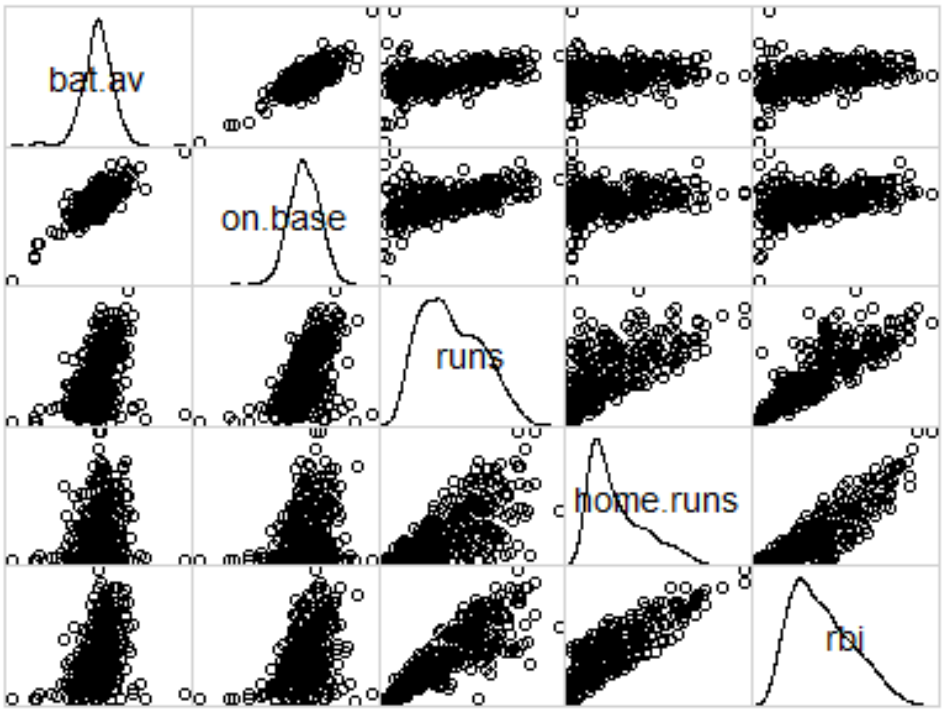
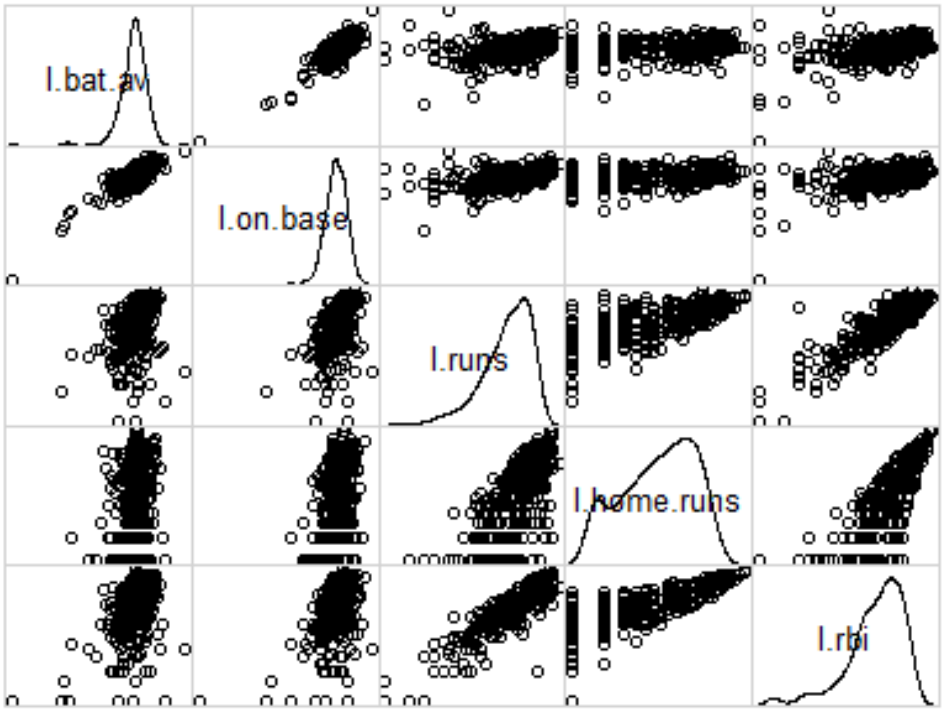
```
# Load dataset
setwd("C:/Users/TMICHAEL/Google Drive/00projects@tk563/baseball")
bball = read.csv("bball.csv")
head(bball)

##   salary bat.av on.base runs hits doubles triples home.runs rbi walks
## 1   3300  0.272   0.302   69  153      21       4        31 104    22
## 2   2600  0.269   0.335   58  111      17       2        18  66    39
## 3   2500  0.249   0.337   54  115      15       1        17  73    63
## 4   2475  0.260   0.292   59  128      22       7        12  50    23
## 5   2313  0.273   0.346   87  169      28       5         8  58    70
## 6   2175  0.291   0.379  104  170      32       2        26 100    87
##   strikes stolen errors free.elig free.agent arb.elig arb
## 1      80      4      3         1          0        0   0
## 2      69      0      3         1          1        0   0
## 3     116      6      5         1          0        0   0
## 4      64     21     21         0          0        1   0
## 5      53      3      8         0          0        1   0
## 6      89     22      4         1          0        0   0
##              player
## 1 Andre Dawson
## 2 Steve Buchele
## 3 Kal Daniels
## 4 Shawon Dunston
## 5 Mark Grace
## 6 Ryne Sandberg

## Distributions of Independent Variables

library(corrgram)
# Vector of variable names
corrdat = c("bat.av", "on.base", "runs", "home.runs", "rbi")
# Creates scatterplots of variables in corrdat
corrgram(bball[, corrdat], panel = "panel.pts", diag.panel = "panel.density")
```



```
# Creating log variables in the dataset
bball[, 19:23] = c(log(bball$bat.av), log(bball$on.base), log(bball$runs), log(bball$home.runs),
   log(bball$rbi))

# Assigning names to these variables
names(bball)[19:23] = c("l.bat.av", "l.on.base", "l.runs", "l.home.runs", "l.rbi")

# Vector of variable names
corrdat2 = c("l.bat.av", "l.on.base", "l.runs", "l.home.runs", "l.rbi")
# Creates Scatterplots
corrgram(bball[, corrdat2], panel = "panel.pts", diag.panel = "panel.density")
```



