STAT2170 Assignment

2022-10-20
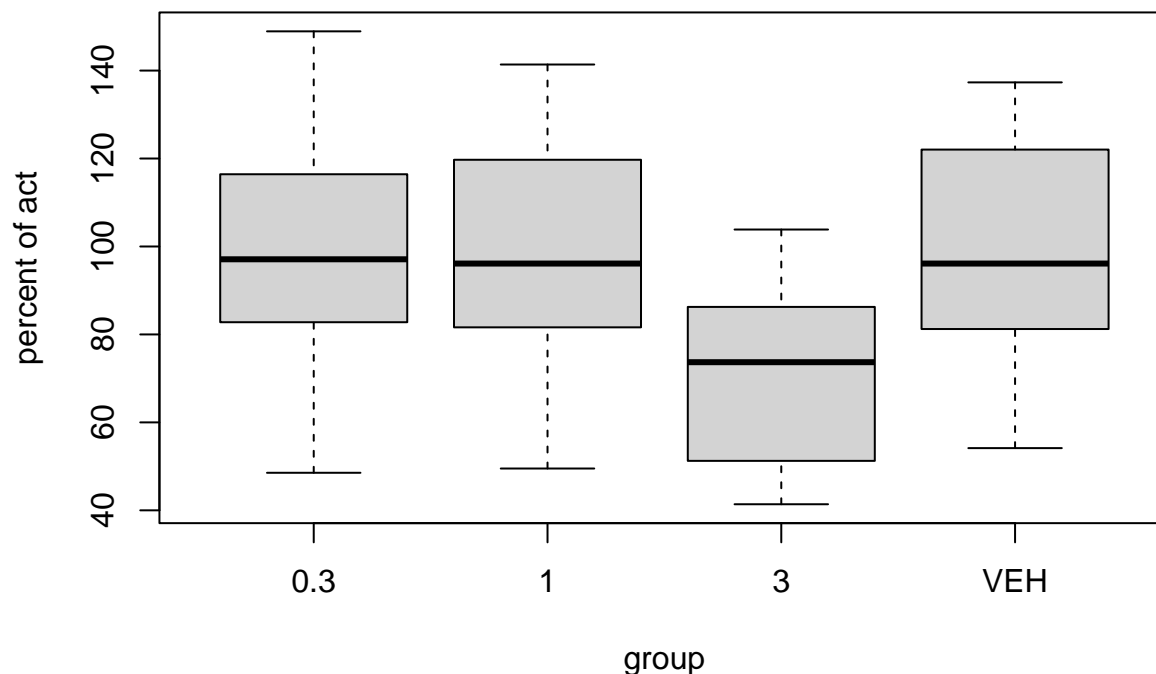
Question 1

```r
micepot = read.table("mice_pot.txt", header=TRUE)
head(micepot)
```

```
##   group percent_of_act
## 1   0.3        98.82671
## 2   0.3       116.42599
## 3   0.3       132.09128
## 4   0.3        84.12842
## 5   0.3       148.91697
## 6   0.3        48.54306
```

    a. Boxplots

```r
boxplot(percent_of_act ~ group, data = micepot, main="Marijuana on Mice",
        xlab="group", ylab="percent of act")
```

Marijuana on Mice

From the boxplots, we can see that:

1. The variability between the group looks pretty similar 2. Group 3 position is much lower than the other groups

b. Conducting the ANOVA

```
micepot.aov = aov(percent_of_act~group, data = micepot)
summary(micepot.aov)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## group         3   6329  2109.7   3.126 0.0357 *
## Residuals    42  28344   674.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let $Y =$ be the observed *percentage movement relative to the untreated mice* for a randomly selected sample. We fit the model, $Y_{ij} = \mu_i + \epsilon_{ij}$ where μ_i is the mean movement percentage at level i and the random variation ϵ_{ij} $N(0, \sigma2)$. Conducting an F-test for the group effects

- **Hypotheses**: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$; H_1 : not all means are equal.
- **Test statistic**: $F_obs = 3.126$
- **Null distribution**: If H_0 is true, F_obs behaves like a $F_{3,42}$ distribution
- **P-Value**: $P(F_{3,42} \geqslant 3.126) = 0.0357 < 0.05$
- **Conclusion**: Since the P-Value is less than the significance level of 0.05 (5%) we have evidence to reject H_0 in favour of H_1. That is, we have evidence that the percentage movement relative to the untreated mice is not all the same for all groups.

c. General Contrast

Hypotheses: $H_0 : \mu_4 = \frac{\mu_1 + \mu_2 + \mu_3}{3}$; $H_1 : \mu_4 \neq \frac{\mu_1 + \mu_2 + \mu_3}{3}$

