

Atf İçin: Kılıç D., Akdur H. T. K. ve Bayrak H., (2023). Poisson ve Negatif Binom Regresyon Modellerine Alternatif Olarak Bell Regresyon Modelinin İncelenmesi: Hastanede Kalış Süresine Etki Eden Faktörlerin Belirlenmesi Türkiye Örneği. *İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(2): 1256-1265.

To Cite: Kılıç D., Akdur H. T. K. & Bayrak H., (2023). Investigation of Bell Regression Model as an Alternative to Poisson and Negative Binomial Regression Models: Determining the Factors Affecting Length of Staying Hospital with The Case of Türkiye. *Journal of the Institute of Science and Technology*, 13(2): 1256-1265.

**Poisson ve Negatif Binom Regresyon Modellerine Alternatif Olarak Bell Regresyon Modelinin İncelenmesi:
Hastanede Kalış Süresine Etki Eden Faktörlerin Belirlenmesi Türkiye Örneği**

Duygu KILÇ^{1*}, Hatice Tül Kübra AKDUR², Hülya BAYRAK³

Öne Çıkanlar:

- Sayma Regresyon Modelleri
- Artık Analizi
- Hastanede Kalış Süresi

Anahtar Kelimeler:

- Artık analizi
- Bell regresyon
- Pearson artıklar
- Rasgele Kantil Artıklar
- Sayma verisi

ÖZET:

Sağlık hizmetleri bir bütün olarak düşünüldüğünde akla ilk gelen hizmetlerden bir tanesi hastanede kalınan süredir. Bireylerin hastanede kaldıkları süreye birçok farklı faktör etki eder. Bu çalışmada bireylerin hastanede kalış süresine etki eden faktörler sayma regresyon modelleri kullanılarak incelenmiştir. Ayrıca, sayım verilerinin modellenmesi için Bell regresyon modeli olarak adlandırılan yeni bir regresyon modeli yakın zamanda Castellares vd. (2018) tarafından istatistik literatürüne kazandırılmıştır. En sık kullanılan sayma regresyon modeli olan Poisson regresyon modelin aksine, Bell regresyon modeli aşırı yayılıma izin vermektedir. Ayrıca Bell dağılımı tek parametrelidir olduğundan çeşitli istatistiksel çıkarımları yapmak aşırı yayılıma izin veren diğer bir model olan negatif binom modeline göre daha kolaydır. Bu nedenle, yeni model aşırı yayılımlı veri setlerini modellemek için iyi bir seçenektir. Bell regresyon modelinin TÜİK tarafından yapılan Türkiye Sağlık Araştırması veri seti kullanılarak bir uygulaması verilmiştir. Bell regresyon, Poisson ve Negatif binom regresyon modelleriyle karşılaştırılmıştır. Elde edilen bulgular, Bell regresyon modelinin kullanılan veri seti için diğer regresyon modellerinden daha iyi uyum sağladığını göstermektedir. Bu sonuç Pearson ve Rasgele Kantil artık türleri incelenerek desteklenmiştir.

**Investigation of Bell Regression Model as an Alternative to Poisson and Negative Binomial Regression Models:
Determining the Factors Affecting Length of Staying Hospital with The Case of Turkey**

Highlights:

- Count Data Regression Models
- Residual Analysis
- Randomized Quantile Residual

Keywords:

- Residual analysis
- Bell regression
- Pearson residual
- Randomize quantile residual
- Count data

ABSTRACT:

When health services are considered as a whole, one of the first services that comes to mind is the length of stay in the hospital. Many different factors affect the length of stay of individuals in the hospital. In this study, the factors affecting the length of hospital stay of individuals were examined using count regression models. Also, a new regression model called Bell regression model for modeling count data was recently developed by Castellares et al. (2018). Unlike the Poisson regression model, which is the most commonly used count regression model, the Bell regression model allows for overdispersion. In addition, since the Bell distribution is single-parameter, it is easier to make various statistical inferences than the negative binomial model, which is another model that allows overdispersion. Therefore, the new model is a good option for modeling overdispersed datasets. An application of the Bell regression model is provided using the Turkey Health Survey data set collected by TURKSTAT. Bell regression was compared with Poisson and Negative binomial regression models for this dataset. The findings show that the Bell regression model fits better than other regression models for the data set used. This conclusion was supported by examining Pearson and Random quantile residual types.

¹ Duygu KILIÇ (Orcid ID: 0000-0002-3972-6648), Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

² Hatice Tül Kübra AKDUR (Orcid ID: 0000-0003-2144-0518), Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

³ Hülya BAYRAK (Orcid ID: 0000-0001-5666-4250), Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara, Türkiye

