



GİRESUN
ÜNİVERSİTESİ



FEN BİLİMLERİ ENSTİTÜSÜ

CONWAY-MAXWELL-POISSON
REGRESYON MODELİ

MATEMATİK
ANABİLİM DALI

Yüksek Lisans Tezi

Bahar ÇELİK

162110002

Mayıs 2019

GİRESUN

**T.C.
GİRESUN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**CONWAY-MAXWELL-POİSSON
REGRESYON MODELİ**

YÜKSEK LİSANS TEZİ

Bahar ÇELİK

Enstitü Anabilim Dalı : Matematik
Tez Danışmanı : Dr. Öğretim Üyesi Esin AVCI

Mayıs 2019

T.C.
GİRESUN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

CONWAY-MAXWELL-POISSON
REGRESYON MODELİ

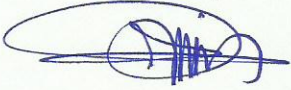
YÜKSEK LİSANS TEZİ

Bahar ÇELİK

Enstitü Anabilim Dalı : Matematik

Bu tez 24/05/2019 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

Doç. Dr.
İmdat İŞCAN



Jüri Başkanı

Dr. Öğretim Üyesi
Mehmet KORKMAZ



Üye

Dr. Öğretim Üyesi
Esin AVCI



Üye

Doç. Dr. Bahadır KOZ

Enstitü Müdürü

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Bahar ÇELİK
24/05/2019

TEŞEKKÜR

Bu çalışmanın gerçekleştirilmesinde değerli bilgilerini benimle paylaşan, kendisine ne zaman danışsam bana kıymetli zamanını ayırıp sabırla ve büyük bir ilgiyle bana faydalı olabilmek için elinden gelenin fazlasını sunan, her sorun yaşadığımdan yanına çekinmeden gidebileceğim, güler yüzünü ve samimiyetini benden esirgemeyen, gelecekteki mesleki hayatımda da bana verdiği değerli bilgilerden faydalanacağımı düşündüğüm kıymetli ve danışman hoca statüsünü hakkıyla yerine getiren Dr. Öğretim Üyesi Esin AVCI'ya teşekkürü bir borç biliyor ve şükranlarımı sunuyorum.

Lisans ve yüksek lisans eğitimim boyunca yardım, bilgi ve tecrübeleri ile bana sürekli destek olan Matematik ve İstatistik bölümündeki tüm hocalarıma teşekkür ederim.

Çalışmalarım boyunca maddi manevi destekleriyle beni hiçbir zaman yalnız bırakmayan aileme de sonsuz teşekkürler ederim.

İÇİNDEKİLER

TEŞEKKÜR.....	I
İÇİNDEKİLER	II
SİMGELER VE KISALTMALAR LİSTESİ	IV
ŞEKİLLER LİSTESİ	V
TABLOLAR LİSTESİ.....	VI
ÖZET.....	VII
SUMMARY	VIII
BÖLÜM 1. GİRİŞ.....	1
BÖLÜM 2. KAYNAK ARAŞTIRMASI.....	3
BÖLÜM 3. MATERYAL VE YÖNTEM.....	6
3.1. Genelleştirilmiş Lineer Modeller	6
3.1.1. Genelleştirilmiş Lineer Modellerin Yapısı	8
3.1.2. Genelleştirilmiş Lineer Modellerde Olabilirlik Fonksiyonu.....	11
3.1.2.1. Genelleştirilmiş Lineer Modellerde Parametre Tahmini	12
3.1.3. Uyum İyiliği.....	15
3.1.3.1. Sapma Ölçüsü	15
3.1.3.2. Pearson Ki-Kare İstatistiği	16
3.1.3.3. Akaike Bilgi Kriteri (AIC)	17
3.1.4. Genelleştirilmiş Lineer Modeller İçin Sonuç Çıkarımı.....	18
3.1.4.1. Wald İstatistiği	18
3.1.4.2. Olabilirlik Oran Testi	19
3.1.5. Genelleştirilmiş Lineer Modellerde Artık Analizi	19
3.1.5.1. Pearson Artıkları	19
3.1.5.2. Sapma Artıkları	20

3.1.6. Aşırı-Az Yayılım Durumu	20
3.2. Sayımla Elde Edilmiş Veriler için Regresyon Modelleri.....	22
3.2.1. Poisson Dağılımı ve Regresyon Modeli.....	22
3.2.1.1. Poisson Regresyon Modeli.....	23
3.2.1.2. Poisson Regresyon için En Çok Olabilirlik Tahmin Edicisi	23
3.2.2. Yarı (Quasi) Olabilirlik Yaklaşımı	25
3.2.3. Negatif Binom Regresyon Modeli	27
3.3. Conway-Maxwell-Poisson (COM-Poisson)	28
3.3.1. Conway-Maxwell-Poisson (COM-Poisson) Dağılımı	28
3.3.2. COM-Poisson Dağılımının Parametre Tahmini.....	30
3.3.2.1. COM-Poisson Dağılımının En Çok Olabilirlik Tahmin Edicisi	30
3.3.2.2. COM-Poisson Dağılımının Ağırlıklandırılmış En Küçük Kareler Tahmini .	32
3.3.2.3. COM-Poisson Dağılımının Bayesci Tahmini	33
3.3.2.4. COM-Poisson Dağılımının Yarı (Quasi) Tahmini.....	33
3.3.3. COM-Poisson Regresyon Modeli	36
3.3.4. Yayılım Parametresinin Testi.....	37
3.3.5. Tahmin Değerlerinin Elde Edilmesi	38
3.3.6. Katsayı Anlamlılığı	38
3.3.7. Katsayıların Yorumlanması	39
 BÖLÜM 4. ARAŞTIRMA BULGULARI.....	40
4.1. TRSM Verileri	41
4.1.1. TRSM Verileri İçin Modeller	45
4.2. Sosyal Medya Verileri.....	57
4.2.1. Sosyal Medya Verileri İçin Modeller.....	59
 BÖLÜM 5. TARTIŞMA VE SONUÇLAR.....	64
 KAYNAKLAR	68
ÖZGEÇMİŞ	73

SİMGELER VE KISALTMALAR LİSTESİ

AIC	: Akaike Bilgi Kriteri
COM-Poisson	: Conway-Maxwell-Poisson
PET	: Pozitron Emisyon Tomografisi
TRSM	: Toplum Ruh Sağlığı Merkezi



ŞEKİLLER LİSTESİ

Şekil 4.1. TRSM'ye geliş sayısı dağılımı	42
Şekil 4.2. TRSM'ye kayıtlı hastaların yaş dağılımı	42
Şekil 4.3. TRSM'de kayıtlı hastaların cinsiyete göre dağılımı	43
Şekil 4.4. TRSM'de kayıtlı hastaların eğitim düzeylerine göre dağılımı	43
Şekil 4.5. TRSM'de kayıtlı hastaların medeni durumuna göre dağılımı	44
Şekil 4.6. TRSM'de kayıtlı hastaların sosyal güvencesine göre dağılımı	44
Şekil 4.7. TRSM'de kayıtlı hastaların madde bağımlılık göre dağılımı	45
Şekil 4.8. Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi	47
Şekil 4.9. Negatif binom regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi.....	51
Şekil 4.10. COM-Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi.....	55
Şekil 4.11. Öğretmenlerin sosyal medya hesap sayılarına göre dağılımı	58
Şekil 4.12. Öğretmenlerin okul türüne göre dağılımı	58
Şekil 4.13. Öğretmenlerin cinsiyetlere göre dağılımı	59
Şekil 4.14. Öğretmenlerin yaş gruplarına göre dağılımı	59
Şekil 4.15. Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi	60
Şekil 4.16. COM-Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi.....	62

TABLolar LİSTESİ

Tablo 3.1. Üstel Ailenin Bazı Dağılımları İçi Kanonik Link (Bağıntı) Fonksiyonları ve Elde Edilen Modeller.....	10
Tablo 4.1. Bağımsız değişken kategorileri.....	41
Tablo 4.2. TRSM'ye geliş sayıları ve yaş için özet istatistikler.....	41
Tablo 4.3. Tüm olası alt kümeler yöntemine göre Poisson regresyon sonuçları	46
Tablo 4.4. Poisson regresyon modeli için katsayıların tahmini	48
Tablo 4.5. Tüm olası alt kümeler yöntemine göre negatif binom regresyon sonuçları	49
Tablo 4.6. Negatif binom regresyon modeli için katsayıların tahmini	52
Tablo 4.7. Tüm olası alt kümeler yöntemine göre COM-Poisson regresyon sonuçları	53
Tablo 4.8. COM Poisson regresyon modeli için katsayıların tahmini	56
Tablo 4.9. Bağımsız değişken kategorileri.....	57
Tablo 4.10. Öğretmenlerin sosyal medya hesap sayısı için özet istatistikler.....	57
Tablo 4.11. Tüm olası alt kümeler yöntemine göre Poisson regresyon sonuçları	60
Tablo 4.12. Poisson regresyon modeli için katsayıların tahmini	61
Tablo 4.13. Tüm olası alt kümeler yöntemine göre COM-Poisson regresyon sonuçları	62
Tablo 4.14. COM Poisson regresyon modeli için katsayıların tahmini	63

CONWAY-MAXWELL-POISSON REGRESYON MODELİ

ÖZET

Bir bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki neden-sonuç ilişkisinin ortaya çıkarılmasında regresyon modelinden yararlanılmaktadır.

Bağımlı değişkenin sayma/kesikli olma durumuyla birçok araştırmada karşılaşılmaktadır. Bu durumda klasik regresyon modeli yerine Poisson, negatif binom vb. gibi birçok dağılımı içeren genelleştirilmiş lineer modellerin kullanılması daha uygun olmaktadır. Her ne kadar Poisson regresyon modeli sayma verilerinin analizinde sıklıkla kullanılsa da ortalama ile varyans eşitliği varsayımı birçok deneysel veri için sağlanmamaktadır. Aşırı/az yayılım olarak tanımlanan bu durumlarda alternatif regresyon modelleriyle analiz edilmesi daha uygun olmaktadır. Genel olarak aşırı yayılım durumuyla daha çok karşılaşıldığından negatif binom regresyon modeli yaygın olarak kullanılmaktadır.

Bu tez çalışmasında, hem aşırı hem de az yayılım durumu için uygun analiz olanağı veren Conway-Maxwell-Poisson Regresyon modeli ayrıntılı olarak tanıtılmıştır. Conway-Maxwell-Poisson dağılım bilgisinin yanı sıra, parametre tahmini, anlamlılık ve yorumlanmasına yer verilmiştir. Poisson, negatif binom ve COM-Poisson regresyon modelleri önce aşırı yayılım gösteren 2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan hastaların, merkeze gelme sayılarına etki eden sosyo-demografik faktörlerin modellenmesinde, daha sonra az yayılım gösteren öğretmenlerin sosyal medya hesap sayılarına etki eden okul türü, cinsiyet ve yaş grubu faktörlerinin modellenmesinde uygulanmıştır. Her iki veri seti için değişken seçim yöntemi olarak tüm olası alt kümeler yöntemi kullanılmıştır. En küçük AIC değerli model en iyi model olarak seçilmiştir. Her iki veri setinde de COM-Poisson regresyon en iyi modeli oluşturmuştur.

Analizler R programının “MASS” ve “COMpoissonReg” paketlerinden ve “glm” fonksiyonundan yararlanılarak yapılmıştır.

Anahtar kelimeler: Sayma veri, Poisson regresyon, Conway-Maxwell-Poisson, Aşırı-az yayılım, Negatif binom

CONWAY-MAXWELL-POISSON REGRESSION MODEL

SUMMARY

The regression model is used to determine the relationship between dependent variable and one or more independent variables.

Counting/discrete dependent variable is encountered in many studies. In this case, instead of the classical regression model, It is more appropriate to use generalized linear models that including many distributions Poisson, negative binomial et al. Although the Poisson regression model is frequently used in the analysis of counting data, the assumption of the equality of mean and variance is not provided for many experimental data. It is more appropriate to analyze with alternative regression models in case of over-under dispersion. Generally, the negative binomial regression model is widely used for overdispersed data.

In this thesis, the Conway-Maxwell-Poisson Regression model which is suitable for analysis of both over-under dispersed data is presented in detail. Beside Conway-Maxwell-Poisson distribution information, parameter estimation, significance and interpretation are given. Poisson, negative binomial and COM-Poisson regression models were used to determine the effect of socio-demographic factors on the number of visits Giresun Community Mental Health Center (TRSM) between 2011 and 2014. Then, three regression models applied to model school type, gender, and age group factors with the number of social media accounts of the teachers. All possible subset regression approach was used as a variable selection method for both data sets. The smallest AIC-valued model was selected as the best model. COM-Poisson regression was the best model in both datasets.

The analyses were performed by using the "MASS "and "COMPoissonReg" packages and the "glm" function on the R program.

Keywords: Count data, Poisson regression, Conway-Maxwell-Poisson, Over-under dispersion, Negative binomial

BÖLÜM 1. GİRİŞ

Bir bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki neden-sonuç ilişkisinin ortaya çıkarılmasında regresyon modelinden yararlanılmaktadır.

Normal dağılıma sahip olan bağımlı değişkenin bağımsız değişken(ler) ile arasındaki ilişkinin ortaya çıkarılmasında klasik regresyon analizi kullanılmaktadır, ancak normal dağılıma sahip olmayan bağımlı değişkenin regresyon modelinin oluşturulmasında birçok dağılım seçeneğini kapsayan genelleştirilmiş lineer modeller kullanılabilmektedir. Genelleştirilmiş lineer modellerde hatalar normal dağılımının yanı sıra üstel, gamma, Poisson, binom vb. dağılımlara sahip olmaktadır.

Sağlık çalışmalarında, biyoistatistik, aktüeryal bilimler ile demografide sıklıkla negatif olmayan tam sayılardan oluşan sayımla elde edilen verilerle karşılaşmaktadır. Bu tür verilerin dağılımının Poisson dağılımına sahip olduğu varsayılarak ona etki eden değişken(ler) arasındaki ilişkinin analizinde Poisson regresyon modeli sıklıkla kullanılmaktadır. Ancak uygulamada sayımla elde edilen verilerde Poisson dağılımının varsayımı olan ortalamanın varyansa eşitliği çoğu zaman ihlal edilmektedir. Bu varsayımın ihlal edilmesi regresyon modelini oluşturan katsayıların standart hatalarının yanılgılı olmasına dolayısıyla da modelde yer almaması gereken değişken(ler) modelde yer almasına ve sonuç olarak da hatalı modelin oluşturulmasına neden olmaktadır. Varyansın ortalamadan büyük olması aşırı, küçük olması durumuna az yayılım adı verilmektedir.

Varyansın ortalamadan büyük olması durumunda (aşırı yayılım) negatif binom regresyon modelleri yaygın olarak kullanılmaktadır. Bir diğer yaklaşım ise ortalama ile varyans arasındaki fonksiyonel ilişkiden yararlanan yarı (Quasi) olabilirlik yöntemidir.

Bu tez çalışmasının ikinci bölümünde lineer regresyon hakkında genel bir bilgi verildikten sonra genelleştirilmiş lineer modeller için tahmin, güven aralıkları, model seçimi, hipotez testleri, artık analizleri ve uyum iyiliği testleri açıklanmıştır.

Üçüncü bölümde sayımla elde edilen veriler için literatürde yaygın olarak kullanılan Poisson regresyon, yarı (Quasi) olabilirlik yaklaşımı ve aşırı yayılım gösteren veriler için negatif binom regresyon tanıtılmıştır.

Dördüncü bölümde ise verilerin aşırı-az yayılım göstermesi durumunda etkili olan COM-Poisson regresyon modeline ait olabilirlik ve yarı (Quasi) fonksiyonu ile parametre tahmini, katsayı anlamlılığı ve yorumlanması verilmiştir.

COM-Poisson regresyon modelinin aşırı ve az yayılım gösteren veriler için uygunluğunun ortaya konulması amaçlandığından uygulama kısmında aşırı ve az yayılım gösteren iki veri seti analiz edilmiştir. Öncelikle aşırı yayılım gösteren 2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan hastaların, merkeze gelme sayılarına etki eden sosyo-demografik faktörlerin saptanması amaçlanmıştır. İkinci kısmında ise, az yayılım gösteren 2015-2016 yılları arasında Giresun Üniversitesi Bilimsel Araştırma Proje desteği ile (SOS-BAP-A-200515-43) gerçekleştirilen proje verilerinin bir kısmı olan öğretmenlerin sosyal medya hesap sayılarına etki eden okul türü, cinsiyet ve yaş grubu faktörleri incelenmiştir.

Analizler R programının “MASS” ve “COMPOissonReg” paketlerinden ve “glm” fonksiyonundan yararlanılarak yapılmıştır.

BÖLÜM 2. KAYNAK ARAŞTIRMASI

1962 yılında Conway ve Maxwell tarafından tanımlanan ve Shmueli ve ark. (2005) tarafından geliştirilen Conway-Maxwell-Poisson (COM-Poisson) dağılımı, Poisson dağılımının aşırı veya az yayılım gösteren verileri için genelleştirilmiş bir dağılımdır [1, 2].

Michener ve Tighe (1992), ABD’de otobanda meydana gelen ölümcül kazalara etki eden faktörleri Poisson regresyon ile incelemişlerdir [3].

Englin ve Shonkwiler (1995), potansiyel talep ile tüketici rantı ve popülasyon ortalamaları arasındaki ilişkiyi Poisson ve negatif binom regresyon modelleri ile analiz etmişlerdir [4].

Van Der Burg ve ark. (1998), cinsiyete göre akademisyenlerin akademik iş gücü piyasasındaki yükselmelerini Poisson regresyon kullanarak incelemişlerdir [5].

McKean ve Taylor (2000), Moskova’da rekreasyon gezi sayısı ile gezi maliyeti ve geçirilen zamanı ilişkilendirmek için Poisson ve negatif binom regresyon modellerini kullanmışlardır [6].

Moons ve ark. (2001), seyahat zamanı ile maliyet farkı arasındaki ilişkiyi belirlemek için yarı logaritmik, negatif binom ve kesik negatif binom modellerini kullanmışlardır. Negatif binom regresyonunda maliyet istatistiksel olarak anlamlı bulunmuştur [7].

Sezgin ve Deniz (2004), aşırı yayılım gösteren Türkiye’deki grev sayılarına etki eden faktörler ile ilişkilendirmek için negatif binom regresyon modelini kullanmışlardır [8].

Yeşilyurt (2005), en uygun model olarak Poisson regresyonu kullanarak yaş bağımsız değişkeninin Türkiye’deki boşanma verilerini açıkladığını göstermiştir [9].

Bilgiç ve Florkowski (2007), ABD’de levrek avı gezisine katılma sayısını etkileyen faktörleri incelemek için negatif binom regresyon modelini ele almışlardır [10].

Literatürde sayma verileri için yapılan uygulamalardan bazıları aşağıdaki gibi verilmiştir. Mouatassim ve Ezzahid (2012), Poisson bol sıfırlı regresyon modeline özel sağlık sigortası verilerini yaş, medeni hal ve cinsiyet bağımsız değişkenleri ile ilişkilendirmişlerdir [11].

Arı ve Önder (2012), lineer, lojistik, negatif binom, Poisson, Temel bileşen, Probit, Ridge ve Cox regresyon yöntemlerinin kullanılabileceği durumlarla ilgili öneride bulunmuşlardır [12].

Sellers ve ark. (2012), COM-Poisson regresyon modelini işletme, ulaşım ve biyoloji gibi farklı alanlarda uygulamışlardır [13].

Zhu (2012), COM-Poisson dağılımını zaman serisinde kullanarak yeni bir INGARCH modeli önermiştir [14].

Avşar (2014), Türkiye’deki antidamping soruşturma sayısına yerel sanayiye ait üretim ve ithalat değişkenlerinin etkisini negatif binom regresyon analizi ile incelemiştir [15].

Santarelli ve ark. (2016), COM-Poisson regresyon modelinin pozitron emisyon tomografisinde (PET) görüntü rekonstrüksiyon algoritmalarında ve niceliksel PET veri analizinde kullanımının uygun olduğunu belirtmişlerdir [16].

Avcı (2018), Giresun ilinde Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı şizofren hastaların hastaneye yatış sayısını etkileyen demografik faktörleri Poisson, negatif binom ve COM-Poisson ile incelemiştir [17].

Avcı (2018), alg sayısına etki eden mevsim ve istasyonları Poisson ve COM-Poisson ile karşılaştırmıştır [18].



