Term Project of Yiğit Güler

Yiğit Güler

08 06 2021

library(car)

## Loading required package: carData

library(rmarkdown)  
library(haven)  
library(effects)

## Registered S3 methods overwritten by 'lme4':  
## method from  
## cooks.distance.influence.merMod car   
## influence.merMod car   
## dfbeta.influence.merMod car   
## dfbetas.influence.merMod car

## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.4

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':  
##   
## recode

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

ceosal <- read\_dta("ceosal.DTA")  
  
summary(ceosal)

## salary pcsalary sales roe   
## Min. : 223 Min. :-61.00 Min. : 175.2 Min. : 0.50   
## 1st Qu.: 736 1st Qu.: -1.00 1st Qu.: 2210.3 1st Qu.:12.40   
## Median : 1039 Median : 9.00 Median : 3705.2 Median :15.50   
## Mean : 1281 Mean : 13.28 Mean : 6923.8 Mean :17.18   
## 3rd Qu.: 1407 3rd Qu.: 20.00 3rd Qu.: 7177.0 3rd Qu.:20.00   
## Max. :14822 Max. :212.00 Max. :97649.9 Max. :56.30   
## pcroe ros indus finance   
## Min. :-98.9 Min. :-58.0 Min. :0.0000 Min. :0.0000   
## 1st Qu.:-21.2 1st Qu.: 21.0 1st Qu.:0.0000 1st Qu.:0.0000   
## Median : -3.0 Median : 52.0 Median :0.0000 Median :0.0000   
## Mean : 10.8 Mean : 61.8 Mean :0.3206 Mean :0.2201   
## 3rd Qu.: 19.5 3rd Qu.: 81.0 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :977.0 Max. :418.0 Max. :1.0000 Max. :1.0000   
## consprod utility lsalary lsales   
## Min. :0.0000 Min. :0.0000 Min. :5.407 Min. : 5.166   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:6.601 1st Qu.: 7.701   
## Median :0.0000 Median :0.0000 Median :6.946 Median : 8.217   
## Mean :0.2871 Mean :0.1722 Mean :6.950 Mean : 8.292   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:7.249 3rd Qu.: 8.879   
## Max. :1.0000 Max. :1.0000 Max. :9.604 Max. :11.489

1. We have lower p-value than our threshold so we accept the the null hypothesis in our model **ros** is not significant and can be accepted as zero.

reg.ceo <- lm(lsalary ~ lsales + roe + ros, data = ceosal)  
reg.ceo

##   
## Call:  
## lm(formula = lsalary ~ lsales + roe + ros, data = ceosal)  
##   
## Coefficients:  
## (Intercept) lsales roe ros   
## 4.3117125 0.2803149 0.0174168 0.0002417

summary(reg.ceo)

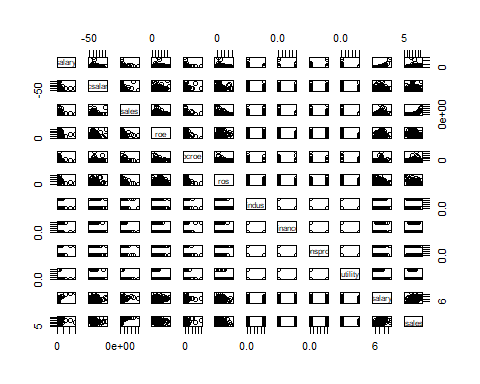
##   
## Call:  
## lm(formula = lsalary ~ lsales + roe + ros, data = ceosal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.96060 -0.27144 -0.03264 0.22563 2.79805   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.3117125 0.3154329 13.669 < 2e-16 \*\*\*  
## lsales 0.2803149 0.0353200 7.936 1.34e-13 \*\*\*  
## roe 0.0174168 0.0040923 4.256 3.17e-05 \*\*\*  
## ros 0.0002417 0.0005418 0.446 0.656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4832 on 205 degrees of freedom  
## Multiple R-squared: 0.2827, Adjusted R-squared: 0.2722   
## F-statistic: 26.93 on 3 and 205 DF, p-value: 1.001e-14

1. We have 4 variables showing the industry of the companies which are showed with 0 and 1.

summary(ceosal)

