

**ÖLÇÜM HATALI AÇIKLAYICI DEĞİŞKENLİ COX
REGRESYON MODELİNE BAYESCİ YAKLAŞIM**

**BAYESIAN APPROACH TO COX REGRESSION MODEL
WITH COVARIATE SUBJECT TO MEASUREMENT
ERROR**

HATİCE IŞIK

PROF. DR. DURU KARASOY

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

İstatistik Anabilim Dalı için Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

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 olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

30 / 06 /2020

HATİCE IŞIK

YAYINLANMA FİKRİ MÜLKİYET HAKKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve 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 kullanması zorunlu metinlerin yazılı izin alarak 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.

30 / 06 /2020



HATİCE IŞIK

ÖZET

ÖLÇÜM HATALI AÇIKLAYICI DEĞİŞKENLİ COX REGRESYON MODELİNE BAYESÇİ YAKLAŞIM

Hatice IŞIK

Doktora, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Duru KARASOY

Haziran 2020, 64 sayfa

Cox regresyon modeli, ölçüm hatalı açıklayıcı değişken içermesi durumunda bilinen kısmi olabilirlik fonksiyonunun maksimizasyonu ile çözümlendiğinde, katsayı kestirimlerinde önemli bir yana neden olabilir. Bu durumda kısmi olabilirlik yaklaşımına alternatif yeni tahmin yöntemleri önerilmiş ve bu yöntemlerin etkinliği tartışılmıştır. Ölçüm hatası, sürekli ölçekle ölçülmüş açıklayıcı değişkenlerin gerçek değerlerinin elde edilemediği durumlarda ortaya çıkar. Yaşam çözümlemesi çalışmalarında da sıklıkla kullanılan sistolik kan basıncı, CD4 sayısı gibi sürekli ölçekle ölçülmüş biyolojik ölçümler çoğunlukla ölçüm hatası içeren değişkenlerdir. Bayesci yaklaşım, ölçüm hatalı Cox regresyon modelinde yanı azaltmak için kullanılan yöntemlerden biridir. Bu tez çalışmasında ölçüm hatası için düzeltilmiş poligonal önselli Bayesci Cox regresyon modeli, literatürde yer alan kısmi olabilirlik fonksiyonuna, Bayesci çözümlemesine ve ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modeline alternatif olarak önerilmiş ve etkinliği incelenmiştir.

Kısmi olabilirlik fonksiyonundan elde edilen Cox regresyon modeli, Bayesci çözümlemesi, ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modeli ve ölçüm hatası için düzeltilmiş poligonal önselli Bayesci Cox regresyon modeli sonuçları karşılaştırılmıştır. Bu yöntemlerle elde edilen tahmin edicilerin performansını değerlendirmek için farklı senaryolar ile benzetim çalışmaları yapılmıştır. Analizler sonucunda riskin yüksek olduğu durumlarda ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modelinin tüm senaryolarda en iyi sonucu verdiği, riskin düşük olduğu durumda ise ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modelinin neredeyse tüm senaryolarda uygun bir alternatif olmadığı sonucuna ulaşılmıştır. Ölçüm hatası için düzeltilmiş poligonal önselli Bayesci Cox regresyon modelinin düşük riskte, düşük ve orta güvenilirlik durumlarında, büyük örneklem ve düşük durdurma hariç daha iyi sonuç verdiği, yüksek güvenilirlikteki senaryolarda ise Cox regresyon modeline ve Bayesci çözümlemesine düzeltilmiş yöntemlerden en az birinin alternatif bir yöntem olduğu sonucuna ulaşılmıştır. Ancak ölçüm hatalı açıklayıcı değişken olması durumunda örneklem büyüklüğü arttıkça klasik yöntem ve Bayesci çözümlemesi ile elde edilen katsayılar için kapsama olasılıkları özellikle düşük ve orta güvenilirlik durumlarında normal düzeyinin oldukça altına düşmektedir.

Yöntemlerin gerçek verilerde uygulanabilirliğini göstermek amacıyla R programında yer alan Mayo Klinik çalışmasından elde edilen primer biliyer sirozu verisinde sonuçlar elde edilmiş, karşılaştırılmış ve yorumlanmıştır. İkinci bir uygulama olarak R programında yer alan Goldman ve ark.'nın (1996) kullanmış oldukları AIDS verisinden rastgele seçilen farklı örneklemeler için model sonuçları elde edilmiştir. Benzetim çalışması sonucunda ulaşılan bilgiye dayanarak ölçüm hatalı açıklayıcı değişken olması durumunda Cox ve Bayesci Cox çözümlemelerinin uygun olmadığı düşünüldüğünde her iki uygulama için en iyi model poligonal önselli Bayesci Cox regresyon modeli sonucunda elde edilmiştir.

Anahtar Kelimeler: Cox regresyon modeli, orantılı tehlikeler modeli, Bayesci yaklaşım, poligonal fonksiyon önseli, ölçüm hatası.

ABSTRACT

BAYESIAN APPROACH TO COX REGRESSION MODEL WITH COVARIATE SUBJECT TO MEASUREMENT ERROR

Hatice IŞIK

Doctor of Philosophy, Department of Statistics

Supervisor: Prof. Dr. Duru KARASOY

June 2020, 64 pages

The Cox regression model with the measurement error covariate can cause an important bias when estimating coefficients by maximizing the known partial likelihood function. In this case, alternative estimation methods have been proposed and the effectiveness of these methods has been discussed. The measurement error occurs when the actual values of the covariates measured on a continuous scale cannot be obtained. Biological measurements such as systolic blood pressure and CD4 count, which are frequently used in survival analysis, are mostly variables with measurement error. The Bayesian approach is one of the methods used to reduce the bias arising from measurement error in the Cox regression model. In this thesis, the corrected Bayesian Cox regression model with polygonal prior to measurement error was proposed as an alternative to the partial likelihood function method, its Bayesian approach and the corrected Bayesian Cox regression model with Gamma prior to measurement error and its effectiveness was investigated.

The results of the Cox regression model obtained from the partial likelihood function, its Bayesian analysis, the corrected Bayesian Cox regression model with Gamma prior to measurement error, and the corrected Bayesian Cox regression model with polygonal prior to measurement error were compared. Simulation studies were performed with different scenarios to evaluate the performance of the estimators obtained with these methods. It is concluded that as a result of the analysis, in cases where the risk is high, the corrected Bayesian Cox regression model with Gamma prior to measurement error gives the best results in all scenarios, and in the case of low risk, not a suitable alternative in almost all scenarios. It is also shown that if the risk is low, the corrected Bayesian Cox regression model with polygonal prior to measurement error gives better results in low and medium reliability situations except for low censoring and large sample size, and at least one of the corrected models is an alternative method to Cox regression model and its Bayesian analysis in high reliability scenarios. However, if the model has the explanatory variable subject to measurement error, as the sample size increases, the coverage possibilities for the coefficients obtained by the classical method and its Bayesian analysis decrease well below the normal level especially in low and medium reliability cases.

To demonstrate the applicability of the methods in real data, the results of the primary biliary cirrhosis data from the Mayo Clinic study in R were obtained, compared, and interpreted. As a second application, model results were obtained for different samples randomly selected from AIDS data used by Goldman et al. (1996) in R. Based on the information obtained as a result of the simulation study, considering that the Cox and Bayesian Cox regression models are not suitable in the presence of measurement error covariate, the best model for both applications was obtained with the corrected Bayesian Cox regression model with polygonal prior to measurement error.

Keywords: Cox regression model, proportional hazards model, Bayesian inference, polygonal function prior, measurement error.

TEŞEKKÜR

Tezimin her aşamasında değerli fikirleri ile bana destek olan ve tezime katkı sağlayan, mesleki gelişimimde büyük emekleri olan değerli danışmanım Prof. Dr. Duru KARASOY'a,

Tez izleme komiteleri boyunca önerileri ile bana yol gösteren, daha iyi bir çalışma ortaya koymamda yardımcı olan değerli hocalarım Prof. Dr. Mehtap AKÇİL OK'a ve Doç. Dr. Nihal ATA TUTKUN'a ve her toplantıda bize katılarak önemli katkılarda bulunan değerli hocam Doç. Dr. Uğur KARABEY'e,

Çalışmam boyunca gösterdiği sonsuz sevgi, sabır ve desteğiyle yanımda olan, motivasyonumu yükseltmeme daima yardımcı olan sevgili eşim Yasin IŞIK'a ve varlığı ve sevgisiyle bana destek olan canım oğlum Yaman IŞIK'a,

Tez çalışmam boyunca ve akademik çalışma hayatım boyunca fikirleri ve eleştirileri ile bana yol gösteren, manevi desteklerini her zaman hissettiğim sevgili arkadaşlarım Derya TURFAN'a, Dr. Ezgi YILMAZ'a ve Dr. Hatice ÖNCEL ÇEKİM'e,

Tez çalışmamda sorularımı geri çevirmeyerek bana yardımcı olan sevgili hocalarım Doç. Dr. Derya ERSEL'e ve Dr. Öğr. Üyesi Ceren Eda CAN'a,

Hayatım boyunca gösterdikleri sevgi ile her daim yanımda olan sevgili AİLEM'e,

İÇTENLİKLE TEŞEKKÜR EDERİM...

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	iii
TEŞEKKÜR	v
İÇİNDEKİLER.....	vi
ÇİZELGELER DİZİNİ.....	ix
ŞEKİLLER DİZİNİ	viii
SİMGELER VE KISALTMALAR.....	x
1. GİRİŞ	1
2. YAŞAM ÇÖZÜMLEMESİ.....	4
2.1. Cox Regresyon Modeli.....	6
2.2. Cox Regresyon Modeli İçin Sayma Süreci.....	7
3. ÖLÇÜM HATASI MODELLERİ.....	11
3.1. Klasik Ölçüm Hatası Modeli	12
3.2. Gerçek Etkiye Bağlı Sistematik Ölçüm Hatası Modeli	13
3.3. Değişen Varyanslı Ölçüm Hatası Modeli	13
3.4. Diferansiyel Ölçüm Hatası Modeli.....	14
4. ÖLÇÜM HATALI AÇIKLAYICI DEĞİŞKENLİ COX REGRESYON MODELİ İÇİN BAYESCİ YAKLAŞIM.....	15
4.1. Cox Regresyon Modelinde Bayesci Çözümleme	17
4.2. Ölçüm Hatası Modellerinde Bayesci Çözümleme.....	23
4.3. Ölçüm Hatalı Açıklayıcı Değişkenli Cox Regresyon Modelinin Düzeltilmiş Gamma Önselli Bayesci Çözümlemesi	26
4.4. Ölçüm Hatalı Açıklayıcı Değişkenli Cox Regresyon Modelinin Düzeltilmiş Poligonal Önselli Bayesci Çözümlemesi	27

5. BENZETİM ÇALIŞMASI VE UYGULAMA	29
5.1. Benzetim Çalışması.....	29
5.1.1. Benzetim Çalışması Sonuçları	31
5.2. Uygulama	41
5.2.1. Uygulama 1.....	41
5.2.2. Uygulama 2.....	52
6. SONUÇ VE TARTIŞMA	56
7. KAYNAKLAR	59
EKLER.....	62
EK 1 - Tezden Türetilmiş Bildiriler	62
EK 2 - Tez Çalışması Orjinallik Raporu	63
ÖZGEÇMİŞ	64

ÇİZELGELER DİZİNİ

Çizelge 5.1. $\beta=0.5$ için HKO ve KO sonuçları	32
Çizelge 5.2. $\beta=1$ için HKO ve KO sonuçları	35
Çizelge 5.3. $\beta=1.5$ için HKO ve KO sonuçları	39
Çizelge 5.4. PBC verisinde sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler	42
Çizelge 5.5. PBC verisinde kullanılan nitel değişkenler	42
Çizelge 5.6. PBC verisinde orantılı tehlikeler varsayımı test sonuçları	43
Çizelge 5.7. DGB_CRM için Heidelberg-Welch yakınsaklık testi sonucu.....	47
Çizelge 5.8. DPB_CRM için Heidelberg-Welch yakınsaklık testi sonucu	49
Çizelge 5.9. B_CRM için Heidelberg-Welch yakınsaklık testi sonucu.....	51
Çizelge 5.10. PBC verisi için model sonuçları	51
Çizelge 5.11. AIDS verisinde sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler	53
Çizelge 5.12. AIDS verisinde kullanılan nitel değişkenler.....	53
Çizelge 5.13. AIDS verisinde orantılı tehlikeler varsayımı test sonuçları.....	54
Çizelge 5.14. AIDS verisinde örneklem büyüklüğü 30, durdurma oranı 0.60 olan örneklemelere ilişkin sonuçlar.....	55

ŞEKİLLER DİZİNİ

Şekil 3.1. Rastgele ve Sistemik Hata.....	11
Şekil 5.1. $\beta=0.5$ için HKO ve KO sonuçlarının grafiği.....	33
Şekil 5.2. $\beta=1.0$ için senaryoların HKO ve KO sonuçlarının grafiği	36
Şekil 5.3. $\beta=1.5$ için senaryoların HKO ve KO sonuçlarının grafiği	40
Şekil 5.4. PBC verisi için süreye ilişkin dağılım grafiği	43
Şekil 5.5. Log-serum bilirubin tekrarlarının ortalamalarına karşı standart sapma grafiği	44
Şekil 5.6. Log-serum bilirubin değişkeninin tekrarlı gözlemleri arasındaki farkın normal QQ grafiği	45
Şekil 5.7. DGB_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri	46
Şekil 5.8. DPB_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri	48
Şekil 5.9. B_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri	50
Şekil 5.10. AIDS verisinde süreye ilişkin dağılım grafiği	53

SİMGELER VE KISALTMALAR

Simgeler

β	Regresyon katsayısı
X	Ölçüm hatası içeren açıklayıcı değişkenlerin gerçek değerine ilişkin vektör
Z	Ölçüm hatası içermeyen açıklayıcı değişkenler vektörü
W	Ölçüm hatası içeren açıklayıcı değişkenlerin ölçüm hatalı değerine ilişkin vektör
ε	Hata terimi

Kısaltmalar

HKO	Hata Kareler Ortalaması
JAGS	Jags Another Gibbs Sampling
MCMC	Monte Carlo Markov Zinciri
DGB_CRM	Ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modeli
DPB_CRM	Ölçüm hatası için düzeltilmiş Poligonal önselli Bayesci Cox regresyon modeli
B_CRM	Bayesci Cox regresyon modeli
CRM	Cox regresyon modeli
KO	Kapsama olasılığı
AIDS	Acquired Immune Deficiency Syndrome (Edinilmiş Bağışıklık Eksikliği Sendromu)
ddC	Didanosine
ddI	Zalcitabine
DIC	Deviance Information Criteria

1. GİRİŞ

Yaşam çözümlemesinde en çok kullanılan model Cox (1972) tarafından önerilen Cox regresyon modeli (CRM) ya da diğer adıyla orantılı tehlikeler modelidir. Model kolay anlaşılabilir, herhangi bir dağılıma bağlı olmayan ve uygulanması kolay bir regresyon modelidir. Modelin temel varsayımı, yaşam süresi üzerinde etkili olduğu düşünülen değişkenlere ait tehlike oranının zaman boyunca sabit olmasıdır.

Birçok alanda, gelecek zamana ait verileri kestirmek/öngörmek büyük önem taşır çünkü öngörü, gelecekteki olası olaylar ve onların sonuçları hakkında belli bilgiler ortaya koymaktadır. Açıklayıcı değişkenlerin alacağı değerlere göre tehlike fonksiyonunun bilinmeyen değerlerinin kestirimi ile ilgilendiğimizde CRM'nin uygun bir kestirim denklemi üretebilmesi öncelikli gereksinimdir. Bu nedenle, regresyon katsayılarının tahmininde farklı yöntemler önerilmiştir. Literatürde CRM'de regresyon katsayılarının tahmininde en çok karşımıza çıkan yaklaşım kısmi olabilirlik yaklaşımıdır.

Açıklayıcı değişkenlerin gerçek değerlerinin elde edilemediği durumlarda ölçüm hatası ortaya çıkar. Ölçüm hatası, laboratuvar hatası, ölçüm cihazlarındaki kalibrasyon hatası ya da örneklemeden kaynaklı meydana gelebilir (Keogh ve White, 2014). Yaşam çözümlemesi çalışmalarında da sıklıkla kullanılan kan basıncı, CD4 sayısı gibi sürekli ölçekle ölçülmüş biyolojik ölçümler çoğunlukla ölçüm hatası içeren değişkenlerdir. Açıklayıcı değişkenlerde sıklıkla görülebilen ölçüm hatası sorununun çözümü yaşam çözümlemesinde de karşılaşılan zorluklardandır. Açıklayıcı değişkenlerin ölçüm hatası içeren değerleri klasik çözümlemelerde kullanıldığında yanıltıcı sonuçlara neden olacaktır (Nakamura, 1992). Ölçüm hatası dikkate alınmazsa CRM'de kısmi olabilirlik yöntemi ile elde edilen sonuçlarda ciddi bir yana neden olacaktır (Yi ve Lawless, 2007).

Ölçüm hatası probleminin çözümü için birçok yöntem önerilmiştir. Bunlar temel olarak olabilirlik temelli yöntemler, skor fonksiyonu, yarı parametrik ve parametrik olmayan yöntemler ile Bayesci çözümlemelerdir.

İlk olarak Prentice (1982) CRM'de ölçüm hatalı açıklayıcı değişkenin etkisini düzeltmek için regresyon kalibrasyonu yöntemini önermiştir. Clayton (1992), her bir risk kümesiyle

regresyon kalibrasyon yöntemini kullanarak Prentice'in (1982) yöntemini güncellemiştir. Nakamura (1992), ölçüm hatalı açıklayıcı değişken içeren CRM'de β katsayılarının kestirimi için yaklaşık kısmi olabilirlik skor fonksiyonu önermiştir. Buzas (1998), benzer şekilde β için yansız skor fonksiyonu önermiştir. Wang ve ark. (1997), test (validation) verisi olduğu durum için regresyon kalibrasyon yöntemini önermişlerdir. Hu, Tsiatis ve Davidian (1998) tek açıklayıcı değişken için tam olabilirlik fonksiyonu yöntemini önermişlerdir. Zhou ve Wang (2000) parametrik olmayan bir yaklaşım önermişler ancak bu yöntem ölçüm hatası içeren açıklayıcı değişkenler için tekrarlı ölçümler gerektirmektedir. Xie, Wang ve Prentice (2001) yine tekrarlı ölçümleri elde edilebilen açıklayıcı değişkenlerin var olması durumunda en küçük kareler yöntemi ile tehlike fonksiyonu kalibrasyonunu uygulamışlardır. Hu ve Lin (2002), Nakamura (1992) ve Buzas'ın (1998) çalışmalarını genelleştirerek CRM'de regresyon parametresi β ve birikimli tehlike fonksiyonu $H_0(t)$ için tutarlı tahmin ediciler elde etmişlerdir. Augustin (2004) çalışmasında Nakamura'nın (1992) elde ettiği yaklaşımın yanlılığını giderecek yansız tahmin edicilerin elde edildiği düzeltilmiş tam olabilirlik fonksiyonunu önermiş ve farklı durumlar için genişletmiştir. Song ve Huang (2005), β 'yı elde etmek için koşullu skor yöntemi önermişlerdir. Yi ve Lawless (2007), Nakamura'nın (1992) düzeltilmiş skor yöntemini kullanarak β ve birikimli tehlike fonksiyonunu birlikte çıkarsamışlardır. Grace ve Lawless (2007) çalışmalarında yine düzeltilmiş bir olabilirlik fonksiyonu önermişler ve "soldan kesilmiş" ve "tekrarlı durumlar" için yaklaşımlarını genişletmişlerdir. Bartlett ve Keogh (2018) ölçüm hatalı CRM'de Bayesci yaklaşımı ele almışlardır.

Bayesci yöntemler CRM'de ölçüm hatası sorununun çözümü için son yıllarda önerilen yöntemlerden biridir. CRM'de yaşam süresi üzerinde etkisi olan değişkenlerin ve bu değişkenlerin etkilerinin büyüklüğünün doğru olarak belirlenmesi çok önemlidir. Veri setinin büyüklüğü, içerdiği durdurma oranı ve güvenilirlik düzeyine göre kullanılan çözümleme ölçüm hatasının yarattığı yanlılığı etkileyebilir. Bayesci düzeltme ve kısmi olabilirlik fonksiyonu ile elde edilen sonuçlarda ölçüm hatasının tüm bu parametrelere göre ne kadar yanlılık yarattığı literatürde karşılaştırılmamıştır. Bu nedenle doktora tez çalışmasında CRM'de regresyon katsayılarının tahmininde kullanılan, ölçüm hatasını dikkate almayan klasik yöntemler ile ölçüm hatası için kullanılan Bayesci düzeltmenin hangi örneklem büyüklüğünde, hangi durdurma oranında ve hangi güvenilirlikte daha iyi sonuç verdiğinin incelenmesi ve önerilecek yeni yöntem ile daha iyi katsayı

kestirimlerinin elde edilmesi hedeflenmektedir. Yöntemlerin farklı yapıda veri setlerindeki işlevselliği için incelemeler yapılacak ve daha doğru kestirimler için katkıda bulunulacaktır.

Çalışmanın izleyen ikinci bölümünde yaşam çözümlemesi ve CRM için temel tanım ve kavramlar verilmiştir.

Çalışmanın üçüncü bölümünde ölçüm hatası hakkında bilgi verildikten sonra ölçüm hatası modelleri tanıtılmıştır.

Çalışmanın dördüncü bölümünde ölçüm hatalı açıklayıcı değişkenli CRM’de Bayesci çözümlemenin uygulama adımları için öncelikle CRM’de Bayesci çözümleme tanımlanmıştır. İkinci aşamada ölçüm hatası için Bayesci çözümlemenin nasıl uygulandığı, üçüncü aşamada CRM’de ölçüm hatasının düzeltilmesinde gamma önselli modelin tanımlanması ve son olarak tez kapsamında önerilen ölçüm hatasının düzeltilmesinde poligonal önselli modelin tanımlanması sunulmuştur.

Çalışmanın beşinci bölümünde benzetim çalışması ve gerçek veri uygulaması yapılmıştır. Çalışma kapsamında tanımlanan farklı regresyon katsayısına, örneklem büyüklüğüne, durdurma oranına ve güvenilirlik düzeyine göre veri setleri üretilmiştir. CRM’nin kısmi olabilirlik fonksiyonu ile elde edilen sonuçları, ölçüm hatasını göz ardı eden Bayesci çözümlemesi sonuçları, ölçüm hatası için düzeltilmiş gamma önselli ve tez kapsamında önerilen ölçüm hatası için düzeltilmiş poligonal önselli model sonuçları karşılaştırılmıştır. Sonrasında çalışmaya uygun gerçek veri setlerinde model sonuçları elde edilmiş ve önerilen modelin uygulanabilirliği gösterilmiştir.

Son bölümde ise sonuçlar hakkında genel değerlendirme yapılmıştır.

