

## Assignment-5

İlyürek Kılıç

2023-06-08

### PART1

Set the working directory

```
getwd()

## [1] "/Users/ilyurekkilic/R works/STAT364-Reg"

setwd("/Users/ilyurekkilic/R Works/STAT364-Reg")
```

Read data

```
heart<-read.csv("heart.csv")
# view the first rows of the data
head(heart)

##   age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 1  63  1  3   145   233   1       0    150     0     2.3    0  0    1
## 2  37  1  2   130   250   0       1    187     0     3.5    0  0    2
## 3  41  0  1   130   204   0       0    172     0     1.4    2  0    2
## 4  56  1  1   120   236   0       1    178     0     0.8    2  0    2
## 5  57  0  0   120   354   0       1    163     1     0.6    2  0    2
## 6  57  1  0   140   192   0       1    148     0     0.4    1  0    1
##   target
## 1      1
## 2      1
## 3      1
## 4      1
## 5      1
## 6      1
```

Frequency of the response variable 0's means less chance of heart attack 1's means higher chance of heart attack

```
prop.table(table(heart$target))

##
##           0           1
## 0.4554455 0.5445545

# %45 of the data consist of 0's
# %55 of the data consist of 1's
```

Correlation between covariates and detect variables

cor(heart)

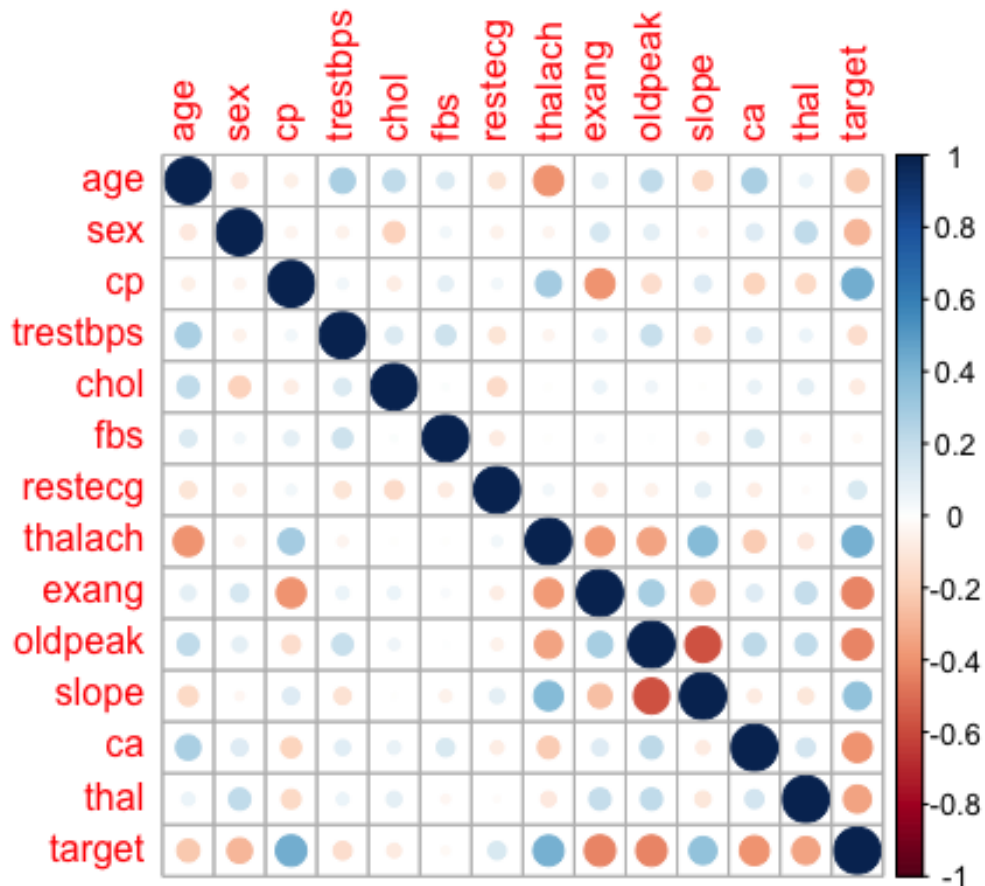
```
##          age          sex          cp      trestbps          chol
## age      1.00000000 -0.09844660 -0.06865302  0.27935091  0.213677957
## sex     -0.09844660  1.00000000 -0.04935288 -0.05676882 -0.197912174
## cp      -0.06865302 -0.04935288  1.00000000  0.04760776 -0.076904391
## trestbps 0.27935091 -0.05676882  0.04760776  1.00000000  0.123174207
## chol     0.21367796 -0.19791217 -0.07690439  0.12317421  1.000000000
## fbs      0.12130765  0.04503179  0.09444403  0.17753054  0.013293602
## restecg  -0.11621090 -0.05819627  0.04442059 -0.11410279 -0.151040078
## thalach  -0.39852194 -0.04401991  0.29576212 -0.04669773 -0.009939839
## exang     0.09680083  0.14166381 -0.39428027  0.06761612  0.067022783
## oldpeak  0.21001257  0.09609288 -0.14923016  0.19321647  0.053951920
## slope    -0.16881424 -0.03071057  0.11971659 -0.12147458 -0.004037770
## ca        0.27632624  0.11826141 -0.18105303  0.10138899  0.070510925
## thal      0.06800138  0.21004110 -0.16173557  0.06220989  0.098802993
## target   -0.22543872 -0.28093658  0.43379826 -0.14493113 -0.085239105
##          fbs      restecg      thalach      exang      oldpeak
## age      0.121307648 -0.11621090 -0.398521938  0.09680083  0.210012567
## sex      0.045031789 -0.05819627 -0.044019908  0.14166381  0.096092877
## cp        0.094444035  0.04442059  0.295762125 -0.39428027 -0.149230158
## trestbps  0.177530542 -0.11410279 -0.046697728  0.06761612  0.193216472
## chol      0.013293602 -0.15104008 -0.009939839  0.06702278  0.053951920
## fbs       1.000000000 -0.08418905 -0.008567107  0.02566515  0.005747223
## restecg  -0.084189054  1.00000000  0.044123444 -0.07073286 -0.058770226
## thalach  -0.008567107  0.04412344  1.000000000 -0.37881209 -0.344186948
## exang     0.025665147 -0.07073286 -0.378812094  1.00000000  0.288222808
## oldpeak  0.005747223 -0.05877023 -0.344186948  0.28822281  1.000000000
## slope    -0.059894178  0.09304482  0.386784410 -0.25774837 -0.577536817
## ca        0.137979327 -0.07204243 -0.213176928  0.11573938  0.222682322
## thal     -0.032019339 -0.01198140 -0.096439132  0.20675379  0.210244126
## target   -0.028045760  0.13722950  0.421740934 -0.43675708 -0.430696002
##          slope          ca          thal      target
## age     -0.16881424  0.27632624  0.06800138 -0.22543872
## sex     -0.03071057  0.11826141  0.21004110 -0.28093658
## cp       0.11971659 -0.18105303 -0.16173557  0.43379826
## trestbps -0.12147458  0.10138899  0.06220989 -0.14493113
## chol     -0.00403777  0.07051093  0.09880299 -0.08523911
## fbs      -0.05989418  0.13797933 -0.03201934 -0.02804576
## restecg  0.09304482 -0.07204243 -0.01198140  0.13722950
## thalach  0.38678441 -0.21317693 -0.09643913  0.42174093
## exang    -0.25774837  0.11573938  0.20675379 -0.43675708
## oldpeak  -0.57753682  0.22268232  0.21024413 -0.43069600
## slope     1.00000000 -0.08015521 -0.10476379  0.34587708
## ca       -0.08015521  1.00000000  0.15183213 -0.39172399
## thal     -0.10476379  0.15183213  1.00000000 -0.34402927
## target    0.34587708 -0.39172399 -0.34402927  1.00000000
```

