

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**POISSON REGRESYON ANALİZİNDE LOG-DOĞRUSAL
MODELLERİN KULLANILMASI**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Eylem ŞİMŞEK
DANIŞMAN: Doç. Dr. Abdullah YEŞİLOVA

VAN-2017

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**POISSON REGRESYON ANALİZİNDE LOG-DOĞRUSAL
MODELLERİN KULLANILMASI**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Eylem ŞİMŞEK

Bu çalışma VAN YYÜ Bilimsel Araştırma Projeleri Başkanlığı tarafından **FYL-2017-5852** nolu proje olarak desteklenmiştir.

VAN-2017

KABUL VE ONAY SAYFASI

Zootekni Anabilim Dalı'nda Doç. Dr. Abdullah YEŞİLOVA danışmanlığında, Eylem ŞİMŞEK tarafından sunulan “Poisson regresyon analizinde log-doğrusal modellerin kullanılması” isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince... / ... / ... tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan: Doç. Dr. Abdullah YEŞİLOVA

İmza:

Üye: Yrd. Doç. Dr. Yılmaz KAYA

İmza:

Üye: Yrd. Doç. Dr. Hasan KOYUN

İmza:

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun / / tarih ve sayılı kararı ile onaylanmıştır.

Enstitü Müdürü

Prof. Dr. Suat ŞENSOY

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

İmza

Eylem ŞİMŞEK

ÖZET

POISSON REGREYON ANALİZİNDE LOG-DOĞRUSAL MODELLERİN KULLANILMASI

ŞİMŞEK, Eylem

Yüksek Lisans Tezi, Zootekni Anabilim Dalı

Tez Danışmanı: Doç. Dr. Abdullah YEŞİLOVA

Ağustos 2017, 50 sayfa

Poisson regresyonunda bağımlı değişken Poisson dağılımı gösterir. Poisson regresyonu sayıma dayalı olarak elde edilen bağımlı değişkenin modellenmesinde sıklıkla kullanılmaktadır. Bununla birlikte, bağımlı değişkenin sayıma dayalı olarak elde edildiğinde, bağımsız değişken kategorileri için relatif risk değerini de hesaplamaktadır. Bu çalışmada, YGS’de 180 puan altı alan öğrenciler ve 180 puan ve üstü alan ancak herhangi bir programa yerleşemeyen öğrencilerin relatif riskleri bölgelere ve yıllara göre tahmin edildi. YGS’de, öğrencilerin 180 altında puan alması bakımından; Güneydoğu Anadolu’daki öğrencilerin relatif riski iç Anadolu bölgesindeki öğrencilerin 1.499 katıdır ($p<0.01$). YGS’de, öğrencilerin 180 ve üstü puan alıp fakat herhangi bir programa yerleşmemesi bakımından; Güneydoğu Anadolu’daki öğrencilerin relatif riski Marmara bölgesindeki öğrencilerin 1.316 katı olmuştur ($p<0.01$). Bu çalışmada, iki yönlü tabloda kullanılan değişkenler yıllar ve bölgeler olmuştur. Sonuç olarak Poisson log-doğrusal regresyonu, iki yönlü çapraz sınıflandırılmış tabloları analizinde oldukça etkili bir yöntemdir.

Anahtar kelimeler: En çok olabilirlik yöntemi, Poisson log-doğrusal regresyonu, Poisson regresyonu, Relatif risk.



ABSTRACT

USING LOG-LINEAR MODELS IN THE ANALYSIS OF POISSON REGRESSION

ŞİMŞEK, Eylem

M.Sc.Thesis, Animal Science

Supervisor: Assoc. Prof. Dr. Abdullah YEŞİLOVA

August 2017, 50 pages

In Poisson regression, dependent variable based on count data display a Poisson distribution. Poisson regression is frequently used for analyzing based on count data. However, in the Poisson regression, the dependent variable is a relative risk can be estimated as well as count data. Dependent variable is very appropriate for modeling the occasion of rare events. In the this study, relative risk of students who score under 180 and relative risk of students who score 180 and above in the YGS, but do not fit in any program were estimated according to independent variables such as regions and years. Year 2009 and South Eastern Anatolia region were taken as a referances category. The risk of getting the South Eastern Anatolia region was 1.499 times that of the risk of getting the Central Anatolia region for the students who score under 180 in the YGS ($p<0.01$). The risk of getting the South Eastern Anatolia region was 1.316 times that of the risk of getting the sea of Marmara for the students who fit any program in the YGS ($p<0.01$). Two-way were location and varieties, respectively in this study. As a result, Poisson log-linear regression model is a very effective method for analysis of two-way contingency table.

Key words: Maximum likelihood method, Poisson log-linear regression, Poisson regression, Relative risk.



ÖN SÖZ

Tezimin her aşamasında, benimle deneyimlerini ve bilgisini paylaşan, tez konumun belirlenmesine katkı sunan danışman hocam sayın Doç. Dr. Abdullah YEŞİLOVA'ya sonsuz teşekkürlerimi sunarım.

Yüksek lisans tezimin her aşamasında beni yalnız bırakmayan, gösterdiği sabır ve verdiği maddi ve manevi destek için başta babam Murat ŞİMŞEK 'e ve aileme en içten duygularıyla teşekkürlerimi sunarım. Ayrıca, yüksek lisans tezime maddi destekte bulunan Yüzüncü Yıl Üniversitesi Bilimsel Araştırma Proje Koordinatörlüğüne teşekkür ederim.

2017

Eylem ŞİMŞEK



İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	iii
ÖNSÖZ	v
İÇİNDEKİLER	vii
ÇİZELGELER LİSTESİ.....	ix
ŞEKİLLER LİSTESİ.....	xi
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ	1
2. KAYNAK BİLDİRİŞLERİ	4
3. MATERYAL VE YÖNTEM	10
3.1. Materyal	10
3.2. Yöntem	10
3.2.1. Genelleştirilmiş doğrusal modeller	10
3.2.2. Genelleştirilmiş doğrusal modellerde bağlantı fonksiyonu tanımı.....	12
3.2.3. Log- doğrusal modellerin genel yapısı	12
3.2.4. Poisson regresyonu	17
3.2.5. Poisson regresyonunda parametre tahmini	18
3.2.6. Poisson log linear Regresyon analizi	18
3.2.7. Model uyumu.....	20
3.2.8. En çok olabilirlik tahminleri için hipotez testi.....	21
3.2.9. Relatif risk.....	21
3.2.10. Pseudo - R^2 belirleme katsayısı.....	22
4. BULGULAR	24
4.1. Tanıtıcı İstatistikler	24
4.2. Poisson Log-Doğrusal Regresyon Analizi Sonuçları	29
4.2.1. YGS sınavında 180 altında puan alan öğrenciler için Poisson log doğrusal regresyon modeli sonuçları	29
4.2.2 Herhangi bir lisans programına yerleşemeyen öğrenciler için Poisson log doğrusal regresyon modeli sonuçları	34
4. TARTIŞMA VE SONUÇ.....	44

KAYNAKLAR.....	47
ÖZ GEÇMİŞ.....	51



ÇİZELGELER LİSTESİ

Çizelge

Sayfa

Çizelge 4.1.1. 2010 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler	25
Çizelge 4.1.2. 2011 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler.....	26
Çizelge 4.1.3. 2012 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler.....	27
Çizelge 4.1.4. 2013 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler	28
Çizelge 4.2.1. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı.....	30
Çizelge 4.2.2. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrenciler için en çok olabilirlik parametre tahminler.....	33
Çizelge 4.2.3. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrenciler için Log-oran testi sonuçları.....	34
Çizelge 4.2.2.1. Yıllara ve bölgelere göre herhangi bir lisans programına yerleşemeyen öğrenciler.....	35
Çizelge 4.2.2.2. Yıllara ve bölgelere göre YGS sınavında lisans programına yerleşemeyen öğrenciler için en çok olabilirlik parametre tahminleri.....	39
Çizelge 4.2.2.3. Yıllara ve bölgelere göre YGS sınavında yerleşemeyen öğrenciler için log-oran testi sonuçları.....	40
Çizelge 4.2.2.4. Yıllara ve bölgelere göre YGS sınavında 180 altı puan alan öğrenciler ile herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekanslar.....	41



ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 4.2.1. Yıllara göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı.....	30
Şekil 4.2.2. Bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı.....	31
Şekil 4.2.3. Yıllara göre YGS sınavında herhangi bir programa yerleşemeyen öğrencilerin dağılımı.....	36
Şekil 4.2.4. Bölgelere göre YGS sınavında herhangi bir programa yerleşemeyen öğrencilerin dağılımı.....	36
Şekil 4.2.5. YGS sınavında 180 altı puan alan öğrenciler için gözlenen ve beklenen frekansların uyumu.....	42
Şekil 4.2.6. YGS sınavında 180 üstü puan alıp ancak herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekansların uyumu.....	43



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklama

α

Aşırı Yayılım Parametresi

β

Regresyon Katsayıları

μ

Beklenen Değer

ω

Sıfır Değer Ağırlıklı Parametre

ϕ

Aşırı yayılım Parametresi

Kısaltmalar

Açıklama

YGS

Yüksek Öğretime geçiş Sınavı

LYS

Lisan Yerleştirme Sınavı

TÜİK

Türkiye İstatistik Kurumu

PR

Poisson Regresyon

ÖSYM

Öğrenci Seçme Yerleştirme
Merkezi

GLM

Genelleştirilmiş Doğrusal
Modellerde

1. GİRİŞ

Sağlık ve davranış bilimleri gibi birçok bilim dalında yapılan çalışmalardan elde edilen veriler genellikle sayıma dayalı olarak elde edilmektedir. Bu tip kategorik veriler normal dağılım göstermediklerinden dolayı parametrik testler kullanılamaz. Bu bağlamda, çapraz olarak sınıflandırılmış (contingency tables) sayıma dayalı olarak elde edilen değişkenler arasındaki ilişkiyi ya da bağımsızlığı incelemek için birçok yöntem geliştirilmiştir. Bu açıdan bakıldığında, çalışmalarda doğru sonuçların edilebilmesi için kullanılacak istatistiksel yöntemler oldukça önem taşımaktadır. Bu nedenle, değişkenlerin elde ediliş biçimlerine göre farklı istatistiksel yöntemler kullanılmaktadır. Sayıma dayalı olarak elde edilen, binary, nominal ve ordinal gibi kategorik değişkenler olarak tanımlanan değişkenler için parametrik olmayan testlerin kullanılması daha uygundur. Bazı durumlarda, sürekli değişkenlerde belirlenen aralıklarla gruplara ayrılarak kategorik değişkene dönüştürülebilir (Upton, 2017).

Kategorik değişkenler arasındaki ilişkiyi incelemede iki yönlü çapraz sınıflandırılmış tablolar sıklıkla kullanılmaktadır. Bu tip düzenlenen tabloların bağımsızlık analizinde genellikle Pearson ki-kare Testi kullanılmaktadır. Ki-kare testi 2X2'lik çapraz sınıflandırılmış tablolar için yapılan hipotez testlerinde yeterlidir. Ancak daha büyük boyutlu (2X2'den fazla) çapraz sınıflandırılmış tablolarda ilişkinin kaynaklarını belirlemede yetersiz kalmaktadır. Çünkü böyle durumlarda ki-kare testi ile sadece değişkenler arasında bir ilişki olup olmadığı belirlenmektedir. Bu nedenler üç ya da daha çok boyutlu çapraz sınıflandırılmış tablolardaki ilişkileri incelemede uyum analizi (correspondance analysis) ve log-doğrusal modeller kullanılmaktadır. Uyum analizinde, iki ya da daha çok değişkene sahip çapraz tablolarda her bir değişkenin kategorileri arasındaki uyum grafiksel olarak yorumlanır. Bu nedenle, uyum analizinde anlamlılık testi yoktur. Log-doğrusal modeller çapraz sınıflandırılmış tabloları oluşturan değişkenler arasında bağımlı ve bağımsız değişken ayırımı yapmadan genel yapısal ilişkiyi incelemektedir. Bununla beraber, eğer bağımlı ve bağımsız değişken ayırımı yapılarak analiz yapılmak istenirse, log-doğrusal modeller Poisson regresyon modeline dönüşür (SAS, 2017).

Log-doğrusal modeller, bağımlı değişken binary (binom dağılımı gösterdiğinde) olduğunda, lojit modele dönüşür. Bununla birlikte, bağımlı değişken Poisson dağılımı gösterdiğinde Poisson log-doğrusal regresyona dönüşmektedir (Eye ve Mun, 2013).

Regresyon analizinde, veri setinin yapısına uygun regresyon modellerin uygulanması elde edilen sonuçları tutarlılığı açısından oldukça önemlidir. Regresyon modelleri, bağımlı değişkenin göstermiş olduğu dağılıma göre değişmektedir. Genel olarak regresyon modelleri doğrusal (linear) ve doğrusal olmayan (nonlinear) olarak iki başlık altında toplanmaktadır. Örneğin; bağımlı sürekli bir yapıda olup normal dağılımı gösterdiğinde doğrusal regresyon (linear regression), bağımlı ikili (binary) bir yapıda olup binom dağılımı gösterdiğinde lojistik regresyon (logistic regression) ve bağımlı kesikli bir yapıda (sayılarak elde edildiğinde) olup Poisson dağılımı gösterdiğinde Poisson regresyon yaygın olarak kullanılan yöntemlerdir. Poisson regresyon analizi, bağımsız değişken(ler) ile sayıya dayalı olarak elde edilen bağımlı değişken arasındaki ilişkiyi açıklayan bir yöntemdir. Poisson regresyonunda bağlantı fonksiyonu logaritmik dönüşüm ile verilmektedir.

Genelleştirilmiş doğrusal modellerde, bağımlı değişkenin göstermiş olduğu dağılıma uygun bir link fonksiyonu kullanılarak regresyon analizi uygulanmaktadır. Poisson regresyonu için log link fonksiyonu kullanılmaktadır. Bu tip doğrusal olmayan regresyon yöntemlerinde normalite varsayımı aranmadığından dolayı, veri setinin orijinal yapısı üzerinden her hangi bir işlem yapmaksızın olduğu gibi analize edilmektedir. Bu durumda, tüm gözlemlerin olduğu gibi kullanılması, elde edilen sonuçların tutarlılığı açısından oldukça önemlidir. Poisson regresyonunda, bağımlı değişkenin sayıya dayalı olarak elde edildiğinde kullanılacağı gibi, risk oranları (ratio risk) ya da nispi risk (relative risk) olduğu durumlarda da kullanılmaktadır. Örneğin farklı ilaçların farklı zamanlarda uygulandığı hastalardan, iyileşme oranları bakımından ilaçların ve zamanların etkinliklerinin karşılaştırılması durumunda log-doğrusal Poisson regresyonu kullanılabilir. İki yönlü çapraz sınıflandırılmış tablolarda kategori sayısı ikiden fazla olduğunda ki-kare testi sonucunda bir bağımlılık sonucuna varıldığında bağımlılığın hangi kategori seviyesinden kaynaklandığı konusunda bilgi vermemektedir. Özellikle kategori sayısı ikiden fazla olan iki yönlü çapraz sınıflandırılmış tabloların analizinde log-doğrusal Poisson modeller daha açıklayıcı bilgiler vermektedir. Poisson log doğrusal regresyon modeller hem iki yönlü çapraz

sınıflandırılmış tabloların analizinde hem de regresyon analizinde kullanıldığından dolayı, kullanıcılara daha açıklayıcı bilgiler sunmaktadır (Stata, 2012).

Bu çalışmada, Türkiye’de 81 ilde 2010-2013 yılları arasında ÖSYM’nin düzenlediği Yüksek Öğrenime Giriş Sınavına (YGS) giren; 180 puan altı alan öğrenciler ile 180 ve üzeri bir puan alıp ancak herhangi bir programa yerleşmeyen öğrenciler bazında bölgelere ve yıllara göre relatif risklerin tahmin edilmesi amaçlanmıştır. Bu amaçla, önce Poisson log-doğrusal regresyon modeli ile ilgili teorik bilgiler verilmiştir; sonrasında ise, Poisson log-doğrusal regresyon modeline ilişkin relatif risk tahminleri yapılmıştır.



