Car insurance premium calculation

ECMI Modelling Week 2023, Szeged

Instructor: Dora Selesi

Arianna Zanin
Koppány Dodony
Filip Rydin
Tiago Miranda
Roeland Hooijmans

Motivation

For insurance companies it is crucial to determine the right premiums for policyholders.

Why?

- Fair price that reflects risk associated to a policyholder
- Financial sustainability in a competitive market.

Challenges

Estimating the level of ‘risk’ of policyholders without discriminating people.

How?

Explainable and statistically rigorous methods.

Introduction

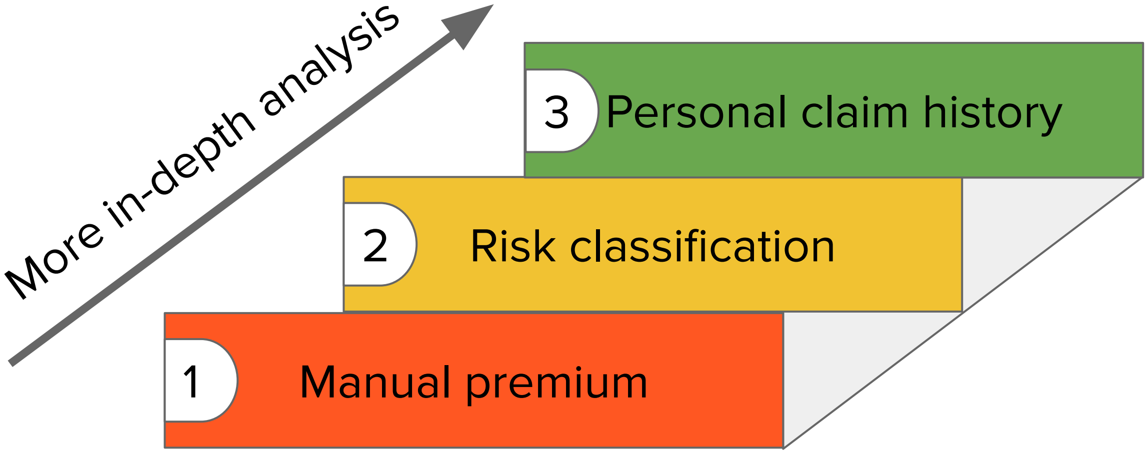
- Premium is calculated based on the total claim size (S). X_i represents the i-th claim of an individual policyholder and N represents the number of claims that a policyholder files per year.

$$S = X_1 + X_2 + \cdots + X_N$$

- Assumptions:
  - x_i's are i.i.d. nonnegative continuous random variables
  - N is a discrete nonnegative random variables

Insurance companies wants to make predictions for the total claim amount of their policyholders to set the right premium.

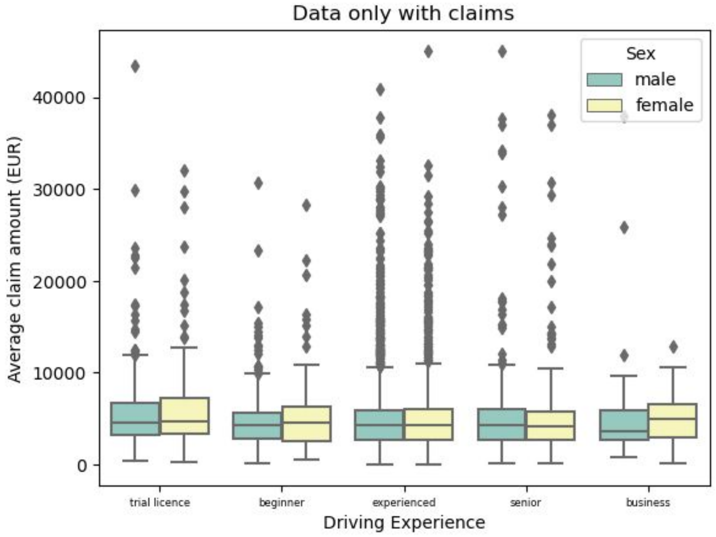
Project Outline



Data (1)

Information of 100.000 policyholders of the insurance company BlackCat in the state of Bandrika.

