

RiseInTemp_dataset

IREM KOYUNLU

15 05 2020

ILERI REGRESYON ANALIZI

KURESEL ISINMA (IKLIM DEGISIKLIGI)

KAYNAK:<https://www.kaggle.com/vageeshabudanur/riseintemp-dataset>



```
library(readr)
TempData=read.csv("C:/Users/CASPER/Desktop/climate_change.csv", header=T)

head(TempData,10)
```

##	Year	Month	MEI	CO2	CH4	N2O	CFC.11	CFC.12	TSI	Aerosols
## 1	1983	5	2.556	345.96	1638.59	303.677	191.324	350.113	1366.102	0.0863
## 2	1983	6	2.167	345.52	1633.71	303.746	192.057	351.848	1366.121	0.0794
## 3	1983	7	1.741	344.15	1633.22	303.795	192.818	353.725	1366.285	0.0731
## 4	1983	8	1.130	342.25	1631.35	303.839	193.602	355.633	1366.420	0.0673
## 5	1983	9	0.428	340.17	1648.40	303.901	194.392	357.465	1366.234	0.0619
## 6	1983	10	0.002	340.30	1663.79	303.970	195.171	359.174	1366.059	0.0569
## 7	1983	11	-0.176	341.53	1658.23	304.032	195.921	360.758	1366.107	0.0524
## 8	1983	12	-0.176	343.07	1654.31	304.082	196.609	362.174	1366.061	0.0486
## 9	1984	1	-0.339	344.05	1658.98	304.130	197.219	363.359	1365.426	0.0451
## 10	1984	2	-0.565	344.77	1656.48	304.194	197.759	364.296	1365.662	0.0416
##	Temp									
## 1	0.109									
## 2	0.118									
## 3	0.137									
## 4	0.176									
## 5	0.149									
## 6	0.093									
## 7	0.232									
## 8	0.078									
## 9	0.089									
## 10	0.013									

Veri Seti Aciklamasi ;

Gecen yuzyl boyunca ortalama kuresel sicakligin artmakta oldugunu belgeleyen birçok calisma yapilmistir. Kuresel sicaklikta devam eden artisin sonuclari korkunc olacaktır. Artan deniz seviyeleri ve artan asiri hava olaylari sikligi milyarlarca insani etkileyecektir. Bu veri, ortalama kuresel sicaklik ve diger birçok faktor arasindaki iliskiyi icermektedir. 308 gozlem 11 sutun bulunmakta.

Climate_change.csv dosyasi, Mayıs 1983 ile Aralık 2008 arasindaki iklim verilerini icerir.

CO2, N2O ve CH4 ppmv cinsinden ifade edilir (hacimce milyonda parca sayisi;Yani, 397 ppmv CO2, CO2'nin atmosferin toplam hacminin 397 milyonuncu kismini olusturdugu anlamina gelir). CFC.11 ve CFC.12, ppbv cinsinden ifade edilir (hacim basina milyarda parca).

Degiskenler:

Year (Yil): Verinin gozlem yili.

Month (Ay): Verinin gozlem ayi.

MEI: Çok degiskenli El Nino Guney salinim indeksi (MEI). Gucunun bir olcusu; El Nino / La Nina-Guney salinim (kuresel sicakliklari etkileyen Pasifik Okyanusu'nda bir hava etkisi) Guney Amerika'nin Bati kiyisindaki Buyuk Okyanus'ta yuzey sularinin isinmasina Ispanyolca'da "erkek

cocuk” anlamına gelen El Nino, sogumasina ise “kiz cocuk” anlamına gelen La Nina deniliyor. El Nino, ormanların ve bitki ortusunun karbondioksit emilimini azaltiyor. Bu nedenle daha fazla karbondioksit atmosferde kalıyor bu da kuresel sicakligi etkiliyor.

CO2: Atmosferik karbondioksit konsantrasyonları (CO2). Hava sicak oldugunda karbondioksit miktarinin fazla, soguk oldugunda ise karbondioksit miktarinin az oldugunu bilinmekte.Kuresel sicaklik artisinin en buyuk etkenlerinden biridir.

CH4: Metan (CH4). Karbondioksitten 23 kat daha guclu olan metan gazi, kuresel sicaklik artisinin en onemli unsurlarından biri.

N2O: Nitroz oksit (Azot Protoksit). Nitroz oksit gazi da dogal kaynaklar , suni gubre kullanimi ve endustri gibi insani kaynaklardan olusuyor, kuresel sicaklik artisini etkiliyor.

CFC.11: Trikloroflorometan (CCl3F; yaygin olarak CFC-11 olarak adlandirilir.) Ozon tabakasinin delinmesine yol acan kimyasal gaz (evlerin yalitiminda kullaniliyor). kuresel sicaklik artisini etkiliyor.

CFC.12: Diklorodiflorometan (CCl2F2; yaygin olarak CFC-12 olarak adlandirilir.) Sera etkisi yaratan zararlı gaz,kuresel sicaklik artisini etkiliyor.

TSI:W/m2 cinsinden toplam gunes isinimi (TSI). (gunes enerjisinin birim alan basina biriktirdigi oran). Gunes lekeleri ve diger gunes olaylari nedeniyle, gunes tarafından verilen enerji miktarı zamanla buyuk olcude degisir.

Aerosols: 550 nm’de ortalama stratosferik aerosol optik derinligi. Bu degisken volkanlarla baglantilidir, cunku volkanik patlamalar atmosfere yeni parcacikların eklenmesiyle sonuclanir, bu da gunes enerjisinin ne kadarinin uzaya yansidigini etkiler.

Temp:Bu veride ortalama kuresel sicaklik ve referans degeri arasindaki santigrat derece farki. Bu bagimli degiskendir.

Degiskenleri Aciklamada Kullanilan Ek Bilgi Kaynakları:

<https://www.bbc.com/turkce/haberler-46308251> http://www.imo.org.tr/resimler/ekutuphane/pdf/16154_50_07.pdf

Degisken turleri dogru mu diye kontrol icin ozetine bakalim;

```
summary(TempData)
```

```
##      Year      Month      MEI      CO2
## Min.   :1983 Min.   : 1.000 Min.   : -1.6350 Min.   :340.2
## 1st Qu.:1989 1st Qu.: 4.000 1st Qu.: -0.3987 1st Qu.:353.0
## Median :1996 Median : 7.000 Median : 0.2375 Median :361.7
## Mean   :1996 Mean   : 6.552 Mean   : 0.2756 Mean   :363.2
## 3rd Qu.:2002 3rd Qu.:10.000 3rd Qu.: 0.8305 3rd Qu.:373.5
## Max.   :2008 Max.   :12.000 Max.   : 3.0010 Max.   :388.5
##      CH4      N2O      CFC.11      CFC.12      TSI
## Min.   :1630 Min.   :303.7 Min.   :191.3 Min.   :350.1 Min.   :1365
## 1st Qu.:1722 1st Qu.:308.1 1st Qu.:246.3 1st Qu.:472.4 1st Qu.:1366
## Median :1764 Median :311.5 Median :258.3 Median :528.4 Median :1366
## Mean   :1750 Mean   :312.4 Mean   :252.0 Mean   :497.5 Mean   :1366
## 3rd Qu.:1787 3rd Qu.:317.0 3rd Qu.:267.0 3rd Qu.:540.5 3rd Qu.:1366
## Max.   :1814 Max.   :322.2 Max.   :271.5 Max.   :543.8 Max.   :1367
##      Aerosols      Temp
## Min.   :0.00160 Min.   : -0.2820
## 1st Qu.:0.00280 1st Qu.: 0.1217
## Median :0.00575 Median : 0.2480
## Mean   :0.01666 Mean   : 0.2568
## 3rd Qu.:0.01260 3rd Qu.: 0.4073
## Max.   :0.14940 Max.   : 0.7390
```

Summary kodu ile bakildiginda month duzeltilmeli.Kategorik degisken icin mean hesaplanmaz.

```
TempData$Month<-factor(TempData$Month)
summary(TempData)
```

```
##      Year      Month      MEI      CO2      CH4
## Min.   :1983    5      : 26   Min.   : -1.6350   Min.   :340.2   Min.   :1630
## 1st Qu.:1989    6      : 26   1st Qu.: -0.3987   1st Qu.:353.0   1st Qu.:1722
## Median :1996    7      : 26   Median : 0.2375   Median :361.7   Median :1764
## Mean   :1996    8      : 26   Mean   : 0.2756   Mean   :363.2   Mean   :1750
## 3rd Qu.:2002    9      : 26   3rd Qu.: 0.8305   3rd Qu.:373.5   3rd Qu.:1787
## Max.   :2008   10      : 26   Max.   : 3.0010   Max.   :388.5   Max.   :1814
##
##      (Other):152
##      N2O      CFC.11      CFC.12      TSI
## Min.   :303.7   Min.   :191.3   Min.   :350.1   Min.   :1365
## 1st Qu.:308.1   1st Qu.:246.3   1st Qu.:472.4   1st Qu.:1366
## Median :311.5   Median :258.3   Median :528.4   Median :1366
## Mean   :312.4   Mean   :252.0   Mean   :497.5   Mean   :1366
## 3rd Qu.:317.0   3rd Qu.:267.0   3rd Qu.:540.5   3rd Qu.:1366
## Max.   :322.2   Max.   :271.5   Max.   :543.8   Max.   :1367
##
##      Aerosols      Temp
## Min.   :0.00160   Min.   : -0.2820
## 1st Qu.:0.00280   1st Qu.: 0.1217
## Median :0.00575   Median : 0.2480
## Mean   :0.01666   Mean   : 0.2568
## 3rd Qu.:0.01260   3rd Qu.: 0.4073
## Max.   :0.14940   Max.   : 0.7390
##
```

