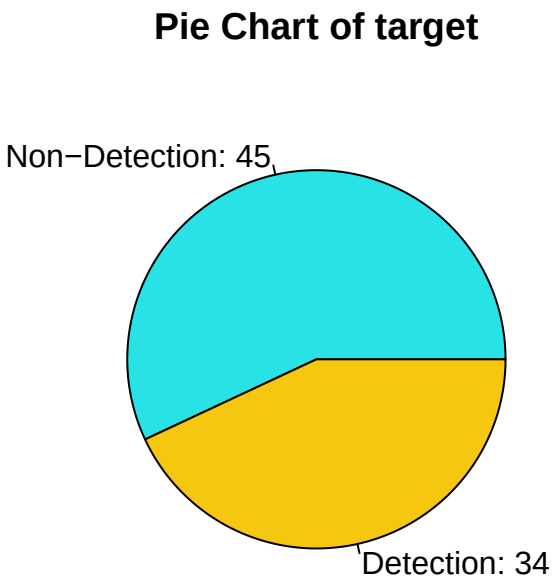
BST223 Project1

Shutong Gu

2/20/2024

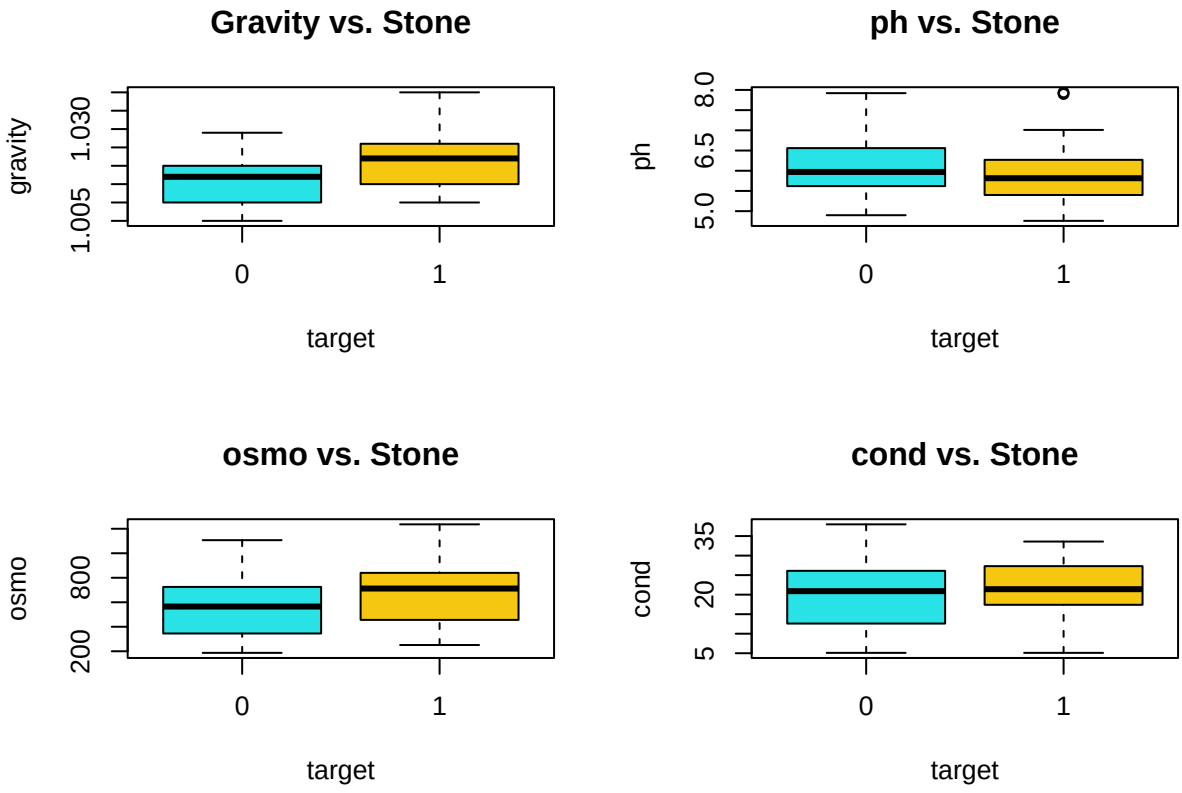
```
stone <-read.csv("~/Desktop/Data/BST223/kindey stone urine analysis.csv")
stone$target <- factor(stone$target)
```

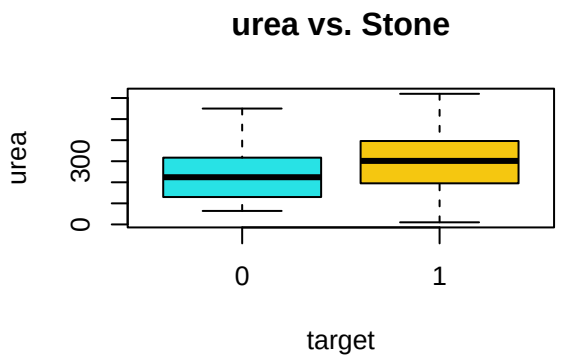
```
n_0 <- nrow(stone[stone$target=='0',])
n_1 <- nrow(stone[stone$target=='1',])
category_counts=c("Non-Detection" = n_0,"Detection"=n_1)
pie(category_counts,
    main = "Pie Chart of target",
    col = c(5,7),
    labels = paste(names(category_counts), ": ", category_counts, sep=""))
```