Numerical Features	*Categorical Features*
Income (EUR)	Sex
No. of family members	Territory
Mileage (km)	Driving Experience
Number of claims	Education
Claim amount (per claim)	Vehicle production year
	Vehicle color
	Manufacturer



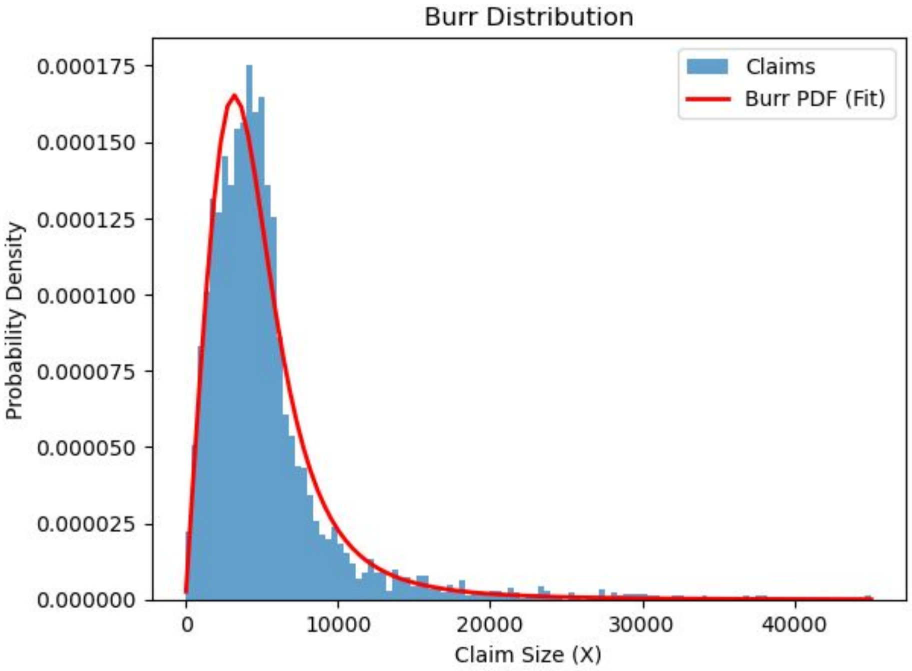
1. Manual premium (M) calculation

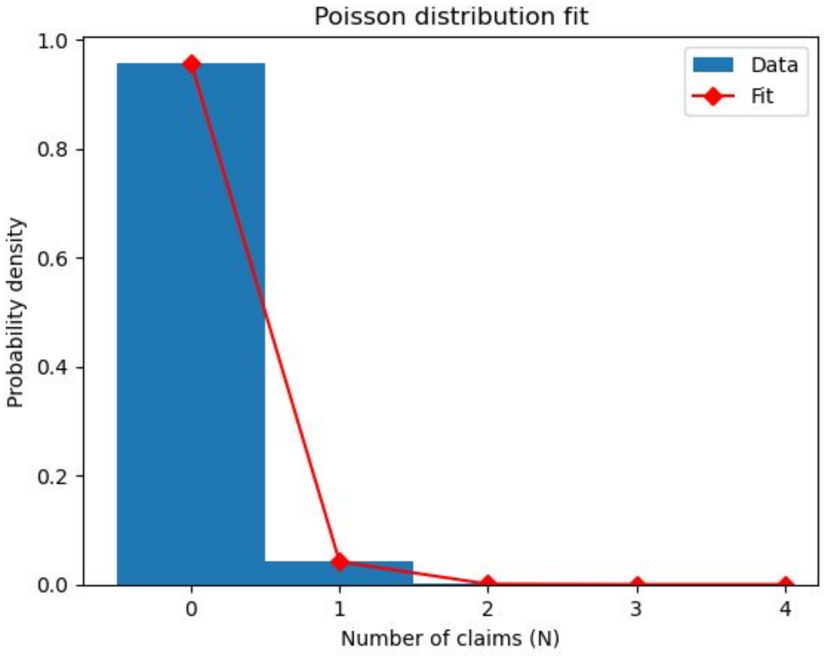
$$M = E[S] = E[N]\ E[X]$$

$$M_S(t) = E[(M_X(t))^N] = G_N(M_X(t))$$

Final result: M=241€

1. Fitting distributions for X and N





Data (2)