Correlation matrix visualization

```
library(corrplot)

## corrplot 0.92 loaded

corrplot(cor(heart))
```



According to correlation matrix above, oldpeak and slope variable seems highly negatively correlated it can be cause multicollinearity problem but except these ones, there is no serious correlation problme between the covariates.

3-Conduct Logistic Regression using all covariates

```
#Change categorical covariates as factor
```

```
heart$sex<-as.factor(heart$sex)
heart$cp<-as.factor(heart$cp)
heart$fbs<-as.factor(heart$fbs)
heart$restecg<-as.factor(heart$restecg)
heart$exang<-as.factor(heart$exang)
heart$ca<-as.factor(heart$ca)
heart$thal<-as.factor(heart$thal)
```

```
#Conduct Logistic Regression
```

```

logit<- glm(target ~.,data=heart,family=binomial(link="logit"))
summary(logit)

##
## Call:
## glm(formula = target ~ ., family = binomial(link = "logit"),
##      data = heart)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.467022   3.292725  -0.446 0.655934
## age          0.024660   0.025286   0.975 0.329453
## sex1        -1.714248   0.553647  -3.096 0.001960 **
## cp1          0.972519   0.572697   1.698 0.089481 .
## cp2          1.943936   0.518371   3.750 0.000177 ***
## cp3          2.254634   0.697451   3.233 0.001226 **
## trestbps    -0.025698   0.011837  -2.171 0.029929 *
## chol        -0.005024   0.004197  -1.197 0.231291
## fbs1         0.447680   0.572197   0.782 0.433987
## restecg1     0.476621   0.393723   1.211 0.226069
## restecg2    -0.659531   2.893323  -0.228 0.819686
## thalach      0.023946   0.011604   2.064 0.039057 *
## exang1      -0.806692   0.446215  -1.808 0.070629 .
## oldpeak     -0.375557   0.245798  -1.528 0.126535
## slope        0.857564   0.386330   2.220 0.026434 *
## ca1         -2.225797   0.512356  -4.344 1.40e-05 ***
## ca2         -3.309532   0.775646  -4.267 1.98e-05 ***
## ca3         -2.200315   0.908507  -2.422 0.015440 *
## ca4          0.897954   1.709134   0.525 0.599316
## thal1        2.539354   2.255967   1.126 0.260328
## thal2        2.326324   2.154730   1.080 0.280304
## thal3        0.834980   2.159955   0.387 0.699072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 417.64  on 302  degrees of freedom
## Residual deviance: 184.10  on 281  degrees of freedom
## AIC: 228.1
##
## Number of Fisher Scoring iterations: 6

```

sex,cp,thalach,slope and ca covariates are significant at 0.05 alpha level.

