

## Aşırı Dağılımlı Sayılabilir Verilerin GLMM ile Modellemesi

Adile TATLIYER TUNAZ<sup>✉</sup>

Kahramanmaraş Sütçü İmam Üniversitesi, Ziraat Fakültesi Zootekni Bölümü, Avşar Kampüsü, Onikişubat Kahramanmaraş  
<https://orcid.org/0000-0002-4239-7072>  
✉: [atatliyer@ksu.edu.tr](mailto:atatliyer@ksu.edu.tr)

### ÖZET

Bu çalışmanın amacı, varyansın ortalamadan büyük olduğu Poisson dağılımlı verilerde, aşırı yayılım sorununun giderilmesidir. Bu amaçla Türkiye İstatistik Kurumu (TÜİK) hayvancılık istatistiklerinden elde edilen, IBBS-1 düzeyinde arıcılık yapan işletme varlığına ilişkin 2004 yılı ile 2022 yılları arasındaki veriler kullanılmıştır. Aşırı yayılım probleminin giderilmesi için, genelleştirilmiş doğrusal karma modellerde (GLMM) farklı model algoritmaları kullanılarak dört model oluşturulmuştur. Modellemelerde birinci model olarak aşırı yayılım durumu kontrol edilmiştir. Aşırı yayılım varlığı tespit edildikten sonra Poisson ve Negatif Binom dağılımı varsayımı ile birlikte modellere şansa bağlı etkiler ilave edilmiştir. Arıcılık yapan işletme varlığına ilişkin verilerde, aşırı yayılımın giderildiği modellerde (Model 2, 3, 4) sabit etki olan yıl etkisi önemsiz bulunurken, aşırı yayılımın görüldüğü Model 1'de yıl etkisi çok önemli bulunmuştur ( $p < 0.0001$ ). Modellemeler ise "Genelleştirilmiş Ki-Kare/Sd" uyum istatistiğine göre karşılaştırılmıştır. Sonuç olarak, aşırı yayılımın olduğu verilerde GLMM'de Negatif Binom dağılımının kullanılması ya da Poisson dağılımında şansa bağlı etkilerin modele dahil edilmesi aşırı yayılım sorununu ortadan kalkabilir.

### Modelling Over-Dispersion Countable Data with GLMM

#### ABSTRACT

This scholarly investigation aims to surmount the predicament of over-dispersion in datasets with Poisson distributed that exhibit a variance larger than the mean. For the aforementioned objective, information regarding the number of beekeeping enterprises at the IBBS-1 level from the years 2004 to 2022, acquired from the TUIK's animal husbandry statistics, has been utilized. To address the issue of overdispersion, four distinct models were developed through the utilization of varied model algorithms within the framework of generalized linear mixed models (GLMM). As the first model in the modeling, overdispersions were checked. Upon identifying the presence of overdispersion, random effects were incorporated into the models with the premise of Poisson and Negative Binomial distribution. In analyzing the data about beekeeping enterprises, it was observed that the fixed effect (year) was found non-significant in the models wherein over-spread was eliminated (Models 2, 3, 4). Conversely, in Model 1 where over-spread was observed, the effect of year was found highly significant ( $p < 0.0001$ ). The models were compared using the "Generalized Chi-Square/Df" fit statistic. The utilization of the Negative Binomial distribution in the GLMM or the incorporation of random effects in the Poisson distribution within the model can effectively address the issue of overdispersion.

**Atıf Şekli:** Tathyier-Tunaz, A. (2024) Aşırı Dağılımlı Sayılabilir Verilerin GLMM ile Modellemesi. *KSÜ Tarım ve Doğa Derg* 27(5), 1194-1201. <https://doi.org/10.18016/ksutarimdog.vi.1357418>  
**To Cite :** Tathyier-Tunaz, A. (2024). Modelling Over-Dispersion Countable Data with GLMM. *KSU J. Agric Nat* 27(5), 1194-1201. <https://doi.org/10.18016/ksutarimdog.vi.1357418>

### Biyometri

### Araştırma Makalesi

### Makale Tarihçesi

Geliş Tarihi : 08.09.2023

Kabul Tarihi : 03.04.2024

### Anahtar Kelimeler

GLMM

Aşırı yayılım

Poisson dağılım

### Biometrics

### Research Article

### Article History

Received : 08.09.2023

Accepted : 03.04.2024

### Keywords

GLMM

Overdispersion

Poisson distribution

## GİRİŞ

Varyans analizinin temel faraziyelerinden başlıcaları: gözlemlerin ve hataların normal dağılması, varyansların homojenliği ve kovaryansın sıfır olması, ortalama ve varyansların bağımsızlığı ve son olarak da toplanabilirliğidir (Mendes, 2002). Bu faraziyelerden birinin ya da birkaçının sağlanmaması durumunda uygulanan istatistik analizde testin gücü zayıf olup, tahminlerin doğruluğu da yetersiz olmaktadır (Akbulut, 2024). Ancak çoğu araştırmalarda verilerin normal dağılmadığı durumlar görülebilmektedir. Örneğin, parsel başına yabancı ot sayısı (Poisson veya negatif binom), çimlenen tohumların yüzdesi (binom), çiçeklenme zamanı (üstel veya gama) gibi. Verilerin normal dağılmadığı durumda, varyans ortalamaya bağlı olmakla birlikte normallik varsayımı ihlal edilir ve dolayısı ile homojen varyans varsayımı da ihlal edilmiş olunur (Stroup, 2015). 2000'li yılların sonuna doğru, istatistiksel metodolojideki gelişmeler ve aynı zamanda bilgisayar teknolojisindeki patlayıcı artışla birlikte Genelleştirilmiş Doğrusal Karma Modeller (GLMM) ortaya çıkmıştır. GLMM, heterojen varyansa sahip olabilen, korelasyonlu ve normal dağılmayan verileri kapsayacak şekilde ANOVA teorisini genişletmiştir. Aynı zamanda GLMM, üstel dağılım ailesinde bulunan dağılımlara sahip olan sayımları ve oranları içeren tahmin edicilere verilen yanıtları bulmaya çalışır.

