



HACETTEPE ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

İST 366 – REGRESYON ÇÖZÜMLEMESİ ÖDEVİ

Prof. Dr. Meral ÇETİN

Arş. Gör. Serhan TUNÇEL

11.05.2022

ECEM ÇIRAKOĞLU / 21821665

Veri Seti Hakkında Bilgi

Veri setimde bir tane bağımlı değişken (y), üç tane bağımsız nicel değişken (x1, x2, x3) ve bir tane de nitel değişken bulunmaktadır.

Veri seti bir e-ticaret sitesine ait verileri içermektedir. Bağımlı y değişkeni, e-ticaret sitesindeki ürünlerin satış fiyatını (dolar), x1 değişkeni ürünlerin günlük satış miktarını, x2 değişkeni ürünlerin stok yenilenme süresini (gün), x3 değişkeni ürünlerin her birinin aldığı oy miktarı ve x4 değişkeni ise ürünün kategorisini göstermektedir (1:Giyim 2:Kozmetik 3:Aksesuar)

DEĞİŞKEN	AÇIKLAMA	DÜZEY
y	Ürün satış fiyatı(dolar)	
x1	Ürünün günlük satış miktarı	
x2	Ürünün stok yenilenme süresi (gün)	
x3	Ürünün aldığı oy miktarı	
x4	Ürünün kategorisi	1:Giyim 2:Kozmetik 3:Aksesuar

Tanımlayıcı İstatistikler

```
> summary(veri)
   fiyat      satis_miktari    stok_yenileme     oy_miktari    kategori
Min.   : 33.59  Min.   : 6.064  Min.   :0.1697  Min.   :1.250  1:50
1st Qu.: 52.30  1st Qu.: 8.342  1st Qu.:2.4489  1st Qu.:3.333  2:50
Median : 58.90  Median : 8.968  Median :2.9037  Median :4.013  3:50
Mean   : 60.30  Mean   : 8.978  Mean   :2.9332  Mean   :3.995
3rd Qu.: 65.84  3rd Qu.: 9.598  3rd Qu.:3.4783  3rd Qu.:4.702
Max.   :117.19  Max.   :11.448  Max.   :6.0065  Max.   :6.023
```

“Fiyat” değişkeni, y bağımlı değişkeni temsil etmektedir. Görüldüğü üzere sitedeki minimum fiyatın 33.59 \$, maksimum fiyatın 117.19 \$ ve ürünlerin sahip olduğu ortalama fiyatın 60.30 \$ olduğunu söyleyebiliriz.

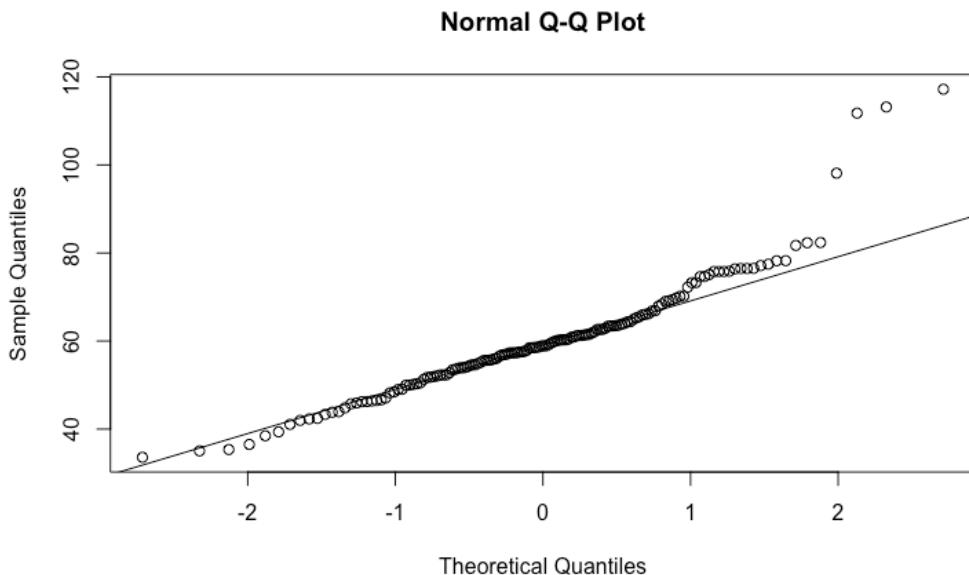
“Satis_miktari” değişkeni x1 değişkenini temsil etmektedir. Bir gün içerisinde minimum satış miktarının 6, maksimum satış miktarının 11 ve ortalama satış miktarının $8.97 \cong 9$ olduğunu söyleyebiliriz.

“Stok_yenileme” değişkeni x2 nicel değişkenini temsil etmektedir. Ürünlerin stok yenilenme süresini göstermektedir. Buna göre bir ürünün en kısa yaklaşık her gün stoğunun yenilendiğini ve en uzun 6 günde bir yenilendiğini ve ürünlerin ortalama stokların da $2.93 \cong 3$ günde bir yenilendiğini söyleyebiliriz.

“Oy_miktari” değişkeni x3 nicel değişkenini temsil etmektedir. Sitedeki minimum oya sahip ürünün oyu 1.250 iken sitedeki en yüksek oyun ise 6.023 olduğunu söyleyebiliriz. Ürünlerin ortalama oy miktarı $3.99 \cong 4$ ’ tür.

“Kategori” değişkeni x4 nitel değişkenini temsil etmektedir. Sitedeki 50 ürünün giyim sektöründen, 50 ürünün kozmetik sektöründen ve geriye kalan 50 ürününde aksesuar sektöründen olduğunu söyleyebiliriz.

Bağımlı Değişken İçin Normallik Varsayıımı



Q-Q plot grafiğine bakıldığında verilerin çoğunun 45 derecelik doğru etrafında dağıldığını söyleyebiliriz ve grafiğe bakınca 4 gözlem aykırı değer almış gibi gözükmektedir. Ancak grafik, bağımlı değişkenin normalliği konusunda bize net bir bilgi vermemektedir. Bir de normallik testini uygulayalım.

H_0 : Fiyatların dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Fiyatların dağılım ile normal dağılım arasında fark vardır.

```
> ks.test(fiyat, "pnorm")  
  
One-sample Kolmogorov-Smirnov test  
  
data: fiyat  
D = 1, p-value < 2.2e-16  
alternative hypothesis: two-sided
```

```
> shapiro.test(fiyat)  
  
Shapiro-Wilk normality test  
  
data: fiyat  
W = 0.91507, p-value = 1.033e-07
```

Kolmogorov-Simirnov testine göre elde edilen p değeri 0.05 oldukça küçük olduğundan H_0 hipotezi reddedilecektir. Aynı şekilde Shapiro-Wilks testinden elde edilen p değeri 0.05'ten oldukça küçük olduğundan H_0 hipotezi reddedilecektir. Böylece fiyatların dağılıminın normal dağılım göstermediğini %95 güven düzeyinde söyleyebiliriz.

Normallik koşulunu saglatabilmek için fiyat değişkenine logaritmik dönüşüm uygulayalım.

H_0 : Fiyatların dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Fiyatların dağılım ile normal dağılım arasında fark vardır.

```
> yeni_fiyat <- log(fiyat)  
> yeni_veri <- cbind(yeni_fiyat,satis_miktari,stok_yenileme,oy_miktari,kategori)  
> shapiro.test(yeni_fiyat)  
  
Shapiro-Wilk normality test  
  
data: yeni_fiyat  
W = 0.97568, p-value = 0.009177
```

```
> ks.test(yeni_fiyat, "pnorm")  
  
One-sample Kolmogorov-Smirnov test  
  
data: yeni_fiyat  
D = 0.99978, p-value < 2.2e-16  
alternative hypothesis: two-sided
```

Her iki testten elde edilen p değeri 0.05'ten küçük olduğundan H_0 hipotezi yine reddedilecektir. Böylece fiyatların dağılıminın hala normal dağılım göstermediğini %95 güven düzeyinde söyleyebiliriz.

Normallik sağlanmadığı için bir de karekök dönüşümünü uygulayalım.

H_0 : Fiyatların dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Fiyatların dağılım ile normal dağılım arasında fark vardır.

```

> yeni_fiyat <- sqrt(fiyat)
> yeni_veri <- cbind(yeni_fiyat,satis_miktari,stok_yenileme,oy_miktari,kategori)
> qqnorm(yeni_fiyat)
> qqline(yeni_fiyat)
> shapiro.test(yeni_fiyat)

Shapiro-Wilk normality test

data: yeni_fiyat
W = 0.9553, p-value = 9.264e-05

```

```

> ks.test(yeni_fiyat,"pnorm")

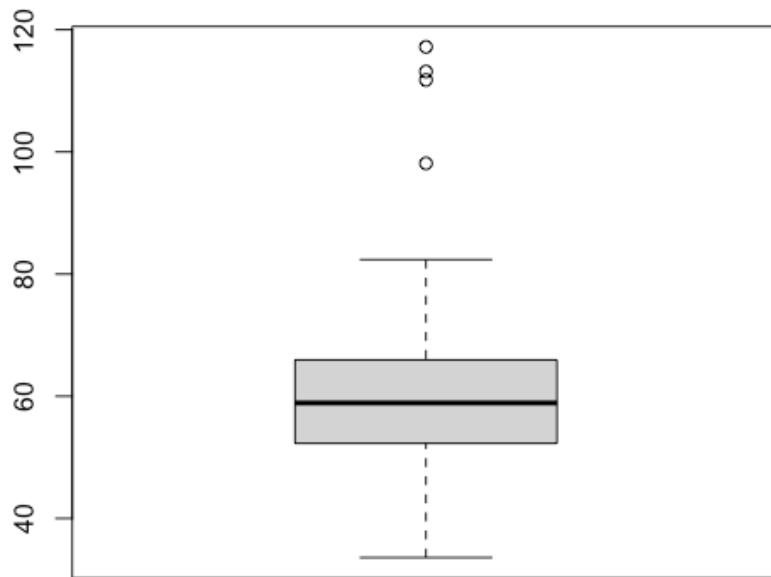
One-sample Kolmogorov-Smirnov test

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

```

Her iki testten elde edilen p değeri 0.05'ten küçük olduğundan H_0 hipotezi yine reddedilecektir. Böylece fiyatların dağılımının hala normal dağılım göstermediğini %95 güven düzeyinde söyleyebiliriz.

