KANTİL REGRESYON VE NEGATİF BİNOMİAL REGRESYON İLE İLLERDE KULLANILAN İLAÇ SAYISINA ETKİ EDEN FAKTÖRLERİN İNCELENMESİ

Kübra ELMALI

Yüksek Lisans Tezi
Ekonometri Anabilim Dalı
Yrd. Doç. Dr. Ömer ALKAN
2014

T.C.
ATATÜRK ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI

Kübra ELMALI

KANTİL REGRESYON VE NEGATİF BİNOMİAL REGRESYON İLE İLLERDE KULLANILAN İLAÇ SAYISINA ETKİ EDEN FAKTÖRLERİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

TEZ YÖNETİCİSİ
Yrd. Doç. Dr. Ömer ALKAN

ERZURUM - 2014

T.C.
ATATÜRK ÜNİVERSİTESİ
SOSYAL BİLİMLERİ ENSTİTÜSÜ

TEZ BEYAN FORMU

11 /07/ 2014

SOSYAL BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

BİLDİRİM

Atatürk Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum " Kantil Regresyon ve Negatif Binomial Regresyon ile İllerde Kullanılan İlaç Sayısına Etki Eden Faktörlerin İncelenmesi " adlı tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Atatürk Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

Lisansüstü Eğitim-Öğretim yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

☑☐Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐☐Tezim/Raporum sadece Atatürk Üniversitesi yerleşkelerinden erişime açılabilir.

☐☐Tezimin/Raporumun ..... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

11/07/2014

Kübra ELMALI

K. Elmalı

T.C.
ATATÜRK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

TEZ KABUL TUTANAĞI

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Yrd. Doç. Dr. Ömer ALKAN danışmanlığında, Kübra ELMALI tarafından hazırlanan bu çalışma 11 / 07 / 2014 tarihinde aşağıdaki jüri tarafından Ekonometri Anabilim Dalı'nda Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan : Prof. Dr. Erkan OKTAY İmza:

Jüri Üyesi : Doç. Dr. Bekir ELMAS İmza:

Jüri Üyesi : Yrd. Doç. Dr. Ömer ALKAN İmza:

Yukarıdaki imzalar adı geçen öğretim üyelerine aittir. ..... / ..... / .....

Prof. Dr. Mustafa YILDIRIM
Enstitü Müdürü

İÇİNDEKİLER

DÖRDÜNCÜ BÖLÜM

UYGULAMA

SONUÇ VE DEĞERLENDİRME

ÖZET

YÜKSEK LİSANS TEZİ

KANTİL REGRESYON VE NEGATİF BİNOMİAL REGRESYON İLE İLLERDE KULLANILAN İLAÇ SAYISINA ETKİ EDEN FAKTÖRLERİN İNCELENMESİ

Kübra ELMALI

Tez Danışmanı: Yrd. Doç. Dr. Ömer ALKAN

2014, 83 Sayfa

Jüri: Yrd. Doç. Dr. Ömer ALKAN

Prof. Dr. Erkan OKTAY

Doç. Dr. Bekir ELMAS

Bu çalışmada, 2010-2011-2012 ve 2013 yılları esas alınarak illerin ilaç tüketiminde etkili olan faktörler incelenmiştir. Klasik regresyon varsayımlarının sağlanmaması durumunda çoklu doğrusal regresyona alternatif olarak sunulan robust regresyon yöntemlerinden olan "Kantil Regresyon" kullanılmış ve çeşitli kantil dilimleri esas alınarak analiz yapılmıştır.

Ayrıca, alternatif regresyon yöntemlerinden varyansın ortalamadan büyük olduğu aşırı yayılım durumunda kullanılan ''Negatif Binomial Regresyon'' analizleri de kullanılmıştır. Analiz sonucunda 2010-2011 yıllarında log-nüfus, log-hekim, log-eczacı ve deniz etkisi değişkenleri, 2012 yılında log-nüfus, log-hekim, log-eczacı, nüfus artışı ve deniz etkisi değişkenleri, 2013 yılında log-nüfus, log-eczacı, nüfus artışı ve deniz etkisi değişkenleri anlamlı bulunmuştur. Çeşitli karşılaştırma yöntemleri ile de negatif binomial regresyon analizinin kantil regresyon analizine göre daha etkin sonuç verdiği görülmüştür.

Anahtar Kelimeler: Kantil Regresyon, Poisson Regresyon, Negatif Binomial Regresyon, Stata programı.

ABSTRACT

MASTER'S THESIS

QUANTILE REGRESSION AND NEGATIVE BINOMIAL REGRESSION WITH THE NUMBER OF DRUGS USED IN CITIES TO EXAMINE THE FACTORS AFFECTING

Kübra Elmalı

Advisor: Assist. Prof. Dr. Ömer ALKAN

2014, 83 Pages

Jury: Assist. Prof. Dr. Ömer ALKAN (Advisor)
Prof. Dr. Erkan OKTAY
Assoc. Prof. Dr. Bekir ELMAS

In this study, it is investigated factors affecting on number of drug by taking base 2010-2011-2012 and 2013 years. Even if assumption is not provide it is used "Quantile Regression" which is robust regression by alternatively multilineer regression and it is analysis by various Quantile slice.

Alternatively, it is used ''Negative Binomial Regression'' which variance is greater than mean and uses overdispersion to offer more extensive regression. We found that log-population, log-physician, log-pharmacist and the marine influence variables in 2010-2011 are significant. In addition to the log-population, log-physician, log-pharmacist, population growth and the marine influence variables in 2012, log-population log-pharmacist, population growth and the marine influence variables in 2013 are found significant. It is seen that Negative Binomial Regression gives more effective result than Quantile Regression by various comparison methods.

Keywords: Quantile Regression, Poisson Regression, Negative Binomial Regression, Stata program.

KISALTMALAR DİZİNİ

QR : Kantil Regresyon
LAD : En küçük mutlak sapma
CDF : Kümülatif Dağılım Fonksiyonu
QDF : Kantil Yoğunluk Fonksiyonu
PDF : Olasılık Yoğunluk Fonksiyonu
MSE : Hata kareler ortalaması
MAD : Mutlak sapma ortalaması
LP : Doğrusal Programlama
PS : Poisson Regresyon
NB1 : Negatif Binomial Regresyon-1
NB2 : Negatif Binomial Regresyon-2
AIC : Akaike Bilgi Kriteri
BIC : Bayesian Bilgi Kriteri
MAE : Mutlak Hata Ortalaması

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

GRAFİKLER DİZİNİ

ÖNSÖZ

İlk olarak tez konusunun belirlenmesinden, tezin tamamlanmasına kadar benden hiçbir yardımını esirgemeyen tez danışmanım Sayın Yrd. Doç. Dr. Ömer ALKAN başta olmak üzere lisans eğitiminde danışmanım olan ablam Sayın Yrd. Doç. Dr. Ceren Sultan ELMALI'ya ve her zaman yanımda olan değerli aileme teşekkürü bir borç bilirim.

Erzurum-2014 **Kübra ELMALI**

GİRİŞ

Regresyon terimi ilk kez Francis Galton tarafından kullanılmış olup regresyon analizi çeşitli bilim dallarında yaygın bir şekilde kullanılmaktadır. Ekonometrik çalışmalarda en çok kullanılan araçlardan biri olan regresyon analizinde önemli olan fonksiyonel ya da kesin ilişkiler olmayıp istatistiksel ilişkilerdir ve değişkenler arasındaki ilişkiler matematiksel bir model ile ortaya konmaktadır. Regresyon analizinde her ne kadar bir değişkenin başka değişkenlere bağlılığıyla uğraşılsa da mutlaka bir nedensellik ilişkisi ifade etmez. Yani, mutlaka bağımsız değişkenin sebep ve bağımlı değişkenin sonuç olduğu anlamına gelmez.

Regresyon analizinin başlıca amaçları:

1. Bağımsız değişkenlerin verilen değerleri ile bağımlı değişkenin ortalama değerini tahmin etmek.
2. Bağımsız değişkenlerin, bağımlı değişken üzerinde önemli bir etkiye sahip olup olmadığını araştırmak.
3. Bağımsız değişkenlerin verilen değerleri ile bağımlı değişkenin ortalama değerini öngörmek veya gelecekte alacağı değeri tahmin etmek.

Regresyon analizi bazı varsayımlara dayanmaktadır. Bu varsayımlar arasında en önemli olanlarında biride bağımlı ve bağımsız değişken arasındaki ilişkinin fonksiyonel kalıbının biliniyor olmasıdır. Varsayımların sağlanmadığı durumlarda yapılmış olan tahminlerinde iyi bir tahmin olmaları mümkün değildir. Bu durumda alternatif regresyon modelleri gerekli olabilir.

Bu çalışmada ilk olarak robust regresyon anlatıldı. En küçük mutlak sapma (LAD) regresyon yönteminden bahsedildi. Basit ve çoklu doğrusal LAD regresyon için algoritmaları verildi.

İkinci bölümde, kantil kavramı, kantil fonksiyonu ve kantil regresyon kavramları hakkında genel bilgiler verildi. Parametrik regresyon modelleri normal dağılım varsayımını gerektirdiğinden dağılımın normal olmadığı durumlarda kullanılan alternatif regresyon modellerinden kantil regresyon modeli anlatıldı.

Üçüncü bölümde, poisson regresyon analizinden bahsedildi. Poisson regresyon analizinde ortalama ve varyansın eşitliği gerektiğinden ve veriler için uygun

olmadığından varyansın ortalamadan büyük olma şartını sağlayan negatif binomial regresyon uygulandı.

Son olarak da uygulama ve sonuç kısmı yer almaktadır. Uygulama kısmında bağımlı değişken olarak 2010-2011-2012 ve 2013 yıllarını esas alan Türkiye'de illerde kullanılan ilaç sayısı esas alındı. Böylece ilaç tüketimine etki eden faktörlerin ne derece anlamlı olduğu araştırıldı.

Bağımsız değişken olarak ise özel ya da devlet ayırt etmeksizin hastane sayısı, eczacı sayısı, hekim sayısı, illerin nüfusu, illerin nüfus artış hızı ve deniz etkisi alındı. Hastane sayısı ve nüfus artış hızı hariç diğer değişkenlerin logaritması alınarak çoklu doğrusal bağlantı azaltılmaya çalışıldı. Ayrıca deniz etkisi de kukla değişken olarak alındı ve denize kıyısı olan illere 1 denize kıyısı olmayan illere 0 verilerek analize dahil edildi. Bu verilere kantil ve negatif binomial regresyon yöntemleri uygulanarak hangi regresyonun daha etkili sonuç verdiği incelendi.

Son olarak ise negatif binomial regresyon ve kantil regresyon analizlerini karşılaştırmak için Tahminlerin Standart Hatası, Akaike Bilgi Kriteri, Hata Kareler Ortalaması ve Ortalama Mutlak Hata yöntemleri kullanıldı. Bu yöntemlerle en etkin sonucun hangi analiz ile elde edildiği tespit edildi.

LİTERATÜR TARAMASI

(Koenker ve Basett, 1978) kantil regresyon analizi alanında ilk çalışmayı yapmışlardır. Koşullu kantil fonksiyonlarının tahmin modeli için uygun bir yöntem önermişlerdir. Çalışmalarında örnek quantili kavramından yola çıkarak basit minimizasyonu problemini, lineer modeller için genelleştirmişler ve bunu "regresyon kantili" olarak adlandırmışlardır. Kantil Regresyonda tahmin ediciler hataların mutlak toplamlarını minimize etmektedir. Bu çalışmada, ayrıca regresyon kantillerinin ortak asimptotik dağılımları da oluşturulmuştur.

(Englin ve Shonkwiler, 1995) rekreasyonel yürüyüş parkurlarında sayma veri modellerini kullanarak uzun dönemli sosyal refah tahmini yapmışlardır. Yarı logaritmik fonksiyon ile poisson ve negatif binomial regresyon modellerini kullanarak tüketici rantını ve popülasyon ortalamalarını kullanarak potansiyel talebi hesaplamışlardır.

(Lingren, 1997) parametrik kantil fonksiyon tahminleri için yeni bir yöntem önermiştir. Kantil fonksiyonunun tahmini, simetrik olmayan L1 teknikleri ile birlikte sansür limitleri dağılımının ağırlıklı Kaplan-Meier tahmini olarak ele alınmıştır.

(Eide ve Showalter, 1998) "Test skoru kazanımı" durumunun koşullu dağılımında farklı noktalardaki okul kalitesi ve performans arasındaki ilişkinin farklı olup olmadığını standartlaştırılmış testlerde kantil regresyon yöntemini kullanarak tahmin etmişlerdir.

(McKean ve Taylor, 2000) Moskova'da nisan ve kasım aylarında rekreasyon gezisine ödeme istekliliğini ölçmek için poisson ve negatif binomial regresyon yöntemlerini uygulamışlardır. Bağımlı değişken olarak alana yapılan gezi sayısı, bağımsız değişken olarak ise gezi maliyeti, zamanı, alternatif olarak geçirilen zaman, yıllık hane halkı geliri ve ziyaretçinin yıllık serbest zamanı alınmıştır.

(Moons vd, 2001) seyahat zamanı ve maliyeti konusunda bir çalışma yapmıştır. Hesaplanan seyahat zamanı ve maliyeti coğrafi bilgi sistemi yazılımları ile algılanan seyahat zamanı ve maliyeti ise anket yoluyla elde edilmiştir. Seyahat maliyetini açıklamak için uzaklık, zaman, grup büyüklüğü, sosyal sınıf, öğrenim düzeyi, ziyaretin memnunluk düzeyi ve ziyaret süresi kullanılmıştır. Seyahat zamanı ve maliyet farkı arasındaki etkiyi belirlemek için yarı logaritmik, negatif binomial ve kesik negatif binomial modeller kullanılmıştır. Bu fark uzaklık ve ziyaret sıklığı ile ters orantılı olarak elde edilmiştir. Negatif binomial regresyon formunda olan maliyetin istatistiksel olarak anlamlı olduğu elde edilmiştir.

(Kan ve Tsai, 2004) obeziteye neden olan değişkenlerin etkilerini incelemişlerdir. Taiwan'dan alınan anket verileri temel alınmış ve Kantil Regresyon Tekniği kullanılarak analiz edilmiştir. Sonuçlar ilişkinin var olduğunu, kadın ve erkekler arasında BMI açısından farklı ilişki olduğunu göstermişlerdir.

(Baur ve ark, 2004) Atina'da dört gözlem kulesinden 1992-1999 yılları arasında alınan verilerle, ozon konsantrasyonun koşullu dağılımını etkileyen bağımsız değişkenlerin, farklı ozon düzeylerinde farklı derecede öneme sahip olduğunu göstermiştir.

(Sezgin ve Deniz, 2004) 1964-2000 yılları dönemi için Türkiye'de grev sayılarını etkileyen faktörleri poisson regresyon modeli aracılığıyla incelemişlerdir. Analiz

sonucunda verilerde aşırı yayılım gözlendiği için tahmin etkinliği artırılmak istendiğinden negatif binomial regresyon uygulanmıştır. Grevin belirleyicileri olarak ise işsizlik oranı sendikalaşma oranı, çalışan başına milli gelirin değişimi olarak belirlenmiştir.

(Koenker, 2005) kantil regresyonun özellikleri, asimptotik teori özellikleri, ağırlıklandırılmış kantil regresyon özellikleri, kantil regresyonun hesaplama yöntemleri ve belirsizlik durumundan son olarak da kantil regresyona ait sonuç çıkarımından bahsetmiştir. Kantil değerini değişkenin dağılımında yer alan ve dağılımı kendisinden büyük ve küçük olanlar diye ikiye bölen bir değer olduğunu belirtmiştir.

(Chen, 2005) özellikle uç noktaların önemli olduğu durumlarda kantil regresyon yönteminin kullanışlı olduğundan bahsetmiş ve halk sağlığı açısından çevresel çalışmalarda üst quantillerde hava kirliliğinin kritik seviyeye geldiği durumları örnek olarak göstermiştir.

(Saçaklı, 2005) En Küçük Kareler Regresyonu, En Küçük Mutlak Sapma Regresyonu, En Küçük Medyan Kare Regresyonu, M regresyon ve Rank Regresyonundan bahsedip bu alternatif regresyon modellerini karşılaştırmıştır. Ayrıca, kantil regresyon ‘un özellikleri vermiş ve kantil regresyon analizine yönelik bir uygulama yapmıştır.

(Yeşilova ve diğ, 2006) Norduz erkek kuzularının bazı kesikli üreme davranış özelliklerinin analizinde doğrusal olmayan regresyon modellerin kullanılması çalışmalarında genelleştirilmiş doğrusal modelleri esas alan poisson ve negatif binom regresyonları kullanarak genelleştirilmiş tahmin modelleri elde etmişlerdir. Çalışmada model uyumu için sapma ve Pearson χ^2 uyum ölçütleri kullanılmıştır. Elde edilen yayılım parametre değeri 1 değerine yakın olduğundan model uyumu açısından negatif binomial regresyonun daha iyi sonuç verdiği saptanmıştır. Bunun sonucu olarak her iki modeldeki parametre değerlerinin birbirinden farklı olduğu belirtilmiştir.

(Bilgiç ve Florkowski, 2007) ABD’de levrek avı talebi konusunda negatif binomial regresyon yöntemlerini ele almışlardır. Geziye katılma sıklığı ve geziye katılma kararını etkileyen faktörler incelenmiştir. Değişkenler olarak yaş, medeni durum, yaşın karesi, cinsiyet, ırk, öğrenim durumu, gezi maliyeti ve yakalanan balık sayısı, balık büyüklüğü ve balık sayısının karesi alınmıştır. Sıfır değerinin aşırılığı ve

heterojenlik durumu olmadığı için negatif binomial regresyon ile sağlıklı sonuçlar elde edilmiştir.

(Li ve ark, 2009) Çin Borsasında yer alan finansal olmayan 643 şirket örneklemi kullanarak şirketlere ait performans üzerinde hükümet iştirakçilerinin etkilerini değerlendirmişlerdir. Literatürdeki tartışmalı gözlem bulguları ve en küçük kareler regresyonu kısıtlamaları yüzünden kantil regresyon metodu benimsenmiş, hükümet iştirakçileri ile şirket performansları arasında önemli bir negatif ilişki olduğunu rapor etmişlerdir. Eldeki yeni bulguların koşullu ortalama regresyondan elde edilemediğini ve ilişki parametresinin performans değişkenlerinin dağılımında quantiller boyunca değiştiğini ele almışlardır.

(Meligkotsidou ve ark, 2009) risk faktörlerini kullanarak yatırım fonu getirilerinin koşullu quantillerle modellenmesi fikrini tanıtmışlardır. Kantil regresyon analizi, yatırım fonu getirisi ve risk faktörlerinin değişimi arasındaki ilişkinin koşullu getiri dağılımı boyunca nasıl değiştiğini anlamak için bir yol sunmuşlardır. En uygun risk faktörleri, farklı kantiller için tanımlanmış ve koşullu beklenen modelden elde edilenle karşılaştırılmıştır. Getirilerdeki faktör etkisindeki farklılığı kantiller ile elde etmişlerdir.

(Stifel ve Averett, 2009) ABD'de aşırı kilolu (obez) çocukların yaygınlığının son yirmi yılda çarpıcı biçimde arttığı ve bunun halk sağlığı problemlerine neden olduğu, üstelik fazla kilolu olmayan çocukların da ağırlığının artması problemini ele almışlardır. Bunun için Beden Kitle Endeksi (BMI) ile ölçülen ağırlık durumu ve aşırı kilo arasındaki ilişkiyi bulmak için kantil regresyon yöntemini kullanmışlardır. Sonuç olarak kantil regresyon yönteminin daha detaylı bilgi verdiğini göstermişlerdir.

(Çalışkan, 2009) 21 OECD ülkesinde yaşam süresi ve kullanılan ilaç arasındaki ilişkiyi incelemiştir. 1985-2002 yıllarını esas almış olup panel veri analizi yapmıştır. 40-60 ve 65 yaş grubunu esas almış olup bayan ve erkek olarak ayrı bir değerlendirme yapmıştır. Bağımsız değişken olarak toplum ve kişisel ilaç harcamalarını, meyve ve sebze tüketimini, 1000 kişi başına düşen doktor sayısını ve şehirleşme derecesini esas almışlardır. Sonuç olarak kadınlar ve erkekler arasında farklı anlamlılıklar elde edilmesinin yanı sıra doktor değişkeni, kişisel ilaç kullanımı gibi değişkenler de yaş gruplarında farklı önem düzeylerinde anlamlı bulunmuştur.

(Demirbağ ve Timur, 2011) Doğu Karadeniz Bölgesi'nde bir şehre bağlı Aile Sağlığı Merkezi'ne gelen 'Bir grup yaşlının ilaç kullanımı ile ilgili bilgi, tutum davranışlarını' değerlendirmişlerdir. Araştırma, 1 Eylül-1 Ekim 2010 tarihleri arasında yapılmıştır. Araştırmanın evrenini, karşılıklı iletişim sağlanabilen (işitme, görme, mental problemi olmayan) ve 65 yaş üstü yaşlılar oluşturmaktadır. Bu sürede seçilen Aile Sağlığı Merkezi'nde 165 yaşlı araştırmayı kabul etmiştir. Elde edilen veriler, hata kontrolleri, tablolar istatistiksel analizler paket program aracılığıyla değerlendirilmiş olup, istatistiksel analizinde aritmetik ortalama, sayı, yüzde hesaplamaları ve Ki-Kare testi kullanılmıştır.

(Kurtoğlu, 2011) kantil regresyon yöntemi ile kantil regresyonun özel bir hali olan En Küçük Mutlak Sapma (LAD) yöntemini ele almış ve bu yöntemlerle elde edilen sonuçları EKK regresyon yöntemi ile karşılaştırmıştır. Bu veri seti diyabet ile şişmanlık arasındaki ilişkiyi incelemektedir. Veri seti R paket programının *Hmisc* paketi ile çalışmaktadır. Veri setinde 19 değişken, 403 gözlem değeri bulunmaktadır.

(Ghitay ve diğ, 2012) çok değişkenli karışımlı poisson regresyon modellerini sınıflandırmak için maksimum olabilirlik tahminini kolaylaştırmada genel bir EM algoritması önermişlerdir. Çok değişkenli negatif binomial ve literatürde yeni bulunan çok değişkenli poisson ters guassian ve poisson log normal regresyon modelleri hakkında bilgi verilmektedir.

(Arı ve Önder, 2012) regresyon yöntemlerinden; Doğrusal regresyon, Lojistik regresyon, Negatif binom regresyon, Poisson regresyon, Temel bileşenler regresyonu, Probit regresyon, Ridge regresyonu ve Cox regresyon yöntemlerinin hangi durumlarda kullanılabileceği konusunda öneride bulunmuşlardır.

(Avşar, 2014) yerel sanayiye ait üretim ve ithalat verileri ile Türkiye'deki antidamping soruşturma sayısını etkileyen faktörler; analiz etmiştir. Yerel sanayinin toplam istihdamı, üretim miktarlarındaki ve ürettikleri ürünlerin ithalatındaki yüzdelik değişiklik açıklayıcı değişken olarak kullanılarak tahmin edilen negatif binomial regresyon sonuçlarına göre, yerel sanayilerin büyüklüğü, üretimlerindeki azalma ve iç piyasada ithalattan dolayı artan rekabetin soruşturma sayılarını arttırdığı bulunmuştur.

BİRİNCİ BÖLÜM

1.1. ROBUST REGRESYON

Çoklu doğrusal regresyon yöntemiyle elde edilen parametreler ile ilgili hipotez testleri yapabilmek ve güven aralığını belirleyebilmek için hatanın normal dağılıma uyması gerekmektedir. Hata paylarının normal dağılmaması durumunda, hipotez test sonuçları ve bulunan güven aralıkları hatalı sonuçlar vermektedir. Ayrıca, veri kümesinde sapan değerlerin olması durumunda da, çoklu doğrusal regresyon tahmin edicileri ve hipotez testlerinin sonuçları etkilenmekte ve böylece 'robust' olmaktadır. Varsayımlara bağlı olmayan, özellikle normallik varsayımına duyarsız yaklaşımlar "robust (sağlam)" olarak adlandırılmıştır.

