Lack of fit of regression model

Chetana Jamadar

2022-12-19

1. Consider the following data,(DATA given by code) Fit the linear regression model of Y on X1 and X2. Obtain an estimate of pure error term and hence perform the lack of fit test.

#------------Que.2---------------------  
y=c(26,24,175,160,163,55,62,100,26,30,70,71)  
x1=c(1,1,1.5,1.5,1.5,1,2,0.5,1,1,1,1)  
x2=c(1,1,4,4,4,2,5,3,2,2,3,3)  
M1=lm(y~x1+x2) #Reduced  
M1

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Coefficients:  
## (Intercept) x1 x2   
## 3.941 -27.910 38.395

M2=lm(y~0+as.factor(x1)+as.factor(x2)) #Full  
M2

##   
## Call:  
## lm(formula = y ~ 0 + as.factor(x1) + as.factor(x2))  
##   
## Coefficients:  
## as.factor(x1)0.5 as.factor(x1)1 as.factor(x1)1.5 as.factor(x1)2   
## 54.5 25.0 166.0 62.0   
## as.factor(x2)2 as.factor(x2)3 as.factor(x2)4 as.factor(x2)5   
## 12.0 45.5 NA NA

anova(M1,M2)

## Analysis of Variance Table  
##   
## Model 1: y ~ x1 + x2  
## Model 2: y ~ 0 + as.factor(x1) + as.factor(x2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 9 16184.6   
## 2 6 622.5 3 15562 49.999 0.0001227 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#-------------Que.4------------------  
y=c(16.68,11.50,12.03,14.88,13.75,18.11,8.00,17.83,79.24,21.50,40.33,21.00,13.5,19.75,24.00,  
 29.00,15.35,19.00,9.50,35.10,17.90,52.32,18.75,19.83,10.75)  
x1=c(7.00,3.00,3.00,4.00,6.00,7.00,2.00,7.00,30.00,5.00,16.00,10.00,4.00,6.00,9.00,10.00,  
 6.00,7.00,3.00,17.00,10.00,26.00,9.00,8.00,4.00)  
x2=c(560,226,340,80,150,330,110,210,1460,605,688,215,255,462,448,776,200,132,36  
 ,770,140,810,450,635,150)  
M1=lm(y~x1+x2)  
M1

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Coefficients:  
## (Intercept) x1 x2   
## 2.33366 1.61618 0.01439

M2=lm(y~0+as.factor(x1)+as.factor(x2)) #Full  
M2

##   
## Call:  
## lm(formula = y ~ 0 + as.factor(x1) + as.factor(x2))  
##   
## Coefficients:  
## as.factor(x1)2 as.factor(x1)3 as.factor(x1)4 as.factor(x1)5   
## 8.00 9.50 16.75 21.50   
## as.factor(x1)6 as.factor(x1)7 as.factor(x1)8 as.factor(x1)9   
## 19.75 16.68 19.83 18.75   
## as.factor(x1)10 as.factor(x1)16 as.factor(x1)17 as.factor(x1)26   
## 29.00 40.33 35.10 52.32   
## as.factor(x1)30 as.factor(x2)80 as.factor(x2)110 as.factor(x2)132   
## 79.24 -1.87 NA 2.32   
## as.factor(x2)140 as.factor(x2)150 as.factor(x2)200 as.factor(x2)210   
## -11.10 -6.00 -4.40 1.15   
## as.factor(x2)215 as.factor(x2)226 as.factor(x2)255 as.factor(x2)330   
## -8.00 2.00 -3.25 1.43   
## as.factor(x2)340 as.factor(x2)448 as.factor(x2)450 as.factor(x2)462   
## 2.53 5.25 NA NA   
## as.factor(x2)560 as.factor(x2)605 as.factor(x2)635 as.factor(x2)688   
## NA NA NA NA   
## as.factor(x2)770 as.factor(x2)776 as.factor(x2)810 as.factor(x2)1460   
## NA NA NA NA

