GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER İÇİN SINIRLI DALGALANMALI KREDİBİLİTE YAKLAŞIMI

LIMITED FLUCTUATION CREDIBILITY APPROACH FOR GENERALIZED LINEAR MODELS

ÖVGÜCAN GÖNENÇ KARADAĞ

Doç. Dr. MERAL SUCU

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin

Aktüerya Bilimleri Anabilim Dalı İçin Öngördüğü

YÜKSEK LİSANS TEZİ

olarak hazırlanmıştır.

2014

ÖVGÜCAN GÖNENÇ KARADAĞ'ın hazırladığı **"Genelleştirilmiş Doğrusal Modeller İçin Sınırlı Dalgalanmalı Kredibilite Yaklaşımı"** adlı bu çalışma aşağıdaki jüri tarafından **AKTÜERYA ANABİLİM DALI**' nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan

Doç. Dr. Serpil AKTAŞ ALTUNAY

Danışman

Doç. Dr. Meral SUCU

Üye

Yrd. Doç. Dr. Yasemin GENÇTÜRK

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

Prof. Dr. Fatma SEVİN DÜZ

Fen Bilimleri Enstitüsü Müdürü

ETİK

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

- 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 tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

26/ 06/ 2014

Övgücan Gönenç KARADAĞ

ÖZET

GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER İÇİN SINIRLI DALGALANMALI KREDİBİLİTE YAKLAŞIMI

ÖVGÜCAN GÖNENÇ KARADAĞ,

Yüksek Lisans, Aktüerya Bilimleri Bölümü

Tez Danışmanı: Doç. Dr. MERAL SUCU

Haziran 2014, 99 sayfa

Bu çalışmada hayat dışı sigorta ürünlerinin fiyatlandırılmasında sıklıkla kullanılan Genelleştirilmiş Doğrusal Modeller (GDM) ve Kredibilite Kuramı birlikte ele alınmıştır. Üstel Dağılım Ailesi (ÜDA)'ndeki dağılımlar için GDM'ler incelenmiştir. Tam Kredibilite Standardı kullanılarak, GDM'ler ile Sınırlı Dalgalanmalı Kredibilite yaklaşımı arasında ilişki kurulmuştur. GDM'deki açıklayıcı değişkenler yardımıyla oluşturulan risk sınıflarının güvenilirliğinin analizinde kullanılabilecek karşılaştırma kriterleri belirlenmiştir.

Özel bir sigorta şirketinin bir yıllık kasko sigortası poliçelerinden gelen hasar sayısı verisi ile iki aşamalı bir uygulama yapılmıştır. Birinci aşamada, model açıklayıcı değişkenlerinin belirlenmesi için modelleme basamakları teker teker uygulanmış, parametre tahminleri yapılmış ve belli bilgi kriterlerine göre model seçimi yapılarak, veri GDM ile modellenmiştir. İkinci aşamada modelin açıklayıcı değişkenlerine göre oluşturulmuş risk sınıflarının aktüeryal değerleme için uygunluğu ile tam kredibilite sağlayan risk sınıfının olup olmadığı analiz edilmiştir. Oluşturulan modelde tam kredibilite sağlayan bir risk sınıfı olduğu görülmüştür. Bu risk sınıfı küçük motor hacimli arabaya sahip erkek sigortalıları temsil etmektedir ve bu sınıfın hasar deneyiminin fiyatlama için güvenilir olduğu saptanmıştır. Tam kredibilitenin sağlanmadığı risk sınıfları kendi aralarında karşılaştırılmış

ve tam güvenilir olmaları için gereken minimum gözlem sayısı belirlenmiştir. Minimum gözlem sayısı Sınırlı Dalgalanmalı Tam Kredibilite yaklaşımı kullanılarak; dağılım bilgisine ve hata-tahmin toleransına göre hesaplanmaktadır. Ayrıca hasar sayısının, poliçe sayısının, örneklem büyüklüğünün ve GDM bileşenlerinden açıklayıcı değişkenler ile bağ fonksiyonunun kredibilite üzerindeki etkileri incelenmiştir. Hasar sayısı, poliçe sayısı ve örneklem sayısı arttıkça güvenilirlik artmakta ve risk sınıfı tam kredibiliteye yaklaşmaktadır. Açıklayıcı değişkenin etkisini görmek için, birçok açıklayıcı değişkenden oluşan modelde, etkisi incelenmek istenen açıklayıcı değişken dışında diğer açıklayıcı değişkenler aynı alınarak, açıklayıcı değişkenlerin risk sınıflarının güvenilirliği üzerinde etkileri incelenmiş ve açıklayıcı değişkenin güvenilirlikte etkili bir unsur olduğu görülmüştür. Model bileşenlerinden bağ fonksiyonu yapısının ve türünün kredibiliteyi etkileyip etkilemediğinin cevabını aramak için bağ fonksiyonunun bir sabit ile çarpılarak yapısı, logaritmik bağ fonksiyonu yerine birim bağ fonksiyonu alınarak türü değiştirilmiştir. Karşılaştırma sonucu bağ fonksiyonunun yapısının ve türünün değişiminin kredibiliteyi etkilemediği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Genelleştirilmiş Doğrusal Model, Tam Kredibilite, Üstel Dağılım Ailesi, Log- Olabilirlik Fonksiyonu

ABSTRACT

LIMITED FLUCTUATION CREDIBILITY APPROACH FOR GENERALIZED LINEAR MODELS

ÖVGÜCAN GÖNENÇ KARADAĞ,

Master of Science, Departmant of Actuarial Sciences

Supervisor: Assoc. Prof. Dr. MERAL SUCU

June 2014, 99 pages

In this study, Generalized Linear Models (GLM) examined and Credibility Theory which are frequently used in non-life insurance pricing are combined. GLMs are examined for distributions in Exponential Family (EF). Using Full Credibility Standart, GLMs are associated with Limited Fluctuation Credibility approach. By GLM explanatory variables, comparison criterias which can be used in analyses of risk classes's credibility are defined.

Two-stage application is performed by using one-year claim number data of a special insurance company policies. At first stage, modelling steps are performed one by one for the determination model explanatory variables, paramaters are estimated and by choosing a model according to some criteria, data are modeled by GLM. At second stage, suitability of risk classes set with explanatory variables for actuarial valuation and whether there are any full credibility risk classes or not are analized. To create the model, one full credibility risk class is determined. This risk class represents insured males who have small engine size cars and the claim experience of this class is found to be credible for pricing. Risk classes not providing full credibility are compared with each other and minimum observation number for full credibility is determined. Using Limited Fluctuation Full Credibility approach, minimum observation number is calculated according to the distribution and estimation-error tolerance. Furthermore, the effect of the claim number, the number of policy, sample size and components of GLM (explanatory variable and link function) on credibility is analized. The more the claim numbers, the number of policy and sample size are, the more credible the study gets which leads risk class to be approximately full credible. To observe the effect of explanatory variables, the effects of explanatory variables on credibility of risk classes are analized by supposing that in the model that constitutes lots of explanatory variables, the explanatory variables apart from the ones whose effects are desired to be observed are taken as the same and it is observed that explanatory variable is an efficient factor at credibility. To find the answer whether structure and kind of link function, one of the model components, affect the credibility, kind of link funtion are changed by multiple with a constant structure of link function and by taking unit link function instead of log-link funtion. After this comparison, it is observed that changing structure and kind of link function do not affect the credibility.

Keywords: Generalized Linear Model, Full Credibility, Exponenxial Family, Log-Likelihood Function

TEŞEKKÜR

Tez çalışmamın her aşamasında yol gösteren, bilgi ve deneyimiyle karşılaşılan güçlüklerin giderilmesinde yardımcı olan danışmanım Sayın Doç. Dr. Meral SUCU'ya,

Göstermiş olduğu anlayış ve destek için Bölüm Başkanımız Sayın Doç. Dr. Kasırga YILDIRAK'a,

Değerli eleştiri ve katkıları için tez jüri üyelerim Sayın Doç. Dr. Serpil AKTAŞ ALTUNAY'a ve Sayın Yrd. Doç. Dr. Yasemin GENÇTÜRK'e,

Desteklerini esirgemeyen bölüm hocalarıma ve arkadaşlarıma,

Tezimin uygulama kısmında kullanılan veriye ulaşmamı sağlayan Taha GÜNDOĞDU'ya,

Destek ve arkadaşlıkları için Arş. Gör. Murat KIRKAĞAÇ'a, Arş. Gör. Rümeysa KARATAŞ'a ve sevgili oda arkadaşım Arş. Gör. Aslıhan ŞENTÜRK ACAR'a,

Akademisyenlik mesleğinde en başından beri en büyük destekçim ve örnek aldığım insan babam Prof. Dr. Erdener KARADAĞ'a, yeri doldurulamaz öğütleriyle hayattaki zorluklara karşı daha dayanıklı olmamı sağlayan annem Selma KARADAĞ'a ve babamızın yolundan beraber gittiğimiz meslektaşım ablam Arş. Gör. Özge KARADAĞ'a,

Hayatımın her aşamasında gösterdiği anlayış ve destek için Ömer ERDEMİR'e,

Evdeki çalışmalarımda beni yalnız bırakmayan minik dostum Çikocan'a,

Teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ŞEKİLLER

ÇİZELGELER

Sayfa

KISALTMALAR

Kısaltmalar

AEKK	Ağırlıklandırılmış En Küçük Kareler Tahmini
EÇO	En Çok Olabilirlik Tahmini
EKK	En Küçük Kareler Tahmini
GDKM	Genelleştirilmiş Doğrusal Karma Model
GDM	Genelleştirilmiş Doğrusal Model
GEKK	Genelleştirilmiş En Küçük Kareler Tahmini
ÜDA	Üstel Dağılım Ailesi

1. GİRİŞ VE ÖNCEKİ ÇALIŞMALAR

1.1. Giriş

6102 sayılı yeni Türk Ticaret Kanunu (TTK)’nun 1401. maddesinde sigorta tanımı şöyle yapılmaktadır: “ Sigorta sözleşmesi, sigortacının bir prim karşılığında, kişinin para ile ölçülebilir bir menfaatini zarara uğratan tehlikenin, rizikonun, meydana gelmesi hâlinde bunu tazmin etmeyi ya da bir veya birkaç kişinin hayat süreleri sebebiyle ya da hayatlarında gerçekleşen bazı olaylar dolayısıyla bir para ödemeyi veya diğer edimlerde bulunmayı yükümlendiği sözleşmedir. ” Bu tanım hayat ve hayat dışı sigortaları kapsar. Hayat sigortaları bireylerin yaşam veya ölüm durumları, hayat dışı sigortalar ise hasar sayısı ve hasar tutarı dikkate alınarak fiyatlandırılır. Genellikle aktüerler hayat dışı sigortaların fiyatlandırılmasında Genelleştirilmiş Doğrusal Model (GDM) gibi birçok fiyatlandırma faktörünün tahmin edildiği çoklu modellerden yararlanırlar [1].

Çoklu doğrusal modeller; regresyon modelleri, GDM’ler, Genelleştirilmiş Doğrusal Karma Modeller (GDKM) gibi çeşitlilik gösterir. Bu modellerden klasik regresyon modelleri sigorta verisi için uygun olmayan modellerdir. Analizler yapılırken hasar sayılarının Poisson dağılımına sahip olduğu varsayımının yapılması, yani normal dağılmaması klasik regresyon çözümlemesi yapılmasına engel olur. Bu durumda GDM kullanılır. Çünkü bu modellerde yanıt değişkeninin Üstel Dağılım Ailesi (ÜDA)(Poisson, Gamma, Normal, Binom, Negatif Binom, Ters Gauss) içinde yer alan bir dağılıma sahip olması varsayımı yeterlidir. Bu nedenle sigorta veri kümelerinin istatistiksel analizinde GDM kullanımı son yıllarda önem kazanmaktadır.

Sigorta ürünlerinin fiyatlandırmasında kullanılan diğer bir yaklaşım ise Kredibilite Kuramı’dır. Kredibilite Kuramı, sigorta ürününün önceki dönemlerine ait verilerden yararlanarak, gelecek dönemlere ait beklenen hasar veya prim tahmininde kullanılır. Antonie ve Beirlant [2]’a göre kredibilite fiyatlandırma problemi, belli bir grubun hasar deneyimi ile ilişkili risk sınıfının deneyimi birleştirilerek risk primine karar verir. Bu kuram özellikle hayat dışı sigortaların fiyatlandırılmasında tercih edilen bir yöntemdir.

1.2. Önceki Çalışmalar

GDM ile ilgili ilk çalışma 1972 yılında Nelder ve Wedderburn tarafından yapılmıştır [3]. McCullagh ve Nelder [4] GDM Teorisi'ni temel hatlarıyla vermişlerdir. Aktüerya alanındaki ilk GDM çalışmaları 1991 yılında Renshaw [5] ve 1994 yılında Renshaw ve Verall [6] tarafından yapılmıştır. 1996 yılında Haberman ve Renshaw [7] aktüeryal veri analizinde GDM'lerin kullanımını detaylı olarak incelemişlerdir ve GDM'nin hayat ile hayat dışı sigorta alanlarında kullanımlarını göstermişlerdir. GDM'in sıklıkla kullanıldığı aktüerya bilimleri konuları; yaşam modellemesi, ölümlülük, çoklu-durum modelleri (özellikle sağlık sigortalarında), lapse (primlerin ödenmemesi nedeniyle sigorta hakkının kaybedilmesi), risk sınıflandırması (sigara içenlerin ölümlülüğünün yüksek olması, bağımlı yaşayan durumunda ölümlülük), hayat dışı sigortalarda hasar büyüklüğü için kayıp (loss) dağılımlarının uyumu, prim fiyatlandırması (hasar sıklığı ve hasar tutarı), hayat-dışı sigortalarda rezervler ve kredibilitedir. Anderson et al. [8] GDM'lerin uygulamalarını anlatan; Jong ve Heller [9] ise sigorta veri analizinde kullanılan GDM'lere ilişkin kapsamlı birer kitap yayınlamışlardır. GDKM, GDM'in rastgele etkilerle genişletilmiş halidir. Lee ve Nelder [10] GDKM üzerine çalışmışlardır. Antonie ve Beirlant [2] Kredibilite Kuramı, ölümlülük hesabı, hasar rezervleri ve kredi riskinin modellenmesi gibi çeşitli aktüeryal problemlerde GDKM kullanımı göstermişlerdir. Hayat dışı sigorta ürünlerinin fiyatlandırılmasında GDM'lerin dışında tercih edilen diğer bir yaklaşım olan Kredibilite kavramı ilk olarak 1914 yılında Mowbray tarafından yapılan çalışma ile gündeme gelmiştir [11]. Mowbray [11] iş kazası sigortalarında tam güvenilirlik için gereken minimum gözlem sayısı üzerine bir çalışma yapmış; bu çalışmalarda gözlem sayısı için belirlenen rastlantı değişkeni binom dağılımlı varsayılmış ve normal yakınsama ile tam güven için gözlem sayısının alt sınırı bulunmuştur. Tam güvenilirliliğin sağlanmasının zor olacağı düşüncesiyle Whitney [12] Kısmi Kredibilite kavramını gündeme getirmiş, tüm risk sınıfı verisi ile bireysel veriyi ağırlıklandırarak deneyim fiyatlandırması yapmıştır. Whitney'in Kısmi Kredibilite yaklaşımı, homojenliği temel alması nedeniyle En Büyük Doğruluk Kredibilitesi'nin ilk basamağını oluşturur [13]. Temel anlamda Bailey [14]'in çalışması En Büyük Doğruluk Kredibilitesi'ni Bayes Teoremi ile ilişkilendirerek farklı bir bakış açısı getirmiştir. Bailey [14], önsel dağılımı önemsemeden sadece verilerin bilgisine göre kredibilite prim tahmini yapılması doğru gelmediğinden bayesci bir yaklaşım benimsenmiştir. Bühlmann [15] dağılımdan bağımsız bir şekilde En Küçük Kareler (EKK) yöntemini kullanarak Bühlmann Kredibilite yaklaşımını oluşturmuştur. Bühlmann [16] ve

Bühlmann-Straub [17] kredibiliteye daha istatistiksel ve teorik olarak yaklaşmışlar; Hachemeister [8] ve Jewell [19] ise hiyerarşik kredibilite üzerine çalışmalar yapmışlardır. Goulet [13] Sınırlı Dalgalanmalı Kredibilite, En Büyük Doğruluk Kredibilitesi, Bayesci Kredibilite, Bühlmann- Straub Kredibilite, Hiyerarşik Kredibilite ve Çapraz Sınıflandırma Kredibilite Modelleri gibi tüm kredibilite modellerine ilişkin çalışmasında, parametre tahminleri yapmış ve modellerin hangi durumlarda kullanılacağını açıklamıştır. Norberg [20] kredibilite yaklaşımlarından Sınırlı Dalgalanmalı Kredibilite yaklaşımını kredibilitenin "Sabit Etki" teorisi ve En Büyük Doğruluk Kredibilite yaklaşımını da kredibilitenin "Rastgele Etki" teorisi olarak adlandırılabileceğini ileri sürmüştür. Dannenburg et al. [21] aktüeryal kredibilite modelleri üzerine çalışmalar yapmışlardır. Kaas et al. [22] kredibilite problemlerini bayesci tahmin problemleri olarak ele almış ve En Büyük Doğruluk Kredibilite yaklaşımı üzerinde durmuşlardır. Campbell [23] Kredibilite Kuramı'nı araba modellerinin sınıflandırılmasında kullanmıştır. Sundt [24]'ın 1987 yılındaki çalışması Kredibilite Kuramı'nın araç sınıflandırmasında kullanıldığı diğer bir çalışmadır. Schmitter 2004 yılında [25] ihtiyaç olan hasar sayısını risk faktörlerine göre tahmin etmek için bir yöntem geliştirmiş, hasarların Poisson dağılımlı olduğu varsayımı altında Poisson parametresi tahmini yaparak minimum gözlem sayısı için bir üst sınır hesaplamıştır. Mahler ve Dean [26] Kredibilite Kuramı'nın temeli, önemli varsayımları ve model çeşitleri üzerinde durmuşlardır. Şentürk [27] 2010 yılı tez çalışmasında Kredibilite Kuramı'nı temel olarak anlatmış ve yaklaşımlarına değinmiştir. Şentürk çalışmasının uygulama kısmında grup verilerini bireysel veri gibi düşünerek panel veri analizine zemin hazırlamış ve Türkiye'deki tüm iller için Kredibilite primi hesabı yapmıştır. Panel veri analizi bir değişkenin belli zaman dilimlerindeki gözlemlerini içeren zaman serisi analizleri ile seçilen bir zamanda birden çok değişkeninin gözlemini içeren kesit verisi analizlerini birleştiren bir yöntemdir.

Hayatdışı sigorta ürünlerinin fiyatlandırılmasında kullanılan bu iki yaklaşımı birleştirmek de mümkümdür. İlk olarak 1997 yılında Nelder ve Verall [28] Kredibilite Kuramı ile GDM'leri birlikte ele alırken Bühlmann Kredibilite ile Hiyerarşik Genelleştirilmiş Doğrusal Modellerler (HGDM) arasında ilişki kurmuşlardır. HGDM-Kredibilite modelleri tüm ortalamayı sabit etki olarak ve bağ fonksiyonuna göre belirlenen ortalamayı rastgele etki olarak alan modellerdir. Hiyerarşik denmesinin nedeni sabit-rastgele etki parametreleri ile yayılım parametresinin iki eşitlik arasında tekrarlı (iterative) bir şekilde tahmin etmesidir. Nelder ve Verall [28]'ın oluşturduğu bu modeller prim fiyatlandırması ve hasar rezervlerinin hesaplanmasında kullanılmaktadır. Antonie ve Beirlant [2] aktüerya istatistiği üzerine yaptıkları çalışmada, GDKM'leri belli varsayımlara göre Kredibilite modellerine dönüştürmüşlerdir. Bühlmann Model, Bühlmann-Straub Model, Hachemeister Model, Jewell'ın Hiyerarşik Kredibilite Modeli ve Dannenburg'un Çapraz Sınıflandırma Modelleri'nin GDKM tanımlarını yaptıktan sonra, $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta + T'_{ij}u_i$ biçimindeki GDKM için β ve u_i parametreleri ile X'_{ij} - T'_{ij} açık layıcı değişkenlerinin tahminlerini belli varsayımlara göre yaparak bu GDKM'leri Kredibilite modellerine dönüştürmüşlerdir [2]. Çoklu faktörleri içeren hayat dışı fiyatlandırma problemlerinde GDM ile Kredibilite Kuramını birleştiren diğer bir çalışma 2008 yılında Ohlsson [1] tarafından yapılmıştır. Değişkene ilişkin veri az sayıda (örneğin 2500 araba modeli) ve çok faktörlü bir değişken ise modellemenin yapılabilmesi için GDM'den ziyade Kredibilite Kuramına ihtiyaç duyulabilir. Ohlsson [1] yaptığı çalışmada bir İşviçre sigorta şirketinden alınan kasko sigortası verisi ile trafik sigortalarında araba modeli sınıflandırması yaparken bu iki modeli birleştirmiştir. Garrido ve Zhou [29] 2009 yılında yaptıkları çalışmada bir GDM tahmin edicisinin güvenilir olması için en az kaç tane gözleme ihtiyaç vardır sorusundan hareket etmişlerdir. GDM ve GDKM tahmin edicilerinin güvenilirliğinde model bileşenlerinin etkisini incelemişler; GDM tahmin edicisinin güvenilirliğinin örneklem boyutuna, açıklayıcı değişkenlere ve bağ fonksiyonunun seçimine bağlı olduğu sonucuna ulaşmışlardır. Ayrıca örneklem boyutu arttıkça tam kredibiliteye yaklaşılmaktadır. 2011 yılında Zhou [30], 2009 yılındaki Tam Kredibilite-GDM-GDKM çalışmasının devamı olan doktora tezinde GDM-GDKM için hem Tam Kredibilite hem de Kısmi Kredibilite yaklaşımını ele alan, ayrıca hasar rezervlerinde GDM kullanımını inceleyen ve sigorta uygulamalarını veren daha kapsamlı bir çalışma yapmıştır. Klinker [31] Bühlmann-Straub Kredibilite yaklaşımı ile GDKM'i birlikte ele almıştır.

Bu tez çalışmasında belli risk faktörlerine göre sınıflandırmış verilerin GDM ile modellenmesi durumunda Sınırlı Dalgalanmalı Kredibilite yaklaşımı ele alınmıştır. GDM'lerin güvenilirliğini araştırmak için Tam Kredibilite Standardı'nın kullanımı detaylı olarak incelenmiş, belirli karşılaştırma kriterleri belirlenmiştir. Özel bir sigorta şirketinden alınan hasar sayısı verisi adil prim hesabında önemli bir nokta olan sınıflandırma işleminden sonra analiz edilmiş ve elde edilen sonuçlar yorumlanmıştır.

Bu tez çalışması şu şekilde oluşturulmuştur:

İkinci Bölümde doğrusal modeller ve varsayımları, ÜDA ve özellikleri, GDM bileşenleri ve çeşitleri, model parametre tahmininde kullanılan yaklaşımlar, model uyum iyiliği testleri ve risk sınıflandırması temel olarak anlatılmıştır.

Üçüncü Bölümde Kredibilite Kuramı ve yaklaşımları, özellikle Sınırlı Dalgalanmalı Kredibilite yaklaşımlarından birisi olan Tam Kredibilite Yaklaşımı altında Tam Kredibilite Standardı üzerinde durulmuştur.

Dördüncü Bölümde GDM ve Kredibilite Kuramı beraber ele alınmış, Tam Kredibilite Standardı kullanılarak GDM'lerin güvenilirliğini test etmek için belli karşılaştırma kriterleri oluşturulmuştur.

Çalışmanın Beşinci Bölümde özel bir sigorta şirketinden alınan bir veri ile uygulama yapılmıştır.

Son bölüm olan 6. Bölümde sonuçlara ve önerilere yer verilmiştir.

2. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

2.1. Doğrusal Modeller

Aktüeryal çalışmalarda varlık ve yükümlülüğe ilişkin finansal dengenin sağlanmasında sigortacının üstlendiği riske ilişkin belirsizliğin belirgin hale getirilmesi gerekmektedir. Belirsizliği en aza indirmek açısından modelleme önemlidir. Bu amaçla önceki dönemlere ait veri toplanır, analiz edilir, model oluşturulur, çeşitli yöntemler ile parametreler tahmin edilir, duyarlılık analizleri ile modelin parametreleri incelenerek optimal model belirlenmeye çalışılır [7]. Değişkenler arasındaki ilişkileri incelemek için birçok model vardır. Doğrusal modeller bunlardan birisidir. Doğrusal modellerdeki doğrusallık varsayımı, parametreler arasındaki ilişkilerin doğrusal olmasından kaynaklanmaktadır. $\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p$, $\beta_0+\beta_1x_1+\beta_2x_1^2$ veya $\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1x_2$ modellerinde parametreler arasında doğrusal ilişki olduğundan bu modeller doğrusal modellere örnek olarak verilebilir. Açıklayıcı değişkenler polinomiyal ya da çarpımsal olabilir [9]. Bu modellerdeki amaç gözlemlenen bir değişkendeki değişimi bir veya daha çok gözlemlenen değişken yardımıyla açıklamaktır. Bu değişimi bir değişken yardımıyla açıklayan modeller basit doğrusal model, birden çok değişken ile açıklayan modeller ise çoklu doğrusal modellerdir. Y yanıt değişkenine farklı kaynaklarda bağımlı değişken veya sonuç değişkeni; x açıklayıcı değişkenine ise eş değişken, bağımsız değişken, risk faktörü, tahmin edici, regresör veya dışsal (exogenous) değişken denilmektedir. Genel olarak bir doğrusal model,

$$E\left(y|x\right)=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p \qquad (2.1)$$

şeklindedir. Eşitlik 2.1 farklı bir şeklide aşağıdaki gibi de ifade edilebilir.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p+\varepsilon\,, \qquad E(\varepsilon)=0$$

Bu modelin matris gösterimi $\boldsymbol{Y=X\beta+\varepsilon}$ şeklindedir. Burada $\boldsymbol{Y}=\{y_1,y_2,...,y_n\}'$ şeklinde gözlem vektörü, $\boldsymbol{X}$ nxp boyutlu açıklayıcı değişken matrisi, $\boldsymbol{\beta}=\{\beta_0,\beta_1,\beta_2,...,\beta_p\}'$ parametre vektörü ve $\boldsymbol{\varepsilon}=\{\varepsilon_1,\varepsilon_2,...,\varepsilon_n\}'$ hata vektörüdür. $\boldsymbol{\beta}=\{\beta_0,\beta_1,\beta_2,...,\beta_p\}'$ parametrelerinin bulunmasına ilişkin 1980'li yıllarda yapılan ilk yaklaşım, hata terimlerinin kareler toplamı olan $\varepsilon'\varepsilon=\sum_i\varepsilon_i^2$ ifadesini minimum yapan β değerlerini bu parametrelerin tahmini olarak kabul etmek olmuştur. Daha sonraları hata teriminin ortalaması ve varyansına dair varsayımlar yapılarak farklı yaklaşımlar önerilmiştir [4]. Model varsayımları $\varepsilon = Y - X\beta$ hata terimi kullanılarak yapılır. Klasik doğrusal bir modelin aşağıdaki özellikleri sağladığı varsayılır.

1) **Sabit Varyanslılık (Homoskedastic):** $Var(\varepsilon) = \sigma_\varepsilon^2$, σ^2 sonlu ve sabit bir değerdir.
2) **Normal Dağılımlılık (Normal):** $\varepsilon \sim N(0, \sigma_\varepsilon^2)$
3) **İlişkisizlik (Uncorrelated):** Her bir y gözlemi birbirinden bağımsızdır. ε hata terimleri de bağımsız ya da en azından ilişkisizdir [9].

Doğrusal model varsayımları genelikle hata terimi üzerinden verilmektedir. Ancak bazı kaynaklarda bağımlı değişken ile ilişkilendirilen varsayımlar da bulunmaktadır. Bu karmaşıklığı gidermek adına doğrusal model çeşitlerinden biri olan çoklu doğrusal regresyon modeli için hem hata terimine hem de bağımlı değişkene göre yapılan varsayımlar Çizelge 2.1'de belirtilmiştir [32].

Çizelge 2.1. Çoklu Regresyon Modellerinin Gözlemlere ve Hata Terimine Göre Varsayımları

Gözlemlere göre Varsayımlar	**Hata Terimine göre Varsayımlar**
$E(Y_i \mid X_i) = \beta_0 + \beta_0 x_{i1} + ... + \beta_k x_{ik}$	$Y_i = \beta_0 + \beta_0 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i$
$Y_i \mid X_i \sim N(X_i\beta, \sigma^2)$	$\varepsilon_i \sim N(0, \sigma^2)$
$(x_{i1}, ..., x_{ik})$ değişkenleri stokastik değil.	$(x_{i1}, ..., x_{ik})$ değişkenleri stokastik değil.
$E(Y_i \mid X_i) = X_i\beta$ $Var(Y_i \mid X_i) = \sigma^2$	$E(\varepsilon_i) = 0$ ve $Var(\varepsilon_i) = \sigma_\varepsilon^2$
Y_i R.D.'leri bağımsız.	ε_i 'ler bağımsız.
Y_i R.D.'leri normal dağılımlı.	ε_i 'ler normal dağılımlı.