Modelimizi oluşturalım .Bu verimizde bağımlı değişkenimiz Temp tir(yanıt değişkeni). Year ve month değişkenleri hariç diğer değişkenler açıklayıcı değişkenimiz olacaktır.

```
lmod1 <- lm( Temp~
             MEI+
             CO2+
             CH4+
             N2O+
             CFC.11+
             CFC.12+
             Aerosols+
             TSI ,data=TempData) #Olusturdugum modeli lmod1 diye adlandirdim.
```

Olusturdugum modeli lmod1 diye adlandirdim.

```
head(fitted(lmod1),15) #Head komutu ile sadece 15 tanesini gosterdik.
```

```
##      1      2      3      4      5      6
## 0.16307026 0.14825658 0.14019475 0.11349428 0.04886278 0.01390213
##      7      8      9     10     11     12
## 0.02015506 0.02967373 -0.03122330 -0.01550142 0.08620368 0.05100123
##      13     14     15
## 0.06264959 0.07189982 0.06501410
```

Fitted komutu lm ile kurulan modelin fit edilmiş y sapka değerlerini verir.

```
head(residuals(lmod1),15) #Head komutu ile sadece 15 tanesini gosterdik.
```

```
##      1      2      3      4      5      6
## -0.054070260 -0.030256582 -0.003194751 0.062505724 0.100137219 0.079097870
##      7      8      9     10     11     12
## 0.211844939 0.048326265 0.120223295 0.028501421 -0.037203683 -0.070001228
##      13     14     15
## 0.002350413 -0.087899818 -0.089014096
```

Kurulan regresyon modelinin artıklarını verir.

1- HATA İLE İLGİLİ VARSAYIMLAR

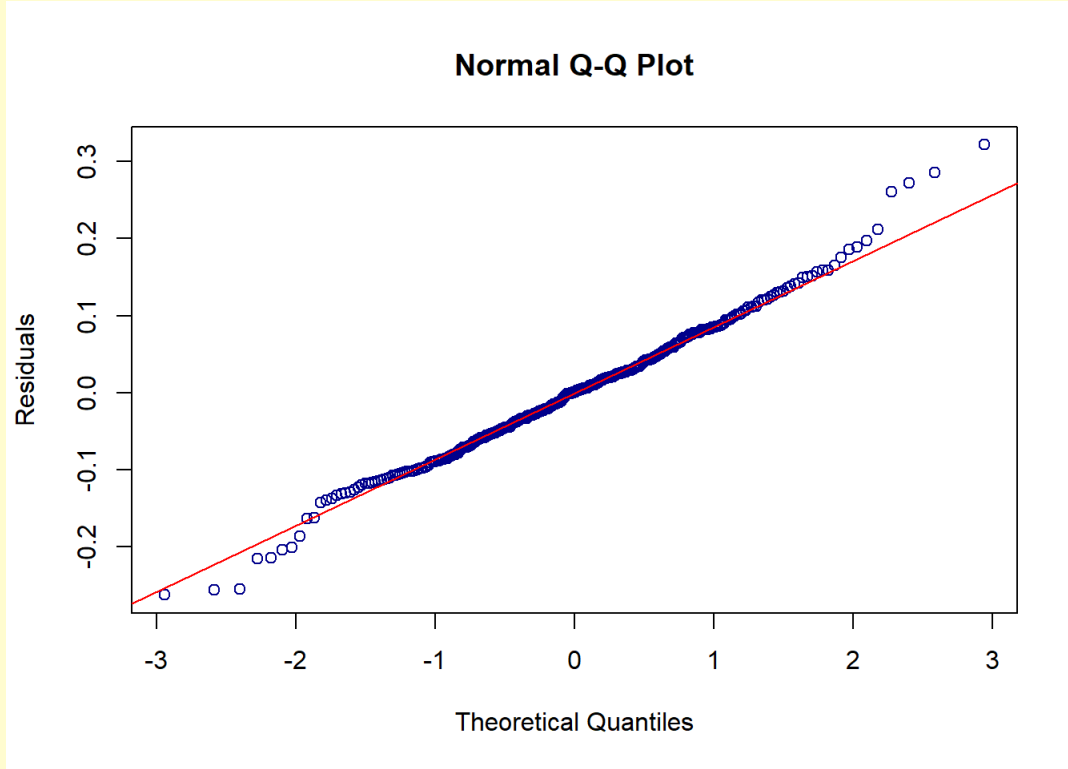
A- NORMALLİK VARSAYIMININ TESTLERİ

Regresyon analizinde kullandığımız tüm hipotez testleri ve güven aralıkları normallik varsayımına dayalıdır.Bu nedenle normallik

varsayiminin kontrolu onemlidir.Normallik varsayimi artiklar uzerinden Q-Q plot yardimi ile inceleyelim;

Onceklikle Q-Q Plot cizdirelim ;

```
qqnorm(residuals(lmod1),ylab="Residuals",col="darkblue")
qqline(residuals(lmod1),col="red")
```



Q-Q Plot grafigi artiklara karsilik cizilir. Grafigi cizdirdigimizde artiklar Q-Q Plot cizgisi uzerinde olur ise verimiz normal dagiliyor demektir. Grafikler yaniltici oldugundan testlere bakilir. Verimiz Q-Q Plot cizgisi uzerinde ciktiigi icin normal dagildigini varsayariz.Grafikler yaniltici oldugundan diger testlere bakmalyiz. Normallik sinamasi icin Shapiro-Wilk Testi (gozlem sayisi 50'nin altindaysa) ve Kolmogorov-Smirnov Testi (gozlem sayisi 50'nin ustundeyse) kullanilir. Simdi testlere bakalim ;Gozlem sayimiz 50 den fazla oldugu icin Kolmogorov-Smirnov Testine bakilmasi gerekir.

KOLMOGOROV SMIRNOV TEST

H0:Artiklar normal dagiliyor. H1:Artiklar normal dagilmiyor.

Normal dagilimin kontrolu icin gelistirilmis bir test. Gozlem sayisi 50 nin uzerinde ise kullanilir.

```
ks.test(residuals(lmod1),"pnorm",
        mean=mean(residuals(lmod1)),
        sd=sd(residuals(lmod1)))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: residuals(lmod1)
## D = 0.033681, p-value = 0.8758
## alternative hypothesis: two-sided
```

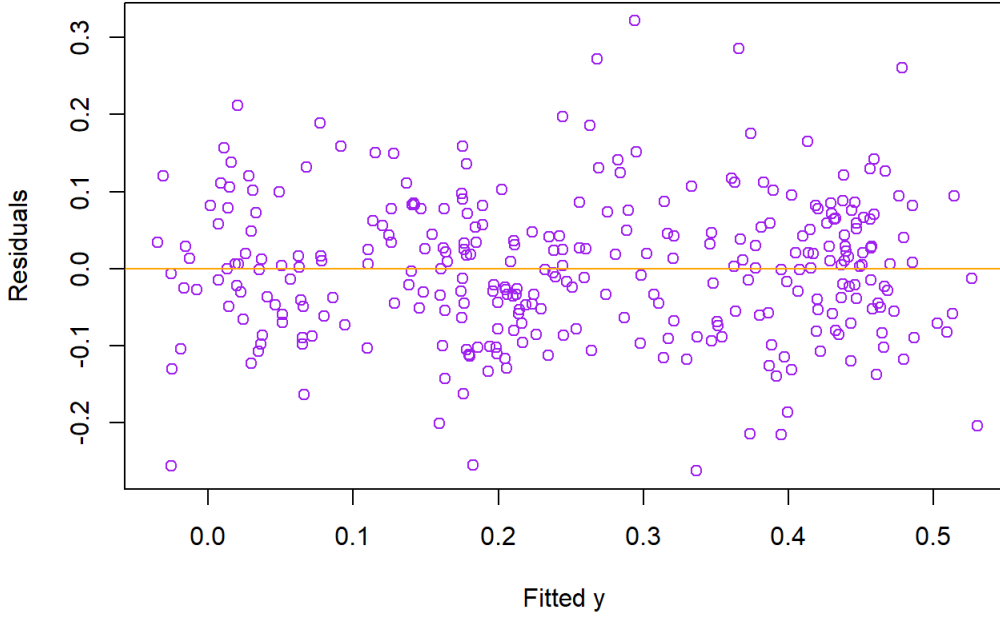
Veri setimde gozlem sayim 50 nin uzerinde oldugundan Kolmogorov Smirnov Testi yapilir.Test sonucumda p-value=0.8758 cikti. 0.05 ' den buyuk oldugu icin H0 hipotezi kabul edilir yani artiklarin dagilimi normaldir.

