

## İçindekiler

1. Veri ve Senaryosu .....	3
2. Tanımlayıcı İstatistikler .....	4
3. Normallik ve Doğrusallık Varsayımları .....	5
3.1 Normallik .....	5
3.2 Doğrusallık .....	9
4. Çoklu Regresyon Modeli ve Artık İncelemesi .....	10
4.1 Regresyon Modeli.....	10
4.2 Artık İncelemesi .....	12
4.3 Artıksız Normallik .....	16
4.4 Artıksız Doğrusallık .....	17
4.5 İkinci Artık İncelemesi .....	19
5. Çoklu Regresyon Modeli ve Anlamlılığı.....	22
5.1 Regresyon Modeli.....	22
5.2 Anlamlılık ve Kestirim Denklemi.....	24
6. Katsayı Anlamlılıkları ve Yorumları.....	25
7. Belirtme Katsayısının Yorumlanması .....	27
8. %99 Güven Aralıkları ve Yorumlanması .....	27
9. Değişen Varyanslılık Sorunu .....	29
10. Öz İlişki Sorunu .....	30
11. Çoklu Bağlantı Sorunu.....	31
11.1 VİF ve Koşul Sayısı .....	31
11.2 Özdeğer ve Özvektör.....	32
12. Uyum Kestirimi.....	34
13. Ön Kestirim .....	34
14. Uyum Kestirimi ve Ön Kestirim için GA.....	35
15. Değişken Seçimi- En İyi Modelin Bulunması.....	36
15.1 İleriye Doğru Seçim .....	36
15.2 Geriye Doğru Seçim .....	38
15.3 Adımsal Seçim .....	39
16. Ridge Regresyon Modeli .....	40
Kaynakça.....	41

# 1. Veri ve Senaryosu

Bir online alışveriş sitesi yöneticileri; web sitelerinin tasarımını değiştirmek istemektedir. Bunun için yaptığı çalışma sonucunda üç farklı web site tasarımı arasında seçim yapılması gerekiyor.

Müşteri etkinliği ve site tasarımları arasındaki ilişkiyi anlamak için, her bir site tasarımı belli bir süre uygulanacak ve müşteri etkinliği izlenecektir. Üç farklı site tasarımıyla birlikte; gönderilen öneri bildirimlerinin sıklığı, indirimli ürünlerin oranı ve müşteri geçmişi üzerinden (arama geçmişi, satın alma geçmişi, liste geçmişleri) tahminler yapan yapay zekanın tahmin oranı da bu izlenmeye dahil tutulacaktır. Yapılan incelemelerin bir kısmı aşağıdaki gibidir.

y	Yapay zekâ	İndirimli ürün	Bildirim sıklığı	Site
Müşteri etkinliği	oranı (%)	oranı (%)	(günlük ortalama)	Tasarımı
52.03710	12.192298	2.2533662	4.2662135	1
45.31281	10.460990	0.5147431	2.6325714	1
41.70349	10.193229	0.6273021	2.3787847	2
35.01419	9.276898	0.7697428	1.2559291	2
46.58245	12.249632	1.8403595	3.4824209	3
46.42294	12.486872	1.9821151	4.7902894	3

❖ Görüldüğü üzere verimizdeki değişkenler,

- Müşteri Etkinliği → Bağımlı Değişken
- Yapay Zekâ Oranı → Nicel Bağımsız Değişken
- İndirimli Ürün Oranı → Nicel Bağımsız Değişken
- Bildirim Sıklığı → Nicel Bağımsız Değişken
- Site Tasarımı → Nitel Bağımsız Değişkendir.

➤ Verimizi R Studio'ya aktarıp kullanılabılır hale getirelim:

```
ham_veri <- read.csv("C:/Users/gungo/Desktop/regresyon/ham_veri.txt", sep="")
```

```
head(ham_veri)
```

```
##      y      yapay_zeka  indirimli_urun  bildirim  site_tasarimi
## 1 52.03710  12.19230    2.2533662    4.266214      1
## 2 45.31281  10.46099    0.5147431    2.632571      1
## 3 57.55408  12.20494    2.5136276    4.044828      1
## 4 53.08547  11.42794    0.9101550    2.595185      1
## 5 50.80389  11.11258    1.5749445    3.071317      1
```

```
attach(ham_veri)
```

```
names(ham_veri)
```

```
## [1] "y"      "yapay_zeka"  "indirimli_urun"  "bildirim"  "site_tasarimi"
```

```
site_tasarimi <- as.factor(site_tasarimi) (Nitel Değişken olduğu için)
```

## 2. Tanımlayıcı İstatistikler

➤ Bağımlı ve Bağımsız Değişkenlerimiz için Tanımlayıcı İstatistikler:

```
library(pastecs)
stat.desc(ham_veri)

##          y yapay_zeka indirimli_urun  bildirim
## nbr.val  100.0000000 100.0000000 100.0000000 100.0000000
## nbr.null  0.0000000 0.0000000 0.0000000 0.0000000
## nbr.na    0.0000000 0.0000000 0.0000000 0.0000000
## min       32.3949616 8.1789956 -2.0492834 0.6966112
## max       148.8041980 13.2693439 3.4554261 5.2446430
## range     116.4092364 5.0903483 5.5047095 4.5480318
## sum       4631.8471149 1106.1698541 95.8013057 298.5208271
## median    43.3765103 11.0401421 0.9901032 3.0706802
## mean      46.3184711 11.0616985 0.9580131 2.9852083
## SE.mean    1.4714556 0.1114857 0.1062103 0.1009239
## CI.mean.0.95 2.9196872 0.2212118 0.2107443 0.2002548
## var       216.5181608 1.2429057 1.1280630 1.0185625
## std.dev    14.7145561 1.1148568 1.0621031 1.0092386
## coef.var    0.3176822 0.1007853 1.1086520 0.3380798

summary(site_tasarimi)

## 1 2 3
## 25 35 40
```

✚ Çıktıda her bir değişken için özet istatistikler verilmiştir. Buna göre;

- ✓ Müşteri etkinliği (y) için minimum değer 32.39, maksimum değer 148.80, ortalama değer 46.32, varyans 216.518 ve range değeri 116.409'dur.
- ✓ Yapay zeka oranı (yapay\_zeka) için minimum değer 8.179, maksimum değer 13.269, ortalama değer 11.062, varyans 1.24 ve range değeri 5.09'dur.
- ✓ İndirimli ürün oranı (indirimli\_urun) için minimum değer -2.0493, maksimum değer 3.4554 ortalama değer 0.9580, varyans 1.128 ve range değeri 5.504'tür.
- ✓ Bildirim sıklığı (bildirim) için minimum değer 0.6966, maksimum değer 5.2446, ortalama değer 2.9852, varyans 1.018 ve range değeri 4.548'dir
- ✓ Site tasarımı ise nitel bir değişkendir. Görüldüğü üzere; 1.site tasarımı için 25 veri, 2.site tasarımı için 35 veri ve 3.site tasarımı için 40 veri bulunmaktadır.

### 3. Normallik ve Doğrusallık Varsayımlarının İncelenmesi

#### 3.1 Normallik Varsayımı

➤ Bağımlı değişkenin normal dağılım gösterip göstermediğini inceleyelim:

Analizimize başlamadan önce Bağımlı değişkenin yani Müşteri Etkinliği (y)'nin Normal Dağılım gösterdiğine emin olmalıyız.

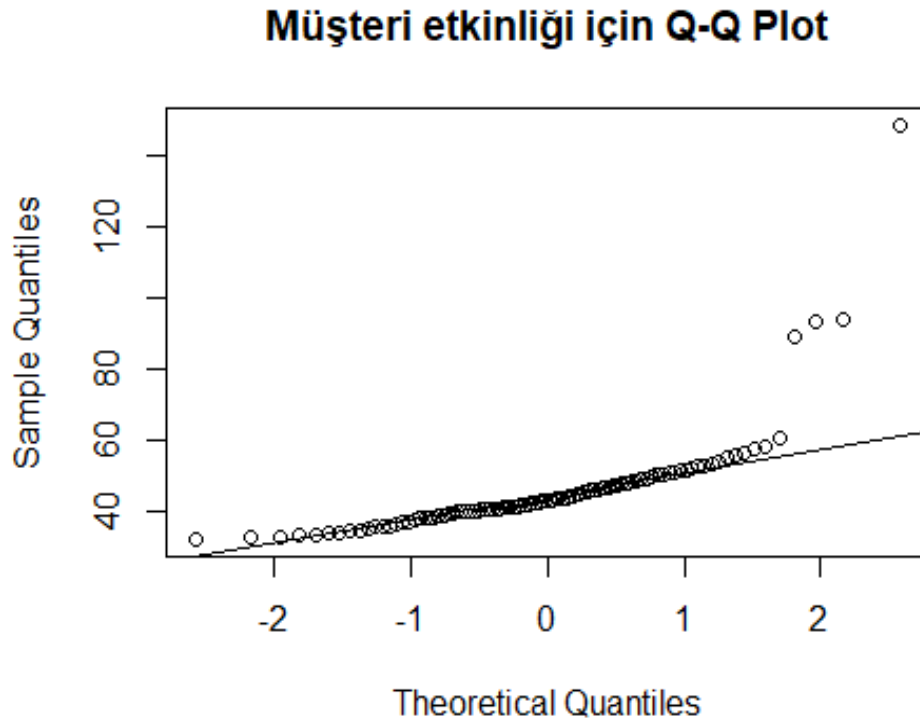
Bunun için çeşitli yöntemleri kullanabiliriz.

- Q-Q plot ile:

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
qqnorm(y, main = "Müşteri etkinliği için Q-Q Plot")  
qqline(y)
```



- ✓ Gözlem noktaları 45 derecelik bir açı etrafında dağılmış olsaydı bağımlı değişken (y) normal dağılmış olurdu. Fakat q-q plottan görüldüğü üzere değerlerimizin normal dağılım göstermediğini rahatlıkla söyleyebiliriz.

Verimizdeki gözlem sayısı ( $n=100 > 50$ ) olduğu için Kolmogrov-Smirnov testi ile normallik kontrolü yapabiliriz.

- **Kolmogrov-Smirnov Testi ile:**

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
library(nortest)
lillie.test(y)

##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: y
## D = 0.2028, p-value = 6.279e-11
```

✓  $p = 0 < \alpha = 0,05$  olduğundan  $H_0$  RED. Yani bağımlı değişken (y) normal dağılım göstermiyor.

- ❖ **Bu durumda dönüşüm uygulamamız gerekiyor.**

- **Ln Dönüşümü uygulandı.**

```
ln_y <- log(y)
```

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

- **Shapiro-Wilk Testi ile:**

```
shapiro.test(ln_y)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: ln_y  
## W = 0.81681, p-value = 8.408e-10
```

- **Anderson-Darling Testi ile:**

```
ad.test(ln_y)
```

```
##  
## Anderson-Darling normality test  
##  
## data: ln_y  
## A = 3.3813, p-value = 1.612e-08
```

- **Kolmogorov-Smirnov Testi ile:**

```
lillie.test(ln_y)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: ln_y  
## D = 0.11251, p-value = 0.003298
```

✓ Değişkenin ln dönüşümü uygulanmış hali ile yapılan tüm testlerin sonucunda  $H_0$  REDDEDİLDİ. Yani gözlemlerimiz normal dağılım göstermiyor.

