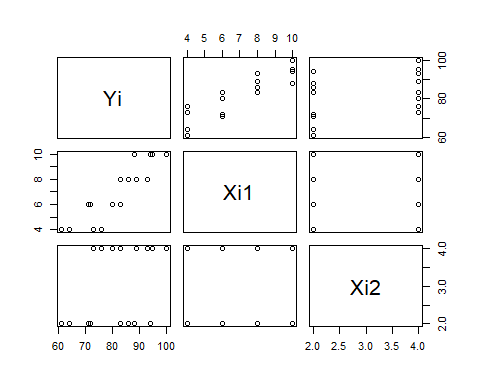
Chad Huntebrinker’s HW8

Chad Huntebrinker

2024-11-06

Problem 6.5 and 6.6

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Brand\_preference\_data.xlsx")  
  
#Problem 6.5a  
pairs(Yi~Xi1+Xi2,data=excel\_data)



cor(excel\_data)

## Yi Xi1 Xi2  
## Yi 1.0000000 0.8923929 0.3945807  
## Xi1 0.8923929 1.0000000 0.0000000  
## Xi2 0.3945807 0.0000000 1.0000000

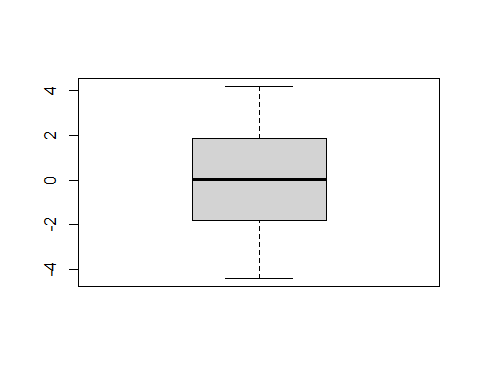
#We see that there is a strong linear correlation between Yi and Xi1, but not so much  
#anywhere else.  
  
#Problem 6.5b  
model\_1 <- lm(Yi~Xi1+Xi2, data = excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
sum\_of\_model\_1

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.400 -1.762 0.025 1.587 4.200   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.6500 2.9961 12.566 1.20e-08 \*\*\*  
## Xi1 4.4250 0.3011 14.695 1.78e-09 \*\*\*  
## Xi2 4.3750 0.6733 6.498 2.01e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.693 on 13 degrees of freedom  
## Multiple R-squared: 0.9521, Adjusted R-squared: 0.9447   
## F-statistic: 129.1 on 2 and 13 DF, p-value: 2.658e-09

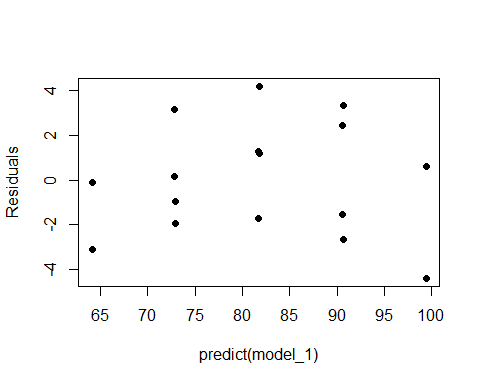
#Yi1 = 37.65 + 4.425 \* Xi1 + 4.375 \* Xi2  
  
#Problem 6.5c  
sum\_of\_model\_1$residuals

## 1 2 3 4 5 6 7 8 9 10 11 12 13   
## -0.10 0.15 -3.10 3.15 -0.95 -1.70 -1.95 1.30 1.20 -1.55 4.20 2.45 -2.65   
## 14 15 16   
## -4.40 3.35 0.60

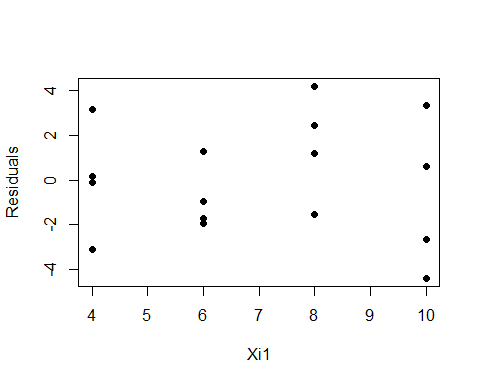
boxplot(sum\_of\_model\_1$residuals)



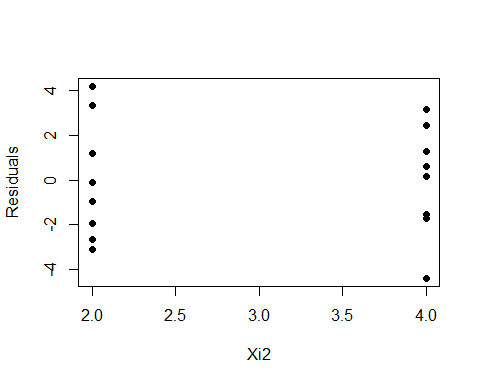
#We see the residuals are fairly well distrubted.  
  
