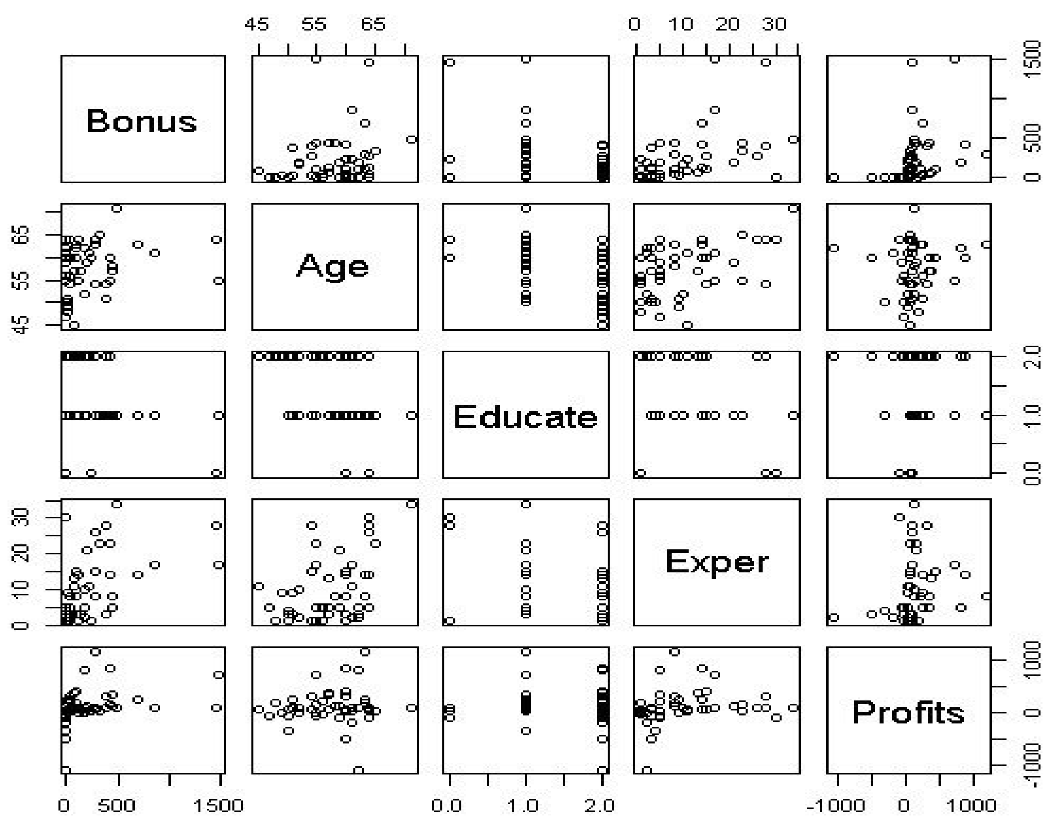
Executive compensation

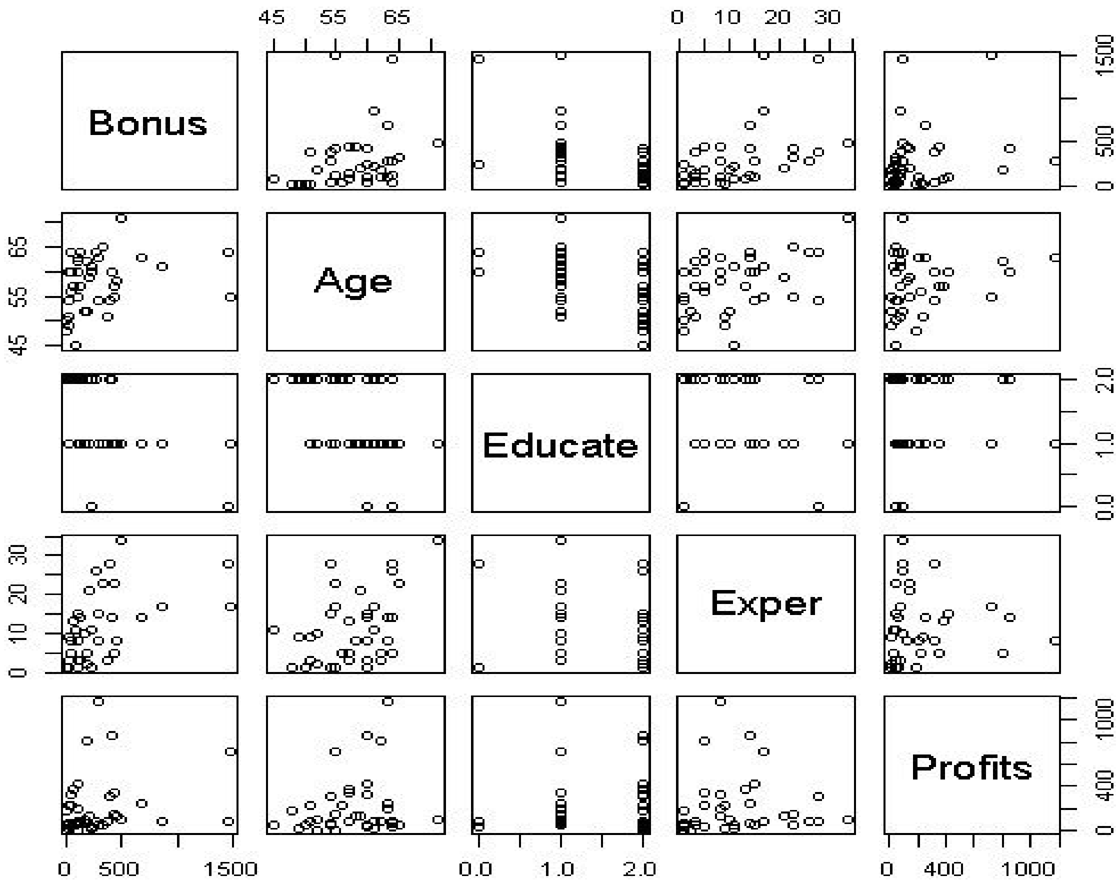
Executive compensation or bonus is composed of the financial compensation and other non-financial awards received by an executive from their firm for their service to the organization. It is typically a mixture of salary, bonuses, shares of or call options on the company stock, benefits, and perquisites, ideally configured to take into account government regulations, tax law, the desires of the organization and the executive, and rewards for performance.[1] It would be very interesting to analyze bonuses for CEOs and what factors does a bonus depends on. Let's consider the CEO dataset of compensation for CEOs of 48 U.S. companies (Appendix 1). As we can see from the table we have data for bonus, age, experience, education of CEOs and profit of the companies which are most obvious factors that could impact on bonuses. Variates of our model are:

Bonus paid ($1000's)
Age of CEO (years)
Education level (0 = no college/university degree, 1 = undergraduate degree, 2 = graduate degree)
Experience (years)
Profit (millions)

Before we dive into the regression analysis, it would be beneficial to make a scatter plot to reveal relations visually between the variates.



