Inclass_Assignment_MLR

S20426

2025-03-19

1 Load the Dataset

```
library(AppliedPredictiveModeling)
```

```
## Warning: package 'AppliedPredictiveModeling' was built under R version 4.3.3
```

```
data(abalone)
attach(abalone)
```

2 Graphical Interpretation

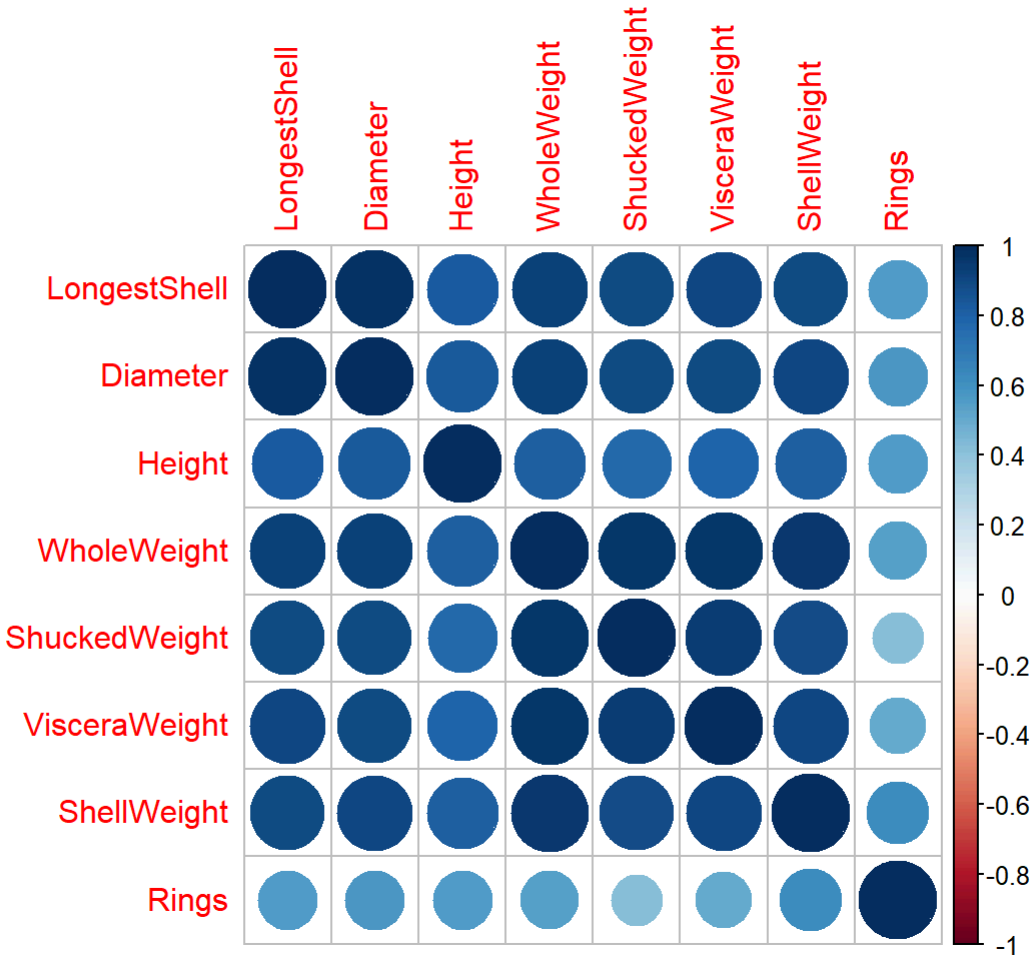
Draw a correlogram using corrplot() function

```
# install.packages("corrplot")
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.3
```