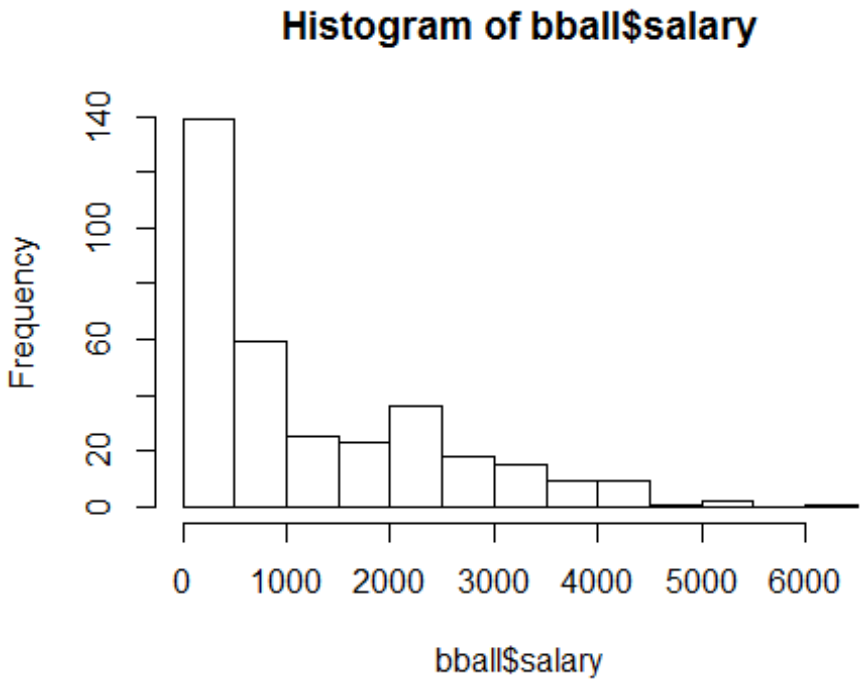

```
# Covariance matrix
cor(bball[,corrdat])
```

```
##              bat.av   on.base      runs home.runs       rbi
## bat.av    1.0000000 0.8059576 0.4367424 0.2127320 0.3695426
## on.base   0.8059576 1.0000000 0.5135739 0.3112102 0.3994226
## runs      0.4367424 0.5135739 1.0000000 0.6810578 0.8334819
## home.runs 0.2127320 0.3112102 0.6810578 1.0000000 0.8773785
## rbi       0.3695426 0.3994226 0.8334819 0.8773785 1.0000000
```

```
# three most corrleated pairs of variables are home.runs & rbi, runs & rbi, bat.av & on.base
```

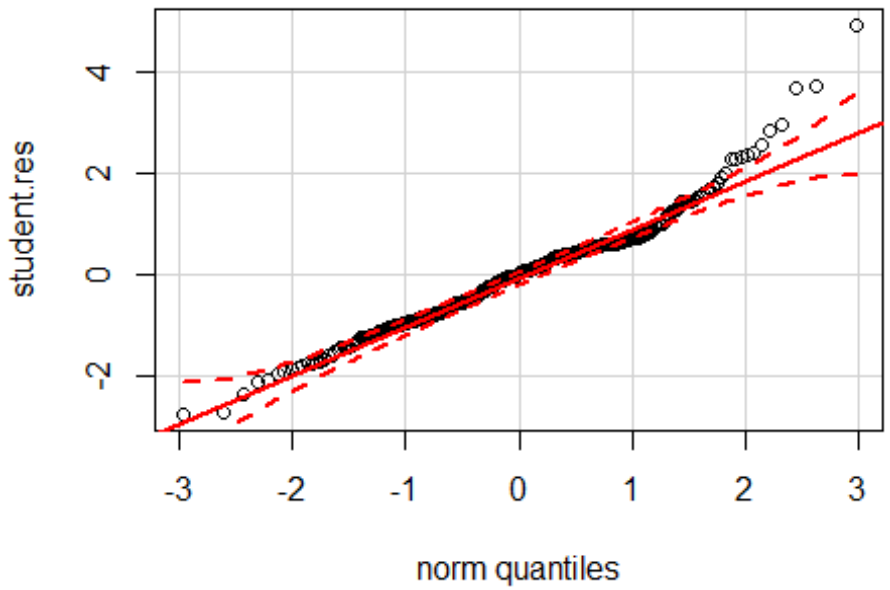
Distributions of Response Variable

```
hist(bball$salary)
```

