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

```
# Loading data and converting to factors
filter <- read.table("filter.txt", header=TRUE)
filter$SIZE <- factor(filter$SIZE)
levels(filter$SIZE) <- c("small", "medium", "large")
filter$TYPE <- factor(filter$TYPE)
levels(filter$TYPE) <- c("standard", "octel")
filter$SIDE <- factor(filter$SIDE)
levels(filter$SIDE) <- c("right", "left")

# Summary of the data
summary(filter)
```

```
##      NOISE           SIZE          TYPE       SIDE
##  Min.   :760.0   small :12   standard:18   right:18
##  1st Qu.:782.5   medium:12   octel   :18   left :18
##  Median :820.0   large :12
##  Mean   :810.1
##  3rd Qu.:827.5
##  Max.   :855.0
```

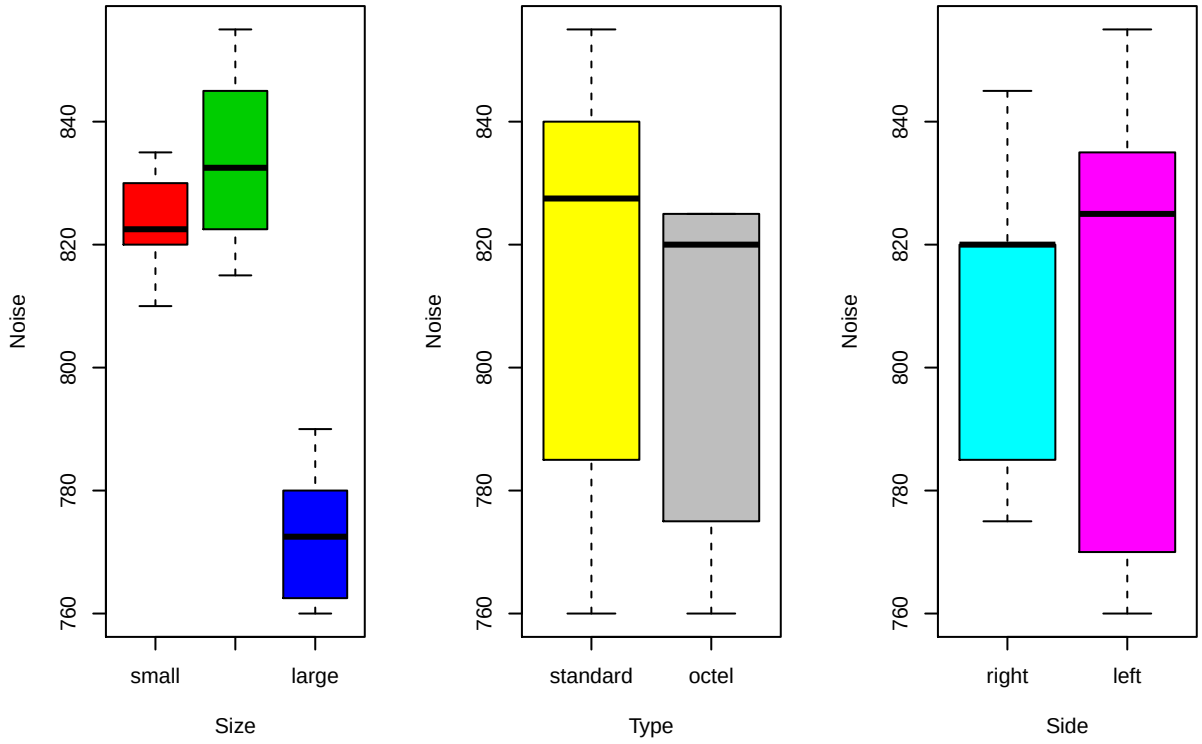
```
# Structure of the data
str(filter)
```

