Regression Analysis on FActors Affecting Global Crude Oil

2023-02-08

#Data Preprocessing

```
da <- read.csv("First.csv")
df<-da
str(df)
```

```
## 'data.frame':    38 obs. of  6 variables:
##  $ Oprice  : num  27 15.1 19.2 16 19.6 ...
##  $ Oprod   : num  19.7 20.6 20.7 21.4 21.8 ...
##  $ Gprice  : num  2.28 1.76 1.7 1.89 1.92 2.04 2 2.07 2.15 1.88 ...
##  $ Gprod   : num  8.04 8.51 8.76 9.2 9.22 ...
##  $ Oreserve: num  699 700 889 907 1002 ...
##  $ Wrate   : num  3.76 3.24 3.58 4.48 3.57 2.83 1.22 1.68 1.62 3.04 ...
```

##Correlation

```
cor(df)
```

```
##              Oprice      Oprod    Gprice      Gprod   Oreserve       Wrate
## Oprice   1.00000000  0.74841465 0.4524395  0.75338058  0.8141924  0.05410912
## Oprod    0.74841465  1.00000000 0.3645036  0.96709783  0.9491385 -0.03757295
## Gprice   0.45243949  0.36450363 1.0000000  0.38869656  0.2672351  0.18667611
## Gprod    0.75338058  0.96709783 0.3886966  1.00000000  0.8971570 -0.03501935
## Oreserve 0.81419238  0.94913847 0.2672351  