2. KAYNAK BİLDİRİŞLERİ

Log-doğrusal modeller kategorik değişkenler arasındaki ilişkiyi incelemektedir. Bununla birlikte, çapraz sınıflandırılmış çok yönlü tablolarda karmaşık ilişkileri de ortaya koymaktadır. Log- doğrusal modelleme sınıflama amaçlı kullanıldığında, sınıflandırma amaçlı kullanılan bütün değişkenler bağımsız değişkenler, çapraz tablolardaki hücrelerde yer alan gözlemler ise bağımlı değişkeni göstermektedir. Bu yönü ile log-doğrusal modeller çoklu regresyon analizine benzetilmektedir. Log-doğrusal modeller, kategorik değişkenler arasındaki ilişkiyi, bazı durumlarda yardımcı bir değişken kullanarak araştırmaktadır. Örneğin, kişilerin eğitim düzeyleri esas alınarak bir kursun değişik açılardan etkilerinin incelenmesi (Evers ve Namboediri, 1978; Eye ve Mair, 2008).

Log- doğrusal modellemeyi, lojistik regresyon ayıran temel özellik yalnızca kategorik verileri modellemek için kullanılmasıdır. Bununla birlikte Log-doğrusal modelleme bağımlı ve bağımsız değişken arasında herhangi bir ayırım yapmamaktadır. Log-doğrusal modellemede, örneklem hacminin yeterli büyüklükte olması gerekmektedir. Genellikle, incelenen denek sayısının, tablodaki hücre sayısının beş katı olması istenmektedir (Krauth, 2005; Eye ve Mun, 2013).

Log-doğrusal modeller hiyerarşik ve hiyerarşik olmayan olmak üzere iki başlık altında incelenmektedir. Hiyerarşik log-doğrusal modellemede, modeldeki en yüksek dereceli parametreyi oluşturan değişkene bakılır ve daha sonra bu değişkenlere ait daha düşük dereceli parametreler de modele alınır. Hiyerarşik olmayan log-doğrusal modellemede böyle bir zorunluluk yoktur (Mieser, 1996; Eye ve Schuster, 2000; SAS, 2017).

Hiyerarşik log-doğrusal modelleme, Log-doğrusal modellemenin genelleştirilmiş hali olup, ana etkileri yanında ayrıca interaksiyon etkilerini de içermektedir. İki yönlü çapraz sınıflandırılmış tablolarda satır ve sütun etkileri ana etkiler, bunlar arasındaki etkileşimler ise interaksiyon etki olarak adlandırılmaktadır (Stata, 2012).

Uygun modelin seçiminde, belirli bir derece teriminin modele olan katkısı sistematik olarak test edilebilir. Önce interaksiyon teriminin modele uyumu daha sonra

sadece ana etkilerinin modele uyumuna bakılmaktadır. Verilerin modele ne kadar uygun olduğunun belirlemenin diğer bir yolu ise her bir hücredeki gözlenen ve beklenen değerler arasındaki farklar olarak adlandırılan artık değerlerin (residuals) incelenmesidir. Artık değerlerin oldukça küçük ve grafiklerinin belirli bir biçim göstermemeleri gerekmektedir. Eğer standartlaştırılmış artık değerler, normal dağılım gösteriyorsa artıkların grafiği normal ya da yaklaşık normal dağılım göstermelidir (Glück ve Von, 2000; Eye ve Gardiner, 2004; Galindo ve Vermunt, 2004).

Risk faktörü, bireylerin herhangi bir olaya (hastalık ya da okula gidememe durumu gibi) maruz kalma durumunu artırabilen, ancak söz konusu olayın çıkmasına doğrudan neden olmamaktadır. Birçok değişik risk ölçüleri söz konusudur. Bunların içinden oransal ya da relatif risk (relative risk); söz konusu risk faktörüne maruz kalanların kalmayanlara göre olaya yakalanma açısından ne kadar (kaç kat) daha fazla risk altında olduklarını tahmin etmede kullanılmaktadır. Relatif risk, hem ileriye yönelik (prospektif) hem de geriye dönük (retrospektif) araştırmalarda kullanılmaktadır (SAS, 2017).

Bağımlı değişkenin Poisson dağılımı gösterdiği durumda genelleştirilmiş doğrusal modelleri esas alan Poisson regresyonu kullanılmaktadır. Genelleştirilmiş doğrusal modellerde (GLM), veri setinin orijinal yapısı üzerinde her hangi bir varsayıma gerek duyulmadan, üssel formda tanımlanmaktadır. Bağımlı değişkenin göstermiş olduğu dağılışa, uygun bir link fonksiyonu kullanılarak regresyon analizi uygulanmaktadır. Poisson regresyonunda, bağımlı değişkenin sayıya dayalı olarak elde edildiğinde kullanılacağı gibi risk oranları (ratio risk) ya da nispi risk (relative risk) olduğu durumlarda da kullanılmaktadır. İki yönlü çapraz sınıflandırılmış tablolarda (contingency tables) kategori sayısı ikiden fazla olduğu durumlarda ki-kare testiyle bir bağımlılık sonucuna varıldığında, bağımlılığın hangi kategori seviyesinden kaynaklandığı konusunda bilgi vermemektedir. Özellikle kategori sayısı ikiden fazla olan iki yönlü çapraz sınıflandırılmış tabloların analizinde log-doğrusal Poisson modeller daha açıklayıcı bilgiler vermektedir. Ayrıca Poisson log doğrusal modeller ile regresyon analizi yapılarak diğer yöntemlere nazaran daha çok bilgi elde edilmektedir. Bu çalışmada amaç, bağımlı değişkenin oran (proportion) biçiminde elde edildiğinde, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi açıklamada log-doğrusal Poisson regresyonun kullanılmasıdır. (Eye ve Mun, 2013).

Sayıma dayalı olarak elde edilen veriler genellikle normal dağılım göstermezler. Bu tip veriler çok geniş bir aralıkta değiştiğinden dolayı, dağılımın sağa doğru çarpık olmaktadır. Bu nedenle bu tür verilere uygulanan t-testi ve varyans analizi gibi teknikleri sapmalı parametre tahminleri ve standart hataların elde edilmesine neden olmaktadır (Stokes ve ark., 2000; SAS, 2017).

Genelleştirilmiş doğrusal modellerde, veri setinin orijinal yapısı üzerinde her hangi bir varsayıma gerek duyulmadan, üssel forma tanımlanmaktadır (Nelder ve Wedderburn, 1972; McCullagh ve Nelder, 1989; Dobson, 2002; Luo ve Qu, 2015).

Doğrusal modellerde, gözlem değerlerinin ve bunlara ait hataların normal dağılışa sahip olduğu varsayılmaktadır. Bazı durumlarda, hata terimi binomial, Poisson ve gamma dağılışları gösterebilir. Böyle durumlarda, normallik varsayımını sağlamak için bazı dönüşümler kullanılmaktadır. Ancak dönüşüm her zaman varsayımları sağlayamamaktadır. Böyle durumlarda genelleştirilmiş doğrusal modeller kullanılmaktadır. Genelleştirilmiş doğrusal modellerde, verilerin dağılışına uygun bir bağlantı (link) fonksiyonu tanımlanarak dağılış üstel formda yazılmaktadır (Nelder ve Wedderburn, 1972; McCullagh ve Nelder, 1989; Long ve Freese , 2006).

Log-doğrusal modelleme kavramı, ilk kez 1969 yılında Bishop ve Fienberg tarafından kullanılmıştır. İlk kullanılışından bu yana log-doğrusal modeller ile ilişkili birçok yöntem geliştirilerek frekans verilerinin analizinde yoğun olarak kullanılmaya başlandı. Özellikle bu geliştirilen log-doğrusal modeller çok değişkenli frekans tabloların analizinde (çok değişkenli çapraz sınıflandırılmış tablolar) yoğun olarak kullanılmıştır. Log-doğrusal modelleme çapraz sınıflandırmada ortak dağılımı tanımlamak için gerekli bulunan ana etkileri ve interaksiyonları tanımlamak için geliştirilmiştir. İnteraksiyon etkileri önemli bulunmasıyla beraber, ana etkilere olan ilgide başlamıştır. Çok değişkenli çapraz sınıflandırılmış tablolarda, ana etkiler ve interaksiyonlara ek olarak kovariyetler de incelenmeye başlanmıştır.

Log-doğrusal modelleme özellikle kategorik olarak tanımlanan verilerin analizinde kullanılmaktadır. Çapraz olarak sınıflandırılmış kategorik veriler değişik formatta olabilirler. Veri yapısı; sürekli, sınırlandırılmış sürekli (sensing ve truncated), kategorik, nominal kategorik ve ordinal kategorik yapıları içermektedir. Bu tip verilerin kullanıldığı çapraz sınıflandırılmış tablolarda, çok değişkenli formatta olabilir. Bu

tanımlanan çok değişkenli çapraz sınıflandırılmış tablolarda, ana etkilerin, interaksiyon etkilerin ve kovarietlerin test edilmesi oldukça zaman alabilmektedir (Stata, 2012).

Goodman'ın (1979) log-doğrusal modellemede üç yaklaşımı bulunmaktadır. Bunlar;

1. İki değişkenli çapraz sınıflandırılmış tablolarda ortak dağılımın incelenmesi
2. Bağımlı değişkeninin yapısının incelenmesi
3. İki bağımlı değişken arasındaki ilişkinin araştırılması

Değişken sayısı ikiden fazla olduğu durumlarda, Goodman'ın tanımladığı log-doğrusal modellemede her üç yaklaşım değişken sayısının durumuna göre güncellenebilir. Değişken sayısının ikiden fazla olması durumu ilk olarak Hand ve Viciotti (2003), daha sonra Havranek ve Lienert (1984) lokal modelleri incelemişlerdir. Lokal modeller tablonun tamamı yerine, yalnızca bir kısmını içermektedir (SAS, 2017).

Log-doğrusal modellemede, kullanılan formülasyon bazı avantajlara sahiptir. Bunlardan ilki, genelleştirilmiş doğrusal modellerde kullanılacak dağılımlar için basit tanımlamaların yapılmasını sağlamaktadır. İkincisi, log-doğrusal modellerin kullandığı formlarının varyans analizi ile paralellik göstermesidir. Üçüncüsü, log-doğrusal modeller çoğunlukla eklemeli terimleri içermektedir. Dördüncü avantajı ise odds oranları ile olan ilişkisini kolayca açıklayabilmesidir (Bishop ve ark., 1975; SAS, 2017).

Değişkenlerin ortak dağılımının modellenmesine gereksinim duyulduğunda, analiz sonuçları ortak frekans dağılımının tanımlanmasıyla elde edilmektedir. Örneğin iki yönlü çapraz sınıflandırılmış değişkenler tablonun ana diyagonaluna göre simetrik olabilir. Örneğin, IXI boyutlu bir kare çapraz sınıflandırılmış tablosu ele alındığında, her bir hücredeki olasılık değeri π_{ij} (i ve j=1,2,...,I) bu çapraz sınıflandırmada, $\pi_{ij} = \pi_{ji}$ olduğunda eksensel simetrik (axial symmetric) olur. Bununla birlikte, ortak dağılımların birçok formları incelenmiştir (DeCoster ve ark., 2009; Stata, 2012).

Değişkenler arasındaki bağımlılık ilişkileri modellendiğinde, lojit modellerden elde edilen koşullu dağılımları, odds oranları ya da regresyon parametreleri kullanılmaktadır. İki yönlü çapraz sınıflandırılmış verilerin analizinde ki-kare testi yaygın olarak kullanılmaktadır. Ki-kare istatistiği yaklaşık ki-kare dağılım gösteren bir istatistiktir. Ki-kare testinde 3 temel varsayım vardır. Bunlar sırasıyla; her bir hücredeki

gözlemlerin birbirinden bağımsız olması, benzer dağılım göstermeleri ve yeterli sayıda gözlem olması şeklinde yazılabilir (Gardiner, 1999; Beh ve Farver, 2009).

Bağımlı değişken sayıma dayalı olarak (negatif olmayan gerçek 0,1 gibi gerçek değer) elde edildiğinde, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi tanımlamada, normal dağılımı esas alan doğrusal modeller uygun olmamaktadır. Bununla birlikte, sayıma dayalı olarak elde edilen bağımlı değişken Poisson dağılımı gösterdiğinden dolayı, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi açıklamada Poisson regresyonun kullanılması daha uygundur (Frome ve ark., 1973; Cameron ve Trivedi, 1998; Lawal, 2003; Everitt ve Hothorn, 2006, Luo ve Qu, 2013).

Poisson regresyon modeli sayıma dayalı olarak elde edilen çalışmalara yaygın bir şekilde uygulanmaktadır (Box ve ark., 2005).

Cameron ve ark. (1998) yaptıkları çalışmada, hastane servislerine bireylerin kaç defa geldiklerini ve bunu etkileyen kişilerin sağlık durumları ve sigortalı olup olmadıklarını gösteren bağımsız değişkenlerini esas alan veri setini modellemek için PR modelini uygulamışlardır.

Lerner ve ark. (2010)'da yaptığı çalışmada, winkelman üzerinde belirli aralıklarda, uçaklar tarafından yaşanan kazaların sayısını ve bu kazalara etki eden değişkenleri modellemek için PR modelini uygulamışlardır.

Loken (2010) tarafından yapılan çalışmada, Poisson regresyon modelini, bilinen bir hastalığa maruz kalan ve kalmayan bireylerin epidemiyoloji oluşumunu araştırmayı içeren biyomedikal çalışmalara uygulamasını yapmışlardır.

Eye ve Spiel (1996), erkek İsveç veterinerler ile diğer veteriner endüstrisinde çalışanlar arasındaki kanser oluş sayılarını modellemek için Poisson regresyon modelini kullanmışlardır.

Poisson regresyon modeli, bağımlı değişkenin ortalama ile varyansının eşit olduğunu varsaymaktadır. Ancak pratikte bu uygulamayı sağlamak her zaman mümkün olmamaktadır ve çoğunlukla sayıma dayalı olarak elde edilen bağımlı değişkenin ortalaması varyansından küçük olmaktadır. Poisson regresyon modelinde bu durum aşırı yayılım olarak adlandırılmaktadır. Bağımlı değişkende aşırı yayılım ya da az yayılım olduğunda PR modeli yerine genelleştirilmiş Poisson regresyonu kullanılmaktadır (Cox, 1983; Lawles, 1987; Dean, 1992; Christensen, 1997; Hilbe, 2007; Wang ve ark., 2014; Brian, 2017).

Genelleştirilmiş Poisson regresyon modeli ilk olarak Consul ve Famoye (1992) tarafından tanıtılmıştır. Genelleştirilmiş Poisson regresyon modeli genelleştirilmiş Poisson dağılımı gösteren bağımlı değişkenin modellenmesinde kullanılmaktadır. Bu modelin; hane veri setine (Wang ve Famoye, 1997), yaralama çalışmalarına (Wulu ve ark., 2002) ve özellikle aşırı yayılım gösteren veri setlerine (Breslow, 1996; Yeşilova ve ark., 2007; Yeşilova ve ark., 2010; Yeşilova ve Denizhan, 2016) uygulaması yapılmıştır.

Ki-kare test istatistiği ile log-doğrusal ana etkileri modeline benzemektedir. IXJ boyutlu bir tablo için ki-kare test istatistiği aşağıdaki gibi yazılabilir,

$$\chi^2 = \sum_i \sum_j \frac{(m_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

Bu eşitlikte, $\hat{m}_{ij} = \frac{m_{i.}m_{.j}}{N}$ ya da $\hat{m}_{ij} = \pi_{i.}\pi_{.j}N$ biçiminde yazılabilir.

$\hat{m}_{ij} = \pi_{i.}\pi_{.j}N$ eşitliğinin logaritması alındığında,

$$\begin{aligned} \log(\hat{m}) &= \log N + \log \pi_{i.} + \log \pi_{.j} \\ &= \lambda + \lambda_i^A + \lambda_j^B \end{aligned}$$

ki-kare istatistiği log-doğrusal ana etkileri modeline dönüşür. Başka bir değişle ki-kare test istatistiği ana etkileri modeli olarak tanımlanabilir. Bu model marjinal olasılıkları esas almaktadır (Eye ve Mum, 2013).

Bu çalışmada, çapraz sınıflandırılmış tablolarda log-doğrusal modelleme yaklaşımının kullanılması amaçlanmıştır. Birçok çalışmada, bu tip verilerin analizi için odds ratio analizi, genelleştirilmiş doğrusal modelleri ve varyans analiz tekniğini kullanmışlardır. Bu çalışmada ise bu yöntemlerin aksine, log-doğrusal modelleme için desen matrisi (design matrix) yaklaşımı kullanılmıştır. Desen matrisinde vektörler tarafından tanımlanan log-doğrusal modelleme için hipotez etkileri incelenmiştir (Krauth, 2005).

