



Kantil Regresyon

Quantile Regression

Arzu ALTIN YAVUZ¹, Ebru GÜNDOĞAN AŞIK²

¹Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, 26480 ESKİŞEHİR

²Karadeniz Teknik Üniversitesi, Fen Fakültesi, 61080 TRABZON

Başvuru/Received: 13/03/2017

Kabul/Accepted: 18/05/2017

Son Versiyon/Final Version: 15/06/2017

Öz

Regresyon analizi uygulama alanı en geniş olan istatistiksel analiz yöntemlerinden biridir. Birçok alanda tekniğinde olduğu gibi mühendislik alanında da yaygın olarak kullanılmaktadır. Regresyon analizinde kullanılan En Küçük Kareler (EKK) tekniğinin çıkarsama amaçlı kullanılabilmesi bazı varsayımların sağlanmasını zorunlu kılar. EKK tekniğinde hata terimleri dağılımının normal dağılıma sahip olmaması ve modelin aykırı değerler içermesi durumunda EKK tahmin edicileri etkinlik özelliklerini kaybetmektedir. Bu durumda alternatif regresyon tekniklerine başvurulmaktadır. Alternatif regresyon yöntemlerinden biri olan Kantil regresyon, klasik regresyon yöntemlerinin bazı sınırlamalarının üstesinden gelmektedir. Bu çalışmada Kantil regresyon yöntemi tanımlanmış ve bir mühendislik uygulaması üzerinde EKK tahmin edicileri ile karşılaştırılmıştır. Beton kırma deneyi için elde edilen sonuçlara göre, EKK yöntemi ile elde edilen modelin çıkarsama amaçlı kullanılamayacağı tespit edilmiştir. Bu durumda $\tau=0.75$ 'inci ve $\tau=0.25$ 'inci kantil değerine göre kurulan regresyon denklemi çıkarsama amaçlı kullanılabilir.

Anahtar Kelimeler

“Aykırı değer, Regresyon analizi, En Küçük Kareler, Kantil regresyon”

Abstract

Regression analysis is one of the most widely used statistical analysis methods. It is widely used in the engineering field as it is in many areas. The fact that the Least Squares (LS) technique used in regression analysis can be used for inference makes it necessary to provide some assumptions. In the LS, if the distribution of error terms does not have normal distribution and if the model contains outliers, the least squares estimators lose their efficiency properties. In this case, alternative regression techniques are applied. Quantile regression, one of the alternative regression methods, comes from overcoming some of the limitations of classical regression methods. In this study, the method of quantile regression is introduced and on an engineering application is compared with the estimators of the LS. According to the results obtained for the concrete breaking test, it has been determined that the model obtained by the method of LS can not be used for inference. In this case, it can be used for inference the regression equation established for $\tau=0.75$ th and $\tau=0.25$ th quantile value.

Key Words

“Outlier, Regression analysis, Least Squares, Quantile regression”

1. GİRİŞ

Mühendislik alanında yapılan birçok bilimsel çalışmada yoğun bir şekilde istatistiksel yöntemler kullanılmaktadır. Yapılan bilimsel çalışmalardan elde edilen sonuçların geçerliliği ve güvenilirliği, doğru yöntemlerin kullanılması ile yakından ilişkilidir. Regresyon analizi mühendislik çalışmalarında kullanılan en önemli istatistiksel yöntemlerden biridir. Regresyon analizi, bağımlı ve bağımsız değişken(ler) arasındaki fonksiyonel ilişkiyi inceler ve bu ilişkiyi ifade eden matematiksel modelin parametrelerini tahmin etmeyi amaçlar. Regresyon analizinde yaygın olarak kullanılan En Küçük Kareler (EKK) tahmin edicilerinin etkin olması ve elde edilen modelin çıkarsama amaçlı kullanılabilmesi bazı varsayımların sağlanmasına bağlıdır. Bu varsayımlar hataların bağımsız olması, sıfır ortalamalı ve eşit varyanslı ($Var(\epsilon_i)=\sigma^2$) Normal dağılıma sahip olması, bağımsız değişkenler arasında çoklu doğrusal bir ilişkinin olmaması olarak sıralanabilir (Montgomery vd, 2013).

Regresyon analizinde EKK yönteminin yaygın olarak kullanılmasının başlıca nedeni diğer regresyon yöntemlerine göre daha kolay hesaplanmasıdır. EKK yönteminin amacı hataların kareler toplamını minimize etmektir. Ayrıca hatalar normal dağıldığında sapmasız tahmin ediciler arasında minimum varyanslı (UMVUE) olan tahmin edici EKK tahmincisidir. Özellikle hataların normal dağılıma uygun olmadığı ve aykırı değerler barındırdığı durumlarda EKK tahmin edicileri etkinlik özelliklerini kaybetmektedir. EKK regresyonunun bu dezavantajı tahmin edilen katsayıların gerçeği yansıtmamasına sebep olmaktadır. Bu durumlarda alternatif regresyon tekniklerine başvurulması gerekmektedir.