B- SABIT VARYANS

Modelimin oncelikle plotuna bakiyorum; Sabit varyansi incelemek icin y sapkaya karsi artiklarin(e) grafigini cizdirmeliyiz;

```
plot( fitted(lmod1), residuals(lmod1),xlab="Fitted y ",ylab="Residuals",main="Artiklar-Tahmin Grafigi",col="purple")
abline(h=0,col="orange")
```

Artiklar-Tahmin Grafigi



0 etrafında nasıl dağıldığını görmek için $h=0$ ile yataya çizgi ekledik. Düzgün bir grafik çıkmadığı için sabit varyanslı mı emin olamıyoruz. Güvenilir olması için değişken varyanslılık testlerine de bakmalıyız.

- DEĞİSKEN VARYANS TESTLERİ

BREUSCH PAGAN TEST

H0:Heteroscedasticity (Değişken Varyanslılık) problemi yok. H1:Heteroscedasticity (Değişken Varyanslılık) problemi vardır.

```
library(lmtest)
bptest(lmod1, data=TempData)
```

```
##
## studentized Breusch-Pagan test
##
## data: lmod1
## BP = 7.444, df = 8, p-value = 0.4896
```

Breusch Pagan Test sonucunda $p\text{-value}=0.4896$ çıktı. 0.05 'den büyük olduğu için H_0 hipotezi kabul edilir yani HETEROSCEDASTICITY(değişken varyanslılık) problemi yoktur.

WHITE TEST

```
wmod<-lm(residuals(lmod1)^2~fitted(lmod1)+fitted(lmod1)^2, TempData)
summary(wmod)
```

```
##
## Call:
## lm(formula = residuals(lmod1)^2 ~ fitted(lmod1) + fitted(lmod1)^2,
##     data = TempData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.008283 -0.007459 -0.004823  0.001987  0.095287
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0080401  0.0014953   5.377 1.51e-07 ***
## fitted(lmod1) 0.0005614  0.0049923   0.112   0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01351 on 306 degrees of freedom
## Multiple R-squared:  4.133e-05, Adjusted R-squared: -0.003227
## F-statistic: 0.01265 on 1 and 306 DF, p-value: 0.9105
```

White Test sonucunda p-value= 0.9105 çıktı. 0.05 ' den büyük olduğu için H0 hipotezi kabul edilir yani HETEROSCE DOSTICITY(degisken varyanslilik) problemi yoktur.Breusch Pagan Test sonucu ,White Testten daha güvenilirdir.

2- AYKIRI GOZLEMLERIN BELIRLENMESI (OLAGANDISI GOZLEMLER BULMA)

3 farklı tip aykırı gözlem vardır;

Outlier;Hataları büyük olan gözlemlere denir.(Model tarafından iyi tahmin edilemez.) Influential;Parametreleri (betaları) değiştirir. Leverage; x uzayının merkezine uzak olan gözlemlere denir.

A)LEVERAGE

Leverage tespiti için hat matrisinin köşegen elemanları kullanılır. Herhangi bir gözlemin elemanı cut point değerinden büyükse o gözleme Leverage şüphesiyle yaklaşılar.

Önce regresyon modeli kurulur hat valuesle değerler hesaplanır; n: Gözlem sayısı. p: Degisken sayısı+1, Cut point değerim; $2p/n$. Cut point değerimden büyük olan hat matrisinin köşegen elemanı leverage adayı demektir.

```
hatv <- hatvalues(lmod1) #sapka matrisinin köşegen elemanları
head(hatv,15)
```

```
##      1      2      3      4      5      6      7
## 0.08867497 0.07791777 0.07136349 0.06936140 0.06045417 0.06358753 0.05525247
##      8      9     10     11     12     13     14
## 0.04947510 0.05329869 0.04864966 0.04431911 0.04177722 0.03935979 0.04815078
##      15
## 0.05079924
```

```
sum(hatv)
```

```
## [1] 9
```

cut point değerim $2p/n$ (p: degisken sayısı+1 ,n:gozlem sayısı) den büyükse o gözlem leverage olabilir diye düşünüyoruz. Kurduğum regresyon modelinde 8 degiskenim olduğu için $p:8+1=9$ (Beta0 ile birlikte) tane gözlemim var.

Cut point değerim : $2p/n$ olduğu için. $2.p=2.9=18$

```
cut <- 18/length(TempData$Temp)
cut
```

```
## [1] 0.05844156
```

Cut point değerim 0.05844156 çıktı.

```
which(hatv>cut)
```

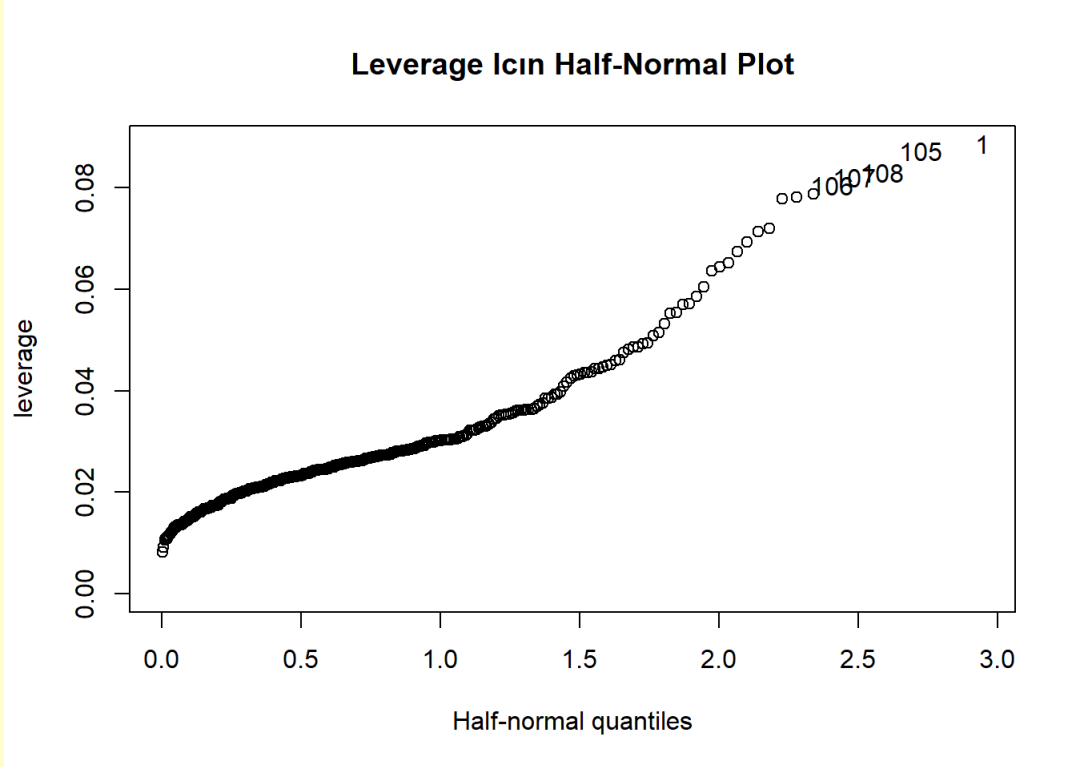
```
## 1 2 3 4 5 6 103 104 105 106 107 108 109 110 112 113 173
## 1 2 3 4 5 6 103 104 105 106 107 108 109 110 112 113 173
```

Cut Pointimden büyük olan hat matrisinin köşegen elemanları, Leverage adaylarıdır. 1,2,3,4,5,6,103,104,105,106,107,108,109,110,112,113 ve 173. gözlemler leverage adaylarıdır.

Half-Normal Plot ile görselleştirme yapalım;

HALF-NORMAL PLOT

```
library(faraway)
halfnorm(hatv,5,ylab="leverage",main="Leverage İçin Half-Normal Plot") # sıra dışı gözlemlerimin 5 ini gösterdim
```



Hat matrisi için çizdiğimiz half-normal plot grafiğinde sadece en aykırı 5 gözleminin gösterilmesini istedik. Cut point sonucumuz ile Half-Normal grafiğinde görülen sonuçlar aynı çıkmıştır.

