

Modèles de durée et tables de mortalité
Examen du 10 janvier 2024 - Durée : 1h
Les codes R commentés seront récupérés sur clé USB.
L'utilisation d'Internet est interdite durant l'épreuve.

On s'intéresse à la base **pbc** de la librairie **randomForestSRC** de R, que l'on peut charger avec la commande :

```
data(pbc, package = "randomForestSRC")
```

L'objectif est d'étudier le temps de survie en jours (*days*) pour les patients atteints de cirrhose biliaire primitive, à partir des données PBC (Primary Biliary Cirrhosis). La cirrhose biliaire primitive est une maladie chronique du foie rare mais mortelle, de cause inconnue, avec une prévalence d'environ 50 cas par million d'habitants. Entre janvier 1974 et mai 1984, la Mayo Clinic a mené un essai randomisé en double aveugle sur la cirrhose biliaire primitive du foie (CBP), comparant le médicament D-pénicillamine (DPCA) à un placebo.

Pour cette analyse, l'état de la maladie et de la survie en juillet 1986 a été enregistré pour le plus grand nombre possible de patients. À cette date, 125 des 312 patients étaient décédés, dont 11 seulement n'étaient pas attribuables à la CBP. Huit patients avaient été perdus de vue et 19 avaient subi une transplantation hépatique.

1. Considérer les 312 premiers individus de la base et supprimer ensuite de la base les variables présentant au moins une donnée manquante (valeur NA pour au moins un individu).

Combien de variables contient alors la base de données ?

2. Tracer (sur ordinateur) l'estimateur de Kaplan-Meier de la fonction de survie des individus, en prenant en compte les censures éventuelles, en distinguant selon le type de traitement (variable *treatment*). Que remarquez-vous ?

Au seuil de rejet de 5%, doit-on rejeter l'hypothèse : la fonction de survie d'un sujet ne change pas selon le type de traitement ?

3. Tracer (sur ordinateur) l'estimateur de Kaplan-Meier de la fonction de survie des individus, en prenant en compte les censures éventuelles, en distinguant selon le sexe (0 = Homme, 1 = Femme). Que remarquez-vous ?

Au seuil de rejet de 5%, doit-on rejeter l'hypothèse : la fonction de survie d'un sujet ne change pas selon le sexe ?

4. On cherche à expliquer la durée de survie en fonction des variables explicatives par un modèle de Cox.

- Calibrer un modèle de Cox en prenant l'ensemble des variables explicatives.

- Déterminer pour ce modèle la probabilité de survivre au moins 400 jours pour un individu ayant les mêmes caractéristiques que le premier individu de la base.

5. Proposer un modèle de Cox en conservant seulement 7 variables explicatives, qui vous semblent les plus pertinentes.

- Pensez-vous que ce modèle est adapté ? On pourra vérifier l'hypothèse de hasard proportionnel pour chaque variable explicative retenue.

- Déterminer pour ce modèle la probabilité de survivre au moins 400 jours pour un individu ayant les mêmes caractéristiques que le premier individu de la base.
6. On cherche à expliquer la durée de survie en fonction des variables explicatives par une forêt aléatoire de survie.
- Calibrer une forêt aléatoire de survie en prenant l'ensemble des variables explicatives.
 - Calculer le C-index pour ce modèle et interpréter le résultat.
 - Etudier l'importance des variables : quelles variables contribuent le plus dans ce modèle ? Comparer avec les variables retenues pour le modèle de Cox.
 - Déterminer pour ce modèle la probabilité de survivre au moins 400 jours pour un individu ayant les mêmes caractéristiques que le premier individu de la base.
 - Comparer graphiquement les fonctions de survie des deux modèles de Cox et de la forêt aléatoire.
 - Analyser les résultats obtenus.