

Code ▼

Progetto Modelli Statistici Per Le Scienze Attuariali

ANALISI DATASET

step1. aprire il dataset

visualizzo il dataset:

Hide

```
str(ds)
```

```
'data.frame':  393071 obs. of  12 variables:
 $ Gender      : Factor w/ 3 levels "Corporate","Female",...: 2 3 3 3 2 3 2 2 1 2 ...
 $ DrivAge     : Factor w/ 5 levels ">55","18-25",...: 4 1 5 1 5 2 3 1 NA 2 ...
 $ VehYear    : int   2011 2007 2006 2010 2009 2007 2006 2010 2009 2009 ...
 $ VehModel   : Factor w/ 4259 levels "Acura - Legend 3.2/3.5",...: 3996 2869 501 696 3095 290
8 1385 625 1236 1162 ...
 $ VehGroup   : Factor w/ 436 levels "Acura","Agrale - Marrua",...: 417 324 67 100 354 327 170
90 147 143 ...
 $ Area       : Factor w/ 40 levels "Acre","Alagoas",...: 21 8 11 11 6 8 25 30 22 9 ...
 $ State      : Factor w/ 27 levels "Acre","Alagoas",...: 19 6 16 16 24 6 24 25 24 16 ...
 $ StateAb    : Factor w/ 27 levels "AC","AL","AM",...: 19 6 18 18 24 6 24 26 24 18 ...
 $ ExposTotal : num   0.98 1.22 0.12 0 10.48 ...
 $ SumInsAvg  : num  28393 27555 27801 0 30186 ...
 $ ClaimNb    : num   0 0 0 2 1 0 1 0 0 0 ...
 $ ClaimAmount: num   0 0 0 4018 1177 ...
```

verifica se ci sono osservazioni = 0 visto che non comportano info utile

Hide

```
summary(ds$ExposTotal)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.400	0.530	3.212	1.800	2346.500

convertendo i valori in scala giornaliera eliminale esposizioni pari a 0, verifica se l'osservazione minima è != 0

Hide

```
summary(ds$ExposTotal)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000008	0.001205	0.001589	0.009037	0.005151	6.428767

Hide

```
nrow(ds)
```

```
[1] 382693
```

Hide

```
summary(ds$VehYear)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1881	2003	2007	2005	2009	2012

Hide

```
summary(ds$SumInsAvg)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	19627	30477	40474	48959	125000

discretizza la variabile VehYear utilizzando il criterio dei quartili

Hide

```
levels(ds$VehYear)
```

```
[1] "< 2003" "2003-2007" "2007-2009" ">2009"
```

rimuovo na

Hide

```
summary(ds)
```

Gender	DrivAge	VehYear			
VehModel					
Corporate: 30861	>55 :71527	< 2003 :95112	Outros		
: 3681					
Female :118650	18-25:23585	2003-2007:61938	Fiat - Uno Mille 1.0 Fire/ F.flex/ Econom		
y 4p: 1486					
Male :158868	26-35:61938	2007-2009:78533	Ford - Ecosport Xls 1.6/ 1.6 Flex 8v 5p		
: 1337					
	36-45:78533	>2009 :72796	Toyota - Corolla Xei 1.8/1.8 Flex 16v Au		
t. : 1334					
	46-55:72796		Ford - Ecosport Xlt 1.6/ 1.6 Flex 8v 5p		
: 1286					
			Fiat - Uno Mille 1.0 Fire/ F.flex/ Econom		
y 2p: 1210			(Other)		
:298045					
VehGroup		Area			
State					
Vw Volkswagen Gol 1.0 : 8877	Met. de Sao Paulo	: 23804	Sao Paulo		
:70726					
Fiat Palio 1.0 : 8617	Demais regioes	: 23557	Minas Gera		
is :31861					
Gm Chevrolet Corsa 1.0: 7778	Ribeirao Preto e Demais Mun. de Campinas:	19059	Parana		
:31127					
Ford Eco Sport : 6914	Met. do Rio de Janeiro	: 14632	Santa Cata		
rina :28073					
Gm Chevrolet S-10 : 6808	Met. Porto Alegre e Caxias do Sul	: 14315	Rio Grande		
do Sul:27020					
Peugeot 206 : 6551	Met. Curitiba	: 13677	Rio de Jan		
eiro :22879					
(Other) :262834	(Other)	:199335	(Other)		
:96693					
StateAb	ExposTotal	SumInsAvg	ClaimNb	ClaimAmount	
SP :70726	Min. :0.000008	Min. : 0	Min. : 0.0000	Min. : 0	
MG :31861	1st Qu.:0.001260	1st Qu.: 18975	1st Qu.: 0.0000	1st Qu.: 0	
PR :31127	Median :0.001753	Median : 28953	Median : 0.0000	Median : 0	
SC :28073	Mean :0.009648	Mean : 38001	Mean : 0.2293	Mean : 1280	
RS :27020	3rd Qu.:0.005507	3rd Qu.: 45597	3rd Qu.: 0.0000	3rd Qu.: 0	
RJ :22879	Max. :6.428767	Max. :953514	Max. :213.0000	Max. :493800	
(Other):96693					

CREAZIONE CELLE TARIFFARIE

convertiamo le colonne 'ClaimNb' e 'ClaimAmount' del dataset in interi, ottenendo una migliore struttura del dataset

[Hide](#)

```
str(ds)
```

```
'data.frame':  308379 obs. of  12 variables:
 $ Gender      : Factor w/ 3 levels "Corporate","Female",...: 2 3 3 2 3 2 2 2 3 2 ...
 $ DrivAge     : Factor w/ 5 levels ">55","18-25",...: 4 1 5 5 2 3 1 2 1 4 ...
 $ VehYear     : Factor w/ 4 levels "< 2003","2003-2007",...: 3 1 4 4 1 2 1 1 1 3 ...
 $ VehModel    : Factor w/ 4259 levels "Acura - Legend 3.2/3.5",...: 3996 2869 501 3095 2908 13
85 625 1162 1372 3546 ...
 $ VehGroup    : Factor w/ 436 levels "Acura","Agrale - Marrua",...: 417 324 67 354 327 170 90
143 170 388 ...
 $ Area        : Factor w/ 40 levels "Acre","Alagoas",...: 21 8 11 6 8 25 30 9 5 2 ...
 $ State       : Factor w/ 27 levels "Acre","Alagoas",...: 19 6 16 24 6 24 25 16 5 2 ...
 $ StateAb     : Factor w/ 27 levels "AC","AL","AM",...: 19 6 18 24 6 24 26 18 5 2 ...
 $ ExposTotal  : num  0.002685 0.003342 0.000329 0.028712 0.000548 ...
 $ SumInsAvg   : num  28393 27555 27801 30186 35401 ...
 $ ClaimNb     : int  0 0 0 1 0 1 0 0 0 0 ...
 $ ClaimAmount: int  0 0 0 1177 0 1362 0 0 0 0 ...
- attr(*, "na.action")= 'omit' Named int [1:74314] 8 12 14 17 19 29 33 34 39 50 ...
..- attr(*, "names")= chr [1:74314] "1619759" "1737752" "1568700" "1933063" ...
```

