TD9

Charles Vin

2022-04-25

```
library(car)
```

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

```
library(emmeans)
```

Question 1

```
df = read.csv("./fultonfish.csv", sep=",")
head(df)
```

```
##     date     lprice      quan    lquan mon tue wed thu stormy mixed rainy cold
## 1 911202 -0.4307829  8058.003 8.994421   1   0   0   0      1     0     1    0
## 2 911203  0.0000000  2224.001 7.707063   0   1   0   0      1     0     0    0
## 3 911204  0.0723207  4231.001 8.350194   0   0   1   0      0     1     1    1
## 4 911205  0.2471390  5749.998 8.656955   0   0   0   1      1     0     0    1
## 5 911206  0.6643268  2551.001 7.844241   0   0   0   0      1     0     0    1
## 6 911209 -0.2065143 10952.000 9.301277   1   0   0   0      0     0     0    0
##    totr       diff change
## 1  7232  -826.0029      1
## 2  2110  -114.0012      0
## 3  5247  1015.9990      1
## 4  1290 -4459.9980      1
## 5  1717  -834.0010      1
## 6 11643   691.0020      1
```

Question 2

```
df$day = "fri"
df$day[which(df$mon==1)] = "mon"
df$day[which(df$tue==1)] = "tue"
df$day[which(df$wed==1)] = "wed"
df$day[which(df$thu==1)] = "thu"
df$day = as.factor(df$day)
df$day
```

```
##   [1] mon tue wed thu fri mon tue wed thu fri mon tue wed thu fri mon tue thu
##  [19] fri mon tue thu fri mon tue wed thu fri mon tue wed thu fri mon tue wed
##  [37] thu fri mon tue wed thu fri mon tue wed thu fri mon tue wed thu fri tue
##  [55] wed thu fri mon tue wed thu fri mon tue wed thu fri mon tue wed thu fri
##  [73] mon tue wed thu fri mon tue wed thu fri mon tue wed thu fri mon tue wed
```

```
##  [91] thu fri mon tue wed thu fri tue wed thu fri mon tue wed thu fri mon tue
## [109] wed thu fri
## Levels: fri mon thu tue wed
```

Question 3

```
df$wind = "low"
df$wind[which(df$stormy==1)] = "high"
df$wind[which(df$mixed==1)] = "medium"
df$wind = as.factor(df$wind)
df$wind
```

```
##   [1] high   high   medium high   high   low    medium low    medium low   
##  [11] high   high   high   high   high   medium medium medium low    low   
##  [21] low    low    low    high   high   high   medium medium medium medium
##  [31] high   high   high   high   medium low    low    low    high   medium
##  [41] low    low    low    high   high   high   high   medium high   high  
##  [51] high   medium medium low    medium low    medium low    low    medium
##  [61] high   high   medium high   medium medium low    medium medium medium
##  [71] high   high   low    low    medium low    high   low    medium medium
##  [81] medium medium high   high   medium low    low    low    low    low   
##  [91] low    low    low    low    low    low    low    low    low    low   
## [101] low    low    low    low    low    low    low    medium medium medium
## [111] medium
## Levels: high low medium
```

Question 4

```
df$rainy = as.factor(df$rainy)
df$cold = as.factor(df$cold)
str(df)
```

```
## 'data.frame':    111 obs. of  17 variables:
##  $ date  : int  911202 911203 911204 911205 911206 911209 911210 911211 911212 911213 ...
##  $ lprice: num  -0.4308 0 0.0723 0.2471 0.6643 ...
##  $ quan  : num  8058 2224 4231 5750 2551 ...
##  $ lquan : num  8.99 7.71 8.35 8.66 7.84 ...
##  $ mon   : int  1 0 0 0 0 1 0 0 0 0 ...
##  $ tue   : int  0 1 0 0 0 0 1 0 0 0 ...
##  $ wed   : int  0 0 1 0 0 0 0 1 0 0 ...
##  $ thu   : int  0 0 0 1 0 0 0 0 1 0 ...
##  $ stormy: int  1 1 0 1 1 0 0 0 0 0 ...
##  $ mixed : int  0 0 1 0 0 0 1 0 1 0 ...
##  $ rainy : Factor w/ 2 levels "0","1": 2 1 2 1 1 1 1 2 1 1 ...
##  $ cold  : Factor w/ 2 levels "0","1": 1 1 2 2 2 1 1 1 1 1 ...
##  $ totr  : int  7232 2110 5247 1290 1717 11643 9640 9347 3890 16318 ...
##  $ diff  : num  -826 -114 1016 -4460 -834 ...
##  $ change: int  1 0 1 1 1 1 1 0 0 1 ...
##  $ day   : Factor w/ 5 levels "fri","mon","thu",..: 2 4 5 3 1 2 4 5 3 1 ...
##  $ wind  : Factor w/ 3 levels "high","low","medium": 1 1 3 1 1 2 3 2 3 2 ...
```

