



HACETTEPE
ÜNİVERSİTESİ

REGRESYON ÇÖZÜMLEMESİ

Prof. Dr. DURU KARASOY

Arş. Gör. Dr. Hatice IŞIK

YİĞİT SAÇIK

21936277

İçindekiler Tablosu

VERİ SETİ	2
Tanımlayıcı İstatistikler :	2
Normallik ve Doğrusallık Varsayımlarının İncelenmesi :	3
Çoklu Regresyon Modeli ve Artık İncelemesi :	7
Çoklu Regresyon Model Denklemi :	8
Regresyon Katsayıları :	17
Belirtme Katsayısı :	18
Güven Aralıkları :	18
Değişen Varyanslılık Sorunu :	19
Öz İlişki Sorunu :	20
Çoklu Bağlantı Sorunu :	21
Uyum Kestirimi :	22
Ön Kestirim :	23
Değişken Seçimi :	24
Ridge Regresyon Modeli :	27
Kaynakça:	29

VERİ SETİ

Veri Setimizi Tanıyalım :

Veri setimizde bir adet bağımlı değişken (y), üç adet nicel değişken (x1, x2, x3) ve bir adet nitel değişken (x4) bulunmaktadır. Veri seti bir otele ait odaların sahip olduğu özelliklerden oluşmaktadır. Bağımlı y değişkeni odanın bir gecelik konaklama ücretini, x1 bağımsız değişkeni odanın denize olan uzaklığını, x2 bağımsız değişkeni odanın sıcaklığını, x3 bağımsız değişkeni odanın m^2 cinsinden büyüklüğünü ve x4 nitel değişkenimiz odanın kaç kişilik olduğu bilgisini içermektedir

Değişken Adı	Açıklama	Düzeyler
Y	Konaklama Ücreti	
X1	Denize Olan Uzaklık	
X2	Sıcaklık	
X3	Büyüklük(m^2)	
X4	Oda Tipi	1:Tek Kişilik Oda 2:İki Kişilik Oda 3:Aile

Veri Setini Yükleme :

```
#Veriler import dataset ile yuklenmistir.  
#Tanımlayıcı istatistikler  
names(regresyon_veriler)  
names(regresyon_veriler)<-c("y", "x1", "x2", "x3", "x4")  
attach(regresyon_veriler)  
nitel<-as.factor(x4)
```

Veri girişini tanımladıktan sonra “as.factor” kodu yardımıyla nitel değişken olan y değişkenimiz R tarafından otomatik oluşturulmuştur. Kılavuz değişkeni de R kendisi belirler.

Tanımlayıcı İstatistikler :

```
> summary(regresyon_veriler)  
      y      x1      x2  
Min.   : 9.249   Min.   :2.303   Min.   : -1.7294  
1st Qu.:26.238   1st Qu.:4.201   1st Qu.: 0.3404  
Median :32.438   Median :5.044   Median : 1.0288  
Mean   :33.971   Mean   :4.986   Mean   : 0.9928  
3rd Qu.:39.604   3rd Qu.:5.733   3rd Qu.: 1.7341  
Max.   :84.626   Max.   :7.350   Max.   : 3.3259  
      x3      x4  
Min.   :0.2923   Min.   :1.000  
1st Qu.:2.2357   1st Qu.:1.000  
Median :3.0602   Median :2.000  
Mean   :2.9418   Mean   :2.038  
3rd Qu.:3.6030   3rd Qu.:3.000  
Max.   :5.6461   Max.   :3.000  
> library(moments)  
> skewness(regresyon_veriler)  
      y      x1      x2      x3  
1.58887207 -0.12896850 -0.21361302 -0.21219693  
      x4  
-0.06997085  
> kurtosis(regresyon_veriler)  
      y      x1      x2      x3      x4  
7.349099 2.599315 2.545785 2.781742 1.536073
```

Ücret için (y):

Bağımlı değişkenimiz olan Ücret'e ne ilişkin tanımlı istatistiklere baktığımız zaman en düşük kalitenin 9.24 en yüksek şarap kalitesinin ise 84.62 olduğunu söyleyebiliriz.

Çeyrekler açıklığını 3rd Qu-1st Qu işlemi sonucunda bulabiliriz. Bu açıklık bize en ucuz ve en pahalı otel odaları arasındaki uzaklığı gösterir.

- Çeyrekler açıklığı=39.60–26.23= 13.37

Ortalama ücret medyandan büyük olduğu için veri kümesi sağa çarpıktır diyebiliriz.

Uzaklık için(x1):

Otel odalarının denize uzaklığı en az 2.30 en fazla ise 7.35 olduğunu söyleyebiliriz.

Otel odalarının denize olan ortalama uzaklığı 4.98'dir. Medyanın ortalamadan büyük olması bizlere Otel odalarının denize olan uzaklığına ilişkin verilerin sola çarpık olduğu yorumunu yapabiliriz.

Sıcaklık için(x2):

Odaların ortalama sıcaklığının 0.99 olduğunu söyleyebiliriz. En sıcak ve en soğuk odalar arasındaki fark ise 5.04'tür. Otel odalarının sıcaklığına ilişkin verilerin sola çarpık olduğu yorumunu yapabiliriz.

Büyüklik için(x3):

Odaların ortalama büyüklüğü $2.94 m^2$, en küçük oda ise $0.29 m^2$ ve en büyük odanın ise $5.64 m^2$ olduğunu söyleyebiliriz. Otel odalarının büyüklüğüne ilişkin verilerin sola çarpık olduğu yorumunu yapabiliriz.

Tip için(x4):

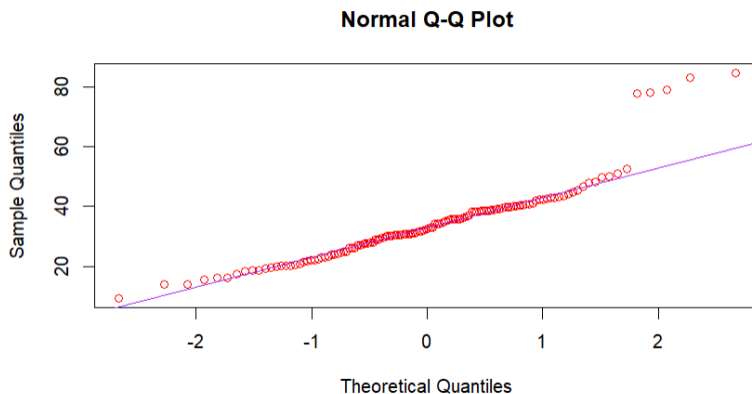
Otelde tek kişilik 40 adet, çift kişilik 45 adet, aile odasından da 45 tane bulunmaktadır.

```
> table(x4)
```

```
x4  
Tek Kişilik Oda Çift Kişilik Oda Aile  
40 45 45
```

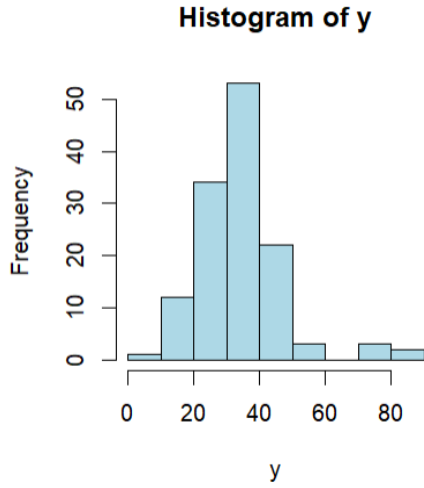
Normallik ve Doğrusallık Varsayımlarının İncelenmesi :

Normallik Varsayımı:



Eğer konaklama ücretimiz 45'lik bir açı ile dağılırsa verilerimizin normal dağıldığını söyleyebiliriz.

Veri setimizdeki bağımlı değişken olan Ücret (y) için çizilen Q-Q plot grafiğinde verilerin çoğunluğu oluşturulan çizgi çevresinde olsa da sapan değerlerimizin olduğunu görüyoruz. Shapiro-wilk testi yaparız.



Histogram grafiğimize baktığımızda verilerimizin sağa çarpık bir dağılım gösterdiğini görüyoruz.

Shapiro-wilk ve Anderson Darling testi:

H_0 : Otel odalarında konaklama ücreti verileri ile normal dağılım arasında fark yoktur
 H_1 : Otel odalarında konaklama ücreti verileri ile normal dağılım arasında fark vardır.

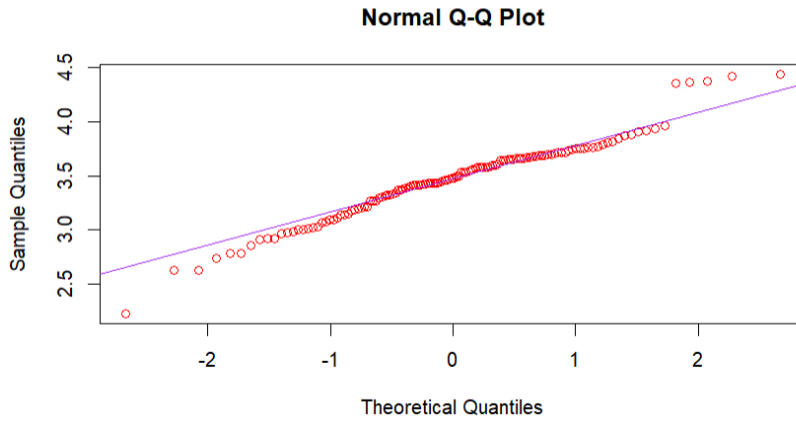
`> shapiro.test(y)`

Shapiro-wilk normality test	Anderson-Darling normality test
data: y	data: y
W = 0.87174, p-value = 3.182e-09	A = 3.0176, p-value = 1.289e-07

p-value < $\alpha=0.05$ olduğundan yokluk hipotezimiz reddedilir. %95 güven düzeyinde otel odasındaki konaklama ücretlerine ilişkin verilerin normal dağılmadığı yorumu yapılmaktadır. Dönüşüm yapılmalıdır.

Dönüşüm:

1. Logaritmik Dönüşüm: Otel odası ücretleri verilerine logaritmik dönüşüm uyguluyoruz. Yeni verilerimizle Q-Q plot grafiği çiziyoruz.



Q-Q plot grafiğimde verilerimin line üzerinde olması dönüşüm yapılmış otel odası ücretlerinin normal dağılıma uymadığını düşündürmektedir, aykırı değerler vardır.

H_0 : Otel odalarında konaklama ücreti(lny) verileri ile normal dağılım arasında fark yoktur.

H_1 : Otel odalarında konaklama ücreti(lny) verileri ile normal dağılım arasında fark vardır.