```
aggregate(percent_of_act ~ group, mean, data = micepot)
```

```
##    group percent_of_act
## 1    0.3       97.32250
## 2      1       99.05235
## 3      3       70.66787
## 4    VEH      100.00000
```

Raw test statistic: $c = \sum_{i=1}^{g} k_i \bar{y}_i$

$c = 1 \times 70.66787 + \left(-\frac{1}{3}\right) \times 100.00000 + \left(-\frac{1}{3}\right) \times 97.32250 + \left(-\frac{1}{3}\right) \times 99.05235$

$= 70.66787 - 33.33333 - 32.44083 - 33.0174$

$= -28.12374$

Standard error: $s.e.(c) = s_p \sqrt{\sum_{i=1}^{g} \frac{k_i^2}{n_i}}$

$s.e.(c) = \sqrt{674.8} \times \sqrt{\frac{1^2}{10} + \frac{\left(-\frac{1}{3}\right)^2}{15} + \frac{\left(-\frac{1}{3}\right)^2}{9} + \frac{\left(-\frac{1}{3}\right)^2}{12}}$

$= 9.33046$

t-stat: $t_{obs} = \frac{c}{s.e.(c)}$ under H_0 follows t_{N-g}

$= \frac{-28.12374}{9.33046}$

$= -3.01419$

P-value:

```
pval <- 2*pt(3.01419, 42, lower.tail = FALSE)
pval
```

```
## [1] 0.004356163
```

$P - value = 0.004356163 < 0.05$
associated df $= 3,42$ Since the P-Value is less than the significance level of 0.05 (5%) we have evidence to reject H_0 in favour of H_1. That is, we have evidence that the the group with 3mg/kg dose is significantly different from the other groups.

Question 2