BÖLÜM 3. MATERYAL VE YÖNTEM

3.1. Genelleştirilmiş Lineer Modeller

Regresyon analizi, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki nedensel ilişkiyi tanımlama, tahmin etme veya kestirmek için kullanılan istatistiksel modeller sınıfı şeklinde tanımlanabilir. Bağımlı değişken Y ve bağımsız değişkenler kümesi $X_1, X_2, \dots, X_k$ (k bağımsız değişken sayısıdır) olarak tanımlanır. Y ve $X_1, X_2, \dots, X_k$ arasındaki ilişki,

$$Y = f(X_1, X_2, \dots, X_k) + \varepsilon \quad (3.1)$$

olarak ifade edilir. Buradaki ε terimi hata vektörüdür ve uyum eksikliğini ifade etmektedir. Y ve $X_1, X_2, \dots, X_k$ arasındaki ilişkiyi tanımlamak için $f(X_1, X_2, \dots, X_k)$ regresyon fonksiyonu kullanılabilir ve (3.1) eşitliği,

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (3.2)$$

şeklinde bir lineer regresyon modeli yazılabilir. Burada $\beta_1, \beta_2, \dots, \beta_k$ regresyon parametreleri veya katsayıları olarak adlandırılır ve veriden tahmin edilen bilinmeyen sabitlerdir.

Uygulamalarda bağımlı değişkenin sürekli olmasına rağmen normal dağılıma sahip olmadığı, ayrıca kesikli veya iki uçlu değer (başarı ve başarısızlık) alması durumlarıyla sıklıkla karşılaşilmektedir. Bu tür verilerin dağılımlarının genel olarak üstel aileye (binom, Poisson, gamma, normal, negatif binom) sahip olduğu görülmüştür. Üstel dağılım ailesine sahip bağımlı değişkenin bağımsız değişkenler

ile olan ilişkisinin ortaya konulmasında genelleştirilmiş lineer modellerin kullanılması uygun olmaktadır. Bu modeller ayrıca bağımlı değişkenin normal dağılım göstermediği durumlar için veri dönüşümüne alternatif olarak kullanılabilir. Buradan elde edilen sonuçların veri dönüşümüne göre daha dar güven aralığına sahip olduğu ve ters dönüşüme gerek olmaması nedeniyle daha doğru sonuçlar verdiği görülmüştür [19, 20]. Genelleştirilmiş lineer modellerde varyans, ortalamanın bir fonksiyonudur [21].

Genelleştirilmiş lineer modeller ilk defa Nelder ve Wedderburn (1972) tarafından ele alınmıştır [21]. Daha sonraki yıllarda McCullagh ve Nelder (1989) [20], Aitken M. (1989) [22], Lindsey (1997) [23], Uusipaikka (2000) [24], McCulloch ve Searle (2001) [25], Myers ve ark. (2001) [26], Dobson (2002) [27], Duntelman ve Ho (2006) [28], genelleştirilmiş lineer modeller teorisi hakkında çalışmaya devam etmişlerdir.

Üstel ailenin genel formu,

$$f(y; \theta, \Phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\Phi)} + c(y, \Phi) \right\} \quad (3.3)$$

şeklinde. Burada $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ belirli fonksiyonlardır. θ parametresi konum parametresidir, Φ parametresi yayılım parametresidir. $a(\Phi) = \Phi w$ şeklindedir. Gruplandırılmamış veriler için $w=1$ iken gruplandırılmış veriler için $w=n$ olarak alınır.

Üstel ailenin üyesi olan dağılımlardan normal, binom, Poisson ve gamma dağılımları için olasılık yoğunluk fonksiyonları aşağıdaki gibi verilebilir:

μ ve σ^2 parametrelili normal dağılıma sahip Y rastgele değişkeni için olasılık yoğunluk fonksiyonu,

$$f(y; \mu, \sigma) = \exp \left\{ -\frac{[y - \mu]^2}{2\sigma^2} \right\} \cdot \frac{1}{\sqrt{2\pi}\sigma}$$

$$= \left\{ (y\mu - \mu^2/2)/\sigma^2 - \frac{1}{2}[y^2/\sigma^2 + \ln(2\pi\sigma^2)] \right\} \quad (3.4)$$

dir. Fonksiyon (3.3)'de $\theta = \mu$, $b(\theta) = \mu^2/2$, $a(\Phi) = \Phi$, $(\varphi = \sigma^2)$ ve $c(y, \Phi) = -\frac{1}{2}[y^2/\sigma^2 + \ln(2\pi\sigma^2)]$ olarak ifade edilir. Kısaca konum parametresi μ ve doğal ölçek parametresi σ^2 dir.

Binom dağılımının olasılık yoğunluk fonksiyonu şu şekildedir:

$$f(y; p) = \exp \left\{ y \ln \left[\frac{p}{1-p} \right] + n \ln(1-p) + \ln \binom{n}{y} \right\} \quad (3.5)$$

Buradan, $\theta = \ln \left[\frac{p}{1-p} \right]$, $b(\theta) = n \ln(1 + \exp[\theta])$, $a(\Phi) = 1$, $\Phi = 1$ ve $c(y) = \ln \binom{n}{y}$ olur.

Poisson dağılımında ise olasılık yoğunluk fonksiyonu şöyledir;

$$f(y; \mu) = \exp[y \ln \mu - \mu - \ln(y!)] \quad (3.6)$$

Buradan, $\theta = \ln(\mu)$, $\mu = \exp(\theta)$, $a(\Phi) = 1$, $b(\theta) = \mu$ ve $c(y, \Phi) = -\ln(y!)$ olur.

Gamma dağılımının olasılık yoğunluk fonksiyonu şu şekildedir;

$$f(y; \mu, \Phi) = \exp \left\{ \frac{y/\mu - (-\ln \mu)}{-\Phi} + \frac{\Phi+1}{\Phi} \ln y - \frac{\ln \Phi}{\Phi} - \ln \Gamma \left(\frac{1}{\Phi} \right) \right\} \quad (3.7)$$

Buradan, $\theta = 1/\mu$, $a(\Phi) = -\Phi$, $b(\theta) = \ln(\theta)$ ve $c(y, \Phi) = \frac{\Phi+1}{\Phi} \ln y - \frac{\ln \Phi}{\Phi} - \ln \Gamma \left(\frac{1}{\Phi} \right)$ olur.

3.1.1. Genelleştirilmiş Lineer Modellerin Yapısı

Genel olarak genelleştirilmiş lineer modeller üç bileşenden oluşmaktadır. Bunlar;

- a. Bağımlı değişkenin dağılımı. Bağımlı değişken gözlemleri $(y_1, y_2, \dots, y_n)$ sırasıyla $\mu_1, \mu_2, \dots, \mu_n$ ortalamalı birbirinden bağımsız gözlemlerdir ve y_i gözlemleri üstel aile üyesi olan bir dağılıma sahiptir.
- b. Lineer tahminlerin bulunduğu sistematik kısım. Modelin sistematik kısmı $x_1, x_2, \dots, x_k$ bağımsız değişkenlerini içermektedir.
- c. Link (bağıntı) fonksiyonu. Model bir link (bağıntı) fonksiyon aracılığıyla lineer tahminlerin bulunduğu sistematik kısım ile bağımlı değişken ortalamasını ilişkilendirir.

$$\eta = X'\beta = \beta_0 + \sum_{i=1}^k \beta_i x_i = g(\mu_i), \quad i=1,2,\dots,n.$$

Yukarıda tanımlanan link(bağıntı) terimi ortalama ve lineer tahmin edici arasındaki ilişki fonksiyonundan türetilmiştir. Bağımlı değişkenin beklenen değeri,

$$E(Y_i) = g^{-1}(\eta_i) = g^{-1}(X'\beta) \quad (3.8)$$

şeklindedir. Çoklu lineer regresyonda,

$$\mu_i = \eta_i = X'\beta, \quad i = 1, 2, \dots, n \quad (3.9)$$

yani, $g(\mu_i) = \mu_i$ 'dir. $g(\mu_i)$ 'nin μ_i 'ye eşit olması bağıntı fonksiyonuna birim (özdeş) link fonksiyonu adını alır. Link (bağıntı) fonksiyonunu diferansiyellenebilen monoton bir fonksiyon olmalıdır [26].

Link fonksiyonu için birçok alternatif bulunmaktadır. İki tip link fonksiyonu bulunmaktadır. Bunlar, kanonik ve kanonik olmayan link fonksiyonlarıdır. η_i 'ye $\eta_i = \theta_i$ ise kanonik ve $\eta_i \neq \theta_i$ ise kanonik olmayan link denir.

Bazı durumlar için kanonik link fonksiyonu Tablo 3.1.'de verilmiştir [29].

Tablo 3.1. Üstel Ailenin Bazı Dağılımları İçin Kanonik Link (Bağıntı) Fonksiyonları ve Elde Edilen Modeller

Dağılım	Konum Parametresi	Link (bağıntı) Fonksiyonu	Model
Normal	$\eta_i = \theta_i$	$\mu_i = x'_i \beta_i$ (birim link(bağıntı))	$\mu = x' \beta$
Binom	$\theta_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right)$	$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = x'_i \beta_i$ (lojistik link(bağıntı))	$p = \frac{\exp\{x' \beta\}}{1 + \exp\{x' \beta\}}$
Poisson	$\theta_i = \ln(\mu_i)$	$\ln(\mu_i) = x'_i \beta_i$ (log link (bağıntı))	$\mu = \exp(x' \beta)$
Negatif Binom	$\frac{-\log(1 - \exp \theta_i)}{k}$	$\log \left(\frac{k \mu_i}{1 + k \mu_i} \right) = x'_i \beta_i$	$\mu = \exp(x' \beta)$
Gamma ve Üstel	$\theta_i = 1/\mu_i$	$(\mu_i)^{-1} = x'_i \beta_i$ (ters link (bağıntı))	$\mu = 1/x' \beta$

Genelleştirilmiş lineer modeller için diğer link (bağıntı) fonksiyonları;

- 1) Probit link (bağıntı) $\eta_i = \Phi^{-1}[E(y_i)]$ dir. Burada Φ , kümülatif standart normal dağılımdır.
- 2) Tamamlayıcı log-log link (bağıntı), $\eta_i = \ln\{\ln[1 - \mu_i]\}$,
- 3) Güç aile link (bağıntı),

$$\eta_i = \begin{cases} \mu_i^\lambda, & \lambda \neq 0 \\ \ln[\mu_i], & \lambda = 0 \end{cases} \quad (3.10)$$

şeklindedir.

Probit ve tamamlayıcı log-log link fonksiyonları sadece Bernoulli ve binom dağılımları ile birlikte kullanılmaktadır. Güç aile linki bağımlı değişkenin normal dağılıma uygunluk varsayımlarını yerine getirmemesi durumunda yapılan Box-Cox dönüşümüyle aynıdır. Genelleştirilmiş lineer modellerde link fonksiyonu bağımlı değişkenin dağılımının taşıdığı özellikleri üstlenmektedir. Link fonksiyonunun yanlış seçimi ile yapılan analizler yanlış sonuçlara götürmektedir.

3.1.2. Genelleştirilmiş Lineer Modellerde Olabilirlik Fonksiyonu

Genelleştirilmiş lineer modellerde parametre tahminleri en çok olabilirlik fonksiyonuna dayanmaktadır. Parametre tahminlerinin en çok olabilirlik tahmin edicileri iterasyonlu yeniden ağırlıklandırılmış en küçük kareler, Newton-Raphson yöntemi ve Fisher-Scoring algoritmalarıyla bulunmaktadır.

Üstel ailenin genel formu, $f(y, \theta, \Phi)$ olmak üzere, Y rastgele değişkenin gözlenen bir y_i değeri için olabilirlik fonksiyonu;

$$L = (\theta, \Phi, Y) = \prod_{i=1}^n f(y_i, \theta, \Phi) \quad (3.11)$$

şeklinde tanımlanır. L olabilirlik fonksiyonu θ ve Φ parametrelerinin bir fonksiyonudur.

Matematiksel istatistikten üstel aile dağılımının şu özelliklere sahip olduğu görülmektedir [30].

$$1) \ E \left[\frac{\partial \ln L(\theta)}{\partial \theta} \right] = 0 \quad (3.12)$$

$$2) \ E \left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \right] + E \left\{ \left[\frac{\partial \ln L(\theta)}{\partial \theta} \right]^2 \right\} = 0 \quad (3.13)$$

(3.12) ve (3.13) eşitlikleri kullanılarak ortalama ve varyans sırasıyla tek bir gözlem için aşağıdaki şekilde elde edilir.

$$f(Y; \theta, \Phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\Phi)} + c(y, \Phi) \right\} \quad (3.14)$$

$$\frac{\partial L(\theta)}{\partial \theta} = \frac{y - b'(\theta)}{a(\Phi)} \quad (3.15)$$

$$\frac{\partial^2 L(\theta)}{\partial \theta^2} = \frac{-b''(\theta)}{a(\Phi)} \quad (3.16)$$

Üstel aile üyelerinin ortalama ve varyansı,

$$E(Y) = b'(\theta) \quad (3.17)$$

$$var(Y) = a(\Phi)b''(\theta) \quad (3.18)$$

şeklindedir.

3.1.2.1. Genelleştirilmiş Lineer Modellerde Parametre Tahmini

Genelleştirilmiş lineer modellerin parametre tahminleri, en çok olabilirlik fonksiyonundan oluşturulan skor denklem sisteminin çözümlenmesiyle elde edilmektedir. Denklem sistemlerinin çözümlenmesinde iterasyonlu yeniden ağırlıklandırılmış en küçük kareler, Newton-Raphson ve Fisher-Scoring algoritmalarından yararlanılmaktadır. Newton-Raphson algoritması Hessian matrisinin kendisinden, Fisher-Scoring algoritması Hessian matrisinin beklenen değerinden yararlanarak sayısal çözümleme sağlar. Kanonik link kullanılması durumunda Newton-Raphson ve Fisher-Scoring algoritmaları aynı sonucu vermektedir.

Kanonik ve kanonik olmayan link (bağıntı) için en çok olabilirlik yöntemi sırasıyla aşağıdaki gibi tanımlanır.

Kanonik link kullanılması durumunda, $\eta_i = g(\mu_i) = x_i' \beta = \sum_{j=1}^p x_{ij} \beta_j$ olur. En çok olabilirlik fonksiyonunun logaritmasından yararlanarak β parametresinin en çok olabilirlik tahminleri,

$$\frac{\partial L_i(\beta)}{\partial \beta} = 0 \quad (3.19)$$

eşitliğiyle elde edilir. Burada,

$$\begin{aligned}\frac{\partial L(\beta)}{\partial \beta} &= \frac{\partial L}{\partial(\theta_i)} \cdot \frac{\partial(\theta_i)}{\partial(\beta)} \\ &= \sum_{i=1}^n \frac{1}{a(\Phi)} \left[y_i - \frac{db(\theta_i)}{d(\theta_i)} \right] x_i\end{aligned}\quad (3.20)$$

$$L = \log f(y, \theta, \Phi) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\Phi)} + c(y_i; \Phi) \right\} = \sum_{i=1}^n L_i \quad (3.21)$$

ile tanımlanır. Olabilirlik fonksiyonu β parametresinin bir fonksiyonu olarak yazılabilmektedir.

$$\begin{aligned}L(\beta) &= \sum_{i=1}^n L_i(\beta) \\ &= \sum_{i=1}^n \frac{1}{a(\Phi)} [y_i - \mu_i] x_i = 0\end{aligned}\quad (3.22)$$

Birçok durumda $a(\Phi)$ sabit olarak ele alındığından (3.22) denklemi,

$$\sum_{i=1}^n [y_i - \mu_i] x_i = 0 \quad (3.23)$$

dönüşür. Bu aslında $p=k+1$ denklem sisteminin her bir parametre için çözümüdür. Matris formatı ile gösterimi aşağıdaki gibidir.

$$X'(y - \mu) = 0 \quad (3.24)$$

Burada $\mu' = [\mu_1, \mu_2, \dots, \mu_p]$ dir. Bu denklem sistemine skor denklem sistemi adı verilir ve yeniden ağırlıklandırılmış en küçük kareler yöntemi ile çözümlenir. Bu iterasyonun sonucunda elde edilen b değeri, β parametresinin en çok olabilirlik tahmin değeridir. Bu tahmin değerinin beklenen değer ve varyansı aşağıdaki gibi tanımlanır.

$$E(b) = \beta \quad (3.25)$$

Varyans, bilgi matrisinden yararlanılarak elde edilir.

$$var(b) = I^{-1}(b) = [X'VX]^{-1}[a(\Phi)]^2 \quad (3.26)$$

Burada $V = diag\{\sigma_i^2\}$, σ_i^2 bağımlı değişkenin dağılımına bağlı olarak μ_i 'nin bir fonksiyonudur [26].

Kanonik link (bağıntı) fonksiyonunun kullanılmaması durumunda $\eta_i \neq \theta_i$, log olabilirlik fonksiyonu,

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^n \frac{\partial L}{\partial(\theta_i)} \cdot \frac{\partial(\theta_i)}{\partial(\eta_i)} \cdot \frac{\partial(\eta_i)}{\partial \beta} \quad (3.27)$$

olarak elde edilir. Burada,

$$\frac{dL}{d\theta_i} = \sum_{i=1}^n \frac{1}{a(\Phi)} \left[y_i - \frac{db(\theta_i)}{d(\theta_i)} \right] = \frac{1}{a(\Phi)} \sum_{i=1}^n (y_i - \mu_i) \quad (3.28)$$

ve

$$\frac{\partial(\eta_i)}{\partial \beta} = x_i \quad (3.29)$$

şeklindedir. Dolayısıyla,

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \frac{y_i - \mu_i}{a(\Phi)} \frac{d\theta_i}{d\eta_i} x_i \quad (3.30)$$

olarak elde edilir. Yine $a(\Phi)$ sabit olarak ele alınırsa skor denklemleri,

$$X' \Delta(y - \mu) = 0 \quad (3.31)$$

olarak bulunur. Burada $\Delta = diag\{\frac{d\theta_i}{d\eta_i}\}$ dir. β parametrelerinin en çok olabilirlik tahmini olan b 'lerin yaklaşık varyans-kovaryans matrisi bilgi matrisinden yararlanılarak aşağıdaki gibi elde edilir [26].

$$var(b) = I^{-1}(b) = [X'\Delta V\Delta X]^{-1}[a(\Phi)]^2 \quad (3.32)$$

β parametresinin en çok olabilirlik tahmin edicisinin yaklaşık (asimptotik) normal dağılımlı olduğu ve ortalamasının β , varyans-kovaryans matrisinin $I^{-1}(b)$ 'e eşit olduğu varsayılmaktadır. Yani $b \approx AN[\beta, I^{-1}(b)]$ şeklindedir [31].

3.1.3. Uyum İyiliği

Oluşturulan modelin doğruluğu, model içindeki değişimleri sabitleyen test istatistikleri aracılığıyla belirlenmektedir. Özellikle modelin uyum iyiliği için en yararlı iki test istatistiği; Sapma ölçüsü ve Pearson ki-kare istatistiğidir. Sapma ölçüsü, gözlemlerin en büyük değeri ile beklenen olabilirlik fonksiyonları arasındaki farkı ölçerken, Pearson ki-kare istatistiği gözlenen ve modellenen değerler arasındaki nispi farkı ölçmektedir [32]. Her iki istatistik, ilgili serbestlik dereceleriyle ki-kare dağılımına sahip olduğu varsayılmaktadır. Büyük bir sapma ya da Pearson ki-kare test istatistik değeri kötü bir uyumun belirtisidir.

3.1.3.1. Sapma Ölçüsü

Genelleştirilmiş lineer modelde gözlem sayısının parametre sayısı ile aynı sayı içermesine doymuş model adı verilir. Sapma istatistiği, doymuş olmayan (indirgenmiş) her bir alt model için doymuş modelin uyum iyiliğindeki kaybı için ölçülür ve aşağıdaki gibi tanımlanır.

$$D(\hat{\beta}, Y) = -2\ln \left[\frac{L(\Phi, \hat{\beta})}{L_{max}(\Phi, Y)} \right]$$

$$= 2[\log L_{max}(\Phi, Y) - \log L(\Phi, \hat{\beta})] \quad (3.33)$$

Burada, $L_{max}(\Phi, Y)$ doymuş modelin maksimum olabilirlik fonksiyonunun, $L(\Phi, \hat{\beta})$ ise verilen q parametrelili ($q < n$) doymuş olmayan (indirgenmiş) bir model için β parametrelerine göre maksimum olabilirlik fonksiyonunu göstermektedir ve $D(y, \hat{\beta}) \sim \chi^2_{n-q}$ dur. θ kanonik parametresinin doymuş ve indirgenmiş model için tahminleri sırasıyla $\tilde{\theta}$ ve $\hat{\theta}$ ile gösterilmesi durumunda üstel dağılım olabilirlik fonksiyonu için sapma istatistiği aşağıdaki gibi yazılabilir.

$$\begin{aligned} l(\theta, \Phi, y) &= \sum_{i=1}^n \left[\frac{\theta_i Y_i - b(\theta_i)}{a(\Phi)} + c(y; \Phi) \right] \\ &= 2 \sum_{i=1}^n \left[\frac{(\tilde{\theta}_i - \hat{\theta}_i) Y_i - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a(\Phi)} \right] = \frac{D(\hat{\beta}, Y)}{a(\Phi)} \end{aligned} \quad (3.34)$$

$$D^*(\hat{\beta}, Y) = \frac{D(\hat{\beta}, Y)}{\Phi} \quad (3.35)$$

$\frac{D^*(\hat{\beta}, Y)}{n-q}$ ifadesine ortalama sapma değeri denir. Ortalama sapma değerinin 1'den büyük olması, varsayılan modelin uyumunun iyi olmadığını gösterir. Model karşılaştırmalarında en küçük sapma değerine sahip en basit modelin bulunması amaçlanır [31].