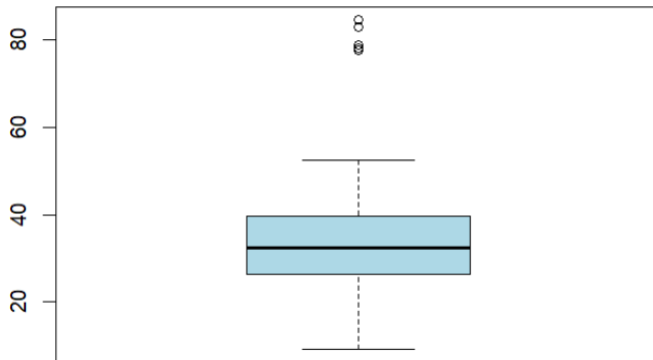
Shapiro-wilk normality test

data: lny
w = 0.9696, p-value = 0.005118

Anderson-Darling normality test

data: lny
A = 1.2394, p-value = 0.003043

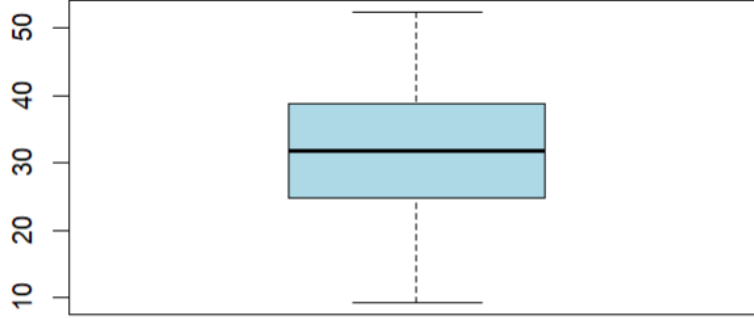
p-value < $\alpha=0.05$ olduğundan yokluk hipotezimiz reddedilmektedir. %95 güven düzeyinde otel odasındaki konaklama ücretlerine ilişkin verilerin normal dağılıma uygun olmadığı yorumu yapılır. Bu durumda aykırı değerlerimizi test etmek için box-plot grafiğine bakıp yeniden normallik testi yapmamız gerekir.



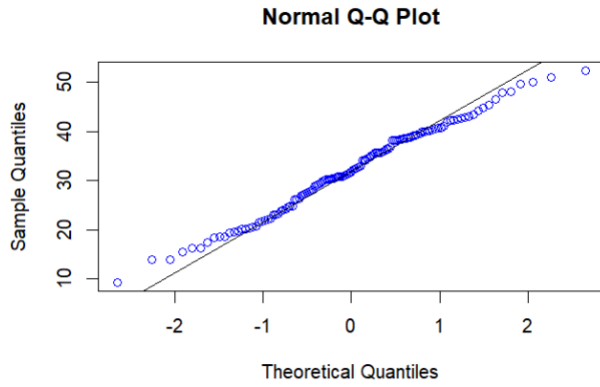
Bağımlı y değişkenimizin kutu grafiğine baktığımızda tam simetrik bir görünüm olmadığını aykırı değerlerin olduğunu görmekteyiz. Aykırı değerlerin olduğu 5 satırı silerek yeniden test yapmamız gerekmektedir.

```
> which(y%in%boxplot(y)$out)
[1] 38 43 67 90 109
```

```
yeni_data<-regresyon_veriler[-c(38,43,67,90,109),]
names(yeni_data)<-c("Ücret", "Uzaklık", "Sıcaklık", "Büyüklik", "Tip")
nitel<-as.factor(x4)
attach(yeni_data)
```



Aykırı değerlerini çıkarttığımız yeni verilerimizin box-plot grafiğine baktığımızda simetrik bir görüntü olduğunu görüyoruz. Verilerimizin normal dağıldığını söyleyebiliriz.



Q-Q Plot grafiğine baktığımızda da verilerimizin 45 derecelik açı etrafında dağıldığını görmekteyiz. Bu grafiğe bakarak da verilerin normal dağıldığını söylemek mümkündür.

H_0 : Otel odalarında konaklama ücreti verileri ile normal dağılım arasında fark yoktur

H_1 : Otel odalarında konaklama ücreti verileri ile normal dağılım arasında fark vardır.

Shapiro-Wilk normality test

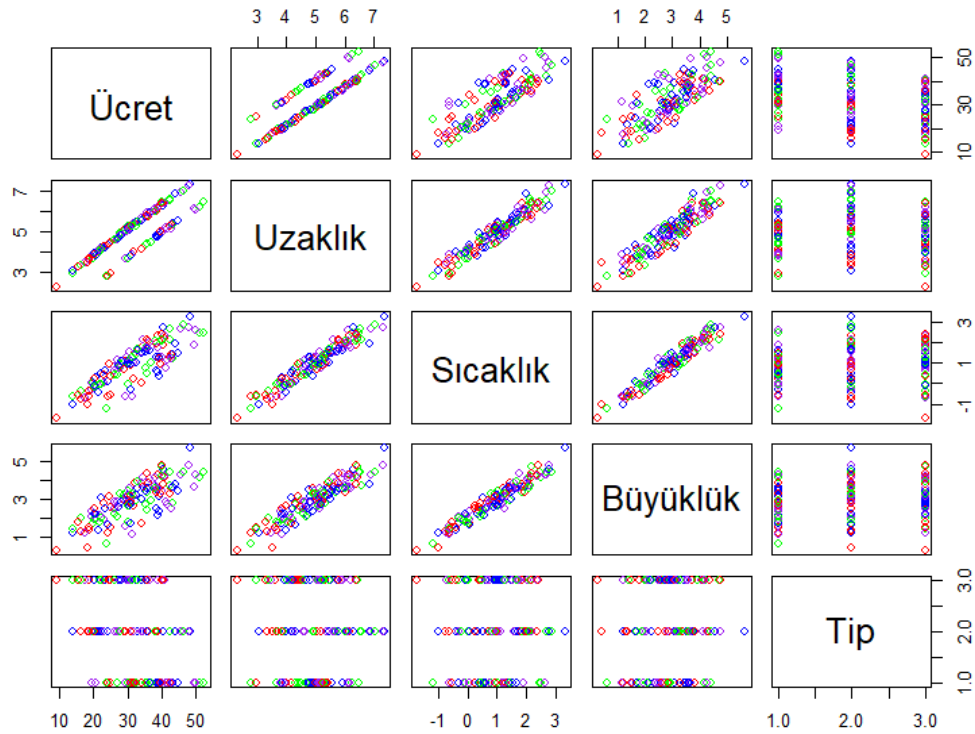
data: Ücret
W = 0.98911, p-value = 0.4275

Anderson-Darling normality test

data: y
A = 0.45236, p-value = 0.2683

Yorum: Shapiro-wilk testine göre $p = 0.4275 > \alpha$ olduğundan aynı şekilde Anderson-Darling testine göre $p=0.2683 > \alpha$ olduğundan H_0 reddedilemez. Bağımlı değişken (y) verilerinin dağılışı normal dağılıma uygundur.

Doğrusallık Varsayımı:



Bağımlı değişkenimiz otel odası ücreti ile bağımsız değişkenlerimiz arasındaki ilişki incelendiğinde doğrusallığın sağlandığını görürüz. Tip değişkenimiz nitel değişken olduğu için doğrusallık incelemesi yapılmaz.

Çoklu Regresyon Modeli

Call:

```
lm(formula = y ~ x1 + x2 + x3 + nitel)
```

Coefficients:

(Intercept)	x1	x2	x3	nitel2	nitel3
9.653	6.241	3.710	-2.104	-8.842	-8.870

> summary(sonuc)

Call:

```
lm(formula = y ~ x1 + x2 + x3 + nitel)
```

Residuals:

Min	1Q	Median	3Q	Max
-9.3625	-0.2229	0.0919	1.4365	2.5848

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	9.6533	3.5538	2.716	0.00759	**
x1	6.2407	0.7587	8.226	2.79e-13	***
x2	3.7102	1.1468	3.235	0.00157	**
x3	-2.1038	0.8337	-2.524	0.01294	*
nitel2	-8.8424	0.5475	-16.149	< 2e-16	***
nitel3	-8.8703	0.5496	-16.138	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.428 on 119 degrees of freedom

Multiple R-squared: 0.9315, Adjusted R-squared: 0.9287

F-statistic: 323.9 on 5 and 119 DF, p-value: < 2.2e-16

Yorum: Belirtme katsayısına ($R^2 = 0.9315 \cong 0.93$) bakarak bağımlı değişkendeki değişimin %93'ünün bağımsız değişken tarafından açıklanabildiğini söyleyebiliriz. Yani konaklama ücretinin %93'ü uzaklık, sıcaklık, büyüklük ve oda tipi değişkenleri tarafından açıklandığı söylenebilmektedir. Açıklanmayan kısım %7'dir. Gözlem sayısının arttırılması ya da anlamlı başka bağımsız değişkenin modele eklenmesi belirtme katsayısını arttıracaktır.

H_0 : Model anlamsızdır.

H_1 : En az bir β_j sıfırdan farklıdır. (Model anlamlıdır.)

Yorum: $p=0.00 < \alpha=0.05$ olduğundan H_0 reddedilir. Kurulan regresyon modelinin anlamlı olduğu %95 güven düzeyinde söylenebilmektedir. En az bir bağımsız değişken, bağımlı değişkeni açıklamaktadır. Nitell yani tek kişilik oda değişkeni kılavuz değişken olduğu için modelde görünmemektedir.

Kestirim Denklemi:

$$\text{Konaklama ücreti} = 9.6533_{3.5538} + 6.2407_{0.7587} \text{uzaklık} + 3.7102_{1.1468} \text{sıcaklık} - 2.1038_{0.8337} \text{büyüklük} - 8.8424_{0.5475} \text{cift kişilik} - 8.8703_{0.5496} \text{aile} \pm 2.428$$

Modelimiz anlamlı çıktığına göre aykırı değer incelenmesine geçilebilir.