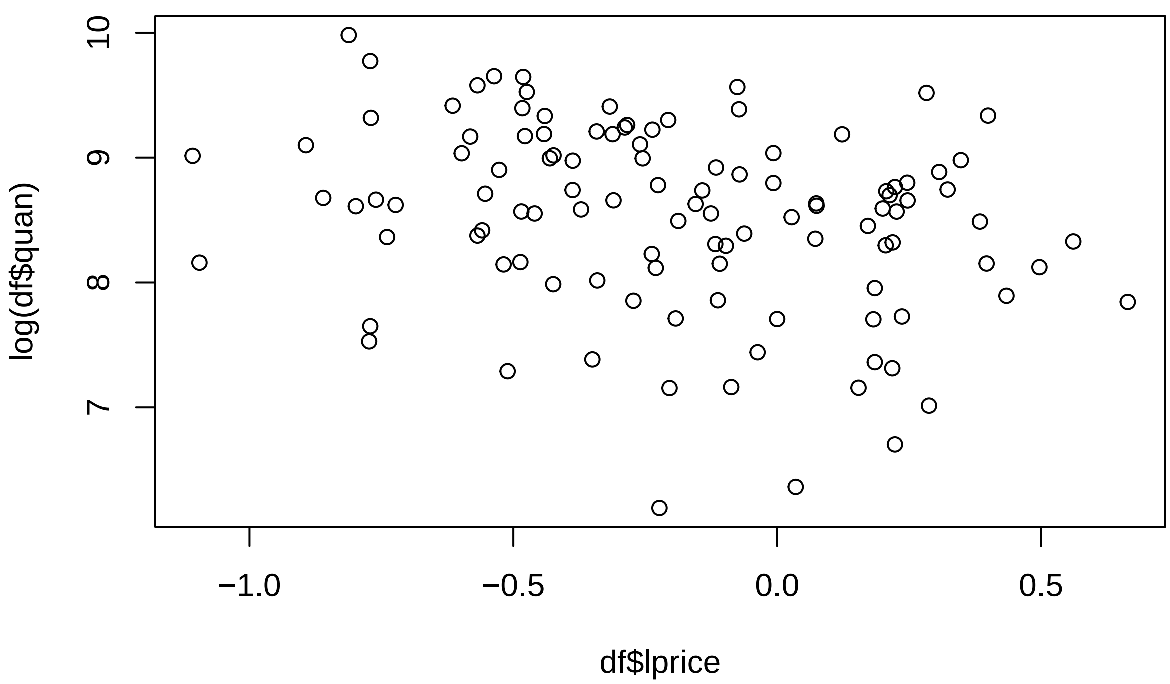
```
df = df[,-c(5,6,7,8,9,10)]
str(df)
```

```
## 'data.frame':    111 obs. of  11 variables:
##  $ date  : int  911202 911203 911204 911205 911206 911209 911210 911211 911212 911213 ...
##  $ lprice: num  -0.4308 0 0.0723 0.2471 0.6643 ...
##  $ quan  : num  8058 2224 4231 5750 2551 ...
##  $ lquan : num  8.99 7.71 8.35 8.66 7.84 ...
##  $ rainy : Factor w/ 2 levels "0","1": 2 1 2 1 1 1 1 2 1 1 ...
##  $ cold  : Factor w/ 2 levels "0","1": 1 1 2 2 2 1 1 1 1 1 ...
##  $ totr  : int  7232 2110 5247 1290 1717 11643 9640 9347 3890 16318 ...
##  $ diff  : num  -826 -114 1016 -4460 -834 ...
##  $ change: int  1 0 1 1 1 1 1 0 0 1 ...
##  $ day   : Factor w/ 5 levels "fri","mon","thu",..: 2 4 5 3 1 2 4 5 3 1 ...
##  $ wind  : Factor w/ 3 levels "high","low","medium": 1 1 3 1 1 2 3 2 3 2 ...
```

Question 5

Relation prix (en log) et demande (en log)

```
plot(df$lprice, log(df$quan))
```



On remarque une sorte de corrélation négative. Si le prix augmente la demande diminue

Question b

```
res = lm(log(df$quan)~df$lprice)
res
```

```
##
## Call:
## lm(formula = log(df$quan) ~ df$lprice)
##
## Coefficients:
## (Intercept)    df$lprice
##      8.4187      -0.5409
```

```r
summary(res)
```

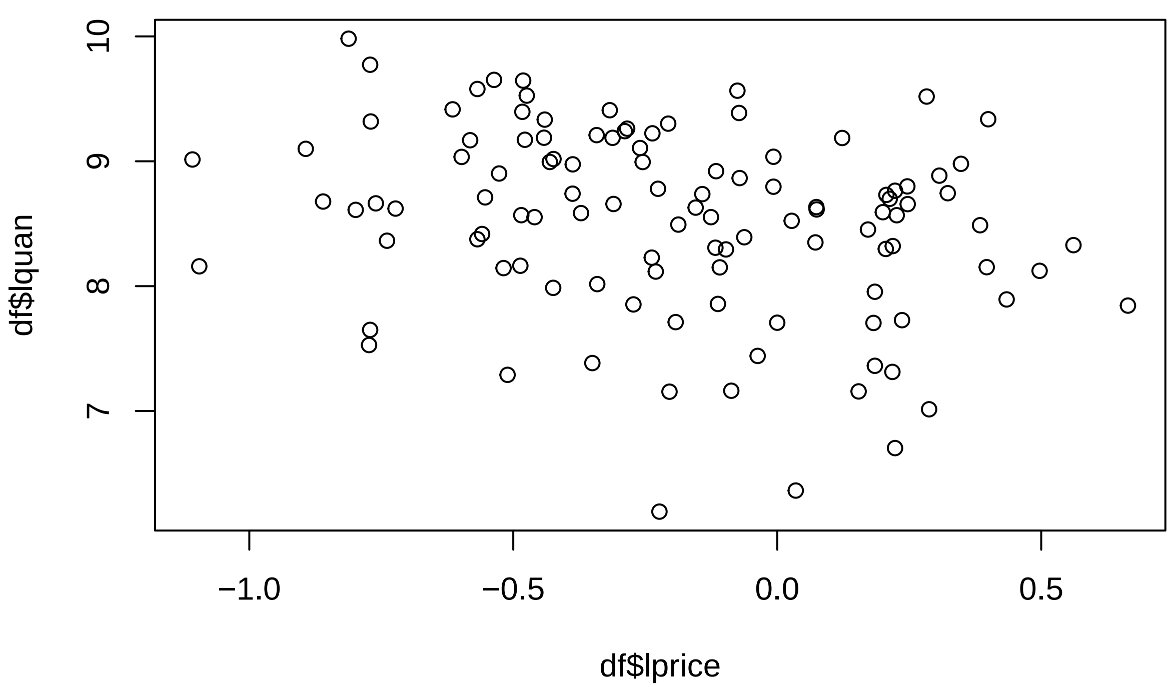
```
##
## Call:
## lm(formula = log(df$quan) ~ df$lprice)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3450 -0.3569  0.1193  0.4976  1.2528
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.41867    0.07622 110.445  < 2e-16 ***
## df$lprice   -0.54087    0.17864  -3.028  0.00308 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7156 on 109 degrees of freedom
## Multiple R-squared:  0.07758,    Adjusted R-squared:  0.06912
## F-statistic: 9.167 on 1 and 109 DF,  p-value: 0.003075
```

