

Burak Uğur

21822197

İST366 REGRESYON ÇÖZÜMLEMESİ ÖDEVİ

İçindekiler

[Veri Hikayesi 2](#_Toc72949262)

[Tanımlayıcı İstatistikler 2](#_Toc72949263)

[Varsayımlar 2](#_Toc72949264)

[1. Normallik 2](#_Toc72949265)

[2. Doğrusallık 4](#_Toc72949266)

[Model Kurulması 5](#_Toc72949267)

[Artık İncelemesi 6](#_Toc72949268)

[Aykırı Gözlemlerin Ayrılması & Yeni Modelin Kurulması 10](#_Toc72949269)

[Aykırı Gözlem Analizi Yeniden 12](#_Toc72949270)

[Model Yorumları 13](#_Toc72949271)

[Güven Aralıkları 13](#_Toc72949272)

[Değişen Varyans 14](#_Toc72949273)

[Öz İlişki Sorunu 15](#_Toc72949274)

[Çoklu Bağlantı Sorunu 15](#_Toc72949275)

[Uyum Kestirimi 16](#_Toc72949276)

[Ön Kestirim 16](#_Toc72949277)

[En İyi Model 17](#_Toc72949278)

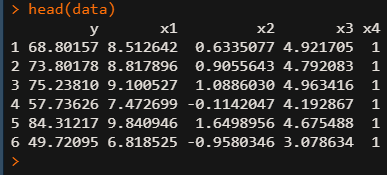
[Ridge Regresyon 21](#_Toc72949279)

[Sonuç 22](#_Toc72949280)

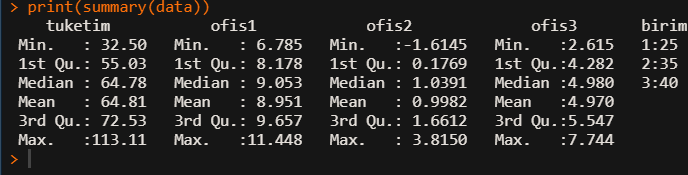
# Veri Hikayesi

Veriler SRSTech firmasının Ankara, San Francisco ve Madrid şehirlerinde bulunan ofislerinin aylık gider/gelirlerinden oluşmaktadır. Birim özniteliği firmanın sahip olduğu birimleri tutmakta olup;

* 1 = Yazılım
* 2 = Pazarlama
* 3 = Halka İlişkiler departmanlarını temsil etmektedir.



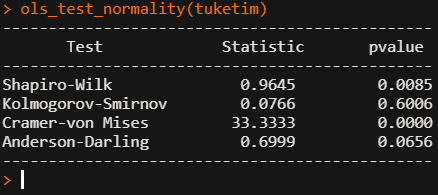
## Tanımlayıcı İstatistikler



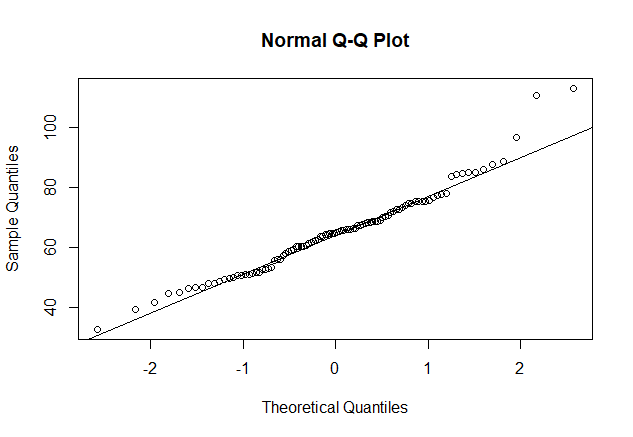
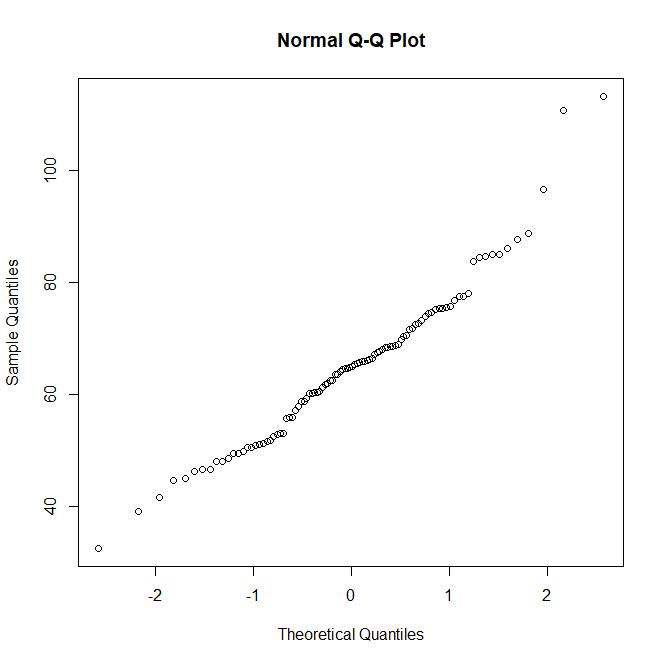
* Ortalama 64.78 Mn TL kazanç sağlanmaktadır.
* Firmanın San Francisco ofisi belirli bir ayda 1.61 Mn TL zarar etmiştir. Ortalamada 1Mn TL kar elde etmektedir.
* Ankara’da bulunan ofis ise 8.9 Mn TL ile ortalamada en yüksek karı elde etmiştir.