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu çalışmada, Türkiye’de 81 ilde 2010-2013 yılları arasında ÖSYM’nin düzenlediği Yüksek Öğrenime Giriş Sınavına (YGS) giren; 180 ve üzeri puan alarak ikinci basamak sınavı olan Lisans Yerleştirme Sınavına (LYS) girmeye hak kazanan öğrencilerden Lisans, Açık Öğretim Fakültesi ve Meslek Yüksek Okuluna yerleşen öğrencilerin ve hiçbir bölüme yerleşemeyen öğrencilere ilişkin bilgiler TUIK (Türkiye İstatistik Kurumu)’ten ve ÖSYM arşivinden değişkenlerle ilgili veriler toplanmıştır.

ÖSYM sınav kılavuzu incelenerek değerlendirmelerin yapılmasında bu kılavuzdaki esaslar baz alınarak incelemeler yapılmıştır. TUIK ve ÖSYM arşivinden alınan verilerle oluşturulan değişkenler şu şekilde belirlenmiştir:

- 1) YGS’ ye giren öğrenci sayısı
- 2) YGS’ de 180 altında puan alan öğrenci sayısı
- 3) LYS sonucu Lisansa yerleşemeyen öğrenci sayısı

3.2. Yöntem

3.2.1. Genelleştirilmiş doğrusal modeller

Genelleştirmiş doğrusal modellerde, bağımlı değişkenin normal dağılışı göstermesi şart değildir. Veri seti üssel formda tanımlanarak veri yapısına uygun bir şekilde link fonksiyonu tanımlanmaktadır. Böylece, orijinal verilerin yerine, onların doğrusal kombinasyonlarının beklenen değerleri kullanılmaktadır. Veri setinin yapısına uygun bir link fonksiyonu kullanılarak söz konusu doğrusal olamayan model doğrusallaştırılmaktadır.

Genelleştirilmiş doğrusal modellerin genel formu aşağıdaki gibi yazılabilir.

$$f(\theta, \phi; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

Bu eşitlikte,

θ doğal parametreyi (natural parameter) ve ϕ ise skala parametresini (scale parameter) göstermektedir. Eşitlikte, a,b ve c dağılım şeklini gösteren parametrelerdir. $Y_1, Y_2, \dots, Y_n$ üssel dağılım familyasındaki bir dağılım gösteren rasgele örneklem olsun. Bu durumda, bu dağılıma ait olabilirlik fonksiyonu :

$$l(\theta_i, \phi_i; y_i) = \frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \quad (1)$$

biçiminde yazılabilir. Bu durumda $y_1, y_2, \dots, y_n$ için ortak olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$l(\theta, \phi; y_1, y_2, \dots, y_n) = \sum \left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right]$$

Aynı zamanda ortak olabilirlik fonksiyonu matris formatında,

$$l(\theta, N; y) = y' A^{-1} \theta - (b_\theta^{1/2})' A^{-1} b_\theta^{1/2} + 1' c$$

biçiminde yazılabilir. Bu eşitlikte,

A: $a(\phi_i)$ 'nin i'ninci bir diyagonal matrisi

θ : θ_i için i'ninci vektörü

b_θ : θ_i için i'ninci vektörü

c: $c(Y_i, \phi_i)$ için i'ninci vektörü

1: n boyutlu bir vektörü ve $b_\theta^{1/2}$ notasyonu ise b_θ unsuruna karşılık gelen karelerinin karekökünü göstermektedir (Dobson, 2002; SAS, 2017).

Yaygın olarak kullanılan üç dağılım için fonksiyonlar aşağıda verilmiştir.

	Binomial/n	Poisson	Normal
ortalama	π	λ	μ
$\theta(\mu)$	$\log[\pi / (1 - \pi)]$	$\log(\lambda)$	μ
$a(\phi)$	1/n	1	σ^2
$V(\mu)$	$\pi(1 - \pi)$	λ	1
$Var[Y]$	$\pi(1 - \pi) / n$	λ	σ^2

Genelleştirilmiş doğrusal modellerde bağlantı fonksiyonu tanımı

Doğrusal modellerde bağlantı fonksiyonu birim matris olup (1'eşit) $\theta = \mu$ şeklinde gösterilmektedir. Poisson dağılım gösteren verilerin dağılımı doğrusal model ailesine ait olmadığından dolayı bu dağılımı doğrusallaştırmak için log bağlantı fonksiyonu kullanılarak aşağıdaki gibi yazılabilir.

$$\theta = \log(\lambda)$$

Burada $\log(\lambda)$ aynı zamanda log-doğrusal modeli de göstermektedir. Bununla birlikte binom dağılımı da genelleştirilmiş doğrusal modeller ailesine ait bir dağılımdır. Binom dağılımını doğrusallaştırmak için,

$$\theta = \log[\pi/(1-\pi)]$$

bağlantı fonksiyonu kullanılır. Bu bağlantı fonksiyonu aynı zamanda lojistik regresyondaki lojit (logit) dönüşüme de karşılık gelmektedir.

Bu çalışmada kullanılan Poisson dağılımının üssel formda aşağıdaki gibi yazılabilir. Poisson dağılımının olasılık fonksiyonu,

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

biçiminde yazılabilir. Buradan log-olabilirlik fonksiyonu da aşağıdaki gibi yazılabilir (SAS, 2017).

$$l(\lambda; y) = y \log(\lambda) - \lambda - \log(y!)$$

3.2.3. Log- doğrusal modellerin genel yapısı

IXJ boyutlu bir tablo göz önüne alınsın (Eye ve Mun, 2013).

$\hat{\pi}_{ij} = \pi_{i.} \pi_{.j}$ olsun. Bu durumda,

$$\hat{m}_{ij} = \frac{m_{i.} m_{.j}}{N} = \hat{m}_{ij} = \pi_{i.} \pi_{.j} N$$

IXJ boyutlu bir tablonun her bir hücresindeki beklenen frekansları gösterebilir. 2X2 Çapraz sınıflandırılmış tabloda, ilk değişken A ile ($i=1,2,\dots,I$) ve ikinci değişken ise B ile ($j=1,2,\dots,J$) gösterilsin.

$\hat{m}_{ij} = \pi_i \pi_j N$ eşitliğinin doğal logaritması aşağıdaki gibi yazılabilir. Burada $\hat{m}_{ij}$ (i, j) hücresindeki gözlemlerinin olasılığının N ile çarpılmış halini göstermektedir.

$$\log(\hat{m}_{ij}) = \log N + \log \pi_i + \log \pi_j$$

Bu eşitlik log-doğrusal formda,

$$\log(\hat{m}_{ij}) = \lambda + \lambda_i^A + \lambda_j^B$$

biçiminde de yazılabilir. Log-doğrusal model formu kullanılarak sıfır modeli (null model) ile A ve B değişken kategorileri için yalnızca ana etkileri içeren ana etkileri modelleri yazılabilir. Sıfır modeli,

$$\log(\hat{m}_{ij}) = \log N = \lambda$$

A değişken kategorisi için ana etkileri modeli

$$\log(\hat{m}_{ij}) = \log N + \log \pi_i = \lambda + \lambda_i^A$$

B değişken kategorisi için ana etkileri modeli

$$\log(\hat{m}_{ij}) = \log N + \log \pi_j = \lambda + \lambda_j^B$$

biçiminde yazılabilirler. Yukarıda tanımlanan A ve B değişken kategorileri için ana etkileri modelini birlikte alındığında aşağıdaki gibi yazılabilir.

$$\log(\hat{m}_{ij}) = \log N + \log \pi_i + \log \pi_j = \lambda + \lambda_i^A + \lambda_j^B$$

Böylece, log-doğrusal formu kullanarak A ve B değişkenleri arasındaki ilişkiyi hesaba katan modeli,

$$\nu_{ij} = \log(\hat{m}_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad (2)$$

biçiminde yazılabilir. Bu eşitlikteki λ_{ij}^{AB} terimi A ve B değişkenleri arasındaki interaksyonu göstermektedir. A ve B değişkenlerinin için yazılan model ana etkileri modeli iken, iki değişken arasındaki ilişkiyi açıklayan model ise doymuş model (saturated model) olarak adlandırılmaktadır. Doymuş modelin serbestlik derecesi sıfıra eşittir. Olası tüm hiyerarşik etkileri hesaba katmaktadır.

Hem ana etkiler hem de interaksyon etkiyi ye ait λ parametrelerinin tahmin edilmesi gerekmektedir. Bu parametrelerin sayısı çok olabilir. Örneğin 2X2 boyutlu doymuş bir model ele alındığında, tabloda dört hücre olacaktır. Bu durumda doymuş model için tahmin edilmesi gereken parametreler sırasıyla,

$$\lambda, \lambda_1^A, \lambda_2^A, \lambda_1^B, \lambda_2^B, \lambda_{11}^{AB}, \lambda_{12}^{AB}, \lambda_{21}^{AB} \text{ ve } \lambda_{22}^{AB}$$

biçiminde yazılabilir. 2X2 boyutlu tabloda dört hücre olmasına rağmen tahmin edilmesi gereken dokuz parametre vardır. Parametre sayısı olasılıklardan daha fazladır. Bu tür modeller aşırı parametre içeren modeller olarak da adlandırılmaktadır. Bu problemi çözmek için parametrelerde kısıtlama yapılır. Değişkenleri sınırlamak için en popüler yaklaşım toplamalarının sıfıra eşitlenmesidir. Örneğin, İXJ boyutlu bir tabloda kısıtlama işlemi aşağıdaki gibi yapılabilir.

$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ji}^{AB} = 0$$

Buna benzer olarak, 2X2 boyutlu bir tabloda kısıtlama işlemi,

$$\left\{ \begin{array}{ll} \lambda_1^A + \lambda_2^A = 0 & \\ \lambda_1^B + \lambda_2^B = 0 & v_{11} = \mu \\ \lambda_{11}^{AB} + \lambda_{12}^{AB} = 0 & v_{i1} = \mu + \lambda_i^A, i \neq 1 \\ \lambda_{21}^{AB} + \lambda_{22}^{AB} = 0 & v_{1j} = \mu + \lambda_j^B, j \neq 1 \\ \lambda_{11}^{AB} + \lambda_{21}^{AB} = 0 & v_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, i \neq 1, j \neq 1 \\ \lambda_{12}^{AB} + \lambda_{22}^{AB} = 0 & \end{array} \right. = \quad (3)$$

biçiminde yazılabilir. Yukarıdaki eşitliğin ilk sütununda verilen son dört kısıtlamadan her hangi biri işlem dışı bırakılır ve geriye 5 kısıtlama kalır. Böylece doymuş model dokuz eşitlikten 4 eşitliğe inmiş olur (Eye ve Mun, 2012). Yukarıda verilen eşitlikte, bazı matematiksel düzenlemeler yapıldığında (Upton, 2017),

$$\lambda_i^A = v_{i1} - \mu = v_{i1} - v_{11} = \ln \left(\frac{p_{i1}}{p_{11}} \right)$$

$$\lambda_j^B = v_{1j} - \mu = v_{1j} - v_{11} = \ln \left(\frac{p_{1j}}{p_{11}} \right)$$

Daha sonra bu yazılan eşitlik kullanılarak kısıtlamalar interaksiyon etki aşağıdaki gibi yazılabilir,

$$\begin{aligned}
\lambda_{ij}^{AB} &= v_{ij} - v_{11} - (v_{i1} - v_{11}) - (v_{j1} - v_{11}) \\
&= v_{ij} - v_{j1} - v_{i1} + v_{11} \\
&= \ln \left(\frac{p_{ij} p_{11}}{p_{i1} p_{11}} \right)
\end{aligned}$$

Burada λ_i^A ve λ_j^B parametrelerinin ana etkileri oddsun logaritması olarak ifade edilmekte olup,

$$\frac{p_{ij} p_{11}}{p_{i1} p_{11}} = \frac{\frac{p_{11}}{p_{i1}} \frac{p_{ij}}{p_{ij}}}{\frac{p_{11}}{p_{i1}} \frac{p_{ij}}{p_{ij}}} \quad (4)$$

biçiminde yazılabilir. Yukarıda yapılan kısıtlamalar özetlenecek olursa; kısıtlamalar sonucu aşağıdaki parametreler sıfır olarak alındı.

$$\lambda_2^A = \lambda_2^B = \lambda_{12}^{AB} = \lambda_{21}^{AB} = \lambda_{22}^{AB} = 0$$

kısıtlamalar sonucunda geriye kalan parametreler aşağıdaki gibi yazılabilir (Eye ve Mun, 2013).

$$\lambda, \lambda_1^A, \lambda_1^B, \lambda_{11}^{AB}$$

Log-doğrusal modelin daha iyi anlaşılabilmesi için aşağıdaki örnek verilmiştir. Kişilerin sigara alışkanlığı ile cinsiyetleri sorulmuş ve elde edilen sonuçlar aşağıda verilmiştir.

	Sigara içmeyen	Sigara içen	Toplam
Kadın	17	122	139
Erkek	31	109	140
Toplam	48	231	279

SAS programı ve sonuçlar aşağıda verilmiştir.

```

data a;
input cinsiyet sigara sayi;
cards;
1 1 17
1 2 122
2 1 31
2 2 109
;
run;

proc genmod data=a;
class cinsiyet sigara;
model sayi=cinsiyet sigara cinsiyet*sigara/link=log dist=poi;
run;

```

Analysis Of Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.6913	0.0958	4.5036	4.8791	2398.95	<.0001
cinsiyet	1	1	0.1127	0.1318	-0.1456	0.3710	0.73	0.3926
cinsiyet	2	0	0.0000	0.0000	0.0000	0.0000	.	.
sigara	1	1	-1.2574	0.2035	-1.6563	-0.8584	38.16	<.0001
sigara	2	0	0.0000	0.0000	0.0000	0.0000	.	.
cinsiyet*sigara	1 1	1	-0.7134	0.3293	-1.3589	-0.0680	4.69	0.0303
cinsiyet*sigara	1 2	0	0.0000	0.0000	0.0000	0.0000	.	.
cinsiyet*sigara	2 1	0	0.0000	0.0000	0.0000	0.0000	.	.
cinsiyet*sigara	2 2	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Burada;

$$\mu = \lambda = 4.6913$$

$$\lambda_1^{cinsiyet} = 0.1127$$

$$\lambda_1^{sigara} = -1.2574$$

$$\lambda_{11}^{cinsiyet*sigara} = -0.7134 \text{ ya da}$$

$$\log(OR) = \log\left(\frac{17 * 109}{31 * 122}\right) = \log(0.499) = -0.7134$$

olarak elde edilir.

3.2.4. Poisson regresyonu

Poisson regresyonunda (PR), ilgilenilen olayın gözlenen sayısı olan y_i bağımlı değişkenin Poisson dağılışına sahip olduğu varsayılmaktadır. Poisson ortalaması olan μ 'nün logaritmasının, bağımsız değişkenlerin bir doğrusal fonksiyonu olduğu varsayılmaktadır (Stokes ve ark., 2000; SAS, 2017). Log bağlantı fonksiyonlu Poisson regresyon modeli,

$$\Pr[Y_i = y_i] = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, y_i = 0, 1, 2.. \quad (5)$$

Bişiminde yazılabilir. PR modelinin yaygın olarak kullanım biçimi aşağıdaki gibi verilebilir,

$$E[y_i | x_i] = \lambda_i = \exp(x_i' \beta) \quad (6)$$

Eşitlik 5’de verilen PR regresyon modelini doğrusallaştırmak için log link fonksiyonu kullanılarak log-doğrusal (log-linear) yaklaşımı,

$$\log(E[y_i | x_i]) = \log(\lambda_i) = x_i' \beta \quad (7)$$

Poisson dağılışında her bir olayın gerçekleşme sayısının bekleneni,

$$E[y_i | x_i] = \lambda_i = \exp(x_i' \beta)$$

Bişiminde yazılabilir. Bu herhangi bir olayın beklenen gerçekleşme sayısını elde etmek için bağımsız değişkenlere göre kısmi türev alınarak aşağıdaki gibi elde edilmektedir.

$$\frac{\partial E[y_i | x_i]}{\partial x_i} = \lambda_i \beta_i$$

PR modelinin temel varsayımı, beklenen değer (ortalama) ile varyansın birbirine eşit olmasıdır. PR’ de bu eşitlik aşağıdaki gibi yazılabilir.

$$E[y_i | x_i] = \lambda_i = e^{x_i' \beta} = Var[y_i | x_i] \quad (8)$$

Ancak uygulamada bu eşitliği sağlamak her zaman mümkün değildir. Eğer,

$Var[y] > E[y]$ ise aşırı yayılım

$Var[y] < E[y]$ ise az yayılım

olarak adlandırılmaktadır. Az ya da aşırı yayılım durumunu test etmek için diyagnostik test (diagnostic testing) kullanılmaktadır.