## salary pcsalary sales roe   
## Min. : 223 Min. :-61.00 Min. : 175.2 Min. : 0.50   
## 1st Qu.: 736 1st Qu.: -1.00 1st Qu.: 2210.3 1st Qu.:12.40   
## Median : 1039 Median : 9.00 Median : 3705.2 Median :15.50   
## Mean : 1281 Mean : 13.28 Mean : 6923.8 Mean :17.18   
## 3rd Qu.: 1407 3rd Qu.: 20.00 3rd Qu.: 7177.0 3rd Qu.:20.00   
## Max. :14822 Max. :212.00 Max. :97649.9 Max. :56.30   
## pcroe ros indus finance   
## Min. :-98.9 Min. :-58.0 Min. :0.0000 Min. :0.0000   
## 1st Qu.:-21.2 1st Qu.: 21.0 1st Qu.:0.0000 1st Qu.:0.0000   
## Median : -3.0 Median : 52.0 Median :0.0000 Median :0.0000   
## Mean : 10.8 Mean : 61.8 Mean :0.3206 Mean :0.2201   
## 3rd Qu.: 19.5 3rd Qu.: 81.0 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :977.0 Max. :418.0 Max. :1.0000 Max. :1.0000   
## consprod utility lsalary lsales   
## Min. :0.0000 Min. :0.0000 Min. :5.407 Min. : 5.166   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:6.601 1st Qu.: 7.701   
## Median :0.0000 Median :0.0000 Median :6.946 Median : 8.217   
## Mean :0.2871 Mean :0.1722 Mean :6.950 Mean : 8.292   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:7.249 3rd Qu.: 8.879   
## Max. :1.0000 Max. :1.0000 Max. :9.604 Max. :11.489

pairs(ceosal)



glimpse(ceosal)

## Rows: 209  
## Columns: 12  
## $ salary <dbl> 1095, 1001, 1122, 578, 1368, 1145, 1078, 1094, 1237, 833, ...  
## $ pcsalary <dbl> 20, 32, 9, -9, 7, 5, 10, 7, 16, 5, 7, -3, -9, 9, 49, 4, 12...  
## $ sales <dbl> 27595.0, 9958.0, 6125.9, 16246.0, 21783.2, 6021.4, 2266.7,...  
## $ roe <dbl> 14.1, 10.9, 23.5, 5.9, 13.8, 20.0, 16.4, 16.3, 10.5, 26.3,...  
## $ pcroe <dbl> 106.4, -30.6, -16.3, -25.7, -3.0, 1.0, -5.9, -1.6, -70.2, ...  
## $ ros <dbl> 191, 13, 14, -21, 56, 55, 62, 44, 37, 37, 109, -10, 41, 44...  
## $ indus <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...  
## $ finance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...  
## $ consprod <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...  
## $ utility <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...  
## $ lsalary <dbl> 6.998509, 6.908755, 7.022868, 6.359574, 7.221105, 7.043160...  
## $ lsales <dbl> 10.225389, 9.206132, 8.720281, 9.695602, 9.988894, 8.70307...

1. If we look at the F-stat we can say that our model is significant with 205 degrees of freedom. The Coefficients are significant except **ros**. Adjusted R-squared is 0.2722. That means we can explain %27,22 of the model.

summary(reg.ceo)

##   
## Call:  
## lm(formula = lsalary ~ lsales + roe + ros, data = ceosal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.96060 -0.27144 -0.03264 0.22563 2.79805   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.3117125 0.3154329 13.669 < 2e-16 \*\*\*  
## lsales 0.2803149 0.0353200 7.936 1.34e-13 \*\*\*  
## roe 0.0174168 0.0040923 4.256 3.17e-05 \*\*\*  
## ros 0.0002417 0.0005418 0.446 0.656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4832 on 205 degrees of freedom  
## Multiple R-squared: 0.2827, Adjusted R-squared: 0.2722   
## F-statistic: 26.93 on 3 and 205 DF, p-value: 1.001e-14

salvars <- coef(reg.ceo)  
salvars

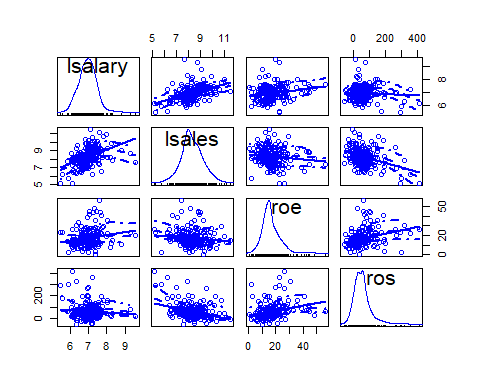
## (Intercept) lsales roe ros   
## 4.3117124632 0.2803148895 0.0174167522 0.0002416552

cor.data <- cor(ceosal[,c(4,6,11,12)])  
print(cor.data)