#### 4-) Check Multicollinearity

```

car::vif(logit)

##              GVIF Df GVIF^(1/(2*Df))
## age          1.524008 1          1.234507

```

```
## sex      1.653682  1      1.285956
## cp       1.843714  3      1.107343
## trestbps 1.273765  1      1.128612
## chol     1.262295  1      1.123519
## fbs      1.204880  1      1.097670
## restecg  1.145217  2      1.034480
## thalach  1.528622  1      1.236375
## exang    1.166147  1      1.079883
## oldpeak  1.558793  1      1.248516
## slope    1.556834  1      1.247731
## ca       2.015097  4      1.091533
## thal     1.562172  3      1.077180
```

VIFs<-10 there is no multicollinearity

5-) Remove variables that are insignificant

*#chance categorical variables as factor*

```
heart$sex<-as.factor(heart$sex)
heart$cp<-as.factor(heart$cp)
heart$ca<-as.factor(heart$ca)
```

*#conduct logistic regression*

```
logit<-glm(target~sex+cp+thalach+slope+ca,data=heart,family
=binomial(link="logit"))
summary(logit)
```

```
##
## Call:
## glm(formula = target ~ sex + cp + thalach + slope + ca, family =
binomial(link = "logit"),
##      data = heart)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.913203   1.324900  -2.954 0.003141 **
## sex1        -1.846896   0.408759  -4.518 6.23e-06 ***
## cp1         1.704935   0.508913   3.350 0.000808 ***
## cp2         2.174338   0.434641   5.003 5.66e-07 ***
## cp3         2.199837   0.615908   3.572 0.000355 ***
## thalach     0.023863   0.009187   2.598 0.009389 **
## slope       1.191253   0.300244   3.968 7.26e-05 ***
## ca1        -2.000142   0.445878  -4.486 7.26e-06 ***
## ca2        -2.892959   0.623524  -4.640 3.49e-06 ***
## ca3        -2.125174   0.755818  -2.812 0.004927 **
## ca4         0.521773   1.432272   0.364 0.715636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 417.64 on 302 degrees of freedom
## Residual deviance: 222.18 on 292 degrees of freedom
## AIC: 244.18
##
## Number of Fisher Scoring iterations: 5
```

The estimated coefficient for sex1 = -1.846896 The estimated coefficient for cp1 = 1.70935  
The estimated coefficient for cp2 = 2.174338 The estimated coefficient for cp3 = 2.199837  
The estimated coefficient for thalach = 0.023863 The estimated coefficient for slope =  
1.191253 The estimated coefficient for ca1 = -2.000142 The estimated coefficient for ca2 = -  
2.892959 The estimated coefficient for ca3 = -2.125174 The estimated coefficient for ca4 =  
0.521773

Indicating for one unit increase in "thalach" the log odds of the target variable increase by  
0.023863 Indicating for one unit increase in "slope" the log odds of the target variable  
increase by 1.191253

P values are extremely small. This suggests strong evidence to reject the null hypothesis of  
no effect and supports the idea that both the intercept and covariates have a significant  
impact on the target variable.

The null deviance is 417.64, the residual deviance 222.18 AIC 244.18

6-) Check overdispersion problem

```
logit$deviance/logit$df.residual
## [1] 0.7608845
```

There is no overdispersion problem since it is smaller than 1.

7-) If there is overdispersion problem, fix this problem

*#there is no overdispersion problem*

8-) Obtain the confusion matrix and model evaluation metrics

```
library(caret)
## Loading required package: ggplot2
## Loading required package: lattice

prediction <- predict(logit, newdata=heart)
preds <- factor(ifelse(prediction <= 0.5, 0, 1))
confusionMatrix(as.factor(preds), as.factor(heart$target))

## Confusion Matrix and Statistics
##
```

```
##           Reference
## Prediction    0    1
##           0 113  35
##           1  25 130
##
##           Accuracy : 0.802
##           95% CI : (0.7526, 0.8454)
##           No Information Rate : 0.5446
##           P-Value [Acc > NIR] : <2e-16
##
##           Kappa : 0.6031
##
## Mcnemar's Test P-Value : 0.2453
##
##           Sensitivity : 0.8188
##           Specificity : 0.7879
##           Pos Pred Value : 0.7635
##           Neg Pred Value : 0.8387
##           Prevalence : 0.4554
##           Detection Rate : 0.3729
##           Detection Prevalence : 0.4884
##           Balanced Accuracy : 0.8034
##
##           'Positive' Class : 0
##
```

Accuracy is 0.80 and No Information Rate is 0.60. Accuracy is bigger than no information rate and confidence interval for accuracy does not include the no information rate, so we can say accuracy is satisfactory. Our model can categorize 80% of the data as less chance of heart attack and more chance to heart attack correctly.

Sensitivity is 0.81 which is the ability of a test to correctly classify an individual as “more chance of heart attack”.

Specificity is 0.78 which is the ability of test to correctly classify an individual as “less chance of heart attack”.