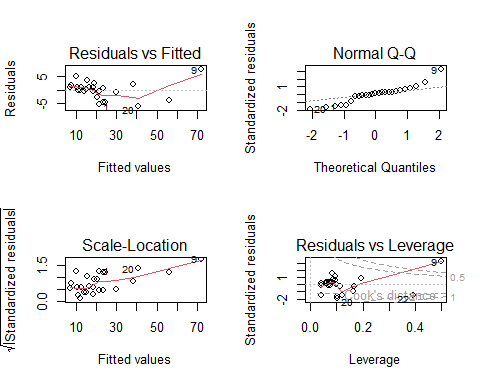
anova(M1,M2)

## Analysis of Variance Table  
##   
## Model 1: y ~ x1 + x2  
## Model 2: y ~ 0 + as.factor(x1) + as.factor(x2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 22 233.54   
## 2 0 0.00 22 233.54 NaN NaN

fit=lm(y~x1+x2)  
fit

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Coefficients:  
## (Intercept) x1 x2   
## 2.33366 1.61618 0.01439

par(mfrow=c(2,2))  
plot(M1)



anova(M1)

## Analysis of Variance Table  
##   
## Response: y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## x1 1 5382.4 5382.4 507.034 < 2.2e-16 \*\*\*  
## x2 1 168.6 168.6 15.882 0.0006253 \*\*\*  
## Residuals 22 233.5 10.6   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

hist(fit$residuals)  
plot(x1,fit$residuals)  
plot(x2,fit$residuals)  
#R-statistic  
summary(fit)

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.7883 -0.6613 0.4414 1.1614 7.4129   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.333663 1.096678 2.128 0.044790 \*   
## x1 1.616184 0.170542 9.477 3.18e-09 \*\*\*  
## x2 0.014389 0.003611 3.985 0.000625 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.258 on 22 degrees of freedom  
## Multiple R-squared: 0.9596, Adjusted R-squared: 0.956   
## F-statistic: 261.5 on 2 and 22 DF, p-value: 4.645e-16

#To find Standardise residuals  
e=fit$residuals  
e

## 1 2 3 4 5 6   
## -5.02476844 1.06588009 -0.04446194 4.93048546 -0.43910919 -0.28530646   
## 7 8 9 10 11 12   
## 0.85118359 1.16136936 7.41294145 2.38009521 2.23779214 -0.58912609   
## 13 14 15 16 17 18   
## 1.03241656 1.07153368 0.67442857 -0.66133554 0.44144256 3.45370865   
## 19 20 21 22 23 24   
## 1.79978347 -5.78828655 -2.60995370 -3.68949703 -4.60434936 -4.57012437   
## 25   
## -0.20674210

d=e/3.258  
d

## 1 2 3 4 5 6   
## -1.54228620 0.32715779 -0.01364700 1.51334729 -0.13477876 -0.08757104   
## 7 8 9 10 11 12   
## 0.26125954 0.35646696 2.27530431 0.73053874 0.68686069 -0.18082446   
## 13 14 15 16 17 18   
## 0.31688660 0.32889309 0.20700693 -0.20298820 0.13549495 1.06007018   
## 19 20 21 22 23 24   
## 0.55241973 -1.77663798 -0.80109076 -1.13244230 -1.41324413 -1.40273922   
## 25   
## -0.06345675

d>2

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE FALSE   
## 14 15 16 17 18 19 20 21 22 23 24 25   
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE

h=hatvalues(fit)  
r=rstandard(fit) #This command is for studantised residuals  
r

## 1 2 3 4 5 6   
## -1.62709851 0.33945481 -0.01437309 1.58243944 -0.14013894 -0.08950705   
## 7 8 9 10 11 12   
## 0.27264246 0.36839862 3.21274953 0.81465593 0.71846932 -0.19206876   
## 13 14 15 16 17 18   
## 0.32701985 0.34252603 0.21138680 -0.22222561 0.13970959 1.11512385   
## 19 20 21 22 23 24   
## 0.58117478 -1.87441475 -0.87682859 -1.45147120 -1.44325709 -1.49559885   
## 25   
## -0.06568264

