

Sleep-Time Prediction

Eren Talha Temur

01/01/2025

Önce gerekli kütüphaneleri indirmeliyiz

```
options(warn=-1)
options(message=FALSE)
library(ggplot2) # Veri görselleştirme ve grafik oluşturmak için
library(dplyr) # Veri manipülasyonu ve temizleme işlemi için kullanırız.
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, lag
```

```
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union
```

```
library(broom) # regresyon sonuçlarını tabloya dönüştürmek için kullanırız
library(ggpubr) # ggplot2'ye yardımcı kütüphanedir.
library(ISLR) # istatistiksel öğrenme için kullanılan veri seti koleksiyonudur.
library(PerformanceAnalytics) # Finansal veri analizi ve zaman serisi görselleştirme araçları sağlar. R
```

```
## Zorunlu paket yükleniyor: xts
```

```
## Zorunlu paket yükleniyor: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
## as.Date, as.Date.numeric
```

```
##
## ##### Warning from 'xts' package #####
## #
```

```
## # The dplyr lag() function breaks how base R's lag() function is supposed to #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or #
## # source() into this session won't work correctly. #
## # #
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop #
## # dplyr from breaking base R's lag() function. #
## # #
## # Code in packages is not affected. It's protected by R's namespace mechanism #
## # Set 'options(xts.warn_dplyr_breaks_lag = FALSE)' to suppress this warning. #
## # #
## #####
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
## first, last
```

```
##
## Attaching package: 'PerformanceAnalytics'
```

```
## The following object is masked from 'package:graphics':
##
## legend
```

```
library(lmtest) # doğrusal model üstüne hipotez testleri yapar
library(faraway) # regresyon modelleme ve analiz uygulamaları için araçlar
library(mice) # eksik verileri düzenlemek ve işlemek için kullanırız
```

```
##
## Attaching package: 'mice'
```

```
## The following object is masked from 'package:faraway':
##
## mammalsleep
```

```
## The following object is masked from 'package:stats':
##
## filter
```

```
## The following objects are masked from 'package:base':
##
## cbind, rbind
```

```
library(caret) # makine öğrenmesi modeli oluşturup eğitmek, analiz etmek için kullanırız.
```

```
## Zorunlu paket yükleniyor: lattice
```

```
##
## Attaching package: 'lattice'

## The following object is masked from 'package:faraway':
##
##      melanoma
```

```
options(scipen = 999)
```