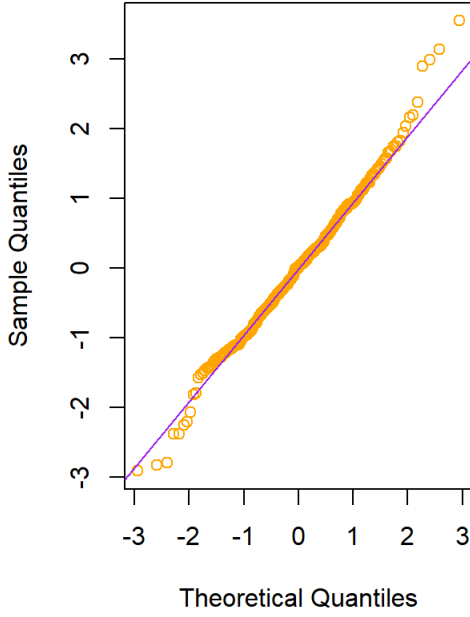
VARSAYIM BOZUNUMU KONTROLÜ

Verimizde varsayım bozunumu yok ise ham artıklar üzerinden çizdirilen Q-Q Plot ile standartlaştırılmış artıklar üzerinden çizdirilmiş olan Q-Q Plot görsel olarak birbirlerine benzemeli.

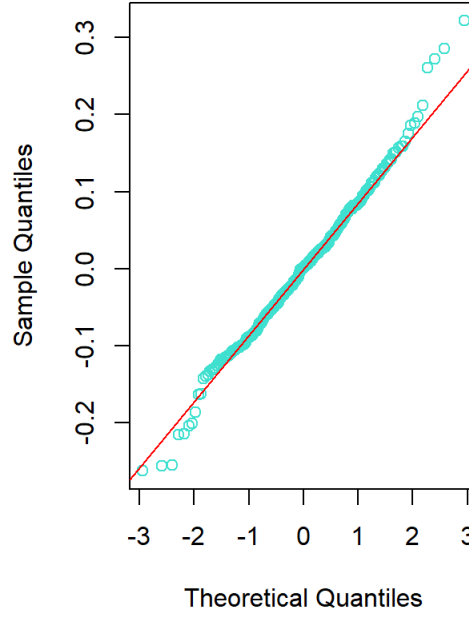
```
par(mfrow=c(1,2)) #iki grafiğe yan yana gösterir
qqnorm(rstandard(lmod1),main="Stand. Artıkların Q-Q Plotu",col="orange") #standartlaştırılmış artıkların qq-plot u
qqline(rstandard(lmod1),col="purple")

qqnorm(residuals(lmod1),main="Ham Artıkların Q-Q Plotu",col="turquoise") #ham artıkların qq-plot u
qqline(residuals(lmod1),col="red")
```

Stand. Artikların Q-Q Plotu



Ham Artikların Q-Q Plotu



Grafiklerde de gözüktüğü üzere varsayım bozunumu yoktur. İki grafikte de değişkenlerin çoğu çizginin üstünde çıkmıştır ve birbirine benzemektedir.

B) OUTLIER

STUDENTIZED ARTIKLAR BENFERRONI TESTİ

```
stud <- rstudent(lmod1)
stud[which.max(abs(stud))] ]
```

```
##      184
## 3.614831
```

Standartlaştırılmış artıklarımın mutlaka en büyüğüne bakıyoruz. En büyük standartlaştırılmış artık (rstudent) değerim 184. gözlem olup değeri 3.614831 dir.

Cut point belirlemeliyiz, cut point değeri $\alpha/2n$ dir. Rstudentler $(n-p-1)$ serbestlik dereceli t-dağılımına sahiptir.

```
qt(0.05/(length(stud)*2) , (length(TempData$Temp)-9-1)) # (alfa/2n) , (n-p-1)
```

```
## [1] -3.820115
```

Mutlakca cut point değerinden büyük olan standartlaştırılmış artık değeri bizim için outlier olmaktadır. Verimde en büyük mutlak standartlaştırılmış artık değerini 184. gözlem = 3.614831 olarak bulduk. 184. gözlem = $|3.614831| < |3.820115|$ olduğundan 184. gözlem bizim için outlier değildir.

Outlier rstudent tipi artıklar için halfnorm çizdirmeye gerek yok.

C) INFLUENTIAL

Influential Parametreleri (betaları) değiştirir. Bir etkili gözlem aynı zamanda bir outlier veya leverage olabilir. Etkinliği ölçmek için cook-distance yöntemini kullanmalıyız.

COOK-DISTANCE

```
cook <- cooks.distance(lmod1)
head(cook, 15)
```



```
##           1           2           3           4           5           6
## 4.114053e-03 1.105695e-03 1.113158e-05 4.123764e-03 9.050653e-03 5.979512e-03
##           7           8           9          10          11          12
## 3.661465e-02 1.685484e-03 1.132834e-02 5.754802e-04 8.851882e-04 2.938438e-03
##          13          14          15
## 3.105404e-06 5.411798e-03 5.887856e-03
```

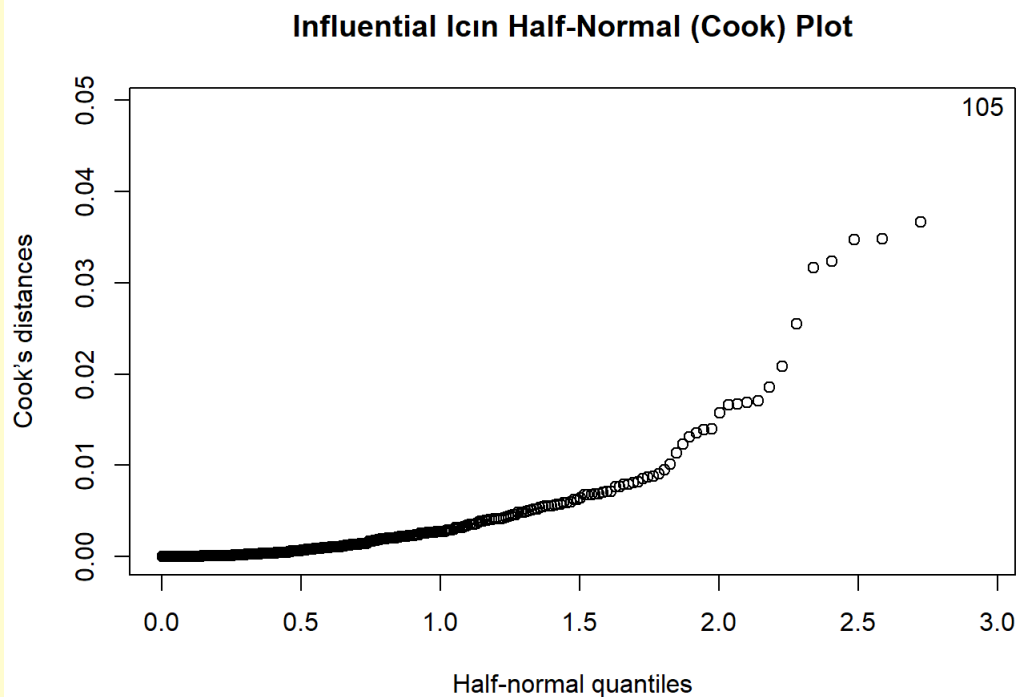
```
max(cook)
```

```
## [1] 0.04936055
```

En buyuk cook distance degerim 0.04936055 dir ve bu deger 105. gozlemde vardır.

HALF-NORMAL PLOT

```
halfnorm(cook,1,ylab="Cook's distances",main="Influential Icin Half-Normal (Cook) Plot")
```



```
cook[105]
```

```
##           105
## 0.04936055
```

Cook distance icin cizdirdigimiz half-normal plot grafiginde en etkili gozlemimin gosterilmesini istedik.En buyuk cook distance degerim 105.gozleme aittir. Cook Distance sonucum ile half-normal plot grafigindeki sonucum ayni ciktigindan 105. gozlemim etkili gozlemdir.

Bu gozlemi veriden cikardigimizda modelde ki degisimin nasil olacagini inceleyelim;

```
lmod2 <- lm(Temp~
             MEI+
             CO2+
             CH4+
             N2O+
             CFC.11+
             CFC.12+
             Aerosols+
             TSI ,data=TempData, subset = cook < max(cook))
```

lki model arasi degisime bakarsak;

```
summary(lmod1)
```

```
##
## Call:
## lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
##     Aerosols + TSI, data = TempData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26228 -0.05868  0.00051  0.05718  0.32170
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.277e+02  1.919e+01  -6.654 1.36e-10 ***
## MEI          6.632e-02  6.186e-03  10.722 < 2e-16 ***
## CO2          5.207e-03  2.192e-03   2.375  0.0182 *
## CH4          6.371e-05  4.977e-04   0.128  0.8982
## N2O         -1.693e-02  7.835e-03  -2.161  0.0315 *
## CFC.11       -7.278e-03  1.461e-03  -4.980 1.07e-06 ***
## CFC.12       4.272e-03  8.763e-04   4.875 1.77e-06 ***
## Aerosols    -1.582e+00  2.099e-01  -7.535 5.86e-13 ***
## TSI          9.586e-02  1.401e-02   6.844 4.38e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09182 on 299 degrees of freedom
## Multiple R-squared:  0.744, Adjusted R-squared:  0.7371
## F-statistic: 108.6 on 8 and 299 DF,  p-value: < 2.2e-16
```

```
summary(lmod2)
```

```
##
## Call:
## lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
##     Aerosols + TSI, data = TempData, subset = cook < max(cook))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26006 -0.05951  0.00090  0.05723  0.31874
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.259e+02  1.909e+01  -6.593 1.96e-10 ***
## MEI          6.606e-02  6.149e-03  10.744 < 2e-16 ***
## CO2          5.267e-03  2.179e-03   2.417  0.0162 *
## CH4         -1.310e-06  4.956e-04  -0.003  0.9979
## N2O         -1.797e-02  7.802e-03  -2.304  0.0219 *
## CFC.11      -7.517e-03  1.457e-03  -5.161 4.49e-07 ***
## CFC.12      4.447e-03  8.746e-04   5.085 6.52e-07 ***
## Aerosols    -1.690e+00  2.145e-01  -7.878 6.29e-14 ***
## TSI          9.482e-02  1.393e-02   6.807 5.50e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09125 on 298 degrees of freedom
## Multiple R-squared:  0.748, Adjusted R-squared:  0.7412
## F-statistic: 110.5 on 8 and 298 DF,  p-value: < 2.2e-16
```