"p-value: 0.003075" le modèle est utile très significativement.

$$Y_i = \alpha + \beta x_i + E_i Y_i = \text{quantité en log} x_i = \text{prix en log} H_0 = \{\beta = 0\} H_1 = \{\beta \neq 0\} F = \frac{SCM/1}{SCR/n-2}$$

Question c

```r
plot(df$lprice, df$lquan)
```




```r
# lines(c(-1.5,1), predict(res, )
```

Question 6

Lorsqu'il y a de mauvaise condition météo, les prix augmentent.

Question 6.a

Il faut vérifier la phrase précédant. On est face à une ANOVA, 3 facteur sur une variable quantitative. Est-ce que j'ai des interaction

$$Y_i = a + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \delta_{ik} + \delta_{jk} + \delta_{ijk} + \delta_{ij?} i = vent j = pluie k = froid$$

```
res_weather = lm(lprice~wind*rainy*cold, data=df)
anova(res_weather)
```

```
## Analysis of Variance Table
##
## Response: lprice
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## wind             2  3.6383 1.81914 16.1017 8.824e-07 ***
## rainy            1  0.0012 0.00125  0.0110   0.91655
## cold             1  0.0725 0.07250  0.6417   0.42502
## wind:rainy       2  0.9333 0.46667  4.1306   0.01893 *
## wind:cold        2  0.0778 0.03888  0.3442   0.70966
## rainy:cold       1  0.0907 0.09071  0.8029   0.37240
## wind:rainy:cold  2  0.0475 0.02375  0.2102   0.81077
## Residuals       99 11.1848 0.11298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(res_weather)
```

```
## Anova Table (Type II tests)
##
## Response: lprice
##                  Sum Sq Df F value    Pr(>F)
## wind             2.7753  2 12.2824 1.719e-05 ***
## rainy            0.0004  1  0.0032   0.95492
## cold             0.0664  1  0.5874   0.44524
## wind:rainy       0.9822  2  4.3468   0.01551 *
## wind:cold        0.0815  2  0.3606   0.69818
## rainy:cold       0.0907  1  0.8029   0.37240
## wind:rainy:cold  0.0475  2  0.2102   0.81077
## Residuals       11.1848 99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

cold dans un modèle avec wind et rainy n'est pas significatif, ni ces interractions -> on peut le mettre à la poubelle On remarque également que l'intéraction wind:rainy est significative malgrè la non significativité pas additivité de rainy.

