02441 Applied Statistics and Statistical Software

Exercise 3C - Filter

The data are from a statement by Texaco, Inc. to the Air and Water Pollution Subcommittee of the Senate Public Works Committee on June 26, 1973. Mr. John McKinley, President of Texaco, cited the Octel filter, developed by Associated Octel Company as effective in reducing pollution. However, questions had been raised about the effects of pollution filters on aspects of vehicle performance, including noise levels. He referred to data presented in the datafile associated with this story as evidence that the Octel filter was at least as good as a standard silencer in controlling vehicle noise levels. The dataset filter constitute a 3-way factorial experiment with 3 replications. The factors are type of filter (2 types), vehicle size (3 sizes), and side of car (two sides).

Variable name	Description
NOISE	Noise level reading (decibels)
SIZE	Vehicle size: 1 small 2 medium 3 large
TYPE	1 standard silencer 2 Octel filter
SIDE	1 right side 2 left side of car

1. Determine whether size, type and side influence on the noise level by doing a graphical comparison

Start by loading and converting data to right class

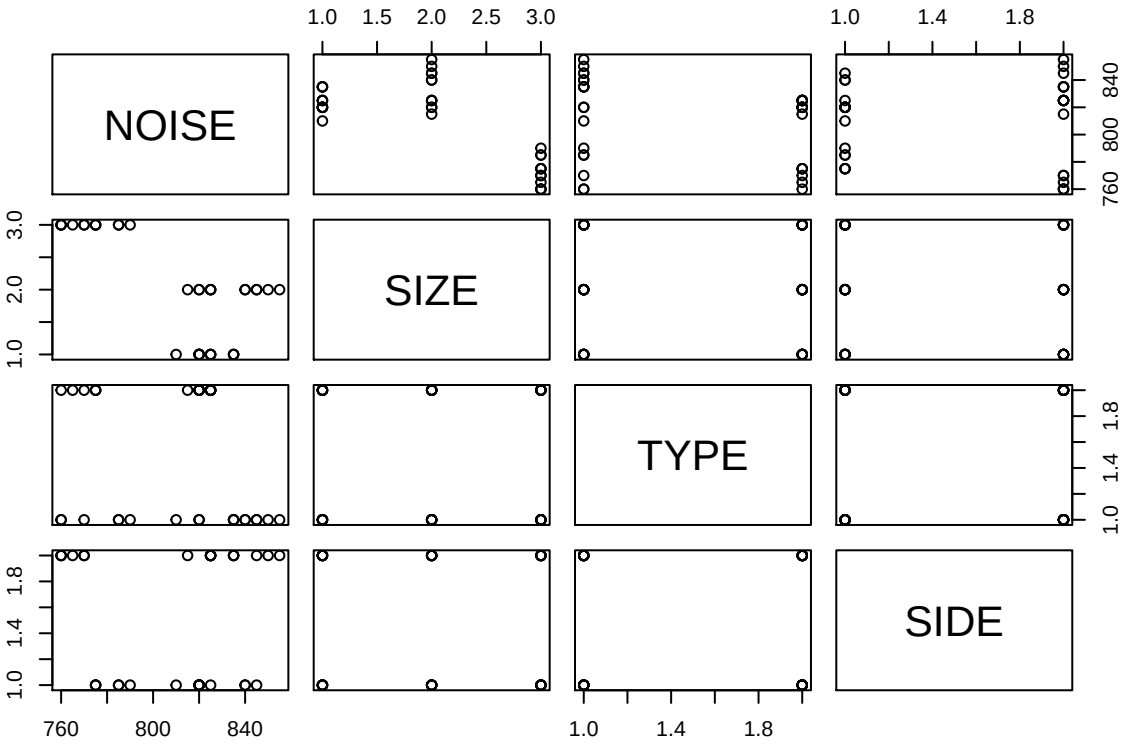
```
noi <- read.table("filter.txt", header = TRUE, colClasses=
                    c('numeric','factor','factor','factor'))
str(noi)
```

```
## 'data.frame':    36 obs. of  4 variables:
##  $ NOISE: num  810 820 820 840 840 845 785 790 785 835 ...
##  $ SIZE : Factor w/ 3 levels "1","2","3": 1 1 1 2 2 2 3 3 3 1 ...
##  $ TYPE : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ SIDE : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 2 ...
```

