İstatistikçiler Dergisi: İstatistik & Aktüerya

Journal of Statisticians: Statistics and Actuarial Sciences

IDIA 9, 2016, 1, 37-46

Geliş/Received:25.02.2016, Kabul/Accepted: 05.06.2016

www.istatistikciler.org

Araştırma Makalesi / Research Article

Genelleştirilmiş Doğrusal Model Bileşenlerinin Kredibilite Üzerindeki Etkilerinin Araştırılması

Övgücan Karadağ Erdemir

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Çankaya, Ankara, Türkiye
ovgucan@hacettepe.edu.tr

Meral Sucu

Hacettepe Üniversitesi
Aktüerya Bilimleri Bölümü
06800-Çankaya, Ankara, Türkiye
msucu@hacettepe.edu.tr

Öz

Bu çalışmada, genelleştirilmiş doğrusal model (GDM) bileşenlerinin modelin kredibilitesini nasıl etkilediği araştırılmıştır. Bağ fonksiyonu yapısının ve türünün kredibiliteye etkisi incelenmiştir. Bağ fonksiyonunun kredibilite üzerindeki etkilerini görmek amacıyla; logaritmik bağ fonksiyonu yerine birim bağ fonksiyonu ve bir sabit ile çarpılmış bağ fonksiyonu kullanılmıştır. Karşılaştırma işlemi, tam kredibilite standardı yardımıyla tanımlanan asimptotik varyans ve kredibilite olasılığı ile yapılmıştır. Sayısal analizler özel bir sigorta şirketinden alınan bir yıllık kasko sigortası poliçelerinden gelen hasar sayısı verisi kullanılarak yapılmıştır. Analizler sonucunda açıklayıcı değişken seçiminin kredibiliteyi etkilediği, ancak bağ fonksiyonu yapısı ve türünün kredibiliteyi etkilemediği sonucuna ulaşılmıştır.

Anahtar sözcükler: Genelleştirilmiş doğrusal model, tam kredibilite, kredibilite olasılığı, asimptotik varyans, bağ fonksiyonu

Abstract

Investigating the Effects of Generalized Linear Model Components on Credibility

In this study, it has been investigated how the components of generalized linear model (GLM) affect the credibility of model. It is analyzed how the structure and the type of link function affect the credibility. Instead of logarithmic link function, unit link function and link function multiplied by a constant are used in order to see their effects on credibility. The comparison is conducted by asymptotic variance and credibility probability which are defined using full credibility standard. Numerical analyses are carried out by using one-year claim frequency data of motor own damage insurance policies from an insurance company. It is concluded that the selection of explanatory variables have effect on credibility, but the structure and the type of link function has no effect on credibility.

Keywords: *Generalized linear model, full credibility, credibility probability, asymptotic variance, link function*

1. Giriş

Aktüeryal çalışmalarda, özellikle hayat dışı sigorta fiyatlandırmasında, ölümlülük modellemesinde, hasar rezervi hesaplamalarında, risk kabul sürecinde (underwriting), kredi riskinin modellenmesinde ve kredibilite kuramında GDM'lerden sıklıkla yararlanılmaktadır [1, 6]. Sigorta verilerinin istatistiksel analizinde GDM'nin klasik doğrusal modellere tercih edilmesinin birçok nedeni vardır. Bunlar; sigorta verilerinin her zaman sabit varyanslılık, normallik ve ilişkisizlik gibi doğrusal model varsayımlarını sağlamaması olarak sıralanabilir [10]. Sigorta hasar tutarı verisinin dağılımının genellikle sağa çarpık olması nedeniyle simetrik bir dağılım olan normal dağılıma uyum göstermesi zordur.

Model oluşturma işlemi; yanıt değişken için uygun parametrik dağılım, bu parametrik dağılım için uygun bağ fonksiyonu seçimi ile başlayan, açıklayıcı değişkenlerin belirlenmesi ve parametre tahminlerinin yapılması ile devam eden bir süreçtir. Oluşturulan GDM'nin doğruluğunun test edilmesi gerekir. Modelin fiyatlandırmada kullanılabilecek bir model olup olmadığı, diğer bir deyişle güvenilir bir model olup olmadığı kredibilite yaklaşımlarından biri olan tam kredibilite yaklaşımı ile test edilebilir.