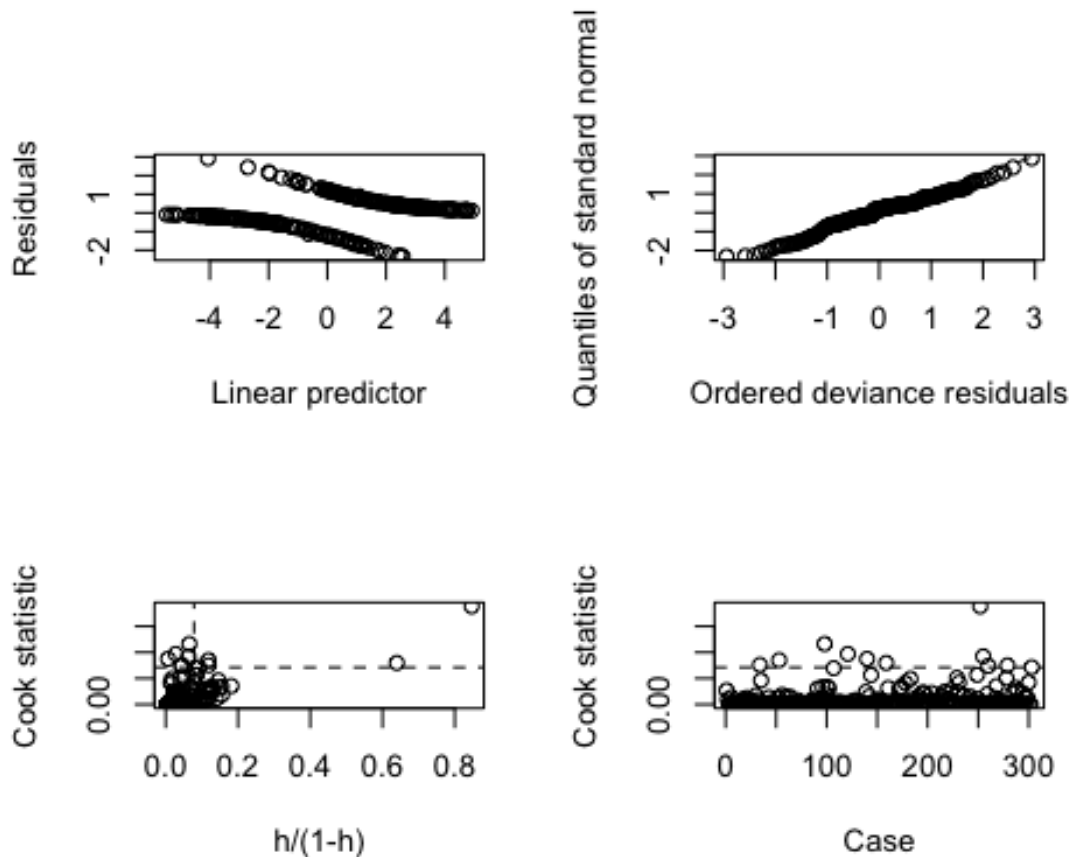
9-) Check diagnostics of the model (outliers, leverage points and influential.)

```
library(boot)

##
## Attaching package: 'boot'

## The following object is masked from 'package:lattice':
##
##      melanoma

glm.diag.plots(logit, glmdiag = glm.diag(logit))
```



Top

right plot is normal QQ plot of the standardized deviance residuals. For logistic regression there is no normality assumption bt we can see that most of the data point is on the dotted line which is the expected line if the standardized residuals are normmally distributed.

The bottom two panels are plots of the Cook statistics. 8 observations cause high influence on the model. These points are on the dotted lines.

10) Obtain deviance residuals

```
dev_res<-residuals(logit,type="deviance")
head(dev_res)

##           1           2           3           4           5           6
## 1.1693022 0.8342586 0.1659333 0.3831510 0.4250177 1.6374281
```

PART-2

Read the data

```
library(readxl)
birds<-read_excel("birds.xlsx")
head(birds)

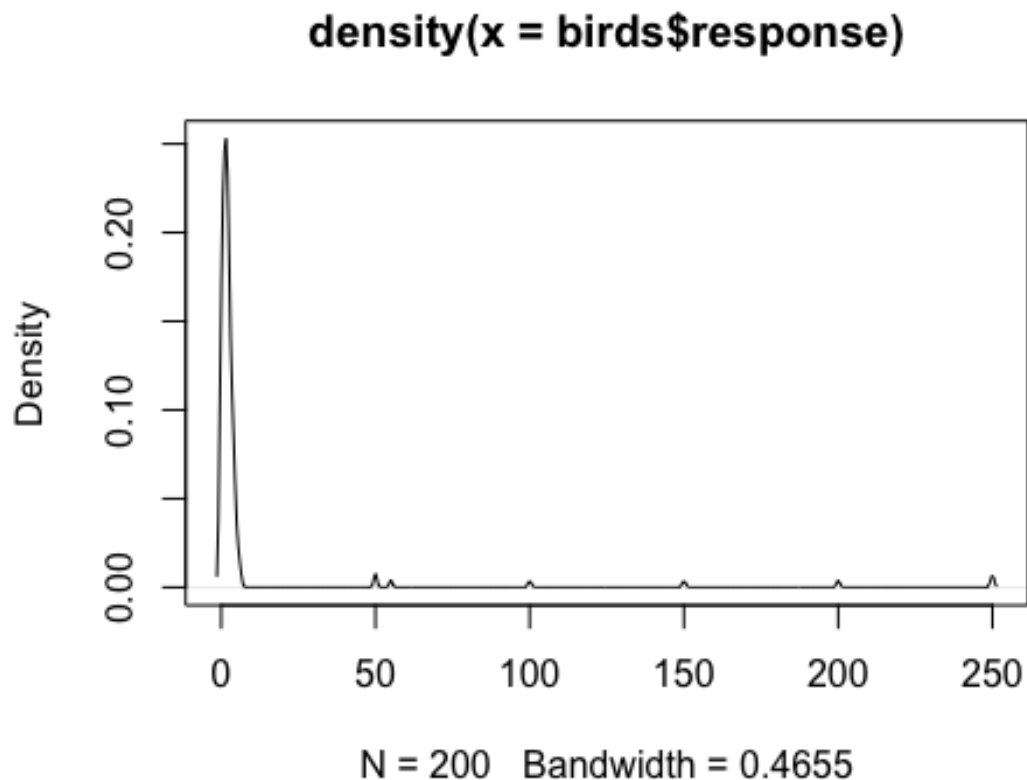
## # A tibble: 6 × 8
##   response covariate1 covariate2 covariate3 covariate4 covariate5
##   covariate6
```



```
##      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
<dbl>
## 1         2      0.916      2.20       96.3      1.07      0.356
0
## 2         6      0.801      1.31       41.6     -0.0273    -0.658
0
## 3         1     -0.937     -0.265      68.3     -0.0333     0.855
0
## 4         2     -1.40      0.543      98.6     -1.52      1.15
1
## 5         1      0.160     -0.414     134.      0.790      0.276
1
## 6         2     -0.274     -0.476      17.5     -0.211      0.144
0
## # i 1 more variable: covariate7 <dbl>
```