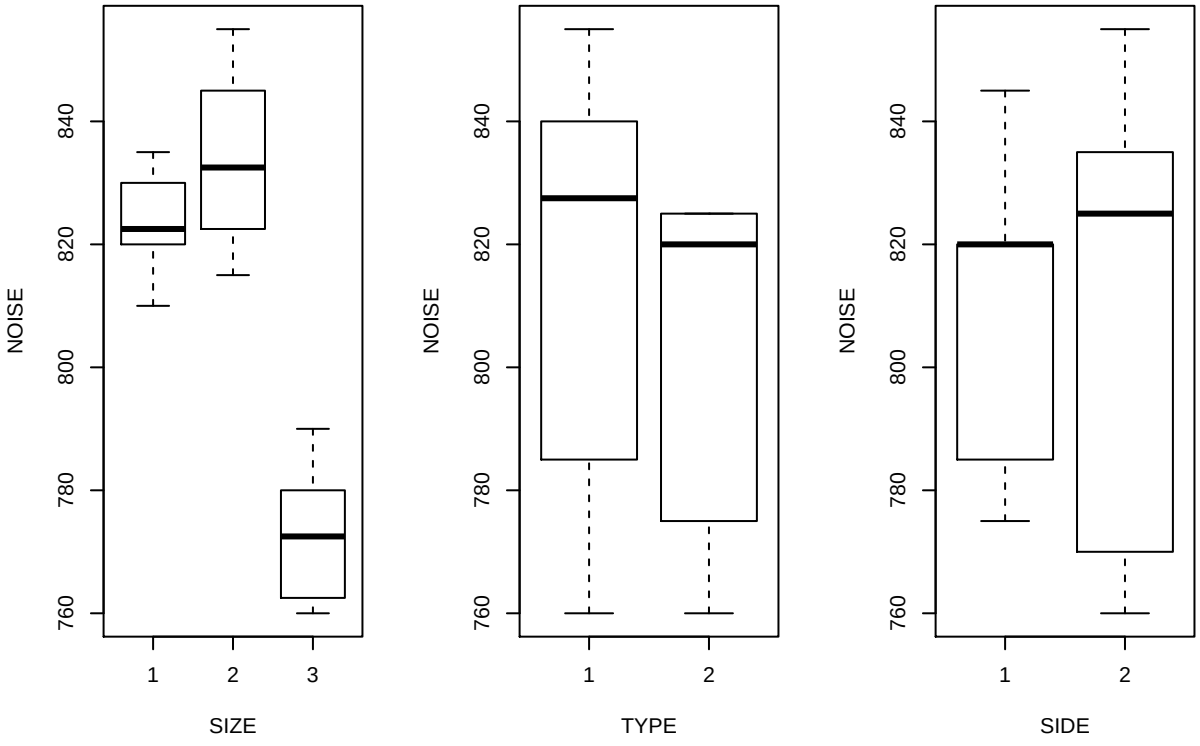
It is important to read categorical variables as factors. Check if this is done right using the `str` command.

Visual inspection of data

```
pairs(noi)
```




```
par(mfrow=c(1,3))
plot(NOISE~SIZE+TYPE+SIDE, noi)
```