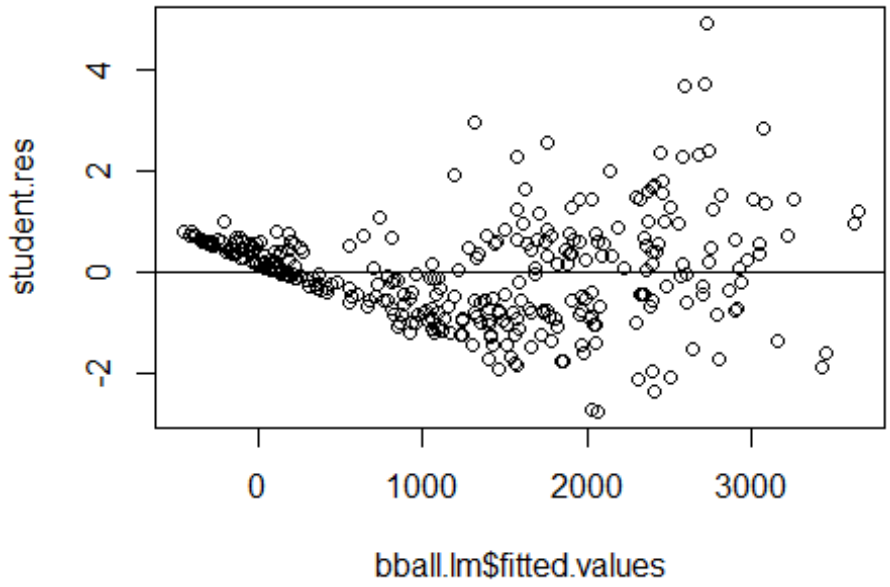


Initial Model

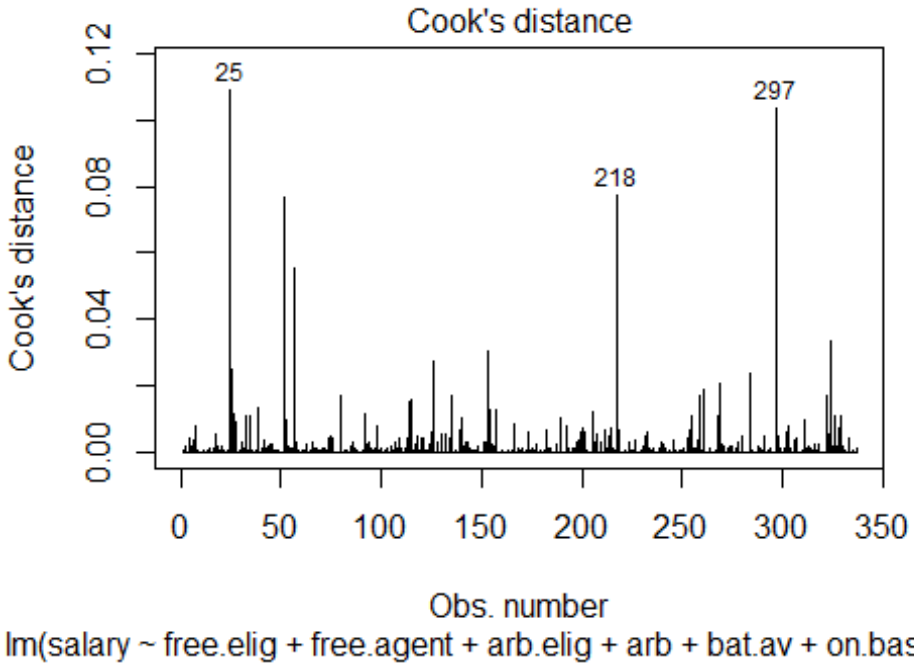
```
# denote all of the categorical predictors as factors
bball$free.elig = as.factor(bball$free.elig)
bball$free.agent = as.factor(bball$free.agent)
bball$arb.elig = as.factor(bball$arb.elig)
bball$arb = as.factor(bball$arb)

# fit a multiple linear regression model
bball.lm = lm(salary~free.elig+free.agent+arb.elig+arb+bat.av+on.base+runs+home.runs+rbi,data=bball)

# MLR model diagnostics: error distribution 1) normality, 2) homogeneous variance

library(MASS)
student.res = studres(bball.lm)
library(car)
qqPlot(student.res)
```



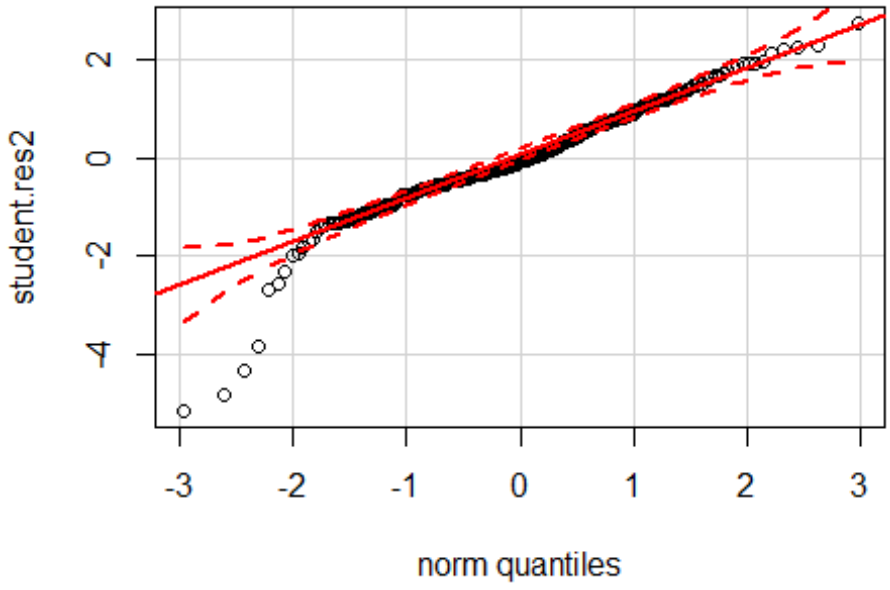

```
plot(bball.lm$fitted.values,student.res)
abline(0,0)
```



