), Look ve diğ. (2007), Berger ve diğ. (2003), Bellera ve diğ. (2010), Bucholz ve Sauerberi (2011) Zhang ve diğ. (2014), Bucholz ve diğ. (2014) alıřmaları ile geliřtirilmiřtir.

Bu alıřmanın amacı, kesirli polinomlu Cox regresyon modelini incelemek ve gerek veriler kullanarak uygulamasını yapmaktır. alıřmanın ikinci blmnde klasik CRM hakkında temel bilgiler verilmiřtir. nc blmnde ise kesirli polinomlar verilmiř ve CRM'de kesirli polinomlar incelenmiřtir. Drdnc blmnde meme kanseri ve prostat kanseri verilerine klasik CRM ve kesirli polinomlu CRM uygulanarak elde edilen sonular yorumlanmıřtır.

2. COX REGRESYON MODELİ

Cox (1972) tarafından önerilen Cox regresyon modeli, yaşam süresi verilerini incelemek için sağlık araştırmalarında çok kullanılan bir modeldir. Tehlike fonksiyonu ve açıklayıcı değişkenlerle ifade edilen bir regresyon modelidir. Matematiksel olarak CRM,

$$h(t,x)=h_0(t)\exp\{b_1x_1+ b_2x_2+ \dots + b_px_p \} \quad (2.1)$$

biçimindedir. Tehlike fonksiyonu $h(t)$, katsayıların b_1, b_2, \dots, b_p büyüklüğü ile ölçülebilen bir dizi x_1, x_2, \dots, x_p değişkenlerine bağlıdır.

h_0 terimi, temel tehlike olarak bilinir ve tüm x_i ' ler sıfıra eşit olduğundaki tehlike fonksiyonudur. $h(t)$ deki "t", tehlikenin zaman içinde değişebileceğini göstermektedir. CRM'nin önemli bir özelliği, temel tehlike fonksiyonunun parametrik olmayan olarak tahmin edilmesidir ve bu nedenle diğer istatistiksel modellerin aksine, yaşam sürelerinin belirli bir istatistiksel dağılıma sahip olduğu varsayılmaz. Ancak model, orantılı tehlikeler (OT) varsayımına sahiptir. Bu varsayım, değişken düzeyleri için tehlike fonksiyonlarının birbirine orantılı olması anlamına gelmektedir. $\exp(b_i)$ 'nin değeri tehlike oranı olarak bilinir. b_i 'nin değeri sıfırdan büyükse ya da birden büyük bir değere eşit ise, i . değişkenin etkisi ve tehlike oranı artar ve bu yüzden yaşam süresinin uzunluğu azalır. Başka bir ifadeyle birin üzerinde olan tehlike oranı olayın gerçekleşme olasılığı ile pozitif ilişkilidir ve yaşam süresinin uzunluğuyla negatif ilişkilidir (Bradburn ve ark., 2003).

CRM, bir ya da daha fazla değişken kümesinin ve tehlike oranının ilişkisini modellemektedir. Değişkenler kesikli veya sürekli olabilir. Çeşitli faktörlerin yaşam süresi üzerindeki etkisini incelemek için kullanılabilir. Örneğin bir bireye kesin hastalık teşhisi konduktan sonra beslenme tarzının, yaşın, egzersiz miktarının ve uyku miktarının yaşam süresi ile ilgisi araştırılabilir. Normal koşullar altında, kullanılan yaşam süresi ve çeşitli açıklayıcı değişkenler arasındaki ilişkiyi incelemek için kullanılan istatistiksel yöntem çoklu regresyon çözümlemesi olacaktır. Ancak, yaşam verilerinin özel yapısı nedeniyle, çoklu regresyon modeli uygun değildir.

Yaşam verileri genellikle durdurulmuş veriler içerir ve yaşam sürelerinin dağılımı genellikle çarpıktır. Bu iki problem çoklu regresyon modelinin kullanımını geçersiz kıldığından, alternatif regresyon modelleri önerilmiştir. Bunlardan en popüler olanı CRM'dir. Cox (1972)'den sonra Collett (1994)'in çalışmaları da bu yaşam çözümlemesine kapsamlı bir katkı sağlamıştır. Klein and Moeschberger (1997), Hosmer ve Lemeshow (1999), Therneau ve Grambsch (2000) da yaşam çözümlemesi ve CRM üzerine önemli çalışmalar yapmıştır.

Yaşam çözümlemesi, ilgilenilen olay gerçekleşene kadar geçen sürenin çözümlemesini ifade eder. Yanıt değişkeni, zamanın başlangıcıyla bitişi arasındaki süredir. Süre, ölüm ya da başarısızlık olarak adlandırılan ilgilenilen olayın ortaya çıkması ya da birimin çalışmadan çıkmasıyla sonlanmaktadır. Bu geçen süre standart istatistiksel teknikleri (t testi, varyans analizi, çoklu regresyon modeli vb.) geçersiz kılacak iki tane özelliğe sahiptir. Bunlardan birincisi, sürenin dağılımı genellikle çarpıktır. Standart istatistiksel teknikler, verilerin normal olarak dağılmasını gerektirir. İkinci özellik ise yaşam verileri durdurulmuş veriler içermektedir.

Yaşam verilerini çözümlerken, iki ana fonksiyon vardır. Bu fonksiyonlar, yaşam fonksiyonu ve tehlike fonksiyonudur. Yaşam süresi olan T, zamanın başlangıcından itibaren geçen süredir. T'nin değerleri olasılık dağılımı olarak düşünülebilir. T rastlantı değişkeninin olasılık yoğunluk fonksiyonunun f(t) ile verildiğini varsayalım. T'nin dağılım fonksiyonu,

$$F(t) = P(T < t) \quad (2.2)$$

biçiminde tanımlanabilir. Yaşam fonksiyonu ise aşağıdaki gibidir:

$$S(t) = P(T > t) = 1 - F(t). \quad (2.3)$$

Tehlike fonksiyonu, birim zaman aralığının [t ile t+Δt] başlangıcından itibaren hayatta kaldığında, belirtilen zaman aralığı içerisinde başarısız olma riskinin bir tanımıdır. Tehlike fonksiyonunun matematiksel gösterimi ise,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < (t + \Delta t) | t \leq T)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (2.4)$$

biçimindedir.

Birikimli tehlike fonksiyonu $H(t)$ ise, sıfırdan t zamanına kadar birim tehlike oranlarının toplamıdır. Birikimli tehlike fonksiyonunun formülü ise aşağıdaki gibidir:

$$H(t) = \int_0^t h(u)du. \quad (2.5)$$

Birikimli tehlike fonksiyonunun birikimli yaşam fonksiyonu ile arasındaki ilişki ise;

$$S(t) = e^{-H(t)} \quad (2.6)$$

veya

$$H(t) = -\ln(S(t)) \quad (2.7)$$

biçiminde ifade edilir.

Dağılım fonksiyonunun, tehlike fonksiyonunun ve yaşam fonksiyonunun matematiksel olarak ilişkili olduğu görülmektedir. Kolaylık ve pratiklik açısından, temel regresyon modelinde tehlike fonksiyonu kullanılır. Cox regresyon modeli kullanılarak tehlike oranı ve açıklayıcı değişkenler arasındaki ilişki;

$$\ln[h(t)] = \ln[h_0(t)] + \sum_{i=1}^p x_i \beta_i \quad (2.8)$$

veya

$$h(t) = h_0(t) e^{\sum_{i=1}^p x_i \beta_i} \quad (2.9)$$

olarak ifade edilmiştir.

Burada $x_1, x_2, x_3, \dots, x_p$ açıklayıcı değişkenlerdir. $\beta_1, \beta_2, \dots, \beta_p$ modelin katsayılarıdır. T yaşam süresidir. $h_0(T)$ ise temel tehlike oranıdır. Böylece regresyonun doğrusal biçimi;

$$\ln \left[\frac{h(t)}{h_0(t)} \right] = \sum_{i=1}^p x_i \beta_i \quad (2.10)$$

olarak ifade edilir. Eşitlik 2.10'nun her iki tarafının üsteli alındığında, gerçek tehlike oranı ile temel tehlike arasındaki oran elde edilir ve bu oran orantılı risk olarak bilinmektedir. Model tekrar düzenlendiğinde,

$$\frac{h(t)}{h_0(t)} = \exp\left(\sum_{i=1}^p x_i \beta_i\right) = e^{x_1 \beta_1} e^{x_2 \beta_2} \dots e^{x_p \beta_p} \quad (2.11)$$

biçiminde elde edilir. Değişken katsayıları bir birim arttırıldığında oluşan orantılı risk olarak yorumlanabilir. Eşitliğin sağ tarafı t'yi içermemektedir. Bir başka ifadeyle, orantılı risk tüm zaman noktaları için sabittir. Bu nedenle modele orantılı tehlikeler modeli de denmektedir.

CRM altında, birikimli tehlike fonksiyonu,

$$\begin{aligned} H(t, X) &= \int_0^t h(u, X) du = \int_0^t h_0(u) e^{\sum_{i=1}^p x_i \beta_i} du = e^{\sum_{i=1}^p x_i \beta_i} \int_0^t h_0(u) du \\ &= H_0(t) e^{\sum_{i=1}^p x_i \beta_i} \end{aligned} \quad (2.12)$$

biçiminde ifade edilir. Yaşam süresi, $e^{\sum_{i=1}^p x_i \beta_i}$ de değil, $H_0(t)$ ' de mevcuttur. t zamanına

kadar birikimli tehlike, $e^{\sum_{i=1}^p x_i \beta_i}$ değişkenleriyle $H_0(t)$ 'nin çarpılmasıyla elde edilmektedir.

Cox regresyon modeli altında, birikimli yaşam fonksiyonu,

$$S(t, X) = \exp(-H(t, X)) = \exp(-H_0(t) e^{\sum_{i=1}^p x_i \beta_i}) = \left[e^{-H_0(t)} \right] \left[e^{\sum_{i=1}^p x_i \beta_i} \right]$$

$$= S_0(t) e^{\sum_{i=1}^p x_i \beta_i} \quad (2.13)$$

biçimindedir.

Yaşam süresi t , $e^{\sum_{i=1}^p x_i \beta_i}$ ifadesinde değil, $S_0(t)$ fonksiyonunda yer almaktadır.

2.1. En Çok Olabilirlik Tahmini

t_1, t_2, \dots, t_m başarısızlık süreleri olsun. Bu süreler, tekrarlanan süreleri ve durdurulmuş gözlemleri içermemektedir. t_i ($i=1, \dots, M$) zamanında meydana gelen tüm başarısızlıklar D_i olarak ifade edilsin. T_i zamanından hemen önce risk altında olan tüm birimlerin kümesi R_i olarak ifade edilsin. Bu küme t_i zamanından sonra başarısız olan tüm birimleri ve t_i ' den daha sonraki bir zamanda durdurulmuş veya başarısız olanları içerir. R_i kümesinin üyeleri $r=1, \dots, n_i$ şeklinde olsun. X , açıklayıcı değişkenleri gösterecek. Bu değişkenler i, j, k tarafından indekslenir. T_d başarısızlık süresinde değişkenlerin değerleri, $x_{1d}, x_{2d}, \dots, x_{id}$ şeklinde yazılır. Regresyon katsayıları ise; $\beta_1, \beta_2, \dots, \beta_p$ olarak tahmin edilir.

Eş zamanlı gözlem olmaması durumunda, log-olabilirlik fonksiyonu Kalbfleisch ve Prentice (1980) tarafından aşağıdaki gibi tanımlanmıştır:

$$\log(\beta) = \sum_{t=1}^M \sum_{i=1}^p (x_{it} \beta_i) - \ln \sum_{r \in R_t} \exp \sum_{i=1}^p x_{ir} \beta_i = \sum_{t=1}^M \left\{ \sum_{i=1}^p x_{it} \beta_i - \ln(G_{R_t}) \right\}. \quad (2.14)$$

Burada G_R ,

$$G_R = \sum_{r \in R} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \quad (2.15)$$

biçiminde ifade edilir. Birinci ve ikinci dereceden kısmi türevler, bu alt bölümdeki türevleri elde ederken yararlı olmaktadır ve

$$H_{jR} = \frac{\partial G_R}{\partial \beta_j} = \sum_{r \in R} x_{jr} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \quad (2.16)$$

$$A_{jkr} = \frac{\partial^2 G_R}{\partial \beta_j \partial \beta_k} = \frac{\partial H_{jR}}{\partial \beta_k} = \sum_{r \in R} x_{jr} x_{kr} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \quad (2.17)$$

biçiminde ifade edilir. En çok olabilirlik (EÇÖ) çözümü, Newton-Raphson yöntemi ile yapılmaktadır. Bu yöntem, birinci ve ikinci dereceden kısmi türevlere ihtiyaç duymaktadır. Birinci dereceden kısmi türev;

$$U_j = \frac{\partial \log(\beta)}{\partial \beta_j} = \sum_{i=1}^M \left\{ x_{jt} - \frac{H_{jR_t}}{G_{R_t}} \right\} \quad (2.18)$$

biçimindedir. Bilgi matrisini oluşturan ikinci dereceden kısmi türev ise aşağıdaki gibidir:

$$I_{jk} = \sum_{t=1}^M \frac{1}{G_{R_t}} \left(A_{jkr_t} - \frac{H_{jR_t} H_{kR_t}}{G_{R_t}} \right). \quad (2.19)$$

Log olasılık fonksiyonu için Breslow'un yaklaşımı, Kalbfleisch and Prentice (1980) tarafından aşağıdaki gibi ifade edilmiştir:

$$\log(\beta) = \sum_{t=1}^M \left\{ \sum_{d \in D_t} \sum_{i=1}^p x_{id} \beta_i - m_t \ln \left[\sum_{r \in R_t} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \right] \right\}$$

$$= \sum_{t=1}^M \left\{ \sum_{d \in D_t} \sum_{i=1}^p x_{id} \beta_i - m_t \ln(G_{R_t}) \right\}. \quad (2.20)$$

Burada,

$$G_R = \sum_{r \in R} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) \quad (2.21)$$

biçiminde yazılabilir. EÇO çözümü Newton-Raphson yöntemiyle elde edilmektedir. Bu yöntem birinci ve ikinci dereceden kısmi türevlere ihtiyaç duymaktadır. Birinci dereceden kısmi türev;

$$U_j = \frac{\partial \log(\beta)}{\partial \beta_j} = \sum_{t=1}^M \left\{ \left(\sum_{d \in D_t} x_{jd} \right) - \left(m_t \frac{H_{jR_t}}{G_{R_t}} \right) \right\} \quad (2.22)$$

biçimindedir. Bilgi matrisini oluşturan ikinci dereceden negatif kısmi türev,

$$I_{jk} = \sum_{t=1}^M \frac{m_t}{G_{R_t}} \left(A_{jkR_t} - \frac{H_{jR_t} H_{kR_t}}{G_{R_t}} \right) \quad (2.23)$$

biçimindedir.

