```
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 3.6.3
```

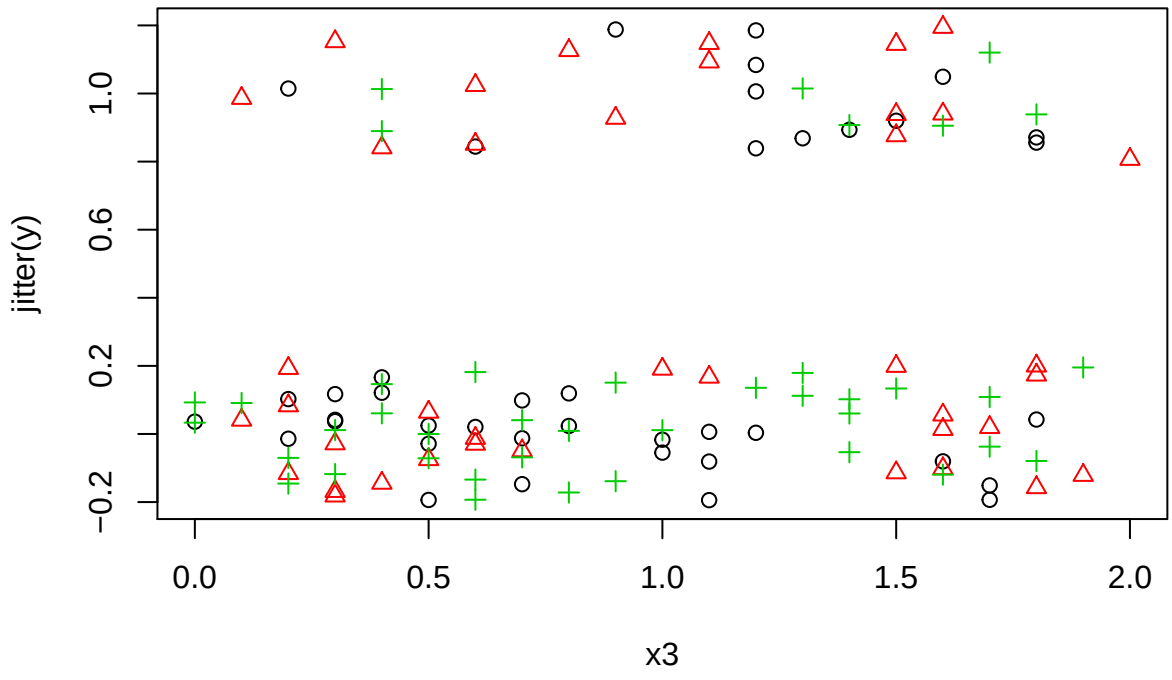
```
library(car)
```

```
## Loading required package: carData
```

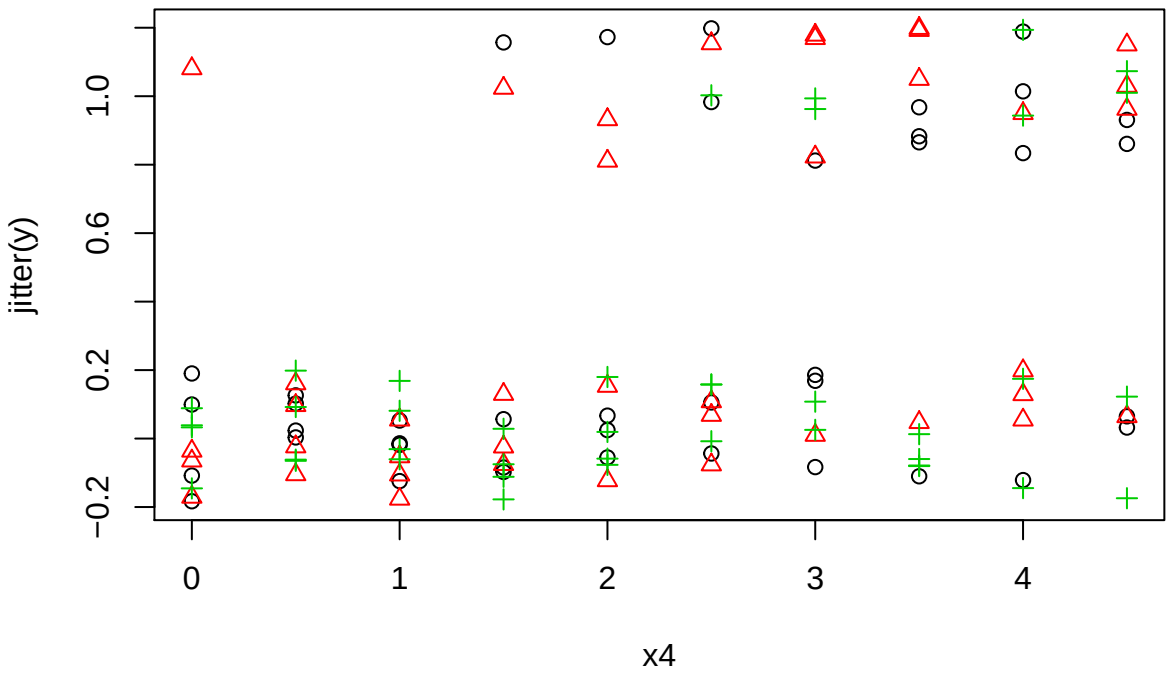
```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 3.6.3
```