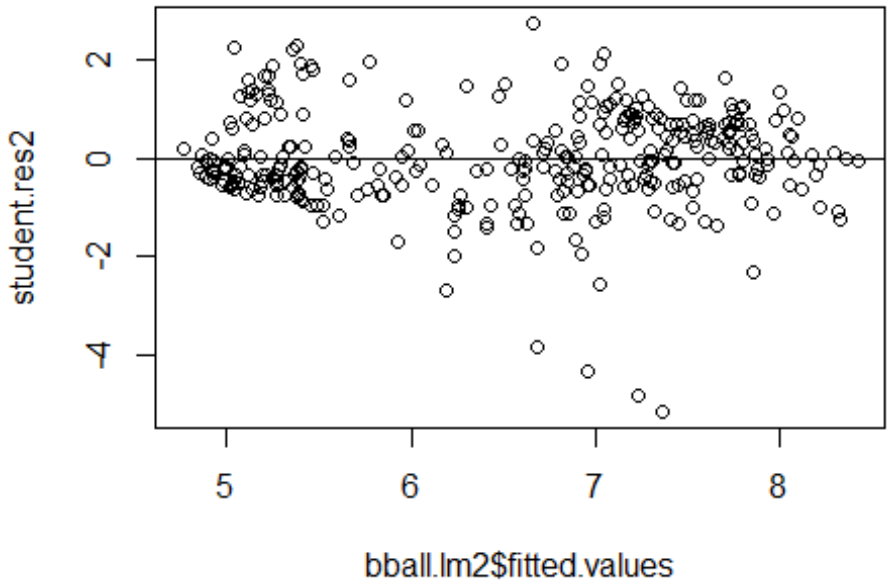

```
# Store Cook's distances for every observation in a vector
cooks_d = cooks.distance(bball.lm)
# Creates plot of Cook's distances
cutoff=1
plot(bball.lm, which=4, cook.levels=cutoff)
```



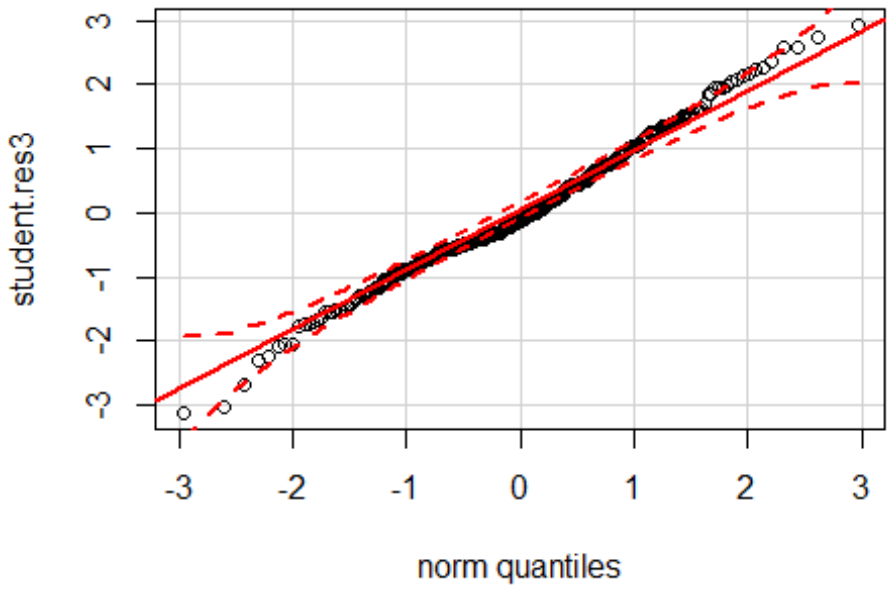

```
## New Model

bball.lm2 = lm(log(salary)~free.elig+free.agent+arb.elig+arb+bat.av+on.base+runs+home.runs+rbi,data=bball)

library(MASS)
student.res2 =studres(bball.lm2)
library(car)
qqPlot(student.res2)
```