```
mykml = read.table("kml.dat", header=TRUE)
head(mykml)
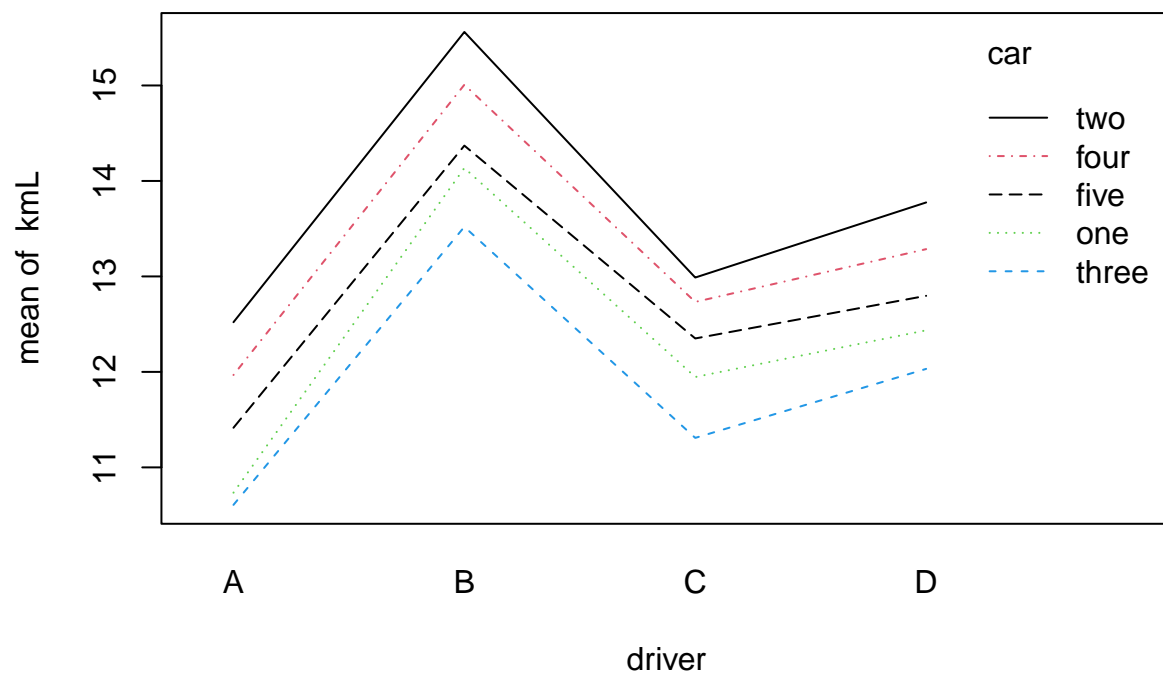```

```
kmL driver   car
1 10.75614   A   one
2 10.71363   A   one
3 12.28666   A   two
4 12.75432   A   two
5 10.54357   A three
6 10.67111   A three
```
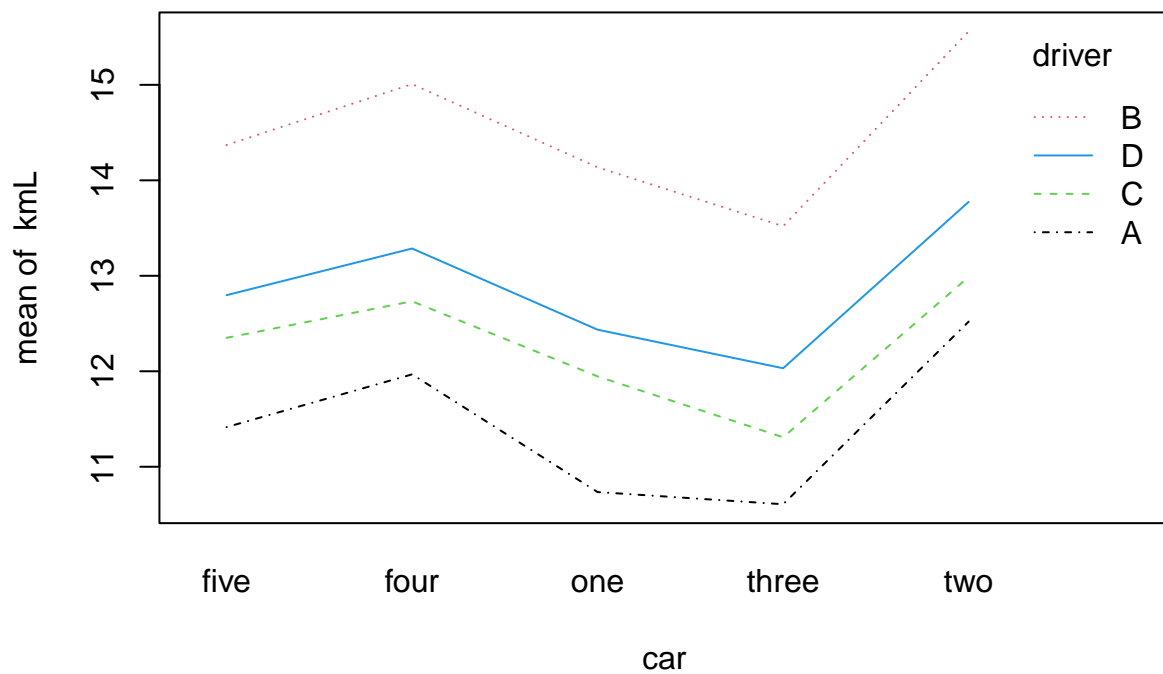
3

```
table(mykml[, 2:3])
```

```
##        car
## driver five four one three two
##      A    2    2   2     2   2
##      B    2    2   2     2   2
##      C    2    2   2     2   2
##      D    2    2   2     2   2
```

