

# İÇİNDEKİLER

<b>VERİ KÜMESİ</b>	<b>2</b>
<b>A) VARSAYIMLARIN İNCELENMESİ</b>	<b>5</b>
<b>B) ARTIK DEĞER İNCELEMESİ</b>	<b>9</b>
<b>C) DEĞİŞEN VARYANSLILIK:</b>	<b>14</b>
□ <b>BRUGE-PAGAN TESTİ</b>	<b>14</b>
□ <b>WHITE TESTİ</b>	<b>14</b>
<b>D) ÖZ İLİŞKİ SORUNU :</b>	<b>15</b>
<b>Durbin-Watson Testi</b>	<b>15</b>
<b>E) ÇOKLU BAĞLANTI SORUNU</b>	<b>15</b>
Özdeğer ve Özvektörlerin Bulunması	15
<b>F) MODEL ANLAMLILIĞI :</b>	<b>16</b>
<b>G) KATSAYI YORUMLARI</b>	<b>18</b>
<b>H) UYUM KESTİRİMİ</b>	<b>19</b>
<b>İ) ÖN KESTİRİM</b>	<b>19</b>
<b>J) REGRESYON KATSAYILARI İÇİN GÜVEN ARALIKLARI VE YORUMU</b>	<b>19</b>
<b>K) ÖN KESTİRİM VE UYUM KESTİRİMİ İÇİN GÜVEN ARALIKLARI</b>	<b>19</b>
<b>L) EN İYİ MODEL BULUNMASI</b>	<b>21</b>
<b>L.1) İleriye Doğru Seçim Yöntemi</b>	<b>21</b>
<b>L.2) Geriye Doğru Seçim Yöntemi</b>	<b>22</b>
<b>L.3) Adımsal Seçim Yöntemi</b>	<b>23</b>
<b>M) RİDGE REGRESYON</b>	<b>25</b>
<b>RAPOR</b>	<b>26</b>
<b>KAYNAKÇA</b>	<b>26</b>

# REGRESYON ANALİZİ

## (Rasgele Veri Kümesi Örneği)

Ne kadar para kazandığınız kadar o parayı nasıl harcadığınız, o parayla nasıl geçindiğiniz ve o parayla nasıl birikim, yatırım yaptığınız da önemli. Tam bu noktada finansal okuryazarlık devreye giriyor. Bu konu üzerine yapılan bir araştırmada finansal okur-yazarlık dersi alan 100 öğrenci rasgele seçilip günlük birikim miktarını (y) etkileyen faktörlerin aileden alınan para miktarı (x1), bir işte çalışarak kazanılan para miktarı (x2), günlük harcama miktarı (x3) ve eğitim düzeyi (x4) gibi değişkenler ile ilişkili olup olmadığı araştırılmak isteniyor.

Y: Günlük birikim miktarı (TL)

X1: Aileden alınan günlük para miktarı (TL)

X2: Bir işte çalışarak kazanılan günlük para miktarı (TL)

X3: Günlük harcama miktarı (TL)

X4: Eğitim düzeyi  $\begin{cases} 1 = \text{ilköğretim} \\ 2 = \text{lise} \\ 3 = \text{üniversite} \end{cases}$  ve N=100

## VERİ KÜMESİ

	y	x1	x2	x3	x4
1	12,653813133594700	8,18428841477259	2,837331190168170	7,74713777417576	1
2	9,860315287992020	8,63319543656707	2,970008750523710	8,48844916698387	1
3	9,099250980732910	8,25952759784279	3,863641480890320	9,15899105240744	1
4	5,933376833186100	7,43633171575873	2,431605053334620	7,89487596527556	1
5	11,282007024824400	8,43768380323681	3,960969646892280	9,13178800700680	1
6	-1,747192482055190	7,60658672542246	3,640557198234180	10,54853067411100	1
7	5,071862969293850	6,69098985731639	1,373992729063970	7,50601350991081	1
8	,245806404201752	6,30265815287014	2,017663881544150	8,79347923556040	1
9	12,145819442444900	9,18626824907389	4,978813465806210	8,99345960039765	1
10	3,441057714663650	7,74994767985487	2,520989706631390	9,42830816616939	1
11	10,072241816402100	8,21870598393488	2,332901034186140	8,51648154892722	1
12	9,238652034732510	7,09316931127151	2,660276953927840	7,53888731940226	1
13	14,273599528894400	8,10265207622816	2,237187188210160	6,95584891953417	1
14	,564570518152261	6,34661047370194	1,867756142215600	8,56625026656925	1
15	8,838720523800210	7,97072355122767	2,871038454889100	8,56365355071689	1
16	12,407793002078100	9,13434289107812	3,093163840717830	8,75274966945486	1
17	8,003793805037140	7,87351863422770	3,387817026902610	9,42373630689416	1
18	9,581149349257620	8,96023905539236	3,455618940956390	9,11964973655651	1
19	2,452133962163980	6,90440807859401	2,777549560170050	9,04273504946870	1
20	-,898569076917819	8,54810843526788	4,495680776764110	11,58549335509410	1
21	12,472767040829800	8,42692855821283	3,748605865153910	8,67113663121600	1
22	51,688630193122700	6,75316792895998	2,199126299203060	8,25567251908212	1
23	10,931957677263800	8,74487507694592	4,331717923842660	9,18779292141520	1

24	4,538775091204520	6,36408549809724	1,358092539764210	7,10515984225476	1
25	9,257177757794850	8,52921292733311	3,074658532876990	8,78171492611018	1
26	2,468342211641750	8,99495134072085	3,512889004691830	9,06132356534645	2
27	3,266385446785090	7,32004350517938	2,169597483975140	7,26475721606499	2
28	2,108109564826450	8,76433781272841	3,324402720700750	8,78150682875810	2
29	-,750837977834689	9,60032794743248	5,020627408585380	10,80325871449820	2
30	-3,001028728996300	6,91127596460406	1,692292048793960	7,78353809608042	2
31	4,703662141845000	8,34722477614756	2,983115305462870	8,04102318583023	2
32	-1,655359019245940	7,67523922355196	2,259557461892110	8,96971607931314	2
33	-6,237772096174310	7,33581032063657	2,237962561763740	9,51087523370848	2
34	-3,677586131981690	7,16929349656451	2,955923489697940	9,40352138590981	2
35	-5,725261227446390	6,97508005619956	2,991170110140400	8,98051621748129	2
36	1,353147281602590	9,46241608955735	3,137894939604640	10,44922786406180	2
37	-3,742710264780900	7,69692548034644	2,620330042636580	8,66873413624575	2
38	,176085744841702	7,94876060201556	2,777025454331330	8,86118785141585	2
39	-3,413586380792840	7,74313197177561	3,140420689521230	9,61476328426897	2
40	,170177451630661	7,91512564537927	2,430222308644980	8,16320577777974	2
41	-6,952130015427540	8,09743331437687	2,902877706708040	10,81082627986140	2
42	-3,191342355855770	7,13138776074953	1,601431485522090	8,44185531169482	2
43	2,809791043942910	7,71399904214284	2,897210297658690	7,72976368799612	2
44	-,597565523533555	9,21373697457949	4,606135857846680	10,49439179622210	2
45	-2,694244085493160	8,97721570052633	5,077252544859140	11,36852599208900	2
46	11,002600545473900	8,85503285041240	3,030645511340680	7,38230104643818	2

47	-2,879316399072730	7,74482832873161	2,345397895388240	9,06730677147853	2
48	,548050727619677	9,62620543897487	3,649294246532870	10,15965238952460	2
49	-,125759873291533	7,43620222792254	2,446664086431240	8,18673211068352	2
50	5,360511154097100	10,23863003667250	4,450034605156400	9,57543353916646	2
51	51,655374725181000	6,75934092361557	2,130202122267030	7,06134441509922	2
52	-1,713745041468920	6,66741058181785	,793351407365085	7,41497379722334	2
53	-6,745013515800320	8,22820323498772	3,715366497195480	10,34934986245130	2
54	49,562520036312800	8,42012053216054	4,271382380663500	9,37052684296332	2
55	1,114935379882280	6,67322381315363	1,251301939497690	6,72369314269823	2
56	-3,167013911095490	7,63135518824530	3,371415844905250	9,60086888234360	2
57	-4,938333889873730	7,76045408135650	3,601269422854450	10,18835267614050	2
58	-3,253235971904090	8,41482951433512	3,200216825186710	9,66461870215483	2
59	3,270241688870800	9,72909265849031	4,236743605360310	9,75359358418531	2
60	-,786161208247254	7,66008306554824	2,373490159428670	8,45781322559459	2
61	3,722790294984590	9,13059137619422	3,032580165612050	8,67328035532784	3
62	2,256911736141690	8,68967603648708	2,769904187945220	8,60106818493376	3
63	-1,674242653846500	7,92317532215002	2,438851487024240	8,77250439383170	3
64	-,177912593837074	9,78606773667207	4,577627948396670	11,00597876467400	3
65	4,157428757337490	9,64508506010463	4,025150128470830	9,78587182691988	3
66	1,705415309140120	9,18144591954656	3,751540978817520	9,53875593156542	3
67	-4,235142782640340	6,30301188943691	2,257565928172110	8,56138107689818	3
68	-1,039475197441490	8,38171310844190	2,729207592268130	8,59394811925562	3
69	1,407428766189680	7,66541733854573	2,187842428619460	7,88906837066395	3
70	1,336460524083970	7,56201338303741	2,094389346169400	7,91661112369535	3
71	6,806868049822630	9,10217681063978	3,391815718207450	8,77675446617250	3
72	3,131844532852940	8,57466115377285	2,633952766699760	8,98919906605579	3
73	-2,300900985893420	8,48165431928777	3,095346072923710	9,87218432162006	3
74	,109575399700523	8,62738321508849	2,123886481596750	9,33566308860909	3
75	-1,596386265143460	8,47252470098623	3,383727459587510	9,67240581998214	3
76	-4,858923349134760	6,97390969071212	3,260438850870460	9,19713275038653	3
77	-11,555838656427500	7,22835119987011	3,634612146804090	11,21179410614160	3
78	-4,735582549726700	7,18627424877464	2,422458349341960	9,03918770619264	3
79	-3,119857269910300	8,61403994777739	4,059060069968390	10,37644603718920	3
80	-7,278343917495530	7,16894377240538	3,025553880830970	9,58827788183780	3
81	2,097016477263960	6,96185448836519	1,354562533294290	7,11210928022279	3
82	-3,711131618890740	7,95621924627032	3,012433927323940	8,73004639393704	3
83	-3,761014440908340	7,36817262550653	2,699065299055050	8,73212811364834	3
84	52,602517054769900	9,66481934865487	4,804138092796120	10,00607190488890	3
85	-5,658775573573070	8,48048839418853	3,949496265301200	10,35508791277930	3
86	-6,284584552996810	7,29351378465162	3,133731014574010	9,68813158072628	3
87	-4,636965469567090	6,75967642741896	3,135575300621000	9,12295693401791	3
88	-1,639420598893500	8,61358388717632	2,617577097178450	9,80755623675598	3
89	-4,363932570736120	8,23859716106758	4,123836427945180	10,49770789167080	3
90	-5,158293371068580	6,12160242224422	1,236818620359440	7,84246840651076	3
91	6,809713164011060	8,79098618403068	3,409190871621510	8,54321697652104	3
92	-7,197596369832910	7,25674558799692	4,038479443091830	9,81211131515446	3
93	5,047660033878870	8,70412155439853	3,266810007964170	8,54085747574689	3
94	50,709502523636800	6,65194479959626	1,955966218547550	6,97442491093682	3
95	-8,167028377090700	6,16570596590852	2,153856241203930	8,71190183778376	3
96	-1,694158185179690	7,74518923324535	2,322694847942110	8,27637641362981	3
97	-2,341728494184950	8,07444391370293	1,784295538853330	8,72173356997538	3
98	3,854441993205070	8,03561768027438	1,992735408256850	7,21302186196167	3
99	-4,569950600388980	7,49074032959382	2,044702509022930	8,75191626121348	3
100	-11,821313542052300	7,10219700213436	2,423082726893560	9,59135029971599	3

