

PROJETS 2021 D'ACTUARIAT DE L'ASSURANCE NON-VIE

Consignes

Le projet à réaliser en binôme consiste à répondre a minima à la liste des étapes mentionnées dans le sujet détaillé ci-dessous. Chaque binôme devra rendre plusieurs fichiers :

1. un rapport (**pdf**) synthétique et organisé présentant la méthodologie utilisée et les résultats obtenus de **4 pages** (hors annexe). Les sorties R doivent être clairement choisies tandis que les figures doivent être mis en annexe. Aucune démonstration n'est demandée.
2. un **fichier R principal** appelant le code utilisé dans le rapport et commenté brièvement.
3. un ou plusieurs fichiers R pour les fonctions créées pour le projet et commenté(s) brièvement.
4. un fichier csv **TARIFPUR.csv*** avec votre tarif en prime pure.
5. un fichier csv **TARIFCOM.csv*** avec votre tarif en prime pure.
6. éventuellement des fichiers csv complémentaires utiles pour le rapport.

L'ensemble doit être **zippé** dans un fichier nommé selon la convention **NUMXXYY.zip**, avec **XX/XX** les deux premières lettres de vos (premiers) noms de famille, par exemple **01DUBA.zip** pour le binome DUTANG/BARADEL.

Le travail est à envoyer **au plus tard** pour le **10 janvier 2022 à 23h59** par mail à l'adresse dutang@ceremade.dauphine.fr.

Tarification en assurance IARD

L'assurance est un contrat par lequel, moyennant le versement d'une prime dont le montant est fixé a priori (en début de période de couverture), l'assureur s'engage à indemniser l'assuré pendant toute la période de couverture (disons un an). Cette prime est censée refléter le risque associé au contrat (on peut renvoyer à Denuit & Charpentier (2004) sur la théorie du calcul des primes, et à Denuit & Charpentier (2005) pour les considérations économiques).

Pour chaque police d'assurance, la prime est une fonction de variables dites tarifaires (permettant de segmenter la population en fonction de son risque). Le but de ce projet est de proposer un tarif en se basant sur l'approche fréquence – sévérité et les modèles de régression vus en cours. Au moins deux modèles doivent être testés.

Vos données sont composées d'une base de 100 000 polices ***-PG_2017_YEAR0.csv** et d'une base de sinistres associés ***-PG_2017_CLAIMS_YEAR0.csv** qui vous permettront de calibrer vos modèles. La base de validation sera aussi de 100 000 polices ***-PG_2017_CLAIMS_YEAR1.csv**. Les données sont issues du 3ème jeu de tarification actuariel : elles sont décrites en détails dans le document **3rdPricingGame.pdf**.

*Merci d'UTILISER la fonction **write.csv** pour obtenir un format csv anglais.

Les étapes suivantes devront être abordées

1. Explorez les données `*YEAR0.csv`.
2. Présenter la méthodologie utilisée (lois calibrées, variables retenues, adéquation, ...).
3. Présenter les résultats et conclure sur le modèle le plus approprié.
4. Donnez votre tarif **en prime pure** pour les polices des données de validation sous la forme d'un fichier `TARIFPUR.csv`* composé de **QUATRE COLONNES** : `id_client`, `id_vehicle`, `id_policy` et `premium`.
5. Simuler la charge sinistre globale du portefeuille de validation sur 10^4 scénarios. Ensuite, réaliser et commenter l'histogramme et la fonction de répartition empirique de cette charge.
6. Construire votre tarif **en prime commerciale** incluant une marge pour risque (permettant de couvrir la charge de portefeuille dans **95%** des cas) mais **SANS FRAIS NI TAXE** (on précisera la méthodologie choisie). Fournir pour les polices des données de validation sous la forme d'un fichier `TARIFCOM.csv`* composé de **QUATRE COLONNES** : `id_client`, `id_vehicle`, `id_policy` et `premium`.

Conseils

- citer vos sources bibliographiques de manière très précise,
- faire des graphiques lisibles,
- faire de jolis tableaux pour les quantités numériques (paramètres, indicateurs, ...) plutôt que de simples sorties R, par exemple `xtable(coef(summary(monmodel)), digits=3)` produira un joli tableau $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$,
- si besoin, utiliser le format Rmarkdown (de Rstudio) pour mélanger textes, codes et équations, voir <http://rmarkdown.rstudio.com/>,
- pour les équations, utilisez latex ou mathtype pour word 2007.

A NE PAS faire

- plagier (i.e. recopier) des sections entières d'un document au mot près voire à la lettre près,
- ne pas citer ces sources,
- oublier de justifier les paragraphes.

Références autres que le polycopié de cours :

1. Denuit & Charpentier (2004, 2005), Mathématiques de l'Assurance non-vie, Tome 1 et 2.
2. Goulet (2017), Programmer avec R, <https://vigou3.github.io/programmer-avec-r/>.
3. Charpentier (2014), Computational Actuarial Science with R.
4. Charpentier (2017), cours ENSAE, <https://freakonometrics.hypotheses.org/category/courses/ensae>.

*Merci d'UTILISER la fonction `write.csv` pour obtenir un format csv anglais.