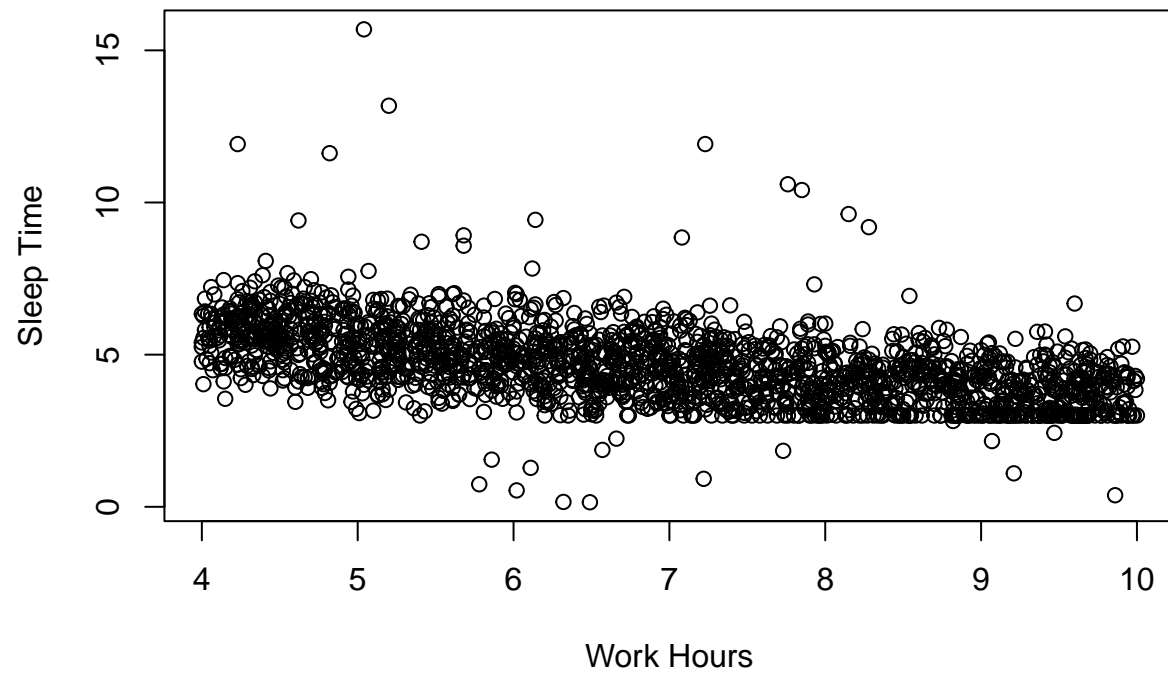
Veriyi hazırlamak ve Overfitting riskini azaltmak için test-eğitim parçalaması yapıp, gerekli işlemleri uyguluyoruz.

```
model_data <- read.csv("sleep_time_prediction_dataset.csv")
model_data <- model_data %>% filter(WorkoutTime + PhoneTime + WorkHours + RelaxationTime + SleepTime <= 24)
#Veriye ön inceleme uyguladığımızda, bazı satırlarda
#verilen zaman dilimleri toplamlarının 24 saati geçtiğini gözlemledik.
#Bunun üzerine daha doğru tahminler yapabilmek adına veriyi filtreledik.
set.seed(145)
sampleindex <- sample(1:nrow(model_data), size = 0.8 * nrow(model_data))
egitimset <- model_data[sampleindex, ]
testset <- model_data[-sampleindex, ]

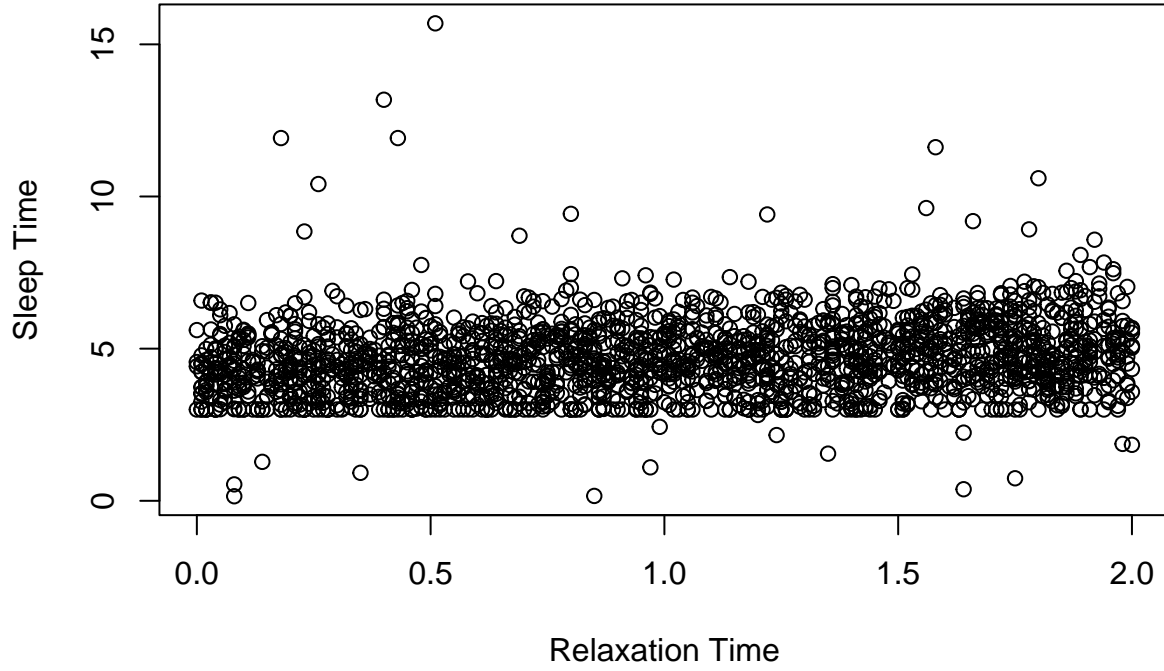
# Eğitim verisinde değişkenler arası korelasyonu görelim.
cor(egitimset)
```

```
##      WorkoutTime  ReadingTime  PhoneTime  WorkHours
## WorkoutTime    1.000000000    0.007175988  0.029281774  0.04376092
## ReadingTime     0.007175988    1.000000000  0.033629960 -0.01914353
## PhoneTime       0.029281774    0.033629960  1.000000000  0.02286417
## WorkHours       0.043760918   -0.019143527  0.022864168  1.00000000
## CaffeineIntake  -0.067099394   -0.009501900 -0.005923921 -0.01217535
## RelaxationTime  0.016013230    0.024638305 -0.037737110  0.03820581
## SleepTime       0.279740772    0.120733096 -0.529166482 -0.52246264
##
##      CaffeineIntake  RelaxationTime  SleepTime
## WorkoutTime       -0.067099394      0.01601323  0.2797408
## ReadingTime        -0.009501900      0.02463830  0.1207331
## PhoneTime          -0.005923921     -0.03773711 -0.5291665
## WorkHours          -0.012175350      0.03820581 -0.5224626
## CaffeineIntake      1.000000000     -0.04260465 -0.1447115
## RelaxationTime     -0.042604651      1.00000000  0.1799388
## SleepTime          -0.144711509      0.17993878  1.0000000
```

```
#Bazı değişkenler arası ilişkiyi görselleştirelim
#Görselleştirme üzerinden aykırı değer varlığı gözlemledik.
plot(model_data$WorkHours, model_data$SleepTime, xlab = "Work Hours", ylab = "Sleep Time")
```



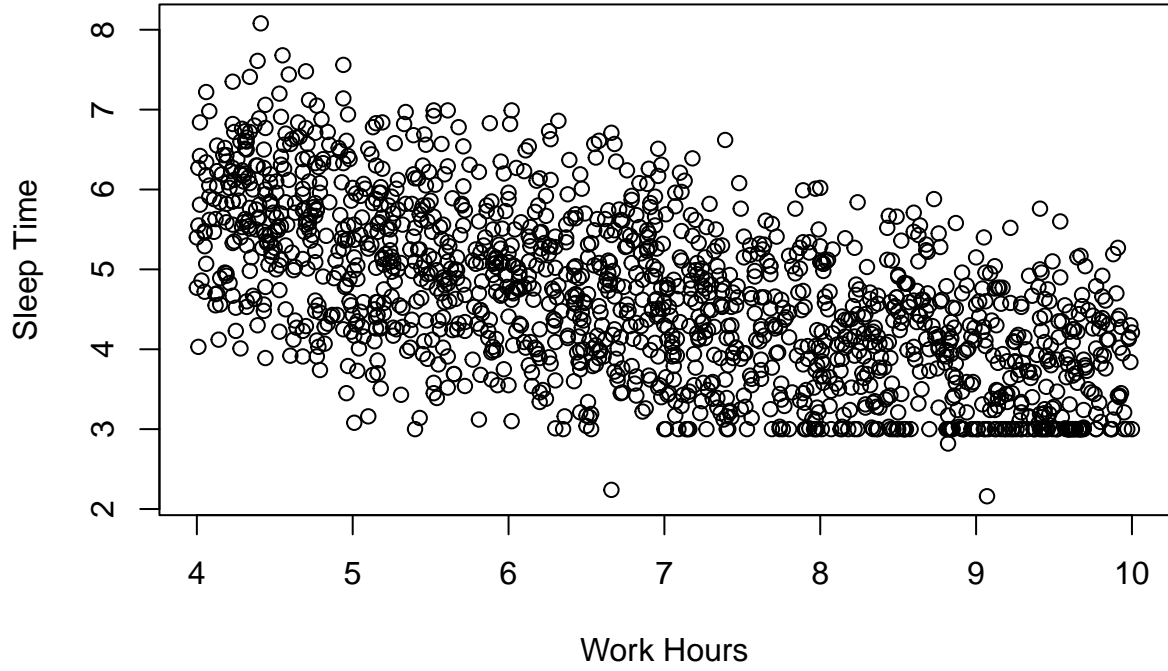
```
plot(model_data$RelaxationTime, model_data$SleepTime,xlab = "Relaxation Time", ylab = "Sleep Time")
```



