

PAPER DETAILS

TITLE: POISSON REGRESYON VE EVLİLİK İSTATİSTİKLERİNE UYGULANMASI

AUTHORS: Ozan ALTUNBAS,Özlem DENİZ BASAR

PAGES: 55-67

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/1955897>



İSTANBUL TİCARET ÜNİVERSİTESİ FEN BİLİMLERİ DERGİSİ

İstanbul Commerce University Journal of Science

<http://dergipark.gov.tr/ticaretfbd>



Araştırma Makalesi / Research Article

POISSON REGRESYON VE EVLİLİK İSTATİSTİKLERİNE UYGULANMASI*

POISSON REGRESSION AND APPLICATION TO MARRIAGE STATISTICS

Ozan ALTUNBAŞ¹

Özlem DENİZ BAŞAR²

<https://doi.org/10.55071/ticaretfbd.990732>

Sorumlu Yazar / Corresponding Author
ozan.altunbas@hotmail.com

Geliş Tarihi / Received
23.09.2021

Kabul Tarihi / Accepted
28.12.2021

Öz

Regresyon analizlerinde bağımlı değişkenler normal dağılım gösterebilmekle birlikte iki değere sahip olabilirler. Bağımlı değişkenlere ilişkin durumlar nicel veya nitel olarak ölçülmesiyle veri setleri oluşturulur. Söz konusu gruplar içerisinde yer alan verilerin değerlendirilmesi ve modellenmesi için çeşitli yöntemler bulunmaktadır. Bu yöntemlerden birisi ise poisson regresyon modelidir. Araştırmada Türkiye’de gerçekleşen evlilik sayıları, ortalama evlilik yaşı ve bunlara etken unsurlar incelenmektedir. Bu bağlamda TÜİK’ten 2001-2010 yılları arasında Türkiye’de gerçekleşen evlilik verileri analiz edilmektedir. Verilerin analizinde evlilik sayısı bağımlı değişken olarak, yaş ve evliliğin gerçekleştiği yıllar ise bağımsız değişken olarak kabul edilmiştir. Analizler sonucunda 2001-2010 yılları arasında erkeklerde en çok evlilik sayısının gerçekleştiği yaş grubunun 25-29 yaş grubu olduğu, kadınlarda ise 20-24 yaş grubu olduğu tespit edilmiştir.

Anahtar Kelimeler: Evlilik, genelleştirilmiş doğrusal modeller, Poisson regresyon modeli.

Abstract

Dependent variables in regression analyzes may show normal distribution but may have two values. Data sets are created by measuring the conditions of dependent variables quantitatively or qualitatively. There are various methods for evaluating and modeling the data in these groups. One of these methods is the poisson regression model. In the research, the number of marriages in Turkey, the average age of marriage and the factors affecting them are examined. In this context, the marriage data that took place in Turkey between the years 2001-2010 are analyzed from TUIK. In analysis, number of marriages was accepted as the dependent variable, age and years of marriage as the independent variable. As a result of the analyzes, it has been determined that the age group with the highest number of marriages for men between the years 2001-2010 is the 25-29 age group, and the 20-24 age group for women.

Keywords: Generalized linear models, marriage, Poisson regression model.

*Bu yayın Ozan ALTUNBAŞ isimli öğrencinin İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Programındaki Lisansüstü tezinden üretilmiştir.

¹İstanbul Ticaret Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Sütluçe, İstanbul, Türkiye.
ozan.altunbas@hotmail.com, Orcid.org/ 0000-0003-0438-485X.

²İstanbul Ticaret Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, İstatistik Bölümü, Sütluçe, İstanbul, Türkiye.
odeniz@ticaret.edu.tr, Orcid.org/0000-0002-9430-8975.

1. GİRİŞ

Veri setleri kapsamında bağımlı ve bağımsız değişkenler arasındaki ilişkinin belirlenmesi amacıyla regresyon analizi yapılmaktadır. Bu bağlamda bağımlı ve bağımsız değişkenler arasında neden-sonuç ilişkisi oluşturulabilmektedir.

Sayılabilir verilerin analizinde en kullanışlı ve kolay yöntemlerden biri genelleştirilmiş doğrusal modellerden biri olan Poisson Regresyon modelidir. Poisson regresyon modeli oluşturulurken verilerin olasılıklarını belirlemek amacıyla poisson dağılımı kullanılır. Söz konusu dağılımın önemli özelliklerinden birisi elde edilen sonucun ortalamasının varyansı ile eşit olması olmaktadır. Bununla birlikte genel olarak gerçekleştirilen uygulamalarda varyansın ortalama değerini aştığı gözlemlenmektedir. Bu durum aşırı dağılım (overdispersion) olarak tanımlanmaktadır. Söz konusu durumlarda negatif binom regresyon modelleri işlev göstermektedir (Kabacoff, 2015).

Klasik regresyon modeli ve poisson regresyon modeli arasındaki en önemli fark poisson regresyon modelinde bağımlı değişken için oluşturulan poisson dağılımında değerler negatif olmayan tamsayılardan oluşmaktadır. Bununla birlikte kesikili bir dağılım göstermektedir. Poisson regresyon modelinin en sık kullanıldığı alanlar demografik incelemeler, sağlık araştırmaları, biyoistatistik ve aktüeryal bilimlerdir.

Graff ve ark. (2020) tarafından gerçekleştirilen “Forecasting Daily Wildfire Activity Using Poisson Regression” adlı çalışmada bir ile beş gün arasındaki orman yangınlarının etkinliğini tahmin etmek için iki farklı poisson regresyon modeli oluşturulmuştur. Oluşturulan regresyon modellerinin yangınları tahmin etmede kalıcı modellerden daha doğru sonuç verdiği görülmüştür (Graff ve ark., 2020).

