

Projet Final

```
chooseCRANmirror(graphics=FALSE, ind=1)
knitr::opts_chunk$set(echo = TRUE)

library(summarytools)

## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp

## For best results, restart R session and update pander using devtools:: or remotes::install_github('rapporter/pander')

install.packages("actuar")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(actuar)

##
## Attaching package: 'actuar'

## The following object is masked from 'package:grDevices':
##
##      cm

install.packages("contrib.url")

## Warning: package 'contrib.url' is not available (for R version 3.6.2)

install.packages("Hmisc")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:summarytools':
##
##      label, label<-

## The following objects are masked from 'package:base':
##
##      format.pval, units

install.packages("corrplot")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(corrplot)

## corrplot 0.84 loaded

install.packages("sandwich")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(sandwich)
install.packages("lmtest")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(lmtest)

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

install.packages("summarytools")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages
```

```
library(summarytools)

install.packages("tree")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library("tree")
install.packages("AER")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(AER)

## Loading required package: car
## Loading required package: carData

install.packages("MASS")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(MASS)
install.packages("pscl")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(pscl)

## 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

install.packages("vcd")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages
```

```

library(vcd)

## Loading required package: grid

install.packages("car")

##
## The downloaded binary packages are in
## /var/folders/c6/qc0vj5mj08b03yl61kddtpsh0000gn/T//RtmpRucQnw/downloaded_packages

library(car)

setwd("/Users/Deneux/Desktop/Jeremy&Pierre")
base_sinistre<- read.csv(file="Jeremy&Pierre-PG_2017_CLAIMS_YEAR0.csv", sep = ";")
base_client <- read.csv(file="Jeremy&Pierre-PG_2017_YEAR0.csv", sep = ";")
base_sinistrePositive <- subset(base_sinistre, claim_amount >=0)

```

On remarque que certaines données de claims_year0 sont des characters et nous aimerions des entiers en particulier claim_amount. Or, les décimaux sont en virgules et pour le programme il faut des points, nous devons donc avant interchanger ces deux caractères

```

base_sinistrePositive$claim_amount <- gsub(",", ".", base_sinistrePositive$claim_amount)
base_sinistrePositive$claim_amount <- as.numeric(as.character(base_sinistrePositive$claim_amount))

```

On crée deux variables dans lesquelles d’une part on somme le coût total des sinistres connus dans l’année par véhicule et d’autre part le nombre de sinistre connus dans l’année par ce même véhicule (en ne prenant que les montants positifs).

```

library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##      select

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

## The following objects are masked from 'package:Hmisc':
##
##      src, summarize

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

```

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

montant_sinistre_annuel <- base_sinistrePositive %>% group_by(id_policy) %>% filter(claim_amount >= 0) %>% summarise(claim_amount = sum(claim_amount))

## `summarise()` ungrouping output (override with `.groups` argument)

nb_sinistres <- base_sinistrePositive %>% group_by(id_policy) %>% filter(claim_amount >= 0) %>% summarise(claim_amount = n())

## `summarise()` ungrouping output (override with `.groups` argument)
```

On rajoute nos deux variables sous forme de colonne à base_client sur une base que l'on appelle Base_etude. Je rajoute à la table la colonne Si NA écrire 0, donc attribué 0 à tout ceux qui ont pas eu de sinistre.

On effectue un changement du nom de la colonne par montant_sinistre_annuel

```
base_etude <- base_client %>% full_join(montant_sinistre_annuel, by="id_policy")
base_etude$claim_amount[is.na(base_etude$claim_amount)] <- 0
names(base_etude)[names(base_etude)=="claim_amount"] <- "montant_sinistre_annuel"
```

Je rajoute à la table la colonne nb_sinistres Attribue un 0 lorsqu'il n'y a pas de valeur en nombre de sinistre Changement du nom de la colonne

```
base_etude <- base_etude %>% full_join(nb_sinistres, by="id_policy")
base_etude$claim_amount[is.na(base_etude$claim_amount)] <- 0
names(base_etude)[names(base_etude)=="claim_amount"] <- "nb_sinistres"
```

Statistique descriptives et proportion

```
summary(base_etude$montant_sinistre_annuel)

##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
##      0.0      0.0      0.0     194.4      0.0 234104.0

table(base_etude$claim_amount)/nrow(base_etude)*100

## numeric(0)

table(base_etude$drv_sex1)/nrow(base_etude)*100

##
##      F      M
## 39.766 60.234
```

```

table(base_etude$drv_sex2)/nrow(base_etude)*100

##
##           F           M
## 66.810 20.325 12.865

str(base_etude)

## 'data.frame':    100000 obs. of  33 variables:
##  $ id_year          : Factor w/ 1 level "Year 0": 1 1 1 1 1
## 1 1 1 1 1 ...
##  $ id_policy         : Factor w/ 100000 levels "A00000004-V0
## 1",...: 1 2 3 4 5 6 7 8 9 10 ...
##  $ 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          : Factor w/ 2 levels "F","M": 2 1 2 2 1
## 1 1 1 2 1 ...
##  $ 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          : Factor w/ 3 levels "", "F", "M": 2 3 2
## 2 3 3 2 2 3 3 ...
##  $ drv_drv2          : Factor w/ 2 levels "No", "Yes": 2 1 1
## 1 1 2 1 1 2 2 ...
##  $ id_client         : Factor w/ 59833 levels "A00000004","A
## 00000007",...: 1 1 2 3 4 4 5 6 6 7 ...
##  $ vh_age            : int   16 16 7 11 6 6 5 13 13 1 ...
##  $ vh_cyl            : int  1781 1781 1870 1595 1997 1997 17
## 98 1905 1905 1560 ...
##  $ vh_din            : int   90 90 108 101 90 90 127 68 68 10
## 9 ...
##  $ vh_fuel           : Factor w/ 3 levels "Diesel","Gasoline
## ",...: 2 2 1 2 1 1 2 1 1 1 ...
##  $ vh_make           : Factor w/ 93 levels "ACL","ALFA ROMEO
## ",...: 90 90 69 7 64 64 52 20 20 20 ...
##  $ vh_model          : Factor w/ 934 levels "04-avr","09-mai
## ",...: 472 472 547 200 660 660 634 918 918 303 ...
##  $ 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           : Factor w/ 2 levels "Commercial","Tour
## ism": 2 2 2 2 2 2 2 2 2 2 ...
##  $ vh_value          : int  14407 14407 22450 20535 18550 18
## 550 22450 14773 14773 27100 ...
##  $ vh_weight         : int  1020 1020 1350 1195 1110 1110 10
## 80 1106 1106 1530 ...
##  $ id_vehicule       : Factor w/ 10 levels "V01","V02","V03"
## ,...: 1 2 1 1 1 2 1 1 2 1 ...
##  $ 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          : Factor w/ 4 levels "Maxi","Median1",.
.: 1 1 1 1 1 1 1 1 2 1 ...
## $ 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          : Factor w/ 4 levels "Biannual","Monthl
y",...: 4 1 2 4 2 2 4 4 3 2 ...
## $ pol_payd              : Factor w/ 2 levels "No","Yes": 1 1 1
1 1 1 1 2 1 1 ...
## $ pol_usage              : Factor w/ 4 levels "AllTrips","Profes
sional",...: 4 4 4 2 4 4 4 3 4 4 ...
## $ pol_insee_code        : Factor w/ 14759 levels "10003","1001"
,...: 8331 5680 3945 10380 9871 10244 8433 2928 9260 963 ...
## $ montant_sinistre_annuel: num   0 0 0 0 0 0 0 0 0 0 ...
## $ nb_sinistres          : num   0 0 0 0 0 0 0 0 0 0 ...

```

```
table(base_etude$pol_coverage)/nrow(base_etude)*100
```

```

##
##      Maxi Median1 Median2      Mini
## 64.854   9.320  17.316   8.510

```

```
table(base_etude$pol_usage)/nrow(base_etude)*100
```

```

##
##      AllTrips Professional      Retired  WorkPrivate
##         0.099         7.202      26.673         66.026

```

```
table(base_etude$vh_make)/nrow(base_etude)*100
```

```

##
##      ACL      ALFA ROMEO      ALPINE      APAL
ARO
##      0.010      0.400      0.002      0.001      0
.005
##      ASIA      AUDI      AUSTIN AUSTIN HEALEY      AUTOBIA
NCHI
##      0.002      1.504      0.014      0.001      0
.008
##      AUVERLAND      BABOULIN      BERTONE      BMW      BRE
MACH
##      0.005      0.005      0.001      1.522      0
.002
##      BUICK      CADILLAC      CHEVROLET      CHRYSLER      CIT
ROEN
##      0.002      0.001      0.273      0.195      16
.068
##      COUNIL      DACIA      DAEWOO      DAF      DAIH
ATSU
##      0.003      1.214      0.141      0.007      0
.073

```

## RARI ## .003	DAIMLER 0.010	DATSUN 0.004	DODGE 0.025	EBRO 0.039	FER 0
## ONDA ## .640	FIAT 2.930	FORD 4.379	FSO 0.002	GME 0.010	H 0
## GUAR ## .090	HOTCHKISS 0.019	HYUNDAI 0.625	ISUZU 0.025	IVECO 0.115	JA 0
## OVER ## .345	JEEP 0.172	KIA 0.429	LADA VAZ 0.077	LANCIA 0.148	LAND R 0
## RATI ## .004	LDV 0.003	LEXUS 0.019	LOTUS 0.008	MAHINDRA 0.007	MASE 0
## MG ## .032	MATRA 0.002	MAZDA 0.436	MEGA MERCEDES BENZ 0.001	3.047	0
## SSAN ## .761	MINI 0.396	MITSUBISHI 0.348	MORGAN 0.013	MORRIS 0.004	NI 1
## GGIO ## .004	OM 0.001	OPEL 3.590	PANHARD 0.001	PEUGEOT 19.780	PIA 0
## ILEY ## .001	PININFARINA 0.001	PONTIAC 0.012	PORSCHE 0.073	RENAULT 26.441	R 0
## SEAT ## .040	ROVER 0.234	SAAB 0.127	SANTANA 0.140	SAVIEM 0.038	1
## PUCH ## .003	SIMCA 0.014	SKODA 0.402	SMART 0.185	SSANGYONG 0.038	STEYR 0
## YOTA ## .295	SUBARU 0.042	SUZUKI 0.819	TALBOT 0.011	TEILHOL 0.007	TO 3
##	TRIUMPH	UMM	UNIC	VD 4 ROUES	VOLKSW


```

AGEN
##          0.022          0.002          0.103          0.070          5
.367
##          VOLVO          VW PORSCHE          WILLYS
##          0.494          0.003          0.013

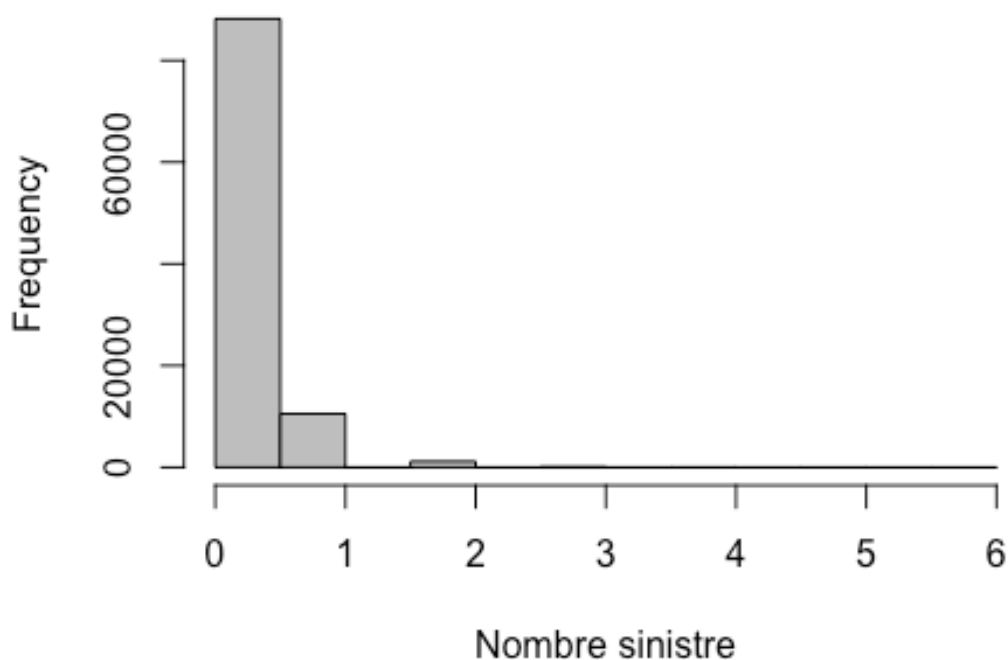
table(base_etude$vh_fuel)/nrow(base_etude)*100

##
## Diesel Gasoline Hybrid
## 55.256 44.666 0.078

hist(base_etude$nb_sinistres,freq=TRUE,plot=TRUE,xlab="Nombre sinist
re", main="Repartition du nombre des sinistres", col="grey")

```

Repartition du nombre des sinistres



important nombre du sinistre par age de la voiture #important diminution du nombre de sinistre en fonction de l'age du véhicule

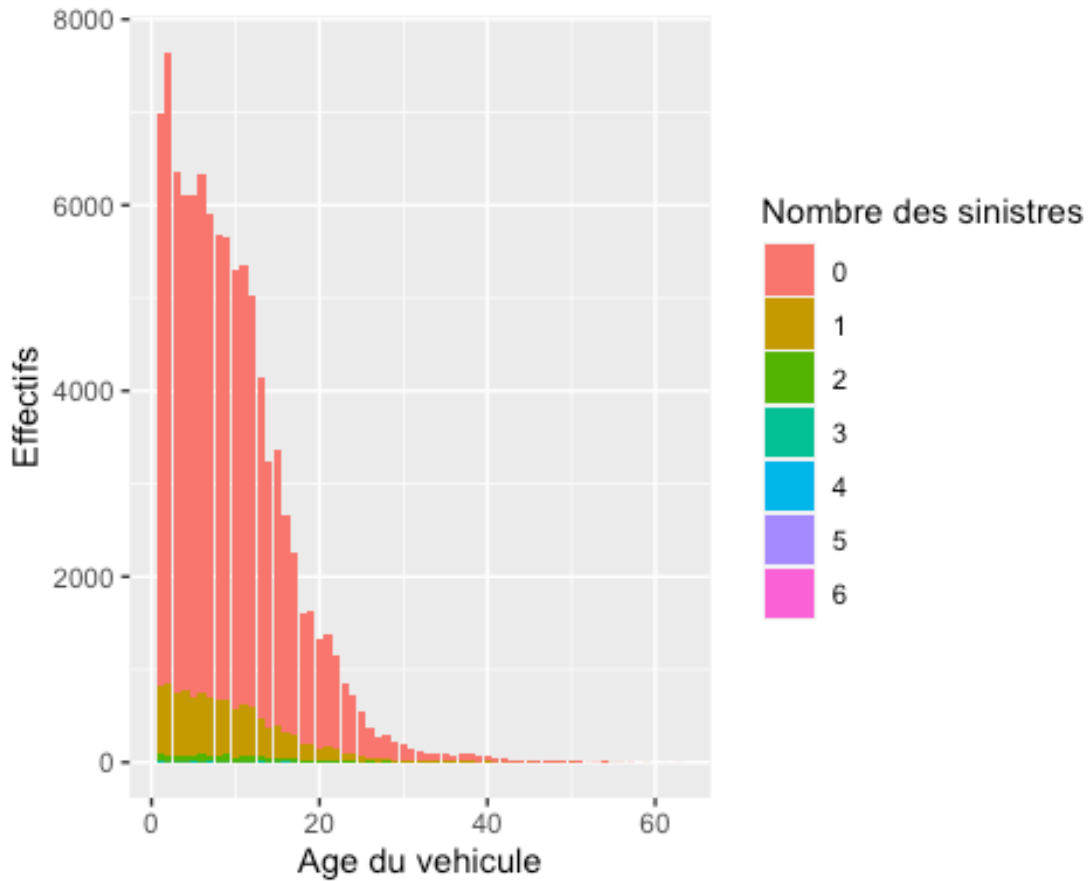
```

base_etude$nombre_sinistre <- as.character(base_etude$nb_sinistres)
base_etude$nombre_sinistre[base_etude$claim_nb == 0] <- "0"
base_etude$nombre_sinistre[base_etude$claim_nb == 1] <- "1"
base_etude$nombre_sinistre[base_etude$claim_nb == 2] <- "2"
base_etude$nombre_sinistre[base_etude$claim_nb == 3] <- "3"
base_etude$nombre_sinistre[base_etude$claim_nb == 4] <- "4"
base_etude$nombre_sinistre[base_etude$claim_nb == 5] <- "5"
base_etude$nombre_sinistre[base_etude$claim_nb == 6] <- "6"

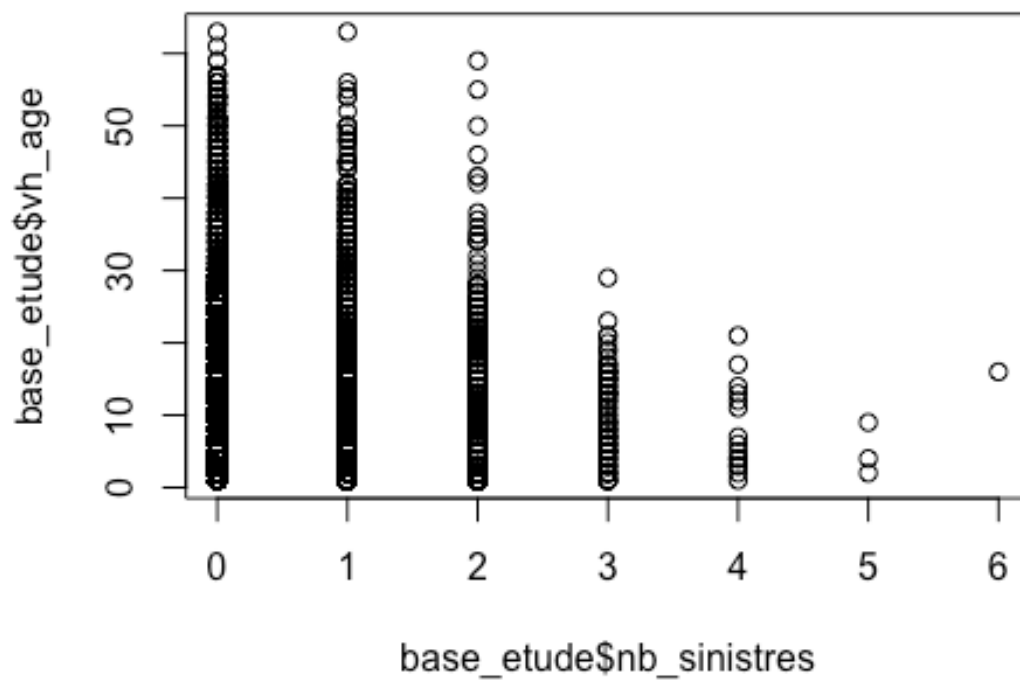
```

```
ggplot(base_etude) +
  aes(x = vh_age, fill = nombre_sinistre) +
  geom_bar() +
  xlab("Age du vehicule") +
  ylab("Effectifs") +
  labs(fill = "Nombre des sinistres")
```

Warning: position_stack requires non-overlapping x intervals



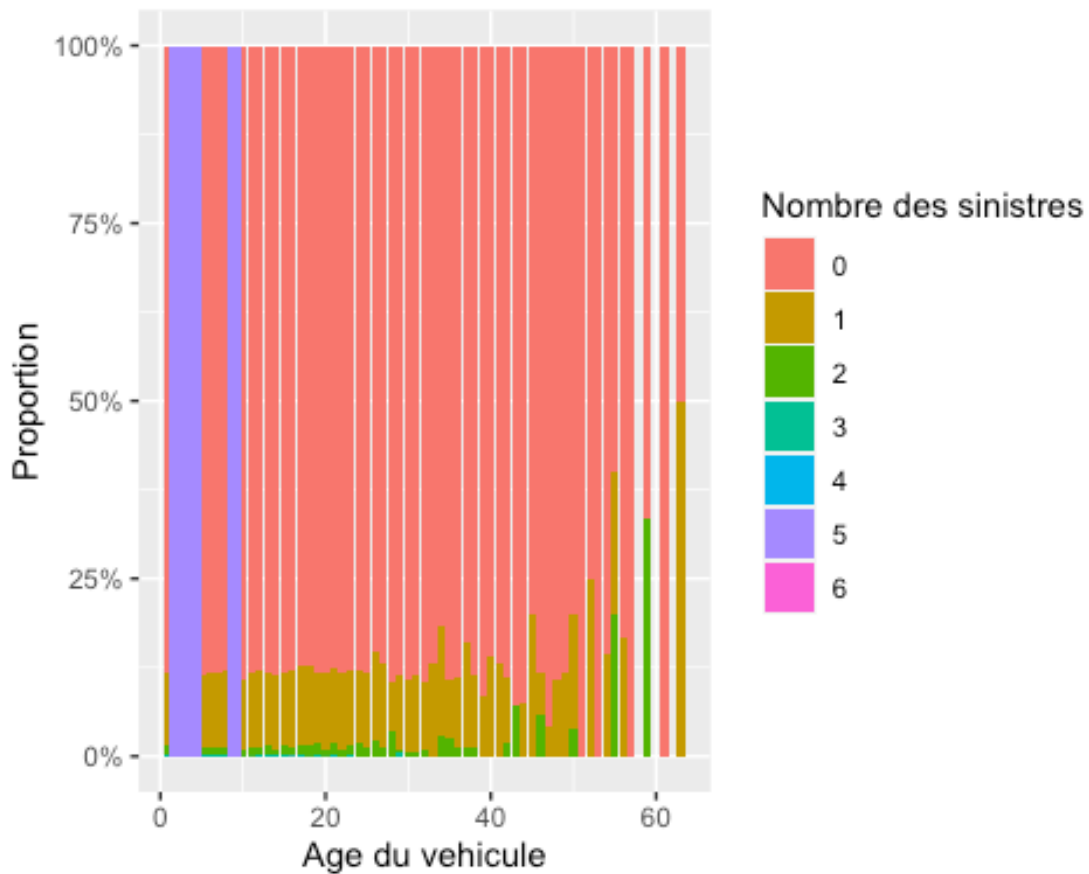
```
plot(base_etude$vh_age ~ base_etude$nb_sinistres)
```



proportion du nombre de sinistre comme au dessus mais en fonction de l'age