İstatistikte "robust" kelimesi "varsayımlardan sapmalardan fazla etkilenmeyen" anlamında kullanılmaktadır. Robust tahminciler üzerine çalışmalar 1757'de Roger Joseph Boscovich tarafından En Küçük Mutlak Sapmalar Regresyonunun bulunmasıyla başlamış olup daha sonra Laplace tarafından yeniden ele alınmış ve güncellenerek kullanılmaya başlanmıştır (Nevitt ve Tam, 1998: 54-55).

Sapan değer, bir veri kümesinde gözlemlerin çoğunun sahip olduğu dağılıma veya modele uymayan gözlemler olarak ifade edilebilir (Barnet ve Levis, 1994:7).

Sapan değer içeren veri kümesinde varsayımların sağlanamamasından dolayı kurulan regresyon modelinden alınan sonuçlarda yanıltıcı olmaktadır Bu gibi durumlarda, sapan değerlerin etkisini azaltmak ve sapmalardan daha az etkilenmek için alternatif tahmin ediciler kullanılabilir (Goodal, 1983:211).

Çoklu doğrusal regresyon yönteminin güvenilir sonuçlar vermediği durumlarda daha güvenilir sonuçlar elde edebilmek amacıyla ortaya konulmuş alternatif regresyon yöntemleri önerilmektedir. Robust regresyon, bu alternatif yöntemlerden biridir.

1.1.1. En Küçük Mutlak Sapmalar Yöntemi (Least Absolute Deviation-LAD)

LAD tekniği çoklu doğrusal regresyon tekniğinden 50 yıl önce Boscovich tarafından ortaya atılmış daha sonra Laplace tarafından yeniden ele alınmış ve güncellenerek kullanılmaya başlanmıştır (Nevitt ve Tam, 1998:56-57).

Hata paylarının normal dağılmaması veya veri kümesi içinde sapan değerlerin bulunması durumunda en küçük mutlak sapmalar yöntemi diğer klasik tahmin yöntemlerine üstünlük göstermektedir.

1.1.2. Basit En Küçük Mutlak Sapmalar Regresyon Doğrusunun Tahmin Edilmesi

Çoklu doğrusal regresyon yöntemi β_0 ve β_1 tahmincilerini, hata terimlerinin karelerinin toplamını ($\sum \hat{u}_i^2$) minimum yapacak şekilde hesaplamaya dayanmaktadır. LAD yöntemi ise, hata terimlerinin mutlak değerlerinin toplamını minimum yapan parametre değerleri seçilir. Bu durum aşağıdaki şekilde ifade edilir:

$$\sum\left|\widehat{u_i}\right|$$

$\sum|\widehat{u_\iota}| = min\sum\left|y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right)\right|$ olarak yazılabilmektedir. Bu ifadede yer alan y ve x değerleri belli olduğuna göre, ifadenin minimum olması $\beta_0$ve β_1 değerlerinin minimum olmasına bağlıdır. Buradan hareketle LAD yönteminin amacının β_0 ve β_1 $'in$ $\sum|\widehat{u_\iota}|$ 'yi minimum yapacak şekilde tahmin etmek olduğu söylenebilir

LAD tekniğin amacı, mutlak sapmalar toplamını minimum yapan regresyon doğrusunun seçimidir. Bu doğru ise herhangi bir (x_0, y_0) veri noktasından geçen doğrular arasından en iyi sonucu veren doğru olacaktır.

Her aşamada toplam mutlak sapması daha küçük bir doğru bulunarak, bulunan bir doğru, bir önceki adımda bulunan doğru ile aynı çıkana kadar işlemler devam ettirilmektedir. Aynı doğru ardı ardına iki defa elde edildiğinde ise yapılan işleme son verilmektedir. Bulunan bu doğru, tüm doğrular arasında toplam mutlak sapması en küçük olanıdır dolayısıyla en iyisidir ve aranan "En Küçük Mutlak Sapmalar Regresyon Doğrusu" dur.

(x_0, y_0) noktasından geçen tüm doğrular arasından en iyi doğruyu bulmak için tanımlanan algoritmanın işleyişi aşağıdaki gibidir. İki noktadan geçen doğrunun eğimi; $\frac{(y_i - y_0)}{(x_i - x_0)}$ formülünden yararlanılarak yapılır. Hesaplamalar sırasında, x_i, x_0' a eşit

olduğunda eğim tanımlanamayacağı için $x_i = x_0$ eşitliğini noktalar işleme tabi tutulmamaktadır. Bütün eğimler hesaplandıktan sonra;

$$\frac{(y_1 - y_0)}{(x_1 - x_0)} \leq \frac{(y_2 - y_0)}{(x_2 - x_0)} \leq \ldots \leq \frac{(y_n - y_0)}{(x_n - x_0)}$$

olarak artan şekildedir.

$$T = \sum |x_i - x_0|$$

olmak üzere;

$$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| \langle \frac{1}{2} T$$

$$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| + |x_k - x_0| \rangle \frac{1}{2} T$$

$|x_k - x_0|$ ifadesinin eklenmesi ile $|x_i - x_0|$ toplamı T/2 'yi geçmektedir. Böylece (x_0, y_0) noktasından geçen en iyi doğrunun (x_k, y_k) noktasından geçtiği bulunarak bu doğrunun eğimi;

$$\beta^* = \frac{(y_k - y_0)}{(x_k - x_0)}$$

dir.

Aynı işlemler (x_k, y_k) noktası için yapılarak, (x_k, y_k) noktasından geçen tüm doğrular arasında en iyi doğru bulunur. Bu işlemler, ard arda birbirinin aynısı iki doğru bulunana kadar devam etmektedir. Bulunan bu doğru veri kümesini en iyi temsil eden doğrudur ve "En Küçük Mutlak Sapmalar Regresyon Doğrusu" olarak tanımlanmaktadır.

(x_0, y_0) noktasından geçen en iyi doğru;

$$\hat{Y} = \beta_0^* + \beta_1^* X$$

dir ve burada,

$$\beta_0^* = y_0 - \beta_1^* x_0$$

bulunan k indisi tekrarlandığında algoritma durdurulur (Birkes ve Dodge, 1993:58).

Daha öncede belirtildiği gibi algoritmanın uygulanabilmesi için; bir noktadan geçen bir en iyi doğru vardır ve bir noktadan geçen en iyi doğru aynı zamanda sadece bir noktadan daha geçer varsayımları kabul edilir. Algoritma ileriye doğru adım adım oluşturulmakta olup her adımda verilen noktadan geçen, sapmayı minimize eden en iyi doğru bulunmaktadır. Varsayımın aksine bir noktadan geçen birden fazla en iyi doğru olabilir, bir noktadan geçen en iyi doğru aynı zamanda iki ya da daha fazla noktadan geçebilir.

Bu algoritma, veri kümesinin, tek olmama ve dejenere olma kavramlarına dayanmadığını varsaymaktadır, çünkü bu kavramlar teknik problemlere sebep olabilmektedir.

1.1.3. En Küçük Mutlak Sapma Regresyonunda Tek Olmama ve Bozulma Sorunu

Tek olmama bir gözlem noktasından geçen birden fazla en iyi doğru olmasıdır. Bozulma ise, bir gözlem noktasından geçen en iyi doğrunun iki veya daha fazla gözlem noktasından da geçmesidir. Bu durumda bir sonraki adım için gözlem noktası yanlış seçilebilir algoritma bir döngüye girebilir veya seçilen doğru LAD regresyon doğrusu olmayabilir. Bu sorunu gidermek için, tüm gözlem çiftleri için mümkün olan LAD regresyon doğruları bulunur. Aralarından en küçük mutlak sapma toplamını veren doğru seçilir. Ancak bu yöntemin uygulanabilirliği gözlem sayısı n 'ye bağlıdır.

Verilen (x_k, y_k) gözlem noktası için eğim,

$$b_1^* = \frac{(y_k - y_0)}{(x_k - x_0)} = \frac{(y_{k+1} - y_0)}{(x_{k+1} - x_0)} = \frac{(y_{k-1} - y_0)}{(x_{k-1} - x_0)}$$

eşitliginden hesaplanır.

Tek bir doğru olmaması sorununda ise LAD regresyon doğrularından herhangi biri seçilir (Birkes ve Dodge, 1993:59).

1.2. EGİM PARAMETRESİNİN ANLAMLILIK TESTİ

Eğim parametresini test etmek için;

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_1 \neq 0$$

hipotezi kurulur. $e_i = Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right)$ artıkları hesaplanır, bunlar küçükten büyüğe doğru sıralanır. En ortadan bağımsız değişken sayısından bir fazla (p+1) gözlem değeri atılır, basit regresyonda bir bağımsız değişken olduğundan $m = n - 2$ olur, burada n başlangıç gözlem sayısını m de sıfıra eşit olmayan yeni gözlem sayısını verir. Test istatistiği,

$$|t| = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \sim t_{\alpha/2;(n-2)}$$

olarak hesaplanır. Burada;

$$S_{\hat{\beta}} = \frac{\hat{\tau}}{\sqrt{\sum \left(X_i - \overline{X}\right)^2}}$$

olacaktır. Parametrenin standart hatasının hesaplamasında kullanılan $\hat{\tau}$

$$\hat{\tau} = \frac{\sqrt{m}\left[e_{(k_2)} - e_{(k_1)}\right]}{4}$$

olarak hesaplanır. Burada,

$k_1 : \left(\frac{m+1}{2} - \sqrt{m}\right)$'e yakın tam sayı değeridir.

$k_1 : \left(\frac{m+1}{2} + \sqrt{m}\right)$'e yakın tam sayı değeridir.

$\hat{\beta}_1$ tahmini değerinin $\hat{\beta}_1$'ya yakın olması beklenir. β_1 ve $\hat{\beta}_1$ arasındaki fark 1 ya da 2 standart sapmadan büyük olmamalıdır. $|t|$ değerinin büyük olması, $\hat{\beta}_1$ ile sıfır

arasındaki mesafenin $S_{\hat{\beta}_1}$ 'den büyük olmasını, böylelikle $\beta_1 \neq 0$ hipotezinin reddedilmesi yönünde karar verilmez ve H_0 reddedilir.

En küçük mutlak sapmalar yöntemindeki $\hat{\tau}$ niceliği, çoklu doğrusal regresyon yönetimindeki σ ile benzer görevdedir. Çoklu doğrusal regresyon için $\hat{\beta}$' nın standart hatası $\frac{\sigma}{\sqrt{\sum(x_i - \overline{x})^2}}$ iken; en küçük mutlak sapmalar yöntemi için $\hat{\beta}$' nın standart hatası yaklaşık $\frac{\hat{\tau}}{\sqrt{\sum(x_i - \overline{x})^2}}$ tır. Örnek hacmi büyüdükçe bu değer, parametre değeri olan $\beta_{E.K.M.S}$ 'a yaklaşmaktadır.

1.3. τ PARAMETRESI

Regresyon doğrusunun eğiminin tahmin edilmesinde, MSE ve LAD regresyon yöntemlerinden hangisinin daha iyi sonuç verdiği $\frac{\tau}{\sigma}$ oranına bağlıdır. τ ve σ rasgele hata paylarının büyüklüklerinin ölçüsüdür. ϑ medyan etrafındaki hata paylarının dağılımının olasılık yoğunluk fonksiyonu olmak üzere, $\tau = \frac{1}{2\vartheta}$'dir. σ büyük olduğunda, hata payları geniş bir alana yayılır. Bu durumda medyan etrafındaki olasılık yoğunluk küçüktür, bu sebeple ϑ da küçüktür. Yukarıdaki eşitlikten de görüldüğü üzere, ϑ'nın küçük olması, τ 'nun büyük olması anlamına gelir. Kısaca; σ büyük olduğunda da τ 'da büyük ve σ küçük olduğunda da τ 'da küçüktür. Fakat $\frac{\tau}{\sigma}$ 'nın gerçek oranı, anakütle hata paylarının dağılımının şekline bağlıdır. Hata payları normal dağılıyorsa $\frac{\tau}{\sigma} \rangle 1$ olmaktadır. Bu durumda, en azından büyük örnekler için, çoklu doğrusal regresyon yöntemi, en küçük mutlak sapmalar yönteminden daha iyi sonuç vermektedir. Hata paylarının normal dağılmaması halinde ise, $\frac{\tau}{\sigma} \langle 1$ olmaktadır (Temiz, 2006:10).

1.4. ÇOKLU DOĞRUSAL EN KÜÇÜK MUTLAK SAPMA REGRESYONU

LAD regresyon L_1-regresyon olarak da adlandırılır çünkü, $\sum\left|Y_i-\left(a+bX_i\right)\right|$ sapma vektörünün L_1 normudur. Bir v vektörünün L_1 normu $\sum\left|v_i\right|$' dir. Benzer şekilde EKK regresyonu da L_2-regresyonu olarak adlandırılır, çünkü sapma vektörlerinin L_2 normunu minimize eder. v vektörünün L_2 normu $\sqrt{\sum v_i^{\,2}}$ 'dür.

Basit doğrusal regresyonda LAD regresyon doğrusu iki noktadan geçmekteydi. p tane açıklayıcı değişkene sahip çoklu regresyonda LAD regresyon denklemi (p+1) gözlem noktasını sağlar.

Çoklu doğrusal regresyonda $\sum\left|e_i\right|$ ifadesini en küçük yapan

$$\min\left|y_i-\left(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+...+\beta_px_{ip}\right)\right|$$

biçiminde verilen optimizasyon probleminin çözülmesi gerekmektedir. Minimizasyonu için herhangi bir formül bulunmadığından bir algoritma kullanılır. Basit regresyonda olduğu gibi algoritmanın uygulanabilmesi için bir noktadan geçen bir en iyi doğru vardır ve bir noktadan geçen en iyi doğru aynı zamanda sadece bir noktadan daha geçer varsayımları kabul edilir.

Vektör yöntemi ile;

$$\hat{\beta}=\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_k\end{bmatrix} \quad \text{ve} \quad X_i=\begin{bmatrix}1\\ X_{i1}\\ X_{i2}\\ \vdots\\ X_{ik}\end{bmatrix}$$

olarak verilir.

Bu durumda mutlak sapma; $\sum\left|Y_i-\hat{\beta}^{'}X_i\right|$ olarak yazılır. Bunu minimize edecek $\hat{\beta}^{'}$ vektörünü bulmak amaçtır. Basit LAD regresyonunda olduğu gibi, Çoklu LAD

regresyonu da iteraktif olarak çözülür. $\hat{\beta}'$ vektörüyle işlemlere başlanır, sonra $\sum\left|Y_i-\hat{\beta}'X_i\right|$'nin minimum değerini veren daha iyi bir vektör bulunur. Sonunda $\hat{\beta}$'nın en iyi vektörü bulunmuş olur. Her adımda, $\hat{\beta}$ tahminleri vektöründe daha iyi $\hat{\beta}^*$ vektörü,

$$\hat{\beta}^* = \hat{\beta} + td$$

olarak bulunur. Bu vektörün bulunması için yön vektörü *d* ve *t* değerlerinin elde edilmesi gerekir. Minimumluğu sağlayacak *t'*yi bulmak için bir yöntem geliştirilir. Minimize edilecek ifade;

$$\sum\left|Y_i-\left(\hat{\beta}+td\right)'X_i\right|$$

olacağından, burada

$$Z_i = Y_i - \hat{\beta}'X_i, \quad W_i = d'X_i$$

dönüşümü yapılarak,

$$\sum\left|Z_i - tW_i\right|$$

olarak elde edilir. Bu gösterim, $\sum\left|\left(Y_i-Y_0\right)-\hat{\beta}\left(X_i-X_0\right)\right|$' u minimize edecek $\hat{\beta}$' yı bulmakla aynıdır. Z_i/W_i oranları hesaplanıp, artan sıraya göre dizilir. Z ve W' yi yeniden indeksleyerek k indeksi bulunur.

$$\left|W_1\right|+\left|W_2\right|+\ldots+\left|W_{k-1}\right|\langle\frac{1}{2}T$$

$$\left|W_1\right|+\left|W_2\right|+\ldots+\left|W_{k-1}\right|+\left|W_k\right|\rangle\frac{1}{2}T$$

Burada $T=\sum\left|W_i\right|$'dir. t'nin minimum yapan değeri

$$Z_k/W_k$$

dır. Algoritmanın her bir adımında *p* açıklayıcı değişken sayısından bir fazla (P+1) yön vektörü vardır. Her bir d_j vektörü için (+) pozitif yön söz konusu olduğu gibi (-)

negatif yön de söz konusudur. Bu nedenle açıklayıcı değişken sayısının bir fazlasının iki katı sayıda yön olacaktır. Bunlar arasında $\sum\left|Y_i-\left(\hat{\beta}+td\right)' X_i\right|$ değerini mümkün olduğunca hızlı $t=0$ değerine ulaştıran yön seçilir. Bu değerin nasıl hızla azaldığını belirlemek için sağ tarafın $t=0$'daki türevini alınır.

$\sum\left|Z_i-tW_i\right|$ ifadesinde $t=0$ ' da sağ tarafın türevi, $W_-+W_0-W_+$'dır. Burada;

$$W_-:\frac{Z_i}{W_i}\text{negatifken}\ \sum\left|W_i\right|\text{'lerin toplamı}$$

$$W_0:\frac{Z_i}{W_i}\ \text{sıfırken}\ \sum\left|W_i\right|\text{' lerin toplamı}$$

$$W_+:\frac{Z_i}{W_i}\ \text{pozitifken}\ \sum\left|W_i\right|\text{' lerin toplamıdır.}$$

Tüm yön vektörleri için bu türevler hesaplanır. Türevi negatiflerin en küçüğü olan yön en uygun yöndür. Tüm türevler pozitifse, bu durumda geçerli $\hat{\beta}$ vektörü β'nın en iyi tahminidir ve işlemler bu noktada son bulur (Birkes ve Dodge, 1993:60-65).

1.4.1. Çoklu En Küçük Mutlak Sapma Regresyonunda Tek Olmama Ve Bozulma Sorunu

Tek olmama ve bozulma sorunu türevi negatiflerin en küçüğü olan yön vektörünün birden fazla olması veya en küçük türevin sıfır olması durumunda çoklu LAD regresyonda da ortaya çıkar.

Algoritma türevlerinin tamamının pozitif değerler ve sıfırlardan oluştuğu bir adıma gelirse o andaki tahminlerin vektörü bir LAD vektörüdür. Ancak türevi sıfır olan diğer yönlerden de LAD vektörü elde edilebilir. Bazen de algoritma tüm türevlerin en küçük olduğu bir sonuç verebilir ki buna döngü denir.

Teorik olarak, tahminlerin LAD vektörü en az (p+1) gözlem noktasından geçmek zorunda olduğundan, tüm mümkün (p+1) gözleme sahip altkümeler değerlendirilerek mutlak sapmalar toplamını en küçük yapan vektör seçilebilir. Fakat bu çözüm büyük veri kümeleri için uygun olmaz.

Döngü meydana geldiği zaman uygulanabilecek daha uygun bir yol, algoritmaya farklı bir başlangıç vektörü ile başlama olabilir. Diğer bir yöntem de, çok küçük rasgele sayılar üretip, bunları açıklayıcı değişkenlere ekleyerek veriyi değiştirmek olabilir. Bu yöntem döngüyü giderir ve değiştirilen veriden elde edilen LAD vektörünün elde edildiği (p+1) gözlem asıl verinin LAD vektörünü de belirler. Bozulma sorunu da,

$$\sum \left| Y_i - \left(\hat{\beta} + td \right)' X_i \right|$$

eşitliğindeki t'nin en küçük değerinin $t^* = 0$ olması durumudur. Böyle bir durumda, farklı bir d yön vektörü kullanılarak sorun giderilebilir (Birkes ve Dodge, 1993:60-65).

1.4.2. Çoklu En Küçük Mutlak Sapma Regresyonunda Katsayıların Anlamlılığının Testi

LAD regresyonda katsayıların anlamlılığının test edilmesi indirgenmiş model ve tüm modelin artıklarının mutlak değerlerinin toplamları ile yapılmaktadır. Tüm modelde parametre sayısı p, indirgenmiş modelde ise parametre sayısı q' dur. İki model tahmin edilip artıklarının mutlak toplamları bulunarak iki model arasındaki farkı oluşturan (p-q) sayıda parametrenin anlamlılığı birlikte test edilir. Bu durumda,

$$H_0 : \beta_{q+1} = ... = \beta_p = 0$$

$$H_1 : En\ az\ bir \beta_i \neq 0$$

Hipotezi oluşturulur ve test istatistiği;

$$F_{LAD} = \frac{SAR_{reduced} - SAR_{full}}{(p-q)(\hat{\tau}/2)}$$

olarak hesaplanır. Burada

$$SAR = \sum |e_i|$$

olup ayrıca;

$$\hat{\tau} = \frac{\sqrt{m}\left[e_{(k_2)} - e_{(k_1)} \right]}{4}$$

olarak elde edilir. m değeri sıfır olmayan artık sayısı olup

$$m = n - (p+1)$$

dir . $e_{(k_1)}$ ve $e_{(k_2)}$ basit LAD regresyonunda açıklandığı gibi hesaplanır (Birkes ve Dodge, 1993:60-65).

İKİNCİ BÖLÜM

KANTİL REGRESYON

2.1. KANTİL KAVRAMI

Bu bölümde, kantil kavramı, kantil dağılım fonksiyonu, kantil yoğunluk fonksiyonu, kantil regresyon, kantil regresyonun doğrusal programlama gösterimi ele alınacaktır.

Serileri iki, dört, on ve yüz eşit parçaya ayıran değerler genel olarak kantil olarak adlandırılmaktadır. Seriyi iki eşit parçaya bölmek için hesaplanan değerlere medyan, dört eşit parçaya bölmek için hesaplanan değerlere çeyreklik (quartile), on eşit parçaya bölmek için hesaplanan değerlere ondalık (desil) ve yüz eşit parçaya bölmek için hesaplanan değerlere santil adı verilmektedir (Serper, 2004:123).

a) Kartiller

Kartiller, büyüklük sırasına konulmuş bir seriyi dört eşit kısma bölen değerlerdir. Bir seride üç kartil mevcuttur ve ikinci kartil medyandır. Birinci kartil Q_1, ikinci kartil Q_2 ve üçüncü kartil Q_3 ile gösterilir. Basit ve sınırlandırılmış serilerde,

$$Q_1 = \frac{N+1}{4} \text{nci değer}$$

$$Q_2 = \frac{2(N+1)}{4} \text{nci değer} = \frac{N+1}{4} \text{ inci değer ve}$$

$$Q_3 = \frac{3(N+1)}{4} \text{nci değerdir.}$$

b) Desiller

Desiller, büyüklük sırasına konulmuş bir seriyi 10 eşit kısma bölen değerlerdir. Dokuz adet deşil vardır. Birinci desil D_1, ikinci desil D_2,.., dokuzuncu desil D_9 ile gösterilir. Basit serilerde,

$$D_1 = \frac{N+1}{10} \text{uncu değer,}$$

$$D_2 = \frac{2(N+1)}{10} = \frac{N+1}{5} \text{inci değer}$$

$$D_5 = \frac{5(N+1)}{10} = \frac{N+1}{2} \text{inci değer } = Q_2 = \text{Medyan}$$

$$D_9 = \frac{9(N+1)}{10} \text{ uncu değer}$$

c) Pörsentil

Pörsentiller, büyüklük sırasına konulmuş bir seriyi 100 eşit kısma bölen değerlerdir. Toplam 99 adet pörsentil vardır. En küçüğü birinci pörsentil (P_1) en büyüğü doksandokuzuncu (P_{99}) pörsentildir. Basit serilerde,

$$P_1 = \frac{N+1}{100} \text{ üncü değer,}$$

$$P_2 = \frac{2(N+1)}{100} = \frac{N+1}{50} \text{ inci değer,}$$

$$P_{50} = \frac{50(N+1)}{100} = \frac{N+1}{2} \text{inci değer } = Q_2 = D_5 = Medyan$$

$$P_{99} = \frac{99(N+1)}{100} \text{üncü değerdir (Oktay ve Başar, 2010:30-33).}$$

2.2. ANAKÜTLENİN MODELLENMESİ

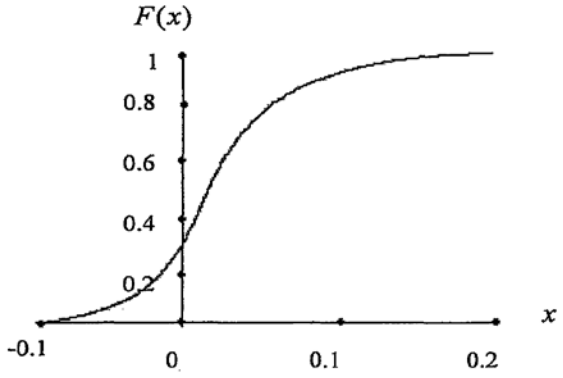
Herhangi bir dağılıma sahip örnek yapısı gösterilirken kümülatif dağılım fonksiyonu, olasılık yoğunluk fonksiyonu, kantil fonksiyonu ve kantil yoğunluk fonksiyonu olmak üzere dört farklı yol kullanılır.