Sex	Territory	Driving Experience	Education	Income (EUR)	No. of family members	Vehicle production year	Vehicle color	Manufacturer	Mileage (km)	Nuber of claims	1. claim amount	2. claim amount	3. claim amount	4. claim amount	Total claim amount (EUR)
male	B	experienced	bachelor	3009.0	1.0	before 2007	yellow	BMW	14831.0	0.0	NaN	NaN	NaN	NaN	0
female	A	experienced	high school	3331.0	3.0	before 2007	red	Fiat	4789.0	0.0	NaN	NaN	NaN	NaN	0
female	A	experienced	high school	1229.0	1.0	2018-2022	yellow	Hyundai	2994.0	0.0	NaN	NaN	NaN	NaN	0
female	A	experienced	master degree	3513.0	1.0	2018-2022	black	Skoda	14600.0	0.0	NaN	NaN	NaN	NaN	0
male	A	experienced	master degree	3597.0	1.0	2018-2022	grey	Opel	8288.0	0.0	NaN	NaN	NaN	NaN	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

2. Risk classification

- Correlation analysis (Pearson's coefficient)
- Discretization of continuous variables
- One-way ANOVA testing
- Two-way ANOVA testing
- Tukey testing
- Contingency tables and χ^2 testing

Nuber of claims	0.0	1.0	2.0	3.0	4.0	All
Driving Experience						
trial licence	3134	288	28	3	0	3453
beginner	3915	261	13	0	1	4190
with experience	88624	3593	136	4	0	92357
All	95673	4142	177	7	1	100000

Chi_square value 359.60920790077495

p value 2.0503261701464603e-67

2. Risk classification

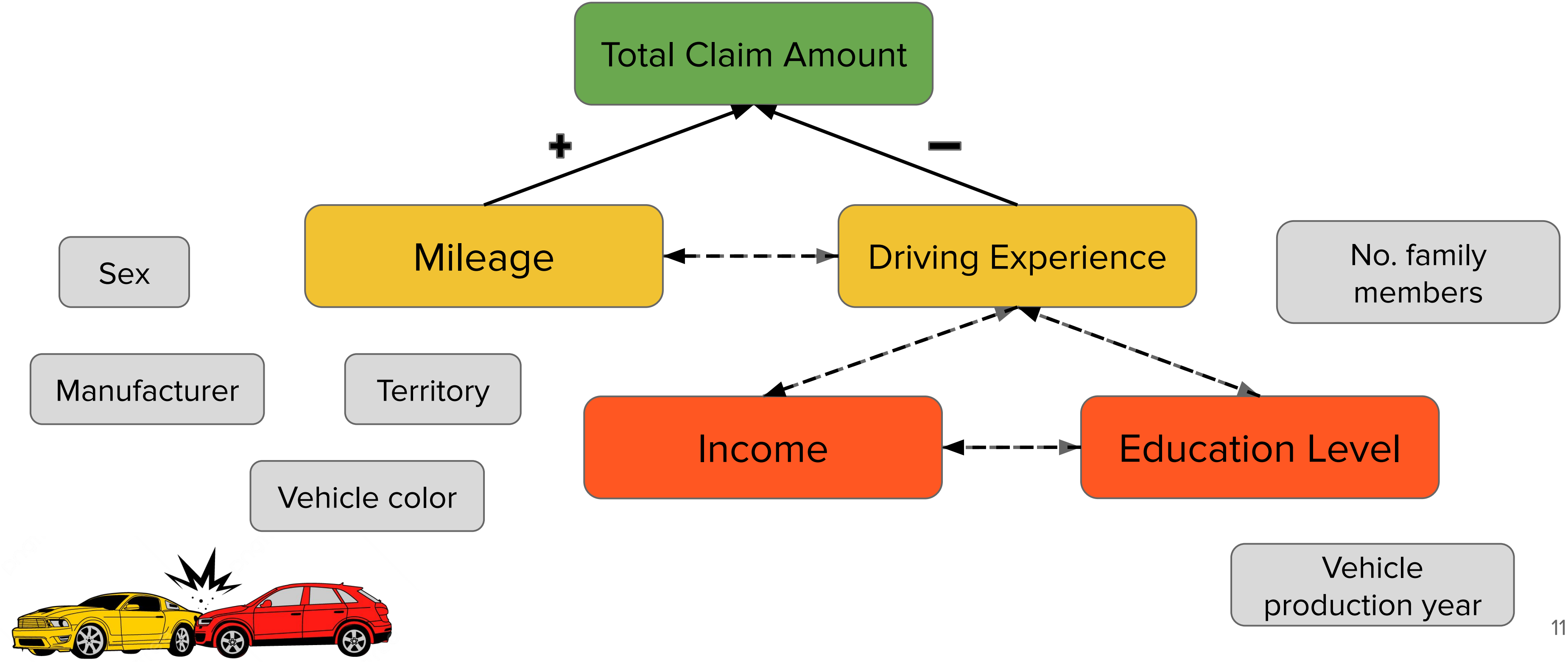
- Strong correlation between Driving Experience, Education and Income.
- Moderate correlation between Mileage and Driving Experience.
- Experience and Mileage are the best predictors for Number of Claims and Total Claim Amount.
- Other features (Sex, Manufacturer, Vehicle Color, Vehicle Production Year, etc.) showed no significant influence on Number of Claims or Total Claim Amount.
- Both Driving Experience and Mileage showed significant difference in mean Total Claim Amount between groups.
- The interaction/combination between Driving Experience and Mileage showed no significant influence on Total Claim Amount.