1-) Show density of response variable

```
plot(density(birds$response))
```



We can understand that density of response variable belongs the Poisson Dist.

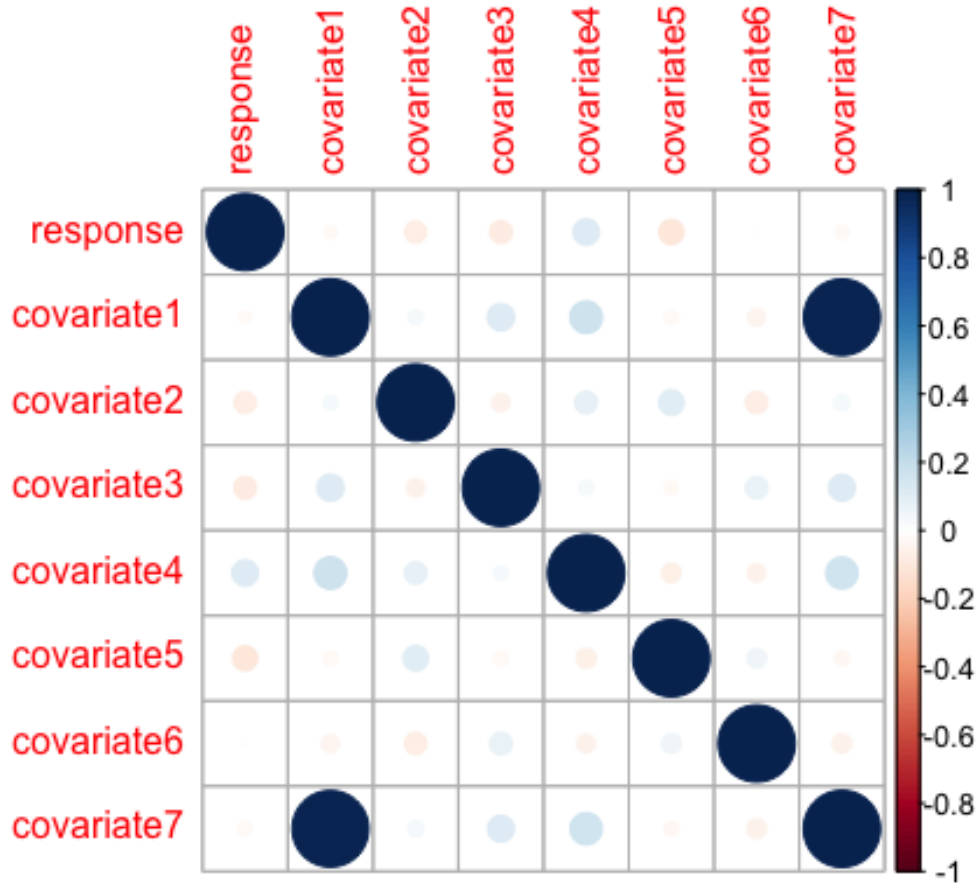
2-) Correlation Matrix

```
cor(birds)
```

```
##           response covariate1 covariate2 covariate3 covariate4
## response    1.000000000 -0.02220487 -0.07590014 -0.08056737  0.11224407
## covariate1 -0.022204873  1.000000000  0.03238508  0.11993587  0.17340935
## covariate2 -0.075900143  0.03238508  1.000000000 -0.05141456  0.08015345
## covariate3 -0.080567374  0.11993587 -0.05141456  1.000000000  0.03197187
## covariate4  0.112244069  0.17340935  0.08015345  0.03197187  1.000000000
## covariate5 -0.100346783 -0.02898979  0.10403897 -0.02826934 -0.06049721
## covariate6 -0.003283895 -0.04872897 -0.07781909  0.07979179 -0.05283958
## covariate7 -0.028173785  0.98934634  0.03871752  0.11598168  0.16961345
##           covariate5 covariate6 covariate7
## response   -0.10034678 -0.003283895 -0.02817379
## covariate1 -0.02898979 -0.048728965  0.98934634
## covariate2  0.10403897 -0.077819092  0.03871752
## covariate3 -0.02826934  0.079791786  0.11598168
## covariate4 -0.06049721 -0.052839578  0.16961345
## covariate5  1.000000000  0.059443934 -0.03143313
## covariate6  0.05944393  1.000000000 -0.05991752
## covariate7 -0.03143313 -0.059917523  1.000000000
```

Visualization of the correlation matrix

```
library(corrplot)
corrplot(cor(birds))
```



According to below graph, Covariate1 and Covariate7 seems highly positively correlated. It will cauess a multicollinearity problem.