2. YAŞAM ÇÖZÜMLEMESİ

Yaşam çözümlemesi, ilgilenilen olay ile karşılaşınca kadar geçen sürenin çözümlemesi için istatistiksel yöntemlerin gelişimini inceleyen bir alandır (Lawless, 2003). Yaşam çözümlemesi sağlık bilimleri, biyoloji, mühendislik, ekonomi gibi birçok alanda kullanılan önemli bir çözümlemedir. Yaşam çözümlemesindeki önemli farklılık ilgilenilen sürenin sonunda tanımlanan olay dışında ortaya çıkabilecek durdurulmuş gözlemleri de çözümlemeye dahil etmesidir.

T_i yaşam süresini ve C_i durdurma süresini göstermek üzere gözlenen minimum süre (T_i, C_i) için yaşam süresi ve durdurma süresinin bağımsızlığı varsayımı altında durdurma göstergesi $\delta_i = I(T_i \leq C_i)$ ile gösterilir ve i . gözlem durdurulmuş ise 0, olay gözlemlenmiş ise 1 değerini alır.

T rastlantı değişkeninin t 'den büyük olma olasılığı yaşam fonksiyonu olarak tanımlanır ve

$$S(t) = P(T > t) = \int_t^{\infty} f(x) dx, \quad 0 < t < \infty$$

biçiminde gösterilir. Yaşam fonksiyonu için $F(t)$, t 'nin dağılım fonksiyonu, $f(t)$ olasılık yoğunluk fonksiyonunu göstermek üzere,

$$F(t) = P(T < t) = \int_0^t f(u) du$$

eşitliği ile yaşam süresinin t 'den küçük olma olasılığı elde edilir. Buradan,

$$S(t) = 1 - F(t)$$

eşitliği de yazılabilir. Monoton azalan soldan sürekli bir fonksiyon olan yaşam fonksiyonunun $t = 0$ için $S(0) = 1$ ve $t = \infty$ için $S(\infty) = 0$ değerini alır. Tehlike fonksiyonu ise t zamanına kadar tanımlanan olay ile karşılaşmamış birimin t zamanında başarısız olma riskidir ve $h(t)$ ile gösterilir (Collett, 1994). $h(t)$, yaşam fonksiyonunun aksine başarısızlığa odaklanır. Tehlike fonksiyonu için yaşam fonksiyonunun verdiği bilginin tam tersi olarak düşünülebilir. Tehlike fonksiyonu,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

biçimindedir (Kleinbaum ve Klein, 2010). Buradan tehlike fonksiyonundaki paydaki olasılık, koşullu olasılığın tanımından aşağıdaki gibi de ifade edilebilir:

$$\frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{P(T \geq t)}$$

Bu durumda tehlike fonksiyonu,

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)}$$

biçiminde de yazılabilir. Burada $F(t)$ 'nin t 'ye göre türev tanımının $f(t)$ 'ye eşit olduğu bilinmektedir. Bu durumda

$$\lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} = f(t)$$

eşitliği yazıldığında tehlike fonksiyonu,

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

elde edilir. Ayrıca tehlike fonksiyonu ve yaşam fonksiyonu arasındaki fonksiyonel ilişkiyi ifade eden diğer bir eşitlik,

$$h(t) = \frac{d}{dt} \{\log(S(t))\}$$

biçimindedir. Buradan,

$$S(t) = \exp(-H(t))$$

yazılabilir. $H(t)$ birikimli tehlike fonksiyonunu göstermek üzere,

$$H(t) = \int_0^t h(u) du = -\log S(t)$$

eşitliği ile elde edilir (Collet, 1994).

2.1. Cox Regresyon Modeli

Literatürde yaygın olarak kullanılan ve orantılı tehlikeler modeli olarak da adlandırılan CRM yarı parametrik bir modeldir ve aşağıdaki biçimde ifade edilebilir (Cox, 1972):

$$h(t, \mathbf{X}) = h_0(t) \exp\left(\sum_{i=1}^p \beta_i x_i\right) \quad (2.1)$$

Burada modelin ilk çarpanı temel tehlike fonksiyonunu, $\mathbf{X}$ açıklayıcı değişkenler vektörünü, β regresyon katsayılar vektörünü gösterir. CRM, tanımlanan açıklayıcı değişkenler vektörüne sahip bir birimin t zamanındaki riskini modeller.

CRM'deki katsayıların tahmininde sıklıkla kullanılan yaklaşım kısmi olabilirlik fonksiyonu yaklaşımıdır. CRM'de çıktı değişkeninin dağılımı için her hangi bir biçim tanımlanmadığı için tam olabilirlik fonksiyonu kullanılamaz. Kısmi olabilirlik yaklaşımında olabilirlik fonksiyonuna durdurulmuş gözlemler hesaba katılmayıp yalnızca başarısız olan gözlemler için olasılıklar dahil edilir. Olabilirlik fonksiyonu için tüm birimler düşünülmediği için kısmi olabilirlik olarak adlandırılır (Kleinbaum ve Klein, 2010).

δ_i başarısızlık için 1, durdurma için 0 değerini alan durdurma göstergesi, $t_{(1)}, t_{(2)} \dots, t_{(k)}$ k tane başarısızlık için sıralı başarısızlık süreleri, t_i durdurma sürelerini gösteren rastgele değişken olsun. $R(t) = \{i : t_i \geq t\}$ t zamanında riskte olan birimlerin kümesini gösterebilir. Kısmi olabilirlik fonksiyonu,

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta' X_j)}{\sum_{\ell \in R(t_j)} \exp(\beta' X_\ell)}$$

biçimindedir. Burada paydadaki toplam, t_i zamanda riskte olan birimler üzerinden $\exp(\beta' x_i)$ değerlerinin toplamıdır (Cox, 1972). CRM'de, β parametrelerinin en çok olabilirlik tahminleri Newton-Raphson tekniği gibi yöntemler kullanılarak log-olabilirlik fonksiyonunun maksimize edilmesiyle bulunmaktadır (Collett, 1994).

2.2. Cox Regresyon Modeli İçin Sayma Süreci

Yaşam modellerinin çözümlemesinde kullanılan pratik ve alternatif yollardan biri sayma sürecidir (counting process). Daha karmaşık yaşam çözümlemesi modelleri için kullanılan bir yöntemdir (Kleinbaum ve Klein, 2010). Sayma süreci yaklaşımı ilk defa Aalen (1975) tarafından stokastik integrasyonun bileşenlerini, sürekli zaman martingale teoremini ve sayma sürecini birleştiren bir yöntem olarak geliştirilmiştir. Sayma süreci Poisson sürecinin genelleştirilmiş halidir. CRM ise Poisson regresyon modelinin oranlar için genişletilmiş durumudur.

CRM sıralı başarısızlık sürelerinin sayısını sayan sayma sürecinin özel bir durumu olarak tanımlanabilir. Sayma süreci formülasyonu ve güçlü martingale teorisi göz önüne alındığında, CRM ve asimptotik stokastiğini belirlemek için sayma süreci işlemleri kolayca kullanılabilir. Örneğin, log olabilirlik fonksiyonunun kısmi türevleri olan skor fonksiyonu $\tilde{U}(\beta_0, \cdot)$, temelde martingale merkezi limit teoreminin uygulandığı yerel bir martingaledir (Liu, 2012).

Sayma süreci verisinin analizi genellikle yoğunluğun modellenmesine bağlıdır. T ilgilenilen olayın gerçekleştiği ana kadar geçen süre olsun, $S(t) = P(T > t)$ yaşam fonksiyonunu versin. Δt küçük olmak üzere yaşam süresinin Δt uzunluğundaki birçok küçük eşit aralığa bölündüğünü düşünelim. T sürekli olsun. t süresine kadar olayın gerçekleşmediği ancak sonlu $[t, t + \Delta t]$ zaman aralığında olayın gerçekleşme olasılığı hesapladığında tüm birimler için ortak olan temel tehlike fonksiyonu;

$$h_0(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}$$

ile tanımlanabilir. $h_0(t)$, t süresine kadar olay ile karşılaşmamış birimin olay ile karşılaşmasının anlık riski olarak yorumlanabilir.

Eşitlik (2.1)' de verilen modelde, modelin parametrik olmayan kısmı $h_0(t)$, parametrik kısmı ise $\beta'X = \sum_j \beta_j X_{ij}$ ile ifade edilir. Burada β ve X 'in t zamanı ile sabit olduğu ve p boyutlu

regresyon parametresi β ve temel tehlike fonksiyonu $h_0(t)$ 'nin yerel olarak integrallenebilir olduğu varsayılır. Yani,

$$\int_0^{T_{\max}} h_0(t)dt < \infty$$

sağlanmalıdır.

Anderson ve Gill (1982), yaşam süresi verileri için sayma sürecinin kullanımını genişletmişlerdir. n birim için, $I_i(t)$, sayma sürecinde verilen açıklayıcı değişkenler vektörü $X_i = (X_{i1}, \dots, X_{ip})$ ve $Y_i(t)$ risk göstergesi için yoğunluk sürecini gösterir. Burada $Y_i(t)$, T_i zamanında hala riskte olan, t zamanından hemen önce olay gözlemlenmemiş ya da durdurulmamış gözlemler içindir. $N_i(t)$ süreci $[0, t]$ aralığında gerçekleşen olayları sayar. Bu süreç sabittir ve olaylar arası 0 değerini alır, her bir olay zamanında bir birim artar. Burada stokastik süreç $\{N_i(t), t \geq 0\}$ ile gösterilir ve aşağıdaki koşulları sağlar:

- i. $N_i(t) \geq 0$,
- ii. $N_i(t)$ tam sayı değeri alır,
- iii. $s < t$ ise $N_i(s) < N_i(t)$, $[s, t]$ zaman aralığında meydana gelen olay sayısını gösterir.

Yeni bir olay oranı $I_i(t) = Y_i(t) h(t | X_i)$ ile ifade edilir. $I_i(t)$, $[t, t+\Delta t)$ gibi bir zaman aralığında daha önce gerçekleşmeyen ilgilenilen olay ile karşılaşma olasılığını göstermektedir. Bu da yaklaşık olarak aşağıdaki eşitlik ile ifade edilebilir:

$$dN_i(t) \approx h(t | X_i)dt = h_0(t) \exp[\beta'X_i] = I_i(t)$$

Burada $dN_i(t)$, $[t, t+\Delta t)$ kadar bir zaman aralığında $N_i(t)$ 'deki artışı gösterir, yani bu aralıktaki olay sayısını verir. Bu durumda çarpımsal yoğunluk,

$$I_i(t) = Y_i(t)h(t | X_i) = Y_i(t)h_0(t)\exp[\beta'X_i] \quad (2.2)$$

biçiminde ifade edilebilir. Bu durumda yoğunluk süreci gözlenen süreç ile gözlenmeyen fonksiyonun çarpımıdır. Cox model altında $N_i(t)$ için yoğunluk süreci,

$$I_i(t)dt = Y_i(t)\exp[\beta'X_i]dH_0(t)$$

biçimindedir. Burada $dH_0(t)$, t zamanında riskte olan birimin hemen sonraki $[t, t+\Delta t)$ aralığında olayla karşılaşmasının anlık riskini gösterir.

Yoğunluk sürecinin diğer bir tanımı aşağıdaki gibi ifade edilebilir:

$$I_i(t)dt = P(dN_i(t) = 1 | \text{geçmiş}) = P([N_i(t + \Delta t) - N_i(t)] = 1 | \text{geçmiş}) = E(dN_i(t) = 1 | \text{geçmiş})$$

Buradan da,

$$E(dN_i(t) - I_i(t)dt | \text{geçmiş}) = 0$$

yazılabilir. $dN_i(t)$ 'nin yaklaşık eşitliği dikkate alındığında,

$$M_i(t) = N_i(t) - \int_0^t I_i(s)ds$$

biçiminde yazılabilir. Bu da “hata = (gözlenen – model)” olarak düşünülebilir. Buradan da $E(dM_i(t) | \text{geçmiş}) = 0$ eşitliği yazılabilir. Sayma süreci ve integrallenmiş yoğunluk süreci arasındaki farkın verilen yoğunluk süreci $I_i(t)$ 'ye eşitliği bir martingaledir. Yaşam fonksiyonu için Gill (2005) tarafından verilen,

$$S(t) = \prod_{(0,t]} \{1 - dH(s)\}$$

eşitliği düşünölsün. Birikimli yoğunluk,

$$dH_i(s) = \int_0^t I_i(s)ds$$

biçiminde ifade edilebildiği için,

$$\prod_{(0,t]} \{1 - dH_i(s)\} = \exp\left(-\int_0^t I_i(s)ds\right)$$

olur. Elde edilen veri $D = \{N_i(t), Y_i(t), X_i, i = 1, \dots, n\}$ ile gösterilsin, β ve $H_0(t)$ bilinmeyen parametreler olmak üzere bilgi içermeyen durdurma altında olabilirlik fonksiyonu,

$$L_i(D | \beta, H_i(t)) = \exp \left(- \int_{t \geq 0} I_i(t) dt \right) \prod_{t \geq 0} (I_i(t))^{dN_i(t)}, i = 1, \dots, n$$

biçimindedir. Bu durumda bileşik dağılım (joint distribution) aşağıdaki gibi elde edilir (Mostafa ve Ghorbal, 2011):

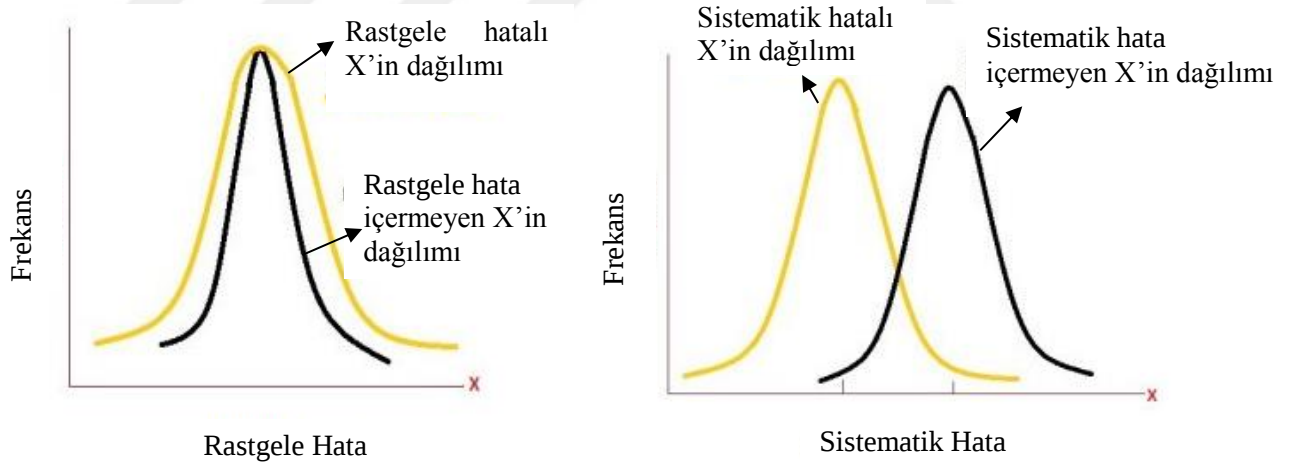
$$L(D | \beta, H(t)) = \prod_{i=1}^n L_i(D | \beta, H_i(t)) \quad (2.3)$$



3. ÖLÇÜM HATASI MODELLERİ

Modele girecek bir ya da birden çok açıklayıcı değişkenin gerçek değerinin elde edilemediği durumlarda ölçüm hatası meydana gelir. Ölçüm hatasının birçok nedeni olabilir ancak en çok karşılaşılanı araç-gereç hatası ve örnekleme hatasıdır (Buonaccorsi, 2010). Ölçüm hatası rastgele hata ve sistematik hata olmak üzere iki bileşenden oluşur. İlk bileşen olan rastgele hata ölçümü rastgele etkileyen laboratuvar ölçümü yapan araç gerecin yanlış ölçüm yapmasından ya da çevresel koşullardaki rastgele değişimlerden kaynaklanabilir. İkinci bileşen olan sistematik hata ise yanlış kalibrasyon ya da hatalı cihaz kullanımı gibi rastgele olmayan problemlerden kaynaklı olarak tüm ölçümleri etkileyebilir (Espino-Hernandez, Gustafson, Burstyn, 2011).

Rastgele ve sistematik hatanın dağılım üzerindeki etkisi Şekil 3.1’de verilmiştir. Rastgele hata değişkenin yalnızca yayılımı üzerinde fark yaratırken, sistematik hata değişkenin konumunu etkileyerek yan yaratmaktadır.



Şekil 3.1. Rastgele ve Sistematik Hata

Bir modelde ölçüm hatalı açıklayıcı değişken olmasının üç etkisi vardır. Bunlar,

- İstatistiksel modelde parametre tahminlerinin yanlış olması,
- Değişkenler arası ilişkilerin ortaya çıkarılmasında gücün düşmesi,

- Verilerin özelliklerini gizlemesi ve grafiksel modellemenin zorlaşmasıdır (Carroll ve ark., 2006).

Ölçüm hatası, ekonomik, sosyal, tıbbi ve epidemiyolojik araştırmalar dahil olmak üzere pek çok araştırma alanında tahmin sürecinde önemli bir zorluktur (Gray, 2018). Klasik çözümlemelerde ölçüm hatasının neden olduğu yanı gidermek ya da azaltmak için literatürde birçok yöntem önerilmiştir. Bunlar temel olarak olabilirlik temelli yöntemler, fonksiyonel tahmin yöntemleri (düzeltilmiş skor gibi) ve Bayesci çözümlemelerdir (Schneeweiss ve Augustin, 2006).

Ölçüm hatasından kaynaklı yanlışlığın miktarı ve yapısı ölçüm hatasının tipine bağlıdır. Literatürde ölçüm hatası modellemeleri klasik ölçüm hatası, sistematik ölçüm hatası, değişen varyanslı ölçüm hatası ve differansiyel ölçüm hatası modelleri olarak verilir.

3.1. Klasik Ölçüm Hatası Modeli

Toplamsal ölçüm hatası olarak da adlandırılan klasik ölçüm hatası modelinde yanın sistematik olmadığı, hatanın kaynağının rastgele olduğu varsayılır (Gray, 2018). Klasik ölçüm hatası modeli,

$$W_{ij} = X_i + \varepsilon_{ij}, \quad j=1, 2$$

biçimindedir. Burada X_i , ölçüm hatalı açıklayıcı değişkenin gerçek, gözlemlenemeyen değeri; W_i , ölçüm hatalı açıklayıcı değişkenin hata içeren gözlemlenen değeridir. Burada hata terimi ε_{ij} , 0 ortalama ve σ_ε^2 varyansa sahiptir. ε_{i1} ve ε_{i2} sıfır korelasyona sahiptir. ε ; X_i , Z_i (ölçüm hatası içermeyen açıklayıcı değişken) ve Y_i ’ den (ilgilenilen çıktı değişkeni) bağımsızdır (Keogh ve White, 2014).

Klasik ölçüm hatası modeline uyumu test etmek için bazı görsel yöntemler kullanılabilir. Tekrarlı ölçümleri elde edilebilen W için her bir birey için elde edilen standart sapmasının, ortalamalarıyla ilişkisiz ve ölçüm hatası içermeyen Z değişkeninden bağımsız olduğu varsayılır. Bu amaçla ortalamalara karşı standart sapmaların çizimi arasında belirgin bir eğilim yoksa ölçüm hatası varyansının X ’den bağımsız olduğu söylenebilir. Her bir birey için W ’nin tekrarlı

ölçümleri için elde edilen standart sapmanın, o bireye karşılık gelen ölçüm hatası içermeyen Z açıklayıcı değişkenine karşı çiziminde yine belirgin bir eğilim yoksa Z'nin ölçüm hatası varyansına bağımlı olmadığı söylenebilir. Her bir gözlem için W'nin tekrarları arasındaki farkın normal QQ grafiği çiziminde normal dağılımdan belirgin bir fark yoksa ölçüm hatalarının normal dağılım gösterdiği söylenebilir (Carroll ve ark., 2006).

3.2. Gerçek Etkiye Bağlı Sistemik Ölçüm Hatası Modeli

Çok sayıda tekrarlı ölçümün ortalaması X_i 'nin tahminini sağladığı için klasik ölçüm hatası modeli ile elde edilen W_{ij} , X_i 'nin yansız ölçümleri olarak tanımlanır. Ancak bazı durumlarda klasik ölçüm hatası modeli uygun olmayabilir. Beslenme epidemiyolojisinde gıda kayıt ölçümleri gerçek değerlerin yanı sıra tahminleridir. Hem X_i 'nin gerçek ölçümleri için hem de bireylere özgü hatalar içermektedir. Bu durumda bir birey üzerinden tekrarlı ölçümler arasında ilişki ortaya çıkmaktadır. Bu durumda gerçek etkiye bağlı sistemik hata için ölçüm hatası modeli,

$$W_{ij} = \psi + \theta X_i + \varepsilon_{ij}, \quad j = 1, 2$$

biçimindedir. Burada ψ , sabit değişim (constant shift), $\theta \neq 1$ gerçek etkiye bağlı hata ve hatalar arası ilişki genel olarak 0 olmayabilir. $\text{corr}(\varepsilon_{i1}, \varepsilon_{i2}) = \rho$ hata korelasyonunu gösterebilir. ε_{ij} hata terimi yine 0 ortalamaya, sabit varyansa sahiptir ve X_i , Z_i ve Y_i 'den bağımsızdır. $\psi = 0$, $\theta = 1$ ve $\rho = 0$ olduğunda gerçek etkiye bağlı sistemik hata modeli klasik ölçüm hatası modeline eşit olacaktır (Keogh ve White, 2014).

3.3. Değişken Varyanslı Ölçüm Hatası Modeli

Diğer ölçüm hatası modellerinde hata terimi sabit varyanslı tanımlanmıştır. Değişken varyanslı hata (heteroscedastic error) modeli hataların X_i 'ye göre değiştiği modeldir. Genellikle de X_i 'nin artışıyla artan, gerçek etki daha büyük olduğunda hatanın da büyüdüğü durumlarda değişken varyanslı hata ortaya çıkar. Bu durumda ölçüm hatası modeli, $E(\varepsilon_{ij} | X_i) = 0$, $V(\varepsilon_{ij} | X_i) = \sigma_\varepsilon^2(X_i)$

biçimindedir. Gerçek X_i gözlemlerinin bir alt grubu varsa X 'in değişimi ile W_{i1} 'in değişip değişmediği grafiksel olarak ortaya koyulabilir. X_i gözlemlenemiyorsa W_{i1} etkisinin tekrarlı ölçümü varsa her bir birim için $\{W_{i1}, W_{i2}\}$ ortalamasına karşılık $\{W_{i1}, W_{i2}\}$ 'nin standart sapması çizilerek klasik hata ölçümü için kısmi test yapılabilir. Bu hata varyansının gerçek etkiye bağlı olup olmadığını gösterir. Sabit hata varyansının istenen özelliklere sahip değerlerini elde etmek için orijinal W_{ij} değişkenine dönüşüm uygulamak mümkündür. Uygun olan en basit dönüşüm log dönüşümüdür. Daha genel olarak Box-Cox dönüşümü uygulanabilir (Keogh ve White, 2014).