Artıkların İncelenmesi :

```
> inf
$std.dev
[1] 2.428128

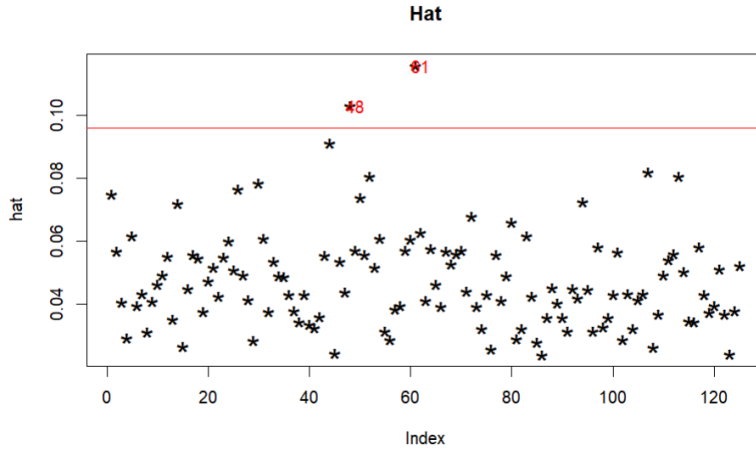
$hat
[1] 0.07491417 0.05684081 0.04083754 0.02946225 0.06181559 0.03969578 0.04336745
[8] 0.03118845 0.04092492 0.04650023 0.04945720 0.05528249 0.03534742 0.07210474
[15] 0.02680747 0.04493265 0.05596494 0.05481121 0.03763607 0.04758166 0.05187871
[22] 0.04262876 0.05493148 0.06022263 0.05101656 0.07663067 0.04924434 0.04157980
[29] 0.02863420 0.07852513 0.06093166 0.03783318 0.05373195 0.04904131 0.04873965
[36] 0.04322924 0.03811042 0.03463029 0.04326236 0.03379466 0.03254425 0.03616214
[43] 0.05545498 0.09120839 0.02451001 0.05355746 0.04408478 0.10301753 0.05731710
[50] 0.07388911 0.05589827 0.08054951 0.05179143 0.06094883 0.03157213 0.02897073
[57] 0.03860197 0.03974404 0.05710808 0.06072659 0.11577264 0.06299441 0.04114400
[64] 0.05782757 0.04637673 0.03935112 0.05698612 0.05293347 0.05621580 0.05707091
[71] 0.04436062 0.06787068 0.03949959 0.03237462 0.04305649 0.02588004 0.05575587
[78] 0.04126947 0.04902265 0.06619343 0.02902872 0.03236231 0.06179389 0.04257625
[85] 0.02801323 0.02399122 0.03590224 0.04523832 0.04044055 0.03573507 0.03154878
[92] 0.04493526 0.04208596 0.07256950 0.04485298 0.03152623 0.05817624 0.03296510
[99] 0.03594779 0.04309261 0.05677020 0.02874737 0.04339586 0.03234415 0.04153831
[106] 0.04340560 0.08207517 0.02649855 0.03682354 0.04927995 0.05435306 0.05612443
[113] 0.08079355 0.05055809 0.03477038 0.03460077 0.05835239 0.04323397 0.03737222
[120] 0.03953832 0.05126804 0.03700746 0.02425023 0.03806209 0.05237042
```

Gözlem Uzaklığı (h_{ii}):

$$h_{ii} > \frac{2(k+1)}{n} = \frac{2p}{n} = \frac{2(5+1)}{125} = 0.096$$

0.096'den büyük olan verilerde uç değer vardır. Gözden kaçırmamak için verilerimizin grafiğini çizdirerek ve R kodu ile kaçınıcı gözlemlerin uç değer olduğunu bulalım.

```
> #Gözlem uzaklığı
> hii<-which(inf$hat>0.096)
> hii
[1] 48 61
```



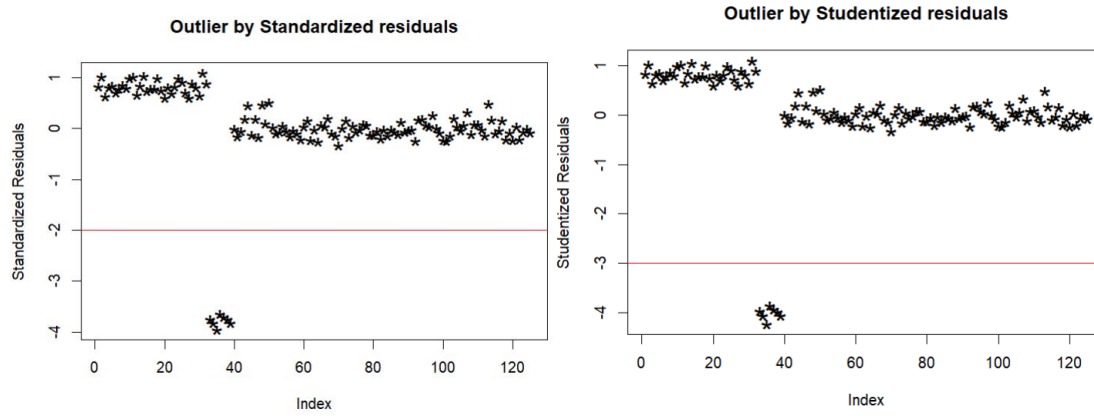
Kodda ve grafikte de görüldüğü gibi 48,61 verilerde **uç değer** vardır.

Aykırı Değer (r_i):

```
$std.res
[1] 0.835700950 1.022782478 0.641808356 0.812447890 0.842335572 0.716620323 0.780049640
[8] 0.858697848 0.801416765 0.995588325 1.018971480 0.669492403 0.847771084 1.044854772
[15] 0.744122028 0.786040632 0.791823184 0.999752457 0.750223310 0.607277671 0.812375632
[22] 0.702641988 0.825953292 0.994134096 0.916258200 0.718835089 0.606756847 0.890463454
[29] 0.817926152 0.650826645 1.098528661 0.892153058 -3.739670314 -3.817439698 -3.953379890
[36] -3.642263797 -3.708190700 -3.745853063 -3.814474408 -0.000395342 -0.149759101 -0.040243391
[43] 0.198288026 0.457402711 -0.116178339 0.196858161 -0.169700328 0.479087542 0.103150398
[50] 0.514953869 0.029692483 -0.095748305 -0.066382285 0.061102125 -0.049953761 -0.144031244
[57] -0.030534147 -0.088173341 -0.214801512 0.039326133 0.163360172 -0.224275255 -0.016047886
[64] -0.249993870 0.069031488 0.039847043 0.203336160 -0.088778162 -0.135906853 -0.330241325
[71] 0.020627168 0.166874492 -0.158260860 0.056935784 -0.054387828 -0.018754171 0.087321870
[78] 0.079885128 -0.116507935 -0.119771195 -0.051440518 -0.195215904 -0.021731970 -0.131917667
[85] -0.021323391 -0.053049082 -0.099897480 0.139058026 -0.056047569 -0.030574287 -0.013479177
[92] -0.234555909 0.178788184 0.196513711 0.079445076 0.038457802 0.259302523 0.010437418
[99] -0.072039709 -0.218659489 -0.234762122 -0.139275305 0.209427175 0.046144596 -0.028460619
[106] 0.066420101 0.330019764 -0.106867448 0.085898987 0.090576232 -0.022758917 -0.136255248
[113] 0.482996155 0.181626274 -0.088635200 -0.032048263 0.163864646 -0.203848491 -0.067964622
[120] -0.226374826 0.037357228 -0.204013064 -0.056689692 0.004069518 -0.073425300
```

```
$stud.res
[1] 0.8346349813 1.0229822565 0.6402150044 0.8112801830 0.8413007280 0.7151477403 0.7787587538
[8] 0.8577438191 0.8002047244 0.9955511872 1.0191368995 0.6679325641 0.8467624432 1.0452610922
[15] 0.7427188518 0.7847709281 0.7905746064 0.9997503599 0.7488374504 0.6056599126 0.8112076268
[22] 0.7011394335 0.8248432922 0.9940848252 0.9156358143 0.7173675786 0.6051388630 0.8896831568
[29] 0.8167813948 0.6492428154 1.0994923749 0.8913826358 -3.9641395029 -4.0579479083 -4.2238709427
[36] -3.8477358116 -3.9263897845 -3.9715686156 -4.0543565034 -0.0003936774 -0.1491425888 -0.0400742174
[43] 0.1974857527 0.4558777179 -0.1156957264 0.1960612082 -0.1690062488 0.4775310711 0.1027206708
[50] 0.5133579306 0.0295675707 -0.0953488249 -0.0661040032 0.0608458065 -0.0497439499 -0.1434372973
[57] -0.0304057003 -0.0878049518 -0.2139385598 0.0391608026 0.1626905803 -0.2233781480 -0.0159803329
[64] -0.2490066556 0.0687422043 0.0396795297 0.2025151879 -0.0884072852 -0.1353451157 -0.3290016231
[71] 0.0205403531 0.1661913053 -0.1576110831 0.0566968252 -0.0541594992 -0.0186752333 0.0869569832
[78] 0.0795509007 -0.1160239909 -0.1192740826 -0.0512244948 -0.1944250730 -0.0216405096 -0.1313718277
[85] -0.0212336489 -0.0528263411 -0.0994810286 0.1384837677 -0.0558123153 -0.0304456718 -0.0134224328
[92] -0.2336223120 0.1780593047 0.1957180409 0.0791126672 0.0382961115 0.2582836963 0.0103934755
[99] -0.0717379472 -0.2177825680 -0.2338277991 -0.1387001846 0.2085838149 0.0459507134 -0.0283408807
[106] 0.0661416626 0.3287806916 -0.1064225845 0.0855399574 0.0901979662 -0.0226631385 -0.1356921241
[113] 0.4814346069 0.1808866017 -0.0882649110 -0.0319134605 0.1631930998 -0.2030256285 -0.0676797679
[120] -0.2254702173 0.0372001521 -0.2031895948 -0.0564517593 0.0040523839 -0.0731177954
```

Standartlaştırılmış hataların (-2,+2) aralığında, Student tipi artıkların ise (-3,+3) aralığında olması istenir. Bu aralığın dışında kalan değerlerin aykırı değer olduğu söylenmektedir.