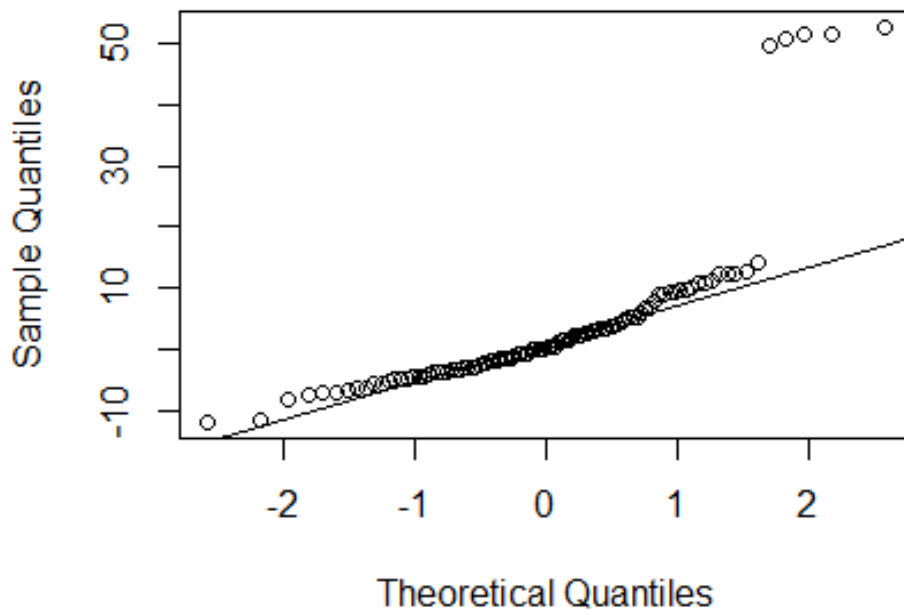
## A) Varsayımların incelenmesi

```
Console Terminal x
C:/Users/User/Desktop/regresyon ödevi/merve/ ↗

[workspace loaded from C:/Users/User/Desktop/regresyon ödevi/merve/.RData]

> veri <- read.csv("C:/Users/User/Desktop/veri.txt", comment.char="#")
> view(veri)
> veri <- read.csv2("C:/Users/User/Desktop/veri.txt", sep="")
> view(veri)
> names(veri)
[1] "y" "x1" "x2" "x3" "x4"
> attach(veri)
> x4<-as.factor(x4)
> qqnorm(y)
> qqline(y)
```

### Normal Q-Q Plot



Bağımlı değişken (y)'nin normal dağılım gösterip göstermediği incelendi. Grafik üzerinden incelemesi yapıldığında 45 derecelik normallik eğrisi üzerinde dağılım gösterse de tamamen normal dağıldığını söyleyemeyiz, eğri dışında noktalar bulunmakta ve  $n > 50$  olduğundan Kolmogorov-Smirnov testi ile normalliği test etmeliyiz.

```
> lillie.test(y)

      Lilliefors (Kolmogorov-Smirnov) normality test

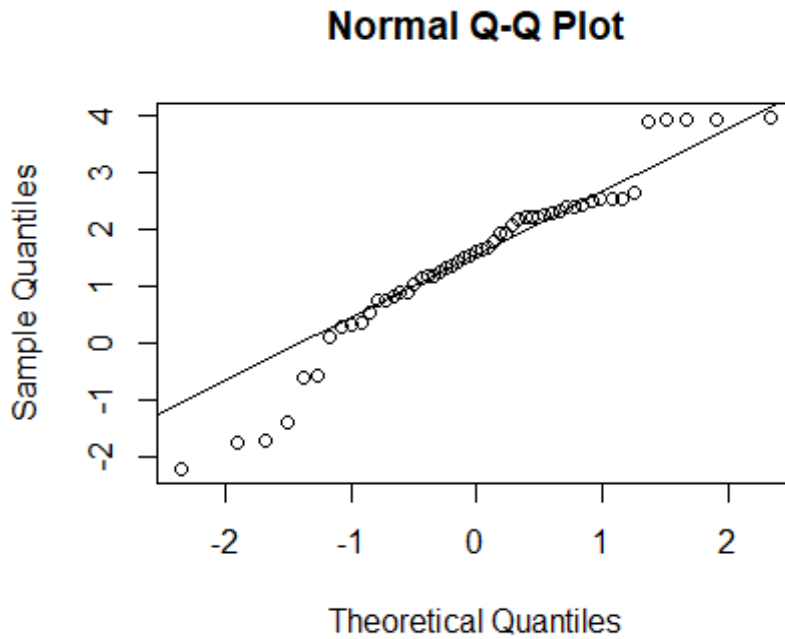
data:  y
D = 0.19663, p-value = 3.072e-10
```

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

**YORUM:**  $P\text{-value} = 3,072e-10 < \alpha = 0,05$  olduğundan  $H_0$  reddedilir. Verilerin dağılımı ile normal dağılım arasında fark vardır.

Verilerin normal dağılması için dönüşüm yapılır.

```
> lny<-log(y)
> qqnorm(lny)
> qqline(lny)
```



Grafik incelendiğinde y değerlerine göre daha iyi bir hal almış olsa da çizginin dışında olan değerler hala bulunmaktadır. Tekrar normalliği Kolmogorov-Smirnov testi ile test edersek;

```
> lillie.test(lny)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  lny
D = 0.1138, p-value = 0.08409
```

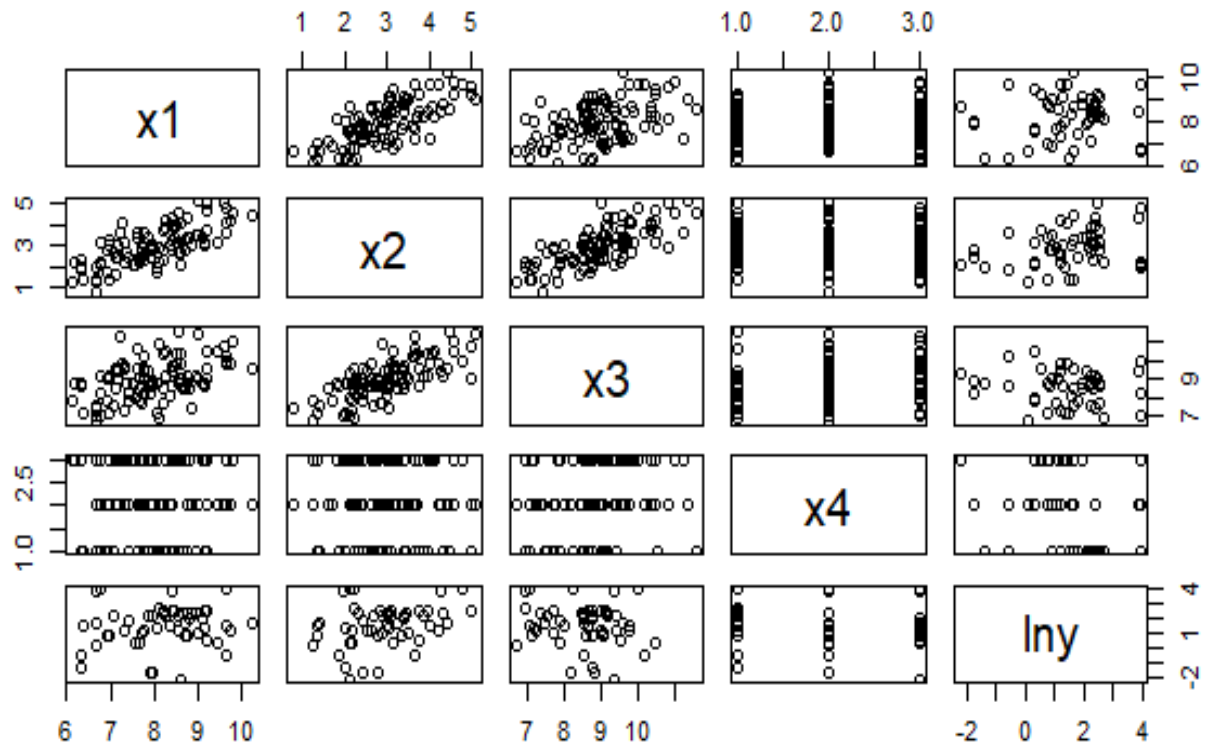
- $H_0$ : Verilerin dağılışı ile normal dağılım arasında fark yoktur.
- $H_S$ : Verilerin dağılışı ile normal dağılım arasında fark vardır.