#to find PRESS residuals  
e\_=e/(1-h)  
rstudent(fit) #To get r student

## 1 2 3 4 5 6   
## -1.69494024 0.33252218 -0.01404270 1.64234829 -0.13697808 -0.08746507   
## 7 8 9 10 11 12   
## 0.26682513 0.36104391 4.30821938 0.80820950 0.71033352 -0.18781032   
## 13 14 15 16 17 18   
## 0.32028054 0.33554673 0.20673674 -0.21736039 0.13655804 1.12164572   
## 19 20 21 22 23 24   
## 0.57222228 -1.99777777 -0.87204194 -1.49130133 -1.48197640 -1.54168736   
## 25   
## -0.06417879

#To delete exceed observations  
y=y[-9]  
x1=x1[-9]  
x2=x2[-9]  
y=y[-20]  
x1=x1[-20]  
x2=x2[-20]  
length(y)

## [1] 23

length(x1)

## [1] 23

length(x2)

## [1] 23

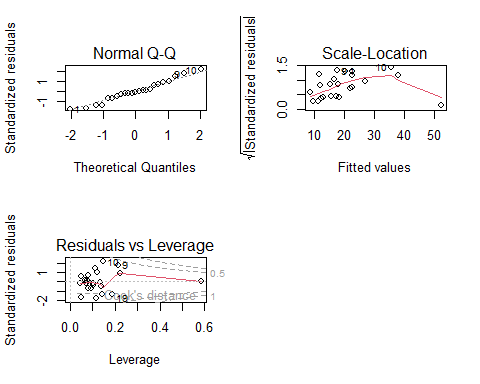
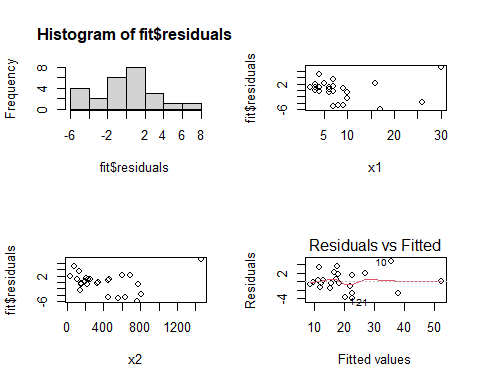
fit\_=lm(y~x1+x2)  
fit\_

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Coefficients:  
## (Intercept) x1 x2   
## 4.669145 1.554694 0.008889

summary(fit\_)

##   
## Call:  
## lm(formula = y ~ x1 + x2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.9112 -1.2992 -0.1532 1.5014 4.6704   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.669145 0.948771 4.921 8.24e-05 \*\*\*  
## x1 1.554694 0.134072 11.596 2.48e-10 \*\*\*  
## x2 0.008889 0.002988 2.975 0.00748 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.381 on 20 degrees of freedom  
## Multiple R-squared: 0.953, Adjusted R-squared: 0.9483   
## F-statistic: 202.7 on 2 and 20 DF, p-value: 5.266e-14

plot(fit\_)



1. The yield of a chemical process is related to the concentration of the reactant and the operating temperature. An experiment has been conducted with the following results.
2. Fit the linear regression model of Y on X1 and X2.
3. Obtain an estimate of pure error term and hence perform the lack of fit test.

y=c(81,89,83,81,79,88,82,90)  
x1=c(1,1,2,2,1,1,2,2)  
x2=c(150,180,150,180,150,180,150,180)  
fit=lm(y~x1+x2)  
Fit1=lm(y~0+as.factor(x1)+as.factor(x2))  
anova(fit,Fit1) #lack of fit test

## Analysis of Variance Table  
##   
## Model 1: y ~ x1 + x2  
## Model 2: y ~ 0 + as.factor(x1) + as.factor(x2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 5 58.625   
## 2 5 58.625 0 7.1054e-14