3.1.3.2. Pearson Ki-Kare İstatistiği

Bir diğer uyum iyiliği ölçüsü Pearson ki-kare istatistiğidir. Genelleştirilmiş lineer modellerde varyans ortalamanın bir fonksiyonudur ($var(Y_i) = a(\Phi)b''(\theta) = a(\Phi)V(\mu_i)$) bu nedenle istatistik aşağıdaki gibi tanımlanmaktadır.

$$X^2_S = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{var(Y_i)} = \frac{1}{a(\Phi)} \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (3.36)$$

Y rastgele değişkeninin gerçekleşen bir y değeri için ve $a(\Phi)=1$ olması durumunda;

$$X_s^2 = \sum_{i=1}^n \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (3.37)$$

elde edilir. Bu ifadeye ölçeklenmiş Pearson ki-kare istatistiği denir.

Büyük n değeri için ($n \rightarrow \infty$), X_s^2 yaklaşık olarak X_{n-q}^2 olarak dağılımlıdır. X^2 değerinin büyük olması kötü uyumun belirtisidir.

Yayılım parametresi tahmini için ortalama sapma veya ölçeklenmiş Pearson ki-kare istatistiği kullanılmaktadır. Yayılım parametresinin özellikle binom ve Poisson dağılımlarında 1 olduğu kabul edilmektedir [33].

3.1.3.3. Akaike Bilgi Kriteri (AIC)

1973 yılında Akaike tarafından ortaya atılan bilgi kriteri yaklaşımına göre bir modelin uyumunu ölçmek için aşağıdaki formül kullanılmaktadır. Sapma ölçümüne dayanarak bilgi kriteri aşağıdaki gibi tanımlanır.

$$D_c = D - \alpha q \Phi \quad (3.38)$$

Burada D sapma, q modeldeki parametre sayısı ve Φ dağılım parametresidir. Eğer Φ sabitse, sırasıyla $\alpha \approx 4$ ve $\alpha \approx 2$ olarak ele alınması durumunda, % 5 seviyesinde bir parametrenin test edilmesine eş değer ve minimum tahmin hatalarına sebep olduğu belirtilmiştir. Uygun modelin seçiminde Akaike bilgi kriteri sıklıkla kullanılır. Tercih edilecek model, küçük D_c değerine göre belirlenmektedir [34].

Log-olabilirlik ve parametre sayısına dayanan bilgi kriteri ise,

$$AIC = -2\log L - 2q \quad (3.39)$$

şeklindedir. Burada L tahmin edilen model için olabilirlik fonksiyonunun maksimum değeri, q ise modeldeki parametre sayısıdır. Bu yaklaşımda fazla parametre sayısına sahip model cezalandırılmaktadır.

Uyum ölçüsünün seçiminde özellikle iç içe (nested) modellerin karşılaştırılmasında sapma ölçüsü uygun iken Collet (1991) ikili veriler için uygun olmadığını belirtmiştir [35]. Büyük örneklerde sapma ölçüsü ve Pearson ki-kare kullanılması uygundur. Akaike bilgi kriteri (AIC) ise modellerin birbiri ile karşılaştırılmasında kullanılabilir.

3.1.4. Genelleştirilmiş Lineer Modeller İçin Sonuç Çıkarımı

İstatistiksel sonuç çıkarımının iki temel yöntemi, güven aralıkları ve hipotez testleridir. Genelleştirilmiş lineer modellerde en sık kullanılan hipotez testleri, Wald ki-kare istatistiği, Olabilirlik oran testidir.

3.1.4.1. Wald İstatistiği

Lineer regresyona benzer olarak genelleştirilmiş lineer modellerde regresyon katsayılarının anlamlılığı ($H_0: \beta_i = 0$) hipotezinin test edilmesinde,

$$z_i^2 = \left(\frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \right)^2 \quad (3.40)$$

formülü ile verilen Wald istatistiği kullanılmaktadır. Wald istatistiği, asimtotik normal dağılım yaklaşımını kullanmaktadır. Büyük örnek için (3.40)'deki istatistik, ki-kare dağılımına sahiptir.

Parametrelerin güven aralıklarının elde edilmesinde de Wald istatistiği kullanılmaktadır [20, 36]. β_j için $100 \times (1-\alpha)$ güven aralığı, $\hat{\beta}_j \pm Z_{\alpha/2} S_{\hat{\beta}_j}$ ile bulunur.

3.1.4.2. Olabilirlik Oran Testi

Genel olarak yokluk hipotezinin (H_0) alternatif hipoteze (H_1) karşı testi için kullanılan bir yöntemdir. Parametre kümelerinin $M_2 \subseteq M_1$ olması durumunda olabilirlik oran test istatistiği,

$$\Lambda = \frac{L(M_2)}{L(M_1)} \quad (3.41)$$

olarak elde edilir. $n \rightarrow \infty$ ise,

$$G^2 = -2 \ln \Lambda \quad (3.42)$$

olur. Λ uyum iyiliği olarak adlandırılır. Burada $L(M_1)$ 'e tam model olabilirlik ve $L(M_2)$ incelenen modelin olabilirlik fonksiyonudur. Λ asimtotik olarak X^2_{n-q} dağılımına sahiptir. Burada $n - q$ iki modeldeki parametrelerin sayısı arasındaki farktır [37].

3.1.5. Genelleştirilmiş Lineer Modellerde Artık Analizi

Artıklar yardımıyla seçilen link (bağıntı) fonksiyonunun uygunluğu ve varsayımların doğruluğu gibi modelin yeterliliğini etkileyen etmenler incelenmektedir.

Standart artıklar ($\hat{e}_i = y_i - \hat{y}_i$), varyans sabit olmadığından dolayı genelleştirilmiş lineer modellerde teknik olarak uygun değildir. Genelleştirilmiş lineer modeller için yaygın olarak iki tür artık kullanılmaktadır.

3.1.5.1. Pearson Artıkları

Pearson artıklar, y_i gözlemi için $\hat{e}_i = y_i - \hat{y}_i$ şeklinde tanımlanan standart artığın standart sapması ile standartlaştırılmasıyla elde edilen artıklardır.

$$e_{i,Pearson} = \frac{y_i - \hat{y}_i}{\sqrt{var(\hat{y}_i)}} \quad (3.43)$$

Pearson artıkları, Pearson ki-kare ($\chi^2 = \sum_{i=1}^n e_i^2$) ile ilişkilidir. Büyük örneklemede genellikle sabit varyans ile normal dağıldığı kabul edilmektedir [26].

3.1.5.2. Sapma Artıkları

Toplamları modelin uyum iyiliği ölçüsü $D = \sum_{i=1}^n d_i$ olarak kullanılan bu artıklar aşağıdaki gibi tanımlanır.

$$e_{i,sapma} = sign(y_i - \hat{y}_i) \sqrt{d_i} \quad (3.44)$$

Pierce ve Schafer (1986), McCulloch ve Nelder (1989), üstel aile modeli için artıkları incelemiş ve sapma artıklarının lineer modellerde kullanılan artıklara oldukça benzerlik gösterdiği, bu nedenle tanışal grafiklerde kullanılmasını önermişlerdir [38, 20].

3.1.6. Aşırı-Az Yayılım Durumu

Temel olarak aşırı yayılım (over-dispersion) bağımlı değişkenin varyansının, bağımlı değişken dağılımının beklediğinden daha büyük olması durumudur. Benzer olarak, az yayılım (under-dispersion) bağımlı değişkenin dağılımının beklediğinden daha küçük olmasıdır. Genel olarak aşırı yayılım durumuyla karşılaşıldığından bu problemin çözümüne yönelik çalışmalar yapılmıştır.

Aşırı-az yayılım durumu, uyum eksikliğine neden olan kaynakların (bağımlı değişken dağılım varsayımının hatalı olması, bağımsız değişken(ler)in modelde hatalı bir fonksiyonla yer alması (örneğin, daha yüksek mertebeden yer alması gereken bir bağımsız değişkenin doğrusal olarak modelde yer verilmesi) ve veride aykırı değer olması) saptanmasında kullanılan Hosner-Leweshow veya en kolay

yöntem olan ortalama sapma (sapma/serbestlik derecesi) değerinin 1'e yakın olma kriterinin kontrol edilmesiyle ortaya konulabilmektedir.

Seçilen dağılımın ve modelin uygun olması ve veride aykırı değer olmamasına rağmen ortalama sapma değerinin 1'i aşması durumuna aşırı yayılım, 1'den küçük olması durumuna da az yayılım adı verilmektedir.

Aşırı-az yayılım durumunda varyans fonksiyonunu bir Φ çarpımsal yayılım etkenine sahip olduğu varsayılmaktadır. Yaygın olarak kullanılan binom ve Poisson için varyans fonksiyonları aşağıdaki gibi tanımlanmaktadır.

$$\text{Binom Dağılımı; } \quad \text{Var}(y) = \Phi\mu(1 - \mu) \quad (3.45)$$

$$\text{Poisson Dağılımı; } \quad \text{Var}(y) = \Phi\mu \quad (3.46)$$

$\Phi > 1$ ise aşırı, $\Phi < 1$ ise az ve $\Phi = 1$ olması durumunda eş yayılım adı verilmektedir. Aşırı-az yayılım durumunda model parametrelerinin tahmin değerleri Φ değerinden etkilenmez. Ancak model parametrelerinin standart sapması Φ değerinden etkilenir. Φ parametresi değeri biliniyorsa ya da bazı veri noktalarında tekrarlanıyorsa doğrudan belirlenebilir aksi takdirde ortalama sapma değeri,

$$\sqrt{\text{sapma}/n - q} \quad (3.47)$$

ya da ölçeklendirilmiş Pearson ki-kare değeri,

$$\sqrt{x^2/(n - q)} \quad (3.48)$$

Φ 'ın nokta tahmin edicisi olarak kullanılır.

Parametrelerin anlamlılıklarının hipotez testinde, (3.47) veya (3.48) denklemlerinden elde edilen değer parametre tahminlerinin kovaryans matrisi ile çarpılarak ölçeklenmiş standart sapma değeri kullanılır [26].

3.2. Sayımla Elde Edilmiş Veriler için Regresyon Modelleri

Birçok araştırmada incelenen bağımlı değişkenin, negatif olmayan tamsayılardan oluşan kesikli bir değer alması durumuyla karşılaşılmaktadır. Bu tür sayımla elde edilmiş bağımlı değişkenlere etkisi olduğu düşünülen bağımsız değişkenler ile ilişkilendirilmesinde normal dağılımlı lineer regresyon modelinin kullanılması doğru olmamaktadır. Bu tür verilerin analizinde sayıma dayalı bağımlı değişkenlerin modellenmesinden yararlanılmaktadır. Yaygın olarak; Poisson, negatif binom ve yarı (Quasi) Poisson regresyon modelleri kullanılmaktadır.

3.2.1. Poisson Dağılımı ve Regresyon Modeli

Bağımlı değişkenin Poisson dağılımına sahip olması durumunda Poisson regresyon modelinin kullanılması uygun olmaktadır. Poisson dağılımı, çoğunlukla nadir olayların oluş sayısının modellenmesinde kullanılan, belirli bir zamanda, alanda veya hacimde gerçekleşen olay sayısını tanımlayan bir dağılımdır. İlk defa 1838’de yayınlanan eserinde Siméon-Denis Poisson tarafından ortaya atılmıştır [39].

Poisson dağılımının olasılık fonksiyonu;

$$P(y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0, 1, \dots \quad (3.49)$$

olarak tanımlanır. Poisson dağılımının ortalaması ile varyansı birbirine eşittir ($E(Y_i) = V(Y_i) = \lambda_i$).

3.2.1.1. Poisson Regresyon Modeli

Poisson regresyon modeli birçok alanda kullanılmıştır. Poisson regresyon modeli genelleştirilmiş lineer modellerin bir üyesidir.

Poisson regresyon modeli kesikli ve negatif olmayan verileri içermesinden dolayı bağımlı değişkenin logaritması, bağımsız değişkenlerin doğrusal fonksiyonu ile bağlantılıdır. Genel anlamda model,

$$\log(E(Y_i)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (3.50)$$

ile tanımlanmaktadır. Bu modele istatistikte “log-linear” model adı da verilmektedir. Burada $\log(E(Y_i))$ bağımlı değişkenin tahmin edilen değeri, β_0 doğrunun y eksenini kestiği nokta; $\beta_0, \beta_1, \dots, \beta_k$ bağımsız değişkenlerin katsayıları, $X_1, X_2, \dots, X_k$ bağımsız değişkenleri ifade eder.

Poisson regresyon modelinin kullanılabilmesi için bağımlı değişkenin ortalaması ve varyansının birbirine eşit, çok az sayıda sıfır değerine sahip olması, ayrıca gözlemlerin bağımsız ve gözlem sırasının eşit olduğu varsayımlarının sağlanması gerekmektedir. Varyansın ortalamaдан büyük (aşırı yayılım) ve bağımlı değişkenin çok fazla sıfırlı olması durumunda Poisson regresyon analizi yerine negatif binom regresyon analizinin kullanılması gerekmektedir [40].

3.2.1.2. Poisson Regresyon için En Çok Olabilirlik Tahmin Edicisi

$y_1, y_2, \dots, y_n$ rasgele bağımlı değişkenlerinin $E(y_i) = \lambda_i$ ve $V(y_i) = \lambda_i$ varyanslı bağımsız Poisson dağılıma sahip olması durumunda $x_1, x_2, \dots, x_k$ bağımsız değişkenlerin bağımlı değişken ortalaması ile olan koşullu ilişkileri,

$$E(Y_i|x_i) = \lambda_i = e^{X_i' \beta} \quad i = 1, 2, \dots, n \quad (3.51)$$

şeklindedir. Burada $x'_i = [1, x_{i1}, x_{i2}, \dots, x_{in}]$ olarak tanımlanır. Bağımsızlık varsayımından $y_1, y_2, \dots, y_n$ 'nin log olabilirlik fonksiyonu,

$$\ln L(\beta_i y) = \log \prod_{i=1}^n \left[\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \right] \quad (3.52)$$

şeklindedir.

$$\ln L(y; \beta) = \sum_{i=1}^n \left[-e^{x'_i \beta} + y_i x'_i \beta - \ln y_i! \right] \quad (3.53)$$

β 'nın en çok olabilirlik tahmin edicisinin belirlenmesi için β 'ya göre birinci dereceden kısmi türevi alınıp sıfıra eşitlendiğinde,

$$\frac{\partial \ln L(\beta_i y)}{\partial \beta} = \sum_{i=1}^n (y_i x_i - e^{x'_i \beta} x_i) \quad (3.54)$$

olur. Bu durumda skor denklemleri;

$$\sum_{i=1}^n (y_i x_i - e^{x'_i \beta} x_i) = 0 \quad (3.55)$$

$$\sum_{i=1}^n (y_i - \lambda_i) x_i = 0 \quad (3.56)$$

$$x'(y - \lambda) = 0 \quad (3.57)$$

ile elde edilir. β 'nın tahmin edicileri iterasyonlu yöntemlerle (Newton-Raphson, Fisher-Scoring algoritması vb.) elde edilir.

İterasyonlu yöntemlerle elde edilen β_i parametre tahminleri, x_i 'deki bir birimlik değişime bağlı olan $E(Y_i | x_i)$ 'deki nispi değişimi ölçmektedir [41].

3.2.2. Yarı (Quasi) Olabilirlik Yaklaşımı

Yarı (Quasi) olabilirlik, dağılımın ortalama ve varyansının birbiri ile ilişkili olması durumunda birinci ve ikinci momentleri kullanan bir yaklaşımdır. Yarı (Quasi) olabilirlik yaklaşımında,

$$Var(Y_i) = \Phi f(\lambda_i) \quad (3.58)$$

olarak tanımlanır. Burada Φ yayılım parametresi ve $f(.)$ uygun şekilde seçilmiş bir fonksiyondur. Bağımsız değişkenlerin etkisini yarı (Quasi) yaklaşımında tahmin etmek için yarı (Quasi) olabilirlik fonksiyonu,

$$Q(y_i, \lambda_i) = \int_{y_i}^{\lambda_i} \frac{y_i - \lambda'_i}{v(\lambda'_i)} d\lambda'_i + f(y_i) \quad (3.59)$$

veya eşit olarak,

$$\frac{\partial Q(y_i, \lambda_i)}{\partial \lambda_i} = \frac{y_i - \lambda_i}{v(\lambda_i)} \quad (3.60)$$

tanımlanır.

Eğer y_i 'ler tek parametrelili üstel dağılım ailesinden ise Q , log olabilirlik fonksiyonuna eşittir [42]. Log olabilirlik fonksiyonu ve log-Quasi olabilirlik fonksiyonu aynı özelliklere sahiptir ve β parametreleri yaklaşık normal dağılıma sahiptir [43]. Tek parametrelili üstel dağılım ailesi için log olabilirlik ve log-Quasi olabilirlik aynı olduğundan, tek parametrelili Poisson dağılımı için de geçerlidir.

Z vektörünün elemanları $\frac{\partial Q}{\partial \beta_i}$ 'lerden oluşuyorsa yarı (Quasi) olabilirlik fonksiyona göre Z , sıfır ortalamalı ve varyansı aşağıdaki matrisin elemanlarıdır.

$$-E \left(\frac{\partial^2 Q}{\partial \beta_i \partial \beta_j} \right) \quad (3.61)$$

Ortalama ile varyans arasındaki ilişki tam olarak bilinmediğinde ancak varyans bilindiğinde $\hat{\beta}$ hesaplanabilmektedir.

Varyansın ortalamadan büyük veya küçük olması parametre tahmini üzerinde çok az bir etkisi olmasına rağmen standart hata anlamlılık testleri ve güven aralıklarının oluşturulmasında oldukça etkilidir [44]. Bu durumda varyans,

$$Var(Y_i) = \Phi V(\lambda_i) = \Phi \lambda_i \quad (3.62)$$

olarak ele alınır. Burada $V(\lambda_i)$ bilinen fonksiyon ve Φ bilinmeyen parametredir. $\Phi > 1$ ise aşırı, $\Phi < 1$ ise az ve $\Phi = 1$ olması durumunda eş yayılıma sahiptir. β parametresinin en çok yarı (Quasi) olabilirlik tahmin değerleri Φ değerlerinden etkilenmediğinden $\Phi = 1$ olarak ele alınır. Ancak hata tahminlerinin elde edilmesinde Φ 'nın bilinmesi gerekmektedir. λ yaklaşık olarak β için lineer olduğundan,

$$E \left(\sum_{i=1}^n \left(\frac{(y_i - \lambda_i)^2}{Var \lambda_i} \right) \right) \cong n - p \quad (3.63)$$

dolayısıyla,

$$\hat{\Phi} = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \lambda_i)^2}{Var \lambda_i} \quad (3.64)$$

elde edilir. Normal lineer modeller için bu değer bilinen varyans tahmin değeridir [42].

Yarı (Quasi) olabilirlik, regresyon katsayılarının değişen varyans durumunda bile (varyans fonksiyonunun hatalı tanımlanmasında bile) tutarlı tahmin edilmesini sağlar [44].

3.2.3. Negatif Binom Regresyon Modeli

Sayımla elde edilen bağımlı değişkenin çok fazla sıfırlı veya varyansının ortalamasından büyük olması durumlarında nedensellik ilişkisi kurulan bağımsız değişkenlere ait katsayı tahminlerinin tutarlı ancak yeterli olmamasına yol açtığından Poisson regresyon modeli ile analiz edilmesi uygun değildir. Bu durumda negatif binom regresyon modeli alternatif olarak kullanılabilir. Poisson regresyon modeline benzer olarak (log) link (bağıntı) fonksiyonunu kullanarak bağımlı değişkenle bağımsız değişkenleri ilişkilendirmektedir. Lawless (1987) yaptığı çalışmada aynı veri setine Poisson ve negatif binom regresyonunu uygulamış, analiz sonucunda negatif binom regresyonunun daha küçük sapma değerine sahip olduğunu göstermiştir [45].

