

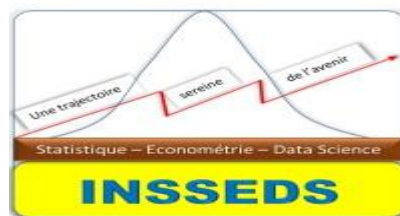


Ministère de l'Enseignement
Supérieur et de la
Recherche Scientifique

REPUBLIQUE DE CÔTE D'IVOIRE



Union – Discipline – Travail



**INSTITUT SUPERIEUR DE STATISTIQUE, D'ECONOMETRIE
ET DE DATASCIENCE**

MASTER 2

STATISTIQUE ECONOMETRIE DATA SCIENCE

**MINI PROJET ECONOMETRIE DES VARIABLES
QUALITATIVES**

**ETUDE DES VARIABLES INFLUENCANT LE
RISQUE D'ACCIDENT A L'AIDE DE
METHODES ECONOMETRIQUE**

MODELE LOGISTIQUE ET MODELE POISSON

ANNEE UNIVERSITAIRE : 2024-2025

ETUDIANT :

N'DRI ONESIME

ENSEIGNANT :

AKPOSSO DIDIER MARTIAL

AVANT-PROPOS :

L'association du théorique au pratique, des connaissances aux compétences et des savoir-faire aux savoirs est la principale tendance récente dans le secteur technique. Dans ce contexte, l'INSEDS (Institut Supérieur de la Statistique, d'Econométrie et de la Data Science), dans sa formation en master professionnel en statistique, économie et science des données, impose que les divers crédits soient validés en effectuant un mini-projet à la fin de chaque module. Le projet est donc structuré et supervisé de cette manière, visant principalement à faire de chaque élève un participant dynamique, engagé et libre dans la vie active.

Ce document est un rapport de projet Muni axé sur l'économétrie des variables quantitatives. Il se divise principalement en trois parties : Prétraitements des données, Analyse univariée, Analyse bivariée, Estimation de la proportion des accidentés, Test de conformité à la proportion d'accidentés, Test de liaison de la variable cible avec les autres variables, Modélisation LOGISTIQUE, Modélisation POISSONNIENNE.

En règle générale, toutes les analyses et conclusions présentées dans ce rapport relèvent de la responsabilité de l'auteur, qui ne sollicite ni Autrui ni l'INSEDS (Institut Supérieur de Statistique d'Econométrie et de Data Science).

Table des matières

INTRODUCTION	5
Contexte et justification de l'étude	5
Problématique	5
Principaux résultats attendus	6
Méthodologie	6
Description du jeu de données : dictionnaire des données	6
I- Prétraitements des données	7
<input type="checkbox"/> Visualisation des données.....	7
<input type="checkbox"/> Transformation de certaines variables en variables catégorielles	8
<input type="checkbox"/> Resumé numérique	8
<input type="checkbox"/> Traitement des doublons.....	9
<input type="checkbox"/> Traitement des valeurs manquantes.....	9
<input type="checkbox"/> Traitement des valeurs abberantes.....	10
II- ANALYSE UNIVARIEE.....	11
II-1 Variables quantitatives	11
II-2 Variables qualitatives.....	17
III- ANALYSE BIVARIEE	25
IV- ESTIMATION DE LA PROPORTION DES ACCIDENTES ..	45
V- TEST DE CONFORMITE	45
VI- TEST DE LIAISON.....	46
V- REGRESSION LOGISTIQUE	62
<input type="checkbox"/> Specification du modele.....	62
<input type="checkbox"/> Anova	62
<input type="checkbox"/> Estimation des paramètres	63
<input type="checkbox"/> ODD-RATIO	64

<input type="checkbox"/>	Interpretation des Odd Ratios.....	64
<input type="checkbox"/>	Effets marginaux.....	65
<input type="checkbox"/>	Taux de mauvais classement (TMC)	67
	V- REGRESSION POISSON	67
<input type="checkbox"/>	Importation du jeu de données.....	67
<input type="checkbox"/>	Graphique.....	68
<input type="checkbox"/>	Test d'adéquation à la loi de poisson	68
<input type="checkbox"/>	ESTIMATION DU MODELE	69
<input type="checkbox"/>	Test d'anova permet de voir les variables les plus liées à la cible	70
<input type="checkbox"/>	Reprenons le modèle avec ces variables	71
<input type="checkbox"/>	Anova	72
<input type="checkbox"/>	Prenons l'exponentiel des coefficients	73
<input type="checkbox"/>	Interpretation	74
<input type="checkbox"/>	ANALYSE DES RESIDUS.....	75
<input type="checkbox"/>	FAIRE LA PREVISION	76
	CONCLUSION GENERALE.....	77
	Recommandations pour la Société d'Assurance	78
	POWER BI	79

INTRODUCTION

Contexte et justification de l'étude

L'industrie de l'assurance automobile est en pleine mutation, avec une demande croissante pour des modèles de tarification plus précis et personnalisés. Les assureurs cherchent à mieux comprendre les facteurs qui influencent le risque d'accident et la gravité des sinistres afin d'ajuster leurs primes en conséquence. Le dataset `assurance_auto_makani.csv` offre une opportunité unique d'explorer ces facteurs à travers une variété de variables, allant des caractéristiques démographiques des conducteurs aux conditions environnementales lors des accidents. Cette étude vise à exploiter ces données pour fournir des insights utiles aux assureurs et aux parties prenantes du secteur.

Problématique

La problématique centrale de cette étude est de déterminer quels sont les facteurs les plus influents dans la survenue et la gravité des accidents automobiles. En particulier, nous cherchons à répondre aux questions suivantes:

- Quelles sont les caractéristiques des conducteurs (âge, sexe, etc.) les plus associées à un risque accru d'accident ?
- Comment les caractéristiques du véhicule (âge, marque, type de carburant, etc.) influencent-elles la probabilité d'accident et la gravité des sinistres ?
- Quel est l'impact des conditions environnementales (météo, éclairage, état de la route) sur la survenue des accidents ?
- Existe-t-il des combinaisons de facteurs qui augmentent significativement le risque d'accident ou la gravité des sinistres ?

Principaux résultats attendus

Les résultats attendus de cette étude incluent :

- **L'identification des variables les plus prédictives du risque d'accident et de la gravité des sinistres.**
- **La construction de modèles de prédiction permettant d'estimer la probabilité d'accident et la gravité des sinistres en fonction des caractéristiques des conducteurs, des véhicules et des conditions environnementales.**
- **Des recommandations pour les assureurs sur la manière d'ajuster leurs primes en fonction des risques identifiés.**
- **Des insights sur les interactions entre les différentes variables et leur impact combiné sur le risque d'accident.**

Méthodologie

La méthodologie de cette étude comprendra les étapes suivantes :

La méthodologie proposée comprend : - Prétraitement des données - Analyse univariée et bivariée – Estimation du nombre accident- Test de liaison - Modélisation de régression logistique - Modélisation poisson

Description du jeu de données : dictionnaire des données

- **Age: Age**
- **Sexe: Sexe**
- **Vehicle_Age: Age du véhicule**
- **MARQUES: Marques**
- **Couleur: Couleur**
- **fuel_type: Type de carburant**
- **seat_count: Nombre de sièges**
- **door_count: Nombre de portes**

- **manufacture_year**: Année de fabrication
- **transmission**: Transmission
- **ACCIDENT**: Accident (Oui/Non)
- **GRAVITÉ**: Gravité
- **Trajet**: Trajet
- **light_conditions**: Conditions d'éclairage
- **meteo_conditions**: Conditions météorologiques
- **Road_surface_conditions** : Conditions de la surface de la route
- **manv**: Manœuvre
- **fréquence**: Fréquence
- **Annual_Premium**: Prime annuelle

I- Prétraitements des données

Avant l'entame de la deuxième partie du Prétraitement des données, il est important de signifier que quelques modifications ont été appliqués sur le jeu de données telles que la modification des types de données, le renommage des modalités de quelques variables, le traitement des doublons. Tout cela s'est fait dans le logiciel ExcelPower-Query.

Visualisation des données

```
tibble [382,154 × 18] (S3: tbl_df/tbl/data.frame)
  Age                : num [1:382154] 22 22 22 22 22 22 22 22 22 22 ...
  Sexe               : chr [1:382154] "femme" "femme" "homme" "homme" ..
  Vehicle_Age        : chr [1:382154] "moins de 5 ans" "plus de 5 ans" "plus de 5
ans" "moins de 5 ans" ...
  Marque             : chr [1:382154] "NISSAN" "NISSAN" "NISSAN" "NISSAN" ...
  Couleur            : chr [1:382154] "gris" "argent" "argent" "gris" ..
  fuel_type          : chr [1:382154] "Gazoil" "Gazoil" "Gazoil" "Gazoil" ...
  nb_siege           : num [1:382154] 21 21 21 21 21 21 21 21 21 21 ...
  nb_porte           : num [1:382154] 4 4 4 4 4 4 4 4 4 4 ...
  Annee_fabric       : num [1:382154] 2007 2008 2000 2010 2006 ...
  Transmission       : chr [1:382154] "man" "man" "man" "man" ...
  Gravite_accident   : chr [1:382154] "grave" "modere" "grave" "grave" .
  Trajet             : chr [1:382154] "trajet 1" "trajet 1" "trajet 3" "trajet 3"
...
  Condition_eclairage : chr [1:382154] "eclairage 3" "eclairage 2" "eclairage 2"
"eclairage 1" ...
  Condition_meteo     : chr [1:382154] "meteo 1" "meteo 1" "meteo 1" "meteo 1" ...
  Condition_surface_route: chr [1:382154] "surface 1" "surface 1" "surface 1" "surface
1" ...
  manv               : num [1:382154] 1 17 1 9 1 15 8 1 10 1 ...
```

```
Annual_Premium      : num [1:382154] 2630 43327 35841 27645 29023 ...
accident            : chr [1:382154] "Oui" "Oui" "Oui" "Oui" ...
```

Ce jeu de donnée comprend 382154 observations et 18 variables dont la majorité est en chaîne de caractère que nous devons convertir en variable catégorielle.

Transformation de certaines variables en variables catégorielles

```
tibble [382,154 × 18] (S3: tbl_df/tbl/data.frame)
 Age                : num [1:382154] 22 22 22 22 22 22 22 22 22 22 ...
 Sexe              : Factor w/ 2 levels "femme","homme": 1 1 2 2 1 2 1
1 2 1 ...
 Vehicle_Age       : Factor w/ 2 levels "moins de 5 ans",...: 1 2 2 1 1
1 2 2 2 ...
 Marque            : Factor w/ 13 levels "AUDI","BMW","DAF",...: 10 10
10 10 10 10 10 10 13 13 ...
 Couleur           : Factor w/ 11 levels "argent","blanc",...: 5 1 1 5 3
5 8 2 3 5 ...
 fuel_type         : Factor w/ 2 levels "Gazoil","Super": 1 1 1 1 1 1 1
1 1 1 ...
 nb_siege          : num [1:382154] 21 21 21 21 21 21 21 21 21 21 ...
 nb_porte          : num [1:382154] 4 4 4 4 4 4 4 4 4 4 ...
 Annee_fabric      : Factor w/ 21 levels "2000","2001",...: 8 9 1 11 7 8
11 8 8 11 ...
 Transmission      : Factor w/ 2 levels "auto","man": 2 2 2 2 2 2 2 2 2
2 ...
 Gravite_accident  : Factor w/ 4 levels "aucun","grave",...: 2 4 2 2 2 2
4 4 4 2 ...
 Trajet            : Factor w/ 7 levels "trajet 0","trajet 1",...: 2 2 4
4 2 5 4 6 2 4 ...
 Condition_eclairage : Factor w/ 6 levels "eclairage 0",...: 4 3 3 2 2 2 2
2 2 2 ...
 Condition_meteo   : Factor w/ 5 levels "meteo 0","meteo 1",...: 2 2 2 2
2 2 2 2 2 ...
 Condition_surface_route: Factor w/ 5 levels "surface 0","surface 1",...: 2 2
2 2 2 2 2 2 2 ...
 manv              : num [1:382154] 1 17 1 9 1 15 8 1 10 1 ...
 Annual_Premium    : num [1:382154] 2630 43327 35841 27645 29023 ...
 accident          : Factor w/ 2 levels "Non","Oui": 2 2 2 2 2 2 2 2 2
2 ...
```

Résumé numérique

Age	Sexe	Vehicle_Age	Marque
Min. :20.0	femme:205603	moins de 5 ans:166353	TOYOTA :194602
1st Qu.:25.0	homme:176551	plus de 5 ans :215801	SUZUKI : 57938
Median :36.0			NISSAN : 21701
Mean :38.5			IVECO : 16133
3rd Qu.:49.0			MERCEDES: 15817
Max. :80.0			KIA : 15254
			(Other) : 60709
Couleur	fuel_type	nb_siege	nb_porte


```

noir :224205 Gazoil:243589 Min. : 3.00 Min. :2.00
blanc : 50147 Super :138565 1st Qu.: 5.00 1st Qu.:4.00
gris : 31826 Median : 5.00 Median :4.00
argent : 30886 Mean :11.17 Mean :3.98
bleu : 21707 3rd Qu.:21.00 3rd Qu.:4.00
rouge : 9799 Max. :47.00 Max. :4.00
(Other): 13584
Annee_fabric Transmission Gravite_accident Trajet
2010 : 86023 auto:146811 aucun :143473 trajet 0:143473
2015 : 47711 man :235343 grave :195418 trajet 1: 19682
2012 : 32593 leger : 2955 trajet 2: 3494
2014 : 23630 modere: 40308 trajet 3: 46143
2008 : 21829 trajet 4:107919
2007 : 21607 trajet 5: 49111
(Other):148761 trajet 6: 12332
Condition_eclairage Condition_meteo Condition_surface_route manv
eclairage 0:143473 meteo 0:143473 surface 0:143473 Min. : 0.000
eclairage 1:171492 meteo 1:196256 surface 1:177890 1st Qu.: 0.000
eclairage 2: 49029 meteo 2: 26656 surface 2: 58859 Median : 1.000
eclairage 3: 1628 meteo 3: 14652 surface 3: 241 Mean : 2.177
eclairage 4: 11917 meteo 4: 1117 surface 4: 1691 3rd Qu.: 1.000
eclairage 5: 4615 Max. :18.000

Annual_Premium accident
Min. : 2630 Non:143473
1st Qu.: 24546 Oui:238681
Median : 31692
Mean : 30711
3rd Qu.: 39448
Max. :540165

```

Traitement des doublons

Nombre de doublons : 7897

Traitement des valeurs manquantes

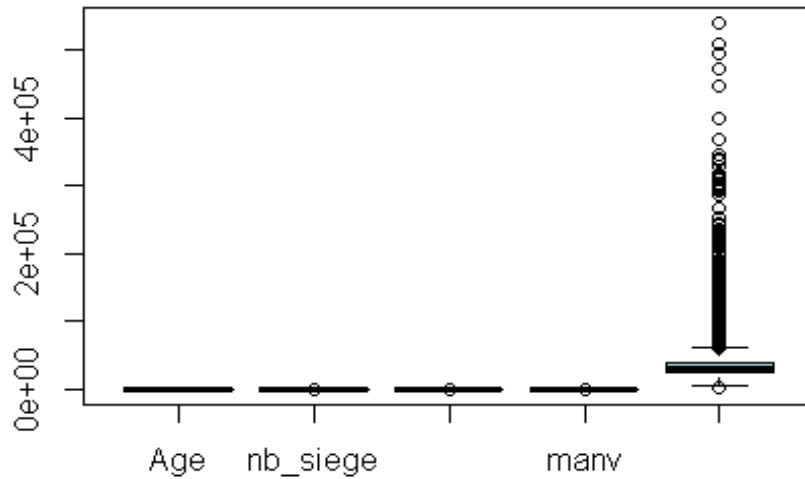
Nombre de valeurs manquantes par variable :

Age	Sexe	Vehicle_Age
0	0	0
Marque	Couleur	fuel_type
0	0	0
nb_siege	nb_porte	Annee_fabric
0	0	0
Transmission	Gravite_accident	Trajet
0	0	0
Condition_eclairage	Condition_meteo	Condition_surface_route
0	0	0
manv	Annual_Premium	accident
0	0	0

Traitement des valeurs abberantes

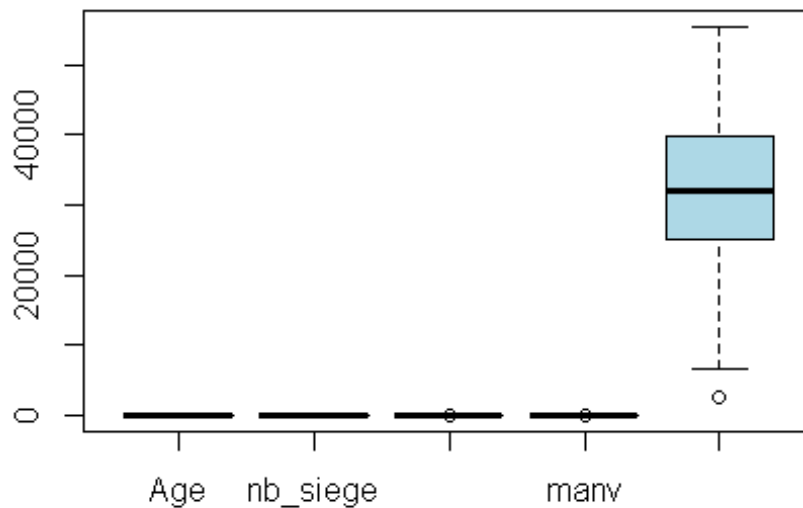
- Visualisation des valeurs abberantes avant traitement

Boîtes à moustache



- Visualisation des valeurs abberantes après traitement

Boîtes à moustache



Le prétraitement des données ainsi terminé nous pouvons passer aux différentes analyses en commençant par l'analyse univariée de la variable cible qui est accident et des autres variables d'intérêt.

II- ANALYSE UNIVARIEE

II-1 Variables quantitatives

a) Variable Age

• Tableau

Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Frequence	Freq_Cum_crois	Freq_Cum_decrois
20	5699	5699	374257	0.0152	0.0152
21	15670	21369	373377	0.0419	0.0571
22	21330	42699	372516	0.0570	0.1141
23	24262	66961	371413	0.0648	0.1789
24	26014	92975	370152	0.0695	0.2484
Freq_Cum_decrois					
20		1.0000			
21		0.9976			
22		0.9953			
23		0.9924			
24		0.9890			

92975 assurés ont au plus 24 ans

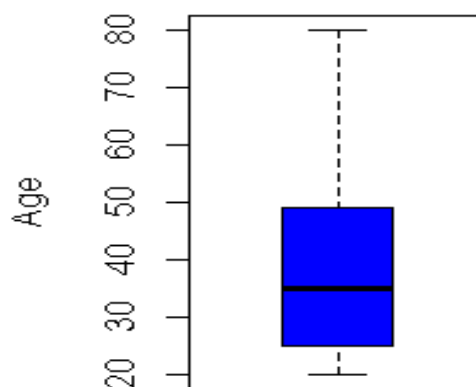
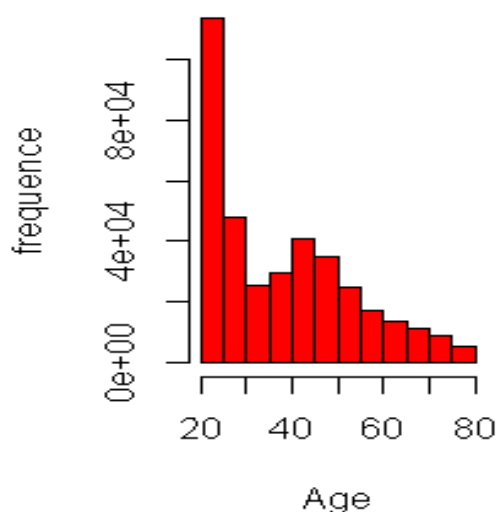
370152 assurés ont au moins 24 ans

24,84% des assurés auto ont au plus 24 ans

98,90% des assurés auto ont au moins 25 ans

• Graphique

histogramme des ages boites à moustaches ag



- Resume numériques

```

minimum
[1] 20

maximum
[1] 80

mode
[1] 24

médiane
[1] 35

moyenne
[1] 38.35317

quantile
  0%  25%  50%  75% 100%
  20  25  35  49  80

coefficient_variation
[1] 39.71871

variance
[1] 232.056

ecart_type
[1] 15.23339

coefficient_assymetrie
[1] 0.7159275

interpretation_skewness
[1] "distribution etalee a droite"

coefficient_applatissage
[1] 2.533606

interpretation_kurtosis
[1] "distribution platikurtique"

```

b) Variable nb_siege

- Tableau

	Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Frequence	Freq_Cum_crois
5	226468	226468	374257	0.6051	0.6051
8	7530	233998	233998	0.0201	0.6252
21	140259	374257	226468	0.3748	1.0000
	Freq_Cum_decrois				
5	1.0000				
8	0.6252				
21	0.6051				

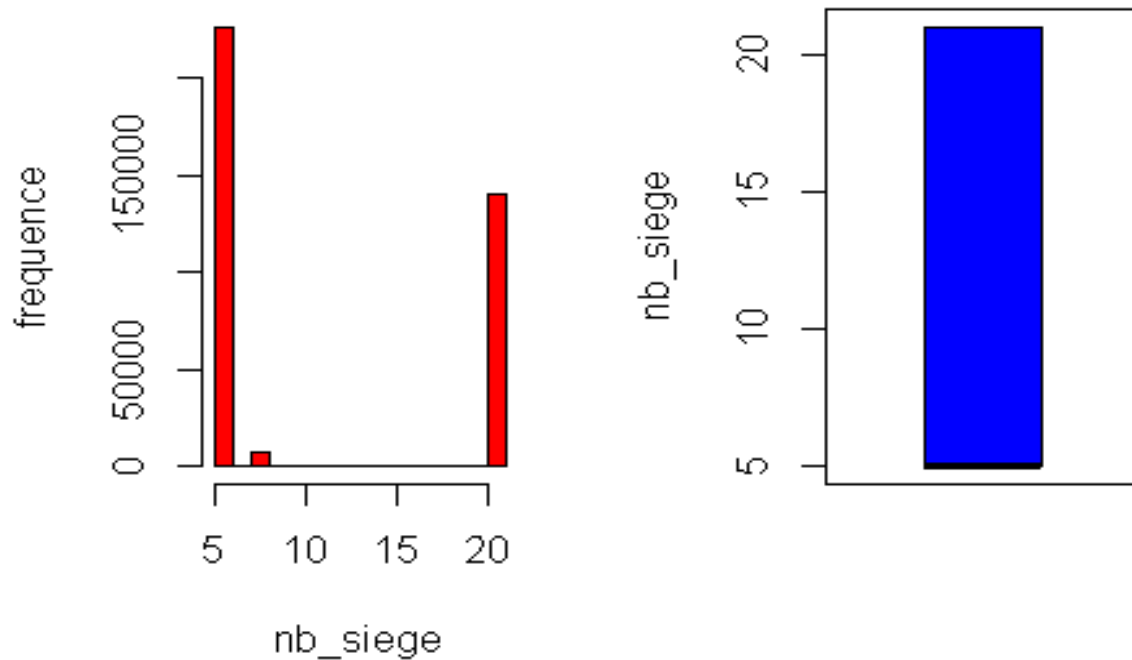
226468 assurés auto ont des véhicules qui ont au plus 5 sièges

374257 assurés auto ont des véhicules qui ont au moins 5 sièges

60,51% des assurés auto ont des véhicules qui ont au plus 5 sièges

- Graphique

histogramme des sièges boîtes à moustaches sie



- Resume numériques

```

minimum
[1] 5

maximum
[1] 21

mode
[1] 5

mediane
[1] 5

moyenne
[1] 11.05662

quantile
  0%  25%  50%  75% 100%
   5   5   5  21  21

coefficient_variation
[1] 69.72883

```

```

variance
[1] 59.43877

ecart_type
[1] 7.709655

coefficient_assymetrie
[1] 0.5093754

interpretation_skewness
[1] "distribution etalee a droite"

coefficient_applatissage
[1] 1.267924

interpretation_kurtosis
[1] "distribution platikurtique"

```

c) Variable nb_porte

- Tableau

	Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Frequence	Freq_Cum_crois
2	3806	3806	374257	0.0102	0.0102
4	370451	374257	3806	0.9898	1.0000
	Freq_Cum_decrois				
2		1.0000			
4		0.0102			

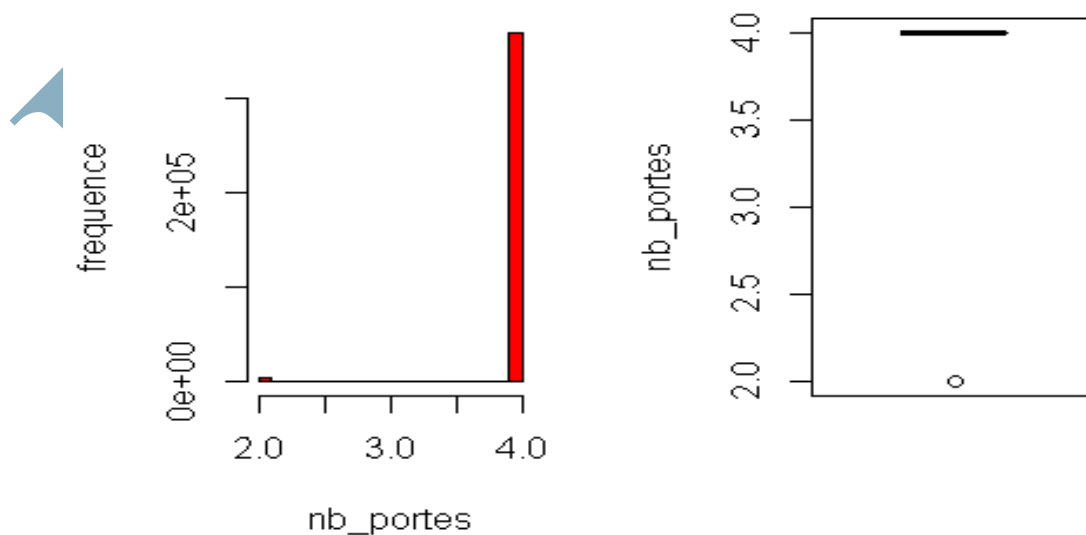
3806 assurés auto ont des véhicules qui ont au plus 2 portières

374257 assurés auto ont des véhicules qui ont au moins 2 portières

1,02% des assurés auto ont des véhicules qui ont au plus 2 portières

- Graphique

histogramme des porte boites à moustaches por



- Resume numerique

```

minimum
[1] 2

maximum
[1] 4

mode
[1] 4

mediane
[1] 4

moyenne
[1] 3.979661

quantile
  0%  25%  50%  75% 100%
   2   4   4   4   4

coefficient_variation
[1] 5.042133

variance
[1] 0.04026436

ecart_type
[1] 0.2006598

coefficient_assymetrie
[1] -9.76441

interpretation_skewness
[1] "distribution etalee a gauche"

coefficient_applatissage
[1] 96.34369

interpretation_kurtosis
[1] "distribution leptokurtique"