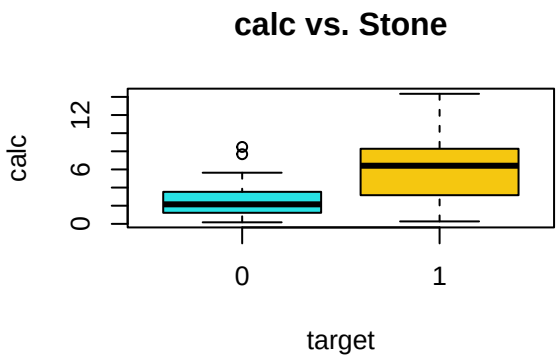


```
cor(stone[,-7])
```

```
##              gravity          ph       osmo        cond       urea       calc
## gravity  1.0000000 -0.25334018  0.8614657  0.55906435  0.8234770  0.5256987
## ph      -0.2533402  1.00000000 -0.2388411 -0.09767955 -0.2755569 -0.1194878
## osmo     0.8614657 -0.23884108  1.0000000  0.81277999  0.8711785  0.5229794
## cond     0.5590643 -0.09767955  0.8127800  1.00000000  0.4954049  0.3510295
## urea     0.8234770 -0.27555694  0.8711785  0.49540493  1.0000000  0.5023267
## calc     0.5256987 -0.11948777  0.5229794  0.35102955  0.5023267  1.0000000
```








```
model_pre <- glm(target~.,data=stone,family = binomial)
summary(model_pre)
```

```
##
## Call:
## glm(formula = target ~ ., family = binomial, data = stone)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5803  -0.6054  -0.2887   0.3993   2.6034
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.631e+02  1.763e+02  -3.194  0.00140 **
## gravity      5.621e+02  1.761e+02   3.192  0.00141 **
## ph          -3.201e-01  5.476e-01  -0.585  0.55885
## osmo        -5.340e-03  7.128e-03  -0.749  0.45373
## cond        -1.391e-01  1.187e-01  -1.171  0.24141
## urea        -1.489e-02  8.601e-03  -1.731  0.08341 .
## calc         7.218e-01  2.210e-01   3.266  0.00109 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 107.981  on 78  degrees of freedom
## Residual deviance:  60.189  on 72  degrees of freedom
## AIC: 74.189
##
## Number of Fisher Scoring iterations: 6
```

```
library(MASS)
stepAIC(model_pre)
```

```
## Start:  AIC=74.19
## target ~ gravity + ph + osmo + cond + urea + calc
##
##           Df Deviance    AIC
## - ph       1   60.535 72.535
## - osmo     1   60.847 72.847
## - cond     1   61.336 73.336
## <none>         60.189 74.189
## - urea     1   62.658 74.658
## - gravity  1   75.704 87.704
## - calc     1   81.452 93.452
##
## Step:  AIC=72.54
## target ~ gravity + osmo + cond + urea + calc
##
##           Df Deviance    AIC
## - osmo     1   61.128 71.128
## - cond     1   61.673 71.673
## <none>         60.535 72.535
## - urea     1   62.820 72.820
## - gravity  1   76.104 86.104
## - calc     1   81.774 91.774
##
## Step:  AIC=71.13
## target ~ gravity + cond + urea + calc
##
##           Df Deviance    AIC
## <none>         61.128 71.128
## - urea     1   69.858 77.858
## - cond     1   73.293 81.293
## - gravity  1   78.207 86.207
## - calc     1   82.612 90.612

##
## Call:  glm(formula = target ~ gravity + cond + urea + calc, family = binomial,
##     data = stone)
##
## Coefficients:
## (Intercept)      gravity         cond         urea         calc
##  -505.48108    502.54060     -0.20836     -0.01776      0.73170
##
## Degrees of Freedom: 78 Total (i.e. Null);  74 Residual
## Null Deviance:       108
## Residual Deviance: 61.13     AIC: 71.13
```