Negatif binom regresyon modelinde bağımlı değişkenin negatif binom dağılımına sahip olduğu varsayılmaktadır. Negatif binom dağılımı, Poisson dağılım parametresi olan ortalamanın Gamma dağılımına sahip olması varsayımına dayanmaktadır. Yani,

$$y \sim \text{Poisson}(\lambda) \quad (3.65)$$

ve λ , Gamma dağılımına sahip bir değişken ise,

$$y|\lambda \sim \text{Poisson}(\lambda) \quad (3.66)$$

$$\lambda \sim \text{Gamma}(\alpha, \beta) \quad (3.67)$$

olur. Burada λ , $\alpha\beta$ ortalamalı ve $\alpha\beta^2$ varyanslı Gamma dağılımına sahiptir.

Ortalaması λ_i ve yayılım parametresi α olan negatif binom dağılımına ($\text{NB}(\lambda, \alpha)$) sahip $y_1, y_2, \dots, y_n$ rasgele değişkenlerinin olasılık fonksiyonu aşağıdaki şekilde tanımlanır.

$$P(Y = y|\lambda, \alpha) = \frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})} \left[\frac{\alpha^{-1}}{\alpha^{-1}+\lambda} \right]^{\alpha^{-1}} \left[\frac{\lambda}{\alpha^{-1}+\lambda} \right]^y \quad (3.68)$$

olarak tanımlanır. Sırasıyla ortalama ve varyans,

$$E(Y_i) = \lambda_i \text{ ve } Var(Y_i) = \lambda_i(1 + \alpha\lambda_i^k) \quad (3.69)$$

dir. $\alpha = 0$ olması durumunda negatif binom dağılımı Poisson dağılımına dönüşmektedir [46]. $E(Y_i)$, birim zamanda bağımlı değişkeninin ortalama ortaya çıkma sayısıdır. Negatif binom regresyon modeli aşağıdaki gibi tanımlanır [47].

$$\log(E(Y_i)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \quad (3.70)$$

Burada $\log(E(Y_i))$ bağımlı değişkenin tahmin edilen değeri, β_0 doğrunun y eksenini kestiği nokta; $\beta_0, \beta_1, \dots, \beta_k$ bağımsız değişkenlerin katsayıları ve $X_1, X_2, \dots, X_k$ bağımsız değişkenleri ifade eder.

3.3. Conway-Maxwell-Poisson (COM-Poisson)

3.3.1. Conway-Maxwell-Poisson (COM-Poisson) Dağılımı

Conway ve Maxwell tarafından 1962 yılında tanımlanan bir dağılımdır [2]. Conway-Maxwell-Poisson (COM-Poisson) dağılımı, Poisson dağılımının aşırı-az yayılım gösteren verileri için genelleştirilmiş bir dağılımdır. Dağılımın olasılık fonksiyonu, Poisson dağılımından yararlanılarak elde edilir. Poisson (λ) dağılımına sahip Y rasgele değişkenine ait olasılık fonksiyonu (3.49) eşitliği ile tanımlıdır. Poisson dağılımının özelliklerinden biri ardışık olasılıkların y'ye göre lineer olmasıdır.

$$\frac{P(Y=y-1)}{P(Y=y)} = \frac{y}{\lambda} \quad (3.71)$$

Ancak bazı uygulamalarda bu oran y 'ye göre lineer olarak azalmaz. Bu durumda dağılım Poisson dağılımından daha kalın veya ince kuyruklu bir dağılım yapısına dönüşür. (3.71) eşitliği aşağıdaki gibi tanımlanır.

$$\frac{P(Y=y-1)}{P(Y=y)} = \frac{y^v}{\lambda} \quad (3.72)$$

(3.72) ile tanımlanmış olan eşitliğin dağılımına Conway-Maxwell-Poisson (COM-Poisson) olarak adı verilir ve aşağıdaki gibi olasılık fonksiyonu tanımlanır [48].

$$P(Y = y|\lambda, v) = \frac{\lambda^y}{(y!)^v} \frac{1}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}} \quad y = 0, 1, 2, \dots \quad (3.73)$$

Burada $\lambda > 0$ ve $v \geq 0$ 'dır. $\frac{\lambda^j}{(j!)^v}$ serisi herhangi bir $\lambda > 0$ ve $v > 0$ için yakınsaklık gösterir. $j \rightarrow \infty$ iken serideki iki ardışık terim için oran $\frac{\lambda}{j^v}$ sifıra yaklaşır. (3.73) eşitliğindeki payda,

$$Z(\lambda, v) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v} \quad (3.74)$$

ile tanımlanır. Burada $Z(\lambda, v)$ 'ye normalleştirme katsayısı ve v 'ye ise yayılım parametresi adı verilir. Veri $v < 1$ olduğunda aşırı, $v > 1$ olduğunda az yayılım göstermektedir. v 'nün değerlerine bağlı olarak COM-Poisson bilinen bazı kesikli dağılımların genelleştirilmesidir. Bu dağılımlar;

1. $v = 1$ olduğunda $Z(\lambda, v) = e^\lambda$ olur. (3.73) eşitliği Poisson (λ)
2. $v \rightarrow \infty$ olduğunda ise $Z(\lambda, v) \rightarrow 1 + \lambda$ olur. (3.73) eşitliği başarı olasılığı $\frac{\lambda}{1+\lambda}$ Bernoulli
3. $v = 0$ ve $\lambda < 1$ iken $Z(\lambda, v) = \sum_{j=0}^{\infty} \lambda^j = \frac{1}{1-\lambda}$ olur. (3.73) eşitliği geometrik

$$P(Y = y|\lambda, v) = \lambda^y(1 - \lambda) \quad y = 0, 1, 2, \dots \quad (3.75)$$

şeklinde olmaktadır. $v = 0$ ve $\lambda \geq 1$ iken $Z(\lambda, v)$ yakınsaklık göstermediğinden herhangi bir dağılım tanımlanamamaktadır. COM-Poisson; ($v = 0$) iken geometrik (ilk başarı sağlanıncaya kadar başarısız deneme sayısı), ($v = 1$) Poisson ve ($v \rightarrow \infty$) Bernoulli dağılımları arasında sürekli bir köprü görevi görmektedir. Benzer biçimde n tane bağımsız COM-Poisson rasgele değişkeni de bu üç bilinen kesikli dağılımlar arasında sürekli bir köprü olmaktadır.

1. $v = 1$ için toplam n λ parametrelili Poisson
2. $v \rightarrow \infty$ için n ve $\lambda/1 + \lambda$ parametrelili binom
3. $v = 0$ ve $\lambda < 1$ için geometrik dağılımlı n tane değişkenin toplamı, n ve $1 - \lambda$ parametrelili negatif binom dağılımına dönüşür [13].

3.3.2. COM-Poisson Dağılımının Parametre Tahmini

Shmueli vd. (2005) iki parametrelili COM-Poisson dağılımının tahmin edilmesinde üç farklı tahmin yöntemi önermiştir. Birincisi en çok olabilirlik tahmin yöntemi, ikinci ağırlıklandırılmış en küçük kareler yöntemi ve üçüncüsü ise Bayesci tahmin yöntemidir [1].

Khan ve Khan (2010)'da parametre tahmini için Wedderburn (1974)'nün yarı (Quasi) tahmin yöntemini geliştirerek λ ve v parametrelerinin yarı (Quasi) tahminlerini elde etmişlerdir [49, 42].

3.3.2.1. COM-Poisson Dağılımının En Çok Olabilirlik Tahmin Edicisi

$y_1, y_2, \dots, y_n$ bağımsız COM-Poisson dağılımı için log-olabilirlik fonksiyonu,

$$L(\lambda, v) = n\bar{Y} \log(\lambda) - n \overline{\log(Y!)} v - n \log Z(\lambda, v) \quad (3.76)$$

ile tanımlanır.

$$E(Y) = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (3.77)$$

$$E(\log(Y!)) = \overline{\log(Y!)} = \frac{1}{n} \sum_{i=1}^n \log(Y_i!) \quad (3.78)$$

Bu olabilirlik $\bar{Y}$ ve $\overline{\log(Y!)}$ yeterli istatistiklerinin bir fonksiyonudur ve en çok olabilirlik tahminleri bu istatistiklerin beklenen değerlerinin iterasyonlu olarak veya doğrudan olabilirlik fonksiyonunun optimizasyon programı ile çözülmesiyle elde edilir. Katsayıların en çok olabilirlik tahmin edicileri doğrudan $v \geq 0$ koşulu altında lineer olmayan optimizasyondan yararlanarak (3.76) denkleminin maksimize edilmesi ile elde edilir. Momentler normalleştirme katsayısı Z 'nin kısmi türevinin alınmasıyla elde edildiğinden yaklaşık bir sonuç vermektedir.

$$E[Y] = \lambda \frac{d \log Z(\lambda, v)}{d \lambda} \approx \lambda^{\frac{1}{v}} - \frac{v-1}{2v} \quad (3.79)$$

$$E[\log(Y!)] = -\frac{d \log Z(\lambda, v)}{d v} \approx \frac{1}{2v^2} \log \lambda + \lambda^{\frac{1}{v}} \left(\frac{\log \lambda}{v} - 1 \right) \quad (3.80)$$

$v \leq 1$ ve $\lambda > 10^v$ için bu yaklaşımlar iyi sonuç vermektedir.

Olabilirlik $\theta = (\log(\lambda), v)$ parametresine göre konkav olduğundan Newton iterasyon yöntemine göre hızlı ve güvenilir bir şekilde maksimum değeri bulunmaktadır. Log-olabilirlik fonksiyonunun gradienti,

$$\nabla L(\theta) = n \left[\frac{\bar{Y} - E[Y]}{-(\log(Y!) - E[\log(Y!)])} \right] \quad (3.81)$$

ve ikinci türev matrisi,

$$\nabla^2 L(\theta) = n \begin{pmatrix} -var(Y) & cov(Y, \log(Y!)) \\ cov(Y, \log(Y!)) & -var(\log(Y!)) \end{pmatrix} \quad (3.82)$$

ile tanımlanır.

$$var(Y)var(\log(Y!)) > cov(Y, \log(Y!))^2 \quad (3.83)$$

olduğundan (3.82) eşitliği ile tanımlanan matris pozitif determinanta sahiptir. Ancak her iki özdeğerler negatif olduğundan negatif iz değerine sahiptir. Bu da olabilirliğin konkav olduğunu gösterir.

3.3.2.2. COM-Poisson Dağılımının Ağırlıklandırılmış En Küçük Kareler Tahmini

İki parametrelili COM-Poisson dağılımının tahmin edilmesinde ikinci tahmin yöntemi ağırlıklandırılmış en küçük kareler yöntemidir. Bu yöntem, ardışık olasılıkların arasındaki ilişkinin lineer olmasına dayandırılmıştır. Aşağıdaki gibi tanımlanmaktadır.

$$\log \left[\frac{P(Y=y-1)}{P(Y=y)} \right] = -\log \lambda + v \log(y) \quad (3.84)$$

Burada $P(Y = y)$, y 'nin veriden tahmin edilen olasılığını göstermektedir. Ağırlıklandırılmış en küçük kareler (3.84) eşitliğinde tanımlanan oranın sıfır olmayan kovaryans ve sabit olmayan varyans için düzeltilmeye ihtiyacı vardır.

Yukarıda tanımlanan oranın bağımlı değişken değerleri olan y_i 'ye karşı çizilmesiyle elde edilen doğrunun eğiminin 1'e eşit olması Poisson dağılımına sahip olduğunu göstermektedir.

3.3.2.3. COM-Poisson Dağılımının Bayesci Tahmini

Kadane vd. (2005) üstel dağılım ailesinden yararlanarak COM-Poisson dağılımının eşlenik önselini aşağıdaki gibi tanımlamıştır.

$$h(\lambda, v) = \lambda^{a-1} e^{-vb} Z^{-c} \kappa(abc) \quad (3.85)$$

Burada $\lambda > 0$, $v \geq 0$ ve $\kappa(a, b, c)$ normalleştirme katsayılarıdır. Eşlenik önsel dağılım olması nedeniyle sadece önsel dağılıma ait hiperparametrelerin (a,b,c) tanımlanmasıyla sonsal dağılımın elde edilmesi kolaydır. Sonsal dağılım $a' = a + S_1$, $b' = b + S_2$, $c' = c + n$ parametreleri ile (3.85) dağılımına sahip olmaktadır [50].

3.3.2.4. COM-Poisson Dağılımının Yarı (Quasi) Tahmini

Khan ve Khan (2010), Wedderburn (1974)'ün bileşik yarı (Quasi) olabilirlik tahminini geliştirerek marjinal yarı (Quasi) olabilirlik tahminini elde etmişlerdir [49,42]. Bileşik yarı (Quasi) olabilirlik denklemi aşağıdaki gibi verilmiştir.

$$\sum_{i=1}^f D_i^T V_i^{-1} (f_i - \mu_i) = 0 \quad (3.86)$$

Burada,

$$f_i = (y_i, y_i^2)^T$$

$$\mu_i = E(f_i)$$

$$V_i = cov(f_i)$$

$$D_i = \frac{\partial E(f_i)}{\partial (\lambda, v)}$$

matris bileşenleri aşağıdaki gibidir,

$$D_i = \begin{pmatrix} \partial \theta_i / \partial \lambda & \partial \theta_i / \partial v \\ \partial m_i / \partial \lambda & \partial m_i / \partial v \end{pmatrix}$$

Burada,

$$\partial \theta_i / \partial \lambda = \frac{\lambda_i^{\frac{1}{v}}}{v} \quad (3.87)$$

$$\partial \theta_i / \partial v = \frac{1}{2} \frac{v-1}{2v^2} - \frac{1}{2v} - \frac{\lambda_i^{\frac{1}{v}}}{v} \quad (3.88)$$

$$\partial m_i / \partial \lambda = \left(\frac{2\lambda_i^{\frac{1}{v}} + 2v\lambda_i^{\frac{1}{v}} + 2v\lambda_i^{\frac{1}{v}}}{v^2} \right) \quad (3.89)$$

$$\partial m_i / \partial v = \frac{1}{2v^3} \left[\begin{array}{c} 2\lambda_i^{\frac{1}{v}} v \ln(\lambda_i) + v - 1 \\ -4\lambda_i^{\frac{1}{v}} \ln(\lambda_i) v - 4\lambda_i^{\frac{1}{v}} v - 4\lambda_i^{\frac{1}{v}} \ln(\lambda_i) \end{array} \right] \quad (3.90)$$

f_i 'in kovaryans matrisi aşağıdaki gibi tanımlanır.

$$V_i = \begin{pmatrix} \text{var}(Y_i) & \text{cov}(Y_i, Y_i^2) \\ & \text{var}(Y_i^2) \end{pmatrix}$$

V_i 'deki her eleman, y_{it} 'nin moment üreten fonksiyonunun iterasyon sonucundan elde edilmiştir.

$$E[Y_i^{r+1}] = \lambda \frac{d}{d\lambda} E[Y^r] + E[Y]E[Y^r] \quad (3.91)$$

Y_i^2, Y_i^3 ve Y_i^4 momentleri yardımıyla varyans ve kovaryans aşağıdaki gibi elde edilir.

$$\begin{aligned} \text{cov}(Y_i, Y_i^2) &= E(Y_i^3) - E(Y_i)E(Y_i^2) \\ &= \frac{2\lambda_i^{\frac{1}{v}} + 2v\lambda_i^{\frac{1}{v}} + 2v\lambda_i^{\frac{1}{v}}}{v^2} \end{aligned} \quad (3.92)$$

$$\text{Var}(Y_i^2) = E(Y_i^4) - E(Y_i^2)^2$$

$$= \frac{\lambda_i^{\frac{1}{v}} v^2 + 4\lambda_i^{\frac{3}{v}} v^2 + 10\lambda_i^{\frac{2}{v}} v - 4\lambda_i^{\frac{1}{v}} v + 4\lambda_i^{\frac{1}{v}} - 4\lambda_i^{\frac{2}{v}} v^2}{v^3} \quad (3.93)$$

λ ve v yarı (Quasi) olabilirlik tahmin değerleri (3.88) denkleminin Newton-Raphson iterasyon yöntemini r . iterasyonunun yakınsaklığıyla aşağıdaki gibi elde edilir.

$$\begin{pmatrix} \hat{\lambda}_{r+1} \\ \hat{v}_{r+1} \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_r \\ \hat{v}_r \end{pmatrix} + [\sum_{i=1}^I D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^I D_i^T V_i^{-1} (f_i - \mu_i)]_r \quad (3.94)$$

Burada $\hat{\lambda}_r$, r . iterasyondaki $\hat{\lambda}$ değerini ve $[\cdot]_r$, r . iterasyondaki değeri ifade eder. Bu tahminler tutarlıdır ve $I \rightarrow \infty$ iken $I^{-\frac{1}{2}} \left((\hat{\lambda}, \hat{v}) - (\lambda, v) \right)^T$ yaklaşık olarak sıfır ortalamalı ve kovaryans matrisi $I [\sum_{i=1}^I D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^I D_i^T V_i^{-1} (f_i - \mu_i) (f_i - \mu_i)^T V_i^{-1} D_i] [\sum_{i=1}^I D_i^T V_i^{-1} D_i]^{-1}$ olan normal dağılıma sahiptir.

COM-Poisson regresyon modeli için marjinal yarı (Quasi) olabilirlik fonksiyonu kullanılarak önce λ daha sonra yayılım parametresi olan v için tahmin elde edilir. λ için yarı (Quasi) olabilirlik tahmini,

$$\sum_{i=1}^I D_{i,\lambda}^T V_{i,\lambda}^{-1} (y_i - \theta_i) = 0 \quad (3.95)$$

elde edilir. Burada,

$$V_{i,\lambda} = \frac{\lambda_i^{\frac{1}{v}}}{v} \quad (3.96)$$

ve

$$D_{i,\beta} = \frac{\partial \theta_i}{\partial \lambda} = \frac{\lambda_i^{\frac{1}{v}}}{v} \quad (3.97)$$

v için yarı (Quasi) olabilirlik tahmini,

$$\sum_{i=1}^I D_{i,\alpha}^T V_{i,\alpha}^{-1} (y_i^2 - \eta_i) = 0 \quad (3.98)$$

elde edilir. Burada,

$$\eta_i = E(Y_i^2) = \frac{\lambda_i^{\frac{1}{v}}}{v} + \left[\lambda_i^{\frac{1}{v}} - \frac{v-1}{2v} \right]^2 \quad (3.99)$$

$$= \sum_{i=1}^I D_{i,\alpha}^T V_{i,\alpha}^{-1} (y_i^2 - \eta_i) = 0$$

ve

$$D_{i,v} = \frac{1}{2v^3} \left[2\lambda_i^{\frac{1}{v}} \ln(\lambda_i) + v - 1 - 4\lambda_i^{\frac{2}{v}} \ln(\lambda_i)v \right] + \frac{1}{2v^3} \left[-4\lambda_i^{\frac{1}{v}}v - 4\lambda_i^{\frac{1}{v}} \ln(\lambda_i) \right] \quad (3.100)$$

$V_{i,v}$, Y_i^2 'in varyansıdır ve şu şekilde elde edilir.

$$V_{i,v} = E(Y_i^4) - E(Y_i^2)^2 \quad (3.101)$$

Momentler Newton-Raphson iterasyon yöntemi ile elde edilmiştir. r. iterasyon değerleri λ ve v 'nün değerleri aşağıdaki gibi olmaktadır.

$$(\hat{\lambda}_{r+1}) = (\hat{\lambda}_r) + [\sum_{i=1}^I D_{i,\lambda}^T V_{i,\lambda}^{-1} D_{i,\lambda}]_r^{-1} \sum_{i=1}^I D_{i,\lambda}^T V_{i,\lambda}^{-1} (y_i - \theta_i) \quad (3.102)$$

$$(\hat{v}_{r+1}) = (\hat{v}_r) + [\sum_{i=1}^I D_{i,v}^T V_{i,v}^{-1} D_{i,v}]_r^{-1} \sum_{i=1}^I D_{i,v}^T V_{i,v}^{-1} (y_i^2 - \eta_i) \quad (3.103)$$

Burada $\hat{\lambda}_r$ ve $\hat{v}_r$ sırasıyla $\hat{\lambda}$ ve $\hat{v}$ 'nın r. iterasyondaki değerini ve $[\cdot]_r$, r. iterasyondaki değeri ifade eder. Tahmin ediciler tutarlıdır ve yaklaşık olarak normal dağılıma sahiptir [49].