Dönüşümlerin uygulanmasına rağmen normallik elde edilemediğinden aykırı değer çıkarımını yapabiliriz.



Görüleceği üzere fiyat değişkenine ilişkin 4 adet aykırı değer bulunmaktadır. Aşamalı olarak 5., 14., 75. ve 13. Gözlemlerini veri setinden çıkartalım.

H_0 : Fiyatların dağılımı ile normal dağılım arasında fark yoktur.

H_1 : Fiyatların dağılım ile normal dağılım arasında fark vardır.

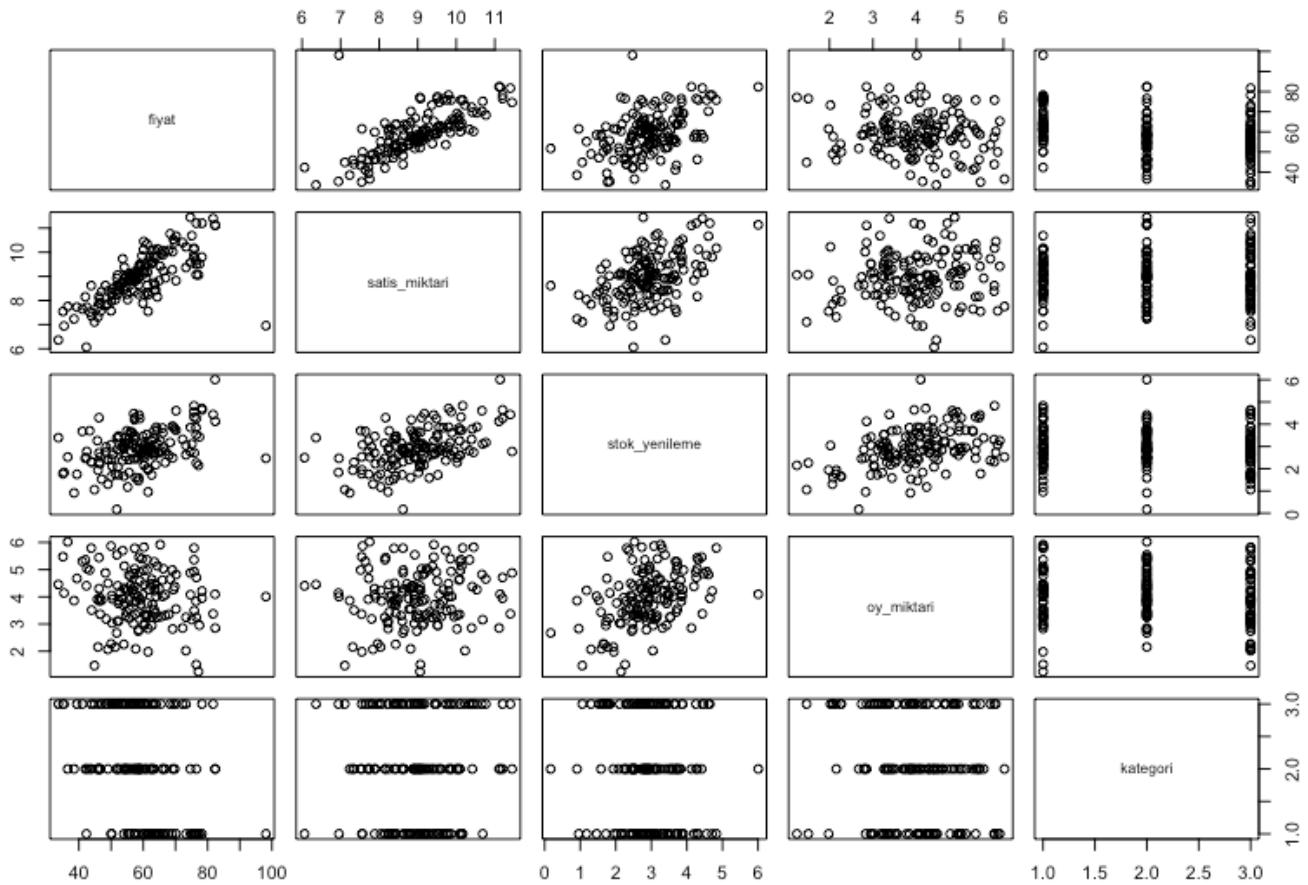
```
> yeni_veri <- veri[-c(5,14,75,13),]  
> shapiro.test(yeni_veri$fiyat)  
  
Shapiro-Wilk normality test  
  
data: yeni_veri$fiyat  
W = 0.98728, p-value = 0.2017
```

$p = 0.2017 > \alpha = 0.05$ olduğundan yokluk hipotezi reddedilemez. Fiyat değişkeninin aykırı değer çıkarımları sonucunda normal dağılım gösterdiğini söyleyebiliriz.

5., 14. , 75. ve 13. Gözlemlerin veri setinden çıkartılması sonucunda normal dağılımın elde edildiğini söyleyebiliriz. (n değerimiz artık 146)

*Burada bütün aşamalara yer vermedim 5. gözlem (en büyük değer) tek başına çıkartıldığında normallik koşulu sağlanmadı. Ardından 14.gözlem de çıkartıldığında (5. gözlemden sonraki en büyük gözlem) yine normallik sağlanmadı. Ancak 75. Gözlemi çıkartmamla birlikte normallik koşulu sağlandı.

Doğrusallık Varsayıımı



X-Y saçılım grafiklerinden görüleceği üzere ürünlerin fiyatı ile satış miktarları arasında doğrusal bir ilişki olabileceğini söyleyebiliriz. Oy miktarı ile değişkenler arasındaki ilişkinin diğer ilişkilere göre az olduğunu söyleyebiliriz.

*Nitel değişkenler için doğrusallık incelemesi yapılmamaktadır.

Çoklu Regresyon Modelinin Kurulması

```
> yeni_veri$kategori <- as.factor(yeni_veri$kategori)
> sonuc <- lm(fiyat~satis_miktari+stok_yenileme+oy_miktari+kategori,data=yeni_veri)
> summary(sonuc)
```

```
> summary(sonuc)

Call:
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +
    kategori, data = yeni_veri)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.767 -0.902  0.281  1.604 48.033 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.0883    4.1096   1.968   0.051 .  
satis_miktari 7.1747    0.4752  15.098 < 2e-16 *** 
stok_yenileme 2.8363    0.5929   4.784 4.28e-06 *** 
oy_miktari    -3.7217   0.4681  -7.950 5.43e-13 *** 
kategori2     -9.2599   1.0279  -9.009 1.32e-15 *** 
kategori3     -10.7619   1.0287 -10.461 < 2e-16 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.041 on 141 degrees of freedom
Multiple R-squared:  0.7991,    Adjusted R-squared:  0.7919 
F-statistic: 112.1 on 5 and 141 DF,  p-value: < 2.2e-16
```

$$\hat{y} = 8.08 + 7.17x_1 + 2.83x_2 - 3.72x_3 - 9.25T_2 - 10.76T_3 \pm 5.041$$

H_0 : Model anlamlı değildir.

H_1 : Model anlamlıdır.

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Modelin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

R^2 (Çoklu belirtme katsayısı) = 0.7991 tüm değişkenler birlikte fiyat üzerindeki değişimin %79.91'unu açıklamaktadır. Açıklanamayan kısım için farklı etkenler(değişkenler) olabilir.

→ T_1 değişkeni modelde bulunmadığından kılavuz değişken olduğunu söyleyebiliriz.

