

TP1 - Modélisation non-paramétrique

Exercice 1

1. Calculer « à la main » l'estimateur de Kaplan-Meier et les fonctions de survie des données ci-dessous dans les groupes A et B. ('+' désigne que l'évènement n'a pu être observé).
2. Représenter sur le même graphique les deux courbes de survie en indiquant par un '+' les dates où ont lieu les censures.
3. Donner une estimation de la médiane et du premier quartile dans chaque groupe.
4. Refaire le travail sur R en utilisant la librairie `survival`.
5. Faire le test du log-rang « à la main » pour comparer les deux groupes.
6. Refaire le test du log-rang avec la librairie `survival`.

Individu	1	2	3	4	5	6	7
Groupe A	3	5	7	9	18	18	20
Individu	8	9	10	11	12	13	14
Groupe B	12	33+	19	20	20+	25+	26

Exercice 2

On étudie la base de données `divorce.csv`. Cette base contient des suivis de couples mariés aux États-Unis. La variable d'intérêt est le temps écoulé depuis le mariage jusqu'à la date de divorce en années. La variable `event` indique si le couple a divorcé (égal à 1) ou si le couple n'a jamais divorcé (égal à 0). La variable `time` représente la dernière date à laquelle les couples ont été suivis et pour laquelle on sait s'ils ont divorcé (`event = 1`) ou s'ils sont toujours mariés à cette date (`event = 0`). On étudie également la variable `fac_heduc` qui représente le nombre d'années d'éducation du mari (trois modalités possibles : moins de 12 ans, entre 12 et 15 ans, 16 ans ou plus).

1. Tracer l'estimateur de Kaplan-Meier pour estimer la fonction de survie du temps écoulé jusqu'au divorce. Donner l'estimation de la probabilité qu'un couple soit toujours marié après 10 ans, puis après 20 ans. Donner également les intervalles de confiance.
2. Donner l'estimation du premier quartile avec intervalles de confiance.
3. Estimer la fonction de survie du temps écoulé jusqu'au divorce en fonction de l'éducation du père. Tracer les trois courbes. Comparer les probabilités qu'un couple soit toujours marié après 10 ans et après 20 ans pour chacun des trois groupes. Comparer l'estimation du premier quartile entre les trois groupes.
4. Implémenter le test du log-rang pour évaluer l'effet du nombre d'années d'éducation du père sur le risque de divorce.

Exercice 3

On utilise à présent la base de données `myel_comp.txt`. Ce cancer est caractérisé par une prolifération maligne de plasmocytes dans la moelle osseuse. Le myélome s'exprime par des troubles qui traduisent une maladie « chimique » autant que tumorale : une altération de l'état général (pâleur, asthénie) ; une atteinte rénale : albuminurie, insuffisance rénale ; une atteinte sanguine (anémie, hémorragies cutanées) ; une atteinte nerveuse (compression de la moelle épinière ou de nerfs comme le nerf médian au poignet) ; une infection favorisée par une baisse des immunoglobulines normales ; des signes d'hypercalcémie. Le diagnostic biologique de myélome repose sur les plasmocytes cancéreux démasqués par l'examen au microscope de la moelle osseuse (myélogramme). L'immunoglobuline pathologique est précisée par immunoélectrophorèse et peut être retrouvée dans les urines : on parle alors de protéinurie de Bence-Jones. Elle doit être chiffrée car son taux renseigne sur le stade de la maladie et permet de suivre l'évolution, comme un marqueur. La chimiothérapie est le principal traitement ; on surveille la tolérance du malade et les principaux paramètres du myélome (la douleur, la calcémie, l'anémie, l'insuffisance rénale, les immunoglobulines dans le sang et les urines, les plasmocytes dans la moelle). Malgré d'incontestables progrès, le myélome reste une maladie grave.

1. Importer la base de données (voir le script fourni). On se focalisera dans ce TP sur les variables **Indice de Bence-Jones** et **Sexe**.
2. Analyse de survie :
 - (a) Calculer la médiane de survie des patients atteints de myélome et les quartiles. Donner l'intervalle de confiance.
 - (b) On cherche à mettre en évidence une éventuelle différence de survie dans le groupe protéinurie de *Bence-Jones*. Remplir le tableau ci-dessous. Comparer aux paramètres calculés en 1) sur la population globale.

	Avec protéinurie	Sans protéinurie	Total
N (taille d'éch.)			
Nb événements			
% censures			
Médiane [IC]			
Survie à 12 mois			

- (c) Comparer graphiquement les courbes de survie dans les deux 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 ».
- (e) Refaire la même étude en fonction de la variable **Sexe**.