

Statistique des assurances - Projet

Isabelle Ajtay (41010932) Smail Chabane (38012939) Yuxuan Zhang (38019811)

20 février 2023

Table des matières

1. Description des données	1
Analyses univariées	2
Analyses bivariées	3
2. Modélisation des sinistres et des primes pures	10
2.1 Problème d'endogénéité dans les variables	10
Tester les problèmes d'endogénéité de certaines variables	10
2.2 Modélisation de Sinistre0	16
2.3 Modélisation de Sinistre 1 ou 2 ou 3 (au moins un)	16
2.4 Modèle pour le prix de Police 1 ou 2 ou 3 (au moins un)	17
2.5 Modèle retenu au final	17
IV regressions	17
3. Modélisation du nombre de sinistres et tarification des nouveaux arrivants	17
3.1 Modèle pour le nombre de sinistres, NSin	17
3.2 Méthode de tarification pour les nouveaux arrivants	17
4. Estimation des durées	17
4.1 Estimateur de Kaplan-Meier	17
4.2 Modèle de Cox	17

Plan d'analyse

L'approche que nous allons suivre pour cette étude :

- préparation des données pour traitement : importer, vérifier et découper les données
- analyses univariées et bivariées, et des corrélations des variables deux à deux
- analyse en composantes principales
- classification hiérarchique ascendante.

1. Description des données

Import des données

La procédure pour lire ligne par ligne ces données est longue. Donc nous les avons exportées dans un fichier .txt pour aller plus vite.

```
data = read.table("data.txt", sep = " ", header=T, encoding = "UTF-8")
#data = read.table("assurance_complete_corrige.R") #, sep = "", header=T)
```

Analyses univariées

On a utilisé str pour afficher les informations simples concernant les variables, et summary pour afficher les données statistiques pour chaque variable.

On observe que les variables *region*, *crevpp*, *agecat* et *habi* sont qualitatives, malgré leur apparence numérique; on les convertit en facteurs

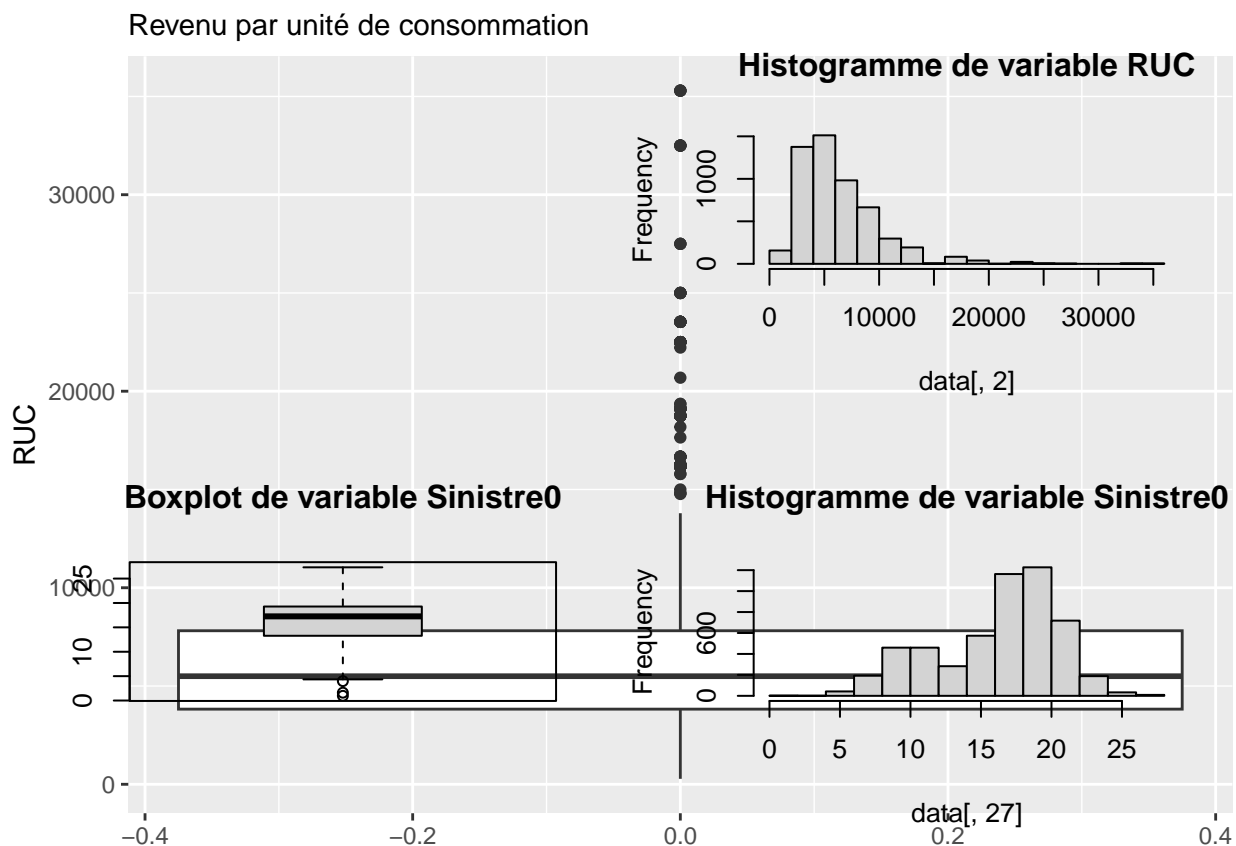
```
## [1] 1 2 3 4 5 7 8 9
```

On a représenté les boxplots des variables RUC et Sinistre0. Pour écraser les grandes valeurs, on utilise la fonction log.

```
par(mfrow=c(2,2))
boxplot(log(data[,2]),main="Boxplot de variable RUC")
ggplot(data, aes(y=RUC, fill=Durée)) + geom_boxplot(orientation = "x") + labs(subtitle = "Revenu par un")
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
## variable into a factor?
```