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

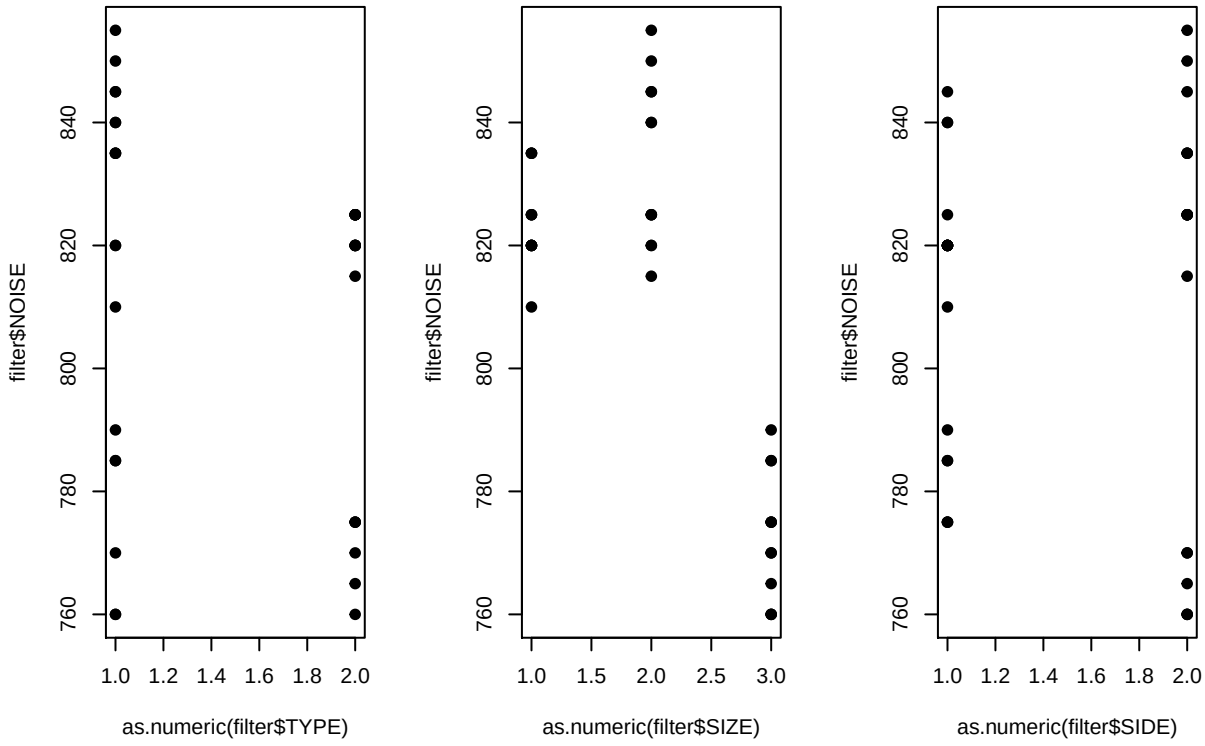
```
##  $ TYPE : Factor w/ 2 levels "standard","octel": 1 1 1 1 1 1 1 1 1 1 ...
##  $ SIDE : Factor w/ 2 levels "right","left": 1 1 1 1 1 1 1 1 1 2 ...
```