Artık İncelemesi

```
> influence.measures(sonuc)
Influence measures of
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +      kategori, data = yeni_veri) :

      dfb.1_  dfb.sts_  dfb.stk_  dfb.oy_m  dfb.ktg2  dfb.ktg3      dffit cov.r  cook.d    hat inf
1 -0.092221  0.058331  5.55e-02  8.46e-02 -0.126911 -1.06e-01  0.242018 1.032 9.74e-03 0.0433
2  0.017520 -0.051409 -6.21e-03  1.41e-01 -0.109869 -8.81e-02  0.216150 1.052 7.79e-03 0.0470
3  0.059653 -0.005531 -8.15e-02  1.87e-02 -0.113458 -1.19e-01  0.185518 1.026 5.73e-03 0.0299
4  0.069552 -0.008631 -5.22e-02 -2.42e-02 -0.109838 -1.19e-01  0.177628 1.024 5.25e-03 0.0278
6  0.019132  0.012435 -2.88e-02  1.67e-02 -0.111770 -1.12e-01  0.157148 1.018 4.11e-03 0.0220
7  0.038752  0.027803 -1.71e-01  7.79e-02 -0.127288 -1.32e-01  0.248330 1.037 1.03e-02 0.0467
8  0.019044  0.001042 -4.47e-02  5.71e-02 -0.112235 -1.08e-01  0.165211 1.023 4.55e-03 0.0251
9  0.074195 -0.057433  1.05e-01 -7.79e-02 -0.091461 -9.38e-02  0.179073 1.041 5.35e-03 0.0350
10 0.119596 -0.088676  7.99e-02 -8.37e-02 -0.089212 -9.58e-02  0.180134 1.044 5.41e-03 0.0362
11 0.000151  0.009737  7.35e-02 -3.16e-02 -0.107853 -1.05e-01  0.179663 1.025 5.38e-03 0.0286
```

```
37 -0.085905  0.058469 -7.48e-02  1.92e-01 -0.132671 -1.08e-01  0.258126 1.043 1.11e-02 0.0512 *
38  0.256964 -0.235354  3.92e-02 -2.28e-01  0.314070  2.77e-01 -0.547605 0.788 4.77e-02 0.0379 *
39  0.063635 -0.319133  3.71e-01  8.18e-02  0.296206  3.44e-01 -0.618116 0.816 6.10e-02 0.0511 *
40 -0.053419  0.152225 -3.25e-01 -9.84e-02  0.317563  2.69e-01 -0.615806 0.739 5.97e-02 0.0394 *
41 -0.078802  0.023020  4.44e-02 -8.85e-02  0.312187  3.03e-01 -0.442758 0.746 3.10e-02 0.0224 *
42  0.216892 -0.386084  9.21e-02  1.68e-01  0.274116  3.05e-01 -0.582412 0.841 5.45e-02 0.0507 *
43 -0.051938 -0.163651  2.01e-01  1.38e-01  0.294591  3.36e-01 -0.522338 0.781 4.34e-02 0.0341 *
44 -0.113213 -0.036140  1.27e-01  4.78e-02  0.302512  3.23e-01 -0.458522 0.758 3.33e-02 0.0249 *
45 -0.044903  0.075009 -1.37e-01 -1.14e-01  0.315218  2.83e-01 -0.498442 0.741 3.92e-02 0.0273 *
46 -0.092956 -0.041613 -4.76e-02  1.80e-01  0.292633  3.16e-01 -0.465509 0.758 3.43e-02 0.0256 *
47 -0.190795  0.241049 -4.15e-01  8.51e-02  0.306879  2.80e-01 -0.626961 0.732 6.18e-02 0.0398 *
48 -0.171985  0.073651  1.21e-01 -7.10e-02  0.311775  3.14e-01 -0.475363 0.747 3.57e-02 0.0256 *
49  0.152753 -0.331907  2.01e-01  7.80e-02  0.288818  3.18e-01 -0.536374 0.813 4.60e-02 0.0400 *
50  0.119395 -0.198039  1.20e-01 -8.53e-02  0.306052  3.02e-01 -0.459947 0.776 3.36e-02 0.0267 *
51 -0.000139  0.013832 -3.27e-02 -4.21e-04  0.011892 -4.29e-03  0.040838 1.160 2.80e-04 0.1004 *
52 -0.001595 -0.000941  8.72e-05  5.76e-03  0.005014  6.25e-04  0.009901 1.081 1.65e-05 0.0345
53 -0.008798  0.008231  7.58e-03 -6.50e-03  0.019293  6.42e-05  0.031663 1.072 1.68e-04 0.0273
54 -0.030560  0.027292  6.46e-03 -8.30e-04  0.030082  1.61e-04  0.057193 1.076 5.49e-04 0.0340
55 -0.047057  0.061780  4.63e-02 -8.08e-02  0.055741 -4.23e-03  0.143652 1.115 3.46e-03 0.0739
56 -0.025347  0.016876 -1.71e-03  1.92e-02  0.021756  1.51e-03  0.044485 1.080 3.32e-04 0.0357
57  0.000469 -0.000451  3.74e-04 -3.62e-04  0.000712  1.56e-05  0.001105 1.073 2.05e-07 0.0271
58 -0.000203 -0.000260  5.85e-04 -4.91e-04  0.008418  1.29e-04  0.012127 1.066 2.47e-05 0.0206
59  0.000246 -0.000513  7.46e-04 -7.85e-05  0.001118  9.24e-05  0.001749 1.072 5.14e-07 0.0258
60  0.001505 -0.001595  1.62e-04  1.59e-04  0.004756  6.81e-05  0.006970 1.067 8.16e-06 0.0218
61 -0.000681  0.007179 -6.44e-03 -8.81e-03  0.018671 -1.87e-03  0.030301 1.071 1.54e-04 0.0266
62 -0.012969  0.031248 -4.71e-02  1.82e-03  0.029569 -6.15e-03  0.068531 1.091 7.88e-04 0.0470
63 -0.008156  0.008855 -1.07e-02  7.92e-03 -0.008054 -6.26e-04  0.015924 1.098 4.26e-05 0.0495
64 -0.016520  0.011872  6.88e-03  2.93e-03  0.020814  9.19e-04  0.036863 1.074 2.28e-04 0.0297
65  0.003781 -0.003391  6.11e-04 -1.20e-03  0.002849  1.26e-05  0.005431 1.089 4.95e-06 0.0417
66 -0.029811  0.050220 -3.85e-02 -1.24e-02  0.038095 -7.00e-03  0.082385 1.083 1.14e-03 0.0427
67 -0.018637  0.018817 -1.22e-02  9.81e-03  0.022771 -8.06e-04  0.040163 1.071 2.71e-04 0.0274
68  0.000434 -0.000489  8.99e-05  7.22e-05  0.000542  2.83e-05  0.000920 1.076 1.42e-07 0.0302
69 -0.002621  0.000755 -1.23e-03  5.24e-03  0.009198  3.97e-04  0.014707 1.069 3.63e-05 0.0236
70 -0.006649  0.010688 -3.91e-03 -6.02e-03  0.020330 -1.33e-03  0.032081 1.068 1.73e-04 0.0245
71  0.003835 -0.003280 -1.45e-03  3.77e-04  0.005066 -7.16e-05  0.008664 1.076 1.26e-05 0.0299
72 -0.082327  0.070971  1.36e-02  7.18e-03  0.048124  1.03e-03  0.118620 1.092 2.36e-03 0.0540
73  0.001467 -0.001272  2.67e-03 -2.87e-03  0.002816  4.32e-05  0.004943 1.084 4.10e-06 0.0368
74 -0.011828  0.011505 -1.10e-02  1.03e-02  0.018296 -4.62e-04  0.031580 1.070 1.67e-04 0.0262
76 -0.000720  0.001898 -5.92e-04 -2.18e-03  0.012419 -3.43e-04  0.018016 1.066 5.45e-05 0.0211
77 -0.024112  0.007580  8.13e-02 -4.08e-02  0.034522  5.37e-03  0.107885 1.179 1.95e-03 0.1182 *
```

Gözlem Uzaklığı (h_{ii})

```
> inf <- ls.diag(sonuc)
> inf
$std.dev
[1] 2.880771

$hat
[1] 0.04328186 0.04701467 0.02993243 0.02775433 0.02204163 0.04673023 0.02511659 0.03495020 0.03621950 0.02857709
[11] 0.04969352 0.03193021 0.03010657 0.05825851 0.02676185 0.03089123 0.05934384 0.02839835 0.03097671 0.02652355
[21] 0.07648153 0.06661759 0.03934219 0.08515050 0.02330644 0.03673921 0.02356731 0.03043792 0.06062942 0.06004275
[31] 0.03254501 0.06235022 0.04240529 0.05116975 0.03791437 0.05109752 0.03941944 0.02239823 0.05072669 0.03412811
[41] 0.02485953 0.02729092 0.02561251 0.03980090 0.02560872 0.04002123 0.02670093 0.10036890 0.03451046 0.02727366
[51] 0.03398935 0.07390546 0.03565207 0.02707107 0.0205002 0.02583342 0.02178303 0.02657632 0.04704455 0.04951741
[61] 0.02972875 0.04165543 0.04266058 0.02738780 0.03021280 0.02358545 0.02449071 0.02989754 0.05401365 0.03678180
[71] 0.02615727 0.02105664 0.11820996 0.06403925 0.02519645 0.02675025 0.02120100 0.03952998 0.02415842 0.04454538
[81] 0.04283313 0.03868070 0.05829676 0.08963227 0.03710948 0.04052968 0.02243426 0.06005337 0.06496607 0.04202815
[91] 0.03093548 0.03455977 0.02913893 0.06442054 0.04308787 0.02077998 0.06183164 0.03786568 0.02749182 0.04783518
[101] 0.03082623 0.03032140 0.02945925 0.02230250 0.04471401 0.03278376 0.02888354 0.06373199 0.07276326 0.02937871
[111] 0.02460962 0.05006754 0.04200526 0.07177944 0.08362773 0.05101113 0.03152428 0.02754334 0.02123591 0.03323274
[121] 0.05750342 0.03971590 0.03815158 0.02419597 0.10211069 0.05175373 0.02192605 0.02226014 0.03096478 0.02781114
[131] 0.04980184 0.03094951 0.07202654 0.04308452 0.04081972 0.05310115 0.06472270 0.02848803 0.07659090 0.02359478
[141] 0.03298297 0.04278275 0.04978422 0.06424773 0.04164334 0.02642655
```

$$h_{ii} > \frac{2(k+1)}{n} = \frac{2*(4+1)}{146} = 0.068$$

```
> x <-which(inf$hat>0.06849)
> x
[1] 21 24 48 52 73 84 109 114 115 125 133 139
```

Çıktıda verilen gözlemlerin üç değerler olduğunu söyleyebiliriz.