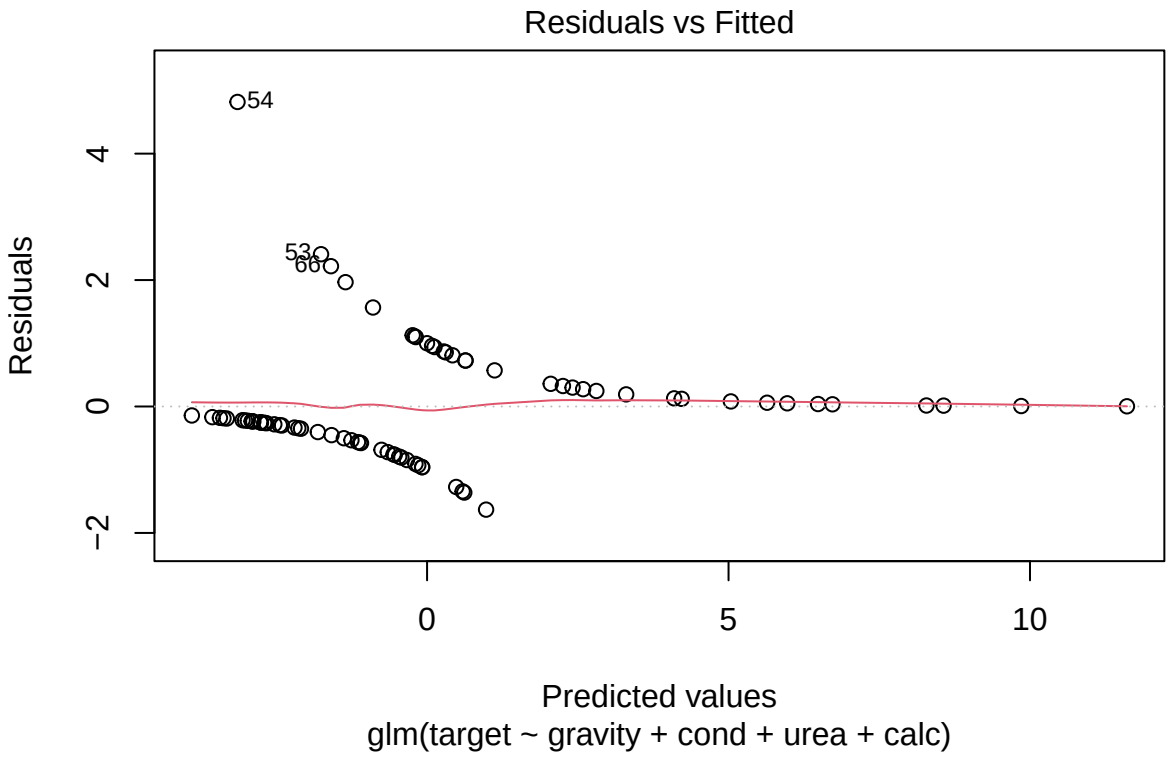
```r
model_pre_AIC <- glm(formula = target ~ gravity + cond + urea + calc, family = binomial,
    data = stone)
summary(model_pre_AIC)
```

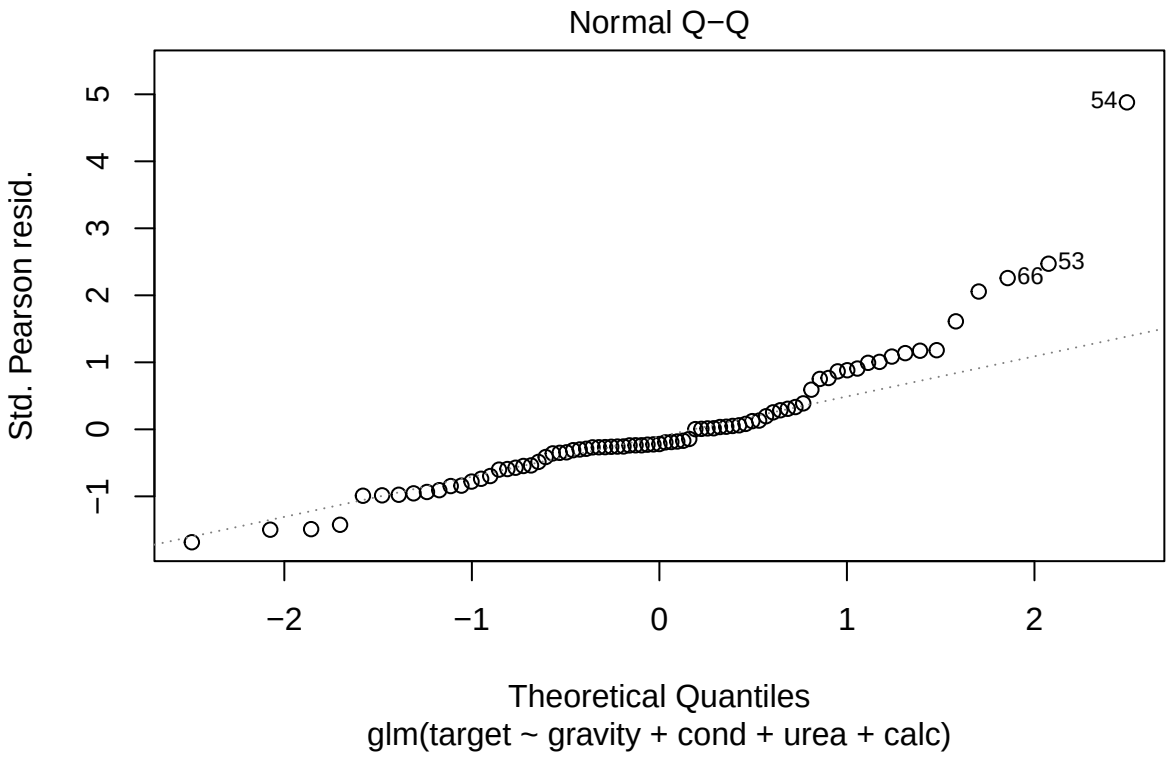
```
##
## Call:
## glm(formula = target ~ gravity + cond + urea + calc, family = binomial,
##     data = stone)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6111  -0.6404  -0.3027   0.3974   2.5246
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.055e+02  1.613e+02  -3.133 0.001730 **
## gravity      5.025e+02  1.608e+02   3.125 0.001776 **
## cond        -2.084e-01  7.157e-02  -2.911 0.003598 **
## urea        -1.776e-02  6.924e-03  -2.565 0.010313 *
## calc         7.317e-01  2.194e-01   3.335 0.000852 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.981  on 78  degrees of freedom
## Residual deviance:  61.128  on 74  degrees of freedom
## AIC: 71.128
##
## Number of Fisher Scoring iterations: 6
```

