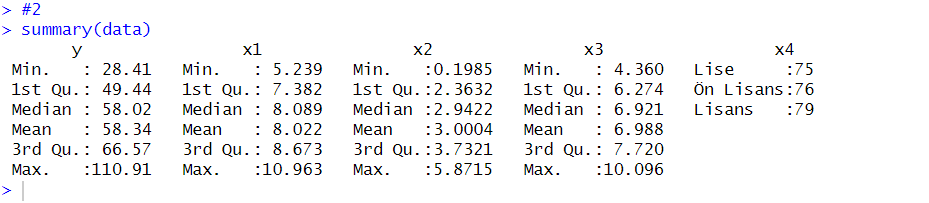
#1  
Bir eğitim kurumunda öğrencilerin genel başarı düzeyi olan **"y"** değişkeni; öğrencilerin ders çalışma süresi (**x1**), internet kullanımı (**x2**), uyku süresi (**x3**) ve kayıtlı olduğu sınıf düzeyi (**x4**, 1: Lise, 2: Ön Lisans, 3: Lisans) değişkenlerine bağlı olarak incelenmek istenmektedir. Amaç, başarıyı etkileyen faktörleri belirlemek ve bu başarı düzeyini öngörebilecek bir regresyon modeli kurmaktır.

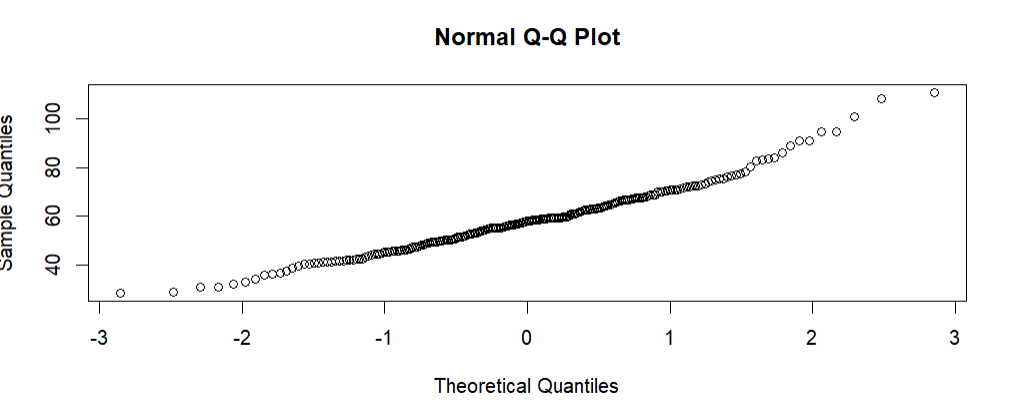
#2

summary(data)

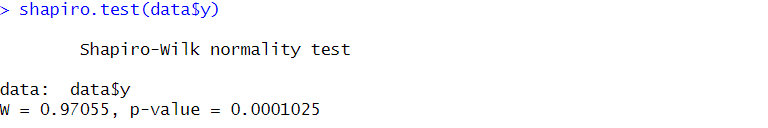


#3

qqnorm(data$y)



shapiro.test(data$y)



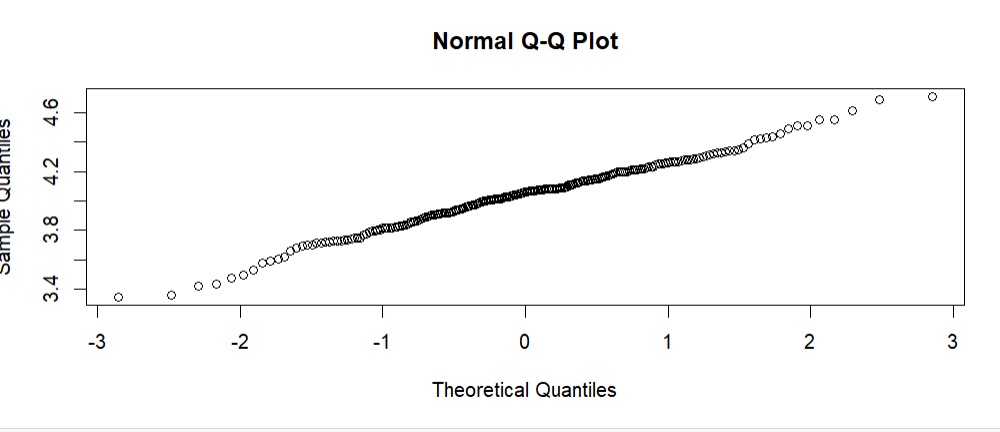
H0=Verilerin dağılışı ile normal dağılımı arasında fark yoktur.

Yorum=p=0.0001025<0.05 olduğundan H0 kabul edilemiyor. Verim normal dağılmıyor.