```

d) Variable annual premium

- Tableau

	Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Frequence	Freq_Cum_crois
2630	54997	54997	374257	0.1469	0.1469
6466	1	54998	355533	0.0000	0.1470
7670	1	54999	355530	0.0000	0.1470
9816	1	55000	355529	0.0000	0.1470
10004	1	55001	355528	0.0000	0.1470
	Freq_Cum_decrois				

2630	1.00
6466	0.95
7670	0.95
9816	0.95
10004	0.95

54997 assurés auto ont au plus une prime de 2630

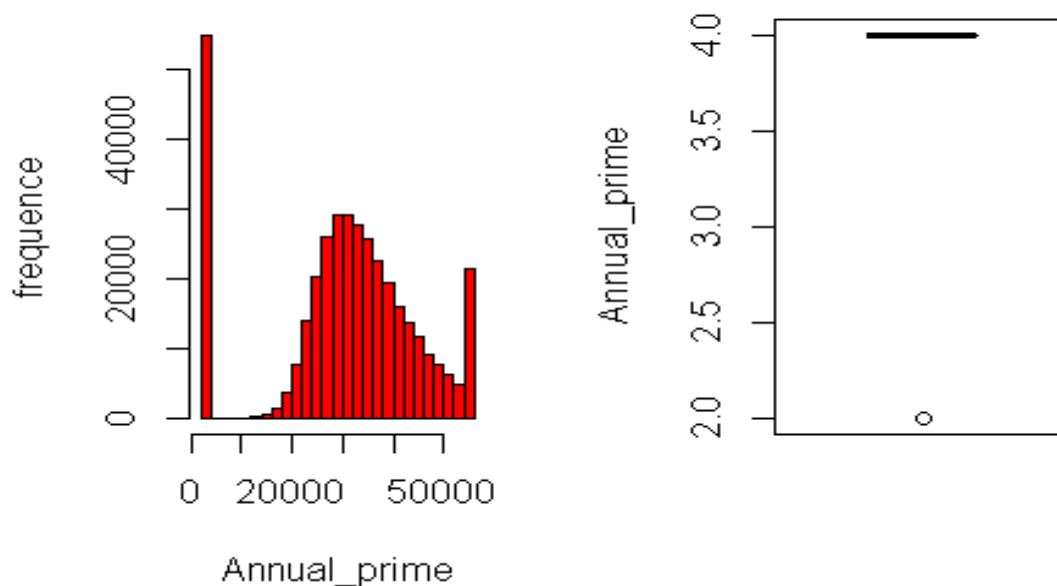
374257 assurés auto ont au moins une prime de 2630

14,69% des assurés auto ont au plus une prime de 2630

100% des assurés auto ont au moins une prime de 2630

- Graphique

histogramme des prime boîtes à moustaches prin



- Résumé numériques

```

minimum
[1] 2630

maximum
[1] 55409

mode
[1] 2630

mediane
[1] 31969

moyenne
[1] 30638.01

quantile
  0%   25%   50%   75%  100%
2630 25146 31969 39664 55409

```



```

coefficient_variation
[1] 47.38861

variance
[1] 210799251

ecart_type
[1] 14518.93

coefficient_assymetrie
[1] -0.5457333

interpretation_skewness
[1] "distribution etalee a gauche"

coefficient_aplatissement
[1] 2.829238

interpretation_kurtosis
[1] "distribution platikurtique"

```

II-2 Variables qualitatives

a) Variable Sexe

- Tableau

	Effectif	Frequence
femme	200419	0.5355117
homme	173838	0.4644883

53,55% des assurés auto sont des femmes

46,45% des assurés auto sont des hommes

- Graphique

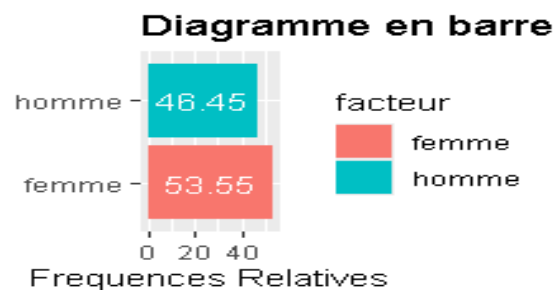
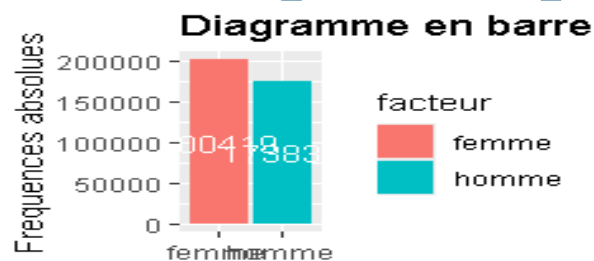


Diagramme en secteur

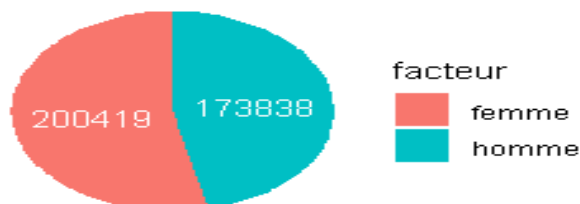


Diagramme en secteur

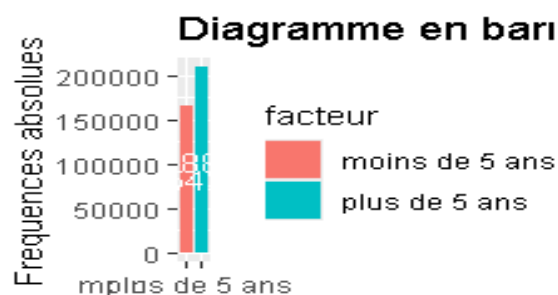


b) Variable vehicle Age

• Tableau

	Effectif	Frequence
moins de 5 ans	165428	0.4420171
plus de 5 ans	208829	0.5579829

• Graphique



44,20% des assurés auto ont des véhicules de moins de 5 ans

55,80% des assurés auto ont des véhicules de plus de 5 ans

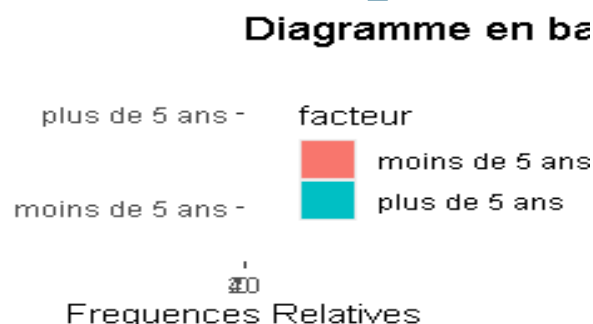


Diagramme en secteur



Diagramme en secteur



c) Variable Marque

• Tableau

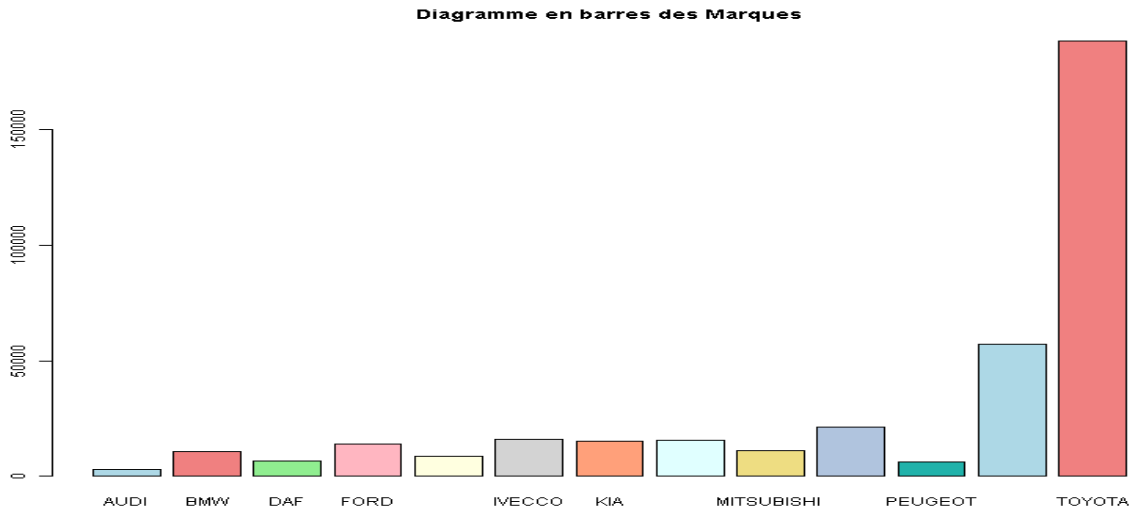
	Effectif	Frequence
AUDI	3055	0.00816284
BMW	10714	0.02862739
DAF	6737	0.01800100
FORD	13961	0.03730324
HYUNDAI	8575	0.02291206
IVECCO	16042	0.04286359
KIA	15165	0.04052028
MERCEDES	15710	0.04197650
MITSUBISHI	11278	0.03013437
NISSAN	21536	0.05754335
PEUGEOT	6128	0.01637378
SUZUKI	56981	0.15225099
TOYOTA	188375	0.50333060

15,23% des assurés auto ont des véhicules SUZUKI

50,33% des assurés auto ont des véhicules TOYOTA

4,20% des assurés auto ont des véhicules MERCEDES

- Graphique



d) Variable Couleur

- Tableau

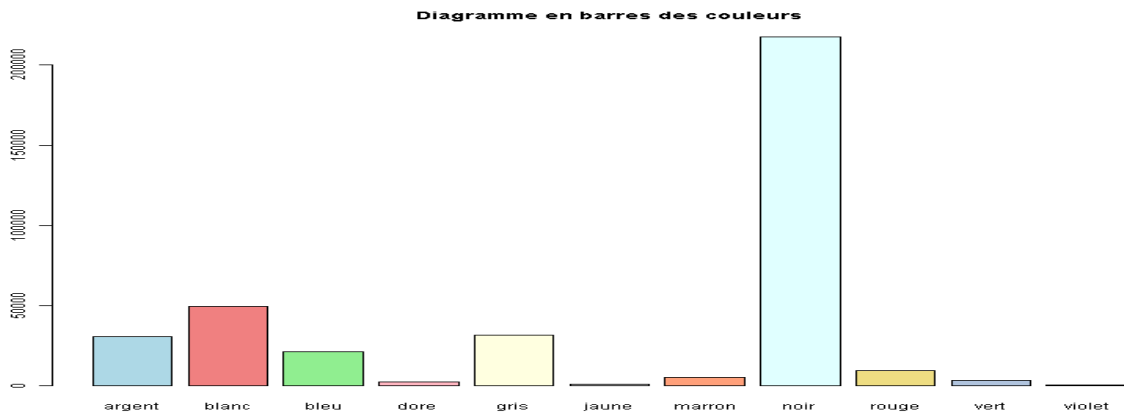
	Effectif	Frequence
argent	30592	0.081740622
blanc	49540	0.132368934
bleu	21551	0.057583425
dore	2673	0.007142151
gris	31517	0.084212186
jaune	961	0.002567754
marron	5532	0.014781287
noir	217728	0.581760662
rouge	9769	0.026102384
vert	3629	0.009696545
violet	765	0.002044050

58,18% des assurés auto ont des véhicules de couleur noir

13,24% des assurés auto ont des véhicules de couleur blanche

8,42% des assurés auto ont des véhicules de couleur grise

- Graphique



e) Variable fuel_type

• Tableau

	Effectif	Frequence
Gazoil	237835	0.6354858
Super	136422	0.3645142

• Graphique

63,55% des assurés auto ont des véhicules qui roulent au Gazoil

36,45 des assurés auto ont des véhicules qui roulent au Super

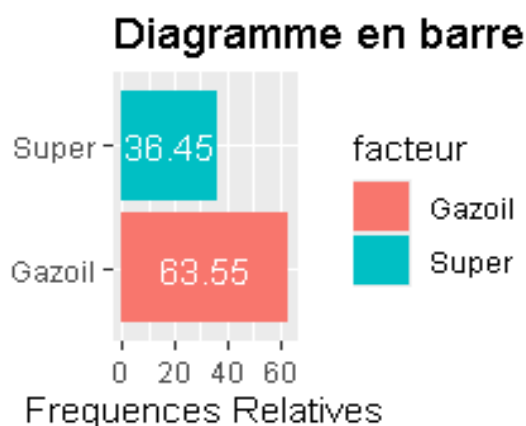
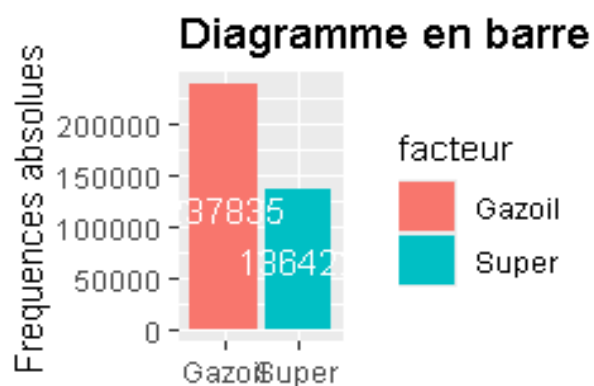


Diagramme en secteur

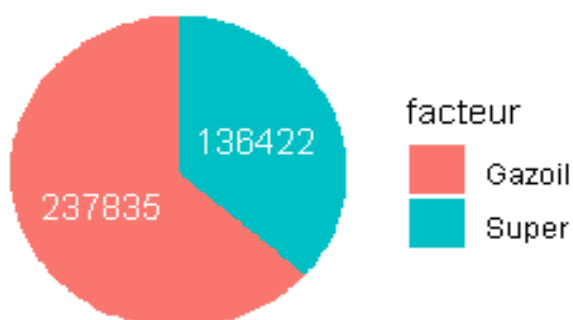
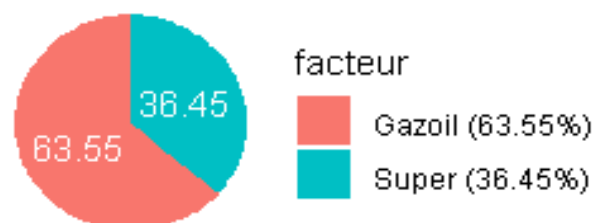


Diagramme en secteur



f) Variable Annee_fabric

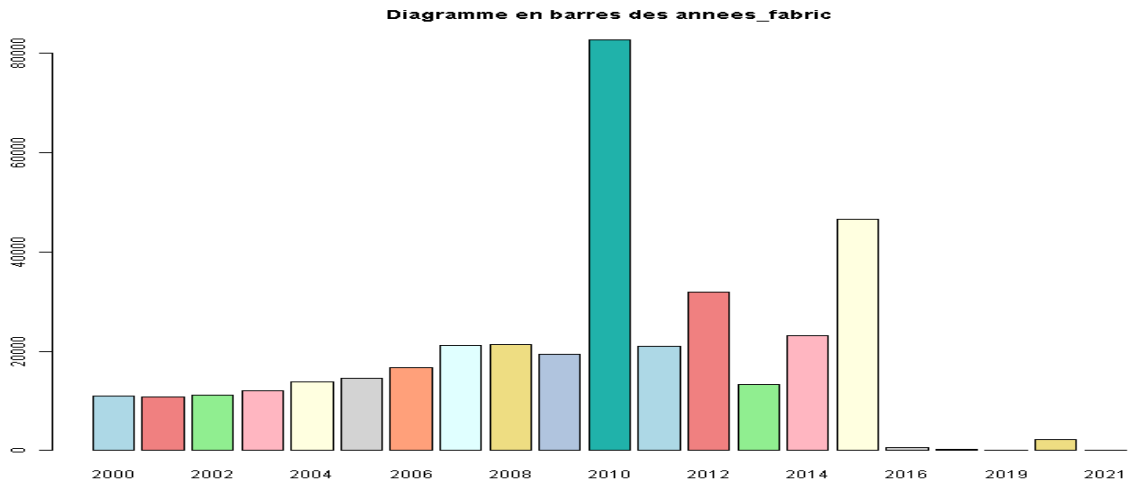
• Tableau

	Effectif	Frequence
2000	11001	0.02939424
2001	10742	0.02870220
2002	11220	0.02997940
2003	12079	0.03227461
2004	13857	0.03702536

2,87% des assurés auto ont des véhicules fabriqués en 2001

3,70% des assurés auto ont des véhicules fabriqués en 2004

- Graphique



g) Variable Transmission

- Tableau

	Effectif	Frequence
auto	143387	0.3831244
man	230870	0.6168756

38,31% des assurés auto ont des véhicules de Transmission auto

61,69% des assurés auto ont des véhicules de Transmission man

- Graphique

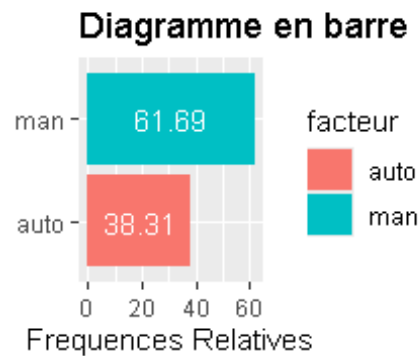
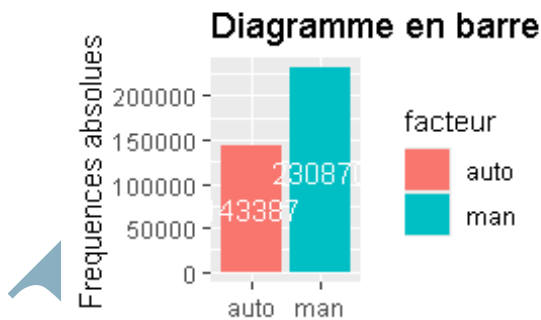


Diagramme en secteur

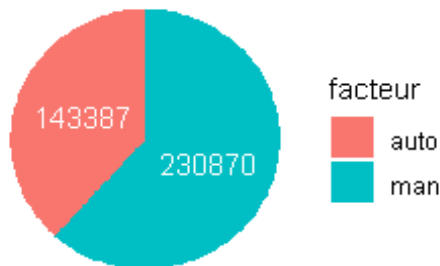
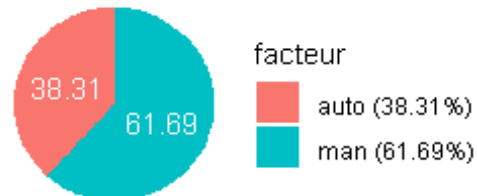


Diagramme en secteur



h) Variable Trajet

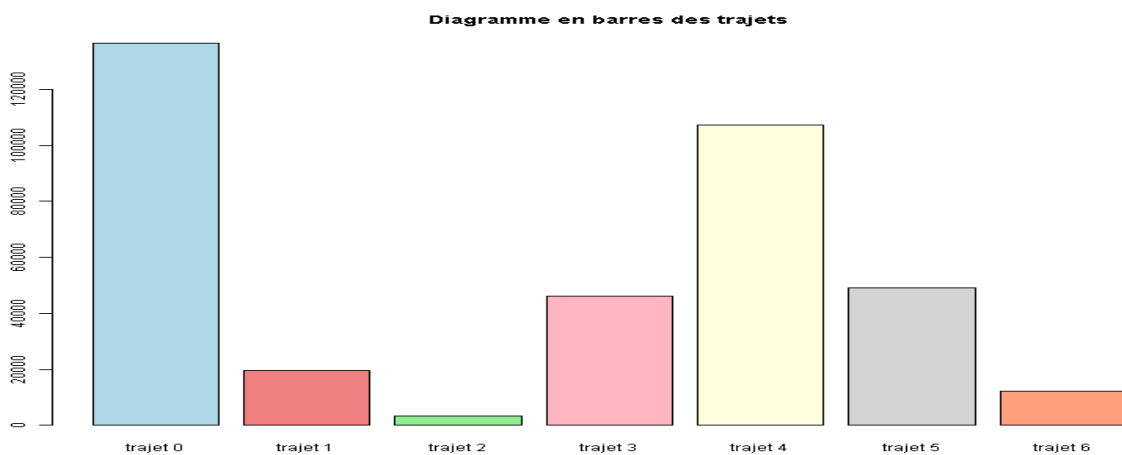
- Tableau

	Effectif	Frequence
trajet 0	136594	0.364973801
trajet 1	19667	0.052549451
trajet 2	3493	0.009333159
trajet 3	46049	0.123041119
trajet 4	107140	0.286273871
trajet 5	48993	0.130907371
trajet 6	12321	0.032921228

28,63% des assurés auto empruntent le trajet 4

12,30% des assurés auto empruntent le trajet 3

- Graphique



i) Variable conditions eclairage

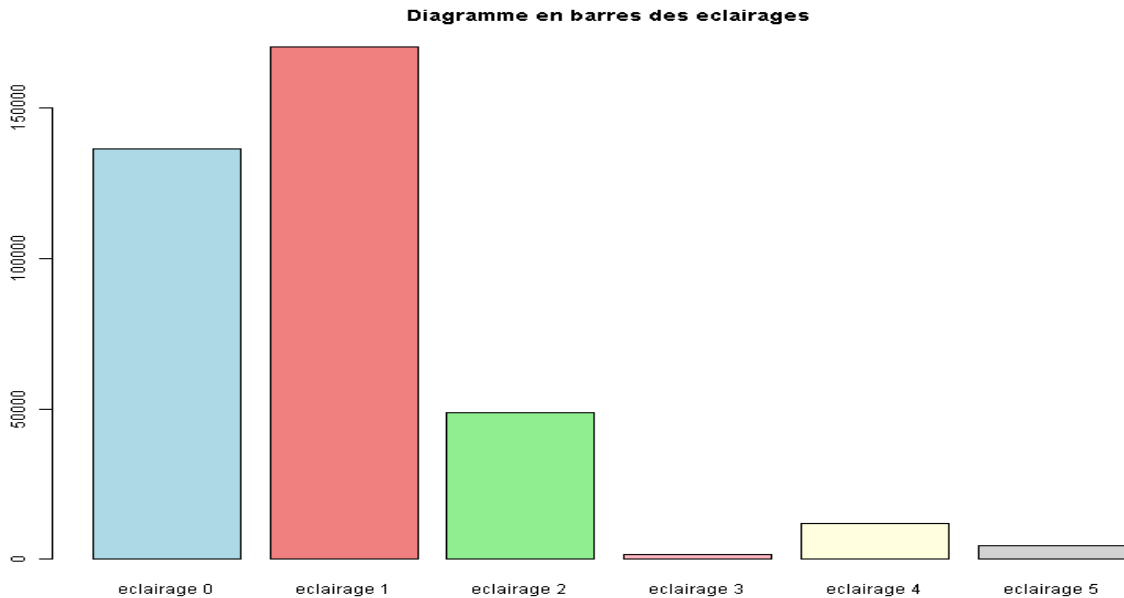
- Tableau

	Effectif	Frequence
eclairage 0	136594	0.364973801
eclairage 1	170539	0.455673508
eclairage 2	48968	0.130840572
eclairage 3	1628	0.004349952
eclairage 4	11913	0.031831068
eclairage 5	4615	0.012331099

13,08% des assurés auto ont eu des accidents dans des conditions d'eclairage 2

45,57% des assurés auto ont eu des accidents dans des conditions d'eclairage 1

- Graphique



j) Variable condition meteo

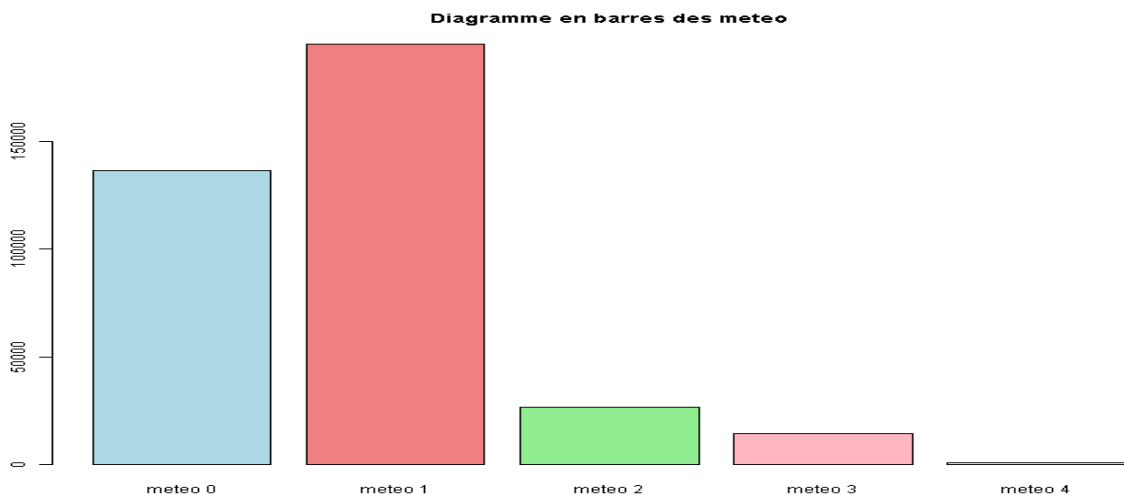
- Tableau

	Effectif	Frequence
meteo 0	136594	0.36497380
meteo 1	195257	0.52171903
meteo 2	26639	0.07117836
meteo 3	14650	0.03914422
meteo 4	1117	0.00298458

52,17% des assurés auto ont eu des accidents dans des conditions meteo 1

36,50% des assurés auto ont eu des accidents dans des conditions meteo 0

- Graphique



k) Variable condition surface route

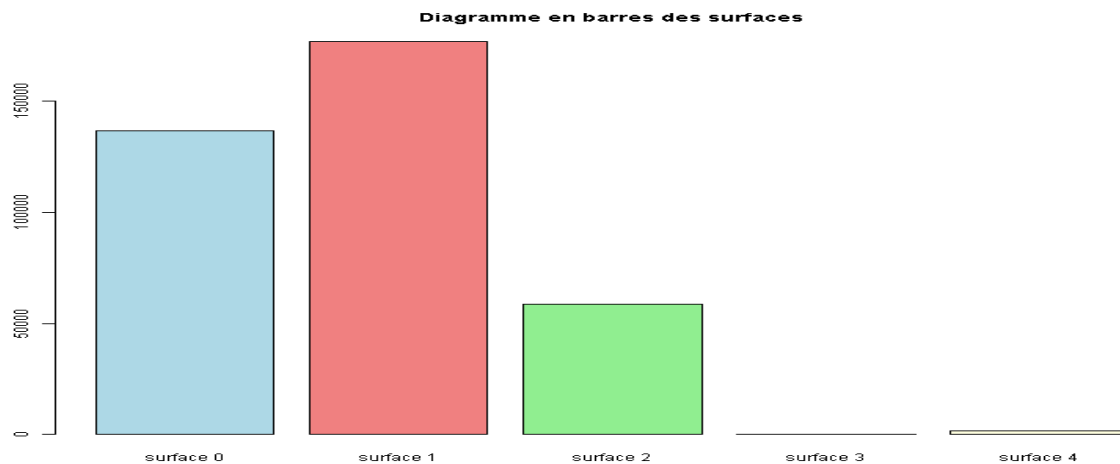
• Tableau

	Effectif	Frequence
surface 0	136594	0.3649738014
surface 1	176914	0.4727072573
surface 2	58817	0.1571567132
surface 3	241	0.0006439425
surface 4	1691	0.0045182856