```
res_weather = lm(lprice~wind*rainy, data=df)
summary(res_weather)
```

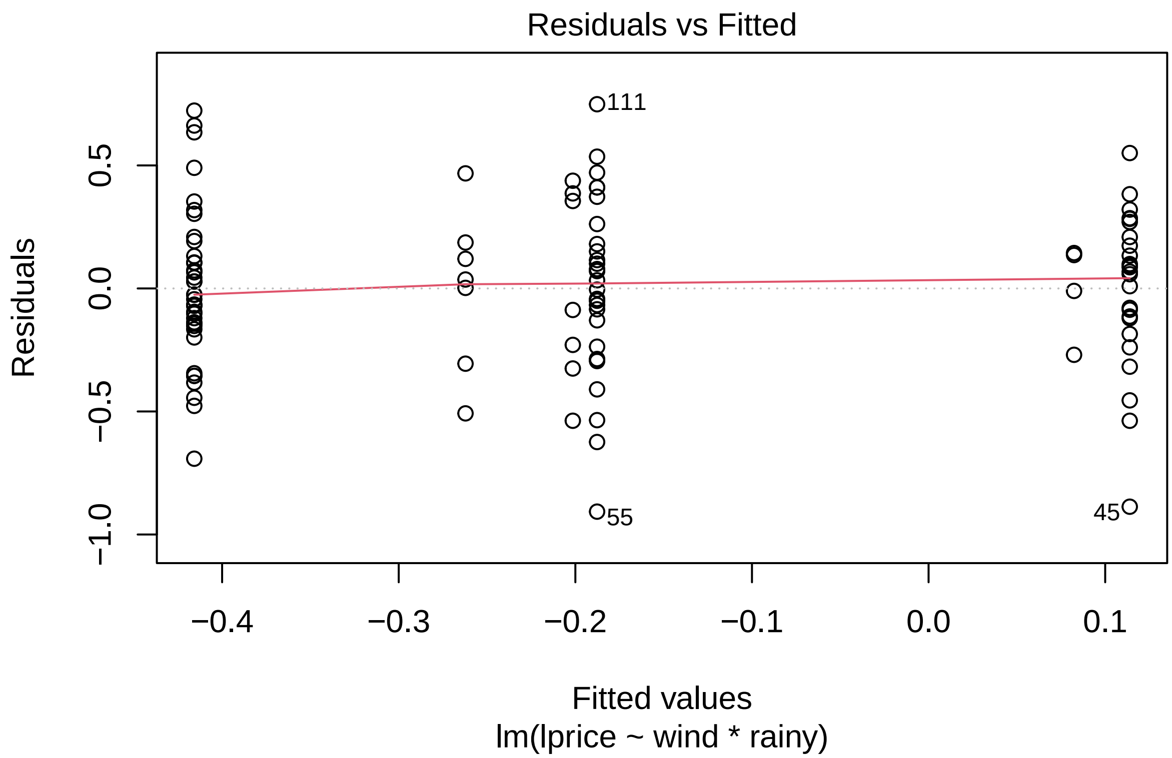
```
##
## Call:
## lm(formula = lprice ~ wind * rainy, data = df)
##
```

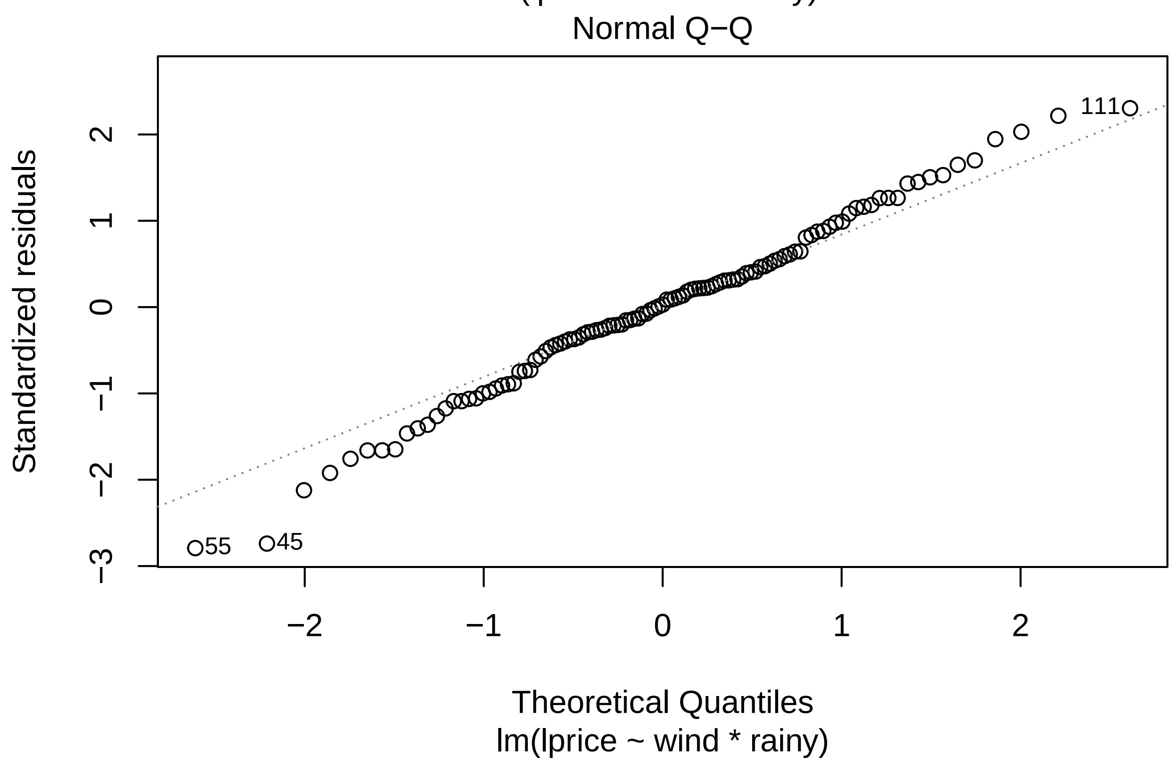
```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90708 -0.17555  0.00915  0.18365  0.74881
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.11391    0.06609   1.724  0.08774 .
## windlow           -0.52969    0.08510  -6.224 1.01e-08 ***
## windmedium        -0.30161    0.08949  -3.370  0.00105 **
## rainy1            -0.31536    0.14132  -2.232  0.02777 *
## windlow:rainy1     0.46894    0.19608   2.392  0.01855 *
## windmedium:rainy1  0.58542    0.22564   2.594  0.01083 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3305 on 105 degrees of freedom
## Multiple R-squared:  0.2854, Adjusted R-squared:  0.2513
## F-statistic: 8.386 on 5 and 105 DF,  p-value: 1.058e-06
```

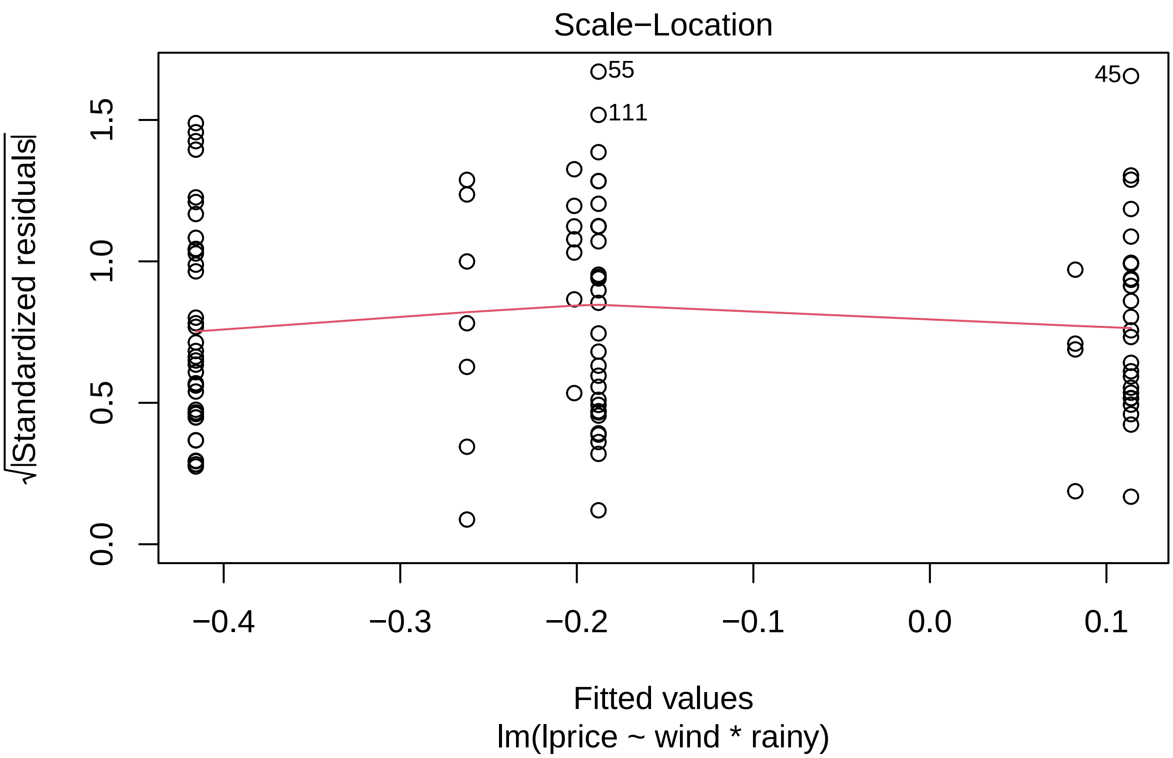
```
anova(res_weather)
```

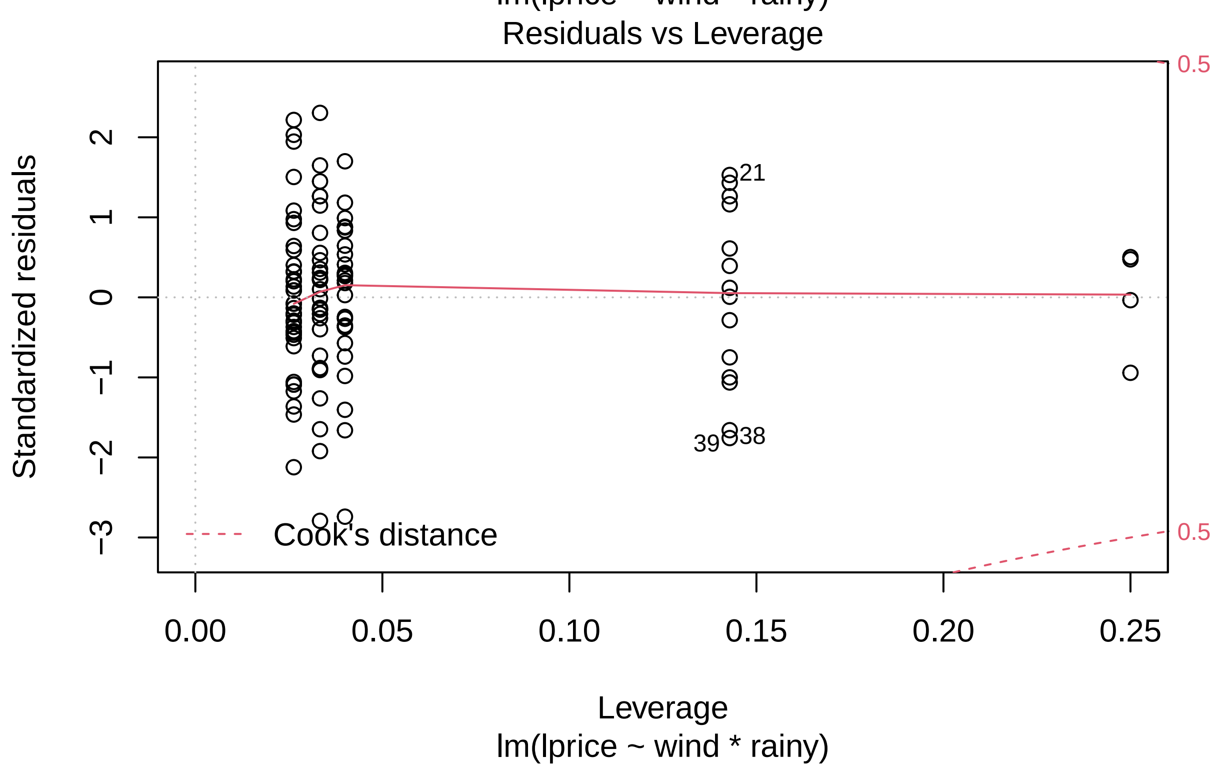
```
## Analysis of Variance Table
##
## Response: lprice
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## wind         2  3.6383 1.81914 16.6571 5.211e-07 ***
## rainy        1  0.0012 0.00125  0.0114   0.91512
## wind:rainy   2  0.9395 0.46973  4.3011   0.01602 *
## Residuals  105 11.4672 0.10921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(res_weather)
```


Residuals vs Fitted
Residuals
0.5
0.0
−0.5
−1.0
111
55
45
−0.4
−0.3
−0.2
−0.1
0.0
0.1
Fitted values
lm(lprice ~ wind * rainy)


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
111
55
45
−2
−1
0
1
2
Theoretical Quantiles
lm(lprice ~ wind * rainy)





Les HP sont vérifiés.

On a bien un effet du vent sur le prix .

Question c