3.4. Diferansiyel Ölçüm Hatası Modeli

Daha önceki hata modellerinde hata çıktıdan bağımsızdır. Diferansiyel (differential) hata modelinde çıktıya bağımlıdır ve Y_i 'nin iki düzeyli (binary) değişken olduğu durumda karşımıza çıkar. Örneğin klasik hata modeli altında $Y_i = 0, 1$ değerleri için ε_{ij} hatalarının değişimi değişebilir. Sistemik hata modeli Y_i 'nin değerlerine bağlı olarak sistemik yana izin veren bir model olarak aşağıdaki gibi genişletilebilir:

$$W_{ij} = \psi_{Y_i} + \theta_{Y_i} X_i + \varepsilon_{ij}, \quad j=1, 2; Y_i = 0, 1$$

Ayrıca hata varyansı Y 'ye bağlı olarak

$$\text{var}(\varepsilon_{ij}) = \sigma_{\varepsilon_y}^2, Y_i = y$$

biçiminde, buna ek olarak hata ilişkileri Y 'ye bağlı olarak

$$\text{corr}(\varepsilon_{ij}, \varepsilon_{ik}) = \rho_y, Y_i = y, j \neq k$$

biçiminde ifade edilebilir. Farklı diferansiyel hata modelleri de yazılabilir. Basit olarak $Y_i = 1$ olan birimler $Y_i = 0$ olan birimlere göre W_{ij} 'de ortalama değişime sahipse aşağıdaki model ile ifade edilebilir (Keogh ve White, 2014):

$$W_{ij} = \psi + \theta X_i + \lambda Y_i + \varepsilon_{ij}, \quad j=1, 2; Y_i = 0, 1$$

4. ÖLÇÜM HATALI AÇIKLAYICI DEĞİŞKENLİ COX REGRESYON MODELİ İÇİN BAYESÇİ YAKLAŞIM

Bayesci çıkarsama, bir veri kümesine bir olasılık modelinin uyarlanması ve model parametrelerinin bir olasılık dağılımı ile sonuçlarının özetlenmesi sürecidir. Bayesci yöntemlerin temel özelliği, istatistiksel veri analizine dayanan çıkarsamalardaki belirsizliğin ölçülmesi için olasılığın kullanımıdır (Gelman ve ark., 2013).

Bayesci yöntemleri kullananların artışı ve hesaplamaları kolaylaştıran teknolojik gelişmelerin artışıyla Bayesci yaklaşım, epidemolojik ve klinik çalışmaların da dahil olduğu bir çok farklı alanda artarak kullanılmaya başlanmıştır (Lesaffre ve Lawson, 2012).

Bayesci çözümleme üç temel adımda tanımlanabilir. İlk adım tam olasılık dağılımının belirlenmesi, gözlenebilen ve gözlemlenemeyen tüm bilgiden birleşik bir olasılık dağılımının oluşturulmasıdır. Burada model, temelde yatan bilimsel sorundan ve veri toplama sürecinden gelen bilgi ile tutarlı olmalıdır. İkinci adım, uygun sonsal dağılımın hesaplanması ve yorumlanmasıdır. Sonsal dağılım, verilen gözlemlenen veriye bakıldığında gözlemlenememiş nihai bilginin koşullu olasılık dağılımıdır. Üçüncü adım ise modelin uygunluğunun ve elde edilen sonsal dağılımın etkilerinin değerlendirilmesidir. Modelin verilere ne kadar uyduğu, sonuçların mantıklı olup olmadığı ve birinci adımdaki model varsayımlarına sonuçların ne kadar hassas olduğu kontrol edilir. Bu soruların cevaplarına göre model değiştirilebilir ya da genişletilebilir ve bu üç adım tekrar uygulanabilir (Gelman ve ark., 2013).

Bayesci çözümlemede veri ve parametre vektörü θ rastgele değişkenlermiş gibi düşünülür. Verinin ürettiği olasılıksal modelin yanında önsel bilgiyi ifade eden $p(\theta)$ önsel dağılımı tanımlanır. Önsel dağılım, D toplanan veriyi göstermek üzere $L(\theta) = L(\theta|D)$ olan olabilirlik fonksiyonu ile birleştirilir. Parametrelerin sonsal dağılımı olan $p(\theta|D)$, Bayes kuralına göre,

$$p(\theta | D) = \frac{L(\theta)p(\theta)}{\int_{\Theta} L(\theta)p(\theta)d\theta} \propto L(\theta)p(\theta)$$

biçimindedir.

Bayesci çözümleme sonsal dağılım olarak ortaya çıkan olasılık dağılımlarının bir ölçümünü ya da beklenen değerinin hesaplanmasını gerektirir. Sonsal ortalama ya da sonsal varyans olarak hesaplanan standart Bayes tahminleri birer beklenen değerdir. Ancak sonsal dağılımlar her zaman bilinen bir olasılık dağılımı olarak elde edilemediğinden Bayesci çözümlemeler için özel teknikler gerekir. İstatistiksel örnekleme, kapalı formda elde edilemeyen bir beklenen değer sayısal integrasyonu ya da analitik çözümü için alternatif bir yöntemdir. Kitle ortalamasının ya da kitle oranının tahmini için bu kitleden çekilen büyük bir örneklemden elde edilen örnekleme ortalaması ya da örnekleme oranının kullanılması istatistiksel yaklaşımlardan biridir. Büyük sayılar kanunu, yeterli büyüklükte alınan örneklemeler için bu şekilde elde edilen tahminlerin uygun olacağını garanti eder (Ghosh, Delampady ve Samanta, 2007). Bu amaçla Monte Carlo Markov Zinciri (MCMC) teknikleri analitik çözümü olmayan istatistiksel modelleme problemlerinin üstesinden gelmek için örnekleme algoritmalarının bir sınıfı olarak ortaya çıkmıştır. Gibbs örnekleme ve Metropolis-Hastings algoritması iki önemli MCMC tekniğidir (Lesaffre ve Lawson, 2012).

Birçok çok boyutlu problemde kullanışlı bir MCMC algoritması Gibbs örneklemesidir. θ parametre vektörü $\theta = (\theta_1, \dots, \theta_d)$ olacak şekilde d bileşene bölünür. Verilen başlangıç noktası $\theta^0 = (\theta_1^0, \dots, \theta_d^0)^T$ olan Gibbs örnekleme aşağıdaki iteratif şemaya sahiptir ve (k+1) iterasyonunda aşağıdaki d adım gerçekleşir.

1. $p(\theta_1 | \theta_2^k, \dots, \theta_{(d-1)}^k, \theta_d^k, y)$ koşullu dağılımından $\theta_1^{(k+1)}$ örneklenir.
2. $p(\theta_2 | \theta_1^{(k+1)}, \theta_3^k, \dots, \theta_d^k, y)$ koşullu dağılımından $\theta_2^{(k+1)}$ örneklenir.
- $\vdots$
- d. $p(\theta_d | \theta_1^{(k+1)}, \dots, \theta_{(d-1)}^{(k+1)}, y)$ koşullu dağılımından $\theta_d^{(k+1)}$ örneklenir.

Burada θ_j diğer tüm parametrelere bağlı koşullu olarak elde edildiği için $p(\theta_j | \theta_1^k, \theta_2^k, \dots, \theta_{(j-1)}^k, \theta_{(j+1)}^k, \dots, \theta_{(d-1)}^k, \theta_d^k, y)$ koşullu dağılımı, tam koşullu dağılım olarak adlandırılır. Üretilen $\theta^k, \theta^{k+1}, \dots$ sonsal dağılımdan elde edilen gözlemler olarak kabul edilebilir.

Gibbs örneklemesinde tam koşullu dağılımlar birleşik dağılımı oluşturur (Lesaffre ve Lawson, 2012).

MCMC sonsal dağılımdan örneklemeler üretir. Burada üretilen örneklemeler ile istenen sonsal dağılıma ulaşıp ulaşılmadığını belirlemek için yakınsama ölçütleri kullanılır. Yakınsama ölçütleri durağan bir dağılımdan tahminleri elde edebilmek için MCMC'nin ne kadar uzunlukta çalışacağını ve oluşturulan zincirlerin yeteri kadar karışıp karışmadığını değerlendiren yöntemlerdir. Bu amaçla birçok istatistiksel yöntem önerilmiştir. İz Grafiği (trace plot), otokorelasyon grafiği gibi grafik yöntemlerinin yanında, Gelman-Rubin, Geweke, Heidelberger-Welch, Raftery-Lewis testleri gibi test yöntemleri yakınsamanın sağlanıp sağlanmadığının kontrolünde kullanılan yöntemlerdir (Lesaffre ve Lawson, 2012).

4.1. Cox Regresyon Modelinde Bayesci Çözümleme

Esnek model yapısının ve Bayesci hesaplamalarda mevcut yazılım programlarının artan avantajıyla, Bayesci yöntemler son zamanlarda yaşam verilerinin çözülmesinde de tercih edilmeye başlamıştır. CRM için kullanılan başlıca Bayesci süreçler Gamma süreci, Beta süreci, İlişkili Gamma süreci ve Dirichlet sürecidir. Süreç genel olarak regresyon katsayıları ve temel tehlike fonksiyonu ya da birikimli temel tehlike fonksiyonu için önsel dağılımın belirlenmesi, olabilirlik fonksiyonunun oluşturulması, sonsal dağılımın elde edilmesi ve çıkarımların yapılması şeklindedir (İbrahim, Chen ve Sinha, 2001).

CRM için sıklıkla kullanılan süreçlerden biri Gamma sürecidir. $G(\alpha, \lambda)$, $\alpha > 0$ şekil parametresi ve $\lambda > 0$ ölçek parametresi ile Gamma dağılımını gösterebilir. $\alpha(t)$, $t \geq 0$ $\alpha(0) = 0$ olan soldan artan sürekli fonksiyon, $Z(t)$, $t \geq 0$ aşağıdaki özelliklerle stokastik bir süreç olsun.

- (i) $Z(0) = 0$,
- (ii) $Z(t)$ ayrık aralılarda bağımsız,
- (iii) $t > s$ için $Z(t) - Z(s) \sim G(c(\alpha(t) - \alpha(s)), c)$.

Gamma süreci $Z(t) \sim GP(c\alpha(t), c)$ ile gösterilir. Burada c , ortalama için ağırlık ya da güven parametresini göstermektedir.

CRM için X açıklayıcı değişkenler vektörüne sahip n birimin yaşamlarına ait birleşik olasılık,

$$P(T > t | X, \beta, H_0) = \exp \left\{ - \sum_{j=1}^n \exp(x_j' \beta) H_0(t_j) \right\} \quad (4.1)$$

biçiminde verilsin.

İlk yaklaşım olarak Gamma süreci birikimli temel tehlike fonksiyonu $H_0(t)$ için önsel dağılım olarak kullanılır. $H_0(t) \sim GP(c_0 H^*, c_0)$ sürecine sahiptir ve $H^*(t)$, $H^*(0)=0$ olan artan bir fonksiyondur.

H^* hiperparametre vektörü γ_0 ile bilinen parametrik bir fonksiyon olduğu varsayılır. H^* üstel dağılım gösteriyorsa $H^*(t) = \gamma_0 t$ ile ifade edilir ve γ_0 belirlenmiş bir hiperparametredir. H^* Weibull dağılımlı alınırsa, $H^*(t) = \eta_0 t^{k_0}$ olur ve burada hiperparametrelerin belirlenen vektörü $\gamma_0 = (\eta_0, k_0)'$ biçimindedir. $A_j = \sum_{\ell \in R_j} \exp(x_\ell' \beta)$ olmak üzere marjinal yaşam fonksiyonu,

$$P(T > t | \beta, X, \gamma_0, c_0) = \prod_{j=1}^n \left[\phi(iA_j) \right]^{c_0 (H^*(t_{(j)}) - H^*(t_{(j-1)}))}$$

biçimindedir.

İkinci yaklaşımda ise tehlike oranı üzerinden gamma süreci önsel dağılım olarak kullanılır. Bu yaklaşımda genişletilmiş gamma sürecinin kesikli yaklaşımı düşünülür. Burada parçalı sabit tehlike fonksiyonunun olabilirlik fonksiyonu kullanılarak çözümleme geliştirilmiş ve yalnızca başarısızlık süresinin hangi aralığa düştüğü bilgisi kullanılır. $0 = s_0 < s_1 < \dots < s_J$ sonlu olmak üzere,

$$\delta_j = h_0(s_j) - h_0(s_{j-1})$$

olsun ve $(s_{j-1} - s_j]$ ($j = 1, 2, \dots, J$) aralığında temel tehlikedeki artışı gösterebilir.

Kitledeki bir birim için t zamanındaki yaşam fonksiyonu,

$$S(t | x) = \exp \left\{ \eta \int_0^y h_0(u) du \right\} \approx \exp \left\{ -\eta \left(\sum_{i=1}^J \delta_i (t - s_{i-1})^+ \right) \right\}$$

biçimindedir. Burada $h_0(0)=0$, $\eta = \exp(x'\beta)$ ve $(u)^+=u$ $u>0$ ise, aksi takdirde 0'dır.

p_j , $(s_{j-1} - s_j]$ aralığında başarısızlık olasılığını göstermek üzere,

$$p_j = S(s_{j-1}) - S(s_j) \approx \exp \left\{ -\eta \sum_{\ell=1}^{j-1} \delta_\ell (s_{j-1} - s_{\ell-1}) \right\} \left[1 - \exp \left\{ -\eta (s_j - s_{j-1}) \right\} \sum_{\ell=1}^J \delta_\ell \right]$$

eşitliği ile elde edilir. j . aralık $(s_{j-1} - s_j]$ 'de bir başarısızlığın olabilirlik fonksiyonuna katkısı, p_j ve $S(s_j)$ 'dir. d_j başarısızlık sayısı, D_j başarısız birimlerin kümesi, c_j sağdan durdurulmuş gözlemlerin sayısı, C_j durdurulmuş birimlerin kümesi olsun. $D = (n, t, X, \delta)$ veriyi gösterebilir.

$D_0 = (n_0, t_0, X_0, v_0)$ daha önceki çalışmalardan elde edilen yaşam verisini gösterebilir. $\pi_0(\beta, \delta)$, (β, δ) için başlangıç önselini gösterebilir. (β, δ) için önsel dağılım,

$$\pi(\beta, \delta | D_0, a_0) \propto \{L(\beta, \delta | D_0)\}^{a_0} \pi_0(\beta, \delta)$$

biçimindedir. Burada $L(\beta, \delta | D_0)$ geçmiş veriye bağlı olabilirlik fonksiyonunu göstermektedir.

Önsel belirlemeyi basitleştirmek için c_0 ve θ_0 sabit hiperparametreler olmak üzere,

$$\pi_0(\beta, \delta) = \pi_0(\beta | c_0) \pi_0(\delta | \theta_0)$$

alınır. Burada $\pi_0(\beta | c_0)$ p boyutlu, 0 ortalama ve $c_0 W_0$ varyans-kovaryans matrisine sahip çok değişkenli normal dağılım $N_p(0, c_0 W_0)$ alınır. $\pi_0(\delta | \theta_0)$ ise f_{0i}/g_{0i} ortalamalı, f_{0i}/g_{0i}^2 varyanslı M bağımsız gamma yoğunluğunun çarpımına eşittir ve aşağıdaki gibi gösterilir:

$$\pi_0(\delta | \theta_0) \propto \prod_{i=1}^M \delta_i^{f_{0i}-1} \exp(-\delta_i g_{0i})$$

Burada $\theta_0 = (f_{01}, g_{01}, \dots, f_{0M}, g_{0M})'$ biçimindedir. Önsel belirleme işlemi a_0 için önselin belirlenmesi ile son bulur. a_0 ($0 \leq a_0 \leq 1$) için Beta önseli,

$$\pi(a_0 | \alpha_0, \lambda_0) \propto a_0^{\alpha_0-1} (1-a_0)^{\lambda_0-1}$$

biçiminde tanımlanır.

Son olarak birleşik önsel aşağıdaki gibi yazılabilir:

$$\pi(\beta, \delta, a_0 | D_0) \propto \{L(\beta, \delta | D_0)\}^{a_0} \pi_0(\beta | c_0) \pi_0(\delta | \theta_0) \pi(a_0 | \alpha_0, \lambda_0)$$

Birleşik önselin kapalı bir formda elde edilmesi ya da hesaplanabilmesi mümkün olmayabilir. Bunun yerine temel tehlike oranı ve regresyon katsayıları arasında bağımsızlık varsayılır. Bu durumda (β, δ) için birleşik olasılık yoğunluk fonksiyonu,

$$\pi(\beta, \delta | D_0) = \pi(\beta | D_0) \pi(\delta | D_0)$$

ile elde edilir. β için çok boyutlu normal dağılım önseli,

$$\pi(\beta | a_0, \mu_0, I_0) \sim N_p(\mu_0, a_0^{-1} I_0^{-1})$$

biçimindedir. Burada μ_0 , önsel ortalamayı göstermektedir. μ_0 , D_0 veri seti kullanılarak β için kısmi olabilirlikten elde edilen çözümdür. I_0 ise kısmi olabilirlik fonksiyonuna bağlı bilgi matrisidir ve aşağıdaki gibi elde edilir:

$$I_0 = \left[\frac{\partial^2}{\partial \beta \partial \beta'} \log(PL(\beta | D_0)) \right] \bigg|_{\beta=\mu_0}$$

Süreç boyunca temel amaç, β için çıkarımlar yapmaktır. δ parametresi için bilgi içermeyen önsel dağılım kullanılması uygun olabilmektedir. İkinci bir yaklaşım da $\pi(\delta | a_0, D_0)$ ortalaması ϕ_i ve varyansı $a_0^{-1} \gamma_i$ olan M bağımsız gamma yoğunluğunun çarpımı olarak belirlenebilir. $\hat{\delta}_{0i}$, D_0 verisinden elde edilen en çok olabilirlik tahmin edicisi olmak üzere $\phi_i = \hat{\delta}_{0i}$ ve $\gamma_i = \text{Var}(\delta_{0i})$ olarak alınabilir (İbrahim, Chen ve Sinha, 2001).

Bayesci çözümlemenin olabilirlik fonksiyonuyla olan ilişkisine bakıldığında $(0, s_1], (s_1, s_2], \dots, (s_{j-1}, s_j]$ ayrık zaman aralıkları, $t_{(1)}, t_{(2)}, \dots, t_{(n)}$ sıralı başarısızlık ya da durdurulmuş süreler olmak üzere $H_0(t) \sim GP(c_0 H^*, c_0)$ sürecine sahip olsun. $h_j = H_0(t_{(j)}) - H_0(t_{(j-1)})$ olduğundan,

$$h_j \sim G(c_0 h_{0j}, c_0)$$

olur. Burada $h_{0j} = H^*(t_{(j)}) - H^*(t_{(j-1)})$ 'dir. $A_j = \sum_{\ell \in R_j} \exp(x_{\ell} \beta)$ ve E_{GP} ; gamma süreci önseline göre

beklenen değeri gösterebilir. Eşitlik (4.1) aynı zamanda

$$P(T > t | X, \beta, H_0) = \exp \left\{ - \sum_{j=1}^n h_j \sum_{\ell \in R_j} \exp(x_{\ell}' \beta) \right\}$$

biçiminde yazılabilir. Bu durumda,

$$E_{GP}[P(T > t | X, \beta, H_0) | H^*] = \prod_{j=1}^n \left(\frac{c_0}{c_0 + A_j} \right)^{c_0 h_{0j}} = \prod_{j=1}^n \exp \left\{ c_0 H^*(t_{(j)}) \log \left(1 - \frac{\exp(x_j' \beta)}{c_0 + A_j} \right) \right\}$$

eşitliğindeki gibi elde edilir. Temel tehlike fonksiyonu,

$$h_0(t) = \frac{d}{dt} H^*(t)$$

olmak üzere, $\theta = (\beta', h_0, c_0)'$ olsun.

Olabilirlik fonksiyonu;

$$\begin{aligned} L(\theta | D) &= \prod_{j=1}^n \exp \left\{ c_0 H^*(t_{(j)}) \log \left(1 - \frac{\exp(x_j' \beta)}{c_0 + A_j} \right) \right\} \times \left\{ -c_0 \frac{dH^*(t_{(j)})}{dt_{(j)}} \log \left(1 - \frac{\exp(x_j' \beta)}{c_0 + A_j} \right) \right\}^{\delta_j} \\ &= \prod_{j=1}^n \exp \left\{ H^*(t_{(j)}) \log \left(1 - \frac{\exp(x_j' \beta)}{c_0 + A_j} \right)^{c_0} \right\} \times \left\{ -c_0 h_0(t_{(j)}) \log \left(1 - \frac{\exp(x_j' \beta)}{c_0 + A_j} \right) \right\}^{\delta_j} \end{aligned}$$

biçimindedir.

Bu durumda aşağıdaki limit durumu elde edilir:

$$\lim_{c_0 \rightarrow 0} \exp \left(H_0(t_{(j)}) \log \left(1 - \frac{\exp(x'_j \beta)}{c_0 + A_j} \right)^{c_0} \right) = 0 \quad j=1, 2, \dots, n \quad (4.2)$$

Burada,

$$\lim_{c_0 \rightarrow 0} \log \left(1 - \frac{\exp(x'_j \beta)}{c_0 + A_j} \right) = \log \left(1 - \frac{\exp(x'_j \beta)}{A_j} \right) \approx - \frac{\exp(x'_j \beta)}{A_j} \quad j=1, 2, \dots, n-1 \quad (4.3)$$

yazılabilir. Eşitlik (4.2) ve Eşitlik (4.3)'ten

$$\lim_{c_0 \rightarrow 0} \frac{L(\theta | D)}{c_0^d \log(c_0) h^*} \approx \prod_{j=1}^n \left(\frac{\exp(x'_j \beta)}{A_j} \right)^{\delta_i} \quad (4.4)$$

yazılabilir. Burada

$$d = \sum_{i=1}^n \delta_i$$

durdurma sayısını,

$$h^* = \sum_{j=1}^n (h_0(t_{(j)}))^{\delta_i}$$

eşitliğini göstermektedir.

Eşitlik (4.4)'ün sağ yanına bakıldığında Cox kısmi olabilirlik fonksiyonu görülebilir. $c_0 \rightarrow \infty$ durumunda (β, h_0) 'a bağlı olabilirlik fonksiyonu;

$$\lim_{c_0 \rightarrow \infty} \left[\exp \left\{ H^*(t_{(j)}) \log \left(1 - \frac{\exp(x'_j \beta)}{c_0 + A_j} \right)^{c_0} \right\} \times \left\{ -c_0 h_0(t_{(j)}) \log \left(1 - \frac{\exp(x'_j \beta)}{c_0 + A_j} \right) \right\}^{\delta_i} \right]$$

$$= \exp \{ -H^*(t_{(j)}) \exp(x'_j \beta) \} \{ h_0(t_{(j)}) \exp(x'_j \beta) \}^{\delta_i}$$

biçimdedir. Buradan,

$$\lim_{c_0 \rightarrow \infty} L(\beta, c_0, h_0 | D) = \prod_{j=1}^n \left(\exp \{ -H^*(t_{(j)}) \exp(x'_j \beta) \} \{ h_0(t_{(j)}) \exp(x'_j \beta) \}^{\delta_i} \right) \quad (4.5)$$

yazılabilir. Eşitlik (4.5)'in orantılı tehlikeler için (β, h_0) 'ın olabilirlik fonksiyonuna eşit olduğu görülmektedir.