İki model için summary kodunu kullandığımızda, modeller arasında değişimler vardır. Adjusted R-squared değerimiz az da olsa artmıştır. CH4 değişkeninin katsayısında gözle görülür bir değişim olmuştur. Tahminlerin tek bir gözleme bu kadar hassas olması istenmeyen bir durumdur.

Bu amaçla dfbeta değerine bakmalıyız.

Dfbeta her bir değişken için ayrı ayrı dfbeta değeri veriyor.

```
dfbeta <- dfbeta(lmod1)
head(dfbeta, 5)
```

```
##      (Intercept)      MEI      CO2      CH4      N2O
## 1  0.70766159 -4.413596e-04 -1.229765e-04 -5.039192e-06  6.670995e-04
## 2  0.37573414 -1.916761e-04 -5.346524e-05  3.905179e-06  3.171070e-04
## 3  0.05268937 -1.544568e-05 -2.751617e-06  7.040433e-07  2.673505e-05
## 4 -1.18165430  1.485248e-04 -2.271089e-05 -2.320427e-05 -3.438887e-04
## 5 -1.06768933 -2.957679e-05 -1.525495e-04  5.951604e-06 -4.849339e-04
##      CFC.11      CFC.12      Aerosols      TSI
## 1  1.377566e-04 -5.740004e-05 -0.0146299794 -6.358033e-04
## 2  7.437055e-05 -3.495304e-05 -0.0073017061 -3.392874e-04
## 3  7.955493e-06 -3.935016e-06 -0.0006713937 -4.488509e-05
## 4 -1.590302e-04  8.511891e-05  0.0118507352  9.777218e-04
## 5 -2.684273e-04  1.244021e-04  0.0195977105  9.296338e-04
```

dfbeta için cut point değeri hesaplamalıyız. cut point: $2/\sqrt{n}$

```
cut <- 2/sqrt(length(TempData$Temp))
cut
```

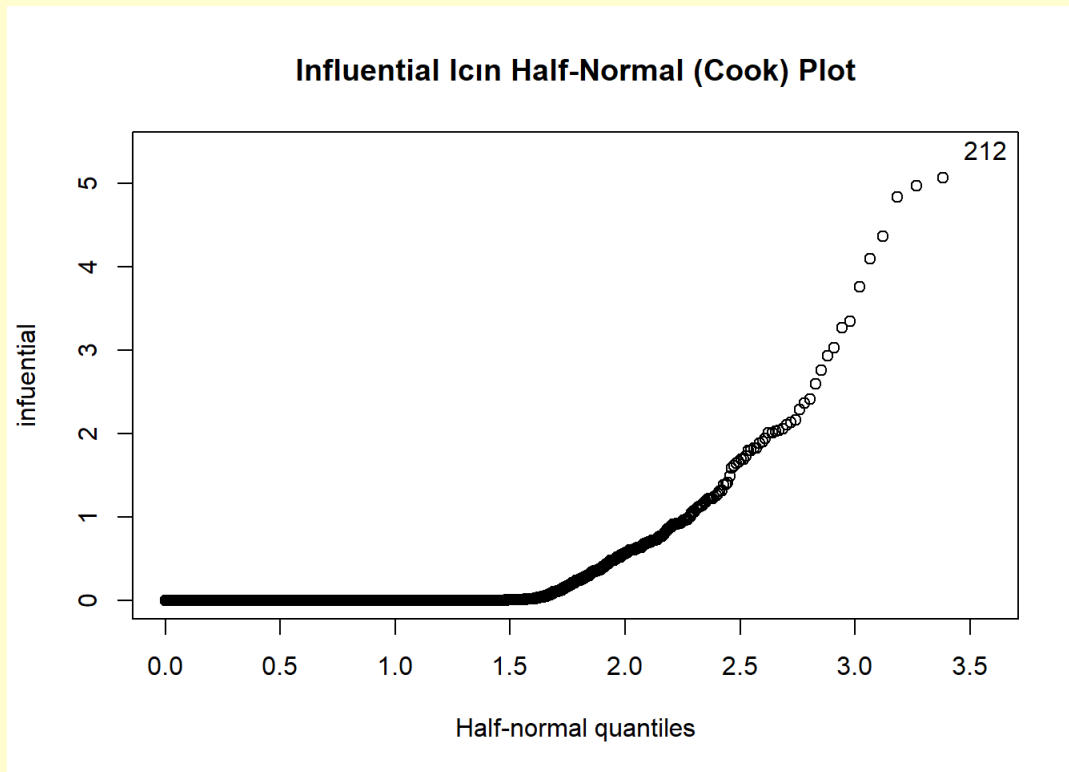
```
## [1] 0.1139606
```

Cut point değerim 0.1139606 çıktı. Cut pointten büyük olan dfbeta değerleri bizim için etkili gözlemdir.

dfbeta için half normal plot'a çizdirirsek;

HALF-NORMAL PLOT

```
halfnorm(dfbeta,1,ylab="influential",main="Influential İçin Half-Normal (Cook) Plot")
```

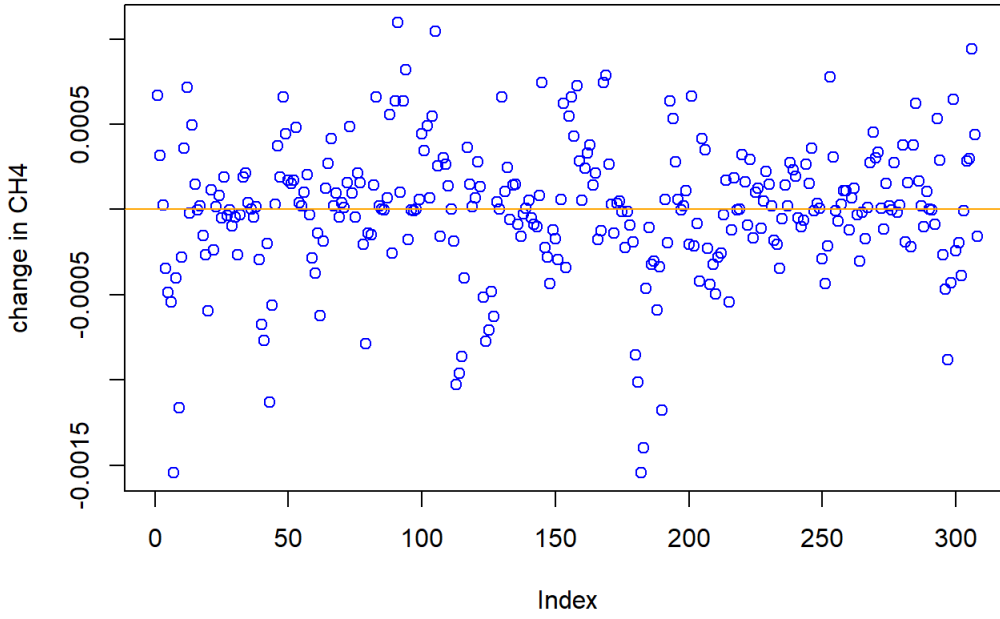


Dfbeta için half normal plot çizdirdiğimizde 212. gözlemimiz etkili gözlemdir. Diğer gözlemlerde etkilidir ama sadece en etkili olanı gösterdik.

Dfbeta da her bir değişkenin etkisine ayrı ayrı plot çizdirerek bakmalıyız;

```
plot(dfbeta(lmod1)[,5],ylab = "change in CH4",main="CH4 İçin Plot",col="blue") # CH4 in model üzerine etkisi ne bakıyoruz.
abline(h=0,col="orange")
```

CH4 İcin Plot



Dfbeta kullanımı zordur. Burada dfbetaya göre modelde CH4 için etkinliğe bakıyoruz. Dfbeta da her bir değişkenin değişimini ayrı ayrı bakmak gerekiyor bu Dfit kullanalım;

Dffit : Betalar arasındaki fark yerine tahminler arasındaki farka dayalı bir ölçüttür.