Sayımla elde edilen veriler, negatif olmayan, kesikli (integer) olup, sağa çarpık Poisson dağılımı göstermektedirler. Poisson dağılımında ortalama ve varyans birbirine eşittir. Ancak, uygulamada Poisson dağılımı gösteren verilerde, ortalama ve varyans kimi zaman eşit olamamaktadır. Çoğu veri analisti, Poisson dağılımı gösteren sayımla elde edilen verilere karekök transformasyonu ya da logaritmik transformasyon uygulamaktadır. Ancak, Stroup (2015) transformasyon uygulanmasının biyolojik verilerin doğasına ters düştüğünü bildirerek ortalama ve varyansın zoraki eşitlenmesinin problem oluşturduğunun altını çizmiştir. Poisson dağılımında, varyans ortalamadan büyük ise aşırı yayılım (overdispersion) sorunu ortaya çıkmaktadır (Stroup 2013). Sayımla elde edilen verilerde aşırı dağılıma neden olan sebepler: bağımlı değişkenin (sayımla elde edilmiş) varyansının çok büyük olması, gözlemlerin kümelenmesi ve aralarındaki heterojenlik, yanlış model kullanılması ve veri setinde çok sayıda sıfırın bulunması olarak gösterilebilir (Morel, 2014; Ser & Yeşilova, 2016). Aşırı dağılımın ihmal edilmesi durumunda, 1. Tip hata oranı artar ve güven aralıkları da yetersiz olmaktadır. Böyle bir durumda, Negatif Binom dağılımının kullanılması alternatif bir çözüm sunmaktadır. Poisson dağılımlı sayımlar elde edilen verilerinin istatistiksel analizinde, genelleştirilmiş doğrusal model (Generalized Linear

Model, GLM) veya şansa bağlı etkilerin modele eklenmesi ile geliştirilen genelleştirilmiş doğrusal karışık model (generalized linear mixed model, GLMM) yaklaşımları kullanılmaktadır.

Bu çalışmanın amacı, sayımla elde edilmiş Poisson dağılımlı aşırı yayılım gösteren verilerin, genelleştirilmiş doğrusal karışık modellerle (generalized linear mixed model, GLMM) ile farklı algoritmalar kullanılarak aşırı yayılım sorununa çözüm üretmektir.

## MATERYAL ve METOD

Bu çalışmanın verisini Türkiye İstatistik Kurumunun (TUIK) hayvancılık istatistikleri göstergelerinden elde edilmiş olup, IBBS-1 düzeyinde arıcılık yapan işletmelerin 2004 yılı ile 2022 yılları arasındaki sayıma dayalı veriler oluşturmuştur (TUIK, 2023).

### İstatistik Analizler

Çalışmada, IBBS-1 düzeyindeki bölgeler (12 farklı bölge) ile yıllar (2004 ile 2022 yılları arası) bağımsız değişkenler olarak modele dahil edilirken, arıcılık yapan işletme sayısı ise bağımlı değişken olarak modele dahil edilmiştir.

### Aşırı yayılım durumu

Poisson dağılımı " $Var(Y) = E(Y) = \mu$ " olarak ifade edilip ortalaması varyansına eşit olan bir dağılımdır. Ortalamanın varyanstan büyük ya da küçük olması dağılımın aşırı ya da az olması durumunu göstermektedir. Aşırı yayılım (overdispersion) durumunda " $Var(Y) > \mu$  (varyans ortalamadan büyük)" şeklinde ifade edilirken, az yayılımda " $Var(Y) < \mu$  (varyans ortalamadan küçük)" şeklinde ifade edilir. Hilbe (2014), aşırı yayılımın veride birbiri ile ilişkili gözlemlerin olması durumunda ya da yine verideki heterojenlik durumu söz konusu olduğunda görülebildiğini bildirmiştir. Bununla birlikte, yayılım parametresi ( $\phi$ ) aşağıdaki gibi hesaplanır (Ser ve Yeşilova, 2016):

Yayılım parametresi (ya da ölçeklendirilmiş sapma) = Sapma (Devians)/serbestlik derecesi

Yukarıda bahsedilen ifadeye göre; Işık (2011) yayınladığı çalışmada  $\phi > 1$  ise aşırı yayılım (overdispersion),  $\phi < 1$  ise az yayılım (underdispersion),  $\phi = 1$  ise aşırı yayılım olmadığını bildirmiştir. Bu çalışmada sadece aşırı yayılım gösterme eğilimi incelenmiştir.

### Genelleştirilmiş Lineer Karma Model (GLMM)

Bağımlı değişkenin normal dağılım göstermediği durumda kullanılabilen Genelleştirilmiş Lineer Karma Modeller (Proc Glimmix), adından da

anlaşılabileceği gibi lineer modellerde sabit etkileri içeren lineer tahmin edicilere ek olarak şansa bağlı etkilerin de modele dahil edilmesi ile oluşturulan bir istatistiksel modeldir (Authement and Knauer, 2023). Diğer bir ifade ile GLMM, doğrusal karışık model ile genelleştirilmiş doğrusal modelin birleşimidir. GLMM' in genel formu şu şekildedir:

$$g(E(Y | u)) = X\beta + Zu \text{ veya}$$

$$E(Y | u) = g^{-1}(X\beta + Zu) = g^{-1}(\eta) = \mu;$$

$$E(u) = 0 \text{ ve } Var(u) = G \text{ şeklindedir.}$$

Burada;  $Y(n \times 1)$  boyutlu gözlem değerleri vektörü,  $X(n \times p)$  boyutlu sabit etkiler desen matrisi ve  $Z(n \times q)$  boyutlu şansa bağlı etkiler desen matrisidir.  $\beta(p \times 1)$  sabit etkiler katsayı vektörü ve  $u(q \times 1)$  şansa bağlı etkiler katsayı vektörüdür.  $g^{-1}(\cdot)$  ise türevlenebilen monoton link fonksiyonunun tersidir. Şansa bağlı etkiler çok değişkenli normal dağılıma sahip olup ortalaması 0, varyans-kovaryans matrisi  $G$ 'dir. Kısaca  $u \sim MVN(0, G)$  şeklinde ifade edilir (Koç & Cengiz, 2012). Bağımlı değişken, dağılımı üstel aileden olduğu sürece sürekli, ikili, sayılı, sıralı, kategorik vb. olabilir. Üstel dağılım ailesi, sayım verileri için Poisson ve Negatif Binom; kesikli veriler için Binom, Bernoulli ve Geometrik; sürekli yanıt veri setinin incelenmesi için Gamma, Normal, Ters Gauss, Beta ve Üstel gibi uygun modelleme için değerli olan çok sayıda dağılımı içerir. (Dobson ve Barnett, 2008). Poisson süreci, sayısal verilerin stokastik farklarını teorik bir beklenti etrafında modellemek için başlangıç noktası olarak kullanılır. Modelin aşırı dağılımı, varyansın beklenen değişimine yönelik farklı model varsayımları nedeniyle dikkate alınmalıdır. İstatistiksel çıkarımların değeri açısından, bu varsayımların seçimi önemli sonuçlara sahiptir. Bu nedenle, negatif binom dağılım parametresi önerilmektedir çünkü bu yöntem, en yaygın kullanılan yaklaşımlarda varsayılan karesel ortalama-varyans ilişkilerini içeren çeşitli yöntemler tanımlar (Yirga ve ark., 2020)

Ortalama ve doğrusal tahminleyici arasındaki bağlantının fonksiyonu olarak ifade edilen link fonksiyonu hem Poisson dağılımda hem de negatif binom dağılımda "log" bağlantı fonksiyonu olarak kullanılmaktadır. Buna göre negatif binom ya da poisson dağılımlı regresyon modelinin ifadesi:

$$\log(\mu_i) = \sum_{i=1}^n B_0 + B_1 X_{i1} + \dots + B_j X_{ij} \text{ şeklindedir.}$$

Çalışmada, öncelikle veri setinin aşırı yayılım gösterip göstermediğine ilişkin uygulanan Model 1, "Poisson Dağılım Varsayımlı Genelleştirilmiş Lineer Karma Model (Poisson Glimmix/ P-Glimmix)" olarak tanımlanmıştır. Model 1'de uygulanan algoritmalar, Stroup (2013) ile Gbur ve arkadaşlarının (2012) kendi

çalışmalarında kullandıkları algoritmalarıdır.

Model 2, aşırı dağılıma alternatif çözüm getirebilen bir model olup, negatif binom dağılıma sahip bir modeldir ve "Negatif binom dağılım varsayımlı Genelleştirilmiş Lineer Karma Model (Negative Binom Glimmix/NB-Glimmix)" olarak tanımlanmıştır (Yirga ve ark., 2020). Model 3 ve Model 4 de ise Model 1 ve Model 2' ye ilave olarak rastgele kesim ve eğim noktası eklenmiştir. Model 3 "Genelleştirilmiş Poisson Dağılım Varsayımlı Genelleştirilmiş Lineer Karma Model (Poisson Glimmix/ GP-Glimmix)" ve Model 4 "Genelleştirilmiş Negatif Binom Dağılım Varsayımlı Genelleştirilmiş Lineer Karma Model (Poisson Glimmix/ GNB-Glimmix)" olarak tanımlanmıştır.

SAS 9.4 yazılımı aracılığı ile "Proc Glimmix" prosedürü kullanılarak Genelleştirilmiş Karışık Doğrusal Modelde (GLMM) analizler yapılmıştır. GLMM'de iki farklı dağılımın kullanıldığı dört farklı modelde aşırı yayılımın belirlenmesi ve modellerin mukayesesi uyum iyiliği kriterlerine göre (genelleştirilmiş ki-kare/sd veya Pearson ki-kare/sd) kıyaslanmıştır. Hilbe (2014) çalışmasında, uyum iyiliği kriterlerinin 1'den büyük olmasının verinin aşırı dağılım gösterdiği anlamına geldiğini bildirmiştir. Bununla birlikte link fonksiyonunun yanlış tercih edilmesi, gözlemlerdeki değişkenliği (varyansı) fazla olması aşırı yayılımın sebeplerindedir (Gbur ve ark., 2012).

Genelleştirilmiş Lineer Karışık Modeller (GLMM) uygulanırken, veriler orijinal ölçekleriyle (veri ölçeği olarak) korunur. Ancak, ortalama değerlerin açıklayıcı değişkenlerin lineer fonksiyonundan tahmin edilmesi durumunda, bunlar model ölçeğinde kullanılır. Bu ölçekler arasında bir uyum sağlamak için, bir bağlantı fonksiyonu kullanılarak model ölçeğindeki ortalama değerler orijinal veri ölçeğine geri bağlanır. Bu metodoloji, orijinal ölçümleri doğrudan farklı bir ölçek düzeyine dönüştürmekten farklıdır (Dawis, 2020). Bununla birlikte, Proc glimmix içindeki dağılımların varsayılan bağlantı fonksiyonları Çizelge 1'de gösterilmiştir.

Çalışmada kullanılan modeller ise aşağıda belirtildiği gibidir. Öncelikle Model 1, aşırı yayılım durumunun belirlenmesi için kurulmuştur. Kategorik değişkenlerin (bölge ve yıl) tanımlanması için "Class" ifadesi kullanılmıştır. Bağımsız ve bağımlı değişkenlerin tanımlanması için ise "Model" ifadesi kullanılmıştır. Syntax kısmında belirtilen "Solution" ifadesi ise regresyon parametre tahminlerinin elde edilmesini sağlamaktadır. Verinin dağılımı için ise, "Dist=" opsiyonu ile birlikte bu opsiyondaki "link=" opsiyonunda da her dağılıma göre özelleştirilen bağlantı fonksiyonu tanımlanmaktadır (Çizelge 1 de ayrıntılı bir şekilde belirtilmiştir). Şansa bağlı değişkenlerin regresyon katsayıları için "Random" ifadesi tanımlanmıştır. Çalışmada bu modelde sadece

kesim (intercept) şansa bağlı olarak kabul edilmesinde dolayı, bu model algoritması aynı zamanda rasgele kesim noktası modelidir. Tekrarlanan (ardışık) ölçümlerin alındığı değişkeni

tanımlamak için ise “Subject” ifadesi kullanılmıştır. Kısaca modeller aşağıda belirtilen şekilde oluşturulmuştur.

Çizelge 1. Tarımsal denemelerde kullanılan ortak dağılımlar ve bağlantı fonksiyonları.

Table 1. Common distributions and link functions for agricultural experiments.