```
library(scales)
ggplot(base_etude) +
  aes(x = vh_age, fill = nombre_sinistre) +
  geom_bar(position = "fill") +
  xlab("Age du vehicule") +
  ylab("Proportion") +
  labs(fill = "Nombre des sinistres") +
  scale_y_continuous(labels = percent)

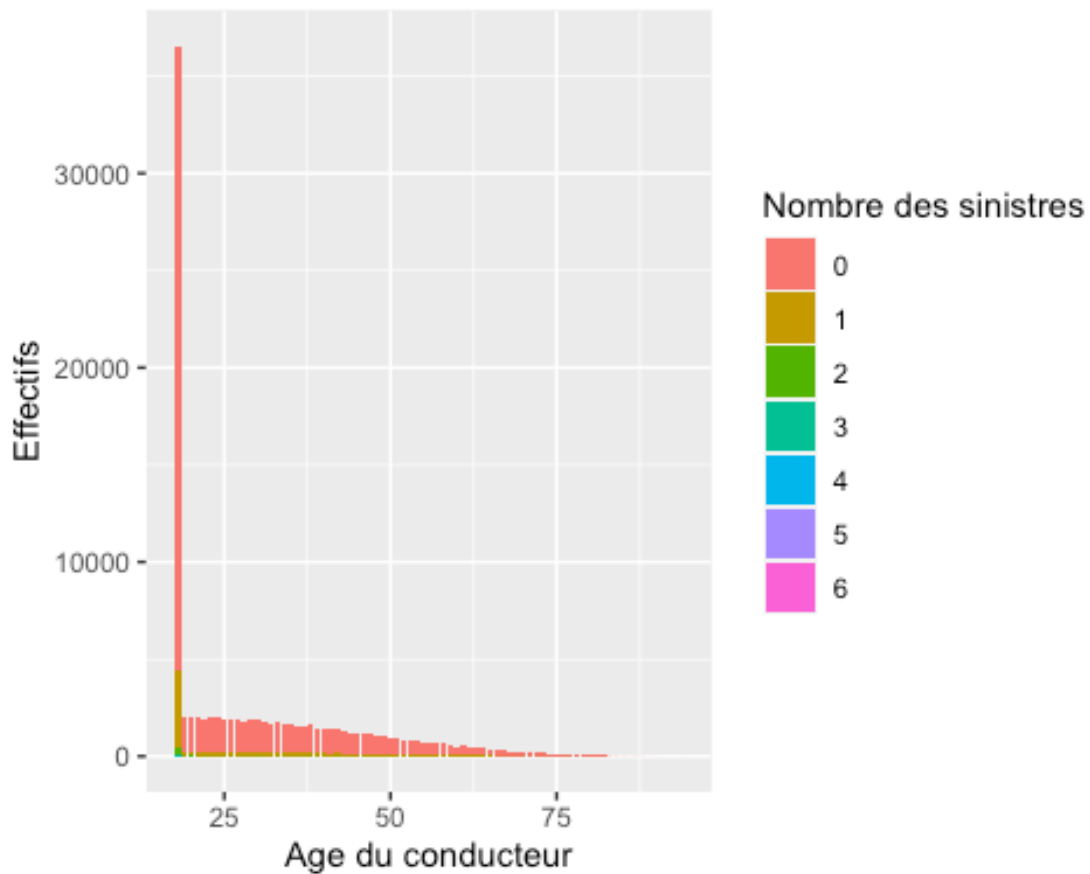
## Warning: position_stack requires non-overlapping x intervals
```



#nombre de sinistre par age en affectif puis en proportion.

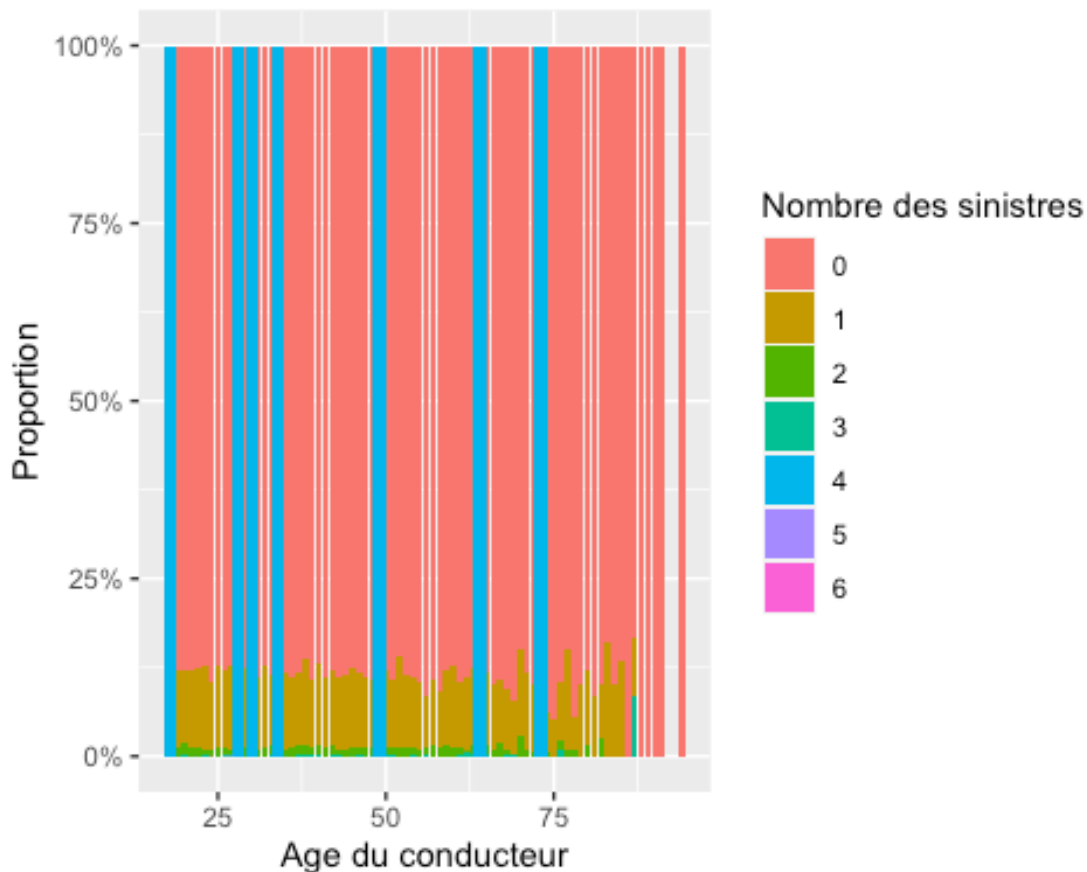
```
ggplot(base_etude) +
  aes(x = drv_age1, fill = nombre_sinistre) +
  geom_bar() +
  xlab("Age du conducteur") +
  ylab("Effectifs") +
  labs(fill = "Nombre des sinistres")
```

Warning: position_stack requires non-overlapping x intervals



```
ggplot(base_etude) +
  aes(x = drv_age1, fill = nombre_sinistre) +
  geom_bar(position = "fill") +
  xlab("Age du conducteur") +
  ylab("Proportion") +
  labs(fill = "Nombre des sinistres") +
  scale_y_continuous(labels = percent)

## Warning: position_stack requires non-overlapping x intervals
```



```
library(emmeans)

##
## Attaching package: 'emmeans'

## The following object is masked from 'package:actuar':
##
##      emm

moyenne <- emm(base_etude$nb_sinistres, 1)
variance <- emm(base_etude$nb_sinistres, 2)
summary(variance)

##      Length Class  Mode
## [1,] 100000 -none- numeric
## [2,]      1 -none- numeric

summary(moyenne)

##      Length Class  Mode
## [1,] 100000 -none- numeric
## [2,]      1 -none- numeric

base_etude$Dummy_sin[base_etude$nb_sinistres >=1] <- 1
base_etude$Dummy_sin[is.na(base_etude$Dummy_sin)] <- 0
sum(base_etude$Dummy_sin)/nrow(base_etude)
```

```
## [1] 0.11811

by(base_etude[,c("pol_bonus", "pol_duration", "pol_sit_duration", "drv_age1", "drv_age2",
                 "drv_age_lic1", "drv_age_lic2", "vh_age", "vh_cyl", "vh_din", "vh_sale_begin",
                 "vh_sale_end", "vh_speed", "vh_value", "vh_weight")], 1,
list(nbr_sin=base_etude$nb_sinistres), summary) #important

## nbr_sin: 0
##      pol_bonus      pol_duration      pol_sit_duration      drv_age1
## Min.      :0.500    Min.      : 1.00    Min.      : 1.000    Min.      :18.00
## 1st Qu.:0.500    1st Qu.: 4.00    1st Qu.: 1.000    1st Qu.:18.00
## Median :0.500    Median : 9.00    Median : 2.000    Median :25.00
## Mean      :0.537    Mean      :11.09    Mean      : 2.732    Mean      :30.72
## 3rd Qu.:0.500    3rd Qu.:16.00    3rd Qu.: 3.000    3rd Qu.:40.00
## Max.      :1.570    Max.      :40.00    Max.      :24.000    Max.      :94.00
##      drv_age2      drv_age_lic1      drv_age_lic2      vh_age
## Min.      : 0.00    Min.      : 0.00    Min.      : 0.000    Min.      : 1.000
## 1st Qu.: 0.00    1st Qu.: 0.00    1st Qu.: 0.000    1st Qu.: 4.000
## Median : 0.00    Median :24.00    Median : 0.000    Median : 8.000
## Mean      :15.54    Mean      :22.04    Mean      : 8.938    Mean      : 9.523
## 3rd Qu.:34.00    3rd Qu.:38.00    3rd Qu.:14.000    3rd Qu.:13.000
## Max.      :99.00    Max.      :71.00    Max.      :80.000    Max.      :63.000
##      vh_cyl      vh_din      vh_sale_begin      vh_sale_end
## Min.      : 0      Min.      :15.00    Min.      : 1.00    Min.      : 1.000
## 1st Qu.:1360    1st Qu.: 68.00    1st Qu.: 6.00    1st Qu.: 4.000
## Median :1587    Median : 87.00    Median :10.00    Median : 7.000
## Mean      :1647    Mean      : 91.42    Mean      :11.63    Mean      : 8.654
## 3rd Qu.:1910    3rd Qu.:109.00    3rd Qu.:15.00    3rd Qu.:12.000
## Max.      :6217    Max.      :555.00    Max.      :74.00    Max.      :55.000
##      vh_speed      vh_value      vh_weight
## Min.      : 25.0    Min.      : 0      Min.      : 0
## 1st Qu.:157.0    1st Qu.:11950    1st Qu.: 950
## Median :170.0    Median :16280    Median :1130
## Mean      :170.7    Mean      :18077    Mean      :1128
## 3rd Qu.:185.0    3rd Qu.:22100    3rd Qu.:1320
## Max.      :310.0    Max.      :145000    Max.      :7901
## -----
## nbr_sin: 1
##      pol_bonus      pol_duration      pol_sit_duration      drv_age1
## Min.      :0.5000    Min.      : 1.00    Min.      : 1.00    Min.      :18.00
## 1st Qu.:0.5000    1st Qu.: 4.00    1st Qu.: 1.00    1st Qu.:18.00
## Median :0.5000    Median : 9.00    Median : 2.00    Median :25.00
## Mean      :0.5391    Mean      :11.23    Mean      : 2.73    Mean      :30.21
## 3rd Qu.:0.5000    3rd Qu.:17.00    3rd Qu.: 3.00    3rd Qu.:39.00
## Max.      :1.6500    Max.      :38.00    Max.      :23.00    Max.      :87.00
##      drv_age2      drv_age_lic1      drv_age_lic2      vh_age
## Min.      : 0.00    Min.      : 0.00    Min.      : 0.000    Min.      : 1.000
## 1st Qu.: 0.00    1st Qu.: 0.00    1st Qu.: 0.000    1st Qu.: 4.000
```

```

## Median : 0.00      Median :23.00      Median : 0.000      Median : 8.000
## Mean    :15.72      Mean    :21.25      Mean    : 9.065      Mean    : 9.546
## 3rd Qu.:35.00      3rd Qu.:37.00      3rd Qu.:15.000      3rd Qu.:13.000
## Max.    :97.00      Max.    :73.00      Max.    :78.000      Max.    :63.000
##      vh_cyl      vh_din      vh_sale_begin      vh_sale_end
vh_speed
## Min.    : 0      Min.    : 15.00      Min.    : 1.00      Min.    : 1.000
Min.    : 25
## 1st Qu.:1360      1st Qu.: 68.00      1st Qu.: 6.00      1st Qu.: 4.000
1st Qu.:157
## Median :1587      Median : 87.00      Median :10.00      Median : 7.000
Median :170
## Mean    :1649      Mean    : 91.77      Mean    :11.65      Mean    : 8.663
Mean    :171
## 3rd Qu.:1910      3rd Qu.:109.00      3rd Qu.:15.00      3rd Qu.:12.000
3rd Qu.:185
## Max.    :5700      Max.    :486.00      Max.    :69.00      Max.    :55.000
Max.    :310
##      vh_value      vh_weight
## Min.    : 648      Min.    : 0
## 1st Qu.: 11952      1st Qu.: 950
## Median : 16300      Median :1138
## Mean    : 18192      Mean    :1136
## 3rd Qu.: 22300      3rd Qu.:1320
## Max.    :107900      Max.    :7901
## -----
## nbr_sin: 2
##      pol_bonus      pol_duration      pol_sit_duration      drv_age1
## Min.    :0.5000      Min.    : 1.00      Min.    : 1.000      Min.    :18.00
## 1st Qu.:0.5000      1st Qu.: 4.00      1st Qu.: 1.000      1st Qu.:18.00
## Median :0.5000      Median : 9.00      Median : 2.000      Median :25.00
## Mean    :0.5351      Mean    :11.13      Mean    : 2.722      Mean    :30.34
## 3rd Qu.:0.5000      3rd Qu.:17.00      3rd Qu.: 3.000      3rd Qu.:39.00
## Max.    :1.4700      Max.    :38.00      Max.    :19.000      Max.    :82.00
##      drv_age2      drv_age_lic1      drv_age_lic2      vh_age
## Min.    : 0.00      Min.    : 0.00      Min.    : 0.000      Min.    : 1.00
## 1st Qu.: 0.00      1st Qu.: 0.00      1st Qu.: 0.000      1st Qu.: 5.00
## Median : 0.00      Median :24.00      Median : 0.000      Median : 9.00
## Mean    :15.45      Mean    :21.42      Mean    : 8.821      Mean    :10.12
## 3rd Qu.:35.00      3rd Qu.:37.00      3rd Qu.:15.000      3rd Qu.:14.00
## Max.    :87.00      Max.    :66.00      Max.    :68.000      Max.    :59.00
##      vh_cyl      vh_din      vh_sale_begin      vh_sale_end
## Min.    : 479      Min.    : 15.00      Min.    : 1.00      Min.    : 1.000
## 1st Qu.:1360      1st Qu.: 68.00      1st Qu.: 6.00      1st Qu.: 4.000
## Median :1587      Median : 86.00      Median :11.00      Median : 8.000
## Mean    :1643      Mean    : 89.86      Mean    :12.24      Mean    : 9.228
## 3rd Qu.:1929      3rd Qu.:109.00      3rd Qu.:16.00      3rd Qu.:13.000
## Max.    :4266      Max.    :279.00      Max.    :74.00      Max.    :54.000
##      vh_speed      vh_value      vh_weight

```



```

## Min. : 90.0 Min. : 648 Min. : 0
## 1st Qu.:156.0 1st Qu.:11831 1st Qu.: 932
## Median :170.0 Median :16000 Median :1110
## Mean :169.8 Mean :17737 Mean :1112
## 3rd Qu.:183.0 3rd Qu.:21400 3rd Qu.:1320
## Max. :250.0 Max. :58700 Max. :2800
## -----
## nbr_sin: 3
## pol_bonus pol_duration pol_sit_duration drv_age1
## Min. :0.5000 Min. : 1.00 Min. : 1.00 Min. :18.00
## 1st Qu.:0.5000 1st Qu.: 4.00 1st Qu.: 1.00 1st Qu.:18.00
## Median :0.5000 Median :10.50 Median : 2.00 Median :25.50
## Mean :0.5366 Mean :11.77 Mean : 2.67 Mean :31.25
## 3rd Qu.:0.5000 3rd Qu.:18.75 3rd Qu.: 3.00 3rd Qu.:41.50
## Max. :0.9500 Max. :29.00 Max. :17.00 Max. :87.00
## drv_age2 drv_age_lic1 drv_age_lic2 vh_age
## Min. : 0.00 Min. : 0.00 Min. : 0.000 Min. : 1.000
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000 1st Qu.: 5.000
## Median : 0.00 Median :27.00 Median : 0.000 Median : 9.000
## Mean :15.73 Mean :22.07 Mean : 8.792 Mean : 9.698
## 3rd Qu.:35.75 3rd Qu.:39.00 3rd Qu.:16.000 3rd Qu.:14.750
## Max. :79.00 Max. :60.00 Max. :59.000 Max. :29.000
## vh_cyl vh_din vh_sale_begin vh_sale_end
## Min. : 602 Min. : 29.00 Min. : 1.00 Min. : 1.000
## 1st Qu.:1232 1st Qu.: 66.50 1st Qu.: 7.00 1st Qu.: 4.250
## Median :1560 Median : 82.00 Median :11.00 Median : 8.000
## Mean :1608 Mean : 93.33 Mean :11.55 Mean : 8.849
## 3rd Qu.:1909 3rd Qu.:109.00 3rd Qu.:15.75 3rd Qu.:14.000
## Max. :2996 Max. :265.00 Max. :41.00 Max. :30.000
## vh_speed vh_value vh_weight
## Min. :115.0 Min. : 3299 Min. : 0.0
## 1st Qu.:160.2 1st Qu.:11288 1st Qu.: 892.5
## Median :170.0 Median :15816 Median :1105.0
## Mean :173.6 Mean :17956 Mean :1115.4
## 3rd Qu.:186.8 3rd Qu.:23488 3rd Qu.:1318.2
## Max. :250.0 Max. :55190 Max. :1954.0
## -----
## nbr_sin: 4
## pol_bonus pol_duration pol_sit_duration drv_age1
## Min. :0.5000 Min. : 1.0 Min. : 1.000 Min. :18.00
## 1st Qu.:0.5000 1st Qu.: 5.5 1st Qu.: 1.000 1st Qu.:18.00
## Median :0.5000 Median :10.0 Median : 2.000 Median :18.00
## Mean :0.5227 Mean :11.4 Mean : 3.267 Mean :29.33
## 3rd Qu.:0.5200 3rd Qu.:13.5 3rd Qu.: 4.000 3rd Qu.:32.00
## Max. :0.6000 Max. :30.0 Max. :15.000 Max. :73.00
## drv_age2 drv_age_lic1 drv_age_lic2 vh_age
## Min. : 0.00 Min. : 0.0 Min. : 0.0 Min. : 1.0 Min
. : 954

```

