Mid-Exam

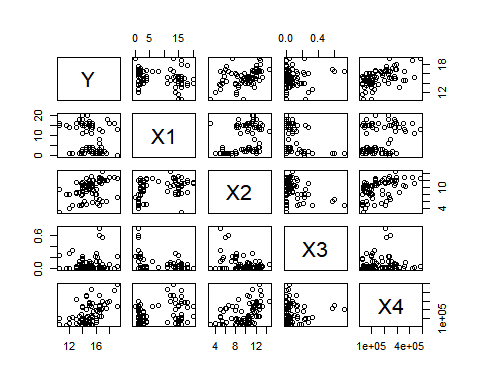
Sumanth Donthula

2022-10-19

Question 1)

From the scattered plots we can see that the rental rates(Y) is modarately correlated with operating\_expenses(X2) and square\_footage(X4).Y is weakly related with age(X1) and vacancy rates(X3).

Data=read.table("Commercial\_Property.txt", header = TRUE, sep = "");  
Y=Data$Y;  
X1=Data$X1;  
X2=Data$X2;  
X3=Data$X3;  
X4=Data$X4;  
  
  
par(mfrow=c(3,2))  
pairs(Data)

 Question 2)

The model is

Y=12.220-0.142*X1+0.238*X2+0.619*X3+0.000007*X4 or Y=1.220e+01-1.420e-01*X1+2.820e-01*X2+6.193e-01*X3+7.924e-06*X4

Model1=lm(Y~X1+X2+X3+X4)  
Model1

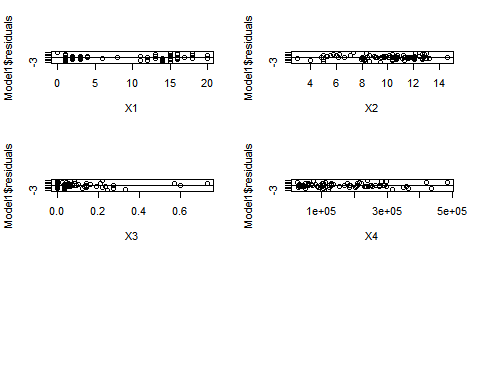
##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3 + X4)  
##   
## Coefficients:  
## (Intercept) X1 X2 X3 X4   
## 1.220e+01 -1.420e-01 2.820e-01 6.193e-01 7.924e-06

Question 3)

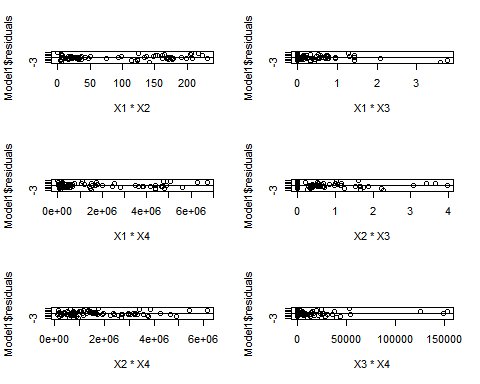
In the first plot i.e, residuals against individual predictors, the residuals appears to form a systematic patterns and i.i.d. with normal distribution.

In the second plot i.e, residuals against two factor interaction the systematic pattern for residuals look like i.i.d. normally distributed in X2*X4 and X1*X2 interaction terms.

par(mfrow=c(3,2))  
  
plot(X1,Model1$residuals)  
abline(0,0)  
plot(X2,Model1$residuals)  
abline(0,0)  
plot(X3,Model1$residuals)  
abline(0,0)  
plot(X4,Model1$residuals)  
abline(0,0)  
  
  
par(mfrow=c(3,2))



plot(X1\*X2,Model1$residuals)  
abline(0,0)  
plot(X1\*X3,Model1$residuals)  
abline(0,0)  
plot(X1\*X4,Model1$residuals)  
abline(0,0)  
plot(X2\*X3,Model1$residuals)  
abline(0,0)  
plot(X2\*X4,Model1$residuals)  
abline(0,0)  
plot(X3\*X4,Model1$residuals)  
abline(0,0)

 Question 4)

The F ratio is greater than F statistic for all coefficients, so we reject null hypothesis and conclude that none of the coeficients are 0.

anova(Model1)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## X1 1 14.819 14.819 11.4649 0.001125 \*\*   
## X2 1 72.802 72.802 56.3262 9.699e-11 \*\*\*  
## X3 1 8.381 8.381 6.4846 0.012904 \*   
## X4 1 42.325 42.325 32.7464 1.976e-07 \*\*\*  
## Residuals 76 98.231 1.293   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Fs=qf(0.95,1,76)  
Fs

## [1] 3.96676

Question 5) R Square defines the variance of dependent variable that can be explained by independent variables.