Dffit için cut point= $2 \cdot \sqrt{p/n}$ Cut point değerinden büyük olanlar etkili gözlemdir denir.

```
dffit <- dffits(lmod1)
head(dffit,15)
```

```
##      1      2      3      4      5      6
## -0.192222963 -0.099608653 -0.009992479  0.192487549  0.285532194  0.231901186
##      7      8      9     10     11     12
##  0.578565284  0.123017783  0.319739400  0.071859227 -0.089132562 -0.162514829
##     13     14     15
##  0.005277805 -0.220681000 -0.230193284
```

Dffit için cut point değerim;

```
cut <- 2*sqrt(9/length(TempData$Temp))
cut
```

```
## [1] 0.3418817
```

Cut point değerim 0.3418817 çıktı. Cut point değerimden büyük olan gözlemlere bakalım;

```
which(abs(dffit)>cut)
```

```
##  7  20  63  83 105 106 115 127 178 183 184 190 210 211 212 236 297 299
##  7  20  63  83 105 106 115 127 178 183 184 190 210 211 212 236 297 299
```

Dffit için cut pointi asan etkili gözlemleri gösterdik. 7,20,63,83,105,106,115,127,178,183,184,190,210,211,212,236,297 ve 299. gözlemlerim etkili gözlemdir.

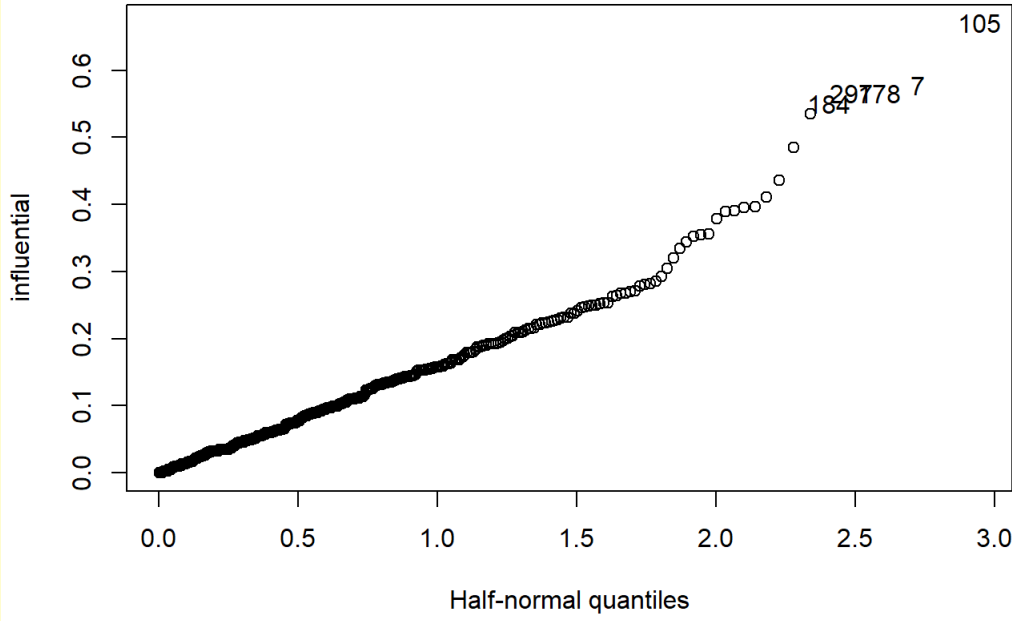
Half normal plot çizdirelim;

HALF-NORMAL PLOT

Dffit için half normal plot'a bakarsak;

```
halfnorm(dffit,5,ylab="influential",main="Influential İcin Half-Normal (Dffit) Plot")
```

Influential Icin Half-Normal (Dffit) Plot



Dffit icin half normal plot uzzerinden en etkili 5 gozleme baktim. Diger gozlemlerde etkilidir ama sadece en etkili 5 tanesini gosterdik.

3)ACIKLAYICI DEGIsKENLERLE ILGILI PROBLEMLER

A)IC ILISKI (COLLINEARITY)

lki degisken arasi lineer iliskiyi gosterir. Eger bir aciklayici degisken ve bir diger aciklayici degiskenin veya degiskenlerin lineer bir kombinasyonlari ise bu durumda x transpoz x matrisi ($X'X$) singular olur ve tersi alinamaz. Bu durumdan ilgili degiskenlerden biri modelden cikartilarak kurtulunur. Gozlem sayisi artikca ic iliski durumu azalir.

```
lmod1 <- lm( Temp~
  MEI+
  CO2+
  CH4+
  N2O+
  CFC.11+
  CFC.12+
  Aerosols+
  TSI ,data=TempData)

summary(lmod1)
```

```
##
## Call:
## lm(formula = Temp ~ MEI + CO2 + CH4 + N2O + CFC.11 + CFC.12 +
##     Aerosols + TSI, data = TempData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26228 -0.05868  0.00051  0.05718  0.32170
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.277e+02  1.919e+01  -6.654 1.36e-10 ***
## MEI          6.632e-02  6.186e-03  10.722 < 2e-16 ***
## CO2          5.207e-03  2.192e-03   2.375  0.0182 *
## CH4          6.371e-05  4.977e-04   0.128  0.8982
## N2O         -1.693e-02  7.835e-03  -2.161  0.0315 *
## CFC.11       -7.278e-03  1.461e-03  -4.980 1.07e-06 ***
## CFC.12       4.272e-03  8.763e-04   4.875 1.77e-06 ***
## Aerosols    -1.582e+00  2.099e-01  -7.535 5.86e-13 ***
## TSI          9.586e-02  1.401e-02   6.844 4.38e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09182 on 299 degrees of freedom
## Multiple R-squared:  0.744, Adjusted R-squared:  0.7371
## F-statistic: 108.6 on 8 and 299 DF,  p-value: < 2.2e-16
```

Regresyon modelimiz için p-value değeri yaklaşık=0 yani 0.05 den küçük olduğundan H_0 hipotezi reddedilir yani kurulan model anlamlıdır. Fakat model anlamlı iken açıklayıcı değişkenlerimizden CH4 değişkeninin p-value değeri 0.05 den büyüktür yani değişken anlamsızdır.

x in korelasyon matrisine bakalım;

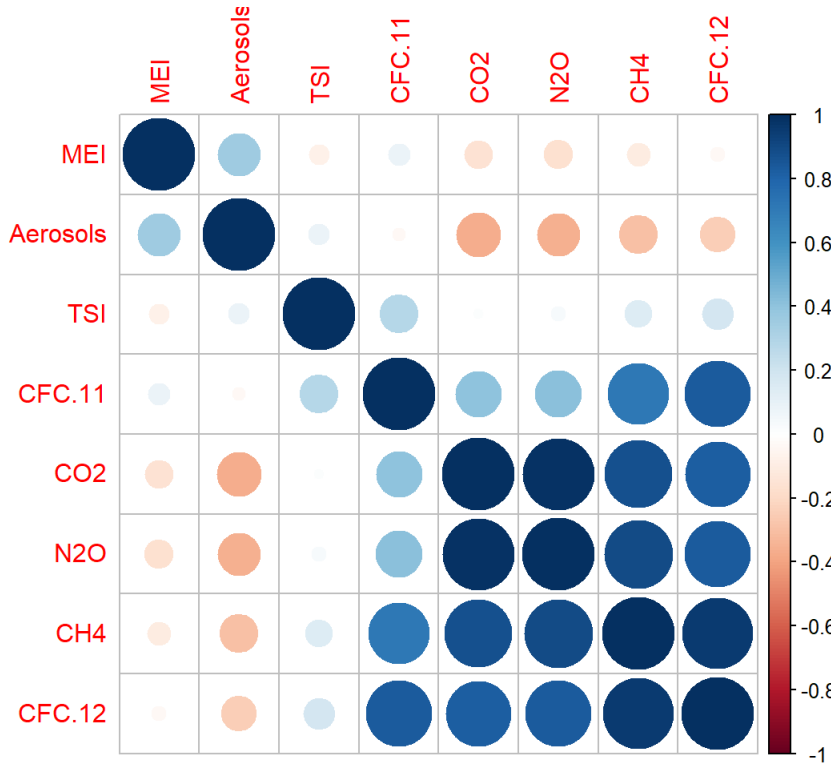
```
cor(TempData[, -c(1,2,11)])
```

```
##           MEI          CO2          CH4          N2O          CFC.11          CFC.12
## MEI      1.00000000 -0.15291104 -0.1055547 -0.16237531  0.08817074 -0.03983567
## CO2     -0.15291104  1.00000000  0.8722531  0.98113544  0.40128447  0.82321031
## CH4     -0.10555472  0.87225311  1.00000000  0.89440921  0.71350408  0.95823718
## N2O     -0.16237531  0.98113544  0.8944092  1.00000000  0.41215475  0.83929545
## CFC.11   0.08817074  0.40128447  0.7135041  0.41215475  1.00000000  0.83138131
## CFC.12  -0.03983567  0.82321031  0.9582372  0.83929545  0.83138131  1.00000000
## TSI     -0.07682560  0.01786672  0.1463349  0.03989183  0.28462884  0.18927009
## Aerosols 0.35235073 -0.36926514 -0.2903814 -0.35349882 -0.03230227 -0.24378508
##           TSI          Aerosols
## MEI     -0.07682560  0.35235073
## CO2      0.01786672 -0.36926514
## CH4      0.14633495 -0.29038142
## N2O      0.03989183 -0.35349882
## CFC.11   0.28462884 -0.03230227
## CFC.12   0.18927009 -0.24378508
## TSI      1.00000000  0.08323812
## Aerosols 0.08323812  1.00000000
```

Görüldüğü üzere bağımsız değişkenler arasındaki ilişkinin korelasyonu en yüksek 0.98113544 N2O ve CO2 arasında çıkmıştır ,bağımsız değişkenimizle ilişkilidir.