```

## 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0 1st Qu.: 3.5 1st
Qu.:1298
## Median :21.00 Median : 0.0 Median : 1.0 Median : 6.0 Med
ian :1461
## Mean :25.67 Mean :13.4 Mean :15.2 Mean : 8.2 Mea
n :1554
## 3rd Qu.:51.50 3rd Qu.:25.5 3rd Qu.:32.0 3rd Qu.:12.5 3rd
Qu.:1903
## Max. :81.00 Max. :51.0 Max. :61.0 Max. :21.0 Max
. :2184
## vh_din vh_sale_begin vh_sale_end vh_speed
## Min. : 51.00 Min. : 3.00 Min. : 1.000 Min. :137.0
## 1st Qu.: 71.50 1st Qu.: 5.00 1st Qu.: 2.500 1st Qu.:159.5
## Median : 86.00 Median : 9.00 Median : 5.000 Median :165.0
## Mean : 89.47 Mean :10.07 Mean : 7.067 Mean :168.8
## 3rd Qu.: 97.50 3rd Qu.:14.00 3rd Qu.:10.500 3rd Qu.:176.0
## Max. :150.00 Max. :25.00 Max. :21.000 Max. :210.0
## vh_value vh_weight
## Min. : 8700 Min. : 815
## 1st Qu.:14136 1st Qu.: 994
## Median :15250 Median :1080
## Mean :18494 Mean :1167
## 3rd Qu.:24202 3rd Qu.:1322
## Max. :31700 Max. :1590
## -----
## nbr_sin: 5
## pol_bonus pol_duration pol_sit_duration drv_age1
drv_age2
## Min. :0.5000 Min. :1 Min. :1.000 Min. :18.00
Min. : 0.0
## 1st Qu.:0.5700 1st Qu.:3 1st Qu.:1.000 1st Qu.:22.00
1st Qu.: 0.0
## Median :0.6400 Median :5 Median :1.000 Median :26.00
Median : 0.0
## Mean :0.5933 Mean :5 Mean :1.333 Mean :23.67
Mean :17.0
## 3rd Qu.:0.6400 3rd Qu.:7 3rd Qu.:1.500 3rd Qu.:26.50
3rd Qu.:25.5
## Max. :0.6400 Max. :9 Max. :2.000 Max. :27.00
Max. :51.0
## drv_age_lic1 drv_age_lic2 vh_age vh_cyl
vh_din
## Min. : 0.00 Min. : 0.00 Min. :2.0 Min. :1149 Min
. : 55.00
## 1st Qu.: 5.00 1st Qu.: 0.00 1st Qu.:3.0 1st Qu.:1274 1st
Qu.: 57.00
## Median :10.00 Median : 0.00 Median :4.0 Median :1398 Med
ian : 59.00
## Mean :15.33 Mean :10.67 Mean :5.0 Mean :1472 Mea

```

```

n      : 81.67
## 3rd Qu.:23.00    3rd Qu.:16.00    3rd Qu.:6.5    3rd Qu.:1634    3rd
Qu.: 95.00
## Max.      :36.00    Max.      :32.00    Max.      :9.0    Max.      :1870    Max
.      :131.00
##  vh_sale_begin    vh_sale_end        vh_speed        vh_value
## Min.      : 2.000    Min.      :1.000    Min.      :151.0    Min.      :10138
## 1st Qu.: 4.000    1st Qu.:1.000    1st Qu.:152.5    1st Qu.:10994
## Median : 6.000    Median :1.000    Median :154.0    Median :11850
## Mean      : 6.333    Mean      :2.333    Mean      :166.7    Mean      :16863
## 3rd Qu.: 8.500    3rd Qu.:3.000    3rd Qu.:174.5    3rd Qu.:20225
## Max.      :11.000    Max.      :5.000    Max.      :195.0    Max.      :28600
##    vh_weight
## Min.      : 823.0
## 1st Qu.: 851.5
## Median : 880.0
## Mean      :1044.3
## 3rd Qu.:1155.0
## Max.      :1430.0
## -----
## nbr_sin: 6
##    pol_bonus    pol_duration pol_sit_duration    drv_age1    drv
_age2
## Min.      :0.6    Min.      :9    Min.      :3    Min.      :24    Min.
:0
## 1st Qu.:0.6    1st Qu.:9    1st Qu.:3    1st Qu.:24    1st Qu
.:0
## Median :0.6    Median :9    Median :3    Median :24    Median
:0
## Mean      :0.6    Mean      :9    Mean      :3    Mean      :24    Mean
:0
## 3rd Qu.:0.6    3rd Qu.:9    3rd Qu.:3    3rd Qu.:24    3rd Qu
.:0
## Max.      :0.6    Max.      :9    Max.      :3    Max.      :24    Max.
:0
##    drv_age_lic1  drv_age_lic2    vh_age    vh_cyl    vh_d
in
## Min.      :10    Min.      :0    Min.      :16    Min.      :1870    Min.      :
66
## 1st Qu.:10    1st Qu.:0    1st Qu.:16    1st Qu.:1870    1st Qu.:
66
## Median :10    Median :0    Median :16    Median :1870    Median :
66
## Mean      :10    Mean      :0    Mean      :16    Mean      :1870    Mean      :
66
## 3rd Qu.:10    3rd Qu.:0    3rd Qu.:16    3rd Qu.:1870    3rd Qu.:
66
## Max.      :10    Max.      :0    Max.      :16    Max.      :1870    Max.      :
66

```

```

##   vh_sale_begin  vh_sale_end    vh_speed      vh_value      vh_we
ight
##   Min.      :21    Min.      :15    Min.      :161    Min.      :12120    Min.
:905
##   1st Qu.:21    1st Qu.:15    1st Qu.:161    1st Qu.:12120    1st Qu.
:905
##   Median :21    Median :15    Median :161    Median :12120    Median
:905
##   Mean   :21    Mean   :15    Mean   :161    Mean   :12120    Mean
:905
##   3rd Qu.:21    3rd Qu.:15    3rd Qu.:161    3rd Qu.:12120    3rd Qu.
:905
##   Max.    :21    Max.    :15    Max.    :161    Max.    :12120    Max.
:905

by(base_etude[,c("pol_coverage", "pol_pay_freq", "pol_payd", "pol_insee
_code", "drv_drv2",
                 "drv_sex1", "drv_sex2", "vh_fuel", "vh_make", "vh_model
", "vh_type")], list(nbr_sin=base_etude$nb_sinistres), summary)

## nbr_sin: 0
##   pol_coverage      pol_pay_freq    pol_payd      pol_insee_code  dr
v_drv2
##   Maxi      :57230    Biannual :25949    No :84480    31555 : 457    No
:59015
##   Median1: 8262    Monthly  :26673    Yes: 3709    67482 : 446    Ye
s:29174
##   Median2:15176    Quarterly: 2216                6088 : 406
##   Mini    : 7521    Yearly    :33351                59350 : 314
##                                     75115 : 310
##                                     75116 : 295
##                                     (Other):85961
##   drv_sex1  drv_sex2      vh_fuel      vh_make      vh_mo
del
##   F:34988    :58948    Diesel  :48706    RENAULT    :23312    CLIO    :
6390
##   M:53201    F:17894    Gasoline:39421    PEUGEOT    :17482    206      :
3503
##               M:11347    Hybrid   : 62    CITROEN    :14166    TWINGO   :
3150
##               VOLKSWAGEN: 4752    SCENIC   :
2934
##               FORD      : 3847    MEGANE   :
2381
##               OPEL      : 3193    307      :
2272
##               (Other)   :21437    (Other):
67559
##           vh_type
##   Commercial: 8851

```

```

## Tourism :79338
##
##
##
##
##
## -----
## nbr_sin: 1
## pol_coverage pol_pay_freq pol_payd pol_insee_code drv_
drv2
## Maxi :6785 Biannual :3220 No :10100 31555 : 53 No :
7021
## Median1: 963 Monthly :3201 Yes: 433 67482 : 51 Yes:
3512
## Median2:1908 Quarterly: 257 44109 : 50
## Mini : 877 Yearly :3855 59350 : 48
## 6088 : 44
## 75115 : 44
## (Other):10243
## drv_sex1 drv_sex2 vh_fuel vh_make vh_model
## F:4270 :7014 Diesel :5843 RENAULT :2779 CLIO : 736
## M:6263 F:2158 Gasoline:4674 PEUGEOT :2048 206 : 436
## M:1361 Hybrid : 16 CITROEN :1698 TWINGO : 381
## VOLKSWAGEN: 545 SCENIC : 353
## FORD : 478 MEGANE : 287
## TOYOTA : 403 307 : 260
## (Other) :2582 (Other):8080
## vh_type
## Commercial:1015
## Tourism :9518
##
##
##
## -----
## nbr_sin: 2
## pol_coverage pol_pay_freq pol_payd pol_insee_code drv_drv2
drv_sex1
## Maxi :764 Biannual :338 No :1110 31555 : 9 No :770
F:451
## Median1: 85 Monthly :344 Yes: 43 67482 : 8 Yes:383
M:702
## Median2:201 Quarterly: 45 49007 : 6
## Mini :103 Yearly :426 33063 : 5
## 37261 : 5
## 75113 : 5
## (Other):1115
## drv_sex2 vh_fuel vh_make vh_model

```

```

vh_type
## :770 Diesel :649 RENAULT :311 CLI0 : 92 Commer
cial: 123
## F:242 Gasoline:504 PEUGEOT :230 206 : 51 Touris
m :1030
## M:141 Hybrid : 0 CITROEN :188 TWINGO : 44
## VOLKSWAGEN : 62 C3 : 36
## FORD : 49 307 : 31
## MERCEDES BENZ: 44 SCENIC : 31
## (Other) :269 (Other):868
## -----
## nbr_sin: 3
## pol_coverage pol_pay_freq pol_payd pol_insee_code drv_drv2
drv_sex1
## Maxi :67 Biannual :22 No :104 26281 : 2 No :69
F:48
## Median1: 5 Monthly :34 Yes: 2 29019 : 2 Yes:37
M:58
## Median2:27 Quarterly: 2 64445 : 2
## Mini : 7 Yearly :48 10344 : 1
## 11286 : 1
## 13056 : 1
## (Other):97
## drv_sex2 vh_fuel vh_make vh_model vh_typ
e
## :68 Diesel :48 RENAULT :33 CLI0 :10 Commercial:
4
## F:25 Gasoline:58 PEUGEOT :17 TWINGO : 7 Tourism :1
02
## M:13 Hybrid : 0 CITROEN :14 106 : 6
## OPEL : 7 MEGANE : 5
## AUDI : 5 306 : 4
## VOLKSWAGEN: 5 CORSA : 4
## (Other) :25 (Other):70
## -----
## nbr_sin: 4
## pol_coverage pol_pay_freq pol_payd pol_insee_code drv_drv2 d
rv_sex1
## Maxi :5 Biannual :5 No :15 13211 :1 No :7 F
:6
## Median1:5 Monthly :2 Yes: 0 16393 :1 Yes:8 M
:9
## Median2:3 Quarterly:1 27198 :1
## Mini :2 Yearly :7 38185 :1
## 39552 :1
## 44128 :1
## (Other):9
## drv_sex2 vh_fuel vh_make vh_model vh_type
## :7 Diesel :7 RENAULT :3 106 :1 Commercial: 0

```

```

## F:6      Gasoline:8  VOLKSWAGEN:3  307      :1  Tourism    :15
## M:2      Hybrid   :0  CITROEN     :2  ALFA 159:1
##          PEUGEOT   :2  C1         :1
##          ALFA ROMEO:1  C3         :1
##          DACIA     :1  CLIO       :1
##          (Other)   :3  (Other)    :9
## -----
## nbr_sin: 5
## pol_coverage    pol_pay_freq pol_payd pol_insee_code drv_drv2 d
rv_sex1
## Maxi      :3      Biannual  :0      No :3      33448   :1      No :2      F
:3
## Median1:0      Monthly   :2      Yes:0      35289   :1      Yes:1     M
:0
## Median2:0      Quarterly:0                      44162   :1
## Mini      :0      Yearly   :1                      10003   :0
##          :0                      1001     :0
##          :0                      10018   :0
##          :0                      (Other):0
## drv_sex2      vh_fuel      vh_make      vh_model      vh_type
## :2      Diesel   :2      RENAULT   :2      107     :1      Commercial:0
## F:0      Gasoline:1      PEUGEOT   :1      SCENIC  :1      Tourism   :3
## M:1      Hybrid   :0      ACL       :0      TWINGO  :1
##          :0                      ALFA ROMEO:0      04-avr   :0
##          :0                      ALPINE    :0      09-mai   :0
##          :0                      APAL     :0      09-mars  :0
##          :0                      (Other)  :0      (Other):0
## -----
## nbr_sin: 6
## pol_coverage    pol_pay_freq pol_payd pol_insee_code drv_drv2 d
rv_sex1
## Maxi      :0      Biannual  :1      No :1      93049   :1      No :1      F
:0
## Median1:0      Monthly   :0      Yes:0      10003   :0      Yes:0     M
:1
## Median2:1      Quarterly:0                      1001     :0
## Mini      :0      Yearly   :0                      10018   :0
##          :0                      10019   :0
##          :0                      10020   :0
##          :0                      (Other):0
## drv_sex2      vh_fuel      vh_make      vh_model      vh_type
## :1      Diesel   :1      RENAULT   :1      CLIO    :1      Commercial:0
## F:0      Gasoline:0      ACL       :0      04-avr   :0      Tourism   :1
## M:0      Hybrid   :0      ALFA ROMEO:0      09-mai   :0
##          :0                      ALPINE    :0      09-mars  :0
##          :0                      APAL     :0      10       :0
##          :0                      ARO      :0      100      :0
##          :0                      (Other)  :0      (Other):0

```

```

library(Hmisc)

e_etude$pol_usage,
ofessional" = "AllTrips")

#vh_make

base_etude$vh_make2 <- as.character(base_
e_etude$vh_make)

base_etude$vh_make2[!(base_etude$vh_mak
e %in% c("RENAULT", "PEUGEOT", "CITROEN", "VOLKSWAGEN", "FORD", "OPEL", "T
OYOTA", "MERCEDES BENZ", "FIAT"))] <- "AUTRE"
base_etude$vh_make2 <- as.factor(base_e
tude$vh_make2)

base_etude$pol_duration2 <- cut(base_et
ude$pol_duration, c(-Inf, 21, Inf))
base_etude$pol_usage1.2 <- fct_recode(b
ase_etude$pol_usage, "Professional" = "AllTrips")
base_etude$drv_age1.2 <- cut(base_etude
$drv_age1, c(-Inf, 35, 41, 46, 50, 54, 58, 63, 68, 75, Inf))
base_etude$vh_fuel2 <- fct_recode(base_
etude$vh_fuel, "Diesel" = "Gasoline")
base_etude$vh_age2 <- cut(base_etude$vh
_age, c(-Inf, 17, Inf))

names(base_etude)

## [1] "id_year" "id_policy"
## [3] "drv_age1" "drv_age_lic1"
## [5] "drv_sex1" "drv_age2"
## [7] "drv_age_lic2" "drv_sex2"
## [9] "drv_drv2" "id_client"
## [11] "vh_age" "vh_cyl"
## [13] "vh_din" "vh_fuel"
## [15] "vh_make" "vh_model"
## [17] "vh_sale_begin" "vh_sale_end"
## [19] "vh_speed" "vh_type"
## [21] "vh_value" "vh_weight"
## [23] "id_vehicle" "pol_bonus"
## [25] "pol_coverage" "pol_duration"
## [27] "pol_sit_duration" "pol_pay_freq"
## [29] "pol_payd" "pol_usage"

```



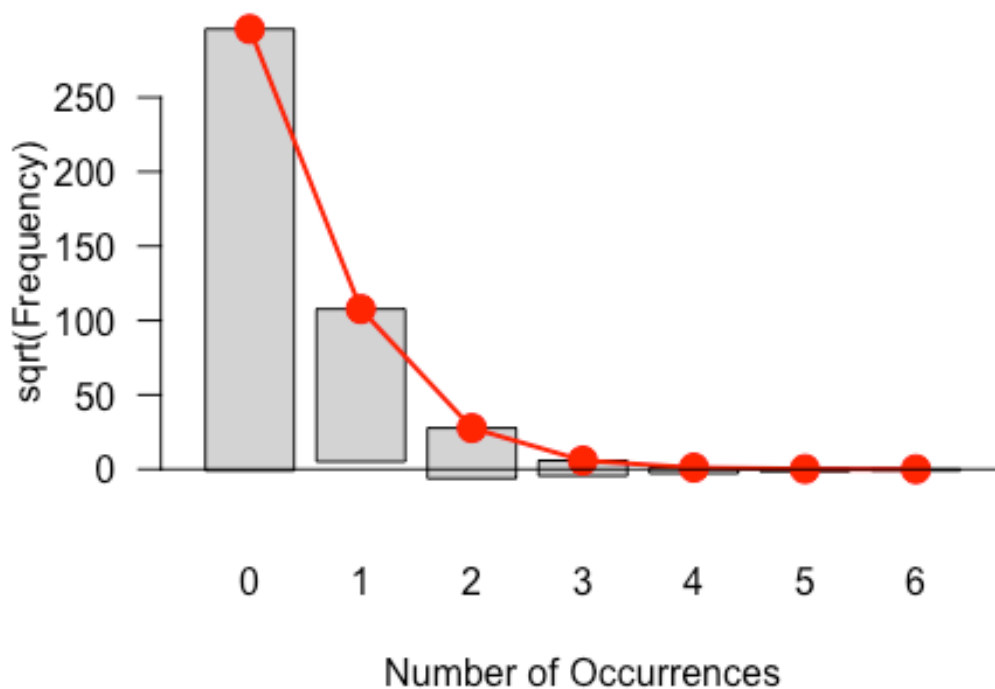
```
## [31] "pol_insee_code"      "montant_sinistre_annuel"
## [33] "nb_sinistres"        "nombre_sinistre"
## [35] "Dummy_sin"          "pol_usagee"
## [37] "vh_make2"           "pol_duration2"
## [39] "pol_usage1.2"        "drv_age1.2"
## [41] "vh_fuel2"           "vh_age2"

nom_variable <- c("vh_age2", "drv_drv2",
"vh_dinn", "vh_valuee", "pol_coverage",
"pol_payd", "vh_fuel2",
, "vh_weightt", "pol_usage1.2",
"pol_usage1.2", "drv_a
ge1.2", "vh_agee", "vh_make2", "vh_speedd"
, "drv_sex2", "vh_cyll"
, "vh_model", "vh_type"
, "pol_pay_freq")

v<-base_etude$nb_sinistres
moyenne <- emm(v, order=1)
variance <- emm(v, order=2)

goodfit_pois<- goodfit(base_etude$nb_sinistres, type = c("poisson")
,
method = c("ML",
"MinChisq"), par = NULL)

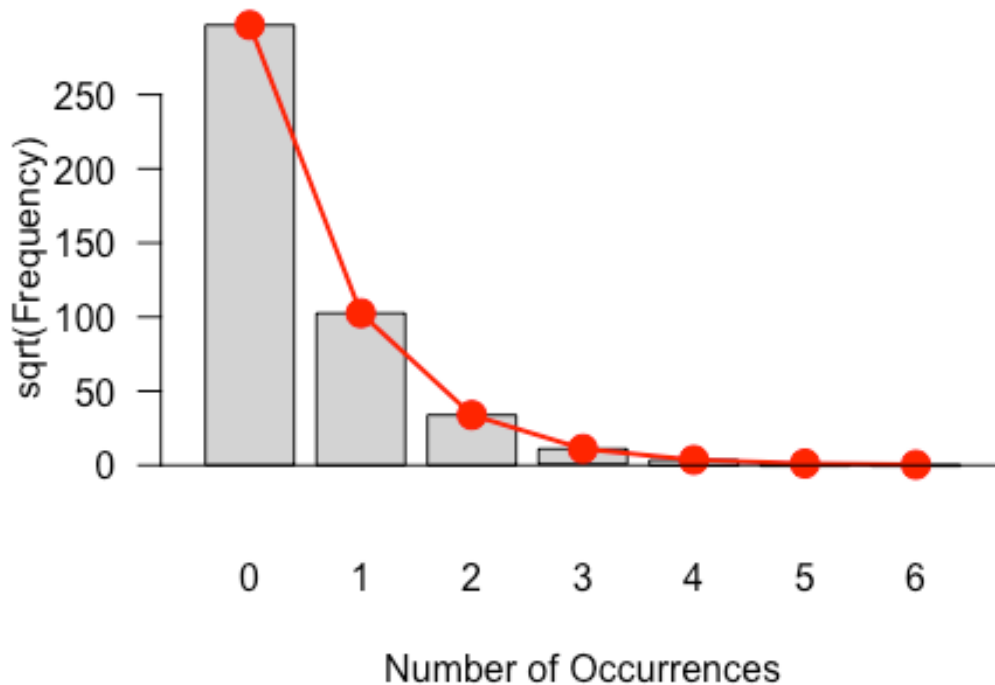
plot(goodfit_pois) #loi de poisson
```



```

goodfit_NB <- goodfit(base_etude$nb_sin
istres, type = c( "nbinomial"),
                                method = c("ML",
"MinChisq"), par = NULL)
plot(goodfit_NB)

```



cherche le modèle avec critère AIC le plus faible en faisant une régression de poisson

```

fpois <- glm(nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_b
egin + vh_value +
                                pol_coverage + pol_payd
+ drv_age2 + vh_fuel + vh_sale_end +
                                vh_weight + pol_duration
+ pol_usage + drv_age1 + drv_age_lic2 +
                                vh_age + vh_speed + pol
_sit_duration + drv_age_lic1 +
                                drv_sex2 + vh_cyl + vh_
type + pol_bonus + pol_pay_freq, family=poisson("log"), data=base_et
ude)
summary(fpois)