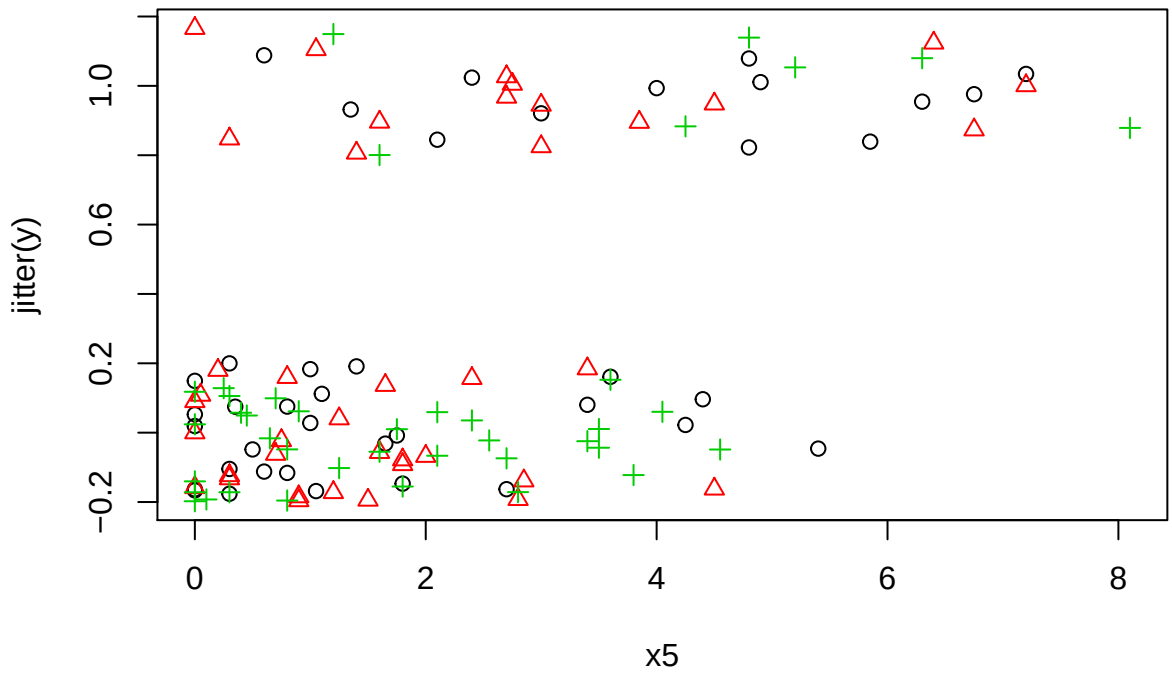
```
##read in data
data = read.table('C:/Users/Ryan/Desktop/Stats/572/exam/draxinus_mod.csv', header = TRUE, sep = ",", de

##plot data raw
myind = c(rep(1,40), rep(2,40), rep(3,40))
plot(jitter(y)~x3, pch=myind, col=myind, data=data)
```




```
plot(jitter(y)~x4, pch=myind, col=myind, data=data)
```




```
plot(jitter(y)~x5, pch=myind, col=myind, data=data)
```




```
##create logistic regression models
fullmodel = glm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11, family = binomial(logit), data = data)

fullmodel2 = glm(y~x1+x2+x3+x4+x6+x7+x9+x10, family = binomial(logit), data = data)

sameintercept = glm(y~x3+x4+x6+x7+x9+x10, family = binomial(logit), data = data)

nodiff = glm(y~x3+x4, family = binomial(logit), data = data)

onlyinteraction = glm(y~x5, family = binomial(logit), data = data)

new = glm(y~x1+x2+x3+x4+x5, family = binomial(logit), data = data)

##summary stats for each model
summary(fullmodel)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +
##     x10 + x11, family = binomial(logit), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5846  -0.6826  -0.3409   0.3861   2.2926
##
## Coefficients:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.1575     3.6322  -1.145    0.252
## x1            1.6374     4.1894   0.391    0.696
## x2            3.3961     3.9240   0.865    0.387
## x3           -1.2036     3.0842  -0.390    0.696
## x4            0.2713     1.1097   0.245    0.807
## x5            1.0599     1.0517   1.008    0.314
## x6            0.3673     3.8544   0.095    0.924
## x7           -0.1086     1.3291  -0.082    0.935
## x8           -0.2345     1.3207  -0.178    0.859
## x9           -0.6673     3.4269  -0.195    0.846
## x10          -0.4775     1.2312  -0.388    0.698
## x11           0.1718     1.2483   0.138    0.891
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 144.873  on 119  degrees of freedom
## Residual deviance:  97.281  on 108  degrees of freedom
## AIC: 121.28
##
## Number of Fisher Scoring iterations: 6
```

```
summary(fullmodel2)
```

```
##
## Call:
## glm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x9 + x10, family = binomial(logit),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5279  -0.7031  -0.3224   0.6379   2.3894
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -7.9186     2.9222  -2.710  0.00673 **
## x1            3.3912     3.2474   1.044  0.29636
## x2            4.7172     3.1999   1.474  0.14044
## x3            2.0070     1.1571   1.735  0.08282 .
## x4            1.4401     0.6081   2.368  0.01788 *
## x6           -0.5364     1.4285  -0.375  0.70731
## x7           -0.5486     0.7032  -0.780  0.43532
## x9           -1.2130     1.3345  -0.909  0.36337
## x10          -0.6555     0.6853  -0.956  0.33883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 144.87  on 119  degrees of freedom
## Residual deviance: 103.77  on 111  degrees of freedom
## AIC: 121.77
##
## Number of Fisher Scoring iterations: 6
```

```r
summary(sameintercept)
```

```
##
## Call:
## glm(formula = y ~ x3 + x4 + x6 + x7 + x9 + x10, family = binomial(logit),
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5349  -0.7173  -0.3209   0.5962   2.7631
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.60079    0.91593  -5.023 5.08e-07 ***
## x3           0.94021    0.68713   1.368   0.1712
## x4           0.81702    0.28067   2.911   0.0036 **
## x6           0.55792    0.89347   0.624   0.5323
## x7           0.08765    0.33301   0.263   0.7924
## x9           0.40235    0.75320   0.534   0.5932
## x10          0.25724    0.28630   0.899   0.3689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 144.87  on 119  degrees of freedom
## Residual deviance: 106.62  on 113  degrees of freedom
## AIC: 120.62
##
## Number of Fisher Scoring iterations: 5
```

```r
summary(nodiff)
```

```
##
## Call:
## glm(formula = y ~ x3 + x4, family = binomial(logit), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4869  -0.7371  -0.3758   0.6902   2.7362
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.4311     0.8720  -5.082 3.74e-07 ***
## x3            1.1860     0.4397   2.697  0.00699 **
## x4            0.9038     0.2038   4.435 9.19e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 144.87  on 119  degrees of freedom
```

```
## Residual deviance: 111.16  on 117  degrees of freedom
## AIC: 117.16
##
## Number of Fisher Scoring iterations: 5
```

```
summary(onlyinteraction)
```

```
##
## Call:
## glm(formula = y ~ x5, family = binomial(logit), data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6602  -0.6702  -0.4354   0.5185   2.2962
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.5621     0.4311  -5.943 2.79e-09 ***
## x5            0.6759     0.1367   4.945 7.60e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 144.87  on 119  degrees of freedom
## Residual deviance: 109.93  on 118  degrees of freedom
## AIC: 113.93
##
## Number of Fisher Scoring iterations: 4
```