Log olabilirlik fonksiyonu için Efron'un yaklaşımı, Kalbfleisch and Prentice (1980) tarafından aşağıdaki gibi ifade edilmiştir:

$$\begin{aligned} \log(\beta) &= \sum_{t=1}^M \left\{ \sum_{d \in D_t} \sum_{i=1}^p x_{id} \beta_i - \sum_{d \in D_t} \ln \left[\sum_{r \in R_t} \exp\left(\sum_{i=1}^p x_{ir} \beta_i\right) - \frac{d-1}{m_t} \sum_{c \in D_t} \exp\left(\sum_{i=1}^p x_{ic} \beta_i\right) \right] \right\} \\ &= \sum_{t=1}^M \left\{ \sum_{d \in D_t} \sum_{i=1}^p x_{id} \beta_i - \sum_{d \in D_t} \ln \left[G_{R_t} - \frac{d-1}{m_t} G_{R_t} \right] \right\} \end{aligned} \quad (2.24)$$

EÇO fonksiyonunun çözümü Newton-Raphson yöntemiyle bulunmaktadır. Bu yöntem birinci ve ikinci dereceden kısmi türevlere ihtiyaç duymaktadır. Birinci dereceden kısmi türev;

$$\begin{aligned}
 U_j &= \frac{\partial LL(\beta)}{\partial \beta_j} = \sum_{t=1}^M \left\{ \sum_{d \in D_t} x_{jd} - \frac{H_{jR_t} - \left(\frac{d-1}{m_t}\right) H_{jD_t}}{G_{R_t} - \left(\frac{d-1}{m_t}\right) G_{D_t}} \right\} \\
 &= \sum_{t=1}^M \sum_{d \in D_t} x_{jd} - \sum_{t=1}^M \sum_{d=1}^{m_t} \left(\frac{H_{jR_t} - \left(\frac{d-1}{m_t}\right) H_{jD_t}}{G_{R_t} - \left(\frac{d-1}{m_t}\right) G_{D_t}} \right) \quad (2.25)
 \end{aligned}$$

biçimindedir. İkinci dereceden kısmi türev, tahmin edilen regresyon katsayılarının kovaryans matrisi, tahmin eden bilgi matrisi kullanılarak elde edilir.

İkinci kısmi türevlerin negatifleri,

$$I_{jk} = -\frac{\partial^2 \log(\beta)}{\partial \beta_j \partial \beta_k} = \sum_{t=1}^M \sum_{d=1}^{m_t} \frac{\left(G_{R_t} - \left(\frac{d-1}{m_t}\right) G_{D_t} \right) \left(A_{jR_t} - \left(\frac{d-1}{m_t}\right) A_{jD_t} \right) - \left(H_{jR_t} - \left(\frac{d-1}{m_t}\right) H_{jD_t} \right) \left(H_{kR_t} - \left(\frac{d-1}{m_t}\right) H_{kD_t} \right)}{\left(G_{R_t} - \left(\frac{d-1}{m_t}\right) G_{D_t} \right)^2} \quad (2.26)$$

biçimindedir.

2.2. Yaşam ve Tehlike Fonksiyonlarının Tahmini

EÇO tahminlerini elde ettikten sonra, yeni veya mevcut bir birimin yaşam olasılığını belirli bir zaman noktasında belirli ortak değişkenlerle tahmin etmek istenebilir. Kalbfleisch ve Prentice (1980) tarafından önerilen yöntemler yaşam olasılıklarını tahmin etmek için kullanılmaktadır.

Değişkenlerinin tamamı sıfıra eşit olan bir birimin birikimli yaşam fonksiyonunu tahmin eder. X_0 değişkeni için yaşam fonksiyonu,

$$S\left(\frac{t}{X_0}\right) = \exp\left(H\left(\frac{t}{X_0}\right)\right) = \exp\left(H_0\left(\frac{t}{X_0}\right)\right) \exp\sum_{i=1}^p x_{0i}\beta = [S_0(t)]^{\exp\sum_{i=1}^p x_{0i}\beta} \quad (2.27)$$

biçiminde tanımlanabilir. Temel yaşam fonksiyonunun $S(t)$ tahmini, birikimli tehlike fonksiyonundan hesaplanır ve aşağıdaki gibidir:

$$S_0(t_0) = \prod_{T_t \in T_0} \alpha_t \quad (2.28)$$

Burada,

$$\alpha_t = \frac{S(t_t)}{S(t_{t-1})} = \left[\frac{S_0(t_t)}{S_0(t_{t-1})} \right]^{\theta_t} \quad (2.29)$$

biçiminde ifade edilebilir. t zamanında koşullu temel yaşam olasılığı olan α_t değeri, koşullu olabirlik denkleminin çözümü aşağıdaki gibidir:

$$\sum_{d \in D_t} \frac{\theta_d}{1 - (\alpha_t)^{\theta_d}} = \sum_{r \in R_t} \theta_r \quad (2.30)$$

Burada $\theta_r = \exp\left(\sum_{i=1}^p x_{ir}\beta_i\right)$ biçiminde gösterilebilir. Eş zamanlı gözlemler yok iken, D_t bir birimi içerir ve yukarıdaki denklemin doğrudan çözümü,

$$\hat{\alpha}_t = \left[1 - \frac{\hat{\theta}_t}{\sum_{r \in R_t} \hat{\theta}_r} \right]^{\hat{\theta}_t^{-1}} \quad (2.31)$$

biçiminde yazılabilir. Hosmer ve Lemeshow (1999), temel tehlike oranını aşağıda gösterildiği gibi hesaplamaktadır:

$$h_0(t_t) = 1 - \alpha_t \quad (2.32)$$

Birikimli tehlike fonksiyonunun tahmini $\hat{H}_0(T)$, birikimli tehlike fonksiyonu ve birikimli yaşam fonksiyonu arasındaki ilişkiden doğmuştur. Temel yaşam fonksiyonu tahmini ile birikimli tehlike fonksiyonu arasındaki bağıntı,

$$\hat{H}_0(t) = -\ln(\hat{S}_0(t)) \quad (2.33)$$

$$\hat{H}(t) = -\exp\left(x_i \hat{\beta}_i\right) \ln(\hat{S}_0(t)), \quad i=1, \dots, p \quad (2.34)$$

biçiminde ifade edilebilir. Bir birimin X_0 'ın değişkenler kümesinin birikimli yaşam fonksiyonunun tahmini aşağıdaki gibi elde edilir:

$$\hat{S}\left(\frac{t}{X_0}\right) = \hat{S}_0(t)^{\exp \sum_{i=1}^p x_{i0} \hat{\beta}_i}. \quad (2.35)$$

2.3. İstatiksel Hipotez Testleri ve Güven Aralıkları

Modelde bir veya daha fazla açıklayıcı değişkenin önemliliğini test etmek için olabilirlik oran testi (LR) ve Wald testi olmak üzere iki test vardır. Simülasyon çalışmaları genellikle olabilirlik oranı testinin Wald testinden daha iyi gerçekleştiğini göstermektedir. Ancak, Wald testi hala hesaplama kolaylığı nedeniyle regresyon katsayılarının önemliliğini test etmek için kullanılmaktadır. Olabilirlik oran test istatistiği, iki modelin log olabilirliği arasındaki farkın -2 katıdır. LR istatistiğinin dağılımı, büyük örneklem için ki-kare dağılımı ile yakından ilgilidir. Yaklaşık ki-kare dağılımının serbestlik derecesi, iki modeldeki regresyon katsayılarının sayısındaki farka eşittir. İki log olabilirlik fonksiyonu arasındaki fark iki olabilirlik oranının logaritmasına eşit olduğu için test, bir farktan ziyade bir oran olarak adlandırılır. Yani, $LR = -2[L_{alt \text{ küme}} - L_{tam \text{ küme}}] = -2[\ln(I_{alt \text{ küme}} / I_{tam \text{ küme}})]$ biçiminde ifade edilebilir. Olabilirlik oran testi, CRM'de tercih edilen testtir. Çoklu regresyonda, belirli bir regresyon katsayısının önemini test etmek için kullanılan t testi, bir Wald testidir. Wald istatistiğinin formülü, $(z_j)^2 = (b_j / s_{b_j})^2$ biçiminde ifade edilebilir. Burada s_{b_j} , b_j 'nin standard hatasının tahminidir. Büyük örneklem için z_j 'nin dağılımı, normal dağılıma yaklaşmaktadır.

2.4. Orantılı Tehlikeler (OT) Varsayımı

OT varsayımı, tehlike oranının zaman içinde sabit olmasını gerektirmektedir. Tehlike oranı, bir birimin tehlike fonksiyonunun diğer birimin tehlike fonksiyonuna bölünmesi olarak tanımlanır. X^* , daha büyük tehlikeye sahip gruba ait açıklayıcı değişkenler ve X ise daha küçük tehlikeye karşılık gelecek açıklayıcı değişkenler biçimde tanımlandığı varsayılın. $X^* = x_1^*, x_2^*, \dots, x_p^*$ ve $X = x_1, x_2, \dots, x_p$ olmak üzere tehlike oranı,

$$\hat{TO} = \frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \exp\left(\sum_{i=1}^p \beta_i (X_i^* - X_i)\right) \quad (2.36)$$

biçimindedir. Burada üstel ifadeyi $\hat{\theta}$ ile gösterirsek tehlike oranının tahmini için üstel ifadenin zamana bağlı olmayan bir sabit olduğu görülmektedir. Örneğin, sadece bir X değişkeni olduğunu varsayalım, X_1 , (0,1) durumunu gösterecek, böylece $p=1$ olacaktır, $X = (X_1, X_2, \dots, X_p) = (X_1)$ olur. $X_1^* = 1$ ve $X_1 = 0$ iken tehlike oranı,

$$\hat{TO} = \exp\left[\hat{\beta}_1(X_1^* - X_1)\right] = \exp\left[\hat{\beta}_1(1-0)\right] = e^{\hat{\beta}_1} \quad (2.37)$$

olur. (Anonim, 2019).

OT varsayımının değerlendirmesinde çoğunlukla Schoenfeld artıklarından yararlanılmaktadır. Açıklayıcı değişkenin değeri ile risk kümesinin ortalamasının arasındaki fark Schoenfeld artıkları olarak tanımlanmaktadır. Bu artıklar sıfır ya da sıfırın etrafında dağılıyorsa OT varsayımı sağlanmaktadır (Kleinbaum ve Klein, 2005).

2.5. Model Seçim Yöntemleri

Model seçim sürecindeki ilk adım CRM'nin doğrusal bileşenine dahil edilecek olan açıklayıcı değişkenleri belirlemektir. En iyi olarak nitelendirilen tek bir model yerine, birçok iyi modelin değerlendirilmesi amaçlanır.

Etkileşim terimlerini, doğrusal olmayan terimleri vb. içeren açıklayıcı değişkenlerin sayısı çok fazla değilse, terimlerin tüm mümkün birleşimlerini elde etmek mümkündür. Seçenek iç içe modeller, modele terim ekleyerek ya da modelden terim çıkararak $-2\log\hat{L}$ 'nin değerindeki değişikliği inceleyerek karşılaştırılır. İç içe olması gerekmeyen

olası modeller arasındaki karşılaştırmalar için Akaike bilgi kriteri (AIC) veya Bayesci bilgi kriteri (BIC) kullanılmaktadır. AIC veya BIC değeri düşük olan model en uygun model olarak seçilmektedir (Klein ve Moeschberger, 1997). CRM için Akaike bilgi kriteri,

$$AIC = -2\log L + 2p \quad (2.38)$$

biçiminde hesaplanmaktadır. Burada p modeldeki bilinmeyen parametrelerin sayısıdır (Klein ve Moeschberger, 1997; Collett, 1994). Bayesci bilgi kriteri ise,

$$BIC = -2\log L + p\log(n) \quad (2.39)$$

biçimindedir. Burada p modeldeki bilinmeyen parametrelerin sayısını ve n toplam gözlem sayısını belirtmektedir.

3. KESİRLİ POLİNOMLU YAŞAM MODELLERİ

3.1. Giriş

Araştırmacılar regresyon modellerinde çoğu zaman sonuçları etkileyebilecek birçok değişken ile karşı karşıya kalmaktadır. Değişkenlerin seçiminde kullanılan stratejiler, 'önemli' değişkenlerin alt kümesini tanımlamak için kullanılır. Sıralı yöntemler (örn. adımsal ya da geriye doğru seçim yöntemleri) ya da farklı optimizasyon kriterleri ile tüm alt küme seçimi gibi yöntemler ile ilişkili zorluklar, aşırı uyum, yetersizlik, regresyon parametrelerinin yanlışlığı ve yeni veride yeni regresyon parametrelerinin yeniden üretilebilirliğinin olmamasıdır. Her ne kadar konu bilgisi değişken seçimine rehberlik etse de, bazı değişkenleri modele dahil etmek ya da çıkarmak için p değerinin kullanılması kaçınılmaz olacaktır. Yeni verilerde iyi tahmin özelliklerine sahip bir model üretmek için "en iyi" stratejinin tanımı zordur (Miller, 1990).

Model oluşturmanın ikinci bir zorluğu ise sonuç ile sürekli bir değişken arasındaki ilişkide doğrusal olmama sorunu ile nasıl başa çıkılacağıdır. Geleneksel doğrusallık varsayımı yanlış olabilir. İlgili değişkenin modele dahil edilmemesi yanlış bir son modele yol açabilir, çünkü varsayılan fonksiyonel biçim bilinmeyen gerçek biçiminden büyük ölçüde farklı olabilir. Alternatif olarak, sürekli değişkenler iki veya daha fazla kategoriye ayrılarak kategorik değişkenlere dönüştürülebilir.

Kesim noktasının seçiminde, en tatmin edici sonuç veren bir değer arayışı vardır. Tüm olası kesim noktaları denenir ve istatistiksel önemi en üst düzeye çıkaracak değer seçilir. (Altman ve ark., 1994; Miller ve Siegmund, 1982; Lausen ve Schumacher, 1996). Seçilen kesim noktası geniş bir güven aralığına sahip olacak ve hiçbir anlam ifade etmeyecektir. Alternatif bir yaklaşım, değişkeni sürekli tutmak ve doğrusal olmama durumuna izin vermektir. Karesel veya kübik polinomlar kullanmak yerine, Royston ve Altman (1994) tarafından genel bir regresyon modeli önerilmiştir. Bu model kesirli polinom fonksiyonları kullanmaktadır ve kesirli polinomlu regresyon modeli olarak adlandırılır.

