Body\_fat\_Prediction

Naveen

5/26/2022

rm(list = ls())  
## Read data from a .csv file ##  
women <- read.csv("C:/Users/morla/Downloads/dataset\_RG.csv")  
Dat <- women[,c(1,2,3,4,5,6,7)]

summary(Dat)

## WomenBodyFat Age Height WaistCircumference  
## Min. :0.1300 Min. :13.00 Min. :150.0 Min. : 57.00   
## 1st Qu.:0.2900 1st Qu.:21.00 1st Qu.:160.0 1st Qu.: 73.00   
## Median :0.3250 Median :28.00 Median :163.0 Median : 78.00   
## Mean :0.3302 Mean :29.88 Mean :163.5 Mean : 79.84   
## 3rd Qu.:0.3825 3rd Qu.:38.00 3rd Qu.:168.0 3rd Qu.: 87.00   
## Max. :0.5600 Max. :62.00 Max. :176.0 Max. :105.00   
## HipCircumference ChestCircumference UpperarmCircumference  
## Min. : 84.0 Min. : 72.00 Min. :21.00   
## 1st Qu.: 96.0 1st Qu.: 86.00 1st Qu.:27.00   
## Median :102.0 Median : 92.50 Median :29.00   
## Mean :102.3 Mean : 93.46 Mean :29.46   
## 3rd Qu.:108.0 3rd Qu.: 99.25 3rd Qu.:31.25   
## Max. :133.0 Max. :121.00 Max. :40.00

#norm\_scale <- predict(process, as.data.frame(Dat))  
data1 <- as.data.frame(scale(Dat))  
  
## Fit MLR model  
mlr <- lm(WomenBodyFat ~ . , data = Dat)  
summary(mlr)

##   
## Call:  
## lm(formula = WomenBodyFat ~ ., data = Dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.062404 -0.007196 -0.000059 0.006747 0.112077   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0039905 0.0498647 0.080 0.9363   
## Age 0.0001805 0.0001717 1.051 0.2950   
## Height -0.0028361 0.0003054 -9.286 5.02e-16 \*\*\*  
## WaistCircumference 0.0023879 0.0004856 4.917 2.62e-06 \*\*\*  
## HipCircumference 0.0046686 0.0003609 12.936 < 2e-16 \*\*\*  
## ChestCircumference 0.0010471 0.0004786 2.188 0.0305 \*   
## UpperarmCircumference 0.0006234 0.0009176 0.679 0.4981   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0195 on 129 degrees of freedom  
## Multiple R-squared: 0.9444, Adjusted R-squared: 0.9418   
## F-statistic: 365 on 6 and 129 DF, p-value: < 2.2e-16

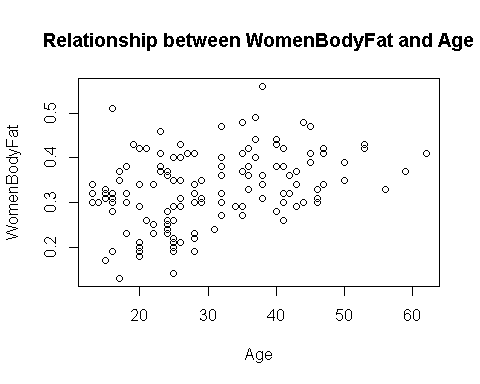
mlr5 <- lm(WomenBodyFat ~ . , data = data1)  
summary(mlr5)

##   
## Call:  
## lm(formula = WomenBodyFat ~ ., data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.77210 -0.08903 -0.00073 0.08348 1.38667   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.817e-17 2.069e-02 0.000 1.0000   
## Age 2.495e-02 2.373e-02 1.051 0.2950   
## Height -2.031e-01 2.187e-02 -9.286 5.02e-16 \*\*\*  
## WaistCircumference 2.929e-01 5.956e-02 4.917 2.62e-06 \*\*\*  
## HipCircumference 5.605e-01 4.333e-02 12.936 < 2e-16 \*\*\*  
## ChestCircumference 1.297e-01 5.930e-02 2.188 0.0305 \*   
## UpperarmCircumference 2.910e-02 4.284e-02 0.679 0.4981   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2413 on 129 degrees of freedom  
## Multiple R-squared: 0.9444, Adjusted R-squared: 0.9418   
## F-statistic: 365 on 6 and 129 DF, p-value: < 2.2e-16

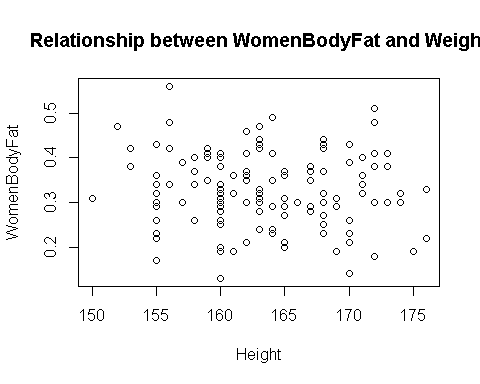
## 95% Confidence interval for model parameters ##  
confint(mlr, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) -0.0946680223 0.1026490455  
## Age -0.0001591942 0.0005202857  
## Height -0.0034404359 -0.0022318536  
## WaistCircumference 0.0014270470 0.0033487352  
## HipCircumference 0.0039545603 0.0053826017  
## ChestCircumference 0.0001001777 0.0019940727  
## UpperarmCircumference -0.0011922024 0.0024389608

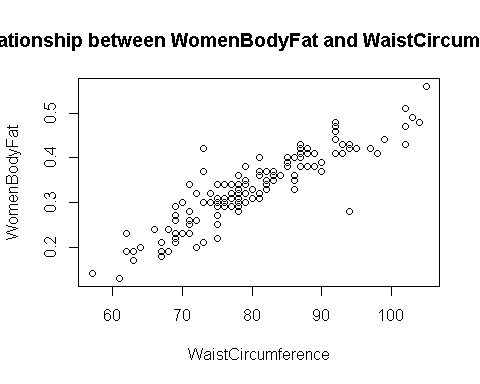
## Scatter plot ##  
plot(Dat$Age, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and Age",  
 xlab = "Age", ylab = "WomenBodyFat")



