



HACETTEPE ÜNİVERSİTESİ

Regresyon Çözümlemesi Dönem Ödevi

Prof. Dr. Duru KARASOY
Arş. Gör. Dr. Hatice IŞIK

Sena POLAT

21822087

İçindekiler Tablosu

1-Verilerin Hikayesi:.....	2
2-Tanımlayıcı İstatistikler.....	3
4-Çoklu Regresyon Modeli ve Artık İncelemesi.....	11
5. Çoklu Regresyon Model Denklemi.....	18
6. Regresyon Katsayıları.....	19
7. Belirtme Katsayısı	20
8. Güven Aralıkları	21
9. Değişen Varyanslılık Sorunu	22
10. Öz İlişki Sorunu	24
11.Çoklu Bağlantı Sorunu	25
12.Uyum Kestirimi	28
13.Ön Kestirim	28
14.Uyum Kestirimi ve Ön Kestirim için Güven Aralıkları	29
15.Değişken Seçimi.....	29
16.Ridge Regresyon Modeli.....	34

1-Verilerin Hikayesi:

Şarap üretimi yapan bir fabrika müşterilerini memnun etmek için en kaliteli Pinot Noir üretimi yapmak istemektedir. Bu amaçla dünyanın farklı coğrafi bölgelerinden getirdiği Pinot Noir üzümlerinden şarap üretmiştir. Üretilen şarapların aromasına, lezzetine ve gövdesine ilişkin verileri kaydederek şarap kalitesini incelemiştir. 1 yıl sonunda analizlerini gerçekleştirerek, şarap üretimine başlamıştır.

Y: Şarap Kalitesi (Bağımlı Değişken)

X1: Aroma

X2: Lezzet

X3: Gövde

X4: Bölge (Nitel Değişken)

Bölgeler: 1<- Fransa- Burgonya

2<- ABD- Willamette Vadisi

3<- Yeni Zelanda- Martinborough

Not: Bu veri setine uygulanan her analiz %95 güven düzeyinde uygulanmaktadır. ($\alpha = 0.05$)

Veri Setini Yükleme

```
> veri <- read.csv("C:/Users/teknosa/Desktop/REGRESYON/ödev/veri.txt", sep="")
> View(veri)
> attach(veri)
The following objects are masked from veri (pos = 3):
    x1, x2, x3, x4, y
> names(veri)<-c("kalite","Aroma","Lezzet","Gövde","Bölge")
> names(veri)
[1] "kalite" "Aroma" "Lezzet" "Gövde" "Bölge"
> bölge<-as.factor(Bölge)
```

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

2-Tanımlayıcı İstatistikler

- Kalite için:

```
> summary(Kalite)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 13.12  21.20   26.45   27.62  30.58   84.63
> |
```

Bağımlı değişkenimiz olan “Şarap Kalitesi”ne ilişkin tanımlı istatistiklere baktığımız zaman en düşük kalitenin 13.12 en yüksek şarap kalitesinin ise 84.63 olduğunu söyleyebiliriz.

Çeyrekler açıklığını 3rd Qu-1st Qu işlemi sonucunda bulabiliriz. Bu açıklık bize en düşük ve en yüksek şarap kalitesinin birbirinden ne kadar uzak olduğunu gösterir.

- Çeyrekler açıklığı= $30.58 - 21.20 = 9.38$

Ortalama şarap kalitesi medyandan büyük olduğu için veri kümesi sağa çarpıktır diyebiliriz.

- Aroma için:

```
> summary(Aroma)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.857  4.314   5.009   4.975  5.537   7.873
> summary(Legzet)
```

Şarapların içerdiği en düşük aroma miktarının 1.857 en yüksek aroma miktarının ise 7.873 olduğunu söyleyebiliriz.

Şaraplardaki ortalama aroma miktarını ise 4.975'tir. Ortancanın ortalamadan büyük olması bizlere şarapların içerdiği aroma miktarına ilişkin verilerin sola çarpık olduğu yorumunu yaptırır.

- Lezzet için:

```
> summary(Lezzet)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.512   4.258   4.991   4.996   5.630   7.792
```

Pinot Noir şaraplarının ortalama lezzetinin 4.996 olduğunu söyleyebiliriz. En lezzetli ve en lezzetsiz şaraplar arasındaki değişim ise 1,372'dir.

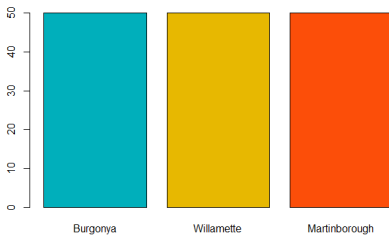
- Gövde için:

```
> summary(Gövde)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.973   5.313   5.985   6.011   6.658   8.601
```

Şarapların üretiminde kullanılan Pinot Noir üzümlerinin ortalama gövde uzunluğunun 6.011 , en kısasının 2.973 , en uzunun ise 8.601 olduğunu söyleyebiliriz.

- Bölge için:

```
> levels(bölge) <- c("Burgonya", "willamette", "Martinborough")
> summary(bölge)
  Burgonya  willamette Martinborough
        50         50         50
```



Pinot Noir şaraplarının 50 tanesi Burgonya'dan, 50 tanesi Willamette Vadisi'nden ve son 50 tanesinin de Martinborough'den gelen üzümlerden üretildiğini söyleyebiliriz

.3-Normallik ve Doğrusallık varsayımlarının incelenmesi:

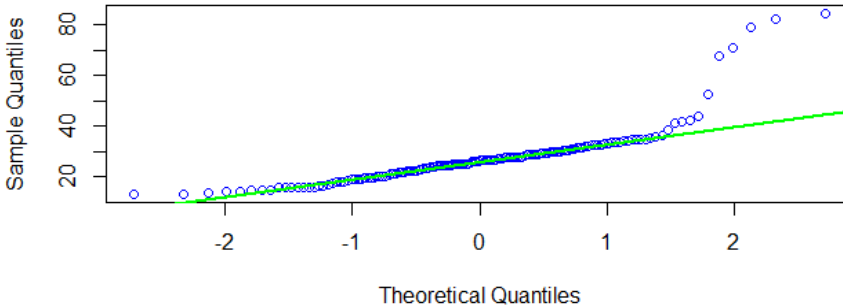
Normallik Varsayımı:

```
qqnorm(kalite,col="blue")  
qqline(kalite,col="green", lwd = 2)
```

col fonksiyonu yardımıyla grafikte renklendirme yapılmıştır.

Eğer kalite değerlerimiz 45^o’lık bir açı ile dağılırsa verilerimizin normal dağıldığını söyleyebiliriz.

Normal Q-Q Plot



Veri setimizdeki bağımlı değişken olan Kalite (y) için çizilen Q-Q plot grafiğinde verilerin çoğunluğu oluşturulan çizgi çevresinde olsa da sapan değerlerimizin olduğunu görüyoruz. Kesin bir fikre ulaşmak **Kolmogrov-Smirnow veya Anderson-Darling normallik testi** uygulanmalıdır.

H_0 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark vardır.

```
> ks.test(kalite,"pnorm")
One-sample Kolmogorov-Smirnov test

data: kalite
D = 1, p-value < 2.2e-16
alternative hypothesis: two-sided

> ad.test(kalite)
Anderson-Darling normality test

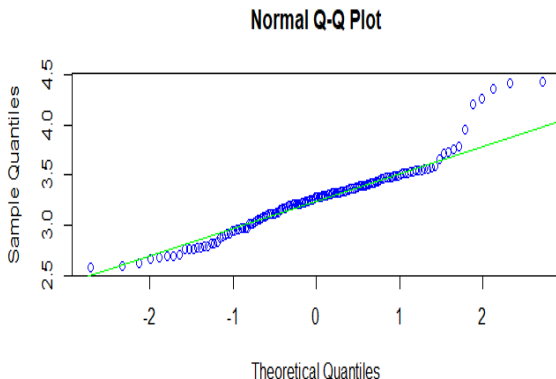
data: kalite
A = 8.2438, p-value < 2.2e-16
```

Her iki testte de $p\text{-value} < \alpha=0.05$ olduğundan yokluk hipotezimiz reddedilir. Yüzde 95 güven düzeyinde şarap kalitesine ilişkin verilerin normal dağılmadığı yorumu yapılmaktadır. Dönüşüm yapılmalıdır.