Bu varsayımlara göre yapılan parametre tahmin yöntemleri şöyledir;

1) **En Çok Olabilirlik Tahmini:** $L(\beta)$ olabilirlik fonksiyonunu maksimum yapan $\hat{\beta}$ değeri bulunur. Bir EÇO tahmin edicisi değişmezlik (invariance), asimptotik yansızlık (asymtotically unbiased), tutarlılık (consistent) ve minimum varyanslılık (minimum variance) özelliklerini sağlamalıdır.

2) En küçük Kareler Tahmini: $\sum_{i=1}^{p}(y_i - \hat{y}_i)^2$ değerini minimum yapan $\hat{\beta} = (X'X)^{-1}X'Y$ değeri bulunur. EKK yöntemine göre $\hat{\beta}$ tahmin edicisi $\hat{\beta} \sim N\{\beta, \sigma^2(X'X)^{-1}\}$ varsayımını sağlar [9]. Bu varsayıma göre $\hat{\beta}$ tahmin edicisi yansızdır ($E(\hat{\beta}) = \beta$) ve bu yansızlık $E(\varepsilon_i) = 0$ durumunda sağlanır. EÇO tahmin değeri ile EKK tahmin değeri birbiriyle örtüşür.

3) Ağırlıklandırılmış En Küçük Kareler Tahmini: Gözlemler az sayıda olduğu zaman w_i ağırlık faktörü modele dâhil edilerek tahmin yapılır. w_i ağırlık değeri genellikle riske maruz kalan birim sayısı yani exposure olarak alınır. $Var(\varepsilon_i) = \frac{\sigma^2}{w_i}$ olduğundan doğrusal model varsayımlarından birincisi olan hata teriminin varyansının σ^2 olması için tüm model ağırlığın karekökü ile çarpılır. **W**, köşegen elemanları w_i ağırlıkları olan köşegen matris olmak üzere, ağırlıklandırılmış model ve parametre tahminleri aşağıdaki gibidir.

$$\sqrt{w_i}y_i = \sqrt{w_i}\beta_0 + \sqrt{w_i}\beta_1 x_{i1} + \sqrt{w_i}\beta_2 x_{i2} + ... + \sqrt{w_i}\beta_p x_{ip} + \sqrt{w_i}\varepsilon_i$$

$$\hat{\beta}_{ağırlıklandırılmış} = (X'WX)^{-1}X'WY$$

$$\hat{\beta}_{ağırlıklandırılmış} \sim N\{\beta, \sigma^2(X'WX)^{-1}\}$$

Ağırlıklandırılmış tahminler özellikle sınıflanmış verilerin modellenmesinde tercih edilmektedir. EKK tahmin edicilerinin yansızlık ve hata teriminin varyansıyla ilgili sağladığı özellikleri AEKK tahmin edicileri de sağlar.

Doğrusal modellerden biri olan regresyon modelleri, bağımlı ve bağımsız değişkenler arasındaki ilişkileri ifade etmek için kullanılır. Regresyon modellerinin kullanılması için ön koşul, modelin hata teriminin normal dağılıma uyması ve sabit varyansa sahip olmasıdır. Bu koşulları sağlayamayan veri kümesinde dönüşümler yapılarak veri modele uygun hale getirilmeye çalışılır. Normal dağılıma uyum için verilerin logaritmasının alınması gibi doğrusal dönüşümler uygulanabilir. Sabit varyanslılık için EKK veya AEKK yöntemi kullanılabilir. Bu dönüşümler yerine kullanılabilecek alternatif bir yol ise GDM’ler olacaktır.

Sigorta hasar sayısı verisi gibi aktüeryal veri kümeleri genellikle simetrik olamayan sağa doğru çarpık dağılımlıdır. Uç değerlerin (kuyruk) bulunduğu bu tür veri kümelerinde modelleme yapılırken simetrik bir dağılım olan normal dağılım varsayımı sağlanmayabilir. Ayrıca hasar sayısı verisi kesikli dağıldığından hata terimleri normal dağılıma uymazlar. Bu nedenle klasik doğrusal modellerin kullanılması doğru olmaz. Bu tür verilerin analizinde hata teriminin normal dağılımlı veri olması varsayımı yerine, yanıt değişkenin ÜDA'den herhangi bir dağılıma uyması durumunda, GDM'ler kullanılabilir.

GDM'ler klasik doğrusal modellerin genelleştirilmiş halidir [7]. GDM açıklayıcı değişkenlerin doğrusal bir fonksiyonu olarak dönüştürülmüş ortalamayı modeller [2]. Bağımlı değişkeni yanıt değişken, bağımsız değişkenleri ise açıklayıcı değişkenler olarak ele alan GDM'ler için yanıt değişkenin ÜDA'den (binom, Poisson, normal, gamma, ters Gauss, negatif binom) gelmesi yeterlidir. Bu noktada ÜDA'ne ait dağılımlardan, varsayımlardan ve özelliklerinden bahsetmek faydalı olacaktır.

2.2. Üstel Dağılım Ailesi

Hayat dışı sigortaların fiyatlandırılmasında hasar sayısı ve hasar tutarının dağılımları önemlidir. Hasar sayıları için genellikle, binom, Poisson, negatif binom gibi kesikli dağılımlar kullanılır. Hasar tutarı için ise normal, ki-kare, gamma, ters-Gauss, dirichlet, log-normal gibi sürekli dağılımlar tercih edilir. Bu dağılımlardan binom, Poisson, negatif binom, normal, gamma, ters-Gauss dağılımları ÜDA'dan gelen dağılımlar olarak ortak bir çatı altında toplanabilir. ÜDA, GDM'lerde anahtar noktadır. ÜDA için olasılık yoğunluk fonksiyonu;

$$f(y) = c(y,\phi)\exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\} \tag{2.2}$$

biçimindedir. θ kanonik parametresi (veya doğal parametre) ve ϕ yayılım parametresi (veya ölçek parametresi)'dir. θ kanonik parametresi $y_1, y_2, ..., y_n$ gözlemlerine bağlıdır, ϕ yayılım parametresi ise her gözlem için sabittir [9]. Genel olarak Eşitlik 2.2 ÜDA'deki tüm dağılımları temsil etmektedir ancak ϕ, θ, $a(\theta)$ ve $c(y,\phi)$ değerleri, kullanılan dağılıma göre seçilerek özelleştirilebilir. ÜDA'deki her bir dağılım için θ kanonik parametresinin, ϕ yayılım parametresinin ve $a(\theta)$ alacağı değerler Çizelge 2.2.'de gösterilmiştir [9] .Çizelge 2.2'de beklenen değer ve varyans ifadeleri verilen θ parametresine göre hesaplanmıştır. Örneğin Poisson dağılımı için $\theta = \ln \mu$ olduğunda beklenen değer ve varyans μ 'ye bağlıdır.

Çizelge 2.2. Üstel Dağılım Ailesi ve Parametreleri

Dağılım Türü	θ	$a(\theta)$	ϕ	$E(y)$	$V(\mu) = \frac{\text{var}(y)}{\phi}$
Binom (n,p)	$\ln \frac{p}{1-p}$	$n\ln(1+e^{\theta})$	1	np	$np(1-p)$
Poisson(μ)	$\ln \mu$	e^{θ}	1	μ	μ
Normal(μ , σ^2)	μ	$\frac{1}{2}\theta^2$	σ^2	μ	1
Gamma(α , β)	$-\frac{1}{\alpha}$	$-\ln(-\theta)$	$\frac{1}{\beta}$	α	α^2
Inv. Gau.(α , β)	$-\frac{1}{2\alpha^2}$	$-\sqrt{-2\theta}$	β	α	α^3
Neg. Bin.(n,p)	$\ln \frac{np}{1+np}$	$-\frac{1}{p}\ln(1-pe^{\theta})$	1	n	$n(1+np)$

ÜDA için genel ortalama ve varyans;

$$\mu = E(y) = \frac{\partial a(\theta)}{\partial \theta} \tag{2.3}$$

$$Var(y) = \phi \frac{\partial^2 a(\theta)}{\partial \theta^2} \tag{2.4}$$

eşitlikleri ile verilebilir.

Varyans Fonksiyonu: $V(.)$ fonksiyonu ile gösterilen varyans fonksiyonu, ortalama ile varyans arasındaki ilişkiyi gösterir. Varyans fonksiyonları iki açıdan önemlidir. Birincisi direkt gözlemlerin ortalaması ile ilişkili olmasıdır. Diğeri ise, parametre tahmininde kullanılacak yöntemin belirlenmesinde kullanılmasıdır. $V(\mu)$ ÜDA'ndeki herhangi bir dağılım için özelleştirilemediği zaman ve aşırı yayılım durumunda parametre tahmininde log-olabilirlik fonksiyonu yerine, yarı-olabilirlik fonksiyonu tercih edilir. Varyans fonksiyonu,

$$\frac{\partial^2 a(\theta)}{\partial \theta^2} = \frac{\partial \left(\frac{\partial a(\theta)}{\partial \theta} \right)}{\partial \theta} = \frac{\partial \mu}{\partial \theta} = V(\mu)$$

şeklinde ifade edilir [9]. Eşitlik 2.3 ile belirtilen yanıt değişkenin varyansı, varyans fonksiyonu yardımıyla aşağıdaki gibi ifade edilir.

$$Var(y) = \phi V(\mu)$$

Üstel Dağılım Ailesi için Parametre Tahmini

ÜDA'nde θ ve ϕ parametrelerinin tahmininde momentler yöntemi ve EÇO tahmin yöntemleri kullanılmaktadır. İyi bir EÇO tahmin edicisinin sağlaması gereken özellikler şöyle sıralanabilir;

1) **Değişmezlik**: Bir parametrenin fonksiyonunun (monoton fonksiyon) tahmin edicisi o parametrenin tahmin edicinin fonksiyonu olur. g monoton bir fonksiyon olduğunda θ parametresinin EÇO tahmin edicisi $\hat{\theta}$ îse, $g(\theta)$'nın EÇO tahmin edicisi de $g(\hat{\theta})$ olur.
2) **Asimptotik Yansızlık:** Örneklem büyüklüğü arttıkça tahmin edicinin beklenen değerinin kendisine yaklaşması beklenir. $E(\hat{\theta}) = \theta$ olduğunda asimptotik yansızlık sağlanmış olur.
3) **Tutarlılık:** Örneklem büyüklüğü arttıkça tahmin edicinin dağılımı parametrenin dağılımına yaklaşır.
4) **Küçük Varyanslılık:** Varyans değerleri ile parametrelerin güvenilirlikleri test edilebililir. Küçük varyans durumunda tahmin edicilerin güvenilirliği artar. Örneklem büyüklüğü arttıkça daha küçük varyans elde etme olasılığı artar.

Bu özelliklerin çoğunda örneklem büyüklüğünün fazla olması veya sonsuza gitmesi koşulu aranır. Bu durumda özellikler asimptotik özellik olarak adlandırılır.

Aktüerya uygulamalarında genellikle tek-parametreli ÜDA kullanılmaktadır [28].

2.3. Genelleştirilmiş Doğrusal Modeller

Başlangıç noktası 1972 yılına dayanan GDM ilerleyen zaman içerisinde aktüerya alanında da yaygın olarak kullanılmaya başlamıştır [2]. Haberman ve Renshaw [7] doğrusal regresyonu, varyans analiz modellerini, olumsallık tablolarının analizindeki log-linear modelleri ve ikili verilerdeki logit modelleri GDM kümesine dahil eder. McCullagh ve Nelder [4] ise GDM'leri; klasik regresyon modelleri, varyans modelleri, var-yok yanıt değişkenleri için logit ve probit modeller, log-linear modeller, yaşam ve hasar verileri için kullanılan çarpımsal yanıt değişken modelleri olarak sınıflandırmışlardır.

GDM'lerdeki genelleştirme yanıt değişkenin dağılımı ve bağ fonksiyonu ile iki şekilde yapılır. Yanıt değişkenin dağılımı ÜDA'den herhangi birini kapsayacak şekilde genelleştirilir [4]. Bağ fonksiyonunun etkisini anlamak için klasik doğrusal model ile GDM arasındaki farka bakmak yeterli olacaktır. Klasik doğrusal modelde y yanıt değişkenin ortalaması E(y) ile x açıklayıcı değişkenleri arasında ilişki kurulur [9]. GDM'de ise y yanıt değişkenin ortalaması E(y)'nin fonksiyonu g{E(y)} ile x açıklayıcı değişkenleri arasında ilişki kurulur.

Sigorta verisiyle çalışılırken klasik doğrusal modellere göre GDM'lerin kullanımının birçok önemli nedeni vardır [29]. Bunlar şu şekilde sıralanabilir:

- Doğrusal modellerde sabit varyanslılık, normallik ve ilişkisizlik varsayımları vardır [4]. Bu varsayımlar her zaman sigorta verilerinde sağlanamayabilir.
- Sigorta verileri genellikle sağa doğru çarpık dağılımlıdır ve simetrik bir dağılım olan normal dağılıma uyum göstermeleri zordur.
- Sigortadaki hasar sayıları kesikli dağılıma sahiptir ve sürekli bir dağılım olan normal dağılıma uyum gösteremezler, ancak Merkesi Limit Teoremi'nden yararlanılarak normal dağılıma uyum sağlanabilir.
- GDM'ler ortalama yerine, ortalamanın bir dönüşümünü veya fonksiyonunun kullanılmasına imkân tanımaktadırlar. Bunun nedeni, yanıt değişkenin ortalamasının bir dönüşümünün açıklayıcı değişkenler ile doğrusal ilişkili olmasıdır [9].

Doğrusal modeller ile GDM'ler arasındaki farklılıklar Çizelge 2.3'te verilmiştir.

Çizelge 2.3. Doğrusal Modeller ile GDM'lerin Karşılaştırılması

	Doğrusal Modeller	**GDM**
Yanıt Değişkenin veya Hata Teriminin Dağılımı	Hata terimi normal dağılımlıdır.	Yanıt değişken üstel dağılım ailesinden herhangi bir dağılıma sahiptir.
Varyans Durumu	Sabit varyanslılık vardır. (Homoskedastic)	Değişen varyanslılık vardır. (Heteroskedastic)
Ortalamanın Dönüşümü	Kullanılmaz.	Kullanılabilir.
Bağ Fonksiyonu	Birim	Birim, Logaritmik, Üstel, Karekök, Logit, Ters

GDM'in aktüerya bilimlerinde kullanım alanı geniştir. Kredibilite kuramında, ölümlülüğün modellenmesinde, hasar rezervi hesaplanmasında, hayat-dışı ürün fiyatlandırması, kredi riskinin modellenmesinde GDM kullanılır [7]. Anderson et. al [8] risk kabul süreci (underwriting) ve fiyatlandırmada GDM kullanmışlardır.

GDM'ler; bir bağ fonksiyonu yardımıyla doğrusal tahmin ediciye bağlı olarak kitle ortalaması hesaplayan klasik doğrusal modellerin genelleştirilmiş halidir [29]. GDM varsayımları aşağıdaki gibidir;

1) Modellemede kullanılacak örneklem, bir parametreli ÜDA'nden seçilir.
2) Bağ fonksiyonu türevlenebilen ve monoton bir fonksiyondur.
3) Yanıt değişkenin ortalamasının bir dönüşümü, açıklayıcı değişkenler ile doğrusal olarak ilişkilidir [7].

GDM'lerde doğrusal tahmin edici $\eta_i = X_i\beta$ ve g bağ fonksiyonu $g(\mu_i) = \eta_i$ biçimindedir. GDM'ler model bileşenleri yardımıyla ifade edilirler.

Genelleştirilmiş Doğrusal Modellerin Bileşenleri

1) Üstel Dağılım Ailesinden Y Yanıt Değişkeninin Olasılık Yoğunluk Fonksiyonu

Kesim 2.2'de ÜDA'den herhangi bir dağılım için olasılık yoğunluk fonksiyonu; θ kanonik parametre, ϕ yayılım parametresi ve c bir sabit olmak üzere $f(y) = c(y,\phi)\exp\left\{\frac{y\theta - a(\theta)}{\phi}\right\}$ şeklindedir. Bu tanım beklenen değer $E(y) = \mu = \mu(\theta)$ ve verilen bir V varyans fonsiyonu yardımıyla varyans $Var(y) = \phi V(\mu)$'ı içerecek şekilde aşağıdaki gibi ifade edilebilir [29].

$$f(y;\theta,\phi) = \exp\left\{\int \frac{[y - \mu(\theta)]}{\phi V(\mu)} d\mu(\theta) + c(y,\phi)\right\} \qquad (2.5)$$

2) Doğrusal Bileşen

i. gözlem için $\boldsymbol{\beta_i} = (\beta_{i1},...,\beta_{ip})'$ ve $\boldsymbol{X_i} = (x_{i1},...,x_{ip})'$ olmak üzere $\eta_i = \sum_{j=1}^{p} x_{ij}\beta_{ij} = X_i'\beta_i, \quad i = 1,...,n$ şeklindedir. Doğrusal tahmin edici olarak da adlandırılan doğrusal bileşen, açıklayıcı değişkenlerin doğrusal bir kombinasyonudur ve bağ fonksiyonu aracılığıyla ortalama ile ilişkilendirilir.

3) Bağ Fonksiyonu

Yanıt değişkenin beklenen değeri ile doğrusal tahmin edici arasında bağ kuran g bağ fonksiyonu $g(\mu_i) = \eta_i$, $i = 1,...,n$ şeklindedir. Doğrusal Modeller ile GDM'ler arasında bir farklılıkta bağ fonksiyonu ile yapılabilir. Standart Doğrusal Modeller'de bağ fonksiyonu birim bağ fonksiyonudur. Ortalamanın dönüşümünü kullanmaya imkân tanımadığı için doğrudan ortalama alınır. GDM'ler ile bağ fonksiyonunun seçimi genişletilmiş olup, tek koşul monoton ve türevlenebilir bir fonksiyon olmasıdır. Bağ fonsiyonları birim, logaritmik, üstel, kakekök ve logit olmak üzere çeşitlere ayrılır [9].

3.1) Birim (Identity) Bağ Fonksiyonu

$g(\mu_i) = \mu_i$

Normal dağılımlı yanıt değişken için kullanımı uygundur. Bazı durumlarda Poisson dağılımlı yanıt değişken için de kullanılır.

3.2) Logaritmik (Log) Bağ Fonksiyonu

$$g(\mu_i) = \ln \mu_i$$

Poisson dağılımlı yanıt değişken için kullanımı uygundur.

3.3) Üstel (Power) Bağ Fonksiyonu

$$g(\mu_i) = \mu_i^{\,p}$$

p değeri negatif olduğunda bu bağ fonksiyonu Ters (Reciprocal) Bağ fonkisyonu olarak adlandırılır. Gamma dağılımlı yanıt değişken için $p = -1$ kabul edilir ve bağ fonksiyonu $g(\mu_i) = \frac{1}{\mu_i}$ olur. Ters-Gauss dağılımlı yanıt değişken için $p = -2$ kabul edilir ve bağ fonksiyonu $g(\mu_i) = \frac{1}{\mu_i^{\,2}}$ olur.

3.4) Karekök (Square Root) Bağ Fonksiyonu

$$g(\mu_i) = \sqrt{\mu_i}$$

Çok yaygın olmamakla birlikte Poisson dağılımlı yanıt değişken için kullanımı uygundur.

3.5) Logit Bağ Fonksiyonu

$$g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i}$$

Binom dağılımlı yanıt değişken için kullanımı uygundur.

$g(\mu_i) = \theta$ ise g fonksiyonu kanonik bağ fonksiyonu olarak adlandırılır ve $a(\theta)$ ile gösterilir. Nelder ve Verall [28] çalışmalarında kanonik bağ fonksiyonu kullanmışlardır.

Kredibilite kuramında yaygın olarak kullanılan GDM'lerdeki dağılımlar normal, gamma, Poisson ve binom dağılımlarıdır. Çizelge 2.4'de bu modellerin bileşenleri gösterilmiştir. Bu modeller dışında ters-Gauss ve negatif binom dağılımının da GDM bileşenleri vardır.

Çizelge 2.4. Kredibilite Kuramında Yaygın Olarak Kullanılan GDM'ler ve Bileşenleri [30]

Y	$Normal(\mu,\sigma^2)$	$Gamma(\alpha,\beta)$	$Poisson(\mu)$	$Binom(n, p)$
ϕ	σ^2	α^{-1}	1	1
$V(\mu)$	1	θ^{-2}	$e^{\theta}=\mu$	$np(1-p)$
$E(y)=\mu=\mu(\theta)$	$\theta=\mu$	$-\theta^{-1}=\frac{\alpha}{\beta}$	$e^{\theta}=\mu$	np
$Var(y)=\phi V(\mu)$	σ^2	$\frac{1}{\theta^2\alpha}=\frac{\alpha}{\beta^2}$	$e^{\theta}=\mu$	$np(1-p)$
Bağ Fonk. Türü	Birim	Ters	Logaritmik	Logit

2.3.1. Modellemenin Önemi ve Modelleme Basamakları

Modelleme yapılmasındaki amaç, mevcut veriden çıkarımlar yaparak geleceğe dair öngörü yapabilmektir. Veri kümesine uygun modelleme yapısını seçmek en önemli noktalardan biridir. Bu nedenle ilk yapılacak işlem verinin doğru olarak incelenerek ne tür bir veri olduğu ve ne tür bir modellemenin uygun olduğuna karar vermek olacaktır.

GDM'in sigorta uygulamalarında Poisson, gamma, bileşik Poisson (Tweedie dağılımları) yoğun olarak kullanılan dağılımlardır [33]. Sektördeki GDM uygulamalarında da hasar sayısı için Poisson dağılımı, hasar tutarının modellenmesinde ise, Gamma dağılımı tercih edilir.

Model tanımı yanıt değişken ile açıklayıcı değişkenler arasındaki ilişkiyi ve yanıt değişkenin olasılık yoğunluk fonksiyonunu ifade eder. Daha sonra modelin parametre tahminleri yapılır. Model yeterliliği test edilir ve en son güven aralıkları-hipotez testleri gibi istatistiksel işlemlerle çıkarımlar yapılır [34]. Bu modelleme basamakları adım adım şu şekilde uygulanır:

1) Yanıt değişken için uygun bir parametrik dağılım (Poisson, binom, normal…) seçilir. Seçilen $f(y)$ dağılımına göre $a(\theta)$ fonksiyonu da Çizelge 2.2'de belirtildiği gibi bulunur.

2) $g(\mu)$ bağ fonksiyonu yanıt değişkenin dağılımına göre seçilir.
3) x_i açıklayıcı değişkenleri seçilir. Bunlar kesikli, sürekli veya kategorik olabilir.
4) $x_1, x_2, ..., x_n$ açıklayıcı değişkenlerine karşılık gelen $y_1, y_2, ..., y_n$ gözlemleri bulunur.
5) Parametre tahminleri yapılır.
6) Uyum iyiliği testleri ile model test edilir [9].

2.3.2. Model Parametrelerinin Tahmini

Parametre tahminlerinde En Çok Olabilirlik (EÇO) yöntemi sıklıkla kullanılmaktadır. β katsayısı EÇO eşitliği yardımıyla Newton-Raphson gibi tekrarlı hesaplamalar ile bulunur. Log-olabilirlik fonksiyonları model parametrelerinin tahmininin yapılması açısından önemlidirler. Normal, Poisson, binom ve gamma dağılımları için log-olabilirlik fonksiyonları ayrı ayrı gösterilebilir. Bağ fonksiyonu her bir dağılım için farklılık gösterir. Örneğin normal dağılım için birim bağ fonsiyonu, Poisson dağılımı için logaritmik bağ fonsiyonu, binom dağılımı için logit bağ fonsiyon ve gamma dağılımı için ters bağ fonsiyonu kullanılır.

1) Normal Dağılım için Log-Olabilirlik Fonksiyonu

$$L(\beta) = \frac{\mu_i y_i - \frac{1}{2}\mu_i^2}{\sigma^2} - \frac{y_i^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$$

2) Poisson Dağılımı için Log-Olabilirlik Fonksiyonu

$$L(\beta) = y_i \log \mu_i - \mu_i - \log y_i!$$

Binom dağılımı ve gamma dağılımı için için log-olabilirlik fonksiyonlarının kapalı formları oldukça karışıktır [7], [28]. EÇO tahmin yönteminin kullanılabilmesi için dağılım bilgisi veya varsayımı gerekir. Bu bilgiye sahip olunmadığı durumlarda EKK yöntemi veya yarı-EÇO yöntemi kullanılır. EÇO tahmin edicileri hesaplanırken iteratif yöntemlerden yararlanılır. Newton- Raphson yöntemi de bu yöntemlerden birisidir.

Newton-Raphson Yöntemi

EÇO yöntemiyle β parametresini bulmak için log-olabilirlik fonksiyonunun bilinmeyen parametreye göre türevi her zaman 0'a eşitlenerek kolayca çözülemez. Bu durumda çözüme ulaşmak için iteratif yöntemler gerekir. Burada iteratif yöntemin kullanılmasının nedeni; (m+1). adımdaki değere ulaşmak için m. adımdaki değerin bilinmesinin gerekliliğidir. β 'nın bir fonksiyonu olan log-olabilirlik fonksiyonu olan $L(\beta)$ yardımıyla aşağıdaki eşitliğe ulaşılır [9].

$$\beta^{(m+1)} = \beta^{(m)} - \frac{L'(\beta^{(m)})}{L''(\beta^{(m)})} = \beta^{(m)} - \frac{\dfrac{\partial L}{\partial \beta_j}}{\dfrac{\partial^2 L}{\partial \beta_j \partial \beta_k}}$$

Burada β_j tahmin edilmek istenen parametre, $\frac{\partial L}{\partial \beta_j}$ Skor vektörü ve ikinci türevleri içeren $\frac{\partial^2 L}{\partial \beta_j \partial \beta_k}$ matrisi Hessian matristir. Newton-Raphson yönteminde β_j parametresine ulaşmak için Skor ve Hessian matrisinin tekrarlı eşitlikler ile hesaplanması gerekmektedir.

Fisher Skoru

Fisher, Newton-Raphson yöntemindeki Hessian matrisini, beklenen değeri ile yer değiştirerek yeni bir yöntem sunmuştur. $-E\left(\frac{\partial^2 L}{\partial \beta_j \partial \beta_k}\right)$ matrisi Fisher Bilgi Matrisi olarak adlandırılır. Fisher Skoru parametrenin kaç tekrarlı adımda tahmin edildiğinin bilgisini verir.

Genelleştirilmiş Doğrusal Modellerin En Çok Olabilirlik Tahmin Edicilerinin Asimptotik Özellikleri

ÜDA'nde EÇO tahmin edicilerinin özellikleri değişmezlik, asimptotik yansızlık, tutarlılık ve küçük varyanslılıktır. Bu özellikler büyük örneklem varsayımı altında $\hat{\beta}$ EÇO tahmin edicisi için özelleştirilebilir.

1) $\hat{\beta}$ EÇO tahmin edicisi β parametresinin asimptotik-yansız tahmin edicisidir ($E\left(\hat{\beta}\right) = \beta, n \to \infty$). Küçük örneklemler için EÇO tahmin edicileri genellikle yanlı olur.
2) $\hat{\beta}$ EÇO tahmin edicisi β parametresinin tutarlı tahmin edicisidir.
3) $\hat{\beta}$ EÇO tahmin edicisinin varyansı $n \to \infty$ iken $V(\hat{\beta}) = X'WX = \sum$ 'dır.

   Bu varyans ifadesi **W** köşegen elemanları w_i ağırlıkları olan köşegen matris olmak üzere AEKK yöntemi ile açıklanabilir.

4) 1. ve 3. özellikler kullanılarak, $n \to \infty$ durumunda $\hat{\beta}$'nın β ortalama, $\sum$ varyans ile normal dağılımlı olduğu ($\hat{\beta} \to N(\beta, \sum)$) varsayılabilir [29].

GDM'lerde bu parametre tahmin yöntemleri dışında bir ω_i ağırlık faktörü ile tahmin yapan AEKK Tahmin Yöntemi ve modeldeki her bir parametre için önsel-sonsal dağılımlar yardımıyla tahmin yapan Bayesci Tahmin Yöntemi (Bayesian Estimation Method)'de mevcuttur.

Modellemesi yapılacak verinin türüne göre GDM'ler çeşitlilik gösterir. Sayma verileri, kategorik veriler ve sürekli veriler için GDM'ler ayrı ayrı Çizelge 2.5'teki gibi incelenebilir [9].

Çizelge 2.5. Veri Türüne göre GDM'ler

Sayma Verileri için Genelleştirilmiş Doğrusal Modeller	**Kategorik Veriler için Genelleştirilmiş Doğrusal Modeller**	**Sürekli Veriler için Genelleştirilmiş Doğrusal Modeller**	**Yaşam Verileri için Genelleştirilmiş Doğrusal Modeller**
Poisson Regresyon Modeli	Logistik Regresyon	Gamma Regresyon	
Negatif Binom Regresyon Modeli	Probit Modeller (İkili(Binary)Verilerde)	Ters Gauss Regresyon	
Genelleş. Poisson Regresyon Modeli		Tweedie Regresyon	
Aşırı-Yayılımlı(Over-Disperson) Poisson		Çoklu Regresyon	
Sıfır-Yığılmalı Modeller		Varyans Analizi-Covaryans Analizi	
Örnek: Hasar Frekansı-Sayma Verileri	Örnek: Hasar Olasılıkları	Örnek: Hasar Tutarı	Örnek: Yaşayan Kişi Sayısı

Schirmacher'in veri türüne göre model bileşenlerini de içerecek şekilde daha kapsamlı olarak yaptığı sınıflandırma Çizelge 2.6'da verilmiştir [35] .