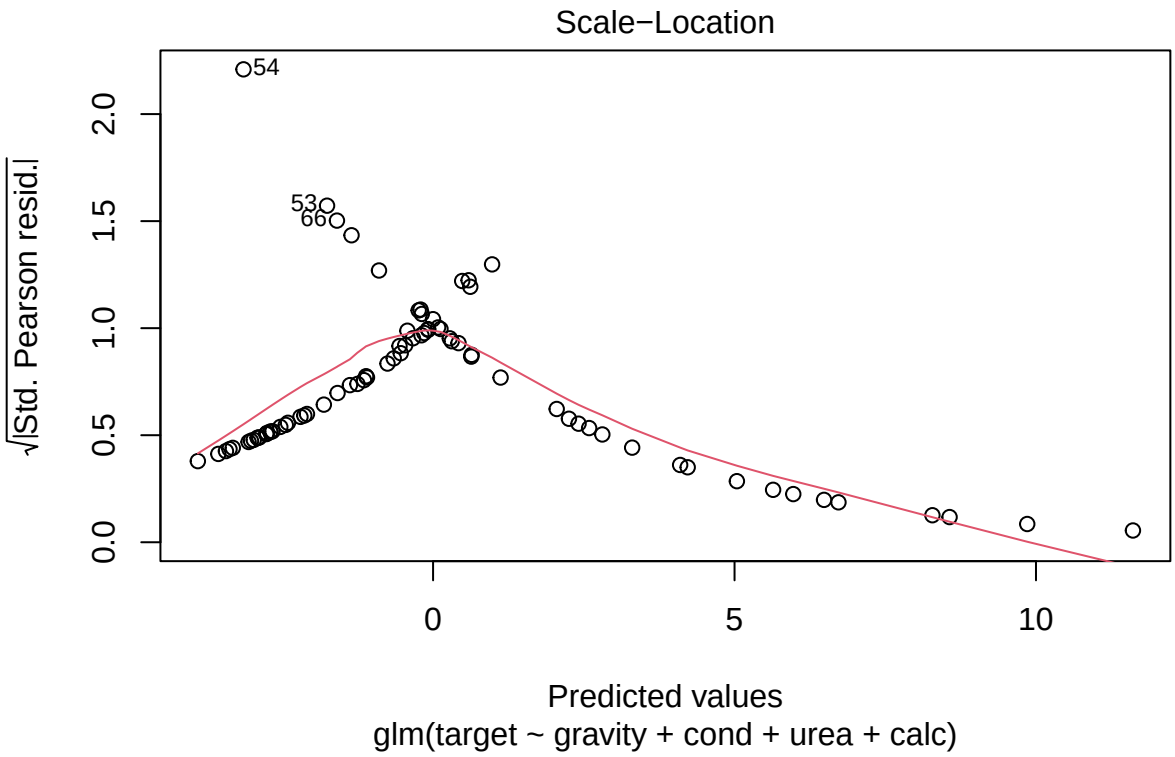
```r
anova(model_pre, model_pre_AIC, test="Chisq")
```

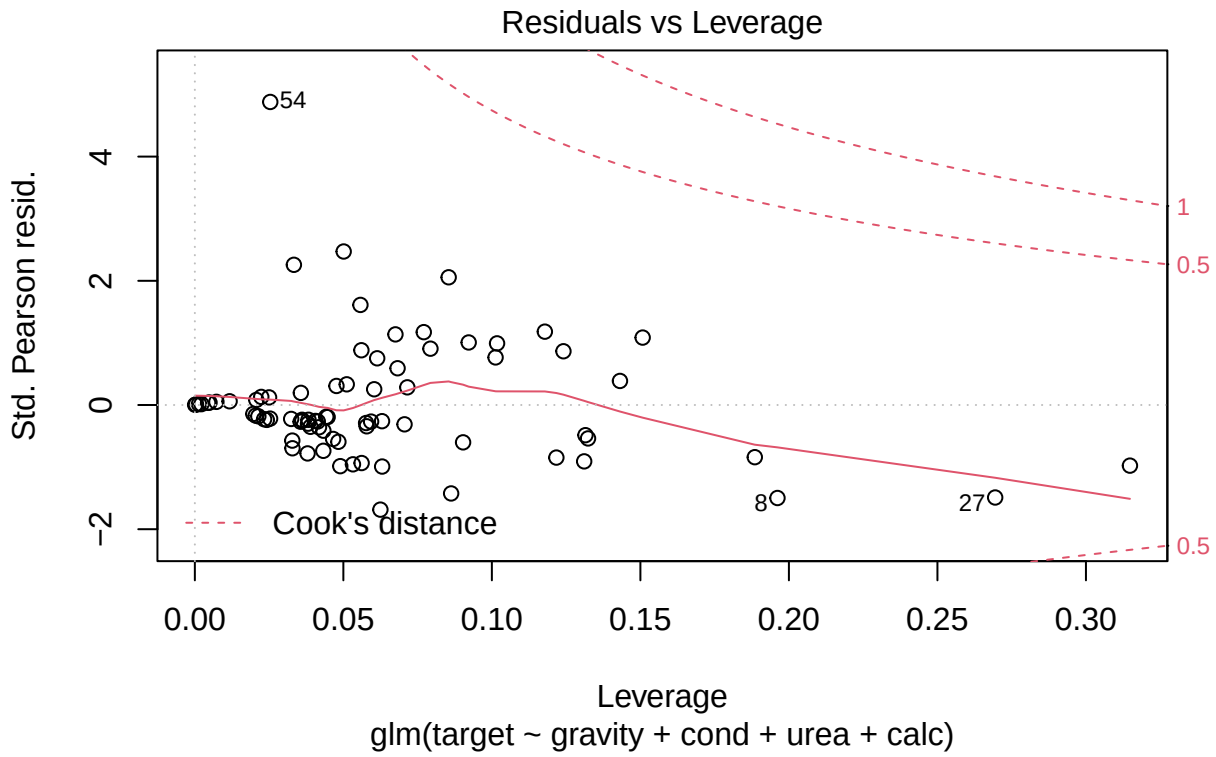
```
## Analysis of Deviance Table
##
## Model 1: target ~ gravity + ph + osmo + cond + urea + calc
## Model 2: target ~ gravity + cond + urea + calc
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        72     60.189
## 2        74     61.128 -2 -0.93909   0.6253
```

```r
plot(model_pre_AIC)
```