Aykırı Değer tespiti ve verinin doğru hale getirilmesi

```
dist <- mahalanobis(egitimset, center = colMeans(egitimset), cov = cov(egitimset))
cutoff <- qchisq(p = 0.95, df = ncol(egitimset))
index <- which(dist > cutoff)
egitimsettemiz <- egitimset[-index, ]

# Temizlenmiş veriyi görselize etme, inceleme
plot(egitimsettemiz$WorkHours, egitimsettemiz$SleepTime,
     xlab = "Work Hours", ylab = "Sleep Time")
```



```
cor(egitimsettemiz)
```

```
##          WorkoutTime  ReadingTime  PhoneTime  WorkHours
## WorkoutTime      1.000000000  0.007765291  0.030239860  0.052998592
## ReadingTime      0.007765291  1.000000000  0.032941095 -0.015537517
## PhoneTime        0.030239860  0.032941095  1.000000000  0.021800270
## WorkHours        0.052998592 -0.015537517  0.021800270  1.000000000
## CaffeineIntake   -0.066698795 -0.014342645 -0.003289419 -0.007251107
## RelaxationTime   0.019059974  0.027963592 -0.038856265  0.036523758
## SleepTime        0.339599596  0.147583173 -0.614798454 -0.601077626
##
##          CaffeineIntake  RelaxationTime  SleepTime
## WorkoutTime      -0.066698795    0.01905997  0.3395996
## ReadingTime      -0.014342645    0.02796359  0.1475832
## PhoneTime        -0.003289419   -0.03885627 -0.6147985
## WorkHours        -0.007251107    0.03652376 -0.6010776
## CaffeineIntake    1.000000000   -0.04108626 -0.1863547
## RelaxationTime   -0.041086255    1.00000000  0.2239364
## SleepTime        -0.186354715    0.22393644  1.0000000
```

Modeli kurma, aykırı değerlerden sıyrılmış veri ile aykırı değerleri içeren modelin karşılaştırılması

```
model1 <- lm(SleepTime ~ ., data = egitimset)
model2 <- lm(SleepTime ~ ., data = egitimsettemiz)
```

```
summary(model1)
```

```
##
## Call:
## lm(formula = SleepTime ~ ., data = egitimset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3068 -0.1091 -0.0525  0.0068  9.6978
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   8.0064918  0.1014926  78.887 <0.0000000000000002 ***
## WorkoutTime    0.4360857  0.0197203  22.114 <0.0000000000000002 ***
## ReadingTime    0.2608034  0.0298484   8.738 <0.0000000000000002 ***
## PhoneTime     -0.5687980  0.0149964 -37.929 <0.0000000000000002 ***
## WorkHours     -0.3823076  0.0099867 -38.282 <0.0000000000000002 ***
## CaffeineIntake -0.0018353  0.0002026  -9.061 <0.0000000000000002 ***
## RelaxationTime 0.3564244  0.0295111  12.078 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6772 on 1561 degrees of freedom
## Multiple R-squared:  0.7027, Adjusted R-squared:  0.7015
## F-statistic: 614.9 on 6 and 1561 DF,  p-value: < 0.00000000000000022
```

```
summary(model2)
```

```
##
## Call:
## lm(formula = SleepTime ~ ., data = egitimsettemiz)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.94239 -0.06900 -0.02109  0.03742  1.88351
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   7.86436842  0.02821111  278.77 <0.0000000000000002 ***
## WorkoutTime    0.45226208  0.00550179   82.20 <0.0000000000000002 ***
## ReadingTime    0.26959562  0.00832913   32.37 <0.0000000000000002 ***
## PhoneTime     -0.56081761  0.00418470 -134.02 <0.0000000000000002 ***
## WorkHours     -0.37472624  0.00277972 -134.81 <0.0000000000000002 ***
## CaffeineIntake -0.00195116  0.00005655  -34.51 <0.0000000000000002 ***
## RelaxationTime 0.37096117  0.00824821   44.98 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1866 on 1527 degrees of freedom
```

```
## Multiple R-squared:  0.9685, Adjusted R-squared:  0.9684
## F-statistic:  7819 on 6 and 1527 DF,  p-value: < 0.00000000000000022
```

#MODEL 1 İÇİN

```
#WorkoutTime (beta1):Egzersiziz yapılan süre 1 birim arttıkça, uyku süresi ortalama 0.436 saat artıyor.
#ReadingTime (beta2):Okuma yapılan süre 1 birim arttıkça, uyku süresi 0.2608 saat artıyor.
#PhoneTime (beta3):Telefon kullanım 1 birim arttıkça, uyku süresi -0.5688 saat azalıyor.
#WorkHours (beta4):Çalışma süresi 1 birim arttıkça, uyku süresi -0.382 saat azalıyor.
#CaffeineIntake (beta5):Kafein tüketimi 1 birim arttıkça, uyku süresi -0.0018 saat azalıyor
#RelaxationTime (beta6):Dinlenme süresi 1 birim arttıkça, uyku süresi 0.3564 saat artıyor.
```