- **1/y Dönüşümü uygulandı.**

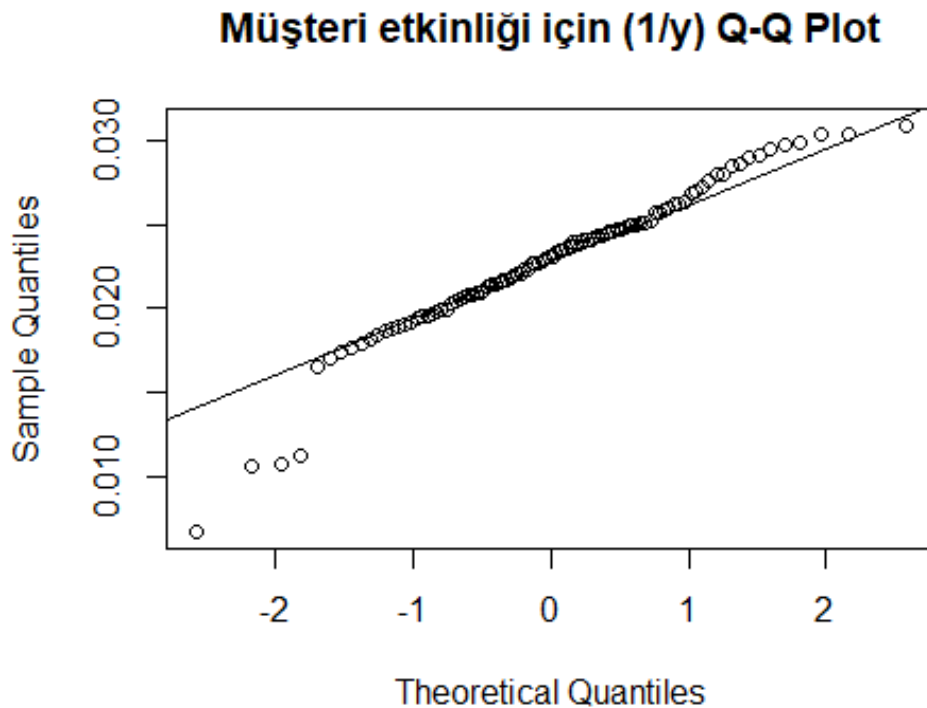
```
y2 <- 1/y
```

- **Q-Q plot ile:**

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
qqnorm(y2 main = "Müşteri etkinliği için (1/y) Q-Q Plot")  
qqline(y2)
```



- ✓ Q-Q plotta gözlemlerin dağılımı net belli olmuyor. Test ile control etmek daha mantıklı olacaktır.

- **Kolmogrov-Smirnov Testi ile:**

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
lillie.test(y2)
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: y2  
## D = 0.059555, p-value = 0.5171
```

- ✓  $p = 0.5171 > \alpha = 0,05$  olduğundan  $H_0$  REDDEDİLEMEZ. Yani bağımlı değişken (y) normal dağılıma uygundur.

## 3.2 Doğrusallık İncelemesi

- Müşteri etkinliği (y) ile Bağımsız değişkenler arasında doğrusal bir ilişki var mı inceleyelim:

Bağımlı değişken yani Müşteri Etkinliği (y) ile bağımsız değişkenlerin arasında doğrusal bir ilişki bulunması gerekmektedir.

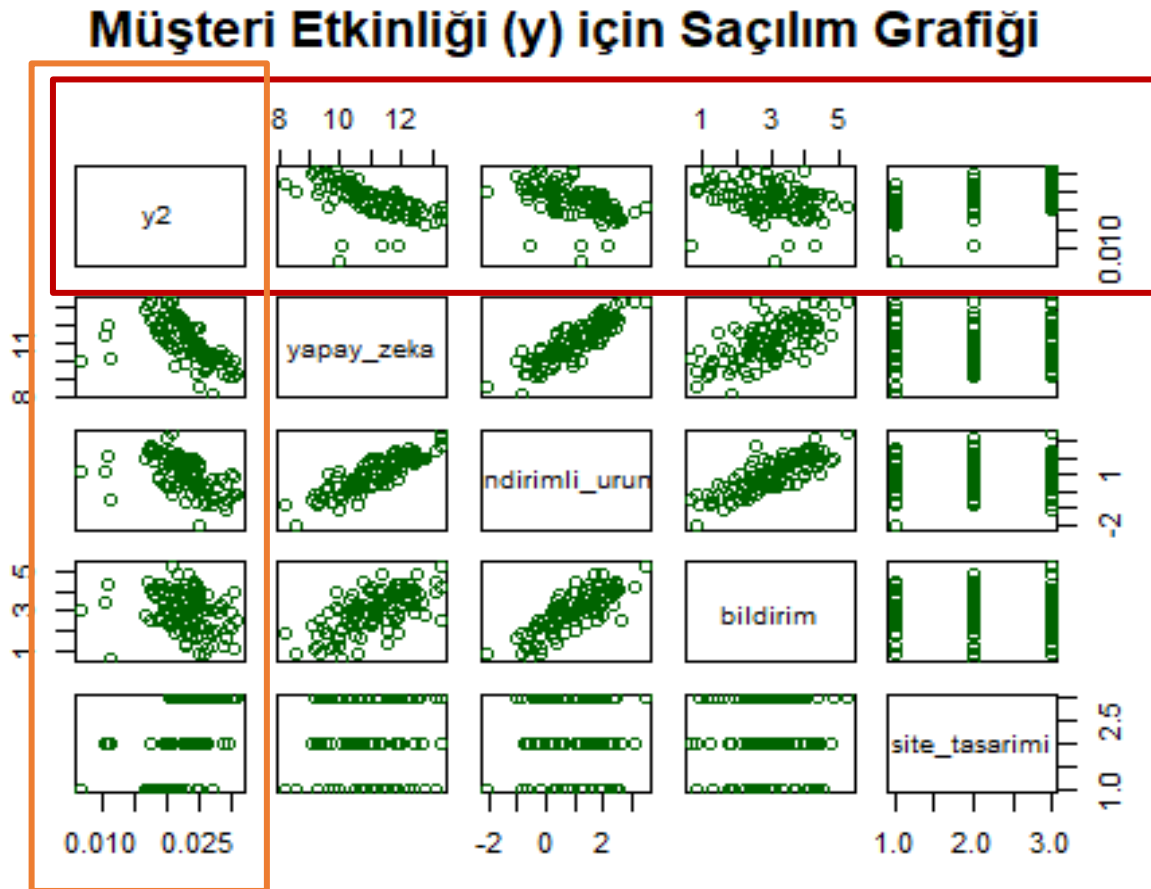
Bunu anlamak için öncelikle saçılım grafiğine bakmamız gerekmektedir.

- ✚ y değişkenine dönüşüm uyguladığımız için ham\_veri ile değil, dönüşüm uygulanan hali ile devam etmemiz gerekmektedir. Bunun için kod aşağıdadır.

```
yeni_veri <- cbind(y2,yapay_zeka,indirimli_urun,bildirim,site_tasarimi)
```

- Saçılım Grafiği:

```
pairs(yeni_veri,col="darkgreen", main="Müşteri Etkinliği(y)için Saçılım Grafiği")
```





- ✓ Müşteri etkinliği için doğrusallık incelendiğinde; yapay zeka oranı ile müşteri etkinliği arasında doğrusal bir ilişkinin olduğu, indirimli ürün oranı ve bildirim sıklığının ise yapay zeka oranına göre daha az bir ilişki olduğu söylenebilir. Fakat genel olarak tüm bağımsız nicel değişkenlerle müşteri etkinliği arasında doğrusal bir ilişki gözükmemektedir.
- ✓ Site tasarımı nitel değişken olduğu için ( göstermelik değişken olduğu için ) ilişkiye bakmaya gerek yoktur.

## 4. Çoklu Regresyon Modelinin kurulması ve Artık İncelemesinin Yapılması

### 4.1 Regresyon Modeli

```
model <- lm(y2 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi)
summary(model)
```

```
##
## Call:
## lm(formula = y2 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.0146560	-0.0004046	0.0004746	0.0011228	0.0028541

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.0368285	0.0055263	6.664	1.79e-09 ***
yapay_zeka	-0.0017790	0.0005088	-3.496	0.000722 ***
indirimli_urun	-0.0014549	0.0006967	-2.088	0.039477 *
bildirim	0.0013092	0.0004803	2.726	0.007654 **
site_tasarimi2	0.0027187	0.0007237	3.757	0.000299 ***
site_tasarimi3	0.0053948	0.0007047	7.655	1.68e-11 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002761 on 94 degrees of freedom
## Multiple R-squared:  0.605, Adjusted R-squared:  0.584
## F-statistic: 28.8 on 5 and 94 DF, p-value: < 2.2e-16
```

- ✓  $R^2 = 0.605 \rightarrow$  alışveriş sitesinin müşteri etkinliğindeki değişimin %60.5 'i; sitedeki yapay zekanın kullanım oranı, günlük uygulanan indirimli ürün oranı, günlük ortalama bildirim sayısı ve kullanılan site tasarımı tarafından açıklanabilmektedir.
- ✓ Değişimin geriye kalan diğer %39.5'lik kısmı ise farklı etkenler açıklıyor olabilir.
- ✓ Site Tasarımı 1'in kılavuz değişken olduğu elde edilen modelde olmadığından görülebilir.

- **Model Anlamlılık Testi:**

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (Model anlamsızdır.)

$H_1$  :En az bir  $\beta_i$  sıfırdan farklı. (Model anlamlıdır.)

- ✓  $p = 0.00 < \alpha = 0.05$   $H_0$  **RED**. Kurulan regresyon modelinin **anlamlı** olduğu %95 güven düzeyinde söylenebilir.
- ✓ En az bir tane bağımsız değişken bağımlı değişkeni açıklamaktadır. Bağımsız ve bağımlı değişkenler arasında doğrusal bir ilişki kurulabilir.

$$\frac{1}{y_i} = 0.0368285 - 0.0017790x_{1i} - 0.0014549x_{2i} + 0.0013092x_{3i} + 0.0027187x_{4i} + 0.0053948x_{5i} \mp 0.002761$$

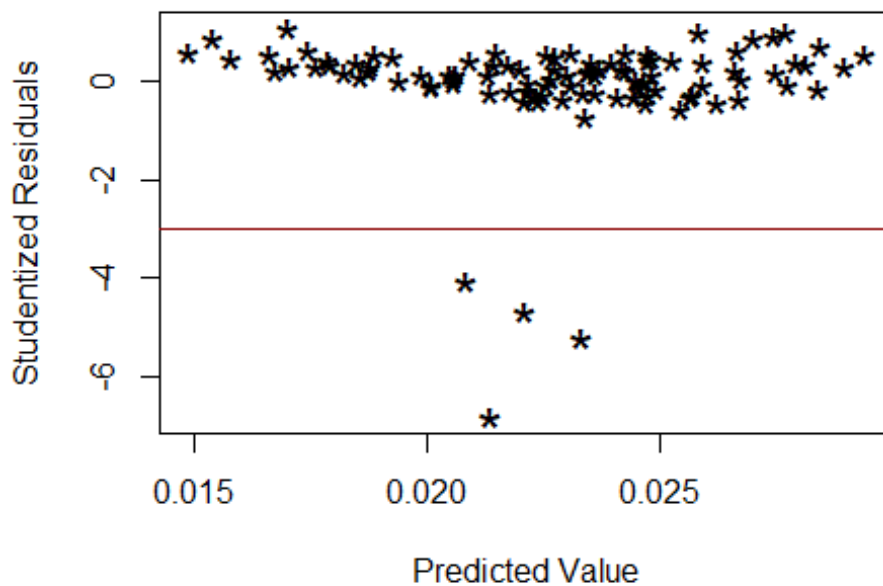
(0.0055263)    (0.0006967)    (0.0004803)    (0.0007237)    (0.0007047)