Gao ve ark. (2021) tarafından hazırlanan “Dispersion modelling of outstanding claims with double Poisson regression models” adlı çalışmada sigortacılıkta tazminat karşılıklar için sıklıkla kullanılmakta olan aşırı dağınık Poisson zincir-merdiven modellerinin hali hazırda bulunan kısıtlamalarının da göz ardı edilerek çift Poisson zincir-merdiven modelinde yeni bir dağılım yapısı geliştirilmesi amaçlanmıştır. Önerilen yöntemin mevcut kullanılan yöntemlerden çok daha esnek olduğu sonucuna varılmıştır (Gao ve ark., 2021).

Benz ve ark. (2021) tarafından gerçekleştirilen “Estimating the change in soccer’s home advantage during the Covid-19 pandemic using bivariate Poisson regression” adlı çalışmada seyircili oynanan maçlarda ev sahibi takımın büyük bir avantaja sahip olduğu ortamda taraftarsız oynanan maçların ev sahibi takımı ne derece etkilediği araştırılmıştır. Bu amaçla 17 farklı ligden veriler alınarak iki değişkenli Poisson regresyon modelleri kullanılmıştır. Araştırma sonucunda bulguların karışık olduğu, bazı liglerde avantajın ortadan kalktığı, bazılarında ise avantajın arttığı görülmektedir (Benz ve ark., 2021).

Ercan (2021) tarafından hazırlanan “Hanehalklarının İletişim ve Bilgi Teknolojilerine Erişimi: Bir Poisson Regresyon Analizi” adlı çalışmada hanehalklarının bilgi ve iletişim teknolojileri araçlarının sayısını etkileyecek faktörlerin Poisson regresyon modeli ile araştırılması amaçlanmıştır. Çalışmada yaşanan şehir, okullara ulaşabilmenin zor olması, gelir durumu, öğrenci sayısı gibi faktörlerin teknolojik aletlerin sayısını etkilediği sonucuna varılmıştır (Ercan, 2021).

Vicuña ve ark. (2021) tarafından hazırlanan “Forecasting the 2020 COVID-19 Epidemic: A Multivariate Quasi-Poisson Regression to Model the Evolution of New Cases in Chile” adlı çalışmada Covid-19’un Şili’deki durumunu anlayıp yayılmasını önlemek ve olabilecek en az hasarla süreci geçirebilmek için gerekli olabilecek alternatif yolların Quasipoisson regresyon

modelleri oluturularak analiz edilmesi amaçlanmıştır. Çıkan sonuçlara göre hastalığın yayılımının daha fazla olması bekleniyordu ancak beklenenin aksine ülkede uygulanan karantina politikaları sayesinde yayılım hızının ileriki süreçte azaldığı görüldü (Vicuña ve ark., 2021).

İşçi ve ark. (2021) tarafından hazırlanan “Comparison of Some Count Models in Case of Excessive Zeros: An Application” adlı çalışmada sayım verisinin çok sıfırlı olması durumunda çok sıfırlı poisson regresyon ve poisson hurdle regresyon modellerinin kullanıldığı, aşırı yayılım durumunda da negatif binom regresyon ve negatif binom hurdle regresyon modellerinin kullanıldığı belirtilmiştir. Örnek bir veri seti kullanılarak bu modellerin karşılaştırılmaları yapılmıştır. (İşçi, 2021).

2. POISSON REGRESYON

Poisson regresyon modeli bağımlı değişkenin sayılabilir olduğu durumlarda kullanılabilir. Bir analiz sürecinde gelişen durumlar ve tanımlanan bağımsız değişkenler arasındaki ilişkiyi incelemek amacıyla poisson regresyon modeli kullanılmaktadır. Poisson regresyon modeli devamlı olmaması ve negatif olmayan sonuçlar oluşturması nedeniyle tahmin edilen sayıların logaritması genel anlamda bağımsız değişkenlerin bir fonksiyonu olarak kabul görmektedir.

2.1. Poisson Regresyon Modeli

Poisson regresyon modeli parametre μ ve bağımsız değişken $X' = [1, X_{i1}, X_{i2}, \dots, X_{ip}]_{n \times (p+1)}$ arasındaki ilişkiyi ifade eden Poisson dağılımı tarafından açıklanmaktadır. Söz konusu model poisson dağılımının ortalaması ile ifade edilmektedir. Poisson regresyon modelinin formülü;

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

şeklinde gösterilmektedir (Cameron & Trivedi, 2013). Bu kısımda μ_i ,

$$E(y_i|x_i) = \mu(x_i) = c_i f(x_i, \beta) = \mu_i \quad i = 1, 2, \dots, n \quad (2)$$

olarak tanımlanmaktadır.

2.2. Çok Sıfırlı Poisson Regresyon Modeli

Poisson regresyon modeli kapsamında gerçekleşen aşırı dağılımın değerler kapsamında yer alan sıfırların çokluğu nedeniyle oluştuğu, genelleştirilmiş poisson regresyonu ve çok sıfırlı regresyonu kapsayan negatif binom regresyon modelinde bir varsayım olduğunun düşünüldüğü, çok sıfırlı bir varsayım bulunmaktadır. Bu bağlamda çoklu sıfır regresyon modellerinde ortalama aynı değerlerde olmasına karşın varyansları değişiklik göstermektedir. Sıfır olan sayım verilerini analiz etmek amacıyla oluşturulan çok sıfırlı poisson regresyon modeli şu şekilde gösterilmektedir (Cameron & Trivedi, 2013; Agresti, 2015):

$$P(Y_i = 0) = \pi_i + (1 - \pi_i)e^{-\mu_i} \mu_i^{y_i} = 0 \quad (3)$$

$$P(Y_i = y_i) = (1 - \pi_i) \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \mu_i^{y_i} > 0 \quad (4)$$

Çok sıfırlı poisson regresyon modeli için tahmin edilen değer ve varyans şu şekilde gösterilmektedir:

$$E(Y_i) = (1 - \pi_i)\mu_i \quad (5)$$

$$V(Y_i) = \mu_i(1 - \pi_i)(1 + \mu_i\pi_i) \quad (6)$$

$\pi_i = 0$ çok sıfırlı bir poisson regresyonu kapsamında gerçekleştirildiğinde poisson regresyona girer ve $\pi_i > 0$ olduğunda ise aşırı dağılım göstermektedir.