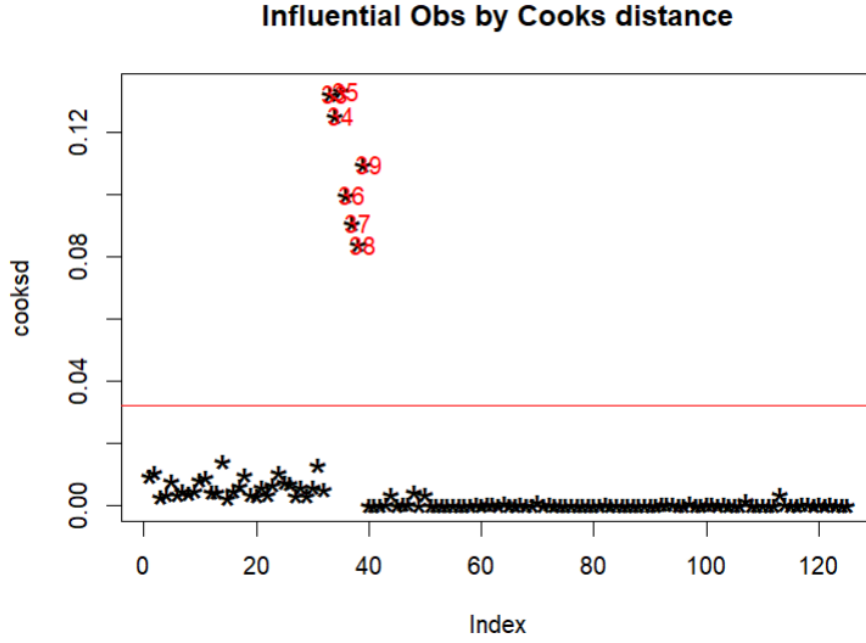


Tablolarda ve grafiklerde gördüğümüz gibi 33,34,35,36,37,38,39. gözlem değeri hem standartlaştırılmış artıklara hem student tipi artıklara göre aykırı değer olmaktadır.

Cook Uzaklığı (Di):

$n > 50$ olduğundan, $D_i > \frac{4}{n}$ formülü kullanılır. Bu durumda $4/125 = 0.032$ olduğundan $D_i > 0.032$ olan gözlemler etkili gözlemlerdir.

```
> #Etkin Değer(cook uzaklığı)
> cook<-which(inf$cooks>4/n)
> cook
[1] 33 34 35 36 37 38 39
```



Yukarıda ve grafikte görüldüğü gibi $D_{33}, D_{34}, D_{35}, D_{36}, D_{37}, D_{38}, D_{39} > 4/n$ olduğundan, 33, 34, 35, 36, 37 ve 38. gözlem değerlerinin etkin değerler olduğu söylenebilmektedir.

Aykırı değerler veri setinden çıkarılmalıdır. Aykırı değerler veri setinden çıkarıldıktan sonra normallik varsayımı, doğrusallık varsayımı, çoklu regresyon incelemesi yeni veri setine uygulanmalıdır.

```
#verileri ayıklama
veri<-yeni_data[-c(33, 34, 35, 36, 37, 38, 39, 48, 61),]
```

Verileri çıkarttıktan sonra geriye kalan 116 veri ile yeniden model kuruyoruz.

```
Call:
lm(formula = y ~ x1 + x2 + x3 + nitel)

Coefficients:
(Intercept)      x1      x2      x3    nitel2    nitel3
      7.096     7.158     2.390    -1.642    -10.932    -10.984
```

```
> summary(sonuc2)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3 + nitel)

Residuals:
    Min       1Q   Median       3Q      Max
-0.24161 -0.14945 -0.08962  0.07521  1.26493

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.09568    0.36668   19.35  <2e-16 ***
x1             7.15782    0.08025   89.19  <2e-16 ***
x2             2.38987    0.11753   20.34  <2e-16 ***
x3            -1.64212    0.08706  -18.86  <2e-16 ***
nitel2        -10.93230    0.05769 -189.50  <2e-16 ***
nitel3        -10.98374    0.05760 -190.70  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2392 on 110 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9993
F-statistic: 3.343e+04 on 5 and 110 DF,  p-value: < 2.2e-16
```

H_0 : Model anlamsızdır.

H_1 : En az bir β_j Sıfırdan farklıdır. (Model anlamlıdır.)

$p=0.00 < \alpha=0.05$ olduğundan H_0 reddedilir. Kurulan regresyon modelinin anlamlı olduğu %95 güven düzeyinde söylenebilmektedir. En az beş bağımsız değişken, bir bağımlı değişkeni açıklamaktadır. Bu durumda yeniden aykırı değer incelemesi yapmalıyız.

$$12/116 = 0,1034 \cong 0.103$$

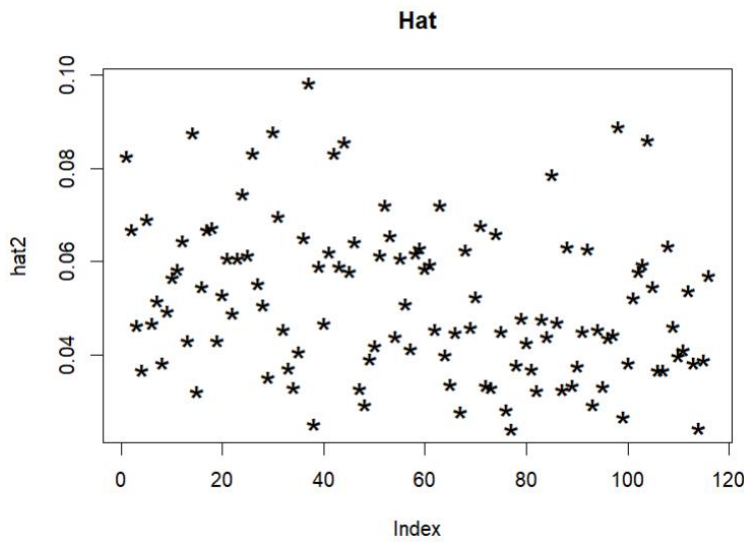
Gözlem Uzaklığı (h_{ii}):

$$h_{ii} > \frac{2(k+1)}{n} = \frac{2p}{n} = \frac{2(5+1)}{116} = 0,103$$

0.103'ten büyük olan verilerde uç değer vardır. Gözden kaçırmamak için verilerimizin grafiğini çizdirerek ve R kodu ile kaçınıcı gözlemlerin uç değer olduğunu bulalım.

```
$hat
[1] 0.08257301 0.06697822 0.04637770 0.03677066 0.06912757 0.04686852 0.05166766 0.03832873 0.04945428
[10] 0.05664342 0.05828650 0.06443812 0.04317531 0.08768116 0.03220167 0.05464222 0.06683152 0.06727868
[19] 0.04298948 0.05302900 0.06087436 0.04906747 0.06087519 0.07461310 0.06140294 0.08330884 0.05537295
[28] 0.05066968 0.03522819 0.08783349 0.06976872 0.04545082 0.03727834 0.03302846 0.04065018 0.06515642
[37] 0.09836250 0.02519634 0.05912536 0.04677226 0.06210266 0.08331187 0.05905296 0.08567209 0.05791130
[46] 0.06430597 0.03278458 0.02927105 0.03923785 0.04192956 0.06141140 0.07223292 0.06548339 0.04401633
[55] 0.06067831 0.05095069 0.04138764 0.06187367 0.06293952 0.05863572 0.05942248 0.04541108 0.07220877
[64] 0.04003044 0.03375988 0.04475038 0.02771139 0.06257431 0.04594006 0.05251207 0.06788540 0.03354330
[73] 0.03304109 0.06612161 0.04504236 0.02826197 0.02408301 0.03775089 0.04794287 0.04271816 0.03706288
[82] 0.03241263 0.04762653 0.04390830 0.07871798 0.04702907 0.03259265 0.06326172 0.03357244 0.03769630
[91] 0.04497008 0.06269442 0.02938983 0.04552355 0.03321065 0.04367078 0.04450289 0.08900366 0.02673477
[100] 0.03838869 0.05235803 0.05804807 0.05942170 0.08603107 0.05467463 0.03677015 0.03670558 0.06337672
[109] 0.04607269 0.03990308 0.04102383 0.05370739 0.03818859 0.02432467 0.03897613 0.05716782
```

```
> #Gözlem uzaklığı 2
> hii2<-which(inf2$hat>0.103)
> hii2
integer(0)
```

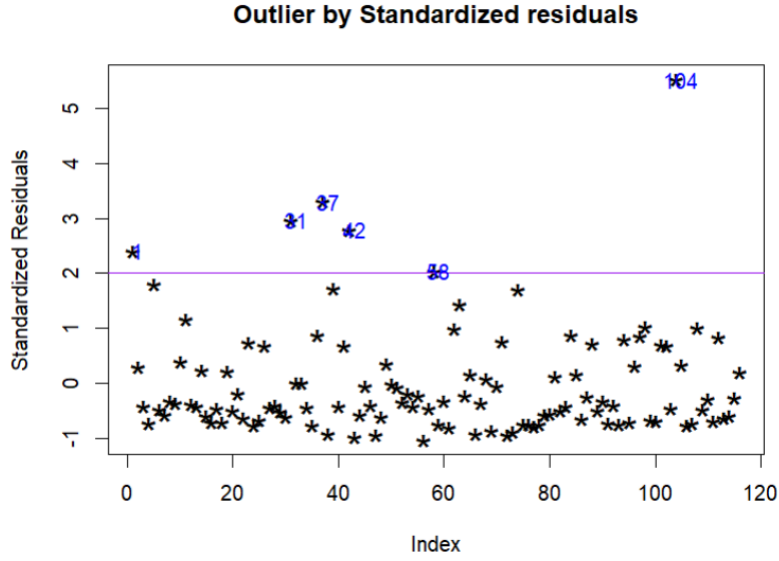


Kodda ve grafikte de görüldüğü gibi **uç değer** yoktur.