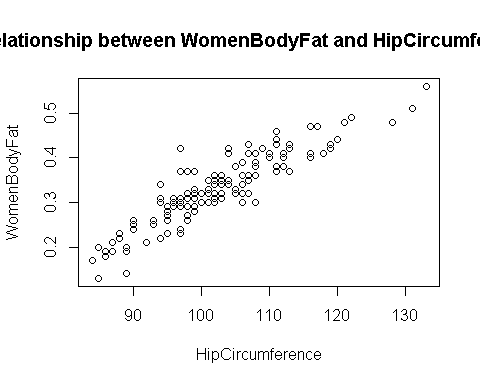
plot(Dat$Height, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and Weight",  
 xlab = "Height", ylab = "WomenBodyFat")



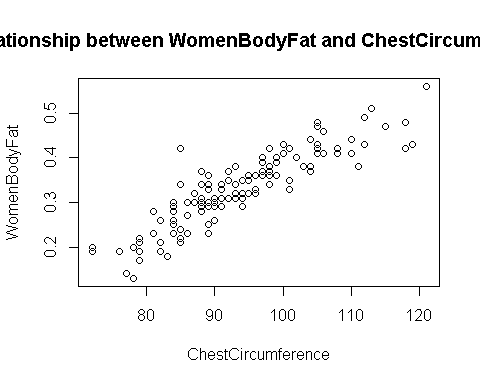
plot(Dat$WaistCircumference, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and WaistCircumference",  
 xlab = "WaistCircumference", ylab = "WomenBodyFat")



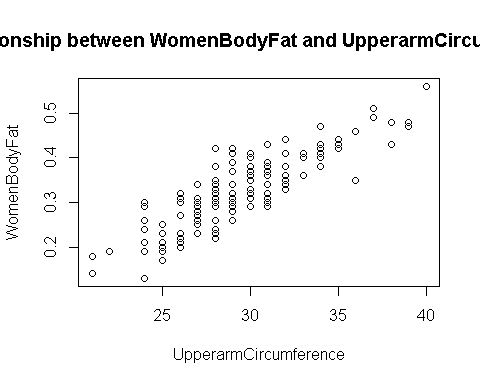
plot(Dat$HipCircumference, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and HipCircumference",  
 xlab = "HipCircumference", ylab = "WomenBodyFat")



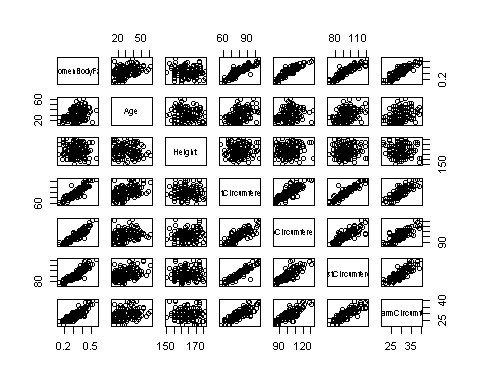
plot(Dat$ChestCircumference, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and ChestCircumference",  
 xlab = "ChestCircumference", ylab = "WomenBodyFat")



plot(Dat$UpperarmCircumference, Dat$WomenBodyFat, main = "Relationship between WomenBodyFat and UpperarmCircumference",  
 xlab = "UpperarmCircumference", ylab = "WomenBodyFat")



pairs(Dat)



## Type I ANOVA ##  
res.aov.IA <- aov(WomenBodyFat ~ ., data = Dat)  
summary(res.aov.IA)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 0.1169 0.1169 307.518 < 2e-16 \*\*\*  
## Height 1 0.0092 0.0092 24.226 2.57e-06 \*\*\*  
## WaistCircumference 1 0.6044 0.6044 1589.501 < 2e-16 \*\*\*  
## HipCircumference 1 0.1000 0.1000 263.055 < 2e-16 \*\*\*  
## ChestCircumference 1 0.0021 0.0021 5.491 0.0206 \*   
## UpperarmCircumference 1 0.0002 0.0002 0.461 0.4981   
## Residuals 129 0.0491 0.0004   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

res.aov.IB <- aov(WomenBodyFat ~ ., data = Dat)  
summary(res.aov.IB)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Age 1 0.1169 0.1169 307.518 < 2e-16 \*\*\*  
## Height 1 0.0092 0.0092 24.226 2.57e-06 \*\*\*  
## WaistCircumference 1 0.6044 0.6044 1589.501 < 2e-16 \*\*\*  
## HipCircumference 1 0.1000 0.1000 263.055 < 2e-16 \*\*\*  
## ChestCircumference 1 0.0021 0.0021 5.491 0.0206 \*   
## UpperarmCircumference 1 0.0002 0.0002 0.461 0.4981   
## Residuals 129 0.0491 0.0004   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Type II and III ANOVA ##  
library(car)

## Loading required package: carData

Anova(mlr, type=2)

## Anova Table (Type II tests)  
##   
## Response: WomenBodyFat  
## Sum Sq Df F value Pr(>F)   
## Age 0.000420 1 1.1055 0.29502   
## Height 0.032788 1 86.2278 5.020e-16 \*\*\*  
## WaistCircumference 0.009193 1 24.1772 2.621e-06 \*\*\*  
## HipCircumference 0.063635 1 167.3520 < 2.2e-16 \*\*\*  
## ChestCircumference 0.001820 1 4.7866 0.03048 \*   
## UpperarmCircumference 0.000175 1 0.4615 0.49815   
## Residuals 0.049052 129   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Anova(mlr, type=3)

## Anova Table (Type III tests)  
##   
## Response: WomenBodyFat  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 0.000002 1 0.0064 0.93634   
## Age 0.000420 1 1.1055 0.29502   
## Height 0.032788 1 86.2278 5.020e-16 \*\*\*  
## WaistCircumference 0.009193 1 24.1772 2.621e-06 \*\*\*  
## HipCircumference 0.063635 1 167.3520 < 2.2e-16 \*\*\*  
## ChestCircumference 0.001820 1 4.7866 0.03048 \*   
## UpperarmCircumference 0.000175 1 0.4615 0.49815   
## Residuals 0.049052 129   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