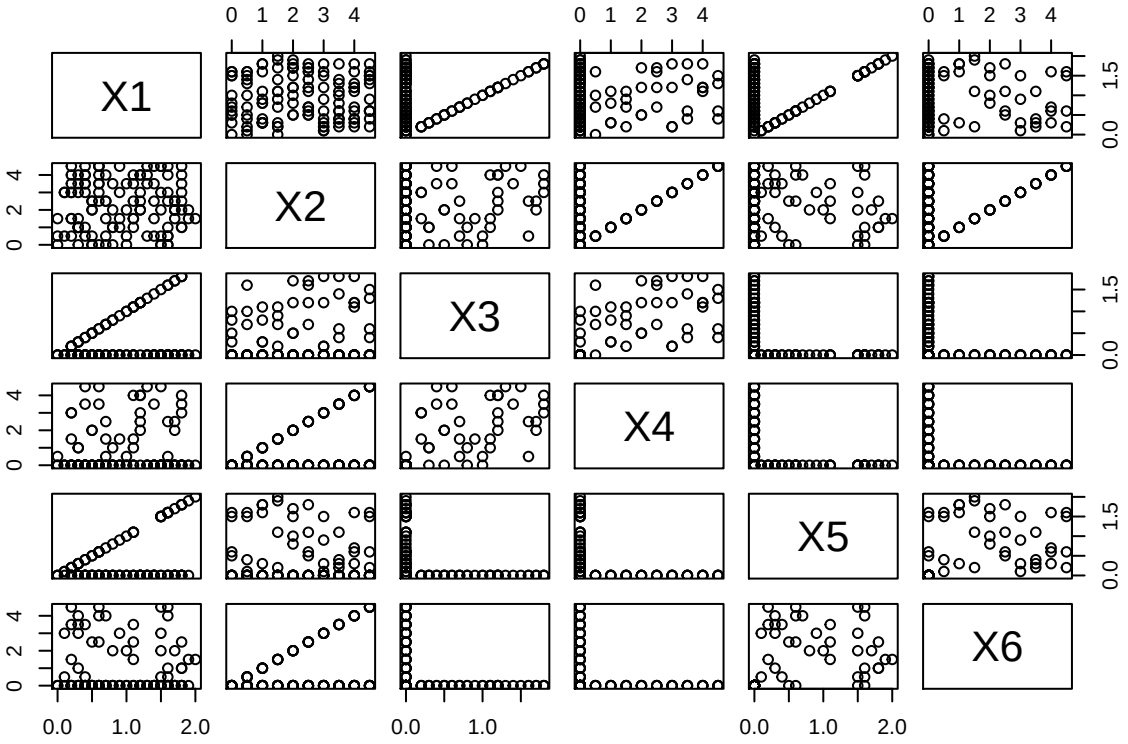
```
##test for sigificant differences between full model and reduced models (parsimony test)

fullvsfull2 = 1-pchisq((103.8-97.281),3)
full2vssameintercept = 1-pchisq((106.6-103.8),2)

#perform exhaustive search#
results <- regsubsets(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11, data = data, nvmax = 8)
sum.results = summary(results)
mods = sum.results$which
bic = sum.results$bic
results.table = cbind(mods,bic)
results.table[order(bic, decreasing = F),]
```

```
##   (Intercept) x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11        bic
## 2           1  0  0  0  0  1  0  0  0  0   1   0 -31.262376
## 1           1  0  0  0  0  1  0  0  0  0   0   0 -30.835841
## 3           1  0  1  1  0  1  0  0  0  0   0   0 -30.107971
## 4           1  1  1  1  0  1  0  0  0  0   0   0 -27.380324
## 5           1  1  1  1  0  1  0  0  1  0   0   0 -22.788693
## 6           1  0  1  1  1  1  0  1  0  0   1   0 -18.110338
## 7           1  1  1  1  0  1  0  0  1  1   0   1 -13.505745
## 8           1  1  1  1  1  1  0  0  1  1   0   1  -8.748061
```

```
results <- regsubsets(y~x1+x2+x3+x4+x6+x7+x9+x10, data = data, nvmax = 8)
sum.results = summary(results)
mods = sum.results$which
bic = sum.results$bic
results.table = cbind(mods,bic)
results.table[order(bic, decreasing = F),]
```

```
##   (Intercept) x1 x2 x3 x4 x6 x7 x9 x10        bic
## 2           1  0  0  1  1  0  0  0   0 -19.042430
## 3           1  0  0  0  1  1  0  1   0 -17.386696
## 1           1  0  0  0  1  0  0  0   0 -16.723444
## 4           1  0  1  1  1  1  0  0   0 -15.227099
## 5           1  0  1  1  1  1  0  0   1 -10.616254
## 6           1  0  1  1  1  1  1  0   1  -6.070687
## 7           1  0  1  1  1  1  1  1   1  -1.294087
## 8           1  1  1  1  1  1  1  1   1   3.486896
```

