

Analyse des Facteurs de Gravité des Accidents Routiers (ONISR 2021)

Régression linéaire

Contents

Introduction	2
Préparation des Données	2
Importation et Nettoyage	2
Modèle de Régression Linéaire	11
Modèle 1	11
Modèle 2	16
Modèle 3	21
Facteurs les plus influents sur la gravité	23

```
# Chargement des librairies utiles
if(!require(dplyr)) install.packages("dplyr")

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##     filter, lag

## The following objects are masked from 'package:base':
##     intersect, setdiff, setequal, union

if(!require(stringr)) install.packages("stringr")

## Loading required package: stringr
```

```

if(!require(car)) install.packages("car")

## Loading required package: car

## Warning: package 'car' was built under R version 4.3.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.3.3

## 
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
## 
##     recode

if(!require(ggplot2)) install.packages("ggplot2")

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.3.3

library(dplyr)
library(stringr)
library(car)
library(ggplot2)

```

Introduction

L'objectif est de déterminer les facteurs qui ont le plus d'impact sur la sévérité d'un accident de la route en France en 2021. Nous cherchons donc à expliquer un **Score de Sévérité** à partir de variables issues de la base de données ONISR 2021 (Observatoire National Interministériel de la Sécurité Routière).

Préparation des Données

Importation et Nettoyage

```

df <- read.csv("ONISR-2021.csv")
head(df)

##           Num_Acc   jour   mois   an   hrmn      lum   dep   com
## 1 N202100000001    30    11 2021 07:32    crepuscule ou aube  30 30319
## 2 N202100000002    25     9 2021 14:20        plein   jour  51 51544
## 3 N202100000003    15     7 2021 07:55        plein   jour  85 85048

```

```

## 4 N202100000004 27 3 2021 19:45 nuit avec eclairage allume 93 93005
## 5 N202100000005 25 2 2021 07:20 nuit avec eclairage allume 76 76429
## 6 N202100000006 23 11 2021 11:10 plein jour 68 68004
##           agg   int      atm          col
## 1 hors agglo aucune normale deux veh - frontal
## 2 hors agglo en T   normale   deux veh - cote
## 3 en agglo aucune eblouissant autre
## 4 en agglo en X   pluie forte autre
## 5 en agglo aucune normale deux veh - arriere
## 6 en agglo en T   normale deux veh - arriere
##                                     adr      lat      long
## 1                               CD 981 44.03896 4.348022
## 2 Aire de repos croisement D20E9 après sortie D977 49.24213 4.554546
## 3                               15 rue François Nicolas 46.92195 -0.964460
## 4                               Route de Mitry 48.94936 2.519664
## 5                               PARIS. ROUTE DE 49.40838 1.145810
## 6 Rue Charles Edouard Amiot 47.61424 7.234336
##           date      catr          voie v1 v2
## 1 2021-11-30 departementale         981 -1 N/A
## 2 2021-09-25 departementale         20  0 E
## 3 2021-07-15  communale GENERAL GALLIENI (BOULEVARD DU) 0 N/A
## 4 2021-03-27 departementale GENERAL GALLIENI (BOULEVARD DU) 0 N/A
## 5 2021-02-25 metropole            PARIS. ROUTE DE -1 N/A
## 6 2021-11-23 communale             N/A -1 N/A
##           circ nbv      vosp prof pr pr1      plan lartpc larrout      surf
## 1 bidirectionnel 2 sans objet plat NA  NA rectiligne  NA  NA normale
## 2 bidirectionnel 2 sans objet plat 0   10 rectiligne  NA  NA normale
## 3 bidirectionnel 2 sans objet plat NA  NA rectiligne  NA  NA normale
## 4 bidirectionnel 4 sans objet plat 0   1 rectiligne  NA  NA normale
## 5 bidirectionnel 2 <NA> plat 11 150 rectiligne  NA  NA normale
## 6 bidirectionnel 2 sans objet plat NA  NA rectiligne  NA   4 normale
##           infra   situ vma nb_indemne nb_tue nb_hospitalise nb_blesse nb_velo nb_moto
## 1 aucun chaussee 80    1   0       1   0     1   1   0
## 2 aucun chaussee 80    0   0       1   1     0   0   0
## 3 aucun chaussee 50    1   0       1   0     0   0   0
## 4 aucun chaussee 50    1   1       0   0     0   0   0
## 5 aucun chaussee 50    0   0       0   0     1   0   0
## 6 aucun chaussee 50    1   0       0   0     1   1   0
##           nb_voiture nb_utilitaire nb_camion nb_tracteur nb_autre nb_transport
## 1           1        0        0        0        0        0
## 2           2        0        0        0        0        0
## 3           1        0        0        0        0        0
## 4           1        0        0        0        1        0
## 5           2        0        0        0        0        0
## 6           1        0        0        0        0        0

```

Pour le traitement des données, nous allons commencer par définir un **Score de Sévérité** pondéré par le nombre de tué, hospitalisé et blessé. Cette pondération est basé sur l'échelle AIS (Abbreviated Injury Scale); 1 pour une blessure légère, 3 pour une blessure sérieuse et 6 pour une blessure maximale.

```

# On ajoute ce score de sévérité dans notre df
df <- df %>%
  mutate(
    Score_Severite = 1*nb_blesse + 3*nb_hospitalise + 6*nb_tue,
  )

```

```

)
colnames(df)

## [1] "Num_Acc"      "jour"        "mois"        "an"
## [5] "hrmn"         "lum"         "dep"         "com"
## [9] "agg"          "int"         "atm"         "col"
## [13] "adr"          "lat"         "long"        "date"
## [17] "catr"         "voie"        "v1"          "v2"
## [21] "circ"         "nbv"         "vosp"        "prof"
## [25] "pr"           "pr1"         "plan"        "lartpc"
## [29] "larrouut"     "surf"        "infra"       "situ"
## [33] "vma"          "nb_indemne"  "nb_tue"      "nb_hospitalise"
## [37] "nb_blesse"    "nb_velo"     "nb_moto"     "nb_voiture"
## [41] "nb_utilitaire" "nb_camion"   "nb_tracteur" "nb_autre"
## [45] "nb_transport"  "Score_Severite"

# Histogramme du score de sévérité

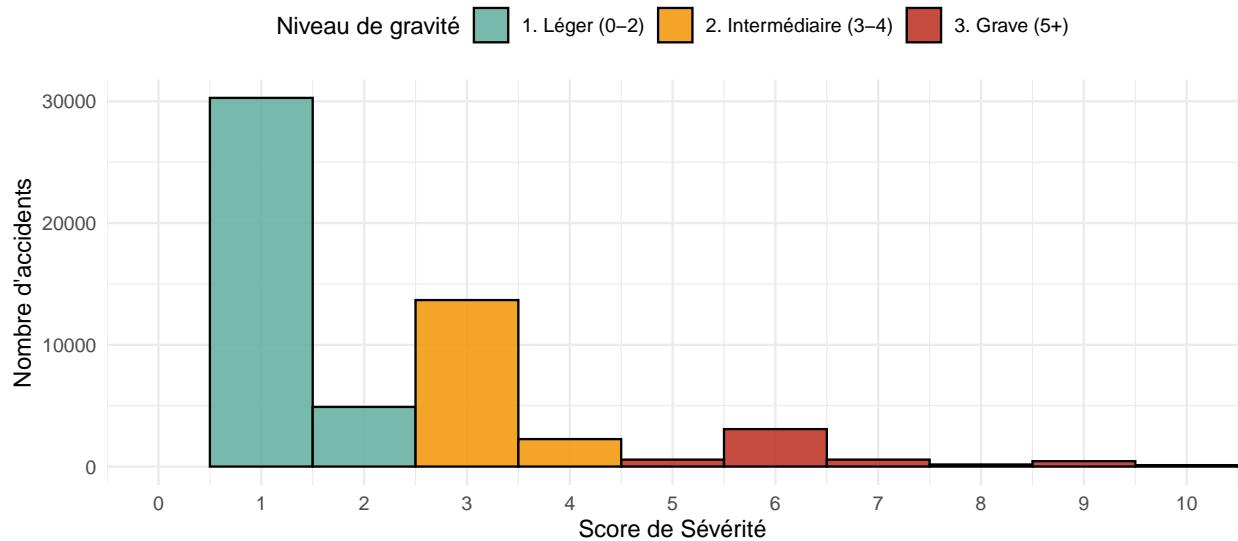
df %>%
  # Séparation du score en fonction de la gravité
  mutate(Tranche_Gravite = case_when(
    Score_Severite <= 2 ~ "1. Léger (0-2)",
    Score_Severite <= 4 ~ "2. Intermédiaire (3-4)",
    Score_Severite >= 5 ~ "3. Grave (5+)"
  )) %>%

ggplot(aes(x = Score_Severite)) +
  # On utilise la nouvelle colonne 'Tranche_Gravite' pour la couleur (fill)
  geom_bar(aes(fill = Tranche_Gravite),
            width = 1,
            color = "black",
            alpha = 0.9) +
  scale_fill_manual(values = c(
    "1. Léger (0-2)" = "#69b3a2",
    "2. Intermédiaire (3-4)" = "#F39C12",
    "3. Grave (5+)" = "#C0392B"
  )) +
  # Axe des x gradués de 1 par 1 jusqu'à 10
  scale_x_continuous(breaks = seq(0, 10, by = 1)) +
  coord_cartesian(xlim = c(0, 10)) +
  labs(title = "Distribution du Score de Sévérité",
       x = "Score de Sévérité",
       y = "Nombre d'accidents",
       fill = "Niveau de gravité") +
  theme_minimal() +
  theme(plot.title = element_text(size = 20, hjust = 0.5),