Alternatif regresyon tekniklerinden biri Kantil regresyondur. Koşullu kantil fonksiyonlarının tahmini için uygun bir yöntem olarak Koenker ve Basset (1978) tarafından geliştirilmiştir. Kantil regresyon, bağımlı değişkenin dağılımındaki herhangi bir kantil ile bağımsız değişkenler arasındaki fonksiyonel ilişkinin tahmini için geliştirilmiştir (Tareghian & Rasmussen, 2013). Kantil regresyon, özellikle koşullu kantillerin değişkenlik gösterdiği durumlarda kullanışlıdır. Kantillere bağlı olarak regresyon katsayıları belirlenir (Chen & Wei, 2005). Kantil regresyon ilk olarak regresyondaki klasik varsayımlardan hata terimlerinin normal dağılması varsayımını ihmal eden sağlam bir regresyon yöntemi olarak ortaya çıkmıştır. (Chen & Wei, 2005). Kantil regresyon modeli, EKK yönteminden daha fazla esnek ve bağımlı değişkenin dağılımına ait kovaryans etkilerini incelemeye olanak sağlamaktadır (Çağlayan & Arikan, 2011). Kantil regresyon modeli, geleneksel ortalama regresyondan daha eksiksiz bir model sunmaktadır (Yu vd, 2003). Çalışmalar örneğin, ücretler (Buchinsky, 1994; Machado & Mata, 2005; Martins & Pereira, 2004), yaşam analizi (Crowley & Hu, 1977; Koenker & Geling, 2001), finansal analiz (Bassett & Chen, 2001), ekonomik araştırmalar (Hendricks & Koenker, 1992; Wang vd, 2013), çevre çalışması (Pandey & Nguyen, 1999), internet ve iletişim teknolojisi (BİT) benimsenmesi (Yu, 2011), sağlık harcamaları (Yu vd, 2011), küçük işletme performansları (Seo vd, 2014) ve bunun gibi çeşitli sorunları yorumlamak için Kantil regresyon modeli uygulanır.

Kantil regresyon, klasik regresyon modellerinin bazı sınırlamalarının üstesinden gelmektedir ve diğer regresyon modellerinin gözden kaçırmış olabileceği değişkenler arasındaki ilişkinin daha kapsamlı bir resmini sağlamaktadır. Kantil regresyon modelleri koşullu ortalama fonksiyonları ile koşullu kantil fonksiyonları için tahmin yapılmasında kullanılır. Deneysel çalışmalarda görülen problemlerden biri aykırı değer problemidir. Bu problem söz konusu olduğunda çalışmalarda ya bu değerler göz ardı edilmekte ya da bu değerler çalışmadan atılmaktadır. Özellikle deneysel çalışmalarda bağımlı değişkenin sadece ortalamasının değil, diğer kantil değerlerinin tahmininin de büyük bir öneme sahip olduğu düşünülürse, büyük maddi imkanlarla yürütülen mühendislik çalışmalarının değerlendirilmesinde Kantil regresyonun kullanılması oldukça faydalı olacaktır.

Kantil regresyon, ilk olarak Koenker ve Schorfheide (1994) tarafından iklim çalışmalarından birinde kullanılmıştır. Ayrıca, Kantil regresyon ekolojik, biyolojik, küresel sıcaklık değişiklikleri, ücret eşitsizliklerinin ve gelir düzeyinin belirlenmesi, akademik başarıda eğitim eşitsizliklerinin etkilerinin belirlenmesi, sağlık çalışmalarında (uzunluk, kilo, hastalıkların artışı, uyuşturucu kullanım ile ilgili) tedavi yöntemlerinin belirlenmesi gibi konularda kullanılmıştır. Mühendislik alanındaki çalışmalarda yeteri kadar uygulama söz konusu değildir.

Bu çalışmanın amacı EKK tahmin edicisinin dayandığı model varsayımlarının sağlanamaması durumunda kullanılacak Kantil regresyon yöntemini tanıtmak ve bir mühendislik uygulaması üzerinde EKK tahmin edicileri ile karşılaştırmaktır. Çalışmanın ikinci bölümünde EKK yönteminin varsayımlarının sağlanmasını etkileyen hataların analizi konusuna değinilecek ve çeşitli hata türleri ile bunları tespit yöntemleri aktarılacaktır. Üçüncü bölümde, Kantil regresyon, EKK yöntemi ile karşılaştırmalı olarak tanıtılacaktır. Dördüncü bölümde, bir beton kırma deneyinde açığa çıkan enerji miktarı ile darbe değeri arasında ilişki EKK ve Kantil regresyon yardımıyla incelenmiştir. Son bölümde elde edilen sonuçlar, farklı sağlam regresyon yöntemleri ile karşılaştırmalı olarak tartışılmıştır.

2. HATALARIN ANALİZİ VE AYKIRI DEĞERLERİ BELİRLEME YÖNTEMLERİ

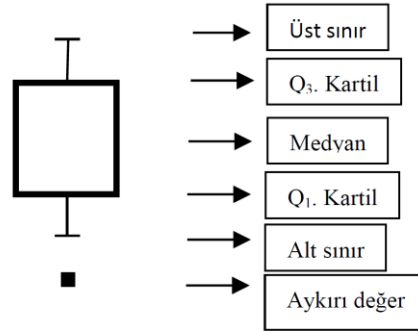
Değişkenler arasındaki ilişkilerin incelenmesinde yaygın olarak kullanılan EKK yöntemi verilerin dağılımının normal olmasına ve aykırı değerlerin bulunmamasına bağlı olarak etkin sonuçlar vermektedir. Ancak günümüzde elde edilen birçok veri seti normal dağılım göstermemekte ve içlerinde diğer gözlemlerden önemli derecede farklılık gösteren aykırı değerler bulundurmaktadır. Aykırı değerler diğer gözlem değerlerinden uzakta bulunan ve onlarla tutarlılık göstermeyen değerlerdir. Aykırı değerler konumlarına göre analiz sonuçlarını etkileyebilecekleri gibi, analiz sonuçları üzerinde herhangi bir etki de yapmayabilirler. Bu

durumda aykırı değerlerin araştırılması gerekir. Veri setinde aykırı değer olması durumunda varsayımsal sapmalardan çok fazla etkilenmeyen sağlam istatistiksel yöntemlerin kullanılması gerekmektedir.