2.2.1. Kümülatif Dağılım Fonksiyonu

Kümülatif dağılım fonksiyonu *F(X)*'le gösterilir; verilen *x* değerinde küçük ya da eşit olan *X* değişkeninin olasılığı olarak tanımlanır ve

$$F\left(X\right) = O\text{lasılık}\left(\text{tesadüfi değişken } X \le x\right)$$

olarak ifade edilebilir. Azalan fonksiyon olan kümülatif dağılım fonksiyonu aşağıda gösterilmiştir.



Şekil 2.1.Kümülatif Dağılım Fonksiyonu

Görüldüğü gibi fonksiyon sol taraftan '0'olasılık değerinden başlayarak, sağ tarafta bulunan '1' olasılığına kadar devam etmektedir. Herhangi bir olasılığın θ örnekten elde edilen değerinin *x*'e göre çizimini verir. Kümülatif dağılım fonksiyonu (CDF) dağılımının tek yönlü tanımlamasını yapar (Gilchrist, 2000: 10).

2.2.2. Olasılık Yoğunluk Fonksiyonu

Bir değişkenin alabileceği değerlerle bu değerleri alma olasılıkları arasındaki bağıntıyı gösteren fonksiyona 'Olasılık Yoğunluk Fonksiyonu (PDF)' denir (Saraçoğlu ve Çevik, 1995: 67). *f (x)* ile gösterilen olasılık yoğunluk fonksiyonu,

$$f\left(x\right)dx = O\text{lasılık}\left(x \le X \;\; tesadüfi\;\; değişkeni \le x + dx\right)$$

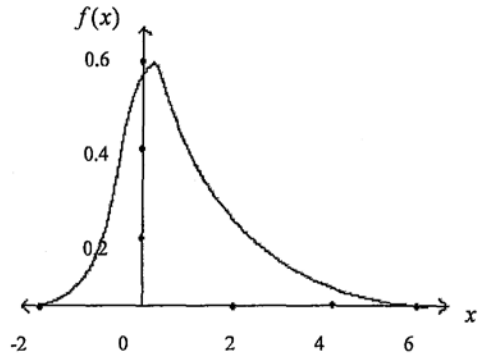
olarak tanımlanır. Burada *dx, x*' in sonsuza doğru küçük aralığıdır. *f (x)* eğrisinin altındaki alan, herhangi gözlenen değerin toplam olasılığı, 1 olmalıdır. Kümülatif dağılım fonksiyonu ve olasılık fonksiyonu arasındaki ilişki,

$$f(x)dx = F(x+dx) - F(x) = dF(x)$$

olacaktır. Olasılık fonksiyonu, kümülatif dağılım fonksiyonunun türevine eşit olup

$$f(x) = dF / dx$$

olarak elde edilir.

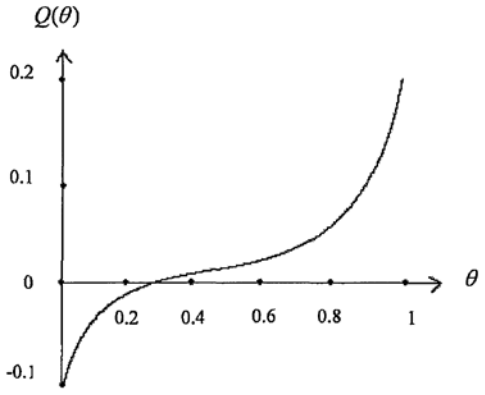


Şekil 2.2. Olasılık Yoğunluk Fonksiyonu

2.2.3. Kantil Fonksiyon

Kantil fonksiyon QF ve $\mathrm{Q}(\theta)$ ile gösterilmektedir. Kantil değeri değişkenin dağılımında yer alan ve dağılımı, kendisinden büyük olanlar ve kendisinden küçük olanlar diye ikiye bölen herhangi değerdir. Yani değerlerin % θ' sı, θ. kantilden daha küçüktür.(θ olasılık değerini ifade etmektedir)

$x_\theta = (X \le x_\theta)$ olasılığı için x' in değeridir



Şekil 2.3. Kantil Fonksiyon

x_θ 'nın değeri, kitlenin θ.'cı kantil olarak adlandırılır. $x_\theta = Q(\theta)$ fonksiyonu, θ. kantil, θ' nın bir fonksiyonu olarak ifade edilir ve kantil fonksiyon olarak adlandırılır.

(0,1) aralığındaki her θ ve rasgele Y değişkeni için, Y'nin θ. kantili her $\varepsilon_\theta \epsilon R$

$$P(Y < \varepsilon_\theta) \leq \theta \leq P(Y < \varepsilon_\theta)$$

olarak tanımlanır.

Kantil fonksiyon ve kümülatif dağılım fonksiyonu, herhangi (*x*, θ) çifti için $x_\theta = Q(\theta)$ ve $\theta = F(x)$ şeklinde yazılabilir. Bu fonksiyonlar birbirlerinin tersine eşit ve sürekli artan fonksiyonlardır. Böylece;

$$Q(\theta) = F^{-1}(\theta) \text{ ve } F(x) = Q^{-1}(\theta)$$

şeklinde de gösterilir (Gilchrist, 2000: 13).

Q(θ) kantil fonksiyonu ise, θ' nın tüm olasılıkları için, $0 \leq \theta \leq 1$, kantil değerlerini verir. Medyanda $Q(0,5)$ tir. Benzer şekilde $Q(1/4)$ ve $Q(3/4)$ kantilleridir. Hesaplamalar için normal dağılım tablosundan faydalanılır. Kullanılan normal tablolar standart normal dağılım için kantil fonksiyon tablolarıdır.

Dağılımları modelleyebilmek için kantil fonksiyon kullanılabilir. x verilmişken y' nin θ. kantili;

$$Q_y(\theta / x) = \beta x + \eta S(\theta)$$

olarak gösterilir. Burada;

$\eta S(\theta) =$ artık terimi ifade etmektedir.

S(θ)= simetrik olması gerekmeyen kantil fonksiyonudur.

η =ölçek parametresidir.

$'y'nin$ x üzerindeki kantil regresyon fonksiyonu'' ya da 'şartlı kantil fonksiyonu' olarak adlandırılır (Gilchrist, 2000: 13).

2.2.4. Kantil Yoğunluk Fonksiyonu

Dağılımları modelleyebilmek için, dağılım fonksiyonunun türevini alarak olasılık yoğunluk fonksiyonu elde edildiği gibi, *QF* 'in de türevi alınarak kantil yoğunluk fonksiyonu (*QDF* **)** belirlenebilir ve

$$q(\theta) = qQ(\theta) / d\theta$$

olarak gösterilir. $Q(\theta)$ azalmayan bir fonksiyon olduğu için eğimi $q(\theta)$ negatif değildir, her zaman $0 \le \theta \le 1$ birim aralığında yer almakta iken, olasılık yoğunluk fonksiyonu *f (x)* ise sonsuz tanım aralığında yer almaktadır.

Serinin mod değerinin olasılığı $p - \text{mod} \ge 0,5$ ise dağılım sola çarpıktır ve $q(\theta)$ yoğunluk fonksiyonu $q(\theta) \le q(1-\theta)$ durumunu sağlar, $0 \le \theta \le 0,5$'tir. Kantil fonksiyonu da $Q(\theta) + Q(1-\theta) \le 2\theta(0,5)$ durumunu sağlar ve ortalama ≤ medyan ≤ mod sıralaması sağlanır.

Benzer şekilde serinin mod değerinin olasılığı $p - \text{mod} \ge 0,5$ ise dağılım sağa çarpıktır ve $q(\theta)$ yoğunluk fonksiyonu işaretler durumunu sağlar, $0 \le \theta \le 0,5$'tir.

Kantil fonksiyonu da $Q(\theta) + Q(1-\theta) \ge 2\theta(0,5)$ durumunu sağlar ve ortalama ≥ medyan ≥ mod sıralaması sağlanır

2.3. KANTİL REGRESYON

Kantil regresyon, ilk olarak regresyondaki klasik varsayımlardan hata terimlerinin normal dağılması varsayımını ihmal eden robust (sağlam) bir regresyon tekniği olarak ortaya çıkmıştır. Gelir dağılımında ki eşitsizlik ya da ücretlerde ki farklılık gibi dağılımın bozulduğu konularda kullanımı yaygın olan Kantil regresyon, daha kapsamlı bir regresyon görüntüsü sunmak amacıyla tasarlanan bir yöntemdir (Koenker, 2005:112).

Uygulamalı istatistiğin önemli bir kısmı lineer regresyon modeli ve bu modelin tahmininde sıklıkla kullanılan "En küçük Kareler" tahmin metotlarının detaylı bir şekilde incelenmesi olarak görülebilir (Koenker, 2005:115)

Kantil regresyon modelleri şartlı ortalama fonksiyonları ve şartlı kantil fonksiyonları için tahmin yapılmasında kullanılır. Mostseller ve Tukey; ‘‘regresyon eğrisinin yaptığı şey x’lerin kümesine karşılık gelen dağılımların ortalamasının özet bilgisini vermektedir. Daha fazla bilgi için dağılımların çeşitli yüzde puanlarına denk gelen farklı regresyon eğrileri hesaplanabilir, bu sayede de kümenin daha detaylı bilgisi elde edilebilir.’’ şeklinde ifade etmişlerdir.

Kantil Regresyon, özellikle koşullu kantillerin değişkenlik gösterdiği durumlarda kullanışlıdır. Kantillerin bağlı olarak regresyon katsayılarını belirler (Chen, 2005).

Çoklu doğrusal regresyon modelinde hata teriminin değişkenlerin değerinden bağımsız olduğu (varyanslar homojen) varsayılır. Tam tersine kantil regresyon modelinde hata terimlerinin değişkenliğine izin verilir ve varyans yapısına ilişkin herhangi bir varsayımı bulunmamaktadır (Baur ve ark. 2004:4695).

Bağımsız değişken x’ in değişen değerlerine karşı kantil regresyonlarla analiz yapmak, aşırı değerlerin varlığı durumunda daha etkin sonuçlar ortaya koymaktadır. Çoklu doğrusal regresyon doğrusu, aşırı değerleri yakalayamazken, farklı katillerdeki kantil regresyon doğruları aşırı değerleri rahat bir şekilde yakalayabilir (Wang, 2007:9-10).

Kantil regresyon modeli aslında bir yerleşim modelidir. Basit yerleştirme modeli,

$$Y_t = \beta + e_t$$

olarak ifade edildiğinde, burada yer alan Y_t, simetrik F dağılım fonksiyonuna sahip, bağımsız, özdeş dağılımlı, β medyanlı tesadüfi değişkendir. Bu modelde θ. örnek quantili

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i:y_i \geq \beta} \theta \left| y_i - \beta \right| + \sum_{i:y_i < \beta} (1-\theta) \left| y_i - \beta \right| \right\}$$

ifadesinin minimizasyonu ile ifade edilir (Judge, 1985:834). Bu doğrusal regresyon modeli;

$$y_i = x_i^{'} \beta + e_i$$

θ. kantil regresyon gözlem değerlerinin işaretlerine bağlı olarak

$$\theta \min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(\theta - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(y_i - x_i^{'} \beta \right) \left(y_i - x_i^{'} \beta \right) \right)$$

şeklinde tahmin edilir. Burada, $\operatorname{sgn}(\alpha)$ α 'nın işaretidir ve α pozitif ise 1, negatif veya 0 ise -1 değerini alır. Tahmincilerin gözlem değerlerinin büyüklüğü yerine işaretlerine dayalı olması kantil regresyonun robust bir yöntem olmasını sağlamaktadır. Minimizasyon için birinci mertebe koşulunun sağlanması gerektiğinden birinci mertebe koşulunun $K \times 1$ vektörü,

$$\theta \min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(\theta - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(y_i - x_i^{'} \beta \right) \left(y_i - x_i^{'} \beta \right) \right) = 0$$

olarak gösterilir. Bu ifade birinci mertebe koşulu genelleştirilmiş momentler yöntemi (GMM)'ne uyan bir moment fonksiyonudur. Moment fonksiyonu,

$$\psi\left(x_i, y_i, \beta \right) = \left(\theta - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(y_i - x_i \beta \right) \right) x_i$$

olarak tanımlanabilir. $\psi(.)$ 'nın fonksiyon olarak geçerli olabilmesi için belirli düzenleme şartları altında

$$E\left[\psi\left(x_i, y_i, \beta_\theta \right) \right] = 0$$

olması gerekir. Asimtotik dağılımın genel gösterimi;

$$E\left[f_{u_\theta}\left(0 / x_i \right) x_i x_i^{'} \right] = \partial E\left[\psi\left(x_i, y_i, \beta_\theta \right) \right] / \partial \beta_\theta^{'}$$

$$\theta(1-\theta) E\left[x_i x_i^{'} \right] = E\left[\psi\left(x_i, y_i, \beta_\theta \right) \psi\left(x_i, y_i, \beta_\theta \right)^{'} \right]$$

şeklinde olup genelleştirilmiş momentler yöntemi kullanılarak elde edilen parametre tahmincileri tutarlı ve asimtotik olarak normal olacaktır.

Belirli düzenleme şartları altında,

$$\sqrt{n}\left(\hat{\beta}_\theta - \beta_\theta \right) \xrightarrow{L} N\left(0, \Lambda_\theta \right)$$

olarak gösterilebilir. Burada,

$$\Lambda_\theta = \theta(1-\theta) \left(E\left[f_{u_\theta}\left(0 / x_i \right) x_i x_i^{'} \right] \right)^{-1} E\left[x_i x_i^{'} \right] \left(E\left[f_{u_\theta}\left(0 / x_i \right) x_i x_i^{'} \right] \right)^{-1}$$

olarak tanımlanır (Koenker ve Bassett, 1978:48).

Olasılık değeri 1 olduğunda ve $f_{u_\theta}(0/x_i) = f_{u_\theta}(0)$ ise, yani hata teriminin yoğunluğu sıfır etrafındaysa ve *x*' ten bağımsızsa, Λ_θ ,

$$E(y_i/x_i) = \mu(x_i, \beta) = \mu_i$$

şeklinde sadeleştirilebilir. $f_{u_\theta}(./x)$ *x*'ten bağımsız olduğunda, tüm quantillerin parametre ve vektörleri sadece kesim noktalarında farklılık gösterir.

Kantil katsayılarını yorumlayabilmek için y'nin k açıklayıcı değişkenine göre şartlı kantil 'nin kısmi türevi alınmaktadır. Türev alındığında,

$$\partial Quant_\theta(y_i / x_i) / \partial x_{ik}$$

olacaktır. Bu türev, *x*' in k.'ncı değerindeki marjinal değişime göre, θ.'cı şartlı kantil deki marjinal değişimi vermektedir (Behr, 2008:570).

Kantil regresyon parametre tahminleri, açıklayıcı değişkendeki bir birim değişme karşısındaki y'nin belli bir katilindeki değişmeyi gösterir (www.cscu.cornell.edu/, 2007).

2.3.1. Kantil Regresyon Yönteminin Özellikleri

Kantil regresyonun en önemli özellikleri aşağıda sıralanabilir (Leping, 2005:15).

1) Çoklu Doğrusal Regresyon yöntemi *y*' nin koşullu dağılımının ortalaması hakkında bilgi vermekte, kantil regresyon ise farklı kantil değerleri için *y*'nin *x*' e göre koşullu dağılımının tümü hakkında bilgi vermektedir.

2) Kantil Regresyon' da; $\min_b \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(y_i - x_i'b)$ ifadesinin minimizasyonu, doğrusal programlama(LP) gösterimidir, bu durum da tahmin kolaylaşmaktadır

3) Kantiller monoton dönüşümlere olanak verirler. Herhangi h(.) monoton fonksiyonu için $Q_{h(y)/x}(\tau / x) = h(\theta_{y/x}(\tau / x))$ olur.

4) Kantiller *y*' deki aşırı değerlere karşı kararlıdırlar (robust).

5) Hata terimi normal dağılmadığında, kantil regresyon tahmin edicileri çoklu doğrusal regresyon tahmin edicilerinden çok daha etkin olabilmektedir.

6) Kantil regresyon değişen varyansın belirlenmesine imkân vermektedir.

7) Kantil regresyon amaç fonksiyonu için tahmin edilen katsayı vektörü, bağımlı değişkendeki aşırı değerlere duyarlı değildir ve yerleşimin robust bir ölçüsüdür.

8) Farklı kantillerde farklı sonuçlar çıkması, bağımlı değişkenin koşullu dağılımının farklı noktalarındaki bağımsız değişkenlerdeki değişikliklere farklı tepki vermesi olarak yorumlanabilir.

9) Kantil regresyon analizinde $\tau = 0,5$ olması durumunda LAD regresyon analizi elde edilmektedir.

2.3.2. Kantil Regresyonun Doğrusal Programlama Gösterimi

Kantil regresyon tahmin edicileri doğrusal programlama problemi olarak formüle edilebilir ve artıkların iki parçalı doğrusal amaç fonksiyonu optimize edilerek simpleks veya sınır metodu gibi yöntemlerle çözülebilir (Koenker ve Hallock, 2001: 153).

F dağılım fonksiyonuna sahip *Y* bağımlı değişken, b, tahmin edilecek katsayı vektörü ve $e_i = y_i - x_i\beta$ hata değeri olmak üzere, θ. regresyon kantili $(0 < \theta < 1)$;

$$\min_{\beta} \frac{1}{n}\left\{ \sum_{i:y_i \geq x_i\beta} \theta\left|y_i - x_i\beta\right| + \sum_{i:y_i < x_i\beta} (1-\theta)\left|y_i - x_i\beta\right| \right\}$$

$$\min \frac{1}{n}\left\{ \sum_{i=1}^{n} \rho_\theta\left(y_i - x_i\beta\right) \right\}$$

Minimizasyon ile tahmin edilir. y'nin θ. kantil olarak adlandırılır. Kantil regresyonun bu şekilde gösterimi doğrusal programlama gösterimidir. B parametre tahmincileri;

$$\widehat{\beta}(\theta) = \arg\min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} \rho_\theta\left(y_i - x_i\beta\right) \right\}$$

çözümü ile tahmin edilebilir. Burada $0 < \theta < 1$'dir.

y_i sadece pozitif elemanların fonksiyonu;

$$y_i = \sum_{j=1}^{K} x_{ij}\beta_{\theta_j} + u_{\theta_i} = \sum_{j=1}^{K} x_{ij}\left(\beta^1{}_{\theta_j} - \beta^2{}_{\theta_j}\right) + \left(\varepsilon_{\theta_i} - v_{\theta_i}\right)$$

$\beta^1{}_{\theta_j} \geq 0,\ \beta^2{}_{\theta_j} \geq 0$ (j=1,...,K) ve $\varepsilon_{\theta_i} \geq 0,\ v_{\theta_i} \geq 0$ (i=1,…,n) olarak yeniden yazılabilir (Buchinsky, 1998:91-92).

Daha etkili bir tahmin için;

$$\min_{\beta} \frac{1}{n}\left\{\sum_{i=1} f_{u_\theta}(0/x)\left(\theta - \frac{1}{2} + \frac{1}{2}\operatorname{sgn}(y_i - x_i\beta)\right)(y_i - x_i\beta)\right\}$$

prosedür bilinmeyen yoğunluk $f_{u_\theta}(0/x)$ için bir tahmin kullanılmasını gerektirir (Newey ve Powel, 1990:295-300).

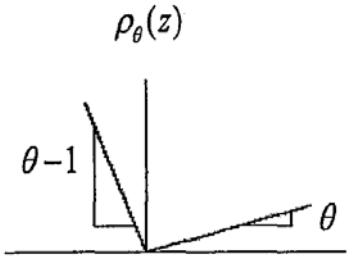
Burada I karakteristik fonksiyon, ρ_θ ise kontrol fonksiyonudur ve,

$$\rho_\theta(z) = z\left(\theta - I(z \leq 0)\right)$$

veya

$$\rho_\theta(z) = \begin{cases} \theta z, & z \geq 0 \\ (\theta - 1)z, & z < 0 \end{cases}$$

olarak tanımlanır. Bu fonksiyon;



Şekil 2.4. Kantil Regresyon Amaç Fonksiyonu

şeklinde gösterilir (Koenker ve Hallock, 2001:143). θ=0,5 olması durumunda kantil regresyon amaç fonksiyonu LAD amaç fonksiyonuna ve I_1 regresyona eşittir (Lee,

2004:1). Kantil regresyon amaç fonksiyonu mutlak sapmaların ağırlıklandırılmış toplamıdır.

2.3.3. Kantil Regresyon Parametre Tahminleri

Kalıntıların kareleri toplamını minimize eden parametre değerlerini alarak $\hat{\beta}_0$ ve $\hat{\beta}_1$ için çoklu doğrusal regresyon parametreleri çözer.

$$\min \sum_i \left(y_i - \left(\beta_0 + \beta_1 x_i \right) \right)^2$$

Bu ifade y= $\hat{\beta}_0 + \hat{\beta}_1 x_i$ doğrusu ve $\left(x_i, y_i \right)$ veri noktaları arasındaki dikey mesafe karelerin toplamıdır. Bu denklemin sıfıra eşitlenerek kısmi türevi alındığında iki bilinmeyenli iki denklem elde edilir. Denklem çözüm sistemleri;

$$\hat{\beta}_1 = \frac{\sum_i^n \left(x_i - \overline{x} \right)\left(y_i - \overline{y} \right)}{\sum_i^n \left(x_i - \overline{x} \right)^2}, \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

dir.

Kantil regresyonda, çoklu lineer regresyon parametre tahminlerinin kantil regresyon parametre tahminlerinden farkı bir doğrudan noktaya olan uzaklığın dikey mesafenin toplam etkisi kullanılarak ölçülmesidir. Burada etki doğru üzerindeki noktalar için p ve tahmini doğru altındaki noktalar için (1-p) dir. p oranı için her seçim, örneğin; p=.10, .25, .50 farklı tahmini koşullu- kantil fonksiyona neden olur. Amaç her p olasılığı için özellik tahmini bulmaktır.

Medyan regresyon modeli için tahmini düşünüldüğünde medyan regresyonda da aynı durum söz konusudur. Saf kalıntıların toplamını minimize etme tercih edilir. Diğer bir ifade ile saf kalıntıların toplamını minimize ederek katsayıları bulunur. β_s için tahmini çözüm

$$\sum_i \left| y_i - \beta_0 - \beta_1 x_i \right|$$

dür. Bu ifade minimize edildiğinde çözüm sonucu olarak medyan regresyon doğrusuna bakarız. Regresyon doğrusu üzerinde uydurma kalıntı verilerinin yarısı ve aşağı düşen

diğer yarısı ile veri noktalarının bir çifti ile geçmelidir. Yani kalıntıların yarısı pozitif yarısı negatiftir (Hao ve Naiman, 2007:33).