```
emmeans(res_weather, ~ wind+rainy)
```

```
##  wind   rainy  emmean     SE  df lower.CL upper.CL
##  high   0      0.1139 0.0661 105  -0.0171   0.2450
##  low    0     -0.4158 0.0536 105  -0.5221  -0.3095
##  medium 0     -0.1877 0.0603 105  -0.3073  -0.0681
##  high   1     -0.2014 0.1249 105  -0.4491   0.0462
##  low    1     -0.2622 0.1249 105  -0.5099  -0.0145
##  medium 1      0.0824 0.1652 105  -0.2453   0.4100
##
## Confidence level used: 0.95
```

Le prix en log augmente lorsque le vent est fort et q'uil ne pleut pas est de 0.1139 en log. Le prix lorsque qu'estimer pas le modèle est 0.1139, en terme mathématique c'est $\hat{\mu} + \hat{\alpha_{high}} + \hat{\beta_{no}} + \hat{\delta_{high,no}}$

```
by(df$lprice, list(df$rainy, df$wind), FUN=mean)
```

```
## : 0
## : high
## [1] 0.1139144
## ---------------------------------------------------------------
## : 1
## : high
## [1] -0.2014474
## ---------------------------------------------------------------
## : 0
## : low
## [1] -0.4157789
## ---------------------------------------------------------------
## : 1
## : low
## [1] -0.2622
## ---------------------------------------------------------------
## : 0
## : medium
## [1] -0.187693
## ---------------------------------------------------------------
## : 1
## : medium
## [1] 0.082364
```

ça c'est les moyenne empiriques.

Complément, de base on estime

$$Y = X\beta + E\hat{\beta} = (X^T X)^{-1} X^T Y$$

