tarification.R

hp

2024-01-03

```
## tarification assurance non vie
### auteur: zakaria sarhan
### importation des packages
########### projet final#################
chooseCRANmirror(graphics=FALSE, ind=1)
knitr::opts_chunk$set(echo = TRUE)
if(!require(readr)) install.packages("readr")
```

```
## Le chargement a nécessité le package : readr
```

```
## Warning: le package 'readr' a été compilé avec la version R 4.3.2
```

```
library (readr)
if(!require(stats)) install.packages("stats")
library(stats)
if(!require(car)) install.packages("car")
```

```
## Le chargement a nécessité le package : car
```

```
## Warning: le package 'car' a été compilé avec la version R 4.3.2
```

```
## Le chargement a nécessité le package : carData
```

```
library(car)
if(!require(ggplot2)) install.packages("ggplot2")
```

```
## Le chargement a nécessité le package : ggplot2
```

```
## Warning: le package 'ggplot2' a été compilé avec la version R 4.3.2
```

```
library(ggplot2)
if(!require(forcats)) install.packages("forcats")
```

```
## Le chargement a nécessité le package : forcats
```

```
## Warning: le package 'forcats' a été compilé avec la version R 4.3.2
```

```r
library(forcats)
if(!require(MASS)) install.packages("Mass")
```

```
## Le chargement a nécessité le package : MASS
```

```r
library(MASS)
if(!require(readr)) install.packages("readr")
library(readr)
if(!require(corrplot)) install.packages("corrplot")
```

```
## Le chargement a nécessité le package : corrplot

## Warning: le package 'corrplot' a été compilé avec la version R 4.3.2

## corrplot 0.92 loaded
```

```r
library(corrplot)
if(!require(gtsummary)) install.packages("gtsummary")
```

```
## Le chargement a nécessité le package : gtsummary

## Warning: le package 'gtsummary' a été compilé avec la version R 4.3.2

## #StandWithUkraine

##
## Attachement du package : 'gtsummary'

## L'objet suivant est masqué depuis 'package:MASS':
##
##     select
```

```r
library(gtsummary)
if(!require(effects)) install.packages("effects")
```

```
## Le chargement a nécessité le package : effects

## Warning: le package 'effects' a été compilé avec la version R 4.3.2

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```r
library(effects)
if(!require(tidyverse)) install.packages("tidyverse")
```

```
## Le chargement a nécessité le package : tidyverse

## Warning: le package 'tidyverse' a été compilé avec la version R 4.3.2
```

```
## Warning: le package 'dplyr' a été compilé avec la version R 4.3.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::recode() masks car::recode()
## x dplyr::select() masks gtsummary::select(), MASS::select()
## x purrr::some()   masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidyverse)
if(!require(actuar)) install.packages("actuar")
```

```
## Le chargement a nécessité le package : actuar

## Warning: le package 'actuar' a été compilé avec la version R 4.3.2

##
## Attachement du package : 'actuar'
##
## Les objets suivants sont masqués depuis 'package:stats':
##
##     sd, var
##
## L'objet suivant est masqué depuis 'package:grDevices':
##
##     cm
```

```
library(actuar)
if(!require(Hmisc)) install.packages("Hmisc")
```

```
## Le chargement a nécessité le package : Hmisc

## Warning: le package 'Hmisc' a été compilé avec la version R 4.3.2

##
## Attachement du package : 'Hmisc'
##
## Les objets suivants sont masqués depuis 'package:dplyr':
##
##     src, summarize
##
## Les objets suivants sont masqués depuis 'package:base':
##
##     format.pval, units
```

```r
library(Hmisc)
if(!require(summarytools)) install.packages("summarytools")
```

```
## Le chargement a nécessité le package : summarytools

## Warning: le package 'summarytools' a été compilé avec la version R 4.3.2

##
## Attachement du package : 'summarytools'
##
## Les objets suivants sont masqués depuis 'package:Hmisc':
##
##     label, label<-
##
## L'objet suivant est masqué depuis 'package:tibble':
##
##     view
```

```r
library(summarytools)
if(!require(sandwich)) install.packages("sandwich")
```

```
## Le chargement a nécessité le package : sandwich

## Warning: le package 'sandwich' a été compilé avec la version R 4.3.2
```

```r
library(sandwich)
if(!require(AER)) install.packages("AER")
```

```
## Le chargement a nécessité le package : AER

## Warning: le package 'AER' a été compilé avec la version R 4.3.2

## Le chargement a nécessité le package : lmtest
## Le chargement a nécessité le package : zoo
##
## Attachement du package : 'zoo'
##
## Les objets suivants sont masqués depuis 'package:base':
##
##     as.Date, as.Date.numeric
##
## Le chargement a nécessité le package : survival
```

```r
library(AER)
if(!require(tree)) install.packages("tree")
```

```
## Le chargement a nécessité le package : tree

## Warning: le package 'tree' a été compilé avec la version R 4.3.2
```

```r
library(tree)
if(!require(vcd)) install.packages("vcd")
```

```
## Le chargement a nécessité le package : vcd
```

```
## Warning: le package 'vcd' a été compilé avec la version R 4.3.2
```

```
## Le chargement a nécessité le package : grid
```

```r
library(vcd)
if(!require(pscl)) install.packages("pscl")
```

```
## Le chargement a nécessité le package : pscl
```

```
## Warning: le package 'pscl' a été compilé avec la version R 4.3.2
```

```
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
```

```r
library(pscl)
if(!require(lmtest)) install.packages("lmtest")
library(lmtest)
library(dplyr)
##### importation des données #############
sinistre=read.csv("base sinistre.csv",sep=";",header=TRUE)
production=read.csv("base production.csv",sep=";",header=TRUE)
head(sinistre)
```

```
##       id_policy id_client id_vehicle id_claim id_year claim_amount claim_nb
## 1 A00000022-V01 A00000022        V01     CL01  Year 0         1181        1
## 2 A00000022-V01 A00000022        V01     CL02  Year 0          215        1
## 3 A00000022-V01 A00000022        V01     CL03  Year 0          314        1
## 4 A00000028-V01 A00000028        V01     CL01  Year 0         6203        1
## 5 A00000034-V01 A00000034        V01     CL01  Year 0           49        1
## 6 A00000044-V02 A00000044        V02     CL01  Year 0          919        1
```

```r
attach(sinistre)

sinistre_positive=subset(sinistre,claim_amount>=0)
# on voit que certain donnes en virgule et nous aimerons des points ##########
sinistre_positive$claim_amount=gsub(",",".",sinistre_positive$claim_amount)
str(sinistre_positive)
```

```
## 'data.frame':    13238 obs. of  7 variables:
##  $ id_policy   : chr  "A00000022-V01" "A00000022-V01" "A00000022-V01" "A00000028-V01" ...
##  $ id_client   : chr  "A00000022" "A00000022" "A00000022" "A00000028" ...
```

```
##  $ id_vehicle  : chr  "V01" "V01" "V01" "V01" ...
##  $ id_claim    : chr  "CL01" "CL02" "CL03" "CL01" ...
##  $ id_year     : chr  "Year 0" "Year 0" "Year 0" "Year 0" ...
##  $ claim_amount: chr  "1181" "215" "314" "6203" ...
##  $ claim_nb    : int  1 1 1 1 1 1 1 1 1 1 ...
```

```
sinistre_positive$claim_amount=as.numeric(sinistre_positive$claim_amount)
# rassemnler les base de sinistre et production en fonction de id policy#############
sinistre_total=sinistre_positive %>% group_by(id_policy) %>% filter(claim_amount>=0) %>%
  summarise(claim_amount=sum(claim_amount))
nbr_total=sinistre_positive %>% group_by(id_policy) %>% filter(claim_amount>=0) %>%
  summarise(claim_nb=n())

base_etude=production %>% full_join(sinistre_total,by="id_policy")
base_etude$claim_amount[is.na(base_etude$claim_amount)]=0
base_etude=base_etude %>% full_join(nbr_total,by="id_policy")
base_etude$claim_nb[is.na(base_etude$claim_nb)]=0
sum(is.na(base_etude))
```

```
## [1] 0
```

```
# renamer les nouvelles variables###########
base_etude=base_etude %>% rename("montant_sinistre"=claim_amount,"nombre_sinistre"=claim_nb)
######## pretraitement des données #################
str(base_etude)
```

```
## 'data.frame':    100000 obs. of  33 variables:
##  $ id_year         : chr  "Year 0" "Year 0" "Year 0" "Year 0" ...
##  $ id_policy       : chr  "A00000004-V01" "A00000004-V02" "A00000007-V01" "A00000008-V01" ...
##  $ drv_age1        : int  50 26 18 45 18 18 18 50 18 20 ...
##  $ drv_age_lic1    : int  30 22 0 37 0 0 0 30 0 42 ...
##  $ drv_sex1        : chr  "M" "F" "M" "M" ...
##  $ drv_age2        : int  26 0 0 0 0 63 0 0 32 24 ...
##  $ drv_age_lic2    : int  7 0 0 0 0 44 0 0 14 5 ...
##  $ drv_sex2        : chr  "F" "M" "F" "F" ...
##  $ drv_drv2        : chr  "Yes" "No" "No" "No" ...
##  $ id_client       : chr  "A00000004" "A00000004" "A00000007" "A00000008" ...
##  $ vh_age          : int  16 16 7 11 6 6 5 13 13 1 ...
##  $ vh_cyl          : int  1781 1781 1870 1595 1997 1997 1798 1905 1905 1560 ...
##  $ vh_din          : int  90 90 108 101 90 90 127 68 68 109 ...
##  $ vh_fuel         : chr  "Gasoline" "Gasoline" "Diesel" "Gasoline" ...
##  $ vh_make         : chr  "VOLKSWAGEN" "VOLKSWAGEN" "RENAULT" "AUDI" ...
##  $ vh_model        : chr  "GOLF" "GOLF" "LAGUNA" "A4" ...
##  $ vh_sale_begin   : int  18 18 10 16 9 9 6 14 14 3 ...
##  $ vh_sale_end     : int  15 15 6 13 7 7 3 13 13 1 ...
##  $ vh_speed        : int  180 180 193 191 163 163 196 162 162 180 ...
##  $ vh_type         : chr  "Tourism" "Tourism" "Tourism" "Tourism" ...
##  $ vh_value        : int  14407 14407 22450 20535 18550 18550 22450 14773 14773 27100 ...
##  $ vh_weight       : int  1020 1020 1350 1195 1110 1110 1080 1106 1106 1530 ...
##  $ id_vehicle      : chr  "V01" "V02" "V01" "V01" ...
##  $ pol_bonus       : num  0.5 0.57 0.5 0.5 0.68 0.95 0.5 0.5 0.6 0.5 ...
##  $ pol_coverage    : chr  "Maxi" "Maxi" "Maxi" "Maxi" ...
##  $ pol_duration    : int  4 5 13 25 5 1 25 17 9 2 ...
```

```
##  $ pol_sit_duration: int  1 2 2 2 1 1 5 1 2 1 ...
##  $ pol_pay_freq    : chr  "Yearly" "Biannual" "Monthly" "Yearly" ...
##  $ pol_payd        : chr  "No" "No" "No" "No" ...
##  $ pol_usage       : chr  "WorkPrivate" "WorkPrivate" "WorkPrivate" "Professional" ...
##  $ pol_insee_code  : chr  "59063" "42225" "33090" "67496" ...
##  $ montant_sinistre: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ nombre_sinistre : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
summary(base_etude)
```

```
##    id_year           id_policy            drv_age1      drv_age_lic1
##  Length:100000      Length:100000      Min.   :18.00   Min.   : 0.00
##  Class :character   Class :character   1st Qu.:18.00   1st Qu.: 0.00
##  Mode  :character   Mode  :character   Median :25.00   Median :24.00
##                                        Mean   :30.66   Mean   :21.94
##                                        3rd Qu.:40.00   3rd Qu.:38.00
##                                        Max.   :94.00   Max.   :73.00
##    drv_sex1            drv_age2      drv_age_lic2      drv_sex2
##  Length:100000      Min.   : 0.00   Min.   : 0.000   Length:100000
##  Class :character   1st Qu.: 0.00   1st Qu.: 0.000   Class :character
##  Mode  :character   Median : 0.00   Median : 0.000   Mode  :character
##                     Mean   :15.56   Mean   : 8.951
##                     3rd Qu.:34.00   3rd Qu.:14.000
##                     Max.   :99.00   Max.   :80.000
##    drv_drv2          id_client             vh_age           vh_cyl
##  Length:100000      Length:100000      Min.   : 1.000   Min.   :   0
##  Class :character   Class :character   1st Qu.: 4.000   1st Qu.:1360
##  Mode  :character   Mode  :character   Median : 8.000   Median :1587
##                                        Mean   : 9.532   Mean   :1647
##                                        3rd Qu.:13.000   3rd Qu.:1910
##                                        Max.   :63.000   Max.   :6217
##      vh_din          vh_fuel            vh_make            vh_model
##  Min.   : 15.00   Length:100000      Length:100000      Length:100000
##  1st Qu.: 68.00   Class :character   Class :character   Class :character
##  Median : 87.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 91.44
##  3rd Qu.:109.00
##  Max.   :555.00
##  vh_sale_begin    vh_sale_end        vh_speed       vh_type
##  Min.   : 1.00   Min.   : 1.000   Min.   : 25.0   Length:100000
##  1st Qu.: 6.00   1st Qu.: 4.000   1st Qu.:157.0   Class :character
##  Median :10.00   Median : 7.000   Median :170.0   Mode  :character
##  Mean   :11.64   Mean   : 8.661   Mean   :170.7
##  3rd Qu.:15.00   3rd Qu.:12.000   3rd Qu.:185.0
##  Max.   :74.00   Max.   :55.000   Max.   :310.0
##     vh_value        vh_weight     id_vehicle          pol_bonus
##  Min.   :     0   Min.   :   0   Length:100000      Min.   :0.5000
##  1st Qu.: 11950   1st Qu.: 950   Class :character   1st Qu.:0.5000
##  Median : 16280   Median :1130   Mode  :character   Median :0.5000
##  Mean   : 18085   Mean   :1129                      Mean   :0.5372
##  3rd Qu.: 22106   3rd Qu.:1320                      3rd Qu.:0.5000
##  Max.   :145000   Max.   :7901                      Max.   :1.6500
##  pol_coverage        pol_duration   pol_sit_duration pol_pay_freq
##  Length:100000      Min.   : 1.00   Min.   : 1.000   Length:100000
```

```
##  Class :character   1st Qu.: 4.00   1st Qu.: 1.000   Class :character
##  Mode  :character   Median : 9.00   Median : 2.000   Mode  :character
##                     Mean   :11.11   Mean   : 2.732
##                     3rd Qu.:17.00   3rd Qu.: 3.000
##                     Max.   :40.00   Max.   :24.000
##    pol_payd          pol_usage         pol_insee_code     montant_sinistre
##  Length:100000      Length:100000      Length:100000      Min.   :     0.0
##  Class :character   Class :character   Class :character   1st Qu.:     0.0
##  Mode  :character   Mode  :character   Mode  :character   Median :     0.0
##                                                           Mean   :   194.4
##                                                           3rd Qu.:     0.0
##                                                           Max.   :234104.0
##  nombre_sinistre
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.1324
##  3rd Qu.:0.0000
##  Max.   :6.0000
```

```r
base_etude= base_etude %>% select(-id_vehicle,-id_year)
base_etude= base_etude %>% select(-vh_model)
base_etude= base_etude %>% select(-pol_insee_code)
base_etude$drv_sex1=as.factor(base_etude$drv_sex1)
base_etude$drv_sex2=as.factor(base_etude$drv_sex2)
base_etude$vh_fuel=as.factor(base_etude$vh_fuel)
base_etude$vh_make=as.factor(base_etude$vh_make)
base_etude$pol_usage=as.factor(base_etude$pol_usage)
base_etude$pol_coverage=as.factor(base_etude$pol_coverage)
base_etude$pol_pay_freq=as.factor(base_etude$pol_pay_freq)
base_etude$vh_type=as.factor(base_etude$vh_type)
############## statistique descriptives########
################### statistique univariée##########
theme_set(theme_bw())
theme_gtsummary_mean_sd()
base_etude %>% select(drv_age1,drv_sex1,drv_age2,drv_sex2,vh_age,vh_cyl,vh_fuel,vh_make,pol_bonus,
                      pol_usage,pol_pay_freq,montant_sinistre,nombre_sinistre) %>% tbl_summary(
                        digits = all_categorical()~c(1,0),
                        statistic = all_categorical()~"{p}%[{n}]"
                      )
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

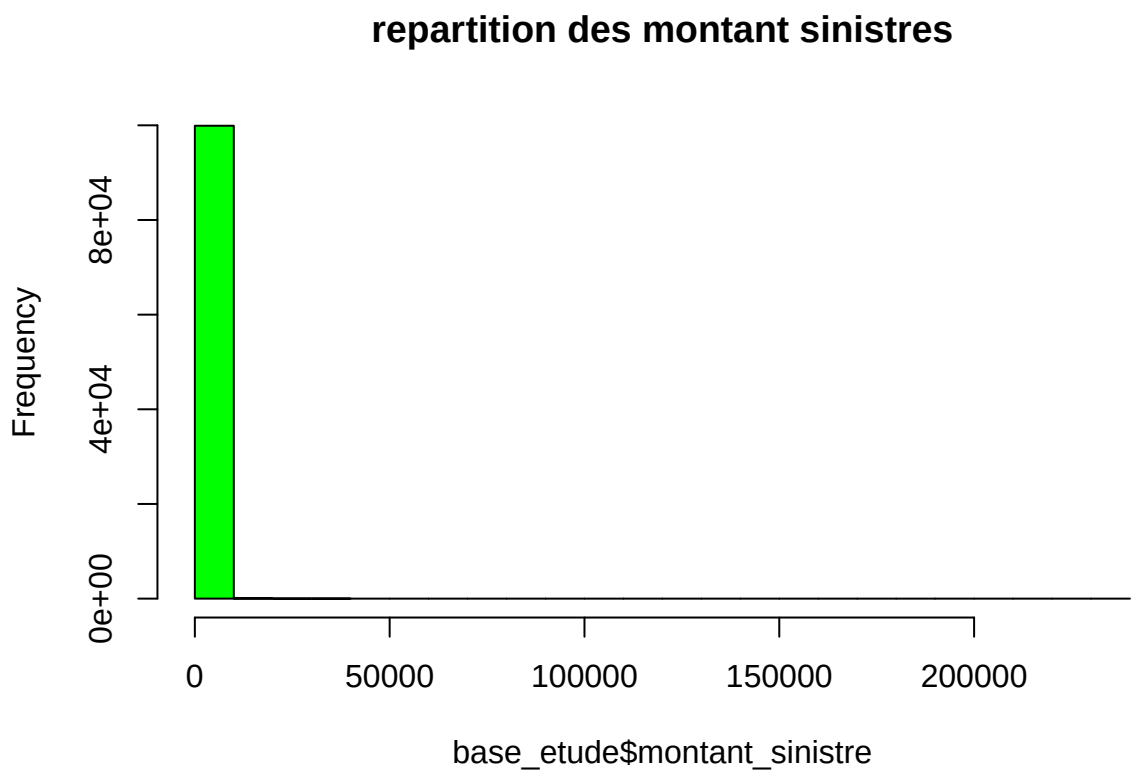
Characteristic	**N = 100,000**
drv_age1	31 (15)
drv_sex1	
F	39.8%[39,766]
M	60.2%[60,234]
drv_age2	16 (24)
drv_sex2	

Characteristic	**N = 100,000**
	66.8%[66,810]
F	20.3%[20,325]
M	12.9%[12,865]
vh_age	10 (7)
vh_cyl	1,647 (462)
vh_fuel	
Diesel	55.3%[55,256]
Gasoline	44.7%[44,666]
Hybrid	0.1%[78]
vh_make	
ACL	0.0%[10]
ALFA ROMEO	0.4%[400]
ALPINE	0.0%[2]
APAL	0.0%[1]
ARO	0.0%[5]
ASIA	0.0%[2]
AUDI	1.5%[1,504]
AUSTIN	0.0%[14]
AUSTIN HEALEY	0.0%[1]
AUTOBIANCHI	0.0%[8]
AUVERLAND	0.0%[5]
BABOULIN	0.0%[5]
BERTONE	0.0%[1]
BMW	1.5%[1,522]
BREMACH	0.0%[2]
BUICK	0.0%[2]
CADILLAC	0.0%[1]
CHEVROLET	0.3%[273]
CHRYSLER	0.2%[195]
CITROEN	16.1%[16,068]
COURNIL	0.0%[3]
DACIA	1.2%[1,214]
DAEWOO	0.1%[141]
DAF	0.0%[7]
DAIHATSU	0.1%[73]
DAIMLER	0.0%[10]
DATSUN	0.0%[4]
DODGE	0.0%[25]
EBRO	0.0%[39]
FERRARI	0.0%[3]
FIAT	2.9%[2,930]
FORD	4.4%[4,379]
FSO	0.0%[2]
GME	0.0%[10]
HONDA	0.6%[640]
HOTCHKISS	0.0%[19]
HYUNDAI	0.6%[625]
ISUZU	0.0%[25]
IVECO	0.1%[115]
JAGUAR	0.1%[90]
JEEP	0.2%[172]
KIA	0.4%[429]

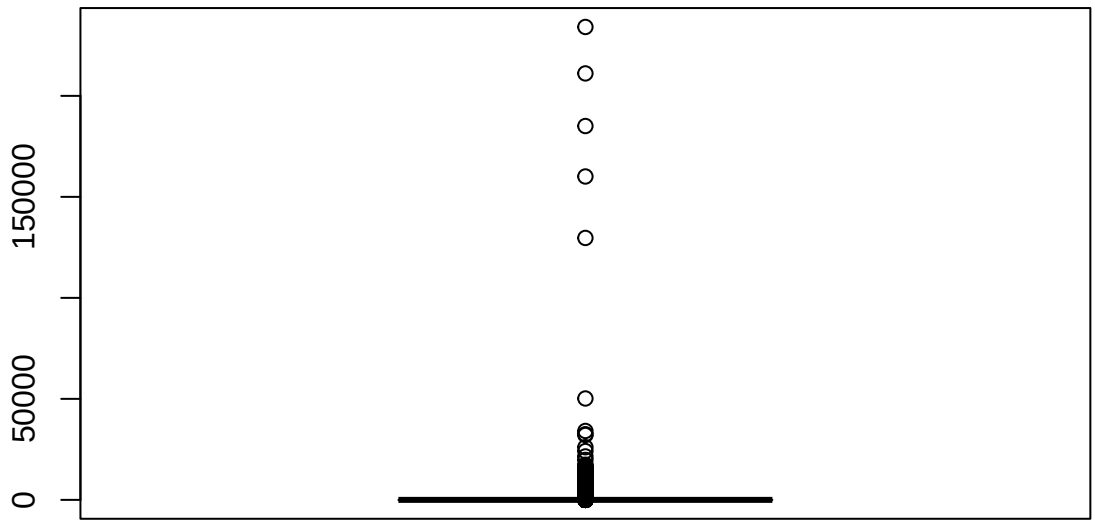
Characteristic	N = 100,000
LADA VAZ	0.1%[77]
LANCIA	0.1%[148]
LAND ROVER	0.3%[345]
LDV	0.0%[3]
LEXUS	0.0%[19]
LOTUS	0.0%[8]
MAHINDRA	0.0%[7]
MASERATI	0.0%[4]
MATRA	0.0%[2]
MAZDA	0.4%[436]
MEGA	0.0%[1]
MERCEDES BENZ	3.0%[3,047]
MG	0.0%[32]
MINI	0.4%[396]
MITSUBISHI	0.3%[348]
MORGAN	0.0%[13]
MORRIS	0.0%[4]
NISSAN	1.8%[1,761]
OM	0.0%[1]
OPEL	3.6%[3,590]
PANHARD	0.0%[1]
PEUGEOT	19.8%[19,780]
PIAGGIO	0.0%[4]
PININFARINA	0.0%[1]
PONTIAC	0.0%[12]
PORSCHE	0.1%[73]
RENAULT	26.4%[26,441]
RILEY	0.0%[1]
ROVER	0.2%[234]
SAAB	0.1%[127]
SANTANA	0.1%[140]
SAVIEM	0.0%[38]
SEAT	1.0%[1,040]
SIMCA	0.0%[14]
SKODA	0.4%[402]
SMART	0.2%[185]
SSANGYONG	0.0%[38]
STEYR PUCH	0.0%[3]
SUBARU	0.0%[42]
SUZUKI	0.8%[819]
TALBOT	0.0%[11]
TEILHOL	0.0%[7]
TOYOTA	3.3%[3,295]
TRIUMPH	0.0%[22]
UMM	0.0%[2]
UNIC	0.1%[103]
VD 4 ROUES	0.1%[70]
VOLKSWAGEN	5.4%[5,367]
VOLVO	0.5%[494]
VW PORSCHE	0.0%[3]
WILLYS	0.0%[13]
pol_bonus	0.54 (0.10)

Characteristic	N = 100,000
pol_usage	
AllTrips	0.1%[99]
Professional	7.2%[7,202]
Retired	26.7%[26,673]
WorkPrivate	66.0%[66,026]
pol_pay_freq	
Biannual	29.5%[29,535]
Monthly	30.3%[30,256]
Quarterly	2.5%[2,521]
Yearly	37.7%[37,688]
montant_sinistre	194 (1,571)
nombre_sinistre	
0	88.2%[88,189]
1	10.5%[10,533]
2	1.2%[1,153]
3	0.1%[106]
4	0.0%[15]
5	0.0%[3]
6	0.0%[1]

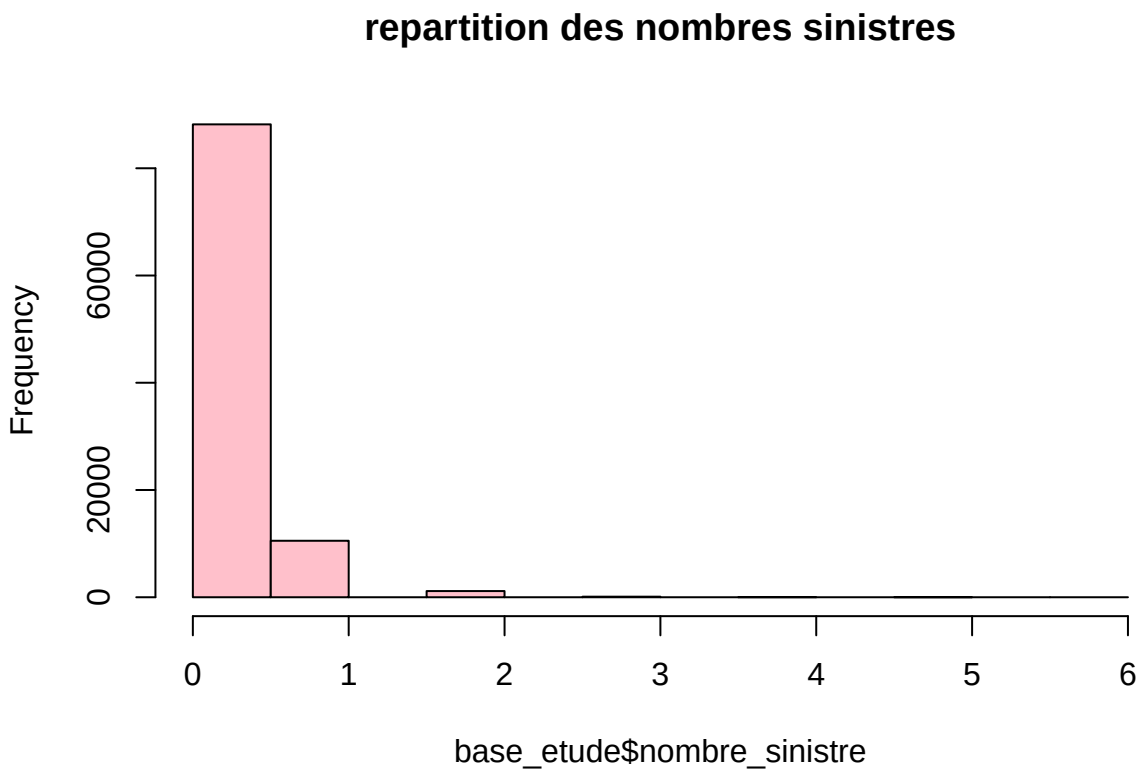
```
########## visualisation des données univariée variables quantitatives#########
hist(base_etude$montant_sinistre,plot = TRUE,main="repartition des montant sinistres",col="green")
```



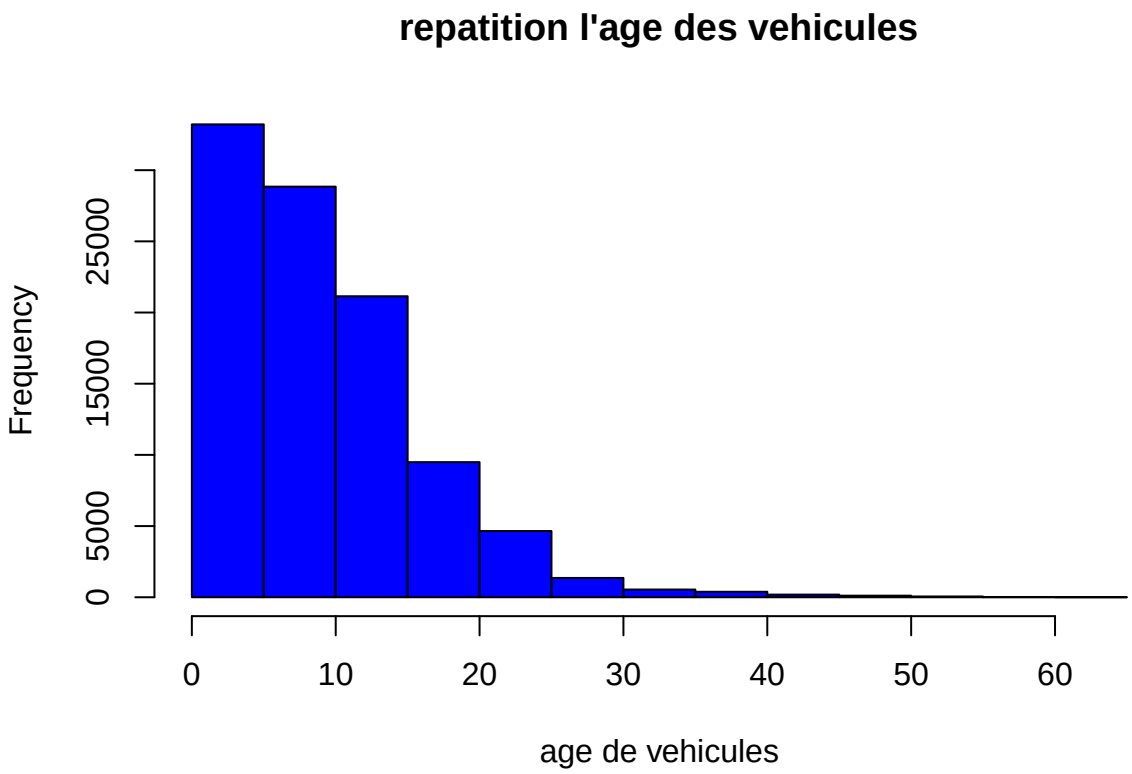
```r
boxplot(base_etude$montant_sinistre)
```



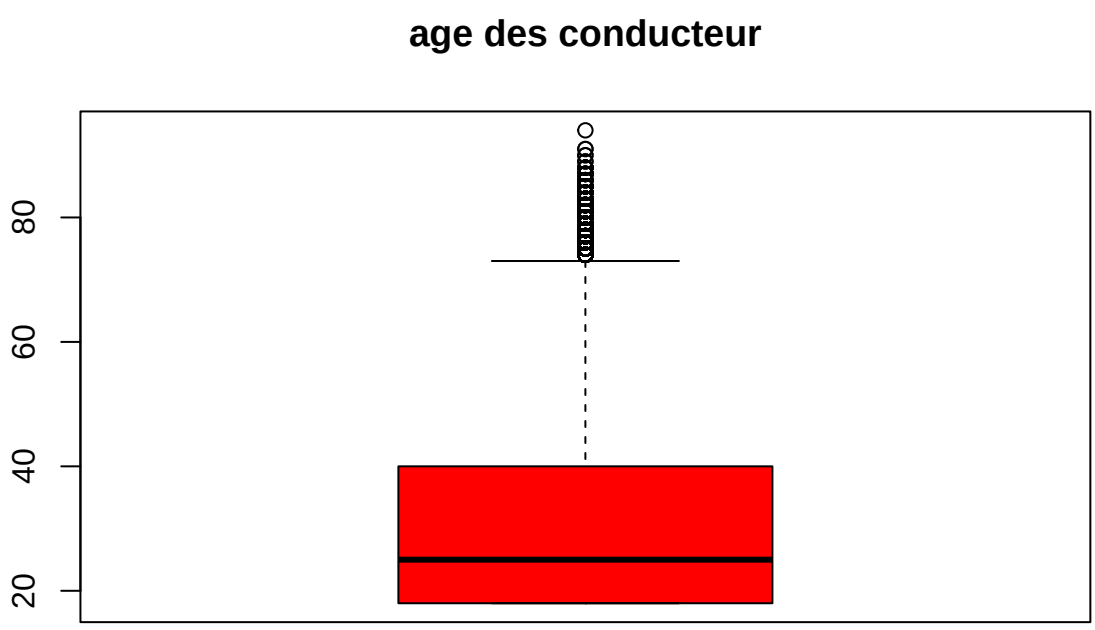