Standartlaştırılmış Artıklar

```
$std.res
[1] 1.136656374 0.973337873 1.055689573 1.050923322 1.046405772 1.120567897 1.029064365 0.941367262 0.929660544
[10] 1.047136989 1.021648835 1.085049432 1.004760975 1.111884583 0.934122559 0.997850151 1.197463028 1.148278493
[19] 1.220049724 0.957134207 1.225764380 1.197136424 0.895231470 1.119708052 1.014054321 1.093770784 1.039644834
[28] 1.069667621 1.045512616 1.140140250 1.075302185 0.981789959 1.052077463 1.110590336 -2.695602293 -2.607515843
[37] -2.954171122 -2.849217433 -2.472677271 -2.714389645 -2.800194987 -2.895628406 -2.799712922 -2.990199040 -2.855764231
[46] -2.573288294 -2.712687965 0.122695419 0.052555728 0.189749905 0.305895518 0.509865297 0.232148191 0.006650568
[55] 0.084018816 0.010778884 0.046878495 0.184017305 0.309439821 -0.070013406 0.211317284 0.026141316 0.391461358
[64] 0.240148597 0.005230574 0.094966203 0.203167625 0.049529814 0.497761453 0.025384446 0.193353569 0.123276461
[73] 0.295623714 0.166872833 0.022475049 0.070947620 0.096406081 0.009533164 0.164581242 0.374926631 0.028448883
[82] 0.234023455 -0.062794036 0.783335233 0.056740224 0.004464580 -0.579762593 -0.803611255 -0.647999861 -0.648533899
[91] -0.545068912 -0.510942094 -0.480234509 -0.656610739 -0.762409403 -0.629393175 0.227742262 -0.009031404 -0.084237965
[100] -0.077186572 -0.026356551 0.002049984 0.134577185 -0.073981114 0.073908372 -0.202474777 0.134381964 -0.227565264
[109] 0.128759581 -0.160448752 -0.136641106 -0.007178009 -0.043845112 0.391658193 0.271922348 0.187488176 -0.163078618
[118] -0.138831860 -0.116292086 -0.202899807 0.208170506 0.003071029 0.157443650 -0.035283927 -0.136582707 -0.089775437
[127] -0.114324516 0.025291677 -0.104788100 -0.019576942 -0.197797880 0.055672058 -0.067124514 -0.048594898 -0.125862646
[136] 0.115628401 0.285884355 -0.065685340 0.011113706 0.037993139 0.106549871 0.201952969 0.308574472 -0.048480371
[145] -0.203151584 -0.118852982
```

```
> x <-which((inf$std.res< (-2)) | inf$std.res>2)
> x
[1] 35 36 37 38 39 40 41 42 43 44 45 46 47
```

Çıktıda verilen değerlerin [-2,2] aralığının içerisinde olmadığını birim normal sapmaya göre aykırı değer olduklarını söyleyebiliriz.

Student Türü Artıklar

```
$stud.res
[1] 1.137852106 0.973153714 1.056124575 1.051318358 1.046763408 1.121599890 1.029282735 0.940982054 0.929206977
[10] 1.047500648 1.021809714 1.085742233 1.004795473 1.112830841 0.933694719 0.997834735 1.199336482 1.149596502
[19] 1.222199350 0.956845497 1.227986034 1.199005983 0.894592738 1.120731474 1.014157589 1.094544067 1.039947366
[28] 1.070222852 1.045862913 1.141372282 1.075907165 0.981662533 1.052482129 1.111523972 -2.758497347 -2.663672578
[37] -3.039873794 -2.925097163 -2.519460008 -2.778787171 -2.871755444 -2.975753684 -2.871231735 -3.079459905 -2.932233551
[46] -2.626959873 -2.776948442 0.122263010 0.052368209 0.189095330 0.304902988 0.508513429 0.231362140 0.006626775
[55] 0.083720321 0.010740324 0.046711138 0.183381101 0.308438195 -0.069764131 0.210594816 0.026047851 0.390274429
[64] 0.239338688 0.005211861 0.094629478 0.202470576 0.049353037 0.496420014 0.025293683 0.192687513 0.122842067
[73] 0.294658006 0.166292329 0.022394677 0.070695052 0.096064345 0.009499059 0.164008265 0.373772902 0.028347180
[82] 0.233231783 -0.062570251 0.782248758 0.056537867 0.004448607 -0.578383036 -0.802589301 -0.646651913 -0.647186444
[91] -0.543695956 -0.509589375 -0.478910935 -0.655271248 -0.761263627 -0.628030463 0.226969486 -0.008999094 -0.083938703
[100] -0.076912048 -0.026262317 0.002042649 0.134104365 -0.073717863 0.073645377 -0.201779903 0.133909805 -0.226793024
[109] 0.128306498 -0.159889395 -0.136161307 -0.007152329 -0.043688541 0.390470883 0.271021037 0.186840834 -0.162510587
[118] -0.138344667 -0.115881610 -0.202203598 0.207457818 0.003060041 0.156894234 -0.035157843 -0.136103105 -0.089456811
[127] -0.113920800 0.025201245 -0.104417281 -0.019506925 -0.197117738 0.055473487 -0.066885430 -0.048421442 -0.125419427
[136] 0.115220204 0.284944695 -0.065451338 0.011073948 0.037857401 0.106172959 0.201259734 0.307575059 -0.048307322
[145] -0.202454585 -0.118433722
```

```
> x <-which((inf$stud.res< (-3)) | inf$stud.res>3)
> x
[1] 37 44
```

Çıktıda verilen değerlerin [-3,3] aralığının içerisinde olmadığını student türü standartlaştırmaya göre aykırı değer olduklarını söyleyebiliriz.

Cook Uzaklılığı (D_i)

Cook uzaklığı bize değişkenlerin model üzerinde etkinliği hakkında bilgi vermektedir.

```
$cooks
[1] 9.741572e-03 7.789743e-03 5.731407e-03 5.254671e-03 4.113129e-03 1.025905e-02 4.547176e-03 5.348930e-03 5.413298e-03
[10] 5.376076e-03 9.096788e-03 6.472066e-03 5.222897e-03 1.274664e-02 3.999019e-03 5.289835e-03 1.507710e-02 6.423151e-03
[19] 7.930578e-03 4.160070e-03 2.073833e-02 1.704768e-02 5.470276e-03 1.944894e-02 4.089675e-03 7.604791e-03 4.347969e-03
[28] 5.986677e-03 1.175855e-02 1.383942e-02 6.482813e-03 1.068275e-02 8.169255e-03 1.108616e-02 4.772550e-02 6.102123e-02
[37] 5.968934e-02 3.099929e-02 5.445389e-02 4.338962e-02 3.331585e-02 3.920754e-02 3.433967e-02 6.177042e-02 3.572300e-02
[46] 4.601023e-02 3.364561e-02 2.799240e-04 1.645478e-05 1.682537e-04 5.487265e-04 3.457648e-03 3.320706e-04 2.051118e-07
[55] 5.468493e-05 5.135054e-07 8.156036e-06 1.540845e-04 7.878398e-04 4.256228e-05 2.280354e-04 4.950552e-06 1.138118e-03
[64] 2.706613e-04 1.420568e-07 3.630755e-05 1.727141e-04 1.260085e-05 2.357816e-03 4.101025e-06 1.673620e-04 5.448045e-05
[73] 1.952613e-03 3.175475e-04 2.176067e-06 2.305834e-05 3.355216e-05 6.233983e-07 1.117631e-04 1.092280e-03 6.036307e-06
[82] 3.672774e-04 4.068327e-05 1.006912e-02 2.067944e-05 1.403304e-07 1.285627e-03 6.876619e-03 4.862480e-03 3.075400e-03
[91] 1.580723e-03 1.557534e-03 1.153645e-03 4.947757e-03 4.362227e-03 1.401063e-03 5.697260e-04 5.350193e-07 3.343295e-05
[100] 4.988472e-05 3.682516e-06 2.190136e-08 9.162195e-05 2.080844e-05 4.261341e-05 2.315932e-04 8.951790e-05 5.875137e-04
[109] 2.168350e-04 1.298687e-04 7.851233e-05 4.526060e-07 1.404856e-05 1.977024e-03 1.124649e-03 3.149201e-04 1.442777e-04
[118] 9.098574e-05 4.890372e-05 2.358611e-04 4.406574e-04 6.501015e-08 1.638721e-04 5.144972e-06 3.535807e-04 7.331356e-05
[127] 4.883332e-05 2.427216e-06 5.847915e-05 1.827286e-06 3.417616e-04 1.652273e-05 5.828651e-05 1.772058e-05 1.123601e-04
[136] 1.249620e-04 9.426398e-04 2.108627e-05 1.707456e-06 5.813597e-06 6.453722e-05 3.038138e-04 8.314540e-04 2.689537e-05
[145] 2.988874e-04 6.390585e-05
```

$$n > 50 \quad \frac{4}{146} = 0.02739$$

```
> x <- which(inf$cooks>0.02739)
> x
[1] 35 36 37 38 39 40 41 42 43 44 45 46 47
```

Çıktıda verilen gözlemlerin etkin bir aykırı değer olduğunu söyleyebiliriz.