Hide

```
list(
  Gender_Levels = levels(ds$Gender),
  DrivAge_Levels = levels(ds$DrivAge),
  VehModel_Levels = head(levels(ds$VehModel)),
  DrivAge_Summary = summary(ds$DrivAge)
) #list usato per printare tutto assieme
```

```
$Gender_Levels
[1] "Corporate" "Female"      "Male"

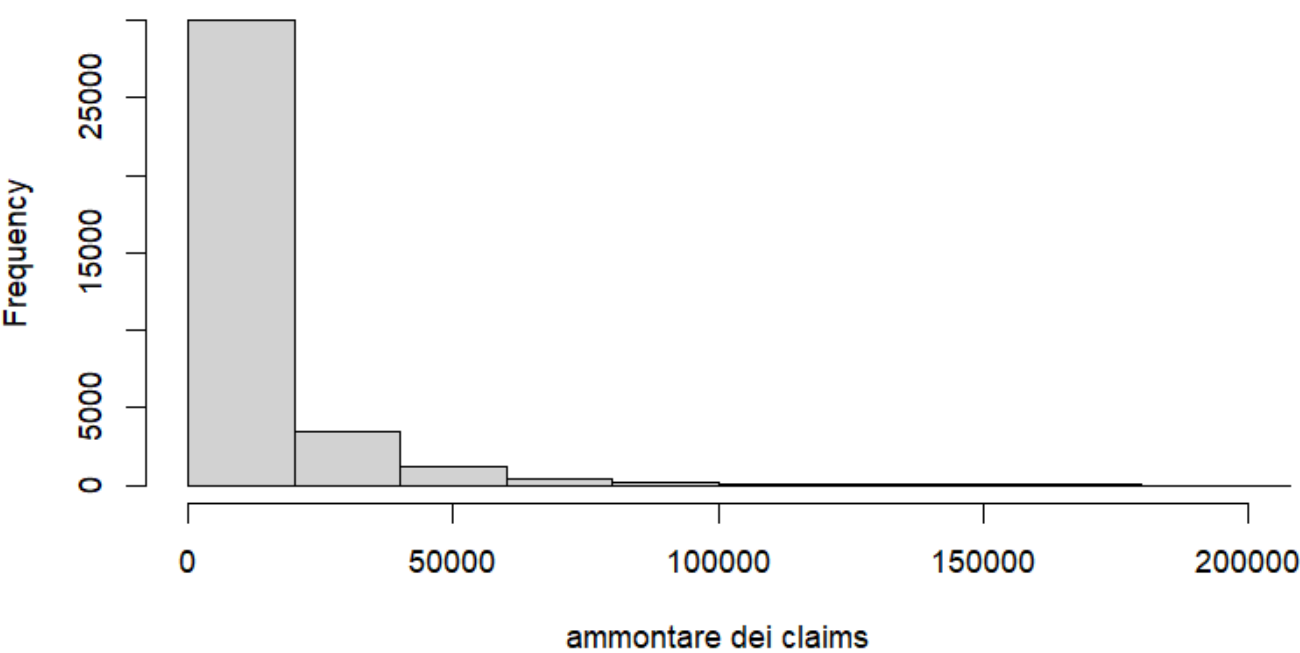
$DrivAge_Levels
[1] ">55"      "18-25" "26-35" "36-45" "46-55"

$VehModel_Levels
[1] "Acura - Legend 3.2/3.5"          "Acura - Nsx 3.0"
[3] "Agrale - 13000 Turbo 2p (diesel)" "Agrale - 13000 Turbo 3-eixos 2p (diesel)"
[5] "Agrale - 1600 D-rd 2p (diesel)"  "Agrale - 1600 D-rs 2p (diesel)"

$DrivAge_Summary
  >55 18-25 26-35 36-45 46-55
71527 23585 61938 78533 72796
```

Istogramma della variabile 'ClaimAmount' per valori maggiori di 0

Istogramma della variabile 'ClaimAmount'



numero totale di sinistri, l'esposizione totale per riuscire a trovare la frequenza dei sinistri

Hide

Freq

[1] 23.76307

aggregato il numero di sinistri per area

Area<fctr>	ClaimNb<int>
Acre	108
Alagoas	557
Amapa	56
Amazonas	354
Bahia	2318
Blumenau e demais regioes	1846
Brasilia	2739
Ceara	1521
Demais regioes	4325
Espirito Santo	1282
1-10 of 40 rows	
Previous1234Next	

tabelle per il numero di sinistri e per l'esposizione totale per area,

attraverso tabFreq -> calcolato la frequenza dei sinistri per area

[Hide](#)

tabFreq

```
[1] 42.12380 27.80555 34.73429 32.63643 22.61070 25.60309 31.32281 30.00638 25.62964 29.2621
1 34.16323 32.04994
[13] 18.14898 20.51431 22.37937 33.43668 38.06377 39.22224 26.88333 17.65129 14.70588 28.1087
5 23.43448 26.30141
[25] 21.33720 27.31729 29.89599 20.59683 36.87276 27.01009 32.20455 43.13469 19.85097 37.5081
2 32.52679 24.26934
[37] 34.27092 34.58284 30.40732 21.43101
```

Creazione delle celle tariffarie aggregando numero di sinistri ed esposizione totale per area celle + verifica presenza NA -> se non ci sono NA nel dataframe -> dati completi.

[Hide](#)

```
if (length(na_positions) > 0) {
  print(paste("NA found alla posizione:", toString(na_positions)))
} else {
  print("No NA found.")
}
```

```
[1] "No NA found."
```

Calcolo della frequenza dei sinistri per le celle tariffarie

Area <fctr>	ClaimNb <dbl>	ExposTotal <dbl>	Freq <dbl>
Acre	108	2.5638713	42.12380
Alagoas	557	20.0319700	27.80555
Amapa	56	1.6122394	34.73429
Amazonas	354	10.8467739	32.63643
Bahia	2318	102.5178307	22.61070
Blumenau e demais regioes	1846	72.1006740	25.60309
Brasilia	2739	87.4442698	31.32281
Ceara	1521	50.6892254	30.00638
Demais regioes	4325	168.7499452	25.62964
Espirito Santo	1282	43.8109142	29.26211
1-10 of 40 rows	Previous	1 2 3 4	Next

analisi di regressione sul numero di risarcimenti e quindi, sul numero di sinistri, utilizzando le variabili Gender+DrivAge+State

[Hide](#)

```
model_1 <- glm(ClaimNb ~ Gender+DrivAge+State+offset(log(ExposTotal)), family = poisson(link  
="log"),  
              data = ds, subset=ExposTotal>0, x = TRUE)  
model_1 <- glm(ClaimNb ~ Gender+DrivAge+State+offset(log(ExposTotal)), family = poisson(link  
="log"),  
              data = ds, subset=ExposTotal>0, x = TRUE)  
summary(model_1)
```

Call:

```
glm(formula = ClaimNb ~ Gender + DrivAge + State + offset(log(ExposTotal)),
     family = poisson(link = "log"), data = ds, subset = ExposTotal >
     0, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-12.37183	39.22792	-0.315	0.752471
GenderFemale	15.86536	39.22780	0.404	0.685888
GenderMale	15.91923	39.22780	0.406	0.684878
DrivAge18-25	0.59464	0.01862	31.933	< 2e-16 ***
DrivAge26-35	0.23273	0.01208	19.272	< 2e-16 ***
DrivAge36-45	0.35193	0.01086	32.408	< 2e-16 ***
DrivAge46-55	0.15206	0.01205	12.623	< 2e-16 ***
StateAlagoas	-0.41925	0.10515	-3.987	6.69e-05 ***
StateAmapa	-0.18272	0.16467	-1.110	0.267172
StateAmazonas	-0.24168	0.10993	-2.199	0.027913 *
StateBahia	-0.59874	0.09845	-6.082	1.19e-09 ***
StateCeara	-0.31840	0.09959	-3.197	0.001388 **
StateDistrito Federal	-0.28074	0.09811	-2.862	0.004215 **
StateEsperito Santo	-0.33499	0.10021	-3.343	0.000829 ***
StateGoias	-0.26501	0.09812	-2.701	0.006918 **
StateMaranhao	-0.21599	0.10659	-2.026	0.042726 *
StateMato Grosso	-0.12126	0.10056	-1.206	0.227891
StateMato Grosso do Sul	-0.07278	0.10075	-0.722	0.470039
StateMinas Gerais	-0.37803	0.09698	-3.898	9.69e-05 ***
StatePara	-0.42412	0.10366	-4.091	4.29e-05 ***
StateParaiba	-0.32448	0.10402	-3.119	0.001812 **
StateParana	-0.39675	0.09711	-4.086	4.39e-05 ***
StatePernambuco	-0.70145	0.09942	-7.055	1.72e-12 ***
StatePiaui	-0.09067	0.10871	-0.834	0.404290
StateRio de Janeiro	-0.93561	0.09743	-9.603	< 2e-16 ***
StateRio Grande do Norte	-0.27085	0.10239	-2.645	0.008164 **
StateRio Grande do Sul	-0.55268	0.09723	-5.684	1.31e-08 ***
StateRondonia	0.03024	0.18032	0.168	0.866799
StateRoraima	-0.76349	0.20376	-3.747	0.000179 ***
StateSanta Catarina	-0.45934	0.09761	-4.706	2.53e-06 ***
StateSao Paulo	-0.68630	0.09644	-7.117	1.11e-12 ***
StateSergipe	-0.11330	0.10442	-1.085	0.277881
StateTocantins	-0.20918	0.11328	-1.847	0.064805 .

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136071 on 308378 degrees of freedom

Residual deviance: 127539 on 308346 degrees of freedom

AIC: 211580

Number of Fisher Scoring iterations: 17

variabile Gender è poco significativa

procedere con altro modello senza var Gender


