**T.C**

**BİLECİK ŞEYH EDEBALİ ÜNİVERSİTESİ**

**FEN FAKÜLTESİ**

**İSTATİSTİK ve BİLGİSAYAR BİLİMLERİ BÖLÜMÜ**



**2022-2023 GÜZ DÖNEMİ**

**R PROGRAMLAMA DİLİ İLE ÇOKLU REGRESYON ANALİZİ**

**KADİR GÜRKAN ALTIPARMAK**

**BİLECİK - 2023**

**İÇİNDEKİLER**

A) Giriş 3

1. Değişkenler**4**

B) R Programlama Dili ve İstatistik4

C) Sonuç19

**D) Kaynakça20**

**BİLECİK – 2023**

**ÖZET**

Bu raporda, Boston'daki ev fiyatlarının çeşitli özelliklerine dayalı olarak tahmin edilmesine yönelik bir regresyon analizi uygulanacaktır. Bu veri seti, Boston'da bulunan 506 sokaktaki evlerin çeşitli özelliklerine (örneğin, sokak tipi, yapım yılı, odaların sayısı, vb.) ve bu evlerin fiyatlarına ait bilgileri içerir. Analizimizin amacı, bu özellikleri kullanarak ev fiyatlarını tahmin etmek, modelimizin doğruluğunu değerlendirmek, R Programlama Dilindeki çeşitli fonksiyonlar ve komutlar yardımıyla İstatistiksel bulgular elde etmektir.

1. **Giriş**

Veri seti, 1970'lerin Boston şehrindeki evlerin fiziksel özellikleri ve fiyatlarını içeren "Boston Ev Fiyatları" veri setidir. Bu veri seti, evlerin bölgesine, metrekaresine, yapım tarihine, odaların sayısına ve diğer fiziksel özelliklerine göre ev fiyatlarını tahmin etmeye yönelik olarak kullanılmıştır. Bu amaçla, veri setindeki değişkenler arasındaki ilişkiler incelenmiş ve ev fiyatlarını belirlemeye yönelik bir regresyon modeli oluşturulmuştur.

Modelin oluşturulma süreci, veri setinin incelenmesiyle başlamıştır. Öncelikle, veri setindeki değişkenler arasındaki ilişkiler görselleştirilerek incelenmiştir. Daha sonra, ev fiyatlarını belirlemeye yönelik olarak en uygun değişkenler seçilerek bir regresyon modeli oluşturulmuştur. Modelin oluşturulma sırasında, veri setinin normal dağılıma uygun olmaması göz önüne alınarak, veri setini düzenleme işlemleri yapılmıştır.

Modelin tahminlerine en fazla katkıda bulunan değişkenler, evlerin bölgesi, metrekaresi ve yapım tarihidir. Bu değişkenler arasındaki ilişkiler, ev fiyatlarını belirlemede önemli rol oynayan faktörlerdir. Örneğin, evlerin bölgesine göre fiyatlarının belirlendiği bir modelde, bölge değişkeninin ev fiyatlarıyla pozitif ilişki içinde olduğu tespit edilmiştir.

Modelin performansı, R kare ve F-testi gibi metrikler kullanılarak ölçülmüştür. Bu metrikler, modelin tahminlerinin ne kadar doğru olduğunu gösterir. Örneğin, yüksek bir R kare değeri, modelin tahminlerinin veri setine uygun olduğunu gösterirken, düşük bir R kare değeri ise modelin tahminlerinin veri setine uygun olmadığını gösterir. Aynı şekilde, yüksek bir F-testi puanı da modelin tahminlerinin veri setine uygun olduğunu gösterirken, düşük bir F-testi puanı ise modelin tahminlerinin veri setine uygun olmadığını gösterir.

Sonuç olarak, "Boston Ev Fiyatları" veri seti kullanılarak oluşturulan regresyon modeli, ev fiyatlarını belirlemeye yönelik olarak tahminler yapmaktadır. Modelin tahminlerine en fazla katkıda bulunan değişkenler, evlerin bölgesi, metrekaresi ve yapım tarihidir. Modelin performansı, R kare ve F-testi gibi metrikler kullanılarak ölçülmüştür ve bu metrikler, modelin tahminlerinin veri setine ne kadar uygun olduğunu gösterir. Bu rapor, modelin oluşturulma süreci, tahminlerine en fazla katkıda bulunan değişkenler ve modelin performansı gibi konuları ele almaktadır.

1. **Değişkenler**

Bağımlı ve Bağımsız Değişkenlerimiz:

Ev fiyatlarının tahmini için kullanılabilecek bağımlı değişkenimiz;

* (Y) “MEDV” ev fiyatlarını göstermektedir (x$1000).