Şimdi Korelasyon plotuna bakalım;

```
library(corrplot)
corrplot(cor(TempData[, -c(1,2,11)]), method = "circle", order="hclust")
```



Yukarıdaki korelasyon grafiğinde Pozitif korelasyonlar mavi, negatif korelasyonlar kırmızı renkte görülmektedir. Örneğin N2o ve CO2 arasındaki ilişkinin korelasyonu yüksek çıkmıştır. Fakat CO2 ile MEI arasındaki ilişkinin korelasyonu düşüktür.

Verimizi matrix haline donusturelim ;

```
x <- model.matrix(lmod1)[,-11]
head(x, 5)
```

```
##      (Intercept)    MEI      CO2      CH4      N2O      CFC.11      CFC.12      Aerosols      TSI
## 1      1 2.556 345.96 1638.59 303.677 191.324 350.113    0.0863 1366.102
## 2      1 2.167 345.52 1633.71 303.746 192.057 351.848    0.0794 1366.121
## 3      1 1.741 344.15 1633.22 303.795 192.818 353.725    0.0731 1366.285
## 4      1 1.130 342.25 1631.35 303.839 193.602 355.633    0.0673 1366.420
## 5      1 0.428 340.17 1648.40 303.901 194.392 357.465    0.0619 1366.234
```

Verimizden yanıt degiskenini cikarip sadece aciklayici degiskenlerden olusan matrix formuna donusturduk.

x transpoz x in eigen valuelerini hesaplayalım;

```
e <- eigen(t(x)%*%x)$values
e
```

```
## [1] 1.685102e+09 1.002420e+06 6.284630e+04 1.365592e+04 2.009339e+03
## [6] 2.519922e+02 1.303370e+02 1.937460e-01 2.289110e-05
```

KOSUL INDEKSI

Kapa degeri : $\sqrt{\text{en büyük özdeğer} / \text{en küçük özdeğer}}$

```
k <- sqrt(max(e)/min(e))
k
```

```
## [1] 8579852
```

Kapa degeri > 30 ise orta derece collinearity (ic iliski) , Kapa degeri > 100 ise guclu collinearity (ic iliski) oldugunu gosterir. Verimizde cikan sonucumuza gore guclu collinearity (ic iliski) problemi vardır. Bagimsiz degiskeni diger degiskenler çok fazla etkiliyor.

VIF

Vif degerlerine bakmak icin car paketi icinde olan hazır kod kullanilir.

```
library(car)
vif(lmod1)
```

##	MEI	CO2	CH4	N2O	CFC.11	CFC.12	Aerosols	TSI
##	1.225696	27.996103	19.129508	61.037453	31.829321	93.498182	1.354470	1.140976

Vif degerinin 10 dan buyuk olmasi collinearity probleminin oldugunu soyler. CO2,CH4,N2O,CFC.11,CFC.12 degiskenlerimin degerleri 10 dan buyuk cikti collinearity(ic iliski) problemi vardır.

Hata ile ilgili varsayimlardan Otokorelasyon testlerini farkli bir veri üzerinden inceleyelim;

1- HATA ILE ILGILI VARSAYIMLAR

C- OTOKORELASYON

Hatalar ile bir kaydirilmis hatalar arasindaki otokorelasyondur.Korelasyon iki tane degisken vektor arasinda bakilir. Zaman serilerinde ,mekansal verilerde ve panel(kesit) verilerinde bu durum incelenmistir. Onceden fitted y lere karsilik grafik cizdiriyorduk simdi modelimizdeki zaman degiskenine karsi cizdirecegiz.

National Stock Exchange : Time Series

KAYNAK: <https://www.kaggle.com/atulanandjha/national-stock-exchange-time-series>



```
NSTS=read.csv("C:/Users/CASPER/Desktop/infy_stock.csv", header=T)
head(NSTS,10)
```

##	Date	Symbol	Series	Prev.Close	Open	High	Low	Last	Close
## 1	2015-01-01	INFY	EQ	1972.55	1968.95	1982.00	1956.90	1971.00	1974.40
## 2	2015-01-02	INFY	EQ	1974.40	1972.00	2019.05	1972.00	2017.95	2013.20
## 3	2015-01-05	INFY	EQ	2013.20	2009.90	2030.00	1977.50	1996.00	1995.90
## 4	2015-01-06	INFY	EQ	1995.90	1980.00	1985.00	1934.10	1965.10	1954.20
## 5	2015-01-07	INFY	EQ	1954.20	1965.00	1974.75	1950.00	1966.05	1963.55
## 6	2015-01-08	INFY	EQ	1963.55	1985.60	1997.00	1950.00	1979.25	1973.45
## 7	2015-01-09	INFY	EQ	1973.45	1980.10	2109.00	1913.05	2075.30	2074.45
## 8	2015-01-12	INFY	EQ	2074.45	2092.00	2119.20	2075.00	2112.95	2115.95
## 9	2015-01-13	INFY	EQ	2115.95	2107.80	2107.80	2075.00	2092.00	2088.90
## 10	2015-01-14	INFY	EQ	2088.90	2098.50	2133.00	2092.60	2129.00	2128.65
##	VWAP	Volume	Turnover	Trades	Deliverable	Volume	X.Deliverble		
## 1	1971.34	500691	9.870306e+13	14908		258080	0.5154		
## 2	2003.25	1694580	3.394669e+14	54166		1249104	0.7371		
## 3	2004.59	2484256	4.979911e+14	82694		1830962	0.7370		
## 4	1954.82	2416829	4.724458e+14	108209		1772070	0.7332		
## 5	1962.59	1812479	3.557162e+14	62463		1317720	0.7270		
## 6	1972.78	3391230	6.690160e+14	92752		2686012	0.7920		
## 7	2037.69	11215832	2.285439e+15	359214		3369489	0.3004		
## 8	2099.40	3189722	6.696516e+14	107209		1818800	0.5702		
## 9	2089.42	2200309	4.597374e+14	66676		1385009	0.6295		
## 10	2110.88	2480315	5.235638e+14	53263		1832958	0.7390		

Veri Seti Aciklamasi ;

Ulusal Borsa: Zaman Serisi.Zaman serisi analizi icin Hint bilisim sirketlerinin ulusal borsa veri setidir. Hindistan Ulusal Borsasi (NSE) Mumbai, Maharashtra, Hindistan'da bulunan bir Hint borsasidir. Ulusal Menkul Kiymetler Borsasi (NSE) 1992 yilinda yetkisiz bir elektronik borsa olarak kuruldu. Hindistan Hukümeti'nin talebi uzerine onde gelen finans kuruluslari tarafından tesvik edildi. Hindistan'in ciro ile yaptigi en buyuk borsa. 1994 yilinda elektronik ekran tabanlı ticareti baslattı. Daha sonra, 2000 yilinda ulkede turunun ilk ornegi olan endeks futures ve internet ticareti baslattı. 248 gozlem 15 sutun bulunmakta.

Degiskenler:

Date: Verilerin kaydedildiği tarih.

Symbol: Stokun NSE (Hindistan Ulusal Borsası) sembolü.

Series: Bu hisse senedinin serisi. (EQ,BE,BL,BT,GC,IL)

Prev Close: Son gün kapanış noktası.

Open: Mevcut gün açılış noktası.

High: Mevcut günün en yüksek noktası.

Low: Mevcut günün en düşük noktası.

Last: Son işlem gününde belirli bir hisse senedi veya borsa endeksi için son teklif edilen işlem fiyatı.

Close: Geçerli gün için kapanış noktası.

VWAP: Hacim ağırlıklı ortalama fiyat anlamına gelir, bir varlığın belirli bir zaman aralığı için hacme göre ağırlığı alınmış ortalama fiyatıdır.

Volume: Belirli bir zaman diliminde işlem gören menkul kıymet tutarı. Her alıcı için bir satıcı var ve her işlem toplam hacim sayısına katkıda bulunuyor.

