STAT GR5205 – Section 005 HW 8

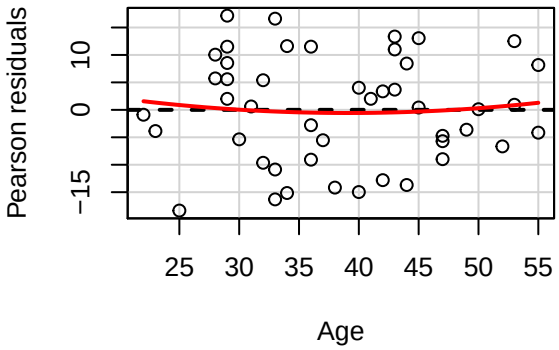
Bo Rong br2498

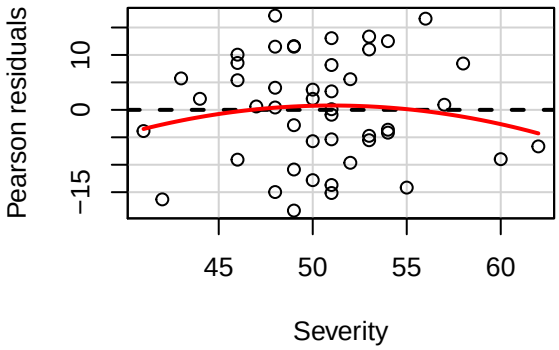
Dec. 10th, 2016

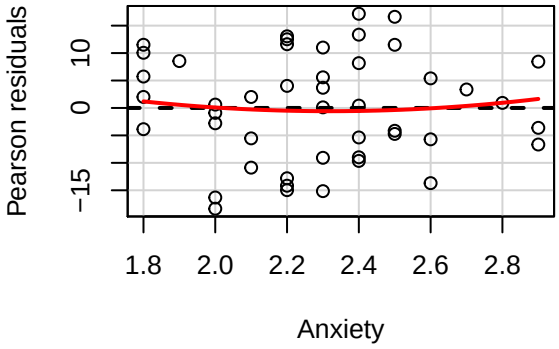
```
#1.
#(a)
filename <- "~/Downloads/patient_satisfaction.txt"
PS<-read.table(file=filename, header=T)
mlack <-lm(Y ~ Age + Severity + Anxiety, data=PS)
summary(mlack)
```

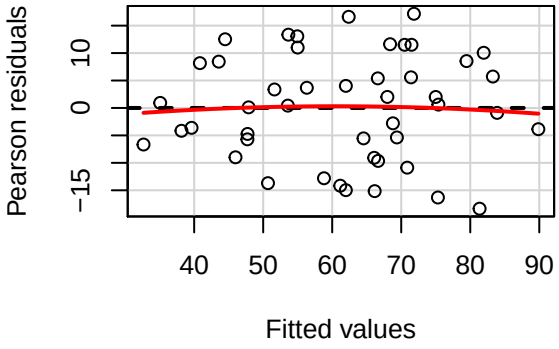
```
##
## Call:
## lm(formula = Y ~ Age + Severity + Anxiety, data = PS)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## Age          -1.1416     0.2148  -5.315 3.81e-06 ***
## Severity     -0.4420     0.4920  -0.898   0.3741
## Anxiety     -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