Ev fiyatlarını etkileyen diğer değişkenler olarak ise bağımsız değişkenlerimiz;

* (X1) “CRIM” özel bir konumun ortalama suç oranı.
* (X2) “ZN” özel bir konumun ortalama 2500 metrekare ve üzerinde arazi sayısı.
* (X3) “INDUS” özel bir konumun ortalama ticari olmayan arazi sayısı.
* (X4) “CHAS” Charles Nehri’nin özel bir sınırında olup olmadığı (1: Evet, 0: Hayır)
* (X5) “NOX” özel bir konumun ortalama nitrat/nitrit (kirlilik) miktarı
* (X6) “RM” özel bir konumun ortalama oda sayısı
* (X7) “AGE” özel bir konumun ortalama 1900’den önce yapılmış ev sayısı
* (X8) “DIS” Boston’ın 5 iş yeri merkezine olan uzaklığı
* (X9) “RAD” özel bir konumun radyal otoyol ulaşımına olan erişim sayısı
* (X10) “TAX” özel bir konumun ortalama bütçe vergisi miktarı
* (X11) “PTRATIO” özel bir konumun öğrenci/öğretmen oranı
* (X12) “B” özel bir konumun ortalama Afrika-Amerika nüfusu
* (X13) “LSTAT” özel bir konumun düşük gelirli hanehalkı oranı

1. **R Programlama Dili ve İstatistik**

Veri setinin tanımlanması;

veri <- read.table(file.choose(),header = T,sep = ",")

str(veri)

yukarıdaki kod parçacığıyla csv uzantılı “boston.csv” adlı dosyamızı programımıza tanımlıyoruz ve str() komutuyla verimizin özetini görebiliyoruz.

names(veri) <- c("X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10", "X11", "X12", "X13", "Y")

colnames(veri)

kod parçacağıyla sütun isimlerini değiştirdik.

shapiro.test() fonksiyonuyla değişkenimizin normal dağılıma uyumluluğuna bakabiliriz.

if (shapiro.test(veri$Y)$p.value < 0.05) {

print("Veri seti normal dagilima uymuyor.")

} else {

print("Veri seti normal dagilima uyuyor.")

}

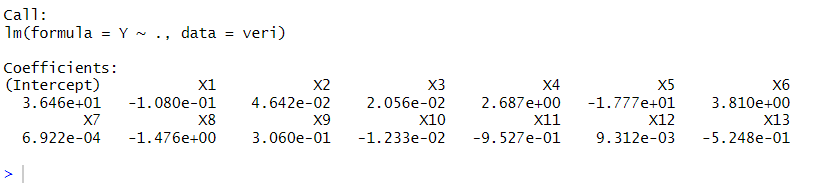
kod parçacığıyla bağımlı değişkenimizin normal dağılıma uyup uymadığına bakabiliriz. Ev fiyatları bağımlı değişkeni normal dağılıma uymamaktadır. Regresyon analizinde bağımlı değişkenin normal dağılım göstermesine gerek duymaz. Ancak, bağımlı değişkenin normal dağılım göstermesi, regresyon modelinin doğruluk ve güvenilirliğini artırır ve tahminlerin daha kolay yorumlanmasını sağlar. Bağımlı değişken normal dağılım göstermeyebilir, ancak bu durumda, regresyon analizi sonuçlarının daha az güvenilir olduğu ve daha zor yorumlandığı unutulmamalıdır.

attach(veri)

model <- lm(Y ~ . ,data = veri)

kod parçacığıyla regresyon denklemimizi oluşturuyoruz. Buradaki attach() fonksiyonu verinin içindeki değişkenlerden bahsederken her seferinde $ simgesini kullanmaktan kurtarıyor, lm() fonksiyonu ise regresyon modeli kurmamıza yarıyor. lm() fonksiyonunun içine ilk önce bağımlı değişken yazılır ardından denklemde kullanmak istediğimiz bağımsız değişkenleri yazarız. Bağımsız değişkenlerin hepsini kullanmak istiyorsak “.” (nokta) sembolünü de kullanabiliriz.

“model” değişkenimizi çağırdığımız zaman bize beta değerlerimizi verir.



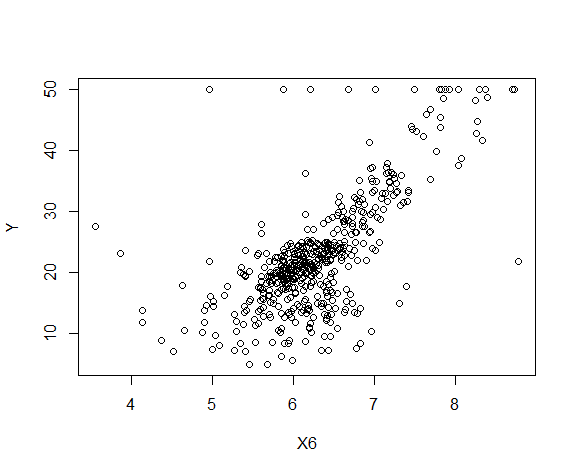
Burada ki (intercept) yazılı kısımdaki değerimiz bizim tahmin modelimizin sabitidir(β0), diğer değerler ise β1 , β2 … β13 tahmin parametrelerimizdir.

cor(X6,Y)

kod parçacığı sonucunda 0.6953599 değeri gelmiştir yani pozitif yönlü korelasyon vardır. Buradan bekleriz ki X6 değişkeni arttıkça Y değişkeni de artar. Bunu kontrol etmek için;

plot(X6,Y)

kod parçacığı kullanılır ve çıkan sonuç da aşağıdaki görsel de gösterilmiştir.