3.2. Kesirli Polinomlu Cox Regresyon Modeli

Kesirli polinomlar, ilk kez Royston ve Sauerbrei (1994) tarafından regresyon modelinde sürekli açıklayıcı değişkenler ile sonuç değişkeni arasındaki ilişkinin modellenmesi için önerilmiştir. Sauerbrei ve diğ. (1999) tarafından meme kanseri verisinde semptom ve tanı verileri modellenmiştir. Berger ve diğ. (2000) mide kanseri verilerini kullanarak kesirli polinomlarda dinamik CRM'yi kullanarak model oluşturmuşlardır. Klasik yöntemler ile sonuçları karşılaştırmışlardır. Berger ve diğ. (2003) mide kanseri verisi için CRM'yi kullanarak yaşam verilerindeki dinamik etkilerin modellenmesi için bir yöntem önermişler ve klasik yöntemler ile karşılaştırma yapmışlardır. Look ve diğ. (2007) Rotterdam akciğer kanseri verilerini kullanarak zamana bağlı değişkenleri olan kesirli polinomlarda dinamik CRM'yi kullanarak model oluşturmuşlardır. Bellera ve diğ. (2010) meme kanseri veri kümesi için dinamik CRM'yi kullanarak kanserin yayılma belirtilerini incelemişlerdir. Buchholz ve Sauerbrei (2011) meme kanseri ve akciğer kanseri verilerini kullanarak çok değişkenli modellerde doğrusal olmayan ve zamanla değişen etkileri değerlendirmek için model oluşturma yöntemlerini karşılaştırmışlardır. Zhang ve ark. (2014) ilk laktat ve normalleşme süresinin tehlike fonksiyonu ile olan ilişkisini araştırırken kesirli polinomlu CRM'yi kullanılmışlar ve en uygun modeli bulmaya çalışmışlardır. Buchholz ve diğ. (2014) Rotterdam akciğer kanseri verilerini kullanarak zamana bağlı değişkenlere ait yaşam eğrilerini kesirli polinomlarda dinamik CRM ile incelenmiştir.

3.3. Kesirli Polinomlar

Kesirli Polinom (KP) yöntemi, açıklayıcı bir değişkenin model için önemli olup olmadığını ve fonksiyonel formunun belirlenmesine izin veren bir yöntemdir. Kesirli polinomlar doğrusallık sağlanmadığında sürekli değişkenlerin yapısını korumak için kullanılır.

Bir sonuç değişkenine, tek bir sürekli değişken X 'e ve bunları ilişkilendiren uygun bir regresyon modeline sahip olduğumuzu varsayalım. Başlangıç noktası doğrusal model olan $\beta_1 X$ 'dir. Çoğu zaman bu, ilişkinin yeterli bir açıklaması olacaktır, ancak diğer modeller, olası iyileştirmeler için araştırılmalıdır. Doğrusal modelin dönüşüm formülü $\beta_1 X^p$ 'dir. Bu model değişik p değerleriyle kullanılmaktadır. Model, Royston ve Altman (1994) tarafından birinci dereceden kesirli polinom veya KP1 şeklinde adlandırılmış ve

$$KP1 = \beta_1 x^p. \quad (3.1)$$

biçiminde ifade edilmiştir. p değerleri sınırlı bir S adı verilen diziden seçilir. Royston ve Altman (1994), X^0 'ın $\log(x)$ 'i ifade ettiği S kümesi için $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ değerlerini önermişlerdir. S kümesi, logaritmik, karekök vb. dönüşümleri içermez. Model uyumu için 8 tane p değeri denenir. En uygun model, en yüksek olasılığı veren değerdir. Doğrusal olmama, $p=1$ 'e karşı $p \neq 1$ 'in anlamlılık testidir. Bu test, sapma farkını serbestlik derecesi 1 olan ki-kare dağılımı ile karşılaştırarak gerçekleştirilmektedir. Eğer test anlamlı değilse doğrusal model kabul edilir, yoksa doğrusal olmayan kesirli polinom tercih edilir.

Polinomiyal regresyonda olduğu gibi, KP1 fonksiyonundan daha karmaşık olan KP2 fonksiyonuna geçilir. $\beta_1 X_1 + \beta_2 X_2$ denklemini kullanmak yerine, KP2 fonksiyonları S kümesinden alınan p_1 ve p_2 değerleriyle $\beta_1 x^{p_1} + \beta_2 x^{p_2}$ olarak tanımlanır. KP2 ise,

$$\begin{aligned} KP2 &= \beta_1 x^{p_1} + \beta_2 x^{p_2} & p_1 \neq p_2 \\ &= \beta_1 x^{p_1} + \beta_2 x^{p_2} \ln x & p_1 = p_2 = p \end{aligned} \quad (3.2)$$

biçimindedir (Ata Tutkun, 2015). Eğer $p_2 = p_1$ ise; Royston ve Altman (1994) tekrarlı üslü adını verdikleri $\beta_1 x^{p_1} + \beta_2 x^{p_2} \ln x$ fonksiyonunu önermişlerdir. S kümesinden elde edilen 36 tane üs kombinasyonu arasından en yüksek olasılık değeri olarak tanımlanır. KP modelinin genel denklemi ise,

$$\begin{aligned} \varphi_m^*(x; p) &= \beta_0 + \varphi_m(x; p) = \sum_{j=1}^m \beta_j h_j(x), \quad j=1, \dots, m \\ h_j(x) &= x^{p_j} & p_j \neq p_{j-1} \\ &= h_{j-1} \log x & p_j = p_{j-1} \end{aligned} \quad (3.3)$$

biçimindedir (Royston ve Sauerbrei, 2008).

3.3.1. Model Seçiminde Kullanılacak Yöntemler

KP1 için 8 ve KP2 için 36 olmak üzere toplam 44 model veri kümesi için çalıştırılır ve bu modellerden en uygun olanının seçilmesi gerekmektedir. Bu amaçla, kapalı test yöntemi(RA2 yöntemi) ve ardışık yöntem olmak üzere iki yöntem önerilmiştir.

3.3.1.1. Kapalı Test Yöntemi

Bu algoritma Ambler ve Royston (2001) ve Sauerbrei ve Royston (2002) tarafından tanımlanmıştır. Bu algoritma RA2 olarak bilinir. Bu test her bir değişken için yaklaşık 1.tip hata değerini verir. Bu yöntem, aday modeller arasında, belirlenmiş bir minimumdan (yokluk modeli), test sonuçlarının düzenli bir sıralamasına göre öngörülen maksimuma (kp) kademeli olarak artmasına izin verir. Algoritma adımları aşağıdaki gibidir:

1. En iyi ikinci dereceden kesirli polinom modeli 4 serbestlik derecesinde alfa anlamlılık düzeyinde yokluk modeline göre test edilir. Eğer test anlamlı değilse x değişkeni önemli değildir ve bir sonraki aşamaya geçilmez, x değişkeni anlamlıysa bir sonraki aşamaya geçilir.
2. En iyi ikinci dereceden kesirli polinom modeli 3 serbestlik derecesinde alfa anlamlılık düzeyinde doğrusal modele göre test edilir. Eğer test anlamlı değilse bir sonraki aşamaya geçilmez yani son model doğrusal modeldir yoksa bir sonraki aşamaya geçilir.
3. En iyi ikinci dereceden kesirli polinom modeli 2 serbestlik derecesinde alfa anlamlılık düzeyinde en iyi birinci dereceden kesirli polinom modeline göre test edilir. Eğer test anlamlı değilse en iyi model birinci dereceden kesirli polinom modelidir, yoksa en iyi model ikinci dereceden kesirli polinom model olarak alınır.

Sırasıyla 1., 2. ve 3. adımdaki testler genel ilişkiyi, doğrusal olmamayı ve daha basit veya karmaşık kesirli polinom modelleri arasındaki seçimi inceler.

3.3.1.2. Ardışık Yöntem

En iyi modeli seçmek için kullanılan ikinci alternatif algoritma olan ardışık yöntemin adımları aşağıdaki gibidir:

1. En iyi ikinci dereceden kesirli polinom modeli 2 serbestlik derecesinde alfa anlamlılık düzeyinde birinci dereceden kesirli polinom modeline göre test edilir. Eğer test anlamlıysa bir sonraki aşamaya geçilmez ve en iyi model ikinci dereceden kesirli polinom modeli olarak alınır.

2. En iyi birinci dereceden kesirli polinom modelini 1 serbestlik derecesinde alfa anlamlılık düzeyinde doğrusal modele göre test edilir. Eğer test anlamlıysa bir sonraki aşamaya geçilmez ve en iyi model birinci dereceden kesirli polinom olarak alınır. Yoksa bir sonraki aşamaya geçilir.
3. Doğrusal modeli alfa anlamlılık düzeyinde, 1 serbestlik derecesinde x'in olmadığı modele göre test edilir. Eğer test anlamlıysa en iyi model doğrusal model olarak alınır. Yoksa x modele dahil edilmez.

Doğru ilişki doğrusal model olduğunda 1. tip hata değeri nominal alfa değerini aşabilir. Bu yüzden prosedür basit olanlara kıyasla daha karmaşık modelleri tercih etmektedir. Amber ve Royston (2001)'nin yapmış olduğu simülasyon çalışmasının sonuçlarına dayanarak en uygun fonksiyon seçimi için kapalı test yöntemi (RA2 yöntemi) önerilir (Sauerbrei ve ark., 2006).

3.3.2. Ölçekleme ve Odaklama

Dönüştürülmüş x^p değerlerinin hassas ölçeklendirilmiş ve ortalanmış (centred) olmasını sağlamak için sayısal ve kavramsal avantajlar vardır. Ölçeklendirme, model tahmininde belirsizlikleri ve zorlukları azaltır. Merkezleme (centring), CRM'de β_0 kesim noktasının veya temel tehlike fonksiyonunun anlamlı bir şekilde yorumlanmasını sağlar. Ölçekleme işlemi aşağıdaki gibi ifade edilebilir:

$$\text{Irange} = \log_{10} [\max(x) - \min(x)] \quad (3.4)$$

$$\text{ölçek} = 10^{\text{sign(Irange)} \cdot \text{int}(|\text{Irange}|)} \quad (3.5)$$

$$x^* = x / \text{ölçek} \quad (3.6)$$

Burada $\text{int}(x)$, x'in altındaki en yakın tam sayıyı gösterir. Miktar ölçeği 10'un üssü olan, bir tamsayıdır. X^* 'in KP dönüşümü daha sonra x^* gözlenen değerlerinin ortalama sına odaklanır. Örneğin, KP1 modeli için $\beta_0 + \beta_1 x^p$, STATA programı tarafından

$$\beta_0 + \beta_1 \left(\left(\frac{x}{\text{ölçek}} \right)^p - \left(\frac{1}{n} \sum_{i=1}^n x_i^* \right)^p \right)$$

biçiminde uyarlanır.

İşlem sırası ilk olarak ölçekleme, ikincisi kesirli polinom üslerini tahmin etme ve son olarak merkezlemedir. İlk aşamada merkezleme farklı üslere yol açabilir ve bu yüzden ilk aşamada kullanılması tavsiye edilmemektedir. Üslerin tahmini ölçeklemeden etkilenmez. Ölçeğin herhangi bir pozitif değeri için aynı üsler $x/\text{ölçek}$ için bulunur. Ölçekleme, modelin doğruluğunu korumak için gereklidir (Royston ve Sauerbrei, 2008).

3.3.3. Kesirli Polinomlar ile Çok Değişkenli Model Oluşturma

Model oluşturmada amaç uygun, yorumlanabilir çok değişkenli bir model türetmek ve daha güçlü tahmin ediciler elde etmektir. Sürekli değişkenler için doğru fonksiyonel biçimlerin belirlenmesi de önemlidir. Doğrusal olmayan modeller için uygun modeli oluştururken değişken seçim yöntemi olarak geriye doğru seçim yöntemi de ileriye doğru değişken seçim yöntemi ile birlikte kullanılır. Sauerbrei ve Royston (1999) bunu çok değişkenli kesirli polinom (Multivariate Fractional Polynomial, MFP) olarak adlandırmışlardır. Çok değişkenli kesirli polinom, yalnızca regresyon modelleri oluşturma hakkında genel bilgi gerektirir. İki anlamlılık seviyesine ihtiyaç vardır: Geriye doğru seçim ile değişkenleri seçmek için α_1 ve ileriye doğru seçim yöntemi içindeki işlevlerin uyumunu karşılaştırmak için α_2 dir. Çok değişkenli kesirli polinom (α_1, α_2) notasyonu, Geriye doğru seçim için α_1 ve ileriye doğru seçim yöntemi için α_2 kullanılarak seçilen bir modeli gösterir. Ortak durum $\alpha_1 = \alpha_2 = \alpha$, çok değişkenli kesirli polinom (α) olarak kısaltılmıştır.

Çok değişkenli kesirli polinom algoritması, yani MFP (α_1, α_2) prosedürünün adımları aşağıdaki gibidir. Tüm anlamlılık testleri ki-kare olasılık oranını esas almaktadır.

1. Nominal p değerleri α_1, α_2 olarak seçilmiştir. $\alpha_1=\alpha_2=0.05$ 'dir. Değerler değişkenler arasında değişebilir. Bir değişken için $\alpha_1=1$ alındığında değişken modele alınır. Sürekli bir değişken için $\alpha_2=1$ olduğunda kesirli polinom fonksiyonu için modele alınır.
2. Maksimum serbestlik dereceli kesirli polinom fonksiyonları seçilir. Örneğin; serbestlik derecesi=2 ve serbestlik derecesi=1 sırasıyla KP2, KP1 ve doğrusal fonksiyonlar anlamına gelir.
3. Tahmin edicilerin modele alınma sırası, her bir tahmin edicinin modelden çıkarılması için p değerine göre belirlenir. En önemli tahmin edici, ilk sırada modele alınır. Prosedürün tüm döngülerinde tutulan x_1, \dots, x_k değişkenlerinin bu sırayla düzenlendiği varsayılır.

4. Döngü sayacı $c=0$ olarak başlatılsın.
5. Her döngü içindeki değişken sayacını başlatmak için $j = 1$ olsun.
6. Eğer x_j sürekli değişkense 7. adıma gidilir; değilse x_j değişkeni kategorik ya da ikili değişkendir. Kukla değişkenlerin ortak önemi α_1 anlamlılık düzeyinde test edilir. Diğer tüm değişkenler varsayım olarak modele dahil edilmiştir. Eğer x_j önemliyse modelde tutulur; değilse modelden çıkarılır. 8.adıma gidilir.
7. Fonksiyon seçim prosedürünün 1. adımı $\alpha = \alpha_1$ seviyesinde x_j 'ye uygulanır. Eğer x_j önemli değilse modelden çıkarılır; önemliyse fonksiyon seçim prosedürünün 2. ve 3. adımları $\alpha = \alpha_2$ seviyesinde, kesirli polinom veya doğrusal fonksiyonu seçmek için uygulanır. Diğer tüm değişkenler varsayım olarak modele dahil edilmiştir. p üslerinin seçildiği kesirli polinom durumunda, x_j dönüştürülmüş x_j^p olarak temsil edilir.
8. x_j 'nin modele dahil edilmesi veya modelden çıkarılması, bir sonraki döngüde ele alınana kadar geçerlidir.
9. $j=j+1$ olsun. Eğer $j \leq k$ ise, sonraki tahmin edici için 6. adıma geri dönülür; değilse 10. adıma gidilir.
10. $c=c+1$ olsun. c . döngüde adımlar tamamlanır. Eğer $c > c_{\max}$ ise dur. c_{\max} döngülerinde, algoritma başarısız olur. Aksi halde, dahil edilen değişkenlerin ve FP dönüşümlerinin $c-1$ döngüsünden c döngüsüne değişip değişmediği kontrol edilir. Eğer öyleyse, yeni bir döngü başlatmak için 5. adıma dönülür. Yoksa, döngü durdurulur ve mevcut model tahminleri kullanılır (Royston ve Sauerbrei, 2008).