Distribution	Link Function	Syntax dist=	Syntax Link=
Beta	Logit	dist=beta	link=logit
Binomial	Logit	dist=binomial   bin   b	link=logit
Normal	Identity	dist=gaussian   g   normal   n	link=identity   id
Multinomial	Cumulative logit	dist=multinomial   multi   mult	link=cumlogit   clogit
Negative binomial	Log	dist=negbinomial   negbin   nb	link=log
Poisson	Log	dist=poisson   poi	link=log

Model 1: Aşırı yayılım durumunun belirlenmesi amacıyla kurulmuştur.

Model 2: Model 1’den farklı olarak verilerin Negatif Binom dağılımına sahip olduğu varsayılmıştır.

Model 3: Poisson dağılımı varsayımı altında “random” ifadesine, kesimin (intercept) yanı sıra bir şansa bağlı terim (yıl) daha eklenerek rasgele kesim ve eğim modeli oluşturulmuştur.

Model 4: Negatif Binom dağılımı varsayımı altında “random” ifadesine, kesimin (intercept) yanı sıra bir şansa bağlı terim (yıl) daha eklenerek rasgele kesim ve eğim modeli oluşturulmuştur.

## BULGULAR ve TARTIŞMA

Çalışmada arıcılık yapan işletme sayısı verilerinin

her bölge için ortalama ve varyans değerlerine bakılmış ve Çizelge 2’de gösterilmiştir. Elde edilen sonuçlar incelendiğinde, her bölge için arıcılık yapan işletme sayısına ilişkin sayım yoluyla elde edilmiş veride varyansın ortalamadan büyük olduğu görülmüştür. Dolayısı ile aşırı bir dağılım söz konusudur. Şekil 1’de verilerin varyansının ortalamadan ne kadar uzakta olduğunu grafiksel gösterimi sunulmaktadır, Şekil 2’de ise verilerdeki büyük varyanstan dolayı dağılımın normal olmadığı hatta sağa çarpık olduğu görülmektedir. Araştırmacılar çoğu zaman normal dağılmayan verilere log transformasyonu uygulamaktadır. Ancak, Gbur ve arkadaşları (2012) çalışmalarında, log transformasyonun aşırı dağılıma bir çözüm olmadığını bildirmişlerdir.

Çizelge 2. Arıcılık yapan işletme sayısı verilerinin her bölge için ortalama ve varyansı

Table 2. Mean and variance of the number of beekeeping enterprises for each region

Region	M ± SE	Variance
Akdeniz	6496.68 ± 1018.25	19699943.90
Batı Anadolu	2240.21 ± 268.06	1365289.06
Batı Karadeniz	8024.11 ± 819.36	12755523.88
Batı Marmara	3704.11 ± 480.02	4377930.77
Doğu Karadeniz	7169.37 ± 1107.18	23291120.25
Doğu Marmara	3519.58 ± 415.57	3281190.48
Ege	3519.58 ± 415.57	3281190.48
Güneydoğu Anadolu	2973.26 ± 430.22	3516764.09
İstanbul	831.89 ± 146.40	407232.99
Kuzeydoğu Anadolu	3710.32 ± 463.35	4079105.67
Orta Anadolu	4582.63 ± 639.70	7775042.02
Ortadoğu Anadolu	4015.74 ± 557.93	5914499.87

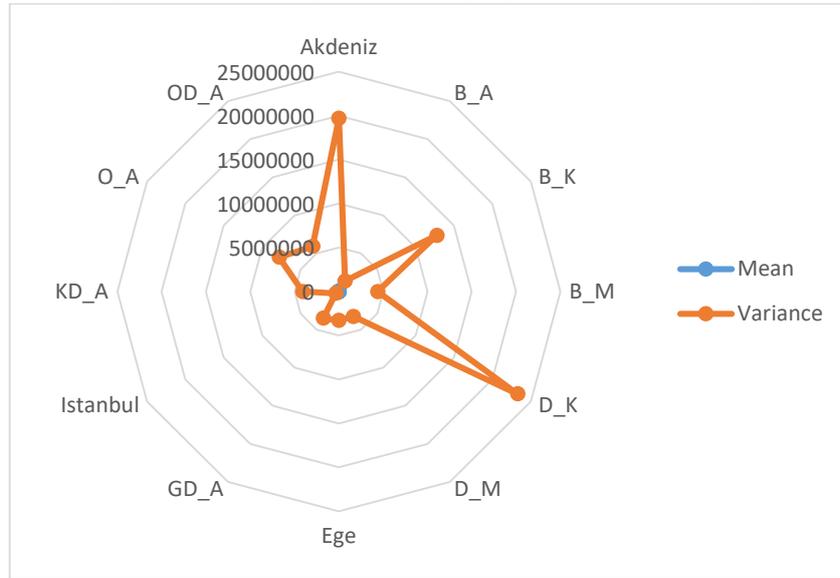
M: Mean (Ortalama); SE: Standard Error (Standart hata)

Çizelge 3’de de görüleceği üzere, Türkiye’de IBBS-1 düzeyine göre gruplandırılmış arıcılık yapan işletme verisinin aşırı yayılımına ilişkin sonuçlarda Model 2, 3 ve 4’deki uyum iyiliği istatistiklerinden elde edilen değerler Model 1’dekine göre oldukça düşüktür. Daha önce de bahsedildiği gibi Model 1, veri setinin aşırı dağılım dağılmadığına ilişkin oluşturulan bir modeldir. Poisson dağılımlı Model 1’deki uyum istatistiği (Genelleştirilmiş Ki-Kare/Sd değeri) 75.94