3.2.5. Poisson regresyonunda parametre tahmini

PR’de, parametre tahmini en çok olabilirlik yöntemi (Maximum likelihood estimation=MLE) kullanılarak elde edilmektedir. Doğrusal olmayan PR modeli için olabilirlik fonksiyonu (likelihood function),

$$\ln L = \sum_{i=1}^n [-\lambda_i + y_i x_i' \beta - \ln y_i!] = \sum_{i=1}^n [-e^{x_i' \beta} + y_i x_i' \beta - \ln y_i!] \quad (9)$$

biçiminde yazılabilir. Eşitlik 9’da, bilinmeyen parametreler en çok olabilirlik yöntemi kullanılarak aşağıdaki gibi elde edilmektedir.

$$\frac{\partial L}{\partial \beta} = - \sum_{i=1}^n [\mathbf{x}_i [e^{x_i' \beta} - y_i]] = 0$$

3.2.6. Poisson log linear regresyon analizi

Poisson dağılımının ortalaması μ olmak üzere, veri setimiz için; “n” sırasıyla “YGS’ye giren öğrenci sayısı” ve “180 ve üzeri puan alan öğrenci sayısı” iken, “p” sırasıyla “180 puan altı alan öğrenci sayısı” ve “180 ve üzeri puan alıp ancak herhangi bir programa yerleşemeyen öğrenci sayısı” olsun. Bilindiği gibi “n” değeri çok büyük ve “p” değeri çok küçük olduğunda, binom dağılımı yaklaşık olarak Poisson dağılımına yaklaşmaktadır. Bu nedenle, “180 puan altı alan öğrenci sayısı” ve “180 ve üzeri puan alıp ancak herhangi bir programa yerleşemeyen öğrenci sayısı” için,

$$\mu = \frac{p}{n}$$

biçiminde yazılabilir.

İki yönlü çapraz tablonun genel formatı aşağıdaki gibi yazılabilir. y_{ij} i'ninci satır ve j'inci sütunda elde edilen gözlenen değer ve n_{ij} , i'ninci satır ve j'inci sütunda elde edilen kitle değeri olsun ($i = 1, \dots, r$ ve $j = 1, \dots, c$).

		Factor B			
		1	2	3C	
Factor A	1	$y_{11} \quad n_{11}$	$y_{12} \quad n_{12}$	$y_{13} \quad n_{13}$	$y_{1c} \quad n_{1c}$
	2	$y_{21} \quad n_{21}$	$y_{22} \quad n_{22}$	$y_{23} \quad n_{23}$	$y_{2c} \quad n_{2c}$
	:	:	:	:	:
	r	$y_{r1} \quad n_{r1}$	$y_{r2} \quad n_{r2}$	$y_{r3} \quad n_{r3}$	$y_{rc} \quad n_{rc}$

p_{ij} , her bir hücredeki olasılığı göstermek üzere aşağıdaki gibi yazılabilir.

$$p_{ij} = y_{ij} / n_{ij}$$

Buradan, Poisson log-doğrusal regresyon modeli aşağıdaki gibi yazılabilir.

$$\log [p_{ij}] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

ya da

$$p_{ij} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k}$$

Bu modeli örneklem bazında ise,

$$\hat{p}_{ij} = e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_k X_k}$$

biçiminde yazılabilir. Bu eşitlikte, $e^{\hat{\beta}_i}$ en çok olabilirlik regresyon katsayısı, X_i ($i = 1, \dots, k$) bağımsız değişkendeki bir birimlik değişimin Y bağımlı değişkende yaratacağı riski göstermektedir. Böylece, iki bağımsız değişken kategorisi arasındaki düzeltilmiş relatif risk istatistiği,

$$\hat{RR}_{X_1=1 \text{ vs } X_1=0} = \frac{\hat{Risk}(X_1 = 1, X_2, \dots, X_k)}{\hat{Risk}(X_1 = 0, X_2, \dots, X_k)} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k}}{e^{\hat{\beta}_0 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k}} = e^{\hat{\beta}_1}$$

biçiminde yazılabilir.

3.2.7. Model uyumu

Genelleştirilmiş doğrusal modellerde model uyum ölçütü olarak Devians İstatistiği yaygın olarak kullanılmaktadır. Devians (deviance) İstatistiği aynı zamanda sapma istatistiği olarak da adlandırılmaktadır. Bağımlı değişkenin sayıma dayalı olarak elde edilmesi durumunda kullanılacak regresyon yöntemlerinden hangisinin uygun olduğunu belirlemek için kullanılan bir uyum iyiliği testidir. Devians uyum iyiliği istatistiği aşağıdaki gibi yazılabilir (SAS, 2017).

Devians uyum ölçütü aşağıdaki gibi yazılabilir.

$$D = 2 \sum_{i=1}^r \sum_{j=1}^c [y_{ij} \log\left(\frac{y_{ij}}{\hat{\mu}_{ij}}\right) - (y_{ij} - \hat{\mu}_{ij})]$$

bu eşitlikte; $\hat{\mu}_{ij} = \hat{p}_{ij} * n_{ij}$ her bir hücredeki beklenen değeri biçiminde hesaplanmaktadır. Serbestlik derecesi ise, hücre sayısından parametre sayısı (sabit terim dahil) çıkartılarak elde edilir. Yukarıdaki eşitlikte de görüldüğü gibi, hücrelerdeki gözlenen değerler ile beklenen değerler arasındaki uyumu ölçmektedir. Devians ölçütü standart çoklu regresyon analizinde hata kareler toplamına karşılık gelmektedir. Model uyumu tam olduğunda devians ölçütü sıfıra eşit olur ($D = 0$). İki yönlü tablolardaki hücrelerde, gözlenen ve beklenen değerler arasındaki uyum büyüdükçe, devians değeri de büyümektedir. Eğer, $\hat{\mu}_{ij} > 3$ ise devians ölçütü yaklaşık olarak Pearson ki-kare istatistiğine yaklaşabilir.

A modeli “p” tane bağımsız değişkene ve B modeli de “k” bağımsız değişkene sahip olsun. Bu durumda olabilirlik ki-kare oran istatistiği (likelihood Chi-Square ratio statistics),

$$G^2 = (-2 \log L_A) - (-2 \log L_B)$$

Yazılabilir. Bu eşiklik,

$$Devians_A - Devians_B$$

Eşitliğine benzemektedir. Her iki istatistikte, (k-p) serbestlik dereceli X^2 dağılımı gösterir.

3.2.8. En çok olabilirlik tahminleri için hipotez testi

Regresyon katsayıları için hipotezler “Wald ki-kare testi” kullanılarak test edilmektedir. Wald ki-kare test istatistiğinin aşamaları aşağıda verilmiştir.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Z – test istatistiği,

$$Z_{hesap} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

$$P - \text{value: } 2P(Z \geq |Z_{hesap}|)$$

χ^2 – test istatistiği,

$$X_{hesap}^2 = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2$$

$$P - \text{value } (\chi^2 \text{ için}): P(\chi_1^2 \geq X_{hesap}^2) \quad (i = 1, \dots, r; j = 1, \dots, c)$$

3.2.9. Relatif Risk

Eğer iki oranın değerleri çok küçük olduğunda, bunların farkı da küçük olacaktır. Yinede bu küçük farklılıklar önemli olabilir. Örneğin, A tedavisi ile hayatta kalma olasılığı %2 iken B tedavisi ile hayatta kalma olasılığı %1 olduğunda, iki oran arasındaki fark %1 olsa dahi A tedavisinin kullanılması gerektiği söylenebilir. Başka bir deyişle, A tedavi yönteminin etkinliğinin, B tedavi yönteminin iki katıdır.

Özellikle sağlık alanında, relatif risk etki altında kalan bireylerin insidansının etki altında kalmayan birey insidansına oranı biçiminde tanımlanmaktadır. Bu başarı olasılığı,

$$R = \frac{P_1}{P_2}$$

Olarak yazılabilir ve bu relatif risk olmaktadır. 2X2 çapraz sınıflandırılmış tabloda relatif risk aşağıdaki gibi yazılabilir.

$$\hat{R} = \frac{a/m}{c/n} = \frac{an}{cm}$$

Burada,

$\hat{R}$ istatistiği kesikli bir dağılım göstermiş olup, p_1 ve p_2 olasılıklarına bağlı olarak değişmektedir. Bununla birlikte, normal yaklaşım $\ln(\hat{R})$ 'nin dağılımı için kullanılabilir. Aşağıda verilen yaklaşık standart normal dağılım gösteren güven aralığı eşitliği,

$$\text{Re } xp \left(\ln(\hat{R}) \pm \sin h^{-1} \left\{ \frac{1}{2} z_0 \sqrt{\frac{1}{a} + \frac{1}{b} - \frac{1}{m+1} - \frac{1}{n+1}} \right\} \right) \quad (1)$$

Bu eşitlikte z_0 standart normal dağılım kritik değerini göstermektedir.

Örneğin, yapılan bir çalışmada (Eye ve Mun, 2013), stres faktörünün prostat ve meme kanserlerini etkileyip etkilemediği araştırılmıştır. Bunun için elde edilen sonuçlar aşağıdaki 2X2 çapraz sınıflandırılmış tablo halinde göz önüne alındığında; erkeklerin %22.4'ü (37 kişi) stresin kansere neden bir faktör olduğunu, kadınların ise %37.9'u (165 kişi) stresin kansere neden bir faktör olduğunu düşünmektedirler. Bu durumda relatif risk, erkeklere nazaran kadınların stresi kanser nedeni olarak görme oranları;

$$\hat{R} = \frac{165/416}{37/165} = 1.77$$

şeklinde yazılabilir. Bu durumda kadınların stresi kansere neden bir faktör olarak görme relatif riski erkeklerden 1.77 kat daha fazladır. Eşitlik 1 kullanılarak yaklaşık %95'lik güven sınırları (1.30, 2.40) biçiminde elde edilebilir. Güven aralığı 1 değerini kapsamadığı için stresi risk faktörü görme açısından erkek ile kadın oranları arasındaki farkın istatistiksel olarak anlamlıdır.

3.2. 10. Pseudo- R^2 belirleme katsayısı

Genelleştirilmiş doğrusal modelleri esas lojistik ve Poisson gibi regresyon modellerinde, R^2 belirleme katsayısı elde edilememektedir. Bu tip doğrusal olmayan regresyon modellerinde parametreler iteratif bir sürecin sonucunda en çok olabilirlik

yöntemi ile tahmin edilir. ML varyansı minimize etmez. Bundan dolayı en küçük kareler yaklaşımı uyum iyiliği için uygulanmaz. Bu nedenle, R^2 belirleme katsayısı yerine bir çok Pseudo- R^2 katsayısı, bu tip regresyon modelleri için geliştirilmiştir. Bu Pseudo- R^2 katsayıları bilinen R^2 belirleme katsayısına benzemektedir. Bunlar aynı skalada olup 0-1 arası değişmektedir. Pseudo- R^2 'si yüksek olan modelin daha iyi olduğu söylenebilir ancak en küçük karelerdeki R^2 gibi yorumlanmaz. Pseudo- R^2 istatistiğinin genel hesaplama formu aşağıda verilmiştir.

$$R^2 = 1 - \frac{\ln \hat{L}(M_{model})}{\ln \hat{L}(M_{sabit})}$$

Eşitlikte $\hat{L}$ tahmin edilen olabilirlik değerini göstermektedir. Pseudo- R^2 belirleme katsayısını kendi modelimize uyarladığımızda, aşağıdaki gibi yazılabilir.

$$\text{Pseudo-} R^2 = \frac{D_A - D_B}{D_A}$$

4. BULGULAR

Çalışmada gerekli istatistiksel analizler SAS 9.1.4 istatistik yazılım programı kullanılarak yapılmıştır. SAS'ta “proc GENMOD” alt programı kullanılmıştır. Poisson log-doğrusal regresyon analizi iki aşamada yapılmıştır. Bu aşamalar;

1. Poisson log-regresyon modeline “180 puan altı alan öğrenci sayısı” bağımlı değişken olarak, 7 farklı bölge ve 4 farklı yıl ise bağımsız değişkenler olarak alınmıştır.
2. Poisson log-regresyon modeline “herhangi bir lisans programına yerleşemeyen öğrenci sayısı” bağımlı değişken olarak, 7 farklı bölge ve 4 farklı yıl ise bağımsız değişkenler olarak alınmıştır.

Bulgular iki başlık halinde verilmiştir. Bunlar; veri seti hakkında genel bilgileri veren “tanıtıcı istatistikler” ve “Poisson log-doğrusal regresyon analizi sonuçları” şeklinde sıralanmıştır.

4.1. Tanıtıcı İstatistikler

Bu çalışmada, YGS’de, “180 puan altı alan öğrenci sayısı” ve “herhangi bir lisans programına yerleşemeyen öğrenci sayısı” relatif riskleri esas alınarak Poisson log-doğrusal regresyon uygulandığından dolayı, söz konusu bu risk faktörlerinin bölgelere ve yıllara göre dağılımları Çizelge 4.1.1. –Çizelge 4.1.4 arasında verilmiştir.

Çizelge 4.1.1. 2010 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler

Bölgeler	Yıllar	Değişkenler	Ortalama (s.hata)	Toplam
Akdeniz	2010	YGS'ye giren öğrenci sayısı	10996.130 (2323.580)	87969.000
		180 puan altı alan öğrenci sayısı	1947.000 (389.554)	15576.000
		Lisans programına yerleşemeyen öğrenci sayısı	6486.630 (1403.370)	51893.000
Doğu Anadolu	2010	YGS'ye giren öğrenci sayısı	3310.000 (649.468)	46340.000
		180 puan altı alan öğrenci sayısı	678.928 (117.647)	9505.000
		Lisans programına yerleşemeyen öğrenci sayısı	2084.360 (425.456)	29181.000
Ege	2010	YGS'ye giren öğrenci sayısı	10325.000 (3438.000)	82600.000
		180 puan altı alan öğrenci sayısı	1762.000 (523.415)	14096.000
		Lisans programına yerleşemeyen öğrenci sayısı	5478.630 (1835.390)	43829.000
Güney Doğu	2010	YGS'ye giren öğrenci sayısı	6366.560 (1421.290)	57299.000
		180 puan altı alan öğrenci sayısı	2445.330 (1164.820)	22008.000
		Lisans programına yerleşemeyen öğrenci sayısı	3335.440 (811.289)	30019.000
Karadeniz	2010	YGS'ye giren öğrenci sayısı	4157.110 (662.026)	74828.000
		180 puan altı alan öğrenci sayısı	820.055 (116.408)	14761.000
		Lisans programına yerleşemeyen öğrenci sayısı	2356.220 (394.772)	42412.000
Marmara	2010	YGS'ye giren öğrenci sayısı	17202.360 (9563.150)	189226.000
		180 puan altı alan öğrenci sayısı	3289.640 (1783.850)	36186.000
		Lisans programına yerleşemeyen öğrenci sayısı	8850.270 (4898.120)	97353.000
İç Anadolu	2010	YGS'ye giren öğrenci sayısı	8730.770 (3555.170)	113500.000
		180 puan altı alan öğrenci sayısı	1426.080 (554.758)	
		Lisans programına yerleşemeyen öğrenci sayısı	4879.690 (1854.560)	18539.000 63436.000