Dönüşüm:

1. **Logaritmik Dönüşüm:** İlk olarak şarap kalitesi verilerimize logaritmik dönüşüm uyguluyoruz. Yeni verilerimizle q-q plot grafiği çiziyoruz.

```
> lnkalite<-log(kalite)
> qqnorm(lnkalite,col="blue")
> qqline(lnkalite,col="green")
```



Q-Q Plot grafiğini incelediğimizde yine verilerimizin çoğu doğru hizasında olsa da sapan değerler olduğunu görmekteyiz. Emin olmak için Normallik Testlerinden uygulamalıyız.

H_0 : Şarap kalitesi($\log(y)$) verilerinin dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Şarap kalitesi($\log(y)$) verilerinin dağılımı ile normal dağılım arasında fark vardır.

```
> ad.test(lnkalite)

Anderson-Darling normality test

data:  lnkalite
A = 1.8629, p-value = 8.865e-05
```

p-value < $\alpha=0.05$ olduğundan yokluk hipotezimiz reddedilir. Yüzde 95 güven düzeyinde logaritmik dönüşüm yapılmış şarap kalitesine ilişkin verilerin normal dağılmadığı yorumu yapılmaktadır.

2. **Ters Dönüşüm:** Şarap kalitesi verilerimize bu sefer de ters dönüşüm yaparak Normallik testi yapalım.

H_0 : Şarap kalitesi($1/y$) verilerinin dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Şarap kalitesi($1/y$) verilerinin dağılımı ile normal dağılım arasında fark vardır.

```
> ad.test(tkalite)

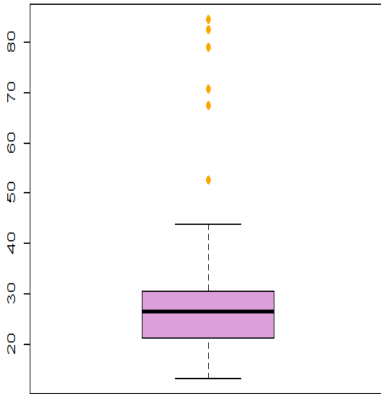
Anderson-Darling normality test

data:  tkalite
A = 2.6738, p-value = 9.056e-07
```

p-value < $\alpha=0.05$ olduğundan yokluk hipotezimiz reddedilir. Yüzde 95 güven düzeyinde ters dönüşüm yapılmış şarap kalitesine ilişkin verilerin normal dağılmadığı yorumu yapılmaktadır.

Dönüşüm yapılmasına rağmen şarap kalitesine ilişkin veriler normal dağılıma uyum göstermemiştir. Normallik varsayımının sağlanması için y değişkenimizde ki sapan değerleri veri setimizden çıkarmalıyız. Bunun içinde **box plot** grafiği çizilmelidir.

```
> boxplot(kalite,col="plum",outcol="orange",outpch=19)
```



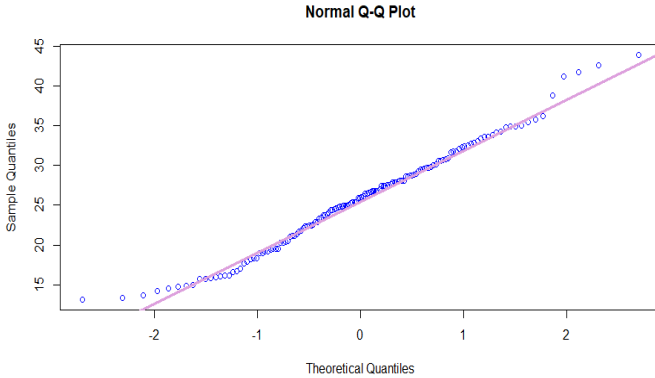
Box Plotu incelediğimiz zaman 6 tane aykırı değerimiz olduğunu görüyoruz. Bu aykırı değerlerin ne olduğunu görmek için aşağıdaki kodu kullanırız. Normalliğin sağlanması amacıyla bu sapan değerler veri setinden çıkartılarak analize öyle devam edilir.

```
> boxplot.stats(kalite)$out  
[1] 52.50738 84.62933 82.41882 70.76303 78.92126 67.45629
```

Aykırı Değerleri silme:

```
> veri2 <- read.delim2("c:/users/teknosa/Desktop/REGRESYON/ödev/yeniveri.txt", header=TRUE)  
> view(veri2)  
> names(veri2)<-c("Kalite","Aroma","Lezzet","Gövde","Bölge")  
> names(veri2)  
[1] "Kalite" "Aroma" "Lezzet" "Gövde" "Bölge"  
> attach(veri2)  
The following objects are masked from veri (pos = 3):  
Aroma, Bölge, Gövde, Kalite, Lezzet  
> bölge<-as.factor(Bölge)
```

Yeni veri setimiz üzerinden normallik varsayımını tekrar kontrol edelim.



Q-Q Plot grafiğini incelediğimizde çoğu gözlemin normallik çizgisinde olduğunu görüyoruz.

H_0 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark vardır.

```
> library(nortest) #Anderson-Darling testi için kullanılan kütüphane  
> ad.test(yeniveri$kalite)
```

Anderson-Darling normality test

```
data: yeniveri$kalite  
A = 0.34772, p-value = 0.4739
```

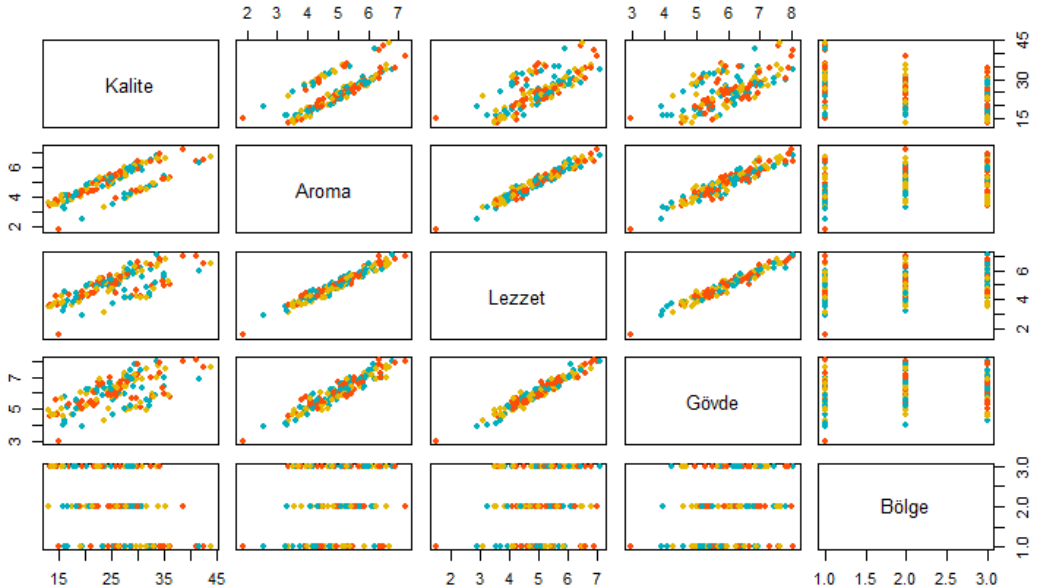
%95 güven düzeyinde $p\text{-value} > 0.05$ olduğundan yokluk hipotezimiz reddedilemez. Yani, şarap kalitesi verilerimizin normal dağılıma uyuma istatistiksel olarak anlamlıdır.

Doğrusallık Varsayımı:

my_cols değişkeni ile grafikte renklendirme yapılmıştır.

```
my_cols <- c("#00AFBB", "#E7B800", "#FC4E07")  
pairs(yeniveri, pch = 19, cex = 0.8, main="İkili Korelasyon Grafikleri",  
      col = my_cols)
```

İkili Korelasyon Grafikleri



Bağımlı değişkenimiz şarap kalitesi ile bağımsız değişkenlerimiz arasındaki ilişki incelendiğinde doğrusallığın sağlandığını görürüz. Ayrıca ilişkilerin pozitif yönlü olduğunu da görmekteyiz. Bölge değişkenimiz nitel değişken olduğu için doğrusallık incelemesi bu yöntem ile yapılmaz.

