

Fiche technique

Utilisation des GLM en tarification IARD



Juillet 2025

La théorie des GLM

Une prime commerciale est composée d'une prime pure, de chargements et de taxes réglementaires.

La prime pure correspond à un montant estimé par l'assureur qui doit permettre de faire face en moyenne à l'ensemble des prestations attendues.

Elle est généralement estimée par **un modèle fréquence x coût moyen**.

$$\text{Prime pure} = E[N] \times E[C]$$

Où **N** et **C** sont des variables aléatoires représentant respectivement la fréquence et le coût.

Le modèle le plus couramment utilisé en tarification pour estimer la fréquence et le coût moyen est **le modèle linéaire généralisé (GLM)**. Ce modèle permet d'établir une liaison entre une variable aléatoire notée **Y** (dans notre cas, **Y** correspond à la fréquence ou au coût moyen) et **M** variables explicatives déterministes et **indépendantes** X_i .

La liaison entre $E[Y | X_1, \dots, X_M]$ et les variables explicatives est définie par une fonction de lien **g**

$$g(E[Y | X_1, \dots, X_M]) = \beta_0 + \beta_1 X_1 + \dots + \beta_M$$

Il convient alors de choisir deux éléments déterminants pour la construction du modèle :

- ☐ La distribution de la variable aléatoire réponse **Y**.
- ☐ Le choix de la fonction **g** permettant de définir le lien entre la variable aléatoire **Y** et la **combinaison linéaire** des variables explicatives X_i .

Pour estimer le nombre de sinistres (pour rappel, $\text{fréquence} = \frac{\text{Nombre de sinistres}}{\text{exposition}}$), on utilise généralement une loi discrète comme **la loi de Poisson** ou **la loi binomiale négative**.

Pour estimer le coût, on utilise généralement une loi continue, à valeurs positives et asymétriques comme **la loi gamma**.

Pour modéliser la fréquence et le coût moyen, une fonction de lien souvent utilisée est la **fonction logarithmique**. Elle permet une interprétation simple des résultats car chaque variable a un effet multiplicatif sur la variable réponse.

Avantages et Inconvénients



AVANTAGES :

- ✓ Interprétabilité
- ✓ Facilement modulable et modifiable
- ✓ Adaptable à différente distribution
- ✓ Facilité de mise en production (R, Python)
- ✓ Méthode de référence sur le marché



INCONVENIENTS :

- ✓ Dépend fortement de la qualité des données
- ✓ Hypothèse forte d'indépendance entre les variables explicatives
- ✓ Surajustement aux données pouvant nuire au pouvoir prédictif du modèle



Il existe des méthodes alternatives de machine learning (arbres de décision, forêts aléatoires, réseaux de neurones, etc.). Ces méthodes peuvent présenter un gain de pouvoir prédictif mais perdre parfois en interprétabilité.



SAVIEL
ACTUARIAT

Les grandes étapes de la tarification

Cas pratique d'une tarification en assurance MRH

1.

Traitements des bases de données

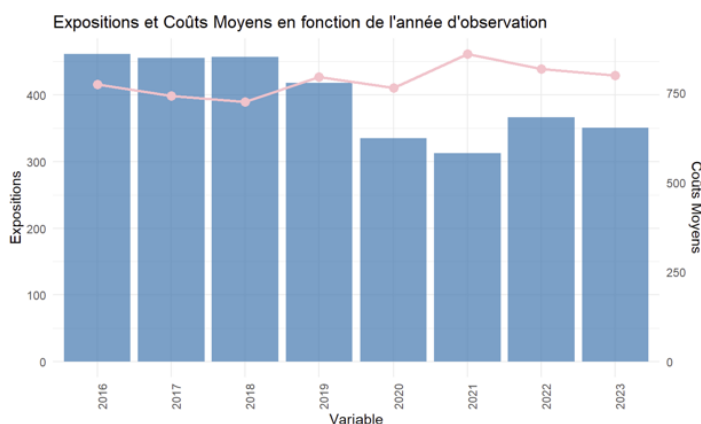
a. Choix de l'historique

Le choix de l'historique est une étape cruciale dans l'exercice de tarification.

