Modèle de régression simple - Pollution de l'air et trafic

Sarah Ouadah

Module DYST, M2 BEE, Septembre 2023

Données

On s'intéresse ici à l'effet du trafic routier ou de certaines conditions météorologiques sur la pollution de l'air en dioxyde d'azote via l'étude de données publiques dont la description est la suivante (http://lib.stat.cmu.edu/datasets/NO2.dat) :

"The data are a subsample of 500 observations from a data set that originate in a study where air pollution at a road is related to traffic volume and meteorological variables, collected by the Norwegian Public Roads Administration. The response variable (column 1) consist of hourly values of the logarithm of the concentration of NO2 (particles), measured at Alnabru in Oslo, Norway, between October 2001 and August 2003. The predictor variables (columns 2 to 8) are the logarithm of the number of cars per hour, temperature 2 meter above ground (degree C), wind speed (meters/second), the temperature difference between 25 and 2 meters above ground (degree C), wind direction (degrees between 0 and 360), hour of day and day number from October 1. 2001."

```r
pollution <- read.table("NO2.txt")
names(pollution) <- c("NO2", "logCar", "temp", "wind", "temp_diff",
                      "wind_direc", "hour", "day")
head(pollution)
```

```
##       NO2  logCar temp wind temp_diff wind_direc hour day
## 1 3.71844 7.69120  9.2  4.8      -0.1       74.4   20 600
## 2 3.10009 7.69894  6.4  3.5      -0.3       56.0   14 196
## 3 3.31419 4.81218 -3.7  0.9      -0.1      281.3    4 513
## 4 4.38826 6.95177 -7.2  1.7       1.2       74.0   23 143
## 5 4.34640 7.51806 -1.3  2.6      -0.1       65.0   11 115
## 6 4.16044 7.67183  2.6  1.6       0.3      224.2   19 527
```

```r
str(pollution)
```

```
## 'data.frame':    500 obs. of  8 variables:
##  $ NO2       : num  3.72 3.1 3.31 4.39 4.35 ...
##  $ logCar    : num  7.69 7.7 4.81 6.95 7.52 ...
##  $ temp      : num  9.2 6.4 -3.7 -7.2 -1.3 2.6 -7.9 -4.1 -12.7 -1.6 ...
##  $ wind      : num  4.8 3.5 0.9 1.7 2.6 1.6 1.6 3.8 5.2 3 ...
##  $ temp_diff : num  -0.1 -0.3 -0.1 1.2 -0.1 0.3 0.3 -0.1 -0.1 0.4 ...
##  $ wind_direc: num  74.4 56 281.3 74 65 ...
##  $ hour      : int  20 14 4 23 11 19 5 4 12 3 ...
##  $ day       : int  600 196 513 143 115 527 502 453 462 554 ...
```

Dans la suite, nous étudierons l'influence du trafic `logCar` sur la pollution `NO2`.

Statistiques descriptives

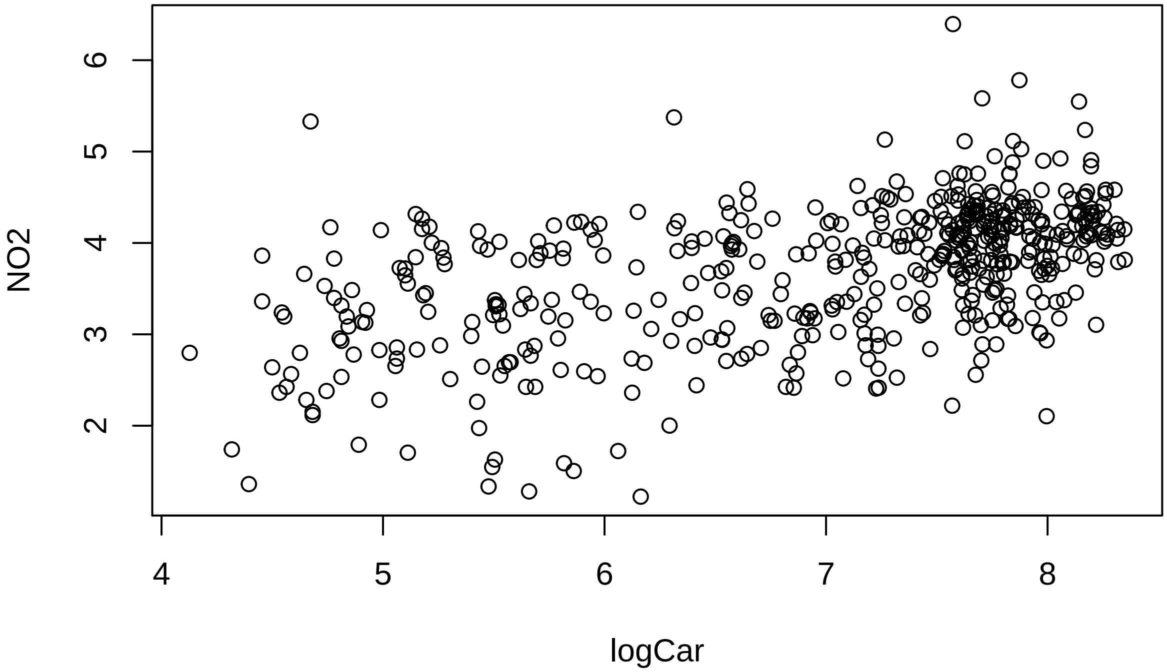
```r
summary(pollution$logCar)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.127   6.176   7.425   6.973   7.793   8.349
```

```r
summary(pollution$NO2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.224   3.214   3.848   3.698   4.217   6.395
```

```r
plot(pollution$logCar,pollution$NO2, xlab="logCar", ylab="NO2")
```




```r
cor(pollution$logCar,pollution$NO2)
```