Çizelge 2.6. Veri Türüne göre Model Bileşenleriyle GDM'ler

	Veri Tipi		
	Sayma Verisi	**Tutar**	**Olasılık**
Bağ Fonksiyonu	$\log(\mu)$	$\log(\mu)$	$\log it(\mu)$
Hata Terimi	Poisson	Gamma	Binomial
Varyans	μ	μ^2	$\mu(1-\mu)$
Ağırlık	1	Hasar Sayısı	1
Offset	$\log(\exp osure)$	0	0

Bu tez çalışmasında sayma verileri için GDM'ler detaylı olarak ele alınmıştır.

2.3.3. Sayma Verileri için Genelleştirilmiş Doğrusal Modeller

Sigorta verisi hem kesikli hem de sürekli dağılıma sahip veriden oluşur. Sayma verisi pozitif tamsayı değerleri alır. Belli bir zaman diliminde veya belli bir bölgede meydana gelen olay sayısı olarak ifade edilen raslantı değişkenleri sayma raslantı değişkeni olarak bilinir. Kesikli dağılıma sahip olan sayma verisinin analizi için özel GDM'ler vardır. Bu modellemeler yapılırken genellikle Binom, Poisson ve Negatif Binom gibi kesikli dağılımlar tercih edilir. Bu modelleri incelemeden önce sigorta sektöründe karşılaşılabilecek aktüeryal sayma verilerine birkaç örnek vermek faydalı olacaktır.

- **Ölüm Sayısı**
  Mortalite çalışmalarında ölüm sayısını yaş, cinsiyet ve yaşam biçimi gibi açıklayıcı değişkenler ile açıklamaya çalışan GDM'ler oluşturulabilir.
- **Sağlık Hasar Sayısı**
  Sağlık sigortalarında bireye veya bir gruba ait hasar sayısını yaş, cinsiyet ve meslek gibi açıklayıcı değişkenler ile açıklamaya çalışan GDM'ler oluşturulabilir.
- **Araç Hasar Sayısı**
  Araç sigortalarında hasar sayısını sigortalının yaşı, sigortalının cinsiyeti, araba rengi, motor gücü, sigortalının önceki hasar deneyimi, yaşanan bölge ve benzer birçok risk faktörü ile açıklamaya çalışan GDM'ler oluşturulabilir [9].

Bu örnekler genel olarak hayat dışı sigortaları içerecek şekilde arttırılabilir. Aktüeryal sayma verisinin analizlerinde kullanılan modeller daha çok logaritmik-doğrusal (log-linear) modellere dayandırılır. McCullagh ve Nelder [4] logaritmik-doğrusal modeller ile klasik doğrusal modelleri Çizelge 2.7'deki gibi karşılaştırmışlardır.

Çizelge 2.7. Klasik Doğrusal Modeller İle Logaritmik-Doğrusal Model Arasındaki Farklar [4]

	Klasik Doğrusal Model	**Logaritmik-Doğrusal Model**
Sistematik Etkiler	Toplamsal	Çarpımsal
Hata Dağılımı	Normal	Poisson

Sayma verisi için kullanılan GDM'ler temel olarak Poisson Regresyon Modeli, Negatif Binom Regresyon Modeli ve Genelleştirilmiş Poisson Regresyon Modelidir. Bu tez çalışmasında Poisson Regresyon Modeli ayrıntılı olarak incelenecek, diğer GDM'lere ise kısaca değinilecektir.

2.3.3.1. Poisson Regresyon Modeli

Sayma verisinin modellenmesinde Poisson dağılımı sıklıkla tercih edilmektedir. Zorunlu Trafik Sigortası ve Kasko gibi araç sigortalarında hasar tutarı yanında hasar sayısı söz konusudur. Hasar sayısı ve tutarı yaş, cinsiyet, motor gücü, kullanım tipi, bölge, araç marka ve modeli gibi kategorik değişkenler kullanılarak modellenirler. Hasar sayısının bu açıklayıcı değişkenleri içerecek şekilde modellenmesinde Poisson Regresyon modeli kullanılabilir.

Poisson dağılım varsayımı için;

- Sıklık tahmin edilmeye çalışılmalıdır.
- Sıklıklar poisson süreci ile verilmelidir (Ortalama=Varyans= μ).
- Normal yakınsamayı kullanabilmek için merkezi limit teoremi gereği yeterli sayıda hasar verisi olmalıdır [26].

Poisson Regresyon'da çoğunlukla logaritmik bağ fonksiyonu kullanılmakla beraber, birim veya karekök bağ fonksiyonları da kullanılabilir [36]. Ortalamanın pozitif değerli olması logaritmik bağ fonksiyonunun kullanımına olanak tanır. Bazı durumlarda ortalama değeri logaritma alma işlemine uygun olmayabilir. Bu durumda birim bağ fonksiyonu tercih edilir [37]. Bağ fonksiyonunun farklılaşması açıklayıcı değişkenlerin modeldeki etkisini de belirlemektedir. Çizelge 2.7'den de görüleceği üzere birim bağ fonksiyonlarının kullanıldığı modellerde açıklayıcı değişkenlerin etkisi toplamsal, logaritmik bağ fonksiyonlarının kullanıldığı modellerde ise açıklayıcı değişkenlerin etkisi çarpımsaldır.

Poisson Regresyon modeline göre i. birim için yanıt değişken y_i'ler birbirinden bağımsız ve μ_i ortalama ile Poisson dağılımına ($y_i \sim Poisson(\mu_i)$) sahiptir. GDM'lerde bağ fonksiyonu yardımıyla model $g(\mu_i) = x_i'\beta$ olarak oluşturuluyordu. Bağ fonksiyonu birim bağ fonksiyonu olduğunda $\mu_i = x_i'\beta$, logaritmik bağ fonksiyonu olduğunda $\ln \mu_i = x_i'\beta$ eşitlikleri elde edilir. Tek açıklayıcı değişkenli bir örnek üzerinden birim ve logaritmik bağ fonksiyonunun etkisi kolayca görülebilir. i. birim için tek açıklayıcı değişkenli açıklayıcı değişken vektörü $\boldsymbol{x_i} = (1, x_{i1})'$, parametre vektörü $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ ve bağ fonksiyonu $g(\mu_i) = \beta_0 + \beta_1 x_{i1}$ şeklinde oluşturulsun. Birim bağ fonksiyonu için model $\mu_i = \beta_0 + \beta_1 x_{i1}$, logaritmik bağ fonksiyonu için ise model $\mu_i = e^{\beta_0 + \beta_1 x_{i1}}$ olacaktır. x_{i1} açıklayıcı değişkeni 1 birim artırılarak her iki model yeniden oluşturulduğunda bağ fonksiyonlarının model üzerindeki etkileri açıkça görülecektir. Birim bağ fonksiyonu için model $\mu_i = \beta_0 + \beta_1(x_{i1} + 1) = (\beta_0 + \beta_1 x_{i1}) + \beta_1$ olacak ve etkinin toplamsal olduğu görülecektir. Logaritmik bağ fonksiyonu için model $\mu_i = e^{\beta_0 + \beta_1(x_{i1}+1)} = e^{(\beta_0 + \beta_1 x_{i1})} e^{\beta_1}$ olacak ve etkinin çarpımsal olduğu görülecektir. Tek bir açıklayıcı değişken yerine birçok açıklayıcı değişken ile de aynı sonucu görmek mümkündür [9].

Aktüeryal veri analizinde logaritmik bağ fonsiyonunun daha çok tercih edilmesi nedeniyle çalışmanın bundan sonraki kısmında Poisson Regreson modelinin logaritmik bağ fonksiyonu kullanılarak oluşturulduğunu varsayarak çalışmalar yapılacaktır.

Hasar sayısı veya bir risk grubundaki ölen kişi sayısı gibi sayma verilerinin modellenmesinde riske maruz kalan birim sayısı n için offset denilen bir dengeleme veya düzeltme faktörü eklenir [38]. Bunun sebebi gözlemler arasındaki farkı veya sınıf büyüklüğünü düzeltmektir. i. birim için ortalama μ_i ise, μ_i yerine $\frac{\mu_i}{n}$ gibi meydana gelme olasılığı alınarak offset oluşturulur. Logaritmik bağ fonksiyonu için offset $\ln n$'dir.

$$g\left(\frac{\mu_i}{n}\right) = x_i'\beta$$

$$g\left(\frac{\mu_i}{n}\right) = \ln\left(\frac{\mu_i}{n}\right) = x_i'\beta$$

$$\ln \mu_i = \ln n + x_i'\beta \tag{2.6}$$

Logaritmik bağ fonksiyonundaki offset terimi $\ln n$, katsayısı 1 olan yeni bir açıklayıcı değişken gibi düşünülebilir. Bu offset ile y_i gözleminin beklenen değeri μ_i, riske maruz kalan birim sayısı n ile doğru orantılı olarak ifade edilir.

$$\mu_i = ne^{x_i'\beta}$$

Bu şekilde bir düzeltme yapılmış olur [9]. Poisson Regresyon Modeli için daha doğru bir model tanımı Eşitlik 2.6 ile yapılabilir. i. birim için yanıt değişken y_i’ler birbirinden bağımsız ve μ_i ortalama ile Poisson dağılımına sahip, n gözlemli bir poisson regresyon modeli,

$$y_i \sim Poisson(\mu_i) \qquad \ln \mu_i = \ln n + x_i'\beta$$

biçimindedir. Poisson Regresyon modelinde parametre tahmini yapılırken EÇO yöntemi, Yarı-Olabilirlik yöntemi ve GEKK yöntemi kullanılmaktadır [37].

Sayma verilerinin GDM ile analizinde Poisson Regresyon dışında Negatif Binom Regresyon Modeli ve Genelleştirilmiş Poisson Regresyon Modeli de kullanılmaktadır [9], [37].

2.4. Genelleştirilmiş Doğrusal Karma Modeller

GDKM, GDM’nin rastgele etkilerle genişletilmiş halidir. GDM uygulanırken tüm açıklayıcı değişkenler sabit etki değişkenleri olarak düşünülürken, GDKM uygulanırken hem sabit hem de rastgele etkiler düşünülür. GDKM’ler panel verilerin ve kümelenmiş verilerin analizinde yoğun olarak kullanılır. Aktüerya verileri hem sabit hem de rastgele etkiler içerir. Bu sebepten dolayı GDKM ile modellenmesi uygun olur.

Nelder ve Verall [4] bir GDM’in rastgele etkilerin de dikkate alındığı bir regresyon kredibilite modeli olarak ele alınabileceğini göstermişledir. GDM rastgele etkilerle

GDKM'e dönüşür ve bu GDKM'de bir regresyon kredibilite modeli olarak ifade edilir. GDKM ile ilgili bu çalışmaları Antonio ve Beirlant [2] ve Ohlsson'un [39] çalışmaları takip etmiştir.

GDKM'ler de GDM gibi olasılık yoğunluk fonksiyonu, doğrusal bileşen ve bağ fonksiyonu gibi model bileşenlerinden oluşmaktadırlar. Karma modellerde rastgele etki değişkeni devreye sokulmakta, ikili ve sınıflandırılmış verilerin analizinde önem kazanmaktadır [29]. GDKM'de rastgele etkilerin kullanılması, β regresyon katsayılarının bir alt kümesi için gözlemler arasında heterojenlik olduğu düşüncesinden kaynaklanmaktadır [40]. Antonio ve Beirlant [2] GDKM'lerin Aktüerya Bilimleri alanında kullanımı üzerine bir çalışma yapmıştırlar.

GDM'de gözlemlerin birbirinden bağımsız olduğu varsayılmaktadır, ancak risk sınıfları birbiriyle bağımlıdır ve risk sınıflarıyla çalışırken GDKM'lerden yararlanılmalıdır. Bu şekildeki verilerde bağımlılık sorunu ile karşı karşıya kalınabilir. Oluşturulan sınıflar arasında bağımlılık durumunun ortadan kaldırılması için GDKM'lerde sınıflara özgü rastgele etki değişkeni belirlenmekte, bu rastgele etki değişkeninin bilindiği varsayımı altında sınıflar arasında bağımsızlık koşulu sağlanmaktadır. Herhangi bir veri dönüşümüne ihtiyaç duyulmaması ve üstel aileden farklı dağılımlar için kullanılabilir olması açısından GDKM de GDM gibi önemli bir alternatif yöntemdir [40].

Dannenburg et al. [21], kredibilite faktörleri cinsinden risk primi için tahminleri elde etmede kredibilite hesaplamaları üzerinde durmuş ve farklı gözlemler için ağırlıklandırma yapmıştır. Yapılan bu çalışmalar bilinmeyen varyans bileşenleri ve β'nın fonksiyonlarıdır. Karma modeller (mixed models) böyle fonksiyonların modellenmesinde faydalıdır. Çünkü;

- Rastgele etkiler tahminleri etkileyebilir.
- Bu modeller regresyon katsayısı (β) ve varyans bileşenlerini modele aynı anda dahil ederek daha kapsamlı modelleme imkânı sunar. [40].

Sigorta portföyünün birden fazla risk sınıfından oluşması durumunda, risk sınıflarının gerçekleşmiş hasar sayısına olan etkisinin incelenmesi amacıyla GDKM'ler kullanılabilir.

GDM'lerde parametreler, sabit etkilerken, GDKM'lerde sabit ve rastgele etkiledir. GDKM'lerde β ve u_i rastgele etki parametrelerinin tahmini için Sınırlandırılmış Sözde-Olabilirlik (Restricted Pseudo-Likelihood), Genelleştirilmiş Tahmin Eşitliği (Generalized Estimating Equation), Cezalı Yarı- Olabilirlik (Penalized Quasi-Likelihood(PQL)), Gauss-

Hermite Tümlevi (Gauss-Hermite Quadrature), Laplace Yakınsaması, EÇO yaklaşımları için Monte- Carlo Yöntemi, Bayesci Yöntem, doğrusallaştırma temelli Sözde-Olabilirlik (Pseudo-Likelihood) gibi birçok yöntem vardır [2], [29], [30], [41].

2.5. Model Seçimi

Modelleme yaparken parametre tahmini yapıldıktan sonra oluşan model için uyum iyiliği testleri yapılmalıdır. Model seçiminde ölçüt çok önemlidir. Örneğin sadece Y mi, Y'nin bir dönüşümü mü (örneğin logY) daha uygun olur sorusuna verilecek yanıt önemlidir [4].

GDM'lerde model sonuçlarının yorumlanmasında, en iyi modeli bulmak amacıyla bilgi kriterlerinden yararlanılabilir. Modeller karşılaştırılırken L olabilirlik değeri, p parametre sayısı, n örneklem büyüklüğü, S artık kareler toplamı ve σ^2 varyansı göstermek üzere kullanılacak kriterler,

- **Log-Olabilirlik Değeri**
- **AIC:** Akaike Bilgi Kriteri ($-2L+2p$) veya ($\frac{S}{\sigma^2}+2p$)
- **BIC:** Bayesci Bilgi Kriteri ($-2L+p\log n$) veya ($\frac{S}{\sigma^2}+p\log n$)

  $\frac{S}{\sigma^2}$ terimi, gözlemlenmiş verilerin modellenmesinde yanlılığı ölçen bir terimdir ve küçük olması beklenir.
- **AICC:** Akaike Bilgi Kriterinin Küçük Örneklem Yanlı Düzeltilmiş Hali ($-2L+2p^n/(n-p-1)-2L$)
- **CAIC:** Tutarlı Akaike Bilgi Kriteri ($-2L+p(\log n+1)$)
- **HQIC:** Hannan ve Quinn Bilgi Kriteri($-2L+2p\log n$)
- **Genelleştirilmiş Pearson Ki-Kare**
- **Sapma Değeri**
- **Artık Değerleri** : Pearson Artığı, Anscombe Artığı, Yanıt Artığı, Kısmi Artık

biçimindedir.

Model seçiminde bu değerlerden ve bilgi kriterlerinden yararlanılır [41], [4] **.** En iyi modellerler bu kriterlerden Log-Olabilirlik değeri dışında en küçüğünü sağlayanlardır. Log- Olabilirlilik Değeri ile karşılaştırma yaparken büyük Log-Olabilirlik değeri daha iyi modeli temsil eder.

Modele hangi değişkenlerin alınacağı, hangilerinin model dışında bırakılacağı önemlidir. Bütün açıklayıcı değişkenleri içeren modellerin en iyi modeller olduğu düşünülebilir ancak yanıt değişken üzerinde etkisi az olan değişkenlerin eklenmesi, model uyumunu azaltacaktır. Buradan bazı değişkenlerin beraber modele alınması durumunda birbirlerinin etkilerini azalttığı sonucuna ulaşılabilir.
Zhou [30], GDM'ler ile ilgili olarak farklı bir uyum iyiliği analizi çalışması yapmıştır. Çalışmada, GDM tahmin edicilerini, kısmi kredibilitenin ağırlıklı tahmin edicileri ile karşılaştırmış ve kredibilite tahmin edicilerinin belli standartlara göre GDM tahmin edicilerinden daha üstün olduğu sonucuna ulaşmıştır. Bu standartlar Yanıt Artığı, Pearson Artığı, Anscombe Artığı, Kısmi Artık, Sapma Artığı, Düzeltilmiş Sapma Artığı, Olabilirlik Artığı, Skor Artığı ve Çalışan Artık'tır.

GDM bireysel veri kullanılarak oluşturulabileceği gibi sınıflanmış veri kullanılarak da yapılabilir. Sigorta verileri genellikle sınıflanmış ya da sınıflamaya uygun verilerdir. Sınıflanmış veri kullanıldığı zaman her bir sınıftaki birey sayısı ağırlık faktörleri ile devreye sokulur. Logaritmik bağ fonksiyonunun kullanıldığı GDM'lerden biri olan Poisson Regresyon modellerinde ise bu ağırlıklar offset terimi ile modele dâhil edilirler.

2.6. Risk Sınıflandırması

Risk sınıflandırması (Risk Faktörlerinin Tanımı) yapmak heterojen risklerden oluşan bir portföyü benzer risk sınıflarına bölerek, adil prim üretimini sağlamak amacıyla önemlidir [40] . Bir trafik sigortası belli risk faktörlerine göre sınıflandırılabilir. Bunlardan yaş, cinsiyet, medeni hal, kaza geçmişi, yaşanan bölge, kullanım tarzı, hasarsızlık indirimi gibi faktörler sürücü özellikleri iken; motor hacmi (silindir hacmi), motor gücü, araç yaşı, araç bedeli, araç ağırlığı, yakıt tipi, araç rengi gibi faktörler araç ile ilgili olan fiyatlandırma faktörleridir.

Risk faktörleri ile önsel bir sınıflandırma yapılabilir ama tam homojen bir sınıflandırma elde edilemez. Daha sonra geçmiş hasar deneyim bilgisi kullanılarak daha güçlü bir sınıflandırma yapmak gerekebilir. Açıklayıcı değişkenler sınıflandırılırken belli ölçümler kullanılır. Veri analizinde dikkat edilmesi gereken bir nokta yapılan ölçümlerde kullanılan ölçüttür. Kullanılacak veri kategorik veya sürekli olabilir. Kategorik bir veri ile çalışılıyorsa belirlenen değişkenlerin nominal mi ordinal mi olduğuna karar vermek önemlidir.

Nominal Sınıflandırma: İkili (binary), iki terimli (binomial) değişkenler veya çoklu(multinominial) veri kümeleri için uygundur. Örneğin; kadın-erkek, yeşil-mavi-kırmızı, evet-hayır, biliniyor-bilinmiyor vb.

Sıralı (Ordinal) Sınıflandırma: Bir sıralamaya ya da derecelendirmeye göre yapılır. Örneğin; genç-orta yaşlı-yaşlı, kan basıncı değerleri $\leq 70, 71-90, 91-110, 111-130, \geq 130$ mm Hg, motor hacmi $\leq 1600, 1600-2000, \geq 2000$ vb.

Nominal ve ordinal sınıflandırmada kullanılan veriler kategorik veri kümeleri olarak adlandılır ve gözlemlerin sıklık veya sayı değerleri ile ilgilenilir [34].

Bu tez çalışmasında kategorik bir veri kümesi kullanıldığından, sınıflandırma yapılırken hem nominal hem de ordinal bir sınıflandırma yapılmıştır. Örneğin cinsiyet için nominal, motor gücü için ordinal bir sınıflandırma yapılmıştır. Bu çalışmada sınıflanmış veri kullanılmasının önemi şöyle açıklanabilir; sınıflar arası karşılaştırma ile bir güvenilirlik analizi yapılmıştır. Bölüm 5'te veri seti ile uygulama yapılırken, veri belli risk faktörlerine göre sınıflanmış olmasaydı 87904 gözlem için karşılaştırma gerekebilirdi. Veriler cinsiyet ve motor gücüne göre sınıflandırıldığı için güvenilirliği analiz edilecek ve karşılaştırılacak birim sayısı 6'ya düşmüştür.

3. KREDİBİLİTE KURAMI

Hayat dışı sigortalarda karşılaşılacak hasar sayısı ve tutarındaki belirsizlik gelecekle ilgili doğru tahminler yapılarak en aza indirilebilir. Kredibilite Kuramı, sigorta ürününe ilişkin önceki dönemlerin verilerinden yararlanarak, gelecek dönemlere ait beklenen hasar ve prim tahmini yapılmasında kullanılır [42].

Kredibilitenin diğer bir adı da deneyim değerlendirmesidir. Bu şekilde adlandırılması oldukça mantıklıdır, çünkü sigortalıların ne kadar deneyimli, diğer bir deyişle elde bulunan verinin ne kadar gerçeğe yakın olduğunun analizinde kullanılır. Kredibilite tahminleri genellikle gözlemler arasında doğrusallık olduğu varsayımı altında ağırlıklı ortalama yöntemi ile yapılır [33]. Kredibilite tahmini gözlemlerin doğrusal bir fonksiyonudur [26].

Tahminleri ağırlıklandıran temel kredibilite formülü;

Tahmin=Z*(Gözlem)+(1-Z)*Diğer Bilgiler,   0≤Z≤1

ya da matematiksel olarak

$$\hat{\mu}_{kredibilite(i)} = Z_i \overline{X}_{gözlem} + (1 - Z_i)\overline{X}_{diğer} \tag{3.1}$$

şeklinde yazılabilir. Eşitlik 3.1'de Z_i i. birim için kredibilite faktörüdür ve 0 ile 1 arasında bir değer alır. Ağırlıklı ortalaması alınan $\overline{X}_{gözlem}$ ve $\overline{X}_{diğer}$ ortalamaları farklı farklı ifade edilebilir. Bir bireyin bilgisi ile örneklem bilgisi ağırlıklandırılabilir, örneklem bilgisi ile teorik olarak dağıldığı varsayılan dağılımdan gelen bilgi ağırlıklandırılabilir, geçmişe dair veri ile güncel veri ağırlıklandırılabilir ya da özel bir grubun verisi ile tüm portföy ağırlıklandırılarak tahminler yapılabilir. Elde bulunan veri kümesine göre uygun bir kredibilite modeli seçilerek analizler yapılmalıdır.

Kredibilite Kuramı'nın temelinde normal dağılıma yakınsama veya diğer adı ile normalleştirme varsayımı bulunulur. Mahler ve Dean [26] Kredibilite Kuramı'nın çıkış noktası olan normal dağılıma yakınsama ve güven aralıkları ile ilişkili bir çalışmalar yapmışlardır. Normal dağılıma yakınsama, gözlemlerin ortalamadan ne kadar sapacağının tahmininde kullanılır. Ortalama ve standart sapma bilgileri yardımıyla güven aralığı tahminlerine benzeyen eşitlikler yazılır.

Kredibilite fiyatlandırma probleminde; belli bir sınıfın hasar deneyimi ile tüm risk sınıfının deneyimini birleştirerek, belli bir sınıf için risk priminin ne olacağına karar verilmesi ile ilgilenir. Bu yaklaşım belli bir risk sınıfının kendine has özellikleri olduğu gibi diğer risk sınıflarıyla da ortak özellikler taşıdığı için mantıklıdır [2].

3.1. Kredibilite Yaklaşımları

Kredibilite Kuramı farklı ölçütlere önem vermesine göre gruplandırılır. Bu gruplandırma farklı kaynaklarda farklı şekillerde yapılmıştır.

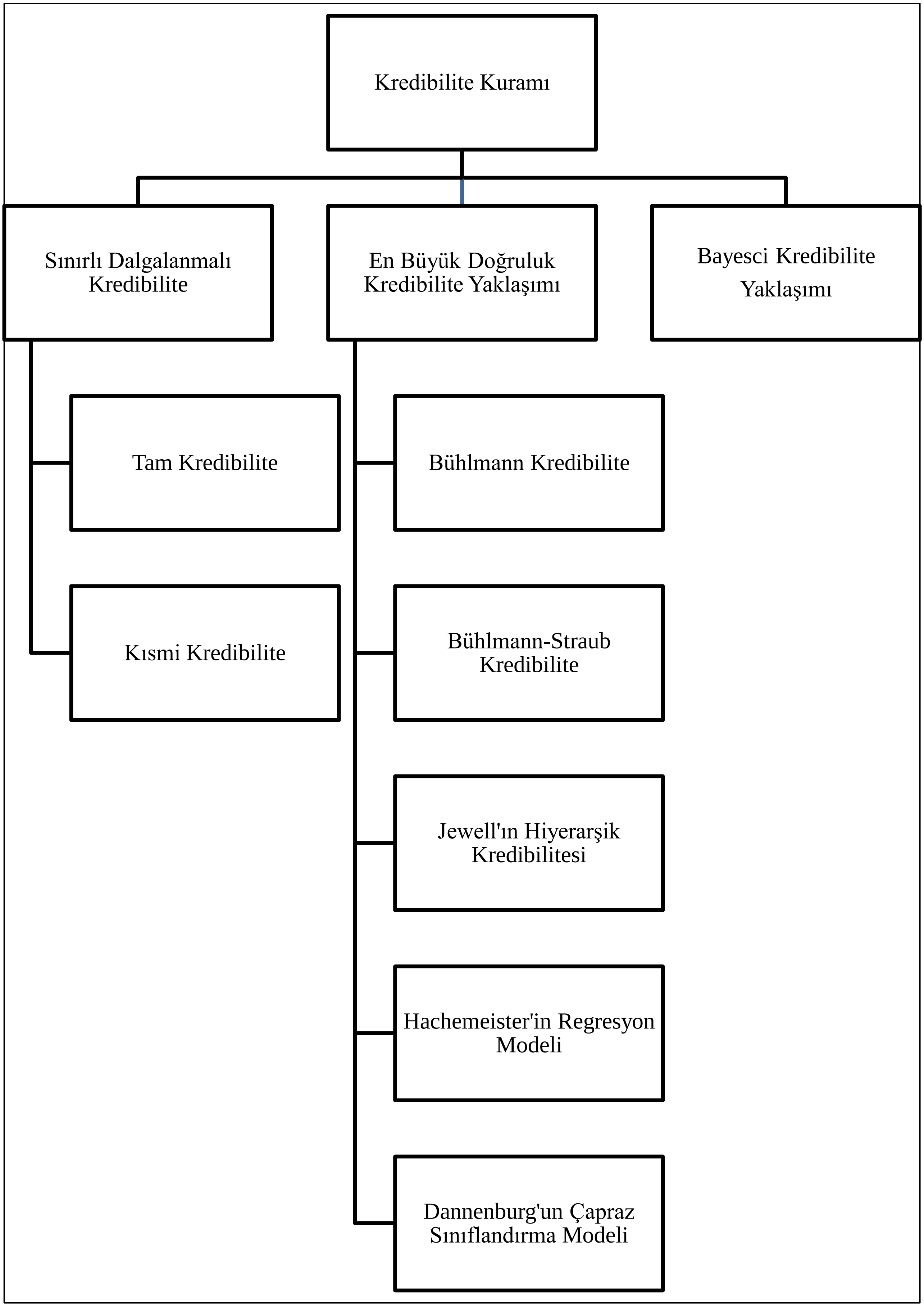
Goulet [13] kredibiliteyi, Sınırlı Dalgalanmalı Kredibilite (Amerikan Kredibilitesi) ve En Büyük Doğruluk Kredibilitesi (Avrupa Kredibilitesi) olmak üzere ikiye ayırmıştır. Daha sonra En Büyük Doğruluk Kredibilitesini Bayesci Kredibilite ve Bühlmann-Straub Kredibilite olarak alt dallara ayırmıştır. Hiyerarşik Kredibilite Modelini ve Çapraz Sınıflandırma Kredibilite Modelini Bühlmann-Straub Kredibilite altında incelenmiştir.

Dannenburg et al.[21] 1996 yılındaki çalışmalarında kredibiliteyi Sınırlı Dalgalanmalı Kredibilite, En Büyük Doğruluk Kredibilitesi, Bühlmann-Straub Model, Jewell'ın Hiyerarşik Modeli ve Hachemeister Model olmak üzere 5 alt bölümde incelemiştir.