- ❖ **Modelin istatistiksel olarak anlamlı olduğu görüldü. Kabaca çıktıya bakıldığında tüm bağımsız değişkenlerin de anlamlı olduğu söylenebilir. Fakat yine de artık incelemesi yapılmasında fayda vardır.**

## 4.2 Artık İncelemesi

- Aykırı Değer ( $r_i$ ):

```
inf <- ls.diag(model)
plot(predict(model), (inf$stud.res), ylab="Studentized Residuals", xlab="Predicted Value")
abline(h=-3, col="darkred")
```



```
which(inf$stud.res<(-3))
```

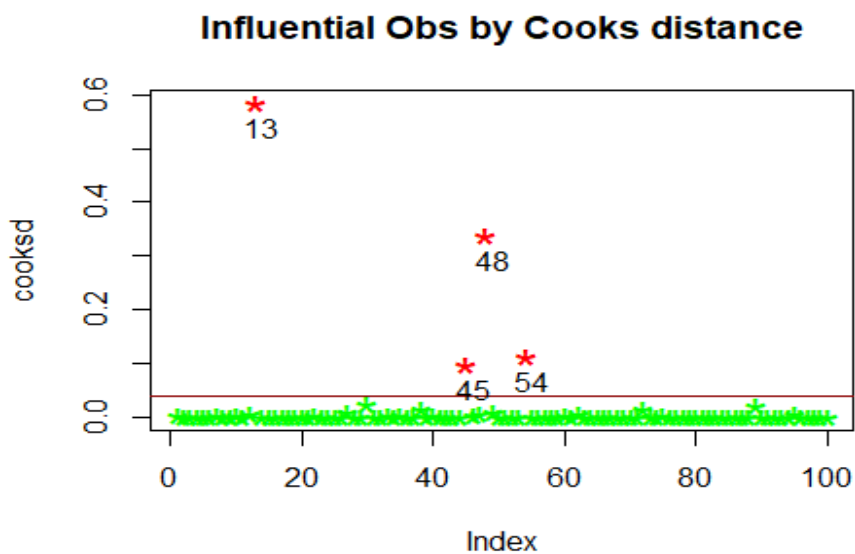
```
## [1] 13 45 48 54
```

- ✓ Student tipi artıkların (-3,3) aralığında olması istenir. Grafiğe bakıldığında 13, 45, 48 ve 54. gözlemlerin bu aralığa uymadığı görülüyor.
- ✓ Bu durumda bu gözlemler aykırı değerdir ve veriden çıkartılmalı.

- Cook Uzaklığı (Di):

```
n <- length(y2)
k <- length(ham_veri)-1

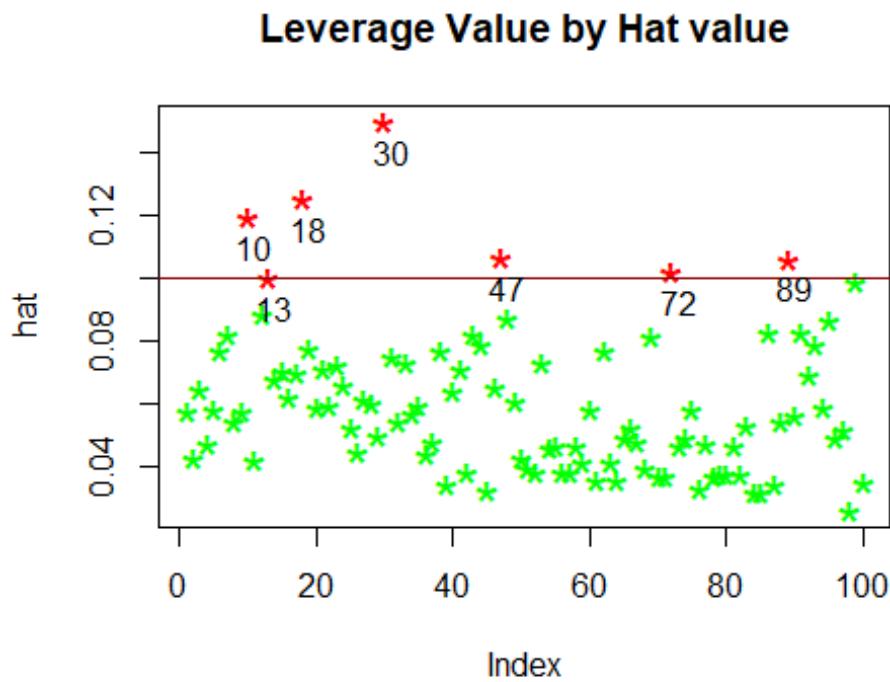
cooks_d <- cooks.distance(model)
plot(cooks_d, pch="*", cex=2, main="Influential Obs by Cooks distance", col=
ifelse(cooks_d>4/n, 'red', 'green'))
abline(h = if (n>50) 4/n else 4/(n-k-1) , col="darkred")
text(x=1:length(cooks_d)+1, y=cooks_d, labels=ifelse(cooks_d>if (n>50) 4/n else
4/(n-k-1), names(cooks_d), ""), col="black", pos = 1)
```



- ✓  $n = 100 > 50 \rightarrow \frac{4}{n} = \frac{4}{100} = 0.04 < D_{13}, D_{45}, D_{48}, D_{54}$  olduğundan ve grafiğe bakıldığında; 13, 45, 48 ve 54. Gözlemlerin etkin aykırı değer oldukları görülmektedir.
- ✓ Bu durumda bu gözlemler aykırı değerdir ve veriden çıkartılmalı.

- Gözlem Uzaklığı (Hii):

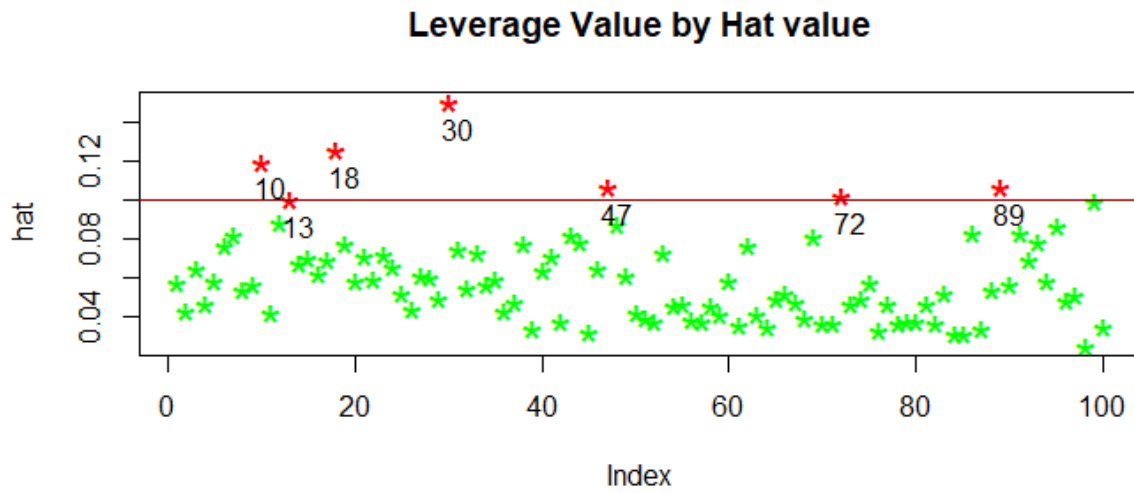
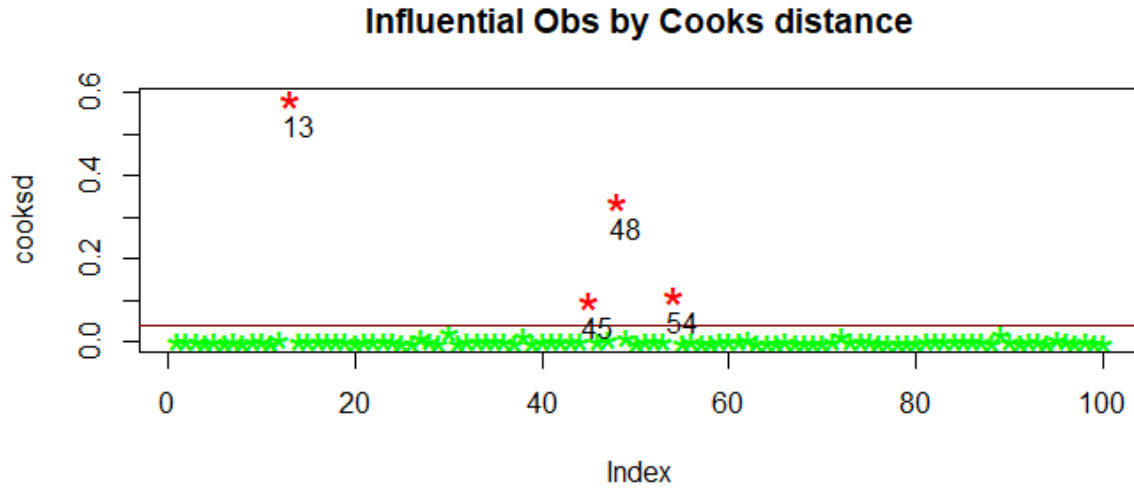
```
library(zoo)
hat <- inf$hat
plot(hat, pch="*", cex=2, main="Leverage Value by Hat value", col=ifelse(hat > 2*(k+1)/n, 'red', 'green'))
abline(h = 2*(k+1)/n , col="darkred")
text(x=1:length(hat)+1, y=hat, labels=ifelse(hat > 2*(k+1)/n, index(hat), ""), col="black", pos = 1)
```



✓  $\frac{2(k+1)}{n} = \frac{10}{100} = 0.01 < H_{ii}$  koşulunu sağlayan  $H_{ii}$  değerleri uç değerlerdir.

✓ Grafiğe bakıldığında; 10, 13, 18, 30, 47, 72 ve 89. Gözlemler uç değerlerdir.

✓ Bu durumda bu gözlemler veriden çıkartılmalı.



- ✓ Sonuç olarak; aykırı, etkin ve uç gözlemlerin ( 10, 13, 18, 30, 45, 47, 48, 54, 72, 89) tamamı veriden çıkartılmalı ve en baştan adımlar uygulanmalıdır.

```
son_veri <- yeni_veri[-c(10, 13, 18, 30, 45, 47, 48, 54, 72, 89),]
```

### 4.3 Artıksız Normallik Varsayımı

- Bağımlı değişkenin normal dağılım gösterip göstermediğini inceleyelim:

Bağımlı değişkenin yani Müşteri Etkinliği (y)'nin Normal Dağılım gösterdiğine emin olmalıyız.

