



SİVAS CUMHURİYET ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Ekonometri Ana Bilim Dalı

**SAYMA VERİLERİ İLE REGRESYON ANALİZİ VE
TALEP VERİLERİNE UYGULANMASI**

Yüksek Lisans Tezi

Diclehan ŞEKER

Sivas
Eylül 2022

SİVAS CUMHURİYET ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Ekonometri Ana Bilim Dalı

**SAYMA VERİLERİ İLE REGRESYON ANALİZİ VE
TALEP VERİLERİNE UYGULANMASI**

Diclehan ŞEKER

Tez Danışmanı:
Prof. Dr. Hakan TRKAY

Sivas
Eyll 2022

ETİK İLKELERE UYGUNLUK BEYANI

Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü bünyesinde hazırladığım bu Yüksek Lisans tezinin bizzat tarafımdan ve kendi sözcüklerimle yazılmış orijinal bir çalışma olduğunu ve bu tezde;

- 1- Çeşitli yazarların çalışmalarından faydalandığımda bu çalışmaların ilgili bölümlerini doğru ve net biçimde göstererek yazarlara açık biçimde atıfta bulunduğumu;
- 2- Yazdığım metinlerin tamamı ya da sadece bir kısmı, daha önce herhangi bir yerde yayımlanmışsa bunu da açıkça ifade ederek gösterdiğimi;
- 3- Başkalarına ait alıntılanan tüm verileri (tablo, grafik, şekil vb. de dâhil olmak üzere) atıflarla belirttiğimi;
- 4- Başka yazarların kendi kelimeleriyle alıntıladığım metinlerini, tırnak içerisinde veya farklı dizerek verdiğim yine başka yazarlara ait olup fakat kendi sözcüklerimle ifade ettiğim hususları da istisnasız olarak kaynak göstererek belirttiğimi,

beyan ve bu etik ilkeleri ihlal etmiş olmam halinde bütün sonuçlarına katlanacağımı kabul ederim.

27/09/2022

Diclehan ŞEKER

TEŞEKKÜR

Lisans ve Yüksek Lisans eğitimim boyunca bilgi ve tecrübeleriyle bana yol gösteren ve destek olan, öğrencisi olduğum için gurur duyduğum değerli danışman hocam Prof. Dr. Hakan TÜRKAY'a

Hayatım boyunca maddi ve manevi destekleriyle her zaman yanımda olan babam Kerametdin ŞEKER'e, annem Mehtap ŞEKER'e ve Ablalarıma,

Katkı ve desteklerinden dolayı teşekkürlerimi sunarım.

Diclehan ŞEKER



İÇİNDEKİLER

İÇİNDEKİLER	i
KISALTMALAR	v
TABLO LİSTESİ	vii
ŞEKİL LİSTESİ	ix
ÖZET	xi
ABSTRACT	xiii
GİRİŞ	1
1. BÖLÜM	3
1. GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER	3
1.1. Genelleştirilmiş Doğrusal Modellerin Bileşenleri	4
1.2. Genelleştirilmiş Doğrusal Modellerin Yapısı	5
1.3. Model Parametre Tahmini	5
2. BÖLÜM	9
2. SAYMA VERİSİ İLE REGRESYON MODELLERİ	9
2.1. Poisson Regresyon Modeli	10
2.1.1 Poisson Dağılımı	10
2.1.2. Poisson Regresyon Modeli (PR)	11
2.1.3. Poisson Regresyon İçin En Çok Olabilirlik Tahmin Edici (Poisson-EÇO)	13
2.1.4 Aşırı Yayılım (Overdispersion)	13
2.1.4.1 Aşırı Yayılım Testleri	14
2.2 Negatif Binom Regresyon Modeli	18
2.2.1 Negatif Binom Dağılımı	18

2.2.2 Negatif Binom Regresyon Modeli (NBR)	19
2.2.3 Doğrusal Varyans Fonksiyonu İle NB1	22
2.2.4 Karesel Varyans Fonksiyonu İle NB2	23
2.2.5. Negatif Binom Regresyon İçin En Çok Olabilirlik Fonksiyonu.....	23
2.3. Genelleştirilmiş Poisson Regresyon Modeli (GPR)	24
2.4. Sıfır Yığılmalı Regresyon Modelleri (Zero-Inflated)	25
2.4.1 Sıfır Yığılmalı Poisson Regresyon Modeli (ZIPR).....	26
2.4.2 Sıfır Yığılmalı Negatif Binom Regresyon Modeli (ZINB).....	28
2.4.3 Sıfır Yığılmalı Genelleştirilmiş Poisson Regresyon Modeli (ZIGP).....	29
2.5. Hurdle Regresyon Modeli	31
2.5.1 Hurdle Poisson Regresyon Modeli	32
2.5.2 Hurdle Negatif Binom Regresyon Modeli	33
2.6. Literatür Taraması	34
3. BÖLÜM.....	37
3. MODEL SEÇİMİ	37
3.1. Uyum İyiği Ölçümü.....	37
3.1.1. Genelleştirilmiş Pearson Ki-Kare Testi	37
3.1.2. Sapma İstatistiği.....	38
3.1.3.Sözde R^2 Ölçümü (Pseudo R^2)	38
3.2 Model Seçim Kriterleri.....	39
3.2.1. Akaike Bilgi Kriteri (AIC).....	40
3.2.2. Bayesyen (Schwarz) Bilgi Kriteri (BIC).....	40
3.3. Katsayıların Yorumlanması	41
4. BÖLÜM.....	43
4.UYGULAMA.....	43

4.1. Birinci Veri Seti İçin Çalışma	43
4.1.1. Tanımlayıcı İstatistikler	43
4.1.2 Durağanlık Testi	46
4.1.3 Regresyon Analiz Sonuçları	47
4.1.3.1 Poisson Regresyon Analizi.....	47
4.2.3.2 Negatif Binom Regresyon Analizi	49
4.1.3.3 Genelleştirilmiş Poisson Regresyon Analizi	50
4.1.3.4 Model Seçimi	51
4.1.3.5. Uygun Model İçin Katsayı Yorumlama	51
4.2. İkinci Veri Seti İçin Çalışma	52
4.2.1. Tanımlayıcı İstatistikler	52
4.2.2. Durağanlık Testi	54
4.2.3. Regresyon Analiz Sonuçları	56
4.2.3.1. Poisson Regresyon Analizi.....	56
4.2.3.2 Negatif Binom Regresyon Analizi	57
4.2.3.3 Genelleştirilmiş Poisson Regresyon Analizi	58
4.2.3.4 Model Seçimi	59
4.2.3.5. Uygun Model İçin Katsayı Yorumlama	59
SONUÇ.....	61
KAYNAKÇA	63
ÖZ GEÇMİŞ.....	69



KISALTMALAR

OLS	: Ordinary Least Squares (Sıradan En Küçük Kareler)
GDM	: Genelleştirilmiş Doğrusal Modeller
EÇO	: En Çok Olabilirlik
PR	: Poisson Regresyon
EKK	: En Küçük Kareler
GPR	: Genelleştirilmiş Poisson Regresyon
NBR	: Negatif Binom Regresyon
ZIPR	: Sıfır Yığılmalı Poisson Regresyon
ZINB	: Sıfır Yığılmalı Negatif Binom
NBD	: Negatif Binom Dağılımı
NB	: Negatif Binom
ZIGP	: Sıfır Yığılmalı Genelleştirilmiş Poisson
PH	: Poisson Hurdle
NBH	: Negatif Binom Hurdle
ML	: Maksimum Likelihood
AIC	: Akaike Bilgi Kriteri
BIC	: Bayesyen (Schwarz) Bilgi Kriteri
GSYİH	: Gayrisafi Yurt içi Hâsıla
ASAE	: Araç Satın Alımı Endeksi
TCMB	: Türkiye Cumhuriyeti Merkez Bankası
EVDS	: Elektronik Veri Dağıtım Sistemi
ADF	: Augmented Dickey-Fuller
KFE	: Konut Fiyat Endeksi



TABLO LİSTESİ

Tablo 1.1. Doğrusal Modeller ile Genelleştirilmiş Doğrusal Modellerin Karşılaştırılması.....	4
Tablo 1.2. Üstel Dağılım Ailesinin Bazı Dağılımları İçin Bağ Fonksiyonları.....	5
Tablo 1.3. GDM'in Veri Türüne Göre Çeşitleri.....	8
Tablo 2.1. Negatif Binom Regresyon Modelleri.....	22
Tablo 2.2. Sayma Veri Regresyon Modelleri.....	25
Tablo 4.1. GSYİH Değişkenine Ait Tanımlayıcı İstatistikler.....	44
Tablo 4.2. Araç Satış Fiyat Endeksi (ASFE) Değişkenine Ait Tanımlayıcı İstatistikler.....	44
Tablo 4.3. Bağımlı Değişkene Ait Tanımlayıcı İstatistikler.....	45
Tablo 4.4. ADF (Durağanlık) Birim Kök Testi Sonuçları.....	46
Tablo 4.5. Kırılmalı Birim Kök Testi Sonuçları.....	47
Tablo 4.6. Poisson Regresyon Modeli Sonuçları.....	47
Tablo 4.7. Sapma Uyum İyiliği Sonucu.....	48
Tablo 4.8. Pearson Ki-Kare Uyum İyiliği Sonucu.....	48
Tablo 4.9. Negatif Binom Regresyon Modeli Sonuçları.....	49
Tablo 4.10. Genelleştirilmiş Poisson Regresyon Modeli Sonuçları.....	50
Tablo 4.11. Birinci Veri Seti İçin Model Seçim Kriterleri.....	51
Tablo 4.12. Katsayı Yorumları İçin Negatif Binom Regresyon Modeli Sonuçları...	51
Tablo 4.13. GSYİH Değişkenine Ait Tanımlayıcı İstatistikler.....	52
Tablo 4.14. Fiyat (Konut Fiyat Endeksi) Değişkenine Ait Tanımlayıcı İstatistikler	53
Tablo 4.15. Türkiye Konut Satış Sayısı Değişkenine Ait İstatistikler.....	54
Tablo 4.16. ADF (Durağanlık) Birim Kök Testi Sonuçları.....	55
Tablo 4.17. Kırılmalı Birim Kök Testi Sonuçları.....	55

Tablo 4.18. Poisson Regresyon Modeli Sonuçları	56
Tablo 4.19. Sapma Uyum İyiliği Sonucu	57
Tablo 4.20. Pearson Ki-Kare Uyum İyiliği Sonucu	57
Tablo 4.21. Negatif Binom Regresyon Modeli Sonuçları	57
Tablo 4.22. Genelleştirilmiş Poisson Regresyon Modeli Sonuçları	58
Tablo 4.23. İkinci Veri Seti İçin Model Seçim Kriterleri.....	59
Tablo 4.24. Katsayı Yorumu İçin Negatif Binom Regresyon Modeli Sonuçları	60



ŞEKİL LİSTESİ

Şekil 4.1. GSYİH (1998 yılı sabit fiyatları) Verisinin Dönemlere Göre Grafiği.....	44
Şekil 4.2. Araç Satın Alımı Fiyat Endeksi (ASAFE) Verisinin Dönemlere Göre Grafiği.....	45
Şekil 4.3. Otomobil Satış Sayısının Dönemlere Göre Grafiği.....	46
Şekil 4.4. GSYİH (1998 yılı sabit fiyatları) Verisinin Dönemlere Göre Grafiği.....	53
Şekil 4.5. Fiyat (Konut Fiyat Endeksi)Verisinin Dönemlere Göre Grafiği	53
Şekil 4.6. Türkiye Konut Satış Sayısının Dönemlere Göre Grafiği.....	54



ÖZET

Regresyon analizi iki veya daha fazla değişken arasındaki ilişkiyi açıklamak için kullanılan istatistiksel bir yöntemdir. Regresyon analizinde bağımlı değişkenin sayım yoluyla elde edilmesi durumunda sayma verileri ile regresyon modellerinin kullanılması gerekmektedir. Sayma verileri ile regresyon modellerinin temelini Poisson regresyon oluşturmaktadır.

Bu tez çalışmasında sayma verisi regresyon modellerinin tanıtılması ve otomobil talebi ile konut talebi verilerine uygulanması gerçekleştirilmiştir. Otomobil talebini ve konut talebini etkileyen faktörlerin açıklayıcı değişken olarak kullanıldığı regresyon modelleri kurularak çözümlenmiştir. Bu amaç doğrultusunda çalışmada ilk olarak genelleştirilmiş doğrusal modellere yer verilmiştir. İkinci olarak sayma verisi ile regresyon modellerinden Poisson regresyon, Negatif Binom regresyon, Genelleştirilmiş Poisson regresyon, Sıfır Yığılmalı regresyon ve Hurdle regresyon modellerinin yapısı ve işleyişleri açıklanmıştır. Bu modellere ek olarak aşırı yayılım durumundan da bahsedilmiştir. Daha sonra sayma verisi regresyon modelleri için kullanılan uyum iyiliği testleri ve bilgi kriterleri açıklanmıştır.

Çalışmanın uygulama kısmında talep verisi olarak otomobil satış ve konut satış sayıları gelir ve fiyat endekslerinin açıklayıcı değişken olarak kullanıldığı regresyon modelleri kurularak analizler yapılmıştır. Modelleme denemeleri STATA 16 ve EVIEWS 10 paket programları ile yapılmıştır. Tahmin edilen sonuçlara göre Negatif Binom regresyon modelinin hem otomobil satış sayıları hem de konut satış sayıları için en uygun model olduğu görülmüştür.

Anahtar Kelimeler: Sayma Verileri İle Regresyon, Sıfır Yığılmalı Regresyon, Poisson Regresyon, Hurdle Regresyon, Negatif Binom Regresyon



ABSTRACT

Regression analysis is a statistical method used to describe the relationship between two or more variables. In the regression analysis, counting data and regression models must be used if the dependent variable is obtained by counting. Poisson regression is the basis for counting data and regression models.

In this thesis study, the presentation of the counting data regression models and the application of the car request and housing request data was carried out. The regression models are established and analyzed by using the factors that affect the car demand and housing demand as descriptive variables. For this purpose, the study first included the generalized linear models. Secondly, the structure and operation of the Poisson regression, negative Binom regression, Generalized Poisson regression, Zero-Inflated regression and Hurdle regression models are explained by counting data and regression models of regression. In addition to these models, there is also a mention of the over-spread situation. The compatibility goodness-of-fit test and information criteria used for counting data regression models are then explained.

In the application of the study, the number of car sales and housing sales as claim data was established by establishing regression models where income and price indexes are used as descriptive variables. Modeling trials are done with STATA 16 and EVIEWS 10 package programs. According to the estimated results, the negative Binom regression model was found to be the best model for both car sales and housing sales numbers.

Keywords: Regression with Counting Data, Zero-Inflated Regression, Poisson Regression, Hurdle Regression, Negative Binomial Regression,



GİRİŞ

Regresyon analizi iki veya daha fazla değişken arasındaki ilişkiyi açıklamak için kullanılan istatistiksel bir yöntemdir. Bu yöntem ekonometrinin temelini oluşturmaktadır. Regresyon analizinde en önemli aşamalardan birisi regresyon modelinin seçimidir. Model seçilirken değişkenlerinin yapısı, sınırları ve toplanması gibi unsurlar göz önüne alınarak model seçimi yapılmaktadır. Değişkenler elde edilme şekillerine göre yapısı bakımından sürekli ve kesikli olmak üzere iki kısımdan oluşmaktadır. Sürekli yapıdaki değişkenler ölçüm yoluyla elde edilirken, kesikli yapıya sahip değişkenler sayım yoluyla elde edilmektedir. Sayım yoluyla elde edilen bu değişkenler için regresyon analiz yöntemlerinin özel bir türü olan sayma verisi modelleri kullanılmaktadır. Bu sebeple bu modeller için farklı çözüm yöntemleri gerekmektedir.

Sayma verisi ile regresyon modellerinde bağımlı değişkenin dağılımı dikkate alınarak analiz yapılmaktadır. Bağımlı değişkenin dağılımı Poisson ya da Negatif Binom dağılımlarından gelmektedir. Sayma verisine Poisson regresyon, Negatif Binom regresyon, Genelleştirilmiş Poisson regresyon, Sıfır Yığılmalı regresyon ve Hurdle regresyon modelleri uygulanabilir. Sayma verisi ile regresyon modellerinin temeli Poisson regresyona dayanmaktadır. Poisson regresyon bağımlı değişkenin Poisson dağılması sonucu kullanılabilir. Fakat burada dikkat edilmesi gereken husus Poisson regresyonun eşit yayılım ilkesidir. Eşit yayılımın sağlanamaması durumunda aşırı yayılım ya da yetersiz yayılım durumları ortaya çıkar. Bu durumlar sonucunda aşırı yayılım varsa Negatif Binom regresyon, yetersiz yayılım olması durumunda ise genelleştirilmiş Poisson regresyona başvurulmaktadır. Sayma verilerinde bazen bağımlı değişkenin gözlemlenmesi sonucunda beklenenden fazla sifıra rastlanabilir, bu durumda Sıfır Yığılmalı regresyon ve Hurdle regresyon modellerinin kullanılması daha uygun olacaktır.

Talep piyasada belirli bir ürüne yönelik olarak ortaya çıkan satın alma isteğidir. Toplumda talepten söz edilirken ürünü satın almak için gerekli paraya sahip olan ve gerekli paraya sahip olmayan ancak sadece satın almayı isteyen kişiler de bulunur. Diğer taraftan satın alma isteği, satın alma gücü ile desteklenmediği sürece

talep sayılmaz. Bu nedenle ekonomik anlamda satın alma isteğinin satın alma gücü ile desteklenmesi gerekmektedir. Belirli bir ürününden talep edilen miktar, ürünün fiyatının yanı sıra tüketicinin gelirine, tercihlerine, nüfusa ve gelir dağılımına vb. gibi faktörlere bağlıdır. Çalışmada talep verisi olarak otomobil talebi ve konut talebi ele alınmıştır. Otomobil ve konut talebini mevcut otomobil ve konut fiyatları, Gayri Safi Yurt İçi Hâsıla (GSYİH), hane halkı tercihleri, faiz oranları ve gelecekteki beklenen fiyatlar gibi birçok ekonomik faktör etkilemektedir. Otomobil ve konut talebindeki eğilimleri görebilmek için bahsettiğimiz ekonomik değişkenleri dikkate almamız gerekmektedir.

Bu tez çalışmasında, otomobil satış sayısı ve konut satış sayısı talep verileri ile bu talep verilerini etkileyen ekonomik faktörlerin arasındaki ilişkilerin, sayma verisi regresyon modelleriyle açıklanması ile bu modellerin tanıtılması amaçlanmıştır. Bu amaç doğrultusunda, çalışma 4 bölümden oluşmaktadır. Birinci bölümde sayma verileri ile regresyon analizinin daha kolay anlaşılması için geliştirilmiş doğrusal modellere yer verilmiştir. İkinci bölümde sayma verisi ile regresyon modelleri yapısı ve işleyişi açıklanmıştır. Üçüncü bölümde sayma verisi ile regresyon analizi için model seçim kriterleri ve uyum iyiliği ölçütleri hakkında bilgi verilmiş olup son olarak dördüncü bölümde ise otomobil satış sayısı ve konut satış sayısı için 2 ayrı veri seti oluşturulmuş ve bu veri setleri hakkında istatistiksel bilgiler verilmiştir. Bağımlı değişkenin dağılımı da dikkate alınarak sayma veri modelleri tahmin edilerek uygun modeli seçebilmek adına uyum iyiliği testleri ve model seçim kriterleri hesaplanmıştır. Son olarak uygun modeller için tahmin edilen parametreler yorumlanmıştır.

1. BÖLÜM

1.GENELLEŞTİRİLMİŞ DOĞRUSAL MODELLER

Genelleştirilmiş doğrusal modeller (GDM), doğrusal modellerin bir uzantısıdır. Doğrusal modellerde regresyon analizi belirli varsayımlar altında gerçekleşmektedir. Bu varsayımlar hata teriminin normal dağılımlı olması, modelin doğrusal olması, eş varyanslı olması gibi birçok varsayımdan oluşmaktadır. Ancak kullanılan verilerin hata terimlerinin normal dağılmaması, doğrusal olmaması gibi durumlarda doğrusal modellerin varsayımları dışına çıkılması sonucunda modellemede bazı hatalar ve sorunlar ortaya çıkmaktadır. Bu durumda tahminlerde sapmalar ve tutarsız sonuçlar gibi durumlarla karşılaşılması sonucunda en küçük kareler yöntemi (EKK) ile analizin yapılması uygun olamamaktadır.

Genelleştirilmiş doğrusal modeller ilk olarak Nelder ve Wedderburn (1972) çalışmasında ortaya atılmıştır. İlerleyen süreçte ise Nelder (1977), Mccullagh ve Nelder (1989) modelin geniş bir araştırmasını yapmışlardır.