Aykırı değerler görsel olarak histogram ve kutu grafiği yardımıyla belirlenebilir. Sınıflandırılmış serilerin grafikleri bir koordinat sistemi üzerinde çizilirse, çizilen bu frekans serilerinin grafiklerine histogram denir ve aykırı değerleri belirlemede kullanılabilir. Kutu grafiği yardımıyla ise öncelikle kartiller arası fark,

$$IQR = Q_3 - Q_1 \quad (1)$$

elde edilir. Q1 ve Q3 en düşük ve en yüksek kartili ifade eder. IQR genişliği yardımıyla alt ve üst sınırlar belirlenir(Tukey, 1977; Walfish, 2006). Üst sınır $Q_3 + 3/2 * IQR$ ve alt sınır $Q_1 - 3/2 * IQR$ olarak hesaplanır ve bu sınırlar dışındaki değerler aykırı değer olarak belirlenir. Şekil 1’ de aykırı değer belirleme yöntemi gösterilmiştir. (Ovla & Taşdelen, 2012).



Şekil 1. Kutu Grafiği ile Aykırı Değer Belirleme

Regresyon analizinde ise aykırı değerleri belirlemek için en etkin ve basit yol artıkları ele almaktır(Alpar, 2013). Standartlaştırılmış artıklar, student türü artıklar, vb. değişik artık türlerine örnek oluşturlar.

2.1. Standartlaştırılmış Artıklar

$e_i \sim N(0, \sigma^2)$ sahip ise, e_i / σ 'nın standart normal dağılıma sahiptir. Bu nedenle e_i artıkları s ' ye bölünerek standartlaştırılmış artıklar elde edilir. Standartlaştırılmış artıklar

$$e_{is} = \frac{e_i}{s} \quad i=1,2,\dots,n \quad (2)$$

olarak hesaplanır. Bu değer $[-2, +2]$ aralığı dışında ise artıklar, aykırı gözlem şüphesiyle incelenir (Alpar, 2013).

2.2. Student Türü Artıklar

e_{is} artıklar her zaman standart normal dağılıma sahip olmadıklarından standartlaştırma işlemi için e_i hataları kendi standart sapması olan $s\sqrt{1 - p_{ii}}$ 'ye bölünür. Student türü artıklar

$$r_i = \frac{e_i}{s\sqrt{1 - p_{ii}}} \quad i=1,2,\dots,n \quad (3)$$

olarak elde edilir. Burada p_{ii} gözlem uzaklığı değeridir ve $0 < p_{ii} < 1$ arasında değer alır. Genellikle $[-2, +2]$ sınırların dışına çıkan gözlemlerin aykırı gözlem olabileceği düşünülür ve bu gözlemlerin incelenmesi gerekir(Alpar, 2013).

2.3. Etkili Gözlemler

Etkili gözlemler, veride bulunması veya veriden çıkartılması durumunda regresyon modelinin katsayılarını ve tahminlerini etkileyen gözlemlerdir (Montgomery vd, 2013). Veri setindeki bir gözlemin etkili gözlem olup olmadığının belirlemek için Cook Uzaklığı, DFITS ölçüsü ve DFBETAS ölçüsü kullanılır.

2.3.1. Cook Uzaklığı (C_i^2)

Cook Uzaklığı,

$$C_i^2 = \frac{1}{p+1} r_i^2 \frac{p_{ii}}{1-p_{ii}} \quad (4)$$

r student türü artıklar yardımıyla Cook uzaklığı hesaplanır. C_i^2 'nin büyük değerlerine karşılık gelen gözlemler etkili gözlemlerdir ve gözlem sayısı küçük olduğu durumlarda $4/(n-p-1)$ değeri ile karşılaştırılır.

2.3.2. DFBETAS Ölçüsü

DFBETAS_{j,i}'nin büyük değerleri i 'inci gözlemin j 'inci regresyon katsayısı üzerine dikkate değer bir şekilde etkili olduğunu göstermektedir.

$$DFBETAS_{j,i} = \frac{r_{ji}}{\sqrt{r_j' r_j}} \frac{r_i^*}{\sqrt{1-p_{ii}}} \quad (5)$$

olarak hesaplanır. X $n \times (p+1)$ boyutlu veri matrisi olmak üzere, $R = (X'X)^{-1}X'$ $(p+1) \times n$ boyutlu matristir. r_j' , R matrisinin j . satırını göstermektedir. r_i^* , student türü silinmiş artıkları göstermektedir. durumunda bu gözlemin incelenmesi gerekir (Alpar, 2013).