*Sorumlu Yazar/Corresponding Author: Duygu KILIÇ, e-mail: duygukilic4@gmail.com

GİRİŞ

Sağlık hizmetleri, sağlıklı bir yaşamın sürdürülmesi için bir gereklilik olarak önemli bir rol oynamaktadır. Hastaneler, sağlık hizmetlerinin sunulduğu başlıca merkezlerdir. Bazı hastalar hastaneye geldiklerinde tanı veya tedavi için gecelik ya da daha uzun süreli kalmaya ihtiyaç duymadan aynı gün içerisinde ayakta hasta olarak hastanelere gelip yatmadan ayrılabilirken bazı hastalar ise bir gecelik veya daha uzun süreli olarak hastaneye yatırılmaktadır. Bir hastanın hastaneye başvurması ile taburcu edilmesi arasında geçen zaman olarak tanımlanabilen “hastanede yatış süresi” çoğu çalışmada kalite, performans ve verimlilik ile ilgili önemli bir sonuç ölçüsü olarak düşünülmektedir. Yatış süresinin kısalığı hastanın kısa sürede etkin şekilde tedavi edilmiş olmasından kaynaklanabilir ve bu durum kalite göstergesi olarak kullanılabilir (Aydan ve Arıkan, 2021). Bu gibi nedenlerle hastanede kalış süresine etki eden faktörlerin araştırılması önemlidir. Bu araştırmaların analizleri istatistiksel yöntemler kullanılarak yapılmaktadır. Doğru istatistiksel analiz yöntemi ile problemlere etki eden faktörler belirlenebilmektedir. Sağlık ve spor gibi bilimsel alanlardan elde edilen verilerin genellikle sayıya dayalı verilerdir. Örneğin spor yapılan gün sayısı, yürüyüş yapılan gün sayısı, hastanede kalınan gün sayısı gibi. Böyle sayma verilerinin analizleri sayma regresyon modelleri kullanılarak yapılır. Klasik doğrusal regresyon varsayımları sağlanamadığından klasik regresyon gibi analizler yapılırsa hatalı sonuçlar elde edilir.

Sayma verilerine sadece sağlık ve spor gibi alanlarda rastlanmaz. Bunun dışında sigortacılık, eğitim ve psikoloji gibi birçok farklı disiplinde sayma verilerinin modellenmesine yönelik araştırmaların sayısı son yıllarda ciddi bir artış göstermektedir. Poisson regresyon modeli ve Negatif Binom regresyon modeli bu verilerin modellenmesinde sıklıkla kullanılan bir modellerdir. Sayma verilerinde aşırı yayılım görülmektedir. Aşırı yayılım sayma verilerinde sıklıkla rastlanan bir durumdur ve varyansın ortalamadan büyük olması şeklinde tanımlanır. Poisson dağılımı beklenen değer ile varyansın birbirine eşit olması varsayımı nedeniyle aşırı yayılım olduğu durumlarda tercih edilmemelidir. Eğer böyle durumlarda Poisson dağılımı tercih edilirse standart hataların var olandan daha az tahmin edilmesine neden olacaktır ayrıca regresyon parametreleri de hatalı tahmin edilmiş olacaktır. Aşırı dağılımı modellemek için yarı Poisson regresyon modeli, Poisson-inverse Gauss, Poisson-Lognormal ve Negatif Binom regresyon modelleri gibi alternatif modeller önerilmiştir (Altun, 2018).

Sayma verisi kullanılarak yapılan çalışmalara kısaca göz atılmak istenirse:

Sezgin ve Deniz (2004) Türkiye’de grev sayılarına erki eden faktörleri bulmak için negatif binom regresyon modelini kullanmışlardır. Özmen ve Famoye (2007), çalışmalarında Dalyan Plajı’nda 1991-1993 yılları arasında güneşe maruz kalarak ölen caretta caretta yavrularının sayısının modellenmesinde sayma regresyon modellerini kullanmışlardır. Yapmış oldukları analiz sonucunda Genelleştirilmiş Poisson Regresyon modelinin daha iyi uyum sağladığını tespit etmişlerdir. Tamar (2013), yapmış olduğu tez çalışmasında trafik kazalarına etki eden faktörleri belirlemeyi amaçlamıştır. Bu amaçla Aksaray Emniyet Müdürlüğü Trafik Şube Müdürlüğü’ne ait veriyi Poisson regresyon kullanarak analiz etmiştir. Avcı (2018), çalışmasında şizofren hastalara ilişkin verilere sayma verisi regresyon modellerini uygulamıştır. Kullandığı regresyon modelleri Poisson, Negatif Binom ve Conway-Maxwell-Poisson (COM-Poisson) regresyondur. Analiz sonuçlarına göre; veri setine en iyi uyan modelin Negatif Binom Regresyon modeli olduğunu göstermiştir. Güneri ve Durmuş (2020), çalışmalarında veride aşırı yayılım ya da eksik yayılım olması durumunda kullanabilecek modeller olan Poisson ve negatif binom regresyon modellerini karşılaştırmışlardır. Bu karşılaştırmayı yapmak için ABD Tıbbi Harcama Paneli Araştırmasından (MEPS) elde edilen bir veri seti kullanmışlar ve doktor ziyaret sayısını etkileyen faktörlerin sayma modelleri aracılığı ile incelemişlerdir. Çalışma sonucunda negatif binom regresyon

modelinin daha uygun olduğunu göstermişlerdir. Durmuş vd. (2021), yaptıkları çalışmalarında bireylerin sigara içme alışkanlıklarını etkileyen faktörleri belirlemek amacıyla negatif binom regresyon modelini kullanmışlardır. Çalışmalarında Ekim-Aralık 2019 tarihleri arasında gönüllü katılımcıların dijital ortamda yapılan anket sorularına verdikleri cevaplardan elde edilen verileri kullanmışlardır.

Bu makalenin amacı, sayma verilerinin modellenmesinde kullanılan yeni ve alternatif bir regresyon modeli olan Bell regresyon modeli hakkında bilgi vermektir. Ayrıca diğer bir amaç Bell regresyon modelini Poisson ve Negatif Binom regresyon modelleriyle karşılaştırmaktır. Bu amaçla, gerçek bir veri Poisson, Negatif binom ve Bell regresyon modellerine uygulanmıştır. Hastanede kalınan gün sayısını etkileyen faktörler bu üç regresyon modeli kullanılarak incelenmiştir. Akdur ve ark. (2023) Bell ve sıfır yığılmalı Bell regresyon modellerinde Pearson, sapma (deviance) artıklarının yanı sıra rasgele kantil artıklarını uyarlayarak simülasyon çalışmalarında ve gerçek bir veri üzerinde incelemişlerdir. Bell regresyon modelinde de diğer sayma regresyon modellerinde olduğu gibi rasgele kantil artıklarının model uyumunu belirlemekte Pearson ve sapma artıklardan daha iyi olduğu gösterilmiştir (Akdur ve ark., 2023). Sayma regresyon modellerinde daha iyi sonuç verdiği bilinen rasgele kantil artıkları ile birlikte Pearson artıkları bu makalede model uyumunu incelemek için kullanılmıştır.