CRM'de sayma süreci, t zamanına kadar gerçekleşen olay sayısını sayar, $dN_i(t)$, $[t, t+\Delta t)$ aralığında $N_i(t)$ 'deki küçük bir artışı göstermektedir. $N_i(t)$ ve $dN_i(t)$ $[0, t)$ aralığında olay gerçekleşmişse 1, gerçekleşmemişse 0 değerini almaktadır. $dN_i(t)$ tüm i ve t 'ler için en az 1 değerini alıyor olsa da, $dN_i(t)$, $[t, t+\Delta t)$ aralığında olabilirlik fonksiyonuna katkısı $I_i(t)dt$ ortalamalı bağımsız Poisson rastgele değişkeni olarak düşünülür. Modeli bu şekilde tanımlamak, bu aralıkta yoğunluğun sabit olmasını sağlar (Clayton, 1991). CRM'de Bayesci çözümlemede etkili sonsal hesaplamalar için Eşitlik (2.3)'de verilen olabilirlik fonksiyonu altında $dN_i(t)$,

$$dN_i(t) \sim \text{Poisson}(I_i(t)dt)$$

biçimindedir. Veri (D) elde edildikten sonra, sonsal dağılım olan $P(\beta, H_0(t) | D)$ olasılığının elde edilmesine geçilir. Bayes teoreminden,

$$P(\beta, H_0(t) | D) \propto P(D, \beta, H_0(t))$$

yazılabilir. Bayesci yaklaşım ile bileşik sonsal dağılım aşağıdaki gibi yazılabilir:

$$P(\beta, H_0(t) | D) \propto L(D | \beta, H_0(t))P(\beta)P(H_0(t)) \quad (4.6)$$

Burada öncelikli olarak ilgilenilen parametre β 'dir. Diğer bir parametre ise $H_0(t)$ 'dir.

4.2. Ölçüm Hatası Modellerinde Bayesci Çözümleme

Bayesci yaklaşımlar, önsel dağılım ve uzman bilgisini de dahil ederek gerçek değerleri gözlenemeyen açıklayıcı değişkenlerin ölçüm hatasını hesaba katarak çözümlemeye esnek bir çerçeve sağlamaktadır (Muff ve ark., 2015). Son yıllarda MCMC yöntemlerindeki gelişmeyle birlikte ölçüm hatası problemlerinde de Bayesci yaklaşımın kullanımı artış göstermiştir.

Ölçüm hatası modellerinde Bayesci çözümlemelerde ilk adım olabilirlik yaklaşımında olduğu gibi verinin her bir bileşeni için parametrik bir model belirlenir ve olabilirlik çözümlemesi

gerçek X_i değerleri gözlenebilirmiş gibi modelde kullanılır. İkinci adımda ise ölçüm hatası modelinin belirlenmesi yer alır. Klasik ölçüm hatası modeli mi, Berkson hata modeli mi yoksa her ikisinin kombinasyonu olan bir model mi kullanılacağına karar verilir ve gözlenen modeldeki diğer Z açıklayıcı değişkenler koşulunda gözlemlenemeyen X açıklayıcı değişkenleri için bir dağılım belirlenir. Üçüncü adımda klasik Bayesci çözümlemede X kayıp gözlem gibi alınır ve diğer tüm değişkenler göz önüne alındığında X 'in koşullu dağılımından yararlanılarak birden çok kez yerine koyularak işlem yapılır. W 'yi içeren tüm verinin olabilirliği X ulaşılabilmemiş gibi elde edilir. Dördüncü adımda, Bayesci yaklaşımda klasik olabilirlik yöntemlerinden temel farklılık olarak parametreler rastgeleymiş gibi ele alınır. Bu da modelde önsel dağılım olarak adlandırılır. Son adımda ise gözlenen tüm verilerden parametrelerin sonsal dağılımları elde edilerek Bayesci sonuçlar yorumlanır (Carroll ve ark., 2006).

Ölçüm hatası modeli verilen gerçek değişkenin yanlış ölçümü için bir modeldir ve aşağıdaki gibi üretilir:

$$W_i | X_i \sim N(X_i, \sigma_e^2)$$

Bu klasik ölçüm hatası modelinin bir sonucudur. Açıklayıcı değişken ölçüm hatası içeriyorsa $E(W_i | x_i) = x_i$ olur. Ölçüm hatalı değişken W_i , gerçek ölçüm değişkeni olan X_i 'nin yansız bir tahmin edicisidir (Weeding, 2016). Ölçüm hatası modellerinde en büyük zorluk ölçüm hatası varyansının tahminidir (Muff ve ark., 2015). Ölçüm hatalı değişken için tekrarlı gözlem varsa bu bilgi modelden tahmin edilebilen ölçüm hatası varyansı için önsel bir dağılım üretmekte kullanılabilir. Ancak bu yalnızca Bayesci çözümleme için geçerlidir. Diğer düzeltme yöntemleri ölçüm hatası varyansının bilinmesini ya da tekrarlı gözlemlerden tahmin edilmesini gerektirir. Gözlenen bir birim için olabilirlik,

$$f(Y, W, X | Z, \Omega) = f_{Y|Z, X}(Y | Z, X, \beta) \times f_{W|Z, X}(W | Z, X, \alpha_1) f_{X|Z}(X | Z, \alpha_2)$$

biçiminde gösterilebilir ve burada Ω bütün bilinmeyen parametreleri göstermektedir. Ölçüm hatası modellerinin Bayesci çözümlemesi için dördüncü adım buradaki Ω için $\pi(\Omega)$ biçiminde gösterilen önsel bir dağılım tanımlanır. Sonrasında tüm verinin olabilirliği,

$$\pi(\Omega) \prod_{i=1}^n f(Y_i, W_i, X_i | Z_i, \Omega)$$

biçiminde ifade edilebilir. $f(Y, W, X | Z, \Omega) = \prod_{i=1}^n f(Y_i, W_i, X_i | Z_i, \Omega)$ olmak üzere Ω 'nın sonsal dağılımı ise,

$$f(\Omega | Y, W, Z) = \frac{\pi(\Omega) f(Y, W, X | Z, \Omega) d\Omega}{\int \pi(\omega) f(Y, W, X | Z, \omega) d\omega} \quad (4.7)$$

biçimindedir.

Ancak Eşitlik (4.7)'nin integrasyonu Ω az sayıda bilinmeyen parametre içeriyor dahi olsa analitik olarak mümkün değildir. Gibbs örnekleme bu sorun için kullanılan çözümlerden biridir. Gibbs örnekleme Ω 'nın başlangıç değerleri ile başlayan, aşağıdaki adımlardan oluşan yinelemeli MCMC yöntemidir. İlk adım olarak gözlemlenmemiş X değerlerinin bir örnekleme, şu anki değeri verilen Ω 'nın sonsal dağılımlarından örnekleyerek oluşturulur ve X_i 'nin sonsal dağılımı,

$$f(X_i | Y_i, W_i, Z_i, \Omega) = \frac{f(Y_i, W_i, X_i | Z_i, \Omega)}{\int f(Y_i, W_i, x | Z_i, \Omega) dx}$$

biçimindedir. İkinci adımda, sonsal dağılımından Ω 'nın yeni bir değeri üretilir ve üretilen X değeri aşağıdaki gibidir:

$$f(\Omega | Y, W, Z, X) = \frac{\pi(\Omega) f(Y, W, X | Z, \Omega)}{\int \pi(\omega) f(Y, W, X | Z, \omega) d\omega}$$

Genellikle bir kerede diğerleri sabit tutularak Ω 'nın içerdiği yalnız bir parametre için bu işlemler yapılır. Ω 'nın j . değeri ω_j olsun, Ω 'nın diğer bileşenleri $\Omega_{(-j)}$ ile gösterilsin. Bu durumda sonsal dağılım,

$$f(\omega_j | Y, W, Z, X, \Omega_{(-j)}) = \frac{\pi(\omega_j, \Omega_{(-j)}) f(Y, W, X | Z, \omega_j, \Omega_{(-j)})}{\int \pi(\omega_j^*, \Omega_{(-j)}) f(Y, W, X | Z, \omega_j^*, \Omega_{(-j)}) d\omega_j^*}$$

biçimindedir. Üçüncü adımda bu işlem ilk üretilen bir kısım örneklem yakılmak (burn in) üzere yakınsama sağlanana kadar birçok kez tekrarlanır. Son adımda ise sonsal ortalamadan tahmin edilmek istenen parametre değeri elde edilir (Carroll ve ark., 2006).

4.3. Ölçüm Hatalı Açıklayıcı Değişkenli Cox Regresyon Modelinin Düzeltilmiş Gamma Önselli Bayesci Çözümlemesi

X_i ölçüm hatası içeren açıklayıcı değişken, Z_i ölçüm hatası içermeyen açıklayıcı değişken, T_i gözlenen yaşam süresi, δ_i durum göstergesi, $Y_i=(T_i, \delta_i)$ bileşenlerini içeren çıktı değişkeni, β regresyon katsayıları ve $H_0(t)$ birikimli temel tehlike fonksiyonu olmak üzere $(Y_i, X_i, W_{i1}, W_{i2} | Z_i)$ için birleşik parametrik model,

$$f(Y_i | X_i, Z_i, \beta, H_0(t)) f(W_i | X_i, \sigma_\epsilon^2) f(X_i | Z_i, \gamma)$$

olarak ifade edilebilir. Burada ilk bileşen öncelikli amaç olan $\beta=(\beta_x, \beta_z)$ regresyon parametrelerini ve birikimli temel tehlike fonksiyonunu içerir. İkinci bileşen klasik (toplamsal) hata modeline sahip ve hataların 0 ortalama, σ_ϵ^2 varyansı ile normal dağılım gösterdiği, ölçüm hatası için yazılan kısmı ifade eder. Son bileşen ise Z_i koşulu altında gözlemlenemeyen X_i açıklayıcı değişkeni için tanımlanan kısımdır. Üç alt model parametreleri için önsellerin bağımsızlığı varsayılır. Burada, Z_i 'nin dağılımını modellemekten kaçınmak için tüm Z_i gözlemlerinin elde edildiği, hatanın Y_i ve Z_i 'ye göre türevlenemediği varsayılır (Bartlett ve Keogh, 2018).

Burada ilk kısım için Eşitlik (4.6)'da öncelikli olarak ilgilenilen parametre β , diğer bir parametre ise $H_0(t)$ 'dir. β için $P(\beta | H_0(t), D)$, $H_0(t)$ için $P(H_0(t) | D, \beta)$ tam koşullu sonsal dağılımların elde edilmesi için Gibbs örnekleme uygulanması gerekir. Winbugs, R gibi programlar otomatik olarak bu dağılımların elde edilmesini sağlamaktadır. Eşitlik (4.6)'nın farklı Bayesci çözümlemelerinde $H_0(t)$ 'nin parametrik olmayan tanımlaması farklılık gösterir. Bayesci çözümlemelerin birçoğunda $H_0(t)$ için Gamma süreci önsel dağılım olarak tanımlanmıştır (Bartlett ve Keogh, 2018).

Bu tez çalışmasında, ölçüm hatası için düzeltilmiş CRM’de $H_0(t)$ için Bayesci çözümlemede kullanılan gamma önsel dağılımı yerine poligonal fonksiyonun kullanımı alternatif bir yöntem olarak düşünülmüştür.

4.4. Ölçüm Hatalı Açıklayıcı Değişkenli Cox Regresyon Modelinin Düzeltilmiş Poligonal Önselli Bayesci Çözümlemesi

Beamonte ve Bermúdez (2003), CRM’nin parametrik olmayan kısmı olan temel tehlike fonksiyonu $h_0(t)$ için köşe noktaları $a_0 = 0 < a_1 < \dots < a_T < a_{T_{\max}+1}$ zaman noktalarında olacak şekilde negatif olmayan poligonal fonksiyonu kullanmışlardır. Burada poligonal $\tau_0 = 0 < \tau_1 < \dots < \tau_T < \tau_{T_{\max}+1}$ değerlerini almaktadır ve $a_{T_{\max}+1}$ zaman noktasından sonra sabit olacak biçimde ele alındığında temel tehlike fonksiyonu aşağıdaki gibi yazılabilir:

$$h_0(t) = \begin{cases} \tau_j + \frac{(\tau_{j+1} - \tau_j)(t - a_j)}{a_{j+1} - a_j} & \text{if } a_j \leq t \leq a_{j+1}, j = 1, \dots, T_{\max} \\ \tau_{T_{\max}+1} & \text{if } t \geq T_{\max} + 1 \end{cases} \quad (4.8)$$

Temel tehlike fonksiyonu için Eşitlik (4.8)’in kullanılması ile Eşitlik (2.2)’deki $I_i(t)$ sürekli olacaktır.

Gamma sürecinde birikimli tehlike artışlarının bağımsız olduğu varsayılır. Ancak bu birçok uygulamada gerçekçi olmayabilir. Ayrıca bitişik aralıklar arası etkinliği yansıtmaya izin vermez. Birinci dereceden otokorelasyon süreci olarak tanımlanabilen τ vektörü için önsel belirlenerek düzeltirme fikri Gamerman (1991) tarafından benzer bir çalışmada önerilmiştir. Gamerman’nın (1991) çalışmasında τ vektörü,

$$\tau_{j+1} = \tau_j \exp(e_j), \quad j = 1, \dots, T_{\max}$$

biçiminde tanımlanır. Burada,

$$\tau_1 \sim \text{Gamma}(\tau_1 | a_\tau, b_\tau)$$

dağılımına sahiptir.

τ vektöründeki $(e_1, \dots, e_{T_{\max}})$, sıfır ortalama ve σ_e^2 varyansı ile bağımsız, log-normal dağılıma sahip olduğu varsayılır. Burada varyans terimi,

$$\sigma_e^2 \sim \text{Gamma}(\text{tau.e} | a_e, b_e)$$

dağılımı göstermektedir ve

$$\text{tau.e} = 1 / \sigma_e^2$$

olarak tanımlanır.

Burada $a_e = b_e = 0.001$ ve $a_\tau = b_\tau = 0.01$ alınması τ_1 parametresi için önselin 1 ortalama ve 500 varyansa sahip olmasını gösterir (Mostafa ve Ghorbal, 2011).

İlk olarak Gamerman (1991) tarafından önerilen τ vektörü için önsel belirlenerek düzleştirme fikri, Beamonte ve Bermúdez (2003) tarafından geliştirilerek temel tehlike fonksiyonu için tanımlanan poligonal fonksiyonun, birikimli tehlike artışlarının bağımsızlığını varsayan gamma süreci yerine kullanılması, ölçüm hatası içeren CRM’de kullanılarak ölçüm hatası düzeltilmesinde kullanılan Bayesci yaklaşımlara alternatif bir yöntem olarak önerilmiştir. Önerilen yeni yöntem ile ölçüm hatalı açıklayıcı değişken içeren CRM’de beta için daha az yanlı katsayı kestirimlerinin elde edilmesi hedeflenmektedir.

5. BENZETİM ÇALIŞMASI VE UYGULAMA

Bu bölümde ölçüm hatalı açıklayıcı değişkenli CRM’de literatürde verilen ve önerilen yöntemlerin karşılaştırılması için uygulanan benzetim süreci hakkında bilgi verilmiş ve sonuçları sunulmuştur. Ardından literatürde yer alan 1974 ve 1984 yılları arasında yapılan Mayo klinik çalışmasına ait verilerle ve Goldman ve ark.’nın (1996) çalıştığı AIDS veri setinde modellerin sonuçları elde edilerek karşılaştırılmış ve yorumlanmıştır.

5.1. Benzetim Çalışması

T_i , yaşam süresi, orantılı tehlikeler varsayımını sağlama koşulu altında CRM’den, belirlenen kitle parametrelerine göre Bender, Augustin ve Bettner (2005) tarafından önerilen fonksiyon yardımı ile üretilmiştir. X_i ölçüm hatası içeren açıklayıcı değişkenin gerçek değeri ve Z_i ölçüm hatası içermeyen açıklayıcı değişken olmak üzere ortalaması 0, varyansı 1 olan çok değişkenli normal dağılımdan, aralarındaki kovaryansları 0.25 olacak şekilde üretilmiştir. Örneklem büyüklüğü (n) 30, 50 ve 100 için senaryolar oluşturulmuştur. β_X , $\beta_Z = 0.5, 1.0, 1.5$ ve 0.10 (düşük), 0.30 (orta) ve 0.60 (yüksek) durdurma oranları için senaryolar üretilmiştir.

Ölçüm hatalı değişken $W_{i1} = X_i + \varepsilon_{i1}$ eşitliğinden elde edilmiştir. Burada $\varepsilon_{i1} \sim N(0, \sigma_\varepsilon^2)$ dağılımından üretilmiştir. Ölçümlerden rastgele %20’si, aynı hata varyansına sahip olacak şekilde ölçüm hatalı değişkenin ikinci tekrarlı ölçümü olarak elde edilmiştir. Bu hata varyansı koşullu olmayan güvenilirlik $r = \sigma_X^2 / (\sigma_X^2 + \sigma_\varepsilon^2)$ olacak biçimde seçilmiş ve sırasıyla düşük, orta ve yüksek güvenilirliği göstermek üzere 0.5, 0.7 ve 0.9 için sonuçlar elde edilmiştir.

CRM’nin Bayesci çözümlemesinde regresyon parametreleri ve bazı alt model parametreleri için önsel dağılım belirlenmesine ek olarak birikimli temel tehlike fonksiyonu $H_0(t)$ için önsel dağılım belirlemeyi gerektirir. $H_0(t)$ için belirlenen önsel dağılımın β dahil diğer önsel dağılımlardan bağımsız olduğu varsayılır. Burada daha önce önerilen ölçüm hatası için düzeltilmiş Bayesci CRM ve Bayesci CRM’de $H_0(t)$ için Gamma önseli kullanılmıştır. Tez kapsamında önerilen model için ölçüm hatası için düzeltilmiş Bayesci modelde $H_0(t)$ için poligonal önseli kullanılmıştır.

Benzetim sürecinde bütün model parametreleri için Gustafson (2003) tarafından önerilen bilgi içermeyen önseller tanımlanmıştır.

β_x ve β_z için 0 ortalama ve 1.38 varyansı ile bağımsız normal dağılım önseli tanımlanmıştır. Bu şekilde belirlenmesi parametre tahmininin sabitlenmesine (stabilizing) yardımcı olmaktadır (Hamra, MacLehose ve Cole, 2013). Bu önselin kullanımı %95 olasılıkla $\exp(\beta_x)$ ve $\exp(\beta_z)$ 'nin 0.1 ile 10 arasında olduğunu var sayar. γ_z için ise 0 ortalamalı 10000 varyanslı normal dağılım önseli tanımlanmıştır. Tüm varyans parametreleri için ters Gamma dağılımı $IG(0.5, 0.5)$ önseli kullanılmıştır. Bu önsel dağılım ile varyans için en iyi tahmin elde edilebilir (Gustafson, 2003).

Tüm modeller beş paralel zincir ile çalıştırılmıştır. Gelman-Rubin yakınsama testi yakınsama ölçütü olarak kullanılmış ve 1.01 ve daha küçük değeri için zincirlerin yeteri kadar karıştığı varsayılmıştır. Bütün senaryolar için yakınsamanın sağlandığı iterasyon sayısı ile çalışılmıştır. Her bir senaryo için 500 tekrar alınmıştır. Bayesci çıkarsamalar için R paket programında Just Another Gibbs Sampler (JAGS) programı kullanılmıştır.

Kısmi olabilirlik fonksiyonundan elde edilen CRM, Bayesci çözümlemesi (B_CRM), ölçüm hatası için düzeltilmiş Gamma önselli Bayesci CRM (DGB_CRM) ve ölçüm hatası için düzeltilmiş poligonal önselli Bayesci CRM (DPB_CRM) sonuçları karşılaştırılmıştır. Bu yöntemler için tahmin edicilerin performansını karşılaştırmak amacı ile hata kareler ortalaması (HKO) ve kapsama olasılığı (KO) hesaplanmıştır. KO Bayesci yöntemler için üretilen 500 benzetim senaryosunda %95 Sonsal Bayes güven aralıklarının (posterior credible intervals) parametrenin gerçek değerini içerme oranı olarak elde edilmiştir. Burada %95 Sonsal Bayes güven aralıkları 0.025 ve 0.975 sonsal yüzdelerinden hesaplanmıştır. Klasik model için ise benzetim ile üretilen 500 senaryoda parametre için tahmin edilen katsayının güven aralığının gerçek değeri içerme oranı olarak elde edilmiştir.

5.1.1. Benzetim Çalışması Sonuçları

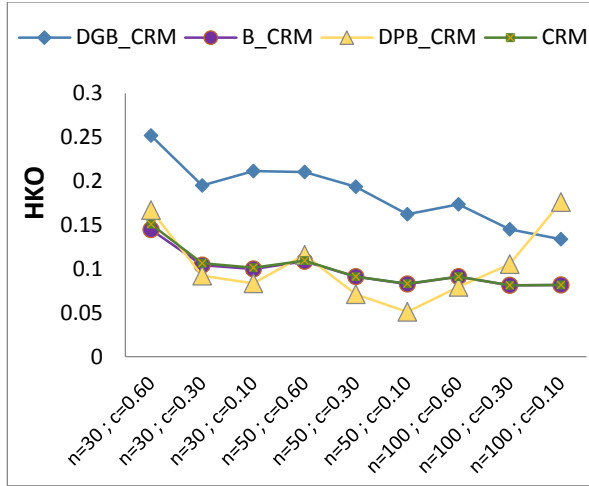
Benzetim çalışmalarının HKO ve KO değerleri $\beta=0.5$ için Çizelge 5.1’de ve Şekil 5.1’de, $\beta=1.0$ için Çizelge 5.2’de ve Şekil 5.2’de ve $\beta=1.5$ için Çizelge 5.3’te ve Şekil 5.3’te yer almaktadır.