## Varsayımlar

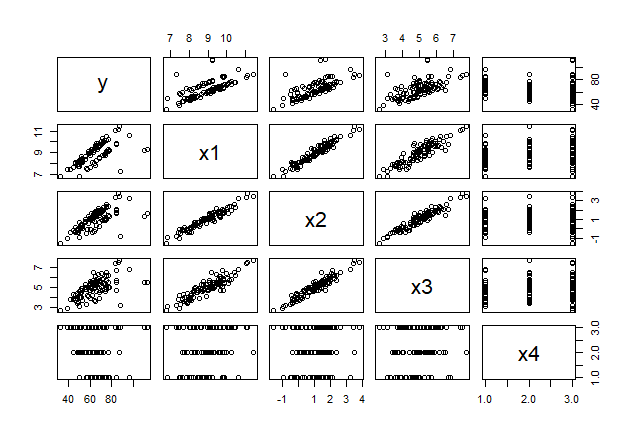
### Normallik



Anderson Darling veya Kolomograv Smirnov testlerine göre tahmin değişkenimiz olan tüketim değişkeni normal dağılmaktadır. Dönüşüm yapmaya gerek yoktur.

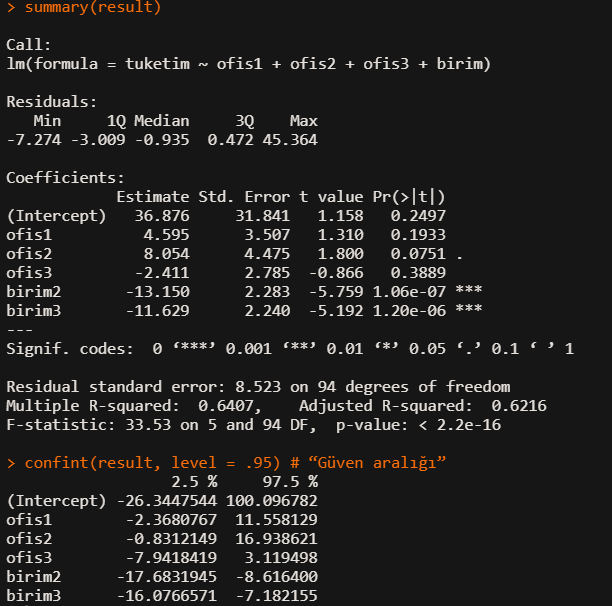


### Doğrusallık



Verilerin doğrusal olduğu görülmektedir. Çoklu doğrusal bağıntı probleminden şüphelenenilir kanıtlarımız var.

