

Projet - Modélisation de la sévérité et de la fréquence des accidents routiers

Contact: Christophe Dutang (christophe.dutang@univ-grenoble-alpes.fr)

Contexte

Depuis les années 1980, le nombre de tués sur les routes n'a cessé de diminuer, bien que le parc automobile a augmenté jusque dans les années 2000, sous l'effet des avancées technologiques et de la réglementation, voir wikipedia sécurité routière.

Cependant, ces dernières années, le nombre de tués se stabilisent, voir ré-augmentent, voir par exemple le Monde. D'autres facteurs que la vitesse se sont rajoutés, par exemple la drogue, l'alcool, l'usage du téléphone.

Problématiques

L'objet est d'étudier l'accidentologie routière sous plusieurs angles. Tout d'abord, on étudie la sévérité des accidents routiers en construisant un score de sévérité. Une fois un score de sévérité défini par vos soins, déterminer le meilleur modèle linéaire pour prédire ce score en fonction des variables explicatives de la base ONISR. Ensuite, on cherchera à déterminer des regroupements à l'aide de méthode de partitionnement. Il faudra déterminer le nombre de groupes ainsi que leur constitution. Enfin, on déterminera la fréquence des accidents de la route. Pour ce faire, on cherchera à déterminer le traffic pour les accidents étudiés dans la précédente base sur une zone délimitée. On proposera un modèle de fréquence d'accident compte tenu des variables explicatives.

Modélisation

Les modélisations envisagées pour répondre aux problématiques sont

- modèle linéaire:

$$S = X\beta + \epsilon,$$

où S le score d'accident, X la matrice des variables explicatives issues des variables explicatives x_1, \dots, x_p , β les paramètres et ϵ le bruit gaussien.

- méthode de partitionnement en K ensembles de centre c_1, \dots, c_k :

$$\min_{c_1, \dots, c_k} \sum_{k=1}^K \sum_{i, G(i)=k} d(x^{(i)}, c_k),$$

où $d(x^{(i)}, c_k)$ est une distance entre le point $x^{(i)}$ et le centre c_k .

- modèle de fréquence: modèle binomial ou Poisson en tenant compte de l'exposition.

Les applications numériques sont à faire en R de préférence mais Julia peut être utilisé.

Données

Les bases de l'ONISR recueille tout accident corporel provoquant au moins une victime, survient sur une voie (publique ou privée) ouverte à la circulation publique, impliquant au moins un véhicule et n'étant pas provoqué délibérément. Sont donc exclus tous les accidents strictement matériels ainsi que les accidents survenus sur des voies fermées à la circulation publique ou ceux qui n'impliquent aucun véhicule. Parmi les accidents corporels, on distingue

- un accident mortel comporte au moins une personne tuée,
- un accident grave comporte au moins une personne tuée ou un blessé hospitalisé,
- un accident léger ne comporte ni personne tuée ni blessé hospitalisé.

Un accident corporel implique un certain nombre d'usagers :

- les usagers indemnes : usagers impliqués non décédés et dont l'état ne nécessite aucun soin médical,
- les victimes : usagers impliqués non indemnes.

Parmi les victimes ou les blessés, on distingue :

- les personnes tuées : victimes décédées sur le coup ou dans les 30 jours qui suivent l'accident,
- les blessés : victimes non décédées, dont l'état nécessite des soins médicaux (c'est-à-dire prodigués par un professionnel de la santé, quel que soit le cadre de ces soins : hôpital, cabinet médical, voire sur place)
- les blessés hospitalisés : blessés dont l'état nécessite plus de 24 heures d'hospitalisation,
- les blessés légers : blessés dont l'état nécessite un soin médical mais qui, en cas d'hospitalisation, ne sont pas hospitalisés plus de 24 heures.

Chaque groupe travaillera sur une année d'accident différente.

Références bibliographiques

- Cours de Probabilité & Statistiques 1, 1A
- Probabilités, analyse des données et statistique, G. Saporta, 2011
- Practical Guide To Cluster Analysis in R, A. Kassambara, 2017
- Geocomputation with R, R. Lovelace, J. Nowosad, J. Muenchow, 2025, site