#MODEL 2 İÇİN

```
#WorkoutTime (beta1):Egzersiziz yapılan 1 birim arttıkça, uyku süresi ortalama 0.452 saat artıyor.
#İlk modele göre etkisi biraz daha yüksek.
#ReadingTime (beta2):Okuma yapılan süre 1 birim arttıkça, uyku süresi 0.2696 saat artıyor.
#PhoneTime (beta3):Telefon kullanımı 1 birim arttıkça, uyku süresi -0.5608 saat azalıyor.
#WorkHours (beta4):Çalışma süresi 1 birim arttıkça, uyku süresi -0.3747 saat azalıyor.
#CaffeineIntake (beta5):Kafein tüketimi 1 birim arttıkça, uyku süresi -0.0019 saat azalıyor.
#RelaxationTime (beta6):Dinlenme süresi 1 birim arttıkça, uyku süresi 0.371 saat artıyor.
#Telefon kullanımı ve çalışma saatleri uyku süresi üzerinde en negatif etkilerden biri.
#Egzersiziz, okuma ve dinlenme ise olumlu etkiler sağlıyor.
```

```
##Nihai model için hataların aykırı değer kontrolü
```

```
standartlasma_hata<-rstandard(model2) #hataları standartlaştırdık
summary(standartlasma_hata) #Değerlendirme için özetini aldık.
```

```
##      Min.    1st Qu.      Median        Mean     3rd Qu.       Max.
## -10.42625  -0.37089  -0.11321    0.00011    0.20087    10.11648
```

```
#SUMMARY ÜZERİNDEN YORUM;
#REFERANS OLARAK -2/+2 ARALIĞINI ALARAK,HANGİ GÖZLEMLERİN BU ARALIK DIŞINDA KALDIĞINI BELİRLEYECEĞİZ.
indeksler<-which(abs(standartlasma_hata)>2)
length(indeksler)
```

```
## [1] 49
```


#49 tane hatamız OUTLIER olma potansiyeline sahip.

*#ŞİMDİ COOK'S DISTANCE KULLANARAK AYKIRI DEĞER TESPİTİ YAPACAĞIZ VE
#BU İKİ ÖLÇÜM SONUCUNDA MODELDEN TEMİZLENMESİ GEREKENLERİ BELİRLEYECEĞİZ.*

```
distance<-cooks.distance(model2)
```

```
olcut_1<-mean(distance)*3
```

```
olcut_2<-4/length(distance)
```

#Eğer bir gözlemin Cook's distance mesafesi burada verdiğimiz ölçütlerden herhangi birini aşar ise, o gözlem OUTLIER'dir.

#TESTLER SONUCU BELİRLENEN AYKIRILARIN TEMİZLENMESİ

```
aykiri<-which(distance>olcut_1&&abs(standartlasma_hata)>2)
```

```
egitimsettemiz<-egitimsettemiz[-aykiri,]
```

```
nrow(egitimsettemiz)
```

```
## [1] 1485
```

#VERİMİZDEN HATALARIN AYKIRISINI DA TEMİZLEDİKTEN SONRA 1534'TEN 1485'E BİR DÜŞÜŞ GÖZLEMLEDİK.

#ŞİMDİ TEMİZ VERİMİZİ KULLANARAK NİHAİ MODEL OLAN MODEL2'Yİ TEKRAR OLUŞTURUYORUZ.

```
model2 <- lm(SleepTime ~ ., data = egitimsettemiz)
```

```
summary(model2)
```

```
##
```

```
## Call:
```

```
## lm(formula = SleepTime ~ ., data = egitimsettemiz)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -0.30123 -0.02806 -0.01144  0.00939  0.49125
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.93958235  0.00992939   799.60 <0.0000000000000002 ***
## WorkoutTime    0.48491897  0.00194418   249.42 <0.0000000000000002 ***
## ReadingTime    0.29090489  0.00292359    99.50 <0.0000000000000002 ***
## PhoneTime     -0.58563755  0.00148976  -393.11 <0.0000000000000002 ***
## WorkHours     -0.39007576  0.00098815  -394.75 <0.0000000000000002 ***
## CaffeineIntake -0.00197283  0.00001985   -99.41 <0.0000000000000002 ***
## RelaxationTime 0.39145936  0.00289373   135.28 <0.0000000000000002 ***
```

```
## ---
```

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

```
##
```

```
## Residual standard error: 0.0645 on 1478 degrees of freedom
```

```
## Multiple R-squared:  0.9961, Adjusted R-squared:  0.996
```

```
## F-statistic: 6.215e+04 on 6 and 1478 DF, p-value: < 0.00000000000000022
```

#MODEL 2'NİN YENİDEN YORUMLANMASI

*#WorkoutTime (beta1):Egzersiz yapılan 1 birim arttıkça, uyku süresi ortalama 0.484 saat artıyor.
#İlk modele göre etkisi biraz daha yüksek.*

#ReadingTime (beta2):Okuma yapılan süre 1 birim arttıkça, uyku süresi 0.2909 saat artıyor.

#PhoneTime (beta3):Telefon kullanımı 1 birim arttıkça, uyku süresi -0.5856 saat azalıyor.

```
#WorkHours (beta4):Çalışma süresi 1 birim arttıkça, uyku süresi -0.3900 saat azalıyor.
#CaffeineIntake (beta5):Kafein tüketimi 1 birim arttıkça, uyku süresi -0.0019 saat azalıyor.
#RelaxationTime (beta6):Dinlenme süresi 1 birim arttıkça, uyku süresi 0.391 saat artıyor.
#Telefon kullanımı ve çalışma saatleri uyku süresi üzerinde en negatif etkilerden biri.
#Egzersiz, okuma ve dinlenme ise olumlu etkiler sağlıyor.
```

Modellerin performans metrikleri üzerinden kıyaslanması

```
# AIC ve BIC tanımı:::
AIC(model1,k=8)
```

```
## [1] 3284.432
```

```
AIC(model2,k=8)
```

```
## [1] -3869.843
```

```
## ## AIC değeri, modelin uygunluğunu ve basitliğini değerlendirir.
## Daha düşük bir AIC değeri, modelin veri setine daha iyi uyduğunu gösterir.
## o yüzden -3869.843 olan AIC (model 2 için) 3284.432 olan AIC (model1)'den daha iyidir

# AIC, yalnızca modelin ne kadar iyi tahmin ettiğini değil,
# aynı zamanda gereksiz karmaşıklıktan ne kadar kaçındığını da değerlendirir.
# İyi bir model ise hem basit hem de etkili olmalıdır.
```

```
BIC(model1)
```

```
## [1] 3279.293
```

```
BIC(model2)
```

```
## [1] -3875.417
```

```
## model1 = "3279.293" Yüksek BIC değeri, modelin veriye uyum sağlayamadığını göstermekte.
#Yani olumsuz etkilemiş
## model2 = "-3875.417" Düşük BIC değeri, modelin veri setine hem daha iyi uyum sağladığını
#hem de daha sade olduğunu kanıtlıyor.

# AIC VE BIC NEDEN ÖNEMLİ HANGİSİ?
# Küçük veri setinde: AIC genelde daha iyi sonuç verir çünkü karmaşıklığı daha AZ cezalandırır.
# Büyük veri setinde: BIC genelde daha iyidir çünkü daha fazla örnek varsa karmaşık
#modelleri AĞIR cezalandırır.
```