```r
hist(base_etude$nombre_sinistre,freq=TRUE,plot = TRUE,main="repartition des nombres sinistres",col="pink
```



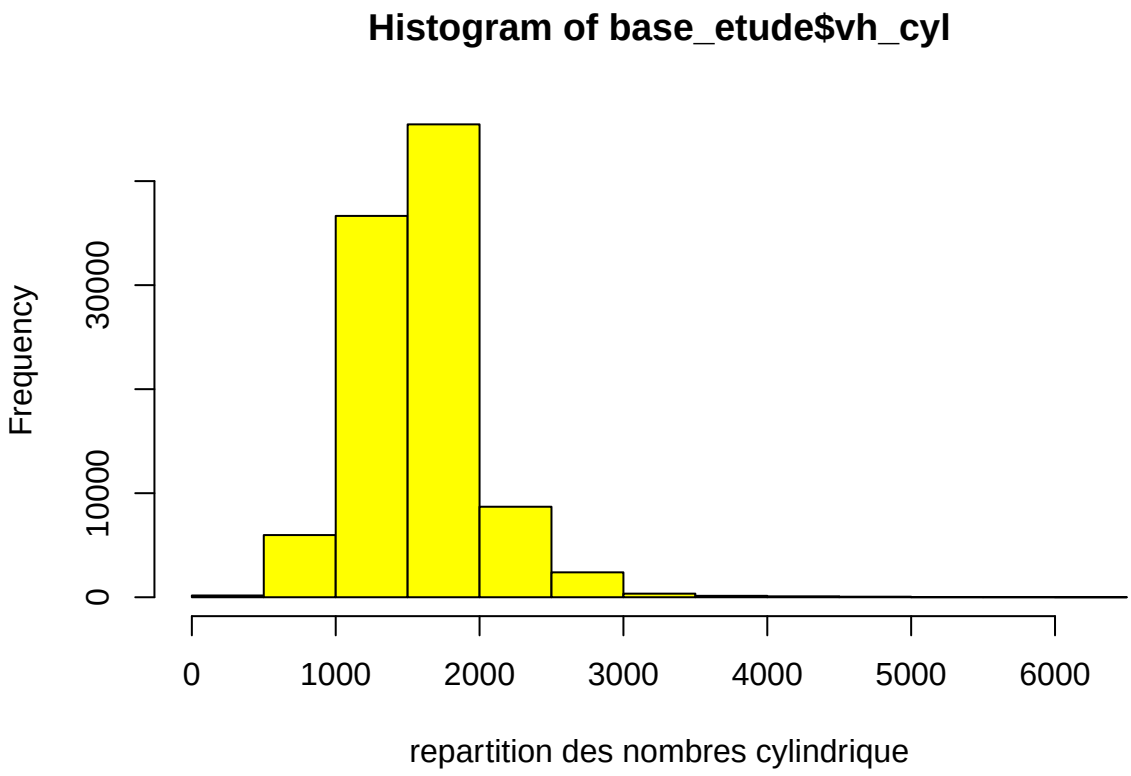

```r
hist(base_etude$vh_age,xlab="age de vehicules",plot=TRUE,main="repatition l'age des vehicules",col="blu
```




```r
boxplot(base_etude$drv_age1,main="age des conducteur",col="red")
```



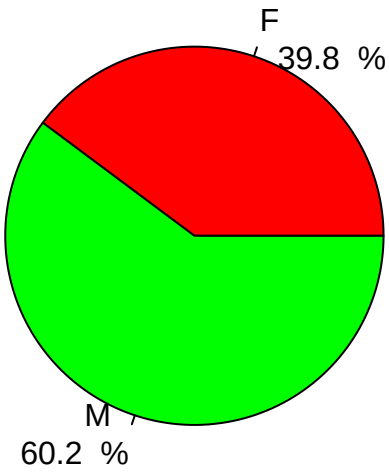

```r
hist(base_etude$vh_cyl,xlab='repartition des nombres cylindrique',plot=TRUE,col="yellow")
```



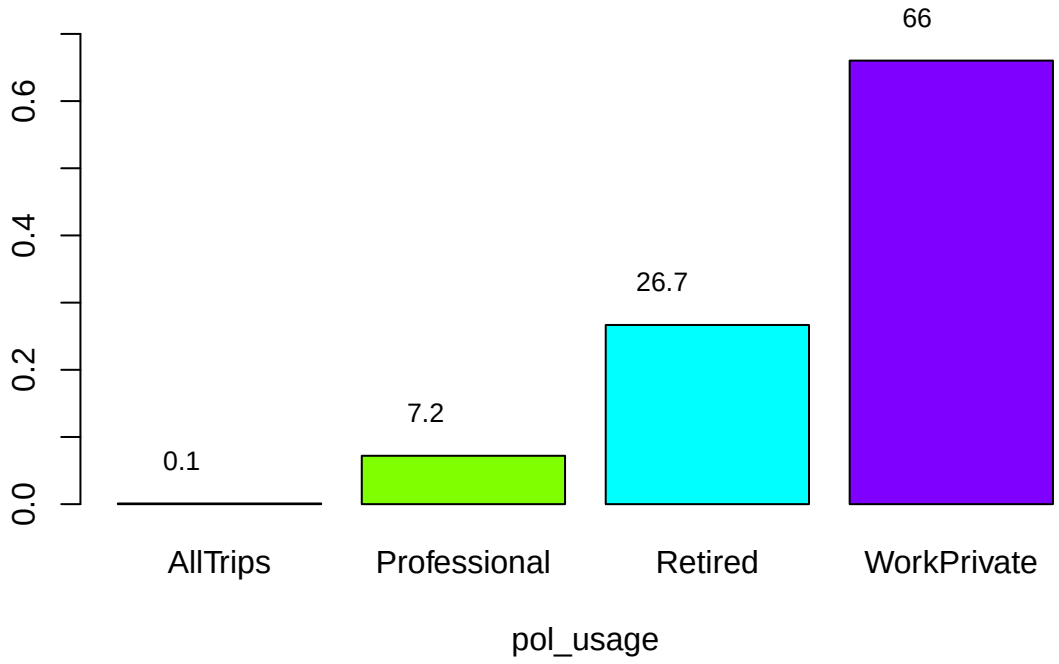

```
################ statistique univariée pour variables qualitatives##########
tt=table(base_etude$drv_sex1)
tab=round(prop.table(tt)*100,1)
print(tab)
```

```
##
##    F    M
## 39.8 60.2
```

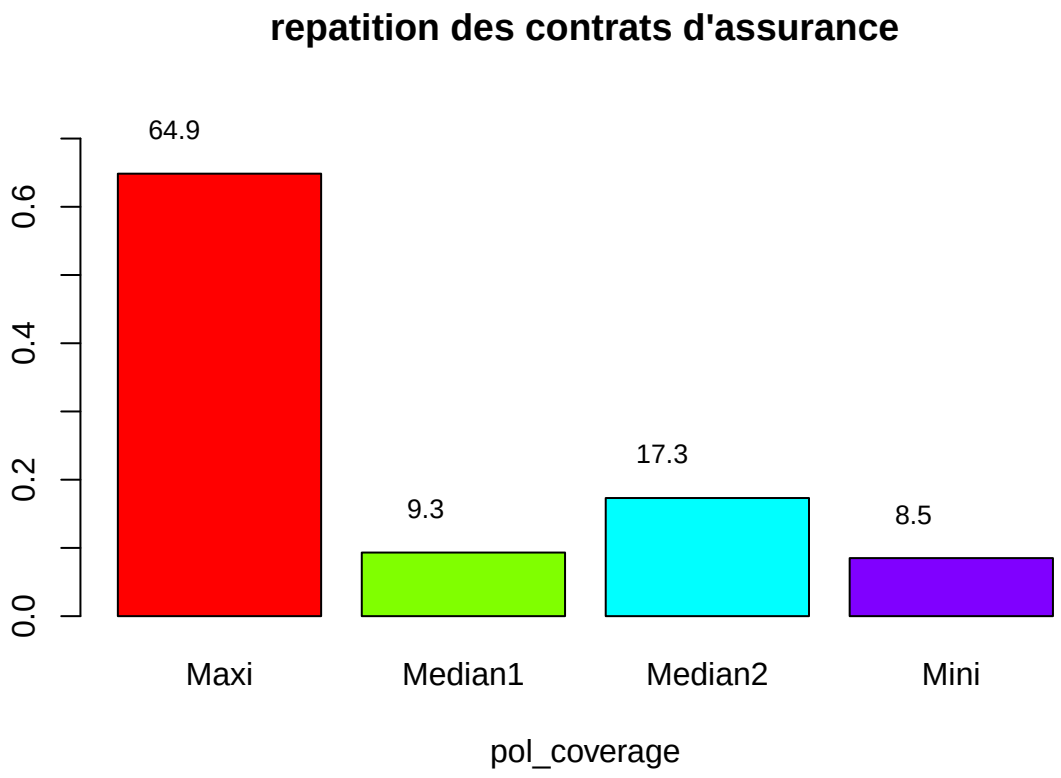
```
label=paste(names(tt),"\n",tab,"%",sep="  ")
pie(tab,labels = label,main="repartition de genre ",col=c("red","green"))
```



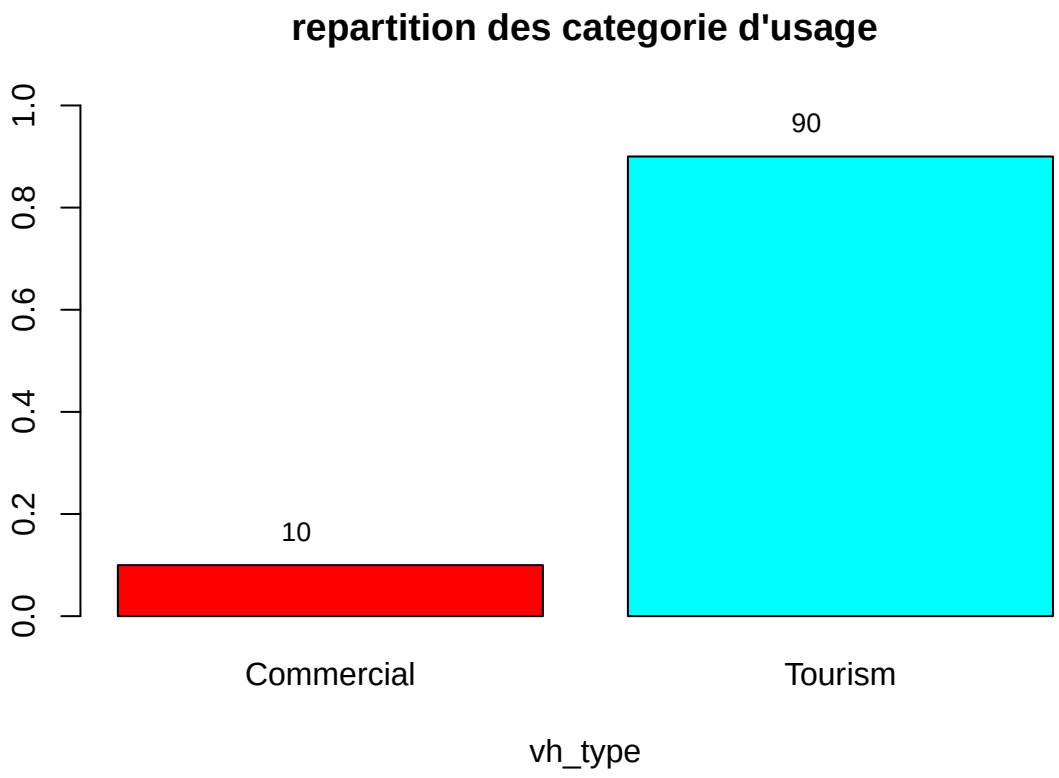

```
#### on va creer une fonction pour visualiser les variables qualitative
bar_plot=function(variable,data,title){
  tta=table(data[[variable]])
  pop=prop.table(tta)
  bp=barplot(pop,main=title,xlab=variable,col=rainbow(length(pop)),ylim=c(0,max(pop)+0.1))
             text(bp,pop+0.06,round(pop*100,1),cex=0.8,pos=2)
}
bar_plot("pol_usage",base_etude,"repartition des usage des vehicules")
```



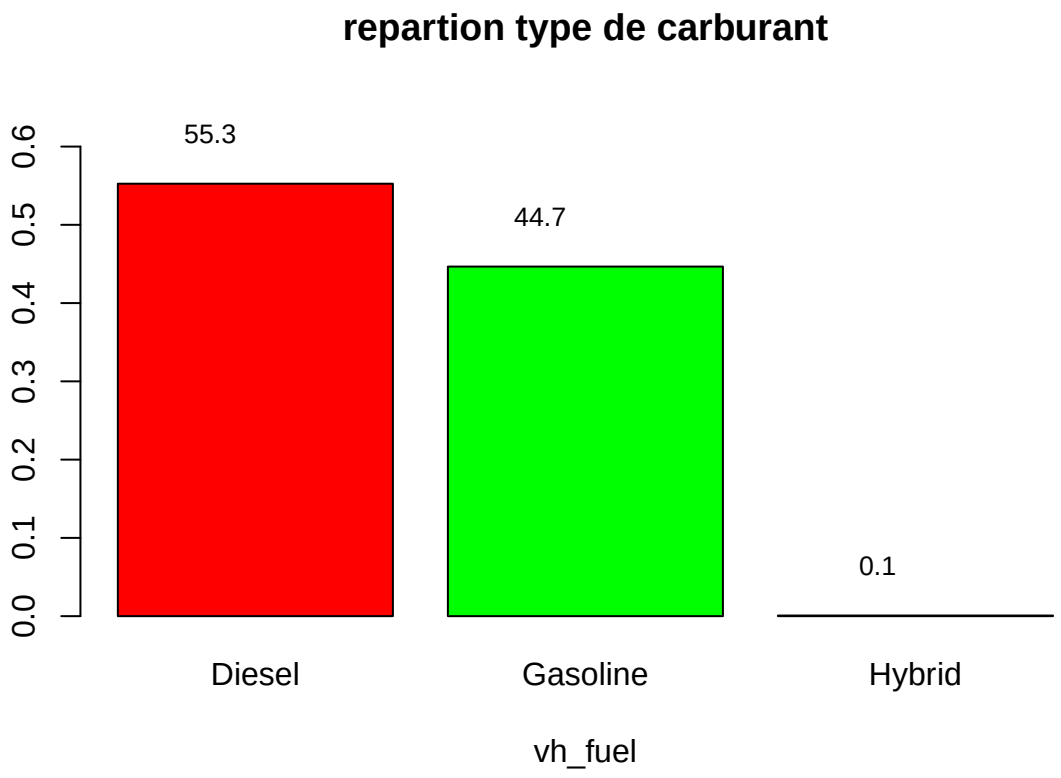

```
bar_plot("pol_coverage",base_etude,"repatition des contrats d'assurance")
```



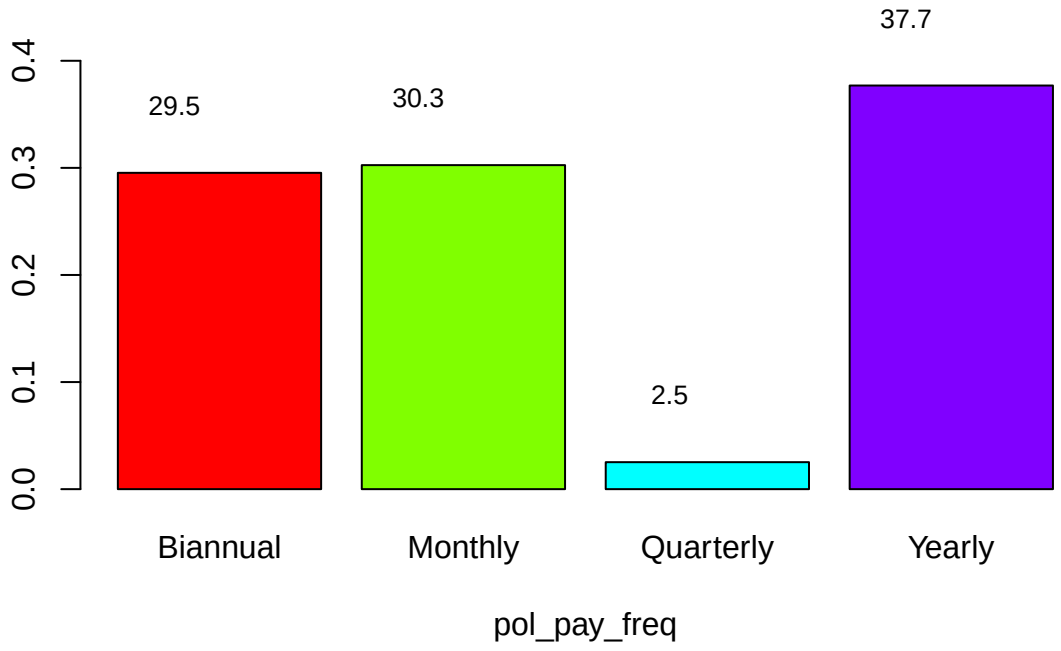