#Problem 6.5d  
# Plot residuals versus each X to examine the relationship  
plot(model\_1$residuals~predict(model\_1),pch=16,ylab="Residuals", data = excel\_data)



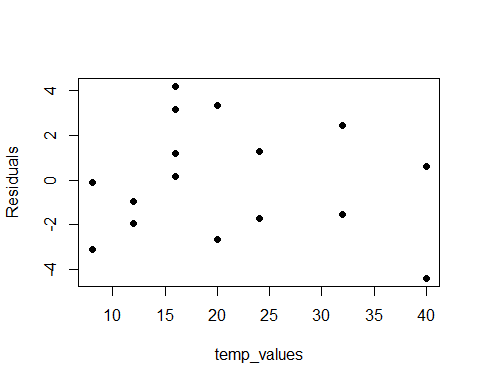
plot(model\_1$residuals~Xi1,pch=16,ylab="Residuals", data = excel\_data)



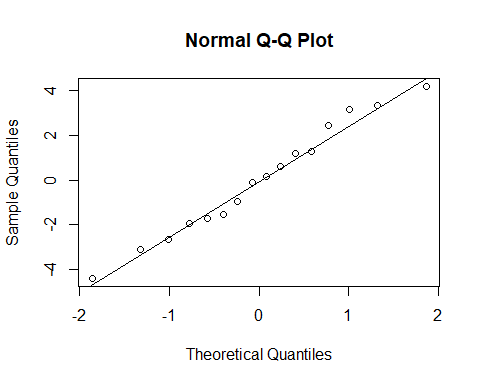
plot(model\_1$residuals~Xi2,pch=16,ylab="Residuals", data = excel\_data)



temp\_values <- excel\_data$Xi1 \* excel\_data$Xi2  
plot(model\_1$residuals~temp\_values,pch=16,ylab="Residuals", data = excel\_data)



qqnorm(residuals(model\_1))  
qqline(residuals(model\_1))



#What we find is:  
#Yhat seems to have more of a curve.  
#Xi1 seems to be distributed well, but only at values 4, 6, 8, and 10  
#Xi2 is the same but only at values 2 and 4  
#Xi1 \* Xi2 seems to be distributed well  
#The normal plot seems reasonable  
  
#Problem 6.6a  
#H0: B1 = B2 = 0  
#Ha: not all Bk = 0 (k = 1 and 2)  
sum\_of\_model\_1

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.400 -1.762 0.025 1.587 4.200   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.6500 2.9961 12.566 1.20e-08 \*\*\*  
## Xi1 4.4250 0.3011 14.695 1.78e-09 \*\*\*  
## Xi2 4.3750 0.6733 6.498 2.01e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.693 on 13 degrees of freedom  
## Multiple R-squared: 0.9521, Adjusted R-squared: 0.9447   
## F-statistic: 129.1 on 2 and 13 DF, p-value: 2.658e-09

#F score = 129.1 on 2 and 13 DF  
129.1 > qf(1 - 0.01, 2, 13)

## [1] TRUE

#Ha is true. Meaning that both B1 and B2 != 0  
  
#Problem 6.6b  
#P-value = 2.658e-09 (which is basically 0+)  
  
#Problem 6.6c  
confint(model\_1, level=0.995)

## 0.25 % 99.75 %  
## (Intercept) 27.545738 47.754262  
## Xi1 3.409483 5.440517  
## Xi2 2.104236 6.645764

Problem 6.18-6.21

#Chad Huntebrinker  
  
library(readxl)  
library(lawstat)  
  
excel\_data <- read\_excel("Commercial\_properties\_data.xlsx")  
  
#Problem 6.18a  
stem(excel\_data$Xi1)

##   
## The decimal point is at the |  
##   
## 0 | 0000000000000000  
## 2 | 00000000000000000000000  
## 4 | 00000  
## 6 | 0  
## 8 | 0  
## 10 | 00  
## 12 | 00000  
## 14 | 0000000000000  
## 16 | 0000000000  
## 18 | 000  
## 20 | 00

stem(excel\_data$Xi2)

##   
## The decimal point is at the |  
##   
## 2 | 0  
## 4 | 080003358  
## 6 | 012613  
## 8 | 00001223456001555689  
## 10 | 013344566677778123344666668  
## 12 | 00011115777889002  
## 14 | 6

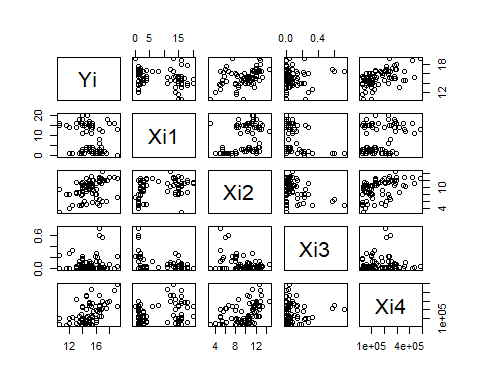
stem(excel\_data$Xi3)