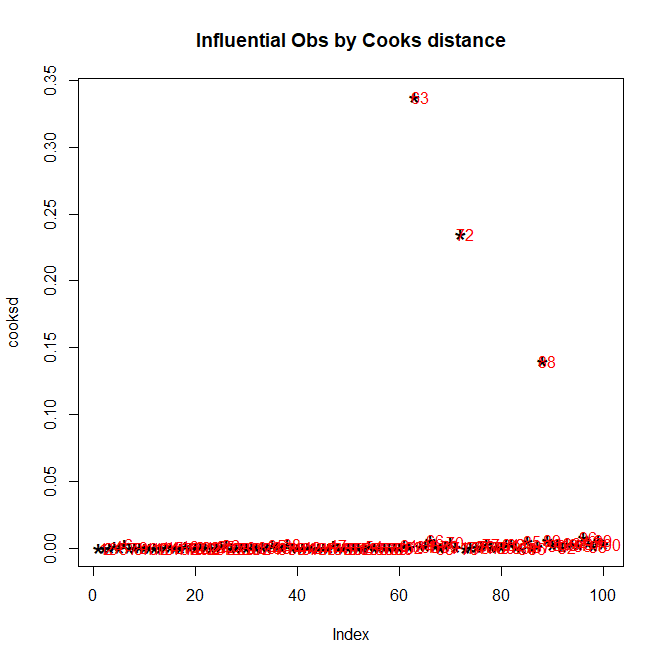
## Model Kurulması



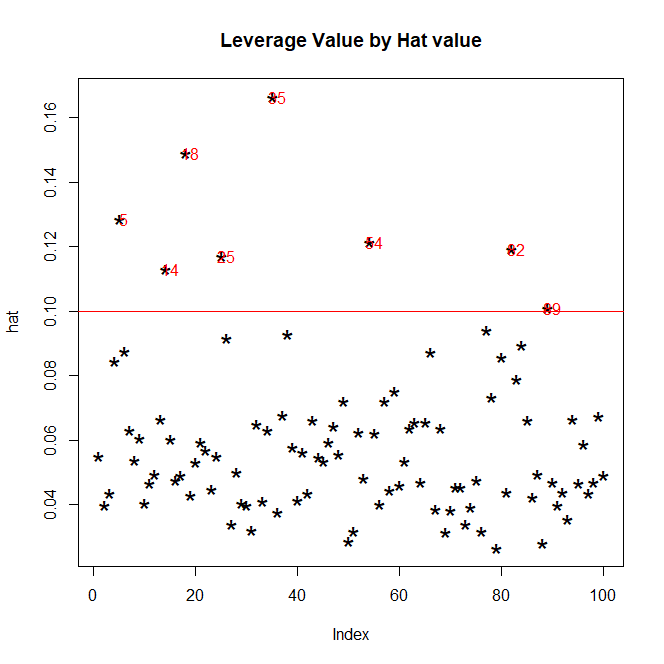
Modelin istatistiksel olarak anlamlı olduğu görülmektedir (p=0.000). Ancak bağımsız değişkenlerin ofis3 ve birim anlamlıdır. Ayrıca bağımlı değişkenin %64’ü bağımsız değişkenlerce açıklanmaktadır.

### Artık İncelemesi

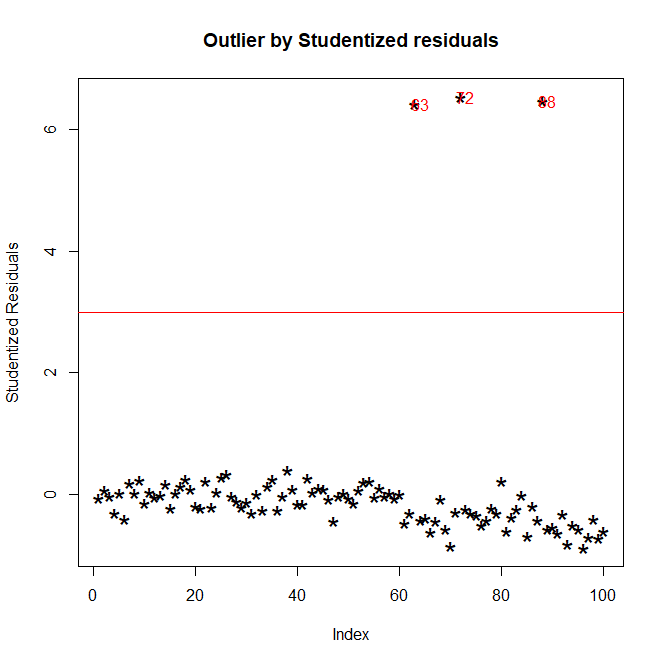




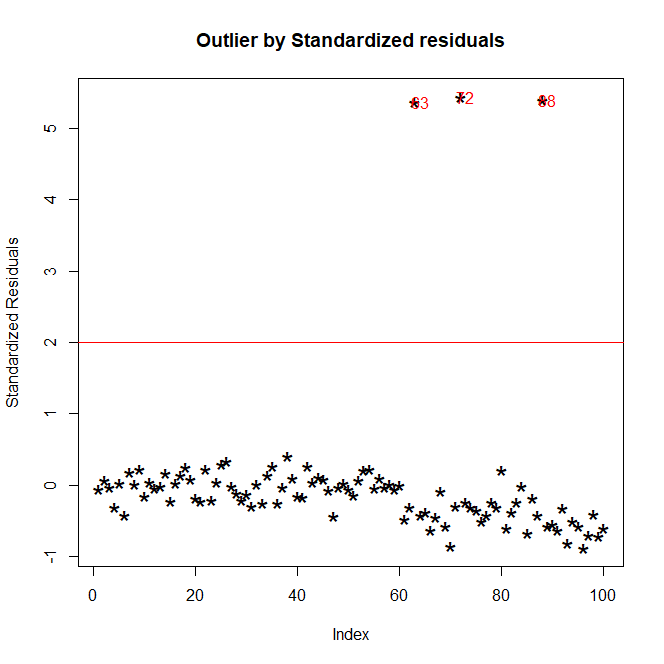
Cook uzaklığında bakıldığında 63. 72. 88’inci gözlemlerin veri setinden ayrılması gerektiğini söyleriz.