a. Here we can see that the group sizes are same for different levels of each factor combination. So the group design is balanced.

```
with(mykml, interaction.plot(driver, car, kmL, col = 1:4))
```
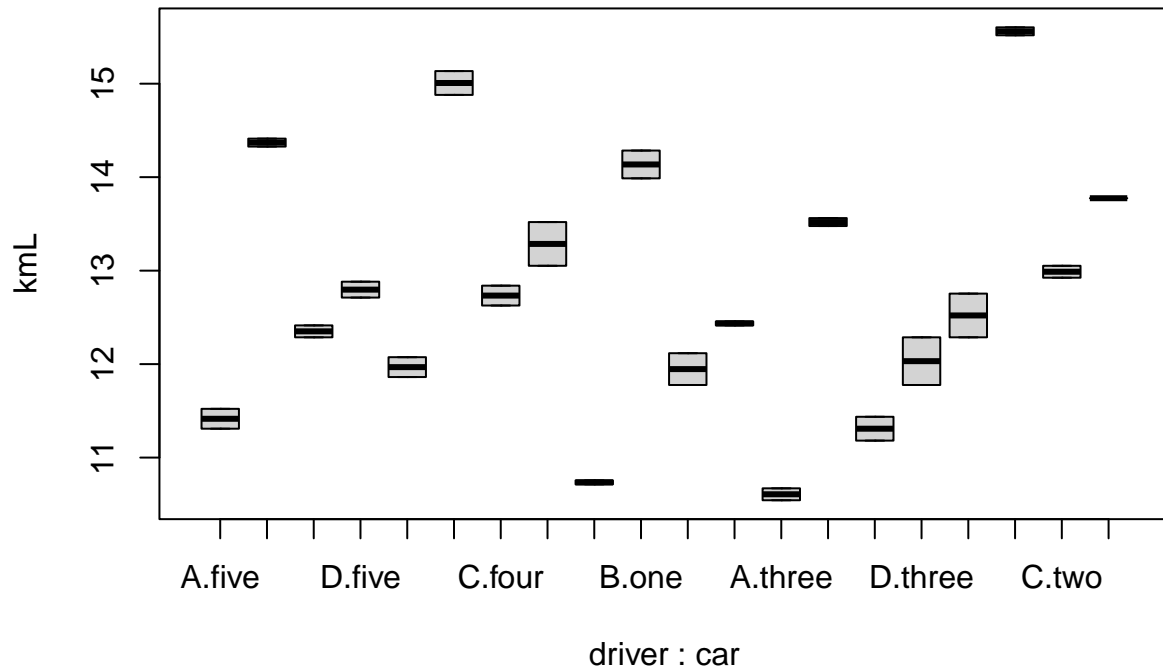


```
with(mykml, interaction.plot(car, driver, kmL, col = 1:5))
```

b. From the plot there could be a possible interaction between the factors and the response since the lines have slightly different slopes between levels of each factor.

```
boxplot(kmL~ driver + car, data = mykml)
```

From the plot, we could see that there are different variability in each group

c. **Hypotheses**

1. Interaction: $H_0 : \gamma_{ij} = 0$ for all i, j; H_A : not all $\gamma_{ij} = 0$

2. Main effect A: $H_0 : \alpha_i = 0$ for all i; H_A : not all $\alpha_i = 0$

3. Main effect B: $H_0 : \beta_j = 0$ for all j; H_A : not all $\beta_j = 0$

```
mykml.aov = aov(kmL ~ driver*car,
data = mykml)
summary(mykml.aov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## driver        3  50.66  16.887  531.60  < 2e-16 ***
## car           4  17.12   4.280  134.73 3.66e-14 ***
## driver:car   12   0.44   0.037    1.16    0.371
## Residuals    20   0.64   0.032
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table we have the F-statistics for this context to be 1.16 with $P - Value = P(F_{12,20} \geqslant 1.16) = 0.371$ which means we do not have enough evidence to reject H_0, the interaction between driver and car is not related to the efficiency of the car.

From the test above, we know the interaction is insignificant and can be dropped.