2.3.3. DFITS Ölçüsü

DFITS_i ölçüsü, C_i^2 'ölçüsüne benzemektedir. i . gözlemin \hat{y}_i kestirim değeri üzerine etkisini incelemek için kullanılır.

$$DFITS_i = r_i^* \sqrt{\frac{p_{ii}}{1-p_{ii}}} \quad (6)$$

olarak hesaplanır. $|DFITS_i| > 2\sqrt{(p+1)/n}$ koşulunu sağlayan gözlemler, etkili gözlem olarak sınıflandırılır (Alpar, 2013). Tablo 1' de etki ölçüleri verilmiştir.

Tablo 1. Etki ölçüleri

Etki ölçüsü	Formül	Karar kriteri
Cook Uzaklığı	$C_i^2 = \frac{1}{p+1} r_i^2 \frac{p_{ii}}{1-p_{ii}}$	$4/(n-p-1)$
DFBETAS	$DFBETAS_{j,i} = \frac{r_{ji}}{\sqrt{r_j' r_j}} \frac{r_i^*}{\sqrt{1-p_{ii}}}$	$2/\sqrt{n}$
DFITS	$DFITS_i = r_i^* \sqrt{\frac{p_{ii}}{1-p_{ii}}}$	$2\sqrt{(p+1)/n}$

3. KANTİL REGRESYON

Regresyon analizinde EKK yöntemi ile yapılan tahminler varsayımlarının sağlanamadığı durumlarda etkin tahmin olma niteliği taşımazlar. Bu durumda alternatif regresyon modellerine ihtiyaç duyarız. Parametrik modellerde en uygun matematiksel modelin seçilmesi gibi, alternatif regresyon modellerinin de en uygun modelin seçilip kullanılması gerekir. Alternatif regresyon modellerinden biri de Kantil regresyondur ve diğer regresyon modellerinde olduğu gibi bu yönteminde amacı değişkenler arasındaki ilişkiyi açıklamaktır.

Kantil regresyon, bağımlı değişkeninin koşullu dağılımının seçilen kantilleri için geliştirilmiştir. Klasik regresyon modelinin aksine hata varyansının homojenliği ve hataların dağılımı hakkında herhangi bir varsayımı yoktur. Bu durumlardan dolayı lineer regresyondan daha esnek bir yaklaşım olarak kabul edilebilir. Klasik regresyon bağımlı değişkeninin koşullu beklenen değeri için bir model ararken, Kantil regresyon bağımlı değişkeninin koşullu dağılımında seçilen kantiller için model belirler. Klasik regresyon bağımlı değişkeninin koşullu ortalaması ile artık kareler toplamını minimizasyonuna dayanırken, Kantil Regresyon fonksiyonları mutlak artıkların ağırlıklandırılmış toplamının minimizasyonuna dayanmaktadır. Kantil regresyon modeli,

$$y_i = x_i' \beta_\theta + u_{\theta i} \tag{7}$$

şeklinde ifade edilir. Burada, x_i ($k \times 1$) boyutlu bağımsız değişkenler vektörüdür ve bağımlı değişken y_i 'nin koşullu dağılımının θ 'nci kantili ile bağımsız değişkenler arasındaki doğrusal regresyonu göstermektedir. β_θ , θ 'nci kantil regresyona ile ilgili parametreler vektörüdür. $u_{\theta i}$ hata vektörüdür. y_i 'nin θ 'nci koşullu kantili,

$$Q(y_i / x_i) = x_i' \beta_\theta \tag{8}$$

Kantil regresyonunun amaç fonksiyonu, mutlak sapmaların ağırlıklandırılmış toplamlarıdır. θ 'inci kantil regresyon için amaç fonksiyonu

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{\theta y_i < x_i' \beta} \theta |y_i - x_i' \beta| + \sum_{\theta y_i \geq x_i' \beta} (1-\theta) |y_i - x_i' \beta| \right\} \text{ dir ve} \tag{9}$$

$$\hat{\beta}_\theta = \frac{1}{n} \sum_{i=1}^n \rho_\theta(y_i - x_i' \beta) = \sum_i \rho_\theta(u_{\theta i}) \text{ minimizasyonu ile tahmin edilir. Kantil regresyonun bu formda gösterimi}$$

doğrusal programlama gösterimidir. $p\theta$ fonksiyonu,

$$i \quad i \quad \square \quad \square$$

$$i \square 1 \quad i$$

gösterimi doğrusal programlama gösterimidir. $p\theta$ fonksiyonu,

$$u_{\theta i} < 0 \text{ yani } (y_i < x_i' \beta) \text{ ise } (1-\theta)u \tag{10}$$

$$u_{\theta i} \geq 0 \text{ yani } (y_i \geq x_i' \beta) \text{ ise } \theta u \tag{11}$$

olarak belirlenir. Kantil regresyonda farklı kantillerde farklı sonuçlar elde edilmektedir. Bu durum bağımlı değişkeninin koşullu dağılımının farklı noktalarındaki açıklayıcı değişkenlerdeki değişikliklere farklı tepki vermesi olarak yorumlanabilir.

Kantil Regresyonun özellikleri kısaca aşağıda verilmiştir.

- EKK regresyonu y 'nin koşullu dağılımının ortalaması hakkında bilgi verirken, Kantil regresyon ise farklı kantil değerleri için y 'nin x 'e göre koşullu dağılımının tümü hakkında bilgi vermektedir.
- Kantiller y 'deki sapan değerlere karşı kararlıdır.
- Hata terimi normal dağılmadığında, Kantil regresyon tahmin edicileri EKK tahmin edicilerinden çok daha etkin olabilir.
- Kantil regresyonda ki,

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_\theta(y_i - x_i' \beta) \tag{12}$$

bu minimisasyon doğrusal programlama gösterimidir ve tahmini kolaylaştırır.