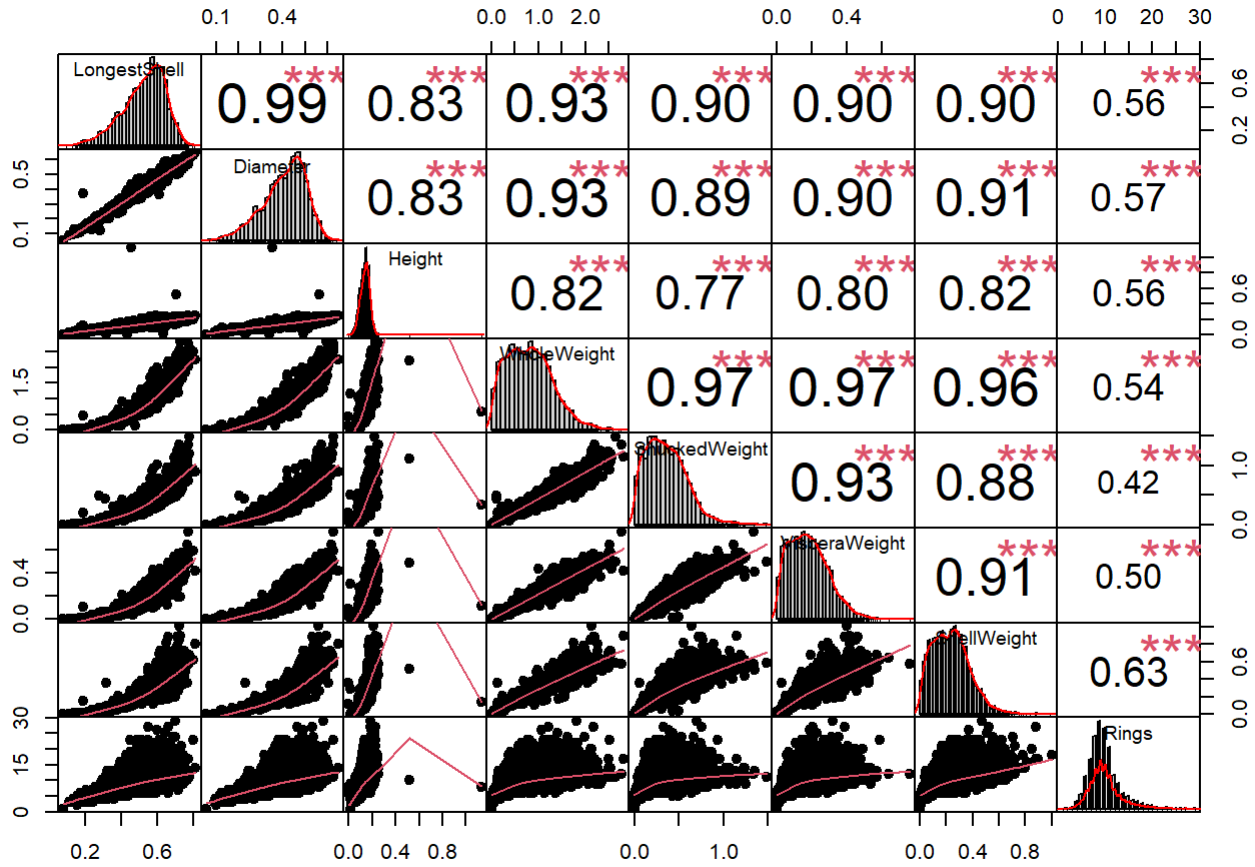
```
## corrplot 0.95 loaded
```

```
corrplot(cor(abalone[,-1]), tl.col = "red")
```



Draw scatter plots using chart.Correlation():

```
# install.packages("PerformanceAnalytics")
# library("PerformanceAnalytics")
chart.Correlation(abalone[,-1], histogram=TRUE, pch=19)
```



02. Parameter Estimation

Check For Missing Values

```
sum(is.na(abalone))
```

```
## [1] 0
```

```
model1 <- lm(Rings ~ as.factor(Type) + LongestShell + Diameter +
             WholeWeight + ShuckedWeight + VisceraWeight + ShellWeight,
             data = abalone)
model1
```

```
##
## Call:
## lm(formula = Rings ~ as.factor(Type) + LongestShell + Diameter +
##     WholeWeight + ShuckedWeight + VisceraWeight + ShellWeight,
##     data = abalone)
##
## Coefficients:
##       (Intercept)  as.factor(Type)I  as.factor(Type)M      LongestShell
##           4.24786          -0.87919           0.04437          -0.05352
##          Diameter       WholeWeight     ShuckedWeight     VisceraWeight
##          12.73487           9.06362         -19.95631         -10.18938
##       ShellWeight
##           9.57238
```