YORUM:  $P\text{-value} = 0,08409 > \alpha = 0,05$  olduğundan  $H_0$  kabul edilir. Verilerin dağılımı ile normal dağılım arasında fark yoktur, lny normal dağılım göstermektedir.

- Bağımlı (y) ve bağımsız değişkenler arasındaki ilişkinin doğrusal olup olmadığını incelemek varsayımlarımızdan bir diğeridir. Bunu test edecek olursak;

```
> datayeni<-cbind(x1,x2,x3,x4,lny)
> pairs(datayeni)
```





**YORUM:** Saçılım grafiği incelendiğinde bağımlı değişken ile bağımsız değişkenler arasında ilişki vardır. Ancak bağımsız değişkenler de birbirleriyle ilişkili durumdadır. X2 ile X3 ilişkili durumdadır.

### MODEL:

- $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (Model anlamsızdır.)
- $H_S: \text{En az bir } \beta_j \neq 0$  (Model anlamlıdır.)

```

Console Terminal x
C:/Users/User/Desktop/yeniregresyon/
> sonuc<-lm(lny~x1+x2+x3+x4)
> summary(sonuc)

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

Residuals:
    Min       1Q   Median       3Q      Max
-2.30814 -0.43550 -0.09605  0.35526  2.62630

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.5184     1.9647   4.336 7.62e-05 ***
x1            -0.1132     0.2958  -0.383 0.70360
x2             1.4967     0.3240   4.620 3.00e-05 ***
x3            -1.1781     0.2476  -4.759 1.89e-05 ***
x4            -1.1050     0.3852  -2.868 0.00616 **
x43            -0.5124     0.4061  -1.262 0.21323
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.08 on 47 degrees of freedom
(47 observations deleted due to missingness)
Multiple R-squared:  0.4756,    Adjusted R-squared:  0.4198
F-statistic: 8.526 on 5 and 47 DF, p-value: 8.382e-06

```

Katsayıları yorumlayacak olursak;

Bağımlı değişken (y) günlük birikim miktarı (TL) olmak üzere; x1=aileden alınan günlük para miktarı, x2=çalışarak(iş vb.) kazanılan günlük para miktarı, x3= günlük harcama, x42=eğitim düzeyi lise, x43=eğitim düzeyi üniversite değişkenleri girilmiştir.

- X41= eğitim düzeyi ilköğretim kılavuz değişken olduğunu elde edilen modelde olmadığından görülebilir. P-value =  $8,382e-06 < \alpha = 0,05$  olduğundan H0 reddedilir. Modelimiz istatistiksel olarak anlamlıdır. Ayrıca bağımlı değişkenin %47,56'sının bağımsız değişkenlerce açıklandığı yorumlaması yapılabilir.

$\beta_0$ = Bağımsız değişkenlerimiz 0 olduğunda günlük birikim miktarının aldığı ortalama değer 8,52 TL olduğunu söyleyebiliriz.

$$H_0 : \beta_1 = 0$$

p-value=0,7 >  $\alpha = 0,05$  olduğundan H0 kabul.

(Aileden alınan günlük para miktarı model üzerinde anlamsızdır.)

$$H_0 : \beta_2 = 0$$

p-value=0,000 <  $\alpha = 0,05$  olduğundan H0 red.

(Çalışarak (iş vb.) kazanılan günlük para miktarı model üzerinde anlamlıdır.)

$$H_0 : \beta_3 = 0$$

p-value=0,000 <  $\alpha = 0,05$  olduğundan H0 red.

(Günlük harcama miktarı model üzerinde anlamlıdır.)

$$H_0 : \beta_4 = 0$$

p-value=0,006 <  $\alpha = 0,05$  olduğundan H0 red.

Günlük para birikimi yapma bakımından eğitim düzeyi ilköğretim olanla eğitim düzeyi lise olan arasında istatistiksel olarak farklılık bulunmaktadır.

$$H_0 : \beta_5 = 0$$

p-value=0,21 >  $\alpha = 0,05$  olduğundan H0 kabul.

Günlük para birikimi yapma bakımından eğitim düzeyi ilköğretim olanla eğitim düzeyi üniversite olan arasında istatistiksel olarak farklılık bulunmamaktadır.)



## B) ARTIK DEĞER İNCELEMESİ

```
> influence.measures(sonuc)
```

```
Influence measures of  
lm(formula = lny ~ x1 + x2 + x3 + x4)
```