```
library(car)
residualPlots(mlack)
```










```
##             Test stat Pr(>|t|)
## Age             0.387    0.701
## Severity       -0.802    0.427
## Anxiety         0.475    0.637
## Tukey test     -0.225    0.822
```

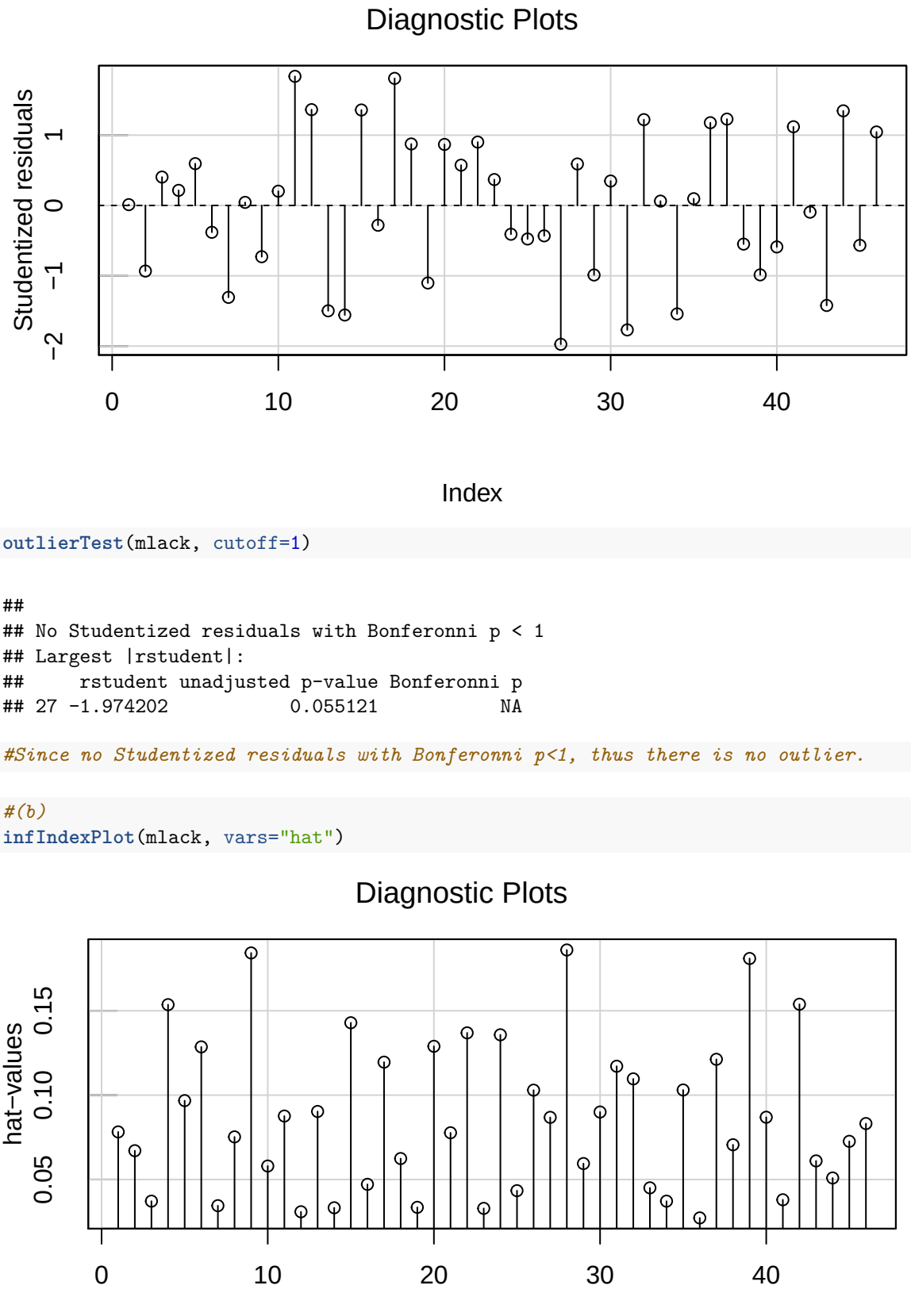
```
# No evidence of lack of fit in linear mean function.
```

```
#(b)
#H0:Var(Y|X=x)=sigma^2  H1:Var(Y|X=x)=sigma^2exp{gamma1x1 +gamma2x2 +gamma3x3}
#not all gamma is 0.
ncvTest(mlack, ~ Age + Severity + Anxiety)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ Age + Severity + Anxiety
## Chisquare = 1.25157    Df = 3     p = 0.7406642
```

```
#The P -value is 0.7406642, so there is no evidence of nonconstant variance.
```

```
#2.
#(a)
infIndexPlot(mlack, vars="Studentized")
```




```
outlierTest(mlack, cutoff=1)
```

```
##
## No Studentized residuals with Bonferonni p < 1
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferonni p
## 27 -1.974202           0.055121           NA
```

```
#Since no Studentized residuals with Bonferonni p<1, thus there is no outlier.
```

```
#(b)
infIndexPlot(mlack, vars="hat")
```

```
#The three highest hat-values are for cases 9, 28, and 39.No high leverage cases.
```