35 36 37 38 39 40 41 42 43 44 45 46 47 21 24 48 52 73 84 109 114

115 125 133 139 gözlemleri çıkartıldığında n = 121

Artık İncelenmesi Sonunda Elde Edilen 1. Model

```
> yeni_sonuc <- lm(fiyat~satis_miktari+stok_yenileme+oy_miktari+kategori,data=artik_temizlenmis_veri)
> summary(yeni_sonuc)

Call:
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +
    kategori, data = artik_temizlenmis_veri)

Residuals:
    Min      1Q      Median      3Q      Max 
-1.69460 -0.08884  0.01319  0.30562  0.84455 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.60496   0.53375   1.133   0.259    
satis_miktari 8.40446   0.06384 131.644 <2e-16 ***  
stok_yenileme 2.39226   0.07622   31.387 <2e-16 ***  
oy_miktari    -3.76933   0.05769  -65.333 <2e-16 ***  
kategori2     -11.40071   0.12190  -93.523 <2e-16 ***  
kategori3     -12.97104   0.12502 -103.748 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5256 on 115 degrees of freedom
Multiple R-squared:  0.9973,    Adjusted R-squared:  0.9972 
F-statistic: 8526 on 5 and 115 DF,  p-value: < 2.2e-16
```

Elde Edilen Model Üzerinden Tekrar Aykırı Değer İncelemesi

Gözlem Uzaklığı (h_{ii})

$$h_{ii} > \frac{2(k+1)}{n} = \frac{2*(4+1)}{121} = 0.08264$$

```
$hat
[1] 0.05981177 0.06335146 0.04429613 0.04138853 0.03274881 0.06959618 0.03610954 0.05174165 0.05329760 0.04239346 0.07266840 0.04533102 0.04172915
[14] 0.07887382 0.03824338 0.04237534 0.08887821 0.04504825 0.05023098 0.03745187 0.0972166 0.05739121 0.03363090 0.05680320 0.03441176 0.04600865
[27] 0.08688432 0.08890635 0.05013553 0.08684336 0.05950276 0.06987377 0.03974315 0.03555544 0.04714166 0.04544642 0.03099155 0.02252712 0.02950439
[40] 0.02323288 0.03339284 0.06577008 0.06473482 0.03842521 0.05021037 0.06265634 0.03566707 0.03431913 0.02628807 0.03137777 0.03367687 0.07916143
[53] 0.04689267 0.03249393 0.02429495 0.08219303 0.02910608 0.02988816 0.02394340 0.04721904 0.02919544 0.06477073 0.05252676 0.04972992 0.07674825
[66] 0.04594773 0.04896020 0.02571589 0.07815344 0.08140091 0.05055870 0.03755589 0.04445786 0.03872305 0.08164315 0.05426418 0.02268499 0.08726526
[79] 0.05162702 0.03274365 0.06864720 0.03723089 0.03763653 0.03661670 0.02500741 0.05490788 0.04230296 0.03679165 0.09344992 0.03562947 0.02929835
[92] 0.06195276 0.05104206 0.07173553 0.03878396 0.03361724 0.02492537 0.04284057 0.07416855 0.04768801 0.04870857 0.02734205 0.07551458 0.02657582
[105] 0.02636335 0.04114560 0.03172432 0.06603315 0.03939150 0.05194633 0.05209760 0.06859037 0.08665333 0.03278816 0.02755338 0.04255926 0.05491181
[118] 0.06922287 0.09144371 0.05527861 0.03398589
```

```
> #Gozlem Uzakligi
> x3 <- which(inf2$hat>0.0826)
> x3
[1] 17 21 27 28 30 78 89 113 119
```

Çıktıda verilen gözlemlerin üç değerler olduğunu söyleyebiliriz.

Standartlaştırılmış Artıklar

```
$std.res
[1] -0.0002222483 0.2960901980 0.0256761012 -0.0482894533 -0.1718556849 0.1271042916 -0.0821580295 -0.2399476589 0.0228989784 -0.2462009527
[14] -0.2350900523 -0.106319883 0.0128518739 0.0689930850 -0.0213532811 -0.1468031341 0.1855014509 -0.1991920620 -0.1218666718 0.1031746158
[21] -0.1004797695 0.3107704575 -0.0845133896 -0.2769007503 -0.1263066385 -0.2494352696 0.4592470592 0.2869255600 -0.2804424053 0.9991796200
[31] -0.1322404521 0.0143228298 0.7901443650 0.6113660675 0.8509511516 0.7943976398 0.6963761577 0.5845488732 0.6179556341 0.6895848563
[41] 0.5816844357 0.7429122509 0.7668895676 0.7038724549 0.4524831653 0.8432700059 0.6809975545 1.0943700302 0.6369555017 0.5907667660
[51] 1.0375125299 1.4128404936 0.5013105040 0.6414052895 0.5585367988 1.6743429251 0.5729872852 0.8202094152 0.5606753342 1.3707200354
[61] 0.6018533847 0.8175710627 0.7065739474 0.7743536780 1.6721652360 0.5909826380 1.2538666939 -3.2563725354 -2.7193370766 -2.5193844030
[71] -2.802868715 -3.1947337388 -3.2820449379 -2.3695110835 -2.7163537358 -3.2611254427 0.1673824216 0.0103282155 -0.0275958007
[81] 0.4497060469 -0.2629904413 -0.1046618212 -0.1216693283 -0.2914410044 -0.0764741742 -0.1288027149 -0.1221869249 0.7747234713 -0.2190038250
[91] -0.2706013019 0.2791940861 0.1520716759 0.041835221 -0.0822517576 -0.2739816385 -0.2185581170 -0.1243378158 0.0211668282 -0.0359692546
[101] 0.0655038673 -0.2456017982 1.1542256397 -0.1217424117 -0.2409545590 -0.0554612861 -0.2491081220 -0.0279856181 -0.1626913627 0.1036507226
[111] -0.2260985699 0.1225807167 0.4803755287 -0.3208457066 -0.2434013740 -0.0692880272 0.0363356439 0.3377890858 0.1196960059 0.1900085082
[121] 0.0918316356
```

```
> #Birim normal sapma
> x1 <- which((inf2$std.res< (-2)) | inf2$std.res>2)
> x1
[1] 68 69 70 71 72 73 74 75 76 77
```

Çıktıda verilen değerlerin [-2,2] aralığının içerisinde olmadığını birim normal sapmaya göre aykırı değer olduklarını söyleyebiliriz.

Student Türü Artıklar

```
$stud.res
[1] -0.0002212799 0.2949124721 0.0255642955 -0.0480795282 -0.1711288309 0.1265593476 -0.0818024415 -0.2389619561 0.0227992522 -0.2451928040
[14] -0.2341219544 -0.1058619211 0.0127958833 0.0686938818 -0.0212602802 -0.1461771635 0.1847208005 -0.1983583405 -0.1213434954 0.1027298058
[21] -0.1000463389 0.3095463390 -0.0841477510 -0.2757861582 -0.1257650039 -0.2484156092 0.4576658411 0.2857776434 -0.2793159560 0.9991724326
[31] -0.1316742502 0.0142604334 0.7888456659 0.6096937581 0.8499233579 0.7931153171 0.6948083295 0.5828683844 0.6162870769 0.6880040436
[41] 0.5809537255 0.7414565182 0.7655079118 0.7023199382 1.4596046347 0.842205365 0.6794015175 1.0953199392 0.6353017244 0.5890871754
[51] 1.0378602296 1.4190538570 0.4996724057 0.6397558378 0.5568858832 1.6877457995 0.5713066937 0.8190346680 0.5589968350 1.3760345093
[61] 0.6001768828 0.8163846667 0.7050272026 0.7729974558 1.6854959392 0.5893030893 1.2570252535 -3.4028632821 -2.7989737688 -2.5806350774
[71] -2.8902895984 -3.3321117323 -3.3448526674 -3.4324511845 -2.4189754252 -2.7956924070 -3.408357122 0.1666733871 0.0102832169 -0.0274756480
[81] 0.4481407509 -0.2619232835 -0.1042107396 -0.1211469738 -0.2902783211 -0.0761428880 -0.1282507316 -0.1216624150 0.7733685377 -0.2180950403
[91] -0.2695080225 0.2780718057 0.1514242771 0.0416515570 -0.0818957694 -0.2728768874 -0.2176509972 -0.1238043582 0.0210746387 -0.0358127265
[101] 0.0652196628 -0.2445957894 1.1559112384 -0.1212197529 -0.2399652253 -0.0552203621 -0.2480896229 -0.2071183165 -0.1620011087 0.1032039077
[111] -0.2251634374 0.1220545675 0.4787629667 -0.3195907519 -0.2424032437 -0.0689875571 0.0361775256 0.3364841999 0.1191818765 0.1892102832
[121] 0.0914348484
```

```
> #Student turu artik
> x2 <- which((inf2$stud.res< (-3)) | inf2$stud.res>3)
> x2
[1] 68 72 73 74 77
```

Çıktıda verilen değerlerin [-3,3] aralığının içerisinde olmadığını student türü standartlaştırmaya göre aykırı değer olduklarını söyleyebiliriz.

Cook Uzaklılığı (D_i)

$$n > 50 \quad \text{Cook değeri} = \frac{4}{121} = 0.0330$$

```
$cooks
[1] 5.237181e-10 9.882726e-04 5.092715e-06 1.677995e-05 1.666605e-04 2.014110e-04 4.214471e-05 5.235949e-04 4.920113e-06 4.472393e-04 7.218182e-04
[12] 8.945843e-05 1.198762e-06 6.765147e-05 3.021824e-06 1.589415e-04 5.594513e-04 3.119531e-04 1.309098e-04 6.903122e-05 1.863883e-04 9.800362e-04
[23] 4.142812e-05 7.696044e-04 9.475803e-05 5.001030e-04 3.344700e-03 1.338927e-03 6.918603e-04 1.582440e-02 1.843983e-04 2.568495e-06 4.306619e-03
[34] 2.296573e-03 5.970828e-03 5.007535e-03 2.584951e-03 1.312476e-03 1.934890e-03 1.884360e-03 1.948170e-03 6.475874e-03 6.784496e-03 3.299665e-03
[45] 1.858818e-02 7.922245e-03 2.858777e-03 7.093815e-03 1.825556e-03 1.884293e-03 6.252368e-03 2.859994e-02 2.060751e-03 2.302832e-03 1.294768e-03
[56] 1.3484285e-02 1.640404e-03 3.454424e-03 1.285235e-03 1.551924e-02 1.815571e-03 7.715437e-03 4.612938e-03 5.229958e-03 3.873963e-02 2.803433e-03
[67] 1.348950e-02 4.664798e-02 1.044875e-01 9.374344e-02 6.968507e-02 6.637746e-02 7.969599e-02 7.232006e-02 8.319065e-02 7.056101e-02 4.114222e-02
[78] 4.464422e-04 9.678258e-07 4.296553e-06 2.484362e-03 4.457691e-04 7.139960e-05 9.377586e-05 3.630926e-04 5.662900e-05 1.221353e-04 9.504455e-05
[89] 1.031168e-02 2.953368e-04 3.683544e-04 8.580194e-04 2.073130e-04 2.254033e-05 4.549570e-05 4.352157e-04 2.035103e-04 1.153257e-04 5.982024e-06
[100] 1.079796e-05 3.661630e-05 2.826069e-04 1.813681e-02 6.743993e-05 2.620129e-04 2.199882e-05 3.388581e-04 5.097367e-04 1.808980e-04 9.811047e-05
[111] 4.682723e-04 1.844232e-04 3.648884e-03 5.816164e-04 2.797717e-04 3.556700e-05 1.278521e-05 1.414307e-03 2.403311e-04 3.520855e-04 4.944799e-05
```

```
> #Cook uzakligi
> x4 <- which(inf2$cooks>0.0330)
> x4
[1] 56 65 68 69 70 71 72 73 74 75 76 77
```

Çıktıda verilen gözlemlerin etkin bir aykırı değer olduğunu söyleyebiliriz.