```
table(df$wind, df$rainy)
```

```
##
##           0  1
##   high   25  7
##   low    38  7
##   medium 30  4
```

Le fait abscence de pluie est plus important on a plus d'exemple.

Question 7

On a vu que le vent et la pluis on une influence sur le prix. On a vu que le prix à une influence sur la demande. Maintenant est-ce que le jour de la semaine à une influence sur la demande

```
res2 = lm(log(df$quan)~df$lprice*df$day)
res2
```

```
##
## Call:
## lm(formula = log(df$quan) ~ df$lprice * df$day)
##
## Coefficients:
##         (Intercept)            df$lprice            df$daymon
##             8.59941             -0.60379              0.02581
##           df$daythu            df$daytue            df$daywed
##             0.08021             -0.50749             -0.51997
## df$lprice:df$daymon  df$lprice:df$daythu  df$lprice:df$daytue
##             0.05715             -0.02630              0.04736
## df$lprice:df$daywed
##             0.18011
```

Effet significatif de certain jour -> à garder + jour a une influence Voir modèle mathématique dans onenote + graphique de l'ancova L'avantage par rapport à une simple regression c'est d'estimer toute les erreurs en même temps ICI ON FAIT UNE ANCOVA

```
anova(res2)
```

```
## Analysis of Variance Table
##
## Response: log(df$quan)
##                   Df Sum Sq Mean Sq F value   Pr(>F)
## df$lprice          1  4.694  4.6942 10.0667 0.002001 **
## df$day             4  8.647  2.1618  4.6359 0.001775 **
## df$lprice:df$day   4  0.070  0.0175  0.0374 0.997286
## Residuals        101 47.097  0.4663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L'intéraction n'est pas nécéssaire.

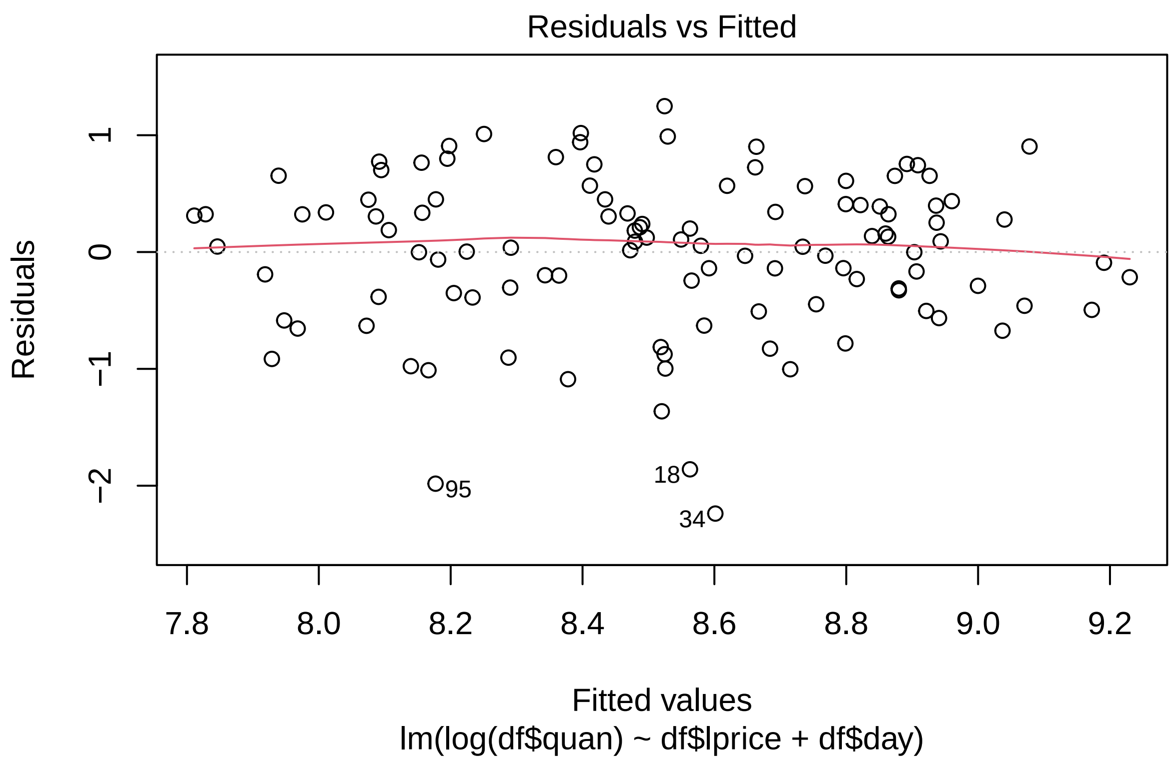
```
res2 = lm(log(df$quan)~df$lprice+df$day)
```

