

# Modèle linéaire généralisé et Choix de modèles

## Devoir maison obligatoire

Robin Ryder et Vincent Rivoirard

Vous traiterez les deux sujets.

Rapport à envoyer par courriel à Robin Ryder ([ryder@ceremade.dauphine.fr](mailto:ryder@ceremade.dauphine.fr)) pour le 30 octobre.

### 1 Survenance d'accidents automobiles

Le fichier `baseassuranceauto.csv` contient des données d'assurance automobile. Vous allez modéliser la survenance d'accidents : au choix, la colonne `Surv1` ou la colonne `Surv2`. Les colonnes correspondent à :

- `PolNum` : Numéro de police
- `CalYear` : Année calendaire de souscription
- `Gender` : Genre du conducteur
- `Type` : Type de véhicule
- `Category` : Catégorie de véhicule
- `Occupation` : Profession
- `Age` : Âge du conducteur
- `Group1` : Groupe du véhicule
- `Bonus` : Bonus/Malus
- `Poldur` : Ancienneté du contrat
- `Value` : Valeur du véhicule
- `Adind` : Indicateur d'une garantie dommages
- `SubGroup2` : Sous-région d'habitation
- `Group2` : Région d'habitation
- `Density` : Densité de population
- `Expdays` : Exposition (en jours)
- `Nb1` : Nombre de sinistres RC (responsabilité civile) matériels
- `Nb2` : Nombre de sinistres RC corporels

- **Surv1** : Survenance de sinistres RC matériels (indicatrice 0/1)
- **Surv2** : Survenance de sinistres RC corporels (indicatrice 0/1)

Si vous le souhaitez, vous pouvez également chercher à expliquer les variables **Nb1** et **Nb2**.

Source des données : Arthur Charpentier

## 2 Analyse de données sur la tuberculose

Le fichier **tb\_real.xlsx** contient des données sur le nombre de cas de tuberculose animale dans des fermes canadiennes entre 1985 et 1994. Chacune des 134 lignes correspond à une observation pour un groupe d'animaux. Les colonnes correspondent à :

- **farm\_id** : l'identifiant de la ferme
- **type** : 1=vache laitière ; 2=vache à viande ; 3=cervidé ; 4=autre
- **sex** : 0=femelle ; 1=mâle
- **age** : 0=entre 0 et 12 mois ; 1=entre 12 et 24 mois ; 2=plus de 24 mois
- **reactor** : le nombre de cas de tuberculose dans le groupe
- **par** : nombre d'animaux-jours à risque dans le groupe

Proposez et validez un modèle pour expliquer le nombre de cas de tuberculose.

Source des données : Dohoo, Martin & Stryhn, *Veterinary Epidemiologic Research - 2nd Edition* (2009).