

TP1 : Estimateur de Kaplan-Meier et tests du log-rang

Exercice 1 (*R*)

On s'intéresse au temps de survie de patients atteints de leucémie myéloïde aiguë. Les données se trouvent sous R, en tapant `aml` après avoir chargé la librairie `survival`. Le jeu de données comprend les informations suivantes :

- `time` : temps de survie
- `status` : 0 pour une observation censurée et 1 sinon
- `x` : maintien d'un traitement par chimiothérapie.

1. Après avoir chargé la librairie `survival`, récupérer le jeu de données `aml`, attacher les informations concernant ce jeu de données et regarder les premières observations contenues dans ce jeu de données.
2. Créer un objet de type `Surv` sous R à partir de cette base. Pour cela, on utilisera la fonction `Surv` avec la syntaxe `base=Surv(time,status==1)`. Les données censurées sont marquées par un `+`.
3. Calculer l'estimateur de Kaplan Meier de la survie globale, puis ceux des deux groupes décrits par la variable `x` maintien du traitement par chimiothérapie. On pourra utiliser la fonction `survfit` de R. Calculer les quartiles de la survie globale.
4. Représenter sur un même graphique ces deux courbes de survie en indiquant par un `+` les dates où ont lieu les censures. Tracer les intervalles de confiance en rajoutant l'option `conf.int`. Comparer graphiquement les deux courbes de survie.
5. On cherche à mettre en évidence une éventuelle différence de survie entre les deux groupes décrits par la variable `x`. Calculer les paramètres (moyenne, médiane, quartiles) dans les deux groupes. Comparer avec les résultats sur la population globale. les deux groupes". On précisera le (les) test(s) utilisé(s), la statistique de test, sa loi sous l'hypothèse nulle, sa zone de rejet. Mettre en oeuvre le(s) test(s) avec R en utilisant la fonction `survdif`. On utilisera l'option `rho` de la fonction `survdif`.
6. Proposer deux estimateurs du risque cumulé pour la population globale. Les tracer.
7. Sur un même graphique, tracer l'estimateur du risque cumulé pour chaque groupe décrit par la variable "maintien du traitement par chimiothérapie" (on choisira un des deux estimateurs possibles du risque cumulé).

Exercice 2 (SAS)

On utilise la base de données `myel_comp.txt` qui porte sur le myélome.

Ce cancer est caractérisé par une prolifération maligne de plasmocytes dans la moelle osseuse. Le taux de protéinurie de Bence-Jones renseigne sur le stade de la maladie et permet de suivre l'évolution. La chimiothérapie est le principal traitement.

1. Importer la base de données. On se focalisera dans ce TP sur les variables Age, Sexe, calcium et indice de Bence-Jones.
2. Représentation des données.
 - (a) Calculer les statistiques descriptives des différentes variables.
 - (b) Faire une représentation graphique de l'âge en fonction du sexe et de l'indice Bence Jones.
 - (c) Comparer la moyenne d'âge entre les hommes et les femmes.
 - (d) L'âge moyen est-il le même quelque soit l'indice de Bence-Jones ?
 - (e) Etudier le lien entre les variables sexe et Bence-Jones.
 - (f) Comparer la moyenne d'âge entre les différents niveaux de la consommation de calcium.
3. Analyse de survie
 - (a) Calculer la moyenne de survie des patients atteints de myélome, la médiane de survie, les quartiles. Donner l'intervalle de confiance. Que peut-on dire de la distribution de la variable de survie de ces patients ?
 - (b) On cherche à mettre en évidence une éventuelle différence de survie dans le groupe *protéinurie de Bence-Jones*. Calculer les paramètres (moyenne, médiane, quartiles) dans les deux groupes avec ou sans protéinurie de Bence-Jones. Comparer aux paramètres calculés en 1) sur la population globale.
 - (c) Comparer graphiquement les courbes de survie dans les 2 groupes (avec ou sans protéinurie de Bence-Jones).
 - (d) Mettre en place un test de l'hypothèse "Les courbes de survie sont égales dans les 2 groupes" (Indiquer quels tests peuvent être mis en place)
 - (e) On cherche maintenant à mettre en évidence une éventuelle association entre la survie et la variable quantitative CALCIUM. Mettre en place un test de l'hypothèse " La survie n'est pas liée à la consommation de calcium". (deux façons peuvent être envisagées, soit en stratifiant la variable CALCIUM et en testant l'égalité des survies sur les strates, ou bien directement en testant l'association de la variable CALCIUM avec la variable de survie).
 - (f) Etudier l'effet sur la survie des autres variables qualitatives.
 - (g) Comment peut-on étudier l'effet sur la survie de variables quantitatives ?