İlk artık temizlenmesi sonucu elde edilen verinin model denklemini tekrar vermek gerekirse,

```
> yeni_sonuc <- lm(fiyat~satis_miktari+stok_yenileme+oy_miktari+kategori,data=artik_temizlenmis_veri)
> summary(yeni_sonuc)

Call:
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +
    kategori, data = artik_temizlenmis_veri)

Residuals:
    Min      1Q      Median      3Q      Max 
-1.69460 -0.08884  0.01319  0.30562  0.84455 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.60496   0.53375   1.133   0.259    
satis_miktari 8.40446   0.06384 131.644 <2e-16 ***  
stok_yenileme 2.39226   0.07622  31.387 <2e-16 ***  
oy_miktari   -3.76933   0.05769 -65.333 <2e-16 ***  
kategori2    -11.40071   0.12190 -93.523 <2e-16 ***  
kategori3    -12.97104   0.12502 -103.748 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5256 on 115 degrees of freedom
Multiple R-squared:  0.9973,    Adjusted R-squared:  0.9972 
F-statistic: 8526 on 5 and 115 DF,  p-value: < 2.2e-16
```

$$\hat{y} = 8.40 + 2.39x_1 + 2.83x_2 - 3.76x_3 - 11.40T_2 - 12.97T_3 \pm 0.5256$$

$$(0.533) \quad (0.063) \quad (0.076) \quad (0.057) \quad (0.121) \quad (0.125)$$

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$ (Model anlamlı değildir.)

$H_1 : \text{En az bir } \beta_j \text{ farklıdır. (Model anlamlıdır.)} \quad j=1,2,3,4,5$

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Modelin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

R^2 (Çoklu belirtme katsayısı) = 0.9973. Tüm değişkenler birlikte fiyat üzerindeki değişimin %99.73'ünü açıklamaktadır. Düzeltilmiş belirtme katsayısı ise %99.72'dir.

→ T_1 değişkeni modelde bulunmadığından “giyim” kategorisinin kılavuz değişken olduğunu söyleyebiliriz.

$H_0 : \beta_1 = 0$ (Satış miktarının model üzerindeki etkisi anlamlı değildir.)

$H_s : \beta_1 \neq 0$ (Satış miktarının model üzerindeki etkisi anlamlıdır.)

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Satış miktarının model üzerindeki etkisinin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

$H_0 : \beta_2 = 0$ (Stok yenileme değişkeninin model üzerindeki etkisi anlamlı değildir.)

$H_s : \beta_2 \neq 0$ (Stok yenileme değişkeninin model üzerindeki etkisi anlamlıdır.)

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Stok yenilemenin model üzerindeki etkisinin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

$H_0 : \beta_3 = 0$ (Oy miktarının değişkeninin model üzerindeki etkisi anlamlı değildir.)

$H_s : \beta_3 \neq 0$ (Oy miktarının değişkeninin model üzerindeki etkisi anlamlıdır.)

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Oy miktarının model üzerindeki etkisinin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

$H_0 : \beta_4 = 0$ ($T_1=T_2$)

$H_s : \beta_4 \neq 0$ ($T_1 \neq T_2$)

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Giyim ve kozmetik kategorisinin ortalama fiyat açısından farklı olduğunu %95 güven düzeyinde söyleyebiliriz.

$H_0 : \beta_5 = 0$ ($T_1=T_3$)

$H_s : \beta_5 \neq 0$ ($T_1 \neq T_3$)

$p < \alpha = 0.05$ olduğundan H_0 hipotezi reddedilir. Giyim ve aksesuar kategorisinin ortalama fiyat açısından farklı olduğunu %95 güven düzeyinde söyleyebiliriz.

→ R^2 (Çoklu belirtme katsayısı) = 0.9973. Tüm değişkenler birlikte fiyat üzerindeki değişimin %99.73'ünü açıklamaktadır. Düzeltilmiş belirtme katsayısı ise %99.72'dir.

```
> confint(yeni_sonuc, level=0.99)
          0.5 %    99.5 %
(Intercept) -0.7930696  2.002985
satis_miktari  8.2372429  8.571682
stok_yenileme  2.1926245  2.591892
oy_miktari     -3.9204424 -3.618209
kategori2      -11.7200085 -11.081418
kategori3      -13.2985131 -12.643568
```

Diğer değişkenler “0” değerini aldıklarında (mesela satış miktarı 0 olduğunda) fiyatın -0.79 ile 2 dolar arasında olacağını %99 güven düzeyinde söylenebilir. Ancak sabit terim denklem gereği bir katsayı olduğundan bu yorum anlamsızdır, fiyat negatif olamaz.

Diğer değişkenler sabit tutulup satış miktarı bir birim arttığında fiyatındaki artışın 8.23 ile 8.57 dolar arasında olacağını %99 güven düzeyinde söyleyebiliriz.

Diğer değişkenler sabit tutulup stok yenileme süresi bir gün arttırıldığında fiyatındaki artışın 2.19 ile 2.59 dolar arasında olacağını %99 güven düzeyinde olabileceğini söyleyebiliriz.

Diğer değişkenler sabit tutulup oy miktarı bir birim arttırıldığında fiyatındaki artışın -3.92 ile -3.61 dolar arasında olacağını %99 güven düzeyinde olabileceğini söyleyebiliriz.

Giyim kategorisi ve kozmetik kategorisi arasındaki ortalama fiyat farkının -11.72 dolar ile -11.08 dolar arasında olacağını %99 güven düzeyinde olabileceğini söyleyebiliriz.

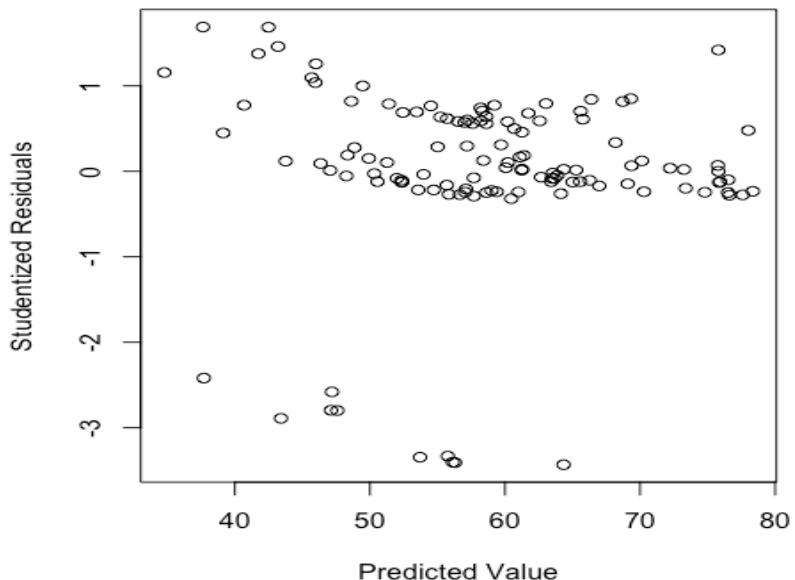
Giyim kategorisi ve aksesuar kategorisi arasındaki ortalama fiyat farkının -13.29 dolar ile -12.64 dolar arasında olacağını %99 güven düzeyinde olabileceğini söyleyebiliriz.

Güven aralıklarına bakıldığı zaman sabit değer ve stok yenileme değişkeni dışında diğer değişkenlerin güven aralığı 0 değerini içermediğinden anlamlı olmadıklarını güven aralıklarına bakarak da söyleyebiliriz.

Değişen Varyanslılık İncelemesi

Değişen varyanslılık varsayıminın bozulup bozulmadığını görebilmek için grafik ve test yöntemlerini uygulayalım.

```
#Degisen varyanslik incelemesi
par(mfrow=c(1,1))
plot(predict(yeni_sonuc),inf2$stud.res,xlab="Predicted Value",ylab="Studentized Residuals")
```



Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olmadığı, gözlemlerin nokta olarak ifade edildiği durumda doğrusal olmayan bir dağılım gösterdiği görülmektedir. Ancak görsel analiz tek başına yeterli değildir. Bu nedenle test edelim.

Brue-Pagan Testi

```
> bptest(yeni_sonuc)

studentized Breusch-Pagan test

data: yeni_sonuc
BP = 31.879, df = 5, p-value = 6.279e-06
```

H_0 = Varyanslar homojendir.

H_s = En az bir varyans farklıdır.

Burada BP testi sonucunda elde edilen değer p değerinden oldukça küçük olduğundan değişen varyanslılık sorunu olduğunu söyleyebiliriz.