47,27% des assurés auto ont eu des accidents dans des conditions de surface 1

36,50% des assurés auto ont eu des accidents dans des conditions de surface 0

• Graphique



l) Variable accident

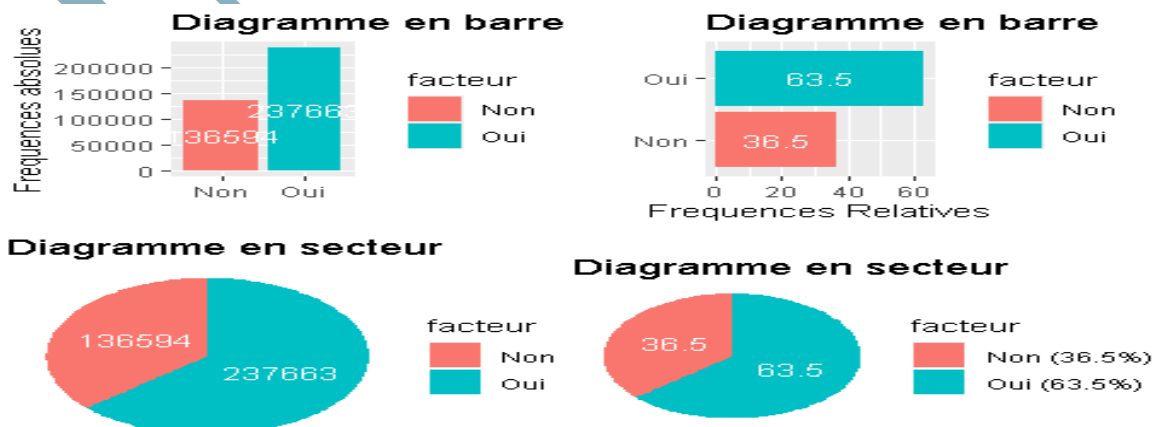
• Tableau

	Effectif	Frequence
Non	136594	0.3649738
Oui	237663	0.6350262

36,50% des assurés auto ont eu des accidents

63,50% des assurés auto ont eu des accidents

• Graphique



III- ANALYSE BIVARIEE

a) Variables accident et Age

- Tableau

	Non	Oui
20	0	5699
21	0	15670
22	0	21330
23	0	24262
24	0	26014
25	0	20923
26	45	13876
27	0	10969
28	0	9202
29	0	7605

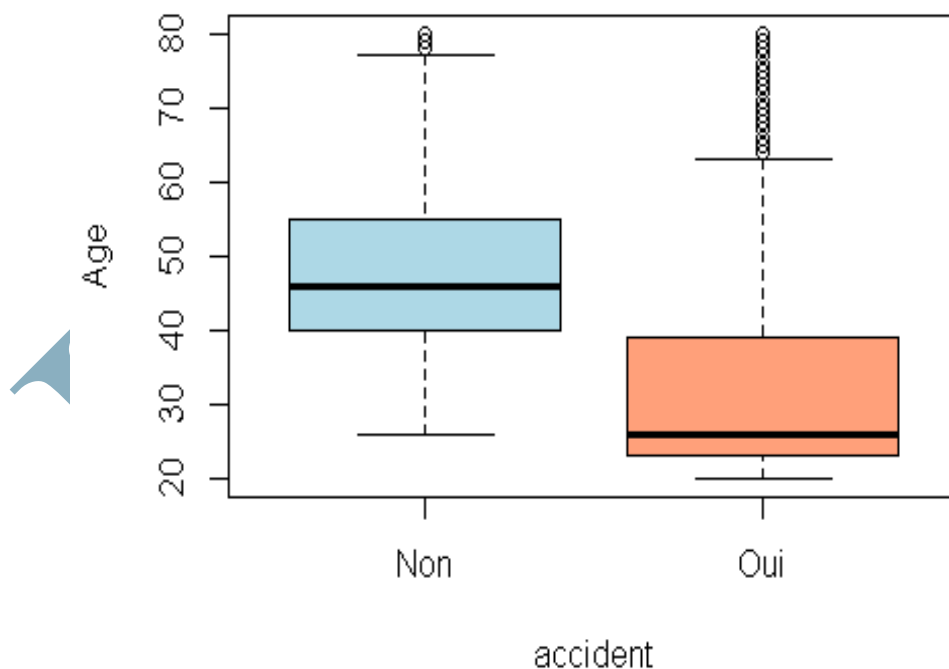
Observons les 10 premières lignes de ce tableau:

Les assurés auto dont l'âge varie entre 20 et 29 ans ont fait un accident

Parmi les assurés auto âgée de 26 ans, 45 n'ont pas fait d'accident tandis que 13876 en ont fait

- Graphique

boîtes à moustaches



On peut soupçonner une liaison entre ces deux variables à travers ces deux boîtes à moustaches

- Liaison

Rapport_Correlation

[1] 0.2406044

Resultat_Test_Anova

Analysis of Variance Table

Response: vecteur

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
facteur	1	20896101	20896101	118578	< 2.2e-16 ***
Residuals	374255	65952261	176		

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

Anova.P.value

[1] 0

Significativite_TestAnova

[1] "liaison significative, les deux variables sont liees"

Intensite_liaison

[1] "liaison faible"

Remarque

[1] "Si la liaison n'est pas significative, Ne pas tenir compte de son intensite"

b) Variables accident et Sexe

- Tableau

Tableau_Contingence

facteur2

facteur1 femme homme

Non 84356 52238

Oui 116063 121600

Tableau_Frequence

facteur2

facteur1 femme homme

Non 0.23 0.14

Oui 0.31 0.32

Tableau_Profil_Ligne

facteur2

facteur1 femme homme

Non 0.62 0.38

Oui 0.49 0.51

Tableau_Profil_Colonne

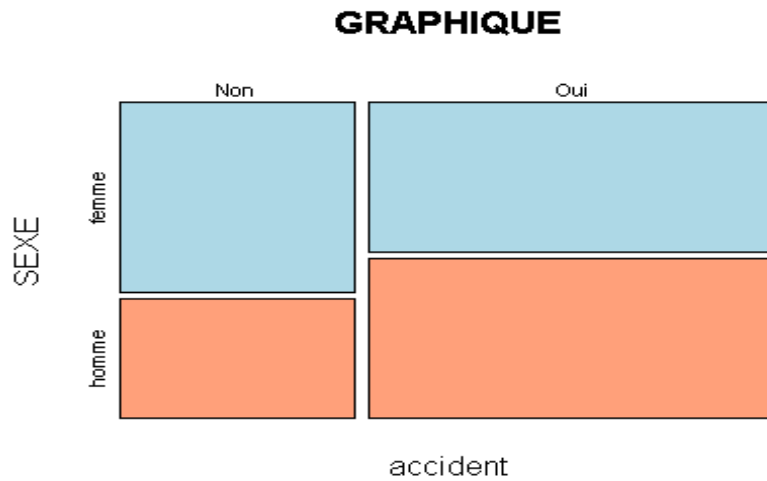
facteur2

facteur1 femme homme

Non 0.42 0.30

Oui 0.58 0.70

- Graphique



- Liaison

```
Effectif_Theorique
vecteur2
```

```
vecteur1    femme    homme
Non    73147.68  63446.32
Oui   127271.32 110391.68
```

```
Resultat_Test_KhiDeux
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: table(vecteur1, vecteur2)
X-squared = 5822, df = 1, p-value < 2.2e-16
```

```
Resultat_Test_Fisher
```

```
Fisher's Exact Test for Count Data
```

```
data: table(vecteur1, vecteur2)
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.669115 1.714979
sample estimates:
odds ratio
 1.691839
```

```

Khi_Deux
X-squared
5822.037

V_Cramer
[1] 0.1247303

Khi2.P.value
[1] 0

Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liees"

Fisher.P.value
[1] 0

Significativite_TestFisher
[1] "liaison significative, les deux variables sont liees"

Intensite_liaison
[1] "liaison faible"

Remarque
[1] "Si la liaison n'est pas significative, Ne pas tenir compte de
son intensité"

```

c) Variables accident et Vehicle_Age

- Tableau

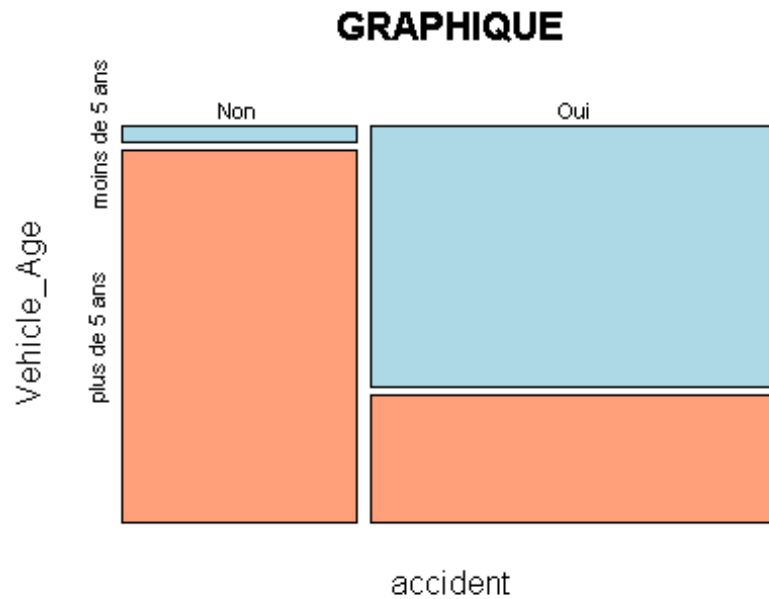
Tableau_Contingence			
	facteur2		
facteur1	moins de 5 ans	plus de 5 ans	
Non	5821	130773	
Oui	159607	78056	

Tableau_Frequence			
	facteur2		
facteur1	moins de 5 ans	plus de 5 ans	
Non	0.02	0.35	
Oui	0.43	0.21	

Tableau_Profil_Ligne			
	facteur2		
facteur1	moins de 5 ans	plus de 5 ans	
Non	0.04	0.96	
Oui	0.67	0.33	

Tableau_Profil_Colonne			
	facteur2		
facteur1	moins de 5 ans	plus de 5 ans	
Non	0.04	0.63	
Oui	0.96	0.37	

- Graphique



- Liaison

Effectif_Theorique

vecteur2

vecteur1 moins de 5 ans plus de 5 ans

Non 60376.89 76217.11

Oui 105051.11 132611.89

Resultat_Test_KhiDeux

Pearson's Chi-squared test with Yates' continuity correction

data: table(vecteur1, vecteur2)

X-squared = 139121, df = 1, p-value < 2.2e-16

Resultat_Test_Fisher

Fisher's Exact Test for Count Data

data: table(vecteur1, vecteur2)

p-value < 2.2e-16

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

0.02114381 0.02239297

facteur1	PEUGEOT	SUZUKI	TOYOTA
Non	0.01	0.06	0.18
Oui	0.01	0.10	0.32

Tableau_Profil_Ligne
facteur2

facteur1	AUDI	BMW	DAF	FORD	HYUNDAI	IVECCO	KIA	MERCEDES	MITSUBISHI	NISSAN
Non	0.01	0.03	0.02	0.04	0.02	0.04	0.04	0.04	0.03	0.06
Oui	0.01	0.03	0.02	0.04	0.02	0.04	0.04	0.04	0.03	0.06

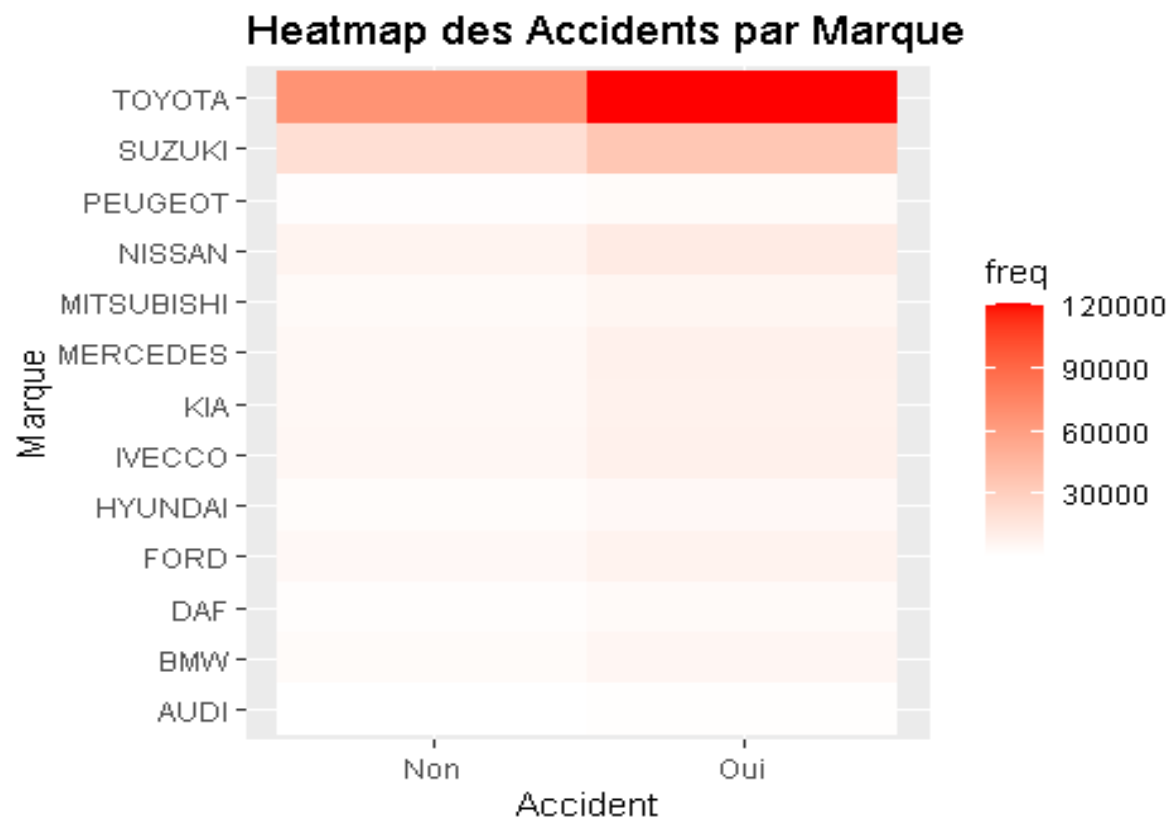
facteur1	PEUGEOT	SUZUKI	TOYOTA
Non	0.02	0.15	0.49
Oui	0.02	0.15	0.51

Tableau_Profil_Colonne
facteur2

facteur1	AUDI	BMW	DAF	FORD	HYUNDAI	IVECCO	KIA	MERCEDES	MITSUBISHI	NISSAN
Non	0.37	0.37	0.38	0.38	0.38	0.38	0.37	0.37	0.37	0.37
Oui	0.63	0.63	0.62	0.62	0.62	0.62	0.63	0.63	0.63	0.63

facteur1	PEUGEOT	SUZUKI	TOYOTA
Non	0.37	0.37	0.36
Oui	0.63	0.63	0.64

- Graphique



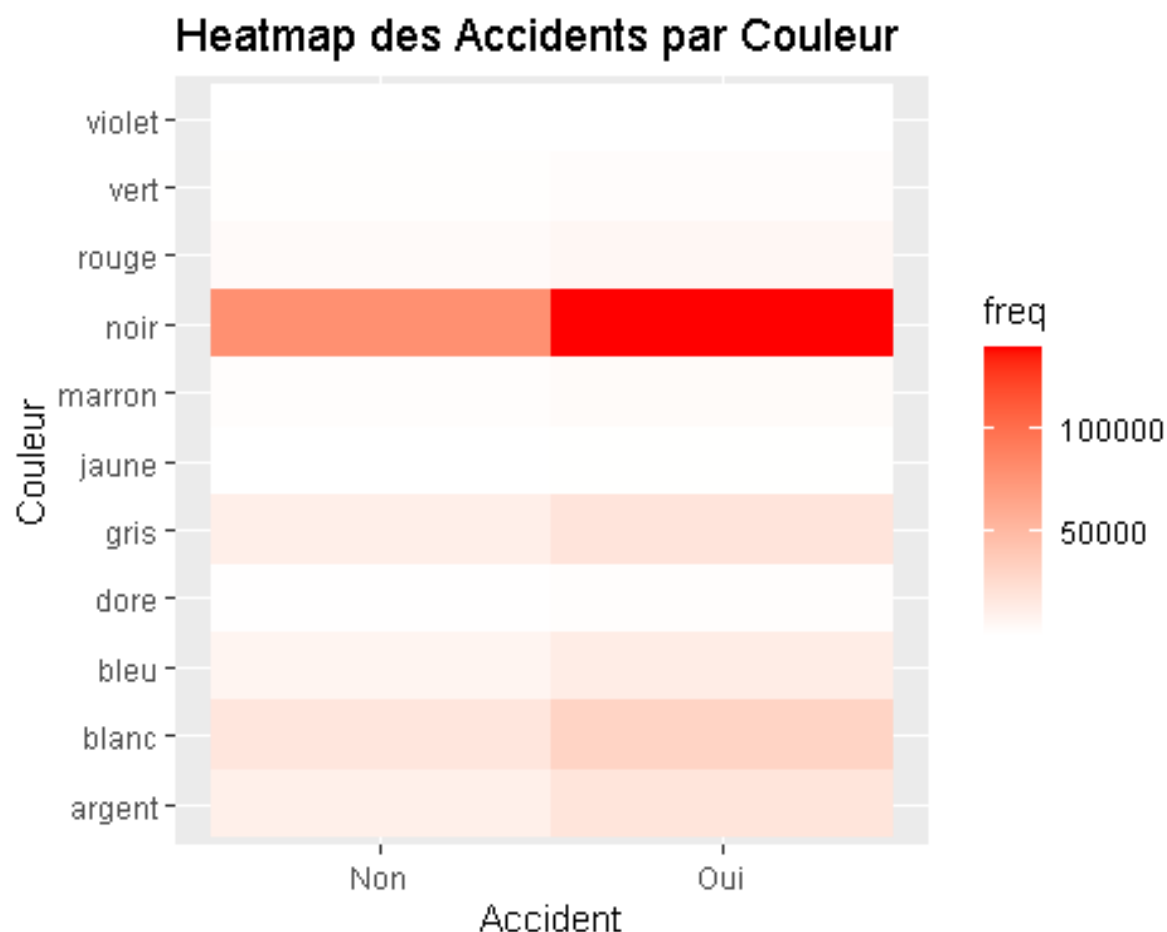
On verra la liaison entre ces deux variables plutard avec les tests de liaison

e) Variables accident et Couleur.

- Tableau

		argent	blanc	bleu	dore	gris	jaune	marron	noir	rouge
vert 1352	Non	11270	18218	7863	1002	11748	357	2075	78816	3597
	Oui	19322	31322	13688	1671	19769	604	3457	138912	6172
2277										
violet										
	Non	296								
	Oui	469								

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

f) Variables accident et fuel_type

- **Tableau**

Tableau_Contingence

		facteur2	
facteur1		Gazoil	Super
Non	86479	50115	
Oui	151356	86307	

Tableau_Frequence

		facteur2	
facteur1		Gazoil	Super
Non	0.23	0.13	
Oui	0.40	0.23	

Tableau_Profil_Ligne

		facteur2	
facteur1		Gazoil	Super
Non	0.63	0.37	
Oui	0.64	0.36	

Tableau_Profil_Colonne

		facteur2	
facteur1		Gazoil	Super
Non	0.36	0.37	
Oui	0.64	0.63	

- **Graphique**

Diagramme er

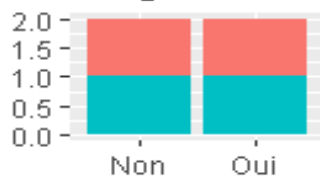
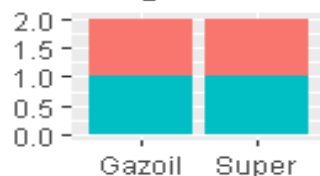
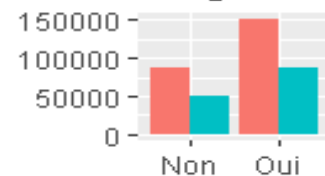


Diagramme er



Diagramme

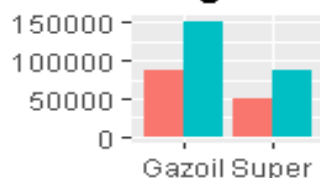


facteur2 ■ Gazoil ■

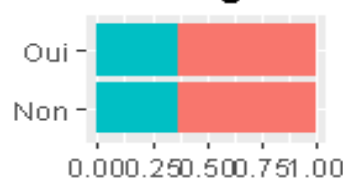
facteur1 ■ Non ■

facteur2 ■ Gazoil ■

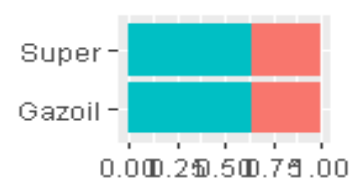
Diagramme



Profil ligne



Profil colonne



facteur1 ■ Non ■ variable

■ Gazoil ■

variable ■ Non ■

- **Liaison**

```
Effectif_Theorique
      vecteur2
```

```
 vecteur1   Gazoil   Super
      Non  86803.54 49790.46
      Oui 151031.46 86631.54
```

```
Resultat_Test_KhiDeux
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: table(vecteur1, vecteur2)
X-squared = 5.2259, df = 1, p-value = 0.02225
```

```
Resultat_Test_Fisher
```

Fisher's Exact Test for Count Data

```
data: table(vecteur1, vecteur2)
p-value = 0.02207
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.9704621 0.9976933
sample estimates:
odds ratio
 0.9839815
```

```
Khi_Deux
X-squared
 5.225942
```

```
V_Cramer
[1] 0.003742544
```

```
Khi2.P.value
[1] 0.0222524
```

```
Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liees"
```

```
Fisher.P.value
[1] 0.02206549
```

```
Significativite_TestFisher
[1] "liaison significative, les deux variables sont liees"
```

```
Intensite_liaison
[1] "liaison tr?s faible"
```

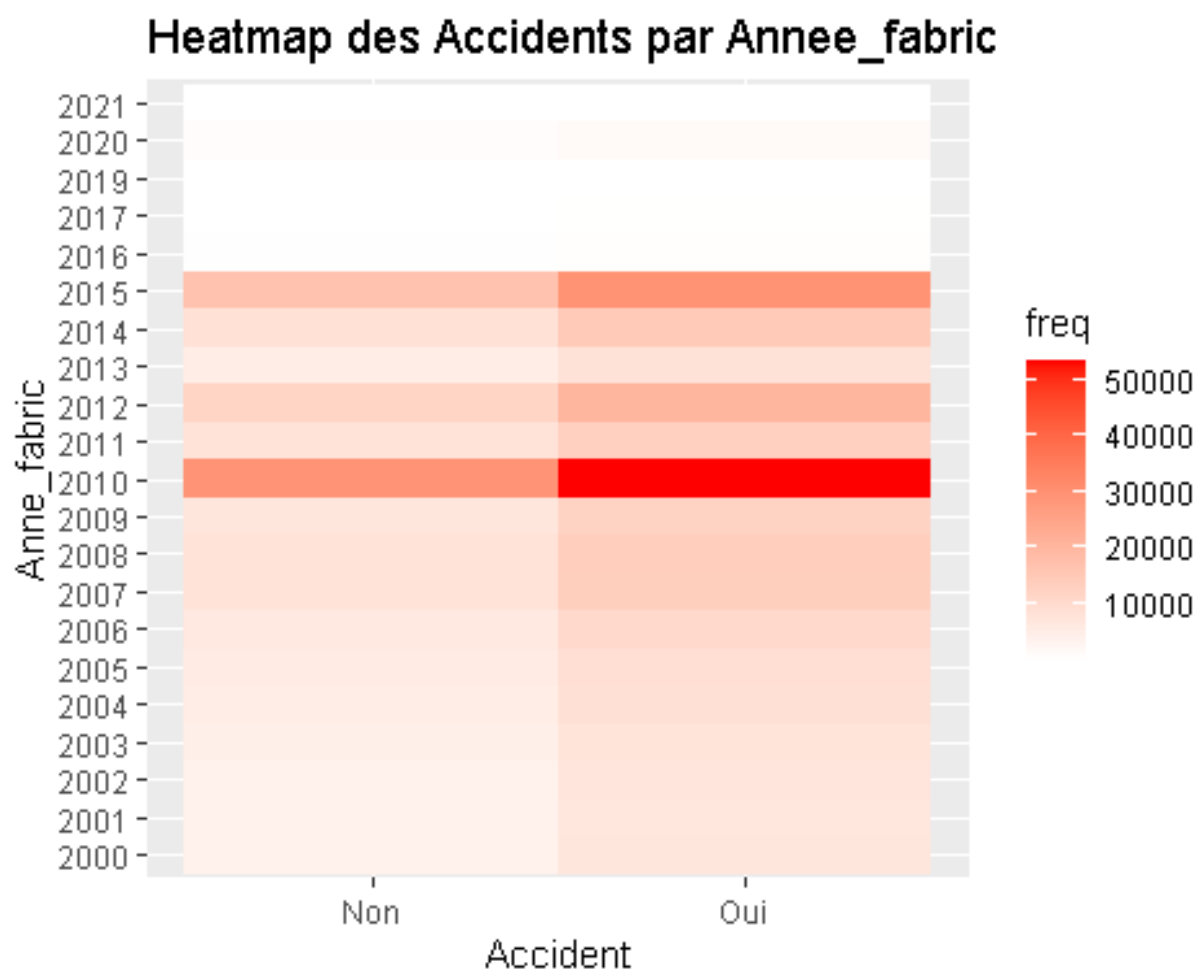
```
Remarque
[1] "Si la liaison n'est pas significative, Ne pas tenir compte de son intensité"
```

g) Variable accident et annee_fabric

- Tableau

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
2010	29435	3983	4008	4516	5063	5477	6285	7804	7829	7120
Non	4061	3983	4008	4516	5063	5477	6285	7804	7829	7120
Oui	6940	6759	7212	7563	8794	9174	10491	13456	13631	12291
2011	7766	6759	7212	7563	8794	9174	10491	13456	13631	12291
Non	11739	5031	8471	16911	201	59	34	799	2	
Oui	20224	8199	14740	29665	391	121	43	1393	3	
2012										
2013										
2014										
2015										
2016										
2017										
2019										
2020										
2021										