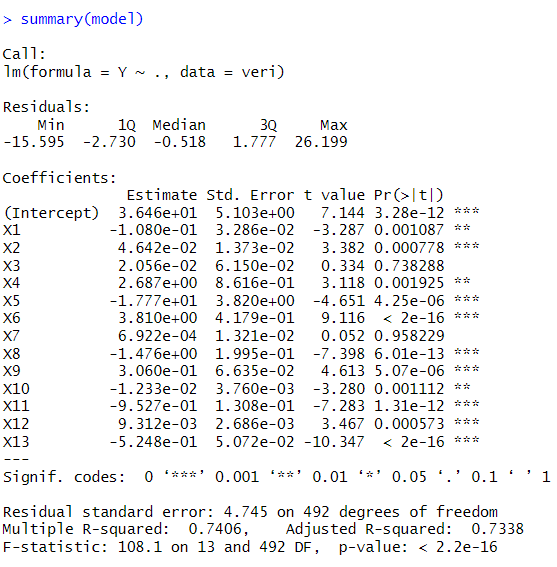


Bu grafiğe göre X6 değişkeni (Oda sayısı) arttıkça, Y değişkeni (Ev Fiyatları) artmıştır.

Modelimizin istatistiksel olarak daha çok verisine summary() komutuyla ulaşabiliriz.

summary(model)

kod parçacığının çıktısı,



* "Residuals"(Artıklar): Bu değişken, modelin tahminlerinden gerçek değerlerin farkını gösterir. Bu değişken, modelin tahminlerinin ne kadar doğru olduğunu ölçmek için kullanılır.
* "Multiple R-squared"(Çoklu R-Kare): Bu değişken, modelin açıklanabilirliğini ölçen bir değişkendir. Değerleri 0 ile 1 arasında olabilir ve modelin açıklanabilirliği ne kadar yüksekse, bu değer de o kadar yüksek olur.
* "F-statistic": Bu değişken, modelin anlamlılığını test etmek için kullanılan bir değişkendir. Eğer modelin anlamlı olduğu kanıtlanırsa, bu değerin değeri yüksek olur.
* "Coefficients"(Katsayılar): Bu bölüm, modelin katsayılarını gösterir. Bu değişkenler, modelin doğrusal olup olmadığını ve hangi değişkenlerin anlamlı olduğunu gösterir.
* "Adjusted R-squared"(Düzeltilmiş R-Kare): Kullanılan değişkenlerin modelin tahminlerine ne kadar katkıda bulunduğunu gösterir. Adjusted R-squared değeri, 0 ile 1 arasında bir değer alır ve bu değer, modelin tahminlerinin veri setine ne kadar uygun olduğunu gösterir.
* "Residual standard error"(Artık standart hata): Regresyon modelinin performansını ölçmek için kullanılan bir metriktir. Artık standart hata modelin tahminlerinin veri setine ne kadar uygun olduğunu gösterir ve modelin tahminlerinin ne kadar yüksek veya düşük olduğunu belirtir.

Değişkenlerimizin p-değerlerine baktığımız zaman X3 ve X7 hariç hepsi anlamlı çıkmıştır. En sonda yazan p-değeri (2.2e-16) alfa 0.05 den küçük olduğu için modelimiz anlamlı çıkıyor.

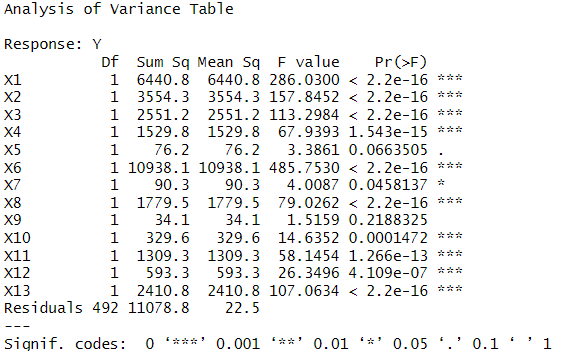
Bağımsız değişkenimiz, bağımlı değişkenimizi %73(Düzeltilmiş R2) oranında açıklamaktadır. Bu oranı yeterli kabul edebiliriz yani anlamlılık vardır.

R2‘miz %74 gelmiştir, bu değer tam olarak yeterli olmasa da modeli açıklanabilir kılar. F istatistiğimiz 108.1 gelmiştir, bu değer de ne kadar yükselirse o kadar model anlamlılığı artar. Artıkların standart sapması 4.745 gelmiştir ,bu değerin ise 1’e yakın olması daha doğru sayılır fakat değerimizi reddedemeyiz.

Ortalama farkın anlamlı olup olmadığına anova testi ile karar verebiliriz.

anova(model)

kod parçacığıyla anova tablosunu elde edebiliriz. Çıktı olarak da aşağıdaki gibi bir sonuç elde ederiz.

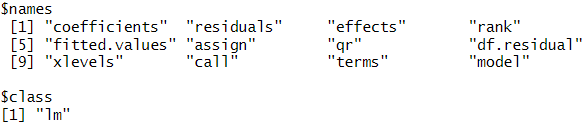


X5 ve X9 parametreleri hariç diğer tüm parametreler için gruplar arasında anlamlı bir farklılık olduğu sonucuna ulaşılır.