Genelleştirilmiş doğrusal modeller (GDM) en çok olabilirlik yöntemi ile tahmin edilir. GDM bağımlı değişkenin dağılış biçimini dikkate alarak ortalamanın yerine bir dönüşümünün veya fonksiyonunun kullanılması temeline dayanarak işlem yapar. Genelleştirilmiş doğrusal modellerde hataların normal dağıldığı varsayımı kullanılmaz.

Genelleştirilmiş doğrusal modellerde tahmin dağılımının olasılık yoğunluk fonksiyonunu üstel (exponential) formunda tanımlar. Sonra verilerin beklenen değerleri değil de beklenen değerlerinin doğrusal kombinasyonunu kullanarak tahmin yapar. Burada $E(y_i) = \mu$ ifadesinin yerine, $\eta = X\beta$ kullanılır ve $\eta = \log E(y_i)$ olur. η doğrusal tahminci olup $\log E(y_i)$ modelle beklenen değer arasında bir bağlantı fonksiyonudur (Gökdere, Akkol 2004: 1).

Doğrusal modelle genelleştirilmiş doğrusal modeller arasındaki farkı daha net ortaya koymak amaçlı aşağıdaki tabloda düzenlenmiştir.

Tablo 1.1. Doğrusal Modeller ile Genelleştirilmiş Doğrusal Modellerin Karşılaştırılması

	Doğrusal Model	Genelleştirilmiş Doğrusal Model
Yanıt değişkeni veya Hata terimi dağılımı	Normal dağıldığı varsayılır.	Yanıt(Bağımlı) değişkeni dağılımı üstel dağılım ailesine ait herhangi bir dağılıma sahiptir.
Varyans	Sabit varyans	Değişen varyans
Ortalama Dönüşüm	Kullanılmaz	Kullanılabilir
Bağ Fonksiyonu	Birim	Birim, logaritmik, üstel, ters, karekök, logit

Kaynak: (Karadağ 2014)

Genelleştirilmiş doğrusal modeller 3 temel bileşene dayanmaktadır. Modele geçişte bu 3 bileşen üzerinden işlem yapılmaktadır (Nelder, Wedderburn 1972: 372). Bu bileşenler;

- Rasgele bileşen
- Sistematik bileşen
- Bağ fonksiyonu

Bu bileşenleri ayrıntılı olarak bir alt başlıkta ele alınmıştır.

1.1.Genelleştirilmiş Doğrusal Modellerin Bileşenleri

1. Rasgele bileşen: Y'nin bileşenleri $E(Y) = \mu$ ortalama ve (σ^2) sabit varyans ile normal dağılıma sahiptir.

2. Sistematik bileşen (Doğrusal Bileşen): $x_1, \dots, x_p$ değişkenlerinin birleşimi sonucu η doğrusal bağımsız değişkenleri elde ederiz:

$$\eta = \sum_{j=1}^p x_j \beta_j \quad (1.1)$$

Sistematik bileşen dışsal değişkenler arasındaki ilişkinin doğrusal bir birleşimi olup bağ fonksiyonu ile ortalamayı ilişkilendirilir.

3. Bağ fonksiyonu: Rasgele bileşen ve sistematik bileşen arasındaki ilişki $\mu = \eta$ genellemesi ile gösterilir.

$$\eta = g(\mu_i) = X\beta \quad (1.2)$$

η : doğrusal ön kestirim, μ : y rasgele değişkeninin beklenen değerini ifade etmektedir.

Rasgele bileşen ile sistematik bileşenin kombinasyonu sonucunda bağ fonksiyonunu elde ederiz. Burada $g(\cdot)$ bağlantı fonksiyonudur. Bağ fonksiyonu birim fonksiyona sahip olan doğrusal modellerde ortalama dönüşümüne imkân verilmediğinden direk olarak ortalama alınır. Genelleştirilmiş doğrusal modellerde ise fonksiyon açısından çeşitlendirilmiş olup tek koşulu türevlenebilir bir fonksiyon olmasıdır. Bağ fonksiyonları birim, logit, karekök gibi çeşitlere ayrılır.

Tablo 1.2. Üstel Dağılım Ailesinin Bazı Dağılımları İçin Bağ Fonksiyonları

Dağılım	Bağ Fonksiyonu	$g(\mu)$
Normal	Birim	μ_i
Poisson	Log	$\ln(\mu_i)$
Binom	Logit	$\ln(\mu_i / (1-\mu_i))$
Gamma	Ters	$1/\mu_i^p$

1.2.Genelleştirilmiş Doğrusal Modellerin Yapısı

Y'nin bileşenlerinin herhangi bir üstel dağılım ailesinden geldiği düşünüldüğünde olasılık yoğunluk fonksiyonu aşağıdaki gibidir.

$$f_y(y; \theta, \phi) = \exp \left\{ \frac{y(\theta) - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (1.3)$$

Burada ϕ yayılım (ölçek) parametresi, θ kanonik (doğal) parametresidir.

ϕ yayılım parametresi, θ kanonik parametresi ve $c(y, \phi)$ üstel dağılımın herhangi bir dağılımına göre düzenlenerek kullanılabilir.

Eğer ϕ biliniyorsa, dağılım θ parametrelili üstel dağılım ailesidir. Eğer ϕ bilinmiyorsa, dağılım iki parametrelili üstel dağılım ailesi olabilir (Dinarcan 2018: 5).

1.3.Model Parametre Tahmini

Modellere ait parametrelerin tahmini yapılırken farklı yöntemlere başvurulabilir. Yanıt değişkeninin farklı dağılımlara sahip olduğu durumlarda parametre tahmini için en yaygın olan En Çok Olabilirlik (EÇO) yöntemi kullanılır.

En çok olabilirlik yöntemi ekonometri ve istatistik alanlarında en yoğun kullanılan nokta tahmin yöntemlerinden biridir.

“Olabilirlik fonksiyonu; $f(y, \theta)$ olarak gösterilmektedir. Ancak uygulamadaki kolaylığı nedeniyle olabilirlik fonksiyonu yerine, log-olabilirlik fonksiyonu kullanılmaktadır.” (Dinarcan, 2018: 5).

Genelleştirilmiş modellerde olabilirlik fonksiyonu,

$$L(\beta) = \prod_{i=1}^n \left(\frac{y_i \theta - b(\theta)}{a(\theta)} \right) + c(y, \theta) \quad (1.4)$$

Eşitlik (1.4)’de verildiği gibidir. Modelin log olabilirlik fonksiyonu, $l = l(y; \theta, \emptyset) = \ln L(y; \theta, \emptyset)$ kabul edilirse Log olabilirlik fonksiyonu,

$$l(\beta) = \sum_{i=1}^n l_i(\beta) = \sum_{i=1}^n \left(\frac{y_i \theta - b(\theta)}{a(\theta)} + c(y, \theta) \right) \quad (1.5)$$

$i = 1, 2, \dots, n$ şeklindedir.

En çok olabilirlik tahmin yöntemiyle işlem yaparken dağılım bilgisine ihtiyaç duyulur. Eğer dağılım bilgisi yoksa EKK ve Yarı EÇO kullanılır. En Çok Olabilirlik yönteminde β katsayısını bulabilmek için iteratif yöntemler gerekir. *Fisher Skoru* bu yöntemlerden birisidir.

Genelleştirilmiş doğrusal modellerde beklenen değerin ve varyansın genel formuna Fisher Skoru ile ulaşılabilir.

l' nin θ ’ ye göre birinci türevinin alınması ile ortalama,

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\theta)} \quad (1.6)$$

İkinci türevinin alınmasıyla da varyansı,

$$\frac{\partial^2 l}{\partial \theta^2} = - \frac{b''(\theta)}{a(\theta)} \quad (1.7)$$

elde edilir. Eşitlik (1.6)’ yi sıfıra eşitlediğimizde $a(\emptyset) = 1$ olması şartıyla,

Fisher Skorlama ile

$$E \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad (1.8)$$

$$-E \left(\frac{\partial^2 l}{\partial \theta^2} \right) = E \left(\frac{\partial l}{\partial \theta} \right)^2 \quad (1.9)$$

elde edilir. Eşitlik (1.6)'yı alıp Fisher skarlama ile elde ettiğimiz formülde yerine yazarsak;

$$E \left(\frac{y-b'(\theta)}{a(\emptyset)} \right) = 0 \quad (1.10)$$

$$E(y) - b'(\theta) = 0 \quad (1.11)$$

Buradan genelleştirilmiş modellerin beklenen değerin genel formunu

$$E(y) = b'(\theta) = \mu \quad (1.12)$$

elde ederiz. Aynı şekilde eşitlik (1.6) ve eşitlik (1.7)'yi alıp Fisher skarlama ile elde ettiğimiz formülde yerine yazalım.

$$E \left(-\frac{b''(\theta)}{a(\emptyset)} \right) = E \left(\frac{y-b'(\theta)}{a(\emptyset)} \right)^2 \quad (1.11)$$

$$\frac{-b''(\theta)}{a(\emptyset)} = \frac{V(y)}{a(\emptyset)^2} \quad (1.12)$$

Böylece genelleştirilmiş doğrusal modellerin varyansının genel formuna ulaşmış oluruz.

$$V(y) = a(\emptyset)b''(\theta) = \frac{\partial E(y)}{\partial n_i} = \frac{\partial \mu}{\partial \theta} \quad (1.13)$$

Buradan varyansın, ortalamanın bir fonksiyonu olduğu sonucuna ulaşılır.

$a(\emptyset)$ fonksiyonunun genel olarak gösterimi;

$$a(\emptyset) = \emptyset/w \quad (1.14)$$

şeklindedir. Burada $\emptyset$ parametresi, σ^2 olarak gösterilen yayılım parametresi; w ise gözlemden gözleme değişen önsel ağırlıktır (Dinarcan 2018: 7).

Genelleştirilmiş doğrusal modellemesinde veri türüne göre çeşitlerine ayrılmaktadır. Sürekli veriler, kesikli veriler(sayma verisi için), kategorik veriler gibi. Aşağıda veri türüne göre GDM çeşitleri verilmektedir.

Tablo 1.3. GDM'in Veri Türüne Göre Çeşitleri

Sayma Veri İçin GDM	Sürekli Veri İçin GDM	Kategorik Veri İçin GDM
Poisson Regresyon Modeli	Lojistik regresyon Modeli	Gamma Regresyon
Negatif Binom Regresyon Modeli	Probit Regresyon Modeli	Ters Gauss Regresyon
Aşırı Yayılım Regresyon Modeli		Tweedie Regresyon
Sıfır Yığılmalı Regresyon Modelleri		Çoklu Regresyon
Genelleştirilmiş Poisson Modeli		Varyans Analizi Covaryans Analizi

2. BÖLÜM

2.SAYMA VERİSİ İLE REGRESYON MODELLERİ

Sayma veri modellerine başlamadan önce sayma verilerinin ne anlama geldiğini anlamak önemlidir. "Sayma" kelimesi genellikle birimleri, öğeleri veya olayları numaralandırmak için kullanılır. Sayma verileri ise, sayma yoluyla ifade edilebilen olaylar veya öğeler hakkında yapılan gözlemleri ifade etmek için kullanılır (Kılıç, Bayrak 2020 : 86). Sayma veri olayların ve öğelerin gerçekleşme sayısı ile ilgilenir.

Sayısal veriler elde edilme şekillerine göre yapı bakımından sürekli ve kesikli olmak üzere iki kısma ayrılır. Sürekli yapıya sahip veriler ölçüm yoluyla elde edilirken, kesikli yapıya sahip veriler sayım yoluyla elde edilmektedir. Bu sebeple kesikli yapıdaki sayılabilen veriler sayma verileri olarak adlandırılmaktadır. Kesikli yapıdaki değişkenlere örnek verecek olursak, bir hafta içindeki markete gitme sayısı, gün içinde hastaneye gelen hasta sayısı, belirli bir rotada trende yolculuk eden kişi sayısı gibi verileri ele alabiliriz.

“Sayma veri modelleri regresyon modellerinin özel bir türüdür. Bu nedenle bu modeller için farklı çözüm yöntemleri gerekmektedir. Poisson ve Negatif Binom regresyon modelleri sıklıkla kullanılan yöntemlerdir. Her iki yöntemde de bilinen regresyon modellerinde olduğu gibi bağımsız değişkenler ile sayım veri niteliğindeki bağımlı değişken arasındaki ilişkiyi araştırmak amaçlanmaktadır.” (Cameron, K.Trivedi 1998: 1-8).

Sayma verilerinin modellenmesi sürecinde bağımlı değişkenin dağılımı büyük önem gösterir. Bağımlı değişkenin sahip olduğu dağılım Poisson Ve Negatif Binom dağılımlarından gelmekte ve kullanılacak yöntemin belirlenmesinde dağılımın da etkisi göze alındığında sayma verilerinin modellenmesinde Poisson Ve Negatif Binom dağılımları üzerinde durulmuştur.

Sayma verileri genellikle çok fazla dağılır, bu da hesaplanan Poisson maksimum olabilirlik t değerlerinin önemli ölçüde aşırı şişirilmesine neden olur. Bu,

regresörlerin (ortak değişkenlerin) istatistiksel önemi konusunda çok hatalı ve aşırı iyimser sonuçlara yol açabilir (Cameron, Trivedi 1998: 68). Sayma verilerinin olduğu durumlarda dağılım genellikle sağa çarpıktır. Bu durum sonucunda normallik varsayımının ihlali söz konusu olduğundan klasik regresyon yöntemlerini kullanamayız. Eğer bu durum göz ardı edilip klasik yöntemler kullanılırsa yanlış, tutarsız ve etkin olmayan tahminlerle karşılaşabiliriz.

Sayma regresyon modellerin her biri Poisson modeline dayanmaktadır (Long 1997: 217-218). Bu bölümde sayma verisinin modellenmesi için kullanılan regresyon modellerine değinilecektir.

2.1.Poisson Regresyon Modeli

2.1.1 Poisson Dağılımı

Dağılım ilk olarak Fransız matematikçi Siméon-Denis Poisson tarafından 1938’de ortaya çıkartılmıştır. Poisson dağılımı belirli bir zaman aralığında olayların meydana gelme sayısını modellemek için kullanılan tek parametrelili bir dağılım türüdür.

Y rasgele değişkeni için poisson dağılımı olasılık fonksiyonu $\mu > 0$ olmak üzere,

$$P(Y; \mu) = \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots \quad (2.1)$$

şeklindedir.

μ birim zamanda meydana gelmesi beklenen ortalama oluş sayısını, y belli bir zamanda meydana gelen olay sayısını ifade etmektedir. Poisson dağılımının şartlı beklenen değeri $E(y_i) = \mu_i$ ’dir. μ_i ’nin açıklayıcı değişkenlere bağlı olarak tanımlanması durumunda Poisson regresyon modeli elde edilmektedir. μ_i genellikle $\mu_i = e^{x'\beta}$ şeklinde tanımlanır. Burada x açıklayıcı değişken vektörü ve β ise tahmin edilecek parametre vektörünü göstermektedir (Memiş, Önder 2018: 142).

Poisson dağılımının özellikleri aşağıda sunulmuştur.

- ✓ Olaylar birbirinden bağımsızdır.
- ✓ Nadir görülür.

- ✓ Ortalama ve varyans eşittir. (Eşit Yayılım)

$$E[Y] = \mu, \quad V[Y] = \mu \quad (2.2)$$

- ✓ Belirli bir zaman aralığında gerçekleşir.
- ✓ Sağa çarpık bir dağılım olması sebebiyle dağılım pozitifdir.
- ✓ Poisson dağılımı tek parametrelidir.
- ✓ Dağılımın parametresinin sıfırdan büyük olma şartı mevcuttur. ($\mu > 0$) (Avcı 2020: 30).

2.1.2. Poisson Regresyon Modeli (PR)

Poisson regresyon modeli sayma verileri ile modellemenin temelini oluşturmakta olup sayma verisi ile regresyon için sıklıkla kullanılan en basit yöntemdir. Modelde bağımlı değişkenin sayma verisi olması değişkenin dağılımı ile belirlenir. PR, bağımsız değişken ile bağımlı değişken arasındaki ilişkiyi ele alırken bağımlı değişkeni sayıma dayalı veriler kullanarak analiz işlemini yapar.

Poisson regresyon analizinde Y bağımlı değişkeni küçük değerlerde yoğunlaşan, negatif olmayan Y: 0,1,2, şeklinde tam sayı değerleri almaktadır (İşbilen 2021: 9).

Sayma veriler için standart bir model olan Poisson regresyon modeli doğrusal olmayan bir regresyon modelidir. Bu regresyon modeli yoğunluk parametresi μ 'nin ortak değişkenlere bağlı olmasına izin vererek Poisson dağılımından türetilmiştir (Cameron, Trivedi 1998: 9).

Poisson dağılımından türetilen PR modeli :

$$P(y_i|x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad i = 1, 2, \dots \quad (2.3)$$

ve koşullu ortalama parametresi,

$$E(y_i|x_i) = \exp(x_i'\beta) = \mu_i \quad (2.4)$$

şeklinde gösterilir ve “üstel ortalama fonksiyonu” olarak ifade edilir. İstatistik literatüründe bu fonksiyon ayrıca; “log-doğrusal fonksiyon” olarak da ifade edilir. Çünkü koşullu ortalamanın logaritması, parametreleri doğrusal olarak vermektedir (Deniz 2005: 61).

Eşitlik (2.4)'teki ortalama fonksiyonunu şu şekilde yazarsak,

$$\mu_i = e^{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n} \quad (2.5)$$

X'ler artığında μ_i ' lerde üstel bir şekilde artmaktadır. Yukarıdaki eşitliği doğrusal bir hale getirirsek,

$$\ln(\mu_i) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n \quad (2.6)$$

Yukarıda da görüldüğü gibi koşullu ortalamanın(μ) logaritmik dönüşümü ile denklem doğrusal bir hale gelmektedir.

$x\beta$ 'nin $\exp(x_i'\beta)$ olarak alınması, beklenen sayma değerinin pozitif olmasını gerektirmektedir (Selim, Üçdörük 2003: 15).

$$E(y_i|x_i) = V(y_i|x_i) = \mu_i \quad (2.7)$$

Bu özellik doğrultusunda varyansın sabit olmadığı ve heteroskadezik (değişken) olduğu anlaşılmaktadır. Karma Poisson dağılımlarında Poisson dağılımının parametresi (μ), gözlemler arası değişkenlik göstermektedir ve bu değişkenlik gözlemlenmemiş değişkenlik (unobserved heterogeneity) terimi ile açıklanmaktadır. (Tüzel, Sucu 2012: 25). Negatif binom regresyon bir karma poisson yaklaşımıdır.

Dağılım Eşitlik (2.7) teki gibi ortalama ve varyans eşitliğine dayanmaktadır.(Eşit yayılım). Poisson modelinin en önemli özelliklerinden olan bu eşitlik ilkesi, uygulama aşamasında çoğunlukla sağlanamaz. Bu durumda sayım değerlerinin ortalamadan büyük varyansa (aşırı yayılım, overdispersion) ya da ortalamadan küçük varyansa (yetersiz yayılım, underdispersion) sahip olması gibi durumlar ortaya çıkarır.

EKK yönteminin kendine özgü varsayımları olduğu gibi En Çok Olabilirlik Tahmin yöntemini kullanan poisson regresyon modelinin de kendisine özgü bazı varsayımları bulunmaktadır. Model bu varsayımlar altında ele alınır.

Modelin en çok olabilirlik tahmini için varsayımları;

- Koşullu ortalamanın doğru tanımlanmasında bağımlılık şartı sağlanmalıdır. Ayrıca bağımlı değişken y 'nin Poisson dağılması gereklidir.

- En çok olabilirlik standart hataları ve t istatistikleri kullanılarak hesaplanan istatistiksel sonuçlar, hem koşullu ortalama, hem varyansın doğru tanımlanmasını gerektirmektedir. Burada istenen koşul, koşullu varyans ve ortalamanın eşit olmasıdır.
- Veriler için koşullu varyans ve koşullu ortalamanın eşit olmaması durumunda, en çok olabilirlik yönteminin uygulanması ile elde edilmiş istatistiksel sonuçlar, koşullu ortalamanın doğru tanımlandığının ispat edildiği durumlarda geçerli ve doğrudur.
- Veriler için koşullu varyans ve ortalamanın eşit olmaması durumunda, Poisson en çok olabilirlik tahmin edicisinden daha etkin tahmin ediciler kullanılabilir. (Deniz 2005: 61)

2.1.3. Poisson Regresyon İçin En Çok Olabilirlik Tahmin Edici (Poisson-EÇO)

En çok olabilirlik (maximum likelihood)fonksiyonu,

$$\ln L(\beta) = \sum_{i=1}^n \{y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!\} \quad (2.8)$$

$\exp(x_i' \beta) = \mu_i$, $\hat{\beta}$ tahmincisine göre birinci dereceden türevi alınarak sıfıra eşitlenir. (Cameron ve K. Trivedi 1998: 61)

$$\sum_{i=1}^n (y_i - \exp(x_i' \beta)) x_i = 0 \quad (2.9)$$

Ortalamanın üstel olması sebebi ile Log-olabilirlik fonksiyonu doğrusal değildir. Bu durumda β 'lar için EÇO yöntemi ile tek adımda çözüme ulaşamayabiliriz. Çözüme ulaşabilmek için iteratif yöntem olan Newton-Rahpson yöntemi kullanılmaktadır.