```
#(c)
X <- as.matrix(PS[,-1])
solve(t(X) %*% X)
```

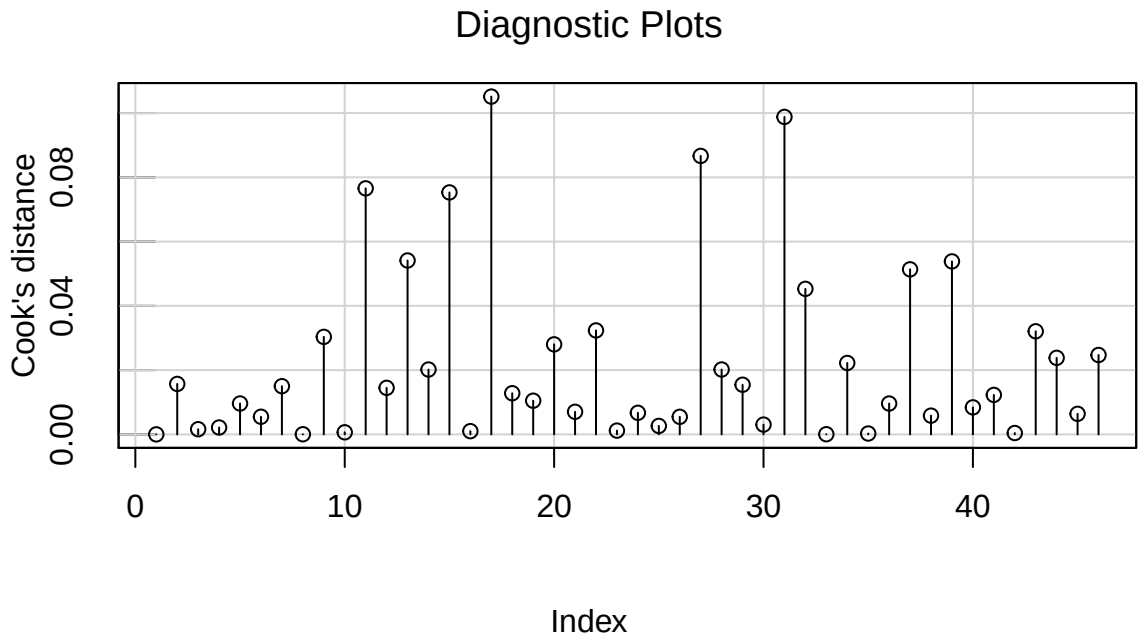
```
##                    Age      Severity      Anxiety
## Age        0.0004299003 -0.0001257215 -0.004471192
## Severity -0.0001257215  0.0009716065 -0.019117741
## Anxiety  -0.0044711916 -0.0191177410  0.496863170
```

```
x <- c(30, 58, 2)
as.numeric(t(x) %*% solve(t(X)%*%X, x))
```

```
## [1] 0.2334774
```

```
#The new case leverage value for x = (58, 30, 2.0) is 0.2334774, which is greater than
#the highest hat-value. So this estimate involve a "hidden extrapolation".
```

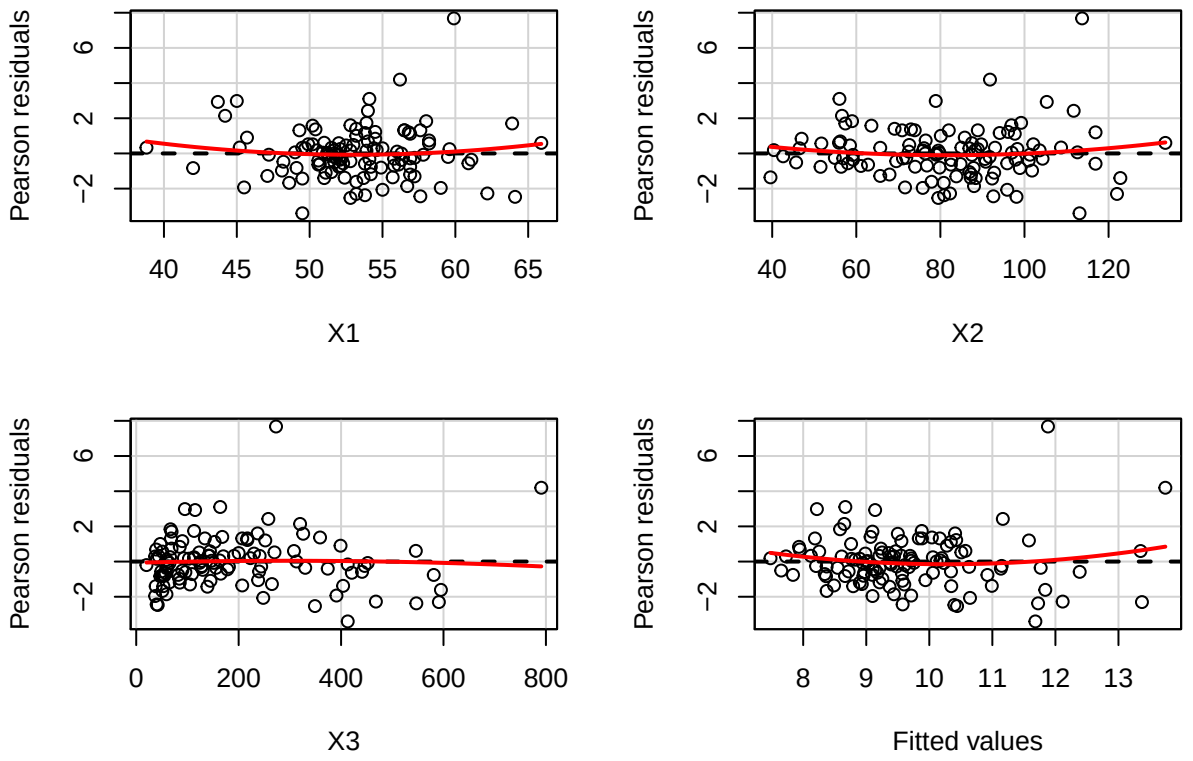
```
#(d)
infIndexPlot(mlack, vars="Cook")
```