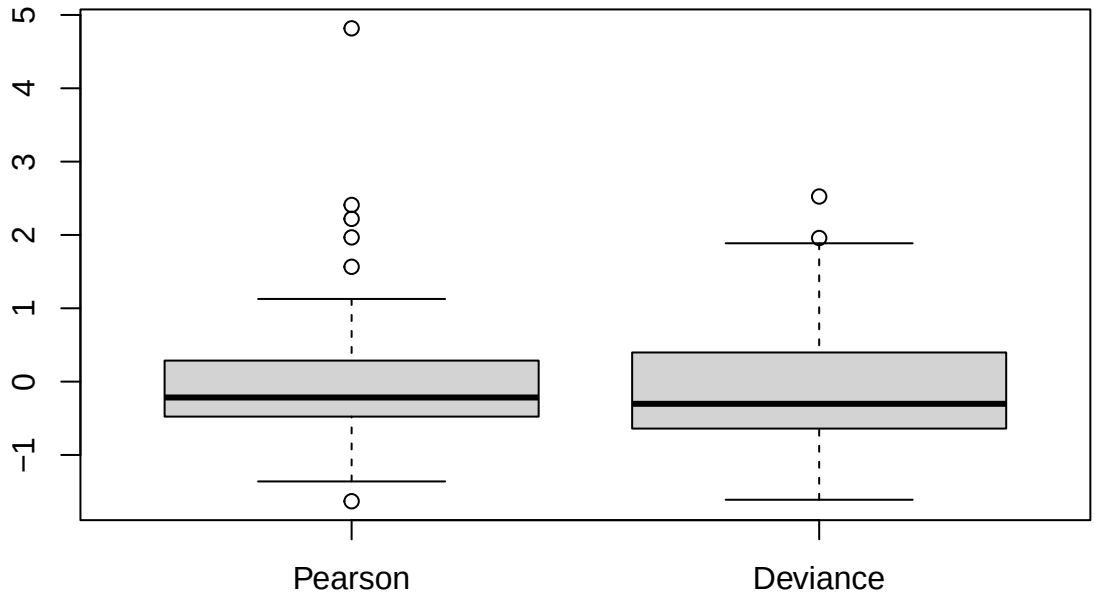

Residuals vs Fitted
54
53
66
Residuals
4
2
0
−2
0
5
10
Predicted values
glm(target ~ gravity + cond + urea + calc)


Normal Q−Q
Std. Pearson resid.
−1
0
1
2
3
4
5
54
66
53
−2
−1
0
1
2
Theoretical Quantiles
glm(target ~ gravity + cond + urea + calc)


Scale−Location
54
53
66
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
0
5
10
Predicted values
glm(target ~ gravity + cond + urea + calc)