Aykırı değer

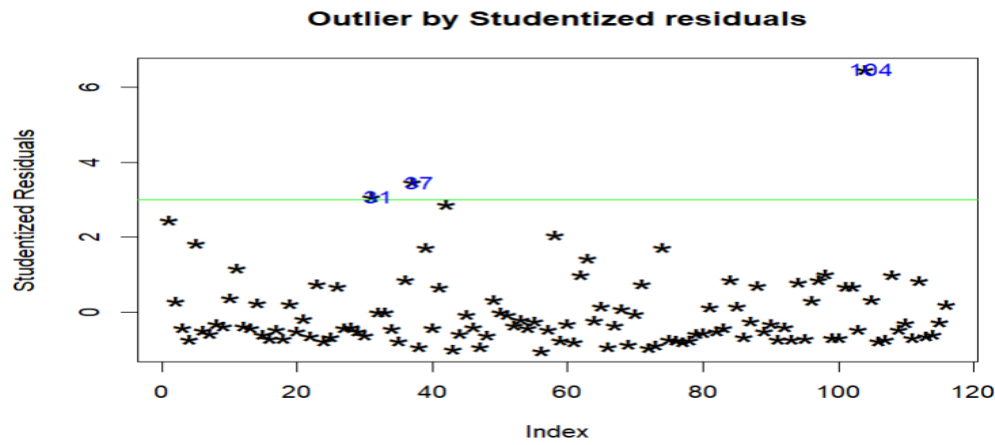
Standartlaştırılmış hataların (-2,+2) aralığında, Student tipi artıkların ise (-3,+3) aralığında olması istenir. Bu aralığın dışında kalan değerlerin aykırı değer olduğu söylenmektedir.

```
$std.res
[1] 2.4101462875 0.2934448325 -0.4115814827 -0.7344843078 1.8099713835 -0.4822762343 -0.5661663884
[8] -0.3135950994 -0.3642708099 0.3888651299 1.1662171566 -0.3829091700 -0.4193374905 0.2418243931
[15] -0.5869001391 -0.7118511010 -0.4626192547 -0.7080102349 0.2266246645 -0.4984747216 -0.1843297355
[22] -0.6361795276 0.7495846927 -0.7715439754 -0.6637996323 0.6870179521 -0.4290634549 -0.3975893271
[29] -0.5153767665 -0.6143050952 2.9673376336 0.0088092442 0.0007822679 -0.4446650296 -0.7678693060
[36] 0.8674105029 3.3031761809 -0.9161513989 1.7168485458 -0.4168150866 0.6735866514 2.7896809070
[43] -0.9829300457 -0.5656473410 -0.0591301405 -0.3972630362 -0.9371753861 -0.6131898475 0.3453017452
[50] -0.0040550232 -0.0701686124 -0.3469782365 -0.1867637037 -0.4190121198 -0.2338464702 -1.0366209841
[57] -0.4570898675 2.0386812265 -0.7518614857 -0.3177810809 -0.8000419561 0.9938681776 1.4267006926
[64] -0.2252381614 0.1631910104 -0.9187505430 -0.3535575642 0.0918742664 -0.8617198162 -0.0471427843
[71] 0.7592532631 -0.9486351265 -0.8772955122 1.7145783126 -0.7386869510 -0.7465561636 -0.7932850082
[78] -0.7528405360 -0.5816187398 -0.5460027271 0.1317626230 -0.5050656069 -0.4212438695 0.8735175320
[85] 0.1662156170 -0.6530026192 -0.2474225432 0.7153964878 -0.4976231263 -0.3169644813 -0.7262852479
[92] -0.3912968269 -0.7376210763 0.7971576554 -0.7133214852 0.3155673308 0.8592569868 1.0263052957
[99] -0.6764056270 -0.6882297752 0.6924388596 0.6814532248 -0.4655438149 5.5304440423 0.3410787747
[106] -0.7638426207 -0.7350671061 1.0014854159 -0.4740334057 -0.2794342860 -0.6794852409 0.8408020319
[113] -0.6326037404 -0.5956047190 -0.2687533248 0.1913591072
```

Tabloda ve grafikte gösterildiği gibi 1,31,37,42,58,104. gözlemler standartlaştırılmış artıklara göre aykırı değerlerdir.

```
$stud.res
[1] 2.465137471 0.292222348 -0.410022224 -0.732937583 1.829169297 -0.480587424 -0.564409986
[8] -0.312306047 -0.362830161 0.387359867 1.168148139 -0.381418982 -0.417761105 0.240786695
[15] -0.585143192 -0.710245844 -0.460960281 -0.706396048 0.225644883 -0.496765138 -0.183518305
[22] -0.634449445 0.748082750 -0.770115570 -0.662102905 0.685359983 -0.427466566 -0.396062663
[29] -0.513649317 -0.612558061 3.079646617 0.008769114 0.000778704 -0.443037573 -0.766427890
[36] 0.866426995 3.464433305 -0.915476938 1.732395180 -0.415244196 0.671905045 2.880744678
[43] -0.982777463 -0.563891039 -0.058861689 -0.395737158 -0.936652629 -0.611442167 0.343915050
[50] -0.004036549 -0.069850499 -0.345586634 -0.185942323 -0.417436439 -0.232838988 -1.036975819
[57] -0.455440174 2.068854290 -0.750366701 -0.316478628 -0.798724295 0.993812443 1.433526024
[64] -0.224263736 0.162467208 -0.918094225 -0.352146960 0.091459212 -0.860703999 -0.046928484
[71] 0.757782456 -0.948199875 -0.876369973 1.730041444 -0.737152223 -0.745044866 -0.791939495
[78] -0.751348862 -0.579861286 -0.544253241 0.131172687 -0.503348587 -0.419663380 0.872569559
[85] 0.165479149 -0.651291232 -0.246363891 0.713799730 -0.495914547 -0.315664630 -0.724716142
[92] -0.389785517 -0.736083272 0.795827982 -0.711719723 0.314271946 0.858227427 1.026556327
[99] -0.674728703 -0.686574105 0.690791390 0.679785054 -0.463880088 6.479241978 0.339704558
[106] -0.762387275 -0.733522024 1.001499074 -0.472356500 -0.278260009 -0.677813601 0.839674037
[113] -0.630870324 -0.593849586 -0.267616806 0.190519021
```



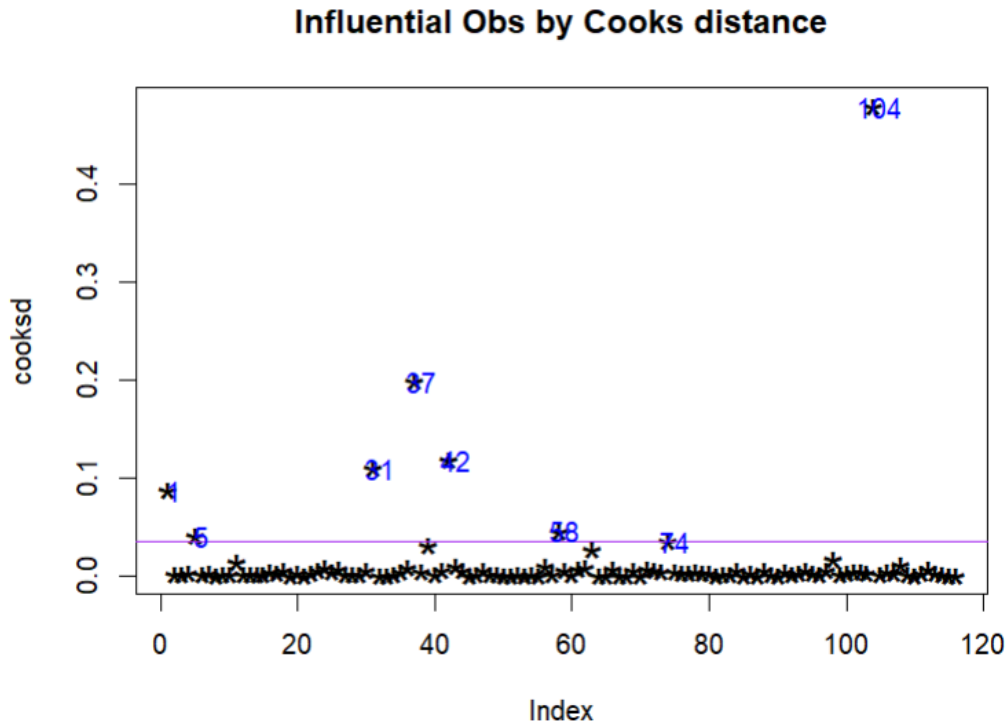
Tabloda ve grafikte görüldüğü gibi 31,37 ve 104. Gözlemler olmak üzere 3 adet verimiz Student tipi artıklara göre artık değerlerdir.

Cook uzaklığı (Di):

$n > 50$ olduğundan, $Di > \frac{4}{n}$ formülü kullanılır. Bu durumda $4/116 = 0.0344$ olduğundan $Di > 0.034$ olan gözlemler etkili gözlemlerdir.

Cook

```
$cooks
[1] 8.713691e-02 1.030252e-03 1.373072e-03 3.432302e-03 4.054650e-02 1.906202e-03 2.910685e-03
[8] 6.532584e-04 1.150610e-03 1.513283e-03 1.402997e-02 1.683103e-03 1.322450e-03 9.367174e-04
[15] 1.910164e-03 4.881566e-03 2.554568e-03 6.026337e-03 3.845108e-04 2.319059e-03 3.670712e-04
[22] 3.480584e-03 6.070259e-03 7.999482e-03 4.804329e-03 7.149126e-03 1.798577e-03 1.406206e-03
[29] 1.616457e-03 6.056239e-03 1.100658e-01 6.158423e-07 3.949259e-09 1.125614e-03 4.163982e-03
[36] 8.740099e-03 1.983855e-01 3.615793e-03 3.087129e-02 1.420781e-03 5.007149e-03 1.178808e-01
[43] 1.010579e-02 4.996635e-03 3.582103e-05 1.807684e-03 4.961774e-03 1.889640e-03 8.115880e-04
[50] 1.199384e-07 5.369184e-05 1.562246e-03 4.073595e-04 1.347303e-03 5.887480e-04 9.615016e-03
[57] 1.503418e-03 4.568693e-02 6.328201e-03 1.048357e-03 6.739543e-03 7.831623e-03 2.640305e-02
[64] 3.525864e-04 1.550805e-04 6.590582e-03 5.937889e-04 9.390652e-05 5.959321e-03 2.052885e-05
[71] 6.997279e-03 5.205596e-03 4.383156e-03 3.469097e-02 4.289500e-03 2.701637e-03 2.588244e-03
[78] 3.705907e-03 2.839138e-03 2.217232e-03 1.113716e-04 1.424192e-03 1.478964e-03 5.840352e-03
[85] 3.934357e-04 3.507239e-03 3.437459e-04 5.760565e-03 1.433717e-03 6.559284e-04 4.139710e-03
[92] 1.706904e-03 2.745792e-03 5.051357e-03 2.913163e-03 7.579077e-04 5.731308e-03 1.715115e-02
[99] 2.094636e-03 3.151515e-03 4.415202e-03 4.769577e-03 2.282023e-03 4.798358e-01 1.121406e-03
[106] 3.712120e-03 3.431435e-03 1.131105e-02 1.808818e-03 5.408782e-04 3.291828e-03 6.687208e-03
[113] 2.648231e-03 1.474031e-03 4.882261e-04 3.700533e-04
```



> cook2

```
[1] 1 5 31 37 42 58 74 104
```

Tabloda grafikte ve kodda gösterildiği gibi 1,5,31,37,42,58,74,104.gözlemlerimiz etkin gözlemlerdir.

*Aykırı gözlemler sorunu çözülerek yeni model kurulup işlemlerimize devam etmemiz gerekmektedir. Fakat burada aykırı değerler belirlenerek devam edilmiştir. Son incelenen aykırı değerler veriden çıkartılmamıştır.