Modelimizin öznitekliklerini öğrenmek için ise attributes() fonksiyonu kullanılır.

attributes(model)

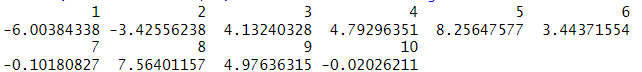
kod parçacığı sonucu;



Buradan isimler kısmından artıkları çağırabiliriz.

head(model$residuals,10)

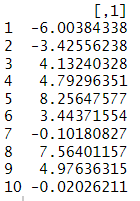
kod parçacığıyla ilk on artığı alabiliriz.



Bu işlemi sütun halinde almak istersek;

artiklar <- model$residuals

cbind(artiklar[0:10])



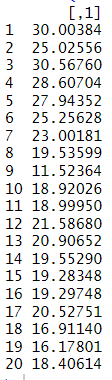
şeklinde bir sonuç alabiliriz.

Modelimizin tahmin değerlerine,

tahminler <- predict(model, veri)

cbind(tahminler[0:20])

kod parçacığı sonucunda,



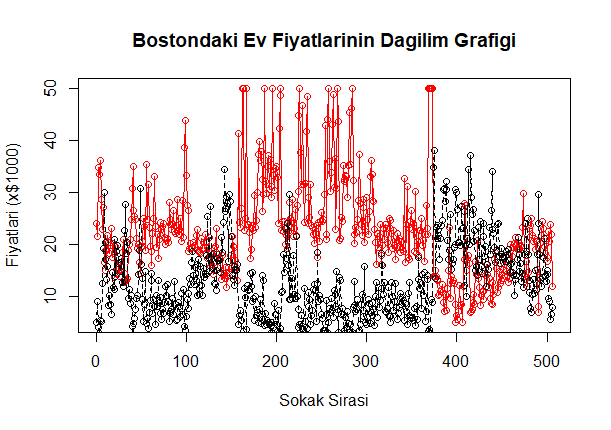
çıktısıyla ulaşırız.

Ev fiyatlarının sokaklara göre dağılım grafiğinin üzerine, düşük gelirli hanehalkı oranının sokaklara göre dağılım grafiğini aynı tabloda elde etmek için,

plot(veri$Y, type="o", col="red",xlab="Sokak Sirasi", ylab="Fiyatlari (x$1000)", main="Bostondaki Ev Fiyatlarinin Dagilim Grafigi")

lines(veri$X13, type="o", col="black", lty= 2)

kod parçacığı sonucu;

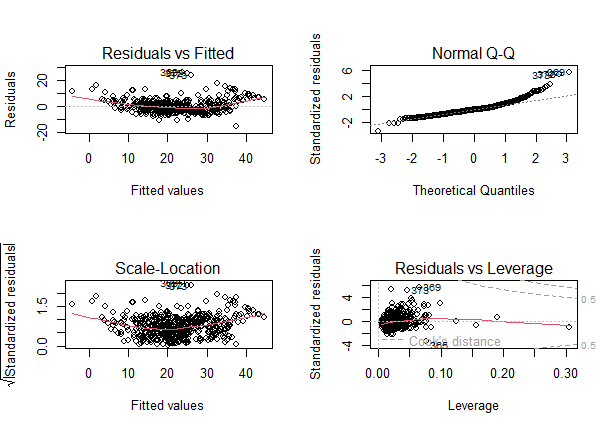


“model” için grafik yorumlayalım.

par(mfrow = c(2,2)) # cikti ekranımızı 4'e boler (2x2'lik matris)

plot(model)

kod parçacığı sonucunda çıktı ekranı 2’ye 2’lik matris şeklinde toplam 4’e bölünür.



Artıklar ve uygun değerler tablosunda, kırmızı çizginin sıfır noktasına yakın ve paralel düz bir şekilde olmaması lineer olmadığını gösteriyor, noktaların bu çizgi üzerinde dağılması beklenir fakat burada taşan noktalar gözlendiği için sabit varyans yoktur.

Normal Q-Q tablosunda hataların normal dağılıp dağılmadığına bakıyoruz. Belli bir kısma kadar güzel gelmiş fakat sona doğru bozulma meydana geldiği için hatalar normal dağılmamaktadır.

Ölçekli konum tablosunda çizgimiz düz fakat noktalar rastgele dağılmıyor.

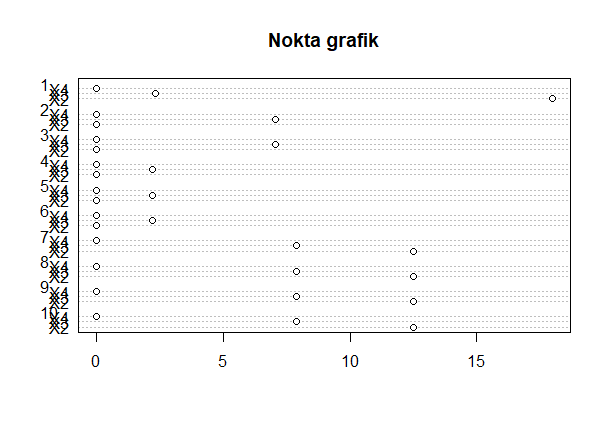
Artıklar ve kaldıraç tablosunda etkili gözlem olup olmadığına bakabiliriz. Kırmızı çizgimiz kesikli çizgiye yakın ve noktalar etrafında rastgele dağılıyor fakat taşan değerler olduğu için etkili gözlem yoktur diyebiliriz.