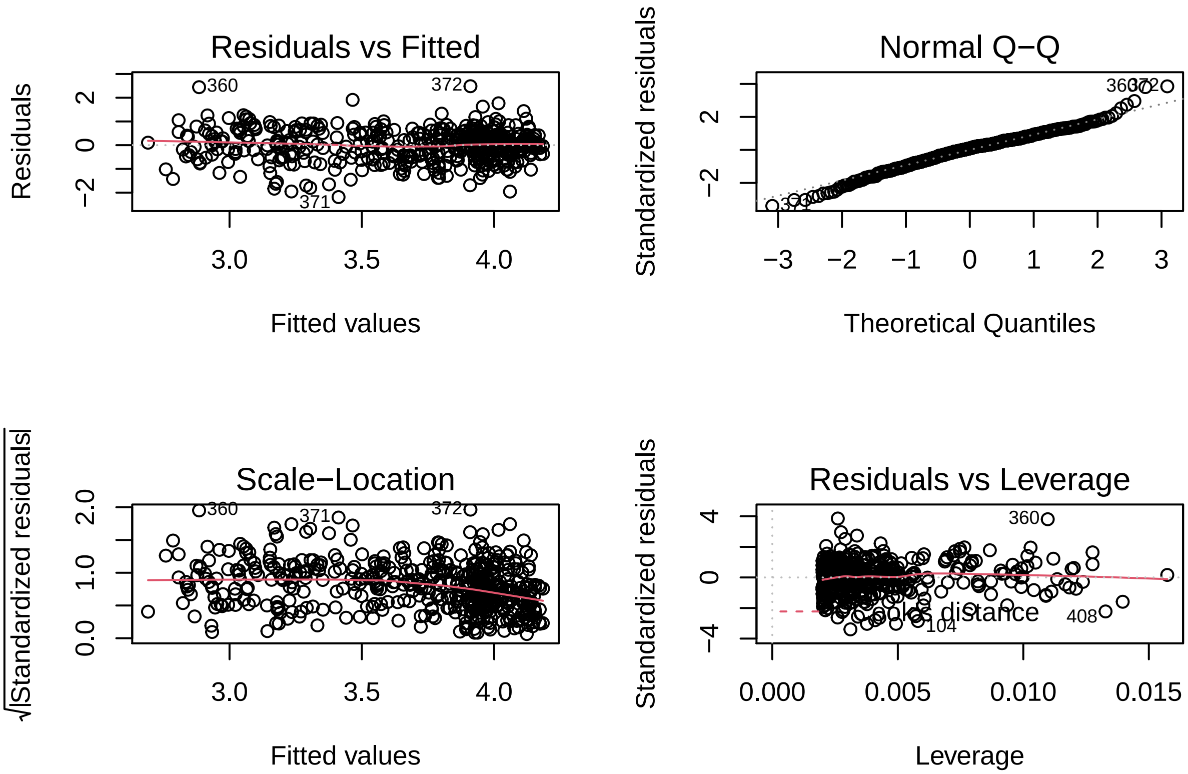
```
## [1] 0.5120504
```

Modèle

```r
pol.lm <- lm(NO2~logCar, data=pollution)
```

Hypothèses du modèle

```r
par(mfrow=c(2,2))
plot(pol.lm)
```



Test du modèle - estimation et tests des paramètres

```
anova(pol.lm)
```

```
## Analysis of Variance Table
##
## Response: NO2
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## logCar      1  73.712  73.712  176.98 < 2.2e-16 ***
## Residuals 498 207.422   0.417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(pol.lm)
```

```
##
## Call:
## lm(formula = NO2 ~ logCar, data = pollution)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.18822 -0.40071  0.06428  0.40362  2.48472
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.23310    0.18755   6.575 1.23e-10 ***
## logCar       0.35353    0.02657  13.303  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.6454 on 498 degrees of freedom
## Multiple R-squared:  0.2622, Adjusted R-squared:  0.2607
## F-statistic:   177 on 1 and 498 DF,  p-value: < 2.2e-16
```

```
plot(pollution$logCar,pollution$NO2, xlab="logCar", ylab="NO2")
abline(coef(pol.lm),col="blue")
```