Literatürde GDM ile kredibiliteyi ilişkilendiren birçok çalışma vardır. Bu çalışmalardan ilki Nelder ve Verrall [11] tarafından yapılmış ve prim ile hasar rezervlerinin hesaplanmasında kullanılan modeller geliştirilmiştir. Antonie ve Beirlant [2] farklı kredibilite modellerine ilişkin genelleştirilmiş doğrusal karma modeller (GDKM) oluşturarak, bunların aktüerya alanında kullanımı üzerine bir çalışma yapmışlardır. Ohlsson [12] hayat dışı ürünlerin fiyatlandırmasında GDM ile kredibilite kuramını birleştirmiştir. Klinker [8] Bühlmann-Straub kredibilite yaklaşımı ile GDKM'yi birlikte ele almıştır. Bu çalışmalarda GDM'ler ve GDKM'ler kredibilite modellerine dönüştürülerek fiyatlandırmada kullanılmıştır. Bu çalışmalarda genellikle Bühlmann kredibilitesi ile GDKM birlikte ele alınmıştır.

GDM ile sınırlı dalgalanmalı kredibilite yaklaşımını birlikte ele alan çalışmalar da yapılmıştır. İlk çalışma Schmitter [14] tarafından, fiyatlandırmada güvenilir bir tahmin edici elde etmek için gerekli minimum hasar sayısının belirlenmesi amacıyla yapılmıştır. Garrido ve Zhou [5] ise güvenilir bir tahmin edici elde etmek için gerekli minimum gözlem sayısından hareket ederek model bileşenlerinin kredibiliteye etkisini incelemişlerdir. Zhou [15], 2009 yılında Garrido ile yaptığı çalışmaya kısmi kredibilite ve hasar rezervlerinin GDM'ler ile analizini ekleyerek daha kapsamlı bir çalışma sunmuştur. Yukarıda söz edilen çalışmalarda kredibilite kuramı, bir modelin güvenilirliğinin testinde kullanılmaktadır.

Bu çalışmada bir yıllık kasko poliçelerinden gelen hasar sayısı verisi kullanılarak, GDM ile kredibilite kuramı arasında Garrido ve Zhou [5, 15]'nun çalışmalarına benzer bir ilişki kurulup, model bileşenlerinin kredibiliteye etkisi incelenmiştir. Sınıflandırılmış veriyle çalışmak, benzer risklere maruz kalan risk sınıfları açısından adil prim hesabında önemlidir. Bu nedenle, çalışmada bireysel veri yerine belirlenen iki risk faktörüne göre sınıflandırılan veri GDM ile modellenerek, sınırlı dalgalanmalı kredibilite yaklaşımı kullanılmıştır.

Çalışmanın ikinci bölümünde model bileşenleri hakkında bilgi verilip, en çok olabilirlik (EÇO) tahmin edicilerinin asimptotik özelliklerine değinilmiş ve özel bir GDM olan Poisson regresyon verilmiştir. Üçüncü bölümde GDM'lerde sınırlı dalgalanmalı kredibilite yaklaşımı ele alınmış ve karşılaştırma kriterleri tanımlanmıştır. Dördüncü bölümde model bileşenlerinin kredibiliteye etkisi gösterilmiştir. Beşinci bölümde gerçek bir veri kümesi ile sayısal analizler yapılmış ve çıktılar yorumlanmıştır. Son bölümde ise elde edilen sonuçlar verilmiştir.

2. Genelleştirilmiş doğrusal model

2.1. Genelleştirilmiş doğrusal model bileşenleri

GDM, bir bağ fonksiyonu yardımıyla doğrusal tahmin ediciye bağlı olarak kitle ortalamasının hesaplandığı klasik doğrusal modellerin genelleştirilmiş halidir [5]. Bu genelleştirme iki yönlü yapılabilir: Yanıt değişkenin dağılımı için normal dağılım yerine üstel dağılım ailesinden herhangi bir dağılım, bağ fonksiyonu için ise birim bağ fonksiyonu yerine çeşitli bağ fonksiyonlarından biri kullanılabilir. GDM; üstel dağılım ailesinden yanıt değişken, doğrusal bileşen ve bağ fonksiyonu olmak üzere üç temel bileşenden oluşur. θ kanonik parametre, ϕ yayılım parametresi, $c(y,\phi)$ ve $a(\theta)$ bilinen fonksiyonlar olmak üzere üstel dağılım ailesinden yanıt değişkenin olasılık yoğunluk fonksiyonu,

$$f(y;\theta,\phi)=c(y,\phi)\exp\left\{\frac{y\theta-a(\theta)}{\phi}\right\} \tag{1}$$

şeklindedir. i. gözleme ilişkin doğrusal bileşen; açıklayıcı değişken vektörü $X_i=\{x_{i1},x_{i2},...,x_{ip}\}'$ ve parametre vektörü $\beta=\{\beta_1,\beta_2,...,\beta_p\}'$ olmak üzere,

$$\eta_i = X_i'\beta=\sum_{j=1}^{p}\beta_j x_{ij} \quad , \quad i=1,2,...,n \tag{2}$$