```

```
legend.position = "top")
```

Distribution du Score de Sévérité



```
library(ggplot2)
library(dplyr)
library(scales) # Pour le formatage des %
```

```
## Warning: package 'scales' was built under R version 4.3.3
```

```
# Calcul des %
df_pie <- df %>%
  # On recrée les catégories
  mutate(Tranche_Gravite = case_when(
    Score_Severite <= 2 ~ "1. Léger (0-2)",
    Score_Severite <= 4 ~ "2. Intermédiaire (3-4)",
    Score_Severite >= 5 ~ "3. Grave (5+)"
  )) %>%
  # On compte le nombre d'accidents par catégorie
  count(Tranche_Gravite) %>%
  # On calcule le pourcentage
  mutate(prop = n / sum(n),
        label_percent = percent(prop, accuracy = 0.1)) # Formatage "12.5%"

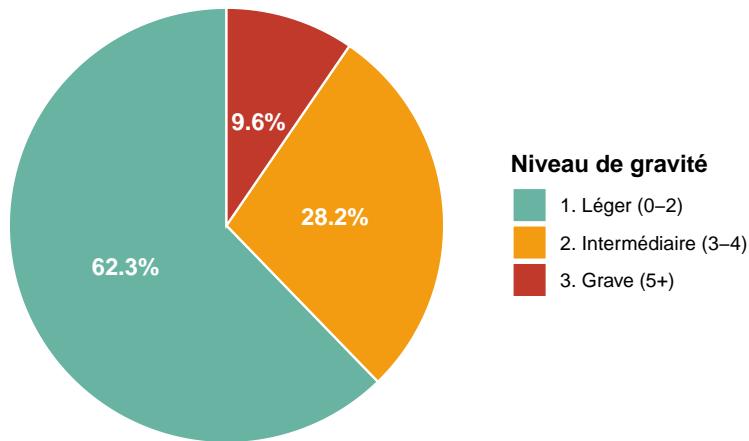
# CRÉATION DU CAMEMBERT
ggplot(df_pie, aes(x = "", y = prop, fill = Tranche_Gravite)) +
  # Création des parts (barre unique empilée)
  geom_bar(stat = "identity", width = 2, color = "white") +
  # Transformation en cercle
  coord_polar("y", start = 0) +
  # Ajout du texte (Pourcentages) au milieu des parts
```

```

geom_text(aes(label = label_percent),
          position = position_stack(vjust = 0.5),
          color = "white", fontface = "bold", size = 4) +
scale_fill_manual(values = c(
  "1. Léger (0-2)" = "#69b3a2",
  "2. Intermédiaire (3-4)" = "#F39C12",
  "3. Grave (5+)" = "#C0392B"
)) +
# Titres
labs(title = "Répartition des accidents par Gravité",
     fill = "Niveau de gravité") + # Titre de la légende
# Thème vide (indispensable pour un beau pie chart)
theme_void() +
theme(
  plot.title = element_text(size = 18, face = "bold", hjust = 0.5, margin = margin(b = 20)),
  legend.position = "right",
  legend.title = element_text(face = "bold")
)

```

Répartition des accidents par Gravité



```
summary(df$Score_Severite)
```

```

##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##    1.000   1.000   1.000   2.271   3.000  34.000

```

```
str(df)
```

```
## 'data.frame': 56518 obs. of 46 variables:
```

```

## $ Num_Acc      : chr  "N202100000001" "N202100000002" "N202100000003" "N202100000004" ...
## $ jour         : int  30 25 15 27 25 23 13 21 29 16 ...
## $ mois         : int  11 9 7 3 2 11 12 11 12 10 ...
## $ an           : int  2021 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
## $ hrmn         : chr  "07:32" "14:20" "07:55" "19:45" ...
## $ lum           : chr  "crepuscule ou aube" "plein jour" "plein jour" "nuit avec eclairage allume"
## $ dep           : chr  "30" "51" "85" "93" ...
## $ com           : chr  "30319" "51544" "85048" "93005" ...
## $ agg           : chr  "hors agglo" "hors agglo" "en agglo" "en agglo" ...
## $ int           : chr  "aucune" "en T" "aucune" "en X" ...
## $ atm           : chr  "normale" "normale" "ebiouissant" "pluie forte" ...
## $ col           : chr  "deux veh - frontal" "deux veh - cote" "autre" "autre" ...
## $ adr           : chr  "CD 981" "Aire de repos croisement D20E9 après sortie D977" "15 rue François ...
## $ lat           : num  44 49.2 46.9 48.9 49.4 ...
## $ long          : num  4.348 4.555 -0.964 2.52 1.146 ...
## $ date          : chr  "2021-11-30" "2021-09-25" "2021-07-15" "2021-03-27" ...
## $ catr          : chr  "departementale" "departementale" "communale" "departementale" ...
## $ voie          : chr  "981" "20" "GENERAL GALLIENI (BOULEVARD DU)" "GENERAL GALLIENI (BOULEVARD DU ...
## $ v1            : int  -1 0 0 0 -1 -1 0 0 0 0 ...
## $ v2            : chr  "N/A" "E" "N/A" "N/A" ...
## $ circ          : chr  "bidirectionnel" "bidirectionnel" "bidirectionnel" "bidirectionnel" ...
## $ nbv           : int  2 2 2 4 2 2 0 2 3 4 ...
## $ vosp          : chr  "sans objet" "sans objet" "sans objet" "sans objet" ...
## $ prof          : chr  "plat" "plat" "plat" "plat" ...
## $ pr             : int  NA 0 NA 0 11 NA 0 5 0 NA ...
## $ pr1            : int  NA 10 NA 1 150 NA 1 500 0 NA ...
## $ plan          : chr  "rectiligne" "rectiligne" "rectiligne" "rectiligne" ...
## $ lartpc         : num  NA NA NA NA NA NA NA NA NA ...
## $ larrouut       : num  NA NA NA NA NA 4 NA NA NA NA ...
## $ surf           : chr  "normale" "normale" "normale" "normale" ...
## $ infra          : chr  "aucun" "aucun" "aucun" "aucun" ...
## $ situ           : chr  "chaussee" "chaussee" "chaussee" "chaussee" ...
## $ vma            : int  80 80 50 50 50 50 50 40 50 50 ...
## $ nb_indemne     : int  1 0 1 1 0 1 1 1 2 1 ...
## $ nb_tue         : int  0 0 0 1 0 0 0 0 0 0 ...
## $ nb_hospitalise: int  1 1 1 0 0 0 0 1 0 0 ...
## $ nb_blesse      : int  0 1 0 0 1 1 2 1 1 4 ...
## $ nb_velo        : int  1 0 0 0 0 1 0 0 0 0 ...
## $ nb_moto        : int  0 0 0 0 0 0 0 1 0 0 ...
## $ nb_voiture     : int  1 2 1 1 2 1 2 1 2 0 ...
## $ nb_utilitaire : int  0 0 0 0 0 0 0 0 0 0 ...
## $ nb_camion      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ nb_tracteur    : int  0 0 0 0 0 0 0 0 0 0 ...
## $ nb_autre        : int  0 0 0 1 0 0 0 0 0 0 ...
## $ nb_transport   : int  0 0 0 0 0 0 0 0 0 1 ...
## $ Score_Severite: num  3 4 3 6 1 1 2 4 1 4 ...

```

On observe qu'un certain nombre de variables ne nous sera pas utile pour notre modèle de régression pour différentes raisons. En effet, les variables qui servent d'identifiants, celles ayant trop de catégories ou encore les variables avec une grande proportion de valeurs manquantes. Pour les variables ayant trop de modalités, il est préférable de regrouper les différentes modalités car juste supprimer ces variables pourrait faire perdre beaucoup d'informations pour notre modèle.

```

# On crée donc une fonction permettant de regrouper les départements (+100) en régions (13 métropolitaines)

get_region <- function(dep_code) {
  # On s'assure que l'entrée est un chr
  d <- as.character(dep_code)

  case_when(
    # Gestion de la Corse (2A et 2B)
    d %in% c("2A", "2B", "20") ~ "Corse",

    # Gestion des DOM (codes commençant par 97)
    substr(d, 1, 2) == "97" ~ "Outre-Mer",

    # Départements standards (on gère les cas "1" et "01")
    # Auvergne-Rhône-Alpes
    d %in% c("01", "1", "03", "3", "07", "7", "15", "15", "26", "38", "42", "43", "63", "69", "73", "74") ~ "Auvergne-Rhône-Alpes",

    # Hauts-de-France
    d %in% c("02", "2", "59", "60", "62", "80") ~ "Hauts-de-France",

    # PACA
    d %in% c("04", "4", "05", "5", "06", "6", "13", "83", "84") ~ "PACA",

    # Grand Est
    d %in% c("08", "8", "10", "51", "52", "54", "55", "57", "67", "68", "88") ~ "Grand Est",

    # Occitanie
    d %in% c("09", "9", "11", "12", "30", "31", "32", "34", "46", "48", "65", "66", "81", "82") ~ "Occitanie",

    # Normandie
    d %in% c("14", "27", "50", "61", "76") ~ "Normandie",

    # Nouvelle-Aquitaine
    d %in% c("16", "17", "19", "23", "24", "33", "40", "47", "64", "79", "86", "87") ~ "Nouvelle-Aquitaine",

    # Centre-Val de Loire
    d %in% c("18", "28", "36", "37", "41", "45") ~ "Centre-Val de Loire",

    # Bourgogne-Franche-Comté
    d %in% c("21", "25", "39", "58", "70", "71", "89", "90") ~ "Bourgogne-Franche-Comté",

    # Bretagne
    d %in% c("22", "29", "35", "56") ~ "Bretagne",

    # Pays de la Loire
    d %in% c("44", "49", "53", "72", "85") ~ "Pays de la Loire",

    # Île-de-France
    d %in% c("75", "77", "78", "91", "92", "93", "94", "95") ~ "Île-de-France"
  )
}