Makalenin şu şekilde düzenlenmiştir: materyal ve metod bölümünde sayma regresyon modelleri ve bazı artık türleri anlatılmıştır. Bulgular ve tartışmalar bölümünde önerilen modelin Poisson ve Negatif binom regresyon modellerine karşı kullanılabilirliğini göstermek için gerçek veri üzerinde çalışılmıştır. Son bölümde ise sonuçlar verilmiştir.

MATERYAL VE METOT

Adından da anlaşılacağı gibi sayma verileri sayımdan kaynaklanan negatif olmayan tamsayı değerli bir rasgele değişkenin sonucudur. Böyle bir rasgele değişkenin değerleri kesikli tamsayılardır. Genellikle alt limiti sıfırdan başlar kesinlikle negatif olmaz ve bir üst limit yoktur. Sayma verilerinin alt sınırdan kesilen çarpık bir dağılıma sahip olması yaygın bir durumdur. Sayma değerleri büyüdükçe veri heterojenleşerek artan bir varyans durumu oluşturmaktadır. Bu nedenle, sıradan en küçük kareler regresyonu gibi standart modeller, artıkların ortalama sıfır ve bir standart sapma ile normal olarak dağıldığını varsaydıklarından uygun değildir. Cameron ve Trivedi (2013), standart OLS regresyonunun kullanımının “sayma verisinin ortalaması yüksek olmadığı sürece önemli eksikliklere yol açtığını, bu durumda normal yaklaşım ve ilgili regresyon yöntemlerinin tatmin edici olabileceğini” belirtmiştir. Sayma regresyon modelleri genelleştirilmiş lineer model ailesinin bir alt kümesini oluşturur. Sayma regresyon modellerinden en yaygın kullanılan Poisson, negatif binom regresyon modellerine bu bölümde yer verilmekle birlikte Bell regresyon modeli de bu bölümde kısaca açıklanacaktır.

Poisson Regresyon

Poisson regresyon, sayıma dayalı olarak elde edilen verilerde en sık kullanılan yöntemlerin başındadır. Bağımlı değişken Poisson dağılımından gelmektedir. Poisson dağılımında log bağlantı fonksiyonu kullanıldığı için Loglineer model olarak da adlandırılmaktadır (Agresti, 2002). Poisson regresyon tıptan ekonomiye kadar birçok disiplinde kullanılmaktadır.

x_i ve y_i bir veri setini oluşturan gözlemler olsun. x_i bağımsız değişkenlerin bir vektörü ve y_i de bağımlı değişkenlerin bir vektörü olmak üzere Poisson modele ilişkin olasılık fonksiyonu Eş.1’de verildiği gibidir.

$$f(y_i|x_i) = \frac{\lambda_i e^{-\lambda_i}}{y_i!}, y_i = 0,1,2, \dots \quad (1)$$

Poisson regresyonu y_i bağımlı değişkenin Poisson dağılımına uyduğunu varsayar ve λ_i dağılımın ortalamasıdır. Poisson dağılımı sağa çarpıktır. Ancak λ_i büyüdükçe dağılım normal dağılıma yaklaşır.

y_i 'nin beklenen değeri ve varyansı Eş.2'de verilmiştir.

$$\lambda_i = E(y_i|x_i) = Var(y_i|x_i) \quad (2)$$

y_i 'nin beklenen değerinin negatif değerler almamasını sağlamak için, beklenen değer ile bağımsız değişkenler arasındaki ilişkiyi gösteren link fonksiyonu (Cameron ve Trivedi, 2013)

$$\log(\lambda_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (3)$$

veya

$$\lambda_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m) = e^{x_i' \beta} \quad (4)$$

şeklinde. Burada $\beta_0, \beta_1, \dots, \beta_m$ bilinmeyen parametreleri temsil eder. Poisson regresyon analizinde, y_i bağımlı değişkeninin dağılımına dayalı olarak $\hat{\beta}$ tahmin edicileri hesaplamak için en çok olabilirlik metodu başta olmak üzere birçok yöntem vardır.

Negatif Binom Regresyon

Poisson regresyonda ortalama ve varyans birbirine eşittir. Ancak bu durum uygulamada her zaman mümkün olmamaktadır. Veri setinde aşırı yayılım olması durumunda, bu yayılımı dikkate alan negatif binom regresyonun kullanılması daha uygun olmaktadır (Sileshi, 2008). Negatif binom regresyonda modele aşırı yayılımdan kaynaklanan etki için yeni bir parametre eklenerek analiz yapılır. Aşırı yayılım durumu gözardı edilirse yanlış parametre tahminleri ve tutarsız sonuçlar elde edilir.

Negatif binom regresyon, Poisson modelindeki varyansın ortalamaya eşit olduğu şeklindeki kısıtlayıcı varsayımı genişleten Poisson regresyonunun bir genellemesidir. Bu model, Poisson-gamma karma dağılımına dayanmaktadır (Cameron ve Trivedi, 2013).

$$P(y_i; \lambda_i, k|x_i) = \frac{\Gamma(y_i+k)}{\Gamma(k)\Gamma(y_i+1)} \left(\frac{\lambda_i}{\lambda_i+k}\right)^{y_i} \left(\frac{k}{\lambda_i+k}\right)^k \quad (5)$$

Burada k aşırı yayılım parametresidir ve $k = 0$ durumunda fonksiyon Poisson dağılımına indirgenir.