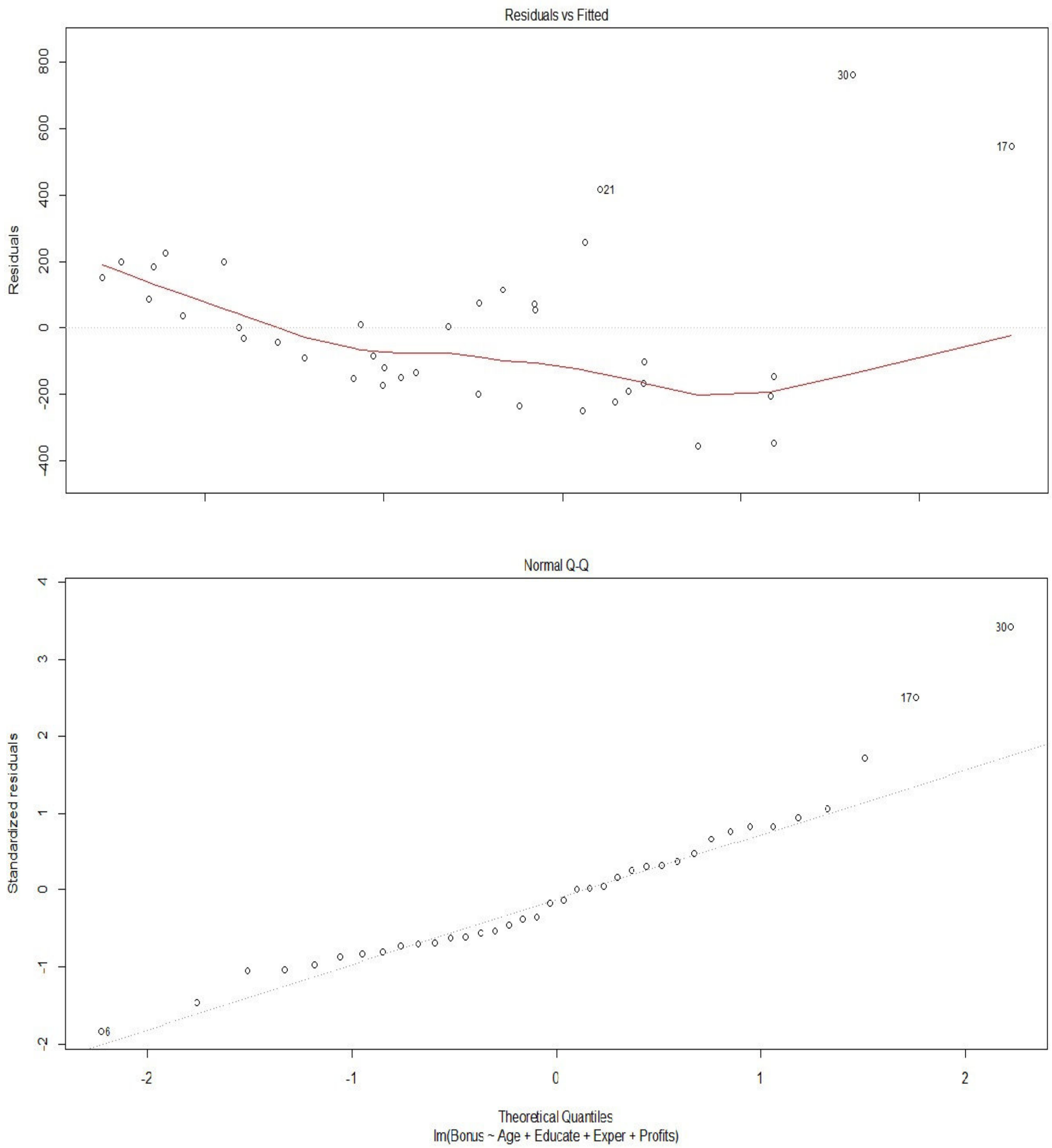
There is not obvious strong relationship between variates. However, it looks like there is a weak positive correlation between bonus and age, bonus and experience. Apparently the companies with negative profits give no bonuses to the CEOs. We can ignore these companies since our analysis will not hold for them. In other words, by restricting data in this way, we can make an inference only for companies with positive profits. So scatter plot after the restriction is as follows.



