Insurance Data Science

Group Assignment - 2023/2024

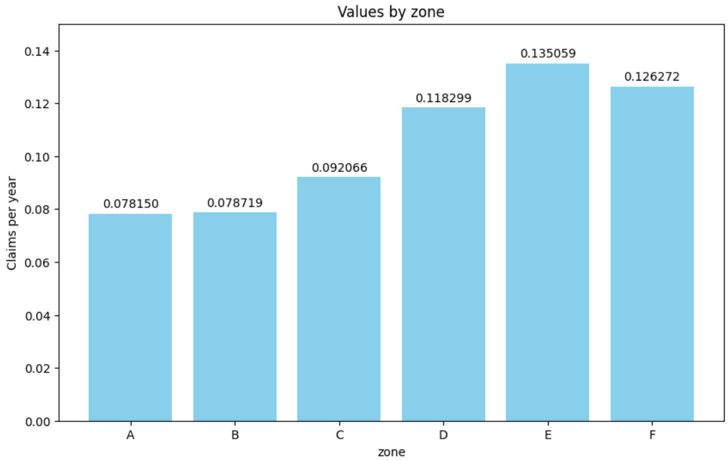
Fernando Reis – nº 20231535
Luis Ribeiro – nº 20231536
Thiago Bellas – nº 20231131
Renato Morais – nº 20231135
Saad Islam – nº 20230513

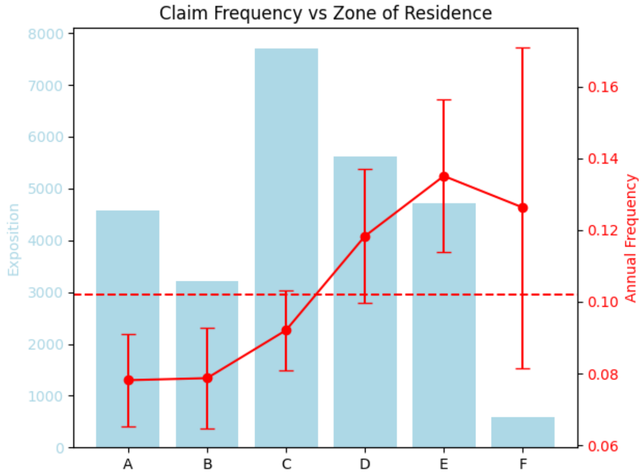
NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Contents

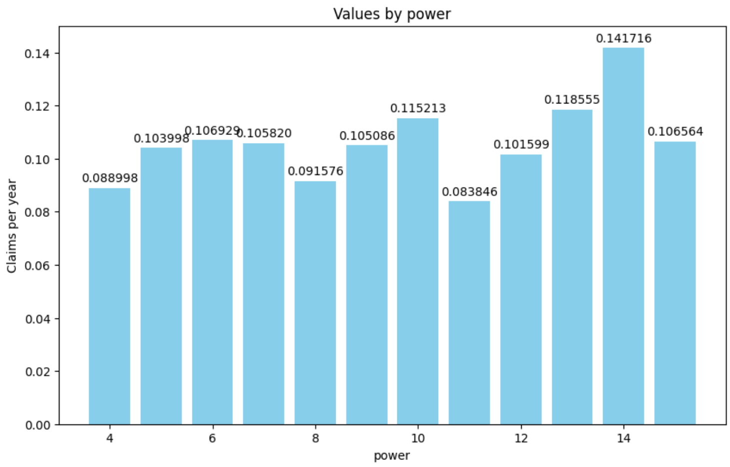
Part 1 – Statistical analysis

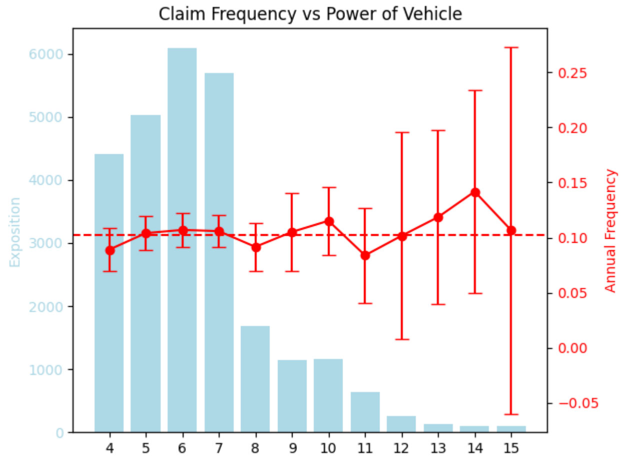
Task 1 – Number of Claims



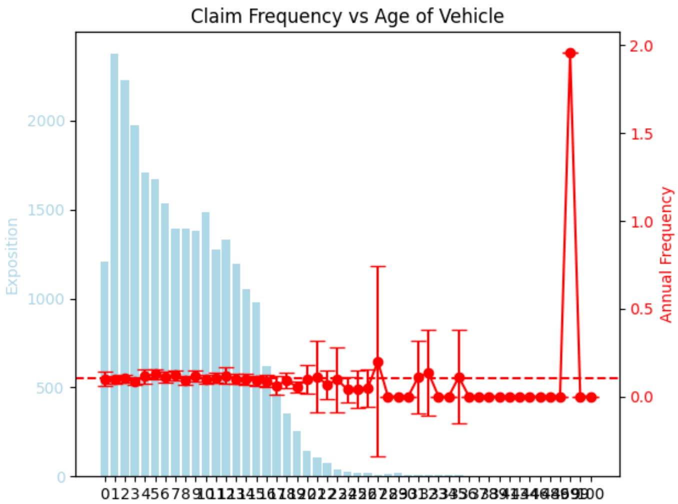


Starting with Zone of Residence, these graphs illustrate the variation in car insurance claims across different zones. The first bar chart shows that Zones E and F have the highest annual claim rates (0.135 and 0.126, respectively), while Zones A and B have the lowest (around 0.078). The second graph, combining claim frequency and exposure, highlights that despite moderate exposure, Zone E has the highest claim frequency, corroborating the first chart. Zones with higher exposures, like C and D, do not necessarily have the highest claim frequencies. This data suggests significant geographical differences in claim rates, with Zones E and F being particularly high-risk areas.