```
bar_plot("vh_type",base_etude,"repartition des categorie d'usage")
```




```
bar_plot("vh_fuel",base_etude,"repartion type de carburant")
```




```
bar_plot("pol_pay_freq",base_etude,"repartion des frequence de paiement")
```




```
################## statistique bivariée##############
############## variable quantitative###########
montant_sin_diesel=mean(base_etude$montant_sinistre[base_etude$vh_fuel=="Diesel"])
print(montant_sin_diesel)
```

```
## [1] 186.8386
```

```
montant_sin_gasoline=mean(base_etude$montant_sinistre[base_etude$vh_fuel=="Gasoline"])
print(montant_sin_gasoline)
```

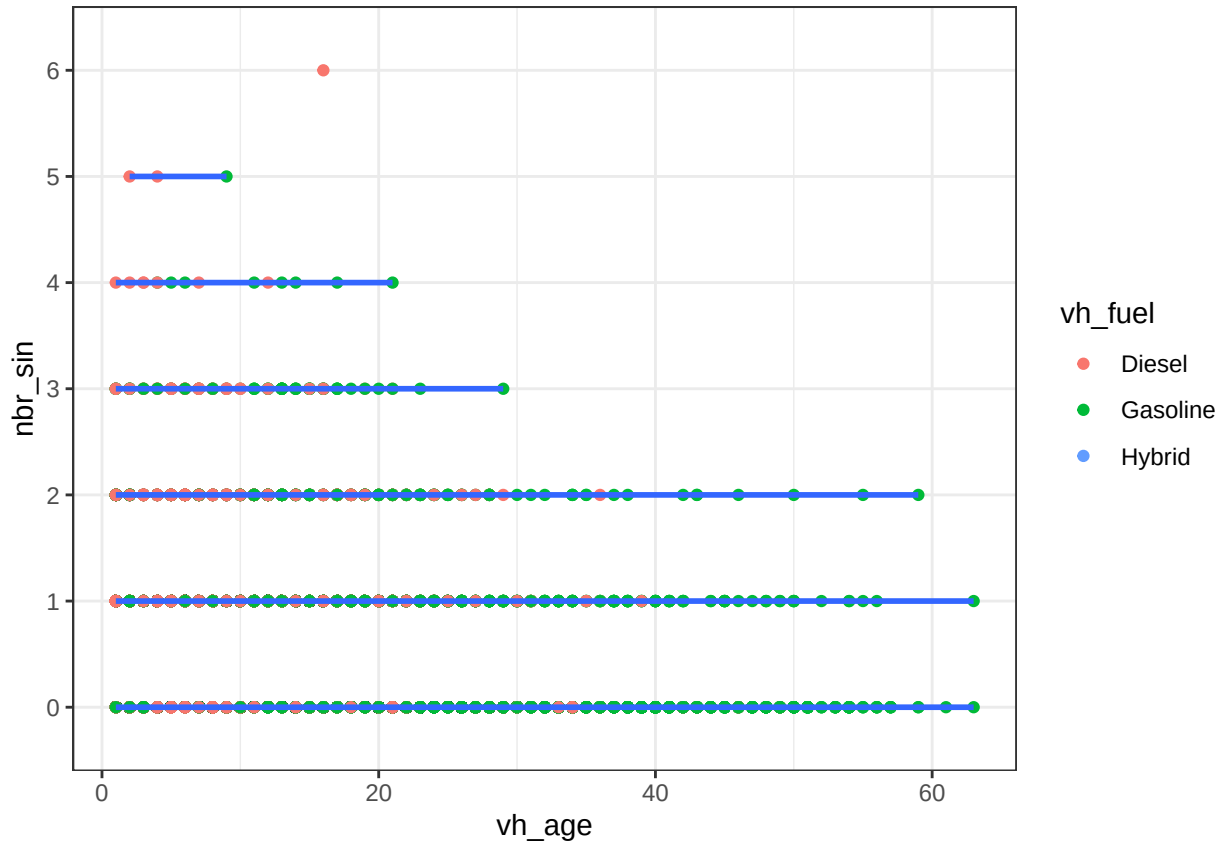
```
## [1] 203.9556
```

```
montant_sin_hybrid=mean(base_etude$montant_sinistre[base_etude$vh_fuel=="Hybrid"])
print(montant_sin_hybrid)
```

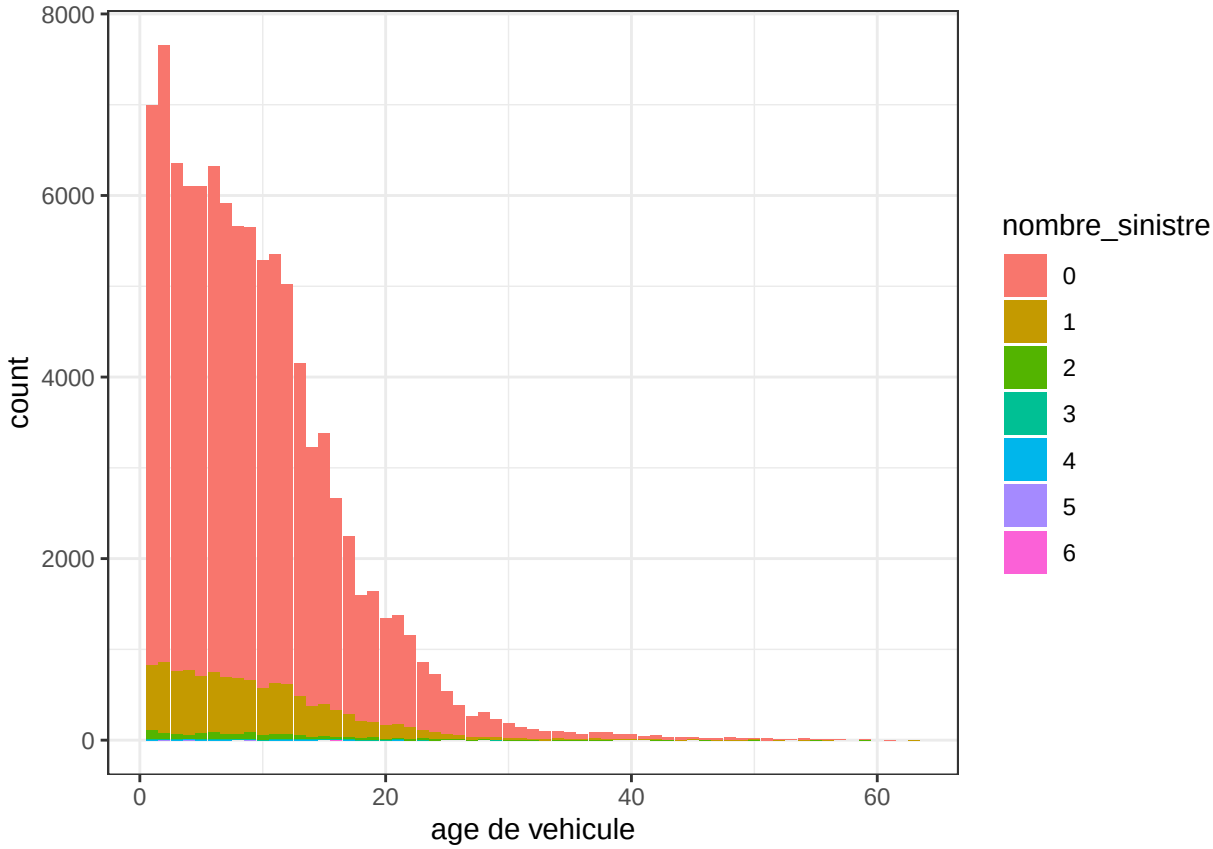
```
## [1] 132.8077
```

```
# visualisation des variables quantitative/quantitative
#base_etude$nombre_sinistre=as.factor(base_etude$nombre_sinistre)
base_etude$nbr_sin=as.factor(base_etude$nombre_sinistre)
ggplot(base_etude,aes(y=nbr_sin,x=vh_age))+geom_point(aes(col=vh_fuel))+geom_smooth(method="lm",se=T)
```

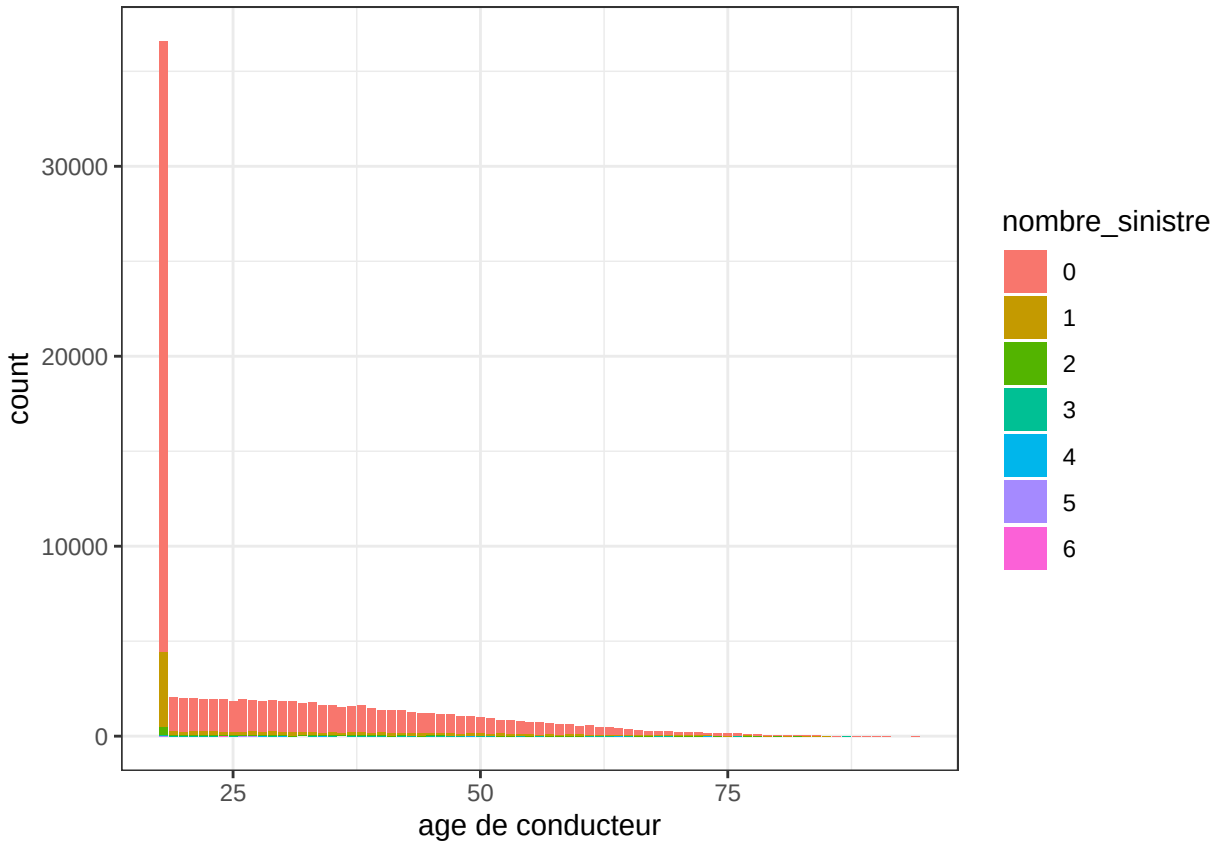
```
## `geom_smooth()` using formula = 'y ~ x'
```



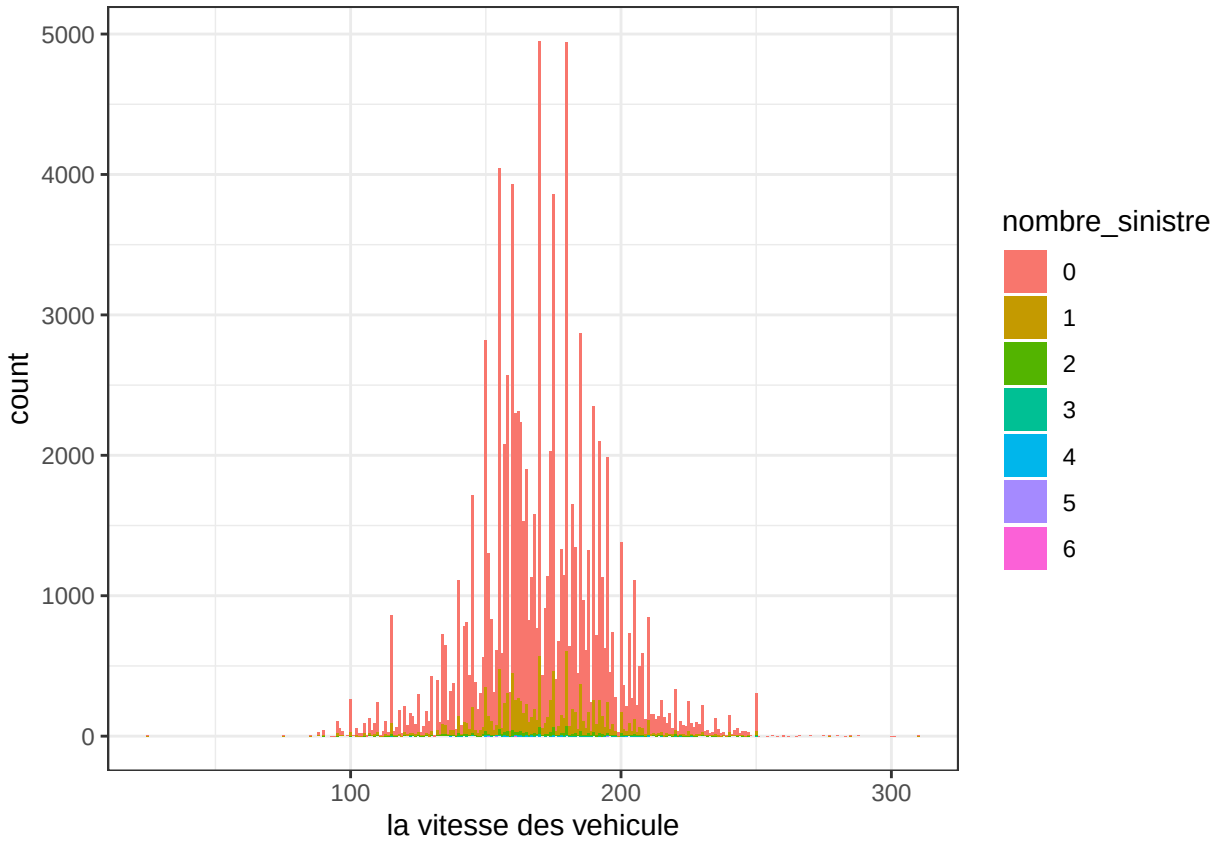

```
ggplot(base_etude,aes(x=vh_age,fill=nbr_sin))+geom_bar()+xlab("age de vehicule")+labs(fill="nombre_sinis
```




```
ggplot(base_etude,aes(x=drv_age1,fill=nbr_sin))+geom_bar()+xlab("age de conducteur")+labs(fill="nombre_
```

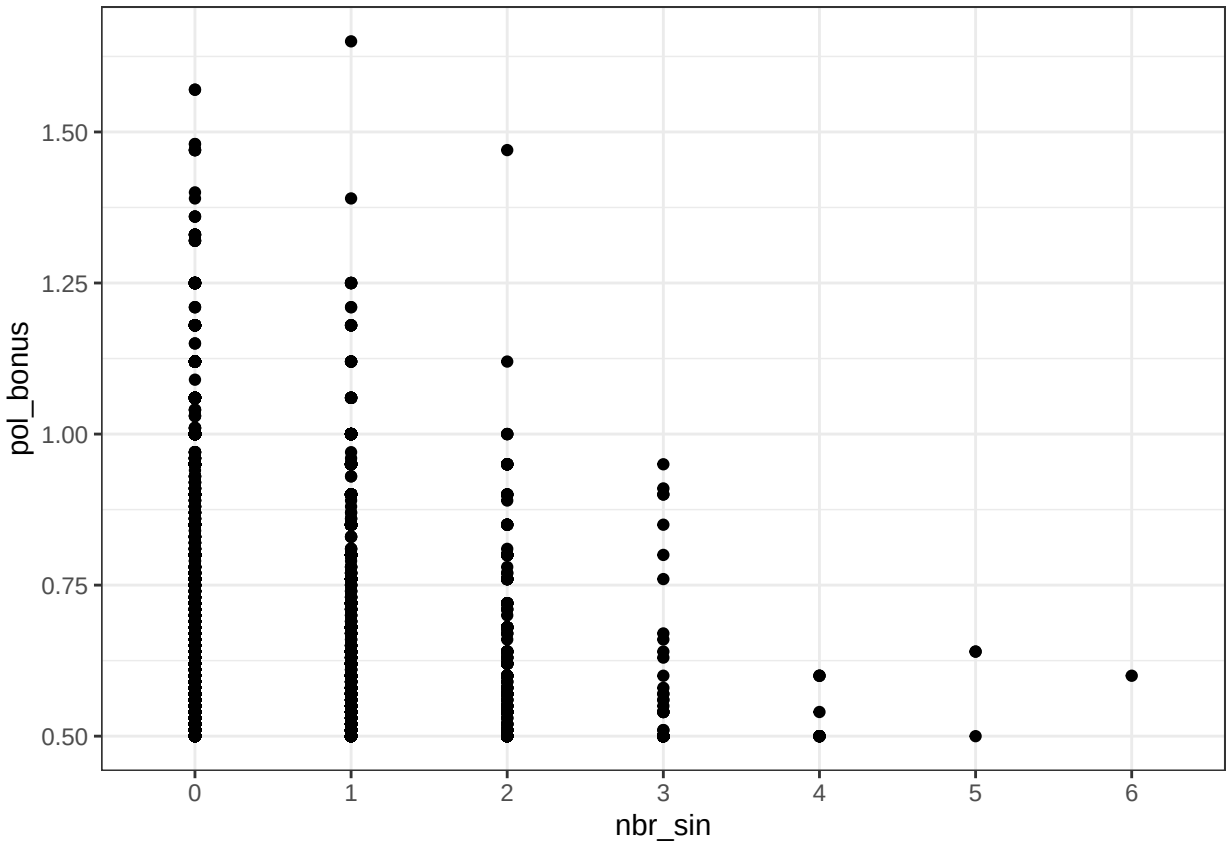



```
ggplot(base_etude,aes(x=vh_speed,fill=nbr_sin))+geom_bar()+xlab("la vitesse des vehicule")+labs(fill="n
```



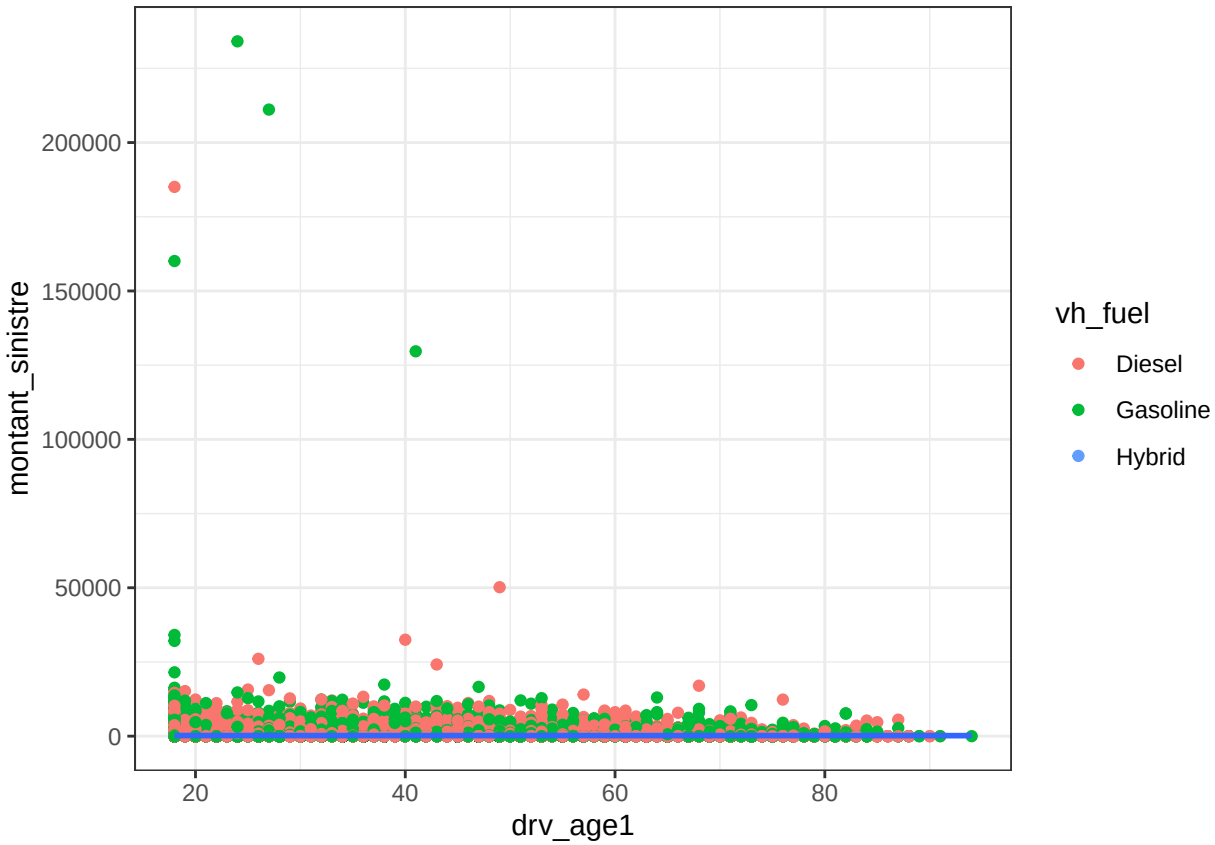

```
ggplot(base_etude,aes(y=pol_bonus,x=nbr_sin))+geom_point()+geom_smooth(method = "lm",se=T)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