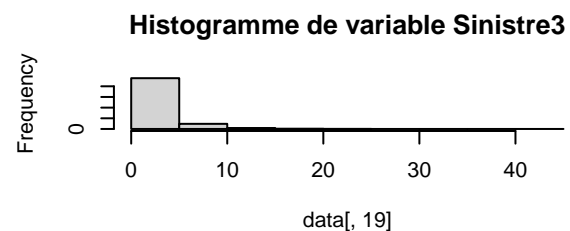
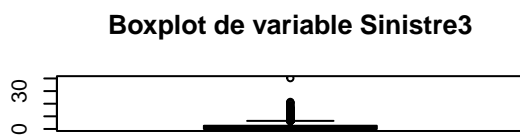
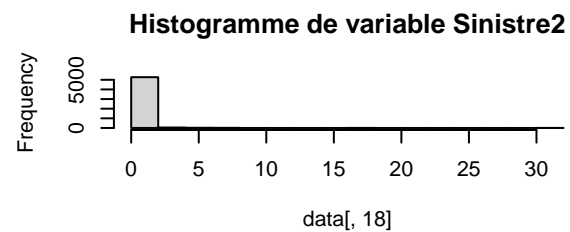
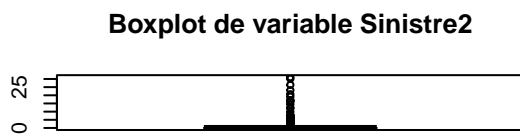
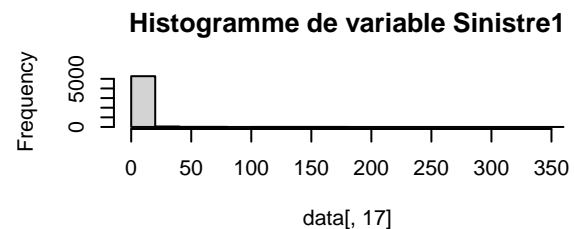
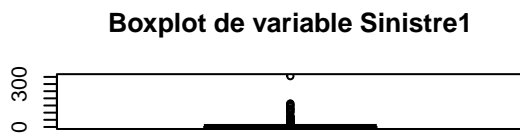
```
hist(data[,2],main="Histogramme de variable RUC")
# erreur ici chez Eva : stat_density requires an x or y aesthetic ggplot(data = data.frame(data[,2]))
boxplot(data[,27],main="Boxplot de variable Sinistre0")
hist(data[,27],main="Histogramme de variable Sinistre0")
```



```
density(data[,27])
```

```
##
## Call:
## density.default(x = data[, 27])
##
## Data: data[, 27] (5352 obs.); Bandwidth 'bw' = 0.6943
##
##      x              y
## Min.   :-1.118   Min.   :1.270e-06
## 1st Qu.: 6.509   1st Qu.:8.995e-04
## Median :14.136   Median :2.309e-02
## Mean   :14.136   Mean    :3.275e-02
## 3rd Qu.:21.763   3rd Qu.:4.731e-02
## Max.   :29.390   Max.    :1.226e-01
```

```
par(mfrow=c(3,2))
boxplot(data[,17],main="Boxplot de variable Sinistre1")
hist(data[,17],main="Histogramme de variable Sinistre1")
boxplot(data[,18],main="Boxplot de variable Sinistre2")
hist(data[,18],main="Histogramme de variable Sinistre2")
boxplot(data[,19],main="Boxplot de variable Sinistre3")
hist(data[,19],main="Histogramme de variable Sinistre3")
```

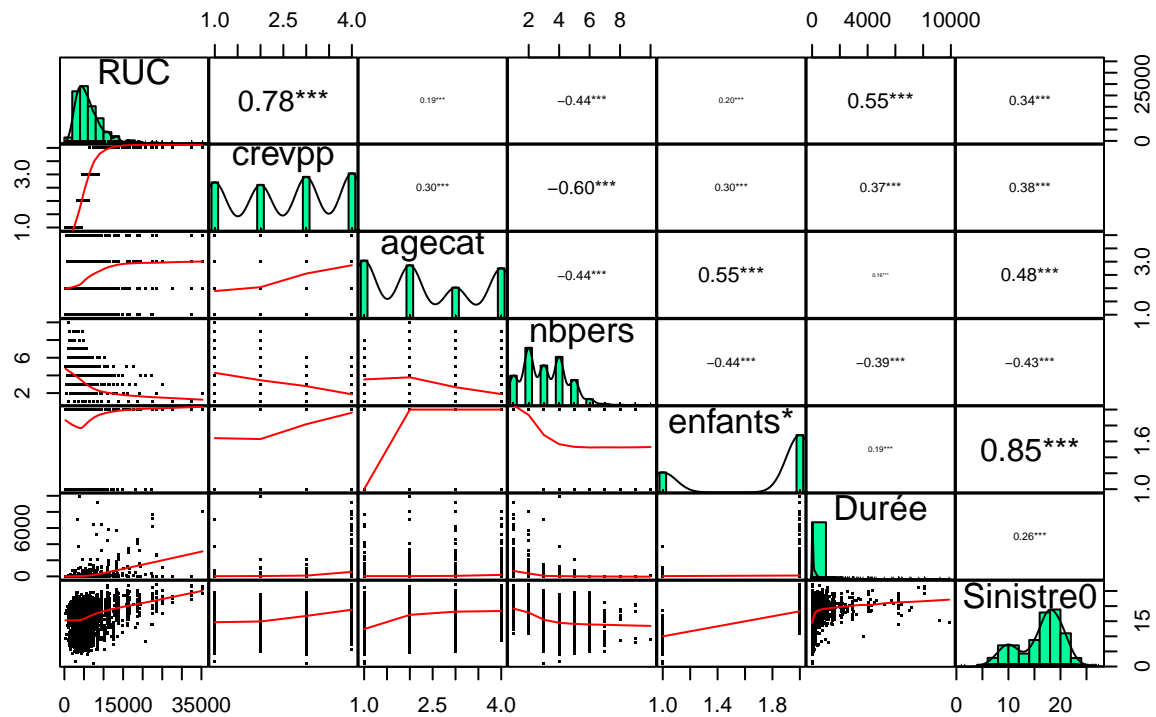


Analyses bivariées

A travers un graphique des variables numériques deux à deux, nous regardons comment évoluent les variables ensemble, et s'il y a des "tendances" reconnaissables. Par exemple, sur le graphique ci-dessous, qui contient les analyses bivariées complètes, on voit une tendance linéaire croissante (et corrélation positive significative)

entre *pib* et *recc*. (fonction trouvée à ref. 7: analyse bi + corrélations).

orrelations les plus significatives de Sinistre0 (variables deux par deux)



On peut constater que les variabilités des Sinistres des 3 types sont toutes grandes.