Klugman et al. [42], Mahler ve Dean [26] ile Herzog [43]'a göre kredibilite yaklaşımları Sınırlı Dalgalanmalı Kredibilite (Tam ve Kısmi), Bayesci Kredibilite ve En Büyük Doğruluk Kredibilitesi (Bühlman ve Bühlmann-Straub) olmak üzere üçe ayrılır.

Kaas [22] kredibiliteyi Sınırlı Dalgalanmalı Kredibilite ve En Büyük Doğruluk Kredibilitesi **(**Bühlmann**,** Bühlmann-Straub ve Bayesci) olmak üzere ikiye ayırır.

Bu çalışmada ise Şekil 3.1'deki gibi Sınırlı Dalgalanmalı Kredibilite, En Büyük Doğruluk Kredibilitesi ve Bayesci Kredibilite şeklinde bir sınıflandırma kullanılmıştır. GDM ile Sınırlı Dalgalanmalı Kredibilite yaklaşımı beraber ele alınacağından Sınırlı Dalgalanmalı Kredibilite Yaklaşımı ayrıntılı bir şekilde incelenecektir.



Şekil 3.1. Kredibilite Yaklaşımları

3.1.1. Sınırlı Dalgalanmalı Kredibilite

Sınırlı Dalgalanmalı Kredibilite (Klasik Kredibilite-Amerikan Kredibilitesi) 20. yüzyılın başlarında iş kazası sigortalarında sigortalıların deneyimlerinden faydalanarak prim hesabı yapılmasıyla gelişmeye başlamıştır. Diğer kredibilite türlerine göre daha basit bir yapısı vardır. Goulet [13]'e göre Sınırlı Dalgalanmalı Kredibilite sadece sigortalıların deneyimlerine göre hesaplanır. n gözlem sayısı ve Z kredibilite faktörü Sınırlı Dalgalanmalı Kredibilite yaklaşımında önemli kavramlardır. Klugman et al. [42], Sınırlı Dalgalanmalı Kredibilite tanımı yaparken zamana göre ve gruplara göre iki farklı yaklaşım kullanılmıştır. Eşitlik 3.1 ile bulunan kredibilite priminde zamana göre hesap yaparken, $\bar{X}_{gözlem}$ değerini geçmiş dönemlere ait deneyim, $\bar{X}_{diğer}$ değerini tüm zamanları içeren portföy deneyimi olarak almışlardır. Gruplara göre hesap yaparken, $\bar{X}_{gözlem}$ değerini belli bir risk sınıfına ait deneyim, $\bar{X}_{diğer}$ değerini ise tüm sınıfları içeren portföy deneyimi olarak almışlardır. $\bar{X}_{gözlem}$ genellikle gözlenen verinin ortalamasıdır.

3.1.1.1. Tam Kredibilite

Bir sigortalının hasar deneyimi bir zaman diliminden diğerine çok fazla dalgalanma göstermiyorsa tam kredibilitenin sağlandığı söylenebilir ve Sınırlı Dalgalanmalı Kredibilite yaklaşımlarından biri olarak Tam Kredibilite'nin kelime anlamı buradan gelir [13].

Tam kredibilite kavramında %100 güven için gerekli olan gözlem sayısı ile ilgilenilir. Tam kredibilite yaklaşımına göre en az belirli bir sayıda gözlem varsa yapılan tahmin güvenilirdir. Tam kredibilite yaklaşımında normal dağılıma yakınsama kullanılmaktadır. Bilindiği üzere normal yakınsama gözlemlenmiş sonuçların ortalamadan ne kadar sapacağını tahmin etmede kullanılmaktadır. Tam kredibilite yaklaşımında kredibilite faktörü $Z_i = 1$'dir ve bu durumda kredibilite primi $\hat{\mu}_{kredibilite(i)} = \bar{X}_{gözlem}$ eşitliği ile sadece örneklem ortalamasına eşit olur. Hangi durumda bu eşitlik sağlanır sorusuna verilecek cevaba göre bir kriter belirlenebilir. Tahmin edilen değer ile gerçek değer arasındaki fark ne kadar az ise tahmin o kadar güvenilirdir. Bu güven ölçüsünü ifade etmek için $r > 0$ olacak şekilde bir hata-tahmin toleransı ve $0 < p < 1$ güven olasılığı değerleri ele alınsın. p güven olasılığı X_i gözlemlerinin ortalamasının $E(X_i)$ ortalamadan r kadar az ya da çok olma olasılığı olsun. Bu olasılığın istatistiksel gösterimi;

$$P\left(-rE(X_i) \le \bar{X}_i - E(X_i) \le rE(X_i)\right) \ge p \tag{3.2}$$

şeklindedir [42]. Bu standart $\bar{X} = \hat{\mu}_i$ ve $E(X_i) = \mu_i$ bilgisiyle i. zaman ya da i. risk sınıfı için özelleştirilirse;

$$\pi_i = \mathrm{P}\{|\hat{\mu}_i - \mu_i| \le r\mu_i\} \ge p \quad , \text{i=1,2,....,n} \tag{3.3}$$

eşitliği elde edilir. π_i tam kredibilite standardı ya da tam kredibilite olasılığı olarak adlandırılmaktadır. Bu tez çalışmasında ise i risk sınıflarının ifade edilmesinde kullanılmış ve risk sınıfları arasında kredibilite ilişkisi kurulmuştur.

Deneyim fiyatlandırmasında kullanılan örneklemin yeteri kadar büyük olması gerekir. Ancak bu durumda tam kredibilite söz konusu olur. Eğer yeterli sayıda gözlem varsa doğru risk sınıflandırması yapılarak adil prim hesaplanabilir [13]. Deneyim değerlendirmesi dikkate alınarak yapılan risk sınıflandırması ve kredibilite birbiri ile ilişkili iki kavramdır. Deneyim fiyatlandırması; iş kazası sigortaları, mali sorumluluk (zorunlu trafik) sigortası ve kasko sigortaları gibi hasar sıklığının yani deneyimin yüksek olduğu sigorta türlerinde uygulanır. Hayat sigortaları, konut sigortaları gibi düşük hasar sıklığına sahip sigorta türlerinde kullanılmaz.

Kredibilite primi hesaplanacak birim, $X_i, i = 1,...,n$ olarak belirlenir. X_i her bir periyotta gözlemlenen hasar sayısı, her bir periyotta gözlemlenen bütünleşik hasar tutarı ya da her bir hasar için hasar tutarı olabilir [26].

Tam kredibilite için gereken minimum gözlem sayısı n'in sayısal olarak hesaplanması için formüller geliştirilmiştir [13], [42], [26], [43]. Bunun için öncelikle güvenilirliği analiz edilecek büyüklüğe (hasar sayısı, bütünleşik hasar tutarı, hasar tutarı) göre $X_1, X_2, ..., X_n$ belirlenir. Bir risk grubu içinde aynı dağılıma sahip ve bağımsız $X_1, X_2, ..., X_n$ olmak üzere n adet gözlem olduğu durumda bu örnekleme ait ortalama ve varyans değerleri tam kredibilite için gereken minimum gözlem sayısı n'in hesabında kullanılmaktadır.

X_i i. risk sınıfına ait hasar deneyimi olmak üzere, $E(X_i) = \xi$ bu hasar deneyiminin beklenen değeri ve $Var(X_i) = \sigma^2$ ise varyansı olsun. n sınıftan oluşan bir örneklemde $X_1, X_2, ..., X_n$ gözlem için $\bar{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ olacaktır. X_i'lerin birbirinden bağımsız olduğu varsayımından hareketle örneklem ortalamasının beklenen değeri $E(\bar{X})=\xi$ ve varyansı $Var(\bar{X})=\dfrac{\sigma^2}{n}$ olur. Bu bilgiler ışığında Eşitlik 3.2 yeniden ele alınırsa,

$$P\left(-r\xi \le \bar{X}-\xi \le r\xi\right) \ge p$$

$$P\left(\left|\bar{X}-\xi\right| \le r\xi\right) \ge p \tag{3.4}$$

olur. Standart normal dağılıma yakınsama kullanılarak Eşitlik 3.4 standart sapmaya bölünürse,

$$P\left(\frac{\left|\bar{X}-\xi\right|}{\frac{\sigma}{\sqrt{n}}} \le \frac{r\xi}{\frac{\sigma}{\sqrt{n}}}\right) \ge p \tag{3.5}$$

elde edilir. Eşitlik 3.5 te en küçük $y=\dfrac{r\xi\sqrt{n}}{\sigma}$ değeri için olasılık y_p gibi özel bir olasılık ile ifade edilsin.

$$y_p = \inf_y \left\{ P\left(\frac{\left|\bar{X}-\xi\right|}{\frac{\sigma}{\sqrt{n}}} \le y\right) \ge p \right\}$$

$\bar{X}$ sürekli bir dağılıma sahip ise "$\ge$" işareti " =" işareti ile yer değiştirebilir [42].

$$P\left(\frac{\left|\bar{X}-\xi\right|}{\frac{\sigma}{\sqrt{n}}} \le y_p\right) = p \tag{3.6}$$

Eşitlik 3.5'te "$\le$" işaretinin sağındaki ifade $\dfrac{r\xi}{\frac{\sigma}{\sqrt{n}}}$ ile Eşitlik 3.6'da "$\le$" işaretinin sağındaki ifade y_p arasında $\dfrac{r\xi\sqrt{n}}{\sigma} \ge y_p$ ilişkisi kurulabilir. Bu ilişki i. risk grubu için $\sigma^2 = Var(X_i)$ ve $\xi = E(X_i)$ bilgileri ile genelleştirilerek, tam kredibilitesi incelenen büyüklüğe göre özelleştirilebilir ve tam kredibilite için gerekli minimum gözlem sayısı n için bir alt sınırı elde edilir [42], [20].

$$n \geq \left(\frac{y_p}{r}\right)^2 . \left(\frac{Var(X_i)}{E(X_i)}\right)^2 \tag{3.7}$$

y_p'nin sayısal olarak bulunması için Eşitlik 3.6'nın yeniden düzenlenmesi gerekir. $E\left(\bar{X}\right)=\xi$ ve $Var\left(\bar{X}\right)=\frac{\sigma^2}{n}$ olarak alındığında standartlaştırma sonucu,

$$p = P\left(\frac{\left|\bar{X}-\xi\right|}{\frac{\sigma}{\sqrt{n}}} \leq y_p\right) = P\left(\left|Z\right| \leq y_p\right)$$

$$p = P\left(-y_p \leq Z \leq y_p\right)$$

$$= \Phi\left(y_p\right) - \Phi\left(-y_p\right)$$

$$= \Phi\left(y_p\right) - 1 + \Phi\left(y_p\right)$$

$$= 2\Phi\left(y_p\right) - 1$$

$$\Phi\left(y_p\right) = \frac{(1+p)}{2}$$

elde edilir [42].

Örneğin, $p = 0,90$ ise $y_p = y_{0,9} = 1,645$ olur.

3.1.1.2. Kısmi Kredibilite

Sınırlı Dalgalanmalı Kısmi Kredibilite, gözlemlerdeki rastgele dalgalanmanın tahminlere etkisini sınırlandırmak diğer bir deyişle sadece gözlemler kullanılarak tahmin yapılmasını engellemek için kullanılır [26]. Kısmi kredibilite için genel tanım tahmin edilen birime ait gözlemlerin ortalaması ile kitle ortalaması da denilebilecek tüm portföye ait ortalamanın ağırlıklı olarak formüle edilmesidir [42]. Gözlemlerin yeteri kadar güvenilir olduğu düşünülmediğinden, kitleye veya tüm portföye ait ortalamanın etkisi, ağırlıklı ortalama ile tahmine yansıtılır. Sahip olunan veri setinin tam güvenilirlik için yeterli olmaması durumunda kısmi kredibilite söz konusu olur. Kısmi kredibilite de varyans tam kredibilite varyansına göre büyüktür. Bu durum dalgalanmaların nedeninin bir göstergesidir. Tam kredibilite yaklaşımında sadece var olan veriye veya gözlemlere göre tahmin yapılırken, kısmi kredibilitede ağırlıklandırılmış ortalama kullanılarak gözlem ve tüm kitle bilgisi hesaba katılır. Bu durumda da güvenilirliğin ölçüsü yani kredibilite parametresi üzerinde durulur, yani güvenilirliğin yüzde kaç olduğu ile ilgilenilir. Tam ve Kısmi Kredibilite arasındaki farklar Çizelge 3.1 ile verilmiştir.

Çizelge 3.1. Tam ve Kısmi Kredibilite Arasındaki Farklar

Tam Kredibilite	Kısmi Kredibilite
▪ Tam güven için gerekli gözlem sayısı n ile ilgilenilir.	▪ Tam güven için gerekli gözlem sayısı n'den daha az sayıda gözlem olduğundan Z ile ilgilenilir.
▪ n veya daha çok veri var	▪ n'den az sayıda gözlem var
▪ Z=1	▪ $0 \leq Z \leq 1$

Tam kredibilite yaklaşımında i. birim için kredibilite faktörü olan Z_i değeri 1'dir. Kısmi kredibilite yaklaşımında ise Z_i değeri 0 ile 1 arasında bir değer alır ve bu değerin tahmini en önemli noktadır. Bu faktörün sayısal olarak hesaplanması üzerine çeşitli hesaplama yöntemleri mevcuttur. Z_i'nin hesabı için Herzog [44], Mahler ve Dean [26], Bühlmann [44] ve Klugman [42] matematiksel formüller geliştirmişlerdir. Bu yöntemler teorik yaklaşımlardan çok sezgisel yaklaşımlardır. Bunlardan ilki $Z = \dfrac{n}{n+k}$ şeklindedir. Bu eşitlikte n gözlem sayısı, k ise karar verilmesi gereken bir sabittir. Bu kısmi kredibilite faktörü hesabı En Çok Doğruluk Kredibilite yakşamında, özellikle de Bühlmann Model'de tercih edilmektedir. Kesim 3.1.2.1'de k parametresinin hesabı için bir eşitlik verilmiştir. Sınırlı Dalgalanmalı Kredibilite yaklaşımında Z kredibilite faktörü için Klugman [42] tarafından, $Z=\min\left\{1,\frac{\xi}{\sigma}\sqrt{\frac{n}{\left(\frac{y_p}{r}\right)^2}}\right\}$ şeklinde parantez içindeki değerin 1'i aşması riskine karşı bir eşitlik yazılmıştır. Klugman [42] bu eşitliği elde ederken, ağırlıklandırılmış kredibilite eşitliğininden yaralanmıştır ve aşağıdaki eşitlikleri elde etmiştir.

$$Var(\hat{\mu}_{kredibilite})=Var\left(Z\overline{X}_{gözlem}+(1-Z)\overline{X}_{diğer}\right)$$

$$=Z^2\frac{\sigma^2}{n}+(1-Z)^2.0$$

$$=Z^2\frac{\sigma^2}{n}+(1-Z)^2.0$$

$$=Z^2\frac{\sigma^2}{n}=\frac{\xi^2}{\left(\frac{y_p}{r}\right)^2}$$

Kredibilite priminin varyansı $Var\left(\hat{\mu}_{kredibilite}\right)=\frac{\xi^2}{\left(\frac{y_p}{r}\right)^2}$ bilgisi ile $\overline{X}_{diğer}$'in standart normal dağılıma sahip olduğu varyasımından hareket ederek Z değerini çekmiştir.

Sınırlı Dalgalanmalı Kredibilite hasar deneyiminin yeterli olduğu durumlarda kullanılır ve kullanım alanları teorik açıdan sınırlıdır. Retrospective (geriye doğru) yöntemi kullanarak prim hesabı yapan şirketlerde toplam hasar tutarı biliniyorsa yıl sonunda sigortalının priminin yeniden belirlenmesinde kullanılabilir [13]. Kredibilite kullanılarak prim hesabı yapılacaksa En Büyük Doğruluk Kredibilite yaklaşımı daha doğru sonuçlar verebilir. Bu tez çalışmasında güvenilirlik analizi yapılacağı için Sınırlı Dalgalanmalı Kredibilite yaklaşımı modellerin karşılaştırılmasında kullanılmıştır.

3.1.1.3. Poisson Dağılımlı Veriler için Tam Kredibilite Yaklaşımı

Bu tez çalışmasında Poisson dağılımlı hasar sayısı verisi için uygulama yapılmıştır. Bu nedenle Poisson dağılan veri kümesinde tam kredibilite koşulunun sağlanması için sahip olunması gereken gözlem sayısı ve kısmi kredibilite durumunda Z kredibilite faktörü için eşitlikler yazılabilir. Poisson dağılımın en önemli özelliklerinden biri $E(X_i) = Var(X_i)$ eşitliğidir. Bu bilgi ile Eşitlik 3.7 deki $\frac{Var(X_i)}{E(X_i)}$ oranı 1 olur ve bu durumda tam kredibilite için gereken minimum gözlem sayısı için alt sınır aşağıdaki gibi olur [26].

$$n_{poisson} \geq \left(\frac{y_p}{r} \right)^2$$

Örneğin $p = 0,90$ ($y_p = y_{0,9} = 1,645$) ise ve r=0,01 seçilirse Poisson dağılımlı verinin tam güvenilir olabilmesi için bildirilmesi gereken hasar sayısı en az,

$$n_{poisson} \geq \left(\frac{y_p}{r} \right)^2 = \left(\frac{1,645}{0,01} \right)^2 = 27060,25 \approx 27060 \text{ olur.}$$

3.1.2. En Büyük Doğruluk Kredibilitesi

En Büyük Doğruluk Kredibilitesi (Modern Kredibilite) (Avrupa Kredibilitesi), Bühlmann tarafından 1960'lı yıllarda geliştirilmiş modele dayalı çözüm sunan bir kredibilite yaklaşımıdır. Kredibilite faktörü bulunurken bayesci modeldeki katsayılar ve varyans bileşenleri dikkate alınır [21]. Kredibilite ile prim hesabında En Büyük Doğruluk Kredibilite yaklaşımı yaygın olarak kullanılmaktadır [21], [22], [26], [42]). Dannenburg et al. [21] En Büyük Doğruluk Kredibilitesi için temel bir kaynaktır. Şentürk [27] tez çalışmasında panel veri analizi yöntemi ile En Büyük Doğruluk Kredibilitesi çalışarak il bazında trafik sigortalarının güvenilirliğini analiz etmiştir.

Bühlmann [15], [16] Kredibilite Kuramı'na daha istatistiksel olarak yaklaşmıştır. Bu istatistiksel çalışmalar Bühlmann ve Straub'un çalışması [17] ve Hachemeister veri seti olarak adlandırılan ve birçok çalışmada yararlanılan verinin tanıtıldığı Hachemeister'in çalışmasıyla [18] devam etmiştir.

3.1.2.1. Bühlmann Kredibilite

Bu yaklaşıma En Küçük Kareler Kredibilitesi de denmektedir. Bu kredibilite modelinde amaç; daha önce tahmin edilen tutar ile gerçek değeri arasındaki hatayı minimize etmektir. Bühlmann kredibilitede süreç varyansının beklenen değeri ve hipotetik ortalamalarının varyansı kullanılır. Sınırlı Dalgalanmalı Kredibilite yaklaşımı süreç varyansının beklenen değeri veya hipotetik ortalamasının varyansı bilinmediği durumlarda tercih edilebilir.

Bu kredibilite yaklaşımında deneyimlerin homojen olmasına önem verilir. Heterojen bir portföyde bireysel deneyimler daha da önemli hale gelir. Bühlmann kredibilitede gözlemlenen veri ve önsel bilgi ağırlıklandırılır. Gözlemlenen veri tek başına güvenilir olmadığından önsel bilgi kullanılarak daha güvenli veri elde edilir. Önsel ortalamanın rolü, tüm veri hakkındaki bilginin ölçütü olmasıdır [26]. Bühlmann kredibilite faktörü,

$$\hat{\mu}_{BühlmannKredibilite(i)} = Z_{Bühlmann}\overline{X}_{gözlem} + (1 - Z_{Bühlmann})\overline{X}_{diğer}$$

eşitliğinden bulunur. Bühlmann Kredibilite yaklaşımında kredibilite faktörü süreç varyanslarının beklenen değerinin hipotetik ortalamaların varyansına oranı ile bulunur.

$$Z = \frac{n}{n + k_{Bühlmann}}$$

$$k_{Bühlmann} = \frac{SüreçVaryansınınBeklenenDeğeri}{HipotetikOrtalmanınVaryansı}$$

formülü ile hesaplanır [43]. $k_{Bühlmann}$ Bühlmann-Kredibilite parametresidir ve model parametreleri kullanılarak elde edilir. n gözlem sayısı arttıkça Z değeri 1'e yaklaşır ancak hiçbir zaman 1 değerini almaz. Z=1 olma durumu sadece tam kredibilite için geçerlidir. Bühlmann Kredibilite önsel dağılımla ilişkilendirildiğinden bazı kaynaklarda Bayesci kredibilite ile birlikte ele alınır. Bu çalışma da Bayesci Kredibilite Bühlmann Kredibiliteden ayrı ele alınmıştır.

3.1.2.2. Bühlmann Straub Kredibilite

Bühlmann-Straub Kredibilite [17], Bühlmann Kredibilite [15]'nin bir genellemesidir. Bühlmann-Straub modelinin Bühlmann modelinden farkı varyansların heterojen olması nedeniyle ağırlık faktörlerinin kullanılmasıdır.

3.1.2.3. Jewell'ın Hiyerarşik Kredibilitesi

Hiyerarşik kredibilite 1995 yılında Jewell [19] tarafından kredibilite primi hesabında ana veriden ziyade yardımcı verinin modellenmesini yapmak amacıyla kullanılmaya başlanmıştır. Heterojen sigorta verilerinde primin adil bir şekilde dağılımında kullanılan bir yöntemdir. (Belhadj) [45]. Bühlmann ve Bühlmann-Straub model bir-aşamalı (one-level) hiyerarşik modellerdir. Hiyerarşik kredibilitede koşullu dağılımlar devreye girerek hiyerarşik bir yapı oluşturulur.

3.1.3. Bayesci Kredibilite

Bühlmann kredibiliteye benzerlik gösterir ve Herzog [43] kredibilite priminin Bayesci tahmini için Bühlmann Kredibiliteyi en iyi yakınsama olarak kabul eder. Bu kredibilite yaklaşımında optimal katsayılar bulunmaya çalışılır. Mahler ve Dean [26] Bayesci kredibiliteyi, Bühlmann ve En Küçük Kareler Kredibilitesi ile eş düşünmüşlerdir. Bayesci Kredibilite yaklaşımında binom-beta, poisson-gamma, normal-normal modeller söz konusudur. Bayesci kredibilitenin olumsuz yanı önsel dağılım hakkında bilgi sahibi olunması gerekliliğidir.

4. GENELLEŞTİRİLMİŞ DOĞRUSAL MODEL İÇİN SINIRLI DALGALANMALI KREDİBİLİTE KAVRAMI

Kredibilite Kuramı ile GDM'leri ilişkilendiren ilk çalışma Nelder ve Verrall [28] tarafından yapılmış ve prim hesaplamaları ile hasar rezervlerinin hesaplanmasında kullanılan modeller geliştirilmiştir. Antonie ve Beirlant [2] farklı kredibilite modellerine ilişkin GDKM'leri oluşturarak, GDKM'lerin aktüerya alanında kullanımı üzerine bir çalışma sunarken; 2008 yılında Ohlsson [1] çoklu faktörleri içeren hayat dışı fiyatlandırma problemlerinde GDM ile Kredibilite Kuramı'nı birleştirmiştir. 2011 yılında Klinker (2011) [31] Bühlmann-Straub kredibilite yaklaşımı ile GDKM'i birlikte ele almıştır. Bu çalışmalarda GDM'ler ve GDKM'ler kredibilite modellerine dönüştürülerek prim fiyatlandırmasında kullanılmaktadırlar. Bu çalışmalarda genellikle En Büyük Doğruluk Kredibilitesi ile GDKM birlikte ele alınmıştır.

GDM ile krediblliteyi birlikte ele alan diğer bir yaklaşım Sınırlı Dalgalanmalı Kredibilite Yaklaşımı ile GDM'leri birlikte ele alan çalışmalardır. Bunlardan ilki 2004 yılında sigorta fiyatlandırmasında bir tahmin edicisinin güvenilir olması için gereken hasar sayısıyla ilgili olarak Schmitter [25] tarafından yapılmıştır. Garrido ve Zhou [29] 2009 yılında yaptıkları çalışmada bir GDM tahmin edicisinin güvenilir olması için en az kaç tane gözleme ihtiyaç vardır sorusundan hareket etmişler ve model tahmin edicilerinin güvenilirliliğinde model bileşenlerinin etkisini incelemişlerdir. Zhou [30], 2009 yılındaki tam kredibilite-GDM-GDKM çalışmasına kısmi kredibilite ve hasar rezervlerinin GDM'ler ile analizini ekleyerek daha kapsamlı bir çalışma sunmuştur. Bu çalışmalarda ise Kredibilite Kuramı, bir GDM oluşturulurken kullanılan örneklem sayısına göre modelin güvenilirliğini test ederken kullanılmaktadır. Bu tez çalışmasında GDM ile Kredibilite Kuramı arasında benzer bir ilişki kurulmuştur.

Sınıflandırılmış verilerle çalışmak benzer risklere maruz kalan risk sınıfları açısından adil prim hesabında önemlidir. Bu nedenle çalışmanın bu bölümünde bireysel veri yerine belli risk faktörlerine göre sınıflandırılmış verilerin GDM ile modellenmesi durumunda Sınırlı Dalgalanmalı Kredibilite Yaklaşımı ele alınmıştır.

4.1. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğinin Tam Kredibilite Standardı ile İncelenmesi

Bir GDM tahmin edicisi tam kredibilite standardını sağlıyorsa, güvenilir bir tahmin edici olduğu söylenebilir. Tahmin edilen değer ile gerçek değer arasındaki fark ne kadar az ise tahmin o kadar güvenilirdir. Eşitlik 3.3 ile verilen π_i tam kredibilite standardı tam kredibilite olasılığı olarak da adlandırılmaktadır. Bu kredibilite olasılığı güven aralığı ifadesine benzemekle birlikte tam bir güven aralığı değildir. Burada bir risk sınıfı için kredibilite olasılığı, parametre tahmin edicisinin ortalamadan r kat az ya da çok olması olasılığını ifade eder. Ayrıca Poisson dağılımlı veride model parametresinin ortalama olması da tam kredibilite standardını güven aralığına benzetmektedir. Mahler ve Dean [26] çalışmalarında mevcut verinin güvenilirliğinin bir ölçüsü olarak örneklem bilgilerini kullanarak tam kredibilite standardını yazmışlardır. Gözlemlerin ortalamadan belli bir katsayı kadar az ya da çok olma olasılıkları bir güven ölçüsü olarak belirlenmiştir. Bu çalışma da ise Zhou [30]'nun 2011 yılı çalışmasındakine benzer şekilde tahmin edici için bir güven ölçütü belirlenmiştir. GDM ile tam kredibilite standardı arasında ilişki kurmak için Scmitter [25] 2004 yılında, Garrido ile Zhou [29] 2009 yılında ve Zhou [30] 2011 yılı çalışmalarında belli önermeler yapmışlardır. Bu önermelerden yararlanarak GDM'deki risk sınıflarının güvenilirliğini test edecek karşılaştırma kriterleri olan kredibilite olasılığı ve asimptotik varyans için eşitlikler yazılabilir.

i. Sınıf için Kredibilite Olasılığı π_i 'nin Bulunuşu

Eşitlik 4.1' deki mutlak değer ifadesi açılırsa tam kredibilite olasılığı,

$$\pi_i = \mathrm{P}\left(\left|\hat{\mu}_i - \mu_i\right| \le r\mu_i\right) = \mathrm{P}\left((1-r)\mu_i \le \hat{\mu}_i \le (1+r)\mu_i\right)$$

eşitliği ile yazılabilir. Monoton artan $g(\mu_i) = \eta_i$ bağ fonksiyonuna sahip bir GDM model için, her iki tarafın g fonksiyonu alındığında eşitsizlik yön değiştirmeden kalır ve tam kredibilite ile GDM'in birlikte ele alınması için ilk adım atılmış olur.

$$\pi_i = \mathrm{P}\left(g[(1-r)\mu_i] \le g(\hat{\mu}_i) \le g[(1+r)\mu_i]\right)$$

GDM tahmin edicisinin güven aralığına ulaşmak için eşitsizliğin her iki tarafından $g(\mu_i)$ çıkarılır. $g(\mu_i)=\eta_i=X_i'\beta$ olduğundan kredibilite olasılığı, g bağ fonsiyonu ve β parametresi cinsinden,

$$\pi_i = \mathrm{P}\left(g[(1-r)\mu_i]-X_i'\beta \le X_i'\hat{\beta}-X_i'\beta \le g[(1+r)\mu_i]-X_i'\beta\right) \tag{4.1}$$

olarak bulunur.

i. Sınıf için Asimptotik Varyans s_i^2 'nin Bulunuşu

$\hat{\beta}$ tahmin edicisinin Kesim 2.3.2'de değinilen asimptotik özelliklerinden ikincisi β tahmin edicisinin varyansının n sonsuza giderken Σ değerine yakınsamasıdır ($n\to\infty, V(\hat{\beta})\to\Sigma$). i. sınıf için asimptotik varyansı,

$$s_i^2 = V(\hat{\mu}_i)=V(X_i'\hat{\beta})=V(x_{i1}\hat{\beta}_1+\ldots+x_{ip}\hat{\beta}_p)$$

şeklinde bulunabilir. Tahmin edicinin varyansı ile ilgilenilidiğinden $\hat{\beta}$ tahmin edicileri birer raslantı değişkeni gibi ele alınabilirler. Bu durumda varyans fonksiyonunun içindeki açıklayıcı değişkenler sabit birer katsayı gibi düşünülerek varyans fonksiyonunun dışına kareleri alınacak şekilde çıkarılabilir.

$$s_i^2 = V(\hat{\mu}_i)=V(X_i'\hat{\beta})=V(x_{i1}\hat{\beta}_1+\ldots+x_{ip}\hat{\beta}_p)=\sum_{j=1}^{p}\sum_{k=1}^{p}x_{ij}x_{ik}Cov(\hat{\beta}_j,\hat{\beta}_k)$$

$Cov(\hat{\beta}_j,\hat{\beta}_k)=\sigma_{ij}$ ve $\sigma_{ij}=\Sigma$ bilgisinden hareketle,

$$s_i^2 = \sum_{j=1}^{p}\sum_{k=1}^{p}x_{ij}x_{ik}\sigma_{ij} \tag{4.2}$$

elde edilir. Eşitlik 4.2'nin matris gösterimi,

$$s_i^2 = X_i'\Sigma X_i \tag{4.3}$$

şeklindedir. Bu eşitlikler $\hat{\beta}$ tahmin edicisinin asimptotik özellikleri kullanılarak yazıldığından örneklemin asimptotik özelliklerin sağlanabileceği kadar büyük bir örneklem olduğu varsayılmalıdır. Bu durumda $n \to \infty$ iken i. risk grubu için asimptotik varyans $s_i^2 = V(\hat{\mu}_i) = V(X_i'\hat{\beta}) = X_i' \sum X_i$ olarak elde edilir.