############################  
### Removing high cor regrossor ###  
############################  
  
library(car)  
vif(mlr)

## Age Height WaistCircumference   
## 1.305736 1.109352 8.227047   
## HipCircumference ChestCircumference UpperarmCircumference   
## 4.354315 8.155432 4.255732

cor(Dat)

## WomenBodyFat Age Height WaistCircumference  
## WomenBodyFat 1.0000000 0.36413427 -0.11345341 0.89865035  
## Age 0.3641343 1.00000000 -0.03102759 0.44337051  
## Height -0.1134534 -0.03102759 1.00000000 0.02629859  
## WaistCircumference 0.8986503 0.44337051 0.02629859 1.00000000  
## HipCircumference 0.9091104 0.25180713 0.13329028 0.81356458  
## ChestCircumference 0.8838725 0.40320626 0.07143423 0.92534829  
## UpperarmCircumference 0.8530102 0.32938842 -0.04379011 0.82455232  
## HipCircumference ChestCircumference UpperarmCircumference  
## WomenBodyFat 0.9091104 0.88387249 0.85301018  
## Age 0.2518071 0.40320626 0.32938842  
## Height 0.1332903 0.07143423 -0.04379011  
## WaistCircumference 0.8135646 0.92534829 0.82455232  
## HipCircumference 1.0000000 0.82711585 0.81793094  
## ChestCircumference 0.8271158 1.00000000 0.82357949  
## UpperarmCircumference 0.8179309 0.82357949 1.00000000

library(StepReg)  
stepwise(Dat, y="WomenBodyFat", selection = "forward", select = "adjRsq")

## $process  
## Step EffectEntered EffectNumber Select  
## 1 0 intercept 1 0.0000000  
## 2 1 HipCircumference 2 0.8251867  
## 3 2 WaistCircumference 3 0.8997966  
## 4 3 Height 4 0.9398642  
## 5 4 ChestCircumference 5 0.9419745  
## 6 5 Age 6 0.9420327  
##   
## $variate  
## [1] "intercept" "HipCircumference" "WaistCircumference"  
## [4] "Height" "ChestCircumference" "Age"

Dat<-Dat[,c(1,2,3,4,5)]  
## Fit MLR model  
mlr2 <- lm(WomenBodyFat ~ . , data = Dat)  
summary(mlr2)

##   
## Call:  
## lm(formula = WomenBodyFat ~ ., data = Dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.080091 -0.007232 -0.000112 0.006759 0.111017   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0139131 0.0498106 0.279 0.780   
## Age 0.0002042 0.0001740 1.173 0.243   
## Height -0.0028416 0.0003000 -9.471 <2e-16 \*\*\*  
## WaistCircumference 0.0032698 0.0003289 9.941 <2e-16 \*\*\*  
## HipCircumference 0.0050213 0.0003142 15.980 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01979 on 131 degrees of freedom  
## Multiple R-squared: 0.9418, Adjusted R-squared: 0.94   
## F-statistic: 530.1 on 4 and 131 DF, p-value: < 2.2e-16

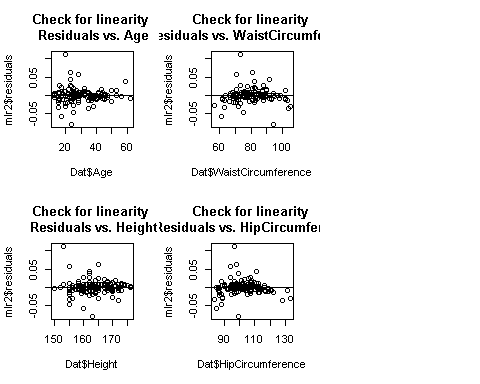
vif(mlr2)

## Age Height WaistCircumference HipCircumference   
## 1.301575 1.039263 3.663524 3.204528

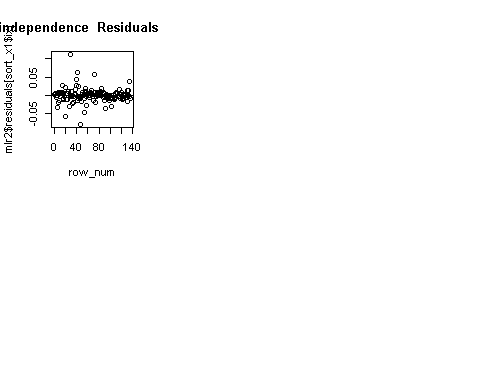
library(car)  
Anova(mlr2, type=2)

## Anova Table (Type II tests)  
##   
## Response: WomenBodyFat  
## Sum Sq Df F value Pr(>F)   
## Age 0.000539 1 1.377 0.2427   
## Height 0.035135 1 89.695 <2e-16 \*\*\*  
## WaistCircumference 0.038710 1 98.821 <2e-16 \*\*\*  
## HipCircumference 0.100026 1 255.351 <2e-16 \*\*\*  
## Residuals 0.051315 131   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

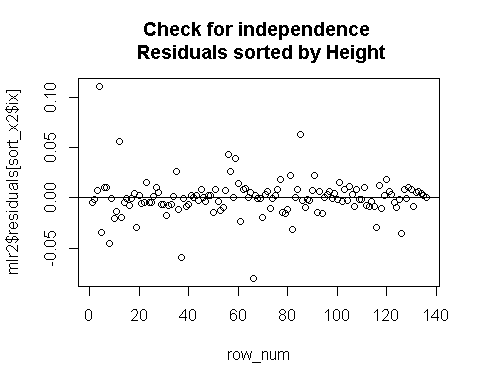
par(mfcol = c(2,3))  
plot(Dat$Age, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. Age")  
abline(h=0)  
plot(Dat$Height, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. Height")  
abline(h=0)  
plot(Dat$WaistCircumference, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. WaistCircumference")  
abline(h=0)  
plot(Dat$HipCircumference, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. HipCircumference")  
abline(h=0)