4-Çoklu Regresyon Modeli ve Artık İncelemesi

Çoklu Regresyon Modelinin Kurulması

```
> sonuc<-lm(Kalite~Aroma+Lezzet+Gövde+Bölge)
> summary(sonuc)
```

Call:
lm(formula = Kalite ~ Aroma + Lezzet + Gövde + Bölge)

Residuals:

Min	1Q	Median	3Q	Max
-8.980	-3.305	-0.511	0.679	50.700

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.8498	5.8756	1.506	0.1342
Aroma	6.9677	3.7026	1.882	0.0619 .
Lezzet	-0.3062	5.0102	-0.061	0.9514
Gövde	-0.7720	3.5399	-0.218	0.8277
Bölge	-4.8619	0.9628	-5.050	1.31e-06 ***

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

Residual standard error: 9.542 on 145 degrees of freedom
Multiple R-squared: 0.3245, Adjusted R-squared: 0.3058
F-statistic: 17.41 on 4 and 145 DF, p-value: 1.096e-11

$p=0.00<0.05$ olduğu için modelin istatistiksel olarak anlamlı olduğu görülmektedir. Bağımlı değişkenin sadece %32.45'inin bağımsız değişkenlerce açıklanması modelde varsayım bozulmalarının olduğunu ve bazı bağımsız değişkenlerin anlamsız olması düşündürür.

Artıkların İncelenmesi

```
install.packages("zoo")
library(zoo)
inf<-ls.diag(sonuc)
inf
```

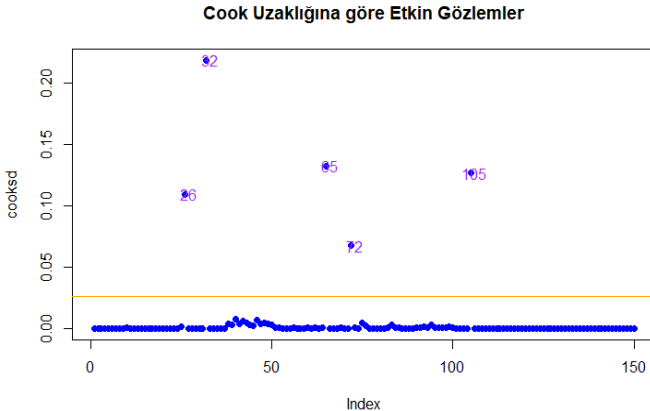
- Cook Uzaklığı için inceleme:

Cook uzaklığı verideki etkin değerleri verir. Cook değeri Di ile ifade edilir.

$$Dh = \frac{4}{n} = \frac{4}{144} \cong 0,028$$

Di>Dh olan gözlemler etkili gözlemlerdir.

```
cooksds <- cooks.distance(sonuc)
plot(cooksds, pch=19, cex=1, col="blue", main="Cook Uzaklığına göre Etkin Gözlemler")
abline(h = if (length(kalite)>50) 4/length(kalite) else 4/(length(kalite)-(length(yeniveri)-1)-1) ,
       col="orange")
text(x=1:length(cooksds)+1, y=cooksds, labels=ifelse(cooksds>if (length(kalite)>50) 4/length(kalite)
                                                    else 4/(length(kalite)-(length(yeniveri)-1)-1), names(cooksds), ""), col="purple")
```



Cook değerinden büyük değerler etkin değerleri verir. 26,32,65,72 ve 105. gözlemlerin etkin gözlem olduğu görülmektedir. Etkin gözlemler katsayı kestirimlerimizi etkiler.

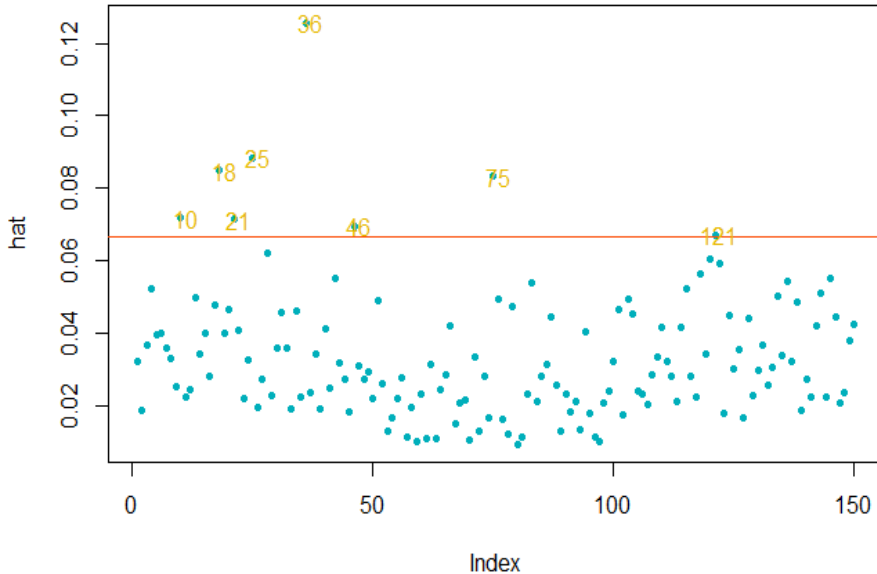
- Uç değ er i in inceleme:

h_{ii} , her g zlemin merkeze olan uzaklıđı olup $[0,1]$ aralıđında değ er alır.

$h_{ii} > \frac{2(k+1)}{n} = \frac{2.5}{144} \cong 0,07$ değ erinden b y k g zlemler u  değ eridir.

```
hat<-inf$hat
plot(hat, pch=20, cex=1,col="#00AFBB", main="G zlem Uzaklıđına g re U  Deđerler")
abline(h = 2*length(yeniveri)/length(kalite) , col="#FC4E07")
text(x=1:length(hat)+1, y=hat, labels=ifelse(hat>2*length(yeniveri)/length(kalite),index(hat),""), col="#E7B800")
```

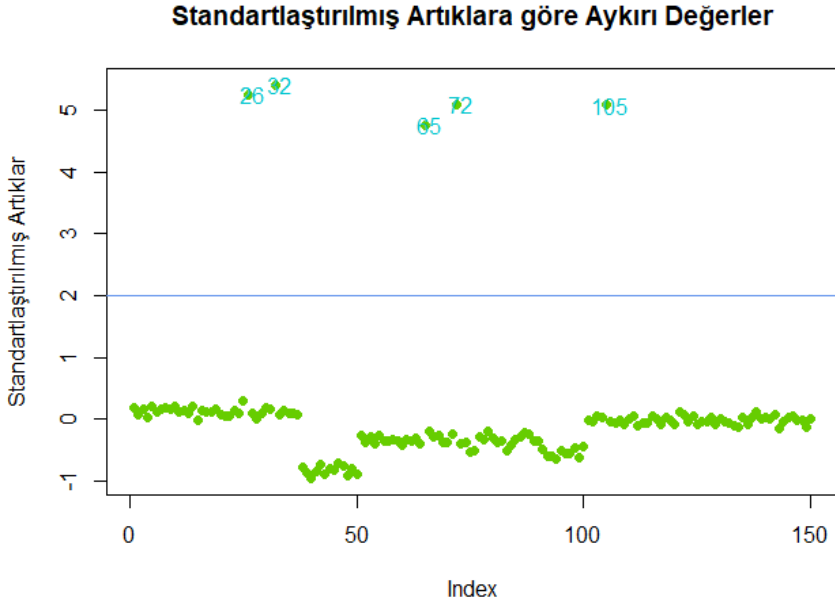
G zlem Uzaklıđına g re U  Deđerler



10,18,21,25,36,46,75 ve 121. g zlemler u  değ erlerdir.