```
#The most influential cases are 17, 27 and 31. And all three cases are outlierness.
```

```
#3.
#(a)
filename <- "~/Downloads/SENIC.txt"
SENIC<-read.table(file=filename, header=T)
Y <- SENIC$Stay
X1 <- SENIC$Age
```

```r
X2 <- SENIC$Xray
X3 <- SENIC$Cen
m1 <- lm(Y ~ X1 + X2 + X3)
residualPlots(m1)
```




```
##            Test stat Pr(>|t|)
## X1             0.906    0.367
## X2             0.951    0.344
## X3            -0.289    0.773
## Tukey test     1.310    0.190
```

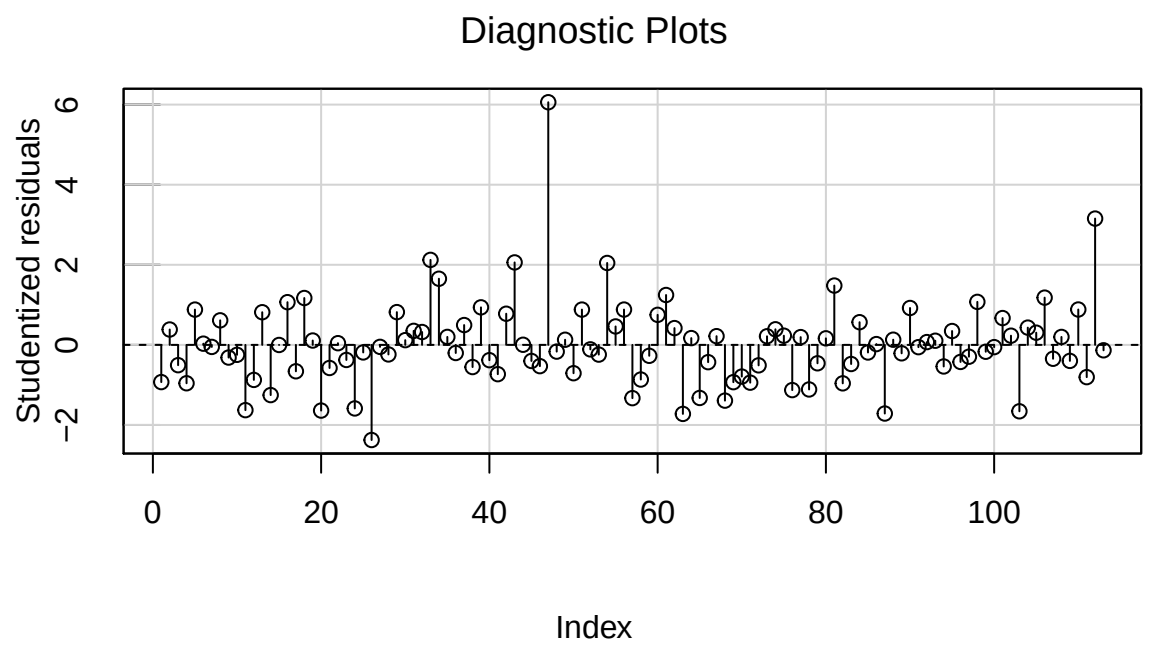
```r
#No indication of curvature in the residual plots, thus no evidence of lack of fit for the linear mean
```