Son veri için kurulan model ve anlamlılığı:

```
Call:
lm(formula = y ~ x1 + x2 + x3 + nitel)

Residuals:
    Min       1Q   Median       3Q      Max
-9.3625 -0.2229  0.0919  1.4365  2.5848

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.6533     3.5538   2.716  0.00759 **
x1             6.2407     0.7587   8.226 2.79e-13 ***
x2             3.7102     1.1468   3.235  0.00157 **
x3            -2.1038     0.8337  -2.524  0.01294 *
nitel2        -8.8424     0.5475 -16.149 < 2e-16 ***
nitel3        -8.8703     0.5496 -16.138 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.428 on 119 degrees of freedom
Multiple R-squared:  0.9315,    Adjusted R-squared:  0.9287
F-statistic: 323.9 on 5 and 119 DF,  p-value: < 2.2e-16
```

H_0 : Model anlamsızdır.

H_1 : En az bir β_j sıfırdan farklıdır. (Model anlamlıdır.)

Yorum: $p=0.00 < \alpha=0.05$ olduğundan H_0 reddedilir. Kurulan regresyon modelinin anlamlı olduğu %95 güven düzeyinde söylenebilmektedir. Burada nitel2 ile gösterilen değişken iki kişilik oda tipi, nitel3 ise aile odasını göstermektedir. Nitel1 yani tek kişilik oda tipi değişkeni kılavuz değişken olduğu için modelde görünmemektedir. Belirtme katsayısına ($R^2=0.93$) bakarak bağımlı değişkendeki değişimin %93'unu bağımsız değişkenler tarafından açıklanabildiğini söyleyebiliriz. Yani konaklama ücretinin %93'ü denize olan uzaklık, sıcaklık, büyüklük ve oda tipi ile açıklandığı söylenebilmektedir. Açıklanmayan kısım %7'dir. Gözlem sayısının artırılması ya da anlamlı başka bağımsız değişkenin modele eklenmesi gibi yöntemler belirtme katsayısını arttıracaktır. (Belirtme katsayısı ve düzeltilmiş (adjusted R^2) belirtme katsayısı neredeyse aynı çıktığı için belirtme katsayısı yorumlanmıştır.)

Kestirim Denklemi:

$$\text{Konaklama Ücreti} = 9.65_{3.55} + 6.24_{0.75} \text{Denize Olan Uzaklık} + 3.71_{1.14} \text{Sicaklik} - 2.10_{0.83} \text{Buyukluk} - 8.84_{0.54} \text{iki kisilik oda} - 8.87_{0.54} \text{Aile tipi} \pm 2.42$$

Model Anlamlılığı:

$H_0: \beta_1=\beta_2=\beta_3=\beta_4=\beta$ (Model anlamsızdır.)

H_1 : En az bir β_i sıfırdan farklıdır.

p-value=00000< $\alpha=0.05$ olduğu için yokluk hipotezi reddedilir. %95 güven düzeyinde doğrusal regresyon modelimiz istatistiksel olarak anlamlıdır. Bağımlı değişkenimizdeki değişim en az bir tane bağımlı değişken tarafından açıklanabilmektedir.

Regresyon Katsayıları :

β_0 için:

$H_0: \beta_0=0$

p-value = 0.007 < α olduğundan dolayı H_0 reddedilir. Sabit terimin modele katkısı %95 güven düzeyinde anlamlıdır.

$H_1: \beta_0 \neq 0$

Yorum: Otel odasının büyüklüğünün 0 olması anlamlı olmadığı için denklem gereği bir katsayıdır.

β_1 için:

$H_0: \beta_1 = 0$

p-value = 2.79×10^{-13} < α olduğundan dolayı H_0 reddedilir. Otel odasının denize uzaklığının modele katkısı istatistiksel olarak anlamlıdır.

$H_1: \beta_1 \neq 0$

Yorum: Diğer değişkenlerin etkisi sabit tutulduğunda Otel odasının denize uzaklığındaki 1 br'lik artış konaklama ücretini ortalama 6.24 br artırır.

β_2 için:

$H_0: \beta_2 = 0$

p-value = 0.001 < α olduğundan dolayı H_0 reddedilir. Otel odasının sıcaklığının modele katkısı istatistiksel olarak anlamlıdır.

$H_1: \beta_2 \neq 0$

Yorum: Diğer değişkenlerin etkisi sabit tutulduğunda Otel odasının sıcaklığındaki 1 br'lik artış konaklama ücretini ortalama 3.71 br artırır.

β_3 için:

$H_0: \beta_3 = 0$

p-value 0.012 < α olduğundan dolayı H_0 reddedilir. Otel odasının büyüklüğünün modele katkısı istatistiksel olarak anlamlıdır.

$H_1: \beta_3 \neq 0$

Yorum: Diğer değişkenlerin etkisi sabit tutulduğunda otel odasının büyüklüğündeki 1 br'lik artış konaklama ücretini ortalama -2.10 br azaltır.

β_4 için:

$H_0: \beta_4 = 0$

p-value 2×10^{-16} < α olduğundan dolayı H_0 reddedilir. Tek kişilik oda ücreti ile Çift kişilik oda ücreti arasında istatistiksel olarak fark vardır.

$H_1: \beta_4 \neq 0$

Yorum: Otel odası ücreti bakımından tek kişilik odaların çift kişilik odalara göre daha etkili olduğunu söyleyebiliriz .

β_5 için:

$H_0: \beta_5 = 0$

$H_1: \beta_5 \neq 0$

p-value $2e^{-16} < \alpha$ olduğundan dolayı H_0 reddedilir. Tek kişilik oda ücreti ile Aile odası ücreti arasında istatistiksel olarak fark vardır

Yorum: . Otel odası ücreti bakımından tek kişilik odaların aile odasına göre daha etkili olduğunu söyleyebiliriz.

Belirtme Katsayısı :

$R^2 = 0.9315$ ve düzeltilmiş belirtme katsayısı $= 0.9287$ değerleri birbirlerine yakın olduğu için çoklu belirtme katsayısı kullanılır. Otel odalarının konaklama ücretlerindeki değişimin %93.15'i bağımsız değişkenlerimiz tarafından açıklanabilmektedir.

Güven Aralıkları :

```
> confint(model, level = .99)
              0.5 %      99.5 %
(Intercept)  0.3501795 18.9564891
x1           4.2546156  8.2268400
x2           0.7081984  6.7122899
x3          -4.2862558  0.0785719
nitel2       -10.2757836 -7.4090878
nitel3       -10.3091751 -7.4315090
```

β_0 : $P(0.350 < \beta_0 < 18.95) = 0.95$

Denklem gereği bir katsayı olduğu için yorum yapamayız.

β_1 (Uzaklık): $P(4.254 < \beta_1 < 8.22) = 0.95$

%95 güvenle diğer değişkenlerin etkisi sabitken otel odasının denize uzaklığındaki 1 birimlik artış otel odasındaki ücreti ortalama **4.254** ile **8.22** birim arasında artırır.

β_2 (Sıcaklık): $P(0.708 < \beta_2 < 6.71) = 0.95$

%95 güvenle diğer değişkenlerin etkisi sabitken otel odasının sıcaklığındaki 1 birimlik artış otel odasındaki ücreti ortalama **0.708** ile **6.71** birim arasında artırır.

β_3 (Büyüklik): $P(-4.28 < \beta_3 < 0.078) = 0.95$

%95 güven düzeyinde diğer değişkenlerin etkisi sabitken otel odasındaki büyüklükteki 1 birimlik artış otel odasındaki ücreti ortalama **-4.28** ile **-0.078** birim arasında azaltır.

β_4 (Tip2): $P(-10.27 < \beta_4 < -7.40) = 0.95$

%95 güven düzeyinde Tip2 olan Çift kişilik oda Tek kişilik odaya göre daha az arttığı yorumunu her iki ucunun negatif olmasından yararlanarak yapabiliriz.

β_5 (Tip3): $P(-10.30 < \beta_5 < -7.43) = 0.95$

%95 güven düzeyinde Aile odası Tek kişilik odaya göre otel odası ücretini daha az arttırır. En uygun fiyatlı olan oda Tek kişilik olandır.

Değişen Varyanslılık Sorunu :

Değişen varyanslılık test yöntemi ve grafik ile incelenebilmektedir.

Test ile incelenmesi:

```
> library(lmtest)
> bptest(model)
```

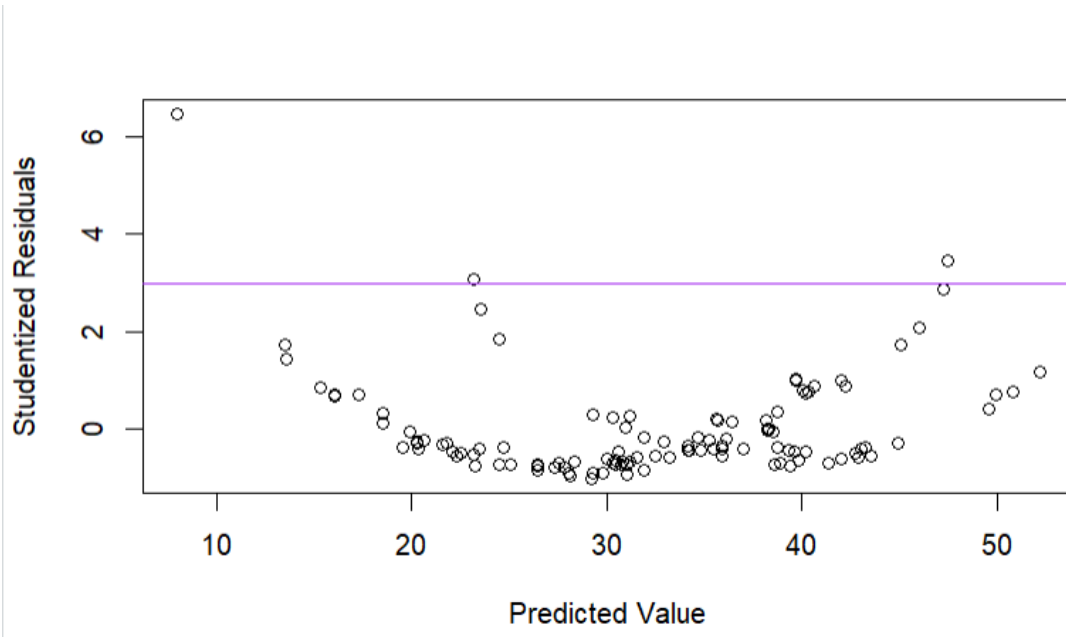
studentized Breusch-Pagan test

```
data: model
BP = 4.3152, df = 5, p-value = 0.505
```

H_0 : Varyanslar homojendir.

Bu durumda H_0 reddedilemez, değişen varyans sorunu bulunamadığını söyleyebiliriz. ($p=0.505 > \alpha=0.05$).