0.89715705  1.0000000 -0.10453334
## Wrate    0.05410912 -0.03757295 0.1866761 -0.03501935 -0.1045333  1.00000000
```

```
library(ggcorrplot)
```

```
## Loading required package: ggplot2
```
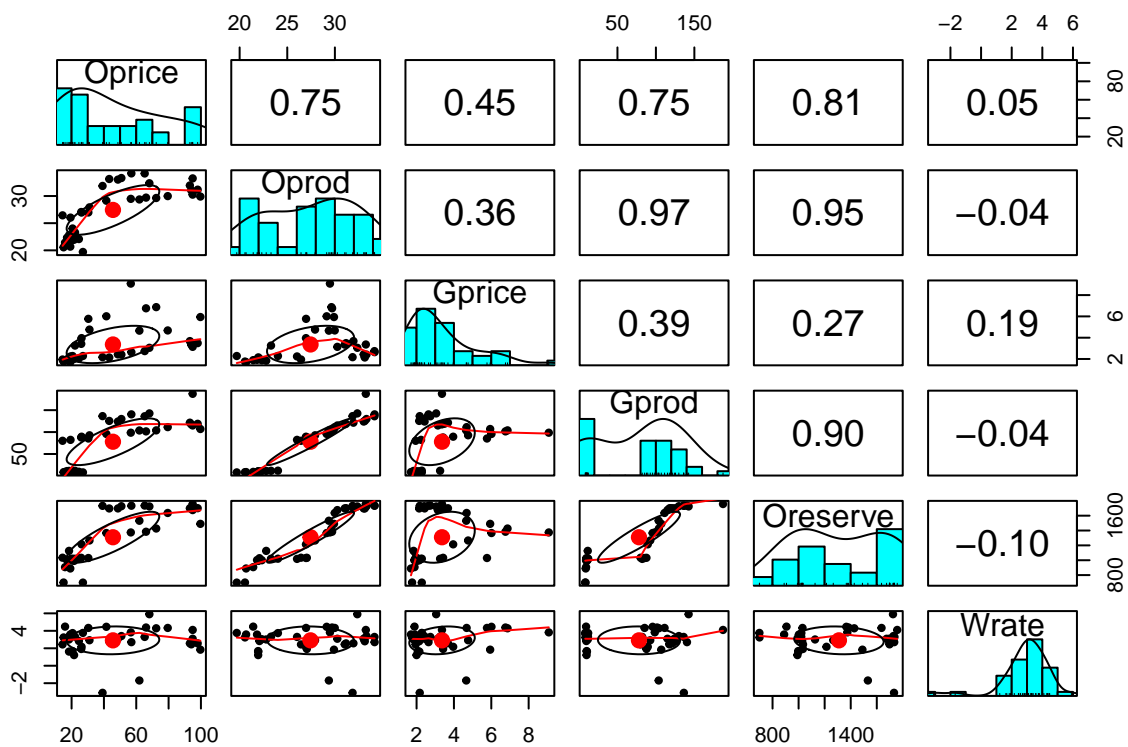
```
ggcorrplot(cor(df))
```
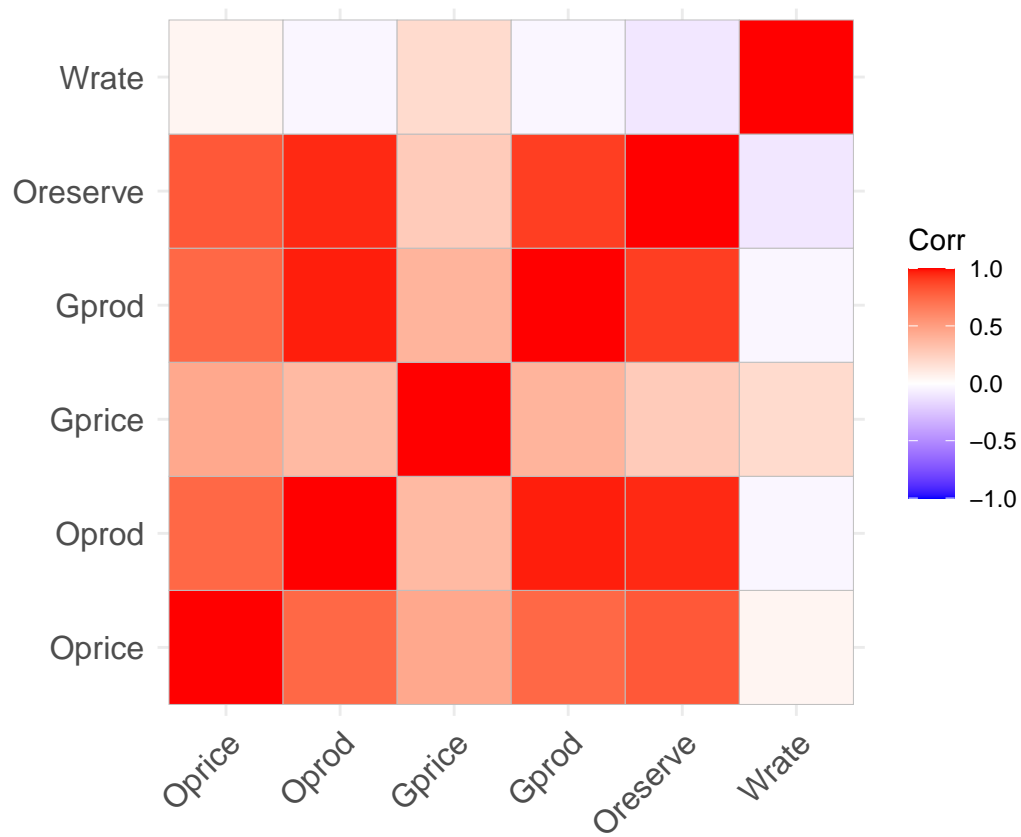
```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```
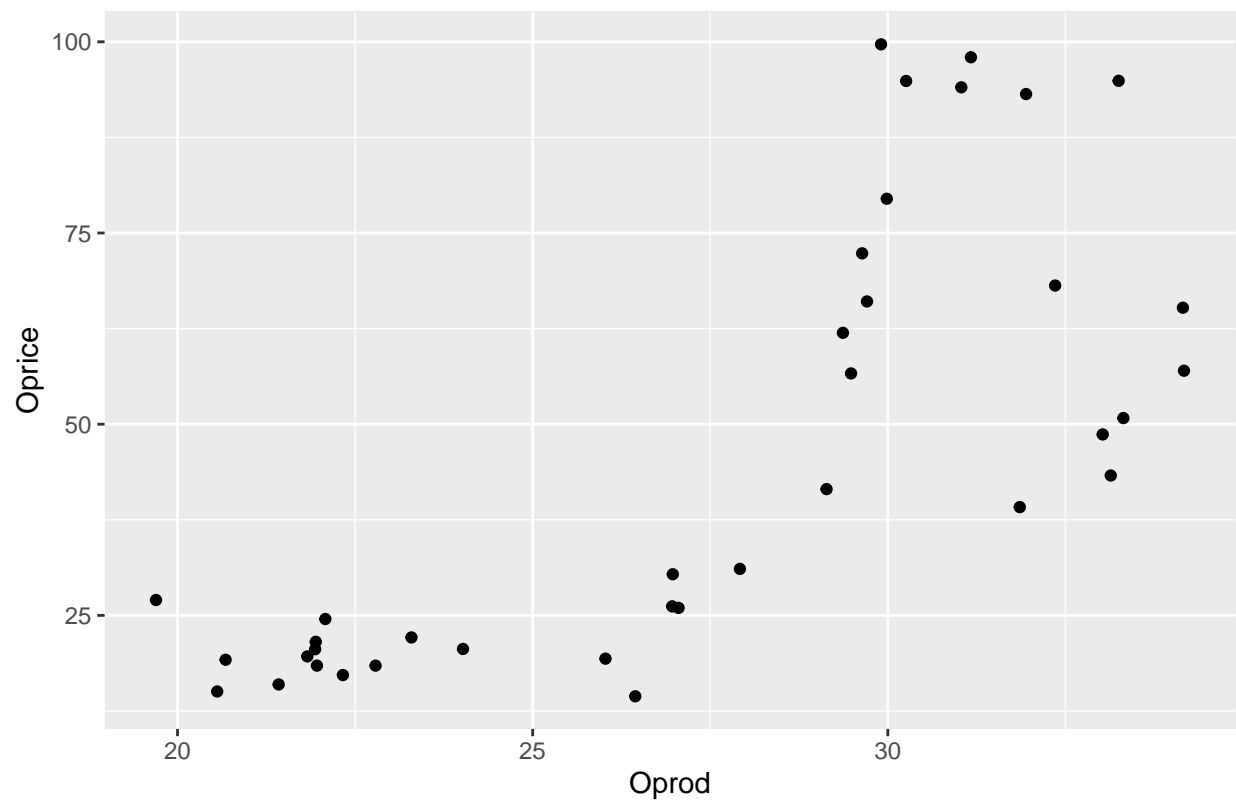
```
library(ggplot2)
pairs.panels(df)
```
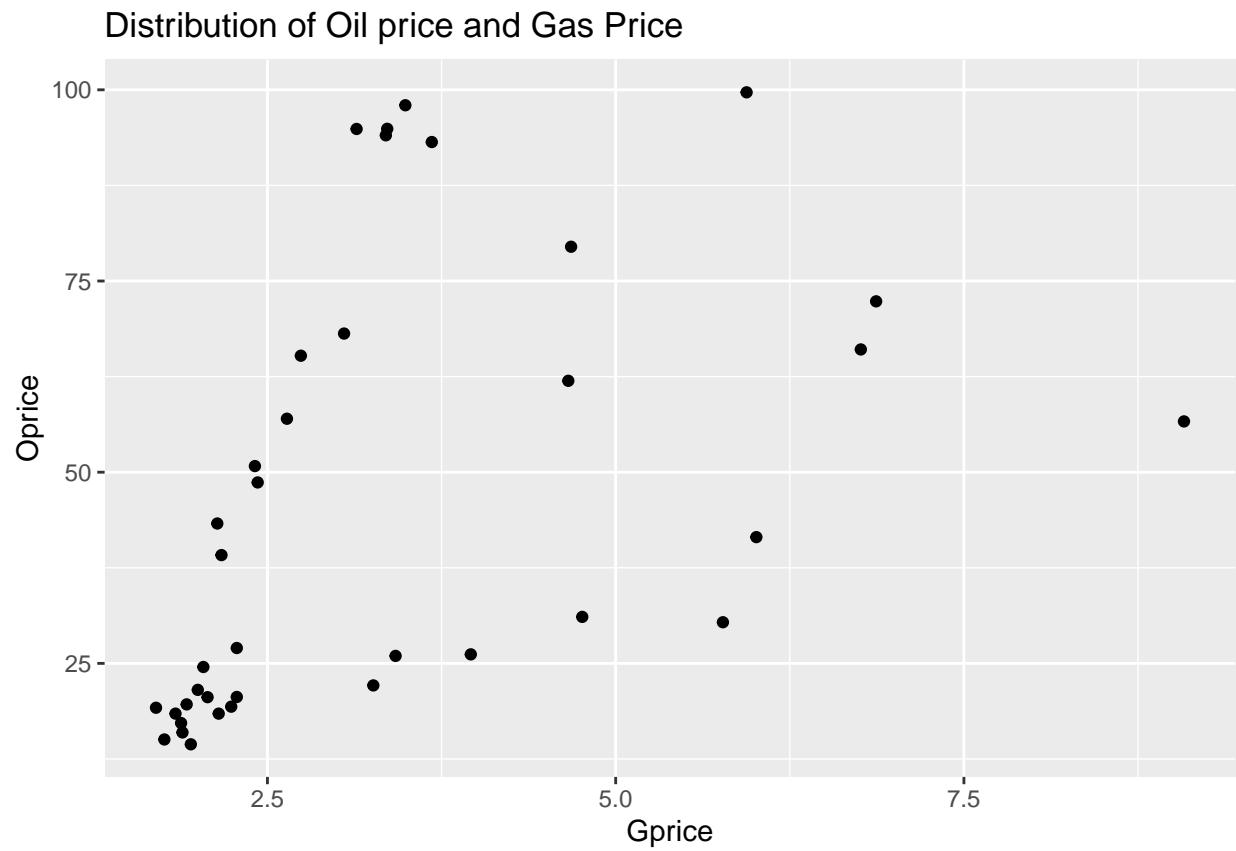
#Scatter Plot