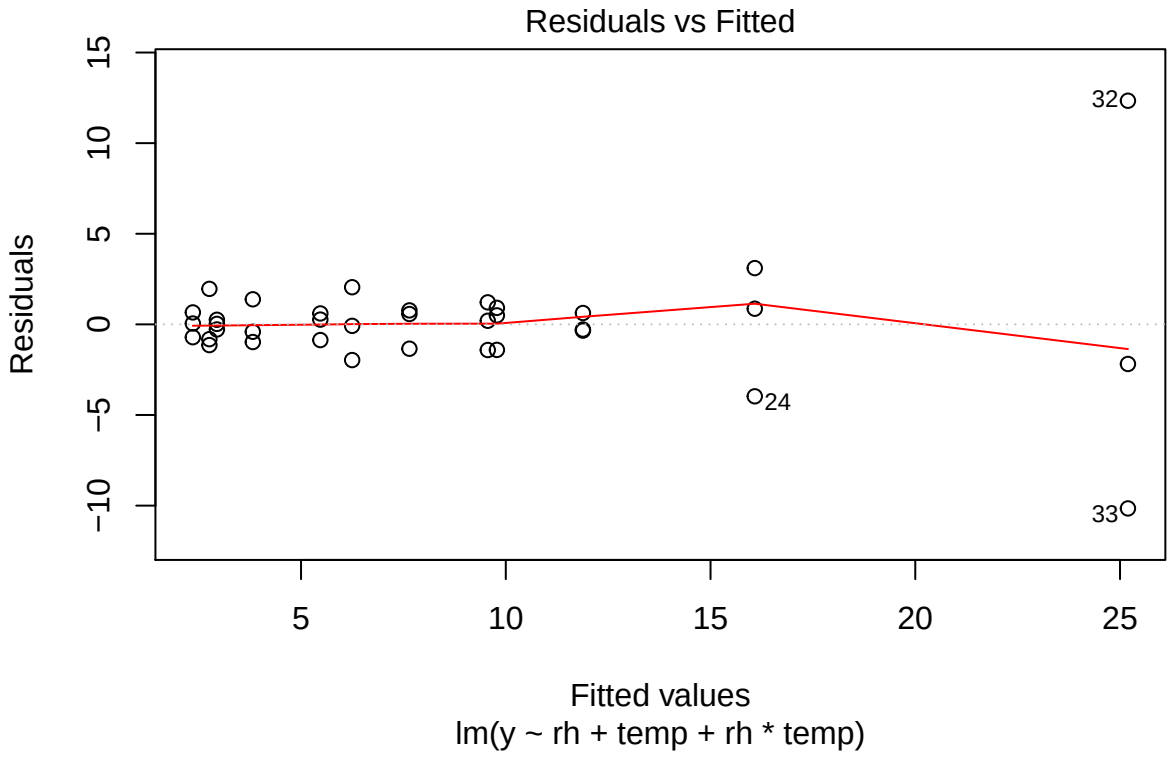
```
##Assumptions and Multicollinearity
subset = cbind(data$x3, data$x4, data$x6, data$x7, data$x9, data$x10)
new = data.frame(subset)
plot(new)
```