2. Determine whether size, type and side influence on the noise level by the appropriate statistical analysis

```
lm1a <- lm(NOISE~.,noi)
anova(lm1a)
```

```
## Analysis of Variance Table
##
## Response: NOISE
##           Df  Sum Sq Mean Sq  F value    Pr(>F)
## SIZE       2 26051.4 13025.7 145.9872 < 2.2e-16 ***
## TYPE       1  1056.2  1056.2  11.8381  0.001679 **
## SIDE       1     0.7     0.7   0.0078  0.930268
## Residuals 31  2766.0    89.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm1b <- update(lm1a, .~.-SIDE)
anova(lm1b)
```

```
## Analysis of Variance Table
##
## Response: NOISE
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## SIZE       2 26051.4 13025.7 150.659 < 2.2e-16 ***
## TYPE       1  1056.2  1056.2  12.217  0.001411 **
## Residuals 32  2766.7    86.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

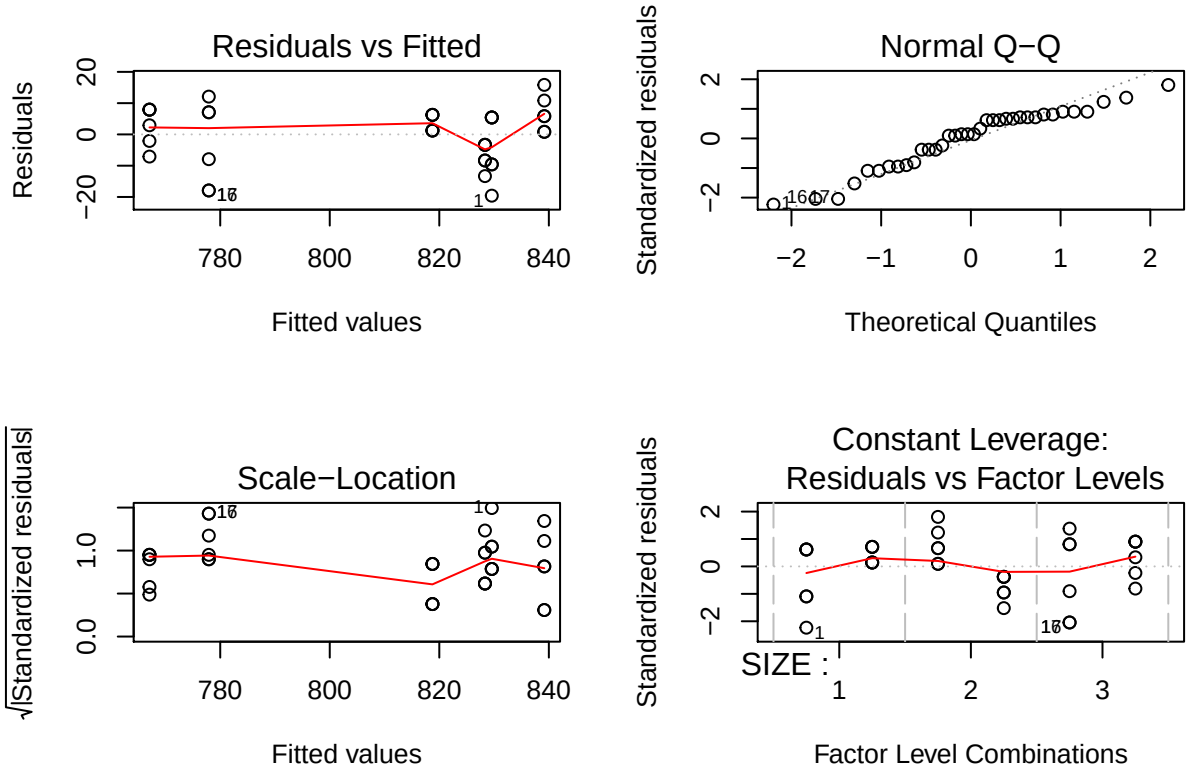
```
summary(lm1b)
```

```
##
## Call:
## lm(formula = NOISE ~ SIZE + TYPE, data = noi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.583  -7.292   1.250   6.250  15.833
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  829.583      3.099 267.657  < 2e-16 ***
## SIZE2          9.583      3.796   2.525  0.01674 *
## SIZE3        -51.667      3.796 -13.611  7.4e-15 ***
## TYPE2        -10.833      3.099  -3.495  0.00141 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.298 on 32 degrees of freedom
## Multiple R-squared:  0.9074, Adjusted R-squared:  0.8987
## F-statistic: 104.5 on 3 and 32 DF,  p-value: < 2.2e-16
```

SIZE and TYPE influence noise level. SIDE does not. Significance level for the statistical test was $\alpha = 5\%$. Roughly 91% of variance in NOISE are explained by `lm1b`.

Check model assumptions

```
par(mfrow=c(2,2))
plot(lm1b)
```



Everything looks fine.

3. Are there any interaction effects between size and type?

```
lm2 <- lm(NOISE~.^3,noi)
anova(lm2)
```

```
## Analysis of Variance Table
##
## Response: NOISE
##                Df  Sum Sq Mean Sq  F value    Pr(>F)
## SIZE            2 26051.4 13025.7 893.1905 < 2.2e-16 ***
## TYPE            1  1056.2  1056.2  72.4286 1.038e-08 ***
## SIDE            1     0.7     0.7   0.0476 0.8291042
## SIZE:TYPE       2   804.2   402.1  27.5714 6.048e-07 ***
## SIZE:SIDE       2  1293.1   646.5  44.3333 8.730e-09 ***
## TYPE:SIDE       1    17.4    17.4   1.1905 0.2860667
## SIZE:TYPE:SIDE  2   301.4   150.7  10.3333 0.0005791 ***
## Residuals      24   350.0    14.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# backwards model selection is easy using drop1
drop1(lm2, test = 'F')
```

```
## Single term deletions
##
## Model:
## NOISE ~ (SIZE + TYPE + SIDE)^3
##                Df Sum of Sq    RSS    AIC F value    Pr(>F)
## <none>                      350.00 105.88
## SIZE:TYPE:SIDE  2    301.39 651.39 124.24  10.333 0.0005791 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The third order interaction is significant, hence the model cannot be reduced any further. The function `drop1` can help you perform model selection. It only displays effects eligible for reduction. If all effects shown in `drop1` output are significant you can't reduce the model any further.