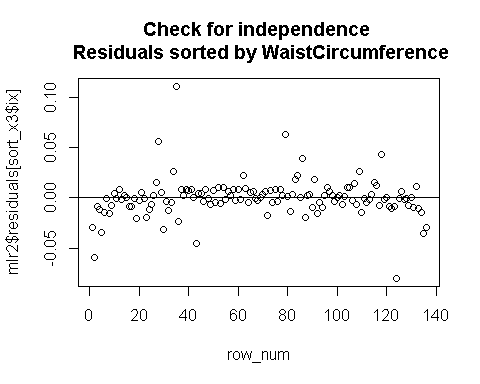
## Check for independence of random error  
par(mfcol = c(2,3))  
row\_num <- c(1:nrow(Dat))  
sort\_x1 <- sort(Dat$Age, index.return=TRUE)  
plot(row\_num, mlr2$residuals[sort\_x1$ix],   
 main = "Check for independence Residuals sorted by Age")  
abline(h=0)



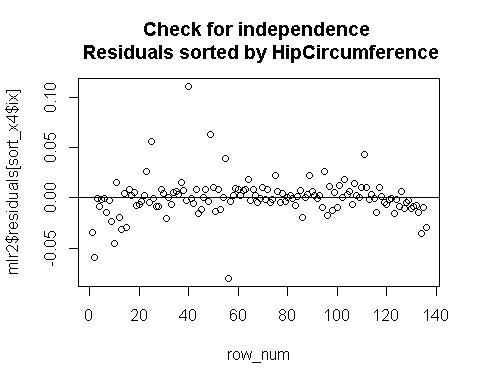
sort\_x2 <- sort(Dat$Height, index.return=TRUE)  
plot(row\_num, mlr2$residuals[sort\_x2$ix],   
 main = "Check for independence \n Residuals sorted by Height")  
abline(h=0)



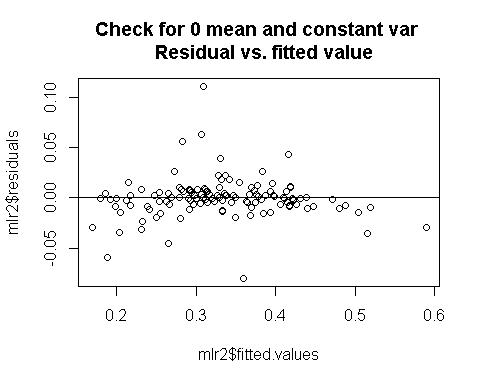
sort\_x3 <- sort(Dat$WaistCircumference, index.return=TRUE)  
plot(row\_num, mlr2$residuals[sort\_x3$ix],   
 main = "Check for independence \n Residuals sorted by WaistCircumference")  
abline(h=0)



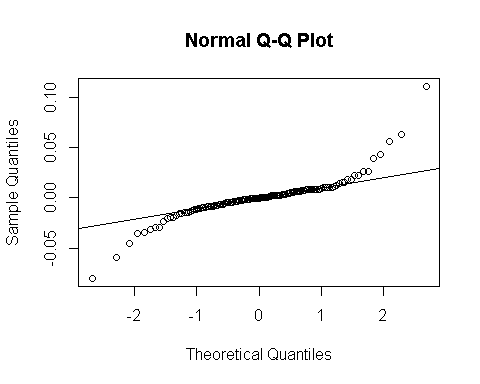
sort\_x4 <- sort(Dat$HipCircumference, index.return=TRUE)  
plot(row\_num, mlr2$residuals[sort\_x4$ix],   
 main = "Check for independence \n Residuals sorted by HipCircumference")  
abline(h=0)



## Check for zero mean and constant variance of random error  
plot(mlr2$fitted.values, mlr2$residuals,   
 main = "Check for 0 mean and constant var \n Residual vs. fitted value")  
abline(h=0)



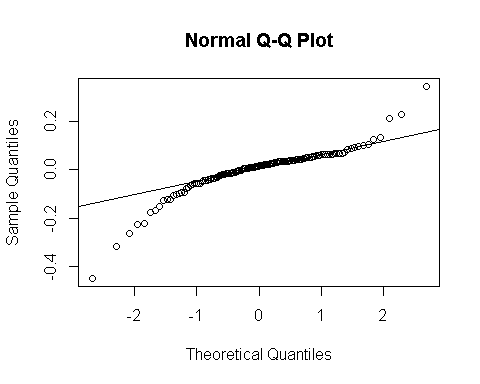
## Check for normality of random error  
qqnorm(mlr2$residuals)  
qqline(mlr2$residuals)



############################  
### Apply Transformation(we are not applying this because this is not useful for us) ###  
############################  
  
mlr2 <- lm(log(WomenBodyFat) ~ ., data=Dat)  
summary(mlr2)

##   
## Call:  
## lm(formula = log(WomenBodyFat) ~ ., data = Dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.44858 -0.03140 0.01434 0.04297 0.34298   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.1314946 0.2319437 -9.190 7.72e-16 \*\*\*  
## Age 0.0011161 0.0008103 1.378 0.171   
## Height -0.0091251 0.0013972 -6.531 1.32e-09 \*\*\*  
## WaistCircumference 0.0101175 0.0015316 6.606 9.04e-10 \*\*\*  
## HipCircumference 0.0160438 0.0014632 10.965 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09216 on 131 degrees of freedom  
## Multiple R-squared: 0.8832, Adjusted R-squared: 0.8796   
## F-statistic: 247.6 on 4 and 131 DF, p-value: < 2.2e-16

## Check for normality of random error  
qqnorm(mlr2$residuals)  
qqline(mlr2$residuals)



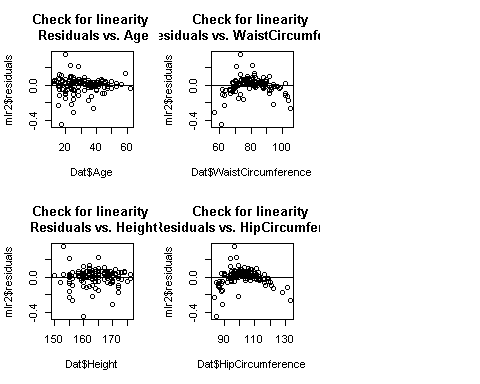
vif(mlr2)