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

h) Variable accident et Transmission

- Tableau

Tableau_Contingence

facteur2		
facteur1	auto	man
Non	67241	69353
Oui	76146	161517

Tableau_Frequence

facteur2		
facteur1	auto	man
Non	0.18	0.19
Oui	0.20	0.43

Tableau_Profil_Ligne

facteur2		
facteur1	auto	man
Non	0.49	0.51
Oui	0.32	0.68

Tableau_Profil_Colonne

facteur2		
facteur1	auto	man
Non	0.47	0.30
Oui	0.53	0.70

- Graphique

Diagramme er

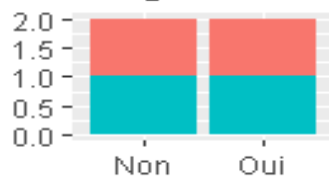
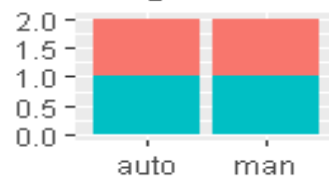
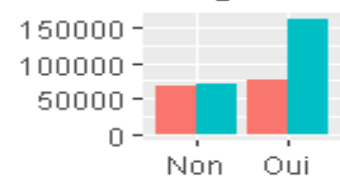


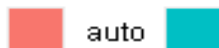
Diagramme er



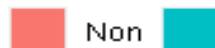
Diagramme



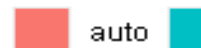
facteur2



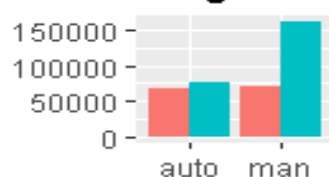
facteur1



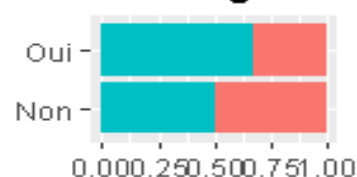
facteur2



Diagramme



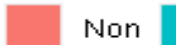
Profil ligne



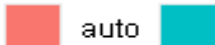
Profil colonne



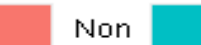
facteur1



variable



variable



- Liaison

```
Effectif_Theorique
      vecteur2
vecteur1  auto      man
      Non 52332.5  84261.5
      Oui 91054.5 146608.5

Resultat_Test_KhiDeux

Pearson's Chi-squared test with Yates' continuity correction

data:  table(vecteur1, vecteur2)
X-squared = 10841, df = 1, p-value < 2.2e-16

Resultat_Test_Fisher

Fisher's Exact Test for Count Data

data:  table(vecteur1, vecteur2)
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.028620 2.084709
sample estimates:
odds ratio
 2.056559

Khi_Deux
X-squared
10841.22

V_Cramer
[1] 0.1702036

Khi2.P.value
[1] 0

Significativite_TestKhi2
[1] "liaison significative, les deux variables sont liees"

Fisher.P.value
[1] 0

Significativite_TestFisher
[1] "liaison significative, les deux variables sont liees"

Intensite_liaison
[1] "liaison faible"

Remarque
[1] "Si la liaison n'est pas significative, Ne pas tenir compte de son
intensité"
```

i) Variables accident et Trajet

- Tableau

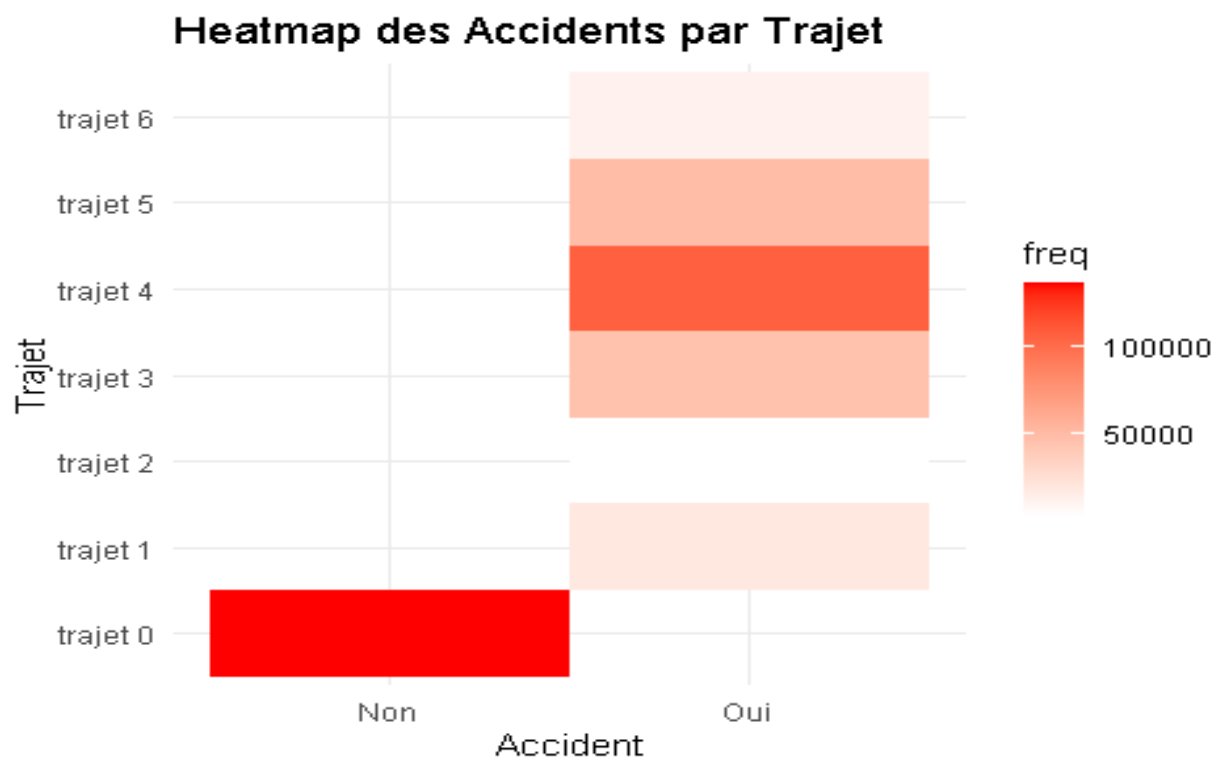
Tableau_Contingence								
facteur2								
facteur1	trajet 0	trajet 1	trajet 2	trajet 3	trajet 4	trajet 5	trajet 6	
Non	136594	0	0	0	0	0	0	
Oui	0	19667	3493	46049	107140	48993	12321	

Tableau_Frequence								
facteur2								
facteur1	trajet 0	trajet 1	trajet 2	trajet 3	trajet 4	trajet 5	trajet 6	
Non	0.36	0.00	0.00	0.00	0.00	0.00	0.00	
Oui	0.00	0.05	0.01	0.12	0.29	0.13	0.03	

Tableau_Profil_Ligne								
facteur2								
facteur1	trajet 0	trajet 1	trajet 2	trajet 3	trajet 4	trajet 5	trajet 6	
Non	1.00	0.00	0.00	0.00	0.00	0.00	0.00	
Oui	0.00	0.08	0.01	0.19	0.45	0.21	0.05	

Tableau_Profil_Colonne								
facteur2								
facteur1	trajet 0	trajet 1	trajet 2	trajet 3	trajet 4	trajet 5	trajet 6	
Non	1	0	0	0	0	0	0	
Oui	0	1	1	1	1	1	1	

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

j) Variable accident et Condition eclairage

- Tableau

Tableau_Contingence

facteur2		eclairage 0	eclairage 1	eclairage 2	eclairage 3	eclairage 4
facteur1						
Non		136594	0	0	0	0
Oui		0	170539	48968	1628	11913

facteur2		eclairage 5
facteur1		
Non		0
Oui		4615

Tableau_Frequence

facteur2		eclairage 0	eclairage 1	eclairage 2	eclairage 3	eclairage 4
facteur1						
Non		0.36	0.00	0.00	0.00	0.00
Oui		0.00	0.46	0.13	0.00	0.03

facteur2		eclairage 5
facteur1		
Non		0.00
Oui		0.01

Tableau_Profil_Ligne

facteur2		eclairage 0	eclairage 1	eclairage 2	eclairage 3	eclairage 4
facteur1						
Non		1.00	0.00	0.00	0.00	0.00
Oui		0.00	0.72	0.21	0.01	0.05

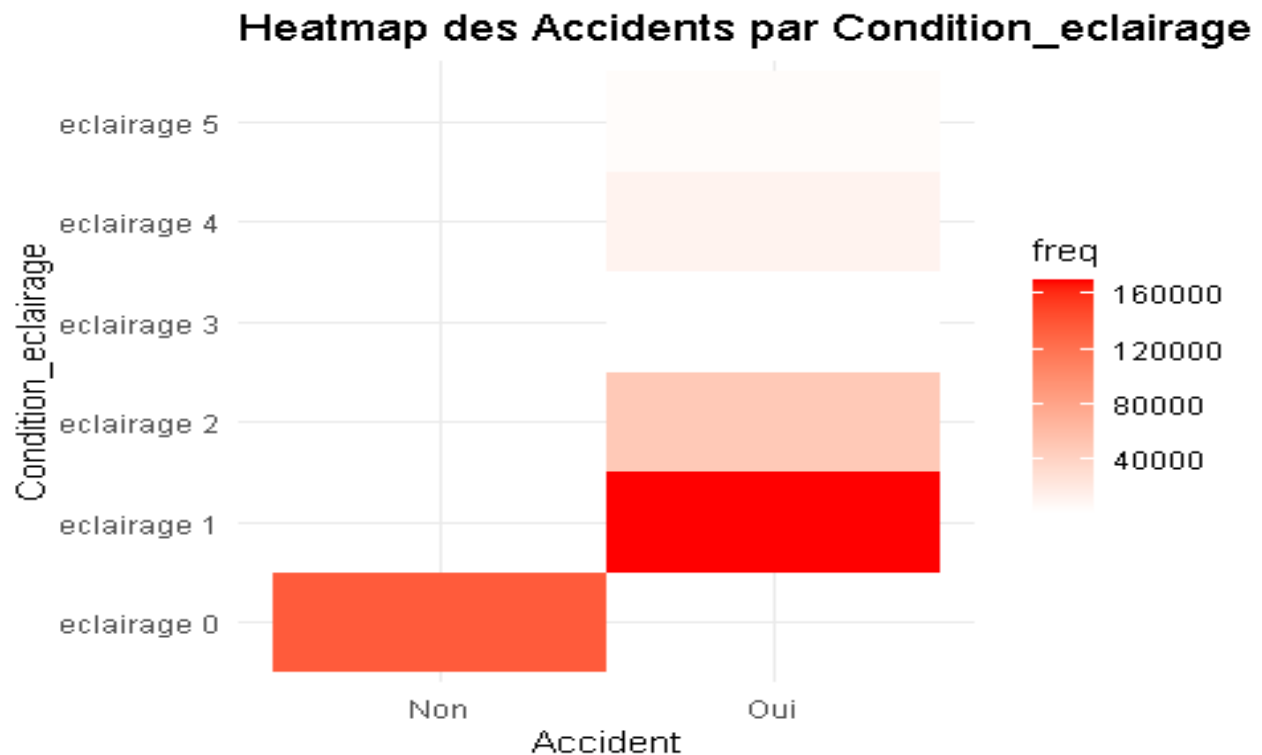
facteur2		eclairage 5
facteur1		
Non		0.00
Oui		0.02

Tableau_Profil_Colonne

facteur2		eclairage 0	eclairage 1	eclairage 2	eclairage 3	eclairage 4
facteur1						
Non		1	0	0	0	0
Oui		0	1	1	1	1

facteur2		eclairage 5
facteur1		
Non		0
Oui		1

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

k) Variable accident et condition_meteo

- Tableau

Tableau_Contingence

facteur2		meteo 0	meteo 1	meteo 2	meteo 3	meteo 4
facteur1	Non	136594	0	0	0	0
	Oui	0	195257	26639	14650	1117

Tableau_Frequence

facteur2		meteo 0	meteo 1	meteo 2	meteo 3	meteo 4
facteur1	Non	0.36	0.00	0.00	0.00	0.00
	Oui	0.00	0.52	0.07	0.04	0.00

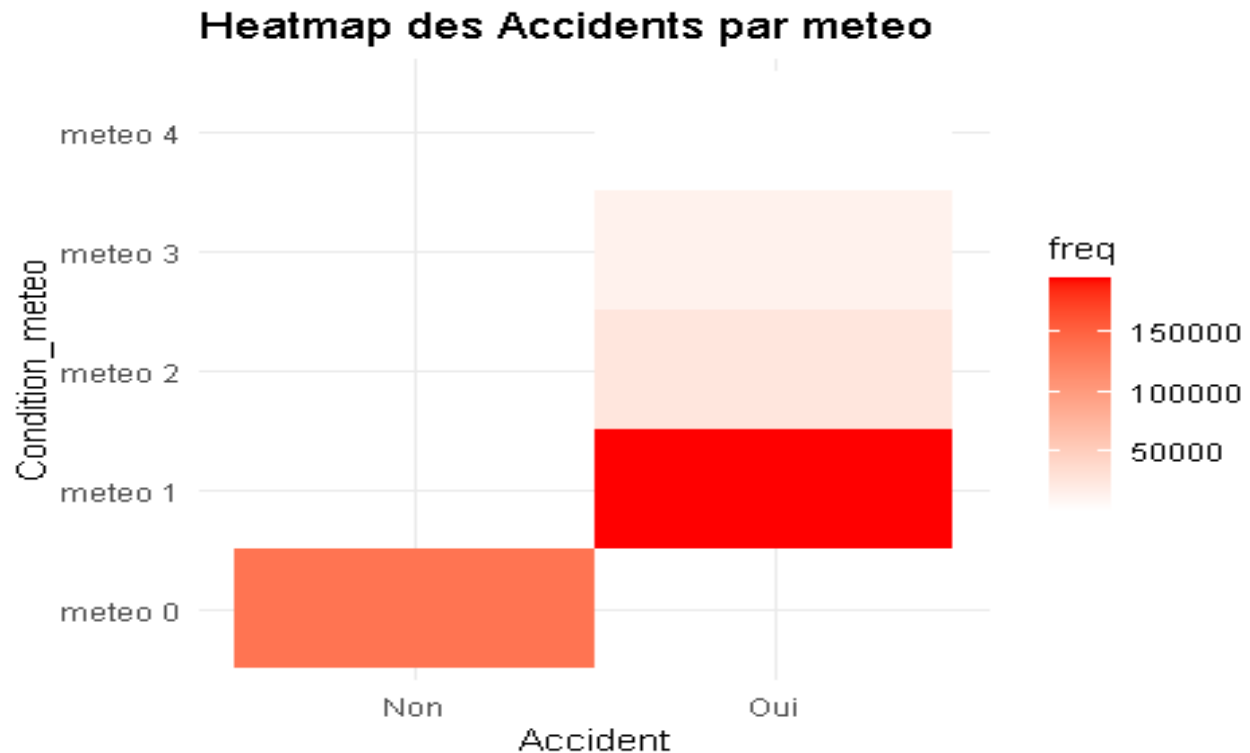
Tableau_Profil_Ligne

facteur2		meteo 0	meteo 1	meteo 2	meteo 3	meteo 4
facteur1	Non	1.00	0.00	0.00	0.00	0.00
	Oui	0.00	0.82	0.11	0.06	0.00

Tableau_Profil_Colonne

facteur2		meteo 0	meteo 1	meteo 2	meteo 3	meteo 4
facteur1	Non	1	0	0	0	0
	Oui	0	1	1	1	1

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

- 1) Variables accident et condition_surface_route

- Tableau

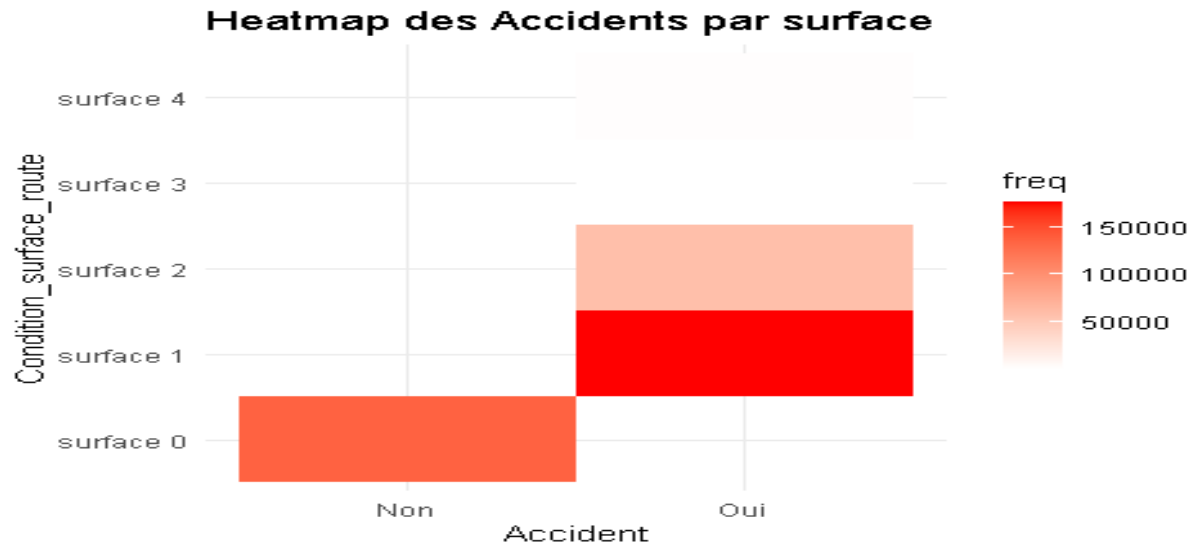
Tableau_Contingence						
	facteur2					
facteur1	surface 0	surface 1	surface 2	surface 3	surface 4	
Non	136594	0	0	0	0	
Oui	0	176914	58817	241	1691	

Tableau_Frequence						
	facteur2					
facteur1	surface 0	surface 1	surface 2	surface 3	surface 4	
Non	0.36	0.00	0.00	0.00	0.00	
Oui	0.00	0.47	0.16	0.00	0.00	

Tableau_Profil_Ligne						
	facteur2					
facteur1	surface 0	surface 1	surface 2	surface 3	surface 4	
Non	1.00	0.00	0.00	0.00	0.00	
Oui	0.00	0.74	0.25	0.00	0.01	

Tableau_Profil_Colonne						
	facteur2					
facteur1	surface 0	surface 1	surface 2	surface 3	surface 4	
Non	1	0	0	0	0	
Oui	0	1	1	1	1	

- Graphique



On verra la liaison entre ces deux variables plutard avec les tests de liaison

- m) Variable accident et many

- Tableau

	Non	Oui
0	136594	0
1	0	174271
2	0	12771
3	0	998
4	0	491

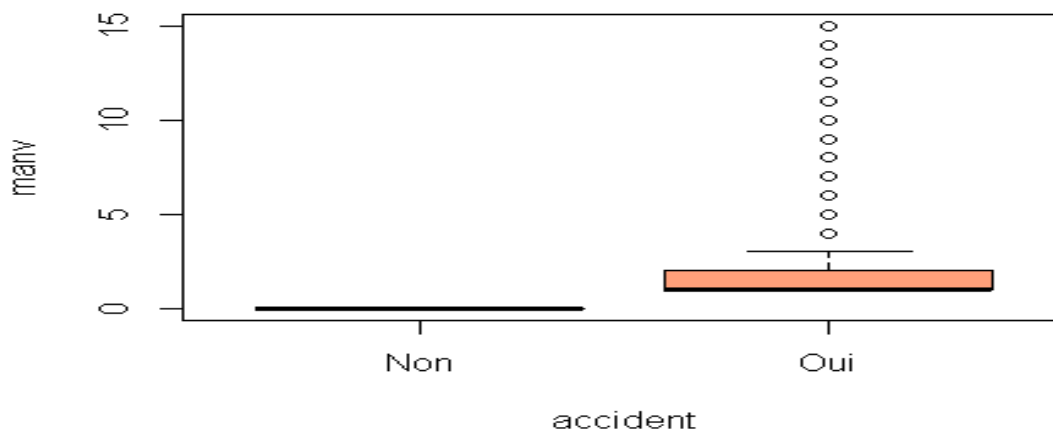
Observons les 5 premières lignes du tableau

Parmi ceux qui n'ont pas fait d'accident, 136594 n'ont fait aucune manœuvre

Parmi ceux qui ont fait au moins un accident, 12771 ont fait 3 manœuvres

- Graphique

boites à moustaches



- Liaison

Rapport_Correlation

[1] 0.1587972

Resultat_Test_Anova

Analysis of Variance Table

Response: vecteur

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
facteur	1	1001634	1001634	70650	< 2.2e-16 ***
Residuals	374255	5305992	14		

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

Anova.P.value

[1] 0

Significativite_TestAnova

[1] "liaison significative, les deux variables sont liees"

Intensite_liaison

[1] "liaison faible"

Remarque

[1] "Si la liaison n'est pas significative, Ne pas tenir compte de son intensite"

n) Variable accident et Annual premium

- Tableau

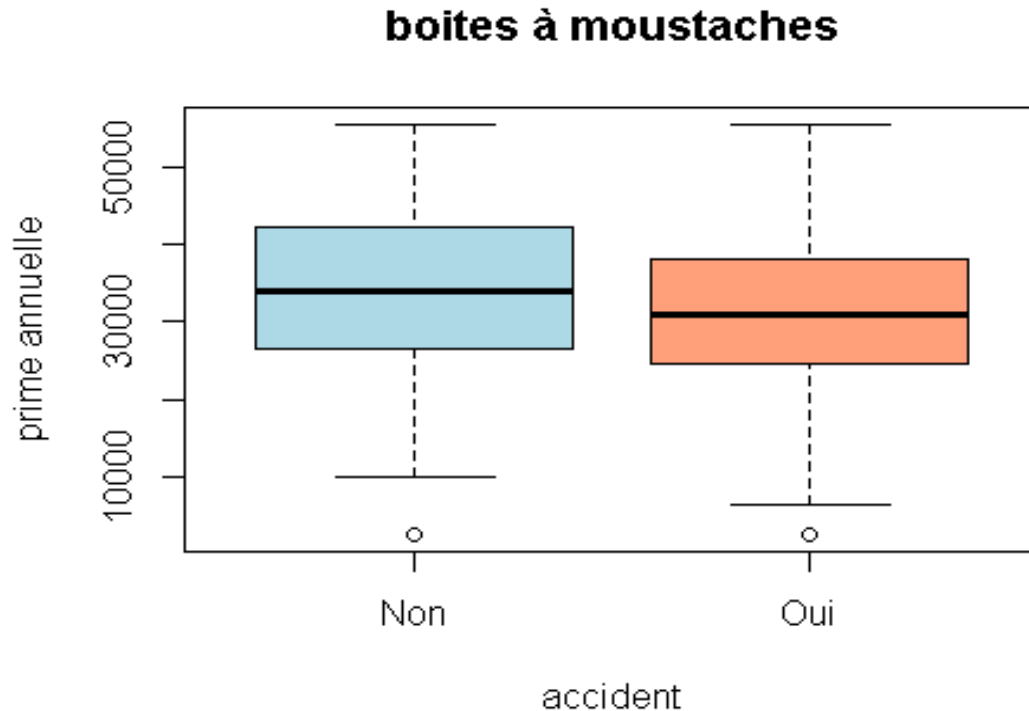
	Non	Oui
2630	21146	33851
6466	0	1
7670	0	1
9816	0	1
10004	0	1

Observons les 5 premières lignes du tableau

Parmi ceux qui n'ont pas fait d'accident, 21146 ont une prime annuelle de 2630

Parmi ceux qui ont fait au moins un accident, 33851 ont une prime annuelle de 2630

- Graphique



- Liaison

```
Rapport_Correlation
[1] 0.00592342

Resultat_Test_Anova
Analysis of Variance Table

Response: vecteur
      Df    Sum Sq   Mean Sq F value    Pr(>F)
facteur  1 4.6732e+11 4.6732e+11  2230.1 < 2.2e-16 ***
Residuals 374255 7.8426e+13 2.0955e+08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova.P.value
[1] 0

Significativite_TestAnova
[1] "liaison significative, les deux variables sont liees"

Intensite_liaison
[1] "liaison très faible"

Remarque
[1] "Si la liaison n'est pas significative, Ne pas tenir compte de son intensite"
```

IV- ESTIMATION DE LA PROPORTION DES ACCIDENTES

- Visualisation des proportions

```
Non  Oui  
0.36 0.64
```

- Conditions

```
Non      Oui  
136594 237663
```

$237663 > 10$ et $136594 > 10$ les conditions sont respectées ont peu ainsi faire l'estimation

- Estimation

```
[1] 0.63 0.64  
attr(,"conf.level")  
[1] 0.95
```

On a 95% de chance que la vraie proportion des assurés auto ayant déjà eu un accident soit comprise 63% et 64%. Soit un intervalle de confiance de : [63% ; 64%] a 95% de confiance.

V- TEST DE CONFORMITE

Testons si la proportion des assurés auto ayant eu un accident est égale à 63%

- Condition

$237663 > 10$ et $136594 > 10$ les conditions sont respectées ont peu ainsi faire l'estimation

- Test

$H_0 : P = 63\%$

$H_1 : P \neq 63\%$

Exact binomial test

```
data: 237663 and 374257
number of successes = 237663, number of trials = 374257, p-
value =
1.847e-10
alternative hypothesis: true probability of success is not
equal to 0.63
95 percent confidence interval:
0.6334814 0.6365689
sample estimates:
probability of success
0.6350262
```

Conclusion : la p-value < 0.05 donc on rejette H_0 , la proportion des assurés auto ayant eu un accident est significativement différente de 63%

$H_0 : P = 63\%$

$H_1 : P > 63\%$

Exact binomial test

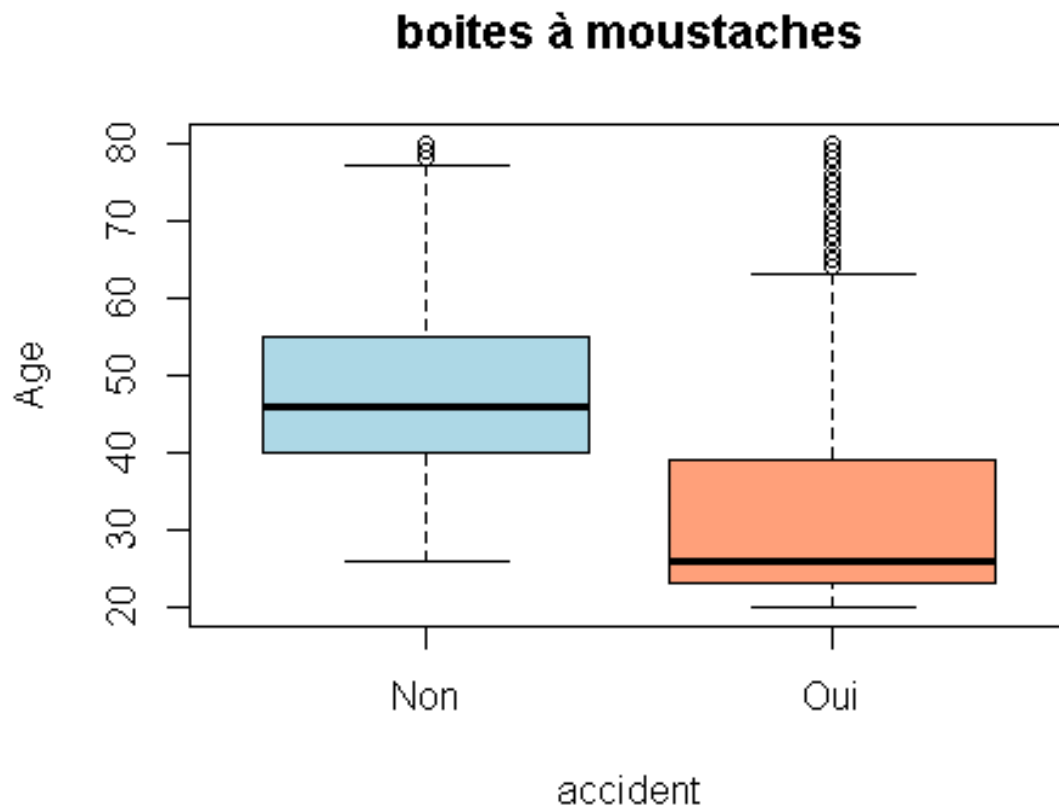
```
data: 237663 and 374257
number of successes = 237663, number of trials = 374257, p-
value =
9.273e-11
alternative hypothesis: true probability of success is greater
than 0.63
95 percent confidence interval:
0.6337297 1.0000000
sample estimates:
probability of success
0.6350262
```

Conclusion : la p-value < 0.05 donc on rejette H_0 , la proportion des assurés auto ayant eu un accident est significativement supérieur de 63%

VI- TEST DE LIAISON

a) Variable accident et Age

- Graphique



On peut soupçonner une liaison entre ces deux variables à travers ce graphique

- Normalité et Homoscédasticité

H0 : la variable “Age” suit une loi normale

H1 : la variable “Age” ne suit pas une loi normale

```
data: assurance_auto_makAge  
D = 0.14706, p-value < 2.2e-16  
alternative hypothesis: two-sided
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que la variable “Age” ne suit pas une loi normale.

H0 : Homoscédasticité

H1 : Hétéroscédasticité

```

Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   1  321.89 < 2.2e-16 ***
      374255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut qu'il y a Hétéroscédasticité.