Dağılımın beklenen değer ve varyansı sırasıyla Eş.6 ve Eş.7'de verilmiştir.

$$E(y_i) = \lambda_i \quad (6)$$

$$Var(y_i) = \lambda_i + \frac{\lambda_i^2}{k} \quad (7)$$

Negatif binom regresyonda, Poisson regresyon için yazılan Eş.3 ve Eş.4'de verilen log link fonksiyonu kullanılarak beklenen değer ile bağımsız değişkenler arasındaki ilişki ifade edilebilir.

Bell Regresyon Modeli ve Bell Dağılımı

Bell regresyon yanıt değişkenin sayılabilir ve tek parametrelili Bell dağılımından geldiği durumda kullanılır. Bell dağılımına ilişkin bazı özellikler şu şekildedir:

- Tek parametrelili bir dağılımdır.
- Üstel dağılım ailesinden gelmektedir.
- Poisson dağılımı ile iç içe (nested) dağılımlar olmamasına rağmen küçük değerler için Poisson dağılımına yakınsar.
- Sonsuz derecede bölünebilir (infinitely divisible).
- Bell dağılımı aslında çoklu Poisson sürecinin (multiple Poisson process) özel bir durumudur. Bell regresyon modeli karmaşık çıkarımlar içermediğinden pratik ve uygulanabilir bir yöntemdir.

Tek parametrelili Bell dağılımına ilişkin olasılık fonksiyonu Eş.8'de gösterildiği gibidir.

$$P(Y = y) = \frac{\theta^y e^{-e^\theta + 1} B_y}{y!}, y = 0, 1, 2, \dots \quad (8)$$

Burada $\theta > 0$ ve B_y Bell sayılarını ifade etmektedir. Bell sayıları genel olarak aşağıda verildiği gibidir (Bell, 1934).

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!} \quad (9)$$

B_n sayıları $B_0 = 1, B_1 = 1$ ile başlar. $B_2 = 2, B_3 = 5, B_4 = 15, B_5 = 52$ şeklinde devam eder. Bell dağılımının beklenen değer ve varyansı ise sırasıyla $E(Y) = \theta e^\theta, Var(Y) = \theta(1 + \theta)e^\theta$ şeklindedir. Buradan da açıkça görüldüğü üzere varyans beklenen değerden büyüktür yani aşırı yayılım durumunda kullanıma uygundur.

Eş.8'de verilen olasılık fonksiyonunu Lambert fonksiyonu yardımı ile de göstermek mümkündür.

$$xe^x = a \quad (10)$$

Yukarıda verilen eşitlik Euler ve Lambert tarafından incelenmiştir. Bu eşitliğin sol tarafta verilen fonksiyonun tersi Lambert fonksiyonu olarak ifade edilmektedir. Bu fonksiyon W_0 ile gösterilmektedir.

Regresyon modeli oluşturulurken genellikle yanıt değişkeninin ortalaması modellenir. Bell dağılımının ortalamasına ilişkin bir regresyon yapısı elde etmek için, Bell olasılık fonksiyonu üzerinde yeniden parametrelendirme işlemi yapılır. Lambert fonksiyonu ve Bell dağılımı ilişkisi kullanılır. $\mu = \theta e^\theta$ olmak üzere burada $\theta = W_0(\mu)$ şeklinde gösterilebilirken, W_0 Lambert fonksiyonunu ifade etmektedir. (Castellares vd., 2018).

Bell dağılımına ilişkin olasılık fonksiyonu Lambert fonksiyonu kullanılarak ifade edilirse yeni oluşan olasılık fonksiyonu Eş.11'de verilmiştir. x_i bağımsız değişkenlerin bir vektörü ve y_i de bağımlı değişkenlerin bir vektörü olmak üzere Bell regresyon modeli için olasılık fonksiyonu

$$f(y_i|x_i) = \exp(1 - e^{W_0(\mu_i)}) \frac{W_0(\mu_i)^{y_i} B_{y_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (11)$$

x_i ve y_i bir veri setini oluşturan gözlemler olsun. Bell regresyon modeli y_i bağımlı değişkeninin Bell dağılımına uyduğunu varsayan parametrik bir modeldir. Poisson ve negatif binoma benzer olarak log link fonksiyonu ile aşağıdaki eşitliklerde olduğu gibi yazılabilir:

$$\log(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (12)$$

veya

$$\mu_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m) = e^{x_i' \beta} \quad (13)$$

Sabit terimler hariç Bell regresyonu için log-olabilirlik fonksiyonu $l(\beta) = \sum_{i=1}^n [y_i \log(W_0(\mu_i)) - e^{W_0(\mu_i)}]$ olarak ifade edilir. En çok olabilirlik metodu, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_m)^T$ Bell regresyon modelinin parametre vektörünü tahmin etmek için kullanılır.

Artıkların İncelenmesi

Artıklar, istatistiksel modellerin varsayımlarının geçerliliğini değerlendirmek için sıklıkla kullanılır. Ayrıca model seçimi içinde kullanmak mümkündür. Lineer modeller için homoskedastisite, aykırı değerlerin varlığı, normalligi ve hataların bağımsızlığını doğrulamak için artıklar kullanılabilir (Nobre ve Singer, 2007). Literatürde tanımlanmış farklı artık türleri vardır. Bunlardan bazıları Pearson, sapma (deviance) ve rasgele kantil (RQR) artıklardır. Pearson ve sapma artıklar, genellikle genelleştirilmiş lineer modellerde kullanılmaktadır. RQR ise yine doğrusal modellerde kullanılmakla birlikte sayma verilerinde de kullanılabilen bir yöntemdir ve literatürde sıklıkla kullanıldığı görülmektedir.