	dfb.1	dfb.x1	dfb.x2	dfb.x3	dfb.x42	dfb.x43	dffit	cov.r	cook.d	hat	inf
1	-0.01067	-0.026430	0.002588	0.034018	0.032130	0.031903	-0.05297	1.248	4.78e-04	0.0918	
2	-0.02497	0.061258	-0.034918	-0.021892	-0.059188	-0.063648	0.08833	1.229	1.33e-03	0.0847	
3	-0.02469	0.043049	-0.060777	-0.005570	0.034789	0.021258	-0.10428	1.218	1.85e-03	0.0809	
4	-0.01763	-0.003741	0.003321	0.014050	0.028308	0.027780	-0.05178	1.193	4.56e-04	0.0523	
5	-0.01928	0.026573	-0.051086	0.004514	0.033551	0.022238	-0.08776	1.223	1.31e-03	0.0807	
7	0.02304	0.029148	-0.111334	0.010653	-0.076554	-0.093193	0.21238	1.203	7.62e-03	0.1019	
8	-0.11641	0.582482	0.018559	-0.589343	0.007166	-0.007550	-0.91061	0.853	1.30e-01	0.1770	
9	-0.37303	0.159514	-0.774537	0.466536	0.309762	0.158874	-1.02531	0.886	1.65e-01	0.2122	
10	-0.15424	-0.010569	-0.162966	0.249566	-0.094845	-0.114929	0.33167	1.156	1.84e-02	0.1163	
11	-0.19565	0.277771	-0.307982	0.042659	-0.240701	-0.287081	0.42452	1.042	2.97e-02	0.1043	
12	-0.13439	0.059517	-0.085688	0.084613	0.055974	0.038367	-0.17422	1.193	5.13e-03	0.0857	
13	-0.00643	-0.032393	0.012752	0.033574	0.025289	0.027338	-0.04853	1.414	4.01e-04	0.1964	*
14	-0.05491	0.249593	0.050374	-0.272051	0.026092	0.026020	-0.44851	1.148	3.34e-02	0.1472	
15	-0.00255	0.011963	-0.011033	0.001706	-0.045419	-0.045208	0.07233	1.174	8.89e-04	0.0449	
16	-0.13117	0.227510	-0.144090	-0.047213	-0.174106	-0.195001	0.27905	1.231	1.31e-02	0.1340	
17	-0.00139	-0.071329	0.034436	0.075416	-0.034145	-0.023819	0.14226	1.209	3.43e-03	0.0854	
18	-0.03931	0.051180	-0.025691	-0.001003	-0.059480	-0.061712	0.09095	1.218	1.41e-03	0.0777	
19	-0.02853	0.097754	-0.033362	-0.074337	0.006527	-0.002442	-0.13344	1.295	3.02e-03	0.1345	
21	-0.03574	0.008012	-0.060686	0.041845	-0.057512	0.044132	-0.11675	1.200	2.31e-03	0.0733	
22	0.25599	-0.415588	0.024394	0.270757	-0.161165	-0.138661	0.79684	0.512	9.33e-02	0.0801	*
23	-0.06667	0.064144	-0.178611	0.052504	0.098165	0.060624	-0.27085	1.180	1.23e-02	0.1084	
24	0.00740	-0.000953	-0.004403	-0.002698	-0.005325	-0.005904	0.01745	1.269	5.19e-05	0.1038	
25	-0.03428	0.054957	-0.036236	-0.003341	-0.069436	-0.073181	0.10280	1.190	1.79e-03	0.0633	
26	0.00486	-0.002438	0.000652	-0.002142	-0.012521	0.001157	-0.01983	1.228	6.69e-05	0.0739	
27	-0.01538	0.004590	-0.002977	0.010617	-0.019162	-0.002306	-0.03009	1.274	1.54e-04	0.1076	
28	0.00925	-0.007407	0.000507	-0.000696	-0.052734	0.003402	-0.07881	1.208	1.06e-03	0.0688	
31	0.00575	0.002066	0.002658	-0.008820	0.015319	-0.000763	0.02440	1.231	1.01e-04	0.0770	
36	-0.80272	0.296202	-0.633434	0.703931	0.316035	-0.160283	1.00962	1.049	1.63e-01	0.2552	
38	0.04690	0.199491	0.054716	-0.287731	-0.479250	-0.085968	-0.62702	0.771	6.17e-02	0.0920	
40	-0.01454	-0.014043	0.165348	-0.044359	-0.472410	0.011177	-0.67483	0.632	6.94e-02	0.0775	
43	-0.14370	0.086050	-0.108081	0.087293	-0.166796	-0.043134	-0.24357	1.187	9.99e-03	0.1036	
46	0.01352	0.031820	0.001111	-0.048153	0.017089	-0.013877	0.06696	1.363	7.63e-04	0.1674	
48	0.12947	-0.050284	0.076771	-0.099504	-0.070447	0.026173	-0.17791	1.315	5.37e-03	0.1543	
50	-0.00695	0.007684	0.000404	-0.002158	0.005459	-0.003469	0.01797	1.318	5.50e-05	0.1373	
51	0.77386	-0.514339	0.385503	-0.326742	0.713055	0.248707	1.09545	0.595	1.78e-01	0.1523	*
54	0.33671	-0.723520	0.734816	0.156461	0.675001	0.346855	1.07053	0.753	1.76e-01	0.1866	
55	-0.08598	0.007411	0.045787	0.051742	-0.115077	-0.002882	-0.21810	1.309	8.06e-03	0.1591	
59	-0.00304	0.001163	0.000496	0.001337	0.003716	-0.000535	0.00768	1.274	1.01e-05	0.1069	
61	-0.00193	0.003003	-0.001675	-0.000761	-0.001105	0.002461	0.00641	1.245	7.00e-06	0.0864	
62	0.01355	-0.015905	0.015481	-0.002002	0.004956	-0.030990	-0.05896	1.220	5.91e-04	0.0731	
65	-0.00190	0.000236	0.000629	0.001235	-0.000258	0.002735	0.00526	1.275	4.71e-06	0.1075	
66	0.05553	0.035618	-0.049419	-0.069105	-0.006118	-0.162688	-0.25240	1.159	1.07e-02	0.0938	
69	-0.06054	0.038283	0.003896	0.012603	-0.019070	-0.155549	-0.22133	1.151	8.24e-03	0.0808	
70	-0.04805	0.040875	0.011253	-0.005171	-0.020909	-0.137257	-0.19505	1.178	6.42e-03	0.0845	
71	-0.00523	0.010887	0.005672	-0.009624	-0.005734	0.031627	0.05713	1.227	5.55e-04	0.0777	
72	-0.09404	0.033425	-0.093491	0.092429	-0.003246	0.116493	0.22754	1.166	8.72e-03	0.0889	
74	0.46011	-0.223781	0.512954	-0.410541	0.033484	-0.170207	-0.71011	1.060	8.20e-02	0.1876	
81	-0.01646	0.003970	0.008762	0.007096	-0.003572	-0.025029	-0.04684	1.300	3.73e-04	0.1265	
84	0.01109	-0.278203	0.478900	0.083466	0.047862	0.475414	0.80930	0.955	1.05e-01	0.1808	
91	-0.00757	0.002712	-0.013911	0.010249	0.000699	-0.024689	-0.03719	1.237	2.36e-04	0.0824	
93	-0.00992	0.003785	-0.018223	0.013180	0.000816	-0.042545	-0.06224	1.223	6.59e-04	0.0754	
94	0.71399	-0.510496	0.410781	-0.286327	0.181029	0.650647	0.95042	0.808	1.41e-01	0.1743	
98	-0.01701	-0.032153	0.018230	0.043187	0.010586	-0.028371	-0.08537	1.290	1.24e-03	0.1241	

```

> inf<-ls.diag(sonuc)
> inf
$`std.dev`
[1] 1.079822

$hat
[1] 0.09180260 0.08471365 0.08092628 0.05233688 0.08066346 0.10187015 0.17695187 0.21219188 0.11634193
[10] 0.10432636 0.08573027 0.19635654 0.14716044 0.04494100 0.13403871 0.08539488 0.07773163 0.13449440
[19] 0.07326241 0.08010793 0.10839762 0.10381503 0.06333460 0.07387677 0.10761058 0.06877716 0.07698862
[28] 0.25522728 0.09204566 0.07752718 0.10356211 0.16741762 0.15425648 0.13728090 0.15225516 0.18656986
[37] 0.15910399 0.10688362 0.08643100 0.07306281 0.10746935 0.09382735 0.08081152 0.08448649 0.07770497
[46] 0.08886809 0.18760225 0.12645861 0.18083808 0.08241480 0.07536435 0.17427622 0.12411061

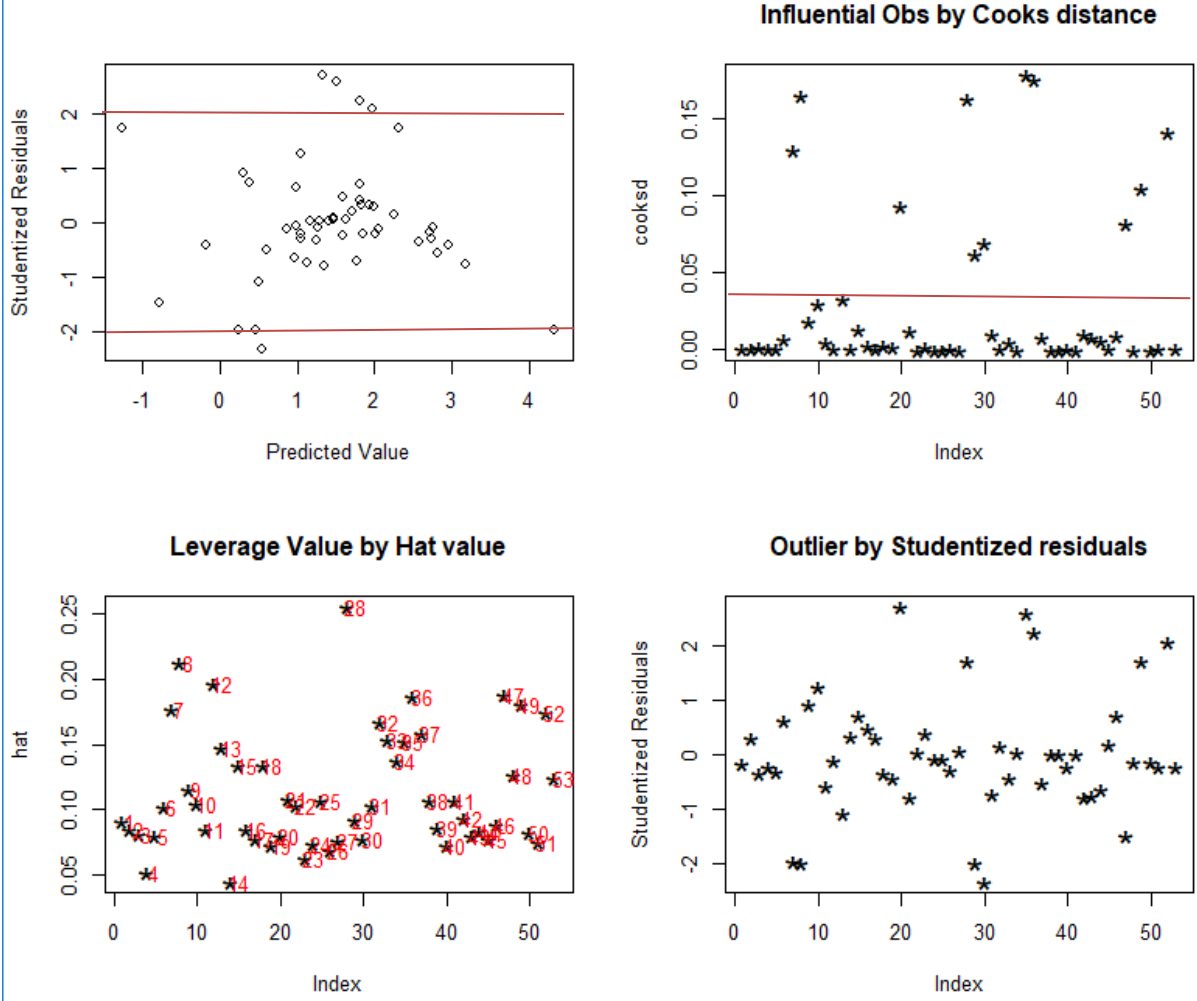
$std.res
[1] -0.16835923 0.29322000 -0.35473832 -0.22261292 -0.29917933 0.63468624 -1.90679740 -1.91727799
[9] 0.91568781 1.23669135 -0.57308703 -0.09922425 -1.07781758 0.33664885 0.71306413 0.46950662
[17] 0.31634684 -0.34175365 -0.41894432 2.53584709 -0.78010278 0.05182550 0.39892183 -0.07095138
[25] -0.08757370 -0.29284422 0.08539207 1.68953946 -1.91164972 -2.22552655 -0.72035165 0.15090022
[33] -0.42029604 0.04553691 2.44150785 2.14594001 -0.50544795 0.02244786 0.02106319 -0.21218352
[41] 0.01531255 -0.78760330 -0.75000543 -0.64613654 0.19886001 0.73225596 -1.45945197 -0.12441616
[49] 1.68751447 -0.12542722 -0.22023581 2.00016584 -0.22910608

$stud.res
[1] -0.16660879 0.29034956 -0.35141498 -0.22034817 -0.29626170 0.63060617 -1.96389462 -1.97560221
[9] 0.91408445 1.24386980 -0.56894893 -0.09817328 -1.07971686 0.33345049 0.70928462 0.46557812
[17] 0.31329708 -0.33851930 -0.41523955 2.70023934 -0.77680464 0.05127266 0.39532500 -0.07019628
[25] -0.08664412 -0.28997678 0.08448531 1.72466301 -1.96931299 -2.32779122 -0.71661402 0.14932244
[33] -0.41658437 0.04505087 2.58486846 2.23531881 -0.50140653 0.02220789 0.02083800 -0.21001473
[41] 0.01514881 -0.78437287 -0.74646414 -0.64208391 0.19681592 0.72859211 -1.47771502 -0.12310575
[49] 1.72246252 -0.12410648 -0.21799279 2.06877874 -0.22678235

$cooks
[1] 4.775263e-04 1.326271e-03 1.846733e-03 4.561455e-04 1.308923e-03 7.615085e-03 1.302830e-01
[8] 1.650165e-01 1.839906e-02 2.969037e-02 5.132746e-03 4.009280e-04 3.340896e-02 8.888234e-04
[15] 1.311710e-02 3.430287e-03 1.405776e-03 3.024889e-03 2.312521e-03 9.333263e-02 1.233107e-02
[22] 5.185591e-05 1.793416e-03 6.692827e-05 1.541334e-04 1.055632e-03 1.013686e-04 1.630380e-01
[29] 6.174541e-02 6.937687e-02 9.991221e-03 7.631368e-04 5.369874e-03 5.499417e-05 1.784316e-01
[36] 1.760375e-01 8.056388e-03 1.005083e-05 6.995602e-06 5.914501e-04 4.705494e-06 1.070489e-02
[43] 8.242259e-03 6.421255e-03 5.552935e-04 8.716441e-03 8.197807e-02 3.734800e-04 1.047765e-01
[50] 2.355001e-04 6.589007e-04 1.407292e-01 1.239600e-03

$sdfits
[1] -0.052970647 0.088332267 -0.104277324 -0.051782900 -0.087755967 0.212379269 -0.910611538
[8] -1.025306390 0.331674683 0.424518934 -0.174222087 -0.048527095 -0.448509605 0.072333188
[15] 0.279052822 0.142262904 0.090954997 -0.133444473 -0.116751081 0.796840820 -0.270854508
[22] 0.017450873 0.102797504 -0.019825934 -0.030087728 -0.078805862 0.024400064 1.009615280
[29] -0.627024349 -0.674829747 -0.243571041 0.066959476 -0.177912037 0.017971053 1.095448301
[36] 1.070533058 -0.218101863 0.007682611 0.006409442 -0.058962080 0.005256653 -0.252395621
[43] -0.221331534 -0.195053188 0.057128088 0.227544722 -0.710109547 -0.046839342 0.809300566
[50] -0.037194085 -0.062235732 0.950420327 -0.085366816