Grafik ile:



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olmadığı görülmektedir. Bu durumda değişen varyanslılık olmadığı yorumunu yapabiliriz. Fakat görsel analiz tek başına yeterli olmadığından test etmemiz gerekir.

```
> summary(lm(abs(residuals(model)) ~ fitted(model)))

Call:
lm(formula = abs(residuals(model)) ~ fitted(model))

Residuals:
    Min       1Q   Median       3Q      Max
-0.17522 -0.08042 -0.02433  0.02417  1.05012

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.230482   0.055942   4.120 7.21e-05 ***
fitted(model) -0.001962   0.001665  -1.179   0.241
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1629 on 114 degrees of freedom
Multiple R-squared:  0.01204, Adjusted R-squared:  0.003372
F-statistic: 1.389 on 1 and 114 DF, p-value: 0.241
```

H_0 : Model anlamsızdır.

$p=0.241 > \alpha=0.05$ olduğundan H_0 reddedilemez. Artıklarla kestirim değerleri arasında oluşturulan modelin anlamsız olduğu söylenebilir. Değişen varyanslılık sorunu yoktur.

Öz ilişki Sorununun İncelenmesi:

```
dwtest(model)

Durbin-Watson test

data: model
DW = 2.0889, p-value = 0.6165
alternative hypothesis: true autocorrelation is greater than 0
```

$H_0: \rho = 0$ (özilişki yoktur)

$H_s: \rho > 0$ (artı özilişki vardır)

Öz ilişki sorunu için Durbin Watson değeri üzerinden inceleme yapılabilmektedir. Model özeti tablosunda test değeri $d=2.0889$ bulunmuştur. Alternatif hipotez kurulmasında bu değer önemlidir. $d=2(1-r)$ yaklaşık formülü dikkate alındığında ilişkinin yönü tablo değerine göre yorumlanabilmektedir. 0 ve 2 arasındaki değerlerde eğer bir öz ilişki varsa bu öz ilişkinin pozitif olacağı biliniyor. Tablo değerine bakıldığında $d_l \cong -0.0889$ ve $d_u \cong 4.0889$ olarak bulunmuştur. Test sonucuna göre $d > d_l$ değeri olduğundan H_0 reddedilemez ve d değeri için verilen p değeri incelendiğinde (-) yönlü bir öz ilişki olduğu görülmektedir ($p=0.61 > \alpha=0.05$)

Bu durumda dönüşüm yapılmalıdır. Fakat burada devamlılık açısından ek bir işlem yapılmadan çözümleme devam etmiştir

Çoklu Bağlantı Sorununun İncelenmesi :

```
> cor(veri)
      y      x1      x2      x3      x4
y  1.0000000 0.83802589 0.83000790 0.76903509 -0.41153866
x1 0.8380259 1.00000000 0.96368920 0.92987158 0.06366982
x2 0.8300079 0.96368920 1.00000000 0.96553674 0.02629917
x3 0.7690351 0.92987158 0.96553674 1.00000000 0.07591728
x4 -0.4115387 0.06366982 0.02629917 0.07591728 1.00000000
```

VIF:

	Variables	Tolerance	VIF
1	x1	0.06959745	14.368343
2	x2	0.03372670	29.650099
3	x3	0.06495168	15.396062
4	nitel2	0.64875693	1.541409
5	nitel3	0.63757991	1.568431

VIF değerlerine baktığımızda büyük veriler vardır. Bu durumda güçlü çoklu bağlantının olduğunu söyleyebiliriz.

Koşul Sayısı:

	Eigenvalue	Condition Index	intercept	x1	x2
1	4.412389176	1.000000	1.726977e-04	1.492414e-04	0.0006554839
2	1.009499504	2.090662	1.137254e-05	9.314341e-07	0.0001795091
3	0.392486294	3.352931	8.275050e-04	5.117047e-06	0.0257538981
4	0.178602980	4.970414	5.187170e-03	1.391472e-03	0.0105545583
5	0.005435606	28.491343	1.778793e-02	1.204244e-01	0.1718772992
6	0.001586439	52.738187	9.760133e-01	8.780289e-01	0.7909792514

	x3	nitel2	nitel3
1	3.364758e-04	9.083420e-03	0.007592203
2	1.020769e-07	1.850331e-01	0.214913911
3	3.169937e-04	2.340667e-01	0.089276041
4	9.690712e-04	5.638410e-01	0.647425098
5	8.547945e-01	4.586321e-05	0.011430645
6	1.435829e-01	7.929913e-03	0.029362102

Koşul sayısı kadar çoklu bağlantı olduğu bilinmektedir. Koşul sayısı 30'dan büyük olduğu durumda çoklu bağlantıdan etkilenilmektedir. Bu durumda 1 tane güçlü çoklu bağlantının, 1 tane güçlü olmayan çoklu bağlantının olduğu söylenebilir.

Çoklu bağlantının yaratmış olduğu sorundan x1,x3ve nitel2 bağımsız değişkenlerinin etkilenmiş olabileceği söylenebilir.

Özdeğer ve Özvektör:

```
[1] 2.9600 1.5200 0.4230 0.0698 0.0223
> signif(sm$eigval,3)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -0.5650  0.0974 -0.0157 -0.72000  0.3900
[2,] -0.5730  0.0868 -0.0752  0.02340 -0.8110
[3,] -0.5630  0.1200 -0.0223  0.69300  0.4330
[4,] -0.1690 -0.6770  0.7160  0.01500 -0.0192
[5,]  0.0812  0.7140  0.6940 -0.00692 -0.0454
```

Özdeğerler arasında sıfıra çok yakın bir değer olmadığından çoklu bağlantının olmadığı söylenebilir.

```
> V<-sm$eigvec
> t(V)%*%V
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.000000e+00 0.000000e+00 -6.938894e-18 4.770490e-17 -1.483189e-16
[2,] 0.000000e+00 1.000000e+00 -1.110223e-16 1.908196e-17 4.857226e-17
[3,] -6.938894e-18 -1.110223e-16 1.000000e+00 -6.071532e-18 9.020562e-17
[4,] 4.770490e-17 1.908196e-17 -6.071532e-18 1.000000e+00 -2.096847e-16
[5,] -1.483189e-16 4.857226e-17 9.020562e-17 -2.096847e-16 1.000000e+00
> V %*% diag(sm$eigval) %*% t(V)
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.00000000 0.96368920 0.92987158 0.1764438 -0.03468878
[2,] 0.96368920 1.00000000 0.96553674 0.1744085 -0.06457191
[3,] 0.92987158 0.96553674 1.00000000 0.1517057 -0.01232276
[4,] 0.17644383 0.17440846 0.15170568 1.00000000 -0.56745849
[5,] -0.03468878 -0.06457191 -0.01232276 -0.5674585 1.00000000
```

En son kod ile özvektörler ve özdeğerlerden ilişki matrisinin elde edilmesi gösterilmiştir.

Uyum Kestirimi :

Kestirim denklemi $y_i = 9.6533_{3.55} + 6.24_{0.75}x_1 + 3.71_{1.14}x_2 - 2.10_{0.83}x_3 - 8.84_{0.54}x_4 - 8.87_{0.54}x_5 \pm 2.428$ yani :

Konaklama Ücreti = $9.65_{3.55} + 6.24_{Denize\ Olan\ Uzaklik_{0.75}} + 3.71_{Sicaklik_{1.14}} - 2.10_{Buyukluk_{0.83}} - 8.84_{iki\ kisilik\ oda_{0.54}} - 8.87_{Aile\ tipi_{0.54}} \pm 2.42$

Olarak elde edilmiştir. Uyum Kestirimi için 58. gözlem değeri alındığında yani

$x_1=7.013, x_2=2.888, x_3=4.411, x_4=2$ (iki kişilik oda) olan gözlem kümesi için uyum kestirimi :

$$y_i = 9.65 + 6.24x_1 + 3.71x_2 - 2.10x_3 - 8.84x_4 - 8.87x_5 \pm 2.428$$

$$y_i = 9.65 + 6.24*7.01 + 3.71*2.88 - 2.10*4.41 - 8.84*1 - 8.87*0 \pm 2.428$$

$$\hat{y}_i = 45.97 \pm 2.428 \text{ olarak bulunmuştur.}$$

Gerçek y değerimizin 46.494 olduğu bilinmektedir. Bu durumda yuvarlamaları da göz önüne alarak modelimizin başarılı olduğu söylenebilir.

Ön Kestirim :

$$y_i = 9.6533_{3.55} + 6.24_{0.75}x_1 + 3.71_{1.14}x_2 - 2.10_{0.83}x_3 - 8.84_{0.54}x_4 - 8.87_{0.54}x_5 \pm 2.428$$

$x_1 = 2, x_2 = 4, x_3 = 3, x_4 = 3$ (aile tipi oda) için ön kestirim elde edelim:

$$y_i = 9.65 + 6.24 \cdot 2 + 3.71 \cdot 4 - 2.10 \cdot 3 - 8.84 \cdot 0 - 8.87 \cdot 1 \pm 2.428$$

$\hat{y} = 21.8 \pm 2.428$ olarak bulunmuştur.

Uyum Kestirimi için Güven Aralığı :

```
> #Uyum Kestirimi
> predict(model,interval = "confidence",level=0.95)[58,]
      fit      lwr      upr
46.02172 45.90378 46.13966
```

Burada $x_1 = 7.013, x_2 = 2.888, x_3 = 4.411, x_4 = 2$ (iki kişilik oda) olan 58. gözlem değerimizin verileri alındığında yani denize olan uzaklık 7.013km, sıcaklık yaklaşık 2 iken, büyüklük 4.411 ve iki kişilik oda durumu göz önüne alındığında konaklama ücreti indeksinin 45.90 ve 46.13 arasında olduğunu dolayısıyla $E(\hat{y})$ için güven aralığının [45.90, 46.13] olduğunu %95 güven düzeyinde söyleyebiliriz.

Ön Kestirim için Güven Aralığı :

```
> #Ön Kestirim
> onkestirim<-data.frame(x1=3,x2=4,x3=10,x4=x43)
> predict(model,onkestirim,interval = "prediction",level=0.95)
      fit      lwr      upr
1  21.70736 20.545451 22.86927
```

Burada $x_1 = 3, x_2 = 4, x_3 = 10, x_4 = 3$ (aile oda tipi) olarak alındığında denize olan uzaklık 2 km., sıcaklık 4 iken, büyüklük 3 ve oda tipi göz önüne alındığında hava konaklama ücretinin 20.54 ve 22.86 arasında olduğunu dolayısıyla $E(\hat{y})$ için güven aralığının [20.54, 22.86] olduğunu %95 güven düzeyinde söyleyebiliriz.