La base retenue doit être :

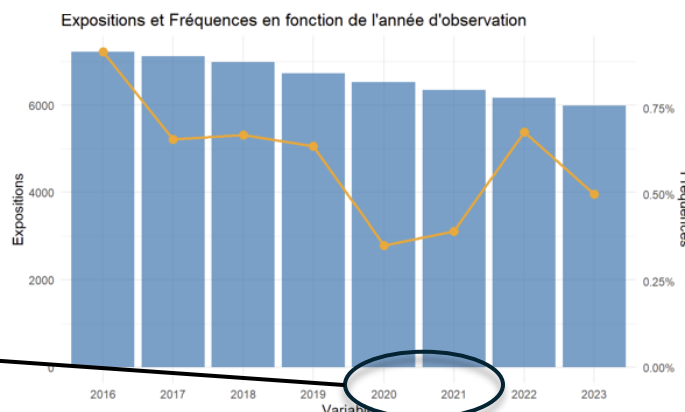
- ☐ Représentative du portefeuille actuel
- ☐ Statistiquement stable
- ☐ Non polluée par des événements exogènes majeurs. (Exemple : COVID, inflation, etc.)

Exemple sur la garantie vol d'un portefeuille MRH :



Le coût moyen est plutôt stable sur l'historique.
✓ Nous conservons toutes les années d'observation.

Conséquence COVID : Le confinement a eu un impact **ponctuel** à la baisse sur la fréquence des cambriolages.
✓ Nous enlevons ces années non représentatives qui pourraient biaiser les prédictions.





Les grandes étapes de la tarification

Cas pratique d'une tarification en assurance MRH

b. Traitement des données aberrantes et manquantes

Nous disposons d'une base de données avec les **variables tarifaires** suivantes :

- Age de l'assuré
- Nature du bien (Maison, Appartement, etc.)
- Numéro étage
- Usage (Résidence principale, secondaire, etc.)
- Qualité de l'occupant (Propriétaire, Locataire, etc.)
- Nombre de pièces
- Superficie habitable du bien
- etc.

○ Tests de cohérence

Quelques exemples simples de **tests de cohérence** et de **retraitements** appliqués aux variables tarifaires :

- ☐ Superficie habitable : vérification que les terrains ou les dépendances ont une superficie égale à 0.
- ☐ Nombre de pièces : vérification qu'un studio a une seule pièce et traitement des valeurs extrêmes (négatives par exemple)
- ☐ Numéro de l'étage :
 1. Zoom sur les valeurs négatives et les étages élevés (> 20)
 2. Vérification que les maisons ont un étage égal à 0
- ☐ Test de cohérence temporelle : stabilité des modalités dans le temps. Par exemple, analyse sur les changements de nature du bien (passage de maison à appartement) .

○ Remplacement des valeurs manquantes

En cas de **valeurs manquantes**, si la valeur de la variable concernée est renseignée pour une autre année d'observation, nous utilisons cette valeur. Dans le cas contraire, nous recourons à des méthodes de substitution, telles que l'utilisation de la moyenne ou de la médiane.



SAVIEL
ACTUARIAT

Les grandes étapes de la tarification

Cas pratique d'une tarification en assurance MRH

c . Traitement des sinistres graves

Distinguer les sinistres graves des sinistres attritionnels : certains sinistres sont particulièrement coûteux et peu fréquents. Ils nécessitent une modélisation particulière pour ne pas biaiser les estimations.

Par exemple, sur l'année d'observation 2021, 5% des sinistres de la garantie incendie représentaient plus de 60% de la charge totale de sinistralité.

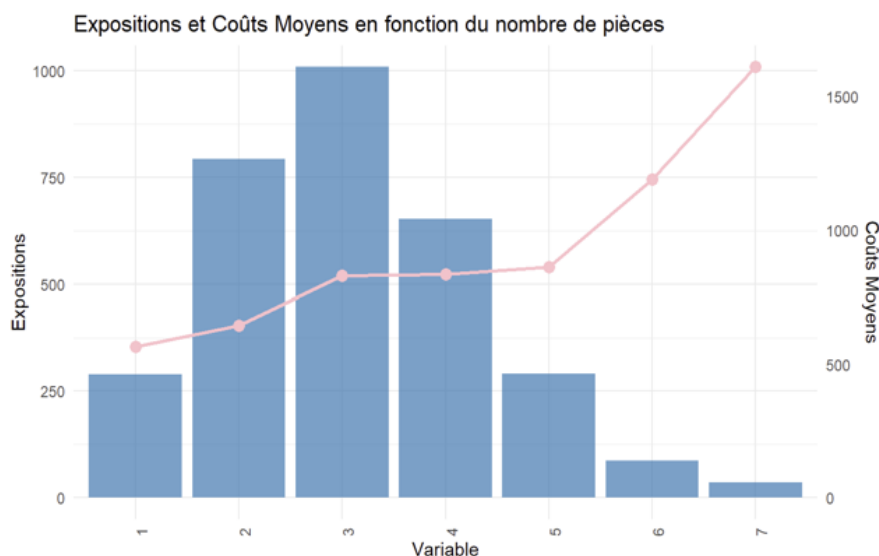


2. Choix à priori des variables explicatives

a. Analyses graphiques

L'analyse graphique des données donne des intuitions sur le pouvoir explicatif des variables tarifaires et permet une **première sélection des variables**.