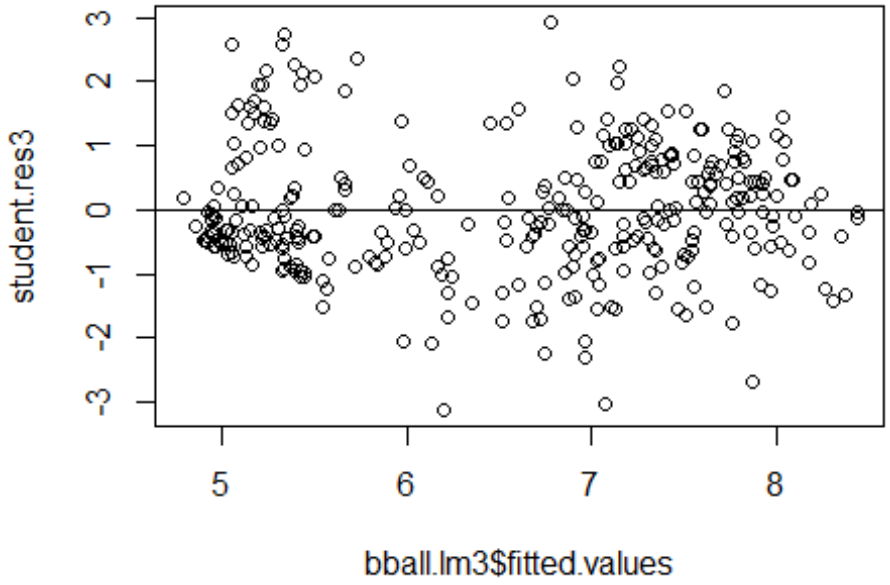

```
plot(bball.lm2$fitted.values,student.res2)
abline(0,0)
```




```
# Determines which observations have large residuals
outliers=which(abs(student.res2)>3)
outliers  # Both rows are the observation numbers of the outliers

## 205 268 284 322
## 205 268 284 322

bball$salary[outliers] # Salary of outliers

## [1] 109 109 109 109

#  Determines observations we want to include
not_outlier = which(abs(student.res2) <= 3)

# Subset option allows only a subset of the data to be analyzed
bball.lm3 = lm(log(salary) ~ free.elig + free.agent + arb.elig + arb + bat.av +
    on.base + runs + home.runs + rbi, data = bball, subset = not_outlier)

library(MASS)
student.res3 = studres(bball.lm3)
library(car)
qqPlot(student.res3)
```




```
plot(bball.lm3$fitted.values, student.res3)
abline(0, 0)
```




```
## Examine the Model w/o outliers, w/ log transform of the response

summary(bball.lm3)

##
## Call:
## lm(formula = log(salary) ~ free.elig + free.agent + arb.elig +
##     arb + bat.av + on.base + runs + home.runs + rbi, data = bball,
##     subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42021 -0.27859 -0.06005  0.31637  1.36010
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.002368   0.198839  25.158  < 2e-16 ***
```

```
## free.elig1   1.730249   0.072302  23.931  < 2e-16 ***
## free.agent1 -0.271726   0.095059  -2.859 0.004533 **
## arb.elig1    1.405013   0.080562  17.440  < 2e-16 ***
## arb1        -0.142809   0.165768  -0.861 0.389602
## bat.av       0.464908   1.201726   0.387 0.699110
## on.base     -0.783290   1.033402  -0.758 0.449020
## runs         0.008787   0.001847   4.757 2.96e-06 ***
## home.runs   -0.008491   0.006400  -1.327 0.185588
## rbi          0.009999   0.002673   3.740 0.000217 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4785 on 323 degrees of freedom
## Multiple R-squared:  0.8361, Adjusted R-squared:  0.8316
## F-statistic: 183.1 on 9 and 323 DF,  p-value: < 2.2e-16
```

```
vif(bball.lm3)
```

```
##  free.elig free.agent   arb.elig        arb     bat.av    on.base
##   1.814218   1.328492   1.465419   1.164046   3.284002   3.472249
##       runs  home.runs        rbi
##   4.193382   5.139175   9.082227
```

```
summary(bball.lm3, correlation=TRUE)
```

```
##
## Call:
## lm(formula = log(salary) ~ free.elig + free.agent + arb.elig +
##     arb + bat.av + on.base + runs + home.runs + rbi, data = bball,
##     subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42021 -0.27859 -0.06005  0.31637  1.36010
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.002368   0.198839  25.158  < 2e-16 ***
## free.elig1   1.730249   0.072302  23.931  < 2e-16 ***
## free.agent1 -0.271726   0.095059  -2.859 0.004533 **
## arb.elig1    1.405013   0.080562  17.440  < 2e-16 ***
## arb1        -0.142809   0.165768  -0.861 0.389602
## bat.av       0.464908   1.201726   0.387 0.699110
## on.base     -0.783290   1.033402  -0.758 0.449020
## runs         0.008787   0.001847   4.757 2.96e-06 ***
## home.runs   -0.008491   0.006400  -1.327 0.185588
## rbi          0.009999   0.002673   3.740 0.000217 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4785 on 323 degrees of freedom
## Multiple R-squared:  0.8361, Adjusted R-squared:  0.8316
## F-statistic: 183.1 on 9 and 323 DF,  p-value: < 2.2e-16
##
## Correlation of Coefficients:
##             (Intercept) free.elig1 free.agent1 arb.elig1 arb1  bat.av
## free.elig1  -0.05
```

```
## free.agent1 -0.07       -0.45
## arb.elig1   -0.10        0.39      -0.03
## arb1         0.02        0.03      -0.01       -0.29
## bat.av      -0.22        0.11      -0.02       -0.04     -0.02
## on.base     -0.42       -0.10       0.05        0.07      0.02 -0.78
## runs         0.24       -0.15       0.11       -0.07     -0.08  0.07
## home.runs   -0.03       -0.05       0.15        0.09      0.02  0.28
## rbi         -0.02       -0.04      -0.11       -0.13      0.01 -0.26
##             on.base runs  home.runs
## free.elig1
## free.agent1
## arb.elig1
## arb1
## bat.av
## on.base
## runs        -0.27
## home.runs   -0.19    0.19
## rbi          0.21   -0.62 -0.78
```

 # Keep track of sums of squares

```
anova(bball.lm3)
```

```
## Analysis of Variance Table
##
## Response: log(salary)
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## free.elig    1 178.890 178.890 781.2627 < 2.2e-16 ***
## free.agent   1   6.668   6.668  29.1200 1.319e-07 ***
## arb.elig     1 137.247 137.247 599.3972 < 2.2e-16 ***
## arb          1   0.047   0.047   0.2043 0.6515463
## bat.av       1   8.553   8.553  37.3531 2.837e-09 ***
## on.base      1   0.957   0.957   4.1786 0.0417472 *
## runs         1  40.302  40.302 176.0104 < 2.2e-16 ***
## home.runs    1   1.517   1.517   6.6247 0.0105028 *
## rbi          1   3.203   3.203  13.9904 0.0002174 ***
## Residuals  323  73.959   0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reduced Model w/o home.runs and bat.av

```
bball.lm3b = lm(log(salary)~free.elig+free.agent+arb.elig+arb+on.base+runs+rbi,data=bball,subset=not_outlier)
summary(bball.lm3b) # on.base NOT stat significant
```

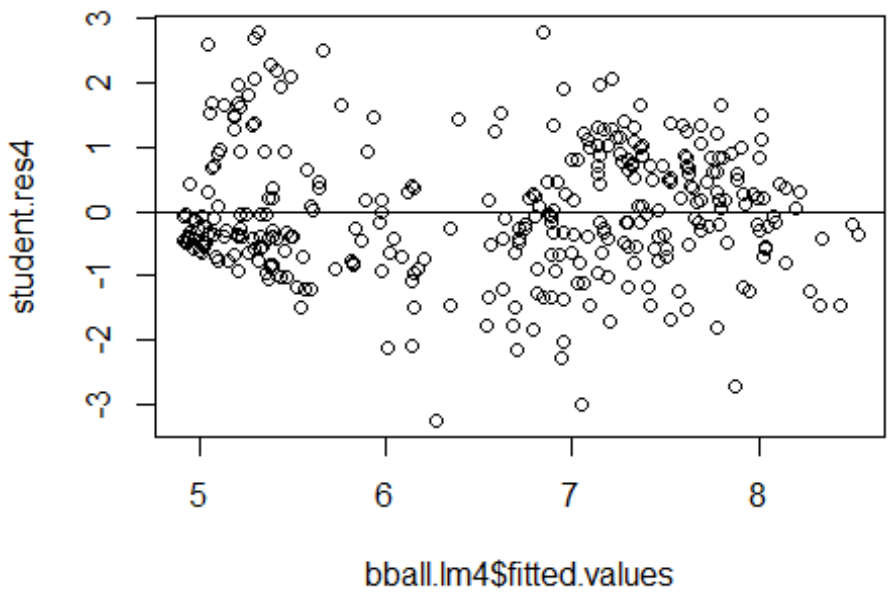
```
##
## Call:
## lm(formula = log(salary) ~ free.elig + free.agent + arb.elig +
##     arb + on.base + runs + rbi, data = bball, subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40839 -0.28611 -0.07282  0.32843  1.31939
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.029440   0.193968  25.929  < 2e-16 ***
## free.elig1   1.718820   0.071711  23.969  < 2e-16 ***
```