Nokta grafiği için,

par(mfrow = c(1,1))

dotchart(t(veri[0:10,2:4]),main="Nokta grafik")

kod parçacığında ise ilk önce çıktı ekranımızı eski haline getirdik ardından nokta grafiğini aşağıdaki gibi elde ettik.



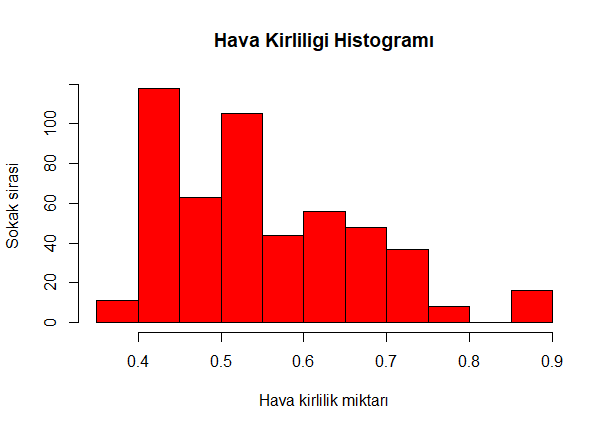
Histogram grafiğini çıkarttırıp bilgisayarımıza kaydedebilmek için,

png("C:/Users/Kadir Altıparmak/Desktop/Hava Kirliligi Histogramı.png")

hist(veri$X5, col="red",main="Hava Kirliligi Histogramı", xlab = "Hava kirlilik miktarı", ylab = "Sokak sirasi")

dev.off()

kod parçacığı sonucunda;



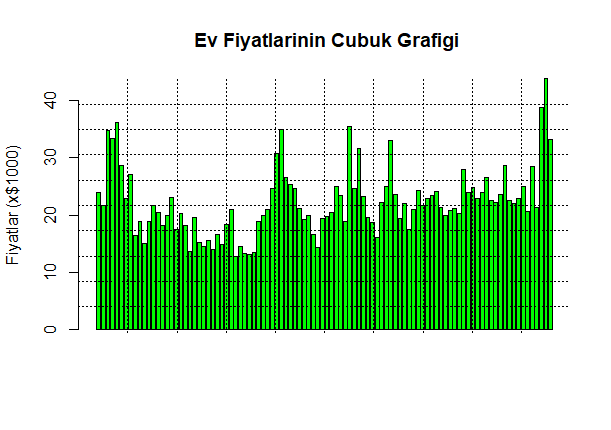
kod parçacığı sonucunda havanın en çok hangi bölgede kirli olduğu görülmektedir.

Çubuk grafiği için,

barplot(Y[0:100],col = "green", main=" Ev Fiyatlarinin Cubuk Grafigi", ylab= "Fiyatlar (x$1000)")

grid(10,10,col = "black")

kod parçacığında ilk 100 sokaktaki ev fiyatlarının çubuk grafiği verilmiştir ve grafiğin okunabilirliğini çizgilerle destekliyoruz. Çıktı olarak da,



Şeklinde sonuç elde edilir.

Kutu grafiği için,

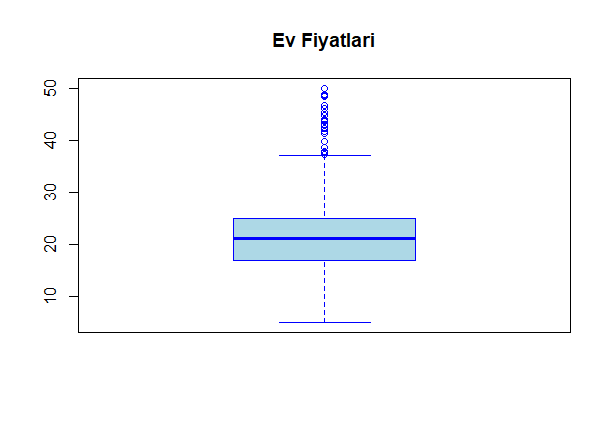
boxplot(veri$Y,

main="Ev Fiyatlari",

col="lightblue",

border = "blue")

kod parçacığı sonucunda,



Kutumuzun sınırlarına çeyreklik değerleri denir, bunları da quantile() fonksiyonuyla öğrenebiliriz.

quantile(veri$Y)

kod parçacığı sonucunda,



Birinci çeyreklik 17.025 , medyan 21.2 ve üçüncü çeyreklik değerimiz 25’dir.