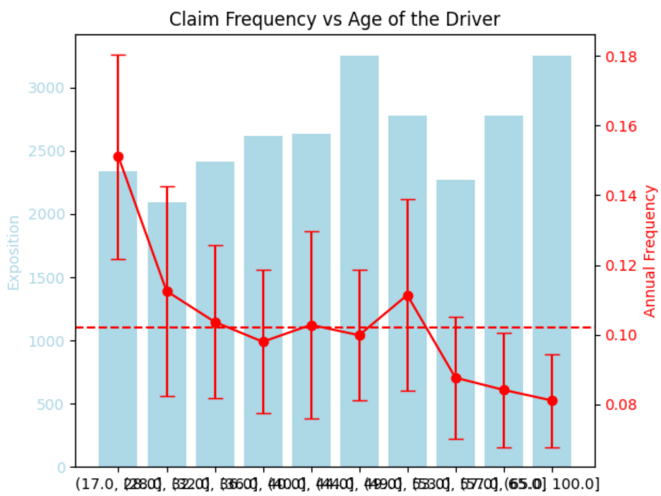


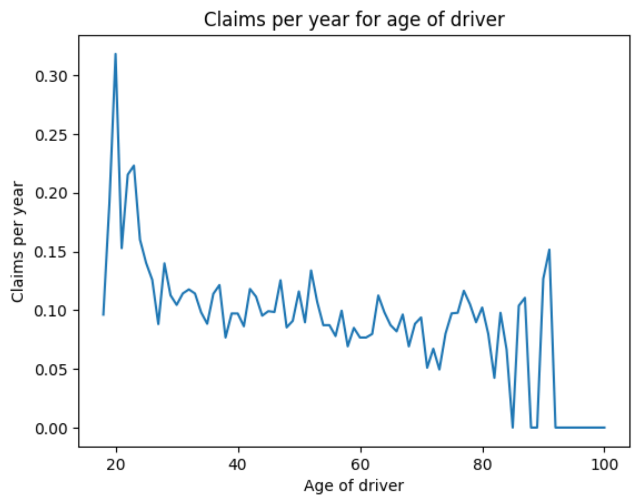


Looking at power, the first bar chart shows the claims per year based on the vehicle's power. Vehicles with higher power (14 and 10) have the highest claim rates (0.142 and 0.115 respectively), while those with power 4 and 11 have the lowest (0.089 and 0.084). The second graph, combining claim frequency and exposure, indicates that despite varying exposure levels, the claim frequency remains relatively consistent across different power levels, with a noticeable spike for higher-power vehicles (14). This suggests that higher-powered vehicles tend to have higher claim rates, making them a higher-risk category for insurers.

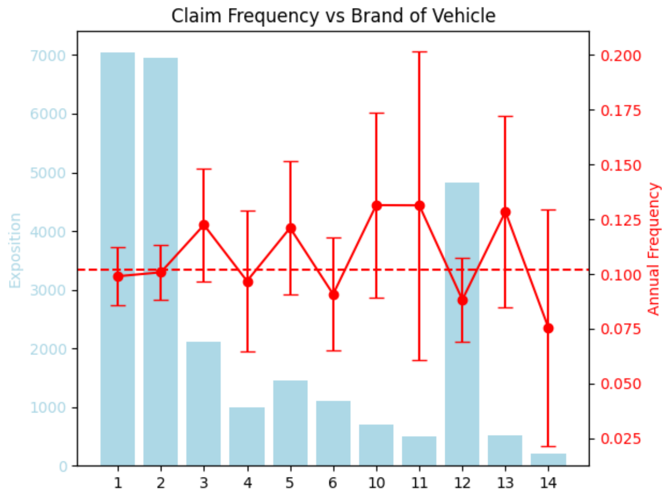


Now for Age of Vehicle, this graph combines claim frequency and exposure based on the age of the vehicle. The blue bars indicate the exposure, which decreases significantly as the vehicle age increases. The red line represents the annual claim frequency, showing a relatively stable claim rate for most vehicle ages. However, there is a notable spike in claim frequency for vehicles around 90 years old, likely due to very low exposure resulting in higher variability. Overall, the claim frequency remains constant across different vehicle ages, suggesting that vehicle age has a minimal impact on claim rates, except for very old vehicles which show erratic behavior due to lower sample sizes.

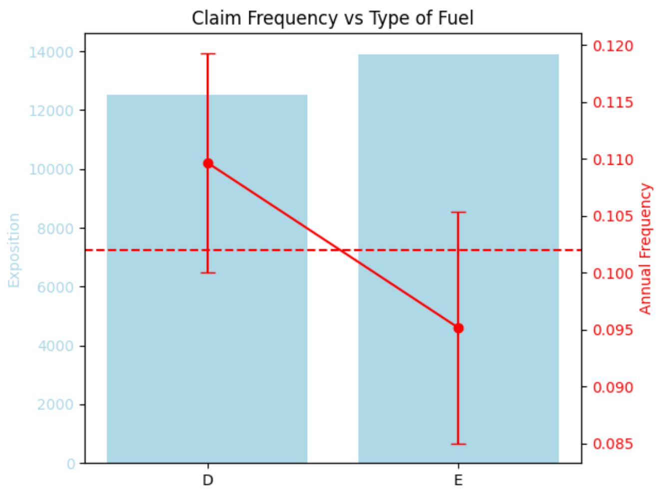




For the age of the Driver, the first graph illustrates claim frequency and exposure across different driver ages, while the second graph focuses on claims per year by age. Both graphs show that young drivers (especially those around 17 to 22 years) have the highest claim rates, which then steadily decrease and stabilize for middle-aged drivers (30 to 65 years). The claim rates slightly drop for older drivers. These trends highlight that younger drivers are at higher risk, possibly due to inexperience, while older drivers exhibit lower claim rates, likely due to more cautious driving behavior. This data underscores the importance of considering driver age in risk assessment for car insurance.