- 3.1 adımında bağımlı değişken (y) normal dağılmadığı için (1/y) dönüşümü yapılmıştı. Fakat artık incelemesinden sonra çıkarılan gözlemlerle (y) değişkeni normal dağılım göstermektedir. Bu yüzden dönüşüme (1/y<sup>2</sup>) dönüşümü uygulanarak eski haline geri getirilmiştir.

$y_3 \leftarrow 1/y_2$

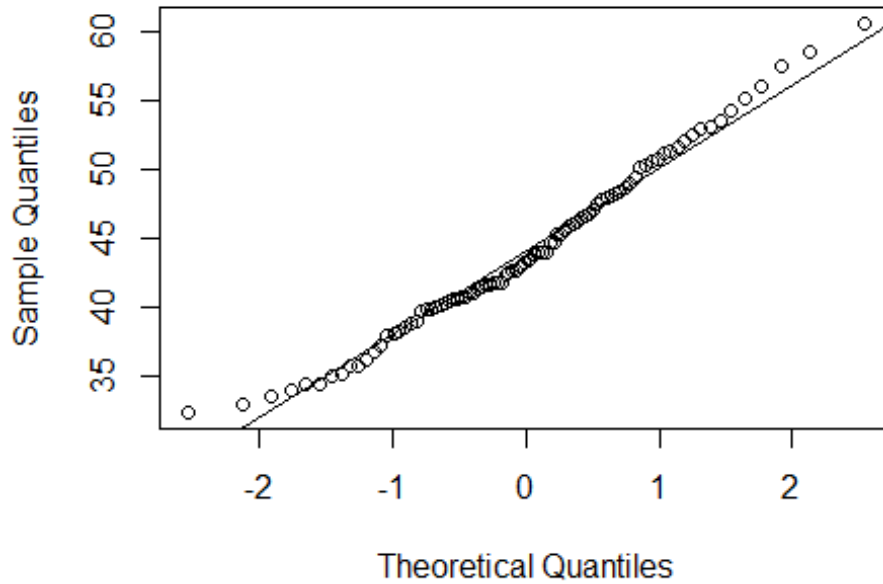
- Q-Q plot ile:

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.

$H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
qqnorm(y3, main = "Artıksız Müşteri etkinliği için (y3) Q-Q Plot")  
qqline(y3)
```

Artıksız Müşteri etkinliği için (y3) Q-Q Plot



- ✓ Görüldüğü üzere; gözlem noktaları 45 derecelik bir açı etrafında dağılmıştır. Grafiğe bakarak bağımlı değişkenin **normal dağılım gösterdiğini** söyleyebiliriz.

Verimizdeki gözlem sayısı ( $n=90 > 50$ ) olduğu için Kolmogrov-Smirnov testi ile normallik kontrolü yapabiliriz.

- **Kolmogrov-Smirnov Testi ile:**

$H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.  
 $H_1$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

```
library(nortest)
lillie.test(y3)

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  y3
## D = 0.073912, p-value = 0.2623
```

✓  $p = 0.2623 > \alpha = 0,05$  olduğundan  $H_0$  REDDEDİLEMEZ. Yani bağımlı değişken (y) **normal dağılım gösteriyor**.

## 4.4 Artıksız Doğrusallık İncelemesi

➤ Müşteri etkinliği (y3) ile Bağımsız değişkenler arasında doğrusal bir ilişki var mı inceleyelim:

Bağımlı değişken yani Müşteri Etkinliği (y) ile bağımsız değişkenlerin arasında doğrusal bir ilişki bulunması gerekmektedir.

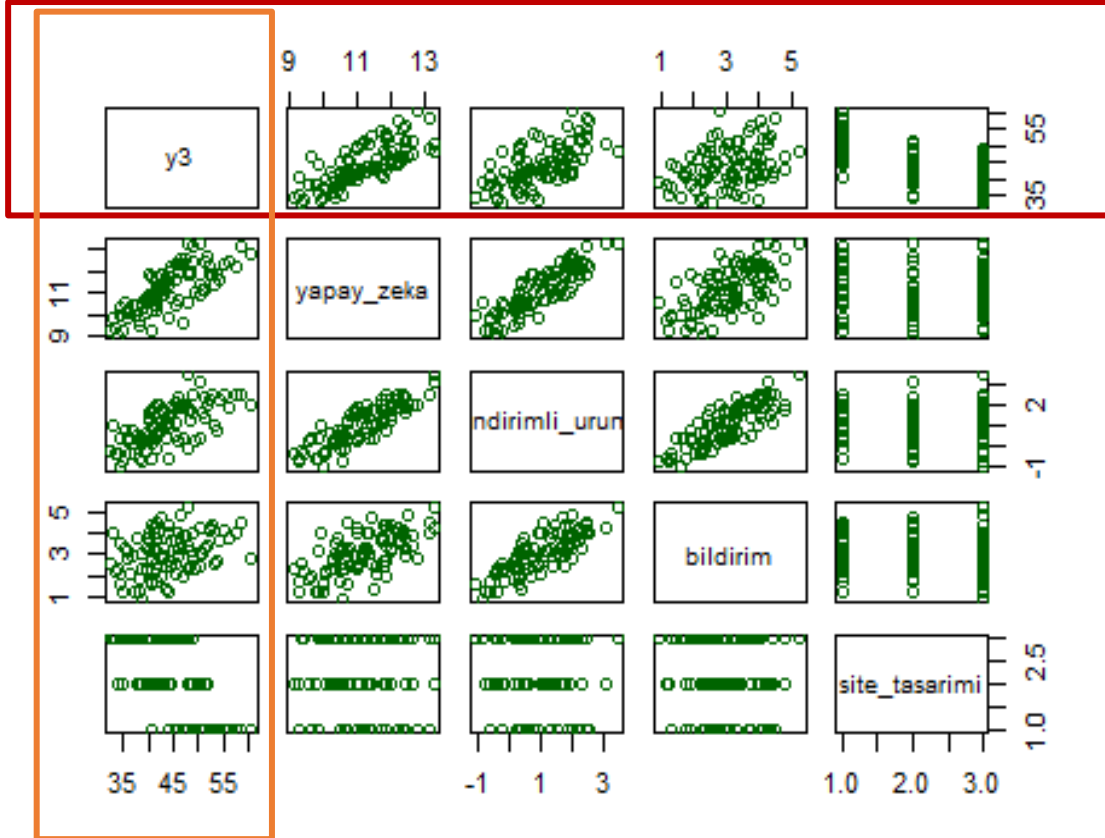
Bunu anlamak için öncelikle saçılım grafiğine bakmamız gerekmektedir.



- Saçılım Grafiği:

```
pairs(son_veri,col="darkgreen", main="Artıksız Müşteri Etkinliği(y3) için S  
açılım Grafiği")
```

### Artıksız Müşteri Etkinliği(y3) için Saçılım Grafiği

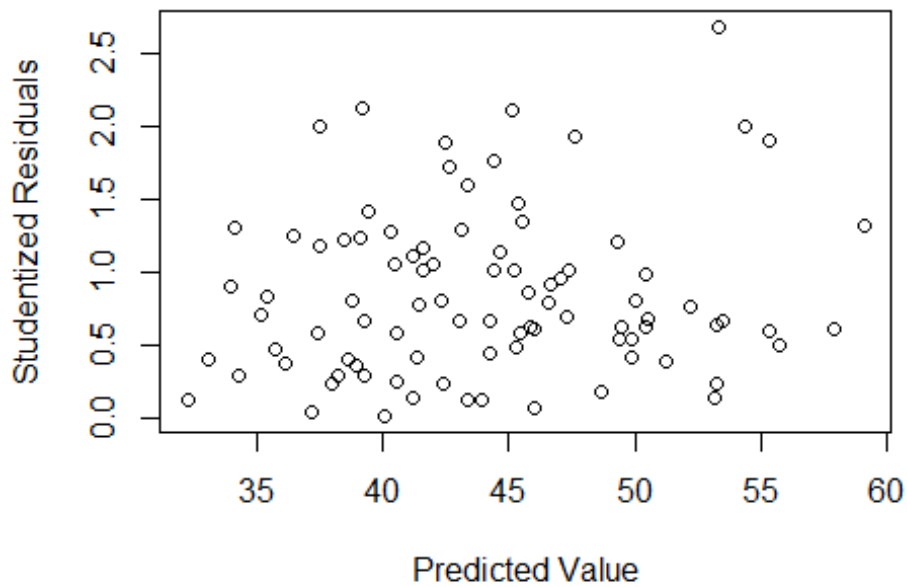


- ✓ Artıksız veride Müşteri etkinliği için doğrusallık incelendiğinde; yapay zeka oranı ve indirimli ürün oranı ile müşteri etkinliği arasında doğrusal bir ilişki olduğu görülmektedir. Bildirim sayısı ile müşteri etkinliği arasındaki ilişkinin ise diğer bağımsız değişkenlerden daha az olduğu söylenebilir.
- ✓ Site tasarımı nitel değişken olduğu için ( göstermelik değişken olduğu için ) ilişkiye bakmaya gerek yoktur.

## 4.5 İkinci Artık İncelemesi

- Aykırı Değer ( $r_i$ ):

```
inf2 <- ls.diag(model2)
plot(predict(model2), abs(inf2$stud.res), ylab="Studentized Residuals", xlab="Predicted Value")
abline(h=-3, col="darkred")
```



```
which(inf2$stud.res<(-3))
```

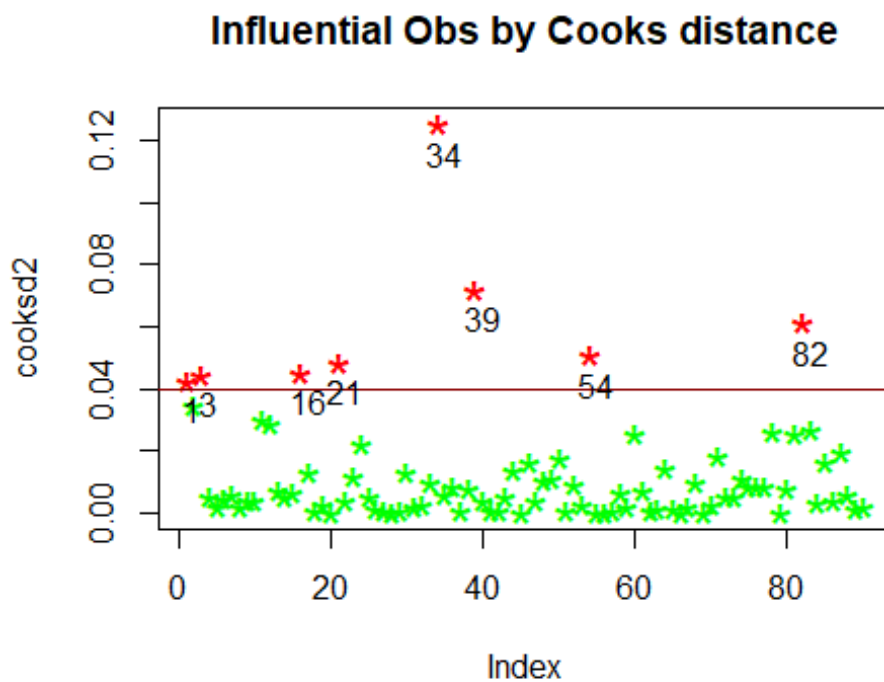
```
## integer(0)
```

- ✓ Student tipi artıkların (-3,3) aralığında olması istenir. Grafiğe bakıldığında bu değerlerin dışına çıkan gözlem yoktur.

- Cook Uzaklığı (Di):