3.3.3. COM-Poisson Regresyon Modeli

Sellers ve Shmueli (2010), McCullagh ve Nelder (1997)'ın geliştirilmiş lineer model yaklaşımına dayandırarak COM-Poisson regresyon modelini log link (bağıntı) fonksiyonunu kullanarak aşağıdaki gibi tanımlamışlardır.

$$\log(E(Y_i)) = \eta_i = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k = X' \beta \quad (3.104)$$

Burada $\log(E(Y_i))$ bağımlı değişkenin tahmin edilen değeri, β_0 doğrunun y eksenini kestiği nokta; $\beta_0, \beta_1, \dots, \beta_k$ bağımsız değişkenlerin katsayıları, $X_1, X_2, \dots, X_k$ bağımsız değişkenleri ifade eder. $E(Y)$ ile $X'\beta$ arasında lineer olmayan bir ilişki olduğundan β ve v , $\beta^{(0)}$ ve $v^{(0)} = 1$ başlangıç değerleri kullanılarak normal denklemler yardımıyla iterasyonlu olarak yeniden ağırlıklandırılmış en küçük kareler ile çözümlenir veya optimizasyon programı kullanarak en çok olabilirlik fonksiyonu maksimizasyonu yardımıyla tahmin edilir. Tahmin edilen katsayılara ait standart hatalar Fisher bilgi matrisinden yararlanarak elde edilir. β ve v tahminleri için bilgi matrisi,

$$I = \begin{pmatrix} I^\beta & I^{\beta,v} \\ I^{\beta,v} & I^v \end{pmatrix} \quad (3.105)$$

elde edilir. Burada I^β , β 'nin tahmin edicisi olan $\hat{\beta}$ 'nin varyans-kovaryans tahminidir. I^v , v 'nin tahmin edicisi olan $\hat{v}$ 'nin varyans tahmini ve $I^{\beta,v}$, $\hat{\beta}$ ve $\hat{v}$ arasındaki kovaryans tahminidir [51, 52, 53].

3.3.4. Yayılım Parametresinin Testi

Sellers ve Shmueli (2010), veri analizinde yayılım parametresinin istatistiksel olarak anlamlı olup olmadığını test etmek için bir test istatistiği geliştirmişlerdir. Başka bir ifadeyle önerilen test, veri analizinde Poisson regresyon ($v = 1$) ile COM-Poisson regresyon modelinin kullanımının uygunluğunu test edilmektedir ($v \neq 1$). Test istatistiği aşağıdaki gibi tanımlanmaktadır,

$$C = -2 \log \Lambda = -2[\log L(\hat{\beta}^{(0)}, \hat{v} = 1) - \log L(\hat{\beta}, \hat{v})] \quad (3.106)$$

Burada Λ , olabilirlik oran test istatistiğini, $\hat{\beta}^{(0)}$ ise $v = 1$ yokluk hipotezi (Poisson dağılımı) altında elde edilen en çok olabilirlik tahmin edicisini ve $(\hat{\beta}, \hat{v})$ ise COM-Poisson dağılımı altında en çok olabilirlik tahmin edicisini göstermektedir. Yokluk hipotezinin doğruluğu altında C yaklaşık olarak 1 serbestlik dereceli χ^2 dağılımına

sahiptir. Küçük örneklemeler için test istatistiğinin dağılımı Bootstrap yöntemi ile tahmin edilebilmektedir [51].

3.3.5. Tahmin Değerlerinin Elde Edilmesi

COM-Poisson regresyon modeli tahmin edildikten sonra tahmin değeri ($\hat{y}_i$) iki farklı şekilde elde edilebilir.

1. Ortalama tahmini yardımıyla: (3.79) denklemindeki yaklaşık ifadeden yararlanarak tahmin değeri,

2.

$$(\hat{y}_i|x_i) = \hat{\lambda}_i^{1/\hat{v}} - \frac{\hat{v}-1}{2\hat{v}} \quad (3.107)$$

elde edilir. Burada $\hat{\lambda}_i = \exp(x_i'\hat{\beta})$ 'dır. Ortalama tahmini yaklaşımı $v \leq 1$ ve $\lambda_i > 10^v$ durumunda geçerlidir [54].

3. Medyan tahmini: Ortalama yaklaşımı geçerli olmadığından $(\hat{y}_i|x_i)$ ve $\hat{v}$ 'nın ters kümülatif yoğunluk fonksiyonunu kullanarak tahmin değerlerinin yüzdelikleri elde edilebilir.

3.3.6. Katsayı Anlamlılığı

Genelleştirilmiş lineer modellere benzer olarak her bir bağımsız değişkenin anlamlılığı,

$$z_i^2 = \left(\frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \right)^2 \quad (3.108)$$

yaklaşık standart normal dağılım yaklaşımından yararlanarak elde edilir. Küçük örneklem durumunda yine Bootstrap'tan yararlanarak ilgilenilen katsayının dağılımı elde edilebilmektedir [35].

3.3.7. Katsayıların Yorumlanması

Katsayıların yorumlanmasında iki farklı yaklaşım bulunmaktadır. Birinci yaklaşım, her bir bağımsız değişkendeki bir birimlik değişimin ortalamaya göre koşullu değişiminin saptanmasını incelemektedir $(E(Y|X_j = x_j, X_{i \neq j} = x))$ ve $E(Y|X_j = x_j + 1, X_{i \neq j} = x)$. İkinci yaklaşım ise lineer olmayan regresyon modeli için katsayı yorumlanmasıdır. İkinci yaklaşımda tahmin değeri ile ve bağımsız değişkendeki değişim doğrudan incelenir [55]. COM-Poisson regresyonda, ortalamayla bağımsız değişkenler arasındaki ilişkinin ne toplamsal ne de çarpımsal olmasından dolayı, yani koşullu ortalamanın oranı bağımsız değişkenlerle lineer olmayan bir ilişki içerdiğinden $(E(Y^v) = \lambda)$ katsayıların yorumlanmasında ilgili katsayının v 'ye bölünerek yorumlanması önerilmektedir [51].

BÖLÜM 4. ARAŞTIRMA BULGULARI

Bu çalışmada 2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan hastaların, merkeze gelme sayılarına etki eden sosyo-demografik faktörlerin saptanması amaçlanmıştır.

Toplum Ruh Sağlığı Merkezi (TRSM), coğrafi bölgelerdeki ağır ruhsal rahatsızlığı olan hastaların (özellikle şizofreni hastaların) ve ailelerin bilgilendirildiği hastanın ayakta tedavisinin yapıldığı ve takip edildiği; rehabilitasyon, psikoeğitim, iş-uğraş terapisi, grup veya bireysel terapi gibi yöntemlerin kullanılarak hastanın toplum içinde yaşama becerilerinin artmasını hedefleyen, psikiatri klinikleri ile ilişki içinde çalışan ve gerektiğinde mobilize ekiplerle hastanın yaşadığı yerde takibini yapan birimler olarak planlanmıştır. Böylece TRSM'ler ağır ruhsal bozukluğu olan hastalar için verilen hizmetin ana düzenleyicisi ve kontrol merkezi olmaktadır. Merkezin ekibi ruh sağlığı ve hastalıkları uzmanı, sosyal hizmet uzmanı, psikolog, hemşire, iş-uğraş terapisti ya da usta öğretici, tıbbi sekreter, idari ve teknik personel, şoför, temizlik elemanı ve güvenlik görevlisinden oluşmaktadır. Ülkemizde 78 ilde 163 TRSM aktif olarak hizmet vermektedir [56].

2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan 244 hastanın cinsiyet, yaş, medeni durum, eğitim düzeyi, sosyal güvenceye sahip olma ve madde bağımlılık durumları incelenmiştir. Analizler R programının “MASS” ve “COMPOissonReg” paketlerinden ve “glm” fonksiyonundan yararlanılarak yapılmıştır.

4.1. TRSM Verileri

2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezinden (TRSM) 244 kayıtlı hasta bulunmaktadır. Bu hastaların TRSM'ye geliş sayısına etki ettiği düşünülen bağımsız değişkenler,

X_1 : Cinsiyet

X_2 : Eğitim Düzeyi

X_3 : Yaş

X_4 : Medeni Durum

X_5 : Sosyal Güvencesi

X_6 : Madde Bağımlılığı

şeklinde ele alınmıştır. Yaş değişkeni dışında diğer değişkenler kategoriktir. Kategorik olan değişkenler sırasıyla cinsiyet 2, eğitim düzeyi 6, medeni durum 4, sosyal güvencesi 2 ve madde bağımlılığı 4 kategoriden oluşmaktadır ve Tablo 4.1.'de özetlenmiştir.

Tablo 4.1. Bağımsız değişken kategorileri

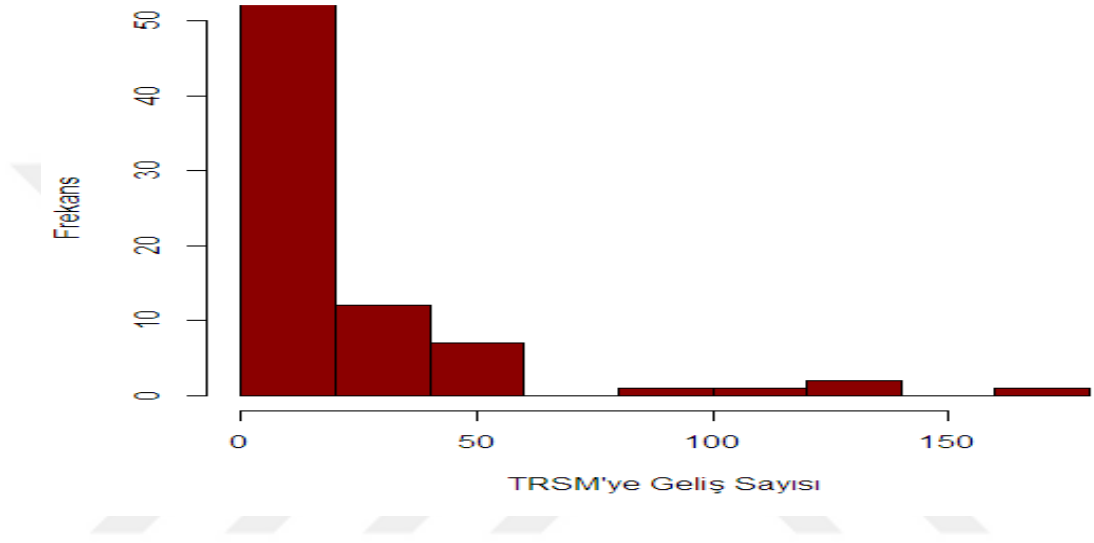
Değişkenler	Kategori Tanımları
Cinsiyet	1: Erkek 2: Kadın
Eğitim Düzeyi	1: Okur-yazar değil 2: İlkokul 3: Ortaokul 4: Lise 5: Önlisans 6: Lisans
Medeni Durum	1: Bekar 2: Evli 3: Boşanmış 4: Dul
Sosyal Güvencesi	1: Yok 2: Var
Madde Bağımlılığı	1: Yok 2: Sigara 3: Alkol 4: Sigara ve Alkol

Modellemede kullanılacak bağımsız değişkenlere ilişkin tanımlayıcı istatistikler Tablo 4.2.'de verilmiştir.

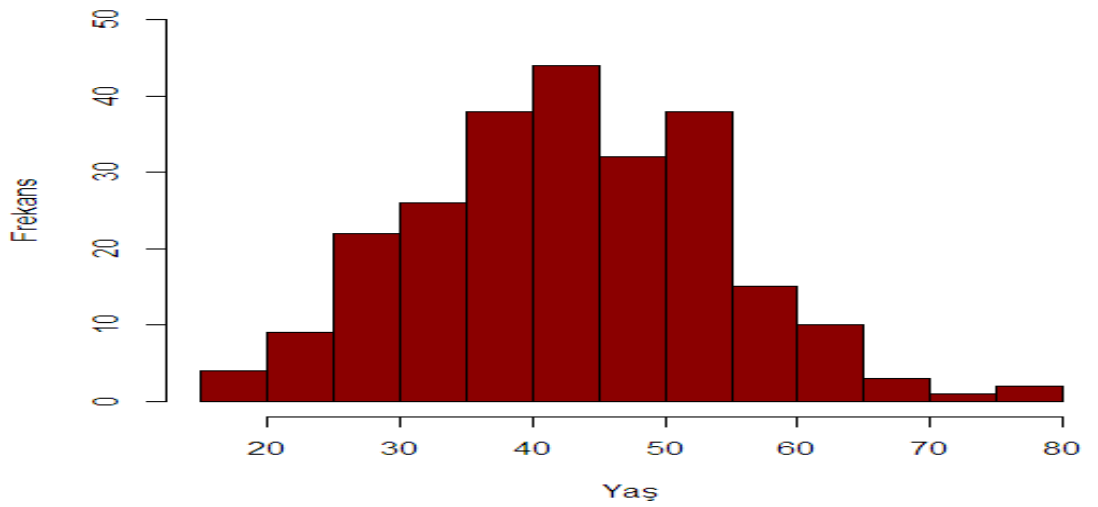
Tablo 4.2. TRSM'ye geliş sayıları ve yaş için özet istatistikler

	Ortalama $\pm$ Standart Sapma	Medyan (Min-Max)
Geliş Sayıları	10,30 $\pm$ 20,1277	5 (0-161)
Yaş	43,47 $\pm$ 11,2467	44 (17-78)

Tablo 4.2.'den geliş sayılarının ortalaması 10,30 ve standart sapması 20,1277 olup ortancası 5 ve aralığı (range) ise 0 ile 161 arasında değişmektedir. Yaş değişkeni için ise ortalaması 43,47 ve standart sapması 11,2467 olup ortanca değeri 44 ve aralığı (range) ise 17 ile 78 arasında değişmektedir. Sırasıyla TRSM'ye geliş sayısı ve yaş için histogram grafikleri Şekil 4.1. ve Şekil 4.2.'de verilmiştir.

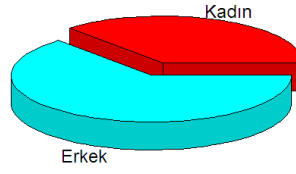


Şekil 4.1. TRSM'ye geliş sayısı dağılımı



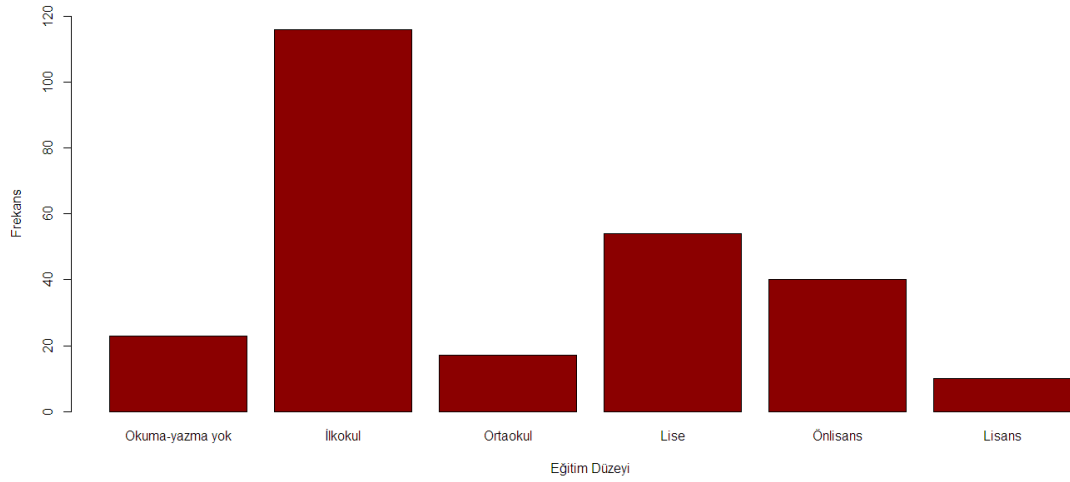
Şekil 4.2. TRSM'ye kayıtlı hastaların yaş dağılımı

Kategorik bağımsız değişkenlere ait daire ve çubuk grafikleri aşağıdadır. Daire grafikleri, cinsiyet (Şekil 4.3.) ve sosyal güvence (Şekil 4.6.) değişkenleri için çizilmiştir. Çubuk grafikleri, eğitim düzeyi (Şekil 4.4.), medeni durum (Şekil 4.5.) ve madde bağımlılık durumu (Şekil 4.7.) değişkenleri için oluşturulmuştur.



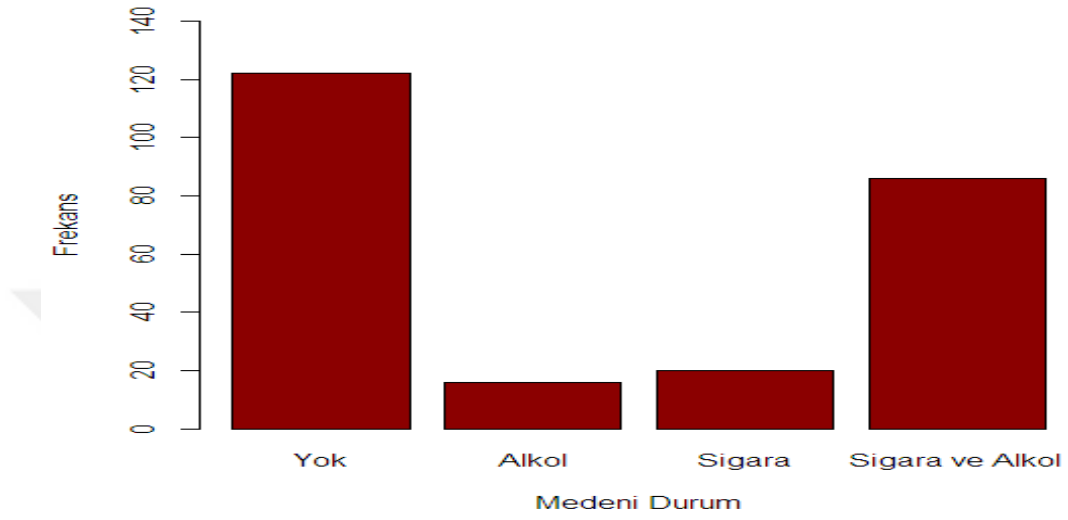
Şekil 4.3. TRSM’de kayıtlı hastaların cinsiyete göre dağılımı

TRSM’yi ziyaret eden hastaların yaklaşık olarak %63’ü erkek %37’si kadındır.



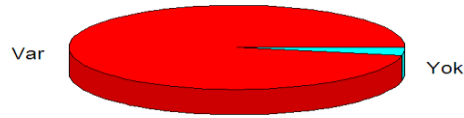
Şekil 4.4. TRSM’de kayıtlı hastaların eğitim düzeylerine göre dağılımı

TRSM'yi ziyaret eden hastaların yaklaşık olarak %9'u okur-yazar değil, %45'i ilkokul, %16'i ortaokul, %19'i lise, %4'ü önlisans ve %7'si lisans eğitim düzeyine sahiptir.



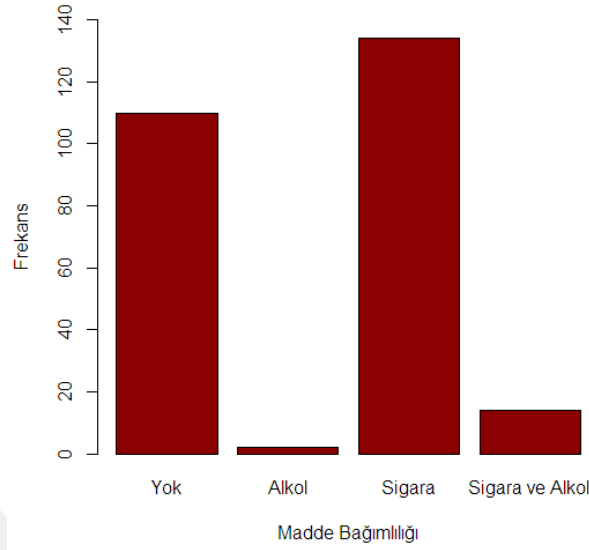
Şekil 4.5. TRSM'de kayıtlı hastaların medeni durumuna göre dağılımı

TRSM'yi ziyaret eden hastaların yaklaşık olarak %50'si bekar, %35'ü evli, %7'si boşanmış ve %8'i duldur.



Şekil 4.6. TRSM'de kayıtlı hastaların sosyal güvencesine göre dağılımı

TRSM'yi ziyaret eden hastaların yaklaşık olarak %3'ünün sosyal güvencesi bulunmamaktadır ve %97'sinin sosyal güvencesi bulunmaktadır.



Şekil 4.7. TRSM’de kayıtlı hastaların madde bağımlılık göre dağılımı

TRSM’yi ziyaret eden hastaların yaklaşık olarak %42’si herhangi bir madde bağımlılığı bulunmamakta, %52’si sigara, %0,08’i alkol, %5’i hem sigara hem alkol kullanmaktadır.

4.1.1. TRSM Verileri İçin Modeller

TRSM’ye geliş sayıları için ele alınacak bağımsız değişkenler, adımsal yöntemlere göre daha güvenilir ve araştırmacıya en iyi alt kümeyi seçme olanağı sağlaması nedeniyle tüm olası alt kümeler yöntemi tercih edilmiştir. Tüm olası alt kümeler yöntemi ele alınarak aşağıdaki modeller oluşturulmuştur. Bu modellere sırasıyla Poisson, negatif binom ve COM-Poisson regresyon uygulanarak en uygun olanı seçilmiştir.

Tüm alt küme yöntemine göre Poisson regresyon sonuçları Tablo 4.3.’te verilmiştir.