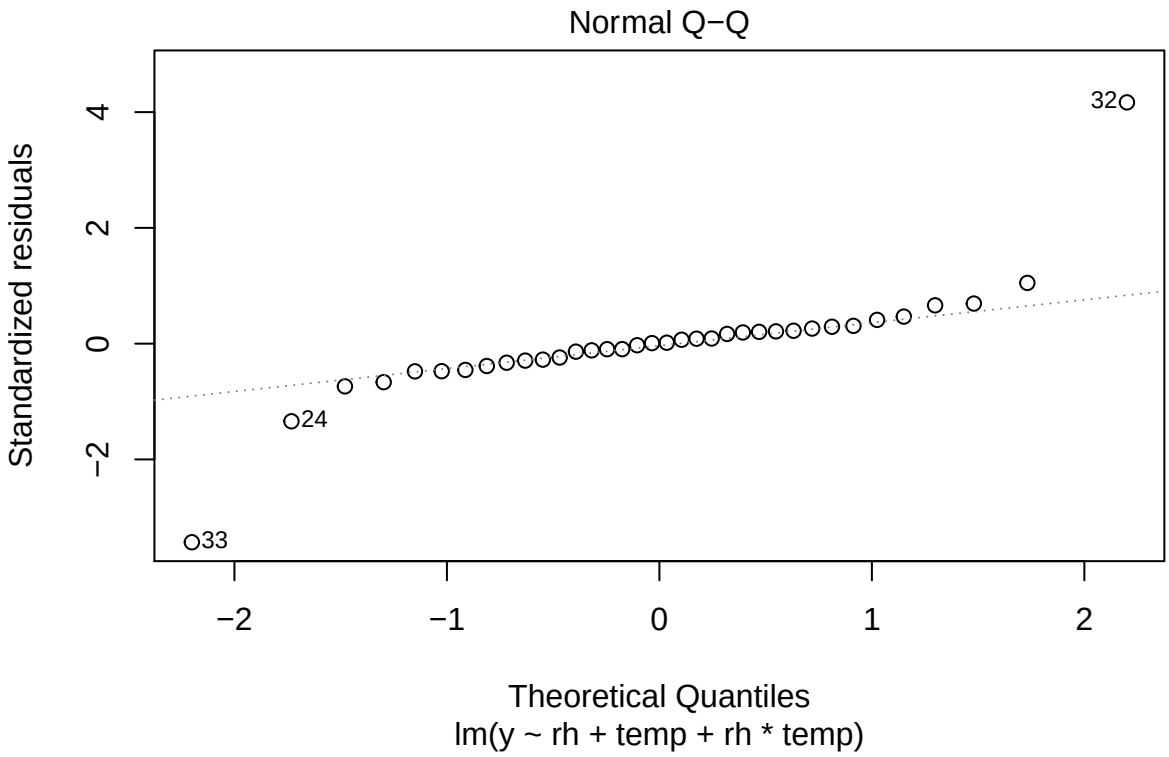


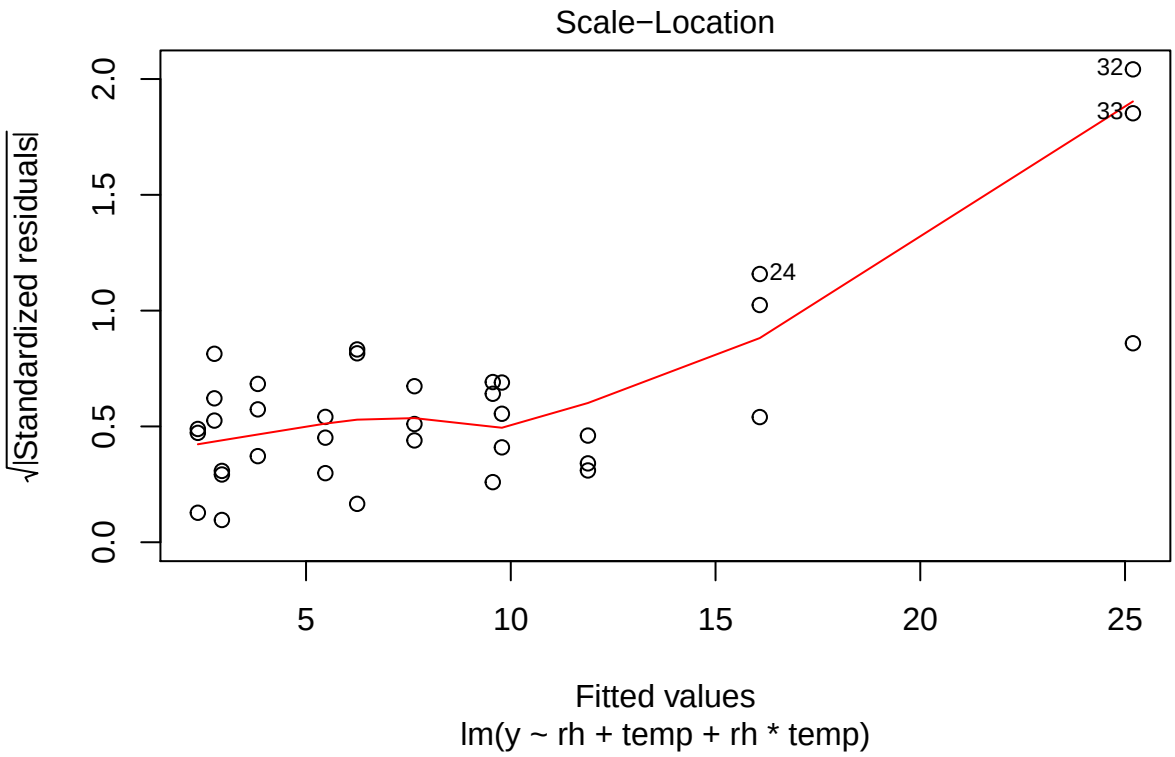

```
library(MASS)
```

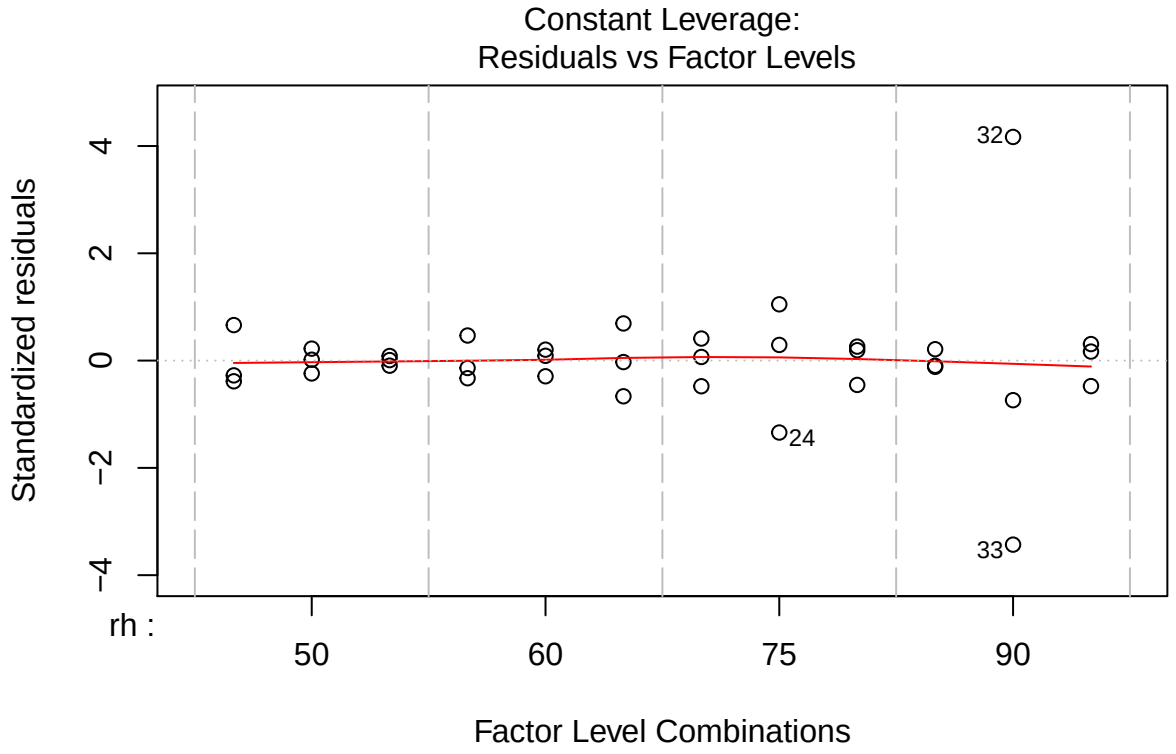
```
## Warning: package 'MASS' was built under R version 3.6.3
```

```r
data = read.table('C:/Users/Ryan/Desktop/Stats/572/exam/spores.csv', header = TRUE, sep = ",", dec = "."