Check Model Summary

```
summary(model1)
```

```
##
## Call:
## lm(formula = Rings ~ as.factor(Type) + LongestShell + Diameter +
##     WholeWeight + ShuckedWeight + VisceraWeight + ShellWeight,
##     data = abalone)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1813 -1.3130 -0.3437  0.8690 13.7654
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.24786    0.28883  14.707  < 2e-16 ***
## as.factor(Type)I  -0.87919    0.10269  -8.562  < 2e-16 ***
## as.factor(Type)M   0.04437    0.08380   0.529    0.597
## LongestShell      -0.05352    1.81860  -0.029    0.977
## Diameter          12.73487    2.22738   5.717 1.16e-08 ***
## WholeWeight        9.06362    0.72947  12.425  < 2e-16 ***
## ShuckedWeight    -19.95631    0.82169 -24.287  < 2e-16 ***
## VisceraWeight    -10.18938    1.29997  -7.838 5.76e-15 ***
## ShellWeight        9.57238    1.12490   8.510  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.207 on 4168 degrees of freedom
## Multiple R-squared:  0.5324, Adjusted R-squared:  0.5315
## F-statistic: 593.3 on 8 and 4168 DF,  p-value: < 2.2e-16
```

Create 2 columns For Type

```
abalone$Type.I=ifelse(abalone$Type=="I",1,0)
abalone$Type.M=ifelse(abalone$Type=="M",1,0)
head(abalone)
```

```
##   Type LongestShell Diameter Height WholeWeight ShuckedWeight VisceraWeight
## 1    M        0.455    0.365  0.095      0.5140        0.2245        0.1010
## 2    M        0.350    0.265  0.090      0.2255        0.0995        0.0485
## 3    F        0.530    0.420  0.135      0.6770        0.2565        0.1415
## 4    M        0.440    0.365  0.125      0.5160        0.2155        0.1140
## 5    I        0.330    0.255  0.080      0.2050        0.0895        0.0395
## 6    I        0.425    0.300  0.095      0.3515        0.1410        0.0775
##   ShellWeight Rings Type.I Type.M
## 1       0.150    15      0      1
## 2       0.070     7      0      1
## 3       0.210     9      0      0
## 4       0.155    10      0      1
## 5       0.055     7      1      0
## 6       0.120     8      1      0
```

```
full_model <-  lm(Rings~Type.I + Type.M+ LongestShell+ Diameter+ WholeWeight+ ShuckedWeight+ VisceraWeight+ ShellWeight,data=abalone)
summary(full_model)
```

```
##
## Call:
## lm(formula = Rings ~ Type.I + Type.M + LongestShell + Diameter +
##     WholeWeight + ShuckedWeight + VisceraWeight + ShellWeight,
##     data = abalone)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1813 -1.3130 -0.3437  0.8690 13.7654
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.24786    0.28883  14.707  < 2e-16 ***
## Type.I         -0.87919    0.10269  -8.562  < 2e-16 ***
## Type.M          0.04437    0.08380   0.529    0.597
## LongestShell   -0.05352    1.81860  -0.029    0.977
## Diameter       12.73487    2.22738   5.717 1.16e-08 ***
## WholeWeight     9.06362    0.72947  12.425  < 2e-16 ***
## ShuckedWeight -19.95631    0.82169 -24.287  < 2e-16 ***
## VisceraWeight -10.18938    1.29997  -7.838 5.76e-15 ***
## ShellWeight     9.57238    1.12490   8.510  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.207 on 4168 degrees of freedom
## Multiple R-squared:  0.5324, Adjusted R-squared:  0.5315
## F-statistic: 593.3 on 8 and 4168 DF,  p-value: < 2.2e-16
```