2.1.4 Aşırı Yayılım (Overdispersion)

Poisson regresyon analizinde, bağımlı değişkenin Poisson dağılıma dayanması sebebiyle dağılımın, en önemli özelliklerinden biri olan ortalama ve varyansın eşitlik ilkesidir. Ancak bu eşitlik ilkesini uygulamada sağlamak bazen mümkün olamamaktadır. Bu durumda veriler ortalamadan daha büyük varyansa sahip olduğunda aşırı yayılım (overdispersion) göstermektedir.

Aşırı yayılım durumu göz ardı edilerek parametre tahminleri yapılması sonucunda standart hataların olduğundan daha küçük olması durumu ortaya çıkacaktır. Bu durum sonucunda standart hatalar yanlış olacak ve büyük ölçüde şişirilmiş t istatistikleri hatalı çıkacaktır. Bu da açıklayıcı değişkenlerin hakkında istatistiksel olarak yanlış çıkarımlar yapmamıza sebep olacaktır.

“Aşırı yayılım sorunu genellikle veride kümelenmeler var ise ortaya çıkmaktadır. Kümelenme neticesinde veride sıfır enflasyonu denen durum (veride peş peşe sıfırlar, uzun boşluklar) ile birlikte aşırı yayılıma sebep olan diğer hususlar; eksik tanımlama (bağımlı değişkeni açıklamada önemli olan açıklayıcı değişkenlerin modele alınmamış olmaları, sabit terimin modele alınmamış olması) veya modelin doğrusal olmadığı halde doğrusal gibi tahmin yapılmasından kaynaklanmış olabilir.” (İşbilen 2021: 13).

Aşırı yayılım durumunda Poisson Regresyon Modelinden elde edilen tahminler tutarlı fakat etkin değildir (Arı, Önder 2013: 169). Aşırı yayılım ve yetersiz yayılım ile karşılaşılması durumunda PR modeli yerine Negatif Binom regresyon veya Genelleştirilmiş Poisson regresyon kullanılabilir. GPR (Genelleştirilmiş Poisson Regresyon) modeli hem aşırı yayılım hem de yetersiz dağılım durumunda PR modelinin alternatifidir. NBR modeli ise aşırı yayılım durumunda PR modeline göre daha büyük varyans ile baş edebilir.

2.1.4.1 Aşırı Yayılım Testleri

Verilerin aşırı yayılım ya da yetersiz yayılıma sahip olup olmadığını belirlemek için Poisson modelinin varyansı ortalamasına oranlanarak yayılım parametresi elde edilir. Elde edilen parametre değeri eğer 1’den büyükse aşırı yayılım, 1’den küçükse de yetersiz dağılım durumu söz konusudur.

Aşırı yayılımı tespit etmek için bazı test istatistikleri kullanılmaktadır. Genellikle aşırı yayılım tespitinde Cameron-Trivedi testi kullanılmaktadır (Cameron, Trivedi 1990: 347-364).

Cameron-Trivedi testinin yokluk hipotezi ve alternatif hipotezi;

$$H_0: \alpha = 0$$

$$H_1: \alpha \neq 0$$

şeklinde kurulmaktadır.

$$V(y_i|x_i) = \mu_i + \alpha g(\mu_i) \quad (2.10)$$

Burada α bilinmeyen bir parametre olup, $g(\cdot)$ bilinen bir fonksiyondur. $g(\mu)=\mu$ veya $g(\mu)=\mu^2$ olarak kabul edilmektedir. Böylece $\alpha=0$ olduğunda $V(y_i|x_i) = \mu_i$ olur. Bu sebeple hipotezler yukarıda gösterildiği gibi olacaktır.

$\hat{\mu} = \exp(x_i'\beta)$ olmak üzere;

$$\frac{(y_i - \mu_i)^2 - y_i}{\mu_i} = \alpha \frac{g(\mu_i)}{\mu_i} + u_i \quad (2.11)$$

u_i hata terimi, klasik EKK tahmini yapıldığında elde ettiğimiz t istatistikleri eğer normalliği sağlarsa H_0 hipotezini kabul ederiz. Yani aşırı yayılım yoktur. Bu durumda PR analizi kullanılabilir. Ancak H_0 hipotezi reddedilirse $\alpha \neq 0$ olacağından poisson dağılmayacaktır. Eşitlik (2.11) sonucunda varyans ortalamayı aşacaktır. Bu sebeple PR analizi yerine NBR analizi tercih edilir

i. Skor testi

Aşırı yayılımın belirlenmesinde kullanılabilen testlerden biri poisson yaklaşımının test edilmesi için geliştirilen skor testidir. Test poisson yaklaşımın aşırı yayılıma sahip olup olmadığını belirler.

Büyük örneklem için kısmi skor test istatistiği;

$$T = \frac{1}{2} \sum_{i=1}^n \{(Y_i - \hat{\mu}_i)^2 - Y_i\} \quad (2.12)$$

şeklindedir.

T istatistiğinin standartlaştırılmış hali olarak gösterilen T_1 test istatistiği;

$$T_1 = \frac{\sum_{i=1}^n \{(Y_i - \hat{\mu}_i)^2 - Y_i\}}{(2 \sum_{i=1}^n \hat{\mu}_i^2)^{1/2}} \quad (2.13)$$

şeklinde gösterilmektedir.

Burada $\hat{\mu}_i$ poisson yaklaşımı altında μ_i 'nin en çok olabilirlik tahmin edicisidir.

“T istatistiğinin pozitif büyük değer alması verinin Poisson regresyon modeline göre aşırı yayılım, negatif büyük değer alması ise verinin Poisson

regresyon modeline göre az yayılım gösterdiğini ifade etmektedir.” (Tüzel ve Sucu, 2012: 27).

“ T_1 test istatistiği, yeterince büyük gözlem sayısı için standart normal dağılıma yakınsadığından bu istatistik için standart normal dağılım tablo değeri ile karşılaştırılarak aşırı yayılımın varlığı test edilmektedir.” (Dinarcı 2018: 26).

Skor testinden ayrı olarak aşırı yayılım olup olmadığı yayılım parametresi değerleri üzerinden de kontrol edilebilir. NBR modelinin yayılım parametresi α 'nın sifıra eşit olması, GPR modeli yayılım parametresinin ϕ 'nin bire eşit olması durumunda eldeki verinin aşırı yayılım göstermediği anlaşılmaktadır. Bu sebeple Poisson regresyon kullanılmasında bir sakınca olmayacaktır.

Negatif binom regresyon modelinin yayılım parametresi için yokluk hipotezi ve alternatif hipotezi;

$$H_0: \alpha = 0$$

$$H_1: \alpha > 0$$

Genelleştirilmiş poisson regresyon modelinin yayılım parametresinin yokluk hipotezi ve alternatif hipotezi;

$$H_0: \phi = 1$$

$$H_1: \phi > 1$$

şeklinde gösterilmektedir.

Yokluk hipotezlerinin reddedilmesi durumunda aşırı yayılım ortaya çıkacak ve PR yerine NBR veya GPR kullanılması daha uygun olacaktır.

Yokluk hipotezlerinin testinde Olabilirlik Oran Testi ve Wald Testi kullanılabilir.

ii. Olabilirlik Oran Testi (LR Testi)

Olabilirlik oran testi, Negatif binom ve poisson regresyonda ortaya çıkan yayılımın varlığını test etmek için kullanılır. Test kıyaslanacak olan modelin benzerlik oranlarını bulup log benzerlik oranını hesaplar. LR test istatistiği;

$$LR = -2(\ln L_1 - \ln L_0) \quad (2.14)$$

L_0 : Yokluk hipotezi doğru olduğu varsayımı altında hesaplanan en çok olabilirlik fonksiyonu, L_1 : Alternatif hipotezin doğru olduğu varsayımı altında hesaplanan en çok olabilirlik fonksiyonudur.

LR testi ki-kare dağılımına uymaktadır. Serbestlik derecesi ise kıyaslanan modellerin parametre sayılarının farkı kadardır. (Gujarati, Porter 2012: 259-260)

iii. Wald Testi

Wald testi, aşırı yayılımın olup olmadığını test etmek için kullanılmaktadır. Yayılım parametresi α ise Wald test istatistiği;

$$W = \frac{\hat{\alpha}^2}{V(\hat{\alpha})} \quad (2.15)$$

biçiminde hesaplanmaktadır.

$\hat{\alpha}$: yayılım parametre tahmini, $V(\hat{\alpha})$: Tahminin varyansını göstermektedir.

Wald test istatistiği ki-kare dağılımına sahiptir. (Gujarati, Porter 2012: 259-260)

iv. Vuong Testi

Aşırı yayılım durumunda olan modellerin Poisson regresyon modeli mi yoksa sıfır yığılmalı regresyon (ZIPR) model mi olup olmadığına karar vermek için vuong testi uygulanır.

Test istatistiği Kullback-Leiber bilgi ölçümü ile hesaplanmaktadır.

s : örneklem varyansı kareköküdür. $m: \frac{n}{2}$ den küçük pozitif bir tamsayıdır.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (2.16)$$

Eğer $i < 1$ ise $x_i = x_1$ 'dir ve $i > 1$ ise $x_i = x_n$ ' dir. K_{mn} 'nin değeri küçüldükçe asimptotik normallikten uzaklaşmaktadır.

$$K_{mn} = \frac{n}{2ms} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)})^{1/n} \right\} \quad (2.17)$$

Vuong istatistiği modelleri birbirleri ile karşılaştırmak için kullanılır (Vuong 1989: 307-333). Vuong istatistiği,

$$Q_v = \frac{(Q - \theta)}{\{2\sqrt{V(Q)n}\}} \quad (2.18)$$

Q_v asimptotik normal dağılıma sahiptir. θ düzeltme terimidir.

$\alpha=0,5$ anlam düzeyinde;

$Q_v > 1.96$ ise ZIPR, ZINB regresyon modelleri, poisson modeli ve negatif binom regresyon modelinden daha uygun olacaktır. Çünkü modelde aşırı yayılım mevcuttur. $Q_v < 1.96$ olduğunda ise Poisson model, Negatif binom regresyon modelleri, ZIPR ve ZINB regresyon modellerinden daha uygun olur. Çünkü bu durumda aşırı yayılım mevcut değildir. Mevcut olsa bile istatistiksel olarak anlamlı bir seviyede değildir (İşbilen 2021: 32-34).

2.2 Negatif Binom Regresyon Modeli

2.2.1 Negatif Binom Dağılım

Negatif binom dağılımı, ilk kez Pascal tarafından 1679 yılında ortaya atılmıştır. Dağılım aynı zamanda “Pascal Dağılımı” olarak da anılmaktadır. NBD poisson ve gamma dağılımlarının karışımından oluşmaktadır. Başka bir tanımlama ile NBD’nın Poisson dağılımlarının sürekli bir bileşiğinden ortaya çıktığı ve bu bileşiği temin eden Poisson oranlı dağılımın bir gamma dağılım olması gerçeğine dayanır. Bu tanıma göre NBD formülü;

$$P(Y = y) = \int_0^\infty \text{Poisson}(\mu) \cdot \text{Gamma}(\alpha, \theta) d\mu \quad (2.19)$$

“Negatif binom (NB) dağılımı iki parametrelili bir dağılımdır. Pozitif tamsayı y için, her bir denemede başarı olasılığının p olduğu y başarıları gerçekleşmeden önce, bir deneme sırasındaki başarı sayısının dağılımıdır. Dağılım, herhangi bir pozitif y için tanımlanmıştır. $y_i \sim \text{Poisson}(\mu)$ olduğunu varsayalım, ancak μ gamma dağılımına sahip herhangi bir değişken olsun. $y/\mu \sim \text{Poisson}(\mu)$, $\mu \sim \text{Gamma}(\alpha, \theta)$ biçimindedir. Buna ek olarak $\text{Gamma}(\alpha, \theta)$ ortalaması $\alpha\theta$, $\text{Gamma}(\alpha, \theta)$ varyansı $\alpha\theta^2$ sahiptir.” (Güney, Abar 2020: 168).

Olasılık yoğunluk fonksiyonu,

$$P(\mu) = \begin{cases} \frac{\mu^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} \exp(-\mu/\theta) & \mu \geq 0 \\ 0 & \mu < 0 \end{cases} \quad (2.20)$$

$P(\mu)$, eşitlik (2.19)'da yerine yazılırsa

$$\begin{aligned} P(Y = y) &= \int_0^\infty \text{Poisson}(\mu) \cdot \text{Gamma}(\alpha, \theta) d\mu \\ &= \int_0^\infty \frac{e^{-\mu} \mu^y}{y!} \frac{\mu^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} \exp(-\mu/\theta) d\mu \end{aligned} \quad (2.21)$$

$$\begin{aligned} &= \int_0^\infty \frac{e^{-\mu} \mu^y}{y!} \frac{1}{\theta^\alpha \Gamma(\alpha)} \mu^{\alpha-1} e^{(-\mu/\theta)} d\mu \\ &= \frac{1}{\Gamma(\alpha) y!} \frac{1}{\theta^\alpha} \int_0^\infty \mu^{y+\alpha-1} e^{-\mu(1+\frac{1}{\theta})} d\mu \end{aligned} \quad (2.22)$$

$$= \frac{1}{\Gamma(\alpha)} \frac{1}{\theta^\alpha} \left(\frac{\theta}{\theta+1} \right)^{y+\alpha} \Gamma(y+\alpha) \quad (2.23)$$

$$= \frac{\Gamma(\alpha+y)}{\Gamma(\alpha) y!} \left(\frac{\theta}{\theta+1} \right)^y \left(\frac{1}{\theta+1} \right)^\alpha \quad (2.24)$$

2.2.2 Negatif Binom Regresyon Modeli (NBR)

Aşırı yayılım (overdispersion) mevcutken poisson regresyon uygulanması sonucunda elde edilen tahminler tutarlı ancak etkin olmayacaktır. Bu sebeple alternatif olarak varyansın daha büyük olmasına imkân veren negatif binom regresyon ve genelleştirilmiş poisson regresyonun kullanımı önerilmektedir.

1960 yılında Leroy Simon, Poisson ve negatif binom modellerini ayıran ilk çalışmasının ardından, negatif binom için en çok olabilirlik algoritmasını yayınlayan ilk kişidir (Dinarcı 2018: 14).

Negatif binom regresyonu elde etmek için çeşitli yöntem bulunmaktadır. Bu yöntemlerden en yaygın olarak kullanılanı PR modelinin koşullu ortalamasına eklenen bir hata terimi (ε_i) ile gerçekleşir. Poisson regresyon gözlemlenen heterojenlik üzerinden hesaplama yapıyorken, eklenen hata terimi (ε_i) sayesinde Negatif binom modeli gözlemlenemeyen heterojenliğide hesaba katmış olur.

$$\begin{aligned} \hat{\mu}_i &= \exp(x_i' \beta + \varepsilon_i) \\ \hat{\mu}_i &= \exp(x_i' \beta) \exp(\varepsilon_i) = \mu_i \delta_i \end{aligned} \quad (2.25)$$

$$\exp(\varepsilon_i) = \delta_i \quad (2.26)$$

δ_i , gamma dağılımına sahip olasılık yoğunluk fonksiyonu,

$$g(\delta_i) = \frac{v_i^{v_i}}{\Gamma(v_i)} \delta^{(v-1)} e^{-\delta_i v_i} \quad (2.27)$$

Beklenen değeri: $E(\delta_i) = 1$,

Varyansı : $Var(\delta_i) = 1/v_i$

PR modelinde $\hat{\mu}_i$ yerine yazılır ise;

$$P(Y_i|x_i, \delta_i) = \frac{e^{-\mu_i \hat{\mu}_i} \mu_i^{y_i}}{y_i!} = \frac{e^{-(\mu_i \delta_i)} (\mu_i \delta_i)^{y_i}}{y_i!} \quad (2.28)$$

$$E(\hat{\mu}_i) = E(\mu_i \delta_i) = \mu_i E(\delta_i) = \mu_i$$

Poisson-gamma bileşiği ile negatif binom regresyon modeli aşağıdaki gibi bulunmaktadır (Kıran, 2019: 20).

$$P(y|\mu, v) = \int_0^\infty P(y|x, \delta) q(\delta) d\delta \quad (2.29)$$

$$= \int_0^\infty \frac{e^{-(\mu\delta)} (\mu\delta)^y}{y!} \frac{v^v}{\Gamma(v)} \delta^{(v-1)} e^{-\delta v} d\delta \quad (2.30)$$

$$= \int_0^\infty \frac{e^{-\delta(\mu+v)} \mu^y \delta^{y+v-1} v^v}{y! \Gamma(v)} d\delta \quad (2.31)$$

$$= \frac{\mu^y v^v}{y! \Gamma(v)} \int_0^\infty e^{-\delta(\mu+v)} \delta^{y+v-1} d\delta \quad (2.32)$$

$$= \frac{\mu^y v^v}{y! \Gamma(v)} \frac{\Gamma(y+v)}{(\mu+v)^{y+v}} \quad (2.33)$$

$$= \frac{\Gamma(y+v)}{\Gamma(y+1)\Gamma(v)} \left(\frac{\mu}{\mu+v}\right)^y \left(\frac{v}{\mu+v}\right)^v \quad (2.34)$$

Negatif binom regresyon modeli ,

$$P(y_i|\mu, v_i) = \binom{y_i+v_i-1}{v_i-1} \left(\frac{\mu_i}{\mu_i+v_i}\right)^{y_i} \left(\frac{v_i}{\mu_i+v_i}\right)^{v_i} \quad (2.35)$$

Negatif binom regresyon modelinin ortalaması,

$$E[y_i|x_i] = \exp(x_i' \beta) = \mu_i \quad (2.36)$$

Varyans,

$$Var[y_i|x_i] = \mu_i \left(1 + \frac{\mu_i}{v_i}\right) = \exp(x_i'\beta) \cdot \left(1 + \frac{\exp(x_i'\beta)}{v_i}\right) \quad (2.37)$$

olarak tanımlanmaktadır.

v_i 'nin bireysel olarak değişmesi gözlemden daha fazla parametre olması sorununu yaratmaktadır. Dolayısıyla en yaygın belirleyici varsayım, v_i 'nin tüm bireyler için aynı olmasıdır (Kıran 2019: 22).

v_i gamma ölçek parametresidir ve aşırı yayılım parametresinin (α 'nın) tersine eşittir. Poisson regresyon modeli, $\alpha=0$ için Negatif binom regresyonunun özel bir halidir.

$$v_i = 1/\alpha, \alpha > 0$$

Bu varsayıma göre δ_i varyansın homoskedastik yani sabit olduğunu gösterir. Eklenen hata terimi sayesinde varyans şu şekilde değişir:

$$Var[y_i|x_i] = \mu_i \left(1 + \frac{\mu_i}{1/\alpha}\right) = \mu_i(1 + \alpha\mu_i) \quad (2.38)$$

$$= \mu_i + \alpha\mu_i^2 \quad (2.39)$$

Bu durumda model düzenlendiğinde NBR modeli

$$P(y_i|\mu_i, \alpha) = \frac{\Gamma(y_i + (1/\alpha))}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{\mu_i}{\mu_i + 1/\alpha}\right)^{y_i} \left(\frac{1/\alpha}{\mu_i + 1/\alpha}\right)^{1/\alpha} \quad (2.40)$$

Sayma verisi için kullanılan birçok negatif binom regresyon modeli bulunmaktadır. Tablo 2.1'de NBR modellerinin bazıları verilmiştir.

Tablo 2. 1. Negatif Binom Regresyon Modelleri

Negatif Binom Regresyon Modelleri

- 1) Basit seçim Negatif Binom
- 2) NB2
- 3) NB1
- 4) Sıfır Değerli Negatif Binom (ZINB)
- 5) Negatif Binom - Hurdle
- 6) Negatif Binom-C < Kanonik>
- 7) Negatif Binom-P
- 8) Negatif Binom-H <Heterojen>
- 9) Kesikli-budanmış- Negatif Binom; NB1;NB-C
- 10) Endojen tabakalı Negatif Binom
- 11) Gizli sınıf Negatif Binom
- 12) Sansürlü Negatif Binom
- 13) İki değişkenli Negatif Binom
- 14) Kantil sayım
- 15) Araç değişkenler
- 16) Koşullu sabit etkiler
- 17) Sonlu karma modeller
- 18) Karma ve çok düzeyli etkiler
- 19) Rasgele etkiler
- 20) Genelleştirilmiş moment metodu

Kaynak: (Hilbe 2007).