* Leverage uzaklığına *which* fonksiyonu baktığımızda *5 18 35 54*’üncü gözlemlerin veri setinden ayrılması gerekmektedir.



* Student artıklara baktığımızda ise 63 72 88’inci gözlemlerin artık olduğu görülmektedir. Veri setinden ayrılmalıdır.

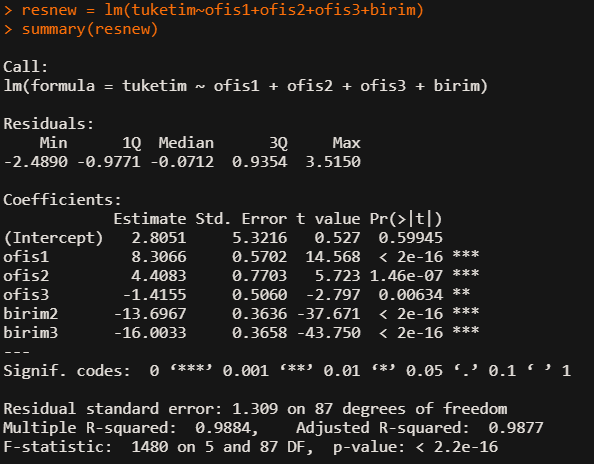


Standartlaşmış artıklara baktığımızda ise 63 72 88’inci gözlemlerin aykırı olduğu görülmektedir. Veri setinden ayrılmalıdır.

### Aykırı Gözlemlerin Ayrılması & Yeni Modelin Kurulması

datanew ­<- data[-c(63,72,88,5,18,54,35),]

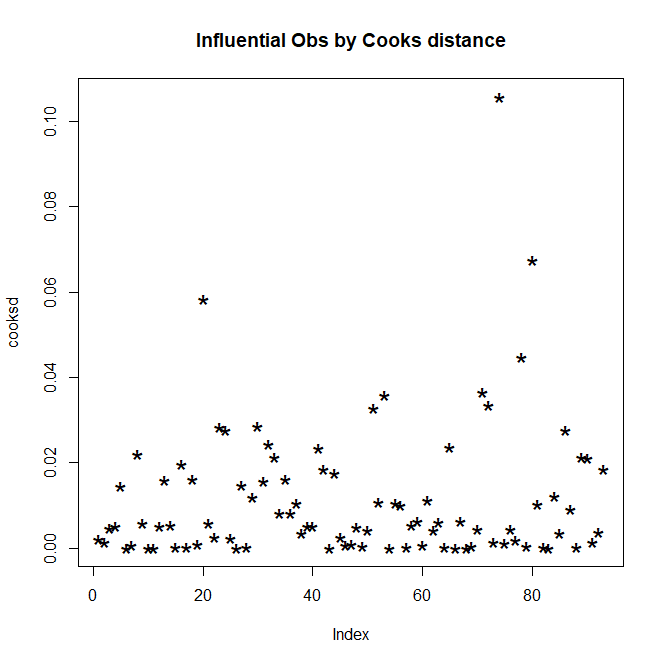
str(datanew)



Yeni Model

Y = 8.3\*(ofis1) + 4.4\*(ofis2) - 1.41\*(ofis3)-13.69\*(birim2) -16\*(birim3) + 2.80

## Aykırı Gözlem Analizi Yeniden

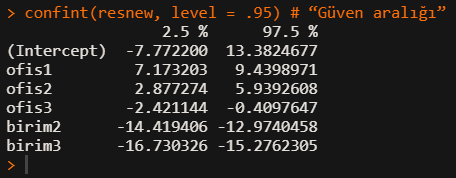


* Yeni oluşturulan veri setinde daha fazla ayrılacak gözlem bulunmamaktadır.

