Assignment 6

Apurva Hari

August 9, 2016

2. If you're feeling more adventurous, try the code from Chapter 8 in Kabacoff's R In Action, which offers further diagnostics and more exploration related to linear regression, such as doing cross-validation or trying to evaluate the relative importance of predictors. Not all of it will work; in particular, I think some of the functions in the "car" package are deprecated. Try FOUR from the following:

Listing 8.6

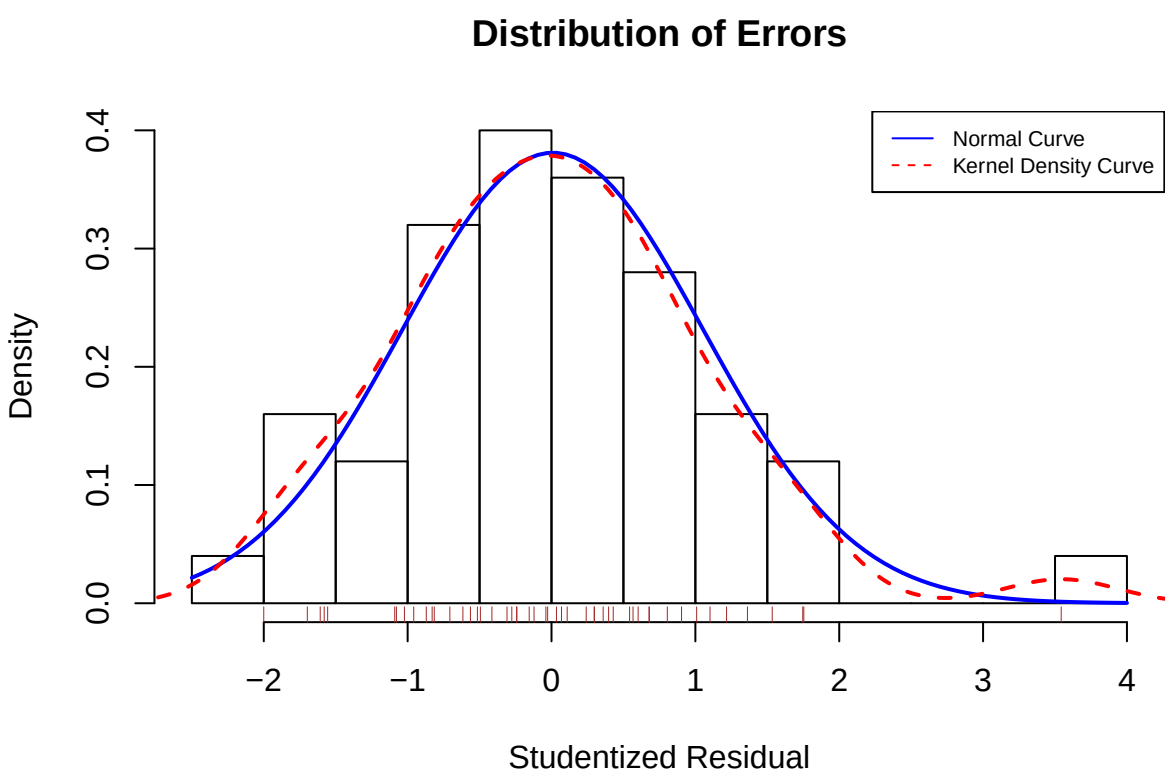
Function for plotting studentized residuals

```
states <- as.data.frame(state.x77[,c("Murder","Population","Illiteracy","Income",
                                     "Frost")])
fit <- lm(Murder ~ Population + Illiteracy + Income + Frost, data = states )

residplot <- function(fit,nbreaks =10) {

        z <-rstudent(fit)
        hist(z,breaks=nbreaks,freq= FALSE, xlab ="Studentized Residual",
        main = "Distribution of Errors")
        rug(jitter(z),col ="brown")
        curve(dnorm(x,mean =mean(z),sd =sd(z)),
        add = TRUE, col ="blue" ,lwd = 2)
        lines(density(z)$x,density(z)$y,
        col ="red",lwd = 2, lty = 2)
        legend("topright",
        legend = c("Normal Curve", "Kernel Density Curve"),
        lty = 1:2, col =c("blue","red"), cex =.7)
}

residplot(fit)
```



Listing 8.8 ####Global Test of Linear Model Assumptions

```
library(gvlma)
gvmodel <- gvlma(fit)

summary(gvmodel)
```

```
##
## Call:
## lm(formula = Murder ~ Population + Illiteracy + Income + Frost,
##     data = states)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7960 -1.6495 -0.0811  1.4815  7.6210
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.235e+00  3.866e+00   0.319   0.7510
## Population  2.237e-04  9.052e-05   2.471   0.0173 *
## Illiteracy  4.143e+00  8.744e-01   4.738 2.19e-05 ***
## Income      6.442e-05  6.837e-04   0.094   0.9253
## Frost       5.813e-04  1.005e-02   0.058   0.9541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.535 on 45 degrees of freedom
## Multiple R-squared:  0.567,  Adjusted R-squared:  0.5285
## F-statistic: 14.73 on 4 and 45 DF,  p-value: 9.133e-08
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                     Value p-value                Decision
## Global Stat        2.7728  0.5965 Assumptions acceptable.
## Skewness           1.5374  0.2150 Assumptions acceptable.
## Kurtosis           0.6376  0.4246 Assumptions acceptable.
## Link Function      0.1154  0.7341 Assumptions acceptable.
## Heteroscedasticity 0.4824  0.4873 Assumptions acceptable.
```

```
gvlma(x=fit)
```

```
##
## Call:
## lm(formula = Murder ~ Population + Illiteracy + Income + Frost,
##     data = states)
##
## Coefficients:
## (Intercept)   Population   Illiteracy       Income        Frost
##   1.235e+00    2.237e-04    4.143e+00    6.442e-05    5.813e-04
##
##
## ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
## USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
## Level of Significance =  0.05
##
## Call:
##  gvlma(x = fit)
##
##                     Value p-value                Decision
## Global Stat        2.7728  0.5965 Assumptions acceptable.
## Skewness           1.5374  0.2150 Assumptions acceptable.
## Kurtosis           0.6376  0.4246 Assumptions acceptable.
## Link Function      0.1154  0.7341 Assumptions acceptable.
## Heteroscedasticity 0.4824  0.4873 Assumptions acceptable.
```

Listing 8.11

Comparing nested models using anova()

```
states <- as.data.frame(state.x77[,c("Murder","Population","Illiteracy","Income",
                                     "Frost")])
fit1 <- lm(Murder ~ Population + Illiteracy + Income + Frost, data = states )
```

```
fit2 <- lm(Murder ~ Population + Illiteracy, data = states )

anova(fit2,fit1)
```

```
## Analysis of Variance Table
##
## Model 1: Murder ~ Population + Illiteracy
## Model 2: Murder ~ Population + Illiteracy + Income + Frost
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     47 289.25
## 2     45 289.17  2  0.078505 0.0061 0.9939
```

Listing 8.12

Comparing models with AIC

```
states <- as.data.frame(state.x77[,c("Murder","Population","Illiteracy","Income",
                                     "Frost")])
fit1 <- lm(Murder ~ Population + Illiteracy + Income + Frost, data = states )
fit2 <- lm(Murder ~ Population + Illiteracy, data = states )

AIC(fit2,fit1)
```

```
##      df      AIC
## fit2  4 237.6565
## fit1  6 241.6429
```