Pearson Artıklar

Yanıt değişkeninin tahmini standart sapması ile ölçeklenen klasik artıklar olarak tanımlamak mümkündür. Yani Pearson artıklar, gözlenen ve tahmin edilen olasılıklar arasındaki farkın, tahmin edilen olasılığın standart sapmasına bölümüdür. Genel gösterimi Eş.14'de verilmiştir.

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(y_i)}} \quad (14)$$

Çizelge 1.'de kullanılan regresyon modellerine ilişkin artıklar verilmiştir. Literatürde Pearson artıklara ilişkin yapılan çalışmalar incelendiğinde Bell regresyon modeline ait Pearson artıkların varlığına rastlanmamıştır.

Çizelge 1. Poisson, Negatif Binom ve Bell regresyon modellerine ait Pearson artıklar

Model	Pearson Artıklar
Poisson	$r_i^p = \frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}$
Negatif Binom	$r_i^p = \frac{y_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i + \hat{\lambda}_i^2/k}}$
Bell	$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i[1 + W_0(\hat{\mu}_i)]}}$

Rasgele Kantil Artıklar (Randomize Quantile Residuals-RQR)

RQR yöntemi, yanıt değerinin sayma verileri olduğu regresyon modellerinden elde edilen sonuçların değerlendirilmesinde geleneksel artıkların kullanılmasındaki zorlukların üstesinden gelmek için önerilen bir yöntemdir (Dunn vd.,1996). RQR'yi hesaplamak için gereken tek bilgi, yanıt değişkeninin kümülatif dağılım fonksiyonudur. Tanımdan da anlaşılacağı gibi RQR yöntemi diğer yöntemlere göre daha basittir. RQR yöntemi temelde bir dönüşüm yapmaktadır. $F(y_i, \hat{\mu}_i, \hat{\phi})$, regresyon modelinde $\hat{\mu}_i$ ile parametrelenen rastgele değişken y_i için kümülatif dağılım fonksiyonunu gösterebilir. $\hat{\phi}$ dağılım parametresi bağımsız değişkenler kümesine bağlı değildir. Dağılım fonksiyonu sürekli ise, $F(y_i, \hat{\mu}_i, \hat{\phi})$ fonksiyonu (0,1) üzerinde düzgün bir dağılım gösterir.

RQR ayrıca $q_i = \Phi^{-1}\{F(y_i, \hat{\mu}_i, \hat{\phi})\}$ şeklinde tanımlanabilir, burada $\Phi^{-1}(\bullet)$ standart normal dağılımın kantil (quantile) fonksiyonudur.

Çizelge 2. Poisson, Negatif Binom ve Bell regresyon modellerine ait Pearson artıklar

Model	RQR
Poisson	$z_i^q = \Phi^{-1}(ppois(y_i - 1; \hat{\lambda}_i) + u_i dpois(y_i; \hat{\lambda}_i))$
Negatif Binom	$z_i^q = \Phi^{-1}(pnb(y_i - 1; \hat{\lambda}_i; \hat{k}) + u_i dnb(y_i; \hat{\lambda}_i, \hat{k}))$
Bell	$z_i^q = \Phi^{-1}(pbell(y_i - 1; W(\hat{\mu}_i)) + u_i dbell(y_i; W(\hat{\mu}_i)))$

Burada p ile başlayan ifadeler kümülatif olasılık dağılım fonksiyonlarını gösterirken d 'ler olasılık fonksiyonlarını göstermektedir.

Kümülatif dağılım fonksiyonu kesikli ise, fonksiyonu sürekli hale getirmek için rasgeleleştirme terimi eklenir. Kesikli durum için $p(y_i, \hat{\mu}_i, \hat{\phi})$, $F(y_i, \hat{\mu}_i, \hat{\phi})$ 'nin olasılık fonksiyonu olsun. u_i , [0, 1]'deki düzgün (uniform) dağılımdan gelen rastgele bir sayı olmak üzere dağılım fonksiyonu şu şekilde edilir:

$$F^*(y_i, u_i, \hat{\mu}_i, \hat{\phi}) = F^{-1}(y_i - y_i, \hat{\mu}_i, \hat{\phi}) + u_i \cdot p(y_i, \hat{\mu}_i, \hat{\phi}) \quad (15)$$

Burada $F^{-1}(y_i^-, y_i, \hat{\mu}_i, \hat{\phi})$ alt sınırdır.

RQR, $q_i = q\{F(y_i; \hat{\mu}_i, \hat{\phi})\} = \Phi^{-1}(F^*(y_i; \hat{\mu}_i, \hat{\phi}, u_i))$ şeklinde de ifade edilebilir (Feng vd, 2020). Bell regresyonu için Akdur vd. (2023)'de uyarlanan rasgele kantil artıkları formülü ve diğer regresyon modellerinin rasgele kantil artık formülleri Çizelge 2.'de verilmiştir.