- Standartlaştırılmış Artıkların İncelenmesi:

```
std<-inf$std.res
plot(std, pch=19, cex=1,col="chartreuse3", main="Standartlaştırılmış Artıklara göre Aykırı Değerler",
      +ylab="Standartlaştırılmış Artıklar", xlab="Index")
abline(h = c(-2,2) ,col="cornflowerblue")
text(x=1:length(std)+1, y=std, labels=ifelse(std<-2 & std>2,index(std),""), col="turquoise3")
```

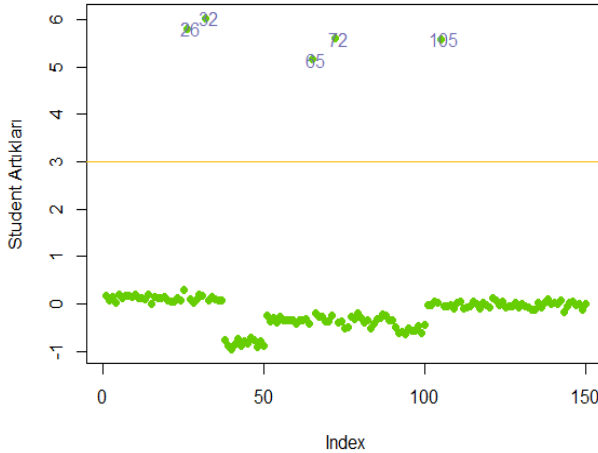


Kurmuş olduğumuz modelde .std.resid tipi artıkların (-2 ve 2) aralığı dışında kalan artıklar aykırı değerlerdir. 26,32,65,72 ve 105. İndexteki gözlemler aykırı değerlerdir.

- Student türü artıklar ile aykırı değer incelemesi:

```
stud<-inf$stud.res
plot(stud, pch=19, cex=1,col="chartreuse3", main="Student türü Artıklara göre Aykırı Değerler",
      ylab="Student Artıkları", xlab="Index")
abline(h = c(-3,3) ,col="darkgoldenrod1")
text(x=1:length(stud)+1, y=stud, labels=ifelse(stud<-3 & stud>3,index(stud),""), col="#7570B3")
```

Student türü Artıklara göre Aykırı Değerler



Student türü artıklara göre $[-3, +3]$ dışındaki artıklara denk gelen gözlemler aykırı değerdir. Grafiği incelediğimizde 26,32,65,72 ve 105. Gözlemlerin artık değerler olduğunu görmekteyiz.

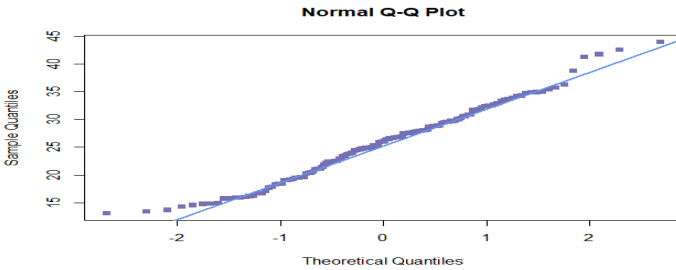
Not: Aykırı değer indexleri; Cook's ve Students tiplerinde aynı değerlere karşılık gelmektedir.

Tespit edilen aykırı değerler, uç değerler ve etkin değerler veri setinden çıkarıldıktan sonra tekrar işlem yapılır. Bu değerler veri setinden çıkarıldıktan sonra normallik varsayımı, doğrusallık varsayımı, çoklu regresyon incelemesi yeni veri setine uygulanmalıdır. Aykırı değer temizleme işlemi 7 kez tekrarlanmıştır.

TEMİZLENEN VERİ SETİ:

- Normallik Varsayımı

```
> data <- read.csv("C:/Users/teknosa/Desktop/REGRESYON/odev/data.txt", sep="")
> view(data)
> names(data)<-c("Kalite", "Aroma", "Lezzet", "Gövde", "Bölge")
> names(data)
[1] "Kalite" "Aroma" "Lezzet" "Gövde" "Bölge"
> bölge<-as.factor(Bölge)
>
> qqnorm(data$kalite,col="#7570B3",pch=15)
> qqline(data$kalite,col="cornflowerblue",lwd=2)
>
```



Çizilen Q-Q plot grafiğinde şarap kalitesine ilişkin verilerin normal dağıldığı söylenebilir fakat Anderson Darling normallik testi ile kontrol edilmelidir.

H_0 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Şarap kalitesi verilerinin dağılımı ile normal dağılım arasında fark vardır.

```
> ad.test(data$kalite)

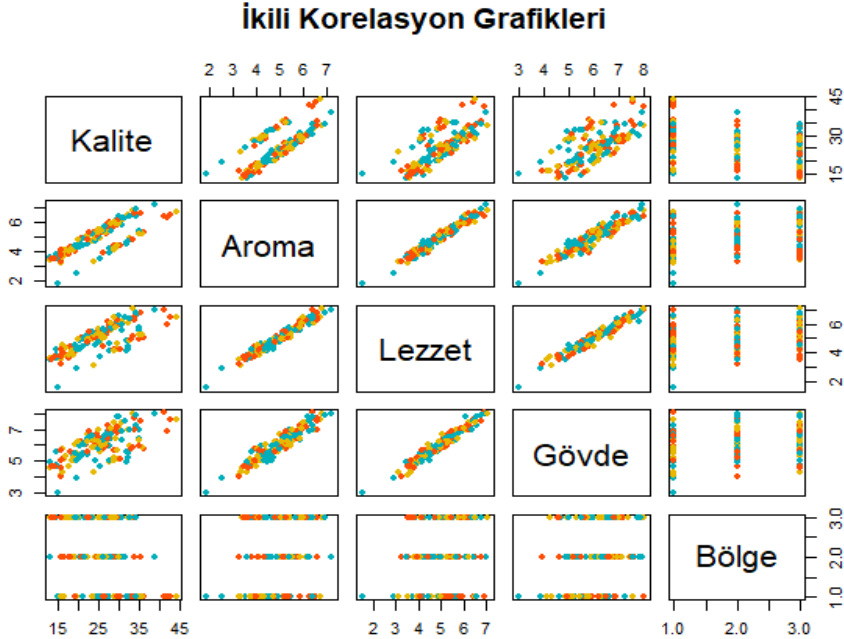
Anderson-Darling normality test

data: data$kalite
A = 0.37851, p-value = 0.4022
```

%95 güven düzeyinde $p\text{-value} > 0.05$ olduğundan yokluk hipotezimiz reddedilemez. Yani, şarap kalitesi verilerimizin normal dağılıma uyuma istatistiksel olarak anlamlıdır.

- Doğrusallık Varsayımı

```
my_cols <- c("#00AFBB", "#E7B800", "#FC4E07")
pairs(data, pch = 19, cex = 0.8, main="İkili Korelasyon Grafikleri",
      col = my_cols)
```



Artıklardan temizlenmiş veri setindeki bağımlı değişken kalite için doğrusallık incelendiğinde bağımsız değişkenler ile doğrusal ilişkisi olduğu yorumu yapılabilir.

- Y(Kalite) ile x1(Aroma) arasında pozitif yönlü doğrusal ilişki vardır.
- Y(Kalite) ile x2(Lezzet) arasında güçlü doğrusal ilişki vardır.
- Y(Kalite) ile x3(Gövde) arasında doğrusal ilişki vardır.
- Nitel veri olan bölge değişkeni göstermelik değişken olduğundan doğrusallık incelemesi yapılamaz.

5. Çoklu Regresyon Model Denklemini

```
> levels(bölge)<-c("1","2","3")
> sonuc<-lm(Kalite~Aroma+Lezzet+Gövde+bölge)
> summary(sonuc)

Call:
lm(formula = Kalite ~ Aroma + Lezzet + Gövde + bölge)

Residuals:
    Min       1Q   Median       3Q      Max
-6.9817 -0.1663  0.2084  0.8480  3.2857

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.9918     1.5012   3.325  0.00114 **
Aroma         7.2709     0.9906   7.340 1.91e-11 ***
Lezzet        1.2905     1.3526   0.954  0.34179
Gövde        -2.7672     0.9385  -2.949  0.00377 **
bölge2       -6.9005     0.5261 -13.116 < 2e-16 ***
bölge3       -8.3937     0.5140 -16.331 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.433 on 133 degrees of freedom
Multiple R-squared:  0.8684,    Adjusted R-squared:  0.8635
F-statistic: 175.6 on 5 and 133 DF,  p-value: < 2.2e-16
```

Bağımlı değişken şarap kalitesi olmak üzere şarapların aroması, lezzeti, gövdesi ve kullanılan üzümün yetiştirildiği bölge bağımsız değişkendir. 1.bölge olan Burgonya kılavuz değişken olarak alınmıştır.

Kestirim Denklemini :

$Kalite = 4.9918 + 7.2709Aroma + 1.2905Lezzet - 2.7672Gövde - 6.9005Bölge2 - 8.3937Bölge3 \pm 2.433$

(1.5012) (0.9906) (1.3526) (0.9385) (0.5261) (0.5140)

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< α =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.