Visual inspection of data

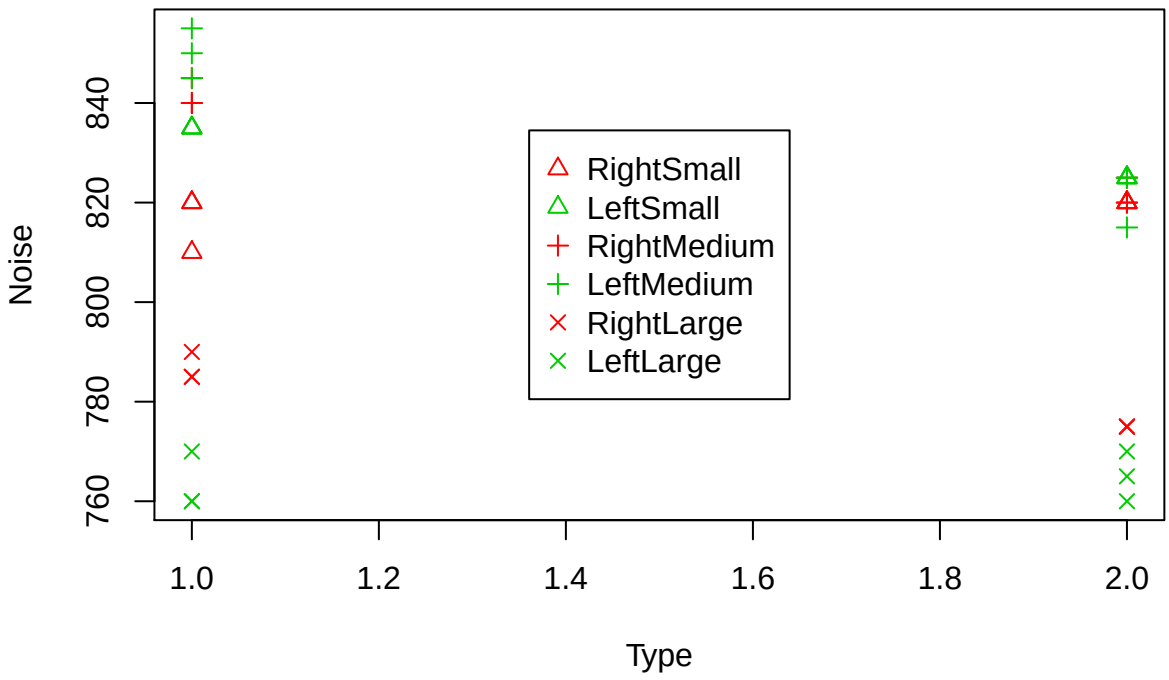
```
# Boxplot
par(mfrow = c(1,3))
boxplot(filter$NOISE~filter$SIZE, filter, xlab = "Size", ylab = "Noise", col = c(2,3,4))
boxplot(filter$NOISE~filter$TYPE, filter, xlab = "Type", ylab = "Noise", col = c(7,8))
boxplot(filter$NOISE~filter$SIDE, filter, xlab = "Side", ylab = "Noise", col = c(5,6))
```




```
# Scatter plot
par(mfrow = c(1,3))
plot(as.numeric(filter$TYPE), filter$NOISE, pch=19)
plot(as.numeric(filter$SIZE), filter$NOISE, pch=19)
plot(as.numeric(filter$SIDE), filter$NOISE, pch=19)
```




```
# Scatter plot combined
par(mfrow=c(1,1))
plot(as.numeric(filter$TYPE), filter$NOISE, col = as.numeric(filter$SIDE)+1,
     pch = as.numeric(filter$SIZE)+1, xlab = "Type", ylab = "Noise")
legend("center", legend = c("RightSmall", "LeftSmall", "RightMedium", "LeftMedium", "RightLarge", "Left
       col = c(2,3), pch = c(2,2,3,3,4,4))
```