4. UYGULAMA

4.1. Meme Kanseri Veri Kümesi

Hacettepe Üniversitesi Onkoloji Bölümü'nde Ocak 1980 ile Eylül 1991 tarihleri arasında meme kanseri tanısı ile hastaneye başvuran ve 5 yıl izlenen 124 hastaya ait veri kümesi kullanılmıştır (Erdoğan,1993). Ameliyat sonrası hastalığın nüksetmesi başarısızlık olarak tanımlanmıştır. 124 hastadan 80'i (%64.52) başarısız ve 44'ü (%35.48) ise durdurulmuştur. Yaşam süresi, hastalığın nüksetmesine veya durdurulmasına kadar geçen süre (ay) olarak ele alınmıştır.

Çalışmada analizler için STATA paket programı kullanılmıştır. Yaş, menapoz, tümör çapı, aksiller lenf nodu, radyoterapi, toksidasyon, müdahale tipi, hastalığın evresi ve tedavi türü değişkenleri incelenmiş ve bu değişkenlere ait bilgiler Çizelge 4.1' de verilmiştir. Yaş değişkeni sürekli bir değişken, diğer değişkenler ise kategorik değişken olarak ele alınmıştır. Hastaların yaş ortalaması 50.59 ± 1.103 (min=25, max=78), ortalama yaşam süresi 81 ± 6.104 ay ve 5 yıllık yaşam olasılığı %59.4 olarak bulunmuştur.

Çizelge 4.1. Meme kanseri veri kümesi için kullanılan değişkenler ve düzeyleri

Değişken	Değişken Düzeyleri	n (%)	Başarısız Gözlemlerin Sayısı	Durdurulmuş Gözlemlerin Sayısı
Kategorik yaş*	1: 40 yaş ve altı	26 (%46.2)	14	12
	2: 41 - 50 yaş	44 (%31.8)	30	14
	3: 51 - 60 yaş	29 (%37.9)	18	11
	4: 61 yaş ve üzeri	25 (%28.0)	18	7
Menapoz	0: Girmiş	66 (%33.3)	44	22
	1: Girmemiş	58 (%37.9)	36	22
Tümör Çapı	0:2 cm'den küçük	31 (%25.8)	23	8
	1:2-5 cm arası	66 (%31.8)	45	21
	2:5 cm'den büyük	27 (%55.6)	12	15
Aksiller Lenf Nodu	0: yok	50 (%32.0)	34	16
	1:1-3 adet	52 (%32.7)	35	17
	2:4'ten fazla	22 (%50.0)	11	11
Radyoterapi	0: Almış	66 (%22.7)	51	15
	1: Almamış	58 (%50.0)	29	29
Toksidasyon	0: Herhangi bir şey yok	76 (%30.3)	53	23
	1: Bulantı,Kusma, Sıcak basması	48 (%43.8)	27	21
Müdahele tipi	0: Sabit mastektomi	21 (%52.4)	10	11
	1: Modifiye radikale mastektomi	48 (%31.2)	33	15
	2: Radikale mastektomi	55 (%32.7)	37	18
Hastalığın evresi	0: Evre 1	42 (%33.3)	28	14
	1: Evre 2	82 (%36.6)	52	30
Tedavi türü	0: TMX	59 (%30.5)	41	18
	1: L-PAM	65 (%40.0)	39	26

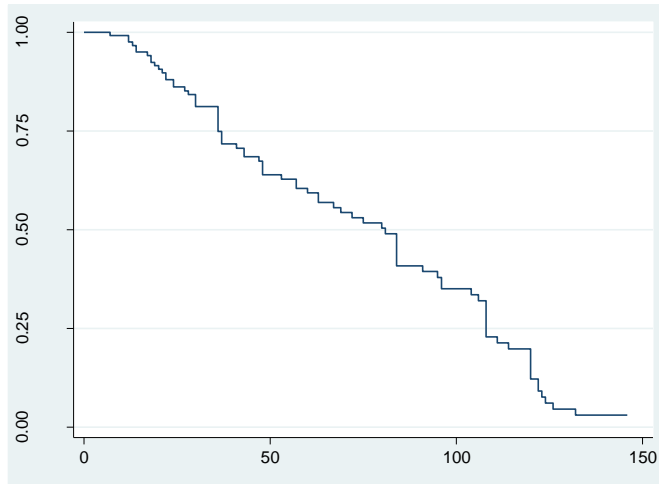
*: Yaş değişkeni kategorik olarak alındığında kullanılan sınıflandırma

Bu çalışmada klasik CRM ve kesirli polinomlu CRM veriye uygulanırken kullanılacak model kurma aşamaları aşağıdaki gibi özetlenebilir:

1. Modele tüm değişkenler dahil edilerek sadece yaş değişkenini sürekli kabul ederek klasik Cox regresyon çözümü yapılmıştır.
2. Sadece sürekli yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümü yapılmıştır.
3. Sadece kategorik yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümü yapılmıştır.
4. Modele tüm değişkenler var iken yaş kategorik değişken olarak modele dahil edilip klasik Cox regresyon çözümü yapılmıştır.
5. Sadece sürekli yaş değişkeni ele alınarak kesirli polinomlu Cox regresyon çözümü yapılmıştır.
6. Bütün değişkenler dahil edilerek ve sürekli yaş değişkeni alınarak kesirli polinomlu Cox regresyon çözümü yapılmıştır.

4.1.1. Meme Kanseri Veri Kümesi için Parametrik Olmayan Yaşam Çözümü Sonuçları

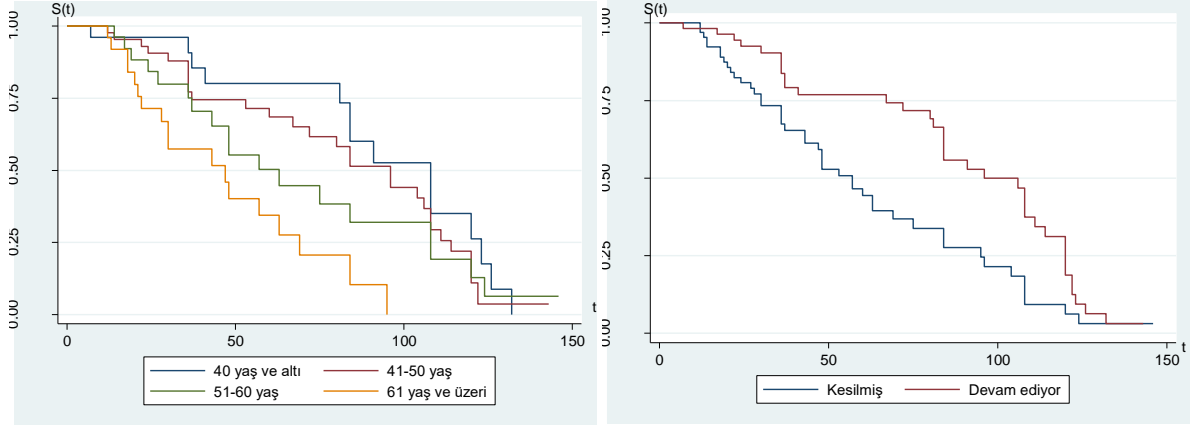
Şekil 4.1'de meme kanseri veri kümesine ait KM yaşam eğrisi verilmiştir.



Şekil 4.1. Meme kanseri verisi için KM yaşam eğrisi

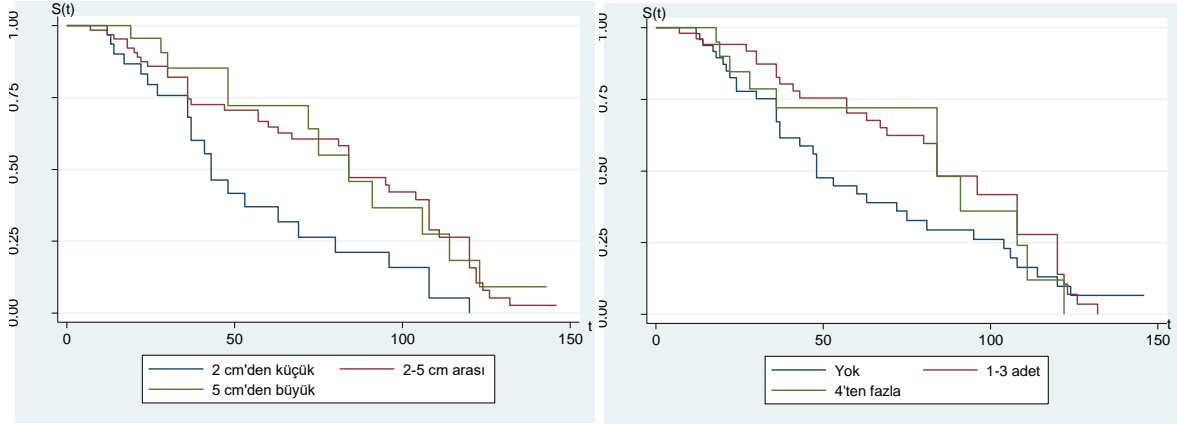
KM yaşam eğrisine bakıldığında yaşam olasılığının zaman ilerledikçe azaldığı görülmektedir.

Çalışmada kullanılan her bir açıklayıcı değişken için KM yaşam eğrileri ise Şekil 4.2'de verilmiştir.



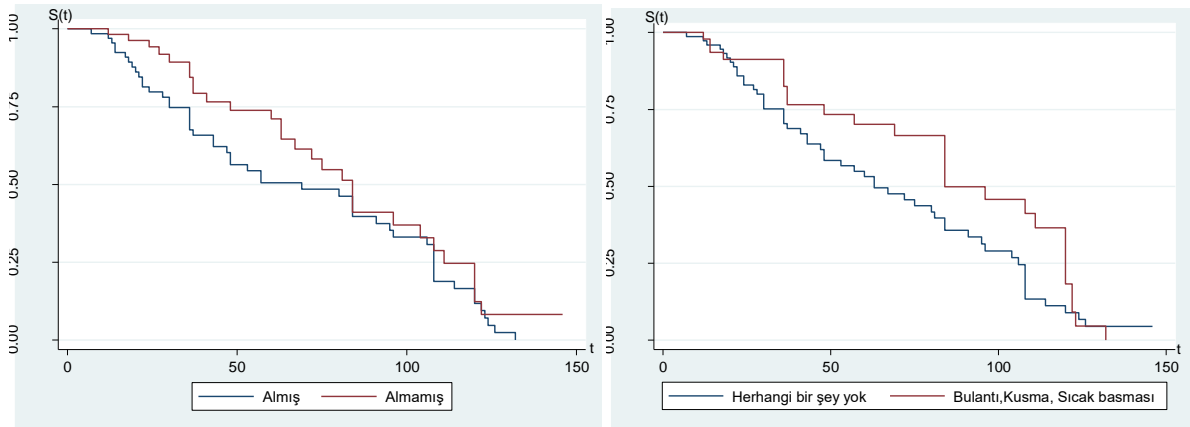
(a) Yaş

(b) Menapoz



(c) Tümör çapı

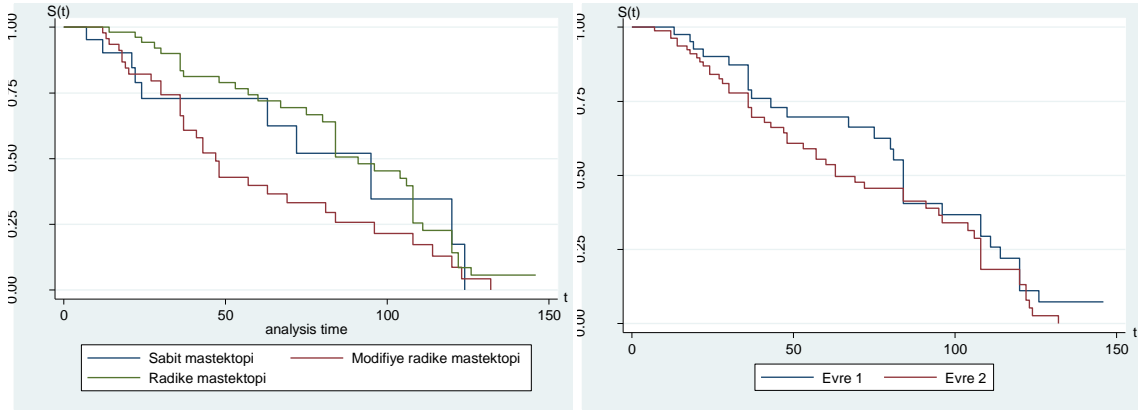
(d) Aksiller lenf nodu



(e) Radyo terapi

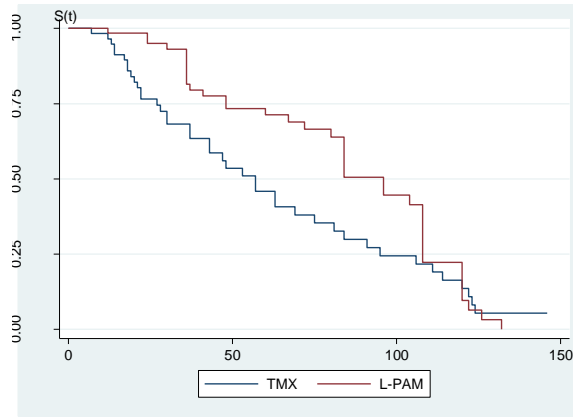
(f) Toksikasyon

Şekil 4.2. Meme kanseri verisinde kullanılan açıklayıcı değişkenlerin KM eğrileri



(g) Müdahale Tipi

(h) Evre



(i) Tedavi türü

Şekil 4.2. Meme kanseri verisinde kullanılan açıklayıcı değişkenlerin KM eğrileri (devam ediyor)

Şekil 4.2 incelendiğinde, radyoterapi, toksidasyon ve evre değişkenlerinde değişken düzeylerinin yaşam eğrileri birbirine çok yakın olarak elde edilmiştir. Diğer değişkenler için değişken düzeyleri arasında farklılığın daha fazla olduğu görülmektedir. Yaşı 40 yaş ve altı olan hastaların yaşam olasılığının diğer düzeylere göre daha yüksek olduğu ve daha sonrasında sırasıyla 41-50 yaş, 51-60 yaş, 61 yaş ve üzerindeki hastaların takip ettiği görülmektedir. Menapoza girmemiş hastaların yaşam olasılıklarının, menapoza girmiş hastalara göre daha yüksek olduğu söylenebilir. Tümör çapı 2-5 cm arası olan hastalar ile tümör çapı 5 cm'den büyük hastaların yaşam olasılıklarının birbirine yakın olduğu ve tümör çapı 2 cm'den küçük olan hastaların yaşam olasılıklarından yüksek olduğu söylenebilir. Aksiller lenf nodu 1-3 adet ve aksiller lenf nodu 4'ten fazla olan hastaların yaşam olasılıkları birbirine yakın ve aksiller lenf nodu olmayan hastaların yaşam olasılıklarından yüksek olduğu söylenebilir. Müdahale tipi radike mastektomi ile sabit mastektomi olan hastaların yaşam olasılıkları birbirine yakın ve müdahale

tipi modifiye radike mastektomi olan hastaların yaşam olasılıklarından yüksek olduğu söylenebilir. Tedavi türü L-PAM olan hastaların yaşam olasılıklarının tedavi türü TMX olan hastalara göre yüksek olduğu söylenebilir.