```
n2 <- length(y3)
k2 <- length(son_veri)-1

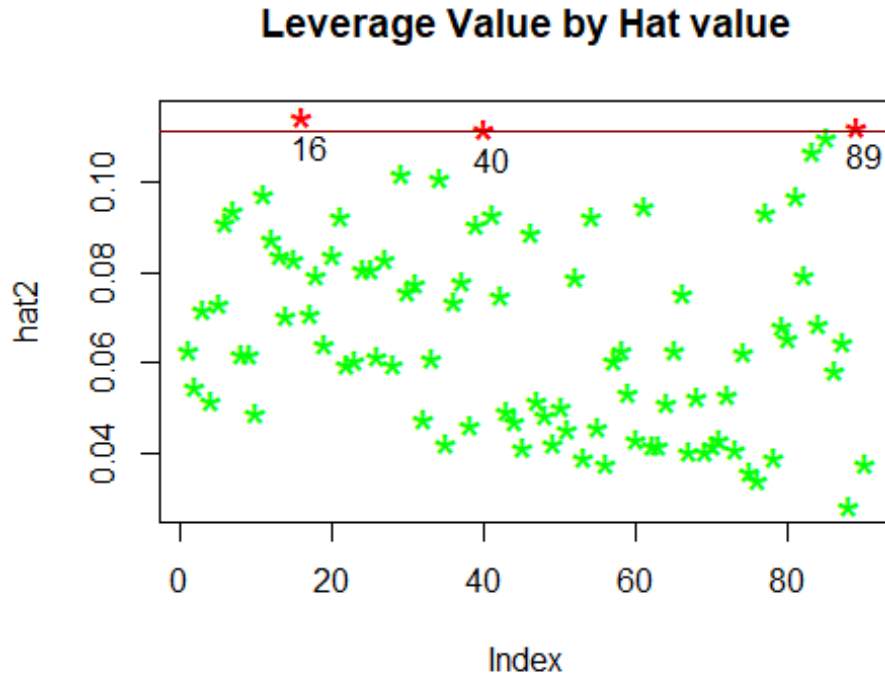
cooks2 <- cooks.distance(model2)
plot(cooks2, pch="*", cex=2, main="Influential Obs by Cooks distance", col=
  ifelse(cooks2>4/n, 'red', 'green'))
abline(h = if (n>50) 4/n else 4/(n-k-1) , col="darkred")
text(x=1:length(cooks2)+1, y=cooks2, labels=ifelse(cooks2>if (n>50) 4/n
  else 4/(n-k-1), names(cooks2), ""), col="black", pos = 1)
```



- ✓  $n = 90 > 50 \rightarrow \frac{4}{n} = \frac{4}{90} = 0.0444 < D_1, D_3, D_{16}, D_{21}, D_{34}, D_{39}, D_{54}, D_{82}$  olduğundan ve grafiğe bakıldığında; 1, 3, 16, 21, 34, 39, 54 ve 82. Gözlemlerin etkin aykırı değer oldukları görülmektedir.
- ✓ Bu durumda bu gözlemler aykırı değerdir ve veriden çıkartılmalıdır.

- Gözlem Uzaklığı (Hii):

```
library(zoo)
hat2 <- inf2$hat
plot(hat2, pch="*", cex=2, main="Leverage Value by Hat value", col=ifelse(hat2>2*(k2+1)/n2, 'red', 'green'))
abline(h = 2*(k2+1)/n2 , col="darkred")
text(x=1:length(hat2)+1, y=hat2, labels=ifelse(hat2>2*(k2+1)/n2, index(hat2), ""), col="black", pos = 1)
```



- ✓  $\frac{2(k+1)}{n} = \frac{10}{90} = 0.0111 < H_{ii}$  koşulunu sağlayan  $H_{ii}$  değerleri uç değerlerdir.
- ✓ Grafiğe bakıldığında; 16, 40 ve 89. Gözlemler uç değerlerdir.
- ✓ Bu durumda bu gözlemler veriden çıkartılmalı.
- ✓ Sonuç olarak; aykırı, etkin ve uç gözlemlerin ( 1, 3, 16, 21, 34, 39, 40, 54, 82 ve 89) tamamı veriden çıkartılmalı ve en baştan adımlar uygulanmalıdır.
- ✓ Fakat biz bir sefer bu işlemi yaptığımız için devam edeceğiz.

## 5. Çoklu Regresyon Modeli ve Anlamlılığı

### 5.1 Regresyon Modeli

```
model2 <- lm(y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi)
summary(model2)

##
## Call:
## lm(formula = y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0010 -0.7621  0.1386  0.8872  2.4322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.1329     2.9548   3.768 0.000305 ***
## yapay_zeka      3.8963     0.2666  14.614 < 2e-16 ***
## indirimli_urun  1.9169     0.3567   5.373 6.79e-07 ***
## bildirim       -2.0161     0.2450  -8.230 2.10e-12 ***
## site_tasarimi2  -6.7985     0.3500 -19.423 < 2e-16 ***
## site_tasarimi3  -9.5264     0.3297 -28.898 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.226 on 84 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9617
## F-statistic: 448.5 on 5 and 84 DF, p-value: < 2.2e-16
```

- ✓  $R^2 = 0.9639 \rightarrow$  alışveriş sitesinin müşteri etkinliğindeki değişimin %96.39 'u; sitedeki yapay zekanın kullanım oranı, günlük uygulanan indirimli ürün oranı, günlük ortalama bildirim sayısı ve kullanılan site tasarımı tarafından açıklanabilmektedir.
- ✓ Değişimin geriye kalan %3.61'lik kısmı ise farklı etkenler açıklıyor olabilir.
- ✓ Site Tasarımı 1'in kılavuz değişken olduğu elde edilen modelde olmadığından görülebilir.

### Değişken Etkileşimlerinin Modele Etkisi (Nicel\*Nitel):

- ✓ Nitel ve nicel değişkenler arasındaki **etkileşimlere** tek tek bakıldı fakat hiçbir etkileşim aşağıda da görüldüğü üzere belirtme katsayısını etkilemedi ve anlamlı çıkmadı. Bu yüzden modele dahil edilmedi.

```
model2 <- lm(y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi+yapay_zeka*site_tasarimi+indirimli_urun*site_tasarimi+bildirim*site_tasarimi)
summary(model2)
```

```
##
## Call:
## lm(formula = y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi +
##      yapay_zeka * site_tasarimi + indirimli_urun * site_tasarimi +
##      bildirim * site_tasarimi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4590 -0.7762  0.1052  0.9377  2.4241
##
## ## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.0284     5.4880   1.098 0.275377
## yapay_zeka        4.5967     0.5229   8.791 2.74e-13 ***
## indirimli_urun    2.5584     0.7037   3.635 0.000496 ***
## bildirim         -3.1771     0.7478  -4.248 5.90e-05 ***
## site_tasarimi2    -2.5930     7.6454  -0.339 0.735403
## site_tasarimi3    -0.6779     7.3649  -0.092 0.926899
## yapay_zeka:site_tasarimi2 -0.6571     0.7307  -0.899 0.371244
## yapay_zeka:site_tasarimi3 -1.0073     0.6726  -1.498 0.138226
## indirimli_urun:site_tasarimi2 -1.0740     0.9424  -1.140 0.257897
## indirimli_urun:site_tasarimi3 -0.2550     0.9463  -0.269 0.788261
## bildirim:site_tasarimi2  1.4283     0.8235   1.735 0.086777
## bildirim:site_tasarimi3  0.9357     0.8565   1.092 0.278010
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.217 on 78 degrees of freedom
## Multiple R-squared:  0.9669, Adjusted R-squared:  0.9623
## F-statistic: 207.4 on 11 and 78 DF, p-value: < 2.2e-16
```

## 5.2 Model Anlamlılığı ve Kestirim Denklemi

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.1329     2.9548   3.768 0.000305 ***
## yapay_zeka      3.8963     0.2666  14.614 < 2e-16 ***
## indirimli_urun  1.9169     0.3567   5.373 6.79e-07 ***
## bildirim       -2.0161     0.2450  -8.230 2.10e-12 ***
## site_tasarimi2  -6.7985     0.3500 -19.423 < 2e-16 ***
## site_tasarimi3  -9.5264     0.3297 -28.898 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.226 on 84 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9617
## F-statistic: 448.5 on 5 and 84 DF, p-value: < 2.2e-16
```

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (Model anlamsızdır.)

$H_1$  : En az bir  $\beta_i$  sıfırdan farklı. (Model anlamlıdır.)

- ✓  $p = 0.00 < \alpha = 0.05$   $H_0$  **RED**. Kurulan regresyon modelinin **anlamlı** olduğu %95 güven düzeyinde söylenebilir.
- ✓ En az bir tane bağımsız değişken bağımlı değişkeni açıklamaktadır. Bağımsız ve bağımlı değişkenler arasında **doğrusal bir ilişki kurulabilir**.

**Müşteri Etkinliği (y) = 11.133 + 3.90 Yapay Zeka + 1.92 İndirimli Ürün – 2.02 Bildirim**

**(2.95) (0.27) (0.36) (0.25)**

**-6.80 Site Tasarımı2 -9.53 Site Tasarımı3 + 1.226**

**(0.35) (0.33)**

## 6. Regresyon Katsayılarının Anlamlılığı ve Yorumlanması

- $H_0 : \beta_0 = 0$  (Sabit terimin modele katkısı anlamlı değildir.)  
 $H_1 : \beta_0 \neq 0$  (Sabit terimin modele katkısı anlamlıdır.)

$$b_0 = 11.133 \quad S_{b_0} = 0.95 \quad t = 3.77 \quad p = 0.0003$$

- ✓  $p = 0.0003 < \alpha = 0.05$   $H_0$  **RED.** Sabit terimin modele katkısı **anlamlıdır.**
- ✓ Bağımsız değişkenlerin “0” olmasında herhangi bir mantıksal sorun olmayacağı için, **sabit terim anlamlıdır.**
- ✓ Sitedeki o gün kullanılan yapay zeka oranı ‘0’ olduğunda, indirimli ürün oranı ‘0’ olduğunda, gönderilen bildirim sayısı ‘0’ olduğunda ve varsayılan site tasarımı kullanıldığında; sitedeki ortalama müşteri etkinliği **11.133’tür.**

- $H_0 : \beta_1 = 0$  (Günlük kullanılan yapay zeka oranının modele katkısı anlamlı değildir.)  
 $H_1 : \beta_1 \neq 0$  (Günlük kullanılan yapay zeka oranının modele katkısı anlamlıdır.)

$$b_1 = 3.90 \quad S_{b_1} = 0.27 \quad t = 14.6 \quad p = 0$$

- ✓  $p = 0.0 < \alpha = 0.05$   $H_0$  **RED.** Günlük kullanılan yapay zeka oranının **modele katkısı anlamlıdır.**
- ✓ İndirimli ürün oranı, ortalama bildirim sayısı ve site tasarımı sabit tutulduğunda; günlük kullanılan yapay zeka oranı bir birim arttırıldığında müşteri etkileşimi **ortalama 3.90 birim artar.**