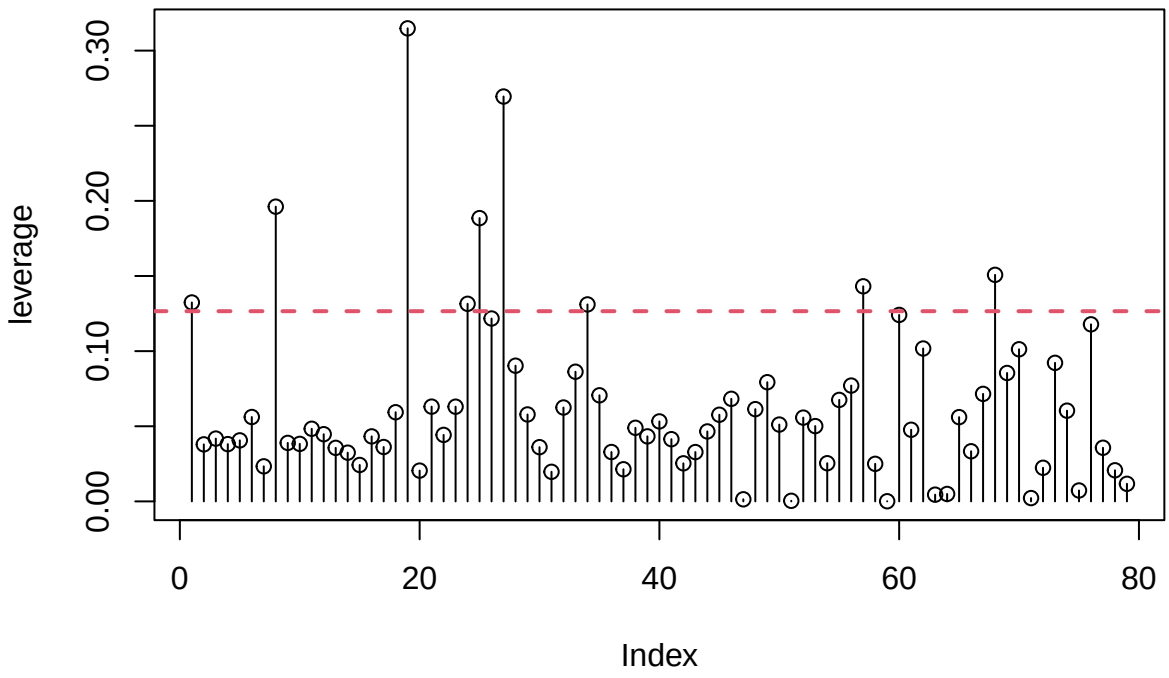

```
pear_res <- residuals(model_pre_AIC, "pearson")
dev_res <- residuals(model_pre_AIC, "deviance")
boxplot(cbind(pear_res,dev_res), names = c("Pearson","Deviance"))
```



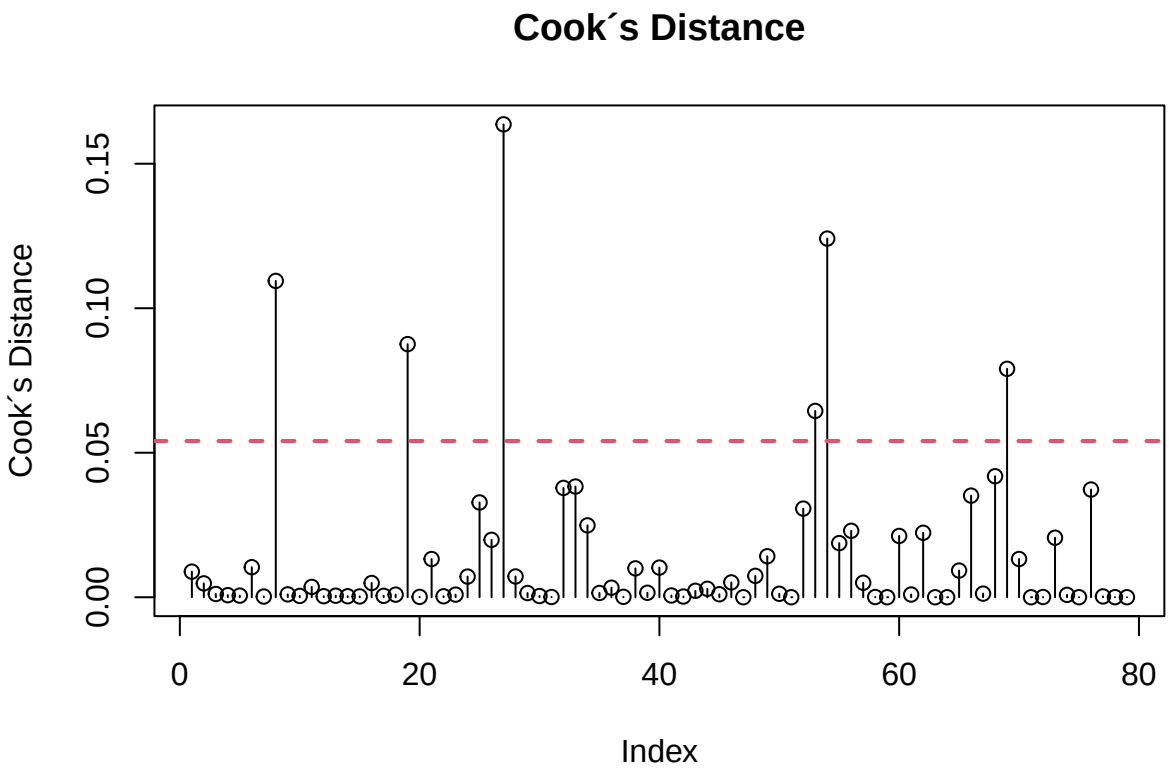

```
anova(model_pre_AIC, test="Chi")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: target
##
## Terms added sequentially (first to last)
##
##
##         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                       78    107.981
## gravity  1  15.0994        77     92.881  0.000102 ***
## cond     1   5.4252        76     87.456  0.019848 *
## urea     1   4.8439        75     82.612  0.027744 *
## calc     1  21.4836        74     61.128 3.569e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#par(mfrow=c(1,2))
leverage <- hatvalues(model_pre_AIC)
plot(names(leverage), leverage, xlab="Index", type="h")
points(names(leverage), leverage)
abline(h=2*length(model_pre_AIC$coefficients)/nrow(stone),col=2,lwd=2,lty=2)
```




```
cooks = cooks.distance(model_pre_AIC)
plot(names(cooks),cooks, xlab="Index", type="h",main="Cook's Distance", ylab="Cook's Distance")
points(names(cooks), cooks)
abline(h = 4/(79-length(model_pre_AIC$coefficients)) ,col=2,lwd=2,lty=2)
```




```
stone[c(8,19,27,54,69),]
```

```
##    gravity   ph osmo cond urea calc target
## 8    1.029 5.67 1107 35.9  550 8.48      0
## 19   1.008 6.88  395 26.1   95 7.68      0
## 27   1.017 7.61  527 25.8   75 2.17      0
## 54   1.011 7.01  443 21.4  124 1.27      1
## 69   1.025 6.90  945 33.6  396 4.18      1
```