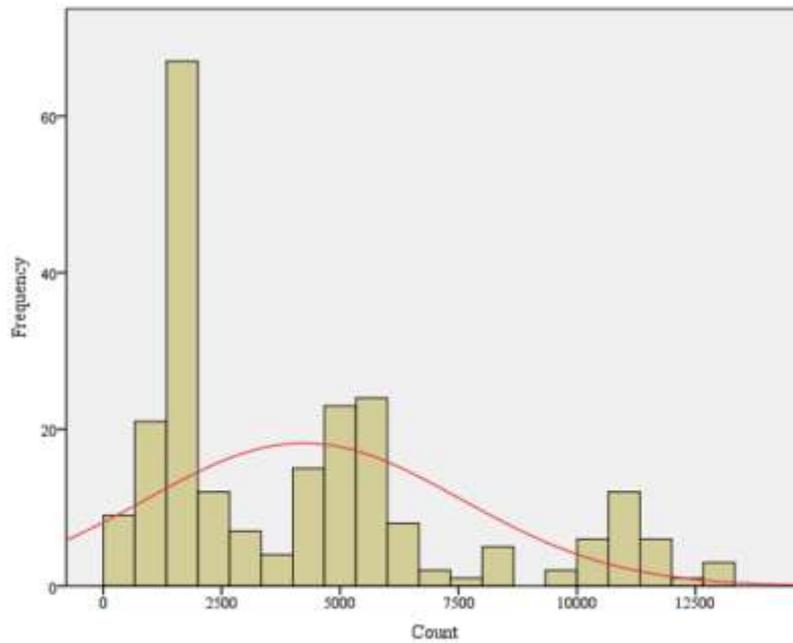
olup 1’den oldukça büyüktür ve dolayısıyla, dağılımın aşırı yayılım gösterdiği sonucuna varılmıştır. Dağılımın aşırı yayılım göstermesi durumunda alternatif bir yaklaşım olan Model 2’de negatif binom dağılıma göre model oluşturulmuştur. Model 2’de uyum istatistiği 1’e yaklaşmış olup, aşırı dağılım sorunu ortadan kaldırılmıştır. Stroup (2013) ile Ser ve Yeşilova (2016), yaptıkları çalışmada, genelleştirilmiş lineer karma modellerde negatif

binom dağılımının kullanılması ile aşırı dağılım sorunun ortadan kalktığını bildirmişlerdir.

Araştırmacıların bildirdikleri bahsi geçen kanı, bu çalışmanın sonucu ile uyumludur.



B\_A: Batı Anadolu, B\_K: Batı Karadeniz, B\_M: Batı Marmara, D\_K: Doğu Karadeniz, D\_M: Doğu Marmara, GD\_A: Güneydoğu Anadolu, KD\_A: Kuzeydoğu Anadolu, O\_A: Orta Anadolu ve OD\_A: Ortadoğu Anadolu bölgesi  
Şekil 1. Her bölge için sayılabilir verilerin ortalama ve varyans arasındaki uzaklık ilişkisi  
Figure1. Distance relationship between mean and variance of countable data for each region



Şekil 2. Aşırı yayılım gösteren sayılabilir verilerin histogram grafiği  
Figure 2. Histogram plot of countable data showing overdispersion

Çizelge 3. Aşırı yayılım durumuna ilişkin uyum istatistiği sonuçları

Table 3. Fit statistic results for the case of overdispersion

Model	Dağılım	Uyum istatistiği	Hesap değeri
Model	Distribution	Fit statistics	P value
Model 1 (P-Glimmix)	Poisson	Genelleştirilmiş Ki-Kare/Sd1	75.94
Model 2 (NB-Glimmix)	Negatif Binom	Genelleştirilmiş Ki-Kare/Sd	1.02
Model 3 (GP-Glimmix)	Poisson	Genelleştirilmiş Ki-Kare/Sd	1.05
Model 4 (GNB-Glimmix)	Negatif Binom	Genelleştirilmiş Ki-Kare/Sd	1.05

SD: Serbestlik derecesi (Degrees of Freedom)

Az yayılım (underdispersion) gösteren Poisson verilerinde kullanılmayan Negatif binom dağılımı, verideki heterojenliği düzelteren dağılıma sahiptir. Bu nedenle, aşırı yayılım probleminde alternatif bir çözüm sunmaktadır (Hilbe, 2014).

Model 3 ve 4'de kesim (intercept) ile birlikte yıl etkisi şansa bağlı terim olarak modellere ilave edilmiştir. Çünkü Gbur ve ark. (2012), şansa bağlı etkilerden dolayı ortaya çıkan değişimin önemli olması durumunda, bu etkileri göz ardı etmek de aşırı yayılıma neden olduğunu bildirmiştir. Bu sebepten dolayı, Model 3 ve 4'de şansa bağlı etkiler dikkate alınmıştır. Şansa bağlı etkilerin dahil edildiği Model 3 ve 4'de Genelleştirilmiş Ki-Kare/Sd değeri aynı çıkmış olup 1'e yakındır.

Ser ve Yeşilova (2016), aşırı yayılım gösteren verilerde yaptıkları modellemelere integral

yaklaşımlarını da modele dahil etmişler, ancak çalışmalarında elde ettikleri sonuçlar az yayılım durumu ortaya çıkmıştır. Bu çalışmada aşırı yayılım durumu incelendiğinden, integral yaklaşımları ihmal edilmiştir.

Çizelge 4'de GLMM ile yapılan dört farklı modelleme ile sabit etkilere ilişkin sonuçlar verilmiştir. Buna göre, Model 1'den elde edilen standart hatalar, Model 2, 3, ve 4'den oldukça küçüktür. Bu sonuç, Ser ve Yeşilova (2016) ile Koç ve arkadaşlarının (2013) yaptıkları çalışmalardan elde edilen sonuçlarla tutarlıdır. Bununla birlikte Koç ve ark. (2013), aşırı yayılım durumunda standart hataların küçük çıkabileceğini bildirmiştir. Bununla birlikte elde edilen sonuçlarda 2022 yılında sabit etkilere ait tahminler 0 olarak bulunmuş olup, Çizelge 4'de yer verilmemiştir.