```

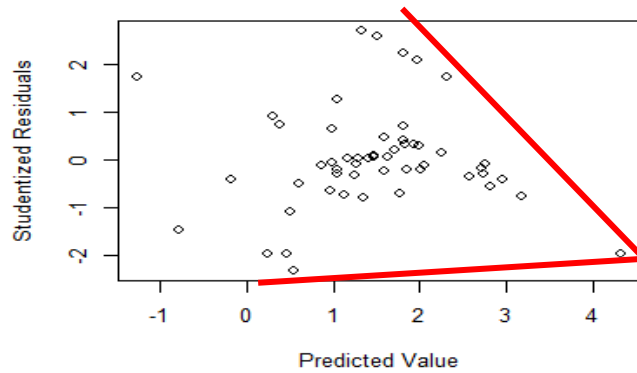


**NOT:**Kullandığım kodu tekrar tekrar kontrol etmeme rağmen indeksimi 100 yerine 50 olarak gösterdiği için analizlerime devam ettim. Artık incelemesinde de influence.measures(sonuc) kodu çıktısı üzerinden artık temizlemesi yaptım.

### 1.ADIM:

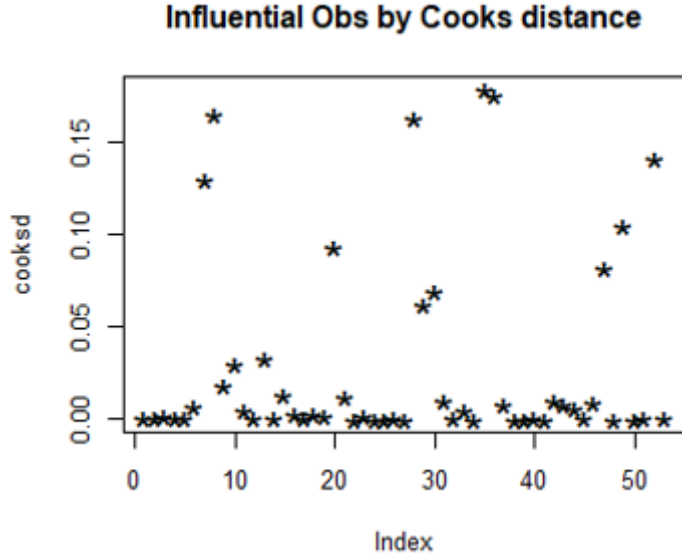
#### Aykırı Değer ( $r_i$ ) :

Standartlaştırılmış hataların  $(-2, +2)$  aralığında olması istenir. std.res kısmındaki verilere bakarsak 20, 30, 35, 52' inci veriler aykırıdır. Student tipi artıkların ise  $(-3, +3)$  arasında olması istenir. Student tipi artıklar için sınırları aşan değer bulunmamaktadır.



### Cook Uzaklığı(Di):

$n > 50$   $4/100 = 0,04$  olduğundan  $D_i > 0,04 \rightarrow D7, D8, D20, D28, D29, D30, D35, D36, D47, D49, D52$  gözlemleri etkin aykırı değer olduğu görülmektedir.



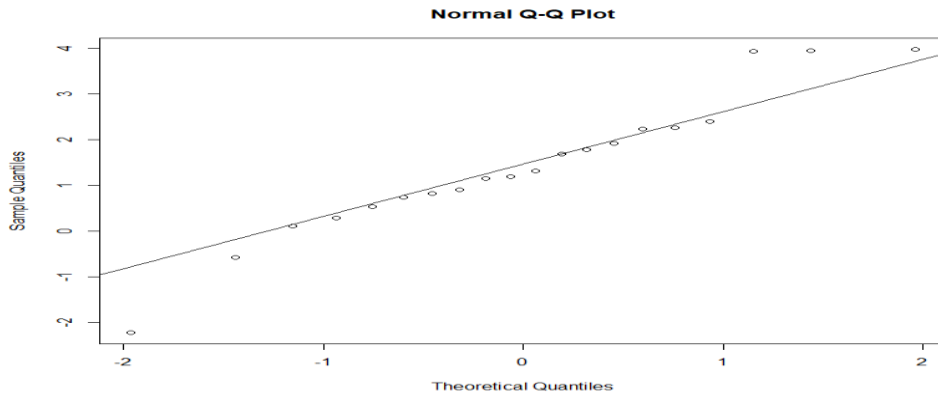
### Gözlem Uzaklığı (hii):

$h_{ii} > 2(k+1)/n \rightarrow 2(3+1)/100 = 0,08$  buna göre 1,2,3,5,6,7,8,9,10,11,12,13,15,16,18,20,21,22,25,28,29,31,32,33,34,35,36,37,38,39,41,42,43,44,46,47,48,49,50,52,53,55,59,61,65,66,69,70,72,74,81,84,91,94,98'inci değerler uç değerdir.

### 2.ADIM:

**Veriler normalliği sağlamadı, ln dönüşümü yapıp normallik sağlandıktan sonra aykırı değerler incelendi. 1.adımdaki bu değerlerimizi çıkardığımızda geriye 44 verimiz bulunmaktadır.**

- ❖ Veri kümesinin 30 veriden büyük olmasını istediğimiz için 1. adımdaki aykırı değerler çıkarıldıktan sonra sonlandırılmıştır. Son kalan 44 veri üzerinden tekrar veri okutma, normallik, vs. testleri yapılır
- ❖  $H_0$  = Verilerin dağılışı ile normal dağılım arasında fark yoktur.
- ❖  $H_5$  = Verilerin dağılışı ile normal dağılım arasında fark vardır.



❖

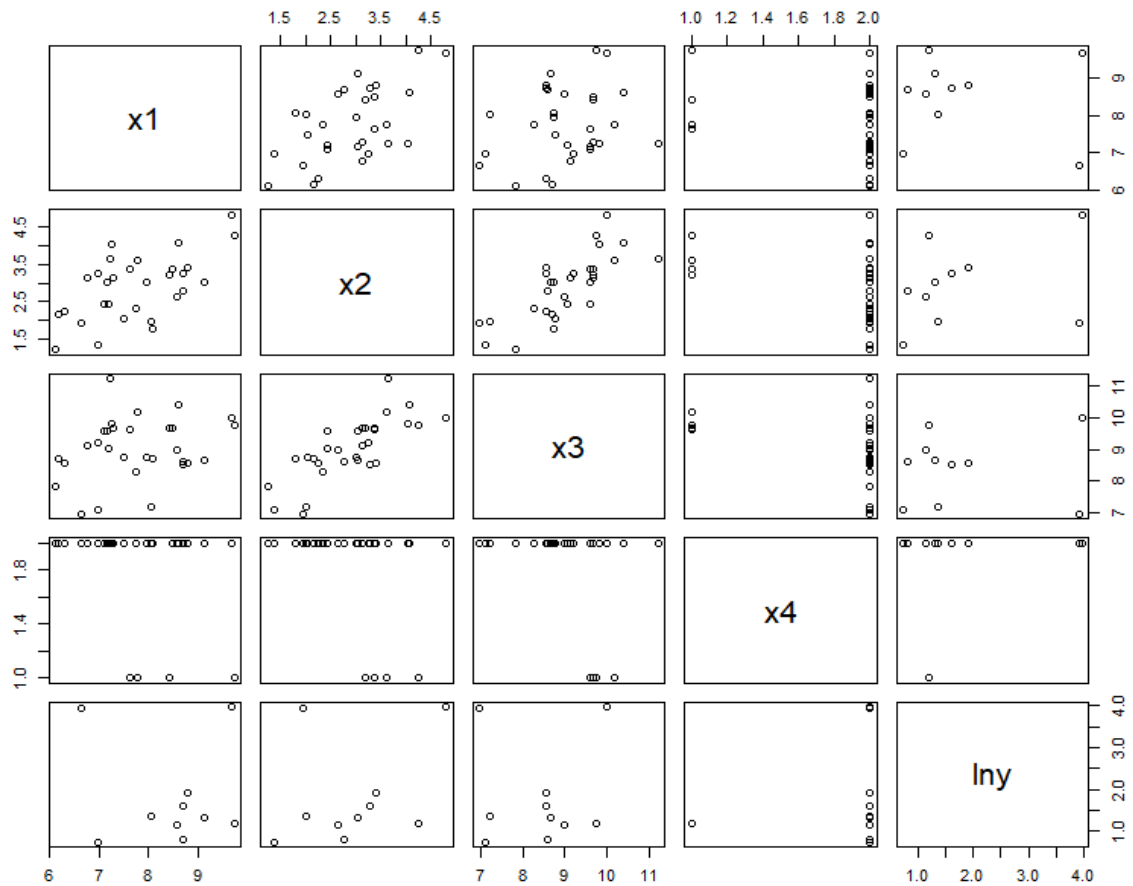
```
> lny<-log(y)
warning message:
In log(y) : NaNs produced
> shapiro.test(lny)
```

shapiro-wilk normality test

```
data: lny
W = 0.95292, p-value = 0.4136
```

- ❖ p-value=0,4136 >  $\alpha = 0,05$   $H_0$  kabul edilir. Verilerin dağılışı ile normal dağılım arasında fark yoktur.
- ❖ 

```
datayeni<-cbind(x1,x2,x3,x4,lny)
pairs(datayeni)
```



Bağımlı değişken ile bağımsız değişkenler arasında ilişki var ve doğrusal model oluşturulabilir.