2.3. Genelleştirilmiş Çok Sıfırlı Poisson Regresyon Modeli

Bir araştırmada sayım verileri kapsamında yer alan sıfırların çokluğunu belirlemek ve çok sıfırlı negatif iki terimli regresyona alternatif olacak şekilde aşırı dağılımı engellemek amacıyla bir model önerilmiştir. Söz konusu model için genelleştirilmiş poisson dağılımı kullanılmıştır (Consul & Famoye, 1992).

$Y_i, i = 1, 2, \dots, n$ bağımlı değişkenlerin ölçümünü tanımlamaktadır. Dolayısıyla Y_i olasılık fonksiyonu şu şekilde tanımlanabilmektedir:

$$f(Y_i > y_i) = \left(\frac{\mu_i}{1+\theta\mu_i}\right)^{y_i} \left(\frac{1+\theta y_i}{y_i!}\right)^{y_i-1} \exp\left\{\frac{-\mu_i(1+\theta y_i)}{1+\theta\mu_i}\right\}, y_i > 0 \quad (7)$$

Ortalama ve varyans genelleştirilmiş poisson dağılımı için şu şekilde tanımlanmaktadır:

$$E(Y_i) = \mu_i \quad (8)$$

$$V(Y_i) = \mu_i(1 + 0\mu_i)^2 \quad (9)$$

Çok sıfırlı poisson regresyon modelinden farklı olarak bu modelde θ parametresi için aşırı ya da zayıf bir dağılım koşulu söz konusu olmaktadır. Bu bağlamda modelde $\theta > 0$ olduğu takdirde aşırı dağılım gerçekleşmesine karşın $\theta < 0$ olduğu durumda ise zayıf dağılım gerçekleşmektedir (Famoye & Singh, 2006). Genelleştirilmiş poisson regresyon modelinde tahmin edilenden daha çok sıfır bulunduğu durumlarda kullanımı uygun bulunmamaktadır (Cameron & Trivedi, 2013; Agresti, 2015).

2.4. En Çok Olabilirlik Tahmin Edici

Log-olabilirlik fonksiyonunun optimal duruma getirebilmek ve en çok olabilirlik yaklaşımında $\hat{\beta}$ tahminleri tercih edilmektedir. Poisson regresyon modeli kapsamında ortalama parametresi şu şekilde formüle edilmektedir:

$$E(y_i|x_i) = \mu_i = \exp(x_i'\beta) \quad (10)$$

Söz konusu denklem en çok olabilirlik fonksiyonunda ise şu şekilde yer almaktadır:

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

Bağımsız incelemeler için en çok olabilirlik işlevi şu şekilde yazılabilir:

$$L(y, \beta) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = \frac{(\prod_{i=1}^n \mu_i^{y_i}) e^{-\sum_{i=1}^n \mu_i}}{\prod_{i=1}^n (y_i!)} \quad (12)$$

Bu bağlamda Log-olabilirlik fonksiyonu şu şekilde tanımlanmaktadır:

$$l(\beta) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n (\mu_i) - \sum_{i=1}^n \ln(y_i!) \quad (13)$$

Denklemden $\mu_i = \exp(x_i' \beta)$ yerine koyulduğunda;

$$l(\beta) = \sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n e^{x_i' \beta} - \sum_{i=1}^n \ln(y_i!) \quad (14)$$

sonucuna ulaşılır.

En çok olabilirlik tahmin edici için log-olabilirlik fonksiyonunun β 'ya göre birinci dereceden kısmi türevi alınır ve sifıra eşitlenir.

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^n y_i x_i' - \sum_{i=1}^n x_i e^{x_i' \beta} = \sum_{i=1}^n x_i (y_i - \mu_i) = 0 \quad (15)$$

$$\sum_{i=1}^n x_i' = (y_i - \exp(x_i' \beta)) = 0 \quad (16)$$

$$\sum_{i=1}^n x_i' (y_i - \mu_i) = 0 \quad (17)$$

biçiminde yazılır (Cameron & Trivedi, 2013).

Yukarıda yer alan denklemde görüldüğü üzere açıklayıcı değişkenler sabit bir ifade ile tanımlandığı durumlarda $(y_i - \mu_i)$ artıkların toplamı sıfır olmaktadır. En çok olabilirlik yaklaşımında üstel ortalama parametresi ve log-olabilirlik fonksiyonlarının doğrusal olmaması nedeniyle tek aşamalı bir çözüm oluşturulamamaktadır.

Ampirik bir yöntem ile $\hat{\beta}_p$ 'nin değerinin belirlenememesi nedeniyle aşamalı ağırlıklı küçük kareler (IWLS) yönteminden yararlanılabilir. Aşamalı ağırlıklı en küçük kareler yöntemi algoritması şu şekildedir:

$$\beta_{ML} = (x' \hat{W} \hat{Z})^{-1} x' W Z \quad (18)$$

Burada $W = \text{diag}[\hat{\mu}_i]$ ve $\hat{Z}$,

$$\hat{Z}_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \quad (19)$$

En çok olabilirlik tahmin edici için asimptotik olarak normal dağılım gösteren kovaryans matrisi şu şekilde gösterilmektedir:

$$\text{Var}(\hat{\beta}_{ML}) = [-E \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k'} \right)]^{-1} = (x' \hat{W} x)^{-1} \quad (20)$$

Bununla birlikte MSE eşitliği şu şekilde tanımlanmaktadır:

$$E(L_{ML}^2) = E(\hat{\beta}_{ML} - \beta)' (\hat{\beta}_{ML} - \beta) = \text{tr} \left[(x' \hat{W} x)^{-1} \right] = \sum_{j=1}^j \frac{1}{\lambda_j} \quad (21)$$

$\lambda_j, x' \hat{W} x$, matrisin j .inci özdeğeridir.