Öz İlişki İncelemesi

```
> dwtest(yeni_sonuc)

  Durbin-Watson test

data:  yeni_sonuc
DW = 0.47219, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

H_0 : Artıklar arasında öz ilişki yoktur.

H_1 : Artıklar arasında pozitif öz ilişki vardır.

p değeri 0.05'ten oldukça küçük olduğundan artıklar arasında pozitif yönlü bir öz ilişkinin olduğunu %95 güven düzeyinde söyleyebiliriz.

Çoklu Bağlantının İncelenmesi

```
> vif(yeni_sonuc)
    satis_miktari  stok_yenileme   oy_miktari      kategori2      kategori3
        1.4057          1.7369       1.3464       1.5202       1.5841
```

Hiçbir bir VIF değerinin 10'dan büyük olmamasından dolayı çoklu bağlantı olmadığı varsa da hiçbir değişkenin çoklu bağlantısından etkilenmediğini düşünebiliriz.

```
> ols_eigen_cindex(yeni_sonuc)
    Eigenvalue Condition Index  intercept satis_miktari stok_yenileme   oy_miktari      kategori2      kategori3
1 4.681438163      1.000000 3.590202e-04  3.116593e-04  1.847419e-03 1.774440e-03 0.0079471015 0.0070026166
2 1.002122945      2.161370 2.242603e-06  3.714519e-06  8.977952e-07 5.490217e-05 0.1950984604 0.2102951372
3 0.241224943      4.405331 5.484521e-04  6.303478e-04  1.534128e-02 1.055242e-02 0.7386940584 0.6335410875
4 0.039409110     10.899117 5.106189e-02  1.495959e-02  6.911335e-01 1.036533e-02 0.0556691080 0.0827131270
5 0.031957178     12.103346 1.920329e-02  3.385951e-02  6.455640e-02 9.044015e-01 0.0020008885 0.0659370599
6 0.003847661     34.881187 9.288251e-01  9.502352e-01  2.271205e-01 7.285143e-02 0.0005903833 0.0005109718
```

Ancak koşul sayılarına bakıldığından 30'dan büyük olan 1 tane durum olduğundan 1 tane güçlü çok bağlantı vardır diyebiliriz. Bir de özdeğer ve özvektörleri inceleyelim.

```
> signif(sm$values,3)
[1] 1.870 1.560 0.786 0.425 0.358
> signif(sm$vectors,3)
      [,1]   [,2]   [,3]   [,4]   [,5]
[1,] -0.348  0.4980  0.6100  0.189 -0.473
[2,] -0.540  0.4000 -0.0194 -0.343  0.656
[3,] -0.520  0.0782 -0.7140  0.327 -0.326
[4,] -0.373 -0.5500  0.3220  0.581  0.342
[5,]  0.421  0.5320 -0.1180  0.634  0.351
```

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

Uyum Kestirimi

```
> newdata = data.frame(satis_miktari=9.46,stok_yenileme=2.67,oy_miktari=3.34,kategori="3")
> predict(yeni_sonuc.lm,newdata ,interval = "confidence")
    fit      lwr      upr
1 60.93791 60.76517 61.11065
```

Veri setimizde bulunan $x_i' = (1 \ 9.46 \ 2.67 \ 3.34 \ "3")$ gözlemi için

$\hat{y}_i' = 60.93791$ çıkmaktadır. Kurduğumuz modele göre günlük satış miktarının 9.46 birim olduğu, stok yenileme süresinin yaklaşık 3 gün olduğu ve 3.34 oy miktarına sahip aksesuar kategorisindeki bir ürünün fiyatının ortalama 60.93\$ olması beklenir. Bu gözlem için y_i değerimiz ise 60.86959'dur.

Ön Kestirim

```
> #ON KESTIRIM
> newdata = data.frame(satis_miktari=6,stok_yenileme=2,oy_miktari=2.5,kategori="1")
> predict(yeni_sonuc.lm,newdata ,interval = "prediction")
    fit      lwr      upr
1 46.39293 45.27946 47.50641
```

Veri setimizde bulunmayan. $\tilde{x}_i' = (1 \ 6 \ 2 \ 2.5 \ "1")$ gözlemi için

$\tilde{y}_i = 46.39293$ çıkmaktadır. Günlük satış miktarının 6 birim olduğu, stok yenileme süresinin 2 gün olduğu ve 2.5 oy miktarına sahip giyim kategorisindeki bir ürünün fiyatının ortalama 46.39\$ olması beklenir.

```
> #GUVEN ARALIGI UYUM KESTIRIMI
> newdata = data.frame(satis_miktari=9.46,stok_yenileme=2.67,oy_miktari=3.34,kategori="3")
> predict(yeni_sonuc.lm,newdata ,interval = "confidence",level=0.95)
    fit      lwr      upr
1 60.93791 60.76517 61.11065
>
> #GUVEN ARALIGI ON KESTIRIMI
> newdata = data.frame(satis_miktari=6,stok_yenileme=2,oy_miktari=2.5,kategori="1")
> predict(yeni_sonuc.lm,newdata ,interval = "prediction",level=0.95)
    fit      lwr      upr
1 46.39293 45.27946 47.50641
```

Uyum kestirimi için güven aralığı

$$P(60.76517 < E(Y / X=x_i) < 61.11065) = 0.95$$

Günlük satış miktarının 9.46 birim olduğu, stok yenileme süresinin yaklaşık 3 gün olduğu ve 3.34 oy miktarına sahip aksesuar kategorisindeki bir ürünün fiyatının 60.76\$ ile 61.11\$ arasında olacağını %95 güven düzeyinde söyleyebiliriz.

Ön kestirim için güven aralığı

$$P(45.27946 < E(Y / X=\tilde{x}_i) < 47.50641) = 0.95$$

Günlük satış miktarının 6 birim olduğu, stok yenileme süresinin 2 gün olduğu ve 2.5 oy miktarına sahip giyim kategorisindeki bir ürünün fiyatının 45.27\$ ile 4750\$ arasında olacağını %95 güven düzeyinde söyleyebiliriz.

Değişken Seçimi

İleriye Doğru Seçim

```
Start: AIC=556.29
artik_temizlenmis_veri$fiyat ~ 1

              Df Sum of Sq    RSS    AIC
+ artik_temizlenmis_veri$satis_miktari  1    7251.2  4558.4 443.10
+ artik_temizlenmis_veri$kategori      2    3030.3  8779.3 524.41
+ artik_temizlenmis_veri$stok_yenileme  1    2605.5  9204.1 528.13
<none>                                         11809.6 556.29
+ artik_temizlenmis_veri$oy_miktari     1    146.0   11663.6 556.78

Step: AIC=443.1
artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari

              Df Sum of Sq    RSS    AIC
+ artik_temizlenmis_veri$kategori      2    3344.9  1213.4 286.96
+ artik_temizlenmis_veri$oy_miktari     1    580.0   3978.4 428.63
<none>                                         4558.4 443.10
+ artik_temizlenmis_veri$stok_yenileme  1     72.1   4486.3 443.17

Step: AIC=286.96
artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari +
  artik_temizlenmis_veri$kategori

              Df Sum of Sq    RSS    AIC
+ artik_temizlenmis_veri$oy_miktari     1    909.47  303.96 121.45
<none>                                         1213.43 286.95
+ artik_temizlenmis_veri$stok_yenileme  1     2.35   1211.08 288.72

Step: AIC=121.46
artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari +
  artik_temizlenmis_veri$kategori + artik_temizlenmis_veri$oy_miktari

              Df Sum of Sq    RSS    AIC
+ artik_temizlenmis_veri$stok_yenileme  1    272.19  31.774 -149.80
<none>                                         303.963 121.45

Step: AIC=-149.8
artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari +
  artik_temizlenmis_veri$kategori + artik_temizlenmis_veri$oy_miktari +
  artik_temizlenmis_veri$stok_yenileme
```

```
> forward

Call:
lm(formula = artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari +
  artik_temizlenmis_veri$kategori + artik_temizlenmis_veri$oy_miktari +
  artik_temizlenmis_veri$stok_yenileme)

Coefficients:
              (Intercept)  artik_temizlenmis_veri$satis_miktari
                           0.605                      8.404
  artik_temizlenmis_veri$kategori2      artik_temizlenmis_veri$kategori3
                           -11.401                     -12.971
  artik_temizlenmis_veri$oy_miktari  artik_temizlenmis_veri$stok_yenileme
                           -3.769                      2.392
```

Bu yönteme göre öncelikle x1 (satış miktarı) değişkeni modele alınmıştır. Ardından en küçük AIC değerine sahip x4 (kategori) değişkeni modele alınmıştır. Ardıncan x3(oy miktarı) değişkeni ve en sonunda da x2(stok yenileme) değişkeni modele alınıyor. Bu şekilde bütün değişkenlerin anlamlı olduğundan denkleme eklenildiğini, en iyi modelin tam kümeye modeli olduğunu söyleyebiliriz.

```
> summary(forward)

Call:
lm(formula = artik_temizlenmis_veri$fiyat ~ artik_temizlenmis_veri$satis_miktari +
   artik_temizlenmis_veri$kategori + artik_temizlenmis_veri$oy_miktari +
   artik_temizlenmis_veri$stok_yenileme)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.69460 -0.08884  0.01319  0.30562  0.84455 

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      0.60496   0.53375   1.133   0.259    
artik_temizlenmis_veri$satis_miktari 8.40446   0.06384 131.644 <2e-16 ***
artik_temizlenmis_veri$kategori2    -11.40071  0.12190 -93.523 <2e-16 ***
artik_temizlenmis_veri$kategori3    -12.97104  0.12502 -103.748 <2e-16 ***
artik_temizlenmis_veri$oy_miktari   -3.76933  0.05769 -65.333 <2e-16 ***
artik_temizlenmis_veri$stok_yenileme 2.39226  0.07622  31.387 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5256 on 115 degrees of freedom
Multiple R-squared:  0.9973, Adjusted R-squared:  0.9972 
F-statistic:  8526 on 5 and 115 DF,  p-value: < 2.2e-16
```

$$\text{En iyi model : } y_i = 0.604 + 8.40x_1 + 2.39x_2 - 3.76x_3 - 11.40T_2 - 12.97T_3 \pm 0.5256$$

$$(0.533) \quad (0.063) \quad (0.076) \quad (0.057) \quad (0.121) \quad (0.125)$$

Modelin ve değişkenlerin anlamlı olduğunu %95 güven düzeyinde söyleyebiliriz.

Geriye Doğru Seçim

```
> #GERIYE DOGRU SECIM
> backward<-step(yeni_sonuc,direction="backward")
Start:  AIC=-149.8
fiyat ~ satis_miktari + stok_yenileme + oy_miktari + kategori

          Df Sum of Sq    RSS    AIC
<none>                 31.8 -149.80
- stok_yenileme  1     272.2  304.0  121.46
- oy_miktari     1    1179.3 1211.1  288.72
- kategori       2    3462.4 3494.2  414.93
- satis_miktari  1    4788.2 4820.0  455.85
```

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlıyor. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. Son modelde hiçbir değişken denklemden çıkartılmamıştır.