data$rh = factor(data$rh)
data$temp = factor(data$temp)

model = lm(y~rh+temp+rh*temp, data = data)

plot(model)
```

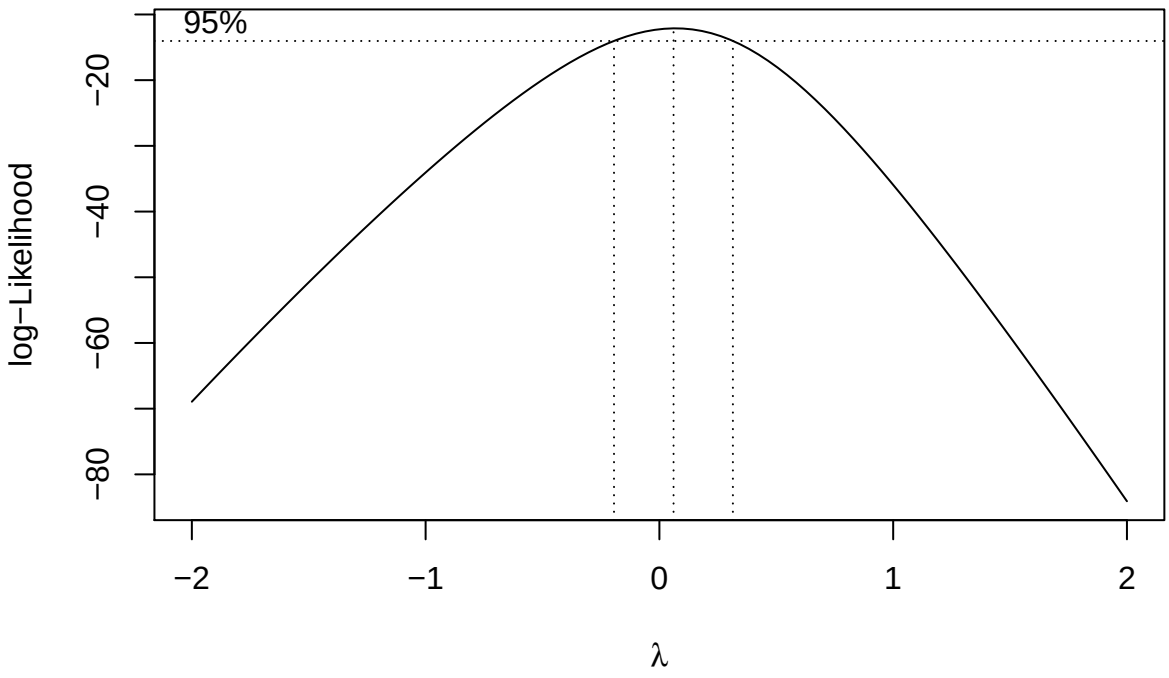



Normal Q−Q
Standardized residuals
4
2
0
−2
32
24
33
−2
−1
0
1
2
Theoretical Quantiles
lm(y ~ rh + temp + rh * temp)

