

Analyse des durées de survie

Projet

On étudie le jeu de données *ovarian*, du package `survival` composé de 26 patientes atteints du cancer des ovaires. Plus particulièrement, on s'intéresse à la durée de survie de ces patients. Pour chaque patient, nous disposons des variables suivantes.

- la variable *futime* : la durée de survie du patient si *fustat* = 1 ou durée de censure si *fustat* = 0. Cette durée est mesurée en jours.
- la variable *fustat* : 1 si patient décédé du cancer des poumons, 0 si patient censuré.
- la variable *resid.ds* : 1 si la patiente est toujours malade, 2 sinon.
- la variable *rx* le groupe de traitement : 1 pour le traitement A, 2 pour le traitement B
- la variable *ecog.ps* : la valeur du score ECOG, 1 est la meilleure valeur.
- la variable *age* : âge du patient (en années).

Dans la suite on note T le temps de survie d'un patient, C sa variable de censure et δ l'indicateur de censure associé. On observe alors un échantillon $(X_i, \delta_i), 1 \leq i \leq 26$, de variables aléatoires indépendantes et de même loi que (X, δ) où

$$X = \min(T, C) \text{ et } \delta = 1_{T \leq C}.$$

On note également $Z^{(i)} = (Z_1^{(i)}, \dots, Z_p^{(i)})^T$ le vecteur des p covariables de l'individu i , $1 \leq i \leq 26$.

I. Estimateur de Kaplan-Meier et tests de comparaison

1. Quel est le pourcentage d'observations censurées dans le jeu de données ?
2. Quelle quantité est estimée par l'estimateur de Kaplan-Meier ? Commenter le graphique représentant l'estimateur de Kaplan-Meier pour tous les groupes confondus. On donnera notamment les estimations des quartiles de la loi de T (c'est à dire les quantiles d'ordre 25%, 50% et 75%).
3. A combien estimez-vous la probabilité qu'une patiente atteinte du cancer des ovaires vive moins de 200 jours ? Qu'elle vive plus de 600 jours ?
4. On s'intéresse dans un premier temps à l'influence du traitement sur la survie des patients. D'un point de vue descriptif vous semble-t-il que le traitement permette d'augmenter la durée de vie des patients ?
5. Quel test statistique proposez-vous pour comparer la survie des patients en fonction du traitement ? On donnera les hypothèses nulle et alternative. Commenter le résultat du test.
6. On s'intéresse enfin à la survie des patients en fonction de la variable *resid.ds*. Commenter le graphique ainsi que le test associé permettant de comparer la survie des patients en fonction de cette variable. Quel est l'impact de la présence de la maladie sur la durée de vie d'un patient ? Par exemple, au bout de 400 jours, donner une estimation de la probabilité d'être toujours en vie en fonction de la présence au nom de *residus* de maladie.

II. Modèle de Cox

1. On propose maintenant d'expliquer la variable *temps de survie* en fonction de la variable *rx* par un modèle de Cox. Écrire le modèle de Cox correspondant.
2. Donner la définition du rapport des risques (ou hazard ratio) comparant le groupe ayant eu une thérapie antérieure avec le groupe n'ayant pas eu de thérapie antérieure. Donner son interprétation en fonction du (des) paramètre(s) du modèle de Cox. Donner l'estimation du rapport de risque instantané comparant le groupe ayant eu le traitement B par rapport au traitement A. Commenter.
3. Donner les hypothèses de test de l'influence de la variable *rx* sur le temps de survie. Proposer trois tests de l'influence de la variable *rx* sur le temps de survie. Commenter le programme correspondant et les résultats obtenus. Que concluez-vous ?
4. Commenter brièvement les deux modèles de Cox univariés pour les variables *age* et *ecog.ps*. Ces deux variables ont-elles un effet sur la survie des patients ?
5. Écrire le modèle de Cox complet (avec toutes les variables explicatives). Discuter les effets des différentes variables.
6. Proposer une méthode de sélection de variable pour le modèle de Cox. Quelles sont les variables sélectionnées par ce critère ?
7. Quelles sont les hypothèses de modélisation que doit vérifier un modèle de Cox ? Détaillez votre réponse.
8. Implémenter sous R cette validation pour le modèle final. Commenter.