```
library(ggplot2)
dd<- ggplot(data=df, aes(x=Oprod, y=Oprice ))
dd +geom_point()+ggtitle("Distribution of Oil price and Oil Production")
```

Distribution of Oil price and Oil Production


```r
dd2<- ggplot(data=df, aes(x=Gprice, y=Oprice))
dd2 +geom_point() +ggtitle("Distribution of Oil price and Gas Price")
```

Distribution of Oil price and Gas Price


```r
dd3<- ggplot(data=df, aes(x=Gprod, y=Oprice))
dd3 +geom_point() + ggtitle("Distribution of Oil price and Gas Production")
```
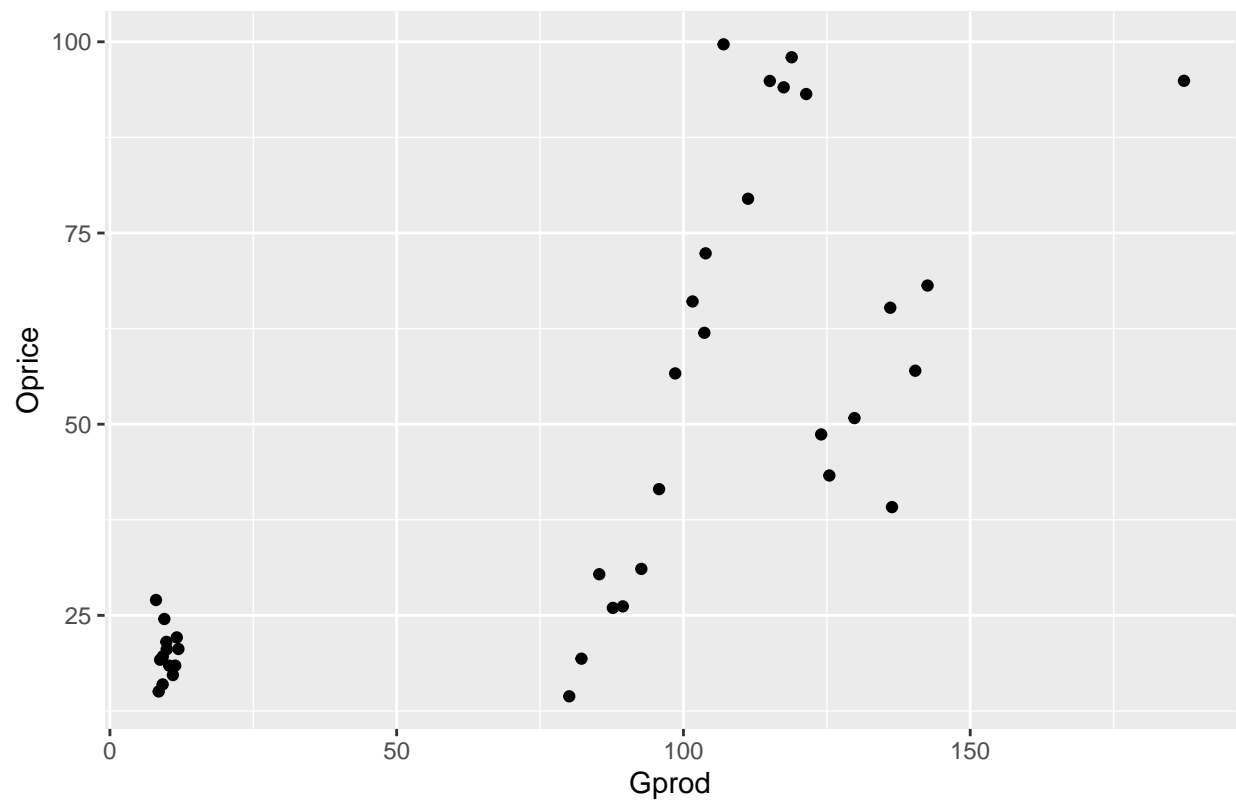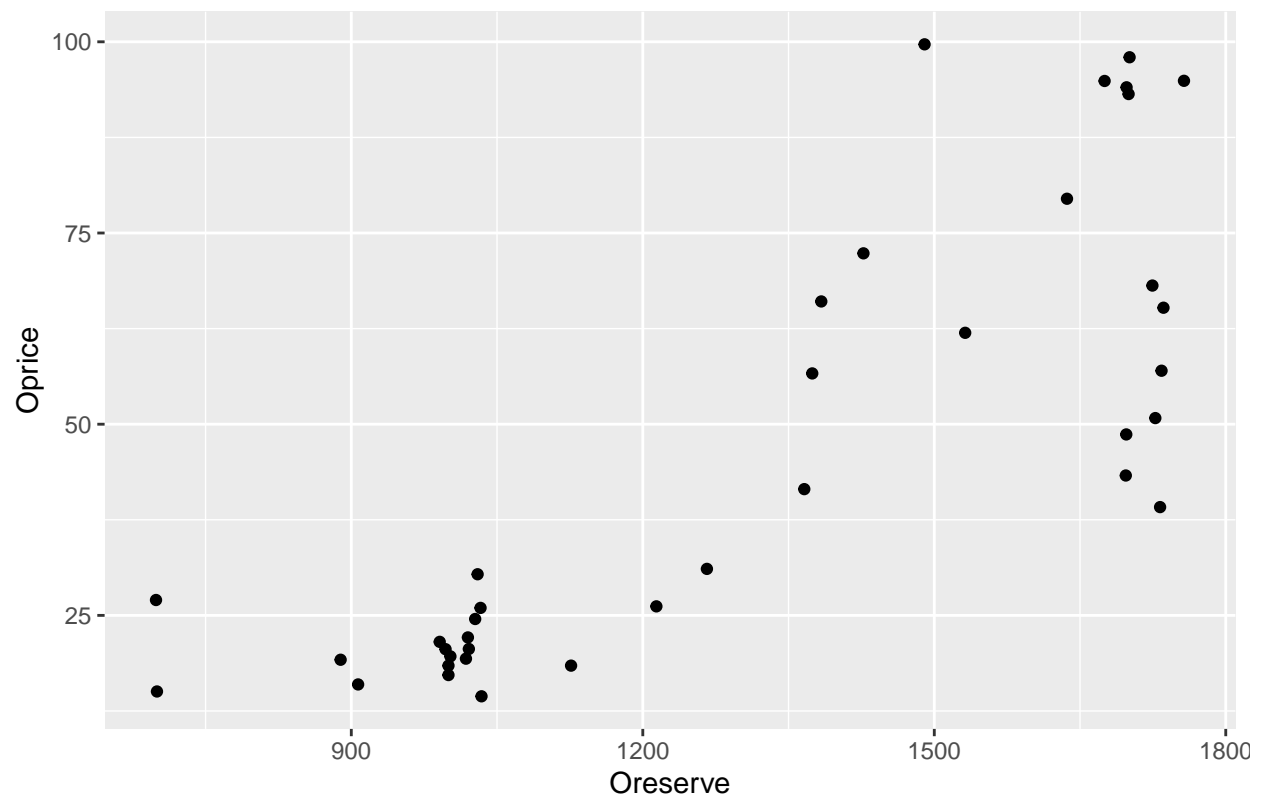
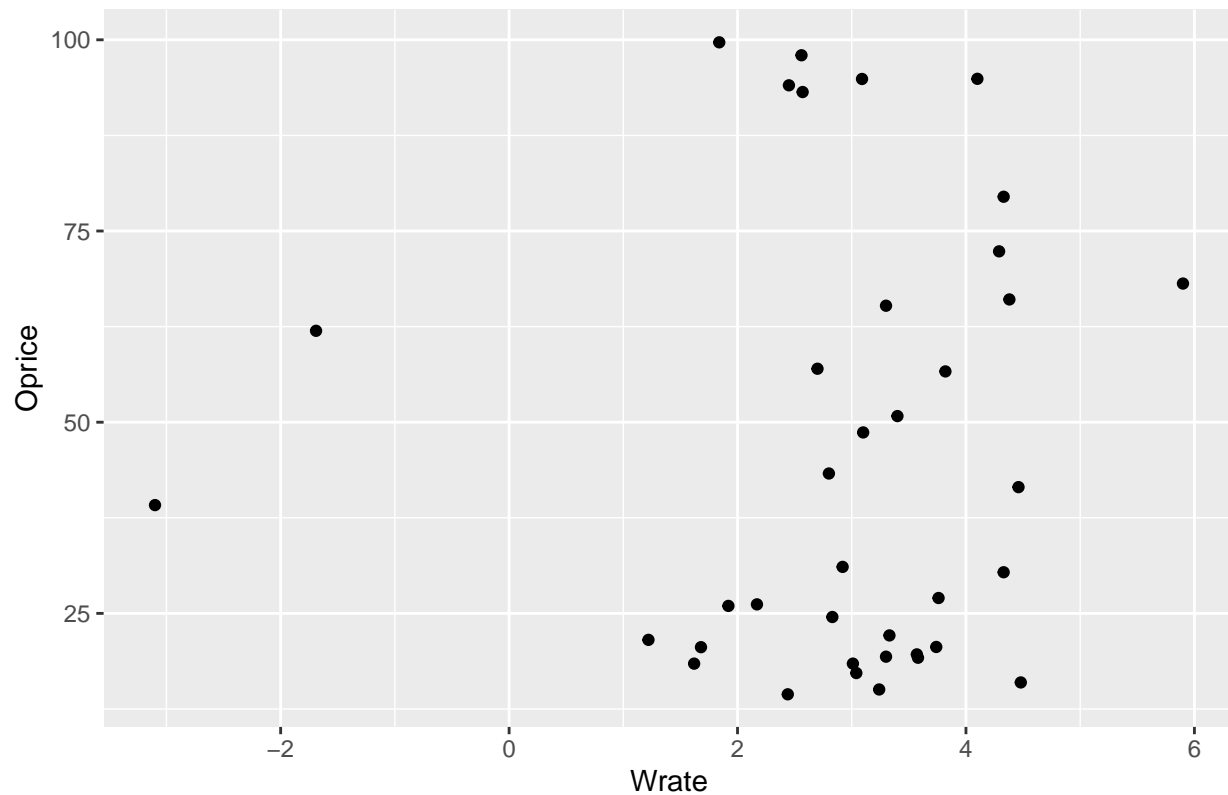## Distribution of Oil price and Gas Production


