



## Hasar sıklıkları için sıfır yığılmalı kesikli modeller

Sema Tüzel

Hacettepe Üniversitesi  
Aktüerya Bilimleri Bölümü  
06800-Beytepe, Ankara, Türkiye  
[sematurzel@hacettepe.edu.tr](mailto:sematurzel@hacettepe.edu.tr)

Meral Sucu

Hacettepe Üniversitesi  
Aktüerya Bilimleri Bölümü  
06800-Beytepe, Ankara, Türkiye  
[msucu@hacettepe.edu.tr](mailto:msucu@hacettepe.edu.tr)

### Özet

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. Bu çalışmada, literatürde yaygın olarak kullanılan sıfır yığılmalı regresyon modelleri ele alınmış ve Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası'nda belirli bir sürede meydana gelen hasar sayıları sıfır yığılmalı regresyon ile modellenerek uygun dağılım belirlenmiştir. Çalışmanın uygulama bölümünde Trafik Sigortaları Bilgi Merkezi'nden (TRAMER) alınan veri kümesi kullanılmıştır. 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 yığılmanın açıklanmasında yetersiz olduğunu, hasar sayılarının modellenmesinde sıfır yığılmalı regresyon modellerinin klasik sayı modellerine tercih edilmesi gerektiğini göstermiştir.

**Anahtar sözcükler:** Sıfır yığılmalı regresyon modelleri; Aşırı yayılım; Hasar sayısı dağılımları; Zorunlu trafik sigortası; Genelleştirilmiş doğrusal modeller.

### Abstract

#### Zero-inflated discrete models for claim frequencies

Zero-inflated regression models are a special type of generalized linear models and obtained by reweighting for zero when data has excess zeros. In this study, the use of zero-inflated regression models to model claim frequencies for motor third party liability insurance is investigated. In application, data taken from Motor Third Party Liability Insurance Center (TRAMER) is used. Two regression models are constructed for two different variable sets. The best (or the most suitable) model is chosen 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.

### 1. Giriş

Hasar sayısının modellenmesi, sigorta alanında poliçe priminin hesaplanması için en temel unsurlardan biridir. Poliçe risk primi, gözlemlenen ve değişkenlik gösterebilen risk karakteristikleri yardımıyla hasar sayılarının koşullu beklenen değerinin elde edilmesi ve bu değer beklenen hasar tutarı ile aktüeryal denge gözönünde bulundurularak birleştirilmesi ile hesaplanmaktadır. Risk primine 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.

Hayat dışı sigortalar içerisinde en yaygın olanları, Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası (Zorunlu Trafik Sigortası) ve Kara Taşıtları Kasko Sigortasıdır. Prim ile bireysel hasar bilgisi arasında ilişki kurulan sistemler, her iki sigorta türü için birçok ülkede 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 değerlendirilmesi (experience rating), liyakat derecelendirmesi (merit rating) ve Bonus-Malus sistemidir [3]. 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. Hasarsızlık indirimi, Bonus-Malus ve muafiyet gibi uygulamaların bir sonucu olarak, hasar kaydedilmeyen poliçe sayısının tüm portföy içindeki oranı oldukça yüksektir.

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 [9], sosyoloji [10], ekonometri [12], ekoloji [4,17], tarım ve tıp [14] gibi birçok alanda kullanılmıştır.

## 2. Hasar sıklığı için kullanılan regresyon modelleri

Hasar sıklığının modellenmesi amacıyla genellikle Poisson, negatif binom ve genelleştirilmiş Poisson dağılımları kullanılmaktadır. Sigortalı bireylerin belirli özellikleri ve çevre koşullarının da hasar sıklığını etkileyebileceği düşünülerek, bireysel ve çevresel faktörlerin (açıklayıcı değişkenler) de dikkate alındığı genelleştirilmiş doğrusal regresyon modelleri kullanılmaktadır. Genelleştirilmiş doğrusal regresyon modelleri; rastgele, sistematik ve bağ fonksiyonu olmak üzere üç temel bileşenden oluşmaktadır. Rastgele bileşen, veri kümesindeki tüm bağımsız değişkenlerin birbirinden bağımsız olduğu ve üstel dağılım ailesindeki dağılımların herhangi birinden geldiği varsayımdır. Sistematik bileşende, modelin ortalama parametresinin hesaplanmasında kullanılan doğrusal önkestim, açıklayıcı değişkenlerin doğrusal kombinasyonu biçiminde ifade edilmektedir.  $n$  sayıda gözlemden ve  $k$  sayıda açıklayıcı değişkenden oluşan bir veri kümesinde  $(n \times k)$  boyutlu açıklayıcı değişkenler girdi matrisi  $X$  ve bilinmeyen regresyon parametrelerinden oluşan  $(k \times 1)$  boyutlu vektör  $\beta$  ile ifade edildiğinde doğrusal önkestim;

$$\eta = X\beta \quad (1)$$

biçiminde tanımlanmaktadır.