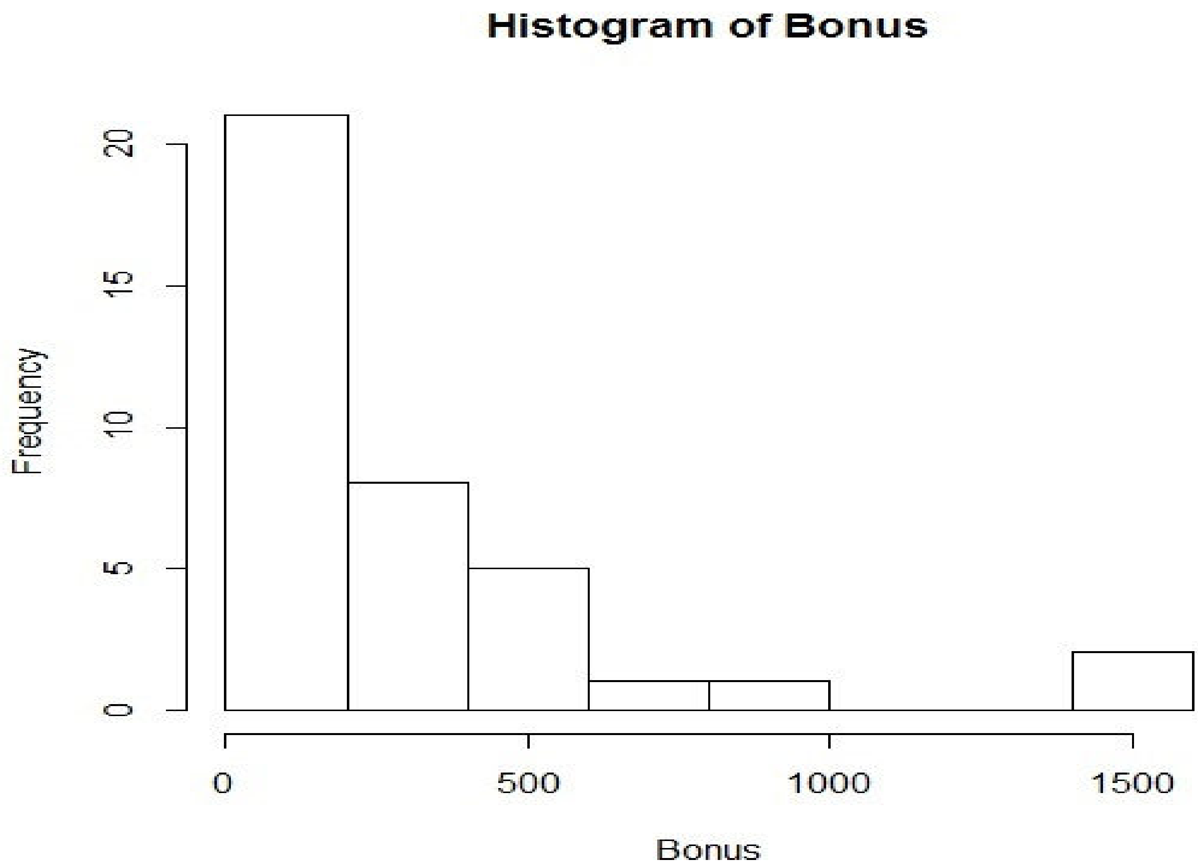
Restricting data almost did not change the relationships between variates. We can now fit the linear regression model to the revised dataset in R.

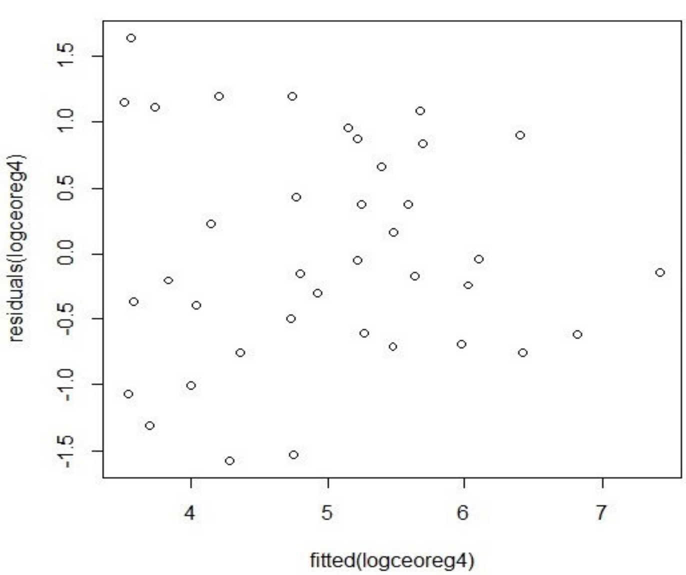
```
Call:
lm(formula = Bonus ~ Age + Educate + Exper + Profits)

Residuals:
    Min      1Q  Median      3Q     Max
-354.50 -162.39  -38.21  107.57  761.99

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1074.2378   525.5778   2.044 0.049010 *
Age          -10.7765     8.6398  -1.247 0.221065
Educate     -298.1100    75.2347  -3.962 0.000374 ***
```