```
summary(model_2)
```

Call:

```
glm(formula = ClaimNb ~ DrivAge + State + offset(log(ExposTotal)),
     family = poisson(link = "log"), data = ds, subset = ExposTotal >
     0, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.506679	0.096570	36.312	< 2e-16 ***
DrivAge18-25	0.568716	0.018581	30.608	< 2e-16 ***
DrivAge26-35	0.219596	0.012007	18.290	< 2e-16 ***
DrivAge36-45	0.342111	0.010821	31.616	< 2e-16 ***
DrivAge46-55	0.136923	0.012020	11.391	< 2e-16 ***
StateAlagoas	-0.394917	0.105145	-3.756	0.000173 ***
StateAmapa	-0.190576	0.164672	-1.157	0.247147
StateAmazonas	-0.245248	0.109931	-2.231	0.025686 *
StateBahia	-0.592137	0.098446	-6.015	1.80e-09 ***
StateCeara	-0.307606	0.099587	-3.089	0.002010 **
StateDistrito Federal	-0.269377	0.098107	-2.746	0.006037 **
StateEsperito Santo	-0.333175	0.100203	-3.325	0.000884 ***
StateGoias	-0.254278	0.098123	-2.591	0.009558 **
StateMaranhao	-0.205611	0.106588	-1.929	0.053727 .
StateMato Grosso	-0.107698	0.100563	-1.071	0.284192
StateMato Grosso do Sul	-0.063164	0.100751	-0.627	0.530702
StateMinas Gerais	-0.387855	0.096975	-4.000	6.35e-05 ***
StatePara	-0.425763	0.103661	-4.107	4.00e-05 ***
StateParaiba	-0.314212	0.104021	-3.021	0.002522 **
StateParana	-0.404710	0.097108	-4.168	3.08e-05 ***
StatePernambuco	-0.684682	0.099419	-6.887	5.70e-12 ***
StatePiaui	-0.088315	0.108713	-0.812	0.416582
StateRio de Janeiro	-0.928535	0.097429	-9.530	< 2e-16 ***
StateRio Grande do Norte	-0.256620	0.102390	-2.506	0.012200 *
StateRio Grande do Sul	-0.550651	0.097226	-5.664	1.48e-08 ***
StateRondonia	0.006278	0.180323	0.035	0.972229
StateRoraima	-0.757129	0.203756	-3.716	0.000203 ***
StateSanta Catarina	-0.476157	0.097605	-4.878	1.07e-06 ***
StateSao Paulo	-0.687949	0.096434	-7.134	9.76e-13 ***
StateSergipe	-0.110208	0.104416	-1.055	0.291211
StateTocantins	-0.203138	0.113275	-1.793	0.072921 .

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136071 on 308378 degrees of freedom
 Residual deviance: 131355 on 308348 degrees of freedom
 AIC: 215393

Number of Fisher Scoring iterations: 6

regressione stepwise utilizzando il criterio di selezione AIC

Hide

```
model_step <- step(model_2, direction = "both")
```

Start: AIC=215392.5

ClaimNb ~ DrivAge + State + offset(log(ExposTotal))

	Df	Deviance	AIC
<none>		131355	215393
- DrivAge	4	132890	216919
- State	26	134329	218314

Hide

```
summary(model_step)
```

Call:

```
glm(formula = ClaimNb ~ DrivAge + State + offset(log(ExposTotal)),
     family = poisson(link = "log"), data = ds, subset = ExposTotal >
     0, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.506679	0.096570	36.312	< 2e-16 ***
DrivAge18-25	0.568716	0.018581	30.608	< 2e-16 ***
DrivAge26-35	0.219596	0.012007	18.290	< 2e-16 ***
DrivAge36-45	0.342111	0.010821	31.616	< 2e-16 ***
DrivAge46-55	0.136923	0.012020	11.391	< 2e-16 ***
StateAlagoas	-0.394917	0.105145	-3.756	0.000173 ***
StateAmapa	-0.190576	0.164672	-1.157	0.247147
StateAmazonas	-0.245248	0.109931	-2.231	0.025686 *
StateBahia	-0.592137	0.098446	-6.015	1.80e-09 ***
StateCeara	-0.307606	0.099587	-3.089	0.002010 **
StateDistrito Federal	-0.269377	0.098107	-2.746	0.006037 **
StateEsperito Santo	-0.333175	0.100203	-3.325	0.000884 ***
StateGoias	-0.254278	0.098123	-2.591	0.009558 **
StateMaranhao	-0.205611	0.106588	-1.929	0.053727 .
StateMato Grosso	-0.107698	0.100563	-1.071	0.284192
StateMato Grosso do Sul	-0.063164	0.100751	-0.627	0.530702
StateMinas Gerais	-0.387855	0.096975	-4.000	6.35e-05 ***
StatePara	-0.425763	0.103661	-4.107	4.00e-05 ***
StateParaiba	-0.314212	0.104021	-3.021	0.002522 **
StateParana	-0.404710	0.097108	-4.168	3.08e-05 ***
StatePernambuco	-0.684682	0.099419	-6.887	5.70e-12 ***
StatePiaui	-0.088315	0.108713	-0.812	0.416582
StateRio de Janeiro	-0.928535	0.097429	-9.530	< 2e-16 ***
StateRio Grande do Norte	-0.256620	0.102390	-2.506	0.012200 *
StateRio Grande do Sul	-0.550651	0.097226	-5.664	1.48e-08 ***
StateRondonia	0.006278	0.180323	0.035	0.972229
StateRoraima	-0.757129	0.203756	-3.716	0.000203 ***
StateSanta Catarina	-0.476157	0.097605	-4.878	1.07e-06 ***
StateSao Paulo	-0.687949	0.096434	-7.134	9.76e-13 ***
StateSergipe	-0.110208	0.104416	-1.055	0.291211
StateTocantins	-0.203138	0.113275	-1.793	0.072921 .

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136071 on 308378 degrees of freedom
 Residual deviance: 131355 on 308348 degrees of freedom
 AIC: 215393

Number of Fisher Scoring iterations: 6

Definizione del modello iniziale con tutte le variabili per ClaimAmount tranne Gender perchè non significativa

Hide

summary(model_3)

Call:

```
glm(formula = ClaimAmount ~ DrivAge + State + offset(log(ClaimNb)),
     family = Gamma(link = "log"), data = ds_1, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.6371471	0.1951631	44.256	< 2e-16 ***
DrivAge18-25	0.0484893	0.0427143	1.135	0.25630
DrivAge26-35	0.0615008	0.0300156	2.049	0.04047 *
DrivAge36-45	-0.0396177	0.0276392	-1.433	0.15176
DrivAge46-55	0.0867542	0.0292380	2.967	0.00301 **
StateAlagoas	-0.0237251	0.2158939	-0.110	0.91250
StateAmapa	-0.0006123	0.3302303	-0.002	0.99852
StateAmazonas	-0.2155279	0.2245954	-0.960	0.33725
StateBahia	0.1649021	0.2014968	0.818	0.41314
StateCeara	-0.0553773	0.2045401	-0.271	0.78659
StateDistrito Federal	-0.0445728	0.2002334	-0.223	0.82384
StateEsperito Santo	0.1439286	0.2045129	0.704	0.48159
StateGoias	0.1683011	0.2003655	0.840	0.40093
StateMaranhao	0.0591854	0.2179770	0.272	0.78599
StateMato Grosso	0.2041662	0.2069704	0.986	0.32392
StateMato Grosso do Sul	0.0308122	0.2072825	0.149	0.88183
StateMinas Gerais	0.0780021	0.1965909	0.397	0.69154
StatePara	0.1098927	0.2133828	0.515	0.60655
StateParaiba	0.0257795	0.2131742	0.121	0.90375
StateParana	0.0795892	0.1966845	0.405	0.68573
StatePernambuco	0.0243540	0.2042765	0.119	0.90510
StatePiaui	0.2173356	0.2214577	0.981	0.32641
StateRio de Janeiro	0.2787102	0.1974907	1.411	0.15818
StateRio Grande do Norte	0.0367002	0.2101874	0.175	0.86139
StateRio Grande do Sul	0.1377249	0.1970165	0.699	0.48452
StateRondonia	-0.2765508	0.3522459	-0.785	0.43240
StateRoraima	-0.2952017	0.3936149	-0.750	0.45327
StateSanta Catarina	0.0938223	0.1975988	0.475	0.63492
StateSao Paulo	0.0841068	0.1951013	0.431	0.66640
StateSergipe	-0.0188034	0.2149877	-0.087	0.93030
StateTocantins	0.4266206	0.2319535	1.839	0.06589 .

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

(Dispersion parameter for Gamma family taken to be 3.28117)

Null deviance: 52351 on 35666 degrees of freedom

Residual deviance: 52017 on 35636 degrees of freedom

AIC: 723523

Number of Fisher Scoring iterations: 7

variabile State poco significativa

Hide

```
modello_step_2 <- step(model_3, direction = "both")
```

Start: AIC=723523.2

ClaimAmount ~ DrivAge + State + offset(log(ClaimNb))

	Df	Deviance	AIC
<none>		52017	723523
- DrivAge	4	52107	723543
- State	26	52264	723547

[Hide](#)

summary(modello_step_2)

Call:

```
glm(formula = ClaimAmount ~ DrivAge + State + offset(log(ClaimNb)),
     family = Gamma(link = "log"), data = ds_1, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.6371471	0.1951631	44.256	< 2e-16 ***
DrivAge18-25	0.0484893	0.0427143	1.135	0.25630
DrivAge26-35	0.0615008	0.0300156	2.049	0.04047 *
DrivAge36-45	-0.0396177	0.0276392	-1.433	0.15176
DrivAge46-55	0.0867542	0.0292380	2.967	0.00301 **
StateAlagoas	-0.0237251	0.2158939	-0.110	0.91250
StateAmapa	-0.0006123	0.3302303	-0.002	0.99852
StateAmazonas	-0.2155279	0.2245954	-0.960	0.33725
StateBahia	0.1649021	0.2014968	0.818	0.41314
StateCeara	-0.0553773	0.2045401	-0.271	0.78659
StateDistrito Federal	-0.0445728	0.2002334	-0.223	0.82384
StateEsperito Santo	0.1439286	0.2045129	0.704	0.48159
StateGoias	0.1683011	0.2003655	0.840	0.40093
StateMaranhao	0.0591854	0.2179770	0.272	0.78599
StateMato Grosso	0.2041662	0.2069704	0.986	0.32392
StateMato Grosso do Sul	0.0308122	0.2072825	0.149	0.88183
StateMinas Gerais	0.0780021	0.1965909	0.397	0.69154
StatePara	0.1098927	0.2133828	0.515	0.60655
StateParaiba	0.0257795	0.2131742	0.121	0.90375
StateParana	0.0795892	0.1966845	0.405	0.68573
StatePernambuco	0.0243540	0.2042765	0.119	0.90510
StatePiaui	0.2173356	0.2214577	0.981	0.32641
StateRio de Janeiro	0.2787102	0.1974907	1.411	0.15818
StateRio Grande do Norte	0.0367002	0.2101874	0.175	0.86139
StateRio Grande do Sul	0.1377249	0.1970165	0.699	0.48452
StateRondonia	-0.2765508	0.3522459	-0.785	0.43240
StateRoraima	-0.2952017	0.3936149	-0.750	0.45327
StateSanta Catarina	0.0938223	0.1975988	0.475	0.63492
StateSao Paulo	0.0841068	0.1951013	0.431	0.66640
StateSergipe	-0.0188034	0.2149877	-0.087	0.93030
StateTocantins	0.4266206	0.2319535	1.839	0.06589 .

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

(Dispersion parameter for Gamma family taken to be 3.28117)

Null deviance: 52351 on 35666 degrees of freedom

Residual deviance: 52017 on 35636 degrees of freedom

AIC: 723523

Number of Fisher Scoring iterations: 7

per le celle scegliamo come variabili DrivAge + State. creiamo la colonna frequenza nelle celle tariffarie come rapporto tra ClaimNb ed ExposTotal

DrivAge	State	ClaimNb	ExposTotal	ClaimAmount	Frequenza						
<fctr>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>						
>55	Acre	20	0.51577857	94155	38.776330						
18-25	Acre	3	0.16292663	19276	18.413196						
26-35	Acre	22	0.59302646	78273	37.097839						
36-45	Acre	46	0.82585813	253242	55.699639						
46-55	Acre	17	0.46628148	117683	36.458665						
>55	Alagoas	122	4.87688722	709667	25.015957						
18-25	Alagoas	27	0.80222931	132650	33.656212						
26-35	Alagoas	72	3.03265566	368341	23.741568						
36-45	Alagoas	236	6.68768437	1089597	35.288747						
46-55	Alagoas	100	4.63251342	722166	21.586554						
1-10 of 135 rows		Previous	1	2	3	4	5	6	...	14	Next

inserimento frequenze in dataset
media e la varianza del numero di sinistri

Hide

```
cat("avg =",media_sinistri," variance =",varianza_sinistri,"phi =",varianza_sinistri/media_si
nistri,"\n")
```

avg = 23.76307 variance = 3265.839 phi = 137.4334

MODELLO PER FREQUENZA

Hide

```
#nobs(mod_frequenza)
summary(mod_frequenza)
```

Call:

```
glm(formula = ClaimNb ~ State + DrivAge + offset(log(ExposTotal)),
     family = poisson(link = "log"), data = ds, subset = ExposTotal >
     0, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.506679	0.096570	36.312	< 2e-16 ***
StateAlagoas	-0.394917	0.105145	-3.756	0.000173 ***
StateAmapa	-0.190576	0.164672	-1.157	0.247147
StateAmazonas	-0.245248	0.109931	-2.231	0.025686 *
StateBahia	-0.592137	0.098446	-6.015	1.80e-09 ***
StateCeara	-0.307606	0.099587	-3.089	0.002010 **
StateDistrito Federal	-0.269377	0.098107	-2.746	0.006037 **
StateEsperito Santo	-0.333175	0.100203	-3.325	0.000884 ***
StateGoias	-0.254278	0.098123	-2.591	0.009558 **
StateMaranhao	-0.205611	0.106588	-1.929	0.053727 .
StateMato Grosso	-0.107698	0.100563	-1.071	0.284192
StateMato Grosso do Sul	-0.063164	0.100751	-0.627	0.530702
StateMinas Gerais	-0.387855	0.096975	-4.000	6.35e-05 ***
StatePara	-0.425763	0.103661	-4.107	4.00e-05 ***
StateParaiba	-0.314212	0.104021	-3.021	0.002522 **
StateParana	-0.404710	0.097108	-4.168	3.08e-05 ***
StatePernambuco	-0.684682	0.099419	-6.887	5.70e-12 ***
StatePiaui	-0.088315	0.108713	-0.812	0.416582
StateRio de Janeiro	-0.928535	0.097429	-9.530	< 2e-16 ***
StateRio Grande do Norte	-0.256620	0.102390	-2.506	0.012200 *
StateRio Grande do Sul	-0.550651	0.097226	-5.664	1.48e-08 ***
StateRondonia	0.006278	0.180323	0.035	0.972229
StateRoraima	-0.757129	0.203756	-3.716	0.000203 ***
StateSanta Catarina	-0.476157	0.097605	-4.878	1.07e-06 ***
StateSao Paulo	-0.687949	0.096434	-7.134	9.76e-13 ***
StateSergipe	-0.110208	0.104416	-1.055	0.291211
StateTocantins	-0.203138	0.113275	-1.793	0.072921 .
DrivAge18-25	0.568716	0.018581	30.608	< 2e-16 ***
DrivAge26-35	0.219596	0.012007	18.290	< 2e-16 ***
DrivAge36-45	0.342111	0.010821	31.616	< 2e-16 ***
DrivAge46-55	0.136923	0.012020	11.391	< 2e-16 ***

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

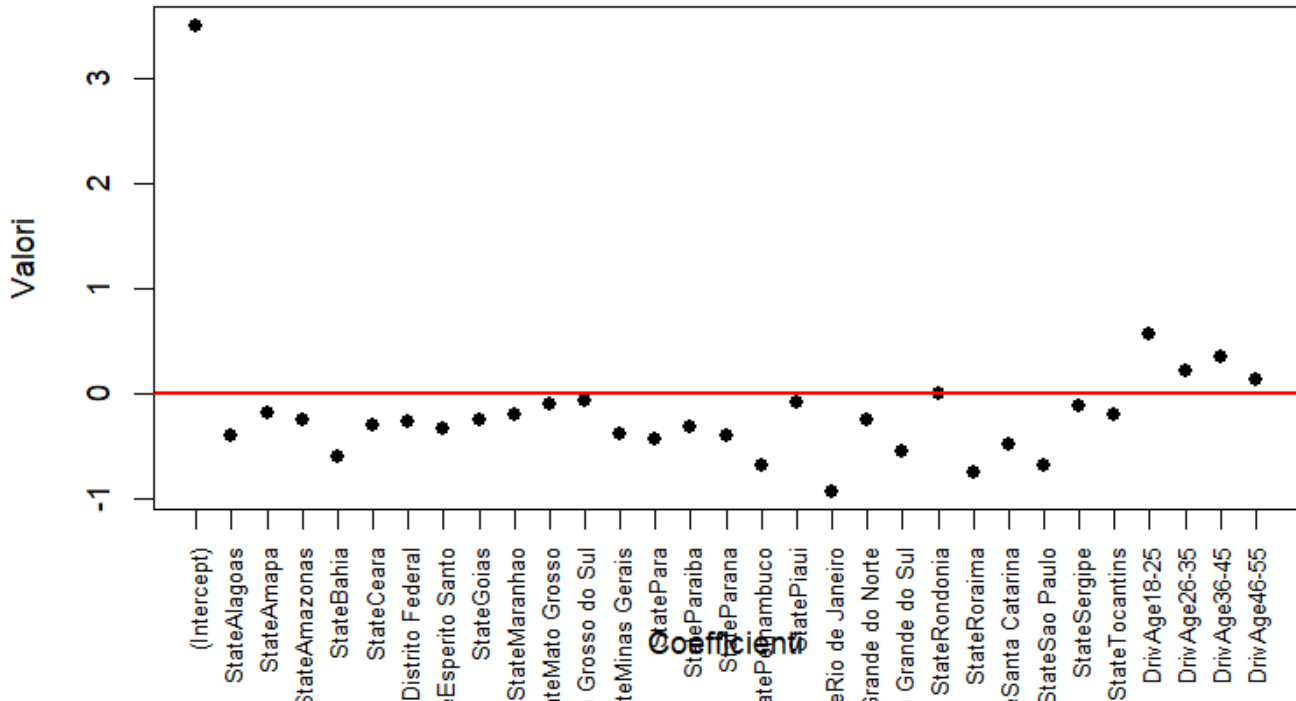
(Dispersion parameter for poisson family taken to be 1)

Null deviance: 136071 on 308378 degrees of freedom
 Residual deviance: 131355 on 308348 degrees of freedom
 AIC: 215393

Number of Fisher Scoring iterations: 6

molte delle variabili risultano essere statisticamente significative. Il segno dei coefficienti riflette l'effetto sull'entità del logaritmo del numero di sinistri rispetto al gruppo di riferimento.

Coefficienti del modello



Le categorie delle variabili tariffarie che presentano coefficienti superiori aumentano la frequenza di sinistri. ex- il coefficiente associato alla variabile “DrivAge” ha una tendenza a essere maggiore -> minore è il range, quindi, per questa categoria la frequenza dei sinistri aumenta di più rispetto alla fascia di età maggiore.

modello moltiplicativo

utilizzo della distribuzione quasi-Poisson [utilizzata per gestire la sovradisersione (condizione in cui la varianza dei dati è maggiore della media)]

[Hide](#)

```
mod_frequenza_2 <- glm(ClaimNb ~ State + DrivAge + offset(log(ExposTotal)), family = quasipoisson(link="log"),
                        data = ds, subset=ExposTotal>0, x = TRUE)
mod_frequenza_2 <- glm(ClaimNb ~ State + DrivAge + offset(log(ExposTotal)), family = quasipoisson(link="log"),
                        data = ds, subset=ExposTotal>0, x = TRUE)
summary(mod_frequenza_2)
```