##
## Call:
## glm(formula = nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sa
le_begin +
##     vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +

```

```

##      vh_sale_end + vh_weight + pol_duration + pol_usage + drv_age1
+
##      drv_age_lic2 + vh_age + vh_speed + pol_sit_duration + drv_age_
_lic1 +
##      drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_freq, famil
y = poisson("log"),
##      data = base_etude)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.6943  -0.5235  -0.5110  -0.4970   5.7432
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.332e+00  3.291e-01  -7.087 1.37e-12 ***
## drv_sex1M       -2.586e-02  1.773e-02  -1.459  0.1447
## drv_drv2Yes      1.255e-01  3.848e-01   0.326  0.7442
## vh_din          -5.462e-04  9.314e-04  -0.586  0.5576
## vh_sale_begin   -2.239e-03  4.583e-03  -0.489  0.6252
## vh_value         2.080e-06  2.658e-06   0.783  0.4337
## pol_coverageMedian1 -4.962e-02  3.143e-02  -1.578  0.1145
## pol_coverageMedian2  5.330e-02  2.326e-02   2.291  0.0219 *
## pol_coverageMini   -3.556e-03  3.231e-02  -0.110  0.9124
## pol_paydYes       -5.674e-02  4.587e-02  -1.237  0.2161
## drv_age2         -5.962e-03  1.082e-02  -0.551  0.5816
## vh_fuelGasoline   -2.370e-02  2.547e-02  -0.930  0.3521
## vh_fuelHybrid      4.150e-01  2.514e-01   1.651  0.0988 .
## vh_sale_end      -3.409e-05  4.697e-03  -0.007  0.9942
## vh_weight         1.452e-05  3.150e-05   0.461  0.6448
## pol_duration      2.321e-03  1.114e-03   2.082  0.0373 *
## pol_usageProfessional 2.028e-01  3.033e-01   0.669  0.5038
## pol_usageRetired    1.728e-01  3.021e-01   0.572  0.5674
## pol_usageWorkPrivate 1.781e-01  3.018e-01   0.590  0.5551
## drv_age1         -1.515e-04  9.295e-04  -0.163  0.8705
## drv_age_lic2       5.974e-03  1.081e-02   0.552  0.5807
## vh_age            6.543e-03  4.978e-03   1.314  0.1887
## vh_speed          8.377e-04  8.530e-04   0.982  0.3261
## pol_sit_duration  -3.234e-03  3.993e-03  -0.810  0.4181
## drv_age_lic1      -1.733e-03  7.115e-04  -2.436  0.0149 *
## drv_sex2F         1.566e-02  3.169e-01   0.049  0.9606
## drv_sex2M        -6.974e-03  3.168e-01  -0.022  0.9824
## vh_cyl           -3.447e-05  4.013e-05  -0.859  0.3903
## vh_typeTourism     3.097e-02  3.869e-02   0.800  0.4235
## pol_bonus         7.124e-02  9.388e-02   0.759  0.4480
## pol_pay_freqMonthly -1.968e-02  2.297e-02  -0.857  0.3915
## pol_pay_freqQuarterly 5.097e-02  5.555e-02   0.918  0.3588
## pol_pay_freqYearly -3.434e-02  2.153e-02  -1.595  0.1107
## ---
## 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: 57612  on 99967  degrees of freedom
## AIC: 82138
##
## Number of Fisher Scoring iterations: 6

      stepwise <- step(fpois)

## Start:  AIC=82137.74
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_sale_end + vh_weight + pol_duration + pol_usage + drv_age1
##      +
##      drv_age_lic2 + vh_age + vh_speed + pol_sit_duration + drv_age
##_lic1 +
##      drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##
##      Df Deviance  AIC
## - pol_usage      3    57613 82133
## - drv_sex2       2    57612 82134
## - vh_sale_end    1    57612 82136
## - drv_age1       1    57612 82136
## - drv_drv2       1    57612 82136
## - vh_weight      1    57612 82136
## - pol_pay_freq   3    57616 82136
## - vh_sale_begin  1    57612 82136
## - drv_age2       1    57612 82136
## - drv_age_lic2   1    57612 82136
## - vh_din         1    57612 82136
## - pol_bonus      1    57612 82136
## - vh_value       1    57613 82136
## - vh_type        1    57613 82136
## - pol_sit_duration 1    57613 82136
## - vh_cyl         1    57613 82136
## - vh_speed       1    57613 82137
## - vh_fuel        2    57615 82137
## - pol_payd       1    57613 82137
## - vh_age         1    57614 82137
## <none>           57612 82138
## - drv_sex1       1    57614 82138
## - pol_duration   1    57616 82140
## - pol_coverage   3    57621 82141
## - drv_age_lic1   1    57618 82142
##
## Step:  AIC=82132.78
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
```

```

##      vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_lic2 +
##      vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + drv_sex2 +
##      vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##              Df Deviance   AIC
## - drv_sex2      2      57613 82129
## - vh_sale_end    1      57613 82131
## - drv_age1       1      57613 82131
## - drv_drv2       1      57613 82131
## - vh_weight      1      57613 82131
## - vh_sale_begin  1      57613 82131
## - drv_age2       1      57613 82131
## - drv_age_lic2   1      57613 82131
## - vh_din         1      57613 82131
## - pol_pay_freq   3      57617 82131
## - pol_bonus      1      57614 82131
## - vh_value       1      57614 82131
## - vh_type        1      57614 82131
## - pol_sit_duration 1      57614 82131
## - vh_cyl         1      57614 82132
## - vh_speed       1      57614 82132
## - vh_fuel        2      57616 82132
## - vh_age         1      57615 82133
## - pol_payd       1      57615 82133
## <none>          57613 82133
## - drv_sex1       1      57615 82133
## - pol_duration   1      57617 82135
## - pol_coverage   3      57622 82136
## - drv_age_lic1   1      57619 82137
##
## Step:  AIC=82129.31
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_lic2 +
##      vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl
##      +
##      vh_type + pol_bonus + pol_pay_freq
##
##              Df Deviance   AIC
## - vh_sale_end    1      57613 82127
## - drv_age1       1      57614 82127
## - vh_weight      1      57614 82128
## - vh_sale_begin  1      57614 82128
## - drv_age2       1      57614 82128
## - drv_age_lic2   1      57614 82128
## - vh_din         1      57614 82128

```

```

## - drv_drv2          1      57614 82128
## - pol_pay_freq      3      57618 82128
## - pol_bonus         1      57614 82128
## - vh_value          1      57614 82128
## - vh_type           1      57614 82128
## - pol_sit_duration  1      57614 82128
## - vh_cyl            1      57614 82128
## - vh_speed          1      57614 82128
## - vh_fuel           2      57617 82129
## - vh_age            1      57615 82129
## - pol_payd          1      57615 82129
## <none>              57613 82129
## - drv_sex1          1      57616 82129
## - pol_duration      1      57618 82131
## - pol_coverage      3      57623 82133
## - drv_age_lic1      1      57620 82133
##
## Step: AIC=82127.31
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_weight + pol_duration + drv_age1 + drv_age_lic2 + vh_age +
##      vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl + vh_type
##      +
##      pol_bonus + pol_pay_freq
##
##              Df Deviance   AIC
## - drv_age1      1      57614 82125
## - vh_weight      1      57614 82126
## - vh_sale_begin  1      57614 82126
## - drv_age2       1      57614 82126
## - drv_age_lic2   1      57614 82126
## - vh_din         1      57614 82126
## - drv_drv2       1      57614 82126
## - pol_pay_freq   3      57618 82126
## - pol_bonus      1      57614 82126
## - vh_value       1      57614 82126
## - vh_type        1      57614 82126
## - pol_sit_duration 1      57614 82126
## - vh_cyl         1      57614 82126
## - vh_speed       1      57614 82126
## - vh_fuel        2      57617 82127
## - pol_payd       1      57615 82127
## <none>           57613 82127
## - vh_age         1      57616 82127
## - drv_sex1       1      57616 82127
## - pol_duration   1      57618 82129
## - pol_coverage   3      57623 82131
## - drv_age_lic1   1      57620 82131
##

```

```

## Step: AIC=82125.34
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_weight + pol_duration + drv_age_lic2 + vh_age + vh_speed +
##      pol_sit_duration + drv_age_lic1 + vh_cyl + vh_type + pol_bonu
s +
##      pol_pay_freq
##
##              Df Deviance    AIC
## - vh_weight      1      57614 82124
## - vh_sale_begin   1      57614 82124
## - drv_age2        1      57614 82124
## - drv_age_lic2    1      57614 82124
## - vh_din          1      57614 82124
## - drv_drv2        1      57614 82124
## - pol_pay_freq    3      57618 82124
## - pol_bonus       1      57614 82124
## - vh_value        1      57614 82124
## - vh_type         1      57614 82124
## - pol_sit_duration 1      57614 82124
## - vh_cyl          1      57614 82124
## - vh_speed        1      57614 82124
## - vh_fuel         2      57617 82125
## - pol_payd        1      57615 82125
## <none>            57614 82125
## - vh_age          1      57616 82125
## - drv_sex1        1      57616 82125
## - pol_duration    1      57618 82127
## - pol_coverage    3      57623 82129
## - drv_age_lic1    1      57629 82139
##
## Step: AIC=82123.55
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      pol_duration + drv_age_lic2 + vh_age + vh_speed + pol_sit_dur
ation +
##      drv_age_lic1 + vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##              Df Deviance    AIC
## - vh_din          1      57614 82122
## - vh_sale_begin   1      57614 82122
## - drv_age2        1      57614 82122
## - drv_age_lic2    1      57614 82122
## - drv_drv2        1      57614 82122
## - pol_pay_freq    3      57618 82122
## - pol_bonus       1      57614 82122
## - vh_value        1      57614 82122
## - pol_sit_duration 1      57614 82122
## - vh_cyl          1      57615 82122

```

```

## - vh_type          1      57615 82122
## - vh_speed         1      57615 82123
## - vh_fuel          2      57617 82123
## - pol_payd         1      57615 82123
## <none>              57614 82124
## - vh_age           1      57616 82124
## - drv_sex1         1      57616 82124
## - pol_duration     1      57618 82126
## - pol_coverage     3      57623 82127
## - drv_age_lic1     1      57629 82137
##
## Step: AIC=82121.83
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_sale_begin + vh_value +
##   pol_coverage + pol_payd + drv_age2 + vh_fuel + pol_duration +
##   drv_age_lic2 + vh_age + vh_speed + pol_sit_duration + drv_age_
##   _lic1 +
##   vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##               Df Deviance   AIC
## - vh_sale_begin    1      57614 82120
## - drv_age2         1      57614 82120
## - drv_age_lic2     1      57614 82120
## - drv_drv2         1      57614 82120
## - pol_pay_freq     3      57618 82120
## - vh_value         1      57614 82120
## - pol_bonus        1      57615 82120
## - pol_sit_duration 1      57615 82121
## - vh_speed         1      57615 82121
## - vh_type          1      57615 82121
## - vh_cyl           1      57616 82121
## - pol_payd         1      57616 82122
## <none>              57614 82122
## - drv_sex1         1      57616 82122
## - vh_age           1      57616 82122
## - vh_fuel          2      57619 82123
## - pol_duration     1      57618 82124
## - pol_coverage     3      57623 82125
## - drv_age_lic1     1      57629 82135
##
## Step: AIC=82120.12
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_value + pol_coverage +
##   pol_payd + drv_age2 + vh_fuel + pol_duration + drv_age_lic2 +
##   vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl
##   +
##   vh_type + pol_bonus + pol_pay_freq
##
##               Df Deviance   AIC
## - drv_age2         1      57615 82118
## - drv_age_lic2     1      57615 82118

```



```

## - drv_drv2          1      57615 82118
## - pol_pay_freq      3      57619 82118
## - vh_value          1      57615 82119
## - pol_bonus         1      57615 82119
## - pol_sit_duration  1      57615 82119
## - vh_speed          1      57615 82119
## - vh_type           1      57615 82119
## - vh_cyl            1      57616 82120
## - pol_payd          1      57616 82120
## <none>              57614 82120
## - drv_sex1          1      57616 82120
## - vh_fuel           2      57619 82121
## - pol_duration      1      57618 82122
## - pol_coverage      3      57624 82123
## - vh_age            1      57622 82126
## - drv_age_lic1      1      57630 82134
##
## Step: AIC=82118.43
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_value + pol_coverage +
##               pol_payd + vh_fuel + pol_duration + drv_age_lic2 + vh_age +
##               vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl + vh_type
##               +
##               pol_bonus + pol_pay_freq
##
##               Df Deviance   AIC
## - drv_age_lic2      1      57615 82116
## - drv_drv2          1      57615 82117
## - pol_pay_freq      3      57619 82117
## - vh_value          1      57615 82117
## - pol_bonus         1      57615 82117
## - pol_sit_duration  1      57615 82117
## - vh_speed          1      57615 82117
## - vh_type           1      57616 82118
## - vh_cyl            1      57616 82118
## - pol_payd          1      57616 82118
## <none>              57615 82118
## - drv_sex1          1      57617 82119
## - vh_fuel           2      57619 82119
## - pol_duration      1      57619 82121
## - pol_coverage      3      57624 82122
## - vh_age            1      57622 82124
## - drv_age_lic1      1      57630 82132
##
## Step: AIC=82116.43
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_value + pol_coverage +
##               pol_payd + vh_fuel + pol_duration + vh_age + vh_speed + pol_s
## it_duration +
##               drv_age_lic1 + vh_cyl + vh_type + pol_bonus + pol_pay_freq
##

```

```

##          Df Deviance   AIC
## - pol_pay_freq      3    57619 82115
## - vh_value          1    57615 82115
## - drv_drv2          1    57615 82115
## - pol_bonus          1    57615 82115
## - pol_sit_duration  1    57615 82115
## - vh_speed          1    57615 82115
## - vh_type           1    57616 82116
## - vh_cyl            1    57616 82116
## - pol_payd          1    57616 82116
## <none>              57615 82116
## - drv_sex1          1    57617 82117
## - vh_fuel           2    57619 82117
## - pol_duration      1    57619 82119
## - pol_coverage      3    57624 82120
## - vh_age            1    57622 82122
## - drv_age_lic1      1    57630 82130
##
## Step: AIC=82114.78
## nb_sinistres ~ drv_sex1 + drv_drv2 + vh_value + pol_coverage +
##               pol_payd + vh_fuel + pol_duration + vh_age + vh_speed + pol_s
##               it_duration +
##               drv_age_lic1 + vh_cyl + vh_type + pol_bonus
##
##          Df Deviance   AIC
## - vh_value          1    57619 82113
## - drv_drv2          1    57620 82113
## - pol_bonus          1    57620 82113
## - pol_sit_duration  1    57620 82114
## - vh_speed          1    57620 82114
## - vh_type           1    57620 82114
## - vh_cyl            1    57620 82114
## <none>              57619 82115
## - pol_payd          1    57621 82115
## - drv_sex1          1    57621 82115
## - vh_fuel           2    57624 82116
## - pol_duration      1    57624 82117
## - pol_coverage      3    57628 82118
## - vh_age            1    57627 82121
## - drv_age_lic1      1    57635 82128
##
## Step: AIC=82113.17
## nb_sinistres ~ drv_sex1 + drv_drv2 + pol_coverage + pol_payd +
##               vh_fuel + pol_duration + vh_age + vh_speed + pol_sit_duration
##               +
##               drv_age_lic1 + vh_cyl + vh_type + pol_bonus
##
##          Df Deviance   AIC
## - drv_drv2          1    57620 82112

```

```

## - pol_bonus          1      57620 82112
## - pol_sit_duration   1      57620 82112
## - vh_type            1      57620 82112
## - vh_cyl             1      57621 82112
## - vh_speed           1      57621 82113
## <none>                57619 82113
## - pol_payd           1      57621 82113
## - drv_sex1           1      57622 82113
## - vh_fuel            2      57624 82114
## - pol_duration       1      57624 82116
## - pol_coverage       3      57629 82117
## - vh_age             1      57627 82119
## - drv_age_lic1       1      57635 82127
##
## Step: AIC=82111.74
## nb_sinistres ~ drv_sex1 + pol_coverage + pol_payd + vh_fuel +
##      pol_duration + vh_age + vh_speed + pol_sit_duration + drv_age
##      _lic1 +
##      vh_cyl + vh_type + pol_bonus
##
##              Df Deviance   AIC
## - pol_bonus          1      57621 82110
## - pol_sit_duration   1      57621 82111
## - vh_type            1      57621 82111
## - vh_cyl             1      57621 82111
## - vh_speed           1      57622 82111
## <none>                57620 82112
## - pol_payd           1      57622 82112
## - drv_sex1           1      57622 82112
## - vh_fuel            2      57625 82113
## - pol_duration       1      57624 82114
## - pol_coverage       3      57629 82115
## - vh_age             1      57627 82117
## - drv_age_lic1       1      57635 82125
##
## Step: AIC=82110.44
## nb_sinistres ~ drv_sex1 + pol_coverage + pol_payd + vh_fuel +
##      pol_duration + vh_age + vh_speed + pol_sit_duration + drv_age
##      _lic1 +
##      vh_cyl + vh_type
##
##              Df Deviance   AIC
## - pol_sit_duration   1      57622 82109
## - vh_type            1      57622 82110
## - vh_cyl             1      57622 82110
## - vh_speed           1      57622 82110
## <none>                57621 82110
## - pol_payd           1      57623 82111
## - drv_sex1           1      57623 82111

```

```

## - vh_fuel          2      57626 82111
## - pol_duration     1      57625 82112
## - pol_coverage     3      57630 82114
## - vh_age           1      57628 82116
## - drv_age_lic1     1      57638 82126
##
## Step: AIC=82109.38
## nb_sinistres ~ drv_sex1 + pol_coverage + pol_payd + vh_fuel +
##      pol_duration + vh_age + vh_speed + drv_age_lic1 + vh_cyl +
##      vh_type
##
##              Df Deviance   AIC
## - vh_type      1      57623 82108
## - vh_cyl       1      57623 82109
## - vh_speed     1      57623 82109
## - pol_payd     1      57623 82109
## <none>         57622 82109
## - drv_sex1     1      57624 82110
## - vh_fuel      2      57627 82110
## - pol_duration 1      57625 82111
## - pol_coverage 3      57631 82113
## - vh_age       1      57629 82115
## - drv_age_lic1 1      57639 82125
##
## Step: AIC=82108.46
## nb_sinistres ~ drv_sex1 + pol_coverage + pol_payd + vh_fuel +
##      pol_duration + vh_age + vh_speed + drv_age_lic1 + vh_cyl
##
##              Df Deviance   AIC
## - pol_payd     1      57624 82108
## <none>         57623 82108
## - drv_sex1     1      57625 82109
## - vh_cyl       1      57625 82109
## - vh_fuel      2      57628 82109
## - pol_duration 1      57626 82110
## - vh_speed     1      57627 82111
## - pol_coverage 3      57632 82112
## - vh_age       1      57631 82114
## - drv_age_lic1 1      57640 82124
##
## Step: AIC=82108.19
## nb_sinistres ~ drv_sex1 + pol_coverage + vh_fuel + pol_duration +
##      vh_age + vh_speed + drv_age_lic1 + vh_cyl
##
##              Df Deviance   AIC
## <none>         57624 82108
## - drv_sex1     1      57627 82108
## - vh_cyl       1      57627 82109
## - vh_fuel      2      57629 82109

```

```
## - pol_duration 1 57627 82109
## - vh_speed 1 57629 82111
## - pol_coverage 3 57634 82111
## - vh_age 1 57632 82114
## - drv_age_lic1 1 57642 82124

## Call: glm(formula = nb_sinistres ~ drv_sex1 + pol_coverage + vh_fuel +
## pol_duration + vh_age + vh_speed + drv_age_lic1 + vh_cyl,
## family = poisson("log"), data = base_etude)
##
## Coefficients:
## (Intercept) drv_sex1M pol_coverageMedian1
## -2.130e+00 -2.653e-02 -4.823e-02
## pol_coverageMedian2 pol_coverageMini vh_fuelGasoline
## 5.329e-02 -5.626e-03 -3.275e-02
## vh_fuelHybrid pol_duration vh_age
## 4.322e-01 1.798e-03 4.209e-03
## vh_speed drv_age_lic1 vh_cyl
## 1.061e-03 -1.916e-03 -4.062e-05
##
## Degrees of Freedom: 99999 Total (i.e. Null); 99988 Residual
## Null Deviance: 57670
## Residual Deviance: 57620 AIC: 82110
```

l'AIC P <- AIC F(POIDS2)

library(AER)

dispersiontest(fpois)

```
##
## Overdispersion test
##
## data: fpois
## z = 14.819, p-value < 2.2e-16
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
## 1.109391
```