03 Test on Parameters

```
library(MASS)
step_model <- stepAIC(full_model, direction = "both")
```

```
## Start:  AIC=6621.33
## Rings ~ Type.I + Type.M + LongestShell + Diameter + WholeWeight +
##     ShuckedWeight + VisceraWeight + ShellWeight
##
##                 Df Sum of Sq   RSS    AIC
## - LongestShell   1      0.00 20297 6619.3
## - Type.M         1      1.37 20298 6619.6
## <none>                       20297 6621.3
## - Diameter       1    159.19 20456 6652.0
## - VisceraWeight  1    299.18 20596 6680.5
## - ShellWeight    1    352.63 20650 6691.3
## - Type.I         1    356.97 20654 6692.2
## - WholeWeight    1    751.79 21049 6771.2
## - ShuckedWeight  1   2872.42 23169 7172.2
##
## Step:  AIC=6619.33
## Rings ~ Type.I + Type.M + Diameter + WholeWeight + ShuckedWeight +
##     VisceraWeight + ShellWeight
##
##                 Df Sum of Sq   RSS    AIC
## - Type.M         1      1.36 20298 6617.6
## <none>                       20297 6619.3
## + LongestShell   1      0.00 20297 6621.3
## - VisceraWeight  1    302.18 20599 6679.1
## - ShellWeight    1    352.98 20650 6689.3
## - Type.I         1    358.32 20655 6690.4
## - WholeWeight    1    751.92 21049 6769.3
## - Diameter       1    871.45 21168 6792.9
## - ShuckedWeight  1   2889.21 23186 7173.2
##
## Step:  AIC=6617.61
## Rings ~ Type.I + Diameter + WholeWeight + ShuckedWeight + VisceraWeight +
##     ShellWeight
##
##                 Df Sum of Sq   RSS    AIC
## <none>                       20298 6617.6
## + Type.M         1      1.36 20297 6619.3
## + LongestShell   1      0.00 20298 6619.6
## - VisceraWeight  1    304.35 20603 6677.8
## - ShellWeight    1    352.82 20651 6687.6
## - Type.I         1    496.66 20795 6716.6
## - WholeWeight    1    751.52 21050 6767.5
## - Diameter       1    870.30 21169 6791.0
## - ShuckedWeight  1   2892.15 23191 7172.0
```

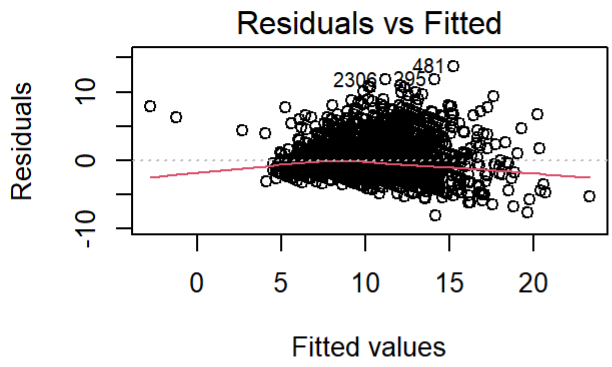
```
summary(step_model)
```

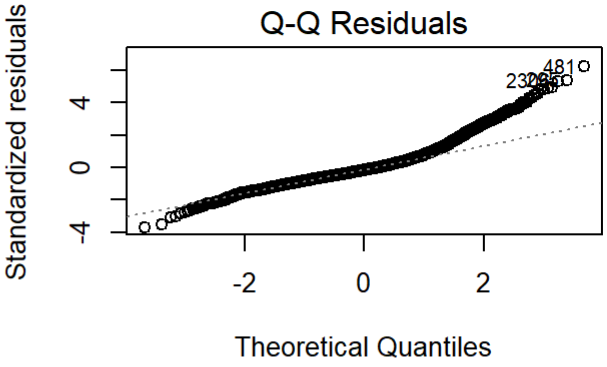
```
##
## Call:
## lm(formula = Rings ~ Type.I + Diameter + WholeWeight + ShuckedWeight +
##     VisceraWeight + ShellWeight, data = abalone)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1597 -1.3126 -0.3417  0.8697 13.7453
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.27930    0.26379  16.222  < 2e-16 ***
## Type.I         -0.90566    0.08966 -10.101  < 2e-16 ***
## Diameter       12.64787    0.94590  13.371  < 2e-16 ***
## WholeWeight     9.06125    0.72926  12.425  < 2e-16 ***
## ShuckedWeight -19.93199    0.81772 -24.375  < 2e-16 ***
## VisceraWeight -10.22104    1.29262  -7.907 3.34e-15 ***
## ShellWeight     9.57105    1.12421   8.514  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.206 on 4170 degrees of freedom
## Multiple R-squared:  0.5324, Adjusted R-squared:  0.5317
## F-statistic: 791.3 on 6 and 4170 DF,  p-value: < 2.2e-16
```