Açıklayıcı değişkenler arasında güçlü bir ilişki oluşturulduğu takdirde ağırlıklı çarpaz-çarpım matrisi $x' \hat{W} x$ zayıf bir şekilde koşullu bir boyut kazanır. Bu durum neticesinde ise varyans oranı yükselir ve vektör en olasılık tahmini sonucu oluşur. Tahmin edilen katsayıların ortalama vektörünün çok uzun olması nedeniyle tahmin edilen katsayının okunması oldukça zordur. Mevcut veri ve modellere göre varyans değeri sonucu (Cameron & Trivedi, 2013; Agresti, 2015):

$$\hat{\beta}_{PR} \sim N[\beta, V_{ML}[\beta_{PR}]]$$

$$V_{ML}[\hat{\beta}_{PR}] = (\sum_{i=1}^n \mu_i x_i x_i')^{-1} \quad (22)$$

2.5. Poisson Regresyon Katsayıları ve Değerlendirilmesi

Poisson regresyon kapsamında elde edilen katsayılar ile bağımlı değişken y_i değerleri için tahminler üretilebilir. Bağımsız değişkenler kapsamında bir birim vektörün etkisi, doğrusal modellerde $E(Y|x_i) = x_i' \beta$ ve β katsayıları şeklinde tanımlanmaktadır. i. bağımsız değişken için üstel koşullu ortalama şu şekilde ifade edilmektedir:

$$\frac{\partial E(y|x_i)}{\partial x_i} = \beta_i \exp(x_i' \beta) \quad (23)$$

Burada $\hat{\beta}_i$ için x_i 'deki bir birimlik değişime bağlı olan $E(Y|x_i)$ 'deki değişimi belirlediği ifade edilebilir.

3. EVLİLİK İSTATİSTİKLERİ ÜZERİNE POISSON REGRESYON ANALİZİ UYGULAMASI

Ulusal çapta gerçekleştirilen araştırmaların yanı sıra dünya genelinde gerçekleştirilen pek çok araştırma sonucunda erken yaşta evlilik sorununun bulunduğu gözlemlenmektedir. Bu bağlamda evlilik yaşı ülkeden ülkeye veya bölgeden bölgeye farklılık göstermesine karşın genel olarak değerlendirildiğinde ortalama evlilik yaşının düştüğü gözlemlenmektedir. Bu durum toplumlar açısından gelecek zaman diliminde sorunlar oluşmasına neden olabilecektir.

UNICEF (United Nations International Children's Emergency Fund - Birleşmiş Milletler Uluslararası Çocuklara Acil Yardım Fonu) tarafından hazırlanan “Çocuk Evliliğini Sona Erdirme” raporu değerlendirildiğinde küresel boyutta 700 milyondan fazla kadının erken yaşta evlendiği gözlemlenmektedir. Bununla birlikte söz konusu grup kapsamında 250 milyon kadının ise 15 yaşından küçükken evlendiği gözlemlenmektedir (UNICEF, 2013: 2). UNICEF tarafından hazırlanan söz konusu raporda erken yaşta evlenen kadınların %42'sinin Güney Asya Bölgesinde yaşadığı gözlemlenmektedir. Bununla birlikte sanayi açısından gelişmiş ülkelerde söz konusu oranın %2, Orta ve Doğu Avrupa Bölgesinde ise %4 olduğu gözlemlenmektedir.

Bireylerin sosyo-demografik özellikleri, evliliği ve evlilik yaşını etkileyen en önemli etmenlerden biri olmaktadır. Bununla birlikte erken yaşta evliliklere etki eden diğer etmenler şu şekilde sıralanabilir (Jensen & Thornton, 2003):

- Ekonomik problemler
- Yetersiz eğitim
- Geleneksel yaklaşım
- Dini pratikler
- Aile içi şiddet
- Toplumsal baskı
- Felaketler

Söz konusu etmenler bireylerde genel anlamda psikolojik olarak derin hasarlar oluşturmakla birlikte toplumsal açıdan eğitimsizliği, yoksulluğu ve anne ya da çocuk ölümlerinin artışına neden olmaktadır (Aydemir, 2011: 3).

Türkiye’de yaş ve yıllar baz alınarak gerçekleştirilen evlilik sayılarının değişiminin regresyon ile incelenmesi gerçekleştirilen evliliklerinin neden ve sonuçlarının belirlenmesi açısından önem

taşımaktadır. Bu bağlamda araştırma kapsamında Poisson Regresyon uygulaması için 2001-2010 yılları arasında Türkiye’de gerçekleştirilen evlilik sayılarından, yaşlarından ve yıllarından yararlanılmıştır.

3.1. Türkiye’de Gerçekleştirilen Evlilik Verileri

Araştırma kapsamında Türkiye’de 2001-2010 yılları arasında gerçekleştirilen evlilik sayıları istatistikleri bağımlı değişken olarak tanımlanmakla birlikte evliliklerin gerçekleştiği yıl ve yaşlar ise bağımsız değişken olarak tanımlanmıştır. Elde edilen veriler doğrultusunda evlilik yaşları için 5 farklı yaş grubu oluşturulmuş ve sınıflandırılmıştır. Söz konusu yaş grupları şunlardan oluşmaktadır:

Tablo 1. Analizde Kullanılan Yaş Aralıkları

18-19 yaş,
20-24 yaş,
25-29 yaş,
30-34 yaş,
35 yaş ve üzeri,

Yıllar bazında hangi yaş aralığında kaç evlilik gerçekleştiğini belirten veri grubu TÜİK’ten alınmıştır.