```
fpois2 <- glm(formula = nb_sinistres ~
pol_coverage + pol_duration2 +
drv_age1.2 + vh_age2 ,
family = poisson("log"),
data = base_etude)
```

summary(fpois2)

```
##
## Call:
## glm(formula = nb_sinistres ~ pol_coverage + pol_duration2 + drv_age1.2 +
##      vh_age2, family = poisson("log"), data = base_etude)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.5532  -0.5244  -0.5131  -0.5067   5.7842
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.0275543   0.0134747 -150.471 < 2e-16 **
##
## pol_coverageMedian1    -0.0481821   0.0312749   -1.541  0.123414
## pol_coverageMedian2     0.0544300   0.0230824    2.358  0.018370 *
## pol_coverageMini      -0.0049535   0.0319516   -0.155  0.876798
## pol_duration2(21, Inf]  0.0479024   0.0229550    2.087  0.036907 *
## drv_age1.2(35,41]       0.0004184   0.0306791    0.014  0.989120
## drv_age1.2(41,46]      -0.0144762   0.0364954   -0.397  0.691619
## drv_age1.2(46,50]      -0.0601519   0.0445638   -1.350  0.177083
## drv_age1.2(50,54]      -0.0252518   0.0483448   -0.522  0.601442
## drv_age1.2(54,58]      -0.2079741   0.0584928   -3.556  0.000377 **
##
## drv_age1.2(58,63]      -0.0223580   0.0556278   -0.402  0.687742
## drv_age1.2(63,68]      -0.0243879   0.0699695   -0.349  0.727427
## drv_age1.2(68,75]      -0.2027856   0.0848667   -2.389  0.016873 *
## drv_age1.2(75, Inf]    -0.1085736   0.1151827   -0.943  0.345875
## vh_age2(17, Inf]       0.0475630   0.0263894    1.802  0.071490 .
## ---
## 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: 57630  on 99985  degrees of freedom
## AIC: 82120
##
## Number of Fisher Scoring iterations: 6

cov.fpois2 <- vcovHC(fpois2, type="HC0")

std.err <- sqrt(diag(cov.fpois2))
r.est <- cbind(Estimate= coef(fpois2),
"Robust SE" = std.err,
"Pr(>|z|)" = 2 * pnorm(
bs(coef(fpois2)/std.err), lower.tail=FALSE),
LL = coef(fpois2) - 1.96
* std.err,
```

UL = coef(fpois2) + 1.96

* std.err)

#AIC_NB<-(regnb) #test d'hypothèse sur quel modèle choisir # test sur la significativité des coefficients sur la prédiction du modèle, le modèle 2 est expliqué seulement par β_0 .

AIC_NB_BIS <-aic(regbn.2)

phi <- 0.9749978

regbn<- glm(formula=nb_sinistres ~ pol_coverage + pol_duration2 + drv_age1.2 + vh_age2,family=negative.binomial(phi), data=base_etude)
summary(regbn)

##

Call:

glm(formula = nb_sinistres ~ pol_coverage + pol_duration2 + drv_age1.2 +

vh_age2, family = negative.binomial(phi), data = base_etude)

##

Deviance Residuals:

Min 1Q Median 3Q Max

-0.5332 -0.5072 -0.4970 -0.4911 4.4260

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.027626 0.014198 -142.814 < 2e-16 ***

pol_coverageMedian1 -0.048070 0.032863 -1.463 0.143542

pol_coverageMedian2 0.054471 0.024381 2.234 0.025476 *

pol_coverageMini -0.004911 0.033647 -0.146 0.883948

pol_duration2(21, Inf] 0.048102 0.024235 1.985 0.047169 *

drv_age1.2(35,41] 0.000341 0.032357 0.011 0.991593

drv_age1.2(41,46] -0.014313 0.038457 -0.372 0.709755

drv_age1.2(46,50] -0.060333 0.046845 -1.288 0.197781

drv_age1.2(50,54] -0.025166 0.050917 -0.494 0.621129

drv_age1.2(54,58] -0.207955 0.061003 -3.409 0.000652 ***

drv_age1.2(58,63] -0.022290 0.058592 -0.380 0.703634

drv_age1.2(63,68] -0.024577 0.073701 -0.333 0.738785

drv_age1.2(68,75] -0.202749 0.088512 -2.291 0.021987 *

drv_age1.2(75, Inf] -0.108551 0.120723 -0.899 0.368562

vh_age2(17, Inf] 0.047708 0.027871 1.712 0.086945 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

(Dispersion parameter for Negative Binomial(0.975) family taken to be 0.9776556)

##

Null deviance: 47324 on 99999 degrees of freedom

Residual deviance: 47290 on 99985 degrees of freedom

AIC: 81692

##

Number of Fisher Scoring iterations: 4

```

base_etude$drv_age1.2_bis <- cut(base_etude$drv_age1, c(-Inf, 35, 41
,46,54,58,63,68,75,Inf))

#Nouvelle regression NB apres Le regroupement d'age
phi2 <- 0.975012
regbn.2 <- glm(formula=nb_sinistres ~ pol_coverage + pol_duration2 +
drv_age1.2_bis +
                                vh_age2,family=negative.binomial(phi2),
data=base_etude)
summary(regbn.2)

##
## Call:
## glm(formula = nb_sinistres ~ pol_coverage + pol_duration2 + drv_a
ge1.2_bis +
##       vh_age2, family = negative.binomial(phi2), data = base_etude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5333  -0.5082  -0.4970  -0.4867   4.4260
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.0276517   0.0141976 -142.817  < 2e-16 *
##
## pol_coverageMedian1    -0.0480475   0.0328622   -1.462  0.143720
## pol_coverageMedian2     0.0545089   0.0243807    2.236  0.025371 *
## pol_coverageMini      -0.0049136   0.0336462   -0.146  0.883891
## pol_duration2(21, Inf]  0.0481349   0.0242348    1.986  0.047015 *
## drv_age1.2_bis(35,41]   0.0003408   0.0323570    0.011  0.991597
## drv_age1.2_bis(41,46]  -0.0143137   0.0384567   -0.372  0.709743
## drv_age1.2_bis(46,54]  -0.0444085   0.0353414   -1.257  0.208917
## drv_age1.2_bis(54,58]  -0.2079551   0.0610020   -3.409  0.000652 *
##
## drv_age1.2_bis(58,63]   -0.0222896   0.0585915   -0.380  0.703631
## drv_age1.2_bis(63,68]  -0.0245779   0.0737004   -0.333  0.738769
## drv_age1.2_bis(68,75]  -0.2027491   0.0885116   -2.291  0.021986 *
## drv_age1.2_bis(75, Inf] -0.1085518   0.1207220   -0.899  0.368555
## vh_age2(17, Inf]       0.0478030   0.0278699    1.715  0.086308 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.975) family taken t
o be 0.9776367)
##
##      Null deviance: 47324  on 99999  degrees of freedom
## Residual deviance: 47290  on 99986  degrees of freedom
## AIC: 81690
##
## Number of Fisher Scoring iterations: 4

```


Regression Binomiale Negative avec Theta estimé package MASS

```
NB1 <- glm.nb(formula=nb_sinistres ~ pol_coverage + pol_duration2 +
drv_age1.2 + vh_age2, data=base_etude)
summary(NB1)
```

```
##
## Call:
## glm.nb(formula = nb_sinistres ~ pol_coverage + pol_duration2 +
##       drv_age1.2 + vh_age2, data = base_etude, init.theta = 1.22495
2652,
##       link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5370  -0.5105  -0.5001  -0.4941   4.5923
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.0276128   0.0141836  -142.955  < 2e-16 **
##
## pol_coverageMedian1    -0.0480905   0.0328471   -1.464   0.14317
## pol_coverageMedian2     0.0544637   0.0243456    2.237   0.02528 *
## pol_coverageMini      -0.0049190   0.0336170   -0.146   0.88367
## pol_duration2(21, Inf]  0.0480660   0.0242020    1.986   0.04703 *
## drv_age1.2(35,41]       0.0003551   0.0323192    0.011   0.99123
## drv_age1.2(41,46]      -0.0143432   0.0384181   -0.373   0.70889
## drv_age1.2(46,50]      -0.0602996   0.0468198   -1.288   0.19778
## drv_age1.2(50,54]      -0.0251819   0.0508707   -0.495   0.62059
## drv_age1.2(54,58]      -0.2079589   0.0610630   -3.406   0.00066 **
##
## drv_age1.2(58,63]      -0.0223020   0.0585376   -0.381   0.70321
## drv_age1.2(63,68]      -0.0245422   0.0736320   -0.333   0.73890
## drv_age1.2(68,75]      -0.2027567   0.0885987   -2.288   0.02211 *
## drv_age1.2(75, Inf]    -0.1085550   0.1207237   -0.899   0.36855
## vh_age2(17, Inf]       0.0476816   0.0278309    1.713   0.08666 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.225) family taken to
be 1)
##
##      Null deviance: 48930  on 99999  degrees of freedom
## Residual deviance: 48895  on 99985  degrees of freedom
## AIC: 81678
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  1.2250
```

```

##          Std. Err.:  0.0740
##
##  2 x log-likelihood:  -81645.5630

NB2 <- glm.nb(formula=nb_sinistres ~ pol_coverage + pol_duration2 +
drv_age1.2_bis + vh_age2, data=base_etude)
summary(NB2)

##
## Call:
## glm.nb(formula = nb_sinistres ~ pol_coverage + pol_duration2 +
##       drv_age1.2_bis + vh_age2, data = base_etude, init.theta = 1.2
24901738,
##       link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5371  -0.5116  -0.5001  -0.4896   4.5922
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.0276389   0.0141837 -142.956 < 2e-16 *
##
## pol_coverageMedian1    -0.0480697   0.0328471   -1.463  0.14335
## pol_coverageMedian2     0.0545006   0.0243456    2.239  0.02518 *
## pol_coverageMini    -0.0049245   0.0336168   -0.146  0.88354
## pol_duration2(21, Inf]  0.0481023   0.0242019    1.988  0.04686 *
## drv_age1.2_bis(35,41]   0.0003548   0.0323193    0.011  0.99124
## drv_age1.2_bis(41,46]  -0.0143438   0.0384182   -0.373  0.70888
## drv_age1.2_bis(46,54]  -0.0443973   0.0353159   -1.257  0.20870
## drv_age1.2_bis(54,58]  -0.2079590   0.0610631   -3.406  0.00066 *
##
## drv_age1.2_bis(58,63]  -0.0223021   0.0585377   -0.381  0.70321
## drv_age1.2_bis(63,68]  -0.0245437   0.0736322   -0.333  0.73889
## drv_age1.2_bis(68,75]  -0.2027569   0.0885988   -2.288  0.02211 *
## drv_age1.2_bis(75, Inf] -0.1085555   0.1207239   -0.899  0.36854
## vh_age2(17, Inf]       0.0477777   0.0278303    1.717  0.08602 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.2249) family taken
to be 1)
##
##      Null deviance: 48930  on 99999  degrees of freedom
## Residual deviance: 48895  on 99986  degrees of freedom
## AIC: 81676
##
## Number of Fisher Scoring iterations: 1
##
##

```

```
##          Theta:  1.2249
##          Std. Err.:  0.0740
##
##  2 x log-likelihood:  -81645.8340
```

#Comparaison des coefficients

```
cbind(regbn$coefficients,fpois2$coefficients) # coefficient de poisson et coefficient binomiaux
```

```
##                                [,1]      [,2]
## (Intercept)                   -2.0276257719 -2.027554336
## pol_coverageMedian1           -0.0480695430 -0.048182094
## pol_coverageMedian2            0.0544709066  0.054429957
## pol_coverageMini              -0.0049113173 -0.004953452
## pol_duration2(21, Inf]         0.0481022072  0.047902424
## drv_age1.2(35,41]              0.0003409501  0.000418354
## drv_age1.2(41,46]             -0.0143132233 -0.014476203
## drv_age1.2(46,50]             -0.0603327945 -0.060151851
## drv_age1.2(50,54]             -0.0251659849 -0.025251828
## drv_age1.2(54,58]             -0.2079551705 -0.207974118
## drv_age1.2(58,63]             -0.0222896270 -0.022358023
## drv_age1.2(63,68]             -0.0245766491 -0.024387873
## drv_age1.2(68,75]             -0.2027490348 -0.202785618
## drv_age1.2(75, Inf]           -0.1085514006 -0.108573642
## vh_age2(17, Inf]              0.0477080606  0.047563020
```

```
#Graphiques des residus pour Poisson et Negative Binomiale par(mfrow = c(1, 2)) plot(regbn$residuals,ylab="ResidusNB")plot(fpois2$residuals,ylab= "Residus Poisson") #Les résidus des modèles poissons et binomiaux
```

Quasi-Poisson

```
quasipois <- glm(formula=nb_sinistres ~ pol_coverage + pol_duration2 + drv_age1.2 +
                vh_age2 ,
                family=quasipoisson("log"), data=base_etude)
```

```
summary(quasipois)
```

```
##
## Call:
## glm(formula = nb_sinistres ~ pol_coverage + pol_duration2 + drv_age1.2 +
##      vh_age2, family = quasipoisson("log"), data = base_etude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5532  -0.5244  -0.5131  -0.5067   5.7842
##
## Coefficients:
```

```
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    -2.0275543   0.0141980  -142.806 < 2e-16 **
*
## pol_coverageMedian1 -0.0481821   0.0329535   -1.462  0.14371
## pol_coverageMedian2  0.0544300   0.0243213    2.238  0.02523 *
## pol_coverageMini    -0.0049535   0.0336665   -0.147  0.88303
## pol_duration2(21, Inf] 0.0479024   0.0241871    1.980  0.04765 *
## drv_age1.2(35,41]     0.0004184   0.0323257    0.013  0.98967
## drv_age1.2(41,46]    -0.0144762   0.0384541   -0.376  0.70658
## drv_age1.2(46,50]    -0.0601519   0.0469557   -1.281  0.20018
## drv_age1.2(50,54]    -0.0252518   0.0509396   -0.496  0.62009
## drv_age1.2(54,58]    -0.2079741   0.0616322   -3.374  0.00074 **
*
## drv_age1.2(58,63]    -0.0223580   0.0586135   -0.381  0.70287
## drv_age1.2(63,68]    -0.0243879   0.0737249   -0.331  0.74080
## drv_age1.2(68,75]    -0.2027856   0.0894217   -2.268  0.02335 *
## drv_age1.2(75, Inf]  -0.1085736   0.1213648   -0.895  0.37100
## vh_age2(17, Inf]     0.0475630   0.0278058    1.711  0.08717 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.11022
5)
##
##      Null deviance: 57668  on 99999  degrees of freedom
## Residual deviance: 57630  on 99985  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

Modele QuasiPoisson avec la variable age modifiée

```
quasipois2 <- glm(nb_sinistres ~ pol_coverage + pol_duration2 + drv_
_age1.2 +
                    vh_age2 ,
                    family=quasipoisson("log"), data=base_etude)

summary(quasipois2)

##
## Call:
## glm(formula = nb_sinistres ~ pol_coverage + pol_duration2 + drv_a
ge1.2 +
##      vh_age2, family = quasipoisson("log"), data = base_etude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5532  -0.5244  -0.5131  -0.5067   5.7842
##
## Coefficients:
```

```
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    -2.0275543   0.0141980  -142.806 < 2e-16 **
*
## pol_coverageMedian1    -0.0481821   0.0329535   -1.462   0.14371
## pol_coverageMedian2     0.0544300   0.0243213    2.238   0.02523 *
## pol_coverageMini      -0.0049535   0.0336665   -0.147   0.88303
## pol_duration2(21, Inf]  0.0479024   0.0241871    1.980   0.04765 *
## drv_age1.2(35,41]       0.0004184   0.0323257    0.013   0.98967
## drv_age1.2(41,46]      -0.0144762   0.0384541   -0.376   0.70658
## drv_age1.2(46,50]      -0.0601519   0.0469557   -1.281   0.20018
## drv_age1.2(50,54]      -0.0252518   0.0509396   -0.496   0.62009
## drv_age1.2(54,58]      -0.2079741   0.0616322   -3.374   0.00074 **
*
## drv_age1.2(58,63]      -0.0223580   0.0586135   -0.381   0.70287
## drv_age1.2(63,68]      -0.0243879   0.0737249   -0.331   0.74080
## drv_age1.2(68,75]      -0.2027856   0.0894217   -2.268   0.02335 *
## drv_age1.2(75, Inf]    -0.1085736   0.1213648   -0.895   0.37100
## vh_age2(17, Inf]       0.0475630   0.0278058    1.711   0.08717 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.11022
5)
##
##      Null deviance: 57668  on 99999  degrees of freedom
## Residual deviance: 57630  on 99985  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```

```
4 Zero- Inflated Poisson AIC_ZIP <- AIC(regZI)
```

```
regZI<- zeroinfl(nb_sinistres ~pol_coverage + drv_age1.2 + vh_age2
+ pol_duration2 | pol_bonus ,data = base_etude,dist = "poisson",link=
"logit")
```

```
summary(regZI)
```

```
##
## Call:
## zeroinfl(formula = nb_sinistres ~ pol_coverage + drv_age1.2 + vh_
age2 +
##      pol_duration2 | pol_bonus, data = base_etude, dist = "poisson
", link = "logit")
##
## Pearson residuals:
##      Min      1Q  Median      3Q      Max
## -0.3846 -0.3527 -0.3447 -0.3417 14.9208
##
## Count model coefficients (poisson with log link):
```

```
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.468463   0.028160 -52.147 < 2e-16 ***
## pol_coverageMedian1 -0.052196   0.032747  -1.594 0.110957
## pol_coverageMedian2  0.052211   0.024259   2.152 0.031376 *
## pol_coverageMini    -0.006449   0.033482  -0.193 0.847268
## drv_age1.2(35,41]    0.004949   0.032271   0.153 0.878122
## drv_age1.2(41,46]   -0.009171   0.038358  -0.239 0.811043
## drv_age1.2(46,50]   -0.054781   0.046725  -1.172 0.241031
## drv_age1.2(50,54]   -0.020234   0.050757  -0.399 0.690147
## drv_age1.2(54,58]   -0.200599   0.060958  -3.291 0.000999 ***
## drv_age1.2(58,63]   -0.016544   0.058395  -0.283 0.776935
## drv_age1.2(63,68]   -0.015952   0.073429  -0.217 0.828017
## drv_age1.2(68,75]   -0.195560   0.088385  -2.213 0.026926 *
## drv_age1.2(75, Inf] -0.099379   0.120396  -0.825 0.409122
## vh_age2(17, Inf]    0.048681   0.027708   1.757 0.078927 .
## pol_duration2(21, Inf] 0.052049   0.024295   2.142 0.032162 *
##
## Zero-inflation model coefficients (binomial with logit link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.08363   0.12941  -0.646  0.5181
## pol_bonus   -0.37491   0.22255  -1.685  0.0921 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 28
## Log-likelihood: -4.083e+04 on 17 Df
```

#Test de Vuong

```
vuong(regZI, fpois2)
```

```
## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
## null that the models are indistinguishable)
## -----
##               Vuong z-statistic                H_A      p-value
## Raw              9.171765 model1 > model2 < 2.22e-16
## AIC-corrected    9.084496 model1 > model2 < 2.22e-16
## BIC-corrected    8.669406 model1 > model2 < 2.22e-16
```

#Modele ZIP avec la variable age modifiée AIC_ZIP2 <- AIC(regZI2)

```
regZI2<- zeroinfl(nb_sinistres ~pol_coverage + drv_age1.2_bis + vh_
age2 + pol_duration2 |pol_bonus ,data = base_etude,dist = "poisson",
link="logit")
```

```
summary(regZI2)
```

```
##
## Call:
## zeroinfl(formula = nb_sinistres ~ pol_coverage + drv_age1.2_bis +
```

```

vh_age2 +
##      pol_duration2 | pol_bonus, data = base_etude, dist = "poisson", link = "logit")
##
## Pearson residuals:
##      Min      1Q  Median      3Q      Max
## -0.3846 -0.3527 -0.3447 -0.3416 14.9207
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.468461    0.028160  -52.146 < 2e-16 ***
## pol_coverageMedian1 -0.052172    0.032747   -1.593 0.111124
## pol_coverageMedian2  0.052243    0.024258    2.154 0.031273 *
## pol_coverageMini   -0.006441    0.033482   -0.192 0.847460
## drv_age1.2_bis(35,41]  0.004947    0.032271    0.153 0.878163
## drv_age1.2_bis(41,46] -0.009171    0.038358   -0.239 0.811031
## drv_age1.2_bis(46,54] -0.039134    0.035298   -1.109 0.267580
## drv_age1.2_bis(54,58] -0.200605    0.060958   -3.291 0.000999 ***
## drv_age1.2_bis(58,63] -0.016549    0.058395   -0.283 0.776873
## drv_age1.2_bis(63,68] -0.015959    0.073429   -0.217 0.827940
## drv_age1.2_bis(68,75] -0.195570    0.088386   -2.213 0.026919 *
## drv_age1.2_bis(75, Inf] -0.099398    0.120397   -0.826 0.409039
## vh_age2(17, Inf]      0.048773    0.027707    1.760 0.078351 .
## pol_duration2(21, Inf] 0.052081    0.024295    2.144 0.032056 *
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.08374    0.12940   -0.647 0.5175
## pol_bonus   -0.37459    0.22253   -1.683 0.0923 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 28
## Log-likelihood: -4.083e+04 on 16 Df

```

3.5 Zero-Inflated Negative Binomiale AIC_ZINB <- AIC(regZINB)