```
mykml.2aov = aov(kmL ~ driver + car, data = mykml)
summary(mykml.2aov)
```
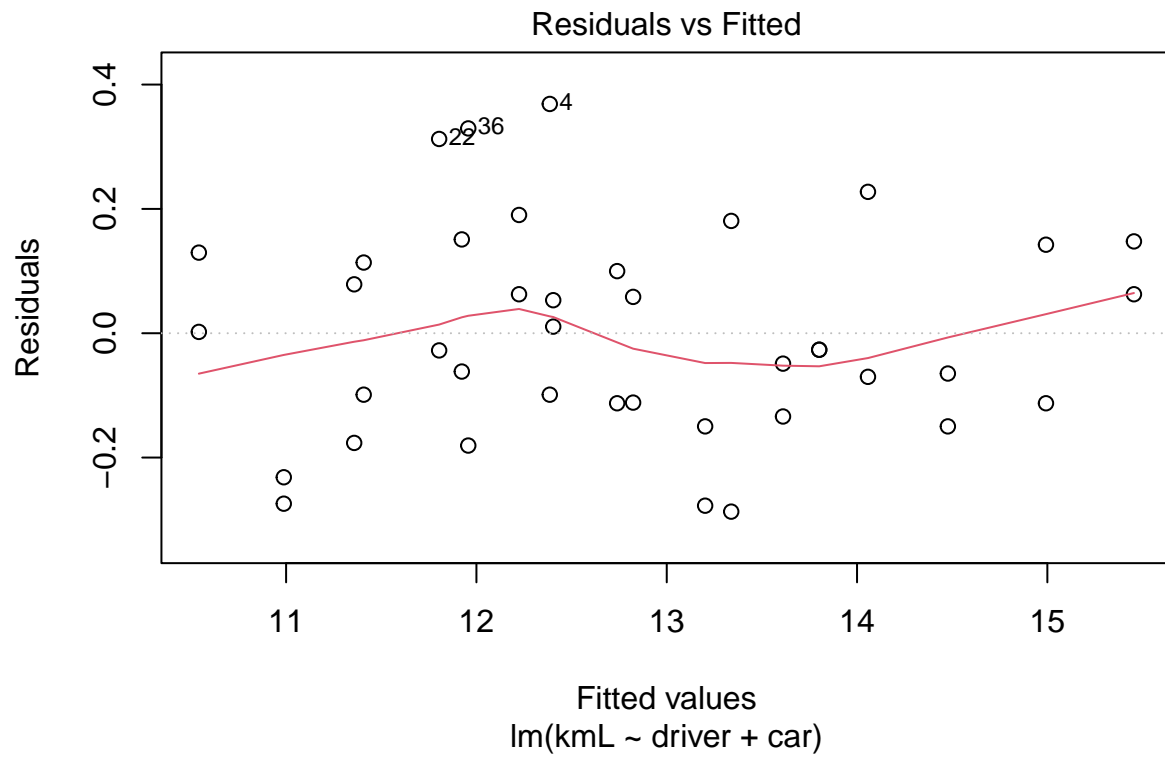
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## driver        3  50.66  16.887   501.5 <2e-16 ***
## car           4  17.12   4.280   127.1 <2e-16 ***
## Residuals    32   1.08   0.034
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
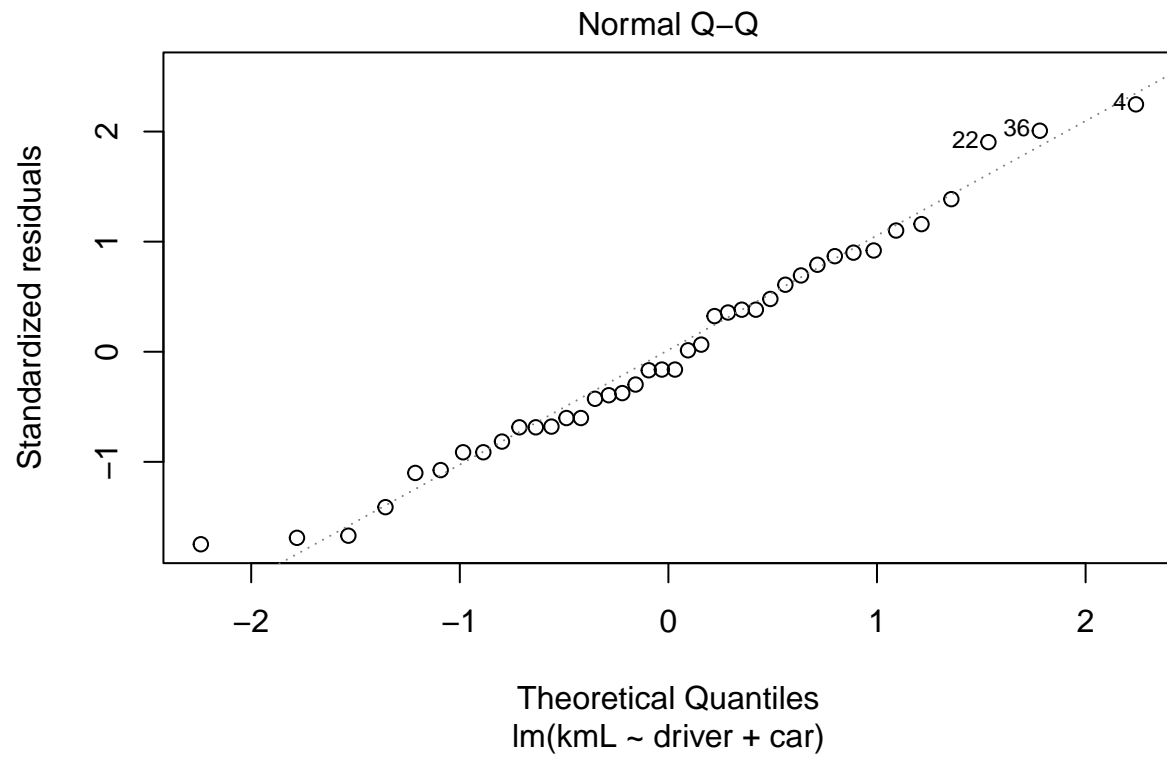
From the ANOVA table we can see the P-value for both driver and car are < 0.05 which means that both