- EKK regresyon modelinde hata terimlerinin varyansları homojen varsayılırken, Kantil regresyon modelinde ise hata terimlerinin değişkenliğine izin verilmekte ve varyans yapısına ilişkin herhangi bir varsayım bulunmamaktadır.

4. UYGULAMA

Deneyisel çalışmaların birçoğunda bağımlı değişkenin ortalama değerini açıklayan değişkenlerin belirlenmesi için EKK regresyonu kullanılmaktadır. Kantil regresyon bağımlı değişkenin sadece ortalama değeri ile değil, tüm dağılımı ile ilgili çıkarılma yapmamıza olanak tanımaktadır. Bağımlı değişkenin ortalama değeri için elde edilen regresyon denklemi istatistiksel olarak anlamlı olmasa bile farklı kantil değerleri için elde edilen regresyon denklemleri anlamlı çıkabilmektedir. Bu da birçok deneysel çalışmanın ortalama etrafındaki regresyonun anlamsız çıkması nedeniyle tekrarlanmasını ya da çalışmanın tamamen çöpe atılmasına engel olmaktadır. Ayrıca deneysel çalışmalarda katlanılan maliyet ve zaman göz önüne alındığında kantil regresyonun EKK regresyonuna göre daha avantajlı bir regresyon tekniği olduğu görülmektedir.

Bu çalışmanın uygulama bölümünde Altındağ (2003)' in yaptığı beton kırma deneyinde elde ettiği sonuçlar kullanılmıştır. Altındağ'ın bu çalışmasında, açığa çıkan enerji miktarı ile darbe değeri arasında ilişki kurulmaya çalışılmıştır. Darbe değeri, gevreklik olarak adlandırılan bir değişken ile ölçülür. Açığa çıkan enerji ise spesifik enerji (MJ/m³) olarak adlandırılır. Çalışmada beton kırılırken açığa çıkan enerji miktarı (spesifik enerji) bağımlı değişken ve gevreklik değişkeni de bağımsız değişken olarak ele alınmıştır. Beton verisi için tanımlayıcı istatistikler Tablo 2' de verilmiştir.

Tablo 3. Beton verisi için tanımlayıcı istatistikler

Variables	Minimum	Q25	Q50 (Median)	Mean	Q75	Maximum
Enerji	1,30	8,16	19,20	27,03	37,50	103,09
BE3	6,58	36,63	77,87	324,36	204,88	2491,24

EKK yöntemi ile spesifik enerji ve gevreklik arasında kurulan regresyon modeli sonuçları Tablo 3'de ve regresyon denklemi eşitlik (13) 'de verilmiştir.

Tablo 3. Beton verisi için elde edilen EKK sonuçları

Coefficients	Estimate	Std. Error	t value	Pr(> t)
İntercept	14,964578	3,492765	4,284	0,000277***
BE3	0,037191	0,005285	7,037	3,6e-07***
Residual standart error: 15,22			Degrees of freedom : 23	
Multiple R-squared: 0,6829			Adjusted R-squared : 0,6691	
F-statistic: 0,6829		Degrees of freedom : 1 and 23		p-value: 3,596e-07

Regresyon denklemi,

$$\hat{y}_i = 14,965 + 0,037x_i \quad i = 1,2,\dots,n \quad (12)$$

Buna göre EKK yöntemi ile kurulan model ve regresyon katsayıları %95 anlam düzeyinde istatistiksel olarak anlamlıdır. Elde edilen bu denklemin çıkarılma amaçlı kullanılabilmesi için daha önce de değinildiği gibi hataların normal dağılıma sahip olması ve aykırı değer içermemesi gerekmektedir. Öncelikli olarak hataların normal dağılıma sahip olmadığı farklı yöntemlerle gösterilmiştir.

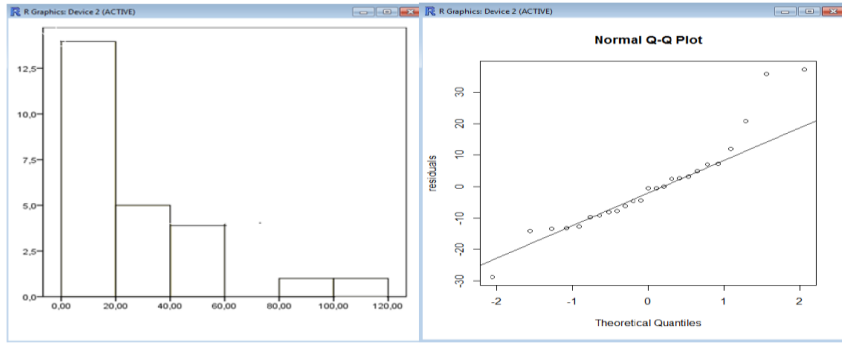
Shapiro-Wilk Testi: Shapiro-Wilk testi en güçlü normallik testlerinden biridir. Örneklem büyüklüğü 3 -50 sınıf aralığına kadar uygulanabilir. Normallik araştırması için hipotezler,

H0: Hatalar normal dağılıma sahiptir

H1: Hatalar normal dağılıma sahip değildir

şeklinde kurulur. Bu test uygulandığında elde edilen p=0,03412 değeri $\alpha=0,05$ anlamlılık düzeyinden büyük olduğu için H0 hipotezi reddedilir ve hataların normal dağılıma sahip olmadığı sonucuna varılır. Hatalara ilişkin histogram ve Q-Q grafiği de Normallik araştırmasında kullanılan görsel yöntemlerdir. Eşitlik (13)' de verilen denklemden elde edilen hatalara ilişkin histogram ve Q-Q grafiği Şekil 2'de verilmiştir.