## Age Height WaistCircumference HipCircumference   
## 1.301575 1.039263 3.663524 3.204528

cor(Dat)

## WomenBodyFat Age Height WaistCircumference  
## WomenBodyFat 1.0000000 0.36413427 -0.11345341 0.89865035  
## Age 0.3641343 1.00000000 -0.03102759 0.44337051  
## Height -0.1134534 -0.03102759 1.00000000 0.02629859  
## WaistCircumference 0.8986503 0.44337051 0.02629859 1.00000000  
## HipCircumference 0.9091104 0.25180713 0.13329028 0.81356458  
## HipCircumference  
## WomenBodyFat 0.9091104  
## Age 0.2518071  
## Height 0.1332903  
## WaistCircumference 0.8135646  
## HipCircumference 1.0000000

par(mfcol = c(2,3))  
plot(Dat$Age, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. Age")  
abline(h=0)  
plot(Dat$Height, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. Height")  
abline(h=0)  
plot(Dat$WaistCircumference, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. WaistCircumference")  
abline(h=0)  
plot(Dat$HipCircumference, mlr2$residuals,   
 main = "Check for linearity \n Residuals vs. HipCircumference")  
abline(h=0)  
#plot(Dat$ChestCircumference, mlr2$residuals,   
 # main = "Check for linearity \n Residuals vs. ChestCircumference")  
abline(h=0)  
#plot(Dat$UpperarmCircumference, mlr2$residuals,   
 # main = "Check for linearity \n Residuals vs. UpperarmCircumference")  
abline(h=0)



############################  
### removing outlier ###  
############################  
  
mlr$residuals

## 1 2 3 4 5   
## 5.956937e-03 -1.332844e-03 -5.251029e-03 -1.039632e-03 1.197011e-02   
## 6 7 8 9 10   
## -1.251717e-02 -6.791130e-03 1.928979e-02 6.188541e-03 1.036856e-02   
## 11 12 13 14 15   
## -4.345616e-03 -7.307195e-03 -2.495303e-02 4.918695e-03 -3.800157e-02   
## 16 17 18 19 20   
## 6.269688e-03 5.547783e-03 1.544445e-02 1.304023e-02 4.432850e-05   
## 21 22 23 24 25   
## 1.875277e-03 -3.501379e-04 -1.833176e-03 4.423430e-03 2.918340e-03   
## 26 27 28 29 30   
## -8.340157e-03 7.695157e-03 3.834305e-03 -8.304792e-03 1.218527e-02   
## 31 32 33 34 35   
## -1.232603e-02 -1.510744e-02 8.842128e-03 3.997699e-02 1.944964e-02   
## 36 37 38 39 40   
## -7.323980e-03 -1.510744e-02 1.218527e-02 -8.473889e-05 4.705266e-03   
## 41 42 43 44 45   
## 2.832812e-02 -7.721004e-03 4.394364e-03 -6.240423e-02 4.647099e-03   
## 46 47 48 49 50   
## -5.245700e-03 -5.270288e-03 -2.494817e-02 -9.101708e-04 -1.863025e-02   
## 51 52 53 54 55   
## 6.322043e-02 -3.853789e-03 1.543100e-04 7.608842e-03 9.529831e-03   
## 56 57 58 59 60   
## 2.019599e-02 -5.601523e-03 -3.417828e-03 8.411472e-03 -3.775943e-02   
## 61 62 63 64 65   
## 4.897513e-03 2.565091e-03 3.549651e-02 -1.365776e-02 2.381819e-03   
## 66 67 68 69 70   
## 1.120767e-01 1.206278e-02 6.671676e-03 -3.109216e-03 -4.055844e-04   
## 71 72 73 74 75   
## -1.915472e-02 -1.746129e-03 -5.871981e-02 5.507945e-03 3.402292e-03   
## 76 77 78 79 80   
## -5.098792e-03 -1.000104e-04 1.997043e-02 1.053151e-02 -3.169265e-02   
## 81 82 83 84 85   
## -8.431075e-04 4.398170e-04 1.249716e-02 -1.140568e-02 -1.709374e-04   
## 86 87 88 89 90   
## -7.090843e-03 -4.175282e-03 -1.331395e-02 -1.030602e-02 1.826647e-02   
## 91 92 93 94 95   
## 1.031556e-02 -5.146534e-03 -1.964057e-02 -1.476825e-02 8.045230e-03   
## 96 97 98 99 100   
## 3.674879e-03 1.261882e-03 5.508613e-02 -1.907924e-02 -1.732636e-03   
## 101 102 103 104 105   
## -3.566094e-02 -1.628274e-05 8.746780e-03 -4.597710e-03 -2.010930e-02   
## 106 107 108 109 110   
## 6.539880e-03 -8.914988e-03 -1.362584e-03 1.246452e-02 -2.296667e-03   
## 111 112 113 114 115   
## 9.461437e-04 5.057741e-03 4.007034e-03 7.017479e-03 1.650925e-02   
## 116 117 118 119 120   
## 3.233252e-03 -5.375014e-03 -2.596098e-03 -1.273549e-03 -2.384207e-02   
## 121 122 123 124 125   
## -3.169982e-02 -1.766917e-02 -7.158312e-03 -3.362470e-05 -1.122663e-02   
## 126 127 128 129 130   
## -5.330584e-03 4.009693e-03 1.626721e-02 -1.064186e-02 6.974544e-03   
## 131 132 133 134 135   
## 6.371222e-03 -4.987573e-03 3.768253e-03 -1.718057e-02 7.100382e-03   
## 136   
## 1.625125e-03

which.max(abs(mlr$residuals))