Call:

```
glm(formula = ClaimNb ~ State + DrivAge + offset(log(ExposTotal)),
     family = quasipoisson(link = "log"), data = ds, subset = ExposTotal >
     0, x = TRUE)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.506679	0.105152	33.349	< 2e-16 ***
StateAlagoas	-0.394917	0.114489	-3.449	0.000562 ***
StateAmapa	-0.190576	0.179306	-1.063	0.287848
StateAmazonas	-0.245248	0.119700	-2.049	0.040477 *
StateBahia	-0.592137	0.107195	-5.524	3.32e-08 ***
StateCeara	-0.307606	0.108437	-2.837	0.004558 **
StateDistrito Federal	-0.269377	0.106825	-2.522	0.011680 *
StateEsperito Santo	-0.333175	0.109108	-3.054	0.002261 **
StateGoias	-0.254278	0.106842	-2.380	0.017317 *
StateMaranhao	-0.205611	0.116060	-1.772	0.076462 .
StateMato Grosso	-0.107698	0.109499	-0.984	0.325339
StateMato Grosso do Sul	-0.063164	0.109705	-0.576	0.564772
StateMinas Gerais	-0.387855	0.105592	-3.673	0.000240 ***
StatePara	-0.425763	0.112873	-3.772	0.000162 ***
StateParaiba	-0.314212	0.113265	-2.774	0.005535 **
StateParana	-0.404710	0.105738	-3.827	0.000129 ***
StatePernambuco	-0.684682	0.108254	-6.325	2.54e-10 ***
StatePiaui	-0.088315	0.118374	-0.746	0.455629
StateRio de Janeiro	-0.928535	0.106087	-8.753	< 2e-16 ***
StateRio Grande do Norte	-0.256620	0.111489	-2.302	0.021350 *
StateRio Grande do Sul	-0.550651	0.105866	-5.201	1.98e-07 ***
StateRondonia	0.006278	0.196348	0.032	0.974495
StateRoraima	-0.757129	0.221863	-3.413	0.000644 ***
StateSanta Catarina	-0.476157	0.106279	-4.480	7.46e-06 ***
StateSao Paulo	-0.687949	0.105004	-6.552	5.70e-11 ***
StateSergipe	-0.110208	0.113695	-0.969	0.332382
StateTocantins	-0.203138	0.123341	-1.647	0.099566 .
DrivAge18-25	0.568716	0.020232	28.110	< 2e-16 ***
DrivAge26-35	0.219596	0.013074	16.797	< 2e-16 ***
DrivAge36-45	0.342111	0.011783	29.035	< 2e-16 ***
DrivAge46-55	0.136923	0.013088	10.461	< 2e-16 ***

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

(Dispersion parameter for quasipoisson family taken to be 1.185629)

Null deviance: 136071 on 308378 degrees of freedom
 Residual deviance: 131355 on 308348 degrees of freedom
 AIC: NA

Number of Fisher Scoring iterations: 6

La distribuzione quasi-Poisson consente la stima del parametro ϕ , che nel modello precedente, con la distribuzione Poisson era fissato a 1. Dalla valutazione dei risultati, il valore di ϕ si avvicina molto a uno, indicando una bassa dispersione nel modello.

modello nullo

Def modello nullo per la frequenza dei sinistri con solo offset

[Hide](#)

```
mod_nullo <- glm(ClaimNb ~ offset(log(ExposTotal)), family = poisson(link = "log"),
                 data = ds, subset=ExposTotal>0, x= TRUE)
mod_nullo <- glm(ClaimNb ~ offset(log(ExposTotal)), family = poisson(link = "log"),
                 data = ds, subset=ExposTotal>0, x= TRUE)

list(
  nobs_mod_freq =nobs(mod_frequenza),
  nobs_mod_nullo =nobs(mod_nullo)
)
```

```
$nobs_mod_freq
[1] 308379
```

```
$nobs_mod_nullo
[1] 308379
```

ANOVA LRT test

il mod nullo(non include alcuna variabile tariffaria), ha lo scopo di effettuare un test ANOVA con likelihood ratio (LRT). Questo test ci consente di confrontare il modello nullo con il modello sviluppato per la frequenza, al fine di valutare se quest'ultimo si adatta meglio ai dati rispetto al modello nullo.

[Hide](#)

```
print(ANOVA_test)
```

Analysis of Deviance Table

Model 1: ClaimNb ~ offset(log(ExposTotal))

Model 2: ClaimNb ~ State + DrivAge + offset(log(ExposTotal))

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	308378	136071			
2	308348	131355	30	4715.9	< 2.2e-16 ***

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

I risultati del test indicano che il modello per il numero di sinistri fornisce un miglior adattamento ai dati rispetto al modello nullo.

Aspetti chiave di questo test:

Likelihood Ratio Test (LRT): Questo test viene utilizzato per confrontare la bontà di adattamento tra due modelli annidati (dove un modello è una versione semplificata dell'altro). L'ipotesi nulla è che il modello più semplice (nel tuo caso, modello_nullo) sia sufficiente per spiegare i dati.

Modelli annidati: Un modello annidato significa che i parametri del modello più semplice sono un sottoinsieme dei parametri del modello più complesso. Nel tuo esempio:

modello_nullo è il modello nullo (più semplice). modello_frequenza è il modello più complesso.

ANOVA con LRT: Confronta le log-likelihood dei due modelli e verifica se il miglioramento nella log-likelihood, passando dal modello nullo a quello più complesso, è statisticamente significativo.

Bilanciamento tra i sinistri previsti e osservati, sia nel complesso che a livello singolo per ciascun livello di ogni fattore.

Durante questa analisi, ho riscontrato un errore trascurabile, attribuibile a un problema di arrotondamento. Calcoliamo il numero di sinistri e della frequenza previsti dal modello:

[Hide](#)

```
# bilanciamento sulla frequenza
cat( sum(celle$Freq * celle$ExposTotal), "\n",
     sum(celle$frequenza_prevista * celle$ExposTotal)
)
```

```
70699
70699
```

Confronto i risultati

[Hide](#)

```
cat(
  "Sinistri previsti totali: ", sinistri_previsti_totali, "\n",
  "Sinistri totali: ", sinistri_totali_reali, "\n",
  "Differenza (arrotondata): ", round(sinistri_totali_reali, digits = 6) - round(sinistri_pre
visti_totali, digits = 6), "\n"
)
```

```
Sinistri previsti totali: 70699
Sinistri totali: 70699
Differenza (arrotondata): 0
```

con un arrotondamento alla sesta cifra decimale l'errore è zero. La differenza tra sinistri totali reali e sinistri previsti ha un errore inferiore alla sesta cifra decimale.

Bilanciamento sul singolo livello di ogni fattore

[Hide](#)

```
t(X) %*% (ds$ClaimNb - sinistri_previsti_singola_polizza)
```

```

                                [,1]
(Intercept)                -8.055973e-08
StateAlagoas                -1.419694e-10
StateAmapa                  -3.869901e-11
StateAmazonas               -4.599528e-11
StateBahia                  -5.392119e-10
StateCeara                  -3.377784e-10
StateDistrito Federal       -5.006572e-10
StateEsperito Santo         -2.915179e-10
StateGoias                  -4.877479e-10
StateMaranhao               -9.176131e-11
StateMato Grosso            -8.200624e-11
StateMato Grosso do Sul     1.483258e-11
StateMinas Gerais           -1.340735e-09
StatePara                   -8.950560e-11
StateParaiba                -1.222851e-10
StateParana                 -8.607412e-10
StatePernambuco             -5.710875e-10
StatePiaui                  -7.524392e-11
StateRio de Janeiro         -1.310466e-09
StateRio Grande do Norte   -1.313899e-10
StateRio Grande do Sul     -8.304377e-10
StateRondonia               -2.485220e-10
StateRoraima                -6.618637e-08
StateSanta Catarina         -1.689355e-09
StateSao Paulo              -4.292653e-09
StateSergipe                -1.522870e-10
StateTocantins              -6.032080e-11
DrivAge18-25                -4.528872e-09
DrivAge26-35                -2.394255e-08
DrivAge36-45                -2.810578e-08
DrivAge46-55                -1.510297e-08

```

prodotto scalare tra il trasposto di X_1 e la differenza tra il numero di sinistri e il numero stimato di sinistri. il risultato è un vettore di 0 → indica che la somma pesata dei residui per il numero di sinistri è 0.

[Hide](#)

```
t(X) %*% ((ds$Frequenza - frequenza_prevista_singola_polizza)* ds$ExposTotal)
```