Şekil 2'ye bakıldığında histogramın Normal dağılım görünümünde olmadığı açıktır. Q-Q grafiğinde ise alt ve üst uçlarda diğer gözlemlerden oldukça farklılık gösteren veriler dikkat çekmektedir. Bu veriler verinin Normal dağılmadığını göstermektedir. Ayrıca bu verilerin dikkatle incelenmesi gerekmektedir. Çeşitli etki ölçülerine göre tüm hata değerleri incelenmiş ve Tablo 4'de verilmiştir.



(a) (b)
Şekil 2. a) Hatalara ilişkin histogram b) Q-Q grafiği

Tablo 4. Modeldeki etkili gözlemler

	COOK	DFITS	DFBETAS
1	0,02212	-0,72375	0,00052
2	0,02191	-0,72743	0,00053
3	0,02499	-0,77252	0,00056
4	0,01046	-0,47806	0,0003
5	0,01013	-0,48821	0,00035
6	0,0054	0,33681	-0,00019
7	0	-0,00377	0
8	0,00003	-0,02619	0,00002
9	0,00075	0,13063	-0,00009
10	0,00004	-0,02863	0,00002
11	0,01632	0,59812	-0,00038
12	0,00105	0,1443	-0,00006
13	0,00076	0,13056	-0,00009
14	0,00798	-0,42971	0,0003
15	0,00765	-0,42993	0,00032
16	0,60071	-7,64862	-0,00521
17	0,00452	-0,31808	0,00021
18	0,01794	-0,62355	0,00038
19	0,00186	-0,18549	0,00001
20	0,00242	0,2186	0,00009
21	0,01196	0,70383	0,00061
22	0,13643	1,62658	-0,00058
23	0,04264	0,90759	-0,00031
24	0,32646	3,81212	0,00326
25	0,17315	-6,97284	-0,00301

COOK uzaklığı için değeri ile, $4 / (n-p-1) = 0,17$ değeri ile,

DFITS ölçüsü için $2\sqrt{(p+1)/n} = 0,57$ değeri ile,

DFBETAS ölçüsü için $2/\sqrt{n} = 0,4$ değeri ile karşılaştırılmıştır ve etkili gözlemler kırmızı renkte yazılmıştır.

EKK regresyon denkleminin çıkarsama amaçlı kullanılması mümkün olmadığından deneyin tekrar edilmesi gerekir. Bu hem zaman hem de maliyet açısından büyük sorunlar yaratmaktadır. Beton verisi için bağımlı değişkenin $\tau = 0.25$, $\tau = 0.50$ ve $\tau = 0.75$ kantil değerleri için elde edilen kantil regresyon sonuçları Tablo 5'te verilmiştir.

Tablo 5. Kantil regresyon sonuçları

$\tau = 0.25$				
Coefficients	Value	Std.Error	t value	Pr(> t)
Sabit (Intercept)	5,80032	3,21197	1,80584	0,08405
BE3	0,03905	0,00882	4,42652	0,00019
R2 = 0,924491				
$\tau = 0.50$				
Coefficients	Value	Std.Error	t value	Pr(> t)
Sabit (Intercept)	14,48486	3,64834	3,97026	0,00061
BE3	0,03557	0,2142	1,66079	0,11033
R2 = 0,861969				
$\tau = 0.75$				
Coefficients	Value	Std.Error	t value	Pr(> t)
Sabit (Intercept)	17,25858	8,19713	2,10544	0,04638
BE3	0,04269	0,01238	3,44856	0,00219
R2 = 0,682869				

$\tau = 0.25$, $\tau = 0.50$ ve $\tau = 0.75$ kantil regresyon denklemleri ise sırasıyla,

$$\hat{y}_i = 5,800 + 0,39x_i \quad i=1,2,\dots,n \quad (14)$$

$$\hat{y}_i = 14,484 + 0,035x_i \quad i=1,2,\dots,n \quad (15)$$

$$\hat{y}_i = 17,258 + 0,042x_i \quad i=1,2,\dots,n \quad (16)$$

şeklinde yazılabilir. Oluşturulan modellerin anlamlığı $\alpha = 0,05$ anlam düzeyinde test edilmiştir. Bağımlı değişkenin $\tau = 0.25$ 'inci kantil değerine göre kurulan model ve BE3 değişkeninin katsayısı istatistiksel olarak anlamlıdır. $\tau = 0.50$ 'e göre kurulan kantil regresyon sonuçları incelendiğinde ise model ve sabit terimin anlamlı olduğu görülmüştür. $\tau = 0.75$ için kurulan kantil regresyon sonuçlarına göre model, sabit terim ve BE3 değişkeninin katsayısı istatistiksel olarak anlamlıdır.