```
## free.agent1 -0.248658   0.093902  -2.648  0.00849 **
## arb.elig1    1.418965   0.080095  17.716  < 2e-16 ***
## arb1        -0.134440   0.165754  -0.811  0.41791
## on.base     -0.437039   0.650805  -0.672  0.50236
## runs         0.009226   0.001814   5.087 6.16e-07 ***
## rbi          0.007323   0.001659   4.414 1.39e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4788 on 325 degrees of freedom
## Multiple R-squared:  0.8349, Adjusted R-squared:  0.8314
## F-statistic: 234.8 on 7 and 325 DF,  p-value: < 2.2e-16

vif(bball.lm3b) # on.base went from 3.45 to 1.38

##  free.elig free.agent   arb.elig        arb    on.base       runs
##   1.782640   1.294851   1.446796   1.162518   1.375549   4.038075
##        rbi
##   3.495335

# Determines observations we want to include
not_outlier = which(abs(student.res2) <= 3)
# Reduced model w/o bat.av, home.runs, on.base
bball.lm4 = lm(log(salary) ~ free.elig + free.agent + arb.elig + arb + runs +
   rbi, data = bball, subset = not_outlier)
summary(bball.lm4)

##
## Call:
## lm(formula = log(salary) ~ free.elig + free.agent + arb.elig +
##     arb + runs + rbi, data = bball, subset = not_outlier)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4357 -0.2888 -0.0740  0.3298  1.3109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.904044   0.052446  93.506  < 2e-16 ***
## free.elig1   1.717885   0.071637  23.980  < 2e-16 ***
## free.agent1 -0.245474   0.093703  -2.620  0.00921 **
## arb.elig1    1.422050   0.079895  17.799  < 2e-16 ***
## arb1        -0.134232   0.165614  -0.811  0.41824
## runs         0.008803   0.001699   5.180 3.89e-07 ***
## rbi          0.007389   0.001655   4.465 1.11e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4784 on 326 degrees of freedom
## Multiple R-squared:  0.8347, Adjusted R-squared:  0.8317
## F-statistic: 274.4 on 6 and 326 DF,  p-value: < 2.2e-16

vif(bball.lm4)

##  free.elig free.agent   arb.elig        arb       runs        rbi
##   1.781968   1.291550   1.442037   1.162514   3.551075   3.483123

anova(bball.lm4)
```

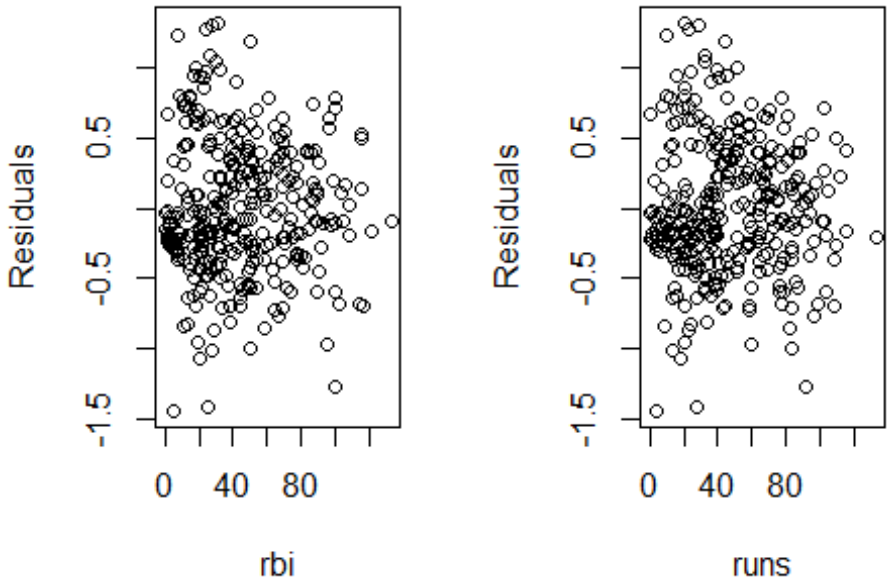
```
## Analysis of Variance Table
##
## Response: log(salary)
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## free.elig    1 178.890 178.890 781.6841 < 2.2e-16 ***
## free.agent   1   6.668   6.668  29.1357 1.301e-07 ***
## arb.elig     1 137.247 137.247 599.7205 < 2.2e-16 ***
## arb          1   0.047   0.047   0.2044    0.6515
## runs         1  49.324  49.324 215.5259 < 2.2e-16 ***
## rbi          1   4.562   4.562  19.9341 1.106e-05 ***
## Residuals  326  74.606   0.229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$F^* = \frac{(SSE(Model2) - SSE(Model1))/(p2 - p1)}{SSE(Model1)/(n - (p2 + 1))} = \frac{(74.6 - 74)/3}{74/323} = 0.87$$

```
Fstar = ((74.6-74)/3)/(74/323)
Fstar

## [1] 0.872973

library(MASS)
student.res4 =studres(bball.lm4)
plot(bball.lm4$fitted.values,student.res4)
abline(0,0)
```