3-) Conduct the regression model suitable for the distribution of the response

```
birds$covariate6<-as.factor(birds$covariate6)
POI.Model<-glm(response~.,family="poisson",data=birds)
summary(POI.Model)
```

```
##
## Call:
## glm(formula = response ~ ., family = "poisson", data = birds)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.3863745  0.0595529  40.072  < 2e-16 ***
## covariate1    0.4376381  0.1895739   2.309  0.02097 *
## covariate2   -0.3444030  0.0291843 -11.801  < 2e-16 ***
## covariate3   -0.0066848  0.0005425 -12.321  < 2e-16 ***
## covariate4    0.4473005  0.0258482  17.305  < 2e-16 ***
## covariate5   -0.3616321  0.0278400 -12.990  < 2e-16 ***
## covariate6    0.0369274  0.0540889   0.683  0.49479
## covariate7   -0.5147673  0.1906544  -2.700  0.00693 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 6171.8  on 199  degrees of freedom
## Residual deviance: 5227.6  on 192  degrees of freedom
## AIC: 5702.4
##
## Number of Fisher Scoring iterations: 7
```

4-) When you look at the model output, do you see a problem with the model coefficients, errors and residuals?

The coefficient for covariate6 is insignificant. Standard errors and p-values are extremely low. This suggests the overdispersion problem should be investigated. Deviance is so high probably there will be overdispersion problem in the model.

5-) Check multicollinearity

```
car::vif(POI.Model)

## covariate1 covariate2 covariate3 covariate4 covariate5 covariate6
## covariate7
## 55.560469  1.084689  1.050950  1.126533  1.051907  1.060592
## 56.528564
```

We can see that covariate1 and covariate7 has a high vif values (>10) so, there is a multicollinearity problem should be handled.

6-) Remove variables that are insignificant and that causes multicollinearity problem

```
POI.Model<-glm(response~covariate2+covariate3+covariate4+covariate5,
               family=poisson,data=birds)
summary(POI.Model)

##
## Call:
## glm(formula = response ~ covariate2 + covariate3 + covariate4 +
##      covariate5, family = poisson, data = birds)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.4251851  0.0552536   43.89  <2e-16 ***
## covariate2   -0.3676834  0.0286500  -12.83  <2e-16 ***
## covariate3   -0.0069837  0.0005247  -13.31  <2e-16 ***
## covariate4    0.4671604  0.0252858   18.48  <2e-16 ***
## covariate5   -0.3660722  0.0277435  -13.20  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##      Null deviance: 6171.8  on 199  degrees of freedom
## Residual deviance: 5244.2  on 195  degrees of freedom
## AIC: 5713
##
## Number of Fisher Scoring iterations: 7
```

Covariate1 and covariate7 cause a multicollinearity problem and covariate6 is insignificant so we discard these variables from the model. Standard Errors and z-values seems very low so that can be overdispersion problem

Let's check if there is an overdispersion problem

```
library(AER)

## Loading required package: car
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:boot':
##
##      logit
## Loading required package: lmtest
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
## Loading required package: sandwich
## Loading required package: survival
##
## Attaching package: 'survival'
## The following object is masked by_ '.GlobalEnv':
##
##      heart
## The following object is masked from 'package:boot':
##
##      aml
```

```
## The following object is masked from 'package:caret':
##
##      cluster

dispersiontest(POI.Model)

##
## Overdispersion test
##
## data: POI.Model
## z = 1.7259, p-value = 0.04218
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
## 108.8454
```

It is strongly bigger than the 1, so we can say that there is an overdispersion problem.

Let's fix the overdispersion problem

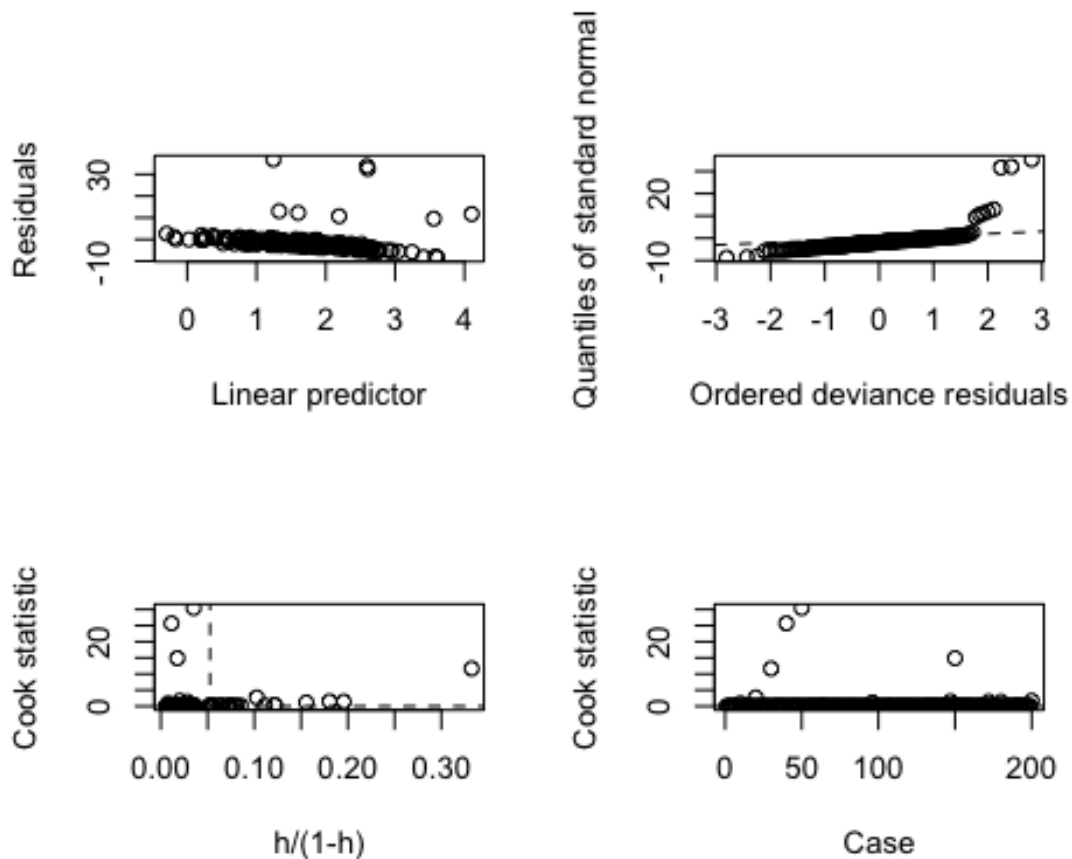
```
QPOI.Model<-glm(response~covariate2+covariate3+covariate4+covariate5,
  family=quasipoisson,data=birds)
summary(QPOI.Model)

##
## Call:
## glm(formula = response ~ covariate2 + covariate3 + covariate4 +
##      covariate5, family = quasipoisson, data = birds)
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.425185   0.584106   4.152 4.92e-05 ***
## covariate2   -0.367683   0.302869  -1.214   0.2262
## covariate3   -0.006984   0.005547  -1.259   0.2095
## covariate4    0.467160   0.267306   1.748   0.0821 .
## covariate5   -0.366072   0.293287  -1.248   0.2135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 111.7539)
##
##      Null deviance: 6171.8  on 199  degrees of freedom
## Residual deviance: 5244.2  on 195  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 7
```