2.3.4. Belirlilik Katsayısı ve Düzeltilmiş Belirlilik Katsayısı

Açıklayıcı sayısı çok olduğu için çoklu korelasyon söz konusudur. Korelasyon katsayısının karesine çoklu belirlilik katsayısı ya da çoklu korelasyon katsayısının karesi adı verilir. Çoklu belirlilik katsayısı $\boldsymbol{R^2}$ ile gösterilir.

$$R^2_{Y,X_1,X_2,\ldots,X_k} = \frac{\widehat{\beta_1}\sum yx_1 + \widehat{\beta_2}\sum yx_2 + \ldots + \widehat{\beta_k}\sum yx_k}{\sum y^2}$$

Fonksiyona yeni değişkenler eklendiği zaman azaldığı apaçık olan serbestlik derecesini hesaba katmak üzere, R^2 'yi düzeltilir. Düzeltilmiş çoklu belirlilik katsayısının ifadesi

$$R^2 = \frac{\min \sum_i \left(x_i\hat{\beta} - \bar{y}\right)^2}{\min \sum_i \left(y_i - \bar{y}\right)^2} = 1 - \frac{\min \sum_i \left(x_i\hat{\beta} - \bar{y}\right)^2}{\min \sum_i \left(y_i - \bar{y}\right)^2}$$

şeklindedir.

Koenker ve Machado (1999) kantil regresyon modelleri için benzer ölçümü aşağıdaki şekilde önerir.

$$R^1(\theta) = \frac{\min \sum_i \rho_\theta \left(x_i\hat{\beta} - \mathrm{Q}_\theta(y)\right)^2}{\min \sum_i \rho_\theta \left(y_i - \mathrm{Q}_\theta(y)\right)^2} = 1 - \frac{\min \sum_i \left(y_i - x_i\hat{\beta}\right)^2}{\min \sum_i \rho_\theta \left(y_i - \mathrm{Q}_\theta(y)\right)^2}$$

R^2 gibi $R^1(\theta)$ değeri de (0,1) arasındadır.

2.3.5. Eşit Varyans Özellikleri

Kantil regresyon eşit varyans özelliği hesaplama kolaylığı sağlayacak bazı özelliklere sahiptir. Aşağıdaki minimizasyonda tanımlanan soruna uygun çözümler kümesi β (θ, Y, X) tarafından ifade edilir.

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i=1} \left(\theta - \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(y_i - x'_i \beta \right) \right) \left(y_i - x'_i \beta \right) \right\}$$

$\hat{\beta}_\theta \equiv \hat{\beta}$ (θ, Y, X) ϵ β (θ, Y, X) ifade edildiğine göre;

$\hat{\beta}$ (θ, λy, X) = λ $\hat{\beta}$ (θ, y, X) λ ϵ [0,

$\hat{\beta}$ (1 − θ, λy, X) = λ $\hat{\beta}$ (θ, y, X) λ ϵ [0,

$\hat{\beta}$ (θ, y+X_γ, X) = $\hat{\beta}$ (θ, y, X) + γ γ ϵ R^k

$\hat{\beta}$ (θ, y, XA) = A^{-1} $\hat{\beta}$ (θ, y, X) A_{k*k} tekil değil

katsayılarının tahmininin eşit varyans özelliklerini gösterir. $\hat{\beta}_\theta$ eşit varyans ölçümüdür. Aşağıda gösterildiği gibi kantil regresyon bir doğrusal programlama sorunudur. Simpleks iterasyonun önemli ölçüde azalan sayısı eşitlikleri ile kullanılabilir (Koenker ve Bassett, 1978:38).

2.3.6. Güven Aralıkları

Kantil regresyon parametresi β(τ) kantil regresyonda güven aralığı hesaplamak için üç metot sağlar: Seyreklik, Rank Metodu ve Yeniden Örnekleme.

Seyreklik metodu en doğrudan ve hızlı olanıdır. Fakat seyreklik fonksiyonunun tahminini içerir. Bu bağımsız ve özdeş olmayan veriler için robust değildir. Bu problemle uğraşarak kantil regresyon prosedürü seyreklik fonksiyonunun bölgesel tahminini kullanarak Huber Sandwich tahminini hesaplar.

Rank metodu, hata yoğunluklarının doğrudan tahminini önleyen bir metottur. Gutenbrunner, Jureckova, Koenker ve Portnoy tarafından 1993 yılında, bağımsız- özdeş dağılımlı hata terimli model için geliştirilmiştir. Rank metodu, rank skor testi ters çevrilerek güven aralıklarını hesaplar. Simpleks algoritma kullanır ve büyük veri setleri ile hesaplama zordur. Bu model;

$$y_i = x'_i \beta + e_i$$

şeklindedir. Daha sonra ise, Koenker ve Machado (1999), bu metodu, konum-ölçek regresyon modelleri için geliştirmişlerdir.

Yeniden örnekleme, bu metot da varyans-kovaryans matrisinin doğrudan tahminini gerektirmemektedir. Yeniden örnekleme metotları; çift örnekleme, tahmin denklemlerini yeniden örnekleme ve Markov Zinciri Marjinal Yeniden Örnekleme şeklinde 3'e ayrılmaktadır. Yeniden örnekleme (bootstrap) örneğine dayalı güven aralığı ile problemlerin üzerinden gelebilir. Fakat küçük veri setleri için kararsız olabilir. Bu özelliklere dayanarak kantil regresyon rank yöntemi ve varsayılan yeniden örnekleme metodu bir arada kullanılır.

2.3.7. Kovaryans ve Parametre Tahminlerinin Korelasyonu

Kantil regresyon kovaryans ve korelasyon matrisleri hesaplamak için iki yöntem sağlar. Asimtotik metot ve bootstrap metodu. Yeniden örnekleme güven aralığı hesapladığında bootstrap kovaryans ve korelasyon matris hesaplar. Aksi takdirde, asimptotik kovaryans ve korelasyon matrisleri hesaplanır.

2.3.7.1. Asimtotik Kovaryans ve Korelasyon

Bu metot güven aralıkları için seyreklik metoduna karşılık gelir. Asimptotik kovaryans ve korelasyon hesaplanmasında seyreklik fonksiyonu için tahminler hesaplanır. Rank metodu kovaryans korelasyon tahmini sağlayamadığından bu metot kullanılarak güven aralığı hesaplandığında asimptotik kovaryans korelasyon hesaplanır.

2.3.7.2. Bootstrap Kovaryans ve Korelasyon

Bu yöntem, güven aralıkları için örnekleme yöntemine karşılık gelir. Markov zinciri marjinal bootstrap (MCMB) yöntemi kullanılır (Colin, 2005:114).

ÜÇÜNCÜ BÖLÜM

SAYMA VERİLİ REGRESYON MODELLERİ

3.1. POİSSON REGRESYON

Poisson regresyon, lojistik regresyondan sonra en genel olan ikinci genelleştirilmiş doğrusal modeldir. Bağımlı değişken oluş sayısı ile belirtilen bir veri olduğunda, yani belirli bir zaman ya da yerde olan olayların sayısı olduğunda yani sayıma dayalı olarak elde edilen bağımlı değişkenin analizinde kullanılmaktadır. Poisson regresyonda, veri olarak genellikle belli kategorilere göre gruplandırılmış tablolardan faydalanılır. Bu tablolardaki hücre değerlerinin belli bir özellik gösteren ve oluş sayısı ile belirtilen bir veri olması gerekmektedir. Bu şekildeki tabloların oluşturduğu verilerin çözümlenmesinde logaritmik doğrusal modeller kullanılmaktadır. Poisson regresyonda bu logaritmik doğrusal modellerden biridir (Cameron ve Trivedi, 1998:2).

Poisson regresyon modelinde, bağımsız değişkenlerin doğrusal yapısını bağımlı değişkenin beklenen değerine bağlayan bağ fonksiyonu logaritmiktir. Bağımlı değişkenlerin regresyon katsayıları, diğer tüm bağımsız değişkenlerin sabit tutulduğu durumda açıklayıcı değişkendeki bir birimin artısından önce ve sonra beklenen sayıların oranının logaritması olarak yorumlanabilir.

Yapılan çalışmalarda poisson dağılımı için ortalama ve varyansın eşit olması varsayımı altında poisson regresyon modeli incelenmiştir. Poisson dağılımında; varyansın ortalamadan büyük olması haline aşırı yayılım ve varyansın ortalamadan küçük olması haline az yayılım denilmektedir (Cox, 1983:269).

Poisson regresyon modeli, cevap değişkeni veya bağımlı değişkenin sayılabilir bir değişken olduğu durumlarda kullanılan bir regresyon modeli olmaktadır. İlgilenilen bir olayın meydana gelme oranları veya bir olayın meydana gelme sayıları poisson regresyon ile modellenebilir.

Dağılım 18. yüzyılda yaşamış olan Fransız matematikçi Poisson tarafından tanıtılmıştır. Sonraki yıllarda farklı bilim adamları tarafından üzerinde çalışılan ve

genişletilerek bugünkü halini alan dağılım ilk temellerini atan ünlü matematikçinin adını taşımaktadır (Gürsakal, 1997: 38).

Poisson regresyon genelleştirilmiş bir modeldir. Bu model μ parametresinin açıklayıcı değişkenlere dayandığı poisson dağılımında elde edilir.

$$y \sim Poisson\ (\mu_i)$$

olduğu kabul edilmekte ve burada açıklayıcı değişkenlerin bir vektörü ile ortalama μ_i arasında ilişki kurulmak istenmektedir.

Poisson regresyon modeli aşağıdaki özelliklere sahip olduğu için model parametrelerinin tahmininde çoklu doğrusal regresyon kullanılmaz. Bunun yerine en çok olabilirlik tahmini kullanılır (Deniz, 2005:61).

- Açıklayıcı değişken *Y*' nin Poisson dağılışı göstermesi gerekmektedir. Ayrıca koşullu ortalamanın tanımlanmasında bağımlılık şartı sağlanmalıdır.
- En çok olabilirlik standart hataları ve *t* istatistikleri kullanılarak hesaplanan istatistiksel sonuçlar, hem koşullu ortalama hem de varyansın doğru tanımlanmasını gerektirmektedir. Burada istenen koşul, koşullu varyans ile ortalamanın eşit olmasıdır. Negatif binomial regresyonda ise ortalamanın varyanstan büyük olmasını gerektirir.
- Veriler için koşullu varyans ve koşullu ortalamanın eşit olmaması durumunda, poisson tahmin edicisinden daha etkin tahmin ediciler kullanılabilir.
- Veriler için koşullu varyans ve koşullu ortalamanın eşit olmaması durumunda en çok olabilirlik yönteminin uygulaması ile elde edilmiş istatistiksel sonuçlar, koşullu ortalamanın doğru tanımlandığının ispat edildiği durumlarda geçerli ve doğrudur.

Poisson regresyon modeli;

$$p(Y;\mu) = \frac{e^{-\mu}\mu^{Y}}{Y!}, Y = 0,1,\ldots,$$

Genelleştirilmiş lineer model de aşağıdaki gibi olmalıdır:

$$p(Y;\mu) = \frac{e^{-\mu}\mu^{Y}}{Y!}, Y = 0,1,\ldots,$$

$$=exp[y_i ln(\mu_i) - \mu_i - \ln(y_i!)]$$

$$=exp[y_i \theta_i - exp(\theta_i) - \ln(y_i!)]$$

olarak verilmektedir (Cameron ve Trivedi, 1986:9-10). Formüldeki μ dağılımın parametresidir. Bu parametre, birimler arasındaki gözlemlenmiş heterojenlik ile ifade edilen regresyon modelinin açıklayıcı değişkenleri x' le değişebilir. Poisson dağılımı tek parametreli bir dağılımdır. Poisson dağılımının ortalaması ve varyansı μ eşittir. Burada μ_i aşağıda tanımlandığı gibidir.

$$E(y_i / x_i) = \mu(x_i, \beta) = \mu_i, i = 1,...,n$$
$$E(y_i / x_i) = \mu_i = \exp(x'_i \beta)$$

şeklinde gösterilir. İstatistik literatür de bu model log-doğrusal model olarak ifade edilmektedir.

Poisson regresyonunda bağımlı değişkenlerin doğrusal yapısını cevap değişkenin beklenen değerine bağlayan bağlantı fonksiyonu, logaritmik dönüşüm ile verilmektedir (Frome, 1983:666-667).

Poisson regresyonun bağlantı fonksiyonu da aşağıda ki logaritmik dönüşüm şeklindedir.

$$\text{Log}(\mu) = X'\beta + e$$

Poisson regresyonunda ilgilenilen olayın sayısı olan Y bağımlı değişkeninin;

$x_1, x_2, \ldots, x_3$ bağımsız değişkenleri verilmişken poisson dağılışına sahip olduğu varsayılmaktadır. Bu durumda poisson ortalaması olan μ'nün logaritmasının bağımsız değişkenlerinin bir doğrusal fonksiyonu olduğu varsayılmaktadır.

Bu fonksiyon aşağıdaki gibidir;

$$\text{Log}(\mu) = b_0 + b_1 x_1 + \cdots + b_m x_m$$

Poisson regresyonunda model uyumu Devians ve Pearson Khi-kare uyum istatistikleri kullanılarak yapılmaktadır. Her iki uyum istatistiğine ilişkin yayılım değerlerinin 1'e eşit olması yayılım olmadığı, 1'den büyük olması aşırı yayılım ve 1' den küçük çıkması da az yayılım olarak tanımlanmaktadır. Çalışmalarda, genellikle aşırı yayılım gözlenmektedir (Wang, 1996; Wang,1998: 273-283).

3.2. QUASİ-POİSSON REGRESYON

Poisson regresyon modelinin kullanılmasındaki en büyük kısıtlayıcı tek bir parametreye sahip olmasıdır. Bu sebeple ortalama ve varyans arasındaki oran 1'e eşittir. Modeldeki bu durum es yayılım anlamına gelmektedir. Genel olarak modellerde gözlenen ise aşırı yayılım olduğu durumlardır. Aşırı yayılım olduğu durumlarda kullanılan modeller;

1) Sosyo ekonomik statü ölçeği; çarpıklık ve χ^2 dağılımı
2) Sosyo ekonomik statü ölçeği; vadeli ölçek
3) Sağlam varyans tahmincileri,
4) Boostrap ya da jacknife,
5) Belirli bir sabit ile varyans çarpımı
6) Tahmin edilen parametre ile varyans çarpımı: quasi-poisson,
7) Genelleştirilmiş poisson,
8) Negatif binomial: NB2
9) Heterojen Negatif Binomial,
10) NB-P; genelleştirilmiş negatif binomial; Waring, Faddy (Hilbe, 2007: 157).

Quasi poisson modelleri poisson modellerden farklı olarak negatif binomial regresyona daha çok benzer özellik göstermektedir. Quasi poisson da aşırı yayılım durumunun söz konusu olduğu durumlarda kullanılır. Bu yöntemde standart poisson model gibi aynı katsayı tahminlerine yol açar fakat sonuç aşırı yayılım özelliği gösterdiği için düzeltilmiştir yani katsayı tahmini değişmez fakat standart hata değişebilir (Zeileis, Kleiber, Jackman, 2007:5)

Quasi- poisson analizi, negatif binomial regresyon modelleri gibi eşit parametre sayısına sahip olup benzer sonuçlar vermektedir. Bu yaklaşımlar dağılımın formuna ve olasılığına bağlıdır. Ancak quasi modeller sadece ortalamaları ve varyansları tarafından

karakterize edilirler. Her iki modelde genelleştirilmiş lineer modeller olarak yapılandırılabilirler.

Tahminlerde kovaryans etki farklılıkları göze çarpabilir. Quasi-poisson modelin varyansı ortalamanın lineer fonksiyonu iken negatif binomial modelin varyansı ortalamanın quadratik fonksiyonudur.

İstatistik literatüründe, negatif binomial ve quasi-poisson modellerine istinaden μ olarak ifade edilen λ' dır. Ayrıca μ genelleştirilmiş lineer modellerde (GLM) ortanca parametre olarak ifade edilir. Poisson ve negatif binomial dağılımları μ ile değişken ilişkilendirilmesine maruz bırakılır.

Y rasgele bir değişken olsun;

$$E(Y)=\mu$$

$$\text{Var}(Y)=v_{POi}(\mu)=\theta_{\mu}$$

E(Y) değeri Y'nin beklenen değeri, Var (Y), Y'nin varyansı, μ>0 ve θ>1dir. θ aşırı yayılım parametresidir.

Poisson model sadece bir bağ fonksiyonunun kullanılması ile quasi-poisson model olarak isimlendirilebilir ve Y~Poi(μ,θ) olarak gösterilir. Her iki model için de ortalama tek parametrelidir (Wedderburn, 1974:439-447).

Farz edilen Y~Poi(μ,θ), gözlemler için kovaryans fonksiyonu i'nin çeşitli gözlemleri için μ_i ortalamasına izin verilmesidir. Çünkü ortalama μ_i>0 ve

$$\mu_i = \exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_P x_{P,i})$$

şeklindedir. Genellikle $\mu=g^{-1}(X\beta)$ parametrelerin ortalama vektörü olarak yazılabilir. Burada g^{-1} örnek bir fonksiyon olarak verilmiştir. $g(\mu)=X\beta$ şeklinde de yazılabilir ki g logaritmik fonksiyondur ve bağ fonksiyonu olarak adlandırılır. Negatif binomial regresyon için de $Y\sim NB(\mu_i, \kappa)$ olup $g(\mu)=X\beta$ bağ fonksiyonudur.

Farklı ortalama ve varyans ilişkilerinden regresyon katsayıları negatif binomial ve quasi-poisson arasında form farkına sebep olabilir çünkü bu modeller çoklu doğrusal

modelin etkisini kullanır ve bu etkiler varyansa ters oranlanır. Böylece negatif binomial ve quasi-poisson gözlemlerle farklı etkilenecektir (VerHoef ve Boveng, 2007:2767).

3.3. NEGATİF BİNOMİAL REGRESYON

Poisson modellerde bağımlı değişkenin ortalama ve varyansının eşit olma durumu genelde sağlanamamakta ve daha çok varyansın ortalamadan büyük olduğu durumlarla karşılaşılmaktadır. Bu durumlarda da negatif binomial regresyon uygulanmaktadır.

Genelde, sayma verilerinin istatistiki bir modeli için temel olarak kullandığı dağılım negatif binomialin parametrelendirilmesidir. Negatif binomial regresyon modeli olasılıklı dağılım fonksiyonu (ODF) altına dayandırılmış olan birçok regresyon modellerinden biridir. Dağılımların karışımı olarak sadece negatif binomial ODF tanımlanabilir.

Geleneksel negatif binomial log bağı, ln (μ), ve üstel ters bağlantı $exp\ (x'\beta)$ almak için kanonik ve ters bağ değerlerini değiştirmektedir. Negatif binom bu formda parametrelendirildiğinde doğrudan Poisson modeli ile ilgili olur.

Sayıma dayalı olarak elde edilen verilerin analizinde, normal dağılım varsayımını sağlamak için kullanılan dönüşümler çoğunlukla yetersiz kalmaktadır. Bundan dolayı üstel dağılım ailesini esas alan negatif binomial regresyon analizleri kullanılmaktadır.

Poisson dağılımın aşırı yayılım gösterdiği durumlarda ortalamanın kuadratik bir fonksiyonu olarak kabul edilen negatif binomial modeli kullanılmıştır. Verinin aşırı yayılım göstermesine gözlemlenen sıfır değerlerin sayısının, poisson modeli ile ortaya konulan sıfır değerlerini aşması ve gözlenmemiş heterojenlik gibi durumlar neden olmaktadır.

Modeldeki aşırı yayılım katsayı tahminini etkilemez, fakat tahminin standart hatasını etkisi altında olmaya sebep verir, böylece modelin güvenirliğini yükseltir.

Negatif binomial regresyon, poisson regresyona alternatif olarak kullanılmaktadır. Zira bu iki yöntem aynı bağlantı fonksiyonunu (log) kullanarak model uyumu yapar. Aynı veri setine poisson ve negatif binom regresyonu uygulandığında, bazen negatif binom ile yapılan uyumda sapma ölçütü daha küçük olmaktadır (Lawless, 1987:209).

Geleneksel negatif binomial model standart maksimum olasılık fonksiyonu kullanılarak tahmin edilebilir ya da genelleştirilmiş doğrusal modeller ilesinin bir üyesi olarak tahmin edilebilir (Hilbe, 2007:3).

Bağımlı değişkenin dağılımı negatif binom dağılımıyla değiştirilir. Böylece "Poisson" dağılımındakinden daha fazla yayılım olması sağlanır.

Bireyler arasındaki gözlemlenemeyen heterojenlik için kullanılan model negatif binomial regresyon modelidir. Bu model bireyler arasındaki diğer heterojenlik kaynaklarının modellenmesine izin vermektedir.

Varyansın farklı değer alabilmesine yönelik olarak parametre tahminlerinin etkinliğini sağlayan bir model olarak geliştirilmiştir olan negatif binomial regresyon varyansın ortalamadan büyük değer aldığı durumlarda kullanılan dağılımlardan biridir.

Farz edelim ki;

$$y \sim Poisson\ (\lambda)$$

ve λ, gama dağılımına sahip rastgele bir değişken olsun.

$$y/\lambda \sim Poisson\ (\lambda)$$

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

Gamma (α,β) ortalaması αβ ve varyansı $\alpha\beta^2$olan gama dağılımıdır. Yoğunluk;

$$P(\lambda) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\lambda/\beta)$$

için 0 ve λ>0'dır. Sonra y'nin şartsız dağılımının negatif binom olduğu görülür.

Bu dağılımın fonksiyonu aşağıdaki gibidir.

$$P(y)=\frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\,y!}\left(\frac{\beta}{1+\beta}\right)^{y}\left(\frac{1}{1+\beta}\right)^{\alpha},\ y=0,1,...,P,\alpha\rangle 0$$

Dağılımın ortalama ve varyansı aşağıdaki gibidir (Lawless, 1987:210).

α aşırı yayılımın ölçüsüdür. Eğer α=0 ise model basit poisson regresyona indirgenir.

$$E\{Y\}=E\{E(y/\lambda\})=E\{E(y/\lambda)\}=\alpha\beta$$

$$\sigma^2\{y\}=E\{\sigma^2\{y/\lambda\}\}+\sigma^2\{E(y/\lambda)\}$$

$$=\sigma^2\{\lambda\}+E\{\lambda\}$$

$$=\alpha\beta+\alpha\beta^2$$

Bir regresyon modeli oluşturmak için μ=αβ ve κ=1/α parametrelerinin terimleri ile negatif binom dağılımı oluşturulabilir. Bu yüzden,

$$E\{Y\}=\mu \text{ ve } \sigma^2\{y\}=\mu+\kappa\mu^2$$

dir. y'nin dağılımı;

$$P(y)=\frac{\Gamma\left(\kappa^{-1}+y\right)}{\Gamma\left(\kappa^{-1}\right)y!}\left(\frac{\kappa\mu}{1+\kappa\mu}\right)^{y}\left(\frac{1}{1+\kappa\mu}\right)^{1/\kappa}$$

κ, 0' a yaklaşırken bu ifade *Poisson(μ)*' ye yaklaşır.