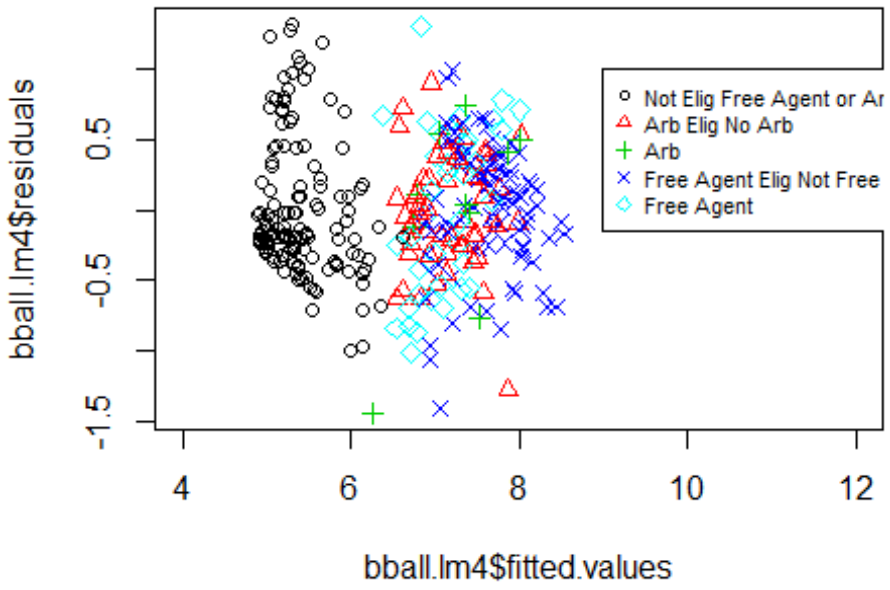



```
res=bball.lm4$residuals #Extracting the residuals
par(mfrow=c(1,2)) # Includes two plots in one figure
plot(bball$rbi[not_outlier],res, xlab='rbi',ylab='Residuals')
plot(bball$runs[not_outlier],res, xlab='runs',ylab='Residuals')
```




```
## Interactions with Contractual Variables

#Note the length of this vector is set to the number of observations in the original dataset
cont.stat = vector(mode="character",length=dim(bball)[1])

for ( i in 1:length(cont.stat)) {
  if ((bball$free.elig[i]=="0")&(bball$arb.elig[i]=="0")) {
    cont.stat[i] = "1"
  }
  if ((bball$arb.elig[i]=="1")&(bball$arb[i]=="0")) {
    cont.stat[i] = "2"
  }
  if ((bball$arb.elig[i]=="1")&(bball$arb[i]=="1")) {
    cont.stat[i] = "3"
  }
  if ((bball$free.elig[i]=="1")&(bball$free.agent[i]=="0")) {
    cont.stat[i] = "4"
  }
  if ((bball$free.elig[i]=="1")&(bball$free.agent[i]=="1")) {
    cont.stat[i] = "5"
  }
}

## Another way via Indexing

cont.stat2=NULL
cont.stat2[bball$free.elig==0 & bball$arb.elig==0] = 1
cont.stat2[bball$arb.elig==1 & bball$arb==0] = 2
cont.stat2[bball$arb.elig==1 & bball$arb==1] = 3
cont.stat2[bball$free.elig==1 & bball$free.agent==0] = 4
cont.stat2[bball$free.elig==1 & bball$free.agent==1] = 5

#Check if these two ways are equivalent:
all(as.numeric(cont.stat)==cont.stat2)

## [1] TRUE

plot(bball.lm4$fitted.values, bball.lm4$residuals, col = cont.stat[not_outlier],
    pch = as.numeric(cont.stat[not_outlier]), xlim = c(4, 12))
```

```
legend(9, 1, c("Not Elig Free Agent or Arb", "Arb Elig No Arb", "Arb", "Free Agent Elig Not Free Agent",
    "Free Agent"), cex = 0.7, col = c(1, 2, 3, 4, 5), pch = c(1, 2, 3, 4, 5))
```




```
# Model with interactions
bball.lm5 = lm(log(salary) ~ free.elig * runs + free.agent * runs + arb.elig *
    runs + arb * runs + free.elig * rbi + free.agent * rbi + arb.elig * rbi +
    arb * rbi, data = bball, subset = not_outlier)

summary(bball.lm5)

##
## Call:
## lm(formula = log(salary) ~ free.elig * runs + free.agent * runs +
##     arb.elig * runs + arb * runs + free.elig * rbi + free.agent *
##     rbi + arb.elig * rbi + arb * rbi, data = bball, subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41496 -0.31130 -0.05962  0.28318  1.48209
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       5.0223496  0.0650315  77.229  < 2e-16 ***
## free.elig1        1.6191259  0.1386085  11.681  < 2e-16 ***
## runs              0.0086673  0.0031495   2.752 0.006264 **
## free.agent1      -0.7540309  0.2075931  -3.632 0.000327 ***
## arb.elig1         1.2982852  0.1631522   7.958 3.12e-14 ***
## arb1             -0.9373129  0.3715933  -2.522 0.012143 *
## rbi               0.0035342  0.0031813   1.111 0.267443
## free.elig1:runs  -0.0001686  0.0040682  -0.041 0.966976
## runs:free.agent1  0.0044133  0.0058824   0.750 0.453651
## runs:arb.elig1    0.0031094  0.0053501   0.581 0.561524
## runs:arb1        -0.0095552  0.0084232  -1.134 0.257483
## free.elig1:rbi    0.0038458  0.0040339   0.953 0.341125
## free.agent1:rbi   0.0066181  0.0052791   1.254 0.210895
## arb.elig1:rbi     0.0008919  0.0051046   0.175 0.861412
## arb1:rbi          0.0238935  0.0094066   2.540 0.011558 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4645 on 318 degrees of freedom
## Multiple R-squared:  0.848,  Adjusted R-squared:  0.8413
## F-statistic: 126.7 on 14 and 318 DF,  p-value: < 2.2e-16
```

```
anova(bball.lm5)
```

```
## Analysis of Variance Table
##
## Response: log(salary)
##                 Df  Sum Sq Mean Sq  F value    Pr(>F)
## free.elig        1 178.890 178.890 829.2835 < 2.2e-16 ***
## runs             1 111.061 111.061 514.8449 < 2.2e-16 ***
## free.agent       1   0.568   0.568   2.6322   0.10571
## arb.elig         1  81.434  81.434 377.5058 < 2.2e-16 ***
## arb              1   0.223   0.223   1.0334   0.31014
## rbi              1   4.562   4.562  21.1480 6.145e-06 ***
## free.elig:runs   1   0.619   0.619   2.8675   0.09136 .
## runs:free.agent  1   1.696   1.696   7.8617   0.00536 **
## runs:arb.elig    1   1.014   1.014   4.6988   0.03093 *
## runs:arb         1   0.508   0.508   2.3558   0.12581
## free.elig:rbi    1   0.190   0.190   0.8798   0.34897
## free.agent:rbi   1   0.339   0.339   1.5716   0.21089
## arb.elig:rbi     1   0.251   0.251   1.1640   0.28145
## arb:rbi          1   1.392   1.392   6.4520   0.01156 *
## Residuals      318  68.598   0.216
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# FINAL Model with only rbi interactions