##   
## The decimal point is 1 digit(s) to the left of the |  
##   
## 0 | 0000000000000000000000000000002333333333334444445555556678889  
## 1 | 023444469  
## 2 | 1223477  
## 3 | 3  
## 4 |   
## 5 | 7  
## 6 | 0  
## 7 | 3

stem(excel\_data$Xi4)

##   
## The decimal point is 5 digit(s) to the right of the |  
##   
## 0 | 333333444444  
## 0 | 555666667778899  
## 1 | 000001111222333334  
## 1 | 578889  
## 2 | 011122334444  
## 2 | 555788899  
## 3 | 002  
## 3 | 567  
## 4 | 23  
## 4 | 8

#These plots provide the range of values that these varaiables are (which we can use  
#to see if there are any outliers or see if it's distributed well).  
  
#Problem 6.18b  
pairs(Yi~Xi1+Xi2+Xi3+Xi4,data=excel\_data)



cor(excel\_data)

## Yi Xi1 Xi2 Xi3 Xi4  
## Yi 1.00000000 -0.2502846 0.4137872 0.06652647 0.53526237  
## Xi1 -0.25028456 1.0000000 0.3888264 -0.25266347 0.28858350  
## Xi2 0.41378716 0.3888264 1.0000000 -0.37976174 0.44069713  
## Xi3 0.06652647 -0.2526635 -0.3797617 1.00000000 0.08061073  
## Xi4 0.53526237 0.2885835 0.4406971 0.08061073 1.00000000

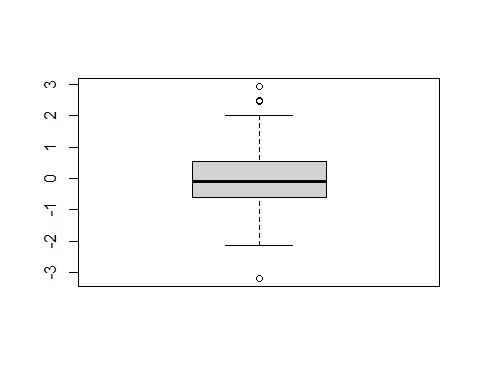
#There seems to be a bit of a linear relationship between Yi and Xi4. Besides that,  
#there really isn't anything  
  
#Problem 6.18c  
model\_2 <- lm(Yi~Xi1+Xi2+Xi3+Xi4, data = excel\_data)  
sum\_of\_model\_2 <- summary(model\_2)  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2 + Xi3 + Xi4, data = excel\_data)  
##   
## 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 \*\*\*  
## Xi1 -1.420e-01 2.134e-02 -6.655 3.89e-09 \*\*\*  
## Xi2 2.820e-01 6.317e-02 4.464 2.75e-05 \*\*\*  
## Xi3 6.193e-01 1.087e+00 0.570 0.57   
## Xi4 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

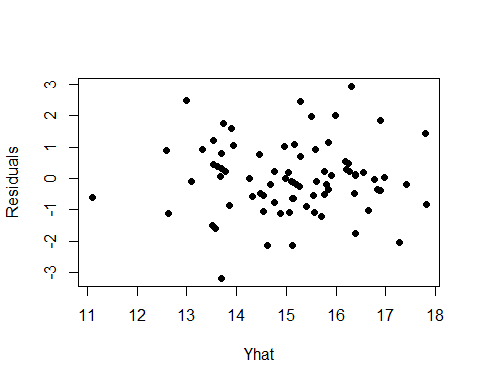
#Yi = 1.220e+01 - 1.420e-01Xi1 + 2.820e-01Xi2 + 6.193e-01Xi3 + 7.924e-06Xi4  
  
#Problem 6.18d  
sum\_of\_model\_2$residuals

## 1 2 3 4 5 6   
## -1.035672440 -1.513806414 -0.591053402 -0.133568082 0.313283765 -3.187185224   
## 7 8 9 10 11 12   
## -0.538356749 0.236302386 1.989220372 0.105829603 0.023124830 -0.337070751   
## 13 14 15 16 17 18   
## 0.717869468 -0.392411015 -0.201019573 -0.814937024 0.101690072 -1.759131637   
## 19 20 21 22 23 24   
## -1.210114916 -0.634341765 -0.366004170 0.288596123 -0.093200248 0.233884284   
## 25 26 27 28 29 30   
## -0.853339941 -2.123934469 0.466014057 -0.573974675 -1.068826727 -0.197717691   
## 31 32 33 34 35 36   
## -1.121737177 -0.173906919 -1.030125636 -0.090953654 0.215053952 0.784804746   
## 37 38 39 40 41 42   
## 1.083920373 -2.132451269 -0.185470952 -1.120385453 -0.012771680 2.500938643   
## 43 44 45 46 47 48   
## -1.582833452 0.929599530 0.394236721 0.117200255 0.815339787 1.605896564   
## 49 50 51 52 53 54   
## 0.557941960 0.494737472 0.207611404 -0.032045798 1.155796537 0.234272601   
## 55 56 57 58 59 60   
## -1.073489739 1.059646672 -0.261711555 1.031651273 -0.345957207 0.203372872   
## 61 62 63 64 65 66   
## 0.917961126 2.944144932 2.459696482 1.859088749 1.451807658 -0.483857748   
## 67 68 69 70 71 72   
## -0.756250356 2.011402309 0.078550427 0.009892809 1.766898426 -0.463930876   
## 73 74 75 76 77 78   
## -0.510410866 -0.106354746 1.209427169 -0.261085606 -0.627547725 0.910085787   
## 79 80 81   
## -0.550846871 -2.030180944 -0.906819056