Tablo 2. 2001-2010 Evlenme Sayıları (Yaşlara ve Cinsiyetlere Göre)

Yıl	Cinsiyet	18-19	20-24	25-29	30-34	35 Yaş ve Üzeri
2010	Erkek	12.824	163.791	237.474	91.054	75.572
	Kadın	89.136	212.923	132.952	45.817	56.149
2009	Erkek	14.171	172.563	240.851	87.490	74.595
	Kadın	95.319	217.547	132.455	42.887	55.675
2008	Erkek	15.616	194.610	264.317	90.898	74.318
	Kadın	108.250	241.925	141.774	44.891	55.430
2007	Erkek	16.036	198.418	269.688	87.497	64.393
	Kadın	113.364	250.828	143.251	43.311	36.834
2006	Erkek	16.332	201.822	264.032	85.970	65.650
	Kadın	113.547	254.430	138.415	42.278	37.085
2005	Erkek	16.682	214.130	255.566	89.404	63.570
	Kadın	115.398	265.144	131.322	43.453	34.309
2004	Erkek	15.943	211.516	237.414	85.995	62.321
	Kadın	108.800	261.813	120.598	41.779	33.087
2003	Erkek	16.688	199.060	215.184	74.323	57.977
	Kadın	101.554	243.954	107.896	35.102	30.981
2002	Erkek	16.018	180.564	191.885	65.308	53.788
	Kadın	91.508	225.043	96.578	30.482	29.281
2001	Erkek	23.352	204.553	194.966	62.432	54.646
	Kadın	99.034	225.813	92.708	28.893	29.656

Tablo 2’de yer alan veriler incelendiğinde 2002 yılında erkeklerde en çok evlilik sayısının 25-29 yaş grubunda, kadınlarda ise 20-24 yaş aralığında gerçekleştiği gözlemlenmektedir. 2006 yılında

gerçekleşen evlilik sayıları incelendiğinde de erkeklerde en çok evlilik sayısının 25-29 yaş grubunda, kadınlarda ise 20-24 yaş aralığında gerçekleştiği gözlemlenmektedir. Bununla birlikte 2010 ve 2006 yıllarında erkeklerde 25-29 yaş grubunda gerçekleşen evlilik sayıları ve kadınlarda 20-24 yaş aralığında gerçekleşen evlilik sayıları birbirine yakın değerlerdedir. 2010 yılında gerçekleşen evlilik sayıları incelendiğinde ise erkeklerde en çok evlilik sayısının 25-29 yaş aralığında ve kadınlarda ise 20-24 yaş aralığında gerçekleştiği gözlemlenmektedir. Bununla birlikte evlilik sayıları incelendiğinde genel olarak evlilik sayılarında bir artış veya düşüş olduğu söylenememektedir.

3.2. Evlilik Verilerinden Elde Edilen Sonuçlar

Evlilik verilerinin düzenlenmesi amacıyla “Tidyverse” kütüphanesi ve Excel formatlarının okunması amacıyla “readxl” kütüphanesi kullanılmıştır. TÜİK’ten elde edilen 2001-2010 yıllarına ilişkin evlilik verileri “read_excel” fonksiyonu kullanılarak okutulmuştur. Okutulan veriler “pipe operatörü” ve “gather fonksiyonu” kapsamında veri seti olarak kullanılmıştır.

Evlilik bağımlı değişkeninin, poisson dağılımına uygunluğu pearson istatistiğiyle kontrol edilmiştir. Pearson istatistiği modeller için uyum iyiliği ölçüm şekli olup seride aşırı yayılım olup olmadığının kontrolünde kullanılır. (Deniz, 2005). Yapılan test sonucunda p-değerinin 0,05’ten küçük olduğu durumlarda aşırı yayılım olduğu söylenir.

Buna göre Pearson istatistiğinin sonucu aşağıdaki tabloda gösterilmiştir.

Tablo 3. Pearson İstatistiği Uyum İyiliği Tablosu

Yıllar	Evlilik Sayısı	Beklenen Yıllar	Beklenen Evlilik Sayıları	Standart Artık Yıllar	Standart Artık Evlilik Sayıları	İstatistik Değeri	P-Değeri
2001	544.322	1996	593.927	0,6883	-0,6883	3,049	0,9623
2002	510.155	1994	593.228	1,1533	-1,1533	3,049	0,9623
2003	565.468	2000	595.116	0,4110	-0,4110	3,049	0,9623
2004	615.357	2006	596.846	-0,2563	0,2563	3,049	0,9623
2005	641.241	2009	597.883	-0,5999	0,5999	3,049	0,9623
2006	636.121	2010	598.024	-0,5270	0,5270	3,049	0,9623
2007	638.311	2011	598.376	-0,5523	0,5523	3,049	0,9623
2008	641.973	2012	598.771	-0,5973	0,5973	3,049	0,9623
2009	591.742	2008	597.608	0,0812	-0,0812	3,049	0,9623
2010	582.715	2009	597.637	0,2065	-0,2065	3,049	0,9623

Pearson Ki-kare testinin sonucu Tablo 3’te de gösterildiği üzere;

p-değerinin 0,05’ten büyük olduğu görülmüştür. Bu sonuca göre bağımlı değişken aşırı yayılım durumu göstermemektedir ve bağımlı değişken poisson dağılımına uygundur.