Yaşam olasılıkları açısından açıklayıcı değişken düzeyleri arasında istatistiksel olarak önemli bir fark olup olmadığı log-rank testi ile incelenmiş, sonuçlar Çizelge 4.2’de verilmiştir.

Çizelge 4.2. Meme kanseri verisi için log-rank testi sonuçları

Değişken	Ki-kare değeri	p değeri
Yaş	18.323	0.000
Menapoz	8.907	0.003
Tümör çapı	9.981	0.007
Aksiller Lenf Nodu	2.739	0.254
Radyoterapi	2.265	0.132
Toksidasyon	2.769	0.096
Müdahale Tipi	6.544	0.038
Evre	1.700	0.192
Tedavi türü	3.358	0.067

Çizelge 4.2’ye göre, %95 güven düzeyinde yaş ($p=0.000$), menopoz ($p=0.003$), tümör çapı ($p=0.007$) ve müdahale tipi ($p=0.038$) değişkenlerinin tüm düzeyleri arasında, yaşam olasılıkları açısından istatistiksel olarak önemli bir fark olduğu söylenebilir.

4.1.2. Meme Kanseri Veri Kümesi için Klasik CRM Sonuçları

Öncelikle, yaş sürekli değişken olarak ele alınmış ve çok değişkenli CRM modeli meme kanseri veri kümesine uygulanmıştır. Elde edilen sonuçlara göre p değeri 0.0002 olduğundan %95 güven düzeyinde kurulan modelin anlamlı olduğu söylenilebilmektedir. Bu model için AIC= 583.640 ve BIC= 617.484 olarak elde edilmiştir. Meme kanseri verisi için CRM’ye ait sonuçlar Çizelge 4.3’de verilmiştir.

Çizelge 4.3. Meme kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan çok değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Yaş	0.430	0.019	0.022	1.530	1.508, 1.566
Menapoz	-0.120	0.447	0.762	0.88	0.567, 1.386
Tümör çapı					
1: 2-5 cm arası	1.118	0.397	0.005	3.060	2.056, 4.549
2: 5 cm'den büyük	0.150	0.369	0.684	1.162	0.803, 1.680
Aksiller lenf nodu					
1:1-3 adet	-0.348	0.177	0.084	0.706	0.591, 0.842
2:4'ten fazla	-0.861	0.564	0.383	0.422	0.240, 0.743
Radyoterapi	0.374	0.168	0.127	1.453	1.228, 1.719
Toksidasyon	0.209	0.216	0.434	1.232	0.993, 1.529
Müdahale tipi					
1: Modifiye radike mastektomi	-0.434	1.172	0.015	0.647	0.200, 2.091
2: Radike mastektomi	0.593	0.631	0.289	1.809	0.962, 3.400
Hastalığın evresi	-0.326	0.380	0.236	0.721	0.493, 1.055
Tedavi türü	-0.088	0.317	0.764	0.915	0.666, 1.257

* Değişkenler için referans kategorisi; menopoz için "0: Girmiş", tümör çapı için "0:0-2 cm'den küçük", aksiller lenf nodu için "0: yok", radyoterapi için "0: Almış", toksidasyon için "0: Herhangi bir şey yok", müdahale tipi değişkeni için "0: Sabit mastektomi", hastalığın evresi değişkeni için "0: Evre 1" olarak alınmıştır.

Çizelge 4.3'den, yaş ($p=0.022$), tümör çapı 2 cm'den küçük ($p=0.002$), tümör çapı 2 ve 5 cm arası ($p=0.005$) ve müdahale tipi modifiye radike mastektomi ($p=0.015$) değişkenlerinin önemli risk faktörleri olduğu görülmüştür.

CRM sonuçlarına göre, müdahale tipi sabit olan hastaların, müdahale tipi modifiye radike mastektomi olan hastalara göre 1.55 kat ($1/(e^{-0.434})=1.55$) hastalığın nüksetme riskinin daha fazla riskli olduğu söylenebilmektedir. Yaş değişkenindeki 1 birimlik artış, hastalığın nüksetme riskini 1.53 birim artırır. Tümör çapı 2 cm-5 cm arası olan hastaların, tümör çapı 2 cm'den küçük hastalara göre 3.060 ($e^{1.118}=3.060$) kat hastalığın nüksetme riskinin daha fazla riskli olduğu söylenebilmektedir.

CRM'nin temel varsayımı OT olduğundan, CRM'nin kullanılabilmesi için bu varsayımın test edilmesi gerekmektedir. Bu varsayımın test edilmesi için çoğunlukla Schoenfeld artıkları yöntemi kullanılmaktadır (Kara, 2017). Bireylerin başarısızlık sürelerinin rankı ile Schoenfeld artıkları arasında korelasyon testi yapılmış ve Çizelge 4.4'de sonuçlar verilmiştir.

Çizelge 4.4. Meme kanseri verisi için Schoenfeld artıkları yöntemi sonuçları

Değişken	r*	Ki-kare	sd	p değeri
Yaş	0.098	0.85	1	0.357
Menapoz	0.067	0.38	1	0.539
Tümör çapı				
1: 2-5 cm arası	-0.164	2.59	1	0.107
2: 5 cm'den büyük	0.004	0.00	1	0.967
Aksiller lenf nodu				
1: 1-3 adet	0.114	1.33	1	0.248
2: 4'ten fazla	0.068	0.46	1	0.495
Radyoterapi	0.092	0.70	1	0.402
Toksidasyon	-0.028	0.07	1	0.787
Müdahale tipi				
1: Modifiye radike mastektomi	0.082	0.65	1	0.419
2: Radike mastektomi	0.129	1.34	1	0.247
Hastalığın evresi	0.028	0.08	1	0.774
Tedavi türü	0.169	2.44	1	0.118

* r: Schoenfeld artıkları ile logt arasındaki Pearson çarpım-moment korelasyon katsayısı

Çizelge 4.4'de tüm değişkenler için $p > 0.05$ olduğundan OT varsayımının sağlandığı yorumu yapılabilmektedir. Buna göre, CRM'nin bu veri için kullanılmasının uygun olduğu sonucuna varılmıştır.

İkinci olarak klasik CRM, sadece sürekli yaş değişkeni için kurulmuştur. Elde edilen sonuçlara göre p değeri 0.000 olduğundan modelin %95 güven düzeyinde anlamlı olduğu söylenebilmektedir. CRM için -2LogL değeri=-584.33, AIC= 586.330 ve BIC= 589.150 olarak elde edilmiş ve sonuçlar Çizelge 4.5'de verilmiştir.

Yaş değişkenindeki 1 birimlik artışın, hastalığın nüksetme riskini 1.037 ($e^{0.036}=1.037$) birim arttırdığı görülmüştür.

Çizelge 4.5. Meme kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Yaş	0.036	0.010	0.000	1.037	1.026, 1.047

Üçüncü olarak, kategorik yaş değişkeni için tek değişkenli klasik CRM kurulmuştur. Model anlamlılığına ilişkin değeri 0.002 olduğundan, %95 güven düzeyinde modelin anlamlı olduğu sonucuna ulaşılmıştır. Bu modele ait AIC= 588.259 ve BIC= 596.720 olarak elde edilmiş ve Çizelge 4.6'da CRM sonuçları verilmiştir.

Çizelge 4.6. Meme kanseri veri kümesi için yaş değişkeni kategorik kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Kategorik yaş			0.001		
41-50 yaş	-1.401	0.443	0.346	0.246	0.158, 0.383
51- 60 yaş	-1.094	0.571	0.189	0.335	0.189, 0.592
61+ yaş	-0.932	1.540	0.000	0.394	0.084, 1.836

*40 yaş ve altı referans kategorisi olarak alınmıştır.

61 yaş ve üzeri yaş grubuna ait p değeri anlamlı ($p=0.000$) bulunmuş ve kategorik yaş değişkeni de anlamlı çıkmıştır ($p=0.001$). 40 yaş ve altında olan hastaların, 61 yaş ve üzerinde olan hastalara göre $1/(e^{-0.932}) = 2.539$ kat hastalığın nüksetme riskinin daha fazla riskli olduğu %95 güven düzeyinde söylenebilmektedir.

Klasik CRM için son olarak modele tüm değişkenler dahil edilerek ve yaş değişkeni kategorik kabul edilerek kurulmuş ve %95 güven düzeyinde modelin anlamlı olduğu görülmüştür ($p=0.0002$). Bu model için AIC=583.640 ve BIC=617.484 olarak elde edilmiş ve Çizelge 4.7'de CRM sonuçları verilmiştir.

Çizelge 4.7. Meme kanseri veri kümesi için modele tüm değişkenler dahil edildiğinde yaş değişkeni kategorik kabul edilerek oluşturulan klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Menapoz	0.268	0.342	0.549	1.308	0.928, 1.840
Tümör çapı					
1: 2-5 cm arası	1.050	0.134	0.007	2.858	2.499, 3.267
2: 5 cm'den büyük	0.222	0.139	0.009	1.249	1.086, 1.434
Aksiller lenf nodu					
1: 1-3 adet	-0.233	0.178	0.076	0.792	0.662, 0.946
2: 4'ten fazla	-0.778	0.516	0.570	0.459	0.274, 0.769
Radyo terapi	0.434	0.169	0.098	1.544	1.303, 1.827
Toksidasyon	0.278	0.204	0.303	1.321	1.076, 1.619
Müdahale tipi					
1: Modifiye radike mastektomi	-0.466	1.163	0.016	0.627	0.196, 2.007
2: Radike mastektomi	0.547	0.666	0.264	1.727	0.887, 3.363
Hastalığın evresi	-0.275	0.370	0.330	0.760	0.524, 1.099
Tedavi türü	-0.068	0.322	0.822	0.934	0.677, 1.289
Kategorik yaş					
2:41 yaş ve 50 yaş arası	-1.193	0.519	0.381	0.303	0.180, 0.509
3:51 yaş ve 60 yaş arası	-0.866	0.786	0.664	0.421	0.191, 0.923
4:61 yaş ve üzerindeki	-0.931	2.091	0.060	0.394	0.048, 3.189

* Değişkenler için referans kategorisi; yaş için 1:"40 yaş ve altı", menopoz için "0: Girmiş", tümör çapı için "0:0-2 cm'den küçük", aksiller lenf nodu için "0: yok", radyoterapi için "0: Almış", toksidasyon için "0: Herhangi bir şey yok", müdahale tipi değişkeni için "0: Sabit mastektomi", hastalığın evresi değişkeni için "0: Evre 1" olarak alınmıştır.

Çizelge 4.7'de açıklayıcı değişkenler için p değerleri incelendiğinde, modele tüm değişkenler dahil edilerek ve yaş değişkeni kategorik kabul edilerek klasik CRM uygulandığında ise yaş değişkeninin anlamlı çıkmadığı görülmektedir.

Tümör çapı 2-5 cm ($p=0.007$), tümör çapı 5 cm'den büyük ($p=0.009$) ve müdahale tipi modifiye radike mastektomi ($p=0.016$) değişkenlerinin, hastalığın nüksetme riski

üzerinde etkili olan anlamlı risk faktörleri olduğu görülmüştür. Müdahale tipi sabit mastektomi olan hastaların, müdahale tipi modifiye radikale mastektomi olan hastalara göre $1.593 (1/e^{-0.466}) = 1.593$ kat hastalığın nüksetme riskinin daha fazla riskli olduğu %95 güven düzeyinde söylenebilmektedir. Tümör çapı 2-5 cm olan hastaların, tümör çapı 2 cm'den küçük hastalara göre $2.858 (e^{1.050} = 2.858)$ kat hastalığın nüksetme riskinin daha fazla riskli olduğu söylenebilmektedir. Tümör çapı 5 cm'den büyük olan hastaların, tümör çapı 2 cm'den küçük hastalara göre $1.249 (e^{0.222} = 1.249)$ kat hastalığın nüksetme riskinin daha fazla riskli olduğu söylenebilmektedir.

4.1.3. Meme Kanseri Veri Kümesi için Kesirli Polinomlu Cox Regresyon Modeli Sonuçları

Yaş değişkeni ile bağımlı değişken arasında doğrusal bir ilişki olup olmadığını görebilmek, değilse de bu ilişkinin fonksiyonel biçimini elde edebilmek için kesirli polinomlu CRM kullanılmıştır. Bu modele ait kapalı test seçim yöntemi ile elde edilen sonuçlar Çizelge 4.8' de verilmiştir.

Çizelge 4.8. Meme kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları

Model	Sapma	p derece	Adım
KP2	583.940	-2,3	1
KP1	584.297	2	2
Doğrusal	584.331	1	3
Yokluk	596.749	-	-

Karşılaştırma	Sapma farkı	p değeri
KP2 ile yokluk	12.809	0.012<0.05
KP2 ile doğrusal	0.391	0.942>0.05
KP2 ile KP1	0.357	0.357>0.05

KP2 ile yokluk modeli karşılaştırıldığında, yaş değişkeninin anlamlı olduğu görülmüştür. KP2 ile doğrusal model karşılaştırıldığında yaş değişkeni ile bağımlı değişken arasında doğrusal bir ilişki olduğu görülmüştür. KP2 ile KP1 karşılaştırıldığında ise test anlamlı çıkmamaktadır bu yüzden en iyi modelin KP1 modeli olduğu söylenebilir. Tablodaki KP2 ve KP1'in sapma değerleri karşılaştırıldığında da en iyi kesirli polinomlu

CRM olarak derecesi 2, sapması 584.297 deęeriyle KP2'nin sapma deęerinden yüksek olmasından dolayı KP1 modeli olduęunu söyleyebiliriz. Bu sonuçlara göre klasik CRM'ye ait sonuçların kullanılması uygundur. Ayrıca elde edilen sonuçlara bakıldığında, kesirli polinomlu CRM'nin veri için uygun olmadığı elde edilen güven aralıklarından da görülebilmektedir.

Kesirli polinomlu CRM bu veri kümesi için uygun olmamakla birlikte, modele ait sonuçlar Çizelge 4.9'da verilmiştir. KP2 modelleri içinde en uygun model Çizelge 4.8'de de görüldüğü gibi $p_1=-2$ ve $p_2=3$ olarak elde edilmiştir. Yani, $KP(2) = \beta_1x^{-2} + \beta_2x^3$ 'dür. Sonuçlara bakıldığında $p>0.05$ olduğu için yaş deęişkeninin anlamlı olmadığı buradan da görülmektedir.

Çizelge 4.9. Meme kanseri veri kümesi için tek deęişkenli kesirli polinomlu CRM sonuçları

Deęişken	$\hat{\beta}$	Standart hata	p deęeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Yaş_1	-7.291	9.792	0.457	0.0006	3.81e-8, 12.194
Yaş_2	0.002	0.001	0.159	1.002	1.001, 1.003

İkinci olarak veri kümesine sürekli yaş deęişkeninin yer aldığı kesirli polinomlu çok deęişkenli CRM uygulanmıştır. Bu modele ait kapalı test seçim yöntemi ile elde edilen sonuçlar Çizelge 4.10'da verilmiştir.