$\hat{\beta}$ tahmin edicisinin üçüncü asimptotik özelliği olan $E(\hat{\beta}) = \beta$ ortalama ve $V(\hat{\beta}) = \sum$ varyansı ile normal dağılıma yakınsamasıdır ($\hat{\beta} \to N(\beta, \sum)$). Dolayısıyla Merkezi limit teoremi kullanılarak $\frac{X_i'\hat{\beta} - X_i'\beta}{s_i} \to N(0,1)$ sonucuna ulaşılır.

Genel olarak bağ fonsiyonu türü belirtmeden ve $X_i'\beta = g(\mu_i)$ bilindiğinden Eşitlik 4.1,

$$\pi_i = \mathrm{P}\left(g[(1-r)\mu_i] - g(\mu_i) \le X_i'\hat{\beta} - X_i'\beta \le g[(1+r)\mu_i] - g(\mu_i) \right)$$

şeklinde yazılabilir. Normal dağılıma yakınsama varsayımıyla tam kredibilite olasılığı,

$$\pi_i = \mathrm{P}\left\{ \frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i} \le \frac{X_i'\hat{\beta} - X_i'\beta}{s_i} \le \frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i} \right\} \tag{4.4}$$

eşitliği ile yazılabilir. Eşitlik 4.4 standart normal kümülatif dağılım fonksiyonları cinsinden;

$$\pi_i = \Phi\left(\frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i} \right) - \Phi\left(\frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i} \right) \tag{4.5}$$

şeklinde yazılır. Bu eşitliklerden de görüleceği üzere s_i küçüldükçe π_i büyümektedir. Dolayısıyla, standart sapması küçük olan ve π_i kredibilite olasılığı büyük olan risk grubu daha güvenilirdir sonucuna ulaşılabilir. Eşitlik 4.1 ile verilen i. sınıf için kredibilite olasılığı π_i ve Eşitlik 4.3 ile verilen i. sınıf varyansı s_i^2 tüm bağ fonsiyonları için kullanılabilir. Belli bir dağılım ve bağ fonkiyonu fonsiyonu şeçilidiğinde özelleşmiş formlar elde edilebilir. Bu özelleşmiş formlar;

- Poisson Dağılımlı Yanıt Değişken - Logaritmik Bağ Fonsiyonu
- Binom Dağılımlı Yanıt Değişken - Lojit Bağ Fonsiyonu
- Normal Dağılımlı Yanıt Değişken - Birim Bağ Fonksiyonu
- Gamma veya Ters Gauss Dağılımlı Yanıt Değişken -Üstel Bağ Fonksiyonu

şeklindedir [9].

Logaritmik Bağ Fonksiyonuna Sahip Poisson Dağılımlı Veriye İlişkin Genelleştirilmiş Doğrusal Modellerin Güvenilirliğinin İncelenmesi

Hasar sayılarının Poisson dağılımlı olduğu varsayıldığında, bağ fonsiyonu logaritmik'tir. Y_i hasar sayısını gösteren raslantı değişkeni olmak üzere , Y_i 'ler birbirinden bağımsız ve Poisson dağılımına sahip olsun. Logaritmik bağ fonsiyonu $g(\mu_i) = \ln \mu_i$ olduğundan,

$$g(\mu_i) = g(E(Y_i)) = \ln \mu_i = X_i{}'\beta = x_{i1}\beta_1 + \ldots + x_{ip}\beta_{ip} \quad \text{ve} \quad \mu_i = E(Y_i) = e^{x_{i1}\beta_1 + \ldots + x_{ip}\beta_{ip}}$$

olur. Eşitlik 4.2'de bağ fonksiyonu logaritmik olarak alınırsa, Poisson dağılımlı verilerin kredibilite olasılığı bulunur.

$$\pi_i = \mathrm{P}\left(\ln[(1-r)\mu_i] - X_i'\beta \le X_i'\hat{\beta} - X_i'\beta \le \ln[(1+r)\mu_i] - X_i'\beta\right)$$

$$\pi_i = \mathrm{P}\left(\ln(1-r) + \ln \mu_i - X_i'\beta \le X_i'\hat{\beta} - X_i'\beta \le \ln(1+r) + \ln \mu_i - X_i'\beta\right)$$

$$\pi_i = \mathrm{P}\left(\ln(1-r) \le X_i'\hat{\beta} - X_i'\beta \le \ln(1+r)\right) \qquad (4.6)$$

Eşitlik 4.6 logaritmik bağ fonksiyonuna sahip poisson dağılımlı verilerin kredibilite olasılığını vermektedir. $\dfrac{X_i'\hat{\beta} - X_i'\beta}{s_i} \to N(0,1)$ olduğundan Eşitlik 4.6 ile verilen kredibilite olasılığı,

$$\pi_i = \mathrm{P}\left(\frac{\ln(1-r)}{s_i} \le \frac{X_i'\hat{\beta} - X_i'\beta}{s_i} \le \frac{\ln(1+r)}{s_i}\right) \qquad (4.7)$$

biçiminde yazılabilir.

Eşitlik 4.7 standart normal dağılım fonksiyonları cinsinden,

$$\pi_i = \Phi\left(\frac{\ln(1+r)}{s_i}\right) - \Phi\left(\frac{\ln(1-r)}{s_i}\right) \qquad (4.8)$$

şeklinde yazılır. Eşitlik 4.8, Eşitlik 4.5'in logaritmik bağ fonksiyonu için özelleşmiş halidir.

Logaritmik bağ fonksiyonuna sahip Poisson dağılımlı verinin varyans ifadesine ulaşmak için Eşitlik 4.6 yeniden düzenlenebilir. Eşitlik 4.6'nin mutlak değeri alınırsa ve $0 < r < 1$ için $|\ln(1+r)| < |\ln(1-r)|$ bilgisinden yararlanılarak,

$$\pi_i = \mathrm{P}\left(|\ln(1-r)| \le \left|X_i'\hat{\beta} - X_i'\beta\right| \le |\ln(1+r)|\right)$$

$$\pi_i \le \mathrm{P}\left(\left|X_i'\hat{\beta} - X_i'\beta\right| \le |\ln(1-r)|\right)$$

olarak yazılabilir [29].

$s_i^2 = V(\hat{\mu}_i) = V(X_i'\hat{\beta})$ olduğundan eşitsizliğin her iki tarafı s standart sapmasına bölünürse standart normal dağılımlı raslantı değişkenine ulaşılır.

$$= \mathrm{P}\left\{\frac{X_i'\hat{\beta} - X_i'\beta}{s} \le \frac{|\ln(1-r)|}{s}\right\}$$

$$= \mathrm{P}\left\{Z_{\pi(tamgüvenilir)} \le \frac{|\ln(1-r)|}{s}\right\}$$

Olasılık ifadesinin içindeki varyans yalnız bırakılırsa logaritmik bağ fonksiyonuna sahip GDM'de oluşturulan risk sınıflarının tam güvenilirliğinin karşılaştırılmasında kullanılacak bir varyans ifadesi

$$s^2 \le \left(\frac{|\ln(1-r)|}{Z_{\pi_{tamgüvenilir}}}\right)^2 = s^2_{tamgüvenilir} \qquad (4.9)$$

olarak elde edilir [25], [29]. Sayısal değerlerle Eşitlik 4.9 daha iyi anlaşılabilir. r=0,01, $\pi(tamgüvenilir) = 0,90$ olarak seçildiğinde, $Z_{\pi(tamgüvenilir)} = Z_{0,90} = 1,645$ ve $s^2_{tamgüvenilir} = \left(\frac{|\ln(1-r)|}{Z_{\pi_{tamgüvenilir}}}\right)^2 = \left(\frac{|\ln(1-0,01)|}{1,645}\right)^2 = 0,000037328$ elde edilir. Herbir risk sınıfı için bulunan s_i^2 asimptotik varyansları $s^2_{tamgüvenilir} = 0,000037328$ varyansıyla, herbir risk sınıfı için bulunan π_i kredibilite olasılığı $\pi_{tamgüvenilir} = 0,90$ olasılığı ile karşılaştırılarak, i.risk sınıfının tam güvenilir olup olmadığına karar verilebilir.

GDM'in güvenilirliği hasar sayısına, poliçe sayısına (exposure), örneklem büyüklüğü, model bileşenlerinden açıklayıcı değişkenlerin ve bağ fonksiyonlarının seçimine bağlı olarak değişir. Bu etkileri incelerken her bir risk grubu için bulunan s_i^2 varyansları ve yine her bir risk grubu için bulunan kredibilite olasılıkları olan π_i değerleri karşılaştırma unsurları olarak kullanılır. Varyans karşılaştırması iki şekilde yapılabilir. Birinci karşılaştırmada s_i^2 varyansları tam kredibilite durumundaki $s_{tamgüvenilir}^2$ varyansı ile karşılaştırılarak o risk grubunun, tam kredibiliteyi sağlayıp sağlamadığına karar verilebilir. İkinci karşılaştırmada s_i^2 varyansları birbirleriyle karşılaştırılarak hangi risk grubunun daha güvenilir olduğu bulunabilir. Küçük s_i^2 değerine sahip olan risk grubu daha güvenilirdir.

Kredibilite olasılıklarının karşılaştırması da iki şekilde yapılabilir. Birinci karşılaştırmada kredibilite olasılıkları olan π_i değerleri tam kredibilite olasılığı $\pi_{tamgüvenilir}$ ile karşılaştırılabilir. Tam güvenilirlik sağlayamayan sınıflar arasındaki ikinci karşılaştırma da π_i varyansları birbirleriyle karşılaştırılarak hangi risk sınıfının daha güvenilir olduğu bulunabilir. Büyük π_i değerine sahip olan risk sınıfı daha güvenilirdir.

4.1.1. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğine Hasar Sayısının Etkisi

Hasar sayılarının etkisi oluşturulan risk sınıfları arasında test edilebilir. GDM'in bağ fonksiyonu ve tam kredibilite standardı yardımıyla bulunmuş s_i^2 değerleri kendi aralarında karşılaştırılarak hasar sayısının kredibiliteye etkisi analiz edilebilir. Küçük hasar sayısına sahip risk grubunda s_i^2 daha büyüktür.

4.1.2. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğine Poliçe Sayısının Etkisi

Hasar sayıları aynı, poliçe sayıları farkı olan risk sınıfları karşılaştırılarak riske maruz kalan birim sayısının kredibilite üzerindeki etkisi incelenebilir. Bu inceleme s_i^2 asimptotik varyansları ve π_i kredibilite oalsılıkları karşılaştırılarak yapılır. Daha fazla poliçeye sahip olan risk grubunda daha küçük s_i^2 ve daha büyük π_i değeri beklenir.

4.1.3. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğine Örneklem Büyüklüğünün Etkisi

Örneklem büyüklüğünün kredibilite üzerinde etkili bir unsur olduğu açıktır. Gözlem sayısı arttıkça daha güvenilir sonuçlar elde edilebilir. Modellemesi yapılacak veri hakkında ne kadar çok deneyim bilgisi var ise tam kredibiliteye o kadar yaklaşılacaktır. Sınırlı dalgalanmalı kredibilite kavramında özellikle de tam kredibilite de örneklem sayısı önemlidir. Mevcut gözlemler kullanılarak hesaplanmış kredibilite olasığı π_i ile tam kredibilite durumundaki $\pi_{tamgüvenilir}$ olasılığı karşılaştırılarak; π_i değerinin $\pi_{tamgüvenilir}$ değerine yaklaşması için poliçe sayısının ne olması gerektiği hesaplanabilir.

4.1.4. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğine Açıklayıcı Değişken Seçiminin Etkisi

Açıklayıcı değişkenlerin güvenilirliliğe etkisini görmek için farklı açıklayıcı değişkenlere göre hesaplanmış karşılaştırma kriterleri s_i^2 ve π_i karşılaştırılabilir. Örneğin cinsiyet ve motor hacmi açıklayıcı değişkenleri kullanılarak oluşturulmuş bir modelde cinsiyet açıklauyıcı değişkeninin etkisini görmek amacıyla aynı motor hacmi açıklayıcı değişkenine ve farklı cinsiyet açıklayıcı değişkenine sahip olan risk sınıfları karşılaştırılabilir.

Zhou [30] açıklayıcı değişkenlerin dağılımının GDM güvenilirliği üzerindeki etkisini görmek için toplam hasar sayısı değişmeyecek şekilde her bir risk grubundaki hasar sayısını değiştirerek daha önceki risk sınıflandırmasından farklı bir sınıflandırma elde etmiştir. Açıklayıcı değişkenlerin dağılımının etkisini önceki ve sonraki s_i^2 değerleri ile test etmiştir. Bu durumda bazı risk sınıfları tam kredibiliteden uzaklaşmış, bazıları ise yakınlaşmıştır

4.1.5. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğine Bağ Fonksiyonun Etkisi

Bir model birleşeni olarak bağ fonksiyonunun GDM tahmin edicilerinin güvenilirliğine etkisi iki şekilde incelenebilir. Bağ fonksiyonunun yapısı veya türü değiştirilebilir. Yapısı bir sabit ile çarpılarak, bölünerek veya bir sabit eklenerek değiştirilebilirken; türü logaritmik, birim, logit bağ fonksiyonları gibi fonksiyonlar yardımıyla değiştirilebilir. Garrido ve Zhou [29] çalışmalarında ele aldıkları bağ fonksiyonunun türünü değiştirmeyip yeniden ölçeklendirerek bağ fonsiyonunun etkisini incelemişler ve bağ fonksiyonu seçiminin GDM tahmin edicilerinin güvenilirliliğini etkilemediğini ileri sürmüşlerdir. Bağ fonksiyonu c gibi bir sabit ile çarpılarak $h(\mu_i) = cg(\mu_i)$ gibi yeni bir bağ fonsiyonu elde edilsin. $g(\mu_i) = \eta_i = X_i'\beta^{(g)}$ iken $h(\mu_i) = cg(\mu_i) = cX_i'\beta^{(g)} = X_i'\beta^{(h)}$ olur. Bu durumda iki farklı bağ fonksiyonuna göre oluşmuş modelde parametreler arasında $\beta^{(h)} = c\beta^{(g)}$ şeklinde bir ilişki doğar. Bu bağ fonsiyonlarına göre asimptotik varyans değerlerine bakılırsa,

$$s_i^{(g)} = \sqrt{V\left(X_i'\beta^{(g)}\right)}$$

$$s_i^{(h)} = \sqrt{V\left(cX_i'\beta^{(g)}\right)} = \sqrt{c^2 V\left(X_i'\beta^{(g)}\right)} = c\sqrt{V\left(X_i'\beta^{(g)}\right)} = cs_i^{(g)}$$

$$s_i^{(h)} = cs_i^{(g)} \tag{4.10}$$

eşitlikleri elde edilir. Eşitlik 4.11 bağ fonksiyonunun bir sabit ile çarpılarak yeniden ölçeklendirilmesinin karşılaştırma kriterlerinden biri olan s_i^2 değerini çarpılan sabit kadar etkilediği sonucunu gösterir. Bağ fonksiyonunun değişiminin diğer karşılaştırma kriteri olan π_i değerini nasıl etkilediğini görmek için, Eşitlik 4.5 farklı g ve h bağ fonksiyonlarına göre $\pi_i^{(g)}$ ve $\pi_i^{(h)}$ olarak elde edilebilir.

$$\pi_i^{(g)} = \Phi\left(\frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right) - \Phi\left(\frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right)$$

$$\pi_i^{(h)} = \Phi\left(\frac{h[(1+r)\mu_i] - h(\mu_i)}{s_i^{(h)}}\right) - \Phi\left(\frac{h[(1-r)\mu_i] - h(\mu_i)}{s_i^{(h)}}\right)$$

$$= \Phi\left(\frac{cg[(1+r)\mu_i] - cg(\mu_i)}{cs_i^{(g)}}\right) - \Phi\left(\frac{cg[(1-r)\mu_i] - cg(\mu_i)}{cs_i^{(g)}}\right)$$

$$= \Phi\left(\frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right) - \Phi\left(\frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right) = \pi_i^{(g)}$$

$$\pi_i^{(h)} = \pi_i^{(g)}$$

Kredibilite olasılıklarına göre bir kaşılaştırma yapılırsa, bağ fonksiyonundaki değişimin kredibiliteyi etkilemediği sonucuna ulaşılır [29].

Bağ fonksiyonun türü yanıt değişkenin dağılıma göre şekillenmektedir. Yanıt değişkenin dağılımının değişmesi, bağ fonksiyonunun türünü de etkileyeceğinden, güvenilirliğe de etkisinin olması beklenmektedir. Daha öncede değinildiği gibi logaritmik bağ (log-link) fonksiyonu genellikle Poisson dağılımlı hasarlarla eşleştirilir [4]. Logaritmik bağ fonksiyonu yerine başka bir bağ fonksiyonu tercih edilmek istenirse birim bağ fonsiyonu kullanılabilir. Poisson dağılımlı veri için logaritmik bağ fonksiyonuna ve birim bağ fonksiyonuna göre hesaplanmış π_i değerleri karşılaştırılabilir. Birim bağ fonksiyonu kullanılarak kredibilite olasılıklarını hesaplamak için Eşitlik 4.5 $g(\mu_i) = \mu_i$ yani birim bağ fonksiyonu için yeniden düzenlendiğinde,

$$\pi_i^{(birim)} = \Phi\left(\frac{(1+r)\mu_i - \mu_i}{s_i}\right) - \Phi\left(\frac{(1-r)\mu_i - \mu_i}{s_i}\right)$$

$$\pi_i^{(birim)} = \Phi\left(\frac{\mu_i + r\mu_i - \mu_i}{s_i}\right) - \Phi\left(\frac{\mu_i - r\mu_i - \mu_i}{s_i}\right)$$

$$\pi_i^{(birim)} = \Phi\left(\frac{r\mu_i}{s_i}\right) - \Phi\left(\frac{-r\mu_i}{s_i}\right)$$

elde edilir. Logaritmik bağ fonksiyonu kullanılarak bulunan kredibilite olasılığının $\pi_i^{(\log)} = \Phi\left(\frac{\ln(1+r)}{s_i}\right) - \Phi\left(\frac{\ln(1-r)}{s_i}\right)$ ile birim bağ fonksiyonu kullanılarak bulunan kredibilite olasılığının $\pi_i^{(birim)} = \Phi\left(\frac{r\mu_i}{s_i}\right) - \Phi\left(\frac{-r\mu_i}{s_i}\right)$ farklı sonuçlar doğuracağı açıktır. Risk sınıflarının kendi aralarında karşılaştırırken farklı bağ fonksiyonlarına göre hesaplanmış kredibilite olasılıkları birbirleriyle karşılaştırılabilir. Değerler farklılaşsa bile güvenilirlik sıralaması değişmeyebilir.

Bağ fonksiyonu yapısının ve türünün GDM'lerin güvenilirliğine etkisi incelendikten sonra farklı g_1 ve g_2 gibi iki bağ fonksiyonu arasında seçim için bir ölçüt belirlenebilir. $\hat{\pi}_i^{(g_1)} < \hat{\pi}_i^{(g_2)}$ ise g_2 bağ fonksiyonu kullanılarak bulunan tahmin edicinin, g_1 bağ fonksiyonu kullanılarak bulunan tahmin ediciye göre daha güvenilir olduğu söylenebilir [29].

4.1.6. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğinde Yanlılık Durumu

GDM'de tam kredibilite yaklaşımı tahmin edicilerin asimptotik özelliklerinin birincisi olan yansızlığın sağlandığı varsayımında yapılmıştır. Bağ fonksiyonunun seçimi GDM tahmin edicilerinin yanlılık durumunu da etkilemektedir [30]. Bağ fonksiyonu doğrusal iken yansız bir tahmin ediciye ulaşılırken, logaritmik bağ fonksiyonları gibi aşağı bükey (downward convex) ve artan bağ fonksiyonları yanlı tahmin edicilere neden olabilmektedir. Yanlılık durumunu analiz atmek için Jensen Eşitsizliği kullanılabilir [29]. Yan, tahmin değeri ile gerçek değer arasındaki farktır.

$$Yan(X) = E(\hat{X}) - E(X)$$

Jensen Eşitsizliği u aşağı bükey bir fonksiyon iken,

$$E\left[u\left(X\right)\right] \le u\left(E\left[X\right]\right)$$

şeklindedir. Jensen eşitsizliği g^{-1} ters-bağ fonsiyonuna sahip GDM için yazılırsa;

$$E\left[g^{-1}\left(X_i'\hat{\beta}\right)\right] \le g^{-1}\left(E\left[X_i'\hat{\beta}\right]\right)$$

$$E\left[\hat{\mu}_i\right] \le g^{-1}\left(X_i'\beta\right) = \mu_i$$

$$E\left[\hat{\mu}_i\right] \le \mu_i$$

$Yan(\mu_i) = E\left[\hat{\mu}_i\right] - E\left[\mu_i\right] = E\left[\hat{\mu}_i\right] - \mu_i$ olarak bulunur.

4.2. Sınıflandırılmış Verilerde Genelleştirilmiş Doğrusal Model'in Güvenilirliğinin Kısmi Kredibilite ile İncelenmesi

Örneklemin yeterli büyüklükte olmadığı durumlarda tam kredibiliteden bahsedilemez. Tam kredibilite yaklaşımında örneklemin yeteri derecede büyük olması normal dağılıma yakınsama ve merkezi limit teoreminin kullanımına olanak tanımaktadır. Ayrıca

$\pi_i = \mathrm{P}\left(\left|\hat{\mu}_i - \mu_i\right| \leq r\mu_i\right) \geq p$ ile oluşturulan tam kredibilite standardında $\hat{\mu}_i$'nin dağılımı bilinmeli ya da bir varsayım ile yakınsama yapılmalıdır. Örneklem yeteri kadar büyük iken dağılım varsayımı yapılabilmektedir. Kesim 4.1'de GDM'lerde tam kredibilite yaklaşımı asimptotik normal yakınsama kullanılarak bulunmuştur. Tam kredibilite sağlanamadığında kısmi kredibilite kavramı devreye girer

GDM'ler için kısmi kredibilite yaklaşımı Zhou [30] tarafından incelenmiş ve GDM'ler için tam kredibilite yaklaşımına göre varsayımlar iki açıdan esnetilmiştir. Bunlardan birincisi yanıt değişkenin dağılımının üstel dağılım ailesinden herhangi bir dağılım olma zorunluluğu kaldırılmıştır. Bunun nedeni, gözlem sayısının yanıt değişkenin dağılımı hakkında doğru bilgi veremeyecek kadar az olmasıdır. Yanıt değişkenin hangi dağılıma uyduğu analiz edilirken; gözlemlenen verilerden yararlanılmakta ve bu gözlemlere göre dağılım varsayımı yapılarak parametreler tahmin edilmektedir. İkinci varsayım esnekliği ise, birinci varsayım esnekliği ile bağlantılı olarak model parametrelerinin tahmininde kullanılan yöntem değişikliği ile yapılmıştır. GDM'lerde kısmi kredibilite yaklaşımında dağılım bilgisi gerektiren EÇO yöntemi yerine EKK parametre tahmin yöntemi tercih edilmektedir. Kısmi kredibilite yaklaşımında kısmi kredibilite faktörü, Hata Kareler Toplamını minimum yapan değerdir. GDM'lerde kısmi kredibilite faktörü, bir kredibilite matrisidir.

Kısmi kredibilite yaklaşımında örneklem boyutunun küçük olması, örneklem boyutunun sonsuza ulaştığı varsayımında sağlanabilen asimptotik özelliklerin kullanımında da sorun yaratacaktır. Kesim 4.1'de GDM için tam kredibilite yaklaşımında GDM tahmin edicileri yansız kabul edilerek s^2 varyansı bir karşılaştırma kriteri olarak kullanılmıştı. Sapma olmadığı durumda Ortalama Hata Karesi (OHK) tahmin edicisi varyansa eşit olacaktır.

Genel olarak varyans ifadesi

$$Var(X) = E\left((X - E(X))^2\right)$$

şeklindedir. OHK tahmin edicisi ise tahmin değeri ile kendi değeri arasındaki değer olan hatanın karesi alınarak (bu değeri minimum yapan tahmin edici) bulunur.

$$OHK(\hat{\mu}) = V(\hat{\mu}) + yan^2 \tag{4.11}$$

Yansız tahmin ediciler için $OHK(\hat{\mu}) = Var(\hat{\mu})$ olur. Bu sonuç s^2 değerinin tam kredibilite yaklaşımında neden bir karşılaştırma kriteri olarak kullanıldığını da açıklamaktadır. Kısmi kredibilite yaklaşımında OHK tahmin yönteminin kullanılmasının bir diğer nedeni de GDM tahmin edicilerinin yanlı olma durumundan gelmektedir. Yanlılık durumunun tespiti için kitle ortalaması μ ve i. birime ait olan μ_i ortalaması arasındaki fark bilinmeli, yan bilgisine ulaşılmalıdır. Gerçek veri ile çalışıken bu bilgiye ulaşmak zordur bu nedenle belli varsayımlar ve yaklaşımlar yapılır. Zhou [30] Hachemeister Regresyon Modeli'ni kullanarak GDM'lerde kısmi kredibilite matrisi için bir tanım yapmaktadır. Tam kredibilite yaklaşımında örneklem büyüklüğü oldukça geniş olduğundan ($n \to \infty$); daha önce bahsedilen EÇO tahmin edicilerinin asimptotik özelliklerinden yararlanarak $yan(\hat{\beta}_i) \to 0$, $Cov(\hat{\beta}_i) \to Cov(\hat{\beta}_i, \beta_i)$ bilgisine ulaşılır ve $Z_i \to \mathrm{I}$ elde edilir. Ancak daha küçük veri setlerinde $\boldsymbol{Z_i}$ kredibilite matrisini elde etmek için $yan(\hat{\beta}_i), Cov(\hat{\beta}_i)$ ve $Cov(\hat{\beta}_i, \beta_i)$ bilgisine ihtiyaç vardır.

4.3. Genelleştirilmiş Doğrusal Karma Model Yardımıyla Yapılan Risk Sınıflandırılmasının Güvenilirliğinin Tam Kredibilite Standardı ile İncelenmesi

GDKM, GDM'in rastgele etkiler ile genelleştirilmiş halidir. Y_{ij}, $i = 1,...,n$ ve $j = 1,...,n_i$ sınıflandırılmış bir veride hasar sayılarını gösteren raslantı değişkeni olsun. GDKM için doğrusal bileşen $\eta_{ij} = X'_{ij}\beta + T'_{ij}u_i$ ve doğrusal bileşen ile kitle ortalaması arasında ilişki kuran bağ fonksiyonu $g(\mu_{ij}) = \eta_{ij}$ biçimindedir. Kesim 4.1'de GDM'ler için oluşturulan tam kedibilite olasılığı formülüne benzer işlemler GDKM için yapılarak GDKM'ler için de tam kredibilite olasılığını bulmak mümkündür. GDM'lerden farklı olarak rastgele etkiler ve rastgele etki durumundaki açıklayıcı değişkenler modele eklenir. $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta + T'_{ij}u_i$ olduğundan kredibilite olasılığı g bağ fonsiyonu, β parametresi ve rastgele etki u_i cinsinden,

$$\pi_{ij} = \mathrm{P}\Big(g[(1-r)\mu_{ij}] - (X'_{ij}\beta + T'_{ij}u_i) \le (X'_{ij}\hat{\beta} + T'_{ij}\hat{u}_i) - (X'_{ij}\beta + T'_{ij}u_i) \le g[(1+r)\mu_i] - (X'_{ij}\beta + T'_{ij}u_i)\Big)$$

biçiminde ifade edilebilir. Antonio ve Beirlant [2] aktüeryal verilerde kredibilite modellerini GDKM'ler ile ele almışlardır. Aktüeryal kredibilite modellerinin, GDKM'lerle yorumlanması klasik doğrusal model yaklaşımlarına göre birçok yönden daha avantajlıdır;

1) GDKM'ler ile aktüeryal literatürde moment-bazlı tahmin yöntemine göre daha iyi bir alternatif olan olabilirlilik-bazlı tahmin kullanılabilir.
2) Sabit etkilerdeki ve kovaryanstaki olabilirliğe dayalı tahmin yapısı (aktüerya literatüründeki klasik moment temelli yönteme alternatif bir yöntem) GDKM ile daha iyi sonuç doğurabilir [2].