- Test

HO : Les variables accident et Age ne sont pas liées

H1 : Les variables accident et Age sont liées

```

data:  assurance_auto_makAge by assurance_auto_makaccident
Kruskal-Wallis chi-squared = 110200, df = 1, p-value < 2.2e-16

```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Age sont liées.

b) Variable accident et Sexe

- Condition

	Non	Oui
femme	73147.68	127271.3
homme	63446.32	110391.7

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

HO : Les variables accident et Sexe ne sont pas liées

H1 : Les variables accident et Sexe sont liées

```

data:  tab
X-squared = 5822, df = 1, p-value < 2.2e-16

```


Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Sexe sont liées.

c) Variable accident et Vehicle_Age

- **Condition**

	Non	Oui
moins de 5 ans	60376.89	105051.1
plus de 5 ans	76217.11	132611.9

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- **Test**

HO : Les variables accident et Vehicle_Age ne sont pas liées

H1 : Les variables accident et Vehicle_Age sont liées

```
data: tab1
X-squared = 139121, df = 1, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Vehicle_Age sont liées.

d) Variable accident et Marque

- **Condition**

	Non	Oui
AUDI	1114.995	1940.005
BMW	3910.329	6803.671
DAF	2458.829	4278.171
FORD	5095.399	8865.601
HYUNDAI	3129.650	5445.350
IVECO	5854.910	10187.090
KIA	5534.828	9630.172
MERCEDES	5733.738	9976.262
MITSUBISHI	4116.175	7161.825

NISSAN	7860.076	13675.924
PEUGEOT	2236.559	3891.441
SUZUKI	20796.572	36184.428
TOYOTA	68751.940	119623.060

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

H0 : Les variables accident et Marque ne sont pas liées

H1 : Les variables accident et Marque sont liées

```
data: tab2
X-squared = 88.851, df = 12, p-value = 8.238e-14
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Marque sont liées.

e) Variable accident et Couleur

- Condition

	Non	Oui
argent	11165.2785	19426.7215
blanc	18080.8021	31459.1979
bleu	7865.5504	13685.4496
dore	975.5750	1697.4250
gris	11502.8793	20014.1207
jaune	350.7398	610.2602
marron	2019.0351	3512.9649
noir	79465.0158	138262.9842
rouge	3565.4291	6203.5709
vert	1324.4899	2304.5101
violet	279.2050	485.7950

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

H0 : Les variables accident et Couleur ne sont pas liées

H1 : Les variables accident et Couleur sont liées

Pearson's Chi-squared test

data: tab3

X-squared = 26.437, df = 10, p-value = 0.003195

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Couleur sont liées.

f) Variables accident et fuel_type

• Condition

	Non	Oui
Gazoil	86803.54	151031.46
Super	49790.46	86631.54

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

• Test

H0 : Les variables accident et fuel_type ne sont pas liées

H1 : Les variables accident et fuel_type sont liées

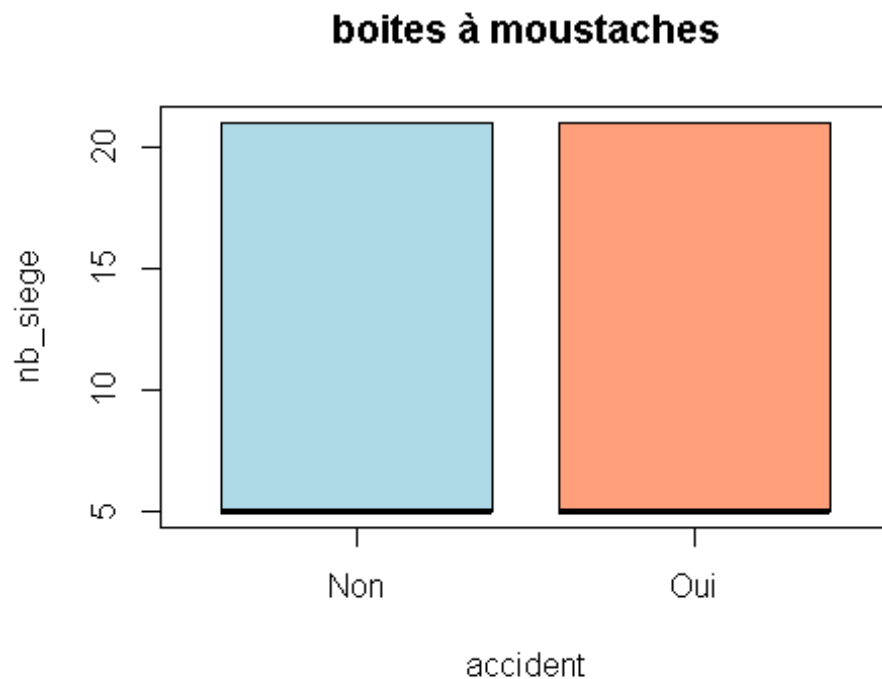
data: tab4

X-squared = 5.2259, df = 1, p-value = 0.02225

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et fuel_type sont liées.

g) Variable accident et nb_siege

• Graphique



On peut soupçonner qu'il n'y a pas de liaison entre ces deux variables à travers ce graphique.

- Normalité et Homoscédasticité

HO : la variable “nb_siege” suit une loi normale

H1 : la variable “nb_siege” ne suit pas une loi normale

```
data: assurance_auto_maknb_siege
D = 0.38906, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que la variable “nb_siege” ne suit pas une loi normale.

HO: Homoscédasticité

H1: Hétéroscédasticité

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  1  4446.8 < 2.2e-16 ***
 374255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion : la $p\text{-value} < 0.05$ donc on rejette H_0 , on conclut qu'il y a Hétéroscédasticité.

- Test

H_0 : Les variables accident et nb_siege ne sont pas liées

H_1 : Les variables accident et nb_siege sont liées

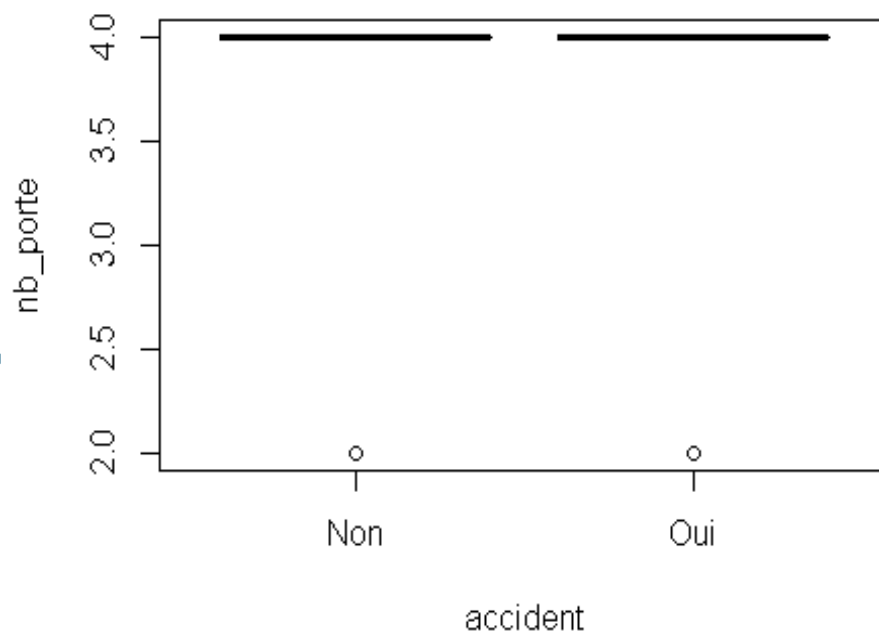
```
data: assurance_auto_maknb_siege by assurance_auto_makaccident
Kruskal-Wallis chi-squared = 4555.3, df = 1, p-value < 2.2e-16
```

Conclusion : la $p\text{-value} < 0.05$ donc on rejette H_0 , on conclut que Les variables accident et nb_siege sont liées.

h) Variables accident et nb_porte

- Graphique

boites à moustaches



On peut soupçonner qu'il n'y a pas de liaison entre ces deux variables a travers ce graphique.

- Normalité et Homoscédasticité

HO : la variable “nb_porte” suit une loi normale

H1 : la variable “nb_porte” ne suit pas une loi normale

```
data : assurance_auto_maknb_porte
D = 0.5302, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que la variable “nb_porte” ne suit pas une loi normale.

HO: Homoscédasticité

H1: Hétéroscédasticité

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   1  40.954 1.56e-10 ***
      374255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut qu'il y a Hétéroscédasticité.

- Test

HO : Les variables accident et nb_porte ne sont pas liées

H1 : Les variables accident et nb_porte sont liées

```
data: assurance_auto_maknb_porte by assurance_auto_makaccident
Kruskal-Wallis chi-squared = 40.95, df = 1, p-value = 1.562e-10
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et nb_siege sont liées.

i) Variables accident et Annee fabric

- Condition

	Non	Oui
2000	4015.076789	6985.923211

2001	3920.548575	6821.451425
2002	4095.006052	7124.993948
2003	4408.518547	7670.481453
2004	5057.441966	8799.558034
2005	5347.231165	9303.768835
2006	6122.800493	10653.199507
2007	7759.343018	13500.656982
2008	7832.337779	13627.662221
2009	7084.506459	12326.493541
2010	30231.144946	52599.855054
2011	7643.646323	13299.353677
2012	11665.657615	20297.342385
2013	4828.603393	8401.396607
2014	8471.406905	14739.593095
2015	16999.019775	29576.980225
2016	216.064490	375.935510
2017	65.695284	114.304716
2019	28.102983	48.897017
2020	800.022573	1391.977427
2021	1.824869	3.175131

Le pourcentage d'effets théoriques supérieur à 5 est :

[1] le pourcentage effets théoriques supérieur à 5 est : [1] 95.2381

- **Test**

HO : Les variables accident et Annee_fabric ne sont pas liées

H1 : Les variables accident et Annee_fabric sont liées

```
data: tab5
X-squared = 77.464, df = 20, p-value = 1.053e-08
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Annee_fabric sont liées.

j) Variables accident et Transmission

- **Condition**

	Non	Oui
auto	52332.5	91054.5
man	84261.5	146608.5

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

HO : Les variables accident et Transmission ne sont pas liées

H1 : Les variables accident et Transmission sont liées

```
data: tab6
X-squared = 10841, df = 1, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Transmission sont liées.

k) Variable accident et Gravité accident

- Condition

	Non	Oui
auto	52332.5	91054.5
man	84261.5	146608.5

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

HO : Les variables accident et Gravite_accident ne sont pas liées

H1 : Les variables accident et Gravite_accident sont liées

```
data: tab11
X-squared = 374257, df = 3, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Gravite_accident sont liées.

l) Variable accident et Trajet

- Condition

	Non	Oui
trajet 0	49853.231	86740.769
trajet 1	7177.940	12489.060
trajet 2	1274.853	2218.147
trajet 3	16806.679	29242.321
trajet 4	39103.293	68036.707
trajet 5	17881.161	31111.839
trajet 6	4496.842	7824.158

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

H0 : Les variables accident et Trajet ne sont pas liées

H1 : Les variables accident et Trajet sont liées

```
data: tab6
X-squared = 10841, df = 1, p-value < 2.2e-16
```

Conclusion : la $p\text{-value} < 0.05$ donc on rejette **H0**, on conclut que Les variables accident et Trajet sont liées.

m) Variables accident et Condition éclairage

- Condition

	Non	Oui
eclairage 0	49853.2314	86740.769
eclairage 1	62242.2671	108296.733
eclairage 2	17872.0371	31095.963
eclairage 3	594.1773	1033.823
eclairage 4	4347.9329	7565.067
eclairage 5	1684.3541	2930.646

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

H0 : Les variables accident et Condition eclaireage ne sont pas liées

H1 : Les variables accident et Condition eclaireage sont liées

```
data: tab7
X-squared = 374257, df = 6, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Condition eclaireage sont liées.

n) Variable accident et Condition meteo

- Condition**

	Non	Oui
meteo 0	49853.2314	86740.7686
meteo 1	71263.6895	123993.3105
meteo 2	9722.5371	16916.4629
meteo 3	5346.8662	9303.1338
meteo 4	407.6757	709.3243

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test**

H0 : Les variables accident et Condition meteo ne sont pas liées

H1 : Les variables accident et Condition meteo sont liées

```
data: tab9
X-squared = 374257, df = 4, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Condition meteo sont liées.

o) Variables accident et Condition surface route

- Condition**

	Non	Oui
surface 0	49853.23143	86740.7686

surface 1	64568.97511	112345.0249
surface 2	21466.66408	37350.3359
surface 3	87.95869	153.0413
surface 4	617.17070	1073.8293

Pourcentage des effectifs théoriques ≥ 5 : 100.00%

La règle de Cochran est respectée.

La règle de Cochran étant respectée nous procéderons au test de khi-deux.

- Test

H0 : Les variables accident et Condition surface route ne sont pas liées

H1 : Les variables accident et Condition surface route sont liées

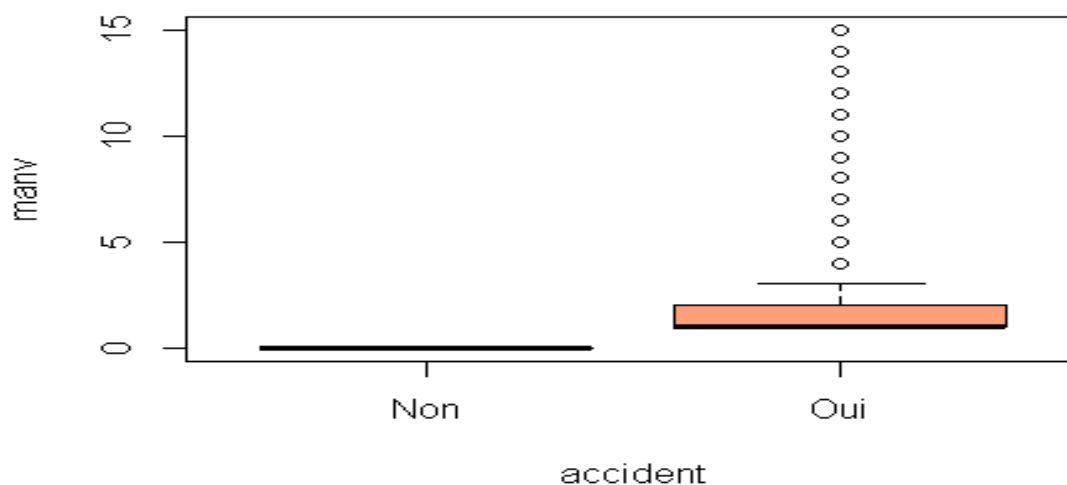
```
data: tab10
X-squared = 374257, df = 4, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Condition surface route sont liées.

p) Variables accident et manv

- Graphique

boîtes à moustaches



On peut soupçonner qu'il y a liaison entre ces deux variables à travers ce graphique.

- Normalité et Homoscédasticité

HO : la variable “manv” suit une loi normale

H1 : la variable “manv” ne suit pas une loi normale

```
data: assurance_auto_makmanv
D = 0.44167, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que la variable “manv” ne suit pas une loi normale.

HO: Homoscédasticité

H1: Hétéroscédasticité

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   1  35187 < 2.2e-16 ***
      374255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut qu’il y a Hétéroscédasticité.

- Test

HO : Les variables accident et manv ne sont pas liées

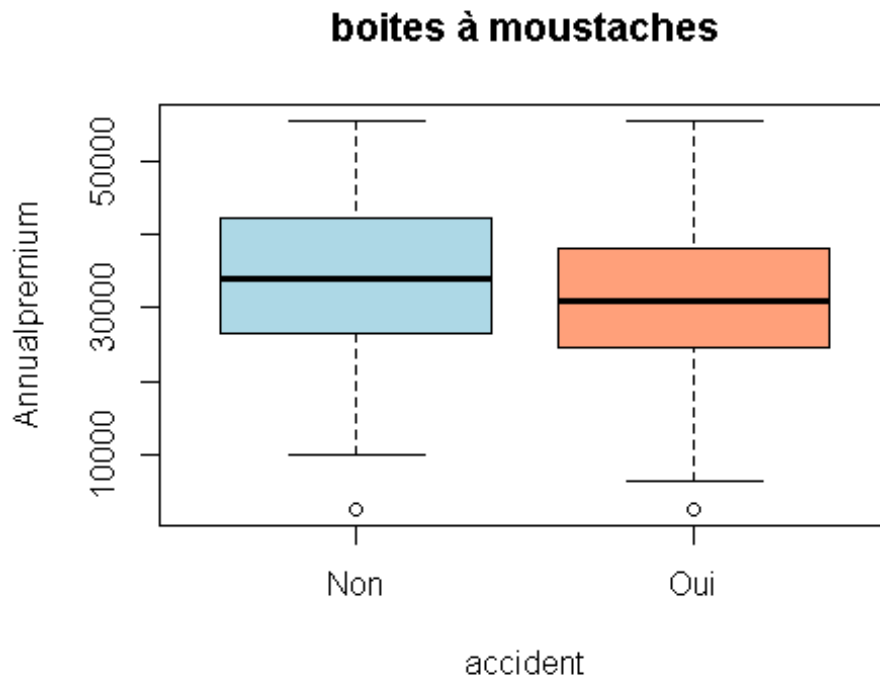
H1 : Les variables accident et manv sont liées

```
data: assurance_auto_makmanv by assurance_auto_makaccident
Kruskal-Wallis chi-squared = 306066, df = 1, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et manv sont liées.

q) Variable accident et Annual premium

- Graphique



On peut soupçonner qu'il y a liaison entre ces deux variables a travers ce graphique.

- Normalité et Homoscédasticité

HO : la variable "Annual premium" suit une loi normale

H1 : la variable "Annual premium" ne suit pas une loi normale

```
data: assurance_auto_makAnnual_Premium
D = 0.12009, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Conclusion : la $p\text{-value} < 0.05$ donc on rejette H_0 , on conclut que la variable "Annual_Premium" ne suit pas une loi normale.

HO: Homoscédasticité

H1: Hétéroscédasticité

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  1    1356 < 2.2e-16 ***
 374255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut qu'il y a Hétéroscédasticité.

- Test

HO : Les variables accident et Annual_Premium ne sont pas liées

H1 : Les variables accident et Annual_Premium sont liées

```
data: assurance_auto_makAnnual_Premium by assurance_auto_makaccident
Kruskal-Wallis chi-squared = 4666.7, df = 1, p-value < 2.2e-16
```

Conclusion : la p-value < 0.05 donc on rejette H0, on conclut que Les variables accident et Annual_Premium sont liées.

V- REGRESSION LOGISTIQUE

+Specification du modele

accident = B0 + B1Age + B2Sexe + B3Vehivle_Age + B4Marque + B5Couleur + B6fuel_type + B7nb_siege + B8nb_porte + B9Annee_fabric + B10Transmission + B11Gravite_accident + B12Trajet + B13 Condition_eclairage + B14Condition_meteo + B15Condition_surface_route + B16manv + B17Annual_premium + ε

+Anova

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			374256	491195	
Age	1	94568	374255	396626	< 2.2e-16 ***
Sexe	1	1365	374254	395261	< 2.2e-16 ***
Vehicle_Age	1	69248	374253	326013	< 2.2e-16 ***
Marque	12	56	374241	325956	1.038e-07 ***
Couleur	10	31	374231	325926	0.0006478 ***
fuel_type	1	14	374230	325912	0.0002208 ***
nb_siege	1	210	374229	325703	< 2.2e-16 ***
nb_porte	1	13	374228	325689	0.0002961 ***
Annee_fabric	20	69	374208	325620	2.283e-07 ***
Transmission	1	683	374207	324937	< 2.2e-16 ***
Gravite_accident	3	324937	374204	0	< 2.2e-16 ***
Trajet	5	0	374199	0	1.0000000
Condition_eclairage	4	0	374195	0	1.0000000
Condition_meteo	3	0	374192	0	1.0000000

Condition_surface_route	3	0	374189	0	1.0000000
manv	1	0	374188	0	1.0000000
Annual_Premium	1	0	374187	0	1.0000000

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

Ainsi, seules les variables significatives seront interprétées et mis dans le modèle à des fins explicatifs et prédictives.

Estimation des paramètres

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.657e+01	1.400e+04	-0.002	0.998
Age	-2.522e-10	7.277e+01	0.000	1.000
Sexehomme	-7.387e-10	1.185e+03	0.000	1.000
Vehicule_Ageplus de 5 ans	-5.406e-08	2.135e+03	0.000	1.000
MarqueBMW	-5.167e-10	7.305e+03	0.000	1.000
MarqueDAF	-8.088e-10	7.769e+03	0.000	1.000
MarqueFORD	-1.359e-11	7.114e+03	0.000	1.000
MarqueHYUNDAI	3.516e-10	7.505e+03	0.000	1.000
MarqueIVECCO	-1.022e-09	7.031e+03	0.000	1.000
MarqueKIA	4.279e-10	7.063e+03	0.000	1.000
MarqueMERCEDES	-2.344e-10	7.044e+03	0.000	1.000
MarqueMITSUBISHI	1.655e-10	7.266e+03	0.000	1.000
MarqueNISSAN	1.497e-09	6.887e+03	0.000	1.000
MarquePEUGEOT	-8.408e-10	7.890e+03	0.000	1.000
MarqueSUZUKI	-2.578e-10	6.615e+03	0.000	1.000
MarqueTOYOTA	-4.089e-10	6.496e+03	0.000	1.000
Couleurblanc	9.290e-10	2.590e+03	0.000	1.000
Couleurbleu	7.915e-10	3.167e+03	0.000	1.000
Couleurdore	1.043e-09	7.183e+03	0.000	1.000
Couleurgris	1.300e-09	2.859e+03	0.000	1.000
Couleurjaune	1.746e-09	1.167e+04	0.000	1.000
Couleurmarron	1.764e-09	5.203e+03	0.000	1.000
Couleurnoir	-1.698e-10	2.177e+03	0.000	1.000
Couleurrouge	-3.815e-10	4.139e+03	0.000	1.000
Couleurvert	7.656e-10	6.253e+03	0.000	1.000
Couleurviolet	1.301e-10	1.304e+04	0.000	1.000
fuel_typeSuper	-1.208e-09	1.262e+03	0.000	1.000
nb_siege	-1.817e-10	8.584e+01	0.000	1.000
nb_porte	-1.226e-10	2.927e+03	0.000	1.000
Transmissionman	4.337e-10	1.523e+03	0.000	1.000
Gravite_accidentgrave	5.313e+01	1.567e+03	0.034	0.973
Gravite_accidentleger	5.313e+01	6.689e+03	0.008	0.994
Gravite_accidentmodere	5.313e+01	2.224e+03	0.024	0.981

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4.9119e+05 on 374256 degrees of freedom

Residual deviance: 2.1713e-06 on 374224 degrees of freedom
AIC: 66

Number of Fisher Scoring iterations: 25

ODD-RATIO

(Intercept)	Age	Sexehomme
2.900702e-12	1.000000e+00	1.000000e+00
Vehicle_Ageplus de 5 ans	MarqueBMW	MarqueDAF
9.999999e-01	1.000000e+00	1.000000e+00
MarqueFORD	MarqueHYUNDAI	MarqueIVECO
1.000000e+00	1.000000e+00	1.000000e+00
MarqueKIA	MarqueMERCEDES	MarqueMITSUBISHI
1.000000e+00	1.000000e+00	1.000000e+00
MarqueNISSAN	MarquePEUGEOT	MarqueSUZUKI
1.000000e+00	1.000000e+00	1.000000e+00
MarqueTOYOTA	Couleurblanc	Couleurbleu
1.000000e+00	1.000000e+00	1.000000e+00
Couleurdore	Couleurgris	Couleurjaune
1.000000e+00	1.000000e+00	1.000000e+00
Couleurmarron	Couleurnoir	Couleurrouge
1.000000e+00	1.000000e+00	1.000000e+00
Couleurvert	Couleurviolet	fuel_typeSuper
1.000000e+00	1.000000e+00	1.000000e+00
nb_siege	nb_porte	Transmissionman
1.000000e+00	1.000000e+00	1.000000e+00
Gravite_accidentgrave	Gravite_accidentleger	Gravite_accidentmodere
1.188481e+23	1.188480e+23	1.188481e+23

Interpretation des Odd Ratios

Intercept : En l'absence de tous les prédicteurs (quand tous les prédicteurs sont à leur valeur de référence ou à zéro), la probabilité de survenue d'un accident est extrêmement faible.

Age : Lorsque l'âge du conducteur augmente d'un an, cela ne change pas le risque d'accident. Ainsi, un conducteur plus âgé d'un an n'augmente pas ses chances d'avoir un accident (odd ratio de 1).

Sexe homme : par rapport à une femme (modalité de référence), Être un homme n'affecte pas le risque d'accident (odd ratio de 1).

Vehicle Age plus de 5 ans : par rapport à des véhicules de moins de 5 ans (modalité de référence), Le fait que le véhicule ait plus de 5 ans n'influence pas le risque d'accident (odd ratio de 1).

Marque des véhicules (BMW, DAF, FORD, HYUNDAI, IVECO, KIA, MERCEDES, MITSUBISHI, NISSAN, PEUGEOT, SUZUKI, TOYOTA) : La

marque du véhicule n'a pas d'effet sur le risque d'accident. Par exemple, posséder un véhicule BMW, par rapport à la modalité de référence (peut-être une marque non spécifiée), ne change pas les chances d'avoir un accident (odd ratio de 1).