Tablo 4. Yaşa Göre Evlilik Sayılarının Regresyon Tablosu

	Tahmin	Standart Hata	z-değeri	p-değeri
(Intercept)	10,1689125	0,001384617	7344,2056	0,000000
Yas18-19	0,8328306	0,001658552	502,1432	0,000000
Yas20-24	2,1188431	0,001465454	1445,8612	0,000000
Yas25-29	1,9343868	0,001481291	1305,8786	0,000000
Yas30-34	0,8491132	0,001654488	513,2180	0,000000
35 Yaş ve Üzeri	0,6951860	0,001695232	410,0830	0,000000

Tablo 4 incelendiğinde poisson regresyon testi uygulanmış p-değerlerinin 0,05’den küçük olduğu gözlemlenmektedir. Bu bağlamda verilerin anlamlı olduğu ifade edilebilir. Söz konusu katsayılar istatistiksel açıdan önemli bir katkı sağlaması nedeniyle model içerisine dahil edilmektedir. Katsayılara bakıldığında 20-24 yaş arasındaki evlilik sayıları modele en çok etki eden yaş değişkeni olup, 20-24 yaş aralığındaki bir birimlik artış modele 2,12 birimlik artı yönlü etki etmektedir.

Yıllara göre gerçekleştirilen evlilik sayılarının regresyon tablosu aşağıda gösterilmektedir.,

Tablo 5. Yıllara Göre Evlilik Sayılarının Regresyon Tablosu

	Tahmin	Standart Hata	z-değeri	p-değeri
(Intercept)	13,20729626	0,0009584226	13780,24254	0,000000
2002	-0,06482638	0,0013779217	-47,04649	0,000000
2003	0,03811272	0,0013426829	28,38550	0,000000
2004	0,12266160	0,0013157167	93,22798	0,000000
2005	0,16386438	0,0013031930	125,74068	0,000000
2006	0,15584781	0,0013055987	119,36885	0,000000
2007	0,15928464	0,0013045655	122,09785	0,000000
2008	0,16500526	0,0013028518	126,64929	0,000000
2009	0,08352975	0,0013279822	62,89975	0,000000
2010	0,06815723	0,0013329015	51,13449	0,000000

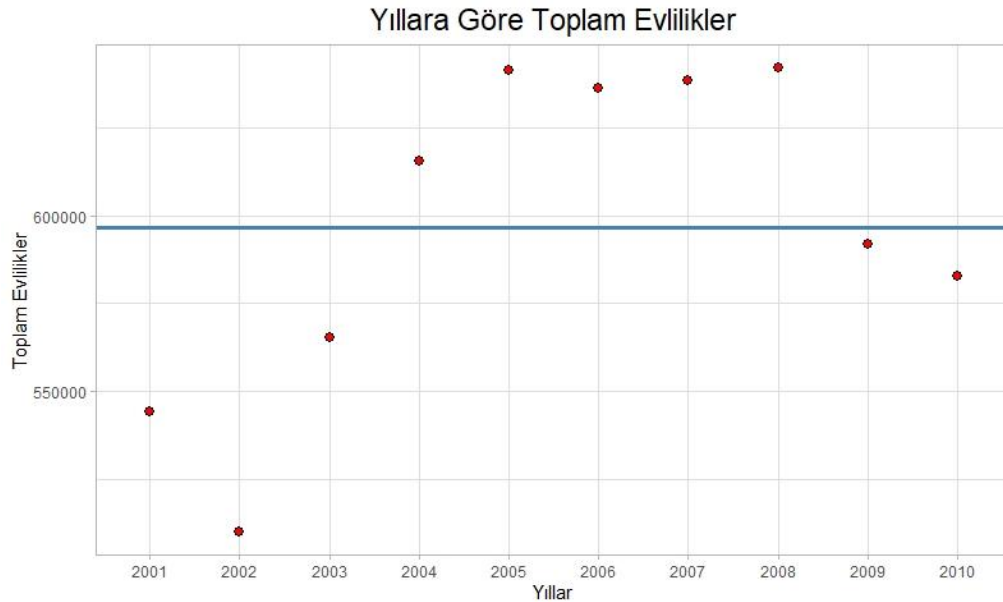
Tablo 5’te yer alan veriler incelendiğinde oluşturulan poisson regresyon modelinde p-değerlerinin 0,05’den küçük olduğu gözlemlenmektedir. Bu durum bağlamında verilerin istatistiksel olarak anlamlı olduğu ifade edilebilir. Bununla birlikte söz konusu katsayılar istatistiksel açıdan önemli bir katkı sunması nedeniyle modele dahil edilmektedir. 2003 yılından itibaren yılların modelde pozitif yönlü bir etkisi olduğu görülmektedir. 2008 yılındaki evlilik sayısındaki bir birimlik artışın modelde 0,17 birimlik artı yönlü bir artışa neden olduğu söylenebilir.

Tablo 6. Evliliğin Yaş ve Yıllar ile Birlikte Regresyonu Tablosu

	Tahmin	Standart Hata	z-değeri	p-değeri
(Intercept)	10,08160078	0,001658783	6077,71085	0,000000
Yas18-19	0,83283060	0,001658552	502,14325	0,000000
Yas20-24	2,11884311	0,001465454	1445,86146	0,000000
Yas25-29	1,93438682	0,001481291	1305,87880	0,000000
Yas30-34	0,84911317	0,001654488	513,21808	0,000000
35 Yaş ve Üzeri	0,6951860	0,001695232	410,0830	0,000000
2002	-0,06482638	0,001377921	-47,04651	0,000000
2003	0,03811272	0,001342682	28,38551	0,000000
2004	0,12266160	0,001315716	93,22802	0,000000
2005	0,16441784	0,001303027	126,18146	0,000000
2006	0,15584781	0,001305598	119,36890	0,000000
2007	0,15928464	0,001304565	122,09790	0,000000
2008	0,14951622	0,001304499	114,61580	0,000000
2009	0,06804070	0,001329598	51,17387	0,000000
2010	0,05266819	0,001334512	39,46626	0,000000