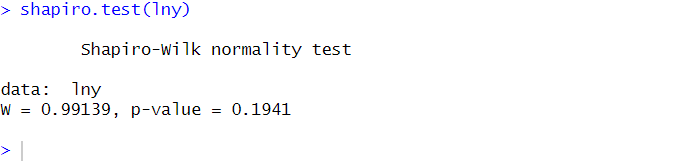
Normallik varsayımı sağlanamadığı için logaritmik dönüşüm yapıyoruz.

lny<-log(data$y)

qqnorm(lny)



shapiro.test(lny)

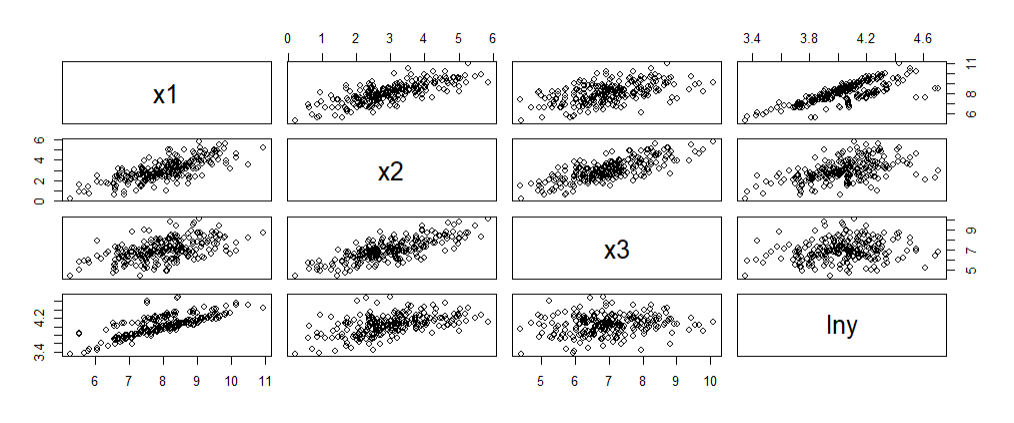


H0=Verilerin dağılışı ile normal dağılımı arasında fark yoktur.

Yorum=p=0.1941>0.05 olduğundan H0 kabul. Verim normal dağılıyor.

df <- data.frame(x1 = data$x1, x2 = data$x2, x3 = data$x3, lny = lny)

pairs(df)



Yorum=x1 ve lny arasında pozitif bir ilişki bulunmakta yani çalışma süresi arttıkça başarıda artıyor.X2 ve lny arasında x1 kadar olmasada pozitif yönlü güçlü bir ilişki bulunmakta.X3 ve lny arasında belirgin bir ilişki bulunmamakta çok zayıf gözükmekte.X4 kategoriktir ve kategorikler ve sürekli değişkenler arasında korelasyon gözlemlenemez.

#4

model<- lm(lny ~ x1 + x2 + x3 + factor(x4),data =data)

summary(model)