Çizelge 4.1.2. 2011 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler

Bölgeler	Yıllar	Değişkenler	Ortalama (s.hata)	Toplam
		YGS'ye giren öğrenci sayısı	12227.880(2633.910)	97823.000
Akdeniz	2011	180 puan altı alan öğrenci sayısı	2822.380 (622.745)	22579.000
		Lisans programına yerleşemeyen öğrenci sayısı	6805.750 (1509.740)	54446.000
		YGS'ye giren öğrenci sayısı	5909.860 (2225.130)	82738.000
Doğu Anadolu	2011	180 puan altı alan öğrenci sayısı	3124.140 (2140.890)	43738.000
		Lisans programına yerleşemeyen öğrenci sayısı	2232.930 (461.998)	31261.000
		YGS'ye giren öğrenci sayısı	11403.250 (3850.720)	91226.000
Ege	2011	180 puan altı alan öğrenci sayısı	2479.750 (793.529)	19838.000
		Lisans programına yerleşemeyen öğrenci sayısı	5813.250 (1984.200)	46506.000
		YGS'ye giren öğrenci sayısı	7443.890 (1668.450)	66995.000
Güney Doğu	2011	180 puan altı alan öğrenci sayısı	2168.670 (497.163)	19518.000
		Lisans programına yerleşemeyen öğrenci sayısı	4348.000 (966.031)	39132.000
		YGS'ye giren öğrenci sayısı	4522.940 (727.341)	81413.000
Karadeniz	2011	180 puan altı alan öğrenci sayısı	1105.330 (169.681)	19896.000
		Lisans programına yerleşemeyen öğrenci sayısı	2439.500 (413.314)	43911.000
		YGS'ye giren öğrenci sayısı	19285.910 (10950.380)	212145.000
Marmara	2011	180 puan altı alan öğrenci sayısı	4639.180 (2593.930)	51031.000
		Lisans programına yerleşemeyen öğrenci sayısı	9393.910 (5297.780)	103333.000
		YGS'ye giren öğrenci sayısı	9639.920 (3930.130)	125319.000
İç Anadolu	2011	180 puan altı alan öğrenci sayısı	2068.380 (812.080)	26889.000
		Lisans programına yerleşemeyen öğrenci sayısı	5109.460 (1944.230)	66423.000

Çizelge 4.1.3. 2012 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler

Bölgeler	Yıllar	Değişkenler	Ortalama (s.hata)	Toplam
		YGS'ye giren öğrenci sayısı	12515.000 (2743.600)	100120.000
Akdeniz	2012	180 puan altı alan öğrenci sayısı	3710.880 (851.8290)	29687.000
		Lisans programına yerleşemeyen öğrenci sayısı	6070.750 (1337.840)	48566.000
		YGS'ye giren öğrenci sayısı	3764.140 (686.044)	52698.000
Doğu Anadolu	2012	180 puan altı alan öğrenci sayısı	1259.710 (201.786)	17636.000
		Lisans programına yerleşemeyen öğrenci sayısı	1925.570 (385.483)	26958.000
		YGS'ye giren öğrenci sayısı	11470.630 (3809.420)	91765.000
Ege	2012	180 puan altı alan öğrenci sayısı	3317.250 (1081.200)	26538.000
		Lisans programına yerleşemeyen öğrenci sayısı	5005.250 (1638.440)	40042.000
		YGS'ye giren öğrenci sayısı	7718.780 (1703.490)	69469.000
Güney Doğu	2012	180 puan altı alan öğrenci sayısı	2843.000 (636.610)	25587.000
		Lisans programına yerleşemeyen öğrenci sayısı	3871.560 (835.300)	34844.000
		YGS'ye giren öğrenci sayısı	4526.110 (729.682)	81470.000
Karadeniz	2012	180 puan altı alan öğrenci sayısı	1389.500 (212.363)	25011.000
		Lisans programına yerleşemeyen öğrenci sayısı	2135.670 (368.655)	38442.000
		YGS'ye giren öğrenci sayısı	19907.910 (11214.890)	218987.000
Marmara	2012	180 puan altı alan öğrenci sayısı	13968.820 (7949.630)	153657.000
		Lisans programına yerleşemeyen öğrenci sayısı	8507.090 (4781.620)	93578.000
		YGS'ye giren öğrenci sayısı	9877.460 (4042.090)	128407.000
İç Anadolu	2012	180 puan altı alan öğrenci sayısı	2701.380 (1054.940)	35118.000
		Lisans programına yerleşemeyen öğrenci sayısı	4575.000 (1735.370)	59475.000

Çizelge 4.1.4. 2013 yılında 7 farklı bölgedeki YGS'ye giren öğrenci sayısı, 180 puan altı alan öğrenci sayısı ve lisans programına yerleşemeyen öğrenci sayısı için tanıtıcı istatistikler

Bölgeler	Yıllar	Değişkenler	Ortalama (s.hata)	Toplam
Akdeniz	2013	YGS'ye giren öğrenci sayısı	12931.880 (2877.450)	103455.000
		180 puan altı alan öğrenci sayısı	4281.000 (1017.820)	34248.000
		Lisans programına yerleşemeyen öğrenci sayısı	5541.380 (1234.290)	44331.000
Doğu Anadolu	2013	YGS'ye giren öğrenci sayısı	4044.710 (760.452)	56626.000
		180 puan altı alan öğrenci sayısı	1490.070 (254.940)	20861.000
		Lisans programına yerleşemeyen öğrenci sayısı	1812.860 (377.956)	25380.000
Ege	2013	YGS'ye giren öğrenci sayısı	11487.000 (3808.370)	91896.000
		180 puan altı alan öğrenci sayısı	3680.000 (1193.360)	29440.000
		Lisans programına yerleşemeyen öğrenci sayısı	4632.130 (1491.670)	37057.000
Güney Doğu	2013	YGS'ye giren öğrenci sayısı	8167.330 (1812.910)	73506.000
		180 puan altı alan öğrenci sayısı	3317.220 (736.978)	29855.000
		Lisans programına yerleşemeyen öğrenci sayısı	3629.220 (793.880)	32663.000
Karadeniz	2013	YGS'ye giren öğrenci sayısı	4592.720 (748.393)	82669.000
		180 puan altı alan öğrenci sayısı	1526.110 (243.408)	27470.000
		Lisans programına yerleşemeyen öğrenci sayısı	1928.220 (328.696)	34708.000
Marmara	2013	YGS'ye giren öğrenci sayısı	20174.550 (11583.820)	221920.000
		180 puan altı alan öğrenci sayısı	6527.360 (3706.900)	71801.000
		Lisans programına yerleşemeyen öğrenci sayısı	7569.640 (4343.660)	83266.000
İç Anadolu	2013	YGS'ye giren öğrenci sayısı	10010.380 (4078.920)	130135.000
		180 puan altı alan öğrenci sayısı	3006.000 (1172.010)	39078.000
		Lisans programına yerleşemeyen öğrenci sayısı	4076.620 (1546.899)	52996.000

4.2. Poisson Log-Doğrusal Regresyon Analizi Sonuçları

Poisson log-doğrusal regresyon analizi sonuçları; YGS sınavında 180 altında puan alan öğrenciler ve herhangi bir lisans programına yerleşemeyen öğrenciler için olmak üzere iki başlık altında verilmiştir.

4.2.1. YGS sınavında 180 altında puan alan öğrenciler için Poisson log-doğrusal regresyon modeli sonuçları

Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımları Çizelge 4.2.1’de verilmiştir. Bu çizelgede; “n”, YGS sınavına giren öğrenci sayısını ve “c” de 180 altında puan alan öğrencilerin sayılarını göstermektedir. YGS sınavında 180 altında puan alan öğrencilerin sayıları için Poisson log-doğrusal regresyon modeli,

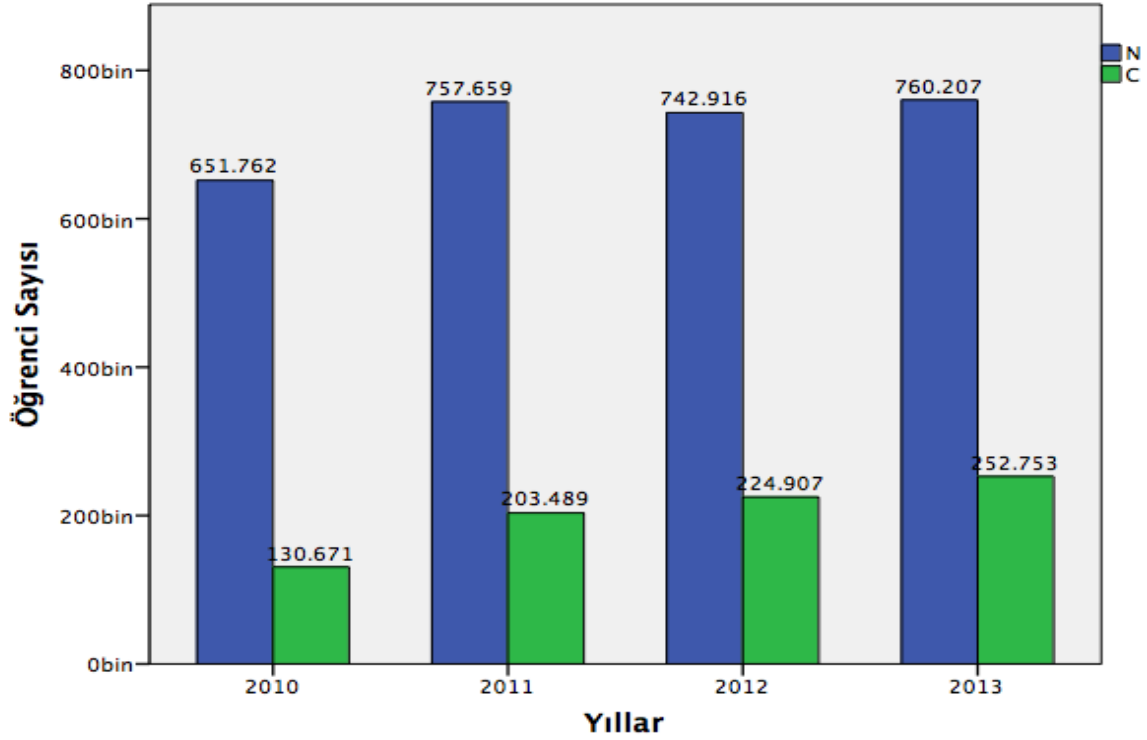
$$\log(p_{ij}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 Z_7 + \beta_8 Z_8 + \beta_9 Z_9 \quad (4.2.1)$$

biçiminde yazılabilir. Eşitlik 4.2.1’de; X_1, X_2, X_3, X_4, X_5 ve X_6 sırasıyla Doğu Anadolu, İç Anadolu, Karadeniz, Akdeniz, Ege ve Marmara bölgelerini; Z_7, Z_8 ve Z_9 sırasıyla 2011, 2012 ve 2013 yıllarını göstermektedir. Eşitlikte, bölgeler içinde Güneydoğu Anadolu yıllar arasında ise 2010 referans kategorisi olarak alınmıştır.

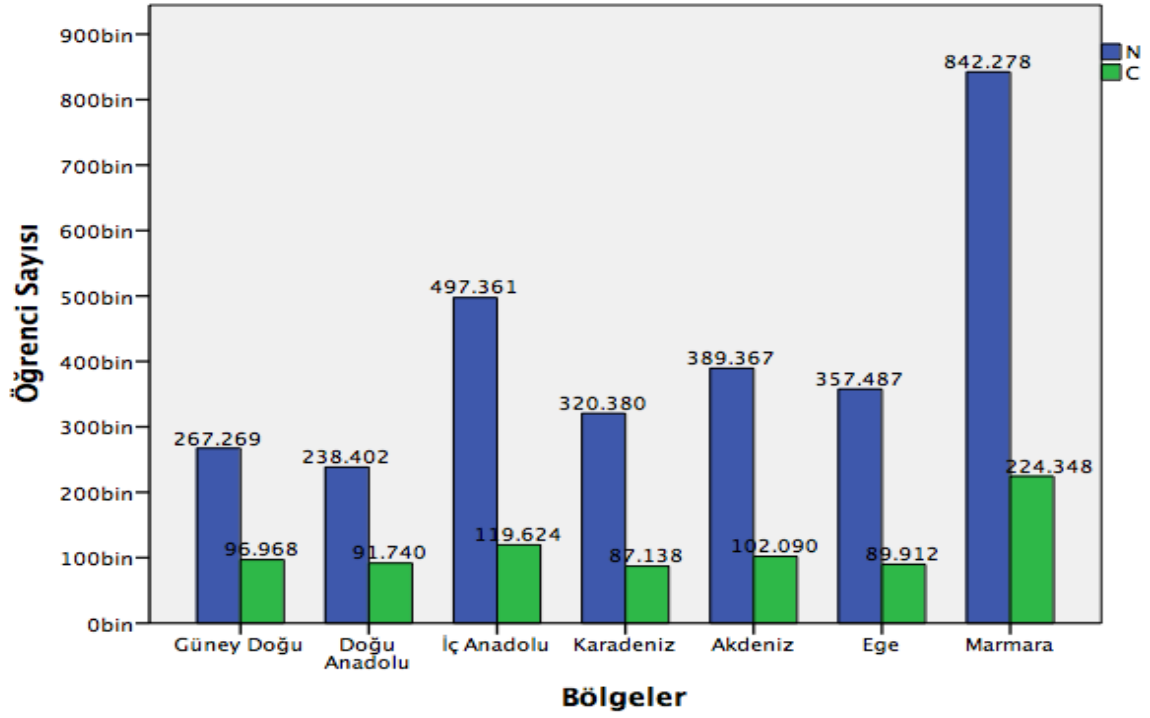
Çizelge 4.2.1. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı

Bölgeler	Yıllar							
	2010		2011		2012		2013	
	n	c	n	c	n	c	n	c
Güney Doğu	57299	22008	66995	19518	69469	25587	73506	29855
Doğu Anadolu	46340	9505	82738	43738	52698	17636	56626	20861
İç Anadolu	113500	18539	125319	26889	128407	35118	130135	39078
Karadeniz	74828	14761	81413	19896	81470	25011	82669	27470
Akdeniz	87969	15576	97823	22579	100120	29687	103455	34248
Ege	82600	14096	91226	19838	91765	26538	91896	29440
Marmara	189226	36186	212145	51031	218987	65330	221920	71801

Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımları şekil 4.2.1 ve şekil 4.2.2 de verilmiştir.



Şekil 4.2.1. Yıllara göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı.



Şekil 4.2.2. Bölgelere göre YGS sınavında 180 altında puan alan öğrencilerin dağılımı.

Poisson log-regresyon modeli kullanılarak elde edilen en çok olabilirlik parameter tahminleri Çizelge 4.2.2’de verilmiştir. Eşitlik 4.2.1 esas alındığında, Çizelge 4.2.2’de verilen en çok olabilirlik parametre tahminleri kullanarak Poisson log-doğrusal regresyon model aşağıdaki gibi yazılabilir.

$$\log(p_{ij}) = -1.343 + 0.069 * Do\text{ğ}o\text{ğ}uanadı - 0.405 * içanadolu - 0.279 * karadeniz - 0.319 * akdeniz - 0.358 * ege - 0.304 * marmara + 0.275 * yu_2011 + 0.410 * yu_2012 + 0.501 * yu_2012 \quad (4.2.1)$$

Bu Poisson log-doğrusal regresyon modelini ayrıca,

$$p_{ij} = \exp \left(\begin{array}{l} -1.343 + 0.069 * Do\text{ğ}o\text{ğ}uanadı - 0.405 * içanadolu - 0.279 * karadeniz \\ - 0.319 * akdeniz - 0.358 * ege - 0.304 * marmara + 0.275 * yu_2011 + \\ 0.410 * yu_2012 + 0.501 * yu_2012 \end{array} \right)$$

biçiminde de yazılabilir.

Bölgeler içinde Güneydoğu Anadolu Bölgesi ve yıllar arasında ise 2010 yılı referans kategorisi olarak alınmıştır. Bu nedenle eşitlik 4.2.1’de bu referans kategorilerine yer verilmemiştir. Gerek Çizelge 4.2.2 gerekse eşitlik 4.2.1 kullanılarak referans kategorilerine göre bölgeler ve yıllar bazında, YGS sınavında öğrencilerin 180

altında puan alma relative risklerinin (RR) hesaplanması bölgeler ve yıllar için aşağıda verilmiştir.