2.2.3 Doğrusal Varyans Fonksiyonu İle NB1

Cameron ve Trivedi (1986) ortalamanın varyansa eşit olduğu varsayımının uygulamada reddedilme durumuna karşı varyansı genişletmişlerdir. Ortalama $\exp(x_i'\beta)$ varsayımı korunarak y_i ' nin koşullu varyansı genel gösterimi:

$$\omega_i = V[y_i|x_i] \quad (2.41)$$

Ortalamanın bir fonksiyonu olarak varyans

$$\omega_i = \omega(\mu_i, \alpha) \quad (2.42)$$

“p” belirtilen bir sabit olmak üzere genellikle aşağıdaki varyans fonksiyonu kullanılmaktadır.

$$\omega_i = \mu_i + \alpha\mu_i^p \quad (2.43)$$

p=1 durumunda NB1 varyans fonksiyonu

$$\omega_i = 1 + \alpha\mu_i \quad (2.44)$$

Bu formülü GDM altında tekrardan şekillendirecek olursak

$$\omega_i = \phi \mu_i \quad (2.45)$$

elde ederiz. Burada $\phi = 1 + \alpha$ dır (Cameron ve K.Trivedi, 1998: 63).

2.2.4 Karesel Varyans Fonksiyonu İle NB2

Negatif binom en yaygın kullanılan şeklidir. $p=2$ olması durumunda varyans ortalamada kareseldir. NB2 varyans fonksiyonu şu şekildedir:

$$\omega_i = \mu_i + \alpha \mu_i^2 \quad (2.46)$$

Her iki durumda da dağılım parametresi α tahmin edilecektir. Bu varyans fonksiyonları iki yerde kullanılabilir. Bunlardan birincisi poisson sözde (quasi) en çok olabilirlik (QMLE)'den daha etkin tahmin ediciler sağlamak için, ikincisi ise poisson QMLE için alternatif standart hatalar elde etmek içindir (Yıldırım 2019: 59)

2.2.5. Negatif Binom Regresyon İçin En Çok Olabilirlik Fonksiyonu

NBR modeli,

$$P(y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + (1/\alpha))}{\Gamma(y_i + 1) \Gamma(1/\alpha)} \left(\frac{\mu_i}{\mu_i + 1/\alpha} \right)^{y_i} \left(\frac{1/\alpha}{\mu_i + 1/\alpha} \right)^{1/\alpha} \quad (2.47)$$

Modelde $\Gamma(\cdot)$ fonksiyonu, y tam sayı olduğunda,

$$\frac{\Gamma(y + \alpha)}{\Gamma(\alpha)} = \prod_{j=0}^{y-1} \ln(j + \alpha) \quad (2.48)$$

şeklinde yazılır.

Negatif binom regresyon modelindeki ifadenin logaritması şu şekilde yazılabilir:

$$\frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})} = \sum_{j=0}^{y-1} (j + \alpha^{-1}) \quad (2.49)$$

Yukarıdaki gösterimi de kullanarak log-olabilirlik fonksiyonu (Cameron, Trivedi 1998: 70-74) :

$$L = \sum_{i=0}^n \{ (\sum_{j=0}^{y-1} \ln(j + \alpha^{-1})) - \ln y_i! - (y_i + \alpha^{-1}) \ln(1 + \alpha \exp(x_i' \beta)) + y_i \ln \alpha + y_i x_i' \beta \} \quad (2.50)$$

şeklinde tanımlanmaktadır.

$\hat{\beta}$, α parametrelerinin tahmini için EÇO fonksiyonu:

$$\sum_{i=1}^n \frac{x_{ij}(y_i - \mu_i)}{1 + \alpha \mu_i} = 0 \quad j = 1, 2 \dots k \quad (2.51)$$

$$\sum_{i=1}^n \left\{ \alpha^{-2} \left(\ln(1 + \alpha \mu_i) - \sum_{j=0}^{y_i-1} \frac{1}{j + \alpha^{-1}} \right) + \frac{y_i - \mu_i}{\alpha(1 + \alpha \mu_i)} \right\} = 0 \quad (2.52)$$

Ortalamanın üstel olması sebebi ile Log-olabilirlik fonksiyonu doğrusal değildir. Bu durumda β 'lar için EÇO yöntemi ile tek adımda çözüme ulaşamayabiliriz. Çözüme ulaşabilmek için iteratif yöntem olan Newton-Raphson yöntemi kullanılmaktadır. (2.51) ve (2.52) çözümü sonucunda $\hat{\beta}$, α parametre tahminlerine ulaşılır (Sezgin, Deniz 2004: 19-21).

2.3. Genelleştirilmiş Poisson Regresyon Modeli (GPR)

GPR modeli Consul ve Famoye (1992) tarafından Poisson regresyon yerine seçenek olarak önerilmiştir. NBR modeli aşırı yayılım durumunda kullanılırken genelleştirilmiş poisson regresyon hem aşırı hem de az yayılım durumunda kullanılabilir. (2.53)

$y_i = 1, 2, \dots$, olmak üzere GPR modeli:

$$f(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha \mu_i} \right) \frac{(1 + \alpha y_i)^{y_i-1}}{y_i!} \exp \left[\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} \right] \quad (2.53)$$

Literatürde belirtildiği gibi (Famoye, 1993) x_i , (k-1) boyutlu bağımsız değişkenlere ilişkin kovaryans vektörü; β , k boyutlu bilinmeyen regresyon parametreleri vektörünü göstermektedir.

y_i değişkeni için ortalama

$$E(y_i | x_i) = \mu_i \quad (2.54)$$

ve varyans:

$$V(y_i | x_i) = \mu_i(1 + \alpha)^2 \quad (2.55)$$

GPR modelinde α , aşırı yayılım parametresi sıfırdan büyük olduğunda ($\alpha > 0$), aşırı yayılım mevcuttur. Aynı şekilde α sıfırdan küçük olduğunda ($\alpha < 0$) az yayılım mevcuttur (Famoye 1993: 1335-1354)

Tablo 2.2. Sayma Veri Regresyon Modelleri

Poisson, NB ve GPR regresyon modelleri		
Regresyon Modeli	Fonksiyon	Ortalama ve Varyans
Poisson	$P(y_i x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$	$E(y_i x_i) = \mu_i$ $V(y_i x_i) = \mu_i$
NB1	$\frac{\Gamma(y_i + \mu_i(1/\alpha))}{y_i! \Gamma(1/\alpha)} \left(\frac{\mu_i}{\mu_i + \mu_i(1/\alpha)} \right)^{y_i} \left(\frac{\mu_i(1/\alpha)}{\mu_i + \mu_i(1/\alpha)} \right)^{\mu_i(1/\alpha)}$	$E[y_i x_i] = \mu_i$ $Var[y_i x_i] = \mu_i(1 + \alpha)$
NB2	$\frac{\Gamma(y_i + (1/\alpha))}{\Gamma(y_i + 1) \Gamma(\mu_i(1/\alpha))} \left(\frac{\mu_i}{\mu_i + 1/\alpha} \right)^{y_i} \left(\frac{1/\alpha}{\mu_i + 1/\alpha} \right)^{1/\alpha}$	$E[y_i x_i] = \mu_i$ $Var[y_i x_i] = \mu_i(1 + \alpha\mu_i)$
NB-P	$\frac{\Gamma(y_i + (1/\alpha)\mu_i^{2-P})}{y_i! \Gamma(1/\alpha \mu_i^{2-P})} \left(\frac{\mu_i}{\mu_i + \mu_i^{2-P} 1/\alpha} \right)^{y_i} \left(\frac{\mu_i^{2-P} 1/\alpha}{\mu_i + \mu_i^{2-P} 1/\alpha} \right)^{\mu_i^{2-P} 1/\alpha}$	$E[y_i x_i] = \mu_i$ $Var[y_i x_i] = \mu_i(1 + \alpha\mu_i^{P-1})$
GPR-1	$\frac{((1-v)\mu_i)((1-v)\mu_i + v y_i)^{y_i-1}}{y_i!} \exp(-(1-v)\mu_i + v y_i)$	$E(y_i x_i) = \mu_i$ $V(y_i x_i) = \mu_i(1-v)^{-2}$
New GPR	$\left(\frac{\mu_i}{1 + \alpha\mu_i} \right) \frac{(1 + \alpha y_i)^{y_i-1}}{y_i!} \exp \left[\frac{-\mu_i(1 + \alpha y_i)}{1 + \alpha\mu_i} \right]$	$E(y_i x_i) = \mu_i$ $V(y_i x_i) = \mu_i(1 + \alpha)^2$
GPR-2	$\frac{\mu_i(\mu_i + \alpha\mu_i y_i)^{y_i-1}}{(1 + \alpha\mu_i)^{y_i} y_i!} \exp \left[-\frac{(\mu_i + \alpha\mu_i y_i)}{1 + \alpha\mu_i} \right]$	$E(y_i x_i) = \mu_i$ $V(y_i x_i) = \mu_i(1 + \alpha\mu_i)^2$
New GPR-P	$\frac{\mu_i(\mu_i + \alpha\mu_i^{P-1} y_i)^{y_i-1}}{(1 + \alpha\mu_i^{P-1})^{y_i} y_i!} \exp \left[-\frac{(\mu_i + \alpha\mu_i^{P-1} y_i)}{1 + \alpha\mu_i^{P-1}} \right]$	$E(y_i x_i) = \mu_i$ $V(y_i x_i) = \mu_i(1 + \alpha\mu_i^{P-1})^2$

Kaynak: (İsmail et al 2013).

2.4. Sıfır Yığılmalı Regresyon Modelleri (Zero-Inflated)

Sayma verileri ile modellemede aşırı yayılımın yanında dikkat edilmesi gereken diğer bir hususta gözlemlenen verinin beklenenden fazla sıfıra sahip olmasıdır. Yoğun sıfıra sahip olan verilerde klasik regresyon yöntemlerinden yararlanmak doğru olmayan sonuçlara sebep olup, yanlış parametre tahminlerine ve tutarsız sonuçlara yol açar.

Gözlemlenen verinin beklenenden fazla sıfıra sahip olması durumunda sıfırları dikkate alan sıfır yığılmalı (zero-inflated) regresyon modelleri ile analiz yapmak uygun olacaktır. Sıfır yığılmalı modeller aynı zamanda iki durumlu süreçler (dual state process) olarak da adlandırılmaktadır.

Sıfır yığılmalı modeller verileri iki kısım şeklinde ele almaktadır. Birinci kısım sayma verilerinden oluşmaktadır ve sıfır içerebilmektedir. İkinci kısım ise

sadece sıfırlardan oluşmaktadır. Verileri iki kısım halinde düşünmek bir varsayımsal bir yaklaşım olup, verileri gerçek anlamda iki kısma ayırma gibi bir durum yoktur.

Beklenenden daha fazla sıfır durumunda ortaya çıkan sıfır değerler elde edilmiş durumlarına göre yapısal sıfır (structural zeros) ve örnekleme sıfırı (sampling zeros) olarak iki şekilde adlandırılmaktadır. Değerlendirilen veri kümesinde elde edilme ihtimali olmadığından yani gözlem yapmanın imkân dâhilinde olmadığı durumlarda hücrenin sıfır değerini almasına *yapısal sıfır*; elde edilmesi mümkün olduğu halde veri setinde gözlenmeyen, bu nedenle sıfır değer alan gözlemlere de *örnekleme sıfırı* denir (Karaca, Olmuş 2018 : 106).

Bu modeller, bağımlı değişkenin sıfır değeri için Bernoulli denemesi yapılarak belirlenen bir yoğunluk (zero point mass) ile sıfır değerini de kapsayan kesikli bir dağılımın karmasından oluşmaktadır (Tüzel, Sucu 2012: 25).

Beklenenden fazla sıfır olması durumunda iki model tahmin edilmektedir. İlk model logit ve probit bağ fonksiyonu ile tahmin edilen yapısal sıfırların olduğu Bernoulli kısmıdır. İkinci model ise Y'nin dağılımına bağlı olarak sayma verilerinden oluşan kısımdır.

Verinin sadece sıfırlardan oluşan kısmına Bernoulli kısmı, diğer kısmını ise sayma verisi kısmı olarak adlandıırırsak Sıfır yığılmalı regresyon modeli:

$$\Pr(Y_i = y_i) \begin{cases} \omega_i + (1 - \omega_i)Pr(j_i = 0) & y_i = 0 \\ (1 - \omega_i)Pr(j_i = y_i) & y_i > 0 \end{cases} \quad (2.56)$$

$0 \leq \omega_i < 1$, j_i rastlantı değişkeni pozitif kesikli dağılıma sahiptir.

ω_i : Bernoulli kısmının olasılığıdır (yapısal sıfırlar)

$(1 - \omega_i)$: sayma verisi kısmının olasılığı

2.4.1 Sıfır Yığılmalı Poisson Regresyon Modeli (ZIPR)

Sıfır yığılmalı poisson regresyon (ZIPR), Lambert tarafından 1992 yılında geliştirilmiştir. ZIPR modeli bir önceki başlıkta da bahsettiğimiz gibi modeli iki kısım şeklinde düşünülmektedir. Burada birinci kısım poisson dağılım gösteren sayma verilerinden oluşan kısımdır.

Sıfır yığılmalı Poisson regresyon modelinde, veride mevcut olan aşırı yayılımın yalnızca sıfır değerindeki yoğunluktan kaynaklandığı varsayımı yapılmaktadır (Tüzel, Sucu 2012: 26).

ZIPR modeli için olasılık fonksiyonu;

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i}, & y_i = 0 \\ (1 - \omega_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, & y_i > 0 \end{cases} \quad (2.57)$$

$0 \leq \omega_i < 1$ ve $\mu_i > 0$ olmak üzere beklenen değeri ve varyansı;

$$E(Y_i) = \mu_i(1 - \omega_i) \quad (2.58)$$

$$V(Y_i) = (1 - \omega_i)\mu_i(1 + \omega_i\mu_i) \quad (2.59)$$

Şeklinde tanımlanır. $\omega_i = 0$ olması durumunda ZIPR modeli Poisson modeline dönüşür. $\omega_i > 0$ durumunda ise aşırı yayılım gösterir.

Modelin log dönüşüm fonksiyonu ile modellenen kısmı poisson dağılımından gelen pozitif sayıları modellemede kullanılırken, logit kısmı veri setindeki sıfırları modellemede kullanılır (Kılıç, Bayrak 2020 : 90).

ZIPR modelinde her iki kısım için farklı bağlantı fonksiyonları kullanılır. Log parçaya ait denklem $\log(\mu) = B\beta$ iken lojit parçasına ait denklem;

$$\text{logit}(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = G\gamma \quad (2.60)$$

$$\omega_i = \frac{\exp(G\gamma)}{1+\exp(G\gamma)} \quad (2.61)$$

Bağlantı fonksiyonlarını kullanarak parametre tahmini yapılır. Burada B ve G ortak değişken matrisleri β ve γ bilinmeyen parametre vektörlerini göstermektedir.

ZIPR modeli log-olabilirlik fonksiyonu;

$$LL(\gamma, \beta, y) = \sum_{y_i=0} \log(e^{G_i\gamma} + \exp(-e^{B_i\beta})) + \sum_{y_i>0} (y_i B_i\beta - e^{B_i\beta}) - \sum_{i=1} \log(1 + e^{G_i\gamma}) - \sum_{y_i>0} \log(y_i!) \quad (2.62)$$

2.4.2 Sıfır Yığılmalı Negatif Binom Regresyon Modeli (ZINB)

ZINB regresyon modeli Greene tarafından 1994 yılında sıfır yığılmalı modellerde beklenenden fazla sıfır barındıran verilerin modellenmesi için ZIPR modeli yerine bir seçenek olarak sunulmuştur.

Bu modelde sıfır yığılmalı poisson regresyon gibi veriyi iki kısımda ele alır. ZINB regresyon modelinde birinci kısım negatif binom dağılımı gösteren sayma verilerinden oluşmaktadır.

ZINB regresyon modeli için olasılık fonksiyonu;

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i) \left(\frac{1/\alpha}{\mu_i + 1/\alpha} \right)^{1/\alpha}, & y_i = 0 \\ (1 - \omega_i) \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{\mu_i}{\mu_i + 1/\alpha} \right)^{y_i} \left(\frac{1/\alpha}{\mu_i + 1/\alpha} \right)^{1/\alpha}, & y_i > 0 \end{cases} \quad (2.63)$$

ZINB regresyon modelinin beklenen değeri ve varyansı Eşitlik (2.64) ve (2.65)' teki gibidir (Long 1997: 240-244).

$$E(Y_i) = \mu_i(1 - \omega_i) \quad (2.64)$$

$$V(Y_i) = (1 - \omega_i)\mu_i(1 + \alpha\mu_i + \omega_i\mu_i) \quad (2.65)$$

ZINB regresyon modelinde de sıfır yığılmalı poisson regresyon modelinde olduğu gibi log ve logit olmak üzere iki parçadan oluşur ve log parçalı kısmı negatif binom dağılımından gelen pozitif sayıları modeller diğer bir kısım olan logit parçası ise sıfırları modellemektedir.

$$\log(\mu) = B\beta \quad (2.66)$$

$$\text{logit}(\omega_i) = \log\left(\frac{\omega_i}{1-\omega_i}\right) = P\gamma \quad (2.67)$$

Bağlantı fonksiyonlarını kullanarak parametre tahmini yapılır. Burada B ve P ortak değişken matrislerini β ve γ bilinmeyen parametre vektörlerini göstermektedir.

ZINB regresyon modeli için log-olabilirlik fonksiyonu;

$$\begin{aligned}
LL = (\beta, \gamma, \alpha) = & \sum_{i=1}^n \log(1 + e^{P_i \gamma}) \\
& - \sum_{i=1: y_i=0}^n \left(e^{P_i \gamma} + \left(\frac{e^{B\beta} + \alpha}{\alpha} \right)^{-\alpha} \right) \\
& + \sum_{i=1: y_i=0}^n \left(\alpha \log \left(\frac{e^{B\beta} + \alpha}{\alpha} \right) \right) \\
& + y_i \log(1 + e^{-B\gamma}) \alpha \\
& + \sum_{i=1: y_i=0}^n \log \Gamma(\alpha) + \log(1 + y_i) - \log \Gamma(\alpha + y_i)
\end{aligned} \tag{2.68}$$

Şeklinde tanımlanmaktadır.

2.4.3 Sıfır Yığılmalı Genelleştirilmiş Poisson Regresyon Modeli (ZIGP)

ZIPG regresyon modeli Famoye ve Singh tarafından sıfır yığılmalı poisson regresyon modelinin yeterli gelmemesi ve ZINB regresyon modeline ek bir seçenek olarak sıfır yığılmalı genelleştirilmiş regresyon modelini önermiştir.

Modelde μ , φ , ω sırasıyla ortalama, aşırı yayılım ve sıfır yoğunluklu parametreleri göstermektedir. ZIGP regresyon modeli olasılık fonksiyonu;

$$P(Y_i = y_i) = \begin{cases} \omega_i + (1 - \omega_i) e^{\frac{-\mu_i}{\varphi}} & , y_i = 0 \\ (1 - \omega_i) \frac{\mu_i(\mu_i + (\varphi - 1)y_i)^{y_i - 1}}{y_i!} \varphi^{-y_i} e^{\left(\frac{-1}{\varphi}(\mu_i(\varphi - 1))\right)y_i} & , y_i > 0 \end{cases} \tag{2.69}$$

$$0 \leq \omega_i < 1$$

ZIGP regresyon modeli beklenen değeri ve varyansı;

$$E(Y_i) = \mu_i(1 - \omega_i) \tag{2.70}$$

$$V(Y_i) = (1 - \omega_i)\mu_i[\varphi^2 + \omega_i\mu_i] \tag{2.71}$$

$\omega_i = 0$ ve $\varphi^2 = 1$ olduğunda ZIGP regresyon modeli Poisson modele dönüşmektedir. $\omega_i = 0$ olduğunda ise model genelleştirilmiş poisson modele dönüşür. Ayrıca $\varphi^2 = 1$ olması durumunda da model ZIPR modeline dönüşmektedir.

Sıfır yığılmalı negatif binom ve sıfır yığılmalı genelleştirilmiş Poisson regresyon modellerinde ise aşırı yayılımın sıfır değerindeki yığılmanın yanı sıra birçok nedenden kaynaklanabileceği varsayımı mevcuttur. Dolayısıyla tüm sıfır yığılmalı regresyon modellerinin ortalaması birbirine eşit iken varyans yapıları değişmektedir (Tüzel, Sucu 2012).

$ZIGP(\mu_i, \varphi_i, \omega_i)$ regresyon modeli tesadüfi, sistematik ve parametrik olmak üzere 3 bileşenden oluşur (Czado vd., 2007: 3).