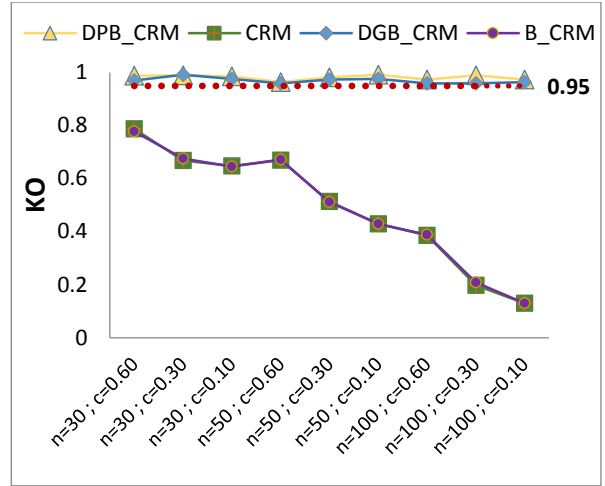
Çizelge 5.1. $\beta=0.5$ için HKO ve KO sonuçları

n	beta	durdurma	güvenirlik	DGB_CRM	KO_DGB	DPB_CRM	KO_DPB	B_CRM	KO_B	CRM	KO_CRM
30	0.5	0.60	0.5	0.25186	0.970	0.16695	0.988	<u>0.14471</u>	0.778	0.15089	0.788
30	0.5	0.60	0.7	0.24814	0.960	0.15581	0.982	<u>0.13046</u>	0.900	0.14405	0.908
30	0.5	0.60	0.9	0.24009	0.964	0.14659	0.974	<u>0.13233</u>	0.938	0.16037	0.938
30	0.5	0.30	0.5	0.19528	0.992	<u>0.09228</u>	0.990	0.10439	0.676	0.10661	0.668
30	0.5	0.30	0.7	0.19388	0.978	<u>0.07532</u>	0.990	0.07564	0.896	0.07858	0.896
30	0.5	0.30	0.9	0.17812	0.970	0.06752	0.988	<u>0.06675</u>	0.956	0.07388	0.95
30	0.5	0.10	0.5	0.21152	0.976	<u>0.08345</u>	0.986	0.09988	0.646	0.10154	0.648
30	0.5	0.10	0.7	0.20869	0.964	<u>0.06895</u>	0.990	0.07176	0.856	0.07430	0.854
30	0.5	0.10	0.9	0.18572	0.946	<u>0.06030</u>	0.996	0.06296	0.950	0.06899	0.948
50	0.5	0.60	0.5	0.21029	0.960	0.11604	0.964	<u>0.10856</u>	0.672	0.10959	0.670
50	0.5	0.60	0.7	0.19221	0.956	0.10406	0.966	<u>0.07794</u>	0.878	0.08005	0.876
50	0.5	0.60	0.9	0.14790	0.946	0.09268	0.960	<u>0.07277</u>	0.940	0.07878	0.942
50	0.5	0.30	0.5	0.19361	0.974	<u>0.07111</u>	0.982	0.09110	0.512	0.09140	0.514
50	0.5	0.30	0.7	0.16969	0.960	0.06021	0.980	<u>0.05465</u>	0.816	0.05505	0.82
50	0.5	0.30	0.9	0.11245	0.938	0.05001	0.982	<u>0.04181</u>	0.946	0.04374	0.946
50	0.5	0.10	0.5	0.16228	0.976	<u>0.05104</u>	0.992	0.08291	0.430	0.08303	0.43
50	0.5	0.10	0.7	0.14346	0.948	<u>0.04320</u>	0.988	0.04657	0.812	0.04650	0.804
50	0.5	0.10	0.9	0.09786	0.924	0.03954	0.978	<u>0.03362</u>	0.926	0.03468	0.926
100	0.5	0.60	0.5	0.17351	0.960	<u>0.07939</u>	0.974	0.09127	0.388	0.09113	0.386
100	0.5	0.60	0.7	0.11334	0.950	0.06271	0.972	0.05175	0.768	<u>0.05144</u>	0.776
100	0.5	0.60	0.9	0.05609	0.946	0.04431	0.960	<u>0.03376</u>	0.938	0.03440	0.942
100	0.5	0.30	0.5	0.14535	0.960	0.10543	0.990	0.08156	0.208	<u>0.08145</u>	0.198
100	0.5	0.30	0.7	0.08670	0.950	0.09014	0.986	0.04036	0.656	<u>0.04005</u>	0.664
100	0.5	0.30	0.9	0.03491	0.956	0.07775	0.974	<u>0.02005</u>	0.922	0.02021	0.932
100	0.5	0.10	0.5	0.13403	0.964	0.17616	0.974	0.08177	0.130	<u>0.08160</u>	0.130
100	0.5	0.10	0.7	0.07528	0.960	<u>0.02944</u>	0.986	0.03627	0.600	0.03601	0.610
100	0.5	0.10	0.9	0.03060	0.940	0.12817	0.970	<u>0.01780</u>	0.920	0.01784	0.920

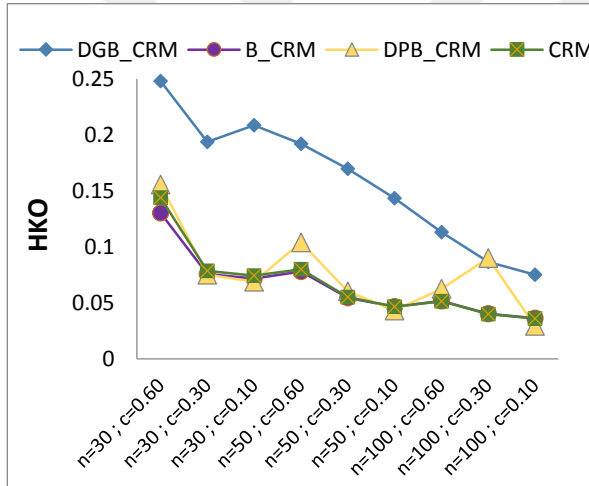
DGB_CRM: Ölçüm hatası için düzeltilmiş Gamma öselli Bayesci Cox regresyon modeli, DPB_CRM: Ölçüm hatası için düzeltilmiş Poligonal öselli Bayesci Cox regresyon modeli, B_CRM: Bayesci Cox regresyon modeli, CRM: Cox regresyon modeli, KO: Kapsama Olasılığı.



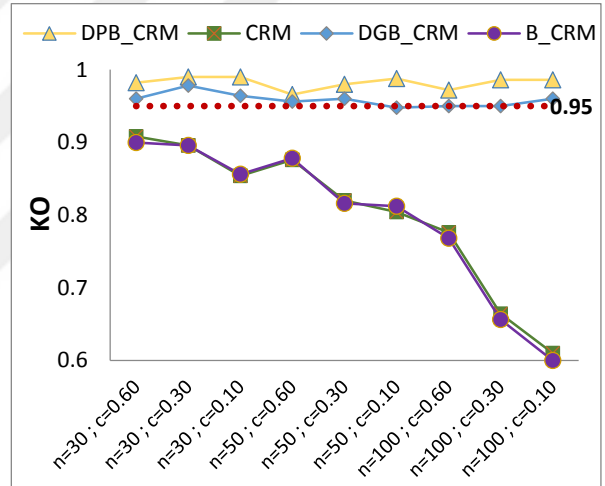
(a) $r = 0.5$ için HKO



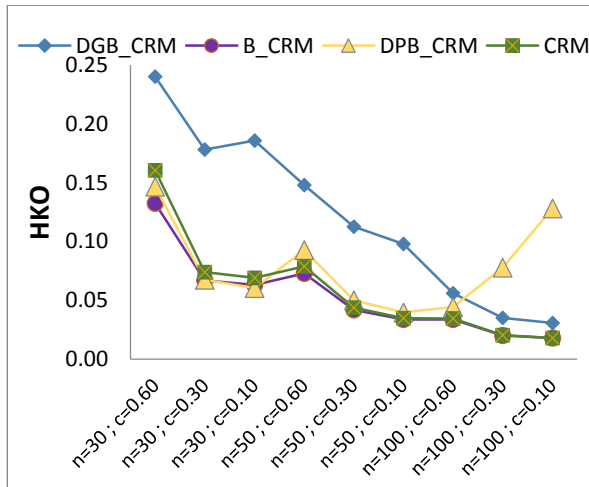
(b) $r = 0.5$ için KO



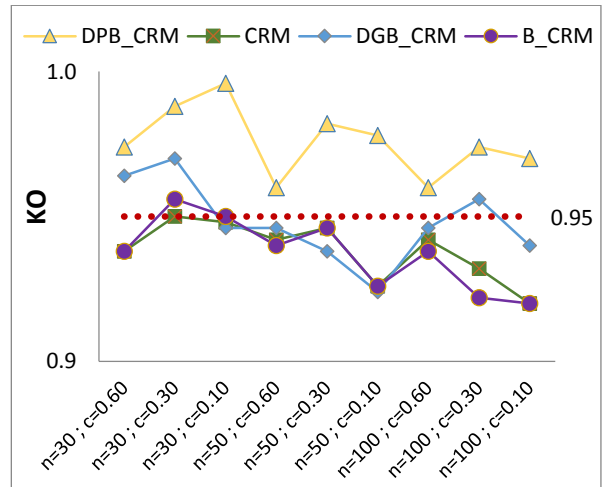
(c) $r = 0.7$ için HKO



(d) $r = 0.7$ için KO



(e) $r = 0.9$ için HKO



(f) $r = 0.9$ için KO

Şekil 5.1. $\beta=0.5$ için HKO ve KO sonuçlarının grafiği

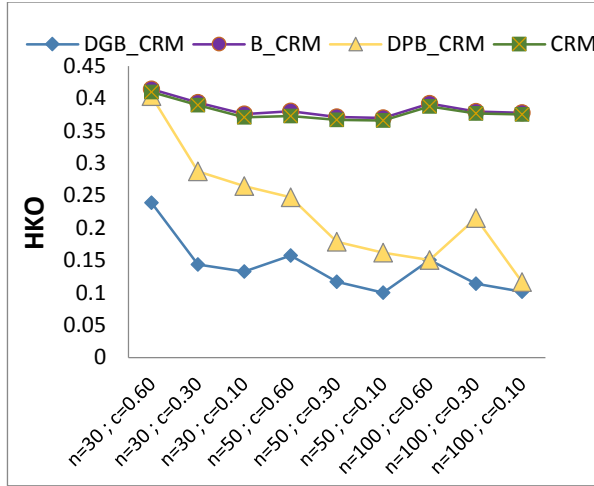
Bilgi içermeyen önsel durumunda B_CRM ile CRM'nin sonuçlarının yakın olduğu literatürde gösterilmiştir (Omurlu, Ozdamar ve Ture, 2009). Ölçüm hatalı açıklayıcı değişken bulunması durumunda da ele alınan tüm senaryolarda B_CRM ile CRM sonuçları yakın elde edilmiştir.

Regresyon katsayısı $\beta=0.5$ için Çizelge 5.1'de verilen sonuçlar yorum kolaylığı için Şekil 5.1'de sunulmuş ve yorumlar Şekil 5.1 üzerinden verilmiştir. DGB_CRM'ye ait HKO değerlerine bakıldığında tüm senaryolarda diğer modellere göre daha büyük HKO değeri elde edilmiştir. Şekil 5.1 (a)'da HKO değerlerine bakıldığında örneklem büyüklüğünün 100, durdurmanın 0.10 olduğu durum hariç DPB_CRM, B_CRM ve CRM yaklaşık olarak benzer sonuçlar vermiştir. Şekil 5.1 (b) incelendiğinde güvenilirliğin 0.5 olduğu durumlarda B_CRM ve CRM için KO değerleri 0.13 ile 0.78 arasındadır ve örneklem büyüklüğü arttıkça KO değerlerinin düştüğü sonucuna ulaşılmıştır. Ancak bu senaryolarda DPB_CRM'nin KO değerlerine bakıldığında diğer modellere göre daha iyi sonuç verdiği görülmektedir. Örneklem büyüklüğünün 100, durdurmanın 0.10 olduğu durumda ise DGB_CRM için uygun KO değeri ile küçük HKO değeri elde edilmiştir. Şekil 5.1 (d) incelendiğinde güvenilirliğin 0.7 olduğu durumlarda B_CRM ve CRM için KO değeri 0.60 ile 0.91 arasındadır. Güvenirliğin 0.5 olduğu durumlara göre yüksek olmasına rağmen diğer modellere göre yine düşüktür ve örneklem büyüklüğü arttıkça KO bu modellerde yine düşmektedir. Şekil 5.1 (c) incelendiğinde 0.7 güvenirlilikte, 0.10 durdurma düzeyinde tüm örneklem büyüklüklerinde DPB_CRM diğer modellere göre daha küçük HKO değerine sahiptir. Örneklem büyüklüğünün 100, durdurma oranının 0.30 olduğu senaryo için DGB_CRM ve DPB_CRM'nin HKO değerleri yakın elde edilmiş ancak KO değerine de bakıldığında DGB_CRM daha iyi sonuç vermiştir. Şekil 5.1 (f) incelendiğinde güvenilirliğin 0.9 olduğu durumlarda B_CRM ve CRM için KO değeri 0.92 ile 0.96 arasındadır. Örneklem büyüklüğünün 30, durdurma oranının 0.10 olduğu durum dışında tüm senaryolarda en küçük HKO, B_CRM'de elde edilmiştir. Örneklem büyüklüğünün 30, durdurma oranının 0.10 olduğu durumda ise en küçük HKO, DPB_CRM için elde edilmiş olmasına rağmen B_CRM ile değerleri çok yakındır ve KO değeri bakımından B_CRM daha uygundur. Diğer senaryolar için ise DPB_CRM için HKO sonuçları B_CRM ve CRM için elde edilen HKO sonuçlarına yakındır. KO değerleri bakımından incelendiğinde ise DPB_CRM'nin alternatif bir model olabileceği söylenebilir.

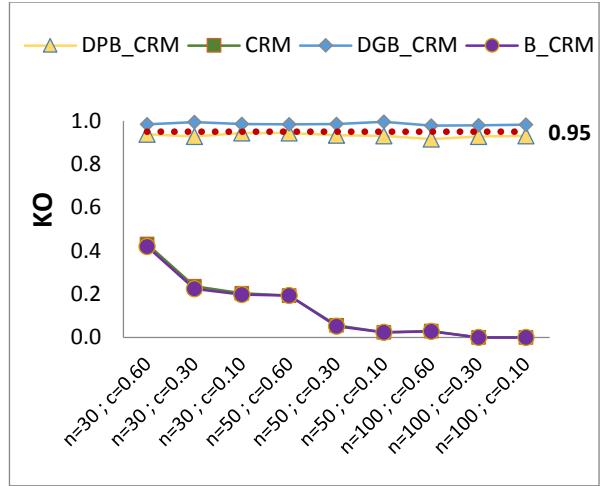
Çizelge 5.2. $\beta=1$ için HKO ve KO sonuçları

n	beta	durdurma	güvenirlik	DGB_CRM	KO_DGB	DPB_CRM	KO_DPB	B_CRM	KO_B	CRM	KO_CRM
30	1	0.60	0.5	<u>0.23882</u>	0.984	0.40375	0.938	0.41435	0.420	0.41023	0.432
30	1	0.60	0.7	<u>0.20561</u>	0.986	0.30457	0.950	0.26102	0.734	0.26867	0.764
30	1	0.60	0.9	0.20247	0.986	0.23426	0.954	<u>0.16687</u>	0.942	0.23016	0.944
30	1	0.30	0.5	<u>0.14382</u>	0.994	0.28770	0.928	0.39388	0.224	0.38956	0.236
30	1	0.30	0.7	<u>0.13670</u>	0.994	0.17827	0.956	0.21270	0.616	0.20764	0.638
30	1	0.30	0.9	0.15763	0.994	0.10842	0.980	<u>0.09892</u>	0.926	0.11552	0.930
30	1	0.10	0.5	<u>0.13284</u>	0.986	0.26458	0.944	0.37577	0.198	0.37102	0.204
30	1	0.10	0.7	<u>0.13341</u>	0.996	0.15917	0.966	0.20163	0.590	0.19718	0.610
30	1	0.10	0.9	0.15789	0.978	<u>0.09792</u>	0.988	0.09988	0.900	0.11561	0.918
50	1	0.60	0.5	<u>0.15744</u>	0.984	0.24733	0.944	0.38022	0.194	0.37310	0.194
50	1	0.60	0.7	<u>0.14078</u>	0.978	0.16746	0.956	0.20374	0.608	0.19417	0.64
50	1	0.60	0.9	0.13055	0.968	0.12177	0.952	<u>0.09877</u>	0.922	0.10737	0.926
50	1	0.30	0.5	<u>0.11709</u>	0.986	0.17896	0.934	0.37169	0.052	0.36713	0.056
50	1	0.30	0.7	0.12198	0.982	<u>0.10703</u>	0.958	0.18040	0.430	0.17274	0.46
50	1	0.30	0.9	0.11932	0.968	0.07050	0.970	0.06584	0.902	<u>0.06538</u>	0.912
50	1	0.10	0.5	<u>0.10038</u>	0.996	0.16203	0.930	0.36995	0.024	0.36587	0.024
50	1	0.10	0.7	0.11125	0.990	<u>0.09099</u>	0.962	0.17515	0.380	0.16870	0.392
50	1	0.10	0.9	0.11314	0.966	<u>0.05416</u>	0.980	0.05632	0.890	0.05596	0.904
100	1	0.60	0.5	0.15078	0.978	<u>0.15043</u>	0.916	0.39246	0.028	0.38801	0.030
100	1	0.60	0.7	0.13310	0.968	<u>0.10971</u>	0.902	0.19872	0.298	0.19197	0.322
100	1	0.60	0.9	0.08653	0.946	0.08036	0.910	0.07274	0.828	<u>0.07032</u>	0.860
100	1	0.30	0.5	<u>0.11388</u>	0.980	0.21527	0.928	0.37981	0.000	0.37695	0.000
100	1	0.30	0.7	<u>0.09226</u>	0.972	0.13662	0.938	0.17861	0.128	0.17380	0.148
100	1	0.30	0.9	0.05419	0.958	0.11255	0.944	0.04706	0.812	<u>0.04432</u>	0.826
100	1	0.10	0.5	<u>0.10162</u>	0.982	0.11653	0.930	0.37770	0.000	0.37534	0.000
100	1	0.10	0.7	<u>0.08195</u>	0.978	0.19606	0.944	0.17792	0.070	0.17418	0.074
100	1	0.10	0.9	0.05213	0.944	0.04765	0.972	0.03992	0.788	<u>0.03746</u>	0.802

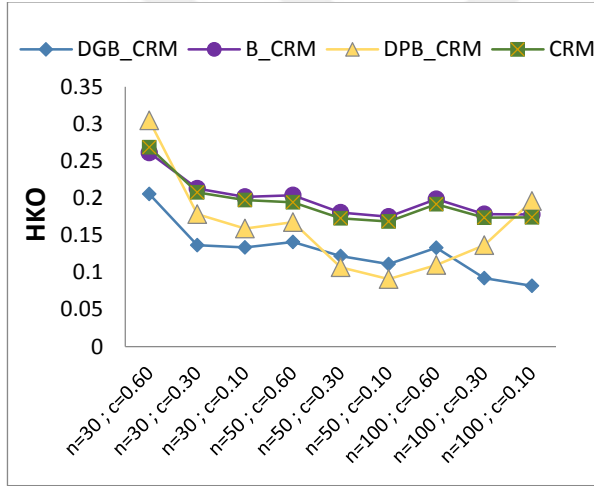
DGB_CRM: Ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modeli, DPB_CRM: Ölçüm hatası için düzeltilmiş Poligonal önselli Bayesci Cox regresyon modeli, B_CRM: Bayesci Cox regresyon modeli, CRM: Cox regresyon modeli, KO: Kapsama Olasılığı.



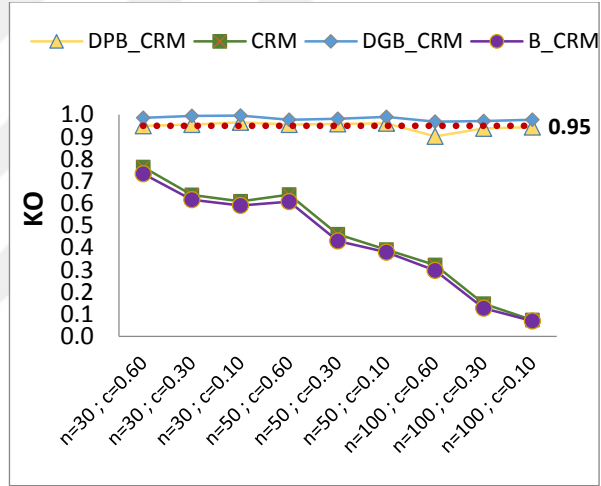
(a) $r = 0.5$ için HKO



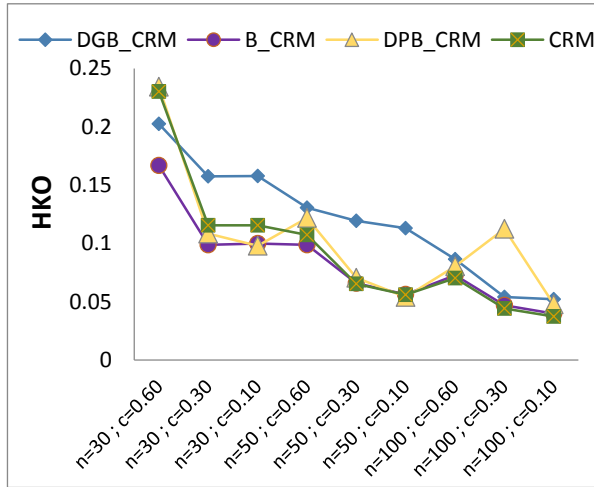
(b) $r = 0.5$ için KO



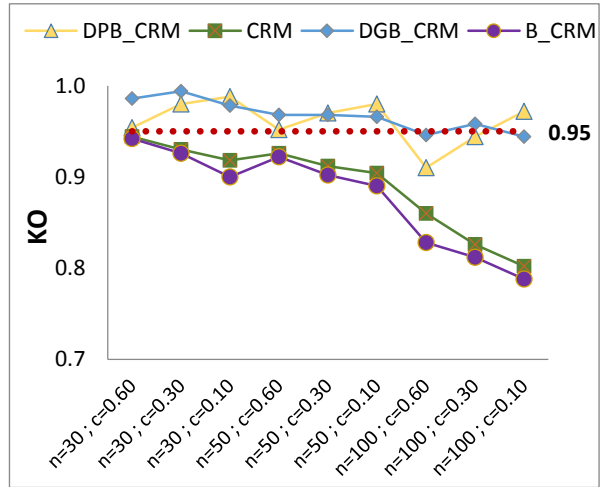
(c) $r = 0.7$ için HKO



(d) $r = 0.7$ için KO



(e) $r = 0.9$ için HKO



(f) $r = 0.9$ için KO

Şekil 5.2. $\beta = 1.0$ için senaryoların HKO ve KO sonuçlarının grafiği

Çizelge 5.2'nin özetlendiği Şekil 5.2 incelendiğinde $\beta=1$ için güvenilirliğin 0.5 olduğu durumda, B_CRM ve CRM için KO sıfır ile 0.43 arasında, güvenilirliğin 0.7 olduğu durumda 0.07 ile 0.76 arasında değişmekte ve örneklem büyüklüğü arttıkça KO düşmektedir. Şekil 5.2 (a) incelendiğinde 0.5 güvenirlilikte örneklem büyüklüğünün 100, durdurma oranının 0.60 ve 0.10 olduğu durumlar dışında tüm senaryolarda en küçük HKO, DGB_CRM için elde edilmiştir. Örneklem büyüklüğünün 100, durdurmanın 0.60 ve 0.30 olduğu durumlar için ise DPB_CRM ve DGB_CRM'nin HKO değerleri yakın elde edilmiştir. KO değerlerine bakıldığında ise iki model için de 0.95'e yakındır. Şekil 5.2 (c) incelendiğinde 0.7 güvenirlilik düzeyinde 50 örneklem büyüklüğü 0.30 ve 0.10 durdurma oranlarında DPB_CRM en küçük HKO değerini veren model olmuş ve KO değeri 0.95 civarındadır. Örneklem büyüklüğünün 100, durdurma oranının 0.60 olduğu durumda ise DPB_CRM en küçük HKO değerine sahip model olmasına rağmen KO değeri 0.90 elde edilmiştir. DGB_CRM için ise bu senaryoda yakın HKO değeri elde edilmiş ve KO değeri de 0.95'e yakındır. Diğer senaryolarda ise DGM_CRM en küçük HKO değerine sahip modeldir. Şekil 5.2 (e) incelendiğinde 0.9 güvenirlilik düzeyinde ise 30 örneklem büyüklüğü, 0.60 durdurma oranı için B_CRM'de HKO ve KO değerlerine bakıldığında en iyi model olarak elde edilmiştir. Ancak örneklem büyüklüğü arttıkça bu modeller bazı senaryolarda daha küçük HKO değeri vermesine rağmen KO değerleri hem B_CRM için hem de CRM için 0.80'e kadar düşmüştür. Örneklem büyüklüğünün 100 olduğu tüm durdurma oranlarında CRM daha küçük HKO değerine sahip olmasına rağmen KO değerleri 0.80 ve 0.86 arasında elde edilmiştir. Durdurma oranının 0.10 olduğu durumda diğer modellerin de HKO değerleri CRM'nin değerine yakın elde edilmiş ancak düzeltilmiş modellerin KO değeri 0.95'e yakın elde edildiği için düzeltilmiş iki modelin daha iyi olduğu söylenebilir. Durdurma oranının 0.30 ve 0.60 olduğu durumlarda ise hem HKO değerleri hem de KO değerleri değerlendirildiğinde DGB_CRM'nin daha iyi sonuç verdiği görülmüştür. Örneklem büyüklüğünün 50 olduğu durumda ise 0.30 ve 0.60 durdurma oranlarında CRM ve B_CRM küçük HKO vermesine rağmen, KO normal seviyesinden düşük sonuç vermiştir. DPB_CRM için bu senaryolarda CRM ve B_CRM ile yakın HKO değeri elde edilirken, KO da 0.95 civarı elde edilmiştir. Durdurma oranının 0.10 olduğu durumda ise uygun KO ile birlikte en küçük HKO değerinin elde edildiği model DPB_CRM'dir. Örneklem büyüklüğünün 30,