```

```

df <- df %>%

# On transforme l'heure en un numeric compris entre 0 et 23
# en extrayant la première partie de hrmn (heure:min)
mutate(
  Heure_Num= as.numeric(str_sub(hrmn, 1, 2))
) %>%

# On crée une variable Moment pour séparer les moments de la journée
mutate(
  Moment = case_when(
    Heure_Num >= 0 & Heure_Num <= 5 ~ "Nuit Profonde",
    Heure_Num >= 6 & Heure_Num <= 10 ~ "Matin (Pointe)",
    Heure_Num >= 11 & Heure_Num <= 16 ~ "Journee (Creuse)",
    Heure_Num >= 17 & Heure_Num <= 21 ~ "Soir (Pointe)",
    Heure_Num >= 22 & Heure_Num <= 23 ~ "Soiree",
    TRUE ~ "Autre" # Sécurité
  ),
  # On transforme en factor et on choisit "Journée" comme référence
  Moment = factor(Moment, levels = c("Journee (Creuse)", "Matin (Pointe)", "Soir (Pointe)", "Soiree"))

# On crée la colonne région
mutate(
  Region = get_region(dep),
  Region = as.factor(Region)
) %>%

# On garde que les variables intéressantes i.e
# en enlevant les identifiants (Num_Acc, adr), les variables quasi vides (pr, lartpc..), ou avec trop
select(
  Score_Severite,
  Region,
  Moment,
  lum, agg, int, atm, col,
  catr, circ, prof, plan, surf, infra, situ,
  vma,
  nb_velo, nb_moto, nb_voiture, nb_utilitaire, nb_camion, nb_transport
) %>%

# On convertit les variables chr en factor pour que r les traitent comme des variables qualitatives
mutate_if(is.character, as.factor) %>%

# On supprime les lignes avec des valeurs manquantes
na.omit

```

```
str(df)
```

```

## 'data.frame': 52250 obs. of 22 variables:
## $ Score_Severite: num 3 4 3 6 1 1 2 1 4 1 ...
## $ Region        : Factor w/ 14 levels "Auvergne-Rhône-Alpes",...: 11 6 14 8 9 6 1 8 8 8 ...
## $ Moment         : Factor w/ 5 levels "Journee (Creuse)",...: 2 1 2 3 2 1 3 1 2 1 ...
## $ lum            : Factor w/ 5 levels "crepuscule ou aube",...: 1 5 5 2 2 5 2 1 5 5 ...

```

```

## $ agg : Factor w/ 2 levels "en agglo","hors agglo": 2 2 1 1 1 1 1 1 1 ...
## $ int : Factor w/ 9 levels "4+ branches",...: 2 4 2 5 2 4 3 5 2 5 ...
## $ atm : Factor w/ 9 levels "autre","brouillard ou fumee",...: 6 6 4 7 6 6 6 8 6 6 ...
## $ col : Factor w/ 7 levels "3+ veh - en chaine",...: 6 5 3 3 4 4 6 5 5 5 ...
## $ catr : Factor w/ 8 levels "autoroute","autre",...: 4 4 3 4 6 3 3 4 7 4 ...
## $ circ : Factor w/ 4 levels "avec voies d'affectation variable",...: 2 2 2 2 2 2 2 2 4 2 ...
## $ prof : Factor w/ 4 levels "base de cote",...: 3 3 3 3 3 3 3 3 3 ...
## $ plan : Factor w/ 4 levels "en courbe a droite",...: 4 4 4 4 4 4 4 4 4 1 ...
## $ surf : Factor w/ 9 levels "autre","boue",...: 8 8 8 8 8 8 7 8 8 ...
## $ infra : Factor w/ 10 levels "aucun","autres",...: 1 1 1 1 1 1 4 1 3 2 ...
## $ situ : Factor w/ 7 levels "accotement","autres",...: 4 4 4 4 4 4 4 7 4 ...
## $ vma : int 80 80 50 50 50 50 50 50 50 ...
## $ nb_velo : int 1 0 0 0 0 1 0 0 0 ...
## $ nb_moto : int 0 0 0 0 0 0 0 0 0 ...
## $ nb_voiture : int 1 2 1 1 2 1 2 2 0 1 ...
## $ nb_utilitaire : int 0 0 0 0 0 0 0 0 0 ...
## $ nb_camion : int 0 0 0 0 0 0 0 0 0 ...
## $ nb_transport : int 0 0 0 0 0 0 0 0 1 0 ...
## - attr(*, "na.action")= 'omit' Named int [1:4268] 8 22 31 40 46 65 77 82 86 89 ...
## ..- attr(*, "names")= chr [1:4268] "8" "22" "31" "40" ...

```

Après ce nettoyage, on a perdu seulement environ 4 000 lignes, ce qui est négligeable puisqu'il nous reste 52 000 lignes et nous avons gagné en qualité.

R crée des levels pour les variables factor afin de comparer ces levels par rapport à un level de référence. Mais initialement R choisit le level de référence par ordre alphabétique, ce qui n'est pas forcément ce que l'on souhaite.

Par exemple, R a choisi la référence "aube ou crépuscule" alors qu'on voudrait mieux comparer par rapport aux autres niveaux.

```

## [1] "crepuscule ou aube"           "nuit avec eclairage allume"
## [3] "nuit avec eclairage non allume" "nuit sans eclairage"
## [5] "plein jour"

```

On affecte le level de référence que l'on souhaite

```

df$lum <- relevel(df$lum, ref = "plein jour", silent=TRUE)
df$atm <- relevel(df$atm, ref = "normale", silent=TRUE)
df$prof <- relevel(df$prof, ref = "plat", silent=TRUE)
df$plan <- relevel(df$plan, ref = "rectiligne", silent=TRUE)
df$surf <- relevel(df$surf, ref = "normale", silent=TRUE)
df$int <- relevel(df$int, ref = "aucune", silent=TRUE)
df$catr <- relevel(df$catr, ref = "departementale", silent=TRUE)
df$situ <- relevel(df$situ, ref = "chaussee", silent=TRUE)
df$col <- relevel(df$col, ref = "sans", silent=TRUE)

```

Vérifions

```

cat("Référence pour LUM :", levels(df$lum)[1], "\n")

```

```

## Référence pour LUM : plein jour

```

```

cat("Référence pour ATM : ", levels(df$atm)[1], "\n")

## Référence pour ATM : normale

cat("Référence pour PROF : ", levels(df$prof)[1], "\n")

## Référence pour PROF : plat

cat("Référence pour PLAN : ", levels(df$plan)[1], "\n")

## Référence pour PLAN : rectiligne

cat("Référence pour SURF : ", levels(df$surf)[1], "\n")

## Référence pour SURF : normale

cat("Référence pour INT : ", levels(df$int)[1], "\n")

## Référence pour INT : aucune

cat("Référence pour CATR : ", levels(df$catr)[1], "\n")

## Référence pour CATR : departementale

cat("Référence pour SITU : ", levels(df$situ)[1], "\n")

```

Ainsi, lorsque l'on va calculer la régression, Intercept correspondra au score de sévérité moyen quand tout est à la référence (Plein jour, Météo normale, route plate, etc.)

Modèle de Régression Linéaire

Modèle 1

```

# Modèle de régression: Score de sévérité expliqué par toutes les colonnes
modele <- lm(Score_Severite ~ ., data = df)

summary(modele)

## 
## Call:
## lm(formula = Score_Severite ~ ., data = df)
## 
## Residuals:

```

```

##      Min     1Q Median     3Q    Max
## -5.021 -0.973 -0.360  0.551 32.291
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.5119133  0.1083028 13.960 < 2e-16 ***
## RegionBourgogne-Franche-Comté 0.4320650  0.0515641  8.379 < 2e-16 ***
## RegionBretagne              -0.1475422  0.0449015 -3.286 0.001017 **
## RegionCentre-Val de Loire   0.1687585  0.0490635  3.440 0.000583 ***
## RegionCorse                  -0.2139013  0.0860399 -2.486 0.012919 *
## RegionGrand Est              -0.0438540  0.0400306 -1.096 0.273298
## RegionHauts-de-France       0.2788928  0.0430864  6.473 9.70e-11 ***
## RegionÎle-de-France         -0.5711555  0.0286188 -19.957 < 2e-16 ***
## RegionNormandie              -0.1198904  0.0448609 -2.672 0.007531 **
## RegionNouvelle-Aquitaine    0.0912208  0.0359579  2.537 0.011187 *
## RegionOccitanie              0.3713846  0.0384095  9.669 < 2e-16 ***
## RegionOutre-Mer              -0.0202551  0.0451463 -0.449 0.653683
## RegionPACA                   0.1361936  0.0352685  3.862 0.000113 ***
## RegionPays de la Loire      0.1916369  0.0470344  4.074 4.62e-05 ***
## MomentMatin (Pointe)        -0.1132069  0.0222258 -5.093 3.53e-07 ***
## MomentSoir (Pointe)         -0.0136886  0.0218733 -0.626 0.531440
## MomentSoiree                 0.2996621  0.0475685  6.300 3.01e-10 ***
## MomentNuit Profonde         0.4473030  0.0414680 10.787 < 2e-16 ***
## lumcrepuscule ou aube       0.0513893  0.0330477  1.555 0.119951
## lumnuit avec eclairage allume 0.0063892  0.0304127  0.210 0.833602
## lumnuit avec eclairage non allume 0.0922829  0.0897972  1.028 0.304104
## lumnuit sans eclairage       0.2837140  0.0359681  7.888 3.13e-15 ***
## agghors agglo                 0.7029752  0.0254128 27.662 < 2e-16 ***
## int4+ branches                 0.3104144  0.1031571  3.009 0.002621 **
## intautre                      -0.0537326  0.0404873 -1.327 0.184466
## inten T                        -0.0524744  0.0275128 -1.907 0.056491 .
## inten X                        0.1100513  0.0271510  4.053 5.06e-05 ***
## inten Y                        0.1222836  0.0529677  2.309 0.020967 *
## intgiratoire                  -0.1917792  0.0441477 -4.344 1.40e-05 ***
## intpassage a niveau           0.3902855  0.2502157  1.560 0.118814
## intplace                       0.0007437  0.0783294  0.009 0.992424
## atmautre                      0.1634957  0.1163053  1.406 0.159806
## atmbrouillard ou fumee       -0.0327726  0.0938602 -0.349 0.726967
## atmcouvert                     0.0001427  0.0427994  0.003 0.997340
## atmefblouissant               0.1829020  0.0599956  3.049 0.002300 **
## atmneige ou grele             0.3544108  0.1644537  2.155 0.031160 *
## atmpluie forte                 -0.1260579  0.0665786 -1.893 0.058314 .
## atmpluie legere                -0.1604635  0.0430115 -3.731 0.000191 ***
## atmvent fort ou tempete       0.2415097  0.1692960  1.427 0.153715
## col3+ veh - en chaine        -0.1107675  0.0668292 -1.657 0.097430 .
## col3+ veh - multiple          0.6540140  0.0679264  9.628 < 2e-16 ***
## colautre                      0.2911522  0.0302296  9.631 < 2e-16 ***
## coldeux veh - arriere         -0.2863944  0.0391961 -7.307 2.78e-13 ***
## coldeux veh - cote            -0.0667604  0.0354050 -1.886 0.059352 .
## coldeux veh - frontal         0.7752125  0.0400147 19.373 < 2e-16 ***
## catrautoroute                 -0.7411083  0.0386307 -19.184 < 2e-16 ***
## catrautre                      -0.3179569  0.1151848 -2.760 0.005775 **
## catrcommunale                 -0.4114663  0.0205995 -19.975 < 2e-16 ***
## catrhors reseau                 -0.2762433  0.3182679 -0.868 0.385421

