

TP: PRÉPARATION DES DONNÉES EN VUE D'UNE TARIFICATION

On travaillera avec le jeu de données *freMTPL*, contenu dans la librairie **R** **CASdatasets**. Ce jeu de données contient deux sous-jeux de données à charger, *freMTPLfreq* et *freMTPLsev*.

- (1) Importer et charger les jeux de données de la base contrat et de la base sinistre. Regarder l'aide de ces jeux de données afin de les comprendre.
- (2) Fusionner ces 2 bases à l'aide de la fonction `merge`, en sommant les sinistres d'un même contrat lorsque cela est nécessaire (avec la fonction `aggregate` par exemple).
- (3) Visualiser la base de données obtenue. Changer certains types de données si nécessaire. Supprimer des données aberrantes (exemple : exposition anormale).
- (4) Traitement de la sinistralité CAT : définir un seuil de grave correspondant au quantile à 95% de la sinistralité. En déduire une base de données des sinistres attritionnels et une base pour les graves, ainsi que la surprime liée aux événements graves. Visualiser les densités empiriques de la sinistralité dans chacune de ces bases. Présentent-elles plusieurs modes ? Quelle serait l'explication ? Quelle est la part des sinistres graves dans la sinistralité globale ?
- (5) Faire des statistiques descriptives univariées et bivariées sur les variables explicatives et la réponse (indicateurs de localisation, de dispersion, de corrélation, ...). Par exemple,
 - sévérité globale moyenne des contrats sinistrés et non-sinistrés (y compris CAT)
 - sévérité attritionnelle moy. des contrats sinistrés/non-sinistrés (lien prime CAT)
 - sévérité attritionnelle moyenne des contrats sinistres uniquement
 - sévérité globale moy. des contrats sinistrés uniquement (y compris CAT)
 - distributions des variables explicatives
 - analyse bivée des liens entre sévérité attritionnelle et variables tarifaires :
 - variables explicatives qualitatives
 - variables explicatives quantitativesRemarque : en fonction de la modélisation choisie : cette étude peut être faite en analysant les liens avec la fréquence au lieu de la sévérité.
- (6) En fonction de l'étude descriptive précédente, procéder
 - à des regroupements de modalités lorsque nécessaire
 - à des créations de classe lorsqu'opportun

- (7) Analyser les corrélations entre les variables présentes dans la base de données.
 - corrélations entre variables qualitatives
 - corrélations entre variables quantitatives
 - corrélations entre variables quantitatives et qualitatives
- (8) Déterminer un profil de référence dans votre population. Quel est-il ? Quelle est la taille de la population qui y est liée ?
- (9) Créer un échantillon d'apprentissage et un échantillon de validation aléatoirement sur la base de données adéquate. En termes de proportion, faire respectivement $(2/3)$ et $(1/3)$.