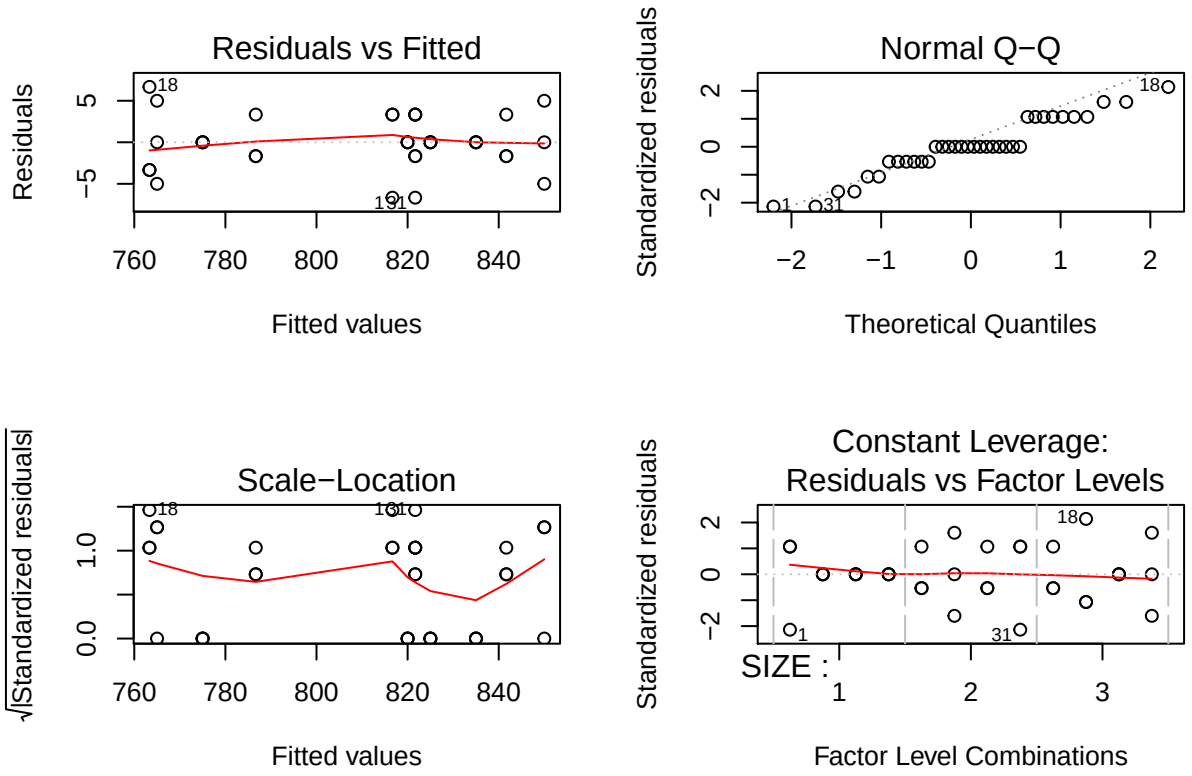
```r
summary(lm2)
```

```
##
## Call:
## lm(formula = NOISE ~ .^3, data = noi)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.667 -1.667  0.000  3.333  6.667
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         816.667      2.205 370.405  < 2e-16 ***
## SIZE2                25.000      3.118   8.018 3.03e-08 ***
## SIZE3               -30.000      3.118  -9.621 1.04e-09 ***
## TYPE2                 3.333      3.118   1.069 0.295684
## SIDE2                18.333      3.118   5.880 4.58e-06 ***
## SIZE2:TYPE2         -23.333      4.410  -5.292 1.99e-05 ***
## SIZE3:TYPE2         -15.000      4.410  -3.402 0.002348 **
## SIZE2:SIDE2         -10.000      4.410  -2.268 0.032627 *
## SIZE3:SIDE2         -41.667      4.410  -9.449 1.47e-09 ***
## TYPE2:SIDE2         -13.333      4.410  -3.024 0.005865 **
## SIZE2:TYPE2:SIDE2     5.000      6.236   0.802 0.430546
## SIZE3:TYPE2:SIDE2    26.667      6.236   4.276 0.000262 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.819 on 24 degrees of freedom
## Multiple R-squared:  0.9883, Adjusted R-squared:  0.9829
## F-statistic:   184 on 11 and 24 DF,  p-value: < 2.2e-16
```

The explained variance (R^2) in NOISE for `lm2` is ca. 8% higher than for `lm1b`.

Check model assumptions

```
par(mfrow=c(2,2))
plot(lm2)
```



The qq-plot for `lm2` looks very structured. This is due to the fact that more model parameters improve the fit, leading to reduced complexity of the residuals.