R-squared: 0.5847,

Adjusted R-Squared Adjusted R-Square is similar to R-Squared but it will consider degrees of freedom of the data points also into account because the R-Squared varies a lot if new dependent variables are added.

Adjusted R-squared: 0.5629 from summary of the model.

summary(Model1)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X3 + X4)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.1872 -0.5911 -0.0910 0.5579 2.9441   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.220e+01 5.780e-01 21.110 < 2e-16 \*\*\*  
## X1 -1.420e-01 2.134e-02 -6.655 3.89e-09 \*\*\*  
## X2 2.820e-01 6.317e-02 4.464 2.75e-05 \*\*\*  
## X3 6.193e-01 1.087e+00 0.570 0.57   
## X4 7.924e-06 1.385e-06 5.722 1.98e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.137 on 76 degrees of freedom  
## Multiple R-squared: 0.5847, Adjusted R-squared: 0.5629   
## F-statistic: 26.76 on 4 and 76 DF, p-value: 7.272e-14

Question 6)

The 95 % confidence and prediction intervals for the data points are as follows:

Data1=data.frame(X1=4,X2=10,X3=0.1,X4=80000)  
Data2=data.frame(X1=6,X2=11.5,X3=0,X4=120000)  
Data3=data.frame(X1=12,X2=12.5,X3=0.32,X4=340000)  
  
  
 writeLines("Confidence Intervals")

## Confidence Intervals

predict(Model1, newdata = Data1, interval = "confidence", level=0.95)

## fit lwr upr  
## 1 15.1485 14.76829 15.5287

predict(Model1, newdata = Data2, interval = "confidence", level=0.95)

## fit lwr upr  
## 1 15.54249 15.15366 15.93132

predict(Model1, newdata = Data3, interval = "confidence", level=0.95)

## fit lwr upr  
## 1 16.91384 16.18358 17.6441

writeLines('-----------------------------')

## -----------------------------

writeLines("Prediction Intervals")

## Prediction Intervals

predict(Model1, newdata = Data1, interval = "prediction", level=0.95)

## fit lwr upr  
## 1 15.1485 12.85249 17.4445

predict(Model1, newdata = Data2, interval = "prediction", level=0.95)

## fit lwr upr  
## 1 15.54249 13.24504 17.83994

predict(Model1, newdata = Data3, interval = "prediction", level=0.95)

## fit lwr upr  
## 1 16.91384 14.53469 19.29299

Question 7)

Model : Y=12.37-0.144*X1+0.267*X2+0.000008\*X4

partial F test F=((SSRF − SSRR)/(dfF − dfr))/MSEf follows F1,n-p

from anova tables:

Model1 Model2 SSRFull=98.231 SSRReduced=98.650 MSEf=1.293 dfF=76 dfr=77

Conclusion: Since F Ratio is less than F statistic we don’t reject null hypothesis i.e, Beta3=0

Model2=lm(Y~X1+X2+X4)  
Model2

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X4)  
##   
## Coefficients:  
## (Intercept) X1 X2 X4   
## 1.237e+01 -1.442e-01 2.672e-01 8.178e-06

anova(Model1)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## X1 1 14.819 14.819 11.4649 0.001125 \*\*   
## X2 1 72.802 72.802 56.3262 9.699e-11 \*\*\*  
## X3 1 8.381 8.381 6.4846 0.012904 \*   
## X4 1 42.325 42.325 32.7464 1.976e-07 \*\*\*  
## Residuals 76 98.231 1.293   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Model2)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## X1 1 14.819 14.819 11.566 0.001067 \*\*   
## X2 1 72.802 72.802 56.825 7.841e-11 \*\*\*  
## X4 1 50.287 50.287 39.251 1.973e-08 \*\*\*  
## Residuals 77 98.650 1.281   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