```

regZINB<- zeroinfl(nb_sinistres ~ pol_coverage +drv_age1.2 + vh_age2
+ pol_duration2 | pol_bonus,data = base_etude,dist = "negbin")

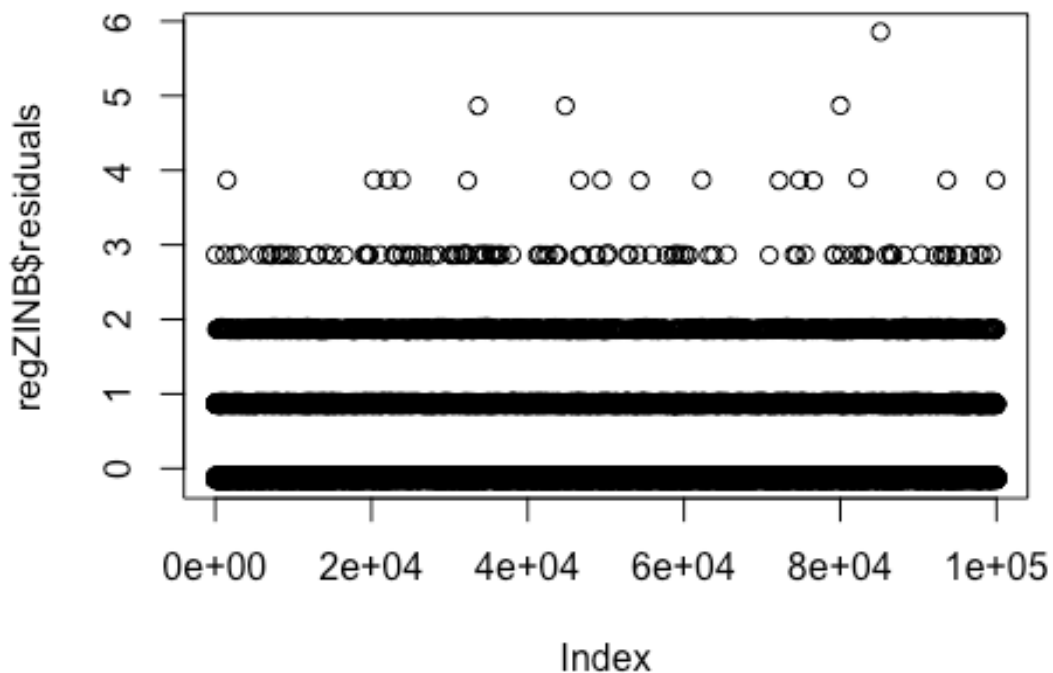
summary(regZINB)

##
## Call:
## zeroinfl(formula = nb_sinistres ~ pol_coverage + drv_age1.2 + vh_age2 +
##      pol_duration2 | pol_bonus, data = base_etude, dist = "negbin")
##

```

```
## Pearson residuals:
##      Min      1Q  Median      3Q      Max
## -0.3762 -0.3514 -0.3431 -0.3402 14.7827
##
## Count model coefficients (negbin with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.983107   0.042419 -46.751 < 2e-16 ***
## pol_coverageMedian1 -0.050689   0.032880  -1.542 0.123165
## pol_coverageMedian2  0.052763   0.024366   2.165 0.030351 *
## pol_coverageMini    -0.006201   0.033621  -0.184 0.853673
## drv_age1.2(35,41]    0.004782   0.032429   0.147 0.882780
## drv_age1.2(41,46]   -0.009268   0.038548  -0.240 0.810009
## drv_age1.2(46,50]   -0.054846   0.046942  -1.168 0.242661
## drv_age1.2(50,54]   -0.019429   0.051005  -0.381 0.703269
## drv_age1.2(54,58]   -0.202096   0.061183  -3.303 0.000956 ***
## drv_age1.2(58,63]   -0.016251   0.058672  -0.277 0.781794
## drv_age1.2(63,68]   -0.018809   0.073735  -0.255 0.798654
## drv_age1.2(68,75]   -0.196961   0.088694  -2.221 0.026373 *
## drv_age1.2(75, Inf] -0.102422   0.120820  -0.848 0.396595
## vh_age2(17, Inf]    0.047858   0.027830   1.720 0.085499 .
## pol_duration2(21, Inf] 0.054022   0.024478   2.207 0.027313 *
## Log(theta)         0.312009   0.116903   2.669 0.007609 **
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.9929     5.7184   0.174   0.862
## pol_bonus     -7.7845    12.4808  -0.624   0.533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 1.3662
## Number of iterations in BFGS optimization: 79
## Log-likelihood: -4.082e+04 on 18 Df
```

```
plot(regZINB$residuals)
```

#Test

de Vuong

```
vuong(regZINB,NB1)
```

```
## Vuong Non-Nested Hypothesis Test-Statistic:
## (test-statistic is asymptotically distributed N(0,1) under the
## null that the models are indistinguishable)
```

```
## -----
##              Vuong z-statistic              H_A      p-value
## Raw              0.8762767 model1 > model2      0.19044
## AIC-corrected    -0.2792729 model2 > model1      0.39002
## BIC-corrected    -5.7756015 model2 > model1 3.8339e-09
```

ZINB avec la variable age modifiée AIC_ZINB2 <- AIC(regZINB2)

```
regZINB2<- zeroinfl(nb_sinistres ~pol_coverage + pol_duration2 + drv
_age1.2_bis + vh_age2 | pol_bonus,data = base_etude,dist = "negbin")
```

```
summary(regZINB2)
```

```
##
## Call:
## zeroinfl(formula = nb_sinistres ~ pol_coverage + pol_duration2 +
##   drv_age1.2_bis +
##     vh_age2 | pol_bonus, data = base_etude, dist = "negbin")
##
## Pearson residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -0.3763 -0.3514 -0.3431 -0.3402 14.7839
##
## Count model coefficients (negbin with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.981658   0.043719 -45.328 < 2e-16 ***
## pol_coverageMedian1 -0.050720   0.032879  -1.543 0.122921
## pol_coverageMedian2  0.052772   0.024365   2.166 0.030320 *
## pol_coverageMini    -0.006242   0.033621  -0.186 0.852722
## pol_duration2(21, Inf]  0.054088   0.024475   2.210 0.027113 *
## drv_age1.2_bis(35,41]  0.004824   0.032428   0.149 0.881736
## drv_age1.2_bis(41,46] -0.009217   0.038547  -0.239 0.811022
## drv_age1.2_bis(46,54] -0.038767   0.035481  -1.093 0.274562
## drv_age1.2_bis(54,58] -0.202038   0.061182  -3.302 0.000959 ***
## drv_age1.2_bis(58,63] -0.016209   0.058671  -0.276 0.782347
## drv_age1.2_bis(63,68] -0.018762   0.073734  -0.254 0.799140
## drv_age1.2_bis(68,75] -0.196875   0.088692  -2.220 0.026435 *
## drv_age1.2_bis(75, Inf] -0.102384   0.120820  -0.847 0.396770
## vh_age2(17, Inf]      0.047985   0.027829   1.724 0.084659 .
## Log(theta)           0.315621   0.119859   2.633 0.008457 **
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.8487     5.4229   0.157   0.876
## pol_bonus    -7.4384    11.9098  -0.625   0.532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 1.3711
## Number of iterations in BFGS optimization: 94
## Log-likelihood: -4.082e+04 on 17 Df

logLik(NB2)

## 'log Lik.' -40822.92 (df=15)

logLik(regbn.2)

## 'log Lik.' -40831.15 (df=14)

logLik(fpois2)

## 'log Lik.' -41044.81 (df=15)

logLik(regZINB2)

## 'log Lik.' -40821.4 (df=17)

logLik(regZI2)

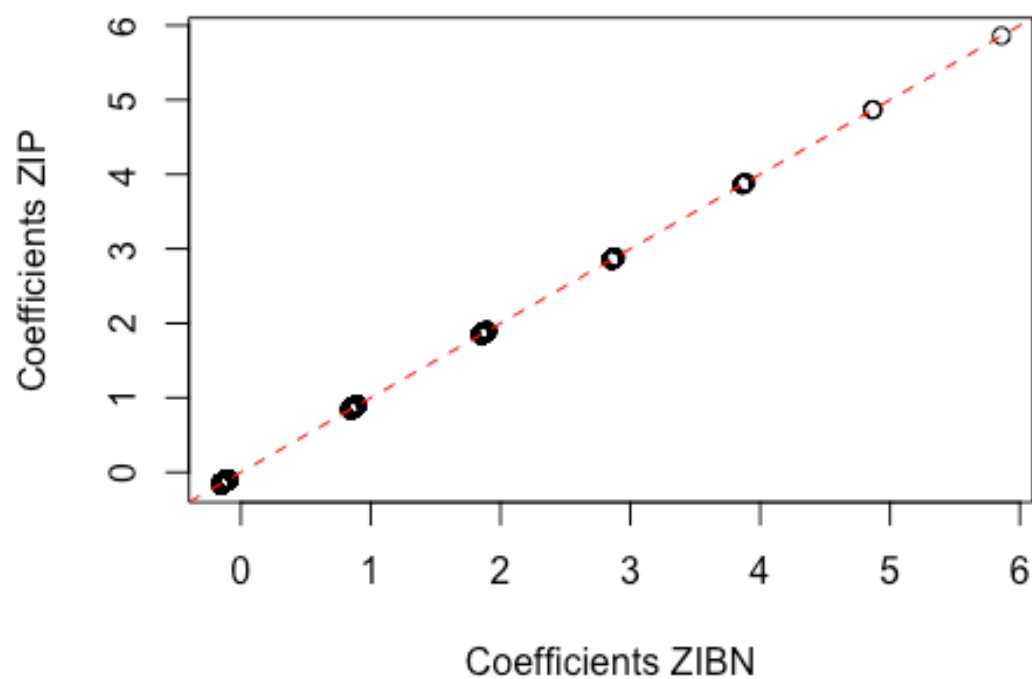
## 'log Lik.' -40834.75 (df=16)

AIC(NB2)
```

```
## [1] 81675.83
AIC(regbn.2)
## [1] 81690.3
AIC(fpois2)
## [1] 82119.63
AIC(regZINB2)
## [1] 81676.8
AIC(regZI2)
## [1] 81701.5
deviance(NB2)
## [1] 48895.48
deviance(regbn.2)
## [1] 47290.03
deviance(fpois2)
## [1] 57629.8
deviance(regZINB2)
## NULL
deviance(regZI2)
## NULL
```

#Comparaison des residus du modele ZIP et ZINB

```
plot(regZINB2$residuals,regZI2$residuals,xlab = "Coefficients ZIBN",
ylab = "Coefficients ZIP")
abline(a=0,b=1,lty=2,col="red")
```

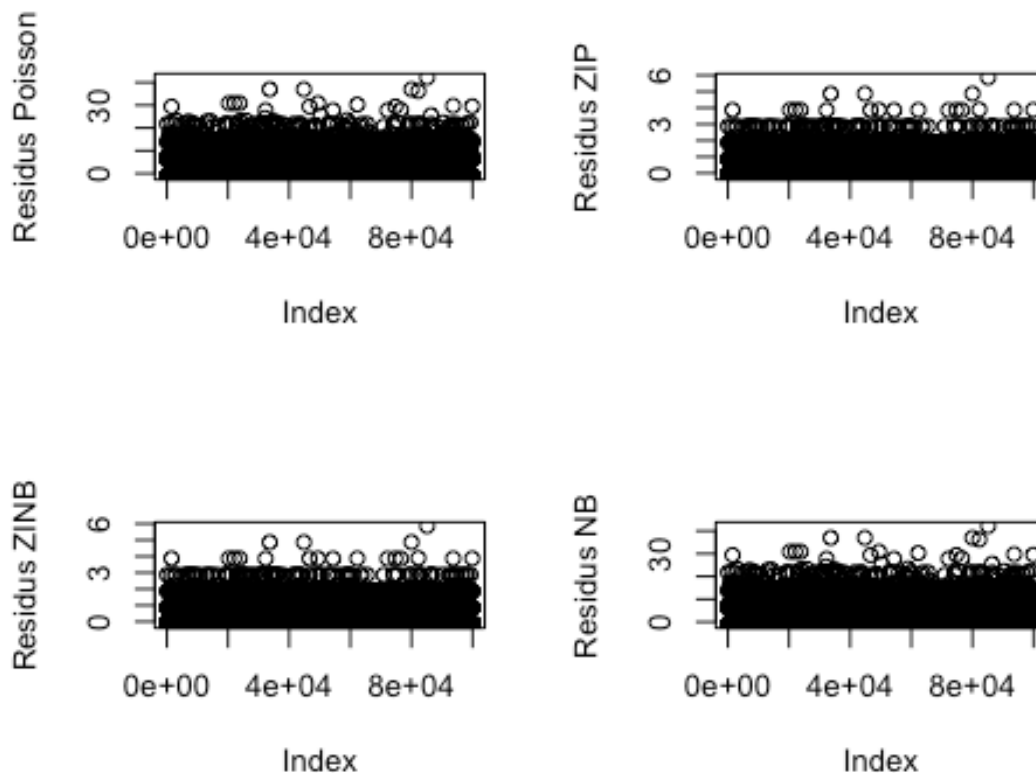


```
cbind(regZINB2$residuals,regZI2$residuals)
```

##	[,1]	[,2]
## 1	-0.1254877	-0.1256347
## 2	-0.1333508	-0.1321282
## 3	-0.1304480	-0.1306487
## 4	-0.1364346	-0.1363768
## 5	-0.1358229	-0.1344414
## 6	-0.1375662	-0.1400480
## 7	-0.1376980	-0.1376333
## 8	-0.1254877	-0.1256347
## 9	-0.1275870	-0.1260118
## 10	-0.1304480	-0.1306487
## 11	-0.1376980	-0.1376333
## 12	-0.1292512	-0.1294560
## 13	-0.1360086	-0.1362982
## 14	-0.1424992	-0.1411611
## 15	-0.1292512	-0.1294560
## 16	2.8689211	2.8687034
## 17	-0.1304480	-0.1306487
## 18	-0.1375169	-0.1376556
## 19	-0.1192817	-0.1192482
## 20	0.8623020	0.8623667

Repartition des residus pour les modeles Poisson, ZIP, ZINB, NB

```
par(mfrow = c(2, 2))
plot(fpois2$residuals, ylab= "Residus Poisson")
plot(regZI2$residuals, ylab = "Residus ZIP")
plot(regZINB2$residuals, ylab = "Residus ZINB")
plot(regbn$residuals, ylab = "Residus NB")
```



3.6

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

```
setwd("/Users/Deneux/Desktop/Jeremy&Pierre")

base_client1 <- read.csv(file="Jeremy&Pierre-PG_2017_YEAR1.csv")

#Categorisation des variables (on utilise la meme categorisation utilisé dans la base d'étude)

base_client1$vh_age2 <- cut(base_client1$vh_age, c(-Inf,17,Inf))
base_client1$pol_duration2 <- cut(base_client1$pol_duration, c(-Inf,
21,Inf))
base_client1$drv_age1.2 <- cut(base_client1$drv_age1, c(-Inf, 35, 41
,46,50,54,58,63,68,75,Inf))
base_client1$drv_age1.2_bis <- cut(base_client1$drv_age1, c(-Inf, 35
, 41,46,54,58,63,68,75,Inf))
```

3.6.2 Prédiction utilisant la Binomiale Negative

```
Prediction_frequence_nb2 <- predict.glm(NB2, base_client1, type="res
ponse")
summary(Prediction_frequence_nb2)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1019  0.1200   0.1285   0.1260  0.1317   0.1531
```

PREDICTION utilisant le ZINB

```
Prediction_frequence_ZINB <- predict(regZINB2, base_client1)
summary(Prediction_frequence_ZINB)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1013  0.1207  0.1284  0.1263  0.1330  0.1563
```

4. MODELISATION DE LA SEVERITE Pour modeliser la severite on utilisera que les individus qui ont eu un montant de sinistre positif

```
base_severite <- subset(base_etude , montant_sinistre_annuel>0 )
base_severite$vh_age <- cut(base_severite$vh_age, c(-Inf,17,Inf))
base_severite$pol_duration<- cut(base_severite$pol_duration, c(-Inf,
21,Inf))
base_severite$drv_age1.2 <- cut(base_severite$drv_age1, c(-Inf, 35,
41,46,50,54,58,63,68,75,Inf))
```

Statistiques descriptives de la base

```
mean(base_severite$montant_sinistre_annuel)
```

```
## [1] 1647.255
```

```
quantile(base_severite$claim_amount,prob=c(.5,.9,.95,.99))
```

```
## 50% 90% 95% 99%
```

```
## NA NA NA NA
```

```
aggregate(base_severite$vh_din ,list(base_severite$drv_age1.2), mean
)
```

```
##      Group.1      x
## 1  (-Inf,35] 91.41280
## 2   (35,41] 91.49486
## 3   (41,46] 89.81293
## 4   (46,50] 93.69118
## 5   (50,54] 92.17370
## 6   (54,58] 90.95539
## 7   (58,63] 96.59866
## 8   (63,68] 93.43094
## 9   (68,75] 92.05556
## 10 (75, Inf] 91.43939
```

```
aggregate(base_severite$vh_speed ,list(base_severite$drv_age1.2), me
an)
```

```
##      Group.1      x
## 1  (-Inf,35] 170.7506
## 2   (35,41] 170.3536
## 3   (41,46] 169.7717
## 4   (46,50] 172.7206
## 5   (50,54] 172.4764
## 6   (54,58] 170.8104
## 7   (58,63] 174.1338
```

```
## 8      (63,68] 171.3702
## 9      (68,75] 171.9921
## 10 (75, Inf] 171.1364

aggregate(base_severite$vh_cyl ,list(base_severite$drv_age1.2), mean
)

##      Group.1      x
## 1  (-Inf,35] 1648.285
## 2   (35,41] 1644.713
## 3   (41,46] 1632.860
## 4   (46,50] 1653.483
## 5   (50,54] 1607.027
## 6   (54,58] 1668.186
## 7   (58,63] 1726.234
## 8   (63,68] 1671.613
## 9   (68,75] 1617.294
## 10 (75, Inf] 1667.788

aggregate(base_severite$vh_value ,list(base_severite$drv_age1.2), me
an)

##      Group.1      x
## 1  (-Inf,35] 18102.23
## 2   (35,41] 17947.92
## 3   (41,46] 17865.32
## 4   (46,50] 18669.68
## 5   (50,54] 18059.16
## 6   (54,58] 17966.11
## 7   (58,63] 19606.13
## 8   (63,68] 18675.97
## 9   (68,75] 18381.87
## 10 (75, Inf] 18847.12

table(base_severite$drv_age1.2,base_severite$vh_fuel)

##
##      Diesel Gasoline Hybrid
##  (-Inf,35]    4519    3659    10
##  (35,41]       581     487     1
##  (41,46]       423     302     2
##  (46,50]       267     209     0
##  (50,54]       214     189     0
##  (54,58]       153     115     1
##  (58,63]       178     120     1
##  (63,68]       103      78     0
##  (68,75]        72      54     0
##  (75, Inf]        35      31     0
```

4.1 Modelisation de la severité par le modele log-normale


```

log_normale <- lm(log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2
+ vh_din + vh_sale_begin + vh_value +
                    pol_coverage + pol_payd + drv_age2 + vh_fuel + vh_sal
e_end +
                    vh_weight + pol_duration + pol_usage + drv_age1 + drv
_age_lic2 +
                    vh_age + vh_speed + pol_sit_duration + drv_age_lic1
+
                    drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_fr
eq, data=base_severite)
stepwise_LN <- step(log_normale)

## Start:  AIC=4223.97
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##      vh_sale_begin + vh_value + pol_coverage + pol_payd + drv_age2
+
##      vh_fuel + vh_sale_end + vh_weight + pol_duration + pol_usage
+
##      drv_age1 + drv_age_lic2 + vh_age + vh_speed + pol_sit_duratio
n +
##      drv_age_lic1 + drv_sex2 + vh_cyl + vh_type + pol_bonus +
##      pol_pay_freq
##
##              Df Sum of Sq  RSS    AIC
## - pol_coverage      3    0.5057 16789 4218.3
## - pol_usage          3    2.9317 16791 4220.0
## - pol_pay_freq       3    3.3567 16792 4220.3
## - drv_sex2           2    1.0746 16790 4220.7
## - drv_age1           1    0.0037 16788 4222.0
## - vh_sale_begin      1    0.0091 16788 4222.0
## - pol_sit_duration   1    0.0273 16788 4222.0
## - vh_type            1    0.0544 16788 4222.0
## - drv_age2           1    0.0994 16788 4222.0
## - drv_age_lic2       1    0.1031 16788 4222.0
## - drv_drv2           1    0.1160 16788 4222.1
## - vh_sale_end        1    0.1279 16788 4222.1
## - vh_speed           1    0.1496 16789 4222.1
## - drv_sex1           1    0.2641 16789 4222.2
## - vh_din             1    0.3763 16789 4222.2
## - pol_payd           1    0.3847 16789 4222.2
## - vh_weight          1    0.5136 16789 4222.3
## - vh_age             1    0.6335 16789 4222.4
## - vh_value           1    0.6592 16789 4222.4
## - vh_cyl             1    0.9290 16789 4222.6
## - pol_duration       1    2.0556 16790 4223.4
## - drv_age_lic1       1    2.3589 16791 4223.6
## <none>                16788 4224.0
## - vh_fuel            2    8.2300 16797 4225.8
## - pol_bonus          1    5.4806 16794 4225.8
##

```