şeklindedir. k düzeye sahip bir kategorik değişken, k-1 tane (0,1) gösterge değişken ile kodlanır. Belli referans düzeyleri belirlenerek kodlama işlemi yapılır. Bu referans düzeyi veride sıklığı en yüksek olan faktöre göre belirlenir [4]. $E(y_i)=\mu_i$ olmak üzere, i. gözlem için monoton ve türevlenebilir bağ fonksiyonu,

$$g(\mu_i)=\eta_i \tag{3}$$

biçimindedir. Yanıt değişkenin dağılımına göre bağ fonksiyonu; birim, logaritmik, üstel, kakekök ve logit olabilir. Modele ilişkin parametre tahminleri çeşitli yöntemler ile yapılabilmektedir, ancak yanıt değişkenin dağılımı bilindiğinde EÇO yönteminden sıklıkla yararlanılmaktadır [3].

2.2. Model parametrelerinin tahmini ve en çok olabilirlik tahmin edicilerinin asimptotik özellikleri

Üstel dağılım ailesi için EÇO tahmin edicisi; değişmezlik, yansızlık, tutarlılık ve minimum varyanslılık özelliklerini sağlar. Bu özellikler büyük örneklem varsayımı altında $\hat{\beta}$ tahmin edicisi için özelleştirilebilir. $\hat{\beta}$, β parametresinin asimptotik-yansız ($E(\hat{\beta})=\beta,\ n\to\infty$) ve tutarlı tahmin edicisidir. W köşegen ağırlık matrisi olmak üzere, $\hat{\beta}$ tahmin edicisinin varyansı $n\to\infty$ iken $V(\hat{\beta})=\Sigma=X'WX$ 'dır. Asimptotik yansızlık ve minimum varyanslılık özellikleri kullanılarak, $n\to\infty$ durumunda $\hat{\beta}\sim N(\beta,\Sigma)$ varsayılabilir [5].

Poisson regresyon modeli, kesikli verilerin modellenmesinde yaygın olarak kullanılan GDM'lerden biridir.

2.3. Poisson regresyon modeli

Genel olarak Poisson dağılımlı yanıt değişkene sahip GDM'ler Poisson regresyon olarak adlandırılır. Poisson regresyonda çoğunlukla logaritmik bağ fonksiyonu kullanılmakla birlikte, birim veya karekök bağ fonksiyonları da kullanılabilir [13]. Hasar sayısı veya bir risk grubundaki ölen kişi sayısı gibi kesikli verilerin modellenmesinde riske maruz kalan birim sayısı (n) için modele bir düzeltme faktörü eklenir [4]. Bunun nedeni gözlemler arasındaki farkı veya sınıf büyüklüğünü düzeltmektir. Logaritmik bağ fonksiyonu için düzeltme faktörü $\ln n$ 'dir. $y_i\sim Poisson(\mu_i)$, $X_i'=\{x_{i1},x_{i2},...,x_{ip}\}$ ve parametre vektörü $\beta=\{\beta_1,\beta_2,...,\beta_p\}'$ olmak üzere Poisson regresyon modeli,

$$\ln\mu_i=\ln n+X_i'\beta$$

biçiminde tanımlanır [4].

3. Genelleştirilmiş doğrusal model için sınırlı dalgalanmalı kredibilite kuramı

Kredibilite kuramı sigorta ürününün önceki dönem verisinden yararlanarak, gelecek dönemlere ait beklenen hasar ve primin tahmininde kullanılır [9]. Kredibilite kuramı yaklaşımlarından biri olan sınırlı dalgalanmalı kredibilite yaklaşımında sadece sigortalıların gözlem sayısı ile ilgilenilir. Bir sigortalının hasar deneyimi, bir zaman diliminden diğerine çok fazla dalgalanma göstermiyorsa tam kredibilitenin sağlandığı söylenebilir. Tam kredibilitede %100 güven için gerekli olan gözlem sayısı ile ilgilenilir. Tam kredibilite yaklaşımında normal dağılıma yakınsama özelliğinden yararlanılır. Tahmin edilen değer ile gerçek değer arasındaki fark ne kadar az ise tahmin o kadar güvenilirdir. $r > 0$ olacak şekilde bir hata-tahmin toleransı ve $0 < p < 1$ güven olasılığı değerlerine göre, $\overline{X} = \hat{\mu}_i$ ve $E(X_i) = \mu_i$ iken i. risk sınıfı için tam kredibilite standardı ya da tam kredibilite olasılığı olarak adlandırılan güven ölçüsü,

$$\pi_i = P\{|\hat{\mu}_i - \mu_i| \leq r\mu_i\} \geq p, \quad i = 1,2,...,n \tag{4}$$

eşitliği ile ifade edilir.