4.4. Genelleştirilmiş Doğrusal Karma Model ile En Büyük Doğruluk Yaklaşımı Arasındaki İlişki

En Büyük Doğruluk Kredibilite yaklaşımında hem sabit hem de rastgele etkilerleri içeren modeller vardır. GDKM'ler de kredibilite modellerine benzer şekilde bu iki etkiyi modeller. Bu açıdan kredibilite ve GDKM'ler benzerlik gösterir [22]. Bu bilgi ışığında kredibilite modellerini GDKM'lerin özel bir hali olarak ele alan çalışmalar mevcuttur [1], [2], [28].

5. UYGULAMA

Kredibilite Kuramı ve GDM genellikle hayat-dışı sigorta ürünlerinin fiyatlandırılmasında kullanılan yaklaşımlardır. Çalışmanın uygulama bölümünde bir sigorta şirketinden alınan kasko sigortası verisi kullanılmıştır. Veri, otomobil, motosiklet, kamyonet, kamyon, minibüs, otobüs gibi araç sınıfları için kasko ve zorunlu trafik sigorta poliçelerinden bir yıllık süre içinde gerçekleşen 161.697 hasarlı poliçeye ilişkin bilgiyi içermektedir. Risk sınıflandırmasından önce veri sadece otomobil sınıfı araçların kasko sigortası poliçelerini kapsayan bir veri kümesine indirgenmiştir. Analizler 2004 ve üstü model olan araçlar, diğer bir ifade ile araç yaşı en çok 10 olan 97.677 otomobil üzerinden yapılmıştır. Risk faktörleri açısından kayıp bilgi içeren poliçeler ile ortalamayı bozan aykırı değerler çıkartılarak, 0,62 parametresi ile Poisson dağılan 86.801 gözlemden oluşan bir hasar sayısı verisi elde edilmiştir.

Sınıflandırmada kullanılacak ve GDM'in olası açıklayıcı değişkenlerini oluşturacak risk faktörleri; sigortalının cinsiyeti, sigortalının yaşı, aracın motor hacmi ve sigortalıların yaşadığı şehir olarak belirlenmiştir.

Sigortalıların cinsiyeti Kadın(K) ve Erkek(E) olmak üzere iki düzeyli;

Sigortalının yaşı 18-29(Genç), 30-44(Orta), 45-58(Yaşlı) olmak üzere üç düzeyli;

Aracın motor hacmi <1600 cc(Küçük), 1600-2000 cc(Orta), >1600 cc (Büyük) olmak üzere üç düzeyli ve

Şehir, Büyük Şehir(bknz. (Ek 1)'de bulunan 30 il) ile Küçük Şehir (bknz. (Ek 1)'de bulunmayan 51 il) olmak üzere 2 düzeyli değişkenler olarak düzenlenmiştir.

Örneklemde her sigorta poliçesinin başlangıç ve bitiş tarihleri farklı olmakla beraber, her bir poliçenin 2013 yılı içinde sonlanacak birer yıllık poliçe olması nedeniyle bu farklılık göz ardı edilmiştir. Sigortalıların yıllık hasar sayısı yanıt değişkeni; GDM ile yaş, cinsiyet, motor hacmi, şehir açıklayıcı değişkenleri yardımıyla modellenmiştir. Kredibilite kavramı bu çalışmada bir karşılaştırma kriteri olarak kullanıldığından bir yıllık hasar sayısı verisi ile çalışmak kredibilite açısından sorun oluşturmamaktadır. Kredibilite Kuramının fiyatlandırmada kullanılması durumunda birkaç yıllık hasar bilgisine gerek duyulur. Panel veri analizi ile kredibilite primi bulunabilir [27].

Bu çalışmada uygulama için R 2.13.0, IMB SPSS Statistics 21 [46] ve Excel programı kullanılmıştır. Excelde verinin düzenlenmesi ve tablolaştırılması işlemleri, SPSS de verinin dağılımına ilişkin uyum iyiliği testleri ve sınıflandırma işlemleri yapılmış; R da ise sınıflandırılmış verinin GDM ile modellenmesi yapılmış ve kredibilite için gerekli değerler hesaplanmıştır.

Yanıt değişkenin Poisson dağılımlı olduğu GDM'ler Poisson Regresyon olarak adlandırılırlar. Poisson Regresyon modelinin bilgisayar ortamında analizinde genellikle SAS ve R programından yararlanılır.

Elde edilen veri seti ile iki aşamalı bir uygulama yapılmıştır. Birinci aşamada model açıklayıcı değişkenleri belirlenmiş, ikinci aşama da ise bu açıklayıcı değişkenlere göre oluşturulan modeldeki risk sınıflarının güvenilirliği Kredibilite Kuramı ile test edilmiştir.

1. Aşama: Model Açıklayıcı Değişkenlerinin Belirlenmesi

Çalışmada risk faktörleri (değişkenler); cinsiyet (Kadın-Erkek), şehir (Büyük Şehir-Küçük Şehir), yaş (Genç, Orta, Yaşlı), motor hacmi (Küçük, Orta, Büyük) olarak belirlenmiştir. Sigorta şirketinin kasko sigortası portföyündeki 2004 ve üstü model otomobilleri olan 18-58 yaş aralığındaki sigortalılarının hasar sayısınının hangi risk faktörleri (değişkenler) ile açıklanacağı belirlenecektir. Bunun için aşağıdaki soruların cevapları araştırılacaktır.

- Model hangi risk faktörlerine göre oluşturulacak? Diğer bir ifade ile modelin açıklayıcı değişkenleri neler olacak?
- Modele katılan açıklayıcı değişkenlerin katsayıları anlamlı mı?
- Oluşan modellerden en küçük bilgi (AIC veya BIC) kriterine veya en büyük log-olabilirlik değerine sahip olan model hangisidir?

Kesim 2.3.1'de belirtilen modelleme basamaklarına göre karşılaştırılacak modellerin elde edilmesinde aşağıdaki adımlar izlenir.

1) Yanıt Değişkenin (Hasar Sayısının) Dağılımı

Yanıt değişkenin ÜDA'ndeki dağılımlardan Poisson dağılıma uyumu incelenmiştir. Poisson dağılımının olasılık fonkiyonu,

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!},x=0,1,2,...$$

olarak yazılır. Yukarıda belirtilen kriterlere göre şeçilen ve 97.677 gözlemden oluşan hasar sayısı verisinin dağılımı Kolmogrov-Smirnov testi ile araştırılmıştır.

H_0 = Hasar sayılarının dağılımı ile Poisson dağılım arasında fark yoktur.

H_S = Hasar sayılarının dağılımı ile Poisson dağılım arasında fark vardır.

One-Sample Kolmogorov-Smirnov Test

		HasarSayısı
N		97677
Poisson Parameter[a,b]	Mean	,57
Most Extreme Differences	Absolute	,015
	Positive	,015
	Negative	-,007
Kolmogorov-Smirnov Z		4,798
Asymp. Sig. (2-tailed)		,000

a. Test distribution is Poisson.

b. Calculated from data.

$\text{Sig}=\text{p}=0<\alpha=0,05$ olduğundan H_0 yokluk hipotezi red edilmiştir. Bunun sebebi veri kümesindeki “0” değerli hasarların çokluğu ve ortalamayı bozan aykırı değerlerdir. “0” hasarlı verilerinin çok olduğu bu tür veri kümeleri temizlenmeden sıfır yığılmalı kesikli modeller ile modellenebilir [37]. Veri, hasar sayılarının Poisson dağılıma uyduğu varsayımından hareketle bu dağılıma uyacak şekilde düzenlenmiştir. Örneğin seçilen örneklemde sıfır hasarlı araç sayısı azaltılmıştır. Belli düzenlemeler ve veri temizliğinden sonra 86.801 gözlemden oluşan 0,62 parametresi ile Poisson dağılan hasar sayıları elde edilmiştir. Bu veri için Kolmogrov-Smirnov testi yapılmıştır.

One-Sample Kolmogorov-Smirnov Test

		HasarSayısı
N		86801
Poisson Parameter[a,b]	Mean	,62
Most Extreme Differences	Absolute	,002
	Positive	,002
	Negative	-,002
Kolmogorov-Smirnov Z		,644
Asymp. Sig. (2-tailed)		,802

a. Test distribution is Poisson.

b. Calculated from data.

$\text{Sig} = \text{p} = 0,802 > \alpha = 0,05$ olduğundan H_0 yokluk hipotezi kabul edilir.

Poisson dağılımının en önemli özelliklerinden birisi ortalama ve varyans değerlerinin birbirine eşit olmasıdır. Ortalama ve varyans değerlerinin eşitliği test edilmek istenirse verinin betimleyici istatistiklerine bakılması gerekmektedir.

Statistics

HasarSayısı

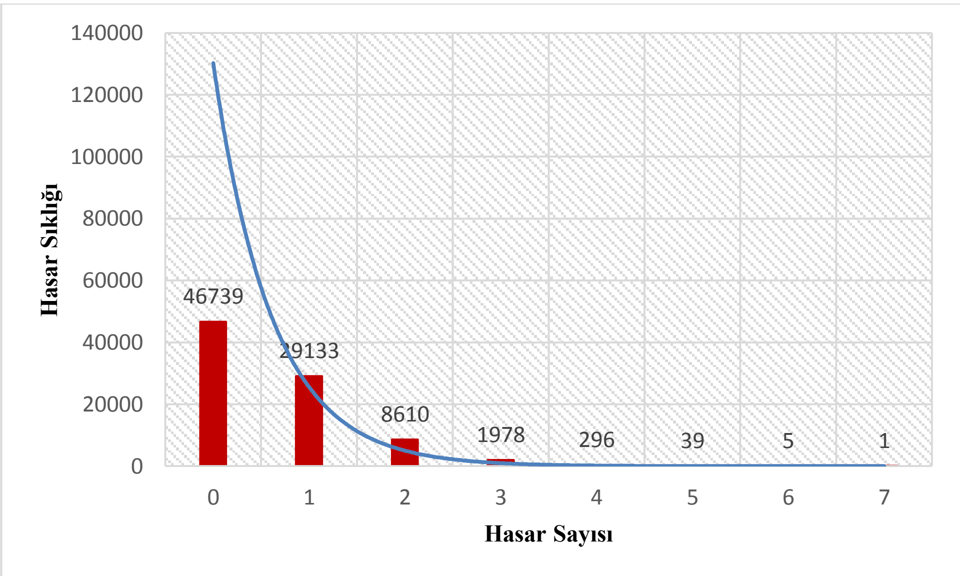
N	Valid	86801
	Missing	0
Mean		$\boxed{,62}$
Std. Error of Mean		,003
Median		,00
Mode		0
Std. Deviation		,789
Variance		$\boxed{,623}$
Skewness		1,305
Std. Error of Skewness		,008
Kurtosis		1,787
Std. Error of Kurtosis		,017
Range		7
Minimum		0
Maximum		7

$Ortalama = 0,62 \cong Varyans = 0,623$ olduğundan ortalama ve varyansın eşitliği kabul edilebilir.

Veri kümesindeki hasar sayıları ve sıklıkları aşağıdaki gibidir.

Hasar Sayısı		
Değer	**Sıklık**	**Yüzdelik**
0	46739	53,84615
1	29133	33,56298
2	8610	9,91924
3	1978	2,27981
4	296	0,34101
5	39	0,04493
6	5	0,00576
7	1	0,00012
Toplam	86801	100

Hasar sayısı dağılımının histogram grafiği ise Şekil 5.1'de verilmiştir. Şekil 5.1'den de görüleceği üzere, hasar sayılarının dağılımı simetrik değildir. Bu histogram sigorta hasar verilerinde doğrusal regresyon modelleri yerine GDM kullanım gerekliliğinin bir göstergesidir.



Şekil 5.1. Hasar Sayısı Dağılımı

2) Bağ Fonksiyon Seçimi

Hasar sayısı yanıt değişkeni Poisson dağılımlı olduğundan bağ fonksiyonu olarak logaritmik, birim veya karekök bağ fonksiyonu seçilebilir. Bu çalışmada bağ fonksiyonu olarak logaritmik bağ fonksiyonu seçilmiştir.

3) Açıklayıcı Değişkenlerin (x_i ’lerin) Seçimi

Uygulamadaki birinci aşamanın amacı açıklayıcı değişkenlerin belirlenmesidir. Dört risk faktörü kullanılarak toplam 15 farklı model oluşturulabilir.

	Yaş	Cinsiyet	Motor Hacmi	Şehir
Model-1	✓	✓	✓	✓
Model-2	✓	✓		✓
Model-3	✓		✓	✓
Model-4	✓	✓	✓	
Model-5		✓	✓	✓
Model-6	✓			✓
Model-7	✓	✓		
Model-8	✓		✓	
Model-9		✓		✓
Model-10			✓	✓
Model-11		✓	✓	
Model-12	✓			
Model-13				✓
Model-14		✓		
Model-15			✓	

Açıklayıcı Değişkenlerin Kodlanması

Modelde olmasına karar verilen faktörlere göre açıklayıcı değişkenlerin belirlenmesi gerekmektedir. Bu faktörler kategorik değişken olduğundan gösterge (dummy-(0,1)) değişkenler kullanılarak kodlanırlar. Gösterge değişkenler gözlemlerin hangi kategoriye ait olduklarını gösterirler ve bu değişkenlerin sayısal olarak önemi yoktur. Bu değişkenler iki veya daha fazla kategoriye sahip olabilirler. Cinsiyet (kadın-erkek) ve sağlık durumu (hasta-hasta değil) iki kategorili değişkenlere örnek verilebilirken; yaşanan bölge (köy, kasaba, kent), motor hacmi (küçük, orta, büyük) ve yaş(genç, orta, yaşlı) daha fazla kategoriye sahip değişkenlere örnek gösterilebilir. k düzeye sahip bir kategorik değişken, k-1 tane (0,1) ikili değişken ile kodlanır. Belli referans düzeyleri belirlenerek kodlama işlemi yapılır. Bu referans düzeyi genelde portföyde sıklığı en yüksek olan faktöre göre belirlenir [38]. Gösterge değişkenler açıklayıcı değişkenler olabildiği gibi, bağımlı değişken de olabilir. Bu durumda bağımlı değişken sadece "evet-hayır" veya "var-yok" gibi iki değer alır. GDM'lerde açıklayıcı değişken bir vektördür ve i. açıklayıcı değişken vektörü $\boldsymbol{X_i} = \left(x_{i1}, x_{i2}, x_{i3} ...\right)$ biçimindedir.

4) Gözlem değerleri bulunur.

Bu adımda herbir risk faktörü için açıklayıcı değişkenlere ($x_1, x_2, ..., x_n$ 'lere) karşılık gelen $y_1, y_2, ..., y_n$ hasar sayısı gözlemleri ve modele offset terimi olarak eklenecek poliçe sayıları bulunmuştur. Ek 2, Model-1; Ek 3, Model-2; Ek 4, Model-3; Ek 5, Model-4; Ek 6, Model-5 ve Ek 7, Model-6; Ek 8, Model-7; Ek 9, Model-8; Ek 10, Model-9; Ek 11, Model-10; Ek 12, Model-11; Ek 13, Model-12; Ek 14, Model-13; Ek 15, Model-14 ve Ek 16, Model-15 için gözlem değerlerini göstermektedir.

5) Parametre tahminleri yapılır.

15 modelin her biri için modeldeki parametrelerin tahmini yapılmış ve anlamlılıkları test edilmiştir. Tüm modellerin parametre tahmin sonuçları Ek 17'de verilmiştir. Tüm modelleri birbiri ile karşılaştırılması yerine sadece tüm açıklayıcı değişkenleri içeren Model-1, açıklayıcı değişkenlerden üçünü içeren Model-3, açıklayıcı değişkenlerden ikisini içeren Model-11 ve tek açıklayıcı değişkeni içeren Model-15 için tahmin sonuçları verilmiştir ve karşılaştırılmıştır.

Çizelge 5.1. Model Parametre Tahminleri ve Tahminler için Fisher Adım Sayısı

	Parametre Tahminleri				
Model-1		**Tahmin**	**Std. Hata**	**z değeri**	**p**
	Sabit	-0,327077	0,007271	-44,984	< 0,01
	Yaş (Genç)	0,327829	0,014405	22,757	< 0,01
	Yaş (Yaşlı)	-0,148196	0,009383	-15,793	< 0,01
	Şehir (Küçük)	-0,268841	0,015635	-17,195	< 0,01
	Cinsiyet (Kadın)	0,044331	0,009163	4,838	< 0,01
	Motor Hacmi (Büyük)	-0,664031	0,022295	-29,783	< 0,01
	Motor Hacmi (Orta)	-0,403867	0,011647	-34,676	< 0,01
	Fisher Adım Sayısı: 4				
Model-3		**Tahmin**	**Std. Hata**	**z değeri**	**p**
	Sabit	-0,310321	0,006366	-48,74	< 0,01
	Yaş (Genç)	0,328058	0,014405	22,77	< 0,01
	Yaş (Yaşlı)	-0,150471	0,009371	-16,06	< 0,01
	Şehir (Küçük)	-0,273791	0,015601	-17,55	< 0,01
	Motor Hacmi (Büyük)	-0,668418	0,022277	-30,00	< 0,01
	Motor Hacmi (Orta)	-0,407004	0,011628	-35,00	< 0,01
	Fisher Adım Sayısı: 4				
Model-11		**Tahmin**	**Std, Hata**	**z değeri**	**p**
	Sabit	-0,380348	0,005844	-65,088	< 0,01
	Motor Hacmi (Büyük)	-0,704897	0,022262	-31,663	< 0,01
	Motor Hacmi (Orta)	-0,425799	0,011624	-36,632	< 0,01
	Cinsiyet (Kadın)	0,067360	0,009131	7,377	< 0,01
	Fisher Adım Sayısı: 4				
Model-15		**Tahmin**	**Std, Hata**	**z değeri**	**p**
	Sabit	-0,356776	0,004846	-73,62	< 0,01
	Motor Hacmi (Büyük)	-0,712521	0,022238	-32,04	< 0,01
	Motor Hacmi (Orta)	-0,430853	0,011603	-37,13	< 0,01
	Fisher Adım Sayısı: 2				

Modeldeki her bir parametrenin anlamlılık testi için,

$H_0 : \beta_{ij} = 0$

$H_S : \beta_{ij} \neq 0$ i=1,3,11,15 ; j=1,…,her bir sınıftaki açıklayıcı değişken sayısı

hipotezleri kurulur.

$Sig = \mathrm{p} < \alpha = 0,05$ ise Ho red edilir ve parametre anlamlıdır yorumu yapılır. Herbir model için $Sig = \mathrm{p} < \alpha = 0,05$ koşulu sağlandığından, parametreler anlamlı bulunmuştur. Sınıflandırılmış veriler ile modelleme yapıldığından bu beklenen bir sonuçtur. Fisher adım sayısı bu parametrelerin tahmini sırasında yapılan yinelemeli işlem sayısını vermektedir. Sınıflandırılmamış veride açıklayıcı değişkenlerin belirlenmesinde farklı seçim yöntemleri kullanılabilir. Lemaire [47], otomobil sigortalarında fiyatlamada kullanılacak değişkenlerin belirlenmesi için eleme yöntemi, ileriye doğru seçim yöntemi ve aşamalı seçim yöntemi olmak üzere üç farklı yöntem üzerinde durmuştur. Bu yöntemlerde regresyon katsayısının anlamlı olup olmadığı hipotezi , Fisher Snedercor'un F testine göre incelenir. Eleme yönteminde başlangıçta bütün açıklayıcı değişkenler modele dahil edilir, anlamlı bir model bulununcaya kadar değişkenler çıkarılır. İleriye doğru seçim yönteminde, modele dışarda önemli bir değişken kalmayacak şekilde değişkenler teker teker eklenir. Ancak bu yöntemde bazı değişkenler eklendikçe ilk eklenen değişkenler gereksiz hale gelebilir. Bu durumda da aşamalı seçim yöntemi kullanılır ve Lemaire'e göre bu yöntemler arasından en iyisi aşamalı seçim yöntemidir [47].

Oluşturulan GDM'ler arasından en iyi GDM'i bulmak için parametrelerin anlamlı olup olmamasına göre bir eleme yapılmadığı için model, bilgi kriterleri karşılaştırılarak seçilmiştir.

6) **Model seçim kriterleri ile model test edilir.**
Tüm modeller için Akaike Bilgi Kriteri (AIC), Bayesci Bilgi Kriteri (BIC) ve Log-Olabilirlik değerleri bulunmuş ve Ek 18'de verilmiştir. Bu kesimde sadece tüm açıklayıcı değişkenleri içeren Model-1, açıklayıcı değişkenlerden üçünü içeren Model-3, açıklayıcı değişkenlerden ikisini içeren Model-11 ve tek açıklayıcı değişkeni içeren Model-15 için sonuçlar verilmiştir.

Çizelge 5.2. Modeller için Akaike Bilgi Kriteri, Bayesci Bilgi Kriteri, Log-Olabilirlik Değeri

	AIC Bilgi Kriteri	**BIC Bilgi Kriteri**	**Log-Olabilirlik Değeri**
Model-1	AIC: 464,8009	BIC: 475,8855	Log-Olabilirlik: -225,4005 (df=7)
Model-3	AIC: 265,3366	BIC: 270,6788	Log-Olabilirlik: -126,6683 (df=6)
Model-11	AIC: 101,5892	BIC: 100,7562	Log-Olabilirlik: -46,79459 (df=4)
Model-15	AIC: 38,93817	BIC: 36,23401	Log-Olabilirlik: -16,46909 (df=3)

Bu dört GDM arasından en iyi sınıflandırmanın yapıldığı model AIC, BIC bilgi kriterleri veya Log-Olabilirlik değeri yardımıyla bulunabilir. Tüm açıklayıcı değişkenler kullanılarak oluşturulan Model-1'in en iyi model olması beklenmektedir, ancak analiz sonuçları öyle çıkmamıştır. Bazı açıklyıcı değişkenler, birlikte modelde yer aldığında etkilerini yitirmektedirler. Risk sınıflarının sayısının azlığından dolayı tek açıklayıcı değişkenli modeller karşılaştırmaya dahil edilmemişlerdir. En küçük AIC ve BIC bilgi kriteri ile en büyük Log-Olabilirlik değerine sahip olan Model-11 seçilmiştir. Bu modelin açıklayıcı değişkenleri motor hacmi ve cinsiyettir.

2. Aşama: Genelleştirilmiş Doğrusal Model'deki Risk Sınıflarının Güvenilirliğinin Tam Kredibilite ile Analizi

Bu aşamada ilk aşamada karar verilen GDM'in güvenilir bir model olup olmadığının tam kredibilite kriteri ile karşılaştırılması ve bu GDM'e göre oluşan risk sınıflarının fiyatlandırmada kullanılabilir olup olmadıklarına karar verilmesi gerekmektedir.

Birinci aşamada karar verilen açıklayıcı değişkenlere ve onların düzeylerine göre oluşturulan risk sınıflarının güvenilirliğinin test edilmesi için her bir risk grubunun verilen uygun hata-tahmin tolerans (r) değeri için Tam Kredibilite yaklaşımıyla karşılaştırma kriterleri olan π_i ve s_i^2 elde edilmiştir. Tam kredibilite yaklaşımı bir bireyin veya bir grup sigortalının deneyimlerine güvenilip güvenilemeyeceğini test ederken kullanılan bir yaklaşımdır. Burada da herbir risk sınıfındaki hasar sayısı deneyiminin güvenilir olup olmadığı test edilmiştir.

İlk aşamada karar verilen model için herbir risk sınıfındaki hasar sayısı ve poliçe sayısı Çizelge 5.3'teki gibidir.

Çizelge 5.3. Motor Hacmi ve Cinsiyete göre Sınıflanmış Poliçe Sayısı ve Hasar Sayısı Verisi

	Poliçe Sayısı	**Hasar Sayısı**	**Motor Hacmi**	**Cinsiyet**
Sınıf 1	20826	15065	Küçük(1)	Kadın(1)
Sınıf 2	40013	27518	Küçük(1)	Erkek(2)
Sınıf 3	5305	2592	Orta(2)	Kadın(1)
Sınıf 4	14472	6405	Orta(2)	Erkek(2)
Sınıf 5	1427	622	Büyük(3)	Kadın(1)
Sınıf 6	4758	1501	Büyük(3)	Erkek(2)

Motor hacmi ve cinsiyete göre sınıflanan veri 6 risk sınıfından oluşmaktadır. Birinci risk sınıfı kadın sürücü ve küçük motor hacimli arabaya sahip sigortalıları kapsar. Bu sınıftaki

20.826 sigortalıdan 15.065 hasar gelmiştir. Diğer risk sınıfları da benzer biçimde açıklanabilir.

Bu değişkenlere göre oluşacak model için açıklayıcı değişkenler $X_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}\}$ biçimindedir ve x_{ij} 'ler referans değere göre kodlanırlar. Kodlama yapılırken referans değerleri (0 alacak değerler) sınıfta en çok bulunan değerlere göre belirlenir. Motor hacmi için referans değer küçük, cinsiyet için referans değer erkektir.

$$x_{i1} = 1$$

$$x_{i2} = \begin{cases} 1, motorhacmi = büyük(3) \\ 0, diğer \end{cases}$$

$$x_{i3} = \begin{cases} 1, motorhacmi = orta(2) \\ 0, diğer \end{cases}$$

$$x_{i4} = \begin{cases} 1, cinsiyet = kadın(1) \\ 0, diğer \end{cases}$$

x_{i1} 'nin tüm i değerleri için 1 şeçilmesinin nedeni modele sabit terimin dahil edilmesidir [25].

Bu kodlamalara göre herbir risk sınıfı için açıklayıcı değişkenler $X_1 = \{1,0,0,1\}$, $X_2 = \{1,0,0,0\}$, $X_3 = \{1,0,1,1\}$, $X_4 = \{1,0,1,0\}$, $X_5 = \{1,1,0,1\}$ ve $X_6 = \{1,1,0,0\}$ şeklindedir. Her bir risk sınıfının sözel ifadeleri ise söyledir;

$X_1 = \{1,0,0,1\}$	Küçük motor hacimli arabaya sahip kadın sigortalıları temsil eder.
$X_2 = \{1,0,0,0\}$	Küçük motor hacimli arabaya sahip erkek sigortalıları temsil eder.
$X_3 = \{1,0,1,1\}$	Orta motor hacimli arabaya sahip kadın sigortalıları temsil eder.
$X_4 = \{1,0,1,0\}$	Orta motor hacimli arabaya sahip erkek sigortalıları temsil eder.
$X_5 = \{1,1,0,1\}$	Büyük motor hacimli arabaya sahip kadın sigortalıları temsil eder.
$X_6 = \{1,1,0,0\}$	Büyük motor hacimli arabaya sahip erkek sigortalıları temsil eder.

Belirlenen açıklayıcı değişkenlere göre oluşan GDM $g(\mu_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$ olarak yazılır. Bağ fonksiyonu logaritmik olduğundan bu eşitlik

$\ln \mu_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}$ eşitliğine dönüşür. Hasar sayısı yanlız bırakılırsa $E[Y_i] = \mu_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4})$ elde edilir. Uygulamanın birinci aşamasında Model-11 için bulunan parametre tahminlerine göre i.risk sınıfı için GDM $E[Y_i] = \mu_i = \exp(-0,380348 - 0,704897 x_{i2} - 0,425799 x_{i3} - 0,067360 x_{i4})$ şeklini alır.