Çizelge 4. Sabit etkilere ilişkin sonuçlar

Table 4. Results for fixed effects

Y	Model 1 (P-Glimmix)			Model 2 (NB-Glimmix)			Model 3 (GP-Glimmix)			Model 4 (GNB-Glimmix)		
	E	SE	Pr >  t	E	SE	Pr >  t	E	SE	Pr >  t	E	SE	Pr >  t
Sabit	8.77	0.17	<.0001	8.79	0.18	<.0001	8.77	0.19	<.0001	8.78	0.19	<.0001
2004	-1.40	0.01	<.0001	-1.50	0.07	0.06	-1.45	0.08	0.052	-1.45	0.10	0.18
2005	-1.38	0.01	<.0001	-1.42	0.07	0.06	-1.43	0.08	0.054	-1.43	0.12	0.18
2006	-1.40	0.01	<.0001	-1.44	0.07	0.06	-1.45	0.08	0.051	-1.45	0.11	0.12
2007	-1.44	0.01	<.0001	-1.48	0.07	0.06	-1.48	0.08	0.053	-1.48	0.13	0.19
2008	-1.46	0.01	<.0001	-1.51	0.07	0.06	-1.50	0.07	0.051	-1.51	0.14	0.26
2009	-1.44	0.01	<.0001	-1.49	0.07	0.06	-1.49	0.07	0.054	-1.49	0.11	0.25
2010	-1.48	0.01	<.0001	-1.53	0.07	0.06	-1.52	0.07	0.055	-1.53	0.15	0.16
2011	-1.45	0.01	<.0001	-1.50	0.07	0.06	-1.49	0.07	0.054	-1.50	0.13	0.12
2012	-1.45	0.01	<.0001	-1.49	0.07	0.06	-1.49	0.07	0.051	-1.49	0.10	0.14
2013	-0.18	0.00	<.0001	-0.21	0.07	0.06	-0.22	0.07	0.057	-0.21	0.13	0.16
2014	-0.17	0.00	<.0001	-0.19	0.07	0.06	-0.18	0.07	0.056	-0.19	0.12	0.15
2015	-0.15	0.00	<.0001	-0.16	0.07	0.06	-0.16	0.07	0.053	-0.16	0.10	0.17
2016	-0.14	0.00	<.0001	-0.15	0.07	0.06	-0.15	0.07	0.052	-0.15	0.10	0.18
2017	-0.16	0.00	<.0001	-0.15	0.07	0.06	-0.15	0.07	0.055	-0.15	0.10	0.18
2018	-0.17	0.00	<.0001	-0.16	0.07	0.06	-0.16	0.07	0.051	-0.16	0.10	0.16
2019	-0.19	0.00	<.0001	-0.18	0.07	0.06	-0.18	0.07	0.054	-0.18	0.11	0.20
2020	-0.16	0.00	<.0001	-0.15	0.07	0.06	-0.15	0.07	0.053	-0.15	0.10	0.19
2021	-0.07	0.00	<.0001	-0.08	0.07	0.24	-0.08	0.07	0.26	-0.08	0.07	0.24

Y: Yıl (Year); E: Tahmin (Estimate); SE: Standart Hata (Standart Error)

Arıcılık yapan işletme varlığına ilişkin verilerde, aşırı yayılımın giderildiği modellerde (Model 2, 3, 4) sabit etki olan yıl etkisi önemsiz bulunurken, aşırı

yayılımın görüldüğü Model 1'de yıl etkisi çok önemli bulunmuştur (P<0.0001). Bununla birlikte, her dört model için en küçük kareler ortalamasına ait sonuçlar Çizelge 5'de verilmiştir.

Çizelge 5. En küçük kareler ortalamasına ait sonuçlar  
 Table 5. Least squares mean results

Y	Model 1 (P-Glimmix)				Model 2 (NB-Glimmix)				Model 3 (GP-Glimmix)				Model 4 (GNB-Glimmix)			
	Model (Log) Ölçeği		Veri ölçeği		Model (Log) Ölçeği		Veri ölçeği		Model (Log) Ölçeği		Veri ölçeği		Model (Log) Ölçeği		Veri ölçeği	
	E	SE	M	SEM	E	SE	M	SEM	E	SE	M	SEM	E	SE	M	SEM
2004	7.37	0.17	1584.85	273.93	7.34	0.19	1539.22	289.80	7.32	0.19	1513.81	288.82	7.34	0.19	1536.68	289.69
2005	7.39	0.17	1620.72	280.12	7.36	0.19	1578.80	297.25	7.35	0.19	1551.17	295.93	7.36	0.19	1575.98	297.10
2006	7.37	0.17	1593.94	275.50	7.34	0.19	1547.46	291.35	7.33	0.19	1524.59	290.87	7.34	0.19	1545.15	291.29
2007	7.34	0.17	1534.64	265.26	7.31	0.19	1487.66	280.10	7.29	0.19	1470.54	280.56	7.30	0.19	1485.91	280.13
2008	7.31	0.17	1501.60	259.55	7.28	0.19	1455.72	274.10	7.27	0.19	1439.21	274.59	7.28	0.19	1454.01	274.12
2009	7.33	0.17	1528.45	264.19	7.30	0.19	1474.57	277.64	7.29	0.19	1460.38	278.63	7.30	0.19	1473.11	277.72
2010	7.30	0.17	1474.24	254.82	7.26	0.19	1425.96	268.50	7.25	0.19	1414.54	269.88	7.26	0.19	1424.75	268.61
2011	7.32	0.17	1510.62	261.11	7.29	0.19	1468.84	276.56	7.28	0.19	1452.91	277.20	7.29	0.19	1467.23	276.61
2012	7.33	0.17	1520.30	262.78	7.30	0.19	1479.66	278.60	7.29	0.19	1464.46	279.40	7.30	0.19	1478.11	278.66
2013	8.60	0.17	5406.89	934.03	8.58	0.19	5300.07	997.23	8.56	0.19	5212.72	993.74	8.57	0.19	5291.26	996.84
2014	8.60	0.17	5457.68	942.80	8.60	0.19	5446.95	1024.85	8.59	0.19	5399.48	1029.33	8.60	0.19	5442.42	1025.31
2015	8.62	0.17	5568.64	961.96	8.63	0.19	5596.43	1052.97	8.62	0.19	5540.34	1056.17	8.63	0.19	5591.00	1053.29
2016	8.63	0.17	5611.43	969.35	8.64	0.19	5627.23	1058.77	8.63	0.19	5583.18	1064.34	8.63	0.19	5623.06	1059.33
2017	8.62	0.17	5527.39	954.84	8.64	0.19	5645.41	1062.18	8.63	0.19	5585.00	1064.68	8.64	0.19	5639.38	1062.40
2018	8.60	0.17	5456.73	942.64	8.63	0.19	5590.27	1051.81	8.62	0.19	5528.79	1053.97	8.63	0.19	5584.03	1051.97
2019	8.59	0.17	5353.04	924.73	8.61	0.19	5467.95	1028.80	8.60	0.19	5408.29	1031.00	8.61	0.19	5462.12	1029.02
2020	8.61	0.17	5512.47	952.26	8.64	0.19	5653.67	1063.74	8.63	0.19	5570.28	1061.87	8.64	0.19	5645.38	1063.53
2021	8.70	0.17	5998.26	1036.16	8.71	0.19	6056.73	1139.56	8.70	0.19	5990.90	1142.05	8.71	0.19	6050.43	1139.82
2022	8.77	0.17	6454.64	1114.98	8.79	0.19	6562.98	1234.78	8.78	0.19	6477.18	1234.73	8.79	0.19	6554.77	1234.81