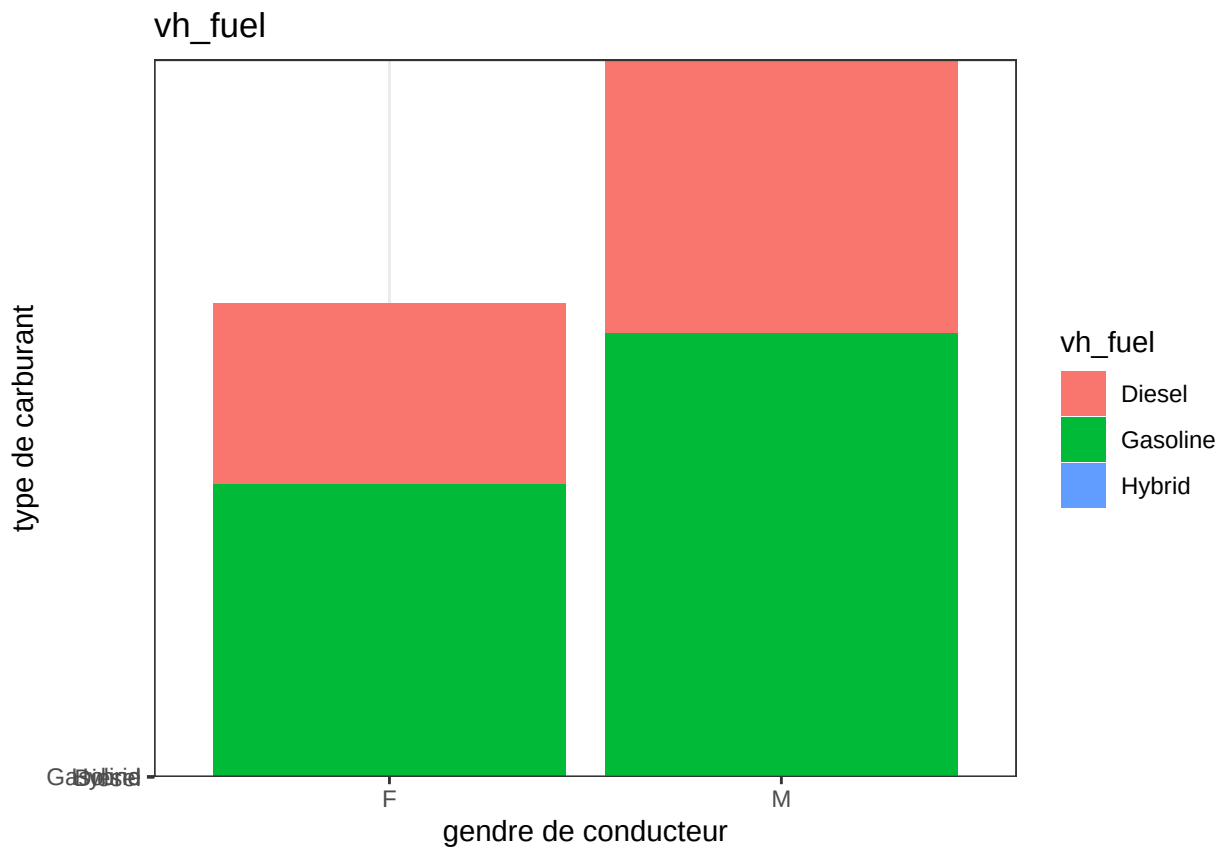

```
### visualisation des variables qualitative/variable qualitative
ggplot(base_etude,aes(x=drv_age1,y=montant_sinistre))+geom_point(aes(col=vh_fuel))+geom_smooth(method="
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

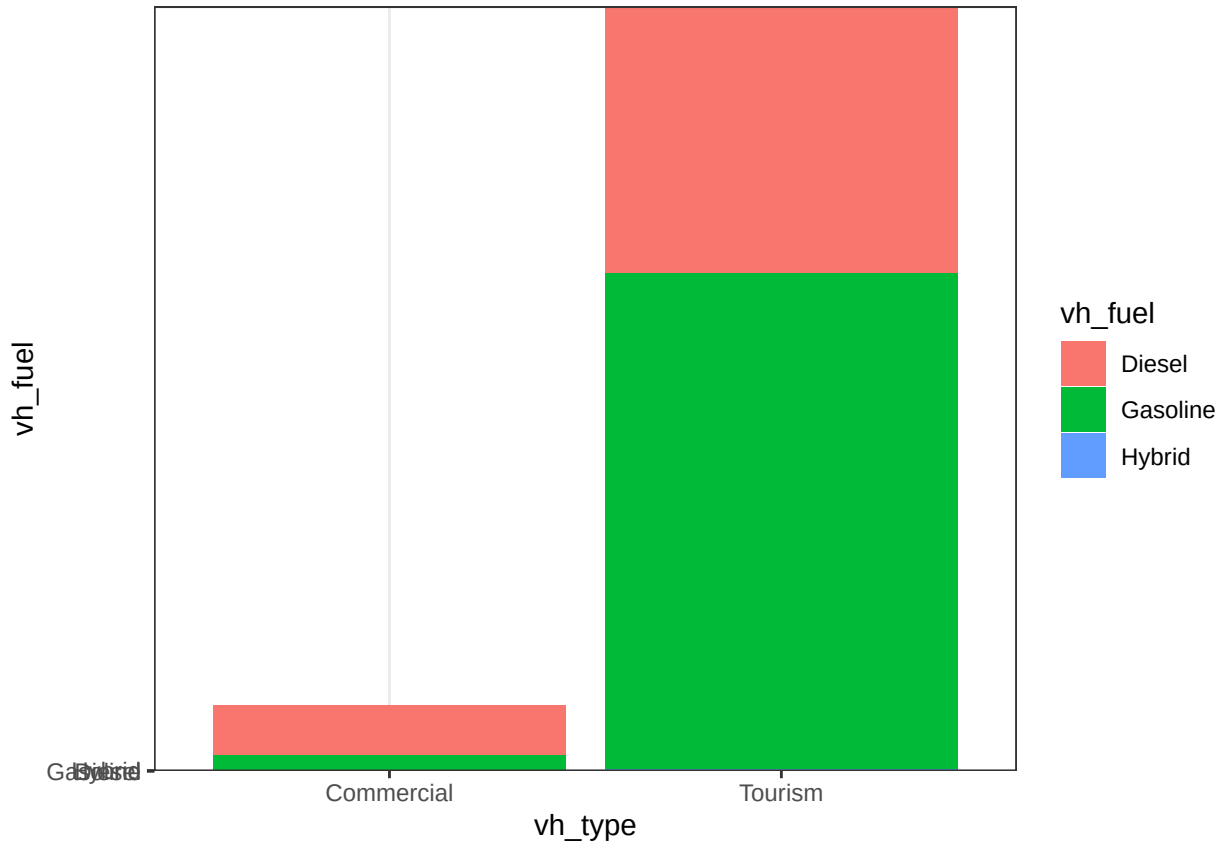



```
ggplot(base_etude,aes(x=drv_sex1,y=vh_fuel))+geom_bar(stat="identity",aes(fill=vh_fuel),cex=0.2)+labs(t
```

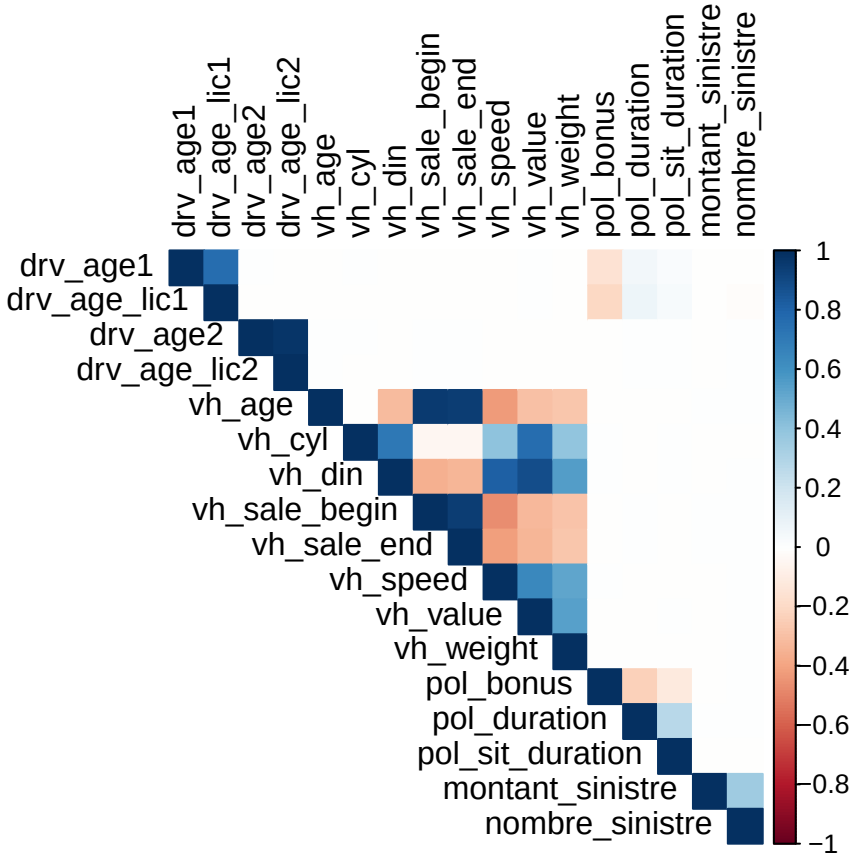
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```




```
ggplot(base_etude,aes(x=vh_type,y=vh_fuel))+geom_bar(stat="identity",aes(fill=vh_fuel),cex=0.2)
```

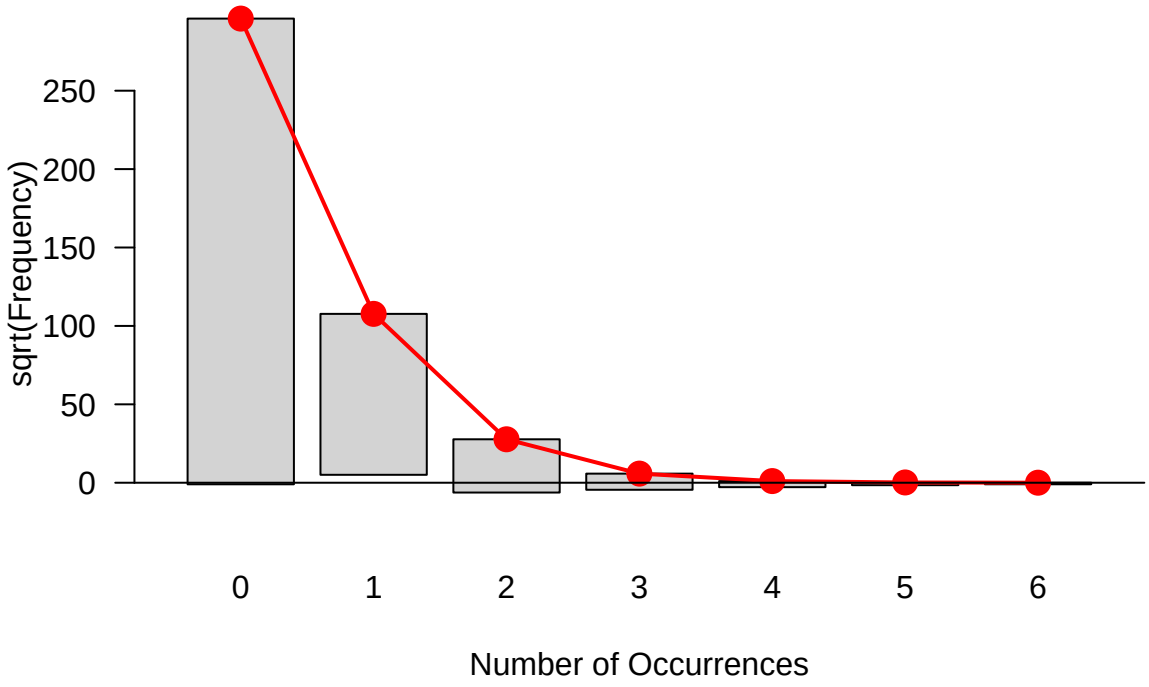



```
################## etude correlation entre variables quantitative##################
var_quantitaive=base_etude %>% select_if("is.numeric")
df_cor=cor(var_quantitaive)
corrplot(df_cor,method = "color",type="upper",tl.col = "black",number.cex = 0.9,number.digits = 3)
```

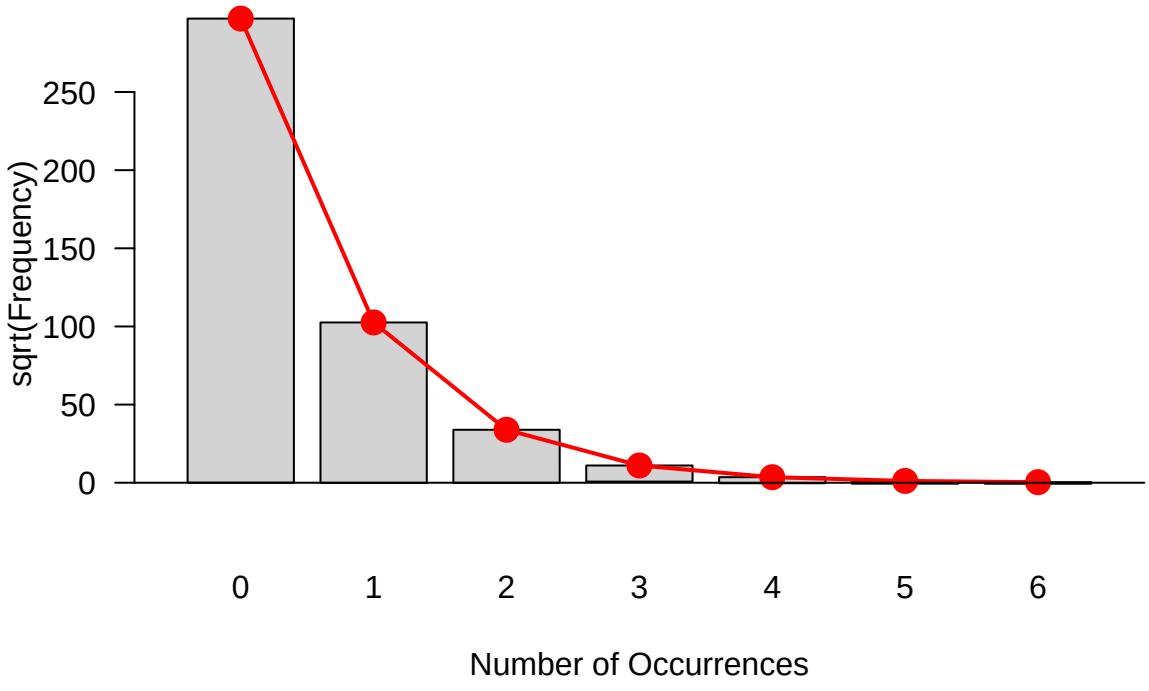



```
######## la partie recodage des variables############
#pol_usage
base_etude$pol_usage2=fct_recode(base_etude$pol_usage,"Professional"="AllTrips")
base_etude$vh_fuel1=fct_recode(base_etude$vh_fuel,"Diesel"="Gasoline")
base_etude$pol_duration2=cut(base_etude$pol_duration,c(-Inf,21,Inf))
base_etude$vh_age22=cut(base_etude$vh_age,c(-Inf,17,Inf))
base_etude$drv_age11=cut(base_etude$drv_age1,c(-Inf,18,22,26,30,35,40,45,50,60,Inf))
base_etude$vh_make2=as.character(base_etude$vh_make)
base_etude$vh_make2[!(base_etude$vh_make %in% c("RENAULT","PEUGEOT","CITROEN","VOLKSWAGEN","FORD","OPEL
OYOTA","MERCEDES BENZ","FIAT"))]="autre_vh"
base_etude$vh_make2=as.factor(base_etude$vh_make2)

############## verifier la proprité de la poisson et la loi binomial negative
#### pour verfier la ajustement des observation avec ces lois
## loi de poisson
base_etude$nombre_sinistre=as.numeric(base_etude$nombre_sinistre)
poiss=goodfit(base_etude$nombre_sinistre,type = c("poisson"),method =c("ML","MinChisq"),par=NULL)
plot(poiss)
```




```
binom=goodfit(base_etude$nombre_sinistre,type = c("nbinomial"),method =c("ML","MinChisq"),par=NULL)
plot(binom)
```




```r
mean(base_etude$nombre_sinistre)
```

```
## [1] 0.13238
```

```r
var(base_etude$nombre_sinistre)
```

```
## [1] 0.146977
```

```r
############ interpretation des graphique goodfit pour 2 lois################
####### : les points rouges représentent
#la loi théorique et les histogrammes les fréquences observées, qui sont collés par le sommet à la loi
#théorique. Tout écart de la base d'un histogramme avec l'axe des abscisses indique donc un mauvais
#ajustement des observations par la loi théorique. On remarque ainsi que l'ajustement par la loi de
#Poisson est peu satisfaisant. En effet les écarts observés sont conséquents et cela quel que soit le
#nombre d'occurrences.
# d'apres ces graphique on constate que la loi binomial une bonne alternative car on voit des ecarts
# entre les histogramme et les absisses cela nous raméne une mauvise ajustement entre les observation e

######### partie modélisation###############
#### glm regréssion poissonienne
base_etude$nombre_sinistre=as.numeric(base_etude$nombre_sinistre)
model1=glm(formula = nombre_sinistre~drv_age2+drv_age_lic1+drv_sex1+drv_age_lic2+drv_sex2+drv_drv2
           +vh_age22+vh_cyl+vh_din+vh_sale_begin+vh_sale_end+vh_speed+vh_type+vh_value
           +pol_bonus+pol_coverage+pol_sit_duration+pol_pay_freq+pol_payd+pol_usage2+vh_fuel1+pol_dura
```

```r
#model1 %>% tbl_regression(intercept = TRUE,exponentiate = TRUE) %>% add_global_p()
summary(model1)
```

```
##
## Call:
## glm(formula = nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + pol_usage2 + vh_fuel1 + pol_duration2 +
##     drv_age11 + vh_make2, family = poisson("log"), data = base_etude)
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -2.148e+00  1.313e-01 -16.363  < 2e-16 ***
## drv_age2               -5.515e-03  1.082e-02  -0.510 0.610241
## drv_age_lic1           -5.095e-03  1.091e-03  -4.671    3e-06 ***
## drv_sex1M              -2.461e-02  1.774e-02  -1.387 0.165393
## drv_age_lic2            5.539e-03  1.081e-02   0.512 0.608526
## drv_sex2F               1.410e-02  3.170e-01   0.044 0.964528
## drv_sex2M              -8.717e-03  3.168e-01  -0.028 0.978048
## drv_drv2Yes             1.179e-01  3.848e-01   0.306 0.759273
## vh_age22(17, Inf]       4.454e-02  3.911e-02   1.139 0.254697
## vh_cyl                 -3.728e-06  3.493e-05  -0.107 0.915012
## vh_din                 -1.159e-03  8.712e-04  -1.330 0.183367
## vh_sale_begin          -1.394e-04  3.934e-03  -0.035 0.971737
## vh_sale_end             1.823e-03  4.336e-03   0.420 0.674192
## vh_speed                1.282e-03  8.931e-04   1.436 0.151135
## vh_typeTourism          3.174e-02  3.809e-02   0.833 0.404703
## vh_value                2.314e-06  2.585e-06   0.895 0.370778
## pol_bonus               1.250e-02  9.393e-02   0.133 0.894129
## pol_coverageMedian1    -5.057e-02  3.138e-02  -1.612 0.106995
## pol_coverageMedian2     5.279e-02  2.326e-02   2.270 0.023234 *
## pol_coverageMini       -5.240e-03  3.225e-02  -0.163 0.870900
## pol_sit_duration       -2.476e-03  3.908e-03  -0.634 0.526250
## pol_pay_freqMonthly    -2.180e-02  2.290e-02  -0.952 0.341139
## pol_pay_freqQuarterly   4.746e-02  5.549e-02   0.855 0.392333
## pol_pay_freqYearly     -3.574e-02  2.148e-02  -1.664 0.096051 .
## pol_paydYes            -5.567e-02  4.583e-02  -1.215 0.224518
## pol_usage2Retired      -2.610e-02  3.653e-02  -0.715 0.474905
## pol_usage2WorkPrivate  -2.131e-02  3.368e-02  -0.633 0.526970
## vh_fuel1Hybrid          4.066e-01  2.523e-01   1.612 0.107033
## pol_duration2(21, Inf]  5.801e-02  2.394e-02   2.424 0.015364 *
## drv_age11(18,22]        1.570e-01  4.535e-02   3.462 0.000535 ***
## drv_age11(22,26]        1.355e-01  4.608e-02   2.940 0.003282 **
## drv_age11(26,30]        1.702e-01  4.691e-02   3.629 0.000285 ***
## drv_age11(30,35]        9.883e-02  4.717e-02   2.095 0.036149 *
## drv_age11(35,40]        1.821e-01  4.920e-02   3.702 0.000214 ***
## drv_age11(40,45]        1.440e-01  5.352e-02   2.691 0.007134 **
## drv_age11(45,50]        1.432e-01  5.763e-02   2.485 0.012946 *
## drv_age11(50,60]        1.278e-01  5.788e-02   2.208 0.027210 *
## drv_age11(60, Inf]      1.623e-01  6.854e-02   2.367 0.017920 *
## vh_make2CITROEN        -1.601e-02  3.184e-02  -0.503 0.615073
```

```
## vh_make2FIAT           -1.590e-02  5.611e-02  -0.283 0.776930
## vh_make2FORD            7.501e-03  4.676e-02   0.160 0.872552
## vh_make2MERCEDES BENZ   6.269e-02  5.499e-02   1.140 0.254245
## vh_make2OPEL           -7.527e-02  5.217e-02  -1.443 0.149083
## vh_make2PEUGEOT        -3.882e-02  3.039e-02  -1.277 0.201530
## vh_make2RENAULT        -8.420e-03  2.819e-02  -0.299 0.765179
## vh_make2VOLKSWAGEN     -4.045e-02  4.393e-02  -0.921 0.357198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 57668  on 99999  degrees of freedom
## Residual deviance: 57585  on 99954  degrees of freedom
## AIC: 82137
##
## Number of Fisher Scoring iterations: 6
```

```
####### qualite de l'ajustement############
# statistique de pearson##
np=(base_etude$nombre_sinistre-model1$fitted.values)^2/model1$fitted.values
total_stat=sum(np)
print(total_stat)
```

```
## [1] 110945.6
```

```
### teste de probabilité#
print(pchisq(total_stat,model1$df.residual,lower.tail = FALSE))
```

```
## [1] 7.375387e-125
```

```
####### on constate que notre model est mal ajusté car p est largement inférieur alpha

############ selection des variables en fonction  de criter de aic
#stepAIC(model1,direction ="backward",k = log(nrow(base_etude)))
mode2=step(model1)
```

```
## Start:  AIC=82136.78
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_sex2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     pol_usage2 + vh_fuel1 + pol_duration2 + drv_age11 + vh_make2
##
##                    Df Deviance   AIC
## - vh_make2          8    57592 82127
## - pol_usage2        2    57585 82133
## - drv_sex2          2    57586 82133
## - vh_sale_begin     1    57585 82135
## - vh_cyl            1    57585 82135
## - pol_bonus         1    57585 82135
## - drv_drv2          1    57585 82135
```

```
## - vh_sale_end       1    57585 82135
## - drv_age2          1    57585 82135
## - drv_age_lic2      1    57585 82135
## - pol_pay_freq      3    57589 82135
## - pol_sit_duration  1    57585 82135
## - vh_type           1    57586 82135
## - vh_value          1    57586 82136
## - vh_age22          1    57586 82136
## - pol_payd          1    57586 82136
## - vh_din            1    57587 82137
## - drv_sex1          1    57587 82137
## <none>                   57585 82137
## - vh_speed          1    57587 82137
## - vh_fuel1          1    57587 82137
## - drv_age11         9    57606 82140
## - pol_coverage      3    57594 82140
## - pol_duration2     1    57591 82141
## - drv_age_lic1      1    57607 82157
##
## Step:  AIC=82127.36
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_sex2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     pol_usage2 + vh_fuel1 + pol_duration2 + drv_age11
##
##                     Df Deviance   AIC
## - pol_usage2         2    57592 82124
## - drv_sex2           2    57592 82124
## - vh_sale_begin      1    57592 82125
## - pol_bonus          1    57592 82125
## - vh_cyl             1    57592 82125
## - drv_drv2           1    57592 82125
## - vh_sale_end        1    57592 82126
## - drv_age2           1    57592 82126
## - drv_age_lic2       1    57592 82126
## - pol_pay_freq       3    57596 82126
## - pol_sit_duration   1    57592 82126
## - vh_type            1    57592 82126
## - vh_age22           1    57593 82127
## - vh_value           1    57593 82127
## - pol_payd           1    57593 82127
## - vh_din             1    57593 82127
## - vh_speed           1    57593 82127
## - drv_sex1           1    57593 82127
## <none>                    57592 82127
## - vh_fuel1           1    57594 82128
## - drv_age11          9    57613 82130
## - pol_coverage       3    57601 82131
## - pol_duration2      1    57597 82131
## - drv_age_lic1       1    57614 82147
##
## Step:  AIC=82123.88
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
```

```
##     drv_sex2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     vh_fuel1 + pol_duration2 + drv_age11
##
##                     Df Deviance   AIC
## - drv_sex2           2    57593 82120
## - vh_sale_begin      1    57592 82122
## - pol_bonus          1    57592 82122
## - vh_cyl             1    57592 82122
## - drv_drv2           1    57592 82122
## - vh_sale_end        1    57592 82122
## - drv_age2           1    57592 82122
## - drv_age_lic2       1    57592 82122
## - pol_sit_duration   1    57593 82122
## - pol_pay_freq       3    57597 82122
## - vh_type            1    57593 82123
## - vh_age22           1    57593 82123
## - vh_value           1    57593 82123
## - vh_din             1    57594 82123
## - vh_speed           1    57594 82124
## - pol_payd           1    57594 82124
## - drv_sex1           1    57594 82124
## <none>                    57592 82124
## - vh_fuel1           1    57594 82124
## - drv_age11          9    57613 82127
## - pol_coverage       3    57602 82127
## - pol_duration2      1    57598 82128
## - drv_age_lic1       1    57614 82144
##
## Step:  AIC=82120.42
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin + vh_sale_end +
##     vh_speed + vh_type + vh_value + pol_bonus + pol_coverage +
##     pol_sit_duration + pol_pay_freq + pol_payd + vh_fuel1 + pol_duration2 +
##     drv_age11
##
##                     Df Deviance   AIC
## - vh_sale_begin      1    57593 82118
## - pol_bonus          1    57593 82118
## - vh_cyl             1    57593 82118
## - vh_sale_end        1    57593 82119
## - drv_age2           1    57593 82119
## - drv_age_lic2       1    57593 82119
## - drv_drv2           1    57593 82119
## - pol_sit_duration   1    57593 82119
## - pol_pay_freq       3    57597 82119
## - vh_type            1    57593 82119
## - vh_age22           1    57594 82120
## - vh_value           1    57594 82120
## - vh_din             1    57594 82120
## - vh_speed           1    57594 82120
## - pol_payd           1    57594 82120
## - drv_sex1           1    57594 82120
```