Değişken Seçimi Yöntemleri ile En İyi Modelin Bulunması

İleriye Doğru Seçim Yöntemi (Forward):

Start: AIC=513.93

y ~ 1

	Df	Sum of Sq	RSS	AIC
+ x1	1	6723.5	2850.2	375.38
+ x2	1	6595.5	2978.3	380.48
+ x3	1	5662.1	3911.7	412.10
+ x4	2	1747.7	7826.0	494.55
<none>			9573.8	513.93

Step: AIC=375.38

y ~ x1

	Df	Sum of Sq	RSS	AIC
+ x4	2	2818.15	32.08	-141.11
+ x2	1	67.44	2782.79	374.60
<none>			2850.23	375.38
+ x3	1	7.39	2842.84	377.08

Step: AIC=-141.11

y ~ x1 + x4

	Df	Sum of Sq	RSS	AIC
+ x2	1	5.4192	26.660	-160.57
+ x3	1	2.1149	29.964	-147.02
<none>			32.079	-141.11

Step: AIC=-160.57

y ~ x1 + x4 + x2

	Df	Sum of Sq	RSS	AIC
+ x3	1	20.363	6.2962	-325.98
<none>			26.6597	-160.57

Step: AIC=-325.98

y ~ x1 + x4 + x2 + x3

> forward

Call:

lm(formula = y ~ x1 + x4 + x2 + x3)

Coefficients:

(Intercept)	x1	x42	x43	x2
7.096	7.158	-10.932	-10.984	2.390
x3				
-1.642				

Birinci adımda modele x1 değişkeni girmiştir. Daha sonra diğer değişkenler modele alındığına hepsi anlamlı çıktığı için modelde değişkenlerin tamamı yer almaktadır.

```

Call:
lm(formula = y ~ x1 + x4 + x2 + x3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.24161 -0.14945 -0.08962  0.07521  1.26493

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.09568    0.36668   19.35  <2e-16 ***
x1             7.15782    0.08025   89.19  <2e-16 ***
x42            -10.93230    0.05769 -189.50  <2e-16 ***
x43            -10.98374    0.05760 -190.70  <2e-16 ***
x2             2.38987    0.11753   20.34  <2e-16 ***
x3            -1.64212    0.08706  -18.86  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2392 on 110 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9993
F-statistic: 3.343e+04 on 5 and 110 DF,  p-value: < 2.2e-16

```

En iyi model: $y_i = 7.09_{3.66} + 7.15_{0.80}x_1 + 2.38_{0.11}x_2 - 1.64_{0.87}x_3 - 10.93_{0.057}x_{42} - 10.98_{0.057}x_{43} \pm 0.2392$

Bağımsız değişkenlerin de p değerlerine baktığımızda hepsinin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

$P=0.00<0.05$ olduğundan modelin anlamlı söylenebilmektedir.

Geriye Doğru Çıkarma Seçim Yöntemi (Backward):

```

Start:  AIC=-325.98
y ~ x1 + x2 + x3 + nitel

```

	Df	Sum of Sq	RSS	AIC
<none>			6.30	-325.98
- x3	1	20.36	26.66	-160.57
- x2	1	23.67	29.96	-147.02
- x1	1	455.31	461.60	170.21
- nitel	2	2631.36	2637.66	370.39

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlıyor. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. Değişkenlerin hepsi anlamlı olduğu için hiçbir modelden çıkmamıştır.

```

Call:
lm(formula = y ~ x1 + x2 + x3 + nitel)

Residuals:
    Min       1Q   Median       3Q      Max
-0.24161 -0.14945 -0.08962  0.07521  1.26493

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.09568    0.36668   19.35  <2e-16 ***
x1             7.15782    0.08025   89.19  <2e-16 ***
x2             2.38987    0.11753   20.34  <2e-16 ***
x3            -1.64212    0.08706  -18.86  <2e-16 ***
nitel2        -10.93230    0.05769 -189.50  <2e-16 ***
nitel3        -10.98374    0.05760 -190.70  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2392 on 110 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.9993
F-statistic: 3.343e+04 on 5 and 110 DF,  p-value: < 2.2e-16

```

Aynı bağımsız değişkenler ile model kurulduğundan ileri ve geriye doğru seçim yöntemleri aynı sonuçları vermiştir. Burada da tüm değişkenlerin olduğu model en iyi model olarak elde edilmiştir. Bağımsız değişkenlerin de p değerlerine baktığımızda hepsinin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Modelin de anlamlı olduğu söylenebilir($p=0.000$).

En iyi model: $y_i = 7.09_{3.66} + 7.15_{0.80}x_1 + 2.38_{0.11}x_2 - 1.64_{0.87}x_3 - 10.93_{0.057}x_4 - 10.98_{0.057}x_5 \pm 0.2392$

Adımsal Regresyon Yöntemi (Stepwise):

Call:

```
lm(formula = y ~ x1 + x2 + x3 + nitel)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.24161	-0.14945	-0.08962	0.07521	1.26493

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.09568	0.36668	19.35	<2e-16 ***
x1	7.15782	0.08025	89.19	<2e-16 ***
x2	2.38987	0.11753	20.34	<2e-16 ***
x3	-1.64212	0.08706	-18.86	<2e-16 ***
nitel2	-10.93230	0.05769	-189.50	<2e-16 ***
nitel3	-10.98374	0.05760	-190.70	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2392 on 110 degrees of freedom

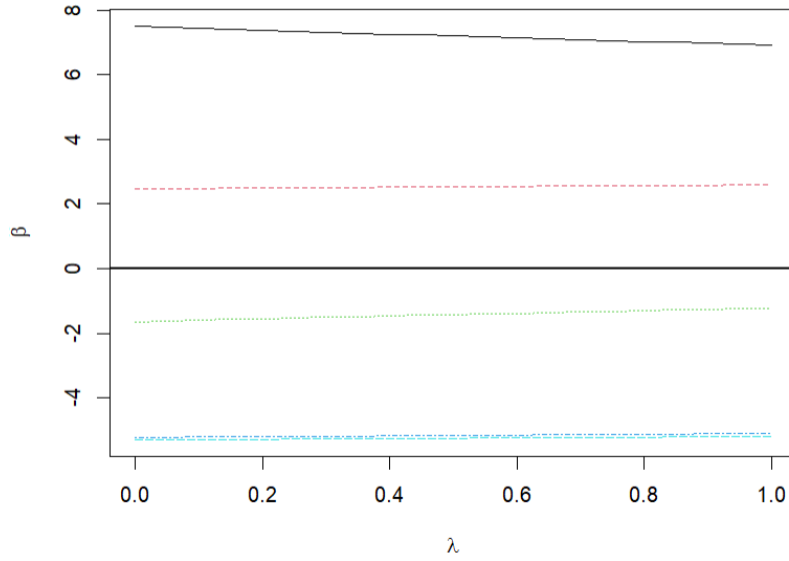
Multiple R-squared: 0.9993, Adjusted R-squared: 0.9993

F-statistic: 3.343e+04 on 5 and 110 DF, p-value: < 2.2e-16

Adımsal seçim yönteminin sonucunun da ileri seçim ve geriye doğru seçim ile aynı olduğu görülmektedir. Model ve bağımsız değişkenler anlamlıdır.

Ridge Regresyon :

```
> ridgeCoef
      0.00      0.05      0.10      0.15      0.20      0.25      0.30
x1  7.509762  7.476230  7.443206  7.410674  7.378621  7.347034  7.315898
x2  2.459591  2.468235  2.476547  2.484544  2.492243  2.499660  2.506809
x3 -1.644000 -1.621369 -1.598904 -1.576609 -1.554487 -1.532540 -1.510770
x42 -5.226081 -5.219685 -5.213315 -5.206971 -5.200652 -5.194358 -5.188088
x43 -5.305032 -5.299445 -5.293882 -5.288345 -5.282832 -5.277341 -5.271873
      0.35      0.40      0.45      0.50      0.55      0.60      0.65
x1  7.285203  7.254937  7.225088  7.195646  7.166602  7.137944  7.109664
x2  2.513702  2.520354  2.526774  2.532976  2.538967  2.544759  2.550361
x3 -1.489178 -1.467766 -1.446534 -1.425482 -1.404611 -1.383922 -1.363413
x42 -5.181841 -5.175618 -5.169418 -5.163241 -5.157085 -5.150951 -5.144839
x43 -5.266426 -5.261000 -5.255594 -5.250208 -5.244840 -5.239492 -5.234161
      0.70      0.75      0.80      0.85      0.90      0.95      1.00
x1  7.081752  7.054201  7.027002  7.000147  6.973629  6.947439  6.921571
x2  2.555780  2.561024  2.566102  2.571021  2.575786  2.580405  2.584883
x3 -1.343085 -1.322938 -1.302969 -1.283180 -1.263568 -1.244134 -1.224875
x42 -5.138747 -5.132676 -5.126626 -5.120596 -5.114585 -5.108595 -5.102623
x43 -5.228848 -5.223553 -5.218274 -5.213011 -5.207765 -5.202534 -5.197319
>
> select(ridge)
modified HKB estimator is 0.00142378
modified L-W estimator is 0.002081927
smallest value of GCV at 0
> ridge$coef[,ridge$lam == 0.4]
      x1      x2      x3      x42      x43
7.254937 2.520354 -1.467766 -5.175618 -5.261000
```



Ridge grafiğimize baktığımızda sıfır eksenine çok yakın değişkenlerin olmadığı görülmektedir. Dolayısıyla değişkenlerimizin hepsinin modele katkısı anlamlıdır. Grafiğimizde hızlı artış ya da azalış da olmadığından çoklu bağlantının olmadığını çoklu bağlantı varsa bile değişkenlerin çok fazla etkilenmediğini söyleyebiliriz. Çoklu bağlantılı değişkenlere karşılık gelen parametre kestirimlerinin çoklu bağlantıdan ne derece etkilendiklerini, dolayısıyla bu katsayıların kararlı olup olmadıklarını gösterir. Lamda parametresine göre katsayılar değişmektedir. Hızlı azalış ve artışların bittiği yerde bir lambda parametresi belirlenerek katsayı kestirimleri elde edilir ve model kurulur. Lamdanın 0,4 olarak alındığı durumun sonuçlarına göre katsayı kestirimleri elde edilmiştir.

Kaynakca :

<https://www.rdocumentation.org/>

[https://www.academia.edu/43213352/R Programlama Grafikleri](https://www.academia.edu/43213352/R_Programlama_Grafikleri)

https://cran.r-project.org/web/packages/olsrr/vignettes/influence_measures.html

Regresyon Çözümlmeleri dersi, ders ve uygulama notları