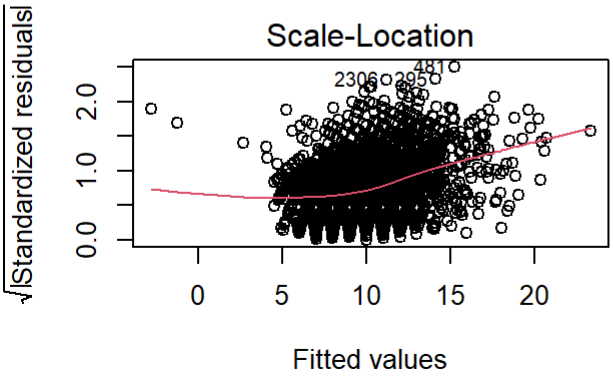
04 Analysis of Variance

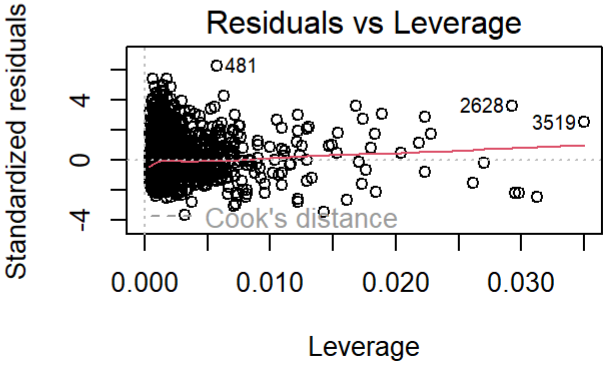
check model assumption

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



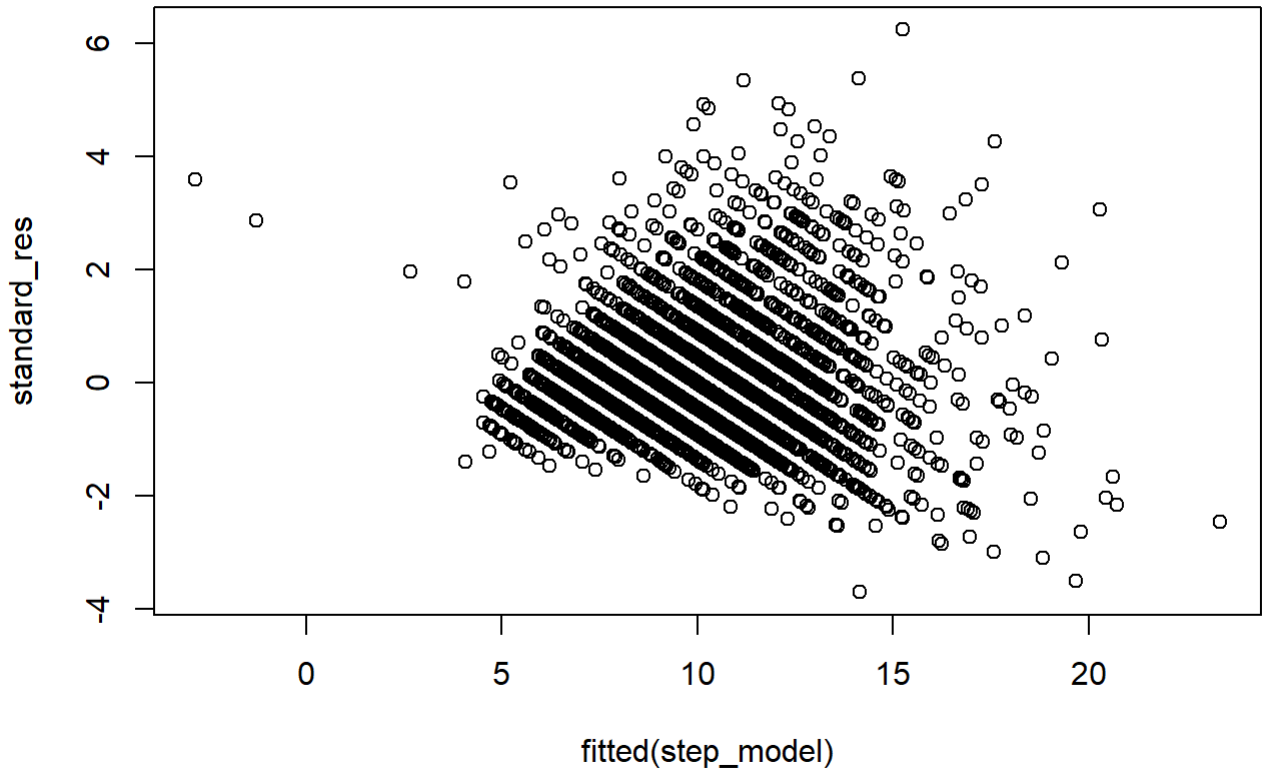