Standard errors seem much more higher than now.

7-) Check the diagnostic of the model.

```
library(boot)
glm.diag.plots(QPOI.Model,glmdiag = glm.diag(POI.Model))
```



Up to 4-5 points need to be investigated since these data points are above the dotted line with respect to cook statistic graphs. In order to handle outlier problem we can conduct robust regression both by strengthening the model and keeping the outliers in the model. (40-50-150 observations are outliers)

8-) Conduct robust poisson regression

```
library(robust)

## Loading required package: fit.models

robust_regression<-
glmRob(response~covariate2+covariate3+covariate4+covariate5,
        family=poisson,data=birds,w = "huber", tuning.constant =
1.345)
summary(robust_regression)

##
## Call: glmRob(formula = response ~ covariate2 + covariate3 + covariate4 +
##   covariate5, family = poisson, data = birds, weights = "huber",
##   tuning.constant = 1.345)
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.15887 -0.77607 -0.06521  0.58696 43.94869
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.5777826   0.137653  4.1974 0.000027
## covariate2    0.1085251   0.058617  1.8514 0.064110
## covariate3    0.0004756   0.001215  0.3915 0.695434
## covariate4    0.0250326   0.056947  0.4396 0.660243
## covariate5   -0.0801080   0.056791 -1.4106 0.158370
##
## (Dispersion Parameter for poisson family taken to be 1 )
##
##      Null Deviance: 9448 on 199 degrees of freedom
##
## Residual Deviance: 7961.202 on 195 degrees of freedom
##
## Number of Iterations: 5
##
## Correlation of Coefficients:
##              (Intercept) covariate2 covariate3 covariate4
## covariate2  -0.09960
## covariate3  -0.90961    0.05275
## covariate4   0.02167   -0.09723   -0.02712
## covariate5  -0.01405   -0.10153    0.03398    0.06862
```

on average, for each unit increase in the covariate2, the predicted response rate increases by 0.1087388.(std=0.058605) on average, for each unit increase in the covariate3, the predicted response rate increases by 0.0004744.(std=0.001215) on average, for each unit increase in the covariate4, the predicted response rate increases by 0.0249564.(std=0.056929) on average, for each unit decrease in the covariate5, the predicted response rate increases by 0.0798389.(std=0.056797)