#scale Data

```
stone_scale <- scale(stone[,-7])
stone_scale <- cbind(stone_scale,stone[,7])
stone_scale <- as.data.frame(stone_scale)
stone_scale[,7] <- as.factor(stone[,7])
names(stone_scale)[7] <- "target"
```

```
model_scale <- glm(target~.,data=stone_scale,family = binomial)
summary(model_scale)
```

```
##
## Call:
## glm(formula = target ~ ., family = binomial, data = stone_scale)
##
```

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5803  -0.6054  -0.2887   0.3993   2.6034
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.08347    0.35096   0.238  0.81202
## gravity      4.06940    1.27495   3.192  0.00141 **
## ph          -0.23186    0.39664  -0.585  0.55885
## osmo        -1.26844    1.69303  -0.749  0.45373
## cond        -1.10425    0.94261  -1.171  0.24141
## urea        -1.95439    1.12892  -1.731  0.08341 .
## calc         2.35317    0.72048   3.266  0.00109 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.981  on 78  degrees of freedom
## Residual deviance:  60.189  on 72  degrees of freedom
## AIC: 74.189
##
## Number of Fisher Scoring iterations: 6
```

```
stepAIC(model_scale)
```

```
## Start:  AIC=74.19
## target ~ gravity + ph + osmo + cond + urea + calc
##
##           Df Deviance    AIC
## - ph       1   60.535 72.535
## - osmo     1   60.847 72.847
## - cond     1   61.336 73.336
## <none>         60.189 74.189
## - urea     1   62.658 74.658
## - gravity  1   75.704 87.704
## - calc     1   81.452 93.452
##
## Step:  AIC=72.54
## target ~ gravity + osmo + cond + urea + calc
##
##           Df Deviance    AIC
## - osmo     1   61.128 71.128
## - cond     1   61.673 71.673
## <none>         60.535 72.535
## - urea     1   62.820 72.820
## - gravity  1   76.104 86.104
## - calc     1   81.774 91.774
##
## Step:  AIC=71.13
## target ~ gravity + cond + urea + calc
##
##           Df Deviance    AIC
## <none>         61.128 71.128
```

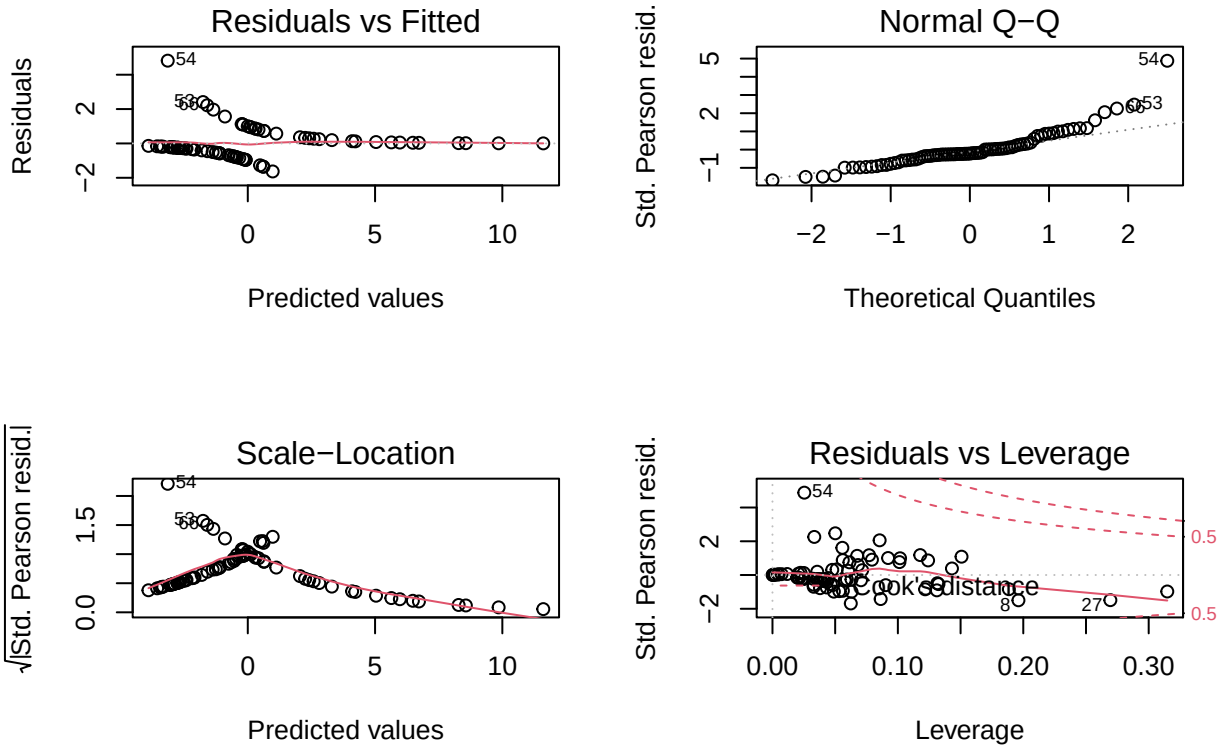
```
## - urea     1   69.858 77.858
## - cond     1   73.293 81.293
## - gravity  1   78.207 86.207
## - calc     1   82.612 90.612
```