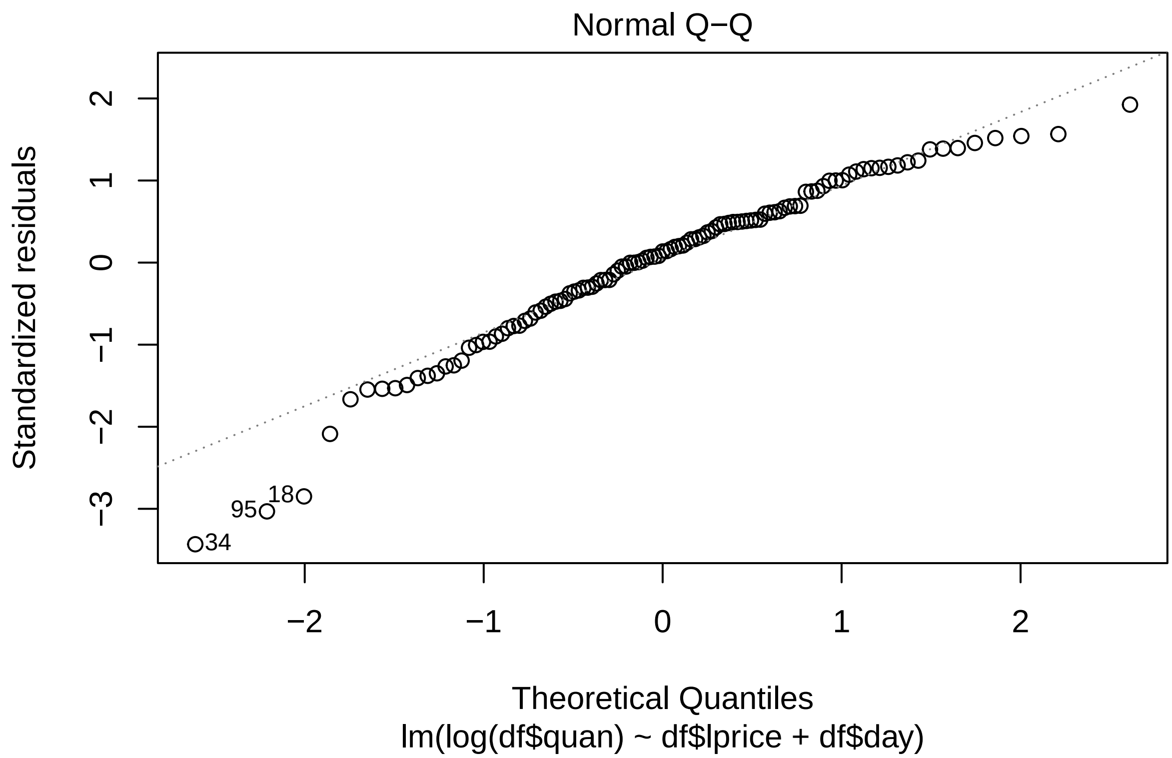
```
summary(res2)
```

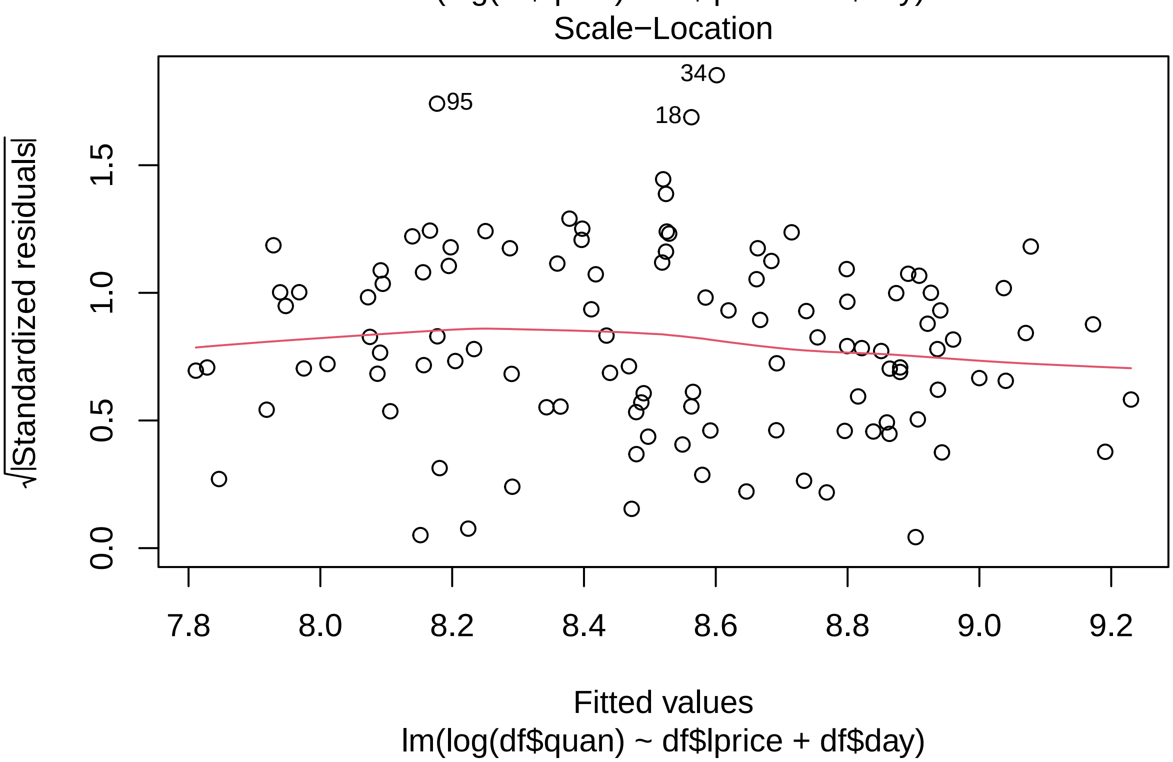
```
##
## Call:
## lm(formula = log(df$quan) ~ df$lprice + df$day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23844 -0.36738  0.08832  0.42304  1.24866
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.60689    0.14304  60.170  < 2e-16 ***
## df$lprice   -0.56255    0.16821  -3.344  0.00114 **
## df$daymon    0.01432    0.20265   0.071  0.94381
```