Modelin test verisi üzerinden değerlendirilmesi

```

model1pred<-predict(model1,testset)

model2pred<-predict(model2,testset)

model1preddata<-data.frame("gercekdegerler"=testset$SleepTime,"tahminler"=model1pred)
model2preddata<-data.frame("gercekdegerler"=testset$SleepTime,"tahminler"=model2pred)

model1hata<-model1preddata$gercekdegerler-model1preddata$tahminler
model2hata<-model2preddata$gercekdegerler-model2preddata$tahminler

```

```

# Model performans metrikleri üzerinden değerlendirme
R2(model1preddata$tahminler, testset$SleepTime)

```

```
## [1] 0.8502043
```

```
R2(model2preddata$tahminler, testset$SleepTime)
```

```
## [1] 0.8497381
```

```
RMSE(model1preddata$tahminler, testset$SleepTime)
```

```
## [1] 0.4489728
```

```
RMSE(model2preddata$tahminler, testset$SleepTime)
```

```
## [1] 0.4552452
```

```
MAE(model1preddata$tahminler, testset$SleepTime)
```

```
## [1] 0.148773
```

```
MAE(model2preddata$tahminler, testset$SleepTime)
```

```
## [1] 0.1094065
```

```

# model1 ve model2 'nin performans metriklerini karşılaştırdığımızda,
#R^2 değerlerinin çok yakın olduğunu görüyoruz
#(model1: 0.8502043, model2: 0.8497381), yani her iki model de
#toplam varyansın %85'ini açıklayabiliyor.

```

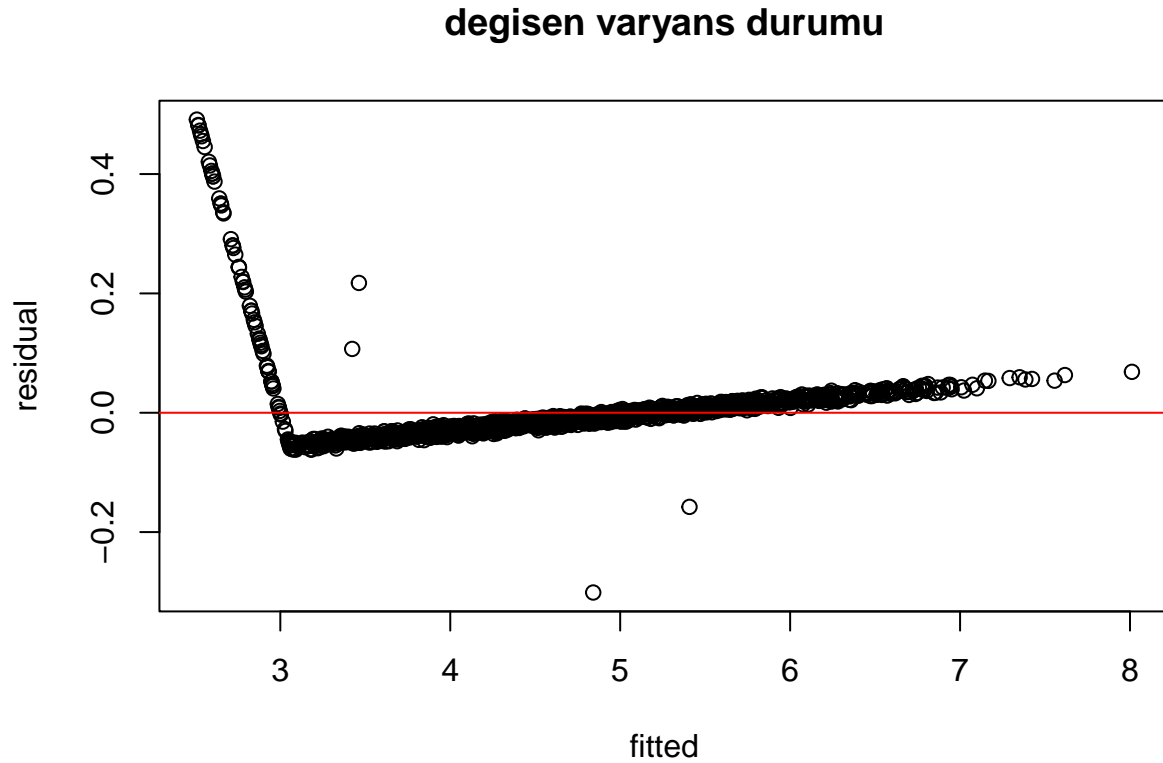
```

#Ancak mutlak hataların küçüklüğü önemsendiği durumda yani minimum hata ile
#çalışmayı amaçladığımızda model2'nin model1'den üstün olduğunu gözlemlemekteyiz;
#> MAE(model1preddata$tahminler, testset$SleepTime)
#> [1] 0.148773
#> MAE(model2preddata$tahminler, testset$SleepTime)
#> [1] 0.1094065

```

Varsayımların Değerlendirilmesi

```
#MODELLERİMİZİ ELDE ETTİK,KARŞILAŞTIRDIK, VARSAYIMLARIN DEĞERLENDİRMESİNİ YAPACAĞIZ.  
# Hatalar ile tahmin edilen değerler arasındaki ilişkiyi gösteren grafik  
plot(model2$fitted.values,model2$residuals,main="degisen varyans durumu",  
      xlab="fitted",ylab="residual")  
abline(h=0,col="red")
```