Regresyon amacı için;

$$y\sim Negbin(\mu_i\kappa)$$

farz edelim ve log-bağ fonksiyonu;

$$\log\mu_i=\eta=x_i^{\tau}\beta$$

olup, negatif binomial dağılım da poisson regresyon ile aynı bağ fonksiyonunu kullandığı için aşağıdaki gibidir.

$$g(\mu)=\text{Log}(\mu)=b_0 + b_1x_1 + \cdots + b_mx_m$$

$$\mu=e^{b_0+b_1x_1+\cdots+b_mx_m}$$

$$\mu=e^{x'\beta} \qquad (x' = [1 \;\; X_1 \;\; X_2 \;\; X_3])$$

$$e_i = \frac{Y_i - \hat{\mu}_i}{\sqrt{V(Y_i)}} = \frac{Y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \hat{\mu}_i^2 / \hat{k}}} \qquad X^2 = \sum e_i^2$$

Poisson-Gamma karma dağılımının diğer bir ismi "Negatif Binom Dağılım" olarak da ifade edilmektedir (Cameron ve Trivedi, 1986:61).

3.3.1. Negatif Binomial Regresyon Türleri

Gerçekte bir kaç negatif binomial regresyon modeli vardır ve her biri negatif binomial modele bağlanır. Negatif binomial dağılım için 13 ayrı tip tanımlamıştır Boswell ve Patil (1970). Diğer bilim adamları ile daha fazla tür olduğu konusunda tartışılmıştır. Genel olarak sayma verili istatistiksel modeller için kullanılır. Poisson model de sıfıra yakın α değerine sahip bir negatif binomial modeldir. Yani poisson model negatif binomial 'in bir türüdür. Negatif binomial modelin özel türleri ise;

0) NB2 V= $\mu+\alpha\mu^2$
1) NB1 V= $\mu+\alpha\mu$
2) NB-C < Kanonik>
3) Kesikli-NB;NB1;NB-C
4) Sıfır-değer ağırlıklı NB (ZINB)
5) Engelli-NB
6) NB-Hurdle
7) NB-P
8) NB-H <heterojen>
9) Sonlu Karma Modeller

10) Koşullu sabit etki
11) Rasgele etki
12) Karışık ve çok düzeyli etkiler
13) NB endojen tabakalaşma
14) Basit seçim Negatif Binomial
15) Latent sınıf Negatif Binomial
16) Kantil sayı
17) Momentlerin genelleştirilmiş metodu
18) Araç değişkenler
19) İki değişkenli Negatif Binomial

şeklinde sıralanabilir (Hilbe,2007:187).

Negatif binomial regresyon modeli yaygın NB2 olarak gösterilir (Cameron ve Trivedi, 1986:61). Poisson-gama karışım dağılımından türetilmiştir.

NB2 varyans fonksiyonu ve kanonik formun karakteristikleri;

$$\mu+\alpha\mu^2$$

şeklinde olup bu değer toplanmış ikinci türev ya da $b''(\theta)$ olarak belirlenebilir.

Lineer negatif binomial yapılı NB, varyans olarak;

$$\mu+\alpha\mu$$

parametrelendirilmiştir. NB1 modeli de Poisson-gamma karışımın bir formu olarak türetilmiştir. Ayrıca genelleştirilmiş negatif binomial;

$$\mu+\alpha\mu^p$$

olarak formüle edilebilir. p tahmin edilmiş olan üçüncü parametredir. NB1 için p=1; NB2 için, p=2 dir. p'nin akla yatkın her değeri için genelleştirilmiş negatif binomial sağlar. Genel olarak NB1 ve NB2 kullanılmaktadır.

Negatif binomialin diğer formu heterojen negatif binomial (NB-H) olarak isimlendirilir. NB-H tipik bir NB2 modelidir fakat α ile parametrelendirilmiştir.

Tahminlerin ikinci bir tablosu olarak temsil edilir. Verideki dağılımın miktarı üzerindeki tahmincilerin etkisi için katsayıları gösterir. Negatif binomial regresyonun diğer türlerin isimleri ise şöyle sıralanabilir (Hilbe,2007:4).

Negatif binomial hurdle yöntemi sayıma dayalı olarak elde edilen bağımlı değişkenin sıfır ya da sıfır değerli olmama sonuçlarını belirleyen binomial olasılık modeli ile pozitif sonuçları tanımlayan sınırlandırılmış sayıma dayalı modeli için verilen log olabilirlik fonksiyonu aşağıdaki gibi yazılabilir.

$$\mathrm{L} = \ln(f(0)) + \left\{\ln\left[1 - f(0)\right] + \ln \mathrm{P(j)}\right\}$$

verilen f(0) modelin binary kısmının olasılığını göstermektedir. P(j) pozitif sayımın olasılığını göstermektedir. Poisson ve negatif binomial hurdle modellerin farklı bağlantı fonksiyonları kullanılarak oluşturulan modellere ait akaike bilgi ölçütleri verilmiştir. Hangi modelin veri kümesini en iyi açıkladığı (ABK) bilgi ölçütüne göre karar verilebilir. Hurdle model iki aşamadan oluşmaktadır. Birincisi, sıfır sayımlara (0) karşı pozitif sayımları (1) gösteren cevaplar; ikincisi ise yalnız pozitif sayımların meydana geldiği süreçtir. Binary cevaplar binary model kullanılarak modellenmektedir.

Veri kümesinde beklenenden fazla sıfır değerin bulunması durumunda veri kümesinin sıfırları göz önünde bulunduran sıfır değer ağırlıklı modeller (zero-inflated models) ile analiz edilmesi daha uygun olmaktadır (Hilbe,2007:4).

3.3.1.1. NB1 Modeli

NB1 modelinde kullanılan varyans fonksiyonu ile doğrusal regresyon yöntemlerinde kullanılan varyans fonksiyonu benzerlik göstermektedir.

$$E\left[y_i;\mu_i,\psi\right] = \mu_i \quad \text{ve} \quad VAR\left[y_i;\mu_i,\psi\right] = \mu_i + \frac{\mu_i}{\psi}$$

NB1 log-olabilirlik fonksiyonu şeklinde ifade edilir.

$$\ln L\ (\psi,\ \beta) = \sum_{i=1}^{n} \left\{ \left(\sum_{j=0}^{y-1} \ln\ (j + \psi\ \mu_i) - \ln\ y_i! - (y_i + \psi\ \mu_i) \ln\left(1 + \psi^{-1}\right) + y_i \ln\ \psi^{-1} \right) \right\}$$

$$\sum_{i=1}^{n} \left\{ \left(\sum_{j=0}^{y-1} \frac{\psi^{-1}\mu_i}{j + \psi^{-1}\mu_i} \right) x_i + \psi^{-1}\mu_i x_i \right\} = 0$$

$$\frac{1}{\psi^2} \left\{ -\left(\sum_{j=0}^{y-1} \frac{\mu_i}{j + \psi} \right) - \psi^{-1}\mu_i \ln + (1 + \psi) - \frac{1}{1 + \psi} + y\psi \right\} = 0$$

NB1 tahmininin ilk sıra koşullu çözümüdür. İlk iki momentine dayalı tahminler en çok olabilirlik tahmini diye adlandırılan Poisson, NB1 en çok olabilirlik tahminlerini verir. NB1 genellikle varyans yakalama konusunda esnektir bu yüzden istatistikçiler ve analistler çok sık kullanmaz (Cameron ve Trivedi, 1998:66)

Poisson regresyon modelde y_i ,eşit ortalama ve varyansa sahiptir. Uygulamada genellikle doğrulanamadığından dolayı ortalamanın $\exp(x_i'\beta)$ varsayımı kabul edilir. Bu nedenle y_i'nin koşullu varyansı için genel gösterim;

$$\omega_j = V\left[{}^{y_j}/_{x_j} \right]$$

biçimindedir.

Ortalamanın bir fonksiyonu olarak varyans modeli

$$\omega_j = \omega\ (\ \mu, \alpha\ )$$

şeklinde formüllenir. α skaler bir parametredir. Çoğu modeller, aşağıda yer alan genel varyans fonksiyonu üzerine kurulur. p belirtilen bir sabit olmak üzere;

$$\omega_j = \ \mu_i + \alpha\mu_j^p$$

NB1 varyans fonksiyonu p=1 olduğu durumlarda belirlenir ve varyans fonksiyonu;

$$\omega_j = \ (1 + \alpha\mu_i)$$

olarak yazılır (Cameron ve Trivedi, 1998:62-63).

3.3.1.2. NB2 Modeli

NB1 modeline benzerdir. Negatif binomial dağılımın en genel uygulaması olarak verilen $\mu+ \kappa\mu^2$ varyans fonksiyonu NB2 modelidir. Modelin ortalama ve varyansı aşağıdaki gibidir;

$$E\left[y_i;\mu_i,\psi\right]=\mu_i \quad \text{ve} \quad VAR\left[y_i;\mu_i,\psi\right]=\mu_i+\frac{\mu_i^2}{\psi}$$

NB2 modelinin olasılık yoğunluk fonksiyonu;

$$f(y/\mu,\psi)=\frac{\Gamma\left(\psi+y_i\right)}{\Gamma\left(\psi\right)\Gamma(y_i+1)}\left(\frac{\psi}{\psi+\mu_i}\right)^{\psi}\left(\frac{\mu_i}{\psi+\mu_i}\right)^{y_i}$$

olarak bulunur.

Burada yer alan Γ aşağıdaki özelliğe sahip gama fonksiyonudur.

$$\Gamma(\alpha)=\int_0^{\infty}e^{-t}t^{\alpha-1}dt, \qquad \alpha>0$$

Söz edilen üstel ortalama $\mu_i = \exp(x_i\beta)$ için log olabilirlik fonksiyonu aşağıdaki gibi tanımlanır.

$$\ln L\ \left(\psi,\beta\right)=\sum_{i=1}^{n}\left\{\left(\sum_{j=0}^{y-1}\ln\left(j+\psi\right)\right)-\ln y_i!-\left(y_i+\psi\right)\ln\left(1+\psi^{-1}\mu_i\right)+y_i\ln\psi^{-1}+y_i\ln\mu_i\right\}$$

NB2 en çok olabilirlik fonksiyonu,

$$\sum_{j=1}^{n}\frac{\left(y_i+\mu_i\right)}{1+\psi\mu_i}x_i=0$$

$$\sum_{i=1}^{n}\left\{\frac{1}{\psi^{2}}\left(\ln\left(1+\psi\mu_{i}\right)-\sum_{j=0}^{y-1}\frac{1}{\left(j+\psi^{-1}\right)}\right)+\frac{\mu_{i}-x_{i}}{\psi\left(1+\psi\mu_{i}\right)}\right\}=0$$

Şeklinde olup ilk sıra koşulunun çözümüdür (Cameron ve Trivedi, 1998:71-72).

NB2 karesel varyans fonksiyonu p=2 olduğu durumlarda belirlenir ve varyans fonksiyonu;

$$\omega_j = \mu_i + \alpha\mu_j^2$$

Böylece α yayılım parametresi tahmin edilir.

3.3.2. Negatif Binomial Regresyon Katsayı Tahmini: Maksimum Olasılık

Joseph Felsenstein tarafından 1981 yılında alternatif olarak ortaya konulmuş, istatistik ve ekonometride en sık kullanılan nokta tahmin yöntemlerinden biridir. Çok iyi temellendirilmiş istatistiğe dayandığından matematiksel olarak daha zahmetlidir. Mevcut modeller içinde en tutarlı olandır. Karakter ve Oran analizlerinde kullanılabilirler.

Anakütleyi betimleyen olasılık yoğunluk fonksiyonunu $f(x,\theta)$ ile gösterildiğinde bu anakütleden çekilmiş n gözlemli X_1 , X_2 ,…, X_n bunun belli bir gerçekleşmesi ise x_1 , x_2 , …, x_n olsun. Maksimum olabilirlik tahmin edicileri olabilirlik fonksiyonunu en yükseğe çıkaran tahmin ediciler olarak tanımlanır. Anakütlenin dağılımının ne olduğu biliniyorsa bu aşağıdaki matematiksel probleme dönüşür. Burada olabilirlik fonksiyonu x verilmişken θ' yı bilinmeyen olarak ifade eden bir fonksiyondur.

$$\max_{\theta} L\left(\theta / x\right)=\prod_{i=1}^{n} f\left(x_i;\theta\right)$$

Bu maksimizasyon probleminin çözümünde kolaylık sağlaması için, ortak olasılık fonksiyonunun e tabanına göre logaritması (ln, doğal log) kullanılabilir:

$$\max_{\theta}\ \ln L\left(\theta / x\right)=\ln\left(\prod_{i=1}^{n} f\left(x_i;\theta\right)\right)=\sum_{i=1}^{n}\ln\left(f\left(x_i;\theta\right)\right)$$

Bu maksimizasyon probleminin çözümü için gerekli ve yeterli koşullar:

$$\frac{\partial}{\partial\theta}\ln\ \mathrm{L}\left(\theta/x\right)=0\quad\text{ve}\quad\frac{\partial^2}{\partial\theta^2}\ln\ \mathrm{L}\left(\theta/x\right)<0$$

θ, k bilinmeyen parametreden oluşuyorsa log-olabilirlik fonksiyonunun bu parametrelere göre birinci türevleri sıfır, ikinci türev matrisi (Hessian) negatif belirli olmalıdır.

Negatif binomial regresyonun katsayılarını tahmin etmede de kullanılan yöntemlerden olan maksimum olasılık tahmin yöntemi (MLE) belli bir örneklem değerlerinin gerçekleşme olabilirliğini en yüksek yapan anakütle parametrelerini bulmaya çalışır.

Maksimum Olasılık Tahminci Özellikleri

1. Çoklu doğrusal regresyon tahmincilerinin özellikleri ile aynı olup yansızdırlar ve bütün yansız doğrusal tahminciler arasında minimum varyansa sahiptirler.

2. Normal dağılan hata terimli regresyon modeli için maksimum olabilirlik tahmincileri diğer özelliklere de sahiptir. Tutarlı ve yeterlidirler.

3. Yansız minimum varyansa sahiptirler. Bu, doğrusal olan ya da olmayan bütün yansız tahminciler içinde minimum varyansa sahip oldukları anlamına gelmektedir.

Negatif binomial regresyon modelinin katsayıları birinci mertebeden koşullu alınarak ve sıfıra eşitlenerek tahmin edilir. İki tane birinci dereceden denkleme sahiptir. Biri model katsayıları için diğeri ise parametre içindir.

$$\sum_{i=1}^{n}\left\{\frac{y_i-\mu_i}{\left(1+\psi^{-1}\mu_i\right)}x_i=0\right\}$$

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

x_i kovaryans vektörüdür (Park ve Lord, 2005:D-7).

3.3.3. Aşırı Yayılım Testleri

Poisson regresyon analiz gücünün zayıf olmasından dolayı hata terim değeri büyük değerler almaktadır. Modelin açıklama gücünün zayıf olması ise aşırı yayılımdan kaynaklanmaktadır. Aşırı yayılımdan kaynaklanan uyum eksikliği ya da regresyon fonksiyonunun yetersizliği poisson regresyon da açıklanamayan değişimin en büyük olmasının sebebidir (Özmen, 1998:13).

Eğer koşullu varyans, koşullu ortalamadan büyük ise veriler aşırı yayılım göstermiştir. Aşırı yayılım ya da az yayılımın önemli belirtisi, bağımlı sayım verilerinin örneklem ortalama ve varyansının karşılaştırılmasıyla elde edilebilir.

Aşırı yayılımın nedenlerinden bazıları, yanlış bağlantı fonksiyonun kullanılması, gözlem değerlerin birbirlerinden bağımsız olmaması, modelde olması gereken önemli terimlerin modele dahil edilmemesi, gözlem değerleri arasındaki varyasyonun büyük olması ve örnek büyüklüğünün yetersiz olması şeklinde verilebilir (Cameron ve Trivedi, 1998:70-71).

Aşırı yayılım testi için 3 istatistik tekniği vardır. Bunlar; Olabilirlik Oranı, Wald ve Lagrange Çarpanı testleridir. Bu testler sınırlamalar için kullanılır. Bu testlerde temel hipotez sınırlamanın veya sınırlamaların geçerli olduğu, alternatif hipotez ise sınırlama veya sınırlamaların geçerli olmadığı şeklindedir (Güriş ve Çağlayan, 2005:377). Ayrıca sadece doğrusal olmayan sınırlamalar için geçerli olmayıp, doğrusal sınırlamalar için de geçerlidir.

3.3.3.1. Benzerlik Oran Testi

Benzerlik Oranı Testi sınırlandırılmış ve sınırlandırılmamış en çok benzerlik tahminlerine dayanmaktadır. Benzerlik yönteminde benzerlik fonksiyonunun logaritmasının tahmin edilecek parametrelere göre kısmi türevi alınarak, fonksiyon maksimize edilir.

Test için sınırlandırılmış modelin tahmini de yapılır ve logaritmik benzerlik fonksiyonunu eğiminin sıfır veya sıfırdan farklı olması durumuna göre sınırlamaların geçerli olup olmayacağına karar verilir. Sınırlandırılmış tahmin için logaritmik benzerlik fonksiyonunu l_R, sınırlandırılmamış tahmin için benzerlik fonksiyonu l_U, ile

ifade edilsin. Sınırlı ve sınırsız tahmin için söz konusu fonksiyonlar maksimize edilerek parametreler tahmin edilecektir. Sınırlı tahmin için belirlenecek l_R, sınırsız tahmin için belirlenecek l_U'dan küçük olacağından fonksiyonların birbirine oranı l_R/l_U 0 ile 1 arasında değer alacaktır. Bu oran benzerlik oranı olarak adlandırılmaktadır.

Fonksiyon logaritmalarını L ile,

$$L_R = logl_R$$

$$L_U = logl_U$$

olarak tanımlanırsa, l_R/l_U oranı birden küçük olacağından logaritmaları, yani (L_R-L_U) negatif olacak ve fark anlamlı ise sınırlama reddedilecektir. Büyük örnekler için,

$$LR= -2 (L_R-L_U)$$

olarak hesaplanır. LR ifadesinin dağılımı h serbestlik dereceli ki-kare dağılımıdır. h sınırlama sayısıdır. Bu değer α hata payı ve h serbestlik derecesi ile ki-kare tablosundan bulunacak kritik değerden büyük ise H_1 hipotezi, küçükse H_0 hipotezi kabul edilir, H_0 hipotezi kabul edilirse sınırlamalar geçerlidir. Aksi söz konusu ise sınırlamalar geçersizdir (Güriş ve Çağlayan, 2005:377).

3.3.3.2. Lagrange Çarpanı Testi

Lagrange Çarpanı Testi sınırlı tahmini gerektirmektedir. En çok benzerlik yönteminde de Lagrange fonksiyonundan yararlanılmakta ve sınırlamaların geçerli olup olmamasına göre Lagrange fonksiyonu etkilenmektedir. Sınırlama geçerli ise Lagrange çarpanının değeri küçük olacaktır. Bu testin temeli sınırlamaların geçeli olması durumunda Lagrange çarpanlarının değeri yeterince büyükse, sınırlamaların reddine dayanır. Lagrange Çarpanı Test istatistiğinde varyans sınırlı modelden hesaplanmaktadır. Bu testte de dağılımı h, sınırlama sayısı serbestlik dereceli ki-kare dağılımıdır.

Lagrange çarpanı testinde logaritmik benzerlik fonksiyonunun eğiminin sıfır veya sıfırdan farklı olması durumlarına göre karar verilmektedir. Sınırlamalar geçersizken, logaritmik benzerlik fonksiyonunun en çok benzerlik yöntemi ile elde edilen tahminlerinde, fonksiyonun eğimi sıfırdır. Sınırlamaların geçerli olması halinde ise eğim sıfırdan farklı olmaktadır. Aradaki farkın istatistiksel olarak anlamlı olması kısıtlamaların geçerli olduğunu belirtmektedir.

Lagrange Çarpanı Test istatistiği LM,

$$LM = \frac{\sum e_{t\,R}^2 - \sum e_{t\,U}^2}{\sum e_{t\,R}^2 / n}$$

olarak hesaplanır (Güriş ve Çağlayan, 2005:378).

Doğrusal sınırlamalar söz konusu olduğunda test istatistiği,

$$LM=nR^2$$

olarak hesaplanabilir. Hipotezler ve hipotezin kabul kararı benzerlik oranı testinde açıklandığı gibidir.

LM testi F testi gibi bağımsız değişken katsayılarının tümünün anlamlılığını test etmek için kullanılabilir. Bu durumda test istatistiği sınırlandırılmamış modelin belirlilik katsayısı R_U^2 kullanılarak,

$$LM=nR_U^2$$

hesaplanır. LM test istatistiğinin dağılımı test edilen parametre sayılı (k-1) serbestlik dereceli ki-kare dağılımıdır.

3.3.3.3. Wald Testi

Wald testi Lagrange çarpanı testi aksine sınırlandırılmamış tahmini gerektirmektedir. Bu test istatistiğinde varyans sınırlandırılmamış modelden hesaplanır. Wald testi ile birden fazla sınırlanma test edilebilir. Bu durumda her bir sınırlama için

ayrı ayrı hesaplama gerekmektedir. Serbestlik derecesi kısıtlama sayısı h olacaktır. Bu testin avantaj sadece sınırlandırılmamış tahmini gerektirmesidir.

Wald test istatistiği ,

$$W = \left(\frac{\hat{\beta}_L . \hat{\beta}_M - 1}{\hat{\sigma}_{\hat{\beta}_L . \hat{\beta}_M}} \right)^2$$

olacaktır. Sınırlama sayısı h=1 olduğunda ki-kare tablosunda 1 serbestlik derecesi ile tablo değeri bulunarak benzerlik oranı testinde olduğu gibi karar verilir. Aynı modelde aynı sınırlamalar için Lagrange çarpanı, Benzerlik oranı, Wald testleri hesaplandığında;

$$LM \leq LR \leq W$$

ilişkisi görülür (Güriş ve Çağlayan, 2005:380).

3.3.4. Regresyon Katsayılarının Anlamlılığının Testi

$\beta_0, \beta_1, \ldots, \beta_k$ şeklinde gösterilmiş olan katsayıların yorumu diğer regresyonlarda olduğu gibi değildir. Çünkü kestirilen değerler, üstel fonksiyon yardımıyla türetilmiştir. Bu yüzden katsayıların anlamlılığını bulmak için hipotez testi oluşturulmalıdır. Katsayıların anlamlılığının testi için kullanılacak hipotezler;

$H_0 : \beta_i = 0$, $(i = 1,..., n)$ (β_i katsayısı anlamsızdır)
$H_0 : \beta_i \neq 0$, $(i = 1,..., n)$ (β_i katsayısı anlamlıdır)

seklindedir. Bu hipotezlerin testinde en sık kullanılan yöntem Wald'ın χ^2 istatistiğidir

$$\chi^2_w = \left(\frac{b_i}{s'_{bi}} \right)^2$$

seklinde hesaplanır. Bu eşitlikte β_i , regresyon katsayılarını; s'_{bi} ise, basit standart hata değerinin ϕ sayısının karekökü ile çarpımı yardımıyla elde edilir.

$$s_{bi}^{'} = s_{b1}\sqrt{\phi}$$

seklinde ifade edilir. Böylece düzeltilmiş standart hata değerine ulaşılır. ϕ sayısı ise, *k* kestirilecek parametre sayısı olmak üzere;

$$\phi = \frac{1}{n-k}\sum_{i=1}^{n}\frac{\left(y_i-\mu_i\right)^2}{\mu_i}$$

hesaplanan Wald'ın χ^2 istatistik değeri, 1 serbestlik dereceli değeriyle χ^2 karşılaştırılır. Eğer hesaplanan değer tablo değerini asıyorsa H_0 hipotezi reddedilir. Yani katsayıların anlamlı olduğuna karar verilir.