```

## Step: AIC=4218.32
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##     vh_sale_begin + vh_value + pol_payd + drv_age2 + vh_fuel +
##     vh_sale_end + vh_weight + pol_duration + pol_usage + drv_age1
+
##     drv_age_lic2 + vh_age + vh_speed + pol_sit_duration + drv_age
_lic1 +
##     drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##           Df Sum of Sq   RSS   AIC
## - pol_usage      3      3.0118 16792 4214.4
## - pol_pay_freq    3      3.3127 16792 4214.7
## - drv_sex2        2      1.0942 16790 4215.1
## - drv_age1        1      0.0048 16789 4216.3
## - pol_sit_duration 1      0.0074 16789 4216.3
## - vh_sale_begin   1      0.0098 16789 4216.3
## - vh_type         1      0.0559 16789 4216.4
## - drv_age2        1      0.0965 16789 4216.4
## - drv_age_lic2    1      0.1005 16789 4216.4
## - drv_drv2        1      0.1234 16789 4216.4
## - vh_sale_end     1      0.1258 16789 4216.4
## - vh_speed        1      0.1433 16789 4216.4
## - drv_sex1        1      0.2696 16789 4216.5
## - pol_payd        1      0.3429 16789 4216.6
## - vh_din          1      0.3850 16789 4216.6
## - vh_weight       1      0.5228 16789 4216.7
## - vh_age          1      0.6255 16790 4216.8
## - vh_value        1      0.6634 16790 4216.8
## - vh_cyl          1      0.9405 16790 4217.0
## - pol_duration    1      1.9097 16791 4217.7
## - drv_age_lic1    1      2.3513 16791 4218.0
## <none>                        16789 4218.3
## - pol_bonus       1      5.2950 16794 4220.0
## - vh_fuel         2      8.2875 16797 4220.1
##
## Step: AIC=4214.44
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##     vh_sale_begin + vh_value + pol_payd + drv_age2 + vh_fuel +
##     vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_l
ic2 +
##     vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + drv_sex
2 +
##     vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##           Df Sum of Sq   RSS   AIC
## - pol_pay_freq    3      2.8283 16795 4210.4
## - drv_sex2        2      1.0849 16793 4211.2
## - drv_age1        1      0.0013 16792 4212.4
## - vh_sale_begin   1      0.0099 16792 4212.4

```

```

## - vh_type 1 0.0485 16792 4212.5
## - pol_sit_duration 1 0.0579 16792 4212.5
## - drv_age2 1 0.0964 16792 4212.5
## - drv_age_lic2 1 0.1007 16792 4212.5
## - drv_drv2 1 0.1197 16792 4212.5
## - vh_sale_end 1 0.1327 16792 4212.5
## - vh_speed 1 0.1404 16792 4212.5
## - drv_sex1 1 0.2721 16792 4212.6
## - vh_din 1 0.4025 16792 4212.7
## - vh_weight 1 0.5250 16792 4212.8
## - pol_payd 1 0.6361 16793 4212.9
## - vh_age 1 0.6690 16793 4212.9
## - vh_value 1 0.6780 16793 4212.9
## - vh_cyl 1 0.9515 16793 4213.1
## - pol_duration 1 1.4178 16793 4213.4
## - drv_age_lic1 1 2.3984 16794 4214.1
## <none> 16792 4214.4
## - pol_bonus 1 4.7644 16797 4215.8
## - vh_fuel 2 8.3374 16800 4216.3
##
## Step: AIC=4210.43
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
## vh_sale_begin + vh_value + pol_payd + drv_age2 + vh_fuel +
## vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_lic2 +
## vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + drv_sex2 +
## vh_cyl + vh_type + pol_bonus
##
## Df Sum of Sq RSS AIC
## - drv_sex2 2 1.0329 16796 4207.2
## - drv_age1 1 0.0044 16795 4208.4
## - vh_sale_begin 1 0.0086 16795 4208.4
## - pol_sit_duration 1 0.0457 16795 4208.5
## - vh_type 1 0.0493 16795 4208.5
## - drv_age2 1 0.0908 16795 4208.5
## - drv_age_lic2 1 0.0953 16795 4208.5
## - drv_drv2 1 0.1052 16795 4208.5
## - vh_sale_end 1 0.1253 16795 4208.5
## - vh_speed 1 0.1333 16795 4208.5
## - drv_sex1 1 0.2863 16795 4208.6
## - vh_din 1 0.4154 16795 4208.7
## - vh_weight 1 0.4685 16795 4208.8
## - pol_payd 1 0.5167 16795 4208.8
## - vh_value 1 0.6514 16795 4208.9
## - vh_age 1 0.6945 16796 4208.9
## - vh_cyl 1 0.9735 16796 4209.1
## - pol_duration 1 1.2616 16796 4209.3
## - drv_age_lic1 1 2.3092 16797 4210.1

```

```

## <none>                                16795 4210.4
## - pol_bonus                          1    4.9811 16800 4211.9
## - vh_fuel                            2    8.3235 16803 4212.3
##
## Step: AIC=4207.16
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_sale_begin + vh_value + pol_payd + drv_age2 + vh_fuel +
##   vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_lic2 +
##   vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl
## +
##   vh_type + pol_bonus
##
##
##           Df Sum of Sq   RSS   AIC
## - drv_age1      1    0.0062 16796 4205.2
## - vh_sale_begin  1    0.0074 16796 4205.2
## - pol_sit_duration 1    0.0408 16796 4205.2
## - vh_type        1    0.0473 16796 4205.2
## - drv_age2        1    0.0780 16796 4205.2
## - drv_age_lic2    1    0.0810 16796 4205.2
## - drv_drv2        1    0.1139 16796 4205.2
## - vh_sale_end     1    0.1216 16796 4205.2
## - vh_speed        1    0.1351 16796 4205.3
## - drv_sex1        1    0.2842 16796 4205.4
## - vh_din          1    0.4174 16796 4205.4
## - vh_weight        1    0.4537 16796 4205.5
## - pol_payd         1    0.5154 16796 4205.5
## - vh_value         1    0.6611 16796 4205.6
## - vh_age          1    0.6922 16796 4205.6
## - vh_cyl          1    0.9527 16797 4205.8
## - pol_duration     1    1.2660 16797 4206.0
## - drv_age_lic1     1    2.2808 16798 4206.8
## <none>                                16796 4207.2
## - pol_bonus        1    5.0298 16801 4208.7
## - vh_fuel          2    8.2951 16804 4209.0
##
## Step: AIC=4205.16
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_sale_begin + vh_value + pol_payd + drv_age2 + vh_fuel +
##   vh_sale_end + vh_weight + pol_duration + drv_age_lic2 + vh_age +
##   vh_speed + pol_sit_duration + drv_age_lic1 + vh_cyl + vh_type
## +
##   pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - vh_sale_begin      1    0.0073 16796 4203.2
## - pol_sit_duration    1    0.0407 16796 4203.2
## - vh_type             1    0.0469 16796 4203.2

```

```

## - drv_age2          1      0.0777 16796 4203.2
## - drv_age_lic2      1      0.0807 16796 4203.2
## - drv_drv2          1      0.1134 16796 4203.2
## - vh_sale_end       1      0.1209 16796 4203.2
## - vh_speed          1      0.1345 16796 4203.3
## - drv_sex1          1      0.2848 16796 4203.4
## - vh_din            1      0.4183 16796 4203.5
## - vh_weight         1      0.4551 16796 4203.5
## - pol_payd          1      0.5161 16796 4203.5
## - vh_value          1      0.6622 16796 4203.6
## - vh_age            1      0.6924 16796 4203.6
## - vh_cyl            1      0.9533 16797 4203.8
## - pol_duration      1      1.2699 16797 4204.1
## <none>              16796 4205.2
## - pol_bonus         1      5.0272 16801 4206.7
## - vh_fuel           2      8.2960 16804 4207.0
## - drv_age_lic1      1      5.7105 16802 4207.2
##
## Step: AIC=4203.16
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_value + pol_payd + drv_age2 + vh_fuel + vh_sale_end +
##   vh_weight + pol_duration + drv_age_lic2 + vh_age + vh_speed +
##   pol_sit_duration + drv_age_lic1 + vh_cyl + vh_type + pol_bonu
s
##
##              Df Sum of Sq   RSS    AIC
## - pol_sit_duration  1      0.0408 16796 4201.2
## - vh_type           1      0.0516 16796 4201.2
## - drv_age2          1      0.0777 16796 4201.2
## - drv_age_lic2      1      0.0806 16796 4201.2
## - drv_drv2          1      0.1134 16796 4201.2
## - vh_speed          1      0.1297 16796 4201.3
## - drv_sex1          1      0.2845 16796 4201.4
## - vh_sale_end       1      0.3250 16796 4201.4
## - vh_din            1      0.4147 16796 4201.5
## - vh_weight         1      0.4589 16796 4201.5
## - pol_payd          1      0.5178 16796 4201.5
## - vh_value          1      0.6573 16796 4201.6
## - vh_age            1      0.8175 16797 4201.7
## - vh_cyl            1      0.9479 16797 4201.8
## - pol_duration      1      1.2682 16797 4202.1
## <none>              16796 4203.2
## - pol_bonus         1      5.0338 16801 4204.7
## - vh_fuel           2      8.2887 16804 4205.0
## - drv_age_lic1      1      5.7118 16802 4205.2
##
## Step: AIC=4201.19
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_value + pol_payd + drv_age2 + vh_fuel + vh_sale_end +

```

```

##      vh_weight + pol_duration + drv_age_lic2 + vh_age + vh_speed +
##      drv_age_lic1 + vh_cyl + vh_type + pol_bonus
##
##              Df Sum of Sq   RSS   AIC
## - vh_type      1    0.0514 16796 4199.2
## - drv_age2      1    0.0792 16796 4199.2
## - drv_age_lic2  1    0.0821 16796 4199.3
## - drv_drv2      1    0.1155 16796 4199.3
## - vh_speed      1    0.1274 16796 4199.3
## - drv_sex1      1    0.2840 16796 4199.4
## - vh_sale_end   1    0.3254 16796 4199.4
## - vh_din        1    0.4145 16796 4199.5
## - vh_weight     1    0.4587 16796 4199.5
## - pol_payd      1    0.4843 16796 4199.5
## - vh_value      1    0.6550 16796 4199.7
## - vh_age        1    0.8134 16797 4199.8
## - vh_cyl        1    0.9470 16797 4199.9
## - pol_duration  1    1.2280 16797 4200.1
## <none>                16796 4201.2
## - pol_bonus     1    4.9932 16801 4202.7
## - vh_fuel       2    8.2851 16804 4203.0
## - drv_age_lic1  1    5.6929 16802 4203.2
##
## Step:  AIC=4199.23
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##      vh_value + pol_payd + drv_age2 + vh_fuel + vh_sale_end +
##      vh_weight + pol_duration + drv_age_lic2 + vh_age + vh_speed +
##      drv_age_lic1 + vh_cyl + pol_bonus
##
##              Df Sum of Sq   RSS   AIC
## - drv_age2      1    0.0799 16796 4197.3
## - vh_speed      1    0.0814 16796 4197.3
## - drv_age_lic2  1    0.0828 16796 4197.3
## - drv_drv2      1    0.1161 16796 4197.3
## - drv_sex1      1    0.2819 16796 4197.4
## - vh_sale_end   1    0.3469 16796 4197.5
## - vh_weight     1    0.4145 16796 4197.5
## - pol_payd      1    0.4812 16796 4197.6
## - vh_din        1    0.4842 16796 4197.6
## - vh_value      1    0.6812 16797 4197.7
## - vh_age        1    0.8297 16797 4197.8
## - vh_cyl        1    0.9027 16797 4197.9
## - pol_duration  1    1.2280 16797 4198.1
## <none>                16796 4199.2
## - pol_bonus     1    4.9873 16801 4200.7
## - vh_fuel       2    8.4142 16804 4201.1
## - drv_age_lic1  1    5.6812 16802 4201.2
##
## Step:  AIC=4197.29

```

```

## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_value + pol_payd + vh_fuel + vh_sale_end + vh_weight +
##   pol_duration + drv_age_lic2 + vh_age + vh_speed + drv_age_lic
1 +
##   vh_cyl + pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - drv_age_lic2  1     0.0051 16796 4195.3
## - vh_speed      1     0.0836 16796 4195.3
## - drv_drv2      1     0.1827 16796 4195.4
## - drv_sex1      1     0.2852 16796 4195.5
## - vh_sale_end   1     0.3486 16796 4195.5
## - vh_weight     1     0.4126 16796 4195.6
## - pol_payd      1     0.4797 16796 4195.6
## - vh_din        1     0.4804 16796 4195.6
## - vh_value      1     0.6783 16797 4195.8
## - vh_age        1     0.8359 16797 4195.9
## - vh_cyl        1     0.9034 16797 4195.9
## - pol_duration  1     1.2253 16797 4196.1
## <none>                      16796 4197.3
## - pol_bonus     1     4.9872 16801 4198.8
## - vh_fuel       2     8.4041 16804 4199.2
## - drv_age_lic1  1     5.6774 16802 4199.3
##
## Step:  AIC=4195.29
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +
##   vh_value + pol_payd + vh_fuel + vh_sale_end + vh_weight +
##   pol_duration + vh_age + vh_speed + drv_age_lic1 + vh_cyl +
##   pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - vh_speed      1     0.0831 16796 4193.3
## - drv_sex1      1     0.2835 16796 4193.5
## - vh_sale_end   1     0.3498 16796 4193.5
## - vh_weight     1     0.4134 16796 4193.6
## - pol_payd      1     0.4797 16796 4193.6
## - vh_din        1     0.4802 16796 4193.6
## - drv_drv2      1     0.6738 16797 4193.8
## - vh_value      1     0.6783 16797 4193.8
## - vh_age        1     0.8367 16797 4193.9
## - vh_cyl        1     0.9017 16797 4193.9
## - pol_duration  1     1.2237 16797 4194.1
## <none>                      16796 4195.3
## - pol_bonus     1     4.9905 16801 4196.8
## - vh_fuel       2     8.4008 16804 4197.2
## - drv_age_lic1  1     5.6773 16802 4197.3
##
## Step:  AIC=4193.35
## log(montant_sinistre_annuel) ~ drv_sex1 + drv_drv2 + vh_din +

```

```

##      vh_value + pol_payd + vh_fuel + vh_sale_end + vh_weight +
##      pol_duration + vh_age + drv_age_lic1 + vh_cyl + pol_bonus
##
##              Df Sum of Sq   RSS   AIC
## - drv_sex1      1    0.2795 16796 4191.5
## - vh_sale_end    1    0.3459 16796 4191.6
## - vh_weight      1    0.4783 16796 4191.7
## - pol_payd       1    0.4820 16797 4191.7
## - drv_drv2       1    0.6692 16797 4191.8
## - vh_age         1    0.7698 16797 4191.9
## - vh_value       1    0.8145 16797 4191.9
## - vh_cyl         1    1.1817 16797 4192.2
## - pol_duration   1    1.2181 16797 4192.2
## - vh_din         1    1.4463 16798 4192.4
## <none>                                16796 4193.3
## - pol_bonus      1    4.9859 16801 4194.9
## - drv_age_lic1    1    5.6613 16802 4195.3
## - vh_fuel        2    8.5404 16805 4195.3
##
## Step:  AIC=4191.54
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_value +
##      pol_payd + vh_fuel + vh_sale_end + vh_weight + pol_duration +
##      vh_age + drv_age_lic1 + vh_cyl + pol_bonus
##
##              Df Sum of Sq   RSS   AIC
## - vh_sale_end    1    0.3424 16797 4189.8
## - vh_weight      1    0.4739 16797 4189.9
## - pol_payd       1    0.4837 16797 4189.9
## - drv_drv2       1    0.6828 16797 4190.0
## - vh_age         1    0.7717 16797 4190.1
## - vh_value       1    0.8002 16797 4190.1
## - vh_cyl         1    1.2032 16798 4190.4
## - pol_duration   1    1.2307 16798 4190.4
## - vh_din         1    1.4448 16798 4190.6
## <none>                                16796 4191.5
## - pol_bonus      1    4.9111 16801 4193.0
## - drv_age_lic1    1    5.5951 16802 4193.5
## - vh_fuel        2    8.5822 16805 4193.6
##
## Step:  AIC=4189.78
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_value +
##      pol_payd + vh_fuel + vh_weight + pol_duration + vh_age +
##      drv_age_lic1 + vh_cyl + pol_bonus
##
##              Df Sum of Sq   RSS   AIC
## - vh_age         1    0.4315 16797 4188.1
## - vh_weight      1    0.4629 16797 4188.1
## - pol_payd       1    0.4848 16797 4188.1
## - drv_drv2       1    0.6746 16797 4188.3

```



```

## - vh_value      1      0.7019 16797 4188.3
## - pol_duration  1      1.2170 16798 4188.6
## - vh_din        1      1.8915 16799 4189.1
## - vh_cyl        1      2.1594 16799 4189.3
## <none>                                16797 4189.8
## - pol_bonus     1      4.8944 16802 4191.2
## - drv_age_lic1  1      5.6078 16802 4191.7
## - vh_fuel       2      9.7184 16806 4192.6
##
## Step: AIC=4188.09
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_value +
##   pol_payd + vh_fuel + vh_weight + pol_duration + drv_age_lic1
##   +
##   vh_cyl + pol_bonus
##
##              Df Sum of Sq   RSS    AIC
## - vh_weight    1    0.4582 16798 4186.4
## - pol_payd     1    0.4824 16798 4186.4
## - drv_drv2     1    0.6629 16798 4186.6
## - vh_value     1    0.7266 16798 4186.6
## - pol_duration  1    1.2384 16798 4187.0
## - vh_din       1    1.6025 16799 4187.2
## - vh_cyl       1    1.7945 16799 4187.3
## <none>                                16797 4188.1
## - pol_bonus    1    4.8293 16802 4189.5
## - drv_age_lic1  1    5.5967 16803 4190.0
## - vh_fuel      2    9.3624 16806 4190.7
##
## Step: AIC=4186.41
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_value +
##   pol_payd + vh_fuel + pol_duration + drv_age_lic1 + vh_cyl +
##   pol_bonus
##
##              Df Sum of Sq   RSS    AIC
## - pol_payd     1    0.4607 16798 4184.7
## - vh_value     1    0.6340 16798 4184.9
## - drv_drv2     1    0.6771 16798 4184.9
## - pol_duration  1    1.2442 16799 4185.3
## - vh_cyl       1    2.1488 16800 4185.9
## - vh_din       1    2.2181 16800 4186.0
## <none>                                16798 4186.4
## - pol_bonus    1    4.8415 16802 4187.8
## - drv_age_lic1  1    5.6033 16803 4188.3
## - vh_fuel      2   10.2290 16808 4189.6
##
## Step: AIC=4184.73
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_value +
##   vh_fuel + pol_duration + drv_age_lic1 + vh_cyl + pol_bonus
##

```

```

##           Df Sum of Sq   RSS   AIC
## - vh_value      1    0.6397 16799 4183.2
## - drv_drv2      1    0.6780 16799 4183.2
## - pol_duration  1    1.1985 16799 4183.6
## - vh_cyl        1    2.1287 16800 4184.2
## - vh_din        1    2.2197 16800 4184.3
## <none>                      16798 4184.7
## - pol_bonus     1    4.7472 16803 4186.1
## - drv_age_lic1  1    5.6222 16804 4186.7
## - vh_fuel       2   10.2889 16808 4188.0
##
## Step:  AIC=4183.18
## log(montant_sinistre_annuel) ~ drv_drv2 + vh_din + vh_fuel +
##   pol_duration + drv_age_lic1 + vh_cyl + pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - drv_drv2      1    0.6741 16799 4181.7
## - pol_duration  1    1.2008 16800 4182.0
## - vh_din        1    1.9430 16801 4182.5
## - vh_cyl        1    2.7441 16801 4183.1
## <none>                      16799 4183.2
## - pol_bonus     1    4.7382 16803 4184.5
## - drv_age_lic1  1    5.6109 16804 4185.1
## - vh_fuel       2    9.7106 16808 4186.0
##
## Step:  AIC=4181.66
## log(montant_sinistre_annuel) ~ vh_din + vh_fuel + pol_duration +
##   drv_age_lic1 + vh_cyl + pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - pol_duration  1    1.2247 16801 4180.5
## - vh_din        1    1.9383 16801 4181.0
## - vh_cyl        1    2.7308 16802 4181.6
## <none>                      16799 4181.7
## - pol_bonus     1    4.7130 16804 4183.0
## - drv_age_lic1  1    5.6589 16805 4183.6
## - vh_fuel       2    9.7300 16809 4184.5
##
## Step:  AIC=4180.52
## log(montant_sinistre_annuel) ~ vh_din + vh_fuel + drv_age_lic1 +
##   vh_cyl + pol_bonus
##
##           Df Sum of Sq   RSS   AIC
## - vh_din        1    1.9200 16802 4179.9
## - vh_cyl        1    2.7321 16803 4180.4
## <none>                      16801 4180.5
## - pol_bonus     1    5.4598 16806 4182.4
## - drv_age_lic1  1    5.8383 16806 4182.6
## - vh_fuel       2    9.7380 16810 4183.4

```