Bir GDM tahmin edicisi tam kredibilite standardını sağlıyorsa, güvenilir bir tahmin edici olduğu söylenebilir. Bu çalışmada Zhou [15]'nun 2011 yılı çalışmasındakine benzer şekilde tahmin edici için bir güven ölçütü belirlenmiştir. Scmitter [14], Garrido ile Zhou [5] ve Zhou [15],GDM ile tam kredibilite standardı arasında ilişki kurulması için önermeler yapmışlardır. Bu önermelerden yararlanarak, GDM'deki risk sınıflarının güvenilirliğininin testi için karşılaştırma kriterleri olan kredibilite olasılığı ve asimptotik varyansa ilişkin eşitlikler yazılabilir.

3.1. Kredibilite olasılığı

Monoton artan $g(\mu_i) = \eta_i$ bağ fonksiyonuna sahip bir GDM için, tam kredibilite olasılığı, Eşitlik 4 kullanılarak bulunabilir. Her iki tarafın g fonksiyonu alındığında, eşitsizlik yön değiştirmeden kalır ve tam kredibilite ile GDM'nin birlikte ele alınması için ilk adım atılmış olur.

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

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

$$\pi_i = P\left(g[(1-r)\mu_i] - X_i'\beta \leq X_i'\hat{\beta} - X_i'\beta \leq g[(1+r)\mu_i] - X_i'\beta\right) \tag{5}$$

olarak bulunur [5].

Eşitlik 5, Poisson dağılımlı yanıt değişken ve logaritmik bağ fonsiyonu kullanılarak, Poisson regresyon modeli için özelleştirilirse;

$$\pi_i = P\left(\ln[(1-r)\mu_i] - X_i'\beta \leq X_i'\hat{\beta} - X_i'\beta \leq \ln[(1+r)\mu_i] - X_i'\beta\right) \tag{6}$$

elde edilir.

3.2. Asimptotik varyans

$n \to \infty$ için $V(\hat{\beta})$ asimptotik olarak Σ'ya yakınsadığından i. sınıf için asimptotik varyans,

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

şeklinde bulunabilir. $Cov(\hat{\beta}_j, \hat{\beta}_k) = \sigma_{ij}$ ve $\sigma_{ij} = \Sigma$ olduğundan $s_i^2 = \sum_{j=1}^{p}\sum_{k=1}^{p} x_{ij}x_{ik}\sigma_{ij}$ 'dır. Bu durumda $n \to \infty$ iken i. risk grubu için asimptotik varyans,

$$s_i^2 = V(\hat{\mu}_i) = V(X_i'\hat{\beta}) = X_i' \Sigma X_i \tag{7}$$

olarak elde edilir.

Yanıt değişkenin normal dağılıma yakınsadığı varsayımıyla, tam kredibilite olasılığı,

$$\begin{aligned}\pi_i &= P\left\{\frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i} \le \frac{X_i'\hat{\beta} - X_i'\beta}{s_i} \le \frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i}\right\} \\ &= \Phi\left(\frac{g[(1+r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right) - \Phi\left(\frac{g[(1-r)\mu_i] - g(\mu_i)}{s_i^{(g)}}\right)\end{aligned}$$

eşitliği ile yazılır. Bu eşitlikten de görüleceği üzere s_i küçüldükçe π_i büyümektedir. Dolayısıyla, asimptotik varyansı küçük olan ve kredibilite olasılığı büyük olan risk grubunun daha güvenilir olduğu söylenebilir [15].

Poisson regresyon modeli için asimptotik varyans hesaplanırken Eşitlik 7'de varyans-kovaryans matrisi Poisson regresyon modelinden yararlanarak bulunur [5].

4. Genelleştirilmiş doğrusal model bileşenlerinin güvenilirliğe etkisi

GDM bileşenleri olan açıklayıcı değişken ve bağ fonksiyonunun modelin kredibilitesini etkileyip etkilemediği incelenebilir. Model bileşenlerinin dışında örneklem büyüklüğü ve riske maruz birim sayısı gibi büyüklüklerin de kredibilite üzerinde etkisi vardır [7]. Örneklem büyüklüğünün etkisi her bir risk sınıfı için bulunan asimptotik varyans s_i^2 yardımıyla karşılaştırılabilir. Yanıt değişkenin hasar sayısı olarak belirlendiği Poisson regresyonda, küçük hasar sayısına sahip risk grubunda s_i^2 daha büyüktür. Örneklem büyüklüğü arttıkça güvenilirliğin artması beklenen bir sonuçtur.