3.3.5. Güven Aralığı

Aşırı yayılım durumunda parametre tahminlerinin ve standart hatalarının sapmalı olmasına neden olmamak için negatif binomial regresyon uygulanmalı Ayrıca regresyon parametreleri için dar güven aralıkların ve çok küçük p değerlerin elde edilmesine neden olmaktadır

Güven aralıkları normal dağılım farz edilen kovaryans matris kullanılarak β ve μ üzerinden hesaplanabilir.

$$\begin{bmatrix}\hat{\beta}\\ \alpha\end{bmatrix} \sim N\left(\begin{bmatrix}\beta\\ \alpha\end{bmatrix},\right)\begin{pmatrix}VAR[\beta] & 0\\ 0 & VAR[\alpha]\end{pmatrix}$$

$$VAR[\beta] = \left(\sum_{i=1}^{n}\frac{\mu_i}{1+\psi^{-1}\mu_i}X_i X_i^{'}\right)^{-1}$$

$$VAR\ [\ \alpha\ ] = \left(\sum_{i=1}^{n}\frac{i}{\left(\psi^{-1}\right)^4}\left(\ln\left(1+\psi^{-1}\mu_i\right)-\sum_{j=0}^{y_{j-1}}\frac{1}{j+\psi}\right)^2+\frac{\mu_i}{\left(\psi^{-1}\right)^2\left(1+\psi^{-1}\mu_i\right)}\right)^{-1}$$

3.3.6. Modelin Değerlendirilmesi

Yaygın olarak kullanılan uygun istatistik modelin değerlendirme yöntemleri; Log-olasılık çarpıklık, Pearson χ^2 çarpıklık, Akaike Bilgi Kriteri (ABK), Bayesian Bilgi Kriteri (BBK) şeklindedir.

Maksimizasyon sürecinin üreticisi olan *L* log olasılığa sahip olunduğundan dolayı çarpıklık D aşağıdaki şekilde;

$$D = 2\sum_{i=1}^{n}\left(L\left(y_i;y_i\right)-L\left(\mu_i;y_i\right)\right)$$

tanımlanır. Burada $L\left(\mu_i;y_i\right)$ log- olasılık fonksiyonudur ve $L\left(y_i;y_i\right)$ μ_i yerine y_i yerleştirilen log-olasılık fonksiyonudur. Bu NB2 modeli için $D=\sum_{i=1}^{n} d_i^2$ 'yi basitleştirir. Burada;

$$d_i^2 = \begin{cases} 2\left(y_i \ln\left(\dfrac{y_i}{\mu_i}\right)-\left(y_i+\dfrac{1}{\alpha}\right)\ln\left(\dfrac{1+\alpha y_i}{1+\alpha\mu_i}\right)\right) & y_i \rangle 0 \\ \dfrac{2}{\alpha}\ln\left(1+\alpha\mu_i\right) & \mathrm{y}_i = 0 \end{cases}$$

dır.

Pearson χ^2 çarpıklık istatistiği $\sum_{i=1}^{n}\dfrac{\left(y_i-\mu_i\right)^2}{\left(\mu_i+\alpha\mu_i^2\right)}$ tarafından verilirken

$$AIC=\frac{2\left(L-\left(p+1\right)\right)}{n} \qquad \text{ve} \qquad BIC = D-\left(n-p-1\right)\ln n$$

şeklinde tanımlanır (Zwilling, 2013:10).

AIC, Akaike tarafından önerilen ve farklı modellerin karşılaştırılmasında yaygın olarak kullanılan ölçüt, Akaike bilgi ölçütü (Akaike Information Criterion: AIC) olarak tanımlanır. En küçük AIC değerine sahip model en iyi model olarak kabul edilmektedir.

Bayes Bilgi Ölçütü, Akaike bilgi ölçütü gibi veriler ve model arasında uygunluğu ölçen yöntemlerden biridir (Ercanlı ve diğ. 2012:82). Pearson istatistiği poisson ve negatif binomial regresyon için aynı formdadır.

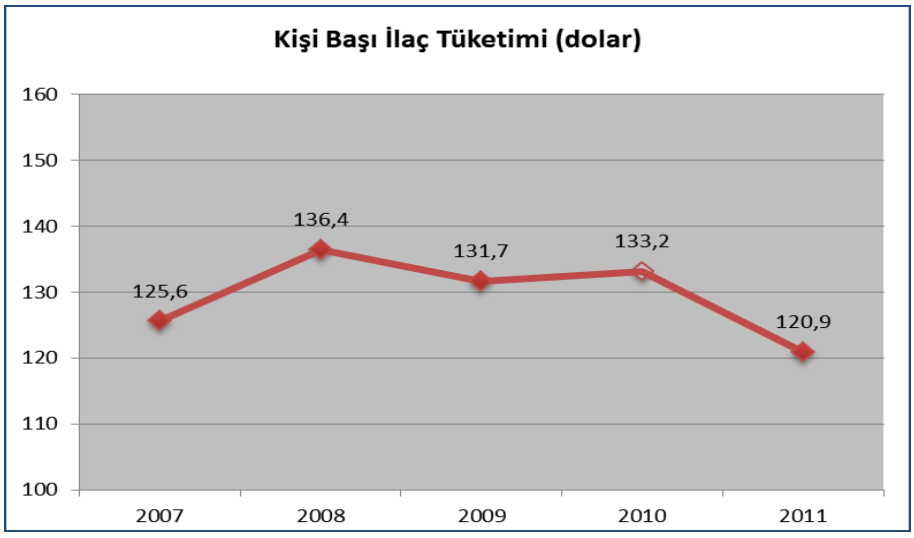
DÖRDÜNCÜ BÖLÜM

BÖLÜM BAŞLIĞI

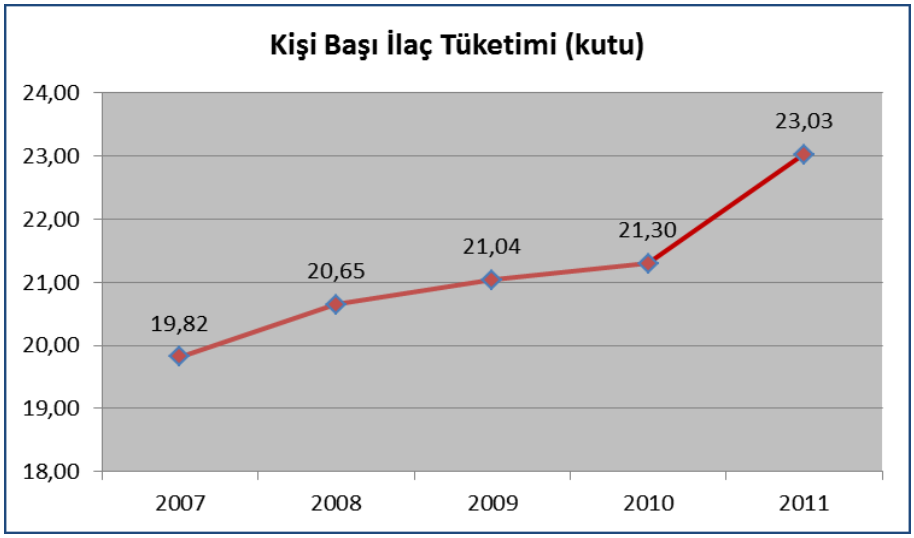
4.1. TÜRKİYE'DE İLAÇ TÜKETİMİ

İlaç insanlarda hastalıklardan korunma, tanı, tedavi veya bir fonksiyonun düzeltilmesi ya da insan yararına değiştirilmesi için kullanılan, genellikle bir veya birden fazla yardımcı madde ile formüle edilmiş etkin madde veya maddeleri içeren bitmiş dozaj şeklidir (Erol ve Dilek, 2001:40). DSÖ: Dünya Sağlık Örgütü ilacı su şekilde tanımlamaktadır; Fizyolojik sistemleri veya patolojik durumları insanın yararı için değiştirmek veya incelemek amacıyla kullanılabilen bir maddedir.

Dünya Sağlık Örgütü'nün araştırmasına göre tüm ilaçların yarısından fazlası uygunsuz şekilde reçete ediliyor, kullanılıyor, dağıtılıyor ya da satılıyor. OECD Sağlık Sistemleri İncelemeleri Türkiye raporuna göre Türkiye'de ilaç giderleri, toplam sağlık harcamalarının yaklaşık üçte birini oluşturmakta olup son 10 yılda üçe katlanan ilaç tüketimi israfın boyutlarını göz önüne sermektedir. Türkiye'de hastalar bu ilaçları serbest eczanelerden temin etmektedirler. Bu yüzden ilaç harcamaları yüksek görünebilmektedir. Türkiye'de kişi başına düşen ilaç harcaması halen OECD ülkeleri içinde oldukça düşük olmasına rağmen gelecek yıllarda kişi başına düşen ilaç harcamalarında artış beklenmektedir.

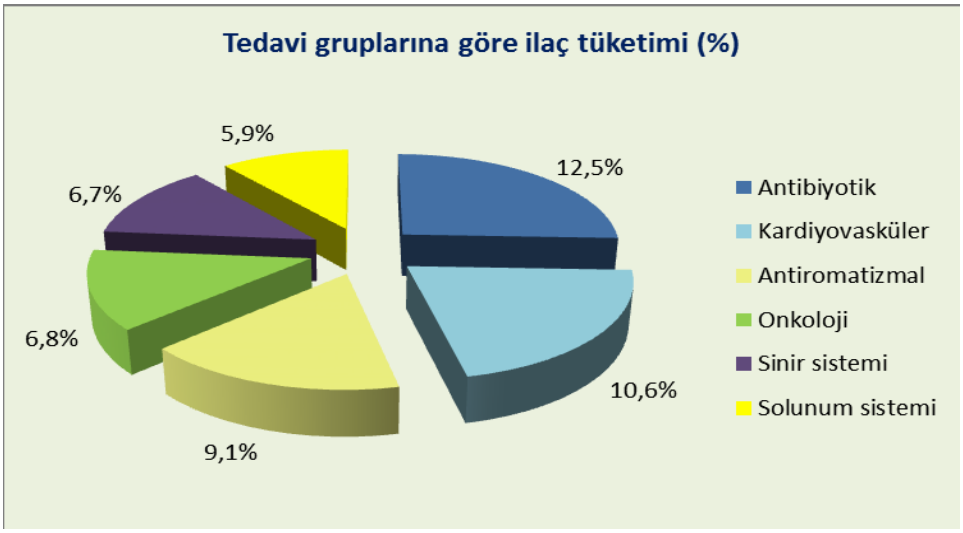


Şekil 4.1. Dolar Bazında Kişi Başı İlaç Tüketimi(IMS)

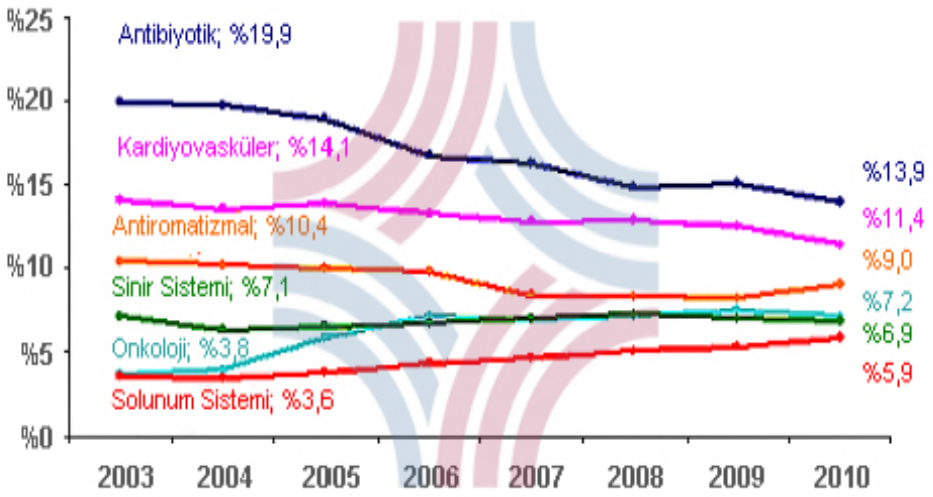


Şekil 4.2. Kutu Bazında Kişi Başı İlaç Tüketimi(IMS)

Sosyal güvenlik kurumu (SGK) 1003 hane ile tek tek görüşme yaptığında 190 milyon kutu ilacı stokta tutulduğu sonucuna ulaşmış oldu. Bu ilaçların 10 milyonu henüz açılmamış. Her evde ortalama 11 kutu ilaç bulunuyor. Nüfusun yüzde 5'inin evinde ise ortalama 30'un üzerinde ilaç bulunuyor. Üçte birinin ise ancak 1-2 adedi kullanılmış. Yüzde 8'inin son kullanma tarihi geçerken vatandaş depoladığı 67 milyon kutu ilacı ihtiyaç halinde kullanmayı düşünüyor. 28 milyon kutu ilacı ise hiç kullanmayı düşünmüyor. Evlerde depolananlar arasında ağrı kesici, antibiyotik, mide ve soğuk algınlığı ilaçları başı çekiyor (SGK).



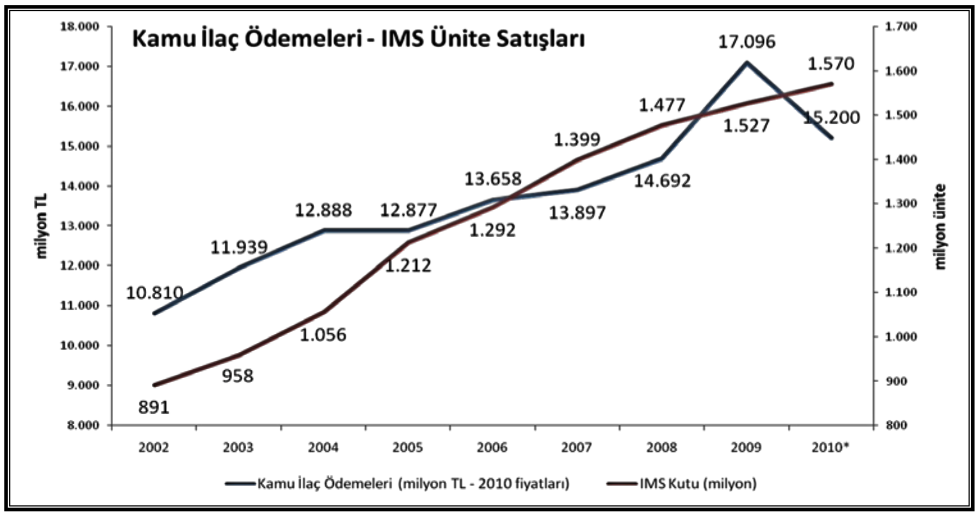
Şekil 4.3. Türkiye İlaç Pazarında Tedavi Gruplarına Göre İlaç Tüketimi(IMS)



Şekil 4.4. Ülkemiz İlaç Pazarında 2003-2010 Yılları Arasında Değişim(IMS)

SGK Türkiye'deki gereksiz ilaç harcamalarının azaltılması için hastane ve eczane bilgi sistemlerinde reformlar yapmasına rağmen, ilaç harcamaları artmaya devam ediyor. 2011 yılında Türkiye'de kullanılan ilaç miktarı 2010 yılına göre yüzde 9,7 oranında artarak 1,7 milyar kutuya ulaştı. İlaç harcamaları ise yüzde 2,7 oranında yükselerek 15,1 milyar TL'ye çıktı. (http://www.farmasyon. com.tr/makale/122/turkiye_ ilac_pazarindan_carpici_rakamlar).

Yılda 160 milyon kutu antibiyotiğe 1.25 milyar TL, 300 milyon kutu ağrı kesiciye 950 milyon lira ödeniyor. Devletin yılda 16 milyar liralık ilaç harcamasının 2 milyarı israf oluyor (Ünal, 2013).



Şekil 4.5. Kamu Tarafından Yapılan İlaç Ödemeleri (IMS)

Son on yılda girilen reform süreci ile birlikte kamu ilaç harcamaları ilaç piyasasının şekillenmesinde en önemli faktör haline gelmiştir. 2009 yılı Aralık ayında global bütçe protokolü imzalanarak 2010-2012 yılları için kamu ilaç bütçesi sabitlenmiş, pazarın yaklaşık %20 daralmasına yol açacak yeni fiyat ve ıskonto düzenlemeleri ile bu yıllardaki hedeflerin tutturulması amaçlanmıştır (Teksöz ve diğ, 2012:).

İlaçla ilgili göz önünde tutulması gereken temel yaklaşım gerektiği zaman, gereken nitelikte, gerektiği kadar ve gerektiği biçimde kullanılabilmesidir. Bu anlamda ilacın sağlık hizmetlerinde vazgeçilmez bir önemi olmakla birlikte, aynı zamanda sağlık sorunları içerisinde de büyük bir yeri bulunmaktadır. Günümüzde ilaç üretim, tüketim, fiyat vb. açılardan toplum sağlığını yakından ilgilendirmektedir (Şemin, 1998:9).

İllerin kullandığı ilaç sayısı bağımlı değişken kabul edilerek özel ya da devlet ayırt etmeksizin illerin hastane sayısı, hekim sayısı, eczacı sayısı ve illerin nüfus sayısı, iller üzerindeki deniz etkisi ve illerin nüfus artış hızı değişkenleri ise bağımsız değişken olarak alındı. Bu değişkenlerden nüfus artış hızı, deniz etkisi ve hastane sayısı hariç diğer değişkenlerin logaritması alındı. Kantil regresyon ve negatif binomial analiz yöntemleri ile karşılaştırıldı. Kantil regresyon analizinde kantil değerleri olarak %10, %50, %75, %90 kantillerine de bakılarak katsayıları ve anlamlılığı yorumlanarak, kantil regresyon yöntemi ile yığılmanın nerede daha çok olduğuna da bakılmıştır.

İlaç sayısının ortalama ve varyans sonuçları incelendiği zaman poisson dağılıma uygun olmadığı ancak varyansın ortalamadan büyük olması sebebi ile negatif binomial regresyon için uygun olduğu görülmüştür. Negatif Binomial Regresyon ile analizi yapılarak elde edilen veriler diğer analizlerle karşılaştırılmıştır.

2010-2011-2012 ve 2013 yılları esas alınarak il il kullanılan ilaç sayısı IMS (İnternational Medical Statistics) Türkiye ofisinden temin edilmiştir. TÜİK' den sağlık istatistikleri bölümünde bu yıllara ait hekim sayısı, eczacı sayısı, hastane sayısı, nüfus artış hızları ve illerin nüfus bilgileri alınmıştır. Hastane sayısı ve nüfus artış hızı hariç diğer değişkenlerin logaritması alındı. Deniz etkisi değişkeni kukla değişken olarak alınmış olup denize kıyısı bulunan illere 1 diğer illere 0 değeri verilmiştir.

Y; İl Bazında Kullanılan İlaç Sayısı

X_1; İllerin Nüfus Miktarının Logaritması

X_2; İl Bazında Hekim Sayısının Logaritması

X_3; İl Bazında Eczacı Sayısının logaritması

X_4; İllerin Nüfus Artış Hızları

X_5; İllerin Deniz Etkisi

X_6; İl Bazında Hastane Sayısı

Aşağıda illerde tüketilen ilaç sayılarına ait tanımlayıcı istatistikler verilmiştir. Standart sapma, veri değerlerinin yayılımının özetlenmesi için kullanılan bir ölçü olup verilerin ortalamadan sapmalarının kareler ortalamasının karekökü olarak tanımlanır ve standart sapma varyansın kare köküdür. Çarpıklık ve basıklık katsayısı ise simetrik durumdan uzaklaşma derecesini belirlemek için kullanılmaktadır. Çarpıklık katsayısının değeri (-1;1) arasında değişmeli ve simetrik bir dağılım için sıfıra eşit olmalıdır. Çarpıklık katsayısı sağa eğik dağılımlarda pozitif, sola eğik dağılımlarda negatif değerler almaktadır. Basıklık katsayısı ise dağılımın sivri ya da basıklığını belirlemektedir. Normal bir dağılımda basıklık katsayısının değeri 3'tür. Bu değerler yardımıyla dağılımın normal dağılım olmadığı söylenebilir.

Tablo 4.1. 2010-2011-2012-2013 Yılı Tanımlayıcı İstatistikler

	Ortalama	Varyans	Standart Sapma	Basıklık	Çarpıklık
2010	193.10^5	$13.\ 10^{14}$	$37.\ 10^6$	45,89	6,22
2011	$212.\ 10^5$	$17.\ 10^{14}$	$41.\ 10^6$	46,18	6,25
2012	$218.\ 10^5$	$18.\ 10^{14}$	$43.\ 10^6$	46,94	6,31
2013	$219.\ 10^5$	$18.\ 10^{14}$	$43.\ 10^6$	46,97	6,31

İlk olarak 2010 yılı incelemeye alınmış olup NB1, NB2 ve kantil regresyon analiz yöntemleri uygulanmıştır. Kantil regresyon analiz yönteminde ara kantil değerleri olan %10, %50, %75 ve %90 kantil dilimleri de ayrı ayrı analiz yapılarak dahil edilmiştir.

Negatif binomial regresyonlarda anlamlı olan değişkenler yıllara göre farklılıklar göstermektedir. Bağımlı değişken üzerindeki bağımsız değişkenlerin etkisi elde edilen katsayı değerleri tarafından belirlendiğinden log-nüfus, log-hekim, log-eczane, nüfus artış hızı, deniz etkisi ya da hastane sayısında meydana gelecek bir birimlik artışın log ilaç sayısında meydana gelecek artma ya da azalmayı ifade etmesi şeklinde yorumlanmaktadır. NB1 analizinde log-nüfus, log hekim, log-eczacı ve deniz etkisi değişkenleri anlamlı olarak elde edilmiştir. NB2 analizinde ise log-nüfus, log-eczacı ve deniz etkisi değişkenleri anlamlı olarak elde edilmiştir. Kantil regresyonda katsayıların anlamlılığı sıradan regresyonun katsayılarının anlamlılığı ile aynı yorumlanmaktadır. Kantil regresyon analizlerinde ise sadece hastane değişkeni anlamlı olarak elde edilmiştir.

Tablo 4.2. 2010 Verilerine Ait Analiz Sonuçları

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
Log-nüfus	1,591*** (0,102)	1,207*** (0,111)	28.10^5 ($2{,}4.10^7$)	10448 (11.10^7)	38.10^5 (85.10^5)	-51.10^5 (23.10^7)
Log-hekim	-0,323*** (0,114)	-0,139 (0,118)	-32.10^5 ($2{,}9.10^7$)	-45.10^5 (12.10^7)	22.10^4 (80.10^5)	23.10^{-7} (20.10^7)
Log-eczacı	1,016*** (0,091)	1,284*** (0,0844)	-12.10^{-7} ($1{,}2.10^7$)	34.10^5 (88.10^5)	-17.10^5 (68.10^5)	-97.10^5 (11.10^7)
Nüfus hızı	56.10^{-5} (7.10^{-4})	83.10^{-5} (5.10^{-4})	42.10^3 (15.10^4)	6358 (42.10^3)	28.10^3 (27.10^3)	17.10^3 (64.10^3)
Deniz etkisi	0,112*** (0,0209)	0,102*** (0,025)	63.10^5 (55.10^4)	94.10^4 (26.10^5)	24.10^5 (18.10^5)	38.10^5 (39.10^5)
Hastane	18.10^{-5} (2.10^{-4})	-91.10^{-5} (5.10^{-4})	15.10^5*** (48.10^3)	14.10^5*** (28.10^3)	14.10^5*** (26.10^3)	13.10^5*** (90.10^3)
Sabit	5,698 (0,388)	6,790 (0,387)	32.10^5 ($8{,}1.10^7$)	-76.10^5 (39.10^7)	-20.10^{-7} (31.10^7)	-97.10^5 (39.10^7)

(***=p<0,01,**=p<0,05,*=p<0,10)

2011 yılı verileri dikkate alınarak, NB1, NB2 ve kantil regresyon analizleri uygulanmıştır. Kantil regresyon analiz yönteminde ara kantil değerleri olan %10, %50, %75 ve %90 kantil dilimleri de ayrı ayrı analiz yapılarak dahil edilmiştir.

NB1 analizinde log-nüfus, log hekim, log-eczacı ve deniz etkisi değişkenleri anlamlı olarak elde edilirken, NB2 analizinde log-nüfus, log-eczacı ve deniz etkisi değişkenleri anlamlı olarak elde edilmiştir. Ayrıca log-hekim değişkeni de %10 önem düzeyinde anlamlı olarak bulunmuştur. %10 kantil regresyon analizinde log-hekim değişkeni %10 önem düzeyinde anlamlı olarak elde edilse de, diğer kantil değerlerinde anlamlı olan tek değişken hastane değişkeni olmuştur.