```

```

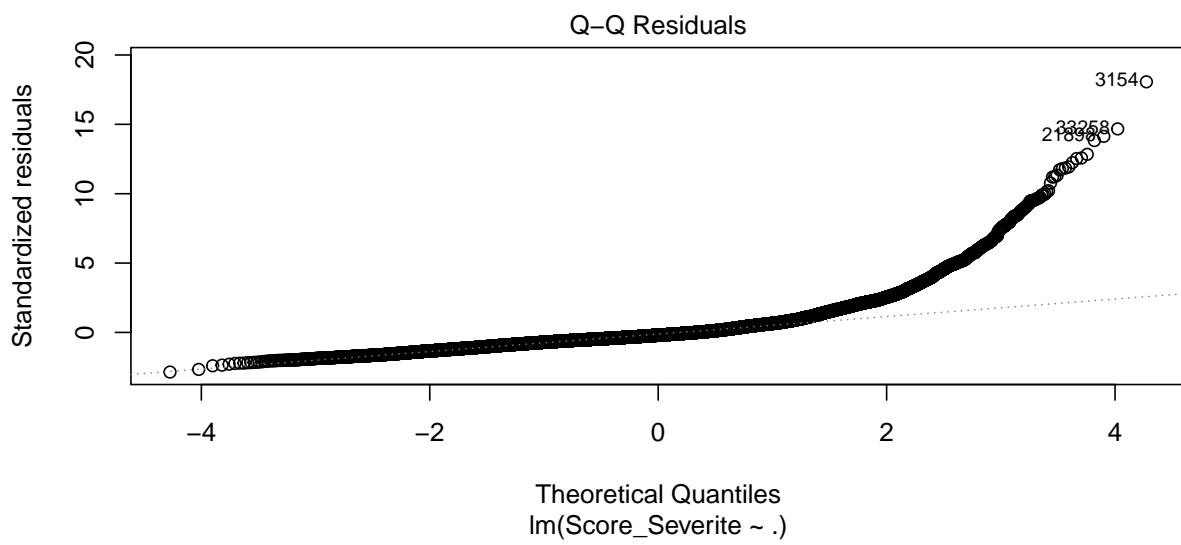
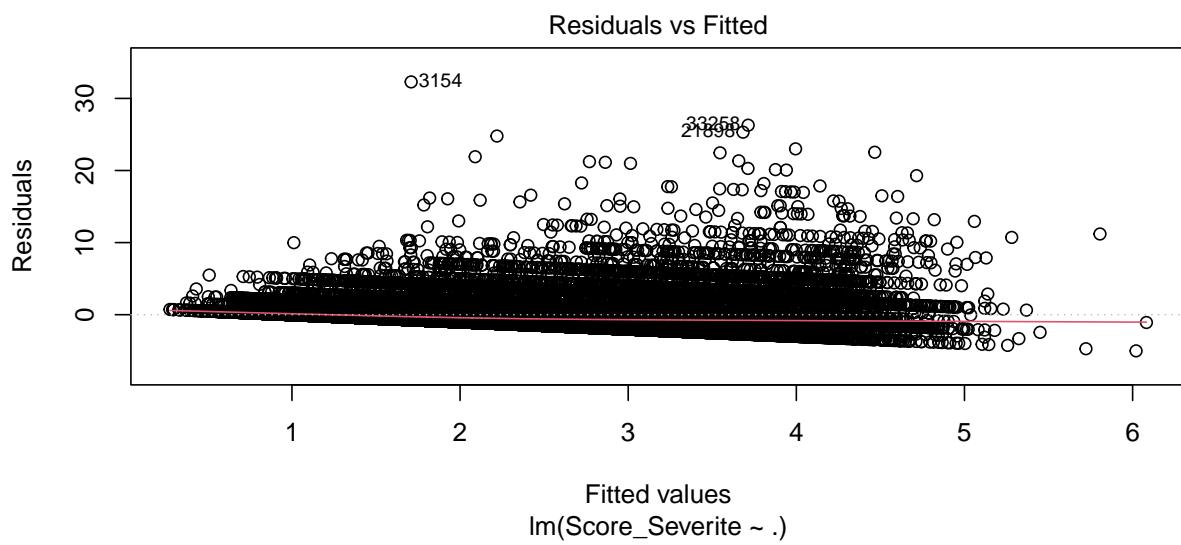
## catrmetropole          -0.4223310  0.0394870 -10.695 < 2e-16 ***
## catrnationale         -0.1676881  0.0378138 -4.435 9.24e-06 ***
## catrparc              -0.1441515  0.4125752 -0.349 0.726795
## circbidirectionnel    0.2007263  0.0961451  2.088 0.036826 *
## circchaussee separees -0.0085460  0.0985431 -0.087 0.930891
## circsens unique       -0.0386818  0.0973214 -0.397 0.691027
## profbase de cote      0.1191376  0.0669393  1.780 0.075117 .
## profpente              0.0284011  0.0221235  1.284 0.199233
## profsommets de cote   0.1734582  0.0619648  2.799 0.005123 **
## planen courbe a droite 0.0238603  0.0293972  0.812 0.416996
## planen courbe a gauche -0.0423865  0.0285490 -1.485 0.137632
## planen S               0.0151489  0.0702303  0.216 0.829220
## surfautre              -0.0501679  0.1141035 -0.440 0.660178
## surfboue                -0.1389618  0.2849530 -0.488 0.625789
## surfcorps gras        -0.2603784  0.2138137 -1.218 0.223312
## surfenneigee            -0.5792076  0.2076905 -2.789 0.005292 **
## surfflaques             -0.0667773  0.2198009 -0.304 0.761275
## surfinondee             0.1916419  0.4057979  0.472 0.636744
## surfmouillee            0.0450494  0.0370246  1.217 0.223708
## surfverglacee           -0.1871131  0.1198388 -1.561 0.118442
## infraautres             0.0770053  0.0453781  1.697 0.089708 .
## infrabretelle ou échangeur -0.0717755  0.0708996 -1.012 0.311374
## infracarrefour          0.0731625  0.0373316  1.960 0.050024 .
## infrachantier           0.0600194  0.0947537  0.633 0.526459
## infrapeage               0.2007572  0.3350390  0.599 0.549039
## infrapont ou autopont   0.1506699  0.0622806  2.419 0.015558 *
## infrasouterrain ou tunnel -0.2503010  0.0757841 -3.303 0.000958 ***
## infravoie ferree         -0.0044992  0.2090870 -0.022 0.982832
## infrazone pietonne       -0.1100168  0.0903706 -1.217 0.223459
## situaccotement          0.4344503  0.0353324 12.296 < 2e-16 ***
## situautres               0.4015617  0.0492058  8.161 3.40e-16 ***
## situbande arret urgence  0.2957791  0.1014175  2.916 0.003542 **
## situpiste cyclable       -0.0945107  0.0590388 -1.601 0.109422
## situtrottoir              0.4870386  0.0548223  8.884 < 2e-16 ***
## situvoie speciale        -0.0085448  0.0768373 -0.111 0.911453
## vma                      0.0037328  0.0004358  8.565 < 2e-16 ***
## nb.velo                  0.1016046  0.0296365  3.428 0.000608 ***
## nb.moto                  0.1574273  0.0232936  6.758 1.41e-11 ***
## nb.voiture                0.1999954  0.0196105 10.198 < 2e-16 ***
## nb.utilitaire             0.2204296  0.0287496  7.667 1.79e-14 ***
## nb.camion                 0.4709500  0.0430879 10.930 < 2e-16 ***
## nb.transport              0.4110409  0.0674353  6.095 1.10e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.793 on 52159 degrees of freedom
## Multiple R-squared:  0.1946, Adjusted R-squared:  0.1932
## F-statistic: 140 on 90 and 52159 DF,  p-value: < 2.2e-16

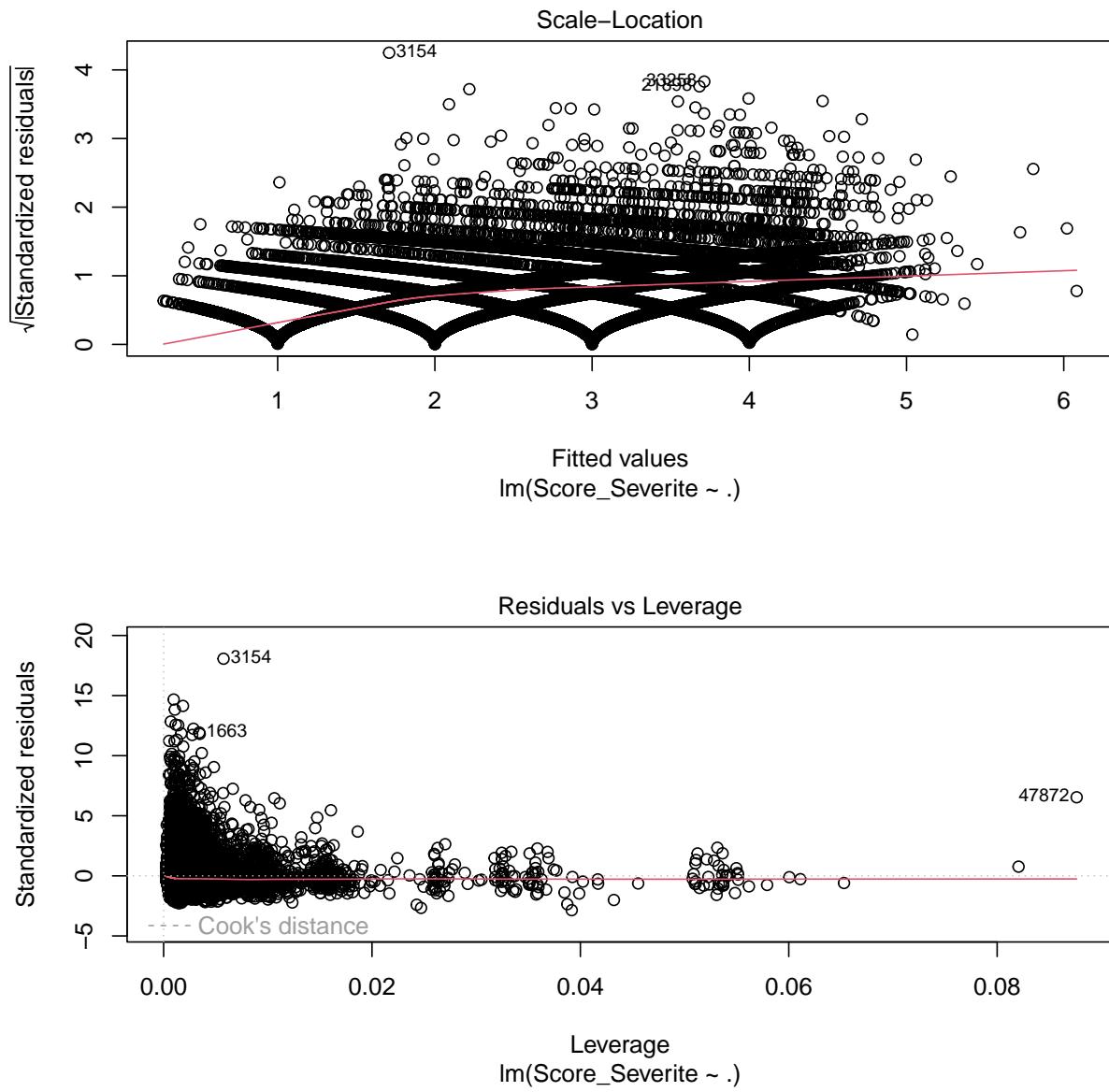
```

```

# On affiche les 4 graphiques de diagnostic
plot(modele)