```r
#(b)
#H0:Var(Y|X=x)=sigma^2  H1:Var(Y|X=x)=sigma^2exp{gamma1x1 +gamma2x2 +gamma3x3}
ncvTest(m1, ~X1+X2+X3)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ X1 + X2 + X3
## Chisquare = 42.96221    Df = 3     p = 2.506893e-09
```

```r
#The P-value is 2.506893e-09. The variance is not constant.
```

```r
#(c)
infIndexPlot(m1, vars="Studentized")
```

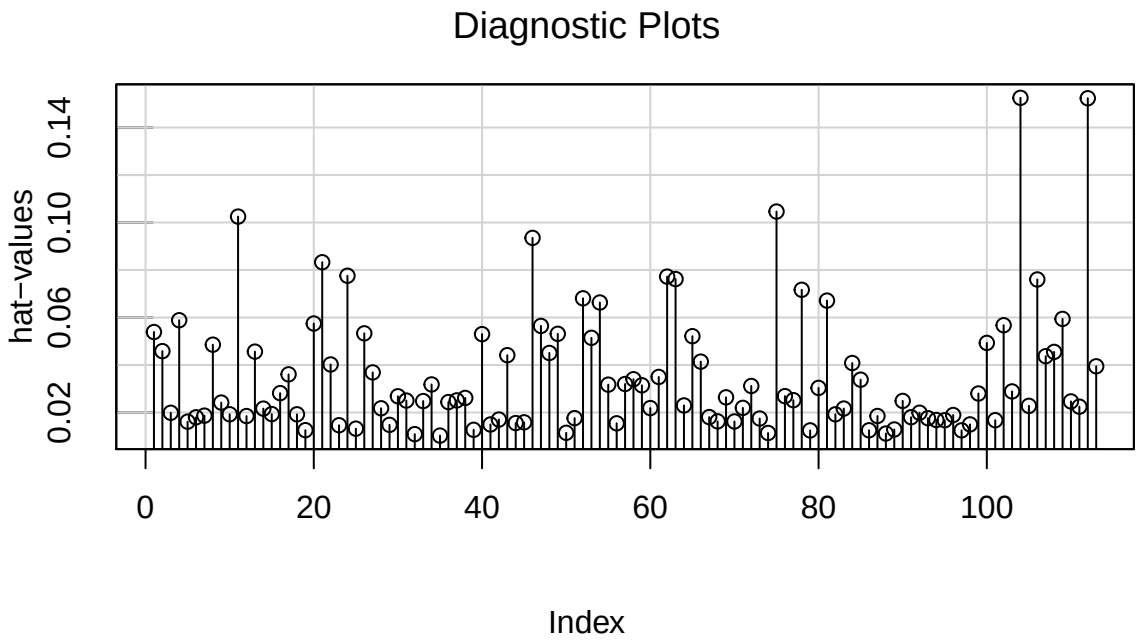