- $H_0 : \beta_2 = 0$  (Günlük indirimli ürün oranının modele katkısı anlamlı değildir.)  
 $H_1 : \beta_2 \neq 0$  (Günlük indirimli ürün oranının modele katkısı anlamlıdır.)

$$b_2 = 1.92 \quad S_{b_2} = 0.36 \quad t = 5.37 \quad p = 0$$

- ✓  $p = 0.0 < \alpha = 0.05$   $H_0$  **RED.** Günlük **indirimli ürün** oranının **modele katkısı anlamlıdır.**
- ✓ Yapay zeka oranı, ortalama bildirim sayısı ve site tasarımı sabit tutulduğunda; günlük indirimli ürün oranı bir birim arttırıldığında müşteri etkileşimi **ortalama 1.92 birim artar.**

- $H_0 : \beta_3 = 0$  (Günlük bildirim sayısının modele katkısı anlamlı değildir.)  
 $H_1 : \beta_3 \neq 0$  (Günlük bildirim sayısının modele katkısı anlamlıdır.)

$$b_3 = -2.02 \quad S_{b_3} = 0.25 \quad t = -8.23 \quad p = 0$$

- ✓  $p = 0 < \alpha = 0.05$   $H_0$  **RED.** Günlük **bildirim sayısının** modele katkısı **anlamlıdır.**
- ✓ Yapay zeka oranı, indirimli ürün oranı ve site tasarımı sabit tutulduğunda; günlük indirimli ürün oranı bir birim arttırıldığında müşteri etkileşimi **ortalama 2.02 birim azalır.**



- $H_0 : \beta_4 = 0$  (2.site tasarımı ile referans olan 1.site tasarımı arasında fark yoktur.)  
 $H_1 : \beta_4 \neq 0$  (2.site tasarımı ile referans olan 1.site tasarımı arasında fark vardır.)

$$b_4 = -6.80 \quad S_{b_4} = 0.35 \quad t = -19.423 \quad p = 0$$

- ✓  $p = 0 < \alpha = 0.05$   $H_0$  **RED.**
- ✓ Müşteri etkinliği bakımından 2.site tasarımı ve 1. Site tasarımı arasında istatistiksel olarak farklılık bulunmaktadır.
- ✓ Müşteri etkinliği bakımından 2.site tasarımının 1.site tasarımına göre daha etkili olduğu söylenebilir.

- $H_0 : \beta_5 = 0$  (3.site tasarımı ile referans olan 1.site tasarımı arasında fark yoktur.)  
 $H_1 : \beta_5 \neq 0$  (3.site tasarımı ile referans olan 1.site tasarımı arasında fark vardır.)

$$b_4 = -9.53 \quad S_{b_4} = 0.33 \quad t = -28.90 \quad p = 0$$

- ✓  $p = 0 < \alpha = 0.05$   $H_0$  **RED.**
- ✓ Müşteri etkinliği bakımından 3.site tasarımı ve 1. Site tasarımı arasında istatistiksel olarak farklılık bulunmaktadır.
- ✓ Müşteri etkinliği bakımından 3.site tasarımının 1.site tasarımına göre daha etkili olduğu söylenebilir.

## 7. Belirtme katsayısının ( $R^2$ ) Yorumlanması

```
##  
## Residual standard error: 1.217 on 78 degrees of freedom  
## Multiple R-squared: 0.9669, Adjusted R-squared: 0.9623  
## F-statistic: 207.4 on 11 and 78 DF, p-value: < 2.2e-16
```

- ✓ Belirtme katsayısı ( $R^2$ ), bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını gösterir. Bu değerin 1'e yakın olması bağımsız değişkenin bağımlı değişkeni iyi bir şekilde açıkladığı anlamına gelir ve bu da istenen bir durumdur.
- ✓ Modelimizin özet çıktısında baktığımızda hem belirtme katsayısı hem de düzeltilmiş belirtme katsayısının neredeyse aynı olduğunu görüyoruz. Herhangi birini yorumlamamızda sakınca yoktur. Biz bu analizde normal  $R^2$ 'yi yorumlayacağız.
- ✓  $R^2 = 0.9669 \rightarrow$  Müşteri etkinliğindeki değişimin %96.7 'si; sitedeki yapay zekanın kullanım oranı, uygulanan indirimli ürün oranı, günlük ortalama bildirim sayısı ve kullanılan site tasarımı tarafından açıklanabilmektedir.
- ✓ Değişimin geriye kalan diğer %3.3'lük kısmı ise farklı etkenler açıklıyor olabilir.

## 8. %99 Güven Aralıkları ve Yorumlanması

```
confint(model2, level = .99)  
##              0.5 %      99.5 %  
## (Intercept)  3.3449779 18.920735  
## yapay_zeka   3.1936033  4.599002  
## indirimli_urun 0.9766416  2.857153  
## bildirim     -2.6617688 -1.370443  
## site_tasarimi2 -7.7209861 -5.875925  
## site_tasarimi3 -10.3952213 -8.657547
```

### • Sabit terim - $\beta_0$ :

- ✓  $\beta_0$  için güven aralığı:  $P(3.35 < \beta_0 < 18.92) = 0.99 \rightarrow 0$  'ı içermiyor. Sabit terimin modele katkısı anlamlıdır.
- ✓ Sitedeki o gün kullanılan yapay zeka oranı '0' olduğunda, indirimli ürün oranı '0' olduğunda, gönderilen bildirim sayısı '0' olduğunda ve varsayılan site tasarımı kullanıldığında; sitedeki ortalama müşteri etkinliğinin 3.35 ile 18.92 arasında olduğu %99 güvenirlikle söylenebilir.

- **Yapay Zeka Oranı -  $\beta_1$  :**

- ✓  $\beta_1$  için güven aralığı:  $P(3.20 < \beta_1 < 4.60) = 0.99 \rightarrow 0$  'ı içermiyor. Günlük kullanılan yapay zeka oranının modele katkısı **anlamlıdır**.
- ✓ İndirimli ürün oranı, ortalama bildirim sayısı ve site tasarımı sabit tutulduğunda; günlük kullanılan yapay zeka oranı bir birim arttırıldığında müşteri etkinliğinin **3.20 ile 4.60** arasında olduğu %99 güvenirlikle **söylenebilir**.

- **İndirimli Ürün Oranı -  $\beta_2$  :**

- ✓  $\beta_2$  için güven aralığı:  $P(0.98 < \beta_2 < 2.86) = 0.99 \rightarrow 0$  'ı içermiyor. Günlük indirimli ürün oranının modele katkısı **anlamlıdır**.
- ✓ Yapay zeka oranı, ortalama bildirim sayısı ve site tasarımı sabit tutulduğunda; günlük indirimli ürün oranı bir birim arttırıldığında müşteri etkinliğinin **0.98 ile 2.86** arasında olduğu %99 güvenirlikle **söylenebilir**.

- **Bildirim Sayısı -  $\beta_3$ :**

- ✓  $\beta_3$  için güven aralığı:  $P(-2.66 < \beta_3 < -1.37) = 0.99 \rightarrow 0$  'ı içermiyor. Günlük bildirim sayısının modele katkısı **anlamlıdır**.
- ✓ Yapay zeka oranı, indirimli ürün oranı ve site tasarımı sabit tutulduğunda; günlük indirimli ürün oranı bir birim arttırıldığında müşteri etkinliğinin **-2.66 ile -1.37** arasında olduğu %99 güvenirlikle **söylenebilir**.

- **Site Tasarımı 2 -  $\beta_4$ :**

- ✓  $\beta_4$  için güven aralığı:  $P(-7.72 < \beta_4 < -5.87) = 0.99 \rightarrow 0$  'ı içermiyor. Müşteri etkinliği bakımından 2.site tasarımı ve 1. Site tasarımı arasında istatistiksel olarak farklılık bulunmaktadır.
- ✓ 2. Site tasarımı müşteri etkinliğinin **-7.72 ile -5.87** arasında olduğu %99 güvenirlikle **söylenebilir**.

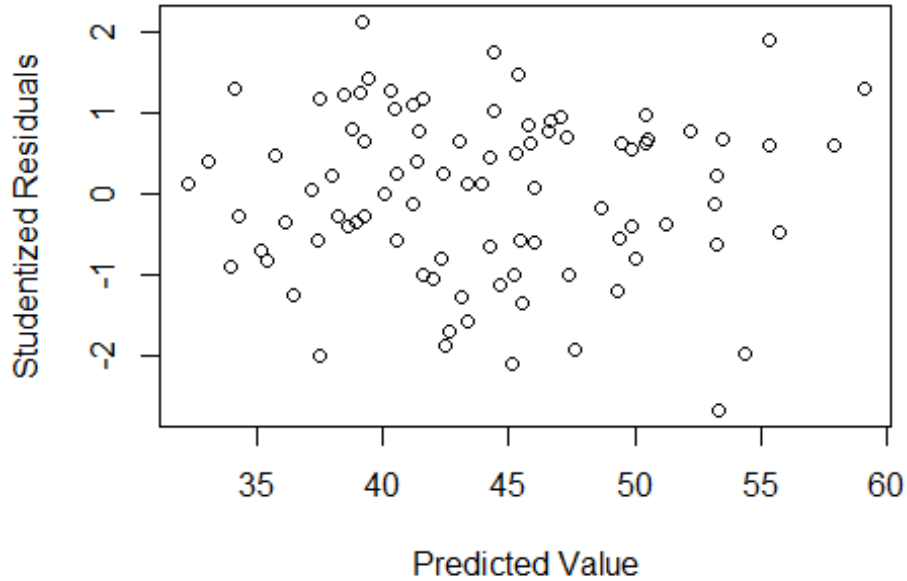
- **Site Tasarımı 3 -  $\beta_5$ :**

- ✓  $\beta_5$  için güven aralığı:  $P(-10.40 < \beta_5 < -8.66) = 0.99 \rightarrow 0$  'ı içermiyor. Müşteri etkinliği bakımından 3.site tasarımı ve 1. Site tasarımı arasında istatistiksel olarak farklılık bulunmaktadır.
- ✓ 3. Site tasarımı müşteri etkinliğinin **-10.4 ile -8.66** arasında olduğu %99 güvenirlikle **söylenebilir**.

## 9. Değişen Varyanslılık Sorununun İncelenmesi

- Grafik ile :

```
plot(predict(model2),inf2$stud.res , ylab="Studentized Residuals", xlab="Predicted Value")
```



- ✓ Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafiği incelediğimizde;
- ✓ Gözlemlerin rasgele bir görüntü oluşturduğu, megafon oluşumunu görmenin zor olduğu görülüyor.
- ✓ Daha kesin sonuç için Bruege-Pagan Testini yapmakta fayda vardır.

- Breusch-Pagan Testi ile :

**H<sub>0</sub> : Varyanslar homojendir.**

**H<sub>1</sub> : Varyanslar homojen değildir.**

```
library(lmtest)
```

```
bptest(model2)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: model2
```

```
## BP = 2.6971, df = 5, p-value = 0.7466
```

- ✓  $p = 0.75 > \alpha = 0.05$  H<sub>0</sub> REDDEDİLEMEZ.
- ✓ Değişen varyanslılık sorununun olmadığını %95 güven düzeyinde söylenebilir.