```

                                [,1]
(Intercept)                -8.055957e-08
StateAlagoas                 -1.419814e-10
StateAmapa                   -3.870167e-11
StateAmazonas                -4.598484e-11
StateBahia                   -5.392155e-10
StateCeara                   -3.377797e-10
StateDistrito Federal        -5.006697e-10
StateEsperito Santo          -2.915412e-10
StateGoias                   -4.877659e-10
StateMaranhao                -9.175331e-11
StateMato Grosso             -8.199158e-11
StateMato Grosso do Sul      1.482703e-11
StateMinas Gerais            -1.340835e-09
StatePara                    -8.950250e-11
StateParaiba                 -1.222780e-10
StateParana                  -8.606932e-10
StatePernambuco              -5.710589e-10
StatePiaui                   -7.524387e-11
StateRio de Janeiro          -1.310438e-09
StateRio Grande do Norte    -1.313876e-10
StateRio Grande do Sul       -8.304581e-10
StateRondonia                -2.485235e-10
StateRoraima                 -6.618637e-08
StateSanta Catarina          -1.689285e-09
StateSao Paulo               -4.292420e-09
StateSergipe                  -1.522922e-10
StateTocantins               -6.031675e-11
DrivAge18-25                 -4.528850e-09
DrivAge26-35                 -2.394251e-08
DrivAge36-45                 -2.810607e-08
DrivAge46-55                 -1.510292e-08

```

prodotto scalare tra il trasposto di X_1 e la differenza tra le frequenze e quelle stimate moltiplicate per l'esposizione. Il risultato è un vettore di 0 → indica che la somma pesata dei residui per la frequenza moltiplicata per l'esposizione è 0.

ANALISI DELLA SEVERITY

1. rimuovere dal dataset tutte le osservazioni con un numero di sinistri pari a zero. Il modello costruito utilizzando un Generalized Linear Model (GLM) della famiglia gamma, con un link logaritmico e le stesse variabili utilizzate per creare le celle tariffarie.

Filtraggio del dataset (records con ClaimNb !=0) Verifica presenza di ClaimAmount =0

Hide

```

if (any(dataset_severity$ClaimAmount == 0)) {
  cat("ClaimAmount =0.\n")
} else {
  cat("ClaimAmount !=0.\n")
}

```

```
ClaimAmount !=0.
```

Definizione del modello GLM per la severity

Hide

summary(mod_severity)

Call:

```
glm(formula = ClaimAmount ~ DrivAge + State + offset(log(ClaimNb)),
     family = Gamma(link = "log"), data = dataset_severity)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	8.6371471	0.1951631	44.256	< 2e-16 ***
DrivAge18-25	0.0484893	0.0427143	1.135	0.25630
DrivAge26-35	0.0615008	0.0300156	2.049	0.04047 *
DrivAge36-45	-0.0396177	0.0276392	-1.433	0.15176
DrivAge46-55	0.0867542	0.0292380	2.967	0.00301 **
StateAlagoas	-0.0237251	0.2158939	-0.110	0.91250
StateAmapa	-0.0006123	0.3302303	-0.002	0.99852
StateAmazonas	-0.2155279	0.2245954	-0.960	0.33725
StateBahia	0.1649021	0.2014968	0.818	0.41314
StateCeara	-0.0553773	0.2045401	-0.271	0.78659
StateDistrito Federal	-0.0445728	0.2002334	-0.223	0.82384
StateEsperito Santo	0.1439286	0.2045129	0.704	0.48159
StateGoias	0.1683011	0.2003655	0.840	0.40093
StateMaranhao	0.0591854	0.2179770	0.272	0.78599
StateMato Grosso	0.2041662	0.2069704	0.986	0.32392
StateMato Grosso do Sul	0.0308122	0.2072825	0.149	0.88183
StateMinas Gerais	0.0780021	0.1965909	0.397	0.69154
StatePara	0.1098927	0.2133828	0.515	0.60655
StateParaiba	0.0257795	0.2131742	0.121	0.90375
StateParana	0.0795892	0.1966845	0.405	0.68573
StatePernambuco	0.0243540	0.2042765	0.119	0.90510
StatePiaui	0.2173356	0.2214577	0.981	0.32641
StateRio de Janeiro	0.2787102	0.1974907	1.411	0.15818
StateRio Grande do Norte	0.0367002	0.2101874	0.175	0.86139
StateRio Grande do Sul	0.1377249	0.1970165	0.699	0.48452
StateRondonia	-0.2765508	0.3522459	-0.785	0.43240
StateRoraima	-0.2952017	0.3936149	-0.750	0.45327
StateSanta Catarina	0.0938223	0.1975988	0.475	0.63492
StateSao Paulo	0.0841068	0.1951013	0.431	0.66640
StateSergipe	-0.0188034	0.2149877	-0.087	0.93030
StateTocantins	0.4266206	0.2319535	1.839	0.06589 .

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

(Dispersion parameter for Gamma family taken to be 3.28117)

Null deviance: 52351 on 35666 degrees of freedom

Residual deviance: 52017 on 35636 degrees of freedom

AIC: 723523

Number of Fisher Scoring iterations: 7

Presenza di regressori non significativi

Verifica con algoritmo Stepwise basato su criterio info AIC.

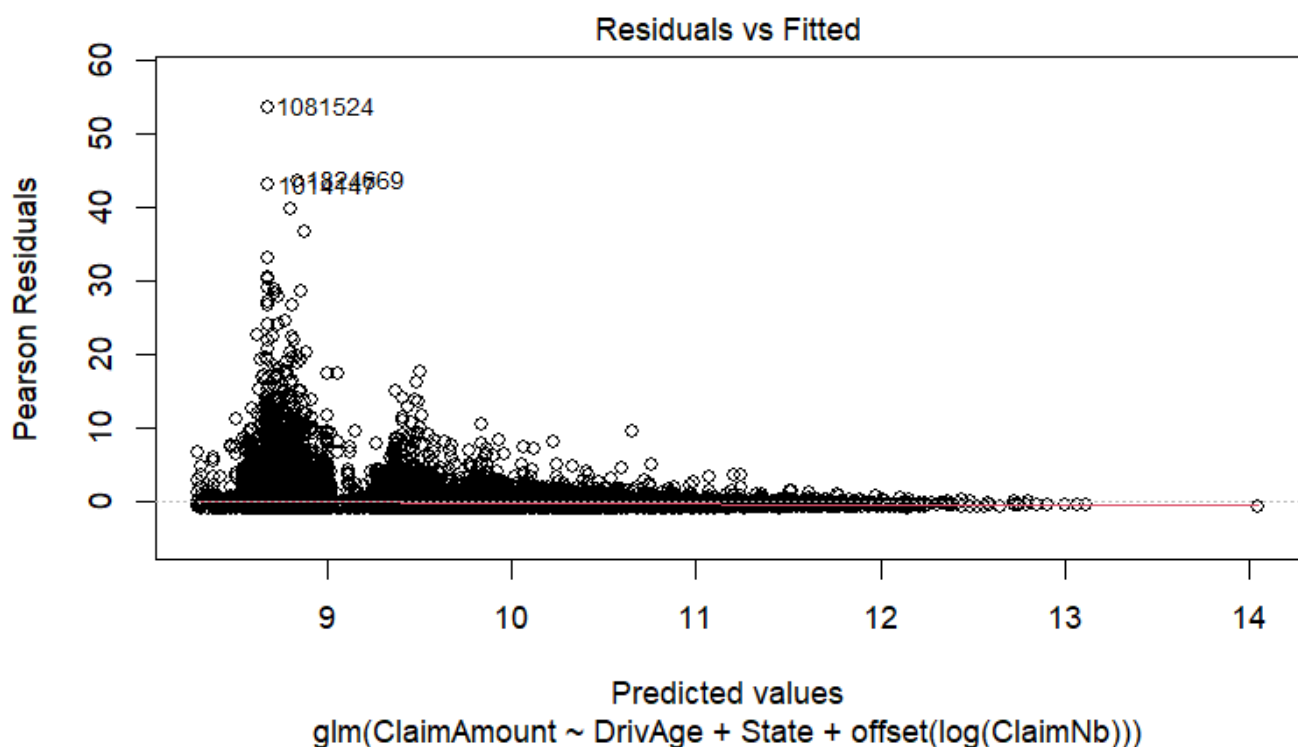
Hide

```
model_step_severity <- step(mod_severity, direction = "both")
```

Start: AIC=723523.2

ClaimAmount ~ DrivAge + State + offset(log(ClaimNb))

	Df	Deviance	AIC
<none>		52017	723523
- DrivAge	4	52107	723543
- State	26	52264	723547