Y: Yıl (Year); E: Tahmin (Estimate); SE: Standart Hata (Standart Error); M: Ortalama (Mean) ; SEM: Ortalamanın standart hatası (Standard Error Mean)

Çizelge 5’den elde edilen en küçük kareler ortalamasına ilişkin sonuçlar SAS yazılımındaki “lsmeans year/ ilink” opsiyonu kullanılarak elde edilmiştir. “ilink” ters bağlantı kodunu ifade etmektedir. Bu kodun modele dahil edilmesi ile, verinin orijinal ölçeğinden ortalama elde edilirken, delta metoduyla ortalamasının standart hatası (veri ölçeği) elde edilmektedir. Sabit etkilere ilişkin elde edilen sonuçlarda olduğu gibi, en küçük kareler ortalamasına ilişkin sonuçlarda da aşırı yayılım durumunda (Model 1) standart hatalar Model 2, 3, ve 4’e göre daha küçük olarak elde edilmiştir. log-bağlantı ölçeği olan Model ölçeğinden ( $\eta$ ) elde edilen tahminlerle, veri ölçeğine ( $\mu$ ) ilişkin ortalamalar elde edilmektedir (Ser ve Yeşilova, 2016; SAS 2008). Bu çalışmada, Model 1’de 2004 yılı için  $\eta = 7.37$  ve  $\mu_{2004} = e^{\eta} = e^{7.37} = 1584.85$ ’dir.

## SONUÇ ve ÖNERİLER

GLMM’ler, üstel dağılımlardan birine sahip olan

bağımlı değişkenler için standart doğrusal modellerin kavramını genişletir. GLMM’ler üç bileşenden oluşmaktadır. Bunlar; bağımlı değişkeninin olasılık dağılımını belirleyen stokastik unsur, açıklayıcı değişkenler tarafından karakterize edilen doğrusal modeli gösteren sistematik bir bileşen olan doğrusal tahminleyici ve bağımlı değişkeninin ortalamasını bağımsız değişkenlerin doğrusal kombinasyonuna bağlayan bir bağlantı (link) fonksiyonudur. Dağılımlar için yaygın olarak kullanılan link fonksiyonları, birçok literatürde tartışılmaktadır (Gill ve Torres, 2019; Stroup, 2013). Genelleştirilmiş lineer karma modellerde parametreler maksimum olabilirlik prensiplerine dayanarak tahmin edilir. Modellerde bağımlı değişkene uygulanan çeşitli transformasyon yöntemleri, modelin bazı varsayımlarını (transforme edilmiş verilerin sabit varyanslı normal dağılıma sahip olduğu) karşılmasını sağlar. Ancak, transformasyonlar, regresyon modellerinde çoğunlukla zorlayıcı olabilmektedir, çünkü transformasyonlar bağımlı değişken ile kovaryanslar arasındaki ilişkiyi etkiler.

Hatta transformasyonlar değişkenler arasındaki ilişkinin ve dağılımın şeklini değiştirir ancak çoğu zaman modeli değiştirdiği algılanmamaktadır (Yirga ve ark., 2020). Transformasyon yönteminin seçimi ise araştırmacıların farklı düşünce şekline bağlı olarak da subjektif olabilmektedir.

GLMM'ler bu sorunlardan kaçınır çünkü veriler artık dönüştürülmez; bunun yerine, ortalamaların bir fonksiyonu, ortak değişkenlerin doğrusal bir kombinasyonu olarak modellenir (Gill ve Torres, 2019). Bazı durumlarda, örneğin, tahmin edilen katsayının büyük değerleri için, bir transformasyonun kullanılması, çıkarsamada GLMM'lerin ve Wald tipi istatistiklerin kullanılmasından daha etkilidir (Menard, 2002). Ancak genel olarak transformasyon yöntemleri GLMM'lerin karşısında zayıf kalmaktadır (AMFAR, 2015).

Bu yüzden bu çalışmada veriler transformasyona tabi tutulmadan arıcılık yapan işletme sayılarına ait verilerin zamana ve bölge etkilerine göre değişimi incelenmiştir. Genelleştirilmiş lineer karma modellerde (GLMM) Sas Proc Glimmix prosedürü kullanılarak dört farklı model oluşturulmuştur. Karışık etkili modeller olarak da adlandırılan boylamsal çalışmalar, ilgili zaman veya mekan boyunca bağımlı değişkendeki değişiklikleri ve farklı faktörlerin bu değişiklikler üzerindeki etkilerini incelemek için kullanılır.

Karışık etkili modellerde Genelleştirilmiş ki-kare/sd, Poisson modeli içindeki aşırı dağılımın değerlendirilmesi için iyi uyum istatistiğini verir. Poisson'un varyansı ve ortalaması eşit olduğundan, ölçek parametresi ( $\alpha$ ) 1'dir. Eğer Poisson varsayımı karşılanıyorsa, Genelleştirilmiş ki-kare/sd'nin 1'e yakın olması beklenir. Tahmini değeri olan 75.94 (Çizelge 3) aşırı dağılıma Poisson modeli altında sağlam bir göstergesidir. Aşırı dağılımın temel ve en önemli sonucu, Çizelge 4'de gösterildiği gibi SE'lerin ve test istatistiklerinin etkisidir. Aşırı dağılım gösteren verilerin analiz (Poisson modeli) sonuçları, yanlış tahminlere ve şişirilmiş test istatistiklerine neden olarak standart hataları yanlış bir şekilde tahmin etmiştir. Morel ve Neerchal (2012), çalışmasında " GLMM uyum sağlandığında aşırı dağılıma kontrol edilmelidir ve uyum sağlanan modelden elde edilen çıkarımların doğru olduğundan emin olunmalıdır." şeklinde açıklamada bulunmuştur. Aşırı dağılıma, kullanılan modelin yanlış olduğunu gösterir ve düzeltmeler gerektirmektedir. Bu istenmeyen sonuçları önlemek amacıyla GLMM'lerde en yaygın olarak kullanılan iki yaklaşım ise, standart hataları ve test istatistiklerini düzelterek modelde aşırı dağılıma bir düzeltme eklemek veya sayıları daha uygun şekilde yaklaşan farklı bir olasılık dağılımını gerektirir. Bu amaçla Poisson'un yerine bir alternatif olarak en önde gelen