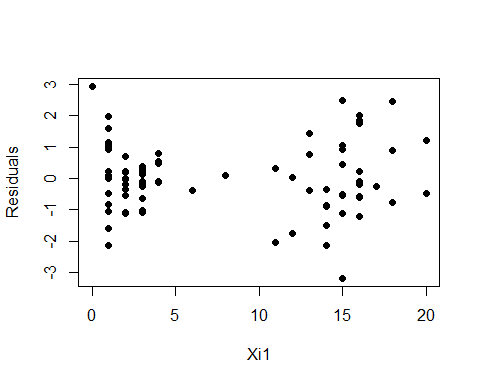
boxplot(sum\_of\_model\_2$residuals)



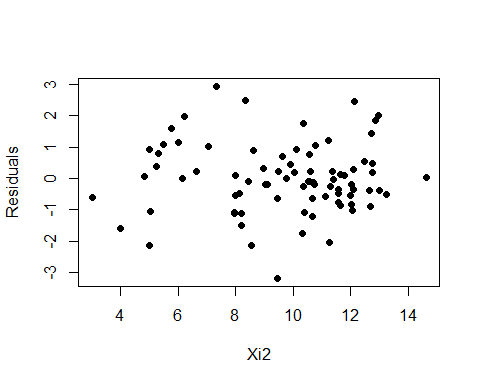
#While it seems to be fairly symmetrical, we see that there are some outlier  
#residuals in the plot  
  
#Problem 6.18e  
plot(model\_2$residuals~predict(model\_2),pch=16,xlab = "Yhat",ylab="Residuals", data = excel\_data)



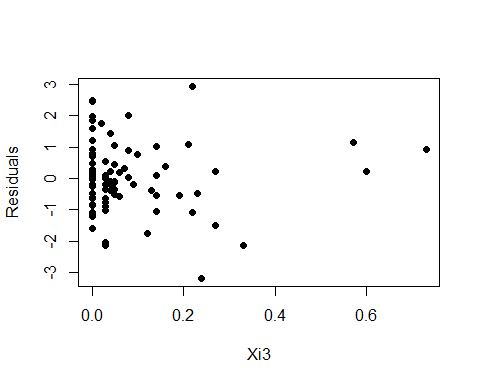
plot(model\_2$residuals~Xi1,pch=16,ylab="Residuals", data = excel\_data)



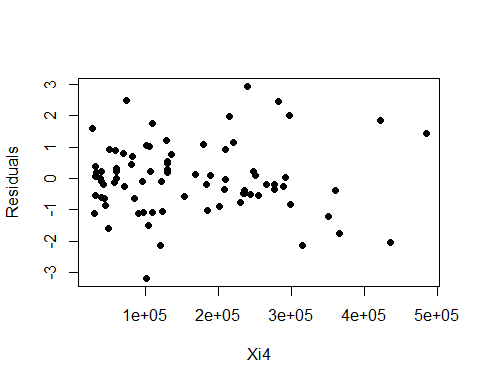
plot(model\_2$residuals~Xi2,pch=16,ylab="Residuals", data = excel\_data)



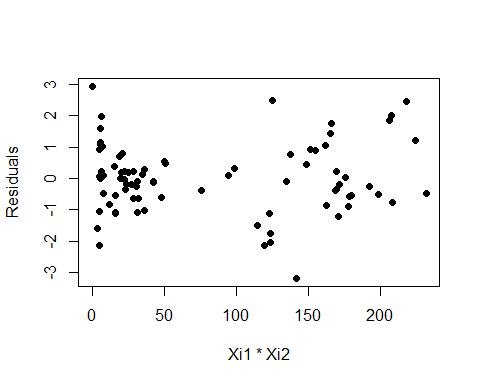
plot(model\_2$residuals~Xi3,pch=16,ylab="Residuals", data = excel\_data)