Tesadüfi bileşen;

$$\{Y_i, 1 \leq i \leq n\}, Y_i \sim ZIGP(\mu_i, \varphi_i, \omega_i)$$

Sistematik Bileşen;

Y_i bağımlı değişkeni üzerinde üç doğrusal tahmin edici olan $\eta_i^\mu(\beta) = x_i^t \beta$, $\eta_i^\varphi(\alpha) = \omega_i^t \alpha$, $\eta_i^\omega(\gamma) = z_i^t \gamma$ etkili olur ve bilinmeyen regresyon parametreleri;

$$\beta = (\beta_0, \beta_1, \dots, \beta_p)^t, \alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)^t, \gamma = (\gamma_0, \gamma_1, \dots, \gamma_p)^t$$

Tasarım matrisleri;

$$X_i = (X_0, X_1, \dots, X_n)^t, W_i = (W_0, W_1, \dots, W_n)^t, Z_i = (Z_0, Z_1, \dots, Z_n)^t$$

Parametrik Bileşen;

$\eta_i^\mu(\beta)$, $\eta_i^\varphi(\alpha)$, $\eta_i^\omega(\gamma)$ doğrusal tahmin edicileri ile $\mu_i(\beta)$, $\varphi_i(\alpha)$, $\omega_i(\gamma)$ parametreleri arasındaki doğrusal fonksiyonlar aşağıda maddeler halinde verilmiştir.

i. Ortalama düzeyi:

$$E(Y_i|\beta) = \mu_i(\beta) = E_i e^{x_i^t \beta} \quad (2.72)$$

$$\Leftrightarrow \eta_i^\mu(\beta) = \log \mu_i(\beta) - \log E_i \quad \text{Log bağlantı}$$

ii. Aşırı yayılım düzeyi:

$$\varphi_i(\alpha) = 1 + e^{w_i^t \alpha} \quad (2.73)$$

$$\eta_i^\varphi(\alpha) = \log(\varphi_i(\alpha) - 1) \quad \text{Düzeltilmiş log bağlantı}$$

iii. Sıfır yığılmalı düzey:

$$\omega_i(\gamma) = \frac{\exp(Z\gamma)}{1+\exp(Z\gamma)} \quad (2.74)$$

$$\eta_i^\omega(\gamma) = \log\left(\frac{\omega_i(\gamma)}{1-\omega_i(\gamma)}\right) \text{ Logit bağlantı}$$

β , α ve γ regresyon parametrelerinin ortak vektörünü δ ile yani $\delta: (\beta^t, \alpha^t, \gamma^t)^t$ ve maksimum olabilirlik (ML) tahminini $\hat{\delta}$ ile gösteririz (Czado vd., 2007: 1-5).

$$\mu_i(\beta) : e^{x_i^t \beta + \log E_i} \quad \varphi_i(\alpha) : 1 + b_i(\alpha)$$

$$\omega_i(\gamma) : \frac{k_i(\gamma)}{1+k_i(\gamma)} \quad b_i(\alpha) : e^{\omega_i^t \alpha} \quad k_i(\gamma) : e^{z_i^t \gamma}$$

ve

$$P_i^0(\delta) = \exp\left(-\frac{E_i x_i^t \beta}{1+e^{\omega_i^t \alpha}}\right) \quad (2.75)$$

ZIGP regresyon modeli için log-olabilirlik fonksiyonu;

$$\begin{aligned} L(\delta) = \sum_{i=1}^n 1_{\{y_i=0\}} & \left[\log\left(e^{z_i^t \gamma} + \exp\left(-\frac{E_i e^{x_i^t \beta}}{1+e^{\omega_i^t \alpha}}\right)\right) - \log(1+e^{z_i^t \gamma}) \right] \\ & + 1_{y_i>0} \left[-\log(1+e^{z_i^t \gamma}) \right. \\ & + \log(E_i) + x_i^t \beta + (y_i - 1) \\ & \left. \times \log(E_i e^{x_i^t \beta} + e^{\omega_i^t \alpha} y_i) - \log(y_i!) - y_i \log(1+e^{\omega_i^t \alpha}) - \frac{E_i e^{x_i^t \beta} + e^{\omega_i^t \alpha}}{1+e^{\omega_i^t \alpha}} \right] \end{aligned} \quad (2.76)$$

2.5. Hurdle Regresyon Modeli

Hurdle regresyon modeli ilk kez 1971 yılında Cragg tarafından sansürlenmiş veriler için ortaya atılmıştır. Ancak sayma verilerine uygulanması ve genel yapısı 1986 yılında Mullahy tarafından açıklanmıştır.

Hurdle (Engel) regresyon modeli sıfır yığılmalı regresyon modellerini alternatif olarak kullanılmaktadır. Hurdle regresyon süreci iki aşamalı bir süreçtir. Modelde değişkenlerin (yanıt) sıfır ya da sıfırdan farklı değer alma durumuna göre

yol izlenir. Bu amaçla hurdle modeller yanıt değişkeninin pozitif değerlere ve sıfıra karşılık olacak şekilde bir olasılık tanımlamaktadır. Böylece model ikili yanıt yöntemi uygulayarak, verideki sıfırlara pozitif cevap vermek koşuluyla sıfır sayımlara 0, karşı pozitif sayım değerlerine ise 1 gibi ikili yanıtlar vermektedir ve bu sürecin ilk aşamasıdır. Modelin ikinci aşaması ise sıfırdan farklı pozitif değerlerin yer aldığı kısımdır ve sıfır değerle sınırlandırılmış (zero-value truncated model) sayma verisi yöntemi kullanılarak modellenir. (Özmen, Famoye 2007: 491-502)

Hurdle regresyon modellerin, iki aşamalı modelleme sürecini kullandığından bahsetmiştik. Bu ikili aşamaların birincisi geçiş aşaması (transition stage) yani cevapların sınırlamanın altına veya üstüne düşmelerini ölçen ikili değişkeni modelleridir. İkincisi ise olay aşaması (event stage) sınırlamanın üstündeki gözlemleri açıklamak için kesikli (truncated) model kullanır. Sınırlama sıfırdır.

İkinci aşama olan olay aşaması poisson ya da negatif binom dağılıbilir ve sıfır değerler hariç elde edilen pozitif değerlerin modellendiği kısımdır. Bu durumda modelin poisson ve negatif binom dağılma sahip olabilir. PH ve NBH modelleri önerilmektedir.

2.5.1 Hurdle Poisson Regresyon Modeli

Pozitif sayma değerlerinin modellendiği olay aşaması kısmının poisson dağılım göstermesi sonucunda poisson hurdle modeli önerilmektedir.

Hurdle poisson regresyon modeli;

$$P(Y_i = y_i) = \begin{cases} \omega & y = 0 \text{ ise} \\ (1 - \omega_i) \frac{e^{-\mu} \mu^y}{(1 - e^{-\mu}) \cdot \Gamma(y+1)} & y \geq 1 \text{ ise} \end{cases} \quad (2.77)$$

şeklinde tanımlanmaktadır. ω sıfır olma ihtimalini gösterir.

y_i , $i = 1, 2 \dots n$ birbirinden bağımsız sayıma dayalı olarak elde edilen gözlem değerleri olsun. $y_i = 0$ olma olasılığı $1 - p(x)$ ve $y_i \approx$ sınırlandırılmış Poisson $\mu(z)$ olma olasılığı $p(x)$ olsun. Burada x ve z ortak değişken matrisleridir. Poisson hurdle model,

$$P(y_i = 0/x) = 1 - p(x) \quad (2.78)$$

$$P(y_i = q|x, z) = \frac{p(x)\exp(-\mu(z))\mu(z)^q}{q!(1-\exp(-\mu(z)))} \quad (2.79)$$

şeklinde yazılabilir. Yukarıda verilen $\mu(z)$, $p(x)$ sırasıyla log-doğrusal ve logit fonksiyonları ile modellenmektedir.

$$\log(\mu(z)) = x_i'\beta \quad (2.80)$$

$$\text{logit}(p_i) = z_i'\alpha \quad (2.81)$$

şeklinde. β ve α bilinmeyen parametre vektörleridir. Bu parametrelerin tahmininde ML yöntemi kullanılmaktadır.

Poisson hurdle regresyon modeli için log-olabilirlik fonksiyonu;

$$\begin{aligned} L &= \sum_{y_i > 0} x_i\beta - \sum_{i=1}^n \log(1 + \exp(x_i\beta)) \\ &\quad + \sum_{y_i > 0} \left[y_i z_i \alpha - \exp(z_i \alpha) - \log(1 - \exp(-\exp(z_i \alpha))) - \log(y_i!) \right] \\ &= L(\beta) + L(\alpha) \end{aligned} \quad (2.82)$$

biçiminde yazılmaktadır. $L(\beta)$ ve $L(\alpha)$ ayrı ayrı maksimize edilerek ML tahminleri elde edilmektedir. (Min, Agresti 2005: 1-19)

2.5.2 Hurdle Negatif Binom Regresyon Modeli

Pozitif sayma değerlerinin modellendiği olay aşaması kısmının negatif binom dağılımı göstermesi sonucunda negatif binom hurdle modeli önerilmektedir.

Hurdle negatif binom regresyon modeli;

$$P(Y_i = y_i) = \begin{cases} \omega & y_i = 0 \text{ ise} \\ (1 - \omega) \frac{\Gamma(y_i + 1/\alpha)^\omega}{\left[1 - \left(\frac{1}{1 + \alpha\mu_i}\right)^{1/\alpha}\right] \Gamma(y_i + 1) \Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} & y_i \geq 1 \text{ ise} \end{cases} \quad (2.83)$$

şeklinde tanımlanmaktadır. α yayılım parametresidir. $\alpha = 0$ durumunda model poisson hurdle modelini dönüşür.

$$LL = \ln(f(0)) + \{\ln[1 - f(0)] + \ln P(j)\} \quad (2.84)$$

$f(0)$, modelin ikili cevap olasılığını göstermektedir. $P(j)$, pozitif sayma olasılığını gösterir.

$$f(0) = P(y_i = 0; x) = 1/1 + \exp(x'_i\beta) \quad (2.85)$$

ve

$$1 - f(0) \text{ ise} \\ \exp(x\beta_1)/(1 + \exp(x\beta_1)) \quad (2.86)$$

Böylece negatif binom Hurdle modelin her iki kısmı için log-olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\text{cond} \{y = 0, \ln(1/1 - \exp(xb1)), \ln(\exp(xb1)/1 + \exp(xb1)) + y^* \ln(\exp(xb)/1 + \exp(xb)) - \\ \ln(1 + \exp(xb))/\alpha + \ln \Gamma(y + 1/\alpha) - \ln(y + 1) - \ln \Gamma(1/\alpha) - \ln(1 - (1 + \exp(xb))(-1/\alpha))\} \quad (2.87)$$

eşitliğinde Log-Olabilirlik fonksiyonunun maksimize edilmesiyle, bilinmeyen parametrelerin en çok olabilirlik tahminleri bulunmaktadır (Kaya, Yeşilova 2012: 55-58).

2.6. Literatür Taraması

Selim ve Üçdoruk (2003), Devlet İstatistik Enstitüsü 1994 yılı hanehalkı gelir dağılımı araştırması verilerini kullanarak çocuk sayısı belirleyicilerini modellenmesi amacıyla sayma veri ile regresyon modellerinden faydalanmışlardır. Çalışmada bağımlı değişkenin eksik yayılım gösterdiği görülmüş ve poisson regresyon modeli yerine poisson quasi maksimum olabilirlik tahmini kullanılmıştır.

Sezgin ve Deniz (2004), 1964-2000 dönemi Türkiye grev sayılarının modellenmesi amacıyla sayma verisi ile regresyon modellerinden poisson ve negatif binom modellerini açıklamış ve grev sayılarında ortaya çıkan aşırı yayılım sonucunda negatif binom regresyon modellerini tercih etmişlerdir.

Özmen ve Famoye (2007), Türkiye Dalyan'da 1991-1993 arası Güneşe maruz kaldığı için ölen caretta caretta yavrularının ölüm hızı sayılarının modellenmesi için

sayma veri regresyon modellerine başvurmuşlardır. Çalışmada verideki sıfır yoğunluğunun yanı sıra aşırı yayılıma rastlanmıştır. Yapılan uyum testleri sonucunda genelleştirilmiş poisson regresyon modeli tercih edilmiştir.

Kaya ve Yeşilova (2012), Yüzüncü Yıl Üniversitesi (YYÜ) e-posta sunucusundan personelin 2009 bahar eğitim öğretim döneminde yaptıkları e-posta trafiği modellenmesi için sayma verisi ile regresyon modellerini kullanmıştır. Bağımlı değişken üzerinde aşırı yayılıma ve sıfır yoğunluğuna rastlanmış ve bunun sonucunda veriye uygun olabilecek ZINB ve NBH modelleri uygulanmıştır. Bu iki modelin karşılaştırması sonucunda ZINB modeli tercih edilmiştir.

Noriszura ve Zamani (2013), İki ayrı veri seti kullanmıştır. Birinci veri seti Malezya otomobil hasar talebi sayıları, ikinci veri seti ise Alman sağlık verileridir. Çalışmada bu veri setlerinin modellenmesi için sayma verisi modelleri kullanılmıştır.

Dinarcen (2018), 2016 yılı TÜİK Hanehalkı İşgücü araştırması verilerini modellemek amacıyla sayma veri modellerinden yararlanmıştır. Çalışmada bağımlı değişken olarak bireylerin haftalık çalışma saatleri ele alınmış ve sektör, yaş, cinsiyet, eğitim meslek, kayıtlılık ve süreklilik açıklayıcı değişken olarak kullanılmıştır. Uygulanan sayma veri modelleri sonucunda bağımlı değişkende aşırı yayılıma rastlanması üzerine uygun model olarak negatif binom regresyon modeli seçilmiştir.

Kılıç ve Bayrak (2019), TÜİK üzerinden alınan 2019 hanehalkı bilişim teknolojileri kullanımı araştırması mikro veri seti kullanılarak e-ticarete yaşanan sorun sayısını modellemek amacıyla sayma verisi ile regresyon modelleri kullanılmıştır. Çalışmada yapılan testler sonucunda ZINB modeli seçilmiştir.

Kıran (2019), TÜİK 2016 yılı Hanehalkı Bütçe Anket araştırması mikro veri setini modellemek amacıyla sayma veri modellerini kullanmıştır. Çalışmada bağımlı değişken olarak hanehalkının sahip olduğu konut sayısı, açıklayıcı değişken olarak ise yaş, cinsiyet, medeni durum, hanehalkı geliri, tasarruf, şans oyunu, iş durumu ve çalışma saati ele alınmıştır. Çalışmada poisson regresyon uygun model olarak ele alınmış ancak verideki sıfır yoğunluğundan dolayı sıfır yığılmalı poisson regresyon modeli uygulanmış ve poisson regresyon ile karşılaştırılması sonucunda ZIPR modeli tercih edilmiştir.

Güney ve Abar (2020), 2016 yılında Türkiye’de meydana gelen ekonomik suç sayılarının bölgesel belirleyicilerinin modellenmesi için sayma verisi ile regresyon modelleri kullanılmıştır. Bağımlı değişken olarak ekonomik suç sayıları, bağımsız değişken olarak ise yabancı göç, nüfus, coğrafi bölge, okuma yazma bilenlerin oranı, bin kişi başına düşen motorlu taşıt sayısı, tasarruf mevduat miktarı, toplam krediler, ortalama günlük kazanç, nüfus yoğunluğu, Sosyal Güvenlik Kurumu kapsamındaki nüfus oranı kullanılmıştır. Çalışma sonucunda NBR, sıfır kesilmiş modeller ve poisson modelleri uygulanmıştır.



3. BÖLÜM

3. MODEL SEÇİMİ

3.1. Uyum İyiliği Ölçümü

Modelden elde edilen sonuçların istatistiksel olarak uygunluğunun test edilmesi gerekmektedir. Modeli en iyi açıklayan modele karar verilmesi gerekmektedir. Bunun içinde model ile verilerin uygunluğuna bakılmalıdır. “Bu amaçla, kullanılan regresyon modellerinde normallik varsayımının sağlanmaması nedeniyle Pearson Ki-Kare, sapma istatistiği, yapay R^2 gibi uyum iyiliği ölçütlerinden yararlanılmaktadır.” (Özmen 1998: 1-109).

3.1.1. Genelleştirilmiş Pearson Ki-Kare Testi

Pearson ki-kare testi elde edilen modelle verilerin arasındaki uyumu araştırmak için kullanılan standart uyum iyiliği testidir. Genelleştirilmiş Ki-Kare test istatistiği;

$$\chi^2 = \frac{\sum(y - \hat{\mu})^2}{V(\hat{\mu})} \quad (3.1)$$

şeklinde tanımlanır.

$V(\hat{\mu})$ ilgili dağılımdan elde edilen varyans fonksiyonudur. n: örneklem büyüklüğü k: sabit ile birlikte açıklayıcı değişken sayısı gösterir.

n-k serbestlik derecesidir.

H_0 : Veriler belirlenen modele uyum sağlar.

H_1 : Veriler belirlenen modele uyum sağlamaz

Hesaplanan pearson ki-kare değeri $> \chi^2_{n-k}$ (tablo değeri) ise H_0 hipotezi reddedilir.

Hesaplanan pearson ki-kare değeri $< \chi^2_{n-k}$ (tablo değeri) ise H_0 hipotezi kabul edilir.

Hesaplanan değer tablo değerinden büyük ise model uyum sağlamamaktadır ve başka modeller denenmelidir. Küçük ise model uyum sağlar ve model geçerlidir. (Sarul 2012: 54-55)

3.1.2. Sapma İstatistiği

Sapma istatistiği uyum iyiliği ölçümü için kullanılan yöntemlerden biridir. Sapma istatistiği aynı zamanda “G kare istatistiği” olarak da adlandırılmaktadır.

$$G^2 = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\mu_i} \right) \quad (3.2)$$

şeklinde tanımlanmaktadır.

y_i : gözlenen bağımlı değişken, μ_i : beklenen bağımlı değişken, n : gözlem sayısını göstermektedir (Deniz 2005: 68-67).

- İstatistik değeri (G^2) eğer 0'a yaklaşıyorsa model uyumu artmaktadır.
- İstatistik değeri (G^2) eğer 0'a eşit ise model uyumu mükemmeldir.

3.1.3.Sözde R^2 Ölçümü (Pseudo R^2)

Standart regresyon analizinde R^2 birden çok bağımsız değişkenin bağımlı değişkeni açıklama miktarını vermektedir. Standart regresyon modelinde R^2 'nın hesaplanması genel kareler toplamının ayrıştırılmasına dayanır. Genel kareler toplamı;

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 + \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{\mu}_i)(\hat{\mu}_i - \bar{y}) \quad (3.3)$$

ile ifade edilir. Burada ilk üç toplam sırasıyla Genel Kareler Toplamı (GKT), Artık Kareler Toplamı (AKT) ve Açıklanmış Kareler Toplamı (RKT) olarak açıklanır. Son ifade eğer model sabit terim içeriyorsa, standart regresyon modelinin en küçük kareler kestirimine göre sifıra eşit olacaktır. Fakat Poisson modelini de kapsayan ve doğrusal olmayan en küçük kareler yöntemi ile üstel koşullu ortalamaya sahip olan tüm kestiricilerin ve modellerin sifıra eşit olmaması söz konusudur.

Bunun olması halinde R^2 'nin $R^2 = 1 - AKT/GKT$ ya da $R^2 = RKT/GKT$ 'den farklı şekilde hesaplanması gerekliliği söz konusudur (Cameron, Trivedi 1998: 153).

Poisson regresyon modelinde normallik varsayımını gerektirmeyen R^2 ölçüsü olabilirlik oran yaklaşımından gelmektedir. Doğrusal regresyon modeline ilişkin en küçük kareler tahmini, artık kareler toplamının en çok olabilirlik tahmini ve sapma değeri ile benzer özellikler göstermesi nedeniyle önerilen R^2 ölçüsü;

$$R_p^2 = 1 - \frac{\log L(y) - \log L(\hat{\mu})}{\log L(y) - \log L(\bar{y})} \quad (3.4)$$

şeklinde tanımlanmaktadır. Bu tanımlamada

- $\log L(y)$: modelin log- olabilirliğini
- $\log L(\hat{\mu})$: değerlendirilen modelin Log-olabilirliğini
- $\log L(\bar{y})$: sadece sabit terimin olduğu modelin Log-olabilirliğini

göstermektedir.

Gözlenen değerler, $y_i \geq 0$ $\hat{\mu}_i = \exp(x_i \hat{\beta})$ ya da $\hat{\mu}_i = c_i \exp(x_i \hat{\beta})$ tahmin edilen değerler ve $\bar{y}_i = \exp(\hat{\beta}_0)$ ya da $\bar{y}_i = c_i \exp(\hat{\beta}_0)$ ortalama değerler olmak üzere log-olabilirlik fonksiyonları;

$$\log L(y) = \sum_{i=1}^n (y_i \log(y_i) - y_i - \log(y_i!)) \quad (3.5)$$

$$\log L(\hat{\mu}) = \sum_{i=1}^n (y_i \log(\hat{\mu}_i) - \hat{\mu}_i - \log(y_i!)) \quad (3.6)$$

$$\log L(\bar{y}) = \sum_{i=1}^n (y_i \log(\bar{y}_i) - \bar{y}_i - \log(y_i!)) \quad (3.7)$$

şeklinde elde edilir. Bu log-olabilirlik fonksiyonları çözümlenirse sözde R^2 ölçüsüne ulaşılmaktadır.