aday olup ve makul ve önerilen diğer bir metodoloji olan negatif binom yaklaşımıdır. Çizelge 3'de Genelleştirilmiş ki-kare/sd'nin nasıl değiştirdiği gösterilmiştir (1.02). Elde edilen bu sonuç, Negatif binomun verilere Poisson modeline göre çok daha iyi uyduğunu gösterir. Aynı şekilde Poisson dağılımı ve negatif binom varsayımı altında "random" ifadesine, kesimin (intercept) yanı sıra bir şansa bağlı terim (yıl) daha eklenerek rasgele kesim ve eğim modeli oluşturulan modellerde (model 3 ve Model 4) aynı sonuçlar elde edilmiştir. Bununla birlikte, yapılan çalışmada aşırı dağılımı bertaraf etmek için uygulanan diğer modellerde Genelleştirilmiş ki-kare/sd'e görülen ani düşüş elde edilmiştir. Bu, aşırı dağılıma ile başa çıkmak için makul bir GLMM yaklaşımıdır.

Sonuç olarak, birçok alanda kullanılabilen Genelleştirilmiş lineer karma modeller (GLMM), genelleştirilmiş karma modeller ile karma modellerin kombine olarak birlikte kullanıldığı bir modeldir. Sayımla elde edilen verilerde aşırı yayılım varlığı Poisson dağılımı altında GLMM'de incelenebilir. Aşırı dağılım durumu olduğunda ise Poisson dağılıma alternatif olarak şansa bağlı etkilerin modele dahil edilmesi ve Negatif binom dağılımının kullanılması aşırı dağılım problemine çözüm olabilir.

#### KAYNAKLAR

- Akbulut, Ö. (2024). Bilimsel Makalelerde Raporlanması Önerilen Çıkarımsal İstatistikler. *Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi*, 27(1), 238-247.
- AMFAR. (2015). The foundation for AIDS research. 'Statistics: Women and HIV/AIDS'. Retrieved from <https://www.amfar.org/about-hiv-and-aids/facts-and-stats/statistics--women-and-hiv-aids/>.
- Authement, M., & Knauer, M. (2023). Associations between sow body lesions with body condition and subsequent reproductive performance. *Open Journal of Veterinary Medicine*, 13, 111-121. doi:10.4236/ojvm.2023.137010.
- Davis, J. W. (2020). Introduction to generalized linear mixed models: a count data example. <https://site.caes.uga.edu> [Accessed: 14 December 2020].
- Dobson, A. J., & Barnett, A. G. (2008). An introduction to generalized linear models. London: Chapman and Hall/CRC.
- Gbur, E. E., Stroup, W. W., McCarter, K. S., Durham, S., Young, L. J., Christman, M., West, M., & Kramer, M. (2012). Analysis of generalized linear mixed models in the agricultural and natural resources sciences. Madison, WI, USA: ASA, CSSA and SSSA, Inc.
- Gill, J., & Torres, M. (2019). Generalized linear models: a unified approach. Sage Publications, Incorporated.
- Hilbe, J. M. (2014). Modelling count Data (1st ed.).



- New York: Cambridge University Press.
- Işık, F. (2011). Generalized Linear Mixed Models: An introduction for tree breeders and pathologists. Statistic Session class notes, Fourth International Workshop on the Genetics of Host-Parasite Interactions in Forestry, July 31 – August 5, Eugene, Oregon, USA.
- Koç, H., Cengiz, M. A., Koç, T., & Dunder, E. (2013). Aşırı yayımlı veriler için genelleştirilmiş Poisson karma modellerin hava kirliliği üzerine bir uygulaması. *International Anatolia Academic Online Journal Sciences Journal*, 1(2), 3-7.
- Koç, T., & Cengiz, M. A. (2012). Genelleştirilmiş lineer karma modellerde tahmin yöntemlerinin uygulamalı karşılaştırılması. *Karaelmas Fen ve Mühendislik Dergisi*, 2(2), 47-52.
- Menard, S. (2002). Applied logistic regression analysis. Thousand Oaks: Sage.
- Mendeş, M. (2002). Normal dağılım ve varyansların homojenliği ön şartlarının gerçekleşmediği durumlarda varyans analizi tekniğinin yerine kullanılabilir bazı parametrik testlerin I. Tip hata ve testin gücü bakımından irdelenmesi. (Tez no 120165). [Doktora Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü Zootekni Ana Bilim Dalı]. Yükseköğretim Kurulu Ulusal Tez Merkezi.
- Morel, J. G., & Neerchal, N. K. (2012). Overdispersion models in SAS. Cary: SAS Publishing.
- Morel, J. G. (2014). Analysis of data with overdispersion using the SAS system. Paper presented at SAS Global Forum, March 23-26, Washington, DC.
- Ser, G., & Yeşilova, A. (2016). Aşırı yayımlı veri setinin genelleştirilmiş doğrusal karışık model yaklaşımı kullanılarak değerlendirilmesi. *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*, 26(2), 266-273.
- Stroup, W. W. (2013). Generalized Linear Mixed Models: modern concepts, methods, and applications. New York: CRC Press.
- Stroup, W. W. (2015). Rethinking the analysis of non-normal data in plant and soil science. *Agronomy Journal*, 107(2), 811-827.
- Yirga, A. A., Melesse, S. F., Mwambi, H. G., & Ayele, D. G. (2020). Negative binomial mixed models for analyzing longitudinal CD4 count data. *Scientific Reports*, 10(1), 16742.