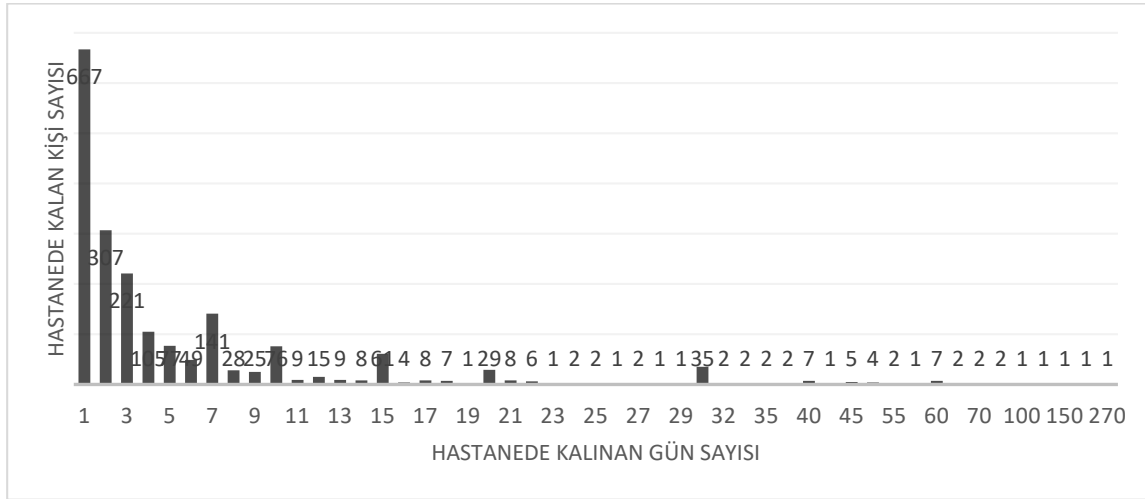
BULGULAR VE TARTIŞMA

2019 Türkiye Sağlık Araştırmasında amaç, ankete katılan kişilerin genel olarak sağlık profilinin ortaya çıkarılması ve buna ek olarak sağlık göstergelerine yönelik bilgilerin elde edilmesidir. Araştırma ile sağlık alanında 0-6, 7-14 ve 15 ve daha büyük yaşta kişiler için çeşitli göstergeler elde edilebilmektedir. Araştırmanın tasarımı Türkiye toplam bazında tahminler üretecek şekilde yapılmıştır.

Çizelge 3. Analizde kullanılan değişken listesi

Değişken Adı	Değişken Açıklaması
Hastanede Kalınan Gün Sayısı (bağımlı değişken)	Bireyin hastanede geçirdiği gece sayısı
Yaş	Kişinin yaşı
Kilo	Kişinin kilosu
Cinsiyet	Kişinin cinsiyeti
Genel sağlık durumu	Kişinin genel sağlık durumu

Bağımlı değişken olarak alınan "Hastanede kalınan gün sayısı" değişkeninin kişilere göre dağılımı Şekil 1'de verilmiştir.



Şekil 1. Gün Süresi Değişkeninin Dağılımı

Yapılan araştırma sonucunda bebek, çocuk ve yetişkinlerin sağlık durumuna ek olarak 15 ve daha büyük yaşta bireylerin sağlık hizmetinden yararlanma durumları, günlük aktivitelerini gerçekleştirirken yaşadıkları zorlukların derecesi buna ek olarak sigara ve alkol kullanma alışkanlıkları gibi göstergeler elde edilmektedir. Bu ankette örnek hacmi olarak 9470 haneye ve toplamda 26075 fert ile anket gerçekleştirilmiştir (Int Kyn. 1). Analiz R Studio programı kullanılarak yapılmıştır.

Çalışmada değişken seçimi sırasında konu ile ilgili yapılan araştırmalar incelenmiştir. İncelenen bu çalışmalardan yola çıkarak belirlenen analizde kullanılan değişkenler aşağıdaki Çizelge 3.'de verilmiştir.

Şekil 1 incelendiğinde hastanede 1 gece geçiren kişi sayısının 667 olduğu görülmektedir. 2 gece geçiren kişi sayısı 307 iken 270 gece geçiren kişi sayısının 1 olduğu görülmektedir.

Yukarıda bahsedilen Poisson model, negatif binom model ve Bell model veri setine uygulanmıştır. Modellerin veri setine olan uygunluğu Akaike Bilgi Kriteri (AIC) ve log olabilirlik (LL) değerleri incelenerek belirlenmiştir. Daha sonra belirlenen en iyi modele ait parametre tahminleri verilmiştir. Modellerle ilişkin oluşan artıklar Pearson ve RQR artık türlerine göre hesaplanarak çizdirilmiştir.

Model Seçimi

Modellerin uygunluğunun karşılaştırılmasında AIC ve log olabilirlik(LL) değerlerine bakılmıştır. AIC değeri en küçük, LL değeri ise en büyük olan model tercih edilmiştir. Çizelge 4 incelendiğinde en düşük AIC ve en büyük LL değerine göre, Bell modelinin verilere diğer modellerden daha iyi uyum sağladığı görülmüştür.

Çizelge 4. Model uyum kriterleri

Model	AIC	LL
Bell regresyon	6546.984	-3265.492
Poisson regresyon	22668.66	-11326.33
Negatif Binom Regresyon	10934.73	-5458.363

Bell modele ilişkin parametre tahmin değerleri Çizelge 5'de verilmiştir:

Çizelge 5. Bell modele ilişkin parametre tahmin değerleri

Log Kısım	Tahmin Değeri	Standart Hata	Z değeri	P değeri	exp(β)
Sabit Terim	1.159	0.136	8.49	< 2.2e-16	3.18
Yaş	0.0092	0.0009	10.18	< 2.2e-16	1.009
Kilo	-0.005	0.0010	-5.09	3.435e-07	0.99
Cinsiyet(kadın)	-0.192	0.030	-6.40	1.486e-10	0.82
Genel sağlık durumu (iyi)	0.190	0.115	1.66	0.09665	1.21
Genel sağlık durumu (orta)	0.655	0.114	5.71	1.123e-08	1.92
Genel sağlık durumu (kötü)	1.052	0.116	9.01	< 2.2e-16	2.86
Genel sağlık durumu (çok kötü)	1.231	0.126	9.69	< 2.2e-16	3.42