```
outlierTest(m1, cutoff=1)
```

```
##     rstudent unadjusted p-value Bonferonni p
## 47  6.059044         2.0360e-08   2.3006e-06
## 112 3.154348         2.0837e-03   2.3546e-01
```

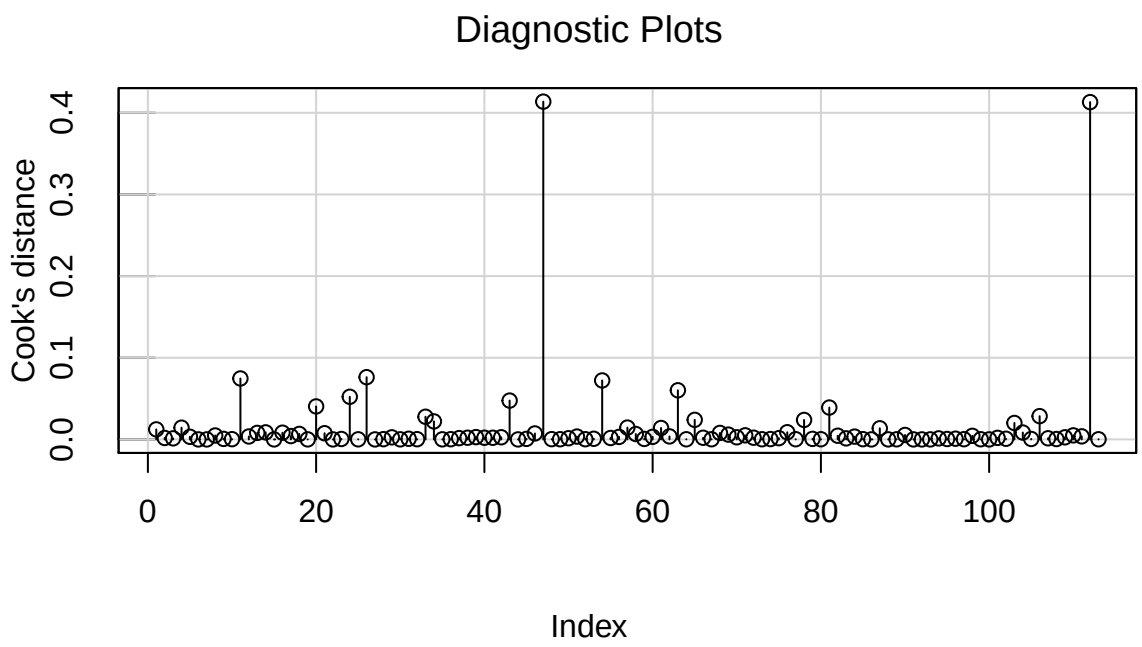
```
#Case 47 and 112 are outliers.
```

```
#(d)
infIndexPlot(m1, vars="hat")
```



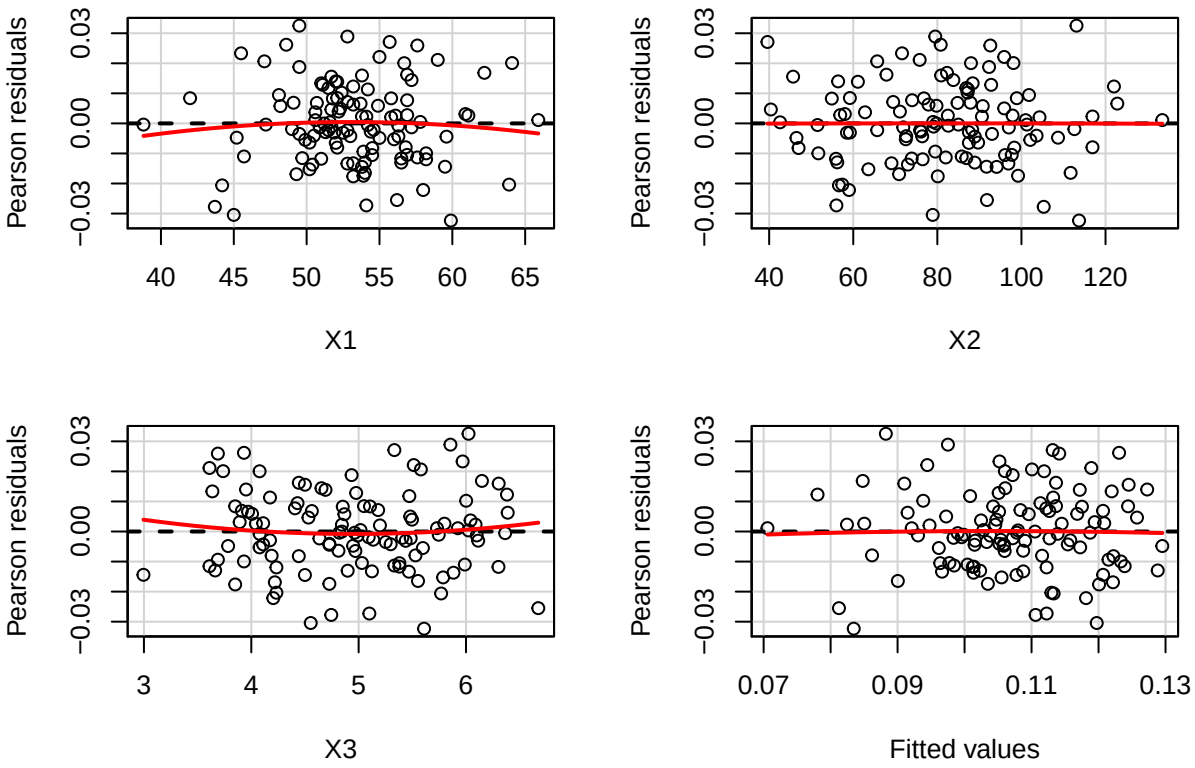
```
#Cases 104 and 112 are high leverage.
```