Regarding the Brand of the Vehicle, the graph shows the relationship between claim frequency and the brand of the vehicle. The blue bars represent the exposure for each brand, while the red line indicates the annual claim frequency, with error bars showing variability. Brands 1 and 2 have the highest exposure but moderate claim frequencies. Claim frequencies for other brands fluctuate around the average line, with significant variability, especially for brands with lower exposure (e.g., 3, 4, 10). This suggests that while some brands with high exposure have moderate claim rates, others show higher variability, indicating a need for brand-specific risk assessment in car insurance policies.



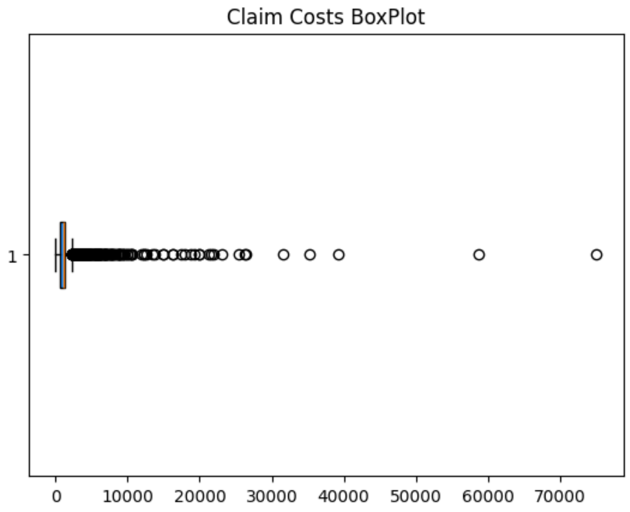
Going for the Type of fuel variable, this graph compares claim frequency and exposure for vehicles using different types of fuel (D and E). The blue bars indicate exposure levels, which are similar for both fuel types. The red line shows the annual claim frequency, with type D having a higher claim frequency than type E. The error bars indicate variability, with type D having a broader range. This suggests that vehicles using fuel type D are associated with a higher risk of claims compared to those using fuel type E, which could be an important factor for insurers to consider in risk assessment and premium calculations.

Task 2 – Severity of Claims

The analysis of claims severity for third-party liability in automobile insurance revealed a right-skewed distribution with significant variability. The mean claim amount is 1715.51, higher than the median of 1172.00, indicating a few high-value claims skewing the average. Most claims are modest, as shown by the 25th and 75th percentiles (662 and 1310, respectively), but the data includes extreme outliers, with claims reaching up to 75,000. This pattern underscores the need for robust risk management to account for occasional but substantial claims impacting overall financial performance.

```
count     1924.000000
mean      1715.510655
std       3449.713150
min          0.010000
25%        661.997500
50%       1172.000000
75%       1309.690000
max      75000.000000
```

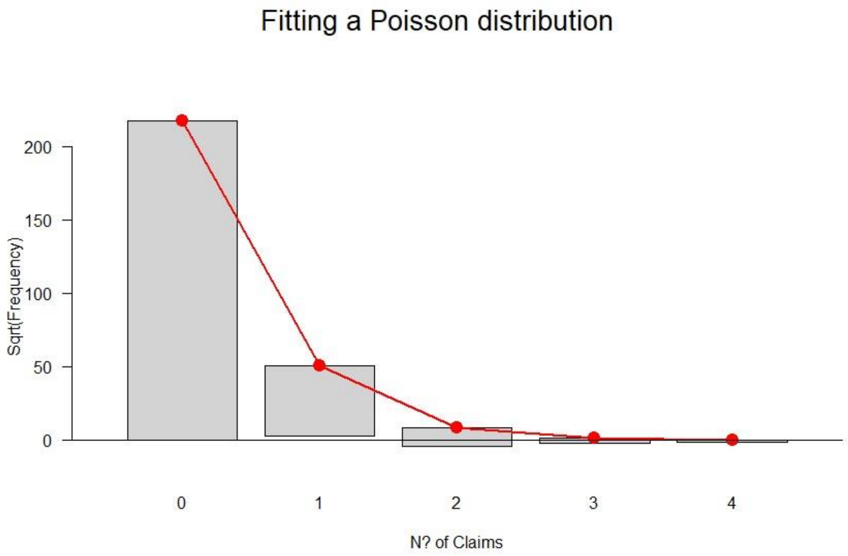
The box plot of claim costs reveals a high concentration of data points at the lower end of the cost spectrum, with a considerable number of outliers extending towards higher values. Most claims are clustered below approximately 10,000, as indicated by the compact box and whiskers. However, the presence of numerous outliers, some exceeding 70,000, underscores the skewed nature of the data. This visualization highlights the variability in claim costs and the impact of high-cost outliers on the overall distribution, emphasizing the need for careful risk management and pricing strategies in automobile insurance.