4.1. Sınıflandırılmış veride açıklayıcı değişkenlerin GDM güvenilirliğine etkisi

Açıklayıcı değişkenlerin kredibiliteye etkisini görmek için farklı açıklayıcı değişkenlere göre hesaplanan s_i^2 ve π_i değerleri karşılaştırılabilir. Örneğin cinsiyet ve motor hacmi açıklayıcı değişkenleri kullanılarak oluşturulmuş bir modelde, cinsiyet açıklayıcı değişkeninin etkisini görmek amacıyla aynı motor hacmi ve farklı cinsiyet açıklayıcı değişkenine sahip olan risk sınıfları karşılaştırılabilir. Böylece açıklayıcı değişkenlerin güvenilirliği nasıl etkilediği incelenebilir.

4.2. Sınıflandırılmış veride bağ fonksiyonunun GDM güvenilirliğine etkisi

Bağ fonksiyonunun GDM'nin kredibilitesine etkisi iki şekilde incelenebilir: Bağ fonksiyonunun yapısı veya türü değiştirilebilir. Yapısı bir sabit ile çarpılarak (bölünerek) veya bir sabit eklenerek (çıkarılarak) değiştirilebilirken; türü logaritmik, birim, logit bağ fonksiyonları gibi fonksiyonlar kullanılarak değiştirilebilir. Garrido ve Zhou [5] çalışmalarında bağ fonksiyonunun türünü değiştirmeyip, yeniden ölçeklendirerek bağ fonsiyonunun etkisini incelemişler ve bağ fonksiyonu seçiminin GDM tahmin edicilerinin güvenilirliğini etkilemediği sonucuna ulaşmışlardır. Bağ fonksiyonu c gibi bir sabit ile çarpılarak $h(\mu_i) = cg(\mu_i)$ gibi yeni bir bağ fonsiyonu elde edilebilir. $g(\mu_i) = \eta_i = X_i'\beta^{(g)}$ olduğundan $h(\mu_i) = cg(\mu_i) = cX_i'\beta^{(g)} = X_i'\beta^{(h)}$ olur. Bu durumda iki farklı bağ fonksiyonuna göre oluşturulmuş modelde parametreler arasında $\beta^{(h)} = c\beta^{(g)}$ şeklinde bir ilişki ortaya çıkar. Bu bağ fonsiyonlarına göre asimptotik varyans değerlerine bakılırsa,

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

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

$$s_i^{(h)} = cs_i^{(g)} \tag{8}$$

eşitlikleri elde edilir. Eşitlik 8 bağ fonksiyonunun bir sabit ile çarpılarak yeniden ölçeklendirilmesinin s_i^2 değerini çarpılan sabit kadar etkilediği sonucunu gösterir. Bağ fonksiyonu değişiminin karşılaştırma kriterinden π_i değerini nasıl etkilediğini görmek için, g ve h bağ fonksiyonlarına göre kredibilite olasılıkları $\pi_i^{(g)}$ ve $\pi_i^{(h)}$,

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

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

elde edilir. $\pi_i^{(h)} = \pi_i^{(g)}$ olduğundan bağ fonksiyonundaki değişimin kredibiliteyi etkilemediği sonucuna ulaşılabilir [5].

Bağ fonksiyonunun türü yanıt değişkenin dağılımına göre şekillenmektedir. Yanıt değişkenin dağılımının değişmesi, bağ fonksiyonunun türünü de etkileyeceğinden, kredibiliteye de etkisinin olması beklenmektedir. Poisson regresyon modeli ile genellikle logaritmik bağ fonksiyonu kullanılmakla birlikte birim ve karakök bağ fonksiyonu da kullanılabilir [4]. Poisson dağılımlı veri için logaritmik bağ fonksiyonuna ve birim bağ fonksiyonuna göre hesaplanmış π_i değerleri karşılaştırılabilir. $g(\mu_i) = \mu_i$ birim bağ fonksiyonu için kredibilite olasılığı,

$$\pi_i^{(birim)} = \Phi\left(\frac{(1+r)\mu_i - \mu_i}{s_i}\right) - \Phi\left(\frac{(1-r)\mu_i - \mu_i}{s_i}\right) = \Phi\left(\frac{r\mu_i}{s_i}\right) - \Phi\left(\frac{-r\mu_i}{s_i}\right) \tag{9}$$

eşitliğinden bulunur. Eşitlik 6'da verilen logaritmik bağ fonksiyonu için kredibilite olasılığı,

$$\pi_i^{(\log)} = \Phi\left(\frac{\ln(1+r)}{s_i}\right) - \Phi\left(\frac{\ln(1-r)}{s_i}\right) \tag{10}$$

elde edilir. Eşitlik 10'da verilen logaritmik bağ fonksiyonu ile elde edilen kredibilite olasılığı ile Eşitlik 9'da verilen birim bağ fonksiyonundan bulunan kredibilite olasılığının farklı sonuçlar doğuracağı açıktır [7].

5. Sayısal Örnek

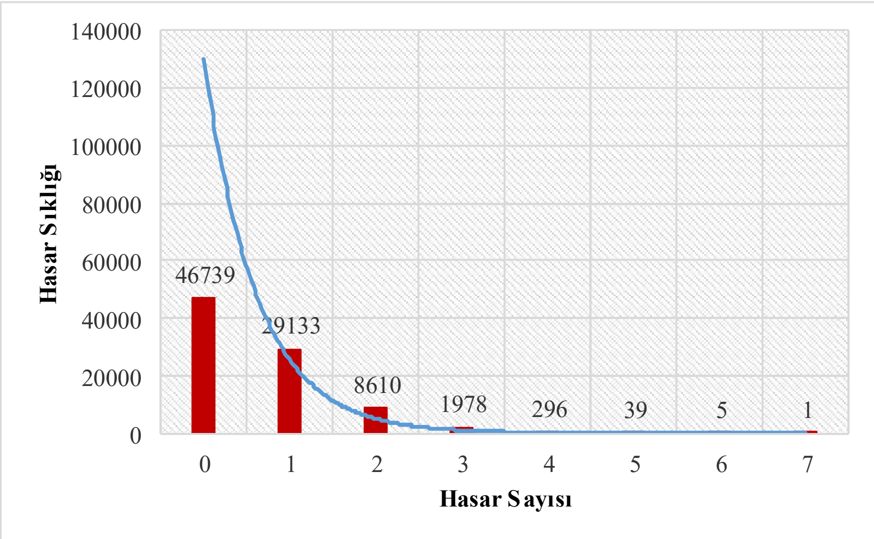
5.1. Veri

Bu çalışmada özel bir sigorta şirketinden alınan; bir yıllık, otomobil sınıfı, 2004 ve üstü model olan 86801 araca ait kasko sigortası verisi ile çalışılmıştır. Örneklemde her sigorta poliçesinin başlangıç ve bitiş tarihleri farklı olmakla birlikte, her bir poliçenin 2013 yılı içinde sonlanan birer yıllık poliçe olması nedeniyle bu farklılık göz ardı edilmiştir. Analizler R.2.13.0 ve IMB SPSS Statistics 21 programları yardımıyla yapılmıştır. Hasar sayısı verisinin Çizelge 1'de verilen istatistiklerinden ortalama ve varyans değerlerine bakıldığında verinin Poisson dağıldığı kabul edilebilir. Kolmogorov-Smirnov testine göre $p = 0{,}802 > \alpha = 0{,}05$ olduğundan, sigortalıların yıllık hasar sayısı yanıt değişkeninin 0,62 parametresi ile Poisson dağıldığı kabul edilmiştir.

Çizelge 1. Hasar sayısı verisinin betimleyici istatistikleri

n	86801
Ortalama	0,620
Standart Hata	0,003
Standart Sapma	0,789
Varyans	0,623
Minimum	0
Maksimum	7

Hasar sayısı dağılımının histogram grafiği Şekil 1'de verilmiştir. Şekil 1'den de görüldüğü üzere, hasar sayılarının dağılımı simetrik değildir. Bu histogram, sigorta hasar verisinde klasik doğrusal modellerin yerine GDM kullanım gerekliliğinin bir göstergesidir.



Şekil 1. Hasar sayısı dağılımı

5.2. Verinin Poisson regresyon ile modellenmesi

Sınıflandırmada kullanılacak açıklayıcı değişkenler, sigortalının cinsiyeti ve aracın motor hacmi olarak belirlenmiştir. Sigortalıların cinsiyeti kadın(K) ve erkek(E) olmak üzere iki düzeyli iken, aracın motor hacmi düşük (<1600 cc), orta (1600-2000 cc), yüksek (>1600 cc) olmak üzere üç düzeylidir. Açıklayıcı değişkenlere göre sınıflandırılmış veri Çizelge 2'de verilmiştir.

Çizelge 2. Motor hacmi ve cinsiyete göre sınıflanmış poliçe ve hasar sayısı verisi

	Poliçe Sayısı	Hasar Sayısı	Motor Hacmi	Cinsiyet
Sınıf 1	20826	15065	Düşük	Kadın
Sınıf 2	40013	27518	Düşük	Erkek
Sınıf 3	5305	2592	Orta	Kadın

Sınıf 4	14472	6405	Orta	Erkek
Sınıf 5	1427	622	Yüksek	Kadın
Sınıf 6	4758	1501	Yüksek	Erkek

Açıklayıcı değişkenlerin kodlaması referans değerlere göre yapılır. Motor hacmi için referans değer düşük, cinsiyet için referans değer erkektir. Her bir 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. x_{i1}'in tüm i değerleri için 1 şeçilmesinin nedeni modele sabit terimin dahil edilmesidir. Sigortalıların yıllık hasar sayısı yanıt değişkeni, Poisson regresyon ile modellenirken logaritmik bağ fonksiyonu kullanılmıştır. Parametre tahminlerinde tüm parametreler anlamlı bulunmuş ve Poisson regresyon modeli,

$$E(Y_i) = \mu_i = \exp(-0{,}380348x_{i1} - 0{,}704897x_{i2} - 0{,}425799x_{i3} - 0{,}067360x_{i4})$$

biçiminde elde edilmiştir. Karşılaştırma kriterlerinin hesaplanmasında kullanılacak varyans kovaryans matrisi, her bir risk sınıfı için ağırlıklar;
$\mathrm{w}_1 = 20826$, $\mathrm{w}_2 = 40013$, $\mathrm{w}_3 = 5305$, $\mathrm{w}_4 = 14472$, $\mathrm{w}_5 = 1427$ ve $\mathrm{w}_6 = 4758$ olmak üzere,

$$\Sigma = \begin{pmatrix} 3{,}4144\mathrm{e}-05 & -2{,}6905\mathrm{e}-05 & -2{,}5748\mathrm{e}-05 & -2{,}9819\mathrm{e}-05 \\ -2{,}6905\mathrm{e}-05 & 4{,}9561\mathrm{e}-04 & 2{,}4210\mathrm{e}-05 & 9{,}5672\mathrm{e}-06 \\ -2{,}5748\mathrm{e}-05 & 2{,}4210\mathrm{e}-05 & 1{,}3511\mathrm{e}-04 & 6{,}3341\mathrm{e}-06 \\ -2{,}9819\mathrm{e}-05 & 9{,}5672\mathrm{e}-06 & 6{,}3341\mathrm{e}-06 & 8{,}3378\mathrm{e}-05 \end{pmatrix}$$

biçimindedir.

5.3. Risk sınıflarının güvenilirlik analizi

Risk sınıflarının güvenilirliği tam kredibilite yaklaşımı kullanılarak karşılaştırılmıştır. Parametrelerin EÇO tahmin edicilerinin asimptotik özellikleri sağladığı varsayılarak karşılaştırma kriterleri hesaplanmıştır. Riske maruz kalan birim sayısı olan poliçe sayıları logaritmaları alınarak modele düzeltme terimi olarak dahil edilmiştir. Böylelikle riske maruz kalan birimlerin etkisi modele yansıtılmıştır. Hata-tahmin tolerans değeri (r) için 0,001; 0,01; 0,1 ve 0,5 gibi değerler denenmiştir. Daha anlamlı sonuçlar vermesinden dolayı r=0,01 olarak seçilmiş ve analizlere bu doğrultuda devam edilmiştir. Açıklayıcı değişken vektörleri, varyans kovaryans matrisi ve belirlenen hata-tahmin tolerans değeri kullanılarak logaritmik bağ fonksiyonuna göre 6 risk sınıfı için hesaplanan asimptotik varyans ve kredibilite olasılıkları Çizelge 3'te verilmiştir.

Çizelge 3. Risk sınıfları için asimptotik varyans ve kredibilite olasılıkları

	Asimptotik Varyans	Kredibilite Olasılığı
Sınıf 1	0,00005	0,81127
Sınıf 2	0,00003	0,91296
Sınıf 3	0,00015	0,57940
Sınıf 4	0,00011	0,64322
Sınıf 5	0,00051	0,33936
Sınıf 6	0,00047	0,35332

Çizelge 3'teki bilgiler kullanılarak risk sınıfları karşılaştırma kriterleri yardımıyla güvenilirliklerine göre

$$s_5^2 > s_6^2 > s_3^2 > s_4^2 > s_1^2 > s_2^2$$

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

biçiminde sıralanabilir. Bu nedenle en güvenilir risk sınıfı, düşük motor hacimli araca sahip erkek sigortalılardan oluşan ikinci risk sınıfı iken, en az güvenilir risk sınıfı ise yüksek motor hacimli araca sahip kadın sigortalılardan oluşan beşinci risk sınıfıdır. GDM bileşenlerinin bu sıralamayı nasıl etkilediği aşağıda incelenmiştir.

5.4. Açıklayıcı değişkenlerin güvenilirliğe etkisi

Açıklayıcı değişkenlerin kredibiliteye etkisini incelemek için cinsiyet ve motor hacmi açıklayıcı değişkenlerine göre asimptotik varyans ve kredibilite olasılığı değerleri karşılaştırılmıştır. Cinsiyet açıklayıcı değişkeninin etkisi, aynı motor hacimli risk sınıfları için (Sınıf 1 ve Sınıf 2, Sınıf 3 ve Sınıf 4, Sınıf 5 ve Sınıf 6); motor hacmi açıklayıcı değişkeninin etkisi ise aynı cinsiyet ve farklı hacimlere ilişkin risk sınıfları (Sınıf 1, Sınıf 3 ve Sınıf 5; Sınıf 2, Sınıf 4 ve Sınıf 6) arasında karşılaştırılarak incelenir. Erkek sigortalıların bulunduğu risk sınıfları daha küçük asimptotik varyans ile daha büyük kredibilite olasılığına sahiptir. Üçüncü ve dördüncü risk sınıfları karşılaştırıldığında, $s_3^2 = 0{,}0001541717 > s_4^2 = 0{,}0001177631$ ve $\pi_3 = 0{,}5794078 < \pi_4 = 0{,}6432215$ sonuçlarına göre, orta motor hacimli araca sahip kadın sigortalıların hasar bilgisinin, orta motor hacimli araca sahip erkek sigortalıların hasar bilgisinden daha güvenilir olduğu sonucuna ulaşılabilir.

5.5. Bağ fonksiyonunun güvenilirliğe etkisi

Bağ fonksiyonu türlerinin kredibilite üzerindeki etkisinin analizinde, Eşitlik 9 ve Eşitlik 10 ile verilen kredibilite olasılıkları hesaplanmış ve karşılaştırılmıştır. Çizelge 4'ten logaritmik bağ fonksiyonuna göre oluşturulan modeldeki risk sınıflarının güvenilirlik sıralamasında en güvenilir risk grubu Sınıf 2 iken en az güvenilir sınıfın Sınıf 5 olduğu görülmektedir. Bu sıralama birim bağ fonksiyonuna göre oluşturulan modelde de aynı kalırsa, bağ fonksiyonu türünün kredibiliteyi etkilemediği yorumu yapılabilir.

Çizelge 4. Logaritmik ve birim bağ fonksiyonuna göre kredibilite olasılıkları

	π_i (Logaritmik Bağ Fonk.)	π_i (Birim Bağ Fonk.)
Sınıf 1	0,81127	0,94276
Sınıf 2	0,91296	0,96437
Sınıf 3	0,57940	0,49275
Sınıf 4	0,64322	0,68658
Sınıf 5	0,33936	0,16582
Sınıf 6	0,35332	0,44790

Çizelge 4'e göre logaritmik ve birim bağ fonksiyonuna göre güvenilirlilik sıralaması;

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

şeklindedir. Her iki bağ fonksiyonuna göre yapılan sıralamanın aynı olması nedeniyle bağ fonksiyonu türünün kredibiliteyi etkilemediği söylenebilir.

Örneklem büyüklüğünün kredibiliteye etkisi, aynı poliçe sayısına sahip risk sınıflarındaki hasar sayılarına göre hesaplanan asimptotik varyans ve kredibilite olasılığı karşılaştırması ile yapılabilir. Ancak çalışılan veri setinde poliçe sayıları diğer bir ifade ile riske maruz kalan birim sayısı aynı 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 orta motor hacimli araca sahip erkek sigortalıların, yüksek motor hacimli araca sahip erkek sigortalılara göre daha küçük asimptotik varyans ve daha büyük kredibilite olasılığına sahip olması beklenir. $s_3^2 = 0{,}0001541717 < s_6^2 = 0{,}0004759506$ ve $\pi_3 = 0{,}5794078 > \pi_6 = 0{,}3533257$ sonuçları bu beklentiyi doğrular.

6. Sonuç ve öneriler

Bu çalışmada, hayat dışı sigortaların istatistiksel analizinde sıklıkla kullanılan GDM ve kredibilite kuramı birlikte ele alınmıştır. GDM yardımıyla özel bir sigorta şirketinden alınan bir yıllık hasar sayısı verisi modellenmiş ve tam kredibilite yaklaşımı kullanılarak güvenilirlik analizi yapılmıştır. Model bileşenlerinin güvenlilirliğe etkisi kredibilite olasılığı ve asimptotik varyans yardımıyla incelenmiş, sonuçlar karşılaştırılmış ve yorumlanmıştır. Model bileşenlerinden doğrusal bileşen içindeki açıklayıcı değişkenin kredibiliteyi etkilediği, ancak bağ fonksiyonu değişiminin kredibiliteyi etkilemediği sonucuna ulaşılmıştır. Bu çalışmanın sınıflandırma ve fiyatlandırma üzerine çalışan araştırmacılara, model bileşenlerinin belirlenmesinde yardımcı olacağı ve daha güvenilir çalışmaların yapılmasına neden olacağı düşünülmektedir.

Kaynaklar

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

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

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

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

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

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

[7] Ö. Karadağ, 2014, Genelleştirilmiş Doğrusal Modeller için Sınırlı Dalgalanmalı Kredibilite Yaklaşımı, Yüksek Lisans Tezi, *Hacettepe Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 99.

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

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

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

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

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

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

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

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