Nuber of claims	0.0	1.0	2.0	3.0	4.0	All
Driving Experience						
trial licence	3134	288	28	3	0	3453
beginner	3915	261	13	0	1	4190
with experience	88624	3593	136	4	0	92357
All	95673	4142	177	7	1	100000

Chi_square value 359.60920790077495

p value 2.0503261701464603e-67

2. Risk classification



2. Risk class clustering

Whole population
premium = **241€**

Trial licence
premium = **614€**

3.4 **%** of drivers

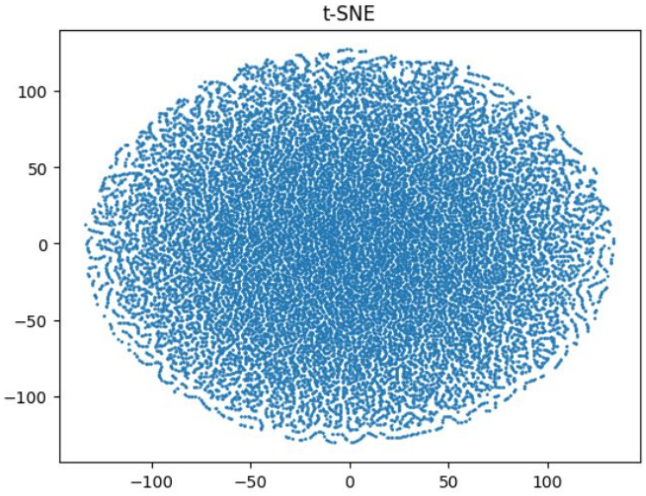
Beginner drivers
premium = **359€**

4.2 **%** of drivers

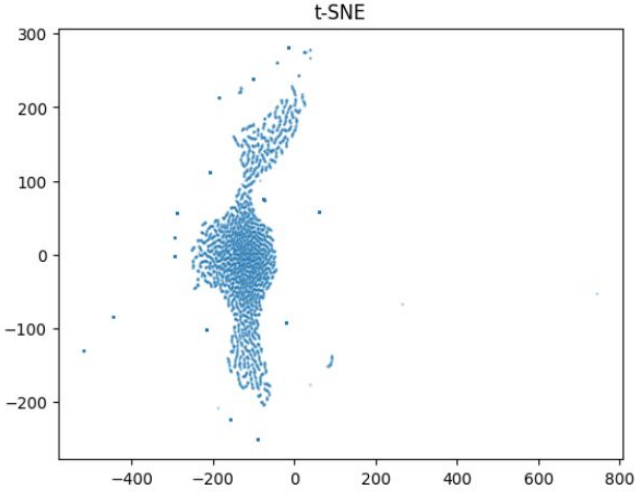
Experienced drivers
premium = **221€**

92.4 **%** of drivers

2. Risk classification



T-SNE whole database



T-SNE on driving experience and number of claims

3. Accounting for personal claim history

- European credibility

$$\text{Premium} = \mathbb{E}_\Lambda\left[\mathbb{E}[S|\Lambda]\right]$$

- American credibility

$$\text{Premium} = Z\bar{S}_k + (1 - Z)M$$

3. Accounting for personal claim history

European method

$$\text{Premium} = \mathbb{E}_\Lambda\left[\mathbb{E}[S|\Lambda]\right]$$

$$\mathbb{P}\left[\ \Lambda \mid N_1, N_2, \cdots, N_k\right] = \frac{\prod_{i=1}^{k} \mathbb{P}(N_i \mid \Lambda)\ \mathbb{P}(\Lambda)}{\mathbb{P}(N)}$$

American method

$$\text{Premium} = \mathcal{Z}\bar{S}_k + (1-\mathcal{Z})M$$

$$\mathbb{P}(|\text{Premium} - M| \leq rM) \geq p$$