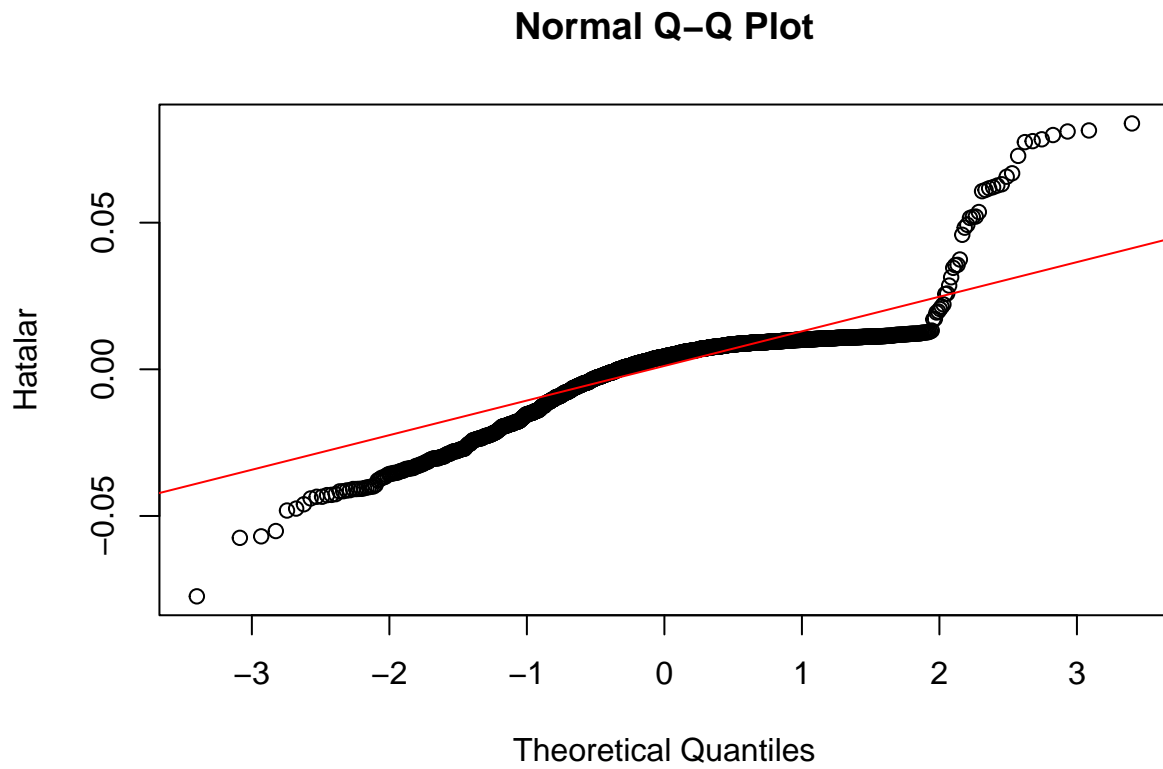


```
#DEĞİŞEN VARYANSLILIK VARSAYIMI;  
#BREUSH-PAGAN TESTİ;  
#H0:DEĞİŞEN VARYANS YOK  
#H1:DEĞİŞEN VARYANS VAR.  
  
model2kok <- lm(sqrt(SleepTime)~.,data=egitimsettemiz)  
  
#bp testi uygulanmadan önce kök dönüşümü uygulandı.  
  
#BREUSH-PAGAN TESTİ  
bptest(sqrt(SleepTime)~.,data=egitimsettemiz)  
  
##  
## studentized Breusch-Pagan test  
##  
## data: sqrt(SleepTime) ~ .
```

```
## BP = 117.45, df = 6, p-value < 0.00000000000000022
```

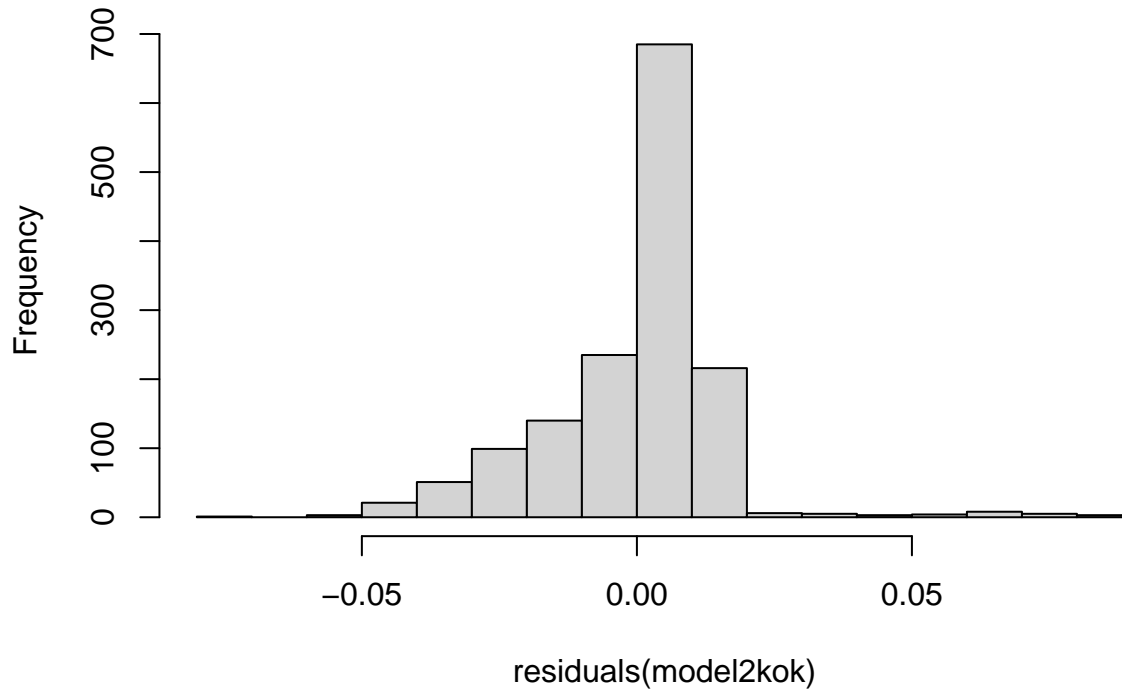
```
# P value 0.05'ten küçük o yüzden  $H_0$  'ı reddettik .  
#Bundan dolayı değişen varyanslılık durumunun kök dönüşümüne  
#rağmen devam ettiğini görüyoruz.
```

```
# Normallik Varsayımı  
#NORMAL DAĞILIM İNCELEMESİNİ ÖNCELİKLE Q-Q PLOT ÜZERİNDEN YAPALIM.  
qqnorm(residuals(model2kok), ylab = "Hatalar")  
qqline(residuals(model2kok), col = "red")
```



```
hist(residuals(model2kok))
```