5. SONUÇ

Birçok mühendislik çalışmasına bakıldığında regresyon analizinin yaygın olarak kullanıldığı görülmüş, ancak regresyon analizinin varsayımlarının kontrolünün yapılmadığı tespit edilmiştir. Gerçek veri seti ile çalışılırken regresyon analizinde karşılaşılan en büyük problem hata terimlerinin normal dağılıma uymamasıdır. EKK

Regresyonunun çıkarsama amaçlı kullanılabilmesini için sağlanması gereken en önemli varsayımlardan biri budur. Bu çalışmada mühendislik alanında regresyon analizi uygulamalarına farklı bir yaklaşım olarak Kantil regresyon yöntemi tanıtılmış ve beton kırma deneyine ilişkin verilerin analizine uygulanmıştır.

EKK tekniği kullanıldığında, gevreklikteki (BE3) bir birim artışın açığa çıkan enerji miktarı üzerinde yaratacağı ortalama artış miktarı 0,037 birim olacağı saptanmıştır. Modelin açığa çıkan enerji miktarını açıklama oranı 0,6829'dur. Model ve katsayılar istatistiksel olarak anlamlıdır. Ancak istatistiksel bir teknikten elde edilen sonuçların güvenilirliği, kullanımı üzerindeki varsayımların sağlanması ile doğru orantılıdır. EKK tekniği ile elde edilen modelin çıkarsama amaçlı kullanılması, hata terimleri normal dağılıma sahip olmadığından ve aykırı değer içerdiğinden uygun değildir. Bu model yardımıyla yapılacak çıkarsamalar gerçeği yansıtmayacaktır.

Kantil regresyon sonuçlarına bakıldığında gevreklikteki (BE3) bir birimlik artışın spesifik enerjinin çeşitli kantilleri üzerinde farklı miktarda artışa neden olduğu Tablo 5'ten görülmektedir. Bağımlı değişkenin $\tau = 0.50$ 'inci kantil değerine (bir başka ifade ile medyan değerine) göre kurulan modelde BE3 değişkeninin anlamsız olduğu belirlenmiştir. Bağımsız değişkenin anlamsız olduğu bu model çıkarsama amaçlı kullanılamaz. Bağımlı değişkenin $\tau = 0.25$ 'inci kantil değerine göre kurulan modelde ise sabit terim anlamsızdır. Modelin açığa çıkan enerji miktarını açıklama oranı 0,9245'tir. $\tau = 0.75$ 'inci kantil değerine göre kurulan modelde hem sabit terim hem de BE3 değişkeni istatistiksel olarak anlamlıdır. Bu modelin bağımlı değişkeni açıklama oranı 0,6829'dur. Kurulan modelde sabit terimin bulunması teknik açıdan gerekli olduğu durumda $\tau = 0.75$ 'inci kantil değerine göre kurulan model, gerekmediği durumda ise $\tau = 0.25$ 'inci kantil değerine göre kurulan model tercih edilmelidir. Her iki modelin bağımlı değişkeni açıklama oranı EKK tekniği ile kurulan modelden az değildir. Ayrıca varsayımsal bozulmalara karşı da dayanıklı olduğundan çıkarsama amaçlı kullanılabilir.

Bu çalışmada bir maden mühendisliği uygulaması olan spesifik enerji miktarı modellenmeye çalışılmıştır. Ayrıca bu çalışma ile mühendislik alanındaki çalışmalarda kullanılan klasik EKK regresyonuna alternatif olarak Kantil regresyon tanıtılmıştır. Mühendislik alanında yapılacak çalışmalarda araştırmacılara yol gösterici bir çalışma olabilir.