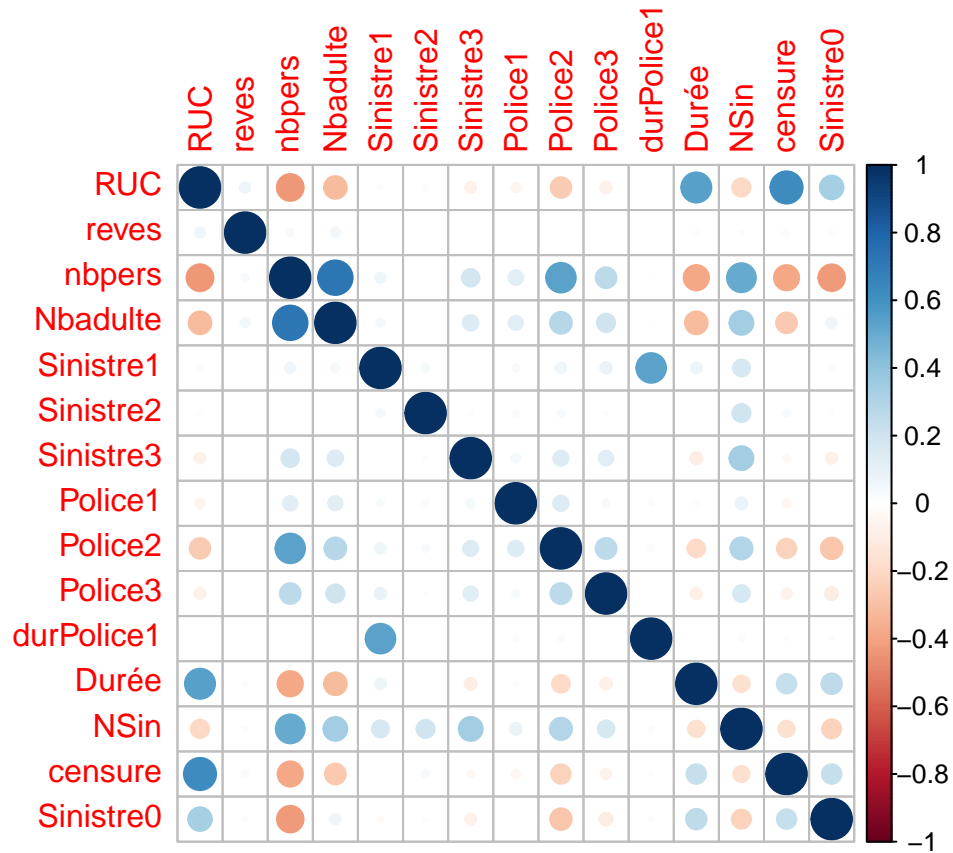
```
aggregate(data[,c(17,18,19,27)],list(data[,1]),mean)
```

```
##           Group.1 Sinistre1 Sinistre2 Sinistre3
## 1 Agr. exploitants 0.4860776 0.009051724 2.983017
## 2 Artisans, comm., chefs d'ent. 0.3559116 0.097458564 1.836381
## 3 Autres pers. sans activite prof. 1.4143169 0.107540984 1.761721
## 4 Cadres et prof. intellectuelles sup. 1.6937083 0.184058333 2.249010
## 5 Employes 1.2392184 0.159409429 1.771824
## 6 Ouvriers 1.6113979 0.128146194 1.841745
## 7 Professions intermediaires 1.7689008 0.207553551 2.214859
## 8 Retraites 0.5480867 0.185876093 1.395999
## Sinistre0
## 1 15.01849
## 2 16.12462
## 3 17.41922
## 4 16.83139
## 5 15.57469
## 6 14.30565
## 7 15.66231
## 8 18.36901
```

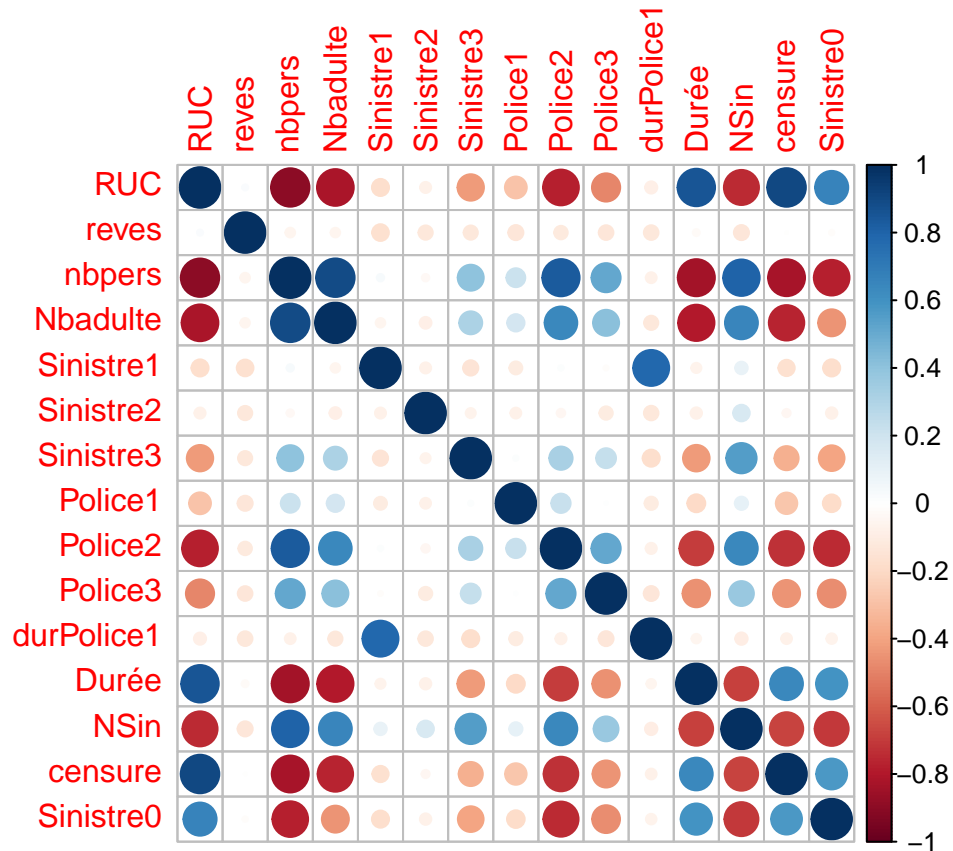
```
variables_quantitatives = data %>% select_if(is.numeric) %>% cor()
kable(variables_quantitatives, digits=3)
```