Histogram of residuals(model2kok)



```
#NORMALLİK VARSAYIMI İÇİN HİPOTEZLERİMİZE KARAR VERELİM VE SHAPIRO-WILK TESTİ UYGULAYALIM.  
#ÖNCE HİPOTEZLERE KARAR VERELİM.  
#H0:NORMAL DAĞILIM VAR  
#H1:SAPMA VAR,N.D. YOK.  
shapiro.test(residuals(model2kok))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(model2kok)  
## W = 0.849, p-value < 0.000000000000000022
```

```
#P DEĞERİ 0.00000000002 ÇIKTIĞI İÇİN H0 REDDEDİLİR. VERİ NORMAL DAĞILMAMAKTADIR.
```

Weighted Least Squares - Ağırlıklandırılmış kare dönüşümü : Değişken varyans durumunu çözebilmek için alternatif bir yol deniyoruz.

```
weights <- 1 / lm(abs(residuals(model2kok)) ~ fitted(model2kok))$fitted.values^2  
wls_model <- lm(sqrt(SleepTime) ~ ., data = egitimsettemiz, weights = weights)  
summary(wls_model)
```

```
##
```

```
## Call:
## lm(formula = sqrt(SleepTime) ~ ., data = egitimsettemiz, weights = weights)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -35.062  -1.039   0.363   1.335   3.430
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  2.877963572  0.002822462 1019.66 <0.0000000000000002 ***
## WorkoutTime  0.103113924  0.000546769  188.59 <0.0000000000000002 ***
## ReadingTime  0.060611307  0.000829410   73.08 <0.0000000000000002 ***
## PhoneTime   -0.128146047  0.000444105 -288.55 <0.0000000000000002 ***
## WorkHours    -0.085462190  0.000295359 -289.35 <0.0000000000000002 ***
## CaffeineIntake -0.000403528  0.000005678  -71.06 <0.0000000000000002 ***
## RelaxationTime 0.079246085  0.000822758   96.32 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.97 on 1478 degrees of freedom
## Multiple R-squared:  0.996, Adjusted R-squared:  0.996
## F-statistic: 6.192e+04 on 6 and 1478 DF, p-value: < 0.00000000000000022
```

```
#WLS DÖNÜŞÜMÜ UYGULADIĞIMIZ MODELİN ÜZERİNDEN TEKRAR BPTEST UYGULAYALIM
#BREUSH-PAGAN TESTİ;
#H0:DEĞİŞEN VARYANS YOK
#H1:DEĞİŞEN VARYANS VAR.
bptest(sqrt(SleepTime) ~ ., data = egitimsettemiz, weights = weights)
```

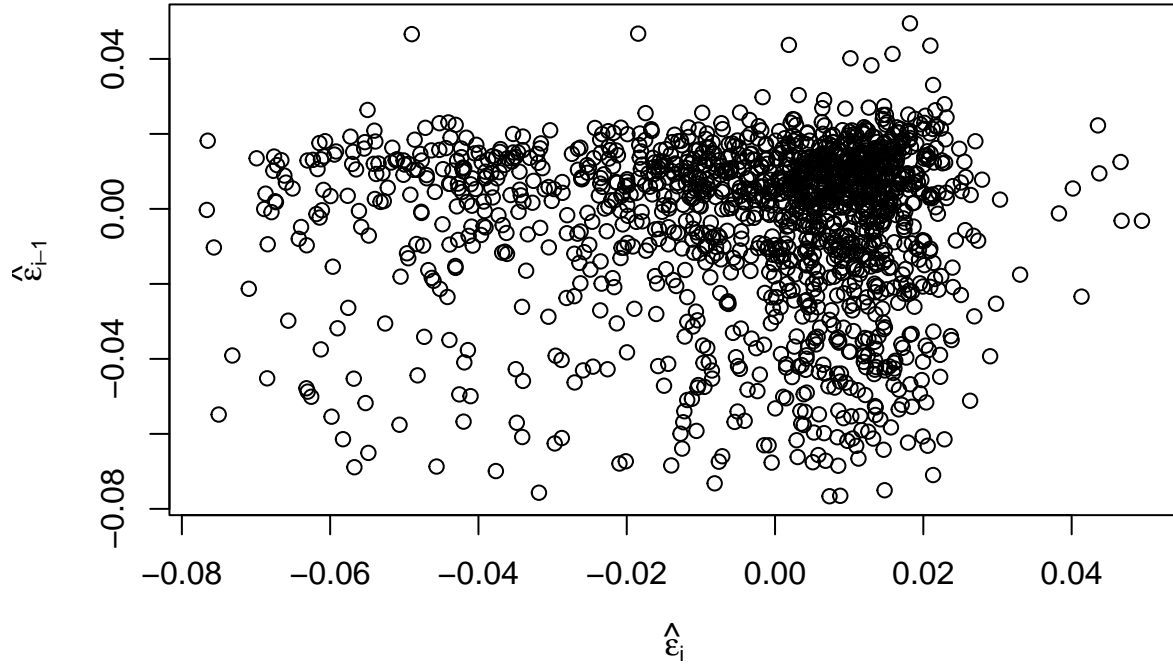
```
##
## studentized Breusch-Pagan test
##
## data: sqrt(SleepTime) ~ .
## BP = 18380304, df = 6, p-value < 0.00000000000000022
```

```
#WLS dönüşümüne rağmen modelimizde bulunan değişen varyans durumu düzeltilememiştir.
```