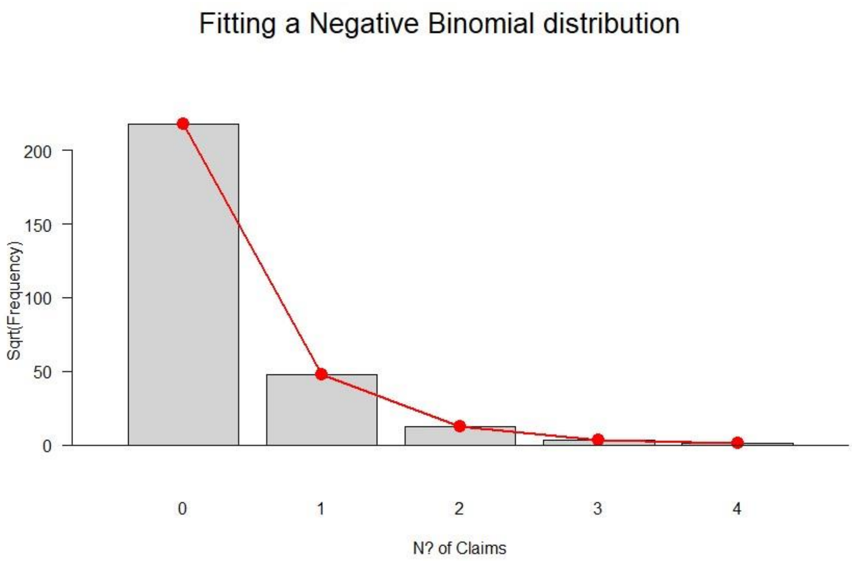

```
0.50     1172.0000
0.90     2963.2200
0.95     5057.5960
0.99    16018.4803
```

Task 3 – Fitting of distributions

For the Number of Claims model, we tested the Poisson distribution, and the binomial negative distribution. The Poisson distribution didn´t fit our data as you can see in the following graph the bars do not align at all with the distribution, most of them being misplaced.



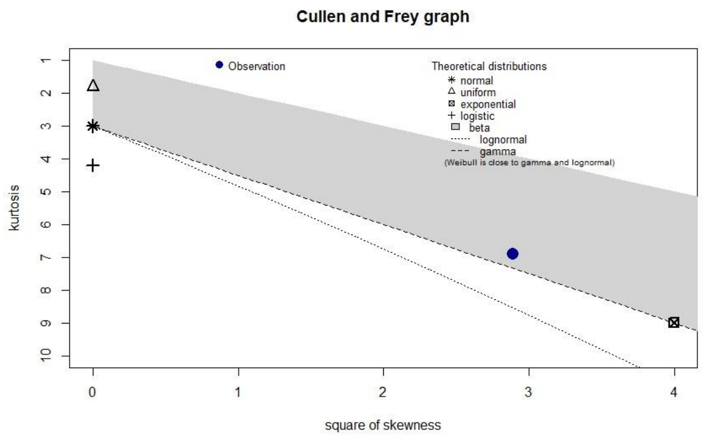
For the Negative Binomial distribution, everything fit perfectly, and this was the distribution fitted to our data.

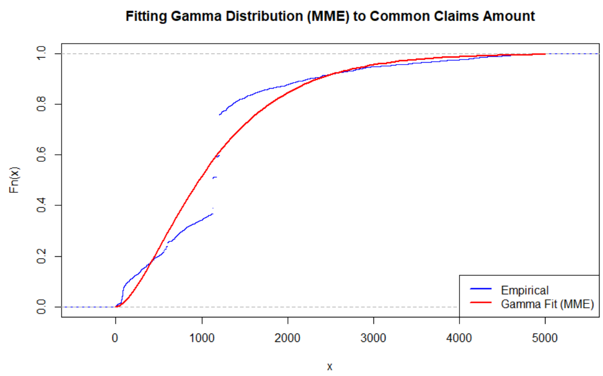


Before fitting we removed the highest outliers in our data as seen in this code snippet.

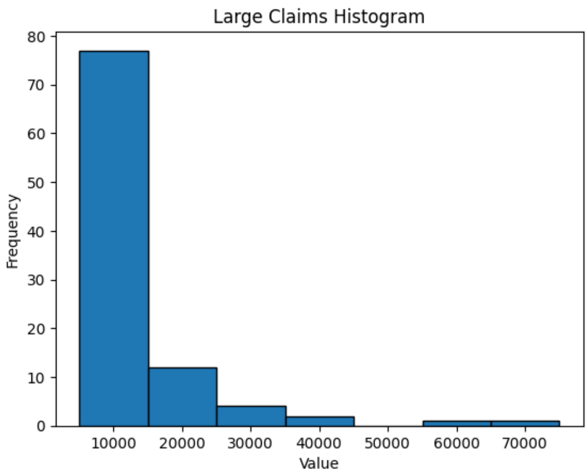
```
#remove outlier in nclaims
baseFREQ <- subset(baseFREQ, subset=nclaims<max(baseFREQ$nclaims))
```

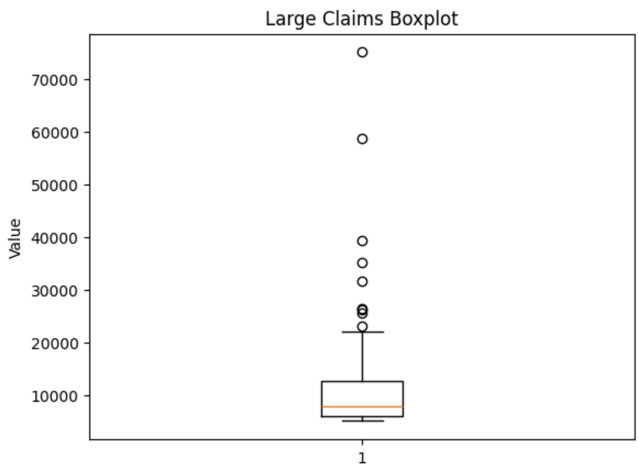
For the distribution fitting of common claims, we first did the Cullen and Frey graph to see if there was an obvious answer to which distribution we should use but got inconclusive results which led us to do some Method of Moments Estimation (MME) tests which led us to conclude that the Gamma distribution was the one that fitted best as it was the one best mirrored in the tests done.





Regarding the common claim limit, we first chose an upper bound to allow us to fit a distribution, the upper bound chosen was 5060€ because as seen before 95% of the claims cost is below 5057.59 so we rounded it up to form a limit. This removed 97 claims from the data, having a mean of 11864.70€ and a standard deviation of 10681.29, these will be our large claims dataset.