Nombre de sièges : Le nombre de sièges dans le véhicule n'influence pas le risque d'accident (odd ratio de 1).

Nombre de portes : Le nombre de portes dans le véhicule n'affecte pas le risque d'accident (odd ratio de 1).

Transmission manuelle : Utiliser une transmission manuelle, par rapport à une transmission automatique (modalité de référence), ne change pas le risque d'accident (odd ratio de 1).

Gravité de l'accident : Si l'on considère les conducteurs sans aucun accident comme le groupe de référence, l'odd ratio de $1.188480e+23$ pour un accident modéré signifie que le fait d'avoir déjà eu un accident modéré est fortement associé à une probabilité extrêmement élevée d'avoir un autre accident dans le futur.

Effets marginaux

Call:

```
logitmx(formula = accident ~ Age + Sexe + Vehicule_Age + Marque +
  Couleur + fuel_type + nb_siege + nb_porte + Transmission +
  Gravite_accident, data = assurance_auto_mak)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z
Age	-1.9288e-13	5.5664e-02	0.0000e+00	1.0000
Sexehomme	-5.6488e-13	9.0616e-01	0.0000e+00	1.0000
Vehicule_Ageplus de 5 ans	-4.1347e-11	1.6328e+00	0.0000e+00	1.0000
MarqueBMW	-3.9535e-13	5.5878e+00	0.0000e+00	1.0000
MarqueDAF	-6.1884e-13	5.9427e+00	0.0000e+00	1.0000
MarqueFORD	-1.0436e-14	5.4412e+00	0.0000e+00	1.0000
MarqueHYUNDAI	2.6912e-13	5.7409e+00	0.0000e+00	1.0000
MarqueIVECO	-7.8193e-13	5.3780e+00	0.0000e+00	1.0000
MarqueKIA	3.2752e-13	5.4029e+00	0.0000e+00	1.0000
MarqueMERCEDES	-1.7941e-13	5.3879e+00	0.0000e+00	1.0000
MarqueMITSUBISHI	1.2657e-13	5.5580e+00	0.0000e+00	1.0000
MarqueNISSAN	1.1446e-12	5.2677e+00	0.0000e+00	1.0000
MarquePEUGEOT	-6.4315e-13	6.0348e+00	0.0000e+00	1.0000
MarqueSUZUKI	-1.9729e-13	5.0595e+00	0.0000e+00	1.0000
MarqueTOYOTA	-3.1264e-13	4.9689e+00	0.0000e+00	1.0000
Couleurblanc	7.1054e-13	1.9810e+00	0.0000e+00	1.0000
Couleurbleu	6.0552e-13	2.4228e+00	0.0000e+00	1.0000

```

Couleurdore          7.9769e-13  5.4943e+00  0.0000e+00  1.0000
Couleurgris          9.9432e-13  2.1866e+00  0.0000e+00  1.0000
Couleurjaune         1.3355e-12  8.9245e+00  0.0000e+00  1.0000
Couleurmarron        1.3494e-12  3.9799e+00  0.0000e+00  1.0000
Couleurnoir          -1.2990e-13  1.6656e+00  0.0000e+00  1.0000
Couleurrouge         -2.9199e-13  3.1658e+00  0.0000e+00  1.0000
Couleurvert          5.8553e-13  4.7827e+00  0.0000e+00  1.0000
Couleurviolet        9.9476e-14  9.9713e+00  0.0000e+00  1.0000
fuel_typeSuper       -9.2382e-13  9.6495e-01  0.0000e+00  1.0000
nb_siege             -1.3899e-13  6.5661e-02  0.0000e+00  1.0000
nb_porte             -1.8730e-13  4.4785e+00  0.0000e+00  1.0000
Transmissionman      3.3185e-13  1.1650e+00  0.0000e+00  1.0000
Gravite_accidentgrave 1.0000e+00  1.3415e-06  7.4545e+05  <2e-16
***
Gravite_accidentleger 1.1638e-03  6.7949e-01  1.7000e-03  0.9986
Gravite_accidentmodere 1.8876e-01  9.6375e+01  2.0000e-03  0.9984
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

dF/dx is for discrete change for the following variables:

```

[1] "Sexehomme"          "Vehicule_Ageplus de 5 ans"
[3] "MarqueBMW"          "MarqueDAF"
[5] "MarqueFORD"         "MarqueHYUNDAI"
[7] "MarqueIVECCO"       "MarqueKIA"
[9] "MarqueMERCEDES"     "MarqueMITSUBISHI"
[11] "MarqueNISSAN"       "MarquePEUGEOT"
[13] "MarqueSUZUKI"       "MarqueTOYOTA"
[15] "Couleurblanc"       "Couleurbleu"
[17] "Couleurdore"        "Couleurgris"
[19] "Couleurjaune"       "Couleurmarron"
[21] "Couleurnoir"        "Couleurrouge"
[23] "Couleurvert"        "Couleurviolet"
[25] "fuel_typeSuper"     "nb_porte"
[27] "Transmissionman"    "Gravite_accidentgrave"
[29] "Gravite_accidentleger" "Gravite_accidentmodere"

```

- Par rapport à quelqu'un qui a fait un accident léger ou modéré, la probabilité de faire un accident pour quelqu'un qui à déjà fait un accident grave augmente de 1 unité, toutes choses égales par ailleurs.
- Lorsque l'âge augmente d'une année supplément, la probabilité de faire un accident diminue de 1.9288e-13 unité, toutes choses égales par ailleurs
- Par rapport à une femme, la probabilité qu'un homme fasse un accident diminue de 5.6488e-13 unité, toutes choses égales par ailleurs

Taux de mauvais classement (TMC)

[1] 100

Modele finale:

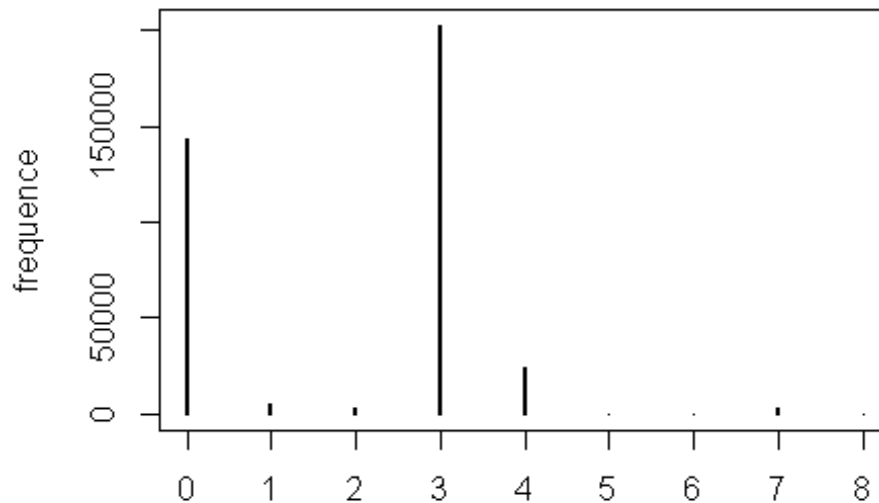
accident = B0 + B1Age + B2Sexe + B3Vehivle_Age + B4Marque + B5Couleur + B6fuel_type + B7nb_siege + B8nb_porte + B9Annee_fabric + B10Transmission + B11Gravite_accident + ε

V- REGRESSION POISSON

Importation du jeu de données

```
tibble [382,154 × 19] (S3: tbl_df/tbl/data.frame)
  Age                : num [1:382154] 22 22 22 22 22 22 22 22 22 22 ...
  Sexe               : chr [1:382154] "femme" "femme" "homme" "homme"
  ...
  Vehicule_Age       : chr [1:382154] "moins de 5 ans" "plus de 5 ans"
"plus de 5 ans" "moins de 5 ans" ...
  Marque             : chr [1:382154] "NISSAN" "NISSAN" "NISSAN"
"NISSAN" ...
  Couleur            : chr [1:382154] "gris" "argent" "argent" "gris"
  ...
  fuel_type          : chr [1:382154] "Gazoil" "Gazoil" "Gazoil"
"Gazoil" ...
  nb_siege           : num [1:382154] 21 21 21 21 21 21 21 21 21 21 ...
  nb_porte           : num [1:382154] 4 4 4 4 4 4 4 4 4 4 ...
  Annee_fabric       : num [1:382154] 2007 2008 2000 2010 2006 ...
  Transmission       : chr [1:382154] "man" "man" "man" "man" ...
  Gravite_accident   : chr [1:382154] "grave" "modere" "grave" "grave"
  ...
  Trajet             : chr [1:382154] "trajet 1" "trajet 1" "trajet 3"
"trajet 3" ...
  Condition_eclairage : chr [1:382154] "eclairage 3" "eclairage 2"
"eclairage 2" "eclairage 1" ...
  Condition_meteo     : chr [1:382154] "meteo 1" "meteo 1" "meteo 1"
"meteo 1" ...
  Condition_surface_route: chr [1:382154] "surface 1" "surface 1" "surface
1" "surface 1" ...
  manv               : num [1:382154] 1 17 1 9 1 15 8 1 10 1 ...
  Annual_Premium     : num [1:382154] 2630 43327 35841 27645 29023 ...
  ACCIDENT           : chr [1:382154] "Oui" "Oui" "Oui" "Oui" ...
  frequence          : num [1:382154] 3 3 4 3 7 4 4 2 3 2 ...
```

Graphique



La distribution n'a pas une allure en « L », calculons maintenant la moyenne et la variance. En effet dans une distribution de poisson ces deux valeurs doivent être égales.

- **Moyenne**

```
[1] 1.934712
```

- **Variance**

```
[1] 2.516623
```

Var > Moyenne Alors on fera un modèle quasi-poisson

Test d'adéquation à la loi de poisson

HO : La variable ne suit pas la loi de poisson

H1 : La variable suit la loi de poisson

```
Chi-squared statistic: 623997.5  
Degree of freedom of the Chi-squared distribution: 5  
Chi-squared p-value: 0
```

Chi-squared table:

	obscounts	theocounts
<= 0	143473.0000	55208.1946
<= 1	5408.0000	106811.9717
<= 2	3530.0000	103325.2163
<= 3	202303.0000	66634.8547
<= 4	23983.0000	32229.8178
<= 7	3132.0000	17603.8560
> 7	325.0000	340.0889

Goodness-of-fit criteria

	1-mle-pois
Akaike's Information Criterion	1443577
Bayesian Information Criterion	1443588

Conclusion : p-value < 0.05. On rejette H0 et on conclut que La variable suit la loi de poisson.

ESTIMATION DU MODELE

frequence = B0 + B1Age + B2Sexe + B3Vehivle_Age + B4Marque + B5Couleur + B6fuel_type + B7nb_siege + B8nb_porte + B9Annee_fabric + B10Transmission + B11Gravite_accident + B12Trajet + B13 Condition_eclairage + B14Condition_meteo + B15Condition_surface_route + B16manv + B17Annual_premium + B18ACCIDENT + ϵ

Call:

```
glm(formula = frequence ~ ., family = quasipoisson, data =  
assurance_poisson)
```

Coefficients: (3 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.938e+07	5.223e+08	0.190	0.849112
Age	2.143e-04	4.479e-05	4.785	1.71e-06 ***
Sexehomme	-9.741e-04	6.777e-04	-1.437	0.150631
Vehivle_Ageplus de 5 ans	3.673e-04	1.238e-03	0.297	0.766778
MarqueBMW	2.199e-02	4.271e-03	5.148	2.64e-07 ***
MarqueDAF	3.477e-03	4.563e-03	0.762	0.446079
MarqueFORD	1.202e-02	4.171e-03	2.881	0.003958 **
MarqueHYUNDAI	1.839e-02	4.397e-03	4.182	2.89e-05 ***
MarqueIVECO	1.298e-02	4.124e-03	3.146	0.001653 **
MarqueKIA	3.729e-02	4.125e-03	9.040	< 2e-16 ***
MarqueMERCEDES	1.772e-02	4.126e-03	4.294	1.75e-05 ***
MarqueMITSUBISHI	2.077e-02	4.251e-03	4.886	1.03e-06 ***
MarqueNISSAN	1.177e-02	4.038e-03	2.914	0.003565 **
MarquePEUGEOT	2.037e-02	4.613e-03	4.414	1.01e-05 ***
MarqueSUZUKI	2.103e-02	3.876e-03	5.425	5.81e-08 ***
MarqueTOYOTA	2.424e-02	3.808e-03	6.365	1.96e-10 ***
Couleurblanc	1.017e-03	1.490e-03	0.682	0.495032
Couleurbleu	7.324e-03	1.815e-03	4.035	5.47e-05 ***
Couleurdore	6.070e-03	4.140e-03	1.466	0.142610
Couleurgris	4.924e-03	1.646e-03	2.992	0.002768 **

Couleurjaune	5.293e-03	6.710e-03	0.789	0.430163	
Couleurmarron	-4.033e-03	3.012e-03	-1.339	0.180662	
Couleurnoir	-1.025e-02	1.273e-03	-8.056	7.90e-16	***
Couleurrouge	3.047e-03	2.379e-03	1.281	0.200296	
Couleurvert	5.671e-03	3.601e-03	1.575	0.115275	
Couleurviolet	2.246e-02	7.517e-03	2.988	0.002812	**
fuel_typeSuper	-1.245e-02	7.431e-04	-16.760	< 2e-16	***
nb_siege	1.283e-03	4.608e-05	27.838	< 2e-16	***
nb_porte	1.380e-02	1.680e-03	8.217	< 2e-16	***
Annee_fabric	-8.859e-04	8.207e-05	-10.795	< 2e-16	***
Transmissionman	2.003e-03	8.717e-04	2.298	0.021573	*
Gravite_accidentgrave	4.549e+08	2.743e+08	1.658	0.097234	.
Gravite_accidentleger	4.549e+08	2.743e+08	1.658	0.097234	.
Gravite_accidentmodere	4.549e+08	2.743e+08	1.658	0.097234	.
Trajettrajet 1	-1.013e-03	1.870e-03	-0.541	0.588226	
Trajettrajet 2	-5.121e-03	3.129e-03	-1.637	0.101644	
Trajettrajet 3	2.898e-03	1.650e-03	1.756	0.079005	.
Trajettrajet 4	-1.867e-02	1.564e-03	-11.941	< 2e-16	***
Trajettrajet 5	4.365e-03	1.639e-03	2.663	0.007746	**
Trajettrajet 6	NA	NA	NA	NA	
Condition_eclairageeclairage 1	9.600e-03	2.471e-03	3.885	0.000102	***
Condition_eclairageeclairage 2	1.418e-02	2.548e-03	5.564	2.63e-08	***
Condition_eclairageeclairage 3	1.815e-03	4.753e-03	0.382	0.702463	
Condition_eclairageeclairage 4	2.423e-03	2.874e-03	0.843	0.399301	
Condition_eclairageeclairage 5	NA	NA	NA	NA	
Condition_meteometeo 1	-5.543e+08	5.891e+08	-0.941	0.346778	
Condition_meteometeo 2	-5.543e+08	5.891e+08	-0.941	0.346778	
Condition_meteometeo 3	-5.543e+08	5.891e+08	-0.941	0.346778	
Condition_meteometeo 4	-5.543e+08	5.891e+08	-0.941	0.346778	
Condition_surface_route4	-1.050e-02	1.139e-02	-0.922	0.356442	
Condition_surface_routesurface 0	-9.938e+07	5.223e+08	-0.190	0.849112	
Condition_surface_routesurface 1	5.327e-03	1.065e-02	0.500	0.616853	
Condition_surface_routesurface 2	4.226e-03	1.063e-02	0.398	0.690847	
manv	1.595e-03	7.097e-05	22.478	< 2e-16	***
Annual_Premium	-6.573e-08	2.096e-08	-3.136	0.001711	**
ACCIDENTOui	NA	NA	NA	NA	

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

(Dispersion parameter for quasipoisson family taken to be 0.08306821)

Null deviance: 726819 on 382153 degrees of freedom

Residual deviance: 30289 on 382101 degrees of freedom

AIC: NA

Number of Fisher Scoring iterations: 25

 **Test d'anova permet de voir les variables les plus liées à la cible**

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			382153	726819	
Age	1	121479	382152	605340	< 2.2e-16 ***

Sexe	1	938	382151	604402	< 2.2e-16	***
Vehicle_Age	1	45539	382150	558863	< 2.2e-16	***
Marque	12	69	382138	558794	< 2.2e-16	***
Couleur	10	125	382128	558669	< 2.2e-16	***
fuel_type	1	107	382127	558563	< 2.2e-16	***
nb_siege	1	2	382126	558561	7.636e-06	***
nb_porte	1	10	382125	558552	< 2.2e-16	***
Annee_fabric	1	25	382124	558527	< 2.2e-16	***
Transmission	1	474	382123	558053	< 2.2e-16	***
Gravite_accident	3	527615	382120	30437	< 2.2e-16	***
Trajet	5	97	382115	30340	< 2.2e-16	***
Condition_eclairage	4	7	382111	30333	< 2.2e-16	***
Condition_meteo	3	0	382108	30332	0.226680	
Condition_surface_route	4	1	382104	30331	0.002848	**
manv	2	42	382102	30290	< 2.2e-16	***
Annual_Premium	1	1	382101	30289	0.001717	**
ACCIDENT	0	0	382101	30289		

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

Reprenons le modèle avec ces variables

Call:

```
glm(formula = frequence ~ Age + Sexe + Vehicle_Age + Marque +
    Couleur + fuel_type + nb_siege + nb_porte + Annee_fabric +
    Transmission + Gravite_accident + Trajet + Condition_eclairage +
    manv + Annual_Premium, family = quasipoisson, data = assurance_poisson)
```

Coefficients: (2 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.961e+01	1.950e+01	-1.006	0.31442
Age	2.149e-04	4.480e-05	4.796	1.62e-06 ***
Sexehomme	-9.632e-04	6.778e-04	-1.421	0.15528
Vehicle_Ageplus de 5 ans	3.502e-04	1.239e-03	0.283	0.77733
MarqueBMW	2.195e-02	4.271e-03	5.138	2.78e-07 ***
MarqueDAF	3.549e-03	4.564e-03	0.778	0.43677
MarqueFORD	1.202e-02	4.171e-03	2.881	0.00397 **
MarqueHYUNDAI	1.839e-02	4.397e-03	4.182	2.89e-05 ***
MarqueIVECO	1.287e-02	4.124e-03	3.121	0.00180 **
MarqueKIA	3.727e-02	4.125e-03	9.034	< 2e-16 ***
MarqueMERCEDES	1.766e-02	4.126e-03	4.279	1.88e-05 ***
MarqueMITSUBISHI	2.072e-02	4.252e-03	4.875	1.09e-06 ***
MarqueNISSAN	1.176e-02	4.038e-03	2.912	0.00359 **
MarquePEUGEOT	2.035e-02	4.614e-03	4.410	1.03e-05 ***
MarqueSUZUKI	2.098e-02	3.877e-03	5.413	6.22e-08 ***
MarqueTOYOTA	2.420e-02	3.809e-03	6.353	2.11e-10 ***
Couleurblanc	1.009e-03	1.490e-03	0.677	0.49815
Couleurbleu	7.333e-03	1.816e-03	4.039	5.37e-05 ***
Couleurdore	6.059e-03	4.141e-03	1.463	0.14336
Couleurgris	4.945e-03	1.646e-03	3.005	0.00266 **
Couleurjaune	5.217e-03	6.709e-03	0.778	0.43682
Couleurmarron	-4.006e-03	3.013e-03	-1.330	0.18365
Couleurnoir	-1.018e-02	1.273e-03	-7.996	1.29e-15 ***
Couleurrouge	2.987e-03	2.379e-03	1.256	0.20923
Couleurvert	5.700e-03	3.601e-03	1.583	0.11345

Couleurviolet	2.242e-02	7.518e-03	2.982	0.00286	**
fuel_typeSuper	-1.248e-02	7.431e-04	-16.798	< 2e-16	***
nb_siege	1.285e-03	4.607e-05	27.894	< 2e-16	***
nb_porte	1.384e-02	1.680e-03	8.242	< 2e-16	***
Annee_fabric	-8.854e-04	8.208e-05	-10.787	< 2e-16	***
Transmissionman	2.018e-03	8.718e-04	2.315	0.02063	*
Gravite_accidentgrave	2.242e+01	1.950e+01	1.150	0.25004	
Gravite_accidentleger	2.242e+01	1.950e+01	1.150	0.25008	
Gravite_accidentmodere	2.243e+01	1.950e+01	1.150	0.24999	
Trajettrajet 1	-1.019e-03	1.870e-03	-0.545	0.58600	
Trajettrajet 2	-5.125e-03	3.129e-03	-1.638	0.10145	
Trajettrajet 3	2.894e-03	1.650e-03	1.754	0.07941	.
Trajettrajet 4	-1.873e-02	1.564e-03	-11.980	< 2e-16	***
Trajettrajet 5	4.359e-03	1.639e-03	2.659	0.00783	**
Trajettrajet 6	NA	NA	NA	NA	
Condition_eclairageeclairage 1	1.025e-02	2.458e-03	4.172	3.02e-05	***
Condition_eclairageeclairage 2	1.453e-02	2.536e-03	5.731	9.97e-09	***
Condition_eclairageeclairage 3	1.884e-03	4.746e-03	0.397	0.69136	
Condition_eclairageeclairage 4	2.662e-03	2.857e-03	0.932	0.35152	
Condition_eclairageeclairage 5	NA	NA	NA	NA	
manv	1.597e-03	7.097e-05	22.506	< 2e-16	***
Annual_Premium	-6.590e-08	2.096e-08	-3.144	0.00167	**

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

(Dispersion parameter for quasipoisson family taken to be 0.08307805)

Null deviance: 726819 on 382153 degrees of freedom
Residual deviance: 30290 on 382109 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 19

 **Anova**

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			382153	726819	
Age	1	121479	382152	605340	< 2.2e-16 ***
Sexe	1	938	382151	604402	< 2.2e-16 ***
Vehicle_Age	1	45539	382150	558863	< 2.2e-16 ***
Marque	12	69	382138	558794	< 2.2e-16 ***
Couleur	10	125	382128	558669	< 2.2e-16 ***
fuel_type	1	107	382127	558563	< 2.2e-16 ***
nb_siege	1	2	382126	558561	7.645e-06 ***
nb_porte	1	10	382125	558552	< 2.2e-16 ***
Annee_fabric	1	25	382124	558527	< 2.2e-16 ***
Transmission	1	474	382123	558053	< 2.2e-16 ***
Gravite_accident	3	527615	382120	30437	< 2.2e-16 ***
Trajet	5	97	382115	30340	< 2.2e-16 ***
Condition_eclairage	4	7	382111	30333	< 2.2e-16 ***
manv	1	42	382110	30291	< 2.2e-16 ***
Annual_Premium	1	1	382109	30290	0.001661 **

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

Prenons l'exponentiel des coefficients

(Intercept)	Age
3.035976e-09	1.000215e+00
Sexehomme	Vehicle_Ageplus de 5 ans
9.990372e-01	1.000350e+00
MarqueBMW	MarqueDAF
1.022188e+00	1.003555e+00
MarqueFORD	MarqueHYUNDAI
1.012089e+00	1.018561e+00
MarqueIVECCO	MarqueKIA
1.012957e+00	1.037974e+00
MarqueMERCEDES	MarqueMITSUBISHI
1.017813e+00	1.020941e+00
MarqueNISSAN	MarquePEUGEOT
1.011829e+00	1.020556e+00
MarqueSUZUKI	MarqueTOYOTA
1.021204e+00	1.024491e+00
Couleurblanc	Couleurbleu
1.001010e+00	1.007360e+00
Couleurdore	Couleurgris
1.006078e+00	1.004957e+00
Couleurjaune	Couleurmarron
1.005231e+00	9.960025e-01
Couleurnoir	Couleurrouge
9.898752e-01	1.002992e+00
Couleurvert	Couleurviolet
1.005717e+00	1.022674e+00
fuel_typeSuper	nb_siege
9.875945e-01	1.001286e+00
nb_porte	Annee_fabric
1.013941e+00	9.991150e-01
Transmissionman	Gravite_accidentgrave
1.002020e+00	5.480185e+09
Gravite_accidentleger	Gravite_accidentmodere
5.470039e+09	5.494084e+09
Trajettrajet 1	Trajettrajet 2
9.989818e-01	9.948886e-01
Trajettrajet 3	Trajettrajet 4
1.002899e+00	9.814409e-01
Trajettrajet 5	Trajettrajet 6
1.004369e+00	NA
Condition_eclairageeclairage 1	Condition_eclairageeclairage 2
1.010308e+00	1.014640e+00
Condition_eclairageeclairage 3	Condition_eclairageeclairage 4
1.001886e+00	1.002665e+00
Condition_eclairageeclairage 5	manv
NA	1.001599e+00
Annual_Premium	
9.999999e-01	

Interpretation

Une augmentation d'un an de l'Age d'un assuré augmente la fréquence d'accident 1.00016

Utiliser la marque BMF augmente la fréquence d'accident de 1.022

Utiliser la marque KIA augmente la fréquence d'accident de 1.038

Utiliser la marque MERCEDES augmente la fréquence d'accident de 1.018

Utiliser la marque SUZUKI augmente la fréquence d'accident de 1.021

Utiliser la marque TOYOTA augmente la fréquence d'accident de 1.024

Utiliser une voiture de couleur bleu augmente la fréquence d'accident de 1.007

Utiliser une voiture de couleur violet augmente la fréquence d'accident de 1.022

Un assuré qui utilise une voiture de carburant super diminue la fréquence d'accident de 0.99

Une augmentation d'un siège supplémentaire à la voiture d'un assuré augmente la fréquence d'accident 1.0011

Une augmentation d'une porte supplémentaire à la voiture d'un assuré augmente la fréquence d'accident 1.013

Utiliser une voiture fabriquée en 2003 diminue la fréquence d'accident de 0.9915

Utiliser une voiture fabriquée en 2010 augmente la fréquence d'accident de 1.005

Utiliser une voiture fabriquée en 2015 diminue la fréquence d'accident de 0.981

Utiliser le trajet 4 diminue la frequence d'accident de 0.982

Utiliser le trajet 5 augmente la frequence d'accident de 1.004

Rouler dans des conditions d'eclairage 1 augmente la frequence d'accident de 1.010