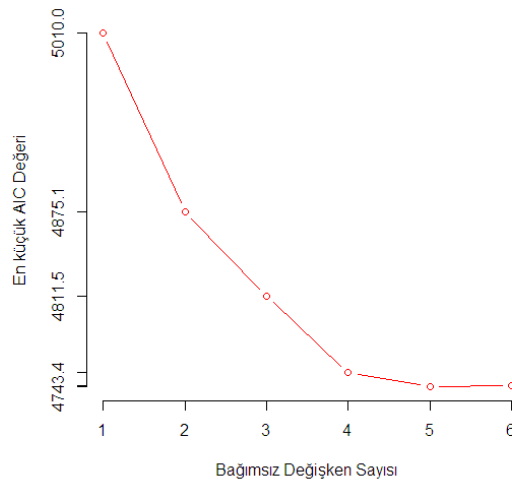
Tablo 4.3. Tüm olası alt kümeler yöntemine göre Poisson regresyon sonuçları

Model	Sapma	Sapma Serbestlik derecesi	AIC	ϕ (yayılm parametresi)
X_1	4346,1	242	5177,8	17,960
X_2	4170,4	238	5010	17,523
X_3	4367,6	242	5199,2	18,048
X_4	4269	240	5104,6	17,788
X_5	4308,8	242	5140,4	17,805
X_6	4369,5	240	5205,1	18,207
X_1, X_2	4098,7	237	4940,3	17,295
X_1, X_3	4330,6	241	5164,2	17,970
X_1, X_4	4233,9	239	5071,5	17,716
X_1, X_5	4276,9	241	5110,5	17,747
X_1, X_6	4326,4	239	5164,1	18,103
X_2, X_3	4168,7	237	5010,3	17,590
X_2, X_4	4029,5	235	4875,1	17,147
X_2, X_5	4113,5	237	4955,1	17,357
X_2, X_6	4161,1	235	5006,7	17,707
X_3, X_4	4267,8	239	5105,4	17,857
X_3, X_5	4295,8	241	5129,4	17,825
X_3, X_6	4355,9	239	5193,5	18,226
X_4, X_5	4182,6	239	5020,2	17,501
X_4, X_6	4252	237	5093,6	17,941
X_5, X_6	4296,1	239	5133,7	17,976
X_1, X_2, X_3	4096,8	236	4940,4	17,360
X_1, X_2, X_4	3963,9	234	4811,5	16,940
X_1, X_2, X_5	4047,9	236	4891,5	17,153
X_1, X_2, X_6	4088,1	234	4935,7	17,471
X_1, X_3, X_4	4232,9	238	5072,5	17,786
X_1, X_3, X_5	4261	240	5096,6	17,755
X_1, X_3, X_6	4311,3	238	5150,9	18,115
X_1, X_4, X_5	4149,3	238	4988,9	17,435
X_1, X_4, X_6	4209,6	236	5053,2	17,838
X_1, X_5, X_6	4257,1	238	5096,7	17,887
X_2, X_3, X_4	4028,6	234	4876,2	17,217
X_2, X_3, X_5	4111,8	236	4955,4	17,423
X_2, X_3, X_6	4159	234	5006,6	17,774
X_2, X_4, X_5	3964,3	234	4811,9	16,942
X_2, X_4, X_6	4014,8	232	4866,4	17,306
X_2, X_5, X_6	4103	234	4950,6	17,535
X_3, X_4, X_5	4181,2	238	5020,8	17,569
X_3, X_4, X_6	4250,5	236	5094,1	18,011
X_3, X_5, X_6	4282,4	238	5122	17,994
X_4, X_5, X_6	4162,1	236	5005,7	17,637
X_1, X_2, X_3, X_4	3962	233	4811,6	17,005
X_1, X_2, X_3, X_5	4045,9	235	4891,5	17,217
X_1, X_2, X_3, X_6	4085,7	233	4935,3	17,536
X_1, X_2, X_4, X_5	3904	233	4753,6	16,756
X_1, X_2, X_4, X_6	3949,1	231	4802,7	17,096
X_1, X_2, X_5, X_6	4037,1	233	4886,7	17,327
X_1, X_3, X_4, X_5	4148	237	4989,6	17,503
X_1, X_3, X_4, X_6	4208,8	235	5054,4	17,910
X_1, X_3, X_5, X_6	4241,6	237	5083,2	17,898
X_1, X_4, X_5, X_6	4124,2	235	4969,8	17,550
X_2, X_3, X_4, X_5	3963,3	233	4812,9	17,010

Tablo 4.3. devamı

Model	Sapma	Sapma Serbestlik derecesi	AIC	ϕ (yayılım parametresi)
X_2, X_3, X_4, X_6	4014,3	231	4867,9	17,378
X_2, X_3, X_5, X_6	4100,7	233	4950,3	17,600
X_2, X_4, X_5, X_6	3946,9	231	4800,5	17,087
X_3, X_4, X_5, X_6	4160,4	235	5006	17,704
X_1, X_2, X_3, X_4, X_5	3902,2	232	4753,8	16,820
X_1, X_2, X_3, X_4, X_6	3947,8	230	4803,4	17,165
X_1, X_2, X_3, X_5, X_6	4034,7	232	4886,3	17,391
X_1, X_2, X_4, X_5, X_6	3887,8	230	4743,4	16,904
X_1, X_3, X_4, X_5, X_6	4123	234	4970,6	17,620
X_2, X_3, X_4, X_5, X_6	3946,4	230	4802	17,159
$X_1, X_2, X_3, X_4, X_5, X_6$	3886,6	229	4744,2	16,973

Buna göre Tablo 4.3.'ten $\phi = \frac{Sapma}{Sapma\ SD}$ değeri 1'den büyük olduğundan yayılım parametresinin tüm modeller için aşırı yayılım gösterdiği görülmüştür. Tüm olası altküne yöntemine göre modeldeki değişken sayısının en küçük AIC değerine karşı çizimi görsel olarak değişimi görmeyi sağlamaktadır (Şekil 4.8.).



Şekil 4.8. Poisson regresyon için altküne yöntemine göre değişken sayısına göre en küçük AIC değeri değişimi

Şekil 4.8.'den görüldüğü gibi modelde bağımsız değişken sayısı arttıkça AIC değerinde azalma görülmektedir. Bu azalma modelde 5 bağımsız değişken

olduğunda en küçük AIC değerine ulaşmakta daha sonra yine artmaktadır. Bu nedenle en uygun model 5 bağımsız değişkeni içeren model olmaktadır. En küçük AIC değerini veren 5 bağımsız değişkenli model,

$$\ln E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 \quad (4.1)$$

şeklinde tanımlanmaktadır. AIC değeri 4743,4 olarak elde edilmiştir.

Katsayıların anlamlılıklarındaki farklılığın gözlemlenebilmesi için modele göre katsayıların sonuçları Tablo 4.4.'te verilmektedir.

Tablo 4.4. Poisson regresyon modeli için katsayıların tahmini

Değişkenler	Tahminler	Standart Hata	z-değeri	p-değeri	Exp(β)
(Sabit)	1,2625	0,1127	11,201	0,000*	3,535
Cinsiyet Erkek (referans)	-				
Kadın	0,3502	0,0452	7,744	0,000*	1,42
Eğitim Düzeyi Okur-yazar değil (referans)	-				
İlkokul	0,8440	0,1080	7,814	0,000*	2,326
Ortaokul	1,1884	0,1138	10,446	0,000*	3,282
Lise	1,2738	0,1107	11,512	0,000*	3,575
Önlisans	0,4689	0,1678	2,794	0,005*	1,599
Lisans	1,2167	0,1232	9,877	0,000*	3,377
Medeni Durum Bekar (referans)	-				
Evli	-0,3992	0,0480	-8,324	0,000*	0,671
Boşanmış	0,4754	0,0712	6,681	0,000*	1,609
Dul	0,0783	0,0760	1,030	0,3029	1,082
Sosyal Güvence Var (referans)	-				
S. G. Yok	-1,5871	0,2694	-5,890	0,000*	0,205
Madde Bağımlılığı Yok (referans)	-				
Sigara	0,1267	0,0439	2,884	0,004*	1,136
Alkol	-0,5037	0,2643	-1,906	0,0567*	0,605
Sigara ve Alkol	-0,0472	0,1036	-0,455	0,6488	0,954

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Tablo 4.4.'ten %5 anlamlılık düzeyine göre cinsiyet, eğitim düzeyi, medeni durum (dul alt kategorisi hariç), sosyal güvence ile madde bağımlılık durumu (sigara ve alkol alt kategorisi hariç) istatistiksel olarak önemlidir, yani TRSM'ye geliş sayısına etki etmektedir. Buna göre; kadınlar erkeklere göre yaklaşık olarak 1,5 kat daha fazla TRSM'ye gelmiştir. Eğitim düzeyi bakımından karşılaştırıldığında referans kategori olan okur-yazar olmayanlara göre tüm eğitim düzeylerinin daha fazla TRSM'ye

geldiği gözlenmiştir. En fazla “Lise” (3,5 kat), en az “Önlisans” (1,6 kat) mezunları TRSM’ye gelmiştir. Medeni durumu evli olanların bekarlara göre yaklaşık olarak yarı yarıya daha az, medeni durumu boşanmış olanların bekarlara göre 1,5 kat daha fazla TRSM’ye geldiği saptanmıştır. Sosyal güvencesi olmayanların olanlara göre yaklaşık olarak 5 kat daha az TRSM’ye geldikleri görülmüştür. Sigara kullananların madde kullanmayanlara göre yaklaşık 1 kat daha fazla, alkol kullananların ise madde kullanmayanlara göre yaklaşık olarak yarı yarıya daha az TRSM’ye geldiği saptanmıştır.

Veriler aşırı yayılım gösterdiğinden alternatif olan negatif binom ve COM-Poisson regresyon sırasıyla uygulanmıştır.

Tüm alt küme yöntemine göre negatif binom regresyon sonuçları Tablo 4.5.’te verilmiştir.

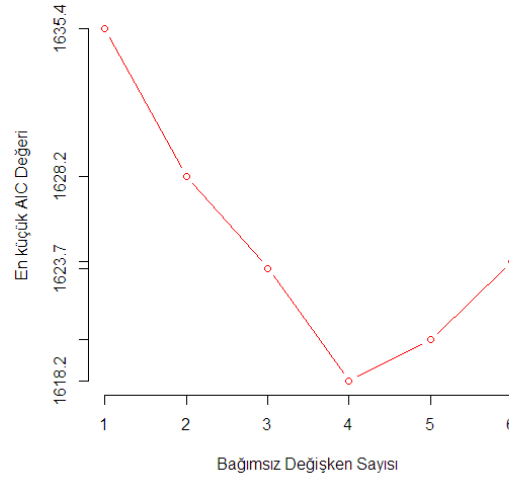
Tablo 4.5. Tüm olası alt kümeler yöntemine göre negatif binom regresyon sonuçları

Model	α	Sapma	Sapma SD	AIC	ϕ (yayılım parametresi)
X_1	0,7518	270,88	242	1641	1,120
X_2	0,7892	268,78	238	1635,4	1,130
X_3	0,7478	271,10	242	1642,5	1,121
X_4	0,7666	270,08	240	1639,6	1,126
X_5	0,7655	269,68	242	1635,5	1,115
X_6	0,7477	271,10	240	1646,2	1,130
X_1, X_2	0,8146	267,05	237	1628,2	1,127
X_1, X_3	0,7571	270,50	241	1641	1,123
X_1, X_4	0,7761	269,50	239	1638,1	1,128
X_1, X_5	0,7714	269,39	241	1635,4	1,118
X_1, X_6	0,7545	270,76	239	1646	1,133
X_2, X_3	0,7892	268,77	237	1637,4	1,135
X_2, X_4	0,8176	267,46	235	1631,8	1,139
X_2, X_5	0,8083	267,38	237	1630,4	1,129
X_2, X_6	0,7910	268,69	235	1640,8	1,144
X_3, X_4	0,7670	270,04	239	1641,4	1,130
X_3, X_5	0,7679	269,54	241	1636,6	1,119
X_3, X_6	0,7495	271,01	239	1647,9	1,134
X_4, X_5	0,7899	268,37	239	1632,8	1,123
X_4, X_6	0,7707	269,81	237	1644	1,139
X_5, X_6	0,7676	269,61	239	1640,8	1,129
X_1, X_2, X_3	0,8152	267	236	1630	1,132
X_1, X_2, X_4	0,8451	265,67	234	1624,4	1,136
X_1, X_2, X_5	0,8327	265,62	236	1623,7	1,126
X_1, X_2, X_6	0,8173	266,87	234	1633,3	1,141
X_1, X_3, X_4	0,7776	269,37	238	1639,5	1,132

Tablo 4.5. devamı

Model	α	Sapma	Sapma SD	AIC	ϕ (yayılım parametresi)
X_1, X_3, X_5	0,7769	269,01	240	1635,3	1,121
X_1, X_3, X_6	0,7595	270,40	238	1646,1	1,137
X_1, X_4, X_6	0,7812	269,19	236	1642,2	1,141
X_1, X_5, X_6	0,7738	269,35	238	1640,6	1,132
X_2, X_3, X_4	0,8177	267,46	234	1633,8	1,143
X_2, X_3, X_5	0,8083	267,38	236	1632,4	1,133
X_2, X_3, X_6	0,7911	268,67	234	1642,7	1,149
X_2, X_4, X_5	0,8407	265,67	234	1625,5	1,136
X_2, X_4, X_6	0,8230	267,05	232	1635,9	1,152
X_2, X_5, X_6	0,8104	267,31	234	1635,7	1,143
X_3, X_4, X_5	0,7903	267,34	238	1634,6	1,124
X_3, X_4, X_6	0,7710	269,79	236	1645,9	1,144
X_3, X_5, X_6	0,7697	269,50	238	1642	1,133
X_4, X_5, X_6	0,7941	268,21	236	1637,4	1,137
X_1, X_2, X_3, X_4	0,8451	265,66	233	1626,3	1,141
X_1, X_2, X_3, X_5	0,8335	265,53	235	1625,4	1,130
X_1, X_2, X_3, X_6	0,8179	266,84	233	1635,1	1,146
X_1, X_2, X_4, X_5	0,8685	263,78	233	1618,2	1,133
X_1, X_2, X_4, X_6	0,8524	265,14	231	1627,8	1,148
X_1, X_2, X_5, X_6	0,8352	265,53	233	1628,9	1,140
X_1, X_3, X_4, X_5	0,8010	267,74	237	1632,9	1,130
X_1, X_3, X_4, X_6	0,7825	269,08	235	1643,7	1,146
X_1, X_3, X_5, X_6	0,7789	269	237	1640,7	1,136
X_1, X_4, X_5, X_6	0,8043	267,67	235	1635,8	1,140
X_2, X_3, X_4, X_5	0,8408	265,68	233	1627,5	1,141
X_2, X_3, X_4, X_6	0,8233	267,05	231	1637,8	1,157
X_2, X_3, X_5, X_6	0,8104	267,31	233	1637,7	1,148
X_2, X_4, X_5, X_6	0,8466	265,33	231	1629,6	1,149
X_3, X_4, X_5, X_6	0,7944	268,19	235	1639,3	1,142
X_1, X_2, X_3, X_4, X_5	0,8685	263,77	232	1620,2	1,137
X_1, X_2, X_3, X_4, X_6	0,8524	265,14	230	1629,8	1,153
X_1, X_2, X_3, X_5, X_6	0,8359	265,46	232	1630,7	1,145
X_1, X_2, X_4, X_5, X_6	0,8754	263,38	230	1622	1,146
X_1, X_3, X_4, X_5, X_6	0,8055	267,59	234	1637,4	1,144
X_2, X_3, X_4, X_5, X_6	0,8468	265,33	230	1631,5	1,154
$X_1, X_2, X_3, X_4, X_5, X_6$	0,8754	267,37	229	1624	1,168

Buna göre Tablo 4.5.'ten $\phi = \frac{\text{Sapma}}{\text{Sapma SD}}$ değeri 1'e yakın olduğundan yayılım parametresinin tüm modeller için eş yayılım gösterdiği görülmüştür. Tüm olası altküme yöntemine göre modeldeki değişken sayısının en küçük AIC değerine karşı çizimi görsel olarak değişimi görmeyi sağlamaktadır (Şekil 4.9.).



Şekil 4.9. Negatif binom regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi

Şekil 4.9.'dan görüldüğü gibi modelde bağımsız değişken sayısı arttıkça AIC değerinde azalma görülmektedir. Bu azalma modelde 4 bağımsız değişken olduğunda en küçük AIC değerine ulaşmakta daha sonra yine artmaktadır. Bu nedenle en uygun model 4 bağımsız değişkeni içeren model olmaktadır. En küçük AIC değerini veren 4 bağımsız değişkenli model,

$$\ln E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 \quad (4.2)$$

şeklinde tanımlanmaktadır. AIC değeri 1618,2 olarak elde edilmiştir. Modelden de görüldüğü gibi aşırı yayılım gösteren Poisson regresyonda anlamlı bulunan madde bağımlılığı (X_6) negatif binom regresyonda anlamlı bulunmamış ve AIC değeri 4743,4'ten 1618,2'e düşmüştür.

Katsayıların anlamlılıklarındaki farklılığın gözlemlenebilmesi için modele göre katsayıların sonuçları Tablo 4.6.'da verilmiştir.

Tablo 4.6. Negatif binom regresyon modeli için katsayıların tahmini

Değişkenler	Tahminler	Standart Hata	z-değeri	p-değeri	Exp(β)
(Sabit)	1,2315	0,2917	4,221	0,000*	3,427
Cinsiyet Erkek (referans)	-				
Kadın	0,5142	0,1587	3,241	0,001*	1,673
Eğitim Düzeyi Okur-yazar değil (referans)	-				
İlkokul	0,7982	0,2859	2,792	0,005*	2,222
Ortaokul	1,2001	0,3252	3,690	0,0002*	3,321
Lise	0,4177	0,3138	4,519	0,000*	1,519
Önlisans	0,5389	0,4692	1,149	0,2508	1,715
Lisans	1,3347	0,3788	3,524	0,0004*	3,799
Medeni Durum Bekar (referans)	-				
Evli	-0,3947	0,1616	-2,443	0,0146*	0,674
Boşanmış	0,5349	0,3032	1,765	0,077	1,708
Dul	0,0161	0,2858	0,056	0,955	1,017
Sosyal Güvence Var (referans)	-				
S. G. Yok	-1,6324	0,5115	-3,191	0,001*	0,196

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Tablo 4.6.'dan %5 anlamlılık düzeyine göre cinsiyet, eğitim düzeyi (önlisans alt kategorisi hariç), medeni durum (Boşanmış ve dul alt kategorileri hariç) ile sosyal güvence istatistiksel olarak önemlidir, yani TRSM'ye geliş sayısına etki etmektedir. Buna göre; kadınlar erkeklere göre yaklaşık olarak 1,5 kat daha fazla TRSM'ye gelmiştir. Eğitim düzeyi bakımından karşılaştırıldığında referans kategori olan okur-yazar olmayanlara göre tüm eğitim düzeylerinin daha fazla TRSM'ye geldiği gözlenmiştir. Ancak Poisson regresyondan farklı olarak en fazla "Lisans" (yaklaşık 4 kat), en az "Lise" (1,5 kat) mezunlarının daha sık TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak medeni durumu evli olanların bekarlara göre yaklaşık olarak yarı yarıya daha az TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak sosyal güvencesi olmayanların olanlara göre yaklaşık olarak 5 kat daha az TRSM'ye geldikleri görülmüştür.

Son olarak tüm alt küme yöntemine göre COM-Poisson regresyon sonuçları Tablo 4.7.'de verilmiştir.