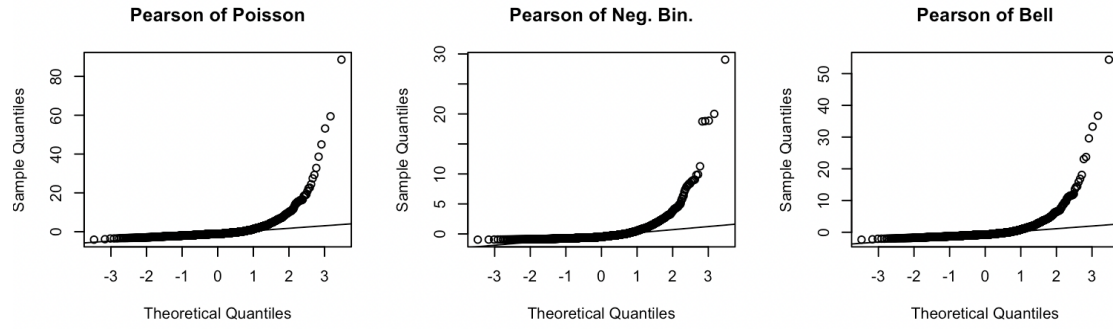
Kurulan regresyon modelinde genel sağlık durumu (iyi) dışında kalan bütün değişkenler anlamlı çıkmıştır ($p < 0.05$). Regresyon denklemi,

$$\mu = \exp(1.159 - 0.0092(\text{yaş}) - 0.005(\text{kilo}) - 0.192(\text{Kadın}) + 0.190 \text{ genel sağlık durumu (iyi)} + 0.655 \text{ genel sağlık durumu (orta)} + 1.052 \text{ genel sağlık durumu (kötü)} + 1.231 \text{ genel sağlık durumu (çok kötü)})$$

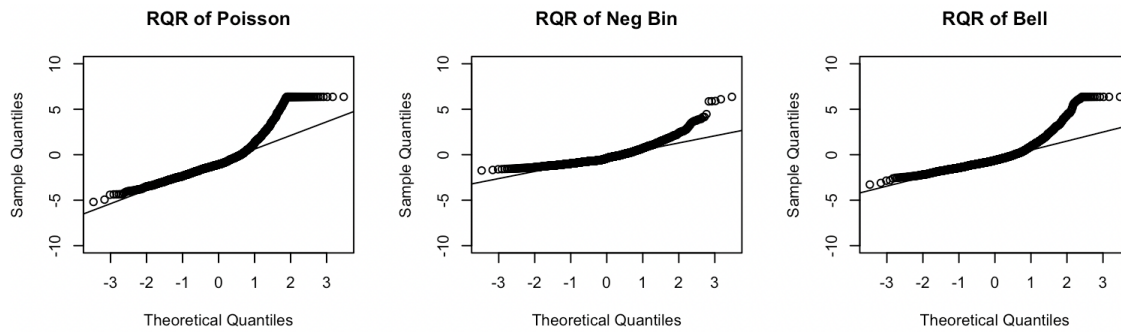
olarak elde edilmiştir. Hastanede kalış süresine yaşın pozitif yönde küçük bir etkisi görülmüştür. Her bir yaşlık artış hastanede bin de 9 daha fazla hastanede kalış süresiyle ilişkilidir. Hastanın cinsiyeti kadın olması hastanede yatış süresinde % 18 (1-0.82) 'lik bir azalma ile ilişkilidir. Diğer değişkenler sabit tutulduğunda, genel sağlık durumu iyi olanlar çok iyi olanlara göre 1.21 kat, genel sağlık durumu orta olanlar çok iyi olanlara göre 1.92 kat, genel sağlık durumu kötü olanlar çok iyi olanlara göre 2.86 kat ve genel sağlık durumu çok kötü olanlar çok iyi olanlara göre 3.42 kat oranında hastanede gece geçirmektedir.

Artıkların İncelenmesi

Bahsedilen üç regresyon modele ait çizilen Pearson ve RQR artıklara ilişkin QQ plotlar aşağıda verilmiştir. Bahsedilen uyum iyiliği ölçütleri sonucu Bell regresyonun veri setine daha iyi uyduğu görülmüştür. Artıklar sonucu elde edilen grafikler de uyum iyiliği testleri gibi verinin bahsedilen regresyon modellerine ne kadar uyum sağladığını göstermenin görsel bir yoludur. Bu çalışmada iki artık türünden bahsedilmiştir ve onlara ilişkin grafikler Şekil 2 ve Şekil 3 de verilmiştir.



Şekil 2. Poisson, negatif binom ve Bell regresyon modelleri için Pearson artıklara ait QQ grafikleri



Şekil 3. Poisson, negatif binom ve Bell regresyon modelleri için rasgele kantil artıklara ait QQ grafikleri

Şekil 2. 'de verilen QQ grafikleri incelendiğinde üç model arası seçim yapmak zordur. Çünkü üç model içinde plotlar birbirine oldukça benzerlik göstermektedir. Buradan da anlaşılacağı gibi Pearson artıkları kullanmak sayma regresyon modelleri için uygun değildir.