Tablo 4.3. 2011 Verilerine Ait Analiz Sonuçları

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
Log-nüfus	1,321*** (0,100)	1,062*** (0,108)	$15.\ 10^{6}$ $(2.\ 10^{7})$	$-47.\ 10^{5}$ $(13.\ 10^{6})$	$-40.\ 10^{5}$ $(13.\ 10^{6})$	$-53.\ 10^{5}$ $(19.\ 10^{6})$
Log-hekim	-0,294*** (0,101)	-0,181* (0,109)	$-32.\ 10^{6}$* $(1{,}7.\ 10^{7})$	$-13.\ 10^{6}$ $(13.\ 10^{6})$	$46.\ 10^{5}$ $(12.\ 10^{6})$	$73.\ 10^{6}$ $(16.\ 10^{6})$
Log-eczacı	1,301*** (0,089)	1,456*** (0,082)	$62.\ 10^{5}$ $(1{,}9.\ 10^{7})$	$10.\ 10^{6}$ $(10.\ 10^{6})$	$63.\ 10^{5}$ $(83.\ 10^{5})$	$32.\ 10^{5}$ $(77.\ 10^{5})$
Nüfus hızı	$-17.\ 10^{-5}$ $(4.\ 10^{-4})$	$-38.\ 10^{-5}$ $(3.\ 10^{-4})$	$13.\ 10^{4}$ (83.935)	87.354* (49.595)	69.918 51.962	31.718 (40.845)
Deniz etkisi	0,056*** (0,0191)	0,071*** (0,0233)	$15.\ 10^{5}$ $(46.\ 10^{7})$	$-31.\ 10^{4}$ $(29.\ 10^{5})$	$16.\ 10^{5}$ (27.10^{5})	38.10^{5} $(32.\ 10^{5})$
Hastane	$-22.\ 10^{-7}$ $(2.\ 10^{-4})$	$-48.\ 10^{-5}$ $(5.\ 10^{-4})$	$16.\ 10^{5}$ (51.989)	$17.\ 10^{5}$*** 34.348	$15.\ 10^{5}$*** 40.695	$15.\ 10^{5}$*** (85.327)
Sabit	6,634 (0,379)	7,454 (0,378)	$-25.\ 10^{6}$ $(7{,}2.\ 10^{7})$	$33.\ 10^{6}$ $(47.\ 10^{6})$	$-49.\ 10^{5}$ $(46.\ 10^{6})$	$31.\ 10^{5}$ $(77.\ 10^{6})$

(***=p<0,01,**=p<0,05,*=p<0,10)

2012 yılı verileri dikkate alınarak, NB1, NB2 ve kantil regresyon analizleri uygulanmıştır. Kantil regresyon analiz yönteminde ara kantil değerleri olan %10, %50, %75 ve %90 kantil dilimleri de ayrı ayrı analiz yapılarak dahil edilmiştir.

NB1 analizinde log-nüfus, log-eczacı ve deniz etkisi değişkenleri %1 önem düzeyinde anlamlı olarak elde edilirken, log-hekim ve nüfus artış hızı değişkenleri de %5 önem düzeyinde anlamlı olarak elde edilmiştir. NB2 analizinde log-nüfus, log-eczacı ve deniz etkisi değişkenleri %1 önem düzeyinde, nüfus artış hızı değişkeni ise %5 önem düzeyinde anlamlı olarak elde edilmiştir. %10 kantil regresyon analizinde hastane değişkeni anlamlı olarak elde edilirken. %50 kantil regresyon analizinde hiçbir değişken anlamlı bulunmamıştır. %75 ve %90 kantil regresyon analizlerinde sadece hastane değişkeni anlamlı olarak elde edilmiştir.

Tablo 4.4. 2012 Verilerine Ait Analiz Sonuçları

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
Log-nüfus	1,228***	0,958***	14.10^5	67.10^5	-38.10^5	-12.10^6
	(0,100)	(0,108)	(21.10^6)	(10.10^6)	(67.10^5)	(14.10^6)
Log-hekim	-0,218**	-0,125	41.10^5	-16.10^6	56.10^5	18.10^6
	(0,106)	(0,114)	(25.10^6)	(11.10^6)	(69.10^5)	(10.10^6)
Log-eczacı	1,292***	1,496***	-27.10^6	43.10^5	-15.10^4	-23.10^5
	(0,094)	(0,086)	(27.10^6)	(83.10^5)	(51.10^5)	(70.10^5)
Nüfus hızı	2.10^{-3}**	19.10^{-4}**	52.10^3	40.10^3	35.10^3	-12.10^3
	(93.10^{-5})	(9.10^{-4})	(22.10^4)	(95.10^3)	(49.10^3)	(11.10^4)
Deniz etkisi	0,070***	0,085***	57.10^5	29.10^5	10.10^5	45.10^5**
	(0,018)	(0,023)	(47.10^5)	(24.10^5)	(16.10^5)	(21.10^5)
Hastane	14.10^{-5}	(-5.10^{-4})	18.10^5***	16.10^5	16.10^5***	15.10^5***
	(20.10^{-5})	(5.10^{-4})	(30.10^4)	(27.10^3)	(23.10^3)	(57.10^3)
Sabit	6,943	7,770	16.10^6	-10.10^6	27.10^5	21.10^6
	(0,375)	(0,382)	(80.10^6)	(37.10^6)	(25.10^6)	(60.10^6)

(***=p<0,01,**=p<0,05,*=p<0,10)

Son olarak 2013 yılı verileri dikkate alınarak NB1, NB2 ve kantil regresyon analizleri uygulanmıştır. Kantil regresyon analiz yönteminde ara kantil değerleri olan %10, %50, %75 ve %90 kantil dilimleri de ayrı ayrı analiz yapılarak dahil edilmiştir.

NB1 analizinde log-nüfus, log-eczacı, log-hekim ve deniz etkisi değişkenleri %1 önem düzeyinde, nüfus artış hızı değişkeni de %10 önem düzeyinde anlamlı olarak elde edilmiştir. NB2 analizinde log-nüfus, log-eczacı ve deniz etkisi değişkenleri %1 önem düzeyinde, nüfus artış hızı değişkeni ise %5 önem düzeyinde anlamlı olarak elde edilmiştir. Bütün kantil regresyon analizlerinde sadece hastane değişkeni anlamlı olarak elde edilmiştir.

Tablo 4.5. 2013 Verilerine Ait Analiz Sonuçları

	NB1	NB2	%10Kantil	%50Kantil	%75 Kantil	%90 Kantil
Log-nüfus	1,391*** (0,117)	1,005*** (0,117)	$55.\ 10^5$ ($12.\ 10^7$)	$25.\ 10^5$ ($11.\ 10^6$)	$-60.\ 10^5$ ($10.\ 10^6$)	$-13.\ 10^6$ ($18.\ 10^6$)
Log-hekim	-0,310*** (0,128)	-0,138 (0,128)	$-31.\ 10^6$ ($17.\ 10^7$)	$-98.\ 10^5$ ($12.\ 10^6$)	$66.\ 10^5$ ($10.\ 10^6$)	$27.\ 10^6$ ($17.\ 10^6$)
Log-eczacı	1,256*** (0,114)	1,504*** (0,101)	$13.\ 10^6$ ($12.\ 10^7$)	$15.\ 10^5$ ($98.\ 10^5$)	$-33.\ 10^5$ ($83.\ 10^5$)	$-10.\ 10^6$ ($16.\ 10^6$)
Nüfus hızı	$17.\ 10^{-4}$* ($10.\ 10^{-4}$)	$17.\ 10^{-4}$** ($8.\ 10^{-4}$)	9.198 ($12.\ 10^5$)	$32.\ 10^3$ ($87.\ 10^3$)	$63.\ 10^3$ ($65.\ 10^3$)	$11.\ 10^4$ ($82.\ 10^3$)
Deniz etkisi	0,068*** (0,0223)	0,083*** (0,026)	$25.\ 10^5$ ($43.\ 10^6$)	$22.\ 10^5$ ($26.\ 10^5$)	$20.\ 10^5$ ($20.\ 10^5$)	$55.\ 10^5$ ($37.\ 10^5$)
Hastane	$-76.\ 10^{-6}$ ($23.\ 10^{-5}$)	$-92.\ 10^{-5}$ ($59.\ 10^{-5}$)	$17.\ 10^5$*** ($44.\ 10^4$)	$16.\ 10^5$*** ($30.\ 10^3$)	$16.\ 10^5$*** ($29.\ 10^3$)	$15.\ 10^5$*** ($78.\ 10^3$)
Sabit	6,360 (0,439)	7,528 (0,413)	$84.\ 10^5$ ($47.\ 10^7$)	$30.\ 10^4$ ($41.\ 10^6$)	$17.\ 10^6$ ($38.\ 10^6$)	$20.\ 10^6$ ($79.\ 10^6$)

(***=p<0,01,**=p<0,05,*=p<0,10)

İncelenen bütün değişkenlerin anlamlılık değerleri yıl olarak incelemiş olup, her değişken için ayrı anlamlılık tablosu yapılarak aşağıdaki verilmiştir.

İlk olarak Log-nüfus değişkeninin anlamlılık tablosu incelendiğinde analiz sonuçları yıllar arasında değişkenlik göstermeyip log-nüfus değişkeninin NB1 ve NB2 analizleri sonucunda anlamlı bir değişken olarak, kantil regresyon sonuçlarına göre ise anlamsız bir değişken olarak elde edilmiştir.

Tablo 4.6. Log-nüfus Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2011	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2012	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2013	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız

Log-hekim değişkeninin anlamlılık tablosu incelendiğinde log-hekim değişkeninin yıllar arasında değişkenlik gösterdiği görülmektedir. Log-hekim değişkeninin 2010 yılında NB1 regresyon analizinde anlamlı olarak elde edilmesine rağmen, NB2 ve kantil regresyon analizleri sonucunda anlamsız bir değişken olarak elde edildiği görülmektedir. 2011 yılında log-hekim değişken NB1 ve NB2 analizlerinin yanında %10 kantil regresyon analizinde anlamlı, diğer kantil regresyon analizlerinde ise anlamsız olarak elde edilmiştir. 2012-2013 yıllarında ki analiz sonucu incelendiğinde 2010 yılı ile aynı olup sadece NB1 regresyonunda anlamlı, diğer analizlerde ise anlamsız olarak elde edilmiştir.

Tablo 4.6. Log-hekim Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamlı -	Anlamsız	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2011	Anlamlı -	Anlamlı -	Anlamlı -	Anlamsız	Anlamsız	Anlamsız
2012	Anlamlı -	Anlamsız	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2013	Anlamlı -	Anlamsız	Anlamsız	Anlamsız	Anlamsız	Anlamsız

Hastane değişkeninin anlamlılık tablosu incelendiğinde hastane değişkeninin yıllar arasında değişkenlik gösterdiği görülmektedir. 2010 yılında kantil regresyon analizleri sonucunda anlamlı, NB1 ve NB2 analizleri sonucuna göre ise anlamsız olarak elde edildiği görülmektedir. 2011 yılında %50, %75 ve %90 kantil regresyon analizleri sonucunda anlamlı, NB1, NB2 ve %10 kantil regresyon analizlerinde ise anlamsız olduğu görüldü. 2012 yılında %10, %75 ve %90 kantil regresyon analizleri sonucunda anlamlı, NB1, NB2 ve %50 kantil regresyon analizlerinde ise anlamsız olduğu görülmüştür. 2013 yılı analiz sonucu incelendiğinde ise 2010 yılı sonuçları ile aynıdır.

Tablo 4.8. Hastane Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamsız	Anlamsız	Anlamlı +	Anlamlı +	Anlamlı +	Anlamlı +
2011	Anlamsız	Anlamsız	Anlamsız	Anlamlı +	Anlamlı +	Anlamlı +
2012	Anlamsız	Anlamsız	Anlamlı +	Anlamsız	Anlamlı +	Anlamlı +
2013	Anlamsız	Anlamsız	Anlamlı +	Anlamlı +	Anlamlı +	Anlamlı +

Log-eczacı değişkeninin anlamlılık tablosu incelendiğinde analiz sonuçları yıllar arasında değişkenlik göstermeyip log-eczacı değişkeninin NB1 ve NB2 analizleri sonucunda anlamlı bir değişken olarak, kantil regresyon sonuçlarına göre ise anlamsız bir değişken olarak elde edildiği görülmektedir.

Tablo 4.9. Log-eczacı Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2011	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2012	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2013	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız

Deniz etkisi değişkeninin anlamlılık tablosu incelendiğinde ise analiz sonuçları yıllar arasında değişkenlik göstermeyip deniz etkisi değişkeninin log-nüfus ve log eczacı değişkenleri analiz sonucu ile paralellik gösterdiği görüldü. Deniz etkisi değişkeninin NB1 ve NB2 analizleri sonucunda anlamlı bir değişken olarak, kantil regresyon sonuçlarına göre ise anlamsız bir değişken olarak elde edildiği görülmektedir.

Tablo 4.10. Deniz Etkisi Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2011	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2012	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2013	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız

Nüfus artış hızı değişkeninin anlamlılık tablosu incelendiğinde ise analiz sonuçları yıllar arasında farklılıklar olduğunu ve nüfus artış hızı değişkeninin 2010 yılında hiçbir analizde anlamlı olmadığı 2011 yılında ise %50 kantil regresyon analizinde anlamlı olarak elde edildiği görülmektedir. 2012-2013 yılları analiz sonuçlarına göre ise NB1 ve NB2 analizleri sonucunda anlamlı, kantil regresyon analiz sonuçlarına göre ise anlamsız bir değişken olarak elde edilmiştir.

Tablo 4.11. Nüfus Artış Hızı Değişkeni Anlamlılık Tablosu

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	Anlamsız	Anlamsız	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2011	Anlamsız	Anlamsız	Anlamsız	Anlamlı +	Anlamsız	Anlamsız
2012	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız
2013	Anlamlı +	Anlamlı +	Anlamsız	Anlamsız	Anlamsız	Anlamsız

4.2. MODEL KARŞILAŞTIRMA KRİTERLERİ

4.2.1. Tahminlerin Standart Hatası

Standart hata gerçek değerlerle tahmini değerler arasındaki farkın ölçüsüdür. Aşağıdaki gibi hesaplanır.

$$S_{YX} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n - 1 - m}}$$

m, regresyon modelindeki parametre sayısıdır. Standart hatası küçük olan regresyon değişkenler arasındaki ilişkiyi daha iyi temsil eder (Başar ve Oktay, 2010:75).

Tahmini Standart Hata Sonuçlarına göre 2010-2011 ve 2012 yıllarında bütün analiz sonuçları içerisinde en etkin sonucu veren analiz NB1 regresyonudur. 2013 yılında da ise en etkin sonuç NB2 modeli ile elde edilmiştir.

Tablo 4.10. Tahmini Standart Hata Sonuçları

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	1.10^6	5.10^6	10.10^6	6.10^6	7.10^6	10.10^6
2011	1.10^6	3.10^6	12.10^6	7.10^6	9.10^6	2.10^8
2012	1.10^6	3.10^6	13.10^6	7.10^6	8.10^6	11.10^6
2013	4.10^6	1.10^6	13.10^6	7.10^6	8.10^6	13.10^6

4.2.2. Akaike Bilgi Kriteri

AIC önemli bir istatistik olup modellerin karşılaştırılması için özellikle kullanılmaktadır.

$$\text{AIC}=n\ ln\left(\frac{\sum(Y_i-\hat{Y}_i)^2}{n}\right)+2p$$

n, gözlem sayısını, p değeri de bağımsız değişken sayısını temsil eder. Amaç minimum AIC değerini bulmaktır(Akaike, 1974:317-332).

Akaike Bilgi Kriteri analiz sonuçlarına göre 2010 yılında en etkin sonucu veren analiz 2.319 değeri ile NB1 regresyonudur. 2011-2012 yıllarında NB1 modeli daha etkin sonuç vermiştir. Son olarak 2013 yılında da NB2 modeli ile daha etkin sonuçlar elde edilmiştir.

Tablo 4.13. Akaike Bilgi Kriteri Analiz Sonuçları

	NB1	NB2	%10 Kantil	%50 Kantil	%75 Kantil	%90 Kantil
2010	2.319	2.504	2.618	2.554	2.576	2.628
2011	2.315	2.424	2.649	2.575	2.607	3.108
2012	2.314	2.449	2.662	2.564	2.596	2.645
2013	2.494	2.350	2.661	2.569	2.596	2.661

4.2.3. Hata Kareler Ortalaması

Model seçiminde kullanılan diğer bir yöntemdir. En küçük olan değer seçilir ve aşağıdaki gibi hesaplanır.

$$MSE = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n - p}$$

n, gözlem sayısı, p parametre sayısıdır (Karadavut, 2005:259).

Hata kareler ortalaması analiz sonuçlarına göre de en etkin sonuç sıfıra en yakın olan model olarak değerlendirilmektedir. Bu modele göre de 2010-2011 ve 2012 yıllarında bütün analiz sonuçları içerisinde en etkin sonucu veren analiz NB1 regresyonudur. 2013 yılında da ise en etkin sonuç NB2 modeli ile elde edilmiştir.

Tablo 4.14. Hata Kareler Ortalaması Analiz Sonuçları

	2010	2011	2012	2013
NB1	2.10^{12}	241.10^{10}	24.10^{11}	22.10^{12}
NB2	25.10^{12}	$9.\ 10^{12}$	12.10^{12}	37.10^{11}
%10 Kantil	10.10^{13}	15.10^{13}	17.10^{13}	17.10^{13}
%50 Kantil	46.10^{12}	59.10^{12}	52.10^{12}	55.10^{12}
%75 Kantil	60.10^{12}	89.10^{12}	77.10^{12}	77.10^{12}
%90 Kantil	11.10^{11}	43.10^{15}	$14.\ 10^{13}$	17.10^{13}

4.2.4. Ortalama Mutlak Hata

Ortalama mutlak hata, tahmin edilen değerlerin gözlenen değerlerden ne kadar uzakta olduğunu gösteren istatistiksel bir ölçümdür. Aşağıdaki gibi hesaplanmaktadır (http://en.wikipedia.org/wiki/Mean_absolute_error).

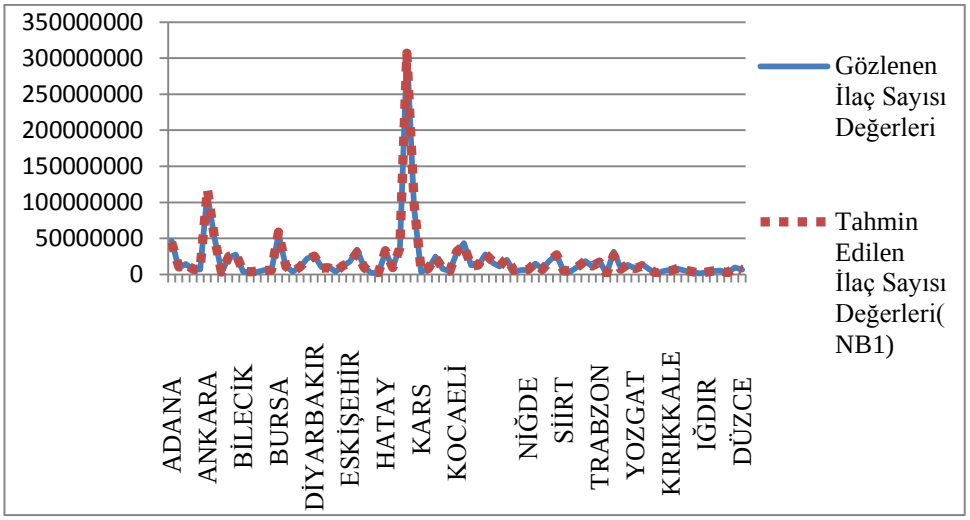
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

Ortalama mutlak hata analiz sonuçlarına göre 2010-011 ve 2012 yılı analiz sonuçlarına göre en etkin sonucu veren analiz NB1 regresyonudur. Son olarak 2013 yılında da en etkin sonuç NB2 modeli ile elde edilmiştir.

Tablo 4.12. Ortalama Mutlak Hata Analiz Sonuçları

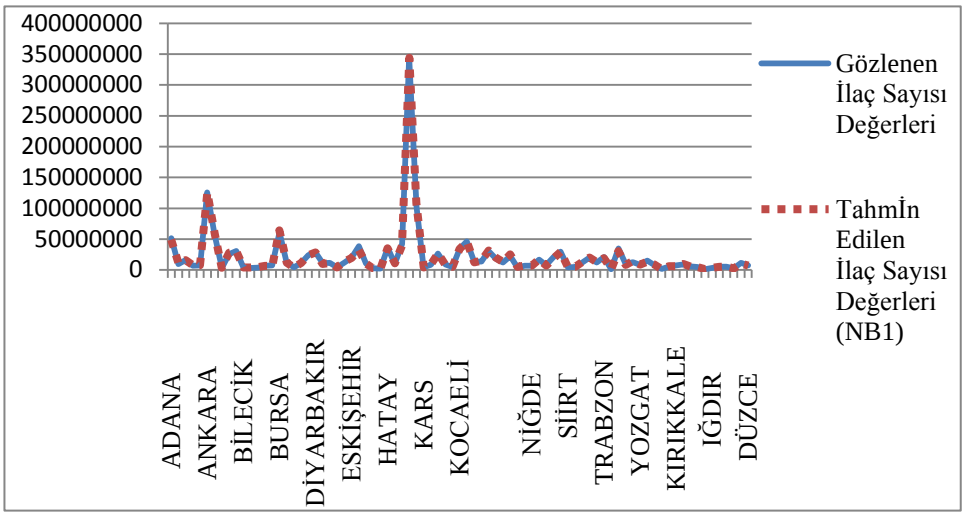
	2010	2011	2012	2013
NB1	10.10^5	933.10^3	969.10^3	173.10^4
NB2	16.10^5	1.10^6	1.10^6	113.10^4
%10 Kantil	6.10^6	8.10^6	9.10^6	9.10^6
%50 Kantil	4.10^6	5.10^6	5.10^6	5.10^6
%75 Kantil	5.10^6	7.10^6	6.10^6	6.10^6
%90 Kantil	8.10^6	19.10^7	8.10^6	10.10^6

Model karşılaştırma kriterleri ile yapılan analizler sonucunda en etkin sonucu veren modelle tahmin edilen ilaç sayısı ve gözlenen değerlerin grafikleri çizildiğinde aşağıdaki sonuçlar elde edilmiştir.



Grafik 4.1. 2010 Yılı NB1 Analiz Tahmini

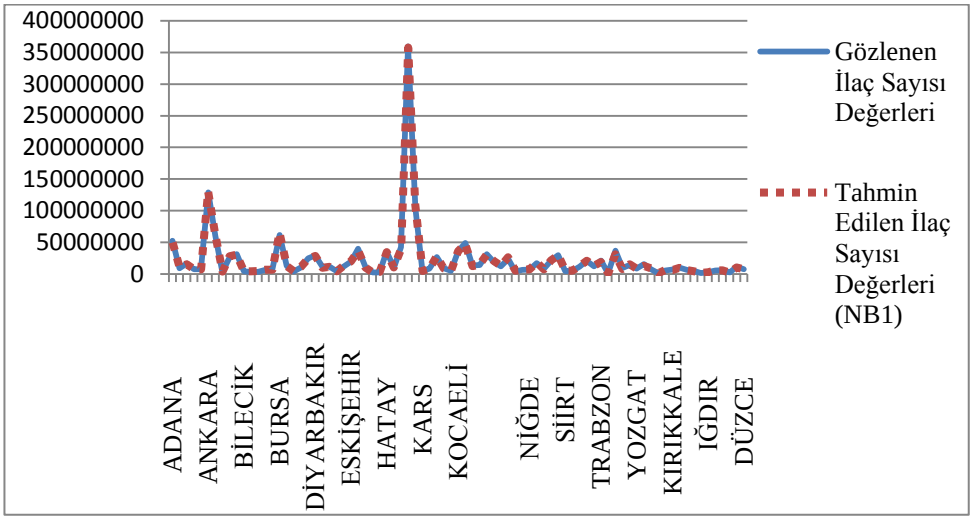
İlk olarak 2010 yılı analizleri sonucunda en etkin sonucun NB1 analizinden elde edildiği, NB1 analizi sonucunda tahmin edilen ve gözlenen ilaç sayısı değerlerine ait grafik çizildiğinde, grafik 4.1'deki gibi olduğu ve NB1 modelinin aşırı yayılım durumlarından etkilenmediği ve gözlenen değerlere oldukça yakın tahminde bulunduğu görülmektedir.