```

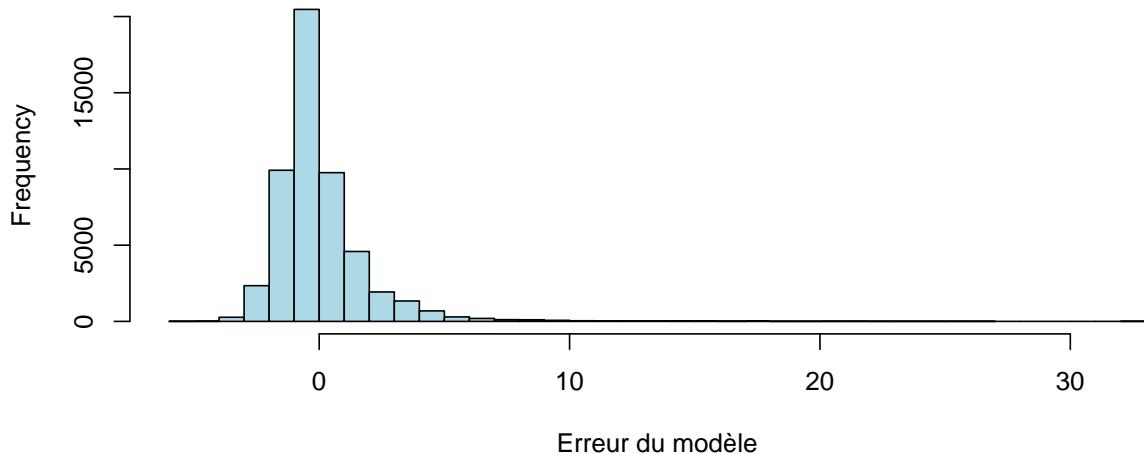




- 1) Les résidus sont globalement centré autour de 0 mais on observe une augmentation de la dispersion des résidus pour les valeurs élevées.
- 2) Les points suivent approximativement la droite normale théorique mais pour les valeurs élevées, les points s'écartent de la droite. Cela indique que les résidus ne suivent pas une distribution normale.
- 3) La variabilité des résidus augmente avec les valeurs prédictes par le modèle (courbe rouge croissante)
- 4) On identifie un point extrême tout à droite

```
# Distribution des résidus
hist(resid(modele),
      breaks = 50,
      col = "lightblue",
      main = "Distribution des Résidus",
      xlab = "Erreur du modèle")
```

Distribution des Résidus



On peut observer un pic autour de 0 et une asymétrie positive, qui confirme visuellement la non-normalité des résidus.

Modèle 2

On a vu que dans la distribution du score de sévérité, il y avait beaucoup d'accidents légers et quelques accidents très graves. Appliquer le log à ce score permet de limiter les scores extrêmes et de rapprocher la distribution empirique d'une loi Normale.

```
# Modèle de régression avec le log: Score de sévérité expliqué par toutes les colonnes
modele2 <- lm(log(Score_Severite+1) ~ ., data = df) #On ajoute le +1 si score =0
```

```
summary(modele2)
```

```
##
## Call:
## lm(formula = log(Score_Severite + 1) ~ ., data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1.20404 -0.27877 -0.08426  0.26017  2.74352 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                9.214e-01  2.426e-02 37.986 < 2e-16 ***
## RegionBourgogne-Franche-Comté 9.870e-02  1.155e-02  8.546 < 2e-16 ***
## RegionBretagne              -3.941e-02  1.006e-02 -3.919 8.92e-05 ***
## RegionCentre-Val de Loire   3.584e-02  1.099e-02  3.262 0.001108 ** 
## RegionCorse                 -4.335e-02  1.927e-02 -2.249 0.024498 *  
## RegionGrand Est             -1.460e-02  8.966e-03 -1.629 0.103374  
## RegionHauts-de-France       5.582e-02  9.650e-03  5.784 7.34e-09 *** 
## RegionÎle-de-France        -1.748e-01  6.410e-03 -27.275 < 2e-16 ***
## RegionNormandie            -3.357e-02  1.005e-02 -3.341 0.000835 ***
```