Şekil 3.' de verilen QQ grafiklerinde de görüldüğü gibi Bell modeli için QQ grafikleri, eğimli düz bir çizgi boyunca düşer ve diyagonal çizgiden sadece birkaç nokta hafifçe sapar, bu da RQR'lerin normal dağıldığını gösterir. Diğer modeller için QQ grafikleri incelendiğinde ise eğrilik ve boşluklar görülmektedir. Yani veri modele iyi bir şekilde uymamaktadır. Model uyum iyiliği sonucunda da görüldüğü gibi QQ grafikleri Bell modeli veri seti için daha uygun bir model olduğunu doğrulamaktadır.

SONUÇ

Hastaların hastanede kaldığı günlerin toplamı hastanede yatılan gün sayısını oluşturmaktadır. Bu nedenle, hastanede kalış süresi gün olarak ifade edilen sayma verisi tipinde bir değişkendir. Bu bağımlı değişken diğer açıklayıcı değişkenlerle birlikte genellikle literatürde Poisson ve negatif binom regresyon modelleriyle analiz edilmiştir. Bu makalenin amacı literatürde en çok kullanılan negatif binom ve Poisson regresyon modellerine yeni bir alternative olan Bell regresyon modeline dikkat çekmektir. Bell regresyon modelinin en önemli avantajı hem tek parametrelidir olması hem de aşırı yayılımı izin vermesiyle Poisson regresyon modelinin sınırlayıcı eşit yayılım varsayımını esnetebilmektedir. Ayrıca, negatif binom regresyonu iki parametrelidir olması sebebiyle daha fazla tahmin edilmesi gereken parametre ortaya çıkmaktadır. Bell regresyon modeli tek parametre ile aşırı yayılım problemine daha basit bir çözüm sunabilmektedir.

Yapılan bu analiz çalışmasında hastanede kalış süresine etki ettiği düşünülen faktörler olarak kilo, yaş, genel sağlık durumu ve cinsiyet değişkenleri alınmıştır. Veri setine en iyi uyum sağlayacak regresyon modeli model uyum kriterleri kullanarak belirlenmiştir. Bell regresyon modelinin bahsedilen diğer iki modele göre bu veri seti için daha iyi olduğu görülmüştür. Bell regresyon modeli oldukça yeni bir regresyon modelidir. Aynı şekilde RQR artıklarının sayma verilerinde kullanımı da yeni bir yaklaşımdır. RQR yaklaşımı kullanarak Bell modelinin veriye uyumunun incelenmesi literatür için bu

makaleyle yeni bir kazanım olacaktır. Bu nedenle yapılan bu çalışmanın sayma regresyon modellerine yeni bir bakış kazandırması açısından öncü olacağı düşünülmektedir.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

- Agresti A. (2002). *Categorical data analysis (Second Edition)*, New Jersey: Wiley & Sons Incorporation.
- Akdur H.T.K., Kılıç D., Bayrak H. (2023). Residual Diagnostic Methods for Bell-Type Count Models. (İncelemede)
- Altun, E. (2018). A new zero-inflated regression model with application. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 11(2), 73-80.
- Avcı, E. (2018). Using Count Regression Models to Determine the Factors which Effects the Hospitalization Number of People with Schizophrenia. *Journal of Data Science*, 16(3), 511-528.
- Bell, E. T. (1934). Exponential numbers. *The American Mathematical Monthly*, 41(7), 411-419.
- Cameron A.C., Trivedi P.K. (2013). *Regression analysis of count data (Second Edition)*. New York: Cambridge University Press.
- Castellares, F., Ferrari, S. L., & Lemonte, A. J. (2018). On the Bell distribution and its associated regression model for count data. *Applied Mathematical Modelling*, 56, 172-185.
- Dunn, P. K., Smyth, G. K., Randomized quantile residuals, *J. Comput. Graph. Stat.*, 5 (3) (1996), 236–244.
- Durmuş, B., Güneri, Ö. İ., & İncekırık, A. (2021). Bireylerin Sigara Alışkanlıklarını Etkileyen Faktörlerin Belirlenmesi: Negatif Binom Regresyon Analizi. *Yönetim ve Ekonomi Dergisi*, 28(4), 717-730.
- Feng, C., Li, L., & Sadeghpour, A. (2020). A comparison of residual diagnosis tools for diagnosing regression models for count data. *BMC Medical Research Methodology*, 20(1), 1-21.
- Güneri, Ö. İ., & Durmuş, B. (2020). Aşırı ya da eksik yayılım durumunda poisson ve negatif binom regresyon modellerinin karşılaştırılması. *Uluslararası Doğu Anadolu Fen Mühendislik ve Tasarım Dergisi*, 2(1), 48-66.
- Lee, I., & Buchner, D. M. (2008). The importance of walking to public health. *Medicine and science in sports and exercise*, 40(7), S512.
- Ozmen, I., & Famoye, F. (2007). Count regression models with an application to zoological data containing structural zeros. *Journal of Data Science*, 5(4), 491-502.
- Santos Nobre, J., & da Motta Singer, J. (2007). Residual analysis for linear mixed models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 49(6), 863-875.
- Sezgin FH, Deniz E. Poisson regresyon modelinde aşırı yayılım durumu ve negatif binomial regresyon analizinin Türkiye grev sayıları üzerine bir uygulaması. *İstanbul Üniversitesi İşletme Fakültesi Yönetim Dergisi* 2004;15(48):17-25.
- Sileshi G (2008). The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. *Pedobiologia* 52: 1–17. <https://doi.org/10.1016/j.pedobi.2007.11.003>
- Tamar, M. (2013). Poisson Regresyonu. *Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ*.
- TUİK. (2019). Türkiye Sağlık Araştırması Haber Bülteni. URL: <https://data.tuik.gov.tr/Bulten/Index?p=Turkiye-Saglik-Arastirmasi-2019-33661>. (Erişim Tarihi: 26.07.2022).