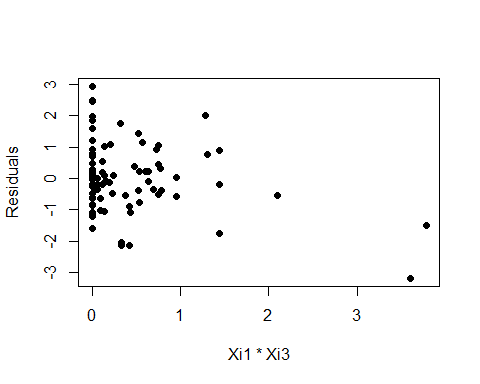
plot(model\_2$residuals~Xi4,pch=16,ylab="Residuals", data = excel\_data)



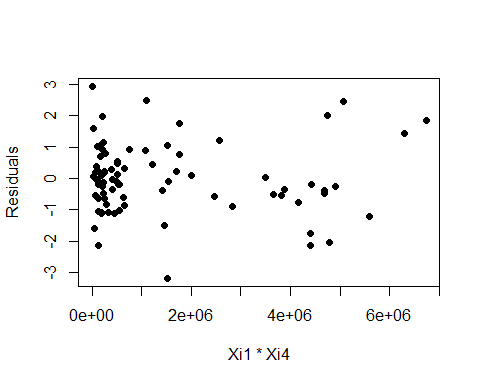
temp\_values <- excel\_data$Xi1 \* excel\_data$Xi2  
plot(model\_2$residuals~temp\_values, xlab = "Xi1 \* Xi2", pch=16,ylab="Residuals", data = excel\_data)



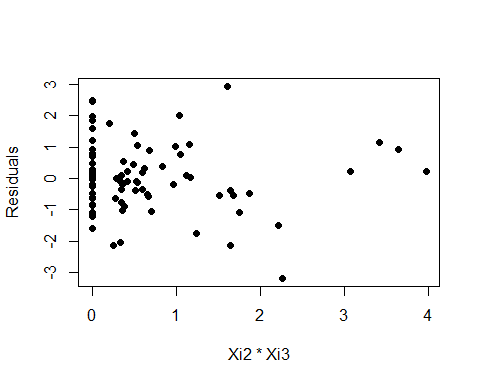
temp\_values <- excel\_data$Xi1 \* excel\_data$Xi3  
plot(model\_2$residuals~temp\_values, xlab = "Xi1 \* Xi3", pch=16,ylab="Residuals", data = excel\_data)



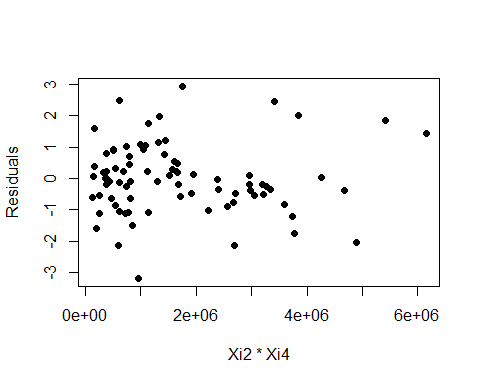
temp\_values <- excel\_data$Xi1 \* excel\_data$Xi4  
plot(model\_2$residuals~temp\_values, xlab = "Xi1 \* Xi4", pch=16,ylab="Residuals", data = excel\_data)



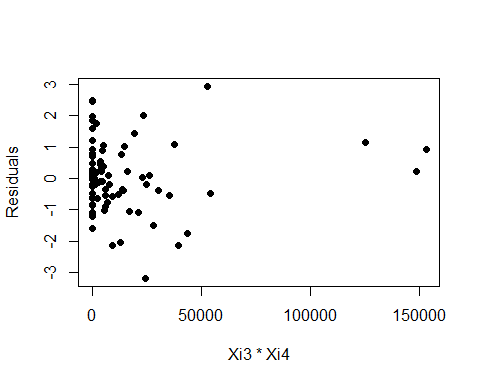
temp\_values <- excel\_data$Xi2 \* excel\_data$Xi3  
plot(model\_2$residuals~temp\_values, xlab = "Xi2 \* Xi3", pch=16,ylab="Residuals", data = excel\_data)



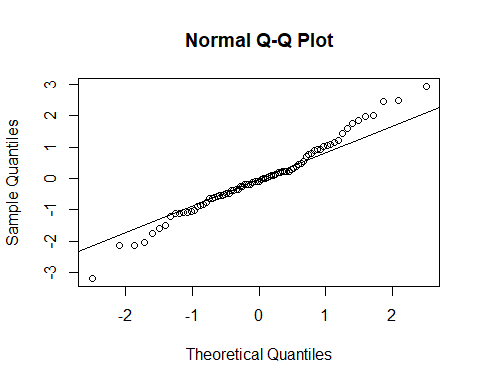
temp\_values <- excel\_data$Xi2 \* excel\_data$Xi4  
plot(model\_2$residuals~temp\_values, xlab = "Xi2 \* Xi4", pch=16,ylab="Residuals", data = excel\_data)



temp\_values <- excel\_data$Xi3 \* excel\_data$Xi4  
plot(model\_2$residuals~temp\_values, xlab = "Xi3 \* Xi4", pch=16,ylab="Residuals", data = excel\_data)



qqnorm(residuals(model\_2))  
qqline(residuals(model\_2))