	RUC	reves	nbpers	Nbadulte	Sinistre1	Sinistre2	Sinistre3	Police1	Police2	Police3	durPolice1	Durée	NSin	censure	Sinistre0
RUC	1.000	0.066	-	-	-	0.017	-	-	-	-	0.005	0.549	-	0.628	0.338
reves	0.066	1.000	0.031	0.053	-	-	0.007	-	0.006	0.005	0.001	0.012	0.013	0.029	0.014
nbpers	-	0.031	1.000	0.720	0.064	0.006	0.188	0.129	0.531	0.262	0.009	-	0.508	-	-
Nbadulte	-	0.053	0.720	1.000	0.042	0.008	0.148	0.124	0.288	0.202	0.008	-	0.346	-	0.066
Sinistre1	-	-	0.064	0.042	1.000	0.042	-	0.030	0.070	0.082	0.534	0.073	0.175	-	-
Sinistre2	-	0.017	-	0.006	0.008	0.042	1.000	0.015	0.029	0.031	0.010	-	0.004	0.200	0.037
Sinistre3	-	0.007	0.188	0.148	-	0.015	1.000	0.048	0.140	0.130	-	-	0.346	-	-
Police1	-	-	0.129	0.124	0.030	0.029	0.048	1.000	0.146	0.035	0.020	0.013	0.087	-	0.000
Police2	-	0.006	0.531	0.288	0.070	0.031	0.140	0.146	1.000	0.265	0.028	-	0.297	-	-
Police3	-	0.005	0.262	0.202	0.082	0.010	0.130	0.035	0.265	1.000	-	-	0.172	-	-
durPolice1	-	0.005	0.001	0.009	0.008	0.534	-	-	0.020	0.028	-	1.000	-	0.021	0.011
Durée	0.549	0.012	-	-	0.073	0.004	-	0.013	-	-	-	1.000	-	0.231	0.256
NSin	-	0.013	0.508	0.346	0.175	0.200	0.346	0.087	0.297	0.172	0.021	-	1.000	-	-
censure	0.628	0.029	-	-	-	0.037	-	-	-	-	0.011	0.231	-	1.000	0.231
Sinistre0	0.338	0.014	-	0.066	-	0.012	-	0.000	-	-	0.011	0.256	-	0.231	1.000

```
corrplot(variables_quantitatives)
```

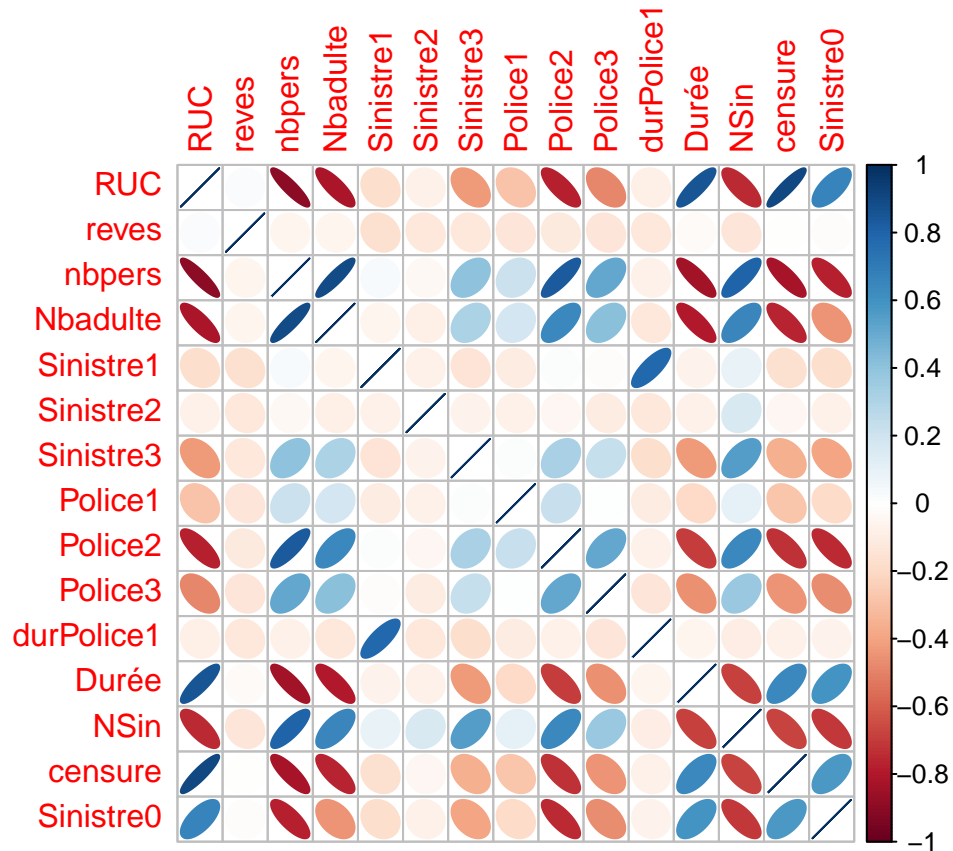


```
corrplot(cor(variables_quantitatives))
```



#Représentation des corrélations : plus l'ellipse ressemble à un cercle et moins les variables sont corrélées. Plus l'ellipse ressemble à une droite et plus les variables sont corrélées.

```
corrplot(cor(variables_quantitatives),method = "ellipse")
```