Turnover: Hisse senedinin o güne kadar toplam cirosu.

Trades: Hisse senedi alım veya satım sayısı.

Deliverable: Bir grup insandan (bugünden önce demat hesabında bu hisseleri olan ve bugün satış yapan) başka bir gruba insana (bu hisseleri satın almış olan ve bu hisseleri T+2 günlerine kadar alacak olan) hareket eden hisse miktarı. Demat hesabı(Hindistan Internet Borsası)).

%Deliverble: Bu hisse senedinin teslimat yüzdesi.

VWAP'yi özellikle güçlü bir göstergesi haline getiren ortalama fiyat hesaplamasında hacmi kullanma şeklidir. VWAP, hacmin gücüyle fiyat hareketini birleştirerek pratik ve kullanımı kolay bir göstergesi yaratır. Alım satım yapanlar VWAP'yi bir trend onaylama ya da giriş ve çıkış noktalarını belirleme aracı olarak kullanabilir. VWAP her günün başında sıfırlanır. Alım satım yapmak istediğimizde, VWAP'in altında almak ve üstünde satmak karlıdır. Fiyatın bugünkü değerinin üstünde mi alım yaptık yoksa altında mı alım yaptık bunu belirlememizi sağlar. Fiyat VWAP üzerinde ise, satmak için iyi bir gün içi fiyattır. Fiyat VWAP'in altındaysa, satın almak için iyi bir gün içi fiyatıdır.

Degiskenleri Aciklamada Kullanilan Ek Bilgi Kaynaklari:

<https://www.investopedia.com/articles/trading/11/trading-with-vwap-mvwmap.asp> <https://www.binance.vision/tr/economics/volume-weighted-average-price-vwap-explained>

Regresyon modelimizi kuralim;

VWAP, hacmin gücüyle fiyat hareketini birleştirerek pratik ve kullanımı kolay bir göstergesi yaratır. Bağımlı değişkenimiz VWAP tir.(Yanıt değişkeni) Açıklayıcı değişken olarak da Volume,High,Low,Close,Last kullanarak regresyon modeli kuralım, zaman serisi için year değişkenini modele dahil etmiyoruz ;

```
model<-lm(VWAP~Volume+
          High+
          Low+
          Close+
          Last,data=NSTS)
```

```
summary(model)
```

```
##
## Call:
## lm(formula = VWAP ~ Volume + High + Low + Close + Last, data = NSTS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.7176  -1.8093  -0.0132   1.9931  23.8569
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.983e+00  1.282e+00   1.547   0.1232
## Volume      -1.505e-07  1.993e-07  -0.755   0.4509
## High         3.824e-01  2.165e-02  17.666 < 2e-16 ***
## Low          2.517e-01  2.317e-02  10.861 < 2e-16 ***
## Close        5.419e-01  8.214e-02   6.597 2.62e-10 ***
## Last        -1.785e-01  8.325e-02  -2.144   0.0331 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.374 on 242 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 7.22e+05 on 5 and 242 DF, p-value: < 2.2e-16
```

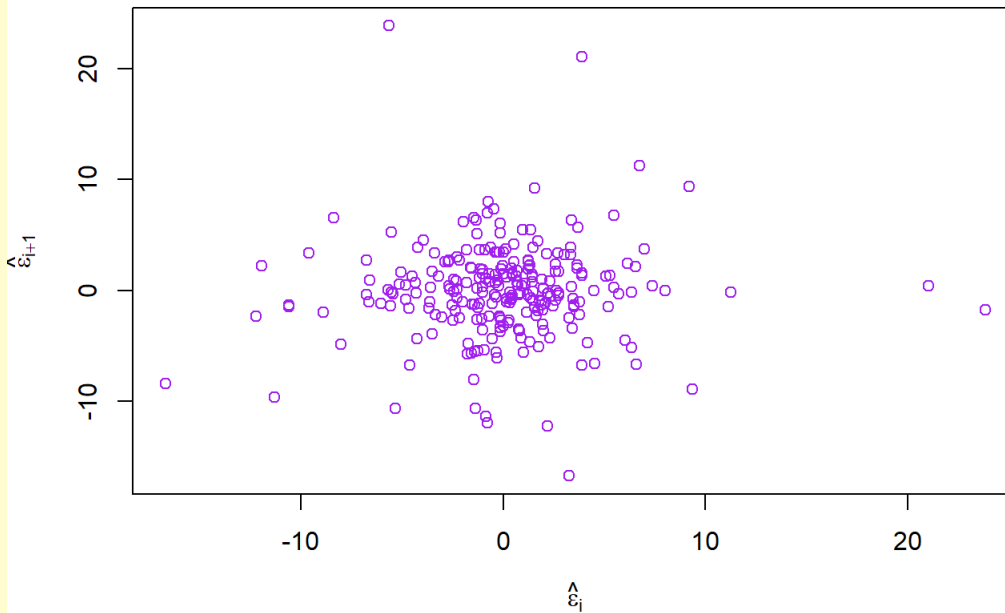
Regresyon modelimiz için p-value değerimiz yaklaşık olarak 0'dır, 0.05'ten küçük olduğu için H_0 hipotezimiz reddedilir yani kurulan regresyon modeli anlamlıdır. Kullandığımız regresyon modelinde VWAP'i en çok etkileyen değişkenleri kullandığımız için Adjusted R-squared değerimiz 0.9999, yani 1'e çok yakın bir değer çıkmıştır.

Otokorelasyon hataları ile bir kaydırılmış hatalar arasındaki otokorelasyondur. Korelasyon iki değişken arasındaki ilişkiyi gösterir. Korelasyon anlamlılığı testine t istatistiği ile bakılır.

Bir gecikmeli ile hataların sacını grafiğini çizdirelim;

```
n<-length(residuals(model))
plot(tail(residuals(model),n-1),
     head(residuals(model),n-1),
     xlab = expression(hat(epsilon)[i]),
     ylab=expression(hat(epsilon)[i+1]),main="Bir Gecimeliler Ile Hatalarin Sacini Grafiği",col="purple")
```

Bir Gecimeliler Ile Hatalarin Sacini Grafiği



Gecikmeli artıklar için kuyruk kısmından ve baş kısmından birer tane atıp ikisi arasında sacını grafiğini çizdirelim. Hiçbir model bilmiyorken ilişkiyi anlamak için birini yanıt değişkeni birini açıklayıcı değişken yapıp regresyon modeli kurulur. Eğer model anlamlı çıkarsa aralarındaki ilişki anlamlıdır denir. Korelasyon testi ile daha net sonuç görülebilir.

```
cor.test(tail(residuals(model),n-1),
         head(residuals(model),n-1))
```

```
##
## Pearson's product-moment correlation
##
## data: tail(residuals(model), n - 1) and head(residuals(model), n - 1)
## t = 0.73268, df = 245, p-value = 0.4645
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.07851958 0.17058208
## sample estimates:
## cor
## 0.04675817
```

Korelasyon testi sonucumuza gore p-value=0.4645 ,0.05 den buyuk oldugu icin hatalar ile bir gecikmeliler arasinda iliski yoktur.

DURBIN -WATSON TEST ISTATISTIGI

H0 : Otokorelasyon yoktur. H1 : Otokorelasyon vardır.

H0 RED ise otokorelasyon vardır. H0 KABUL ise otokorelasyon yoktur. Modelin anlamliligina bakariz.

```
library(lmtest)
dwtest(VWAP~Volume+
       High+
       Low+
       Close+
       Last,data=NSTS)
```

```
##
## Durbin-Watson test
##
## data: VWAP ~ Volume + High + Low + Close + Last
## DW = 1.906, p-value = 0.2024
## alternative hypothesis: true autocorrelation is greater than 0
```

Durbin Watson testimizin sonucuna gore p-value=0.2024,0.05 den buyuk oldugundan H0 hipotezi kabul edilir yani otokorelasyon yoktur.

Durbin-watson testi gecikmelere sadece birinci dereceden bakiyor bu yuzden daha yuksek derecelerden gecikmelere bakmak icin Breusch Godfrey testi yapilir.

BREUSCH-GOLDFREY TEST ISTATISTIGI

H0 : Otokorelasyon yoktur. H1 : Otokorelasyon vardır.

H0 KABUL Otokorelasyon yok. H0 RED Otokorelasyon var. Modelin anlamliligina degil gecikmelilerin kastsayilarinin modele kattigi anlamlara bakariz.

```
library(lmtest)
bgtest(VWAP~Volume+
       High+
       Low+
       Close+
       Last,data=NSTS)
```

```
##
## Breusch-Godfrey test for serial correlation of order up to 1
##
## data: VWAP ~ Volume + High + Low + Close + Last
## LM test = 0.55689, df = 1, p-value = 0.4555
```

Breusch Goldfrey testimizin sonucuna gore p-value=0.4555 ,0.05 den buyuk oldugundan h0 hipotezi kabul edilir yani otokorelasyon yoktur.