Bağ fonksiyonu, modelin sistematik bileşeni ile rastgele bileşeni arasındaki ilişkiyi açıklayan fonksiyondur. Bağımlı değişkenlerin beklenen değeri, doğrusal önkestim bir fonksiyonu biçiminde ifade edilmektedir. Hasar sıklığının modellenmesinde logaritmik bağ fonksiyonu kullanılmaktadır. Bağımlı değişkenlerden oluşan  $(n \times 1)$  boyutlu vektör  $Y$  ve bağımlı değişkenlerin beklenen değeri  $\mu$  ile ifade edildiğinde rastgele bileşen ile sistematik bileşen arasındaki ilişki,  $g(x)$  ile tanımlanan logaritmik bağ fonksiyonu kullanılarak;

$$\mu_i = g^{-1} \left( \sum_{j=1}^k x_{ij} \beta_j \right) = \exp(x_{i1} \beta_1) \cdot \exp(x_{i2} \beta_2) \cdot \dots \cdot \exp(x_{ik} \beta_k) \quad (2)$$

eşitliği ile açıklanmaktadır [11].

Geleneksel olarak hasar sıklığı dağılımının Poisson veya negatif binom dağılımlarından birine uyması beklenir. Ancak 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. Sıfır değerinde yığılma gösteren veri kümelerinin sıfır yığılmalı regresyon modelleri kullanılarak modellenmesi ile, veriye uyumda artış sağlanmaktadır.

### 2.1. Poisson, negatif binom ve genelleştirilmiş Poisson regresyon modelleri

Bağımlı değişkenlerin negatif olmayan kesikli sayılardan oluştuğu ve belirli bir zaman aralığında gerçekleşen olay sayısının küçük olduğu durumda, bu değişken için belirlenen regresyon modellerinden biri Poisson modelidir. Ancak, Poisson regresyonunun kullanımını sınırlandıran bir özellik, ortalamanın varyansa eşit olmasıdır ve bu varsayımın gözlemlenen veri kümeleri üzerinde sağlanması oldukça güçtür [7,8]. Verideki yayılımın Poisson modeliyle tahmin edilen yayılımdan fazla olduğu durum, istatistik literatüründe aşırı yayılım olarak tanımlanmaktadır. Aşırı yayılım sorunuyla karşılaşılan durumlarda Poisson modeli yerine, aşırı yayılımın da dikkate alındığı ve modele eklenen yayılım parametresi ile açıklandığı dağılımlardan herhangi biri kullanılarak modelleme yapılmaktadır. Karma Poisson dağılımlarında Poisson dağılımının parametresi ( $\mu$ ), 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. Genelleştirilmiş Poisson modeli, Poisson modelinin aşırı yayılım veya az yayılım nedeniyle uygun olmadığı durumlarda kullanılmaktadır. Negatif binom modeli yalnızca aşırı yayılım durumunda kullanılırken genelleştirilmiş Poisson modeli, aşırı yayılım ve az yayılım durumlarının her ikisi için de kullanılmaktadır [5].

### 2.2. Sıfır yığılmalı regresyon modelleri

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 [15]. 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 [1,6]. Sıfır yığılmalı modellerde hasar sayısı verisinin dağılımı için belirlenen dağılıma göre hasarı olmayan 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 [13,18]. Negatif olmayan ve sıfır yığılmalı dağılım gösteren kesikli raslantı değişkeni  $Y_i$  olmak üzere; açıklayıcı değişkenlerin hem yapısal sıfır oranı olan  $\omega$  için, hem de temel sayı dağılımı için kullanıldığı sıfır yığılmalı regresyon modelleri;

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i) \Pr(K_i = 0), & y_i = 0 \\ (1 - \omega_i) \Pr(K_i = y_i), & y_i > 0 \end{cases} \quad (3)$$

biçiminde tanımlanmaktadır. Eş.(3)' te tanımlanan  $K_i$  raslantı değişkeni, hasar sayısının modellenmesi 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 modellere 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,

$$E(Y_i) = (1 - \omega_i)E(K) \quad (4)$$

$$Var(Y_i) = (1 - \omega_i) \left\{ Var(K) + \omega_i [E(K)]^2 \right\} \quad (5)$$

biçimindedir.