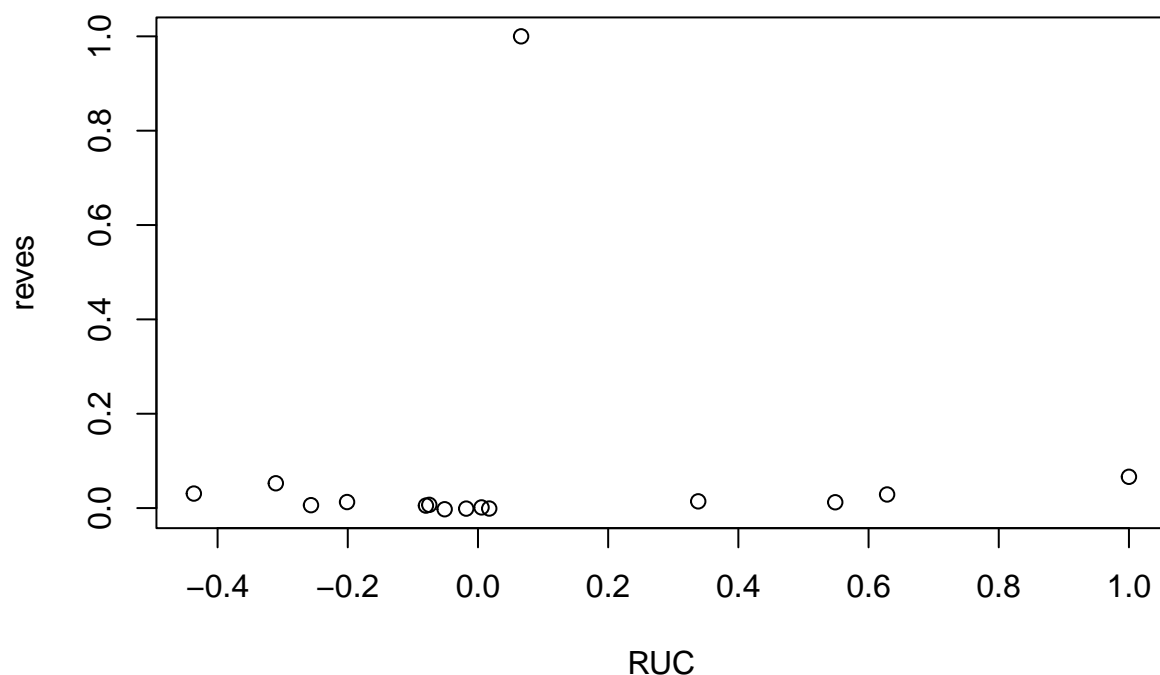
```
# Boxplot
# boxplot(data$column_name, main = "Boxplot of column_name", xlab = "Column name", ylab = "Values")

# Histogram
# hist(data$column_name, main = "Histogram of column_name", xlab = "Values", ylab = "Frequency", col =

# Estimateurs de la densité
# density_plot <- density(data$column_name, main = "Density Plot of column_name", xlab = "Values", ylab
# lines(density_plot, col = "red")

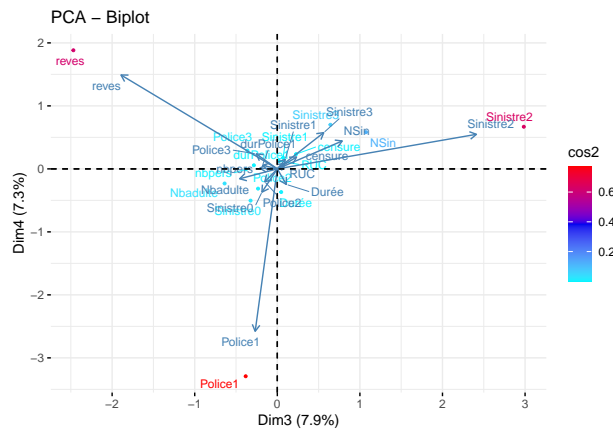
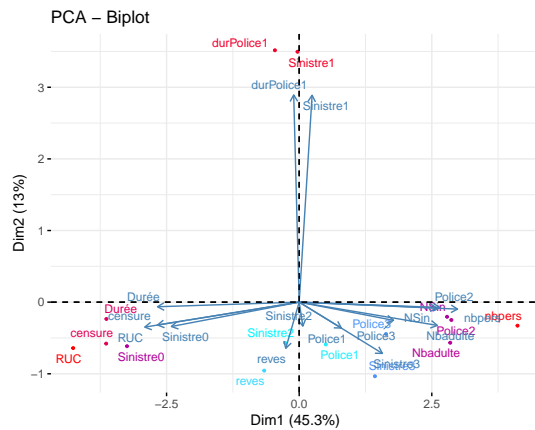
# Statistiques basiques
# summary(data$column_name)

plot(variables_quantitatives)
```

```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 15 individuals, described by 15 variables
## *The results are available in the following objects:
##
##   name                description
## 1  "$eig"              "eigenvalues"
## 2  "$var"              "results for the variables"
## 3  "$var$coord"        "coord. for the variables"
## 4  "$var$cor"          "correlations variables - dimensions"
## 5  "$var$cos2"         "cos2 for the variables"
## 6  "$var$contrib"      "contributions of the variables"
## 7  "$ind"              "results for the individuals"
## 8  "$ind$coord"        "coord. for the individuals"
## 9  "$ind$cos2"         "cos2 for the individuals"
## 10 "$ind$contrib"      "contributions of the individuals"
## 11 "$call"             "summary statistics"
## 12 "$call$centre"      "mean of the variables"
## 13 "$call$ecart.type"  "standard error of the variables"
## 14 "$call$row.w"       "weights for the individuals"
## 15 "$call$col.w"       "weights for the variables"
##
##   eigenvalue percentage of variance cumulative percentage of variance
## comp 1      6.79                  45.28                      45.28
## comp 2      1.94                  12.95                      58.24
## comp 3      1.18                   7.87                      66.10
## comp 4      1.09                   7.28                      73.39
## comp 5      1.02                   6.79                      80.18
```

## comp 6	0.81	5.41	85.59
## comp 7	0.70	4.67	90.26
## comp 8	0.45	2.99	93.25
## comp 9	0.36	2.40	95.65
## comp 10	0.29	1.95	97.60
## comp 11	0.17	1.15	98.75
## comp 12	0.13	0.87	99.62
## comp 13	0.04	0.27	99.89
## comp 14	0.02	0.11	100.00



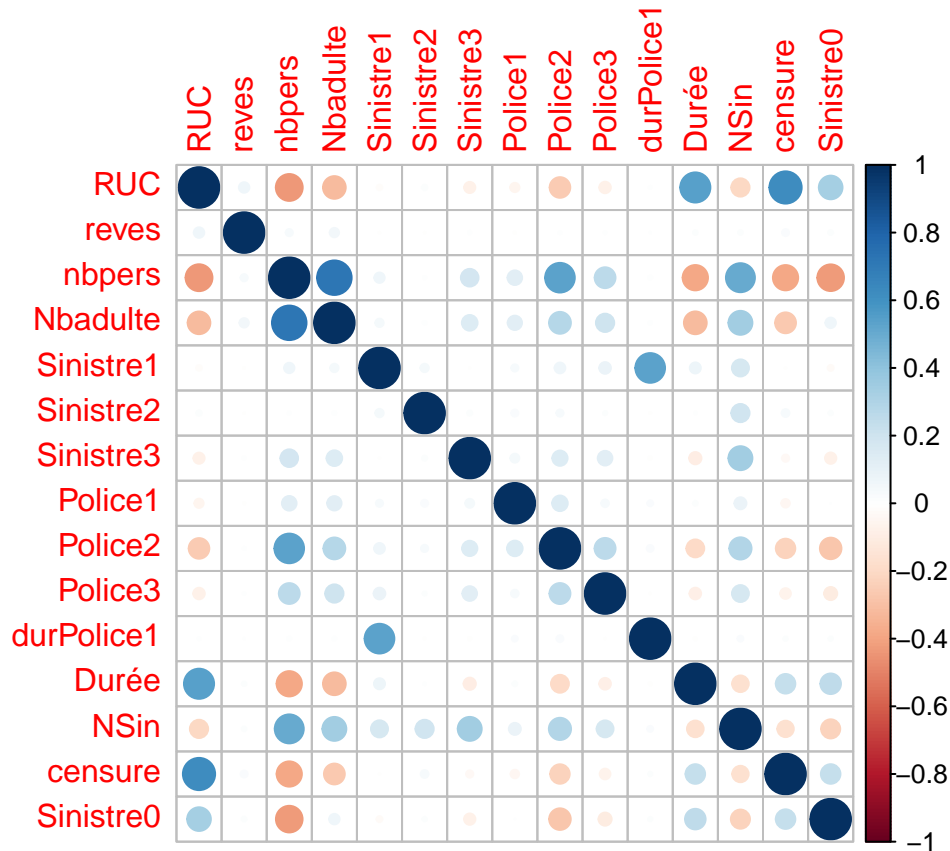
2. Modélisation des sinistres et des primes pures