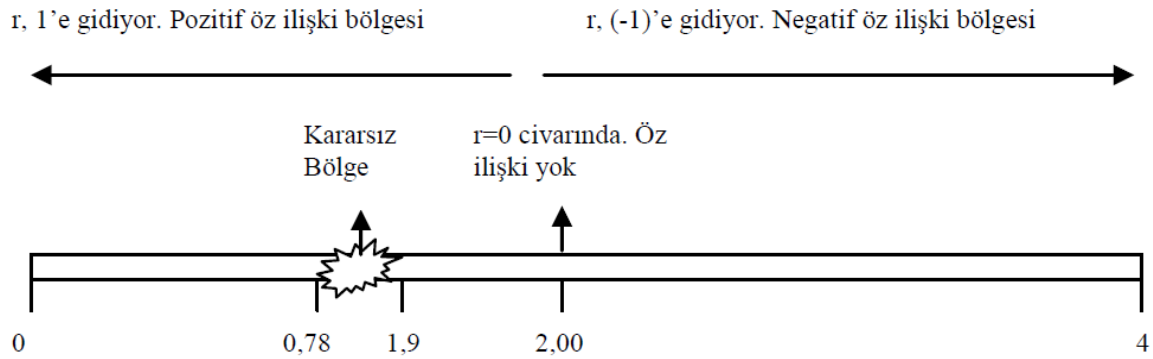
## 10. Öz İlişki Sorununun İncelenmesi

- Durbin-Watson Testi ile :

```
dwtest(model2)
```

```
##  
## Durbin-Watson test  
##  
## data: model2  
## DW = 2.029, p-value = 0.4549  
## alternative hypothesis: true autocorrelation is greater than 0
```

- ✓ Test değeri  $d = 2.03$  bulunmuştur.  $2 < d < 4$  olduğu için alternatif hipotez negatif (-) kurulur. Detay için aşağıdaki tabloya bakılabilir.



$H_0$  : Öz ilişki yoktur. ( $p = 0$ )

$H_1$  : Negatif öz ilişki vardır. ( $p < 0$ )

- ✓  $p = 0.4549 > \alpha = 0.05$   $H_0$  REDDEDİLEMEZ.
- ✓ Test sonucuna göre, negatif öz ilişkinin **olmadığını** söyleyebiliriz.

## 11. Çoklu Bağlantı Sorununun İncelenmesi

### 11.1 VİF ve Koşul Sayısı:

- VİF Testi:

```
library(DAAG)
```

```
vif(model2)
```

```
## yapay_zeka indirimli_urun bildirim site_tasarimi2 site_tasarimi3
## 4.3263 7.5948 3.1625 1.6313 1.5884
```

- ✓ VİF değerleri 10'dan büyük (>10) olan değişkenlerden şüphelenmemiz gerekiyor.
- ✓ Çıktıda da görüldüğü üzere bütün değişkenlerin VİF değeri **10'dan küçük çıkmıştır**.
- ✓ Çoklu bağlantıda sorun olmadığını söyleyebiliriz fakat koşul sayısına bakmakta fayda vardır.

- Koşul Sayısı :

```
library(perturb)
```

```
colldiag(model.matrix(model2),add.intercept=FALSE)
```

```
## Condition
```

```
## Index Variance Decomposition Proportions
```

```
## (Intercept) yapay_zeka indirimli_urun bildirim site_tasarimi2
## 1 1.000 0.000 0.000 0.002 0.001 0.007
## 2 2.092 0.000 0.000 0.000 0.000 0.234
## 3 3.147 0.000 0.000 0.095 0.001 0.098
## 4 5.306 0.002 0.002 0.050 0.005 0.637
## 5 13.947 0.006 0.012 0.181 0.871 0.005
## 6 68.176 0.992 0.987 0.671 0.121 0.018
```

```
site_tasarimi3
```

```
## 1 0.009
```

```
## 2 0.155
```

```
## 3 0.144
```

```
## 4 0.692
```

```
## 5 0.000
```

```
## 6 0.000
```

- ✓ Koşul sayısı, 10'dan küçükse çoklu bağlantı sorunu yok; 10 ile 30 arasındaysa zararsız çoklu bağlantı sorunu; 30'dan büyük ise de çoklu bağlantı sorunu vardır diyoruz.
- ✓ Test sonucuna bakıldığında **bir tane** çoklu bağlantı sorunu **çıkıştır**.
- ✓ Çoklu bağlantının yaratmış olduğu sorundan, **yapay zeka oranı ve indirimli ürün oranı** değişkenlerinin etkilendiği görülmektedir.
- ✓ Bu durumda çoklu bağlantı yapısı, yapay zeka oranı ve indirimli ürün oranı arasında oluşacaktır.

## 11.2 Özdeğer ve Özvektör:

- **Dummy (referans) değişkeni site tasarımı 1 seçilmiştir.**

```
library(fastDummies)

dummy <- dummy_cols(site_tasarimi)

site_tasarimi1 <- dummy$.data_1
site_tasarimi2 <- dummy$.data_2
site_tasarimi3 <- dummy$.data_3

ort1<-mean(yapay_zeka)
kt1<-sum((yapay_zeka-ort1)^2)
skx1<-(yapay_zeka-ort1)/(kt1^0.5)

ort2<-mean(indirimli_urun)
kt2<-sum((indirimli_urun-ort2)^2)
skx2<-(indirimli_urun-ort2)/(kt2^0.5)

ort3<-mean(bildirim)
kt3<-sum((bildirim-ort3)^2)
skx3<-(bildirim-ort3)/(kt3^0.5)

ort_site2<-mean(site_tasarimi2)
kt_site2<-sum((site_tasarimi2-ort_site2)^2)
skx_site2<-(site_tasarimi2-ort_site2)/(kt_site2^0.5)

ort_site3<-mean(site_tasarimi3)
kt_site3<-sum((site_tasarimi3-ort_site3)^2)
skx_site3<-(site_tasarimi3-ort_site3)/(kt_site3^0.5)

x<-cbind(skx1,skx2,skx3,skx_site2,skx_site3)
sm<- eigen (t(x)%*%x)
```

- **Özdeğer:**

```
signif(sm$values,3)
```

```
## [1] 2.5500      1.6000      0.4290      0.3350      0.0835
```

- ✓ Özdeğerlerin terslerinin toplamının büyük sayı olduğu görülmektedir.
- ✓ Özdeğerlerden sıfıra en yakın olan **0.0835'tir**. Yani bir tane çoklu bağlantı olduğu söylenebilir.

- **Özvektör:**

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

```
##      [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] -0.5680  0.00448  0.3810 -0.5280 -0.50300
## [2,] -0.6060 -0.07160  0.0405 -0.0693  0.78800
## [3,] -0.5460 -0.10800 -0.4820  0.5760 -0.35500
## [4,]  0.0958 -0.69900 -0.5210 -0.4800 -0.00515
## [5,] -0.0470  0.70300 -0.5900 -0.3920  0.02370
```

- ✓ Özdeğerlerden sıfıra çok yakın olan 0.0835 değerine denk gelen sütuna (5.sütun) bakıldığında;
- ✓ Yüksek çıkan 1. Ve 2. değişkenler, yani **yapay zeka oranı ve indirimli ürün oranı** değişkenlerinin katsayılarından,  
**-0.503 Yapay Zeka + 0.788 İndirimli Ürün = 0** şeklinde çoklu bağlantının yapısı yazılabilir.



## 12. Uyum Kestiriminin ( $\hat{y}_i$ ) Bulunması

➤ Uyum kestirimi için verideki 45. gözlem seçilmiştir.

	y3	yapay_zeka	indirimli_urun	bildirim	site_tasarimi
45	40.10482	10.566560	0.4971753	3.1566009	2

✓ Nitel değişken için;

- Site tasarımı 1 → 0
- Site tasarımı 2 → 1
- Site tasarımı 3 → 0 kodlaması yapılmıştır.

$$\hat{y}_i = 11.133 + 3.90 * 10.567 + 1.92 * 0.50 - 2.02 * 3.16 - 6.80 * 1 - 9.53 * 0 \mp 1.226$$

(2.95)            (0.27)            (0.36)            (0.25)            (0.35)            (0.33)

$$\hat{y}_i = 40.12 \mp 1.226$$

## 13. Ön Kestiriminin ( $\tilde{y}$ ) Bulunması

➤ Ön kestirimi için aşağıdaki değerler kullanılmıştır.

Yapay_zeka	İndirimli ürün	Bildirim	Site_tasarimi
25.789	0.657	4.332	3

✓ Nitel değişken için;

- Site tasarımı 1 → 0
- Site tasarımı 2 → 0
- Site tasarımı 3 → 1 kodlaması yapılmıştır.

$$\tilde{y} = 11.133 + 3.90 * 25.789 + 1.92 * 0.657 - 2.02 * 4.332 - 6.80 * 0 - 9.53 * 1 \mp 1.226$$

(2.95)            (0.27)            (0.36)            (0.25)            (0.35)            (0.33)

$$\tilde{y} = 94.69 \mp 1.226$$

## 14. $E(\hat{y}_i)$ ve $E(\tilde{y})$ İçin Güven Aralıklarının Bulunması:

- $E(\hat{y}_i)$  için Güven Aralığı:

```
predict(model2, newdata=data.frame(yapay_zeka =10.56,indirimli_urun=0.497,
bildirim= 3.16, site_tasarimi="2"), interval = "confidence")
```

```
##          fit          lwr          upr
## 1  40.06116 39.56322 40.5591
```

- ✓ Çıktıdan da görüleceği üzere  $E(\hat{y}_i) = 40.06$  çıkmıştır.
- ✓ Güven aralığı  $\rightarrow (39,563 - 40.559)$  çıkmıştır.
- ✓ Uyum kestirimi için (45. Gözlem) müşteri etkinliği 40.06 çıkmıştır.
- ✓ 45. Gözlem değerleri için müşteri etkinliğinin 39.563 ile 40.559 aralığında olduğunu %95 güven düzeyinde söyleyebiliriz.

- $E(\tilde{y})$  için Güven Aralığı:

```
predict(model2, newdata=data.frame(yapay_zeka =25.789, indirimli_urun=0.657
, bildirim= 4.332, site_tasarimi="3"), interval = "confidence", level = .95
)
```

```
##          fit          lwr          upr
## 1  94.61385 86.49798 102.7297
```

- ✓ Çıktıdan da görüleceği üzere  $E(\tilde{y}) = 94.61$  çıkmıştır.
- ✓ Güven aralığı  $\rightarrow (86.5 - 102.73)$  çıkmıştır.
- ✓ Ön kestirim (13. Başlıkta tahmin edilmek istenen değişken değerleri) için müşteri etkinliği 94.61 çıkmıştır.
- ✓ Tahmin edilmek istenen değişken değerleri için müşteri etkinliğinin 86.5 ile 102.73 aralığında olduğunu %95 güven düzeyinde söyleyebiliriz.