Yapısal sıfır oranı  $\omega_i$ , regresyon çözümlemesi ile, veya yalnızca lojit dönüşümün uygulandığı tek değişkenli çözümleme ile elde edilebilmektedir. Regresyon çözümlemesi yapılarak elde edilen  $\omega_i$  oranı için;

$$lojit(\omega_i) = \ln(\omega_i / (1 - \omega_i)) = G_i \gamma \quad i = 1, 2, \dots, n \quad (6)$$

eşitliği kullanılmaktadır.

Sıfır yığılmalı modellerden sıfır yığılmalı Poisson (ZIP), sıfır yığılmalı negatif binom (ZINB) ve sıfır yığılmalı genelleştirilmiş Poisson (ZIGP) modelleri, birçok alanda yaygın olarak kullanılmaktadır. Sıfır yığılmalı Poisson, sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modellerinin pozitif sayı değerleri için olasılık fonksiyonu, beklenen değer ve varyansı Tablo 1.'de verilmiştir.

**Tablo 1.** Sıfır yığılmalı regresyon modelleri

Model Adı	$\Pr(Y_i = y_i) (y_i > 0)$	$E(Y_i)$	$Var(Y_i)$
ZIP	$(1 - \omega_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$	$(1 - \omega_i) \mu_i$	$E(Y_i)(1 + \omega_i \mu_i)$
ZINB	$(1 - \omega_i) \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) \Gamma(\alpha^{-1})} \left( \frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}}$	$(1 - \omega_i) \mu_i$	$E(Y_i)(1 + \alpha \mu_i + \omega_i \mu_i)$
ZIGP	$(1 - \omega_i) \frac{\mu_i (\mu_i + (\varphi - 1) y_i)^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp\left(-\frac{1}{\varphi} (\mu_i + (\varphi - 1) y_i)\right)$	$(1 - \omega_i) \mu_i$	$E(Y_i)(\varphi^2 + \omega_i \mu_i)$

Sıfır yığılmalı Poisson regresyon modelinde, veride mevcut olan aşırı yayılımın yalnızca sıfır değerindeki yoğunluktan kaynaklandığı varsayımı yapılmaktadır. Sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modellerinde ise aşırı yayılımın sıfır değerindeki yığılmanın yanı sıra birçok nedenden kaynaklanabileceği varsayımı mevcuttur. Dolayısıyla tüm sıfır yığılmalı regresyon modellerinin ortalaması birbirine eşit iken varyans yapıları değişmektedir.

### 3. Model kurulumu ve seçimi

Veri kümesine en uygun modelin belirlenebilmesi için öncelikle betimsel istatistiklerin yorumlanması ve veri kümesinin modellenmesinde kullanılabilecek olası modellerin belirlenmesi gerekmektedir. Olası modellerin belirlenme sürecinde aşırı yayılım probleminin varlığı test edilmelidir. Aşırı yayılım, Poisson regresyon modelinin varsayımlarından biri olan eşit yayılımın (equidispersion) sağlanmadığı durumda ortaya çıkmaktadır. Verinin aşırı yayımlı olup olmadığına karar vermek için izlenen yollardan biri, kurulan Poisson regresyon 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 geliştirilen  $T$  ve  $T_1$  test istatistiklerinin hesaplanmasıdır [2].  $T$  ve  $T_1$  test istatistikleri sırasıyla;

$$T = \frac{1}{2} \sum_{i=1}^n \left\{ (Y_i - \hat{\mu}_i)^2 - Y_i \right\} \quad (7)$$

$$T_1 = \frac{\sum_{i=1}^n \left\{ (Y_i - \hat{\mu}_i)^2 - Y_i \right\}}{\left( 2 \sum_{i=1}^n \hat{\mu}_i^2 \right)^{1/2}} \quad (8)$$

eşitlikleri ile hesaplanmaktadır.  $T$  istatistiğinin pozitif büyük değer alması verinin Poisson regresyon modeline göre aşırı yayılım, negatif büyük değer alması ise verinin Poisson regresyon modeline göre az yayılım gösterdiğini ifade etmektedir.  $T_1$  test istatistiği, yeterince büyük gözlem sayısı için standart normal dağılıma yakınsamaktadır.

Veri kümesine uygunluğu araştırılan modeller; olabilirlik oran, sapma, genelleştirilmiş Pearson ki-kare gibi uyum iyiliği test istatistikleri kullanılarak belirlenebilmektedir. Sıfır yığılmalı regresyon modellerinin uygunluğu için yapısal sıfır oranı olan  $\omega$ 'nın sıfır değerine eşitliği test edilebilmektedir. Özel olarak Poisson regresyon modeli ile sıfır yığılmalı Poisson regresyon modelinin karşılaştırılması amacıyla skor testi, negatif binom regresyon modeli ile sıfır yığılmalı negatif binom regresyon modelinin ve genelleştirilmiş Poisson regresyon modeli ile sıfır yığılmalı genelleştirilmiş Poisson regresyon modelinin karşılaştırılması amacıyla ise Vuong testi geliştirilmiştir [8,16].

Veri kümesine uygunluğu araştırılan modeller içinden en uygun modelin belirlenmesi amacıyla literatürde yaygın olarak kullanılan Akaike ve Bayes model seçim ölçütlerinden faydalanılmaktadır.

### 4. Uygulama

2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptırmış sigortalıların yaş ve cinsiyet bilgileri ile aracın kullanım tipi, il plaka kodu, sigorta poliçesinin başlangıç tarihi ve yıl içinde gerçekleşen hasar sayısı bilgisini içeren ve TRAMER'den alınan veri kümesi kullanılarak gerekli analizler yapılmıştır. 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. 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.

2008 yılı içinde gerçekleşen hasar sayısı bağımlı değişken olmak üzere, regresyon çözümlemesinde kullanılan açıklayıcı değişkenler yaş, cinsiyet, kullanım tipi ve sektör payıdır. Her bir ilde 2008 yılının ocak ayında yürürlüğe giren poliçe sayısının tüm illerde yürürlüğe giren poliçeler içindeki yüzdesi sektör payı olarak tanımlanmaktadır.

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 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.

Çözümleme sürecinde, incelenen tüm modeller gözlemlenen veri kümesine uygulanmış ve her bir model için anlamsız bulunan açıklayıcı değişkenler çıkartılmıştır. 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.

İlk açıklayıcı değişken kümesi olan yaş, cinsiyet ve araç kullanım tipi değişkenleri kullanılarak oluşturulan regresyon modellerinde kullanım tipi tüm modeller için anlamsız bulunmuş ve çıkartılmıştır. Her bir model için anlamlı bulunan değişkenler ile oluşturulan regresyon modelleri, en çok olabilirlik yöntemi ile elde edilen parametre tahminleri ve hesaplanan model seçim ölçütleri Tablo 2.'de verilmiştir.

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. 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 büyük 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 skor test istatistiği 1114,2 olarak hesaplanmıştır. Dolayısıyla, gözlemlenen hasar sayısının sıfır değerindeki yoğunluk, Poisson modeli varsayımı altında beklenen yoğunluktan fazladır ve sıfır yığılmalı regresyon modellerinin gözlemlenen hasar sayısının modellenmesi amacıyla kullanılması uygun bulunmuştur.

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 2.'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.

**Tablo 2.** Yaş, cinsiyet ve kullanım tipi değişkenlerinden anlamlı bulunan değişkenler ile oluşturulan regresyon modelleri, parametre tahminleri ve model seçim ölçütleri

Açıklayıcı Değişkenler	$\mu$ Regresyonu				
	Poisson	NB	GP	ZIP	ZINB
Sabit terim (intercept)	$\beta_1$ : -2,3233 (0,0399) p<0,1	$\beta_1$ : -2,3282 (0,0412) p<0,1	$\beta_1$ : -2,3233 (0,0412) p<0,1	$\beta_1$ : -1,2797 (0,0518) p<0,1	$\beta_1$ : -2,2559 (0,0955) p<0,1
Yaş	$\beta_2$ : -0,0163 (0,0008) p<0,1	$\beta_2$ : -0,0162 (0,0009) p<0,1	$\beta_2$ : -0,0162 (0,0009) p<0,1	$\beta_2$ : -0,0161 (0,0009) p<0,1	$\beta_2$ : -0,0161 (0,0009) p<0,1
Cinsiyet	$\beta_3$ : -0,0838 (0,0224) p<0,1	$\beta_3$ : -0,0838 (0,0229) p<0,1	$\beta_3$ : -0,0838 (0,0229) p<0,1	$\beta_3$ : -0,0849 (0,0233) p<0,1	$\beta_3$ : -0,1557 (0,0865) p<0,1
Yayılm parametresi (dispersion parameter)	---	$\alpha$ : 0,5125 (0,0263) p<0,1	$\phi$ : 1,0477 (0,0024) p<0,1	---	$\alpha$ : 0,5241 (0,0341) p<0,1
Açıklayıcı Değişkenler	$\omega$ Regresyonu				
Sabit terim (intercept)	---	---	---	$\gamma_1$ : 0,6174 (0,0485) p<0,1	$\gamma_1$ : 0,8150 (0,0122) p<0,1
AIC	112.721,0	111.926,0	111.934,0	111.951,1	<b>111.911,2</b>
BIC	112.752,9	111.968,2	111.975,9	111.993,4	<b>111.964,1</b>
Pearson ki-kare	317.453,8	290.442,7	316.928,0	291.675,3	<b>290.402,2</b>
lnL	-56.357,6	-55.958,9	-55.962,8	-55.971,5	<b>-55.950,6</b>

\*Parantez içinde yer alan değerler, parametrelerin standart sapmasını ifade etmektedir.

İkinci açıklayıcı değişken kümesi olan yaş, cinsiyet ve sektör payı değişkenlerinden anlamlı bulunan değişkenler ile oluşturulan regresyon modelleri, en çok olabilirlik yöntemi ile elde edilen parametre tahminleri ve hesaplanan model seçim ölçütleri Tablo 3.'te verilmiştir.

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 olduğu 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.  $T$ ,  $T_1$  ve skor istatistikleri sırasıyla 639,33; 33,47 ve 954,5 olarak hesaplanmıştır.

Tablo 3.'te görüldüğü üzere, 2008 yılının ocak ayında Karayolları Motorlu Araçlar Zorunlu Mali Sorumluluk Sigortası yaptıran sigortalıların yaş ve cinsiyet bilgileri ile araçların sigortalandığı ilin sektör payları göz önünde bulundurulduğunda bir yıl içinde gerçekleşen hasar sayısının, sıfır yığılmalı genelleştirilmiş Poisson regresyon modeli kullanılarak modellenmesi uygun bulunmuştur. Ancak 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.

**Tablo 3.** Yaş, cinsiyet ve sektör payı değişkenlerinden anlamlı bulunan değişkenler ile oluşturulan regresyon modelleri, parametre tahminleri ve model seçim ölçütleri

Açıklayıcı Değişkenler	μ Regresyonu					
	Poisson	NB	GP	ZIP	ZINB	ZIGP
Sabit terim (intercept)	$\beta_1$ : -2,6527 (0,0370) p<0,1	$\beta_1$ : -2,6619 (0,0387) p<0,1	$\beta_1$ : -2,6528 (0,0381) p<0,1	$\beta_1$ : -1,4816 (0,0711) p<0,1	$\beta_1$ : -2,2143 (0,0742) p<0,1	$\beta_1$ : -2,7498 (0,05015) p<0,1
Yaş	$\beta_2$ : -0,0161 (0,0008) p<0,1	$\beta_2$ : -0,0159 (0,0009) p<0,1	$\beta_2$ : -0,0161 (0,0009) p<0,1	$\beta_2$ : -0,0160 (0,0009) p<0,1	$\beta_2$ : -0,0173 (0,0012) p<0,1	$\beta_2$ : -0,0143 (0,0011) p<0,1
Cinsiyet	---	---	---	---	$\beta_3$ : 0,0630 (0,0312) p<0,1	$\beta_3$ : 0,0674 (0,0294) p<0,1
Sektör Payı	$\beta_3$ : 2,6404 (0,0799) p<0,1	$\beta_3$ : 2,6508 (0,0839) p<0,1	$\beta_3$ : 2,6405 (0,0823) p<0,1	$\beta_3$ : 1,0004 (0,4831) p<0,1	$\beta_4$ : 0,7266 (0,2098) p<0,1	$\beta_4$ : 0,8431 (0,3263) p<0,1
Yayılm parametresi (dispersion parameter)	---	$\alpha$ : 0,5718 (0,0305) p<0,1	$\phi$ : 1,0477 (0,0024) p<0,1	---	$\alpha$ : 0,8101 (0,0612) p<0,1	$\phi$ : 1,0368 (0,0026) p<0,1
Açıklayıcı Değişkenler	ω Regresyonu					
Sabit terim (intercept)	---	---	---	$\gamma_1$ : 0,8171 (0,0894) p<0,1	---	---
Yaş	---	---	---	---	$\gamma_1$ : -0,0073 (0,0042) p<0,1	$\gamma_1$ : 0,0103 (0,0042) p<0,1
Cinsiyet	---	---	---	---	$\gamma_2$ : 0,4723 (0,1674) p<0,1	$\gamma_2$ : -0,3706 (0,1047) p<0,1
Sektör Payı	---	---	---	$\gamma_2$ : -2,6883 (0,8160) p<0,1	$\gamma_3$ : -25,0819 (2,7884) p<0,1	---
AIC	111.700,0	110.993,0	111.001,0	111.006,0	110.813,6	<b>110.805,0</b>
BIC	111.731,3	111.044,2	111.043,5	111.058,8	110.898,3	<b>110.879,0</b>
Pearson ki-kare	315.140,0	290.948,2	314.089,0	290.760,9	288.543,8	<b>314.285,0</b>
lnL	-55.846,8	-55.492,3	-55.496,6	-55.498,0	-55.398,8	<b>-55.395,5</b>

\*Parantez içinde yer alan değerler, parametrelerin standart sapmasını ifade etmektedir.

## 5. Sonuç ve öneriler

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. 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 değerın ekonomik kayıp olarak da tanımlanan beklenen hasar tutarı ile aktüeryal denge gözönünde bulundurularak birleştirilmesi ile 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.

Bu çalışmada, hasar sayılarının modellenmesi amacıyla sıfır yığılmalı regresyon modellerinin kullanımı araştırılmış ve Türkiye'ye ait gözlemlenen veri kümesi kullanılarak uygulanmıştır. Farklı 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.

## Kaynaklar

- [1] 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.
- [2] Dean, C., Lawless, J.F., 1989, Test for Defecting Overdispersion in Poisson Regression Models, *Journal of American Statistical Association*, 84, 467-472.
- [3] Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.F., 2007, Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems, 356p.
- [4] Faddy, M.J., 1998, Stochastic Models for Analysis of Species Abundance Data, *Statistics in Ecology and Environmental Monitoring*, 2, D.J. Fletcher, L. Kavalieris, B.F.J. Manly (eds.), Dunedin: University of Otago Press, 33-40.
- [5] 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.
- [6] Flynn, M., Francis, A., L., 2009, More Flexible GLMs: Zero Inflated Models and Hybrid Models, *CAS Ratemaking and Product Management Seminar*.
- [7] Giles, D., 2010, Notes on the Zero Inflated Poisson Model, Department of Economics, University of Victoria.
- [8] Greene, W. H., 1994, Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Technical report.
- [9] Lambert, D., 1992, Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing, *Technometrics*, 34, No.1.
- [10] 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.
- [11] McCullagh, P., Nelder, J.A., 1989, Generalized Linear Models, 2nd Ed., Chapman and Hall, London, 511p.
- [12] Mullahy, J., 1986, Specification and Testing of Some Modified Count Data Models, *Journal of Econometrics*, 33, 341-365.
- [13] 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.
- [14] Ridout, M., Demetrio, C.G.B., Hinde, J., 1998, Models for Count Data with Many Zeros, International Biometric Conference, Capetown December.
- [15] Shankar, V., Milton, J., Mannering, F., 1997, Modeling Accident Frequencies As Zero-Altered Probability Processes: An Empirical Inquiry, *Accident Analysis and Prevention*, 29, 6, 829-837.
- [16] Van den Broek, J., 1995, A Score Test for Zero-Inflation In a Poisson Distribution, *Biometrics*, 51, 738-743.
- [17] 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.
- [18] 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.