bball.lm6 = lm(log(salary) ~ runs + free.elig * rbi + free.agent * rbi + arb.elig *
   rbi + arb * rbi, data = bball, subset = not_outlier)
```

```
summary(bball.lm6)
```

```
##
## Call:
## lm(formula = log(salary) ~ runs + free.elig * rbi + free.agent *
##     rbi + arb.elig * rbi + arb * rbi, data = bball, subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40857 -0.31281 -0.06346  0.28453  1.46610
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.019323   0.062507  80.301  < 2e-16 ***
## runs            0.009137   0.001661   5.500 7.73e-08 ***
## free.elig1      1.609215   0.128454  12.528  < 2e-16 ***
## rbi             0.003124   0.002159   1.447 0.148832
## free.agent1    -0.689096   0.186951  -3.686 0.000267 ***
## arb.elig1       1.339898   0.149158   8.983  < 2e-16 ***
## arb1           -1.050260   0.356711  -2.944 0.003472 **
## free.elig1:rbi  0.003802   0.002329   1.632 0.103575
## rbi:free.agent1 0.009695   0.003308   2.931 0.003622 **
```

```
## rbi:arb.elig1    0.003270   0.002848   1.148 0.251801
## rbi:arb1         0.015280   0.005457   2.800 0.005415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4629 on 322 degrees of freedom
## Multiple R-squared:  0.8471, Adjusted R-squared:  0.8424
## F-statistic: 178.4 on 10 and 322 DF,  p-value: < 2.2e-16
```

```
anova(bball.lm6)
```

```
## Analysis of Variance Table
##
## Response: log(salary)
##                Df  Sum Sq Mean Sq  F value    Pr(>F)
## runs            1 199.092 199.092 929.0473 < 2.2e-16 ***
## free.elig       1  90.859  90.859 423.9847 < 2.2e-16 ***
## rbi             1   8.585   8.585  40.0596 8.256e-10 ***
## free.agent      1   0.600   0.600   2.8015  0.095146 .
## arb.elig        1  77.452  77.452 361.4217 < 2.2e-16 ***
## arb             1   0.150   0.150   0.7015  0.402886
## free.elig:rbi   1   0.940   0.940   4.3864  0.037008 *
## rbi:free.agent  1   1.841   1.841   8.5927  0.003617 **
## rbi:arb.elig    1   1.140   1.140   5.3217  0.021695 *
## rbi:arb         1   1.680   1.680   7.8414  0.005415 **
## Residuals     322  69.004   0.214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Interpreting the Final Model
```

```
summary(bball.lm6)
```

```
##
## Call:
## lm(formula = log(salary) ~ runs + free.elig * rbi + free.agent *
##     rbi + arb.elig * rbi + arb * rbi, data = bball, subset = not_outlier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40857 -0.31281 -0.06346  0.28453  1.46610
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.019323   0.062507  80.301  < 2e-16 ***
## runs            0.009137   0.001661   5.500 7.73e-08 ***
## free.elig1      1.609215   0.128454  12.528  < 2e-16 ***
## rbi             0.003124   0.002159   1.447 0.148832
## free.agent1    -0.689096   0.186951  -3.686 0.000267 ***
## arb.elig1       1.339898   0.149158   8.983  < 2e-16 ***
## arb1           -1.050260   0.356711  -2.944 0.003472 **
## free.elig1:rbi  0.003802   0.002329   1.632 0.103575
## rbi:free.agent1 0.009695   0.003308   2.931 0.003622 **
## rbi:arb.elig1   0.003270   0.002848   1.148 0.251801
## rbi:arb1        0.015280   0.005457   2.800 0.005415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4629 on 322 degrees of freedom
## Multiple R-squared:  0.8471, Adjusted R-squared:  0.8424
## F-statistic: 178.4 on 10 and 322 DF,  p-value: < 2.2e-16
```

Some interesting points on this model:

- runs and rbi are both positive, indicating that doing better increases your salary.
- Here the reference level for all contract variables is 0. So for example, the positive coefficient for free.elig indicates that being eligible to be a free agent (free.elig = 1) increases your log salary by 1.6.
- Similarly, being eligible for arbitration increases your salary. Interestingly, actually being a free agent or taking arbitration decreases your salary.
- Neither eligibility interacted strongly with rbi, the p-values are not significant and the coefficients are relatively small. So being better doesnâ™t change the boost you get from eligibility status.
- However, the interaction between performance and taking arbitration or becoming a free agent is significant. Since the effect is for taking these statuses, the positive coefficients suggest that if you do go into arbitration or become a free agent, you salary is improved by doing better.