```
## <none>                   57593 82120
## - vh_fuel1          1    57595 82121
## - drv_age11         9    57614 82124
## - pol_coverage      3    57602 82124
## - pol_duration2     1    57598 82124
## - drv_age_lic1      1    57615 82140
##
## Step:  AIC=82118.42
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end + vh_speed +
##     vh_type + vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                     Df Deviance   AIC
## - pol_bonus          1    57593 82116
## - vh_cyl             1    57593 82116
## - drv_age2           1    57593 82117
## - drv_age_lic2       1    57593 82117
## - drv_drv2           1    57593 82117
## - pol_sit_duration   1    57593 82117
## - pol_pay_freq       3    57597 82117
## - vh_type            1    57593 82117
## - vh_sale_end        1    57593 82117
## - vh_value           1    57594 82118
## - vh_age22           1    57594 82118
## - vh_din             1    57594 82118
## - vh_speed           1    57594 82118
## - pol_payd           1    57594 82118
## - drv_sex1           1    57594 82118
## <none>                    57593 82118
## - vh_fuel1           1    57595 82119
## - drv_age11          9    57614 82122
## - pol_coverage       3    57602 82122
## - pol_duration2      1    57598 82122
## - drv_age_lic1       1    57615 82138
##
## Step:  AIC=82116.45
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end + vh_speed +
##     vh_type + vh_value + pol_coverage + pol_sit_duration + pol_pay_freq +
##     pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                     Df Deviance   AIC
## - vh_cyl             1    57593 82114
## - drv_age2           1    57593 82115
## - drv_age_lic2       1    57593 82115
## - drv_drv2           1    57593 82115
## - pol_sit_duration   1    57593 82115
## - pol_pay_freq       3    57597 82115
## - vh_type            1    57593 82115
## - vh_sale_end        1    57594 82115
## - vh_value           1    57594 82116
## - vh_age22           1    57594 82116
## - vh_din             1    57594 82116
```

```
## - vh_speed          1    57594 82116
## - pol_payd          1    57594 82116
## - drv_sex1          1    57594 82116
## <none>                   57593 82116
## - vh_fuel1          1    57595 82117
## - pol_coverage      3    57602 82120
## - drv_age11         9    57614 82120
## - pol_duration2     1    57598 82120
## - drv_age_lic1      1    57616 82137
##
## Step:  AIC=82114.48
## nombre_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_drv2 + vh_age22 + vh_din + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_coverage + pol_sit_duration + pol_pay_freq +
##     pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                    Df Deviance   AIC
## - drv_age2          1    57593 82113
## - drv_age_lic2      1    57593 82113
## - drv_drv2          1    57593 82113
## - pol_sit_duration  1    57593 82113
## - pol_pay_freq      3    57597 82113
## - vh_sale_end       1    57594 82113
## - vh_type           1    57594 82113
## - vh_value          1    57594 82114
## - vh_age22          1    57594 82114
## - pol_payd          1    57594 82114
## - vh_speed          1    57594 82114
## - vh_din            1    57594 82114
## - drv_sex1          1    57595 82114
## <none>                   57593 82114
## - vh_fuel1          1    57595 82115
## - pol_coverage      3    57602 82118
## - drv_age11         9    57614 82118
## - pol_duration2     1    57598 82118
## - drv_age_lic1      1    57616 82135
##
## Step:  AIC=82112.75
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + drv_age_lic2 + drv_drv2 +
##     vh_age22 + vh_din + vh_sale_end + vh_speed + vh_type + vh_value +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     vh_fuel1 + pol_duration2 + drv_age11
##
##                    Df Deviance   AIC
## - drv_age_lic2      1    57593 82111
## - drv_drv2          1    57593 82111
## - pol_sit_duration  1    57593 82111
## - pol_pay_freq      3    57597 82111
## - vh_sale_end       1    57594 82112
## - vh_type           1    57594 82112
## - vh_value          1    57594 82112
## - vh_age22          1    57594 82112
## - pol_payd          1    57595 82112
## - vh_speed          1    57595 82113
```

```
## - vh_din            1    57595 82113
## - drv_sex1          1    57595 82113
## <none>                   57593 82113
## - vh_fuel1          1    57595 82113
## - pol_coverage      3    57602 82116
## - drv_age11         9    57614 82116
## - pol_duration2     1    57599 82116
## - drv_age_lic1      1    57616 82134
##
## Step:  AIC=82110.76
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + drv_drv2 + vh_age22 +
##     vh_din + vh_sale_end + vh_speed + vh_type + vh_value + pol_coverage +
##     pol_sit_duration + pol_pay_freq + pol_payd + vh_fuel1 + pol_duration2 +
##     drv_age11
##
##                    Df Deviance   AIC
## - pol_sit_duration  1    57593 82109
## - pol_pay_freq      3    57597 82109
## - drv_drv2          1    57593 82109
## - vh_sale_end       1    57594 82110
## - vh_type           1    57594 82110
## - vh_value          1    57594 82110
## - vh_age22          1    57594 82110
## - pol_payd          1    57595 82110
## - vh_speed          1    57595 82111
## - vh_din            1    57595 82111
## - drv_sex1          1    57595 82111
## <none>                   57593 82111
## - vh_fuel1          1    57595 82111
## - pol_coverage      3    57602 82114
## - drv_age11         9    57614 82114
## - pol_duration2     1    57599 82114
## - drv_age_lic1      1    57616 82132
##
## Step:  AIC=82109.24
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + drv_drv2 + vh_age22 +
##     vh_din + vh_sale_end + vh_speed + vh_type + vh_value + pol_coverage +
##     pol_pay_freq + pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - drv_drv2       1    57594 82108
## - pol_pay_freq   3    57598 82108
## - vh_type        1    57594 82108
## - vh_sale_end    1    57594 82108
## - pol_payd       1    57595 82109
## - vh_value       1    57595 82109
## - vh_age22       1    57595 82109
## - vh_speed       1    57595 82109
## - vh_din         1    57595 82109
## - drv_sex1       1    57595 82109
## <none>                57593 82109
## - vh_fuel1       1    57596 82110
## - pol_coverage   3    57603 82113
## - pol_duration2  1    57599 82113
```

```
## - drv_age11      9    57615 82113
## - drv_age_lic1   1    57617 82130
##
## Step:  AIC=82107.79
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_din +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_coverage +
##     pol_pay_freq + pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - pol_pay_freq   3    57599 82106
## - vh_sale_end    1    57595 82107
## - vh_type        1    57595 82107
## - pol_payd       1    57595 82107
## - vh_value       1    57595 82107
## - vh_age22       1    57596 82107
## - vh_speed       1    57596 82108
## - vh_din         1    57596 82108
## - drv_sex1       1    57596 82108
## <none>                57594 82108
## - vh_fuel1       1    57596 82108
## - pol_coverage   3    57603 82111
## - pol_duration2  1    57599 82111
## - drv_age11      9    57616 82111
## - drv_age_lic1   1    57617 82129
##
## Step:  AIC=82106.39
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_din +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_coverage +
##     pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - vh_sale_end    1    57599 82105
## - vh_type        1    57599 82105
## - vh_value       1    57600 82106
## - vh_age22       1    57600 82106
## - pol_payd       1    57600 82106
## - vh_speed       1    57600 82106
## - vh_din         1    57600 82106
## - drv_sex1       1    57600 82106
## <none>                57599 82106
## - vh_fuel1       1    57601 82107
## - pol_coverage   3    57608 82110
## - pol_duration2  1    57604 82110
## - drv_age11      9    57620 82110
## - drv_age_lic1   1    57622 82128
##
## Step:  AIC=82105.24
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_din +
##     vh_speed + vh_type + vh_value + pol_coverage + pol_payd +
##     vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - vh_type        1    57600 82104
## - vh_value       1    57601 82104
```

```
## - vh_speed       1    57601 82105
## - vh_din         1    57601 82105
## - pol_payd       1    57601 82105
## - drv_sex1       1    57601 82105
## <none>                57599 82105
## - vh_fuel1       1    57602 82106
## - pol_coverage   3    57609 82109
## - pol_duration2  1    57605 82109
## - vh_age22       1    57605 82109
## - drv_age11      9    57621 82109
## - drv_age_lic1   1    57623 82127
##
## Step:  AIC=82104.17
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_din +
##     vh_speed + vh_value + pol_coverage + pol_payd + vh_fuel1 +
##     pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - vh_value       1    57601 82103
## - pol_payd       1    57602 82104
## - drv_sex1       1    57602 82104
## <none>                57600 82104
## - vh_din         1    57603 82104
## - vh_fuel1       1    57603 82105
## - vh_speed       1    57604 82106
## - pol_coverage   3    57610 82107
## - pol_duration2  1    57606 82108
## - drv_age11      9    57622 82108
## - vh_age22       1    57606 82108
## - drv_age_lic1   1    57624 82126
##
## Step:  AIC=82103.29
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_din +
##     vh_speed + pol_coverage + pol_payd + vh_fuel1 + pol_duration2 +
##     drv_age11
##
##                 Df Deviance   AIC
## - vh_din         1    57603 82103
## - pol_payd       1    57603 82103
## - drv_sex1       1    57603 82103
## <none>                57601 82103
## - vh_fuel1       1    57604 82104
## - vh_speed       1    57605 82104
## - pol_coverage   3    57611 82107
## - vh_age22       1    57607 82107
## - pol_duration2  1    57607 82107
## - drv_age11      9    57623 82107
## - drv_age_lic1   1    57625 82125
##
## Step:  AIC=82102.57
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_speed +
##     pol_coverage + pol_payd + vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
```

```
## - pol_payd       1    57604 82102
## - drv_sex1       1    57605 82102
## <none>                57603 82103
## - vh_speed       1    57605 82103
## - vh_fuel1       1    57605 82103
## - vh_age22       1    57608 82106
## - pol_coverage   3    57612 82106
## - pol_duration2  1    57608 82106
## - drv_age11      9    57624 82106
## - drv_age_lic1   1    57626 82124
##
## Step:  AIC=82102.29
## nombre_sinistre ~ drv_age_lic1 + drv_sex1 + vh_age22 + vh_speed +
##     pol_coverage + vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## - drv_sex1       1    57606 82102
## <none>                57604 82102
## - vh_speed       1    57607 82102
## - vh_fuel1       1    57607 82103
## - vh_age22       1    57609 82105
## - pol_coverage   3    57614 82106
## - pol_duration2  1    57610 82106
## - drv_age11      9    57626 82106
## - drv_age_lic1   1    57628 82124
##
## Step:  AIC=82102.18
## nombre_sinistre ~ drv_age_lic1 + vh_age22 + vh_speed + pol_coverage +
##     vh_fuel1 + pol_duration2 + drv_age11
##
##                 Df Deviance   AIC
## <none>                57606 82102
## - vh_speed       1    57608 82102
## - vh_fuel1       1    57609 82103
## - vh_age22       1    57611 82105
## - pol_coverage   3    57616 82105
## - pol_duration2  1    57612 82105
## - drv_age11      9    57628 82106
## - drv_age_lic1   1    57630 82124
```

```
summary(mode2)
```

```
##
## Call:
## glm(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + vh_speed +
##     pol_coverage + vh_fuel1 + pol_duration2 + drv_age11, family = poisson("log"),
##     data = base_etude)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -2.1234316  0.0708441 -29.973  < 2e-16 ***
## drv_age_lic1           -0.0052340  0.0010718  -4.883 1.04e-06 ***
## vh_age22(17, Inf]       0.0633549  0.0283442   2.235 0.025404 *
## vh_speed                0.0005777  0.0003957   1.460 0.144275
```

```
## pol_coverageMedian1    -0.0496553  0.0312789  -1.588 0.112399
## pol_coverageMedian2     0.0526884  0.0230870   2.282 0.022479 *
## pol_coverageMini       -0.0070349  0.0319557  -0.220 0.825757
## vh_fuel1Hybrid          0.4304714  0.2502040   1.720 0.085345 .
## pol_duration2(21, Inf]  0.0532983  0.0229792   2.319 0.020373 *
## drv_age11(18,22]        0.1604506  0.0449963   3.566 0.000363 ***
## drv_age11(22,26]        0.1393761  0.0457622   3.046 0.002322 **
## drv_age11(26,30]        0.1738288  0.0466947   3.723 0.000197 ***
## drv_age11(30,35]        0.1022269  0.0469946   2.175 0.029608 *
## drv_age11(35,40]        0.1851142  0.0490531   3.774 0.000161 ***
## drv_age11(40,45]        0.1469617  0.0533522   2.755 0.005877 **
## drv_age11(45,50]        0.1463351  0.0574655   2.546 0.010881 *
## drv_age11(50,60]        0.1313750  0.0576737   2.278 0.022732 *
## drv_age11(60, Inf]      0.1658460  0.0682452   2.430 0.015093 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 57668  on 99999  degrees of freedom
## Residual deviance: 57606  on 99982  degrees of freedom
## AIC: 82102
##
## Number of Fisher Scoring iterations: 6
```

```
npp=(base_etude$nombre_sinistre-mode2$fitted.values)^2/mode2$fitted.values
total_staté=sum(npp)
print(pchisq(total_stat,mode2$df.residual,lower.tail = FALSE))
```

```
## [1] 3.180053e-124
```

```
######## statistique deviance############
y=base_etude$nombre_sinistre
res_dev=ifelse(y==0,0,y*log(y/mode2$fitted.values))-(y-mode2$fitted.values)
devaince=2*sum(res_dev)
print(devaince)
```

```
## [1] 57606.35
```

```
### probabilite de statistique
print(pchisq(devaince,mode2$df.residual,lower.tail = FALSE))
```

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

```
############## calcule pseaudo R ################
### calcule la deviance totale de model
a0=log(mean(base_etude$nombre_sinistre))
dev_total=ifelse(y==0,0,y*log(y))-y*a0-(y-exp(a0))
d0=2*sum(dev_total)
print(d0)
```

```
## [1] 57668.21
```

```r
r=(d0-devaince)/d0
print(r)
```

```
## [1] 0.001072605
```

```r
######### interval de confiance des coeficeints####
#Les intervalles de confiance seront affichés sous forme de table,
#fournissant des limites inférieures et supérieures pour chaque coefficient du modèle.
#Vous pouvez interpréter ces intervalles de confiance pour évaluer
#la signification statistique des coefficients estimés.
#Si l'intervalle inclut zéro, le coefficient n'est pas significatif à un niveau de confiance de 95%.
#Sinon, vous avez une indication de la significativité du coefficient.



md=summary(mode2)
coef_md=md$coefficients[,1]
sqrt_md=md$coefficients[,2]
u=qnorm(0.95)
print("borne basse")
```

```
## [1] "borne basse"
```

```r
print(coef_md-u*sqrt_md)
```

```
##            (Intercept)           drv_age_lic1      vh_age22(17, Inf] 
##          -2.239960e+00          -6.997029e-03           1.673282e-02 
##               vh_speed    pol_coverageMedian1    pol_coverageMedian2 
##          -7.311719e-05          -1.011045e-01           1.471372e-02 
##       pol_coverageMini         vh_fuel1Hybrid pol_duration2(21, Inf] 
##          -5.959736e-02           1.892250e-02           1.550083e-02 
##       drv_age11(18,22]       drv_age11(22,26]       drv_age11(26,30] 
##           8.643832e-02           6.410396e-02           9.702283e-02 
##       drv_age11(30,35]       drv_age11(35,40]       drv_age11(40,45] 
##           2.492774e-02           1.044291e-01           5.920518e-02 
##       drv_age11(45,50]       drv_age11(50,60]     drv_age11(60, Inf] 
##           5.181283e-02           3.651025e-02           5.359253e-02
```

```r
print("borne hausse")
```

```
## [1] "borne hausse"
```

```r
print(coef_md+u*sqrt_md)
```

```
##            (Intercept)           drv_age_lic1      vh_age22(17, Inf] 
##          -2.006903489          -0.003471038           0.109976985 
##               vh_speed    pol_coverageMedian1    pol_coverageMedian2 
##           0.001228477           0.001793853           0.090663115 
##       pol_coverageMini         vh_fuel1Hybrid pol_duration2(21, Inf] 
##           0.045527496           0.842020359           0.091095694
```

```
##       drv_age11(18,22]       drv_age11(22,26]       drv_age11(26,30]
##            0.234462810            0.214648329            0.250634846
##       drv_age11(30,35]       drv_age11(35,40]       drv_age11(40,45]
##            0.179526116            0.265799356            0.234718299
##       drv_age11(45,50]       drv_age11(50,60]     drv_age11(60, Inf]
##            0.240857448            0.226239749            0.278099397
```

```
library(MASS)
print(confint.default(mode2,level = 0.90))
```

```
##                                 5 %         95 %
## (Intercept)            -2.239960e+00 -2.006903489
## drv_age_lic1           -6.997029e-03 -0.003471038
## vh_age22(17, Inf]       1.673282e-02  0.109976985
## vh_speed               -7.311719e-05  0.001228477
## pol_coverageMedian1    -1.011045e-01  0.001793853
## pol_coverageMedian2     1.471372e-02  0.090663115
## pol_coverageMini       -5.959736e-02  0.045527496
## vh_fuel1Hybrid          1.892250e-02  0.842020359
## pol_duration2(21, Inf]  1.550083e-02  0.091095694
## drv_age11(18,22]        8.643832e-02  0.234462810
## drv_age11(22,26]        6.410396e-02  0.214648329
## drv_age11(26,30]        9.702283e-02  0.250634846
## drv_age11(30,35]        2.492774e-02  0.179526116
## drv_age11(35,40]        1.044291e-01  0.265799356
## drv_age11(40,45]        5.920518e-02  0.234718299
## drv_age11(45,50]        5.181283e-02  0.240857448
## drv_age11(50,60]        3.651025e-02  0.226239749
## drv_age11(60, Inf]      5.359253e-02  0.278099397
```

```
######## verification hypotheses la variance egale la moyenne #########

# Le test de dispersion en régression de Poisson, également appelé dispersion test,
#évalue si la variance des résidus dans un modèle de régression de Poisson est constante. Dans le contes
#le modèle suppose que la variance de la variable dépendante (qui suit une distribution de Poisson) est
#il peut arriver que la variance soit différente de la moyenne, ce qui peut indiquer une surdispersion

# alors si p est inferieure a 0.05 ce suggere une surdispersion
#Si la valeur de p est grande, cela indique que le modèle de Poisson pourrait être adéquat en termes de
 dispersiontest(mode2)
```

```
##
##  Overdispersion test
##
## data:  mode2
## z = 14.863, p-value < 2.2e-16
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
##   1.109974
```

```
############  cela peut suggérer que le modèle de Poisson standard pourrait ne pas être approprié,
#et des alternatives comme la régression de binomial negative négative pourraient être considérées.


 ######## regression de loi binomial negative############
phi=0.97
Mod_bino=glm(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22
               + pol_duration2 + drv_age11, family =negative.binomial(phi), data = base_etude)
summary(Mod_bino)
```

```
##
## Call:
## glm(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_duration2 +
##     drv_age11, family = negative.binomial(phi), data = base_etude)
##
## Coefficients:
##                          Estimate Std. Error  t value Pr(>|t|)
## (Intercept)            -2.018520   0.015910 -126.870  < 2e-16 ***
## drv_age_lic1           -0.005246   0.001128   -4.651 3.31e-06 ***
## vh_age22(17, Inf]       0.047765   0.027873    1.714 0.086595 .
## pol_duration2(21, Inf]  0.055257   0.024160    2.287 0.022191 *
## drv_age11(18,22]        0.161020   0.047507    3.389 0.000701 ***
## drv_age11(22,26]        0.140151   0.048275    2.903 0.003695 **
## drv_age11(26,30]        0.174106   0.049301    3.531 0.000413 ***
## drv_age11(30,35]        0.102444   0.049502    2.069 0.038503 *
## drv_age11(35,40]        0.185099   0.051754    3.577 0.000348 ***
## drv_age11(40,45]        0.146729   0.056207    2.611 0.009042 **
## drv_age11(45,50]        0.146821   0.060505    2.427 0.015244 *
## drv_age11(50,60]        0.132342   0.060684    2.181 0.029197 *
## drv_age11(60, Inf]      0.165897   0.071774    2.311 0.020813 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.97) family taken to be 0.9770698)
##
##     Null deviance: 47285  on 99999  degrees of freedom
## Residual deviance: 47243  on 99987  degrees of freedom
## AIC: 81681
##
## Number of Fisher Scoring iterations: 4
```

```
Mod_bino %>% tbl_regression(intercept = TRUE,exponentiate = TRUE) %>% add_global_p() %>% add_n()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	N	exp(Beta)	95% CI	p-value
(Intercept)	100,000	0.13	0.13, 0.14	<0.001
drv_age_lic1	100,000	1.0	0.99, 1.00	<0.001

Characteristic	N	exp(Beta)	95% CI	p-value
vh_age22	100,000			0.088
(-Inf,17]		—	—	
(17, Inf]		1.05	0.99, 1.11	
pol_duration2	100,000			0.023
(-Inf,21]		—	—	
(21, Inf]		1.06	1.01, 1.11	
drv_age11	100,000			0.018
(-Inf,18]		—	—	
(18,22]		1.17	1.07, 1.29	
(22,26]		1.15	1.05, 1.26	
(26,30]		1.19	1.08, 1.31	
(30,35]		1.11	1.01, 1.22	
(35,40]		1.20	1.09, 1.33	
(40,45]		1.16	1.04, 1.29	
(45,50]		1.16	1.03, 1.30	
(50,60]		1.14	1.01, 1.29	
(60, Inf]		1.18	1.03, 1.36	

```
####### binomial negative avec theta fixé par mass
mod_binm2=glm.nb(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + vh_speed
             +pol_coverage + vh_fuel1 + pol_duration2 + drv_age11, data = base_etude)
summary(mod_binm2)
```

```
##
## Call:
## glm.nb(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 +
##     vh_speed + pol_coverage + vh_fuel1 + pol_duration2 + drv_age11,
##     data = base_etude, init.theta = 1.229757378, link = log)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -2.1245627  0.0745969 -28.481  < 2e-16 ***
## drv_age_lic1           -0.0052340  0.0011272  -4.643 3.43e-06 ***
## vh_age22(17, Inf]       0.0635988  0.0298866   2.128 0.033337 *
## vh_speed                0.0005840  0.0004167   1.402 0.161062
## pol_coverageMedian1    -0.0493962  0.0328460  -1.504 0.132614
## pol_coverageMedian2     0.0526784  0.0243482   2.164 0.030499 *
## pol_coverageMini       -0.0069019  0.0336163  -0.205 0.837327
## vh_fuel1Hybrid          0.4294567  0.2704487   1.588 0.112300
## pol_duration2(21, Inf]  0.0535292  0.0242236   2.210 0.027120 *
## drv_age11(18,22]        0.1602155  0.0474412   3.377 0.000732 ***
## drv_age11(22,26]        0.1397666  0.0482140   2.899 0.003745 **
## drv_age11(26,30]        0.1735695  0.0492301   3.526 0.000422 ***
## drv_age11(30,35]        0.1023554  0.0494502   2.070 0.038465 *
## drv_age11(35,40]        0.1850550  0.0516872   3.580 0.000343 ***
## drv_age11(40,45]        0.1466766  0.0561489   2.612 0.008994 **
## drv_age11(45,50]        0.1462256  0.0604488   2.419 0.015563 *
## drv_age11(50,60]        0.1312430  0.0606356   2.164 0.030430 *
## drv_age11(60, Inf]      0.1656434  0.0717266   2.309 0.020923 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

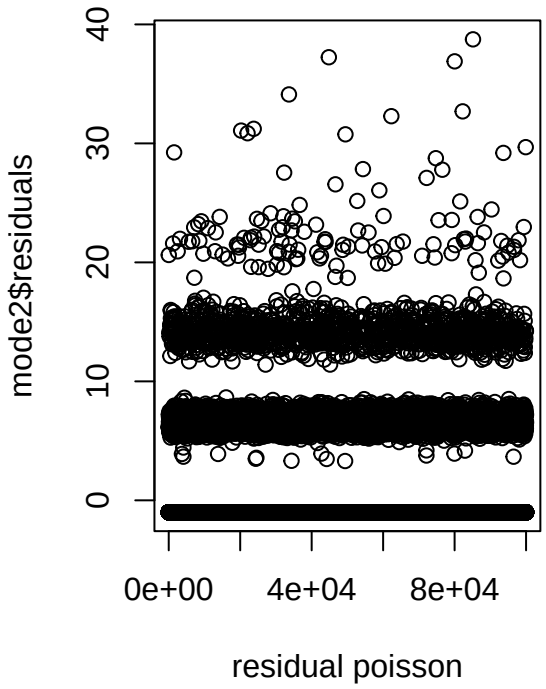
```
##
## (Dispersion parameter for Negative Binomial(1.2298) family taken to be 1)
##
##     Null deviance: 48956  on 99999  degrees of freedom
## Residual deviance: 48900  on 99982  degrees of freedom
## AIC: 81663
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  1.2298
##           Std. Err.:  0.0744
##
##  2 x log-likelihood:  -81624.5010
```

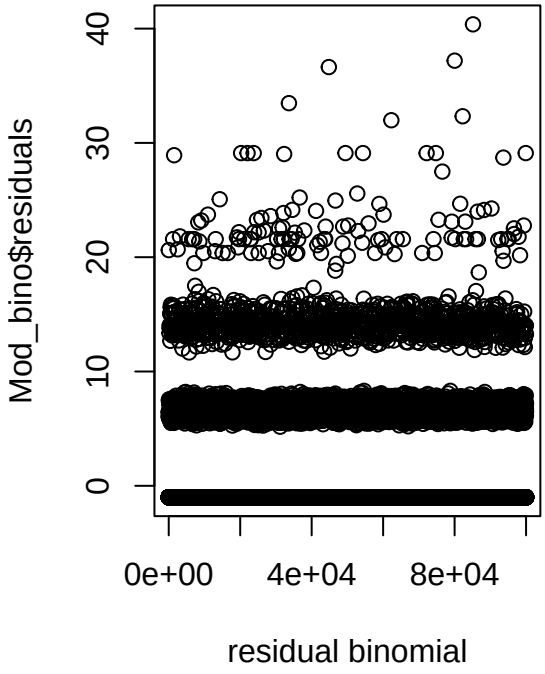
```
########## comparaison des coefficients#####
cbind(mode2$coefficients,Mod_bino$coefficients)
```