Güneydoğu Anadolu Bölgesi referans alındığında diğer bölgelerin RR tahmin değerleri sırasıyla;

$$RR_{\text{doguanadolu vs güneydogu}} = e^{0.069} = 1.071 \text{ ve } \frac{1}{1.071} = 0.934$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski Doğu Anadolu Bölgesindeki öğrencilerin 0.934 katıdır.

$$RR_{\text{icanadolu vs güneydogu}} = e^{-0.405} = 0.667 \text{ ve } \frac{1}{0.667} = 1.499$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski İç Anadolu Bölgesindeki öğrencilere göre 1.499 kat daha fazladır.

$$RR_{\text{karadeniz vs güneydogu}} = e^{-0.279} = 0.757 \text{ ve } \frac{1}{0.757} = 1.321$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski Karadeniz Bölgesindeki öğrencilere göre 1.321 kat daha fazladır.

$$RR_{\text{akdeniz vs güneydogu}} = e^{-0.319} = 0.727 \text{ ve } \frac{1}{0.727} = 1.376$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski Akdeniz Bölgesindeki öğrencilere göre 1.376 kat daha fazladır.

$$RR_{\text{ege vs güneydogu}} = e^{-0.358} = 0.699 \text{ ve } \frac{1}{0.699} = 1.431$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski Ege Bölgesindeki öğrencilere göre 1.431 kat daha fazladır.

$$RR_{\text{marmaraz vs güneydogu}} = e^{-0.304} = 0.738 \text{ ve } \frac{1}{0.738} = 1.355$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında 180 altında bir puan alma relatif riski Marmara Bölgesindeki öğrencilere göre 1.355 kat daha fazladır.

2010 yılı referans alındığında diğer yılların RR tahmin değerleri sırasıyla;

2010 YGS sınavında öğrencileri 180 altında bir puan alma relatif riski 2011 yılındaki öğrencilere göre 0.759 kat daha fazladır.

$$RR_{2011 \text{ vs } 2010} = e^{0.275} = 1.317 \text{ ve } \frac{1}{1.317} = 0.759$$

2010 YGS sınavında öğrencileri 180 altında bir puan alma relatif riski 2012 yılındaki öğrencilere göre 0.664 kat daha fazladır.

$$RR_{2012 \text{ vs } 2010} = e^{0.410} = 1.507 \text{ ve } \frac{1}{1.507} = 0.664$$

2010 YGS sınavında öğrencileri 180 altında bir puan alma relatif riski 2013 yılındaki öğrencilere göre 0.606 kat daha fazladır.

$$RR_{2013 \text{ vs } 2010} = e^{0.501} = 1.650 \text{ ve } \frac{1}{1.650} = 0.606$$

Çizelge 4.2.2. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrenciler için en çok olabilirlik parametre tahminleri

Parametreler	Tahmin	S. hata	%95	Wald	güven	Wald testi	p-değeri
			aralığı				
Sabit	-1.343	0.004	-1.351	-1.335	107112	<0.0001	
Güneydoğu			Referans kategorisi				
Doğu	0.069	0.005	0.060	0.078	221.990	<0.0001	
Anadolu							
İç Anadolu	-0.405	0.004	-0.413	-0.396	8760.580	<0.0001	
Karadeniz	-0.279	0.005	-0.288	-0.270	3568.900	<0.0001	
Akdeniz	-0.319	0.005	-0.328	-0.311	5077.280	<0.0001	
Ege	-0.358	0.005	-0.370	-0.349	5971.860	<0.0001	
Marmara	-0.304	0.004	-0.312	-0.296	6253.230	<0.0001	
2010 yılı			Referans kategorisi				
2011 yılı	0.275	0.003	0.268	0.282	5993.570	<0.0001	
2012 yılı	0.410	0.003	0.403	0.417	13899.600	<0.0001	
2013 yılı	0.501	0.003	0.494	0.508	21613.200	<0.0001	

YGS sınavında 180 altında puan alan öğrenciler göre bölgeler ve yıllar için pseudo R^2 değerlerini hesaplamak için elde edilen log-oran (log ratio) istatistikleri sonuçları Çizelge 4.2.3’de verilmiştir. Bu çizelgede, Bölgeler için pseudo R^2 değeri;

$$pseudo \quad R^2 = \frac{57730.339 - 38323.559}{57730.339} = 0.336 = \%33.6$$

Yıllar için pseudo R^2 değeri;

$$pseudo \quad R^2 = \frac{57730.339 - 13414.087}{57730.339} = 0.768 = \%76.8$$

Bu sonuçlara göre, YGS sınavında 180 altında puan alan öğrencilerin değişimi bakımından bölgeler arası farkın açıkladığı kısım %33.6 iken, yıllar arası farklılığın açıkladığı kısım ise %76.8 olarak saptanmıştır. buna göre YGS sınavında 180 altında puan alan öğrencilerin sayısı yıllar ilerledikçe daha büyük bir varyasyon gösterdikleri saptanmıştır.

Çizelge 4.2.3. Yıllara ve bölgelere göre YGS sınavında 180 altında puan alan öğrenciler için log-oran testi sonuçları

Değişkenler	Devians	Sd	Ki-kare istatistiği	p-değeri
Sabit	57730.339	-	-	-
Bölgeler	38323.559	6	19406.800	<0.0001
Yıllar	13414.087	3	24909.500	<0.0001

Sd: serbestlik derecesi

4.2.2 Herhangi bir lisans programına yerleşemeyen öğrenciler için Poisson log-doğrusal regresyon modeli sonuçları

Yıllara ve bölgelere göre YGS sınavında herhangi bir lisans programına yerleşemeyen öğrencilerin dağılımları Çizelge 4.2.2.1’de verilmiştir. Bu çizelgede; n, YGS sınavında 180 ve üzeri puan alana öğrenci sayısını ve c de herhangi bir programa yerleşemeyen öğrencilerin sayılarını göstermektedir. YGS sınavında herhangi bir programa yerleşemeyen öğrencilerin sayıları için Poisson log-doğrusal regresyon modeli,

$$\log(p_{ij}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 Z_7 + \beta_8 Z_8 + \beta_9 Z_9$$

(4.2.2)

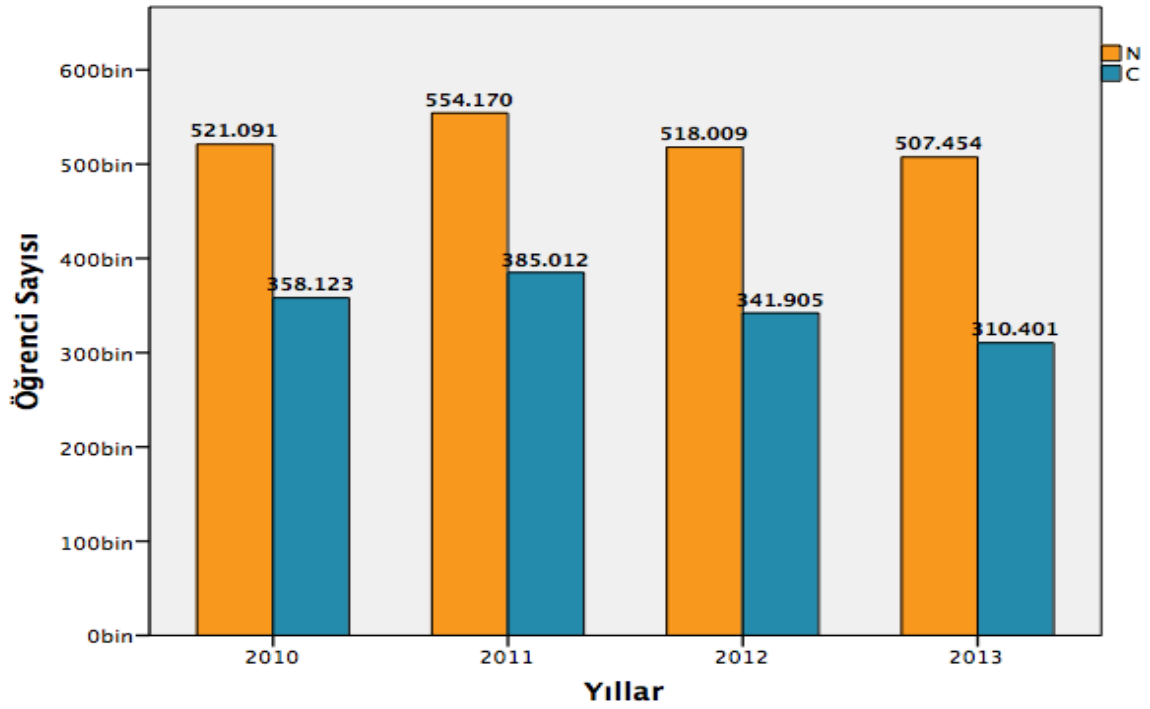
biçiminde yazılabilir. Eşitlik 4.2.2’de; X_1, X_2, X_3, X_4, X_5 ve X_6 sırasıyla Doğu Anadolu, İç Anadolu, Karadeniz, Akdeniz, Ege ve Marmara bölgelerini; Z_7, Z_8 ve Z_9 sırasıyla

2011, 2012 ve 2013 yıllarını göstermektedir. Eşitlikte, bölgeler içinde Güneydoğu Anadolu yıllar arasında ise 2010 referans kategorisi olarak alınmıştır.

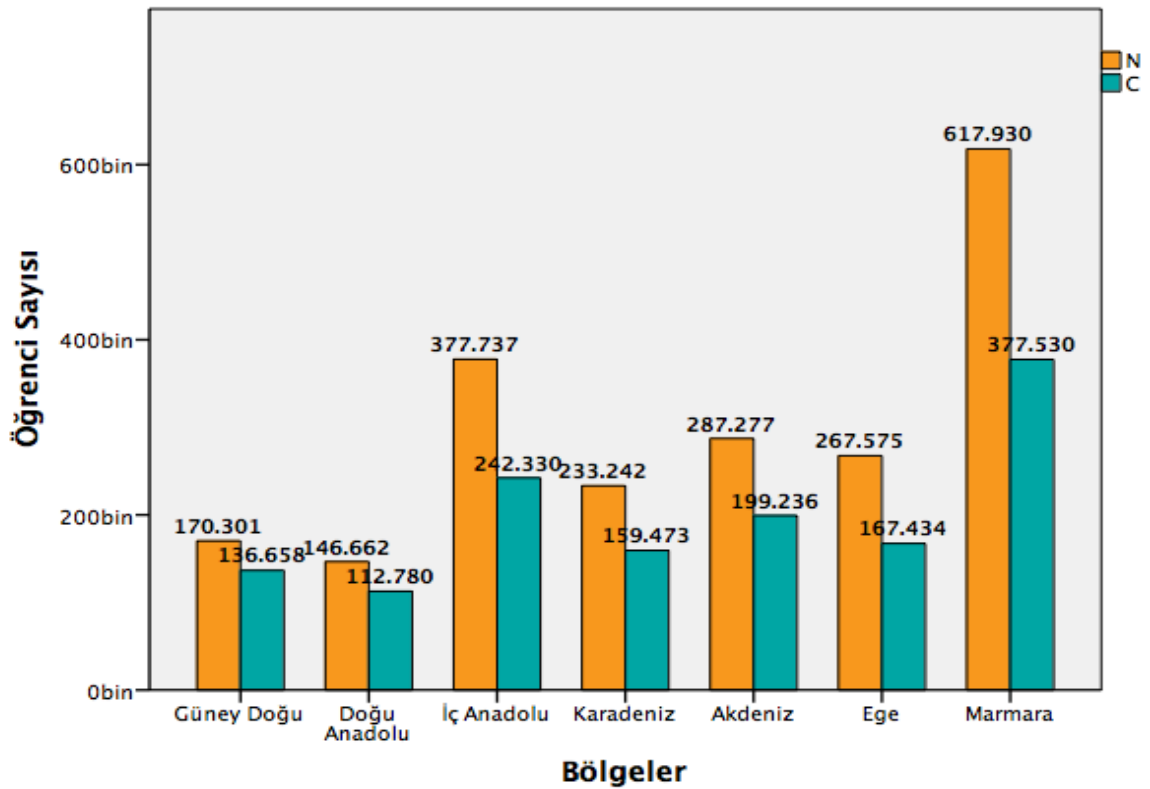
Çizelge 4.2.2.1.Yıllara ve bölgelere göre herhangi bir lisans programına yerleşemeyen öğrenciler

Bölgeler	Yıllar							
	2010		2011		2012		2013	
	n	c	n	c	n	c	n	c
Güney	35291	30019	47477	39132	43882	34844	43651	32663
doğu								
Doğu	36835	29181	39000	31261	35062	26958	35765	25380
Anadolu								
İç	94961	63436	98430	66423	93289	59475	91057	52996
Anadolu								
Karadeniz	60067	42412	61517	43911	56459	38442	55199	34708
Akdeniz	72393	51893	75244	54446	70433	48566	69207	44331
Ege	68504	43829	71388	46506	65227	40042	62456	37057
Marmara	153040	97353	161114	103333	153657	93578	150119	83266

Yıllara ve bölgelere göre YGS’de herhangi bir programa yerleşemeyen öğrencilerin dağılımları şekil 4.2.3 ve şekil 4.2.4’de verilmiştir.



Şekil 4.2.3. Yıllara göre YGS sınavında herhangi bir programa yerleşemeyen öğrencilerin dağılımı.



Şekil 4.2.4. Bölgelere göre YGS sınavında herhangi bir programa yerleşemeyen öğrencilerin dağılımı.

Poisson log-regresyon modeli kullanılarak elde edilen en çok olabilirlik parameter tahminleri Çizelge 4.2.2’de verilmiştir.

Poisson log-regresyon modeli kullanılarak elde edilen en çok olabilirlik parameter tahminleri Çizelge 4.2.2’de verilmiştir. Eşitlik 4.2.2 esas alındığında, Çizelge 4.2.2.2’de verilen en çok olabilirlik parametre tahminleri kullanarak Poisson log-doğrusal regresyon model aşağıdaki gibi yazılabilir.

$$\log(p_{ij}) = -1.343 + 0.069 * Do\ddot{g}o\ddot{g}uanadu - 0.405 * i\c{c}anadolu - 0.279 * karadeniz - 0.319 * akdeniz - 0.358 * ege - 0.304 * marmara + 0.275 * yu_2011 + 0.410 * yu_2012 + 0.501 * yu_2012 \quad (4.2.2)$$

Bu Poisson log-doğrusal regresyon modelini ayrıca,

$$p_{ij} = \exp \left(\begin{array}{l} -1.343 + 0.069 * Do\ddot{g}o\ddot{g}uanadu - 0.405 * i\c{c}anadolu - 0.279 * karadeniz \\ - 0.319 * akdeniz - 0.358 * ege - 0.304 * marmara + 0.275 * yu_2011 + \\ 0.410 * yu_2012 + 0.501 * yu_2012 \end{array} \right)$$

biçiminde de yazmak mümkündür.

Bölgeler içinde Güneydoğu Anadolu Bölgesi ve yıllar arasında ise 2010 yılı referans kategorisi olarak alınmıştır. Bu nedenle eşitlik 4.2.2’de bu referans kategorilerine yer verilmemiştir. Çizelge 4.2.2.2 ve eşitlik 4.2.2 kullanılarak referans kategorilerine göre bölgeler ve yıllar bazında, YGS sınavında lisans programına yerleşemeyen öğrencilerin relative risklerinin (RR) hesaplanması bölgeler ve yıllar için aşağıda verilmiştir.