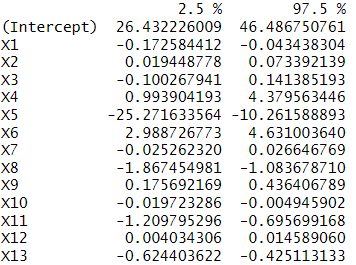
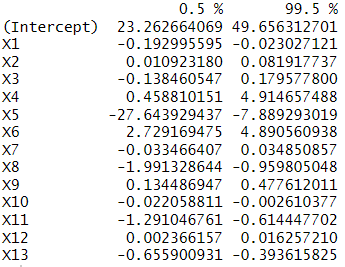
Modelimizin parametreleri için güven aralıklarını bulmak için confint() fonksiyonu kullanılır.

confint(model)

kod parçacığı %95 olasılıkla bulur. Kendimiz aralık belirtmek istiyorsak;

confint(model,level = .99)

kod parçacığı kullanılır, çıktı olarak da,

güven aralıkları elde edilir.

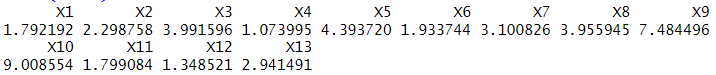
Bağımsız değişkenlerin arasında çoklu doğrusal bağlantı problemine yol açabilecek değişkenleri bulabilmek için varyans artış faktörü(vif) hesaplanır. R programında bu özelliği vif() fonksiyonu kullanılır. Öncelikle kütüphane yükleyip çağırmamız gerekiyor.

install.packages(“car”)

library(car)

vif(model)

kod parçacığı sonucunda,



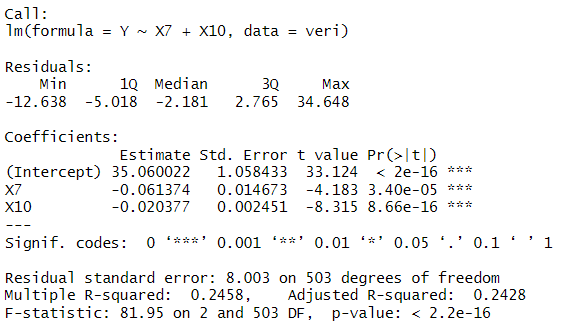
değerlerini elde ederiz. Bu değerlerin 5’den küçük ve 1’e yakın olması beklenir. X9 ve X10 parametreleri çoklu doğrusal bağlantı problemine ol açabilecek değerler almışlar. Bu değerleri çıkartıp yeni bir model kurulabilir.

İki değişken arasında tahmin modeli kuralım;

model2 <- lm(Y ~ X7 + X10 ,data = veri)

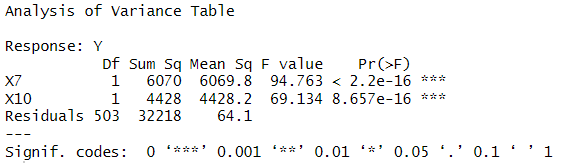
summary(model2)

kod parçacığından çıktı olarak aşağıdaki gibi bir sonuç elde ederiz.



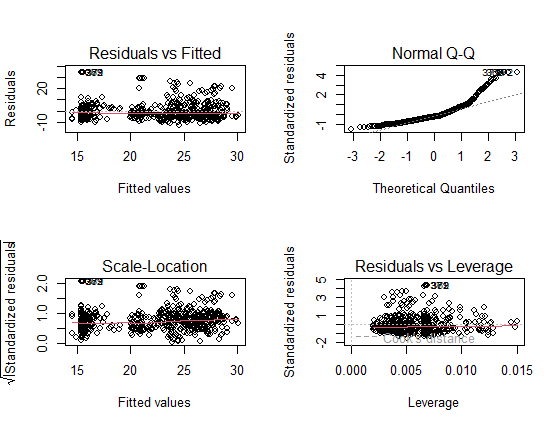
Değişkenlerimiz yani parametrelerimiz anlamlıdır. Düzeltilmiş R-kare %24 olarak çıkmıştır, bu oranı yeterli sayamayız. Hataların standart hatası 8.003’dür, bu değerin 1’e yakın olması beklenir.

“model2” değişkenimizin anova tablosuna da bakalım.



Anova tablosunda göre de paremetrelerimiz anlamlıdır.

“model2” için grafik yorumlayalım.



Artıklar ve uygun değerler tablosunda, kırmızı çizginin sıfır noktasına yakın ve paralel düz bir şekilde olması lineer olduğunu gösteriyor, noktaların bu çizgi üzerinde dağılması beklenir fakat burada taşan noktalar gözlendiği için sabit varyans yoktur.

Normal Q-Q tablosunda hataların normal dağılıp dağılmadığına bakıyoruz. Belli bir kısma kadar güzel gelmiş fakat sona doğru bozulma meydana geldiği için hatalar normal dağılmamaktadır.

Ölçekli konum tablosunda çizgimiz düz fakat noktalar rastgele dağılmıyor.

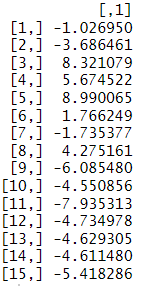
Artıklar ve kaldıraç tablosunda etkili gözlem olup olmadığına bakabiliriz. Kırmızı çizgimiz kesikli çizgiye yakın ve noktalar etrafında rastgele dağılıyor fakat taşan değerler olduğu için etkili gözlem yoktur diyebiliriz.