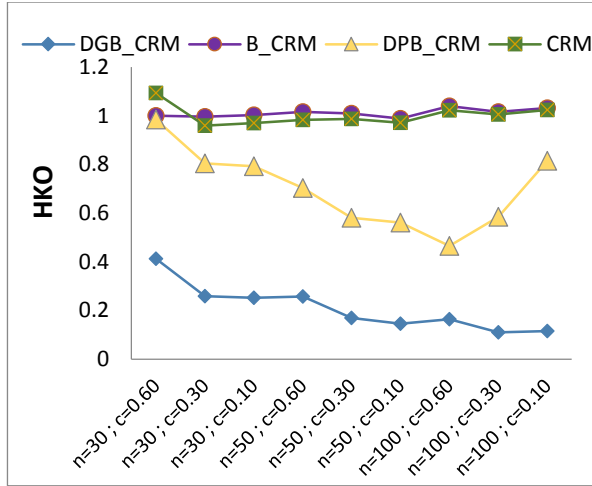
durdurma oranının 0.30 ve 0.60 olduđu senaryolarda ise B_CRM iyi sonuç verirken, 0.10 durdurma oranında DPB_CRM en iyi modeldir.



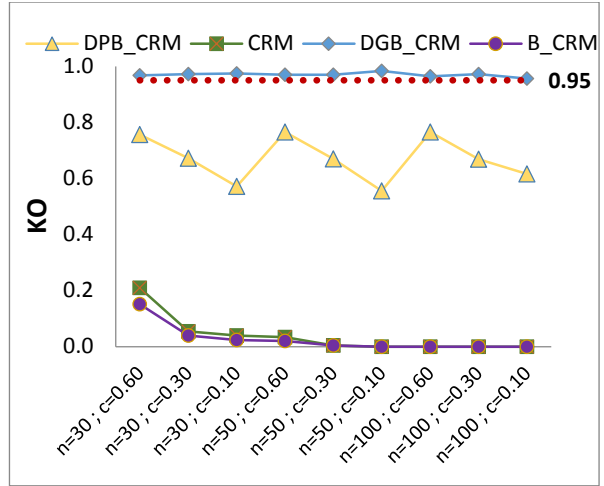
Çizelge 5.3. $\beta=1.5$ için HKO ve KO sonuçları

n	beta	durdurma	güvenirlik	DGB_CRM	KO_DGB	DPB_CRM	KO_DPB	B_CRM	KO_B	CRM	KO_CRM
30	1.5	0.60	0.5	<u>0.41268</u>	0.968	0.98420	0.756	1.00062	0.152	1.09426	0.210
30	1.5	0.60	0.7	<u>0.25756</u>	0.980	0.73055	0.824	0.26374	0.446	0.55016	0.590
30	1.5	0.60	0.9	<u>0.17592</u>	0.988	0.51830	0.884	0.29123	0.852	0.37833	0.890
30	1.5	0.30	0.5	<u>0.25948</u>	0.972	0.80467	0.672	0.99641	0.040	0.95983	0.054
30	1.5	0.30	0.7	<u>0.14210</u>	0.998	0.53196	0.800	0.55439	0.298	0.49537	0.382
30	1.5	0.30	0.9	<u>0.09812</u>	0.992	0.31627	0.920	0.20681	0.840	0.19038	0.898
30	1.5	0.10	0.5	<u>0.25293</u>	0.974	0.79215	0.572	1.00356	0.024	0.96974	0.040
30	1.5	0.10	0.7	<u>0.13339</u>	0.986	0.50993	0.758	0.55324	0.236	0.50112	0.320
30	1.5	0.10	0.9	<u>0.09199</u>	0.996	0.29215	0.898	0.20309	0.756	0.19211	0.832
50	1.5	0.60	0.5	<u>0.25812</u>	0.970	0.70363	0.766	1.01566	0.020	0.98261	0.034
50	1.5	0.60	0.7	<u>0.16129</u>	0.984	0.48350	0.808	0.56902	0.238	0.51320	0.294
50	1.5	0.60	0.9	<u>0.11476</u>	0.986	0.31738	0.838	0.21641	0.794	0.19081	0.858
50	1.5	0.30	0.5	<u>0.16921</u>	0.970	0.58102	0.670	1.00969	0.004	0.98667	0.004
50	1.5	0.30	0.7	<u>0.10089</u>	0.992	0.35408	0.784	0.53608	0.094	0.49505	0.158
50	1.5	0.30	0.9	<u>0.08454</u>	0.992	0.18803	0.876	0.16414	0.724	0.13724	0.808
50	1.5	0.10	0.5	<u>0.14608</u>	0.984	0.56098	0.556	0.98885	0.000	0.97182	0.000
50	1.5	0.10	0.7	<u>0.08522</u>	0.988	0.32578	0.746	0.51263	0.056	0.47932	0.092
50	1.5	0.10	0.9	<u>0.06998</u>	0.998	0.15621	0.882	0.13834	0.684	0.11314	0.772
100	1.5	0.60	0.5	<u>0.16500</u>	0.964	0.46575	0.766	1.04010	0.000	1.02290	0.000
100	1.5	0.60	0.7	<u>0.12501</u>	0.954	0.31382	0.754	0.56659	0.040	0.53415	0.064
100	1.5	0.60	0.9	<u>0.08740</u>	0.968	0.19414	0.758	0.17423	0.628	0.14393	0.712
100	1.5	0.30	0.5	<u>0.11055</u>	0.972	0.58544	0.668	1.01643	0.000	1.00579	0.000
100	1.5	0.30	0.7	<u>0.07722</u>	0.974	0.35858	0.764	0.52822	0.000	0.50707	0.000
100	1.5	0.30	0.9	<u>0.05295</u>	0.986	0.20407	0.844	0.12964	0.498	0.10675	0.618
100	1.5	0.10	0.5	<u>0.11526</u>	0.956	0.81564	0.616	1.03230	0.000	1.02393	0.000
100	1.5	0.10	0.7	<u>0.07216</u>	0.980	0.51111	0.698	0.53771	0.000	0.51992	0.000
100	1.5	0.10	0.9	<u>0.03998</u>	0.994	0.31449	0.862	0.12958	0.392	0.10926	0.480

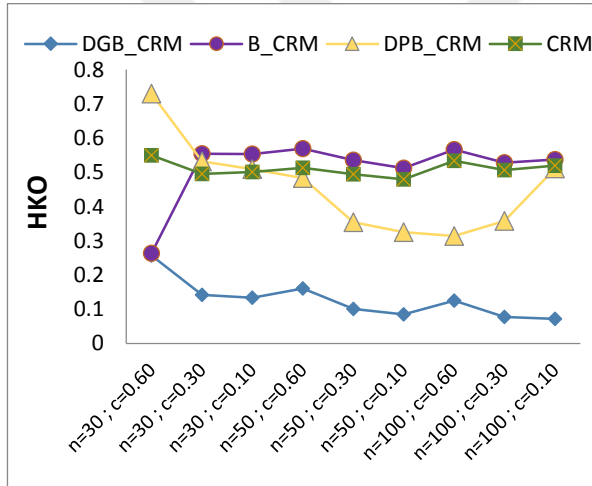
DGB_CRM: Ölçüm hatası için düzeltilmiş Gamma önselli Bayesci Cox regresyon modeli, DPB_CRM: Ölçüm hatası için düzeltilmiş Poligonal önselli Bayesci Cox regresyon modeli, B_CRM: Bayesci Cox regresyon modeli, CRM: Cox regresyon modeli, KO: Kapsama Olasılığı.



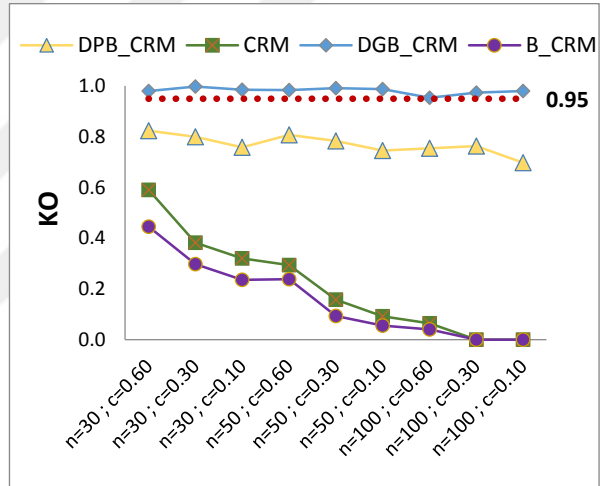
(a) $r = 0.5$ için HKO



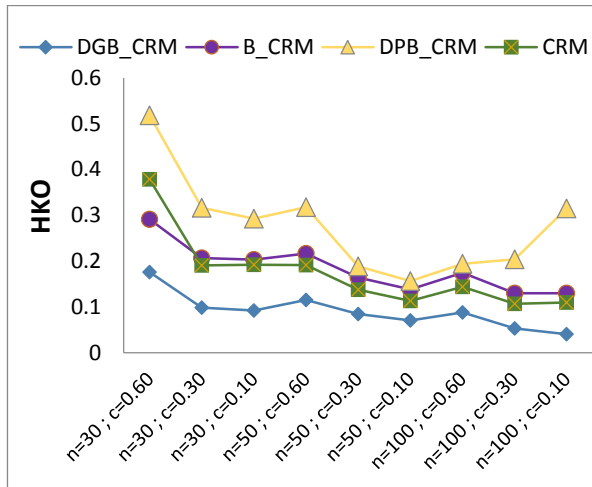
(b) $r = 0.5$ için KO



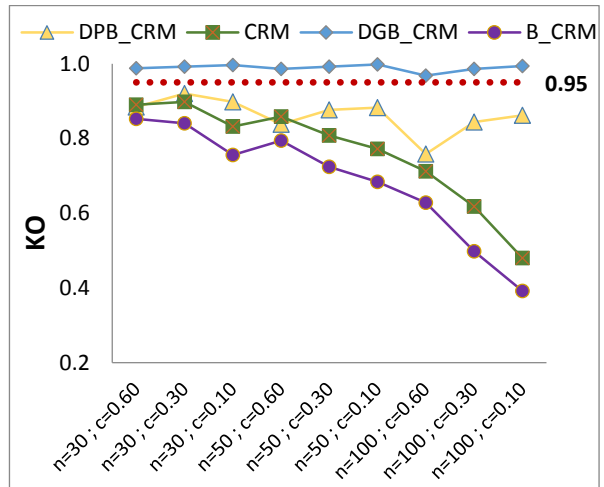
(c) $r = 0.7$ için HKO



(d) $r = 0.7$ için KO



(e) $r = 0.9$ için HKO



(f) $r = 0.9$ için KO

Şekil 5.3. $\beta = 1.5$ için senaryoların HKO ve KO sonuçlarının grafiği

Çizelge 5.3 ve yorumlama kolaylığı için sunulan Şekil 5.3 incelendiğinde $\beta=1.5$ için belirlenen tüm örneklem büyüklükleri, durdurma oranları ve güvenilirlik değerleri için DGM_CRM için en küçük HKO değeri elde edilmiştir. Güvenirliğin 0.5 olduğu, 50 örneklem büyüklüğü, 0.10 durdurma oranı dışında tüm olası senaryolarda bu modelin KO değeri 0.95 civarı elde edilmiştir. Ancak 50 örneklem büyüklüğü, 0.10 durdurma oranında KO değeri normal düzeyinden yüksek elde edilmiştir. Şekil 5.3 (d) incelendiğinde 0.7 güvenirlilik düzeyinde, 100 örneklem büyüklüğünde, 0.30 ve 0.60 durdurma oranında KO yaklaşık 0.95 elde edilirken, diğer tüm durumlarda 0.98-0.99 arasında elde edilmiştir. Şekil 5.3 (f) incelendiğinde 0.9 güvenirlilik düzeyinde ise 100 örneklem büyüklüğünde, 0.60 durdurma oranında KO 0.95 civarı elde edilirken, diğer tüm durumlarda KO yaklaşık 0.99 elde edilmiştir.

5.2. Uygulama

Önerilen modelin uygulanabilirliğini göstermek amacıyla literatürde var olan iki veri seti ele alınmıştır. Bu veri setleri için tez kapsamında ele alınan modeller çalıştırılmış ve karşılaştırılmıştır.

5.2.1. Uygulama 1

Tez çalışması kapsamında analizlerin uygulanabilirliğini göstermek amacı ile uygulamanın birinci kısmında 1974 ve 1984 yılları arasında yapılan karaciğerin primer biliyer sirozunda (Primary Biliary Cirrhosis, PBC) Mayo Klinik çalışmasından elde edilen verileri kullanılmıştır. Bu on yıllık aralıkta D-penisilin ilacının randomize plasebo kontrollü çalışması için kriterleri sağlayan Mayo Klinik' e gönderilen toplam 424 PBC hastası çalışmaya katılmış ancak takibin kaybedilmesi gibi nedenlerden dolayı tam veriyi içeren 312 vaka randomize araştırmaya dahil edilmiştir. Bu klinik çalışmanın amacı D-penisilin ilacının ölüm ya da karaciğer nakli gerçekleşene kadar geçen süre üzerindeki etkisini incelemektir.

PBC, karaciğerin safra kanallarının tahrip olduğu ölümcül bir durumdur. Safra kanalları, yağların düzgün sindirimi, hasarlı veya eski kırmızı kan hücrelerinin çıkarılması ve detoksifikasyon için gereklidir. PBC, vücutta zararlı toksinlerin birikmesine ve ayrıca

karaciğere karşı geri dönüşümsüz hasar oluşumuna yol açar. Mevcut araştırmalar PBC'yi vücudun kendisine saldırdığı bir otoimmün hastalık olarak görmektedir (Ray, 2016).

Çalışmada D-penisilin ilacının hastalık üzerinde etkisini görmek amacı ile 154 hasta plasebo grubu olarak alınmıştır, kalan 158 hastaya ilaç verilmiştir. Ayrıca çalışma boyunca hastaların serum bilirubin seviyeleri ilk 6 ay sonunda ve sonrasında yılda bir olmak üzere kaydedilmiştir.

Çalışmamızda ölçüm hatası içeren açıklayıcı değişken olarak log-serum bilirubin seviyesi ele alınmıştır. Ölçüm hatalı açıklayıcı değişkenin ikinci tekrarlı ölçümü olarak da altı ay sonra kaydedilen ölçümü alınmıştır. Çalışmamızın amacı, PBC hastalarının yaşam süresi üzerinde ilaç (ölçüm hatası içermeyen değişken) ve log-serum bilirubin (ölçüm hatası içeren sürekli ölçekle ölçülmüş değişken) değişkenlerinin etkisinin incelenmesi olarak belirlenmiştir.

Sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler Çizelge 5.4'te, nitel olan değişkenler ise Çizelge 5.5.'te verilmiştir.

Çizelge 5.4. PBC verisinde sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler

Değişken	n	%	Minimum	Maksimum	Ortalama	Ortanca	Standart sapma
Süre	312	100	0	14	6.41	6.00	3.556
log-serum bilirubin_0	312	100	-1	3	.55	0.00	1.116
log-serum bilirubin_6	257	82	-2	3	.42	0.00	1.174

log-serum bilirubin_0: Başlangıçtaki log-serum bilirubin seviyesi, log-serum bilirubin_6: Altı ay sonra elde edilen log-serum bilirubin seviyesi

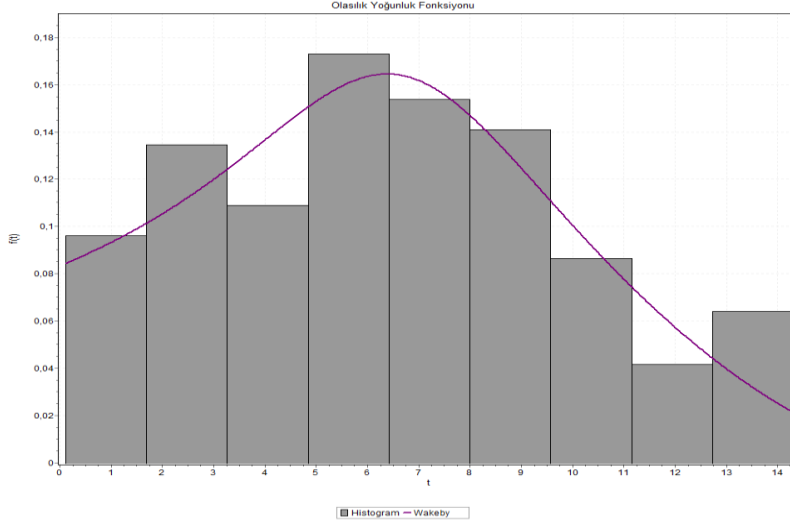
Log-serum bilirubin değişkeninin ikinci ölçümü çalışmaya dahil olan bireylerden %82'si için elde edilmiştir, % 18'i kayıp gözlemdir.

Çizelge 5.5. PBC verisinde kullanılan nitel değişkenler

Değişken	Düzy	Frekans	Yüzde
Durum	0 (durdurma)	172	55.1
	1 (ölüm)	140	44.9
İlaç	D-penicil	158	50.6
	Plasebo	154	49.4

PBC veri setine ilişkin durdurma oranı %55.1'dir.

PBC verisi için süreye ilişkin dağılım grafiği Şekil 5.4'te verilmiştir. Sürenin dağılımı incelendiğinde bilinen parametrik dağılımlardan birini göstermediği, Wakeby dağılımı gösterdiği gözlenmiştir.



Şekil 5.4. PBC verisi için süreye ilişkin dağılım grafiği

Cox regresyon modelinin uygulanabilmesi için orantılı tehlikeler varsayımının sağlanması gerekir. Orantılı tehlikeler varsayımının incelenmesi için Schoenfeld artıklarına dayanan orantılı tehlikeler varsayımı testi sonuçları Çizelge 5.6'da verilmiştir.

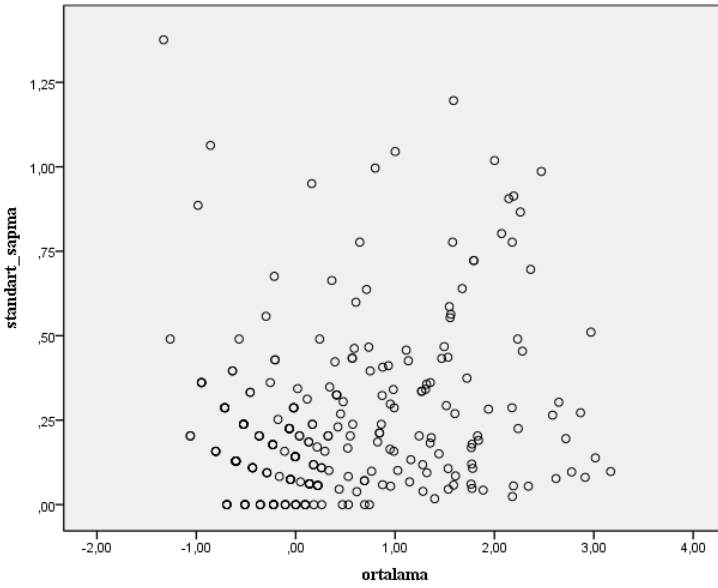
Çizelge 5.6. PBC verisinde orantılı tehlikeler varsayımı test sonuçları

Değişken	ρ	Ki-kare	p
log-serum bilirubin	-0.0661	0.5728	0.449
İlaç	-0.0121	0.0202	0.887
Global	-	0.6290	0.730

Çizelge 5.6 incelendiğinde Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki ilişkinin istatistiksel olarak önemli olmadığı, orantılı tehlikeler varsayımının sağlandığı sonucuna ulaşılmıştır.

Çözümlemede daha hızlı yakınsama elde edilmesi amacıyla log-serum bilirubin değişkeninin standartlaştırılmış değerleri kullanılmıştır.