3.2 Model Seçim Kriterleri

Model seçiminde amaç, elde edilen modeller arasından en iyi performansa sahip modeli seçebilmektir. Bunun için çeşitli model seçim kriterlerinden faydalanılmaktadır.

Model seçim kriterleri ilk olarak 60'ların başı ve 70'lerin sonlarında Akaike'nin FPE (Akaike 1969) ve Mallow'un Cp testleri ile ortaya çıkmıştır. Kullback.-Leibler (Kullback, Leibler 1951) uyumsuzluğuna dayanan ve daha çok zaman serilerinde kullanılan Akaike Bilgi kriterinin (Akaike 1973: 1974) ortaya çıkmasıyla başlayan bu dönüm noktasında 70'lerde bilgi teorileri yaklaşımı görülmeye başlanmıştır. 70'lerin sonlarına gelindiğinde bilgi teorileri alanında çalışmalar artarak gelişmiş ve yeni kriterler doğmuştur (Şengün, Ucal 2006: 43).

3.2.1. Akaike Bilgi Kriteri (AIC)

Akaike bilgi kriteri, japon istatistikçi Hirotugu Akaike tarafından 1973 yılında alternatif modeller arasında en uygun modelin belirlenmesi için ortaya çıkarılmıştır. Modele değişken eklenmesi uyum iyiliğini artırabilir. Bu durumda AIC modele değişkenlerin dahil edilmesine karşın uyum iyiliğini sınırlamaya çalışmaktadır. Akaike bilgi kriteri;

$$AIC = -2 \ln[L(\hat{\theta}|veri)] + 2k \quad (3.8)$$

şeklinde tanımlanır.

k, sabit terim dahil parametre sayısını gösterir. $\ln L$ ise Log olabilirlik fonksiyonudur.

AIC' nin bazı özellikleri şöyle sıralanır:

- Model karşılaştırmalarında her zaman en düşük AIC değerini veren model tercih edilir.
- AIC sadece seçili örnek büyüklüğü içinde değil aynı zamanda seçili örnek büyüklüğü dışındaki gelecek tahmini içinde geçerlidir (Şengün, Ucal 2006: 46).

3.2.2. Bayesyen (Schwarz) Bilgi Kriteri (BIC)

Bayes bilgi kriteri, uygun model seçimi için kullanılan yöntemlerden biri olup AIC' deki gibi modele parametre eklenmesi sonucunda aşırı uyumluluğa sebebiyet verebilir. Bayes bilgi kriteri bu problemi parametre sayısı için bir ceza (penalty) terimi ekleyerek çözmektedir. BIC' de ceza süresi akaike bilgi kriterinden daha

büyüktür. BIC eşitliğin sağ tarafındaki örnek büyüklüğüne bağlı olan ikinci kısım itibariyle AIC'den farklılık gösterir.

$$BIC = -2 \ln L + p(\ln n) \quad (3.9)$$

p: parametre sayısı

n: gözlem sayısı

$\ln L$: log olabilirlik

3.3. Katsayıların Yorumlanması

Regresyon katsayılarının yorumlanması, regresyon analizi sürecindeki en önemli unsurlardan birisidir. Doğrusal modellerde koşullu ortalama $E(y_i|x_i) = x'_i\beta$ şeklinde hesaplanmaktadır. Burada yorumlamayı yapabilmek için β yalnız bırakılıp, $\partial E[y|x]/\partial x_j = \beta_j$ işlemi yapılır. Bu işleme örnek olarak, $\hat{\beta}_j = 0,2$ için “j’ inci bağımsız değişkende oluşan 1 birimlik değişim, koşullu ortalamada 0,2 birimlik artışa neden olur.” şeklinde yorum yapılır. Ancak Poisson modeli üstel bir yapıya sahiptir. Bu sebeple katsayıların yorumlanması daha farklıdır (McCullagh ve Nelder, 1983).

Üstel koşullu ortalama;

$$E(y_i|x_i) = \exp(x'_i\beta)$$

şeklinde tanımlandığından bahsetmiştik. x_j değeri için j’ inci bağımsız değişken olduğu düşünülür ise benzer işlemlerin gerçekleştirilmesiyle;

$$\frac{\partial E[y|x]}{\partial x_j} = \beta_j \exp(x'\beta) \quad (3.10)$$

sonucuna ulaşılır. Yorumlamayı yukarıda bahsettiğimiz örneği ele alırsak, $\hat{\beta}_j = 0,2$ ve $\exp(x'\hat{\beta}) = 2,5$ ise j’ inci bağımsız değişkendeki bir birimlik değişim, y bağımlı değişkeninde 0,5 birimlik artışa neden olacağı, eşitlikten hesaplanabilmektedir.



4. BÖLÜM

4.UYGULAMA

Çalışmanın bu bölümünde daha önce açıklanan sayma verisi modelleri 2 farklı veri setine uygulanmıştır. Bu iki ayrı veri setinin değişkenleri hakkında istatistiksel bilgiler verilmiştir. Çalışmanın ilerleyen kısmında 1.veri setine poisson regresyon, negatif binom regresyon ve genelleştirilmiş poisson regresyon modelleri uygulanmış olup en iyi model seçimi yapılmıştır. 2.veri setine ise poisson, negatif binom, genelleştirilmiş poisson, sıfır yığılmalı poisson ve sıfır yığılmalı negatif binom regresyonları uygulanmıştır. Uygulanan regresyon modelleri arasından en iyi model seçimi yapılmıştır. Tüm çalışmada güven düzeyi: $\alpha=0.05$ olarak alınmıştır.

Çalışmada STATA 16.0 ve EVIEWS 10 paket programları kullanılmıştır.

4.1. Birinci Veri Seti İçin Çalışma

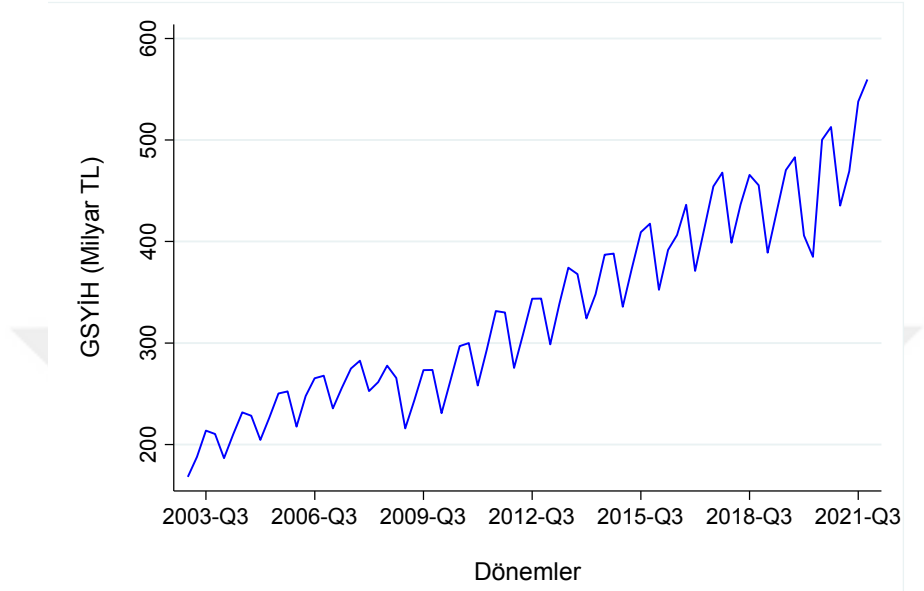
Çalışmada kullanılan 1. Veri setinde bağımlı değişken olarak otomobil talebini göstermek üzere satılan otomobil sayıları ele alınmıştır. Bağımsız değişkenler ise sırasıyla 1998 yılı sabit fiyatları alınarak milyar TL cinsinden Gayrisafi Yurt içi Hasıla (GSYİH) ve Araç Satın Alım Fiyat Endeksi (ASAFE) değişkenleridir. Uygulamada ASAFE üzerinden hesaplanmış değişim oranları kullanılmıştır. Veri seti TCMB (Türkiye Cumhuriyeti Merkez Bankası) elektronik veri dağıtım sistemi (EVDS) üzerinden alınmış olup 2003Q1-2021Q4 yılları arasındaki dönemi kapsamakta ve 76 gözlem sayısından oluşmaktadır.

4.1.1. Tanımlayıcı İstatistikler

1998 yılı sabit fiyatları ile GSYİH değişkeninin istatistiksel bilgileri Tablo 4.1 de verilmiştir.

Tablo 4.1. GSYİH Değişkenine Ait Tanımlayıcı İstatistikler

GSYİH (Milyar TL)								
Gözlem sayısı	Ortalama	Medyan	Maksimum	Minimum	Standart Sapma	Varyans	Çarpıklık	Basıklık
76	332.1955	327.1687	559.6613	168.1574	95.34949	9091.526	0.3673681	2.167072

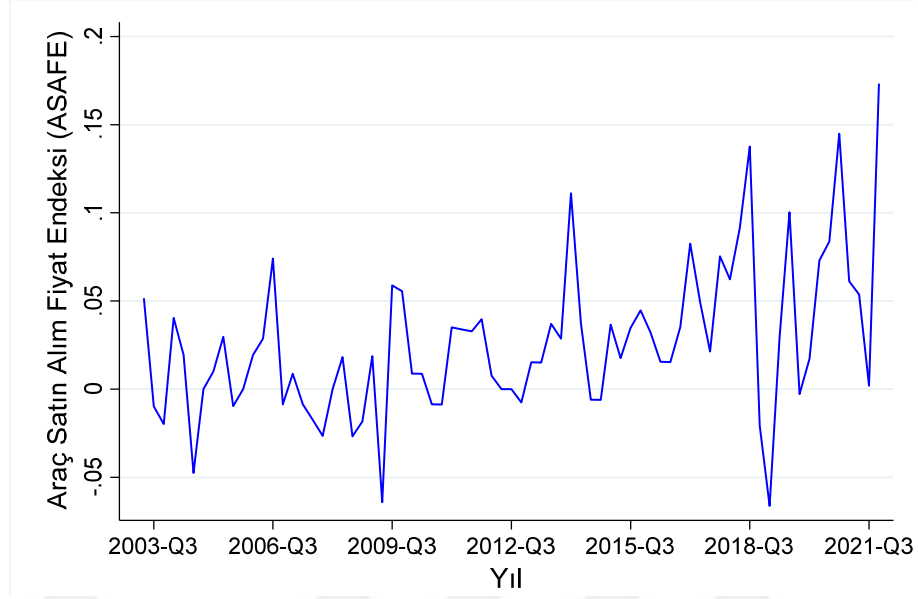


Şekil 4.1. GSYİH (1998 yılı sabit fiyatları) Verisinin Dönemlere Göre Grafiği

Araç Satın Alım Fiyat Endeksi (ASAFE) değişkenine ait istatistiksel bilgileri Tablo 4.2’ de verilmiştir.

Tablo 4.2. Araç Satın Alım Fiyat Endeksi (ASAFE) Değişkenine Ait Tanımlayıcı İstatistikler

Araç Satın Alım Fiyat Endeksi								
Gözlem sayısı	Ortalama	Medyan	Maksimum	Minimum	Standart Sapma	Varyans	Çarpıklık	Basıklık
76	0.0259992	0.018692	0.173258	-.0662651	.0435301	0.0018949	0.8917925	4.529003



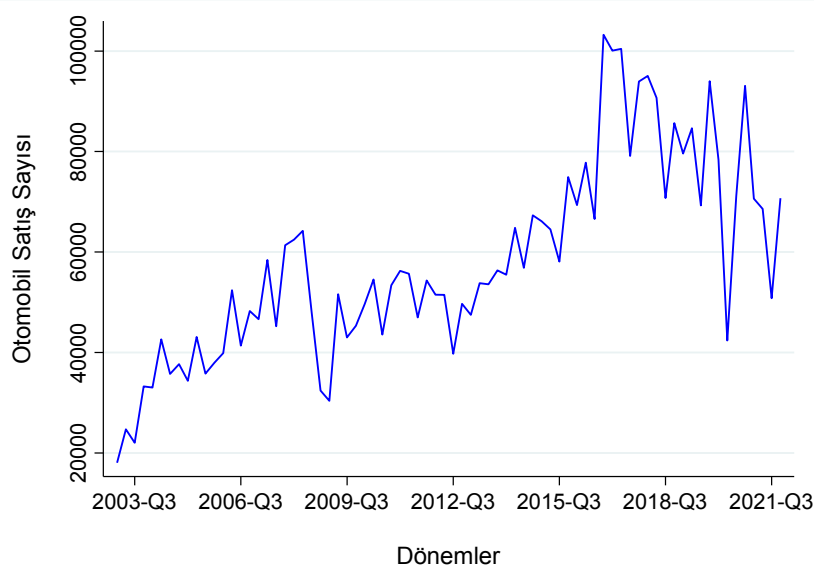
Şekil 4.2. Araç Satın Alımı Fiyat Endeksi (ASAFE) Verisinin Dönemlere Göre Grafiği

Otomobil satış sayısı bağımlı değişkenine ait tanımlayıcı istatistikler Tablo 4.3’de verilmiştir.

Tablo 4.3. Bağımlı Değişkene Ait Tanımlayıcı İstatistikler

Gözlem sayısı	Otomobil Satış Sayısı							
	Ortalama	Medyan	Maksimum	Minimum	Standart Sapma	Varyans	Çarpıklık	Basıklık
76	57913.96	54424.50	103236.0	18046.00	19727.46	3.89e+08	0.446480	2.648910

Tablo 4.3’e bakıldığında bağımlı değişkenin varyansının ($3.89e+08$) ortalamadan (57913.96) büyük olduğu görülmektedir. Bu durumda veride aşırı yayılım olduğuna dair bilgi vermektedir.



Şekil 4.3. Otomobil Satış Sayısının Dönemlere Göre Grafiği

4.1.2 Durağanlık Testi

Durağan olmayan zaman serilerinin kullanılması sonucunda regresyon analizlerinde seriler arasında gerçekte olmayan bir ilişki çıkabilmektedir. Bu duruma sahte (spurious) regresyon denilmektedir. Literatürde bilindiği gibi serilerin durağanlığını belirlemek için birçok birim kök testi bulunmaktadır. Bu testlerden en sık şekilde kullanılan Augmented Dickey-Fuller (ADF) testidir. Çalışmamızda 1. Veri setimizde zaman serisi kullandığımız için verilerimizin durağanlığını test ediyoruz

Tablo 4. 4. ADF (Durağanlık) Birim Kök Testi Sonuçları

Değişkenler	Düzy Değerler				Test Kritik Değerleri					
	Sabitli		Sabit+Trend		Sabitli			Sabit+Trend		
	Test İstatistiği	P*	Test İstatistiği	P*	%1	%5	%10	%1	%5	%10
Otomobil Satış Sayısı	-1.821	0.3701	-4.997	0.0006	-4.095	-3.475	-3.165	-4.095	-3.475	-3.165
GSYİH	0.812	0.9918	-1.936	0.6359	-3.555	-2.916	-2.593	-4.113	-3.483	-3.170
Araç Satın Alım Fiyat Endeksi	-5.667	0.0000	-7.263	0.0000	-3.546	-2.911	-2.590	-4.097	-3.476	-3.166

ADF birim kök testi sonuçlarına baktığımızda otomobil ve Araç Satın Alım Fiyat Endeksi (ASAFE) değişkenlerinin düzey durağan olduğu görülmektedir.

GSYİH ise düzey değerlerde durağan değildir. GSYİH birinci farkında durağan olmaktadır.

Yukarıdaki grafikler incelendiğinde yapısal kırılmalar gözlenmektedir. Bu nedenle yapısal kırılmalı birim kök testi uygulanmıştır.

Tablo 4.5. Kırılmalı Birim Kök Testi Sonuçları

	Düzey Değerler				Test Kritik Değerleri						Kırılma Dönemi
Değişkenler	Sabitli		Sabit+Trend		Sabitli			Sabit+Trend			
	Test İstatistiği	p*	Test İstatistiği	p*	%1	%5	%10	%1	%5	%10	
Otomobil Satış Sayısı	-5.2923	0.01	-5.7750	0.01	-4.9491	-4.443	-4.1936	-5.3475	-4.8598	-4.6073	2014Q1
GSYİH	-1.5797	0.99	-6.6795	0.01	-4.9491	-4.4436	-4.1936	-5.3475	-4.8598	-4.6073	2005Q4
Araç Satın Alım Fiyat Endeksi	-7.029	0.01	-8.0658	0.01	-4.9491	-4.4436	-4.1936	-5.3475	-4.8598	-4.6073	2016Q4

Tablo 4.5’de görüldüğü üzere her üç değişkende kırılmalı birim kök testi sonucunda durağanlaşmıştır.

4.1.3 Regresyon Analiz Sonuçları

4.1.3.1 Poisson Regresyon Analizi

Otomobil satış sayısı bağımlı değişkeni ve GSYİH, Araç Satın Alım Fiyat Endeksi (ASAFE) değişkeni üzerinden hesaplanan değişim oranları ile tahmini yapılan poisson regresyon analiz sonuçları Tablo 4.6’da verilmiştir.

Tablo 4.6. Poisson Regresyon Modeli Sonuçları

Log likelihood = -101215.35			LR istatistiği = 263934.97 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R ²) = 0.5659		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e ^β)	z-değeri	p-değeri
GSYİH	0.002633	5.76e-06	1.002636	457.14	0.000
Araç Satın Alım Fiyat Endeksi	-0.2849837	0.0118159	0.7520265	-24.12	0.000
Sabit	10.07328	0.0019587	23701.29	5142.77	0.000
Yayılım parametresi	0.0448474	0.0072725			

Sonuçlara göre açıklayıcı değişkenlerin katsayılarının her biri istatistiksel olarak anlamlıdır. (p<0,05) Analiz öncesinde bağımlı değişkenin tanımlayıcı istatistiklerinde (Tablo 4.3) varyansın ortalamadan büyük olduğunu belirtmiştik.

Literatürde bilindiği gibi poisson regresyonun eşit yayılım şartı bulunduğundan aşırı yayılım durumunda poisson regresyon yerine negatif binom regresyon ya da genelleştirilmiş poisson regresyon kullanılabilir.

Aşırı yayılım durumdan şüphelendiğimiz durumlarda aşırı yayılım testlerini kullanarak modelin uygunluğunu ölçebiliriz. Analizimizde aşırı yayılımın varlığını test etmek için olabilirlik oran testi (LR) $LR = 263934,97$ olarak hesaplanmıştır. LR testi p-değerine ($p < 0,05$) baktığımızda 0,05 ten küçük olduğundan H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. Böylece H_1 (aşırı yayılım vardır.) hipotezini kabul etmiş oluruz.

Poisson modeli için yayılım parametresi ($\alpha=0.0448474$) sıfırdan farklıdır. Yokluk hipotezinin reddedilir. Bu durumunda aşırı yayılım ortaya çıkacak ve PR yerine NBR veya GPR kullanılması daha uygun olacaktır. Bu sebeple veri setimize aşırı yayılımı dikkate alan NBR ve GPR modelleri uygulanmıştır.

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

Tablo 4.7. Sapma Uyum İyiliği Sonucu

Sapma İstatistiği	201473.8
Prob Değeri	0.0000

Tablo 4.7.'ye bakıldığında prob değeri 0,05 ten küçük olduğundan H_0 (Sapma istatistiği değeri sıfıra eşitse değilse model uygundur.) hipotezi reddedilir. Böylece H_1 (Sapma istatistiği değeri sıfıra eşitse değilse model uygun değildir.) hipotezi kabul etmiş oluruz. Veriler poisson regresyona uygun değildir.

Tablo 4.8. Pearson Ki-Kare Uyum İyiliği Sonucu

Pearson χ^2 Uyum İyiliği İstatistiği	200310.6
Prob Değeri	0.0000

Tablo 4.8'e bakıldığında prob değeri 0,05 ten küçük olduğundan H_0 (Veriler belirlenen modele uyum sağlar.) hipotezi reddedilir. Böylece H_1 (Veriler belirlenen modele uyum sağlamaz) hipotezini kabul etmiş oluruz. Veriler poisson regresyon modeline uygun değildir.

4.2.3.2 Negatif Binom Regresyon Analizi

Otomobil satış sayısı bağımlı değişkeni ve GSYİH, Araç Satın Alım Fiyat Endeksi (ASAFE) değişkeni üzerinden hesaplanan değişim oranları ile tahmini yapılan negatif binom regresyon analiz sonuçları Tablo 4.9.'da verilmiştir.