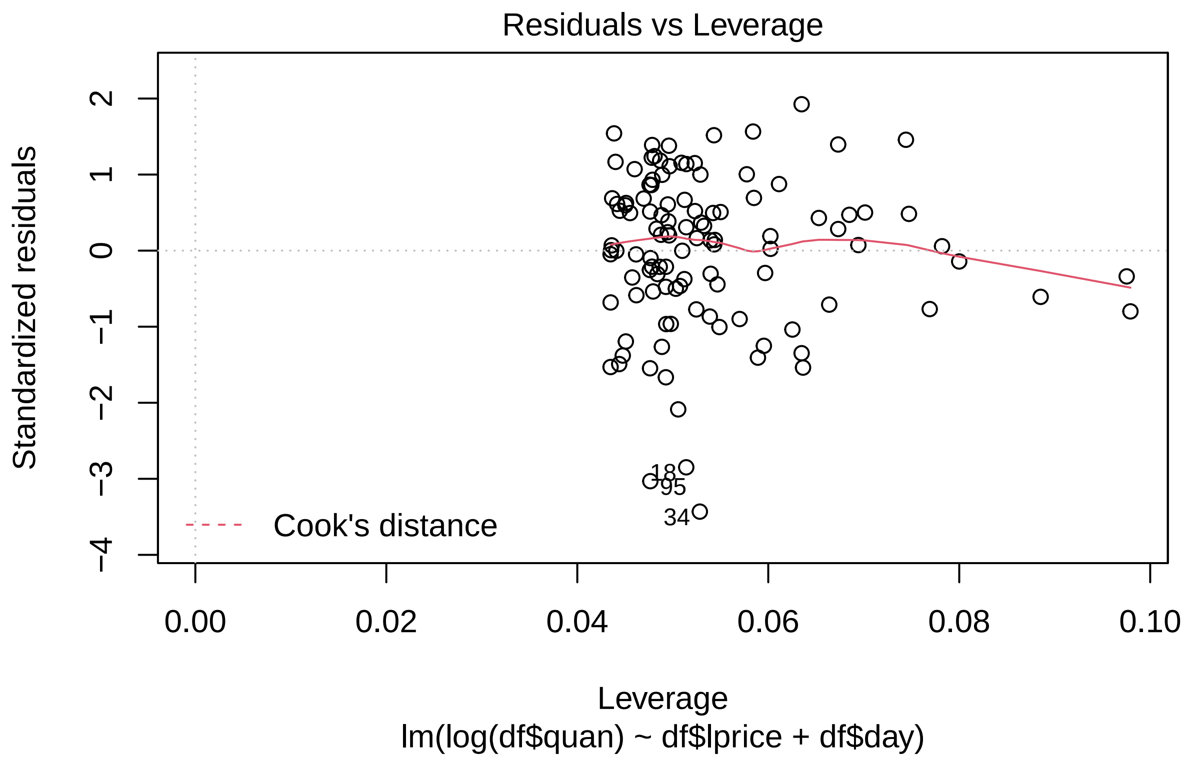
```
## df$daythu    0.08162    0.19782   0.413  0.68073
## df$daytue   -0.51624    0.19769  -2.611  0.01034 *
## df$daywed   -0.55537    0.20232  -2.745  0.00712 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6702 on 105 degrees of freedom
## Multiple R-squared:  0.2205, Adjusted R-squared:  0.1834
## F-statistic:  5.94 on 5 and 105 DF,  p-value: 7.08e-05
```

```
plot(res2)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(df$quan) ~ df$lprice + df$day)
34
95
18
−3
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
Fitted values
lm(log(df$quan) ~ df$lprice + df$day)
34
95
18
0.0
0.5
1.0
1.5
7.8
8.0
8.2
8.4
8.6
8.8
9.0
9.2



quelque pointe abérant mais pas de fort levier mais fort résidue

Question 7.b

Comme on est sans interraction -> les droites sont parallèle, il suffit alors de comparer les ordonnée à l'origine. Mais idk pourquoi elle a dit ça car au final on fait un eemeans

```
emmeans(res2, ~day)
```

```
##  day emmean    SE  df lower.CL upper.CL
##  fri   8.72 0.140 105     8.44     8.99
##  mon   8.73 0.147 105     8.44     9.02
##  thu   8.80 0.140 105     8.52     9.08
##  tue   8.20 0.140 105     7.92     8.48
##  wed   8.16 0.146 105     7.87     8.45
##
## Results are given on the log (not the response) scale.
## Confidence level used: 0.95
```

On voit les min et max ....

Question 7.c

```
contrast(emmeans(res2, ~day), adjust="bonferroni", method="pairwise")
```

```
##  contrast  estimate    SE  df t.ratio p.value
##  fri - mon  -0.0143 0.203 105  -0.071  1.0000
##  fri - thu  -0.0816 0.198 105  -0.413  1.0000
##  fri - tue   0.5162 0.198 105   2.611  0.1034
##  fri - wed   0.5554 0.202 105   2.745  0.0712
##  mon - thu  -0.0673 0.203 105  -0.331  1.0000
##  mon - tue   0.5306 0.202 105   2.621  0.1007
```

```
##  mon - wed   0.5697 0.207 105   2.752  0.0698
##  thu - tue   0.5979 0.198 105   3.019  0.0319
##  thu - wed   0.6370 0.203 105   3.144  0.0217
##  tue - wed   0.0391 0.202 105   0.193  1.0000
##
## Results are given on the log (not the response) scale.
## P value adjustment: bonferroni method for 10 tests
```

Question 7.d

Pour plot elle a fait un truc du futur. Elle a dit qu'elle allait envoyer

Question 7.e

Pourquoi le log ? pour obtenir une élasticité. Voir one note pour un peu plus d'info. L'élasticité est estimé par β

```
summary(res2)
```

```
##
## Call:
## lm(formula = log(df$quan) ~ df$lprice + df$day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23844 -0.36738  0.08832  0.42304  1.24866
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.60689    0.14304  60.170  < 2e-16 ***
## df$lprice   -0.56255    0.16821  -3.344  0.00114 **
## df$daymon    0.01432    0.20265   0.071  0.94381
## df$daythu    0.08162    0.19782   0.413  0.68073
## df$daytue   -0.51624    0.19769  -2.611  0.01034 *
## df$daywed   -0.55537    0.20232  -2.745  0.00712 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6702 on 105 degrees of freedom
## Multiple R-squared:  0.2205, Adjusted R-squared:  0.1834
## F-statistic:  5.94 on 5 and 105 DF,  p-value: 7.08e-05
```

On a $\beta = -0.56255$. Si c'est nul ça veut dire que le prix n'a pas d'influence sur le prix.