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

```
# Linear Models
lmsize <- lm(filter$NOISE~filter$SIZE, filter)
lmtype <- lm(filter$NOISE~filter$TYPE, filter)
lmside <- lm(filter$NOISE~filter$SIDE, filter)
lmall <- lm(filter$NOISE~filter$SIZE+filter$TYPE+filter$SIDE, filter)

# One-way ANOVA of Size
anova(lmsize)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## filter$SIZE  2 26051.4 13025.7  112.44 1.85e-15 ***
## Residuals   33  3822.9   115.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Linear Model of Size
summary(lmsize)
```

```
##
## Call:
```

```
## lm(formula = filter$NOISE ~ filter$SIZE, data = filter)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.7500  -8.7500  -0.8333  10.8333  21.2500
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        824.167      3.107 265.256  < 2e-16 ***
## filter$SIZEmedium    9.583      4.394   2.181   0.0364 *
## filter$SIZElarge   -51.667      4.394 -11.758 2.42e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.76 on 33 degrees of freedom
## Multiple R-squared:  0.872,  Adjusted R-squared:  0.8643
## F-statistic: 112.4 on 2 and 33 DF,  p-value: 1.85e-15
```

```
# One-way ANOVA of Type
anova(lmtype)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##             Df  Sum Sq Mean Sq F value Pr(>F)
## filter$TYPE  1  1056.2 1056.25  1.2462 0.2721
## Residuals   34 28818.1  847.59
```

```
# Linear Model of Type
summary(lmtype)
```

```
##
## Call:
## lm(formula = filter$NOISE ~ filter$TYPE, data = filter)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -55.56 -29.72  15.28  20.28  39.44
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       815.556      6.862 118.849   <2e-16 ***
## filter$TYPEoctel  -10.833      9.704  -1.116    0.272
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.11 on 34 degrees of freedom
## Multiple R-squared:  0.03536,    Adjusted R-squared:  0.006985
## F-statistic: 1.246 on 1 and 34 DF,  p-value: 0.2721
```

```
# One-way ANOVA of Side
anova(lmside)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##             Df  Sum Sq Mean Sq F value Pr(>F)
## filter$SIDE  1     0.7    0.69   8e-04 0.9777
## Residuals   34 29873.6  878.64
```

```
# Linear Model of Side
summary(lmside)
```

```
##
## Call:
## lm(formula = filter$NOISE ~ filter$SIDE, data = filter)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.000 -27.778   9.722  17.500  45.000
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     810.2778     6.9866 115.975   <2e-16 ***
## filter$SIDEleft  -0.2778     9.8806  -0.028    0.978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.64 on 34 degrees of freedom
## Multiple R-squared:  2.325e-05,  Adjusted R-squared:  -0.02939
## F-statistic: 0.0007904 on 1 and 34 DF,  p-value: 0.9777
```

```
# Three-way ANOVA of Size, Type, and Side
anova(lmall)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## filter$SIZE  2 26051.4 13025.7 145.9872 < 2.2e-16 ***
## filter$TYPE  1  1056.2  1056.2  11.8381  0.001679 **
## filter$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
```

```
# Linear model of Size, Type, and Side
summary(lmall)
```

```
##
## Call:
## lm(formula = filter$NOISE ~ filter$SIZE + filter$TYPE + filter$SIDE,
##     data = filter)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -19.722  -7.153   1.111   6.389  15.972
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       829.7222     3.5203 235.698  < 2e-16 ***
## filter$SIZEmedium   9.5833     3.8563   2.485  0.01855 *
## filter$SIZElarge  -51.6667     3.8563 -13.398 1.95e-14 ***
## filter$TYPEoctel  -10.8333     3.1486  -3.441  0.00168 **
## filter$SIDEleft    -0.2778     3.1486  -0.088  0.93027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.446 on 31 degrees of freedom
## Multiple R-squared:  0.9074, Adjusted R-squared:  0.8955
## F-statistic: 75.96 on 4 and 31 DF,  p-value: 1.444e-15
```

Size is statistically significant, i.e. means of the groups are the same. Type is not statistically significant (octel), i.e. means of the groups are not the same. Side is not statistically significant (left) , i.e. means of the groups are not the same. For the three-way anova, side is not statistically significant, i.e. variable side doesn't have the same mean in the groups AND it doesn't affect the noise levels.

We could investigate a new model without the side variable.

```
lmall_side <- lm(filter$NOISE~filter$SIZE+filter$TYPE, filter)
# Three-way ANOVA without Side
anova(lmall_side)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## filter$SIZE  2 26051.4 13025.7 150.659 < 2.2e-16 ***
## filter$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
```

```
# Linear Model of Size and Type
summary(lmall_side)
```