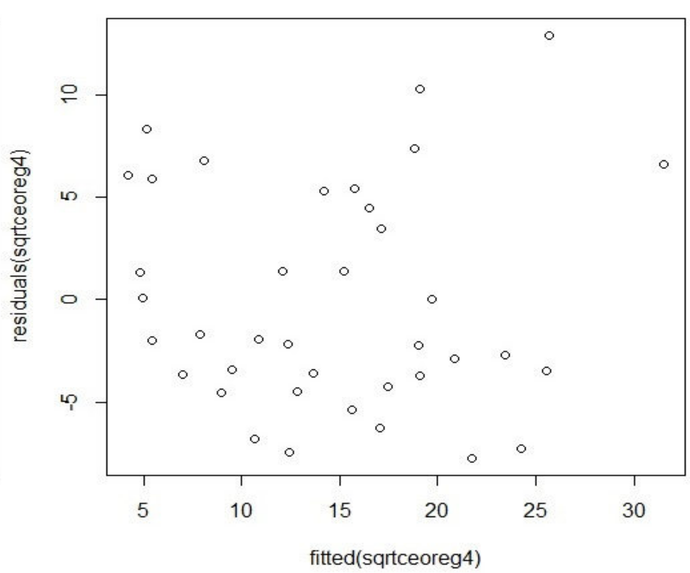
```
Exper        17.4020     5.1757   3.362 0.001968 **
Profits       0.3437     0.1599   2.150 0.039014 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 250.4 on 33 degrees of freedom
Multiple R-squared: 0.5281,     Adjusted R-squared: 0.4709
F-statistic: 9.232 on 4 and 33 DF,  p-value: 4.036e-05
```



Based on the R output, it seems that only Age is an insignificant variate and Education level has the biggest impact on Bonus which makes sense. In order to comment on the adequacy of the fitted model, we can create a plot of the residuals vs the fitted values for this model, as well as a QQ plot. The first plot reflects the problems of non-constant variance and presence of outliers. The second plot shows non-normality of the residuals. In order to stabilize the variance of the residuals we can use appropriate transformation of the response variate. According to the histogram of bonuses, the data is strongly positive skewed. We can use log or square root transformation to improve the model. So we will refit the model with log and square root transformation of response variable (Bonus).







```
Call:
lm(formula = log(Bonus) ~ Age + Educate + Exper + Profits)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5735 -0.6185 -0.1444  0.7931  1.6379 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  5.2342214  1.8569851   2.819  0.00809 ** 
Age          0.0048318  0.0305265   0.158  0.87520    
Educate     -1.0260544  0.2658212  -3.860  0.00050 ***
Exper        0.0630845  0.0182869   3.450  0.00155 ** 
Profits      0.0012037  0.0005649   2.131  0.04067 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8848 on 33 degrees of freedom
Multiple R-squared: 0.5842,	Adjusted R-squared: 0.5338 
F-statistic: 11.59 on 4 and 33 DF,  p-value: 5.468e-06

Call:
lm(formula = sqrt(Bonus) ~ Age + Educate + Exper + Profits)

Residuals:
   Min     1Q Median     3Q    Max 
-7.683 -3.681 -1.949  5.121 12.852 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 27.999296  12.054009   2.323 0.026500 *  
Age         -0.162225   0.198153  -0.819 0.418839    
Educate     -7.699462   1.725491  -4.462 8.91e-05 ***
Exper        0.467135   0.118703   3.935 0.000404 ***
Profits      0.008938   0.003667   2.437 0.020347 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.744 on 33 degrees of freedom
Multiple R-squared: 0.6135,	Adjusted R-squared: 0.5666 
F-statistic: 13.09 on 4 and 33 DF,  p-value: 1.716e-06
```