#Yhat: relatively normal  
#Xi1:relatively normal, but there's a lack of values between 5 and 10  
#Xi2: relatively normal  
#Xi3: relatively normal, but there's some outliers at 0.6  
#Xi4: relatively normal  
#Xi1 \* Xi2: relatively normal, but lack of values between 50 and 100  
#Xi1 \* Xi3: relatively normal, but outliers beyond 1  
#Xi1 \* Xi4: relatively normal  
#Xi2 \* Xi3: relatively normal, but there's some outliers past 2  
#Xi2 \* Xi4: relatively normal  
#Xi3 \* Xi4: relatively normal, but there's some outliers past 50,000  
#Normal plot: relatively normal  
  
#Problem 6.18f  
#Yes, you could.  
  
#Problem 6.18g  
#H0: t\*bf <= qt(0.975,79), the error variance is constant  
#Ha: t\*bf > qt(0.975,79), the error variance is not constant  
mean(model\_2$fitted.values)

## [1] 15.13889

sum(model\_2$fitted.values < mean(model\_2$fitted.values))

## [1] 40

group\_1\_values <- model\_2$fitted.values[model\_2$fitted.values < mean(model\_2$fitted.values)]  
group\_2\_values <- model\_2$fitted.values[model\_2$fitted.values > mean(model\_2$fitted.values)]  
  
fitted\_values <- c(group\_1\_values, group\_2\_values)  
groups <- factor(rep(c("Group1", "Group2"), times = c(40, 41)))  
  
levene.test(fitted\_values, groups)

##   
## Modified robust Brown-Forsythe Levene-type test based on the absolute  
## deviations from the median  
##   
## data: fitted\_values  
## Test Statistic = 1.6323, p-value = 0.2051

#Test if the H0 is true  
1.6323 <= qt(0.975,79)

## [1] TRUE

#It is true, so the error variance is constant  
  
#Problem 6.19a  
#H0: B1 = B2 = 0  
#Ha: not all Bk = 0 (k = 1 and 2)  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2 + Xi3 + Xi4, data = excel\_data)  
##   
## 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 \*\*\*  
## Xi1 -1.420e-01 2.134e-02 -6.655 3.89e-09 \*\*\*  
## Xi2 2.820e-01 6.317e-02 4.464 2.75e-05 \*\*\*  
## Xi3 6.193e-01 1.087e+00 0.570 0.57   
## Xi4 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

#F score = 26.76 on 4 and 76 DF  
26.76 > qf(1 - 0.05, 4, 76)

## [1] TRUE

#Ha is true. Meaning that B1 and B2 and B3 and B4 all do not equal 0  
#p-value: 7.272e-14  
  
#Problem 6.19b  
confint(model\_2, level=0.9875)

## 0.625 % 99.375 %  
## (Intercept) 1.072186e+01 1.367931e+01  
## Xi1 -1.966396e-01 -8.742769e-02  
## Xi2 1.203875e-01 4.436456e-01  
## Xi3 -2.161312e+00 3.399999e+00  
## Xi4 4.381297e-06 1.146731e-05

#Problem 6.19c  
sum\_of\_model\_2$r.squared

## [1] 0.5847496

#This shows there seems to be a decent amount of linear association between  
#Xi1, Xi2, Xi3, and Xi4 combined and Yi  
  
#Problem 6.20  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2 + Xi3 + Xi4, data = excel\_data)  
##   
## 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 \*\*\*  
## Xi1 -1.420e-01 2.134e-02 -6.655 3.89e-09 \*\*\*  
## Xi2 2.820e-01 6.317e-02 4.464 2.75e-05 \*\*\*  
## Xi3 6.193e-01 1.087e+00 0.570 0.57   
## Xi4 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

excel\_data2 <- read\_excel("Mean\_rental\_rates\_data.xlsx")  
  
value1 <- predict(model\_2, data.frame(Xi1 = excel\_data2$Xi1[1], Xi2 = excel\_data2$Xi2[1],  
 Xi3 = excel\_data2$Xi3[1], Xi4 = excel\_data2$Xi4[1]),  
 interval = "confidence", se.fit = TRUE)  
  
value2 <- predict(model\_2, data.frame(Xi1 = excel\_data2$Xi1[2], Xi2 = excel\_data2$Xi2[2],  
 Xi3 = excel\_data2$Xi3[2], Xi4 = excel\_data2$Xi4[2]),  
 interval = "confidence", se.fit = TRUE)  
  
value3 <- predict(model\_2, data.frame(Xi1 = excel\_data2$Xi1[4], Xi2 = excel\_data2$Xi2[3],  
 Xi3 = excel\_data2$Xi3[4], Xi4 = excel\_data2$Xi4[3]),  
 interval = "confidence", se.fit = TRUE)  
  
value4 <- predict(model\_2, data.frame(Xi1 = excel\_data2$Xi1[4], Xi2 = excel\_data2$Xi2[4],  
 Xi3 = excel\_data2$Xi3[4], Xi4 = excel\_data2$Xi4[4]),   
 interval = "confidence", se.fit = TRUE)  
value1$fit