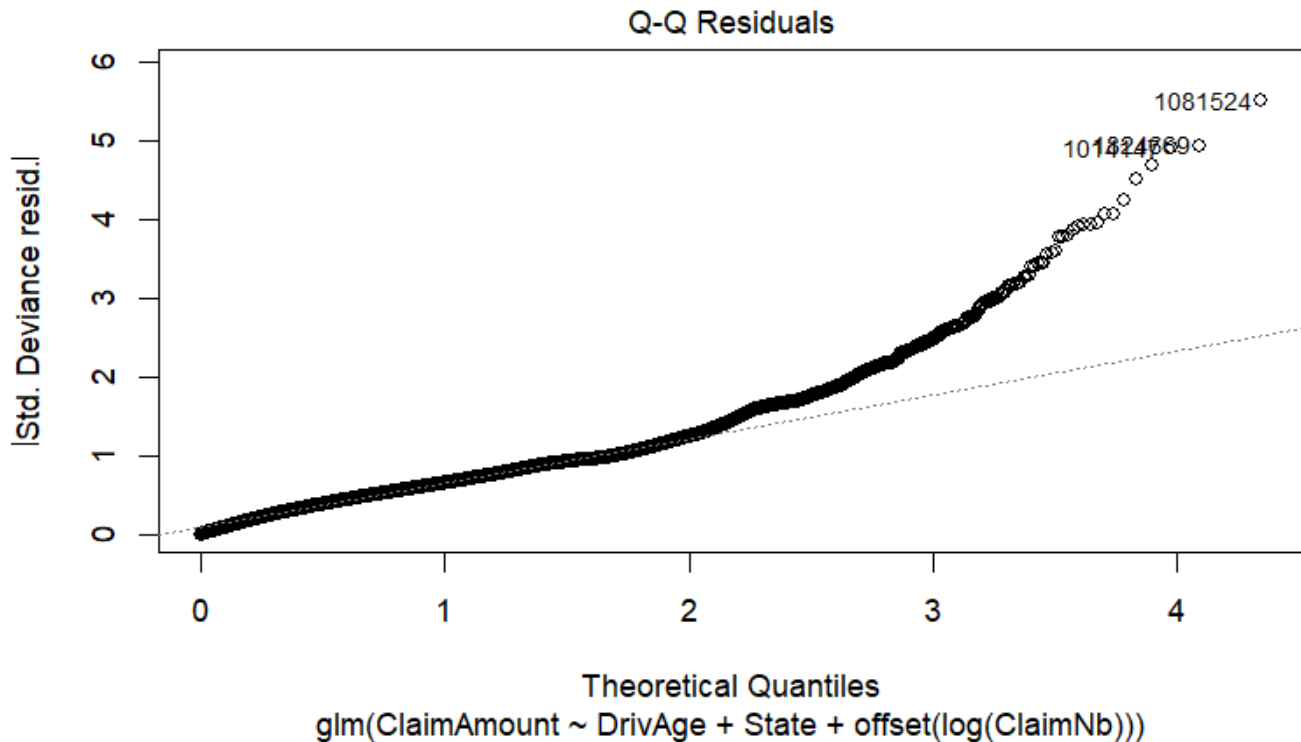
presenza di outlier evidenziati dal grafico

Dal grafico dei residui si nota che i punti sono prevalentemente concentrati intorno allo zero.

Gli outlier, esaminati singolarmente, non rappresentano le osservazioni con il maggior risarcimento totale -> non sono grandi sinistri. La concentrazione dei valori intorno allo zero è un buon segno, suggerisce che il modello prevede accuratamente per molti dati. La minore densità di punti man mano che i residui aumentano può suggerire pochi casi con grandi errori di predizione, (presenza di outliers o parte della variabilità non spiegata dalle nostre variabili è sempre da tenere in considerazione).

La dispersione crescente nel plot (varianza superiore alla media, eterosched) potrebbe spiegare gli elevati residui per i bassi valori predetti.

-> uso di una distribuzione quasi-Poisson o un modello di dispersione negativa potrebbero migliorare il modello.



Il Q-Q plot dei residui:

I punti iniziali che partono da zero indicano che i residui standardizzati sono centrati intorno a zero, il che è atteso per un modello ben adattato. Una salita lenta all'inizio suggerisce che i residui nelle code inferiori (gli estremi negativi) sono vicini ai valori previsti dalla distribuzione normale teorica. I punti iniziano a deviare dalla linea diagonale e salgono più rapidamente, indica che i residui nelle code superiori (gli estremi positivi) si discostano dalla distribuzione normale teorica. Questo pattern può suggerire la presenza di outliers o di una distribuzione asimmetrica dei residui.

Successivamente, abbiamo eseguito un test ANOVA che ci ha consentito di confrontare il modello nullo con il nostro modello: Il modello nullo risulta meno significativo rispetto al modello contenente le variabili.

ANOVA SEVERITY

[Hide](#)

```
anova(modello_nullo_severity, mod_severity, test = 'LRT')
```

Analysis of Deviance Table

Model 1: ClaimAmount ~ offset(log(ClaimNb))

Model 2: ClaimAmount ~ DrivAge + State + offset(log(ClaimNb))

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	35666	52351			
2	35636	52017	30	334.93	8.702e-10 ***

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

P.val = 0

aggiunta colonna severity → rapporto tra severity e numero di sinistri

TARIFFAZIONE

Dopo aver definito il modello per la stima della frequenza e il modello per la severity

Calcolato il premio previsto moltiplicando la frequenza prevista per la severity prevista e il premio basato sulla frequenza osservata e la severity osservata

Hide

```
cat(sum((celle$premio * celle$ExposTotal)[!is.na(celle$premio * celle$ExposTotal)]) - sum(celle$ClaimAmount))
```

0

BILANCIO

Valutare la performance dell'assicurazione: 1. Stimato l'ammontare dei risarcimenti che l'assicurazione deve pagare. 2. I costi totali in termini di risarcimenti ammontano a 394'870'208
3. Le entrate (premi incassati dall'assicurazione) stimate: 449'049'784 4. Calcolo del Bilancio totale: differenza tra le entrate (premi) e le uscite (costi): 54'179'576 [Saldo Positivo]

Hide

```
cat("Costi Totali:", costi_totali, "\n",  
    "Entrate Totali:", entrate_totali, "\n",  
    "Bilancio:", bilancio, "\n")
```

Costi Totali: 394870208
Entrate Totali: 449049784
Bilancio: 54179576

celle.Frequenza <dbl>	celle.sinistri_previsti <dbl>	celle.frequenza_prevista <dbl>	celle.Severity <dbl>
38.776330	17.194703	33.33737	4707.750
18.413196	9.592102	58.87375	6425.333
37.097839	24.624917	41.52415	3557.864
55.699639	38.762661	46.93622	5505.261
36.458665	17.825617	38.22931	6922.529
25.015957	109.537707	22.46058	5816.943
33.656212	31.820706	39.66535	4912.963
23.741568	84.842480	27.97630	5115.847
35.288747	211.482020	31.62261	4616.936
21.586554	119.317087	25.75645	7221.660

1-10 of 135 rows

Previous123456...14Next

celle.Severity_prevista <dbl>	celle.premio_previsto <dbl>	celle.premio <dbl>	celle.differenza <dbl>
5637.224	187930.25	182549.27	2775.3964
5917.305	348373.95	118310.92	37483.3932
5994.801	248928.98	131989.05	69348.4724
5418.257	254312.50	306641.04	-43215.9511
6148.118	235038.27	252386.18	-8089.0068
5505.055	123646.71	145516.39	-106655.9519
5778.568	229208.93	165351.72	51228.1233
5854.247	163780.17	121458.23	128347.8661
5291.221	167322.20	162925.90	29401.0787
6003.970	154640.93	155890.75	-5789.8254
1-10 of 135 rows	Previous	1	2 3 4 5 6 ... 14 Next