```
## Warning in cbind(mode2$coefficients, Mod_bino$coefficients): number of rows of
## result is not a multiple of vector length (arg 2)
```

```
##                                [,1]         [,2]
## (Intercept)            -2.1234316096 -2.018520240
## drv_age_lic1           -0.0052340335 -0.005246193
## vh_age22(17, Inf]       0.0633549022  0.047765299
## vh_speed                0.0005776801  0.055257345
## pol_coverageMedian1    -0.0496553003  0.161019705
## pol_coverageMedian2     0.0526884153  0.140150642
## pol_coverageMini       -0.0070349333  0.174106255
## vh_fuel1Hybrid          0.4304714306  0.102444424
## pol_duration2(21, Inf]  0.0532982619  0.185099184
## drv_age11(18,22]        0.1604505635  0.146728836
## drv_age11(22,26]        0.1393761465  0.146821388
## drv_age11(26,30]        0.1738288385  0.132341653
## drv_age11(30,35]        0.1022269272  0.165896531
## drv_age11(35,40]        0.1851142442 -2.018520240
## drv_age11(40,45]        0.1469617408 -0.005246193
## drv_age11(45,50]        0.1463351407  0.047765299
## drv_age11(50,60]        0.1313749995  0.055257345
## drv_age11(60, Inf]      0.1658459627  0.161019705
```

```
######### verifier normalite des residus
par(mfrow=c(1,2))
plot(mode2$residuals,xlab="residual poisson")
plot(Mod_bino$residuals,xlab="residual binomial")
```






```
######## model quasi poisson
quasi_pos=glm(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22
                +pol_coverage + vh_fuel1 + pol_duration2 + drv_age11,family = quasipoisson("log"), data
summary(quasi_pos)
```

```
##
## Call:
## glm(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_coverage +
##     vh_fuel1 + pol_duration2 + drv_age11, family = quasipoisson("log"),
##     data = base_etude)
##
## Coefficients:
##                          Estimate Std. Error  t value Pr(>|t|)
## (Intercept)            -2.022945   0.017478 -115.744  < 2e-16 ***
## drv_age_lic1           -0.005230   0.001129   -4.631 3.64e-06 ***
## vh_age22(17, Inf]       0.048205   0.027807    1.734 0.082997 .
## pol_coverageMedian1    -0.049634   0.032956   -1.506 0.132052
## pol_coverageMedian2     0.052995   0.024324    2.179 0.029356 *
## pol_coverageMini       -0.006973   0.033669   -0.207 0.835934
## vh_fuel1Hybrid          0.427842   0.263613    1.623 0.104594
## pol_duration2(21, Inf]  0.053257   0.024211    2.200 0.027832 *
## drv_age11(18,22]        0.160153   0.047408    3.378 0.000730 ***
## drv_age11(22,26]        0.139123   0.048214    2.886 0.003909 **
## drv_age11(26,30]        0.173428   0.049197    3.525 0.000423 ***
## drv_age11(30,35]        0.102282   0.049515    2.066 0.038860 *
## drv_age11(35,40]        0.184848   0.051682    3.577 0.000348 ***
```

```
## drv_age11(40,45]        0.146564   0.056210    2.607 0.009123 **
## drv_age11(45,50]        0.146364   0.060545    2.417 0.015631 *
## drv_age11(50,60]        0.131172   0.060766    2.159 0.030880 *
## drv_age11(60, Inf]      0.165529   0.071901    2.302 0.021327 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.110114)
##
##     Null deviance: 57668  on 99999  degrees of freedom
## Residual deviance: 57608  on 99983  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
quasi_pos %>% tbl_regression(intercept = TRUE,exponentiate = TRUE) %>% add_global_p()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	IRR	95% CI	p-value
(Intercept)	0.13	0.13, 0.14	<0.001
drv_age_lic1	1.0	0.99, 1.00	<0.001
vh_age22			0.085
(-Inf,17]	—	—	
(17, Inf]	1.05	0.99, 1.11	
pol_coverage			0.038
Maxi	—	—	
Median1	0.95	0.89, 1.01	
Median2	1.05	1.01, 1.11	
Mini	0.99	0.93, 1.06	
vh_fuel1			0.13
Diesel	—	—	
Hybrid	1.53	0.87, 2.47	
pol_duration2			0.029
(-Inf,21]	—	—	
(21, Inf]	1.05	1.01, 1.11	
drv_age11			0.019
(-Inf,18]	—	—	
(18,22]	1.17	1.07, 1.29	
(22,26]	1.15	1.05, 1.26	
(26,30]	1.19	1.08, 1.31	
(30,35]	1.11	1.00, 1.22	
(35,40]	1.20	1.09, 1.33	
(40,45]	1.16	1.04, 1.29	
(45,50]	1.16	1.03, 1.30	
(50,60]	1.14	1.01, 1.28	
(60, Inf]	1.18	1.02, 1.36	

```r
#Un modèle "zero-inflated" pour la loi de Poisson est utilisé
#pour modéliser des données qui présentent à la fois un excès de zéros
#par rapport à ce qui serait attendu sous une distribution de Poisson pure,
#ainsi que des observations non nulles qui suivent une distribution de Poisson.
#Ce modèle est souvent utilisé pour analyser des données comptables où de nombreux zéros sont observés.

#Le modèle "zero-inflated Poisson" combine deux composantes :

 # Composante de zéro-inflation (Zero-Inflation Component) :
#Cette composante modélise la probabilité que des zéros supplémentaires soient observés par rapport à u
#Cela peut être dû à des facteurs de comportement, des erreurs de mesure ou d'autres raisons.
#Il existe plusieurs distributions que vous pouvez utiliser pour modéliser cette composante,
#mais la plus courante est la distribution de Bernoulli (0 ou 1).

#Composante Poisson (Poisson Component) : Cette composante modélise le nombre d'observations non nulles

#Le modèle est généralement spécifié sous forme de modèle de régression, où les covariables sont utilis

#Voici une équation générale du modèle zero-inflated Poisson :

model_zero=zeroinfl(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_coverage +
                vh_fuel1 + pol_duration2 + drv_age11, dist="poisson", link="logit",
              data = base_etude)
summary(model_zero)
```

```
##
## Call:
## zeroinfl(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_coverage +
##     vh_fuel1 + pol_duration2 + drv_age11, data = base_etude, dist = "poisson",
##     link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -0.4973 -0.3545 -0.3458 -0.3333 14.6340
##
## Count model coefficients (poisson with log link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.435118   0.047872 -29.978   <2e-16 ***
## drv_age_lic1            0.001159   0.002929   0.396   0.6924
## vh_age22(17, Inf]       0.054570   0.076886   0.710   0.4779
## pol_coverageMedian1    -0.193299   0.103018  -1.876   0.0606 .
## pol_coverageMedian2     0.029399   0.064809   0.454   0.6501
## pol_coverageMini       -0.013675   0.096681  -0.141   0.8875
## vh_fuel1Hybrid         -0.113162   0.251657  -0.450   0.6529
## pol_duration2(21, Inf] -0.068976   0.069340  -0.995   0.3199
## drv_age11(18,22]       -0.000956   0.121780  -0.008   0.9937
## drv_age11(22,26]       -0.133555   0.142371  -0.938   0.3482
## drv_age11(26,30]       -0.109128   0.126456  -0.863   0.3882
## drv_age11(30,35]       -0.166749   0.145331  -1.147   0.2512
## drv_age11(35,40]        0.002501   0.135336   0.018   0.9853
## drv_age11(40,45]       -0.035568   0.149170  -0.238   0.8115
## drv_age11(45,50]        0.030811   0.163179   0.189   0.8502
## drv_age11(50,60]       -0.055671   0.165300  -0.337   0.7363
```

```
## drv_age11(60, Inf]      0.066251   0.185598   0.357   0.7211
##
## Zero-inflation model coefficients (binomial with logit link):
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0.222946   0.105066  -2.122   0.0338 *
## drv_age_lic1             0.015187   0.006521   2.329   0.0199 *
## vh_age22(17, Inf]        0.015381   0.170614   0.090   0.9282
## pol_coverageMedian1     -0.364720   0.274960  -1.326   0.1847
## pol_coverageMedian2     -0.055362   0.144081  -0.384   0.7008
## pol_coverageMini        -0.015292   0.211563  -0.072   0.9424
## vh_fuel1Hybrid         -12.071773 455.783915  -0.026   0.9789
## pol_duration2(21, Inf]  -0.298424   0.172173  -1.733   0.0830 .
## drv_age11(18,22]        -0.392842   0.279140  -1.407   0.1593
## drv_age11(22,26]        -0.673352   0.359179  -1.875   0.0608 .
## drv_age11(26,30]        -0.695655   0.316770  -2.196   0.0281 *
## drv_age11(30,35]        -0.656810   0.356637  -1.842   0.0655 .
## drv_age11(35,40]        -0.441580   0.306229  -1.442   0.1493
## drv_age11(40,45]        -0.442297   0.332205  -1.331   0.1831
## drv_age11(45,50]        -0.302456   0.348917  -0.867   0.3860
## drv_age11(50,60]        -0.457516   0.362843  -1.261   0.2073
## drv_age11(60, Inf]      -0.298892   0.389032  -0.768   0.4423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 104
## Log-likelihood: -4.082e+04 on 34 Df
```

```
vuong(mode2,model_zero)
```

```
## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
##  null that the models are indistinguishible)
## -------------------------------------------------------------
##               Vuong z-statistic             H_A    p-value
## Raw                  -9.279257 model2 > model1 < 2.22e-16
## AIC-corrected        -8.590947 model2 > model1 < 2.22e-16
## BIC-corrected        -5.317027 model2 > model1 5.2738e-08
```

```
# on constate que le deuxieme est preferable que le premier car la differance est negative
####### model zero infl binomial########
zero_bin=zeroinfl(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_coverage +
                    vh_fuel1 + pol_duration2 + drv_age11, dist="negbin",
                  data = base_etude)
summary(zero_bin)
```

```
## Warning in sqrt(diag(object$vcov)): Production de NaN
```

```
##
## Call:
## zeroinfl(formula = nombre_sinistre ~ drv_age_lic1 + vh_age22 + pol_coverage +
##     vh_fuel1 + pol_duration2 + drv_age11, data = base_etude, dist = "negbin")
##
```

```
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -0.4536 -0.3531 -0.3468 -0.3317 14.2439
##
## Count model coefficients (negbin with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.949513   0.041013 -47.535  < 2e-16 ***
## drv_age_lic1           -0.003809   0.001594  -2.389  0.01688 *
## vh_age22(17, Inf]       0.023833   0.039296   0.607  0.54418
## pol_coverageMedian1    -0.129371   0.044149  -2.930  0.00339 **
## pol_coverageMedian2     0.066355   0.037284   1.780  0.07513 .
## pol_coverageMini        0.065515   0.071845   0.912  0.36182
## vh_fuel1Hybrid          0.357686   0.268023   1.335  0.18203
## pol_duration2(21, Inf] -0.043882   0.039640  -1.107  0.26828
## drv_age11(18,22]        0.170709   0.070589   2.418  0.01559 *
## drv_age11(22,26]        0.041853   0.068999   0.607  0.54414
## drv_age11(26,30]        0.090090   0.067907   1.327  0.18462
## drv_age11(30,35]        0.113884   0.076538   1.488  0.13677
## drv_age11(35,40]        0.142271   0.078558   1.811  0.07014 .
## drv_age11(40,45]        0.169934   0.091637   1.854  0.06368 .
## drv_age11(45,50]        0.206391   0.094258   2.190  0.02855 *
## drv_age11(50,60]        0.149716   0.095500   1.568  0.11695
## drv_age11(60, Inf]      0.154374   0.124857   1.236  0.21631
## Log(theta)              0.386372   0.091278   4.233 2.31e-05 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -2.65917    0.58943  -4.511 6.44e-06 ***
## drv_age_lic1              0.01925    0.01508   1.277    0.202
## vh_age22(17, Inf]        -0.40522    0.52832  -0.767    0.443
## pol_coverageMedian1     -13.32218 1768.38431  -0.008    0.994
## pol_coverageMedian2       0.18421    0.37844   0.487    0.626
## pol_coverageMini          0.72893    0.52348   1.392    0.164
## vh_fuel1Hybrid           -6.15595        NaN     NaN      NaN
## pol_duration2(21, Inf]  -14.79551  931.63617  -0.016    0.987
## drv_age11(18,22]          0.12535    0.73209   0.171    0.864
## drv_age11(22,26]         -9.16676  151.10152  -0.061    0.952
## drv_age11(26,30]         -1.95074    3.17940  -0.614    0.540
## drv_age11(30,35]          0.13256    0.79203   0.167    0.867
## drv_age11(35,40]         -0.58187    1.02174  -0.569    0.569
## drv_age11(40,45]          0.22713    0.89423   0.254    0.800
## drv_age11(45,50]          0.48426    0.84973   0.570    0.569
## drv_age11(50,60]          0.14627    0.92375   0.158    0.874
## drv_age11(60, Inf]       -0.17739    1.23821  -0.143    0.886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 1.4716
## Number of iterations in BFGS optimization: 125
## Log-likelihood: -4.08e+04 on 35 Df
```

```
vuong(zero_bin,mod_binm2)
```

```
## Vuong Non-Nested Hypothesis Test-Statistic:
```

```
## (test-statistic is asymptotically distributed N(0,1) under the
##  null that the models are indistinguishible)
## -------------------------------------------------------------
##                Vuong z-statistic             H_A  p-value
## Raw                    2.004298 model1 > model2 0.022519
## AIC-corrected         -1.208555 model2 > model1 0.113417
## BIC-corrected        -16.490373 model2 > model1  < 2e-16
```

```
########### comparaison des risidus de model_zero et zero_bin
plot(model_zero$residuals,zero_bin$residuals,xlab="coef_model_pois",ylab="zero_bin")
abline(a=0,b=1,lty=2,col="red")
deviance(model1)
```

```
## [1] 57584.96
```

```
deviance(mode2)
```

```
## [1] 57606.35
```

```
deviance(Mod_bino)
```

```
## [1] 47243.07
```

```
deviance(mod_binm2)
```

```
## [1] 48900.27
```

```
AIC(model1)
```

```
## [1] 82136.78
```

```
AIC(mode2)
```

```
## [1] 82102.18
```

```
AIC(Mod_bino)
```

```
## [1] 81680.67
```

```
AIC(mod_binm2)
```

```
## [1] 81662.5
```

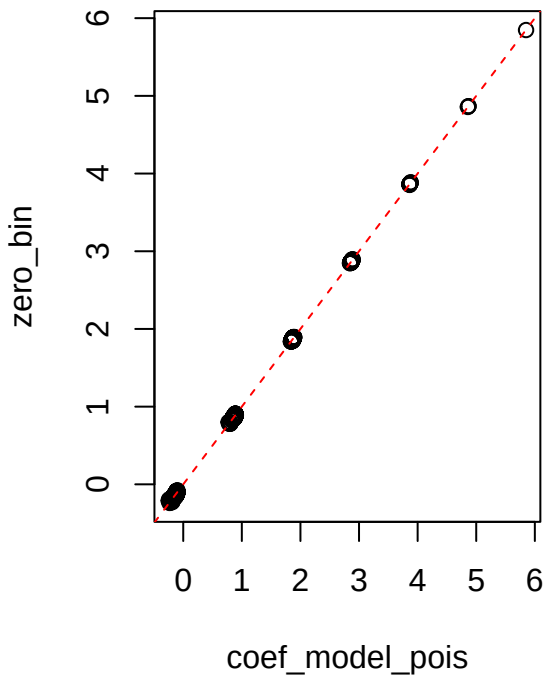
```
AIC(model_zero)
```

```
## [1] 81702.78
```

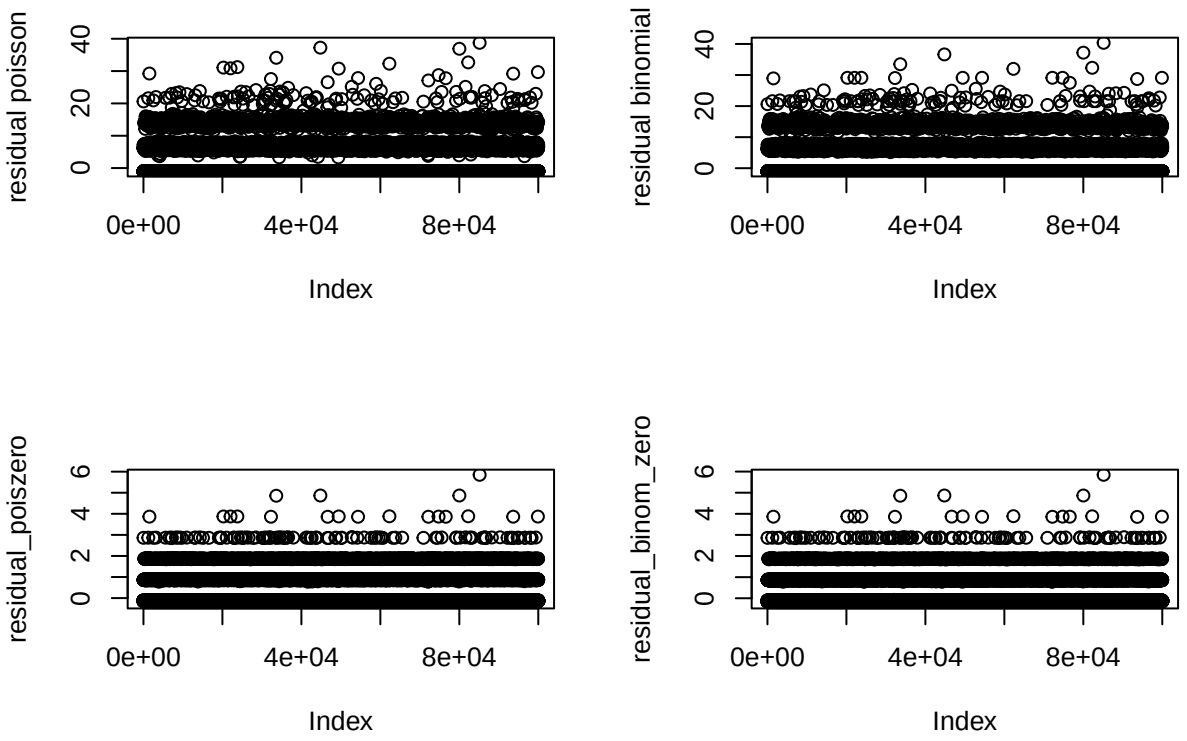
```r
AIC(zero_bin)
```

```
## [1] 81674.54
```

```r
# repartition des residus pour les models
par(mfrow=c(2,2))
```




```r
plot(mode2$residuals,ylab="residual poisson")
plot(Mod_bino$residuals,ylab="residual binomial")
plot(model_zero$residuals,ylab="residual_poiszero")
plot(zero_bin$residuals,ylab="residual_binom_zero")
```




```
########PREDICTION DE LA FREQUENCE SUR LA BASE YEAR1 3.6.1 Preparation de la base#############

setwd("C:/Users/hp/Desktop/wafa assurance/wafa assurance")
pred=read.csv("prediction.csv",sep=",",header=TRUE)
attach(pred)
```

```
## Les objets suivants sont masqués depuis sinistre:
##
##     id_client, id_policy, id_vehicle, id_year
```

```
##Categorisation des variables (on utilise la meme categorisation utilisé dans la base d'etude)
pred$pol_usage2=fct_recode(pred$pol_usage,"Professional"="AllTrips")
pred$vh_fuel1=fct_recode(pred$vh_fuel,"Diesel"="Gasoline")
pred$pol_duration2=cut(pred$pol_duration,c(-Inf,21,Inf))
pred$vh_age22=cut(pred$vh_age,c(-Inf,17,Inf))
pred$drv_age11=cut(pred$drv_age1,c(-Inf,18,22,26,30,35,40,45,50,60,Inf))
pred$vh_make2=as.character(pred$vh_make)
pred$vh_make2[!(pred$vh_make %in% c("RENAULT","PEUGEOT","CITROEN","VOLKSWAGEN","FORD","OPEL","T
OYOTA","MERCEDES BENZ","FIAT"))]="autre_vh"
pred$vh_make2=as.factor(pred$vh_make2)
prediction_freq=predict.glm(Mod_bino,type="response")
summary(prediction_freq)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1025  0.1279  0.1329  0.1324  0.1376  0.1668
```

```r
pred_zero_model=predict(zero_bin,pred,type="response")
summary(pred_zero_model)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.05597 0.12522 0.13254 0.13299 0.13929 0.23992
```

```r
prediction_fre=predict.glm(mode2,type="response")
summary(prediction_fre)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09976 0.12666 0.13227 0.13238 0.13838 0.23307
```

```r
############## MODELISATION DE LA SEVERITE le cout de sinistre###########

base_severete<-subset(base_etude,montant_sinistre>0)

############## statistique descriptive#################

base_severete %>% select(montant_sinistre,vh_age,vh_fuel1,vh_type,pol_coverage,pol_bonus
                         ,pol_duration,drv_age11,vh_cyl) %>% tbl_summary(
                           digits =all_continuous()~c(1,0),
                           statistic = all_continuous()~"{median} ({mean})")
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 11,804**
montant_sinistre	922.0 (1,647)
vh_age	8.0 (10)
vh_fuel1	
Diesel	11,789 (100%)
Hybrid	15 (0.1%)
vh_type	
Commercial	1,141 (9.7%)
Tourism	10,663 (90%)
pol_coverage	
Maxi	7,620 (65%)
Median1	1,057 (9.0%)
Median2	2,139 (18%)
Mini	988 (8.4%)
pol_bonus	0.5 (1)
pol_duration	9.0 (11)
drv_age11	
(-Inf,18]	4,384 (37%)
(18,22]	974 (8.3%)
(22,26]	926 (7.8%)
(26,30]	917 (7.8%)
(30,35]	987 (8.4%)
(35,40]	918 (7.8%)

Characteristic	N = 11,804
(40,45]	745 (6.3%)
(45,50]	609 (5.2%)
(50,60]	807 (6.8%)
(60, Inf]	537 (4.5%)
vh_cyl	1,587.0 (1,648)