KP2 ile yokluk modeli karşılaştırıldığında, yaş deęişkeninin anlamlı olmadığı görülmektedir. KP2 ile doğrusal model karşılaştırıldığında yaş deęişkeninin bağımlı deęişken ile doğrusal bir ilişkisi olduğu görülmüştür. KP2 ile KP1 karşılaştırıldığında ise test anlamlı çıkmamaktadır bu yüzden en iyi modelin KP1 modeli olduğu söylenebilir. Tablodaki KP2 ve KP1'in sapma deęerleri karşılaştırıldığında da en iyi kesirli polinomlu CRM olarak derecesi 1, sapması 559.641 deęeriyle KP2'nin sapma deęerinden yüksek olmasından dolayı KP1 modeli olduęunu söyleyebiliriz.

Çizelge 4.10. Meme kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları

Model	Sapma	p derece	Adım
KP2	559.540	-2,3	1
KP1	559.641	1	2
Doğrusal	559.641	1	3
Yokluk	564.968	-	-

Karşılaştırma	Sapma farkı	p değeri
KP2 ile yokluk	5.427	0.246>0.05
KP2 ile doğrusal	0.100	0.992>0.05
KP2 ile KP1	0.100	0.951>0.05

Çizelge 4.11’da verilen kesirli polinomlu çok değişkenli CRM sonuçlarına bakıldığında yaş değişkeninin anlamlı olmadığı görülmektedir.

Çizelge 4.11. Meme kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için güven aralığı
Yaş_1	-9.653	12.690	0.447	0.00006	1.97e-10, 20.842
Yaş_2	0.002	0.002	0.212	1.002	1.000, 1.004
Menapoz	0.075	0.390	0.846	1.077	0.729, 1.592
Tümör çapı					
1: 2-5 cm arası	-0.954	0.311	0.002	0.385	0.282, 0.525
2: 5 cm’den büyük	-1.109	0.397	0.005	0.329	0.221, 0.490
Aksiller lenf nodu					
1: 1-3 adet	-0.512	0.307	0.096	0.600	0.440, 0.814
2: 4’ten fazla	0.333	0.404	0.410	1.395	0.931, 2.089
Radyoterapi	-0.386	0.245	0.116	0.679	0.532, 0.868
Toksidasyon	-0.197	0.266	0.460	0.821	0.629, 1.071
Müdahale tipi					
1: Modifiye radike mastektomi	1.017	0.419	0.015	2.764	1.818, 4.203
2: Radike mastektomi	0.429	0.415	0.301	1.535	1.014, 2.325
Hastalığın evresi	0.324	0.275	0.238	1.382	1.050, 1.820
Tedavi türü	0.081	0.290	0.779	1.084	0.811, 1.449

* Değişkenler için referans kategorisi; menopoz için “0: Girmiş”, tümör çapı için “0:0-2 cm’den küçük”, aksiller lenf nodu için “0: yok”, radyoterapi için “0: Almış”, toksidasyon için “0: Herhangi bir şey yok”, müdahale tipi değişkeni için “0: Sabit mastektomi”, hastalığın evresi değişkeni için “0: Evre 1” olarak alınmıştır.

Meme kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları elde edildikten sonra, OT varsayımı yeniden incelenmiş ve elde edilen sonuçlar Çizelge 4.12'de verilmiştir. Çizelge 4.12'den, $p > 0.05$ olduğundan OT varsayımının sağlandığı görülmüştür.

Çizelge 4.12. Meme kanseri veri kümesi için kesirli polinomlu çok değişkenli CRM'ye ait Schoenfeld artıkları yöntemi sonuçları

Değişken	r*	Ki-kare	sd	p değeri
Yaş_1	0.042	0.12	1	0.725
Yaş_2	0.117	1.19	1	0.274
Menapoz	0.047	0.18	1	0.673
Tümör çapı				
1: 2-5 cm arası	-0.166	2.62	1	0.105
2: 5 cm'den büyük	0.003	0.00	1	0.971
Aksiller lenf nodu				
1: 1-3 adet	0.086	0.81	1	0.367
2: 4'ten fazla	0.051	0.27	1	0.604
Radyoterapi	0.092	0.70	1	0.402
Toksidasyon	-0.024	0.05	1	0.821
Müdahale tipi				
1: Modifiye radike mastektomi	0.084	0.69	1	0.406
2: Radike mastektomi	0.138	1.58	1	0.208
Hastalığın evresi	0.021	0.05	1	0.830
Tedavi türü	0.173	2.60	1	0.106

* r: Schoenfeld artıkları ile logt arasındaki Pearson çarpım-moment korelasyon katsayısı

Meme kanseri veri kümesi için tek değişkenli ve çok değişkenli kesirli polinomlu Cox regresyon modelleri uygulanmış, ancak her ikisinin sonucunda da doğrusal modelin veri için uygun olduğu görülmüştür. Buna göre meme kanseri veri kümesi için yaş değişkeninin model üzerinde doğrusal bir etkisi olduğunu ve klasik Cox regresyon modellerinin kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

4.2. Prostat Kanseri Veri Kümesi

Bu bölümde, ilk olarak Bayer ve Green (1980) tarafından incelenen ve prostat kanseri hastalığının 3. ve 4. aşamalarında olan ve 5 yıl izlenen 475 hastaya ait bir veri kümesi kullanılmıştır 475 hastadan 338'inde (%71.1) başarısızlık ve 137'sinde (%28.8) durdurma gözlenmiştir. Başarısızlık hastaların ölümü olarak ele alınmıştır.

Çalışmada yaş, tedavi türü ve hastalığın evresi açıklayıcı değişkenler olarak ele alınmıştır. Yaş değişkeni sürekli bir değişken, diğer değişkenler ise kategorik değişken olarak modellere dahil edilmiştir. Hastaların yaş ortalaması 71.56 ± 0.318 (min=48, max=89F), ortanca yaşam süresi 33.5 ± 2.137 ay ve 5 yıllık yaşam olasılığı %25.14 olarak bulunmuştur. Değişkenlere ait bilgiler Çizelge 4.13' de verilmiştir.

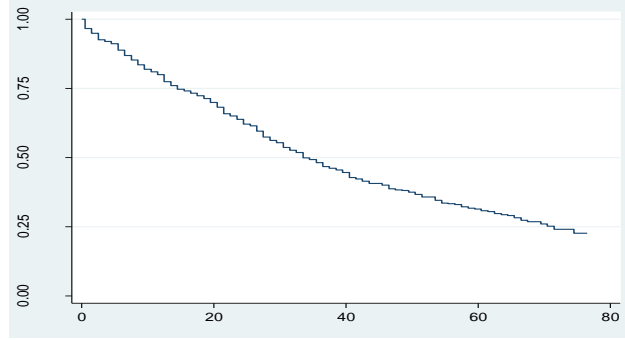
Çizelge 4.13. Prostat kanseri veri kümesi için kullanılan değişkenler ve düzeyleri

Değişken	Değişken Düzeyleri	n (%)	Başarısız Gözlemlerin Sayısı	Durdurulmuş Gözlemlerin Sayısı
Kategorik yaş*	1: 65 yaş ve altı	80 (%35)	52	28
	2: 66-71 yaş	97 (%35.1)	63	34
	3: 72-76 yaş	204 (%27.9)	147	57
	4: 77-80	76 (%22.4)	59	17
	5: 81 yaş ve üzeri	18 (%5.6)	17	1
Hastalığın evresi	Evre 3:Yerel uzak metastaz belirtisi olmayan	272 (%33.5)	181	91
	Evre 4:Uzak metastaz belirtisi olan	203 (%22.7)	157	46
Tedavi türü	0: Placebo	238 (%33.6)	158	80
	1: TRT E	237 (%24.1)	180	57

*: Yaş değişkeni kategorik olarak alındığında kullanılan sınıflandırma

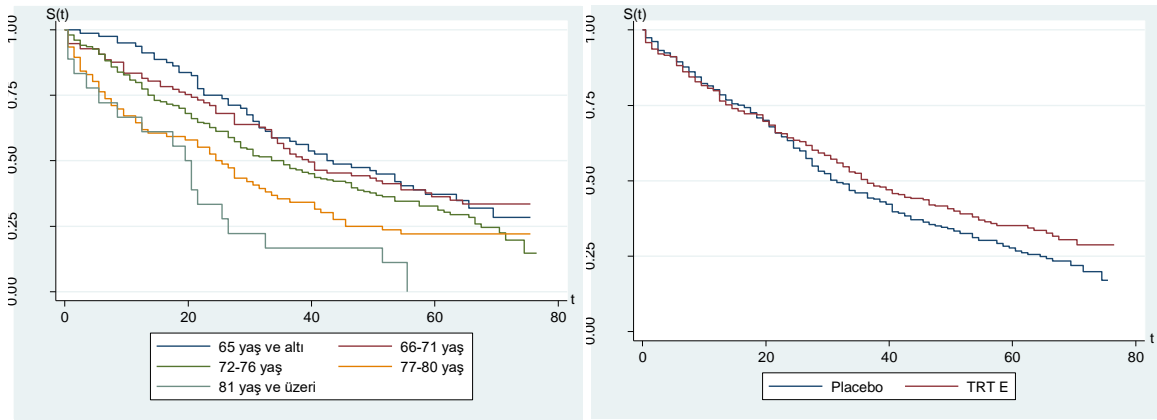
4.2.1. Prostat Kanseri Veri Kümesi için Parametrik Olmayan Yaşam Çözümlemesi Sonuçları

Şekil 4.3'de, prostat kanseri veri kümesine ait KM yaşam eğrisi verilmiştir.



Şekil 4.3. Prostat kanseri verisi için KM yaşam eğrisi

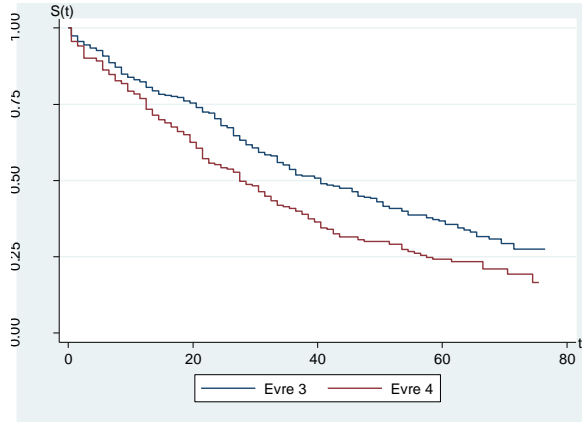
Çalışmada kullanılan her bir açıklayıcı değişken için KM yaşam eğrileri ise Şekil 4.4'de verilmiştir.



(a) Yaş

(b) Tedavi

Şekil 4.4. Prostat kanseri verisi için açıklayıcı değişkenlere ait KM eğrileri



(c) Evre

Şekil 4.4. Prostat kanseri verisi için açıklayıcı değişkenlere ait KM eğrileri (devam ediyor)

Şekil 4.4 incelendiğinde, evre ve tedavi değişkenlerinde değişken düzeylerinin yaşam eğrileri birbirine çok yakın olarak elde edilmiştir. Yaş değişkeni için düzeyleri arasında farklılığın daha fazla olduğu görülmektedir. Yaşı 65 yaş ve altı, 66-71 yaş arası ve 72-76 yaş arası olan hastaların yaşam olasılıklarının birbirine yakın olduğu ve 77-80 yaş arası ile 81 yaş ve üzerinde olan hastaların yaşam olasılıklarından daha yüksek olduğu görülmektedir.

Prostat kanseri veri kümesi için log-rank testi sonuçları Çizelge 4.14'deki gibi elde edilmiştir.

Çizelge 4.14. Prostat kanseri verisi için log-rank testi sonuçları

Değişken	Ki-kare	p değeri
Yaş	24.176	0.000
Evre	10.107	0.001
Tedavi türü	2.663	0.103

Çizelge 4.14'e göre, yaşam olasılıkları açısından yaş ($p=0.0000$) ve evre ($p=0.001$) değişkenlerinin tüm düzeyleri arasında anlamlı bir farklılık olduğu %95 güven düzeyinde söylenebilir.

4.2.2. Prostat Kanseri Veri Kümesi için Klasik Cox Regresyon Modeli Sonuçları

Öncelikle yaş sürekli bir değişken olarak ele alınmış ve çok değişkenli CRM veri kümesine uygulanmıştır. Modelin uygunluğuna ilişkin p değeri 0.000 olarak elde edildiğinden, modelin %95 güven düzeyinde anlamlı olduğu sonucuna ulaşılmıştır. Bu modele ait AIC= 3781.429 ve BIC= 3793.919 olarak elde edilmiştir. Kurulan CRM'ye ait sonuçlar Çizelge 4.15'de verilmiştir.

Çizelge 4.15. Prostat kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan çok değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için %95 güven aralıkları
Yaş	0.031	0.009	0.000	1.031	1.023, 1.039
Evre	0.357	0.109	0.001	1.429	1.281, 1.593
Tedavi türü	-0.163	0.109	0.135	0.849	0.761, 0.947

* Değişkenler için referans düzeyi: hastalığın evresi için "Evre 3", tedavi türü için "0: Placebo" olarak alınmıştır.

Tüm değişkenlerin birinci düzeyi referans kategorisi olarak alınmıştır. Çizelge 4.15'den görüldüğü gibi yaş ($p=0.000$) ve evre ($p=0.001$) değişkenleri önemli risk faktörleri olarak belirlenmiştir. CRM'den elde edilen sonuçlara göre, yaş değişkenindeki 1 birimlik artış ölüm riskini 1.031 ($e^{0.031}=1.031$) birim arttırmaktadır. Evre 4 (uzak metastaz belirtisi olan) aşamasında olan hastaların, evre 3 (yerel uzak metastaz belirtisi olmayan) aşamasında olan hastalara göre 1.429 ($e^{0.357}=1.429$) kat ölüm riskinin daha fazla riskli olduğu görülmüştür.

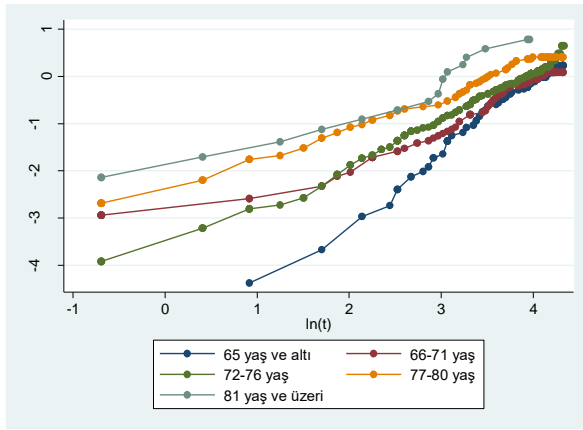
Çalışmada OT varsayımı Schoenfeld artıkları yöntemi ile incelenmiş ve Çizelge 4.16'da sonuçlar verilmiştir. Yaş değişkeni için $p<0.05$ olduğundan, varsayımının sağlanmadığı sonucuna ulaşılmıştır.