## fit lwr upr  
## 1 15.79813 15.24428 16.35198

value2$fit

## fit lwr upr  
## 1 16.02754 15.55765 16.49742

value3$fit

## fit lwr upr  
## 1 16.11666 15.66597 16.56736

value4$fit

## fit lwr upr  
## 1 15.84339 15.32729 16.35948

#Problem 6.21  
excel\_data3 <- read\_excel("No\_rental\_information\_data.xlsx")  
  
value1 <- predict(model\_2, data.frame(Xi1 = excel\_data3$Xi1[1], Xi2 = excel\_data3$Xi2[1],  
 Xi3 = excel\_data3$Xi3[1], Xi4 = excel\_data3$Xi4[1]),  
 interval="prediction", se.fit = TRUE)  
  
value2 <- predict(model\_2, data.frame(Xi1 = excel\_data3$Xi1[2], Xi2 = excel\_data3$Xi2[2],  
 Xi3 = excel\_data3$Xi3[2], Xi4 = excel\_data3$Xi4[2]),  
 interval="prediction", se.fit = TRUE)  
  
value3 <- predict(model\_2, data.frame(Xi1 = excel\_data3$Xi1[3], Xi2 = excel\_data3$Xi2[3],  
 Xi3 = excel\_data3$Xi3[3], Xi4 = excel\_data3$Xi4[3]),  
 interval="prediction", se.fit = TRUE)  
value1$fit

## fit lwr upr  
## 1 15.1485 12.85249 17.4445

value2$fit

## fit lwr upr  
## 1 15.54249 13.24504 17.83994

value3$fit

## fit lwr upr  
## 1 16.91384 14.53469 19.29299

Problem 7.3

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Brand\_preference\_data.xlsx")  
  
model\_1 <- lm(Yi~Xi1+Xi2, data = excel\_data)  
sum\_of\_model\_1 <- summary(model\_1)  
  
reduced\_model\_1 <- lm(Yi~Xi1, data = excel\_data)  
  
#Problem 7.3a  
anova\_table <- anova(reduced\_model\_1, model\_1)  
anova\_table

## Analysis of Variance Table  
##   
## Model 1: Yi ~ Xi1  
## Model 2: Yi ~ Xi1 + Xi2  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 14 400.55   
## 2 13 94.30 1 306.25 42.219 2.011e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Problem 7.3b  
#H0: p-value >= 0.01, X2 can be dropped  
#H1: p-value < 0.01, X2 should not be dropped  
anova\_table$F[2]

## [1] 42.21898

anova\_table$`Pr(>F)`[2]

## [1] 2.011047e-05

anova\_table$`Pr(>F)`[2] < 0.01

## [1] TRUE

#H1 is true, don't drop X2

Problem 7.7 and 7.8

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Commercial\_properties\_data.xlsx")  
  
model\_2 <- lm(Yi~Xi1+Xi2+Xi3+Xi4, data = excel\_data)  
sum\_of\_model\_2 <- summary(model\_2)  
sum\_of\_model\_2

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi2 + Xi3 + Xi4, data = excel\_data)  
##   
## 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 \*\*\*  
## Xi1 -1.420e-01 2.134e-02 -6.655 3.89e-09 \*\*\*  
## Xi2 2.820e-01 6.317e-02 4.464 2.75e-05 \*\*\*  
## Xi3 6.193e-01 1.087e+00 0.570 0.57   
## Xi4 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

#Problem 7.7a  
model\_Xi4 <- lm(Yi ~ Xi4, data = excel\_data)  
model\_Xi1\_Xi4 <- lm(Yi ~ Xi1 + Xi4, data = excel\_data)  
model\_Xi2\_Xi1\_Xi4 <- lm(Yi ~ Xi2 + Xi1 + Xi4, data = excel\_data)  
  
anova\_table\_Xi4 <- anova(model\_Xi4, model\_Xi1\_Xi4)  
anova\_table\_Xi2\_Xi1\_Xi4 <- anova(model\_Xi1\_Xi4, model\_Xi2\_Xi1\_Xi4)  
anova\_table\_Xi3\_Xi2\_Xi1\_Xi4 <- anova(model\_Xi2\_Xi1\_Xi4, model\_2)  
  
anova\_table\_Xi4

## Analysis of Variance Table  
##   
## Model 1: Yi ~ Xi4  
## Model 2: Yi ~ Xi1 + Xi4  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 79 168.78   
## 2 78 126.51 1 42.275 26.065 2.275e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova\_table\_Xi2\_Xi1\_Xi4

## Analysis of Variance Table  
##   
## Model 1: Yi ~ Xi1 + Xi4  
## Model 2: Yi ~ Xi2 + Xi1 + Xi4  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 78 126.51   
## 2 77 98.65 1 27.858 21.744 1.287e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova\_table\_Xi3\_Xi2\_Xi1\_Xi4