```
mean_sini_fuel=base_severete %>% group_by(vh_fuel) %>% summarise(moyenne=mean(montant_sinistre),.groups
mean_sin_age=base_severete %>% group_by(drv_age11) %>% summarise(moyenne=mean(montant_sinistre),.groups

mean_sin_value=base_severete %>% group_by(vh_value) %>% summarise(moyenne=mean(montant_sinistre),.groups

quantile(base_severete$montant_sinistre,probs = c(0.2,0.5,0.7,0.9,0.99))
```

```
##     20%     50%     70%     90%     99%
##  288.60  922.00 1753.00 3745.00 8839.67
```

```
########## recodage des variables pour la base sevérité#########
base_severete$pol_usage2=fct_recode(base_severete$pol_usage,"Professional"="AllTrips")
base_severete$vh_fuel1=fct_recode(base_severete$vh_fuel,"Diesel"="Gasoline")
base_severete$pol_duration2=cut(base_severete$pol_duration,c(-Inf,21,Inf))
base_severete$vh_age22=cut(base_severete$vh_age,c(-Inf,17,Inf))
base_severete$drv_age11=cut(base_severete$drv_age1,c(-Inf,18,22,26,30,35,40,45,50,60,Inf))
base_severete$vh_make2=as.character(base_severete$vh_make)
base_severete$vh_make2[!(base_severete$vh_make %in% c("RENAULT","PEUGEOT","CITROEN","VOLKSWAGEN","FORD"
OYOTA","MERCEDES BENZ","FIAT"))]="autre_vh"
base_severete$vh_make2=as.factor(base_severete$vh_make2)

######## partie modélisation la gamma ##########
gammaa<- glm(formula=montant_sinistre~drv_age2+drv_age_lic1+drv_sex1+drv_age_lic2+drv_sex2+drv_drv2+vh_a

summary(gammaa)
```

```
##
## Call:
## glm(formula = montant_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + pol_usage2 + vh_fuel + pol_duration2 +
##     drv_age11 + vh_make2, family = Gamma(link = "log"), data = base_severete)
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             7.423e+00  2.953e-01  25.133  < 2e-16 ***
## drv_age2                1.782e-03  2.331e-02   0.076  0.93906
## drv_age_lic1            5.901e-03  2.342e-03   2.520  0.01176 *
## drv_sex1M               1.465e-02  3.831e-02   0.382  0.70212
## drv_age_lic2           -1.674e-03  2.331e-02  -0.072  0.94275
## drv_sex2F              -8.783e-02  7.235e-01  -0.121  0.90338
## drv_sex2M              -1.523e-01  7.219e-01  -0.211  0.83293
```

```
## drv_drv2Yes             8.475e-03  8.626e-01   0.010  0.99216
## vh_age22(17, Inf]      -1.687e-01  8.518e-02  -1.980  0.04773 *
## vh_cyl                 -3.140e-05  8.776e-05  -0.358  0.72050
## vh_din                 -3.352e-03  2.073e-03  -1.617  0.10588
## vh_sale_begin          -2.112e-03  8.635e-03  -0.245  0.80679
## vh_sale_end             1.329e-02  9.442e-03   1.408  0.15917
## vh_speed                1.812e-03  1.925e-03   0.941  0.34670
## vh_typeTourism         -7.305e-02  8.141e-02  -0.897  0.36960
## vh_value                1.078e-05  5.989e-06   1.800  0.07196 .
## pol_bonus              -1.741e-01  2.008e-01  -0.867  0.38591
## pol_coverageMedian1    -1.355e-02  6.733e-02  -0.201  0.84047
## pol_coverageMedian2     7.010e-03  5.031e-02   0.139  0.88919
## pol_coverageMini        1.228e-01  6.981e-02   1.759  0.07865 .
## pol_sit_duration       -1.841e-03  8.465e-03  -0.217  0.82783
## pol_pay_freqMonthly    -2.629e-02  4.929e-02  -0.533  0.59382
## pol_pay_freqQuarterly   6.296e-02  1.222e-01   0.515  0.60646
## pol_pay_freqYearly      3.623e-02  4.633e-02   0.782  0.43424
## pol_paydYes             1.631e-01  9.802e-02   1.664  0.09620 .
## pol_usage2Retired      -8.908e-02  7.919e-02  -1.125  0.26066
## pol_usage2WorkPrivate  -8.158e-02  7.291e-02  -1.119  0.26321
## vh_fuelGasoline         9.628e-02  5.420e-02   1.776  0.07569 .
## vh_fuelHybrid          -9.081e-01  5.305e-01  -1.712  0.08698 .
## pol_duration2(21, Inf] -4.844e-02  5.153e-02  -0.940  0.34726
## drv_age11(18,22]       -2.513e-01  9.703e-02  -2.589  0.00963 **
## drv_age11(22,26]       -9.887e-02  9.969e-02  -0.992  0.32134
## drv_age11(26,30]       -1.136e-01  1.006e-01  -1.128  0.25915
## drv_age11(30,35]       -2.245e-01  1.020e-01  -2.202  0.02771 *
## drv_age11(35,40]       -2.511e-01  1.058e-01  -2.373  0.01764 *
## drv_age11(40,45]       -1.582e-01  1.148e-01  -1.378  0.16816
## drv_age11(45,50]       -2.078e-01  1.241e-01  -1.675  0.09393 .
## drv_age11(50,60]       -2.624e-01  1.246e-01  -2.107  0.03514 *
## drv_age11(60, Inf]     -2.551e-01  1.475e-01  -1.730  0.08371 .
## vh_make2CITROEN        -7.066e-02  6.865e-02  -1.029  0.30340
## vh_make2FIAT            8.770e-02  1.205e-01   0.728  0.46661
## vh_make2FORD           -9.673e-02  1.007e-01  -0.961  0.33671
## vh_make2MERCEDES BENZ  -7.253e-02  1.195e-01  -0.607  0.54403
## vh_make2OPEL           -1.966e-02  1.131e-01  -0.174  0.86205
## vh_make2PEUGEOT         2.648e-02  6.552e-02   0.404  0.68608
## vh_make2RENAULT        -9.917e-02  6.100e-02  -1.626  0.10404
## vh_make2VOLKSWAGEN     -1.553e-01  9.527e-02  -1.630  0.10305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 4.143706)
##
##     Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15212  on 11757  degrees of freedom
## AIC: 198390
##
## Number of Fisher Scoring iterations: 8
```

```
gamma3<-step(gammaa)
```

```
## Start:  AIC=198390.3
```

```
## montant_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_sex2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     pol_usage2 + vh_fuel + pol_duration2 + drv_age11 + vh_make2
##
##                     Df Deviance    AIC
## - drv_age11          9    15257 198383
## - vh_make2           8    15253 198384
## - pol_pay_freq       3    15220 198386
## - drv_sex2           2    15216 198387
## - pol_coverage       3    15226 198388
## - pol_usage2         2    15218 198388
## - drv_drv2           1    15212 198388
## - drv_age_lic2       1    15212 198388
## - drv_age2           1    15212 198388
## - pol_sit_duration   1    15212 198388
## - vh_sale_begin      1    15212 198388
## - vh_cyl             1    15213 198388
## - drv_sex1           1    15213 198388
## - pol_bonus          1    15215 198389
## - vh_type            1    15215 198389
## - pol_duration2      1    15216 198389
## - vh_speed           1    15216 198389
## - vh_sale_end        1    15220 198390
## <none>                    15212 198390
## - vh_din             1    15223 198391
## - pol_payd           1    15223 198391
## - vh_value           1    15225 198392
## - vh_fuel            2    15234 198392
## - vh_age22           1    15228 198392
## - drv_age_lic1       1    15238 198394
##
## Step:  AIC=198413.9
## montant_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_sex2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_coverage + pol_sit_duration + pol_pay_freq + pol_payd +
##     pol_usage2 + vh_fuel + pol_duration2 + vh_make2
```

```
summary(gamma3)
```

```
##
## Call:
## glm(formula = montant_sinistre ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + pol_usage2 + vh_fuel + pol_duration2 +
##     vh_make2, family = Gamma(link = "log"), data = base_severete)
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             7.413e+00  3.048e-01  24.321   <2e-16 ***
```

```
## drv_age2                3.110e-03  2.406e-02   0.129   0.8972
## drv_age_lic1            1.163e-03  1.050e-03   1.108   0.2678
## drv_sex1M               1.921e-02  3.955e-02   0.486   0.6272
## drv_age_lic2           -2.890e-03  2.407e-02  -0.120   0.9044
## drv_sex2F              -8.915e-02  7.469e-01  -0.119   0.9050
## drv_sex2M              -1.541e-01  7.453e-01  -0.207   0.8362
## drv_drv2Yes            -2.059e-02  8.908e-01  -0.023   0.9816
## vh_age22(17, Inf]      -1.736e-01  8.797e-02  -1.973   0.0485 *
## vh_cyl                 -3.615e-05  9.064e-05  -0.399   0.6901
## vh_din                 -3.544e-03  2.141e-03  -1.655   0.0979 .
## vh_sale_begin          -1.813e-03  8.919e-03  -0.203   0.8390
## vh_sale_end             1.339e-02  9.752e-03   1.373   0.1698
## vh_speed                1.977e-03  1.989e-03   0.994   0.3201
## vh_typeTourism         -7.590e-02  8.409e-02  -0.903   0.3667
## vh_value                1.145e-05  6.187e-06   1.852   0.0641 .
## pol_bonus              -2.208e-01  2.062e-01  -1.071   0.2843
## pol_coverageMedian1    -1.840e-02  6.953e-02  -0.265   0.7913
## pol_coverageMedian2    -7.346e-04  5.198e-02  -0.014   0.9887
## pol_coverageMini        1.176e-01  7.210e-02   1.632   0.1028
## pol_sit_duration       -1.510e-03  8.744e-03  -0.173   0.8629
## pol_pay_freqMonthly    -3.111e-02  5.091e-02  -0.611   0.5411
## pol_pay_freqQuarterly   5.723e-02  1.263e-01   0.453   0.6504
## pol_pay_freqYearly      3.216e-02  4.785e-02   0.672   0.5016
## pol_paydYes             1.715e-01  1.012e-01   1.695   0.0901 .
## pol_usage2Retired      -9.010e-02  8.180e-02  -1.101   0.2707
## pol_usage2WorkPrivate  -8.305e-02  7.533e-02  -1.103   0.2703
## vh_fuelGasoline         9.819e-02  5.598e-02   1.754   0.0795 .
## vh_fuelHybrid          -9.215e-01  5.479e-01  -1.682   0.0926 .
## pol_duration2(21, Inf] -5.133e-02  5.322e-02  -0.965   0.3348
## vh_make2CITROEN        -7.283e-02  7.091e-02  -1.027   0.3045
## vh_make2FIAT            1.157e-01  1.244e-01   0.930   0.3524
## vh_make2FORD           -1.021e-01  1.040e-01  -0.982   0.3264
## vh_make2MERCEDES BENZ  -8.000e-02  1.234e-01  -0.648   0.5170
## vh_make2OPEL           -2.744e-02  1.168e-01  -0.235   0.8143
## vh_make2PEUGEOT         2.233e-02  6.768e-02   0.330   0.7415
## vh_make2RENAULT        -1.040e-01  6.302e-02  -1.650   0.0990 .
## vh_make2VOLKSWAGEN     -1.546e-01  9.840e-02  -1.571   0.1163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 4.423869)
##
##     Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15257  on 11766  degrees of freedom
## AIC: 198414
##
## Number of Fisher Scoring iterations: 8
```

```
gamma4<-glm(montant_sinistre ~ drv_age_lic1 + pol_bonus +
             vh_fuel1 + drv_age11,family = Gamma(link = "log") ,data = base_severete)
summary(gamma4)
```

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

```
## glm(formula = montant_sinistre ~ drv_age_lic1 + pol_bonus + vh_fuel1 +
##     drv_age11, family = Gamma(link = "log"), data = base_severete)
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.511696   0.134593  55.811   <2e-16 ***
## drv_age_lic1        0.006477   0.002799   2.314   0.0207 *
## pol_bonus          -0.196047   0.233137  -0.841   0.4004
## vh_fuel1Hybrid     -0.865290   0.630672  -1.372   0.1701
## drv_age11(18,22]   -0.283686   0.116124  -2.443   0.0146 *
## drv_age11(22,26]   -0.108174   0.119244  -0.907   0.3643
## drv_age11(26,30]   -0.117126   0.120435  -0.973   0.3308
## drv_age11(30,35]   -0.256693   0.122011  -2.104   0.0354 *
## drv_age11(35,40]   -0.267908   0.126631  -2.116   0.0344 *
## drv_age11(40,45]   -0.171495   0.137367  -1.248   0.2119
## drv_age11(45,50]   -0.239278   0.148482  -1.612   0.1071
## drv_age11(50,60]   -0.294593   0.149115  -1.976   0.0482 *
## drv_age11(60, Inf] -0.284040   0.176446  -1.610   0.1075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 5.952702)
##
##     Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15404  on 11791  degrees of freedom
## AIC: 198499
##
## Number of Fisher Scoring iterations: 6
```

```
deviance(gamma4)
```

```
## [1] 15403.85
```

```
AIC(gamma4)
```

```
## [1] 198498.7
```

```
logLik(gamma4)
```

```
## 'log Lik.' -99235.37 (df=14)
```

```
#residus #
plot(gamma4$residuals)
######## la loi log normal#########
log_nrm=lm(log(montant_sinistre)~drv_age2+drv_age_lic1+drv_sex1+drv_age_lic2+drv_sex2+drv_drv2+vh_age22
summary(log_nrm)
```

```
##
## Call:
## lm(formula = log(montant_sinistre) ~ drv_age2 + drv_age_lic1 +
##     drv_sex1 + drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 +
```

```
##     vh_cyl + vh_din + vh_sale_begin + vh_sale_end + vh_speed +
##     vh_type + vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + pol_usage2 + vh_fuel + pol_duration2 +
##     drv_age11 + vh_make2, data = base_severete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8182 -0.7578  0.0823  0.8482  5.5912
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              6.897e+00  1.731e-01  39.839  < 2e-16 ***
## drv_age2                 2.965e-03  1.366e-02   0.217 0.828185
## drv_age_lic1             5.966e-03  1.373e-03   4.345 1.40e-05 ***
## drv_sex1M               -1.383e-02  2.246e-02  -0.616 0.538134
## drv_age_lic2            -3.086e-03  1.367e-02  -0.226 0.821357
## drv_sex2F               -2.062e-01  4.241e-01  -0.486 0.626788
## drv_sex2M               -2.301e-01  4.232e-01  -0.544 0.586607
## drv_drv2Yes              1.428e-01  5.057e-01   0.282 0.777710
## vh_age22(17, Inf]       -3.827e-02  4.994e-02  -0.766 0.443480
## vh_cyl                   4.518e-05  5.145e-05   0.878 0.379895
## vh_din                  -3.550e-04  1.215e-03  -0.292 0.770161
## vh_sale_begin           -1.610e-03  5.062e-03  -0.318 0.750512
## vh_sale_end              2.649e-03  5.535e-03   0.479 0.632220
## vh_speed                -7.453e-04  1.129e-03  -0.660 0.509059
## vh_typeTourism           1.400e-02  4.773e-02   0.293 0.769283
## vh_value                 6.609e-07  3.511e-06   0.188 0.850673
## pol_bonus               -1.831e-01  1.177e-01  -1.555 0.119931
## pol_coverageMedian1      1.560e-02  3.947e-02   0.395 0.692761
## pol_coverageMedian2      6.499e-03  2.949e-02   0.220 0.825611
## pol_coverageMini         2.050e-02  4.093e-02   0.501 0.616489
## pol_sit_duration        -7.411e-04  4.963e-03  -0.149 0.881285
## pol_pay_freqMonthly      1.011e-02  2.889e-02   0.350 0.726525
## pol_pay_freqQuarterly    6.652e-02  7.165e-02   0.928 0.353215
## pol_pay_freqYearly       3.572e-02  2.716e-02   1.315 0.188491
## pol_paydYes             -3.357e-02  5.746e-02  -0.584 0.559126
## pol_usage2Retired       -3.685e-02  4.642e-02  -0.794 0.427339
## pol_usage2WorkPrivate   -6.046e-03  4.274e-02  -0.141 0.887510
## vh_fuelGasoline          4.647e-02  3.177e-02   1.462 0.143653
## vh_fuelHybrid           -5.764e-01  3.110e-01  -1.853 0.063856 .
## pol_duration2(21, Inf]   3.402e-02  3.021e-02   1.126 0.260125
## drv_age11(18,22]        -2.142e-01  5.688e-02  -3.765 0.000167 ***
## drv_age11(22,26]        -1.915e-01  5.844e-02  -3.277 0.001052 **
## drv_age11(26,30]        -1.787e-01  5.900e-02  -3.029 0.002461 **
## drv_age11(30,35]        -2.110e-01  5.979e-02  -3.529 0.000418 ***
## drv_age11(35,40]        -2.481e-01  6.203e-02  -3.999 6.39e-05 ***
## drv_age11(40,45]        -2.106e-01  6.731e-02  -3.129 0.001760 **
## drv_age11(45,50]        -1.931e-01  7.272e-02  -2.655 0.007942 **
## drv_age11(50,60]        -2.235e-01  7.302e-02  -3.061 0.002211 **
## drv_age11(60, Inf]      -2.107e-01  8.644e-02  -2.437 0.014819 *
## vh_make2CITROEN          1.260e-02  4.024e-02   0.313 0.754285
## vh_make2FIAT            -1.288e-01  7.062e-02  -1.824 0.068110 .
## vh_make2FORD             1.003e-02  5.902e-02   0.170 0.865050
## vh_make2MERCEDES BENZ    7.531e-02  7.008e-02   1.075 0.282553
```

```
## vh_make2OPEL           -2.295e-02  6.633e-02  -0.346 0.729309
## vh_make2PEUGEOT         1.672e-02  3.841e-02   0.435 0.663292
## vh_make2RENAULT         1.136e-02  3.576e-02   0.318 0.750721
## vh_make2VOLKSWAGEN     -1.074e-01  5.585e-02  -1.923 0.054551 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.193 on 11757 degrees of freedom
## Multiple R-squared:  0.004905,   Adjusted R-squared:  0.001012
## F-statistic:  1.26 on 46 and 11757 DF,  p-value: 0.1117
```

```
step(log_nrm)
```

```
## Start:  AIC=4219.71
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_coverage + pol_sit_duration +
##     pol_pay_freq + pol_payd + pol_usage2 + vh_fuel + pol_duration2 +
##     drv_age11 + vh_make2
##
##                     Df Sum of Sq   RSS    AIC
## - pol_coverage       3    0.5266 16743 4214.1
## - vh_make2           8   16.0777 16759 4215.0
## - pol_pay_freq       3    3.3946 16746 4216.1
## - drv_sex2           2    0.9302 16744 4216.4
## - pol_usage2         2    2.0388 16745 4217.1
## - pol_sit_duration   1    0.0318 16743 4217.7
## - vh_value           1    0.0505 16743 4217.7
## - drv_age2           1    0.0671 16743 4217.8
## - drv_age_lic2       1    0.0726 16743 4217.8
## - drv_drv2           1    0.1135 16743 4217.8
## - vh_din             1    0.1216 16743 4217.8
## - vh_type            1    0.1225 16743 4217.8
## - vh_sale_begin      1    0.1440 16743 4217.8
## - vh_sale_end        1    0.3262 16743 4217.9
## - pol_payd           1    0.4859 16743 4218.1
## - drv_sex1           1    0.5398 16743 4218.1
## - vh_speed           1    0.6209 16743 4218.1
## - vh_age22           1    0.8364 16743 4218.3
## - vh_cyl             1    1.0981 16744 4218.5
## - pol_duration2      1    1.8060 16744 4219.0
## <none>                           16743 4219.7
## - pol_bonus          1    3.4442 16746 4220.1
## - vh_fuel            2    8.1278 16751 4221.4
## - drv_age11          9   31.0367 16774 4223.6
## - drv_age_lic1       1   26.8856 16770 4236.6
##
## Step:  AIC=4214.08
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_sit_duration + pol_pay_freq +
##     pol_payd + pol_usage2 + vh_fuel + pol_duration2 + drv_age11 +
```

```
##     vh_make2
##
##                     Df Sum of Sq   RSS    AIC
## - vh_make2           8   16.0765 16759 4209.4
## - pol_pay_freq       3    3.3480 16747 4210.4
## - drv_sex2           2    0.9484 16744 4210.7
## - pol_usage2         2    2.1484 16745 4211.6
## - pol_sit_duration   1    0.0083 16743 4212.1
## - vh_value           1    0.0501 16743 4212.1
## - drv_age2           1    0.0649 16743 4212.1
## - drv_age_lic2       1    0.0706 16743 4212.1
## - drv_drv2           1    0.1207 16743 4212.2
## - vh_type            1    0.1245 16743 4212.2
## - vh_din             1    0.1247 16743 4212.2
## - vh_sale_begin      1    0.1460 16743 4212.2
## - vh_sale_end        1    0.3216 16743 4212.3
## - pol_payd           1    0.4350 16744 4212.4
## - drv_sex1           1    0.5477 16744 4212.5
## - vh_speed           1    0.6105 16744 4212.5
## - vh_age22           1    0.8257 16744 4212.7
## - vh_cyl             1    1.1101 16744 4212.9
## - pol_duration2      1    1.6633 16745 4213.3
## <none>                           16743 4214.1
## - pol_bonus          1    3.2884 16746 4214.4
## - vh_fuel            2    8.1814 16751 4215.8
## - drv_age11          9   31.0058 16774 4217.9
## - drv_age_lic1       1   26.8517 16770 4231.0
##
## Step:  AIC=4209.41
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_sit_duration + pol_pay_freq +
##     pol_payd + pol_usage2 + vh_fuel + pol_duration2 + drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - pol_pay_freq       3    3.1751 16762 4205.6
## - drv_sex2           2    0.9703 16760 4206.1
## - pol_usage2         2    2.1595 16761 4206.9
## - pol_sit_duration   1    0.0124 16759 4207.4
## - vh_sale_begin      1    0.0154 16759 4207.4
## - vh_type            1    0.0185 16759 4207.4
## - drv_age2           1    0.0754 16759 4207.5
## - drv_age_lic2       1    0.0800 16759 4207.5
## - drv_drv2           1    0.1039 16759 4207.5
## - vh_sale_end        1    0.1296 16759 4207.5
## - vh_speed           1    0.2018 16759 4207.6
## - pol_payd           1    0.4110 16760 4207.7
## - drv_sex1           1    0.4729 16760 4207.7
## - vh_value           1    0.5008 16760 4207.8
## - vh_din             1    0.5444 16760 4207.8
## - vh_age22           1    0.7005 16760 4207.9
## - vh_cyl             1    1.2547 16761 4208.3
## - pol_duration2      1    1.7343 16761 4208.6
```

```
## <none>                           16759 4209.4
## - pol_bonus         1    3.3939 16763 4209.8
## - vh_fuel           2    9.0357 16768 4211.8
## - drv_age11         9   31.3132 16791 4213.4
## - drv_age_lic1      1   26.3738 16786 4226.0
##
## Step:  AIC=4205.65
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_sex2 + drv_drv2 + vh_age22 + vh_cyl +
##     vh_din + vh_sale_begin + vh_sale_end + vh_speed + vh_type +
##     vh_value + pol_bonus + pol_sit_duration + pol_payd + pol_usage2 +
##     vh_fuel + pol_duration2 + drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - drv_sex2           2    0.9231 16763 4202.3
## - pol_usage2         2    1.7099 16764 4202.8
## - pol_sit_duration   1    0.0082 16762 4203.7
## - vh_sale_begin      1    0.0131 16762 4203.7
## - vh_type            1    0.0219 16762 4203.7
## - drv_age2           1    0.0698 16762 4203.7
## - drv_age_lic2       1    0.0745 16763 4203.7
## - drv_drv2           1    0.0911 16763 4203.7
## - vh_sale_end        1    0.1212 16763 4203.7
## - vh_speed           1    0.1917 16763 4203.8
## - pol_payd           1    0.3377 16763 4203.9
## - vh_value           1    0.4822 16763 4204.0
## - drv_sex1           1    0.4911 16763 4204.0
## - vh_din             1    0.5521 16763 4204.0
## - vh_age22           1    0.7301 16763 4204.2
## - vh_cyl             1    1.2722 16764 4204.5
## - pol_duration2      1    1.4973 16764 4204.7
## <none>                           16762 4205.6
## - pol_bonus          1    3.5235 16766 4206.1
## - vh_fuel            2    8.9931 16771 4208.0
## - drv_age11          9   31.3606 16794 4209.7
## - drv_age_lic1       1   26.2254 16789 4222.1
##
## Step:  AIC=4202.3
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_sit_duration + pol_payd + pol_usage2 + vh_fuel + pol_duration2 +
##     drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - pol_usage2         2    1.6990 16765 4199.5
## - pol_sit_duration   1    0.0064 16763 4200.3
## - vh_sale_begin      1    0.0115 16763 4200.3
## - vh_type            1    0.0212 16763 4200.3
## - drv_age2           1    0.0588 16763 4200.3
## - drv_age_lic2       1    0.0621 16763 4200.3
## - drv_drv2           1    0.0885 16763 4200.4
## - vh_sale_end        1    0.1172 16763 4200.4
## - vh_speed           1    0.1928 16764 4200.4
```