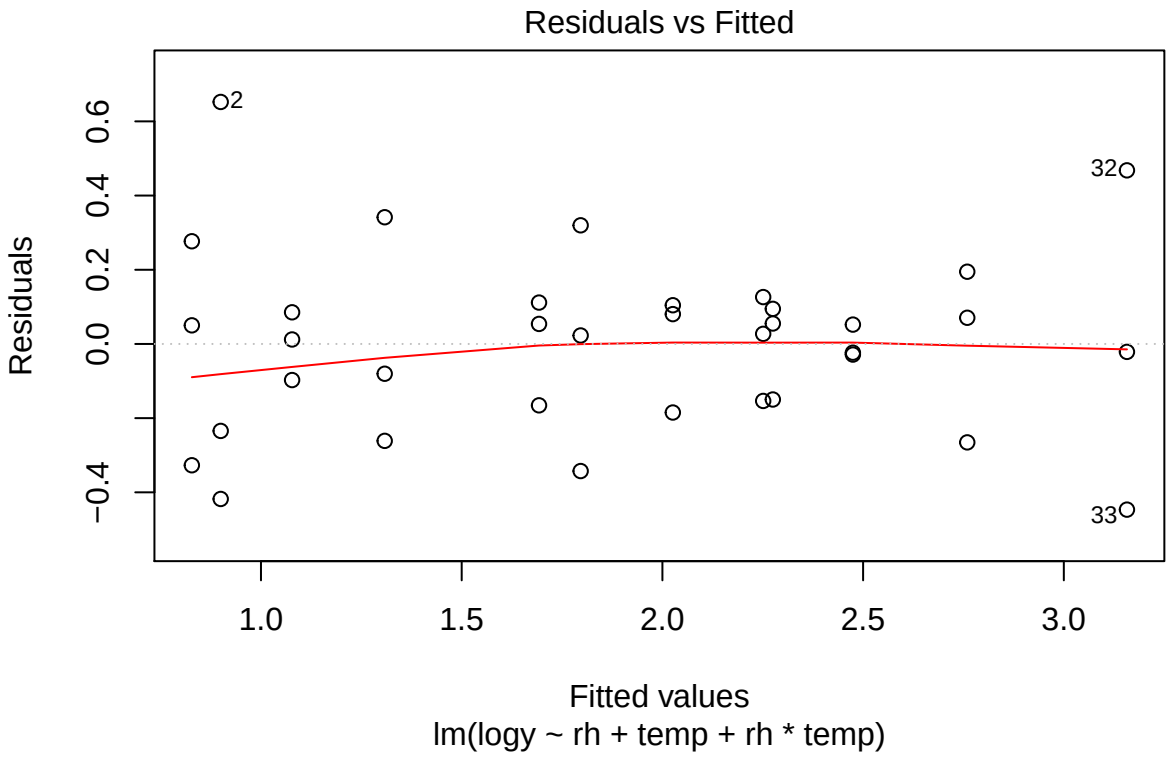

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
5
10
15
20
25
32
33
24
Fitted values
lm(y ~ rh + temp + rh * temp)