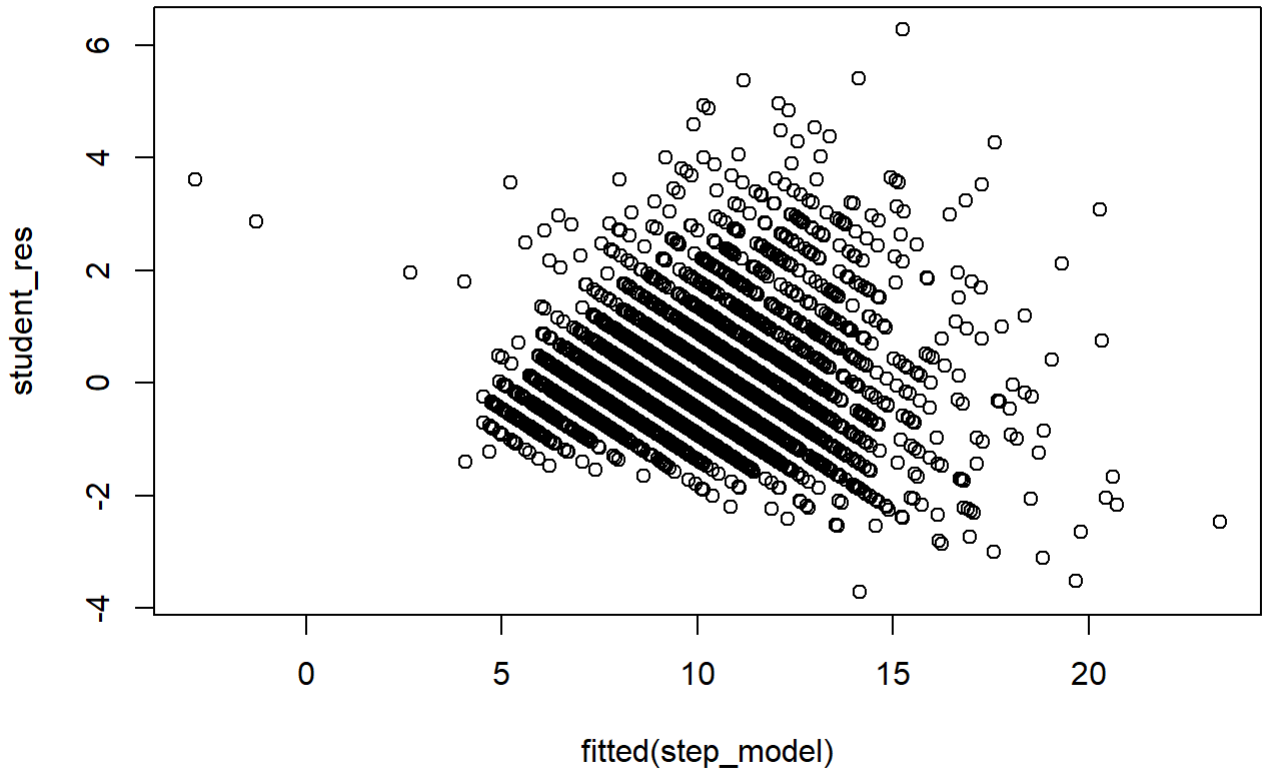
Identifying Influential Outliers

```
standard_res=rstandard(step_model)
plot(fitted(step_model), standard_res)
```



Studentized Residuals

```
student_res=rstudent(step_model)
plot(fitted(step_model), student_res)
```