## 15. Değişken Seçimi ile En İyi Modelin Bulunması ve Yorumlanması

### 15.1 İleriye Doğru Seçim Yöntemi :

```
library(stats)
lm.null <- lm(y3 ~ 1)
forward <- step(lm.null,y3~yapay_zeka + indirimli_urun +bildirim+site_tasarimi, direction = "forward")

## Start: AIC=331.32
## y3 ~ 1
##
##              Df Sum of Sq    RSS    AIC
## + yapay_zeka    1   1944.63 1549.8 260.15
## + site_tasarimi  2   1575.64 1918.8 281.37
## + indirimli_urun 1   1454.00 2040.4 284.90
## + bildirim      1    417.63 3076.8 321.87
## <none>                      3494.4 331.32
##
## Step: AIC=260.15
## y3 ~ yapay_zeka
##
##              Df Sum of Sq    RSS    AIC
## + site_tasarimi  2   1321.66  228.16  91.721
## + bildirim      1     82.36 1467.46 257.233
## <none>                      1549.82 260.148
## + indirimli_urun 1      0.01 1549.81 262.147
##
## Step: AIC=91.72
## y3 ~ yapay_zeka + site_tasarimi
##
##              Df Sum of Sq    RSS    AIC
## + bildirim      1    58.613 169.54 66.998
## <none>                      228.16 91.721
## + indirimli_urun 1     0.244 227.91 93.624
##
## Step: AIC=67
## y3 ~ yapay_zeka + site_tasarimi + bildirim
##
##              Df Sum of Sq    RSS    AIC
## + indirimli_urun 1    43.368 126.18 42.408
## <none>                      169.54 66.998
##
## Step: AIC=42.41
## y3 ~ yapay_zeka + site_tasarimi + bildirim + indirimli_urun
```

```

forward
## Call:
## lm(formula = y3 ~ yapay_zeka + site_tasarimi + bildirim + indirimli_urun
##
## Coefficients:
## (Intercept)      yapay_zeka  site_tasarimi2  site_tasarimi3  bildirim
##      11.133      3.896      -6.798      -9.526      -2.016
## indirimli_urun
##      1.917

```

- ✓ Bağımlı değişken (müşteri etkinliği) y3 olduğu durumda, ilk yapay zeka modele alınmış daha sonra sırasıyla site tasarımı, bildirim ve indirimli ürün alınmıştır; **tüm değişkenler anlamlı çıkmıştır ve modelde yer almıştır.**

```

summary(forward)
## Call:
## lm(formula = y3 ~ yapay_zeka + site_tasarimi + bildirim + indirimli_urun
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0010 -0.7621  0.1386  0.8872  2.4322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.1329     2.9548   3.768 0.000305 ***
## yapay_zeka      3.8963     0.2666  14.614 < 2e-16 ***
## site_tasarimi2 -6.7985     0.3500 -19.423 < 2e-16 ***
## site_tasarimi3 -9.5264     0.3297 -28.898 < 2e-16 ***
## bildirim       -2.0161     0.2450  -8.230 2.10e-12 ***
## indirimli_urun  1.9169     0.3567   5.373 6.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.226 on 84 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9617
## F-statistic: 448.5 on 5 and 84 DF, p-value: < 2.2e-16

```

- ✓ Çıktıya göre en iyi model:

$$\begin{aligned}
 \text{Müşteri Etkinliği (y)} = & 11.133 + 3.90 \text{ Yapay Zeka} - 6.80 \text{ Site Tasarımı2} - 9.53 \text{ Site Tasarımı3} \\
 & (2.95) \quad (0.27) \quad (0.35) \quad (0.33) \\
 & - 2.02 \text{ Bildirim} + 1.92 \text{ İndirimli Ürün} \mp 1.226 \\
 & (0.245) \quad (0.357)
 \end{aligned}$$

- ✓ Modeldeki tüm değişkenlerin **p değerleri '0'a çok yakın** ve  $\alpha=0.05$  olduğu için, **tüm değişkenler %95 güven düzeyinde anlamlıdır.**
- ✓ Model için **p değeri = 0**  $\alpha=0.05$  olduğu için modelin de **anlamlı olduğu** söylenebilir.
- ✓ **R<sup>2</sup> = 0.9640** → alışveriş sitesinin müşteri etkinliğindeki değişimin %96.4 'ü; sitedeki yapay zekanın kullanım oranı, günlük uygulanan indirimli ürün oranı, günlük ortalama bildirim sayısı ve kullanılan site tasarımı tarafından açıklanabilmektedir.

## 15.2 Geriye Doğru Seçim Yöntemi :

```
backward<-step(model12,direction="backward")
```

```
## Start: AIC=42.41
## y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi
##              Df Sum of Sq    RSS    AIC
## <none>                126.18  42.408
## - indirimli_urun    1     43.37  169.54  66.998
## - bildirim          1    101.74  227.91  93.624
## - yapay_zeka        1    320.80  446.98 154.243
## - site_tasarimi     2   1263.88 1390.06 254.356

backward
## Call:
## lm(formula = y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi
## Coefficients:
## (Intercept) yapay_zeka indirimli_urun bildirim site_tasarimi2
##      11.133      3.896      1.917     -2.016     -6.798
site_tasarimi3
##      -9.526
```

- ✓ Bağımlı değişken (müşteri etkinliği) y3 olduğu durumda, ilk adımda değişkenlerin hepsi modele alınmış, sonra sırasıyla indirimli ürün, bildirim, yapay zeka ve site tasarımı çıkartılmışlardır; **tüm değişkenler anlamlı çıkmıştır ve modelde yer almıştır.**

```
summary(backward)
## Call:
## lm(formula = y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.1329     2.9548   3.768 0.000305 ***
## yapay_zeka     3.8963     0.2666  14.614 < 2e-16 ***
## indirimli_urun 1.9169     0.3567   5.373 6.79e-07 ***
## bildirim      -2.0161     0.2450  -8.230 2.10e-12 ***
## site_tasarimi2 -6.7985     0.3500 -19.423 < 2e-16 ***
## site_tasarimi3 -9.5264     0.3297 -28.898 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 1.226 on 84 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9617
## F-statistic: 448.5 on 5 and 84 DF, p-value: < 2.2e-16
```

- ✓ Çıktıya göre en iyi model:

**Müşteri Etkinliği (y) = 11.133 + 3.90 Yapay Zeka + 1.92 İndirimli Ürün – 2.02 Bildirim**

**(2.95) (0.27) (0.36) (0.25)**

**-6.80 Site Tasarımı2 -9.53 Site Tasarımı3 ± 1.226**

**(0.35) (0.33)**

- ✓ Modeldeki tüm değişkenlerin p değerleri '0'a çok yakın ve  $\alpha=0.05$  olduğu için, **tüm değişkenler %95 güven düzeyinde anlamlıdır.**
- ✓ Model için **p değeri = 0**  $\alpha=0.05$  olduğu için **modelin de anlamlı olduğu söylenebilir.**
- ✓ **R<sup>2</sup> = 0.9640** → Modeldeki değişkenler müşteri etkinliğinin %96'4ünü açıklamaktadır.

## 15.3 Adımsal Seçim Yöntemi :

```
library(MASS)

step.model <- stepAIC(model2, direction = "both", trace = FALSE)

summary(step.model)

##
## Call:
## lm(formula = y3 ~ yapay_zeka + indirimli_urun + bildirim + site_tasarimi
## )
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0010  -0.7621   0.1386   0.8872   2.4322
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.1329     2.9548   3.768 0.000305 ***
## yapay_zeka      3.8963     0.2666  14.614 < 2e-16 ***
## indirimli_urun  1.9169     0.3567   5.373 6.79e-07 ***
## bildirim       -2.0161     0.2450  -8.230 2.10e-12 ***
## site_tasarimi2  -6.7985     0.3500 -19.423 < 2e-16 ***
## site_tasarimi3  -9.5264     0.3297 -28.898 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.226 on 84 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.9617
## F-statistic: 448.5 on 5 and 84 DF, p-value: < 2.2e-16
```

- ✓ İlk adımda modele yapay zeka değişkeni alınarak başlanmış daha sonra sırasıyla diğer değişkenler alınmıştır ve teker teker çıkartılmıştır. En iyi model aşağıdaki gibi bulunmuştur:

$$\begin{aligned} \text{Müşteri Etkinliği (y)} = & \text{11.133} + \text{3.90 Yapay Zeka} + \text{1.92 İndirimli Ürün} - \text{2.02 Bildirim} \\ & \text{(2.95)} \quad \text{(0.27)} \quad \text{(0.36)} \quad \text{(0.25)} \\ & - \text{6.80 Site Tasarımı2} - \text{9.53 Site Tasarımı3} \mp \text{1.226} \\ & \text{(0.35)} \quad \text{(0.33)} \end{aligned}$$

- ✓ Modeldeki tüm değişkenlerin **p değerleri '0'a çok yakın** ve  $\alpha=0.05$  olduğu için, **tüm değişkenler %95 güven düzeyinde anlamlıdır.**
- ✓ Model için **p değeri = 0**  $\alpha=0.05$  olduğu için modelin de **anlamlı olduğu** söylenebilir.
- ✓  **$R^2 = 0.9640$**   $\rightarrow$  alışveriş sitesinin müşteri etkinliğindeki değişimin %96.4 'ü; sitedeki yapay zekanın kullanım oranı, günlük uygulanan indirimli ürün oranı, günlük ortalama bildirim sayısı ve kullanılan site tasarımı tarafından açıklanabilmektedir.

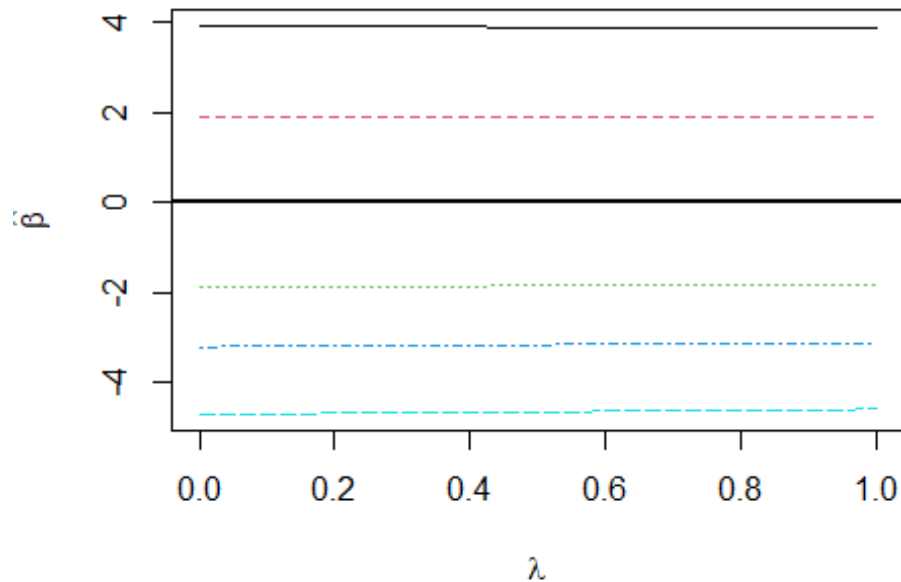
## 16. Ridge Regresyon Modelinin Kurulması

```
library(MASS)

ridge <- lm.ridge(model2 ,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=2)
```



- ✓ Grafikten de görüldüğü üzere; Lamda parametresine göre katsayılar **değişmemektedir**.
- ✓ Hızlı azalış veya artış gösteren katsayı yoktur. Yani **güçlü çoklu bağlantı sorunu yoktur**. 11. Seçenekte de güçlü bir çoklu bağlantı sorununun olmadığını görmüştük
- ✓ Sıfır eksenini etrafında seyreden değişken de **yoktur**. Yani modelde önemsiz-anlamsız değişken yoktur. Zaten kurduğumuz modelde de (model2), tüm değişkenlerin anlamlı olduğu sonucuna varmıştık.