6. Regresyon Katsayıları

β_0 için:

$H_0: \beta_0=0$

$H_1: \beta_0 \neq 0$

p-value = 0.00114 < α olduğundan dolayı H_0 reddedilir. Sabit terimin modele katkısı %5 yanılma düzeyinde önemlidir.

Yorum: Diğer değişkenlerin 0 olması anlamlı olmadığı için denklem gereği bir katsayıdır.

β_1 için:

$H_0: \beta_1=0$

$H_1: \beta_1 \neq 0$

p-value = 1.91e-11 < α olduğundan dolayı H_0 reddedilir. Şarap aromasının modele katkısı istatistiksel olarak anlamlıdır.

Yorum: Diğer değişkenlerin etkisi sabit tutulduğunda şarap aromasındaki 1 br'lik artış şarap kalitesini ortalama 7.2709 br artırır.

β_2 için:

$H_0: \beta_2=0$

$H_1: \beta_2 \neq 0$

p-value = 0.34179 > α olduğundan dolayı H_0 reddedilmez. Şarap lezzetinin modele katkısı istatistiksel olarak anlamlı değildir.

Yorum: Bu değişken anlamlı olmadığı için yorum yapamayız.

β_3 için:

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

p-value $0.00377 < \alpha$ olduğundan dolayı H_0 reddedilir. Gövdenin modele katkısı istatistiksel olarak anlamlıdır.

Yorum: Diğer değişkenlerin etkisi sabit tutulduğunda gövdedeki 1 br'lik artış şarap kalitesini ortalama 2.7672 br azaltır.

β_4 için:

$H_0: \beta_4 = 0$

$H_1: \beta_4 \neq 0$

p-value $2e-16 < \alpha$ olduğundan dolayı H_0 reddedilir. Burgonya bölgesinde(bölge1) yetişen Pinot Noir üzümleri ile Willamette Vadisinde(bölge2) yetişenler arasında istatistiksel olarak fark vardır. Şarap kalitesi bakımından Burgonya'da yetişen üzümlerin Willamette Vadisindekilere göre daha etkili olduğunu söyleyebiliriz.

Yorum: Bölge Willamette ise şarap kalitesi ortalama 6.9005 br azalır.

β_5 için:

$H_0: \beta_5 = 0$

$H_1: \beta_5 \neq 0$

p-value $2e-16 < \alpha$ olduğundan dolayı H_0 reddedilir. Burgonya bölgesinde(bölge1) yetişen Pinot Noir üzümleri ile Martinborough'da (bölge3) yetişenler arasında istatistiksel olarak fark vardır. Şarap kalitesi bakımından Burgonya'da yetişen üzümlerin Martinborough'a göre daha etkili olduğunu söyleyebiliriz.

Yorum: Bölge Martinborough ise şarap kalitesi ortalama 8.3937 br azalır.

7. Belirtme Katsayısı

$R^2 = 0.8684$ ve düzeltilmiş belirtme katsayısı $= 0.8635$ değerleri birbirlerine yakın olduğu için çoklu belirtme katsayısı kullanılır. Şarap kalitesindeki değişimin %86.84'ü bağımsız değişkenlerimiz tarafından açıklanabilmektedir. Aykırı değer, uç değer ve etkin değerlerin çıkarılmasından sonra modelin başarısı artmıştır

8. Güven Aralıkları

```
> confint(sonuc)
```

	2.5 %	97.5 %
(Intercept)	2.022453	7.9611150
Aroma	5.311535	9.2303392
Lezzet	-1.384966	3.9658662
Gövde	-4.623487	-0.9109304
bölge2	-7.941145	-5.8598469
bölge3	-9.410274	-7.3770577

β_0 : $P(2.022453 < \beta_0 < 7.9611150) = 0.95$

Güven aralığı 0'ı içermiyor. Denklem gereği katsayının önemli olduğu %95 güven düzeyinde söylenebilir.

β_1 (Aroma): $P(5.311535 < \beta_1 < 9.2303392) = 0.95$

%95 güvenle diğer değişkenlerin etkisi sabitken aromadaki 1 birimlik artış şarap kalitesini ortalama 5.311535 ile 9.2303392 birim arasında artırır.

β_2 (Lezzet): $P(-1.384966 < \beta_2 < 3.9658662) = 0.95$ Güven aralığının 0'ı içeriyor olması Lezzet bağımsız değişkenin modele katkısı anlamsızdır yorumumuzu desteklemektedir.

β_3 (Gövde): $P(-4.623487 < \beta_3 < -0.9109304) = 0.95$

%95 güven düzeyinde diğer değişkenlerin etkisi sabitken gövdedeki 1 birimlik artış şarap kalitesini ortalama -4.623487 ile -0.9109304 birim arasında azaltır.

β_4 (bölge2): $P(-7.941145 < \beta_4 < -5.8598469) = 0.95$

%95 güven düzeyinde bölge 2 olan Willamette Vadisinde yetiştirilen üzümler Burgonya'da yetiştirilen üzümlere göre şarap kalitesini daha az arttığı yorumunu her iki ucunun negatif olmasından yararlanarak yapabiliriz.

β_5 (bölge3): $P(-9.410274 < \beta_5 < -7.3770577) = 0.95$

%95 güven düzeyinde Martinborough'da yetiştirilen üzümler Burgonya'da yetiştirilen üzümlere göre şarap kalitesini daha az arttırır. En güzel üzümler Burgonya'da yetişmektedir.

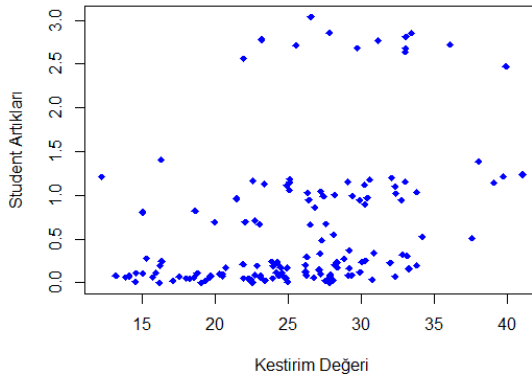
9. Değişen Varyanslılık Sorunu

H_0 : Varyanslar homojendir. ($\alpha=0.05$)

H_1 : Varyanslar homojen değildir.

- **Grafik Yöntemi**

```
> plot(predict(sonuc), abs(inf$stud.res), ylab="Student Artıkları", xlab="Kestirim Değeri", col="blue", pch=18)
```



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olduğunu görürüz. Yapının rasgele olması değişen varyanslılık sorununun olduğuna dair ipucu verir.

- **Breusch ve Pagan Testi**

```
> bptest(sonuc)

studentized Breusch-Pagan test

data: sonuc
BP = 59.052, df = 5, p-value = 1.908e-11
```

$p\text{-value} < \alpha = 0.05$ olduğundan %95 güvenle yokluk hipotezi reddedilir. Değişen varyanslılık sorunu vardır.

- **Model Oluşturma Yöntemi**

Değişen varyanslılığın testi için diğer bir yaklaşım artıklarla kestirim değerleri arasında bir model yazılıp yazılamadığının araştırılmasıdır.

Ho: Model anlamsızdır.

Hs: Model anlamlıdır.

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

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

Residuals:
    Min       1Q   Median       3Q      Max
-2.0965 -1.1672 -0.4934  0.5655  5.3782

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.49272    0.62780  -2.378   0.0188 *
fitted(sonuc)  0.11664    0.02378   4.906  2.6e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

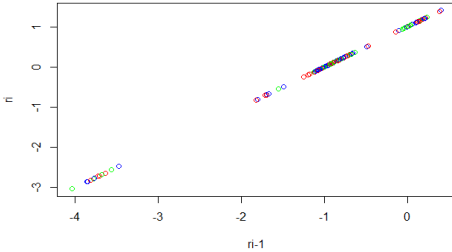
Residual standard error: 1.714 on 137 degrees of freedom
Multiple R-squared:  0.1494,    Adjusted R-squared:  0.1432
F-statistic: 24.07 on 1 and 137 DF,  p-value: 2.597e-06
```

p-value< α =0.05 olduğundan %95 güvenle yokluk hipotezi reddedilir. Model anlamlıdır. Artıklarla kestirim değerleri arasında oluşturulan model anlamlı olduğu için değişen varyanslılık vardır.

10. Öz İlişki Sorunu

Öz ilişki incelemesi yapılırken grafik ve test yöntemi kullanılır. $n \geq 15$ olduğu durumlarda Durbin-Watson, $n < 15$ olduğu durumlarda Von Neuman testi kullanılır.

Grafik Yöntemi: Artıklara karşılık gelen (Artık-1) artığının çizimi rasgelelik dışında bir görünüm taşıyorsa hatalar ilişkilidir. Grafiği incelersek rasgelelik dışında bir desen olduğunu görürüz. Öz ilişki sorunundan bahsedebiliriz.



Test Yöntemi:

Ho: Öz ilişki yoktur.

Hs: Öz ilişki vardır (Pozitif öz ilişki).