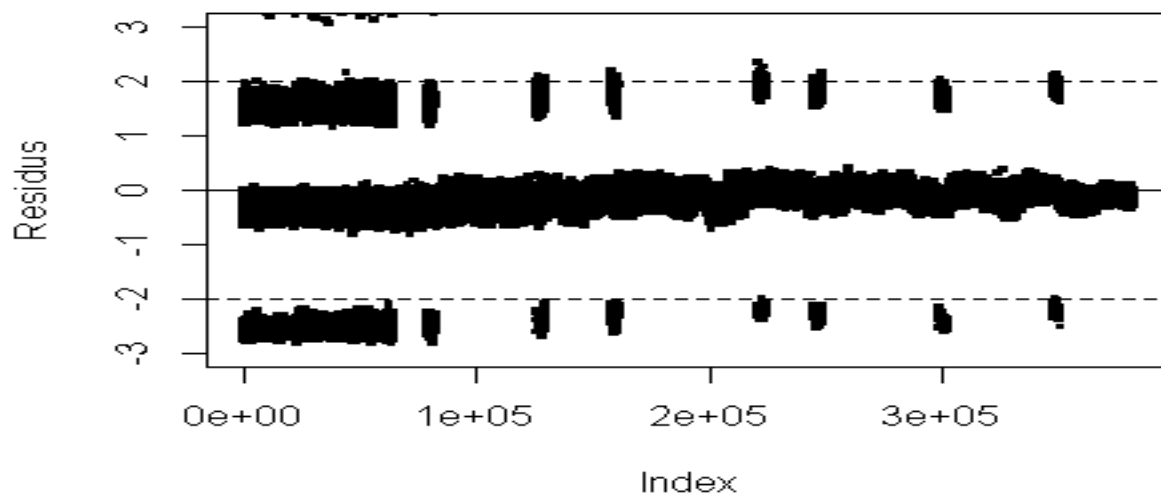
Rouler dans des conditions d'eclairage 2 augmente la frequence d'accident de 1.015

Une augmentation d'une manoeuvre supplementaire augmente la frequence d'accident de 1.001

Une augmentation d'une prime annuelle supplementaire augmente la frequence d'accident de 0.99

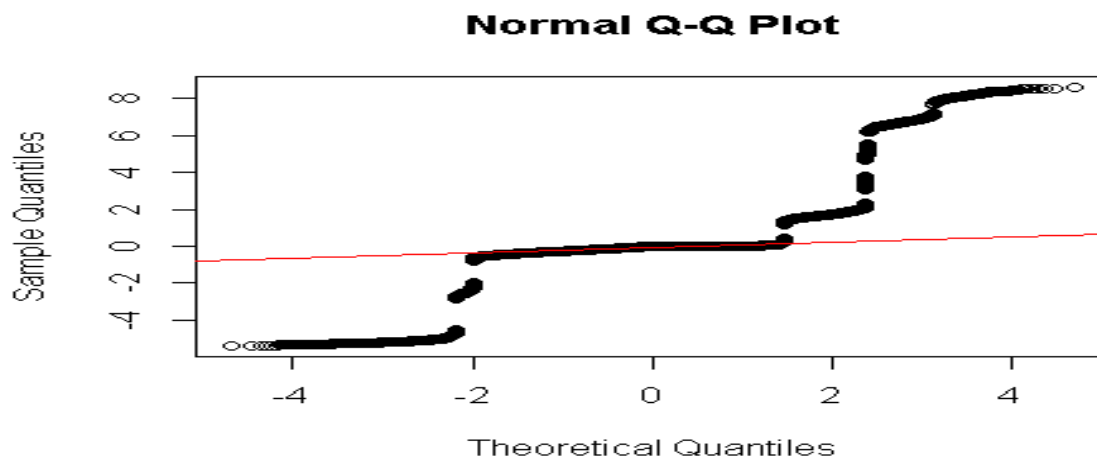
ANALYSE DES RESIDUS

- Graphique



- Test de normalité

- Graphique



– Test

H0 : les résidus suivent une distribution normale

H1 : les résidus ne suivent pas une distribution normale

```
data: res.m
D = 0.37187, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Conclusion : $p\text{-value} < 0.05$ alors on ne peut rejeter H0, le test de Kolmogorov-Smirnov confirme que les résidus ne suivent pas une distribution normale.

FAIRE LA PREVISION

Prédire la fréquence d'accident des individus 10, 11 et 12 :

10	11	12
3.201383	3.171189	3.245014

Intervalle de confiance

10	10	11	11	12	12
3.189411	3.213399	3.151618	3.190881	3.236228	3.253823

CONCLUSION GENERALE

Cette étude avait pour objectif principal de modéliser, dans un premier temps, la probabilité de survenance d'un accident en fonction du profil de l'assuré à l'aide d'une régression logistique. Pour ce faire, la variable cible « fréquence » a été binarisée et renommée « accident », permettant ainsi de distinguer les assurés ayant eu au moins un accident de ceux n'en ayant pas eu. Dans un deuxième temps, l'étude visait à construire un modèle poissonnien pour modéliser le nombre d'accidents en fonction des caractéristiques des assurés et des véhicules. Les résultats de l'analyse ont révélé plusieurs insights clés. Tout d'abord, l'intervalle de confiance à 95% pour la proportion d'assurés ayant déjà eu un accident se situe entre 63% et 64%. Ce résultat est statistiquement significatif, comme en témoigne le test de conformité qui rejette l'hypothèse nulle (H_0) avec une p-value inférieure à 0,05, confirmant que la proportion d'assurés ayant eu un accident est significativement supérieure à 63%. Par ailleurs, les tests de liaison ont montré que toutes les variables explicatives (âge, sexe, âge du véhicule, marque, couleur, type de carburant, nombre de sièges, conditions d'éclairage, conditions météorologiques, etc.) sont liées à la variable cible « accident ». Cela souligne l'importance de ces facteurs dans la survenue des sinistres et confirme leur pertinence dans la modélisation. Cependant, la régression logistique a présenté un taux de mauvais classement de 100%, ce qui suggère une inadéquation du modèle ou des problèmes potentiels liés à la qualité des données (par exemple, des données fictives, déséquilibrées ou non représentatives). Malgré cela, la structure de la régression logistique elle-même est jugée valide, ce qui indique que le problème pourrait résider dans la nature des données plutôt que dans la méthode. En ce qui concerne le modèle poissonnien, les résultats ont montré que les résidus ne suivent pas une distribution normale (p-value < 0,05 selon le test de Kolmogorov-Smirnov), ce qui est une

limitation. Néanmoins, le modèle a permis de réaliser des prévisions fiables pour les individus 10, 11 et 12, avec des intervalles de confiance précis à 95% (respectivement [3,189411 ; 3,213399], [3,151618 ; 3,190881] et [3,236228 ; 3,253823]). Cela démontre que, malgré la non-normalité des résidus, le modèle poissonnien reste utile pour prédire le nombre d'accidents.

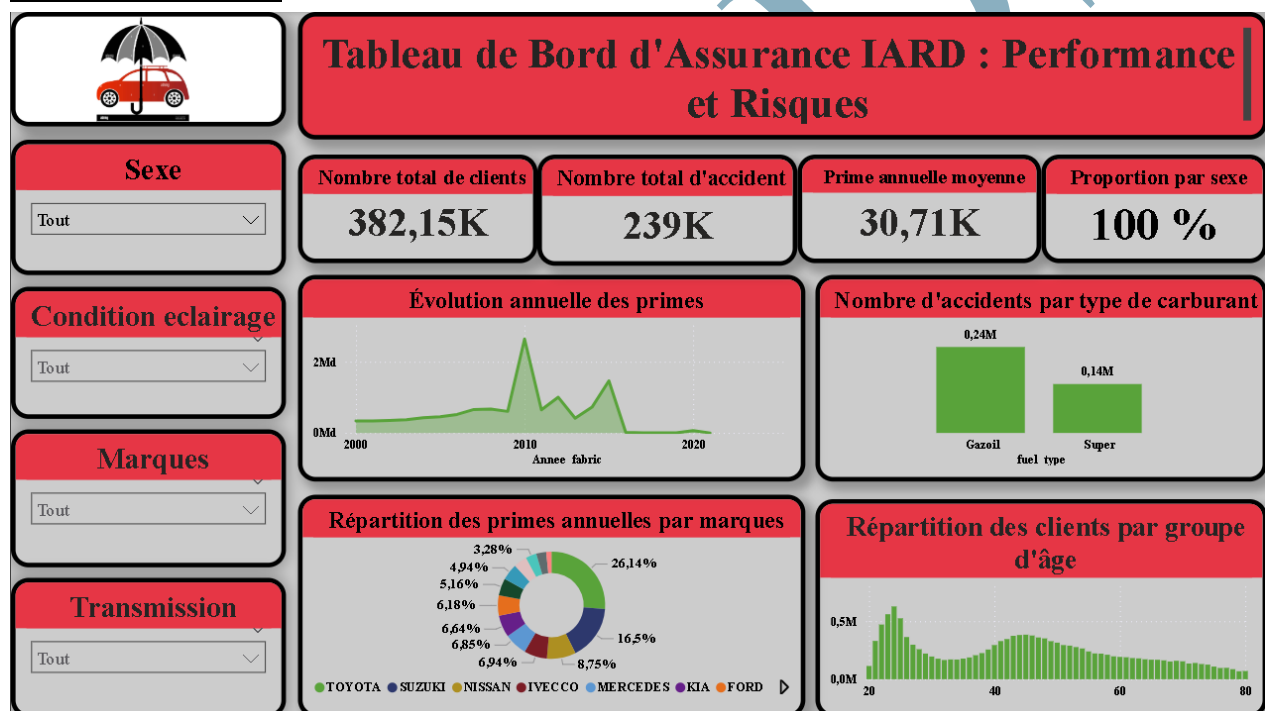
Recommandations pour la Société d'Assurance

- **Améliorer la qualité des données :** Le taux de mauvais classement de 100% dans la régression logistique suggère que les données pourraient être bruitées, déséquilibrées ou fictives. Il est essentiel de vérifier la qualité et la représentativité des données avant de déployer des modèles prédictifs en production.
- **Affiner les modèles prédictifs :** Bien que la régression logistique soit structurellement valide, il serait judicieux d'explorer d'autres méthodes de classification (par exemple, les forêts aléatoires, les machines à vecteurs de support ou les réseaux de neurones) pour améliorer la précision des prédictions. De même, pour le modèle poissonnien, des techniques de correction des résidus pourraient être envisagées pour mieux respecter les hypothèses statistiques.
- **Prendre en compte les facteurs de risque identifiés :** Les variables liées à l'âge de l'assuré, à l'âge du véhicule, aux conditions météorologiques, à l'éclairage et à la surface de la route sont des déterminants clés de la survenue d'accidents. Ces facteurs devraient être intégrés dans les politiques de tarification et de gestion des risques pour mieux segmenter les clients et ajuster les primes en fonction du risque réel.
- **Mettre en place des actions de prévention :** Étant donné que plus de 63% des assurés ont déjà eu un accident, il est crucial de développer des campagnes de sensibilisation et des programmes de prévention ciblés. Par exemple, des alertes en

temps réel sur les conditions de conduite dangereuses (météo, éclairage, état de la route) pourraient réduire la fréquence des sinistres.

- Utiliser les prévisions pour la gestion des sinistres : Les intervalles de confiance obtenus avec le modèle poissonnien permettent d'anticiper le nombre d'accidents pour des individus spécifiques. Ces prévisions peuvent être utilisées pour optimiser la gestion des sinistres, allouer les ressources plus efficacement et améliorer la satisfaction client.

POWER BI



CODE R ET POWER QUERY

I- Prétraitements des données

let

Source

Csv.Document (File.Contents ("C:\Users\HP\Downloads\INSSEDS\cours\E
CONOMETRIE\Mini projet économétrie Reg logic et

```
poisson\assurance_auto_makani.csv"), [Delimiter=";", Columns=19,
Encoding=1252, QuoteStyle=QuoteStyle.None]),
```

```
    #"En-têtes      promus"      =      Table.PromoteHeaders(Source,
[PromoteAllScalars=true]),
```

```
    #"Type      modifié"      =      Table.TransformColumnTypes(#"En-têtes
promus",{{"Age",      Int64.Type},      {"Sexe",      type      text},
{"Vehicle_Age",      type      text}, {"BRANDS",      type      text}, {"Colour",
type      text}, {"fuel_type",      type      text}, {"seat_count",      Int64.Type},
{"door_count",      Int64.Type}, {"manufacture_year",      Int64.Type},
{"transmission",      type      text}, {"ACCIDENT",      type      text},
{"SEVERITY",      Int64.Type}, {"trajet",      Int64.Type},
{"light_conditions",      Int64.Type}, {"weather_conditions",
Int64.Type}, {"road_surface_conditions",      Int64.Type}, {"manv",
Int64.Type}, {"frequence",      Int64.Type}, {"Annual_Premium",
Int64.Type}}),
```

```
    #"Valeur      remplacée"      =      Table.ReplaceValue(#"Type
modifié","1","homme",Replacer.ReplaceText,{"Sexe"}),
```

```
    #"Valeur      remplacée1"      =      Table.ReplaceValue(#"Valeur
remplacée","0","femme",Replacer.ReplaceText,{"Sexe"}),
```

```
    #"Valeur      remplacée2"      =      Table.ReplaceValue(#"Valeur
remplacée1","<      5","moins      de      5
ans",Replacer.ReplaceText,{"Vehicle_Age"}),
```

```
    #"Valeur      remplacée3"      =      Table.ReplaceValue(#"Valeur
remplacée2",">      5","plus      de      5
ans",Replacer.ReplaceText,{"Vehicle_Age"}),
```

```
    #"Colonnes      renommées"      =      Table.RenameColumns(#"Valeur
remplacée3",{{"BRANDS", "Marque"}, {"Colour", "Couleur"}}),
```

```
    #"Valeur      remplacée4"      =      Table.ReplaceValue(#"Colonnes
renommées","black","noir",Replacer.ReplaceText,{"Couleur"}),
```

```
    #"Valeur      remplacée5"      =      Table.ReplaceValue(#"Valeur
remplacée4","white","blanc",Replacer.ReplaceText,{"Couleur"}),
```

```
    #"Valeur      remplacée6"      =      Table.ReplaceValue(#"Valeur
remplacée5","gray","gris",Replacer.ReplaceText,{"Couleur"}),
```

```
    #"Valeur      remplacée7"      =      Table.ReplaceValue(#"Valeur
remplacée6","red","rouge",Replacer.ReplaceText,{"Couleur"}),
```

```
    #"Valeur      remplacée8"      =      Table.ReplaceValue(#"Valeur
remplacée7","silver","argent",Replacer.ReplaceText,{"Couleur"}),
```



```

    # "Valeur remplacée9" = Table.ReplaceValue("#Valeur
    remplacée8", "blue", "bleu", Replacer.ReplaceText, {"Couleur"}),

    # "Valeur remplacée10" = Table.ReplaceValue("#Valeur
    remplacée9", "brown", "marron", Replacer.ReplaceText, {"Couleur"}),

    # "Valeur remplacée11" = Table.ReplaceValue("#Valeur
    remplacée10", "green", "vert", Replacer.ReplaceText, {"Couleur"}),

    # "Valeur remplacée12" = Table.ReplaceValue("#Valeur
    remplacée11", "golden", "dore", Replacer.ReplaceText, {"Couleur"}),

    # "Valeur remplacée13" = Table.ReplaceValue("#Valeur
    remplacée12", "yellow", "jaune", Replacer.ReplaceText, {"Couleur"}),

    # "Colonnes renommées1" = Table.RenameColumns("#Valeur
    remplacée13", {"seat_count", "nb_siege", "door_count",
    "nb_porte"}, {"manufacture_year", "Annee_fabric"}),

    # "Valeur remplacée14" = Table.ReplaceValue("#Colonnes
    renommées1", "Yes", "Oui", Replacer.ReplaceText, {"ACCIDENT"}),

    # "Valeur remplacée15" = Table.ReplaceValue("#Valeur
    remplacée14", "No", "Non", Replacer.ReplaceText, {"ACCIDENT"}),

    # "Colonnes renommées2" = Table.RenameColumns("#Valeur
    remplacée15", {"SEVERITY", "Gravite_accident"}),

    # "Colonnes permutées" = Table.ReorderColumns("#Colonnes
    renommées2", {"Age", "Sexe", "Vehicle_Age", "Marque", "Couleur",
    "fuel_type", "nb_siege", "nb_porte", "Annee_fabric",
    "transmission", "Gravite_accident", "trajet", "light_conditions",
    "weather_conditions", "road_surface_conditions", "manv",
    "frequence", "Annual_Premium", "ACCIDENT"}),

    # "Type modifié1" = Table.TransformColumnTypes("#Colonnes
    permutées", {"Gravite_accident", type text}),

    # "Valeur remplacée16" = Table.ReplaceValue("#Type
    modifié1", "0", "aucun", Replacer.ReplaceText, {"Gravite_accident"}),

    # "Valeur remplacée17" = Table.ReplaceValue("#Valeur
    remplacée16", "0", "aucun", Replacer.ReplaceText, {"Gravite_accident"}),

    # "Valeur remplacée18" = Table.ReplaceValue("#Valeur
    remplacée17", "1", "dommage", Replacer.ReplaceText, {"Gravite_accident"}),

```

```
#"Valeur      remplacée19"      =      Table.ReplaceValue("#Valeur
remplacée18", "2", "dommage
modere", Replacer.ReplaceText, {"Gravite_accident"}),

#"Valeur      remplacée20"      =      Table.ReplaceValue("#Valeur
remplacée19", "3", "dommage
grave", Replacer.ReplaceText, {"Gravite_accident"}),

#"Valeur      remplacée21"      =      Table.ReplaceValue("#Valeur
remplacée20", "aucun
dommage", "aucun", Replacer.ReplaceText, {"Gravite_accident"}),

#"Valeur      remplacée22"      =      Table.ReplaceValue("#Valeur
remplacée21", "dommage
leger", "leger", Replacer.ReplaceText, {"Gravite_accident"}),

#"Valeur      remplacée23"      =      Table.ReplaceValue("#Valeur
remplacée22", "dommage
modere", "modere", Replacer.ReplaceText, {"Gravite_accident"}),

#"Valeur      remplacée24"      =      Table.ReplaceValue("#Valeur
remplacée23", "dommage
grave", "grave", Replacer.ReplaceText, {"Gravite_accident"}),

#"Type      modifié2"      =      Table.TransformColumnTypes("#Valeur
remplacée24", {{ "trajet", type text }}),

#"Valeur      remplacée25"      =      Table.ReplaceValue("#Type
modifié2", "0", "trajet 0", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée26"      =      Table.ReplaceValue("#Valeur
remplacée25", "1", "trajet 1", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée27"      =      Table.ReplaceValue("#Valeur
remplacée26", "2", "trajet 2", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée28"      =      Table.ReplaceValue("#Valeur
remplacée27", "3", "trajet 3", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée29"      =      Table.ReplaceValue("#Valeur
remplacée28", "4", "trajet 4", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée30"      =      Table.ReplaceValue("#Valeur
remplacée29", "5", "trajet 5", Replacer.ReplaceText, {"trajet"}),

#"Valeur      remplacée31"      =      Table.ReplaceValue("#Valeur
remplacée30", "6", "trajet 6", Replacer.ReplaceText, {"trajet"}),

#"Colonnes      renommées3"      =      Table.RenameColumns("#Valeur
remplacée31", {{ "light_conditions", "Condition_eclairage" }}),
```

```

    # "Type    modifié3" = Table.TransformColumnTypes(#"Colonnes
renommées3",{{"Condition_eclairage", type text}}),

    # "Valeur    remplacée32" = Table.ReplaceValue(#"Type
modifié3","1","eclairage
1",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Valeur    remplacée33" = Table.ReplaceValue(#"Valeur
remplacée32","0","eclairage
0",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Valeur    remplacée34" = Table.ReplaceValue(#"Valeur
remplacée33","2","eclairage
2",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Valeur    remplacée35" = Table.ReplaceValue(#"Valeur
remplacée34","3","eclairage
3",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Valeur    remplacée36" = Table.ReplaceValue(#"Valeur
remplacée35","4","eclairage
4",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Valeur    remplacée37" = Table.ReplaceValue(#"Valeur
remplacée36","5","eclairage
5",Replacer.ReplaceText,{"Condition_eclairage"}),

    # "Type    modifié4" = Table.TransformColumnTypes(#"Valeur
remplacée37",{{"weather_conditions", type text}}),

    # "Colonnes    renommées4" = Table.RenameColumns(#"Type
modifié4",{{"weather_conditions", "Condition_meteo"}, {"trajet",
"Trajet"}}),

    # "Valeur    remplacée38" = Table.ReplaceValue(#"Colonnes
renommées4","0","meteo
0",Replacer.ReplaceText,{"Condition_meteo"}),

    # "Valeur    remplacée39" = Table.ReplaceValue(#"Valeur
remplacée38","1","meteo
1",Replacer.ReplaceText,{"Condition_meteo"}),

    # "Valeur    remplacée40" = Table.ReplaceValue(#"Valeur
remplacée39","2","meteo
2",Replacer.ReplaceText,{"Condition_meteo"}),

    # "Valeur    remplacée41" = Table.ReplaceValue(#"Valeur
remplacée40","3","meteo
3",Replacer.ReplaceText,{"Condition_meteo"}),

```

```

    #"Valeur      remplacée42"      =      Table.ReplaceValue("#Valeur
remplacée41", "4", "meteo
4", Replacer.ReplaceText, {"Condition_meteo"}),

    #"Colonnes      renommées5"      =      Table.RenameColumns("#Valeur
remplacée42", {"road_surface_conditions",
"Condition_surface_route"}),

    #"Type      modifié5"      =      Table.TransformColumnTypes("#Colonnes
renommées5", {"Condition_surface_route", type text}),

    #"Valeur      remplacée43"      =      Table.ReplaceValue("#Type
modifié5", "1", "surface
1", Replacer.ReplaceText, {"Condition_surface_route"}),

    #"Valeur      remplacée44"      =      Table.ReplaceValue("#Valeur
remplacée43", "0", "surface
0", Replacer.ReplaceText, {"Condition_surface_route"}),

    #"Valeur      remplacée45"      =      Table.ReplaceValue("#Valeur
remplacée44", "2", "surface
2", Replacer.ReplaceText, {"Condition_surface_route"}),

    #"Colonnes      renommées6"      =      Table.RenameColumns("#Valeur
remplacée45", {"transmission", "Transmission"}, {"frequence",
"accidents"}),

    #"Type      modifié6"      =      Table.TransformColumnTypes("#Colonnes
renommées6", {"accidents", type text}),

    #"Valeur      remplacée46"      =      Table.ReplaceValue("#Type
modifié6", "0", "Non", Replacer.ReplaceText, {"accidents"}),

    #"Valeur      remplacée47"      =      Table.ReplaceValue("#Valeur
remplacée46", "1", "Oui", Replacer.ReplaceText, {"accidents"}),

    #"Valeur      remplacée48"      =      Table.ReplaceValue("#Valeur
remplacée47", "2", "Oui", Replacer.ReplaceText, {"accidents"}),

    #"Valeur      remplacée49"      =      Table.ReplaceValue("#Valeur
remplacée48", "3", "Oui", Replacer.ReplaceText, {"accidents"}),

    #"Valeur      remplacée50"      =      Table.ReplaceValue("#Valeur
remplacée49", "4", "", Replacer.ReplaceText, {"accidents"}),

    #"Valeur      remplacée51"      =      Table.ReplaceValue("#Valeur
remplacée50", "", "Oui", Replacer.ReplaceValue, {"accidents"}),

    #"Valeur      remplacée52"      =      Table.ReplaceValue("#Valeur
remplacée51", "6", "Oui", Replacer.ReplaceText, {"accidents"}),

```

```

    #"Valeur remplacée53" = Table.ReplaceValue(#"Valeur
remplacée52", "7", "Oui", Replacer.ReplaceText, {"accidents"}),

    #"Valeur remplacée54" = Table.ReplaceValue(#"Valeur
remplacée53", "8", "Oui", Replacer.ReplaceText, {"accidents"}),

    #"Colonnes permutées1" = Table.ReorderColumns(#"Valeur
remplacée54", {"Age", "Sexe", "Vehicle_Age", "Marque", "Couleur",
"fuel_type", "nb_siege", "nb_porte", "Annee_fabric",
"Transmission", "Gravite_accident", "Trajet",
"Condition_eclairage", "Condition_meteo",
"Condition_surface_route", "manv", "Annual_Premium", "accidents",
"ACCIDENT"}),

    #"Colonnes supprimées" = Table.RemoveColumns(#"Colonnes
permutées1", {"ACCIDENT"}),

    #"Colonnes renommées7" = Table.RenameColumns(#"Colonnes
supprimées", {"accidents", "accident"}),

    #"Valeur remplacée55" = Table.ReplaceValue(#"Colonnes
renommées7", "5", "Oui", Replacer.ReplaceText, {"accident"}),

    #"Valeur remplacée56" = Table.ReplaceValue(#"Valeur
remplacée55", "3", "surface
3", Replacer.ReplaceText, {"Condition_surface_route"}),

    #"Valeur remplacée57" = Table.ReplaceValue(#"Valeur
remplacée56", "4", "surface
4", Replacer.ReplaceText, {"Condition_surface_route"})

in

    #"Valeur remplacée57"

```

- Visualisation des données

```

library(readxl)

assurance_auto_mak=read_excel("C:/Users/HP/Downloads/INSSSEDs/cours/
ECONOMETRIE/Mini projet économétrie Reg logic et poisson/assurance_auto_mak.xlsx")

str(assurance_auto_mak)

```

- Transformation de certaines variables en variables catégorielles

```

assurance_auto_mak$Sexe = as.factor(assurance_auto_mak$Sexe)

```

```

assurance_auto_mak$Vehicle_Age=as.factor(assurance_auto_mak$Vehicle_Age)

assurance_auto_mak$Marque = as.factor(assurance_auto_mak$Marque)

assurance_auto_mak$Couleur= as.factor(assurance_auto_mak$Couleur)

assurance_auto_mak$fuel_type=as.factor(assurance_auto_mak$fuel_type)

assurance_auto_mak$Transmission=as.factor(assurance_auto_mak$Transmission)

assurance_auto_mak$Gravite_accident=as.factor(assurance_auto_mak$Gravite_accident)

assurance_auto_mak$Trajet = as.factor(assurance_auto_mak$Trajet)

assurance_auto_mak$Condition_eclairage=as.factor(assurance_auto_mak$Condition_eclairage)

assurance_auto_mak$Condition_meteo=as.factor(assurance_auto_mak$Condition_meteo)

assurance_auto_mak$Condition_surface_route=as.factor(assurance_auto_mak$Condition_surface_route)

assurance_auto_mak$accident=as.factor(assurance_auto_mak$accident)

assurance_auto_mak$Annee_fabric=as.factor(assurance_auto_mak$Annee_fabric)

str(assurance_auto_mak)

```

- Résumé numérique

```
summary(assurance_auto_mak)
```

- Traitement des doublons

```

library(akposso)

assurance_auto_mak = traitement_doublons(assurance_auto_mak)

```

- Traitement des valeurs manquantes

```
colSums(is.na(assurance_auto_mak))
```

- Traitement des valeurs abberantes

- **Visualisation des valeurs abberantes avant traitement**

```
library(akposso)

afficher_boites_a_moustache(assurance_auto_mak)
```

- Visualisation des valeurs abberantes apres traitement

```
library(DescTools)

assurance_auto_mak$nb_siege =
Winsorize(assurance_auto_mak$nb_siege)

assurance_auto_mak$manv = Winsorize(assurance_auto_mak$manv)

assurance_auto_mak$Annual_Premium =
Winsorize(assurance_auto_mak$Annual_Premium)

library(akposso)

afficher_boites_a_moustache(assurance_auto_mak)
```

II- ANALYSE UNIVARIEE

II-1 Variables quantitatives

a) Variable Age

- Tableau