## Analysis of Variance Table  
##   
## Model 1: Yi ~ Xi2 + Xi1 + Xi4  
## Model 2: Yi ~ Xi1 + Xi2 + Xi3 + Xi4  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 77 98.650   
## 2 76 98.231 1 0.41975 0.3248 0.5704

#7.7b  
#H0: p-value >= 0.01, X3 can be dropped  
#H1: p-value < 0.01, X3 should not be dropped  
anova\_table\_Xi3\_Xi2\_Xi1\_Xi4$F[2]

## [1] 0.3247534

anova\_table\_Xi3\_Xi2\_Xi1\_Xi4$`Pr(>F)`[2]

## [1] 0.5704457

anova\_table\_Xi3\_Xi2\_Xi1\_Xi4$`Pr(>F)`[2] < 0.01

## [1] FALSE

#H1 is false, so can drop X3  
  
#Problem 7.8  
anova\_table\_Xi1\_Xi4 <- anova(model\_Xi1\_Xi4, model\_2)  
anova\_table\_Xi1\_Xi4

## Analysis of Variance Table  
##   
## Model 1: Yi ~ Xi1 + Xi4  
## Model 2: Yi ~ Xi1 + Xi2 + Xi3 + Xi4  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 78 126.508   
## 2 76 98.231 2 28.277 10.939 6.682e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#H0: p-value >= 0.01, X2 and X3 can be dropped  
#H1: p-value < 0.01, X2 and X3 should not be dropped  
anova\_table\_Xi1\_Xi4$F[2]

## [1] 10.9389

anova\_table\_Xi1\_Xi4$`Pr(>F)`[2]

## [1] 6.682136e-05

anova\_table\_Xi1\_Xi4$`Pr(>F)`[2] < 0.01

## [1] TRUE

#H1 is true, so you shouldn't drop X2 and X3

Problem 7.27

#Chad Huntebrinker  
  
library(readxl)  
  
excel\_data <- read\_excel("Commercial\_properties\_data.xlsx")  
  
#Problem 7.27a  
model\_3 <- lm(Yi~Xi1+Xi4, data = excel\_data)  
sum\_of\_model\_3 <- summary(model\_3)  
sum\_of\_model\_3

##   
## Call:  
## lm(formula = Yi ~ Xi1 + Xi4, data = excel\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.2032 -0.4593 0.0641 0.7730 2.5083   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.436e+01 2.771e-01 51.831 < 2e-16 \*\*\*  
## Xi1 -1.145e-01 2.242e-02 -5.105 2.27e-06 \*\*\*  
## Xi4 1.045e-05 1.363e-06 7.663 4.23e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.274 on 78 degrees of freedom  
## Multiple R-squared: 0.4652, Adjusted R-squared: 0.4515   
## F-statistic: 33.93 on 2 and 78 DF, p-value: 2.506e-11

#Yi = -1.145e-01 \* Xi1 + 1.045e-05 \* Xi4  
  
#Problem 7.27b  
#The coefficients changed a little bit,   
#Xi1 increased positively a little and Xi4 got a little bigger.  
  
#Problem 7.27c  
model4 <- lm(Yi~Xi3, data = excel\_data)  
model5 <- lm(Yi~Xi3 + Xi4, data = excel\_data)  
model6 <- lm(Yi~Xi1, data = excel\_data)  
model7 <- lm(Yi~Xi3 + Xi1, data = excel\_data)  
model8 <- lm(Yi~Xi4, data = excel\_data)  
  
ssr\_X4 <- sum((excel\_data$Yi - mean(excel\_data$Yi))^2) - sum(residuals(model8)^2)  
anova\_table <- anova(model4, model5)  
anova\_table$`Sum of Sq`[2]

## [1] 66.85829

ssr\_X1 <- sum((excel\_data$Yi - mean(excel\_data$Yi))^2) - sum(residuals(model6)^2)  
anova\_table <- anova(model4, model7)  
anova\_table$`Sum of Sq`[2]

## [1] 13.7743

#They don't equal, but they are extremely similar.  
  
  
#Problem 7.27d  
# Yi Xi1 Xi2 Xi3 Xi4  
#Yi 1.00000000 -0.2502846 0.4137872 0.06652647 0.53526237  
#Xi1 -0.25028456 1.0000000 0.3888264 -0.25266347 0.28858350  
#Xi2 0.41378716 0.3888264 1.0000000 -0.37976174 0.44069713  
#Xi3 0.06652647 -0.2526635 -0.3797617 1.00000000 0.08061073  
#Xi4 0.53526237 0.2885835 0.4406971 0.08061073 1.00000000  
  
#It means that X3 doesn't provide any help to X1 and X4 when explaining the variability in  
#the model. This can be seen by the fact that X1 and X4 have a much more varied correlation coefficient  
# then X3 does.