```
##
## Call:  glm(formula = target ~ gravity + cond + urea + calc, family = binomial,
##     data = stone_scale)
##
## Coefficients:
## (Intercept)      gravity         cond         urea         calc
##      0.1226       3.6381      -1.6542      -2.3312       2.3854
##
## Degrees of Freedom: 78 Total (i.e. Null);  74 Residual
## Null Deviance:       108
## Residual Deviance: 61.13     AIC: 71.13
```

```
model_scale_AIC <- glm(target ~ gravity + cond + urea + calc, family = binomial,
    data = stone_scale)
summary(model_scale_AIC)
```

```
##
## Call:
## glm(formula = target ~ gravity + cond + urea + calc, family = binomial,
##     data = stone_scale)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6111  -0.6404  -0.3027   0.3974   2.5246
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.1226     0.3502   0.350 0.726252
## gravity       3.6381     1.1641   3.125 0.001776 **
## cond         -1.6542     0.5682  -2.911 0.003598 **
## urea         -2.3312     0.9088  -2.565 0.010313 *
## calc          2.3854     0.7152   3.335 0.000852 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 107.981  on 78  degrees of freedom
## Residual deviance:  61.128  on 74  degrees of freedom
## AIC: 71.128
##
## Number of Fisher Scoring iterations: 6
```

```
par(mfrow=c(2,2))
plot(model_scale_AIC)
```

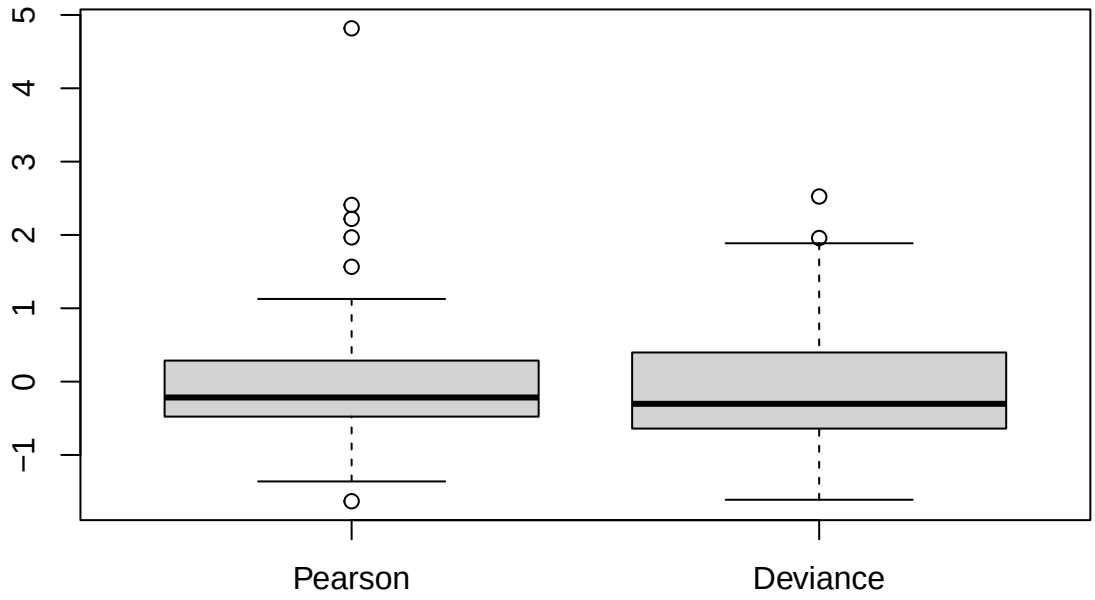


#LRT

```
null_model <- glm(target~ 1, data=stone_scale, family = "binomial")
anova(null_model,model_scale_AIC, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: target ~ 1
## Model 2: target ~ gravity + cond + urea + calc
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1        78    107.981
## 2        74     61.128  4   46.852 1.637e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pear_res <- residuals(model_scale_AIC, "pearson")
dev_res <- residuals(model_scale_AIC, "deviance")
boxplot(cbind(pear_res,dev_res), names = c("Pearson","Deviance"))
```




```
stone_scale[c(8,19,27,54,69),]
```

```
##       gravity         ph       osmo       cond       urea        calc target
## 8   1.5037116 -0.4949296  2.0805103 1.90025035  2.1606484  1.33157800      0
## 19 -1.3970530  1.1756326 -0.9171982 0.66583701 -1.3058981  1.08618309      0
## 27 -0.1538682  2.1834924 -0.3614432 0.62804884 -1.4582737 -0.60397433      0
## 54 -0.9826581  1.3551145 -0.7151055 0.07382245 -1.0849533 -0.88004360      1
## 69  0.9511850  1.2032452  1.3984474 1.61054110  0.9873558  0.01258037      1
```