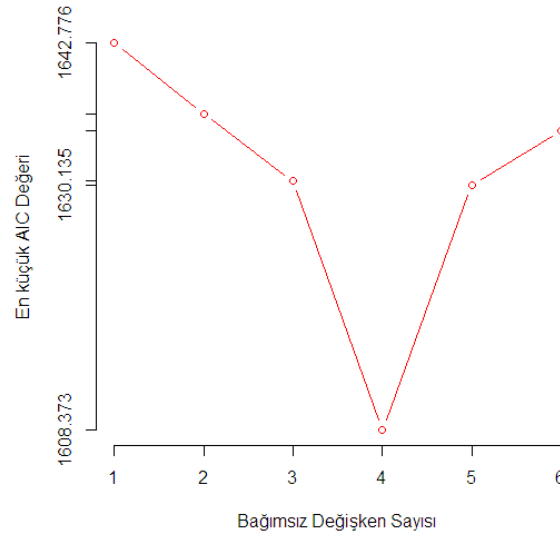
Tablo 4.7. Tüm olası alt kümeler yöntemine göre COM-Poisson regresyon sonuçları

Model	ν	$\nu = 1$ için Test Sonucu (χ^2)	p değeri	AIC
X_1	0,045	3527,7384	0,000*	1652,013
X_2	0,53	3369,2140	0,000*	1642,776
X_3	0,349	3547,3420	0,000*	1653,864
X_4	0,375	3457,6311	0,000*	1649,008
X_5	0,639	3497,3423	0,000*	1645,079
X_6	0,143	3549,1193	0,000*	1658,006
X_1, X_2	0,473	3303,3107	0,000*	1639,028
X_1, X_3	0,269	3513,5659	0,000*	1652,663
X_1, X_4	0,181	3425,4457	0,000*	1648,060
X_1, X_5	0,271	3468,1101	0,000*	1644,352
X_1, X_6	0,056	3509,7615	0,000*	1656,293
X_2, X_3	0,432	3367,6564	0,000*	1644,634
X_2, X_4	0,113	3240,0784	0,000*	1637,042
X_2, X_5	0,135	3320,6479	0,000*	1636,490
X_2, X_6	0,484	3360,7767	0,000*	1647,973
X_3, X_4	0,653	3456,5065	0,000*	1650,903
X_3, X_5	0,419	3485,4833	0,000*	1645,947
X_3, X_6	0,047	3536,7487	0,000*	1658,800
X_4, X_5	0,451	3336,8375	0,000*	1655,596
X_4, X_6	0,677	3442,0092	0,000*	1653,564
X_5, X_6	0,3	3485,8002	0,000*	1649,937
X_1, X_2, X_3	0,194	3301,4817	0,000*	1640,891
X_1, X_2, X_4	0,72	3179,1122	0,000*	1634,369
X_1, X_2, X_5	0,523	3260,0604	0,000*	1633,411
X_1, X_2, X_6	0,447	3293,4257	0,000*	1644,232
X_1, X_3, X_4	0,746	3424,5457	0,000*	1649,986
X_1, X_3, X_5	0,335	3453,5977	0,000*	1644,975
X_1, X_3, X_6	0,402	3495,9097	0,000*	1656,995
X_1, X_4, X_5	0,117	3489,7905	0,000*	1645,596
X_1, X_4, X_6	0,069	3403,1876	0,000*	1652,036
X_1, X_5, X_6	0,105	3450,0827	0,000*	1648,647
X_2, X_3, X_4	0,368	3239,1393	0,000*	1639,025
X_2, X_3, X_5	0,803	3318,9960	0,000*	1638,407
X_2, X_3, X_6	0,319	3358,7655	0,000*	1649,838
X_2, X_4, X_5	0,346	3183,4518	0,000*	1630,478

Tablo 4.7. devamı

Model	ν	$\nu = 1$ için Test Sonucu (χ^2)	p değeri	AIC
X_2, X_4, X_6	0,328	3226,4091	0,000*	1643,000
X_2, X_5, X_6	0,527	3310,8786	0,000*	1641,681
X_3, X_4, X_5	0,204	3220,9349	0,000*	1642,844
X_3, X_4, X_6	0,716	3440,6468	0,000*	1655,435
X_3, X_5, X_6	0,149	3473,2774	0,000*	1650,755
X_4, X_5, X_6	0,642	3363,2167	0,000*	1644,500
X_1, X_2, X_3, X_4	0,194	3177,4932	0,000*	1636,136
X_1, X_2, X_3, X_5	0,719	3258,2974	0,000*	1635,172
X_1, X_2, X_3, X_6	0,292	3291,1243	0,000*	1646,176
X_1, X_2, X_4, X_5	0,074	3127,2581	0,000*	1608,373
X_1, X_2, X_4, X_6	0,156	3165,3772	0,000*	1639,281
X_1, X_2, X_5, X_6	0,112	3250,2067	0,000*	1638,520
X_1, X_3, X_4, X_5	0,299	3185,0367	0,000*	1640,368
X_1, X_3, X_4, X_6	0,327	3402,3950	0,000*	1653,971
X_1, X_3, X_5, X_6	0,305	3435,9135	0,000*	1649,336
X_1, X_4, X_5, X_6	0,189	3328,3469	0,000*	1643,418
X_2, X_3, X_4, X_5	0,568	3182,5037	0,000*	1632,444
X_2, X_3, X_4, X_6	0,383	3225,9498	0,000*	1643,943
X_2, X_3, X_5, X_6	0,44	3308,7113	0,000*	1643,631
X_2, X_4, X_5, X_6	0,194	3167,1209	0,000*	1635,351
X_3, X_4, X_5, X_6	0,352	3361,6694	0,000*	1646,340
X_1, X_2, X_3, X_4, X_5	0,537	3125,6895	0,000*	1630,135
X_1, X_2, X_3, X_4, X_6	0,758	3164,2442	0,000*	1641,113
X_1, X_2, X_3, X_5, X_6	0,008	3247,8312	0,000*	1640,486
X_1, X_2, X_4, X_5, X_6	0,303	3112,3418	0,000*	1633,065
X_1, X_3, X_4, X_5, X_6	0,436	3327,2814	0,000*	1645,338
X_2, X_3, X_4, X_5, X_6	0,01	3166,4895	0,000*	1637,476
$X_1, X_2, X_3, X_4, X_5, X_6$	0,487	3111,1820	0,000*	1635,005

Buna göre Tablo 4.7.'den ν değeri 1'den küçük olduğundan yayılım parametresinin tüm modeller için aşırı yayılım gösterdiği görülmüştür. Tüm olası altküme yöntemine göre modeldeki değişken sayısının en küçük AIC değerine karşı çizimi görsel olarak değişimi görmeyi sağlamaktadır (Şekil 4.10.).



Şekil 4.10. COM-Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi

Şekil 4.10.'dan görüldüğü gibi modelde bağımsız değişken sayısı arttıkça AIC değerinde azalma görülmektedir. Bu azalma modelde 4 bağımsız değişken olduğunda en küçük AIC değerine ulaşmakta daha sonra yine artmaktadır. Bu nedenle en uygun model 4 bağımsız değişkeni içeren model olmaktadır. En küçük AIC değerini veren 4 bağımsız değişkenli model,

$$\ln E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \beta_5 X_5 \quad (4.3)$$

şeklinde tanımlanmaktadır. AIC değeri 1608,373 olarak elde edilmiştir. Modelden de görüldüğü gibi aşırı yayılım gösteren Poisson regresyonda anlamlı bulunan madde bağımlılığı (X_6) COM-Poisson regresyonda anlamlı bulunmamış ve AIC değeri 1618,2'den 1608,373'e düşmüştür.

Katsayıların anlamlılıklarındaki farklılığın gözlemlenebilmesi için modele göre katsayıların sonuçları Tablo 4.8.'de verilmiştir.

Tablo 4.8. COM Poisson regresyon modeli için katsayıların tahmini

Değişkenler	Tahminler	Standart Hata	z-değeri	p-değeri	Exp(β)
(Sabit)	-0,2030	0,0432	-4,6978	0,000*	0,064
Cinsiyet Erkek (referans)	-				
Kadın	0,0236	0,0114	2,0709	0,038*	1,376
Eğitim Düzeyi Okur-yazar değil (referans)	-				
İlkokul	0,0967	0,0433	2,2340	0,025*	3,694
Ortaokul	0,1260	0,0439	2,8685	0,004*	5,489
Lise	0,1295	0,0432	2,9962	0,003*	5,755
Önlisans	0,0573	0,0626	0,9142	0,361	2,169
Lisans	0,1254	0,0455	2,7559	0,006*	5,444
Medeni Durum Bekar (referans)	-				
Evli	-0,0349	0,0150	-2,3232	0,020*	0,624
Boşanmış	0,0255	0,0160	1,5882	0,112	1,411
Dul	0,0083	0,0205	0,4038	0,686	1,119
Sosyal Güvence Var (referans)	-				
S. G. Yok	-0,0775	0,1369	-2,027	0,021*	0,351

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Tablo 4.8.'den %5 anlamlılık düzeyine göre cinsiyet, eğitim düzeyi (önlisans alt kategorisi hariç), medeni durum (Boşanmış ve dul alt kategorileri hariç) ile sosyal güvence istatistiksel olarak önemlidir, yani TRSM'ye geliş sayısına etki etmektedir.

Katsayıların yorumlanmasında ilgili katsayının v 'ye (0,074) bölünmesinden sonra üstelinin alınarak bir yöntem izlenmesi önerilmektedir. Buna göre; kadınlar erkeklere göre yaklaşık olarak 1,5 kat daha fazla TRSM'ye gelmiştir. Eğitim düzeyi bakımından karşılaştırıldığında referans kategori olan okur-yazar olmayanlara göre tüm eğitim düzeylerinin daha fazla TRSM'ye geldiği gözlenmiştir. Ancak Poisson regresyondan farklı olarak en fazla "Lise" (yaklaşık 6 kat), en az "İlkokul" (4 kat) mezunlarının daha sık TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak medeni durumu evli olanların bekarlara göre yaklaşık olarak yarı yarıya daha az TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak sosyal güvencesi olmayanların olanlara göre yaklaşık olarak 3 kat daha az TRSM'ye geldikleri görülmüştür.

4.2. Sosyal Medya Verileri

2015-2016 yılları arasında Giresun Üniversitesi Bilimsel Araştırma Proje desteği ile (SOS-BAP-A-200515-43) gerçekleştirilen proje verilerinin bir kısmı kullanılmıştır. Burada 231 öğretmenin sosyal medya hesap sayıları cinsiyet, yaş grupları (45 yaş altı) ve okul türüne göre incelenmiştir. Analizler R programının “COMPOissonReg” paketlerinden ve “glm” fonksiyonundan yararlanılarak yapılmıştır. Değişkenler aşağıdaki gibi tanımlanmıştır;

X_1 : Okul Türü

X_2 : Cinsiyet

X_3 : Yaş Grupları

İncelenen tüm bağımsız değişkenler kategoriktir. Kategorik olan değişkenler sırasıyla okul türü 2, cinsiyet 2, yaş grupları 5 kategoriden oluşmaktadır ve Tablo 4.9.’da özetlenmiştir.

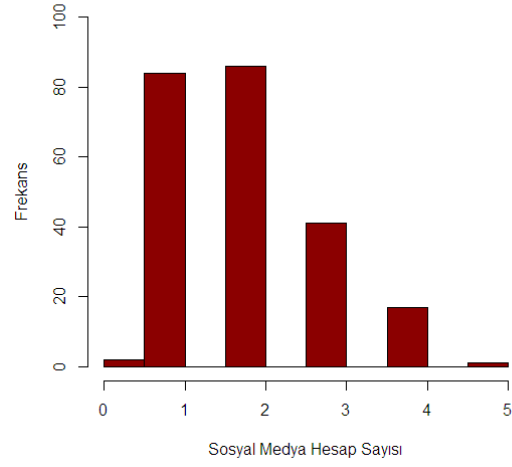
Tablo 4.9. Bağımsız değişken kategorileri

Değişkenler	Kategori Tanımları				
Okul Türü	1: Devlet	2: Özel			
Cinsiyet	1: Erkek	2: Kadın			
Yaş Grupları	1: 18-24	2: 25-29	3: 30-34	4: 35-39	5: 40-44

Tablo 4.10. Öğretmenlerin sosyal medya hesap sayısı için özet istatistikler

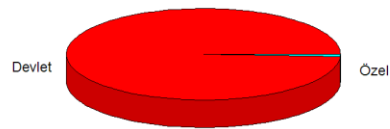
	Ortalama $\pm$ Standart Sapma	Medyan (Min-Max)
Sosyal Medya Hesap Sayısı	1,96 $\pm$ 0,95	2 (0-5)

Tablo 4.10.’dan öğretmenlerin sosyal medya hesap sayısı ortalaması 1,96 ve standart sapması 0.95 olup medyanı 2 ve aralığı (range) ise 0 ile 5 arasında değişmektedir. Sosyal medya hesap sayısı için histogram grafiği Şekil 4.11.’de verilmiştir.



Şekil 4.11. Öğretmenlerin sosyal medya hesap sayılarına göre dağılımı

Kategorik bağımsız değişkenlere ait daire ve çubuk grafikleri aşağıdadır. Daire grafikleri, okul türü (Şekil 4.12.) ve cinsiyet (Şekil 4.13.) değişkenleri için çizilmiştir. Çubuk grafiği, yaş grupları (Şekil 4.14.) değişkeni için oluşturulmuştur.



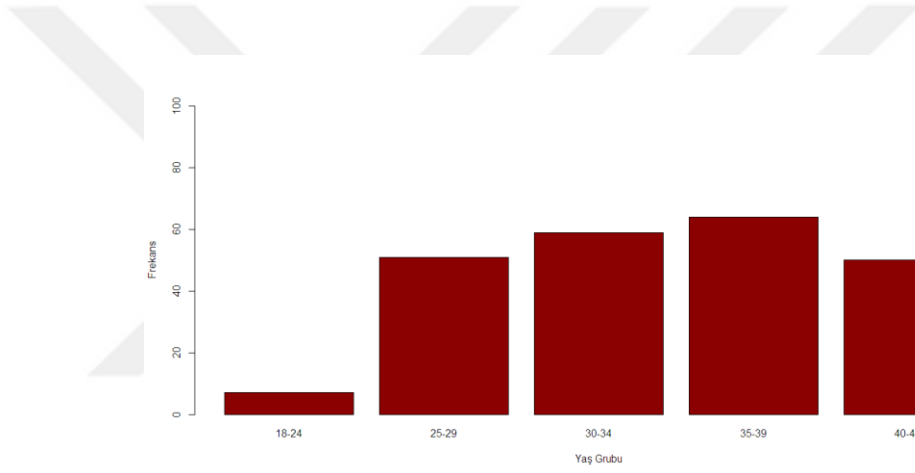
Şekil 4.12. Öğretmenlerin okul türüne göre dağılımı

Öğretmenlerin yaklaşık olarak %99'u devlet, %1'i özel okuldadır.



Şekil 4.13. Öğretmenlerin cinsiyetlere göre dağılımı

Öğretmenlerin yaklaşık olarak %39'u erkek, %61'i kadındır.



Şekil 4.14. Öğretmenlerin yaş gruplarına göre dağılımı

Öğretmenlerin yaklaşık olarak %3'ü “18-24”, %22'si “25-29”, %26'sı “30-34”, %28'i “35-39” ve %21'i “40-44” yaş grubundadır.

4.2.1. Sosyal Medya Verileri İçin Modeller

Öğretmenlerin sosyal medya hesap sayıları için ele alınacak bağımsız değişkenler, adımsal yöntemlere göre daha güvenilir ve araştırmacıya en iyi alt kümeyi seçme olanağı sağlaması nedeniyle tüm olası alt kümeler yöntemi tercih edilmiştir. Tüm olası alt kümeler yöntemi ele alınarak aşağıdaki modeller oluşturulmuştur. Bu

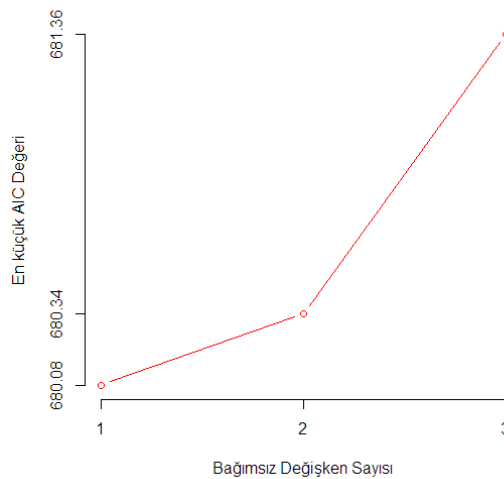
modellere sırasıyla Poisson, negatif binom ve COM-Poisson regresyon uygulanarak en uygun olanı seçilmiştir.

Tüm alt küme yöntemine göre Poisson regresyon sonuçları Tablo 4.11.'de verilmiştir.

Tablo 4.11. Tüm olası alt kümeler yöntemine göre Poisson regresyon sonuçları

Model	Sapma	Sapma SD	AIC	ϕ (yayılım parametresi)
X_1	101,63	229	680,08	0,444
X_2	105,93	229	684,37	0,463
X_3	98,29	226	682,74	0,435
X_1, X_2	101,13	228	681,57	0,444
X_1, X_3	93,90	225	680,34	0,418
X_2, X_3	97,16	225	683,60	0,432
X_1, X_2, X_3	92,92	224	681,36	0,415

Buna göre Tablo 4.11.'den $\phi = \frac{Sapma}{Sapma\ SD}$ değeri 1'den küçük olduğundan yayılım parametresinin tüm modeller için az yayılım gösterdiği görülmüştür. Tüm olası altküme yöntemine göre modeldeki değişken sayısının en küçük AIC değerine karşı çizimi görsel olarak değişimi görmeyi sağlamaktadır (Şekil 4.15.).



Şekil 4.15. Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi

Şekil 4.15.'ten görüldüğü gibi modelde bağımsız değişken sayısı arttıkça AIC değerinde artma görülmektedir. Modelde sadece bir bağımsız değişken olduğunda en küçük AIC değerine ulaşmakta daha sonra yine artmaktadır. Bu nedenle en uygun model tek bağımsız değişkeni içeren model olmaktadır. En küçük AIC değerini veren tek bağımsız değişkenli model,

$$\ln E(Y) = \beta_0 + \beta_1 X_1 \quad (4.4)$$

olarak gösterilir. AIC değeri 680,08 olarak elde edilmiştir.

Katsayıların anlamlılıklarındaki farklılığın gözlemlenebilmesi için modele göre katsayıların sonuçları Tablo 4.12.'de verilmektedir.

Tablo 4.12. Poisson regresyon modeli için katsayıların tahmini

Değişkenler	Tahminler	Standart Hata	z-değeri	p-değeri	Exp(β)
(Sabit)	0,6599	0,0475	13,888	0,000*	1,935
Okul Türü Devlet (referans)	-				
Özel	0,8442	0,3367	2,507	0,0122*	2,327

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Tablo 4.12.'den %5 anlamlılık düzeyine göre sadece okul türü istatistiksel olarak önemlidir, yani sosyal medya hesap sayısına etki etmektedir. Buna göre; özel okulda görev yapan öğretmenlerin devlet okulunda görev yapan öğretmenlere göre yaklaşık 2,5 kat daha fazla sosyal medya hesabı bulunduğu saptanmıştır.

Veriler az yayılım gösterdiğinden dolayı negatif binom regresyon modeli için değerler hesaplanamamaktadır.

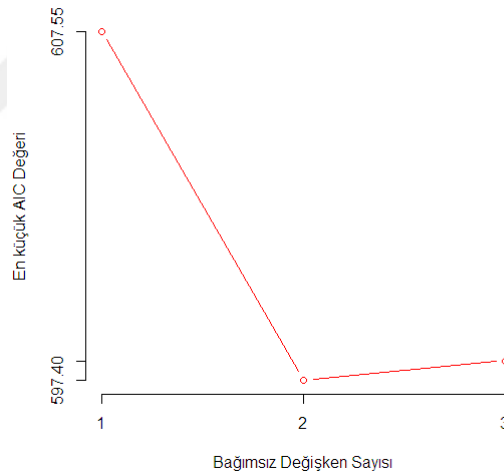
Son olarak tüm alt küme yöntemine göre COM-Poisson regresyon sonuçları Tablo 4.13.'te verilmiştir.

Tablo 4.13. Tüm olası alt kümeler yöntemine göre COM-Poisson regresyon sonuçları

Model	ν	$\nu = 1$ için Test Sonucu (χ^2)	p değeri	AIC
X_1	2,822	73,7053	0,000*	609,37
X_2	2,693	66,5182	0,000*	619,85
X_3	2,921	77,1836	0,000*	607,55
X_1, X_2	2,837	73,4066	0,000*	610,16
X_1, X_3	3,075	84,3789	0,000*	597,40
X_2, X_3	2,957	78,8612	0,000*	606,74
X_1, X_2, X_3	3,110	85,9608	0,000*	597,96

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Buna göre Tablo 4.13.'ten ν değeri 1'den büyük olduğundan yayılım parametresinin tüm modeller için az yayılım gösterdiği görülmüştür. Tüm olası altküme yöntemine göre modeldeki değişken sayısının en küçük AIC değerine karşı çizimi görsel olarak değişimi görmeyi sağlamaktadır (Şekil 4.16.).



Şekil 4.16. COM-Poisson regresyon için altküme yöntemine göre değişken sayısına göre en küçük AIC değer değişimi

Şekil 4.16.'dan görüldüğü gibi modelde bağımsız değişken sayısı arttıkça AIC değerinde azalma görülmektedir. Bu azalma modelde 2 bağımsız değişken olduğunda en küçük AIC değerine ulaşmakta daha sonra yine artmaktadır. Bu nedenle en uygun model 2 bağımsız değişkeni içeren model olmaktadır. En küçük AIC değerini veren 2 bağımsız değişkenli model,

$$\ln E(Y) = \beta_0 + \beta_1 X_1 + \beta_3 X_3 \quad (4.5)$$

şeklinde tanımlanmaktadır. AIC değeri 597,40 olarak elde edilmiştir. Modelden de görüldüğü gibi aşırı yayılım gösteren Poisson regresyonda anlamsız bulunan yaş grupları (X_3) COM-Poisson regresyonda anlamlı bulunmuştur ve AIC değeri 680,08'den 597,40'e düşmüştür.