### Model Yorumları

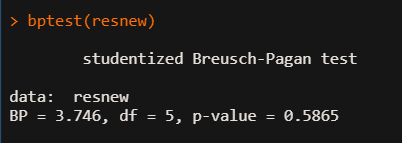
* P-Value < P(0.05) olduğu için model anlamlıdır.
* Bağımsız değişkenle bağımlı değişkeni %98 oranında açıklamaktadır.
* b0 = Satışlar gün geçtikçe negatif etkilenmektedir.
* B1 = Ankara’daki ofis satışları hariç diğer ofisler sabit kaldığında satışlar ortalama 8.3 MnTL arttırmaktadır.
* B2 = S.F Hariç diğer ofisler sabit kaldığında satışlar ortalamada 4.4 MnTL artmaktadır.
* B3 = Madrid ofisi hariç diğer ofisler sabit kaldığında satışlar 1.41 MnTL azalmaktadır.
* B4 = Yazılım birimi sabit kalmak koşulu ile Pazarlama birimi ile halka ilişkiler birimine göre firmanın satışlarına etkisinde fark vardır.
* B5 = Yazılım birimi sabit kalmak koşulu ile halka ilişkiler birimi değişkenler sabitken firma satışlarına etkisinde fark vardır.

### Güven Aralıkları



* P(-7.7 ≤ β0 ≤13.3) = 0.95 => 0 ‘ı içeriyor.
* P(7.1 ≤ β1 ≤9.4) = 0.95 => 0 ‘ı içermez. Ankara’daki ofisin satışlarda etkili olduğu %95 güvenilirlikle söylenebilir.
* P(2.8 ≤ β2 ≤5.93) = 0.95 => 0 ‘ı içermez. San Fransisco ofisin satışlarda etkili olduğu %95 güvenilirlikle söylenebilir.
* P(-2.42 ≤ β3 ≤-0.4) = 0.95 => 0 ‘ı içermez. Madrid ofisin satışlarda etkili olduğu %95 güvenilirlikle söylenebilir.
* P(-14.417 ≤ β4 ≤-12.97) = 0.95 => 0 ‘ı içermez. Pazarlama Departmanının satışlarda etkili olduğu %95 güvenilirlikle söylenebilir.
* P(-16.7 ≤ β5 ≤-15.2) = 0.95 => 0 ‘ı içermez. İnsan Kaynakları Departmanının satışlarda etkili olduğu %95 güvenilirlikle söylenebilir.
* ofis1 finansı sabit tutulduğunda, 1 ay daha geçtiğinde satışların -7.1 ile 9.4 Mn TL arasında olacağı %95 güven düzeyinde söylenebilir
* ofis2 finansı sabit tutulduğunda, 1 ay daha geçtiğinde satışların 2.8 ile 5.93 Mn TL arasında olacağı %95 güven düzeyinde söylenebilir
* ofis3 finansı sabit tutulduğunda, 1 ay daha geçtiğinde satışların -2.42 ile -0.4 Mn TL arasında olacağı %95 güven düzeyinde söylenebilir

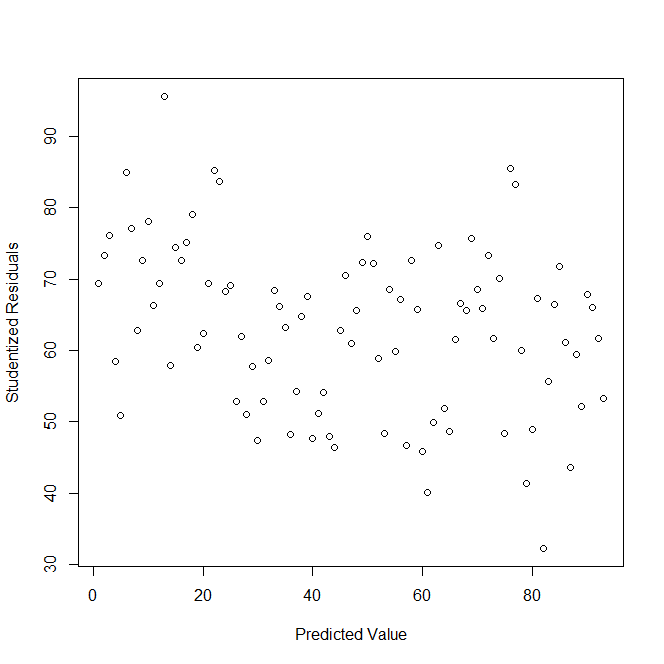
### Değişen Varyans



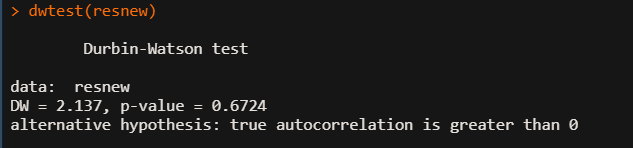
H0: Varyanslar homojendir.