Herbir risk sınıfı için açıklayıcı değişken vektörleri ve kodları belirlendikten sonra bu risk sınıflarının güvenilirliği tam kredibilite yaklaşımı kullanılarak karşılaştırılmıştır. Modelin güvenilirliğine hasar sayısının etkisi, poliçe sayısının etkisi, örneklem büyüklüğünün etkisi, GDM bileşenlerinden açıklayıcı değişkenin etkisi ve bağ fonksiyonun türü incelenmiştir. Bu incelemeler yapılırken parametrelerin Kesim 2.3.2'de söz edilen EÇO tahmin edicilerinin asimptotik özellikleri sağladığı varsayılmaktadır. Bu varsayım yardımıyla karşılaştırma kriterleri olan π_i ve s_i^2 değerlerini hesaplayabilmek için varyans-kovaryans matrisi bulunacaktır. Varyans-kovaryans matrisi olan Σ bulunurken kullanılacak w_i ağırlıkları; herbir risk sınıfındaki poliçe sayıları (exposure) olarak alınmıştır. Ayrıca poliçe sayılarının logaritmaları alınarak modele offset terimi olarak da eklenmiş ve modelde bir dengeleme görevi görmüşlerdir. Herbir risk sınıfı için ağırlıklar $w_1 = 20826$, $w_2 = 40013$, $w_3 = 5305$, $w_4 = 14472$, $w_5 = 1427$ ve $w_6 = 4758$'dir. Σ varyans-kovaryans matrisi R.2.13.0 programı yardımıyla aşağıdaki gibi bulunmuştur.

$$\Sigma = \begin{pmatrix} 3,414793\text{e-}05 & -2,690512\text{e-}05 & -2,574885\text{e-}05 & -2,981917\text{e-}05 \\ -2,690512\text{e-}05 & 4,956129\text{e-}04 & 2,421035\text{e-}05 & 9,567214\text{e-}06 \\ -2,574885\text{e-}05 & 2,421035\text{e-}05 & 1,351129\text{e-}04 & 6,334113\text{e-}06 \\ -2,981917\text{e-}05 & 9,567214\text{e-}06 & 6,334113\text{e-}06 & 8,337871\text{e-}05 \end{pmatrix}$$

Açıklayıcı değişkenler ve varyans-kovaryans matrisinin sayısal değeri bilindiğine göre herbir risk grubu için s_i^2 varyans değerleri matris çarpımı işlemleri ile Eşitlik 4.4'den hesaplanabilir. Her bir risk sınıfı için π_i kredibilite olasılıkları ise Eşitlik 4.12'den hesaplanabilir. Hata-tahmin tolerans değeri 0<r<1 için 0,5; 0,1; 0,01 ve 0,001 gibi değerler denenmiştir. Diğerlerine göre daha anlamlı sonuçlar vermesi nedeniyle r=0,01 olarak alınmıştır.

Cinsiyet ve motor hacmine göre oluşturulan GDM'deki 6 risk sınıfı için asimptotik varyans ve kredibilite olasıklıkları Çizelge 5.4'de gösterilmiştir.

Çizelge 5.4. Asimptotik Varyans ve Kredibilite Olasılıkları (Log-Bağ Fonksiyonuna Göre)

	X_i	s_i^2	π_i
Sınıf-1	(1,0,0,1)	0,0000578883	0,8112701
Sınıf-2	(1,0,0,0)	0,0000341479	0,9129671
Sınıf-3	(1,0,1,1)	0,0001541717	0,5794078
Sınıf-4	(1,0,1,0)	0,0001177631	0,6432215
Sınıf-5	(1,1,0,1)	0,0005188254	0,3393665
Sınıf-6	(1,1,0,0)	0,0004759506	0,3533257

Tam kredibilite için gerekli olan asimptotik varyans değeri $s_*^2 = 0{,}000037328$ ($\left[\frac{\ln(1-r)}{Z_{0,90}}\right]^2 = \left[\frac{\ln(1-0,01)}{1,645}\right]^2 = 0,000037328$) ile her bir risk sınıfının s_i^2 asimptotik varyansları karşılaştırıldığında; $s_*^2 > s_i^2$ koşulunu sağlayan risk sınıfları tam güvenilirdir ve o risk sınıfında bulunan gözlem sayısı prim hesaplaması için uygundur sonucuna varılabilir. Karşılaştırma sonucunda;

$s_1^2 = 0,0000578883 \; > \; s_*^2 = 0,000037328$
$s_2^2 = 0,0000341479 < \; s_*^2 = 0,000037328$
$s_3^2 = 0.0001541717 \; > s_*^2 = 0,000037328$
$s_4^2 = 0.0001177631 \; > s_*^2 = 0,000037328$
$s_5^2 = 0.0005188254 \; > s_*^2 = 0,000037328$
$s_6^2 = 0.0004759506 \; > s_*^2 = 0,000037328$

elde edilir. $s_*^2 > s_i^2$ koşulunu sağlayan ikinci risk sınıfı tam güvenilir bulunmuştur. Diğer risk sınıfları tam güvenilir değildir ancak kendi aralarında karşılaştırılarak daha yüksek güvenilirlik gösteren risk sınıfı bulunabilir. Daha küçük asimptotik varyans değerine sahip olan birinci risk sınıfı diğer risk sınıflarına göre daha güvenilirdir. Poliçe sayısı ve hasar sayısı miktarlarına bakıldığında birinci risk sınıfının diğer risk sınıflarına göre daha fazla gözlem içerdiğinden birinci sınıfın geçmiş hasar deneyiminin fazla olması beklenen bir sonuçtur. Ayrıca π_i kredibilite olasıkları karşılaştırıldığında; ikinci risk sınıfına ait $\pi_2 = 0.9129671 >= \pi = 0{,}90$ olduğundan tam güvenilirlik sağlamaktadır. Birinci risk sınıfına ait kredibilite olasılığı $\pi_1 = 0.8112701$ diğer risk sınıflarının kredibilite olasılıklarının en büyüğüdür ve $\pi = 0{,}90$ tam kredibilite olasılığına yakındır; bu da birinci risk sınıfının diğer risk sınıflarına göre daha güvenilir olduğunu gösterir.

Kesim 3.1.1.3'de Poisson dağılımlı verinin tam güvenilir olabilmesi için bildirilmesi gereken hasar sayısı; $p = 0{,}90$ ($y_p = y_{0,9} = 1{,}645$) ve r=0,01 olarak alınması durumunda en az $n_{poisson} \geq \left(\frac{y_p}{r}\right)^2 = \left(\frac{1{,}645}{0{,}01}\right)^2 = 27060{,}25 \approx 27060$ olarak hesaplanmıştı. İkinci risk sınıfında gözlenen hasar sayısının tam güvenilirlik için gereken minimum hasar sayısından büyük olması da ($27518 \geq 27060$) bu risk sınıfının tam güvenilir olduğunu doğrulamıştır.

5.1. Risk Sınıfında Bulunan Hasar Sayısının Güvenilirliliğe Etkisi

Hasar sayısının güvenilirliliğe etkisini görmek için aynı poliçe sayılarına sahip risk sınıflarında meydana gelen hasarlara göre bulunmuş s_i^2 ve π_i değerlerinin karşılaştırılması gerekmektedir. Ancak veri kümesinde poliçe sayıları eşit olan risk sınıfı bulunmamaktadır. Poliçe sayıları diğerlerine göre daha yakın olan üçüncü ve 6.risk sınıflarının hasar sayılarına bakıldığında; üçüncü risk sınıfında 6. risk sınıfına oranla daha fazla hasar bulunmaktadır. Üçüncü risk grubunda poliçe başına düşen hasar sayısı (hasar sayısı/poliçe sayısı) 6. risk grubunda poliçe başına düşen hasar sayısından fazladır. Bu durumda üçüncü risk sınıfı için 6.risk sınıfına göre daha küçük asimptotik varyans ve daha büyük kredibilite olasılığı beklenir. $s_3^2 = 0.0001541717 < s_6^2 = 0.0004759506$ ve $\pi_3 = 0.5794078 > \pi_6 = 0.3533257$ sonuçları bu beklentiyi doğrular. Risk sınıfında bulunan hasar sayısı güvenilirlik üzerinde etkili bir unsurdur.

5.2. Risk Sınıfında Bulunan Poliçe Sayısının Güvenilirliliğe Etkisi

Poliçe sayısının güvenililirliliğe etkisini görmek için aynı hasar sayılarına sahip risk sınıflarındaki poliçe sayılarına göre bulunan s_i^2 ve π_i değerlerinin karşılaştırmak gerekmektedir. Ancak veri kümesinde hasar sayıları eşit olan risk sınıfı bulunmamaktadır. Hasar yoğunluğu diğerlerine göre daha fazla olan birinci ve ikinci risk sınıflarındaki poliçe sayısı/hasar sayısı oranı ikinci risk grubunda daha fazladır. Bu durumda ikinci risk sınıfı için birinci risk sınıfına göre daha küçük asimptotik varyans ve daha büyük kredibilite olasılığı beklenir. $s_2^2 = 0,0000341479 < s_1^2 = 0,0000578883$ ve $\pi_2 = 0.9129671 > \pi_1 = 0.8112701$ sonuçları bu beklentiyi doğrular. Daha fazla poliçe sayısına sahip olan ikinci risk sınıfı daha güvenilir çıkmıştır. Ayrıca ikinci risk sınıfı tam güvenilir bir risk sınıfı olduğundan bütün risk sınıflarından daha güvenilir olması beklenen bir sonuçtur. Poliçe sayısı, GDM'lerde kredibiliteyi etkileyen bir unsurdur.

Hasar sayısı ve poliçe sayısının güvenilirlik üzerindeki etkileri ayrı ayrı incelendikten sonra ikisinin birlikte güvenirliliği nasıl etkileyeceğinin de analiz edilmesinde fayda vardır.

5.3. Risk Sınıfında Bulunan Örneklem Büyüklüğünün Güvenilirliliğe Etkisi

Örneklem sayısı artıkça oluşturulan GDM'in tam kredibiliteye daha eğilimli olacağı açıktır. Ancak örneklem ne kadar artırılmalıdır ki bir risk sınıfı tam güvenilirlik sağlansın sorusuna verilecek cevap önemlidir. Tam güvenilirliliğe en yakın risk sınıfı olan birinci risk sınıfında örneklem (hem poliçe sayısı hem de hasar sayısı) ne kadar artırılmalı ki risk sınıfı tam güvenilir olsun sorusunun cevabı aranmaktadır. Poliçe sayılarındaki artışa bağlı olarak hasar sayılarının da benzer biçimde artacağı varsayılmıştır. Birinci risk sınıfının asimptotik varyansı ve kredibilite olasılığının, tam güvenilirlilik durumundaki asimptotik varyans $s_*^2 = 0,000037328$ ve %90 düzeyinde tam kredibilite olasılığı $\pi = 0.90$ değerine yaklaşmasına göre örneklem boyutu artırılabilir.

Örneklem boyutu 2 kat arttırıldığında birinci risk sınıfı için $s_1^2 = 0,000028944 < s_*^2 = 0,000037328$ ve $\pi_1 = 0,9369339 > \pi = 0,90$ elde edilir ve tam güvenilirlik koşulu sağlanmış olur.

Bu 6 risk sınıfı içinden en az güvenilir olan beşinci risk sınıfı için örneklem boyutu ne kadar artırılırsa tam güven sağlanmış olur sorusunun cevabını aşağıdaki tablodan rahatlıkla görülebilir.

Arttırılma Miktarı	Asimptotik Varyans	Kredibilite Olasılığı	Tam Kredibilite
2 kat	$s_5^2 = 0{,}0002594127 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}46533 < \pi = 0{,}90$	Sağlanmaz.
5 kat	$s_5^2 = 0{,}000103765 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}6737577 < \pi = 0{,}90$	Sağlanmaz.
10 kat	$s_5^2 = 0{,}000051883 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}8349641 < \pi = 0{,}90$	Sağlanmaz.
12 kat	$s_5^2 = 0{,}000043235 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}8716986 < \pi = 0{,}90$	Sağlanmaz.
13 kat	$s_5^2 = 0{,}000039909 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}8865628 < \pi = 0{,}90$	Sağlanmaz.
14 kat	$s_5^2 = 0{,}000037059 > s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}899551 < \pi = 0{,}90$	Sağlanmaz.
15 kat	$s_5^2 = 0{,}0000345883 < s_*^2 = 0{,}000037328$	$\pi_5 = 0{,}9109316 > \pi = 0{,}90$	Sağlanır.

Altı risk sınıfı arasından tam güvenilirlik sağlamayan ancak tam güvenilirliliğe en yakın olan birinci risk sınıfının tam güvenilir olması için örneklem boyutunun iki kat arttırılması yeterlidir. Tam güvenilirliliğe en uzak olan beşinci risk grubunun tam güvenilir olarak kabul edilebilmesi için ise örneklem boyutu en az 15 katına çıkarılmalıdır. Örneklem büyüklüğü en az 15 kat artırıldığında beşinci risk sınıfı için asimptotik varyans değeri tam kredibilite durumundaki asimptotik varyans değerinden küçük olur ($s_5^2 = 0{,}0000345883 < s_*^2 = 0{,}000037328$) ve $\pi_5 = 0{,}9109316 \approx 0{,}90$ kredibilite olasılığı ile tam kredibilite koşulunu sağlar. Bu sonuçlar, GDM'lerde örneklemin büyüklüğü arttıkça tam kredibilite eğiliminin de arttığını göstermektedir.

5.4. Risk Sınıflarına Göre Açıklayıcı Değişkenlerin Güvenilirliliğe Etkisi

Açıklayıcı değişkenlerin güvenilirliliğe etkisini incelemek için cinsiyet ve motor hacmi açıklayıcı değişkenlerinin dağılımına göre, s_i^2 ve π_i değerleri karşılaştırılmıştır. 6 risk sınıfı arasından cinsiyet açıklayıcı değişkeninin etkisini görmek amacıyla aynı motor hacmi açıklayıcı değişkenine ve farklı cinsiyet açıklayıcı değişkenine sahip olan risk sınıfları karşılaştırılmıştır. Sınıf-1, Sınıf -2 kendi arasında; Sınıf -3, Sınıf -4 kendi arasında ve Sınıf -5, Sınıf -6 kendi arasında karşılaştırıldığında cinsiyet açıklayıcı değişkeninde erkek kategorisine sahip olan risk sınıfları daha küçük asimptotik varyans ile daha büyük kredibilite olasılığına sahiptir. Örneğin orta motor hacmine sahip üçüncü ve dördüncü risk sınıflarında erkek sigortalıları içeren dördüncü risk sınıfının hasar bilgisi kadın sigortalıları içeren üçüncü risk sınıfına daha güvenilir olduğu sonucuna

$s_3^2 = 0,0001541717 > s_4^2 = 0,0001177631$ ve $\pi_3 = 0,5794078 < \pi_4 = 0,6432215$ karşılaştırma sonuçları yardımıyla ulaşılır.

5.5. Bağ Fonksiyonunun Güvenilirliliğe Etkisi

Bağ fonksiyonunun güvenilirliğe etkisi incelenirken hem bağ fonksiyonunun türü hem de bağ fonksiyonunun yapısı değiştirilerek analizler yapılabilmektedir. Zhou ve Garrido [29] bir GDM'in bağ fonksiyonunun yapısının tahmin edicilerin güvenilirliğine etkisi olmadığı sonucuna ulaşmışlardır. Bağ fonksiyonun yapısı, fonksiyona c gibi bir sabit eklemek (veya çıkarmak), bir sabit ile çarpmak (veya bölmek) gibi işlemler sonucu değiştirilebilir. Bu çalışmada ise bağ fonksiyonunun türünün etkisi araştırılmıştır. Bağ fonksiyonunun türünün değişiminin güvenilirlik üzerindeki etkisinin analiz edilmesinde hem logaritmik hem de birim bağ fonksiyonuna göre bulunan model sonuçları (π_i değerleri) karşılaştırılmıştır. Logaritmik bağ fonksiyonuna göre oluşturulan modeldeki risk sınıflarının güvenilirlik sıralaması Sınıf-2>Sınıf-1>Sınıf-4>Sınıf-3>Sınıf-6>Sınıf-5 biçimindedir. Bu sıralamada en güvenilir risk grubu Sınıf-2 iken en az güvenilir sınıf Sınıf-5'tir. Bu sıralama birim bağ fonksiyonu kullanılarak oluşturulmuş modelde de elde edilirse, bağ fonksiyonu türünün güvenilirliliği etkilemediği yorumu yapılabilir. Burada dikkat edilmesi gereken nokta, birim bağ fonksiyonuna göre hesaplanan asimptotik varyans s_i^2 değerleri ile kredibilite olasılıkları π_i değerlerinin formüllerinin değişecek olmasıdır. Ayrıca logaritmik bağ fonksiyonunun kullanıldığı model için tam kredibilite koşulunu sağlayan risk sınıfı için standart bir formül elde edilebilmiştir, çünkü sınıf ortalamasından bağımsız bir olasılık formülü kullanılmıştır. Birim bağ fonksiyonunun kullanıldığı durumda ise tam kredibilite koşulu için standart bir değere ulaşılamaz, çünkü kredibilite formülleri sınıf ortalamasına bağlı bir formülden elde edilir. Bu nedenle sınıflar kendi aralarında karşılaştırılacaktır ve karşılaştırma kriterleri, kredibilite olasılıkları, π_i değerleri olacaktır. Birim bağ fonksiyonu kullanılması varyans yapısını oldukça değiştirmiştir.

Çizelge 5.5. Asimptotik Varyans ve Kredibilite Olasılıkları (Birim-Bağ Fonksiyonuna Göre)

	X_i	s_i^2	π_i
Sınıf-1	(1,0,0,1)	10964,67	0,9427629
Sınıf-2	(1,0,0,0)	11077,22	0,96437
Sınıf-3	(1,0,1,1)	2301,063	0,4927535
Sınıf-4	(1,0,1,0)	2852,476	0,6865829
Sınıf-5	(1,1,0,1)	416,0239	0,1658234
Sınıf-6	(1,1,0,0)	1956,27	0,4479042

Çizelge 5.4 birim bağ fonksiyonuna göre s_i^2 ve π_i değerlerini vermektedir.

Çizelge 5.6. Logaritmik ve Birim Bağ Fonksiyonuna Göre Kredibilite Olasılıkları

	π_i (Logaritmik Bağ Fonk.)	**π_i (Birim Bağ Fonk.)**
Sınıf-1	0,8112701	0,9427629
Sınıf-2	0,9129671	0,96437
Sınıf-3	0,5794078	0,4927535
Sınıf-4	0,6432215	0,6865829
Sınıf-5	0,3393665	0,1658234
Sınıf-6	0,3533257	0,4479042

Logaritmik bağ fonksiyonuna göre güvenilirlilik sıralaması;

$$\pi_2 > \pi_1 > \pi_4 > \pi_3 > \pi_6 > \pi_5$$

şeklindedir. Birim bağ fonksiyonuna göre ise güvenilirlilik sıralaması;

$$\pi_2 > \pi_1 > \pi_4 > \pi_3 > \pi_6 > \pi_5$$

biçimindedir. Bağ fonksiyonu türünün güvenilirliliği etkilemediği söylenebilir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışmasında doğrusal modellerin genelleştirilmiş biçimi olan GDM genel hatlarıyla incelenmiştir. Doğrusal modellerin varsayımları ve parametre tahmin yöntemleri ÜDA içindeki dağılımlar için genelleştirilerek GDM tanımı yapılmış; model bileşenlerinden, parametre tahmin yöntemlerinden ve modelleme basamaklarından bahsedilmiştir. GDM gibi aktüeryal fiyatlandırma yöntemlerinden biri olan Kredibilite Kuramı ve yaklaşımları anlatılmış ancak Sınırlı Dalgalanmalı Kredibilite yaklaşımının üzerinde durulmuştur. Tam kredibilite standardı kullanılarak GDM ve Sınırlı Dalgalanmalı Kredilite yaklaşımı birlikte ele alınmaya çalışılmıştır. GDM'deki açıklayıcı değişkenler yardımıyla oluşturulan risk sınıflarının güvenilirliğinin analizinde kullanılabilecek karşılaştırma kriterleri belirlenmiştir. Bu kriterler asimptotik varyans ve kredibilite olasılığıdır. Küçük asimptotik varyans ve büyük kredibilite olasılığı daha güvenilir risk sınıfı anlamına gelmektedir.

Özel bir sigorta şirketinin bir yıllık kasko sigortası poliçelerinden gelen hasar sayısı verisi kullanılarak yapılan uygulamanın ilk aşamasında model açıklayıcı değişkenlerini belirlemek amacıyla modelleme basamakları kullanılmıştır. Bireysel veri yerine sınıflandırılmış veri kullanılması nedeniyle parametre anlamlılıklarına göre model seçimi yapılamamış; AIC, BIC ve Log-Olabilirlik gibi bilgi kriterlerinden yararlanılmıştır. Modelin açıklayıcı değişkenlerinin sigortalının cinsiyeti ve aracın motor gücü olmasına karar verilmiştir. Cinsiyet ve motor gücü açıklayıcı değişkenlerine göre oluşturulmuş risk sınıflarının aktüeryal değerleme için uygun olup olmadığı, tam kredibilite sağlayan risk sınıfının olup olmadığı karşılaştırma kriterleri yardımıyla belirlenmeye çalışılmıştır. Küçük motor hacimli arabaya sahip erkek sigortalıları temsil eden risk sınıfı tam güvenilir bulunmuştur. Tam kredibilitenin sağlanmadığı risk sınıfları kendi aralarında karşılaştırılmış ve tam güvenilir olmaları için gereken minimum gözlem sayısı belirlenmiştir. Tam güvenilir olmaya en yakın olan küçük motor hacimli arabaya sahip kadın sigortalı sayısının iki katına çıkması bu sigortalıların tam güvenilir olması için yeterli iken; tam güvenilir olmaya en uzak olan büyük motor hacimli arabaya sahip bayan sigortalıların tam güvenilir olması için örneklem sayısı 15 katına çıkarılmalıdır. Ayrıca hasar sayısı, poliçe sayısı ve örneklem büyüklüğü arttıkça güvenilirliğin artacağı karşılatırma kriterleri ile gösterilmiştir. GDM bileşenlerinden açıklayıcı değişkenlerin güvenilirlik üzerinde etkili bir unsur olduğu görülmüştür. Motor hacmi aynı iken, erkek sürücülerin kadınlara oranla daha fazla hasar bildirdikleri gözlemlenmiştir. Ortalama ile doğrusal bileşen arasında bağ kuran bağ fonksiyonu türünün değişimini incelemek amacıyla logaritmik ve birim bağ fonksiyonuna göre oluşturulmuş model sonuçları incelenmiştir. Farklı bağ fonksiyonları kullanıldığında farklı karşılaştırma değerleri hesaplanmakla beraber, risk sınıfları arasındaki güvenilirlik sıralaması değişmemiştir. Bağ fonkiyonunun yapısı teorik olarak incelenmiş ve farklı bağ fonksiyonlarında aynı kredibilite olasılığı elde edilmiştir.

Bu tez çalışmasında GDM için Kısmi Kredibilite kavramına ve GDKM için sadece Tam Kredibilite kavramına yüzeysel olarak yer verilmiştir. Bu konuda çalışacaklar GDM-GDKM modellerinde Kısmi Kredibilite yaklaşımını'na ilişkin çalışabilirler. Çalışmada yoğun olarak Sınırlı Dalgalanmalı Kredibilite yaklaşımına değinilmiş ve güvenilirlik karşılaştırması için tam kredibilite standardı kullanılarak belli kriterler oluşturulmuştur. Diğer kredibilite modellerinin güvenilirlik karşılaştırılmasında ve belli karşılaştırma kriterlerinin belirlenmesinde kullanılıp kullanılamayacağı araştırılabilir. Ayrıca bu çalışmada sadece hasar sayısı verisi kullanılarak GDM ve güvenilirlik çalışması yapılmıştır. Aynı gözlemlerin hasar tutarı verisine ulaşılması durumunda, hasar tutarları GDM (örneğin Gamma GDM) ile modellendikten sonra kredibilite çalışması yapılabilir. En son hem hasar sayısı hem de hasar tutarı verisine sahip olunacağı için bir GDM ile bir fiyatlama ve güvenilirlik çalışması yapılabilir.

KAYNAKLAR

[1] Ohlsson, E., Combining Generalized Linear Models and Credibility Models in Practice, *Scandinavian Actuarial Journal* 4, 301-314, **2008.**

[2] Antonio, K., Beirlant, J., Actuarial Statistics with Generalized Linear Mixed Models, *Insurance: Mathematics and Economics*, 19, **2007.**

[3] Nelder, J. A., Wedderburn, R. W. M., Generalized Linear Models, *Journal of the Royal Statistical Society*, Series A, 135(3), 370-384, **1972.**

[4] McCullagh, P., Nelder, J.A., *Generalized Linear Models*, Chapman and Hall, London, 511, **1989.**

[5] Renshaw, A.E., Actuarial Graduation Practice and Genaralised Linear and Non-Linear Models, *Journal of the Institute of Actuaries*, 118, 295-312, **1991.**

[6] Renshaw, A.E., Verrall, P., A Stochastic Model Underlying The Chain Ladder Technique, *Proceedings of the XXV ASTIN Colloquium*, Cannes, **1994.**

[7] Haberman, S., Renshaw A., E., Generalized Linear Models and Actuarial Science, *The Statistician*, 45(4) , 407–436, **1996.**

[8] Andersen, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., Thandi, N., *A Practitioner's Guide to Generalized Linear Models*, Second Edition, CAS Study Note, **2005.**

[9] De Jong, P., Heller, G., Z., *Generalized Linear Models for Insurance Data*, Cambridge University Press, London, 196, **2008.**

[10] Lee, Y., Nelder, J.A., H*ierarchical Generalized Linear Models (with discussion)*, **1996.**

[11] Mowbray, A.H., *How Extensive A Payroll Exposure İs Necessary To Give A Dependable Pure Premium*, Proc. Of the Casualty Actuarial Society, 1, 24-30, **1914.**

[12] Whitney, A.,W., *The Theory Of Experience Rating*, Proc. of the Casualty Actuarial Society, 4, 274-292, **1918.**

[13] Goulet, V., Principles and Applications of Credibility Theory, *Journal of Actuarial Practice*, 6, 5-62, **1998.**

[14] Bailey, A., L., *Generalized Theory of Credibil*ity, Proc. Casualty Actuarial Soc. 32, 13-20, **1945.**

[15] Bühlmann, H., Experience Rating and Credibility, *I. ASTIN Bulletin*, 4(3), 199–207, **1967.**

[16] Bühlmann, H., Experience Rating and Credibility, *II. ASTIN Bulletin*, 5(2), 157–165, **1969.**

[17] Bühlmann, H., Straub, E, Glaubwürdigkeit für Schadensätze, Mitteilungen der Vereinigung, *Schweizerischer Versicherungsmathematiker*, 70 (1), **1970.**

[18] Hachemeister, C.A., *Credibility For Regression Models With Application To Trend, Credibility, Theory and Application*, Academic Press, New York, 129–163, **1975.**

[19] Jewell, W.S., The Use of Collateral Data in Credibility Theory: Ahierarchical Model, *Giornale dell' Instituto Italiano degli Attuari*, 38,1–16, **1975.**

[20] Norberg, R., *Credibility Theory*, Encyclopedia of Actuarial Science, **2004.**

[21] Dannenburg, D.,R., Kaas, R., Goovaerts, M., J., *Practical Actuarial Credibility Models*, Institute of Actuarial Science and Econometrics, University of Amsterdam, The Netherlands, **1996.**

[22] Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M., *Modern Actuarial Risk Theory: Using R*, Springer Verlang Inc., Berlin, 381, **2008.**

[23] Campbell, M., An İntegrated System For Estimating The Risk Premium Of İndividual Car Models in Motor İnsurance, *ASTIN Bulletin*,16(2), 165-183, **1986.**

[24] Sundt, B., Two Credibility Regression Approaches for the Classification of Passenger Cars in a Multiplicative Tariff, *ASTIN Bulletin*, 17(1), 42-70, **1987.**

[25] Schmitter, H., The Sample Size Needed For The Calculation Of A GLM Tariff, *ASTIN Bulletin*, 34(1), 249–262, **2004.**

[26] Mahler, H.,C., Dean, C.G., *"Chapter 8: Credibility"*, In Foundations of Casualty Actuarial Science, 4th Edition, Arlington, VA: Casualty Actuarial Society, 485-659, **2005.**

[27] Şentürk, A., *Kredibilite Teorisinde Panel Veri Modelleri*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 97, **2010.**

[28] Nelder, J.A., Verall, R.J, Credibility Theory and Generalized Linear Models, *ASTIN Bulletin*, 27(1), 71-82, **1997.**

[29] Garrido, J., Zhou, J., Full Credibility with Generalized Linear and Mixed Models, *ASTIN Bulletin*, 39(1), 61-80, **2009.**

[30] Zhou, J., *Theory and Applications of Generalized Linear Models in Insurance*, Phd Thesis, ConcordiaUniversity, Canada, **2011.**

[31] Klinker, F., *Generalized Linear Mixed models for Ratemaking: A means of Introducing Credibility into a Generalized Linear Model Setting*, Casualty Actuarial Society E-Forum, 2, **2011.**

[32] Fress, E., W., *Regression Modeling With Actuarial and Financial Applications*, Cambridge University Press, New York, 565, **2010.**

[33] Ohlsson, E., Johanson B., *Combining Credibility and GLM for Rating of Multi-Level Factors*, CAS 2004 Discussion Paper Program, http://www.csact.org/pubs/dpp/ppp04, **2004.**

[34] Dobson A.J., *An Introduction to Generalized Linear Models*, Second Edition, Chapman and Hall/CRC , London, **2002.**

[35] Schirmacher, E., GLM II: Basic Modelling Strategy, Liberty Mutual Group, Casualty Actuarial Society, *Ratemaking and Product Development Seminar*, March 15-17, Chicago, IL, **2010.**

[36] Rodriguez, G., Introducing R: Generalized Linear Model, http://data.princeton.edu/R/ (Mayıs, **2014**).