2.1 Problème d'endogénéité dans les variables

Tester les problèmes d'endogénéité de certaines variables

```
# Selection des variables quantitatives
quant_vars <- sapply(data, is.numeric)

# Matrice de corrélation
cor_matrix <- cor(data[, quant_vars])
corrplot(cor_matrix)
```



```
# Tests de causalité de Granger pour toutes les paires de variables
# la fonction grangertest() permet de tester si une variable X est un prédicteur significatif d'une aut
quant_vars <- as.matrix(quant_vars)
d = data[, quant_vars]
for(i in 1:(ncol(d) - 1)){
  for(j in (i + 1):ncol(d)){

    result <- grangertest(d[,i], d[,j], order = 2)

    print(paste("Granger causality test entre ", colnames(d)[i],
                " et ", colnames(d)[j], ":", result[2,4]))
  }
}
```

```
## [1] "Granger causality test entre RUC et reves : 0.0923837756616444"
## [1] "Granger causality test entre RUC et nbpers : 0.00124220485708754"
## [1] "Granger causality test entre RUC et Nbadulte : 0.0172836315703496"
## [1] "Granger causality test entre RUC et Sinistre1 : 0.709483878283177"
## [1] "Granger causality test entre RUC et Sinistre2 : 0.0951611194510821"
## [1] "Granger causality test entre RUC et Sinistre3 : 0.0981000677158548"
## [1] "Granger causality test entre RUC et Police1 : 0.737522698532588"
## [1] "Granger causality test entre RUC et Police2 : 0.0123843096814322"
## [1] "Granger causality test entre RUC et Police3 : 0.912874664962623"
## [1] "Granger causality test entre RUC et durPolice1 : 0.234660821645245"
## [1] "Granger causality test entre RUC et Durée : 0.00278908899804597"
## [1] "Granger causality test entre RUC et NSin : 0.283574612045495"
## [1] "Granger causality test entre RUC et censure : 5.81376380854141e-09"
```

```

## [1] "Granger causality test entre RUC et Sinistre0 : 0.00736870516547412"
## [1] "Granger causality test entre reves et nbpers : 0.18968408454619"
## [1] "Granger causality test entre reves et Nbadulte : 0.36176438407316"
## [1] "Granger causality test entre reves et Sinistre1 : 0.970631057503471"
## [1] "Granger causality test entre reves et Sinistre2 : 0.934632216327838"
## [1] "Granger causality test entre reves et Sinistre3 : 0.87081127385712"
## [1] "Granger causality test entre reves et Police1 : 0.465186907224791"
## [1] "Granger causality test entre reves et Police2 : 0.366262300097807"
## [1] "Granger causality test entre reves et Police3 : 0.38259062929957"
## [1] "Granger causality test entre reves et durPolice1 : 0.99606824048349"
## [1] "Granger causality test entre reves et Durée : 0.89938550170445"
## [1] "Granger causality test entre reves et NSin : 0.956546525314893"
## [1] "Granger causality test entre reves et censure : 0.45554355536769"
## [1] "Granger causality test entre reves et Sinistre0 : 0.365219373654749"
## [1] "Granger causality test entre nbpers et Nbadulte : 0.193083295586577"
## [1] "Granger causality test entre nbpers et Sinistre1 : 0.558110531140403"
## [1] "Granger causality test entre nbpers et Sinistre2 : 0.49305523182955"
## [1] "Granger causality test entre nbpers et Sinistre3 : 0.944191036071153"
## [1] "Granger causality test entre nbpers et Police1 : 0.0381762395658839"
## [1] "Granger causality test entre nbpers et Police2 : 0.38758443522889"
## [1] "Granger causality test entre nbpers et Police3 : 0.79937810796575"
## [1] "Granger causality test entre nbpers et durPolice1 : 0.272894547109501"
## [1] "Granger causality test entre nbpers et Durée : 0.0905118575346298"
## [1] "Granger causality test entre nbpers et NSin : 0.0480562871228446"
## [1] "Granger causality test entre nbpers et censure : 0.00455159012190975"
## [1] "Granger causality test entre nbpers et Sinistre0 : 0.362139149025635"
## [1] "Granger causality test entre Nbadulte et Sinistre1 : 0.358541171759482"
## [1] "Granger causality test entre Nbadulte et Sinistre2 : 0.683643172870649"
## [1] "Granger causality test entre Nbadulte et Sinistre3 : 0.954146981569278"
## [1] "Granger causality test entre Nbadulte et Police1 : 0.130836636428486"
## [1] "Granger causality test entre Nbadulte et Police2 : 0.659774129946827"
## [1] "Granger causality test entre Nbadulte et Police3 : 0.50444501848523"
## [1] "Granger causality test entre Nbadulte et durPolice1 : 0.275432907148821"
## [1] "Granger causality test entre Nbadulte et Durée : 0.0210922967848045"
## [1] "Granger causality test entre Nbadulte et NSin : 0.381605011258915"
## [1] "Granger causality test entre Nbadulte et censure : 0.00262936227638448"
## [1] "Granger causality test entre Nbadulte et Sinistre0 : 0.348576217380067"
## [1] "Granger causality test entre Sinistre1 et Sinistre2 : 0.827711518387258"
## [1] "Granger causality test entre Sinistre1 et Sinistre3 : 0.570989882925704"
## [1] "Granger causality test entre Sinistre1 et Police1 : 0.736836059511227"
## [1] "Granger causality test entre Sinistre1 et Police2 : 0.288643457072958"
## [1] "Granger causality test entre Sinistre1 et Police3 : 0.6554707617976"
## [1] "Granger causality test entre Sinistre1 et durPolice1 : 0.976563675926668"
## [1] "Granger causality test entre Sinistre1 et Durée : 0.473758676644474"
## [1] "Granger causality test entre Sinistre1 et NSin : 0.36196682034159"
## [1] "Granger causality test entre Sinistre1 et censure : 0.38432432846914"
## [1] "Granger causality test entre Sinistre1 et Sinistre0 : 0.134747038950314"
## [1] "Granger causality test entre Sinistre2 et Sinistre3 : 0.300524713276513"
## [1] "Granger causality test entre Sinistre2 et Police1 : 0.10246562700862"
## [1] "Granger causality test entre Sinistre2 et Police2 : 0.158801981749613"
## [1] "Granger causality test entre Sinistre2 et Police3 : 0.370757787884079"
## [1] "Granger causality test entre Sinistre2 et durPolice1 : 0.98055544905637"
## [1] "Granger causality test entre Sinistre2 et Durée : 0.520300685937179"
## [1] "Granger causality test entre Sinistre2 et NSin : 0.498259153148845"