```
> summary(backward)

Call:
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +
    kategori, data = artik_temizlenmis_veri)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.69460 -0.08884  0.01319  0.30562  0.84455 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.60496   0.53375   1.133   0.259    
satis_miktari 8.40446   0.06384 131.644 <2e-16 ***  
stok_yenileme 2.39226   0.07622  31.387 <2e-16 ***  
oy_miktari    -3.76933   0.05769 -65.333 <2e-16 ***  
kategori2     -11.40071   0.12190 -93.523 <2e-16 ***  
kategori3     -12.97104   0.12502 -103.748 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5256 on 115 degrees of freedom
Multiple R-squared:  0.9973,    Adjusted R-squared:  0.9972 
F-statistic: 8526 on 5 and 115 DF,  p-value: < 2.2e-16
```

Bu yöntemde hiçbir değişkeninin denklemden çıkmadığı tam küme denklemi en iyi model olarak elde edilmiştir. p değeri 0.05'ten küçük olduğu için modelin anlamlı olduğunu söyleyebiliriz. Bütün değişkenlerin anlamlı olduğunu da söyleyebiliriz

Adımsal Regresyon Yöntemi

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

Call:
lm(formula = fiyat ~ satis_miktari + stok_yenileme + oy_miktari +
    kategori, data = artik_temizlenmis_veri)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.69460 -0.08884  0.01319  0.30562  0.84455 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.60496   0.53375   1.133   0.259    
satis_miktari 8.40446   0.06384 131.644 <2e-16 ***
stok_yenileme 2.39226   0.07622  31.387 <2e-16 ***
oy_miktari    -3.76933   0.05769 -65.333 <2e-16 ***
kategori2     -11.40071   0.12190 -93.523 <2e-16 ***
kategori3     -12.97104   0.12502 -103.748 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5256 on 115 degrees of freedom
Multiple R-squared:  0.9973, Adjusted R-squared:  0.9972 
F-statistic: 8526 on 5 and 115 DF,  p-value: < 2.2e-16
```

Bu yöntemden elde edilen çıktıya göre de tüm değişkenlerin bulunduğu denklemin en iyi model olduğunu söyleyebiliriz. İleri ve geriye doğru seçim yöntemleri ile aynı sonuçları vermiştir.

Ridge Regresyon

*Ben denemek amaçlı kodları veri setime uyguladım ancak oluşturduğum modelde çoklu bağlantı yok bu nedenle ridge izi grafiğine bakılmasına gerek yok.

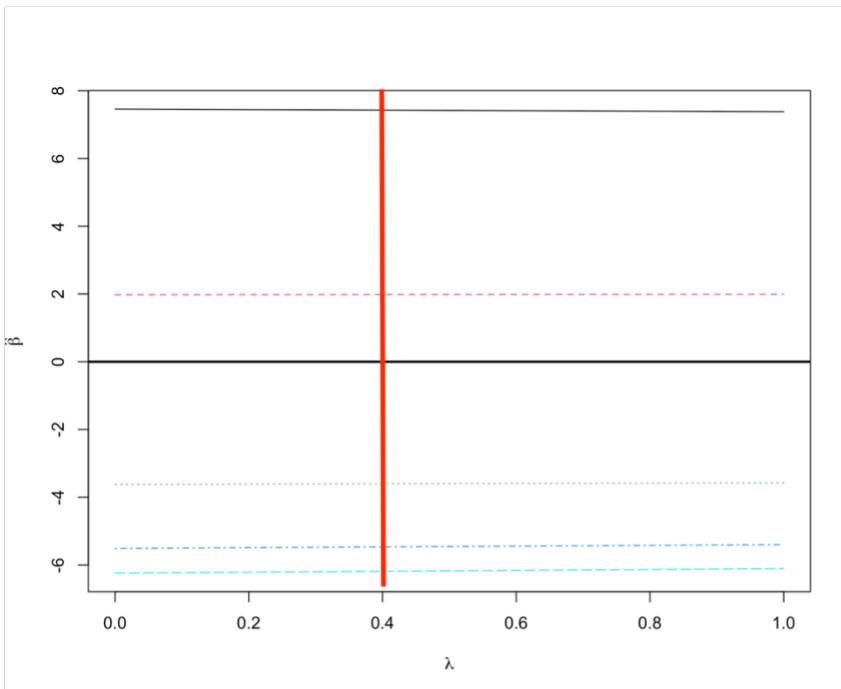
```
> library(MASS)
> ridge <- lm.ridge(artik_temizlenmis_veri$fiyat~artik_temizlenmis_veri$satis_miktari
+                     +artik_temizlenmis_veri$stok_yenileme+artik_temizlenmis_veri$oy_miktari
+                     +artik_temizlenmis_veri$kategori, lambda = seq(0,1,0.05))
> matplot(ridge$lambda, t(ridge$coef), type="l", xlab=expression(lambda)
+           , ylab=expression(hat(beta)))
> abline(h=0, lwd=2)
> ridge$coef[, ridge$lambda==0.4]
```

	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
artik_temizlenmis_veri\$satis_miktari	7.458256	7.454216	7.450181	7.446152	7.442129	7.438113	7.434102	7.430098	7.426099	7.422107	7.418120
artik_temizlenmis_veri\$stok_yenileme	1.976627	1.977326	1.978021	1.978712	1.979400	1.980084	1.980764	1.981441	1.982114	1.982783	1.983449
artik_temizlenmis_veri\$oy_miktari	-3.622479	-3.620059	-3.617642	-3.615227	-3.612815	-3.610405	-3.607998	-3.605594	-3.603191	-3.600792	-3.598394
artik_temizlenmis_veri\$kategori2	-5.510103	-5.504436	-5.498781	-5.493138	-5.487508	-5.481889	-5.476282	-5.470687	-5.465105	-5.459534	-5.453974
artik_temizlenmis_veri\$kategori3	-6.239663	-6.232861	-6.226073	-6.219299	-6.212539	-6.205792	-6.199060	-6.192342	-6.185637	-6.178946	-6.172269
	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
artik_temizlenmis_veri\$satis_miktari	7.414140	7.410165	7.406197	7.402234	7.398277	7.394326	7.390381	7.386442	7.382508	7.378581	
artik_temizlenmis_veri\$stok_yenileme	1.984112	1.984770	1.985426	1.986077	1.986726	1.987370	1.988012	1.988649	1.989284	1.989915	
artik_temizlenmis_veri\$oy_miktari	-3.595999	-3.593607	-3.591217	-3.588830	-3.586445	-3.584063	-3.581683	-3.579305	-3.576930	-3.574557	
artik_temizlenmis_veri\$kategori2	-5.448427	-5.442891	-5.437367	-5.431855	-5.426354	-5.420865	-5.415387	-5.409921	-5.404467	-5.399024	
artik_temizlenmis_veri\$kategori3	-6.165605	-6.158956	-6.152319	-6.145697	-6.139087	-6.132492	-6.125910	-6.119341	-6.112785	-6.106243	

```
> select(ridge)
modified HKB estimator is 0.005839214
modified L-W estimator is 0.008515455
smallest value of GCV at 0
```

```
> ridge$coef[,ridge$lam==0.4]
artik_temizlenmis_veri$satis_miktari artik_temizlenmis_veri$stok_yenileme
                                7.426099                               1.982114
artik_temizlenmis_veri$oy_miktari      artik_temizlenmis_veri$kategori2
                                -3.603191                               -5.465105
artik_temizlenmis_veri$kategori3
                                -6.185637
```

*Burada 0.4 dememizin özel bir sebebi yok sadece değerlerin stabil kaldığı bir değer olmalı.



Ridge izi grafiğine bakacak olursak eğer satış miktarı(x_1), stok yenileme(x_2), oy miktarı(x_3), T_2 ve T_3 değişkenlerinin -sıfır ekseni paralel şekilde ilerlediklerinden- kararlı olduklarını yani çoklu bağlantı varsa bile bundan etkilenmediklerini söyleyebiliriz.

KAYNAKÇA

Ders Notları

<https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>

<http://www.sthda.com/english/articles/40-regression-analysis/166-predict-in-r-model-predictions-and-confidence-intervals/>

[http://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/#:~:text=Multiple%20linear%20regression%20is%20an,distinct%20predictor%20variables%20\(x\).&text=The%20%E2%80%9Cb%E2%80%9D%20values%20are%20called,weights%20\(or%20beta%20coefficients\).](http://www.sthda.com/english/articles/40-regression-analysis/168-multiple-linear-regression-in-r/#:~:text=Multiple%20linear%20regression%20is%20an,distinct%20predictor%20variables%20(x).&text=The%20%E2%80%9Cb%E2%80%9D%20values%20are%20called,weights%20(or%20beta%20coefficients).)