OTOKORELASYON testi.

```
# OTOKORELASYON VARSAYIMI İÇİN DURBIN-WATSON TESTİ KULLANILIR.
# HIPOTEZLERE KARAR VERELİM.
# H0:HATALAR İLİŞKİLİ DEĞİLLERDİR
# H1:HATALAR İLİŞKİLİDİRLER.
# TEST SONUCU YAKLAŞIK 2 İSE OTOKORELASYON YOKTUR.
# 2'DEN OLDUKÇA BÜYÜKSE NEGATİF OTOKORELASYON VARDIR.
# 2'DEN ÇOK KÜÇÜK DEĞERLER İSE POZİTİF OTOKORELASYON VARDIR.

n <- length(residuals(model2))
plot(tail(residuals(wls_model), n - 1),head(residuals(wls_model), n - 1),
     xlab = expression(hat(epsilon)[i]),ylab = expression(hat(epsilon)[i - 1]))
```



```
dwtest(sqrt(SleepTime)~.,data=egitimsettemiz)
```

```
##  
## Durbin-Watson test  
##  
## data: sqrt(SleepTime) ~ .  
## DW = 2.064, p-value = 0.8913  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# DWTEST=2.064~2 olduğundan ötürü Otokorelasyon yoktur.  
#P-VALUE=0.89 olduğundan H0 reddetmek için yeterli kanıt yoktur.  
#Yani Hatalar İlişkili Değillerdir.
```

MULTICOLINEARY TEST

```
#BAĞIMSIZ DEĞİŞKENLERİN İLİŞKİSİZ OLMASI VARSAYIMI.
```

```
# MULTICOLINEARY'NIN NEDEN OLDUĞU SORUNLAR:BETALARA İLİŞKİN VARYANSLARI ŞİŞİRİR.  
#t_hesaplar küçülür.Modelde olması gereken değişkeni  
#yanlışıyla modelden atmak zorunda kalıyoruz.
```

```
#VIF VARYANS ŞİŞİRME FAKTÖRÜ;1/1-R^2
```



```

#BUNU BULMANIN YOLU;4 TANE BAĞIMSIZ DEĞİŞKENİMİZ İÇİM VİF1 BULACAĞIZ
#X1 DEĞERİNİ BAĞIMLI OLARAK ALDIK. GERİ KALAN X2,X3,X4 BAĞIMSIZ OLDU.
#BUNLARI KULLANARAK OLUŞTURDUĞUMUZ MODELİN R^2'Sİ FORMÜLDE KULLANILACAKTIR.
#VIF2'Yİ BULMAK İSTERSEK BU SEFER X2 BAĞIMLI DİĞERLERİ BAĞIMSIZ OLACAKTIR.

#VIF<10 ISE PROBLEM YOKTUR.>10 VE BÜYÜKLÜK NE KADAR ARTARSA SORUN O KADAR ARTAR.

# DEĞİŞKENİ ATMAYIP ÇÖZMEK İSTİYORSAK ÇÖZÜMLER;
# YENİ DEĞİŞKENLER EKLEMEK
# İÇ İLİŞKİYİ ORTADAN KALDIRMAK:TEMEL BİLEŞENLER ANALİZİ YAPMAK
# MODELİ DEĞİŞTİRMEK:REGRESYON MODELİMİZİ DEĞİŞTİRECEĞİZ.
# ---VIF HESAPLATMA---#
vif(wls_model)

```

```

##      WorkoutTime      ReadingTime      PhoneTime      WorkHours CaffeineIntake
## 0.00008616564 0.00008655176 0.00009611659 0.00009685613 0.00008824745
## RelaxationTime
## 0.00008712382

```

#BÜTÜN HEPSİ 10'UN ALTINDA ÇIKTIĞI İÇİN BİR SORUN YOK.MULTICOL PROBLEMİ GÖZLENMEMİŞTİR.

##STEPWISE(AŞAMALI) REGRESYON UYGULAMASI

```

#EN DÜŞÜK AIC DEĞERİNİ ELDE EDECEK BİR MODEL KURMAK İÇİN
#HANGİ SIRAYLA İLERLEMEMİZ GEREKTİĞİ HAKKINDA BİZE YOL GÖSTERMESİ ADINA
#STEPWISE REGRESYON YÖNTEMLERİNDEN "BACKWARD" YÖNTEMİNİ KULLANACAĞIZ.
step(lm(sqrt(SleepTime) ~ ., data = egitimsettemiz, weights = weights),
     direction = "backward",
     scope=~WorkoutTime+ReadingTime+PhoneTime+WorkHours+CaffeineIntake+RelaxationTime)

```

```

## Start:  AIC=2020.44
## sqrt(SleepTime) ~ WorkoutTime + ReadingTime + PhoneTime + WorkHours +
##      CaffeineIntake + RelaxationTime
##
##              Df Sum of Sq    RSS    AIC
## <none>                5735 2020.4
## - CaffeineIntake      1    19595 25330 4224.3
## - ReadingTime         1    20721 26456 4288.9
## - RelaxationTime      1    35996 41731 4965.7
## - WorkoutTime         1   137997 143732 6802.2
## - PhoneTime           1   323059 328794 8031.0
## - WorkHours           1   324855 330590 8039.1

##
## Call:
## lm(formula = sqrt(SleepTime) ~ WorkoutTime + ReadingTime + PhoneTime +
##      WorkHours + CaffeineIntake + RelaxationTime, data = egitimsettemiz,
##      weights = weights)
##
## Coefficients:
##      (Intercept)      WorkoutTime      ReadingTime      PhoneTime      WorkHours

```

```
##      2.8779636      0.1031139      0.0606113      -0.1281460      -0.0854622
## CaffeineIntake RelaxationTime
##      -0.0004035      0.0792461
```

*#AŞAMALI REGRESYON SONUÇLARI İNCELENDİĞİNDE;
#HİÇBİR DEĞİŞKENİN MODELDEN ÇIKARILMASI UYGUN DEĞİLDİR.
#Başlangıç modelimiz optimaldir.
#Özellikle PhoneTime ve WorkHours değişkenleri model için en kritik değişkenlerdir.
#Modelden çıkarılmaları AIC değerini oldukça kötü etkilemiştir.*