Güneydoğu Anadolu Bölgesi referans alındığında diğer bölgelerin RR tahmin değerleri sırasıyla;

Güneydoğu Anadolu Bölgesindeki öğrencilerin, YGS sınavında herhangi bir programa yerleşememe relatif riski Doğu Anadolu Bölgesindeki öğrencilere göre 1.046 kat daha fazladır.

$$RR_{doguanadolu \text{ vs } guneydogu} = e^{-0.045} = 0.956 \text{ ve } \frac{1}{0.956} = 1.046$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında herhangi bir programa yerleşememe relatif riski İç Anadolu Bölgesindeki öğrencilere göre 1.253 kat daha fazladır.

$$RR_{icanadolu \text{ vs } guneydogu} = e^{-0.226} = 0.798 \text{ ve } \frac{1}{0.798} = 1.253$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında herhangi bir programa yerleşememe relatif riski, Karadeniz Bölgesindeki öğrencilere göre 1.176 kat daha fazladır.

$$RR_{\text{karadeniz vs güneydogu}} = e^{-0.163} = 0.850 \text{ ve } \frac{1}{0.850} = 1.176$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında herhangi bir programa yerleşememe relatif riski, Akdeniz Bölgesindeki öğrencilere göre 1.160 kat daha fazladır.

$$RR_{\text{akdeniz vs güneydogu}} = e^{-0.148} = 0.862 \text{ ve } \frac{1}{0.862} = 1.160$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında herhangi bir programa yerleşememe relatif riski Ege Bölgesindeki öğrencilere göre 1.287 kat daha fazladır.

$$RR_{\text{ege vs güneydogu}} = e^{-0.252} = 0.777 \text{ ve } \frac{1}{0.777} = 1.287$$

Güneydoğu Anadolu Bölgesindeki öğrencilerin YGS sınavında herhangi bir programa yerleşememe relatif riski Marmara Bölgesindeki öğrencilere göre 1.316 kat daha fazladır.

$$RR_{\text{marmaraz vs güneydogu}} = e^{-0.275} = 0.760 \text{ ve } \frac{1}{0.760} = 1.316$$

2010 yılı referans alındığında diğer yılların RR tahmin değerleri sırasıyla;

2010 YGS sınavında öğrencileri herhangi bir programa yerleşememe relatif riski 2011 yılındaki öğrencilere göre 1.007 kat daha fazladır.

$$RR_{\text{2011 vs 2010}} = e^{-0.007} = 0.993 \text{ ve } \frac{1}{0.993} = 1.007$$

2010 YGS sınavında öğrencileri herhangi bir programa yerleşememe relatif riski 2012 yılındaki öğrencilere göre 1.044 kat daha fazladır.

$$RR_{\text{2012 vs 2010}} = e^{-0.043} = 0.958 \text{ ve } \frac{1}{0.958} = 1.044$$

2010 YGS sınavında öğrencileri herhangi bir programa yerleşememe relatif riski 2013 yılındaki öğrencilere göre 1.127 kat daha fazladır.

$$RR_{\text{2013 vs 2010}} = e^{-0.120} = 0.887 \text{ ve } \frac{1}{0.887} = 1.127$$

Çizelge 4.2.2.2. Yıllara ve bölgelere göre YGS sınavında lisans programına yerleşemeyen öğrenciler için en çok olabilirlik parametre tahminleri

Parametreler	Tahmin	S. hata	%95 Wald güven aralığı		Wald testi	p-değeri
Sabit	-0.182	0.003	-0.187	-0.175	3409.830	<0.0001
Güneydoğu			Referans kategorisi			
Doğu Anadolu	-0.045	0.004	-0.053	-0.070	123.960	<0.0001
İç Anadolu	-0.226	0.003	-0.233	-0.219	4458.140	<0.0001
Karadeniz	-0.163	0.004	-0.170	-0.156	1965.150	<0.0001
Akdeniz	-0.148	0.004	-0.155	-0.141	1777.190	<0.0001
Ege	-0.252	0.004	-0.259	-0.245	4772.950	<0.0001
Marmara	-0.275	0.003	-0.281	-0.268	7554.640	<0.0001
2010 yılı			Referans kategorisi			
2011 yılı	0.007	0.002	0.002	0.001	8.640	0.003
2012 yılı	-0.043	0.002	-0.048	-0.039	328.890	<0.0001
2013 yılı	-0.120	0.003	-0.125	-0.116	2410.220	<0.0001

YGS sınavında herhangi bir programa yerleşemeyen öğrenciler göre bölgeler ve yıllar için pseudo- R^2 değerlerini hesaplamak için elde edilen log-oran (log ratio) istatistikleri sonuçları Çizelge 4.2.2.3’de verilmiştir.

YGS sınavında 180 puan ve üstü alıp fakat herhangi bir programa yerleşemeyen öğrencilere göre bölgeler ve yıllar için pseudo- R^2 değerlerini hesaplamak için elde edilen log-oran (log ratio) istatistikleri sonuçları Çizelge 4.2.2.3’de verilmiştir. Bu çizelgede,

Bölgeler için pseudo R^2 değeri;

$$pseudo \quad R^2 = \frac{14579.230 - 3568.201}{14579.230} = 0.755 = \%75.5$$

Yıllar için pseudo R^2 değeri;

$$pseudo \quad R^2 = \frac{14579.230 - 119.784}{14579.230} = 0.992 = \%99.2$$

Bu sonuçlara göre, YGS sınavında 180 puan ve üstü alıp fakat herhangi bir programa yerleşemeyen öğrencilerin değişimi bakımından bölgeler arası farkın açıkladığı kısım %75.5 iken, yıllar arası farklılığın açıkladığı kısım ise %99.2 olarak saptanmıştır.

Çizelge 4.2.2.3. Yıllara ve bölgelere göre YGS sınavında yerleşemeyen öğrenciler için log- oran testi sonuçları

Değişkenler	Devians	Sd	Ki-kare istatistiği	p-değeri
Sabit	14579.230	-	-	-
Bölgeler	3568.201	6	11101.400	<0.0001
Yıllar	119.784	3	3448.420	<0.0001

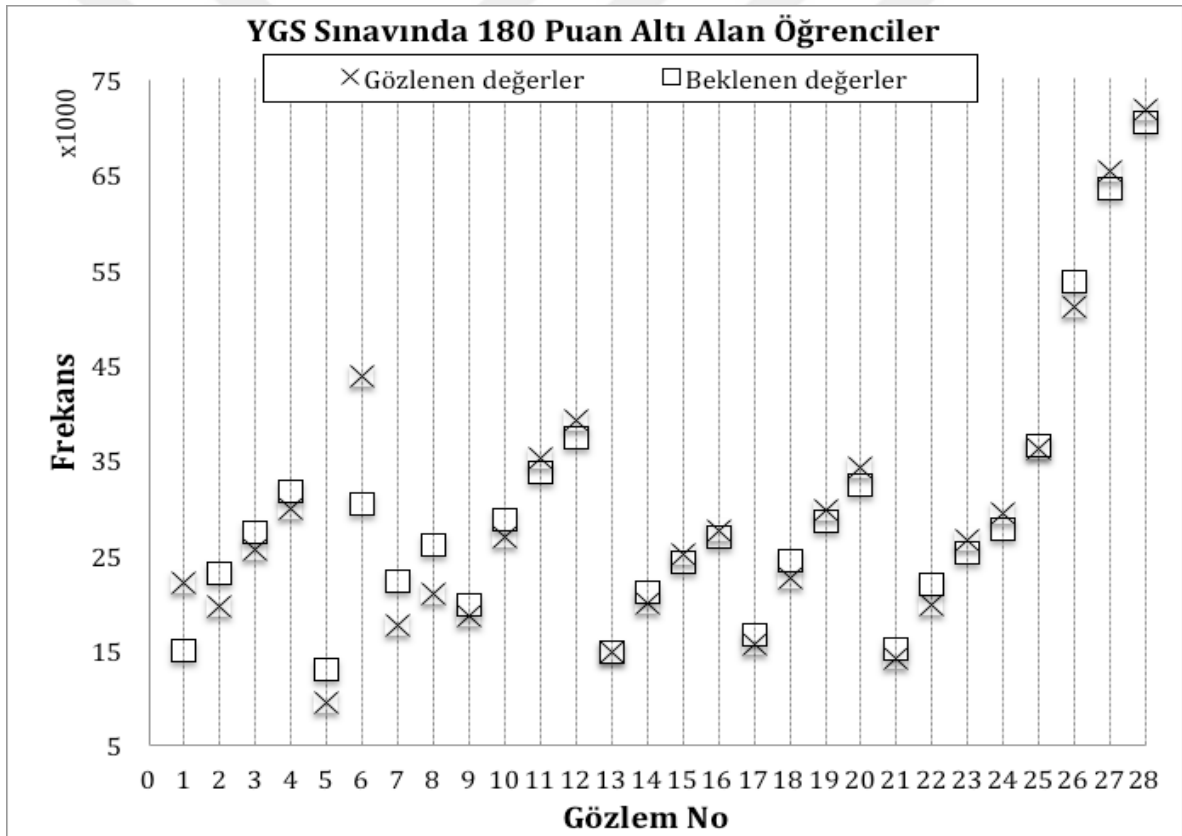
Sd: serbestlik derecesi

Yıllara ve bölgelere göre YGS sınavında 180 altı puan alan öğrenciler ile herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekanslar değerleri Çizelge 4.2.2.4’de verilmiştir.

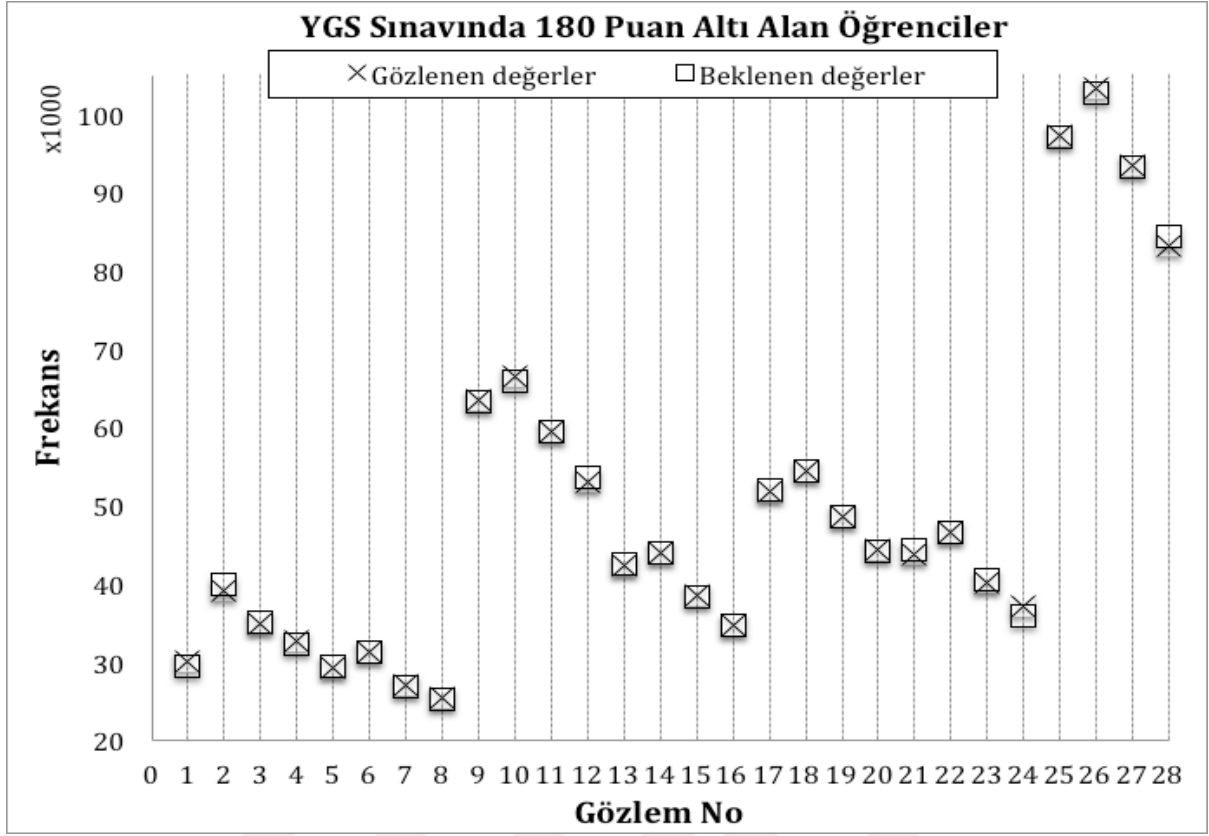
Çizelge 4.2.2.4. Yıllara ve bölgelere göre YGS sınavında 180 altı puan alan öğrenciler ile herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekanslar

YGS sınavında 180 altı puan alan öğrenciler		Herhangi bir programa yerleşemeyen öğrenciler	
Gözlenen değerler	Beklenen değerler	Gözlenen değerler	Beklenen değerler
22008	14956.9	30019	29441
19518	23018.7	39132	39879
25587	27327.1	34844	35054
29855	31665.3	32663	32284
9505	12956.7	29181	29383
43738	30450	31261	31323
17636	22204.5	26958	26781
20861	26128.9	25380	25292
18539	19771.2	63436	63199
26889	28734	66423	65956
35118	33708	59475	59451
39078	37410.8	52996	53724
14761	14779	42412	42573
19896	21164.9	43911	43899
25011	24248.5	38442	38317
27470	26945.6	34708	34684
15576	16682.3	51893	52079
22579	24417.9	54446	54501
29687	28612.3	48566	48518
34248	32377.4	44331	44138
14096	15075.9	43829	44421
19838	21916.2	46506	46608
26538	25239.9	40042	40501
29440	27680	37057	35904
51031	53787.3	103333	102845
65330	63566.7	93578	93283
71801	70545	83266	84375

Yıllara ve bölgelere göre YGS sınavında 180 altı puan alan öğrenciler ile herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekansların uyumu Şekil 4.2.5 ve Şekil 4.2.6’da verilmiştir. Her iki grafiği bakıldığında, YGS sınavında 180 altı puan alan öğrenciler için gözlenen ve beklenen frekansların uyumunun daha iyi olduğu saptanmıştır. Her iki grafikte de hem gözlenen frekanslar hem de beklenen frekansların dağılımı doğrusal değildir. Bu tip veriler Poisson dağılımı gösterdiklerinden dolayı bunların hatalarının dağılımı da Poisson dağılımı gösterir. Bu nedenle hem YGS sınavında 180 altı puan alan öğrencilerin hem de herhangi bir programa yerleşemeyen öğrencilerin gözlemlerine ait hatalar doğrusal olmayan bir forma sahip oldukları belirlenmiştir.



Şekil 4.2.5. YGS sınavında 180 altı puan alan öğrenciler için gözlenen ve beklenen frekansların uyumu.



Şekil 4.2.6. YGS sınavında 180 üstü puan alıp ancak herhangi bir programa yerleşemeyen öğrenciler için gözlenen ve beklenen frekansların uyumu.

4. TARTIŞMA VE SONUÇ

Bu çalışmada, bölgelere ve yıllara göre iki yönlü hazırlanan tabloda, “YGS’de 180 altında puan alma” relatif riskini ve “180 ve üzeri puan alıp fakat herhangi bir programa yerleşmeyenlerin” relatif riskini tahmin etmek için Poisson log-doğrusal regresyon modeli uygulanmıştır.

Bilindiği gibi GLM’i esas alan Poisson ve lojistik regresyon modellerinde, parametre tahminleri elde edilirken iki yönlü kategori seviyelerinde biri referans kategorisi olarak alınmaktadır. Burada tüm yapılan analizlerde bölgeler arasında Güneydoğu Anadolu bölgesi ve yıllar arasında ise 2010 yılı referans kategorisi olarak alınmıştır. Eğer referans kategorisinin seçimi araştırmacı tarafından yapılmazsa, program son seviyeyi otomatik olarak seçmektedir (SAS, 2017).

YGS’de, öğrencilerin 180 altında puan alma bakımından; Güneydoğu Anadolu’daki öğrencilerin relatif riskinin en fazla olduğu bölge İç Anadolu (1.499 kat daha fazla) en az olduğu bölge ise Doğu Anadolu (0.934 kat daha fazla) olmuştur ve bunların tamamı istatistiksel olarak önemli bulunmuştur ($p<0.01$). İç Anadolu bölgesinden sonra, Güneydoğu Anadolu Bölgesinin en büyük relatif riski Ege Bölgesine karşı olmuştur (1.431 kat). Yıllar bazında bakıldığında, 2010 yılındaki relatif riskin en yüksek olduğu yıl 2011 (0.759 kat), en düşük olduğu yıl ise 2013 (0.606 kat) olduğu saptanmıştır ($p<0.01$).

YGS’de, öğrencilerin 180 ve üstü puan alıp fakat herhangi bir programa yerleşmeme bakımından; Güneydoğu Anadolu’daki öğrencilerin relatif riskinin en fazla olduğu bölge Marmara (1.316 kat daha fazla) en az olduğu bölge ise Doğu Anadolu (1.046 kat daha fazla) olmuştur ve bunların tamamı istatistiksel olarak önemli bulunmuştur ($p<0.01$). Marmara Bölgesinden sonra, Güneydoğu Anadolu Bölgesinin en büyük relatif riski ege bölgesine karşı olmuştur (1.287 kat). Yıllar bazında bakıldığında, 2010 yılındaki relatif riskin en yüksek olduğu yıl 2013 (1.127 kat), en düşük olduğu yıl ise 2011 (1.007 kat) olduğu saptanmıştır ($p<0.01$).