```
> dwtest(sonuc)

Durbin-Watson test

data: sonuc
Dw = 0.19817, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Durbin-Watson test istatistiği (0.19817) üzerinden inceleme yapılmaktadır. Alternatif hipotez kurulmasında bu değer önemlidir. $p\text{-value} < \alpha = 0.05$ olduğundan %95 güvenle yokluk hipotezi reddedilir. Öz ilişki sorunu vardır. d değeri için verilen p değeri incelendiğinde (+) yönlü bir özilişki olduğu görülmektedir.

11.Çoklu Bağlantı Sorunu

Çoklu bağlantı (multicollinearity) kavramı, model varsayımlarından bağımsız değişkenlerin ilişkili olmamaları varsayımının bozulması sorunudur. Bağımsız değişkenlerin yüksek ilişkili olmasıdır.

- **Korelasyon İncelemesi**

Çoklu bağlantının ilk göstergeleri basit ilişki katsayılarıdır. İki bağımsız değişken arasındaki basit ilişki katsayısının 0.75'ten büyük olması çoklu bağlantı şüphesi yaratır.

```
> cor(data)
```

	Kalite	Aroma	Lezzet	Gövde	Bölge
Kalite	1.0000000	0.7451392	0.7234063	0.6706258	-0.3808135
Aroma	0.7451392	1.0000000	0.9758015	0.9500291	0.1661607
Lezzet	0.7234063	0.9758015	1.0000000	0.9738143	0.1535553
Gövde	0.6706258	0.9500291	0.9738143	1.0000000	0.1631706
Bölge	-0.3808135	0.1661607	0.1535553	0.1631706	1.0000000

Not: Bölge değişkenimiz nitel olduğu için o sütundaki korelasyon katsayıları yorumlanmaz.

Lezzet ve Aroma arasındaki ilişki katsayısının 0.975, Gövde ve Aroma arasındaki 0.95, Lezzet ve Gövde arasındaki 0.973 olması çoklu bağlantı olduğunu düşündürür.

- **Varyans Şişme Katsayısı(VIF Değeri) İncelemesi**

```
install.packages("DAAG")
library(DAAG) # car kütüphanesi aktif ise nitel katsayıları vermiyor.
detach("package:car", unload=TRUE) #car paketinin kütüphaneden de-aktif yapıyor.
#car paketini yeniden yükleyelim
install.packages("car")
library(car)

> vif(sonuc)
```

Aroma	Lezzet	Gövde	bölge2	bölge3
21.1670	40.2050	19.7150	1.4387	1.4019

Bağımsız değişkenlere(Aroma,Lezzet,Gövde) ait VIF değerlerinin çok büyük olmasından dolayı çoklu bağlantıdan şüpheleniyoruz. Çoklu bağlantıdan en çok Lezzet değişkeni etkilenmektedir.

Not:VIF değeri>10 olduğu durumlarda çoklu bağlantından şüphelenilir. VIF değeri=1 olduğu durumda ise çoklu bağlantı yoktur.

Nitel veriler için VIF:

```
> ols_vif_tol(sonuc)
Variables Tolerance VIF
1 Aroma 0.04724405 21.166686
2 Lezzet 0.02487269 40.204735
3 Gövde 0.05072309 19.714889
4 bölge2 0.69506012 1.438724
5 bölge3 0.71331638 1.401902
```

Nitel verilere ait VIF değerleri 10'dan küçük olduğu için çoklu bağlantı içinde yer almazlar. VIF Değerleri dışında Tolerance değerlerinin 0'a yakın olması güçlü çoklu bağlantı olduğunu söyler. Aroma ,Lezzet ve Gövde değişkenlerinin çoklu bağlantı içinde olduğunu düşünerüz.

- **Koşul Sayısı**

Özdeğerler kullanılarak elde edilen bir ölçüttür.

```
> ols_eigen_cindex(sonuc)
Eigenvalue condition Index intercept Aroma Lezzet Gövde
1 4.7117871867 1.000000 8.266925e-04 7.556592e-05 4.003264e-05 5.819890e-05
2 1.0000231923 2.170640 3.133253e-07 3.274218e-09 1.484391e-10 6.558135e-09
3 0.2614914150 4.244867 3.046468e-03 2.621909e-04 1.492040e-04 2.015989e-04
4 0.0246487668 13.825958 5.601247e-01 8.354963e-03 4.664028e-03 1.122110e-03
5 0.0014949377 56.141148 1.727574e-01 5.548283e-01 8.194181e-03 4.246234e-01
6 0.0005545015 92.181008 2.632445e-01 4.364790e-01 9.869526e-01 5.739946e-01
bölge2 bölge3
1 0.0080942542 0.0083191283
2 0.2352112625 0.2307713199
3 0.7329537613 0.7534144611
4 0.0004418743 0.0000376235
5 0.0001881623 0.0001997247
6 0.0231106853 0.0072577425
```

Eigenvalueler özdeğer olmak üzere Condition Index sütunu koşul sayıları verir.

- Koşul sayısı 30'dan büyük ise ($ksj \geq 30$) güçlü çoklu bağlantı vardır.
- $10 \leq ksj < 30$ ise çoklu bağlantının olduğu ancak çok güçlü olmadığı ifade edilir.
- $ksj < 10$ ise kestirimlere zararlı bir çoklu bağlantı olmadığı yorumlanır.

Koşul sayısı 30'dan büyük olan 2 tane değişkenimiz olduğu için 2 tane güçlü çoklu bağlantı olduğu yorumunu yapabiliriz. Çoklu bağlantının yaratmış olduğu sorundan Aroma, Lezzet ve Gövde bağımsız değişkenlerinin etkilendiği görülmektedir.

- Özdeğer ve Özvektör

Kareler toplamına göre standartlaştırılmış $X'X$ matrisi ilişki matrisini verir.

```
install.packages("fastDummies")
library(fastDummies)
dummy<-dummy_cols(bölge)
x41<-dummy$.data_1
x42<-dummy$.data_2
x43<-dummy$.data_3

#standartlaştırma işlemleri (nitel düzeylerden biri kılavuz değişken olacağı için alınmaz, burada 1. düzey alınmıştır ancak değiştirilebilir)
ort1<-mean(Aroma)
kt1<-sum((Aroma-ort1)^2)
skx1<- (Aroma-ort1)/(kt1^0.5)
ort2<-mean(Lezzet)
kt2<-sum((Lezzet-ort2)^2)
skx2<- (Lezzet-ort2)/(kt2^0.5)
ort3<-mean(Gövde)
kt3<-sum((Gövde-ort3)^2)
skx3<- (Gövde-ort3)/(kt3^0.5)
ort42<-mean(x42)
kt42<-sum((x42-ort42)^2)
skx42<- (x42-ort42)/(kt42^0.5)
ort43<-mean(x43)
kt43<-sum((x43-ort43)^2)
skx43<- (x43-ort43)/(kt43^0.5) |
x<-cbind(skx1,skx2,skx3,skx42,skx43)
sm<- eigen (t(x)%*%x)
```

```
> signif(sm$values,3)
[1] 2.9600 1.5100 0.4650 0.0499 0.0165
```

Özdeğerlerden sıfıra yakın olan ögeleri doğrusal bağımlı değişkenleri gösterir. 0.0499 ve 0.0165 değerleri 0'a yakındır. Bu da demektir ki **2 tane çoklu bağlantı** vardır. Bu durum koşul sayısında yaptığımız yorumu desteklemektedir.