Çizelge 4.16. Prostat kanseri veri kümesi için Schoenfeld artıkları yöntemi sonuçları

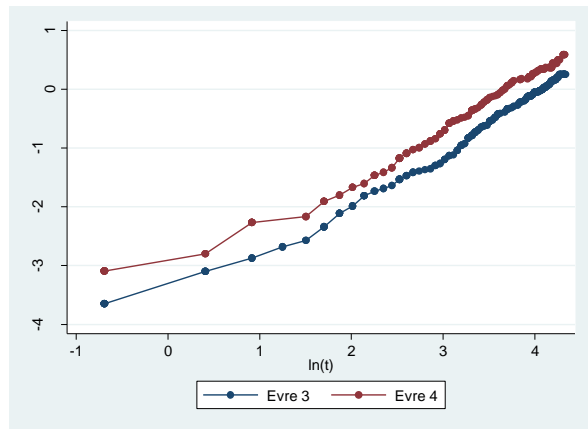
Değişken	r^*	Ki-kare	Sd	p değeri
Yaş	-0.150	9.76	1	0.001
Evre	-0.043	0.65	1	0.420
Tedavi türü	-0.077	2.09	1	0.148

* r : Schoenfeld artıkları ile logt arasındaki Pearson çarpım-moment korelasyon katsayısı

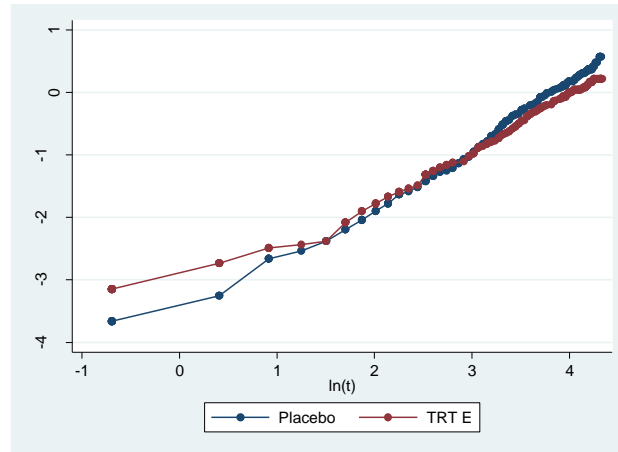
Çalışmada kullanılan prostat kanseri veri kümesine ait OT varsayımı grafiksel yöntemle de incelenmiş olup, sonuçlar Şekil 4.5'de verilmiştir.



(a) Yaş



(b) Evre



(c) Tedavi

Şekil 4.5. Değişkenlerin log-log grafikleri

Şekil 4.5’de değişkenlerin $\ln(t)$ ’ye karşı $-\ln(-\ln S(t))$ grafikleri çizilmiştir. Yaş, tedavi değişkenlerinin düzeyleri çakıştığı için OT varsayımının sağlanmadığı görülmektedir.

Klasik CRM için, ikinci olarak model sadece sürekli yaş değişkeni için kurulmuş ve %95 güven düzeyinde model anlamlı bulunmuştur ($p=0.0001$). Bu modele ait $AIC=3789.696$ ve $BIC= 3793.859$ olarak elde edilmiş ve Çizelge 4.17’de sonuçlar verilmiştir

Çizelge 4.17. Prostat kanseri veri kümesi için yaş değişkeni sürekli kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	$\text{Exp}(\hat{\beta})$	$\text{Exp}(\hat{\beta})$ için %95 güven aralıkları
Yaş	0.033	0.089	0.000	1.030	0.945, 1.129

Bu model için yaş değişkeninin anlamlı olduğu görülmüştür. Yaş değişkenindeki 1 birimlik artış, ölüm riskini 1.030 ($e^{0.033}=1.030$) birim arttırmaktadır.

Üçüncü olarak, klasik CRM için model sadece kategorik yaş değişkeni için kurulmuş ve model %95 güven düzeyinde anlamlı bulunmuştur ($p=0.000$). Bu CRM için $AIC= 3790.376$ ve $BIC= 3807.029$ olarak elde edilmiştir. Meme kanseri verisi için CRM sonuçları Çizelge 4.18’de verilmiştir.

Modelde sadece yaş değişkeni alınarak ve 65 yaş ve altı, 66 yaş ve 71 yaş arası, 72 yaş ve 76 yaş arası, 77 yaş ve 80 yaş arası, 81 yaş ve üzerindeki gruplar olarak kategorik değişken olarak kullanılarak klasik CRM uygulandığında 77-80 yaş grubu ve 81 yaş ve üzeri yaş grubu anlamlı çıkmaktadır.

77-80 yaş arasındaki hastaların, 65 yaş ve altındaki hastalara göre 1.729 ($e^{0.548}=1.729$) kat ölüm riskinin daha fazla riskli olduğu söylenebilmektedir. 81 yaş ve üzeri yaş arasındaki hastaların, 65 yaş ve altındaki hastalara göre 2.989 ($e^{1.095}=2.989$) kat ölüm riskinin daha fazla riskli olduğu söylenebilmektedir.

Çizelge 4.18. Prostat kanseri veri kümesi için yaş değişkeni kategorik kabul edilerek oluşturulan tek değişkenli klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için %95 güven aralıkları
Kategorik yaş			0.000		
66- 71 yaş	0.054	0.197	0.773	1.055	0.866, 1.285
72-76 yaş	0.266	0.210	0.099	1.304	1.057, 1.609
77-80 yaş	0.548	0.329	0.004	1.729	1.244, 2.403
81 yaş ve üzeri	1.095	0.840	0.000	2.989	1.290, 6.924

*65 yaş ve altı referans kategorisi olarak alınmıştır.

Klasik CRM için son olarak modele tüm değişkenler dahil edilerek ve yaş değişkeni kategorik kabul edilerek kurulmuş ve model %95 güven düzeyinde anlamlı bulunmuştur (p=0.000). Bu model için AIC= 3782.703 ve BIC= 3807.683 olarak elde edilmiştir. Prostat kanseri verisi için CRM sonuçları Çizelge 4.19’da verilmiştir.

Çizelge 4.19. Modele tüm değişkenler dahil edildiğinde yaş değişkeni kategorik kabul edilerek oluşturulan klasik CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için %95 güven aralıkları
Kategorik yaş					
66-71 yaş	0.063	0.187	0.736	1.065	0.883, 1.284
72-76 yaş	0.250	0.161	0.121	1.284	1.093, 1.508
77-80 yaş	0.501	0.193	0.090	1.650	1.360, 2.001
81 yaş ve üzeri	1.113	0.284	0.000	3.043	2.291, 4.043
Evre	0.334	0.109	0.002	1.396	1.252, 1.557
Tedavi türü	-0.193	0.824	0.090	0.849	0.361, 1.879

* Değişkenler için referans düzeyi: yaş için “65 yaş ve altı”, hastalığın evresi için “Evre 3”, tedavi türü için “0: Placebo” olarak alınmıştır.

Çizelge 4.19’a göre 81 yaş ve üzeri yaş grubunun anlamlı çıktığı görülmektedir. Ölüm riski üzerinde evre değişkeninin anlamlı bir etkisi olduğu görülmüştür.

Evre 4 (uzak metastaz belirtisi olan) aşamasında olan hastaların, evre 3 (yerel uzak metastaz belirtisi olmayan) aşamasında olan hastalara göre 1.396 ($e^{0.334}=1.396$) kat ölüm riskinin daha fazla riskli olduğu söylenebilir. 81 yaş ve üzerindeki hastaların, 65 yaş ve altındaki hastalara göre 3.043 kat ölüm riskinin daha fazla riskli olduğu görülmektedir.

4.2.3. Prostat Kanseri Veri Kümesi için Kesirli Polinomlu Cox Regresyon Modeli Sonuçları

Öncelikle kesirli polinomlu CRM sadece yaş değişkeni kullanılarak kurulmuş ve kapalı test seçim yöntemi ile elde edilen sonuçlar Çizelge 4.20' de verilmiştir.

Çizelge 4.20. Prostat kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları

Model	Sapma	p derece	Adım
KP2	3778.586	3,3	1
KP1	3784.375	3	2
Doğrusal	3787.696	1	3
Yokluk	3802.495	-	-

Karşılaştırma	Sapma farkı	p değeri
KP2 ile yokluk	23.909	0.000<0.05
KP2 ile doğrusal	9.110	0.028<0.05
KP2 ile KP1	5.790	0.055>0.05

KP2 ile yokluk modeli karşılaştırıldığında, yaş değişkeninin anlamlı olduğu ve KP2 ile doğrusal model karşılaştırıldığında yaş değişkeninin bağımlı değişken ile doğrusal olmayan bir ilişkisi olduğu görülmüştür. KP2 ile KP1 karşılaştırıldığında ise test anlamlı çıkmamaktadır bu yüzden en iyi modelin KP1 modeli olduğu söylenebilir. Tablodaki KP2 ve KP1'in sapma değerleri karşılaştırıldığında da en iyi kesirli polinomlu CRM olarak derecesi 3, sapması 3784.375 değeriyle KP2'nin sapma değerinden yüksek olmasından dolayı KP1 modeli olduğunu söyleyebiliriz.

Tek değişkenli kesirli polinomlu CRM'ye ait sonuçlar Çizelge 4.21'de verilmiştir.

Çizelge 4.21. Prostat kanseri veri kümesi için tek değişkenli kesirli polinomlu CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için %95 güven aralıkları
Yaş_1	-0.038	0.016	0.017	0.962	0.947; 0.978
Yaş_2	0.017	0.007	0.011	1.017	1.010; 1.024

KP2 modelleri için de en uygun model $p_1=p_2=3$ olarak elde edilmiştir. Yani, $KP(2) = \beta_1 x^3 + \beta_2 x^3 \ln x$ 'dir.

İkinci olarak, veri kümesine çok değişkenli kesirli polinomlu uygulanmış ve kapalı test seçim yöntemi ile elde edilen sonuçlar Çizelge 4.22' deki gibi elde edilmiştir.

Çizelge 4.22. Prostat kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM için kapalı test seçim yöntemi sonuçları

Model	Sapma	p derece	Adım
KP2	3766.934	3,3	1
KP1	3772.374	3	2
Doğrusal	3775.429	1	3
Yokluk	3789.494	-	-

Karşılaştırma	Sapma farkı	p değeri
KP2 ile yokluk	22.560	0.000<0.05
KP2 ile doğrusal	8.495	0.037<0.05
KP2 ile KP1	5.440	0.066>0.05

KP2 ile yokluk modeli karşılaştırıldığında, yaş değişkeninin anlamlı olduğu ve KP2 ile doğrusal model karşılaştırıldığında yaş değişkeni ile bağımlı değişken arasında doğrusal olmayan bir ilişkisi olduğu görülmüştür. KP2 ile KP1 karşılaştırıldığında ise test anlamlı çıkmamaktadır bu yüzden en iyi modelin KP1 modeli olduğu söylenebilir. Tablodaki KP2 ve KP1'in sapma değerleri karşılaştırıldığında da en iyi kesirli polinomlu CRM olarak derecesi 3, sapması 3772.374 değeriyle KP2'nin sapma değerinden yüksek olmasından dolayı KP1 modeli olduğunu söyleyebiliriz.

Prostat kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları Çizelge 4.23'deki gibidir.

Çizelge 4.23. Prostat kanseri veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları

Değişken	$\hat{\beta}$	Standart hata	p değeri	Exp($\hat{\beta}$)	Exp($\hat{\beta}$) için %95 güven aralıkları
Yaş_1	-0.037	0.016	0.021	0.963	0.948, 0.979
Yaş_2	0.017	0.007	0.014	1.017	1.010, 1.024
Tedavi türü	-0.194	0.110	0.078	0.823	0.737, 0.919
Evre	0.332	0.109	0.003	1.393	1.249, 1.554

*Evre değişkeni için Evre 3 referans kategorisi olarak alınmıştır.
Tedavi türü değişkeni için 0: Placebo referans kategorisi olarak alınmıştır.

KP2 modelleri için de en uygun model $p_1=p_2=3$ olarak elde edilmiştir. Yani, $KP(2) = \beta_1 x^3 + \beta_2 x^3 \ln x$ 'dir.

Çok değişkenli kesirli polinomlu CRM modelinde yaş ve evre değişkenlerinin anlamlı çıktığı görülmektedir. Evre 4 (uzak metastaz belirtisi olan) olan hastaların, evre 3 (yerel uzak metastaz belirtisi olmayan) aşamasında olan hastalara göre 1.393 ($e^{0.332}=1.393$) kat ölü riskinin daha fazla riskli olduğu söylenebilir.

Prostat veri kümesi için çok değişkenli kesirli polinomlu CRM sonuçları elde edildikten sonra, OT varsayımı yeniden incelenmiş ve elde edilen sonuçlar Çizelge 4.24'de verilmiştir. Çizelge 4.24'den, $p > 0.05$ olduğundan OT varsayımının sağlandığı görülmüştür.

Çizelge 4.24. Prostat kanseri veri kümesi için kesirli polinomlu çok değişkenli CRM'ye ait Schoenfeld artıkları yöntemi sonuçları

Değişken	r^*	Ki-kare	sd	p değeri
Yaş_1	-0.063	1.31	1	0.251
Yaş_2	0.060	1.16	1	0.281
Evre	-0.057	1.11	1	0.291
Tedavi türü	-0.071	1.77	1	0.183

* r: Schoenfeld artıkları ile logt arasındaki Pearson çarpım-moment korelasyon katsayısı

Klasik CRM sonuçlarında prostat veri kümesi için OT varsayımı sağlanmadığından CRM sonuçlarının kullanılması ve yorumlanması uygun değildir. Ancak kesirli polinomlu CRM sonrası yaş değişkeninin bağımlı değişken ile doğrusal olmayan bir ilişkiye sahip olduğu belirlendiğinden ve tekrar Schoenfeld artıkları yöntemi sonuçlarına bakıldığında OT varsayımı sağlandığından kesirli polinomlu CRM'nin prostat veri kümesi için kullanılmasının daha uygun olduğu sonucuna varılmıştır.

5. SONUÇ VE ÖNERİLER

Açıklayıcı değişkenlerin yaşam süresi ile ilişkisini incelemek için birçok yaşam modeli literatürde yer almaktadır. Yarı parametrik bir yaşam modeli olan Cox regresyon modeli yaşam çözümlemesinde en çok kullanılan model olma özelliğine sahiptir. Modele dahil edilen açıklayıcı değişkenlerin etkilerinin doğrusal olmaması durumunda kesirli polinomlu Cox regresyon modelinin kullanımı önerilmiştir.

Uygulamalarda çoğunlukla doğrusal olmayan ilişkilerden kaçınmak için açıklayıcı değişkenler kategorik biçime dönüştürülerek modele dahil edilmektedir. Ancak, veriyi kategorik biçime dönüştürürken, değişken düzeylerinin nasıl belirleneceği konusunda da kararsızlıklar ortaya çıktığı için kesirli polinomlu modelleri kullanmak daha uygundur. Sürekli açıklayıcı değişkenin, kategorik biçime dönüştürülmeden modele dahil edilmesi ve bağımlı değişken ile doğrusal ilişkiye sahip olup olmadığının incelenmesi önerilen modelin uygunluğuna karar verebilmek için önemlidir.