```
library(akposso)

head(akposso.qt.tableau(assurance_auto_mak$Age),5)
```

- Graphique

```
par(mfrow=c(1, 2))

hist(assurance_auto_mak$Age,main = 'histogramme des ages',xlab =
'Age',ylab =

'frequence',col='red',border = 'black')

boxplot(assurance_auto_mak$Age,main='boites à moustaches
ages',ylab = 'Age',col='blue')
```

- Resume numerique

```
library(akposso)

akposso.qt.resume(assurance_auto_mak$Age)
```

b) Variable nb_siege

- Tableau

```
library(akposso)

akposso.qt.tableau(assurance_auto_mak$nb_siege)
```

- Graphique

```
par(mfrow=c(1, 2))

hist(assurance_auto_mak$nb_siege,main = 'histogramme des
sièges',xlab = 'nb_siege',ylab =
'frequence',col='red',border = 'black')

boxplot(assurance_auto_mak$nb_siege,main='boites à moustaches
siege',ylab = 'nb_siege',col='blue')
```

- Resume numerique

```
akposso.qt.resume(assurance_auto_mak$nb_siege)
```

c) Variable nb_porte

- Tableau

```
akposso.qt.tableau(assurance_auto_mak$nb_porte)
```

- Graphique

```
par(mfrow=c(1, 2))

hist(assurance_auto_mak$nb_porte,main = 'histogramme des
portes',xlab = 'nb_portes',ylab =
'frequence',col='red',border = 'black')

boxplot(assurance_auto_mak$nb_porte,main='boites à moustaches
portes',ylab = 'nb_portes',col='blue')
```

- Resume numerique

```
akposso.qt.resume(assurance_auto_mak$nb_porte)
```

d) Variable annual_premium

- Tableau

```
head(akposso.qt.tableau(assurance_auto_mak$Annual_Premium),5)
```

- Graphique


```
par(mfrow=c(1, 2))  
  
hist(assurance_auto_mak$Annual_Premium,main = 'histogramme des  
primes',xlab =  
  
'Annual_prime',ylab = 'frequence',col='red',border = 'black')  
  
boxplot(assurance_auto_mak$nb_porte,main='boites à moustaches  
primes',ylab =  
  
'Annual_prime',col='blue')
```

- Resume numerique

```
akposso.qt.resume(assurance_auto_mak$Annual_Premium)
```

II-2 Variables qualitatives

a) Variable Sexe

- Tableau

```
akposso.q1.tableau(assurance_auto_mak$Sexe)
```

- Graphique

```
library(gridExtra)  
library(akposso)  
library(ggplot2)  
akposso.q1.graph(assurance_auto_mak$Sexe)
```

b) Variable vehicle_Age

- Tableau

```
akposso.q1.tableau(assurance_auto_mak$Vehicle_Age)
```

- Graphique

```
akposso.q1.graph(assurance_auto_mak$Vehicle_Age)
```

c) Variable Marque

- Tableau

```
akposso.q1.tableau(assurance_auto_mak$Marque)
```

- Graphique

```
marque_fabric_table = table(assurance_auto_mak$Marque)
# Créer le diagramme circulaire avec des marques distinctes
barplot(marque_fabric_table,
        main='Diagrammes en barres des Marques',
        col=colors,
        xlab = 'Marques')
```

d) Variable Couleur

- Tableau

```
akposso.q1.tableau(assurance_auto_mak$Couleur)
```

- Graphique

```
couleur_fabric_table = table(assurance_auto_mak$Couleur)
# Créer le diagramme circulaire avec des couleurs distinctes
barplot(couleur_fabric_table,
        main='Diagrammes barres des couleurs',
        col=colors,
        xlab = 'couleurs')
```

e) Variable fuel_type

- Tableau

```
akposso.q1.tableau(assurance_auto_mak$fuel_type)
```

- Graphique

```
akposso.q1.graph(assurance_auto_mak$fuel_type)
```

f) Variable Anne_fabric

- Tableau

```
head(akposso.q1.tableau(assurance_auto_mak$Annee_fabric
), 5)
```

- Gaphique

```
annee_fabric_table = table(assurance_auto_mak$Annee_fabric)
```

```
# Créer le diagramme circulaire avec des couleurs distinctes
barplot(annee_fabric_table,
        main='Diagramme en barre des Années de Fabrication',
        col=colors,
        xlab = 'Année de fabric')
```

g) Variable Transmission

- Tableau

```
akposso.ql.tableau(assurance_auto_mak$Transmission)
```

- Graphique

```
akposso.ql.graph(assurance_auto_mak$Transmission)
```

h) Variable Trajet

- Tableau

```
akposso.ql.tableau(assurance_auto_mak$Trajet)
```

- Graphique

```
Trajet_fabric_table = table(assurance_auto_mak$Trajet)
```

```
# Créer le diagramme circulaire avec des trajets distinctes
```

```
barplot(Trajet_fabric_table,
        main='Diagramme en barre des Trajets',
        col=colors,
        xlab = 'Trajets')
```

i) Variable conditions eclairage

- Tableau

```
akposso.ql.tableau(assurance_auto_mak$Condition_eclairage)
```

- Graphique

```
eclairage_fabric_table  
table(assurance_auto_mak$Condition_eclairage)
```

=

```
# Créer le diagramme circulaire avec des eclairages distinctes
```

```
barplot(eclairage_fabric_table,  
        main='Diagramme en barre des eclairages',  
        col=colors,  
        xlab = 'eclairage')
```

j) Variable condition meteo

- Tableau

```
akposso.ql.tableau(assurance_auto_mak$Condition_meteo)
```

- Graphique

```
meteo_fabric_table = table(assurance_auto_mak$Condition_meteo)  
# Créer le diagramme circulaire avec des meteo distinctes  
barplot(meteo_fabric_table,  
        main='Diagramme en barre des meteo',  
        col=colors,  
        xlab = 'meteo')
```

k) Variable condition_surface_route

-Tableau

```
akposso.ql.tableau(assurance_auto_mak$Condition_surface_route)
```

- Graphique

```
surface_table = table(assurance_auto_mak$Condition_surface_route)  
# Créer le diagramme circulaire avec des surfaces distinctes  
barplot(surface_table,  
        main='Diagramme en barre des surfaces',  
        col=colors,  
        xlab = 'surface')
```

l) Variable accident

- Tableau

```
akposso.ql.tableau(assurance_auto_mak$accident)
```

- Graphique

```
akposso.ql.graph(assurance_auto_mak$accident)
```

III- ANALYSE BIVARIEE

a) variable accident et Age

- Tableau

```
head(table(assurance_auto_mak$Age,assurance_auto_mak$accident),10)
```

- Graphique

```
boxplot(assurance_auto_mak$Age ~ assurance_auto_mak$accident,  
col=c("lightblue","lightsalmon"),main='boîtes à moustaches',xlab  
= 'accident',ylab = 'Age')
```

- Liaison

```
akposso.qtpl.liaison(assurance_auto_mak$Age,assurance_auto_mak$accident)
```

b) Variable accident et Sexe

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Sexe)
```

- Graphique

```
plot(table(assurance_auto_mak$accident,assurance_auto_mak$Sexe),  
main="GRAPHIQUE",col=c("lightblue","lightsalmon"),xlab =  
'accident',ylab = 'SEXE')
```

- Liaison

```
akposso.2ql.liaison(assurance_auto_mak$accident,assurance_auto_mak$Sexe)
```

c) Variable accident et Vehicle_Age

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Vehicle_Age)
```

- Graphique

```
plot(table(assurance_auto_mak$accident,assurance_auto_mak$Vehicle_Age),  
main="GRAPHIQUE",col=c("lightblue","lightsalmon"),xlab =  
'accident',ylab = 'Vehicle_Age')
```

- Liaison

```
akposso.2ql.liaison(assurance_auto_mak$accident,assurance_auto_mak$Vehicle_Age)
```

d) Variable accident et Marque

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Marque)
```

- Graphique

```
library(ggplot2)  
library(dplyr)  
# Calculer les fréquences des combinaisons  
df_freq <- assurance_auto_mak %>%  
  group_by(accident, Marque) %>%  
  summarise(freq = n()) %>%  
  ungroup()  
  
# Créer la heatmap  
ggplot(df_freq, aes(x = accident, y = Marque, fill = freq)) +  
  geom_tile() +  
  xlab("Accident") +  
  ylab("Marque") +  
  ggtitle("Heatmap des Accidents par Marque") +  
  scale_fill_gradient(low = "white", high = "red")
```

e) Variable accident et Couleur

- Tableau

```
# Créer le tableau de contingence  
tableau_contingence=table(assurance_auto_mak$accident,assurance_a  
uto_mak$Couleur)  
  
# Afficher le tableau de contingence  
print(tableau_contingence)
```

- Graphique

```
df_freq1 <- assurance_auto_mak %>%  
  group_by(accident, Couleur) %>%  
  summarise(freq = n()) %>%  
  ungroup()  
  
# Créer la heatmap  
ggplot(df_freq1, aes(x = accident, y = Couleur, fill = freq)) +  
  geom_tile() +  
  xlab("Accident") +  
  ylab("Couleur") +  
  ggtitle("Heatmap des Accidents par Couleur") +  
  scale_fill_gradient(low = "white", high = "red")
```

f) Variable accident et fuel_type

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_ma  
k$fuel_type)
```

- Graphique

```
akposso.2ql.graph(assurance_auto_mak$accident,assurance_auto_mak$  
fuel_type)
```

- Liaison

```
akposso.2ql.liaison(assurance_auto_mak$accident,assurance_auto_ma  
k$fuel_type)
```

g) Variable accident et annee_fabric

- Tableau

```
# Créer le tableau de contingence  
tableau_contingence1=table(assurance_auto_mak$accident,assurance_  
auto_mak$Annee_fabric)  
  
# Afficher le tableau de contingence  
print(tableau_contingence1)
```

- Graphique

```
df_freq2 <- assurance_auto_mak %>%  
  group_by(accident, Annee_fabric) %>%  
  summarise(freq = n()) %>%  
  ungroup()  
  
# Créer la heatmap  
ggplot(df_freq2, aes(x = accident, y = Annee_fabric, fill =  
freq)) +  
  
  geom_tile() +  
  xlab("Accident") +  
  ylab("Annee_fabric") +  
  ggtitle("Heatmap des Accidents par Annee_fabric") +  
  scale_fill_gradient(low = "white", high = "red")
```

h) Variable accident et Transmission

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_ma  
k$Transmission)
```

- Graphique

```
akposso.2ql.graph(assurance_auto_mak$accident,assurance_auto_mak$  
Transmission)
```

- Liaison


```
akposso.2ql.liaison(assurance_auto_mak$accident,assurance_auto_mak$Transmission)
```

i) Variable accident et Trajet

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Trajet)
```

- Graphique

```
df_freq3 <- assurance_auto_mak %>%  
  group_by(accident, Trajet) %>%  
  summarise(freq = n()) %>%  
  ungroup()  
  
# Créer la heatmap  
ggplot(df_freq3, aes(x = accident, y = Trajet, fill = freq)) +  
  geom_tile() +  
  xlab("Accident") +  
  ylab("Trajet") +  
  ggtitle("Heatmap des Accidents par Trajet") +  
  scale_fill_gradient(low = "white", high = "red")+  
  theme_minimal()
```

j) Variable accident et Conditions_eclairage

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Condition_eclairage)
```

- Graphique

```
df_freq4 <- assurance_auto_mak %>%  
  group_by(accident, Condition_eclairage) %>%  
  summarise(freq = n()) %>%  
  ungroup()
```

```
# Créer la heatmap

ggplot(df_freq4, aes(x = accident, y = Condition_eclairage, fill
= freq)) +

  geom_tile() +

  xlab("Accident") +

  ylab("Condition_eclairage") +

  ggtitle("Heatmap des Accidents par Condition_eclairage") +

  scale_fill_gradient(low = "white", high = "red")+

  theme_minimal()
```

k) Variable accident et condition_meteo

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_ma
k$Condition_meteo)
```

- Graphique

```
df_freq5 <- assurance_auto_mak %>%
  group_by(accident, Condition_meteo) %>%
  summarise(freq = n()) %>%
  ungroup()

# Créer la heatmap

ggplot(df_freq5, aes(x = accident, y = Condition_meteo, fill =
freq)) +

  geom_tile() +

  xlab("Accident") +

  ylab("Condition_meteo") +

  ggtitle("Heatmap des Accidents par meteo") +

  scale_fill_gradient(low = "white", high = "red")+

  theme_minimal()
```

l) Variable accident et condition_surface_route

- Tableau

```
akposso.2ql.tableau(assurance_auto_mak$accident,assurance_auto_mak$Condition_surface_route)
```

- Graphique

```
df_freq6 <- assurance_auto_mak %>%  
  group_by(accident, Condition_surface_route) %>%  
  summarise(freq = n()) %>%  
  ungroup()  
  
# Créer la heatmap  
ggplot(df_freq6, aes(x = accident, y = Condition_surface_route,  
  fill = freq)) +  
  geom_tile() +  
  xlab("Accident") +  
  ylab("Condition_surface_route") +  
  ggtitle("Heatmap des Accidents par surface") +  
  scale_fill_gradient(low = "white", high = "red")+  
  theme_minimal()
```

m) Variable accident et manv

- Tableau

```
head(table(assurance_auto_mak$manv,assurance_auto_mak$accident),5  
)
```

- Graphique

```
boxplot(assurance_auto_mak$manv ~ assurance_auto_mak$accident,  
  col=c("lightblue","lightsalmon"),main='boîtes à moustaches',xlab  
  = 'accident',ylab = 'manv')
```

- Liaison

```
akposso.qtql.liaison(assurance_auto_mak$manv,assurance_auto_mak$accident)
```

n) Variable accident et Annual_premium

- Tableau

```
head(table(assurance_auto_mak$Annual_Premium,assurance_auto_mak$accident),5)
```

- Graphique

```
boxplot(assurance_auto_mak$Annual_Premium~assurance_auto_mak$accident,col=c("lightblue","lightsalmon"),main='boites à moustaches',xlab = 'accident',ylab = 'prime annuelle')
```

- Liaison

```
akposso.qtpl.liaison(assurance_auto_mak$Annual_Premium,assurance_auto_mak$accident)
```

IV- ESTIMATION DE LA PROPORTION DES ACCIDENTES

- Visualisation des proportions

```
tab<-table(assurance_auto_mak$accident)
round(prop.table(tab),2)
```

- Conditions

```
tab
```

- Estimation

```
round(binom.test(237663,374257)$conf.int,2)
```

V- TEST DE CONFORMITE

- Test

```
binom.test(237663,374257,p=0.63)
binom.test(237663,374257,p=0.63,alternative = "greater")
```

VI- TEST DE LIAISON

a) Variable accident et Age

- Graphique

```
boxplot(assurance_auto_mak$Age ~ assurance_auto_mak$accident,
col=c("lightblue","lightsalmon"),main='boites à moustaches',xlab = 'accident',ylab = 'Age')
```

- Normalité et homoscedasticité

```
# Chargement de vos données  
# supposez que assurance_auto_mak est déjà chargé  
# Test de Kolmogorov-Smirnov  
ks.test(assurance_auto_mak$Age, "pnorm", mean=mean(assurance_auto_mak$Age), sd=sd(assurance_auto_mak$Age))  
  
library(car)  
  
leveneTest(assurance_auto_mak$Age~assurance_auto_mak$accident, assurance_auto_mak)
```

- Test

```
kruskal.test(assurance_auto_mak$Age~assurance_auto_mak$accident)
```

b) Variable accident et Sexe

- Condition

```
tab = table(assurance_auto_mak$Sexe, assurance_auto_mak$accident)
```

```
chisq.test(tab)$expected
```

- Test

```
chisq.test(tab)
```

c) Variable accident et Vehicle_Age

- Condition

```
tab1=table(assurance_auto_mak$Vehicle_Age, assurance_auto_mak$accident)
```

```
chisq.test(tab1)$expected
```

- Test

```
chisq.test(tab1)
```

d) Variable accident et Marque

- Condition

```
tab2=table(assurance_auto_mak$Marque, assurance_auto_mak$accident)
```

```
chisq.test(tab2)$expected
```

- Test

chisq.test(tab2)

e) Variable accident et Couleur

- Condition

```
tab3=table(assurance_auto_mak$Couleur,assurance_auto_mak$accident  
)
```

```
chisq.test(tab3)$expected
```

- Test

```
chisq.test(tab3)
```

f) Variable accident et fuel_type

- Condition

```
tab4=table(assurance_auto_mak$fuel_type,assurance_auto_mak$accide  
nt)
```

```
chisq.test(tab4)$expected
```

- Test

```
chisq.test(tab4)
```

g) Variable accident et nb_siege

- Graphique

```
boxplot(assurance_auto_mak$nb_siege ~assurance_auto_mak$accident,  
col=c("lightblue","lightsalmon"),main='boites à moustaches',xlab  
= 'accident',ylab = 'nb_siege')
```

- Normalité et homoscedasticité

```
# Chargement de vos données
```

```
# supposez que assurance_auto_mak est déjà chargé
```

```
# Test de Kolmogorov-Smirnov
```

```
ks.test(assurance_auto_mak$nb_siege,"pnorm",mean=mean(assurance_a  
uto_mak$nb_siege), sd=sd(assurance_auto_mak$nb_siege))
```

```
leveneTest(assurance_auto_mak$nb_siege~assurance_auto_mak$accident)
```

- Test

```
kruskal.test(assurance_auto_mak$nb_siege~assurance_auto_mak$accident)
```

h) Variable accident et nb_porte

- Graphique

```
boxplot(assurance_auto_mak$nb_porte~assurance_auto_mak$accident,  
col=c("lightblue","lightsalmon"),main='boîtes à moustaches',xlab  
= 'accident',ylab = 'nb_porte')
```

- Normalité et homoscedasticité

```
# Chargement de vos données
```

```
# supposez que assurance_auto_mak est déjà chargé
```

```
# Test de Kolmogorov-Smirnov
```

```
ks.test(assurance_auto_mak$nb_porte,"pnorm",mean=mean(assurance_a  
uto_mak$nb_porte), sd=sd(assurance_auto_mak$nb_porte))
```

```
library(car)
```

```
leveneTest(assurance_auto_mak$nb_porte~assurance_auto_mak$accident)
```

- Test

```
kruskal.test(assurance_auto_mak$nb_porte~assurance_auto_mak$accident)
```

i) Variable accident et Annee fabric

- Condition

```
tab5=table(assurance_auto_mak$Annee_fabric,assurance_auto_mak$accident)
```

```
chisq.test(tab5)$expected
```

```
# Créez votre tableau de contingence
```

```
tab5=table(assurance_auto_mak$Annee_fabric,assurance_auto_mak$accident)
```

```

# Effectuer le test du chi-carré
chi_result <- chisq.test(tab5)
# Obtenez les fréquences attendues
expected_frequencies <- chi_result$expected
# Comptez le nombre de cellules attendues avec une fréquence >= 5
num_cells_valid <- sum(expected_frequencies >= 5)
# Calculez le pourcentage de cellules qui respectent la condition
de Cochran
percentage_valid<-(num_cells_valid/length(expected_frequencies))
* 100
# Affichez le résultat
print("le pourcentage effets theorique superieur à 5 est :")
percentage_valid

```

- Test

```
chisq.test(tab5)
```

j) Variable accident et Transmission

- Condition

```
tab6=table(assurance_auto_mak$Transmission,assurance_auto_mak$accident)
```

```
chisq.test(tab6)$expected
```

- Test

```
chisq.test(tab6)
```

k) Variable accident et Gravité_accident

- Condition

```
tab11=table(assurance_auto_mak$Gravite_accident,assurance_auto_mak$accident)
```

```
chisq.test(tab6)$expected
```

- Test


```
chisq.test(tab11)
```

l) Variable accident et Trajet

- Condition

```
tab7=table(assurance_auto_mak$Trajet,assurance_auto_mak$accident)  
chisq.test(tab7)$expected
```

- Test

```
chisq.test(tab6)
```

m) Variable accident et Condition eclairage

- Condition

```
tab8=table(assurance_auto_mak$Condition_eclairage,assurance_auto_mak$accident)  
chisq.test(tab8)$expected
```

- Test

```
chisq.test(tab7)
```

n) Variable accident et Condition meteo

- Condition

```
tab9=table(assurance_auto_mak$Condition_meteo,assurance_auto_mak$accident)  
chisq.test(tab9)$expected
```

- Test

```
chisq.test(tab9)
```

o) Variable accident et Condition surface route

- Condition

```
tab10=table(assurance_auto_mak$Condition_surface_route,assurance_auto_mak$accident)  
chisq.test(tab10)$expected
```

- Test

```
chisq.test(tab10)
```

p) Variable accident et manv

- Graphique

```
boxplot(assurance_auto_mak$manv~assurance_auto_mak$accident,col=c("lightblue","lightsalmon"),main='boîtes à moustaches',xlab = 'accident',ylab = 'manv')
```

- Normalité et homoscedasticité

```
# Chargement de vos données
```

```
# supposez que assurance_auto_mak est déjà chargé
```

```
# Test de Kolmogorov-Smirnov
```

```
ks.test(assurance_auto_mak$manv,"pnorm",mean=mean(assurance_auto_mak$manv), sd=sd(assurance_auto_mak$manv))
```

```
library(car)
```

```
leveneTest(assurance_auto_mak$manv~assurance_auto_mak$accident)
```

- Test

```
kruskal.test(assurance_auto_mak$manv~assurance_auto_mak$accident)
```

q) Variable accident et Annual premium

- Graphique

```
boxplot(assurance_auto_mak$Annual_Premium~assurance_auto_mak$accident,col=c("lightblue","lightsalmon"),main='boîtes à moustaches',xlab = 'accident',ylab = 'Annual premium')
```

- Normalité et homoscedasticité

```
# Chargement de vos données
```

```
# supposez que assurance_auto_mak est déjà chargé
```

```
# Test de Kolmogorov-Smirnov
```

```
ks.test(assurance_auto_mak$Annual_Premium,"pnorm",mean=mean(assurance_auto_mak$Annual_Premium),sd=sd(assurance_auto_mak$Annual_Premium))
```

```
library(car)
```

```
leveneTest(assurance_auto_mak$Annual_Premium~assurance_auto_mak$accident)
```

- Test

```
kruskal.test(assurance_auto_mak$Annual_Premium~assurance_auto_mak$accident)
```

V- REGRESSION LOGISTIQUE

- Anova

```
log = glm(accident~.,data = assurance_auto_mak,family = binomial)
print(anova(log,test="Chisq"))
```

- Estimation des parametres

```
# Ajuster le modèle de régression logistique

log_l <- glm(accident ~ Age + Sexe + Vehicle_Age + Marque +
Couleur + fuel_type + nb_siege + nb_porte + Transmission +
Gravite_accident, data = assurance_auto_mak, family = binomial)

# Résumé du modèle

summary(log_l)
```

- ODDRATIO

```
oddratio = exp(log_l$coefficients)

oddratio
```

- Effets marginaux

```
library(mfx)

logitmfx(accident ~ Age + Sexe + Vehicle_Age + Marque + Couleur +
fuel_type + nb_siege + nb_porte+ Transmission + Gravite_accident,
data = assurance_auto_mak)
```

- Taux de mauvais classement

```
# Ajoute la prédiction à l'ensemble de données assurance_auto_mak

assurance_auto_mak$prediction<-predict(log_l, assurance_auto_mak,
type = "response")

# Définir le seuil
```

```
assurance_auto_mak$predicted_class <-
ifelse(assurance_auto_mak$prediction > 0.5, 1, 0)

# Calculer le taux de mauvaise classification

taux_mauvais_classement <-
mean(assurance_auto_mak$predicted_class
assurance_auto_mak$accident)

# Calculer le taux de mauvaise classification en pourcentage

taux_mauvais_classement_pourcentage <- taux_mauvais_classement *
100

taux_mauvais_classement_pourcentage
```

V- REGRESSION POISSON

- Importation du jeu de données

```
library(readxl)

assurance_poisson <-
read_excel("C:/Users/HP/Downloads/INSSEDS/cours/ECONOMETRIE/Mini
projet économétrie Reg logic et poisson/assurance_poisson.xlsx")

str(assurance_poisson)
```

- Graphique

```
plot(table(assurance_poisson$frequence),ylab="frequence")
```

moyenne

```
mean(assurance_poisson$frequence)
```

Variance

```
var(assurance_poisson$frequence)
```

- Test d'adéquation à la loi de poisson

```
library(fitdistrplus)

fpois <- fitdist(assurance_poisson$frequence, "pois")

gofstat(fpois)
```

- ESTIMATION DU MODELE

```
reg=glm(frequence~.,data=assurance_poisson,family = quasipoisson)
```

```
summary(reg)
```

- Test d'anova permet de voir les variables les plus liées à la cible

```
anova(reg, test="Chisq")
```

- reprenons le modèle avec ces variables

```
reg=glm(frequence~Age+Sexe+Vehicle_Age+Marque+Couleur+fuel_type+n  
b_siege+nb_porte+Annee_fabric+Transmission+Gravite_accident+Traje  
t+Condition_eclairage+manv+Annual_Premium,data=assurance_poisson,  
family=quasipoisson)
```

```
summary(reg)
```

- Anova

```
anova(reg, test="Chisq")
```

- Prenons l'exponentiel des coefficients

```
exp(reg$coefficients)
```

- ANALYSE DES RESIDUS

Graphique

```
res.m<-rstudent(reg)  
plot(res.m, pch=15, cex=.5, ylab="Residus", ylim=c(-3,3))  
abline(h=c(-2,0,2), lty=c(2,1,2))
```

- Test de normalité

Graphique

```
qqnorm(res.m)  
qqline(res.m, col="red")
```

Test

```
ks.test(res.m, "pnorm", mean(res.m), sd(res.m))
```

- FAIRE LA PREVISION

Prédire la frequence d'accident des individus 10,11et12:

```
predict.glm(reg, assurance_poisson[10,], type = "response")
```

```
predict.glm(reg, assurance_poisson[11,], type = "response")
predict.glm(reg, assurance_poisson[12,], type = "response")
```

Intervalle de confiance

```
lamb<-predict.glm(reg,assurance_poisson[10,],type = "response")
loglamb<-predict.glm(reg, assurance_poisson[10,],se.fit = TRUE)
icloglamb<-c(loglamb$fit-1.96*loglamb$se.fit,loglamb$fit + 1.96 *
loglamb$se.fit)
ic<-exp(icloglamb)
ic
lamb1<-predict.glm(reg,assurance_poisson[11,],type = "response")
loglamb1<-predict.glm(reg,assurance_poisson[11,],se.fit = TRUE)
icloglamb1<-c(loglamb1$fit-1.96*loglamb1$se.fit,loglamb1$fit+1.96
*loglamb1$se.fit)
ic1 <- exp(icloglamb1)
ic1
lamb2<-predict.glm(reg,assurance_poisson[12,],type= "response")
loglamb2<-predict.glm(reg,assurance_poisson[12,],se.fit = TRUE)
icloglamb2<-c(loglamb2$fit-1.96*loglamb2$se.fit, loglamb2$fit +
1.96 * loglamb2$se.fit)
ic2<-exp(icloglamb2)
ic2
```