```
##
## Step: AIC=4179.87
## log(montant_sinistre_annuel) ~ vh_fuel + drv_age_lic1 + vh_cyl +
##     pol_bonus
##
##           Df Sum of Sq  RSS    AIC
## - vh_cyl      1    0.8702 16803 4178.5
## <none>                        16802 4179.9
## - pol_bonus    1    5.4763 16808 4181.7
## - drv_age_lic1 1    5.7141 16808 4181.9
## - vh_fuel      2    8.6322 16811 4181.9
##
## Step: AIC=4178.48
## log(montant_sinistre_annuel) ~ vh_fuel + drv_age_lic1 + pol_bonus
##
##           Df Sum of Sq  RSS    AIC
## <none>                        16803 4178.5
## - vh_fuel      2    7.7805 16811 4179.9
## - pol_bonus    1    5.4511 16809 4180.3
## - drv_age_lic1 1    5.6948 16809 4180.5

log_normale2 <- lm(log(montant_sinistre_annuel) ~ vh_fuel+ drv_age1
, data=base_severite)
summary(log_normale2)

##
## Call:
## lm(formula = log(montant_sinistre_annuel) ~ vh_fuel + drv_age1,
##     data = base_severite)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.8026 -0.7583  0.0831  0.8526  5.6063
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.6953338   0.0275311  243.191   <2e-16 ***
## vh_fuelGasoline 0.0266606   0.0221229    1.205   0.2282
## vh_fuelHybrid  -0.6070819   0.3085480   -1.968   0.0491 *
## drv_age1       0.0014664   0.0007659    1.915   0.0556 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.194 on 11800 degrees of freedom
## Multiple R-squared:  0.00077, Adjusted R-squared:  0.000516
## F-statistic: 3.031 on 3 and 11800 DF, p-value: 0.02811
```

#Fonction pour calculer le critere AIC

```
loglikg<-logLik(log_normale2)
akaike<-function(npar,loglik,k){-2*loglik+k*npar}
akaike(9,loglikg[1],2)

## [1] 37690.96

#Critère d'adéquation

AIC(log_normale2)

## [1] 37682.96

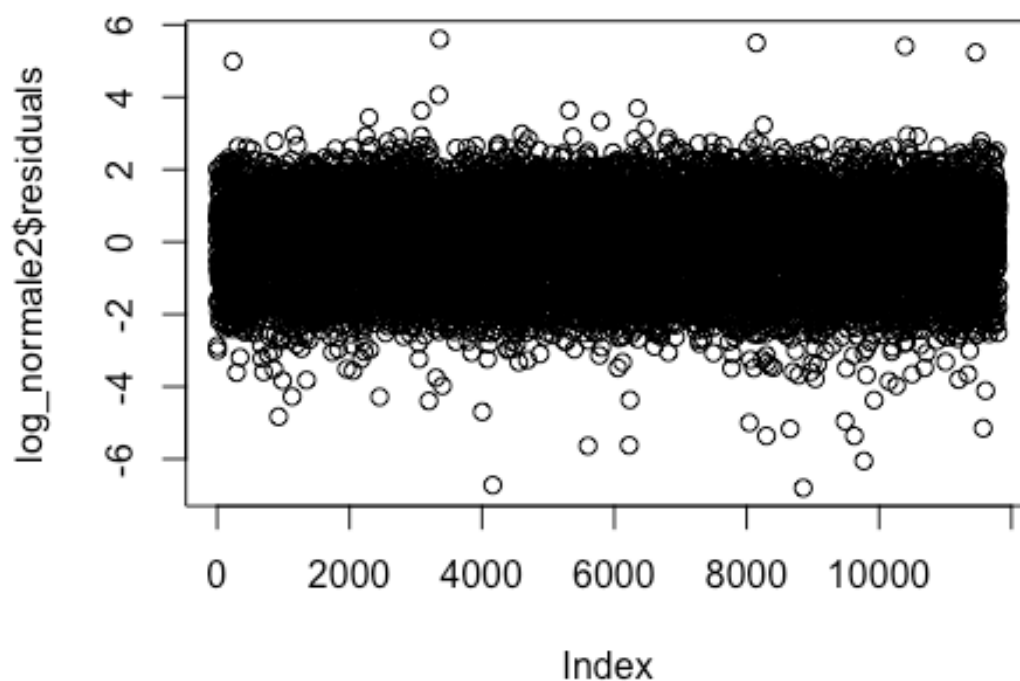
deviance(log_normale2)

## [1] 16812.18

logLik(log_normale2)

## 'log Lik.' -18836.48 (df=5)

plot(log_normale2$residuals)
```



4.2

Modelisation de la severite par le modele Gamma

```
gamma1 <- glm(montant_sinistre_annuel ~ drv_sex1 + drv_drv2 + vh_din
+ vh_sale_begin + vh_value +
pol_coverage + pol_payd + drv_age2 + vh_fuel + vh_sal
e_end +
```

```

      vh_weight + pol_duration + pol_usage + drv_age1 + drv
_age_lic2 +
      vh_age + vh_speed + pol_sit_duration + drv_age_lic1
+
      drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_fr
eq, family=Gamma(link="log"), data=base_severite)
summary(gamma1)

##
## Call:
## glm(formula = montant_sinistre_annuel ~ drv_sex1 + drv_drv2 +
##      vh_din + vh_sale_begin + vh_value + pol_coverage + pol_payd +
##      drv_age2 + vh_fuel + vh_sale_end + vh_weight + pol_duration +
##      pol_usage + drv_age1 + drv_age_lic2 + vh_age + vh_speed +
##      pol_sit_duration + drv_age_lic1 + drv_sex2 + vh_cyl + vh_type
##      +
##      pol_bonus + pol_pay_freq, family = Gamma(link = "log"), data
= base_severite)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -3.5616  -1.1499  -0.5112   0.2226  14.7008
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.220e+00  7.513e-01   9.610  <2e-16 ***
## drv_sex1M         2.195e-02  4.105e-02   0.535  0.5929
## drv_drv2Yes       -4.038e-02  9.243e-01  -0.044  0.9652
## vh_din            -3.176e-03  2.119e-03  -1.499  0.1339
## vh_sale_begin     -7.903e-05  9.088e-03  -0.009  0.9931
## vh_value           1.003e-05  6.159e-06   1.628  0.1035
## pol_coverageMedian1 -2.368e-02  7.215e-02  -0.328  0.7428
## pol_coverageMedian2 -7.988e-03  5.394e-02  -0.148  0.8823
## pol_coverageMini     1.185e-01  7.482e-02   1.584  0.1133
## pol_paydYes         1.896e-01  1.051e-01   1.804  0.0713 .
## drv_age2           3.884e-03  2.497e-02   0.156  0.8764
## vh_fuelGasoline     9.816e-02  5.883e-02   1.668  0.0953 .
## vh_fuelHybrid      -8.569e-01  5.670e-01  -1.511  0.1307
## vh_sale_end         1.146e-02  9.965e-03   1.150  0.2500
## vh_weight          -7.837e-06  7.258e-05  -0.108  0.9140
## pol_duration(21, Inf] -5.203e-02  5.524e-02  -0.942  0.3462
## pol_usageProfessional 1.849e-01  6.954e-01   0.266  0.7903
## pol_usageRetired     7.903e-02  6.927e-01   0.114  0.9092
## pol_usageWorkPrivate  9.124e-02  6.919e-01   0.132  0.8951
## drv_age1           -2.245e-03  2.163e-03  -1.038  0.2993
## drv_age_lic2        -3.701e-03  2.498e-02  -0.148  0.8822
## vh_age(17, Inf]     -1.866e-01  9.095e-02  -2.051  0.0403 *
## vh_speed            1.814e-03  1.966e-03   0.923  0.3562
## pol_sit_duration    -1.392e-03  9.072e-03  -0.153  0.8781
## drv_age_lic1         2.576e-03  1.658e-03   1.553  0.1203

```

```

## drv_sex2F          -8.771e-02  7.751e-01  -0.113   0.9099
## drv_sex2M          -1.488e-01  7.734e-01  -0.192   0.8474
## vh_cyl             -2.555e-05  9.299e-05  -0.275   0.7835
## vh_typeTourism     -5.514e-02  8.825e-02  -0.625   0.5321
## pol_bonus          -2.140e-01  2.140e-01  -1.000   0.3173
## pol_pay_freqMonthly -3.497e-02  5.284e-02  -0.662   0.5081
## pol_pay_freqQuarterly 5.279e-02  1.310e-01   0.403   0.6870
## pol_pay_freqYearly  3.224e-02  4.968e-02   0.649   0.5164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 4.766805)
##
##      Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15296  on 11771  degrees of freedom
## AIC: 198440
##
## Number of Fisher Scoring iterations: 7

stepwise_gamma <- step(gamma1)

## Start:  AIC=198439.8
## montant_sinistre_annuel ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_
begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_sale_end + vh_weight + pol_duration + pol_usage + drv_age1
+
##      drv_age_lic2 + vh_age + vh_speed + pol_sit_duration + drv_age
_lic1 +
##      drv_sex2 + vh_cyl + vh_type + pol_bonus + pol_pay_freq
##
##              Df Deviance    AIC
## - pol_usage      3    15304 198435
## - pol_pay_freq    3    15305 198436
## - drv_sex2        2    15300 198437
## - pol_coverage    3    15310 198437
## - vh_sale_begin   1    15296 198438
## - drv_drv2        1    15296 198438
## - vh_weight        1    15296 198438
## - drv_age_lic2     1    15296 198438
## - pol_sit_duration 1    15296 198438
## - drv_age2         1    15296 198438
## - vh_cyl           1    15296 198438
## - drv_sex1         1    15297 198438
## - vh_type          1    15298 198438
## - vh_speed         1    15300 198439
## - pol_duration     1    15300 198439
## - pol_bonus        1    15301 198439
## - drv_age1         1    15301 198439
## - vh_sale_end      1    15302 198439

```

```

## <none>                15296 198440
## - vh_din               1    15306 198440
## - vh_fuel              2    15317 198440
## - drv_age_lic1         1    15308 198440
## - vh_value             1    15308 198440
## - pol_payd             1    15311 198441
## - vh_age               1    15316 198442
##
## Step:  AIC=198441.1
## montant_sinistre_annuel ~ drv_sex1 + drv_drv2 + vh_din + vh_sale_
begin +
##      vh_value + pol_coverage + pol_payd + drv_age2 + vh_fuel +
##      vh_sale_end + vh_weight + pol_duration + drv_age1 + drv_age_l
ic2 +
##      vh_age + vh_speed + pol_sit_duration + drv_age_lic1 + drv_sex
2 +
##      vh_cyl + vh_type + pol_bonus + pol_pay_freq

summary(stepwise_gamma)

##
## Call:
## glm(formula = montant_sinistre_annuel ~ drv_sex1 + drv_drv2 +
##      vh_din + vh_sale_begin + vh_value + pol_coverage + pol_payd +
##      drv_age2 + vh_fuel + vh_sale_end + vh_weight + pol_duration +
##      drv_age1 + drv_age_lic2 + vh_age + vh_speed + pol_sit_duratio
n +
##      drv_age_lic1 + drv_sex2 + vh_cyl + vh_type + pol_bonus +
##      pol_pay_freq, family = Gamma(link = "log"), data = base_sever
ite)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -3.5631  -1.1528  -0.5114   0.2222  14.6002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.321e+00  3.053e-01  23.982  <2e-16 ***
## drv_sex1M       2.268e-02  4.141e-02   0.548   0.5839
## drv_drv2Yes     -8.929e-02  9.327e-01  -0.096   0.9237
## vh_din          -3.204e-03  2.137e-03  -1.499   0.1339
## vh_sale_begin    1.777e-04  9.170e-03   0.019   0.9845
## vh_value         1.005e-05  6.213e-06   1.617   0.1059
## pol_coverageMedian1 -2.466e-02  7.276e-02  -0.339   0.7347
## pol_coverageMedian2 -1.059e-02  5.436e-02  -0.195   0.8455
## pol_coverageMini   1.169e-01  7.538e-02   1.551   0.1209
## pol_paydYes       1.789e-01  1.049e-01   1.705   0.0883 .
## drv_age2         4.047e-03  2.520e-02   0.161   0.8724
## vh_fuelGasoline    1.009e-01  5.936e-02   1.699   0.0893 .
## vh_fuelHybrid     -8.601e-01  5.721e-01  -1.503   0.1328

```

```

##   vh_sale_end      1.113e-02  1.006e-02   1.107   0.2682
##   vh_weight      -8.281e-06  7.323e-05  -0.113   0.9100
##   pol_duration(21, Inf] -5.763e-02  5.505e-02  -1.047   0.2952
##   drv_age1      -2.377e-03  2.182e-03  -1.089   0.2760
##   drv_age_lic2    -3.822e-03  2.521e-02  -0.152   0.8795
##   vh_age(17, Inf]    -1.884e-01  9.177e-02  -2.053   0.0401 *
##   vh_speed       1.790e-03  1.984e-03   0.902   0.3669
##   pol_sit_duration -2.060e-03  9.105e-03  -0.226   0.8210
##   drv_age_lic1     2.709e-03  1.673e-03   1.620   0.1054
##   drv_sex2F      -4.321e-02  7.821e-01  -0.055   0.9559
##   drv_sex2M      -1.047e-01  7.805e-01  -0.134   0.8933
##   vh_cyl        -2.392e-05  9.382e-05  -0.255   0.7987
##   vh_typeTourism  -5.206e-02  8.905e-02  -0.585   0.5588
##   pol_bonus      -2.104e-01  2.149e-01  -0.979   0.3275
##   pol_pay_freqMonthly -3.845e-02  5.309e-02  -0.724   0.4689
##   pol_pay_freqQuarterly 4.878e-02  1.322e-01   0.369   0.7121
##   pol_pay_freqYearly  2.708e-02  4.995e-02   0.542   0.5877
##   ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 4.853899)
##
##      Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15304  on 11774  degrees of freedom
## AIC: 198441
##
## Number of Fisher Scoring iterations: 7

gamma3 <- glm(montant_sinistre_annuel ~ vh_fuel +drv_age1 , family
= Gamma(link = "log"), data=base_severite )
summary(gamma3)

##
## Call:
## glm(formula = montant_sinistre_annuel ~ vh_fuel + drv_age1, famil
y = Gamma(link = "log"),
##      data = base_severite)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5979  -1.1460  -0.5213   0.2080  16.0774
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.3490502   0.0588792 124.816  <2e-16 ***
## vh_fuelGasoline  0.0966887   0.0473129   2.044   0.041 *
## vh_fuelHybrid  -0.8276307   0.6598731  -1.254   0.210
## drv_age1       0.0004753   0.0016380   0.290   0.772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



```
##
## (Dispersion parameter for Gamma family taken to be 6.516544)
##
##      Null deviance: 15490  on 11803  degrees of freedom
## Residual deviance: 15453  on 11800  degrees of freedom
## AIC: 198526
##
## Number of Fisher Scoring iterations: 6
```

#Critère d'adéquations

```
AIC(gamma3)
```

```
## [1] 198526
```

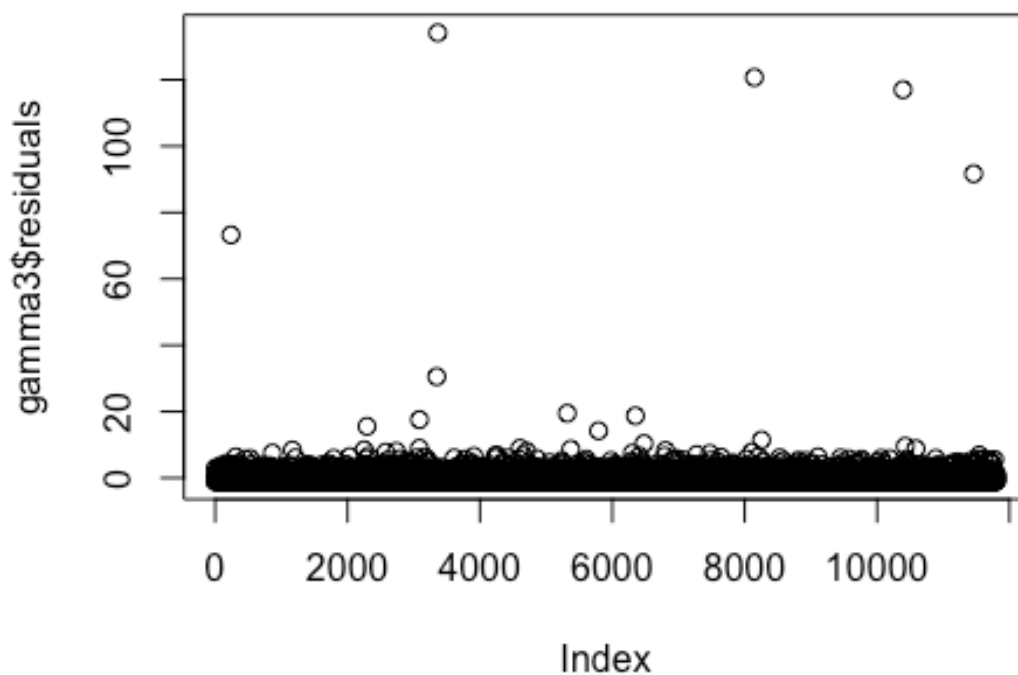
```
deviance(gamma3)
```

```
## [1] 15453.42
```

```
logLik(gamma3)
```

```
## 'log Lik.' -99258.01 (df=5)
```

```
plot(gamma3$residuals)
```



4.4

Prediction de la severite #Prediction Gamma

```
base_client1$Prediction_severite <- predict.glm(gamma3, base_client1
, type="response")
summary(base_client1$Prediction_severite)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  689.6  1593.6  1610.4  1668.8  1757.9  1799.3
```

Etude des couts et valeurs des vehicules suivant la variable vh_fuel

```
aggregate(base_client1$Prediction_severite ,list(base_client1$vh_fue
l), mean)
```

```
##      Group.1      x
## 1   Diesel 1595.0342
## 2 Gasoline 1760.4344
## 3   Hybrid  698.5439
```

```
aggregate(base_client1$vh_value ,list(base_client1$vh_fuel), mean)
```

```
##      Group.1      x
## 1   Diesel 21184.41
## 2 Gasoline 14232.64
## 3   Hybrid 28581.24
```

#Prediction Log-Normal

```
prediction_sev <- predict(log_normale2, base_client1,type="response")
sigma <- summary(log_normale2)$sigma
Pln <- exp(prediction_sev+sigma^2/2)
```

```
mean(Pln)
```

```
## [1] 1810.551
```

```
min(Pln)
```

```
## [1] 940.1988
```

```
max(Pln)
```

```
## [1] 1972.157
```

Prime pure

```
base_client1$tarif <- Prediction_frequence_nb2*Pln
summary(base_client1$tarif)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   101.8   222.4   229.9   227.9   238.4   279.3
```

```
base_tarif <- base_client1[,c("id_policy","tarif")]
write.csv(base_tarif, file = "base_tarif.csv")
```

ETUDE DES SINISTRES GRAVES

```

M <- base_severite[order(-base_severite$montant_sinistre_annuel),c("
montant_sinistre_annuel","nb_sinistres","vh_din","vh_age","drv_age1"
,"vh_type")]
M$SUM <- cumsum(M$montant_sinistre_annuel)/sum(M$montant_sinistre_an
nuel)*100
head(M)

##      montant_sinistre_annuel nb_sinistres vh_din   vh_age drv_a
ge1 vh_type
## 28598          234104          1    101 (-Inf,17]
24 Tourism
## 68972          211112          1     60 (-Inf,17]
27 Tourism
## 87947          185065          1     75 (-Inf,17]
18 Tourism
## 96976          160086          2     51 (-Inf,17]
18 Tourism
## 1850           129665          1     45 (-Inf,17]
41 Tourism
## 28430           50185          1    101 (-Inf,17]
49 Tourism
##              SUM
## 28598 1.203979
## 68972 2.289712
## 87947 3.241487
## 96976 4.064797
## 1850  4.731654
## 28430 4.989752

```