```

```
## [1] "Granger causality test entre Sinistre2 et censure : 0.652951329078083"
## [1] "Granger causality test entre Sinistre2 et Sinistre0 : 0.355856344901385"
## [1] "Granger causality test entre Sinistre3 et Police1 : 0.201632295337263"
## [1] "Granger causality test entre Sinistre3 et Police2 : 0.0553060810276148"
## [1] "Granger causality test entre Sinistre3 et Police3 : 0.688841817622712"
## [1] "Granger causality test entre Sinistre3 et durPolice1 : 0.820036164981182"
## [1] "Granger causality test entre Sinistre3 et Durée : 0.0979144699157492"
## [1] "Granger causality test entre Sinistre3 et NSin : 0.0590920136482579"
## [1] "Granger causality test entre Sinistre3 et censure : 0.420641926037821"
## [1] "Granger causality test entre Sinistre3 et Sinistre0 : 0.541651999430448"
## [1] "Granger causality test entre Police1 et Police2 : 0.667133611759315"
## [1] "Granger causality test entre Police1 et Police3 : 0.00354879403230157"
## [1] "Granger causality test entre Police1 et durPolice1 : 0.785179474093198"
## [1] "Granger causality test entre Police1 et Durée : 0.132813120000447"
## [1] "Granger causality test entre Police1 et NSin : 0.0203246545252461"
## [1] "Granger causality test entre Police1 et censure : 0.204917633529259"
## [1] "Granger causality test entre Police1 et Sinistre0 : 0.846171264937825"
## [1] "Granger causality test entre Police2 et Police3 : 0.310224237087878"
## [1] "Granger causality test entre Police2 et durPolice1 : 0.055655850228255"
## [1] "Granger causality test entre Police2 et Durée : 0.277199668621766"
## [1] "Granger causality test entre Police2 et NSin : 0.0239566356380064"
## [1] "Granger causality test entre Police2 et censure : 0.0692500258718777"
## [1] "Granger causality test entre Police2 et Sinistre0 : 0.269125934485199"
## [1] "Granger causality test entre Police3 et durPolice1 : 0.653405172596898"
## [1] "Granger causality test entre Police3 et Durée : 0.0113542062495392"
## [1] "Granger causality test entre Police3 et NSin : 0.000464286639831717"
## [1] "Granger causality test entre Police3 et censure : 3.1569081679773e-05"
## [1] "Granger causality test entre Police3 et Sinistre0 : 0.0436598567098487"
## [1] "Granger causality test entre durPolice1 et Durée : 0.961630268620227"
## [1] "Granger causality test entre durPolice1 et NSin : 0.939390629133645"
## [1] "Granger causality test entre durPolice1 et censure : 0.567426085872365"
## [1] "Granger causality test entre durPolice1 et Sinistre0 : 0.0396821681991153"
## [1] "Granger causality test entre Durée et NSin : 0.572377359207435"
## [1] "Granger causality test entre Durée et censure : 0.000658970825685853"
## [1] "Granger causality test entre Durée et Sinistre0 : 0.0983144988858764"
## [1] "Granger causality test entre NSin et censure : 0.774359173306188"
## [1] "Granger causality test entre NSin et Sinistre0 : 0.647017398450364"
## [1] "Granger causality test entre censure et Sinistre0 : 0.151342377417034"
```

Si on fixe $\alpha = 0.05$, alors il y a une causalité entre Sinistre0 et les variables suivantes : RUC/durPolice1. La méthode des MCO donne l'estimateur le plus efficient s'il n'y a pas d'endogénéité.

S'il y a de l'endogénéité, OLS (MCO) va donner des résultats inconsistants. L'estimateur des variables instrumentales va être consistant, mais inefficent.

```
# Régression linéaire multiple
modell1 <- lm(Sinistre0 ~ ., data = data)

# Afficher le résumé du modèle
summary(modell1)
```

```
##
## Call:
## lm(formula = Sinistre0 ~ ., data = data)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -9.5038 -1.4182 -0.0375  1.4986  8.0183
##
## Coefficients: (5 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.724e+01  5.071e-01  34.003  <2e-16
## pcsArtisans, comm., chefs d'ent.    1.050e-01  2.608e-01   0.403  0.6873
## pcsAutres pers. sans activite prof.    3.288e-02  2.660e-01   0.124  0.9016
## pcsCadres et prof. intellectuelles sup. 7.633e-02  2.369e-01   0.322  0.7473
## pcsEmployes   -6.494e-02  2.232e-01  -0.291  0.7711
## pcsOuvriers   -1.258e-01  2.145e-01  -0.587  0.5575
## pcsProfessions intermediaires -1.499e-01  2.227e-01  -0.673  0.5009
## pcsRetraites  -3.271e-02  2.485e-01  -0.132  0.8953
## RUC           2.154e-04  2.187e-05   9.849  <2e-16
## csModeste     6.520e-02  2.566e-01   0.254  0.7995
## csMoyenne Inf  1.390e-01  2.043e-01   0.680  0.4963
## csMoyenne Sup  1.672e-01  1.570e-01   1.065  0.2868
## revez        -3.535e-07  3.969e-07  -0.891  0.3731
## crevpp2eme quartile -5.961e-02  1.399e-01  -0.426  0.6701
## crevpp3eme quartile -1.857e-01  1.962e-01  -0.946  0.3440
## crevpp4eme quartile -2.174e-01  2.445e-01  -0.889  0.3741
## region2       -3.011e-01  2.113e-01  -1.425  0.1542
## region3       -2.570e-01  2.304e-01  -1.116  0.2646
## region4       -2.103e-01  2.198e-01  -0.957  0.3388
## region5       -1.379e-01  2.136e-01  -0.645  0.5187
## region7        7.066e-02  2.208e-01   0.320  0.7490
## region8       -2.338e-01  2.167e-01  -1.079  0.2807
## region9       -1.347e-01  2.222e-01  -0.606  0.5443
## habi1         4.859e-02  1.266e-01   0.384  0.7011
## habi2         9.944e-02  1.435e-01   0.693  0.4885
## habi3        -1.135e-01  1.439e-01  -0.789  0.4304
## habi4        -1.486e-01  1.289e-01  -1.153  0.2488
## habi5        -9.448e-02  1.280e-01  -0.738  0.4603
## habi6        -9.277e-03  1.207e-01  -0.077  0.9387
## habi7         5.967e-02  9.343e-02   0.639  0.5231
## habi8        -1.180e-01  2.221e-01  -0.531  0.5954
## AhabiParis + Agglomeration          NA          NA          NA          NA
## AhabiUn. urb. de 10 000 a 99 999 hab.    NA          NA          NA          NA
## AhabiUn. urb. de 100 000 hab. et +      NA          NA          NA          NA
## AhabiUn. urb. de 2 000 a 9 999 hab.      NA          NA          NA          NA
## AtyphNon declare    1.006e-01  2.579e-01   0.390  0.6964
## AtyphProprietaire  -1.998e-02  6.863e-02  -0.291  0.7710
## agecat41-50        2.780e-03  9.272e-02   0.030  0.9761
## agecat51-60       -7.727e-02  1.126e-01  -0.686  0.4925
## agecat61-96       -1.032e-01  1.668e-01  -0.619  0.5362
## AcompmCouple avec enfant(s) -8.070e+00  1.256e-01 -64.245  <2e-16
## AcompmCouple sans enfant    1.130e-01  1.127e-01   1.002  0.3162
## AcompmPersonne seule    1.101e-01  1.628e-01   0.676  0.4988
## nbpers           1.817e-02  6.336e-02   0.287  0.7743
## enfantsPas d'enfants          NA          NA          NA          NA
## AnatMenage francais -2.970e-01  2.053e-01  -1.447  0.1481
## AnatNon declare   -4.437e-01  2.357e-01  -1.882  0.0599
## BautoPas de vehicule -1.930e-02  1.198e-01  -0.161  0.8720
## Nbadulte         -2.155e-02  6.865e-02  -0.314  0.7536