## roe ros lsalary lsales  
## roe 1.0000000 0.27491881 0.20849920 -0.1225531  
## ros 0.2749188 1.00000000 -0.07456453 -0.3503296  
## lsalary 0.2084992 -0.07456453 1.00000000 0.4591482  
## lsales -0.1225531 -0.35032965 0.45914817 1.0000000

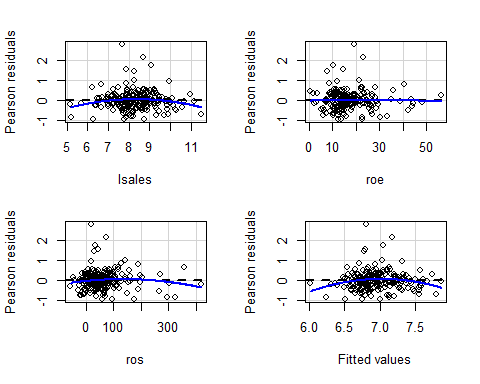
1. We can observe that **ros** and **lsalary**, does not neccesarly have a correlation.

scatterplotMatrix(~ lsalary + lsales + roe + ros, data = ceosal)



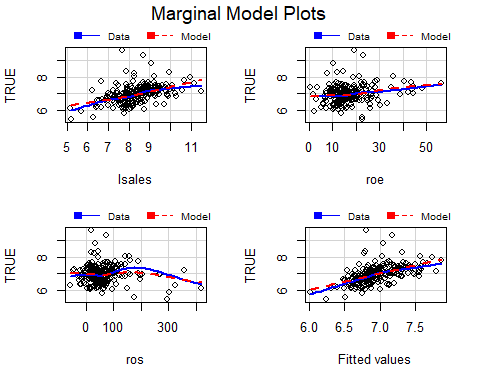
1. Below we can see how the variables placed in our model we can see that **ros** is behaving differently than the others. Our estimation with the variables are not so different than the actual data. The outlier effects is not too heavy to bear.

residualPlots(reg.ceo)

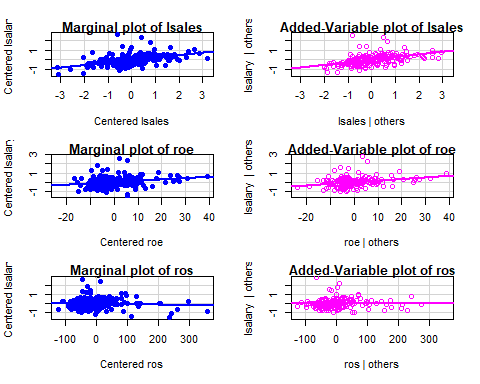


## Test stat Pr(>|Test stat|)   
## lsales -2.2358 0.026451 \*   
## roe -0.2604 0.794824   
## ros -1.4548 0.147249   
## Tukey test -2.7005 0.006923 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

marginalModelPlots(reg.ceo)

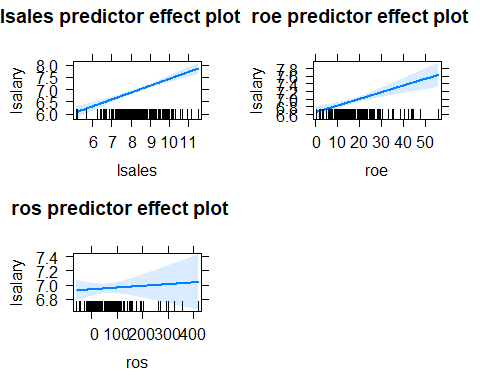


mcPlots(reg.ceo,~lsales + roe + ros, overlaid=FALSE)



We see how well we can estimate the model with taken predictors.

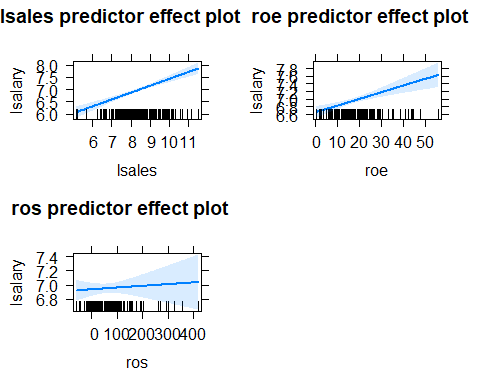
plot(predictorEffects(reg.ceo))



summary(reg.ceo)