Based on the results, we can see that both transformations made a good improvement in the adequacy of our model. There are no longer any major outliers. Since the square root transformation has better R^2 we will proceed with the square root model. As we can see, age is considerably insignificant variate in square root model and we can remove age and refit the model.

```
Call:
lm(formula = sqrt(Bonus) ~ Educate + Exper + Profits)

Residuals:
   Min     1Q Median     3Q    Max
-7.934 -3.891 -1.396  5.229 14.011

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.462930   3.085590   5.984 9.04e-07 ***
Educate     -7.235518   1.621856  -4.461 8.48e-05 ***
Exper        0.432515   0.110376   3.919 0.000409 ***
Profits      0.008350   0.003579   2.333 0.025680 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.716 on 34 degrees of freedom
Multiple R-squared: 0.6056,     Adjusted R-squared: 0.5708
F-statistic: 17.41 on 3 and 34 DF,  p-value: 5.082e-07
```

As a result, we do not have any insignificant variates anymore, R^2 improved and p-value is considerably small. Now we have a reasonable model. If we perform leverage analysis for this model, we will see that there are some leverage points but there are not influential (Cook's Distance). Let's consider an example: Suppose a CEO with a graduate degree, 25 years of experience, and whose company made an annual profit of 89 million dollars receives a bonus of $700,000.

fit	lwr	upr
241.74	9.39	785.78

As we can see, the example $700.000 for this CEO is consistent with our model.

Appendix 1 *(CEO dataset)*

(Source: *Business Forecasting,* by J. Hanke and D Wichern, p. 544. Pearson Prentice Hall, 2009)

ROVSHAN ALIYEV
STUDENT PROJECT

REGRESSIAN ANALYSIS
SPRING 2014 SESSION

Bonus	Age	Educate	Exper	Profits
275	64	2	26	91
429	55	1	23	145
0	47	2	5	-47
325	65	1	23	44
105	63	1	8	201
25	54	2	1	71
0	61	2	2	-187
289	63	1	8	1166
69	57	2	13	377
38	56	2	5	224
129	57	2	3	79
11	48	2	1	189
0	50	1	4	-332
282	54	1	15	55
0	60	2	3	-507
423	60	2	14	856
20	49	2	9	14
0	56	2	1	-29
448	58	1	8	126
12	50	2	1	54
687	63	1	14	249
1452	64	0	28	91
37	60	2	8	322
489	71	1	34	99
0	64	0	30	-99
38	64	2	5	30
0	59	2	5	-85
862	61	1	17	82
221	61	2	11	27
0	55	2	5	-76
391	54	2	28	317
101	60	2	15	417
238	60	0	1	43
25	60	1	3	49
104	62	1	3	81
380	51	1	3	82
107	55	2	1	10
1487	55	1	17	715
198	59	1	21	136
15	51	2	9	237
0	62	2	2	-1086
174	52	1	10	98
80	45	2	11	48
0	50	2	3	-50
440	57	1	5	347
117	64	1	14	63
182	62	2	5	806
183	52	2	2	10

Appendix 2 *(R codes)*

```
A2<-read.table(file.choose(), header=TRUE, sep=",")
attach(A2)
ceo<-as.matrix(A2)
ceo<-subset(ceo, ceo[,5]>0)
Bonus=ceo[,1]
Age=ceo[,2]
Educate=ceo[,3]
Exper=ceo[,4]
Profits=ceo[,5]

ceoreg4<-lm(Bonus~Age+Educate+Exper+Profits)
summary(ceoreg4)
plot(ceoreg4)
hist(Bonus)

logceoreg4<-lm(log(Bonus)~Age+Educate+Exper+Profits)
summary(logceoreg4)
plot(fitted(logceoreg4),residuals(logceoreg4))
plot(logceoreg4)

sqrtceoreg4<-lm(sqrt(Bonus)~Age+Educate+Exper+Profits)
summary(sqrtceoreg4)
plot(fitted(sqrtceoreg4),residuals(sqrtceoreg4))
plot(sqrtceoreg4)

sqrtceoreg3<-lm(sqrt(Bonus)~Educate+Exper+Profits)
summary(sqrtceoreg3)
plot(sqrtceoreg3)

newdata = data.frame(Educate=2,Exper=25,Profits=89)
predict(sqrtceoreg3, newdata, interval="predict")^2
```

References

1. The complete guide to executive compensation By Bruce R. Ellig, 2002 (Wikipedia)
2. Applied Statistics study notes, Master of Actuarial Science Program, University of Waterloo