The analysis of large claims in automobile insurance reveals that most claims cluster around 10,000, with a sharp decline in frequency for higher amounts. The box plot indicates that most claims fall between 9,500 and 15,000, with significant outliers extending up to 75,000. This right-skewed distribution highlights the presence of high-severity, low-frequency claims that can impact the overall cost structure. These patterns underscore the importance of robust risk management and premium calculation strategies to account for the monetary impact of large claims.

To incorporate these cases into the final premium structure, insurers should segment policyholders into different risk categories and formulate two types of claims, each with a respective probability of occurring. The Pure Premium is then the expected value of a common claim multiplied by its probability summed with the expected value of a large claim multiplied by its probability, and all of it is then multiplied by the expected frequency to form the Pure Premium.

Part 2 – Pricing Structure

Task 1 – Fitting GLM to the Number of Claims

The first step we took to improve and simplify the model was categorizing every variable so that it could be grouped into tariffs. **Please note that for the Number of Claims GLM model, we will use a Negative Binomial GLM (shown to be the best fit) with an offset of the logarithm of the exposition.**

Starting with the age of the driver, we observed that lower ages are riskier, specifically from the 18 to 28 years age bracket, but in our view, there is a reasonable discrepancy between someone who's 18 and 28. Therefore, we decided to cut ages from 18 to 28 into smaller 3- or 4-year gaps and then aggregate whenever appropriate, leaving a 10-year gap for the remaining ages.

In other words, this resulted in the following age brackets:

```
> agedriver_levels <- c(18,22,25,28,31,41,51,61,71,81,101)
> baseFREQ$agecut <- cut(baseFREQ$agedriver,breaks=agedriver_levels,right=FALSE)
> agedriver_levels
 [1]  18  22  25  28  31  41  51  61  71  81 101
> levels(baseFREQ$agecut)
 [1] "[18,22)"  "[22,25)"  "[25,28)"  "[28,31)"  "[31,41)"  "[41,51)"
 [7] "[51,61)"  "[61,71)"  "[71,81)"  "[81,101)"
>
```

For categorizing the power, brand, zone, and fuel variables, we opted by making them factors, since they represent "categories" by their value nature.

For the vehicle age, we took a less cautioned approach compared to the age of the driver. The initial analysis of the claim frequency per vehicle age revealed that most claims are done on cars up to 20 years old, then dropping to an average claim frequency below the mean claim frequency of the whole vehicle age feature. We then decided to try and capture the subgroups seen in the initial exposition/claim frequency histogram.

```
> agevehicle_lev<-c(0,4,11,16,20,101)
> baseFREQ$vehcut<-cut(baseFREQ$agevehicle,breaks=agevehicle_lev,right=FALSE)
> levels(baseFREQ$vehcut)
[1] "[0,4)"    "[4,11)"    "[11,16)"   "[16,20)"   "[20,101)"
>
```

The next step we took, now that all variables are categorical, was, before aggregating levels to simplify the model, deciding which variables are not needed. Not all features are statistically relevant to our model. To test them out, we decided to see a summary of the p-value of each feature if it were to explain alone – plus the logarithm of the exposition offset - the "nclaims" target variable. High p-values (below 0.05), in this case, would indicate no statistical relevancy towards the target variable – ***we will call this the GLM test, for simplicity's sake***. **Performing this for all variables, we concluded that the age of the vehicle (p-value of 0.876) and the power (p-value of 0.273) are *possible* removals.**

It's not enough to simply remove a variable entirely because of the previous step. What if there are sub-features/categories within the feature that can prove useful? To decide finally, we used an ANOVA Chi-Square test for a model containing all, but the variable tested for removal against a model containing all variables. **If the p-value is low, it means there is no significant improvement in removing that variable.**

Finally, doing the previous step we concluded: that removing the age of the vehicle feature would not be beneficial, and thus decided to keep it (p-value of 0.009); we would remove the "power" feature was feasible since it does not hold much statistical significance (p-value of 0.405).

The last step to improve and finalize our model, and further simplify it, was grouping tariffs that weren't dissimilar. We started with the "zone" feature **and decided to make Zone C the standard insured zone characteristic since it's the one with the most total exposition.**

Using the GLM test would be slow since we would have to test for each characteristic within a feature, relevel to it, and then see possible groupings. **To perform this quicker, we used the Generalized Linear Hypothesis (GLH) tests to tell us groupings, finding that Zone A & Zone B are very identical (this could be also seen in the initial histogram plots). We opted to cluster them.**

For the fuel feature, there is no necessary work since both Gasoline and Diesel features are already significantly different. For this feature, the standard insured is assumed to have a Diesel engine.

For the brand feature, we took the same approach as for "zone":

- Brands 10 and 11 show a similar risk profile when looking at previous plots, and GLH proves that they are not significantly different. We decided to group them.
- Brands 1 and 4 are similar as well, so we grouped them.
- Brand 13 is like 10 and 11, so we bundled the 3 together to a single tariff composed of all 3.
- Brands 3 and 5 are also similar, so we joined them.
- Brand 1 has a high exposition and thus was made standard insured.
-

For the age of the driver, we used the previous techniques of GLM and GLH testing. **We defined the standard insured as the 18-22 age bracket since it's the most risk-exposed. We found the following (and consequently decided to make grouping decisions matching the findings):**

- The 18-22 age bracket could be coupled with the 22-25 age bracket. This makes sense since they are the same generation and thus will probably have similar behaviors. But it stops there, coupling 18-25 with the 25-28 age bracket would be bad since they are significantly different.
- The age bracket 25-28 could be merged with the 28-31 age bracket, according to GLH (Tukey Contrasts).
- There are more groupings, namely the 31-51 age bracket with the 51-71 age bracket, etc. However, it would result in 20-year gaps which is too much (a generation is about 20 years). We would be joining completely different generations, so we stopped there.