```
## - pol_payd          1    0.3378 16764 4200.5
## - drv_sex1          1    0.4900 16764 4200.6
## - vh_value          1    0.4916 16764 4200.6
## - vh_din            1    0.5530 16764 4200.7
## - vh_age22          1    0.7273 16764 4200.8
## - vh_cyl            1    1.2493 16765 4201.2
## - pol_duration2     1    1.5004 16765 4201.4
## <none>                          16763 4202.3
## - pol_bonus         1    3.5616 16767 4202.8
## - vh_fuel           2    8.9592 16772 4204.6
## - drv_age11         9   31.4688 16795 4206.4
## - drv_age_lic1      1   26.2398 16790 4218.8
##
## Step:  AIC=4199.49
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_begin +
##     vh_sale_end + vh_speed + vh_type + vh_value + pol_bonus +
##     pol_sit_duration + pol_payd + vh_fuel + pol_duration2 + drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - vh_sale_begin      1    0.0095 16765 4197.5
## - vh_type            1    0.0196 16765 4197.5
## - pol_sit_duration   1    0.0502 16765 4197.5
## - drv_age2           1    0.0579 16765 4197.5
## - drv_age_lic2       1    0.0615 16765 4197.5
## - drv_drv2           1    0.0876 16765 4197.6
## - vh_sale_end        1    0.1167 16765 4197.6
## - vh_speed           1    0.1899 16765 4197.6
## - drv_sex1           1    0.4755 16766 4197.8
## - vh_value           1    0.4826 16766 4197.8
## - vh_din             1    0.5499 16766 4197.9
## - pol_payd           1    0.6181 16766 4197.9
## - vh_age22           1    0.7547 16766 4198.0
## - pol_duration2      1    1.0509 16766 4198.2
## - vh_cyl             1    1.2456 16766 4198.4
## <none>                           16765 4199.5
## - pol_bonus          1    3.0756 16768 4199.7
## - vh_fuel            2    9.0186 16774 4201.8
## - drv_age11          9   31.2588 16796 4203.5
## - drv_age_lic1       1   26.0215 16791 4215.8
##
## Step:  AIC=4197.5
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end +
##     vh_speed + vh_type + vh_value + pol_bonus + pol_sit_duration +
##     pol_payd + vh_fuel + pol_duration2 + drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - vh_type            1    0.0228 16765 4195.5
## - pol_sit_duration   1    0.0502 16765 4195.5
## - drv_age2           1    0.0579 16765 4195.5
## - drv_age_lic2       1    0.0614 16765 4195.5
## - drv_drv2           1    0.0876 16765 4195.6
## - vh_speed           1    0.1835 16765 4195.6
```

```
## - vh_sale_end       1    0.2865 16765 4195.7
## - drv_sex1          1    0.4751 16766 4195.8
## - vh_value          1    0.4772 16766 4195.8
## - vh_din            1    0.5456 16766 4195.9
## - pol_payd          1    0.6201 16766 4195.9
## - vh_age22          1    0.8967 16766 4196.1
## - pol_duration2     1    1.0491 16766 4196.2
## - vh_cyl            1    1.2389 16766 4196.4
## <none>                          16765 4197.5
## - pol_bonus         1    3.0810 16768 4197.7
## - vh_fuel           2    9.0092 16774 4199.8
## - drv_age11         9   31.2605 16796 4201.5
## - drv_age_lic1      1   26.0334 16791 4213.8
##
## Step:  AIC=4195.51
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end +
##     vh_speed + vh_value + pol_bonus + pol_sit_duration + pol_payd +
##     vh_fuel + pol_duration2 + drv_age11
##
##                     Df Sum of Sq   RSS    AIC
## - pol_sit_duration  1    0.0500 16765 4193.5
## - drv_age2          1    0.0583 16765 4193.6
## - drv_age_lic2      1    0.0619 16765 4193.6
## - drv_drv2          1    0.0881 16765 4193.6
## - vh_speed          1    0.1639 16765 4193.6
## - vh_sale_end       1    0.3005 16765 4193.7
## - drv_sex1          1    0.4736 16766 4193.8
## - vh_value          1    0.4993 16766 4193.9
## - vh_din            1    0.5969 16766 4193.9
## - pol_payd          1    0.6188 16766 4193.9
## - vh_age22          1    0.9072 16766 4194.2
## - pol_duration2     1    1.0488 16766 4194.3
## - vh_cyl            1    1.2171 16766 4194.4
## <none>                          16765 4195.5
## - pol_bonus         1    3.0779 16768 4195.7
## - vh_fuel           2    9.0783 16774 4197.9
## - drv_age11         9   31.2450 16796 4199.5
## - drv_age_lic1      1   26.0255 16791 4211.8
##
## Step:  AIC=4193.55
## log(montant_sinistre) ~ drv_age2 + drv_age_lic1 + drv_sex1 +
##     drv_age_lic2 + drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end +
##     vh_speed + vh_value + pol_bonus + pol_payd + vh_fuel + pol_duration2 +
##     drv_age11
##
##                   Df Sum of Sq   RSS    AIC
## - drv_age2         1    0.0599 16765 4191.6
## - drv_age_lic2     1    0.0634 16765 4191.6
## - drv_drv2         1    0.0901 16765 4191.6
## - vh_speed         1    0.1610 16765 4191.7
## - vh_sale_end      1    0.3008 16765 4191.8
## - drv_sex1         1    0.4727 16766 4191.9
## - vh_value         1    0.4969 16766 4191.9
```

```
## - pol_payd       1    0.5780 16766 4192.0
## - vh_din         1    0.5963 16766 4192.0
## - vh_age22       1    0.9024 16766 4192.2
## - pol_duration2  1    1.0027 16766 4192.3
## - vh_cyl         1    1.2160 16766 4192.4
## <none>                       16765 4193.5
## - pol_bonus      1    3.0312 16768 4193.7
## - vh_fuel        2    9.0726 16774 4195.9
## - drv_age11      9   31.2354 16796 4197.5
## - drv_age_lic1   1   25.9948 16791 4209.8
##
## Step:  AIC=4191.59
## log(montant_sinistre) ~ drv_age_lic1 + drv_sex1 + drv_age_lic2 +
##     drv_drv2 + vh_age22 + vh_cyl + vh_din + vh_sale_end + vh_speed +
##     vh_value + pol_bonus + pol_payd + vh_fuel + pol_duration2 +
##     drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - drv_age_lic2   1    0.0088 16765 4189.6
## - vh_speed       1    0.1634 16765 4189.7
## - drv_drv2       1    0.1657 16765 4189.7
## - vh_sale_end    1    0.3021 16765 4189.8
## - drv_sex1       1    0.4765 16766 4189.9
## - vh_value       1    0.4949 16766 4189.9
## - pol_payd       1    0.5765 16766 4190.0
## - vh_din         1    0.5927 16766 4190.0
## - vh_age22       1    0.9077 16766 4190.2
## - pol_duration2  1    1.0007 16766 4190.3
## - vh_cyl         1    1.2168 16766 4190.4
## <none>                       16765 4191.6
## - pol_bonus      1    3.0295 16768 4191.7
## - vh_fuel        2    9.0635 16774 4194.0
## - drv_age11      9   31.2535 16796 4195.6
## - drv_age_lic1   1   26.0296 16791 4207.9
##
## Step:  AIC=4189.6
## log(montant_sinistre) ~ drv_age_lic1 + drv_sex1 + drv_drv2 +
##     vh_age22 + vh_cyl + vh_din + vh_sale_end + vh_speed + vh_value +
##     pol_bonus + pol_payd + vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - vh_speed       1    0.1627 16765 4187.7
## - vh_sale_end    1    0.3035 16766 4187.8
## - drv_sex1       1    0.4735 16766 4187.9
## - vh_value       1    0.4948 16766 4187.9
## - pol_payd       1    0.5764 16766 4188.0
## - vh_din         1    0.5925 16766 4188.0
## - drv_drv2       1    0.6678 16766 4188.1
## - vh_age22       1    0.9089 16766 4188.2
## - pol_duration2  1    0.9988 16766 4188.3
## - vh_cyl         1    1.2142 16766 4188.5
## <none>                       16765 4189.6
## - pol_bonus      1    3.0326 16768 4189.7
## - vh_fuel        2    9.0590 16774 4192.0
```

```
## - drv_age11      9   31.2505 16796 4193.6
## - drv_age_lic1   1   26.0305 16791 4205.9
##
## Step:  AIC=4187.71
## log(montant_sinistre) ~ drv_age_lic1 + drv_sex1 + drv_drv2 +
##     vh_age22 + vh_cyl + vh_din + vh_sale_end + vh_value + pol_bonus +
##     pol_payd + vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - vh_sale_end    1    0.2975 16766 4185.9
## - drv_sex1       1    0.4659 16766 4186.0
## - pol_payd       1    0.5775 16766 4186.1
## - vh_value       1    0.6342 16766 4186.2
## - drv_drv2       1    0.6628 16766 4186.2
## - vh_age22       1    0.8023 16766 4186.3
## - pol_duration2  1    0.9922 16766 4186.4
## - vh_cyl         1    1.6868 16767 4186.9
## - vh_din         1    2.1427 16768 4187.2
## <none>                       16765 4187.7
## - pol_bonus      1    3.0273 16768 4187.8
## - vh_fuel        2    9.3680 16775 4190.3
## - drv_age11      9   31.2359 16797 4191.7
## - drv_age_lic1   1   26.0143 16791 4204.0
##
## Step:  AIC=4185.92
## log(montant_sinistre) ~ drv_age_lic1 + drv_sex1 + drv_drv2 +
##     vh_age22 + vh_cyl + vh_din + vh_value + pol_bonus + pol_payd +
##     vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - drv_sex1       1    0.4620 16766 4184.2
## - vh_age22       1    0.5051 16766 4184.3
## - vh_value       1    0.5528 16766 4184.3
## - pol_payd       1    0.5786 16766 4184.3
## - drv_drv2       1    0.6551 16766 4184.4
## - pol_duration2  1    0.9811 16767 4184.6
## - vh_din         1    2.6850 16768 4185.8
## - vh_cyl         1    2.8106 16768 4185.9
## <none>                       16766 4185.9
## - pol_bonus      1    3.0127 16769 4186.0
## - vh_fuel        2   10.7017 16776 4189.5
## - drv_age11      9   31.2732 16797 4189.9
## - drv_age_lic1   1   26.0892 16792 4202.3
##
## Step:  AIC=4184.25
## log(montant_sinistre) ~ drv_age_lic1 + drv_drv2 + vh_age22 +
##     vh_cyl + vh_din + vh_value + pol_bonus + pol_payd + vh_fuel +
##     pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - vh_age22       1    0.5118 16767 4182.6
## - vh_value       1    0.5389 16767 4182.6
## - pol_payd       1    0.5805 16767 4182.7
## - drv_drv2       1    0.6725 16767 4182.7
```

```
## - pol_duration2  1    0.9964 16767 4182.9
## - vh_din         1    2.6753 16769 4184.1
## <none>                       16766 4184.2
## - vh_cyl         1    2.8485 16769 4184.3
## - pol_bonus      1    2.9425 16769 4184.3
## - vh_fuel        2   10.7554 16777 4187.8
## - drv_age11      9   31.0830 16797 4188.1
## - drv_age_lic1   1   25.8757 16792 4200.5
##
## Step:  AIC=4182.61
## log(montant_sinistre) ~ drv_age_lic1 + drv_drv2 + vh_cyl + vh_din +
##     vh_value + pol_bonus + pol_payd + vh_fuel + pol_duration2 +
##     drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - vh_value       1    0.5631 16767 4181.0
## - pol_payd       1    0.5774 16767 4181.0
## - drv_drv2       1    0.6602 16767 4181.1
## - pol_duration2  1    1.0175 16768 4181.3
## - vh_din         1    2.2993 16769 4182.2
## - vh_cyl         1    2.3987 16769 4182.3
## <none>                       16767 4182.6
## - pol_bonus      1    2.8906 16770 4182.6
## - vh_fuel        2   10.3128 16777 4185.9
## - drv_age11      9   30.9980 16798 4186.4
## - drv_age_lic1   1   25.7754 16792 4198.7
##
## Step:  AIC=4181
## log(montant_sinistre) ~ drv_age_lic1 + drv_drv2 + vh_cyl + vh_din +
##     pol_bonus + pol_payd + vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - pol_payd       1    0.5835 16768 4179.4
## - drv_drv2       1    0.6567 16768 4179.5
## - pol_duration2  1    1.0194 16768 4179.7
## - vh_din         1    2.2486 16769 4180.6
## <none>                       16767 4181.0
## - pol_bonus      1    2.8852 16770 4181.0
## - vh_cyl         1    3.0203 16770 4181.1
## - vh_fuel        2    9.8566 16777 4183.9
## - drv_age11      9   31.0690 16798 4184.9
## - drv_age_lic1   1   25.7717 16793 4197.1
##
## Step:  AIC=4179.41
## log(montant_sinistre) ~ drv_age_lic1 + drv_drv2 + vh_cyl + vh_din +
##     pol_bonus + vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - drv_drv2       1    0.6585 16768 4177.9
## - pol_duration2  1    0.9728 16769 4178.1
## - vh_din         1    2.2355 16770 4179.0
## - pol_bonus      1    2.8033 16771 4179.4
## <none>                       16768 4179.4
## - vh_cyl         1    2.9954 16771 4179.5
```

```
## - vh_fuel        2    9.9162 16778 4182.4
## - drv_age11      9   30.9517 16799 4183.2
## - drv_age_lic1   1   25.7104 16793 4195.5
##
## Step:  AIC=4177.88
## log(montant_sinistre) ~ drv_age_lic1 + vh_cyl + vh_din + pol_bonus +
##     vh_fuel + pol_duration2 + drv_age11
##
##                 Df Sum of Sq   RSS    AIC
## - pol_duration2  1    0.9950 16769 4176.6
## - vh_din         1    2.2300 16771 4177.4
## - pol_bonus      1    2.7851 16771 4177.8
## <none>                       16768 4177.9
## - vh_cyl         1    2.9806 16771 4178.0
## - vh_fuel        2    9.9345 16778 4180.9
## - drv_age11      9   30.9673 16799 4181.7
## - drv_age_lic1   1   25.8154 16794 4194.0
##
## Step:  AIC=4176.58
## log(montant_sinistre) ~ drv_age_lic1 + vh_cyl + vh_din + pol_bonus +
##     vh_fuel + drv_age11
##
##                Df Sum of Sq   RSS    AIC
## - vh_din        1    2.2144 16772 4176.1
## <none>                      16769 4176.6
## - vh_cyl        1    2.9836 16772 4176.7
## - pol_bonus     1    3.2811 16773 4176.9
## - vh_fuel       2    9.9437 16779 4179.6
## - drv_age11     9   31.1970 16801 4180.5
## - drv_age_lic1  1   26.0483 16795 4192.9
##
## Step:  AIC=4176.14
## log(montant_sinistre) ~ drv_age_lic1 + vh_cyl + pol_bonus + vh_fuel +
##     drv_age11
##
##                Df Sum of Sq   RSS    AIC
## - vh_cyl        1    0.8661 16773 4174.7
## <none>                      16772 4176.1
## - pol_bonus     1    3.3088 16775 4176.5
## - vh_fuel       2    8.6879 16780 4178.2
## - drv_age11     9   30.9026 16803 4179.9
## - drv_age_lic1  1   25.7303 16797 4192.2
##
## Step:  AIC=4174.75
## log(montant_sinistre) ~ drv_age_lic1 + pol_bonus + vh_fuel +
##     drv_age11
##
##                Df Sum of Sq   RSS    AIC
## <none>                      16773 4174.7
## - pol_bonus     1    3.2905 16776 4175.1
## - vh_fuel       2    7.8296 16780 4176.3
## - drv_age11     9   30.9067 16803 4178.5
## - drv_age_lic1  1   25.6997 16798 4190.8
```

```
##
## Call:
## lm(formula = log(montant_sinistre) ~ drv_age_lic1 + pol_bonus +
##     vh_fuel + drv_age11, data = base_severete)
##
## Coefficients:
##        (Intercept)        drv_age_lic1           pol_bonus     vh_fuelGasoline
##           6.837771            0.005817           -0.173341            0.028455
##      vh_fuelHybrid    drv_age11(18,22]    drv_age11(22,26]    drv_age11(26,30]
##          -0.592239           -0.211740           -0.187256           -0.176323
##   drv_age11(30,35]    drv_age11(35,40]    drv_age11(40,45]    drv_age11(45,50]
##          -0.210638           -0.245605           -0.203503           -0.189959
##   drv_age11(50,60]  drv_age11(60, Inf]
##          -0.218372           -0.200842
```

```
log_nml=lm(log(montant_sinistre) ~ drv_age_lic1 + pol_bonus +
             vh_fuel1 + drv_age11, data = base_severete)
summary(log_nml)
```

```
##
## Call:
## lm(formula = log(montant_sinistre) ~ drv_age_lic1 + pol_bonus +
##     vh_fuel1 + drv_age11, data = base_severete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7955 -0.7571  0.0825  0.8536  5.5893
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.851250   0.065799 104.124  < 2e-16 ***
## drv_age_lic1        0.005787   0.001369   4.229 2.37e-05 ***
## pol_bonus          -0.174586   0.113975  -1.532 0.125600
## vh_fuel1Hybrid     -0.604802   0.308319  -1.962 0.049831 *
## drv_age11(18,22]   -0.210953   0.056770  -3.716 0.000203 ***
## drv_age11(22,26]   -0.186504   0.058295  -3.199 0.001381 **
## drv_age11(26,30]   -0.175898   0.058877  -2.988 0.002818 **
## drv_age11(30,35]   -0.209623   0.059648  -3.514 0.000443 ***
## drv_age11(35,40]   -0.244234   0.061907  -3.945 8.02e-05 ***
## drv_age11(40,45]   -0.203488   0.067155  -3.030 0.002450 **
## drv_age11(45,50]   -0.189267   0.072589  -2.607 0.009135 **
## drv_age11(50,60]   -0.217314   0.072899  -2.981 0.002879 **
## drv_age11(60, Inf] -0.199989   0.086260  -2.318 0.020442 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.193 on 11791 degrees of freedom
## Multiple R-squared:  0.00299,    Adjusted R-squared:  0.001976
## F-statistic: 2.947 on 12 and 11791 DF,  p-value: 0.0004156
```

```
log_nml %>% tbl_regression(intercept = TRUE) %>% add_global_p()
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
```

```
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include ‘message = FALSE‘ in code chunk header.
```

Characteristic	Beta	95% CI	p-value
(Intercept)	6.9	6.7, 7.0	<0.001
drv_age_lic1	0.01	0.00, 0.01	<0.001
pol_bonus	-0.17	-0.40, 0.05	0.13
vh_fuel1			0.050
Diesel	—	—	
Hybrid	-0.60	-1.2, 0.00	
drv_age11			0.011
(-Inf,18]	—	—	
(18,22]	-0.21	-0.32, -0.10	
(22,26]	-0.19	-0.30, -0.07	
(26,30]	-0.18	-0.29, -0.06	
(30,35]	-0.21	-0.33, -0.09	
(35,40]	-0.24	-0.37, -0.12	
(40,45]	-0.20	-0.34, -0.07	
(45,50]	-0.19	-0.33, -0.05	
(50,60]	-0.22	-0.36, -0.07	
(60, Inf]	-0.20	-0.37, -0.03	

```
deviance(log_nml)
```

```
## [1] 16774.82
```

```
AIC(log_nml)
```

```
## [1] 37674.7
```

```
logLik(log_nml)
```

```
## 'log Lik.' -18823.35 (df=14)
```

```
######### etude les risidus##########
ggplot(data_frame(predt=log_nml$fitted.values,reside=log_nml$residuals),aes(x=predt,y=reside))+geom_poi
```

```
## Warning: ‘data_frame()‘ was deprecated in tibble 1.1.0.
## i Please use ‘tibble()‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
  # test pagan pour tester la constate des variance des risidus si p est superiure a 0.05
bptest(log_nml)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  log_nml
## BP = 6.6646, df = 12, p-value = 0.879
```

```r
######### prediction le cout de sinistre#########
##### gamma#####
pred$cou_pred=predict.glm(gamma4,pred,type="response")
summary(pred$cou_pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   561.4  1515.3  1580.0  1579.6  1653.4  2696.7
```

```r
####### log_normal#
sev_pred=predict(log_nml,pred,type="response")
sigma=summary(log_nml)$sigma
pnl=exp(sev_pred+sigma^2/2)
summary(pnl)
```

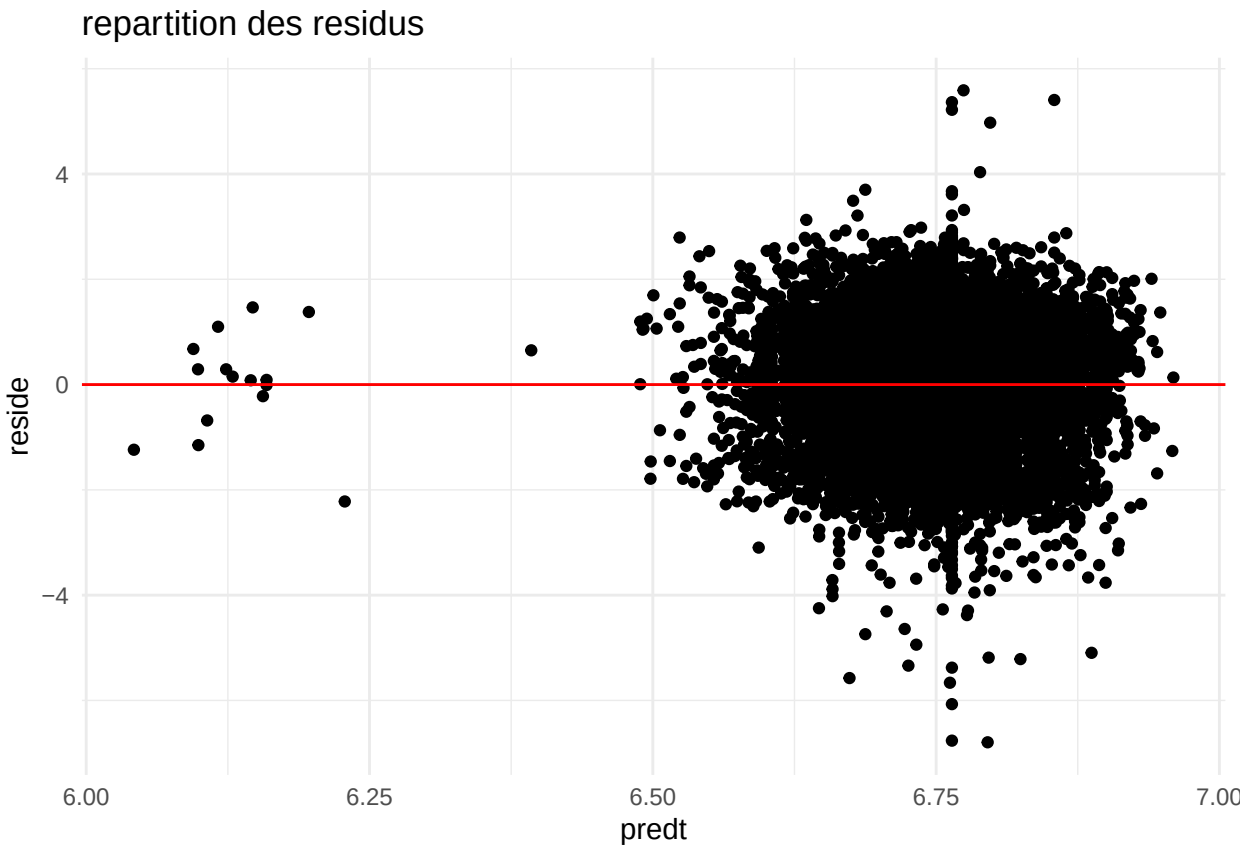
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   808.8  1644.0  1736.5  1737.5  1852.2  2791.1
```

```r
pred$tarif=prim_pure=prediction_freq*pnl
summary(pred$tarif)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   102.8   212.8   229.6   230.0   246.9   387.7
```

```r
base_tarif=pred %>% select(id_policy,tarif)
write.csv(base_tarif,file = "base_tarif.csv")
library(tinytex)
```

```
## Warning: le package 'tinytex' a été compilé avec la version R 4.3.2
```


repartition des residus
reside
predt
4
0
−4
6.00
6.25
6.50
6.75
7.00