
Vous trouverez le fichier de données *leaving_home.rda* dont vous avez besoin pour ce TP sur la page dédiée à ce cours du site web <http://www.andreberchtold.com>.

Exercice 1 Première analyse du “leaving home”

Le fichier *leaving_home.rda* contient un extrait des données du Panel Suisse des Ménages consacrées au processus de départ des enfants de chez leurs parents pour un échantillon total de taille $n=5560$. Nous disposons des variables suivantes :

- *idpers* : identificateur unique de chaque personne.
- *sexe* : masculin, féminin.
- *age* : âge en années au moment de l’enquête.
- *gr_age* : groupe d’âge au moment de l’enquête (20 ans et moins, 21-30, 31-40, 41-50, 51-60, 61 et plus).
- *formation* : formation en cours (non, oui).
- *csp_pere* : catégorie socio-professionnelle du père (8 catégories de “dirigeant” à “travailleur non-qualifié”).
- *treiman* : échelle de Treiman pour le prestige du métier du père.
- *region* : région de résidence actuelle (région lémanique (VD, VS, GE), espace mittelland (BE, FR, SO, NE, JU), Suisse du nord-ouest (BS, BL, AG), Zurich, Suisse orientale (GL, SH, AR, AI, SG, GR, TG), Suisse centrale (LU, UR, SZ, OW, NW, ZG), Tessin).
- *langue* : langue de l’interview (français, allemand, italien).
- *agdeppar* : âge du départ de chez ses parents en années.
- *c_deppar* : variable événement codée 1 pour ceux qui sont partis et 0 pour ceux qui ne sont pas partis.

1. Construire une courbe de survie (table et graphique) selon l’approche de Kaplan-Meier pour l’événement “partir de chez ses parents”. Calculer ensuite l’âge moyen au départ et commenter les résultats.
2. Construire des courbes séparées pour les répondants interviewés dans chaque langue et comparer les courbes ainsi obtenue au moyen d’un test statistique. Commenter les résultats.
3. Estimer un modèle de Cox en n’utilisant que la variable *gr_age* comme facteur explicatif. Commenter le résultat.
4. Estimer un modèle de Cox en utilisant les variables *sexe*, *gr_age*, *formation* et *treiman*, puis effectuer manuellement une procédure de sélection backward de recherche du meilleur modèle et commenter les résultats. Les données manquantes ont-elles une influence sur vos résultats ? Si oui, comment supprimer cette influence ?
5. En utilisant à nouveau les 4 mêmes variables explicatives que ci-dessus, effectuer une sélection automatique du meilleur modèle en utilisant le critère BIC. Comparez avec le résultat obtenu au point précédent.
6. En utilisant le modèle final obtenu au point précédent, calculez et affichez les courbes de survie suivantes :
 - (a) courbe de survie à la valeur moyenne des variables explicatives
 - (b) courbes de survie séparées pour les hommes et les femmes
 - (c) courbe de survie pour une femme de 47 ans