6. REFERANSLAR

- Alpar, R. (2013). Uygulamalı çok değişkenli istatistiksel yöntemler. Ankara, Detay Yayıncılık.
- Altındağ, R. (2003). Correlation of specific energy with rock brittleness concepts on cutting, *The Journal of the South African Institute of Mining and Metallurgy*, 15, 163-171.
- Bassett, G.W. & Chen, H-L. (2001). Quantile style: return-based attribution using regression quantiles, *Physica- Verlag HD, Chicago*, 293-305.
- Buchinsky, M. (1994). Changes in the u.s. wage structure 1963-1987: application of quantile regression, *The Econometric Society*, 62(2), 405-458. doi: 10.2307/2951618.
- Cai, Y. & Reeve, D.E. (2013). Extreme value prediction via a quantile function model. *Coastal Engineering*, 77, 91–98. doi:10.1016/j.coastaleng.2013.02.003.
- Chen, C. & Wei, Y. (2005). Computational issues for quantile regression. special issue on quantile regression and related methods, *The Indian Journal of Statistics*, 67(2), 399-417. doi: 10.2307/i25053424
- Crowley, J.& Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72, 27-36. doi: 10.1080/01621459.1977.10479903
- Çağlayan E. & Arıkan E. (2011). Determinants of house prices in istanbul: a quantile regression approach. *Qual, Quant*, 45, 305-317. doi:10. 1007/s11135-009-9296-x.
- Dehghani, H., Vahidi, B., & Hosseinian, S.H. (2017). Wind farms participation in electricity markets considering uncertainties. *Renewable Energy*, 101, 907-918. doi:10.1016/j.renene.2016.09.049.
- Ergül, B. (2003). Robust regresyon ve uygulamaları. Yüksek Lisans Tezi. Eskişehir Osmangazi Üniversitesi, Eskişehir.
- He, Y., Liu, R., Li, H., Wang, S., & Lu, X. (2016). Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. *Applied Energy*, 185, 254–266. doi:10.1016/j.apenergy.2016.10.079.
- Hendricks, W. & Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American Statistical Association*, 87, 58-68. doi: 10.1080/01621459.1992.10475175.
- Huang, Y. F., Mirzaei, M., & Amin, M.Z.M. (2016). Uncertainty Quantification in Rainfall Intensity Duration Frequency Curves based on Historical Extreme Precipitation Quantiles. *Procedia Engineering*, 154, 426–432. doi:10.1016/j.proeng.2016.07.425.
- Hüdaverdi, T. (2015), Farklı regresyon modelleri ile patlatma kaynaklı yer sarsıntısının tahmin edilmesi, *Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 30(1), 141-150. doi: 22746/242810.
- Koenker R. (2005). *Quantile regression*, Cambridge University Press, NY 10011-4211, New York, USA. Koenker, R. & Bassett, G. (1978). Regression quantiles, *Econometrica*, 46(1), 33-50. doi: 10.2307/1913643.
- Koenker, R. & Geling, O. (2001). A quantile regression survival analysis, *Journal of the American Statistical Association*, 96, 458-468. doi: 10.1198/016214501753168172.
- Koenker, R., & Hallock K., F. (2001). Quantile regression an introduction. *Journal of Economic Perspectives*, 15(4):143–156. doi:10.2307/i346045.
- Koenker, R., & Schorfheide, F. (1994). Quantile spline models for global temperature change. *Climatic Change*, 28, 395-404. doi:10.1007/BF01104081.
- Lv, Z., Zhao, J., Lia, Y., & Vang, W. (2016). Use of a quantile regression based echo state network ensemble for construction of prediction Intervals of gas flow in a blast furnace. *Control Engineering Practice*, 46, 94–104. doi:10.1016/j.conengprac.2015.10.003.
- Machado, A.F. & J, Mata. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression, *Journal of Applied Econometrics*, 20(4), 445-465. doi: 10.1002/jae.788.
- Martins, P.S. & Pereira, P.T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries, *Labour Economics*, 11(3), 355-371. doi: 10.1016/j.labeco.2003.05.003.

- Montgomery, D.C. & Peck, E.A. & Vining, G.G., (2013), Doğrusal regresyon analizine giriş, Ankara, Nobel Akademik Yayıncılık.
- Muraleedharan, G., Lucas, C., & Guedes Soares, C. (2016). Regression quantile models for estimating trends in extreme significant wave heights. *Ocean Engineering*, 118, 204–215. doi:10.1016/j.oceaneng.2016.04.009.
- Muthusamy, M., Godiksen, P.N., & Madsen, H. (2016). Comparison of different configurations of quantile regression in estimating predictive hydrological uncertainty. *Procedia Engineering*, 154, 513–520. doi: 10.1016/j.proeng.2016.07.546.
- Ovla, H.D. & Taşdelen, B. (2012), Aykırı değer yöntemi, Mersin Üniversitesi Sağlık Bilimleri Dergisi, 5(3), 1-8.
- Pandey, G.R. & Nguyen, V.T.V. (1999). A comparative study of regression based methods in regional flood frequency analysis, *Journal of Hydrology*, 225, 92–101. doi:10.1016/S0022-1694(99)00135-3.
- Seo, J.H., Perry, V.G., Tomczyk, D. & Solomon G.T. (2014). Who benefits most? The effects of managerial assistance on high-versus low-performing small businesses, *Journal of Business Research*, 67, 2845-2852. doi: 10.1016/j.jbusres.2012.07.003.
- Tan, X-P., & Wang, X-Y. (2016). Dependence changes between the carbon price and its fundamentals: A quantile regression approach. *Applied Energy*, 190, 306–325. doi:10.1016/j.apenergy.2016.12.116.
- Tareghian R. & Rasmussen, P. (2013). statistical downscaling of precipitation using quantile regression. *Journal of Hydrology*, 487, 122-135. doi:10.1016/j.jhydrol.2013.02.029.
- Tukey, J.W. (1977). *Exploratory data analysis*, Addison-Wesley Publishing Company,
- Walfish, S. (2006). A review of statistical outlier methods. *Pharmaceutical Technology*, 30(11), 82-88.
- Wang, D. H. -M., Yu, T. H. -K., & Liu, H. -Q. (2013). Heterogeneous effect of high-tech industrial R&D spending on economic growth. *Journal of Business Research*, 66(10), 1990–1993. doi:10.1016/j.jbusres.2013.02.023
- Yu, K., Lu, Z. & Stander, J. (2003). Quantile regression: applications and current research areas, *Journal of the Royal Statistical Society: Series D (The Statistician)*. 52,331-350. doi: 10.1111/1467-9884.00363.
- Yu, T.H-K. (2011). Heterogeneous effects of different factors on global ICT adoption, *Journal of Business Research*, 64, 1169-1173. doi: 10.1016/j.jbusres.2011.06.017.
- Yu, T. H. -K., Wang, D. H. -M., & Chang, L. -Y. (2011). Examining the heterogeneous effect of healthcare expenditure determinants. *International Journal of Behavioural and Healthcare Research*, 2(3), 205–213. doi: 10.1504/IJBHR.2011.041044