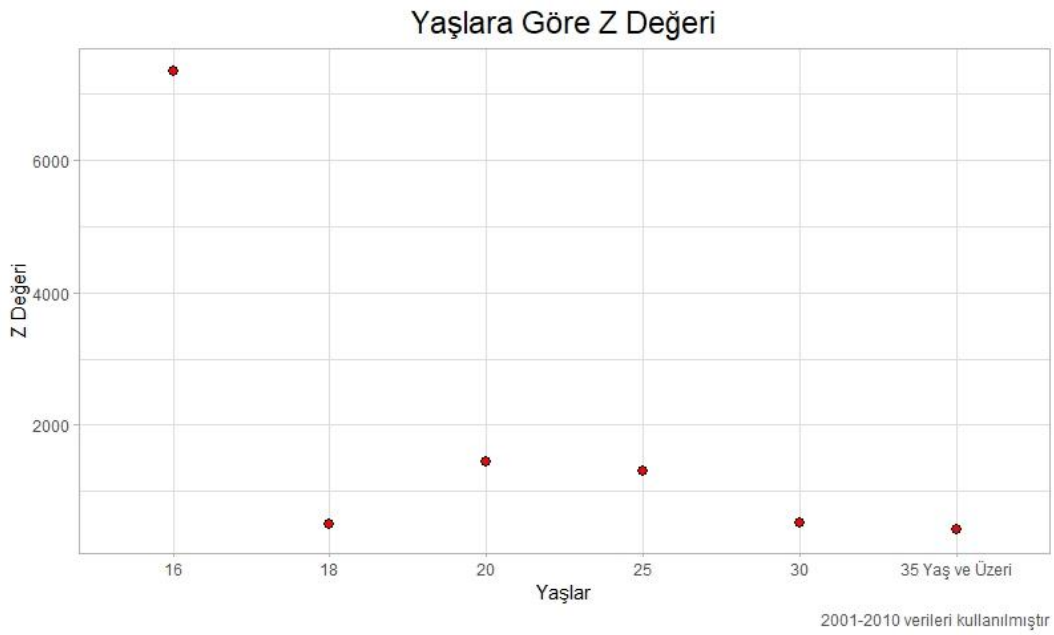
Tablo 6’da yer alan veriler incelendiğinde yaş gruplarına göre gerçekleşen evlilik sayılarında anlamlı bir farklılık olduğunu göstermektedir ($p<0,05$).

Poisson Regresyon uygulaması kapsamında yıllara göre gerçekleştirilen toplam evlilik sayıları Şekil 1’de gösterilmektedir.



Şekil 1. Yıllara Göre Toplam Evlilikler

Poisson Regresyon uygulaması sonucunda yaşlara göre elde edilen Z değerleri Şekil 2’de gösterilmektedir.



Şekil 2. Yaşlara Göre Z Değeri

Evlilikler çoğunlukla 30 yaş öncesinde gerçekleştiği için bu yaşlardaki z değerinin pozitif olduğu ve ortalamadan pozitif yönlü standart saptığı görülmektedir. Sonraki yaşlarda da evlilik sayısının azalmasından dolayı ortalamanın altında standart saptığı sonucuna varılmaktadır.

Bu uygulama kapsamında evlilik oranlarının yanı sıra boşanma oranları da yorumlanabilir. Bu bağlamda teknolojik gelişmeler ve çeşitli sosyo-kültürel, ekonomik ve politik etkiler itibarıyla toplumsal değerler değişiklik göstermektedir. Buna bağlı olarak ise yetişen yeni nesiller kendilerinden önceki nesillere göre farklı bir toplumsal yapıyı barındırmakta iken geçmiş nesiller geleneksel ve kültürel değerlere bağlılıklarını korumaktadırlar (Thornton, 1985: 859). Bu bağlamda evlilik ve boşanma oranlarının değerlendirilmesinde yaş gruplarının dikkate alınması önem arz etmektedir. Bununla birlikte evlilik yaşının boşanma üzerinde etkili olduğu sonucuna ulaşılan pek çok araştırma bulunmaktadır (Bumpass & Sweet, 1972; Booth & Edwards, 1985; Sanizah ve ark., 2014). Bu bağlamda araştırmalar sonucunda evlilik yaşının düşmesinin sonucu olarak toplumsal anlamda boşanma eğiliminin artış gösterdiği gözlemlenmektedir. Yapılan araştırmalar incelendiğinde kadınların erken yaşta evlenmesi ve erkeklerin kendilerinden genç kadınlarla evlenmesi durumunda evliliklerin sürdürülebilirliğinin güçleştiği gözlemlenmektedir. Yapılan bu araştırma kapsamında da kadınların evlilik yaşlarının erkeklerin evlilik yaşlarına göre daha düşük olduğu gözlemlenmektedir. Dolayısıyla Türkiye’de gerçekleşen erken evliliklerde cinsiyetin anlamlı bir etkisinin bulunduğu ifade edilebilir.

4. SONUÇ

Bu araştırmada TÜİK’ten elde edilen veriler doğrultusunda 2001-2010 yılları arasında gerçekleştirilen evlilikler, evlilik yaşları ve bunlara etken unsurlar araştırılmıştır. Bununla birlikte ilk olarak dünya geneli ile kıyaslama yapmak amacıyla 1990-2017 yılları arasında dünya genelinde farklı bölgelerden ülkelerde gerçekleştirilen evlilik sayıları ve evlilik yaşları incelenmiştir. Söz konusu veriler değerlendirildiğinde 1990 yılında erkekler ve kadınlar için ortalama evlilik yaşının en yüksek olduğu ülkenin Danimarka, 2000 yılında İzlanda ve 2017 yılında ise İsveç olduğu gözlemlenmiştir. Bu bağlamda söz konusu ülkeler arasında Kuzey Avrupa Bölgesinde bulunan ülkelerde ortalama evlilik yaşının en yüksek değerde olduğu çıkarımı yapılmıştır. Bununla birlikte

söz konusu veri grubu içerisinde ortalama evlilik yaşının 2017 yılında en düşük olduğu ülkenin Türkiye olduğu tespit edilmiştir.