YGS’de, sadece türkçe, matematik, fen ve sosyal testlerinden 4’er net yapan öğrenciler 140 puan barajını aşarken, bu testlerden 10’ ar net çıkaran adaylar da 200 puan almaktadırlar. Bu açıdan bakıldığında, aslında bu sınavda 180 üstü bir puan almak

için ilgili testlerde 6 ya da 7 net bırakmak yeterli görünmektedir. Ancak buna rağmen yıllara göre bakıldığında 180 puan altı alıp başarısız olan öğrenci sayısında çok ciddi artışlar göstermektedir. Özellikle, referans kategorisi olarak alınan Güneydoğu Anadolu Bölgesi başta olmak üzere Doğu Anadolu Bölgesinde bu artış çok daha yüksektir. Hem sınavda başarısız olma hem de herhangi bir programa yerleşememenin bu bölgeler açısından nedenleri söz konusudur. Ailelerin sahip oldukları ekonomik, sosyal ve kültürel unsurlar bir bütün olarak bireylerin eğitim sürecinde kullanabilecekleri finansal kaynakları, buna bağlı olarak sağlanacak ek materyalleri, anne ve babanın okul hayatına ilgi ve desteğini etkilemektedir. Söz konusu faktörler ise gerek öğrencilerin akademik başarıları, gerekse de üniversite sonrasında elde edilecek kazançlar açısından önemlidir (Bahadır ve Özdemir, 2013; Çiftçi ve Çağlar, 2014). Özellikle matematik alanındaki başarısızlık göz önüne alındığında, ailelerin gelir ve eğitim düzeyleri, öğrencilerin özellikle matematik alanındaki başarıları açısından da oldukça önemlidir. Matematikte başarılı olan öğrencilerin gelecekte elde edecekleri yüksek gelir olasılığı da artmaktadır (Engin, 2009). Çocukların aile özelliklerinin öğrencilerin matematik, okuma becerisi ve bilim konularındaki akademik başarıları üzerindeki etkileri konusunda yapılan çalışmada özellikle eğitim ve gelir düzeylerinin daha düşük olduğu Doğu ve Güney Doğu Anadolu bölgelerinde öğrencilerin genel olarak akademik başarılarının diğer bölgelere göre daha düşük olduğu sonucu ortaya çıkmıştır (Caner ve Ökten, 2010).

Anne ve babasının eğitim ve gelir düzeyi düşük olan, eğitim sürecinde gerekli ilgi ve desteği alamayan öğrencilerin, üniversite sonrasında kazançları görece olarak daha fazla ancak kontenjanları daha az ve giriş puanları daha yüksek olan tıp, mühendislik, diş hekimliği ve hukuk gibi bölümleri kazanabilme olasılıklarının daha düşük olduğunu göstermektedir (Çiftçi ve Çağlar, 2014). Ailelerin eğitim ve ekonomik düzeyleri yaşadıkları bölgeyi, söz konusu bölgenin çevre, okul ve öğretmen özellikleri de öğrencilerin akademik başarısını etkileyebilmektedir. Ailelerdeki eğitim düzeyindeki artış ile birlikte, çocuk sayısında azalma olmaktadır. Bu durum, öğrencilerin okuldaki hayatını ve akademik başarıyı olumlu olarak etkilemektedir. Bu açıdan bakıldığında, bu durum Doğu ve Güney Doğu Anadolu, hatta Karadeniz bölgeleri için risk oluşturmaktadır (Ekinci, 2011).

YGS’de, öğrencilerin 180 altında puan alma bakımından; yılların Pseudo- R^2 değeri (%76.8) bölgelere ait Pseudo- R^2 değerinden (%33.6) oldukça yüksek çıktığı

saptanmıştır ($p<0.01$). Bu açıdan bakıldığında yılların modelde daha etkili olduğu belirlenmiştir.

YGS’de, öğrencilerin 180 ve üzeri puan alıpta herhangi bir programa yerleşmeme bakımından; yılların Pseudo- R^2 (%99.2) değeri bölgelere ait Pseudo- R^2 değerinden daha yüksek çıktığı saptanmıştır ($p<0.01$). Bununla birlikte, bölgelere ait Pseudo- R^2 değerinin de azımsanmayacak kadar yüksek olduğu söylenebilir (%75.5). Buna rağmen, yılların modelde daha etkili olduğu belirlenmiştir.

Yukarıda belirtildiği gibi Pseudo- R^2 değerini, doğrusal regresyonda elde edilen belirleme katsayısı gibi yorumlanmamaktadır. Pseudo- R^2 kategori seviyelerinden hangisinin daha iyi olduğunu belirlemek için kullanılmaktadır (Agresti, 1997; Cameron ve Trivedi, 1998; Beh ve Davy, 2004; Eye ve Mun, 2013).

İki ya da daha çok yönlü çapraz sınıflandırılmış tabloların analizi sıklıkla yapılmaktadır. Özellikle ki-kare analizi iki yönlü tabloların analizinde kullanılmaktadır. İki yönlü tabloların seviyesi 2X2 den büyük olduğunda ki-kare testi yetersiz kalmaktadır. Şöyle ki, bu tip tablolarda sıfır hipotezi reddedildiğinde, farklılığın hangi hücreden kaynaklandığı ki-kare ile doğrudan tespit edilememektedir. Ancak genel bir farklılık olduğu söylenebilir. Bu duruma alternatif olarak uyum analizi kullanılabilir. Uyum analizi sadece görsel bir grafik sunmaktadır. Yani, herhangi bir hipotez testi yapmamaktadır (Eye ve Schuster, 2000; Eye ve Mun, 2003; Long ve Freese, 2006; Mair ve Eye, 2007; Eye ve ark., 2009).

İki yönlü tablolarda daha geniş yorumlara ulaşmak için log-doğrusal modeller kullanılmaktadır. Tablolardaki hücrelerdeki gözlemler sayılar olarak elde edildiğinden dolayı Poisson dağılımı gösterdiğinden dolayı, modeli doğrusallaştırmak için log dönüşümü uygulanır. Bu aynı zamanda Poisson log-doğrusal modeller olarak adlandırılmaktadır. İki yönlü tablonun hücrelerindeki gözlemler bağımlı değişken ve diğer değişkenler bağımsız değişken olarak ele alındığında log-doğrusal model, Poisson log-doğrusal regresyon modeline dönüşmektedir. Analiz ve yorumlama aşamasında iki model arasında farklılıklar mevcuttur (Stokes ve ark., 2000; Stata, 2012; SAS, 2017). Çünkü Poisson log-doğrusal regresyon modelinde hem bağımsız değişkenler için parametre tahmini elde edilirken hem de seçilen referans kategorisine göre diğer kategorilerin relatif risk tahminlerini elde etmek mümkündür.

KAYNAKLAR

- Agresti, A., 1997. **Categorical Data Analysis**. John and Wiley & Sons, Incorporation, New Jersey, Canada.
- Bahadır, E., Özdemir, A. Ş., 2013. Öğretmen adaylarının akademik başarılarının Sınıflandırılmasında Lojistik Regresyon Analizi Yaklaşımı, *Turkish Studies-International Periodical For The Languages, Literature and History of Turkish or Turkic*, **8** (8): 203-214.
- Beh, E. J., Davy, P. J., 2004. A non-iterative alternative to ordinal log-linear models. *Journal of Applied Mathematics and Decision Sciences*, **8**: 67-86.
- Beh, E. J., Farver, T. B., 2009. An evaluation of non-iterativemethods for estimating the linear-by-linear parameter of ordinal log-linear models. *Australian and New Zealand Journal of Statistics*, **51**: 335-352.
- Bishop, Y. M. M., Fienberg, S. E., Holland. P. W., 1975. **Discrete Multivariate Analysis**. MIT Press, Cambridge, MA.
- Box, G. E. P., Hunter, J. S., Hunter, W. G., 2005. **Statistics for Experimenters: Design Innovation, and Discovery** (2nd ed.). Wiley, Hoboken, NJ.
- Brian, C., 2017. **exactLoglinTest: Monte Carlo Exact Tests for Log-linear models, 2006. R package version 1.3.6**, <http://www.biostat.jhsph.edu/bcaffo>. (Erişim Tarihi: 20.07.2017).
- Breslow, N. E., 1996. Statistics in epidemiology: The case-control study. *Journal of the American Statiscal Association*, **91**: 14-26.
- Cameron, A. C., Trivedi, P. K., 1998. **Regression analysis of Count Data**. Cambridge University Press, New York.
- Caner, A., Ökten, C., 2010. Risk and career choice: Evidence from Turkey. *Economics of Education Review*, **29**: 1060–1075.
- Christensen, R., 1997. **Log-linear Models and Logistic Regression**. (2nd ed.). Springer, New York.
- Consul, P., Famoye, F., 1992. Generalized Poisson regression model. *Comunucation and Statistics Theory Methods*, **21** (1): 89-109.
- Cox, R., 1983. Some remarks on overdispersion. *Biometrika*, **70**: 269-274.
- Çiftçi, C., Çağlar, A., 2014. Ailelerin sosyo-ekonomik özelliklerinin öğrenci başarısı üzerindeki etkisi: Fakirlik kader midir? *International Journal of Human Sciences*, **11**(2): 223-231.
- DeCoster, J., Iselin, A. M., Gallucci, M., 2009. A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, **14**: 349-366.
- Dean, C. B., 1992. Testing for overdispersion in Poisson and binomial regression models. *Journal of American Statistical Association*, **87**(418): 453-460.
- Dobson, A., 2002. **An Introduction to Generalized Linear Models**. Chapman & Hall, New York.
- Ekinci, C. E., 2011. Bazı sosyoekonomik etmenlerin Türkiye’de yükseköğretime katılım üzerindeki etkileri. *Eğitim ve Bilim*, **36** (160): 281-297.
- Engin, C., 2009. Factors influencing the academic achievement of the Turkish urban poor. *International Journal of Educational Development*, **29**: 17-29.
- Everitt, B. S., Hothorn., T. A., 2006. **Handbook of Statistcal Analyses Using R**. Chapman & Hall, Boca Raton, FL.

- Evers, M., Namboodiri, N. K., 1978. On the design matrix strategy in the analysis of categorical data. *Sociological Methodology*, **10**: 86-111.
- Eye, V. A., Spiel, C., 1996. Standard and non-standard log-linear symmetry models for measuring change in categorical variables. *The American Statistician*, **50**: 300-305.
- Eye, V. A., Schuster, C., 2000. Log-linear models for rater agreement. *Multiciencia*, **4**: 38-56.
- Eye, V. A., Mun, E. Y., 2003. Characteristics of measures for 2X2 tables. *Understanding Statistics*, **2**: 243-266.
- Eye, V. A., Gardiner, J.C., 2004. Locating deviations from multivariate normality. *Understanding Statistics*, **3**: 313-331.
- Eye, V. A., Mair, P., 2008. A functional approach to configural frequency analysis. *Austrian Journal of Statistics*, **37**:161-173.
- Eye, V. A., Mun, E. Y., Bogat, G.A., 2009. Temporal patterns of variable relationships in person-oriented research: Prediction and auto –association models of Configural Frequency Analysis. *Applied Developmental Science*, **13**: 172-187.
- Eye, A. V., Mun, E. Y., 2013. *Log-Linear Modeling*. John and Wiley & Sons, Inc., Hoboken, New Jersey, Canada.
- Frome, E. D., Kutner, M. H., Beauchamp, J. J., 1973. Regression analysis of Poisson-distributed data. *Journal of American Statistical Association*, **68** (344): 935-940.
- Galindo-Garre F., Vermunt, J. K., 2004. The order-restricted association model: Two estimation algorithms and issues in testing. *Psychometrika*, **69**: 641-654.
- Gardiner, J., Pathak, D., Indurkha, A., 1999. Power calculations for detecting interaction in stratified 2 X 2 tables. *Statistics and Probability Letters*, **41**: 267-275.
- Glück, J., Von, A., 2000. Eye. Including covariates in configural frequency analysis. *Psychologische Beitrage*, **42**: 405-417.
- Goodman, L. A., 1979. Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**: 537-552.
- Hand, D. J., Vinciotti, V., 2003. Local versus global models for classification problems: Fitting models where it matters. *The American Statistician*, **57**: 124-131.
- Havranek, T., Lienert, G. A., 1984. Local and regional versus global contingency testing. *Biometrical Journal*, **26**: 483-494.
- Hilbe, J. M., 2007. *Negative Binomial Regression*. Cambridge, UK.
- Krauth, J., 2005. Paradoxes in multidimensional contingency tables: What does this mean for CFA?, *Psychology Science*, **47**: 304-314.
- Lawal, B., 2003. *Categorical Data Analysis with SAS and SPSS Applications*. Erlbaum, Mahwah, NJ.
- Lawles, J. F., 1987. Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, **15**(3): 209-225.
- Lerner, R. M., Eye, V. A., Lerner, J. V., Lewin-Bizan, S., Bowers, E.P., 2010. The meaning and measurement of thriving: A view of the issues. *Journal of Youth and Adolescence*, **39**: 707-719.
- Loken E., 2010. *Person-specific analysis of the dynamics of weight change*. In P.C.M. Molenaar and K. M. Newell, editors, Individual Pathways of Change: Statistical

- Models for Analyzing Learning and Development, pages 9-22. American Psychological Association, Washington, DC.
- Long, J. S., Freese, J., 2006. **Regression Models for Categorical Dependent Variable Using Stata**. A Stata Pres Publication, StataCorp LD Collage Station, Texas, USA.
- Luo, J., Qu, Y., 2013. Analysis of hypoglycemic events using negative binomial models. **Pharmalogical Statiscs**, **12**(4): 233-42.
- Luo, J., Qu., Y., 2015. Estimation of group means when adjusting for covariates in generalized linear models. **Pharmalogical Statatics**, **14**(1): 56-62.
- Mair, P., Eye, V. A., 2007. Application scenarios for nonstandard log-linear models. **Psychological Methods**, **12**: 139-156.
- Meiser, T., 1996. Loglinear Rasch models for the analysis of stability and change. **Psychometrika**, **61**: 629-645.
- McCullagh, P., Nelder, J. A., 1989. **Generalized Linear Models**. London: Chapman and Hall.
- Nelder, J. A., Wedderburn, R. W. M., 1972. Generalized linear models. **Journal of Royal. Statistics Society, A**, **135**.
- SAS, 2017. **SAS/Stat Software Hangen and Enhanced**, SAS institute incorporation, USA.
- STATA, 2012. **Statistical Software Program**, USA.
- Stokes, M. E., Davis, C. S., Koch, G. G., 2000. **Categorical data analysis using the SAS system**. John and Wiley & Sons Incorporation, USA, 724.
- Upton, G. J. G., 2017. **Categorical Data Analysis by example**. John and Wiley, Inc., Hoboken, Canada.
- Wang, W., Famoye, F., 1997. Modeling household fertility decisions with generalized Poisson regression. **Journal of Population Economy**, **10** (3): 273-83.
- Wang, X. G., Hankey, S., Hoff, K., 2014. Estimating mixed-mode urban trail traffic using negative binomial regression models. **Jornal of Urban Planned Develope**, **140**(1): 512-519.
- Wulu, J. T., Singh, K. P., Famoye, F., McGwin, G., 2002. Regression analysis of count data. **Journal of the Indian Society of Agricultural Statistics**, **55**: 220-231.
- Yeşilova, A., Kaki, B., Kasap, İ., 2007. Sıfır değer ağırlıklı sayıma dayalı olarak elde edilen bağımlı değişkenin modellenmesinde kullanılan regresyon yöntemler. **İstatistik Araştırma Dergisi**, **5** (1): 1-9.
- Yeşilova, A., Kaydan, B., Kaya, Y., 2010. Modelling insect-egg data with excess zeros using zero-inflated regression models. **Hacettepe Journal of Mathematics and Statistics**, **39**(2): 273-282.
- Yeşilova, A., Denizhan, E., 2016. Modeling mite counts Using Poisson and negative binomial regressions. **Fresenius Environmental Bulletin**, **25**(11): 5062-5066.

ÖZGEÇMİŞ

1989 yılında Diyarbakır'ın Ergani ilçesinde doğdu. İlk, orta ve lise öğrenimini Ergani'de tamamladıktan sonra 2008 yılında Yüzüncü Yıl Üniversitesi Fen – Edebiyat Fakültesi Sosyoloji Bölümü 'nü kazandı. 2012 yılında bu bölümden mezun oldu. 2012 yılında Yüzüncü Yıl Üniversitesi Zootekni/ Biyometri ve Genetik Ana Bilim Dalında Yüksek Lisans yapmaya hak kazandı. üniversiteden mezun olduktan sonra Rehber Öğretmen olarak göreve başladı, 2 yıldır Özel Ergani Açık Koleji'nde Rehber Öğretmen olarak göreve devam etmektedir.