Grafik 4.2. 2011 Yılı NB1 Analiz Tahmini

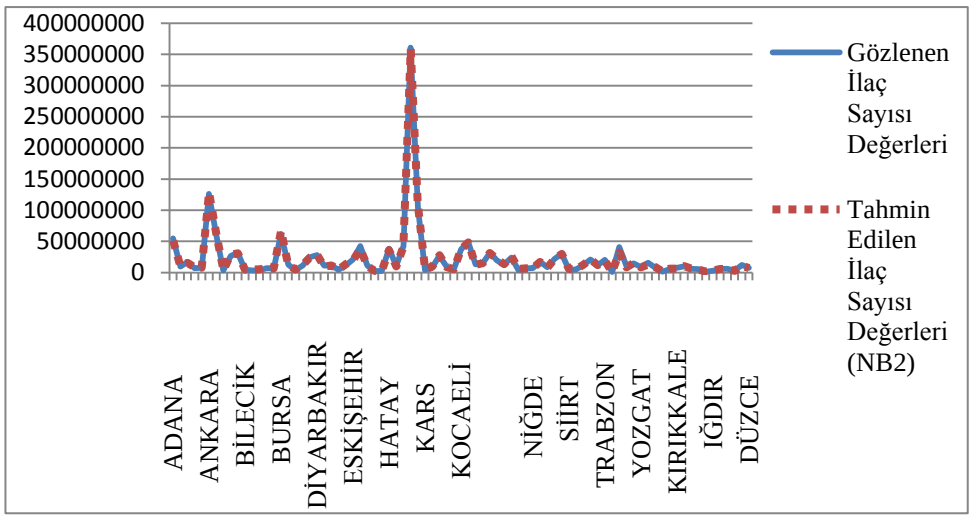
2011 yılı analizleri sonucunda en etkin sonucun 2010 yılında olduğu gibi NB1 analizinden elde edildiği, tahmin edilen ve gözlenen ilaç sayısı değerlerine ait grafik çizildiğinde ise grafik 4.2'deki gibi olduğu ve NB1 modelinin aşırı yayılım

durumlarından etkilenmediği görülmektedir. Özellikle İstanbul iline ait tahmin edilen ilaç sayısı değerlerinin gözlenen değerlere oldukça yakın tahminde bulunduğu gözlenmiştir.



Grafik 4.3. 2012 Yılı NB1 Analiz Tahmini

2012 yılı analizleri sonucunda ise en etkin sonucun 2010-2011 yıllarında olduğu gibi NB1 analizinden elde edildiği, tahmin edilen ve gözlenen ilaç sayısı değerlerine ait grafik çizildiğinde ise grafik 4.3'deki gibi olduğu ve NB1 modelinin aşırı yayılım durumlarından etkilenmediği görülmüştür. Aykırı değerlerin görüldüğü illerde de gözlenen değerlere oldukça yakın tahminde bulunduğu gözlenmiştir.



Grafik 4.4. 2013 Yılı NB2 Analiz Tahmini

Son olarak 2012 yılı analizleri sonucunda ise en etkin sonucun NB2 analizinden elde edildiği, tahmin edilen ve gözlenen ilaç sayısı değerlerine ait grafik çizildiğinde ise, grafik 4.4'deki gibi olduğu ve NB2 modelinin aşırı yayılım durumlarına hassas yaklaştığı görülmüştür.

SONUÇ VE DEĞERLENDİRME

Tüm dünyada olduğu gibi ülkemiz için de ciddi bir sorun olan etkisiz, yanlış ve gereksiz ilaç kullanımının giderek artması, bu konuda yaşanan sorunlara daha ciddi yaklaşılmasını bir zorunluluk haline getirmiştir. İlaçların rahat bir şekilde satın alınabildiği ve bilinçsiz bir şekilde kullanıldığı ülkemizde de OECD raporuna göre her yıl tüketilen ilaç sayısı artmaktadır. İller bazında yapılan IMS kutu ve fiyat karşılaştırmalarında da, belirgin farklılıklar göze çarpmaktadır.

Bu çalışmada 2010-2011-2012 ve 2013 yılları esas alınarak illerde kullanılan ilaç tüketimine etki eden değişkenler incelenmiştir. Bağımsız değişken olarak illerin nüfusu, illerde bulunan eczacı sayısı, özel ya da devlet hastanesi ayırt etmeksizin çalışan hekim sayısı, özel ya da devlet ayırt etmeksizin her ilde bulunan hastane sayısı, illerin nüfus artış hızları ve illerin denizlere olan konumları ele alınmıştır. İllerde tüketilen ilaç miktarı IMS Türkiye bürosundan alınmış olup, diğer veriler TÜİK Sağlık Araştırmaları sayfasından alınmıştır. İllerde bulunan hastane sayısı, deniz etkisi ve illerin nüfus artış hızı değişkenleri hariç, diğer değişkenlerin logaritması alınarak analize dahil edilmiştir. Böylece modellerde altı bağımsız değişken kullanılmıştır. Bu şekilde değişkenler analize dahil edilerek değişkenler arasındaki çoklu bağlantı azaltılmaya çalışılmıştır.

İlk olarak negatif binomial regresyon analizleri son olarak ise kantil regresyon analizleri, son olarak ise kantil regresyon analizleri çeşitli kantil dilimleri alınarak yapılmıştır. Kantil aralıkları %10, %50, %75 ve %90 olarak alınmıştır.

Poisson regresyon analizinin ortalama ve varyans eşitliğini sağlaması gerekmektedir Bu özellik sağlanamadığından dolayı aşırı yayılım durumunda kullanılan negatif binomial regresyon modelleri olan NB1 ve NB2 kullanılmıştır. Bu iki yöntem arasındaki fark yayılım değerinin farklı alınmasından kaynaklanmaktadır. Yapılan analizler sonucunda genel olarak NB1 ve NB2 modellerinde log-nüfus, log-hekim, log-eczacı ve deniz etkisi değişkenleri anlamlı olarak elde edilmiştir. Negatif binomial regresyonlarda bağımlı değişken üzerindeki bağımsız değişkenlerin etkisi de katsayı değerleri tarafından belirlenmektedir. Değişkenler üzerindeki bu yorum log-nüfus, log-hekim, log-eczane, nüfus artış hızı, deniz etkisi ya da hastane sayısında meydana gelecek bir birimlik artışın log ilaç sayısında meydana gelecek artma ya da azalmayı ifade etmesi şeklindedir.

Kantil regresyon analizi ise genelde çoklu doğrusal regresyon analizi ile karşılaştırılır. Çoklu doğrusal regresyondan farklı olarak kantil regresyon analizlerinde normal dağılım şartı aranmamaktadır. Normal dağılım olduğu durumlarda çoklu doğrusal regresyon ile aynı sonucu vermekte olup normal dağılım olmayan durumlarda daha sağlıklı sonuçlar vermektedir. Aykırı değerlerden etkilenmediğinden dolayı aranan bir model olmaktadır. Analizlerde de kantil dilimleri ayrı ayrı yapılarak değerlendirilmiştir. Kantil dilimleri %10, %50, %75 ve %90 alınarak yapılmıştır. Kantil regresyon analizleri sonucunda ise genel olarak illerde bulunan hastane sayısı değişkeni anlamlı olarak elde edilmiştir. Kantil regresyon parametre tahmini çoklu lineer regresyon tahminlerinde olduğu gibi olup bağımsız değişkende meydana gelecek bir birimlik artış bağımlı değişken üzerindeki bir birimlik artışa sebep olmaktadır. Kantil dilimlerine ait analizler tablolaştırılarak bu fark gözlenmiştir.

Her iki modelde de analizlerin aykırı değerlerden etkilenmediği gözlenmiştir. Özellikle İstanbul, Ankara, Bursa, İzmir gibi büyük illerde tüketilen ilaç sayısı diğer illerde tüketilen ilaç sayısına göre oldukça farklı olmaktadır. Ancak kullanılan modellerde bu aykırı durum değerlendirmeye alınmış ve yapılan tahminler sonucunda oldukça yakın değerler elde edilmiştir.

Negatif binomial regresyon ve kantil regresyon analizlerini karşılaştırmak için Tahminlerin Standart Hatası, Akaike Bilgi Kriteri, Hata Kareler Ortalaması ve Ortalama Mutlak Hata yöntemleri kullanılmıştır. Bu yöntemler de paralel sonuçlar elde edilmiştir. En etkin sonuç 2010-2011 ve 2012 yılı analiz sonuçlarına göre NB1 analizinden elde edilirken, 2013 yılında ise NB2 analiz yönteminin en etkin sonuç verdiği görülmüştür.

Son olarak ise karşılaştırılan modeller arasında en yakın tahminde bulunan analiz yönteminin NB1 ve NB2 analizleri tarafından yapıldığı kantil regresyon analizleri sonucunda elde edilen tahminlerin ise negatif binomial regresyon analizleri sonucu elde edilen tahminlerin gerisinde kaldığı görülmüştür.

Değişkenlerin anlamlılıkları yıl olarak esas alındığında farklı sonuçlar elde edilmiştir. 2010 ve 2011 yıllarında log-nüfus, log-eczacı, log-hekim ve deniz etkisi değişkenleri NB1 analizinde anlamlı sonuçlar vermişken hastane ve nüfus artış hızı değişkenleri anlamlı sonuç vermemiştir.

2012 yılı diğer yıllara göre oldukça farklı sonuçlar vermiş olup log-nüfus, log-eczacı, log-hekim ve deniz etkisi değişkenlerinin yanında nüfus artış hızı değişkeni de NB1 analizinde anlamlı sonuçlar vermiştir. Hastane değişkeni yine anlamsız çıkmıştır.

2013 yılı değişkenlerinin anlamlılıkları incelendiğinde log-nüfus, log-eczacı, deniz etkisi ve nüfus artış hızı değişkenleri NB2 analizinde anlamlı sonuçlar vermiş olup hastane ve log-hekim değişkeni anlamsız bulunmuştur.

KAYNAKLAR

Arı, A., Önder, H. (2012). ‘‘Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri’’. Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Anadolu Tarım Bilim. Dergisi, Samsun.

Akaike, H. (1974). “Factor Analysis and AIC”, *Pschychometrika, 52,* 317-332.

Avşar, V.,(2014). ‘‘Türkiye’nin Antidamping Soruşturmalarını Etkileyen Faktörler: Sanayi Verileri İle Ekonometrik Bir Analiz’’. Eskişehir Osmangazi Üniversitesi İibf Dergisi,9(1):41-54.

Başar, A. Oktay E. (1997). *Uygulamalı İstatistik-1*. Aktif yayınları.

Barnett V, Lewıs T, (1994). *Outliers in Statistical Data*. Canada: John Wiley Sons.

Baur, D.,Saisana, M., Niels, N., (2004), ‘‘Modelling The Effects of Meteorological Variables on Ozone Concentration A Quantile Regression Approach’’. Environment pp. 4689–4699.

Behr, A. (2008). “Kantil Regression For Robust Bank Efficiency Score Estimation”, *European Journal of Operational Research, 200*, 568–581.

Bilgiç, A., Florkowski, J. W.(2007). Application of A Hurdle Negative Binomial Count Data Model To Demand For Bass Fishing in The Southeastern United States, Journal Of Environmental Management 83.478-490

Birkes, D., Dodge, Y. (1993). *Alternative Methods of Regression*, Canada: John Wiley Sons, Inc.

Boswell, M. T., G. P. Patil (1970). ‘‘Chance Mechanisms Generating The Negative Binomial Distribution, in Random Counts in Models and Structures’’, Volume 1, ed. G. P. Patil, University Park, PA: Pennsylvania State University Press, pp. 1–22.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963–1987: Application of Quantile Regression. *Econometrica, 62,* 405–458.

Buchinsky, M. (1998). ‘‘Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research’’, The Journal of Human Resources. Vol.33, No.1, 90-101

Cameron, A. C., Trivedi, P.K., (1986). ‘‘Econometrics Models Based On Count Data; Comparison And Applications of Some Estimators and Tests’’, Journal Of Applied Econometrics,Vol,1,s:29-53.

Cameron, A. C., Trivedi, P.K., (1998). ‘‘Regression Analysis of Count Data’’. West Nyack, NY, USA: Cambridge University Press.

Chen, L. (2005). “An Introduction to Kantil Regression and the QUANTREG Procedure”, *Statistics and Data Analysis*, 213-230

Chen, C., Wei, Y. (2005). “Computational Issues for Quantile Regression”. *The Indian Journal of Statistics*, 62(2), 1-19.

Colin, C. (2005). ‘‘An Introduction to Quantile Regression and the Quantreg Procedure’’., SAS Institute Inc., Cary, NC

Cox, D.R., (1983). “Some Remarks on Overdispersion”. *Biometrika 70*, s: 269-274.

Çalışkan, Z.,(2009). ‘‘The Relationship Between Pharmaceutical Expenditure and Life Expectancy: Evidence From 21 OECD Countries’’. Applied Economics Letters, Department of Economics, Hacettepe University, Ankara.

Deniz, Ö. (2005). ‘‘Poisson Regresyon’’. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl: 4, Sayı: 7, s. 59-72

Demirbağ, C. B., Timur, M. (2011). ‘‘Bir Grup Yaşlının İlaç Kullanımı İle İlgili Bilgi, Tutum ve Davranışları’’. Ankara Sağlık Hizmetleri Dergisi, Cilt 11, Sayı 1.

Englin, J. and J. Shonkwiler (1995).‘‘Estimating Social Welfare Using Count Data Models: An Application Under Conditions Of Endogenous Stratification And Truncation’’, Review of Economics and Statistics 77.

Eide, E., Showalter, H., M. (1998). “The Effect of School Quality on Student Performance: A Kantil Regression Approach”. *Economics Letters, 58,* 345–350.

Ercanlı, İ. Kahriman, A. Yavuz, H.(2012). ‘‘Trabzon Orman Bölge Müdürlüğü Doğu Ladini-Sarıçam Karışık Meşcereleri İçin Karışık Etkili Doğrusal Olmayan Regresyon Denklemleri İle Doğu Ladini Çap-Boy Modellerinin Geliştirilmesi’’. SDÜ Orman Fakültesi Dergisi13: 75-84.

Erol, Dilek (2001); " Avrupa Birliği ve İlaç Sanayi", Yeni Türkiye 40, s:1057-1067.

Frome, E. L.(1983). ‘‘The Analysis of Rates Using Poisson Regression Models’’. Medical and Health Sciences Division, Oak Ridge Associated Universities, Oak Ridge, Tennessee 37830, U.S.A. Bıometrıc 39, 665-674.

Ghitay, M. E., Karlis D., Al-Mutairi D. K., Al-Aawadhi F. A., 2012. An EM Algorithm for Multivariate Mixed Poisson Regression Models and its Application. Applied Mathematical Sciences, Cilt:6, no:137, s. 6843-6856.

Gilchrist, Warren G.(2000) Statistical Modelling with Quantile Functions, Florida: Chapman and Hall/ CRC.

Goodal C. (1983). "Understanding Robust and Exploratory Data Analysis''. *John Wiley Sons*, Canada, 211–242.

Gujerati, D., N. (1999). *Temel Ekonometri*. (Çev.: Şenesen, Ü., Şenesen, G., G.), İstanbul: Literatür Yayınları.

Güriş S., Çağlayan E.(2010). '' Ekonometri Temel Kavramlar''. Der Yayınları.

Gürsakal, N.(1997). Bilgisayar Uygulamalı İstatistik 1, Bursa: Marmara Kitabevi, s:38.

Hao, L. ve Naiman, D. Q. (2007). *Quantile Regression.* The United States of America: Sage Publication.

Hilbe, J. (2007). ''Negative Binomial Regression''. Cambridge, UK: Cambridge University Press,

Judge, George G. ve Diğerleri.(1985). ''The Theory and Practice of Econometrics''. 2. Baskı, Canada: John Wiley.

Kan, K., W. D. Tsai.(2004). ''Obesity and Risk Knowledge in Taiwan: A Quantile Regression Analysis''. Journal of Health Economics 23/5:907-34.

Karadavut U. Sinan A., Genç, A., Palta, Ç., Karakoca A. ve Aksoyak Ş. (2005). "Dağdaş Buğday Bitkilerinin Büyüme Eğrilerinin Belirlenmesinde Model Seçimi", *4. İstatistik Kongresi: 8-12 Mayıs 2005,* Belek/Antalya.

Koenker, R., Bassett G. (1978). ''Regression Quantiles'', The Econometric Society, Vol. 46, No.1.pp.33-50.

Koenker, R., Machado, A. F. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," Journal of the American Statistical Association, 94, 1296–1310.

Koenker, R., Hallock, K. F. (2001). "Quantile Regression: An Introduction". *Journal of Economic Perspectives, 15,* 143–156.

Koenker, R. (2005). *Kantil Regression,* USA: Cambridge University Press

Kurtoğlu, F., (2011). ''Quantile Regresyon: Teorisi ve Uygulamaları''. Çukurova Üniversitesi Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yayınlanmış Yüksek Lisans Tezi.

Lawless, J. F.(1987). "Negative Binomial and Mixed Poisson Regression". *The Canadian Journal of Statistics*, 15(3), 209-225.

Lee, S. (2004). ‘‘Quantile Regression’’. Lecture Notes for MECTI, p:1

Leping Kristjan-Olari (2005). “Public-Private Sector Wage Differential in Estonia: Evidence From Quantile Regression, Tartu University”, *Faculty of Economics And Business Administration*, Tartu University Press, Orden No:431, Tartu

Li, T., Sun, L., Zou, L. (2009). “State Ownership and Corporate Performance: A Kantil Regression Analysis of Chinese Listed Companies”, *China Economic Review*, CHIECO-00395.

Lingren, A. (1997). “Kantil Regression With Censored Data Using Generalized L1 Minimization”, *Computational Statistics & Data Analysis, 23*, 509-524.

McKean, J. R., Taylor, R. G. (2000). Outdoor Recreation Use and Value: Snake River Basin of Central Idaho, Idaho Experiment Station Bulletin, Moscow, Idaho, 58p.

Meligkotsidou, A., Vrontos, D., L., Vrontos, D., V. (2009). “Kantil Regression Analysis of Hedge Fund Strategies”, *Journal of Empirical Finance*, 16, 264–279.

Moons, E., Loomis, J., Proost, S., Eggermont, K., Hermy, M. (2001). ‘‘Travel Cost and Time Measurement in Travel Cost Models’’, Faculty of Economics and Applied Economic Science Center For Economic Studies Energy, Transport and Environment, working paper series no : 2001-22.

Nevitt, J., Tam, H (1998). “A Comparison of Robust and Nonparametric Estimators Under The Simple Linear Regression Model”, *Multiple Linear Regression Viewpoints, 25*, 54-69.

Newey, W. K., Powell. J. L.(1990): “Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions,” Econometric Theory, 6(3), 295–317.

Özmen, İ., (1998), ‘‘Poisson Regresyon Çözümleme Teknikleri, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü (Yayınlanmış Doktora Tezi).

Park, B. J., Lord. D.,(2005). ‘‘Negative Binomial Regression Models and Estimation Methods’’.Appendıx-D.

Saçaklı, İ., (2005). ‘‘Kantil Regresyon ve Alternatif Regresyon Modelleri ile Karşılaştırılması’’, (Yayınlanmış Yüksek Lisans Tezi), İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı.

Saraçoğlu, B., Çevik, F. (1995). *Matematiksel İstatistik.* Ankara: Gazi Büro Kitabevi.

Serper, Ö., (2004). Uygulamalı İstatistik 1.5.Bursa: Ezgi Kitabevi.

Sezgin, F.H., Deniz, E.(2004). "Poisson Regresyon Modelinde Aşırı Yayılım Durumu ve Negatif Binomial Regresyon Analizinin Türkiye Grev Sayıları Üzerine Bir Uygulaması", Yönetim Dergisi, Sayı:48, İstanbul.

Şemin, S.(1998).''Sosyal ve Ekonomik Yönüyle İlaç'', Türk Tabipleri Birliği Merkez Konseyi, Ankara.

Stifel, D.C., Averett S.L.(2009). ''Childhood Overweight İn The United States: A Quantile Regression Approach''. Economics & Human Biology Volume 7, Issue 3,387–397.

Teksöz, T., Batıgün, O., Gedikli, C., Fındık, F., Sorgun, S., Daver, D., Taylan, E., Dülger, H., Kayabalı, K.(2012). ''İlaç Sektörünün Son 10 Yılında Yaşanan Gelişmeler''.

Temiz, S. (2006). ''Hata Paylarının Normal Dağılmaması Durumunda Tek Değişkenli Klasik Regresyonda Alternatif Tahmin Yöntemleri''. Yıldız Teknik Üniversitesi, İstatistik Anabilim Dalı İstanbul.

Wang, Peiming, M. Cockburn ve Martin L. P.(1998) "Analysis of Patent-a Mixed-Poisson-Regression-Model Approach", American Statistical Association, No:16, s:40.

Wang, Weiren, Felix, F.(1997) "Modeling Household Fertility Decisions with Generalized Poisson Regression", Journal of Economics, No: 10, s.273-283.

Wang, H. (2007). "*Quantile Regression: Overview and Applications to Risk Assessment*". *North Caroline State University,* 1-26

Wedderburn, R. W. M. (1974). "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method". *Biometrika 61,* 439–447.

Ver Hoef M.Jay, Boveng Peter L.(2007), "Quası-Poısson Vs. Negatıve Bınomıal Regressıon: How Should We Model Overdıspersed Count Data? Ecological Society of America US Department of Commerce". *University of Nebraska - Lincoln Ecology,* 88(11), 2007, 2766–2772

Yeşilova, A., Yılmaz, A., Kaki B.(2006), ''Norduz Erkek Kuzularının Bazı Kesikli Üreme Davranış Özelliklerinin Analizinde Doğrusal Olmayan Regresyon Modellerin Kullanılması''. Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi Tarım Bilimleri Dergisi 2006, 16(2):87-92.

Zeileis A., Kleiber C., Jackman S. (2007). ‘‘Regression Models for Count Data in R’’. Working Papers:24 Faculty of Business and Economics.

Zwilling, M. L.(2013). “Negative Binomial Regression,” The Mathematica Journal, 2013. dx.doi.org/10.3888/tmj.15-6.

İNTERNET KAYNAKLARI

(www.cscu.cornell.edu/,2007)

(http://www.farmasyon.com.tr/makale/122/turkiye_ilaç_pazarından_çarpıcı_rakamlar)

(http://en.wikipedia.org/wiki/Mean_absolute_error).

(http://www.gündem.com.tr/ilaçkullanımı).

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Kübra ELMALI
Doğum Yeri ve Tarihi	26.04.1987
Eğitim Durumu	
Lisans Öğrenimi	Atatürk Üniversitesi Fen Fakültesi Matematik Bölümü (2006-2009)
Y. Lisans Öğrenimi (tezsiz)	Atatürk Üniversitesi Eğitim Fakültesi (2009-2010)
Y. Lisans Öğrenimi	Atatürk Üniversitesi İ.İ.B.F. Ekonometri Bölümü (2011-2014)
Bildiği Yabancı Diller	İngilizce
İletişim	
E-Posta Adresi	k.elmali@hotmail.com
Tarih	