SSRF=98.231  
SSRR=98.650  
dfF=4  
dfr=3  
MSEf=1.293  
  
F=((SSRR - SSRF)/(dfF - dfr))/MSEf  
F

## [1] 0.3240526

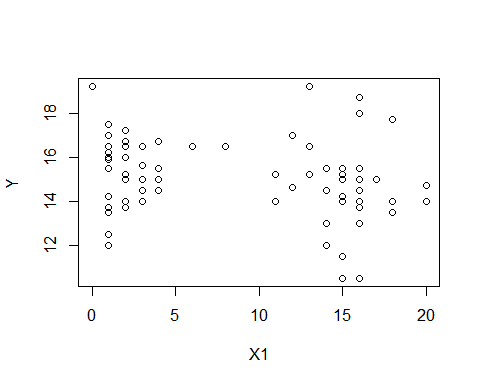
qf(p=.95, dfF-dfr, nrow(Data)-5)

## [1] 3.96676

Question 8)

from the plot we can observe that there is a curvy pattern as the value of Y is increasing wrt X1 till value 10 and it started decreasing after 10

plot(X1,Y)

 Question 9)

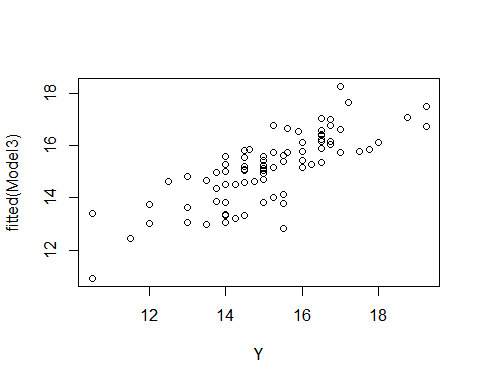
The estimated regression function is Y=12.49-0.4043*X1+0.314*X2+0.00000846*X4+0.0145*X1\*\*2

Model3 is a goodfit.

XSq=X1^2  
Model3=lm(Y~X1+X2+X4+XSq)  
summary(Model3)

##   
## Call:  
## lm(formula = Y ~ X1 + X2 + X4 + XSq)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.89596 -0.62547 -0.08907 0.62793 2.68309   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.249e+01 4.805e-01 26.000 < 2e-16 \*\*\*  
## X1 -4.043e-01 1.089e-01 -3.712 0.00039 \*\*\*  
## X2 3.140e-01 5.880e-02 5.340 9.33e-07 \*\*\*  
## X4 8.046e-06 1.267e-06 6.351 1.42e-08 \*\*\*  
## XSq 1.415e-02 5.821e-03 2.431 0.01743 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.097 on 76 degrees of freedom  
## Multiple R-squared: 0.6131, Adjusted R-squared: 0.5927   
## F-statistic: 30.1 on 4 and 76 DF, p-value: 5.203e-15

plot(Y,fitted(Model3))

 Question 10)

partial F test F=((SSRF − SSRR)/(dfF − dfr))/MSEf follows F1,n-p

from anova tables:

Model2 Model3 SSRFull=91.535 SSRReduced=98.650 MSEf=1.293 dfF=4 dfr=3

Conclusion: Since F Ratio is less than F statistic we dont reject null hypothesis i.e, Beta3=0

anova(Model2)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## X1 1 14.819 14.819 11.566 0.001067 \*\*   
## X2 1 72.802 72.802 56.825 7.841e-11 \*\*\*  
## X4 1 50.287 50.287 39.251 1.973e-08 \*\*\*  
## Residuals 77 98.650 1.281   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(Model3)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## X1 1 14.819 14.819 12.3036 0.0007627 \*\*\*  
## X2 1 72.802 72.802 60.4463 2.968e-11 \*\*\*  
## X4 1 50.287 50.287 41.7522 8.907e-09 \*\*\*  
## XSq 1 7.115 7.115 5.9078 0.0174321 \*   
## Residuals 76 91.535 1.204   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

SSRR=98.650  
SSRF=91.535  
dfF=4  
dfr=3  
MSEf=1.204  
  
F=((SSRR - SSRF)/(dfF - dfr))/MSEf  
F

## [1] 5.909468

qf(p=.95, dfF-dfr, nrow(Data)-5)

## [1] 3.96676