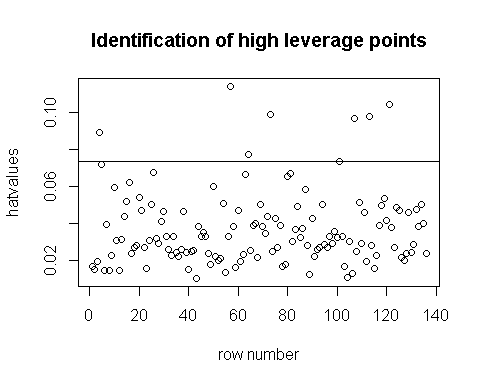
## 66   
## 66

############################  
### removing outlier ###  
############################  
## Find high leverage points  
k=4  
n =136  
hv <- hatvalues(mlr2)  
which(hv > (2\*(k+1)/n))

## 4 57 64 73 107 113 121   
## 4 57 64 73 107 113 121

row\_num <- c(1:136)

plot(row\_num, hv, xlab = "row number",  
 ylab = "hatvalues",   
 main = "Identification of high leverage points")  
abline(h = (2\*(k+1)/n))

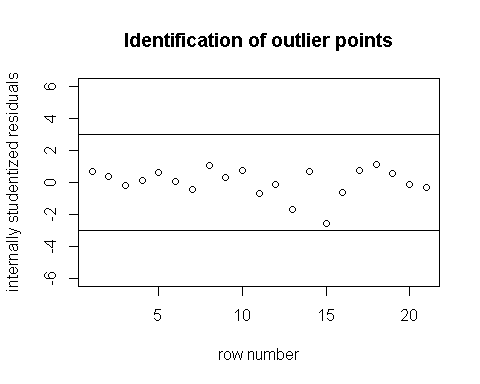


#outliers  
### Use internally standardized residuals (rstandard)  
r1 <- rstandard(mlr2)  
which(abs(r1) > 3)

## 44 66 80 121   
## 44 66 80 121

row\_num <- c(1:136)

plot(c(1,21), c(-6, 6), xlab = "row number",  
 ylab = "internally studentized residuals",   
 main = "Identification of outlier points",  
 type = "n")  
points(row\_num, r1)  
abline(h = 3)  
abline(h = -3)

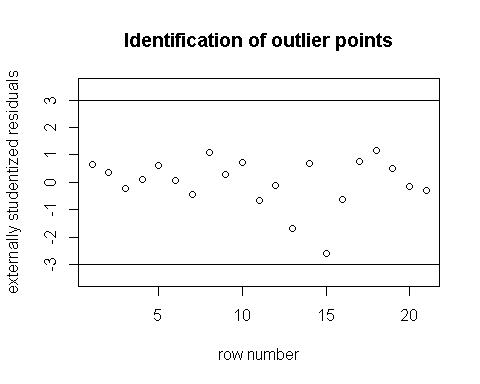


### Use externally standardized residuals (rstudent)  
r2 <- rstudent(mlr2)  
which(abs(r2) > 3)

## 44 66 80 121   
## 44 66 80 121

row\_num <- c(1:136)

plot(c(1,21), c(-3.5, 3.5), xlab = "row number",  
 ylab = "externally studentized residuals",   
 main = "Identification of outlier points",  
 type = "n")  
points(row\_num, r2)  
abline(h = 3)  
abline(h = -3)

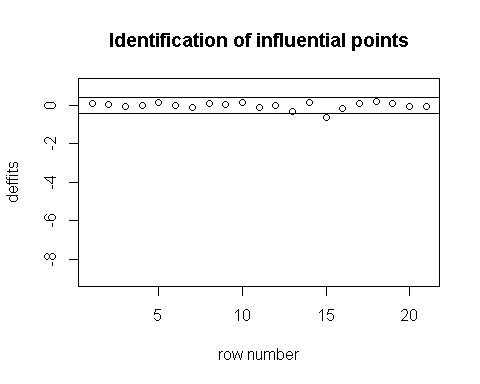


### Find influential points ###  
###############################  
  
### Use difference in fits (dffits)  
d <- dffits(mlr2) #influential if abs(dffits) > 2\*sqrt((k+2)/(n-k-2))  
which(abs(d) > 2\*sqrt((k+2)/(n-k-2)))

## 15 44 57 66 73 80 101 121   
## 15 44 57 66 73 80 101 121

row\_num <- c(1:136)

plot(c(1,21), c(-9, 1), xlab = "row number",  
 ylab = "deffits",   
 main = "Identification of influential points", type = "n")  
points(row\_num, d)  
abline(h = 2\*sqrt((k+2)/(n-k-2)))  
abline(h = - 2\*sqrt((k+2)/(n-k-2)))

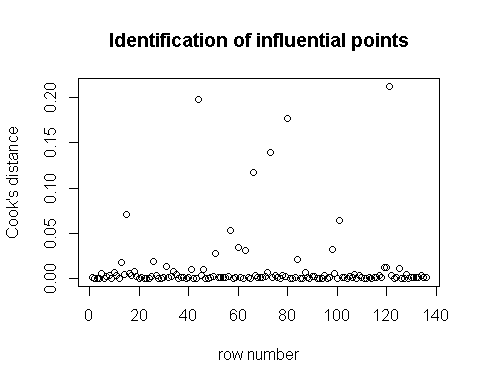


### Use Cook's distance (cooks.distance)  
cd <- cooks.distance(mlr2) #influential if d > 1  
which(cd > 1)

## named integer(0)

row\_num <- c(1:136)

plot(row\_num, cd, xlab = "row number",  
 ylab = "Cook's distance",   
 main = "Identification of influential points")  
abline(h = 1)



### Use COVRATIO (covratio)  
cv <- covratio(mlr2) #influential if covratio > 1 + 3\*(k+1)/n OR covratio < 1 - 3\*(k+1)/n  
which(cv > (1 + 3\*(k+1)/n))