Katsayıların anlamlılıklarındaki farklılığın gözlemlenebilmesi için modele göre katsayıların sonuçları Tablo 4.14.'te verilmektedir.

Tablo 4.14. COM Poisson regresyon modeli için katsayıların tahmini

Değişkenler	Tahminler	Standart Hata	z-değeri	p-değeri	Exp(β)
(Sabit)	3,8457	0,5244	7,3334	0,000*	3,493
Okul Türü Devlet (referans)	-				
Özel	2,1942	0,6081	3,6084	0,000*	2,042
Yaş Grubu 18-24 (referans)	-				
25-29	-1,0288	0,3977	-2,5865	0,01*	0,716
30-34	-1,3538	0,4045	-3,3467	0,000*	0,644
35-39	-1,3562	0,4024	-3,3704	0,000*	0,644
40-44	-1,6862	0,4234	-3,9825	0,000*	0,578

* %5 anlamlılık düzeyine göre istatistiksel olarak önemli

Tablo 4.14.'ten %5 anlamlılık düzeyine göre okul türü ve yaş grubu istatistiksel olarak önemlidir, yani sosyal medya hesap sayısına etki etmektedir.

Katsayıların yorumlanmasında ilgili katsayının v 'ye (3,075) bölünmesinden sonra üstelinin alınarak bir yöntem izlenmesi önerilmektedir. Özel okulda görev yapan öğretmenlerin devlet okulunda görev yapan öğretmenlere göre yaklaşık 2 kat daha fazla sosyal medya hesabı bulunduğu saptanmıştır. Yaş grubu bakımından karşılaştırıldığında referans kategori olan 18-24'a göre tüm yaş gruplarının daha az sosyal medya hesabı bulunduğu saptanmıştır. Yaş grubu arttıkça sosyal medya hesap sayısı azalmaktadır. 40-44 yaş grubundakilerin 18-24 yaş grubuna göre sosyal medya hesap sayısı yaklaşık yarı yarıya düşmektedir. 25-29 yaş grubundakilerin 18-24 yaş grubuna göre sosyal medya hesap sayısı yaklaşık %30 azalmaktadır.

BÖLÜM 5. TARTIŞMA VE SONUÇLAR

Poisson regresyon modeli, bağımlı değişkenin sayıma verisi olması durumunda sıklıkla kullanılmaktadır. Burada regresyon modeli, Poisson dağılımlı bağımlı değişkenin beklenen değerine bağımsız değişkenleri bağlayan bir fonksiyondur. Poisson regresyon modeli uygulamaları sıklıkla belli kategorilere ayrılan sayıma veriler ile yapılmaktadır.

Poisson dağılımı, ortalama ile varyans eşitliğini varsaymaktadır, ancak uygulamalarda genellikle bu varsayım sağlanamamaktadır. Eşitlik varsayımının ihlali nedensellik ilişkisi kurulan bağımsız değişkenlere ait katsayı tahminlerinin tutarlı ancak yeterli olmamasına yol açtığından Poisson regresyon modeli ile analiz edilmesi uygun olmamaktadır. Bu durum alternatif sayma verisi regresyon modellerinin araştırılmasına neden olmuştur.

Varyansın ortalamadan büyük olmasına aşırı, küçük olmasına ise az yayılım adı verilmektedir. Uygulamalarda daha çok varyansın ortalamadan büyük olduğu aşırı yayılım durumu ile karşılaşmaktadır. Aşırı yayılım durumu için genellikle negatif binom regresyon modeli uygulanmaktadır.

Bu amaçla ikinci bölümde Poisson regresyon modelinin üyesi olduğu genelleştirilmiş lineer model hakkında bilgi verilmiştir. Üçüncü bölümde sayımla elde edilen verilerin analizinde sıklıkla kullanılan Poisson ve negatif binom ele alınmıştır. Regresyon modeli, parametre tahmini, anlamlılık ve yorumlanmasından kısaca bahsedilmiştir. Dördüncü bölümde Conway-Maxwell-Poisson (COM-Poisson) dağılımı, ayrıca regresyon modeli, parametre tahmini, anlamlılık testi ve yorumlanma hakkında bilgi verilmiştir.

Bu çalışmadaki analizde, R programının “MASS” ve “COMpoissonReg” paketlerinden ve “glm” fonksiyonundan yararlanılarak modellerdeki parametre tahminleri ve uyum ölçütleri elde edilmiştir.

Uygulamada aşırı ve az yayılım gösteren iki veri seti kullanılmıştır. Aşırı yayılım gösteren 2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan hastaların, merkeze gelme sayılarına etki eden sosyo-demografik faktörlerin saptanması amaçlanmıştır. İkinci veri seti ise, az yayılım gösteren 2015-2016 yılları arasında Giresun Üniversitesi Bilimsel Araştırma Proje desteği ile (SOS-BAP-A-200515-43) gerçekleştirilen proje verilerinin bir kısmı olan öğretmenlerin sosyal medya hesap sayılarına etki eden okul türü, cinsiyet ve yaş grubu faktörleri incelenmiştir.

Bu amaçla birinci veri setine Poisson, negatif binom ve COM-Poisson, ikinci veri setinin az yayılım göstermesi nedeniyle sadece Poisson ve COM-Poisson regresyon modelleri uygulanmıştır. Modeli oluşturacak değişken seçimi için tüm olası alt kümeler yöntemi kullanılmıştır. En uygun model en küçük AIC değerini model olarak saptanmıştır.

2011-2014 yılları arasında Giresun ili Toplum Ruh Sağlığı Merkezine (TRSM) kayıtlı olan 244 hastanın cinsiyet, yaş, medeni durum, eğitim düzeyi, sosyal güvenceye sahip olma ve madde bağımlılık durumları incelenmiştir. Poisson regresyon modelinin tüm olası alt kümeler değişken seçim yöntemine göre cinsiyet, eğitim düzeyi, medeni durum (dul alt kategorisi hariç), sosyal güvencesi ile madde bağımlılık durumu (sigara ve alkol alt kategorisi hariç) istatistiksel olarak önemli bulunmuştur. AIC değeri 4743,4 olarak elde edilmiştir. Ayrıca $\phi = \frac{Sapma}{Sapma\ SD}$ değeri 1’den büyük olduğundan yayılım parametresinin aşırı yayılım gösterdiği görülmüştür. Bu nedenle negatif binom regresyon modeli uygulanmıştır. Negatif binom regresyon modelinde Poisson regresyon modelinden farklı olarak anlamlı bulunan madde bağımlılık durumu anlamsız bulunmuştur. AIC değeri 4743,4’ten 1618,2’e düşmüştür.

COM-Poisson regresyon modelleri uygulandığında v değeri 1'den küçük olduğu saptanmıştır. Dolayısıyla aşırı yayılım gösterdiği görülmüştür. Negatif binom regresyon model sonucuna benzer olarak cinsiyet, eğitim düzeyi, medeni durum ile sosyal güvence istatistiksel olarak önemli olduğu saptanmıştır. AIC değeri 1608,373 olarak elde edilmiştir. Üç regresyon modeli içerisinde en küçük AIC değerini COM-Poisson regresyon modelinden elde edilmiştir.

COM-Poisson regresyon modelinden elde edilen bağımsız değişken katsayıları, ilgili katsayı v 'ye (0,074) bölündükten sonra üsteli alınarak katsayılar yorumlanmıştır. Buna göre; kadınlar erkeklere göre yaklaşık olarak 1,5 kat daha fazla TRSM'ye geldiği saptanmıştır. Eğitim düzeyi bakımından karşılaştırıldığında referans kategori olan okur-yazar olmayanlara göre tüm eğitim düzeyi yükseldikçe TRSM'ye gelme oranında artış gözlenmiştir. Ancak Poisson regresyondan farklı olarak en fazla “Lise” (yaklaşık 6 kat), en az “İlkokul” (4 kat) mezunlarının daha sık TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak medeni durumu evli olanların bekarlara göre yaklaşık olarak yarı yarıya daha az TRSM'ye geldiği saptanmıştır. Poisson regresyondakine benzer olarak sosyal güvencesi olmayanların olanlara göre yaklaşık olarak 3 kat daha az TRSM'ye geldikleri görülmüştür.

Az yayılım gösteren 2015-2016 yılları arasında Giresun Üniversitesi Bilimsel Araştırma Proje desteği ile (SOS-BAP-A-200515-43) gerçekleştirilen proje verilerinin bir kısmı olan öğretmenlerin sosyal medya hesap sayılarına etki eden okul türü, cinsiyet ve yaş grubu faktörleri incelendiğinde, Poisson regresyonda sadece okul türünün anlamlı bulunduğu saptanmıştır. AIC değeri 680,08 olarak elde edilmiştir. $\phi = \frac{Sapma}{Sapma\ SD}$ değeri 1'den küçük olması nedeniyle az yayılım gösterdiği görülmüştür. Az yayılım durumunda negatif binom regresyon modeli hesaplanamamıştır.

COM-Poisson regresyon modelleri uygulandığında v değeri 1'den büyük olduğu saptanmıştır. Dolayısıyla az yayılım gösterdiği görülmüştür. Poisson regresyon modelinden farklı olarak yaş grubu da anlamlı bulunmuştur. AIC değeri 680,08'den

597,40'e düşmüştür. İki regresyon modeli içerisinde en küçük AIC değerini COM-Poisson regresyon modelinden elde edilmiştir.

COM-Poisson regresyon modelinden elde edilen bağımsız değişken katsayıları, ilgili katsayının v 'ye (3,075) bölünmesi sonrasında üsteli alınarak yorumlanmıştır. Buna göre; özel okulda görev yapan öğretmenlerin devlet okulunda görev yapan öğretmenlere göre yaklaşık 2 kat daha fazla sosyal medya hesabı bulunduğu saptanmıştır. Yaş grubu bakımından karşılaştırıldığında referans kategori olan 18-24'a göre tüm yaş gruplarının daha az sosyal medya hesabı bulunduğu saptanmıştır. Yaş grubu arttıkça sosyal medya hesap sayısı azalmaktadır. 40-44 yaş grubundakilerin 18-24 yaş grubuna göre sosyal medya hesap sayısı yaklaşık yarı yarıya düşmektedir. 25-29 yaş grubundakilerin 18-24 yaş grubuna göre sosyal medya hesap sayısı yaklaşık %30 azalmaktadır.

Bu tez çalışmasında son yıllarda hem aşırı hem de az yayılım durumları için sıkça uygulanan Conway-Maxwell-Poisson (COM-Poisson) regresyon modeli tanıtılmıştır. Bu regresyon modeli ile daha önce Türkiye'de tez çalışması yapılmaması nedeniyle literatüre katkı sağlama bakımından önemlidir.

KAYNAKLAR

- [1] Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright P., A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, 54(1),127-142, 2005.
- [2] Conway, R.W., Maxwell, W.L., A queuing model with state dependent service rates, *Journal of Industrial Engineering*, 12, 132-136, 1962.
- [3] Michener, R., Tighe, C., A Poisson regression model of highway fatalities, *American Economic Review Papers and Proceedings*, 82, 452-6, 1992.
- [4] Englin, J., Shonkwiler, J.S., Estimating social welfare using count data models: An application under conditions of endogenous stratification and truncation, *Review of Economics and Statistics*, 77, 104-112, 1995.
- [5] Van Der Burg, B., Siegers, J., Ebmer, R.W., Gender and promotion in theacademic labour market, *Labour*, 12(4), 701-713, 1998.
- [6] McKean, J.R., Taylor, R.G., Outdoor Recreation Use and Value: Snake River Basin of Central Idaho, *Idaho Experiment Station Bulletin*, Idaho, Moscow, 58, 2000.
- [7] Moons, E., Loomis, J., Proost, S., Eggermont, K., Hermy, M., Travel cost and time measurement in travel cost models, *Faculty of Economics and Applied Economic Science Center for Economic Studies Energy, Transport and Environment*, working paper series n°2001-22, 2001.
- [8] Sezgin, H.F., Deniz, E., Poisson regresyon modelinde aşırı yayılım durumu ve negatif binomial regresyon analizinin Türkiye grev sayıları üzerine bir uygulama, *Yönetim Dergisi*, 15(48), 17-25, 2004.
- [9] Yeşilyurt, H., Poisson Regresyon Modeli ve Türkiye'deki Boşanma İstatistiklerine Uygulanması, *Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*, 2005.
- [10] Bilgiç, A., Florkowski, J.W., Application of a hurdle negative binomial count data model to demand for bass fishing in the Southeastern United States, *Journal of Environmental Management*, 83, 478-490, 2007.

- [11] Mouatassim, Y., Ezzahid, H., Poisson regression and zero-inflated Poisson regression: Application to private health insurance data, *European Actuarial Journal*, 10.1007/s13385-012-0056-2, 2012.
- [12] Arı, A., Önder, H., Farklı veri yapılarında kullanılabilir regresyon yöntemleri, *Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Anadolu Tarım Bilim Dergisi*, 28(3), 168-174, 2012.
- [13] Sellers, K.F., Borle, S., Shmueli, G., The COM-Poisson model for count data: A survey of methods and applications, *Applied Stochastic Models in Business and Industry*, 28, 104-116, 2012.
- [14] Zhu, F., Modelling the series of counts with COM-Poisson INGARCH models, *Mathematical and Computer Modelling*, 56, 191-203, 2012.
- [15] Avşar, V., Türkiye'nin antidamping soruşturmalarını etkileyen faktörler: Sanayi verileri ile ekonometrik bir analiz, *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 9(1), 41-54, 2014.
- [16] Santarelli, M.F., Latta, D.D., Scipioni, M., Positano, V., Landini, L., A Conway-Maxwell-Poisson (CMP) model to address data dispersion on positron emission tomography, *Computers in Biology and Medicine*, 77, 90-101. 2016.
- [17] Avcı, E., Using count regression models to determine the factory which effects the hospitalization number of people with schizophrenia, *Journal of Data Science*, 16(3), 511-530, 2018.
- [18] Avcı, E., Flexibility of using COM-Poisson regression model for count data, *Statistics Optimization and Information Computing*, 6, 278-285, 2018.
- [19] Lewis, S.L., Montgomery, D.C., Myers, R.H., Examples of designed experiments with nonnormal responses, *Journal of Quality Technology*, 33, 265-278, 2001.
- [20] McCullagh, P., Nelder, J.A., Background an outline of generalized linear models. İçinde: *Generalized Linear Models*, 2nd ed., Chapman&Hall, London, 30-35, 1989.
- [21] Nelder, A.J., Wedderburn, R.W.M., Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384, 1972.
- [22] Aitken, M., Anderson, D., Francis, B., Hinde, J., *Statistical Modelling in GLIM*. Clarendon Press, Oxford, 1989.
- [23] Lindsey, J.K., *Applying Generalized Linear Model*, Springer Verlag, Newyork, 1997.

- [24] Uusipaikka, E., Confidence Intervals in Generalized Regressions Models. CRC Press, Boca Raton, 2000.
- [25] McCulloch, C.E., Searle, S.R., Generalized Linear and Mixed Effects. John Wiley&Sons, Inc., America, 2001.
- [26] Myers, R.H., Montgomery, D.C., Vining G.G., Generalized Linear Models with Applications in Engineering and Science, John Wiley&Sons INC., 2001.
- [27] Dobson, A.J., An Introduction to Generalized Linear Models (2nd ed), Chapman&Hall, London, 2002.
- [28] Duntelman, G.H., Moon-Ho, H.R., An Introduction to Generalized Linear Models. Sage Publication 145, USA, 2006.
- [29] Dunn, P.K., Smyth, G.K., Generalized Linear Models with Examples in R, Springer, New York, 2018.
- [30] Bickel, P.J., Doksum K.A., Mathematical Statistics: Basic Ideas and Selected Topics, 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, 2001.
- [31] Khuri, A., Linear Models Methodology, Chapman&Hall, USA, 483-485, 2010.
- [32] Benghia, S., Diagnostics for Generalized Linear Models, Concordia University, Master Thesis, Monreal, 34, 2001.
- [33] Korucu, Ö., Genelleştirilmiş Lineer Modellerde Model Seçimi Üzerine Alternatif Bir Yaklaşım. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yayımlanmamış Yüksek Lisans Tezi, 23-25, 2010.
- [34] Olsson, M., Generalized linear models: An applied approach, Student Literature, 46-47, 2002.
- [35] Collett, D., Strategy for model selection. İçinde: Modelling Binary Data, 2nd. ed., Chapman and Hall, London, 240-244, 1991.
- [36] Myers, R.H., Montgomery, D.C., Vining, G.G., Inference and residual analysis for generalized linear models. İçinde: Generalized Linear Models: With Application in Engineering and the Sciences, 2nd ed., John Willey&Sons, New York, 217-220, 2010.
- [37] Rao, C.R., Toutenburg, H., Linear Models: Least Squares and Alternatives, 2. Baskı, Springer, Germany, 1999.
- [38] Pierce, D.A., Schafer, D.W., Residuals in generalized linear models, Journal of American Statistical Association, 81(396), 977-986, 1986.

- [39] Deniz, Ö., Poisson regresyon analizi, İstanbul Ticaret Üniversitesi, Fen Bilimleri Dergisi, 7, 59-72, 2005.
- [40] Sümbüloğlu, K., Akdağ, B., İleri Biyoistatistiksel Yöntemler, Hatiboğlu Yayınları, Ankara, 2009.
- [41] <http://cameron.econ.ucdavis.edu/research/CTE01preprint.pdf>, Cameron, A.C., Trivedi, P.K., Essentials of Count Data Regression, 1999, Erişim Tarihi: 08.03.2019.
- [42] Wedderburn, R.W.M., Quasi-likelihood functions, generalized linear models and the Gauss-Newton method, Biometrika, 61(3), 439-447, 1974.
- [43] McCulloch, P., Quasi-likelihood functions, The Annals of Statistics, 11(1), 59-67, 1983.
- [44] Breslow, N., Biostatistics and bayes, Statistical Science, 5(3), 269-284, 1990.
- [45] Lawless, J.F., Negative binomial and mixed Poisson regression, The Canadian Journal of Statistics, 15(3), 209-225, 1987.
- [46] www.stat.psu.edu.pdf, The Negative Binomial Model, Stat 544, Lecture 17, Erişim Tarihi: 14.02.2019.
- [47] Saraçbaşı, T., Altunay, S.A., Kategorik Veri Çözümlemesi, Hacettepe Üniversitesi Yayınları, Ankara, 2016.
- [48] Winner, G., Altmann, G., Thesaurus of Univariate Discrete Probability Distributions, STAMM, Essen, 865, 1999.
- [49] Khan, N.M., Khan, M.H.M., Model for analyzing counts with over-,equi-and under-dispersion in actuarial statistics, Journal of Mathematics and Statistics, 6(2), 92-95, 2010.
- [50] Kadane, J.B., Shmueli, G., Minka, T.P., Borle, S., Boatwright, P., Conjugate Analysis of the Conway-Maxwell-Poisson distribution, Bayesian Analysis, 1, 363-374, 2005.
- [51] Sellers, K.F., Shmueli, G., A flexible regression model for count data, Annals of Applied Statistics, 4, 943-961, 2010.
- [52] McCullagh, P., Nelder, J.A., Generalized Linear Models, 2nd ed. Chapman&Hall/CRC, London, 1997.
- [53] Sellers, K.F., Shmueli, G., Data dispersion: Now you see it... Now you don't, Communication in Statistics: Theory and Methods, 42(17), 3134-3147, 2013.

- [54] Minka, T.P., Samueli, G., Kadane, J.B., Borle, S., Boatwright, P., Computing with the COM-Poisson distribution. Technical Report 776, Dept. Statistics, Carnegie Mellon Univ., Pittsburg, PA., 2003.
- [55] Long, J.S., Regression Models for Categorical and Limited Dependent Variables. Sage, London, 1997.
- [56] sagligim.gov.tr/ruh-sağlığı/liste/749-toplum-ruh-sağlığı-merkezleri-trsm-tanıyalım.html, Erişim Tarihi: 08.03.2019.



ÖZGEÇMİŞ

Bahar ÇELİK, 1989 yılında Giresun’da doğdu. İlk, orta ve lise eğitimini Giresun’da tamamladı. 2007 yılında Hamdi Bozbağ Anadolu Lisesi’nden mezun oldu. 2008 yılında başladığı Giresun Üniversitesi Matematik Bölümü’nü 2012 yılında bitirdi. 2016 yılında Giresun Üniversitesi Matematik Anabilim Dalında yüksek lisans eğitimine başladı. Halen 2016 yılında kayıt olduğu Giresun Üniversitesi Sağlık Bilimleri Fakültesi Hemşirelik Bölümü’nde okumaktadır.