1. Evolution du coût moyen d'un vol en fonction du nombre de pièces.



Résultat intuitif :

> le coût moyen d'un vol augmente avec le nombre de pièces

> Attention à la fiabilité des observations sur les logements de 6 pièces et + (faible exposition dans le portefeuille)

✓ La variable nombre de pièces semble être significative sur le coût moyen d'un vol

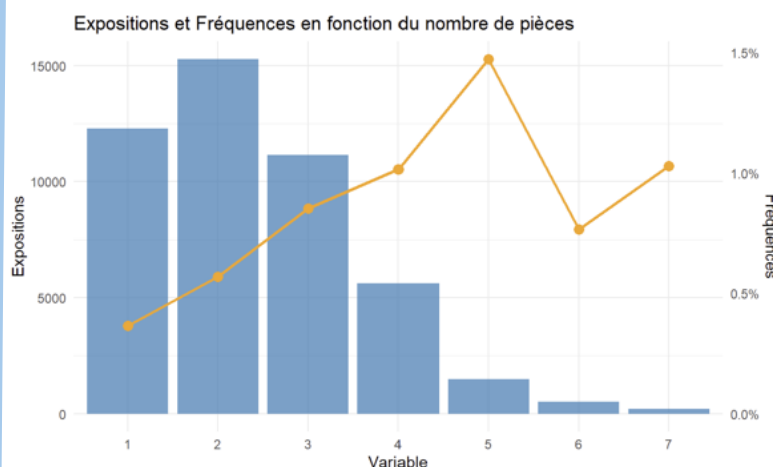


SAVIEL
ACTUARIAT

Les grandes étapes de la tarification

Cas pratique d'une tarification en assurance MRH

2. Evolution de la fréquence d'un vol en fonction du nombre de pièces.



Résultat intuitif :

- La fréquence des vols augmente avec le nombre de pièces
- Même observation que sur le coût moyen pour les logements de 5 pièces et +. Des regroupements et / ou retraitements s'avèrent nécessaires.
- ✓ La variable nombre de pièces semble être aussi une variable significative sur la fréquence d'un vol.

b. Analyse comparative des modèles : méthodes, statistiques et indicateurs décisionnels :

○ Quelques statistiques et indicateurs

Plusieurs tests et indicateurs permettent de juger de la **significativité** des variables explicatives et du **pouvoir prédictif** du modèle. Ils sont disponibles en sortie des GLM sous R ou Python.

Parmi lesquels :

- ☐ Le test de Wald
- ☐ L'écart-type
- ☐ Le critère d'information d'Akaike (AIC), etc



Les grandes étapes de la tarification

Cas pratique d'une tarification en assurance MRH

○ Les méthodes de sélection de variables

Les **méthodes de sélection ascendante / descendante / mixte** sont des méthodes itératives utilisées pour identifier le sous-ensemble de variables explicatives le plus pertinent en se basant sur des critères statistiques tels que l'AIC ou la significativité des coefficients

- > **La méthode ascendante**: ajoute une à une les variables qui améliorent le modèle
- > **La méthode descendante**: part de toutes les variables et retire les moins significatives
- > **La méthode mixte**: combine les deux en ajoutant OU retirant les variables en fonction de leur impact sur le critère choisi

Exemple illustratif de quelques coefficients sur le cout moyen de la garantie Vol

Variable	Valeur des estimateurs β_i
Usage : Résidence secondaire	- 0,20
Usage : Dépendance	- 0,25
Qualité de l'occupant : Locataires	- 0,22
Superficie habitable : (100 ; 150]	+ 0,12
Superficie habitable > 150	+ 0,18

Le tableau ci-dessus présente les résultats des estimateurs β_i obtenus dans le cadre du modèle de coût moyen de la garantie VOL. L'individu de référence (β_0) est un assuré, propriétaire de sa résidence principale comprise entre 50 et 100 m². Nous tirons les conclusions suivantes :

- ❑ le coût moyen est moins élevé pour les résidences secondaires et les dépendances. En général, le capital mobilier est moins important sur ce type de bien que sur une résidence principale.
- ❑ Le coût moyen est moins élevé pour les locataires. En général, les locataires ont des sinistres moins importants que les propriétaires (catégorie socio-professionnelle).
- ❑ Le coût moyen augmente avec la superficie du logement.

CONTACTEZ-NOUS

@ contact@saviel-actuariat.com

9



SAVIEL
ACTUARIAT