```
#(e)
infIndexPlot(m1, vars="Cook")
```




```
#Cases 47 and 112 are the most influential cases in the dataset. Case 47 is an outlier, but not a high
```

```
#4.
#(a)
Y <- 1/SENIC$Stay
X1 <- SENIC$Age
X2 <- SENIC$Xray
X3 <- log(SENIC$Cen)
m2 <- lm(Y ~ X1 + X2 + X3)
residualPlots(m2)
```




```
##             Test stat Pr(>|t|)
## X1             -0.616    0.539
## X2             -0.019    0.985
## X3              0.679    0.498
## Tukey test     -0.151    0.880
```

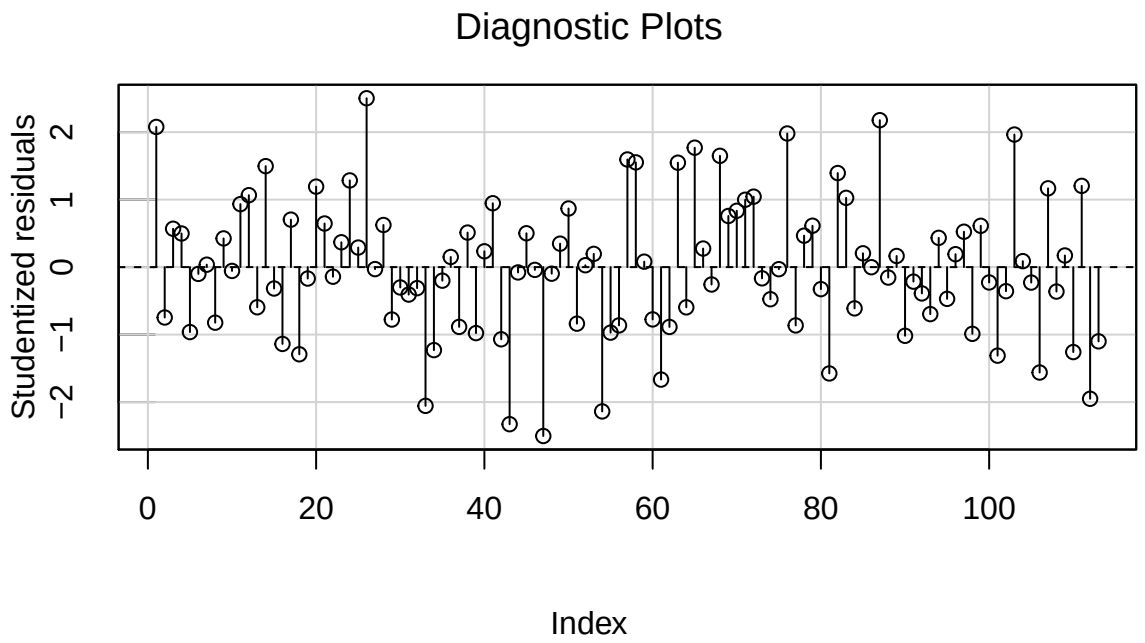
```
#Linear mean function seems reasonable.

#(b)
ncvTest(m2, ~ X1 + X2 + X3)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ X1 + X2 + X3
## Chisquare = 0.1527719    Df = 3     p = 0.9848272
```

```
#The p-value is 0.9848272, so the constant variance assumption is reasonable.