|  |
| --- |
|  |
| H0: β=0 (Model anlamlı değil)  H0:β≠0(Model anlamlıdır)  P=α<0.05 H0 red Model anlamlıdır.    residuals <- resid(model)  fitted\_values <- fitted(model)  plot(fitted\_values, residuals,  xlab = "Tahmin Edilen Değerler",  ylab = "Artıklar",  main = "Artıklar vs Tahmin Edilen Değerler")  abline(h = 0, col = "red", lty = 2)    qqnorm(residuals)  qqline(residuals, col = "red")    hist(residuals, main = "Artıkların Histogramı", xlab = "Artıklar")    std\_resid <- rstandard(model)  plot(std\_resid, main = "Standartlaştırılmış Artıklar")  abline(h = c(-2, 2), col = "red", lty = 2)  metin, ekran görüntüsü, çizgi, yazı tipi içeren bir resim  Yapay zeka tarafından oluşturulan içerik yanlış olabilir.  # Uç değer (leverage)  leverage <- hatvalues(model)  uc\_deger<-which(leverage > 2 \* (length(coef(model))/nrow(df)))  print(uc\_deger)    # Aykırı değer (standartlaştırılmış artık)  aykiri\_deger<-which(abs(std\_resid) > 2)  print(aykiri\_deger)    # Etkin gözlem (Cook's distance)  cooks\_d <- cooks.distance(model)  etkin<-which(cooks\_d > 4 / nrow(df))  print(etkin)    aykiri\_deger <- unique(c(uc\_deger, aykiri\_deger, etkin))  library(olsrr)  ols\_plot\_resid\_fit(model)    ols\_plot\_resid\_qq(model)    ols\_plot\_resid\_hist(model)    ols\_plot\_cooksd\_bar(model)    veri\_temiz <- data[-aykiri\_deger, ]  model\_temiz <- lm(y ~ x1 + x2 + x3 + x4, data = veri\_temiz)  ols\_plot\_resid\_fit(model\_temiz) # Artık vs tahmin    ols\_plot\_resid\_qq(model\_temiz)    ols\_plot\_resid\_hist(model\_temiz)    ols\_plot\_cooksd\_bar(model\_temiz)    ols\_plot\_resid\_stud(model\_temiz)    inf<-ls.diag(model)  inf                influence.measures(model)        Yorum=Artık değerleri veriden çıkardığımda çok fazla bozulmalar olduğu için çıkarmadan devam edilir.  #5  model <- lm(lny ~ x1 + x2 + x3 + x4, data = data)  summary(model)    ŷ=3.07444+0.18874X1+0.03615X2-0.07543X3-0.172770X4.2-0.21473X4.30.1041  (0.08711)+(0.01070)+(0.01305)+(0.01005)+(0.01698)+(0.01699)  #6  summary(model)$coefficients  ,  H0:β=0  H1:β≠0  P<0.05 H0red.Katksı anlamlı.Bu şekilde tüm değişkenlere baktığım zaman hepsinin 0.05den küçük olduğunu görüyorum. Yani tüm değişkenler modele anlamlı bir katkısı var.  #7  summary(model)$r.squared  summary(model)$adj.r.squared    Yorum= ,0.8091 buda demek oluyor ki öğrencilerin başarı düzeyinin %81’i, ders çalışma süresi,internet kullanımı ve eğitim düzeyi değişkenleri ile açıklanmakta.Açıklanamayan kısım %19’dur.  #8  confint(model, level = 0.99)    Yorum=Bütün değişkenlerim 0’ı içermediği için modele anlamlı bir katkı sağlamakta.  #9  # Grafiksel yöntem: Tahminler vs artıklar  plot(model$fitted.values, model$residuals,  xlab = "Tahmin Edilen Değerler", ylab = "Artıklar",  main = "Artıklar vs Tahminler Grafiği")  abline(h = 0, col = "red")    Yorum=X ekseninde modelin tahmin ettiği y değerleri olup Y ekseninde bu tahminlere karşılık gelen artık değerlerim yer almakta. Kırmızı çizgi ise artıkların ortalamasının 0 olduğu noktayı vermekte.Artıklarım çoğunlukla referans çizgime yakın olsada tahminler arttıkça biraz dalgalanmalar gözükmekte o yüzden Breusch-Pagan testi yapıp daha doğru bir yorum yapılır.  # Breusch-Pagan testi  install.packages("lmtest")  library(lmtest)  bptest(model)    Yorum=Bu grafiğe ek olarak pagan testi sonucu p=0.7731 >0.05 olduğundan H0 kabul değişen varyanslılık yoktur.  #10  install.packages("lmtest")  library(lmtest)  # Durbin-Watson testi  dwtest(model)    Yorum=Durbin-Watson testi sonucunda p-değeri 0.05’ten küçük bulunmuştur.Bu sebeple modelde anlamlı bir otokorelasyon olduğu yani modelin artıklarında bağımlılık gözlenmekte.  #11  install.packages("car")  library(car)  # VIF değerleri  vif(model)  library(olsrr)  ols\_eigen\_cindex(model)    #12  # 10. gözlem için uyum kestirimi  model <- lm(lny~ x1 + x2 + x3 + x4, data = data)  exp(predict(model, newdata = data[10, ], interval = "confidence"))  Yorum= Modelim,10.gözlem değerini 46.12816 olarak tahmin etmiş.0.99 güven düzeyinde 44.53084 ün altına düşmeyip yine aynı güven düzeyinde47.78277nin üstüne çıkmayacaktır.  #13  # Ön kestirim  new\_obs <- data.frame(  x1 = 9,  x2 = 3.5,  x3 = 7,  x4 = factor("Lisans", levels = levels(data$x4)) # Bu daha güvenli  )  model <- lm(lny ~ x1 + x2 + x3 + x4, data = data)  exp(predict(model, newdata = new\_obs, interval = "prediction", level = 0.99))    #14  exp(predict(model, newdata = data[10, ], interval = "confidence", level = 0.95))    new\_obs <- data.frame(  x1 = 9,  x2 = 3.5,  x3 = 7,  x4 = factor("Lisans", levels = levels(data$x4)) # güvenli yol  )  exp(predict(model, newdata = new\_obs, interval = "prediction", level = 0.95))    #15  #AIC ile Adım Adım (Stepwise) model seçimi  install.packages("MASS")  library(MASS)  step\_model <- stepAIC(model, direction="both",trace=FALSE)  summary(step\_model)    # geri seçim  backward<-step(model,direction="backward")    #ileri seçim  library(stats)  lm.null <- lm(lny ~ 1,data = data)  lm.full <- lm(lny ~ x1 + x2 + x3 + x4, data = data)  summary(lm.full)    #16  install.packages("glmnet")  library(glmnet)  # x4 faktörü dummy'lere çevrilir, model matrisi oluşturulur  X <- model.matrix(y ~ x1 + x2 + x3 + x4, data = data)[, -1] # sabit terim hariç  y <- data$y  ridge\_model <- glmnet(X, y, alpha = 0)  plot(ridge\_model, xvar = "lambda", label = TRUE,  main = "Ridge Regresyon İz Grafiği")     |  | | --- | |  | |