Tablo 4.9. Negatif Binom Regresyon Modeli Sonuçları

Log likelihood = -809.65571			LR istatistiği = 67.25 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R^2) = 0.0399		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0028642	0.0003097	1.002868	9.25	0.000
Araç Satın Alım Fiyat Endeksi	-0.3109525	0.6583685	0.7327487	-0.47	0.637
Sabit	9.993982	0.0996185	21894.31	100.32	0.000
Yayılım parametresi	0.0448474	0.0072725			
$\alpha=0$ için LR testi $\chi^2(01) = 2.0e+05$			P-değeri $\geq \chi^2 = 0.000$		

NBR sonuçlarına göre açıklayıcı değişkenlerin katsayılarına baktığımızda Araç Satın Alım Fiyat Endeksi (ASAFE) istatistiksel olarak anlamsız ($p > 0,05$) olduğunu diğer değişkenin ise istatistiksel olarak anlamlı ($p < 0,05$) olduğunu görmekteyiz.

Negatif binom regresyon modeli için LR testi ($LR=67.25$) p-değerine baktığımızda ($p < 0,05$) H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. H_1 (aşırı yayılım vardır.) hipotezini kabul etmiş oluruz. Bu durumda aşırı yayılım ortaya çıkacak ve veriler negatif binom regresyon kullanmak için uygun olacaktır.

NBR modeli için α parametresi ($\alpha=0.0448474$) sıfırdan farklıdır. Böylece yokluk hipotezi reddedilir ve bu durumunda aşırı yayılım ortaya çıkacaktır. Bu sebeple veriler negatif binom regresyon için uygundur.

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa

modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

4.1.3.3 Genelleştirilmiş Poisson Regresyon Analizi

Otomobil satış sayısı bağımlı değişkeni ve GSYİH, Araç Satın Alım Fiyat Endeksi (ASAFE) değişkeni üzerinden hesaplanan değişim oranları ile tahmini yapılan genelleştirilmiş poisson regresyon analiz sonuçları Tablo 4.10' da verilmiştir.

Tablo 4.10. Genelleştirilmiş Poisson Regresyon Modeli Sonuçları

Log likelihood = -813.2675			LR istatistiği = 60.78 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R^2) = 0.0360		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.002415	0.0002716	1.002418	8.89	0.000
Araç Satın Alım Fiyat Endeksi	-0.152729	0.5899903	0.8583623	-0.26	0.796
Sabit	10.14695	0.0926244	2363.153	109.55	0.000
Yayılım parametresi	0.9815422	0.0015832			
$\phi=0$ için LR testi $\chi^2(1) = 2.0e+05$ p-değeri $\geq \chi^2 = 0.0000$					

GPR sonuçlarına göre açıklayıcı değişkenlerin katsayılarına baktığımızda Araç Satın Alım Fiyat Endeksi (ASAFE) istatistiksel olarak anlamsız ($p > 0,05$) olduğunu diğer değişkenin ise istatistiksel olarak anlamlı ($p < 0,05$) olduğunu görmekteyiz.

Genelleştirilmiş poisson regresyon modeli için LR testi ($LR=60.78$) p-değerine baktığımızda ($p < 0,05$) H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. H_1 (aşırı yayılım vardır.) H_1 hipotezini kabul etmiş oluruz. Bu durumda aşırı yayılım ortaya çıkacak ve veriler genelleştirilmiş poisson regresyon kullanmak için uygun olacaktır.

GPR modeli için ϕ parametresi ($\phi=0.9815422$) birden farklıdır. Böylece yokluk hipotezi reddedilir ve bu durumunda aşırı yayılım ortaya çıkacaktır. Bu sebeple veriler genelleştirilmiş poisson modeline uygundur.

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

4.1.3.4 Model Seçimi

Yukarıda bahsedilen 1. Veri setine yapılan uyum iyiliği testleri sonrasında en iyi modeli seçebilmek için bilgi kriterleri (AIC, BIC) ve log-olabilirlik değerlerine bakılmalıdır.

Tablo 4.11. Birinci Veri Seti İçin Model Seçim Kriterleri

	AIC	BIC	Log Likelihood
Poisson Regresyon	202436.7	202443.7	-101215.35
Negatif Binom Regresyon	1627.311	1636.581	-809.65571
Genelleştirilmiş Poisson Regresyon	1634.535	1643.805	-813.2675

En iyi modeli seçebilmek için AIC ve BIC değerlerinin en küçük olduğu, log-olabilirlik değerinin ise en büyük olduğu modelin ele alınması gerekmektedir. Veride aşırı yayılım olduğundan poisson modeli dikkate almadan baktığımızda en uygun model olarak seçilmesi gereken model, Tablo 4.12’de görüldüğü üzere en küçük AIC ve BIC değerlerine ve en büyük log-olabilirlik değerine sahip olan Negatif Binom regresyon modelidir.

4.1.3.5. Uygun Model İçin Katsayı Yorumlama

Katsayıları yorumlamak için regresyon modelinin üstel tahmin (e^β) değerleri tahmin edilmelidir. Tablo 4.12’ e bakıldığında IRR (e^β) sonuçları görülmektedir. Katsayıları yorumlarken katsayı değeri eğer sıfırsa ya da istatistiksel olarak anlamsız ise $e^\beta=1$ olacaktır. Açıklayıcı değişkenin bağımlı değişkene olan etkisi pozitif yönde ise $e^\beta>1$, negatif yönde ise $e^\beta<1$ olacaktır.

Tablo 4.12. Katsayı Yorumları İçin Negatif Binom Regresyon Modeli Sonuçları

Log likelihood = -809.65571			LR istatistiği = 67.25 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R ²) = 0.0399		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0028642	0.0003097	1.002868	9.25	0.000
Araç Satın Alım Fiyat Endeksi	-0.3109525	0.6583685	0.7327487	-0.47	0.637
Sabit	9.993982	0.0996185	21894.31	100.32	0.000
Yayılım parametresi	0.0448474	0.0072725			
$\alpha=0$ için LR testi $\chi^2(01) = 2.0e+05$			P-değeri $\geq \chi^2 = 0.000$		

Model sonuçlarına bakıldığında otomobil satış sayıları üzerinde GSYİH değişkeninin pozitif yönde ve istatistiksel olarak anlamlı ($p<0,05$) bir etki gösterdiği

görülmüştür. Araç Satın Alım Fiyat Endeksi (ASAFE) değişkeni ise otomobil değişkeni üzerinde negatif yönde ve istatistiksel olarak anlamsız ($p>0,05$) bir etki göstermektedir. Ancak Normal şartlarda fiyat artışlarının otomobil satışlarını azaltması gerekirken ekonomik dengesizlikler sebebiyle, otomobil fiyatlarının artacağı beklentisi araç satış fiyat endeksinin anlamsız çıkmasına sebep olmuştur. GSYİH’ da her bir milyar artış olduğunda otomobil satışındaki olasılığı ($e^{0.0028} \sim 1.0028$) % 0.3 artırmaya sebep olduğu söylenebilir.

4.2. İkinci Veri Seti İçin Çalışma

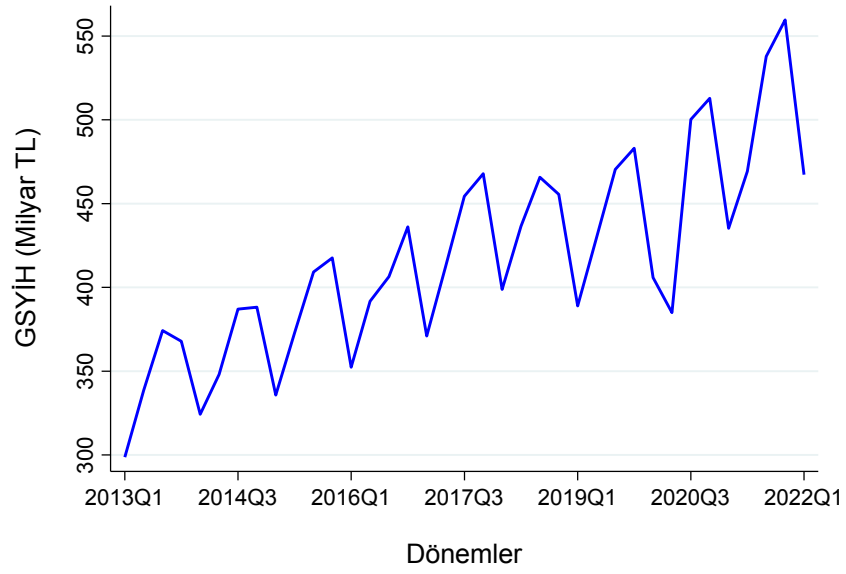
Çalışmada kullanılan 2. Veri setinde bağımlı değişken olarak konut talebini göstermek üzere Türkiye konut satış sayıları ele alınmıştır. Bağımsız değişkenler ise sırasıyla 1998 yılı sabit fiyatları alınarak milyar TL cinsinden Gayrisafi Yurt içi Hâsıla (GSYİH) ve Konut fiyat endeksi (KFE) değişkenleridir. Veri seti TCMB (Türkiye Cumhuriyeti Merkez Bankası) elektronik veri dağıtım sistemi (EVDS) üzerinden alınmış olup 2013Q1-2022Q1 yılları arasındaki dönemi kapsamakta ve 37 gözlem sayısından oluşmaktadır.

4.2.1. Tanımlayıcı İstatistikler

1998 sabit fiyatları ile GSYİH değişkeninin istatistiksel bilgileri tablo 4.14’de verilmiştir.

Tablo 4.13. GSYİH Değişkenine Ait Tanımlayıcı İstatistikler

Gözlem sayısı	GSYİH (Milyar TL)					Standart Sapma	Varyans	Çarpıklık	Basıklık
	Ortalama	Medyan	Maksimum	Minimum					
37	417.7491	409.1662	559.6613	298.6855	60.08664	3610.405	0.3040874	2.676985	

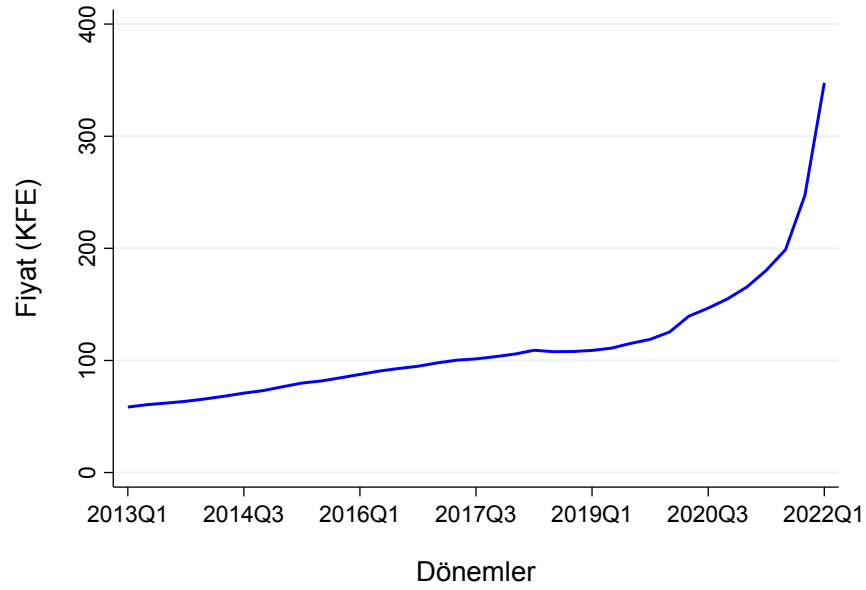


Şekil 4.4. GSYİH (1998 yılı sabit fiyatları) Verisinin Dönemlere Göre Grafiği

Konut fiyat endeksi değişkeninin istatistiksel bilgileri Tablo 4.15’de verilmiştir.

Tablo 4.14. Fiyat (Konut Fiyat Endeksi) Değişkenine Ait Tanımlayıcı İstatistikler

Gözle m sayısı	Fiyat (KFE)							
	Ortalama	Medyan	Maksimum	Minimum	Standart Sapma	Varyans	Çarpıklık	Basıklık
37	113.6405	101.0000	347.7	58.4	56.96595	3245.119	2.325305	9.375008



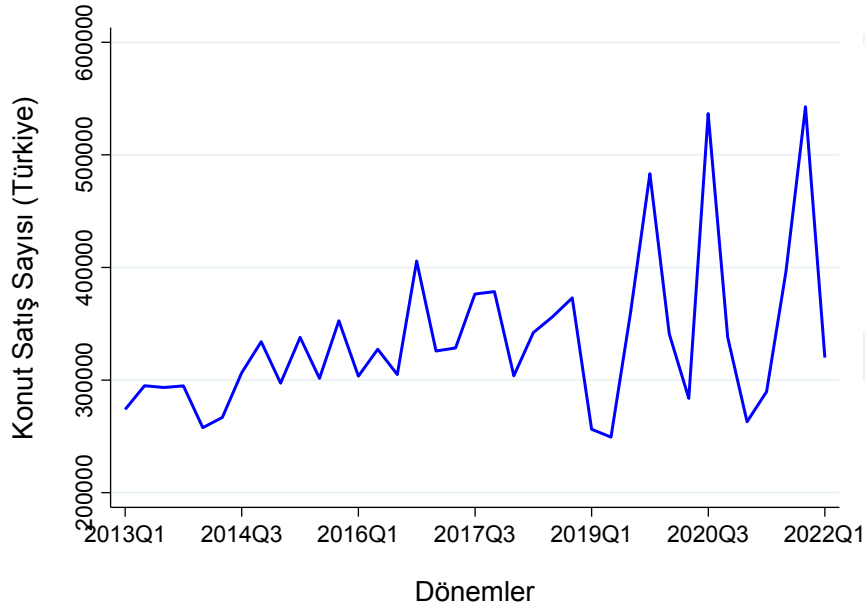
Şekil 4.5 Fiyat (Konut Fiyat Endeksi)Verisinin Dönemlere Göre Grafiği

Türkiye konut satış sayısı değişkeninin istatistiksel bilgileri Tablo 4.16’de verilmiştir.

Tablo 4. 15. Türkiye Konut Satış Sayısı Değişkenine Ait İstatistikler

Gözlem sayısı	Konut Satış Sayısı (Türkiye)							
	Ortalama	Medyan	Maksimum	Minimum	Standart Sapma	Varyans	Çarpıklık	Basıklık
37	335081.6	325780.0	542718	249363	68772.86	4.73e+09	1.563605	5.425944

Tablo 4.16’ya bakıldığında bağımlı değişkenin varyansının (4.73e+09) ortalamadan (335081.6) büyük olduğu görülmektedir. Bu durumda veride aşırı yayılım olduğuna dair bilgi vermektedir.



Şekil 4.6 Türkiye Konut Satış Sayısının Dönemlere Göre Grafiği

4.2.2. Durağanlık Testi

Çalışmamızın 4.1.2’de bahsettiğimiz gibi zaman serilerinde serinin durağanlığını ADF birim kök testi ile test ediyoruz.

Tablo 4. 16. ADF (Durağanlık) Birim Kök Testi Sonuçları

Değişkenler	Düzey Değerler					
	Sabitli		Sabit+Trend		Sabitli+Trendsiz	
	Test İstatistiği	P*	Test İstatistiği	P*	Test İstatistiği	P*
Konut Sayısı	-5.094334	0.0002	-6.9589	0.0000	1.9586	0.9858
GSYİH	0.610609	0.9879	-3.0330	0.1403	4.4693	1.0000
Fiyat (KFE)	0.872035	0.9935	5.5540	1.0000	-3.912006	0.0003

ADF birim kök testi sonuçlarına baktığımızda konut sayısı sabitli ve sabit+trend düzey değerlerinde durağandır. Fiyat (KFE) değişkeninde sabitsiz ve trend olmayan düzey değerlerde durağandır. GSYİH değişken düzeyde durağan olmadığı görülmektedir ve birinci farkında durağanlaşmaktadır.

Yukarıdaki grafikler incelendiğinde yapısal kırılmalar gözlemlenmektedir. Bu nedenle kırılmalı birim kök testi uygulanmıştır.

Tablo 4.17. Kırılmalı Birim Kök Testi Sonuçları

	Düzey Değerler				Test Kritik Değerleri						Kırılma Dönemi
Değişkenler	Sabitli		Sabit+Trend		Sabitli			Sabit+Trend			
	Test İstatistiği	p*	Test İstatistiği	p*	%1	%5	%10	%1	%5	%10	
Konut Sayısı	-6.4985	0.01	-7.9569	0.01	-4.9491	-4.443	-4.193	-5.347	-4.859	-4.607	2019Q1
GSYİH	-3.7851	0.2426	-8.3225	0.01	-4.9491	-4.443	-4.193	-5.347	-4.859	-4.607	2020Q1
Fiyat (KFE)	-1.10533	0.99	-2.4939	0.98	-4.9491	-4.443	-4.193	-5.347	-4.859	-4.607	2021Q4

Tablo 4.17’de görüldüğü üzere konut sayısı ve GSYİH değişkenleri kırılmalı birim kök testi sonucunda durağanlaşmıştır.

4.2.3. Regresyon Analiz Sonuçları

4.2.3.1. Poisson Regresyon Analizi

Türkiye konut satış sayısı bağımlı değişkeni ve 1998 sabit fiyatları ile GSYİH, Fiyat (KFE) açıklayıcı değişkenleri ile tahmini yapılan poisson regresyon analiz sonuçları Tablo 4.18’de verilmiştir.

Tablo 4.18. Poisson Regresyon Modeli Sonuçları

Log likelihood = -101449.23			LR istatistiği = 266616.56 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R ²) = 0.5679		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0029554	6.54e-06	1.00296	451.91	0.000
Fiyat	-0.0009064	6.96e-06	0.999094	-130.15	0.000
Sabit	11.57983	0.0023157	106919.5	5000.48	0.000
Yayılım parametresi	0.0151042	0.0035035			

Sonuçlara göre açıklayıcı değişkenlerin katsayılarının her biri istatistiksel olarak anlamlıdır. ($p < 0,05$) Analiz öncesinde bağımlı değişkenin tanımlayıcı istatistiklerinde (Tablo 4.16) varyansın ortalamadan büyük olduğunu belirtmiştik. Literatürde bilindiği gibi Poisson regresyonun eşit yayılım şartı bulunduğu aşırı yayılım durumunda poisson regresyon yerine Negatif Binom regresyon ya da Genelleştirilmiş Poisson regresyon kullanılabilir.

Aşırı yayılım durumdan şüphelendiğimiz durumlarda aşırı yayılım testlerini kullanarak modelin uygunluğunu ölçebiliriz. Analizimizde aşırı yayılımın varlığını test etmek için olabilirlik oran testi (LR) $LR = 266616.56$ olarak hesaplanmıştır. LR testi p-değerine ($p < 0,05$) baktığımızda 0,05 ten küçük olduğundan H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. Böylece H_1 (aşırı yayılım vardır.) hipotezini kabul etmiş oluruz.

Poisson modeli için yayılım parametresi ($\alpha = 0.0151042$) sıfırdan farklıdır. Yokluk hipotezinin reddedilir. Bu durumda aşırı yayılım ortaya çıkacak ve PR yerine NBR veya GPR kullanılması daha uygun olacaktır. Bu sebeple veri setimize aşırı yayılımı dikkate alan NBR ve GPR modelleri uygulanmıştır.

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa

modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

Tablo 4. 19. Sapma Uyum İyiliği Sonucu

Sapma İstatistiği	202360.4
Prob Değeri	0.0000

Tablo 4.19' e bakıldığında prob değeri 0,05 ten küçük olduğundan H_0 (Sapma istatistiği değeri sıfıra eşitse değilse model uygundur.) hipotezi reddedilir. Böylece H_1 (Sapma istatistiği değeri sıfıra eşitse değilse model uygun değildir.) hipotezi kabul etmiş oluruz. Veriler poisson regresyona uygun değildir.

Tablo 4. 20. Pearson Ki-Kare Uyum İyiliği Sonucu

Pearson χ^2 Uyum İyiliği İstatistiği	202030
Prob Değeri	0.0000

Tablo 4.20'ye bakıldığında prob değeri 0,05 ten küçük olduğundan H_0 (Veriler belirlenen modele uyum sağlar.) hipotezi reddedilir. Böylece H_1 (Veriler belirlenen modele uyum sağlamaz) hipotezini kabul etmiş oluruz. Veriler poisson regresyon modeline uygun değildir.

4.2.3.2 Negatif Binom Regresyon Analizi

Türkiye konut satış sayısı bağımlı değişkeni ve 1998 sabit fiyatları ile GSYİH, Fiyat (KFE) açıklayıcı değişkenleri ile tahmini yapılan negatif binom regresyon analiz sonuçları Tablo 4.21. de verilmiştir.