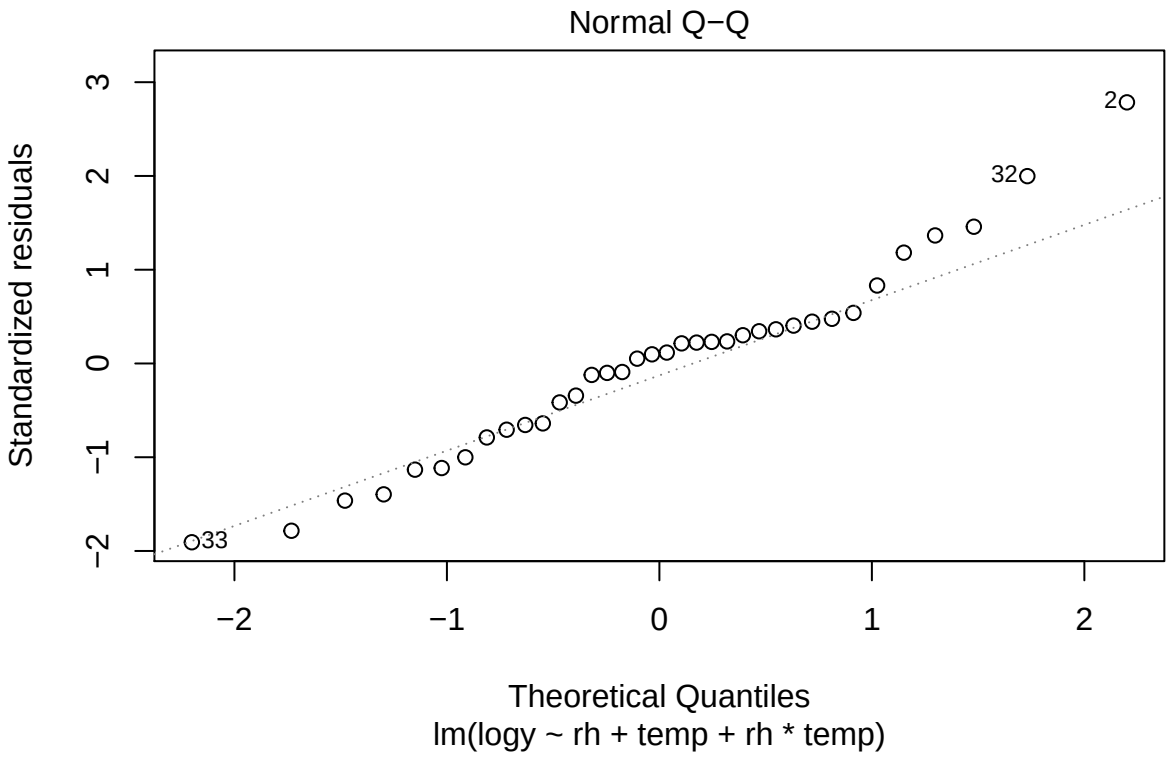


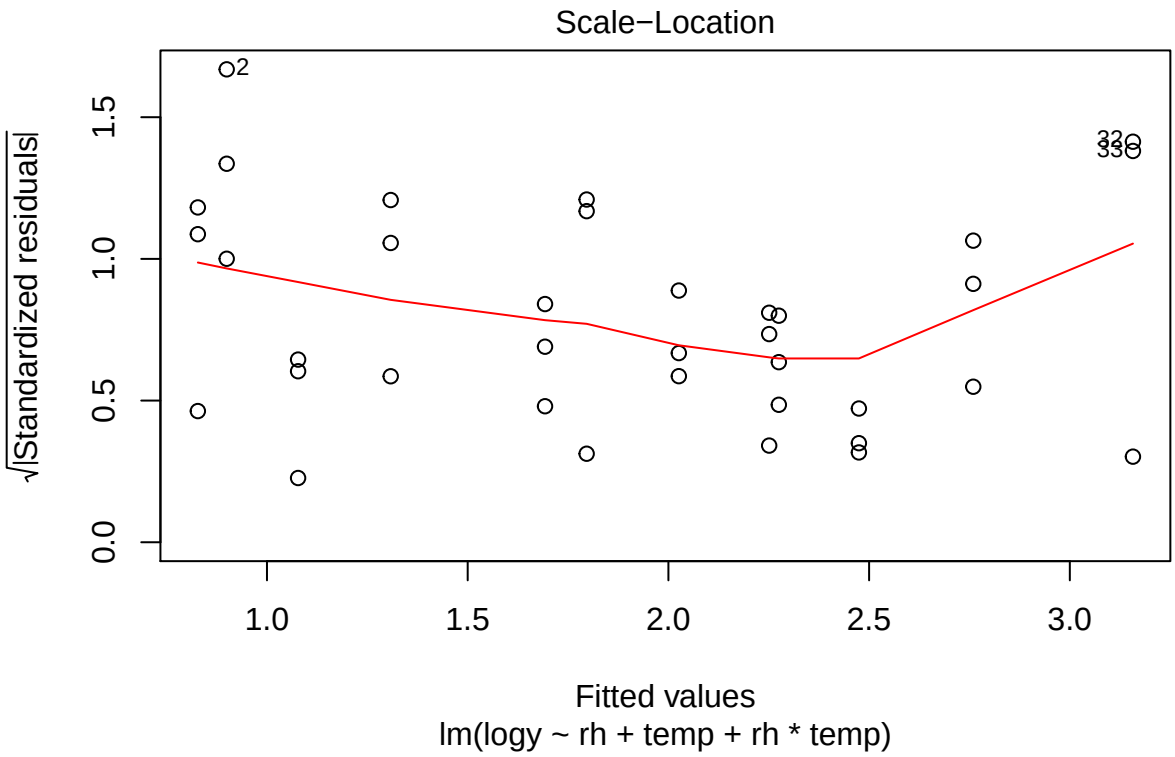

```
boxcox(model)
```

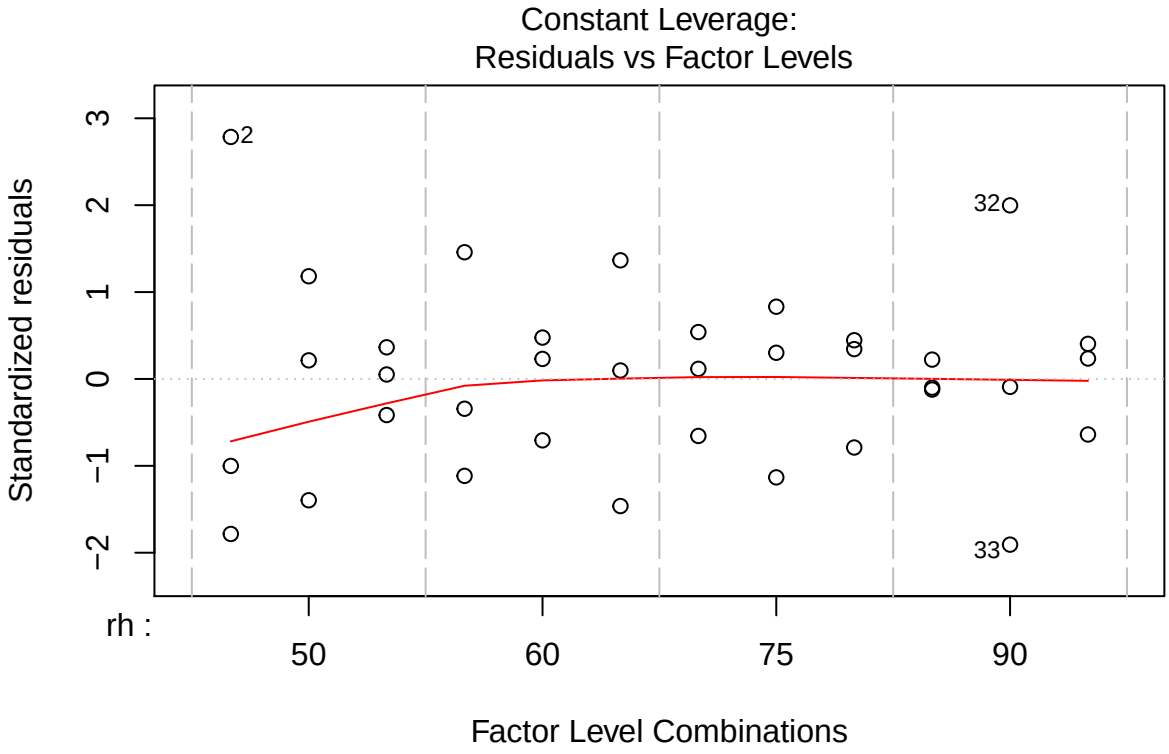



```
log.model = lm(logy~rh+temp+rh*temp, data = data)

plot(log.model)
```


Residuals vs Fitted
Residuals
−0.4
0.0
0.2
0.4
0.6
2
32
33
1.0
1.5
2.0
2.5
3.0
Fitted values
lm(logy ~ rh + temp + rh * temp)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(logy ~ rh + temp + rh * temp)
33
32
2
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
32
33
1.0
1.5
2.0
2.5
3.0
Fitted values
lm(logy ~ rh + temp + rh * temp)


Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
3
2
1
0
−1
−2
2
32
33
rh :
50
60
75
90
Factor Level Combinations