Artıkları elde etmek için,

artiklar <-model2$residuals

cbind(artiklar[0:15])

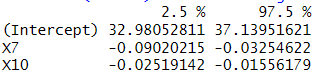
kod parçacığında sadece hataları çağırıp ilk 15 tahmin değerlerini sütun halinde çıkartabiliriz.



Güven aralığı elde edelim.

confint(model2)

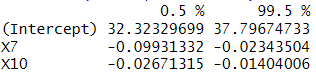
kod parçacığı sonucunda,



çıktısı alınır. Burada %95 olasılıkla güven aralığı elde etmiş oluruz.

confint(model2,level = .99)

kod parçacığı sonucunda,

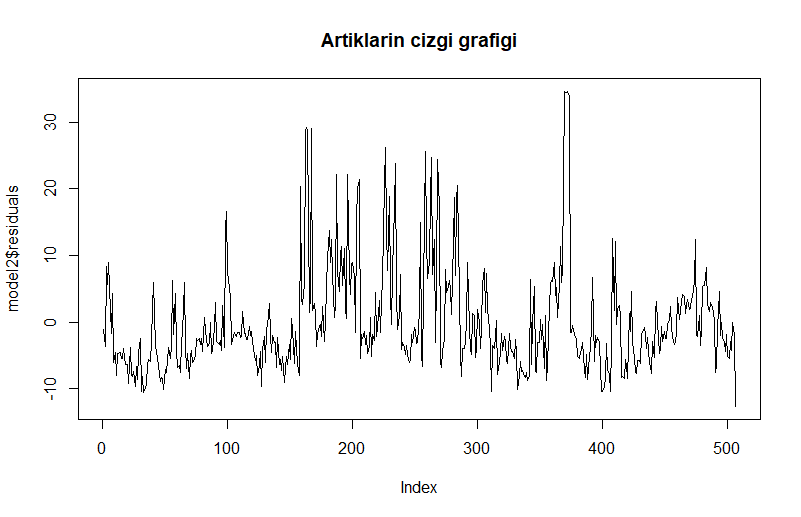


Çıktısı ise %99 olasılıkla güven aralığı elde edilir.

“model2” için artıkların çizgi grafiğin elde edelim.

plot(model2$residuals,type = "l",main = "Artiklarin cizgi grafigi")

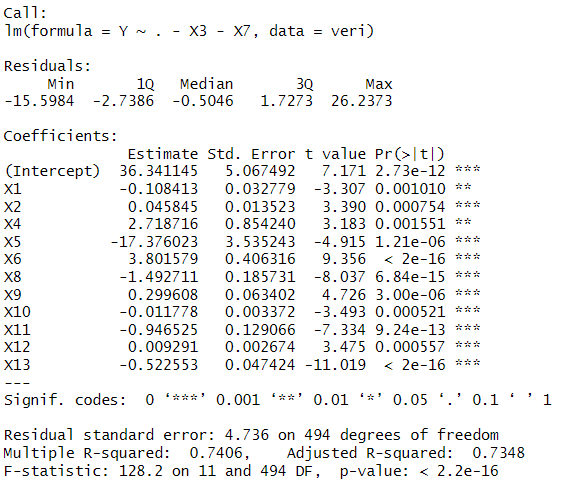
kod parçacığı sonucunda,



grafiği elde edilir.

Bütün bağımsız değişkenleri kullandığımız “model” tahmin modelimizden X3 ve X7 anlamsız parametrelerini çıkartıyoruz.

model3 <- lm(Y ~ . - X3 - X7 ,data = veri)



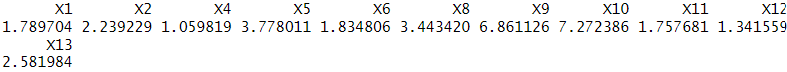
Değişkenlerimiz yani parametrelerimiz anlamlıdır. Düzeltilmiş R-kare %73 olarak çıkmıştır, bu oranı yeterli sayabiliriz. Hataların standart hatası 4.736’dır, bu değerin 1’e yakın olması beklenir.

Çoklu doğrusal bağlantı problemine yol açabilecek değişkenleri bulabilmek için “car” kütüphanesini çağırıp vif() fonksiyonunu kullanalım.

library(car)

vif(model3)

kod parçacığı sonucunda,



değerlerini elde ederiz. X9 ve X10 parametreleri çoklu doğrusal bağlantı problemine ol açabilecek değerler almışlar. Bu değerleri çıkartıp yeni bir model kurulabilir.