Tablo 4. 21. Negatif Binom Regresyon Modeli Sonuçları

Log likelihood = -445.09146			LR istatistiği = 31.73 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R^2) = 0.0344		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0028425	0.0004557	1.002847	6.24	0.000
Fiyat (KFE)	-0.0009006	0.0004831	0.9990998	-1.86	0.062
Sabit	11.62682	0.1593192	112063	72.98	0.000
Yayılım parametresi	0.0151042	0.0035035			
$\alpha=0$ için LR testi $\chi^2(01) = 2.0e+05$			P-değeri $\geq \chi^2 = 0.000$		

NBR sonuçlarına göre açıklayıcı değişkenlerin katsayılarına baktığımızda Fiyat değişkeni ($p > 0,06$) anlamlılık düzeyini göre istatistiksel olarak anlamlıdır. GSYİH değişkeninin ise istatistiksel olarak anlamlı ($p < 0,05$) olduğunu görmekteyiz.

Negatif binom regresyon modeli için LR testi ($LR = 31.73$) p-değerine baktığımızda ($p < 0,05$) H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. H_1 (aşırı yayılım vardır.) hipotezini kabul etmiş oluruz. Bu durumda aşırı yayılım ortaya çıkacak ve veriler negatif binom regresyon kullanmak için uygun olacaktır.

NBR modeli için α parametresi ($\alpha = 0.0151$) sıfırdan farklıdır. Böylece yokluk hipotezi reddedilir ve bu durumunda aşırı yayılım ortaya çıkacaktır. Bu sebeple veriler aşırı yayılıma sahiptir ve negatif binom regresyon için uygundur.

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

4.2.3.3 Genelleştirilmiş Poisson Regresyon Analizi

Türkiye konut satış sayısı bağımlı değişkeni ve 1998 sabit fiyatları ile GSYİH, Fiyat (KFE) açıklayıcı değişkenleri ile tahmini yapılan genelleştirilmiş poisson regresyon analiz sonuçları Tablo 4.22.'de verilmiştir.

Tablo 4. 22. Genelleştirilmiş Poisson Regresyon Modeli Sonuçları

Log likelihood = -446.85272			LR istatistiği = 25.77 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R^2) = 0.0280		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0027552	0.000499	1.002759	5.52	0.000
Fiyat (KFE)	-0.0009142	0.0005342	0.9990863	-1.71	0.087
Sabit	11.66599	0.1770843	116539.5	65.88	0.000
Yayılım parametresi	0.9866268	0.0015805			
$\phi=0$ için LR testi $\chi^2(1) = 2.0e+05$ p-değeri $\geq \chi^2 = 0.0000$					

GPR sonuçlarına göre açıklayıcı değişkenlerin katsayılarına baktığımızda GSYİH ile sabit değişkenin istatistiksel olarak anlamlı ($p < 0,05$) olduğunu görmekteyiz. Fiyat değişkeninin ise istatistiksel olarak anlamsız ($p > 0,05$) olduğu görülmektedir.

Genelleştirilmiş poisson regresyon modeli için LR testi (LR=25.77) p-değerine baktığımızda ($p < 0,05$) H_0 (aşırı yayılım yoktur.) hipotezini reddederiz. H_1 (aşırı yayılım vardır.) hipotezini kabul etmiş oluruz. Bu durumda aşırı yayılım ortaya çıkacak ve veriler genelleştirilmiş poisson regresyon kullanmak için uygun olacaktır.

GPR modeli için ϕ parametresi ($\phi=0.9866268$) birden farklıdır. Böylece yokluk hipotezi reddedilir ve bu durumda aşırı yayılım ortaya çıkacaktır. Bu sebeple veriler genelleştirilmiş poisson modeline uygundur

Sözde R^2 değeri bağımlı değişkenin içindeki varyasyonun yüzde kaçının bağımsız değişken tarafından gösterildiğini açıklamaktadır. 0 ile 1 arasında değer almalıdır. Modelin sözde R^2 değeri ne kadar büyük olursa yani bire yaklaşıyorsa modelin uyumu o kadar iyidir. Eğer sözde R^2 değeri sıfıra yaklaşıyorsa model uyumu iyi değildir.

4.2.3.4 Model Seçimi

Yukarıda bahsedilen 2. Veri setine yapılan uyum iyiliği testleri sonrasında en iyi modeli seçebilmek için bilgi kriterleri (AIC, BIC) ve log-olabilirlik değerlerine bakılmalıdır.

Tablo 4.23. İkinci Veri Seti İçin Model Seçim Kriterleri

	AIC	BIC	Log Likelihood
Poisson Regresyon	202904.5	202909.3	-101449.23
Negatif Binom Regresyon	898.1829	904.6266	-445.09146
Genelleştirilmiş Poisson Regresyon	901.7054	908.1491	-446.85272

En iyi modeli seçebilmek için AIC ve BIC değerlerinin en küçük olduğu, log-olabilirlik değerinin ise en büyük olduğu modelin ele alınması gerekmektedir. Veride aşırı yayılım olduğundan poisson modeli dikkate almadan baktığımızda en uygun model olarak seçilmesi gereken model, Tablo 4.23’de görüldüğü üzere en küçük AIC ve BIC değerlerine ve en büyük log-olabilirlik değerine sahip olan Negatif Binom regresyon modelidir.

4.2.3.5. Uygun Model İçin Katsayı Yorumlama

Katsayıları yorumlamak için regresyon modelinin üstel tahmin (e^β) değerleri tahmin edilmelidir. Tablo 4.24’ e bakıldığında IRR (e^β) sonuçları görülmektedir.

Katsayıları yorumlarken katsayı değeri eğer sıfırsa ya da istatistiksel olarak anlamsız ise $e^\beta=1$ olacaktır. Açıklayıcı değişkenin bağımlı değişkene olan etkisi pozitif yönde ise $e^\beta>1$, negatif yönde ise $e^\beta<1$ olacaktır.

Tablo 4.24. Katsayı Yorumu İçin Negatif Binom Regresyon Modeli Sonuçları

Log likelihood = -445.09146			LR istatistiği = 31.73 P-değeri (LR istatistiği) = 0.0000 Pseudo (Sözde R^2) = 0.0344		
Açıklayıcı değişkenler	Katsayı(β)	Standart Hata	IRR (e^β)	z-değeri	p-değeri
GSYİH	0.0028425	0.0004557	1.002847	6.24	0.000
Fiyat (KFE)	-0.0009006	0.0004831	0.9990998	-1.86	0.062
Sabit	11.62682	0.1593192	112063	72.98	0.000
Yayılm parametresi	0.0151042	0.0035035			
$\alpha=0$ için LR testi $\chi^2(01) = 2.0e+05$			P-değeri $\geq \chi^2 = 0.000$		

Model sonuçlarına bakıldığında konut satış sayıları üzerinde GSYİH değişkeninin pozitif yönde ve istatistiksel olarak anlamlı ($p<0,05$) bir etki gösterdiği görülmüştür. Fiyat (KFE) değişkeni ise otomobil değişkeni üzerinde negatif yönde ve istatistiksel olarak anlamsız ($p>0,05$) bir etki göstermektedir. GSYİH'nın konut satışındaki olasılığı ($e^{0.0028} \sim 1.0028$) % 0.3 artırmıştır. Fiyat (Konut Fiyat Endeksi) değişkeninin bir birimlik artışı konut satışını ($e^{-0.0009} \sim 0.99$) % 1 oranında azaltmıştır.

SONUÇ

Sayma verisi ile regresyon modelleri, sayım yoluyla elde edilen verilerin modellenmesi için kullanılmaktadır. Bu modellerde bağımlı değişkenin dağılımı büyük önem göstermektedir. Bağımlı değişkenin dağılımı Poisson ya da Negatif Binom dağılımlarından gelmektedir. Sayma verisi ile regresyon modelinin temelini poisson regresyona dayandığından bahsetmiştik. Poisson regresyonunun kullanılabilmesi için eşit yayılım şartının sağlanması gerekmektedir. Yani ortalamanın varyansa eşit olması gerekir. Ancak bu şartın sağlanamadığı durumlar yaygındır. Bu durumda Poisson regresyon yerine Negatif Binom ya da genelleştirilmiş Poisson regresyon kullanılmaktadır. Ayrıca bağımlı değişkenin sıfır yoğunluğuna sahip olması durumunda bunun için geliştirilen ZIPR, ZINB ve ZIGP modelleri uygulanmalıdır.

Bu çalışmada iki ayrı veri seti ele alınmış STATA ve EVIEWS paket programlarından yararlanılarak analizler yapılmıştır. Birinci veri seti TCMB elektronik veri dağıtım sistemi üzerinden alınmıştır. Bağımlı değişken olarak otomobil talebini göstermek üzere satılan otomobil sayıları ele alınmıştır. Bağımsız değişkenler ise sırasıyla 1998 yılı sabit fiyatları alınarak milyar TL cinsinden Gayrisafi Yurt içi Hâsıla (GSYİH) ve Araç Satın Alımı Fiyat Endeksi (ASAFE) değişkeni üzerinden hesaplanan değişim oranlarıdır. İkinci veri seti de TCMB elektronik veri dağıtım sistemi üzerinden alınmıştır ve bağımlı değişken olarak konut talebini göstermek üzere Türkiye konut satış sayıları ele alınmıştır. Bağımsız değişkenler ise sırasıyla 1998 yılı sabit fiyatları alınarak milyar TL cinsinden Gayrisafi Yurt içi Hâsıla (GSYİH) ve Konut fiyat endeksi (KFE) değişkenleridir.

Uygulamada ilk olarak birinci veri setine dair analizler yapılmış ve bağımlı değişkenin istatistiksel bilgileri verilmiştir. Bağımlı değişkenin incelenmesi sonucunda varyansının ($3.89e+08$) ortalamadan (57913.96) büyük olduğu görülmüş değişkende aşırı yayılım varlığından şüphelenilmiştir. İlk olarak veri setine Poisson regresyon uygulanmış ve tüm değişkenler istatistiksel olarak anlamlı bulunmuştur. Ancak Poisson regresyon uyum testlerine bakıldığında modelin veri seti için uygun olmadığı görülmüş ve bunun üzerine aşırı yayılım durumunda poisson yerine

uygulanabilecek olan Negatif Binom regresyon ve Genelleştirilmiş Poisson regresyon modelleri uygulanmıştır. Yapılan model seçimine göre en uygun model Negatif Binom regresyon modeli seçilmiştir.

Uygulamada daha sonra ikinci veri setini dair analizler yapılmıştır. Aynı şekilde bağımlı değişkenin istatistiksel bilgileri verilmiş ve varyansının ($4.73e+09$) ortalamadan (335081.6) büyük olduğu görülmüş değişkende aşırı yayılımdan şüphelenilmiştir. İlk olarak veri setine Poisson regresyon uygulanmış ve tüm değişkenler istatistiksel olarak anlamlı bulunmuştur. Ancak Poisson regresyon uyum testlerine bakıldığında modelin veri setini uygun olmadığı görülmüş ve bunun üzerine aşırı yayılım durumunda Poisson yerine uygulanabilecek olan Negatif Binom regresyon ve Genelleştirilmiş Poisson regresyon modelleri uygulanmıştır. Yapılan model seçimine göre en uygun model Negatif Binom regresyon modeli tahmin edilmiştir.

Çalışmada birinci veri seti için seçilen modelin sonuçlarına bakıldığında, GSYİH'nin otomobil satışlarını pozitif yönde etkilediği ve otomobil satışını % 0.3 oranında arttırdığı sonucuna ulaşılmıştır. Araç Satın Alımı Fiyat Endeksi (ASAFE) değişkeninin otomobil satış sayısına etkisinin ise negatif yönde olduğu görülmüştür. Ancak normal şartlarda fiyat artışlarının otomobil satışlarını azaltması beklenirken, son dönemlerde oluşan ekonomik sorunlar ve tüketicinin otomobil fiyatlarının daha da artacağı beklentisi nedeniyle otomobile olan talebin yükselmesi, araç satın alım fiyat endeksinin istatistiksel olarak anlamsız çıkmasına sebep olmuştur. İkinci veri seti için seçilen modelin sonuçlarına göre GSYİH'nin konut satış sayılarına pozitif etki gösterdiği ve konut satışlarını % 0.3 oranında arttırdığı sonucuna ulaşılmıştır. Konut Fiyat Endeksi (KFE) değişkeninin konut satış sayısına etkisi ise negatif yönde etkilediği ve konut satışlarını % 1 oranında azalttığı sonucuna ulaşılmıştır.

KAYNAKÇA

- Arı Arzu ve Önder Hasan (2013). "Farklı veri yapılarında kullanılabilir regresyon yöntemleri". *Anadolu Tarım Bilimleri Dergisi* 28.3: 168-174.
- Avcı, Aziz (2020). *Sayma Verisi Regresyon Modellerini Kullanarak Aile Hekimine Müracaat Sayısını Etkileyen Faktörlerin Analizi*. Yayınlanmamış Doktora Tezi. Atatürk Üniversitesi. Biyoistatistik ve Tıp Bilişimi Anabilim Dalı. Erzurum.
- Akaike, Hirotugu (1969). "Fitting autoregressive models for prediction". *Annals of the institute of Statistical Mathematics* Vol.21 No.1 Page: 243-247.
- Akaike, Hirotugu (1973). *Information theory and an extension of maximum likelihood principle*. Proc. 2nd Int. Symp. on Information Theory. Page: 267-28 Budapest: Akademiai Kiado
- Akaike, Hirotugu (1974). "A new look at the statistical model identification." *IEEE transactions on automatic control* Vol.19 No.6 Page: 716-723.
- Cragg, John G (1971). "Some statistical models for limited dependent variables with application to the demand for durable goods". *Econometrica: Journal of the Econometric Society* . Vol.39 No.5 Page: 829-844.
- Cameron, A. Colin and Pravin K. Trivedi (1990). "Regression-based tests for overdispersion in the Poisson model". *Journal of econometrics*. Vol.46 No.3: 347-364.
- Cameron, A. Colin and Pravin K. Trivedi (1998). *Regression Analysis of Count Data*. United Kingdom: Cambridge University Press.
- Consul PoC and Felix Famoye (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* Vol.21 No.1 Page: 89-109.
- Czado Claudia, Erhardt Vinzenz, Min Aleksey and Wagner Stefan (2007). Dispersion and zero-inflation level applied to patent outsourcing rates Zero-

- inflated generalized Poisson models with regression effects on the mean. *Statistical Modelling*, Vol.7 No.2 Page: 125- 153.
- Deniz, Özlem. (2005). Poisson Regresyon Analizi. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Cilt: 4 Sayı:7 Sayfa: 59-72.
- Dinarcan, Gözdenur (2018). Sayma Verisi İçin Regresyon Modelleri ve Bir Uygulama. Yayınlanmamış Yüksek Lisans Tezi Hacettepe Üniversitesi İstatistik Anabilim Dalı Ankara.
- Famoye, Felix (1993). "Restricted generalized Poisson regression model". *Communications in Statistics-Theory and Methods* Vol.22 No.5 Page: 1335-1354.
- Gökdere Akkol, Suna (2004). *Çok Seviyeli Genelleştirilmiş Doğrusal Modellerde Parametre Tahmlemesinde MQL PQL ve MCMC Yöntemlerinin Karşılaştırılması*. Yayınlanmamış Doktora Tezi. Yüzüncü Yıl Üniversitesi. Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı. Van.
- Gujarati, Damodar N. and Porter, Dawn C(2012). *Temel ekonometri*. (Çevirenler: Ümit ŞENESEN, Gülay Günlük ŞENESEN). İstanbul: Literatür Yayıncılık.
- Güney, Esra ve Abar Hayri (2020). "Türkeyede Ekonomi İle İlgili Suçların Bölgesel Belirleyicileri". *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*. 65: 161-180.
- Greene, William H (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Department of Economics Stern School of Business New York University Page:1-36
- Hilbe, Joseph M. (2007). *Modeling count data*. New York. Cambridge University Press.
- Ismail Noriszura and Hossein Zamani (2013). "Estimation of claim count data using negative binomial, generalized Poisson, zero-inflated negative binomial and zero-inflated generalized Poisson regression models". *Casualty Actuarial Society E-Forum*. Vol.41. No.20. Page:1-29.

- İşbilen Yücel, Leyla (2021). *SPSS ve R Uygulamalı Sayma Veri Regresyon Analizi Modelleri*. İstanbul: Kriter Yayınevi.
- Karaca, Aslı Gizem ve Olmuş Hülya (2018). "Sıfır Değer Ağırlıklı Verilerin Analizinde Sıfır Değer Ağırlıklı Regresyon Modellerinin İncelenmesi". *Trakya Sosyal Bilimler Dergisi*. Cilt:20 Sayı: 2 Sayfa: 105-118.
- Kaya Yılmaz ve Yeşilova Abdullah (2012). "E-Posta Trafikinin Sıfır Değer Ağırlıklı Regresyon Yöntemleri". *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*. Cilt: 13 Sayı: 1 Sayfa: 51-63.
- Kılıç Duygu ve Bayrak Hülya (2020). "Sayma Verisi Modelleri Üzerine Bir Karşılaştırma: E- Ticarete Yaşanan". *Selçuk Üniversitesi Fen Fakültesi Fen Dergisi*. Cilt: 46, Sayı: 2, 85-102.
- Kıran, Gök Nur (2019). *Sayma Veri Modelleri ve Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı. İstanbul.
- Karadağ, Övgücan Gönenç (2014). *Genelleştirilmiş Doğrusal Modeller için Sınırlı Dalgalanmalı Kredibilite Yaklaşımı*. Yayınlanmamış Yüksek Lisans Tezi. Akütery Anabilim Dalı. Ankara.
- Kullback Solomon and Richard A. Leibler (1951). "On information and sufficiency". *The annals of mathematical statistics*. Vol.22 No.1 Page: 79-86.
- Long, J. Scott (1997). *Regression Models For Categorical Dependent And Limited Dependent Variables*. London: SAGE Publications
- Lambert, Diane (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* Vol.34 No.1 Page: 1-14.
- Mullahy, John (1986). "Specification and testing of some modified count data models". *Journal of econometrics*. Vol.33 No.3 Page: 341-365.
- Min, Yongyi ve Alan Agresti (2005). "Random effect models for repeated measures of zero-inflated count data". *Statistical modelling*. Vol.5 No.1 Page: 1-19.

- Memiş, Müslüme ve Önder Hasan (2018). "Poisson Regresyon Tahmin Yöntemlerinin Karşılaştırılması". *Black Sea Journal of Engineering and Science*. Cilt: 1, Sayı: 4. 140-146.
- McCullagh, Peter and John A. Nelder (1983). *Generalized linear models* London Chapman and Hall.
- Nelder, John Ashworth, and Robert WM Wedderburn (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* Vol.135 No.3: 370-384.
- Özmen, İlknur (1998). Poisson Regresyon Çözümleme Teknikleri. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara
- Özmen İlknur and Felix Famoye (2007). "Count regression models with an application to zoological data containing structural zeros". *Journal of Data Science*. Vol.5 No.4 Page: 491-502.
- Selim Sibel ve Üçdoruk Şenay (2003). "Sayma Veri modelleri ile Çocuk Sayısı Belirleyicileri: Türkiye'deki Seçilmiş İller İçin Sosyoekonomik Analizi" . *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* . Cilt: 18, Sayı: 2 13-31.
- Sezgin Funda H. ve Deniz Eylem (2004). "Poisson regresyon modelinde aşırı yayılım durumu ve negatif binomial regresyon analizinin Türkiye grev sayıları üzerine bir uygulaması" . *Yönetim Dergisi: İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadi Enstitüsü*. Cilt 15 Sayı 48 Sayfa: 17-25.
- Sarul, Latife Sinem (2012). *Aktüerya Alanında Genelleştirilmiş Lineer Modeller Kullanılarak Prim Fiyatlama Ve Filo Verisi Üzerinde Bir Uygulama*. Yayınlanmamış Doktora Tezi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı. İstanbul.
- Şengün Ucal, M (2006). "Ekonometrik Model Seçim Kriterleri Üzerine Kısa Bir İnceleme". *C.Ü. İktisadi ve İdari Bilimler Dergisi* . Cilt: 7 Sayı: 2 Sayfa: 87-92.
- Tüzel, Sema ve Sucu Meral (2012). "Hasar Sıklıkları İçin Sıfır Yığılmalı Kesikli Modeller". *İstatistikçiler Dergisi* . Cilt 5 Sayı 1 Sayfa: 23-31.

Vuong, Quang H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses". *Econometrica: Journal of the Econometric Society*. Vol.57 No.2: 307-333.

Yıldırım, Gizem (2019). *Poisson ve Negatif Binom Regresyon Modelleri*, Yayınlanmamış Yüksek Lisans Tezi Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Adana.





ÖZ GEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Diclehan ŞEKER
Uyruğu : T.C

EĞİTİM Derece

Kurum

Mezuniyet Yılı

İŞ TECRÜBESİ

Tarih

Kurum

Görev

YABANCI DİL BİLGİSİ

İngilizce