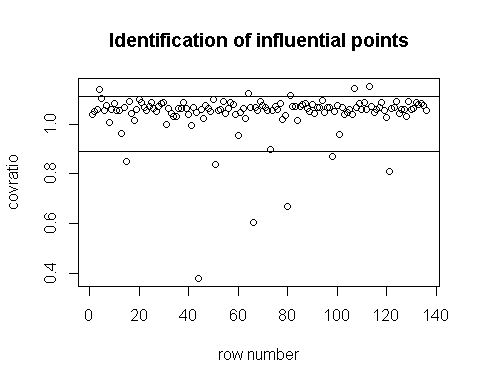
## 4 64 81 107 113   
## 4 64 81 107 113

which(cv < (1 - 3\*(k+1)/n))

## 15 44 51 66 80 98 121   
## 15 44 51 66 80 98 121

row\_num <- c(1:136)

plot(row\_num, cv, xlab = "row number",  
 ylab = "covratio",   
 main = "Identification of influential points")  
abline(h = (1 + 3\*(k+1)/n))  
abline(h = (1 - 3\*(k+1)/n))



cols <- character(nrow(Dat))  
## Remove Obs 66  
dat3 <- Dat[-c(44,51,66,73,98),]  
mlr3 <- lm((WomenBodyFat) ~ . , data = dat3)  
summary(mlr3) #unusual x1 and x2

##   
## Call:  
## lm(formula = (WomenBodyFat) ~ ., data = dat3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.044869 -0.005455 0.000598 0.006653 0.042639   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.0421907 0.0334279 -1.262 0.209   
## Age 0.0001574 0.0001156 1.361 0.176   
## Height -0.0024792 0.0001999 -12.401 <2e-16 \*\*\*  
## WaistCircumference 0.0036511 0.0002265 16.121 <2e-16 \*\*\*  
## HipCircumference 0.0046997 0.0002131 22.055 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.01288 on 126 degrees of freedom  
## Multiple R-squared: 0.9748, Adjusted R-squared: 0.974   
## F-statistic: 1218 on 4 and 126 DF, p-value: < 2.2e-16

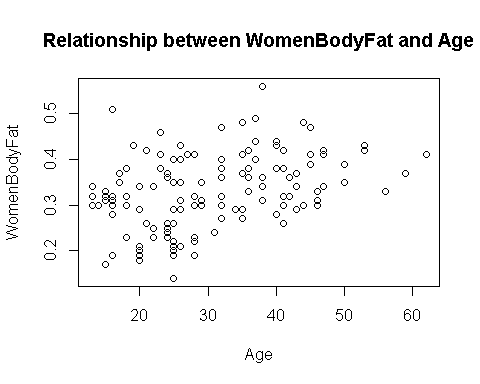
vif(mlr3)

## Age Height WaistCircumference HipCircumference   
## 1.328067 1.040894 3.888220 3.358758

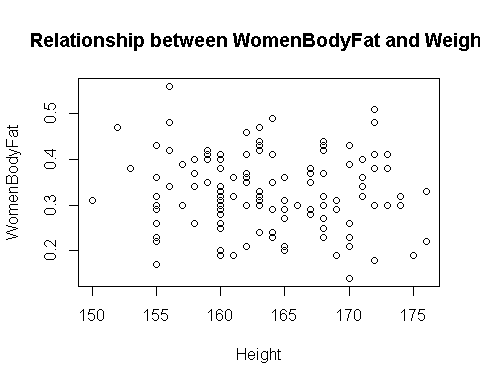
## 95% Confidence interval for model parameters ##  
confint(mlr3, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) -1.083435e-01 0.0239620473  
## Age -7.147381e-05 0.0003862463  
## Height -2.874848e-03 -0.0020835638  
## WaistCircumference 3.202863e-03 0.0040992576  
## HipCircumference 4.277990e-03 0.0051213943

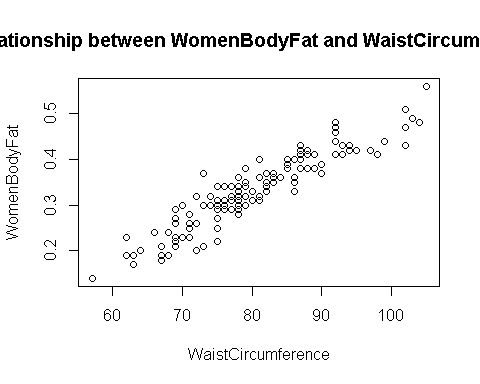
## Scatter plot ##  
plot(dat3$Age, dat3$WomenBodyFat, main = "Relationship between WomenBodyFat and Age",  
 xlab = "Age", ylab = "WomenBodyFat")



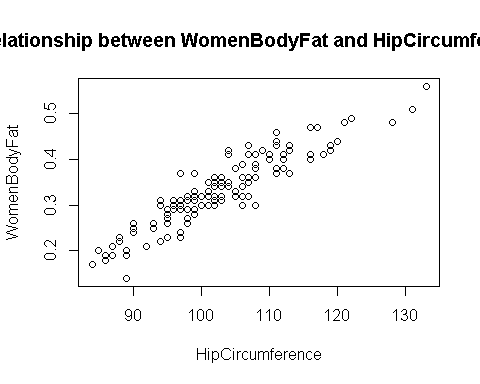
plot(dat3$Height, dat3$WomenBodyFat, main = "Relationship between WomenBodyFat and Weight",  
 xlab = "Height", ylab = "WomenBodyFat")



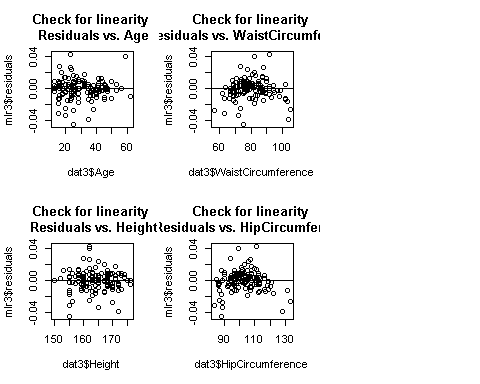
plot(dat3$WaistCircumference, dat3$WomenBodyFat, main = "Relationship between WomenBodyFat and WaistCircumference",  
 xlab = "WaistCircumference", ylab = "WomenBodyFat")