H1 : Varyanslar homojen değildir.

P-value >a(0.05) olduğu için H0 Reddedilemez. Değişken varyanslılık sorunu yoktur.



### Öz İlişki Sorunu



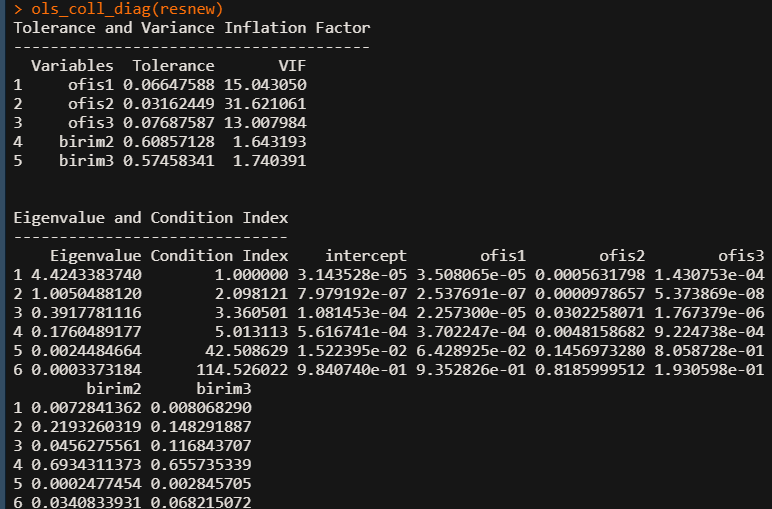
Durbin Watson test değeri d=2.137 bulunmuştur. DW >2 olduğu için negatif yönlü ilişkiden söz edilebilir.

H0: Öz ilişki sorunu yoktur

H1: - yönlü öz ilişki vardır

* P-value > 0.05 olduğu için H0 Reddedilemez. Öz ilişki sorunu yoktur.

### Çoklu Bağlantı Sorunu



VIF değerlerine baktığımızda VIF değerleri büyük olmasından dolayı ofis1, ofis2 ve ofis3 özniteliklerinin çoklu bağlantıdan etkilendiği görülmektedir yani çoklu bağlantıdan şüphelenebiliriz.

### Uyum Kestirimi

Y’ = 8.3\*(8.512642) + 4.4\*(0.63350773)- 1.41\*(4.921705)-13.69\*0 -16\*0 + 2.80



Y’ = 69.34

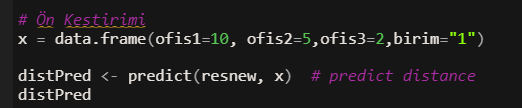
Y Gerçek = 68.80

* SRSTech firmasının yazılım departmanındaki harcamalar Ankara’daki ofisinde 8.5Mn TL S.F’de bulunan ofisinde 0.63 Mn TL Madrid’deki ofisinde de 4.92 Mn TL harcama olduğunda, firmanın toplam satış miktarının 68.68 ile 69.99 arasında olacağı %95 güven ile söylenebilir