Variance inflation factor

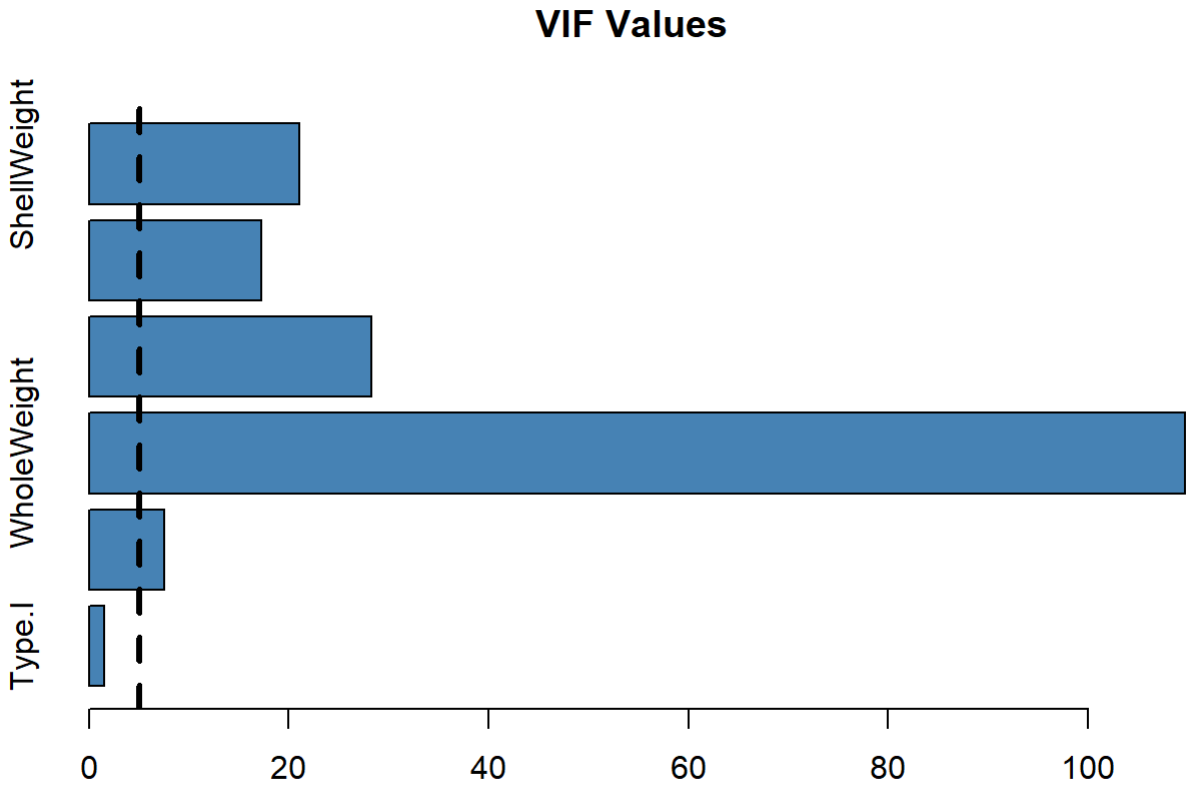
```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

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

```
## Warning: package 'carData' was built under R version 4.3.3
```

```
vif_values <- vif(step_model)
barplot(vif_values, main = "VIF Values", horiz = TRUE, col = "steelblue")
abline(v = 5, lwd = 3, lty = 2)
```



05 Interpretation and Prediction

Testing the multicollinearity among the predictor variables

```
library(mctest)
omcdiag(mod=step_model)
```

```
##
## Call:
## omcdiag(mod = step_model)
##
##
## Overall Multicollinearity Diagnostics
##
##                        MC Results detection
## Determinant |X'X|:         0.0000         1
## Farrar Chi-Square:     45614.2311         1
## Red Indicator:             0.8186         1
## Sum of Lambda Inverse:   185.2768         1
## Theil's Method:            2.3907         1
## Condition Number:         57.9317         1
##
## 1 --> COLLINEARITY is detected by the test
## 0 --> COLLINEARITY is not detected by the test
```

```
# OR

#imcdiag(mod=step.model)
```

model summary()

```
summary(step_model)
```

```
##
## Call:
## lm(formula = Rings ~ Type.I + Diameter + WholeWeight + ShuckedWeight +
##     VisceraWeight + ShellWeight, data = abalone)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1597 -1.3126 -0.3417  0.8697 13.7453
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.27930    0.26379  16.222  < 2e-16 ***
## Type.I         -0.90566    0.08966 -10.101  < 2e-16 ***
## Diameter       12.64787    0.94590  13.371  < 2e-16 ***
## WholeWeight     9.06125    0.72926  12.425  < 2e-16 ***
## ShuckedWeight -19.93199    0.81772 -24.375  < 2e-16 ***
## VisceraWeight -10.22104    1.29262  -7.907 3.34e-15 ***
## ShellWeight     9.57105    1.12421   8.514  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.206 on 4170 degrees of freedom
## Multiple R-squared:  0.5324, Adjusted R-squared:  0.5317
## F-statistic: 791.3 on 6 and 4170 DF,  p-value: < 2.2e-16
```