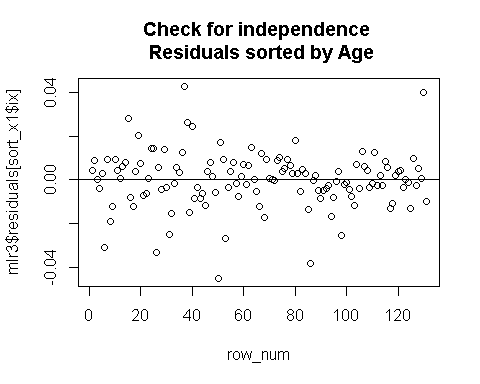
plot(dat3$HipCircumference, dat3$WomenBodyFat, main = "Relationship between WomenBodyFat and HipCircumference",  
 xlab = "HipCircumference", ylab = "WomenBodyFat")



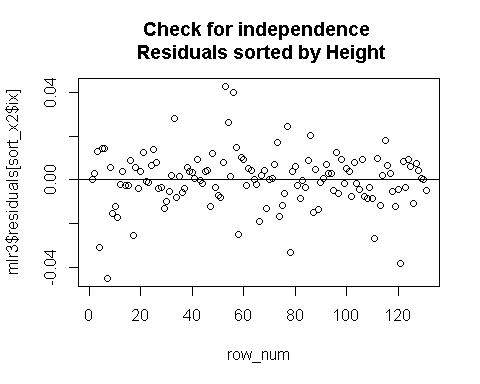
par(mfcol = c(2,3))  
plot(dat3$Age, mlr3$residuals,   
 main = "Check for linearity \n Residuals vs. Age")  
abline(h=0)  
plot(dat3$Height, mlr3$residuals,   
 main = "Check for linearity \n Residuals vs. Height")  
abline(h=0)  
plot(dat3$WaistCircumference, mlr3$residuals,   
 main = "Check for linearity \n Residuals vs. WaistCircumference")  
abline(h=0)  
plot(dat3$HipCircumference, mlr3$residuals,   
 main = "Check for linearity \n Residuals vs. HipCircumference")  
abline(h=0)



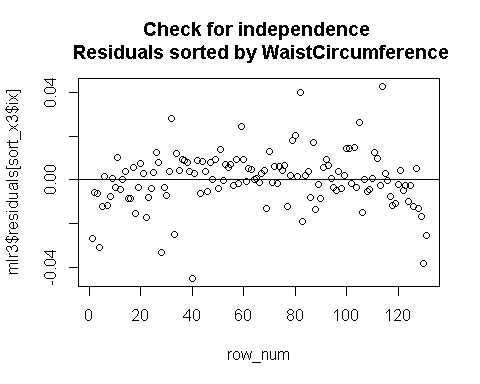
## Check for independence of random error  
par(mfcol = c(1, 1))  
row\_num <- c(1:nrow(dat3))  
sort\_x1 <- sort(dat3$Age, index.return=TRUE)  
plot(row\_num, mlr3$residuals[sort\_x1$ix],   
 main = "Check for independence \n Residuals sorted by Age")  
abline(h=0)



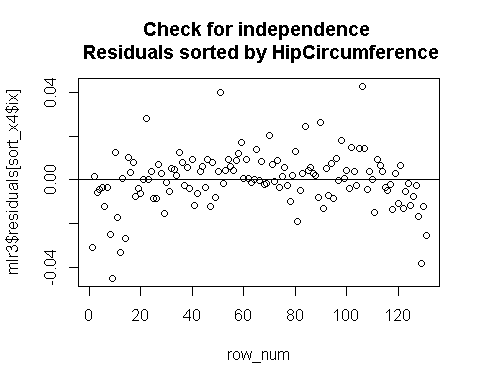
sort\_x2 <- sort(dat3$Height, index.return=TRUE)  
plot(row\_num, mlr3$residuals[sort\_x2$ix],   
 main = "Check for independence \n Residuals sorted by Height")  
abline(h=0)



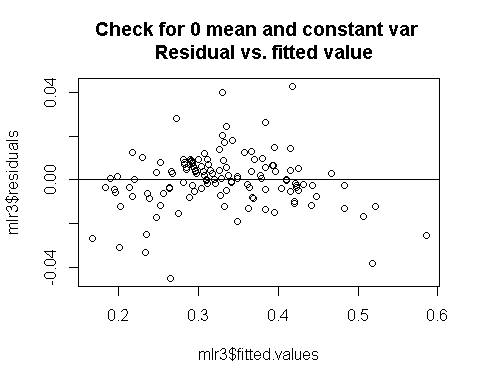
sort\_x3 <- sort(dat3$WaistCircumference, index.return=TRUE)  
plot(row\_num, mlr3$residuals[sort\_x3$ix],   
 main = "Check for independence \n Residuals sorted by WaistCircumference")  
abline(h=0)



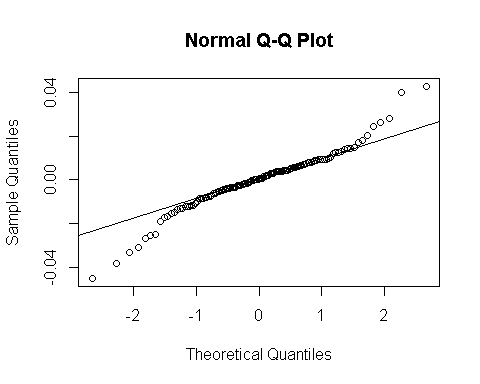
sort\_x4 <- sort(dat3$HipCircumference, index.return=TRUE)  
plot(row\_num, mlr3$residuals[sort\_x4$ix],   
 main = "Check for independence \n Residuals sorted by HipCircumference")  
abline(h=0)



## Check for zero mean and constant variance of random error  
plot(mlr3$fitted.values, mlr3$residuals,   
 main = "Check for 0 mean and constant var \n Residual vs. fitted value")  
abline(h=0)



## Check for normality of random error  
qqnorm(mlr3$residuals)  
qqline(mlr3$residuals)



shapiro.test(mlr3$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: mlr3$residuals  
## W = 0.95067, p-value = 0.0001198