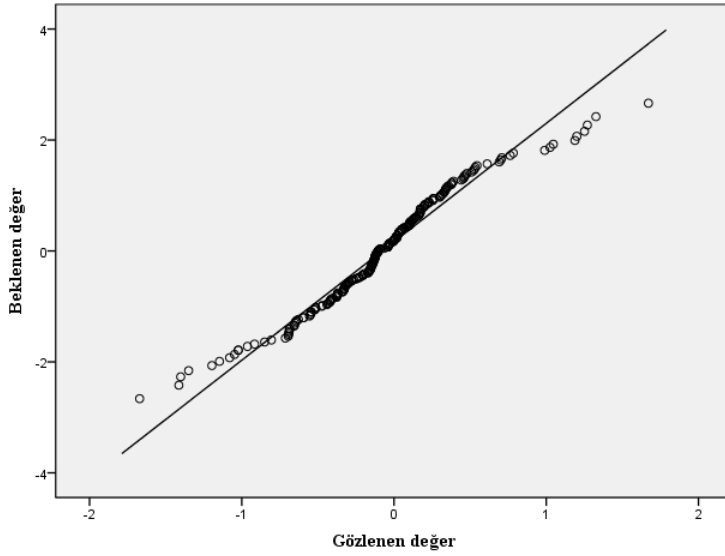
Klasik ölçüm hatası modeline uyumu araştırmak için ölçüm hatalı değişkenin tekrarlı ölçümlerinden elde edilen ortalamalarına karşı standart sapmalarının çizimi Şekil 5.5'te verilmiştir.



Şekil 5.5. Log-serum bilirubin tekrarlarının ortalamalarına karşı standart sapma grafiği

Şekil 5.5 incelendiğinde ortalama ve standart sapma arasında belirgin bir eğilim olmadığı görülmektedir. Ölçüm hatası varyansının log-serum bilirubinin gerçek değerlerinden bağımsız olduğu söylenebilir.

Yine klasik ölçüm hatası modeline uyumu araştırmak için çizdirilen ölçüm hatalı değişkenin tekrarlı ölçümleri arasındaki farkın normal QQ grafiği Şekil 5.6'da verilmiştir.



Şekil 5.6. Log-serum bilirubin değişkeninin tekrarlı gözlemleri arasındaki farkın normal QQ grafiği

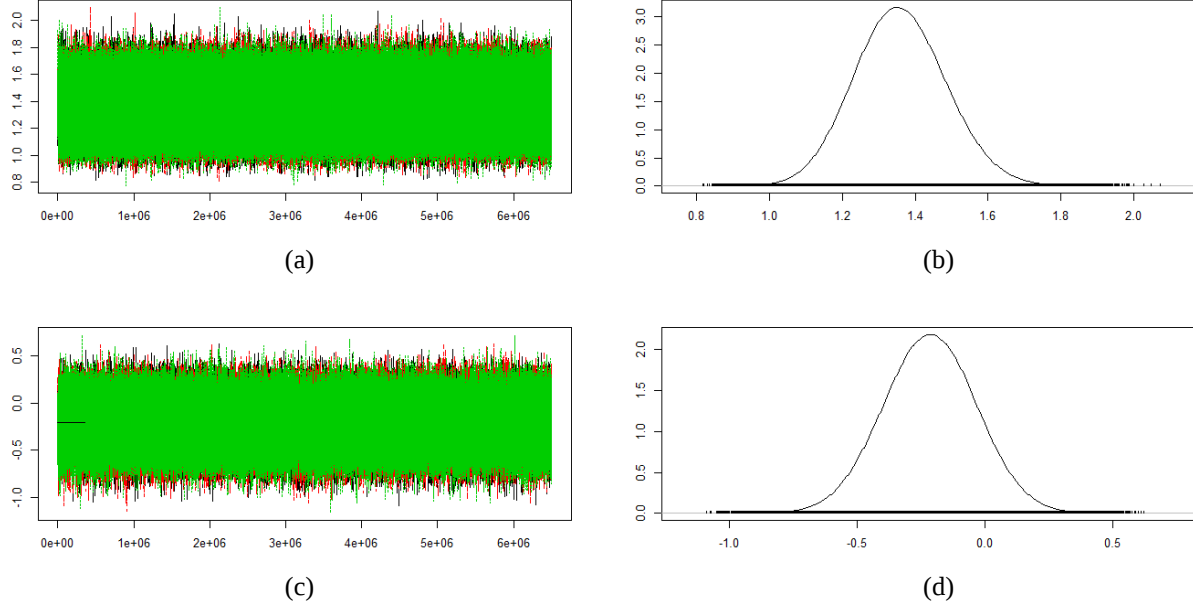
Şekil 5.6 incelendiğinde ise her bir bireye ait log-serum bilirubin değişkeninin tekrarlı gözlemleri arasındaki fark değişkeninin normal QQ grafiğinden normal dağılım gösterdiği söylenebilir. Verinin klasik ölçüm hatası modeline uygun olduğu söylenebilir.

Yakınsamanın elde edilebilmesi için bayesci modellerin her biri için üç zincir çalıştırılmıştır.

Bayesci çözümler için yakınsamanın sağlanıp sağlanmadığının kontrolünde iz ve yoğunluk grafiği çizdirilmiş, Gelman-Rubin yakınsama testi ve Heidelberger-Welch testi yakınsama ölçütü olarak kullanılmıştır. Bütün modeller için yakınsamanın sağlandığı iterasyon sayısı ile çalışılmıştır.

DGB_CRM için elde edilen iz ve yoğunluk grafikleri log-serum bilirubin değişkeni için elde edilen regresyon parametresi β_1 için Şekil 5.7(a, b)'de, ilaç değişkeni için elde edilen regresyon parametresi β_2 için Şekil 5.7(c, d)'de verilmiştir. Gelman-Rubin yakınsama testi sonucu her bir parametre için 1.0 elde edilmiştir. MCMC'nin durağan bir dağılımdan geldiği yokluk hipotezini ve ortalamanın belirlenen bir doğrulukta tahmin edilebilmesi için zincirin kabul edilebilir

uzunlukta olup olmadığını test eden, Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucu Çizelge 5.7'de verilmiştir.



Şekil 5.7. DGB_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri

Şekil 5.7'de iz grafikleri incelendiğinde DGB_CRM için herhangi bir trende rastlanmadığı, parametreye ait zincirlerin yakınsadığı sonucuna ulaşılmaktadır. Yoğunluk grafiğinden de log-serum bilirubin değişkeninin yaşam süresi üzerinde pozitif ve anlamlı bir etkisi olduğu ancak ilaç değişkeninin anlamsız etkiye sahip olduğu görülmektedir.

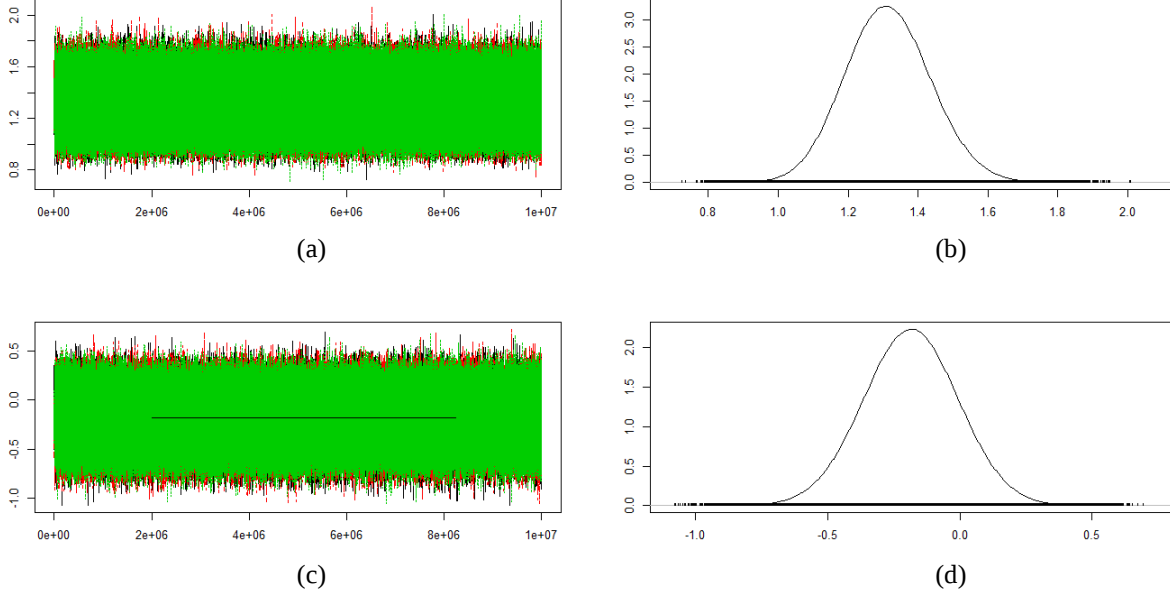
Çizelge 5.7. DGB_CRM için Heidelberger-Welch yakınsaklık testi sonucu

Zincir	Parametre	Heidelberger-Welch Testi				
		Durağanlık Testi		Yarı Genişlik Testi		
		Başlangıç İterasyonu	p- Değeri	Ortalama	Yarı Genişlik Testi	Sonuç
[1]	β_1	1	0.079	1.360	0.000500	Geçti
	β_2	1	0.476	-0.214	0.001630	Geçti
	γ_1	1	0.751	-0.158	0.000180	Geçti
	γ_2	1	0.280	0.117	0.000113	Geçti
	taux	1	0.658	1.07	0.000093	Geçti
	tauu	1	0.384	8.14	0.001460	Geçti
[2]	β_1	1	0.226	1.359	0.000487	Geçti
	β_2	1	0.980	-0.215	0.001523	Geçti
	γ_1	1	0.313	-0.158	0.000178	Geçti
	γ_2	1	0.338	0.117	0.000113	Geçti
	taux	1	0.377	1.07	0.000094	Geçti
	tauu	1	0.552	8.13	0.001460	Geçti
[3]	β_1	1	0.421	1.359	0.000499	Geçti
	β_2	1	0.665	-0.215	0.001704	Geçti
	γ_1	1	0.405	-0.158	0.000181	Geçti
	γ_2	1	0.470	0.117	0.000114	Geçti
	taux	1	0.846	1.07	0.000038	Geçti
	tauu	1	0.282	8.13	0.001460	Geçti

Çizelge 5.7 incelendiğinde Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucunda DGB_CRM 'de tüm parametreler için çalıştırılan üç zincirin yakınsadığı sonucuna ulaşılmıştır.

DPB_CRM için elde edilen iz ve yoğunluk grafikleri log-serum bilirubin değişkeni için elde edilen regresyon parametresi β_1 için Şekil 5.8(a, b)'de, ilaç değişkeni için elde edilen regresyon parametresi β_2 için Şekil 5.8(c, d)'de verilmiştir. Gelman-Rubin yakınsama testi sonucu her bir parametre için 1.0 elde edilmiştir. MCMC'nin durağan bir dağılımdan geldiği yokluk hipotezini ve ortalamanın belirlenen bir doğrulukta tahmin edilebilmesi için zincirin kabul edilebilir

uzunlukta olup olmadığını test eden, Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucu Çizelge 5.8'de sunulmuştur.



Şekil 5.8. DPB_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri

Şekil 5.8'de iz grafikleri incelendiğinde DPB_CRM için herhangi bir trende rastlanmadığı, parametreye ait zincirlerin yakınsadığı sonucuna ulaşılmaktadır. Yoğunluk grafiğinden de log-serum bilirubin değişkeninin yaşam süresi üzerinde pozitif ve anlamlı bir etkisi olduğu ancak ilaç değişkeninin anlamsız etkiye sahip olduğu görülmektedir.

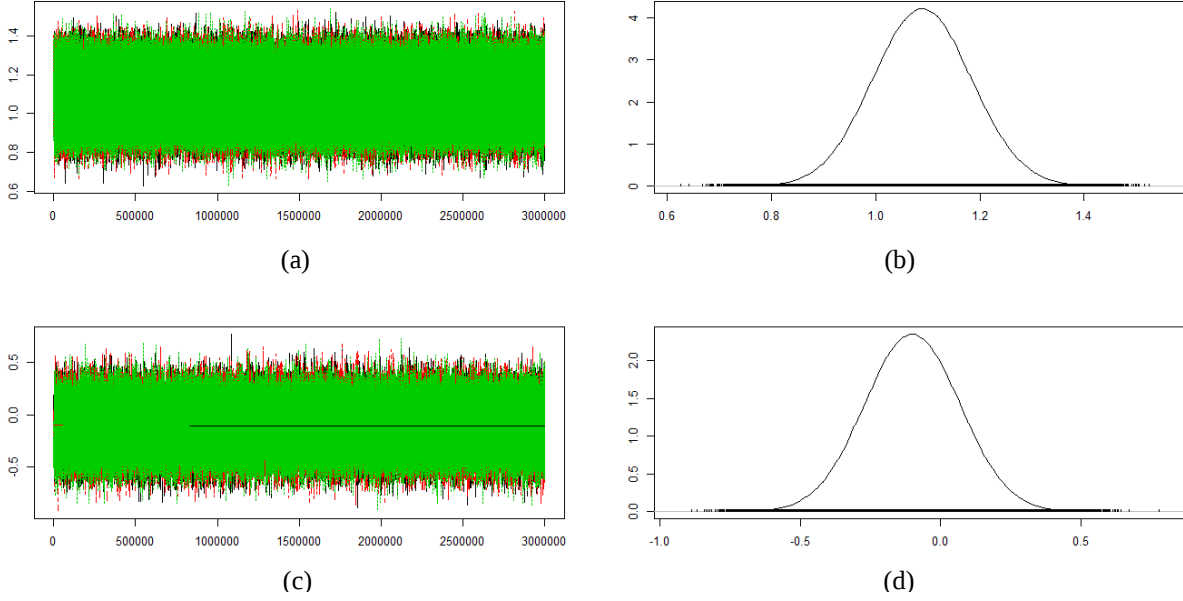
Çizelge 5.8. DPB_CRM için Heidelberger-Welch yakınsaklık testi sonucu

Zincir	Parametre	Heidelberger-Welch Testi				
		Durağanlık Testi		Yarı Genişlik Testi		
		Başlangıç İterasyonu	p- Değeri	Ortalama	Yarı Genişlik Testi	Sonuç
[1]	β_1	1	0.837	1.315	0.000374	Geçti
	β_2	1	0.563	-0.185	0.001010	Geçti
	γ_1	1	0.426	-0.157	0.000138	Geçti
	γ_2	1	0.384	0.117	0.000088	Geçti
	taux	1	0.901	1.100	0.000078	Geçti
	tauu	1	0.551	8.41	0.0012	Geçti
	tau.e	1	0.568	89.9	0.246	Geçti
[2]	β_1	1	0.806	1.315	0.000378	Geçti
	β_2	1	0.945	-0.186	0.001003	Geçti
	γ_1	1	0.460	-0.157	0.000138	Geçti
	γ_2	1	0.679	0.117	0.000088	Geçti
	taux	1	0.682	1.100	0.000078	Geçti
	tauu	1	0.446	8.41	0.0012	Geçti
	tau.e	1	0.194	89.9	0.255	Geçti
[3]	β_1	1	0.570	1.315	0.000381	Geçti
	β_2	1	0.328	-0.186	0.001034	Geçti
	γ_1	1	0.995	-0.157	0.000139	Geçti
	γ_2	1	0.967	0.117	0.000088	Geçti
	taux	1	0.391	1.100	0.000078	Geçti
	tauu	1	0.881	8.41	0.00119	Geçti
	tau.e	1	0.227	90	0.252	Geçti

Çizelge 5.8 incelendiğinde Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucunda DPB_CRM 'de tüm parametreler için çalıştırılan üç zincirin yakınsadığı sonucuna ulaşılmıştır.

B_CRM için elde edilen iz ve yoğunluk grafikleri log-serum bilirubin değişkeni için elde edilen regresyon parametresi β_1 için Şekil 5.9(a, b)'de, ilaç değişkeni için elde edilen regresyon parametresi β_2 için Şekil 5.9(c, d)'de verilmiştir. Gelman-Rubin yakınsama testi sonucu her bir parametre için 1.0 elde edilmiştir. MCMC'nin durağan bir dağılımdan geldiği yokluk hipotezini

ve ortalamanın belirlenen bir doğrulukta tahmin edilebilmesi için zincirin kabul edilebilir uzunlukta olup olmadığını test eden, Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucu Çizelge 5.9'da sunulmuştur.



Şekil 5.9. B_CRM için β_1 ve β_2 parametrelerinin iz ve yoğunluk grafikleri

Şekil 5.9'da iz grafikleri incelendiğinde B_CRM için herhangi bir trende rastlanmadığı, parametrelere ait zincirlerin yakınsadığı sonucuna ulaşılmaktadır. Yoğunluk grafiğinden de düzeltilmiş modellere benzer şekilde log-serum bilirubin değişkeninin yaşam süresi üzerinde pozitif ve anlamlı bir etkisi olduğu ancak ilaç değişkeninin anlamsız etkiye sahip olduğu görülmektedir. Ancak B_CRM ile elde edilen log-serum bilirubin değişkeninin yaşam süresi üzerindeki etkisi düzeltilmiş modellere göre daha küçük elde edilmiştir.

Çizelge 5.9. B_CRM için Heidelberger-Welch yakınsaklık testi sonucu

Zincir	Parametre	Heidelberger-Welch Testi				
		Durağanlık Testi		Yarı Genişlik Testi		
		Başlangıç İterasyonu	p- Değeri	Ortalama	Yarı Genişlik Testi	Sonuç
[1]	β_1	1	0.787	1.090	0.000417	Geçti
	β_2	1	0.510	-0.102	0.001674	Geçti
[2]	β_1	1	0.594	1.090	0.000430	Geçti
	β_2	1	0.549	-0.099	0.001560	Geçti
[3]	β_1	1	0.441	1.090	0.000410	Geçti
	β_2	1	0.783	-0.100	0.001630	Geçti

Çizelge 5.9 incelendiğinde Heidelberger-Welch durağanlık ve yarı genişlik testleri sonucunda B_CRM 'de tüm parametreler için çalıştırılan üç zincirin yakınsadığı sonucuna ulaşılmıştır.

PBC verisi için DGB_CRM, DPB_CRM, B_CRM ile elde edilen beta parametresi için Bayesci tahminleri, standart sapmaları ve %95 Sonsal Bayes güven aralıkları ile kısmi olabilirlik fonksiyonuyla elde edilen CRM için katsayı tahmini ve %95 güven aralığı Çizelge 5.10'da yer almaktadır.

Çizelge 5.10. PBC verisi için model sonuçları

Model	Değişken	Katsayı	Std. Sapma	Alt Sınır	Üst Sınır	DIC
DGB_CRM	Log-serum bilirubin	1.3594	0.1264	1.1191	1.6149	2108
	İlaç	-0.2147	0.1825	-0.5737	0.1415	
DPB_CRM	Log-serum bilirubin	1.3149	0.1228	1.0807	1.5620	1949
	İlaç	-0.1857	0.1797	-0.5399	0.1650	
B_CRM	Log-serum bilirubin	1.0901	0.0948	0.9053	1.2770	1760
	İlaç	-0.1004	0.1694	-0.4324	0.2316	
CRM	Log-serum bilirubin	1.0966	0.0949	0.9105	1.2828	-
	İlaç	-0.1032	0.1708	-0.4380	0.2316	

Çizelge 5.10 incelendiğinde ele alınan tüm modeller için log-serum bilirubin değişkeninin ölüm riski üzerinde etkisinin yüksek olduğu görülmektedir. Benzetim senaryolarıyla karşılaştırıldığında büyük örneklerde, 0.60 durdurma oranında yalnızca 0.9 güvenirlilik düzeyinde CRM için minimum HKO elde edilmiş olmasına rağmen, düzeltilmiş modellerin de

yakın HKO verdiđi sonucu elde edilmiřti. Ölçüm hatalı deđiřken olan log-serum bilirubin deđiřkeni için gerçek deđerler elde edilemediđi için güvenirliliđi hakkında bilgi elde edilemez. Bu yüzden düzeltilmiř modellerin karřılařtırılıp yorumlanması daha güvenilirdir. Düzeltilmiř modeller karřılařtırıldıđında DPB_CRM'nin daha küçük DIC ile daha iyi bir model olduđu sonucuna ulařılmıřtır. Log-serum bilirubin deđeri arttıkça ölüm riskinin arttıđı söylenebilir. Parametrenin 1.0807-1.5620 aralıđına düşmesi olasılıđı % 95'tir.

5.2.2. Uygulama 2

Gerçek veri uygulamasının ikinci kısmında Goldman ve ark. (1996) tarafından kullanılmıř olan ve R'de yer alan veri seti kullanılmıřtır.

Didanosine (ddC) ve zalcitabine (ddI), zidovudin (AZT) tedavisini tolere edemeyen veya buna rađmen hastalık ilerlemesi olan HIV enfeksiyonu olan hastaları tedavi etmek için yaygın olarak kullanılan iki ilaçtır. Bu çalıřma AZT terapisinde bařarısız olan veya intöleransı olan hastaların tedavisinde bu iki antiretroviral ilacın etkinliđini ve güvenirliliđini karřılařtırmak için hem boylamsal hem de yařam verilerinin toplandıđı randomize bir klinik çalıřmadır. Antiretroviral tedavi sırasında ileri HIV enfeksiyonu olan kiřilerde CD4 lenfositinin deđeri, bir belirteç olarak sayılır ve 467 HIV ile enfekte olmuř hastanın AIDS didanosine / zalcitabin klinik çalıřması için Terry Beirn Topluluk Programlarından elde edilen veriler kullanılarak deđerlendirilmiřtir. Bu çalıřmada zidovudin intoleransı veya bařarısızlıđı için spesifik kriterleri yerine getiren AIDS veya iki CD4 sayısı ≤ 300 olan hastalar günde 500 mg ddI veya günde 2.25 mg ddC alacak řekilde randomize edilmiřtir. Çalıřmada CD4 sayıları çalıřmanın bařlangıcında ve sonrasında yapılan dört farklı ölçüm ile kaydedilmiřtir (Goldman ve ark., 1996).

Tez çalıřması kapsamında ölçüm hatası içeren açıklayıcı deđiřken olarak CD4 hücre sayısı ele alınmıřtır. Ölçüm hatalı açıklayıcı deđiřkenin ikinci tekrarlı ölçümü olarak da bir yıl sonra elde edilen ölçümü alınmıřtır. HIV ile enfekte olmuř AIDS hastalarının yařam süresi üzerinde ilaç (ölçüm hatası içermeyen deđiřken) ve CD4 hücre sayısı (ölçüm hatası içeren sürekli ölçekle ölçölmüř deđiřken) deđiřkenlerinin etkisinin incelenmesi olarak belirlenmiřtir.

Sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler Çizelge 5.11’de, nitel olan değişkenler ise Çizelge 5.12’de verilmiştir.