Kesirli polinom yöntemi, açıklayıcı bir değişkenin model için önemli olup olmadığını ve fonksiyonel biçiminin belirlenmesine izin veren bir yöntemdir. Kesirli polinomlar doğrusallık sağlanmadığında sürekli değişkenlerin yapısını korumak için kullanılmaktadır. KP1 ve KP2 olmak üzere 2 modelden oluşmaktadır ve toplamda veri kümesi için 44 model denenmektedir. Polinomun derecesi olarak ifade edilen p değerleri $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ kümesindeki değerlerinden oluşmaktadır. En uygun model en yüksek olasılığı veren modeldir. $p=1$ 'e karşı $p \neq 1$ 'in anlamlılık testi doğrusal olmamanın testidir. Eğer test anlamlı değilse doğrusal model kabul edilir, yoksa doğrusal olmayan kesirli polinomlu model tercih edilir.

Bu tez çalışmasında meme kanseri ve prostat kanseri veri kümeleri kullanılmıştır. Modelde yer alan sürekli değişken dikkate alınarak tek değişkenli ve çok değişkenli klasik ve kesirli polinomlu CRM modelleri veri kümelerine uygulanmıştır. Ayrıca değişkenin anlamlılığı konusunda bir farklılık olup olmadığını görebilmek için de ilgili sürekli değişken kategorik biçime dönüştürülerek de klasik CRM'ye dahil edilmiştir.

Meme kanseri verisi için model kurarken ilk olarak, modele tüm değişkenler dahil edilerek sadece yaş değişkenini sürekli kabul ederek klasik Cox regresyon çözümlemesi yapılmıştır ve OT varsayımını sağladığı görülmüştür. Model sonucunda yaş, tümör

çapı, müdahale tipi sabit mastektomi değişkenlerinin anlamlı olduğu görülmüştür. İkinci olarak, sadece sürekli yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümlemesi yapıldığında yaş değişkeninin anlamlı olduğu görülmektedir. Üçüncü olarak, sadece kategorik yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümlemesi yapıldığında ise 61 yaş ve üzerinde yer alan yaş değişkeninin anlamlı olduğu görülmüştür. Dördüncü olarak, modele tüm değişkenler dahil edilerek ve yaş değişkeni kategorik değişken olarak ele alınıp klasik Cox regresyon çözümlemesi yapıldığında yaş değişkeninin anlamlı çıkmadığı ancak tümör çapı ve müdahale tipi sabit mastektomi olan değişkenin anlamlı çıktığı görülmüştür. Kesirli polinomlu CRM incelendiğinde ise, doğrusal olmayan bir ilişkinin olmadığı, bu nedenle de doğrusal modelin kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

Prostat kanseri veri kümesi için model kurarken ilk olarak, modele tüm değişkenler dahil edilerek sadece yaş değişkenini sürekli kabul ederek klasik Cox regresyon çözümlemesi yapılmıştır ve OT varsayımını sağlamadığı görülmüştür. Ancak kesirli polinomlu Cox regresyon modeli ile klasik Cox regresyon modeli sonuçları arasındaki farkı görebilmek için analize devam edilmiştir. Model sonucunda yaş ve evre değişkenlerinin anlamlı olduğu görülmüştür. İkinci olarak, sadece sürekli yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümlemesi yapıldığında yaş değişkeninin anlamlı olduğu görülmektedir. Üçüncü olarak, sadece kategorik yaş değişkeni modele dahil edilerek klasik Cox regresyon çözümlemesi yapıldığında ise 77-80 yaş grubu ve 81 yaş ve üzeri yaş grubu anlamlı çıkmıştır. Dördüncü olarak, modele tüm değişkenler dahil edilerek ve yaş değişkeni kategorik değişken olarak ele alınıp klasik Cox regresyon çözümlemesi yapıldığında 81 yaş ve üzeri yaş grubunun ve evre değişkeninin anlamlı çıktığı görülmüştür.

Yaş değişkeninin bağımlı değişken ile doğrusal ilişkili olup olmadığını görmek için tek değişkenli ve çok değişkenli kesirli polinom CRM uygulanmıştır. Her iki modelde de yaş değişkeninin anlamlı olduğu ve bağımlı değişken ile doğrusal olmayan bir ilişkiye sahip olduğu sonucuna ulaşılmıştır. Ayrıca kesirli polinomlu CRM kurulduktan sonra, OT varsayımı yeniden incelenmiş ve varsayımın bu defa sağlandığı görülmüştür. Buna göre kesirli polinomlu CRM'nin bu veri kümesi için kullanılması ve yorumlanması daha doğrudur.

Kesirli polinomlu Cox regresyon modelinin kullanıma ilişkin ulusal literatürde bir çalışmaya rastlanmamaktadır. Kesirli polinomlu Cox regresyon modeli kategorik biçime dönüştürme aşamasındaki bilgi kaybını engellemesi, doğrusal olmayan ilişkileri ve fonksiyonel biçimini belirlemesi nedeni ile uygulamacılar için oldukça kullanışlıdır. Ayrıca istatistiksel paket programları için de son yıllarda modele ait kodların geliştirilmesi ile birlikte kullanımı kolaylaştırmıştır.

Paket programlar kullanılarak elde edilen kesirli polinomlu modeller, modelde yer alan tek bir değişken için kesirli polinomların kullanımına izin vermektedir. Bundan sonraki çalışmalarda birden fazla sürekli değişken için modelin kullanılmasına izin veren yazılımlar geliştirilebilir.

KAYNAKLAR

- Altman, D.G., Lausen, B., Sauerbrei, W., Schumacher, M., The dangers of using optimal cutpoints in the evaluation of prognostic factors, *Journal of the National Cancer Institute*, 86:11 (1994) 829–835.
- Ambler, G., Royston, P., Sauerbrei, W., The Use of Fractional Polynomials to Model Continuous Risk Variables in Epidemiology, *US National Library of Medicine National Institutes of Health*, 28:5 (1999) 964-74.
- Ambler, G., Royston, P. Fractional Polynomial Model Selection Procedures: Investigation of Type I Error Rate, *Journal of Statistical Computation and Simulation*, 69:1 (2001) 89-108.
- Anonim, Cox Regression, https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox_Regression.pdf (Erişim tarihi:06.04.2019).
- Ata Tutkun, N., Yaşam Çözümlemesi'nde Orantısız Hazard Modelleri, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2005.
- Ata Tutkun, N., Kesirli Polinomlu Lojistik Regresyon Modeli, *Journal of Statisticians: Statistics and Actuarial Sciences*, 8:2 (2015) 27-35.
- Austin, P.C., Park, L.Y., Juurlink W., DN. Juurlink, Using Fractional Polynomials to Model the Effect of Cumulative Duration of Exposure on Outcomes: Applications to Cohort and Nested Case, Control designs *pharmacoepidemiology and drug safety*, 23 (2014) 819–829.
- Bayer, D.P., Green, S.B., The Choice of Treatment for Cancer Patients Based on Covariate Information: Applications to Prostate Cancer, *Bulletin du Cancer* 67, (1980) 477-490.
- Bellera, C., MacGrogan, G., Debled, M., De Lara, C., Brouste, V., Mathoulin Pélissier, S., Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer, *BMC Medical Research Methodology*, 10:20 (2010) 1471-2288.
- Berger, U., Gerein, P., Ulm, K., Schäfer, J. (2000): On the use of Fractional Polynomials in Dynamic Cox Models. Collaborative Research Center 386, Discussion Paper 207, 2000.
- Berger, U., Schafer J., Ulm K., Dynamic Cox Modelling Based on Fractional Polynomials: Time-Variations in Gastric Cancer Prognosis, *Statistics in Medicine*, 22 (2003) 1163–1180.
- Binder, H., Sauerbrei, W., Royston, P., Comparison Between Splines and Fractional Polynomials for Multivariable Model Building with Continuous Covariates: A Simulation Study With Continuous Response, *Statistics in Medicine*, 32:13 (2012) 2262-2277.
- Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G., Survival Analysis Part II: Multivariate Data Analysis – An Introduction to Concepts and Methods, *British Journal of Cancer*, 89 (2003) 431-436.

- Buchholz, A., Sauerbrei, W., Comparison of Procedures to Assess Non-Linear and Time-Varying Effects in Multivariable Models For Survival Data, *Biometrical Journal*, 53 (2011) 308–331.
- Buchholz, A., Sauerbrei, W., Royston, P., A Measure for Assessing Functions of Time-Varying Effects in Survival Analysis, *Open Journal of Statistics*, 4 (2014) 977-998.
- Collett, D., *Modelling Survival Data in Medical Research*, Chapman & Hall, First Edition, London, Chapter 2, (1994).
- Cox, D.R., Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, 34:2, (1972) 187-220.
- Despa, S., What is Survival Analysis?, <http://www.cscu.cornell.edu/news/stanews/stnews78.pdf> (Erişim tarihi:24.03.2019).
- Dupont, W.D., Review of Multivariable Model-building: A Pragmatic Approach to Regression Analysis, *The Stata Journal*, 10:2 (2010) 297–302.
- Elhan, A.H., Demir, E., Hayme, S., Öztuna D., Klinik Denemelerde Kesirli Polinomların İki Kategorili Lojistik Regresyon Modelinde Kullanımı, XVIII. Ulusal ve I. Uluslararası Biyoistatistik Kongresi Bildiri Özetleri Kitabı, Antalya, 26-29 Ekim 2016, Antalya, 2016, p.131.
- Erdoğan, A., Orantılı Hazard Modeli, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, 1993.
- Ganji, M., Yekaninejad, M.S., Yaseri, M., Comparison Between Fractional Polynomials, Spline Smoothing, and Multiple Logistic Regression Models in The Study of Associated Hypertension Risk Factors, *J Biostat Epidemiol*, 3:2 (2017) 49-59.
- Hosmer, D.W., Lemeshow, S., *Applied Survival Analysis*, Second Edition, John Wiley & Sons, Inc, New York, Chapter 3, 1999.
- Jansen, JP., Network Meta-Analysis of Survival Data with Fractional Polynomials, *BMC Medical Research Methodology*, 11:61 (2011) 1471-2288.
- Kalbfleisch, J., Prentice, R.L., *The Statistical Analysis of Failure Time Data*, Second Edition, John Wiley & Sons, Inc, New York, Chapter 4, 1980.
- Kara, P., Yaşam Çözümlemesinde İyileşme Modelleri, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2017.
- Karunanayake, C., Rennie, DC., Pahwa, P., Chen, Y. ve Dosman, J.A., Relationship Between Lung Function and Hypertension Among Rural Canadians Using Fractional Polynomials, *Sri Lankan Journal of Applied Statistics*, 2012, 41–61.
- Kleinbaum, D.G., Klein, M. *Survival Analysis, A Self Learning Text*, Second Edition, Springer, New York, Chapter 4, 2005.
- Klein, J.P., Moeschberger, M.L., *Survival Analysis Techniques for Censored and Truncated Data*, Second Edition, New York, Chapter 9, 1997.
- Lausen, B., Schumacher, M., 1996. Evaluating the effect of optimized cut-off values in the assessment of prognostic factors. *Computational Statistics Data Analysis* 21:3 (1996) 307–326.
- Marinda, E., Use of Fractional Polynomials in Medical Research, *South African Centre for Epidemiological Modelling and Analysis (SACEMA)*, 9 (2011), 2-4.

- Miller, R., Siegmund, D., Maximally Selected Chi-square Statistics, *Biometrics* 38:4 **(1982)** 1011–1016.
- Miller, A.J., *Subset Selection in Regression*, First Edition, Chapman & Hall, New York, Chapter 2, **1990**.
- Reitz, M., Atzpodien, J., Royston, P., An approach to estimating prognosis using fractional polynomials in metastatic renal carcinoma, *British Journal of Cancer*, 94:12 **(2006)** 1785-1788.
- Royston, P., Altman, D.G., Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling, *Applied Statistics*, 43:3 **(1994)** 429-467.
- Royston, P., Sauerbrei, W., A New Approach to Modelling Interactions Between Treatment and Continuous Covariates in Clinical Trials by Using Fractional Polynomials, *Statistics in Medicine*, 23 **(2004)** 2509–2525.
- Royston, P., Sauerbrei, W., *Multivariable Model-Building*, John Wiley & Sons, Inc, New York, Chapter 4-6, **2008**.
- Sauerbrei, W., Royston, P., Ambler, G., The Use Of Fractional Polynomials to Model Continuous Risk Variables in Epidemiology, *US National Library of Medicine National Institutes of Health*, 28:5 **(1999)** 964-974.
- Sauerbrei, S., Royston, P., Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials, *Journal of the Royal Statistical Society (Series A)*, 165:2 **(2002)** 399–400.
- Sauerbrei, W., Royston, P., Meier-Hirmer, C., Benner, A., *Multivariable Regression Model Building by Using Fractional Polynomials: Description of SAS, STATA, R*, *Computational Statistics & Data Analysis* 50 **(2006)** 3464 – 3485.
- Singh, R., Mukhopadhyay, K., *Survival Analysis in Clinical Trials: Basics and Must Know Areas*, *Perspectives in Clinical Research*, **(2011)** 145-148.
- Therneau, T.M., Grambsch, P.M., *Modelling Survival Data: Extending the Cox Model*, First Edition, New York, Chapter 3, **2000**.
- Zapfen, K., Sauerbrei, W., Royston, P. Detecting an Interaction Between Treatment and A Continuous Covariate: A Comparison of Two Approaches, *Computational Statistics & Data Analysis*, 51:3 **(2007)** 4054-4063.
- Zhang, Z., Chen, K., Ni, H., Fan, H., Predictive Value of Lactate in Unselected Critically Ill Patients: An Analysis Using Fractional Polynomials, *J Thorac Dis.*, 6:7 **(2014)** 995.



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/~~DOKTORA~~ TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 25/06/2019

Tez Başlığı / Konusu: Kesirli Polinomlu Yaşam Modelleri

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 52 sayfalık kısmına ilişkin, 25/06/2019 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %9'dur.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/~~dâhil~~
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.


25.06.2019

Adı Soyadı: Hazal DİNÇ

Öğrenci No: N14321988

Anabilim Dalı: İstatistik

Programı: Yüksek Lisans

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.



Doç. Dr. Nihal ATA TUTKUN

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Hazal DİNÇ
Doğum Yeri : İzmit
Medeni Hali : Bekar
E-posta : hazalldinc@gmail.com
Adresi : Söğütözü Cad. No:4 Söğütözü/Ankara

Eğitim

Lise : 2005-2009 Gölcük İhsaniye Anadolu Lisesi
Lisans : 2009-2014 Hacettepe Üniversitesi İstatistik
Yüksek Lisans : 2015-2019 Hacettepe Üniversitesi İstatistik

Yabancı Dil ve Düzeyi

İngilizce: 73.75

İş Deneyimi

Haziran 2016-Eylül 2016 ING Bank
Aralık 2017- Halen devam ediyor Türkiye Noterler Birliği

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar