

Aşırı ve Az Yayılımlı Sayım Verileri için Regresyon Modellerinin Karşılaştırılması

ANDI ARISYI ZULWAQAR

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

Ocak 2022

Comparison of Regression Models for Over- and Under-Dispersed Count Data

ANDI ARISYI ZULWAQAR

MASTER SCIENCE THESIS

Department of Statistics

January 2022

Aşırı ve Az Yayılımlı Sayım Verileri İçin Regresyon Modellerinin Karşılaştırılması

ANDI ARISYI ZULWAQAR

Eskişehir Osmangazi Üniversitesi

Fen Bilimleri Enstitüsü

Lisansüstü Yöntemliği Uyarınca

İstatistik Anabilim Dalı

İstatistik Teorisi Bilim Dalında

YÜKSEK LİSANS TEZİ

Olarak Hazırlanmıştır

Danışman: Prof. Dr. ÖZLEM ALPU

Ocak 2022

ETİK BEYAN

Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kılavuzuna göre, Prof. Dr. ÖZLEM ALPU danışmanlığında hazırlamış olduğum “Aşırı ve Az Yayılımlı Sayım Verileri için Regresyon Modellerinin Karşılaştırılması” başlıklı YÜKSEK LİSANS tezimin özgün bir çalışma olduğunu; tez çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; tezimde verdiğim bilgileri, verileri akademik ve bilimsel etik ilke ve kurallara uygun olarak elde ettiğimi; tez çalışmamda yararlandığım eserlerin tümüne atıf yaptığımı ve kaynak gösterdiğimi ve bilgi, belge ve sonuçları bilimsel etik ilke ve kurallara göre sunduğumu beyan ederim. 18/01/2022.

ANDI ARISYI ZULWAQAR

İmza

ÖZET

Regresyon analizi, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki nedensel ilişkiyi belirlemek için kullanılan bir yöntemdir. Regresyon analizi, sayım verileri olmak üzere çeşitli veri türleri için kullanılabilir. Sayım verileri, gerçek yaşamda çeşitli durumlarda çok sık bulunan bir veri türüdür. Sayım verilerini analiz etmek için en sık kullanılan regresyon analizi Poisson regresyondur. Varyansın ortalamadan büyük olduğu sayım verilerinde genellikle bir aşırı yayılım problemi söz konusu olurken, varyansın ortalamadan daha küçük olduğu durumda da az yayılım söz konusudur. Bu çalışmada, sayım verisindeki aşırı yayılım ve az yayılım probleminin çözmek için Conway-Maxwell *Poisson* regresyonu ve genelleştirilmiş *Poisson* regresyonu kullanılmıştır. Çalışmada, Conway-Maxwell *Poisson* regresyonunun ve genelleştirilmiş *Poisson* regresyonunun sayım verilerinde aşırı yayılım ve az yayılımın performanslarını belirlemek için simülasyon çalışması ve gerçek bir veri setinde bir uygulaması yapılacaktır. Çalışmanın sonunda Conway-Maxwell *Poisson* regresyon modeli ve genelleştirilmiş *Poisson* regresyon modeli için AIC ve BIC kriterleri sayım verilerinde aşırı yayılım ve az yayılımın üstesinden gelmede hangi modelin daha iyi olduğunu bulmak amacıyla karşılaştırılacaktır.

Anahtar Kelimeler: *Sayım Verisi, Aşırı Yayılım, Az Yayılım, Conway-Maxwell Poisson Regresyonu, Genelleştirilmiş Poisson Regresyonu.*

SUMMARY

Regression analysis is a method used to determine the causal relationship between one dependent variable and one or more independent variables. Regression analysis can be used for various types of data, one of which is count data. Count data is one type of data that is very often run into various cases in real life. The regression analysis commonly used to analyze the count data is *Poisson* regression. The count data often can have an overdispersion problem, where the variance is greater than the mean, or underdispersion problem, where the variance is smaller than the mean. In this thesis, researcher use two methods, namely Conway-Maxwell *Poisson* regression and Generalized *Poisson* regression to overcome the problem of overdispersion and underdispersion in the count data. This study will simulate the performance of the Conway-Maxwell *Poisson* regression and Generalized *Poisson* regression for overdispersed or underdispersed count data. At the end of the thesis, the researcher will compare which model is better at overcoming overdispersion and underdispersion in terms of AIC and BIC criteria in the count data.

Keywords: *Count Data, Overdispersion, Underdispersion, Poisson Regression, Conway-Maxwell Poisson Regression, Generalized Poisson Regression.*

İÇİNDEKİLER

Sayfa

ÖZET	vi
SUMMARY	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
ŞEKİLLER DİZİNİ.....	xi
ÇİZELGELER DİZİNİ.....	xii
1. GİRİŞ VE AMAÇ	1
2. LİTERATÜR ARAŞTIRMASI.....	3
3. MATERYAL VE YÖNTEM	7
3.1. <i>Poisson</i> Regresyonu.....	7
3.1.1. Sapma istatistiği.....	10
3.1.2. Kısmi hipotez testi (Partial testi)	11
3.1.3. Çoklu doğrusallık testi	11
3.1.4. Yayılım testi.....	11
3.2. Genelleştirilmiş <i>Poisson</i> Regresyonu	12
3.3. Conway-Maxwell <i>Poisson</i> Regresyonu.....	13
3.4. Model Seçim Kriterleri	15
4. BULGULAR VE TARTIŞMA	16
4.1. Simülasyon Çalışması.....	16
4.1.1. Aşırı yayılım.....	17
4.1.1.1. <u>Yayılım parametresinin 0.3 olması durumu</u>	17
4.1.1.2. <u>Yayılım parametresinin 0.5 olması durumu</u>	19
4.1.1.3. <u>Yayılım parametresinin 0.75 olması durumu</u>	22
4.1.2. Eşit yayılım.....	24
4.1.3. Az yayılım	27

İÇİNDEKİLER (devam)

Sayfa

4.1.3.1. <u>Yayılım parametresinin 1.25 olması durumu</u>	27
4.1.3.2. <u>Yayılım parametresinin 1.5 olması durumu</u>	30
4.1.3.3. <u>Yayılım parametresinin 2 olması durumu</u>	32
4.2. Gerçek Veri Seti.....	35
4.2.1. <i>Poisson</i> regresyonu modeli tahmin sonuçları	36
4.2.2. Çoklu doğrusallık testi sonuçları	38
4.2.3. <i>Poisson</i> dağılımı testi sonuçları	38
4.2.4. Yayılım testi sonuçları	39
4.2.5. Genelleştirilmiş <i>poisson</i> regresyon modeli tahmin sonuçları	39
4.2.6. Conway-Maxwell <i>poisson</i> regresyon modeli tahmin sonuçları.....	40
4.2.7. Uygun modelin belirlenmesi	41
5. SONUÇ VE ÖNERİLER	42
KAYNAKLAR DİZİNİ.....	44
EK AÇIKLAMALAR.....	48

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
4.1. Yayılım değeri 0.3 olması durumundaki AIC değerlerinin grafiği.....	18
4.2. Yayılım değeri 0.3 olması durumundaki BIC değerlerinin grafiği.....	19
4.3. Yayılım değeri 0.5 olması durumundaki AIC değerlerinin grafiği.....	21
4.4. Yayılım değeri 0.5 olması durumundaki BIC değerlerinin grafiği.....	21
4.5. Yayılım değeri 0.75 olması durumundaki AIC değerlerinin grafiği.....	23
4.6. Yayılım değeri 0.75 olması durumundaki BIC değerlerinin grafiği.....	24
4.7. Yayılım değeri 1 olması durumundaki AIC değerlerinin grafiği.....	26
4.8. Yayılım değeri 1 olması durumundaki BIC değerlerinin grafiği.....	26
4.9. Yayılım değeri 1.25 olması durumundaki AIC değerlerinin grafiği.....	29
4.10. Yayılım değeri 1.25 olması durumundaki BIC değerlerinin grafiği.....	29
4.11. Yayılım değeri 1.5 olması durumundaki AIC değerlerinin grafiği.....	31
4.12. Yayılım değeri 1.5 olması durumundaki BIC değerlerinin grafiği.....	32
4.13. Yayılım değeri 2 olması durumundaki AIC değerlerinin grafiği.....	33
4.14. Yayılım değeri 2 olması durumundaki BIC değerlerinin grafiği.....	34

ÇİZELGELER DİZİNİ

<u>Cizelge</u>	<u>Sayfa</u>
4.1. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 0.3 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerler.....	17
4.2. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 0.5 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	20
4.3. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 0.75 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	22
4.4. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 1 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	25
4.5. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 1.25 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	28
4.6. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 1.5 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	30
4.7. Genelleştirilmiş Poisson regresyon modeli ve Conway-Maxwell Poisson regresyon modeli için 2 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.....	33
4.8. 2015 yılında Orta Java eyaletindeki HIV hasta sayısına ve onu etkileyen değişkenlere ilişkin veriler.....	35
4.9. Poisson regresyon modeli için parametre tahminin değerleri.....	37
4.10. Olabilirlik Oranı Test değeri.....	37
4.11. Varyans şişme faktörü (VIF) değerleri.....	38
4.12. Poisson dağılımı için uyum iyiliği (Goodness of fit) değeri.....	38
4.13. Genelleştirilmiş Poisson regresyon modeli için parametre tahmin değerleri.....	39
4.14. Conway-Maxwell Poisson regresyon modeli için parametre tahmin değerleri.....	40
4.15. AIC ve BIC değerleri.....	41

1. GİRİŞ VE AMAÇ

Poisson regresyon modeli, sayım verilerini analiz edebilen üstel regresyon modellerinden biridir (Berk ve MacDonald, 2008). *Poisson* dağılımına sahip bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki nedensel ilişkiyi belirlemek için kullanılır. Bu regresyon modelinde bağımlı değişkenin *Poisson* dağılımı, sayım verilerinin kullanılması ve beklenen değer varyans değeriyle eşit olması varsayımları söz konusudur (Saffari vd., 2011). Bununla birlikte, *Poisson* regresyon modeli kullanılarak analiz edilen sayım verileriyle ilgili aşırı yayılım ve az yayılım sorunu gibi sorunlar ortaya çıkabilir.

Aşırı yayılım, varyans değerinin beklenen değerden daha büyük olduğu anlamına gelirken, az yayılım, varyans değerinin beklenen değerden daha küçük olduğu anlamına gelmektedir (McCullagh ve Nelder, 1989). Bağımlı değişkendeki eksik veriler, verilerdeki aykırı değerler, gözlemler arasındaki etkileşimler ve link fonksiyonu belirleme hataları nedeniyle aşırı yayılım ve az yayılım meydana gelebilir. Verilerde aşırı yayılım veya az yayılım olduğu halde araştırmacı halen *Poisson* regresyon modeli analizini kullanıyorsa, olması gerekenden düşük veya yüksek tahmin edilen standart hata değerleri nedeniyle modelin doğru yorumu geçersiz hale gelebilecektir. Bu nedenle, aşırı yayılım ve az yayılım sorunlarının üstesinden gelebilecek modellere ihtiyaç duyulur.

Literatürde *Poisson* regresyon modelinde ortaya çıkabilecek bu sorunlarla baş edebilmek amacıyla negatif binom regresyonu, Conway-Maxwell *Poisson* regresyonu, genelleştirilmiş *Poisson* regresyonu, çift *Poisson* regresyonu, sıfır şişirilmiş *Poisson* regresyonu, quasi *Poisson* regresyonu geliştirilmiştir.

Bu tez çalışmasında negatif değer almayan kesikli değerler söz konusu olduğunda kullanılması önerilen sayım verisi modellerinden *Poisson* regresyon, Conway Maxwell *Poisson* regresyonu ve genelleştirilmiş *Poisson* regresyonu modelleri tanıtılmıştır. Ardından *Poisson* modeli ile analiz edilen sayım verilerinde aşırı yayılım ve az yayılım problemlerini aynı anda çözebilmesinden dolayı Conway-Maxwell *Poisson* regresyonu ve genelleştirilmiş *Poisson* regresyonunun performanslarının karşılaştırılması için kullanılacak model seçim kriterlerine yer verilmiştir. Çalışmada, ele alınan regresyon modellerinden daha iyi olan modeli belirlemek amacıyla gerçekleştirilecek simülasyon senaryoları ile belirli şartlarda her

iki modelin bu sorunlarla baş etme performanslarının karşılaştırılması ve yorumlanması ile gerçek veriler üzerinde bu modelleri kullanmak isteyen araştırmacılara yön gösterilmesi hedeflenmektedir.



2. LİTERATÜR ARAŞTIRMASI

Regresyon modelleri günümüzde bağımlı değişken ve bağımsız değişkenler arasındaki ilişkiyi belirlemek için çok sık kullanılmaktadır ve ayrıca regresyon modelleriyle analiz edilebilecek çeşitli veri türleri vardır, bunlardan biri sayım verisidir. Sayım verileri negatif değer almayan kesikli değerlerden oluşur ve bir zaman sürecinde bir durumun gerçekleşme sayısını ifade eder. Sayım verilerini modellemede en sık kullanılan regresyon modellerinden biri *Poisson* regresyon modelidir.

Poisson regresyon modelinin kullanımı için çeşitli koşullar söz konusudur. *Poisson* regresyon modelini kullanmak için karşılanması gereken koşulları sıralamak gerekirse, bağımlı değişken değerlerinin sayım verileri biçiminde olması, bağımlı değişkenin *Poisson* dağılımına sahip olması ve beklenen değer ile varyans değerinin aynı olmasıdır. Gerçek hayatta ise sıklıkla *Poisson* regresyon modeli kullanılarak modellenen sayım verilerinde aşırı yayılım ve az yayılım gibi sorunlar ortaya çıkabilir.

Poisson regresyon modeli kullanılarak analiz edilen sayım verisi modellerini ve problemlerini inceleyen çok sayıda çalışma söz konusudur. Bunlardan bazılarında izleyen satırlarda yer verilmektedir.

Consul ve Jain (1973) *Poisson* dağılımının iki parametreliliği yeni bir genelleştirilmiş halini, genelleştirilmiş negatif binom dağılımının sınırlandırılmış bir şekli olarak elde etmişlerdir. Parametrelerin farklı değerleri için dağılım şeklindeki değişimleri gösteren diyagramlar verilmiştir.

Tripathi ve Gupta (1984) çalışmalarında, genelleştirilmiş *Poisson* dağılımının her iki parametresi için minimum ki-kare tahmin edicileri ve maksimum olabilirlik tahmin edicilerini geliştirmiştir. Bu tahminciler, altta yatan dağılımın *Poisson* olduğu hipotezini test etmek amacıyla bir istatistik önermede kullanılmıştır.

Shmueli vd. (2005) Conway-Maxwell *Poisson* dağılımının istatistiksel ve olasılıksal özelliklerini tanıtmış ve araştırmıştır. Bu dağılımın parametrelerini tahmin etmede kullanılan üç yaklaşımı (ağırlıklı en küçük kareler, en çok olabilirlik ve bayesçi tahmin yöntemi) sunmuşlar ve iki veri seti üzerinde dağılımın esnekliğini göstermişlerdir.

Ismail ve Jemain (2007), *Poisson* regresyon modeli ile analiz edilen sayım verilerindeki aşırı yayılım problemlerinin ele alınmasının, negatif iki terimli ve genelleştirilmiş *Poisson* regresyon modelleri kullanılarak çözülebileceğini belirtmektedir. Çalışmada, Genelleştirilmiş *Poisson*'un en iyi çözüm olduğu vurgulanmaktadır.

Jowaheer ve Khan (2009), Conwall-Maxwell *Poisson* regresyon modelinde tahmin tekniklerinden quasi olabilirliğin en çok olabilirliğe göre performansını araştırmışlardır. Yürüttükleri simülasyon çalışmasında quasi olabilirlik yaklaşımının, en çok olabilirlik yaklaşımına göre etkinlik kaybının ihmal edilebilir olduğunu göstermişlerdir.

Harris vd. (2012) araştırmalarında, *Poisson* regresyon modeli, genelleştirilmiş *Poisson* regresyon modeli ve Quasi *Poisson* regresyon modelini kullanarak az yayılım problemi olan sayım verileri üzerinde simülasyon çalışması gerçekleştirmiştir. Bu araştırma, az yayılım problemi doğru şekilde ele alınmazsa, oluşturulan modeldeki problemin yorumlanmasında kötü sonuçlar verebileceği sonucuna varmaktadır ve genelleştirilmiş *Poisson*'un en iyi çözüm olduğu vurgulanmaktadır.

Zhu (2012) Conway-Maxwell *Poisson* dağılımına dayalı bir zaman serisi sayım verilerini önerdiği bir INGARCH modeli ile analiz etmiştir. Çalışmasında elde edilen sonuçlar genelleştirilmiş *Poisson* ve Conway-Maxwell *Poisson* modellerinin çoğu durumda iyi performans gösterdiğini ve Conway-Maxwell *Poisson* modelinin genelleştirilmiş *Poisson* modeline kıyasla güçlü bir rakip olduğunu göstermektedir.

Sellers vd. (2012) farklı Conway-Maxwell *Poisson* modellerini ve bu modellerin pazarlama, elektronik ticaret, ulaşım ve biyoloji gibi alanlardaki uygulamalarını incelemektedir.

Avcı vd. (2015) araştırmalarında, negatif binom regresyon modeli, quasi *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* modelini kullanarak az yayılım problemi olan sayım verileri üzerinde gerçekleştirilmiştir. Bu araştırma sonucu, Conway-Maxwell *Poisson*, alga verilerindeki aşırı yayılım sorununu çözmek için negatif binom ve quasi *Poisson* modellerinden daha iyi olduğu bulunmuştur.

Yusuf ve Ugalahi (2015) arařtırmalarında, doğum öncesi bakım ziyaretlerinin kullanımını tanımlayan en iyi istatistiksel modeli belirlemek için *Poisson*, negatif binom ve genelleştirilmiş *Poisson* regresyon modellerini karşılařtırmıştır. Bu arařtırma sonucu, genelleştirilmiş *Poisson* regresyonunun, doğum öncesi bakım ziyaretlerinin sayısı ile ilişkili faktörleri belirlemede en iyi regresyon modeli olduğunu ifade etmektedir.

Huang (2017) arařtırmasında, log lineer model gibi modellerin basit ve daha kolay yorumlanabilmesi için Conway-Maxwell *Poisson* dağılımının ortalamaya dayalı nasıl parametrelendirilebileceğini göstermektedir. Bu arařtırma sonucu, analiz edilen sayım verileri ve gerçek veri setlerindeki aşırı yayılım ve az yayılım problemlerinin Conway-Maxwell *Poisson* regresyon modeli kullanılarak çözülebileceğini belirtmektedir.

Sellers ve Morris (2017) arařtırmalarında, seyrek dağılan çeřitli veri kaynaklarını ele alır ve seyrek dağılmış verileri modelleyebilen çeřitli dağılımları inceleyerek, bunların performanslarını uygulanan veri kümeleri üzerinde karşılařtırır. Bu arařtırma sonucu, Conway-Maxwell *Poisson* regresyon modelini ve genelleştirilmiş *Poisson* regresyon modelinin, *Poisson* modelindeki yayılım sorununun üstesinden gelmede iki esnek model olduğu belirtilmiştir.

Avcı (2018) arařtırmasında, Conway-Maxwell *Poisson* regresyon modelini, sayım verilerinde yer alan yayılım problemlerini başka birkaç modeli karşılařtırarak ele almanın esnekliğini göstermektedir. Bu arařtırma, *Poisson* regresyon modeli kullanılarak analiz edilen sayım verilerindeki az yayılım probleminin ele alınmasının Conway-Maxwell *Poisson* regresyon modeli kullanılarak çözülebileceğini belirtmektedir.

Hayati (2018) arařtırmasında, *Poisson* regresyon modeli ile analiz edilen sayım verilerinde yer alan aşırı yayılım ve az yayılım problemlerinin çözümüne yönelik bir simülasyon çalışması gerçekleřtirmiştir. Bu arařtırma, *Poisson* regresyon modeli kullanılarak analiz edilen sayım verilerindeki aşırı yayılım ve az yayılım problemlerinin Conway-Maxwell *Poisson* regresyon modeli kullanılarak çözülebileceğini belirtmektedir.

Chatla ve Shmueli'nin (2018) arařtırmalarında, IRLS (Iterative Reweighted Least Squares) yaklaşımına dayanan bir Conway-Maxwell *Poisson* modelini tahmin etmek için esnek bir tahmin yaklaşımını tanıtmaktadır. Bu arařtırma sonucu, Conway-Maxwell *Poisson* için IRLS algoritması, sıradan bir *Poisson* modeline kıyasla hesaplama açısından yoğun olsa

da, kümülanların hesaplanması gibi bazı hesaplamaların uygun şekilde paralelleştirilmesi ile hesaplama süresinin azaltılabileceği ve bu tür paralel hesaplamaların, özellikle büyük örneklerde faydalı olacağı belirtilmiştir.

Huang ve Kim (2021) Metropolis-Hasting algoritması aracılığıyla az ve aşırı yayımlı sayım verileri için Bayesçi Conway-Maxwell *Poisson* regresyonun istatistiksel çıkarımlarını göstermiştir. Ayrıca bu özellikleri bir simülasyon çalışmasında ve iki veri seti üzerinde incelemişlerdir.

Bu çalışmanın üçüncü bölümünde *Poisson* regresyonu ve aşırı yayılım/ az yayılım ile eşanlı olarak baş edebilen Conway-Maxwell *Poisson* regresyonu ile genelleştirilmiş *Poisson* regresyonu kavramlarına değinilecektir. Dördüncü bölümde ise farklı yayılım türlerine göre oluşturulan Monte Carlo simülasyon çalışmasına ve gerçek hayattan alınan veri seti üzerinde bu yöntemlerin uygulamasına yer verilecektir. Simülasyon çalışması ve gerçek veri seti uygulaması sonucunda hangi regresyon modelinin model seçim kriterleri bakımından daha iyi olduğu araştırılacaktır.

3. MATERYAL VE YÖNTEM

3.1. *Poisson* Regresyonu

Poisson regresyonu, *Poisson* dağılımına sahip olan bağımlı değişken ile bağımsız değişkenler arasındaki bağlantıyı incelemek için kullanılan bir regresyon modelidir. *Poisson* dağılımının olasılık fonksiyonu Eşitlik (3.1)' de

$$P(y_i; \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}; \quad y_i = 0, 1, 2, \dots \quad (3.1)$$

$\lambda_i > 0$ olmak üzere formüle edilir (Lindsey, 1997). Beklenen değer ve varyans değerleri aşağıdaki gibi gösterilir:

$$\begin{aligned} E(Y) &= \sum_{y=0}^{\infty} y \cdot \frac{e^{-\lambda_i} \lambda_i^y}{y!} \\ &= \sum_{y=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i^y}{(y-1)!} \\ &= \sum_{y=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i \lambda_i^{(y-1)}}{(y-1)!} \\ &= \lambda_i \sum_{y=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i^{(y-1)}}{(y-1)!} \\ &= \lambda_i \sum_{y=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i^d}{d!} \\ &= \lambda_i \end{aligned}$$

$$\begin{aligned} Var(Y) &= E(Y^2) - [E(Y)]^2 \\ &= E(Y^2 - Y + Y) - [E(Y)]^2 \\ &= E(Y(Y-1) + Y) - [E(Y)]^2 \\ &= E(Y(Y-1)) + E(Y) - [E(Y)]^2 \\ &= \sum_{y=0}^{\infty} y(y-1) \cdot \frac{e^{-\lambda_i} \lambda_i^y}{y!} + E(Y) - [E(Y)]^2 \\ &= \sum_{y=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i^y}{(y-2)!} + E(Y) - [E(Y)]^2 \\ &= \sum_{y=2}^{\infty} \frac{e^{-\lambda_i} \lambda_i^2 \lambda_i^{(y-2)}}{(y-2)!} + E(Y) - [E(Y)]^2 \\ &= \lambda_i^2 \sum_{y=2}^{\infty} \frac{e^{-\lambda_i} \lambda_i^{(y-2)}}{(y-2)!} + E(Y) - [E(Y)]^2 \end{aligned}$$

$$\begin{aligned}
&= \lambda_i^2 \sum_{y=2}^{\infty} \frac{e^{-\lambda_i} \lambda_i^y}{y!} + E(Y) - [E(Y)]^2 \\
&= \lambda_i^2 \cdot 1 + \lambda_i - \lambda_i^2 \\
&= \lambda_i
\end{aligned}$$

Yukarıda verilen ispat, Poisson dağılımının ortalaması ve varyansının $Var(Y) = E(Y) = \lambda_i$ (eşit yayılıma sahip) olduğunu göstermektedir. Eşitlik (3.1)'i regresyon amacıyla kullanmak için $y_i \stackrel{i.i.d.}{\sim} Poisson(\lambda_i)$, ve λ_i 'nin $1 \times k$ bağımsız değişken vektörünün ve $k \times 1$ katsayı vektörünün bir fonksiyonu olduğu varsayalım. Koşullu ortalama için parametrelili fonksiyonel model şöyledir:

$$E(Y|X) = \lambda_i = \exp(x_i^T \beta)$$

Poisson regresyonu modeli

$$y_i = \lambda_i + \varepsilon_i = \exp\left(\beta_0 + \sum_{j=1}^k \sum_{i=1}^n \beta_j x_{ji}\right) + \varepsilon_i \quad ; j = 1, 2, 3, \dots, k \quad i = 1, 2, \dots, n \quad (3.2)$$

olarak ifade edilebilse de (Francis, et. Al, 2012), bağımlı değişkenin ortalaması bakımından modellenmesi tercih edilir (Myers, Montgomery, Vining, and Robinson, 2010). Özellikle, bağımlı değişkenin ortalamasının doğrusal öntahminci (linear predictor) ile ilişkisini gösteren Eşitlik (3.3)'teki gibi bir bağlantı fonksiyonu kullanılır. Eşitlik (3.2)'de belirtilen gösterimler ise

x_{ji} = j'inci bağımsız değişken, i'inci gözlem

β_0 = sabit terim

β_j = j'inci bağımsız değişkenin parametresi

n = gözlem sayısı

k = parametre sayısı

y_i = i'inci bağımlı değişken değeri (olayların ortaya çıkış sayısı)

λ_i = birim zamanda olayların ortaya çıkış oranı veya dağılımın ortalaması

ε_i = i'inci gözlemin hata terimi

olarak tanımlanır.

Poisson regresyon modeli, sayım verilerini modellemek için kullanılan genelleştirilmiş doğrusal modellerden biridir. Genelleştirilmiş doğrusal modellerde *Poisson* regresyon modelini tanımlamak için, aşağıdaki gibi bir g bağlantı fonksiyonu tanımlanır:

$$g(\lambda_i) = \eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji} \quad (3.3)$$

Böylece, *Poisson* regresyon modelinde ortalama ve doğrusal öntahminci arasındaki ilişkiyi gösteren bağlantı fonksiyonu

$$g(\lambda_i) = \ln \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad (3.4)$$

$$e^{\ln \lambda_i} = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$$

$$\lambda_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji})$$

biçiminde yazılır. *Poisson* dağılımında yaygın olarak kullanılan birkaç bağlantı fonksiyonu vardır. En popüler olan bağlantı fonksiyonu Eşitlik (3.4)'te verilen logaritmik bağlantı (log link) fonksiyonudur. Bu fonksiyon bağımlı değişkenin tahmin edilen değerlerin tamamının pozitif olmasını sağladığı için özellikle tercih edilmektedir.

Poisson regresyonundaki parametreler en sık kullanılan tahmin tekniklerinden biri olan maksimum olabilirlik tahmini ile belirlenebilir (Cahyandari, 2014). Ayrıca pseudo en çok olabilirlik, *Poisson* genelleştirilmiş doğrusal modeller, en küçük kareler ve momentler tahmin teknikleri ile de tahminler yapılabilmektedir (Durmuş vd., 2020; Allo vd. 2019). *Poisson* dağılımının olasılık fonksiyonu, *Poisson* regresyon modelinin bağlantı fonksiyonu ile birleştirilerek Eşitlik (3.5)'teki gibi elde edilir:

$$P(y_i; \boldsymbol{\beta}) = \frac{\lambda_i(X_i; \boldsymbol{\beta})^{y_i} e^{-\lambda_i(X_i; \boldsymbol{\beta})}}{y_i!} \quad (3.5)$$

Yukarıdaki fonksiyondan, *Poisson* regresyon modeli için olabilirlik fonksiyonu Eşitlik (3.6)'te

$$L(y; \boldsymbol{\beta}) = \prod_{i=1}^n P(y_i; \boldsymbol{\beta})$$

$$= \prod_{i=1}^n \frac{\lambda_i(X_i; \boldsymbol{\beta})^{y_i} e^{-\lambda_i(X_i; \boldsymbol{\beta})}}{y_i!} \quad (3.6)$$

olarak ifade edilir. Dolayısıyla, olabilirlik fonksiyonunun doğal logaritması

$$\ln L(y; \boldsymbol{\beta}) = \sum_{i=1}^n (y_i \ln(\lambda_i(X_i; \boldsymbol{\beta})) - \lambda_i(X_i; \boldsymbol{\beta}) - \ln(y_i!)) \quad (3.7)$$

Eşitlik (3.7)'deki gibi bulunur (Sasu vd., 2016). Bu logaritmik olabilirlik fonksiyonu, β tahmin değerini elde etmek için logaritmik olabilirlik fonksiyonunun β parametresine göre türevi alınarak sıfıra eşitlenir (Cahyandari, 2014). Parameter tahminleri elde edildiğinde, tahmin edilen *Poisson* regresyon modeli (fitted *Poisson* regression model)

$$\hat{y}_i = g^{-1}(x_i^T \hat{\beta}) = \exp(x_i^T \hat{\beta})$$

elde edilir.

Poisson regresyon modelinin kullanılabilmesi için gerekli olan birkaç varsayım aşağıda verilmiştir (Roback vd., 2021):

- ***Poisson* değişkeni.** Bağımlı değişken, *Poisson* dağılımına sahip sayım verileri biçimindeki bir değişkendir.
- **Bağımsızlık.** Kullanılan bağımsız değişkenler birbirinden bağımsız olmalıdır.
- **Eşit yayılım.** *Poisson* dağılımına sahip bağımlı değişkenin beklenen değeri ve varyans değeri aynı olmalıdır.
- **Doğrusallık.** Lambda değerinin logaritması ($\log(\lambda)$) y'nin doğrusal bir fonksiyonu olmalıdır.

Oluşturulan *Poisson* regresyon modelinin seçimi için önerilen testlerden bazıları izleyen kısımda verilmiştir.

3.1.1. Sapma istatistiği

Bu test, bağımsız değişkenlerin aynı anda bağımlı değişken üzerinde bir etkisinin olup olmadığını incelemek için yapılır. Bu test için istatistiksel hipotezler aşağıda verilmiştir:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0,$$

$$H_1: \exists \beta_j \neq 0 \quad ; j = 1, 2, \dots, k.$$

Bu eşzamanlı testte kullanılan test istatistiği Eşitlik (3.8)'da (Putri vd., 2019)

$$D(\hat{\beta}) = -2 \ln(\Lambda) = 2(\ln L(\hat{\Omega}) - \ln L(\hat{w})) \quad (3.8)$$

olarak ifade edilir. Burada $L(\hat{w})$, tüm bağımsız değişkenleri içermeyen basitleştirilmiş (kısıtlı) model için olabilirlik değeridir ve $L(\hat{\Omega})$ ise, tüm bağımsız değişkeni içeren (kısıtsız) model için olabilirlik değeridir. Eğer $D(\hat{\beta}) > \chi^2_{(n-k-1;\alpha)}$ ise H_0 hipotezi α anlamlılık düzeyinde reddedilir (Putri vd., 2019).

3.1.2. Kısmi hipotez testi (Partial test)

Bu test, bağımsız değişkenlerin tek başına bağımlı değişkeni etkileyip etkilemediğini incelemek için yapılır. Bu test için istatistiksel hipotezler aşağıda verilmiştir:

$$H_0: \beta_j = 0 \quad (\text{j'inci bağımsız değişken etkisi istatistiksel olarak anlamlı değildir})$$

$$H_1: \beta_j \neq 0 \quad (\text{j'inci bağımsız değişken etkisi istatistiksel olarak anlamlıdır})$$

Bu kısmi testte uygulanan istatistiksel test Eşitlik (3.9)'da

$$W_j = \frac{\hat{\beta}_j^2}{\text{var}(\hat{\beta}_j)} \quad (3.9)$$

olarak ifade edilir (Putri vd., 2019). İstatistiksel karar kuralı ise şöyle ifade edilir, eğer $\alpha = 0,05$ anlamlılık düzeyinde $W_j > \chi^2_{(1;\alpha)}$ ise H_0 hipotezi reddedilir, diğer bir deyişle j'inci bağımsız değişkenin bağımlı değişken üzerinde istatistiksel olarak anlamlı olduğu söylenebilir (Putri vd., 2019).

3.1.3. Çoklu doğrusallık testi

Çoklu doğrusal bağıntı, bağımlı değişken üzerinde etkili olan bağımsız değişkenler arasında güçlü doğrusal ilişki olarak tanımlanır. Veri setinde çoklu bağıntının belirlenebilmesi için, Eşitlik (3.10)'da verilen varyans şişme faktörü (VIF, Variance Inflation Factor) adı verilen istatistik hesaplanır (Dobson, 2001).

$$VIF = \frac{1}{1 - R_j^2} \quad (3.10)$$

burada R_j^2 , bir bağımsız değişken ile diğer birkaç bağımsız değişken arasındaki söz konusu belirlilik katsayısıdır. Eğer her bağımsız değişken için hesaplanan VIF değeri 10'dan küçük ise, bağımsız değişkenler arasında çoklu bağlantının olmadığını belirtmektedir (Putri vd., 2019).

3.1.4. Yayılım testi

Poisson modeli $\text{Var}(Y) = E(Y) = \lambda_i$, eşit yayılım (*equidispersion*) olduğunu varsayarken, gerçekte bu koşul genellikle analiz edilen sayım verilerinde sağlanamayabilmektedir. *Poisson* modeli ile analiz edilen sayım verisinde ortaya çıkan sorun, aşırı yayılım ve az yayılım olarak ifade edilmektedir. $\text{Var}(Y) > E(Y)$, varyansın

beklenen değerdan daha yüksek olduđu bir aşırı yayılım durumudur; $Var(Y) < E(Y)$ ise varyansın beklenen değerdan daha düşük olduđu az yayılım durumudur.

Aşırı yayılım ve az yayılım, Eşitlik (3.11)'da verilen sapma (D) formülü kullanılarak tespit edilebilir (Darnah, 2011):

$$\theta = \frac{D}{db} \quad (3.11)$$

$$D = 2 \sum_{i=1}^n \left(y_i \ln \left(\frac{y_i}{\lambda_i} \right) - (y_i - \lambda_i) \right)$$

burada $db = n - p$ olmak üzere, modelde kullanılan parametre sayısı p ve n ise gözlem sayısıdır. $\theta > 1$ olduğunda veri setinde aşırı yayılma durumunun; $\theta < 1$ iken veri setinde az yayılım durumunun olduğu ifade edilir.

3.2. Genelleştirilmiş *Poisson* Regresyonu

Consul ve Jain'de (1973) tanıtılan ve Consul (1989) tarafından kapsamlı bir şekilde incelenen genelleştirilmiş *Poisson* dağılımından yola çıkılarak geliştirilen genelleştirilmiş *Poisson* regresyonu da, aşırı yayılım veya az yayılım sorunları içeren sayım verisini analiz etmek için kullanılan regresyon modellerinden biridir. Çeşitli genelleştirilmiş *Poisson* dağılımları önerilmiştir; özellikle, Consul (1989)'un çalışmaları, klasik *Poisson* dağılımını esnek bir şekilde genelleştiren λ ve b olmak üzere iki parametrelili bir dağılım ortaya koymuştur. Genelleştirilmiş *Poisson* regresyon modeli (Consul ve Famoye, 1992; Famoye, 1993) için olasılık fonksiyonu (Famoye, 2005) Eşitlik (3.12)'de verilmiştir,

$$f(y_i; \lambda_i, b) = \left(\frac{\lambda_i}{1+b\lambda_i} \right)^{y_i} \frac{(1+b\lambda_i)^{y_i-1}}{y_i!} \exp \left(-\frac{\lambda_i(1+b\lambda_i)}{(1+b\lambda_i)} \right), \quad y_i = 0, 1, 2, \dots \quad (3.12)$$

burada $\lambda_i = \lambda_i(x_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ olarak ifade edilir. Genelleştirilmiş *Poisson* dağılımının varyans ve beklenen değeri sırasıyla,

$$E(Y) = \lambda_i$$

ve

$$Var(Y) = \lambda_i(1 + b\lambda_i)^2$$

olarak formüle edilir.

Genelleştirilmiş *Poisson* regresyon modeli, *Poisson* regresyonunun genelleştirilmiş halidir. Diğer bir deyişle, $b = 0$ olduğunda *Poisson* regresyon modeli elde edilir. Burada $b > 0$ olduğunda, genelleştirilmiş *Poisson* regresyon modeli, aşırı yayılım sorunu içeren sayım verisini modellemede kullanılır. Öte yandan, $b < 0$ olması durumunda genelleştirilmiş *Poisson* regresyon modeli, az yayılım problemi içeren sayım verisini modellemek için kullanılmaktadır (Famoye, 2005).

Bu model için parametre tahmini, maksimum olabilirlik tahmini yöntemi uygulanarak gerçekleştirilmektedir (Consul ve Famoye, 1992). Genelleştirilmiş *Poisson* regresyon modelinin logaritmik olabilirlik fonksiyonu Eşitlik (3.13)'de

$$\ln L(\beta, b) = \sum_{i=1}^n \{y_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - y_i \ln(1 + b \exp(\mathbf{x}_i^T \boldsymbol{\beta})) + (y_i - 1) \ln(1 + b y_i) - \ln(y_i!) - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) (1 + b y_i)(1 + b \exp(\mathbf{x}_i^T \boldsymbol{\beta}))^{-1}\} \quad (3.13)$$

ile gösterilmiştir (Yang vd., 2009). Bu logaritmik olabilirlik fonksiyonu ile, β ve b parametrelerinin tahmin değerini elde etmek için, logaritmik olabilirlik fonksiyonu β ve b değişkenine göre türevi alınarak sıfıra eşitlenir (Famoye, 1993). Parametre tahminleri elde edildiğinde, tahmin edilen genelleştirilmiş *Poisson* regression modeli

$$E(y_i | x_i) = \hat{y}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$$

elde edilir (Durmuş vd., 2020).

3.3. Conway-Maxwell *Poisson* Regresyonu

Conway-Maxwell *Poisson* dağılımı, herhangi bir negatif olmayan tamsayı değerleri alabilen ve verilerdeki aşırı ve az yayılım durumlarını hesaba katan ilave bir yayılım parametresinin eklenmesiyle elde edilen *Poisson* dağılımının bir genellemesidir. Dağılım, Conway ve Maxwell tarafından 1962'de, duruma bağlı hizmet oranlarına (state dependent service rates) sahip kuyruk sistemlerini modellemek için tanıtılmıştır. Aşırı yayılım genellikle satış verilerinde, motorlu taşıt kazası sayılarında vb. bulunur. Az yayılım ise genellikle kelime uzunluğu, hava kargo kesintileri vb. ile ilgili verilerde bulunur (Sellers vd., 2012). Bu dağılımın istatistiksel özellikleri ve parametrelerini tahmin etme yöntemleri Shmueli ve ark. (2005) tarafından oluşturulmuştur.

Göreceli olarak yeni olmasının dışında, dağılımın popüleritesindeki artış, Conway-Maxwell *Poisson* dağılımının üstel aileye ait olmasından kaynaklanmaktadır. Üstel ailenin

bir çok faydalı özelliği bu dağılım için geliştirilmiştir. Ayrıca üstel ailenin kullanımı, dağılımı Bayes analizi ve diğer istatistiksel çıkarımların kullanımı için uygun hale getirmiştir. Dağılımın yeniden kullanılmaya başlanmasından bu yana, motorlu araç kazalarının analizi, elektrik gücü güvenilirliği, hanelerin perakende satış tahminleri ve kanser tekrarının tahmin edilmesi dahil olmak üzere, bu dağılımın uygulamalarında yakın zamanda bir artış olmuştur (Lord vd., 2008; Guikema ve Gofflet, 2008; Boatwright, Borle ve Kadane, 2003; Rodrigues vd., 2009).

Conway-Maxwell *Poisson* regresyon modeli de genelleştirilmiş *Poisson* regresyon gibi aşırı yayılım veya az yayılım sorunları içeren sayım verisini analiz etmek için kullanılan bir regresyon modelidir. Eşitlik (3.14)'de verilen Conway-Maxwell *Poisson* dağılımının olasılık fonksiyonu (Avcı, 2018)

$$P(Y_i=y_i;\lambda_i,v)=\frac{\lambda_i^{y_i}}{(y_i!)^v}\frac{1}{Z(\lambda_i,v)}, \quad y_i = 0,1,2, \dots \quad (3.14)$$

olmak üzere,

$$Z(\lambda_i, v) = \sum_{j=0}^{\infty} \frac{\lambda_i^j}{(j!)^v}, \quad \lambda_i > 0, v > 0$$

olarak ifade edilir. Conway-Maxwell *Poisson* dağılımının varyans ve beklenen değeri (Sellers vd., 2012)

$$E(Y) = \lambda_i \frac{\partial \log Z(\lambda_i, v)}{\partial \lambda} \approx \lambda_i^{\frac{1}{v}} - \frac{v-1}{2v} \quad (3.15)$$

ve

$$Var(Y_i) = \frac{1}{v} \lambda_i^{\frac{1}{v}}$$

olarak formüle edilir. Eşitlik (3.15)'teki yakınsamanın $v \leq 1$ veya $\lambda_i > 10^v$ için iyi sonuç verdiği belirtilmiştir (Minka vd., 2003). Parametre değeri v , veri yayılımının durumunu belirleyen bir parametredir. v parametresi 1 değerine sahipse, Conway-Maxwell *Poisson* regresyon modeli *Poisson* regresyon modeliyle aynıdır. v parametresi 1'den büyük bir değere sahipse, Conway-Maxwell *Poisson* regresyon modeli, az yayılımlı sayım verisini modellemede kullanılır. v parametresi 1'den küçük bir değere sahipse, Conway-Maxwell *Poisson* regresyon modeli aşırı yayılımlı sayım verisini modellemede kullanılır (Huang, 2017).

Çalışmada Conway-Maxwell *Poisson* parametreleri, maksimum olabilirlik tahmini ile tahmin edilir. Conway-Maxwell *Poisson* regresyon modelinin olabilirlik fonksiyonu Eşitlik (3.16)'daki formülde verilmiştir (Sellers vd., 2011)

$$L(y_i|\lambda_i, v) = \exp\left(\sum_{i=1}^n y_i \log \lambda_i - v \sum_{i=1}^n \log(y_i!)\right) Z^{-n}(\lambda_i, v) \quad (3.16)$$

Böylece, Conway-Maxwell *Poisson* regresyon modelinin log- olabilirlik fonksiyonu Eşitlik (3.17)'de

$$\ln L(y_i|\lambda_i, v) = \sum_{i=1}^n (y_i \log \lambda_i - v \log(y_i!) - n \log Z(\lambda_i, v)) \quad (3.17)$$

olarak elde edilir. Yukarıdaki logaritmik olabilirlik fonksiyonundan Conway-Maxwell *Poisson* modelinin tahmini parametre değerleri, Eşitlik (3.17)'de β ve v 'ye göre türev alınıp sıfıra eşitlenerek elde edilebilir (Pattiz, 2009).

Conway-Maxwell *Poisson* regresyon modeli tahmin edildiğinde ise

$$\hat{y}_i|\mathbf{x}_i = \hat{\lambda}_i^{1/\hat{v}} - \frac{\hat{v} - 1}{2\hat{v}}$$

burada $\hat{\lambda}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$ elde edilir (Sellers ve Shmueli, 2010).

3.4. Model Seçim Kriterleri

En iyi modelin belirlenmesi için, maksimum olabilirliğe dayalı ölçülerden yararlanılması söz konusudur. Bu amacı gerçekleştirmek için ise AIC (*Akaike Information Criterion*) ve BIC (*Bayesian Information Criterion*) değerleri karşılaştırılır. AIC ve BIC'e dayalı en iyi model belirleme, analiz edilen modelin sırasına bağlı olmadığından, en iyi modeli seçme konusunda daha yüksek bir potansiyele sahiptir (Yusuf ve Ugalahi, 2015).

Ayrıca, AIC ve BIC kriterleri, belirli bir veri kümesi için regresyon modelinin göreceli kalite ölçüleridir. AIC, verilere en iyi uyan modeli belirlemeyi amaçlarken, BIC en yüksek sonsal (posterior) olasılığı kullanarak en iyi tahmin modelini belirlemeyi amaçlamaktadır (Yusuf ve Ugalahi, 2015).

$$AIC = -2\ln(L(\hat{\Omega})) + 2p \quad (3.18)$$

$$BIC = -2\ln(L(\hat{\Omega})) + p \cdot \log(n) \quad (3.19)$$

Bu kriterlere göre, en düşük AIC ve BIC değerlerine sahip regresyon modeli, verileri modellemek için en iyi regresyon modeli olarak ifade edilir (Durmuş vd., 2020).

4. BULGULAR VE TARTIŞMA

Bu çalışmada, öncelikle Poisson regresyonunda yaşanan aşırı yayılım ve az yayılım problemlerinin aşılmasında en iyi modeli bulabilmek amacıyla bir simülasyon gerçekleştirilecektir. Çalışmada, geliştirilmiş *Poisson* regresyonu ve Conway-Maxwell *Poisson* regresyonu olmak üzere iki regresyon modelinin karşılaştırılması söz konusudur.

Bu bölümde öncelikle aşırı yayılım problemi üzerinde geliştirilmiş *Poisson* regresyonu ile Conway-Maxwell *Poisson* regresyonu arasındaki en iyi modeli bulmada performans kriteri olarak AIC ve BIC değerleri kullanılacaktır. İkinci olarak, az yayılım problemi ile baş etmede aynı adımlar ve simülasyon senaryosu gerçekleştirilecektir.

4.1. Simülasyon Çalışması

Çalışmanın ikinci bölümünde yapılan literatür araştırması sonuçlarından, sayım verilerindeki aşırı yayılım ve az yayılım problemlerinin geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli kullanılarak çözülebileceği belirtilmişti. (Sellers ve Shmueli, 2010; Shmueli vd., 2005; Kadane vd., 2006). Bu nedenle, bu araştırma, Conway-Maxwell *Poisson* regresyon modeli ve geliştirilmiş *Poisson* regresyon modeli kullanılarak modellenen sayım verisi problemini ele alacaktır. Conway-Maxwell *Poisson* regresyon modeli ve geliştirilmiş *Poisson* regresyon modeli, sayım verilerinin sahip olduğu sorunlara göre ayarlanabilen bir yayılım değerine sahiptir, dolayısıyla aşırı yayılım veya az yayılım problemleri üzerinde çalışma yürütülecektir.

Bu çalışmada, bir bağımlı değişken ve bir bağımsız değişken ile türetilmiş veriler kullanılmıştır. Regresyon katsayıları $\beta_0 = \beta_1 = 1$ olarak alınmıştır. Bağımsız değişken $U[-1,1]$ dağılımından türetilmiştir. Yayılım parametresinin değeri eşit yayılım durumunu göstermek üzere $\nu = 1$ alınmış, aşırı ve az yayılım durumları için sırasıyla $\nu > 1$ ve $\nu < 1$ durumları için 3'er değer alınmıştır. Böylelikle Monte Carlo simülasyonunda türetilen bağımlı değişkende aşırı yayılım ve az yayılım sorunları oluşturulmuştur. Bu çalışmada, *Poisson* regresyon analizi, Conway-Maxwell *Poisson* regresyonu ve Geliştirilmiş *Poisson* regresyonu kullanarak verileri analiz etmek için R-studio ver.1.3.959 programında *COMpoissonReg*, *VGAM* paketleri kullanılmıştır. Analizler 1000 tekrarlı olacak şekilde yinelenmiştir.

4.1.1. Aşırı yayılım

Bu bölümde, aşırı yayılım sorununu aşmak için AIC ve BIC değerleri kullanılarak geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyonu arasında bir karşılaştırma yapılmaktadır. Aşırı yayılım probleminin üstesinden gelmede daha iyi modeli belirleme kriteri, hangi modelin en küçük AIC ve BIC değerlerine sahip olduğunu görmektir. Bu simülasyon çalışması 20, 50, 75, 100 ve 500 birimlik türetilmiş veriler üzerinde ve 0.3, 0.5 ve 0.75 yayılım koşullarında gerçekleştirilmiştir.

4.1.1.1. Yayılım parametresinin 0.3 olması durumu

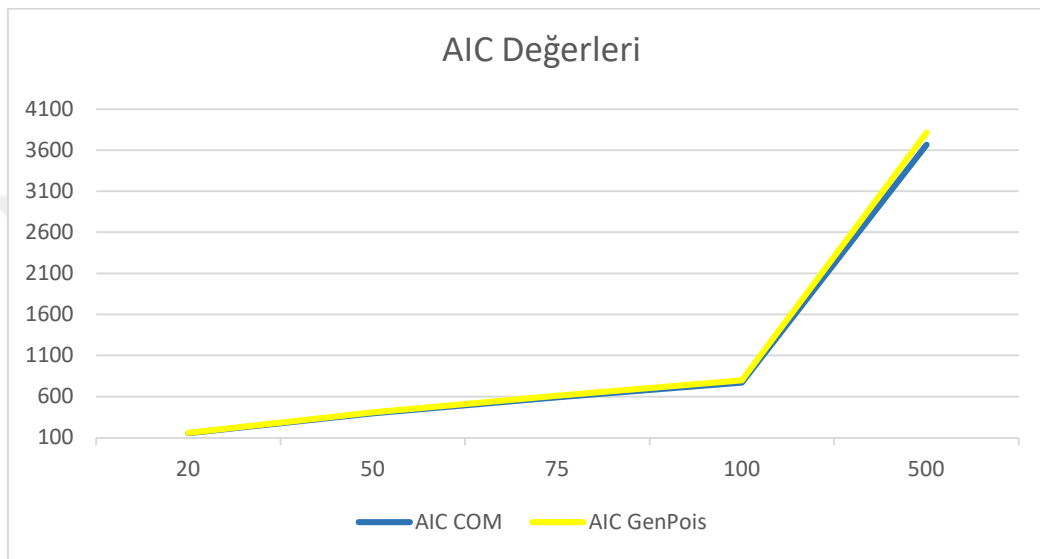
20, 50, 75, 100 ve 500'lük örneklem büyüklüklerindeki 0.3'lük yayılım değeri için simülasyon sonuçları Çizelge 4.1'de gösterilmektedir. Çizelge, geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC ve BIC değerlerini ve model parametrelerinin tahmin değerlerini göstermektedir.

Çizelge 4.1. Geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 0.3 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

	n	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	3.408513	3.245376	157.4084	160.3956
GenPois		3.39973	3.255816	157.4834	160.4706
COM	50	3.396243	3.259222	397.6291	403.3651
GenPois		3.389649	3.267166	398.1262	403.8622
COM	75	3.399185	3.254114	592.5642	599.5167
GenPois		3.393586	3.260946	593.5107	600.4632
COM	100	3.410884	3.239548	770.4131	778.2286
GenPois		3.404238	3.247596	772.0019	779.8174
COM	500	3.418906	3.228545	3668.06	3680.704
GenPois		3.412005	3.236902	3677.15	3689.793

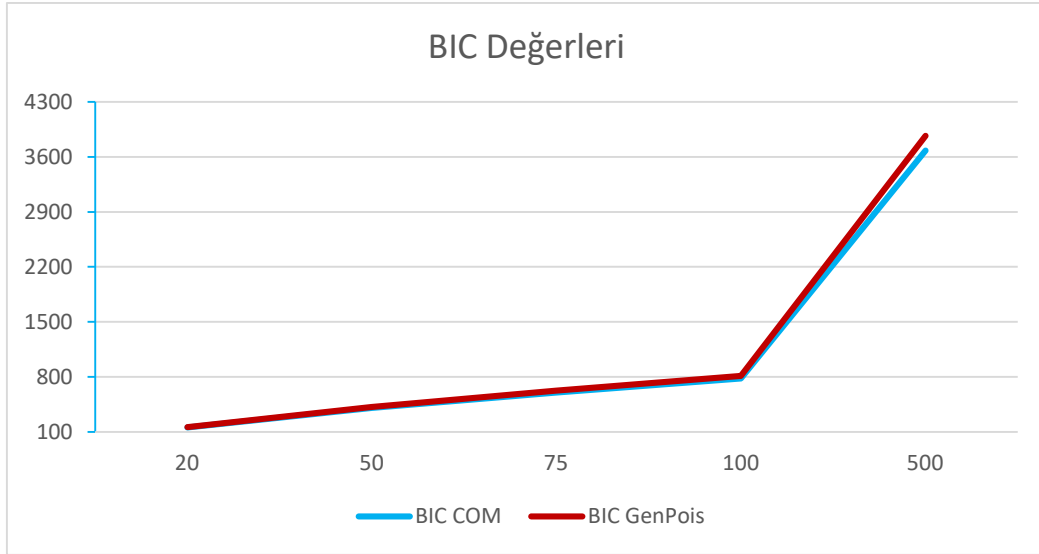
Çizelge 4.1, yayılım değeri 0.3 olması durumunda farklı büyüklüklerdeki örneklem hacimleri için türetilen simülasyon verilerinde, Conway-Maxwell *Poisson*

regresyon AIC değeri genelleştirilmiş *Poisson* regresyonunun AIC değerinden daha küçük olduğu görülmektedir. Dolayısıyla, Çizelge 4.1'e dayanarak, 0.3 yayılım değerinde aşırı yayılım sorununun olması durumunda AIC değerine dayalı en iyi modelin, Conway-Maxwell *Poisson* regresyon modelinin olduğu söylenebilir. Çizelgede sunulan AIC değerleri Şekil 4.1'de gösterilmektedir. Şekle göre örneklem hacmi 100'den sonra her iki modelin AIC değerlerinde sert bir kırılma gözlenmektedir.



Şekil 4.1 Yayılım değeri 0.3 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.1 BIC değerleri bakımından incelendiğinde, 0.3 yayılım değerinde her 20, 50, 75, 100 ve 500 birimlik örneklem hacmine sahip simülasyon verileri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değerinin genelleştirilmiş *Poisson* regresyon modelinin BIC değerinden daha küçük bir BIC değerine sahip olduğu görülmektedir. Hesaplanan BIC değerlerinin, 0.3 yayılım değerindeki AIC değerlerine benzer sonuçlar gösterdiği görülmektedir. Çizelge 4.1'de verilen BIC değerleri Şekil 4.2'deki gibi gösterilebilir. Şekle göre örneklem hacmi 100'den sonra her iki modelin BIC değerlerinde de AIC değerlerinde olduğu gibi sert bir kırılma gözlenmektedir.



Şekil 4.2. Yayılım değeri 0.3 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, 0.3 yayılım değerinde her örneklem hacmi için aşırı yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, $n=20$ 'de oluşturulan Conway-Maxwell *Poisson* modelinin tahmin denklemi olarak formüle edilir.

$$\hat{y} = \exp(3.408513 + 3.245376x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(3.245376) = 25.67136$ kat artıracakı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(3.408513) = 30.22027$ olacaktır. Bu yorum aynı zamanda 0.3 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

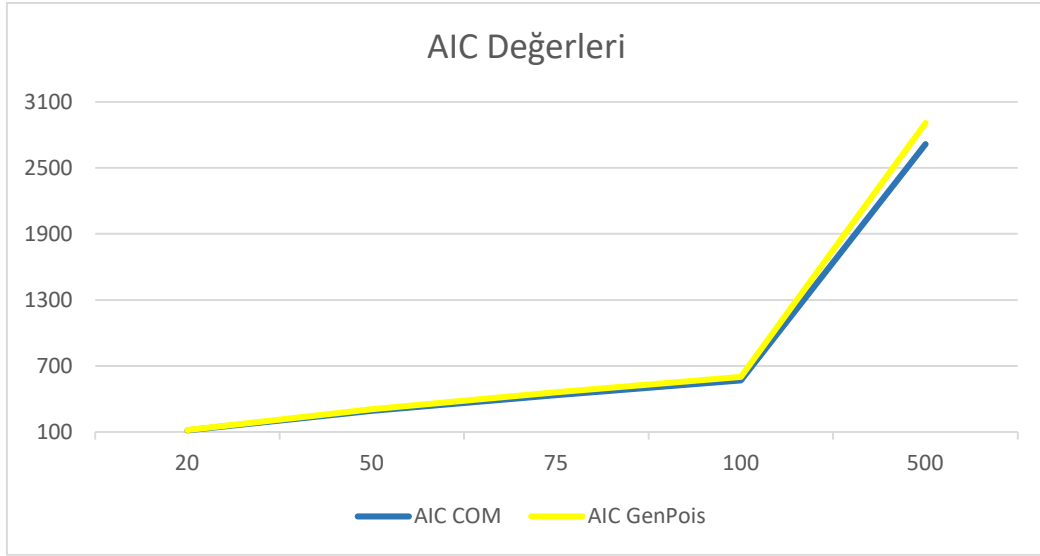
4.1.1.2. Yayılım parametresinin 0.5 olması durumu

20, 50, 75, 100 ve 500 birimden oluşan örneklem hacimlerinde yayılım parametresi değeri 0.5 için simülasyon sonuçları Çizelge 4.2'de gösterilmektedir. Diğer bir deyişle, Çizelge 4.2'de, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC, BIC değerleri ve model parametre tahmin değerleri gösterilmektedir.

Çizelge 4.2. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 0.5 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

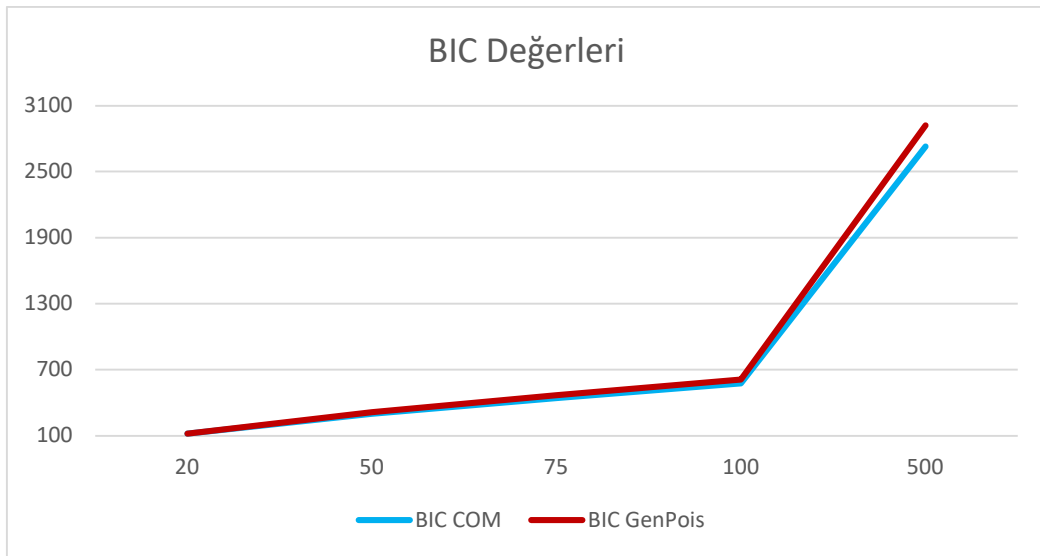
	n	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	2.091188	1.896015	116.042	119.0292
GenPois		2.090565	1.896184	116.264	119.2512
COM	50	2.090496	1.897032	291.8351	297.5712
GenPois		2.090093	1.896852	292.2993	298.0353
COM	75	2.092697	1.893866	434.7205	441.673
GenPois		2.093273	1.892139	435.4335	442.386
COM	100	2.102635	1.880801	566.9747	574.7902
GenPois		2.103326	1.87872	568.0021	575.8176
COM	500	2.106709	1.874348	2715.416	2728.06
GenPois		2.107149	1.872211	2720.963	2733.606

Çizelge 4.2'deki yayılım parametre değeri 0.5 olması durumunda, her örneklem hacminde türetilen simülasyon verileri için, Conway-Maxwell *Poisson* regresyon AIC değerlerinin genelleştirilmiş *Poisson* regresyonunun AIC değerlerinden daha küçük olduğu görülmektedir. Dolayısıyla, Çizelge 4.2'ye dayanarak, 0.5 yayılım değerinde aşırı yayılım sorununun olması durumunda AIC değerine göre, Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* modelinden daha iyi olduğu söylenebilir. Çizelge 4.2'de verilen AIC değerleri Şekil 4.3'te gösterilebilir.



Şekil 4.3. Yayılım değeri 0.5 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.2’de, yayılım değerinin 0.5 olduğu durumda 20, 50, 75, 100 ve 500 birimlik örneklem büyüklüklerinde oluşturulan simülasyon verileri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değerleri genelleştirilmiş *Poisson* regresyon modelinin BIC değerleri ile karşılaştırıldığında daha küçük BIC değerlerine sahip olduğu görülmektedir. Hesaplanan BIC değerlerinin, 0.5 yayılım değerindeki AIC değerlerine benzer sonuçlar gösterdiği görülmektedir. Çizelge 4.2’ye göre, BIC değerine dayalı olarak 0.5’lük bir yayılım değerinde aşırı yayılım sorununun olması durumunda, Conway-Maxwell *Poisson* regresyon modeli genelleştirilmiş *Poisson* modelinden daha iyi olduğu söylenebilir. Çizelge 4.2’de verilen BIC değerlerinin grafiği Şekil 4.4’te gösterilebilir.



Şekil 4.4. Yayılım değeri 0.5 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, her örneklem büyüklüğü için 0.5 yayılım değerinde aşırı yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, $n=20$ 'de oluşturulan Conway-Maxwell *Poisson* modelinin tahmin denklemi olarak formüle edilir

$$\hat{y} = \exp(2.091188 + 1.896015x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(1.896015) = 6.659304$ kat artıracakı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(2.091188) = 8.094526$ olacaktır. Bu yorum aynı zamanda 0.5 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

4.1.1.3. Yayılım parametresinin 0.75 olması durumu

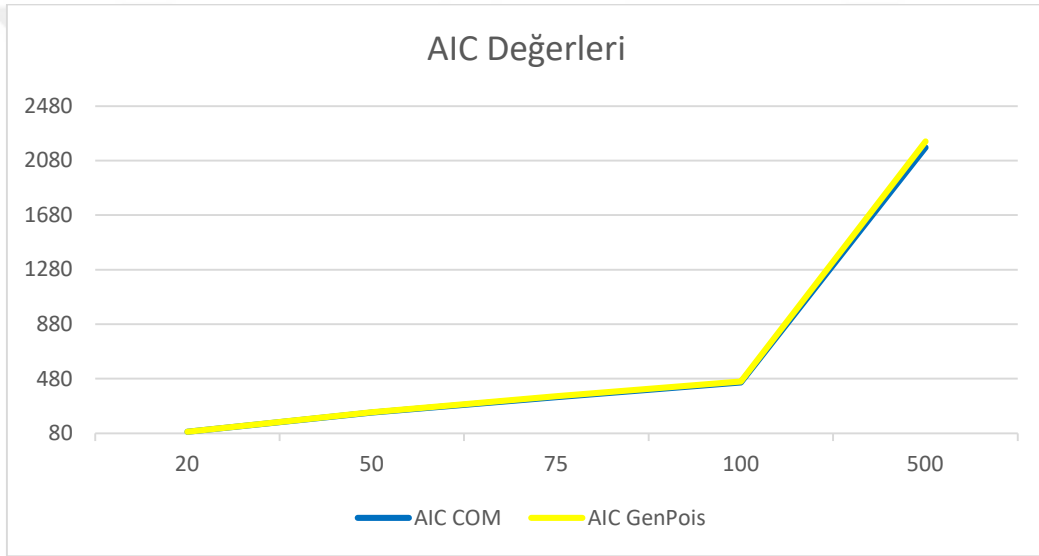
20, 50, 75, 100 ve 500 birimden oluşan örneklem büyüklüklerindeki 0.75'lik yayılım değeri için simülasyon sonuçları Çizelge 4.3'de gösterilmektedir. Çizelge, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC, BIC değerlerini ve model parametrelerinin tahmin değerlerini göstermektedir.

Çizelge 4.3. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 0.75 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

	n	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	1.370356	1.293262	92.76964	95.75684
GenPois		1.37133	1.291325	93.20919	96.19639
COM	50	1.381527	1.282829	232.0477	237.7838
GenPois		1.382392	1.281078	232.2555	237.9916
COM	75	1.383782	1.28186	345.7272	352.6796
GenPois		1.384969	1.279469	345.9693	352.9217
COM	100	1.389578	1.274782	452.0043	459.8198
GenPois		1.390714	1.272386	452.2519	460.0675

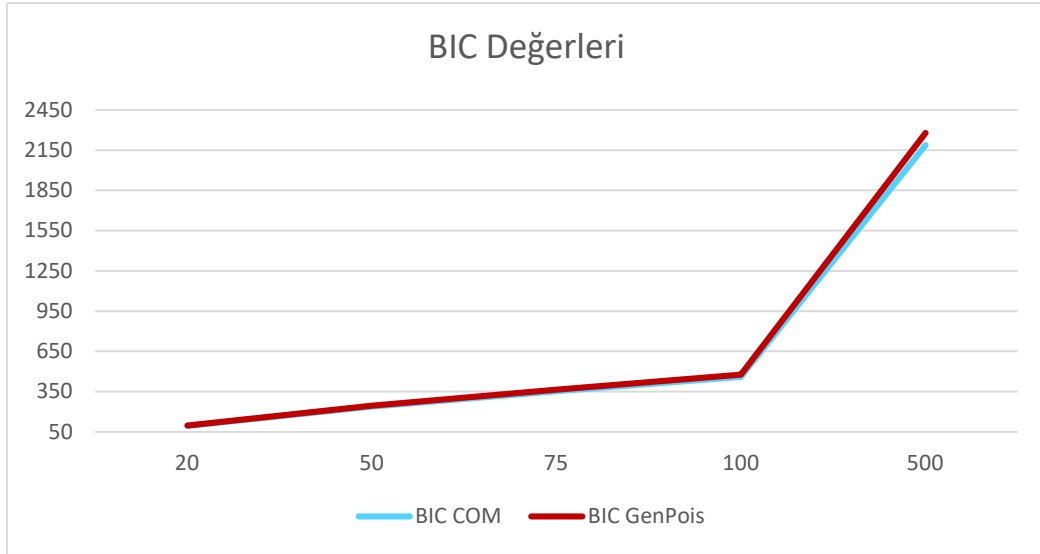
COM	500	1.392184	1.272798	2175.452	2188.096
GenPois		1.392845	1.271005	2176.52	2189.164

Çizelge 4.3, yayılım değeri 0.75 olduğunda her 20, 50, 75, 100 ve 500 birimlik türetilen örneklem verileri için, Conway-Maxwell *Poisson* regresyon AIC değerinin geliştirilmiş *Poisson* regresyonunun AIC değerinden daha küçük olduğunu göstermektedir. Dolayısıyla, Çizelge 4.3'e dayanarak, 0.75 yayılım değerinde aşırı yayılım sorununun olması durumunda, AIC değerine dayalı Conway-Maxwell *Poisson* regresyon modelinin geliştirilmiş *Poisson* modelden daha iyi olduğu söylenebilir. Çizelge 4.3'te sunulan AIC değerlerinin görsel sunumu Şekil 4.5'te gösterilebilir.



Şekil 4.5. Yayılım değeri 0.75 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.3'e göre, 0.75 yayılım değerindeki tüm örneklem büyüklükleri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değeri geliştirilmiş *Poisson* regresyon modelinin BIC değeri ile karşılaştırıldığında daha küçük BIC değerine sahip olduğu görülmektedir. Elde edilen sonuçların, 0.75 yayılım değerindeki AIC sonuçlarına benzer olduğu görülebilmektedir. Çizelge 4.3'e göre, BIC değerine dayalı olarak 0.75'lik bir yayılım değerinde aşırı yayılım sorununun olmasına dayalı olarak, Conway-Maxwell *Poisson* regresyon modelinin geliştirilmiş *Poisson* modelinden daha iyi olduğu söylenebilir. Çizelge 4.3'te verilen BIC değerlerinin grafiksel gösterimi Şekil 4.6'da verilmiştir.



Şekil 4.6. Yayılım değeri 0.75 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, her örneklem hacmi için 0.75 yayılım değerinde aşırı yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, $n=20$ 'de oluşturulan Conway-Maxwell *Poisson* modelinin tahmini olarak formüle edilir

$$\hat{y} = \exp(1.370356 + 1.293262x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(1.293262) = 3.644656$ kat artıracakı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(1.370356) = 3.936752$ olacaktır. Bu yorum aynı zamanda 0.75 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

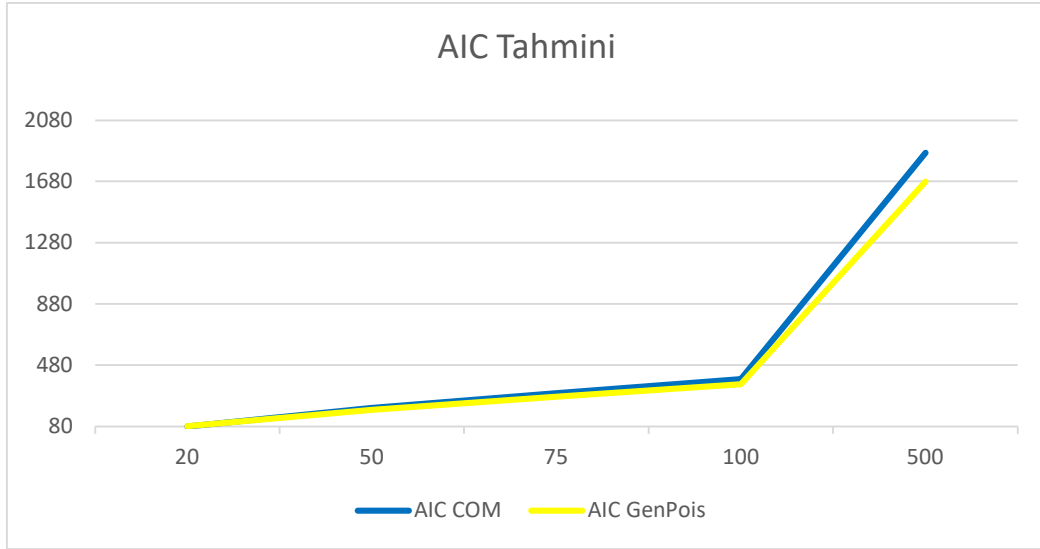
4.1.2 Eşit Yayılım

Çalışmada ayrıca eşit yayılım durumunda genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyonun parametre tahmin değerleri, AIC ve BIC değerlerinin davranışı gösterilmek istenmiştir. Yine 20, 50, 75, 100 ve 500 birimlik türetilmiş veriler üzerinde ve eşit yayılım koşullarında gerçekleştirilen simülasyon sonuçları Çizelge 4.4'te verilmiştir.

Çizelge 4.4. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için eşit yayılım durumundaki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

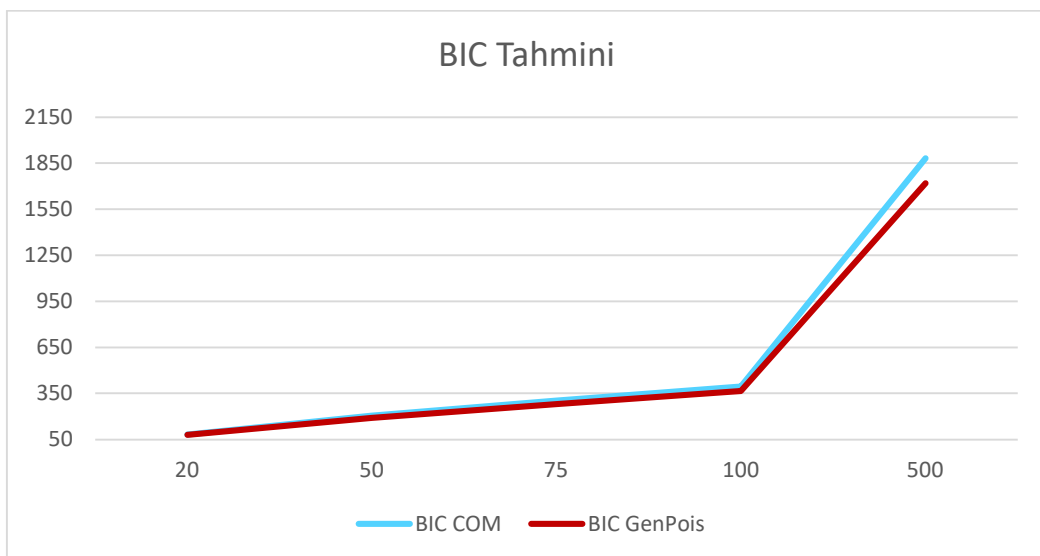
	n	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	0.9729861	1.019235	79.71162	82.69881
GenPois		0.9731877	1.018773	80.70241	83.68961
COM	50	0.9894884	1.004042	198.4989	204.2349
GenPois		0.9896409	1.00377	199.2263	204.9623
COM	75	0.9915833	1.005942	295.426	302.3789
GenPois		0.9917108	1.005706	296.1549	303.1074
COM	100	0.9955086	1.000143	387.0069	394.8224
GenPois		0.9955787	0.99998	387.6847	395.5002
COM	500	0.9986592	1.000066	1868.599	1881.243
GenPois		0.9986603	1.000063	1869.144	1881.788

Çizelge 4.4, yayılım değeri 1 olduğunda her 20, 50, 75, 100 ve 500 birimlik türetilen örneklem verileri için, Conway-Maxwell *Poisson* regresyon AIC değerinin genelleştirilmiş *Poisson* regresyonunun AIC değerinden daha küçük olduğunu göstermektedir. Dolayısıyla, Çizelge 4.4'e dayanarak, 1 yayılım değerinde (diğer bir deyişle, eşit yayılım durumunda) AIC değerine dayalı Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* modelden daha iyi olduğu söylenebilir. Çizelge 4.4'te sunulan AIC değerlerinin görsel sunumu Şekil 4.7'de gösterilmiştir.



Şekil 4.7. Yayılım değeri 1 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.4’de, yayılım değerinin 1 olduğu durumda 20, 50, 75, 100 ve 500 birimlik örneklem büyüklüklerinde oluşturulan simülasyon verileri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değerleri genelleştirilmiş *Poisson* regresyon modelinin BIC değerleri ile karşılaştırıldığında daha küçük BIC değerlerine sahip olduğu görülmektedir. Hesaplanan BIC değerlerinin, 1 yayılım değerindeki AIC değerlerine benzer sonuçlar gösterdiği görülmektedir. Çizelge 4.4’e göre, BIC değerine dayalı olarak eşit yayılım durumunda, Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* modelinden daha iyi olduğu söylenebilir. Çizelge 4.4’de verilen BIC değerlerinin grafiği Şekil 4.8’te gösterilmiştir.



Şekil 4.8. Yayılım değeri 1 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, her örneklem hacmi için 1 yayılım durumunda en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, $n=20$ 'de oluşturulan Conway-Maxwell *Poisson* modelinin tahmini olarak formüle edilir

$$\hat{y} = \exp(0.9729861 + 1.019235x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(0.9729861) = 2.645833$ kat artıracığı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(1.019235) = 2.771074$ olacaktır. Bu yorum aynı zamanda 1 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir. Bölüm 3'ten hareketle Conway-Maxwell *Poisson* modelindeki dağılım değeri 1'e eşit ise model *Poisson* modeli ile benzer şekilde yorumlanır.

4.1.3. Az Yayılım

Bu bölümde de aşırı yayılım durumdakine benzer biçimde az yayılım sorununun üstesinden gelmede geliştirilmiş *Poisson* regresyonu ve Conway-Maxwell *Poisson* regresyonu karşılaştırılması için performans kriteri olarak AIC ve BIC değerleri kullanılacaktır. Az yayılım probleminin üstesinden gelmede en iyi modeli belirleme kriterlerinden ikisi, hangi modelin en küçük AIC ve BIC değerlerine sahip olduğunu belirlemek olabilir. Az yayılım durumunda simülasyon çalışması 20, 50, 75, 100 ve 500 birimlik örneklem verileri üzerinde, ve 1.25, 1.5 ve 2 yayılım koşullarında gerçekleştirilmiştir.

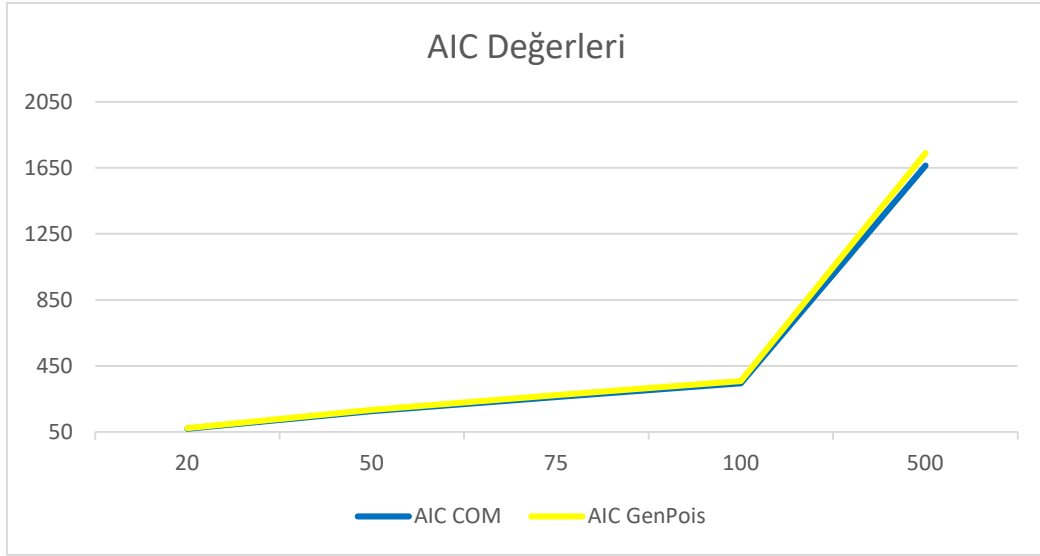
4.1.3.1 Yayılım parametresinin 1.25 olması durumu

20, 50, 75, 100 ve 500 birim olarak türetilen örneklem büyüklüklerindeki 1.25 yayılım değeri için simülasyon sonuçları Çizelge 4.5'te gösterilmektedir. Çizelge, geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC değerini ve model parametre tahmin değerlerini göstermektedir.

Çizelge 4.5. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 1.25 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

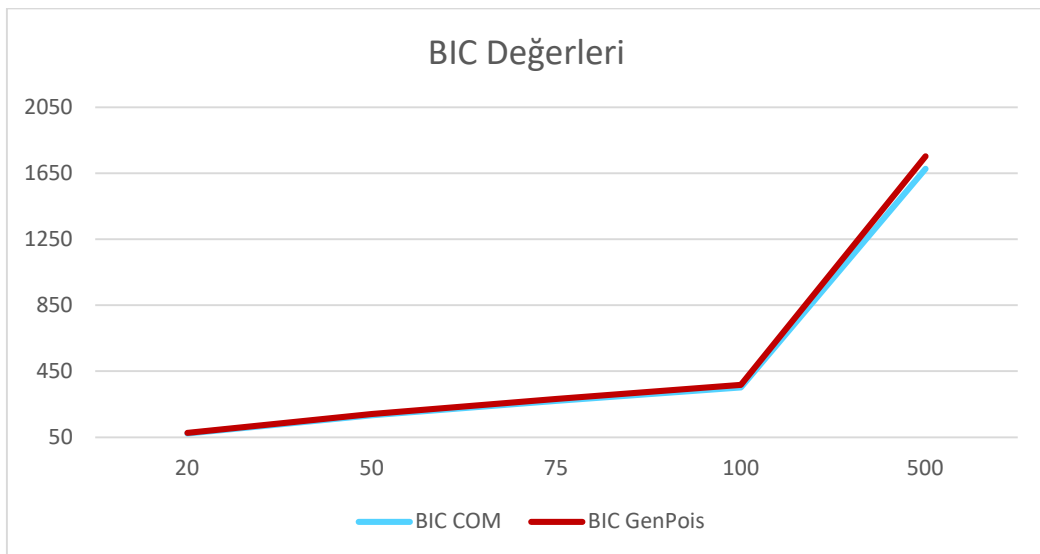
	n	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	0.711550	0.861771	71.03101	74.01821
GenPois		0.711099	0.862547	72.76898	75.75617
COM	50	0.732700	0.844122	176.3526	182.0887
GenPois		0.732337	0.844892	178.3583	184.0944
COM	75	0.733273	0.849204	262.2945	269.247
GenPois		0.733150	0.849503	264.6666	271.6191
COM	100	0.738112	0.843120	343.9901	351.8056
GenPois		0.737923	0.843550	346.7368	354.5523
COM	500	0.740613	0.844603	1664.061	1676.705
GenPois		0.740504	0.844901	1671.955	1684.599

Çizelge 4.5, 1.25 yayılım değerinde farklı büyüklükteki örneklem verileri için, Conway-Maxwell *Poisson* regresyon AIC değeri genelleştirilmiş *Poisson* regresyonunun AIC değerinden daha küçük olduğu göstermektedir. Dolayısıyla, Çizelge 4.5'e göre, az yayılım sorununun olduğu ve 1.25 yayılım değerinde AIC değerine göre Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* modelden daha iyi olduğu söylenebilir. Çizelge 4.5'deki AIC değerlerinin grafiksel gösterimi Şekil 4.9'de verilmiştir.



Şekil 4.9. Yayılım değeri 1.25 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.5'te, yayılım parametresi değeri 1.25 olduğunda tüm örneklem büyüklükleri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değeri geliştirilmiş *Poisson* regresyon modeli BIC değeri ile karşılaştırıldığında daha küçük BIC değerlerine sahip olduğu görülmektedir. Elde edilen sonuçların, 1.25 yayılım değerindeki AIC değerinin sonuçlarına benzer olduğu söylenebilir. Çizelge 4.5'ye göre, BIC değerine dayalı olarak 1.25'lik bir yayılım değerinde az yayılım sorununun olmasına dayalı olarak, Conway-Maxwell *Poisson* regresyon modelinin geliştirilmiş *Poisson* modelinden daha iyi olduğu belirtilebilir. Çizelge 4.5'deki BIC değerlerinin grafiksel gösterimi Şekil 4.10'de verilmiştir.



Şekil 4.10. Yayılım değeri 1.25 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, tüm örneklem büyüklükleri için 1.25 yayılım değerinde az yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, $n=20$ 'de oluşturulan Conway-Maxwell *Poisson* modeli tahmini olarak formüle edilir.

$$\hat{y} = \exp(0.7115503 + 0.8617706x_1)$$

Bu model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(0.8617706) = 2.367349$ kat artıracakı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(0.7115503) = 2.037147$ olacaktır. Bu yorum aynı zamanda 1.25 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

4.1.3.2. Yayılım parametresinin 1.5 olması durumu

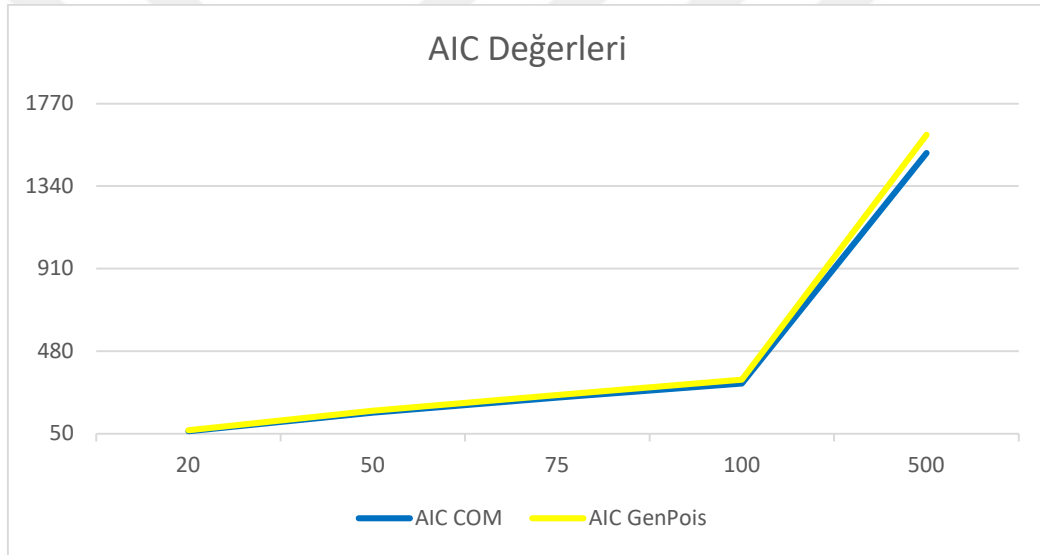
20, 50, 75, 100 ve 500 birimlik örneklem büyüklüklerinde yayılım parametresi değerinin 1.5 olması durumu için simülasyon sonuçları Çizelge 4.6'de gösterilmektedir. Bu çizelgede, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC, BIC değerlerini ve model parametre tahmin değerlerini görülmektedir.

Çizelge 4.6. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 1.5 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

	N	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	0.520279	0.763236	64.65549	67.64268
GenPois		0.519526	0.764767	67.21198	70.19918
COM	50	0.544698	0.7434996	159.7708	165.5069
GenPois		0.544039	0.744897	163.5676	169.3037
COM	75	0.547480	0.745546	237.9363	244.8888
GenPois		0.547072	0.746496	242.8107	249.7631
COM	100	0.550857	0.742695	312.3568	320.1723
GenPois		0.550409	0.743782	318.2297	326.0453

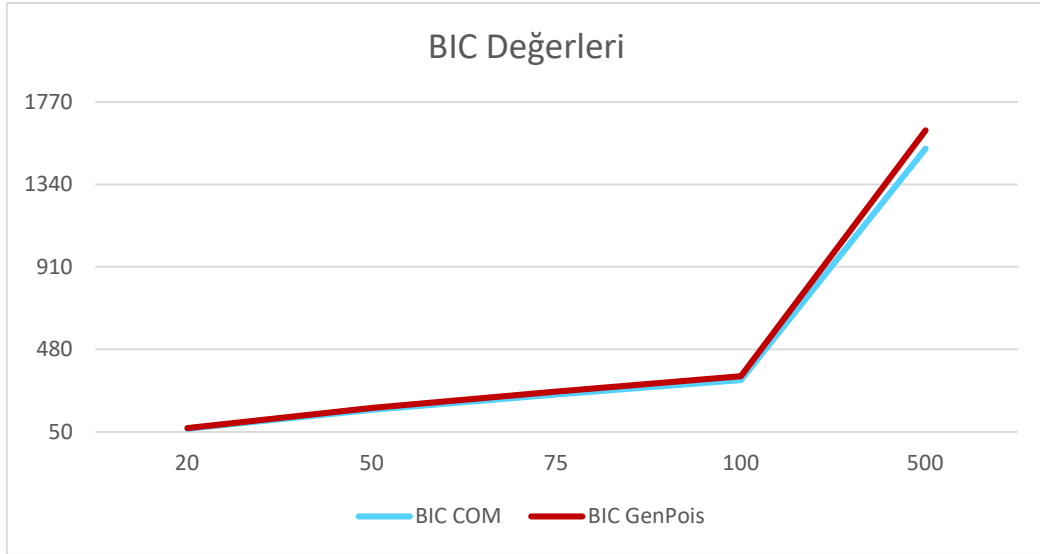
COM	500	0.553888	0.743857	1514.019	1526.663
GenPois		0.553630	0.744644	1535.74	1548.384

Çizelge 4.6, yayılım değerinin 1.5 olduğu durumda farklı büyüklüklerdeki örneklem hacimleri için, Conway-Maxwell *Poisson* regresyon AIC değerinin geliştirilmiş *Poisson* regresyonunun AIC değerinden daha küçük olduğunu göstermektedir. Dolayısıyla, Çizelge 4.6'e göre, 1.5 yayılım değerinde az yayılım sorununun olması durumunda AIC değerine dayalı daha iyi modelin, Conway-Maxwell *Poisson* regresyon modeli olduğu söylenebilir. Çizelge 4.6'deki AIC değerlerinin grafiksel gösterimi Şekil 4.11'da verilmiştir.



Şekil 4.11. Yayılım değeri 1.5 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.6, 1.5 yayılım değerinde ele alınan tüm örneklem hacimlerindeki türetilen veriler için, Conway-Maxwell *Poisson* regresyon modelinin BIC değeri geliştirilmiş *Poisson* regresyon modelinin BIC değeri ile karşılaştırıldığında daha küçük bir BIC değerine sahip olduğunu göstermektedir. Elde edilen sonuçların, 1.5 yayılım değerindeki AIC değerinin sonuçlarına benzer olduğu görülebilmektedir. Çizelge 4.6'e göre, BIC değerine dayalı olarak 1.5 yayılım değerinde az yayılım sorununun olması durumunda, Conway-Maxwell *Poisson* regresyon modelinin geliştirilmiş *Poisson* modelinden daha iyi olduğu söylenebilir. Bu çizelgedeki BIC değerlerinin görsel sunumu Şekil 4.12'da gösterilebilir.



Şekil 4.12. Yayılım değeri 1.5 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, her örneklem büyüklükleri için 1.5 yayılım değerinde az yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, n=20'de oluşturulan Conway-Maxwell *Poisson* modeli tahmini olarak formüle edilir

$$\hat{y} = \exp(0.5202791 + 0.763236x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(0.763236) = 2.145207$ kat artıracığı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(0.5202791) = 1.682497$ olacaktır. Bu yorum aynı zamanda 1.5 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

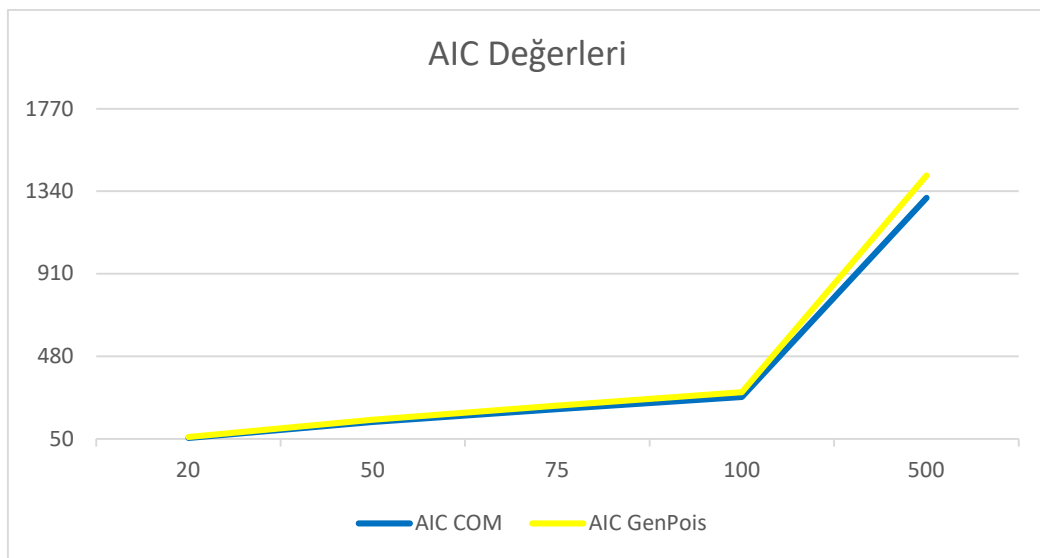
4.1.3.3. Yayılım parametresinin 2 olması durumu

20, 50, 75, 100 ve 500 birimlik örneklem büyüklüklerinde yayılım değerinin 2 olması durumu için simülasyon sonuçları Çizelge 4.7'de gösterilmektedir. Bu çizelge, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için AIC, BIC değerlerini ve model parametre tahmin değerlerini göstermektedir.

Çizelge 4.7. Genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli için 2 yayılım değerindeki AIC, BIC değerleri ve model parametrelerinin tahmin değerleri.

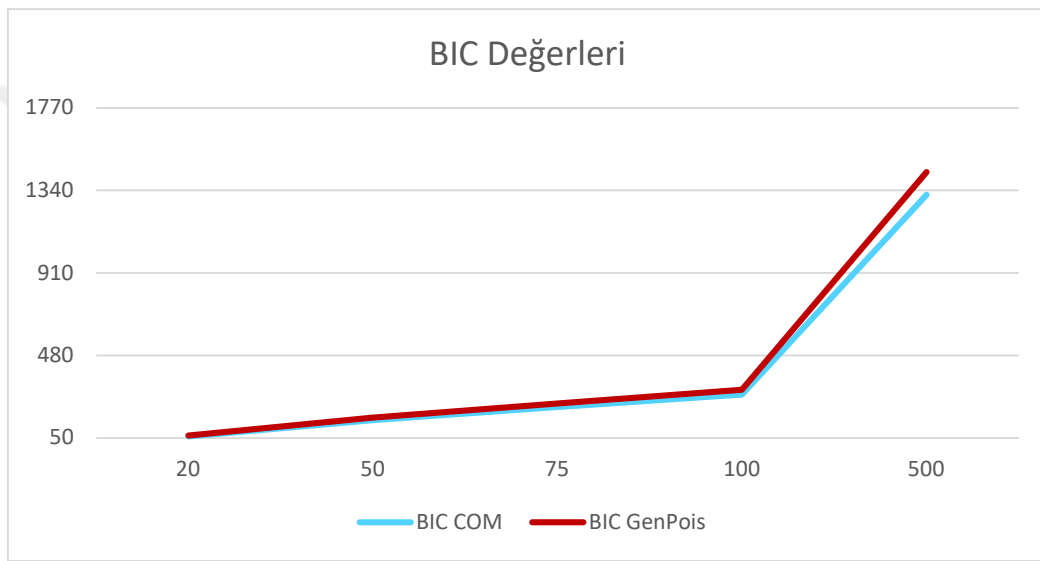
	N	Sabit terim ($\hat{\beta}_0$)	Eğim ($\hat{\beta}_1$)	AIC	BIC
COM	20	0.261128	0.637565	55.77019	58.75739
GenPois		0.259823	0.640556	60.0378	63.025
COM	50	0.287566	0.616132	137.2529	142.989
GenPois		0.286471	0.618627	144.7823	150.5184
COM	75	0.292334	0.617213	204.0332	210.9857
GenPois		0.291495	0.619251	214.5064	221.4589
COM	100	0.293970	0.613558	268.371	276.1865
GenPois		0.293140	0.61570	281.2919	289.1074
COM	500	0.298362	0.614133	1305.478	1318.122
GenPois		0.297896	0.615746	1359.505	1372.149

Çizelge 4.7, 2 yayılım değerinde tüm örneklem hacimleri için, Conway-Maxwell *Poisson* regresyon AIC değerinin genelleştirilmiş *Poisson* regresyonunun AIC değeri ile karşılaştırıldığında daha küçük olduğu görülmektedir. Dolayısıyla, Çizelge 4.7'e göre, 2 yayılım değerinde az yayılım sorununun varlığında AIC değerine dayalı daha iyi modelin, Conway-Maxwell *Poisson* regresyon modeli olduğu söylenebilir. Bu çizelgedeki AIC değerlerinin grafiksel gösterimi Şekil 4.13'te verilmiştir.



Şekil 4.13. Yayılım değeri 2 olması durumundaki AIC değerlerinin grafiği.

Çizelge 4.7'ye göre, 2 yayılım değerinde tüm örneklem büyüklükleri için, Conway-Maxwell *Poisson* regresyon modelinin BIC değeri genelleştirilmiş *Poisson* regresyon modelinin BIC değerinden daha küçük bir BIC değerine sahip olduğu görülebilmektedir. Elde edilen sonuçların, 2 yayılım değerindeki AIC değerinin sonuçlarına benzer olduğu söylenebilir. Çizelge 4.7'ye göre, BIC değerine dayalı olarak 2'lik bir yayılım değerinde az yayılım sorununun varlığında, Conway-Maxwell *Poisson* regresyon modeli genelleştirilmiş *Poisson* modelinden daha iyidir. Bu çizelgedeki BIC değerlerinin grafiksel gösterimi Şekil 4.14'de verilmiştir.



Şekil 4.14. Yayılım değeri 2 olması durumundaki BIC değerlerinin grafiği.

AIC ve BIC değerlerinin sonuçlarına dayanarak, her örneklem büyüklükleri için 2 yayılım değerinde az yayılım sorununun üstesinden gelmede en iyi modelin Conway-Maxwell *Poisson* regresyon modeli olduğu sonucuna varılabilir. Örneğin, n=20'de oluşturulan Conway-Maxwell *Poisson* modeli tahmini olarak formüle edilir

$$\hat{y} = \exp(0.2611282 + 0.6375647x_1)$$

Yukarıdaki model tahminine dayanarak, diğer tüm değişkenler sabit kalmak koşuluyla, x_1 değişkenindeki yüzde 1'lik artışta, ortalama y değişkenini $\exp(0.6375647) = 1.891868$ kat artıracığı söylenebilir. x_1 değişkeni 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(0.2611282) = 1.298394$ olacaktır. Bu yorum aynı zamanda 2 yayılım değerine sahip 50, 75, 100 ve 500 birimlik örneklem büyüklükleri için Conway-Maxwell *Poisson* modelinin yorumlanmasında da kullanılabilir.

Sonuç olarak çalışmaya alınan tüm örneklem büyüklükleri ve farklı yayılım değerlerinde, birbirlerine alternatif olan her iki modelin AIC ve BIC değerlerine bakıldığında, Conway-Maxwell *Poisson* regresyon modelinin model seçim kriterleri bakımından genelleştirilmiş *Poisson* regresyon modelinden daha küçük değerler verdiği görülmüştür. İzleyen satırlarda aynı sonucun gerçek bir veri setinde gerçekleşip gerçekleşmeyeceği incelenecektir.

4.2. Gerçek Veri Seti

Tez çalışmasında ayrıca gerçek veri seti üzerinde ele alınan regresyon modellerinin bir uygulamasının yapılması planlanmıştır. Yapılan uygulamada gerçekleştirilecek işlem adımları simülasyon çalışmasıyla benzer olacaktır. Gerçek verilerin incelendiği çalışmadaki verilerin öncelikle yayılımının belirlenmesi, çoklu doğrusallık varsayımının olup olmadığı ve bağımlı değişkenin Poisson dağılımına sahip olup olmadığı gibi varsayımların test edilmesi gerekmektedir. Böylelikle gerçek verilerde elde edilen sonuçların simülasyon sonuçlarını destekleyip desteklemediğinin görülme imkânı olacaktır. Kullanılan veri kümesi, Hendirawati'nin (2017) araştırmasından elde edilmiş ve Çizelge 4.8'de verilmiştir.

Çizelge 4.8. 2015 yılında Orta Java eyaletindeki HIV hasta sayısına ve onu etkileyen değişkenlere ilişkin veriler.

Orta Java Şehirleri	y	x_1	x_2	x_3	x_4
Cilacap	60	2.35	2.83	5.28	7.055
Banyumas	61	2.41	3.79	9.06	7.03
Purbalingga	9	1.4	3.01	6.28	7.135
Banjarnegara	41	2.06	4.21	4.32	7.16
Kebumen	61	2.22	3.88	2.39	6.78
Purworejo	16	2.42	4.79	5.06	6.415
Wonosobo	45	0.95	3.6	10.94	7.285
Magelang	14	1.67	2.65	1.34	6.665
Genuk	59	0.78	4.05	2.25	6.94
Boyolali	44	2.05	3.97	2.49	7.13
Klaten	40	1.22	2.2	10.18	7.95
Sukoharjo	7	1.79	4.43	14.63	5.82
Wonogiri	20	2	3.39	8.47	7.515
Karanganyar	62	1.81	3.98	5.77	7.025
Sragen	65	0.95	2.74	2.5	7.5
Grobogan	29	1.5	3.76	8.14	7.16
Blora	28	0.69	2.91	2.38	8.08

Rembang	47	1.27	3.08	1.23	7.53
Pati	31	1.76	3.49	6.82	8.21
Kudus	82	1.98	2.44	5.64	7.835
Jepara	61	1.72	2.68	2.41	8.025
Demak	45	0.82	3.1	11.49	7.985
Semarang	32	2.26	3.75	7.96	7.005
Temanggung	38	0.35	3.61	6.99	7.095
Kendal	99	3.11	3.09	3.63	7.685
Batang	31	1.25	3.32	4.56	7.71
Pekalongan	72	1.89	2.25	4.95	7.24
Pamulang	52	0.68	2.53	2.09	7.03
Tegal	37	1.51	2.75	1.68	7.49
Brebes	7	5.97	9.94	100	7.195
Kota Magelang	29	7.41	5.86	100	8.135
Kota Surakarta	18	5.88	7.07	100	8.665
Kota Salatiga	116	5.5	3.76	100	9.09
Kota Semarang	6	4.08	7.42	100	8.335
Kota Pekalongan	3	2.14	4.88	100	8.47
Kota Tegal	60	2.35	2.83	5.28	7.065

Ele alınan değişkenlerin açıklamaları aşağıda ifade edilmiştir:

1. Orta Java eyaletindeki yeni HIV hastalarının sayısı (y -*HIV*)
2. Orta Java eyaletindeki prezervatif kullanan çocuk doğurma yaşındaki çiftlerin oranı (x_1 – *PREZ*)
3. Orta Java eyaletindeki her şehirde 100.000 nüfus başına düşen sağlık çalışanı oranı (x_2 – *SCALIS*)
4. Orta Java eyaletindeki her şehirde kalabalık alanların oranı (x_3 – *ALAN*)
5. Orta Java eyaletindeki her şehirde 25-34 yaş arası nüfusun oranı (x_4 – *NUFUS*)

Veriler *Poisson* regresyon modeli, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli kullanılarak analiz edilmiştir.

4.2.1. *Poisson* regresyon modeli tahmin sonuçları

Orta Java eyaletindeki HIV hastalarına ilişkin verileri kullanılarak elde edilen *Poisson* regresyon analizinin sonuçları Çizelge 4.9'de gösterilmektedir. Çizelgede, *Poisson* regresyon modeli için tahmin edilen parametre tahmin değerleri, standart hataları ve bağımsız değişkenlerin anlamlılık düzeyi gösterilmektedir.

Çizelge 4.9. *Poisson* regresyon modeli için parametre tahmin değerleri.

Değişken	Parametre tahminleri	Standart Hata	<i>p</i> -değeri
Sabit terim	2.620191	0.545199	$1.54e^{-06}$
x_1 (PREZ)	0.318551	0.033652	$2e^{-16}$
x_2 (SCALIS)	-0.350092	0.038948	$2e^{-16}$
x_3 (ALAN)	-0.010447	0.002203	$2.11e^{-06}$
x_4 (NUFUS)	0.249904	0.063898	$9.19e^{-05}$

Çizelge 4.9'e göre, her bir bağımsız değişkeninin bağımlı değişken üzerindeki etkisi 0.05 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.

Her bir bağımsız değişkenin kısmi hipotez testi gerçekleştirildikten sonra eş zamanlı etkiler testi yapılmıştır. Eş zamanlı etkiler testi, tüm bağımsız değişkenlerin birlikte Orta Java eyaletindeki HIV hasta sayısı üzerindeki etkisini belirlemeyi amaçlamaktadır ve testin sonuçları Çizelge 4.10'te gösterilmiştir.

Çizelge 4.10. Olabilirlik Oranı Test değeri

	LogLik	χ^2	<i>p</i> -değeri
Bağımsız değişkensiz model	-388.07		
Bağımsız değişkenli modeller	-245.2	285.73	$< 2.2e^{-16}$

Çizelge 4.10, bağımsız değişkenlerin birlikte Orta Java eyaletindeki HIV hastalarının sayısındaki değişimi etkileyebileceğini göstermektedir. Oluşturulan *Poisson* regresyon modelinin tahmini aşağıdaki gibi ifade edilir.

$$\hat{y} = \exp(2.620191 + 0.318551x_1 - 0.350092x_2 - 0.010447x_3 + 0.249904x_4)$$

Yukarıdaki model tahminine dayanarak, diğer değişkenler sabit kalmak koşuluyla, ortalama y değişkenini x_1 değişkenindeki yüzde 1'lik artışla, $\exp(0.318551) = 1.375134$ kat artıracığı söylenebilir. Bu yorum aynı zamanda x_2 , x_3 ve x_4 değişkenleri için *Poisson* modelinin yorumlanmasında benzer şekilde yapılmaktadır. x_1 , x_2 , x_3 ve x_4 değişkenleri 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(2.620191) = 13.73835$ olacaktır.

Oluşturulan *Poisson* regresyon modelinin çoklu doğrusallık testi, *Poisson* dağılım testi ve yayılım testi olmak üzere çeşitli varsayımları da karşılaması gerekmektedir. Bu varsayımlara ilişkin sonuçlar izleyen kısımlarda verilmiştir.

4.2.2. Çoklu doğrusallık testi sonuçları

Bu test, Orta Java eyaletindeki HIV hastalarına ilişkin verileri modelleyen *Poisson* regresyon modelinin bağımsız değişkenlerinin birbirleriyle korelasyonunun olmadığını ortaya çıkarmayı amaçlamaktadır. Orta Java eyaletindeki HIV hastalarına ilişkin verileri kullanan *Poisson* regresyon modelinin çoklu doğrusallık testi analizinin sonuçları Çizelge 4.11'de gösterilmektedir.

Çizelge 4.11. Varyans şişme faktörü (VIF) değerleri.

Bağımsız Değişkenleri	VIF
x_1 (PREZ)	3.8983
x_2 (SCALIS)	2.5328
x_3 (ALAN)	6.8371
x_4 (NUFUS)	2.3177

Çizelge 4.11, her bir bağımsız değişken için VIF değerinin 10'dan küçük olduğunu göstermektedir. Dolayısıyla, *Poisson* regresyon modelinde yer alan bağımsız değişkenler arasında çoklu doğrusallık durumunun olmadığı sonucuna varılabilir.

4.2.3. *Poisson* dağılımı testi sonuçları

Poisson regresyon modelinin varsayımlarına göre, yeni HIV hastalarının sayısı değişkeni (y) bir *Poisson* dağılımına sahip olmalıdır. *Poisson* dağılım testi analizinin sonuçları Çizelge 4.12'de gösterilmiştir.

Çizelge 4.12. *Poisson* dağılımı için uyum iyiliği (Goodness of fit) değeri

Serbestlik derecesi	p -değeri
27	$7.855726e^{-98}$

Çizelge 4.12'ye göre, p -değeri'nin $\alpha = 0,05$ 'ten küçük olduğu, dolayısıyla, yeni HIV hasta sayısının bir *Poisson* dağılımına sahip olduğu sonucuna varılabilir. Bu nedenle *Poisson* regresyon modelindeki dağılım varsayımının karşılandığı söylenebilir.

4.2.4. Yayılım testi sonuçları

Aşırı yayılım veya az yayılım gibi eşit yayılım varsayımının ihlallerinin meydana gelip gelmediğini belirlemeyi amaçlayan yayılım testi analizinin sonucu 8.4857 olarak bulunmuştur.

1'den büyük olan 8.485711'in yayılım değeri eşit yayılım varsayımının ihlal edildiği, dolayısıyla aşırı yayılım durumunu gösterdiği ifade edilebilir. Diğer bir deyişle, *Poisson* regresyon modelinin, eşit yayılım varsayımını karşılamadığı için Orta Java eyaletindeki HIV hastalarına ilişkin verileri modellemek için uygun olmadığını göstermektedir. *Poisson* regresyon modelinde aşırı yayılımın üstesinden gelmek için genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli sonuçlarına izleyen kısımlarda yer verilmiştir.

4.2.5. Genelleştirilmiş *poisson* regresyon modeli tahmin sonuçları

Poisson regresyon modelinde aşırı yayılımın üstesinden gelmek için kullanılan genelleştirilmiş *Poisson* regresyon modelinin analiz sonuçları Çizelge 4.12'de gösterilmektedir. Çizelgede, genelleştirilmiş *Poisson* regresyon modeli için parametre tahmin değerlerini ve bağımsız değişkenlerin anlamlılık düzeyini göstermektedir.

Çizelge 4.13. Genelleştirilmiş *Poisson* regresyon modeli için parametre tahmin değerleri.

Değişken	Parametre tahminleri	Standart Hata	p -değeri
Sabit terim	0.842737	0.121456	$3.96e^{-12}$
x_1 (PREZ)	0.347232	0.096701	0.00033
x_2 (SCALIS)	-0.245372	0.108795	0.02411
x_3 (ALAN)	-0.016157	0.006438	0.01209
x_4 (NUFUS)	0.370543	0.184823	0.04498

Çizelge 4.13'ye göre, her bir bağımsız değişkeninin p -değeri $\alpha = 0.05$ anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Bu nedenle, oluşturulan genelleştirilmiş *Poisson* regresyon modelinin tahmini aşağıdaki gibi ifade edilir.

$$\hat{y} = \exp(0.842737 + 0.347232x_1 - 0.245372x_2 - 0.016157x_3 + 0.370543x_4)$$

Yukarıdaki model tahminine dayanarak, diğer değişkenler sabit kalmak koşuluyla, ortalama y değişkenini x_1 değişkenindeki yüzde 1'lik artışta, $\exp(0.347232) = 1.415145$ kat artıracığı söylenebilir. Aynı zamanda x_2 , x_3 ve x_4 değişkenleri için *Poisson* modelinin yorumlanması da benzer şekilde yapılmaktadır. x_1 , x_2 , x_3 ve x_4 değişkenleri 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(0.842737) = 2.322716$ olacaktır.

4.2.6. Conway-Maxwell *poisson* regresyon modeli tahmin sonuçları

Poisson regresyon modelinde aşırı yayılımın üstesinden gelmek için Conway-Maxwell *Poisson* regresyon modeli analizinin sonuçları Çizelge 4.13'te gösterilmektedir. Çizelgede, Conway-Maxwell *Poisson* regresyon modeli için parametre tahmin değerlerini, standart hatalarını ve p değerini göstermektedir.

Çizelge 4.14. Conway-Maxwell *Poisson* regresyon modeli için parametre tahmin değerleri.

Değişken	Parametre tahminleri	Standart Hata	p -değeri
Sabit terim	2.491954	1.598307	0.118968
x_1 (PREZ)	0.329898	0.097028	0.000674
x_2 (SCALIS)	-0.340482	0.110012	0.001968
x_3 (ALAN)	-0.011863	0.006422	0.064708
x_4 (NUFUS)	0.262369	0.187873	0.162556

Çizelge 4.14'e göre, x_1 ve x_2 değişkenlerinin etkisi $\alpha = 0,05$ anlamlılık düzeyinde istatistiksel olarak anlamlıdır. Bu nedenle, oluşturulan Conway-Maxwell *Poisson* regresyonu modelinin tahmini aşağıdaki gibi ifade edilir.

$$\hat{y}_i = \exp(2.491954 + 0.329898x_1 - 0.340482x_2 - 0.011863x_3 + 0.262369x_4)$$

Yukarıdaki model tahminine dayanarak, diğer değişkenler sabit kalmak koşuluyla, ortalama y değişkenini x_1 değişkenindeki yüzde 1'lik artışta, $\exp(0.329898) = 1.390826$ kat artıracığı söylenebilir. Aynı zamanda x_2 , x_3 ve x_4 değişkenleri için *Poisson* modelinin yorumlanması da benzer şekilde yapılmaktadır. x_1 , x_2 , x_3 ve x_4 değişkenleri 0 değerini aldığı anda ise, ortalama y değişkeninin değeri $\exp(2.491954) = 12.08487$ olacaktır.

4.2.6. Uygun Modelin Belirlenmesi

Bu bölümde, *Poisson* regresyon modelinin yaşadığı aşırı yayılım sorununun geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* modeli ile çözüldüğünü ve en iyi modelin seçilmesini bulmak için bir analiz yapılmıştır. Bu amaçla AIC ve BIC değerleri kullanılmaktadır. Her model için AIC ve BIC değerlerinin analiz sonuçları Çizelge 4.15'te gösterilebilir.

Çizelge 4.15. AIC ve BIC değerleri.

Model	AIC	BIC
<i>Poisson</i> regresyon modeli	500.41	508.18
Genelleştirilmiş <i>Poisson</i> regresyon modeli	309.76	319.09
Conway-Maxwell <i>Poisson</i> regresyon modeli	309.13	318.46

Çizelge 4.15'te, *Poisson* regresyon modelinin yaşadığı aşırı yayılım sorununun geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* modeli ile çözüldüğünü göstermektedir. Bu sonuç, geliştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* modelinden daha büyük olan *Poisson* regresyon modelinin AIC ve BIC değerleri karşılaştırılarak görülebilir. Çizelge 4.15 ayrıca, Conway-Maxwell *Poisson* regresyon modelinin, *Poisson* regresyon modelinin Orta Java eyaletindeki HIV hastalarına ilişkin verilerde yaşadığı aşırı yayılma sorununun üstesinden gelmede daha iyi model olduğunu göstermektedir. Bu sonuç, geliştirilmiş *Poisson* regresyon modelinden daha küçük olan Conway-Maxwell *Poisson* regresyon modelinin AIC ve BIC değerlerinin karşılaştırılmasıyla elde edilmiştir.

5. SONUÇ VE ÖNERİLER

Poisson regresyon modelinde var olan aşırı yayılım ve az yayılım sorunlarının üstesinden gelmek için çeşitli regresyon modelleri geliştirilmiştir. Bu çalışmada ise *Poisson* regresyon modelinde olan aşırı yayılım ve az yayılım sorunu için, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modelleri kullanılmıştır, çünkü her iki model de aşırı yayılım ve az yayılım sorununun üstesinden gelmede çok esnektir. Bu nedenle çalışmada, genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli kullanılarak R Studio programında gerçekleştirilen simülasyon analizinin sonuçları verilmiştir. Aşırı yayılım ve az yayılım problemlerine göre belirlenen altı farklı yayılım değeri üzerinde gerçekleştirilen 20, 50, 75, 100 ve 500 birimlik simülasyon verileri kullanılarak iki model analiz edilmiştir.

Yapılan analizlerin sonucu Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* regresyon modelinden daha iyi olduğunu göstermiştir. Bu sonuç, her modeldeki AIC ve BIC değerlerine bakılarak görülebilir. 20, 50, 75, 100 ve 500 birim içeren simülasyon verileri türetilerek, ve 0.3, 0.5, 0.75, 1, 1.25, 1.5 ve 2 yayılım değerleri için Conway-Maxwell *Poisson* regresyonunun AIC ve BIC değerleri genelleştirilmiş *Poisson* regresyon modelinin AIC ve BIC değerlerinden daha küçük bulunmuştur. Conway-Maxwell *Poisson* regresyon modelinin, aşırı yayılım ve az yayılım sorunları olan verilerin işlenmesinde daha esnek olduğu benzer sonuçların elde edildiği Hayati vd. (2018) çalışmasında da görülebilmektedir.

Simülasyon sonuçlarına göre örneklem hacmi arttıkça Conway-Maxwell *Poisson* regresyon modelinin ve genelleştirilmiş *Poisson* regresyon modelinin AIC ve BIC değerlerinin büyüdüğü görülmektedir. İki modeldeki AIC ve BIC değerleri ne kadar büyük olursa, AIC ve BIC değerleri arasındaki fark o kadar net olacaktır. Dolayısıyla bu çalışmada, genelleştirilmiş *Poisson* regresyon modelinin ve Conway-Maxwell *Poisson* regresyon modelinin analizinde büyük hacimli örneklem kullanılmasının, *Poisson* regresyon modelinde aşırı yayılımın ve az yayılımın üstesinden gelen uygun modeli belirlemede AIC ve BIC değerlerinde farklılıkları daha net göstereceğini göstermektedir.

Bu çalışmada ayrıca gerçek veriler kullanılarak analiz yapılmıştır. 2015 yılında Orta Java eyaletindeki HIV hastalarının *Poisson* regresyon yöntemini kullanarak modellenmesinde, aşırı yayılım yani yayılım varsayımının ihlali söz konusudur. Böylece, bu çalışmada aşırı yayılım probleminin üstesinden gelmek için genelleştirilmiş *Poisson* regresyon modeli ve Conway-Maxwell *Poisson* regresyon modeli tahmin edilmiştir.

2015 yılında Orta Java eyaletindeki HIV hastalarına ilişkin verileri kullanan analizin sonuçları, Conway-Maxwell *Poisson* regresyon modelinin genelleştirilmiş *Poisson* regresyon modelinden AIC ve BIC kriterleri bakımından daha iyi olduğunu göstermiştir. Conway-Maxwell *Poisson* regresyon modelinde, prezervatif kullanan çocuk doğurma yaşındaki çiftlerin yüzdesi (x_1) ve Orta Java eyaletindeki her şehirde 100.000 nüfus başına düşen sağlık çalışanı oranı (x_2) yeni HIV hastalarının sayısı (y) üzerindeki etkisinin istatistiksel olarak anlamlı olduğunu göstermektedir. Ancak, yoksul Orta Java eyaletindeki her şehirde kalabalık alanların yüzdesi (x_3) ve Orta Java eyaletindeki her şehirde 25-34 yaş arası nüfusun yüzdesi (x_4) değişkenleri istatistiksel olarak anlamlı bulunamamıştır. Bu yüzden, prezervatif kullanan çocuk doğurma yaşındaki çiftlerin oranı yüzde 1 artarsa, Orta Java eyaletindeki HIV hastalarının ortalama sayısını 1.390826 kat daha fazla artıracak ve Orta Java eyaletindeki her şehirde 100.000 nüfus başına düşen sağlık çalışanı oranı yüzde 1 artarsa, Orta Java eyaletindeki HIV hastalarının ortalama sayısını $\exp(-0.340482) = 0.711427$ kat azaltacağı sonucuna varılabilir. Model tahmininde prezervatif kullanan çocuk doğurma yaşındaki çiftlerin yüzdesi (x_1) değişkeni hariç, diğer tüm değişken katsayılarının işaretleri beklenildiği gibi tahmin edilmiştir. x_1 değişkeninin beklentinin tersi yönde çıkmasının sebebinin prezervatif kullanım amacının HIV'i önlemekten ziyade üremeyi engelleme amaçlı olmasıdır (Herindrawati, 2017). Ayrıca HIV'e maruz kalmış biri tarafından prezervatif kullanımının eşine bulaşmasını önlemek amacıyla da olabileceği düşünülmektedir.

Simülasyon analizi ve gerçek veriler kullanılarak yapılan analiz sonuçlarına dayanarak, Conway-Maxwell *Poisson* regresyon modelinin, aşırı yayılım ve az yayılım durumunda büyük ve küçük örneklerle sayım verilerinin modellenmesinde esnek ve tercih edilmesi gereken bir yöntem olduğu sonucuna varılmıştır.

KAYNAKLAR DİZİNİ

- Avcı, E., 2018. Flexibility of Using Com-Poisson Regression Model for Count Data. *Statistics, Optimization and Information Computing*, 6, 278-285.
- Avcı, E., Alturk, S., & Soylu, E. N., 2015. Comparison Count Regression Models for Overdispersed Alga Data. *IJRRAS* 25, 1, 1-5.
- Allo, C. G., Otok, B. W., & Purhadi., 2019. Estimation Parameter of Generalized Poisson Regression Model Using Generalized Method of Moments and Its Application. *IOP Conf. Series: Materials Science and Engineering*, 546, 1-8.
- Berk, R. & MacDonald J.M., 2008. Overdispersion and Poisson regression. *Journal of Quantitative Criminology*, 24, 269-284.
- Cahyandari, R., 2014. Overdispersion Testing on the Poisson Regression Model. *Statistika*, 14, 2, 69 –76.
- Chatla, S. B., & Shmueli, G., 2018. Efficient Estimation of COM-Poisson Regression and A Generalized Additive Model. *Computational Statistics and Data Analysis*, 121, 71-89.
- COMPOissonReg: Conway-Maxwell Poisson(COM-Poisson) Regression. Available online <https://cran.r-project.org/web/packages/COMPOissonReg/index.html> (accessed on 15 October 2021).
- Consul, P.C. & Jain, G.C., 1973. A Generalization of the Poisson Distribution. *Technometrics*, 15, 4, 791-799.
- Consul, P.C., 1989. *Generalized Poisson Distributions: Properties and Applications*. New York. Marcel Dekker.
- Consul, P., & Famoye, F., 1992. Generalized Poisson Regression Model. *Communications in Statistics - Theory and Methods*, 89-109.
- Conway, R. W., and W. L. Maxwell. 1962. A queuing model with state dependent service rates. *Journal of Industrial Engineering* 12: 1, 32.
- Darnah., 2011. Handling Overdispersion on Poisson Regression Models with Generalized Poisson Regression I. *Eksponential Journal*, 2, 2, 5-10.
- Dobson, A.J., 2001. *An introduction to generalized Linear Models*. 2nd Edition. Chapman & Hall / CRC.
- Durmuş, B., & İşçi Güneri, Ö., 2020. An Application of The Generalized Poisson Model for Over Dispersion Data on The Number of Strikes Between 1984 and 2017. *Alphanumeric Journal*, 8, 2, 249-259.
- Famoye, F., 1993. Restricted Generalized Poisson Regression Model. *Communications in Statistics - Theory and Methods*. 22(5): 1335-1354.

KAYNAKLAR DİZİNİ (devam)

- Francis, R. A., Geedipally, S. R., Guikema, S. D., Dhavala, S. S., Lord, D., & LaRocca, S., 2012. Characterizing the Performance of the Conway-Maxwell Poisson Generalized Linear Model. *Risk Analysis*, 32, 1, 167-183.
- Geedipally, S., R., 2008. Examining The Application of Conway-Maxwell Poisson Model for Analyzing Traffic Crash Data. Texas A&M University. Published Doctor of Philosophy Dissertation.
- Harris, T., Yang, Z., & Hardin, J. W., 2012. Modeling Underdispersed Count Data With Generalized Poisson Regression. *The Stata Journal*, 12, 4, 736-747.
- Hayati, M., Sadik, K., & Kurnia, A., 2018. Conway-Maxwell Poisson Distribution: Approach for Over- and Under-Dispersed Count Data Modelling. *The 4th International Seminar on Sciences*. IOP Publishing, doi:10.1088/1755-1315/187/1/012039.
- Herindrawati, A. Y., 2017. Poisson Inverse Gaussian Regression Model (Case Study: Number of New HIV Cases in Central Java Province in 2015). Available online: https://repository.its.ac.id/3479/1/1313100015-Undergraduate_Theses.pdf, (accessed on 14 October 2021).
- Huang, A., 2017. Mean-Parametrized Conway-Maxwell-Poisson Regression Models for Dispersed Counts. *Statistical Modelling*, 1-22.
- Huang, A. & Kim A.S.I, 2021. Bayesian Conway-Maxwell-Poisson regression models for overdispersed and underdispersed counts. *Communications in Statistics-Theory and Methods*, 50, 13, 3094-3105.
- Ismail, N., & Jemain, A.A., 2007. Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Available online: https://www.casact.org/sites/default/files/database/forum_07wforum_07w109.pdf (accessed on 14 October 2021).
- Jowaheer, V. & Khan, N.A.M., 2009. Estimating Regression Effects in COM-Poisson Generalized Linear Model. *International Journal of Mathematical and Computational Sciences*, 3, 5, 362-366.
- Kadane, J.B., Krishnan, R. & Shmueli G., 2006. A Data Disclosure Policy for Count Data Based on the Com-Poisson Distribution, *Management Science*, 52, 10, 1610-1617.
- Lindsey, J. K., 1997. Applying Generalized Linear Models. Belgium: Springer.
- Lord, D., Greedipally, S. R., & Guikema, S. D., 2010. Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Underdispersion. *Risk Analysis*.
- McCullagh, P., & Nelder, J.A., 1989 Generalized Linear Models. 2nd Edition, Chapman and Hall, London.

KAYNAKLAR DİZİNİ (devam)

- Pattiz, B., 2009. Count regression models for recreation demand: an application to Clear Lake. Ames: Iowa State University, Unpublished Master of Science Thesis.
- Putri, A. M., Prahutama, A., & Warsito, B., 2019. Bivariate Poisson Regression Modeling On The Number Of Death Of Pregnancy And Public Mothers In Central Java 2017. *Gaussian Journal*, 8, 3, 317-329.
- Rahayu, L. P., Sadik, K., & Indahwati, 2016. Overdispersion Study of Poisson and Zero-Inflated Poisson regression for Some Characteristics of The Data On Lambda, n, p. *International Journal of Advances in Intelligent Informatics*, 2, 3, 140-148.
- Roback, P., & Legler, J., 2021. Beyond Multiple Linear Regression Applied Generalized Linear Model and Multilevel Models in R. CRC Press.
- Ruliana, Hendikawati, P., & Agoestanto, A., 2016. Generalized Poisson Regression (GPR) Modeling to Overcome Equidispersion Violations in Poisson Regression in Measles Cases in Semarang City in 2013. *UNNES Journal of Mathematics*, 5, 1, 1-8.
- Saffari, S. E., Adnan, R., & Greene, W., 2011. Handling of Over-Dispersion of Count Data via Truncation Using Poisson Regression Model. *Computer Science & Computational Mathematics*, 1, 1, 1-4.
- Sasu, A. O., Isaac, A. M., & Boadi, R., 2016. Statistical Models for Count Data with Applications to Road Accidents in Ghana. *International Journal of Statistics and Application*, 6, 3, 123-137.
- Shmueli G., Minka T.P., Kadane, J.B., Borle S. & Boatwright P., 2005, A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution. *Appl. Statist.*, 54, 1, 127-142.
- Sellers, K. F., Borle, S., & Shmueli, G., 2012. The COM-Poisson Model for Count Data: A Survey of Methods and Applications. *Applied Stochastic Models in Business and Industry*, 28, 104-116.
- Sellers, K. F., & Morris, D. S., 2017. Under-Dispersion Models: Models That Are "Under The Radar". *Communications in Statistics - Theory and Methods*.
- Tripathi, R.C. & Gupta R.C., 1984. Statistical Inference Regarding the Generalized Poisson Distribution. *Sankhya: The Indian Journal of Statistics*, 46,B-2, 166-173.
- VGAM: Vector Generalized Linear and Additive Models. Available online: <https://cran.r-project.org/web/packages/VGAM/index.html> (accessed on 15 October 2021).
- Yang, Z., Hardin, J. W., & Addy, C. L., 2009. A Score Test for Overdispersion in Poisson Regression on The Generalized Poisson-2 Model. *Journal of Statistical Planning and Inference*, 1514-1521.
- Yusuf, O. B., & Ugalahi, L. O., 2015. On the Performance of the Poisson, Negative Binomial and Generalized Poisson Regression Models in Nigeria. *American Journal of Mathematics and Statistics* 2015, 5(3): 128-136.

KAYNAKLAR DİZİNİ (devam)

Zhu F., 2012. Modeling time series of counts with COM-Poisson INGARCH models. Mathematical and Computer Modelling, 56, 9-10, 191-203.



EK AÇIKLAMALAR

Gerçek veri setinde *Poisson* regresyon modeli tahmin sonuçları

```
Call:
glm(formula = Y ~ X1 + X2 + X3 + X4, family = poisson(link = "log"),
    data = Data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-6.1448 -2.6307  0.0457  1.8012  6.1071

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.620191   0.545199   4.806 1.54e-06 ***
X1           0.318551   0.033652   9.466 < 2e-16 ***
X2          -0.350092   0.038948  -8.989 < 2e-16 ***
X3          -0.010447   0.002203  -4.742 2.11e-06 ***
X4           0.249904   0.063898   3.911 9.19e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 590.16  on 34  degrees of freedom
Residual deviance: 304.43  on 30  degrees of freedom
AIC: 500.41

Number of Fisher Scoring iterations: 5
```

Olabilirlik oran testinin sonucu

```
Likelihood ratio test

Model 1: Y ~ 1
Model 2: Y ~ X1 + X2 + X3 + X4
  #Df LogLik Df  Chisq Pr(>Chisq)
1    1 -388.07
2    5 -245.20  4 285.73 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Çoklu doğrusallık testinin sonucu

```
      x1      x2      x3      x4
3.898322 2.532842 6.837092 2.317738
```

Poisson dağılım testinin sonucu

```
Goodness-of-fit test for poisson distribution

      X^2 df      P(> X^2)
Likelihood Ratio 544.9461 27 7.855726e-98
```

EK AÇIKLAMALAR (devam)

Yayılım testinin sonucu

```
Overdispersion test

data: MP1
z = 4.1, p-value = 2.066e-05
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
8.485711
```

Genelleştirilmiş *Poisson* regresyon modelinin tahmin sonuçları

```
Call:
vglm(formula = Y ~ X1 + X2 + X3 + X4, family = "genpoisson1",
      data = Data, trace = TRUE)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  1.381865   1.577739   0.876  0.38111
(Intercept):2  0.842737   0.121456   6.939 3.96e-12 ***
X1              0.347232   0.096701   3.591  0.00033 ***
X2             -0.245372   0.108795  -2.255  0.02411 *
X3             -0.016157   0.006438  -2.510  0.01209 *
X4              0.370543   0.184823   2.005  0.04498 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: loglink(meanpar), logloglink(dispond)

Log-likelihood: -148.8791 on 64 degrees of freedom

Number of Fisher scoring iterations: 10

No Hauck-Donner effect found in any of the estimates
```

Conway-Maxwell *Poisson* regresyon modelinin tahmin sonuçları

```
Call:
glm.CMP(formula.mu = Y ~ X1 + X2 + X3 + X4, formula.nu = Y ~
1, data = Data)

Mean model coefficients (with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.491954   1.598307   1.559 0.118968
X1           0.329898   0.097028   3.400 0.000674 ***
X2          -0.340482   0.110012  -3.095 0.001968 **
X3          -0.011863   0.006422  -1.847 0.064708 .
X4           0.262369   0.187873   1.397 0.162556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Dispersion model coefficients (with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.3102     0.2888    -8.125e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 309.1
```

EK AÇIKLAMALAR (devam)

AIC ve BIC değerlerinin sonucu

```
> AIC(MP1)
[1] 500.4064
> AICv1m(GP2)
[1] 309.7583
> AIC(COM)
[1] 309.1326
> BIC(MP1)
[1] 508.1832
> BICv1m(GP2)
[1] 319.0904
> BIC(COM)
[1] 318.4647
```