## C) DEĞİŞEN VARYANSLILIK:

Student tipi artıklarla kestirim değerleri arasında çizilmiş olan grafik incelendiğinde yapının rasgele olmadığı, gözlemlerin nokta olarak ifade edildiği durumda sola doğru megafon olduğu görülmektedir. Ayrıca değişen varyans incelemesi için aşağıdaki test uygulanabilir.

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

Call:

```
lm(formula = abs(residuals(sonuc)) ~ fitted(sonuc))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.5453	-0.4073	-0.1521	0.2573	1.5693

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.4726	0.1904	2.482	0.0232 *
fitted(sonuc)	0.1581	0.1030	1.535	0.1422

---

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

Residual standard error: 0.5473 on 18 degrees of freedom

Multiple R-squared: 0.1157, Adjusted R-squared: 0.06658

F-statistic: 2.355 on 1 and 18 DF, p-value: 0.1422

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

Call:

```
lm(formula = abs(residuals(sonuc)) ~ fitted(sonuc))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.7418	-0.4764	-0.2419	0.2207	1.8829

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.9223	0.1797	5.132	4.51e-06 ***
fitted(sonuc)	-0.1356	0.1018	-1.333	0.189

---

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

Residual standard error: 0.7175 on 51 degrees of freedom

Multiple R-squared: 0.03365, Adjusted R-squared: 0.0147

F-statistic: 1.776 on 1 and 51 DF, p-value: 0.1886

### • Bruege-Pagan Testi

➤  $H_0$ : Varyanslar homojendir.

$H_s$ : Varyanslar homojen değildir.

```
> library(lmtest)
> bptest(sonuc)

studentized Breusch-Pagan test

data: sonuc
BP = 13.311, df = 5, p-value = 0.02063
```

**Yorum:** Aynca  $p\text{-value} = \text{değeri } 0,066 > 0,05$  olduğundan  $H_0$  kabul  $H_0$  hipotezini ret ederiz. Bu durumda değişen varyans sorunu bulunmamaktadır. bulunduğunu %5 anlamlılık düzeyinde söyleyebiliriz.

```
> bptest(sonuc)

studentized Breusch-Pagan test

data: sonuc
BP = 10.324, df = 5, p-value = 0.06656
```

- White Testi

**Model Denklemi:**

$$e_i^2 : \alpha_1 + \alpha_2 X_1 + \alpha_3 X_2 + \alpha_4 X_3 + \alpha_5 X_1^2 + \alpha_6 X_2^2 + \alpha_7 X_3^2 + \alpha_8 X_1 X_2 + \alpha_9 X_1 X_3 + \alpha_{10} X_2 X_3$$

$$H_o : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = \alpha_{10} \text{ (Değişen varyanslılık yoktur.)}$$

$$H_s : \text{Değişen varyanslılık vardır.}$$

```
sqres=res^2
sqx1=x1*x1
sqx2=x2*x2
sqx3=x3*x3
x1x2x3=x1*x2*x3
WH =lm(sqres ~ x1+x2+x3+sqx1+sqx2+sqx3+x1x2x3)
WHS = summary(WH)
```

## D) ÖZ İLİŞKİ SORUNU :

### Durbin-Watson Testi

➤  $H_o$ :Özilişki sorunu yoktur.

$H_s$ :Özilişki sorunu vardır.

Durbin-watson test

data: sonuc

DW = 2.5898, p-value = 0.8467

alternative hypothesis: true autocorrelation is greater than 0

```
> dwtest(sonuc)
```

Durbin-watson test

data: sonuc

DW = 1.845, p-value = 0.2125

alternative hypothesis: true autocorrelation is greater than 0

---

**Yorum:** Durbin Watson test istatistiği [2,5898](#)[1,845](#) olarak bulunduğu ve 2 değerine yakın bir değer olmasından dolayı ayrıca p değeri  $0.2125 > \alpha=0.05$  koşulunu sağladığı için %5 anlamlılık düzeyinde  $H_o$  hipotezini kabul edip öz ilişki sorunu yoktur diyebiliriz.



## E) ÇOKLU BAĞLANTI SORUNU

```
> library(car)
> vif(sonuc)
      GVIF Df GVIF^(1/(2*Df))
x1 3.870612  1      1.967387
x2 3.737158  1      1.933173
x3 2.290285  1      1.513369
x4 1.341949  2      1.076302
> |
```

Özdeğer ve Özvektörlerin Bulunması

```
> ort1<-mean(x1)
> kt1<-sum((x1-ort1)^2)
> skx1<-(x1-ort1)/(kt1^0.5)
> ort2<-mean(x2)
> kt2<-sum((x2-ort2)^2)
> skx2<-(x2-ort2)/(kt2^0.5)
> ort3<-mean(x3)
> kt3<-sum((x3-ort3)^2)
> skx3<-(x3-ort3)/(kt3^0.5)

> x<-cbind(skx1,skx2,skx3)
> sm<- eigen(t(x)%*%x)
> signif(sm$values,3) #"lamda1=3,8" "lamda2=0,124" "lamda3=0,0789" "lamda4=0,000674" toplamışrareti(1/lamda)>
k ise güçlü çoklu bağlantı vardır. 0 a yakın özdeğer kadar güçlü çoklu bağlantı vardır.
[1] 2.240 0.592 0.166
> signif(sm$vectors,3) #denklem yazarsak (0,712x1-0,701x3=0)
      [,1] [,2] [,3]
[1,] 0.527 0.7770 0.344
[2,] 0.633 -0.0885 -0.769
[3,] 0.568 -0.6230 0.538
> v<-sm$vectors
> t(v)%*%v
      [,1] [,2] [,3]
[1,] 1.000000e+00 2.220446e-16 1.110223e-16
[2,] 2.220446e-16 1.000000e+00 -5.551115e-17
[3,] 1.110223e-16 -5.551115e-17 1.000000e+00
> v %*% diag(sm$values) %*% t(v)
      [,1] [,2] [,3]
[1,] 1.0000000 0.6630273 0.4150063
[2,] 0.6630273 1.0000000 0.7688608
[3,] 0.4150063 0.7688608 1.0000000
```

Özdeğerlerden sıfıra çok yakın olan 0,166'dır ve 1 tane güçlü çoklu bağlantı olduğu söylenebilir. Bu özdeğere ait olan özvektör sütununa bakıldığında x1 ve x3 değişkenleri katsayılarından  $0,344X_1 - 0,538X_3 = 0$  şeklinde çoklu bağlantının yapısı yazılabilir. En son kod ile özvektörler ve özdeğerlerden ilişki matrisinin elde edilmesi gösterilmiştir.

## F) MODEL ANLAMLILIĞI :

```

> sonuc<-lm(lny~x1+x2+x3+x4)
> summary(sonuc)

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

Residuals:
    Min       1Q   Median       3Q      Max
-1.61723 -0.50846 -0.01706  0.49264  2.30169

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.8907     3.4818   3.989 0.001343 **
x1          -0.7450     0.6230  -1.196 0.251652
x2           2.4752     0.5793   4.273 0.000774 ***
x3          -1.5334     0.5818  -2.635 0.019580 *
x42          -1.1468     0.8491  -1.351 0.198246
x43          -0.1753     0.6936  -0.253 0.804148
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.062 on 14 degrees of freedom
(24 observations deleted due to missingness)
Multiple R-squared:  0.6412,    Adjusted R-squared:  0.5131
F-statistic: 5.004 on 5 and 14 DF,  p-value: 0.007752

```

$$\ln y_i = 13,8907 - 0,7450x_{1i} + 2,4752x_{2i} - 1,5334x_{3i} - 1,1468x_{42} - 0,1753x_{43} + 0,6295$$

Std. = (3,5) (0,6) (0,57) (0,8) (0,58) (0,84) (0,6936)

Error

Katsayıları yorumlayacak olursak;

- Bağımlı değişken günlük birikim miktarı (TL) olmak üzere; x1=aileden alınan günlük para miktarı, x2=çalışarak(iş vb.) kazanılan günlük para miktarı, x3= günlük harcama, x42=eğitim düzeyi lise, x43=eğitim düzeyi üniversite değişkenleri girilmiştir.
- X41= eğitim düzeyi ilköğretim kılavuz değişken olduğunu elde edilen modelde olmadığından görülebilir

### MODEL:

- H0:  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$  (Model anlamsızdır.)
- HS: En az bir  $\beta_j \neq 0$  (Model anlamlıdır.)
- P-value = 0,007752 <  $\alpha = 0,05$  olduğundan H0 reddedilir. Modelimiz istatistiksel olarak anlamlıdır. Ayrıca bağımlı değişkenin %67,12'sinin bağımsız değişkenlerce açıklandığı yorumlaması yapılabilir.