### Ön Kestirim

Y’ = 8.3\*(10) + 4.4\*(5)- 1.41\*(2)-13.69\*0 -16\*0 + 2.80

Y’ = 105.08

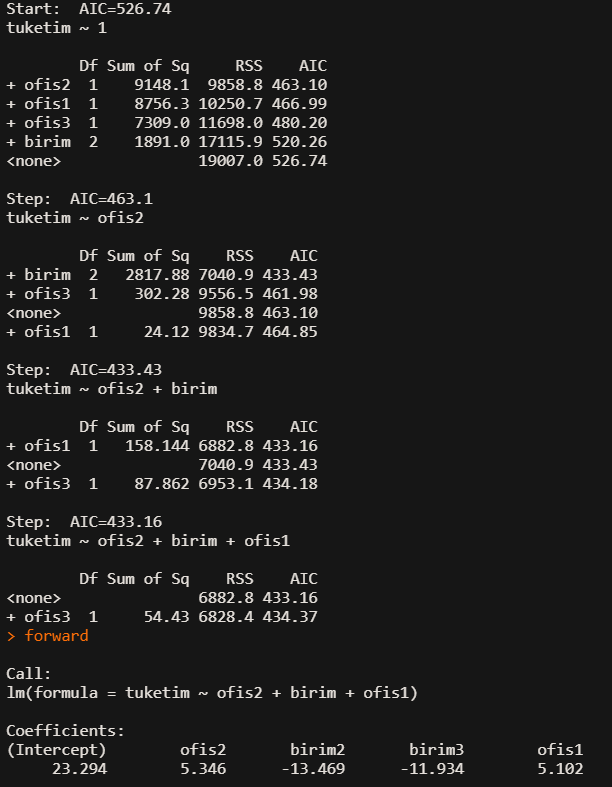


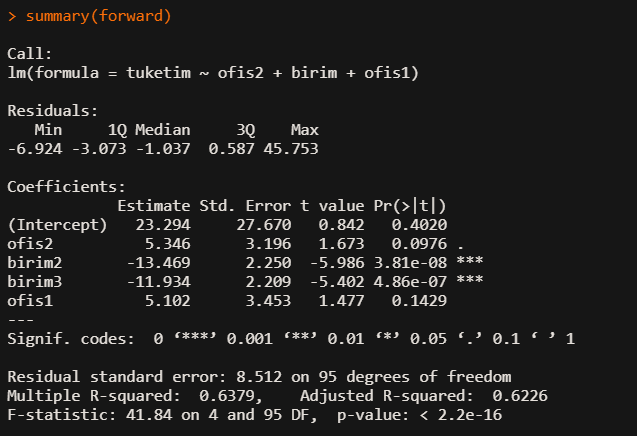


* SRSTech firmasının yazılım departmanındaki harcamalar Ankara’daki ofisinde 10 S.F’daki ofisinde 5 Madrid’deki ofisinde 2 ve olduğunda, firmanın toplam satış miktarının 97.29693 ile 112.8652 arasında olacağı %95 güven ile söylenebilir

### En İyi Model

#### Forward





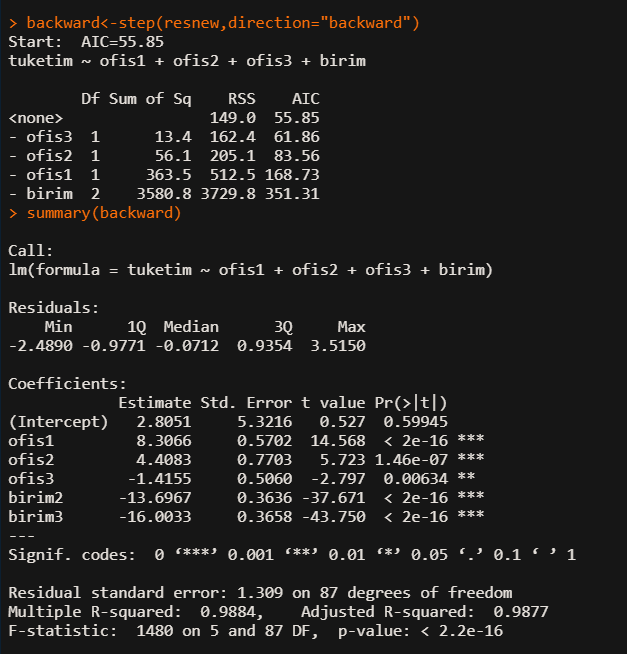
En iyi model

5.10\*(ofis1) + 5.34\*(ofis2)-13.46(birim2)-11.93 (birim3) + 23.294

p değerinden görüldüğü üzere modeldeki bağımsız değişkenlerden ofis1 ,ofis2 birim2 ve birim3 öznitelikleri ile %95 güven düzeyinde anlamlıdır (p=0.000).

Bağımlı Değişken Bağımsız değişkenlerce %63 düzeyinde açıklanmaktadır.

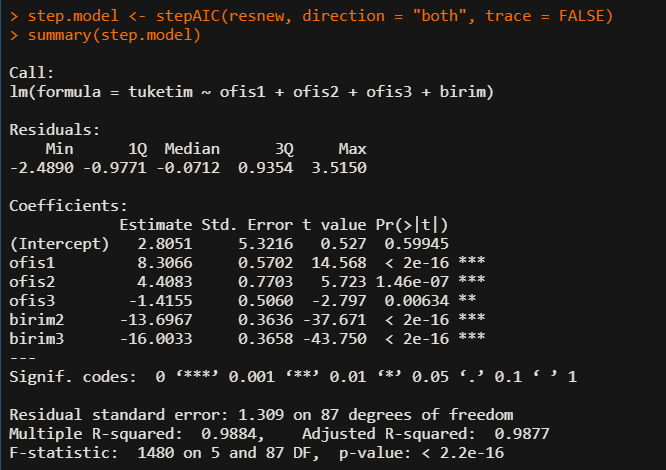
#### Backward



Aynı bağımsız değişkenler ile model kurulduğundan ileri ve geriye doğru seçim yöntemleri farklı sonuçları vermiştir. Backward Elim. Yönteminde tüm özniteliklerin kullanıldığı model en iyi model olarak elde edilmiştir. Bu yöntemden elde edilen modelin anlamlı olduğu söylenebilir. ( p=0.000)