For the age of the vehicle feature, there are groupings, but they would result in large gaps inside the grouping, so we decided to not do anything regarding them. We defined standard insured as the 0-4 vehicle age bracket.

With this, we have our final model whose coefficients can be seen below or in the pricing structure. Our standard insured risk profile is riskier than the average, made in a way to account for as much risk as we could reasonably expect for an average client.

```
> results_nclaims_model
                    Estimate
(Intercept)      -1.67938083
zoneD             0.23397721
zoneE             0.40675555
zoneF             0.41533622
zoneA;B          -0.16554756
brand2            0.05532466
brand6           -0.11715936
brand12          -0.13324057
brand14          -0.21362289
brand10;11;13     0.26427886
brand3;5          0.16295670
fuelE            -0.16885304
agecut[31,41)    -0.68124040
agecut[41,51)    -0.67231526
agecut[51,61)    -0.79561490
agecut[61,71)    -0.79354247
agecut[71,81)    -0.85028884
agecut[81,101)   -1.05206750
agecut[25,31)    -0.57661273
vehcut[4,11)      0.10430701
vehcut[11,16)     0.02603449
vehcut[16,20)    -0.21915696
vehcut[20,101)   -0.12321290
```

<u>**In short, our standard insured: lives in Zone C; drives a car of Brands 1, 4, 7, 8, or 9; has a Diesel Engine; is in the 18-25 age bracket; his vehicle is 0-4 years old; the power of the engine is Power 4.**</u>

<u>**We expect the standard insured to have an expected number of claims of 0.186 claims per year.**</u>

Task 2 – Fitting a GLM to the Severity/Cost of "common" claims

For this task, no new groupings were done. Instead, the groupings will match the one in the Number of Claims model to make it possible to formulate a Pricing Structure. What will be done is to check if all features are necessary, and remove the ones that show significance – again, using the previous tests like ANOVA Chi-square, etc. **The GLM model that is going to be fitted and used for GLM tests is a Gamma GLM with a log link function, with <u>no</u> exposition as the offset.**

Performing a GLM test to see the p-value of each feature, **interestingly almost all of them exceed a threshold of 0.05. However, this does not mean we will remove all of them, but we can eliminate more variables. The first step taken was to categorize all variables and group them in the same manner as for the Number of Claims model.**

We saw that statistically significant variables are the age of the driver and the age of the vehicle, both having a p-value on this test below 0.05. Performing an ANOVA Chi-Square comparing the full Common Claims Model against a model with all features except the one being tested for removal showed that we can remove "zone" and other features, **however, we just removed the "zone" feature since the others seemed reasonable features to estimate a Claim Cost.**

With this, we have obtained the following coefficients for the Common Claims GLM:

(Intercept)	7.277704804
power5	-0.075188989
power6	-0.141010225
power7	-0.102385512
power8	-0.050447159
power9	-0.276240592
power10	-0.083611171
power11	-0.110995543
power12	-0.231768403
power13	0.094797056
power14	-0.071833461
power15	-0.172567344
brand2	0.043494699
brand6	-0.071770064
brand12	0.127862102
brand14	-0.146287625
brand10;11;13	0.024495074
brand3;5	-0.042941897
agecut[31,41)	-0.210359492
agecut[41,51)	-0.166014801
agecut[51,61)	-0.148053399
agecut[61,71)	-0.247046452
agecut[71,81)	-0.011402594
agecut[81,101)	-0.315508860
agecut[25,31)	-0.267184201
vehcut[4,11)	0.053337390
vehcut[11,16)	-0.004808443
vehcut[16,20)	0.231345458
vehcut[20,101)	-0.173907977

Our standard insured has the same previous risk profile as the Number of Claims model and has an expected claim value of 1447.66, which combined with the frequency of 0.186 claims per year means he would have a Pure Premium of 269.97 euros.

Task 3 – Proposed Common Claims Price Structure

Our proposed Common Claims Price Structure can be seen below or in the attached “final_pricing_structure” file.

Risk Profile	beta_N	beta_Y	E(N)	E(Y)	Pure Premium	Tariff
Standard Insured	**-1.679381**	**7.2777048**	**0.186489**	**1447.662**	**269.9735453**	**269.9735**
Zone D	0.233977207	0	0.235650943	1447.661542	341.1428079	1.263615691
Zone E	0.406755548	0	0.280095326	1447.661542	405.4832322	1.501936909
Zone F	0.415336221	0	0.282509074	1447.661542	408.9775217	1.51487999
Zone A & B	-0.165547563	0	0.158036637	1447.661542	228.7835617	0.847429556
Brand 2	0.055324662	0.043494699	0.197097615	1512.016551	298.0148554	1.103866881
Brand 6	-0.117159364	-0.071770064	0.165871781	1347.403572	223.49623	0.827844927
Brand 12	-0.133240575	0.127862102	0.163225695	1645.117198	268.5253977	0.994635965
Brand 14	-0.213622895	-0.146287625	0.150618711	1250.648104	188.371006	0.697738757
Brand 10 & 11 & 13	0.264278859	0.024495074	0.242900843	1483.559992	360.3579731	1.334789943
Brand 3 & 5	0.162956697	-0.042941897	0.219495372	1386.812056	304.3988274	1.127513538
Fuel E	-0.168853044	0	0.157515112	1447.661542	228.0285703	0.844633018
Age 31-41	-0.681240399	-0.210359492	0.094361585	1173.029869	110.6889575	0.409999274
Age 41-51	-0.672315265	-0.166014801	0.095207544	1226.218108	116.7452146	0.432432054
Age 51-61	-0.795614898	-0.148053399	0.08416335	1248.441691	105.0730347	0.389197522
Age 61-71	-0.793542467	-0.247046452	0.084337953	1130.774814	95.3672335	0.353246587
Age 71-81	-0.85028884	-0.011402594	0.079685338	1431.2482	114.0494971	0.422446936
Age 81-101	-1.052067503	-0.31550886	0.065124899	1055.94982	68.76862514	0.25472357
Age 25-31	-0.576612731	-0.267184201	0.104769397	1108.231303	116.1087256	0.430074456
Vehicle Age 4-11	0.10430701	0.05333739	0.206992272	1526.972344	316.0714747	1.170749802
Vehicle Age 11-16	0.026034494	-0.004808443	0.191408319	1440.717253	275.7652678	1.021452926
Vehicle Age 16-20	-0.219156963	0.231345458	0.149787479	1824.479941	273.2842517	1.012263077
Vehicle Age 20-101	-0.123212898	-0.173907977	0.164870703	1216.577451	200.57798	0.742954202
Power 5	0	-0.075188989	0.186489409	1342.804767	250.4188673	0.92756817
Power 6	0	-0.141010225	0.186489409	1257.26572	234.4667411	0.868480431
Power 7	0	-0.102385512	0.186489409	1306.777278	243.7001221	0.90268149
Power 8	0	-0.050447159	0.186489409	1376.44263	256.6919725	0.950804169
Power 9	0	-0.276240592	0.186489409	1098.240038	204.8101356	0.758630389
Power 10	0	-0.083611171	0.186489409	1331.542912	248.3186506	0.919788827
Power 11	0	-0.110995543	0.186489409	1295.574184	241.6108638	0.894942738
Power 12	0	-0.231768403	0.186489409	1148.183492	214.1240608	0.793129788
Power 13	0	0.094797056	0.186489409	1591.610793	296.818556	1.099435708
Power 14	0	-0.071833461	0.186489409	1347.318154	251.2605662	0.930685879
Power 15	0	-0.172567344	0.186489409	1218.209528	227.1831748	0.841501617