TÜİK'ten elde edilen veriler doğrultusunda Türkiye'de gerçekleştirilen evlilik verileri kapsamında Poisson Regresyon modeli oluşturulmuştur. Oluşturulan modelde her bir değişkenin modele uygun olduğu ($p < 0,05$) tespit edilmiştir. Dolayısıyla herhangi bir değişken model kapsamından çıkartılmamıştır.

Oluşturulan Poisson Regresyon modeli değerlendirildiğinde erkeklerde en yüksek evlilik yaşının 25-29 yaş aralığında, kadınlarda ise 20-24 yaş aralığında gerçekleştiği görülmektedir. Bununla birlikte 20-24 ve 25-29 yaş gruplarında gerçekleştirilen evlilik sayısının 2010 yılına doğru ilerledikçe azalma gösterdiği gözlemlenmektedir. Bu bağlamda sosyo-kültürel, teknolojik, ekonomik ve politik etmenler doğrultusunda farklılık gösteren toplumsal yapının etkili olduğu çıkarımı yapılmaktadır. 45-49 yaş grubunda yer alan bireylerin evlilik sayılarının düşmesi ise bekârların az olması veya bekârlığa alışma durumu ile açıklanabilir. Genç yaş grubunda (18-25) yer alan erkeklerin eğitim hayatı, askerlik ve iş bulma kaygıları dolayısıyla evliliği erteledikleri ifade edilebilir.

Çalışma daha kapsamlı olacak şekilde 2010-2020 yılları arasındaki veriler ele alınarak tekrar değerlendirildiğinde toplam evlilik sayılarında azalma olduğu görülmekte olup, evlilik yaşlarına dair sonuçların da benzer olduğu sonucuna varılmıştır. “Poisson regresyon analizi sonuçlarına bakıldığında 25-29 yaş aralığında evlilik oranının erkeklerde daha yüksek olduğu, 20-24 yaş aralığında evlenen kadın sayısının da yüksek olduğu anlaşılmaktadır” (Altunbaş, 2021).

Yazarların Katkısı

Yazarların makaleye katkıları eşit orandadır.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

KAYNAKÇA

- Agresti, A. (2015). Foundations of linear and generalized linear models. *John Wiley & Sons*.
- Altunbaş, O. (2022). *Poisson regresyon analizi ve Türkiye'deki evlilik istatistiklerine uygulanması*. [Yüksek Lisans Tezi]. İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü. İstanbul.
- Aydemir, E. (2011). Evlilik mi evcilik mi? Erken ve zorla evlilikler: Çocuk gelinler. *Uluslararası Stratejik Araştırmalar Kurumu*. Uşak.
- Benz, L.S. & Lopez, M.J. (2021). Estimating the change in soccer's home advantage during the Covid-19 pandemic using bivariate Poisson regression. *AStA Advances in Statistical Analysis: A Journal of the German Statistical Society*. 1-28.
- Booth, A. & Edwards, J.N. (1985). Age at marriage and marital instability. *Journal of Marriage and the Family*, 47(1), 67-75.

- Bumpass, L.L. & Sweet, J.A. (1972). Differentials in marital instability: 1970. *American Sociological Review*, 37(6), 754-766.
- Cameron, A.C. & Trivedi, P.K. (2013). Regression analysis of count data (Vol.53). *Cambridge university press*, Cambridge, 598p.
- Consul, P. & Famoye, F. (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1), 89-109.
- Deniz Başar, Ö. (2005). Poisson regresyon analizi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 4(7), 59-72.
- Ercan, U. (2021). Hanehalklarının İletişim ve Bilgi Teknolojilerine Erişimi: Bir Poisson Regresyon Analizi. *Akdeniz İletisim*. 35, 402-422.
- Gao, G., Meng, S. & Shi, Y. (2021). Dispersion modelling of outstanding claims with double Poisson regression models. In *Insurance Mathematics and Economics November: Part B*. 101,572-586.
- Graff, C.A., Coffield, S.R., Chen, Y., Foufoula-Georgiou, E., Randerson, J.T. & Smyth, P. (2020). Forecasting daily wildfire activity using poisson regression. *IEEE Transactions on Geoscience and Remote Sensing IEEE Trans. Geosci. Remote Sensing Geoscience and Remote Sensing, IEEE Transactions on*. 58(7), 4837-4851.
- İşçi Güneri, Ö., Durmuş, B. & İncekırık, A. (2021). Comparison of some count models in case of excessive zeros: An application. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 20(40), 247-268.
- Jensen, R. & Thornton, R. (2003). Early female marriage in the developing world. *Gender & Development*, 11(2), 9-19.
- Kabacoff, R.I. (2015). R in Action (Second Edt.) Data analysis and graphics with R. *Manning*. Shelter Island.
- Sanizah, A., Hasfariza, F., Rahayu, S.N. & Nasliana, N.N. (2014). Determinants of marital dissolution: A survival analysis approach. *International Journal of economics and Statistics*, 2, 348-354.
- Thornton, A. (1985). Reciprocal influences of family and religion in a changing world. *Journal of Marriage and the Family*, 381-394.
- Vicuña, M. I., Vásquez C. & Quiroga B. F. (2021). Forecasting the 2020 COVID-19 Epidemic: A multivariate quasi-poisson regression to model the evolution of new cases in Chile. *Frontiers in Public Health*, 9, 1-7.
- UNICEF, (2013). Dünya Çocuklarının Durumu 2013 Raporu- Engelli Çocuklar. *UNICEF Türkiye*. Ankara.