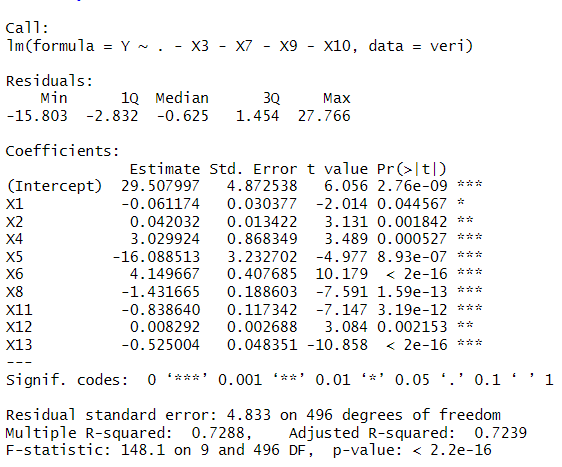
Bu modelimizden çoklu doğrusal bağlantı problemine yol açabilecek değerlere sahip parametrelerimizi de çıkartalım.

model4 <- lm(Y ~ . - X3 -X7 - X9 - X10, data = veri)

summary(model4)

vif(model4)

kod parçacığı sonucunda;





çıktısı elde edilir. Burada görüyoruz ki F istatistik değerimiz artmıştır, modelimiz anlamlıdır ve R2‘miz %72 anlamlıdır.

1. **Sonuç**

Boston’daki Ev Fiyatları veri setinde, ev fiyatlarını tahmin etmek için evlerin bölgesi, yapım yılı ve diğer özelliklerini kullandık.

Ev fiyatları bağımlı değişkeni normal dağılıma uymamaktadır. Regresyon analizinde bağımlı değişkenin normal dağılım göstermesine gerek duymaz. Ancak, bağımlı değişkenin normal dağılım göstermesi, regresyon modelinin doğruluk ve güvenilirliğini artırır ve tahminlerin daha kolay yorumlanmasını sağlar. Bağımlı değişken normal dağılım göstermeyebilir, ancak bu durumda, regresyon analizi sonuçlarının daha az güvenilir olduğu ve daha zor yorumlandığı unutulmamalıdır.

İlk başta oluşturduğumuz modelimizde bütün özellikler ile ev fiyatlarının denklemini çıkarttık. Bu denklemde ticari olmayan arazi sayısı ile 1900’den önce yapılan evlerin değerleri normal dağılıma uymadıkları için çıkartıp modeli daha anlamlı kılabiliriz.

Oda sayısı ile ev fiyatları arasındaki ilişkiyi kontrol etmek için korelasyon testi uyguladık. Bu testin sonucunda iki değişken arasında pozitif yönlü korelasyon olduğunu gözlemledik. Yani evlerdeki oda sayısı arttıkça ev fiyatları da artmaktadır.

Ortalama farkın anlamlılığı için anova testi uyguladık. Bu test sonucunda hava kirliliği miktarı ve otoyola uzanan yol sayısı değişkenlerimiz hariç diğer değişkenlerimiz arasında anlamlı bir ilişki olduğunu gözlemledik. Evlerin yapım yılı ve bölgesi gibi değişkenler, ev fiyatları üzerinde anlamlı bir etkiye sahip olabilir. Yani ev fiyatlarını tahmin etmek için bu değişkenleri kullanmamızın yararlı olacağını söyleyebiliriz.

Kurduğumuz modelde artıklar hakkında şöyle bir yorum yapabiliriz. Artıklarımız eğer normal dağılıyorsa modelin doğruluğundan endişe etmemize gerek yok fakat bizim modelimizde artıklar normal dağılmıyor. Diğer bir gözlem şekli ise artıkların standart hatasına bakmak olur. Modelimizde çıkan değer de 1’e pek yakın olmadığı için modelimizi anlamlı sayamayız.

Ev fiyatlarının sokaklara göre dağılım grafiğinin üzerine, düşük gelirli hanehalkı oranının sokaklara göre dağılım grafiğini aynı tabloda elde ettik. Bu grafik sonucunda ev fiyatlarının sokaklara göre dağılımı ile düşük gelirli hanehalkı oranının sokaklara göre dağılımı arasında pozitif yönlü bir ilişki görülmektedir. Yani düşük gelirli hanehalkı oranı düşüyorsa ev fiyatları da düşüyor.

Evlerin bölgesi, yapım yılı ve diğer değişkenler arasındaki korelasyonun gücünü ölçmek için VIF değerlerini inceledik. Bu değerler sonucunda vergi miktarı ve otoyola uzanan yol sayısı değişkenleri 1 ile 5 arasında değerler almadıkları için modelimizden çıkartıp modelimizi de değişkenlerinin arasındaki korelasyon gücünü kuvvetlendirdik.

En son olarak toparlamamız gerekirse Boston şehrindeki toplamda 506 sokak üzerinden ev fiyatlarını tahmin etmek için evlerin bölgesi, yapım yılı ve diğer özelliklerini kullandık. Bu tahminlerimizi regresyon analizi üzerinde denedik. Sonucunda da modelimizde genel olarak bakındığımızda anlamlılık gözlenmiştir fakat modelimizdeki ev fiyatları normal dağılıma uygun olmadığı için çoklu regresyon tahmin denklemimizi kolaylıkla yorumlayamayız.

1. **Kaynakça**

[**https://www.kaggle.com/datasets/fedesoriano/the-boston-houseprice-data**](https://www.kaggle.com/datasets/fedesoriano/the-boston-houseprice-data)

Rapor hakkında yorum veya düşüncelerinizi belirtmek için [kadirgurkanaltiparmak@gmail.com](mailto:kadirgurkanaltiparmak@gmail.com) e-posta adresinden bana ulaşabilirsiniz.