```
dd4<- ggplot(data=df, aes(x=Oreserve, y=Oprice))
dd4 +geom_point() +ggtitle("Distribution of Oil price and Oil Proved Reserve")
```

Distribution of Oil price and Oil Proved Reserve


```
dd5<- ggplot(data=df, aes(x=Wrate, y=Oprice))
dd5 +geom_point() + ggtitle("Distribution of Oil price and World Economy")
```

Distribution of Oil price and World Economy

#Multiple Regression

```
set.seed(100)
```

##First model

```
mo <- lm(Oprice~., data=df)
summary(mo)
```

```
##
## Call:
## lm(formula = Oprice ~ ., data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.257  -7.475  -0.946   7.279  33.483
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.66326   47.35803   1.936  0.06180 .
## Oprod       -9.77194    2.85836  -3.419  0.00173 **
## Gprice       4.62301    1.45476   3.178  0.00328 **
## Gprod        0.42184    0.17231   2.448  0.02002 *
## Oreserve     0.12742    0.02253   5.655 2.95e-06 ***
## Wrate        2.33496    1.46964   1.589  0.12194
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.67 on 32 degrees of freedom
## Multiple R-squared:  0.8035, Adjusted R-squared:  0.7727
## F-statistic: 26.16 on 5 and 32 DF,  p-value: 1.966e-10
```

```
mean(mo$residuals^2)
```

```
## [1] 157.3943
```

Multicolinearity

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```
vif(mo)
```

```
##     Oprod    Gprice     Gprod  Oreserve     Wrate
## 33.867636  1.291220 16.780103 11.869326  1.087277
```

2nd model

```
model2 <- lm(Oprice~ Gprice+Oprod+Wrate+Oreserve, data=df)
summary(model2)
```

```
##
## Call:
## lm(formula = Oprice ~ Gprice + Oprod + Wrate + Oreserve, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.973  -9.006  -2.494   5.556  33.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.63840   21.26242  -0.641  0.52567
## Gprice        4.97934    1.55311   3.206  0.00299 **
## Oprod        -4.12740    1.81277  -2.277  0.02941 *
## Wrate         2.08010    1.57295   1.322  0.19512
## Oreserve      0.11432    0.02348   4.868 2.72e-05 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.67 on 33 degrees of freedom
## Multiple R-squared:  0.7666, Adjusted R-squared:  0.7384
## F-statistic:  27.1 on 4 and 33 DF,  p-value: 5.097e-10
```

```
mean(model2$residuals^2)
```

```
## [1] 186.8732
```