The Standard Insured risk profile: lives in zone C; drives a vehicle of the brands 1, 4, 7, 8, or 9; has a diesel engine; is 18 to 25 years old; his vehicle is relatively recent, between 0 to 4 years old; has an engine with power 4. Its Pure Premium is calculated at 269.97 euros per year.

The Highest Risk and Lowest Risk profiles can be seen below, along with their Pure Premiums:

Highest Risk Profile Base Premium	269.9735
Zone F	1.51488
Brand 10 & 11 & 13	1.33479
Fuel D	1
Age 18-25	1
Vehicle Age 4-11	1.17075
Power 13	1.099436
Pure Premium	702.6617

Lowest Risk Profile Base Premium	269.9735
Zone A & B	0.84743
Brand 14	0.697739
Fuel E	0.844633
Age 81-101	0.254724
Vehicle Age 20-101	0.742954
Power 9	0.75863
Pure Premium	19.3574

The highest risk profile pays over twice the premium of the Standard Insured due to these factors: they live in the riskiest area (zone F), drive the most dangerous brands (10, 11, 13), have slightly older cars (4-11 years), and use high-power engines (power 13), despite having the same age and engine type. In essence, they’re young, inexperienced drivers with powerful cars.

Conversely, the lowest risk profile is a senior (81+ years) living in the safest areas (zones A and B), driving low-risk, gasoline-powered brands with old cars. This suggests minimal driving and, therefore, lower risk exposure.

Task 4 – Machine Learning Model for Large Claims & Complete Pricing Structure

This task can be divided into two: estimate the probability of making a large claim **<u>(in our case, more than 5060 euros</u>**); and estimate the average cost of a large claim. The integration with the Common Claims model is simply a sum of the expected value of each type of claim.

To start with the easiest part: determining the average cost of a large claim. For simplicity's sake, we've opted to separate the Large Claims set from the Common Claims and fit the "cost" feature to an extreme values distribution. We found that the Pareto distribution is a good fit for this type of claim, then using its parameters calculated the average Large Claim Cost according to the Pareto Expected Value formula.

```
> gofstat(pareto_fit)
Goodness-of-fit statistics
                             1-mle-pareto
Kolmogorov-Smirnov statistic   0.07914557
Cramer-von Mises statistic     0.04921643
Anderson-Darling statistic     0.55292122

Goodness-of-fit criteria
                               1-mle-pareto
Akaike's Information Criterion     1889.324
Bayesian Information Criterion     1894.473
> #we can see that the p-values are above 0.05, indicate it follows a pareto dist.
>
> alpha = pareto_fit$estimate["shape"]
> m = pareto_fit$estimate["scale"]
> pareto_mean = m*alpha/(alpha-1)
> print(paste("Large Claims Mean Cost with Pareto dist. ->", pareto_mean))
[1] "Large Claims Mean Cost with Pareto dist. -> 13995.559902374"
```


The second and last step is fitting it to a Machine Learning Model. We have chosen the Logistic Regression to estimate the probability of making a large claim since it's a robust and commonly used model for risk applications. To fit it, we used all features of the Number of Claims Model, and the Common Claims Model, adding a target feature that indicates if the cost is above the Common Claim limit of 5060 euros.

Fitting it to a Logistic Regression Machine Learning Model, we obtained the following odds ratio:

```
> exp(coefficients(reglogit_model))
   (Intercept)        zoneD         zoneE         zoneF       zoneA;B
  1.951002e-02  2.295543e+00  1.940468e+00  1.729536e+00  1.287498e+00
        power5        power6        power7        power8        power9
  1.076537e+00  9.318494e-01  7.934638e-01  7.489169e-01  1.585449e+00
       power10       power11       power12       power13       power14
  4.627871e-01  2.728011e-01  5.135275e-01  4.112443e-07  5.208426e-07
       power15  vehcut[4,11) vehcut[11,16) vehcut[16,20) vehcut[20,101)
  3.828902e+00  9.057227e-01  7.762046e-01  5.338095e-01  7.233674e-07
 agecut[31,41) agecut[41,51) agecut[51,61) agecut[61,71) agecut[71,81)
  7.986775e-01  9.062123e-01  8.548134e-01  2.457092e-01  8.071429e-01
agecut[81,101) agecut[25,31)        brand2        brand6       brand12
  1.867779e+00  6.090785e-01  1.395321e+00  1.472352e+00  3.364675e+00
       brand14 brand10;11;13      brand3;5         fuelE
  6.204441e+00  2.061245e+00  1.308378e+00  6.628494e-01
```