[37] Tüzel, S., *Sıfır Yığılmalı Kesikli Modeller,* Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 97, **2010.**

[38] Denuit, M., Maréchal, X., Pitrebois, S., Walhin, *J., Actuarial Modelling of Claim Counts*, John Wiley&Sons, **2007.**

[39] Ohlsson, E., Johansson, B., *Non-Life Insurance Pricing with Generalized Linear Models*, Springer, 174, **2010.**

[40] Antonio, K., Beirlant J., Risk Classification in NonLife Insurance, *Proceedings of the 4th Actuarial and Financial Mathematics Day*, Brussels, **2006.**

[41] Koç, T., Cengiz, M. A., Genelleştirilmiş Linear Karma Modellerde Tahmin Yöntemlerinin Uygulamalı Karşılaştırılması, *Karaelmas Fen ve Mühendislik Dergisi*, 2(2), 47-52, **2012.**

[42] Klugman, S.,A., Panjer H., Willmot G.,E., *Loss Models From Data to Decisions*, John Wiley&Sons, New Jersey, 726, **2008.**

[43] Herzog, T., N., *Credibility Theory, ACTEX*, Third Edition, 273, **1999.**

[44] Bühlmann H., Gisler A., *A Course in Credibility Theory and its Applications*, Springer, 327, **2005.**

[45] Belhadj H., Goulet V., Ouellet T., On Parameter Estimation in Hierarchical Credibility, *ASTIN Bulletin*, 39(2), 495-514, **2009.**

[46] IMB SPSS Statistics for Windows, Version 21.0., Armonk, NY, IBM Corp., Released **2012.**

[47] Lemaire, J., *Automobile Insurance Actuarial Models*, Boston: Kluwer-Nijhoff, **1985.**

EKLER

Ek 1. 2013 yılı Türkiye Büyükşehir Belediyeleri ve Nüfus Yoğunlukları

EK 2. Yaş, Şehir, Cinsiyet ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-1 için Poliçe ve Hasar Sayıları

EK 3. Yaş, Şehir ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-2 için Poliçe ve Hasar Sayıları

EK 4. Yaş, Şehir ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-3 için Poliçe ve Hasar Sayıları

EK 5. Yaş, Motor Hacmi ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-4 için Poliçe ve Hasar Sayıları

EK 6. Motor Hacmi, Cinsiyet ve Şehir Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-5 için Poliçe ve Hasar Sayıları

EK 7. Şehir ve Yaş Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-6 için Poliçe ve Hasar Sayıları

EK 8. Yaş ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-7 için Poliçe ve Hasar Sayıları

EK 9. Yaş ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-8 için Poliçe ve Hasar Sayıları

EK 10. Şehir ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-9 için Poliçe ve Hasar Sayıları

EK 11. Şehir ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-10 için Poliçe ve Hasar Sayıları

EK 12. Motor Hacmi ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-11 için Poliçe ve Hasar Sayıları

EK 13. Yaş Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-12 için Poliçe ve Hasar Sayıları

EK 14. Şehir Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-13 için Poliçe ve Hasar Sayıları

EK 15. Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-14 için Poliçe ve Hasar Sayıları

EK 16. Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-15 için Poliçe ve Hasar Sayıları

Ek 17. Model Parametre Tahminleri ve Fisher Adım Sayıları

Ek 18. Modeller için Akaike Bilgi Kriteri, Bayesci Bilgi Kriteri ve Log-Olabilirlik Değeri

EK 1. 2013 yılı Türkiye Büyükşehir Belediyeleri ve Nüfus Yoğunlukları[1,2]

İl Plaka Kodu	İl	Nüfus Yoğunluğu
01	Adana	2.149.260
06	Ankara	5.045.083
07	Antalya	2.158.265
09	Aydın	1.020.957
10	Balıkesir	1.162.761
16	Bursa	1.162.761
20	Denizli	963.464
21	Diyarbakır	1.607.437
25	Erzurum	766.729
26	Eskişehir	799.724
27	Gaziantep	1.844.438
31	Hatay	1.503.066
33	Mersin	1.705.774
34	İstanbul	14.160.467
35	İzmir	4.061.074
38	Kayseri	1.295.355
41	Kocaeli	1.676.202
42	Konya	2.079.225
44	Malatya	762.538
45	Manisa	1.359.463
46	Kahramanmaraş	1.075.706
47	Mardin	779.738
48	Muğla	866.665
52	Ordu	731.452
54	Sakarya	917.373
55	Samsun	1.261.810
59	Tekirdağ	874.475
61	Trabzon	758.237
63	Şanlıurfa	1.801.980
65	Van	1.070.113

TBMM'nin çıkardığı bir yasa[2] ile nüfusu 750.000'i aşan iller Büyükşehir Belediyesi olmuştur. Bu ayırımda Ordu nüfus koşulunu tam olarak sağlamasa da Büyükşehir yapılmıştır.

Kaynaklar:

[1] Türkiye İstatistik Kurumu (TÜİK) 2013 Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları (ADNKS) (http://rapor.tuik.gov.tr/reports/rwservlet?adnksdb2&ENVID=adnksdb2Env&report=wa_buyukbelediye.RDF&p_kod=1&p_yil=2013&p_dil=1&desformat=html

[2] TBMM

(http://www.tbmm.gov.tr/kanunlar/k6360.html)

EK 2. Yaş, Şehir, Cinsiyet ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-1 için Poliçe ve Hasar Sayıları

Model-1 toplam 36 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Şehir**	**Cinsiyet**	**Motor Hacmi**
Sınıf 1	1734	1792	Genç(1)	Büyük(1)	Kadın(1)	Küçük(1)
Sınıf 2	235	137	Genç(1)	Büyük(1)	Kadın(1)	Orta(2)
Sınıf 3	57	19	Genç(1)	Büyük(1)	Kadın(1)	Büyük(3)
Sınıf 4	2808	2935	Genç(1)	Büyük(1)	Erkek(2)	Küçük(1)
Sınıf 5	660	354	Genç(1)	Büyük(1)	Erkek(2)	Orta(2)
Sınıf 6	141	87	Genç(1)	Büyük(1)	Erkek(2)	Büyük(3)
Sınıf 7	150	95	Genç(1)	Küçük(2)	Kadın(1)	Küçük(1)
Sınıf 8	16	6	Genç(1)	Küçük(2)	Kadın(1)	Orta(2)
Sınıf 9	2	0	Genç(1)	Küçük(2)	Kadın(1)	Büyük(3)
Sınıf 10	346	307	Genç(1)	Küçük(2)	Erkek(2)	Küçük(1)
Sınıf 11	95	51	Genç(1)	Küçük(2)	Erkek(2)	Orta(2)
Sınıf 12	38	7	Genç(1)	Küçük(2)	Erkek(2)	Büyük(3)
Sınıf 13	10789	7931	Orta(2)	Büyük(1)	Kadın(1)	Küçük(1)
Sınıf 14	2725	1339	Orta(2)	Büyük(1)	Kadın(1)	Orta(2)
Sınıf 15	724	315	Orta(2)	Büyük(1)	Kadın(1)	Büyük(3)
Sınıf 16	18418	13665	Orta(2)	Büyük(1)	Erkek(2)	Küçük(1)
Sınıf 17	6068	2823	Orta(2)	Büyük(1)	Erkek(2)	Orta(2)
Sınıf 18	1741	595	Orta(2)	Büyük(1)	Erkek(2)	Büyük(3)
Sınıf 19	968	543	Orta(2)	Küçük(2)	Kadın(1)	Küçük(1)
Sınıf 20	231	104	Orta(2)	Küçük(2)	Kadın(1)	Orta(2)
Sınıf 21	30	9	Orta(2)	Küçük(2)	Kadın(1)	Büyük(3)
Sınıf 22	2499	1341	Orta(2)	Küçük(2)	Erkek(2)	Küçük(1)
Sınıf 23	972	384	Orta(2)	Küçük(2)	Erkek(2)	Orta(2)
Sınıf 24	385	82	Orta(2)	Küçük(2)	Erkek(2)	Büyük(3)
Sınıf 25	6700	4465	Yaşlı(3)	Büyük(1)	Kadın(1)	Küçük(1)
Sınıf 26	2013	975	Yaşlı(3)	Büyük(1)	Kadın(1)	Orta(2)
Sınıf 27	602	278	Yaşlı(3)	Büyük(1)	Kadın(1)	Büyük(3)
Sınıf 28	14041	8336	Yaşlı(3)	Büyük(1)	Erkek(2)	Küçük(1)
Sınıf 29	5877	2499	Yaşlı(3)	Büyük(1)	Erkek(2)	Orta(2)
Sınıf 30	2044	672	Yaşlı(3)	Büyük(1)	Erkek(2)	Büyük(3)

Sınıf 31	485	239	Yaşlı(3)	Küçük(2)	Kadın(1)	Küçük(1)
Sınıf 32	85	31	Yaşlı(3)	Küçük(2)	Kadın(1)	Orta(2)
Sınıf 33	12	1	Yaşlı(3)	Küçük(2)	Kadın(1)	Büyük(3)
Sınıf 34	1901	934	Yaşlı(3)	Küçük(2)	Erkek(2)	Küçük(1)
Sınıf 35	800	294	Yaşlı(3)	Küçük(2)	Erkek(2)	Orta(2)
Sınıf 36	409	58	Yaşlı(3)	Küçük(2)	Erkek(2)	Büyük(3)

EK 3. Yaş, Şehir ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-2 için Poliçe ve Hasar Sayıları

Model-2 toplam 12 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Şehir**	**Cinsiyet**
Sınıf 1	2026	1948	Genç(1)	Büyük(1)	Kadın(1)
Sınıf 2	3609	3376	Genç(1)	Büyük(1)	Erkek(2)
Sınıf 3	168	101	Genç(1)	Küçük(2)	Kadın(1)
Sınıf 4	479	365	Genç(1)	Küçük(2)	Erkek(2)
Sınıf 5	14238	9585	Orta(2)	Büyük(1)	Kadın(1)
Sınıf 6	26227	17083	Orta(2)	Büyük(1)	Erkek(2)
Sınıf 7	1229	656	Orta(2)	Küçük(2)	Kadın(1)
Sınıf 8	3856	1807	Orta(2)	Küçük(2)	Erkek(2)
Sınıf 9	9315	5718	Yaşlı(3)	Büyük(1)	Kadın(1)
Sınıf 10	21962	11507	Yaşlı(3)	Büyük(1)	Erkek(2)
Sınıf 11	582	271	Yaşlı(3)	Küçük(2)	Kadın(1)
Sınıf 12	3110	1286	Yaşlı(3)	Küçük(2)	Erkek(2)

EK 4. Yaş, Şehir ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-3 için Poliçe ve Hasar Sayıları

Model-3 toplam 18 tane risk sınıfından oluşmaktadır. Sınıf-1'de hasar sayısı toplamının poliçe sayısından fazla olmasının sebebi "0" hasar oranının az olup, büyük hasarların fazla gerçekleşmesidir.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Şehir**	**Motor Hacmi**
Sınıf 1	4542	4727	Genç(1)	Büyük(1)	Küçük(1)
Sınıf 2	895	491	Genç(1)	Büyük(1)	Orta(2)
Sınıf 3	198	106	Genç(1)	Büyük(1)	Büyük(3)
Sınıf 4	496	402	Genç(1)	Küçük(2)	Küçük(1)
Sınıf 5	111	57	Genç(1)	Küçük(2)	Orta(2)
Sınıf 6	40	7	Genç(1)	Küçük(2)	Büyük(3)
Sınıf 7	29207	21596	Orta(2)	Büyük(1)	Küçük(1)
Sınıf 8	8793	4162	Orta(2)	Büyük(1)	Orta(2)
Sınıf 9	2465	910	Orta(2)	Büyük(1)	Büyük(3)
Sınıf 10	3467	1884	Orta(2)	Küçük(2)	Küçük(1)
Sınıf 11	1203	488	Orta(2)	Küçük(2)	Orta(2)
Sınıf 12	415	91	Orta(2)	Küçük(2)	Büyük(3)
Sınıf 13	20741	12801	Yaşlı(3)	Büyük(1)	Küçük(1)
Sınıf 14	7890	3474	Yaşlı(3)	Büyük(1)	Orta(2)
Sınıf 15	2646	950	Yaşlı(3)	Büyük(1)	Büyük(3)
Sınıf 16	2386	1173	Yaşlı(3)	Küçük(2)	Küçük(1)
Sınıf 17	885	325	Yaşlı(3)	Küçük(2)	Orta(2)
Sınıf 18	421	59	Yaşlı(3)	Küçük(2)	Büyük(3)

EK 5. Yaş, Motor Hacmi ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-4 için Poliçe ve Hasar Sayıları

Model-4 toplam 18 tane risk sınıfından oluşmaktadır. Bazı risk sınıflarında hasar sayısı toplamının poliçe sayısından fazla olmasının sebebi “0” hasar oranının az olup, büyük hasarların fazla gerçekleşmesidir.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Motor Hacmi**	**Cinsiyet**
Sınıf 1	1884	1887	Genç(1)	Küçük(1)	Kadın(1)
Sınıf 2	3154	3242	Genç(1)	Küçük(1)	Erkek(2)
Sınıf 3	251	143	Genç(1)	Orta(2)	Kadın(1)
Sınıf 4	755	405	Genç(1)	Orta(2)	Erkek(2)
Sınıf 5	59	19	Genç(1)	Büyük(3	Kadın(1)
Sınıf 6	179	94	Genç(1)	Büyük(3	Erkek(2)
Sınıf 7	11757	8474	Orta(2)	Küçük(1)	Kadın(1)
Sınıf 8	20917	15006	Orta(2)	Küçük(1)	Erkek(2)
Sınıf 9	2956	1443	Orta(2)	Orta(2)	Kadın(1)
Sınıf 10	7040	3207	Orta(2)	Orta(2)	Erkek(2)
Sınıf 11	754	324	Orta(2)	Büyük(3	Kadın(1)
Sınıf 12	2126	677	Orta(2)	Büyük(3	Erkek(2)
Sınıf 13	7185	4704	Yaşlı(3)	Küçük(1)	Kadın(1)
Sınıf 14	15942	9270	Yaşlı(3)	Küçük(1)	Erkek(2)
Sınıf 15	2098	1006	Yaşlı(3)	Orta(2)	Kadın(1)
Sınıf 16	6677	2793	Yaşlı(3)	Orta(2)	Erkek(2)
Sınıf 17	614	279	Yaşlı(3)	Büyük(3	Kadın(1)
Sınıf 18	2453	730	Yaşlı(3)	Büyük(3	Erkek(2)

EK 6. Motor Hacmi, Cinsiyet ve Şehir Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-5 için Poliçe ve Hasar Sayıları

Model-5 toplam 12 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Motor Hacmi**	**Cinsiyet**	**Şehir**
Sınıf 1	19223	14188	Küçük(1)	Kadın(1)	Büyük(1)
Sınıf 2	1603	877	Küçük(1)	Kadın(1)	Küçük(2)
Sınıf 3	35267	24936	Küçük(1)	Erkek(2)	Büyük(1)
Sınıf 4	4746	2582	Küçük(1)	Erkek(2)	Küçük(2)
Sınıf 5	4973	2451	Orta(2)	Kadın(1)	Büyük(1)
Sınıf 6	332	141	Orta(2)	Kadın(1)	Küçük(2)
Sınıf 7	12605	5676	Orta(2)	Erkek(2)	Büyük(1)
Sınıf 8	1867	729	Orta(2)	Erkek(2)	Küçük(2)
Sınıf 9	1383	612	Büyük(3)	Kadın(1)	Büyük(1)
Sınıf 10	44	10	Büyük(3)	Kadın(1)	Küçük(2)
Sınıf 11	3926	1354	Büyük(3)	Erkek(2)	Büyük(1)
Sınıf 12	832	147	Büyük(3)	Erkek(2)	Küçük(2)

EK 7. Şehir ve Yaş Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-6 için Poliçe ve Hasar Sayıları

Model-6 toplam 6 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Şehir**	**Yaş**
Sınıf 1	5635	5324	Büyük(1)	Genç(1)
Sınıf 2	40465	26668	Büyük(1)	Orta(2)
Sınıf 3	31277	17225	Büyük(1)	Yaşlı(3)
Sınıf 4	647	466	Küçük(2)	Genç(1)
Sınıf 5	5085	2463	Küçük(2)	Orta(2)
Sınıf 6	3692	1557	Küçük(2)	Yaşlı(3)

EK 8. Yaş ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-7 için Poliçe ve Hasar Sayıları

Model-7 toplam 6 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Cinsiyet**
Sınıf 1	2194	2049	Genç(1)	Kadın(1)
Sınıf 2	4088	3741	Genç(1)	Erkek(2)
Sınıf 3	15467	10241	Orta(2)	Kadın(1)
Sınıf 4	30083	18890	Orta(2)	Erkek(2)
Sınıf 5	9897	5989	Yaşlı(3)	Kadın(1)
Sınıf 6	25072	12793	Yaşlı(3)	Erkek(2)

EK 9. Yaş ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-8 için Poliçe ve Hasar Sayıları

Model-8 toplam 9 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**	**Motor Hacmi**
Sınıf 1	5038	5129	Genç(1)	Küçük(1)
Sınıf 2	1006	548	Genç(1)	Orta(2)
Sınıf 3	238	113	Genç(1)	Büyük(3
Sınıf 4	32674	23480	Orta(2)	Küçük(1)
Sınıf 5	9996	4650	Orta(2)	Orta(2)
Sınıf 6	2880	1001	Orta(2)	Büyük(3
Sınıf 7	23127	13974	Yaşlı(3)	Küçük(1)
Sınıf 8	8775	3799	Yaşlı(3)	Orta(2)
Sınıf 9	3067	1009	Yaşlı(3)	Büyük(3

EK 10. Şehir ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-9 için Poliçe ve Hasar Sayıları

Model-9 toplam 4 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Şehir**	**Cinsiyet**
Sınıf 1	25579	17251	Büyük(1)	Kadın(1)
Sınıf 2	51798	31966	Büyük(1)	Erkek(2)
Sınıf 3	1979	1028	Küçük(2)	Kadın(1)
Sınıf 4	7445	3458	Küçük(2)	Erkek(2)

EK 11. Şehir ve Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-10 için Poliçe ve Hasar Sayıları

Model-10 toplam 6 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Şehir**	**Motor Hacmi**
Sınıf 1	54490	39124	Büyük(1)	Küçük(1)
Sınıf 2	17578	8127	Büyük(1)	Orta(2)
Sınıf 3	5309	1966	Büyük(1)	Büyük(3)
Sınıf 4	6349	3459	Küçük(2)	Küçük(1)
Sınıf 5	2199	870	Küçük(2)	Orta(2)
Sınıf 6	876	157	Küçük(2)	Büyük(3)

EK 12. Motor Hacmi ve Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-11 için Poliçe ve Hasar Sayıları

Model-11 toplam 6 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Motor Hacmi**	**Cinsiyet**
Sınıf 1	20826	15065	Küçük(1)	Kadın(1)
Sınıf 2	40013	27518	Küçük(1)	Erkek(2)
Sınıf 3	5305	2592	Orta(2)	Kadın(1)
Sınıf 4	14472	6405	Orta(2)	Erkek(2)
Sınıf 5	1427	622	Büyük(3)	Kadın(1)
Sınıf 6	4758	1501	Büyük(3)	Erkek(2)

EK 13. Yaş Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-12 için Poliçe ve Hasar Sayıları

Model-12 toplam 3 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Yaş**
Sınıf 1	6282	5790	Genç(1)
Sınıf 2	45550	29131	Orta(2)
Sınıf 3	34969	18782	Yaşlı(3)

EK 14. Şehir Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-13 için Poliçe ve Hasar Sayıları

Model-13 toplam 2 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Şehir**
Sınıf 1	77377	49217	Büyük(1)
Sınıf 2	9424	4486	Küçük(2)

EK 15. Cinsiyet Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-14 için Poliçe ve Hasar Sayıları

Model-14 toplam 2 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Cinsiyet**
Sınıf 1	27558	18279	Kadın(1)
Sınıf 2	59243	35424	Erkek(2)

EK 16. Motor Hacmi Açıklayıcı Değişkenlerine Göre Sınıflandırılmış Model-15 için Poliçe ve Hasar Sayıları

Model-15 toplam 3 tane risk sınıfından oluşmaktadır.

	Poliçe Sayısı	**Hasar Sayısı**	**Motor Hacmi**
Sınıf 1	60839	42583	Küçük(1)
Sınıf 2	19777	8997	Orta(2)
Sınıf 3	6185	2123	Büyük(3)

Ek 17. Model Parametre Tahminleri ve Fisher Adım Sayıları

	Parametre Tahminleri				
		Tahmin	**Std. Hata**	**z değeri**	**p**
Model-1	Sabit	-0,327077	0,007271	-44,984	< 0,01
	Yaş (Genç)	0,327829	0,014405	22,757	< 0,01
	Yaş (Yaşlı)	-0,148196	0,009383	-15,793	< 0,01
	Şehir (Küçük)	-0,268841	0,015635	-17,195	< 0,01
	Cinsiyet (Kadın)	0,044331	0,009163	4,838	< 0,01
	Motor Hacmi (Büyük)	-0,664031	0,022295	-29,783	< 0,01
	Motor Hacmi (Orta)	-0,403867	0,011647	-34,676	< 0,01
	Fisher Adım Sayısı: 4				
		Tahmin	**Std. Hata**	**z değeri**	**p**
Model-2	Sabit	-0,445698	0,006898	-64,611	< 0,01
	Yaş (Genç)	0,362573	0,014389	25,197	< 0,01
	Yaş (Yaşlı)	-0,171809	0,009373	-18,330	< 0,01
	Şehir (Küçük)	-0,280902	0,015637	-17,963	< 0,01
	Cinsiyet (Kadın)	0,075274	0,009147	8,229	< 0,01
	Fisher Adım Sayısı: 3				
		Tahmin	**Std. Hata**	**z değeri**	**p**
Model-3	Sabit	-0,310321	0,006366	-48,74	< 0,01
	Yaş (Genç)	0,328058	0,014405	22,77	< 0,01
	Yaş (Yaşlı)	-0,150471	0,009371	-16,06	< 0,01
	Şehir (Küçük)	-0,273791	0,015601	-17,55	< 0,01
	Motor Hacmi (Büyük)	-0,668418	0,022277	-30,00	< 0,01
	Motor Hacmi (Orta)	-0,407004	0,011628	-35,00	< 0,01
	Fisher Adım Sayısı: 4				
		Tahmin	**Std, Hata**	**z değeri**	**p**
Model-4	Sabit	-0,356923	0,007111	-50,195	< 0,01
	Yaş (Genç)	0,329574	0,014405	22,879	< 0,01
	Yaş (Yaşlı)	-0,146087	0,009383	-15,570	< 0,01
	Cinsiyet (Kadın)	0,055998	0,009145	6,124	< 0,01
	Motor Hacmi (Büyük)	-0,671916	0,022293	-30,140	< 0,01
	Motor Hacmi (Orta)	-0,404829	0,011647	-34,758	< 0,01
	Fisher Adım Sayısı: 4				
		Tahmin	**Std, Hata**	**z değeri**	**p**
Model-5	Sabit	-0,351538	0,006029	-58,308	< 0,01
	yşehir5küçük	-0,267130	0,015634	-17,086	< 0,01
	ycinsiyet5kadın	0,055661	0,009150	6,083	< 0,01
	Motor Hacmi (Büyük)	-0,697462	0,022264	-31,327	< 0,01
	Motor Hacmi (Orta)	-0,425093	0,011623	-36,573	< 0,01
	Fisher Adım Sayısı: 4				
		Tahmin	**Std, Hata**	**z değeri**	**p**
Model-6	Sabit	-0,418499	0,006010	-69,63	< 0,01
	Şehir (Küçük)	-0,290038	0,015597	-18,60	< 0,01
	Yaş (Genç)	0,363215	0,014389	25,24	< 0,01
	Yaş (Yaşlı)	-0,176123	0,009358	-18,82	< 0,01
	Fisher Adım Sayısı: 3				

Model		Tahmin	Std, Hata	z değeri	p
Model-7	Sabit	-0,477994	0,006716	-71,177	< 0,01
	Cinsiyet (Kadın)	0,088636	0,009124	9,715	< 0,01
	Yaş (Genç)	0,364581	0,014389	25,337	< 0,01
	Yaş (Yaşlı)	-0,169462	0,009373	-18,080	< 0,01
	Fisher Adım Sayısı: 3				
Model-8	Sabit	-0,336324	0,006232	-53,97	< 0,01
	Motor Hacmi (Büyük)	-0,677894	0,022271	-30,44	< 0,01
	Motor Hacmi (Orta)	-0,408892	0,011628	-35,16	< 0,01
	Yaş (Genç)	0,329919	0,014405	22,90	< 0,01
	Yaş (Yaşlı)	-0,148965	0,009371	-15,90	< 0,01
	Fisher Adım Sayısı: 3				
Model-9	Sabit	-0,483211	0,005530	-87,377	< 0,01
	Cinsiyet (Kadın)	0,090294	0,009132	9,888	< 0,01
	Şehir (Küçük)	-0,278737	0,015638	-17,824	< 0,01
	Fisher Adım Sayısı: 3				
Model-10	Sabit	-0,33149	0,00501	-66,17	< 0,01
	Motor Hacmi (Büyük)	-0,70333	0,02224	-31,62	< 0,01
	Motor Hacmi (Orta)	-0,42917	0,01160	-36,99	< 0,01
	Şehir (Küçük)	-0,27341	0,01560	-17,53	< 0,01
	Fisher Adım Sayısı: 4				
Model-11	Sabit	-0,380348	0,005844	-65,088	< 0,01
	Motor Hacmi (Büyük)	-0,704897	0,022262	-31,663	< 0,01
	Motor Hacmi (Orta)	-0,425799	0,011624	-36,632	< 0,01
	Cinsiyet (Kadın)	0,067360	0,009131	7,377	< 0,01
	Fisher Adım Sayısı: 4				
Model-12	Sabit	-0,447008	0,005859	-76,29	< 0,01
	Yaş (Genç)	0,365452	0,014389	25,40	< 0,01
	Yaş (Yaşlı)	-0,174555	0,009358	-18,65	< 0,01
	Fisher Adım Sayısı: 2				
Model-13	Sabit	-0,452450	0,004508	-100,4	< 0,01
	Şehir (Küçük)	-1,942827	0,015596	-124,6	< 0,01
	Fisher Adım Sayısı: 2				
Model-14	Sabit	-0,514258	0,005313	-96,79	< 0,01
	Cinsiyet (Kadın)	0,103718	0,009107	11,39	< 0,01
	Fisher Adım Sayısı: 2				
Model-15	Sabit	-0,356776	0,004846	-73,62	< 0,01
	Motor Hacmi (Büyük)	-0,712521	0,022238	-32,04	< 0,01
	Motor Hacmi (Orta)	-0,430853	0,011603	-37,13	< 0,01
	Fisher Adım Sayısı: 2				

Ek 18. Modeller için Akaike Bilgi Kriteri, Bayesci Bilgi Kriteri ve Log-Olabilirlik Değeri

	AIC Bilgi Kriteri	**BIC Bilgi Kriteri**	**Log-Olabilirlik Değeri**
Model-1	AIC: 464,8009	BIC: 475,8855	Log-Olabilirlik: -225,4005 (df=7)
Model-2	AIC: 173,7093	BIC: 176,1338	Log-Olabilirlik: -81,85465 (df=5)
Model-3	AIC: 265,3366	BIC: 270,6788	Log-Olabilirlik: -126,6683 (df=6)
Model-4	AIC: 301,7589	BIC: 307,1011	Log-Olabilirlik: -144,8795 (df=6)
Model-5	AIC: 180,0554	BIC: 182,4800	Log-Olabilirlik: -85,02771 (df=5)
Model-6	AIC: 70,49207	BIC: 69,65911	Log-Olabilirlik: -31,24604 (df=4)
Model-7	AIC: 113,8800	BIC: 113,0470	Log-Olabilirlik: -52,93999 (df=4)
Model-8	AIC: 145,3740	BIC: 146,3601	Log-Olabilirlik: -67,68701 (df=5)
Model-9	AIC: 48,95704	BIC: 47,11592	Log-Olabilirlik: -21,47852 (df=3)
Model-10	AIC: 110,5854	BIC: 109,7524	Log-Olabilirlik: -51,29268 (df=4)
Model-11	AIC: 101,5892	BIC: 100,7562	Log-Olabilirlik: -46,79459 (df=4)
Model-12	AIC: 40,29777	BIC: 37,59361	Log-Olabilirlik: -17,14889 (df=3)
Model-13	AIC: 26,88851	BIC: 24,27480	Log-Olabilirlik: -11,44425 (df=2)
Model-14	AIC: 27,96442	BIC: 25,35072	Log-Olabilirlik: -11,98221 (df=2)
Model-15	AIC: 38,93817	BIC: 36,23401	Log-Olabilirlik: -16,46909 (df=3)

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Övgücan Gönenç KARADAĞ

Doğum Yeri : Sivas

Medeni Hali : Bekar

E- Posta : ovgucan@hacettepe.edu.tr

Adresi : Hacettepe Üniversitesi, Aktüerya Bilimleri Bölümü, Beytepe Kampüsü, ANKARA

Eğitim

Lise : 2001-2005 Aydın Süleyman Demirel Anadolu Lisesi

Lisans : 2006-2011 Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü

Yüksel Lisans : 2011-2014 Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü

Yabancı Dil ve Düzeyi

İngilizce, İleri

İş Deneyimi

2013 -… : Hacettepe Üniversitesi, Aktüerya Bilimleri Bölümü, Araştırma Görevlisi

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar