HASAR SIKLIKLARI İÇİN SIFIR YIĞILMALI KESİKLİ MODELLER

ZERO-INFLATED DISCRETE MODELS FOR CLAIM FREQUENCIES

SEMA TÜZEL

Hacettepe Üniversitesi
Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin
AKTÜERYA BİLİMLERİ Anabilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ
Olarak Hazırlanmıştır.

2011

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **AKTÜERYA BİLİMLERİ ANABİLİM DALI 'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan	: .... Doç. Dr. Durdu KARASOY
Üye (Danışman)	: .... Doç. Dr. Meral SUCU
Üye	: .... Dr. Yasemin GENÇTÜRK

ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından ...../...../........ tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca ...../...../......... tarihinde kabul edilmiştir.

Prof.Dr. Adil DENİZLİ
Fen Bilimleri Enstitüsü Müdürü

HASAR SIKLIKLARI İÇİN SIFIR YIĞILMALI KESİKLİ MODELLER

Sema Tüzel

ÖZ

Çalışmada, sıfır yığılmalı regresyon modellerinin, Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası'nda belirli bir sürede meydana gelen hasar sayılarının modellenmesi amacıyla kullanımı araştırılmıştır. Sıfır yığılmalı regresyon modelleri, genelleştirilmiş doğrusal modellerin özelleştirilmiş biçimi olup, veri kümesinin sıfır değerinde aşırı yığılma gösterdiği durumda sıfır değeri için yeniden ağırlıklandırma yapılarak elde edilmektedir. Hasar sayısı verisi için sıfır yığılmalı regresyon modelleri, belirli bir sürede hasar gerçekleşmemesi durumu ile, hasar gerçekleşip hasarsızlık indirimi ve muafiyet gibi uygulamalar nedeniyle sigorta şirketine bildirilmemesi durumunun birbirinden ayrı değerlendirilmesinde önemli bir rol oynamaktadır.

Çalışmanın uygulama bölümünde Trafik Sigortaları Bilgi Merkezi'nden (TRAMER) alınan veri kümesi kullanılmıştır. Veri kümesinde, Türkiye'de 2008 yılının Ocak ayında yürürlüğe giren Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası poliçelerine ilişkin yaş, cinsiyet, kullanılan aracın plaka il kodu, kullanım tipi ve bir yıllık sürede gerçekleşen hasar sayısı bilgileri yer almaktadır.

İki farklı değişken kümesi ile regresyon modelleri oluşturulmuştur. Oluşturulan regresyon modelleri model seçim ölçütleri dikkate alınarak karşılaştırılmış ve veri kümesine en uygun model belirlenmiştir. Her iki değişken kümesi ile elde edilen sonuçlar, klasik sayı modellerinin sıfır değerindeki yığılmayı açıklayamadığını, hasar sayılarının modellenmesinde sıfır yığılmalı regresyon modellerinin klasik sayı modellerine tercih edilmesi gerektiğini göstermiştir.

Anahtar Kelimeler: Sıfır Yığılmalı Regresyon Modelleri, Aşırı Yayılım, Hasar Sayısı Dağılımları, Zorunlu Trafik Sigortası, Genelleştirilmiş Doğrusal Modeller.

Danışman: Doç. Dr. Meral SUCU, Hacettepe Üniversitesi, Aktüerya Bilimleri Bölümü.

ZERO-INFLATED DISCRETE MODELS FOR CLAIM FREQUENCIES

Sema Tüzel

ABSTRACT

In this study, the use of zero-inflated regression models to model claim frequencies for motor third party liability insurance is investigated. Zero-inflated regression models are a special type of generalized linear models and obtained by reweighting for zero when data has excess zeros. For claim frequency data, zero-inflated regresssion models play an important role for evaluation seperately no claim occurrence during a period and claim occurred but not reported due to deductible and no claim discount applications.

In application, data taken from Motor Third Party Liability Insurance Center (TRAMER) is used. This data involves information about age, sex, licence number, type of car usage and observed claim frequency over a year for motor third party liability policies written on January 2008 in Turkey.

Two regression models are constructed for two different variable sets. The best (or the most suitable) model is choosen by taking into account model selection criteria. The results obtained for two variable sets show that classical count models do not explain the density accumulation mass at zero value and zero-inflated regression models should be preferred to classical count models while modeling claim count data.

Keywords: Zero-inflated Regression Models, Overdispersion, Claim Count Distributions, Motor Third Party Liability Insurance, Generalized Linear Models.

Advisor: Assoc. Prof. Dr. Meral SUCU, Hacettepe University, Department of Actuarial Sciences.

TEŞEKKÜR

Tez çalışmamın sonuçlandırılması aşamasında yol gösterici olan danışmanım Sayın Doç. Dr. Meral SUCU' ya,

zamansız kaybettiğim, danışmanlığımı yaptığı kısa sürede kendisinden çok şey öğrendiğim, çalışmam boyunca desteğini ve sevgisini hiçbir zaman esirgemeyen, çok değerli hocam Sayın Prof. Dr. Ömer ESENSOY' a,

değerli katkıları ve destekleri için Sayın Dr. Yasemin GENÇTÜRK' e,

çalışma sürecinde hoşgörülerini esirgemeyen değerli hocalarım ve arkadaşlarıma,

zor anlarımda manevi desteklerini içtenlikle hissettiğim ve her zaman yanımda olan çok değerli arkadaşlarım Tuğba TUNÇ, Uğur KARABEY, Şule ŞAHİN, Görkem DENİZCİ, Damla BARLAS ve Mehmet PIRILDAK' a,

üzerimdeki emeklerini hayatım boyunca ödeyemeyeceğim çok sevdiğim annem, babam, kardeşime ve en değerli varlıklarımdan biri olan tatlı köpeğim Spikey'e,

teşekkürlerimi sunarım.

İÇİNDEKİLER DİZİNİ

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

TABLOLAR DİZİNİ

Sayfa

1. GİRİŞ

Türkiye' de, hayat dışı sigortalarının yıllık prim üretimindeki payının, hayat sigortalarına oranla daha yüksek olduğu görülmektedir. Hayat dışı sigortalar içerisinde en yaygın olanları, Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası (Zorunlu Trafik Sigortası) ve Kara Taşıtları Kasko Sigortasıdır. Her iki sigorta türünün de yaygın olarak kullanılması, geçmiş hasar bilgisinin yanı sıra sigortalıların karakteristik özelliklerinin de kayıt altına alınmasıyla çok sayıda gözlem içeren veri tabanlarının oluşturulmasına olanak sağlamaktadır.

Birçok ülkede otomobil sigortasında, prim ile bireysel hasar bilgisi arasında ilişki kurulan sistemler yaygın olarak kullanılmaktadır. Bu sistemlerde, herhangi bir sigorta poliçesi kapsamında bir veya birden fazla kaza gerçekleşmesi durumunda prim artırımı yapılarak sigortalı cezalandırılırken, kaza gerçekleşmemesi durumunda ise prim indirimi yapılarak sigortalı ödüllendirilmektedir. Uygulanma şekline bağlı olarak değişen bu sistemlerden bazıları; hasarsızlık indirimi, deneyim değerlendirmesi (experience rating), liyakat derecelendirmesi (merit rating) ve Bonus-Malus sistemidir (Denuit et al., 2007).

Sigortalı hasar bildiriminde bulunmadığında primden indirim (bonus) aldığı; ancak, bir veya daha çok hasar yapması halinde birikmiş indirimlerinin bir kısmını veya tümünü kaybettiği ve bazen prime ek bir fiyat yüklemesinin (malus) yapıldığı sisteme Bonus-Malus sistemi denir.

Bonus-Malus sistemindeki başlıca amaç, iyi sürücülerin primini azaltmak, nispeten riskli sürücülerin ise primini artırmaktır. Bu nedenle bir kaza olduğunda sigortalının, bir hak talebinde bulunup daha yüksek prim sınıfına geçme veya zararını kendisi karşılayarak indirimini koruma opsiyonu vardır.

Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası ve Kara Taşıtları Kasko Sigortası'nda yaygın olan bir başka uygulama ise muafiyet uygulamasıdır. Düzenlenen herhangi bir poliçede muafiyet koşulu bulunması durumunda, gerçekleşen hasar miktarı önceden belirlenen muafiyet miktarını aşmadığı takdirde sigorta şirketlerinin sigortalıya karşı sorumluluğu bulunmamaktadır.

Sigorta sektöründe fiyatlandırma sisteminin iyi olması; sigortacının, beklenen kayba, masraflara ve beklenmedik durumlar için gerekli önlemlerin alınmasına ilişkin teminatı rahatlıkla verebilmesinde en önemli unsurlardan biridir. Özellikle otomobil sigortalarında fiyatlandırmaya ilişkin temel bileşenlerin iyi belirlenmesi ve bir sonraki hesap yılına ait öngörünün gerçek değerlere olabildiğince yakın sonuçlar vermesi, sigorta şirketlerinin sorumluluklarının daha tutarlı belirlenebilmesinde etkili olmaktadır.

Fiyatlandırmada ilk adım, hasar sayısı dağılımının modellenmesidir. Geleneksel olarak hasar sayısı dağılımının Poisson veya negatif binom (NB) dağılımlarından birine uyması beklenir. Ancak hasarsızlık indirimi, Bonus-Malus sistemi ve muafiyet gibi uygulamalar, değişen veri yapılarının modellenmesi için farklı dağılımların kullanımına duyulan gereksinimi artırmış ve bu doğrultuda çalışmalar yapılmıştır. Sigorta şirketlerinin belirli bir dönemde kayıt altına aldığı hasar sayıları incelendiğinde gözlemlenen sıfır hasar sayısının, Poisson veya negatif binom dağılımları varsayımı altında beklenen sıfır hasar sayısından fazla olduğu görülmektedir. Bunun nedeni; sigortalıların, maddi kaybın nispeten az olduğu küçük kaza veya kazalar yapması durumunda sigorta şirketinden tazminat talebinde bulunmayarak prim indiriminden yararlanmaya devam etmek istemesi veya önceden belli bir muafiyet miktarı belirlenmişse oluşan maddi kayıp bu muafiyet miktarını aşmadığında sigorta şirketinin sigortalıya karşı herhangi bir sorumluluğunun bulunmamasıdır.

Belirli bir dönemde yapılan gözlem sonucu elde edilen verinin sıfır değerinde aşırı yoğunluk göstererek Poisson veya negatif binom dağılımlarına uygunluk varsayımı altında beklenen sıklıktan fazla olduğu durumlarda kullanılmak üzere sıfır yığılmalı dağılımlar geliştirilmiş ve bu dağılımlar başta aktüerya ve sigorta olmak üzere; üretim (Lambert, 1992), sosyoloji (Land et al., 1996), ekonometri (Mullahy, 1986), ekoloji (Welsh et al., 1996; Faddy, 1998), tarım ve tıp (Ridout et al., 1998) gibi birçok alanda kullanılmıştır.

Hasar sayısının modellenmesi amacıyla kullanılması gereken açıklayıcı değişken sayısının, sigortalıların bireysel özelliklerini yansıtacak ve bu özelliklerin hasar sayısı üzerindeki etkisinin incelenmesine olanak sağlayacak şekilde belirlenmesi uygun olacaktır. Çok sayıda açıklayıcı değişken belirlenerek bağımlı değişken üzerindeki etkilerinin analiz edilmesi amacıyla genelleştirilmiş doğrusal regresyon modelleri geliştirilmiştir. Sıfır yığılmalı dağılımlar, tek değişken belirlenerek kullanılabileceği gibi, genelleştirilmiş doğrusal modellerin özelleştirilmiş biçimi olarak çok sayıda açıklayıcı değişken belirlenerek de kullanılmaktadır.

Genelleştirilmiş doğrusal modellerin Nelder ve Wedderburn tarafından 1972 yılında geliştirilmesi ile birlikte bu modeller, gözlemlenen veriye uygun modelin seçilmesi aşamasında yaygın olarak kullanılmıştır. Sıfır yığılmalı regresyon modellerinden biri olan sıfır yığılmalı Poisson modeli, 1992 yılında Lambert tarafından geliştirilmiş ve baskı devre kartlarının üretimi aşamasında karşılaşılan lehimleme hata sayılarının modellenmesinde kullanılmıştır. Sıfır yığılmalı Poisson regresyon modeli; çevresel çalışmalar (Welsh et al., 1996), otoyolların belirli bölümlerinde meydana gelen kaza sıklıklarının modellenmesi (Shankar et al., 1997), iş güvenliği alanında yapılan müdahalelerin, işçilerin yaralanma sayısı üzerindeki etkisinin incelenmesi (Carrivick et al., 2003), genç sürücülerin yaptığı motorlu araç kaza sayılarının modellenmesi (Lee et al., 2002), karayollarının geometrik tasarımlarının gerçekleşen kamyon kazaları üzerindeki etkisinin incelenmesi (Miaou, 1994), otomobil sigortalarında gerçekleşen hasar sayılarının modellenmesi (Yip and Yau, 2005) gibi birçok farklı alanda kullanılmıştır.

Lambert'in (1992) yapmış olduğu çalışmada sıfır yığılmalı negatif binom modelinin, sıfır yığılmalı Poisson modeline alternatif olarak kullanılabileceğine değinilmiş ve buna paralel olarak sıfır yığılmalı negatif binom modeli 1994 yılında Greene tarafından geliştirilmiştir. Bu model, otomobil sigortaları kapsamında belirli bir dönemde gerçekleşen hasar sayısının modellenmesi (Boucher et al., 2007; Flynn and Francis, 2009), otoyolların belirli bölümlerinde meydana gelen kaza sıklıklarının modellenmesi (Shankar et al., 1997), belirli bir kuş türünün sayısının modellenmesi (Martin et al., 2005), genç bireylerin sağlığa zararlı madde kullanımı nedeniyle oluşan sorunlarının belirlenmesi (Simons et al., 2006) gibi birçok farklı alanda kullanılmıştır.

Sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli, Famoye ve Singh tarafından 2006 yılında geliştirilmiş ve Portland şehrinde gözlemlenen aile içi şiddetin modellenmesinde kullanılmıştır. Sıfır yığılmalı genelleştirilmiş Poisson modeli; bir bölgede yaşayan kanserli hasta sayısının modellenmesi (Rashwan and Kamel, 2011), bir hayvan türünün güneşe maruz kalması sonucu gerçekleşen ölüm sayısının modellenmesi (Famoye and Özmen, 2007; Özmen and Demirhan, 2010) gibi çeşitli alanlarda uygulanmıştır.

Bu çalışmanın amacı, Türkiye' de belirli bir yıl içinde Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptıran sigortalıların hasar sayısına ilişkin veri için sırasıyla Poisson, negatif binom, genelleştirilmiş Poisson, sıfır yığılmalı Poisson, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson modellerinin uygunluğu test edilerek, gözlemlenen veri kümesini en iyi biçimde açıklayan modelin belirlenmesidir.

Çalışmanın İkinci Bölümünde genelleştirilmiş doğrusal modeller ile hasar sıklığının modellenmesinde yaygın olarak kullanılan Poisson, negatif binom ve genelleştirilmiş Poisson regresyon modelleri ele alınmıştır.

Üçüncü Bölümde ise sıfır yığılmalı regresyon modelleri ile sıfır yığılmalı Poisson, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modelleri incelenmiştir.

Çalışmanın Dördüncü Bölümünde önceki bölümlerde ele alınan modeller için en çok olabilirlik yöntemi ile parametrelerin tahmini, modellerin uygunluğunun test edildiği uyum iyiliği testleri ve en iyi modelin belirlenmesi amacıyla kullanılan model seçim ölçütleri ayrıntılı olarak ele alınmıştır.

Beşinci Bölümde Türkiye'de 2008 yılında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası poliçelerinde ihbar edilen hasar sayıları kullanılarak Poisson, negatif binom, genelleştirilmiş Poisson regresyon modelleri ile bu modellerin sıfır yığılmalı karmaları oluşturulmuş ve sıfır yığılmalı regresyon modellerinin klasik sayı modellerine göre etkinliği araştırılmıştır.

Çalışmanın 6. Bölümünde, veri kümesindeki hasar sıklıklarına ilişkin en uygun model sonuçları yorumlanarak, daha sonra yapılacak çalışmalar için önerilerde bulunulmuştur.

2. HASAR SAYISI İÇİN KULLANILAN REGRESYON MODELLERİ

2.1. Genelleştirilmiş Doğrusal Modeller

Fiyatlandırmada kullanılan klasik yöntemler, istatistiksel olarak çok yönlü değildir ve geçmiş dönemlerde birçok sigorta dalında hasar deneyimleri, basit tek yönlü (simple one-way) veya çift yönlü (two-way) analizler kullanılarak elde edilmiştir. Tek yönlü analiz, hasar oranı veya sıklığı gibi sigorta istatistiklerinin özetlenmesinde kullanılır. Bu analizlerde, herhangi bir açıklayıcı değişkenin her bir değeri için istatistikler elde edilirken diğer değişkenlerin bu istatistikler üzerindeki etkisi göz ardı edilir. Açıklayıcı değişkenler nitel (ölçülemeyen) veya nicel (ölçülebilen) olması durumunda farklı isimlendirilmektedir. Eğer açıklayıcı değişkenler nitel ise faktör, faktörlere ilişkin gruplar da faktör düzeyi adını almaktadır. Açıklayıcı değişkenlerin nicel olması durumunda ise açıklayıcı değişkenlere ortak değişken adı verilmektedir. Tek yönlü analizlerde hasar deneyimini etkileyen açıklayıcı değişkenler arasında var olabilecek bağımlılığın incelenmemesi, buna benzer ilişkilerin göz önünde bulundurulmasına ve etkilerinin araştırılmasına izin veren genelleştirilmiş doğrusal modeller gibi çok değişkenli modellerin kullanımına duyulan gereksinimi artırmıştır.

Minimum yanlılık uygulamaları olarak da bilinen ve bir fiyatlandırma tekniği olan adımsal yöntemlerin 1960'lı yıllarda aktüerler tarafından geliştirilmesi ile önemli bir iyileştirme sağlanmış ancak tam anlamıyla bir istatistiksel yöntemler bütünü oluşturulamamıştır. Adımsal yöntemlerde; gözlemlenen veri, fiyatlandırmada kullanılan değişkenler ve önceden kararlaştırılan parametreler kümesiyle ilgili eşitlikler oluşturulur. Oluşturulan bu eşitlikler sistemi, optimal çözüme en yakın değerler elde edilene kadar adımsal olarak çözülür. Minimum yanlılık yöntemlerinde optimal çözüm elde edilmesine rağmen, herhangi bir değişkenin elde edilen sonuç üzerindeki etkisinin anlamlılığına ilişkin testler yapılmaz. Ayrıca parametre tahminleri için güven aralığı da oluşturulamamaktadır.

Klasik doğrusal modeller ve minimum yanlılık uygulamalarının birçoğu, genelleştirilmiş doğrusal modellerin özel durumlarıdır. Genelleştirilmiş doğrusal modeller, sigorta veri yapısına ve verinin bağımlı değişkenlerle (predictive variables) olan ilişkisine dair birtakım varsayımların yapılabilmesine olanak sağlar.

Genelleştirilmiş doğrusal modellerin çözümlenmesi, teknik olarak adımsal yöntemlere göre daha etkindir. Genelleştirilmiş doğrusal modellerin sigortacılıktaki başlıca uygulama alanları, fiyatlandırma ve risk kabul sürecidir.

Genelleştirilmiş doğrusal model, en genel tanımı ile bir doğrusal modelin genelleştirilmiş biçimidir. Genelleştirilmiş doğrusal modellerin yapısının daha iyi anlaşılabilmesi için öncelikle klasik doğrusal modellerin yapısının incelenmesi gerekmektedir.

Klasik ve genelleştirilmiş doğrusal modellerin ortak amacı, gözlemlenen yanıt değişkeni (bağımlı değişken) olan Y ile, açıklayıcı değişkenler olarak adlandırılan X değerleri arasındaki ilişkinin açıklanmasıdır. Y_i, i'nci bağımlı değişkeni göstermek üzere birbirinden bağımsız n sayıda bağımlı değişkenden oluşan vektör Y ile tanımlanmaktadır. Doğrusal modelde, bağımlı değişkenin aldığı değerlerden oluşan vektör Y, açıklayıcı değişkenlerden oluşan vektör X ve bilinmeyen parametreler vektörü β ile gösterildiğinde, bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişki;

$$Y = X\beta + \varepsilon$$

biçiminde tanımlanır.

Eşitlikte yer alan $X\beta$ ifadesi, Y raslantı değişkeninin beklenen değeri olup bu değer, açıklayıcı değişkenlerin doğrusal birleşimi şeklinde hesaplanır. Hata terimi olan ε, ortalaması sıfır ve varyansı σ^2 olan normal dağılıma sahiptir.

Doğrusal modellerde amaç, ε hata terimi vektörü elemanlarının kareler toplamını en küçük yapan β vektörüne ait değerlerin bulunmasıdır. Gözlem sayısı n ve modeldeki parametre sayısı k ise, β vektörü k sayıda, ε vektörü n sayıda bileşenden oluşur (k<n).

Bir doğrusal model, iki temel unsurdan oluşur:

- Bağımlı değişken olan Y ile açıklayıcı değişkenler arasındaki ilişkiye dair varsayımlar kümesi

- Sorunun çözülmesi amacıyla optimize edilmek üzere belirlenen bir amaç fonksiyonu. İstatistik teorisinde bu amaç fonksiyonu, olabilirlik fonksiyonu olarak tanımlanır. Hata teriminin normallik varsayımı altında tanımlanan klasik doğrusal modellerde hata kareler toplamını en küçük yapan (minimize eden) parametreler, aynı zamanda olabilirlik fonksiyonunu da en büyük yapar (maksimize eder).

Klasik doğrusal modeller, belirli varsayımlar altında kullanılabilir modellerdir. Bu modellerde tüm gözlemlerin birbirinden bağımsız olduğu ve her bir gözlemin normal dağılımdan geldiği varsayılır. Ayrıca doğrusal modellerde ortalama, açıklayıcı değişkenlerin doğrusal birleşimi ile elde edilir. Bağımlı değişkenle ilişkili bileşenlerin tümünün sabit ve eşit varyansa sahip olduğu varsayılır.

Özet olarak, doğrusal modeller üç temel varsayıma dayanmaktadır:

1. Rasgele bileşen: Bağımlı değişken olan Y'nin tüm bileşenleri birbirinden bağımsızdır ve normal dağılıma uymaktadır. Bu varsayımda, her bir bileşenin ortalama parametresi (μ_i) değişkenlik gösterebilmektedir ancak tüm bileşenlerin varyansı sabit ve birbirine eşittir.
2. Sistematik bileşen: Tüm açıklayıcı değişkenler birleştirilerek doğrusal önkestirim (linear predictor) olarak adlandırılan η elde edilir.

   $$\eta = X\beta$$

3. Bağ fonksiyonu: Modelin rasgele bileşeni ile sistematik bileşeni arasındaki ilişki, bağ fonksiyonu kullanılarak tanımlanır. Doğrusal modellerde kullanılan bağ fonksiyonu özdeş bağ fonksiyonudur. Diğer bir ifadeyle bağımlı değişkenlerin beklenen değeri, açıklayıcı değişken değerlerinden oluşan X matrisi ile kestirilen parametreler vektörü β'nın çarpımına eşittir ve $E(Y) = \mu = \eta$ biçiminde ifade edilir.

(McCullagh and Nelder, 1989)

Doğrusal model kullanımı için gereken normallik ve sabit varyanslılık varsayımları, uygulamada bu modellerin kullanılabilirliğini kısıtlamaktadır. Normallik ve sabit varyanslılık varsayımlarının sağlanmasının zorluğunun yanı sıra, bağımlı değişken yalnızca negatif olmayan değerler aldığında normallik varsayımı istenen koşulu sağlamamakta ve klasik doğrusal model kullanımı uygun olmamaktadır. Ayrıca bu modellerde, sistematik bileşen ve bağ fonksiyonu varsayımlarında incelenen etkenlerin toplanabilirliği kuralı, birçok uygulama alanında geçerliliğini yitirmiştir. Etkenlerin bağımlı değişken üzerindeki etkileri toplamsal olarak ifade edilebileceği gibi çarpımsal olarak da elde edilebilmektedir. Birçok sigorta riski, fiyatlandırma etkenlerinin değişiminden çarpımsal olarak etkilenmektedir. Bu durum, sigorta verisi için klasik doğrusal modellerin kullanılabilirliğini azaltmaktadır.

Genelleştirilmiş doğrusal modeller, klasik doğrusal modelleri de kapsayan çok sayıda modelden oluşmaktadır. Doğrusal modellerin kısıtlayıcı özellikleri olarak ele alınan normallik, sabit varyanslılık ve etkenlerin toplanabilirliği varsayımları genelleştirilmiş doğrusal modellerde bulunmamaktadır.

Bu modellerde bağımlı değişkenin, üstel dağılım ailesinden olan normal, Poisson, gamma, ters gauss, binom, üstel gibi dağılımlardan birine sahip olduğu varsayılır. Ek olarak, klasik doğrusal modellerde varyans sabit kabul edilirken genelleştirilmiş doğrusal modellerde varyans, ortalamaya bağlı olarak değişim gösterebilmektedir. Bu modelleri klasik doğrusal modellerden ayıran bir diğer özellik, açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkisinin, dönüşüm yapıldıktan sonra toplamsal olduğu varsayımının bulunmasıdır.

Doğrusal modellere ilişkin varsayımlara benzer biçimde genelleştirilmiş doğrusal modeller de üç temel varsayıma dayanmaktadır (McCullagh and Nelder, 1989) :

1. Rasgele bileşen: Bağımlı değişkenin tüm elemanları birbirinden bağımsızdır ve her bir eleman üstel dağılım ailesindeki dağılımlardan herhangi birine uyabilir.
2. Sistematik bileşen: Doğrusal modellere benzer olarak, açıklayıcı değişkenlerin doğrusal birleşimi sonucu doğrusal önkestirim olan η elde edilir ve bu terim,

   $$\eta = X\beta \tag{2.1}$$

   biçimindedir.

3. Bağ fonksiyonu: Rasgele bileşen ile sistematik bileşen arasındaki ilişki, $g(x)$ ile belirtilen uygun bağ fonksiyonu kullanılarak açıklanır. Bağ fonksiyonu, tekdüze ve türevlenebilir bir fonksiyondur. Bağımlı değişken ile doğrusal önkestirim arasındaki ilişki,

$$E(Y) = \mu = g^{-1}(\eta) \tag{2.2}$$

biçiminde tanımlanmaktadır.

Üstel dağılım ailesinin, genelleştirilmiş doğrusal regresyon modellerinin önemli bileşenlerden biri olması nedeniyle genelleştirilmiş doğrusal modellerin yapısının kolay anlaşılabilmesi için üstel dağılım ailesinin özellikleri ve matematiksel gösterimi en genel biçimiyle ele alınmalıdır.

Üstel dağılım ailesi, iki parametreli fonksiyonlar kümesi olup Y bağımlı değişkeninin i sayıda bileşeninden her birine ilişkin olasılık fonksiyonu,

$$f(y_i;\theta,\phi) = exp\left\{\left(y_i\theta_i - b(\theta_i)\right)/a(\phi) + c(y_i,\phi_i)\right\} \tag{2.3}$$

biçiminde tanımlanmaktadır. Eş. 2.3.'teki $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ fonksiyonları özel tanımlı fonksiyonlardır. $a(\cdot)$ fonksiyonu pozitif ve süreklidir. $b(\cdot)$ fonksiyonu, θ'ya göre ikinci türevi pozitif olmak koşuluyla iki kez türevlenebilir bir fonksiyondur ve $c(\cdot)$ fonksiyonu, θ parametresinden bağımsızdır. $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ fonksiyonları, f fonksiyonunun bağımlı değişken sınırları dahilinde integrali 1'e eşit olacak biçimde seçilmektedir (Anderson et al., 2007).

Eş. 2.3.'te yer alan θ parametresi doğal parametre ve ϕ parametresi ölçek veya yayılım parametresi olarak tanımlanmaktadır. Her bir gözlem değeri için doğal parametre değişebilir ancak ölçek parametresi tüm gözlemler için sabittir. $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ fonksiyonlarının tüm gözlemler için aynı olduğu varsayımı altında tüm gözlemlerin aynı dağılıma sahip olduğu kabul edilmektedir. Ancak doğal parametrenin gözlemden gözleme değişmesi, ortalamanın da gözlemden gözleme değişmesine olanak sağlamaktadır.

Bazı dağılımlar için varyansı ölçeklendiren ϕ parametresi 1 değerini alabileceği gibi (Poisson dağılımı), üstel dağılım ailesine mensup dağılımların çoğunda ölçek parametresi bilinmemektedir ve veriden yararlanılarak tahmin edilmesi gerekmektedir. Ölçek parametresi, modelde ek parametre olarak düşünülerek en çok olabilirlik yöntemi ile tahmin edilebilmektedir. Ölçek parametresi parametre olarak modele eklenebileceği gibi, veriden yararlanılarak elde edilen tahmin değerinin kullanılması da mümkündür.

Genelleştirilmiş doğrusal modellerin yapısında da yer alan önsel ağırlıklar (w), gözlem değerlerine ilişkin önceden var olan bilginin modelde yer almasına olanak sağlamaktadır. Özellikle hasar sıklığına ilişkin veride karşılaşılması mümkün önsel bilgilerden bir tanesi, riske maruz kalan birim sayısındaki farklılıktır. Bazı portföyler için, hasarlı bir poliçe, bir aylık süre zarfıyla ilgili iken, portföydeki bir diğer hasarlı poliçe, bir yıllık süre zarfıyla ilgili olabilmektedir. Bu durumda gözlemler arası farklılık gösteren riske maruz kalan birim sayısı, verinin güvenilirliğini belirtmek amacıyla önsel bilgi olarak modele eklenebilmektedir. Ancak hasar sayısı modellemelerinde genellikle önsel ağırlıklar 1'e eşit kabul edilmektedir.

Üstel dağılım fonksiyonunun yapısında var olan $a(\phi)$ fonksiyonu için ölçek parametresinin önsel ağırlığa oranı olan ϕ/w kullanılmaktadır.

Üstel dağılım ailesinin iki önemli özelliği vardır:

- Bu ailedeki her bir dağılım, ortalama ve varyansına göre belirlenmektedir.
- Bağımlı değişkene ilişkin varyans, ortalamanın bir fonksiyonudur.

Üstel dağılım ailesine mensup dağılımlardan birine sahip olan bağımlı değişkene ilişkin ortalama ve varyans sırasıyla,

$$E(Y_i) = \mu_i = b'(\theta) \tag{2.4}$$

$$Var(Y_i) = b''(\theta)a(\phi) \tag{2.5}$$

biçiminde verilmektedir. Varyans eşitliğinden de görüldüğü üzere bağımlı değişkene ilişkin varyans, varyans fonksiyonu olarak adlandrılan ve yalnızca doğal parametreye bağlı olan $b(\theta)$ fonksiyonunun θ parametresine göre ikinci türeviyle ve θ parametresinden bağımsız ve yalnızca ölçek parametresine bağlı olan $a(\phi)$ fonksiyonuyla ilişkilidir. Ortalamanın bir fonksiyonu olarak tanımlanan varyans fonksiyonu $(b''(\theta))$, $V(\mu)$ ile gösterilmektedir (Anderson et al., 2007; De Jong and Heller, 2008).

Tek değişkenli durum göz önüne alınarak, üstel dağılım ailesindeki dağılımlardan bazıları ve ilgili $a(\phi)$, $b(\theta)$ ve $c(y,\phi)$ fonksiyonları Tablo 2.1.'de verilmiştir.

Tablo 2.1. Üstel dağılım ailesindeki dağılımlardan bazıları ve ilgili fonksiyonları

Dağılım	$a(\phi)$	$b(\theta)$	$c(y,\phi)$
Normal	ϕ/w	$\theta^2/2$	$-1/2\left(wy^2/\phi + ln(2\pi\phi/w)\right)$
Poisson	ϕ/w	$exp(\theta)$	$-ln\, y!$
Gamma	ϕ/w	$-ln(-\theta)$	$(w/\phi)ln(wy/\phi) - ln\, y - ln(\Gamma(w/\phi))$
Binom(m deneme)	ϕ/w	$m \cdot ln(1+e^{\theta})$	$ln\binom{m}{y}$
Ters Gauss	ϕ/w	$-\sqrt{-2\theta}$	$-1/2\left\{ln(2\pi\phi y^3/w) + w/(\phi y)\right\}$

Tablo 2.1.'de normal ve ters Gauss dağılımlarına ait $c(y,\phi)$ fonksiyonunda yer alan π parametresi, ilgili dağılımın ortalamasını ifade etmektedir (Anderson et al., 2007).

Genelleştirilmiş doğrusal modeller, veri yapısına bağlı olarak seçilen üstel dağılım ailesindeki dağılımlardan biri kullanılarak oluşturulmaktadır. Y_i, i'nci gözleme ilişkin bağımlı değişkeni, $g(x)$ gözlemlenen etkenlerin doğrusal kombinasyonu ile yanıt değişkeninin beklenen değerini ilişkilendirmek amacıyla belirlenen bağ fonksiyonunu, X_{ij} i'nci gözlemin j'nci etken için değerini, β_j j'nci etken için tahmin edilecek model parametresini, ϕ üstel dağılım ailesi için varyans fonksiyonunu ölçeklendiren parametreyi (ölçek parametresini), $V(x)$ varyans fonksiyonunu, w_i i'nci gözleme ilişkin önsel ağırlık veya güvenilirliğin ölçüsünü göstermek üzere, bir genelleştirilmiş doğrusal modele ilişkin ortalama ve varyans sırasıyla,

$$\mu_i = E(Y_i) = g^{-1}\left(\sum_j X_{ij}\beta_j + \xi_i\right) \qquad i = 1,2,\ldots,n \quad ; \quad j = 1,2,\ldots,k \tag{2.6}$$

$$Var(Y_i) = \frac{\phi V(\mu_i)}{w_i} \tag{2.7}$$

biçiminde tanımlanmaktadır. Genelleştirilmiş doğrusal modellerin yapısında yer alan ξ_i , i'nci gözleme ilişkin bilinen etki (known effect, offset) terimidir. Bu terim, herhangi bir açıklayıcı değişkenin bağımlı değişken üzerindeki etkisinin bilindiği durumlarda kullanılmaktadır. Etkisi bilinen açıklayıcı değişken için bilinmeyen regresyon parametreler vektörü β'nın tahmini yerine bu açıklayıcı değişkene ait bilgi, modele bilinen etki terimi olarak eklenmektedir. Bilinen etki terimi, ilgili modelin doğrusal önkestirim tanımına eklenerek modele dahil edilmektedir. Bilinen etki teriminin doğrusal önkestirime eklenmesiyle,

$$\eta = X\beta + \xi \tag{2.8}$$

eşitliği elde edilmektedir.

Doğrusal önkestirim fonksiyonu Eş. 2.8.'deki gibi tanımlanan bağımlı değişkenin beklenen değeri;

$$E(Y) = \mu = g^{-1}(\eta) = g^{-1}(X\beta + \xi)$$

eşitliği ile elde edilmektedir.

Genelleştirilmiş doğrusal modellerin yapısında bulunan bağımlı değişken, açıklayıcı değişkenler ve önsel ağırlıklar analistin vereceği kararlara bağlı olmakta ve veriye göre değişkenlik gösterebilmektedir. Dolayısıyla bir genelleştirilmiş doğrusal modelin biçimini belirleyen faktörler $g(x)$ bağ fonksiyonu, $V(x)$ varyans fonksiyonu ve tahmin edilen veya bilinen ϕ ölçek parametresidir.

Klasik doğrusal regresyon modellerinin kullanımında veriye çeşitli dönüşümler uygulanarak normallik, sabit varyanslılık ve etkenlerin toplanabilirliği varsayımlarının sağlanması amaçlanmaktadır. Genelleştirilmiş doğrusal modellerde ise yalnızca belirlenen uygun bağ fonksiyonu kullanılarak bağımlı değişkene uygulanan dönüşüm sayesinde etkenlerin toplanabilirliği kuralı geçerli olmaktadır.

i'nci gözleme ilişkin ortalama, i'nci gözleme ilişkin doğrusal önkestirimin bir fonksiyonu olarak ifade edilmektedir. Doğrusal önkestirim için belirlenen bu fonksiyon genellikle bağ fonksiyonunun ters fonksiyonudur ve

$$\mu_i = g^{-1}(\eta_i) \tag{2.9}$$

biçiminde ifade edilir.

Teoride her bir gözlem değeri için farklı bağ fonksiyonu kulllanılması mümkündür fakat uygulamada sık karşılaşılan bir durum değildir. Herhangi bir genelleştirilmiş doğrusal modelde kullanılan bağ fonksiyonunun, türevlenebilirlik ve tekdüzelik koşullarını sağlaması gerekmektedir. Tablo 2.2.'de genelleştirilmiş doğrusal modellerde kullanılan bazı bağ fonksiyonları ve ters fonksiyonları gösterilmiştir (Anderson et al., 2007).

Tablo 2.2. Genelleştirilmiş doğrusal modellerde kullanılan başlıca bağ fonksiyonları ve ters fonksiyonları

Bağ Fonksiyonu	$g(x)$	$g^{-1}(x)$
Birim	x	x
Logaritmik	$ln(x)$	e^x
Lojit	$ln(x/(1-x))$	$e^x/(1+e^x)$
Ters	$1/x$	$1/x$

Genelleştirilmiş doğrusal modellerde kullanılan bağ fonksiyonlarından biri olan logaritmik bağ fonksiyonunda, etkenler bağımlı değişkeni çarpımsal olarak etkilemektedir. Diğer bir ifadeyle modellemede logaritmik bağ fonksiyonu kullanıldığında; etkenlerin bağımlı değişken üzerindeki etkisi, çarpımsal etkinin logaritması biçiminde hesaplanmaktadır.

Logaritmik bağ fonksiyonu $g(x)$ ve ters bağ fonksiyonu $g^{-1}(x) = e^{x}$ ile gösterildiğinde i'nci gözleme ilişkin ortalama ile k sayıda açıklayıcı değişken arasındaki ilişki,

$$\mu_i = g^{-1}\left(\sum_{j=1}^{k} x_{ij}\beta_j\right) = exp(\beta_1 X_{i1}) \cdot exp(\beta_2 X_{i2}) \cdot ... \cdot exp(\beta_k X_{ik}) \tag{2.10}$$

biçiminde açıklanmaktadır. Hasar sıklığına veya hasar sayısına ilişkin verinin modellenmesi amacıyla yaygın olarak kullanılan modellerden bazıları Poisson modeli, negatif binom modeli ve negatif binom dağılımına benzer özellikler gösteren fakat üstel dağılım ailesine mensup olmayan genelleştirilmiş Poisson modelidir.

2.2. Poisson Regresyon Modeli

Bağımlı değişkenin negatif olmayan kesikli sayılardan oluştuğu ve belirli bir zaman aralığında gerçekleşen olay sayısının küçük olduğu durumda, bu değişkenin dağılımı için belirlenen dağılımlardan biri Poisson dağılımıdır. Poisson regresyonunda ortalama (μ), açıklayıcı değişkenler ve uygun bağ fonksiyonu kullanılarak hesaplanmaktadır.

Poisson regresyonunda kullanılan bağ fonksiyonları, özdeş veya logaritmik bağ fonksiyonudur. Logaritmik bağ fonksiyonu kullanıldığında ortalama pozitiftir, ancak özdeş bağ fonksiyonu kullanıldığında ortalamanın pozitif değer alacağı kesin değildir. Ayrıca logaritmik bağ fonksiyonu kullanıldığında, modeldeki herhangi bir açıklayıcı değişkenin değerindeki değişimin bağımlı değişken üzerindeki etkisi çarpımsal iken, özdeş bağ fonksiyonu kullanıldığında bu değişimin bağımlı değişken üzerindeki etkisi toplamsaldır. Dolayısıyla sigorta verisi için Poisson regresyonunda bağ fonksiyonu olarak logaritmik bağ fonksiyonunun kullanılması daha uygundur.

i'nci gözleme ilişkin bağımlı değişken y_i olmak üzere n sayıda bağımlı değişkenden oluşan $(n \times 1)$ boyutlu vektör Y, k sayıda açıklayıcı değişkenden oluşan $(n \times k)$ boyutlu açıklayıcı değişkenler girdi matrisi $X\ (x_{i1} = 1,\ i = 1,2,...,n)$ ve $(k \times 1)$ boyutlu bilinmeyen regresyon parametreleri vektörü β $(\beta_1 = 1)$ olmak üzere genelleştirilmiş doğrusal Poisson modeli;

$$P(Y_i = y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i\,!} \qquad i = 1,2,\ldots,n \ \ ; \ y_i = 0,1,2,3,...$$

ve

$$ln(\mu_i) = \sum_{j=1}^{k} x_{ij} \beta_j \tag{2.11}$$

biçiminde tanımlanmaktadır.

Poisson regresyon modeline göre y_i bağımlı değişkenleri birbirinden bağımsızdır ve her bir değişken μ_i parametresi ile Poisson dağılmaktadır. Bilinmeyen regresyon parametreleri (β), en çok olabilirlik yöntemi, yarı olabilirlik yöntemi veya genelleştirilmiş en küçük kareler yöntemi gibi tahmin yöntemleri kullanılarak tahmin edilmektedir.

Poisson modelinde yer alan bilinmeyen regresyon parametrelerine ilişkin tahmin değerlerini elde etmek amacıyla kullanılan log-olabilirlik fonksiyonu;

$$ln\,L(\beta; y) = \sum_{i=1}^{n} \left[y_i\, ln\, \mu_i - \mu_i - ln(y_i\,!) \right]$$

biçiminde tanımlanmaktadır. Poisson modeline ait log-olabilirlik fonksiyonunun β parametrelerine göre kısmi türevi;

$$\frac{\partial\left[\ln L(\beta;y)\right]}{\partial\beta_r}=\sum_{i=1}^{n}\left(y_i-\mu_i\right)x_i=0 \qquad r=1,2,...,k$$

eşitliği ile belirtilmektedir (Greene, 1994).

Poisson regresyonunun kullanımını sınırlandıran bir özellik, ortalamanın varyansa eşit olmasıdır $\left(E\left(Y_i\,|x_i\right)=Var\left(Y_i\,|x_i\right)=\mu_i\right)$. Ancak gerçek veri kümesi üzerinde bu varsayımın sağlanması oldukça güçtür (Giles, 2010).

Verideki yayılımın Poisson modeliyle tahmin edilen yayılımdan fazla olduğu durum, istatistik literatüründe aşırı yayılım olarak tanımlanmaktadır. Aşırı yayılım problemi Bölüm 2.2.1'de detaylı olarak ele alınmıştır.

2.2.1. Aşırı yayılım

Poisson modelinde de açıklandığı üzere verideki yayılımın Poisson modeliyle tahmin edilen yayılımdan fazla olduğu durum, istatistik literatüründe aşırı yayılım olarak tanımlanmaktadır. Aşırı yayılım sorununa neden olan durumlardan bazıları,

- gözlemlenen bir Poisson sürecinin aralık uzunluğunun rasgele yerine sabit belirlenmesi,
- verinin, her bir olayın toplama rasgele miktarda katkıda bulunduğu kümelenmiş Poisson sürecinden elde edilmesi,
- genel varsayım olan bağımsızlığın sağlanmaması,
- oluşturulan model için önemli açıklayıcı değişkenlerin ölçülmemiş olması ya da
- önemli açıklayıcı değişkenlerin modelden çıkartılmasıdır.

(McCullagh and Nelder, 1989; Miaou, 1994). Poisson regresyon modelinde aşırı yayılımın göz ardı edilmesi, regresyon parametrelerine ilişkin en çok olabilirlik yöntemi gibi yöntemler kullanılarak elde edilen tahminlerin tutarlı olmasına karşın, varyanslarının olması gereken değerinden düşük hesaplanmasına neden olmaktadır. Diğer bir ifadeyle, gözlem sayısı büyük iken, Poisson modeline ait regresyon parametrelerine ilişkin tahmin değerleri ile gerçek parametre değerleri arasındaki farkın az olmasına karşın, tahmin edilen parametrelerin anlamlılık düzeyleri bulunması gerekenden daha yüksektir (Miaou, 1994; Yip and Yau, 2005).

Aşırı yayılım sorunuyla karşılaşılan durumlarda Poisson modeli yerine, aşırı yayılımın da dikkate alındığı ve modele eklenen yayılım parametresi ile açıklandığı dağılımlarından herhangi biri kullanılmaktadır. Karma Poisson dağılımlarında Poisson dağılımının μ parametresi gözlemler arası değişkenlik göstermektedir ve bu değişkenlik gözlemlenmemiş değişkenlik (unobserved heterogeneity) terimi ile açıklanmaktadır. Poisson modeline alternatif olarak kullanılan karma Poisson dağılımlarından en önemlisi negatif binom dağılımıdır.

2.3. Negatif Binom Regresyon Modeli

Negatif binom dağılımı, Poisson modeline göre aşırı yayılım gösteren verinin modellenmesi amacıyla kullanılan ve Poisson dağılımının karması biçiminde elde edilen dağılımlardan biridir (Denuit et al., 2007). Aşırı yayılımın modele dahil edilmesi amacıyla, her bir gözlem için farklılık gösteren ve gözlemlenmemiş değişkenlik terimi olarak tanımlanan θ raslantı değişkeni, Poisson dağılımının μ parametresi ile çarpılmaktadır. Genellikle gözlemlenmemiş değişkenlik teriminin ortalaması 1, varyansı ise sabit ve τ'ya eşit kabul edilmektedir. Dolayısıyla, Poisson modeli kullanılarak elde edilen ortalama değeri değişmezken, varyans ortalamadan yüksek hesaplanmaktadır (Boucher et al., 2007).

Değişkenlik terimi, Gamma, ters Gauss, Log-normal gibi dağılımlardan herhangi birine sahip olabilmektedir. Karma Poisson dağılımları, genellikle değişkenlik teriminin sahip olduğu dağılıma göre adlandırılmaktadır. Değişkenlik teriminin dağılımı Ters Gauss ise, elde edilen karma dağılım Poisson-ters Gauss dağılımı (PIG), Log-normal ise elde edilen dağılım Poisson Log-normal dağılımı (PLN) adını almaktadır. Y_i bağımlı raslantı değişkeni karma Poisson dağılımlarından herhangi birine sahip olduğunda, bu raslantı değişkenine ilişkin ortalama ve varyans sırasıyla,

$$E\left[Y_i\right] = \theta\mu_i \quad ; \quad Var\left[Y_i\right] = \mu_i + \mu_i^2\tau \qquad i = 1,2,\ldots,n \tag{2.12}$$

biçiminde tanımlanmaktadır (Miaou, 1994; Boucher et al., 2007).

Gözlemlenmemiş değişkenlik terimi Θ_i, ortalaması 1 ve varyansı α olan Gamma dağılımına sahip olduğunda, elde edilen dağılım negatif binom dağılımıdır. i'nci gözleme ilişkin bağımlı değişken y_i olmak üzere n sayıda bağımlı değişkenden oluşan $(n \times 1)$ boyutlu vektör Y, k sayıda açıklayıcı değişkenden oluşan $(n \times k)$ boyutlu açıklayıcı değişkenler girdi matrisi X $(x_{i1} = 1,\ i = 1,2,...,n)$ ve $(k \times 1)$ boyutlu bilinmeyen regresyon parametreleri vektörü β $(\beta_1 = 1)$ olmak üzere negatif binom regresyon modeli;

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \quad \begin{array}{l} i = 1,2,...,n \\ y_i = 0,1,2,... \end{array} \qquad (2.13)$$

biçiminde tanımlanmaktadır.

Eş. 2.13.'te $\mu_i = exp(x_i \beta)$ ve $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha - 1} dt$ şeklinde ifade edilmektedir. α parametresi, sıfırdan büyük bir değer olup genellikle yayılım parametresi olarak adlandırılmaktadır (Boucher et al., 2007; Denuit et al., 2007).

Negatif binom modeli, Poisson modelinin karması biçiminde tanımlandığından bağ fonksiyonu olarak logaritmik bağ fonksiyonu kullanılmaktadır. Negatif binom modeli için ortalama ve varyans sırasıyla,

$$E(Y_i | x_i) = exp\left(\sum_{j=1}^{k} x_{ij} \beta_j \right) = \mu_i \qquad (2.14)$$

$$Var(Y_i | x_i) = \mu_i + \alpha \mu_i^2 \qquad i = 1,2,...,n \qquad (2.15)$$

eşitlikleri ile elde edilmektedir. Ortalama ve varyans eşitliklerinden de görüleceği üzere değişkenlik teriminin eklenmesiyle elde edilen negatif binom modeli ile, varyansın ortalamadan yüksek hesaplanmasına olanak sağlanarak aşırı yayılımın modele dahil edilmesi sağlanmaktadır. Ek olarak, yayılım parametresi sıfıra eşit olduğunda $(\alpha = 0)$ negatif binom modeli, Poisson modeline indirgenmektedir.

Negatif binom modelinin Poisson modeline göre uygunluğunu test etmek için negatif binom modeline ilişkin yayılım parametresinin anlamlılığının test edilmesi yeterli olmaktadır. Bunun nedeni, negatif binom modeline ait yayılım parametresi olan α'nın değeri sıfıra eşit olduğunda bu modelin, Poisson modeline indirgenmesidir.

Negatif binom modeli, Cameron ve Trivedi (1986) tarafından genelleştirilmiş ve ortalaması negatif binom modeli ile elde edilen ortalamaya eşit (μ_i) ancak varyansı $\mu_i + \alpha\mu_i^{k}$ biçiminde tanımlanan NBk dağılımları elde edilmiştir. Bu dağılımlar için değişkenlik terimi, ortalaması 1 ve varyansı $\alpha\mu_i^{k-2}$ olmak üzere Gamma dağılmaktadır. $k=2$ için negatif binom modeli (NB2) ve $k=1$ için NB1 modeli elde edilmektedir (Boucher et al., 2007). Bu çalışmanın uygulama bölümünde yalnızca $k=2$ için tanımlanan negatif binom dağılımı kullanılmıştır.

Negatif binom modelinde yer alan parametrelere ilişkin tahmin değerlerini elde etmek amacıyla kullanılan log-olabilirlik fonksiyonu;

$$ln\left(\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(\alpha^{-1})}\right)=\sum_{j=0}^{y_i-1} ln(j+\alpha^{-1})$$

eşitliğinden yararlanılarak;

$$ln\,L(\alpha,\beta;y)=\sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y_i-1} ln\left(j+\alpha^{-1}\right)\right)-ln\,y_i\,!-\left(y_i+\alpha^{-1}\right)ln\left(1+\alpha\mu_i\right)+y_i\,ln\,\alpha+y_i\,ln\,\mu_i\right\}$$

biçiminde elde edilmektedir.

Log-olabilirlik fonksiyonuna ilişkin diğer bir gösterim ise;

$$ln\,L(\alpha,\beta;y)=\sum_{i=1}^{n}\left\{y_i\,ln\left(\frac{\alpha^{-1}\mu_i}{1+\alpha^{-1}\mu_i}\right)-\alpha\,ln(1+\alpha^{-1}\mu_i)+ln\,\Gamma(y_i+1)-ln\,\Gamma(\alpha)\right\}$$

olarak ifade edilmektedir.

Negatif binom modeline ilişkin log-olabilirlik fonksiyonunun α ve β parametrelerine göre kısmi türevleri sırasıyla;

$$\frac{\partial\left[\ln L(\alpha,\beta;y)\right]}{\partial\beta_r}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)x_i}{(1+\alpha\mu_i)}=0 \qquad r=1,2,..,k$$

$$\frac{\partial\left[\ln L(\alpha,\beta;y)\right]}{\partial\alpha}=\sum_{i=1}^{n}\left\{\frac{1}{\alpha^2}\left(\ln(1+\alpha\mu_i)-\sum_{j=0}^{y_i-1}\frac{1}{(j+\alpha^{-1})}\right)+\frac{y_i-\mu_i}{\alpha(1+\alpha\mu_i)}\right\}=0$$

eşitlikleri ile verilmektedir (Lord and Park, 2010).

2.4. Genelleştirilmiş Poisson Regresyon Modeli

Genelleştirilmiş Poisson modeli, Poisson modelinin aşırı yayılım veya az yayılım nedeniyle uygun olmadığı durumlarda kullanılmaktadır. Negatif binom dağılımı yalnızca aşırı yayılım durumunda kullanılırken genelleştirilmiş Poisson dağılımı, aşırı yayılım ve az yayılım durumlarının her ikisi için de kullanılmaktadır (Famoye and Singh, 2006). Açıklayıcı değişkenlerin göz ardı edildiği tek değişkenli genelleştirilmiş Poisson modeli; Janardan, Kerster ve Schaeffer (1979) ile Consul (1979) gibi birçok analist tarafından kullanılmıştır. İlk olarak Consul ve Famoye (1992) tarafından çalışılan ve geliştirilen genelleştirilmiş Poisson regresyon modeli ise, bir veya daha fazla açıklayıcı değişkenden etkilenen bağımlı değişkenin dağılımını tahmin etmek amacıyla, genelleştirilmiş doğrusal modellere benzer biçimde kullanılmıştır. Genelleştirilmiş Poisson regresyon modelinde yer alan yayılım parametresi için regresyon çözümlemesinin yapıldığı model Czado et al. (2007) tarafından geliştirilmiştir.

i'nci gözleme ilişkin bağımlı değişken y_i olmak üzere n sayıda bağımlı değişkenden oluşan $(n\times 1)$ boyutlu vektör Y, k sayıda açıklayıcı değişkenden oluşan $(n\times k)$ boyutlu açıklayıcı değişkenler girdi matrisi X $(x_{i1}=1,\ i=1,2,...,n)$ ve $(k\times 1)$ boyutlu bilinmeyen regresyon parametreleri vektörü β $(\beta_1=1)$ olmak üzere μ_i ve φ parametrelerinin yer aldığı genelleştirilmiş Poisson regresyon modeli;

$$P(Y_i = y_i) = \frac{\mu_i \left(\mu_i + (\varphi - 1) y_i\right)^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp\left(-\frac{1}{\varphi}\left(\mu_i + (\varphi - 1) y_i\right)\right) \qquad \begin{array}{l} i = 1,2,...,n \\ y_i = 0,1,2,3,... \end{array}$$

(2.16)

biçiminde tanımlanmaktadır. Eş. 2.16.'da, $\mu_i > 0$ olmak üzere $\mu_i = \exp(x_i \beta)$ olarak ifade edilmektedir. Burada φ, yayılım parametresidir. Yayılım parametresinin değeri birden küçük olduğunda $(\varphi < 1)$, $\mu_i + m(\varphi - 1) > 0$ koşulunu sağlayan en büyük m değeri için yayılım parametresinin alacağı değerin alt sınırı, $\varphi > \max(1/2, 1 - \mu_i / m)$ olarak belirlenmektedir. Yayılım parametresinin değeri bire eşit olduğunda genelleştirilmiş Poisson modeli, Poisson modeline indirgenmektedir. Kurulan genelleştirilmiş Poisson modelinde yayılım parametresi birden küçük ise az yayılım, büyük ise aşırı yayılım söz konusudur.

Poisson ve negatif binom modellerinde olduğu gibi genelleştirilmiş Poisson modelinde de genellikle logaritmik bağ fonksiyonu kullanılmaktadır. Genelleştirilmiş Poisson modeline ilişkin ortalama ve varyans eşitlikleri;

$$E\left(Y_i \mid x_i\right) = \mu_i \quad ; Var\left(Y_i \mid x_i\right) = \mu_i \varphi^2 \qquad i = 1,2,...,n \tag{2.17}$$

biçimindedir (Czado et al., 2007).

Genelleştirilmiş Poisson modelinin Poisson modeline göre uygunluğunu test etmek için genelleştirilmiş Poisson modeline ilişkin yayılım parametresinin anlamlılığının test edilmesi yeterli olmaktadır. Bunun nedeni, genelleştirilmiş Poisson modeline ait yayılım parametresi olan φ'nin değeri bire eşit olduğunda bu modelin, Poisson modeline indirgenmesidir.

Genelleştirilmiş Poisson modelinde yer alan φ ve β parametrelerinin tahmini için gereken log-olabilirlik fonksiyonu;

$$\ln L(\varphi, \beta; y) = \sum_{i=1}^{n} \left\{ (y_i - 1) \ln\left(\mu_i \left(\mu_i + (\varphi - 1) y_i\right)\right) - y_i \ln(\varphi) - \frac{1}{\varphi}\left(\mu_i + (\varphi - 1) y_i\right) - \ln(y_i!) \right\}$$

biçiminde verilmektedir. Genelleştirilmiş Poisson modeli için log-olabilirlik fonksiyonunun φ ve β parametrelerine göre kısmi türevi alınarak;

$$\frac{\partial\left[\ln L(\varphi,\beta;y)\right]}{\partial\varphi}=\sum_{i=1}^{n}\left\{\frac{y_i(y_i-1)}{\mu_i+y_i(\varphi-1)}-\frac{y_i}{\varphi}+\frac{\mu_i-y_i}{\varphi^2}\right\}=0$$

ve

$$\frac{\partial\left[\ln L(\varphi,\beta;y)\right]}{\partial\beta_r}=\sum_{i=1}^{n}\left\{1+\frac{(y_i-1)\mu_i}{\mu_i+y_i(\varphi-1)}-\frac{\mu_i}{\varphi}\right\}=0 \qquad r=1,2,..,k$$

eşitlikleri elde edilmektedir (Czado et al., 2007).

Parametre tahmini için elde edilen eşitlikler, eş zamanlı olarak adımsal algoritma yardımıyla çözülmektedir. Algoritmada kullanılmak üzere β için belirlenen başlangıç değerleri, verinin Poisson modeli ile modellenmesi sonucu elde edilen β tahmin değerleridir. φ parametresi için başlangıç değeri bir olarak belirlenebileceği gibi Pearson ki-kare istatistiğinin $(n-k)$ serbestlik derecesine eşitlenmesi yoluyla da elde edilebilmektedir. Bu eşitlik;

$$\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i\hat{\varphi}^2}=n-k$$

biçiminde belirtilmektedir. Burada n gözlem sayısını, k ise modeldeki regresyon parametre sayısını göstermektedir (Wang and Famoye, 1997).

3. SIFIR YIĞILMALI REGRESYON MODELLERİ

3.1. Giriş

Hasar sayısının modellenmesi amacıyla ilk olarak Poisson ve negatif binom modelleri kullanılmıştır (Thomas and Samson, 1987; Renshaw, 1994; Haberman and Renshaw, 1996). Ancak veri yapısının geliştirilen uygulamalara bağlı olarak farklılık gösterdiği durumlarda Poisson veya negatif binom modellerinin kullanımı uygun olmamaktadır. Hayat dışı sigortalarından Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası ve Kara Taşıtları Kasko Sigortası (otomobil sigortaları) kapsamında, hasar sayısı verisi kullanılarak yapılan modelleme çalışmalarında oldukça sık karşılaşılan durumlardan biri, verinin aşırı yayılım göstermesinin yanı sıra belirli bir dönemde portföyde yer alan poliçeler içerisinde hasar bildirmeyen poliçe sayısının Poisson veya negatif binom modeline göre beklenen sayıdan yüksek, aynı dönemde bir hasar bildiren poliçe sayısının ise Poisson veya negatif binom modeline göre beklenen sayıdan düşük olmasıdır. Hasar sayısına ilişkin verinin sıfır ve bir değerlerinde gözlemlenen bu farklılığın nedeni, otomobil sigortaları kapsamında yaygın olan muafiyet ve hasarsızlık indirimi uygulamalarıdır. Bu nedenle ortalamanın varyansa eşit olduğu Poisson modeli, hasar sayısı veya sıklığı için uygun olmamaktadır. Poisson modeline alternatif olarak birçok çalışmada kullanılan negatif binom ve genelleştirilmiş Poisson modelleri ile, mevcut aşırı yayılımın modellenmesinin bir sonucu olarak Poisson modeline göre veriye uyumda artış sağlanmasına rağmen, hasar sayısında gözlemlenen sıfır ve bir değerindeki farklılığın açıklanmasında yetersiz kalmaktadır.

Belirli bir dönemde yapılan gözlem sonucu elde edilen verinin aşırı yayılım gösterdiği, aynı zamanda da gözlemlenen sıfır sayısının varsayılan modelde beklenen sıfır sayısından fazla olduğu durumda kullanılmak üzere geliştirilen modellerden biri, sıfır yığılmalı (zero inflated) modellerdir. Sıfır yığılmalı modeller, veride mevcut olan fazla sıfırlara ilişkin yoğunluğun temel sayı dağılımından ayrı tutularak hesaplanması nedeniyle iki durumlu süreçler (dual state process) olarak da adlandırılmaktadır (Shankar et al., 1997). Bu modeller, bağımlı değişkenin sıfır değeri için Bernoulli denemesi yapılarak belirlenen bir yoğunluk (zero point mass) ile sıfır değerini de kapsayan kesikli bir dağılımın karmasından oluşmaktadır. Sıfır yığılmalı modellere benzer amaçla kullanılan bir diğer model, engelli modeldir (hurdle model). Engelli modellerde ise bağımlı değişkenin sıfır değeri için Bernoulli denemesi yapılırken pozitif değerleri için kesikli ve kesilmiş (truncated) bir dağılım kullanılmaktadır (Boucher et al., 2007; Flynn and Francis, 2009).

Sıfır yığılmalı modellerde temel sayı dağılımından, diğer bir ifadeyle hasar sayısı verisinin dağılımı için belirlenen dağılıma göre hasar yapılmamış poliçelerden kaynaklanan sıfır değerleri, yapısal olmayan sıfır hasar sayıları olarak adlandırılmaktadır. Yapısal sıfır hasar sayıları ise otomobil sigortalarında karşılaşılan muafiyet ve hasarsızlık indirimi gibi uygulamaların bir sonucu olarak hasar yapılmış fakat bildirilmemiş olan poliçelerden kaynaklanan ve verinin sıfır değerindeki yığılmanın nedeni olarak ifade edilen değerlerdir (Yip and Yau, 2005; Özmen and Demirhan, 2010).

Poisson, negatif binom ve genelleştirilmiş Poisson modellerine benzer biçimde, sigortalıların bireysel özelliklerinin modele dahil edilmesi amacıyla açıklayıcı değişkenlerin kullanıldığı sıfır yığılmalı regresyon modelleri elde edilmiştir. Negatif olmayan ve sıfır yığılmalı dağılım gösteren kesikli raslantı değişkeni Y_i olmak üzere; açıklayıcı değişkenlerin hem yapısal sıfır oranı olan ω için, hem de temel sayı dağılımı için kullanıldığı sıfır yığılmalı regresyon modelleri;

$$Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i) Pr(K_i = 0), & y_i = 0 \\ (1-\omega_i) Pr(K_i = y_i), & y_i > 0 \end{cases} \tag{3.1}$$

biçiminde tanımlanmaktadır. Eş. 3.1.'de tanımlanan K_i raslantı değişkeni, hasar sayısını modellemek amacıyla kullanılan negatif olmayan kesikli dağılımlardan herhangi birine sahip olabilmektedir. Yapısal sıfır oranı için $0 < \omega_i < 1$ varsayımı yapılarak, sıfır değerindeki yoğunluk modele dahil edilmektedir. $\omega_i < 0$ için elde edilen modeller sıfır değerlerinin çıkartıldığı sayı modelleridir (zero-deflated count distributions). Yapısal sıfır oranı sıfıra eşit olduğunda $(\omega_i = 0)$ sıfır yığılmalı modeller, temel sayı dağılımı kullanılarak elde edilen modele indirgenmektedir.

Herhangi bir sıfır yığılmalı modelde, açıklayıcı değişkenlerin yalnızca bağımlı değişkenin pozitif değerlerine ait dağılım için tanımlanması ve yapısal sıfır oranı için regresyon modeli yerine sabit bir oran belirlenmesi mümkündür. Bu oran, oluşturulan modelin diğer parametreleriyle eş zamanlı olarak tahmin edilir.

Sıfır yığılmalı modellere ilişkin beklenen değer ve varyans sırasıyla,

$$E(Y_i | x_i) = (1-\omega_i)E(K) \tag{3.2}$$

$$Var\left(Y_i | x_i\right) = (1-\omega_i)\left\{Var(K) + \omega_i \left[E(K)\right]^2\right\} \tag{3.3}$$

eşitlikleri ile tanımlanmaktadır. Sıfır yığılmalı modellerden sıfır yığılmalı Poisson (ZIP), sıfır yığılmalı negatif binom (ZINB) ve sıfır yığılmalı genelleştirilmiş Poisson (ZIGP) modelleri, birçok alanda yaygın olarak kullanılmaktadır. Bu bölümde sıfır yığılmalı Poisson, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modellerinin yapısı ve özellikleri açıklanmıştır.

3.2. Sıfır Yığılmalı Poisson Regresyon Modeli

Sıfır yığılmalı Poisson regresyon modeli, ilk olarak Lambert (1992) tarafından geliştirilmiş ve bu model, birçok çalışmada kullanılmıştır. Sıfır yığılmalı Poisson modeline uyan hasar sayısı raslantı değişkenine ilişkin olasılık fonksiyonu;

$$Pr(Y_i = 0) = \omega_i + (1-\omega_i)e^{-\mu_i}$$

$$Pr(Y_i = y_i) = (1-\omega_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \qquad i = 1,2,...,n \qquad y_i = 1,2,3,... \qquad 0 < \omega_i < 1 \tag{3.4}$$

eşitlikleri ile tanımlanmaktadır. Poisson modeline ait ortalama parametresi için logaritmik bağ fonksiyonu kullanılmaktadır. Dolayısıyla Eş. 3.4.'te $\mu_i = exp(X_i\beta)$ biçiminde hesaplanmaktadır.

Yapısal sıfır oranı belirlenirken açıklayıcı değişkenlerin etkisi göz önünde bulundurulduğunda Bernoulli başarı olasılıklarının doğrusallaştırılması ve yapısal sıfır oranının pozitif bulunması amacıyla bağ fonksiyonu olarak genellikle lojit bağ fonksiyonu kullanılmaktadır (Lambert, 1992; Shankar et al., 1997; Famoye and Singh, 2006; Famoye and Özmen, 2007; Giles, 2010). Eş. 3.4.'te yer alan yapısal sıfır oranı için;

$$lojit(\omega_i) = ln\left(\omega_i / (1-\omega_i)\right) = G_i \gamma \qquad i = 1,2,...,n \tag{3.5}$$

dönüşümü yapılmaktadır. Eş. 3.5.'te tanımlanan G matrisi, ω_i için belirlenen açıklayıcı değişkenlerin oluşturduğu matris ve γ, bilinmeyen regresyon parametreler vektörüdür. G ve γ'nın boyutu, ω_i oranının hesaplanmasında kullanılan açıklayıcı değişken sayısına göre belirlenmektedir. Kullanılan açıklayıcı değişken sayısı m ise G matrisi $(n \times m)$ boyutlu ve γ vektörü $(m \times 1)$ boyutludur $(\gamma_1 = G_{i1} = 1,\ i = 1,2,3,..,n)$.

Yapısal sıfır oranı için regresyon çözümlemesi yapılmadığında G matrisi bir değerlerinden oluşan $(n \times 1)$ boyutlu sütun vektörüdür ve sıfır yığılmalı Poisson modeli için tahmin edilen parametre sayısı, Poisson modeli için tahmin edilen parametre sayısından bir fazladır. Yapısal sıfır oranı için regresyon analizi yapılmadığı durumda bu oran;

$$lojit(\omega_i) = ln\left(\omega_i / (1-\omega_i)\right) = \gamma \tag{3.6}$$

eşitliği kullanılarak en çok olabilirlik yöntemi ile tahmin edilmektedir.

ω_i ve μ_i parametresi için aynı açıklayıcı değişkenler kullanılabileceği gibi, yapısal sıfır oranı için daha az sayıda açıklayıcı değişken kullanılması mümkündür. Her iki regresyon modelinde de aynı açıklayıcı değişkenler kullanıldığında μ_i ile ω_i fonksiyonel olarak birbiriyle ilişkili değilse tahmin edilen parametrelere ilişkin β ve γ vektörleri aynı boyutlara sahiptir ve birbirine eşittir (Lambert, 1992).

Sıfır yığılmalı Poisson modelinin her iki bölümünde de aynı açıklayıcı değişkenler kullanılarak regresyon çözümlemesi yapıldığında, μ_i parametresi ile ω_i oranının fonksiyonel olarak ilişkili olduğu düşünülmekte ve bu ilişki bilinmeyen gerçel değerli biçim parametresi τ ile açıklanmaktadır. μ_i ve ω_i arasındaki ilişki τ ile açıklandığında bu parametreler için tanımlanan regresyon modelleri,

$$ln(\mu_i) = X_i \beta \ ; \ lojit(\omega_i) = -\tau X_i \beta \qquad i = 1.2,...,n \tag{3.7}$$

biçiminde belirtilmektedir. Sıfır yığılmalı Poisson dağılımından gelen Y_i raslantı değişkeninin Eş. 3.2. ve Eş. 3.3.'ten yararlanılarak elde edilen beklenen değer ve varyansı sırasıyla,

$$E(Y_i|x_i) = (1-\omega_i)\mu_i \ ; \ Var(Y_i|x_i) = E(Y_i)(1+\omega_i\mu_i) \qquad i = 1.2,...,n \tag{3.8}$$

biçimindedir. Eş. 3.8.'de tanımlanan varyans eşitliği incelendiğinde, sıfır değerindeki yığılma nedeni ile oluşan aşırı yayılımın, $E(Y_i)\omega_i\mu_i$ terimi ile modellendiği görülmektedir.

ω_i ve μ_i parametrelerinin birbiriyle ilişkili olmadığı ve her iki parametre için de regresyon çözümlemesinin yapıldığı sıfır yığılmalı Poisson modelinde parametre tahmini için kullanılan log-olabilirlik fonksiyonu;

$$lnL(y;\gamma,\beta) = -\sum_{i=1}^{n} ln\left(1+e^{G_i\gamma}\right) + \sum_{i=1}^{n}\left\{\sum_{y_i=0} ln\left(e^{G_i\gamma}+e^{-\mu_i}\right) + \sum_{y_i>0}\left[y_i\, ln(\mu_i) - \mu_i - ln(y_i!)\right]\right\} \tag{3.9}$$

biçiminde elde edilmektedir. Açıklayıcı değişkenlerin yapısal sıfır oranı üzerindeki etkisi göz ardı edildiğinde Eş. 3.9. ile belirtilen log-olabilirlik fonksiyonunda $exp(G_i\gamma)$ terimi yerine $exp(\gamma)$ terimi kullanılmaktadır (Lambert, 1992; Famoye and Özmen, 2007; Özmen and Demirhan, 2010).

Sıfır yığılmalı Poisson modeli, verinin sıfır değerindeki yığılmayı açıklayan bir modeldir. Ancak aşırı yayılım problemi, bağımlı değişkenin sıfır değerindeki yoğunluğun yanı sıra, çalışmanın aşırı yayılım bölümünde belirtilen nedenlerden de kaynaklanabilmektedir. Bu durumda, sıfır yığılmalı modellerde temel sayı dağılımı için aşırı yayılımın da açıklandığı dağılımlardan birinin kullanımı uygun olmaktadır.

3.3. Sıfır Yığılmalı Negatif Binom Regresyon Modeli

Sıfır yığılmalı negatif binom regresyon modeli, aşırı yayılım sorununun yanı sıra sıfır değerinde aşırı yığılmanın gözlemlendiği hasar sayısı verisinin modellenmesi amacıyla sıfır yığılmalı Poisson regresyon modeline alternatif olarak 1994 yılında Greene tarafından geliştirilmiştir.

Sıfır yığılmalı Poisson modeline benzer biçimde, ω_i yapısal sıfır oranı olmak üzere sıfır yığılmalı negatif binom modeline uyan hasar sayısı raslantı değişkenine ilişkin olasılık fonksiyonu;

$$Pr\left(Y_i = 0\right) = \omega_i + \left(1-\omega_i\right)\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}$$

$$Pr\left(Y_i = y_i\right) = \left(1-\omega_i\right)\frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(y_i+1)\Gamma(\alpha^{-1})}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}} \quad ; \quad \begin{array}{l} y_i = 1,2,... \\ i = 1,2,...,n \\ 0 < \omega_i < 1 \end{array} \tag{3.10}$$

eşitlikleri ile tanımlanmaktadır. Eş. 3.10.'da yer alan α , negatif binom dağılımının yayılım parametresi ve $\mu_i \left(= exp(X_i\beta)\right)$ ortalama parametresidir $(\alpha, \mu_i > 0,$ $i = 1,2,...,n)$. Yayılım parametresinin değeri sıfıra eşit ise, aşırı yayılım yalnızca sıfır değerindeki yoğunluktan kaynaklanmaktadır. Bu durum, verinin sıfır yığılmalı Poisson modeli kullanılarak modellenmesinin yeterli olduğunu göstermektedir (Yip and Yau, 2005).

Sıfır yığılmalı Poisson modeline benzer biçimde yapısal sıfır oranı regresyon çözümlemesi yapılarak hesaplanabileceği gibi, yalnızca lojit dönüşüm kullanılarak da elde edilebilmektedir. Regresyon çözümlemesi yapılarak ω_i oranı;

$$lojit(\omega_i) = ln\left(\omega_i/(1-\omega_i)\right) = G_i\gamma \qquad i = 1.2,...,n \tag{3.11}$$

eşitliği ile elde edilmektedir. Yapısal sıfır oranı için kullanılan açıklayıcı değişkenlerden oluşan matris G ve bilinmeyen regresyon parametrelerinden oluşan vektör γ'dır. Açıklayıcı değişkenlerin etkisi göz ardı edildiğinde ise, bu oran için $lojit(\omega_i) = \gamma$ eşitliği kullanılmaktadır.

Sıfır yığılmalı Poisson modelinde olduğu gibi, sıfır yığılmalı negatif binom modelinde de ω_i ve μ_i parametreleri aynı açıklayıcı değişkenler kullanılarak tahmin edildiğinde; bu parametrelerin birbiriyle ilişkili olduğu düşünülerek, bu ilişki bilinmeyen gerçel değerli biçim parametresi τ ile açıklanabilmektedir. Bu durumda ω_i ve μ_i parametreleri Eş. 3.7. ile tanımlanmaktadır (Shankar et al., 1997).

Sıfır yığılmalı negatif binom dağılımına uyan Y_i raslantı değişkeninin Eş. 3.2. ve Eş. 3.3.'ten yararlanılarak elde edilen beklenen değer ve varyansı sırasıyla,

$$E(Y_i \mid x_i) = (1-\omega_i)\mu_i \quad ; \; Var(Y_i \mid x_i) = E(Y_i)(1+\alpha\mu_i + \omega_i\mu_i) \quad i = 1.2,...,n \tag{3.12}$$

biçimindedir. Varyans eşitliği incelendiğinde, temel sayı dağılımındaki aşırı yayılımın sıfır değerindeki yığılmayla ilgili olan bölümü $E(Y_i)\omega_i\mu_i$ terimi ile, diğer nedenlerden kaynaklanan bölümü $E(Y_i)(1+\alpha\mu_i)$ terimi ile açıklanmaktadır.

ω_i ve μ_i parametrelerinin birbiriyle ilişkili olmadığı ve her iki parametre için de regresyon çözümlemesinin yapıldığı sıfır yığılmalı negatif binom modelinde parametre tahmini için kullanılan log-olabilirlik fonksiyonu;

$$\begin{aligned} ln\,L(y;\gamma,\beta,\alpha) = &-\sum_{i=1}^{n} ln\left(1+e^{G_i\gamma}\right) + \sum_{i=1}^{n}\left\{ \sum_{y_i=0} ln\left[e^{G_i\gamma} + \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \right] \right. \\ &\left. + \sum_{y_i>0}\left[\sum_{t=0}^{y_i-1} ln(1+t\alpha) + y_i\, ln(\mu_i) - (y_i+\alpha^{-1})ln(1+\alpha\mu_i) - ln(y_i!) \right] \right\} \end{aligned} \tag{3.13}$$

biçiminde elde edilmektedir. Açıklayıcı değişkenlerin yapısal sıfır oranı üzerindeki etkisi göz ardı edildiğinde Eş. 3.13. ile belirtilen log-olabilirlik fonksiyonunda $exp(G_i\gamma)$ terimi yerine $exp(\gamma)$ terimi kullanılmaktadır (Yip and Yau, 2005; Özmen and Demirhan, 2010).

3.4. Sıfır Yığılmalı Genelleştirilmiş Poisson Regresyon Modeli

Sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli, gözlemlenen verinin aşırı yayılım veya az yayılım gösterdiği ve sıfır değerinde aşırı yığılmanın olduğu durumda kullanılmak üzere sıfır yığılmalı negatif binom modeline alternatif olarak 2006 yılında Famoye ve Singh tarafından geliştirilmiştir. Gözlemlenmiş veri kümeleri kullanılarak yapılan bazı modelleme çalışmalarında, sıfır yığılmalı negatif binom regresyon modeli için parametre tahminleri elde edilememiştir (Lambert, 1992; Famoye et al., 2004; Famoye and Özmen, 2007). Sıfır yığılmalı negatif binom regresyonunun parametre tahminlerinin elde edilemediği durumda sıfır yığılmalı genelleştirilmiş Poisson modeli kullanılmaktadır.

Tanımlanan diğer sıfır yığılmalı modellere benzer biçimde; ω_i yapısal sıfır oranı, φ ve μ_i genelleştirilmiş Poisson modelinin parametreleri olmak üzere sıfır yığılmalı genelleştirilmiş Poisson modeline uyan Y_i hasar sayısı raslantı değişkenine ilişkin olasılık fonksiyonu;

$$Pr\left(Y_i = 0\right) = \omega_i + \left(1-\omega_i\right) exp\left(\frac{-\mu_i}{\varphi}\right)$$

$$Pr\left(Y_i = y_i\right) = \left(1-\omega_i\right)\frac{\mu_i\left(\mu_i + (\varphi - 1)y_i\right)^{y_i - 1}}{y_i!}\varphi^{-y_i} exp\left(-\frac{1}{\varphi}\left(\mu_i + (\varphi-1)y_i\right)\right) \quad ; \quad \begin{array}{l} y_i = 1,2,... \\ i = 1,2,...,n \\ 0 < \omega_i < 1 \end{array}$$

(3.14)

eşitlikleri ile tanımlanmaktadır. Eş. 3.14.'te $\mu_i > 0$ olmak üzere $\mu_i = exp(X_i \beta)$ ile hesaplanmaktadır. Yayılım parametresinin değerinin birden küçük olduğu durumda $(\varphi < 1)$, $\mu_i + m(\varphi - 1) > 0$ koşulunu sağlayan en büyük m değeri için yayılım parametresinin alacağı değerin alt sınırı; $\varphi > max(1/2, 1 - \mu_i/m)$ biçiminde belirlenmektedir. Sıfır yığılmalı negatif binom modeline benzer biçimde sıfır yığılmalı genelleştirilmiş Poisson modelinin yayılım parametresi bire eşit ise, aşırı yayılım yalnızca sıfır değerindeki yoğunluktan kaynaklanmakta ve bu yayılım, yapısal sıfır oranıyla açıklanmaktadır. Bu durumda, temel sayı dağılımı için aşırı yayılım parametresi gerekli olmadığından verinin sıfır yığılmalı Poisson modeli kullanılarak modellenmesi yeterli olmaktadır (Czado et al., 2007; Rashwan and Kamel, 2011).

Yapısal sıfır oranı ω_i regresyon çözümlemesi ile, veya yalnızca lojit dönüşümün uygulandığı tek değişkenli çözümleme ile elde edilebilmektedir. Regresyon çözümlemesi yapılarak elde edilen ω_i oranı için Eş. 3.11.' de belirtilen tanım kullanılmaktadır. Açıklayıcı değişkenlerin yapısal sıfır oranı üzerindeki etkisi göz ardı edildiğinde ise bu oran, $lojit(\omega_i) = \gamma$ eşitliği kullanılarak tahmin edilmektedir.

Sıfır yığılmalı genelleştirilmiş Poisson dağılımına sahip Y_i raslantı değişkeninin Eş. 3.2. ve Eş. 3.3.'ten yararlanılarak elde edilen beklenen değer ve varyansı sırasıyla,

$$E(Y_i|x_i) = (1-\omega_i)\mu_i \;;\quad Var\left(Y_i|x_i\right) = E\left(Y_i\right)\left[\varphi^2 + \omega_i\mu_i\right] \quad i = 1.2,...,n \tag{3.15}$$

biçimindedir. Varyans eşitliğinde yer alan $E(Y_i)\varphi^2$ terimi, temel sayı dağılımında başka nedenlerden kaynaklanan aşırı yayılımı; $E(Y_i)\omega_i\mu_i$ terimi ise sıfır değerindeki yığılmadan kaynaklanan aşırı yayılımı açıklamaktadır.

ω_i ve μ_i parametrelerinin birbiriyle ilişkili olmadığı ve her iki parametre için de regresyon çözümlemesinin yapıldığı sıfır yığılmalı genelleştirilmiş Poisson modelinde parametre tahmini için kullanılan log-olabilirlik fonksiyonu;

$$\begin{aligned} \ln L\left(y;\gamma,\beta,\varphi\right) = \sum_{i=1}^{n} 1_{(y_i=0)}\left[\ln\left(e^{G_i\gamma} + \exp\left(-\mu_i/(1+\varphi)\right)\right) - \ln(1+e^{G_i\gamma})\right] \\ + 1_{(y_i>0)}\left[-\ln(1+e^{G_i\gamma}) + \ln\mu_i - \ln(y_i!) - y_i\ln(1+\varphi) + (y_i-1)\ln(\mu_i+\varphi y_i) - \frac{\mu_i+\varphi y_i}{\varphi}\right] \end{aligned} \tag{3.16}$$

biçiminde elde edilmektedir. $1_{(y_i=0)}$ ifadesi, parantez içinde belirtilen koşul sağlandığında bir, sağlanmadığında sıfır değerini almaktadır. Açıklayıcı değişkenlerin yapısal sıfır oranı üzerindeki etkisi göz ardı edildiğinde Eş. 3.16. ile belirtilen log-olabilirlik fonksiyonunda $exp(G_i\gamma)$ terimi yerine $exp(\gamma)$ terimi kullanılmaktadır (Yip and Yau, 2005; Czado et al, 2007; Özmen and Demirhan, 2010).

4. MODEL SEÇİMİ

4.1. Parametre Tahmini

En çok olabilirlik tahmin yöntemi, model parametrelerinin tahmin edilmesi ve istatistiksel sonuçların elde edilmesi amacıyla yaygın olarak kullanılan bir tahmin yöntemidir. Bu yöntem, tutarlı ve asimptotik olarak etkin tahminler vermektedir. En çok olabilirlik yöntemi ile elde edilen tahminler büyük örneklemler için normal dağılıma sahiptir (Harrell, 2001). Bu yöntemde öncelikle kurulan modele ilişkin olabilirlik fonksiyonunun logaritması alınarak elde edilen log-olabilirlik fonksiyonunun her bir parametreye göre kısmi türevi alınır. Kısmi türevlerin alınması sonucunda parametre sayısına eşit sayıda denklemler kümesi oluşturulur. Oluşturulan denklemlerin her biri sıfıra eşitlenerek tüm eşitlikler için eş zamanlı çözüm yapılır. Elde edilen parametre tahminleri, modelin en çok olabilirlik tahmin değerleri olarak ifade edilmektedir. Log-olabilirlik fonksiyonunun parametrelere göre birinci türevi, skor fonksiyonu (score function) olarak adlandırılmaktadır.

n sayıda gözlemden oluşan herhangi bir regresyon modelinde tahmin edilecek regresyon parametreleri vektörü $\beta' = (\beta_1, \beta_2, ..., \beta_k)$, bağımlı değişkenler vektörü $Y' = (Y_1, Y_2, ..., Y_n)$, her bir bağımlı değişkenin olasılık fonksiyonu ve olabilirlik fonksiyonu sırasıyla $f_i(Y_i;\beta)$ ve $L_i(\beta)$ olmak üzere; olabilirlik fonksiyonu ve bulunan olabilirlik fonksiyonunun logaritması alınarak elde edilen log-olabilirlik fonksiyonu sırasıyla;

$$L(\beta) = \prod_{i=1}^{n} f_i(Y_i;\beta) \qquad ; \qquad \ln L(\beta) = \sum_{i=1}^{n} \log L_i(\beta)$$

biçiminde tanımlanmaktadır. β'nın en çok olabilirlik tahmini, β'nın bir fonksiyonu olacak biçimde $\ln L(\beta)$ fonksiyonunu maksimize eden β vektörünün değerleridir. Bilinmeyen regresyon parametrelerinin en çok olabilirlik tahmin değerleri vektörü $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k)$ ile gösterilir. Parametrelerin en çok olabilirlik tahminlerinin elde edilmesi için genellikle adımsal deneme-yanılma yöntemleri gerekmektedir. Bu yöntemlerden Newton-Raphson yöntemi, EM (Estimation and Maximization) algoritması ve Fisher skorlama yöntemi yaygın olarak kullanılmaktadır. Newton-Raphson yönteminde, skor vektörünün sıfır değerini alması veya log-olabilirlik fonksiyonunun maksimize edilmesi amaçlanmaktadır (Faraway, 2006). Skor vektörü, $log L(\beta)$ fonksiyonunun $\beta_1, \beta_2, ..., \beta_k$ parametrelerine göre birinci türevi alınarak;

$$U(\beta)' = (\partial/\partial\beta_1 \ln L(\beta), ..., \partial/\partial\beta_k \ln L(\beta)) = (\partial/\partial\beta_k) \ln L(\beta)$$

biçiminde elde edilmektedir. Newton-Raphson yönteminde kullanılan diğer bir matris, Hessian matrisidir. Bu matris,

$$H(\beta) = \frac{\partial^2 \ln L(\beta)}{\partial\beta\partial\beta'}$$

biçiminde elde edilmektedir. Newton-Raphson yönteminde β_0, β parametrelerine ilişkin başlangıç değerleri olmak üzere her bir adımda;

$$\beta_{i+1} = \beta_i - H^{-1}(\beta_i)U(\beta_i) \quad i = 0,1,2,3,...$$

işlemi tekrarlanarak denklemler kümesi çözülmektedir.

Newton-Raphson yönteminde Hessian matrisi yerine Fisher bilgi matrisinin eksi işaretlisini kullanan adımsal tahmin yöntemi, Fisher skorlama yöntemidir. Tahmin edilen parametrelerin belirsizliğinin derecesinin düşürülerek gerçek değerlere yakın sonuçlar elde edilmesi için gerekli bilgi, Fisher bilgi fonksiyonu kullanılarak elde edilmektedir. Bu yöntem, adımsal olarak yeniden ağırlıklandırılan en küçük kareler yöntemine eşdeğer olup, genelleştirilmiş doğrusal modellerin parametre tahminlerinde yaygın olarak kullanılmaktadır (Faraway, 2006). k parametreli bir regresyon modelinde Fisher bilgi matrisi $\left(I^*(\beta)\right)$, $k \times k$ boyutludur. Bu matrisin elemanları, $log L(\beta)$ fonksiyonunun her bir parametreye göre ikinci kısmi türevinin beklenen değerinin eksi işaretlisi alınarak elde edilmektedir.

$$I^*(\beta) = -\left\{E\left[\left(\partial^2 \ln L(\beta)/\partial\beta_j\partial\beta_p\right)\right]\right\}_{k\times k} = -E\left\{\left(\partial^2/\partial\beta\partial\beta'\right)\ln L(\beta)\right\}$$

Fisher bilgi matrisi ile tahmin edilen $\hat{\beta}$ parametrelerinin varyansı arasında;

$$Var(\hat{\beta}) = I^{*-1}(\hat{\beta})$$

ilişkisi mevcuttur. Fisher bilgi matrisinin beklenen değerinin alınmadığı durumda gözlenmiş bilgi matrisi (observed information matrix) elde edilmektedir. Gözlenmiş Fisher bilgi matrisi;

$$I(\beta) = -\left(\partial^2 / \partial\beta\partial\beta'\right) \ln L(\beta)$$

biçiminde ifade edilmektedir (Harrell, 2001). Fisher bilgi matrisinde β parametrelerinin yerine en çok olabilirlik tahmin değerleri olan $\hat{\beta}$ değerlerinin yazılması ile elde edilen matris ise beklenen bilgi matrisidir. Bu matris, bazı hipotez testlerinde hesaplamada kolaylık sağlaması nedeniyle kullanılabilmektedir (Cavanaugh and Shumway, 1996).

4.2. Aşırı Yayılım Testleri

Aşırı yayılım, Poisson regresyon modelinin varsayımlarından biri olan eşit yayılımın (equidispersion) sağlanmadığı durumda ortaya çıkmaktadır. Verinin aşırı yayılımlı olup olmadığına karar vermek için izlenen yollardan biri, kurulan Poisson modelinin varyansının ortalamasına oranlanmasıdır. Bu oran 1'den büyük ise aşırı yayılım, küçük ise az yayılım mevcuttur. Aşırı yayılım sorununun belirlenmesinde kullanılan bir diğer yöntem, Poisson yaklaşımının test edilmesi amacıyla çeşitli test istatistiklerinin hesaplanmasıdır.

Gözlem sayısı n ve i'nci gözlem için bağımlı değişken Y_i olmak üzere, büyük örneklemler için önerilen ve kısmi skor testi (partial score test) olarak adlandırılan test istatistiği;

$$T = \frac{1}{2}\sum_{i=1}^{n}\left\{\left(Y_i - \hat{\mu}_i\right)^2 - Y_i\right\} \tag{4.1}$$

biçiminde verilmektedir.

Standart normal dağılıma yakınsayan ve T istatistiğinin standartlaştırılmış biçimi olarak ifade edilen başka bir test istatistiği ise;

$$T_1 = \frac{\sum_{i=1}^{n} \left\{ \left(Y_i - \hat{\mu}_i \right)^2 - Y_i \right\}}{\left(2 \sum_{i=1}^{n} \hat{\mu}_i^2 \right)^{1/2}} \tag{4.2}$$

biçiminde tanımlanmaktadır.

T ve T_1 test istatistiklerindeki $\hat{\mu}_i$ terimi, Poisson regresyon modeli altında μ_i'nin en çok olabilirlik tahminini göstermektedir. T istatistiğinin pozitif büyük değerler alması verinin Poisson modeline göre aşırı yayılım, negatif büyük değerler alması ise verinin Poisson modeline göre az yayılım gösterdiğini ifade etmektedir. T_1 test istatistiği, yeterince büyük gözlem sayısı için standart normal dağılıma yakınsadığından bu istatistik, belirlenen yanılma düzeyi için standart normal dağılım tablo değeri ile karşılaştırılarak aşırı yayılımın varlığı test edilmektedir (Dean and Lawless, 1989).

Poisson yaklaşımını veya yayılım durumunu test etmek amacıyla geliştirilen diğer test istatistikleri ise log-olabilirlik oran istatistiğine dayanmaktadır. Daha önce de değinildiği üzere, negatif binom modelinin yayılım parametresinin (α) sıfıra ve genelleştirilmiş Poisson modelinin yayılım parametresinin (φ) bire eşit olması, gözlemlenen verinin aşırı yayılım göstermediğini, dolayısıyla modellemede Poisson regresyonunun kullanılmasının yeterli olacağını belirtmektedir. Bu durumda her iki modele ait yayılım parametrelerinin belirtilen değerlere eşitliği hipotez testleri kullanılarak test edilmektedir.

Negatif binom modelinin yayılım parametresinin anlamlılığı için kullanılan yokluk hipotezi ve seçenek hipotez sırasıyla;

$$H_0 : \alpha = 0 \quad ; \quad H_1 : \alpha > 0$$

biçiminde tanımlanmaktadır.

Genelleştirilmiş Poisson modelinin yayılım parametresinin anlamlılığı için kullanılan yokluk hipotezi ve seçenek hipotez;

$$H_0 : \varphi = 1 \quad ; \quad H_1 : \varphi > 1$$

biçimindedir. Yokluk hipotezlerinin reddedilmesi, yayılım parametrelerinin anlamlı olduğunu, diğer bir ifadeyle Poisson modeli yerine negatif binom modeli veya genelleştirilmiş Poisson modelinin kullanılmasının uygun olduğunu göstermektedir. Bu hipotezler, sıfır yığılmalı Poisson modelinin, aşırı yayılımı da kapsayan sıfır yığılmalı modellere göre etkinliğini ölçmek için de kullanılmaktadır. Yokluk hipotezlerini test etmek amacıyla kullanılan hipotez testleri aşağıdaki gibidir:

- Olabilirlik Oran Testi: Olabilirlik oran test istatistiği; test edilen parametre için yokluk hipotezinin doğru olduğu varsayımı altında hesaplanan olabilirlik fonksiyonunun (L_0), seçenek hipotezin doğru olduğu varsayımı altında test edilen parametrenin en çok olabilirlik tahmin değeri kullanılarak elde edilen olabilirlik fonksiyonuna (L_1) oranının logaritması alınarak hesaplanmaktadır. Özetle olabilirlik oran test istatistiği;

$$LR = -2\ln(L_0 / L_1) \tag{4.3}$$

eşitliği kullanılarak hesaplanmaktadır (Wang and Famoye, 1997). Yalnızca bir parametrenin anlamlılığı test edildiğinde, yeterince büyük örneklemler için, olabilirlik oran test istatistiği serbestlik derecesi 1 olan ki-kare dağılımına sahiptir.

- Wald Testi: Bilinmeyen parametre θ ile ifade edildiğinde; Wald testinde, test edilen parametrenin en çok olabilirlik tahmin değeri $(\hat{\theta})$ ile yokluk hipotezindeki değeri (θ_0) karşılaştırılmaktadır. İki değer arasındaki farkın normal olduğu varsayımı kullanılmaktadır. Wald test istatistiği;

$$W = \frac{(\hat{\theta} - \theta_0)^2}{Var(\hat{\theta})} \tag{4.4}$$

biçiminde hesaplanmaktadır. Olabilirlik oran testine benzer biçimde, Wald test istatistiği de 1 serbestlik dereceli ki-kare dağılımına sahiptir.

- Skor Testi: Test edilen parametrenin en çok olabilirlik tahmin değeri $(\hat{\theta})$, yokluk hipotezinde test edilen değere (θ_0) eşit olduğunda yokluk hipotezindeki parametre değeri için skor fonksiyonunun değeri sıfıra eşittir ve log-olabilirlik fonksiyonu en yüksek değerine ulaşmaktadır (Harrell, 2001). Skor istatistiği, yokluk hipotezi doğru kabul edildiğinde elde edilen skor fonksiyonunun, sıfır değerinden uzaklığını ölçmek için kullanılmaktadır. Skor istatistiği;

$$S = \frac{U(\theta_0)^2}{I(\theta_0)} \tag{4.5}$$

eşitliği kullanılarak hesaplanmaktadır. Eşitlikte yer alan U ve I fonksiyonları sırasıyla skor ve gözlenmiş bilgi fonksiyonlarıdır. Skor testini diğer hipotez testlerinden ayıran bir özelliği, test edilen parametrenin en çok olabilirlik tahmin değerinin kullanılmamasıdır. Yokluk hipotezi doğru kabul edildiğinde skor test istatistiği de 1 serbestlik dereceli ki-kare dağılımına sahiptir.

Hipotez testi için kullanılan üç test de birbirine yakın sonuçlar vermektedir (Wang and Famoye, 1997). Ancak söz konusu testler, test istatistiklerinin hesaplanması açısından farklılık göstermektedir. Olabilirlik oran testi ve Wald testinde, hem yokluk hipotezinde hem de seçenek hipotezde belirtilen modeller uygulanarak parametre tahminlerinin elde edilmesi gerekmektedir. Skor testinde ise yalnızca yokluk hipotezindeki parametre değeri ile ilgilenilmektedir. Skor testinde en çok olabilirlik tahmin değerinin kullanılmaması nedeniyle, olabilirlik oran testi ve Wald testine göre daha yaygın kullanılmaktadır (Harrell, 2001).

4.3. Uyum İyiliğinin Ölçümü

Bir modelden elde edilen sonuçlar modelin geçerliliğine bağlıdır. Gözlemlenen veri kümesinin modellenmesi amacıyla çok sayıda model belirlenebilmektedir. Belirlenen her bir modelin gözlemlenen veri kümesi ile uygunluğunun test edilmesi, istatistiksel modellemenin önemli bir parçasını oluşturmaktadır. Uyum iyiliği testlerinde temel amaç, gözlemlenen değerler ile tahmini değerler arasındaki farkın ölçülmesidir. Bu amaçla; genelleştirilmiş Pearson ki-kare $\left(\chi^2\right)$, sapma, olabilirlik oran gibi uyum iyiliği istatistikleri kullanılmaktadır.

Gözlemlenen verinin sıfır değerindeki yoğunluğun Poisson regresyon modeline göre fazla olduğu durumda sıfır yığılmalı Poisson regresyon modelinin uygunluğunun test edilmesi amacıyla, 1995 yılında Van den Broek tarafından geliştirilen skor testi kullanılmaktadır (Van den Broek, 1995).

Çalışmanın bu bölümünde sırasıyla skor testi, genelleştirilmiş Pearson ki-kare test istatistiği ve olabilirlik oran test istatistiği açıklanmıştır.

4.3.1. Skor testi

Sıfır yığılmalı modellerin geliştirilmesi ve birçok farklı alanda kullanılması ile birlikte, bu modellerin klasik sayı modellerine göre etkinliğinin ölçülmesi amacıyla çok sayıda test istatistiği geliştirilmiştir. Yapılan çalışmalarda öncelikle Poisson regresyon modeli ile bu modele seçenek olarak tanımlanan sıfır yığılmalı Poisson regresyon modelinin karşılaştırıldığı testler kullanılmıştır. Sıfır yığılmalı negatif binom modelinin geliştirilmesi ile birlikte bu modelin, negatif binom modeli ile karşılaştırılması amacıyla Vuong testinin kullanılabilirliği Greene (1994) ve Shankar et al. (1997) tarafından açıklanmıştır. Vuong testi, 1989 yılında Vuong tarafından geliştirilen ve iç içe geçmeyen (non-nested) modellerin karşılaştırılmasında kullanılan hipotez testlerinden biridir. Sıfır yığılmalı genelleştirilmiş Poisson modelinin literatüre girmesiyle birlikte, bu modelin genelleştirilmiş Poisson modeli ile karşılaştırıldığı, genellikle skor testine dayanan hipotez testleri de geliştirilmiştir (Famoye and Singh, 2006; Zhao, 2006).

Gözlemlenen verinin sıfır yığılmalı modellerden biri kullanılarak modellenmesi için, öncelikle Poisson modeli ile sıfır yığılmalı Poisson modeli karşılaştırılmaktadır. Bu karşılaştırma, aynı zamanda yapısal sıfır oranının değerinin sıfıra eşitliğinin test edilmesini ifade etmektedir. İki modelin karşılaştırıldığı yokluk hipotezi ve seçenek hipotez;

$H_0 : \omega_i = 0 \quad ; \quad H_1 : \omega_i > 0$

biçiminde ifade edilmektedir. Sıfır oranı için açıklayıcı değişkenler kullanılarak lojit regresyon çözümlemesinin yapılmadığı ve açıklayıcı değişkenlerin yalnızca Poisson regresyon modelinin ortalama parametresini etkilediği durumda yokluk hipotezi ve seçenek hipotezde yer alan bu oran ω ile ifade edilmektedir. ω oranının test edilmesi amacıyla 1995 yılında Jan Van den Broek tarafından geliştirilen skor testi kullanılmaktadır. Test istatistiği;

$$S(\hat{\beta}) = U'(\hat{\beta},0)\left[J(\hat{\beta},0)\right]^{-1} U(\hat{\beta},0) \qquad (4.6)$$

biçiminde hesaplanmaktadır (Van den Broek, 1995). Eş. 4.6.'da $\hat{\beta}$, $U(\hat{\beta},0)$ ve $J(\hat{\beta},0)$ sırasıyla Poisson regresyon modeli için parametre tahmin değerleri, sıfır oranı sıfır değerini aldığında elde edilen skor vektörü ve beklenen bilgi matrisini göstermektedir.

Yapısal sıfır oranı için açıklayıcı değişkenlerin kullanılmadığı ve bu oranın sabit kabul edildiği durumda Eş. 4.6. ile tanımlanan skor test istatistiği;

$$S(\hat{\beta}) = \left\{\sum_{i=1}^{n} \frac{1_{(y_i=0)} - \hat{p}_{0i}}{\hat{p}_{0i}}\right\}^2 \Bigg/ \left\{\sum_{i=1}^{n} \frac{1-\hat{p}_{0i}}{\hat{p}_{0i}}\right\} - n\bar{y} \qquad (4.7)$$

eşitliği ile hesaplanmaktadır. Poisson modelinin ortalama parametresinin tahmin değeri $\hat{\mu}_i$ olmak üzere, eşitlikte yer alan $\hat{p}_{0i}$ olasılığı $exp(-\hat{\mu}_i)$ biçiminde hesaplanmaktadır. Eşitliğin pay kısmında yer alan $1_{(y_i=0)}$ ifadesi, parantez içinde belirtilen koşul sağlandığında bir, sağlanmadığında sıfır değerini almaktadır. Eşitlikte n gözlem sayısı ve $\bar{y}$, gözlemlenen bağımlı değişken değerlerinin ortalamasıdır. Yokluk hipotezinin doğru olduğu varsayımı altında hesaplanan skor test istatistiği, 1 serbestlik dereceli ki-kare dağılımına sahiptir (Flynn and Francis, 2009).

Van den Broek tarafından geliştirilen skor testine benzer biçimde, farklı durumlar için hesaplanan ve skor testine dayanan test istatistikleri de mevcuttur. Deng ve Paul (2000), sıfır yığılmalı Binom modeline ilişkin skor testini geliştirmiştir. Poisson regresyonunda bilinmeyen etki terimlerini de kullanarak bu modeli sıfır yığılmalı Poisson regresyon modeli ile karşılaştıran skor testi, Lee et al. tarafından 2001 yılında geliştirilmiştir. Yapısal sıfır oranı için lojit bağ fonksiyonu yerine özdeş bağ fonksiyonunun kullanıldığı ve bu oran için de regresyon çözümlemesinin yapıldığı sıfır yığılmalı Poisson modelinin Poisson modeli ile karşılaştırıldığı skor testi, Jansakul ve Hinde (2002) tarafından geliştirilmiştir. Ugarte et al. (2004) ise, sıfır yığılmalı Poisson modelinin uygunluğunun test edilmesi amacıyla geliştirdikleri skor testini, hastalık haritalarının çıkarılması alanında kullanmıştır.

4.3.2. Genelleştirilmiş Pearson ki-kare istatistiği

Belirlenen herhangi bir modelin gözlemlenen veri kümesine uygunluğunu ölçmek için kullanılan diğer bir uyum iyiliği istatistiği genelleştirilmiş Pearson ki-kare istatistiğidir. Y_i i'nci gözleme ilişkin gözlemlenen bağımlı değişken değeri, $\hat{\mu}_i$ uygunluğu test edilen model kullanılarak kestirilen ortalama parametresi olmak üzere; incelenen regresyon modelleri için ki-kare istatistikleri;

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{Var(\hat{y}_i)} \tag{4.8}$$

biçiminde hesaplanmaktadır (McCullagh and Nelder, 1989).

Gözlemlenen ve uygunluğu araştırılan model kullanılarak kestirilen bağımlı değişken değerleri küçük olduğunda bu istatistik büyük bir değer olarak hesaplanmaktadır.

4.3.3. Olabilirlik oran test istatistiği

Olabilirlik oran test istatistiği, gözlemlenmiş veriye uyması beklenen herhangi bir modelin veriye uygunluğunu test etmek amacıyla hesaplanmaktadır. Olabilirlik oran test istatistiğinde uygunluğu araştırılan model, tek parametreli basit model ile karşılaştırılarak "Uygunluğu araştırılan model ile basit model arasında fark yoktur." biçiminde kurulan yokluk hipotezi test edilmektedir. Uygunluğu araştırılan model M ile ifade edildiğinde bu test istatistiği;

$$LR = -2\left(\ln L(\text{basit model}) - \ln L(M)\right) \tag{4.9}$$

eşitliği kullanılarak hesaplanmaktadır. M modelinin parametre sayısı f ve basit modelin parametre sayısı s olarak belirlendiğinde bu test istatistiği, belirli bir yanılma düzeyinde $(f-s)$ serbestlik dereceli ki-kare değeri ile karşılaştırılmaktadır.

Bu test istatistiği, iç içe geçen iki modelin karşılaştırılması amacıyla da kullanılabilmektedir.

4.4. Model Seçim Ölçütleri

Gözlemlenmiş veriye uyması muhtemel modellerin veriye uygunluğu uyum iyiliği testleri kullanılarak araştırılmaktadır. Uyum iyiliği testleri neticesinde veriye uygun olduğu kanıtlanan modellerden oluşan kümeden veriye en uygun modelin seçilmesi, modelleme çalışmalarının son aşamasını oluşturmaktadır. Gözlemlenen veriye en uygun modelin seçilmesi amacıyla kullanılan başlıca ölçütler; Akaike bilgi ölçütü (AIC), Bayes bilgi ölçütü (BIC) ve log-olabilirlik değerleridir.

Akaike bilgi ölçütü ve Bayes bilgi ölçütü, model için kullanılan açıklayıcı değişkenlerin doğru ve yeterli sayıda belirlenerek modele hiçbir katkıda bulunmayan veya modelin veriye uyumunu olumsuz yönde etkileyen açıklayıcı değişkenlerin modelden çıkartılması amacıyla da kullanılmaktadır. Herhangi bir açıklayıcı değişken istatistiksel olarak anlamlı olmasına rağmen, modele eklendiğinde uyum derecesini düşürebilmektedir. Açıklayıcı değişken sayısı artırıldığında parametre sayısı (k) da artarken, gözlemlenen bağımlı değişken değeri ile model kullanılarak tahmin edilen bağımlı değişken değeri arasındaki fark azalmaktadır. Diğer yandan modeldeki parametre sayısı arttıkça, tahmin edilen parametrelerin varyansı artarken modelin uyum derecesi ise azalmaktadır. Kurulan modelin, yeterli sayıda açıklayıcı değişken kullanılarak ve optimal uyum sağlanarak oluşturulması hedeflenmektedir. Akaike ve Bayes bilgi ölçütleri, modellerin parametre sayısı ile uyum iyiliği arasında dengenin kurulması amacıyla kullanılmaktadır (De Jong and Heller, 2008).

Model seçim ölçütleri, gözlemlenen veriye uyum gösteren modellerin en çok olabilirlik değerleri kullanılarak hesaplanmaktadır. n gözlem sayısını, k modeldeki parametre sayısını ve $\ln L$ model için hesaplanan log-olabilirlik fonksiyon değerini göstermek üzere Akaike ve Bayes bilgi ölçütleri sırasıyla;

$$AIC = -2\ln L + 2k \tag{4.10}$$

$$BIC = -2\ln L + k\ln n \tag{4.11}$$

eşitlikleri kullanılarak hesaplanmaktadır (De Jong and Heller, 2008).

Gözlemlenen veri için seçilen en uygun model; Akaike, Bayes bilgi ölçütleri ile genelleştirilmiş Pearson ki-kare istatistiğinin en küçük ve log-olabilirlik fonksiyonunun en büyük değeri aldığı modeldir.

5. UYGULAMA

Çalışmanın uygulama bölümünde, çözümlemede kullanılan veri setinin genel yapısı incelenerek hasar sayısının modellenmesi amacıyla kullanılan Poisson, negatif binom, genelleştirilmiş Poisson regresyon modelleri ile bu modellerin sıfır yığılmalı karmalarının gözlemlenen veri setine uygunluğu araştırılmıştır. Çözümleme sürecinde; verinin incelenmesi ve düzenlenmesi aşamasında Clementine 12.0 programı, parametre ve model tahminlerinde R 2.12.2. programı, görsel ögelerin oluşturulmasında ise Excel programı kullanılmıştır.

5.1. Veri Kaynağı ve Yapısı

Trafik Sigortaları Bilgi Merkezi'nden (TRAMER) alınan veri kümesi, 2008 yılında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptıran otomobil sahibi sigortalılara ve araçlarına ilişkin bilgileri içermektedir. Veri kümesi, sigortalıların yaş ve cinsiyet bilgileri ile aracın kullanım tipi, il plaka kodu, sigorta poliçesinin başlangıç tarihi ve yıl içinde gerçekleşen hasar sayısı bilgisini içermektedir. Veri kümesinde toplam 7.132.960 adet poliçe vardır. Ancak bazı poliçelerde eksik bilgi olduğu görülmüş ve bu poliçeler veri kümesinden çıkarıldığında 2008 yılı için 3.699.157 sigorta poliçesi kalmıştır.

Veri kümesinin poliçe başlangıç tarihi bilgisi incelendiğinde, yıl içinde farklı zamanlarda düzenlenen poliçeler olduğu görülmektedir. Uygulamanın bir yılını tamamlamış poliçeler üzerinden yapılması amacıyla, yalnızca 2008 yılının ocak ayında başlayan poliçelerin dikkate alınmasının uygun olacağı düşünülmüş ve bir yıllık sürede gerçekleşen hasar sayısı, sigortalının yaşı, cinsiyeti, aracın plaka il kodu ve kullanım tipi bilgilerini içeren toplam 289.934 poliçe ile çalışılmıştır.

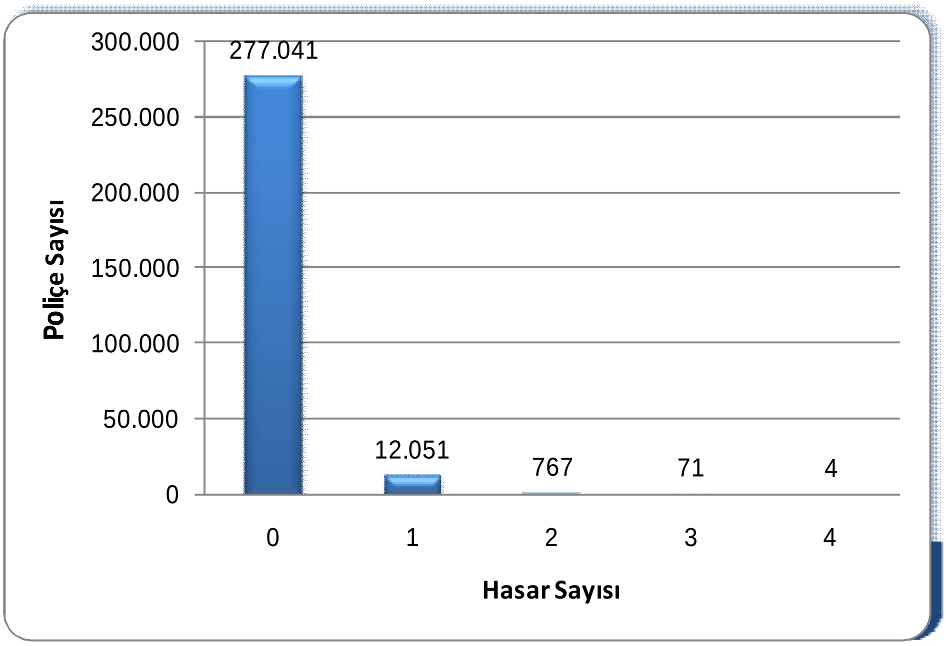
2008 yılı içinde gerçekleşen hasar sayısı bağımlı değişken olmak üzere, regresyon çözümlemesinde kullanılan açıklayıcı değişkenler yaş, cinsiyet, kullanım tipi ve sektör payıdır. Yaş, cinsiyet ve kullanım tipi değişkenlerinin belirlenen her bir kategori için gözlem sayıları ve yüzdeleri Tablo 5.1.'de özetlenmiştir. Her bir ilde 2008 yılının ocak ayında yürürlüğe giren poliçe sayısı ve tüm illerde yürürlüğe giren poliçeler içindeki yüzdesi (sektör payı) Ek.1' de verilmiştir.

Tablo 5.1. Yaş, cinsiyet ve kullanım tipi değişkenlerine ilişkin özet tablo

Cinsiyet		Kullanım Tipi		Yaş		
Kadın	**Erkek**	**Özel**	**Ticari**	**18-29**	**30-44**	**45-58**
47.630 (%16,43)	242.304 (%83,57)	289.880 (%99,98)	54 (%01,86)	42.309 (%14,60)	139.537 (%48,12)	108.088 (%37,28)

Tablo 5.1.'de görüldüğü üzere 2008 yılının ocak ayında sigorta yaptıran araç sahiplerinin cinsiyet dağılımına bakıldığında; 47.630 sigortalının kadın, 242.304 sigortalının ise erkek olduğu görülmüştür. Kullanılan araçlar, kullanım tipi bakımından özel ve ticari olmak üzere ikiye ayrılmaktadır. Buna göre; araçların yaklaşık olarak %99' u özel, %1' i ticari amaçla kullanılmaktadır. Araç sahiplerinin yaş gruplarına göre dağılımı incelendiğinde, sigortalıların 18 ile 58 yaşları arasında olduğu görülmüştür. Bireylerin yaş ortalaması 41 olmakla birlikte, büyük bölümü 30 yaş ve üzerindedir.

Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası kapsamında bir yıl içinde gerçekleşen hasar sayılarına ilişkin çubuk grafiği ve hasar sayılarına göre poliçe adetlerinin dağılımı sırasıyla Şekil 5.1. ve Tablo 5.2.'de verilmiştir.



Şekil 5.1. Hasar sayısı dağılımı

Tablo 5.2. Hasar sayılarına göre poliçe sayıları

Hasar Sayısı	Poliçe Sayısı	Portföy İçindeki Payı (%)
0	277.041	95,553
1	12.051	4,156
2	767	0,265
3	71	0,025
4	4	0,001

Şekil 5.1. ve Tablo 5.2.'de görüldüğü üzere 2008 yılının ocak ayında yürürlüğe giren poliçelerden bir yıl içinde bir hasar gerçekleşen poliçe sayısı 12.051, iki hasar gerçekleşen poliçe sayısı 767, üç hasar gerçekleşen poliçe sayısı 71 ve dört hasar gerçekleşen poliçe sayısı 4'tür. Hasar olmayan poliçe sayısı 277.041 olup, veri kümesinin %95'ini oluşturmaktadır. Bir yıl içinde gerçekleşen hasar sayısının ortalaması 0,0476 ve varyansı 0,052'dir.

Uygulamada öncelikle regresyon modelleri tüm açıklayıcı değişkenler kullanılarak oluşturulmuş, ancak bu regresyon modellerinde yaş ve sektör payı değişkenleri %10 yanılma düzeyinde anlamlı, cinsiyet ve kullanım tipi değişkenleri ise %10 yanılma düzeyinde anlamsız değişkenler olarak bulunmuştur. Prim hesaplamalarında genellikle önemli bir değişken olan cinsiyet faktörünün, tüm değişkenlerin kullanıldığı regresyon modellerinde, anlamlı olmadığı görülmüştür. Regresyon modellerinde cinsiyet faktörünün de dikkate alınması amacıyla açıklayıcı değişkenler kümesi, iki gruba ayrılmıştır. Belirlenen ilk açıklayıcı değişkenler kümesinde; sigortalıların yaşı, cinsiyeti ve araçların kullanım tipi değişkenleri yer almaktadır. İkinci kümede ise sigortalıların yaşı ve cinsiyetinin yanı sıra, il bilgisinin hasar sayısı üzerindeki etkisinin incelenmesi amacıyla, her bir aracın plaka il kodu değişkeni yerine ilgili ilin 2008 yılı için TRAMER tarafından belirlenen sektör payı değişkeni yer almaktadır. Regresyon modelleri iki ayrı açıklayıcı değişken kümesi için elde edilmiş ve iki küme için bulunan sonuçlar ayrı ayrı yorumlanmıştır. Yaş, cinsiyet ve kullanım tipi değişkenleri ile kurulan modeller, Bölüm 5.2.'de açıklanmıştır. Yaş, cinsiyet ve sektör payı değişkenleri ile kurulan modeller, Bölüm 5.3.'te açıklanmıştır.

Çözümleme sürecinde öncelikle incelenen tüm modeller, gözlemlenen veri kümesine uygulanmış ve her bir model için anlamsız bulunan açıklayıcı değişkenler belirlenmiştir. Anlamlı bulunmayan değişkenler ile regresyon modeli oluşturulması uygun olmayacağından her bir model anlamlı bulunan değişkenler kullanılarak elde edilmiştir. Anlamlı bulunan değişkenlerin olduğu regresyon modellerinin elde edilmesi amacıyla geriye doğru ve ileriye doğru seçim yöntemi uygulanabilmektedir. Geriye doğru seçim yönteminde tüm açıklayıcı değişkenler modele alınarak, anlamsız bulunan değişkenler sırayla modelden çıkartılmaktadır. İleriye doğru seçim yönteminde ise hiçbir açıklayıcı değişkenin olmadığı basit model kurularak, açıklayıcı değişkenler anlamlılık düzeylerine göre sırayla modele dahil edilmektedir. Anlamlı bulunan açıklayıcı değişkenlerle oluşturulan regresyon modellerinin gözlemlenen veri kümesine uygunluğu, olabilirlik oran testi ile test edilmiştir. Aşırı yayılım sorununun araştırılması ve sıfır değerindeki yoğunluğun Poisson regresyonu varsayımı altında beklenen yoğunlukla karşılaştırılması amacıyla sırasıyla T , T_1 ve skor test istatistikleri hesaplanmıştır. Son olarak, model seçim ölçütleri dikkate alınmış ve veri kümesine uygun bulunan modeller içinden en uygun model belirlenmiştir.

5.2. Birinci Model Sonuçları

Sigortalı bireylerin yaş ve cinsiyet bilgileri ile araçların kullanım tipi değişkenleri kullanılarak oluşturulan regresyon modellerine ilişkin sonuçlar Tablo 5.3.'te verilmiştir.

Tablo 5.3. Yaş, cinsiyet ve kullanım tipi değişkenleri ile kurulan regresyon modelleri

Açıklayıcı Değişkenler	μ Regresyonu					
	Poisson	NB	GP	ZIP	ZINB	ZIGP
Sabit terim	β_1: -2,3225 (0,0399) p<0,1	β_1: -2,3276 (0,0419) p<0,1	β_1: -2,3225 (0,0412) p<0,1	β_1: -1,2882 (0,1253) p<0,1	β_1: -2,2178 (0,0895) p<0,1	β_1: -1,7946 (0,2809) p<0,1
Yaş	β_2: -0,0163 (0,0008) p<0,1	β_2: -0,0162 (0,0009) p<0,1	β_2: -0,0163 (0,0009) p<0,1	β_2: -0,0145 (0,0025) p<0,1	β_2: -0,0165 (0,0097) p<0,1	β_2: -0,0174 (0,0029) p<0,1
Cinsiyet	β_3: -0,0838 (0,0222) p<0,1	β_3: -0,0844 (0,0233) p<0,1	β_3: -0,0838 (0,0229) p<0,1	β_3: -0,1517 (0,0818) p<0,1	β_3: -0,1819 (0,0769) p<0,1	β_3: -0,4137 (0,1824) p<0,1
Kullanım Tipi	β_4: -0,5121 (0,7072) p>0,1	β_4: -0,4973 (0,7284) p>0,1	β_4: -0,5121 (0,7283) p>0,1	β_4: -0,5394 (0,7320) p>0,1	β_4: -0,5290 (0,7328) p>0,1	β_4: -0,7225 (1,3587) p>0,1
Yayılım parametresi	---	α: 0,5125 (0,0263) p<0,1	φ: 1,0477 (0,0024) p<0,1	---	α: 0,5281 (0,0016) p<0,1	φ: 1,0404 (0,0073) p<0,1
Açıklayıcı Değişkenler	**ω Regresyonu**					
Sabit terim	---	---	---	γ_1: 0,6001 (0,1831) p<0,1	γ_1: -1,4926 (0,8782) p<0,1	γ_1: -0,1624 (0,6634) p>0,1
Yaş	---	---	---	γ_2: 0,0026 (0,0037) p>0,1	γ_2: -0,0205 (0,0192) p>0,1	γ_2: -0,0082 (0,0106) p>0,1
Cinsiyet	---	---	---	γ_3: -0,1024 (0,1190) p>0,1	γ_3: -8,9912 (49,3025) p>0,1	γ_3: -1,3069 (1,3916) p>0,1
Kullanım Tipi	---	---	---	γ_4: -5,8925 (35,9859) p>0,1	γ_4: -3,8677 (33,3361) p>0,1	γ_4: -2,0503 (26,4686) p>0,1

*Parantez içinde yer alan değerler, parametrelerin standart sapmasını göstermektedir.

Hasar sayısının bağımlı değişken, sigortalı bireylerin yaş ve cinsiyet bilgileri ile sigortalı araçların kullanım tipi değişkenlerinin açıklayıcı değişkenler olarak belirlendiği regresyon modellerinde her bir model için açıklayıcı değişkenlerin katsayılarının anlamlılığı Wald tipi test istatistiği kullanılarak test edilmiş ve %10 yanılma düzeyine göre anlamlılıkları belirtilmiştir. Buna göre kurulan tüm modeller için kullanım tipi değişkeninin anlamlı olmadığı sonucuna ulaşılmıştır. Ayrıca sıfır yığılmalı regresyon modelleri incelendiğinde, yapısal sıfır oranının regresyonunda anlamlı bulunan açıklayıcı değişkenler birbirinden farklı elde edilmiştir.

Sıfır yığılmalı modellerde temel sayı dağılımı için kullanılan açıklayıcı değişkenler ile yapısal sıfır oranı regresyonu için kullanılan açıklayıcı değişkenler birbirini etkilemekte ve katsayıların anlamlılığı da buna bağlı olarak değişmektedir. Herhangi bir açıklayıcı değişkenin yapısal sıfır oranı üzerindeki etkisinin incelenmesi amacıyla geriye doğru seçim yöntemi uygulanmalıdır.

Tablo 5.3.'te de görüldüğü üzere sıfır yığılmalı genelleştirilmiş Poisson regresyon modelinin yapısal sıfır oranı için sabit terim de dahil olmak üzere hiçbir açıklayıcı değişken anlamlı bulunmamıştır. Bu model için de geriye doğru seçim yöntemi uygulanmasının ardından öncelikle olabilirlik oran testi hesaplanmış, daha sonra sıfır yığılmalı genelleştirilmiş Poisson regresyon modelinin genelleştirilmiş Poisson regresyon modeli ile karşılaştırılması amacıyla Vuong testi uygulanarak yapısal sıfır oranının anlamlılığı test edilmiştir. Olabilirlik oran testine göre sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli, basit modele göre anlamlı bulunmuş ancak yapısal sıfır oranının sıfıra eşitliğinin test edildiği Vuong testi sonucunda sıfır yığılmalı genelleştirilmiş Poisson modeli yerine genelleştirilmiş Poisson regresyon modelinin kullanılmasının uygun olacağı sonucuna ulaşılmıştır. Sıfır yığılmalı genelleştirilmiş Poisson modeline ilişkin olabilirlik oran testi ve bu modelin genelleştirilmiş Poisson modeli ile karşılaştırıldığı Vuong testi Tablo 5.4.'te özetlenmiştir.

Tablo 5.4. ZIGP olabilirlik oran testi ve ZIGP - GP Vuong testi

ZIGP için LR testi

lnL(basit model)	lnL(ZIGP)	Serbestlik derecesi	LR	p
-56119,29	-55960,4	4	317,78	<0,10

ZIGP-GP Vuong Testi

vuong (gp,zigp)

nu	p
7,68	<0,10

Favour model 1

Tüm regresyon modellerinden öncelikle kullanım tipi değişkeni çıkartılmış ve geriye doğru seçim yöntemi uygulanmıştır. Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerinin tümü, olabilirlik oran uyum iyiliği testi ile test edilmiştir. Her bir model için hesaplanan olabilirlik oran test istatistikleri Tablo 5.5.'te özetlenmiştir.

Tablo 5.5. Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerinin olabilirlik oran test istatistikleri

Model Adı	**lnL(basit model)**	**lnL(model)**	**Serbestlik derecesi**	**LR**	**p**
Poisson	-56535,00	-56358,00	2	354,80	<0,10
NB	-56120,00	-55959,00	2	321,39	<0,10
GP	-56119,16	-55962,80	2	312,72	<0,10
ZIP	-56133,00	-55971,54	2	322,92	<0,10
ZINB	-56133,45	-55950,59	2	365,72	<0,10

Tablo 5.5.'te görüldüğü üzere basit modeller ile uygunluğu araştırılan modellerin log-olabilirlik değerleri arasındaki fark, 2 serbestlik derecesi ve %10 yanılma düzeyi için elde edilen ki-kare değerinden (4,60) oldukça yüksektir. Bu nedenle uygunluğu araştırılan modellerin tümü %10 yanılma düzeyinde anlamlı bulunmuştur.

Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerine ilişkin parametre tahminleri Tablo 5.6.'da yer almaktadır.

Tablo 5.6. Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerinin parametre tahminleri

Açıklayıcı Değişkenler	**μ Regresyonu**				
	Poisson	**NB**	**GP**	**ZIP**	**ZINB**
Sabit terim (intercept)	β_1: -2,3233 (0,0399) p<0,1	β_1: -2,3282 (0,0412) p<0,1	β_1: -2,3233 (0,0412) p<0,1	β_1: -1,2797 (0,0518) p<0,1	β_1: -2,2559 (0,0955) p<0,1
Yaş	β_2: -0,0163 (0,0008) p<0,1	β_2: -0,0162 (0,0009) p<0,1	β_2: -0,0162 (0,0009) p<0,1	β_2: -0,0161 (0,0009) p<0,1	β_2: -0,0161 (0,0009) p<0,1
Cinsiyet	β_3: -0,0838 (0,0224) p<0,1	β_3: -0,0838 (0,0229) p<0,1	β_3: -0,0838 (0,0229) p<0,1	β_3: -0,0849 (0,0233) p<0,1	β_3: -0,1557 (0,0865) p<0,1
Yayılım parametresi (dispersion parameter)	---	α: 0,5125 (0,0263) p<0,1	φ: 1,0477 (0,0024) p<0,1	---	α: 0,5241 (0,0341) p<0,1
Açıklayıcı Değişkenler	**ω Regresyonu**				
Sabit terim (intercept)	---	---	---	γ_1: 0,6174 (0,0485) p<0,1	γ_1: 0,8150 (0,0122) p<0,1

*Parantez içinde yer alan değerler, parametrelerin standart sapmasını göstermektedir.

2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası'nda bir yıllık sürede gerçekleşen hasar sayısı bağımlı değişkeni için açıklayıcı değişken olarak sigortalıların yaşı, cinsiyeti ve aracın kullanım tipi değişkenleri ile kurulan regresyon modellerinin ilk bölümünü oluşturan ortalama regresyonu incelendiğinde; Poisson, negatif binom, genelleştirilmiş Poisson ve sıfır yığılmalı Poisson modellerinde sigortalıların yaş ve cinsiyet bilgilerinin gerçekleşen hasar sayısı üzerinde anlamlı olduğu görülmektedir. Aşırı yayılımın da modele dahil edildiği negatif binom ve genelleştirilmiş Poisson modellerinin yayılım parametreleri sırasıyla sıfır ve birden farklı hesaplanmış, gözlemlenen verinin aşırı yayılım gösterdiği sonucuna ulaşılmıştır. Gözlemlenen veride var olan aşırı yayılımın diğer bir göstergesi olarak tanımlanan T ve T_1 test istatistikleri sırasıyla 665,63 ve 36,22 olarak hesaplanmıştır. T istatistiğinin büyük ve T_1 istatistiğinin %10 yanılma düzeyi için standart normal dağılım tablo değerinden yüksek hesaplanması, veri kümesinin Poisson varsayımına göre aşırı yayılım gösterdiğini belirtmektedir.

Oluşturulan sıfır yığılmalı regresyon modellerinde hiçbir açıklayıcı değişkenin yapısal sıfır oranı üzerinde etkili olmadığı sonucuna ulaşılmış ve bu oran sabit kabul edilmiştir. Poisson regresyon modeli ile sıfır yığılmalı Poisson regresyon modelinin karşılaştırılarak yapısal sıfır oranının sıfıra eşitliğinin test edildiği, Van den Broek tarafından geliştirilen skor test istatistiği 1114,2 olarak hesaplanmıştır. Bu değer, 1 serbestlik dereceli ki-kare değeri ile karşılaştırılarak yokluk hipotezi reddedilmiştir. Gözlenen hasar sayısının sıfır değerindeki yoğunluk, Poisson modeli varsayımı altında beklenen yoğunluktan fazladır ve sıfır yığılmalı regresyon modellerinin gözlemlenen hasar sayısını modellemek amacıyla kullanılması uygun bulunmuştur. Ek olarak, oluşturulan sıfır yığılmalı regresyon modellerinde de görülebileceği üzere; yapısal sıfır oranı için elde edilen sabit terim değerleri, sıfır yığılmalı Poisson ve sıfır yığılmalı negatif binom modelleri için anlamlı bulunmuştur.

Gözlemlenen veri kümesine en uygun modelin seçilmesi amacıyla Akaike, Bayes bilgi ölçütleri, genelleştirilmiş Pearson ki-kare istatistiği ve log-olabilirlik değerlerinin de hesaplandığı sonuç tablosu Tablo 5.7.'de verilmiştir.

Tablo 5.7. Akaike, Bayes, genelleştirilmiş Pearson ki-kare ve log-olabilirlik değerleri

Model Adı	AIC	BIC	Pearson ki-kare	LnL
Poisson	112.721,0	112.752,9	317.453,8	-56.357,6
NB	111.926,0	111.968,2	290.442,7	-55.958,9
GP	111.934,0	111.975,9	316.928,0	-55.962,8
ZIP	111.951,1	111.993,4	291.675,3	-55.971,5
ZINB	111.911,2	111.964,1	290.402,2	-55.950,6

Akaike, Bayes bilgi ölçütleri ve genelleştirilmiş Pearson ki-kare değerinin en küçük, log-olabilirlik değerinin en büyük olduğu model, hasar sayısı bağımlı değişkeni için en uygun model olarak belirlenmektedir. Tablo 5.7.'den de görüldüğü üzere 2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptıran sigortalıların yaş ve cinsiyet bilgileri göz önünde bulundurulduğunda bir yıl içinde gerçekleşen hasar sayısının, sıfır yığılmalı negatif binom regresyon modeli kullanılarak modellenmesi uygun bulunmuştur. Sıfır yığılmalı negatif binom modeli için bulunan katsayılar incelendiğinde, gerçekleşen hasar sayısı ile sigortalı bireylerin yaşı arasında ters ilişki olduğu görülmektedir. Bunun nedeni, genç sürücülerin kurallara uyumunun az, hız yapma eğiliminin fazla olması ve sürücü tecrübesinin ileri yaştaki sigortalılara göre az olması olarak değerlendirilebilir. Sigortalı bireylerin cinsiyet bilgisinin ortalama hasar sayısı üzerindeki etkisi incelendiğinde, erkek sürücülerin yıl içinde gerçekleştirmesi beklenen hasar sayısının kadın sürücülere göre daha düşük olduğu görülmektedir. Gerçekleşen hasar sayısının sıfır yığılmalı negatif binom regresyon modeli ile modellenmesi sonucu yapısal sıfır oranı %69,3 olarak belirlenmiştir. Dolayısıyla bir yıllık sürede hasar gerçekleştiren fakat hasar tutarı muafiyet sınırının altında kalan veya hasarsızlık indirimi nedeniyle düşük miktarda prim ödemeyi tercih ettiği için sigorta şirketini bilgilendirmeyen sigortalıların tüm sigortalılar içindeki oranı %69,3' tür.

5.3. İkinci Model Sonuçları

Sigortalıların yaş ve cinsiyet bilgileri ile araçların bağlı bulundukları il bilgisinin göz önünde bulundurulması amacıyla kullanılan sektör payı değişkenleri ile oluşturulan regresyon modellerine ilişkin özet tablo Tablo 5.8.' de gösterilmiştir.

Tablo 5.8. Yaş, cinsiyet ve sektör payı değişkenleri ile kurulan regresyon modelleri

Açıklayıcı Değişkenler	μ Regresyonu					
	Poisson	**NB**	**GP**	**ZIP**	**ZINB**	**ZIGP**
Sabit terim (intercept)	β_1: -2,6439 (0,0415) p<0,1	β_1: -2,6506 (0,0434) p<0,1	β_1: -2,6439 (0,0428) p<0,1	β_1: -1,4746 (0,1332) p<0,1	β_1: -2,2143 (0,0742) p<0,1	β_1: -2,7598 (0,0511) p<0,1
Yaş	β_2: -0,0161 (0,0008) p<0,1	β_2: -0,0159 (0,0009) p<0,1	β_2: -0,0161 (0,0009) p<0,1	β_2: -0,0163 (0,0026) p<0,1	β_2: -0,0173 (0,0012) p<0,1	β_2: -0,0142 (0,0010) p<0,1
Cinsiyet	β_3: -0,0104 (0,0224) p>0,1	β_3: -0,0134 (0,0234) p>0,1	β_3: -0,0104 (0,0230) p>0,1	β_3: 0,0158 (0,1035) p>0,1	β_3: 0,0630 (0,0312) p<0,1	β_3: 0,0686 (0,0285) p<0,1
Sektör Payı	β_4: 2,6365 (0,0803) p<0,1	β_4: 2,6459 (0,0844) p<0,1	β_4: 2,6365 (0,0828) p<0,1	β_4: 0,8946 (0,6429) p>0,1	β_4: 0,7266 (0,2098) p<0,1	β_4: 5,19851 (0,3272) p<0,1
Yayılım parametresi (dispersion parameter)	---	α: 0,5718 (0,0305) p<0,1	φ: 1,0477 (0,0024) p<0,1	---	α: 0,8101 (0,0612) p<0,1	φ: 1,0368 (0,0026) p<0,1
Açıklayıcı Değişkenler	**ω Regresyonu**					
Sabit terim (intercept)	---	---	---	γ_1: 0,8111 (0,2035) p<0,1	γ_1: -0,2796 (0,2631) p>0,1	γ_1: -10,1267 (13,6431) p>0,1
Yaş	---	---	---	γ_2: -0,0004 (0,0039) p>0,1	γ_2: -0,0073 (0,0042) p<0,1	γ_2: 0,0113 (0,0042) p<0,1
Cinsiyet	---	---	---	γ_3: 0,0451 (0,1675) p>0,1	γ_3: 0,4723 (0,1674) p<0,1	γ_3: -0,3604 (0,0935) p<0,1
Sektör Payı	---	---	---	γ_4: -2,8662 (1,1156) p<0,1	γ_4: -25,0819 (2,7884) p<0,1	γ_4: 38,3999 (52,9273) p>0,1

*Parantez içinde yer alan değerler, parametrelerin standart sapmasını göstermektedir.

Hasar sayısının bağımlı değişken, sigortalıların yaş ve cinsiyet bilgileri ile sektör payı değişkenlerinin açıklayıcı değişkenler olarak belirlendiği regresyon modellerinde her bir model için açıklayıcı değişkenlerin katsayılarının anlamlılığı, Wald tipi test istatistiği kullanılarak test edilmiş ve %10 yanılma düzeyine göre anlamlılıkları belirtilmiştir. Buna göre cinsiyet değişkeni, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modelleri dışındaki tüm modeller için anlamsız bulunmuştur. Poisson, negatif binom ve genelleştirilmiş Poisson modellerinde anlamlı bulunan açıklayıcı değişkenler açısından fark olmamasına rağmen, sıfır yığılmalı modellerde her bir model için anlamlı bulunan açıklayıcı değişkenlerin farklı olduğu görülmektedir.

Sıfır yığılmalı modellerde temel sayı dağılımının regresyonu için kullanılan açıklayıcı değişkenler ile yapısal sıfır oranı regresyonu için kullanılan açıklayıcı değişkenler birbirini etkilemekte ve katsayıların anlamlılığı da buna bağlı olarak değişkenlik göstermektedir. Herhangi bir açıklayıcı değişkenin yapısal sıfır oranı üzerindeki etkisinin incelenmesi amacıyla geriye doğru seçim yöntemi uygulanmalıdır.

Tablo 5.8.'de görüldüğü gibi, sigortalıların yaş ve cinsiyet bilgileri ile araçların kullanım tipi değişkeni açıklayıcı değişkenler olarak belirlendiğinde hasar verisine uyum sağlamadığı belirlenen sıfır yığılmalı genelleştirilmiş Poisson modeli, kullanım tipi değişkeninin yerine sektör payı değişkeni kullanıldığında yapısal sıfır oranı regresyonunda bazı değişkenlerin anlamlı bulunması nedeniyle, anlamlı bir regresyon modeli olarak elde edilmiştir.

Her bir model için hesaplanan olabilirlik oran test istatistikleri Tablo 5.9.'da özetlenmiştir.

Tablo 5.9. Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerinin olabilirlik oran test istatistikleri

Model Adı	**lnL(basit model)**	**lnL(model)**	**Serbestlik derecesi**	**LR**	**p**
Poisson	-56535,00	-55846,80	2	1376,40	<0,10
NB	-56119,63	-55492,32	2	1254,62	<0,10
GP	-56119,20	-55496,60	2	1245,20	<0,10
ZIP	-56132,89	-55497,98	3	1269,82	<0,10
ZINB	-56133,10	-55398,82	5	1468,56	<0,10
ZIGP	-56119,29	-55395,50	4	1447,58	<0,10

Tablo 5.9.'da görüldüğü üzere basit modeller ile uygunluğu araştırılan modellerin log-olabilirlik değerleri arasındaki fark, çeşitli serbestlik dereceleri ve %10 yanılma düzeyi için elde edilen ki-kare değerlerinden yüksektir. Bu nedenle uygunluğu araştırılan modellerin tümü %10 yanılma düzeyinde anlamlı bulunmuştur.

Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerine ilişkin parametre tahminleri Tablo 5.10.'da yer almaktadır.

Tablo 5.10. Anlamlı bulunan açıklayıcı değişkenlerle kurulan regresyon modellerinin parametre tahminleri

Açıklayıcı Değişkenler	**μ Regresyonu**					
	Poisson	**NB**	**GP**	**ZIP**	**ZINB**	**ZIGP**
Sabit terim (intercept)	β_1: -2,6527 (0,0370) p<0,1	β_1: -2,6619 (0,0387) p<0,1	β_1: -2,6528 (0,0381) p<0,1	β_1: -1,4816 (0,0711) p<0,1	β_1: -2,2143 (0,0742) p<0,1	β_1: -2,7498 (0,05015) p<0,1
Yaş	β_2: -0,0161 (0,0008) p<0,1	β_2: -0,0159 (0,0009) p<0,1	β_2: -0,0161 (0,0009) p<0,1	β_2: -0,0160 (0,0009) p<0,1	β_2: -0,0173 (0,0012) p<0,1	β_2: -0,0143 (0,0011) p<0,1
Cinsiyet	---	---	---	---	β_3: 0,0630 (0,0312) p<0,1	β_3: 0,0674 (0,0294) p<0,1
Sektör Payı	β_3: 2,6404 (0,0799) p<0,1	β_3: 2,6508 (0,0839) p<0,1	β_3: 2,6405 (0,0823) p<0,1	β_3: 1,0004 (0,4831) p<0,1	β_4: 0,7266 (0,2098) p<0,1	β_4: 0,8431 (0,3263) p<0,1
Yayılım parametresi (dispersion parameter)	---	α: 0,5718 (0,0305) p<0,1	φ: 1,0477 (0,0024) p<0,1	---	α: 0,8101 (0,0612) p<0,1	φ: 1,0368 (0,0026) p<0,1
Açıklayıcı Değişkenler	**ω Regresyonu**					
Sabit terim (intercept)	---	---	---	γ_1: 0,8171 (0,0894) p<0,1	---	---
Yaş	---	---	---	---	γ_1: -0,0073 (0,0042) p<0,1	γ_1: 0,0103 (0,0042) p<0,1
Cinsiyet	---	---	---	---	γ_2: 0,4723 (0,1674) p<0,1	γ_2: -0,3706 (0,1047) p<0,1
Sektör Payı	---	---	---	γ_2: -2,6883 (0,8160) p<0,1	γ_3: -25,0819 (2,7884) p<0,1	---

*Parantez içinde yer alan değerler, parametrelerin standart sapmasını göstermektedir.

2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası'nda bir yıllık sürede gerçekleşen hasar sayısı bağımlı değişkeni için açıklayıcı değişken olarak sigortalıların yaşı, cinsiyeti ve kullanılan aracın bağlı bulunduğu ilin sektör payı değişkenlerinin belirlendiği regresyon modellerinin ilk bölümü olan temel sayı dağılımının regresyonu incelendiğinde; Poisson, negatif binom, genelleştirilmiş Poisson ve sıfır yığılmalı Poisson modellerinde cinsiyet değişkeninin gerçekleşen hasar sayısı üzerinde etkisinin bulunmadığı, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson modellerinde ise bu değişkenin anlamlı bulunduğu görülmektedir. Yaş değişkeni tüm modeller için anlamlı iken, sektör payı değişkeni ise sıfır yığılmalı Poisson ve sıfır yığılmalı genelleştirilmiş Poisson modelleri dışındaki tüm modellerde anlamlı bulunmuştur. Temel sayı modelleri için belirlenen açıklayıcı değişkenler kümesi aynı iken sıfır yığılmalı regresyon modellerinde her bir model için anlamlı bulunan açıklayıcı değişkenler birbirinden farklıdır.

Aşırı yayılımın da model içinde değerlendirildiği negatif binom ve genelleştirilmiş Poisson modellerinin yayılım parametrelerinin sırasıyla sıfır ve birden farklı hesaplanması nedeniyle, gözlemlenen verinin aşırı yayılım gösterdiği sonucuna ulaşılmıştır. Gözlemlenen verideki aşırı yayılımın diğer bir göstergesi olarak tanımlanan T ve T_1 test istatistikleri sırasıyla 639,33 ve 33,47 olarak hesaplanmıştır. T istatistiğinin büyük ve T_1 istatistiğinin %10 yanılma düzeyi için standart normal dağılım tablo değerinden yüksek hesaplanması, veri kümesinin Poisson varsayımına göre aşırı yayılım gösterdiğini belirtmektedir.

Sıfır yığılmalı Poisson regresyon modelinin kullanımının test edildiği, Van den Broek tarafından geliştirilen skor test istatistiği 954,5 olarak hesaplanmıştır. Bu değer, 1 serbestlik dereceli ki-kare değeri ile karşılaştırılarak yokluk hipotezi reddedilmiştir. Gözlenen hasar sayısının sıfır değerindeki yoğunluk, Poisson modeli varsayımı altında beklenen yoğunluktan fazladır ve sıfır yığılmalı regresyon modellerinin gözlemlenen hasar sayısını modellemek amacıyla kullanılması uygun bulunmuştur.

Gözlemlenen veri kümesine en uygun modelin seçilmesi amacıyla Akaike, Bayes bilgi ölçütleri, genelleştirilmiş Pearson ki-kare istatistiği ve log-olabilirlik değerlerinin de hesaplandığı sonuç tablosu Tablo 5.11.’de verilmiştir.

Tablo 5.11. Akaike, Bayes, genelleştirilmiş Pearson ki-kare ve log-olabilirlik değerleri

Model Adı	**AIC**	**BIC**	**Pearson ki-kare**	**LnL**
Poisson	111.700,0	111.731,3	315.140,0	-55.846,8
NB	110.993,0	111.044,2	290.948,2	-55.492,3
GP	111.001,0	111.043,5	314.089,0	-55.496,6
ZIP	111.006,0	111.058,8	290.760,9	-55.498,0
ZINB	110.813,6	110.898,3	288.543,8	-55.398,8
ZIGP	110.805,0	110.879,0	314.285,0	-55.395,5

Akaike, Bayes bilgi ölçütleri ve genelleştirilmiş Pearson ki-kare değerinin en küçük, log-olabilirlik değerinin en büyük olduğu model, hasar sayısı bağımlı değişkeni için en uygun model olarak belirlenmektedir. Tablo 5.11.'den de görüldüğü üzere genelleştirilmiş Pearson ki-kare değeri yüksek olmasına rağmen diğer ölçütler için gerekli koşulun sağlandığı sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli; 2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası'nda bir yıl içinde gerçekleşen hasar sayısı için en uygun model olarak belirlenmiştir. Sıfır yığılmalı negatif binom regresyon modeli ile sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli için model seçim ölçütleri bakımından yakın sonuçlar elde edilmesi nedeniyle sıfır yığılmalı negatif binom regresyon modeli, sıfır yığılmalı genelleştirilmiş Poisson regresyon modeline alternatif olarak belirlenebilmektedir.

Sıfır yığılmalı genelleştirilmiş Poisson modeli için bulunan katsayılar incelendiğinde, gerçekleşen hasar sayısı ile sigortalı bireylerin yaşı arasında ters ilişki olduğu görülmektedir. Bunun nedeni, genç sürücülerin kurallara uyumunun az, hız yapma eğiliminin fazla olması ve sürücü tecrübesinin ileri yaştaki sigortalılara göre az olması olarak değerlendirilebilir. Sigortalı bireylerin cinsiyet bilgisinin ortalama hasar sayısı üzerindeki etkisi incelendiğinde, kadın sürücülerin yıl içinde gerçekleştirmesi beklenen hasar sayısının erkek sürücülere göre düşük olduğu görülmektedir. Ayrıca 2008 yılının ocak ayında poliçe üretiminin fazla olduğu illerde gerçekleşmesi beklenen hasar sayısı, üretimin az olduğu illerde gerçekleşmesi beklenen hasar sayısından daha yüksektir. Gerçekleşen hasar sayısı sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli ile modellendiğinde yapısal sıfır oranı için de regresyon modeli kurulduğundan, bu oran sabit bir değere eşitlenememekte ve her sigortalı için farklı bir değer hesaplanmaktadır. Bu durumda, hasar ihbarı için bir oran belirlendiğinde sigortalı bireylerin kişisel özellikleri önemli bir rol oynamaktadır.

6. SONUÇ VE ÖNERİLER

Türkiye'de sigorta şirketlerine beklenenden daha az sayıda hasar bildiriminde bulunulduğu görülmektedir. Hasar sayılarının sıfır yığılmalı regresyon modelleri ile modellenmesi; sıfır değerindeki yığılmanın değerlendirilmesi ve kaza gerçekleşme riski seviyesinin sigortalıların bireysel özellikleri ve çevresel unsurlar dikkate alınarak belirlenmesine imkan sağlamaktadır.

Türkiye'de 2008 yılının ocak ayında yürürlüğe giren poliçelerin bir yıllık bilgisi kullanılarak elde edilen sonuçlar; sıfır değerindeki yığılmanın temel sayı modelleri ile yeterli ölçüde açıklanamadığını, bu modeller yerine sıfır yığılmalı regresyon modellerinin kullanımının uygun olacağını göstermiştir. Ek olarak, hasar sayısı bağımlı değişkeni için farklı açıklayıcı değişkenler kullanıldığında, seçilen en uygun model ve parametre tahminleri de farklı elde edilmiştir. Regresyon çözümlemesi sonucunda; yaş, cinsiyet ve sektör payı değişkenlerinin, hasar sayısı bağımlı değişkeninin açıklanmasında önemli risk faktörleri olduğu sonucuna ulaşılmıştır. Yaş, cinsiyet ve kullanım tipi değişkenleri kullanıldığında sıfır yığılmalı negatif binom modeli; yaş, cinsiyet ve sektör payı değişkenleri kullanıldığında ise sıfır yığılmalı genelleştirilmiş Poisson modeli hasar sayıları için en uygun model olarak belirlenmiştir. Sıfır yığılmalı negatif binom regresyon modelinde yapısal sıfır oranı sabit kabul edilmiş ve bu oran, sigortalıların bireysel özelliklerine göre farklı hesaplanmamıştır. Ancak yaş ve cinsiyet değişkenlerinin ortalama hasar sayısı üzerindeki etkisi incelendiğinde, sigortalıların yaşı arttıkça gerçekleşmesi beklenen hasar sayısının azaldığı sonucuna ulaşılmıştır. Ayrıca kadın sigortalılar tarafından gerçekleştirilmesi beklenen hasar sayısının, erkek sigortalılar tarafından gerçekleştirilmesi beklenen hasar sayısından daha fazla olduğu görülmüştür. Sıfır yığılmalı genelleştirilmiş Poisson regresyon modelinde ise yapısal sıfır oranının değeri, sigortalıların yaş ve cinsiyet bilgilerine göre değişmektedir. Sigortalıların yaşı ile yapısal sıfır oranı doğru orantılıdır. Dolayısıyla cinsiyet değişkeni sabit kabul edildiğinde sigortalıların yaşı arttıkça gerçekleşen hasarın sigorta şirketine bildirilmeme olasılığı da artmaktadır. Yaş değişkeni sabit kabul edildiğinde ise kadın sigortalılar tarafından gerçekleştirilen hasarın sigorta şirketine bildirilmeme olasılığı, erkek sigortalılara göre yüksek hesaplanmıştır. Yaş, cinsiyet ve sektör payı değişkenlerinin ortalama hasar sayısı üzerindeki etkisi incelendiğinde, sigortalıların yaşı ile gerçekleşmesi beklenen hasar sayısının ters ilişkili olduğu görülmüştür. Sektör payı ile gerçekleşmesi beklenen hasar sayısı ise doğru orantılıdır. Poliçe üretiminin çok olduğu illerde gerçekleşmesi beklenen hasar sayısı, poliçe üretiminin az olduğu illere göre daha yüksek hesaplanmıştır. Erkek sigortalılar tarafından gerçekleştirilmesi beklenen hasar sayısı ise kadın sigortalılara göre fazladır.

Risk kabul sürecinde poliçelere daha ayrıntılı sigortalı ve araç bilgisinin kaydedilmesi, hasar sıklığının modellenmesinde kullanılmak amacıyla belirlenen açıklayıcı değişken sayısının artmasına neden olacaktır. Açıklayıcı değişken sayısının fazla olduğu ve özellikle sigortalılar tarafından kullanılan araçlara ilişkin değişkenlerin de yer aldığı regresyon modelleri ile, hem sigortalıların hem de kullanılan araçların özelliklerine dayanan sonuçlar elde edilmesi mümkündür.

Hasar sayısının modellenmesi, sigorta alanında poliçe priminin hesaplanması için en temel unsurlardan biridir. Regresyon analizi ile; risk faktörleri belirlenerek sigortalıların bireysel özelliklerinin bilindiği durumda hasar sayılarının beklenen sıklığının tahmini elde edilmektedir. Poliçe risk primi, gözlemlenen ve değişkenlik gösterebilen risk karakteristikleri (yaş, cinsiyet, araç kullanım tipi, araç tipi, hasar kayıtları, meslek vb.) yardımıyla hasar sayılarının koşullu beklenen değerinin elde edilmesi ve bu değerin ekonomik kayıp olarak da tanımlanan beklenen hasar tutarı ile aktüeryal denge gözönünde bulundurularak birleştirilmesi biçiminde hesaplanmaktadır. Risk primine idari masraflar ile satın alma, reasürans ve sermaye maliyetleri gibi şirketlerce belirlenen maliyetler eklenerek brüt prim elde edilmektedir. Prim hesaplanması aşamasında hasar sayıları için kullanılan modelin doğru ve değişen risk faktörlerine duyarlı olması, fiyatlandırmanın iyi yapılabilmesinde önemli rol oynamaktadır.

Çalışmada bir yıllık sürede gerçekleşen hasar sayıları modellendiğinden sigortalıların geçmiş hasar bilgisi dikkate alınmamıştır. Çözümlemede geçmiş hasar bilgisinin kullanılmaması, bağımsız hasar sayılarının modellendiğini ifade etmektedir. Ancak aynı şirkette bir yıldan uzun süre kalan sigortalılar için genellikle geçmiş hasar bilgisine dayanan incelemeler yapılabilmektedir. Bu durumda, her bir sigortalı tarafından gerçekleşen hasar sayıları bağımlı olmaktadır. Hasar sayılarının ilişkili olduğu veri kümelerinin analizi için bağımlılığı inceleyen dağılımların kullanılması gerekmektedir. Çalışmanın sonraki aşamasında sıfır değerinde yığılma gösteren bağımlı hasar sayılarının modellenmesi konusu araştırılabilir. Ayrıca oluşturulan regresyon modellerinde yapısal sıfır oranı regresyonu için lojit bağ fonksiyonu yerine, diğer bağ fonksiyonları da kullanılarak farklı modeller oluşturulabilir.

KAYNAKLAR

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E., Thandi, N., 2007, A Practitioner's Guide to Generalized Linear Models, Casualty Actuarial Society Discussion Paper Program, 1-116.

Boucher, J.-P., Denuit, M., Guillen, M., 2007, Risk Classification for Claim Counts: Mixed Poisson, Zero-Inflated Mixed Poisson and Hurdle Models, North American Actuarial Journal, 11, 4, 110-131.

Carrivick, P.J.W., Lee, A.H., Yau, K.K.W., 2003, Zero-inflated Poisson Modeling to Evaluate Occupational Safety Interventions, Safety Science, 41, 53-63.

Cavanaugh, J.E., Shumway, R.H., 1996, On Computing the Expected Fisher Information Matrix for State-Space Model Parameters, Statistics and Probability Letters, 26, 4, 347-355.

Czado, C., Erhardt, V., Min, A., Wagner, S., 2007, Zero-inflated Generalized Poisson Models with Regression Effects on the Mean, Dispersion and Zero-inflation Level Applied to Patent Outsourcing Rates, Statistical Modelling, 7, 2, 125-153.

Consul, P.C., 1979, Generalized Poisson Distributions: Properties and Applications. Marcel Dekker, New york.

Consul, P.C., Famoye, F., 1992, Generalized Poisson Regression Model, Communications in Statistics-Theory and Method, 21, 1, 89-109.

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

Dean, C., Lawless, J.F., 1989, Test for Defecting Overdispersion in Poisson Regression Models, Journal of American Statistical Association, 84, 467-472.

Deng, D., Paul, S. R., 2000, Score Tests for Zero Inflation In Generalized Linear Models, Canadian Journal of Statistics, 28, 563-570.

Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.F., 2007, Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems, 356p.

Faddy, M.J., 1998, Stochastic Models for Analysis of Species Abundance Data, Statistics in Ecology and Environmental Monitoring, 2, eds. D.J. Fletcher, L. Kavalieris, B.F.J. Manly, Dunedin: University of Otago Press, 33-40.

Famoye, F., Özmen, İ., 2007, Count Regression Models with an Application to Zoological Data Containing Structural Zeros, Journal of Data Science, 5, 491-502.

Famoye, F., Singh, K.P., 2006, Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data, Journal of Data Science, 4, 117-130.

Famoye, F., Wulu, J.T., Singh, K.P., 2004, On The Generalized Poisson Regression Model with an Application to Accident Data, Journal of Data Science, 2, 287-295.

Faraway, J.J., 2006, Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Chapman and Hall, London, 301p.

Flynn, M., Francis, A.,L., 2009, More Flexible GLMs: Zero Inflated Models and Hybrid Models, CAS Ratemaking and Product Management Seminar.

Giles, D., 2010, Notes on the Zero Inflated Poisson Model, Department of Economics, University of Victoria.

Greene, W. H., 1994, Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Technical report.

Haberman, S., Renshaw, A.E., 1996, Generalized Linear Models and Actuarial Science, The Statistician, 45, 407-463.

Harrell, F.E., 2001, Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis, Springer-Verlag New York, Inc. 557p.

Janardan, K.G., Kerster, H.W., Schaeffer, D.J., 1979, Biological Applications of the Lagrangian Poisson Distributions, Bioscience, 29, 599-602.

Jansakul, N., Hinde, J.P., 2002, Score Tests for Zero-Inflated Poisson Models, Computational Statistics and Data Analysis, 40, 75 – 96.

Lambert, D., 1992, Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing, Technometrics, 34, No.1.

Land, K.C., McCall, P.L., Nagin, D.S., 1996, A Comparison of Poisson, Negative Binomial and Semiparametric Mixed Poisson Regressive Models with Empirical Applications to Criminal Careers Data, Sociological Methods and Research, 24, 387-442.

Lee, A.H., Stevenson, M.R., Wang, K., Yau, K.K.W., 2002, Modeling Young Driver Motor Vehicle Crashes: Data with Extra Zeros, Accident Analysis and Prevention, 34, 515-521.

Lee, A.H., Wang, K., Yau, K.K.W., 2001, Analysis of Zero-Inflated Poisson Data Incorporating Extent of Exposure, Biometrical Journal, 43, 8, 963–975.

Lord, D., Park, B.-Y., 2010, CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations, Appendix D: Negative Binomial Regression Models and Estimation Methods.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P., 2005, Zero Tolerance Ecology: Improving Ecological Inference by Modelling the Source of Zero Observations, Ecology Letters, 8, 1235-1246.

McCullagh, P., Nelder, J.A., 1989, Generalized Linear Models, 2nd Ed., Chapman and Hall, London, 511p.

Miaou, S., 1994, The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions, Accident Analysis and Prevention, 26, 4, 471-482.

Mullahy, J., 1986, Specification and Testing of Some Modified Count Data Models, Journal of Econometrics, 33, 341-365.

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

Özmen, İ., Demirhan, H., 2010, A Bayesian Approach for Zero-Inflated Count Regression Models by Using the Reversible Jump Markov Chain Monte Carlo Method and an Application, Communications in Statistics-Theory and Methods, 39, 2109-2127.

Rashwan, N.A., Kamel, M.M., 2011, Using Generalized Poisson Log Linear Regression Models in Analyzing Two-Way Contingency Tables, Applied Mathematical Sciences, 5, 5, 213 – 222.

Renshaw, A.E., 1994, Modelling the Claims Process in The Presence of Covariates, Astin Bulletin, 24, 265-285.

Ridout, M., Demetrio, C.G.B., Hinde, J., 1998, Models for Count Data with Many Zeros, International Biometric Conference, Capetown December.

Shankar, V., Milton, J., Mannering, F., 1997, Modeling Accident Frequencies As Zero-Altered Probability Processes: An Emprical Inquiry, Accident Analysis and Prevention, 29, 6, 829-837.

Simons, J.S., Neal, D.J., Gaher, R.M., 2006, Risk for Marijuana-related Problems Among College Students: An Application of Zero-inflated Negative Binomial Regression, The American Journal of Drug and Alcohol Abuse, 32, 41–53.

Thomas, H., Samson, D., 1987, Linear Models as Aids In Insurance Data Making: The Estimation of Automobile Insurance Claims, Journal of Business Research, 15, 3, 247-256.

Ugarte, M.D., Ibáńez, B., Militino, A.F., 2004, Testng for Poisson Zero Inflation in Disease Mapping, Biometrical Journal, 46, 5, 526-539.

Van den Broek, J., 1995, A Score Test for Zero-Inflation In a Poisson Distribution, Biometrcis, 51, 738-743.

Wang, W., Famoye, F., 1997, Modeling Household Fertility Decisions With Generalized Poisson Regression, Journal of Population Economics, 10, 273-283.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996, Modelling the Abundance of Rare Species: Statistical Models for Counts with Extra Zeros, Ecological Modeling, 88, 297-308.

Yang, Z., 2006, Score Tests for Generalization and Zero-Inflation In Count Data Modeling, Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 131p.

Yip, K.C.H., Yau, K.K.W., 2005, On Modeling Claim Frequency Data In General Insurance With Extra Zeros, Insurance: Mathematics and Economics, 36, 153-163.

EKLER DİZİNİ

Sayfa

EK-1. 2008 yılı ocak ayı otomobil araç türü için sektör payları

Araç Grup Kodu ve İl Bazında Poliçe Üretimi Dağılımı Tarih: 1/2008 - 1/2009 (Cari) Araç Grup Kodu : Otomobil			
İL	**Sektör Payı (%)**	**İL**	**Sektör Payı (%)**
ADANA	2,57	KONYA	2,71
ADIYAMAN	0,34	KÜTAHYA	0,87
AFYON	0,66	MALATYA	0,64
AGRI	0,10	MANISA	1,57
AMASYA	0,40	K.MARAS	0,86
ANKARA	12,63	MARDIN	0,18
ANTALYA	3,80	MUGLA	1,60
ARTVIN	0,11	MUS	0,07
AYDIN	1,39	NEVSEHIR	0,34
BALIKESIR	1,55	NIGDE	0,29
BILECIK	0,23	ORDU	0,54
BINGÖL	0,06	RIZE	0,22
BITLIS	0,08	SAKARYA	0,98
BOLU	0,39	SAMSUN	1,28
BURDUR	0,46	SIIRT	0,06
BURSA	3,75	SINOP	0,22
ÇANAKKALE	0,62	SIVAS	0,60
ÇANKIRI	0,10	TEKIRDAG	0,81
ÇORUM	0,63	TOKAT	0,57
DENIZLI	1,56	TRABZON	0,70
DIYARBAKIR	0,45	TUNCELI	0,02
EDIRNE	0,51	SANLIURFA	0,87
ELAZIG	0,49	USAK	0,50
ERZINCAN	0,21	VAN	0,31
ERZURUM	0,49	YOZGAT	0,35
ESKISEHIR	1,34	ZONGULDAK	0,81
GAZIANTEP	1,55	AKSARAY	0,42
GIRESUN	0,27	BAYBURT	0,05
GÜMÜSHANE	0,07	KARAMAN	0,29
HAKKARI	0,03	KIRIKKALE	0,22
HATAY	1,41	BATMAN	0,15
ISPARTA	0,69	SIRNAK	0,04
IÇEL	1,96	BARTIN	0,22
ISTANBUL	25,52	ARDAHAN	0,02
IZMIR	6,83	IGDIR	0,05
KARS	0,09	YALOVA	0,20
KASTAMONU	0,46	KARABÜK	0,35
KAYSERI	1,90	KILIS	0,06
KIRKLARELI	0,43	OSMANIYE	0,52
KIRSEHIR	0,27	DÜZCE	0,35
KOCAELI	1,60		

Kaynak:
http://www.tramer.org.tr/raporlar/kumul2/istarac.php?ReportType=0&AY0=1&YIL0=2008&AY=1&YIL=2009&ARAC_GRUP_KODU[]=01&PLAKA_IL_KODU[]=-1&x=41&y=15

ÖZGEÇMİŞ

**Adı Soyadı : ** Sema TÜZEL

**Doğum Yeri : ** Konya

**Doğum Yılı : ** 1986

**Medeni Hali : ** Bekar

Eğitim ve Akademik Durumu:

**Lise : ** 2000-2004 Dr. Binnaz Ege - Dr. Rıdvan Ege Anadolu Lisesi

**Lisans : ** 2004-2008 Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü

**Yabancı Dil : ** İngilizce

İş Tecrübesi:

Aralık, 2008 - ... Hacettepe Üniversitesi Aktüerya Bilimleri Bölümü, Araştırma Görevlisi

.