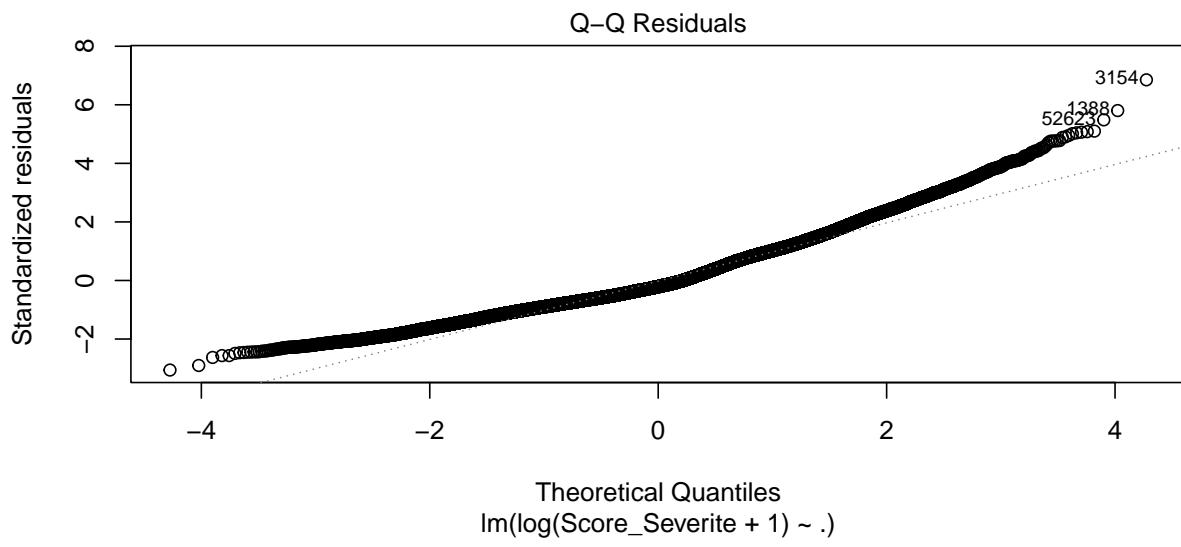
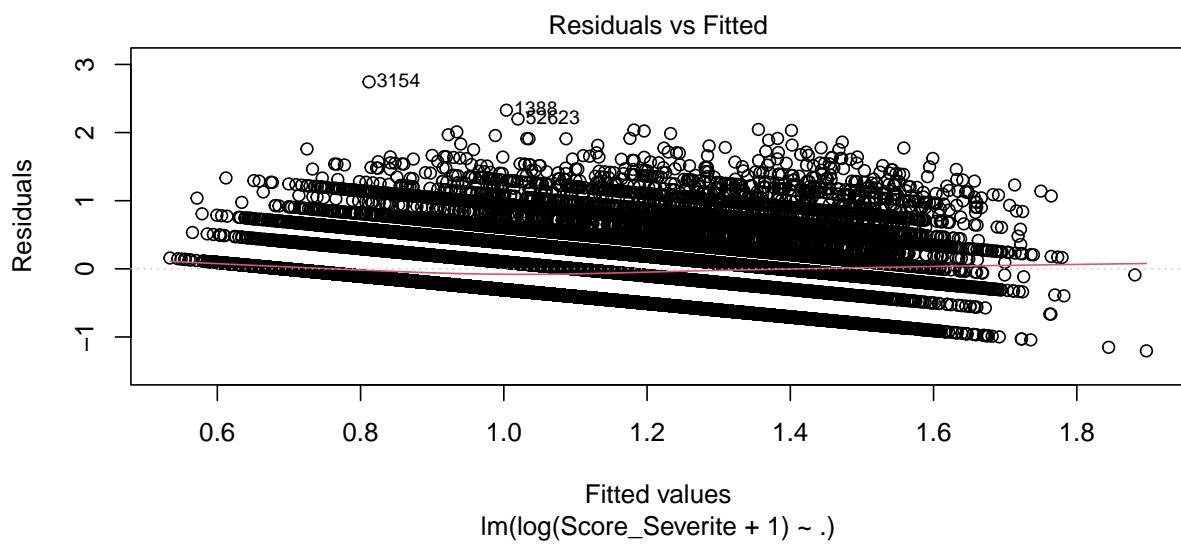
## RegionNouvelle-Aquitaine	1.801e-02	8.054e-03	2.236	0.025326	*
## RegionOccitanie	9.316e-02	8.603e-03	10.829	< 2e-16	***
## RegionOutre-Mer	-1.753e-02	1.011e-02	-1.734	0.082965	.
## RegionPACA	3.426e-02	7.899e-03	4.337	1.45e-05	***
## RegionPays de la Loire	4.974e-02	1.053e-02	4.721	2.35e-06	***
## MomentMatin (Pointe)	-3.099e-02	4.978e-03	-6.225	4.86e-10	***
## MomentSoir (Pointe)	-4.756e-03	4.899e-03	-0.971	0.331627	
## MomentSoiree	5.961e-02	1.065e-02	5.595	2.22e-08	***
## MomentNuit Profonde	9.572e-02	9.288e-03	10.306	< 2e-16	***
## lumcrepuscule ou aube	1.487e-02	7.402e-03	2.009	0.044502	*
## lumnuit avec eclairage allume	5.675e-03	6.812e-03	0.833	0.404793	
## lumnuit avec eclairage non allume	2.225e-02	2.011e-02	1.106	0.268630	
## lumnuit sans eclairage	5.747e-02	8.056e-03	7.134	9.86e-13	***
## agghors agglo	1.752e-01	5.692e-03	30.789	< 2e-16	***
## int4+ branches	4.892e-02	2.311e-02	2.117	0.034256	*
## intautre	-7.572e-03	9.068e-03	-0.835	0.403714	
## inten T	-6.085e-03	6.162e-03	-0.987	0.323440	
## inten X	3.252e-02	6.081e-03	5.347	8.96e-08	***
## inten Y	3.927e-02	1.186e-02	3.310	0.000934	***
## intgiratoire	-5.023e-02	9.888e-03	-5.080	3.79e-07	***
## intpassage a niveau	9.724e-02	5.604e-02	1.735	0.082735	.
## intplace	-6.808e-03	1.754e-02	-0.388	0.697979	
## atmautre	4.964e-02	2.605e-02	1.905	0.056721	.
## atmbrouillard ou fumee	-1.681e-03	2.102e-02	-0.080	0.936258	
## atmcouvert	-3.486e-03	9.586e-03	-0.364	0.716122	
## atmeblouissant	6.046e-02	1.344e-02	4.499	6.84e-06	***
## atmneige ou grele	8.334e-02	3.683e-02	2.263	0.023670	*
## atmpluie forte	-2.412e-02	1.491e-02	-1.618	0.105741	
## atmpluie legere	-3.942e-02	9.634e-03	-4.092	4.28e-05	***
## atmvent fort ou tempete	6.459e-02	3.792e-02	1.703	0.088503	.
## col3+ veh - en chaine	6.711e-03	1.497e-02	0.448	0.653902	
## col3+ veh - multiple	1.461e-01	1.521e-02	9.600	< 2e-16	***
## colautre	7.224e-02	6.771e-03	10.669	< 2e-16	***
## coldeux veh - arriere	-6.495e-02	8.779e-03	-7.398	1.40e-13	***
## coldeux veh - cote	-1.612e-02	7.930e-03	-2.032	0.042141	*
## coldeux veh - frontal	1.683e-01	8.962e-03	18.779	< 2e-16	***
## catrautoroute	-1.842e-01	8.652e-03	-21.292	< 2e-16	***
## catrautre	-6.704e-02	2.580e-02	-2.599	0.009363	**
## catrcommunale	-1.153e-01	4.614e-03	-24.994	< 2e-16	***
## catrhors reseau	-4.971e-02	7.129e-02	-0.697	0.485593	
## catrmetropole	-1.119e-01	8.844e-03	-12.652	< 2e-16	***
## catrnationale	-4.982e-02	8.469e-03	-5.882	4.07e-09	***
## catrparc	-1.116e-02	9.241e-02	-0.121	0.903855	
## circbidirectionnel	4.790e-02	2.153e-02	2.224	0.026121	*
## circchaussee separees	-7.739e-03	2.207e-02	-0.351	0.725882	
## circsens unique	-2.155e-02	2.180e-02	-0.989	0.322843	
## profbase de cote	2.599e-02	1.499e-02	1.733	0.083067	.
## profpente	1.146e-02	4.955e-03	2.312	0.020784	*
## profsommet de cote	3.615e-02	1.388e-02	2.604	0.009207	**
## planen courbe a droite	1.153e-02	6.584e-03	1.752	0.079821	.
## planen courbe a gauche	-7.855e-03	6.394e-03	-1.228	0.219305	
## planen S	-3.055e-03	1.573e-02	-0.194	0.845992	
## surfautre	-3.565e-03	2.556e-02	-0.139	0.889061	
## surfboue	-1.263e-02	6.382e-02	-0.198	0.843131	