3rd model #Backward AIC

```
model3=step(mo,direction = "backward")
```

```
## Start:  AIC=204.23
## Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate
##
##             Df Sum of Sq     RSS    AIC
## <none>                    5981.0 204.23
## - Wrate      1     471.8  6452.8 205.12
## - Gprod      1    1120.2  7101.2 208.76
## - Gprice     1    1887.5  7868.5 212.66
## - Oprod      1    2184.5  8165.5 214.06
## - Oreserve   1    5977.8 11958.8 228.56
```

```
mean(model3$residuals^2)
```

```
## [1] 157.3943
```

```
summary(model3)
```

```
##
## Call:
## lm(formula = Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.257  -7.475  -0.946   7.279  33.483
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.66326   47.35803   1.936  0.06180 .
## Oprod       -9.77194    2.85836  -3.419  0.00173 **
## Gprice       4.62301    1.45476   3.178  0.00328 **
## Gprod        0.42184    0.17231   2.448  0.02002 *
## Oreserve     0.12742    0.02253   5.655 2.95e-06 ***
## Wrate        2.33496    1.46964   1.589  0.12194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

10

```
## Residual standard error: 13.67 on 32 degrees of freedom
## Multiple R-squared:  0.8035, Adjusted R-squared:  0.7727
## F-statistic: 26.16 on 5 and 32 DF,  p-value: 1.966e-10
```

#Forward AIC

```
model4=step(mo, direction="forward")
```

```
## Start:  AIC=204.23
## Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate
```

```
mean(model4$residuals^2)
```

```
## [1] 157.3943
```

```
summary(model4)
```

```
##
## Call:
## lm(formula = Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.257  -7.475  -0.946   7.279  33.483
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.66326   47.35803   1.936  0.06180 .
## Oprod       -9.77194    2.85836  -3.419  0.00173 **
## Gprice       4.62301    1.45476   3.178  0.00328 **
## Gprod        0.42184    0.17231   2.448  0.02002 *
## Oreserve     0.12742    0.02253   5.655 2.95e-06 ***
## Wrate        2.33496    1.46964   1.589  0.12194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.67 on 32 degrees of freedom
## Multiple R-squared:  0.8035, Adjusted R-squared:  0.7727
## F-statistic: 26.16 on 5 and 32 DF,  p-value: 1.966e-10
```

#BOth FOrward and BAckward

```
model5=step(mo, direction="both")
```

```
## Start:  AIC=204.23
## Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate
##
##          Df Sum of Sq    RSS    AIC
## <none>                 5981.0 204.23
## - Wrate   1     471.8 6452.8 205.12
```

```
## - Gprod     1    1120.2  7101.2 208.76
## - Gprice    1    1887.5  7868.5 212.66
## - Oprod     1    2184.5  8165.5 214.06
## - Oreserve  1    5977.8 11958.8 228.56
```

```r
summary(model5)
```

```
##
## Call:
## lm(formula = Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate,
##     data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.257  -7.475  -0.946   7.279  33.483
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 91.66326    47.35803   1.936  0.06180 .
## Oprod       -9.77194     2.85836  -3.419  0.00173 **
## Gprice       4.62301     1.45476   3.178  0.00328 **
## Gprod        0.42184     0.17231   2.448  0.02002 *
## Oreserve     0.12742     0.02253   5.655 2.95e-06 ***
## Wrate        2.33496     1.46964   1.589  0.12194
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.67 on 32 degrees of freedom
## Multiple R-squared:  0.8035, Adjusted R-squared:  0.7727
## F-statistic: 26.16 on 5 and 32 DF,  p-value: 1.966e-10
```

```r
mean(model5$residuals^2)
```

```
## [1] 157.3943
```

```r
#Mean square error
```

```r
mean(model2$residuals^2)
```

```
## [1] 186.8732
```

```r
mean(model3$residuals^2)
```

```
## [1] 157.3943
```

```r
mean(model4$residuals^2)
```

```
## [1] 157.3943
```

```
mean(model5$residuals^2)
```

```
## [1] 157.3943
```

```
anova(mo, model2 )
```

```
## Analysis of Variance Table
##
## Model 1: Oprice ~ Oprod + Gprice + Gprod + Oreserve + Wrate
## Model 2: Oprice ~ Gprice + Oprod + Wrate + Oreserve
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     32 5981.0
## 2     33 7101.2 -1   -1120.2 5.9934 0.02002 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```