## 10-) Compare the models

```
library(car)
compareCoefs(QPOI.Model,robust_regression)

## Warning in compareCoefs(QPOI.Model, robust_regression): models to be
## compared
## are of different classes

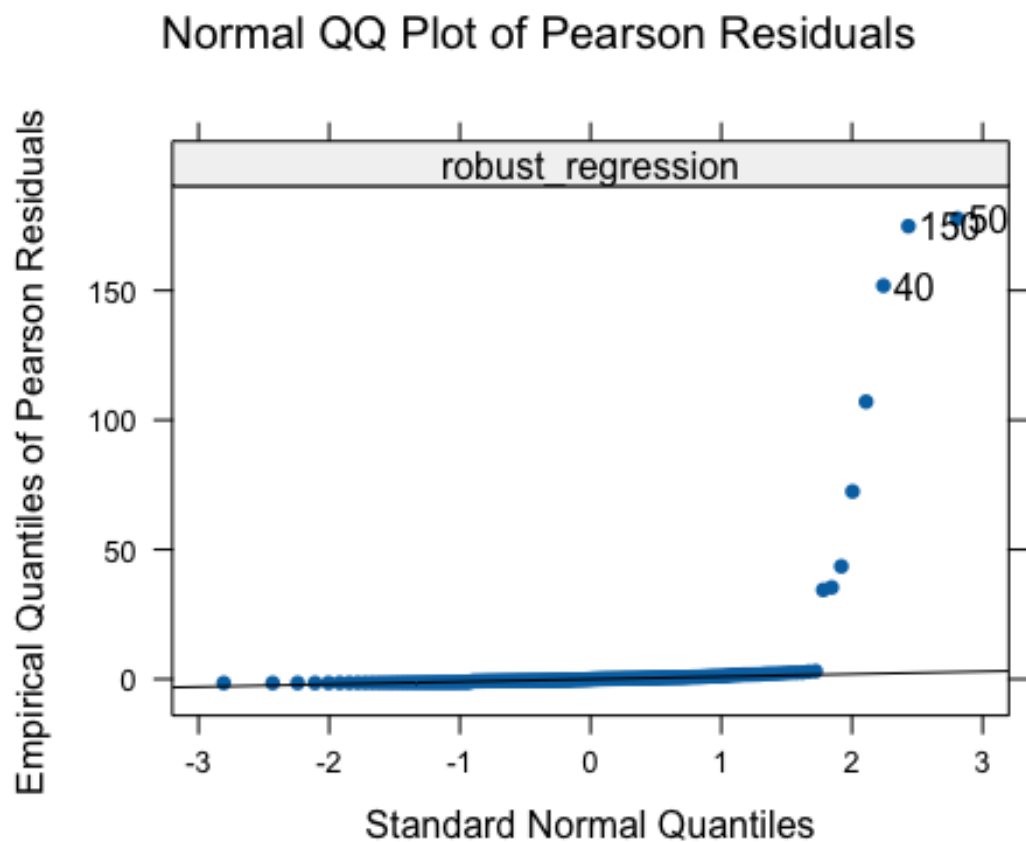
## Calls:
## 1: glm(formula = response ~ covariate2 + covariate3 + covariate4 +
##   covariate5, family = quasipoisson, data = birds)
## 2: glmRob(formula = response ~ covariate2 + covariate3 + covariate4 +
##   covariate5, family = poisson, data = birds, weights = "huber",
##   tuning.constant = 1.345)
##
##              Model 1   Model 2
```



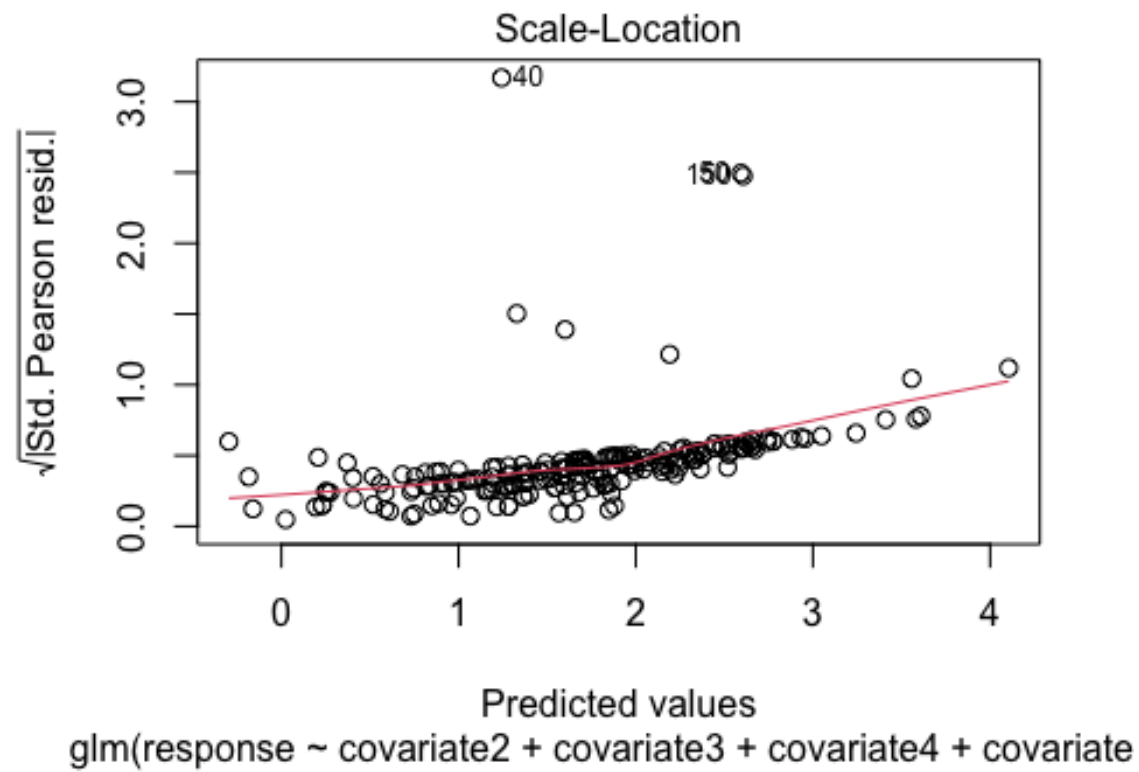
```
## (Intercept)      2.425      0.578
## SE              0.584      0.138
##
## covariate2      -0.3677     0.1085
## SE              0.3029     0.0586
##
## covariate3     -0.006984    0.000476
## SE              0.005547    0.001215
##
## covariate4       0.4672     0.0250
## SE              0.2673     0.0569
##
## covariate5      -0.3661    -0.0801
## SE              0.2933     0.0568
##
```

Standard errors are smaller than to the first model. Smaller standard errors indicate greater precision in the coefficient estimates. Model coefficients are different from each other.

```
#plot diagnostic plots for robust poisson regression model
plot(robust_regression, which=3)
```



```
#plot diagnostic plots for robust poisson regression model  
plot(QPOI.Model,which=3)
```



We can see that the robust regression provides us a better fitted line.