## G) KATSAYI YORUMLARI

$\beta_0$ : Bağımsız değişkenlerimiz 0 olduğunda günlük birikim miktarının aldığı ortalama değer 13,89 TL olduğunu söyleyebiliriz.

$\beta_1$ : Aileden alınan günlük para miktarı 1 TL azaldığında günlük birikim miktarı ortalama 0,74 lira azalmaktadır.

$\beta_2$ : Çalışarak (iş vb) kazanılan para miktarı 1 TL arttığında günlük birikim miktarı ortalama 2,47 TL artmaktadır.

$\beta_3$ : Günlük harcama miktarı 1 TL azaldığında günlük birikim miktarı ortalama 1,53 TL azalmaktadır. (Veriler rasgele olduğundan çok da geçerliliği olmayabilir.)

Etkileşim terimleri, aileden alınan para miktarı ( $x_1$ ), bir işte çalışarak kazanılan para miktarı ( $x_2$ ), günlük harcama miktarı ( $x_3$ ) değişkenlerinin günlük birikim miktarına ( $y$ ) etkisinin kişinin eğitim düzeyine ( $x_4$ ) göre farklılaştığını bize gösteriyor.

Katsayıların anlamlılığına bakarsak;

$$H_0 : \beta_0 = 0$$

p-value=0,001 <  $\alpha = 0,05$  olduğundan  $H_0$  red.

(Sabit terimin modele katkısı anlamlıdır.)

$$H_0 : \beta_1 = 0$$

p-value=0,25 >  $\alpha = 0,05$  olduğundan  $H_0$  kabul.

(Aileden alınan günlük para miktarı model üzerinde anlamsızdır.)

$$H_0 : \beta_2 = 0$$

p-value=0,00077 <  $\alpha = 0,05$  olduğundan  $H_0$  red.

(Çalışarak (iş vb.) kazanılan günlük para miktarı model üzerinde anlamlıdır.)

$$H_0 : \beta_3 = 0$$

p-value=0,195 >  $\alpha = 0,05$  olduğundan  $H_0$  kabul.

(Günlük harcama miktarı model üzerinde anlamsızdır.)

$$H_0 : \beta_4 = 0$$

p-value=0,198 >  $\alpha = 0,05$  olduğundan  $H_0$  kabul.

(Günlük para birikimi yapma bakımından eğitim düzeyi ilköğretim olanla eğitim düzeyi lise olan arasında istatistiksel olarak farklılık bulunmaktadır.)

$$H_0 : \beta_5 = 0$$

p-value=0,804 >  $\alpha = 0,05$  olduğundan  $H_0$  kabul.

(Günlük para birikimi yapma bakımından eğitim düzeyi ilköğretim olanla eğitim düzeyi üniversite olan arasında istatistiksel olarak farklılık bulunmamaktadır.)

## H) UYUM KESTİRİMİ

```
> new.speeds<-merve[7.436332:9,152636]
> predict(sonuc, newdata = new.speeds)
```

1	2	3	4	5	7	8	9	10
2.7112110	1.9856008	2.5754197	2.0146008	2.7329660	0.9742058	0.4647562	4.3345698	0.3062955
11	12	13	14	15	16	17	18	19
1.0459543	2.8151072	2.7544626	0.5031183	1.8238838	1.8018021	1.5950621	1.9317444	1.2402795
21	22	23	24	25	26	27	28	31
2.9590462	1.3189424	3.1870976	1.4596793	1.8084990	0.9772773	1.2730152	1.0509431	1.4597537
36	38	40	43	46	48	50	51	54
-1.2720292	0.2301616	0.5372238	1.7695830	2.2494506	-0.1840122	1.6333873	1.5171862	1.8133143
55	59	61	62	65	66	69	70	71
0.6092914	1.1619562	1.2927341	1.0345889	1.4092757	1.3433986	1.1182239	0.9576131	1.7117104
72	74	81	84	91	93	94	98	
0.3868678	-0.7906914	0.8660812	2.3135217	2.0480880	1.8476033	1.9634966	1.5807595	

```
> |
```

## İ) ÖN KESTİRİM

```
> new.prt <- data.frame(x1=20,x2=5,x3=15,x4=1)
> new.prt$x4<-as.factor(new.prt$x4)
> predict(sonuc, newdata= new.prt)
```

1
-3.93461

```
> |
```

## J) REGRESYON KATSAYILARI İÇİN GÜVEN ARALIKLARI VE YORUMU

```
> confint(sonuc, level = .95)
```

	2.5 %	97.5 %
(Intercept)	4.5660107	12.4708032
x1	-0.7082905	0.4818323
x2	0.8449666	2.1484143
x3	-1.6761619	-0.6800897
x42	-1.8799582	-0.3299774
x43	-1.3294225	0.3045373

```
> |
```

Aileden alınan günlük para miktarı değişkeni(x1) 0 değerini içermektedir. Bu nedenle anlamsızdır diyebiliriz.

## K) ÖN KESTİRİM VE UYUM KESTİRİMİ İÇİN GÜVEN ARALIKLARI

```
#güven aralıkları
predict(sonuc, newdata=new.speeds, interval = "prediction")
predict(sonuc, newdata = new.prt, interval = "prediction")

> predict(sonuc, newdata=new.speeds, interval = "prediction")
      fit      lwr      upr
```

1	2.7112110	0.44136647	4.981055
2	1.9856008	-0.27686279	4.248064
3	2.5754197	0.31690941	4.833930
4	2.0146008	-0.21384171	4.243043
5	2.7329660	0.47473028	4.991202
7	0.9742058	-1.30607980	3.254491
8	0.4647562	-1.89193890	2.821451
9	4.3345698	1.94285318	6.726287
10	0.3062955	-1.98891574	2.601507
11	1.0459543	-1.23687140	3.328780
12	2.8151072	0.55158370	5.078631
13	2.7544626	0.37841924	5.130506
14	0.5031183	-1.82355882	2.829795
15	1.8238838	-0.39671415	4.044482
16	1.8018021	-0.51152998	4.115134
17	1.5950621	-0.66811185	3.858236
18	1.9317444	-0.32342598	4.186915
19	1.2402795	-1.07351732	3.554076
21	2.9590462	0.70855669	5.209536
22	1.3189424	-0.93871284	3.576598
23	3.1870976	0.90006773	5.474127
24	1.4596793	-0.82261787	3.741977
25	1.8084990	-0.43155772	4.048556
26	0.9772773	-1.27385631	3.228411
27	1.2730152	-1.01320257	3.559233
28	1.0509431	-1.19483901	3.296725
31	1.4597537	-0.79463912	3.714147
36	-1.2720292	-3.70583113	1.161773
38	0.2301616	-2.03993552	2.500259
40	0.5372238	-1.71773264	2.792180
43	1.7695830	-0.51245274	4.051619
46	2.2494506	-0.09767948	4.596581
48	-0.1840122	-2.51787432	2.149850
50	1.6333873	-0.68324931	3.950024
51	1.5171862	-0.81465177	3.849024
54	1.8133143	-0.55299059	4.179619
55	0.6092914	-1.72946641	2.948049
59	1.1619562	-1.12351114	3.447424
61	1.2927341	-0.97151976	3.556988
62	1.0345889	-1.21569134	3.284869
65	1.4092757	-0.87679627	3.695348
66	1.3433986	-0.92854967	3.615347
69	1.1182239	-1.14016652	3.376614
70	0.9576131	-1.30461357	3.219840
71	1.7117104	-0.54343212	3.966853
72	0.3868678	-1.87992426	2.653660
74	-0.7906914	-3.15802550	1.576643
81	0.8660812	-1.43950659	3.171669
84	2.3135217	-0.04706102	4.674104
91	2.0480880	-0.21197691	4.308153
93	1.8476033	-0.40508895	4.100296
94	1.9634966	-0.39051808	4.317511
98	1.5807595	-0.72242417	3.883943

```
> predict(sonuc, newdata = new.prt, interval = "prediction")
      fit      lwr      upr
1 -3.93461 -10.31692  2.447697
```

**Yorum:** 0 değeri aralığa girdiği için anlamsızdır yorumu yapılabilir.

## L) EN İYİ MODEL BULUNMASI

### L.1) İleriye Doğru Seçim Yöntemi

```
> library(stats)
> lm.null <- lm(lny ~ 1)
> forward <- step(lm.null, lny~x1+x2+x3, direction = "forward")
Start:  AIC=17.79
lny ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ x2	1	5.7941	38.245	16.966
<none>			44.040	17.787
+ x1	1	0.1191	43.920	19.733
+ x3	1	0.0002	44.039	19.787

```
Step:  AIC=16.97
lny ~ x2
```

	Df	Sum of Sq	RSS	AIC
+ x3	1	14.092	24.154	9.7743
+ x1	1	13.331	24.915	10.3947
<none>			38.245	16.9659

```
Step:  AIC=9.77
lny ~ x2 + x3
```

	Df	Sum of Sq	RSS	AIC
+ x1	1	5.4972	18.657	6.6095
<none>			24.154	9.7743

```
Step:  AIC=6.61
lny ~ x2 + x3 + x1
```

Bağımlı değişken lny(birikim miktarı) olduğu durumda birinci adımda modele x2 değişkeni girmiştir. İkinci adımda x3, üçüncü adımda x1 modele girmiştir. En iyi model olarak bütün değişkenlerin içinde yer aldığı model denklemi aşağıdadır.