#(c)
infIndexPlot(m2, vars="Studentized")
```

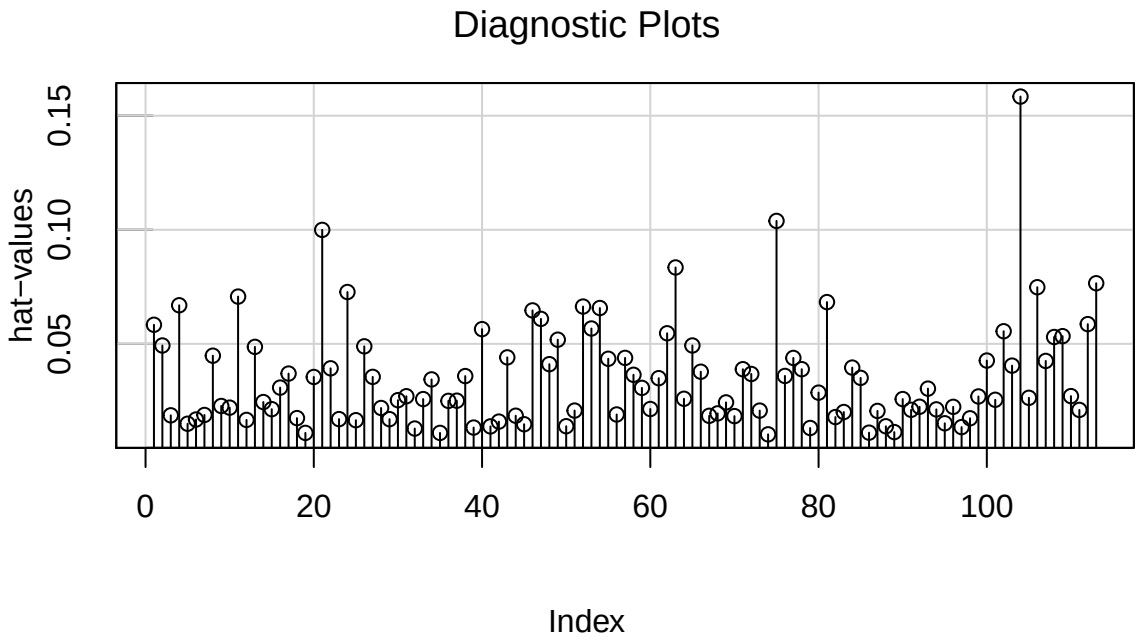



```
outlierTest(m2, cutoff=1)
```

```
##
## No Studentized residuals with Bonferonni p < 1
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferonni p
## 47 -2.502595           0.013827           NA
```

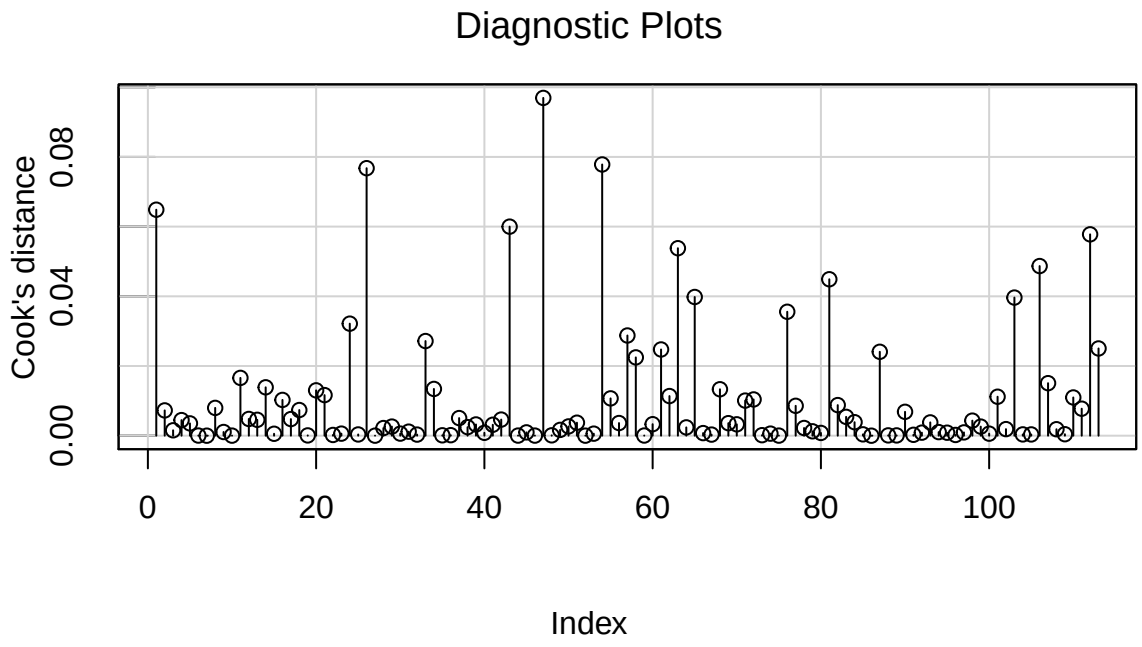
```
#Case 1 and 87 maybe outliers.

#(d)
infIndexPlot(m2, vars="hat")
```



```r
#Cases 104 is high leverage.

#(e)
infIndexPlot(m2, vars="Cook")
```




```r
#Case 47 is the most influential cases in the dataset. But case 47 is not an outlier.
```