##   
## Call:  
## lm(formula = lsalary ~ lsales + roe + ros, data = ceosal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.96060 -0.27144 -0.03264 0.22563 2.79805   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.3117125 0.3154329 13.669 < 2e-16 \*\*\*  
## lsales 0.2803149 0.0353200 7.936 1.34e-13 \*\*\*  
## roe 0.0174168 0.0040923 4.256 3.17e-05 \*\*\*  
## ros 0.0002417 0.0005418 0.446 0.656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4832 on 205 degrees of freedom  
## Multiple R-squared: 0.2827, Adjusted R-squared: 0.2722   
## F-statistic: 26.93 on 3 and 205 DF, p-value: 1.001e-14

plot(predictorEffects(reg.ceo))



# Equation of the model

With using 3 predictors we come up with the model below. Using **reg.ceo**. The interpretaions and way of estimations are above.

1. Since ros is insignificant in our model it does not change the salary and we take it as zero. But 50 units in increase would result to 0.0002417*50 = 0.012085 Of course in the model we have log-level so 100*0.0002417x50 = 1.2085 per cent change It is not that big of a change considering it is also **insignificant**.
2. We can see that we conclude with cor = -0.0336819 at the given confidence intervals. In the light of the previous findings we can accept the Ho in this case. **ros** is not significant in our model. Ergo we cannot use it nor can say it has a positive effect on **salary**.

aa <- ceosal[,c(1,6)]  
  
cor.test(aa$salary,aa$ros, conf.level = 0.90)

##   
## Pearson's product-moment correlation  
##   
## data: aa$salary and aa$ros  
## t = -0.48487, df = 207, p-value = 0.6283  
## alternative hypothesis: true correlation is not equal to 0  
## 90 percent confidence interval:  
## -0.14721939 0.08073166  
## sample estimates:  
## cor   
## -0.0336819

cor.test(aa$salary,aa$ros, conf.level = 0.95)

##   
## Pearson's product-moment correlation  
##   
## data: aa$salary and aa$ros  
## t = -0.48487, df = 207, p-value = 0.6283  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1686257 0.1025013  
## sample estimates:  
## cor   
## -0.0336819

cor.test(aa$salary,aa$ros, conf.level = 0.99)

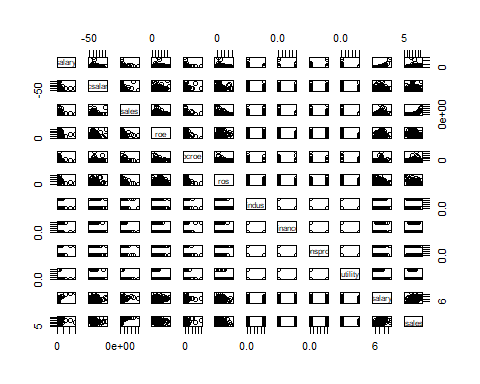
##   
## Pearson's product-moment correlation  
##   
## data: aa$salary and aa$ros  
## t = -0.48487, df = 207, p-value = 0.6283  
## alternative hypothesis: true correlation is not equal to 0  
## 99 percent confidence interval:  
## -0.2099903 0.1447481  
## sample estimates:  
## cor   
## -0.0336819

1. No I would not because of the fact that given findings. There is no situation we can accept **ros** to explain compensation for the CEOs. Since we cannot accept it any model as a variable. The significancy of **ros** is making it impossible to include into our models. Ros predictor is not a good predictor.

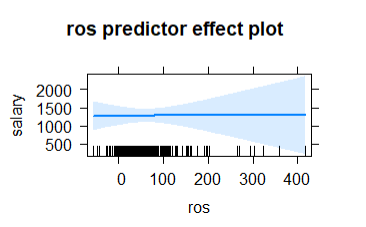
reg.ceo <- lm(lsalary ~ lsales + roe + ros, data = ceosal)  
  
summary(reg.ceo)

##   
## Call:  
## lm(formula = lsalary ~ lsales + roe + ros, data = ceosal)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.96060 -0.27144 -0.03264 0.22563 2.79805   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.3117125 0.3154329 13.669 < 2e-16 \*\*\*  
## lsales 0.2803149 0.0353200 7.936 1.34e-13 \*\*\*  
## roe 0.0174168 0.0040923 4.256 3.17e-05 \*\*\*  
## ros 0.0002417 0.0005418 0.446 0.656   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4832 on 205 degrees of freedom  
## Multiple R-squared: 0.2827, Adjusted R-squared: 0.2722   
## F-statistic: 26.93 on 3 and 205 DF, p-value: 1.001e-14

pairs(ceosal)



# ros image



image