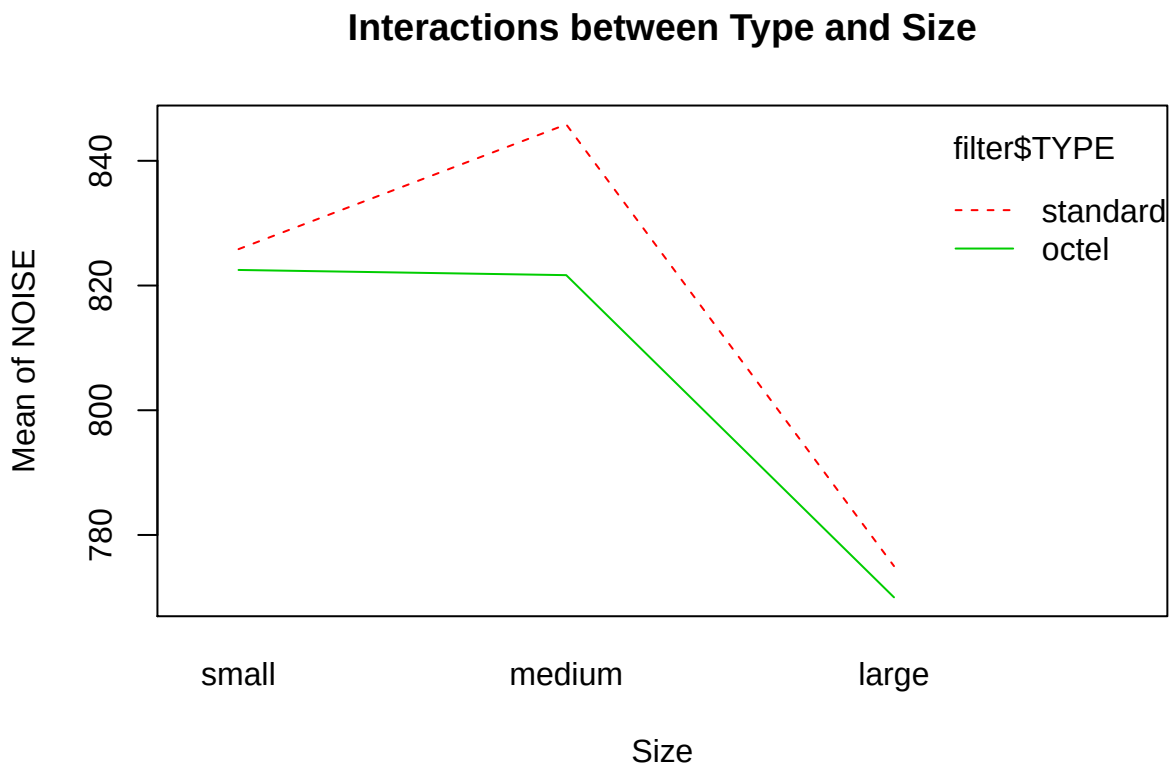
```
##
## Call:
## lm(formula = filter$NOISE ~ filter$SIZE + filter$TYPE, data = filter)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.583  -7.292   1.250   6.250  15.833
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        829.583      3.099 267.657  < 2e-16 ***
## filter$SIZEmedium    9.583      3.796   2.525  0.01674 *
```

```
## filter$SIZElarge   -51.667      3.796 -13.611  7.4e-15 ***
## filter$TYPEoctel   -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
```

Model looks better and all groups have the same mean.

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

```
# Visualize interaction
par(mfrow=c(1,1))
interaction.plot(filter$SIZE, filter$TYPE, filter$NOISE, ylab="Mean of NOISE", xlab="Size", main="Intera
```




```
# Linear Model
lmall_int <- lm(filter$NOISE~filter$SIZE*filter$TYPE, filter)

# Two-way ANOVA
anova(lmall_int)
```

```
## Analysis of Variance Table
##
## Response: filter$NOISE
##                          Df  Sum Sq Mean Sq  F value    Pr(>F)
## filter$SIZE               2 26051.4 13025.7 199.1189 < 2.2e-16 ***
## filter$TYPE               1  1056.2  1056.2  16.1465 0.0003631 ***
## filter$SIZE:filter$TYPE   2   804.2   402.1   6.1465 0.0057915 **
## Residuals                30  1962.5    65.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Linear Model of Size and Type
summary(lmall_int)
```

```
##
## Call:
## lm(formula = filter$NOISE ~ filter$SIZE * filter$TYPE, data = filter)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.8333  -5.2083  -0.4167   5.0000  15.0000
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         825.833      3.302 250.106  < 2e-16 ***
## filter$SIZEmedium                    20.000      4.670   4.283 0.000175 ***
## filter$SIZElarge                    -50.833      4.670 -10.886 6.11e-12 ***
## filter$TYPEoctel                     -3.333      4.670  -0.714 0.480849
## filter$SIZEmedium:filter$TYPEoctel  -20.833      6.604  -3.155 0.003638 **
## filter$SIZElarge:filter$TYPEoctel    -1.667      6.604  -0.252 0.802471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.088 on 30 degrees of freedom
## Multiple R-squared:  0.9343, Adjusted R-squared:  0.9234
## F-statistic: 85.34 on 5 and 30 DF,  p-value: < 2.2e-16
```

Size small for type octel and interaction between size large and type octel are not statistically significant. That means, for size, that it doesn't matter if the type of the filter is standard or octel but the size is the important feature. While the interaction of size large and type octel doesn't affect the model, we thus reject H_0.