Now, we can actively predict the probability of making a large claim. For the Standard Insured, our model suggests roughly a 5% probability of making a large claim:

```
> #Compute probabilities of standard insured
> get_std_insured_rp <- function() {
+   return(data.frame(zone="C", power=as.factor(4), vehcut="[0,4)", agecut="[18,25)",
brand=as.factor(1), fuel="D", exposition=1))
+ }
> predict_risk_profile(get_std_insured_rp())
         1
0.05036281
```

Then, for other features, we did it one by one. For example, for “Zone D”, we assumed all features to have the Standard Insured characteristics, but the zone would be Zone D instead of Standard Insured Zone C. Doing this, we can formulate the final pricing model seen below:

Risk Profile	E(N)	E(Y | Y <= 506	E(Y | Y > 50	P(Y > 50	E(Y)	Pure Premi	Tariff
Standard Insured	0.186489409	1447.661542	13995.5605	0.0503628	2079.609	387.8250518	387.8250518
Zone D	0.235650943	1447.661542	13995.5605	0.1085288	2809.47	662.0542453	1.707095099
Zone E	0.280095326	1447.661542	13995.5605	0.0933079	2618.4799	733.4239803	1.891120692
Zone F	0.282509074	1447.661542	13995.5605	0.0840174	2501.9028	706.8102316	1.822497614
Zone A & B	0.158036637	1447.661542	13995.5605	0.0639166	2249.68	355.531854	0.916732564
Brand 2	0.197097615	1512.016551	13995.5605	0.0689005	2372.1392	467.5429815	1.205551264
Brand 6	0.165871781	1347.403572	13995.5605	0.0724288	2263.4941	375.4498049	0.968090646
Brand 12	0.163225695	1645.117198	13995.5605	0.1514214	3515.2386	573.7772647	1.479474475
Brand 14	0.150618711	1250.648104	13995.5605	0.2475799	4406.0322	663.6308981	1.711160471
Brand 10 & 11 & 13	0.242900843	1483.559992	13995.5605	0.0985432	2716.5326	659.8480493	1.701406462
Brand 3 & 5	0.219495372	1386.812056	13995.5605	0.0648859	2204.9418	483.974518	1.247919689
Fuel E	0.157515112	1447.661542	13995.5605	0.0339596	1873.7829	295.1491269	0.761036776
Age 31-41	0.094361585	1173.029869	13995.5605	0.0406357	1694.0817	159.8562374	0.412186466
Age 41-51	0.095207544	1226.218108	13995.5605	0.045856	1811.7689	172.4940717	0.444772897
Age 51-61	0.08416335	1248.441691	13995.5605	0.0433679	1801.2576	151.5998727	0.390897576
Age 61-71	0.084337953	1130.774814	13995.5605	0.0128633	1296.2579	109.323738	0.281889315
Age 71-81	0.079685338	1431.2482	13995.5605	0.0410487	1946.9966	155.1470857	0.400044002
Age 81-101	0.065124899	1055.94982	13995.5605	0.0901277	2222.167	144.7184034	0.37315383
Age 25-31	0.104769397	1108.231303	13995.5605	0.031291	1511.4882	158.3577081	0.408322534
Vehicle Age 4-11	0.206992272	1526.972344	13995.5605	0.0458324	2098.437	434.3602503	1.119990182
Vehicle Age 11-16	0.191408319	1440.717253	13995.5605	0.0395375	1937.104	370.7778189	0.956044013
Vehicle Age 16-20	0.149787479	1824.479941	13995.5605	0.0275305	2159.5562	323.4744858	0.83407321
Vehicle Age 20-101	0.164870703	1216.577451	13995.5605	0	1216.5775	200.57798	0.517186755
Power 5	0.186489409	1342.804767	13995.5605	0.0540092	2026.1702	377.8592893	0.974303459
Power 6	0.186489409	1257.26572	13995.5605	0.0470922	1857.1398	346.3369018	0.893023543
Power 7	0.186489409	1306.777278	13995.5605	0.0403811	1819.1643	339.2548752	0.874762663
Power 8	0.186489409	1376.44263	13995.5605	0.0382006	1858.5008	346.5907074	0.893677976
Power 9	0.186489409	1098.240038	13995.5605	0.0775608	2098.5664	391.3604076	1.009115852
Power 10	0.186489409	1331.542912	13995.5605	0.0239554	1634.9143	304.8941949	0.78616426
Power 11	0.186489409	1295.574184	13995.5605	0.0142613	1476.6929	275.3875823	0.710081984
Power 12	0.186489409	1148.183492	13995.5605	0.0265122	1488.7962	277.6447298	0.715901999
Power 13	0.186489409	1591.610793	13995.5605	0	1591.6108	296.818556	0.765341369
Power 14	0.186489409	1347.318154	13995.5605	0	1347.3182	251.2605662	0.647870902
Power 15	0.186489409	1218.209528	13995.5605	0.1687869	3374.859	629.3754574	1.622833426

Given the inclusion of Large Claims, we can see that all premiums have changed, generally increasing. Therefore, the extreme risk profiles (highest and lowest risk) have changed as well:

Highest Risk Profile Base Premium	387.825
Zone E	1.89112
Brand 14	1.71116
Fuel D	1
Age 18-25	1
Vehicle Age 4-11	1.11999
Power 15	1.62283
Pure Premium	2281.05

Lowest Risk Profile Base Premium	387.825
Zone A & B	0.91673
Brand 6	0.96809
Fuel E	0.76104
Age 61-71	0.28189
Vehicle Age 20-101	0.51719
Power 14	0.64787
Pure Premium	24.7409