```
> forward
```

```
call:
```

```
lm(formula = lny ~ x2 + x3 + x1)
```

```
Coefficients:
```

(Intercept)	x2	x3	x1
13.657	2.480	-1.252	-1.049

```
> summary(forward)
```

```
call:
```

```
lm(formula = lny ~ x2 + x3 + x1)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.4790	-0.6936	-0.2493	0.8581	2.2539

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	13.6572	3.5049	3.897	0.001283	**
x2	2.4797	0.5361	4.625	0.000281	***
x3	-1.2520	0.5404	-2.317	0.034107	*
x1	-1.0488	0.4831	-2.171	0.045297	*

```
---
```

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

```
Residual standard error: 1.08 on 16 degrees of freedom
```

```
(24 observations deleted due to missingness)
```

```
Multiple R-squared:  0.5764,    Adjusted R-squared:  0.4969
```

```
F-statistic: 7.256 on 3 and 16 DF,  p-value: 0.002733
```

## **MODEL**

$$y = 13.657 - 1.049x_1 + 2.480x_2 - 1.252x_3 \pm 1.08$$

p değerinden görüldüğü üzere modeldeki değişken %95 güven düzeyinde anlamlıdır (p=0.002733). Modelin de anlamlı olduğu görülmektedir.

## **L.2) Geriye Doğru Seçim Yöntemi**



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

```
Start: AIC=7.29
```

```
lmy ~ x1 + x2 + x3 + x4
```

	Df	Sum of Sq	RSS	AIC
- x4	2	2.8564	18.657	6.6095
- x1	1	1.6137	17.414	7.2308
<none>			15.800	7.2859
- x3	1	7.8384	23.639	13.3430
- x2	1	20.6024	36.403	21.9782

```
Step: AIC=6.61
```

```
lmy ~ x1 + x2 + x3
```

	Df	Sum of Sq	RSS	AIC
<none>			18.657	6.6095
- x1	1	5.4972	24.154	9.7743
- x3	1	6.2582	24.915	10.3947
- x2	1	24.9466	43.603	21.5880

Geriye doğru seçim yönteminin özelliğinden tüm değişkenler modelde olarak başlıyor. İlk model tüm bağımsız değişkenlerin modelde bulunduğu durumdur. İkinci modelde x4 değişkeni modelden çıkmış. Son modelde x1, x2, x3 değişkenleri modelde kalmıştır.

```
> summary(backward)
```

```
Call:
```

```
lm(formula = lmy ~ x1 + x2 + x3)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.4790	-0.6936	-0.2493	0.8581	2.2539

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	13.6572	3.5049	3.897	0.001283	**
x1	-1.0488	0.4831	-2.171	0.045297	*
x2	2.4797	0.5361	4.625	0.000281	***
x3	-1.2520	0.5404	-2.317	0.034107	*

```
---
```

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

```
Residual standard error: 1.08 on 16 degrees of freedom
```

```
(24 observations deleted due to missingness)
```

```
Multiple R-squared:  0.5764,    Adjusted R-squared:  0.4969
```

```
F-statistic: 7.256 on 3 and 16 DF,  p-value: 0.002733
```

Aynı bağımsız değişkenler ile model kurulduğundan ileri ve geriye doğru seçim yöntemleri aynı sonuçları vermiştir. Burada da x1, x2, x3 değişkeninin olduğu model en iyi model olarak elde edilmiştir. Modelin anlamlı olduğu söylenebilir (p=0.002733).

### L.3) Adımsal Seçim Yöntemi

```

> library(MASS)
warning message:
package 'MASS' was built under R version 3.5.3
> step.model <- stepAIC(sonuc, direction = "both", trace = FALSE)
> summary(step.model)

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

Residuals:
    Min       1Q   Median       3Q      Max
-1.4790 -0.6936 -0.2493  0.8581  2.2539

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6572     3.5049   3.897 0.001283 **
x1           -1.0488     0.4831  -2.171 0.045297 *
x2            2.4797     0.5361   4.625 0.000281 ***
x3           -1.2520     0.5404  -2.317 0.034107 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.08 on 16 degrees of freedom
(24 observations deleted due to missingness)
Multiple R-squared:  0.5764,    Adjusted R-squared:  0.4969
F-statistic: 7.256 on 3 and 16 DF,  p-value: 0.002733

```

İlk modele  $x_1$ ,  $x_2$ ,  $x_3$  değişkeni alınarak başlamış.  $x_4$  bağımsız değişkeni anlamlı olmadığından modelde yer almamıştır. En iyi model yalnızca  $x_1$ ,  $x_2$ ,  $x_3$ 'ün olduğu modeldir.

## M) RIDGE REGRESYON

```
> ridge$coef
      0.00      0.05      0.10      0.15      0.20      0.25      0.30      0.35      0.40
x1 -0.84163963 -0.80999741 -0.7805863 -0.7531824 -0.7275905 -0.7036395 -0.6811791 -0.6600765 -0.6402140
x2  2.51612225  2.46597804  2.4181705  2.3725179  2.3288585  2.2870481  2.2469572  2.2084693  2.1714791
x3 -1.37771277 -1.35917826 -1.3408130 -1.3226513 -1.3047200 -1.2870399 -1.2696269 -1.2524929 -1.2356464
x42 -0.45872227 -0.46143245 -0.4636412 -0.4654090 -0.4667878 -0.4678227 -0.4685531 -0.4690132 -0.4692331
x43 -0.08764931 -0.09660251 -0.1048033 -0.1123292 -0.1192478 -0.1256181 -0.1314920 -0.1369153 -0.1419285
      0.45      0.50      0.55      0.60      0.65      0.70      0.75      0.80      0.85
x1 -0.6214871 -0.6038023 -0.5870764 -0.5712343 -0.5562087 -0.5419387 -0.5283694 -0.5154510 -0.5031382
x2  2.1358913  2.1016190  2.0685832  2.0367113  2.0059371  1.9761995  1.9474423  1.9196136  1.8926653
x3 -1.2190930 -1.2028364 -1.1868782 -1.1712185 -1.1558564 -1.1407896 -1.1260154 -1.1115301 -1.0973297
x42 -0.4692391 -0.4690545 -0.4686998 -0.4681933 -0.4675510 -0.4667873 -0.4659150 -0.4649457 -0.4638895
x43 -0.1465676 -0.1508648 -0.1548485 -0.1585447 -0.1619762 -0.1651639 -0.1681266 -0.1708812 -0.1734433
      0.90      0.95      1.00
x1 -0.4913897 -0.4801679 -0.4694385
x2  1.8665529  1.8412349  1.8166730
x3 -1.0834097 -1.0697652 -1.0563912
x42 -0.4627555 -0.4615522 -0.4602868
x43 -0.1758267 -0.1780444 -0.1801078
> select(ridge)
modified HKB estimator is 0.3698106
modified L-w estimator is 2.397922
smallest value of GCV at 0.4
> ridge$coef[,ridge$lam == 0.4]
      x1      x2      x3      x42      x43
-0.6402140  2.1714791 -1.2356464 -0.4692331 -0.1419285
> |
```

```
library(MASS)

ridge <- lm.ridge(lm~x1+x2+x3+x4 ,lambda = seq(0,1,0.05))

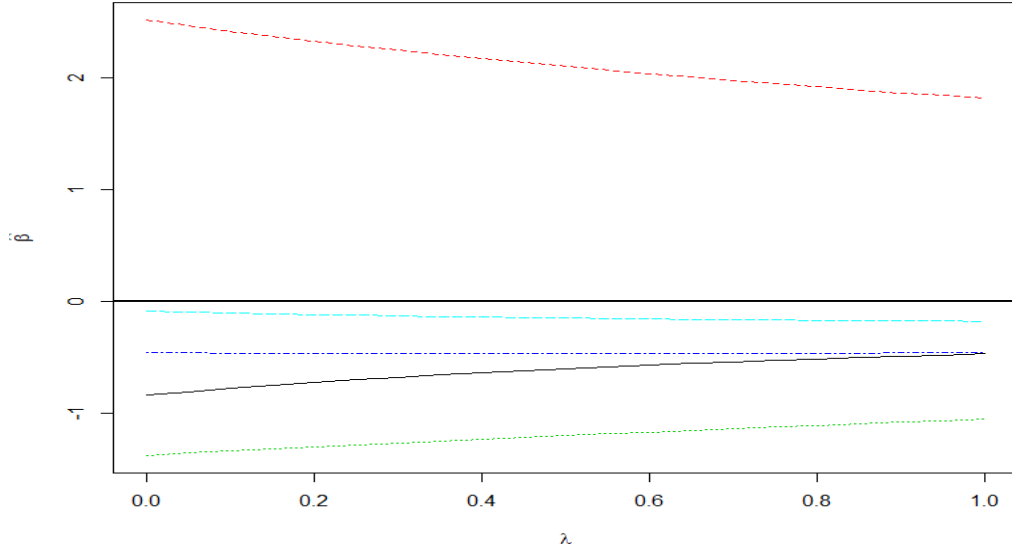
matplot(ridge$lambda,t(ridge$coef),type="l",xlab=expression(lambda),
        ylab=expression(hat(beta)))

abline(h=0,lwd=2)

ridge$coef|

select(ridge)

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



Lamda parametresine göre katsayılar değişmektedir. Hızlı azalış ve artışların bittiği yerde bir lamda parametresi belirlenerek katsayı kestirimleri elde edilir ve model kurulur.

## RAPOR

Bu analizi hazırlarken RStudio programında gerekli kodları kullandım ve çoklu regresyon analizi yaptım. Yaptığım analiz sonuçlarını hem derste öğrendiğimiz hem de uygulama derslerinde yaptığımız yorumlardan yararlanarak yorumladım. Elde ettiğim sonuçlara bağlı olarak varsayım bozulmalarını, artık incelemelerinin gerekli model kurulmasını yaparak analizimi tamamladım.

## KAYNAKÇA

<https://www.rstudio.com/online-learning/>

<http://www.barissanli.com/calismalar/dersler/r/rders1.php>

<https://www.veribilimi.co/r-ile-coklu-dogrusal-regresyonbaglanim-cozumlemesi/>

Dawson B, Trapp RG. Statistical Methods for Multiple Variables. Basic & Clinical Biostatistics. Lange Medical books/McGraw Hill Medical Publishing Division, 2001, USA, 236-242.