Bunlara ek olarak, $\sum \frac{1}{\lambda} = 65 > k$ olduğundan çoklu bağlantı vardır diyebiliriz.

```
> signif(sm$vectors,3)
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	-0.5730	0.0248	-0.0491	0.69400	0.4340
[2,]	-0.5770	0.0292	-0.0795	0.02470	-0.8120
[3,]	-0.5730	0.0158	-0.0438	-0.72000	0.3900
[4,]	-0.0993	-0.7000	0.7070	0.00778	-0.0236
[5,]	-0.0412	0.7130	0.6990	-0.00149	-0.0136

Özvektör matrisinden yararlanarak çoklu bağıntı yapılarını yazalım. λ_4 ve λ_5 'e karşılık gelen 4. ve 5. özvektörün büyük ögeleri ilişkili değişkenleri açıklar. O halde doğrusal bağımlılıklar:

$$1.Yapı: 0.69400Aroma^* + 0.02470Lezzet^* \cong 0$$

$$2.Yapı: 0.02470Aroma^* + 0.3900Gövde^* \cong 0$$

12.Uyum Kestirimi

Üretici, Willamette vadisinde yetişen Pinot Noir üzümü ile üretilen; aroması 6.123612, lezzeti 6.213497 ve gövdesi 7.451064 br olan şarabın kalitesini merak etmiştir. Üreticinin istediği değerler veri setimizde olduğu için uyum kestirimi yaparak onun merakını giderebiliriz.

```
> predict(sonuc, data.frame(Aroma=6.123612,Lezzet=6.213497 ,
+                             Gövde=7.451064 , bölge="2"), interval = 'confidence')
      fit      lwr      upr
1 30.01525 29.08015 30.95035
```

Verilen çıktıya göre şarap üreticisi memnun edecek en uygun şarap kalitesi 30.01525'tir.

13.Ön Kestirim

Şarap üreticisi bu sefer de Burgonya'da yetişen Pinot Noir üzümü ile üretilen; aroması 5.872,lezzeti 7.2521 ve gövdesi 4.3262 br olan şarabın kalitesini merak etmiştir. . Üreticinin istediği değerler veri setimizde olmadığı için ön kestirim yaparak onun merakını giderebiliriz.

```
> predict(sonuc, data.frame(Aroma=5.872,Lezzet=7.2521 ,
+                             Gövde=4.3262 , bölge="1"), interval = 'confidence')
      fit      lwr      upr
1 45.0737 37.37105 52.77636
```

Verilen çıktıya göre şarap üreticisi memnun edecek en uygun şarap kalitesi 45.0737'dir.

14.Uyum Kestirimi ve Ön Kestirim için Güven Aralıkları

Uyum Kestirimi için: $P(30.01525 < E(\hat{y}_i) < 30.95035) = 0.95$

%95 güven düzeyinde Willamette vadisinde yetişen Pinot Noir üzümü ile üretilen; aroması 6.123612, lezzeti 6.213497 ve gövdesi 7.451064 br olan şarabın kalitesi ortalama 30.01525 ile 30.95035 br arasındadır.

Ön Kestirim için: $P(37.37105 < E(\tilde{y}_i) < 52.77636) = 0.95$

%95 güven düzeyinde Burgonya'da yetişen Pinot Noir üzümü ile üretilen; aroması 5.872,lezzeti 7.2521 ve gövdesi 4.3262 br olan şarabın kalitesi ortalama 37.37105 ile 52.77636 br arasındadır.

15.Değişken Seçimi

1-ileriye Doğru Seçim Yöntemi (Forward Selection):

```
> lm.null <- lm(kalite ~ 1)
> forward <- step(lm.null,kalite~Aroma+Lezzet+Gövde+bölge,
+               direction = "forward")
Start: AIC=525.03
kalite ~ 1

Df Sum of Sq  RSS   AIC
+ Aroma      1    3323.9 2662.6 414.41
+ Lezzet      1    3132.8 2853.7 424.04
+ Gövde       1    2692.4 3294.2 443.99
+ bölge       2     918.8 5067.7 505.87
<none>                5986.5 525.03

Step: AIC=414.41
kalite ~ Aroma

Df Sum of Sq  RSS   AIC
+ bölge      2    1810.27  852.34 260.08
+ Gövde       1     85.37 2577.24 411.88
<none>                2662.61 414.41
+ Lezzet      1     1.72 2660.90 416.32

Step: AIC=260.08
kalite ~ Aroma + bölge

Df Sum of Sq  RSS   AIC
+ Gövde       1    59.413 792.93 252.03
+ Lezzet       1    13.321 839.02 259.89
<none>                852.34 260.08

Step: AIC=252.04
kalite ~ Aroma + bölge + Gövde

Df Sum of Sq  RSS   AIC
<none>                792.93 252.03
+ Lezzet      1    5.3896 787.54 253.09
```

```
> forward
Call:
lm(formula = kalite ~ Aroma + bölge + Gövde)

Coefficients:
(Intercept)      Aroma      bölge2      bölge3      Gövde
      4.388       7.949      -6.970      -8.431      -2.148
```

Bağımlı değişken Kalite olduğu durumda 1.adımda modelde Aroma değişkeni girmiştir. Aroma değişkeninin modele girmesi anlamlı çıktığı için işleme devam edilmiştir. Bölge değişkeni modele eklenmiştir o da anlamlı çıktığı için aynı işlem Gövde için tekrarlanmıştır. Gövde değişkeninin de modele girmesi anlamlı çıkmıştır. En son Lezzet değişkenine sıra geldiği zaman değişkenimiz anlamsız çıkmıştır.

Anlamlılık Kontrolü:

```
> summary(forward)

Call:
lm(formula = Kalite ~ Aroma + bölge + Gövde)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3822 -0.2827  0.2345  1.1245  3.5142

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.3876     1.3606   3.225  0.00158 **
Aroma         7.9488     0.6900  11.520 < 2e-16 ***
bölge2       -6.9700     0.5209 -13.381 < 2e-16 ***
bölge3       -8.4311     0.5123 -16.457 < 2e-16 ***
Gövde        -2.1484     0.6780  -3.169  0.00190 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.433 on 134 degrees of freedom
Multiple R-squared:  0.8675,    Adjusted R-squared:  0.8636
F-statistic: 219.4 on 4 and 134 DF,  p-value: < 2.2e-16
```

summary(forward) fonksiyonu ile oluşan regresyon modelinin katsayılarının anlamlı olduğunu doğruluyoruz.(p-value < α=0.05). p-value = 2.2e-16 < α=0.05 olduğundan dolayı model anlamlıdır. Şarap kalitesindeki değişimim %86.75'i Aroma, Gövde ve Bölge ile açıklanabilmektedir.

O halde ileriye dönük seçim ile en iyi regresyon modeli:

$$Kalite = 4.388 + 7.949Aroma - 2.148Gövde - 6.970Bölge2 - 8.431Bölge2 \pm 2.433$$

2-Geriye Doğru Çıkarma Yöntemi (Backward):

```
> backward<-step(sonuc,direction = "backward")
```

```
Start:  AIC=253.09
```

```
Kalite ~ Aroma + Lezzet + Gövde + bölge
```

	Df	Sum of Sq	RSS	AIC
- Lezzet	1	5.39	792.93	252.04
<none>			787.54	253.09
- Gövde	1	51.48	839.02	259.89
- Aroma	1	319.00	1106.54	298.36
- bölge	2	1742.15	2529.69	411.29

```
Step:  AIC=252.04
```

```
Kalite ~ Aroma + Gövde + bölge
```

	Df	Sum of Sq	RSS	AIC
<none>			792.93	252.04
- Gövde	1	59.41	852.34	260.08
- Aroma	1	785.25	1578.18	345.71
- bölge	2	1784.31	2577.24	411.88

```
> backward
```

```
call:
```

```
lm(formula = Kalite ~ Aroma + Gövde + bölge)
```

```
Coefficients:
```

(Intercept)	Aroma	Gövde	bölge2	bölge3
4.388	7.949	-2.148	-6.970	-8.431

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlıyor. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. İkinci modelde Lezzet değişkeni modelden çıkmıştır. Lezzet değişkeninin modele katkısı önemsiz olduğu için işlem sona ermiştir.