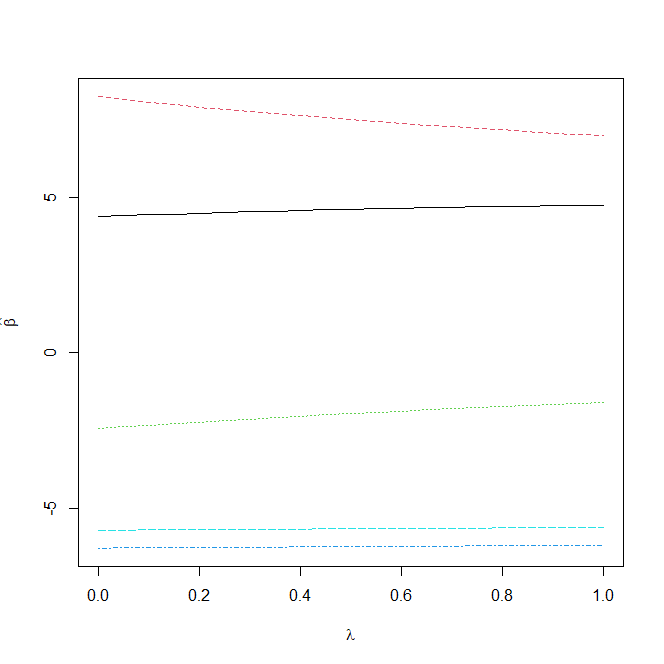
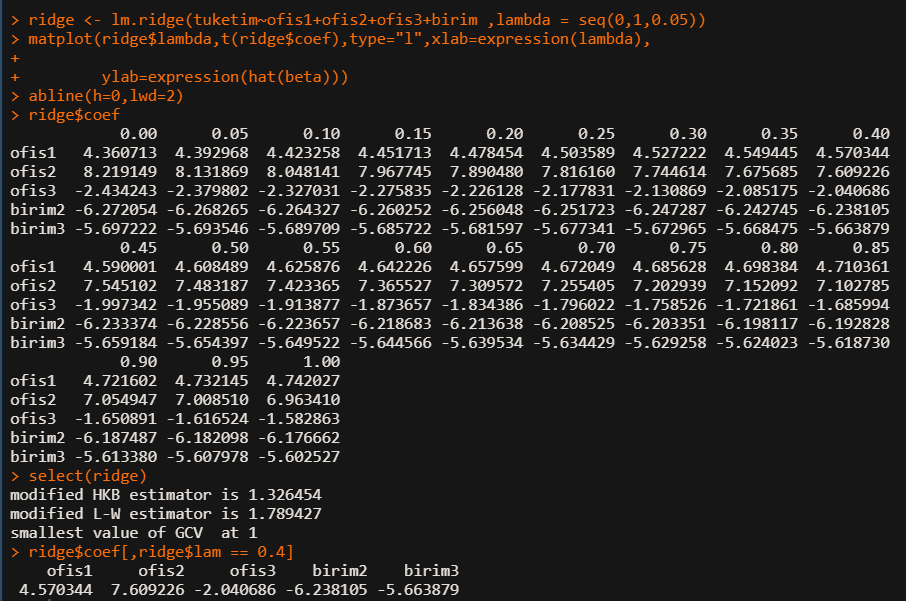
Bağımlı değişkenler SRSTech firmasının satışlarını %98 oranında açıklamaktadır.

#### Adımsal Seçim



* Adımsal seçim yönteminde de tüm değişkenlerin bulunduğu model anlamlıdır.

### Ridge Regresyon



Lambda parametresine göre katsayılar değişir. Hızlı azalış ve artışların bittiği yerde bir lambda parametresi belirlenerek katsayı kestirimleri elde edilir ve model kurulur ancak modelimizde hızlı artış azalış bulunmamaktadır. Tüm değişkenler anlamlıdır.

# Sonuç

En İyi Model:

8.307\*(ofis1) + 4.408\*(ofis2) -1.45\*(ofis3)-13.697\*(birim2)-16\*(birim3)