of them are significant and we have the evidence to reject H_0 in favour of H_1.
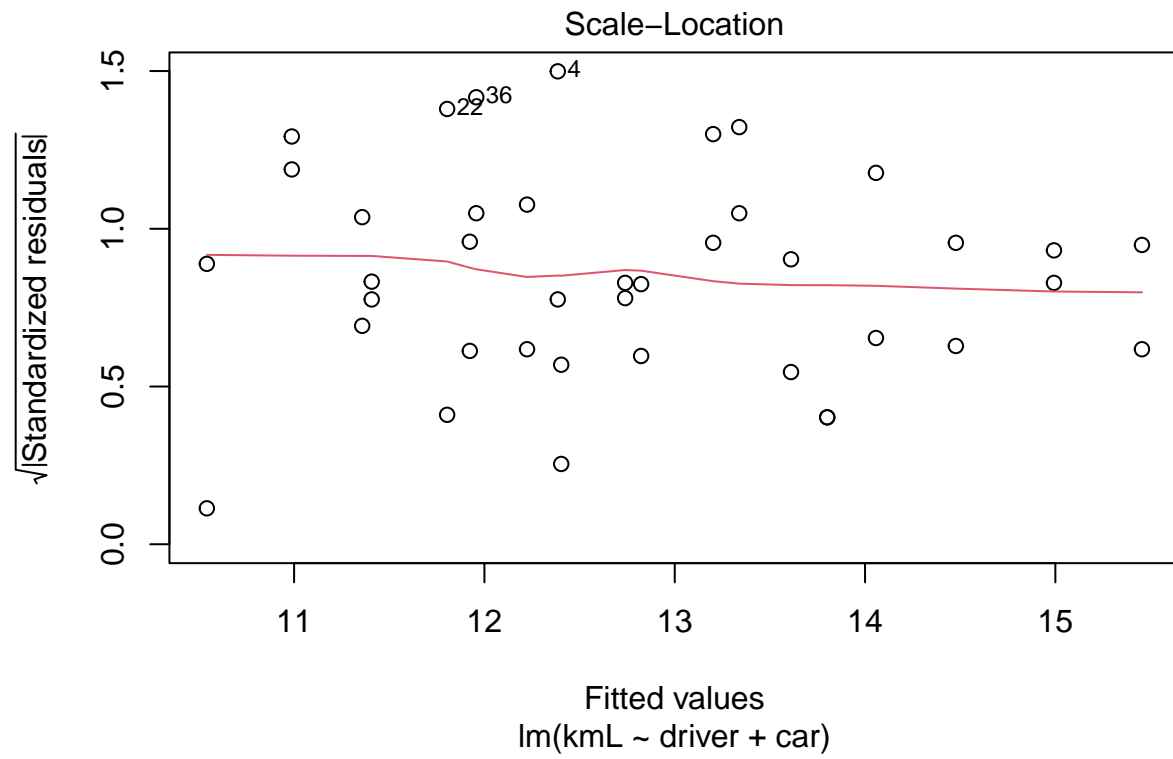
```
kml.final = lm(kmL ~ driver + car, data = mykml)
kml.final
```
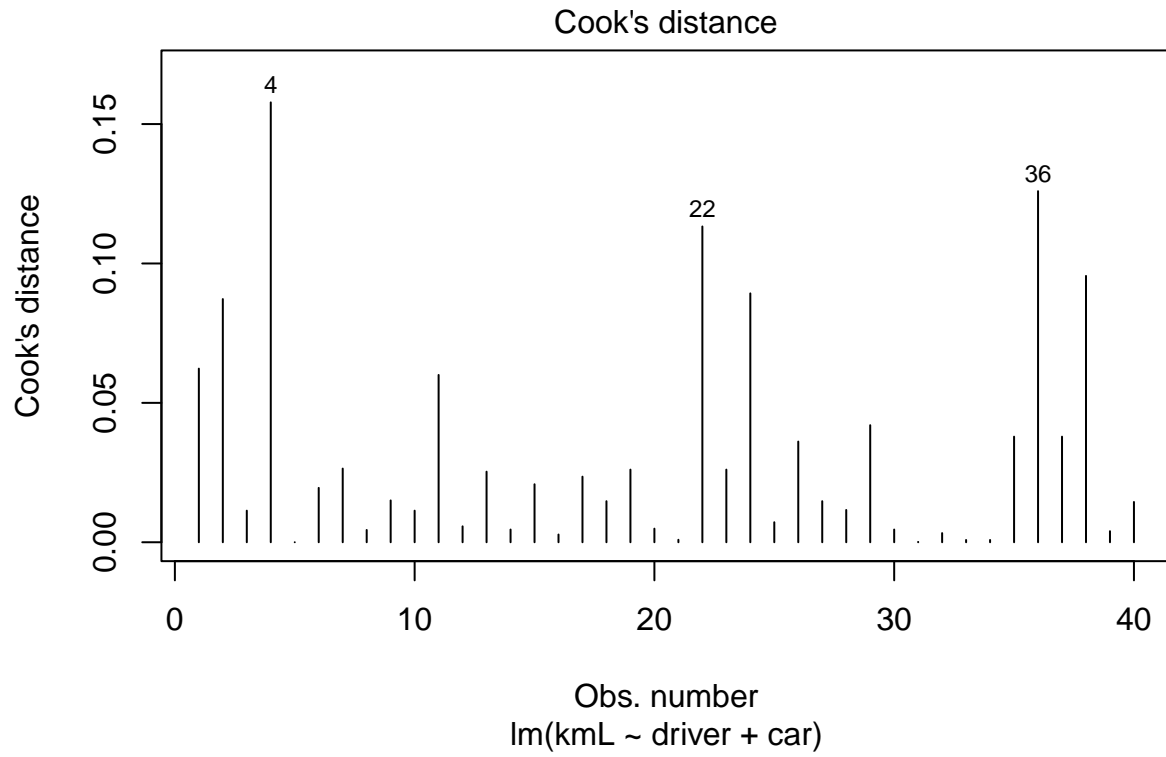
```
##
## Call:
## lm(formula = kmL ~ driver + car, data = mykml)
##
## Coefficients:
## (Intercept)      driverB      driverC      driverD      carfour       carone
##     11.4077       3.0695       0.8163       1.4157       0.5155      -0.4198
##    carthree       cartwo
##     -0.8662       0.9778
```

```
plot(kml.final, which = 1:4)
```

Residuals vs Fitted

Residuals

Fitted values
lm(kmL ~ driver + car)

Normal Q–Q

Theoretical Quantiles
lm(kmL ~ driver + car)

Scale–Location

lm(kmL ~ driver + car)

Cook's distance

Obs. number
lm(kmL ~ driver + car)

d. There is no pattern in residual vs. fitted value, variability seems pretty constant between effects. The normal quantile plot of residuals is very close to linear suggesting residuals are very close to normally distributed.