```
> summary(backward)
```

```
Call:
```

```
lm(formula = Kalite ~ Aroma + Gövde + bölge)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-7.3822	-0.2827	0.2345	1.1245	3.5142

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.3876	1.3606	3.225	0.00158	**
Aroma	7.9488	0.6900	11.520	< 2e-16	***
Gövde	-2.1484	0.6780	-3.169	0.00190	**
bölge2	-6.9700	0.5209	-13.381	< 2e-16	***
bölge3	-8.4311	0.5123	-16.457	< 2e-16	***

```
---
```

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

```
Residual standard error: 2.433 on 134 degrees of freedom
```

```
Multiple R-squared:  0.8675,    Adjusted R-squared:  0.8636
```

```
F-statistic: 219.4 on 4 and 134 DF,  p-value: < 2.2e-16
```

p-value = $2.2e-16 < \alpha = 0.05$ olduğundan dolayı model anlamlıdır. Bağımsız değişkenlere ait p-value'lar 0.05'ten küçük olduğu için onlar da anlamlıdır.

O halde geriye dönük çıkarma ile en iyi regresyon modeli:

$$\text{Kalite} = 4.3876 + 7.9488\text{Aroma} - 2.1484\text{Gövde} - 6.970\text{Bölge2} - 8.4311\text{Bölge3} \pm 2.433$$

Not: Modelimizde çoklu bağlantı olduğu için en iyi seçim yöntemi geriye doğru çıkarma yöntemidir.

3.Adımsal Seçim Yöntemi(Stepwise):

```
> library(MASS)
> step.model <- stepAIC(sonuc, direction = "both", trace = FALSE)
> step.model

Call:
lm(formula = Kalite ~ Aroma + Gövde + bölge)

Coefficients:
(Intercept)      Aroma      Gövde      bölge2      bölge3
      4.388      7.949     -2.148     -6.970     -8.431

> summary(step.model)

Call:
lm(formula = Kalite ~ Aroma + Gövde + bölge)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3822 -0.2827  0.2345  1.1245  3.5142

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.3876     1.3606   3.225  0.00158 **
Aroma         7.9488     0.6900  11.520 < 2e-16 ***
Gövde        -2.1484     0.6780   -3.169  0.00190 **
bölge2       -6.9700     0.5209 -13.381 < 2e-16 ***
bölge3       -8.4311     0.5123 -16.457 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.433 on 134 degrees of freedom
Multiple R-squared:  0.8675,    Adjusted R-squared:  0.8636
F-statistic: 219.4 on 4 and 134 DF,  p-value: < 2.2e-16
```

İlk olarak modele Aroma değişkeni alınarak başlamış. Bağımsız değişkenimiz anlamlı olduğundan Gövde değişkenimiz de eklenmiştir o da anlamlı çıkmıştır. Yeni değişkenimizin olduğu yeni modelde Aroma değişkeni hala anlamlı mı kontrolü yapılmıştır. Sonuç anlamlı çıktığı için Bölge değişkenimiz de eklenmiştir. Aynı işlemler tekrar edilmiştir. Sıra Lezzet değişkenine gelince değişkenimiz anlamsız çıkmıştır. Adımlarımız ileri doğru seçim yöntemi gibi sonra ermiştir.

O halde adımsal yöntem ile en iyi regresyon modeli:

$$\text{Kalite} = 4.3876 + 7.9488\text{Aroma} - 2.1484\text{Gövde} - 6.970\text{Bölge2} - 8.4311\text{Bölge2} \pm 2.433$$

Not: Her üç yöntemde de aynı bağımsız değişkenlerin modele katkısı anlamlı çıkmıştır.

16.Ridge Regresyon Modeli

Ridge regresyon modeli çoklu bağlantı sorununu gidermek için yapılır.

```
> library(MASS)
> ridge <- lm.ridge(Kalite~Aroma+Lezzet+Gövde+bölge ,lambda = seq(0,1,0.05))
> matplot(ridge$lambda,t(ridge$coef),type="l",xlab=expression(lambda),ylab=expression(hat(beta)))
> abline(h=0,lwd=3)
> ridge$coef
```

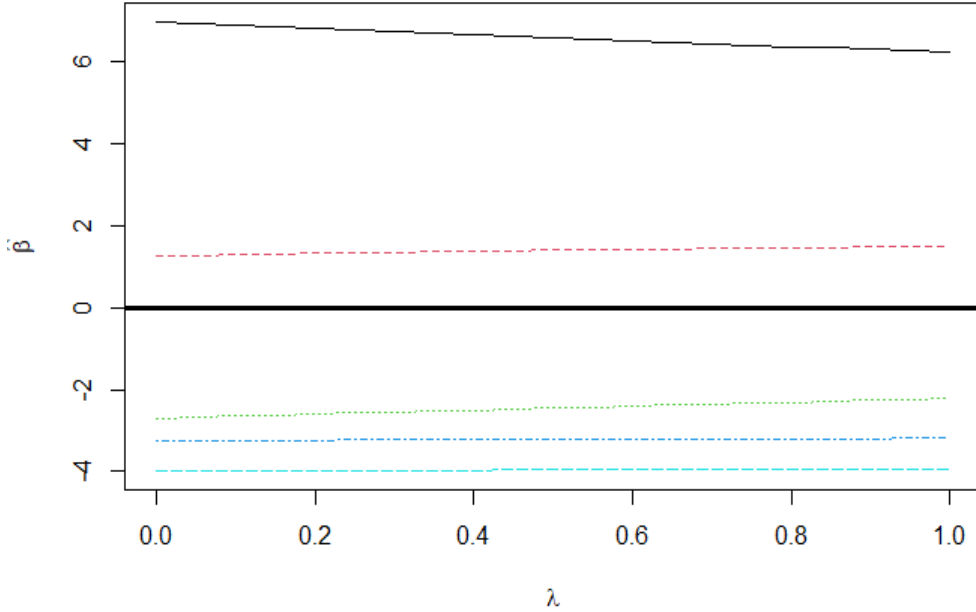
	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Aroma	6.969706	6.925567	6.882328	6.839954	6.798410	6.757664	6.717688
Lezzet	1.248561	1.265698	1.282063	1.297704	1.312666	1.326990	1.340712
Gövde	-2.702198	-2.676649	-2.651218	-2.625917	-2.600757	-2.575749	-2.550900
bölge2	-3.247027	-3.243999	-3.240992	-3.238007	-3.235041	-3.232093	-3.229163
bölge3	-3.990967	-3.987596	-3.984244	-3.980909	-3.977592	-3.974291	-3.971006
	0.35	0.40	0.45	0.50	0.55	0.60	0.65
Aroma	6.678454	6.639935	6.602107	6.564947	6.528432	6.492542	6.457257
Lezzet	1.353869	1.366491	1.378610	1.390251	1.401442	1.412206	1.422565
Gövde	-2.526219	-2.501712	-2.477383	-2.453239	-2.429283	-2.405519	-2.381949
bölge2	-3.226250	-3.223352	-3.220469	-3.217601	-3.214745	-3.211903	-3.209073
bölge3	-3.967736	-3.964481	-3.961240	-3.958012	-3.954797	-3.951596	-3.948406
	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Aroma	6.422558	6.388428	6.354849	6.321807	6.289284	6.257267	6.225742
Lezzet	1.432541	1.442152	1.451416	1.460351	1.468972	1.477294	1.485332
Gövde	-2.358575	-2.335400	-2.312425	-2.289651	-2.267079	-2.244709	-2.222542
bölge2	-3.206255	-3.203448	-3.200652	-3.197867	-3.195091	-3.192325	-3.189568
bölge3	-3.945228	-3.942062	-3.938907	-3.935763	-3.932630	-3.929507	-3.926394

```
> select(ridge)
modified HKB estimator is 0.2117076
modified L-W estimator is 0.4749403
smallest value of GCV at 0.45
```

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. Lamda 0,4 civarında alınabilir.

```
> ridge$coef[,ridge$lam == 0.4]
```

Aroma	Lezzet	Gövde	bölge2	bölge3
6.639935	1.366491	-2.501712	-3.223352	-3.964481



Hızlı artış ya da azalış gösteren katsayılara karşılık gelen değişkenler çoklu bağlantılı değişkenlerdir. Burada Aroma değişkeninin çoklu bağlantıdan en çok etkilenen değişken olduğunu görmekteyiz. Gövde değişkeninin de etkilendiği yorumunu yapabiliriz.

Sıfır eksenin civarında seyreden değişkenler de modelde önemsiz değişkeni göstermektedir.