```

## Sinistre1	-5.460e-03	3.989e-03	-1.369	0.1711
## Sinistre2	2.712e-02	2.643e-02	1.026	0.3049
## Sinistre3	5.726e-03	1.178e-02	0.486	0.6269
## Police1	8.886e-03	6.099e-03	1.457	0.1452
## Police2	2.874e-03	2.730e-03	1.053	0.2924
## Police3	6.160e-03	1.321e-02	0.466	0.6411
## durPolice1	1.137e-12	9.263e-13	1.227	0.2197
## Durée	2.832e-05	6.300e-05	0.450	0.6531
## NSin	1.200e-02	9.962e-03	1.205	0.2284
## censure	9.184e-02	1.205e-01	0.762	0.4459
##				
## (Intercept)		***		
## pcsArtisans, comm., chefs d'ent.				
## pcsAutres pers. sans activite prof.				
## pcsCadres et prof. intellectuelles sup.				
## pcsEmployes				
## pcsOuvriers				
## pcsProfessions intermediaires				
## pcsRetraites				
## RUC		***		
## csModeste				
## csMoyenne Inf				
## csMoyenne Sup				
## reve				
## crevpp2eme quartile				
## crevpp3eme quartile				
## crevpp4eme quartile				
## region2				
## region3				
## region4				
## region5				
## region7				
## region8				
## region9				
## habi1				
## habi2				
## habi3				
## habi4				
## habi5				
## habi6				
## habi7				
## habi8				
## AhabiParis + Agglomeration				
## AhabiUn. urb. de 10 000 a 99 999 hab.				
## AhabiUn. urb. de 100 000 hab. et +				
## AhabiUn. urb. de 2 000 a 9 999 hab.				
## AtyphNon declare				
## AtyphProprietaire				
## agecat41-50				
## agecat51-60				
## agecat61-96				
## AcompmCouple avec enfant(s)		***		
## AcompmCouple sans enfant				
## AcompmPersonne seule				

```

## nbpers
## enfantsPas d'enfants
## AnatMenage francais
## AnatNon declare .
## BautoPas de vehicule
## Nbadulte
## Sinistre1
## Sinistre2
## Sinistre3
## Police1
## Police2
## Police3
## durPolice1
## Durée
## NSin
## censure
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.148 on 5298 degrees of freedom
## Multiple R-squared:  0.7524, Adjusted R-squared:  0.7499
## F-statistic: 303.8 on 53 and 5298 DF,  p-value: < 2.2e-16
summary(selectionAIC)

##
## Call:
## lm(formula = Sinistre0 ~ RUC + Acompm + Police1 + NSin, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4423 -1.4309 -0.0272  1.4871  7.8258
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.682e+01  8.389e-02  200.442  <2e-16 ***
## RUC             2.081e-04  8.701e-06   23.923  <2e-16 ***
## AcompmCouple avec enfant(s) -8.009e+00  7.598e-02 -105.410  <2e-16 ***
## AcompmCouple sans enfant    3.721e-02  8.351e-02    0.446  0.6560
## AcompmPersonne seule     -1.015e-02  1.025e-01   -0.099  0.9211
## Police1         1.050e-02  5.889e-03    1.782  0.0748 .
## NSin           1.602e-02  8.531e-03    1.878  0.0604 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.146 on 5345 degrees of freedom
## Multiple R-squared:  0.7507, Adjusted R-squared:  0.7504
## F-statistic: 2682 on 6 and 5345 DF,  p-value: < 2.2e-16

```

2.2 Modélisation de Sinistre0

2.3 Modélisation de Sinistre 1 ou 2 ou 3 (au moins un)

notamment pour Sinistre1 à 3 on choisira entre modèle gamma combiné à probit/logit, tobit, tobit généralisé ou double hurdle pour des variables bien choisies

2.4 Modèle pour le prix de Police 1 ou 2 ou 3 (au moins un)

2.5 Modèle retenu au final

Le choix du modèle retenu au final et les critères choisis devront être justifiés.

IV regressions

The four kinds of variables in IV

- Y = outcome variables
- X = endogenous, causal variable(s)
- Z = instrument(s): doivent être exogènes, c'est-à-dire leur influence sur Y se fait seulement via leur influence sur X, la var endogène
- W = any exogenous variables not including instruments

3. Modélisation du nombre de sinistres et tarification des nouveaux arrivants

3.1 Modèle pour le nombre de sinistres, NSin

3.2 Méthode de tarification pour les nouveaux arrivants

4. Estimation des durées

4.1 Estimateur de Kaplan-Meier

4.2 Modèle de Cox