$$\mathcal{Z} = \frac{-\sqrt{k}rM}{\sqrt{Var(S)}\ \Phi^{-1}\left(\frac{1-p}{2}\right)}$$

$$r = 0.05\ p = 0.9$$

3. Accounting for personal claim history

Example

Number of claims yearly: 0, 1, 0, 0, 1, 0, 0, 0, 0, 0
Risk class: Experienced

Experienced risk class
premium = **221€**

European method
premium = **271€**

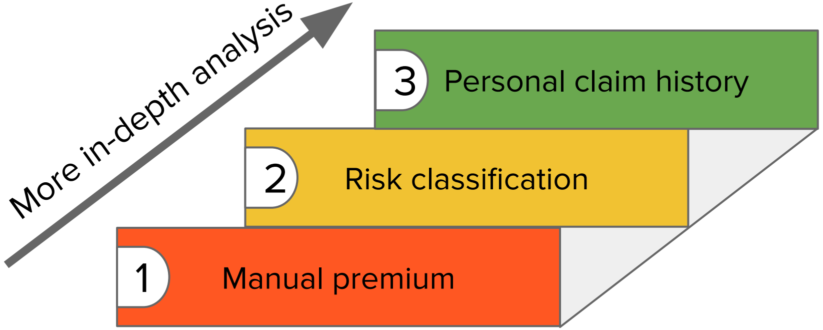
American method
premium = **240€**

3. Accounting for personal claim history

	European method	American method
Initial risk classification	Not required	Required
Explainable results	Yes	Sort of
Susceptible to very risky individuals	Very	No
Bounded premium	Yes	No
Can handle short histories	Yes	Not really

Conclusions

- The manual premium is a good starting point for insurance premiums.
- We found evidence that the essential information needed when profiling individuals into different risk classes is their driving experience.
- Both credibility methods (European and American) are suitable for calculating personalised premiums. It up to the insurance company to decide which method matches their preferences best.



References

Walters, Michael A.. “RISK CLASSIFICATION STANDARDS.” (1999).

Metz, Jason. “How Age and Gender Affect Car Insurance Rates.” Forbes, June 26, 2023. https://www.forbes.com/advisor/car-insurance/rates-age-and-gender/.

Thank you!

Arianna Zanin
Dodony Koppány
Filip Rydin
Roeland Hooijmans
Tiago Miranda