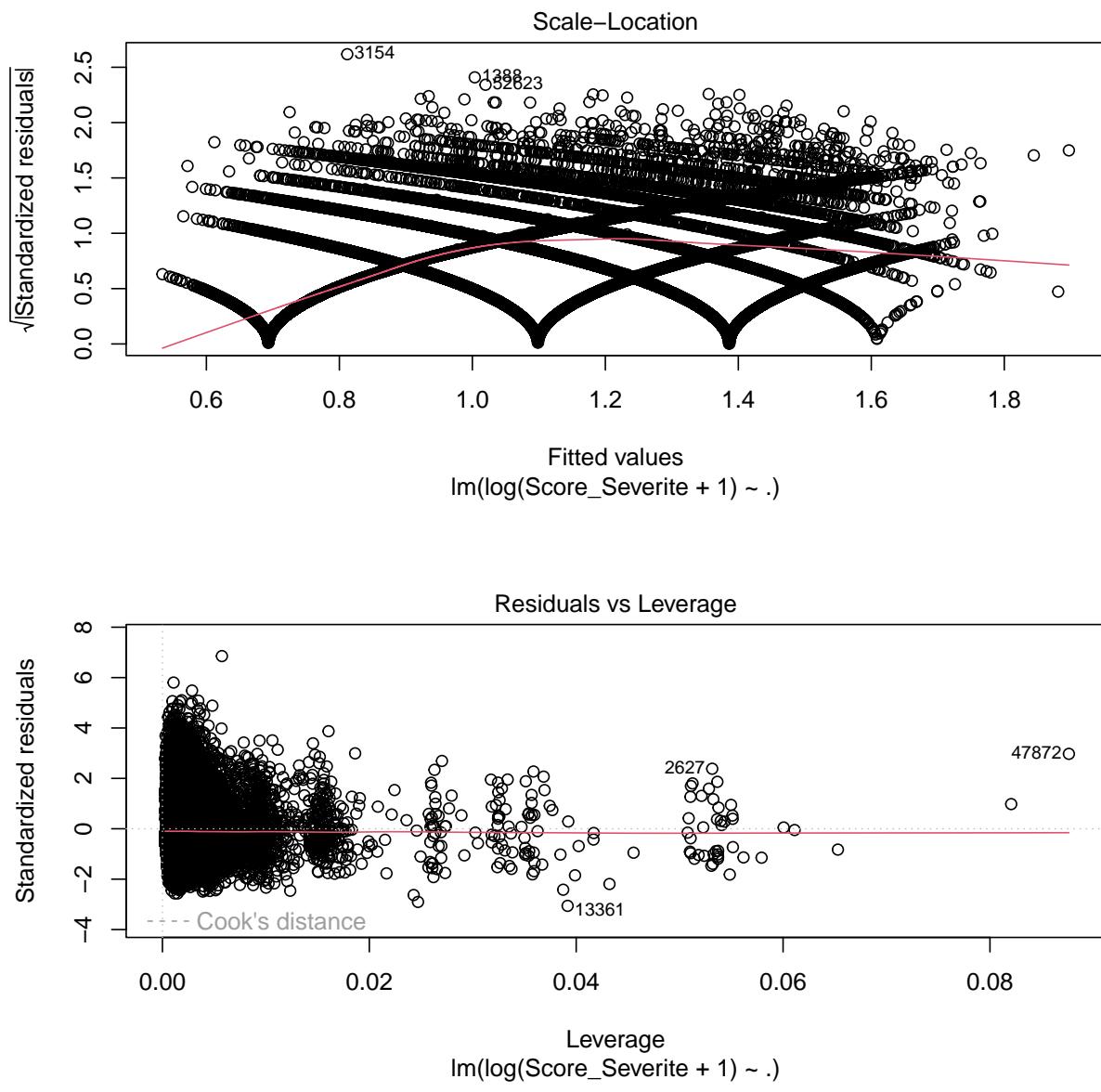
```

## surfcorps gras          -8.103e-02  4.789e-02 -1.692 0.090667 .
## surfenneigee           -1.429e-01  4.652e-02 -3.072 0.002126 **
## surfflaques            -6.111e-03  4.923e-02 -0.124 0.901215
## surfinondee            8.469e-02  9.089e-02  0.932 0.351429
## surfmouillee           5.322e-03  8.293e-03  0.642 0.521035
## surfverglacee          -3.769e-02  2.684e-02 -1.404 0.160315
## infraautres             2.529e-02  1.016e-02  2.488 0.012839 *
## infrabretelle ou échangeur -2.552e-02  1.588e-02 -1.607 0.108070
## infracarrefour          2.334e-02  8.361e-03  2.792 0.005245 **
## infrachantier            1.798e-02  2.122e-02  0.847 0.396782
## infrapeage                5.680e-02  7.504e-02  0.757 0.449124
## infrapont ou autopont      3.745e-02  1.395e-02  2.684 0.007267 **
## infrasouterrain ou tunnel -6.153e-02  1.697e-02 -3.625 0.000290 ***
## infravoie ferree          -5.653e-03  4.683e-02 -0.121 0.903917
## infrazone pietonne        -3.401e-02  2.024e-02 -1.680 0.092956 .
## situaccotement            1.129e-01  7.914e-03 14.261 < 2e-16 ***
## situautres                 9.724e-02  1.102e-02  8.823 < 2e-16 ***
## situbande arret urgence     7.700e-02  2.272e-02  3.390 0.000700 ***
## situpiste cyclable         -3.907e-02  1.322e-02 -2.954 0.003134 **
## situtrottoir                  1.165e-01  1.228e-02  9.484 < 2e-16 ***
## situvoie speciale           1.911e-04  1.721e-02  0.011 0.991142
## vma                         8.216e-04  9.761e-05  8.417 < 2e-16 ***
## nb_velo                      1.583e-02  6.638e-03  2.384 0.017122 *
## nb_moto                      4.509e-02  5.217e-03  8.642 < 2e-16 ***
## nb_voiture                   3.458e-02  4.392e-03  7.873 3.53e-15 ***
## nb_utilitaire                 4.355e-02  6.439e-03  6.763 1.37e-11 ***
## nb_camion                     9.727e-02  9.651e-03 10.079 < 2e-16 ***
## nb_transport                  9.122e-02  1.510e-02  6.040 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4015 on 52159 degrees of freedom
## Multiple R-squared:  0.2361, Adjusted R-squared:  0.2348
## F-statistic: 179.2 on 90 and 52159 DF,  p-value: < 2.2e-16

```

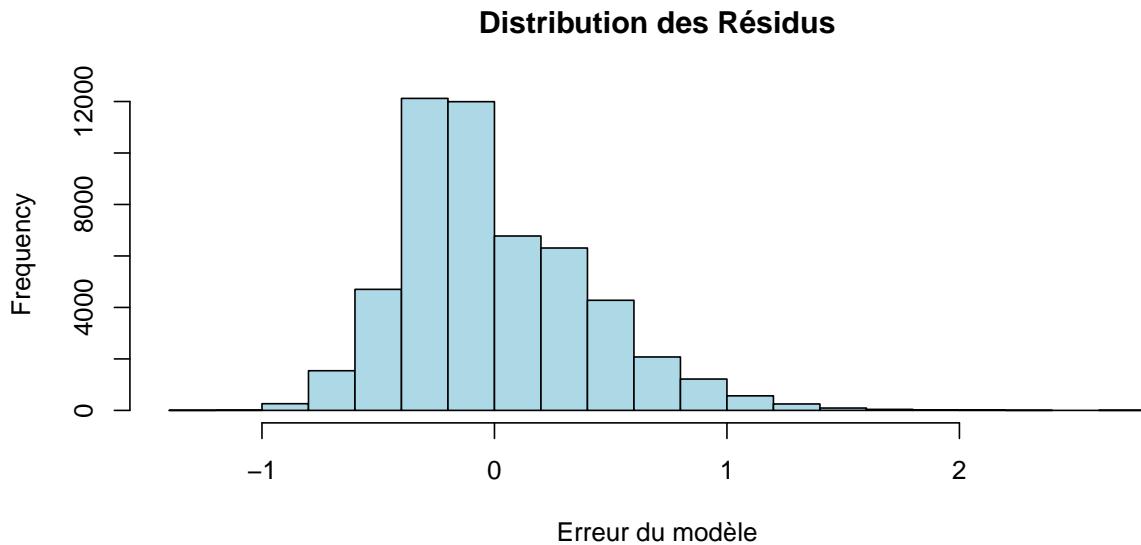
```
plot(modele2)
```





Les points suivent désormais presque parfaitement la droite normale. Cela signifie que l'hypothèse de normalité des résidus est respectée.

```
hist(resid(modele2),
  breaks = 20,
  col = "lightblue",
  main = "Distribution des Résidus",
  xlab = "Erreur du modèle")
```



Graphique un peu plus en forme de cloche qui fait penser à la loi Normale.

Modèle 3

On enlève les variables surf, prof et circ de notre modèle car ces variables ne sont statistiquement pas significatives au seuil 5% sur le score de sévérité (test de Student avec H0: coefficient nul).

```
modele3 <- lm(log(Score_Severite+1) ~ . - surf - prof - circ, data=df)
summary(modele3)
```

```
##
## Call:
## lm(formula = log(Score_Severite + 1) ~ . - surf - prof - circ,
##      data = df)
##
## Residuals:
##      Min        1Q        Median        3Q       Max
## -1.15610 -0.27942 -0.08536  0.26206  2.70565
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                9.631e-01  1.134e-02  84.894 < 2e-16 ***
## RegionBourgogne-Franche-Comté 1.040e-01  1.157e-02   8.990 < 2e-16 ***
## RegionBretagne              -3.675e-02  1.006e-02  -3.652 0.000260 ***
## RegionCentre-Val de Loire   3.931e-02  1.100e-02   3.574 0.000351 ***
## RegionCorse                  -3.155e-02  1.929e-02  -1.636 0.101876
## RegionGrand Est             -1.297e-02  8.979e-03  -1.445 0.148477
## RegionHauts-de-France       5.820e-02  9.654e-03   6.029 1.66e-09 ***
## RegionÎle-de-France         -1.849e-01  6.382e-03 -28.972 < 2e-16 ***
## RegionNormandie              -2.945e-02  1.006e-02  -2.927 0.003421 **
## RegionNouvelle-Aquitaine    2.158e-02  8.055e-03   2.679 0.007378 **
## RegionOccitanie              9.571e-02  8.614e-03  11.111 < 2e-16 ***
## RegionOutre-Mer             -1.197e-02  1.011e-02  -1.183 0.236782
```