Çizelge 5.11. AIDS verisinde sürekli ölçekle ölçülmüş değişkenlere ait tanımlayıcı istatistikler

Değişken	n	%	Minimum	Maksimum	Ortalama	Ortanca	Standart sapma
Süre	467	100	0.47	21.40	12.63	13.20	4.94
CD4_0	467	100	0.00	19.24	7.13	6.08	4.71
CD4_12	226	48.4	0.00	20.44	7.03	5.00	5.27

CD4_0: Başlangıçtaki CD4 ölçümü, CD4_12: Bir yıl sonra elde edilen CD4 ölçümü

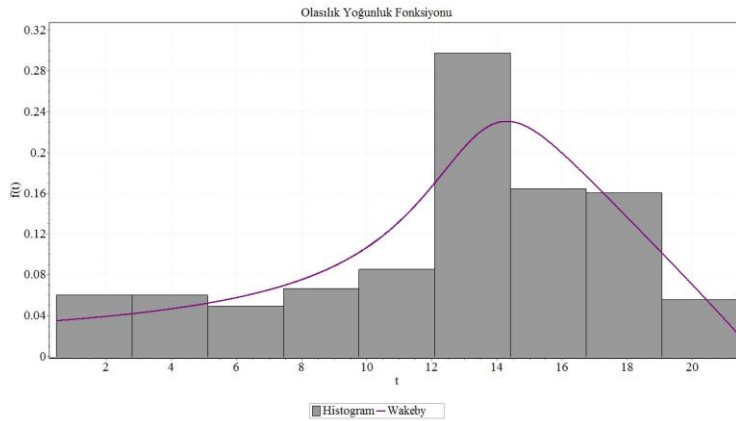
CD4 değişkeninin ikinci ölçümü çalışmaya dahil olan bireylerden %48.4’ü için elde edilmiştir, % 42.6’sı kayıp gözlemdir.

Çizelge 5.12. AIDS verisinde kullanılan nitel değişkenler

Değişken	Düzy	Frekans	Yüzde
Durum	0 (durdurma)	279	59.7
	1 (ölüm)	188	40.3
İlaç	ddC	237	50.7
	ddI	467	49.3

AIDS veri setine ilişkin durdurma oranı yaklaşık olarak %60’tır.

AIDS verisi için süreye ilişkin dağılım grafiği Şekil 5.10’da verilmiştir.



Şekil 5.10. AIDS verisinde süreye ilişkin dağılım grafiği

AIDS verisi için sürenin dağılımı incelendiğinde bilinen parametrik dağılımlardan birini göstermediği, Wakeby dağılımı gösterdiği gözlenmiştir.

Cox regresyon modelinde, incelen süre için her hangi bir dağılım varsayımı yoktur. Ancak orantılı tehlikeler varsayımının sağlanması gerekir.

Orantılı tehlikeler varsayımının kontrolü için Schoenfeld artıklarına dayanan orantılı tehlikeler varsayımı testi sonuçları Çizelge 5.13’de verilmiştir.

Çizelge 5.13. AIDS verisinde orantılı tehlikeler varsayımı test sonuçları

Değişken	rho	Ki-kare	p
İlaç	-0.004	0.003	0.954
CD4_0	0.069	0.979	0.322
Global	-	0.989	0.610

Çizelge 5.13 incelendiğinde Schoenfeld artıkları ile yaşam sürelerinin rankı arasındaki ilişkinin istatistiksel olarak önemli olmadığı, orantılı tehlikeler varsayımının sağlandığı sonucuna ulaşılmıştır.

Bayesci çözümlemeler için yakınsamanın sağlanıp sağlanmadığının kontrolünde iz grafiği çizdirilmiş, Gelman-Rubin yakınsama testi ve Heidelberger-Welch testi yakınsama ölçütü olarak kullanılmıştır. Tüm veri üzerinden çözümlemeler yapıldığında DPB_CRM için yakınsama elde edilememiştir.

Bu çalışmanın yapısına uygun, farklı örneklem büyüklüklerinde gerçek veri setlerine ulaşmak güçtür. Bu yüzden AIDS verisinden benzetim senaryolarına benzer olması için 30 örneklem büyüklüğü ve verinin orijinal yapısına uygun olması amacı ile 0.60 durdurma oranına sahip 11 veri seti rastgele olarak çekilmiştir. Seçilen her bir örneklem için orantılı tehlikeler varsayımı kontrol edilmiş ve orantılı tehlikeler varsayımının sağlandığı veri setleri için çözümlemeler yapılmıştır. Çözümlemelerde yakınsama ölçütlerine bakılmış, her bir veri seti için yakınsamanın

sağlandığı iterasyon sayısı ile 5 paralel zincir çalıştırılmıştır. Seçilen her bir örneklem için modelleri karşılaştırmak amacı ile DIC değeri hesaplanmıştır.

Örneklem büyüklüğünün 30, durdurma oranının 0.60 olduğu durum için seçilen örneklemle ilgili elde edilen katsayı değerleri, katsayıların standart hataları ve modellerin DIC değerleri Çizelge 5.14’de yer almaktadır.

Çizelge 5.14. AIDS verisinde örneklem büyüklüğü 30, durdurma oranı 0.60 olan örneklemle ilgili sonuçlar

Örneklem	Model	CD4_0 (std. sapma)	İlaç (std. sapma)	DIC
1	DGB_CRM	-0.7493 (0.7376)	-0.3857 (0.5913)	173.5
	DPB_CRM	-0.6680 (0.6333)	-0.3913 (0.5805)	<u>162.8</u>
2	DGB_CRM	-1.3817 (0.6556)	0.2504 (0.6369)	153.5
	DPB_CRM	-1.1383 (0.5527)	0.3286 (0.6045)	<u>145.5</u>
3	DGB_CRM	-0.8002 (0.6241)	-0.5028 (0.5730)	165.9
	DPB_CRM	-0.6796 (0.5490)	-0.4944 (0.5646)	<u>154.2</u>
4	DGB_CRM	-1.5802 (0.7527)	-0.5655 (0.6475)	162.9
	DPB_CRM	-1.2966 (0.6715)	-0.4492 (0.6113)	<u>156.4</u>
5	DGB_CRM	-0.9328 (0.6646)	-0.2471 (0.5997)	176.7
	DPB_CRM	-0.7649 (0.5453)	-0.2117 (0.5729)	<u>165.6</u>
6	DGB_CRM	-1.0990 (0.9420)	-0.1592 (0.5857)	132.9
	DPB_CRM	-0.9621 (0.8744)	-0.1148 (0.5742)	<u>123.3</u>
7	DGB_CRM	-0.5613 (0.6351)	-0.0239 (0.5617)	178.6
	DPB_CRM	-0.4677 (0.5558)	-0.0026 (0.5534)	<u>166.9</u>
8	DGB_CRM	-1.3125 (0.7468)	-0.0346 (0.6158)	157.5
	DPB_CRM	-1.1194 (0.6252)	0.0276 (0.5914)	<u>148.7</u>
9	DGB_CRM	-0.8530 (0.5457)	0.2860 (0.5909)	185.5
	DPB_CRM	-0.7423 (0.4939)	0.2429 (0.5753)	<u>174.0</u>
10	DGB_CRM	-1.7932 (0.7234)	-0.7618 (0.6330)	141.4
	DPB_CRM	-1.4614 (0.6090)	-0.7101 (0.5998)	<u>132.7</u>
11	DGB_CRM	-1.2932 (0.6310)	0.1126 (0.6923)	154.9
	DPB_CRM	-1.0556 (0.5421)	0.0558 (0.6689)	<u>146.1</u>

Çizelge 5.14 incelendiğinde rastgele seçilerek çalıştırılan 11 veri seti için de DPB_CRM ile elde edilen DCI, DGB_CRM ile elde edilen DIC değerinden daha küçük elde edilmiştir.

6. SONUÇ VE TARTIŞMA

Modelin ölçüm hatalı açıklayıcı değişken içermesi durumunda, parametre tahminlerinin ölçüm hatasının göz ardı edildiği klasik modellerle elde edilmesi, yana ve değişkenler arası ilişkilerin ortaya koyulması ve yorumlanmasında yanılgıya neden olmaktadır. Bu durumda ölçüm hatası için düzeltilmiş modellerin kullanılması daha uygundur.

Son yıllarda bilgisayar sistemlerinde teknolojik ilerlemeyle birlikte, karmaşık model yapısında analitik olarak çözümlenmesi güç olan Bayesci çıkarsamaların kullanımı, ölçüm hatası probleminin çözümünde de yaygın olarak kullanılmaya başlanmıştır. Bu tez çalışması ile ölçüm hatalı açıklayıcı değişken varlığında CRM ile yapılacak çözümlemelerde kullanılabilecek daha az yanlı parametre tahminlerinin elde edilebileceği alternatif bir yöntem olan DPB_CRM önerilmiş, bu yöntemin literatürde var olan yöntemlerle karşılaştırılması benzetim çalışması ile farklı örneklem büyüklüğü, durdurma oranı, güvenilirlik düzeyi ve regresyon katsayılarında ele alınmış ve etkin olduğu durumlar sunulmuştur.

Literatürde gösterildiği üzere ölçüm hatası içermeyen durumlarda olduğu gibi ölçüm hatalı açıklayıcı değişken olması durumunda da verilen tüm senaryolarda B_CRM ve CRM benzer sonuç vermiştir.

Beta katsayısının 0.5 olduğu yani riskin düşük olduğu durumlarda bir durum hariç DGB_CRM diğer modellerden daha düşük performans gösterdiği için uygun bir alternatif değildir. Güvenirliğin 0.5 ve 0.7 olduğu bazı durumlarda B_CRM için daha küçük HKO elde edilmesine rağmen DPB_CRM'nin hem HKO'su bu modelle yakın değer vermiş hem de KO değeri daha uygun bulunmuştur. Güvenirliğin 0.9 olduğu durumlarda ise klasik model ya da B_CRM daha iyi sonuç vermiştir.

Beta katsayısının 1.0 olduğu durumda 0.5 ve 0.7 güvenilirlik düzeylerinde DGB_CRM genel olarak daha iyi sonuç vermiş, DPB_CRM'nin daha küçük HKO elde edildiği senaryolarda ise DGB_CRM ile değerlerinin çok yakın olduğu, bu nedenle bu modellerin birbirlerine alternatif olduğu söylenebilir. Güvenirliğin 0.9 olduğu durumlarda ise B_CRM ya da CRM bir çok senaryoda HKO değeri bakımından daha iyi sonuç vermiş, bazı durumlarda DPB_CRM'ye ait

HKO daha küçük ancak B_CRM ve CRM ile HKO değerleri yakındır. Küçük örnekleme KO bakımından B_CRM ve CRM normal değerine yakın iken örneklem büyüklüğü arttıkça KO değeri kabul edilebilir sınırların altında kalmış, elde edilen HKO ve KO bakımından DPB_CRM'nin uygun bir alternatif olarak kullanılabileceği sonucuna ulaşılmıştır.

Beta katsayısının 1.5 olduğu yani riskin yüksek olduğu durumlarda ise klasik modelin kullanılmasının ciddi bir yana neden olduğu, güvenilirliğin 0.5 ve 0.7 olduğu durumlarda güvenilirliğin 0.9 olduğu duruma göre daha yanlış olduğu ve örneklem büyüklüğü arttıkça yanlışlığın arttığı benzetim çalışması sonucu ortaya koyulmuştur. Risk yüksek ise ölçüm hatası içeren açıklayıcı değişkenin varlığı durumunda parametre tahminleri yapılırken çok dikkat edilmelidir. Benzetim çalışması sonucu klasik model yerine ölçüm hatası için düzeltilmiş modellerin daha iyi sonuç verdiği, DGB_CRM'nin bu modeller içinde en iyi performans gösteren model olduğu sonucu elde edilmiştir.

Genel olarak yorumlanacak olursa güvenilirliğin düşük ve orta olduğu durumlarda klasik model ve Bayesci yaklaşımının düzeltilmiş yöntemlere göre KO bakımından performansının düşük olduğu söylenebilir. Güvenirliğin yüksek olduğu durumlarda ise daha küçük HKO elde edilmesine rağmen düzeltilmiş modellerden en az biri yakın HKO ile alternatif bir yöntem olmuşlardır. Gerçek veri çözümlemelerinde ölçüm hatası içeren değişkenin gerçek değeri çoğunlukla bilinmediği için güvenilirliği elde edilemez. Benzetim çalışmasında da görüldüğü gibi düşük ve orta güvenilirliklerde olduğu gibi yüksek güvenilirliklerde de çözümlemeye düzeltilmiş modellerin kullanılması klasik modellere göre daha güvenilir sonuçlar elde edilmesini sağlayacaktır. Bu nedenle modelde ölçüm hatalı açıklayıcı değişken bulunması durumunda klasik modeller değil, ölçüm hatası için düzeltilmiş modeller kullanılmalıdır.

Özellikle büyük örneklerde ya da riskin yüksek olduğu durumlarda ölçüm hatalı açıklayıcı değişkenli modellerin çözümlemesinde klasik çözümlemelerin kullanılması araştırılan değişkenin çıktı üzerindeki etkisini yorumlarken araştırmacıyı yanlış sonuçlara götürecektir. Epidemolojik çalışmalarda da sıklıkla karşımıza çıkan ölçüm hatalı değişkenlerin, yaşam çözümlemesi gibi hassas yorumlama gerektiren çalışmalarda kullanılırken parametre tahminlerinde kullanılacak çözümlemeye dikkatle karar verilmesi gerekmektedir.

Bartlett ve Keogh (2018) alıřmalarında lm hatası iin Bayesci dzeltmeyi birikimli tehlike fonksiyonu iin gamma nseli kullanarak uygulamıřlar, regresyon kalibrasyonu yntemi ve Bayesci dzeltme ile elde ettikleri sonuları karřılařtırmıřlardır. Bu alıřmada, farklı durdurma oranlarını ve kk rneklem durumlarını karřılařtırmaya dahil etmemiřlerdir. Bk rneklem durumunda ise kullandıkları senaryolarda nerdikleri yntemin daha iyi ya da regresyon kalibrasyonu yntemi ile benzer sonulara ulařtıklarını belirtmiřlerdir. Literatr incelendiğinde tm bu deėiřenlere gre lm hatası iin Bayesci dzeltmiř model ile klasik yntemleri karřılařtıran bařka alıřma yer almamaktadır. Bu tez alıřması ile lm hatalı aıklayıcı deėiřkenli CRM iin farklı rneklem bklė, durdurma oranı, gvenirlik ve risk dzeylerine sahip veri setlerinde klasik yntemler ile lm hatası iin dzeltmiř Bayesci yntemlerin regresyon katsayısının kestirimlerinde oluřturdukları yanlılıėın bklė detaylıca ortaya koyulmuřtur.

7. KAYNAKLAR

- Aalen, O.O., Statistical inference for a family of counting processes, ProQuest LLC. Ann Arbor, MI, **1975**.
- Andersen P.K., Gill R.D., Cox's regression model for counting processes: A large sample study, *Annals of Statistics*, 10 (**1982**) 1100–1120.
- Augustin, T., An Exact Corrected Log-Likelihood Function for Cox's Proportional Hazards Model under Measurement Error and Some Extensions, *Scandinavian Journal of Statistics*, 31(1) (**2004**) 43-50.
- Bartlett, J.W., Keogh, R.H., Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. *Statistical Methods in Medical Research*, 27(6) (**2018**) 1695-1708.
- Beamonte, E., Bermúdez, J.D., A Bayesian semiparametric analysis for additive hazards models with censored observations, *Test*, 12 (**2003**) 2, 101–117.
- Bender, R., Augustin, T., Blettner, M., Generating survival times to simulate Cox proportional hazards models, *Statistics in Medicine*, 24(11) (**2005**) 1713-1723.
- Buonaccorsi, J.P., Measurement error: models, methods, and applications, Chapman and Hall/CRC, **2010**.
- Buzas, J.S., Unbiased scores in proportional hazards regression with covariate measurement error, *Journal of Statistical Planning and Inference*, 67(2) (**1998**) 247-257.
- Carroll, R.J., Ruppert, D., Crainiceanu, C.M., Stefanski, L.A., Measurement error in nonlinear models: a modern perspective, Chapman and Hall/CRC, **2006**.
- Clayton, D., A Monte Carlo for Bayesian inference in frailty models, *Biometrics*, 47 (**1991**) 467– 485.
- Clayton, D.G., Models for the analysis of cohort and case-control studies with inaccurately measured exposures, In: Dwyer, J. H., Feinlieb, M., Lippert, P. and Ho_meister, H. (Ed.) *Statistical Models for Longitudinal Studies on Health*, New York: Oxford University Press, **1992**.
- Collett, D., Modelling survival data in medical research, CRC press, **1994**.
- Cox, D.R., Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2) (**1972**) 187-202.
- Espino-Hernandez, G., Gustafson, P. and Burstyn, I., Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study. *BMC medical research methodology*, 11(1) 67 (**2011**).

- Gamerman, D., Dynamic Bayesian models for Survival data, *Applied Statistics*, 40 (1991) 63–79.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D., *Bayesian Data Analysis*, New York: Chapman and Hall/CRC, 2013.
- Ghosh, J.K., Delampady, M., Samanta, T., *An introduction to Bayesian analysis: Theory and methods*, Springer Science and Business Media, 2007.
- Gill, R.D., Product-integration, *Encyclopedia of Biostatistics* (eds P. Armitage and T. Colton), 6 (2005) 4246 – 4250.
- Goldman, A., Carlin, B., Crane, L., Launer, C., Korvick, J., Deyton, L., Abrams, K., Response of CD4+ and clinical consequences to treatment using ddI or ddC in patients with advanced HIV infection, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11 (1996) 161–169.
- Grace, Y.Y., Lawless, J.F., A corrected likelihood method for the proportional hazards model with covariates subject to measurement error, *Journal of Statistical Planning and Inference*, 137(6) (2007) 1816-1828.
- Gray, C.M., Use of the Bayesian family of methods to correct for effects of exposure measurement error in polynomial regression models, PhD Thesis, London School of Hygiene and Tropical Medicine, London, 2018.
- Gustafson, P., *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*, Boca Raton, FL, USA: Chapman and Hall/CRC, 2003.
- Hamra, G.B., MacLehose, R.F., Cole, S.R., Sensitivity analyses for sparse-data problems using weakly informative Bayesian priors, *Epidemiology*, 24 (2013) 233–239.
- Hu, C. and Lin, D.Y., Cox regression with covariate measurement error, *Scandinavian Journal of Statistics*, 29(4) (2002) 637-655.
- Hu, P., Tsiatis, A., Davidian, M., Estimating the parameters in the Cox model when covariate variables are measured with error, *Biometrics*, 54(4) (1998) 1407-1419.
- Ibrahim, J.G., Chen, M.H. and Sinha, D., *Bayesian survival analysis*, Springer-Verlag, New York, 2001.
- Keogh, R.H. and White, I.R., A toolkit for measurement error correction, with a focus on nutritional epidemiology, *Statistics in Medicine*, 33(12) (2014) 2137-2155.
- Kleinbaum, D.G. and Klein, M., *Survival analysis* (Vol. 3), New York: Springer, 2010.
- Lawless, J.F., *Statistical Models and Methods for Lifetime Data*, (2nd Ed.) Wiley: New York, 2003.
- Lesaffre, E., Lawson, A.B., *Bayesian Biostatistics*, John Wiley & Sons, 2012.
- Liu, X., *Survival analysis: models and applications*, John Wiley & Sons, 2012.

- Martin-Magniette, M.L. and Taupin, M.L., Estimation of the hazard function in a semiparametric model with covariate measurement error, *ESAIM: Probability and Statistics*, 13 (2009) 87-114.
- Mostafa, A. and Ghorbal, A.. Using WinBUGS to Cox model with changing from the baseline hazard function, *Applied Mathematical Sciences*, 45(5) (2011) 2217-2240.
- Muff, S., Riebler, A., Held, L., Rue, H. and Saner, P., Bayesian analysis of measurement error models using integrated nested Laplace approximations, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2) (2015) 231-252.
- Nakamura, T., Proportional hazards model with covariates subject to measurement error, *Biometrics*, (1992) 829-838.
- Omurlu, I.K., Ozdamar, K. and Ture, M., Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets, *Expert Systems with Applications*, 36(8) (2009) 11341-11346.
- Prentice, R.L., Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika*, 69 (1982) 331-342.
- Ray, A., Primary Biliary Cirrhosis, https://rstudio-pubs-static.s3.amazonaws.com/159812_042b6e22b9cf44639fb26ae8b2df0a98.html (Erişim Tarihi: 6 Mayıs 2020).
- Schneeweiss, H., Augustin, T., Some recent advances in measurement error models and methods, *Allgemeines Statistisches Archiv*, 90(1) (2006) 183-197.
- Song, X. and Huang, Y., On corrected score approach for proportional hazards model with covariate measurement error, *Biometrics*, 61(3) (2005) 702-714.
- Wang, C.Y., Hsu, L., Feng, Z.D., Prentice, R.L., Regression calibration in failure time regression, *Biometrics*, 53(1) (1997) 131-145.
- Weeding, J.L., Bayesian measurement error modeling with application to the area under the curve summary measure, Doctoral dissertation, Montana State University-Bozeman, College of Letters & Science, 2016.
- Xie, S.X., Wang, C.Y., Prentice, R.L., A risk set calibration method for failure time regression by using a covariate reliability sample, *Journal of the Royal Statistical Society Series B*, 63(4) (2001) 855-870.
- Yi, G.Y. and Lawless, J.F., A corrected likelihood method for the proportional hazards model with covariates subject to measurement error, *Journal of Statistical Planning and Inference*, 137 (2007) 1816-1828.
- Zhou, H. and Wang, C.Y., Failure time regression with continuous covariates measured with error, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4) (2000) 657-665.

EKLER

EK 1 - Tezden Türetilmiş Bildiriler

Işık, H. ve Karasoy, D., Cox Regression Model with Covariates Subject to Measurement Error,
11.International Statistics Days Conference, Muğla, 2018.





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

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 06 /07 /2020

Tez Başlığı / Konusu: ÖLÇÜM HATALI AÇIKLAYICI DEĞİŞKENLİ COX REGRESYON MODELİNE BAYESÇİ YAKLAŞIM

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 59 sayfalık kısmına ilişkin, 06/07/2020 tarihinde şahsım/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 8 'dir.

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.

06 /07/2020

Adı Soyadı: Hatice IŞIK
Öğrenci No: N13246256
Anabilim Dalı: İstatistik
Programı: İstatistik
Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Duru KARASOY

ÖZGEÇMİŞ

Adı Soyadı : Hatice IŞIK
Doğum yeri : Ankara
Doğum tarihi : 30 Haziran 1986
Yazışma adresi : Hacettepe Üniversitesi İstatistik Bölümü Beytepe, Çankaya,
ANKARA
Telefon : -
Elektronik posta adresi : haticeyeniay@hacettepe.edu.tr
Yabancı dili : İngilizce

EĞİTİM DURUMU

Lisans : Hacettepe Üniversitesi, İstatistik Bölümü
Yüksek Lisans : Hacettepe Üniversitesi, İstatistik Bölümü
Doktora : Hacettepe Üniversitesi, İstatistik Bölümü

İş Tecrübesi

2011 (Şubat-Haziran) : Formosa Uluslararası Ticaret A.Ş.
2013-... : Hacettepe Üniversitesi İstatistik Bölümü Araştırma Görevlisi