## RegionPACA	3.423e-02	7.897e-03	4.335	1.46e-05	***
## RegionPays de la Loire	5.330e-02	1.054e-02	5.056	4.29e-07	***
## MomentMatin (Pointe)	-3.163e-02	4.980e-03	-6.351	2.16e-10	***
## MomentSoir (Pointe)	-3.938e-03	4.907e-03	-0.803	0.422229	
## MomentSoiree	5.834e-02	1.067e-02	5.468	4.58e-08	***
## MomentNuit Profonde	9.491e-02	9.303e-03	10.202	< 2e-16	***
## lumcrepuscule ou aube	1.518e-02	7.414e-03	2.048	0.040543	*
## lumnuit avec eclairage allume	3.911e-03	6.809e-03	0.574	0.565674	
## lumnuit avec eclairage non allume	2.065e-02	2.015e-02	1.025	0.305546	
## lumnuit sans eclairage	5.770e-02	8.057e-03	7.162	8.07e-13	***
## agghors agglo	1.806e-01	5.683e-03	31.783	< 2e-16	***
## int4+ branches	4.796e-02	2.315e-02	2.072	0.038282	*
## intautre	-1.129e-02	9.079e-03	-1.244	0.213515	
## inten T	-1.578e-03	6.165e-03	-0.256	0.797952	
## inten X	3.207e-02	6.087e-03	5.268	1.39e-07	***
## inten Y	3.811e-02	1.188e-02	3.208	0.001339	**
## intgiratoire	-6.721e-02	9.815e-03	-6.848	7.56e-12	***
## intpassage a niveau	1.043e-01	5.616e-02	1.857	0.063328	.
## intplace	-1.896e-02	1.755e-02	-1.080	0.280055	
## atmautre	3.854e-02	2.435e-02	1.583	0.113411	
## atmbroutilard ou fumee	1.085e-03	2.087e-02	0.052	0.958512	
## atmcouvert	-1.923e-03	9.062e-03	-0.212	0.831962	
## atmeflouissant	6.555e-02	1.346e-02	4.870	1.12e-06	***
## atmneige ou grele	5.452e-03	2.645e-02	0.206	0.836696	
## atmpluie forte	-1.669e-02	1.264e-02	-1.320	0.186825	
## atmpluie legere	-3.404e-02	5.842e-03	-5.827	5.69e-09	***
## atmvent fort ou tempete	6.717e-02	3.796e-02	1.770	0.076812	.
## col3+ veh - en chaine	6.972e-04	1.498e-02	0.047	0.962885	
## col3+ veh - multiple	1.473e-01	1.524e-02	9.669	< 2e-16	***
## colautre	7.177e-02	6.766e-03	10.607	< 2e-16	***
## coldeux veh - arriere	-7.065e-02	8.765e-03	-8.061	7.76e-16	***
## coldeux veh - cote	-1.597e-02	7.920e-03	-2.017	0.043731	*
## coldeux veh - frontal	1.763e-01	8.948e-03	19.698	< 2e-16	***
## catrautoroute	-2.229e-01	8.046e-03	-27.706	< 2e-16	***
## catrautre	-7.633e-02	2.583e-02	-2.955	0.003125	**
## catrcommunale	-1.229e-01	4.587e-03	-26.790	< 2e-16	***
## catrhors reseau	-5.593e-02	7.140e-02	-0.783	0.433488	
## catrmetropole	-1.216e-01	8.822e-03	-13.781	< 2e-16	***
## catrnationale	-7.200e-02	8.279e-03	-8.696	< 2e-16	***
## catrparc	-2.420e-02	9.260e-02	-0.261	0.793876	
## planen courbe a droite	1.332e-02	6.513e-03	2.046	0.040787	*
## planen courbe a gauche	-8.422e-03	6.327e-03	-1.331	0.183117	
## planen S	3.267e-03	1.568e-02	0.208	0.834969	
## infraautres	2.117e-02	1.018e-02	2.080	0.037523	*
## infrabretelle ou échangeur	-4.194e-02	1.582e-02	-2.650	0.008042	**
## infrafcarrefour	2.195e-02	8.366e-03	2.624	0.008690	**
## infrachantier	1.531e-02	2.122e-02	0.721	0.470643	
## infrapeage	4.731e-02	7.514e-02	0.630	0.528946	
## infrapont ou autopont	3.632e-02	1.396e-02	2.603	0.009255	**
## infrasouterrain ou tunnel	-7.258e-02	1.698e-02	-4.276	1.91e-05	***
## infravvoie ferree	-1.518e-02	4.692e-02	-0.324	0.746314	
## infrazone pietonne	-3.876e-02	2.028e-02	-1.911	0.056003	.
## situaccotement	1.188e-01	7.916e-03	15.011	< 2e-16	***
## situautres	9.827e-02	1.104e-02	8.899	< 2e-16	***

```

## situbande arret urgence      6.953e-02  2.275e-02  3.056 0.002241 ***
## situpiste cyclable        -4.322e-02  1.325e-02 -3.263 0.001105 **
## situtrottoir                 1.198e-01  1.230e-02  9.737 < 2e-16 ***
## situvoie speciale       -3.816e-03  1.724e-02 -0.221 0.824858
## vma                         7.456e-04  9.733e-05  7.660 1.89e-14 ***
## nb_velo                      1.709e-02  6.649e-03  2.570 0.010163 *
## nb_moto                       4.705e-02  5.225e-03  9.006 < 2e-16 ***
## nb_voiture                     3.461e-02  4.399e-03  7.869 3.65e-15 ***
## nb_utilitaire                  4.330e-02  6.450e-03  6.713 1.93e-11 ***
## nb_camion                      9.636e-02  9.669e-03  9.966 < 2e-16 ***
## nb_transport                    9.111e-02  1.513e-02  6.020 1.76e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4024 on 52173 degrees of freedom
## Multiple R-squared:  0.2326, Adjusted R-squared:  0.2315
## F-statistic: 208.1 on 76 and 52173 DF,  p-value: < 2.2e-16

```

Le modèle linéaire explique 23,3% de la variabilité du score de sévérité.

Facteurs les plus influents sur la gravité

```

# Variables qui ont le plus d'impact sur la sévérité
coefs <- modele3$coefficients[-1] #on enlève l'intercept
tri_coefs <- coefs[order(abs(coefs), decreasing = TRUE)]
head(tri_coefs, 8)

##          catrautoroute RegionÎle-de-France      agghors agglo
##             -0.2229141           -0.1849009        0.1806154
## coldeux veh - frontal col3+ veh - multiple      catrcommunale
##              0.1762504            0.1473271       -0.1228872
##          catrmetropole      situtrottoir
##             -0.1215731            0.1197871

library(dplyr)
library(ggplot2)
library(broom)

## Warning: package 'broom' was built under R version 4.3.3

# On prend la valeur absolue (abs) pour avoir les facteurs les plus INFLUENTS (positifs ou négatifs)

coef_df <- tidy(modele3) %>%
  filter(term != "(Intercept)") %>%
  arrange(desc(abs(estimate))) %>%
  head(8) %>% # On en garde 8
  mutate(
    Type_Influence = ifelse(estimate > 0, "Aggravant", "Protecteur"), # Pour la couleur
  )

```

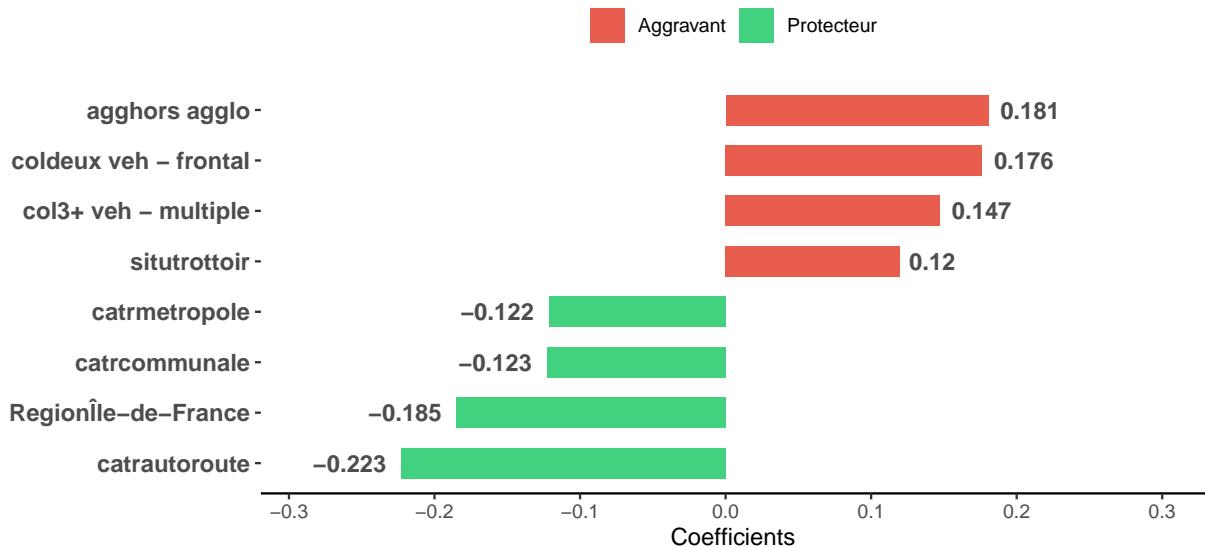
```

ggplot(coef_df, aes(x = reorder(term, estimate), y = estimate, fill = Type_Influence)) +
  geom_col(width = 0.6, alpha = 0.9) +
  # Ajout des valeurs au bout des barres
  geom_text(aes(label = round(estimate, 3)),
            hjust = ifelse(coef_df$estimate > 0, -0.2, 1.2), # Ajuste la position selon le signe
            fontface = "bold", color = "grey30") +
  coord_flip() + # On met à l'horizontale

  scale_fill_manual(values = c("Aggravant" = "#E74C3C", "Protecteur" = "#2ECC71")) +
  scale_y_continuous(limits = c(min(coef_df$estimate) * 1.3, 0.3)) +
  # Titres et Axes
  labs(title = "Facteurs les plus influents sur la Gravité",
       x = "", y = "Coefficients") +
  theme_classic() +
  theme(
    legend.position = "top",
    legend.title = element_blank(),
    axis.text.y = element_text(size = 11, face = "bold"),
    axis.line.y = element_blank(), # Enlève la ligne verticale moche
    plot.title = element_text(face = "bold", size = 16)
  )

```

Facteurs les plus influents sur la Gravité



Les accidents hors agglomération sont en moyenne 20% ($e(0.181) - 1 = 0.198$) plus graves que ceux en agglomération.

Une collision frontale entre deux véhicules augmente le score de sévérité d'environ 19% par rapport aux

accidents sans collision.

Les accidents sur l'autoroute sont en moyenne 20% ($e(-0.223) - 1 = -0.199$) moins graves que les accidents sur une départementale.

Les accidents en Ile-de-France sont en moyenne 17% moins graves que les accidents en Auvergne-Rhône-Alpes.

Remarque: Ces interprétations se font toute choses égales par ailleurs, les autres variables étant constantes. Et la gravité de l'accident est jugé selon notre score de sévérité.